INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI

A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600

Construction and analysis of a model for the transmembrane domain of the serotonin 2C G-protein coupled receptor

by

Juan Antonio Ballesteros

A dissertation submitted to
the Graduate Faculty in Biomedical Sciences
in partial fulfillment of the requirements for
the degree of Doctor of Philosophy,
The City University of New York

1997

This manuscript has been read and accepted for the Graduate Faculty in Biomedical Sciences in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

9/23/97
Date

Roman Osman, Ph.D.
R. Osman
Chair of Examining Committee

9/25/97
Date

Terry Ann Kulwrich, Ph.D.
Executive Officer

Members of the Examining Committee:
Harel Weinstein, D.Sc.
Stuart Sealfon, M.D.
Jonathan Javitch, M.D.
Charles Hutchins, Ph.D.

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Construction and analysis of a model for the transmembrane domain of the serotonin 2C G-protein coupled receptor

by

Juan Antonio Ballesteros

Advisor: Professor Harel Weinstein, D. Sc.

A model of the transmembrane domain of the serotonin 2C G-protein coupled receptor (GPCR) complexed with its natural ligand, serotonin, has been constructed and analyzed by energetic, experimental data, and functional criteria. Interpretation of a thermodynamic framework for receptor activation in terms of energy levels for the principal states suggests a metastable activated receptor in the absence of G-protein. The modeling procedure was based on the analysis of biophysical properties and their conservation patterns observed in a multiple sequence alignment of neurotransmitter GPCRs. The boundaries and orientations of seven transmembrane α-helices (TMH) were predicted based on the analysis of α-helix periodicity in the conservation pattern of several biophysical properties, which can identify and discriminate between interior-facing and lipid-exposed residues. The helical periodicity of non-conserved Arg/Lys occurring at the cytoplasmic boundaries, predicted to

face the mostly negatively charged phospholipid headgroups, was identified as a novel predictive tool which leads to a redefinition of a TMHs. Integration of results from the various criteria results in TMH boundaries and orientations that include most proposed G-protein coupling domains, and are in agreement with results from NMR, SCAM and Spin labeling experiments. Five organizing motifs characterize the TMH domain: The ligand binding site is conected to an "aromatic cluster" motif, characterized by a proposed "aromatic rotamer switch" in H6 which relates to receptor activation. The aromatic cluster is connected to the H2-H7-H1 cluster of polar residues, defined by involving D2.50 and N7.49 double revertant mutant constructs in the GnRH and 5HT2C receptors (Zhou 1994; Sealfon 1995). Two clusters of charged residues at the cytoplasmic boundaries, the "Arg-Cage" involving H3-H5-H6 and another cluster on H1-H2-H7, comprise the most important G-protein coupling regions. Conformational changes leading to receptor activation are described in terms of these microdomains. Highly conserved Pro-kinks are proposed as conformational hinges involved in the activation mechanism. Molecular Dynamics simulations of the 5HT2CR complex with serotonin produce an energy-refined model, which is shown to be fully consistent with available data and to have yielded specific guides for experimental verifications that validate it.

Acknowledgements

I never though of being a scientist. Yet it is not chance alone that has brought me here. Carmen Martinez was crucial in sustaining my efforts in Barcelona, but the real discovery of Science as I now know it and passionately love it came here in New York. Thus my dearest gratefulness is for those who taught me and introduced me into the realm of Science: My advisor Dr. Harel Weinstein, and the Drs. Roman Osman, Juan Luis Pascual-Ahuir, Chung Wong, George Nemethy and Leonardo Pardo. For a dreamer, there isn't better gift than to introduce him to a world where he can dream freely and get paid and appreciated by doing so. A special acknowledgement for Dr. Weinstein, who suffered most those dreams and wisely filtered them into real science.

Thanks to those who have come later and brought me yet to another level of understanding. Foremost among you are our experimental collaborators, Drs. Stuart Sealfon, Saul Mayanni, Barbara Ebersole, Ted Bargiello, Jonathan Javitch, Ulrik Gether and Brian Kobilka. Further doors have been opened by Massimo Sasaroli, Frank Guarnieri, Nina Pastor, and many others. Thanks to Dr. Charles Hutchins for being in my disertation.

To those who have shared this pathway in Mount Sinai, Nina, Boban, Karel, Liisa, Tona, Kombo, I remember that we were here together. To you Nina, the bond knows no borders. To my friends through these years, thanks for understanding my obsessive dedication. And to my family, you were always there. Isdinen, thank you for being here, please stay.

TABLE OF CONTENTS

1. INTRODUCTION 1

2. METHODS 12
2.1- Chapter Neurosciences 12
2.1.1- Primary structure of GPCRs: Conservation patterns and functional divergence. 17
2.1.1.1- Criteria and inferences in the construction of sequence alignments. 19
2.1.1.2- Analysis of the conservation pattern from the sequence alignments. 24
2.1.1.3- Numbering scheme for GPCR sequences. . .29
2.1.2- Secondary structure of GPCRs: Prediction and modeling of individual transmembrane helices. 32
2.1.2.1- The prediction of TMH boundaries. . . .32
2.1.2.1.1- The hydrophobicity profile as parsing criterion. .34
2.1.2.1.2- The use of Arg and Lys positions to define the cytoplasmic ends. 36
2.1.2.1.3- Prediction of TMH segments from the α-helix periodicity of property profiles measured by Fourier transform analysis. 44
2.1.2.1.4- Prediction of TMH segments from the α-helix periodicity identified from surface patches on a helical net representation. 54
2.1.2.1.5- Analysis of the occurrence of Pro residues within putative TMH segments as an indication of helix ends. .55
2.1.2.1.6- Methods for defining TMH ends based on the statistical probability of occurrence of specific AAs in TMHs.56
2.1.2.2- Modeling of the 3D structure of individual transmembrane helices. 57
2.1.2.3- Secondary structure prediction for the loops, N- and C-terminal domains. 62
2.1.3- Modeling the tertiary structure of the three-dimensional model of a GPCR. 63
2.1.3.1- Prediction of TMH-TMH vs. TMH-Lipid interfaces. .63
2.1.3.1.1- The hydrophobic or conservation moment predicts the relative orientation of helices in the transmembrane bundle. 64
2.1.3.1.2- Patches on the TMH surface identify TMH-TMH versus TMH-lipid interfaces. 69
2.1.3.2- Modeling the tertiary structure of GPCR: Packing the TMHs in the transmembrane bundle. 78

2.1.3.2.1- Inferences for TMH packing from methods used to identify the individual TMHs on the projection map of RH .83
2.1.3.2.2- Modeling TMH-TMH packing on the RH template.84
2.1.3.2.3- Assembly of the TMH bundle according to a BR template.. 96
2.1.3.3- Modeling extramembrane segments of the GPCRs. .100
2.1.4- Probing of GPCR models based on molecular energy criteria.102
2.1.5- Computational approaches to probe structural details and functional mechanisms of GPCR. 107
2.1.5.1- Ligand binding properties as criteria for probing GPCR models 108
2.1.5.2- Criteria based on computational simulation of receptor activation mechanisms. 110
2.1.6- Concluding remarks. 115
2.2- New developments in the literature. 116
2.2.1.- Direct structural determination of GPCRs. . . .121
2.2.2.- Indirect structural determination of GPCRs. . . .125
2.2.2.1- Data pertinent to TMH ends and orientations. . .125
2.2.2.2- Experimentally derived TMH-TMH contacts. . .127
2.2.2.3- Proposed ligand-receptor interaction sites. . .134
2.3- New modeling methods 135
2.3.1.- Primary structure analysis: 135
2.3.1.1.- Quantification of the conservation of volume or hydrophobicity (hdp). 136
2.3.1.2.- Quantification of the conservation of properties not present at every AA. 137
2.3.1.3.- The conservation index .142
2.3.2.- Tertiary structure analysis: . . .147
2.3.2.1.- Quantification of criteria for the prediction of TMH-Lipid interfaces. 148
2.3.2.2.- Prediction of the direction and degree of helix tilt .153
2.3.3.- Energy based methods . . .156

3.- A THERMODYNAMIC FRAMEWORK TO MODEL AND INTERPRET STRUCTURE-FUNCTION RELATIONSHIPS ON GPCRS. . .161

3.1- The ternary complex model: .165
3.2- Interpretation of the ternary complex model in terms of energy levels for the principal states:. 168
3.3- Experimental validation of the thermodynamic scheme .177
3.3.1- Fluorescence studies of β2-adrenergic receptor activation. . .177
3.3.2- Laser-induced optoacoustic studies of rhodopsin activation.183

4.- PRIMARY STRUCTURE OF GPCRS: CONSERVATION PATTERNS AND FUNCTIONAL DIVERGENCE. 185

4.1- Construction of a multiple sequence alignment of GPCRs .186
4.2.- Analysis of the conservation pattern. 195
4.3.--Deviations for the 5HT2C receptor 199

5.- SECONDARY STRUCTURE ANALYSIS: 209

5.1- Initial definition of the 5HT2C TMH boundaries. . . .209
5.2- Analysis of the occurrence of insertions or deletions. . .211
5.3.- The Arg /Lys motif .at the cytoplasmic boundary . .212
5.4- α-helix periodicity measured by Fourier transform analysis. .213
5.5- α-helix periodicity measured by analysis of surface patches .216
5.6- Analysis of the occurrence of non-conserved Pro residues. .218
5.7.- Integration of results from the various criteria. . . .221
5.8.- Consistency with experimental data 227

6.- ORIENTATION OF THE 7 TMHS WITHIN THE TMH DOMAIN .239

6.1.- Prediction of TMH-TMH versus TMH-Lipid interfaces:. . .239
6.1.1.- Prediction of AA in TMH-TMH interfaces. . . .240
6.1.2.- Prediction of AA in TMH-Lipid interfaces. . . .245
6.1.2.1.- Prediction of AA facing the hydrophobic core . .245
6.1.2.2.- Arg/Lys at cytoplasmic boundaries facing the phospholipid head-groups 248
6.1.3.- Alternative positioning of {Pro, Ser, Thr, Cys, Gly} . .249
6.1.4.- Integration of the predictions. 252
6.1.5- Consistency with experimental data. 260
6.2.- A sequential and anticlockwise arrangement of the 7 TMHs based on the rhodopsin map. 262
6.2.1.- The prediction of a tertiary environment for AA positions is determining the shape of the TMH domain.. . . .263
6.2.2- Derivation of a sequential and anticlockwise packing. .264
6.2.3- Classification of interior-facing residues into sequential TMH-TMH interfaces 267
6.3.- Prediction of the direction and degree of helix tilt. . .268

7.-MODELING THE TERTIARY STRUCTURE OF THE THREE-DIMENSIONAL MODEL OF THE 5-HT_{2C} GPCR. 279

7.1- Consistency among experimentally derived interactions . .284
7.2- The effect of the Pro-kinks (PK) in H2, H5, H6, and H7: . .286
7.2.1- Extent of the conformational freedom induced by PKs:. .287

7.2.2- The conformation of the PK in H7: 290
7.2.3- The conformation of the PK in H5: . . . 292
7.2.4- The conformation of the PK in H6: 298
7.2.5- Modulation of the H5-H6 PKs in other GPCRs: . .302
7.3.- Organizing motifs within the TMH domain. . . .308
7.3.1.- The H2-H7-H1 cluster of polar residues 309
7.3.2.- Charge-charge interactions at the cytoplasmic side: . .310
7.3.2.1.- The Arg-cage 312
7.3.3.- The aromatic cluster (H6-H5-H4) 333
7.3.3.1- The aromatic rotamer switch in H6 . . .334
7.3.3.2.- A conformationally constrained Trp in H4: . .337
7.4.- Optimization of helix-helix packing for the TMH bundle, including serotonin and LSD as prototype ligands: 338
7.4.1.- Packing H3-H6 at the cytolasmic boundaries: . . .340
7.4.2.- Packing H5 into the H3-H6 cytoplasmic template: . .346
7.4.3.- Modeling the extracellular segments of H3-H5-H6: . .349
7.4.4.- Packing H4 onto the H3-H5-H6 template: . . .354
7.4.5.- Incorporating the ligands in the H3-H4-H5-H6 template. .364
7.4.6.- Incorporation of the extracellular segment of H7, comprising residues 7.35-7.46, into the H3-H4-H5-H6 template. .375
7.4.7.- Incorporation of H2 into the H3-H4-H5-H6-H7ec template .378
7.4.8.- Modeling the NPxVY motif in H7 into the H1-H2-H3-H4-H5-H6-H7(7.35-7.46) template. .383
7.4.9.- Incorporation of H1 into H2-H3-H4-H5-H6-H7(7.35-7.46). .390

8.- EVALUATION OF THE PROPOSED 3D MODEL OF THE TMH DOMAIN OF THE 5HT2C RECEPTOR BY ENERGETIC, EXPERIMENTAL DATA AND FUNCTIONAL CRITERIA. 427

8.1- Evaluation of the 5HT2CR model by energy-based criteria. .427
8.2.- Summary of experimentally validated criteria . . .435
8.3- Functional evaluation of the proposed 5HT2CR model . .441
8.3.1.- Ligand binding: 5HT2C/2A binding specificity. . .442
8.3.2.- The G-protein coupling domain: The H7 extension. .443
8.3.3.- The receptor activation mechanism: 447

9.- INFERENCES FROM THE MODEL 462

9.1- Modulation of the conformation of the aromatic cluster on H6 by incorporation of β-branched residues at positions i-4, I+4. . .462
9.2- Modulation of the H6 PK conformation by 5HT2 specific H-bonds 464
9.3- Arg-cage expanded: Role of E6.30 holding R3.50 in the inactive form 465

9.4- Newly proposed ligand-receptor interactions from the model. .466
9.5- Engineering an all-activated receptor: 467
10.- CONCLUSION 471

REFERENCES. 475

LIST OF TABLES

Table 2.1 Generalized numbering scheme for GPCRs.....30
Table 2.2 Predictions of the hdp plot versus observed TMH ends BR.....42
Table 2.3 Experimentally derived TMH-TMH contacts.....127
Table 2.4 H-bonding capabilities according to side chain position.....140
Table 2.5 Multiple Sequence Alignment PRC.....150
Table 4.1 Multiple Sequence Alignment GPCR.....187
Table 4.2 Conservation of AA properties by % of AA type.....196
Table 5.1 Predicted versus observed GPCR TMH boundaries 231
Table 6.1 Occurrence of {Ser, Thr, Cys} residues in TMHs of GPCRs.....251
Table 7.1 Erms among residues modulating the PK of H6.....303
Table 7.2 Rotamer populations of Arg3.50 in WT and D3.49 mutants.....317
Table 7.3 Rotamer populations of Arg3.50 in WT and I3.54 mutants.....320
Table 7.4 Binding and activation GnRHR Arg-Cage WT and mutants.....322
Table 7.5 Effect of W6.48 X1 rotamer on X1 rotamer of F6.44 and F6.52.....335
Table 7.6 Erms between H4 and H5.....356
Table 7.7 Erms of Glu3.41 in β2-adrenergic receptors.....380
Table 8.1 Proposed TMH-TMH contacs versus MD simulations.....437

LIST OF FIGURES

Figure 2.1 Global conservation at each site of the TMH3 segment.....26
Figure 2.2 Generalized numbering scheme illustrated in a helical net 31
Figure 2.3 Hydrophobicity profile for TMH6 of the human β_2-adrenergic receptor.....31
Figure 2.4 The seven TMH bundle of BR.....37
Figure 2.5 Prediction of TMH6 boundaries by the Arg/Lys cytoplasmic motif.....39
Figure 2.6 Emerging picture of the polarity pattern for a prototype TMH from the inclusion of the phospholipid head-group regions.....42
Figure 2.7 TMH 6 and 7 of BR.....43
Figure 2.8 α-Helix periodicity of the variability profile measured by Fourier transform analysis.....47
Figure 2.9 The helical net representation.....50
Figure 2.10 Non-local deviations from an ideal α-helix induced by a Pro-kink.....51
Figure 2.11 Directionality of the hydrophobic and conservation moments relative to helix-helix or helix-lipid interfaces.....68
Figure 2.12 Prediction of TMH-TMH versus TMH-Lipid interfaces.....76
Figure 2.13 Our model of the gonadotropin releasing hormone receptor 82
Figure 2.14 Predicted degree of surface exposure for each GPCR TMH.....98
Figure 2.15 Differences in the definitions and most appropriate uses of computational approaches to molecular modeling and computational simulation.....113
Figure 2.16 Experimentally derived THM-TMH interactions.....129
Figure 2.17 Polytopes conservation index compared to the number of different AA at a given site.....146
Figure 2.18 Satisfactory prediction of AA facing the lipid chains using the P(TMH-lipid) function.....152
Figure 2.19 Testing the discriminative power of the probability function P(TMH-lipid).....152
Figure 2.20 Helix tilting may be expressed by 4 components.....154
Figure 2.21 Helix tilting may be recognized in helical net representations.155
Figure 3.1 The two-state (or ternary complex) model of receptor activation 168
Figure 3.2 The two-state model of receptor activation in an energetic context.....170
Figure 4.1 TMH conservation index, conservation of volume and hydrophobicity.....202
Figure 5.1 Predicted Hx boundaries for the 5HT-1c GPCR.....233
Figure 5.2 Prediction of Hx boundaries by the presence of insertions or deletions within the alignment.....233

Figure 5.3 Prediction of Hx boundaries by the Arg/Lys cytoplasmic motif 234
Figure 5.4 Hx nets representation of the conserved AA sites by the conservation index....235
Figure 5.5 Hx nets representation of the conserved AA sites in terms of the volume....235
Figure 5.6 Hx nets representation of the conserved AA sites in terms of the hydrophobicity....236
Figure 5.7 Hx nets representation of the conserved AA sites in terms of the H-bonding capabilities....236
Figure 5.8 Hx nets representation of the conserved AA sites in terms of the aromatic character....237
Figure 5.9 Prediction of Hx boundaries by the presence of Pro residues within the alignment....237
Figure 5.10 Consistency of the predicted TMH boundaries and orientations for the 5HT2CR with experimental data....238
Figure 5.11 The effect of the orientation of the Cα-Cβ vector towards the N-terminus of an α-helix on the solvent accessibility....238
Figure 6.1 Prediction of AA sites in THM-TMH interfaces....244
Figure 6.2 The probability function P(TMH-Lipid) predicts as lipid-facing residues those above the threshold value of 6....246
Figure 6.3 Predicted AA sites in TMH-Lipid interfaces....271
Figure 6.4 Occurrence of conserved Pro, Ser/Thr/Cys, and Gly residues within the TMH domain of neurotransmitter GPCRs....272
Figure 6.5 Juxtaposition of the predictions of a tertiary structure environment derived from three applied criteria....273
Figure 6.6 Final prediction of Hx-Hx of Hx-Lipid environments for AA sites....274
Figure 6.7 Continuous TMH-TMH versus TMH-Lipid interfaces....275
Figure 6.8 The extent of azimuthal surface predicted to be exposed to the lipids in each TMH....276
Figure 6.9 Classification of interior-facing residues into sequential TMH-TMH interfaces....277
Figure 6.10 Predicting the TMH tilt for the 5HT2CR....278
Figure 7.1 100 representative conformations of the Pro-kinks for H6 and H7 obtained by Monte-Carlo simulations....398
Figure 7.2 Conformation of the PK in H7 comprising the conserved NP motif....399
Figure 7.3 Helical net comparison of our predictions of the TMH-TMH interfaces and the experimentally observed accessibility to MTSEA by SCAM techniques on the D2 receptor....399
Figure 7.4 Modeled H7 is in agreement with the SCAM data....400
Figure 7.5 Face-twist observed for the PK in H5....401
Figure 7.6 Comparison of the bending induced in the PK of H6 by the

presence of C6.47 402
Figure 7.7 Proposed interactions between the NPxVY motif in H7 403
Figure 7.8 The conservation pattern of this motif is presented in an alignment of several GPCRs 404
Figure 7.9 Spatial proximity between the conserved Arg and the conserved residues surrounding this Arg in TMD3 405
Figure 7.10 Wild type $R^{3.50(139)}$ orientations 406
Figure 7.11 $R^{3.50(139)}$ orientations in the $D^{3.49(138)}N$ mutant 407
Figure 7.12 $R^{3.50(139)}$ orientations in the $I^{3.54(143)}A$ mutant 408
Figure 7.13 Three-dimensional model of TMD2, 3, and 7 of the GnRH receptor 409
Figure 7.14 Proposed role of $I^{3.54(143)}$ in "caging" $R^{3.50(139)}$ through a steric clash 410
Figure 7.15 Alternative conformations of the aromatic cluster in H6 411
Figure 7.16 The Arg-cage motif at the cytoplasmic boundaries of H3-H5-H6 412
Figure 7.17 H3-H5-H6 packing at the extracellular side 413
Figure 7.18 The most salient features of interior facing residues in H4 414
Figure 7.19 Bend in a Hx induced by SC χ1-g- observed in PDB structures 415
Figure 7.20 Alternative H3 orientations induced by Ser/Thr/Cys 416
Figure 7.21 Packing of H4 into the H3-H5-H6 template 417
Figure 7.22 Proposed "pharmacophore" for the 5HT2CR 418
Figure 7.23 Proposed agonist-receptor interactions 419
Figure 7.24 Proposed spatial relationship between 5-HT and LSD 420
Figure 7.25 Packing the extracellular segment of H7 into H3-H4-H5-H6 421
Figure 7.26 The **erm** sites correlated with E3.41 422
Figure 7.27 H2 packing into the H3-H4-H5-H6-H7(7.35-7.46) template 423
Figure 7.28 Modeling the NPxVY motif in TMH template 424
Figure 7.29 Incorporation of H1 into the H2-H3-H4-H5-H6-H7 template. 425
Figure 7.30 Packing of H1 with H2-H7 at the extracellular side 426
Figure 8.1 Rms deviations observed in the MD simulation 453
Figure 8.2 Reorientation of the TMHs by the MD simulation 454
Figure 8.3 Solvent accessibility for the energy-refined model 455
Figure 8.4 5HT-receptor H-bonding interactions from the simulation 456
Figure 8.5 Dynamic evolution of 5HT-receptor H-bonding contacts 457
Figure 8.6 Organizing principles of the 5HT2CR TMH domain 458
Figure 8.7 Interactions of LSD and Mesulergine with Ala/Ser5.46 459
Figure 8.8 Solvent accessibility of G-protein coupling domains 460
Figure 8.9 Hypothetical role of the cytoplasmic extension of H7 461
Figure 9.1 Inactive conformation of the cytoplasmic extension of H7 470

1. INTRODUCTION

The important pharmacological role of G-protein coupled receptors (Ballesteros 1995) as therapeutic targets has driven an increasing interest in their structure as a means for understanding their functional properties. However, the difficulty in determining their structure by experimental techniques has hindered this analysis. The interest in complementing the structure-activity analysis at the level of the ligands with a structural representation of the receptor level has lead to an increasing interest in the use of molecular modeling and simulation techniques (Ballesteros 1995). Such techniques have been applied in this work to provide a three-dimensional model of the transmembrane domain of the serotonin 2C receptor (5HT2CR) and to probe its structural and functional implications.

At the start of this thesis, in 1989, the structural data on neurotransmitter GPCRs available to guide the modeling process was almost nil. The primary sequences of these GPCRs were emerging at a rapid pace, and site directed mutagenesis studies were just starting (Dixon 1988; Lefkowitz 1988). There was only one transmembrane protein structure known at high resolution, the Photosynthetic Reaction Center (PRC) (Rees 1989b). The theoretical methods available to predict the structure of a membrane protein were essentially the hydrophobicity plot (Kyte 1982) and the hydrophobicity moment (Eisenberg 1984). These methods predicted consistently 7

transmembrane α-helices (TMH) for GPCR s. Given these shortcomings, the 9 A resolution structure of Bacteriorhodopsin (BR) published in 1990 by electron cryo-microscopy (Henderson 1990) led to the acceptance of this 7 transmembrane structure as a template for modeling GPCRs (see (Ballesteros 1995) for a review). This assumption was based on the similarity between BR and RH, both of which contain a retinal chromophore and 7 TMHs. However, based on the divergent sequence between these proteins, we proposed that BR was not an appropriate template to model GPCRs (Pardo 1992), a contention later validated by the structure of RH obtained by similar techniques (Schertler 1993; Schertler 1995; Davies 1996).

In the absence of sequence homology with a known structural template and of appropriate modeling methods, we focused our interest on analyzing known structures in the search for structural motifs that could be used as guidelines to model GPCRs. Analysis of the only known TMH protein (the PRC), lead to the discovery of the role of Arg/Lys motif at the cytoplasmic boundaries facing the lipids at the level of the phospholipid headgroups (see Section 2.1). This motif, later confirmed in all known TMH structures, led to a redefinition of transmembrane segment boundaries (Ballesteros 1995). and could be used to refine the prediction of TMH ends of GPCRs (Ballesteros 1995). The TMH ends defined in this work for the 5HT2CR according to such criteria lead to the inclusion of most G-protein coupling domains inside the TMH domain, an inference that was validated recently by experimental data

(Farahbakhsh 1995; Altenbach 1996). Another structural motif analyzed in known structures that proved to be important for modeling the TMH domain is the Pro-kink (Barlow 1988; Williams 1991; Sankararamakrishnan 1992; Ballesteros 1995), for which we discovered the presence of a "face-twist" induced in the TMH structure (Ballesteros 1995).

To guide the modeling of secondary and tertiary structures of GPCRs, we focused our analyses first on the extensive primary structure data available. These analyses involve the construction of a multiple sequence alignment of GPCRs where the conservation pattern can be analyzed in detail (Chapter 4). Analyses of the resulting conservation patterns have led to specific predictions at each structural level, described in Chapter 5 (2^{ry} structure), Chapter 5 (3^{ry} structure), and Chapter 5 (three-dimensional model).

The methods available at the time to derive structural inferences from the analysis of the conservation pattern were few and inconclusive, as indicated in our published review (Ballesteros 1995) presented in Methods (Section 2.1). Thus, new predictive tools were developed through the course of this work. As will be described in Methods (Section 2.3), these include the derivation of a global conservation index, the quantitative analysis of the conservation of physicochemical properties (some not used before for prediction purposes, such as the β-branched character), the Arg/Lys motif mentioned above, a probability function to predict lipid-facing residues, and the prediction of TMH tilting from the pattern of buried versus lipid

accessible residues.

The inconclusive nature of each one of these predictive tools taken individually have led to the most important guiding principle in the modeling process: First, many different and inconclusive methods are applied independently, deriving specific predictions from each one. Second, we integrate these predictions searching for consistency among their results for the derivation of structural inferences. The consequence of this methodological approach is that a very extensive analysis is necessary to derive structural hypotheses, and hence explains the length of this thesis. Two main factors are responsible for this extensive exposition. First, the large number of different methods applied, as they have become available with time, that must be introduced; Second, the use of a variety of very specific criteria required for the comparison of predictions derived from different methods, based on different evolutionary or physicochemical considerations, which need to be applied in order to resolve the inconsistencies found in the integration of the results. Criteria to evaluate the relative strength among predictions based on similar approaches (e.g. the conservation of physico-chemical properties) are developed by reformulating these prediction in a quantitative manner (e.g. the standard deviation of the mean hydrophobicity found at a given locus is an estimate of the conservation of hydrophobic character at this locus). Criteria for comparison among predictions based on qualitatively different approaches required a detailed understanding of the

relative strengths and pitfalls of each method applied. The expertise gained through the application and development of all these methodological approaches allowed us to write a comprehensive review (Ballesteros 1995) of the methods available for modeling GPCRs that is very frequently quoted.

As the analysis and construction of the model progressed at every structural level, an increasing wealth of experimental data pertinent to structure for many different GPCRs has been constantly appearing in the literature. The experimental results were based initially on mutagenesis studies and were later followed by the application of more biophysical approaches. Among this vast amount of data arising from the application of molecular biology techniques to GPCRs, the most informative for structural purposes is the finding of double revertant mutant constructs (Rao 1994; Zhou 1994; Liu 1995; Sealfon 1995; Han 1996), which imply an interrelated role between the residues involved, most likely achieved by spatial proximity. We predicted the first residue-to-residue double revertant mutant in GPCR between H2 and H7, a prediction that was validated experimentally in the GnRH and 5HT2A receptors (Zhou 1994; Sealfon 1995) through a collaboration with the laboratory of Dr. Stuart Sealfon. Among the biophysical data, the most significant have been of course the electron cryo-microscopy studies on bovine rhodopsin, which have provided the footprint of the 7 TMHs onto which all modeling approaches have been based since its publication. More recently, the biophysical methods of ESR and fluorescence

microscopy are providing dynamic insight into structural changes related to receptor activation, as described in a subsequent section that reviews the literature since our review chapter was published (Ballesteros 1995) (Section 2.1).

The continuing development of experimental approaches to study GPCR structure and function amplified the need for a structural context to rationalize and interpret the experimental results. In the absence of a detailed structure of GPCRs of atomic resolution, such a context was sought from modeling studies. Consequently, I am grateful to have lived through and contributed to the emergence of an increasingly synergistic collaboration between experimental and theoretical laboratories, creating a network of such interdisciplinary scientific collaborations to study the structure and function of GPCRs. This has resulted in numerous publications pertaining to combined theoretical and experimental approaches to study different receptor systems, as different as the 5HT2A (Sealfon 1995), 5HT2C (Almaula 1996a), dopamine 2 (Fu 1996), β2-adrenergic (Gether 1996), GnRH (Zhou 1994; Ballesteros 1996), and cannabinoid (Bramblett 1995) receptors. In particular, our initial and long standing collaboration with the laboratory of Dr. Stuart Sealfon has resulted in specific structure-oriented studies such as the double revertant mutants (Zhou 1994; Sealfon 1995), the role of residue 5.46 in ligand selectivity between the 5HT 2C and 2A receptors (Almaula 1996a), and the more recent exploration of the Arg-cage motif in the GnRH receptor (see

Section 7.3.2). The collaboration with Dr. Jonathan Javitch made possible the probing of our predictions on the TMH boundaries and orientation. My predictions described in this thesis were in agreement with the experimental findings from a study of the accessibility of H7 residues to the binding site crevice in the H7 segment of the D2 receptor by the Cys Substituted Accessibility Method (SCAM) (Fu 1996).

The discovery of constitutively activating mutations (CAM) in the β2-adrenergic receptor (Lefkowitz 1993; Samama 1993) lead to a redefinition of the minimal set of different functional states of the receptor that need to be considered to understand experimental data on receptor activation (Lefkowitz 1993; Samama 1993). The result was a proposed thermodynamic model of receptor activation called the ternary complex or two-state model (Lefkowitz 1993; Samama 1993). In order to rationalize available experimental data in a structural context relevant to these models, it became necessary to consider the structural correlates of the defined functional sates of the receptor i.e., the so called "conformational states". In an article published recently in collaboration with the laboratory of Dr. Brian Kobilka (Gether 1996), we reinterpreted this thermodynamic model by analyzing the relative energy levels of the different receptor states, and offered the insight that the activated state of the receptor is a metastable, high-energy, and thus transient conformational state of the receptor in the absence of G-protein. This was validated experimentally by the work of Drs. Ulrik Gether and Brian Kobilka

who showed the differences between wild type and CAM β2-adrenergic receptors in their observed fluorescence changes upon ligand binding, and suggested that these could be rationalized by our proposal of a high-energy, unstable and flexible activated state of the receptor (Gether 1996). These achievements and the approaches used in the work are described in detail in this thesis.

Multiple possible packing interactions have been found for several regions of the receptor model, e.g. Pro kinks. Selecting a particular choice in the modeling process, even in the presence of compelling inferences, does not exclude a role for these other, alternative packing arrangements. Instead, it may indicate either receptor flexibility and/or a role for these alternative packing arrangements in the different conformational states that the receptor may adopt. Thus, a recurrent dilemma throughout the modeling process is the need to select among alternative structural arrangements in several regions of the TMH bundle in order to model their interactions into a single model of the TMH domain. Criterion for this selection is developed by first exploring the conformational preferences of these localized motifs using energy-based methods (e.g. MonteCarlo simulations of the Arg-Cage motif in H3, Section 7.2). Because the possible alternative conformations resulting from this procedure satisfy both structural and energetic criteria, selecting a particular conformation for modeling purposes relies in functional considerations. Functional criterion in this context is defined in terms of

specific structure-function relationships that relate the alternative conformations found to specific functional states of the receptor. For example, in our studies of the Arg-cage motif (Section 7.2) the presence of an ionic bond between D3.49 and R3.50 in MonteCarlo simulations of wild-type and mutant constructs was negatively correlated with receptor activity, suggesting that R3.50 is held in the inactive conformation by D3.49. Thus, this structure-function correlation was used to select the conformation of the conserved Arg3.50 to model the inactive state of the receptor. I have attempted to anchor the selection among energetically feasible conformations on experimental data, which requires the formulation of specific hypothesis relating the observed structural features to functional correlates. In the absence of experimental data to guide the selection process, I have resorted to the structural inferences derived from the theoretical analysis, e.g. the analysis of the conservation pattern.

This illustrates an important concept in modeling studies: first, rationalization of the available experimental data in the context of the model results in the derivation of specific structural hypothesis responsible for the observed phenotype. Second, modeling structural domains where experimental data is not available requires the derivation of specific structural hypothesis based on theoretical considerations. The inferences obtained through each of these approaches is then integrated into a common structural template searching for consistency among the alternative conformations

selected for these structural microdomains. For example, the proposed D3.49-R3.50 interaction in the inactive state of the receptor is consistent with a similar role proposed for the E6.30-R3.50 interaction (Section 7.4.1), and thus both interactions are incorporated into the model. Therefore, rationalization of available experimental data and generation of novel structure-function hypotheses are both intrinsic components of the modeling process, performed simultaneously in a synergistic approach, as described in this thesis. A consequence of this approach is that the construction of a model for the TMH domain of the 5HT2C receptor requires the generation of more novel structure-function relationships that can or has been proven experimentally.

Multiple possible packing interactions have been found for several regions of the receptor model, e.g. Pro kinks. Selecting a particular choice in the modeling process, even in the presence of compelling inferences, does not exclude a role for these other, alternative packing arrangements. Instead, it may indicate either receptor flexibility and/or a role for these alternative packing arrangements in the different conformational states that the receptor may adopt. Interestingly, in addition to the flexibility provided by the Pro-kinks in H5 and H6, I have identified above S3.39 and C3.44 in H3, and the loop following P4.59 in H4, as potential sources of structural heterogeneity. These "conformational hinges" in H3, H4, H5, and H6 define the spatial relationship between the conserved inner core of the TMH domain, defined above by the aromatic cluster (see Figure 7.7), and the more divergent

"binding site crevice" towards the extracellular side (see Figure 7.21). Note the correlation between the partition of the receptor based on the "conformational hinges" identified in this work, and similar partitions of the TMH domain based on conservation and/or functional criteria. This correlation suggests that the sources for structural heterogeneity identified in this work, i.e. the proposed "conformational hinges", could play a role in the interconversions among the "conformational states" defined in Section 3.1, i.e. the different inactivated and activated forms of the receptor. An example is the proposed conformational change of the PK in H6 upon activation (Section 7.2.4). Alternatively, these same "conformational hinges" may have been utilized through evolution as a source of functional specificity, as proposed in Section 7.2.5.

2. METHODS

Most methodological approaches and available experimental data that can be used to model the 5HT2C receptor were reviewed in our published chapter in the series Methods in Neurosciences (Ballesteros 1995). Therefore, the published chapter is presented below (Section 2.1), followed by two sections where I describe new developments in the literature since the chapter was written (Section 2.2), and new methods I developed not included in the review chapter or in any other publication (Section 2.3).

2.1- Chapter neurosciences

Ballesteros, J.A. and H. Weinstein (1995). "Integrated Methods for Modeling G-Protein Coupled Receptors." Methods Neurosci. **25**: 366-428.

Reprinted by permission of the publisher.

Introduction

The rapid growth in cloning and expression of G-protein coupled receptors (GPCR) offers attractive opportunities to probe and reveal the structural basis of signal transduction mechanisms at the level of these cell surface receptors. Major insights have emerged from comparisons and classifications of amino acid (AA) sequences of GPCRs into families defined by evolutionary developments and adapted to perform selective functions. Mutation studies and chimeric constructs of the kinds described elsewhere in this volume continue to serve in probing such insights and in relating them to an understanding of the structural underpinnings of receptor function. Yet it is quite clear that the interpretation of results from such explorations of GPCR structure-function relations, as well as their integration into a mechanistic description of the molecular process in which the ligand binding signal is propagated to activate the receptor for subsequent interaction with G-proteins, depend on a level of structural understanding that is not currently available from direct experimental data.

Based on the increased rate of success achieved by molecular modeling (e.g. see (Benner, 1991; Lesk, 1992; Rost, 1993; Thornton, 1991)) and computational simulation methods (Brooks, 1988; Jahnig, 1992; McCammon, 1987; van Gunsteren, 1992; van Gunsteren, 1989; Weinstein, 1992a; Weinstein, 1992b) in providing structural insights relevant to the functions of

biological molecules, a variety of such theoretical approaches have already been applied to the study of GPCRs (Cronet, 1993; Dahl, 1991; Donnelly, 1994; Findlay, 1990; Hibert, 1991; Huggins, 1993; IJzerman, 1992; Lewell, 1992; Livingstone, 1992; MaloneyHuss, 1992; Nordvall, 1993; Smolyar, 1993; Trumpp-Kallmeyer, 1992; Trumpp-Kallmeyer, 1993; Westkaemper, 1991; Yamamoto, 1993; Zhang, 1993b; Zhou, 1994), as discussed in recent overviews (Findlay, 1994; Hoflack, 1993b; Humblet, 1992; Kontoyianni, 1993b). However, a coherent approach to such modeling and computational probing has not yet emerged, as the early efforts have focused on disparate aspects in the wealth of existing information about the GPCRs and other transmembrane proteins. In contrast, this presentation of methods and approaches applicable to the elucidation of structure-function relations of GPCRs at the atomic level of detail attempts a hierarchical scheme for incorporating existing information into GPCR modeling. Thus, the sequence of steps in the modeling process is organized to reflect the hierarchical but interrelated use of information about the primary, secondary and tertiary structural properties of GPCRs, as well as inferences from the experimental probing and biophysical analysis of transmembrane proteins.

Structural data on GPCRs, based on biochemical, immunological and biophysical approaches, has validated a consensus architecture of GPCRs with an extracellular N-terminus, a cytoplasmic C-terminus and a transmembrane domain comprised of seven transmembrane helical domains connected by

loops. The fact that the transmembrane portion of GPCRs consists of seven helices was recently revealed in the projection map of the electron density of bovine rhodopsin (RH) on the membrane (Schertler, 1993), although the structural details are not discernable in this low resolution (9 Å) structure. To date, the models of GPCR incorporate such a basic informatio, but a major drawback in improving them is that functional data in the form of pharmacological profiles and ligand structure-activity relations (SAR) appear non-discriminant as a test of the models. This shortcoming is evident from the claims of various authors that agreement with such pharmacological data are achieved in spite of the significantly divergent models they present for similar or identical receptors. Such lack of discrimination is especially disturbing when the modes of ligand binding in the various receptor models are compared to expectations from previous SAR studies, because many different combinations of side-chain residues in the receptor can correspond to the same qualitative predictions of a receptor binding pocket. Consequently, very different GPCR models can claim ligand binding arrangements that would be compatible with a given rank order of binding affinities, if only a small number of compounds is being analyzed. The ambiguities are exacerbated by the limited capability of mutagenesis experiments to distinguish between a direct, as opposed to an indirect effect of a mutation on ligand-receptor interaction if the sole assessment is the measurement of affinity in mutant receptors (Colquhoun, 1993). Together,

these difficulties have precluded a conclusive test of modeling hypotheses and of the resulting GPCR models, because a variety of molecular mechanisms can explain the observed changes in the measured properties (Ward, 1990). These apparent ambiguities that burden the pharmacological testing of GPCR models could be resolved with the use of structural information about the receptor, about mutants, and about the changes induced by ligand binding. Because direct structural determination remain impossible by current means, several laboratories have been searching for double-revertant mutants and/or chimeric constructs that impose more stringent criteria for inferences on structural details such as helix-helix interactions and spatial adjacencies of various domains in the GPCR proteins (Pittel, 1994; Rao, 1994; Suryanarayana, 1992; Zhou, 1994). Such details can provide essential information necessary for probing GPCR models at a structural level. At the functional level, novel criteria for probing GPCR models have emerged from the formulation of agonistic activity as a measurable conformational change on the receptor model upon ligand binding. Such considerations may prove to be a much more stringent discriminant of the ability of a GPCR model to agree with pharmacological data than the analysis of ligand binding modes in the model, as suggested by the results of recent simulations of a serotonin 5-HT_2 receptor subtype (Zhang, 1993b). At this time, the bulk of information required to interpret results from both the structural and functional probing of receptor models is

most likely to emerge from molecular modeling and computational simulation, as the structural elucidation of transmembrane protein structures is technically still a formidable task. Given the current state of the art, such applications of molecular modeling will require considerable improvements in techniques of receptor structure modeling, and augmentation with powerful methods of computational simulation of the dynamic, time-dependent behavior of the receptor molecules. Therefore, we have oriented this discussion of methodological approaches towards an understanding of the major capabilities and possible pitfalls of a variety of modeling and computational methods used in the study of structural and functional properties of GPCRs.

2.1.1- Primary structure of GPCRs: Conservation patterns and functional divergence

The primary structural information about GPCRs is encoded in their amino acid (AA) sequences. As it is not feasible to translate directly the AA sequence into three dimensional (3D) structure, sequence comparisons are used to identify the likely determinants for the structural commonality expressed in the template of seven loop-connected transmembrane helices (TMH) shared by the GPCRs. Moreover, such comparisons are used to identify the basis for the differences expressed in the functional properties of these proteins that determine their binding and response properties to the actions

of a large variety of ligands. Consequently, a sequence alignment (SA) of GPCR is one of the first and most crucial steps in the modeling process.

Two basic assumptions underlie the extraction of structural information about GPCRs from a set of aligned sequences (Lesk, 1992): First, that they all share a structural framework, as predicted for the TMH packing arrangement of the GPCR superfamily (Baldwin, 1993; Donnelly, 1989; Findlay, 1990; Trumpp-Kallmeyer, 1992). Second, that highly conserved residues can be considered essential for the structural and/or functional integrity of the receptor. For AA sites with a lower degree of conservation, this second hypothesis implies that the degree of conservation observable through the alignment is proportional to the role of these residues in determining the structure and/or function of the GPCR.

The criteria guiding the construction of a sequence alignment of GPCR are briefly reviewed in Section 2.1.1.1. The alignment makes possible an analysis of the conservation pattern at individual AA sites, as discussed in Section 2.1.1.2. Since the conserved residues are predicted to be molecular determinants of the structure and/or function of the GPCR, their identification should guide the modeling strategy. Identification of structurally or functionally equivalent AA positions in the great variety of GPCR sequences is facilitated by a general numbering scheme that reflects the sequence alignment and makes possible direct comparisons of loci related through this alignment. Such a general numbering scheme is presented in

Section 2.1.1.3.

2.1.1.1- Criteria and inferences in the construction of sequence alignments.

A multiple sequence alignment is constructed so as to maximize similarities at every AA site. This is an area of great general interest in bioinformatics and an increasing variety of computational algorithms is available to perform such tasks -for a review see (Chan, 1992). These computerized methods will not be discussed here. Rather, we focus on options and criteria that are useful to derive structural inferences from a sequence alignment (SA) of GPCRs. These criteria pertain to the selection of correct inputs for these alignments programs, and to structural considerations applicable for checking and refining the SA of GPCRs generated by those programs.

An important consideration is the selection of the proper set of GPCR sequences for the SA. This selection depends on the information that is sought and is determined by the extent of homology among the compared sequences. Alignment of sequences with intermediate homologies (i.e., 30-70%) among the GPCRs can identify continuous patterns of conservation distributed over the entire sequence. Such patterns provide structural inferences based on the relative degree of conservation among all AA sites, such as the helical periodicity in the conservation pattern described in Section 2.1.2.1.3.1. On the other hand, alignments with maximal (i.e., above

80%) or minimal (i.e., below 30%) homologies are useful in pinpointing a few residues that provide key information about structure or function. Thus, divergent sites among highly homologous sequences (above 80%) are either responsible for functional divergence -e.g. ligand binding specificity (Kao, 1992; Link, 1992; Oksenberg, 1992)- or are indicative of functionally nonessential positions -e.g. lipid-facing residues (Baldwin, 1993). In the comparison of highly divergent sequences (below 30%), conserved AA sites are candidate molecular determinants of the 3D structure and/or of receptor activation mechanisms (Baldwin, 1993; Donnelly, 1989; Findlay, 1990; Hibert, 1991), since both ligand binding and G-protein coupling have divergent specificities (Baldwin, 1993).

A quantitative measure of the sequence homology identified by SA is useful in determining the relation between AA conservation and structural determinants. The "mutation matrix" provides such a measure for individual AA-to-AA substitutions. The integration of the pairwise measurements comprised in the mutation matrix for a set of AA substitutions identified from comparing two sequences, provides a quantitative estimate of their overall homology. The AA substitutions (AA_i->AA_j) are quantified by the probability of their occurrence P(AA_i->AA_j), where i and j connote natural amino acids or a gap indicative of an insertion or deletion. The 21x21 possible mutation probabilities P(AA_i->AA_j) define a matrix called the mutation matrix (Dayhoff, 1983) used in all alignment

programs to quantify the extent of sequence homology between two proteins which is proportional to the probability that one of them will convert into the other by a series of mutations. Mutation matrices sometimes contain the inverse of the mutation probabilities, the "mutation penalties", but we will refer here to mutation probabilities for the sake of clarity. It is convenient to think of mutation probabilities as a parametrization derived from a set of sequence alignments that are representative of the proteins under study, in this case the GPCR. Two complementary criteria have been presented for the selection of the most appropriate mutation matrix: one based on the degree of evolutionary divergence, and the other based on structural considerations. According to the first criterion, the mutation probabilities are calculated from a set of protein sequences that are all within a similar evolutionary distance of the GPCR, such as in the mutation matrix pam200 described recently (Altschul, 1991). There are good reasons for using a set of sequences that are all within the same evolutionary distance, based on the argument that a given mutation will have a different probability to occur in 1 million than in 200 million years of evolution. The other criterion for an appropriate selection of the mutation matrix is based on structural considerations. For instance, it has been shown for soluble proteins that mutation probabilities depend on secondary structure, i.e. that they are different for residues in an α-helix compared to a β-sheet or random-coil (Overington, 1992). In addition, mutation probabilities for surface exposed residues in TMHs are different

from those in helices of soluble proteins (Donnelly, 1993b), in accordance with the opposite polarity of the solvent in each case. Both the evolutionary and the structural criteria pertain to the selection of the appropriate mutation matrix for the GPCRs. From the structural standpoint, different mutation matrices should be used for the TMHs and the loops. For the loop regions connecting the TMHs, the mutation matrix could be derived from soluble proteins disregarding structural considerations, and that covers the evolutionary distances found in GPCR, e.g. as in pam200 (Altschul, 1991). For the TMHs, both structural and evolutionary criteria could be met if the mutation matrix is derived directly from a comprehensive alignment of GPCR TMHs. However, because such a mutation matrix has not yet been presented in the literature, a reasonable alternative would be a combination of the mutation matrix for interior residues of soluble proteins in helical segments (Overington, 1992) and the mutation matrix for residues exposed to the lipid milieu from the PRC family (Donnelly, 1993b). The combination can be achieved by averaging the two mutation matrices, or by selecting either of them for individual AA sites depending on their predicted location in interior or lipid-exposed helical faces. Those mutation matrices would fulfill the structural criteria, although not the evolutionary criteria. However, since mutation probabilities are proportional to the evolutionary distances and all GPCR within a SA share a similar evolutionary distance, the absence of an evolutionary criterion would be expected to alter only the absolute

probabilities but not the relative degree of conservation among different AA sites, from which structural inferences are derived. It should be noted that the above considerations apply for the quantification of the extent of homology or conservation within a SA of GPCRs. If the aim is to use a particular GPCR sequence to search for homologous GPCR sequences against the entire protein database then other considerations apply. For instance, the evolutionary criteria becomes much more important because most proteins in the database have evolutionary distances significantly different from other GPCRs. Furthermore, for a database search we can not specify different mutation matrices for TMH and loops with current algorithms, although because TMHs are much more homologous than loops the structural argument would favor a mutation matrix for TMH. Summarizing, for a database search against all known protein sequences both evolutionary and structural criteria are necessary and the best alternative would be to derive a mutation matrix from GPCR TMHs fitted to that task.

In addition to their use in defining appropriate mutation matrices, structural criteria can also be useful in the refinement of SA generated for the GPCR by computer programs. Thus, the alignment can take into consideration that insertions and deletions in helices are very rare events (Pascarella, 1992) and should preferably occur in the loops. Any gaps inserted in predicted TMHs by the computerized SA can probably be relocated to loops without significant loss of homology. Furthermore, insertion of gaps in loop

regions is more advisable where non-conserved Pro or Gly residues occur. The reason is that the high flexibility of Gly or the high rigidity of Pro are often found to have a specific structural role and thus would require conservation. For example, conserved Gly or Pro in loops are suggestive of specific structures such as turns. A non-conserved occurrence of these residues indicates a structurally permissive site where gaps can better be accommodated.

2.1.1.2- Analysis of the conservation pattern from the sequence alignments

Sequence comparisons and inferred degrees of conservation have been used in most attempts to model the 3D structure of GPCRs (e.g., see (Baldwin, 1993; Hibert, 1991)), although few explicit methods for quantifying the degree of conservation and its pattern have been presented. It is noteworthy, however, that because the set of GPCR sequences available for SA is still limited, it remains unclear whether a certain degree of conservation observed from the alignment reflects a real property of that locus in the sequences, or an artifact of the limitation in the size of the database. For example, if a certain site contains 80% Tyr and 20% Leu, it is not clear whether the real evolutionary distribution for the particular family of GPCR being modeled would not be closer to 50:50 or 95:5 ratios if a complete sample were available. Clearly however, the model must be consistent with the information that both Tyr and Leu can be present at this locus. Therefore, the conundrum

arises as to whether the conservation analyses should be based on the nature of the different AA present at a given site, or on the frequencies of occurrence of each AA. In our example, the conservation analyses could be based on the existence of either Tyr or Leu at this locus, or on the 80:20 ratio of Tyr vs. Leu appearance. As illustrated below, both approaches can be applied in a sequential fashion to draw useful inferences. Alternatively, the selection of GPCR sequences for a SA could be done stressing the representativity across the evolutionary spectrum as suggested from evolutionary trees, and then the analysis could more safely be based on the frequencies of occurrence of each AA within the SA.

The first step in the conservation analysis consists of the quantification of the global degree of sequence conservation at a given AA site. This information is obtained from the set of different AAs that are found to occur at this position in the aligned sequences. For GPCRs, a variability profile V was proposed as the number of different AAs at a given position (Donnelly, 1989). The plot of V versus the AA number (naa) -Figure 2.1- provides the global conservation pattern through the sequence alignment; V(naa). However, this quantification ignores the distinctions -both physical and chemical- among the natural AAs, and hence does not provide information about the putative physico-chemical significance of the substitutions. A well known measurement of the quantitative significance of AA substitutions is the "mutation probability" that defines a mutation matrix, as described in

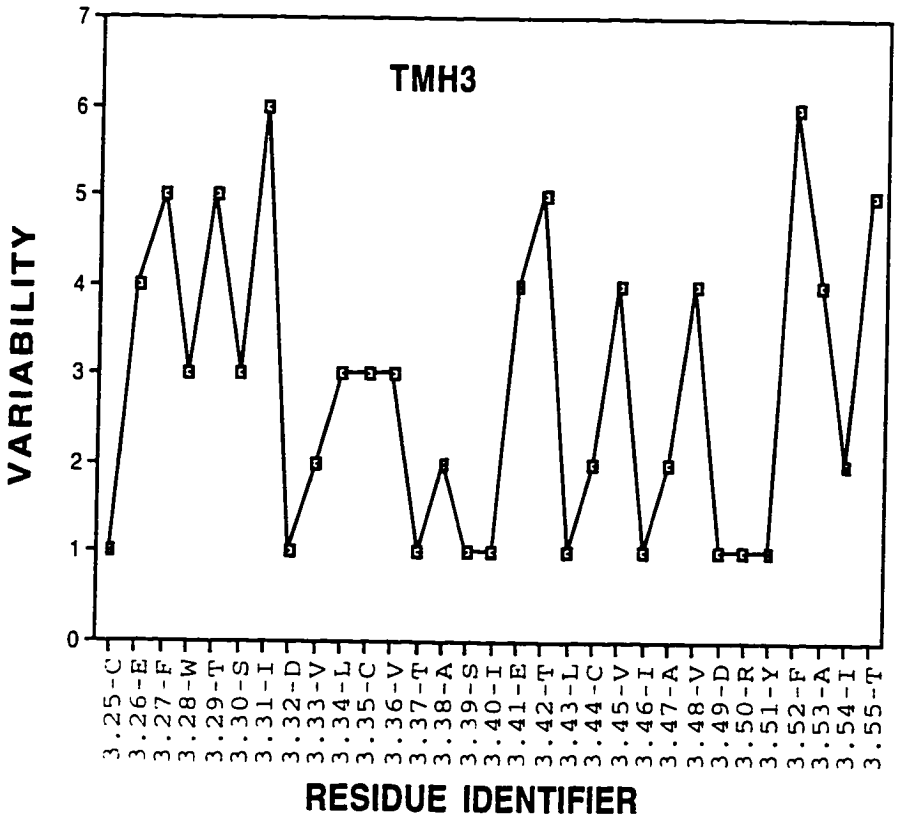


Figure 2.1:
Global conservation at each site of the TMH3 segment quantified through the variability profile V(naa) as the number of different amino acids occurring at the locus (naa). The plot is calculated from an alignment of dopaminergic, adrenergic and serotonergic receptors. The sequence identified on the abcisa corresponds to the human β_2-adrenergic receptor, and the locus is identified by the numbering scheme described in section 2.1.3. Note a periodicity of 3-4 residues for this conservation pattern, suggestive of α-helical character as analyzed in section 2.1.2.1.

Section 2.1.1.1, above. An alternative method to measure the global conservation at a given AA site uses mutation matrices and has been proposed for an analysis of the Photoreaction center II proteins (Svensson, 1991). It involves a similarity index defined as the average of all mutation probabilities between the AA present at a given site (Svensson, 1991).

The conservation pattern obtained from the use of such quantification methods is essential for extracting structural inferences. For example, the analysis of the helical periodicity of a conservation pattern by Fourier transform of the V(naa) (Donnelly, 1989; Rees, 1989b) provides a powerful algorithm for predicting the boundaries of the TMH segments, as discussed below in Section 2.1.2.1.3. Furthermore, Section 2.1.3.1.1 describes the use of the conservation pattern as a guide for modeling of the 3D architecture of GPCRs by suggesting TMH-TMH packings based on the correlation between high conservation and surface inaccessibility (Donnelly, 1993b; Overington, 1990; Rees, 1989b).

The use of the V(naa) defined above as guides for the 3D modeling of GPCR are based on the identification of those AA sites that are conserved beyond a given threshold. However, additional information that is important for the structural predictions resides in the detailed percentage presence of each specific AA at a given position within the conserved set, e.g. that $Tyr_{5.58}$ is 80% conserved because it occurs in 8 out of 10 GPCR aligned. This information is useful because the nature of the AA present at every

conserved site becomes a candidate molecular determinant of the structure and/or function of the GPCR. Since those molecular determinants exert their effects through physico-chemical properties, it is important also to quantify the conservation of a given physico-chemical property at a given site. This analysis should be done independently of the previous analysis of conservation of AA character since there could be positions that conserve a property such as hydrophobicity yet were not considered conserved by the previous analysis. Methodologically, the analysis of conservation of physico-chemical properties has been based in grouping the 20 natural AAs into subclasses according to an associated property. For example, Baldwin (Baldwin, 1993) regrouped the 20 AA into three classes according to their size: small, medium and large. AA sites containing residues belonging to only one size class or two neighboring classes were considered conserved, and thus predicted to lie in helix-helix interfaces as other conserved residues. A clearly significant physico-chemical property for membrane proteins is the hydrophobicity/hydrophilicity. Again, the method used for its conservation analysis was the classification of the 20 AA into polar or hydrophobic, with the AA sites that conserve a polar character predicted in helix-helix interfaces while positions that conserve an apolar character considered more likely to face the lipids (Baldwin, 1993; Trumpp-Kallmeyer, 1992; Zhang, 1993b). Other physico-chemical properties that can be identified by grouping AA are the electrostatic charge (R,K or D,E), analogous H-bonding characteristics (S,T),

aromaticity (F,Y,W,H), and these could serve in similar analyses leading to structural predictions.

2.1.1.3- Numbering scheme for GPCR sequences.

To relate site-defined properties to the sequences of the many different classes of GPCR proteins, we have developed a common numbering scheme that is informative of the relative position of each AA, the AA present at that position, and the real AA number in a particular GPCR. We will refer to these three generalized numbers associated with each AA position in a GPCR sequence as identifiers, and they are derived as follows: Every AA identifier starts with the TMH number -e.g. 4. for TMH4- and followed by its position relative to a reference residue among the most conserved AA in that TMH. That reference residue is arbitrarily assigned the number 50. For example, the most conserved AA in TMH4 is a Trp whose identifier would be 4.50, i.e. $W_{4.50}$. A Ser residue located 5 AA after $W_{4.50}$ will be $S_{4.55}$. This general numbering scheme is illustrated in Figure 2.2 on the helical net representation of the human β_2-adrenergic GPCR. To relate the identifier to the AA sequence of a particular GPCR protein, each identifier defined above can be followed by the numbering in the particular sequence. For example, $W_{4.50}$ is number 136 in the human β_2-adrenergic GPCR and would be referred to as $W_{4.50(158)}$ for that specific receptor.

Mutations are identified in this numbering scheme in the usual manner, with the wild-type identifier followed by the mutant AA. For example W4.50(158)F defines the Phe mutation of the Trp wild type. This identification scheme allows for a systematic comparison of mutations done in different receptors at the same loci.

Table 2.1 lists selected reference AAs in each TMH and illustrates the new identifiers for TMHs of the β_2-adrenergic receptor. The relative position of each reference AA in the GPCR TMHs can be seen in the helical net representation of the human β2-adrenergic GPCR in Figure 2.2.

If the AA is in a loop, this numbering scheme could provide it with 2 different identifiers, e.g. L1.63=L2.34. For a specific AA site, the proximity to a TMH boundary determines which identifier is actually used, although both define uniquely the same position.

Table 2.1

Generalized numbering scheme for GPCR sequences.

TMH	100% conserved in Neurotr. GPCR	AA identifier	AA# in β_2-ADR	AA identifier in β_2-ADR
1	Asn	N1.50	51	N1.50(51)
2	Asp	D2.50	79	D2.50(79)
3	Arg	R3.50	131	R3.50(131)
4	Trp	W4.50	158	W4.50(158)
5	Pro	P5.50	211	P5.50(211)
6	Pro	P6.50	288	P6.50(288)
7	Pro	P7.50	323	P7.50(323)

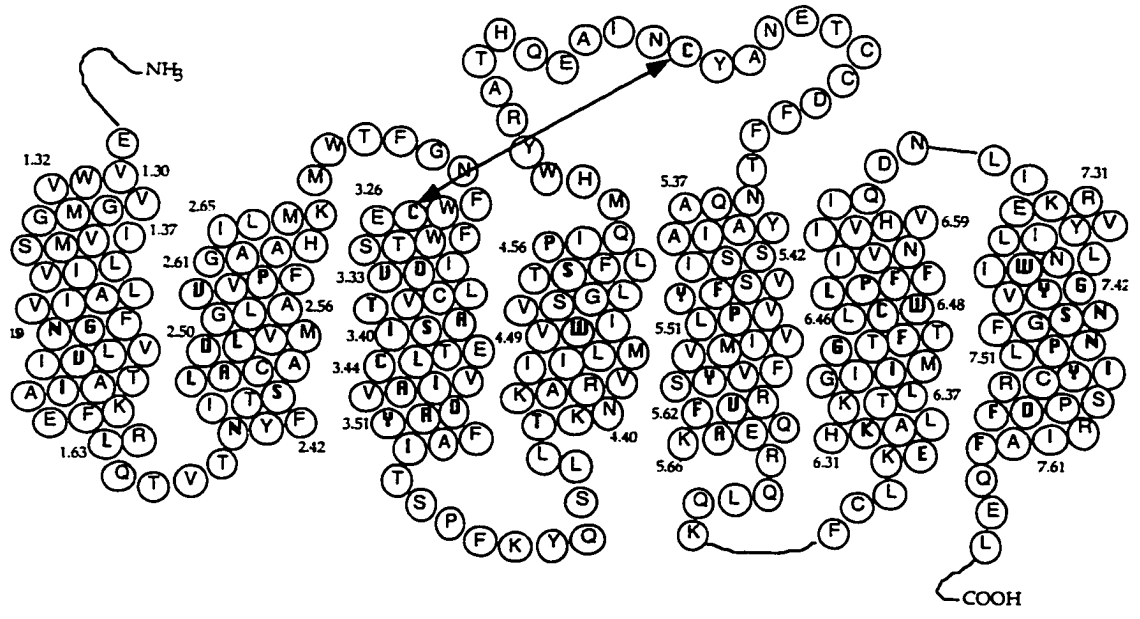


Figure 2.2:
Generalized numbering scheme illustrated on a helical net representation for the human β_2-adrenergic receptor. Highly conserved residues are shown in bold letters. See Figure 2.9 for an explanation of helical net representations. TMH ends are taken from (D. Donnelly, J.B.C. Findlay, A.M. MacLeod, and T.L. Blundell, submitted (1994)).

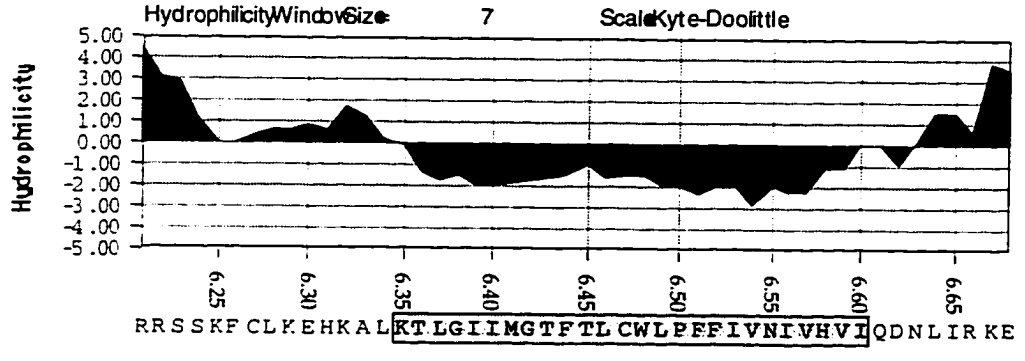


Figure 2.3:
Hydrophobicity profile for TMH6 of the human β_2-adrenergic receptor used to predict TMH boundaries. Note that the predicted helical ends by this criterion would be residues 6.35 and 6.60 for the N- and C-terminus, respectively. The locus is identified by the numbering scheme described in section 2.1.1.3.

2.1.2- Secondary structure of GPCRs: Prediction and modeling of individual transmembrane helices

The helical nature of the transmembrane domains of GPCRs (e.g., see (Findlay, 1990)) was recently confirmed by the 9 Å resolution structure of bovine rhodopsin (RH) (Schertler, 1993), validating the prediction of 7 TMHs for rhodopsin based on the hydrophobicity profile (Findlay, 1984). Consequently, we assume the general validity of the parsing of GPCRs sequences into sections representing the seven TMHs, and concentrate on methods to predict the TMH ends, presented in Section 2.1.2.1. With the TMH ends defined, the essential considerations for the 3D modeling of each TMH at atomic detail are presented in Section 2.1.2.2. Finally, some approaches for secondary structure predictions for the TMH-connecting loops as well as for the N- and C-terminal domains are discussed in Section 2.1.2.3.

2.1.2.1- The prediction of TMH boundaries

The most common approach for predicting the transmembrane helical domains from GPCR sequences rests on the hydrophobicity profile (Kyte, 1982), but at least four other useful approaches are available to achieve such parsing of the primary protein structure. The five methods presented below are qualitatively different, thus providing alternative paths to the prediction of the TMHs. Slightly conflicting predictions are likely to occur due to the

inconclusive nature of each of these methods when applied individually. Consequently, the ability to quantify and compare predictions from each method to pursue convergence among their results becomes a desirable aim in the selection of the predicted TMH boundaries. As described below, the application of these methods leads to a definition of the TMH domain that goes beyond the limits provided by hydrophobicity criteria and extends the TMHs into other physico-chemical environments such as the phospholipid head-groups.

The predictions from the hydrophobicity profile presented in Section 2.1.2.1.1 are challenged in Section 2.1.2.1.2 by the hypothesis that Arg and Lys adjacent to the predicted cytoplasmic ends belong to the TMHs, where they supposedly interact with phospholipid head-groups. A combination of these predictions leads to the redefinition of some TMHs that is supported by the analysis of the α-helical periodicity in the AA sequence. The α-helical periodicity of the sequence has been suggested as yet another approach to the prediction of TMH ends, as discussed in sections 2.1.3 and 2.1.4, below. Other approaches to predict the TMH boundaries rely on structural and evolutionary considerations, such as the position of non-conserved Pro residues (2.1.5). Finally, statistical methods based on the probability of each AA to belong to a TMH are presented in Section 2.1.2.1.6. All these prediction methods use the conservation analysis of the GPCR sequence alignment described in Section 2.1.1, based on the assumption that the aligned GPCRs

share a common structure, in this case similar TMH boundaries. This assumption is verified by the comparison of the L versus M subunit of the Photosynthetic Reaction Center (PRC) in the available crystal structure (Deisenhofer, 1989).

2.1.2.1.1- The hydrophobicity profile as parsing criterion

In this approach, a sequence is scanned with a fixed window size and the average hydrophobicity is calculated within that window (Kyte, 1982). The plot of the average hydrophobicity against the sequence number identifies regions with high overall hydrophobic character such as expected of the TMH domains (Kyte, 1982). A hydrophobic region identified in the plot is predicted as a TMH if it spans at least 18 AA, a length considered to represent the minimum number of residues needed to traverse the cell membrane in an α-helical conformation. This method has been extensively applied to sequences suspected to encode GPCR proteins, although its focus has rarely been to define the exact boundaries of the 7 TMHs. Nonetheless, TMH ends have been inferred as the AA at which the sequence returns to residues with high hydrophilic character, as shown in Figure 2.3, based on the assumption that the hydrophobic environment of the membrane excludes the hydrophilic residues. This assumption is likely to be more correct for the region of the protein facing the lipid chains of the membrane than for the ends of the TMH domains that are likely to be adjacent to the polar lipid

head-group regions of the membrane. Consequently, TMH boundaries predicted by the hydrophobicity profile should represent minimal TMH boundaries, since the helices could actually extend into the more polar lipid headgroup region of the membrane, as discussed in the following Section. Although the hydrophobicity profile was originally derived for a single sequence, it can also be applied through the sequence alignment of GPCRs in the form of average hydrophobicities (Dahl, 1991). The advantage of this "evolutionary averaged" hydrophobicity plots is that they may overcome irrelevant deviations of the TMH hydrophobic pattern present in particular GPCRs. The assumption of common TMH boundaries for the GPCR within the SA allows for the analysis of the evolutionary consistency of these predictions.

The optimal choices of a hydrophobicity scale from which the averaged hydrophobicity is calculated, as well as the optimal window size and the threshold values for the prediction of TMHs using the hydrophobicity profile have been the subject of several studies (Degli Esposti, 1990; Edelman, 1993; Edelman, 1989). The authors suggested the use of different hydrophobicity scales, pursuing convergence among their results as the guiding predictive principle. Notably, scales based on partition coefficients of peptides and/or small molecules between aqueous and non-polar solvents yielded larger errors than scales based on residue solvent accessibility in protein structures or statistical occurrence of each natural AA within TMH segments. The

evaluation of the optimal choice of the window size was also inconclusive; small windows (e.g. 7 AA) offered better resolution but produced more false positives (i.e., TMH-like regions), whereas larger windows reduced false positives but also merged closely spaced TMHs. Since the aim is to predict TMH ends with the total number of helices assumed to be fixed at seven, a shorter window of 7 AA is expected to perform better due to its better resolution.

2.1.2.1.2- The use of Arg and Lys residues to define the cytoplasmic ends.

Predicted to lie outside the TMH by the hydrophobicity profile, the Arg and Lys residues positioned predominantly at the cytoplasmic ends of BR and PRC, were actually found to belong to the TMH domains (Ballesteros, 1992a) in the available structures of these proteins (Deisenhofer, 1989; Deisenhofer 1987; Henderson, 1990a), as shown in Figure 2.4 for BR. It is thus very likely that when such residues appear in GPCRs next to the hydrophobic domains of the putative TMHs, they actually belong to these helices and can be used to define their cytoplasmic end (Ballesteros, 1992a). The mechanistic explanation given for the observation that Arg/Lys residues are part of the TMH region in BR and PRC is that these residues anchor the TMH to the membrane through ionic interactions with phospholipid head-groups, as suggested by NMR studies on other membrane proteins (Deber, 1986; Roux, 1989; Yeagle, 1982).

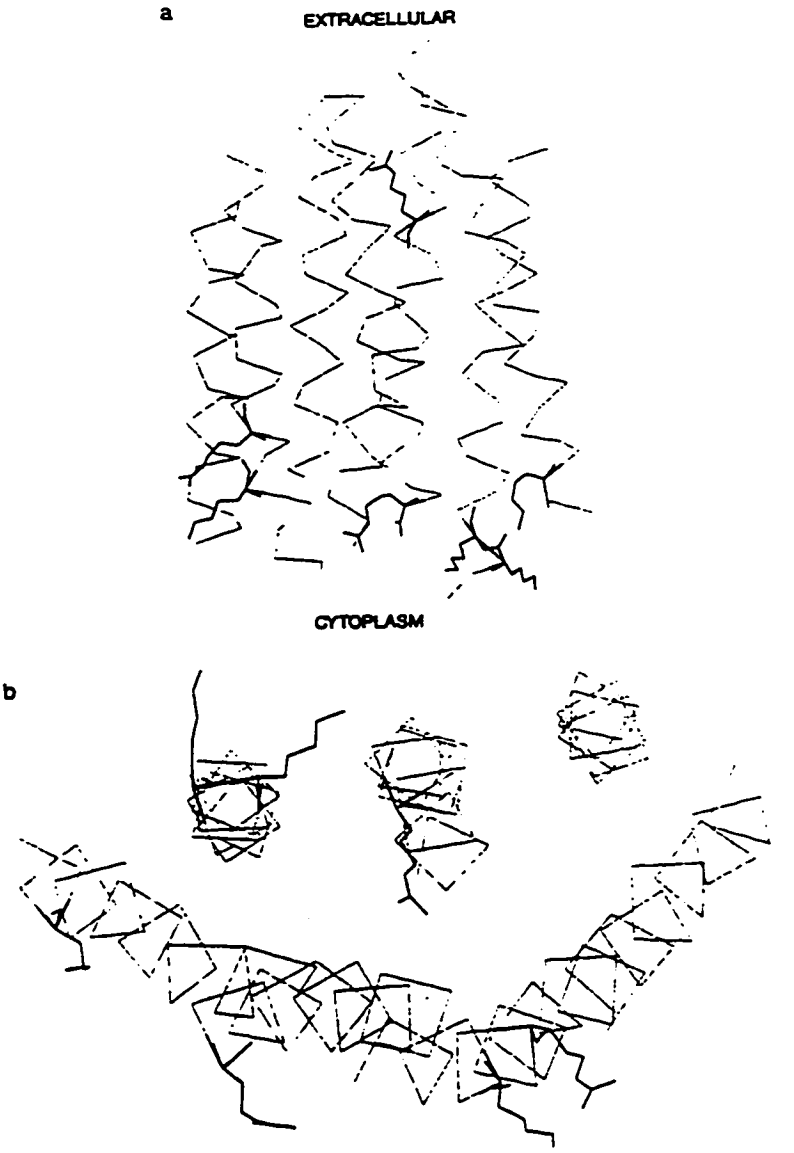


Figure 2.4:
The seven TMH bundle of BR is shown by the C_α trace plus the Arg and Lys residues. a) View parallel to the membrane plane. Note that 6-out-of-7 Arg/Lys residues are concentrated at the cytoplasmic boundaries. b) View along the membrane axis from the cytoplasmic side. All 6 Arg/Lys at the cytoplasmic boundaries shown in (a) belong to the TMH, and are protruding into the lipid phase, presumably to interact with the phospholipid head-groups through ionic bonds.

This would explain the cytoplasmic localization of the Arg/Lys since it is known that the inner leaflet of cell membranes is richer in negatively charged phospholipids than in the outer one. For use as a criterion, it is reasonable to predict any Arg/Lys at the cytoplasmic ends to lie inside a TMH if they appear in the sequence with an α-helical periodicity (i,i+3 or i+4). An example is seen in a helical net representation of TMH6 of the β_2-adrenergic GPCR (see Section 2.1.2.1.4 for definition), shown in Figure 2.5. In that helix, the patch formed by $\{K_{6.29}, K_{6.32}, K_{6.35}\}$ is predicted to be part of the TMH and to face the membrane environment. Such helical patches of Arg/Lys are often observed in the representation of GPCR sequences continuing the hydrophobic patch of the TMH core -see Section 2.1.3.1. This continuity of the helical domain, shown by the analysis of the helical periodicity in Section 2.1.2.1.3 and 2.1.4, integrates the predictions of TMH boundaries provided by the hydrophobicity profile with the Arg/Lys criterion. Since the phospholipid head-groups are considered the limit of the membrane in the intracellular face, the ends predicted by this motif constitute maximal cytoplasmic TMH ends, as opposed to the minimal ends predicted from the hydrophobicity profiles.

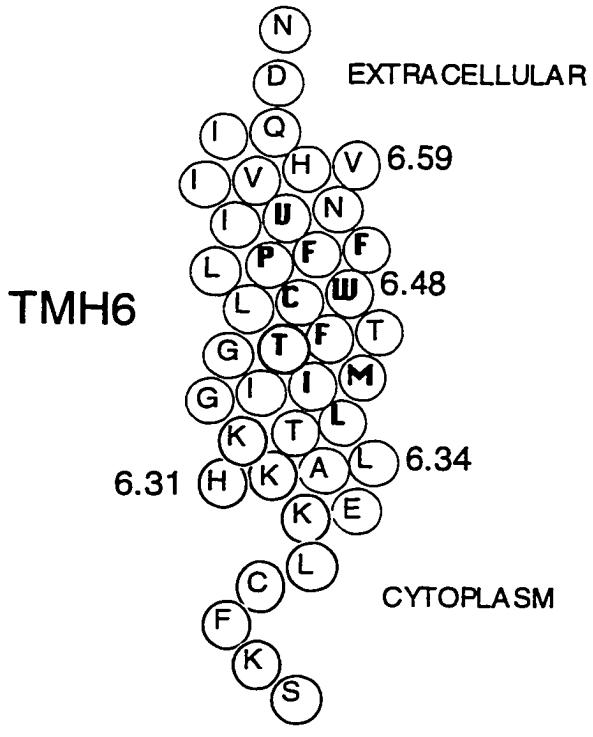


Figure 2.5:
Prediction of TMH6 boundaries by the Arg/Lys cytoplasmic motif. The sequence correspond to the human β_2-adrenergic receptor. Residues highlighted by a thicker circle represent positions where there are Arg or Lys residues in any sequence within an alignment of neurotransmitter GPCRs. Note that beyond the predicted TMH6 cytoplasmic end Arg/Lys occur at consecutive positions, thus no longer consistent with an α-helical periodicity. Highly conserved amino acids in a SA of adrenergic, dopaminergic and serotonergic GPCRs are shown in bold letters.

The extension of the single TMH of M13 filamentous bacteriophage from the hydrophobic core into the Arg/Lys region, in accordance with the considerations described above, was confirmed with NMR measurements (Henry, 1990). Notably, the three Arg/Lys residues present in M13 at the cytoplasmic boundary were shown to be the energetic determinant of M13-lipid interactions because mutation of these Arg/Lys to Glu/Asp

converted M13 into a soluble protein (Gallusser, 1990). The authors concluded from the properties of these mutants that the M13 protein anchors to the membrane through electrostatic interactions, consistent with the proposed role of the Arg/Lys in anchoring the TMH to the membrane through ionic pairs with the phospholipid head-groups. The direct interaction between the Arg/Lys side-chains and phospholipid head-groups has been measured by ^{31}P-NMR for a number of TMH-containing proteins using the signature of the phosphorous group which displays an immobilized component due to the presence of the protein (e.g. glycophorin (Yeagle, 1982)), an effect that can be mimicked by poly-Lysine (Deber, 1986).

The picture of a typical TMH that emerges from this analysis is thus of an hydrophobic core with polar ends facing the phospholipid head-groups, especially in the cytoplasmic extension which is rich in basic residues, as shown in Figure 2.6.

The extent of the refinement in the definition of cytoplasmic ends of TMHs that results from the inclusion of these Arg/Lys in addition to the predictions from the hydrophobic profile is dependent on the TMH topology, i.e. whether the N-terminus of the TMH is at the cytoplasmic side (N_{in}-to-C_{out}) or at the extracellular side (C_{in}-to-N_{out}). TMHs with N_{in}-to-C_{out} topology would have a larger error in the cytoplasmic ends predicted with the hydrophobicity profile than the TMHs with opposite topology (C_{in}-to-N_{out}). The reason, illustrated in Figure 2.7, is that in an α-helix the direction of the

C^{α}->C^{β} bond is always oriented toward the N-terminus. Therefore, Arg/Lys residues for N_{in}->C_{out} TMHs (TMH6 in Figure 2.7) can extend over 2 turns of the helix (7 AAs) towards the cytoplasmic side to interact with the phospholipid head-groups, while the hydrophobic portion of their side-chains can interact with the lipid chains. On the other hand, Arg/Lys residues for C_{in}-->N_{out} TMHs (TMH7 in Figure 2.7) can barely extend one half turn (2 AAs) towards the cytoplasmic side to reach the phosphate groups. This observation explains the major errors committed in predicting TMH ends in proteins for which experimental structural information has become available, as shown for BR in Table 2.2. The differential error in the predicted cytoplasmic ends has a topological dependence that not only agrees with, but can be satisfactorily explained by, the observed presence of the Arg/Lys in the TMHs. The magnitude of the error in the prediction of TMH boundaries using the hydrophobicity profile can now be estimated from Table 2.2. For TMH with an N_{in}-to-C_{out} topology the error is around 7 AAs (TMH4 is not well defined in BR), i.e. 25-30% (7/22) of an average TMH length. For TMH with an N_{out}-to-C_{in} topology the average error is insignificant. As discussed in Section 2.1.3.2.2 below, the topological dependence of this error (i.e., even numbered helices in GPCRs are affected more than the odd numbered ones) could seriously affect the validity of helix-helix packing models if TMH boundaries are taken from the hydrophobicity profile.

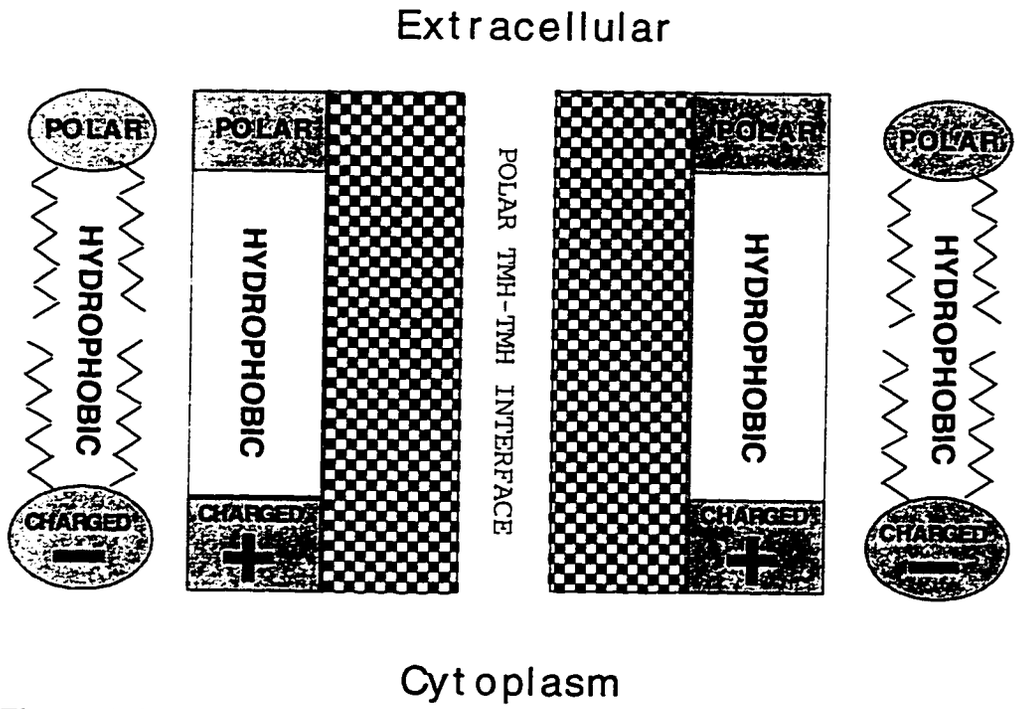


Figure 2.6:
Emerging picture of the polarity pattern for a prototype TMH from the inclusion of the phospholipid head-group regions. Note the complementarity in the polarity distribution pattern at the TMH-Lipid interface, especially the ionic complementarity of the Arg/Lys at cytoplasmic boundaries with the negatively charged character of inner monolayers in biological membranes. (Note that the extracellular polar extension of the TMH is drawn by analogy to the cytoplasmic end, although no clear example has been presented for this portion of the TMH). Residues on interior faces of the TMH bundle have similar polarity properties to those found for soluble proteins.

Table 2.2
Predicted and Observed TMH of BR using the hydrophobicity profile

Helix	Predicted Kyte-Doolittle	Observed	Cytoplasmic Difference
1	10-34	10-32	2
2	44-68	38-62	6
3	78-102	80-101	1
4	106-130	108-127	-2
5	134-158	136-157	1
6	175-199	167-193	8
7	200-224	203-227	-3

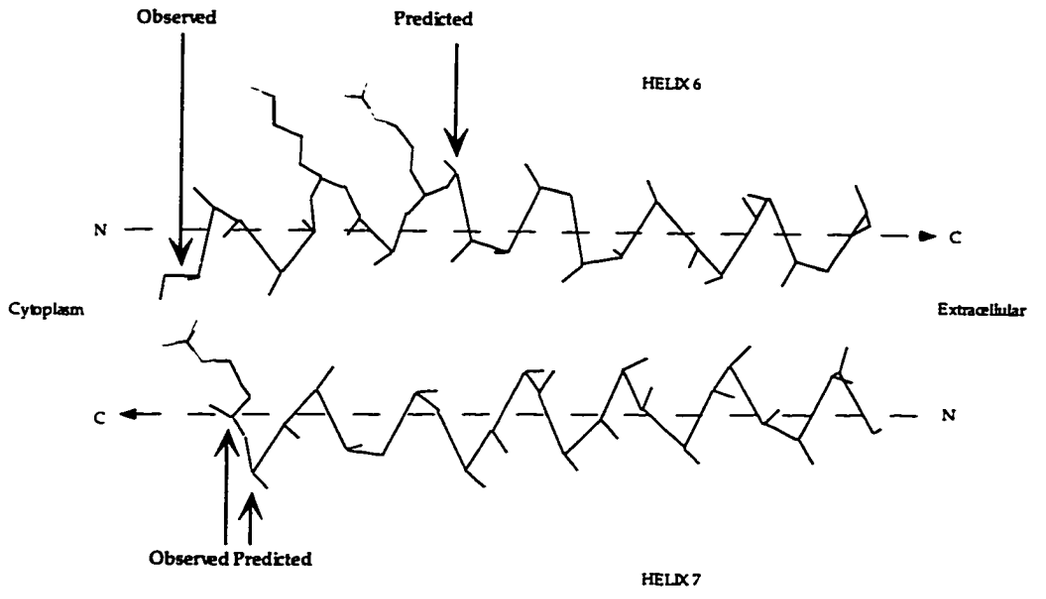


Figure 2.7:
TMH 6 and 7 of BR showing the Cα–Cβ trace plus the Arg/Lys residues, in a view parallel to the membrane. TMH6 has the topology N_{in}->C_{out} while TMH7 has C_{in}-->N_{out}. The topological dependence of the error in the predicted cytoplasmic ends using the hydrophobicity profile can be observed: 8 AAs for TMH6 and 2 AAs for TMH7. This topological dependence can be explained by the Arg/Lys motif. The Cα–>Cβ vector is always oriented toward the N-terminus. Therefore, Arg/Lys residues for N_{in}->C_{out} TMHs (e.g., TMH6) can extend over 2 turns of the helix (7 AA) towards the cytoplasmic side to interact with the phospholipid head-groups. On the other hand, Arg/Lys residues for C_{in}-->N_{out} TMHs (e.g., TMH7) can barely extend half turn (3 AA) towards the cytoplasmic side to reach the phosphate groups. Note that the actual helix-helix interface is offset from the cytoplasmic ends of the hydrophobic segment by 7 AA.

Although the Arg/Lys criterion was originally derived for a single sequence, it can also be applied through the sequence alignment of GPCRs. The assumption of common TMH boundaries for the GPCR within the SA allows for the analysis of the evolutionary consistency of these predictions. For example, the positions where those Arg/Lys occurred within an alignment of neurotransmitter GPCR are shown in Figure 2.5 on an helical

net representation of TMH6 in the β_2-adrenergic receptor. Note the consistent helical periodicity in the evolutionary disposition of the Arg/Lys in the predicted TMH; they always occur in a single face of TMH6 predicted to face the lipids -see Section 2.1.3.1. Beyond the predicted cytoplasmic end of TMH6 there are six Arg/Lys residues occurring at consecutive AA sites, inconsistent with helical periodicity.

2.1.2.1.3- Prediction of TMH segments from α-helix periodicity of properties measured by Fourier transform analysis.

The regularity of α-helical structures, and more specifically their characteristic periodicity of 3.6 AA/turn, can be used for prediction purposes based on the assumption that if the profile of a given property -e.g. hydrophobicity- within a segment of the sequence alignment displays an α-helical periodicity, this segment can be predicted to be a TMH. Properties whose value for every residue along the sequence reflects a degree of surface exposure, such as hydrophobicity (Eisenberg, 1984b; Rees, 1989b) or conservation (Rees, 1989a; Rees, 1989b), would exhibit an amphipathic profile in a TMH and can thus serve in the detection of the α-helical periodicity.

In the computational algorithm developed by Komiya et al. (Komiya, 1988) and Donnelly et al. (Donnelly, 1993a) to take advantage of these criteria for TMH modeling, a property profile U is calculated for a sequence from the individual properties U(j) assigned at each position j in a sequence alignment.

U(j) can be the hydrophobicity H(j) or the conservation as measured by the variability V(j) described in Section 2.1.1.2. Two methods are used to analyze the α-helical periodicity of a property profile U(j): a Fourier transform analysis presented in this Section, and the visual inspection of helical net representations in search for continuous patches along the TMH surface, presented in Section 2.1.2.1.4 below. The Fourier transform analyzes the overall pattern throughout a sequence segment, e.g. the amphiphatic character, whereas the analysis of continuous patches of AA with a common property in a helical net representation identifies residues that fulfil the TMH periodicity if they are separated by 3-4 AAs. Both approaches can be used to predict TMH boundaries, or to define TMH-TMH versus TMH-lipid interfaces as described below in Section 2.1.3.1.

In the Fourier transform analysis of the property profile U(j) over a window size N (Eisenberg, 1984b; Komiya, 1988), the power spectrum P(ω) reveals all existing periodicities ω. If the sequence contained within the window N adopts an α–helical conformation, a peak should appear around ω=105°, the angle between adjacent side-chains for an α-helix viewed down its axis. The power spectrum P(ω) is calculated as:

$$P(\omega)=\left[\sum_{j=1}^{N}(U(j)-U_m)\sin(j\omega)\right]^2+\left[\sum_{j=1}^{N}(U(j)-U_m)\cos(j\omega)\right]^2$$

where U_m is the mean value of U(j) over the window N. If U(j) is the hydrophobicity H(j), the Fourier transform leads to the commonly used hydrophobicity moment (Eisenberg, 1984b). An α-helix periodicity index (AP) calculated from P(ω) describes the extent of the periodicity in the helical region compared to that over the entire spectrum, and is defined as

$$AP = \frac{(1/30)\int_{90^o}^{120^o} P(\omega)d\omega}{(1/180)\int_{0^o}^{180^o} P(\omega)d\omega}$$

A value of AP>2 was suggested as a significant indication of α-helical character based on an analysis of the known TMH structure of the PRC (Rees, 1989b), although parameters such as the window size and number of sequences within the alignment can affect the AP value. Since AP measures the extent of α-helix periodicity, a plot of AP versus the residue number calculated by scanning the sequence or sequence alignment with a constant window size N can be used to identify the TMHs (Rees, 1989b) and to infer their boundaries (Donnelly, 1993a). The sequence segment at which the AP value drops significantly indicates the N or C terminus of the TMH.

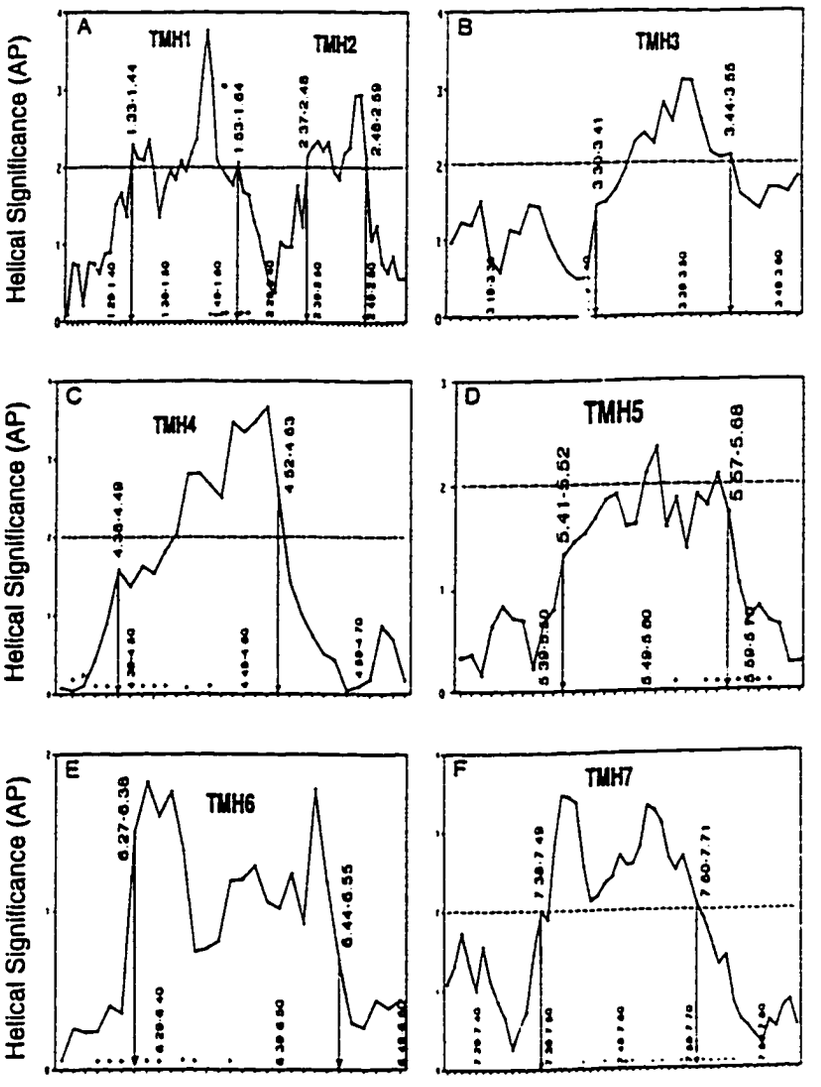


Figure 2.8:

α-Helix periodicity of the variability profile measured by Fourier transform analysis used to predict TMH boundaries for TMH1-2 (in panel A) to TMH7 (in panel F). The plots were calculated with a scanning window of 12 AAs shown at the abscisa. TMH ends predicted by this criterion are indicated by arrows, the occurrence of Arg/Lys residues by positive signs, and the suggested significance threshold for α-helix periodicity is marked by a dotted line. Note the presence of Arg/Lys residues inside predicted TMH segments at the cytoplasmic side, i.e. C-termini for odd TMHs and N-terminal for even TMHs. The degree of α-helical periodicity was calculated from an alignment of adrenergic, dopaminergic, and serotonergic GPCRs.

For an illustrative example of the application of this procedure to define the putative ends of TMHs, we have scanned a sequence alignment of dopaminergic, adrenergic and serotonergic GPCRs calculating the AP values for each consecutive segment of 12 AA. The plots obtained for each of the seven TMHs (Figure 2.8) clearly identify each of them as a region of increased AP values within the AP plot, indicating the presence of significant helix periodicity within these regions. The TMH boundaries predicted by these plots are determined from the first 12 AA segment at which the AP changes from a basal value to one indicating significant helix periodicity and then back again, as indicated by arrows in Figure 2.8. If this change occurs in the 12 AA window at the N-terminus of the TMH, the predicted beginning of this TMH is the first AA within this window, e.g. residue 6.27 in TMH6. Similarly, if the drop in AP value is in the 12 AA window at the C-terminus of the TMH, the predicted end of this TMH is the last AA within this window, e.g. residue 6.55 in TMH6. Note, however, that the predicted C-terminus of TMH6 could be either 6.55 or 6.54 since the drop in the AP plot is not very sharp.

Some additional considerations are necessary in the interpretation of the TMH boundaries predicted by the Fourier transform of the conservation pattern (Figure 2.8). First, the assumption that lipid-exposed residues are less conserved than residues facing the interior of the TMH bundle is based on structural premises. But interior residues that are functionally divergent -e.g determining ligand binding specificity- may induce a similar variability on

both helix faces thus leading to a loss of helical periodicity. This seems to be the case for the C-terminus of TMH2 that shows a subtype-specific AA distribution at positions 2.60 and 2.64. Second, the distortion of α-helix character induced by a Pro residue includes a face-twist of the helical portions N- and C-terminal from the Pro (Ballesteros, 1992a; Sankararamakrishnan, 1992) -see also Section 2.1.2.2.2 and Figure 2.10. The periodicities are offset by this Pro-induced shift, thus diminishing the AP value for window segments that include the Pro-kink region as defined in Section 2.1.2.2.2 below. Another artifact in the AP plot produced by a Pro residue occurs when it is facing the lipids and yet it is highly conserved. Such an apparent departure from the general expectation that lipid facing residues are not conserved, applies specifically to Pro because this AA can exert its structural effects without facing the protein interior -e.g. $Pro_{6.50}$ in TMH6 according to Baldwin (Baldwin, 1993). To correct for this occurrence, Donnelly et al (Donnelly, 1993b) proposed to alter the Pro variability to an average value in order to increase the detection of the helical periodicity. A consequence of the presence of Pro in TMHs 2, 4, 5, and 6 in the GPCR aligned for Figure 2.8 is that the extracellular segments of the helix relative to the Pro residues are shorter than the 12 AA window chosen for the Fourier analysis. This is responsible for decreasing the AP values (e.g., see Figure 2.8) and prevents the accurate prediction of the extracellular TMH boundaries for these helices. Donnelly et al. argued that the presence of AA sites whose variability was significantly

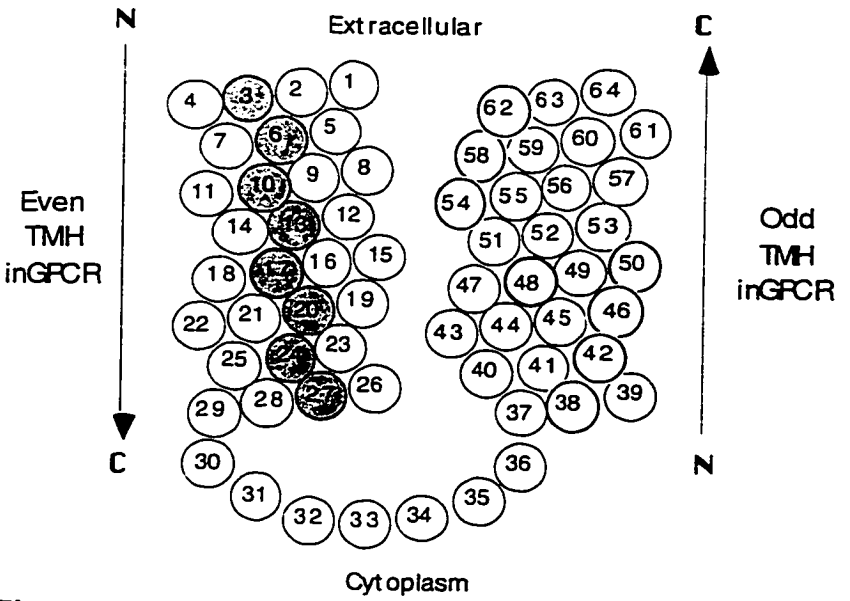


Figure 2.9:
The helical net representation, illustrated by two consecutive TMHs with opposite topologies. Each circle represents the position of a C^{α} of an ideal α-helix on the surface of the cylinder enveloping a helix with the same axis. When this cylinder is unwrapped, it provides a two dimensional surface of the envelope of an ideal α-helix. The positions of the C^{α} on this envelope constitute the helical net. The path described by consecutive residues can be followed in this figure from the residue numbers. Note that consecutive residues on the continuous cylinder surface can appear at the opposite margins of the helical net, e.g. see residues 4 and 5. Thus, residue 5 (AA_5) should be considered for all purposes as if it was positioned next to AA_4. Note the handedness of the helical net progressing from the N- to the C-termini. This reflects the right handed symmetry of the α-helix that is not reversed for an even- relative to an odd-numbered TMH for the GPCRs. The nearer neighbors helix periodicity that positions on the same "face" of the helix those residues separated by 3 to 4 other AAs, is reflected in the helical net by the fact that the C^{α} of such pairs are adjacent in position, above and below, respectively. These representations allow the exploration of a-helical periodicity by visual inspection, as described in section 2.1.3.1.2. For example, note the continuous helical patch defined by residues 3-6-10-13-17-20-24-27 shown shaded, which runs parallel to the TMH axis and thus appears on a single face of the TMH. Helical nets representations also identify continuous patches that run at a certain angle from the TMH axis, as the one defined by residues 38-42-46-50-54-58-62 shown by thicker circles. This later patch would result from a tilted and/or supercoiled TMH relative to the membrane axis. Note that by tilting the helical net, the direction of this patch would become parallel to the membrane axis defined by the first helix.

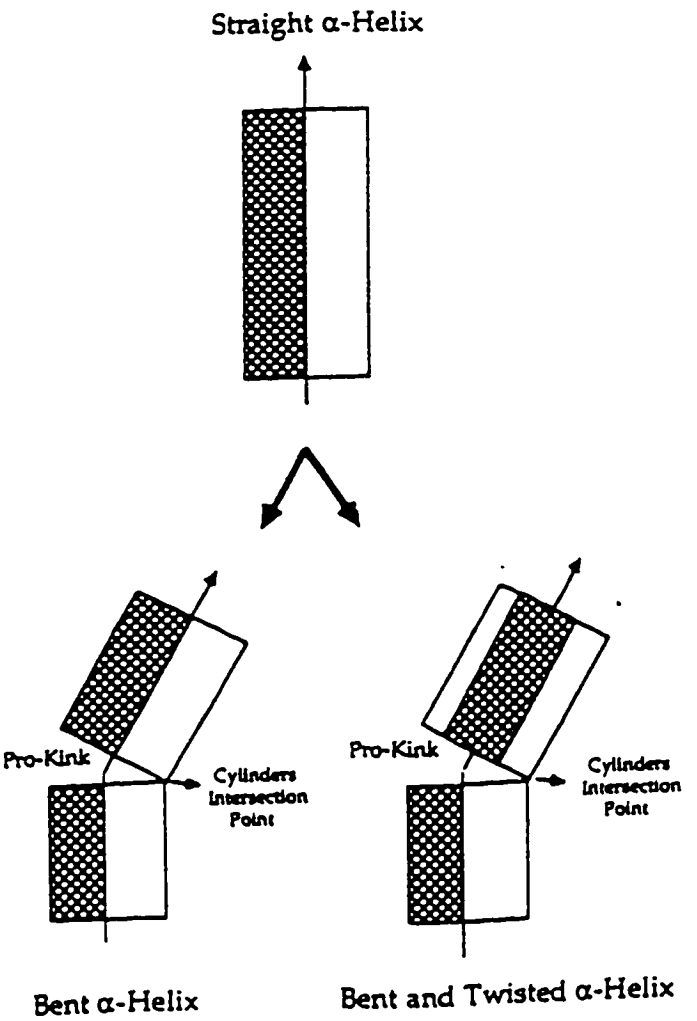


Figure 2.10:
Non-local deviations from an ideal α-helix induced by a Pro-kink, characterized by a bending and twisting of the regular α-helix halves both N- and C-terminal from the Pro-kink region. The TMH is schematically represented by a cylinder with two distinct faces. Note that the twist induced by the Pro-kink on the square face would be expected to offset the periodicities of these two halves, thus hindering the analysis of helical periodicity that assume a continuous helix such as the Fourier transform analysis described in section 2.1.2.1.3.

larger (or smaller) than the average within a window can dominate the Fourier transform and thus induce inaccuracies. They proposed a smoother profile (either conservation or hydrophobicity) by modifying the U(j) values (Donnelly, 1993b).

Based on the special considerations and possible artifacts in the calculation of AP plots, the prediction of TMH boundaries by the Fourier transform method may be incomplete and should be augmented by predictions using other methods presented in this Chapter. Convergence of results from several different methods can be sought to secure the prediction of TMH ends. This is illustrated by a comparison of the TMHs predicted from the AP plots in Figure 2.8 (e.g. 6.27 for TMH6 N-terminus) with the predictions shown in Figure 2.3 obtained using the hydrophobicity profile (6.36 for TMH6 N-terminus). As discussed in Section 2.1.2.1.2 above, the TMH boundaries predicted from AP plots in the cytoplasmic side of TMHs can extend beyond the boundaries predicted by the hydrophobicity profile, due to the Arg/Lys motif. The implication from this comparison is that the hydrophobic core of the TMH predicted by the hydrophobicity profile is an α-helix that is continuous with the cytoplasmic extensions predicted by the Arg/Lys motif since they have the same helical periodicity as measured by Fourier transform. If a break existed in the helix between the hydrophobic core facing the lipid chains and the Arg/Lys region facing the phospholipid head-groups, the corresponding AP values of this non-helical portion would

be significantly smaller. Such a conclusion was reached by Donnelly et al. in their analysis of the TMHs in the antenna complexes (Donnelly, 1993a). Thus, it appears that the redefinition of a TMH as suggested in Figure 2.6 is validated by the analysis of the helical periodicity. Interestingly, the TMH boundaries on the extracellular side show the opposite pattern whereby the AP plots suggest shorter TMHs (e.g. 6.55 for TMH6 C-terminus) than the hydrophobicity plots (6.59 for TMH6 C-terminus). The reason here seems to be the above mentioned inaccuracy in the prediction of extracellular ends by the Fourier transform of TMH 2, 4, 5, and 6 due to the presence of Pro residues close to the end of the helix; in this case predictions made using the hydrophobicity profile would be more reliable than the Fourier transform.

In addition to the use of conservation and hydrophobicity profiles for the analysis of α-helical periodicity using Fourier transform methods, Donnelly et al. proposed the use of another property, namely an estimate of the extent to which a position is exposed to the solvent as evaluated from environment-dependent substitution tables calculated from either lipid facing or aqueous facing residues (Donnelly, 1993a; Donnelly, 1993b). However, no significant improvement was reported for the prediction of TMH boundaries using this set of parameters. Nevertheless, these properties are useful in the prediction of the lipid and interior faces of the TMHs, as described below in Section 2.1.3.1.

2.1.2.1.4- Prediction of TMH segments from α-helix periodicity measured by surface patches on a helical net representation.

Since α-helices have a periodicity of 3.6 AA/turn, individual residues separated by 3 or 4 AA and sharing a given property such as hydrophobicity or conservation would define a continuous patch on the helical surface. Since any such patch is characteristic of an α-helix, the limits of these patches in the sequence are predicted to be the TMH boundaries. For example, Donnelly et al (Donnelly, 1994) predicted TMH ends by the limit of helical patches consisting of AA sites where hydrophilic residues are excluded. This method is analogous to the Fourier transform method described in the previous Section, since both measure the same property -i.e. α-helical periodicity of a given property. The difference between these two methods is that the Fourier transform is applied to a property profile over the entire sequence, rather than to individual residues. This difference allows for a better prediction using helix patches than the Fourier transform for TMH ends close to Pro-kinks. The Fourier transform method has difficulties with such structures for reasons described in the previous Section.

The analysis of helical patches is best undertaken on helical nets, which are a correct two dimensional representation of positions on a three-dimensional cylinder defined by the C_α atoms of an α-helix. Due to their value as analytic tools of helical periodicity as described in this Chapter, Figure 2.9 depicts in detail this mode of representation and its translation to

3-D.

2.1.2.1.5-Analysis of the occurrence of Pro residues within putative TMH segments as indication of helix ends

Due to the characteristic distortion produced by a Pro residue on an α-helix -see Section 2.1.2.2.2 below- single Pro residues within TMHs are generally highly conserved (Williams, 1991). Single Pro residues that are not conserved are common at the ends of helices and can be used for prediction of TMH ends. It has been shown that such single and non-conserved Pro residues can occupy the first three positions at the N-terminus and at most the last position at the C-terminus without significantly distorting the helix (MacArthur, 1991; Strehlow, 1991). The reasons for these position preferences are 1) that an amide hydrogen at the N-terminus in an α-helix is not required to H-bond and 2) that the Pro ring within an α-helix produces a steric clash with the AA in the preceding turn, but not with residues following the Pro. Within an α-helix, Pro-Pro or Pro-X-Pro motifs never occur at the C-terminus and are severely restricted at the N-terminus where they could adopt a poly(Pro) helix conformation (MacArthur, 1991). In fact, the four Pro-Pro motifs that appear in the PRC structure are all in an extended conformation located in extracellular loops. Such a motif appears towards the C-terminus of TMH4 in many neurotransmitter GPCRs.

2.1.2.1.6- Methods for defining TMH ends based on the statistical probability of occurrence of specific AA in TMHs.

A statistical method based on conformational preference functions has been proposed to predict TMHs (Juretic, 1993), in which the probability for each AA in a sequence to belong to a TMH is calculated and used for prediction. The accuracy of this method was around 79% for 5 membrane proteins. Granatir et al (Granatir, 1994) combined this method with hydrophobicity-based methods of the type described in Section 2.1.2.1.1. The authors provided a consensus algorithm where different methods are applied and their respective predictions of TMH ends are compared pursuing a convergent or consensus prediction among them. The statistical combination of different approaches is expected to improve the results due to the inconclusive nature of each of these methods applied individually. A similar conclusion was reached by Edelman (Edelman, 1993) from a more complex mathematical formulation applied to derive TMH predictions based on known structural data. Although all these methods represent algorithmic improvements relative to the hydrophobicity profile, their predictive power does not seem substantially improved. The reason may be that all these methods are restricted to revealing the hydrophobic core of the TMHs either by their definition -e.g., based on calculated hydrophobicity- or through the database from which the frequencies of residue occurrence in TMHs was derived. Because ignoring the Arg/Lys motif can lead to the major errors

described above for BR in the prediction of the TMH boundaries by the hydrophobicity profile, the absence of this additional criterion in all these methods may preclude a significant predictive improvement. For instance, the test of Edelman's method on the BR and PRC structures (29 TMHs total) yielded a total error of 4 residues at the extracellular boundaries, compared to 53 residues at the cytoplasmic side where the Arg/Lys motif is located. It is likely that the addition of the Arg/Lys motif criterion to this type of methods could significantly improve their predictive power.

2.1.2.2- Modeling of the 3D structure of individual transmembrane helices.

Modeling the three-dimensional structure of the helix bundle is greatly facilitated by the availability of structural information about preferred backbone and side-chain dihedral angles in α-helices (Gray, 1984; McGregor, 1987; Piela, 1987a), giving rise to side-chain rotamer libraries (Dunbrack, 1993; Ponder, 1987) that can be used in GPCR modeling (MaloneyHuss, 1992). The allocation of appropriate torsional angles to residues in the TMHs is important because inappropriate rotamers may hinder helix packing interactions. The (φ,ψ) backbone dihedral angles for a TMH are the same as for a soluble protein α-helix judged by the PRC crystal structure (Deisenhofer, 1989), i.e. -60°, -40°. Side-chain dihedral angles can be classified broadly according to their relative energy minima: {-60, 180, 60} for all side-chain angles, except when preceding a bond with pi-electron character such as for

{F,Y,W,H,D,N,E,Q,R} in which case the values are +/-90°. An α-helix environment significantly restricts X_1 for all AA to values around -60° and 180° (McGregor, 1987). The reason is the steric overlap between the side-chain atom in position gamma of residue i and the carbonyl oxygen of the third preceding residue (N-terminal to i), denoted $C{=}O_{i-3}$. The statistical analysis of X_1 distributions in helical environments in high resolution crystal structures clearly supports this generalization, particularly for aromatic side-chains (McGregor, 1987). Two sets of residues have additional specific interactions between their side-chains and the helix backbone that give rise to preferential X_1 rotamers:

Ser/Thr/Cys: The ability of these residues to H-bond their side-chain to the $C{=}O_{i-4}$ of the preceding turn, and the significant preference observed in the PDB for structures with such conformations, constrains the starting rotamers for the modeling purpose to -60° (Gray, 1984; McGregor, 1987). Since this criterion arises from favorable interactions rather than steric hindrances, these residues could also adopt other X_1 rotamers -e.g. X_1=180° for Ser and Cys- although they appear with much lower frequency. Thr residues are generally restricted to X_1=-60° due to the C_β-branched character of the side-chain, as explained below.

Val/Ile/Thr: C_β-branched AA in an α-helix are restricted to one rotamer as

suggested by the relative rotamer populations observed in the PDB structures, as well as by results from computational energy minimizations (see Table III in Ref. (Piela, 1987a)). The reason for this preference is the above mentioned steric clash of a C_γ atom at X_1=60 with the $C{=}O_{i-3}$. Because C_β-branched AAs have two atoms at position γ, the only rotamer that avoids the steric clash at 60° is the one that places the two γ atoms at X_1 values of -60° and 180°. For reasons of nomenclature, this rotamer corresponds to X_1=-60° for Ile and Thr and X_1=180° for Val.

The choices of torsional angles mentioned above as preferred starting values for the construction of TMHs assume standard helical conformations. However, this is not true for TMHs including internal Pro due to the disruption of the helical character by the presence of Pro-kinks (PK), as found in TMHs 2, 5, 6 and 7 of the GPCR. The geometry of the PK regions within α-helices is characterized by backbone dihedral angles that deviate from α-helical values for the residues comprised between the Pro_i and the amino acid at position i-4 from the Pro termed AA_{i-4} (Barlow, 1988). These local distortions produce a highly exposed, and H-bond free $C{=}O_{i-3}$ and $C{=}O_{i-4}$ (Woolfson, 1990). The loss of two helical H-bonds in one turn suggests a significant intrinsic flexibility of the TMH at the PK region. This flexibility has been demonstrated by NMR (Bazzo, 1988; Pastore, 1989) and computational molecular dynamics simulations (Sankararamakrishnan, 1993; Yun, 1991).

There are also non-local deviations from an ideal α-helix induced by a PK, characterized by a bending (Barlow, 1988; Piela, 1987b; Sankararamakrishnan, 1992) and twisting (Ballesteros, 1992a; Sankararamakrishnan, 1992)) of the regular α-helix halves both N- and C-terminal from the PK region, as illustrated in Figure 2.10. The disruption of α-helical character occurs due to steric clashes of Pro rings with the preceding turn, and the loss of a helical backbone H-bond due to the lack of the amide H in Pro. These reasons, and the intrinsic flexibility of a PK, confer to these positions in a TMH some crucial structural, and possibly also functional roles (Ballesteros, 1992b; Sansom, 1992; Williams, 1991; Woolfson, 1991). However, *de novo* structural modelling of PK regions is currently problematic due to three interrelated factors:

1- Their intrinsic flexibility at the level of secondary structure increases significantly the exploration of the conformational space necessary to find the preferred packing.

2- In known structures, PK regions can adopt a variety of (φ,ψ) angles, making it difficult to model PK *de novo* in TMH structures.

3- The non-local disruption of the TMH pattern by PK implies large deviations of the relative 3D localization of the helical portions before and after the PK. The non-local perturbation of helical character can be separated into a bend and a twist of the helical portions N- and C-terminal to the PK, as shown schematically in Figure 2.10. These distortions produce a twisting of

the helix "face" (Ballesteros, 1992a; Sankararamakrishnan, 1992) , as illustrated in Figure 2.10, which may hinder the analysis of helical periodicity as mentioned in Section 2.1.2.1.3 for the Fourier transform method. In known structures, the bending and twisting angles vary between {9°,40°} and {-44°,87°} respectively (Sankararamakrishnan, 1992).

Sankararamakrishnan et al. suggested specific backbone dihedral angles for modeling PKs (Sankararamakrishnan, 1990). Use of these initial angles for the PK is preferable over modeling PK-containing TMHs as ideal α-helices and relying on energy minimizations to produce the appropriate conformations, because the latter approach produces structures that differ from known PK geometries (Nordvall, 1993). Due to the intrinsic flexibility of the PKs, those starting values will usually be modified through the modeling process so as to optimize tertiary structure interactions. For example, a specific H-bond between $Asn_{7.49}$ and the $C{=}O_{i-4}$ from $Pro_{7.50}$ (i.e. $C{=}O_{7.46}$) could dictate the conformation of the PK of TMH7.

The last step in modeling individual TMHs is the conformational search among feasible side-chain rotamers within each TMH, pursuing both the absence of steric conflicts as well as the presence of specific stabilizing intrahelical side-chain to side-chain interactions. For example, Donnelly et al. (Donnelly, 1994) found that $N_{7.49}$ can H-bond the backbone $C{=}O_{7.46}$ of the Pro-kink in TMH7.

2.1.2.3- Secondary structure prediction for the loops, N- and C-terminal domains.

The secondary prediction methods applied to extramembranal portions of GPCRs (Dahl, 1991; MaloneyHuss, 1992) are those commonly used for prediction in soluble proteins, e.g. see (Fasman, 1989; Rost, 1993), and will not be reviewed here. For predicting the conformation of short loop structures connecting TMHs, additional constraints are imposed by the TMHs and these can be used for prediction purposes (see Section 2.1.3.3, below). For example, the identification of sequence motifs favoring specific conformationally defined turns, or the use of conformational selection based on statistical trends, can be combined with energy minimization to model backbone conformations for these regions (Hilbert, 1993; Summers, 1991). A note of caution is necessary in the application of secondary structure prediction methods derived from soluble proteins to extramembranal portions of GPCRs, because interactions between these protein regions and the adjacent membrane environment may lead to physico-chemical situations not considered in the development of such methods for soluble proteins, thus vitiating their predictive powers. This seems to be the case for the cytoplasmic portions of GPCRs, due to the tendency of Arg/Lys residues to be predominantly localized at these regions where they can interact with negatively charged phospholipid head-groups, as discussed above (Section 2.1.2.1.2).

2.1.3- Modeling the tertiary structure of the three-dimensional model of a GPCR.

Modeling the interactions that determine the tertiary structure of the GPCRs can be divided into three procedural steps, discussed below in sections 3.1, 3.2 and 3.3 respectively. The first involves the prediction of the orientation of each TMH relative to the protein and the protein interior. This entails predicting for each TMH the residues that lie in the helix-helix or the helix-lipid interface. Throughout this Chapter, the terms helix-helix or TMH-TMH interfaces is used to refer to the protein interior of the transmembrane helix bundle of GPCRs. The second step involves the packing of the seven TMHs into a bundle that fulfills the predictions from step 1, and represents the transmembrane domain of the GPCR. The third step proceed.- to the completion of the 3D model with protein portions that lie outside the membrane environment, such as the loops and the N- and C-termini.

2.1.3.1- Prediction of TMH-TMH vs. TMH-Lipid interfaces.

All methods for predicting the orientation of TMHs towards the lipid or towards other TMH interfaces rely on the same conservation or polarity criteria that served in the definition of TMH segments (see Sections 2.1.3 and 2.1.4, above). Based on the conservation criterion, the face of the TMH containing more conserved residues is predicted to face the protein interior.

The rationale for this criterion is the common observation that surface-exposed residues are less conserved than interior residues, a prevalent assumption that was corroborated for TMHs in an analysis of the PRC (Rees, 1989b). The second criterion for defining the orientation of each TMH in the helix bundle is based on the complementarity of the surface polarity between a TMH and the membrane lipids. Such complementarity considerations must differ in character for the hydrophobic region of the TMHs compared to the cytoplasmic extensions of these TMHs that contain the Arg/Lys patches: in the hydrophobic core, the more apolar face of the TMH is predicted to be oriented towards the membrane lipids, whereas for the Arg/Lys patches, the more hydrophilic face (containing the basic residues) of the TMH cytoplasmic extension is oriented towards the membrane at the level of the phospholipid head-groups (Figure 2.6). Since all these considerations rely on α-helix periodicity, the methods available for the analysis of the conservation and hydrophobic profile presented above for the detection of helix periodicity (Sections 2.1.3 and 2.1.4) can be used for prediction of TMH-TMH versus TMH-lipid interfaces. While the Fourier transform procedure would thus predict the orientation of a TMH, the identification of TMH patches formed by AAs located at the same interface is very useful in predicting the extent to which each TMH is exposed to the lipids or buried among other TMHs.

2.1.3.1.1- The hydrophobic or conservation moment predicts the relative

orientation of helices in the transmembrane bundle.

The formulation of the Fourier transform analysis of a given property profile -e.g. the conservation U(j), was described in Section 2.1.2.1.3. Donnelly et al. (Donnelly, 1989; Donnelly, 1993b) showed that the same methodology can be applied to predict the TMH orientation. The orientation of the internal, protein-facing side of a TMH can be estimated from the direction of the moment $(P(\omega))^{1/2}$ when $\omega=100^{o}$. This is the moment produced by the property profile U(j) when the sequences form an ideal α-helix. The angle ω describes the direction of the moment relative to the first residue (j=1) and can be calculated as

$$\theta=\begin{cases}\gamma, & \text{if } \left(\sum_{j=1}^{N}(U(j)-U_m)\sin(j\omega)\right)>0\\ 360-\gamma, & \text{if } \left(\sum_{j=1}^{N}(U(j)-U_m)\sin(j\omega)\right)<0\end{cases}$$

$$\gamma=\arccos\left[\frac{\sum_{j=1}^{N}(U(j)-U_m)\cos(j\omega)}{\sqrt{P(\omega)}}\right] \quad \text{where } 0^{o}\le\gamma\le 180^{o}$$

When the conservation or variability profile V(j) is used, the vector of the property moment points towards the protein interior. When the

hydrophobicity profile H(j) is used, the moment points towards the more hydrophobic face of the TMH (Eisenberg, 1984a). However, this direction is not always facing the TMH-lipids interface. Although the hydrophobic core of the TMH the hydrophobicity moment is predicted to point towards the membrane lipid, the Arg/Lys region on the cytoplasmic side the high hydrophilic character of these basic residues inverts the direction of the hydrophobic moment 180 degrees away from the lipid interface. These conflicting orientations predicted for a continuous TMH face explain the surprising observation about the seemingly chaotic direction of the hydrophobic moments calculated for the TMHs in the BR structure (Cronet, 1993). The reason was that the Arg/Lys region of the TMH was included, as it is present in the BR structure -see Figure 2.4- and therefore the calculated moments were a composite of two different moments of opposite direction. as discussed above.

Most authors have applied the hydrophobic moment method to the TMH boundaries as predicted by the hydrophobicity profile, which do not include the Arg/Lys region, resulting in predictions that were consistent with expectations. Thus the surprising situation arises that the hydrophobic moment as originally presented by Eisenberg et al (Eisenberg, 1984a) has yielded useful predictions although in its test with a real TMH structure it failed to predict correctly the TMH orientations (Cronet, 1993). The proper application of this methodology was recently presented by Donnelly et al.

(Donnelly, 1993a) where the presence of the Arg/Lys extensions of the TMH are taken into account and used for prediction purposes. An analysis of the antenna complexes showed the opposite directions of the hydrophobic moment depending on which segment of a TMH is considered. The angle between the moment for the hydrophobic core of the helix and the Arg/Lys extension was indeed close to 180° degrees (Donnelly, 1993a). The authors suggested this 180° difference as a useful tool for prediction purposes. A schematic representation of the relative direction of each property moment is shown in Figure 2.11.

Properties other than the variability and/or hydrophobicity profiles can be used in a similar manner to predict TMH orientation. In particular, Donnelly et al. (Donnelly, 1993a; Donnelly, 1993b) used the estimate of the extent to which a position is exposed to solvent as predicted from environment-dependent substitution tables. These tables were calculated from either lipid-facing (Donnelly, 1993b) or water-facing (Overington, 1992) residues. These two properties are directly related to the prediction of the orientation of TMH with respect to their TMH-TMH versus TMH-lipid interfaces. In fact, the direction of moments calculated for these properties where shown to predict correctly the orientation of the seven TMHs in BR (Donnelly, 1993b).

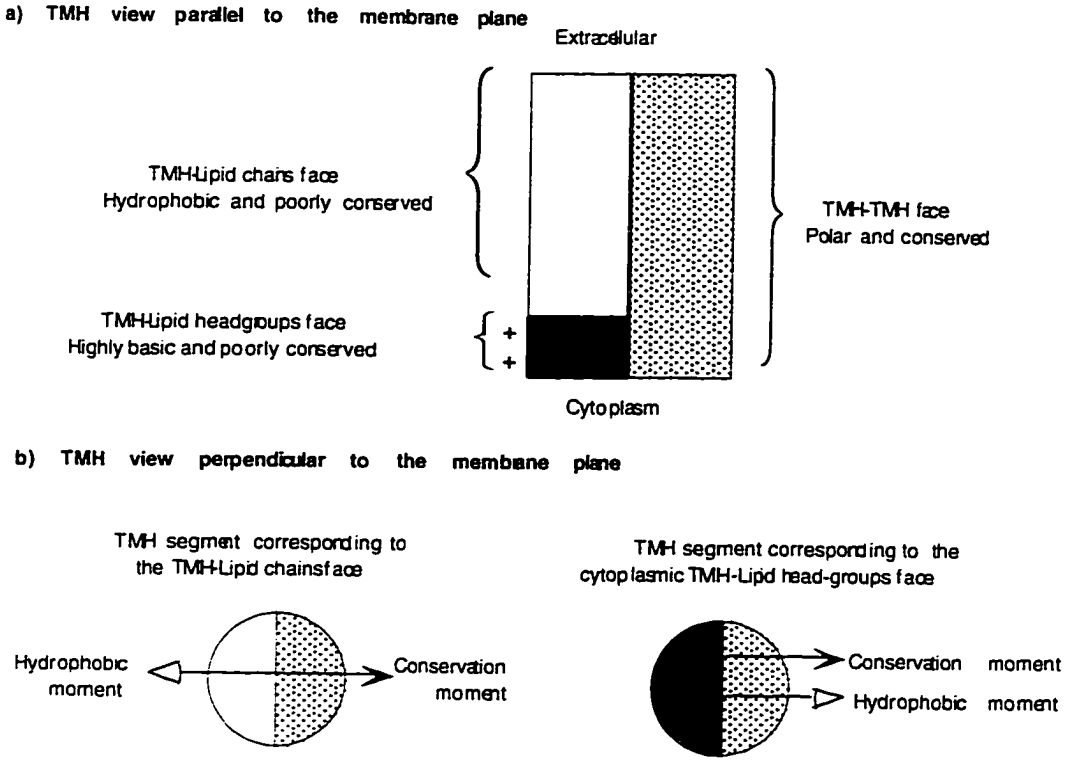


Figure 2.11:
Directionality of the hydrophobic (open arrows) and conservation (closed arrows) moments relative to helix-helix or helix-lipid interfaces. While the conservation moment consistently points towards the protein interior, note the opposite orientation (i.e. 180°) of the hydrophobic moment depending on the helical portion of the TMH. For the hydrophobic core, the hydrophobic moment points towards the TMH-lipid interface. For the Arg/Lys cytoplasmic extensions facing the phospholipid head-groups, the hydrophobic moment point towards TMH-TMH interfaces.

2.1.3.1.2- Patches on the TMH surface identify TMH-TMH versus TMH-lipid interfaces

Similar to their use in identifying helix ends (see Section 2.1.2.1.4) "patches" of AAs sharing a particular property can be useful as an alternative method to predict the face of each TMH that is in contact with either the lipid or the other TMHs. Individual residues separated by 3-4 AAs and predicted from their properties to be likely to face a given environment, e.g. the lipids, will appear as continuous property-sharing patches on a helical net representation. The orientation of such patches defines the predicted interface of a TMH according to the shared property, e.g., the lipid interface identified from a hydrophobic patch. An alternative representation to helical nets for the identification of such patches is the helical wheel representation in which the helix is projected along its axis, as used by Balwin (Baldwin, 1993). To summarize the use of AA patches to identify the nature of TMH interfaces we present first the criteria for predicting a lipid or interior environment for individual AA sites. We then show how these positions can be identified as continuous patches on a helical net representation to become the predicted TMH-TMH or TMH-Lipid interfaces (Baldwin, 1993). Because these predictions are done at the level of single residues, as opposed to a global orientation for a TMH segment provided by the Fourier transform method described above, we will show how this method can be used to estimate the extent to which each TMH is exposed to the lipid phase. This degree of surface

exposure will become a constraint used in the subsequent packing of the TMH into a model of the transmembrane domain of the GPCR, as illustrated in the subsequent Section 2.1.(3.2).

The two major criteria for predicting whether a given residue is facing the lipid milieu or other TMHs, are based on the same two AA properties discussed in previous sections: the conservation and the degree of polarity. Consequently, the same general guidelines apply for the use of these two properties as described in the previous sections. In evolutionary terms, conserved AAs are predicted to be at TMH-TMH interfaces while nonconserved sites are likely to face the lipid environment. The quantification of the degree of conservation as presented in Section 2.1.1 leaves unspecified the criterion of a conservation threshold for prediction purposes, and different authors have used different thresholds. In the most comprehensive conservation analysis presented to date, Baldwin (Baldwin, 1993) considered the extent of conservation at different levels beyond a conservative threshold -i.e. 66% presence of a given AA. Selecting a conservative threshold avoids the exclusion of residues that are significantly, but not absolutely conserved. The subdivision of these selected AA sites into different levels of conservation is useful for the comparison of their respective types of predictions that can be expected: Thus, predictions derived from more conserved positions are stronger than predictions derived from less conserved positions, so that if contradictory predictions arise from any

two AA sites, the prediction derived from the most conserved position prevails. The polar character of an AA site can also serve as a criterion (Zhang, 1994), because polar sites are predicted to be at TMH-TMH interfaces, while hydrophobic AA are predicted to face the lipid. These expectations are validated by data from known structures of transmembrane proteins (e.g. see (Rees, 1989a; Zhang, 1994)). As discussed in the previous sections, this polarity pattern is reversed in the Arg/Lys regions towards the cytoplasmic boundaries of TMHs where these basic residues are predicted to face the membrane environment at the level of the phospholipid head-groups. Although similar conservation or polarity criteria are used in the helical patches method and the Fourier transform method (see Section 2.1.3.1.1), the former has the advantage that the criteria can be applied to only a few AA sites within TMHs, because a continuous profile over a sequence segment is not required. This allows not only the modification of criteria based on conservation or polarity properties, but also makes possible the use of qualitatively different criteria such as experimental data on ligand or lipid interaction sites for GPCR. These three sets of criteria are described below:

(1) The conservation analysis as presented in Section 2.1.1 identifies specific residues that can be predicted to lie either at lipid- or interior-facing positions. For example, Baldwin (Baldwin, 1993) predicted that divergent residues among highly homologous and functionally equivalent GPCR would face the lipid environment, because these sites are likely to occur in

functionally non-important positions. Baldwin also proposed the conservation analysis at the level of physico-chemical properties rather than chemical identity. Thus, she identified residues that conserve a similar size during evolution and predicted them to face TMH-TMH interfaces (Baldwin, 1993). Zhang and Weinstein (Zhang, 1994) analyzed the conservation of polar character at positions where more than 90% of the GPCR sequences considered retained polar residues. These positions, called "polarity conserved positions" (PCPs) were predicted to be at TMH-TMH interfaces, as verified by the structure of BR family (Zhang, 1994).

There are notable exceptions to the validity of predictions using the conservation analysis. An accepted exception to the rule that conserved residues point towards the protein interior is the case of Pro residues because they can exert their effects from a lipid facing position (Baldwin, 1993; Donnelly, 1993b), as discussed in Section 2.1.2.1.3. Similarly, an exception to the rule that a poorly conserved AA site is facing the lipid would be the case of a residue responsible for functional divergence, such as ligand binding; such a residue is expected to face the interior of the TMH bundle in the ligand recognition pocket.

(2) The criterion of a polar complementarity between the TMH and the lipid phase can also identify specific residues predicted to face the lipid or protein interior phase. For example, Donnelly et al. (Donnelly, 1994) identified a continuous patch of AA sites not containing polar residues and

predicted them to face the lipid. Similarly, Arg and Lys residues at the cytoplasmic boundaries are predicted to face the phospholipid head-groups (Ballesteros, 1992a). A note of caution is necessary in defining the polar character of each AA in a TMH. The polar character of the 20 natural AAs that is relevant to the predictions discussed here may differ from the polarity assigned by hydrophobicity scales, due to specific interactions within TMHs. Such is the case of Ser, Thr, Cys, Pro and Tyr. In the first group, {Ser, Thr, Cys} residues are generally found (Gray, 1984) to H-bond back to the carbonyl backbone of the preceding turn, i.e. C^{β}-O(S)$^{\gamma}$-H...O=C_{i-4}. Such as arrangement decreases the polarity of the face they present for interaction so that this AA site can be exposed to the lipid milieu (Gray, 1984). For example, 45% of the Ser residues in BR are exposed to the lipid phase. In the second group, Pro residues induce a characteristic kink in TMH, as described in Section 2.1.2.2, that results in the two carbonyls of the preceding turn being highly exposed and H-bond free. This arrangement enhances the polar effect of a Pro residue relative to the hydrophobicity scale. This indirect polar character of Pro residues would suggest that they are likely to be positioned facing the protein interior, as suggested by Woolfson (Woolfson, 1991) based on an analysis of the hydrophobic moments of Pro-containing α-helices. On the other hand, Tyr residues have been observed facing the lipid head-groups in the structures of BR (Henderson, 1990a) and bacterial porins (Cowan, 1992; Weiss, 1991).

(3) Some experimental data on GPCR are available concerning the exposure of AA sites to the membrane lipid or the protein interior. Such information includes i) the identification of covalent attachment sites for activated probes; ii) the identification of residues proposed from mutagenesis experiments to be in direct contact with the ligand; iii) data on specific TMH-TMH interactions implying that the residues involved should face the protein interior. In particular, experiments on adrenergic GPCR identified positions $W_{7.40}$ (Wong, 1988) and $\{S_{2.63}$ and/or $H_{2.64}\}$ (Dohlman, 1988) as ligand attachment sites from alkylation experiments, thus identifying AAs lying at TMH-TMH interfaces. Covalent attachment AA sites identified for the bovine rhodopsin GPCR belong to two classes, from hydrophilic (Findlay, 1984) and hydrophobic (Davison, 1986a; Davison, 1986b) probes respectively. The hydrophilic probes were predicted to label AA sites exposed to the aqueous phase or at least lying in the phospholipid head-group region. The hydrophobic probes, on the other hand, were predicted to label AA sites exposed to the lipid milieu. The alkylation experiments thus provided a powerful discriminant of TMH-TMH versus TMH-Lipid interfaces. Using similar arguments, AA sites proposed to lie in direct contact with receptor ligands based on mutagenesis experiments were also useful to predict TMH-TMH interfaces, e.g Maloney-Huss et al (MaloneyHuss, 1992), Hibert et al (Hibert, 1991) and Baldwin (Baldwin, 1993). However, the strength of those predictions is less than those obtained from covalent attachment sites because

they are not supported by evidence of direct contact. Finally, some considerations were proposed for predicting specific interactions among TMHs based on interactions of specific AAs. Such is the case of the interaction between the conserved $D_{2.50}$ and $N_{7.49}$ loci. A double-revertant mutant of the gonadotropin-releasing hormone receptor (GnRHR) (Zhou, 1994) showed that the pair $N_{2.50}/D_{7.49}$ in the GnRHR has the same ligand binding properties as the D2.50/N7.49 pair found in most GPCRs, thus identifying a TMH2-TMH7 interaction. These inferences will be discussed below in the context of modeling the packing of TMHs in the transmembrane bundle.

Once the prediction of a tertiary environment has been achieved for individual AA sites, the regularity of an α-helix implies that residues predicted in a common environment -either TMH-TMH or TMH-Lipid- should lie on a common face of the TMH. The consistency of the predicted TMH-TMH or TMH-Lipid environment at the level of individual AA sites was shown by Baldwin using helical wheel representations for the seven TMHs (Baldwin, 1993). As an illustration, her resulting predictions are shown for TMH6 in Figure 2.12.a. Note that when the Arg/Lys criterion is added to the Baldwin predictions, the resulting helix-lipid interface is continuous.

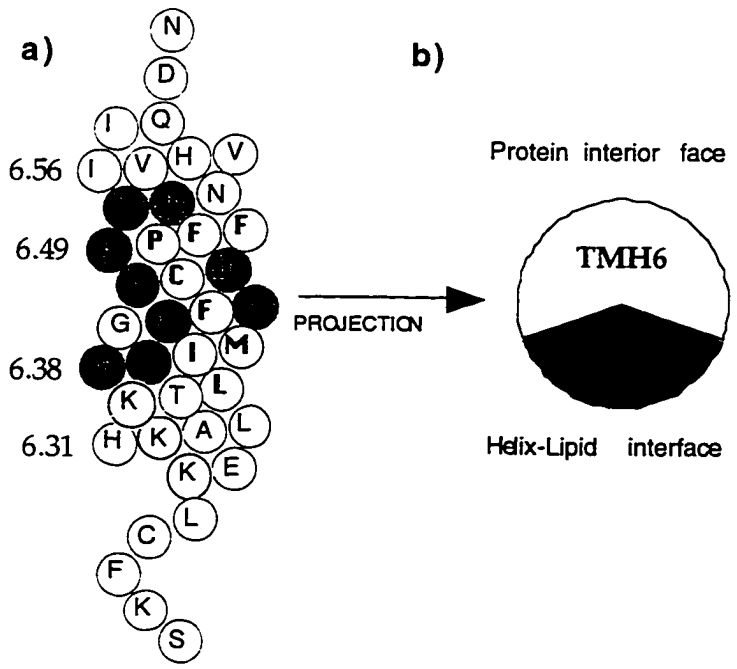


Figure 2.12:
Prediction of TMH-TMH versus TMH-Lipid interfaces shown on a helical net representation (a) and its projection on a helical wheel representation (b) of TMH6 of the human β_2-adrenergic GPCR. a) Highlighted are the predictions made of AA facing the lipids (shaded) or the interior of the TMH bundle (bold letters) by Baldwin (J.M. Baldwin, Embo J 12, 1693 (1993)), as well as the Arg/Lys motif towards the cytoplasmic boundaries (thicker circles) predicted by Ballesteros and Weinstein (J.A. Ballesteros and H. Weinstein, Biophys J 6. 107 (1992)). Note the remarkable consistency among different prediction methods, so that a continuous TMH between the hydrophobic core and the Arg/Lys cytoplasmic extension is supported by the continuity in the helical periodicity suggested by a common TMH-lipid interface. b) A helical wheel projection of the individual AA predictions shown in (a) serves to characterize the degree of surface exposure in section 2.1.3.2.1.

From these predictions of TMH interfaces, one can estimate the extent to which each TMH is exposed or buried within the TMH bundle. A helical wheel projection of the individual AA predictions, as shown in Figure 2.12.b for TMH6, serves to characterize the degree of surface exposure. Because those predictions should be consistent with the final 3D-model, the predicted extent of lipid exposure (or TMH exposure) of any given helix becomes a constraint for modeling the helix packing arrangement of the seven TMHs (see Section 2.1.below). An extensive analysis of a sequence alignment of 204 GPCRs leading to comprehensive predictions of the degree of surface exposure for each TMH has been presented by Baldwin (Baldwin, 1993); the degree of lipid exposure in decreasing order was TMH 1, 4, 5, followed by TMH 6, then TMH 2, 7, and finally TMH 3 which is predicted to have very little lipid-facing area, mostly at the extracellular end. The author used these predictions (combined with other criteria) to identify each TMH on the experimentally determined electron density projection map of rhodopsin at 9 Å resolution. It should be noted that the degree of surface exposure for a TMH may vary along the membrane height, i.e., a TMH could be more exposed on the extracellular than on the intracellular side of the membrane.

No single packing arrangement is expected to fulfill all the predictions for each AA site on a helix for which a TMH-TMH or TMH-Lipid environment have been defined, because the approach is integrating a set of predictions emerging from different criteria and considerations. Therefore, as

for the predictions at the secondary structure level, conclusions are based on the comparison and mutual reinforcement among different methods. Convergent or contradictory structural inferences from such an analysis can be integrated and explored in a 3D-model at atomic detail (see below).

2.1.3.2- Modeling the tertiary structure of GPCR: Packing the TMHs in the transmembrane bundle.

Modeling helix-helix packing interactions among the seven TMHs leads to a 3D-model of the transmembrane domain of a GPCR, a major objective of any receptor modeling strategy. Until recently, this modeling step has been the subject of controversy over the use of the known structure of BR as a suitable template for GPCR (e.g., see (Baldwin, 1993; Hoflack, 1994; Pardo, 1992; Zhang, 1994). The use of the structure of BR as a template for GPCR modeling requires the assumption of a common 3D-fold between these two distant families of transmembrane proteins. Its attractivity is that it significantly simplifies the modeling process, since the seven TMHs of a GPCR can be superimposed onto the BR backbone following an alignment of their respective TMH sequences (Cronet, 1993; Dahl, 1991; Findlay, 1990; Hibert, 1991; IJzerman, 1992; Lewell, 1992; Livingstone, 1992; Nordvall, 1993; Trumpp-Kallmeyer, 1992; Yamamoto, 1993). The arguments supporting the structural similarity are the following: first, BR and the opsin GPCR seem to be evolutionarily related, as reflected by a similar architecture based on seven

TMHs surrounding a retinal chromophore attached through a Schiff-base to a Lys residue on TMH7. This functional similarity in the "ligand" and the conformational changes associated with it during activation by light do not extend to the signal transduction mechanism; BR pumps protons out of the cell while rhodopsin (RH) activates the G-protein transducin. Although there is no detectable sequence homology between the BR and GPCRs (Henderson, 1990b), detectable homologies have been found if evolutionary events such as exon shuffling (Pardo, 1992) and/or gene duplication (Taylor, 1993) are hypothesized. However, these similarities do not in themselves suggest a specific alternative 3D-fold for the GPCR because only few TMHs are involved. The possibility remainded that the AA sequences of BR and GPCR would be similar at the level of physico-chemical properties in the absence of direct AA homology, but a recent report indicates that the pattern of polarity conserved positions through their sequences is clearly divergent (Zhang, 1994). Since helix-helix interactions are significantly driven by complementarity of their polar-apolar surfaces, this result suggests that the TMH packing arrangement is likely to be significantly different between BR and the GPCR.

Recently, the projection map of the electron density of bovine RH confirmed the suspected divergence in the packing arrangement of the seven TMHs in BR compared to GPCRs (Schertler, 1993). However, the low resolution (9 Å) of the RH projection map precluded the experimental

identification of the individual TMHs as well as the identification of the view as cytoplasmic or extracellular. Although the helix packing arrangements appeared to be significantly different in RH compared to BR, the low resolution prevented a more detailed comparison between the structures. A comprehensive attempt to identify the view side and the individual TMHs of RH in the projection map has been undertaken by Baldwin (Baldwin, 1993); the results share with most GPCRs models and with BR the sequential and anti-clockwise order (viewed from the extracellular side) of the seven TMHs. However, the details of the helix arrangement and packing are different, as discussed by Baldwin (Baldwin, 1993).

Currently, a GPCR model is expected to comply with the projection map of RH, although due to the short time since its publication only two such models have been presented to date, Baldwin's (Baldwin, 1993) and ours (Zhou, 1994) (Figure 2.13). The projection map of the seven TMHs for RH thus becomes a starting point for modeling helix-helix interactions. Having modeled the seven individual TMHs as described above, methods can be used to position the individual TMHs on the projection map of RH as described in Section 2.1.3.2.1, as a point of departure for modeling the transmembrane bundle. *De novo* modeling methods that can be used to infer the relative axial displacement among the seven TMHs to produce a specific 3D model at atomic detail are described in Section 2.1.3.2.2. Methods used to model GPCRs based on the structure of BR as a template are described in Section 2.1.3.2.3,

since a similar strategy would be useful once a comparable 3D structure of RH becomes available, and because of the interest these models have evinced as hypotheses generators in mutagenesis studies.

In developing a general method for helix-helix packing in the transmembrane helix bundle of GPCR, a note of caution is necessary regarding the assumption that all GPCRs share a common template. The ability to predict structural similarity is actually modulated by two factors: First, it would depend on the extent of sequence homology across members of the superfamily, 50% having been proposed as a threshold for structurally homologous proteins (Chothia, 1986). Second, it would depend on the extent to which the presence of AA substitutions can be expected to entail significant structural modifications, such as the non-local structural effects of Pro in TMHs discussed above, or new anchors for TMH-TMH interactions. As GPCR subfamilies differ in the degrees of homology among them as well as in the presence and relative positions of Pro residues in TMHs 2, 4 and 5, significant differences in the TMH bundle packing may occur, and these may have been exploited by evolution for functional differentiation.

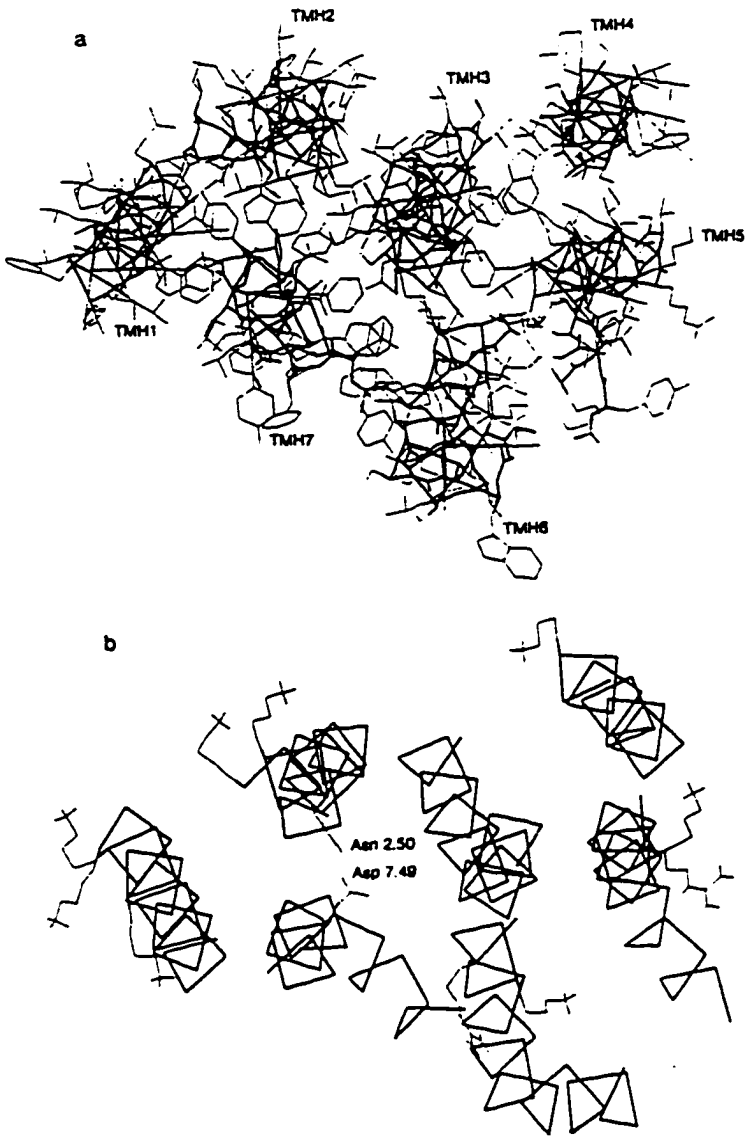


Figure 2.13:
(a) Our model of the gonadotropin releasing hormone receptor. This model was constructed so as to comply with the electron density projection map of RH and was based on criteria described in this Chapter. (b) The criterion that highly conserved residues interact among themselves is illustrated for a proposed interaction between $N_{2.50}$ and $D_{7.49}$ which received experimental support through the double-revertant mutant character observed experimentally (W. Zhou, C. Flanagan, J.A. Ballesteros, K. Konvicka, J.S. Davidson, H. Weinstein, R.P. Millar, and S.C. Sealfon, Mol Pharm 45, 165 (1994)).This criterion focuses attention to clusters of conserved AA in the 3D model that are very useful in predicting the detailed helix packing among the TMHs involved. Note the presence of Arg/Lys at the cytoplasmic side facing the lipid head-groups, as discussed in Sections 2.1.2.1.2 and 2.1.3.1.

2.1.3.2.1- Inferences for TMH packing from methods used to identify the individual TMHs on the projection map of RH.

Baldwin (Baldwin, 1993) identified the TMHs on the RH projection map (Schertler, 1993) based on three criteria: First, the minimum length of all inter-helical loops over the whole GPCR family (6 to 13 AA) was considered short enough to suggest a sequential packing arrangement where each TMH is in contact with the two adjacent TMHs in the sequence. Second, the extent of lipid exposure predicted for each TMH -see Section 2.1.3.1.2- should be consistent with the observed lipid exposure for each TMH on the RH map. Thus TMH3 would be most buried, TMH2, 6 and 7 have intermediate exposure, and TMH1, 4 and 5 are most exposed. Application of the surface exposure criteria assuming a sequential helix packing arrangement led to two alternative assignments of the seven TMHs in the RH map. These alternatives differ in having either clockwise or anti-clockwise connectivity when viewed from the intracellular side of the membrane, as shown in Figure 2.13.a for the clockwise arrangement prepared after Figure 3 of Baldwin (Baldwin, 1993). Note that the solvent accessible surface for each helix around the TMH bundle agrees fairly well with the predicted degree of lipid exposure, perhaps with the exception of TMH2. Third, criteria based on proposed ligand-GPCR interactions favored the clockwise connectivity (Figure 2.13.a) because the anti-clockwise could not explain a variety of proposed sites of interaction for the retinal (in RH) and adrenergic ligands based on

mutagenesis experiments. The required consistency between the predicted (GPCR) and the observed (RH) degree of lipid exposure for each TMH provides yet another criterion for testing the suitability of BR as a template for GPCR (Baldwin, 1993). The predicted degree of surface exposure for each GPCR TMH is not consistent with the observed degree of lipid exposure for each TMH in BR, as shown in Figure 2.13.b (Figure 5 in (Baldwin, 1993)); note that the solvent accessibility surface around the TMH bundle indicates that according to the GPCR properties, TMH 2, 3 and 7 in BR are overexposed while TMH4 is underexposed.

2.1.3.2.2- Modeling TMH-TMH packing on the RH template

Throughout this Chapter several methods incorporating a variety of criteria have been presented at every structural level, up to the level of TMH-TMH packing in three-dimensions in an atomic resolution model of the TMH bundle in GPCR. Having predicted the relative positions of the seven TMHs in a 2-dimensional projection on the membrane plane from the RH map, modeling considerations now face the problem of the axial displacement of each TMH in the TMH bundle, and the detailed residue-to-residue interactions among them. An initial axial positioning of each TMH can be accomplished by aligning the respective TMH ends positions so as to maximize TMH-TMH contacts. If the TMH ends have been predicted from the hydrophobicity profile, the resulting axial displacements of

the TMHs relative to each other corresponded to their hydrophobic cores, thus missing the helical portions facing the phospholipid head-groups defined by the Arg/Lys motif (see Section 2.1.2.1.2). The error in this initial relative axial displacement among adjacent TMH has been estimated to be 3-4 residues (e.g., (Baldwin, 1993)). However, larger errors can be expected due to the topological dependence of the cytoplasmic extensions of TMH based on the Arg/Lys motif as discussed in Section 2.1.2.1.2. This was shown for BR in Table 2.2, where the average error in the predicted TMH end at the cytoplasmic side with an N_{in}-->C_{out} topology is 7 AAs, while for TMHs with an N_{out}-->C_{in} topology the average error is minimal. As a result, because any two TMHs adjacent in the GPCR sequence have an opposite topology, their respective cytoplasmic helix ends (predicted from the hydrophobicity profile) would be in average offset by 7 residues or two turns of an α-helix , thus inducing a similar error in the subsequent helix-helix packing. Notably, if a length of 22-28 AAs is assumed for a TMH, the error in the modeled axial displacement between adjacent TMHs predicted only from the hydrophobicity profile, could amount to 25-30%. This error induced by the Arg/Lys motif would be predicted to affect current GPCR models whose TMH boundaries were derived from hydrophobicity criteria (Cronet, 1993; Findlay, 1990; Hibert, 1991; Hoflack, 1993b; Huggins, 1993; IJzerman, 1992; Kontoyianni, 1993a; Lewell, 1992; Livingstone, 1992; MaloneyHuss, 1992; Nordvall, 1993; Trumpp-Kallmeyer, 1992; Trumpp-Kallmeyer, 1993; Yamamoto, 1993; Zhang, 1993b). In

a most illustrative example of such possibility, this error was shown to vitiate models of the BR TMH bundle in which TMH boundaries were predicted by hydrophobicity criteria before the structural elucidation of BR (Jahnig, 1990). In an interesting molecular dynamics approach to packing the seven TMHs of BR based on the electron density projection map of the helices on the membrane -an approach that is very similar to the current situation with RH-, the authors (Jahnig, 1990) found that the best model could be improved both in energy and in residues involved in retinal binding by shifting TMH6 two turns upwards. Figure 2.7 illustrates that these two turns in TMH6 actually correspond to the helical portion of TMH6 missed due to the Arg/Lys motif. Furthermore, the TMH6-TMH7 packing in the structure of BR shown in Figure 2.7 illustrates that the actual helix-helix interface is offset from the cytoplasmic ends of the hydrophobic segment by seven AAs -see predicted vs. observed TMH boundaries for these. When the same authors repeated these calculations with the same predicted TMH boundaries after the structure of BR was elucidated, they reached similar conclusions (Jahnig, 1992). Interestingly, in this case they could modify the cytoplasmic end of TMH6 to include the previously missed Arg/Lys portion, thus improving the accuracy of their models (Jahnig, 1992). Although we have used the Arg/Lys redefinition of the TMH boundaries as an illustrative case, the general conclusion is that inaccuracies in the prediction of TMH boundaries could seriously affect the modeling of TMH packing arrangements.

The problems described above for the modeling of TMH-TMH packing, can be circumvented with methods to predict interhelical residue-to-residue interactions that could guide the TMH packing independent of the selected TMH boundaries. This is a most difficult step in the modeling process, as illustrated by the fact that few approaches have been offered to determine specific residue to residue contacts among the seven TMHs. Even in cases where specific residues are predicted to interact, or to bind a ligand, the resulting constraints imposed by such information on the possible helix packings at atomic level are rather weak. The ambiguity arises from the intrinsic conformational degrees of freedom of the residue side-chain and/or the ligand, as well as the different possible orientations of the TMHs in a 3D space that can satisfy the proposed interactions. As a consequence, a similar set of criteria considered by several authors has yielded significantly different 3D models of highly homologous GPCR. For example, significant differences are present in the details of models of GPCR that have used a common 3D template, such as BR, to infer the interhelical contacts -see Section 2.1.3.2.3 below- even within the same laboratory (Hibert, 1991; Trumpp-Kallmeyer, 1992). The lack of appropriate theoretical methodologies is echoed in the lack of appropriate experimental data to derive information on inter-helical contacts. Notably, recent attempts to find double-revertant mutants in GPCR has shed some light in what promises to become the most valuable source of experimental testing of GPCR models. The assumption of spatial proximity

between residues shown to display a double-revertant mutant character is based on the hypothesis of direct interaction among them (Ward, 1990). Thus, we present below first those methods that are useful to predict residue-to-residue interhelical contacts, and then a description of the two reported double-revertant mutants for the GPCR family. The methodology used to predict specific interhelical contacts for GPCRs is based on the proposed ligand-interaction sites, as well as polarity or conservation criteria.

Proposed ligand-GPCR contacts have been widely used in modeling TMH-TMH packing arrangements. The most solid case is the use of covalently bound ligands where the attachment site has been identified experimentally. The retinal attachment site ($K_{7.43}$) in rhodopsin is an example in which the combination with the proposed counterion to the Schiff base ($E_{3.28}$) leads to constraints in the relative position of TMH7 and TMH3 -see Oprian (Oprian, 1992) for a review. Similarly, Maloney-Huss et al (MaloneyHuss, 1992) used structural inferences from the binding of a pindolol derivative that labels $Ser_{2.63}$ and/or $His_{2.64}$ of the β_2-adrenergic receptor to infer on the relative positioning of TMH2 and TMH3; this ligand could be accommodated in their model structure with the alkylation group pointing towards residues 2.63-2.64 and the protonated amine making an ionic bond with the $Asp_{3.32}$ in TMH3. The proposed ionic interaction between the protonated amine moiety of neurotransmitters and the corresponding

conserved $Asp_{3.32}$ in the cationic neurotransmitter GPCRs is the most widely accepted ligand-receptor interaction site inferred from mutagenesis studies.

Another set of widely accepted ligand interaction sites from mutagenesis experiments involve Ser and Thr residues occurring at positions 5.42-5.43-5.46 in a variety of neurotransmitter GPCRs (Gantz, 1992; Ho, 1992; Kao, 1992; Link, 1992; Mansour, 1992; Pollock, 1992; Strader, 1989). These three positions, that define a common patch on the TMH surface (see Figure 2.2), are differentially used in homologous receptors for ligand binding and/or activation. The strongest support for a direct interaction of some ligands at this site came from systematic double modification of the hydroxyl groups in the adrenergic catecholamine ligands and the β_2-adrenergic receptor (Strader, 1989).

The other AA site extensively studied by mutagenesis experiments, that has been proposed to be in direct contact with the ligand for a variety of receptors (Smolyar, 1993), is position 7.39. Interestingly, the RH projection map with the TMHs identified by Baldwin (Baldwin, 1993) brings close in space the key residues $D_{3.32}$, $S/T_{5.42\text{-}5.43\text{-}5.46}$ and $AA_{7.39}$, consistent with a large body of pharmacological and structure-activity data pertinent to the nature of the ligand recognition sites on GPCRs.

Molecular models of the recognition site obtained earlier from structure-activity considerations and computational probing have also served as a source for predicted ligand-receptor interaction sites. For example,

following a precise definition of the residues forming the binding site of the 5-TH_2 GPCR, Zhang and Weinstein (Zhang, 1993b) used this approach as a primary criterion to model the TMH bundle.

The criterion of residue polarity predicts a preferential interaction of residues with similar polarity, i.e. polar-polar and apolar-apolar contacts. Yet this criterion in itself is rather general in the context of putative interactions among the many residues of each kind in the TMHs. Additional constraints are needed to predict specific residue-residue contacts. The constraints established so far for the packing arrangement of the seven TMHs significantly restrict the possibilities of residues interacting based on their polar character. A clear example is provided by GPCR modeling based on the BR template -see below- where the superposition of the GPCR TMHs onto the BR backbone severely restricts the possible combinations of polar-polar interactions within the TMH bundle. The same criterion can become a test of the model: Maloney-Huss and Lybrand (MaloneyHuss, 1992) found a glutamic acid in position 3.39 that had no polar counterpart in their model and was significantly exposed to the lipids. An explanation is provided by the projection map of RH which indicates that TMH3 at the level of $E_{3.39}$ is significantly more buried among the other TMHs than initially predicted by these authors, calling for a refinement of the packing in this region of the model. Zhang and Weinstein (Zhang, 1994) proposed to combine the polarity criterion with the conservation analysis by identifying AA sites that

conserved their polarity across more than 90% of the aligned GPCR, and called these loci "polarity conserved positions" (PCPs). When applied to the BR structure, these PCPs were shown to preferentially interact among themselves compared to other polar residues within the sequence (Zhang, 1994).

The last set of criteria to predict residue-to-residue interactions to be discussed here as guides of the packing of the seven TMHs is based on the degree of evolutionary conservation. The final 3D-model is expected to explain the observed degrees of conservation for each AA site, which becomes a constraint for modeling the detailed packing of the seven TMHs. Because highly conserved residues are candidate molecular determinants of the structural and/or functional integrity of the GPCR, the final 3D-model should encompass specific interactions involving these residues that are responsible for their apparent structural and/or functional requirement through evolution. These molecular determinants exert their effects through physico-chemical properties such as charge, aromaticity, volume, H-bonding ability, so that the characteristic properties of each conserved residue should be engaged in specific interactions within the model. For example, a residue that conserves its volume through evolution should be significantly buried, surrounded by a closely packed environment. Similarly, the acidic character of $D_{2.50}$ should have a counterpart within the TMH bundle that stabilizes a charge in a protein interior environment. Because the counterpart of a

conserved residue should be equally preserved through evolution, highly conserved residues are predicted to interact among themselves. This criterion focuses attention to clusters of conserved AA in the 3D model that are very useful in predicting the detailed helix packing among the TMHs involved. An example of such clusters involves TMH1, 2 and 7 at the level of the conserved residues {$N_{1.50}$, $D_{2.50}$, $N_{7.49}$, $P_{7.50}$, $Y_{7.53}$} as proposed by several authors (Donnelly, 1994; MaloneyHuss, 1992; Trumpp-Kallmeyer, 1992; Zhang, 1993b; Zhou, 1994). An illustration of such a predictive criterion is given in Figure 2.13 for a model of the gonadotropin releasing hormone receptor where a proposed interaction between $N_{2.50}$ and $D_{7.49}$ received experimental support trough the double-revertant mutant character observed experimentally (Zhou, 1994), as outlined briefly below in the context of coordinated conservation patterns.

A recent variation of the conservation analysis seems especially well suited for the prediction of tertiary structure interactions. Thus, Benner et al. (Benner, 1991) and Altschuh et al. (Altschuh, 1987) have proposed to identify for this purpose the AA sites where mutations during evolution show a coordinated or parallel pattern in which mutations at one site are accompanied by mutations at a distant position. For example, the conserved AA sites 2.50 and 7.49 appear to be an example of such evolutionary correlation, because when $D_{2.50}$ changes to $N_{2.50}$ in the gonadotropin releasing hormone receptors (GnRHR), then residue $N_{7.49}$ changes to $D_{7.49}$.

Residues that display a coordinated AA substitution pattern are thus predicted to interact among themselves, either through functional coupling and/or through direct contact (Ward, 1990). The initial translation of the steric constraints offered by these coordinated AA sites into 3D structures is based on the simple assumption of direct interactions, because this hypothesis was found to be almost always correct for 5 different protein families with known structures (Altschuh, 1987; Altschuh, 1988; Pastor, 1990). Benner et al. have applied the same criteria for structural predictions, achieving a remarkable success in the secondary structure prediction of the catalytic domain of protein kinases (Benner, 1991). From an evolutionary standpoint, step-by-step AA substitutions imply successive adaptations of the structure, the driving force being to optimize function. Assuming a common framework for the GPCR within the alignment means that the structural changes during this adaptation process are local in character, without gross alterations of TMH-TMH packing arrangements. This hypothesis allows the evaluation of which structural frameworks, i.e. which TMH packing arrangements, are consistent with the set of identified correlated AA sites. It should be noted that the initial criterion that conserved AA interact among themselves can now be seen as a particular case of co-ordinated AA substitution patterns. Although the structural hypothesis of spatial proximity between co-ordinated AA sites has been applied to infer tertiary structure contacts, the correlated mutation pattern may arise also from functional

coupling among non-interacting residues. The identification of such related sites may provide a molecular basis for understanding how functional diversity is achieved across a protein family (Benner, 1991) such as the GPCRs.

Experimentally, tertiary structure interactions can be probed by the construction of double-revertant mutants. If the phenotype of a single mutant can be reverted back to wild-type by a second mutation at another site, it is likely that these two residues interact with one another, the simplest manner being a direct contact. Kobilka's laboratory (Suryanarayana, 1992) proposed an interaction between position 7.39 (TMH7) and TMH1-TMH2 based on chimeric constructs between adrenergic receptors; the deleterious phenotype of a N7.39F mutation in the β_2-adrenergic receptor was suppressed by chimeras containing the first two TMHs of the α_2 subtype. Applying the same logic, Pittel and Wess (Pittel, 1994) proposed the adjacency of TMHs 1 and 7 in the muscarinic GPCR. More recently, double revertant mutagenesis studies on RH by Rao et al. (Rao, 1994) suggested that the retinal attachment site, $K_{7.43(296)}$, is in proximity to $G_{2.57(90)}$ in addition to its wild-type counterion $E_{3.28(113)}$, thus positioning TMH7 adjacent to TMHs 2 and 3. Notably, double revertant mutants can be predicted as a special case of coordinated AA substitutions, when two AA sites predicted to affect function significantly have a coordinated AA substitution pattern that is complementary in nature and is restricted to two residues within the sequence (see above). The simplest

case would be a reciprocal exchange between two conserved AA sites, as the above mentioned $D_{2.50}$->$N_{2.50}$ and $N_{7.49}$->$D_{7.49}$ occurring in the GnRHR. Zhou et al. (Zhou, 1994) showed the double-revertant mutant character between these two loci in the GnRHR, thus providing a bridge between theoretical prediction methods and the equivalent experimental counterpart, used to validate the models. Note that distinct double-revertant mutant experiments involve different residues within the same TMHs (i.e. 7.39, 7.43-2.57-3.28, 7.49-2.50), so that when combined these experiments provide significant structural constraints for modeling the relative placement and mutual interactions of the TMHs involved.

A seemingly powerful and stringent test of a 3D model of the transmembrane domain of a GPCR is to require a packing density for interior residues similar to that commonly observed for known protein structures (Kontoyianni, 1993a; MaloneyHuss, 1992), including membrane proteins as shown for the PRC (Rees, 1989b). These techniques, as implemented in the program QPACK (Gregoret, 1990), have been applied in the Lybrand laboratory to provide with one of the most stringent and helpful tools for the modeling process (Kontoyianni, 1993a; MaloneyHuss, 1992). While it is always possible to pack seven TMHs into a model of the transmembrane domain of a GPCR, it is quite difficult to attain a closely packed interior of the seven helix bundle that resembles the packing density found for known structures (Chothia, 1981). The application of this method to probe and

improve current GPCR models is likely to increase significantly in the near future, since it provides not only a valuable modeling tool, but also a quantitative measurement of the "correctness" of the modeled structures that constitutes an objective measure for comparisons among different models.

2.1.3.2.3- Assembly of the TMH bundle according to a BR template.

As discussed at the beginning of this Section 2.1.(3.2) and illustrated throughout, the use of the helix bundle packing observed for BR (Henderson, 1990a) as a direct template for GPCR modeling is likely to be inappropriate. Nevertheless, the procedural steps and the inferences that can be drawn from the application of a known structural template to the specific modeling of GPCR merit scrutiny because structural information at the same level as currently available for BR is likely to emerge shortly for RH -a member of the GPCR family.

Modeling helix-helix interactions based on the BR template involves three sequential steps, as illustrated in many of the reported modeling studies that have utilized this approach (Cronet, 1993; Dahl, 1991; Findlay, 1990; Hibert, 1991; IJzerman, 1992; Lewell, 1992; Livingstone, 1992; Nordvall, 1993; Smolyar, 1993; Trumpp-Kallmeyer, 1992; Yamamoto, 1993): 1) alignment of TMH sequences for proteins in the BR and GPCR families 2) superposition of the modeled GPCR TMH backbone onto the structurally known BR backbone, and 3) refinement of the side-chain dihedral angles in order to relieve steric

clashes and to improve inter-residue interactions. Although the process seems straightforward, several caveats and considerations apply. First, a meaningful alignment of the GPCR sequences with BR is hindered by the lack of significant sequence homology among them. Solutions to this problem have involved a "structural" alignment whereby residues predicted to face the lipids (or the protein interior) in the GPCR are aligned with residues known to face the lipids (or protein interior) in the BR structure. However, because the degree of predicted lipid exposure for the GPCR sequences does not match the observed exposure in the BR structure (see discussion above and Figure 2.14.b), this "structural" alignment is not unique so that different alignments have been proposed in the literature. The published TMH alignments are offset from each other by 1 to 4 residues (compare (Hibert, 1991) vs. (Hibert, 1993) vs. (Nordvall, 1993)) depending on the particular helix. The structural implications of these differences in the alignments are very significant, and can be estimated from the parameters of an ideal α-helix: 1.5 Å pitch and 105° degrees per residue. When the same GPCR TMH is superimposed on its BR counterpart based on two different alignments, the resulting positioning of the same TMH on the respective GPCR models can deviate up to 6 Å axially, and their faces could be displaced by a rotation of up to 150°. As a result, GPCR models derived from different alignments would position residues critical for ligand binding quite differently. However,

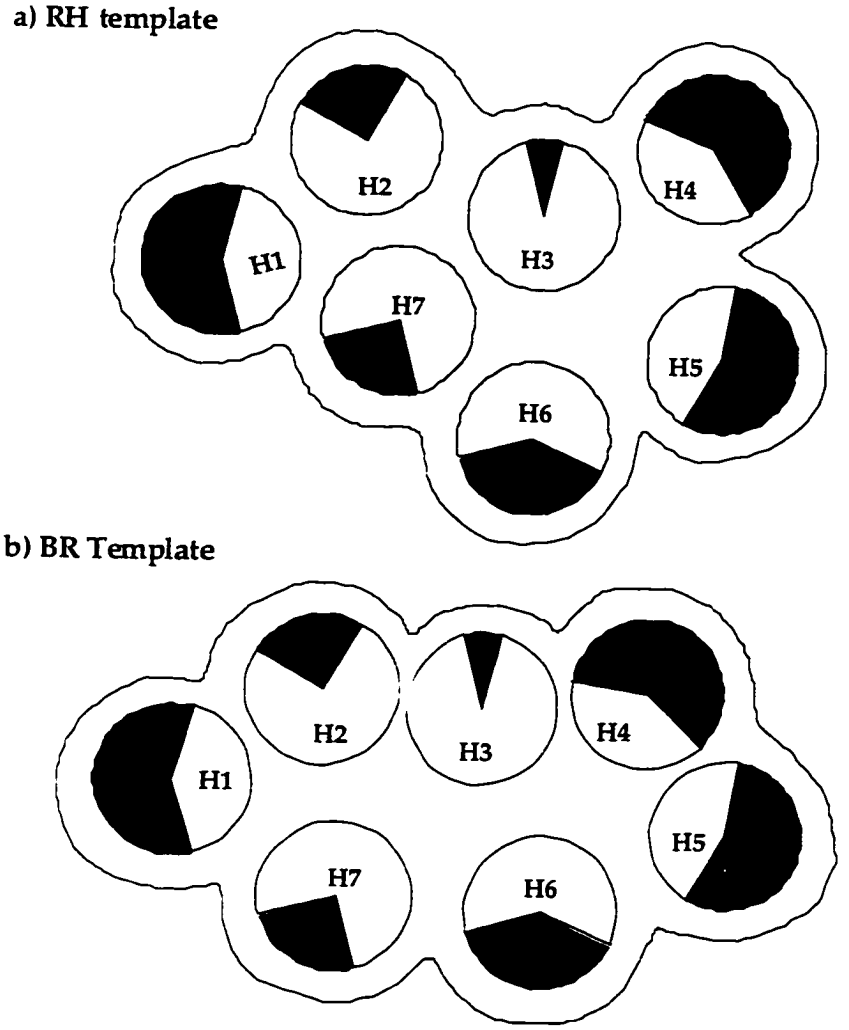


Figure 2.14:
Predicted degree of surface exposure for each GPCR TMH (J.M. Baldwin, Embo J 12, 1693 (1993)) superposed on a cytoplasmic view of the projection map of RH (a) and BR (b), where it can be compared with the extent of solvent accessibility for each TMH, marked with a continuous line surrounding the TMH bundle. (a) The identification of the RH seven TMHs follows that of Baldwin (J.M. Baldwin, Embo J 12, 1693 (1993)) for a sequential and clockwise helix packing arrangement. Note that the solvent accessible surface for each helix around the TMH bundle agrees fairly well with the predicted degree of lipid exposure (see text). (b) The predicted degree of surface exposure for each GPCR TMH is not consistent with the observed degree of lipid exposure for each TMH in BR, thus suggesting the inadequacy of using BR as a template for GPCR modeling (see text).

because these problems are due to an ambiguous alignment of the modeled and template sequences, they should not be expected when the RH template becomes available for use in the modeling of the highly homologous GPCRs.

It is important to note that the superposition of a GPCR TMH backbone onto the BR backbone is even more ambiguous when the position of Pro residues in these two proteins is not conserved, as occurs for all TMHs except TMH1. The kink introduced in the TMH by a Pro residue (see Section 2.1.2.2) renders the two helical portions of the TMH that flank the Pro discontinuous both in the directions of their helical axis and in the orientations of their "faces" (see Figure 2.10). Whether kinked TMHs are superimposed onto straight TMHs or viceversa, the superposition could be done for either helical portion, or for a combination of both. Nordval et al (Nordvall, 1993) found that superimposing the GPCR helices on only one of the two helical portions of kinked helices led to unacceptable steric clashes or large openings between TMHs, thus suggesting an averaged full backbone superposition. Notably, the non-conserved character of Pro-kinks applies also among members of the GPCR family. For example, the Pro in TMH2 occurs at different positions within the alignment and is even absent in some receptors. Since the cytoplasmic half of the TMHs in GPCRs is much more conserved than the extracellular half (Baldwin, 1993), and sequence conservation suggests structural preservation, these helices should probably be modeled superimposing their helical portions closer to the cytoplasmic side. The

consequence is thus that the extracellular portion of TMH2 and others (e.g. TMH5) is structurally more divergent, and significant alterations of the helix-helix packing interactions are likely to occur at these loci. This hypothesis is also consistent with functional considerations, the extracellular half being more likely to have to accommodate chemically divergent ligands while the intracellular half is responsible for the activation mechanism that could be shared by all members of the GPCR family.

2.1.3.3- Modeling extramembrane segments of the GPCRs.

Modeling the interhelical linkers, termed loops, has become increasingly important due to information obtained from mutagenesis experiments that have identified some of these protein segments as binding sites in peptide receptors (Fong, 1992), or G-protein coupling domains (Cheung, 1992). Modeling the N- or C-terminal domains of the GPCR is a very different task from that concerned with interhelical loops, and will not be discussed here. For loop modeling, a number of loop searching algorithms have been used, some of them implemented in standard modeling software packages (e.g. as used by Trump-Kallmeyer et al (Trumpp-Kallmeyer, 1993) to model the neurokinin 1 GPCR). These computational algorithms search structural data bases for protein fragments that have AA sequences homologous to portions of the loops. The yield of such searches is usually low, and the validity of structural predictions for such short, inherently

flexible fragments remains unclear. Maloney-Huss and Lybrand (MaloneyHuss, 1992) generated the structure of interhelical loops using a *de novo* procedure based on molecular dynamics (MD) simulations carried out at a low temperature and applying a series of constraints. Each loop segment was first assigned a completely extended structure(all $\phi,\psi=180°$), and then attached to the end of one target helix via a trans peptide linkage, before a short constant temperature (10-20° K) MD run was initiated. Weak harmonic constraints (5-10Kcal/mol) were used to pull the free end of the loop into position to form a trans peptide bond with the end of the second target helix. The final conformation defined for loops from such a procedure is probably one among many possible structures, due to the predicted flexibility and high degree of dynamical motion observed in such segments of soluble proteins. As a consequence, no reliable methods are currently available to suggest the specific modes of interaction between these loops and the TMH bundle. An exception is a disulfide bridge between the top of TMH3 and the TMH4-TMH5 linker shown in adrenergic (Raymond, 1990) and muscarinic (Savarese, 1992b) receptors, as well as in opsins (Findlay, 1990), and by analogy assumed to exist in most other GPCRs.

Of the six interhelical loop segments, the one that has attracted the most attention is intracellular loop III, whose N- and C- terminal ends have been implicated in the interaction of GPCR with the G-proteins (Savarese, 1992a). These portions of loop III, adjacent to the cytoplasmic ends of TMH5

and TMH6, were predicted to be in an α-helical conformation by Maloney-Huss and Lybrand (MaloneyHuss, 1992). This prediction was substantiated by extensive mutagenesis studies (Cheung, 1992), with the authors concluding that the N- and C-terminal ends of the loop were a continuous helix extending from TMH5 and TMH6 into the cytoplasm where they could interact with the G-protein. Notably, these cytoplasmic extensions of TMH5 and 6 agree with the predictions of TMH boundaries made by the Arg/Lys motif discussed in Section 2.1.2.1.2.

2.1.4- Probing of GPCR models based on molecular energy criteria.

Minimum energy-based criteria can be used throughout the modeling process at every structural level, but it is only when a 3D model of the GPCR is nearly complete that a full energetic refinement of the model can be undertaken. Computational methods ranging from quantum mechanical calculations (Clementi, 1991; Hirst, 1990; Weinstein, 1981) to energy minimizations by molecular mechanics (MM) and the search for structures with the lowest free energy with molecular dynamics (MD) simulations are available for this purpose (Brooks, 1988; Jahnig, 1992; McCammon, 1987; van Gunsteren, 1992; van Gunsteren, 1989; Weinstein, 1992a; Weinstein, 1992b).

The energy-based probing of the GPCR model requires an exploration of the conformational space around the 3D model searching for structures with increasingly lower energies. This procedure is continued until

convergence is reached, i.e. when the lowest energy structure has been identified within the limitations of the computational algorithms used. The two major computational algorithms that have been used for this task, molecular mechanics (MM) minimizations and molecular dynamics (MD) simulations, are applicable in order of increasing complexity. While MM minimization performs an optimization of the structure based on gradients of the potential energy surface, MD simulates the dynamic behavior of the protein where the incorporation of kinetic energy allows the algorithm to cross higher barriers of the potential energy surface exploring a wider portion of the conformational space available to the 3D model towards an "average structure" that corresponds at least to a local free energy minimum. As the computational requirements for MD simulations make this procedure much more costly than MM calculations, most published reports of models of GPCR have used only the former. However, for the purpose of energetic probing of the 3D model, MD simulations are much more stringent than MM: While MM will always reach a local minimum providing a minimized structure, in MD simulations the 3D model may fall apart if it lacks energetic consistency whereby the attractive forces do not prevail over dynamic fluctuations at room temperature. It is noteworthy, however, that Maloney-Huss and Lybrand found that excessive use of energy minimization lead to structures that are too densely packed (MaloneyHuss, 1992), thus hindering the model building process due to the difficulty to "loosen" the model structure for

further modeling explorations. The same effect of excessive compactness was observed for energy minimizations *in vacuo* of the entire model (Kontoyianni, 1993b; MaloneyHuss, 1992), although in this case the problem could be overcome by a treatment of the protein environment (Kontoyianni, 1993a), as described below.

Methodological improvements and a probing of the effects of the environment, both lipid and aqueous, are required for further validation and quantitative evaluation of the GPCR models. Progress in that area is hindered by the lack of structural information at atomic detail of the protein-lipid interface, and by the necessary simultaneous treatment of a lipid and an aqueous environments for a complete model of the receptor structure. An exploration of the effects of the environment on a GPCR model has been presented by Kontoyianni and Lybrand (Kontoyianni, 1993a). MD simulations of a complete model of the β_2-adrenergic receptor were carried with three different considerations of the environments: 1) simulation of the GPCR model *in vacuo*, 2) a simulation *in vacuo* with an explicit model of the membrane consisting of 60 DPPC lipid molecules surrounding the receptor, and 3) a continuum solvent model for the aqueous phase combined with the explicit treatment of the DPPC bilayer as in point 2. Although the last and more sophisticated representation of the environment produced the most physically reasonable behavior for the receptor models, the authors report that no major qualitative differences were observed among these three

simulations except in the loop regions and TMH bundle surface residues. Residues of the TMH portions that would be exposed to lipids were found to be packed against the TMH backbone when explicit lipids were not included. On the other hand, the loop regions collapsed onto the lipid head-group regions when the explicit bilayer model was included without a model for the aqueous phase, producing an unlikely physical situation. As a result, the authors have suggested the application of MD methods with an *in vacuo* explicit bilayer model where only the TMH bundle is included, in order to optimize TMH-TMH packing contacts. The resulting transmembrane models could then be completed by a subsequent addition of the loop regions. Although other authors have used different approaches, such as restricting the model to the transmembrane portion where a low dielectric environment can be assumed (Trumpp-Kallmeyer, 1992; Zhang, 1993b) or simulating the entire receptor model with an homogenous dielectric environment (Dahl, 1991), in the future MD simulation of the form described by Kontoyianni and Lybrand (Kontoyianni, 1993a) are likely to provide GPCR models that are more reliable in their description of the extramembranal and juxtamembranal regions.

An alternative treatment of the membrane environment was proposed by Jahnig and Edholm (Jahnig, 1990; Jahnig, 1992) in a MD simulation study of alternative TMH packing arrangements of BR. To simulate the membrane environment, the authors added to the potential energy function an

hydrophobic potential term that applies only to surface atoms. These MD simulations for BR bear special significance for the study of GPCR models because their aims and procedures parallel those applicable to GPCR models (Jahnig, 1990), but the availability of the BR structure makes possible a test of the adequacy of the structures resulting from the computations (Jahnig, 1992). Using the electron density projection map of BR (Henderson, 1990a) as a starting point for modeling helix-helix packings, the ability of energy-based criteria to distinguish the correct helix packing arrangement among alternative models was tested by MD simulations. Although the sequential and anti-clockwise packing arrangement of the BR structure was found to be lower in energy than any other structural alternative considered by the authors, the energy differences were not significant enough to substantiate the ability of MD simulation to guide the packing of the seven TMHs of BR (Jahnig, 1992). A particularly significant conclusion of these studies is that differences in the relative axial displacements among adjacent helices discussed in Section 2.1.3.2.2 was not overcome by the MD simulations (Jahnig, 1990), thus stressing the importance of the modeling approaches described in Section 2.1.3 of this Chapter to model TMH-TMH packing at atomic detail. However, a caveat of these studies is that the authors did not use the experimentally known TMH boundaries, but rather those predicted by hydrophobicity criteria (see table 4 of (Jahnig, 1992)), which can be in error by up to 7-8 AA as described in Section 2.1.2.1.2. Interestingly, these criteria for

the selection of TMH ends parallel those used for most published GPCR models (Cronet, 1993; Findlay, 1990; Hibert, 1991; Hoflack, 1993b; Huggins, 1993; IJzerman, 1992; Kontoyianni, 1993a; Lewell, 1992; Livingstone, 1992; MaloneyHuss, 1992; Nordvall, 1993; Trumpp-Kallmeyer, 1992; Trumpp-Kallmeyer, 1993; Yamamoto, 1993; Zhang, 1993b), thus providing an indication of the consequences of the expected inaccuracies in the prediction of TMH boundaries. As described in the discussion of the results obtained by Jahnig and Edholm in Section 2.1.3.2.2, errors in the predicted TMH boundaries seriously vitiated the BR models constructed, so that the correct ligand-receptor interactions were obtained only after shifting TMH6 two turns towards the extracellular side (Jahnig, 1990). Note that this correction is precisely what would be expected to be required according to the criterion that calls for including Arg/Lys patches at the cytoplasmic ends of the GPCR helices, as illustrated in Figure 2.7. In summary, it appears that at the present stage of modeling MD simulations are best suited to refine carefully built 3D models, but are not likely to provide a good approach to helix packing from a crude starting structure.

2.1.5- Computational approaches to probe structural details and functional mechanisms of GPCR

Pharmacological profiles, in the form of ligand binding affinities, or data on the efficacy of a ligand in eliciting a response (i.e. degree of agonism or

antagonism), are available for probing 3D models of GPCRs with respect to functional data. The approaches required to relate the models of ligand-receptor complexes to the binding properties are far less elaborate, and turn out to be much more ambiguous than the equivalent methodological approaches required to relate the receptor models to the experimental data on receptor response. Probing the constructed models of GPCRs against rank orders of binding affinity for various ligands is discussed in Section 2.1.5.1. The apparent lack of discriminatory power of these criteria in probing the results of modeling approaches makes necessary the use of the receptor models in the prediction of relative ranks of pharmacological efficacies, as described below in Section 2.1.5.2, in order to achieve more stringent tests of the structural and functional details.

2.1.5.1- Ligand binding properties as criteria for probing GPCR models

The most widely used approaches for probing GPCR models by criteria related to their pharmacological properties relate to their ability to rationalize ligand binding affinities. Models of ligand-receptor complexes have been proposed for each of the modeled receptors, and validation of the models were sought on the basis of a proposed agreement between the ligand binding scheme in the receptor model and the experimentally determined rank order of affinities and stereoselectivities of various ligands (Donnelly, 1994; Findlay, 1990; Hoflack, 1993a; IJzerman, 1992; Lewell, 1992; Livingstone, 1992;

MaloneyHuss, 1992; Nordvall, 1993; Smolyar, 1993; Trumpp-Kallmeyer, 1992; Yamamoto, 1993; Zhang, 1993a; Zhang, 1993b). To translate binding affinities into ligand-receptor interactions at the atomic level, structure activity relationships (SAR) of the ligands were used to infer on the nature of the chemical groups of the receptor that make direct contact with the ligands. For example, an H-bond acceptor or an aromatic ring may be proposed to interact with specific moieties of the ligands under consideration, based on the identified environment of the ligand binding pocket in the receptor model. However, such structural requirements in the ligand binding pocket can be fulfilled by different combinations of side-chains from residues in different TMHs of the receptor model, thus producing a high degree of ambiguity in the fulfillment of structural criteria. This lack of stringency in the criteria applied to receptor probing based on the mere fit of ligands in a predetermined binding pocket may be responsible for the apparent agreement of dissimilar GPCR models with the same affinity data and SAR considerations (cf. claims made for structurally very different models for the β_2-adrenergic receptor (Findlay, 1990; Lewell, 1992; MaloneyHuss, 1992; Trumpp-Kallmeyer, 1992); or for the different models for the 5-TH_2 receptor (Edvardsen, 1992; Trumpp-Kallmeyer, 1992; Zhang, 1993a; Zhang, 1993b)). Consequently, the application of more stringent criteria related to the functional response of the receptor to the binding of agonist appears necessary for the validation of GPCR models. These criteria require the application of a

more complex set of computational approaches in which the effect of agonist binding on the time-dependent structural rearrangements of the receptor model are compared to experimental data about the pharmacological efficacies of the ligands and the structural elements involved in receptor-effector coupling, as described below.

2.1.5.2- Criteria based on computational simulation of receptor activation mechanisms

The most challenging task in computational modeling in support of the efforts to understand structure-function relations of GPCR and their complexes with ligands remains the incorporation of the various available experimental results connecting structural information to functional data for the receptor. New and experimentally testable hypotheses must be derived from such approaches, based on the behavior of models for the receptors and their ligands, and on the computational exploration of their interactions. For this exploration, methods of static energy minimizations and molecular dynamics (MD) simulations have been presented in the literature (Brooks, 1988; McCammon, 1987; van Gunsteren, 1989). A detailed discussion of the computational methods providing functional inferences from the simulation of molecular mechanisms such as ligand-receptor interactions, is beyond the scope of this Chapter. However, the significant methodological and practical differences between molecular modeling of receptor structure, and the

computational simulation of receptor dynamics required to probe receptor function, merit some emphasis. As discussed recently (Weinstein, 1992a) and summarized in Figure 2.15, the computational simulation of the time-dependent behavior of the molecular model of a GPCR in the presence and absence of bound ligand can provide essential clues about the relation between structural details and the properties of the receptor. Consequently, such simulations provide a much more reliable basis for understanding the effects of mutations and chimeric constructs than mere inferences from static models. This is because subtle structural rearrangements could redistribute local interactions at mutated sites so as to compensate for modifications intended to be introduced by the mutations. It follows that modified proteins may maintain structural and functional integrity through rearrangements that can be predicted from computational simulations, but not from the analysis of the static models. Similarly, computational simulations of the signal transduction mechanism triggered in a model of the 5-HT_2 subtype of the serotonin GPCR by the binding of a ligand have revealed a characteristic series of structural rearrangements produced by the binding of agonists that differ from the structural effects of antagonist binding (Zhang, 1993a; Zhang, 1993b). Consequently, it is to be expected that mutations affecting the structural rearrangements, not the actual interactions between ligands and the receptor, would affect the measured affinities of agonists and antagonists. Such mutations would thus be difficult to discern from mutations affecting

the actual binding sites if only affinities are compared, making it difficult to interpret the results of mutagenesis probing of receptor structure (e.g., see (Colquhoun, 1993)). The combined use of ligand affinity and functional probing of the effects of structural modifications of GPCR can overcome such difficulties if the results are interpreted with the help of insight obtained from simulations of the consequences of these modification for both the structural and the dynamic properties of the GPCR.

Computational Simulation Approaches

Offer: **A formal framework for computational explorations of the structure, dynamics and functional properties of specific molecular systems.**

Features include:

- **rigorous definitions of the explored systems (e.g., boundary conditions; nature of the environment);**
- **systematic explorations of specific hypotheses;**
- **control experiments.**

Best applications: **To predict the behavior of the systems in *new experiments* designed to test specific hypotheses resulting from the analysis of their dynamic behavior.**

Figure 2.15:

Differences in the Definitions and most appropiate uses of computational approaches to molecular modeling and computational simulation.

The computational simulation studies that can provide dynamic information about the receptor proteins are complex and fraught with methodological pitfalls and shortcomings related to the special properties of the membrane-bound species (Jahnig, 1992; Kontoyianni, 1993b; Zhang, 1993b). Nevertheless, the validation of some of the inferences obtained from such studies by comparison to the known properties of the simulated system can help substantiate their use in probing the properties of the system. For example, the simulation of the complexes between a model of the transmembrane portion of the 5-HT_2 receptor with ligands chosen to represent full agonists, partial agonists and an antagonist (Zhang, 1993a; Zhang, 1993b) produced results that were in full agreement with the pharmacological properties of the ligands and with experimental data concerning the structural rearrangements expected from receptor activation (Zhang, 1993b). The detailed mechanistic insights provided by these simulations at the atomic level of structure, identified the steps in the propagation of the signal produced by ligand binding to the region of the receptor molecule that governs the interaction with G-proteins. Key residues responsible for these interactions were identified (Luo, 1993) in a manner not feasible with structural inferences alone. It is important to note that experimental probing (e.g, by mutagenesis) of these specific structural and functional inferences should provide valuable information not only about

the validity of the specific 3D model of the 5-HT_2 GPCR, but also about the detailed nature of a process of intramolecular signal transduction that may be common to many if not all GPCR families.

2.1.6- Concluding remarks

By focusing this description of a methodological approach to GPCR modeling and probing on the hierarchical use and intrinsic caveats of each available method, we have attempted to illustrate the extent to which each method is inconclusive taken individually, but can gain great predictive value if applied appropriately in conjunction with complementary approaches. Therefore, major inferences and conclusions guiding the construction and functional probing of 3D models of GPCR should be reached by comparing predictions from different methods with valid experimental data, while pursuing convergence among the results of the various prediction methods. This procedure requires an understanding of the basis and limitations of each available method, as sought from the juxtaposition and discussion of these approaches in the present overview. It is clear, however, that a deeper and more rewarding understanding of the methods, approaches and inferences presented here will be achieved when a correct three dimensional structure of a GPCR (e.g. a refined structure of RH) will be available at atomic detail for use in the exploration of convergent or contradictory structural inferences from the diverse analyses presented here.

This exploration should provide the required preparation for a complete structural and functional modeling of individual proteins from the great variety of GPCRs that are likely to exhibit tantalizing differences related to their specific functions and localizations.

ACKNOWLEDGMENTS

Critical reading of the manuscript and many helpful suggestions by Daniel Strahs and Juan Luis Pascual-Ahuir are gratefully acknowledged. The work was supported in part by NIH grants DA-06620, and Research Scientist Award, DA-00060 (to H.W.), DK-46943 (to Stuart C. Sealfon) and by a Fulbright/M.E.C.(Spain) fellowship (to J.A.B.). Computations for the illustrative examples were performed on the supercomputer systems at the Pittsburgh Supercomputer Center (sponsored by the National Science Foundation), Cornell National Supercomputer Facility (sponsored by the National Science Foundation and IBM), and the Frederick Biomedical Supercomputing Center of the NIH (NCI-Laboratory for Mathematical Biology).

2.2- New developments in the literature since the review chapter was written.

There has been an increasing number of theoretical and experimental approaches to study the structure and function of GPCRs in the years since the

publication of our review chapter (Ballesteros 1995). Fortunately, several databases have been constructed and made available to the scientific community through the World Wide Web (WWW) that organize the enormous wealth of biochemical, pharmacological, genomic, and theoretical data related to these receptors. There are five databases of this kind available in the WWW: the national Institutes of Health (NIH) database constructed by Drs. M. Van Rhee and K. Jacobson at the National Center for Biomedical Applications (NCBI), which uses our proposed numbering scheme for sequence alignments and their mutation database, and which includes an introduction to modeling GPCRs based on the methodological approaches described in our review chapter (Ballesteros 1995) and Section 2.1). The "GRAP" database constructed at the laboratory of Dr. Edvardsen in Norway and which contains quite detailed pharmacological data attached to their mutant database (binding and activation effects), as well as links to the SWISS-PROT database of protein sequences for each GPCR represented. The GCRDB information system set up by Dr. G. Vriend at the EMBL site in Heidelberg (Germany), contains also similar information as the previous databases but is different than the previous two databases in that it contains a database for theoretical models. The mutations and alignment have also links to the SWISS-PROT protein sequence database on the ExPASy Molecular Biology Server. There are also the GCRDb database constructed by Dr. L.F. Kolakowski Jr. at UTHSC, San Antonio, USA, and an Olfactory Receptor

Database (ORD) at Yale, USA.

The most significant novel theoretical approach to model GPCRs is the use of automated computational procedures to model GPCRs based on the large number of experimentally derived TMH-TMH interactions proposed recently in the literature (reviewed below). These approaches utilizes the algorithms developed originally to derive protein structures from NMR constraints (Herzyk 1995; Pogozheva 1997). The particular selection of distance constraints, TMH ends, the representations of the membrane environment, the extent of conformational freedom allowed in these procedures, and other parameters in these published studies is debatable and likely to affect significantly the results, as shown by the different models obtained by two different automated approaches (Herzyk 1995; Pogozheva 1997). The simulated annealing MonteCarlo procedure presented by Herzyk et al. (Herzyk 1995) utilized ideal TMHs thus devoid of the kink induced by Pro residues in these helices and the ensuing flexibility at this locus (see Figure 7.1 for an illustration of the degree of conformational uncertainty and flexibility induced by the PKs present in the 5HT2CR). The more recent distance geometry approach presented by Pogozheva et al. (Pogozheva 1997) represents a significant improvement in that it includes the flexibility induced by PKs and many of the interhelical interactions proposed recently in the literature, used in this thesis and reviewed below. However, none of these studies (Herzyk 1995; Pogozheva 1997) includes the TMH portions at the cytoplasmic

side predicted based on the Arg/Lys motif discovered in this thesis. Because in the procedure used by Pogozheva et al. (Pogozheva 1997) the structure of the TMH domain is established by interhelical H-bonds, they argue that the precise identification of TMH ends was not crucial.

However, the TMH segments at the cytoplasmic side that I have included in the model and are missing in Pogozheva's model (Pogozheva 1997) are very rich in polar (or charged), H-bonding groups proposed to interact with residues included in both models. Thus, inclusion of these TMH extensions present in the model derived in this thesis is likely to change significantly the resulting TMH model applying this automated procedure (e.g. see Arg-cage manuscript in Section 7.3.2). Nonetheless, the debatable applications of automated procedures to model the TMH domain of GPCRs as presented nowadays in the literature do not invalidate these methods. On the contrary, the increasing amount of experimental data pertinent to secondary as well as tertiary structure for these receptors, and the ensuing constraints for modeling purposes, would be expected to facilitate these automated approaches, which are likely to become our main modeling tool in the future.

There are several functional and conformational states of the receptor that need to be considered in order to interpret current experimental data in terms of structure-function relationships, and thus to use this data for modeling purposes. The minimal states of the receptor that need to be considered have been defined by the ternary model of receptor activation

(Lefkowitz 1988), later expanded by the addition of the activated receptor in the absence of G-proteins and renamed the two-state model (Samama 1993). This thermodynamic framework has been extended by the inclusion of multiple G-proteins that can be activated by a single receptor (Costa 1992) as well as other considerations (e.g. sodium modulation (Costa 1992)). Recent reports by Perez et al. (Perez 1996) have shown the inadequacy of this thermodynamic model to explain the presence of multiple and different activated states of the receptor (Perez 1996), as well as ligand induced coupling selectivity (Perez 1996). The changes necessary in the formulation of a thermodynamic model of receptor activation to explain this new data are beyond the scope of this thesis and would thus not be considered. In a collaborative effort with the laboratory of Dr. B. Kobilka (Gether 1996), we have furthered our understanding of the proposed thermodynamic model of receptor activation by considering the different states of the receptor in an energetic context (see Section 3.2). Based on these considerations, we have proposed that the activated state of the receptor, in the absence of the G-protein, is a high energy, metastable and thus transient state, resulting in a more unstable and flexible receptor conformation. The observed changes in fluorescence of a Cys-labeled purified β2-adrenergic receptor upon ligand binding could be rationalized in terms of the energetic levels proposed (Gether 1996), as described in Section 3.3.

The most relevant experimental studies in recent years have been the

advance in the structural characterization of GPCRs by experimental techniques. The new developments can be classified into direct structural determinations (Section 2.2.1) and indirect structural inferences (Section 2.2.2), described below.

2.2.1.- Direct structural determination of GPCRs.

The initial electron density map of bovine rhodopsin obtained by electron cryomicroscopy (Schertler 1993) has been complemented by similar studies on frog (Schertler 1995) and squid (Davies 1996) opsins. The same architecture and salient structural features are observed for all three opsins, the main difference among these species variants being the oligomeric contacts observed in frozen two-dimensional crystals. Because these divergent oligomeric contacts pertain to sequence motifs divergent among the opsins and with respect to other GPCRs (Davies 1996), they are not pertinent to the studies presented in this thesis.

NMR techniques have been applied to isolated segments of a variety of GPCRs. The peptide corresponding to the seventh transmembrane segment of the NK1R, Ac-AMSSTMYNPIIYSSL-NH_2, has been studied by NMR techniques in various solvents of mixed polarity, including Me_2SO perfluoro-tert-butanol); little structure was found for the peptide, although the authors proposed a gamma turn at the NP motif thus creating an Asx-turn (Berlose 1994). This NP conformation induces a sharp bend in the helix of nearly

+-90°. These finding agree with the conformational preference of the NP motif in water-soluble protein of known structure, as substantiated by Karel Konvicka in this laboratory (Konvicka 1995) and others (Berlose 1994), where the NP motif is presented as a strong N-capping motif for α-helices. The agreement of the conformation proposed for the NP motif in quite polar solvents (Me_2SO) with the conformations observed for water-soluble proteins leaves unanswered the question of how this conformation would be maintained in a membrane environment, because it generates several H-bond free backbone groups which would be energetically unfavorable in a hydrophobic environment, but not in the native aqueous medium where these conformations are normally found. Furthermore, the authors (Berlose 1994) observed few constraints, insufficient to delineate uniquely the structure, and observed multiple conformations co-existing thus complicating the analysis because it was not possible to know whether observed NOEs corresponded to the same conformation or to different conformers. They note "a strong CaHAsn8-CdH2Pro9 NOE crosspeak indicates that Asn8 is preferentially in an extended orientation with 120 < psi < 180. The presence of a weak CaHPro9-NHIle11 NOE, daN(i,i+2), suggest the participation of a 3-10 helix or beta-turn structure" (Berlose 1994). I would constrain psi of N7.49 to these boundaries, and will favor their 3-10 helix versus the turn hypothesis (Section 7.4).

Other NMR studies have focused on isolated fragments of the receptor

that belong to extramembranal domains, thus avoiding the difficulties that the membrane environment imposes on NMR techniques. The most significant and conclusive study was performed on the so called "intracellular loop 4" of the avian β-adrenergic receptor (Jung 1996). The peptide comprising residues 7.56(345)-7.70(359) was observed to adopt a regular α-helical conformation in the presence of phospholipid micelles or small unilamellar phospholipid vesicles. This peptide was also shown to strongly inhibit receptor-mediated adenylate cyclase activity (Jung 1996), thus likely to be involved in the G-protein coupling mechanism. Other NMR studies on the intracellular loops (IC) I, II, III, and C-terminal fragment of rhodopsin (Yeagle 1995a; Yeagle 1995b; Yeagle 1997) and the ICIII loop of the β-adrenergic receptor (Jung 1995) resulted in either questionable structural determinations, or were pertinent to receptor domains evolutionary highly divergent and thus with no direct implications for the structure of the 5HT2CR.

Further direct structural determination pertains to the conformation or interactions of specific residues in rhodopsin, mostly by spectroscopic techniques, whose interpretation involved comparison of the conformational changes upon activation, often aided by the analysis of receptor mutant constructs. Because the data *per se* are pertinent to the native receptor structure, they are included in this section. The conserved Asp in H2 (D2.50) has been shown to be much more strongly H-bonded upon rhodopsin photoexcitation (Fahmy 1993; Rath 1993; Lin 1996). Another conserved acidic

group, belonging to the DRY motif at the cytoplasmic boundary of H3 (D3.49), has been shown to undergo protonation upon rhodopsin activation (Fahmy 1993; Arnis 1994). A second protonation site was detected (Arnis 1994), though the specific acidic residue involved was not identified. The side chain conformation of the conserved W6.48 was observed to change upon rhodopsin activation (Lin 1996): The transition dipoles of the indole were reoriented upon activation towards an orientation parallel with the membrane plane. As described in Section 7.3.3, this rearrangement can be interpreted in terms of the side chain dihedral angle $\chi 1$ adopting a *gauche+* rotamer in the inactive and a *trans* rotamer configuration in the active state of the receptor. The indole NH of W6.48 was observed to participate in H-bonding interactions in the inactive, but not in active, state of rhodopsin (Lin 1996). The opposite case was observed for W4.50 using similar spectroscopic approaches; W4.50 was not H-bonded in either state and its side chain conformation did not change significantly upon activation (Lin 1996). Other spectroscopic studies of rhodopsin were pertinent to the retinal binding site and other domains which are structurally divergent from the neurotransmitter GPCRs and thus not considered in this study. An exception would be the proposed interaction between the protonated Schiff base of the retinal linked to K7.43(296), and the E3.28(113); this interaction, which would impose significant constraints for modeling the packing of H3 and H7 in the TMH domain, has been questioned in the recent literature based on NMR

experiments (Han 1993) showing a decreased electron density for the C12, not the Schiff base, related to the counter ion of the protonated retinal, which is likely to involve also structural water molecules. Thus, this constraint was not used for modeling the 5HT2CR.

2.2.2.- Indirect structural determination of GPCRs.

These experimental approaches often involve modification of the amino acids present in the native receptor by either site directed mutagenesis or chemical derivatization. The essential assumption to interpret these sets of data is that the overall structure of the receptor is maintained, which is not always the case as will be discussed in detail for some of this data in the context of a three-dimensional model of the receptor TMH domain (Section 7). These experimentally derived structural inferences can be classified into three classes: experiments pertinent to the TMH boundaries and orientation (Section 2.2.2.1), experiments pertinent to TMH-TMH interactions (Section 2.2.2.2), and proposed ligand-receptor interaction sites (Section 2.2.2.3).

2.2.2.1- Experimental data pertinent to TMH ends and orientations.

Recently, the ability of Cys residues to undergo chemical modification by specific probes has been used to study the structural and dynamic properties of several GPCRs. A continuous stretch of residues is first mutated to Cys one at the time. The accessibility of the substituted Cys is then studied

by the availability of these Cys residues to undergo chemical derivatization by different probes. Chemical derivatization by spin labeling probes can be confirmed spectroscopically by means of ESR techniques, and has been used to characterize the cytoplasmic side of H3-H4-H5-H6 in rhodopsin (Farahbakhsh 1995; Altenbach 1996; Yang 1996b). Spin labelling experiments permit the precise characterization of the environment of the labeled sites in terms of solvent polarity (lipid, interior protein, or water exposed), flexibility of the labeled probe, and the changes in these properties upon rhodopsin photoexcitation. Chemical derivatization by highly hydrophilic and Cys-specific reagents added to intact receptor-expressing cells can be determined by the perturbations induced in ligand binding to the modified receptors, and has been used to characterize residues accessible to the ligand binding crevice for the extracellular side of H3-H5-H7 in the D2 receptor (Javitch 1995a; Javitch 1995b; Fu 1996). The latter technique is called the Substituted Cys Accessibility Method (SCAM).

Gether et al. have used Cys-specific fluorescent probes to label native Cys residues in the purified β2-adrenergic receptor (Gether 1995). The observed increase or decrease in fluorescence upon ligand binding showed a correlation with the intrinsic efficacies of these ligands, thus indicating that the fluorophores were reporting conformational changes in the receptor structure related to the activation mechanism. The Cys responsible for that effect were identified as C3.44(125) and C6.47(285) (Gether 1996). The

fluorescence of the IANBD probes attached to these two Cys indicated a predominantly hydrophobic environment.

2.2.2.2- Experimentally derived TMH-TMH interactions.

Specific TMH-TMH interactions in GPCRs have been proposed based on spin labeling (Farrens 1996; Yang 1996a), fluorescence labeling, Cys crosslinking (Yu 1995; Farrens 1996; Kono 1996), engineered Zn binding sites (Elling 1995; Elling 1996; Sheikh 1996), and double revertant mutant constructs (Rao 1994; Zhou 1994; Liu 1995; Sealfon 1995; Han 1996). The multiple results from these studies are mentioned below in that order, listed in Table 2.1 arranged by the proposed helix-helix contacts, and are also illustrated schematically in Figure 2.16 on a rhodopsin-like arrangement of the seven TMH represented by helical wheels assuming ideal helical conformations. Note the general consistency among these results, as will be explored in detail in the context of a three-dimensional model of the TMH domain of the 5HT2CR (Section 7.1).

Table 2.3

Experimentally derived TMH-TMH interactions ordered by the helix pair involved and shown schematically on a projection map in Figure 2.16. Interactions are indicated by the residue identifier, the AA of the corresponding receptor including the subsequent mutant, receptor involved, conformational state (active or inactive) and experimental technique, the latter indicated by the following codes: Cys-Cys (Cys crosslinking), Zn-His (engineered Zn binding site), Spin (spin labeling of Cys residues), 2rm (double revertant mutants), Lig (Ligand binding residues, included as links to D3.32).

TMH-TMH	AA <--> AA	Technique	GPCR	Conf. state (R,R*)
H1-H7				
	T1.39A <--> H7.36T	2rm	MUSC	R
	H1.60C <--> C7.63	Spin	RHO	d(R)<10 A d(R*)>10 A
H2-H7				
	D2.50 <--> N7.49	2rm	5HT2a,GNRHR	
	K7.43 <--> G2.57G	2rm	RHO	R*
H2-H3				
	Y2.64H <--> H3.28	Zn-His	NK1	R
H3-H5				
	N3.29H <--> E5.35H	Zn-His	NK1	R
	D3.32 <--> S5.42	Lig	DOP/ADR	
	D3.32 <--> ST5.43	Lig	5HT/DOP/ADR	
	D3.32 <--> S5.46	Lig	5HT2a/DOP/ADR	
H3-H6				
	V3.54C <--> E6.30C	Cys-Cys	RHO	R, not R*
	V3.54C <--> E6.31C	Cys-Cys	RHO	R, not R*
	V3.54C <--> E6.32C	Cys-Cys	RHO	R, not R*
	V3.54C <--> E6.33C	Cys-Cys	RHO	R, not R*
	V3.54C <--> E6.34C	Cys-Cys	RHO	R, not R*
	V3.53H <--> T6.34H	Zn-His	RHO	R, not R*
	V3.53H/K3.56H<->T6.34H	Zn-His	RHO	R, not R*
	G3.36L <--> F6.44A	2rm	RHO	R
	D3.32 <--> N6.55	Lig	ADRb2	
	D3.32 <--> F6.52	Lig	5HT2a/ADR	
H3-H7				
	K7.43 <--> E3.28	Lig	RHO	R
	D3.32 <--> W7.40	Lig	ADRa2	R
	D3.32 <--> N7.39	Lig	5HT1a/ADRb2	R
H5-H6				
	Y6.59H <--> E5.35H	Zn-His	NK1	R
	Y6.59H <--> H5.39	Zn-His	NK1	R
	N5.35C <--> F6.59C	Cys-Cys	RHO	
	V5.39C <--> F6.55C	Cys-Cys	RHO	
	V5.39C <--> A6.56C	Cys-Cys	RHO	
	V5.39C <--> F6.59C	Cys-Cys	RHO	
	V5.39C <--> T6.60C	Cys-Cys	RHO	
	I5.40C <--> F6.55C	Cys-Cys	RHO	
	I5.40C <--> A6.56C	Cys-Cys	RHO	
	I5.40C <--> F6.59C	Cys-Cys	RHO	
	F5.43C <--> F6.59C	Cys-Cys	RHO	
	Y5.58C <--> V6.37C	Cys-Cys	RHO	
	G5.59C <--> V6.37C	Cys-Cys	RHO	

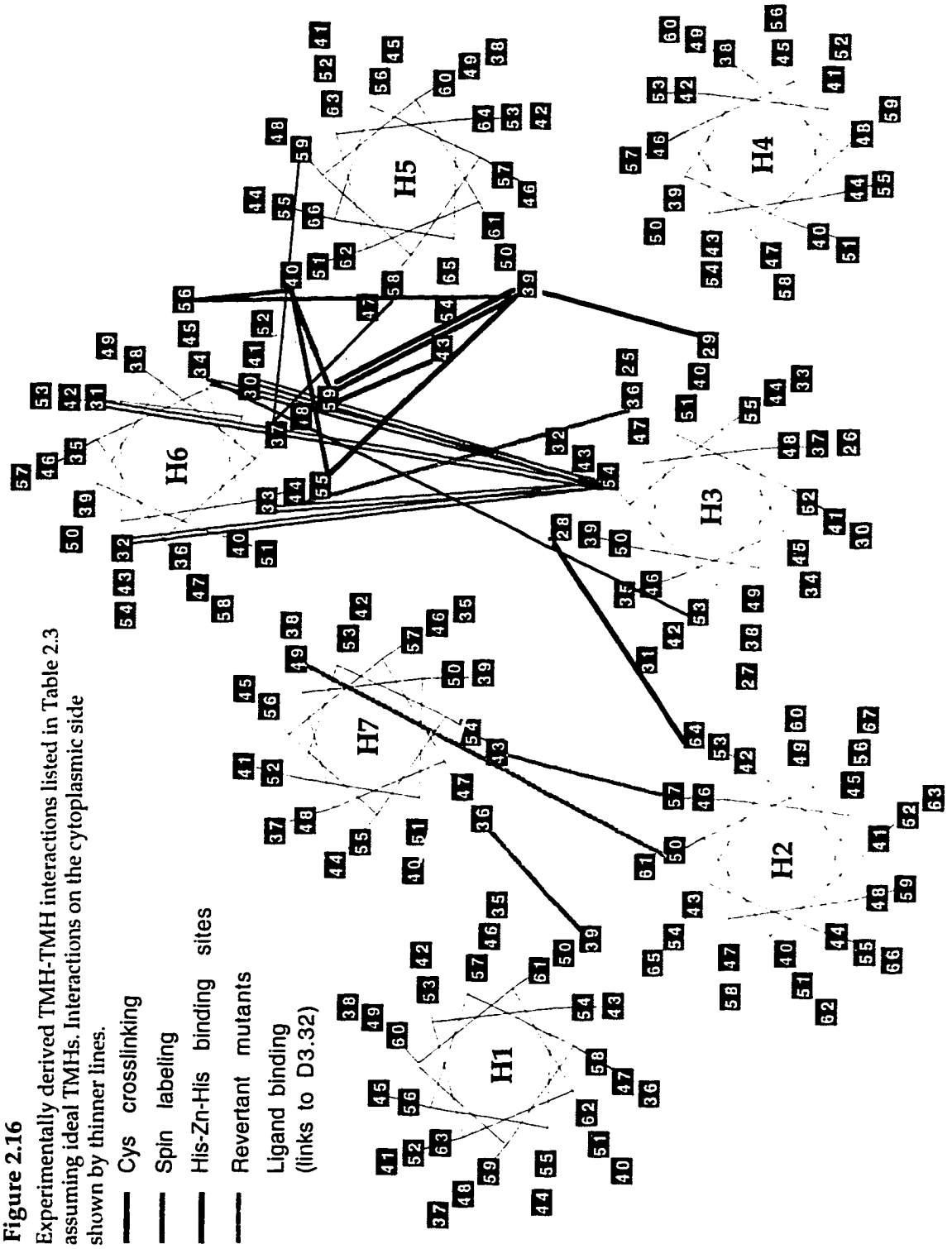


Figure 2.16
Experimentally derived TMH-TMH interactions listed in Table 2.3 assuming ideal TMHs. Interactions on the cytoplasmic side shown by thinner lines.

Spin labeling experiments in rhodopsin: The cytoplasmic portions of H3 (3.54) and H6 (6.30 to 6.35) were shown to be in spatial proximity and to move away upon rhodopsin activation (Farrens 1996). The pattern of distances and the subsequent changes upon activation observed between residues 3.54 and 6.30 to 6.35 led to the proposal that rhodopsin activation involves a movement of H6 away from H3 and a rotation of H6 in an anticlockwise direction viewed from the extracellular side (Farrens 1996). Similar studies by the laboratories of Hubbell and Khorana on rhodopsin have also identified significant movement between residues H1.60(65)C and C7.63(316) which were found to be closer than 10 Å in the basal state but significantly further away upon activation (Yang 1996a).

Fluorescence labeling experiments in the NK2 receptor: A fluorescent unnatural amino acid was incorporated at known sites of the NK2 receptor (Turcatti 1996). This allowed the determination of intermolecular distances by measuring fluorescence transfer between the fluorescent unnatural amino acid and a fluorescently labeled NK2 heptapeptide antagonist (Turcatti 1996). Although this approach seems promising, at the moment the uncertainty in the measured distances is too large to be useful in modeling studies. The reason is the conformational degrees of freedom of the ligand, the side chains in the receptor, and the lack of knowledge about the conformation of the ligand bound to the receptor.

Cys crosslinking experiments in rhodopsin: The spin labeling

experiments on the cytoplasmic boundaries of H3 and H6 described above, were complemented by the analysis of Cys-Cys crosslinking patterns between the C3.54 in H3 and a Cys at positions 6.30 to 6.35 in H6 (Farrens 1996). All crosslinkages found resulted in rhodopsin inactivation, thus consistent with their findings from spin labeling experiments that H6 and H3 move away from each other upon activation. It was found that C3.54 could crosslink to all substituted Cys from position 6.30 to 6.34, involving a whole turn of H6. This is inconsistent with a helical conformation at this locus. However, the same authors proposed an α-helical conformation at this locus based on their prior spin labeling experiments (Farrens 1996). It follows that the energy provided by the covalent bond formed in a Cys disulfide bridge could override the energy required to disrupt significantly structural motifs of the receptor. In order to control for structural perturbations induced by these chemical and mutagenesis modifications of rhodopsin, the authors showed that the absorbance of these constructs was always similar to wild type. However, rhodopsin has been shown to maintain near wild type absorbance under conditions of significant structural perturbations (see Section 7.4.1 and references (Sikora 1994; Strassburger 1997)). Examples include a 25 % loss in α-helical structure upon solubilization and reconstitution (Sikora 1994), as well as thermal stability in the absorbance spectra from 4° up to 90° or more (Shnyrov 1988), depending on the specific conditions. Thus, in interpreting biophysical experiments on rhodopsin such as the spin labeling considered

above, as well as other spectroscopic measurements that require solubilization and reconstitution of rhodopsin for sample preparation, we should expect some degree of structural perturbation in regions far from the retinal binding site, such as the cytoplasmic boundaries considered here.

Another set of Cys crosslinking studies on rhodopsin was performed by the laboratory of Dr. D. Oprian involving multiple sites between H5 and H6 (Yu 1995; Kono 1996), as listed below:

```
N5.35(200)C   <---> F6.59(276)C
V5.39(204)C   <---> F6.55(273)C
V5.39(204)C   <---> A6.56(274)C
V5.39(204)C   <---> F6.59(276)C
V5.39(204)C   <---> T6.60(277)C
I5.40(205)C   <---> F6.55(273)C
I5.40(205)C   <---> A6.56(274)C
I5.40(205)C   <---> F6.59(276)C
F5.43(208)C   <---> F6.59(276)C
Y5.58(223)C   <---> V6.37(254)C
G5.59(224)C   <---> V6.37(254)C
```

The consistency of these multiple crosslinking sites with an α-helical conformation for the H5 and H6 segments studied can be apreciated in Figure 2.16.

Engineered Zn binding sites: Similar studies as described above designed to estimate distances between the cytoplasmic portions of H3 and H6 were done by engineering Zn binding sites (Sheikh 1996), i.e. by incorporation of His residues at known positions through mutagenesis experiments followed by measured the affinities of Zn for the wild type and mutant constructs. From the pattern of Zn sites formed, the authors concluded that V3.53(138)H was close to T6.34(251)H and K3.56(141) (Sheikh 1996). As a

control, the affinity of Zn for rhodopsin was not increased by other His substitutions between these H3-H6 loci, as shown by the constructs {V3.53(138)H, K6.31(248)H, R6.35(252)H} and {E3.49(134)H, K6.31(248)H, R6.35(252)H}. The effect of Zn binding to the {V3.53(138)H, T6.34(251)H} mutant was to prevent rhodopsin activation, thus consistent with the inference from similar spin labeling experiments (Farrens 1996) described above that H3 and H6 move away from each other upon activation.

The method of engineering Zn binding sites in GPCRs by His substitutions was pioneered by Dr. Thue Schwartz on the NK1 receptor. Several publications from this lab (Elling 1995; Elling 1996; Sheikh 1996) using this approach support a proximity between His substituted at positions 2.64-3.28, 3.29-5.35, and 6.59-5.35 (6.59-5.39 to a lesser extent). The connectivity among the 7 TMHs defined by this set of engineered Zn binding sites supported a sequential and anticlockwise packing arrangement (Elling 1996), as can be observed in Figure 2.16. In functional terms, all these engineered Zn sites behaved as an antagonist of the NK1 receptor.

Double revertant mutant constructs: In addition to the revertants described in the published review (Ballesteros 1995) (Section 2.1), which included the 7.39-H1 in adrenergic receptors (Mizobe 1996), K7.43-D2.57-E3.28 in rhodopsin (Oprian 1992; Rao 1994), and N7.49-D2.50 in 5HT2A and GnRH receptors (Zhou 1994; Sealfon 1995), the following studies have been recently published. Residue F7.38 have been proposed to face H3 and/or H6 based on

revertant chimeras in adrenergic receptors (Mizobe 1996). The structural perturbation induced by the H7.36(423)T mutation in muscarinic receptors was reverted by the subsequent mutation of T1.39(37)A (Liu 1995), indicating that H7.36(423) and T1.39(37) are in spatial proximity. Another structural revertant was observed in rhodopsin with the G3.36(121)L and F6.44(261)A mutations (Han 1996). The proposed proximity among these residues derived from the observed revertant phenotypes is illustrated schematically in Figure 2.16.

2.2.2.3- Experimentally proposed ligand-receptor interaction sites.

The direct ligand-receptor interactions within neurotransmitter GPCRs assumed for modeling purposes were reviewed in the published chapter ((Ballesteros 1995), Section 2.1) and included D3.32, G5.42, S5.43, A5.46, and V7.39 based on mutagenesis data and W7.40 (Wong 1988) and S2.63 and/or A2.64 (Dohlman 1988) based on covalent attachment sites for pindolol analogs in adrenergic receptors (amino acids correspond to the 5HT2CR). More recent publications have substantiated a role for S3.36 (Almaula 1996b) and F6.52 (Choudhary 1995) in the 5HT2AR and N6.55 in the β2-adrenergic receptor in ligand recognition (Wieland 1996). The region of the TMH domain defined by these residues can be appreciated in Figure 2.16 on a representation of the 7 TMHs following the rhodopsin map as described in Section 2.1. Note that all these residues cluster in the extracellular boundaries

of H3-H5-H6-H7 thus defining an initial "consensus" binding site for neurotransmitter GPCRs.

2.3.- New methods applied to modeling that were not described in the Review Chapter.

Described below are some methodological approaches that were not described in the published chapter (Section 2.1 above). They involve the derivation of specific mathematical formulations to analyze quantitatively multiple sequence alignments according to criteria and the ensuing structural inferences already described in the chapter. The new procedures applied for primary structure analysis are presented in Section 2.3.1, and those applied for tertiary structure analysis are presented in Section 2.3.2. Energy-based methods used for energetic evaluation of the model (Molecular Dynamics) or conformational explorations of its structural components (MonteCarlo simulations) are described in Section 2.3.3.

2.3.1.- Primary structure analysis: methods applied for the quantification of the conservation profile.

The methods described below represent a quantitative approach to the same criteria discussed in our published review (Ballesteros 1995) (Section 2.1), intended to identify amino acid (AA) positions which are evolutionarily constrained, and to explain these constraints in terms of a conserved

physico-chemical property. In attempting to quantify the degree of conservation for a given property, I distinguish 3 categories according to their formulations. First (Section 2.3.1.1), those that have been quantified for all natural AA, e.g. hydrophobicity (hdp) and volume. Second (Section 2.3.1.2), properties whose quantification is only possible for a subset of AA due to their absence in the others, e.g. charge, aromaticity, etc... . Third (Section 2.3.1.3), the quantification of the conservation index at a given position, requiring the use of pairwise mutation matrices. I describe their quantification separately below.

2.3.1.1.- Quantification of the conservation of volume or hydrophobicity (hdp). This analysis is performed based on the hdp scale of Kyte-Dootlitle (Kyte 1982) and the volumes for each standard AA as derived from known protein and peptide structures (Chothia 1984). A statistical analysis is applied for both the hdp and the volume, computing the arithmetic mean and the standard deviation at each AA position. Since we do not know how representative is a particular alignment, the calculations are based on the different residues present at a given AA position rather than the number of occurrences of each of them (Donnelly 1989); e.g. we consider that both Ala and Tyr are present at 2.64, irrespective of how many of them actually occur at 2.64. The conservation of hdp or volume is quantified through the standard deviation. The calculated mean value is taken as the putative molecular

determinant for those positions with low standard deviation.

What threshold value of the standard deviation represents a conserved position in terms of volume or hdp?. Instead of statistical significance the analysis has relied upon a physico-chemical criterion. The analysis of 10 globin structures showed that for core residues with equivalent environments the mutations were restricted to a volume change of one methyl group (Bordo 1990). Therefore, the volume is considered conserved if the standard deviation is smaller than the volume of a methyl group (22 A^3), i.e. $\sigma(Vol)<22\ A^3$. The analysis of the conservation of hdp character was done by considering subsets of AA instead of a hdp scale (Bordo 1990; Overington 1990), thus precluding a similar threshold as in the analysis of volume conservation described before. Since $\sigma(Vol)<22\ A^3$ represents a 13% deviation over the volume range (66-238), I will considered the hdp conserved for an analogous 13% deviation over the range (-4.5,4.5), that means $\sigma(hdp)<1.2$.

2.3.1.2.- Quantification of the conservation of properties not present at every AA -e.g. aromaticity. In this case the analysis is based on the % presence of a subset of AA types that display such a property, i.e. aromatic character=W%+F%+Y%. The threshold value is arbitrarily chosen as 70%. For the purpose of a systematic analysis I have classified the H-bonding capabilities according to their position in the side-chain, as presented in Table 2.4. Another property I found useful, but for which there is no absolute

statistical evidence that is specifically conserved, is the β-branched character. The conformational role of these conservations within TMHs has been inferred based on the analysis presented here of known structures (Section 7.3.3.2), supported by mutagenesis studies developed through our collaborative efforts (see Arg-cage paper in Section 7.3.2) and by others (Deber 1992), and by the fact that {Val, Ile, Thr} represent 40% of all AA in TMH domains (Deber 1986).

I shall clarify why I search for properties whose specific conservation has not been proven or associated as a structural determinant. (By unproven I mean that it has not been analyzed before, or that it has been analyzed only partially). Given their proven significance in TMH domains, I will explore whether the conservation of those properties at a given site is a reasonable indication of the molecular determinant associated at this AA position. For this exploration I will assume the legitimacy of the proposition, and will as I did for other contrast the results with structural inferences derived from other criteria. Whenever clashes or incompatibilities occur among different criteria, I will acknowledge that some of the criteria are unproven in selecting among alternative TMH packings. Given the explanation above, the specific properties searched are listed below with the AA subset selected.

Description	Code
1) Positively charged groups: {R, K}	+
2) Negatively charged groups: {E, D}	-
3) Aromatic groups: {F, W, Y, H}	arom
4) β-Branched hydrophobic: {I, V}	bbra
5) HBD/HBA at position γ along the side-chain: {S, T, C}	hdag
These 3 are considered as good HBA as HBD because in an α-helix environment they often H-bond back to the C=Oi-4, thus rendering the O or S lone pair available for H-bonding.	
6) HBD at position δ along the side-chain: {N, H}	hbdd
7) HBD at position ε along the side-chain: {Q, H, W, R}	hbde
8) HBD at position ζ along the side-chain: {K}	hbdz
9) HBD at position μ along the side-chain: {Y, R}	hbdm
10) HBA at position α along the side-chain: {G, P}	hbaa
Here I refer to the C=Oi-4 of these residues as HBA, since they render a O lone pair accessible to solvent- see Table 2.4.	
11) HBA at position δ along the side-chain: {N, D, H}	hbad
12) HBA at position ε along the side-chain: {Q, E, H}	hbae

Table 2.4

Classification of the H-bonding capabilities of the 20 standard amino acids according to their position in the side chain.

1) H-bonding donors (HBD)

X_1 / aa \at	α (X_1)	β (X_2)	γ (X_3)	δ (X_4)	ε	ζ	η	HBD Position
Ser	C	C	OH					γ
Cys	C	C	SH					γ
Thr	C	C(OH)(C)						γ
Asn	C	C	C(O)(NH_2)					δ
His	C	C	C(C=N-NH-C)					δ
+His	C	C	C(C=NH-NH-C)					ε
Gln	C	C	C	C(O)(NH_2)				ε
Trp	C	C	C(C=C(C=C)-C=C-C)(C-NH)					ε
Arg	C	C	C	C	NH-C(N)(N)			ε
Lys	C	C	C	C	C	NH_3		ζ
Arg	C	C	C	C	N	C(NH_2)(NH_2)		η
Tyr	C	C	C(C=C-C(OH)-C=C)					η

2) H-bonding acceptors (HBA)

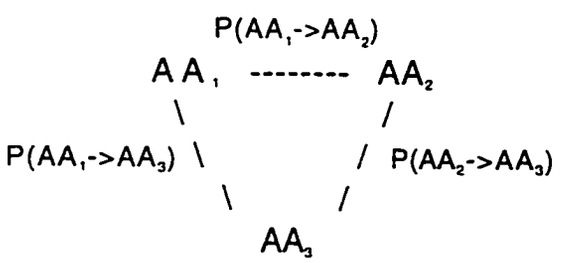
X_1 / aa \at	α	β	γ	δ	ε	ζ	η	HBD Position
Gly	CO							α
Pro	CO							α
Ser	C	C	OH					γ
Cys	C	C	SH					γ
Thr	C	C	OH / C					γ
Asp	C	C	C(=O)O					δ
Asn	C	C	C(=O)NH					δ
His	C	C	C(NH—C=N)					ε
Glu	C	C	C	C(=O)O				ε
Gln	C	C	C	C(=O)NH				ε

2.3.1.3.- Quantification of the conservation at a given position, the definition of a "conservation index". This property has been quantified as the number of different AA at that position (Rees 1989b). I consider this choice unphysical since it assumes that all 20 AA as equally distinct, thus attaching no quantitative significance to the physico-chemical implications of the AA substitutions. I propose a different procedure for the calculation of the conservation index that would take into account the relations among the 20 natural AA. Note that the term "Conservation Index (CI)" as adopted here represents a different quantification of the extent of amino acid variability at any given position within the multiple sequence alignment. Thus, the CI is actually a variability index so that a maximal value for the CI corresponds to a minimally conserved position, and vice versa.

There is a well known measurement of the quantitative significance of AA substitutions: the pairwise substitution probabilities or mutation penalty matrices (Overington 1992). These matrices calculate the probability of mutating a given residue AA_1 -e.g. Ala- into another residue AA_2 -e.g. Glu. I refer to this probability as $\Pi(AA_1->AA_2)$, and to the penalty associated with it as $P(AA_1->AA_2)$; the penalty of a mutation Π is the complementary probability of the probability of a mutation P, i.e. $P(AA_1->AA_2)=1-\Pi(AA_1->AA_2)$. Those probability values are calculated from a multiple sequence alignment of homologous proteins belonging to the same protein family. The

statistics are improved by combining results from a number of such alignments, corresponding to a set of protein families. A mutation matrix specifically calculated for AA located within α-helices (Overington 1992) will be used for these calculations. The use of mutation matrices for the calculation of the conservation index requires the estimation of the penalty for the presence of a set of different AA at a given position $P\{AA_1, AA_2, ..., AA_k\}$, from a set of pairwise substitution penalties $\{P(AA_1->AA_2), P(AA_1->AA_3), ...\}$. There is a well defined mathematical formula for such calculation called the theory of polytopes (Blumenthal 1953). I present the mathematical formulation below, followed by a comparison between the results I obtained with it, and the previous method (which I believe justifies the use of the new formulation).

Let us imagine the set of different AA $\{AA_1, AA_2, ..., AA_k\}$ as points in an euclidean space, and their mutation penalties P(AAi->AAj) as distances among these points, i.e:



If the triangular inequality $P(AA_1->AA_3) > P(AA_1->AA_2) + P(AA_2->AA_3)$ is fulfilled

for every set of $\{AA_1, AA_2, AA_3\}$, condition fulfilled by the mutation matrices, then:

N points $\{AA_1, AA_2, ..., AA_N\}$

↓

$(N*(N-1))/2$ distances $P(AA_i \rightarrow AA_j)$

↓

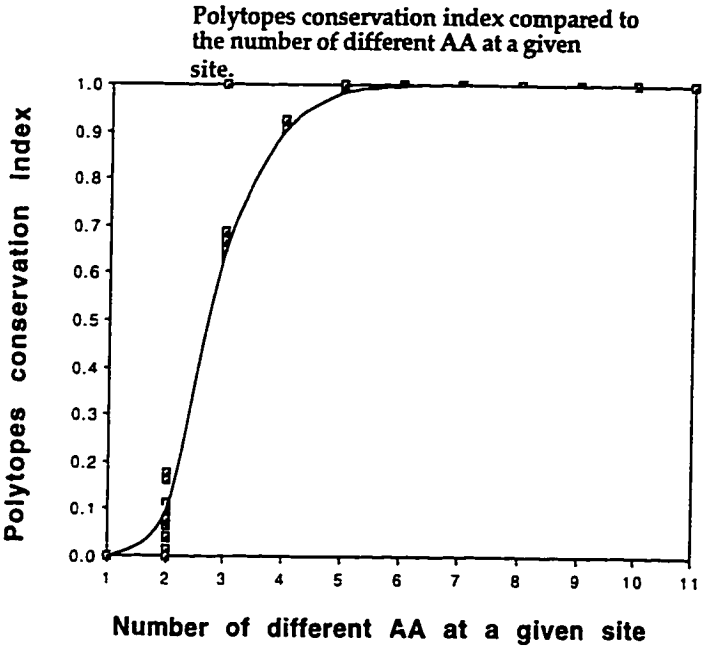
There exists a polytope, defined as the closed geometrical object in a N-1 dimensional space defined by these $(N*(N-1))/2$ distances so that the space contained within it is the minimal possible.

For the above example; 3 AA define 3 distances which define a triangle in a 2-dimensional space (the polytope). The area of the triangle is a quantitative measurement of the conservation at that position. In general the "volume"-or space contained- of the polytope is a quantification of the conservation; i.e. a point has 0 volume and it is 100% conserved, and a line quantifies the conservation having only 2 AA by its length or distance $P(AA_i \rightarrow AA_j)$. The following formula calculates the volume of the polytope in a general case from pairwise $P(AA_i \rightarrow AA_j)=P(i,j)$ mutation matrices, this volume defines the conservation index at a given AA site V(NAA):

$$V\{AA_1,AA_2,\ldots,AA_N\} = \sqrt{\frac{(-1)^N}{2^{N-1}\times((N-0)!)^2}\times \begin{vmatrix} 0 & 1 & 1 & . & 1 \\ 1 & 0 & P^2(1,2) & . & P^2(1,N) \\ 1 & P^2(2,1) & 0 & . & P^2(2,N) \\ . & . & . & . & . \\ 1 & P^2(N,1) & P^2(N,2) & . & 0 \end{vmatrix}}$$

Since this volume correspond to the penalty attached to perform the necessary mutations, and this penalty is equal to one minus the probability, the actual values vary between 0 and 1. Zero being maximal conservation and 1 being minimal conservation. This calculation assumes that $P(AA_1->AA_2)=P(AA_2->AA_1)$, which is not true for mutation matrices -e.g. a Phe to Ala mutant has different penalty than an Ala to Phe mutant. I have selected among both of them the direction that resembles divergent evolution; from the AA with the higher conservation, defined as the number of occurrences, to the AA with the lower conservation at a given position. If both AA occur equally then I calculate the average among the two possible penalties.

Figure 2.17



The conservation index proposed here is compared in Figure 2.17 to the previously proposed estimation (Donnelly 1989), that was based on the number of different AA occurring at a given site. The index proposed here includes not only the variability of the site but also the relation among the replacing AAs. The values are calculated from 224 AA positions in the alignment of GPCRs developed here (see Table 4.1) including only putative TMH domains. This plot shows three marked features all indicating the improvement made with the conservation index. First, instead of a line we observe a sigmoidal curve; i.e. an increment of one in the number of different AA present at a given site contributes differently if it increments from 2 AA to 3 AA or from 9 AA to 10 AA. An important implication of the sigmoidal character of the curve in Figure 2.17 is that positions with more than 5

different AA are considered all maximally variable, i.e. the probability of this site being conserved is already nil for 6 AA so that no information content is found for AA positions with more than 5 AA. Second, the same number of different AA -e.g. 3- corresponds to similar but not identical conservation indexes; this value changes in accordance with the different probabilities of the specific mutations replacing one AA by another. Third, the point {3,1} in Figure 2.17 represents a position where there are 2 different AA and a gap, thus a total of 3 different AA which corresponds to a conserved position according to previous estimates. However, a gap is a strong indication of non-conservative character at that position, as it is correctly indicated by the conservation index -see point {3,1} in Figure 2.17. Conserved positions are defined by a conservation index < 0.70, which corresponds to a probability greater than 30% that this position would be conserved. Note the parallel between the threshold for the conservation index (70%) and the threshold for AA properties quantified through the percentage of AA content (70%).

2.3.2.- New methods applied for tertiary structure analysis:

Two new methods are described. The first constitutes a quantitative formulation of criteria for the prediction of lipid-facing residues in terms of a probability function (Section 2.3.2.1). The second represents a novel graphical approach for the prediction of the direction and degree of helix tilt from the analysis of the helix-helix versus helix-lipid interface patterns (Section

2.3.2.2).

2.3.2.1.- Quantification of criteria for the prediction of TMH-Lipid interfaces: Residue positions highly variable through evolution yet conserving a high hydrophobicity character are proposed to face the lipid chains.

The widespreaded energetic argument that highly hdp AA are likely to face the lipid milieu has been coupled to the low conservation profile of surface-exposed residues to identify (Rees 1989a; Rees 1989b) or predict (Donnelly 1989; Donnelly 1993) TMH-Lipid interfaces. This specific hypothesis has been quantitatively proven in the PRC protein family through the calculation of AA substitution matrices for lipid-exposed residues (Donnelly 1993). Thus, residues highly variable though conserving a high hdp character would be proposed to face the lipid chains. In terms of the quantitative analysis proposed in Section 2.3.1, such residues would have a high conservation index (CI), low hdp standard deviation (σ_{hdp}) and high hdp mean value. This criterion can be quantitated by explicitly defining the probability of an AA to lie in a TMH-Lipid interface, in terms of the properties defined above, as follows:

$$P(\text{TMH-Lipid}) = (1-CI)*(\text{mean}(hdp)+4.5)/\sigma_{hdp}$$

Here, the conservation index is always greater than 0.7 to restrict the analysis to poorly conserved positions. The range of hydrophobicity values

for the 20 natural AAs (-4.5,4.5) is rescaled to (0,9) by adding 4.5 in the formulae (above), in order to avoid negative probabilities.

The predictive power of this probability function was tested on the known structure of the Photosynthetic Reaction Center (PRC) based on an analysis of a multiple sequence alignment of 12 PRC sequences of the homologous L and M subunits, shown in Table 2.5. The satisfactory prediction of AA facing the lipid-chains using the P(TMH-Lipid) function, compared to the % solvent accessibility surface per AA for helix 5 of the PRC L and M subunits combined, is shown in Figure 2.18. Note the discrepancy at the cytoplasmic N-terminus due to Arg/Lys facing the phospholipid head-groups. Although this probability function correctly predicts the solvent accessibility pattern (see Figure 2.18), for modeling purposes it is more useful to select those sites strongly predicted to face the lipids based on this criterion. The proposed function is very discriminant in that only few residues are predicted with high confidence, as shown in Figure 2.19 for the PRC analysis. A threshold value of 6 for the probability function has a very high discriminative power, and is thus selected for prediction purposes.

Table 2.5 :

Multiple sequence alignment of 12 Photosynthetic Reaction Center (PRC) sequences of the homologous L and M subunits. Protein names correspond to their respective codes in the Swissprot sequence database.

```
                1:::::::::11:::::::::21:::::::::31:::::::::41:::::::::51::::::::::
RCEL_RHOVI      1:-----------------------------------ALLSFERKYRVRGGTLIGGDLFDFWVGP
RCEL_RHOSH      1:-----------------------------------ALLSFERKYRVPGGTLVGGNLFDFWVGP
RCEL_RHOCA      1:-----------------------------------ALLSFERKYRVPGGTLIGGSLFDFWVGP
RCEL_RHORU      1:-----------------------------------ALLSFERKYRVRGGTLIGGDLFDFWVGP
PRCLES          1:-----------------------------------MALLSFERKYRVRGGTLVGGDLFDFWVGP
PRCLCA          1:MSRAKAKDPRFPDFSFTVVEGARATRVPGGRTIEEIEPEYKIKGRTTFSAIFRYDPFDFWVGP
PRCMCA          2:----------------------MATINMTPGDLELGRDRGRIGKPIEIPLLENFGFDSQLGP
RCEM_RHOVI      2:------------ADYQTIYTQIQARGPHITVSGEWGDNDRVGKPFYSYWL---GKIGDAQIGP
RCEM_RHOSH      2:------------AEYQNIFSQVQVRGPADLGMTEDVNLANRSGVGPFSTL--LGWFGNAQLGP
RCEM_RHOCA      2:------------AEYQNFFNQVQVAGAPEMGLKEDVDTFERTPAGMFNIL---GWMGNAQIGP
PRCMES          2:----------MYPEYQNIFTQVQVRGTPEMGMDDAGNNMMEERVGKPFFSTLAGLFGNGQIGP
RCEM_RHORU      2:------------SEYQNILTGVQVRTAPHSAPIAKGIFPRLGKPGFSYWL---GKIGDAQIGP

                65:::::::::75:::::::::85:::::::::95:::::::::105:::::::115::::::::::
RCEL_RHOVI      1 YFVGFFGVSAIFFIFLGVSLIGYAASQGPIWDP-------FAISINPPDLKYGLGA-APLLEG
RCEL_RHOSH      1 FYVGFFGVATFFFAALGIILIAWSAVLQGTWNP-------QLISVYPPALEYGLGG-APLAKG
RCEL_RHOCA      1 FYVGFFGVTTIFFATLGFLLILWGAAMQGTWNP-------QLISIFPPPVENGLNV-AALDKG
RCEL_RHORU      1 FYVGFFGVTTLLFTVLGTALIVWGAALGPSWTF-------WQISINPPDVSYGLAM-APMAKG
PRCLES          1 FYVGFFGVTTAFFALLGTILIFWGASQQGTFNP-------WLINIAPPDLSYGLGL-APLLEG
PRCLCA          1 FYVGFWGFVSVIGIIFGSYFYINETILKGPYSI---PQNFFAGRIDPPPPELGLGFAAPGEPG
PRCMCA          2 FYLGFWNAVAYITGGIFTFIWLMVMFAQVNYNPVAFAKYFVVLQIDPPSSRYGLSF-PPLNEG
RCEM_RHOVI      2 IYLGASGIAAFAFGSTAILIILFNMAAEVHFDPLQFFRQFFWLGLYPPKAQYGMGI-PPLHDG
RCEM_RHOSH      2 IYLGSLGVLSLFSGLMWFFTIGIWFWYQAGWNPAVFLRDLFFFSLEPPAPEYGLSFAAPLKEG
RCEM_RHOCA      2 IYLGIAGTVSLAFGAAWFFTIGVWYWYQAGFDPFIFMRDLFFFSLEPPPAEYGLAI-APLKQG
PRCMES          2 YYFGWTSIVAFGTGIAWFVIVGFNMLAQVGWSIPQFIRQLFWLALEPPSPEYGLSM-PPLNDG
RCEM_RHORU      2 IYLGTTGVLSLVFGFFAIEIIGFNLLASVNWSPMEFGRQFFWLGLEPPAAEYGLGF-APLAEG

                128:::::::138:::::::148:::::::158:::::::168:::::::178::::::::::
RCEL_RHOVI      1 GFWQAITVCALGAFISWMLREVEISRKLGIGWHVPLAFCVPIFMFCVLQVFRPLLLGSWGHAF
RCEL_RHOSH      1 GLWQIITICATGAFVSWALREVEICRKLGIGYHIPFAFAFAILAYLTLVLFRPVMMGAWGYAF
RCEL_RHOCA      1 GLWQVITVCATGAFCSWALREVEICRKLGIGFHIPVAFSMAIFAYLTLVVIRPMMMGSWGYAF
RCEL_RHORU      1 GLWQIITFSAIGAFVSWALREVEICRKLGIGYHIPFAFGFAILAYVSLVVIRPVMMGAWGYGF
PRCLES          1 GLWQIITICATGAFISWALREVEICRKLGMGYHVPFGFAAAIIAYMTLVIFRPLLMGAWGHGF
PRCLCA          1 FAWQMTVLFATIAFFGWMMRQVDISMKLDMGYHVPIAFGVAFSAWLVLQVIRPIALGMWHEGF
PRCMCA          2 GWWLIATFFLTVSIFAWYMHIYTRAKALGIKPYLAYGFTGAIALYLVIYIIRPVWMGDWSEAP
RCEM_RHOVI      2 GWWLMAGLFMTLSLGSWWIRVYSRARALGLGTHIAWNFAAAIFFVLCIGCIHPTLVGSWSEGV
RCEM_RHOSH      2 GLWLIASFFMFVAVWSWWGRTYLRAQALGMGKHTAWAFLSAIWLWMVLGFIRPILMGSWSEAV
RCEM_RHOCA      2 GVWQIASLFMAISVIAWWVRVYTRADQLGMGKHMAWAFLSAIWLWSVLGFWRPILMGSWSVAP
PRCMES          2 GWYIIASFFLLVSVMIWLLRAYLLAEQHKMGKHIFWGFAAAVWLFLVLGLFRPILMGSWSEAV
RCEM_RHORU      2 GWWQIAGFFLTTSILLWWVRMYRRARALKMGTHTAWAFASAIFLFLSLGFIRPLLMGNFSESV
```

```
                  191:::::::201:::::::211:::::::221:::::::231:::::::241::::::::::
RCEL_RHOVI      1 PYGILSHLDWVNNFGYQYLNWHYNPGHMSSVSFLFVNAMALGLHGGLILSVANPGDG------
RCEL_RHOSH      1 PYGIWTHLDWVSNTGYTYGNFHYNPAHMIAISFFFTNALALALHGALVLSAANPEKG------
RCEL_RHOCA      1 PYGIWTHLDWVSNTGYTYGNFHYNPFHMLGISLFFTTAWALAMHGALVLSAANPVKG------
RCEL_RHORU      1 PYGFMTHLDWVSNTGYQYANFHYNPAHMLGITLFFTTCLALALHGSLILSAANPGKG------
PRCLES          1 PYGIFSHLDWVSNVGYAYLHFHYNPAHMLAVTLFFTTTLALALHGGLILSACNPEKG------
PRCLCA          1 VLGIMPHLDWVSNFGYRYNNFFYNPFHAIGITGLFASTWLLACHGSLILSAAQYRGP------
PRCMCA          2 AHGIKALLDWTNNVSVRYGNFYYNPFHMLSIFFLLGSTLLLAMHAGTIWALEKYAAHEEWNEI
RCEM_RHOVI      2 PFGIWPHIDWLTAFSIRYGNFYYCPWHGFSIGFAYGCGLLFAAHGATILAVARFGGDREIEQI
RCEM_RHOSH      2 PYGIFSHLDWTNNFSLVHGNLFYNPFHGLSIAFLYGSALLFAMHGATILAVSRFGGERELEQI
RCEM_RHOCA      2 PYGIFSHLDWTNQFSLDHGNLFYNPFHGLSIAALYGSALLFAMHGATILAVTRFGGERELEQI
PRCMES          2 PYGIFPHLDWTTAFSIRYGNLYYNPFHCLSIVFLYGSVLLFCMHGGTILAVTRYGGDRELEQI
RCEM_RHORU      2 PFGIFPHLEWTNSFSLNYGNFFYNPFHMLSIAFLYGSALLSAMHGATILAVSRLGGDREVEQI

                  254:::::::264:::::::274:::::::284:::::::294:::::::304::::::::::
RCEL_RHOVI      1 -DKVKTAEHENQYFRDVVGYSIGALSIHRLGLFLASNIFLTGAFGTIASGPFWTRGWPEWWGW
RCEL_RHOSH      1 -KEMRTPDHEDTFFRDLVGYSIGTLGIHRLGLLLSLSAVFFSALCMIITGTIWFDQWVDWWQW
RCEL_RHOCA      1 -KTMRTPDHEDTYFRDLMGYSVGTLGIHRLGLLLALNAVFWSACCMLVSGTIYFDLWSDWWYW
RCEL_RHORU      1 -EVVKGPEHENTYFQDTIGYSVGTLGIHRVGLILALSAVVWSIICMILSGPIYTGSWPDWWLW
PRCLES          1 -EEAKTPDHEDTFFRDFIGYSVGTLGIHRLGYLLAINAGLWSAICIIISGPVWTAGWPEWWNW
PRCLCA          1 ----EGGDIENVFFRDVQYYSVGESGVHRLGYIFAIGGILSADLCILLSGWPVQDWVSFWNFW
PRCMCA          2 QAPGTGTERAQLFWRWCMGFNANAYSIHLWAFWFAWLCGITGALGVFFSMPDFVNNWFQWGIE
RCEM_RHOVI      2 TDRGTAVERAALFWRWTIGFNATIESVHRWGWFFSLMVMVSASVGILLTGTFV-DNWYLWCVK
RCEM_RHOSH      2 ADRGTAAERAALFWRWTMGFNATMEGIHRWAIWMAVLVTLTGGIGILLSGTVV-DNWYVWGQN
RCEM_RHOCA      2 VDRGTASERAALFWRWTMGFNATMEGIHRWAIWMAVMVTLTGGIGILLSGTVV-DNWYVWAQV
PRCMES          2 YDRGTATERAALFWRWTMGFNATMEGIHRWAWWFAVLTPITGGIGILLTGTVV-DNWFIWAQE
RCEM_RHORU      2 TDRGTAAERAALFWRWTMGFNATMESIHRWAWWFAVLCTFTGAIGILLTGTVV-DNWFEWGVK

                  317:::::::327:::::::
RCEL_RHOVI      1 WLDIPFWS
RCEL_RHOSH      1 WVKLPWWANIPGGING
RCEL_RHOCA      1 WVNMPFWADMAGGING
RCEL_RHORU      1 WQKLPFWNHG
PRCLES          1 WLDMPIWGEPIAVIGGM
PRCLCA          1 NNLPFWSGV
PRCMCA          2 AGINYPQGPTPPVSLP
RCEM_RHOVI      2 HGAAPDYPAYLPATPDPASLPGAPK
RCEM_RHOSH      2 HGMAPLN
RCEM_RHOCA      2 HGYAPVTP
PRCMES          2 HHFAPMYDGSYGYEDYGSYEAFIGKEN
RCEM_RHORU      2 HGLAPAP
```

Figure 2.18

Satisfactory prediction of AA facing the lipid-chains using the P(TMH-Lipid) function, compared to the % solvent accessibility surface per AA for helix 5 of the PRC L and M subunits combined. Note the discrepancy at the cytoplasmic N-terminus due to Arg/Lys facing the phospholipid head-groups.

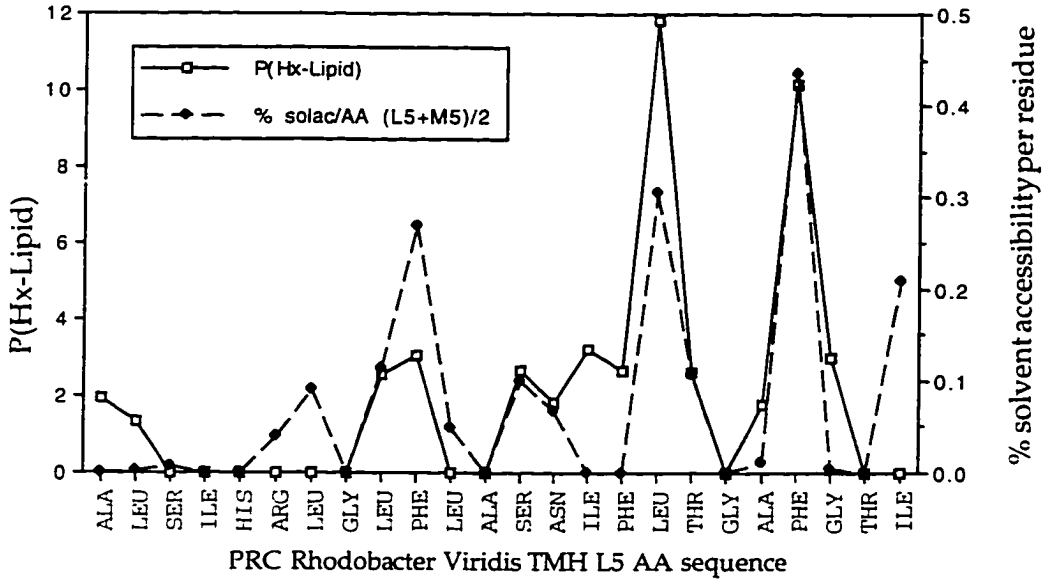
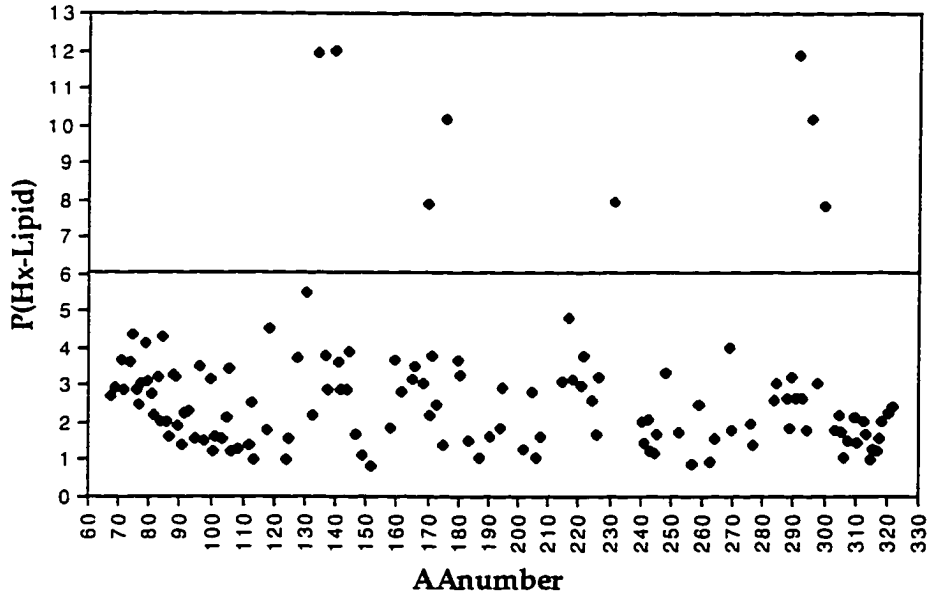


Figure 2.19

Testing the discriminative power of the probability function P(TMH-Lipid) in predicting lipid-facing residues for the PRC L and M subunits. A threshold value of 6 is selected for prediction purposes.



2.3.2.2.- Prediction of the direction and degree of helix tilt from the analysis of the helix-helix versus helix-lipid interface patterns: Definition of the relation between surface patches and α-helix tilting.

Consider an α-helical segment partially exposed to solvent, as it occurs frequently in known protein structures. The residues of this segment facing the protein interior define continuous patches on the surface of the α-helix (Chothia 1981). The shape of these patches would reflect the orientation in which the α-helix is packed against the protein structure. Understanding the relation between the shape of the helical patches and the orientation of the packed helix would allow the *de novo* prediction of the orientation in which an α-helix (i.e., a TMH) packs against a protein (i.e., the interior of the TMH bundle). The orientation of a TMH as it packs against the TMH domain, without considering translations, can be geometrically defined by four tilting components, described in Figure 2.20: the helix can tilt towards the lipids (outwards) or towards the protein interior (inwards), and can tilt in a clockwise (left), or an anticlockwise (right) direction. Note that I have selected a cytoplasmic perspective of the TMH bundle for this analysis. The reason is that the predicted tiltings for the 5HT2CR would be compared to the observed tiltings for the rhodopsin structure, which are always described from a cytoplasmic perspective. Each of these four tilting components results in a distinctive shape of the helical patch defined by interior-facing residues, as illustrated in Figure 2.21 relative to a straight α-helix. The distinct shapes

corresponding to each tilting component can be recognized in helical net representations, as shown in Figure 2.21. These helical nets are generated by opening the cylinder defined by the C^{α} coordinates of the α-helix (Chothia 1981) and are thus a geometrically correct two-dimensional representation of the surface of a three-dimensional object (an ideal α-helix). Therefore, recognition of the distinct shapes described above in the patches of predicted TMH-TMH interfaces can be used to derive the direction and degree of helix tilt for each TMH, as shown in Section 6.3 for the 5HT2CR.

Figure 2.20:
Helix tilting may be expressed by 4 components: (Cytoplasmic view)

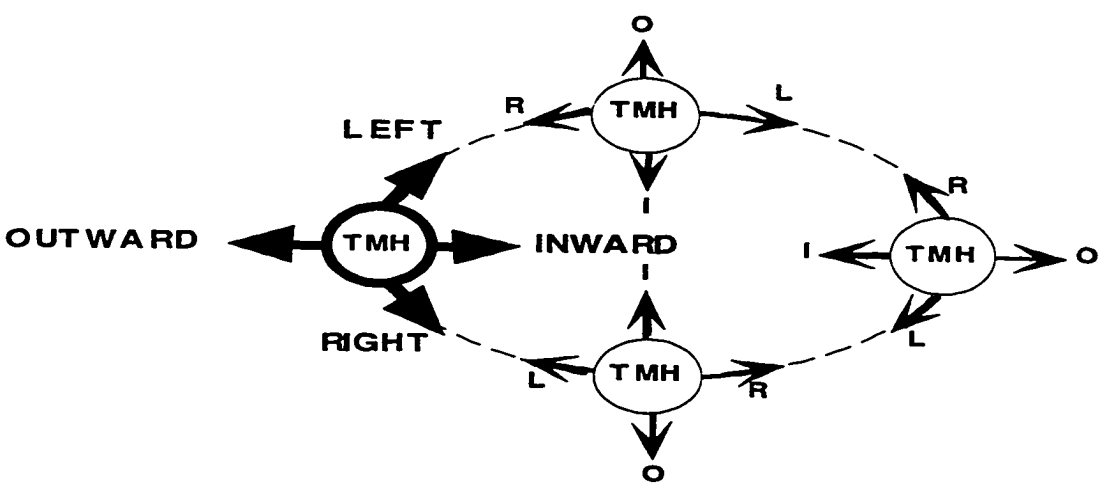
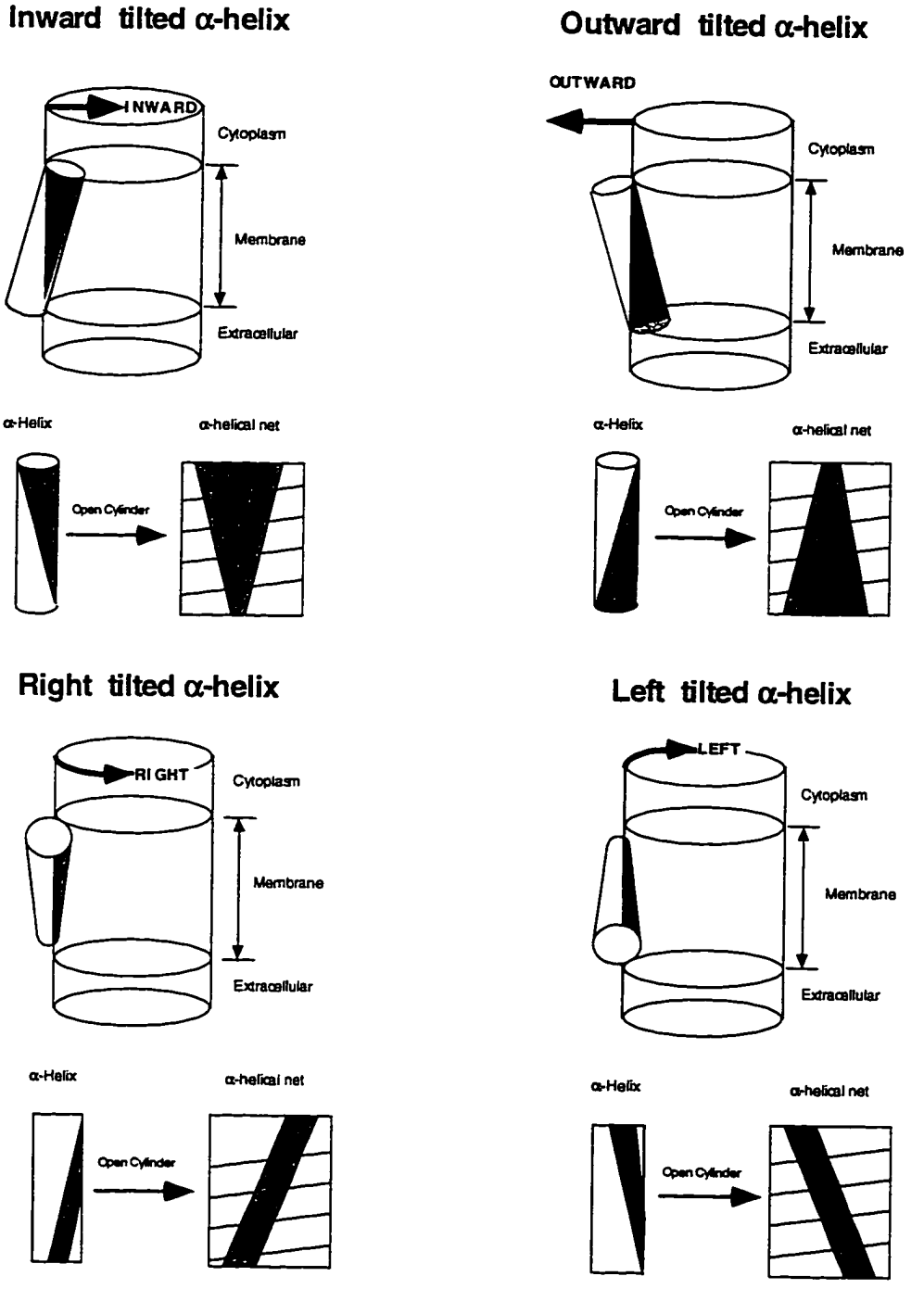


Figure 2.21

Helix tilting may be recognized in helical net representations: 1) Residues facing the protein interior define continuous patches on the helix surface. 2) These patches have distinctive shapes for each tilting component. 3) These shapes can be recognized in helical net representations.



2.3.3.- Energy based methods: MonteCarlo and Molecular Dynamics simulations.

The novel technique of conformational memories (Guarnieri 1996), an energy-based MonteCarlo procedure developed by Dr. Frank Guarnieri, has been used to explore the conformations of several structural motifs. A description of this method as applied for the studies on the "Arg-cage" motif (Section 7.3.2.1) follows. Other applications of this method followed similar methodological procedures unless otherwise stated, with the exception of the mixed water-membrane solvent model used only in the Arg-cage studies.

The segment of TMD 3 from C3.25 to T3.55 was modeled as an α-helix. the helix boundaries, predicted following the methodology described elsewhere (Ballesteros 1995) are in agreement with recent experimental results for other GPCRs (Farahbakhsh 1995; Javitch 1995a). The environment surrounding the Arg residue in TMD3 has been determined using spin labeling experiments in rhodopsin (Farahbakhsh 1995), found to include two distinct phases, a water phase and the membrane-embedded protein, whose boundary was determined between residue 3.52 (in the membrane) and 3.53 (in water). Therefore, the environment of the $R^{3.50(139)}$ was assumed to be similar to the environment of residues buried inside the protein interior, following studies on the known structure of the Photosynthetic Reaction Center (Rees 1989b), and thus modeled by a distance dependent dielectric. To simulate the byphasic environment in our calculations, we used a novel

mixed solvent model composed of a water phase and a distance dependent dielectric phase (Guarnieri 1997). The boundary between the two phases consisted of a plane parallel to the membrane, and TMD3 was initially positioned perpendicular to the membrane plane at the midpoint between residue 3.52 and 3.53 following the experimental observations, but was allowed to freely rotate and move ± 2.3 Å in the direction normal to the membrane plane to prevent arbitrary effects due to the starting point.

To explore the conformational space available for the cytoplasmic end of TMD3, MonteCarlo simulations on the wild type and mutant constructs were performed by varying of the torsional angles for residues $V^{344(133)}$ to $T^{3.55(144)}$. The variation of backbone dihedral angles *phi* and *psi* were restrained to ± 20° from their initial values. Side chain dihedral angles were rotated freely. Extensive simulations are necessary to reach convergence of the resulting conformations at the level of the rotamers analyzed (correlation of four consecutive dihedrals for Arg, see below). Thus, between 100 and 400 rounds of independent random simulations were performed for each TMD3 construct. In each round, repeated runs of MonteCarlo simulated annealing were performed from a starting temperature of T_1=2070K, with a cooling schedule of $T_{n+1} = 0.9*T_n$, and 10,000 steps per temperature to reach 310 K. This implies that between 19 and 76 million conformations were sampled for each wild type and mutant construct. Analysis of the resulting conformations was performed at T=310 K and restricted to backbone conformations within ±

10° from their initial values to maintain an a-helical conformation.

The conformation of the side chains, in particular the $R^{3.50(139)}$ side-chain, are defined by the corresponding dihedral angles $\chi 1$, $\chi 2$, $\chi 3$, $\chi 4$, etc.... These dihedral angles are classified according to three main rotamers: *gauche* plus (g+) centered on -60 ° (and encompassing angle values between -120 ° and 0 °), *gauche* minus (g-) centered on +60° (between 0 ° and 120 °), and *trans* (t) centered on 180° (between 120 ° and -120 °). Evaluation of the preferred conformations was performed by analyzing the populations of each side chain rotamer. For the Arg side chain, rotamers were defined based on its first four dihedral angles: the rotamer state of the Arg side chain is defined by the state of each one of its four dihedral angles, e.g. ($\chi 1$, $\chi 2$, $\chi 3$, $\chi 4$)=(g+, g+, g-, t). Thus there are 81 possible Arg side chain rotamers. These rotamers were grouped according to their spatial orientation towards specific residues. The spatial orientation of each $R^{3.50(139)}$ rotamer was inferred from the average values of each of the four dihedrals within each rotamer.

Energy-based methods used to explore the conformational preferences of the entire TMH domain used Molecular Dynamics procedures rather than the MonteCarlo procedures described above to explore the conformations of individual TMH segments. The three-dimensional model of the TMH domain of the 5HT2CR complexed with 5HT was subjected to Molecular Dynamics simulations. The simulations were done with the CHARMM software package (Brooks 1983) using the CHARMM24 potential (MacKerell Jr

1995) where all H atoms are included. SHAKE was applied to all bonds. A spherical cutoff of 13 Å with shifting functions was used for both the Van der Waals and electrostatic terms. The environment was approximated by scaling the electrostatic interactions with a distance dependent dielectric (ε=1). The leap-frog integrator was used with a time step of 0.5 femtoseconds. Following minimization by multiple rounds of steepest descent and ABNR, the system was heated to 60 K in 1.5 picoseconds (ps), equilibrated at 60 K for 3.5 ps, heated to 180 K in 1.5 ps, equilibrated at 180 K for 3.5 ps, heated to 300 K in 1.5 ps, and equilibrated at 300 K for 1,250 ps.

Because the environment of the TMH domain was represented by a distance dependent dielectric without considering any explicit solvent molecules, charged groups at the cytoplasmic boundaries were neutralized by reducing their partial charges. The ionic interactions proposed among them were emulated by distance constrains using the RDIS command with a force constant of 30 kcal/mol and the minimum at 3.0 Å. These residues and their emulated interactions were D3.49-R3.50, R3.50-E6.30, E6.30-R6.31, E1.60-K1.61. The charge on D2.50 was maintained but a distance constraint as described above was applied between D2.50-N7.49 enforcing our experimentally validated H2-H7 interaction (Zhou 1994; Sealfon 1995). The charge on serotonin and D3.32 was maintained, and the proposed H-bond interactions between 5HT and the receptor sites D3.32, S3.36, A5.46, N6.55 enforced as described above except for 5.46 (3.5 Å) for the first 350 ps. After these 350 ps

and for the rest of the simulation, only the interactions with D3.32 and 5.46 were constrained. During heating and the first 200 ps of equilibration, all *phi* and *psi* angles were constrained within ±10° of the initial values by an harmonic potential with a force constant of 100 kcal/(mol*rad^2). After 200 ps the system became conformationally stable as judged by the low rms deviation (see Figure 8.1), and then these dihedral constrains were reduced at the Pro-kink regions, comprising residues (i, i-4) from the proline, to a force constant of 25 kcal/(mol*rad^2) for the rest of the simulation.

3- A THERMODYNAMIC FRAMEWORK TO MODEL AND INTERPRET STRUCTURE-FUNCTION RELATIONSHIPS ON GPCRS.

There are several functional and conformational states of the receptor that need to be considered in order to interpret current experimental data in terms of structure-function relationships. It follows that experimentally proposed interactions may apply for only one or some of these receptor states but not others, and that not one but several structural models of the 5HT2CR would be needed to rationalize these experimental findings. Therefore, a prerequisite for using experimentally derived interactions in modeling the 5HT2CR is to define the state of the receptor being studied, and to assign experimentally derived structural inferences to specific functional states. This complexity of diverse receptor states needs to be rationalized in a formalistic framework where experimental data and theoretical predictions from various sources and receptors can be integrated into a coherent, self-consistent description of receptor activation. There are two levels of this description; a thermodynamic and a structural level, applied and described below in this order.

A thermodynamic model of receptor activation defines the different functional states and their relationship and is described in this chapter, i.e. as present in the ternary model of receptor activation (Lefkowitz 1993; Samama 1993) described in Section 3.1. This initial framework allows the

interpretation of experimental observations within functionally defined receptor states. However, most experimental findings are obtained under conditions in which several of these states are present simultaneously, i.e. agonist affinity is measured in the presence of both inactivated and activated receptors. Therefore, in order to interpret experimental observations in terms of structure-function relationships, we need to discern which of the existing states of the receptor was responsible for the observed phenotype, i.e. agonist high and low affinity states would correspond to agonist binding to the active and inactive forms of the receptor, respectively. This distinction requires an understanding of which states of the receptor are present under the particular experimental conditions, i.e. which states are significantly populated. The relative populations of the different receptor states considered, under conditions of thermodynamic equilibrium, are determined by the free energy differences among these states. Therefore, the prevaling thermodynamic model of receptor activation, so called the ternary complex or two-state model (Lefkowitz 1993; Samama 1993), needs to be analyzed in an energetic context from which the population of the different receptor states can be derived. I have proposed such an energetic framework to interpret receptor activation, described below in Section 3.2, and novel and significant implications arise from such analysis. The proposed energetic context of the thermodynamic model has been validated experimentally in collaborative work with the laboratory of Dr. Kobilka (Gether 1996), and has received further support from

recent studies on rhodopsin. Experimental validations are described below in Section 3.3.

Once the there is a thermodynamic framework to model and interpret structure-function relationships on GPCRs, the next level of description in modeling the process of receptor activation is the structural level, described in the remaining chapters of this thesis. Based on the thermodynamic model and ensuing population analysis, the available experimental data pertinent to structure can now be ascribed to specific, functionally identified receptor states, i.e. it is found that the agonist-bound form of the receptor in the absence of G-proteins is significantly populated by the inactive state of the receptor. This leads to specific and distinct structural hypotheses for the 5HT2C receptor for the different functional sates considered. In developing structural models for the receptor states present in the thermodynamic model, it should be noted that a given functional state as defined above does not correspond to a single conformation of the receptor. The trivial example is when different chemical species are considered, such as binding of different ligands or different G-proteins. Even for the simplest case of an isolated, uncomplexed receptor in a given functional state, the dynamic properties observed for other proteins suggest the presence of multiple conformations of the receptor coexisting, which could share the same functional phenotype or "state" (e.g., see (Frauenfelder 1988)). This multiplicity of structures for a single state defined based on functional and thermodynamic criteria arises

primarily from thermal fluctuations in the receptor structure unrelated to function, e.g., local rotations of a lipid-exposed residue. Nonetheless, these multiple receptor conformations with a similar functional (and thermodynamic) phenotype are expected to share specific structural components that differentiates them from the sets of multiple conformations present in other functional states. Thus the structural equivalent of functionally and thermodynamically defined receptor states can be better described as "conformational states". The conformational changes upon receptor activation can be considered in a similar manner. If we assume an "activation pathway" for the receptor that is common to agonists and partial agonists, then the term "receptor activation" has a meaning and would consist of a series of conformational changes involving specific residues and movements of helices and Pro-kinks. While many conformational changes occur in the receptor upon ligand binding, some would belong to the activation pathway, and some would be dependent on the chemical nature of the specific ligand but not directly related to the activation mechanism. For example, the Pro-kink in TM VI is likely to be an intrinsic hinge of the activation mechanism (see Sections 8.2.4 and 8.3.3), so that activation can be characterized by the conformational changes at this particular locus. A counter-example would be a direct receptor-ligand contact, for instance with a Leu on the TMIV-TMV loop, which alters the local conformation of the Leu side chain but has no bearing on the activation state of the protein.

3.1.- The ternary complex model: The minimal set of molecular entities necessary to relate the structural model of the receptor to its functions of ligand binding and activation.

In order to rationalize the complexity of different functional states of the receptor into the formalism of a thermodynamic framework, we need to define the minimal set of states necessary to interpret current experimental data in a structural context. Minimally, we need to consider the receptor in its inactive (R) and active forms (R*), its corresponding interactions with a ligand (LR and LR*), and the interactions of these molecular species with the G-protein (LRG and LR*G). A thermodynamic cycle for receptor activation comprising these molecular entities have been proposed based on constitutively active mutants (CAM) of the β2-adrenergic receptor (Lefkowitz 1993; Samama 1993), termed the ternary complex model or the two-state model, described in Figure 3.1. I will use this model throughout this thesis to relate the energy levels of its component states (next section) and to rationalize structure-function relationships in a structural context by considering the conformational states that correspond to the functional states defined in the ternary complex model.

It is clear that this model is an oversimplification of the real receptor activation mechanism, which is likely to be composed of many more substates and variations. There are several receptor states not present in the ternary complex model that I will consider for my structural analysis.

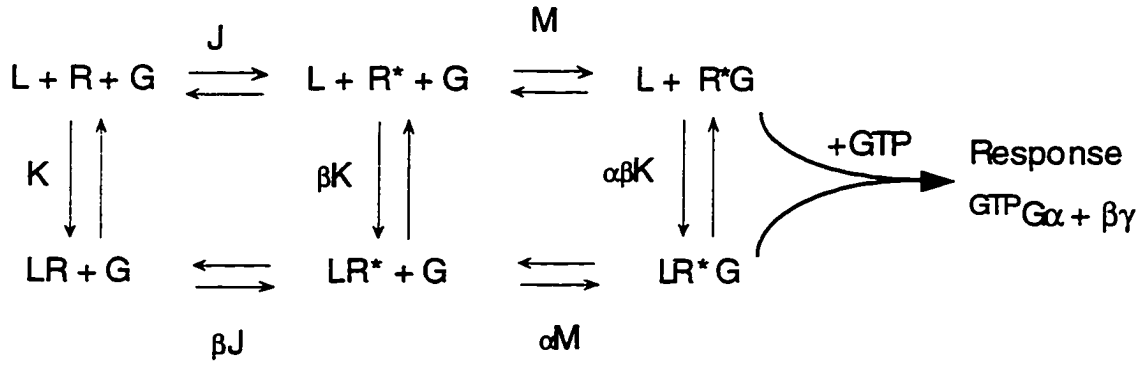
Chemical modification of the receptor that accompanies activation is a clear case for this oversimplification. Palmitoylation has significant structural consequences (Moench 1994; Traxler 1994; Bouvier 1995; Loisel 1996), and its structural and/or functional role will be discussed in the context of the 5HT2C receptor model developed. However, it is not known how states resulting from differential palmitoylation as observed upon receptor activation would fit in the thermodynamic scheme presented above. Another important chemical modification is receptor phosphorylation, which follows receptor activation and induces receptor desensitization. Thus, the pharmacological phenotype of a phosphorylated receptor parallels the inactive form of the receptor in both affinity and activity, but differs from it in that it can no longer interconvert with the active form unless it is chemically modified by a phosphatase. Therefore, the phosphorylated state can be incorporated into the thermodynamic model described above. It has been shown that the same receptor can bind and activate multiple G-proteins (Perez 1996), with different coupling efficacies for the different G-proteins (Perez 1996). This suggests that several different R*G must exist for the different G-proteins, but this variability is not included in the minimal model considered here. This finding parallels the phenotype observed for different ligands binding at the extracellular side of the receptor, which show different affinities and different efficacies. A simple permutation of the specific molecule representing the ligand (L1, L2, ...) or the specific G-protein (G1, G2, ...) in the thermodynamic

scheme considered would simply add mathematical but not conceptual complexity to the model, and has been discussed in detail by Costa et al (Costa 1992; Onaran 1993). The ternary complex model would encompass many ligands and many G-proteins as long as the basic tenet that R* is the same can be maintained. I discussed before that the structural equivalence of a given functional state such as R* is a set of multiple conformations, defined as "activated conformational states", by arguing that the different R* conformations could be ascribed to localized structural fluctuations unrelated to function. In terms of the ligand and/or G-protein complexed receptor this hypothesis suggests that the conformation of the receptor as defined by the structural motifs responsible for a common activation mechanism is maintained for the different complexes available to R*, i.e. L1R* versus L2R*, or R*G1 versus R*G2, etc... . The assumption of a common R* state for different ligands and different G-proteins has been recently contradicted by Perez et al. (Perez 1996) where it has been shown that the CAM C3.35->F in the α1-adrenergic receptor resulted in constitutive activation for only one of the two effector pathways, the other remaining at wild-type levels. This implies that the active conformation of this receptor (R*) is significantly different depending on the specific G-protein that interacts with it, and thus proves the presence of multiple activated states. These same studies by Perez et al. (Perez 1996) also showed that chemically distinct ligands selectively preferred one of the two different activated forms of the receptor stabilized by

different G-proteins; phenethylamines showed higher potencies and intrinsic activities for polyphosphoinositide turnover but not for arachidonic acid release. Imidazolines showed wild-type potencies and intrinsic activities for both pathways. These data indicate that chemically distinct agonists can direct receptor coupling towards specific G-proteins. These findings cannot be rationalized at the level of the simplified ternary complex model, nor do they have a straightforward structural rationale, but they indicate the direction in which our future understanding of the activation mechanism must evolve.

Figure 3.1.

The two-state (or ternary complex) model of receptor activation.



3.2.- Interpretation of the ternary complex model in terms of energy levels for the principal states: A thermodynamic framework for receptor activation suggests an metastable activated receptor in the absence of G-protein.

The derivation of relative energy levels for the receptor states

considered above in the ternary complex model is based on inferences from the wild type receptor as well as on constitutively activated receptor mutants (CAM). Two measurable properties that are shared by the CAM and by the receptor in its complex with agonist are: i) a higher probability of coupling to G-proteins, and ii) a higher affinity for agonists (e.g., see (Kjelsberg 1992; Lefkowitz 1993)). This similarity in the interaction properties of the CAM and activated forms of the receptor led to the assumption that the conformational states of the proteins are similar (Samama 1993) - albeit not necessarily identical, as these complex proteins are likely to achieve similar interaction properties for a family of conformational states.

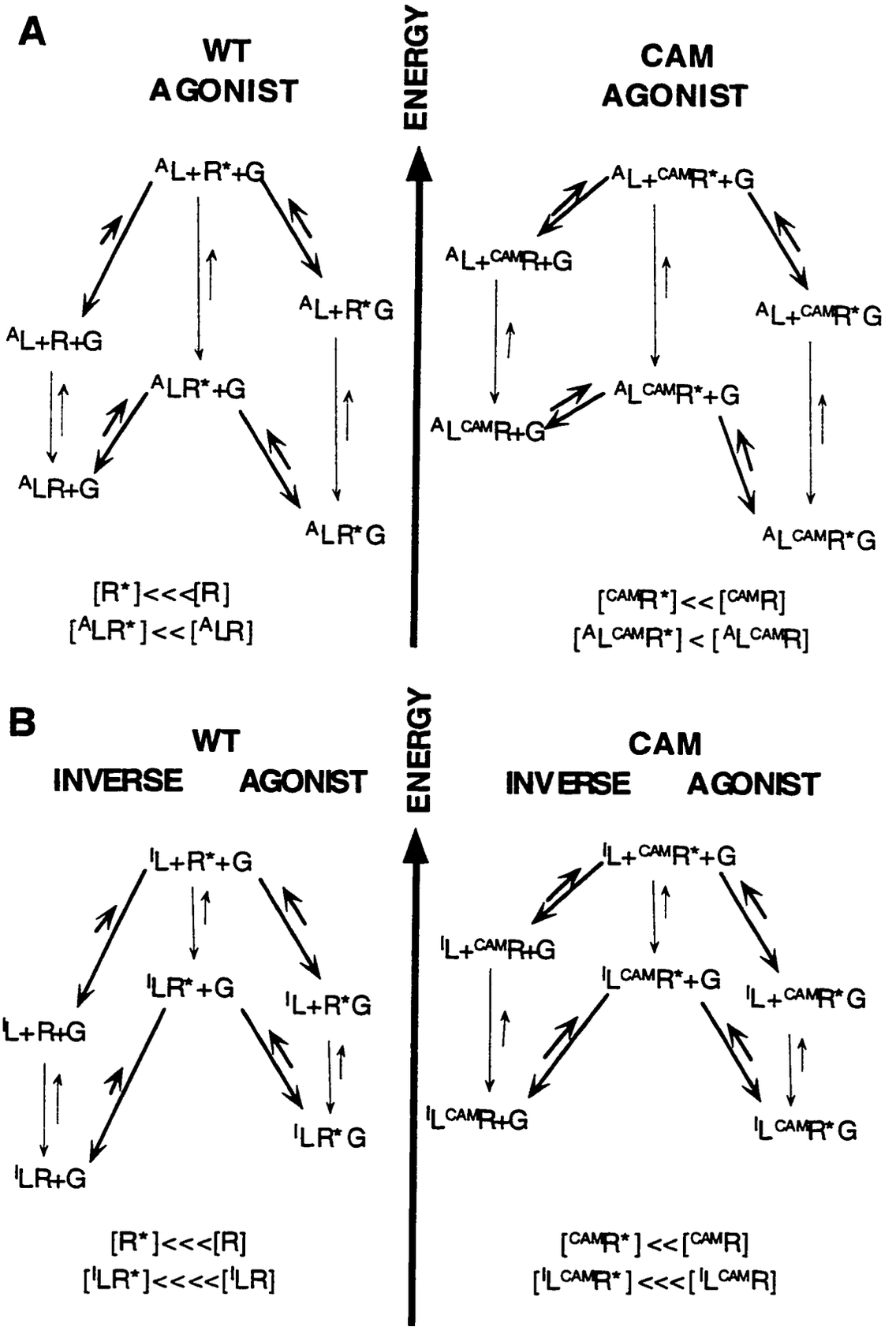
From the dynamic properties of complex proteins (e.g., see (Frauenfelder 1988)) this family of states, designated here as R*, is expected to consist of multiple "activated forms" of the receptor, with different species termed R*i. By definition, the receptor by itself exists in the ground state, which for GPCRs is the inactive state (R). Therefore, the activated state of the receptor protein itself (R*) must correspond to a higher (less stable) free energy level than the unliganded ground state (R) of the wild type receptor (see diagram in Figure 3.2.A). The "basal activity" of wild type protein, measurable in the absence of agonist, can be interpreted as a reflection of the ability of the wild type receptor protein to overcome the energy difference separating the ground state R from an activated state R*i, even in the absence of agonist binding. If equilibrium conditions are established, then the relative

population of the R and R* states will depend on the energy difference between them. To maintain normal signalling, the difference must be high so as to maintain a negligibly low population of R* at equilibrium. Receptors with a high basal level of activity, such as the dopamine D5, would have a significantly lower energy difference for the R->R* transition than most GPCRs.

As for the G-protein coupled form of the receptor in the absence of ligand (R*G), interaction with a heterotrimeric G–protein should stabilize the R* form by virtue of the interaction energy between them, as shown in the unliganded diagram (Figure 3.2.A). Thus the emerging picture of the free energy levels associated with the activation of unliganded receptors, shown in the diagram, is defined by a ground state (R) that can adopt an "excited", higher energy state (R*), which in turn can be stabilized by G-proteins (R*G).

FIGURE 3.2

The two-state model of receptor activation in an energetic context, shown for the wild type and CAM receptors. (A) Ligand is an agonist. (B) Ligand is an inverse agonist. Note the predicted concentrations among activated and inactivated species.



Once it is achieved from an excitation R->R*, the R*i "excited state" of the receptor will have a lifetime that depends strongly on interactions with ligands such as heterotrimeric G–proteins, agonists, and inverse agonists. In this regard, the observation that GTP abolishes the high affinity state of the receptor as well as the coupling to G-proteins indicates that the high-affinity state measured in binding assays corresponds to the active state of the receptor coupled to the G-protein, i.e. ^{A}LR*G. It follows that the low-affinity state of the receptor would correspond to the agonist-bound to the uncoupled and inactive receptor (^{A}LR). Because high-affinity binding is not observed when the receptor is not coupled to the G-protein, it follows that the population of ^{A}LR* is negligible compared to ^{A}LR and hence that the agonist-bound form of the active receptor ^{A}LR* is higher in free energy than the agonist-bound form of the inactive receptor ^{A}LR (see Figure 3.2.A). Conversely, the presence of high-affinity binding in the presence of G-protein under physiological conditions indicates that the agonist-bound form of the active receptor coupled to G-protein (^{A}LR*G) is an energetically stable and favored state, i.e. that (^{A}LR*G) is lower in energy than (^{A}LR*) and (^{A}LR). Thus a consistent picture of the energy levels that accompany agonist-induced receptor activation emerges as shown in Figure 3.2.A, in which the agonist favors the transition of the inactive receptor (^{A}LR) to an active conformation (^{A}LR*) that has the characteristics of a transition state, metastable and with a very low

lifetime, which can be stabilized by G-proteins leading to a stable active form of the receptor consisting of the ternary complex (^{A}LR*G), which is responsible for the agonist-induced activation.

We can now rationalize the effects of antagonists and inverse agonist within the framework of the thermodynamic scheme shown in Figure 3.2. A neutral antagonist would have similar affinities for the active and inactive state of the receptor, and therefore would not affect the equilibrium between the populations of active and inactive states established for the unliganded receptor. In this scenario, the unliganded scheme (Figure 3.2) is still valid as the representation of the free-energy relationships among the conformational states of the receptor, where now R, R*, and R*G would become ^{N}LR, ^{N}LR*, and ^{N}LR*G. Thus the most populated state for a neutral antagonist would be the inactive state (^{N}LR), because the other states were sparsely populated in the absence of ligand.

The case of an inverse agonist is opposite to that of an agonist, in the sense that now the affinity for the uncoupled, inactive state (^{I}LR) is higher than the affinity for the active state (^{I}LR*G). The active state in the absence of G-protein remains an intermediate state, higher in energy than both (^{I}LR) and (^{I}LR*G). The free-energy levels associated with the activation states of an inverse agonist are shown in the inverse agonist diagram in Figure 3.2.B.

The phenotype of the constitutively active mutant (CAM) receptor can

be understood at the molecular level to result from a reduction of the energy gap between the two states in the mutant protein ($^{CAM}R^{*}->^{CAM}R$). These energy changes lead to an increase in the population of the R*i form that has high affinity for G–proteins and for agonists. The observation that all substitutions at 6.34 induce a CAM phenotype (Kjelsberg 1992) was interpreted as a structural perturbation of the inactive state R by these mutations, which in energetic terms translates into a higher energy for (^{CAM}R) than for the wild-type (R). The general relationships that define the energy levels among these states in wild-type as described above are still valid for the CAM receptor. For instance, GTP still abolishes or reduces significantly the high affinity component, indicating that ($^{A}LR^{*}$) in the absence of G-protein is still an "excited", metastable state, and that G-protein binding stabilizes the active state leading to the ternary complex ($^{A}LR^{*}G$). Thus, we can use qualitatively the same energy diagrams as shown for the wild-type simply reducing the energy difference for activation (see diagram in Figure 3.2.A). Note that although it is useful to represent their energy levels in a similar diagram, we do not know the relative energies between the wild type and CAM receptor constructs. In Figure 3.2.A I have arbitrarily adopted an energy level of the inactive and uncoupled state (^{CAM}R) higher than the equivalent wild type state.

The relative energy levels and ensuing populations for the different receptor states described above can be integrated with the ternary complex

model of receptor activation presented in Section 3.1, yielding a thermodynamic framework to interpret structure-function relationships on GPCRs. This is illustrated in Figure 3.2, including the case of agonist and inverse agonist binding to an isolated receptor, the case involved in the experimental validation described below (Gether 1996), and also the case modeled at the structural level in this thesis for the 5HT2C receptor. Note that this thermodynamic model predicts that the agonist-bound receptor is predominantly in the inactive state. This contradicts the previously held notion that the agonist-bound receptor represents the activated form of the receptor. G-protein coupling would change this conclusion because $^{A}LR^{*}G$ is preferred energetically relative to ^{A}LR. However, the net effect of the G-protein will depend on the lifetime of $^{A}LR^{*}G$, which depending on the experimental conditions may be short due to the exchange of GDP by GTP in the $^{A}LR^{*}G$ complex, leading to a desestabilization of the ternary complex that results in $^{A}LR^{*}$ which in turn could decay rapidly to ^{A}LR. This effect is expected to be more pronounced in cell expression systems, the most common procedure to measure ligand affinities nowadays, due to the artificially high receptor to G-protein ratios. It follows that the conformational state of the receptor amenable to pharmacological characterization in terms of ligand binding affinities is highly dependent on the experimental conditions, which for most of the cases I have encountered would favor the inactive form of the receptor.

The recent discovery of constitutively activating mutants (CAM) could be used to circumvent this experimental shortcoming by increasing significantly the population of the activated state of the receptor. However, as discussed earlier and shown in Figure 3.2.A, the conformational state of a CAM receptor would still be dominated by the inactive state. Recently, the activating effect of CAM were shown to be additive for the α1-adrenergic receptor (Hwa 1997), suggesting that the population of activated versus inactivated receptor could be increased significantly by these procedures. Simply increasing the population of the activated state of the receptor is really not the desirable solution for three reasons. First, the extent of the structural perturbations induced by these CAM is presently unknown, but clearly present. Second, current ligand binding analysis tools either can only fit one single "apparent" affinity, or at best they can fit two affinity constants if they are assumed independent from each other, an assumption we know is not true because ^{A}LR and $^{A}LR^{*}$ interconvert in equilibrium (see the ternary complex model in Figure 3.1). It follows that the affinity reported for the "high affinity", supposedly activated receptor, is incorrect and we do not know the extent of the error. Third, increased activation of the CAM receptor results in constitutive desensitization and down-regulation by the cell biochemical machinery (Pei 1994). Because a desensitized receptor has an affinity profile indistinguishable from the inactive receptor, the effect of desensitization imposes a ceiling in the ability of CAM procedures to provide

receptor that is predominantly in the activated form. The conclusion from this analysis is that there is a need for a receptor construct that in its basal form would adopt only the active conformation, but yet that it would not undergo constitutive desensitization, so that we can measure agonist affinities accurately. Based on the understanding of the activation mechanism at the structural level of these receptors gained through the course of this thesis, I have proposed to engineer a Cys-Cys disulfide bridge in the 5HT2CR designed to fit these properties, as discussed at the end of the thesis under inferences from the model (Section 9).

3.3.- Experimental validation of the proposed thermodynamic scheme:

3.3.1- Fluorescence studies of β2-adrenergic receptor activation.

Collaborative work with the laboratory of Dr. Brian Kobilka provided experimental support for the interpretation of the thermodynamic model presented above in terms of energy levels among the various receptor states defined. This work has been published (Gether 1996) and thus I just describe here how the apparently surprising experimental results were readily predicted by the thermodynamic model presented above.

Fluorescent probes were covalently attached to Cys residues on the purified β2-adrenergic receptor. The changes in receptor fluorescence measured upon ligand binding were directly proportional to the

pharmacological efficacies of a variety of ligands (Gether 1996), suggesting that the fluorescent probes were sensing the conformational changes that accompany receptor activation; agonist decreased the fluorescence while inverse agonists increased it (Gether 1996). The surprising observation was that fluorescence labeling of the CAM receptor resulted in larger changes in fluorescence for both agonist and antagonists, as well as an increased instability of the protein. If the CAM induced the active state conformation of the receptor, agonists would have been expected to induce a smaller fluorescent change in the receptor.

The experimental results on the florescence changes between the wild-type and CAM receptors can now be understood in the context of the energy diagrams described above. Thus, interaction with a heterotrimeric G–protein should stabilize the R* form by virtue of the interaction energy between them, as shown in the diagram of Figure 3.2.A, whether the R* is ligand-bound or not. However, in the absence of such a stabilizing complexation with a G–protein, as occurs in the fluorescence experiments where the G-protein is missing, the excitation will decay rapidly, the receptor will return to the more stable R form, and any bound agonist will dissociate from this "low affinity state" of the receptor. Thus, in the absence of G–proteins, the measured affinity of agonists for the CAM receptor construct should appear to be significantly lower than their affinity measured in systems where coupling of R* to the G–protein is possible. Notably, even though the CAM receptor

has a higher probability of attaining the active conformation (R*) in the absence of G-protein, this state is still a metastable state which is not significantly populated.

The observation that the fluorescence change induced by both agonists and antagonists is larger for the CAM receptor than for the wild-type receptor can now be explained in the context the energy levels described above. If the system is in thermodynamic equilibrium, the free-energy levels described for the wild-type and CAM receptors can be directly translated into relative populations (P) of the available states. The available states in the absence of G-protein are (R) and (R*) for the wild-type and the equivalent (^{CAM}R) and (^{CAM}R*) for the CAM receptor. According to the energy levels discussed before, the ensuing populations (P) can be described as:

POPULATIONS (P)

P(R) >>> P(R*)
P(CAMR) >> P(CAMR*)
P(R) > P(CAMR)
P(R*) < P(CAMR*)

If the ligand is a neutral antagonist, there should be no significant change in fluorescence because the populations of the available states do not change significantly. If the ligand is an inverse agonist, it drives the population towards the inactive state from the active state: because the CAM

receptor has more populated the active state than the wild-type, i.e. $P(R^*) < P(^{CAM}R^*)$, the CAM receptor should show a larger fluorescence change than the wild-type receptor, as observed.

Notably, both populations of R* forms (i.e., R* and $^{CAM}R^*$) are very low, so that the differential effect of the inverse agonist would be quite small. In contrast, the effect of the agonist on the fluorescence change observed for the wild-type compared to CAM would be significantly larger. The reason is that the energy difference necessary to change the conformational state of the receptor from the inactive to the active states is significantly smaller in the CAM relative to the wild-type receptor, as discussed above, and thus agonist binding to CAM receptors has a higher probability of producing the active state than the wild-type. It is noteworthy, however, that in case of the CAM receptor the inactive state is slightly less populated than the wild-type, i.e. $P(R) > P(^{CAM}R)$, so that in the case of CAM the pool of receptors whose conformation can be altered would be smaller than the wild-type. However, due to the rapid decay from the R* states, the turnover to R and ^{CAM}R is very rapid.

Because under the experimental conditions of the fluorescence assays the inactive state is overwhelmingly populated in both wild-type and CAM receptors, judged by the lack of high-affinity binding, the lower energy difference for activation of the CAM than the wild-type receptor would prevail, resulting in a larger change in fluorescence for the CAM receptor. We

should mention than in the presence of G-proteins, which would stabilize the active state into a (^{A}LR*G) complex, the relative populations of active versus inactive states for wild-type and mutant receptors should be significantly different. In this case, the differences in the populations of the active state may well prevail over the energy difference required for activation, and then the CAM receptor would show a smaller change in fluorescence than the wild-type for agonists.

The increased instability observed for the CAM receptor (Gether 1996) requires a different explanation that pertains to the states other than the inactive or the G-protein coupled states mentioned above, available once the receptor has been "excited" to the active state. Thus, it is conceivable that another form of decay of the excited state R* is towards a partially denatured form that is likely to be irreversible. Such a transition is made likely by the special conformational and dynamic properties of the R* forms of a G-protein coupled receptor (GPCR). That constitutive activation is likely to be associated with a "release" of the receptor to a more dynamically free form has been proposed from various considerations (e.g., see (Rao 1994)). The dynamic properties of ligand-bound R* forms of GPCRs have also been explored computationally with molecular dynamics simulations of ligand-receptor complexes (e.g., see (Luo 1993; Zhang 1993)). Clear differences emerged between the properties of agonist-bound as compared to antagonist-bound complexes of the 5-HT2A serotonin receptor subtype. In general, the

agonist-bound state of the receptor showed a greatly enhanced flexibility of certain domains and significant structural distortion that loosened the tightly packed ground state conformation of the unliganded wild type receptor. These dynamic properties are acquired as a result of specific structural changes, that have been described for a particular case (Luo 1993). The dynamic properties of the energetically excited $^{A}LR^{*}$ form are likely to be similar for R^{*} in the absence of ligand, but perhaps exaggerated by the lack of restraint from the local interactions with the ligand in the binding site. The free R^{*} species thus become more likely to undergo a transition towards a loosely packed, denatured form of the protein that will not be readily reversible. The higher probability of reaching an R^{*} state in the CAM receptors compared to the wild type, makes the transition to such a denatured form more probable for the mutants than for the wild type.

Given the higher affinity of ligands with inverse agonist properties for the ground state R than for the R^{*} state of a receptor, the binding of such ligands should decrease the population of free R^{*} species in equilibrium with R, and thus stabilize the receptor against denaturation. The stabilizing effect of agonists will be smaller as it is due to the specific interactions that restrain the molecule and prevent the flexible R^{*} from achieving overly large fluctuation amplitudes, rather than to a shift of equilibrium to the stable form R.

3.3.2- Laser-induced optoacoustic studies of rhodopsin activation.

Another experiment that pertains to the thermodynamic scheme in Figure 3.2 is based on the use of Laser-induced optoacoustics to measure volume and changes after photoexcitation of bovine rhodopsin in the temperature range 5-32° C in its natural environment, i.e. washed membranes (Strassburger 1997). The validity of this technique has been corroborated by prior studies on other biological photoreceptors (Strassburger 1997) whose structure is better understood than rhodopsin. The authors reported that the enthalpy of the activated state of rhodopsin was 85 kJ/mol higher in energy than the basal state, in full accordance with the proposed energy levels proposed above. Rhodopsin activation involves several intermediates. In this study, the authors found a much bigger increase in enthalpy upon retinal isomerization (bathorhodopsin) that decayed to a blue-shifted intermediate (BSI), followed by the lumi state that couples to G-protein and thus should be compared with the activated form of other GPCRs. The enthalpy of the BSI state was indistinguishable from the lumi. Rhodopsin activation involved also a significant increase in volume, proposed to arise as a result of increased conformational flexibility, in agreement with our proposals described above based on the thermodynamic scheme.

It is noteworthy that the same experiments were performed on detergent solubilized rhodopsin, due to the previously observed structural

perturbations induced by rhodopsin solubilization (see Section 2.2), and the large number of other structural studies used in this thesis performed on rhodopsin under similar conditions. The enthalpy changes that accompany rhodopsin activation were drastically altered in the detergent-solubilized rhodopsin as compared to rhodopsin in its natural environment. The enthalpy difference between rhodopsin and BSI was now 50 kJ/mol, as opposed to 85 kJ/mol for rhodopsin, and the enthalpy of the lumi state was no longer indistinguishable from the BSI state but 40 kJ/mol lower. This indicates that the interaction between rhodopsin and its surrounding phospholipids modulates dramatically the energetics of the activation mechanism. I shall propose a molecular basis for the observed role of surrounding phospholipids in modulating the energetics of receptor activation (see Section 9).

4.- PRIMARY STRUCTURE OF GPCRS: CONSERVATION PATTERNS AND FUNCTIONAL DIVERGENCE

Derivation of structural inferences from analysis of the conservation pattern performed on a multiple sequence analysis (MSA) requires the assumption of a common fold or structural framework for the GPCRs contained in the homologous alignment. The selection of the GPCR subset for this purpose and the resulting MSA are described in Section 4.1. It should be noted that the 5HT2C and 5HT2A GPCRs are significantly divergent from the rest of the neurotransmitter GPCRs considered. Thus, the 5HT2 receptors are not included in the quantitative analysis of the conservation pattern, presented in Section 4.2. Their exclusion allows a very high homology score among the GPCR selected for this analysis, and thus a higher confidence in the structural inferences derived. In order to extract such information regarding the 5HT2C GPCR, I will identify AA positions in the 5HT2C receptor sequence where there are significant deviations from the pattern of conserved properties explored above (Section 4.3). These sites would become target sites for the study of structural deviations of the 5HT2C receptor from the common fold pursued.

The analysis of the conservation pattern performed in this section, and the derivation of structural inferences in the next two sections, was performed for the rat 5HT2CR, which was the only 5HT2CR sequence

available at the time. However, modeling the three-dimensional structure of the receptor (Section 7) was done using the human 5HT2C sequence. Because the differences between the human and rat receptor sequences, as well as between their pharmacological profiles, is not significant, and because the structural inferences derived here are based on conservation analysis and thus should be valid for all these neurotransmitter GPCRs, I have maintained the rat sequence for illustrative purposes throughout the presentation of the conservation analysis (Sections 4 to 6).

4.1- Construction of a multiple sequence alignment of neurotransmitter GPCRs: Alignment of the subset of GPCR for which a common structural framework with the 5HT2C receptor will be assumed.

I have aligned 55 sequences corresponding to serotonergic, adrenergic and dopaminergic GPCR following the guidelines described in our review chapter (Ballesteros 1995) (Section 2.1). From the sequences of these receptors available at the time (1993) for these three GPCR subgroups, some receptor sequences (e.g. the D4 dopamine) were excluded due to the presence of non-conserved Pro residues at positions predicted to affect significantly the structure (Ballesteros 1995) the receptor because they were not certain to obey the assumption of a common structural framework that is required to derive structural inferences from the MSA. Table 4.1 shows the alignment of the selected receptor sequences with the 3 types of GPCR in the above mentioned

order; the first of them is always the 5HT2CR. Regions far from the TMH domain are not included. Several features are indicated and will be described as they become relevant in subsequent sections.

Table 4.1 :

Sequence alignment of GPCR homologous from the 5HT2C (on top), arranged in the following order: Serotonergic (5HT 1C,2,1A,1D,1B), Adrenergic (A2,B1, B2,B3,A1), and Dopaminergic (D2, D3, D4, D5, D1). Receptor names represent their codes in the Genbank sequence database. Predicted TMHs are indicated by overlines. Highly conserved residues (CI < 0.3) are shown on top. β-branched AA are denoted "B", e.g. B=V,I,(T).

```
                    1         1         1         1         1         1
                    0         1         2         3         4         5
                    0         0         0         0         0         0
                                                                  GN  B    B
RATSR2CA        1 LLVWQFDISISPVAAIVTDTFNSSDGGRLFQFPDGVQNWPALSIVVIIIMTIGGNILVIMAVS
RATSR5HT2       1 NTSEASNWTIDAENRTNLSCEGYLPPTCLSILHLQEKNWSALLTTVVIILTIAGNILVIMAVS
HUM5HT1A        0 -MDVLSPGQGNNTTSPPAPFETGGNTTGISDVTVSYQVITSLLLGTLIFCAVLGNACVVAAIA
HSHTRB          0 -MDVLSPGQGNNTTSPPAPFETGGNTTGISDVTVSYQVITSLLLGTLIFCAVLGNACVVAAIA
HUMSER1DRA      0 SPLNQSAEGLPQEASNRSLNATETSEAWDPRTLQALKISLAVVLSVITLATVLSNAFVLTTIL
HUM5HT1DA       0 SPLNQSAEGLPQEASNRSLNATETSEAWDPRTLQALKISLAVVLSVITLATVLSNAFVLTTIL
RAT5HT1D        0 --MSLPNQSLEGLPQEASNRSLNATGAWDPEVLQALRISLVVVLSIITLATVLSNAFVLTTIL
HSHTRA          0 PAGSETWVPQANLSSAPSQNCSAKDYIYQDSISLPWKVLLVMLLALITLATTLSNAFVIATVY
HUM5HT1BSR      0 PAGSETWVPQANLSSAPSQNCSAKDYIYQDSISLPWKVLLVMLLALITLATTLSNAFVIATVY
HUMSER1DRB      0 PAGSETWVPQANLSSAPSQNCSAKDYIYQDSISLPWKVLLVMLLALITLATTLSNAFVIATVY
MM5HT1B         0 APPPPAASQTGVPLTNLSHNCSADGYIYQDSIALPWKVLLVALLALITLATTLSNAFVIATVY
RAT5HT1BR       0 APPPPATSQTGVPLANLSHNCSADDYIYQDSIALPWKVLLVALLALITLATTLSNAFVIATVY
DRO5HTR         0 TSSNLGDSNTTLVPLSDTPLLLEEFAAGEFVLPPLTSIFVSIVLLIVILGTVVGNVLVCIAVC
DRO5HT2AR       0 EMLANRSGHLDLINGTGGLNVTTSKVAEDDFTQLLRMAVTSVLLGLMILVTIIGNVFVIAAII
DRO5HT2BR       0 QIVWNRSVNGNGNSNTFDLVDDEQERAAVEFWLLVKMIAMAVVLGLMILVTIIGNVFVIAAII
HUMADRA         0 ----MGSLQPDAGNASWNGTEAPGGGARATPYSLQVTLTLVCLAGLLMLLTVFGNVLVIIAVF
HUMADRA2R       0 ----MGSLQPDAGNASWNGTEAPGGGARATPYSLQVTLTLVCLAGLLMLLTVFGNVLVIIAVF
PIGA2AR         0 ----MGSLQPEAGNASWNGTEAPGGGARATPYSLQVTLTLVCLAGLLMLFTVFGNVLVIIAVF
RATRG20         0 ----MGSLQPDAGNSSWNGTEAPGGGTRATPYSLQVTLTLVCLAGLLMLFTVFGNVLVIIAVF
RATRG10         0 AEGPNGSDAGEWGSGGGANASGTDWGPPPGQYSAGAVAGLAAVVGFLIVFTVVGNVLVVIAVL
HUMADRA2C       0 AAGPNASGAGERGSGGVANASGASWGPPRGQYSAGAVAGLAAVVGFLIVFTVVGNVLVVIAVL
HUMADRA2RA      0 ------------------------MDHQDPYSVQATAAIAAAITFLILFTIFGNALVILAVL
RATB1ARA        0 LPDGAATAARLLVLASPPASLLPPASEGSAPLSQQWTAGMGLLLALIVLLIVVGNVLVIVAIA
RATB1AR         0 LPDGAATAARLLVLASPPASLLPPASEGSAPLSQQWTAGMGLLLALIVLLIVVGNVLVIVAIA
HUMADRB1        0 LPDGAATAARLLVPASPPASLLPPASESPEPLSQQWTAGMGLLMALIVLLIVAGNVLVIVAIA
MUSB2ARG        0 ----MGPHGNDSDFLLAPNGSRAPHHDVTQERDEAWVVGMAILMSVIVLAIVFGNVLVITAIA
HUMBARR         0 ----MGQPGNGSAFLLAPNRSHAPDHDVTQQRDEVWVVGMGIVMSLIVLAIVFGNVLVITAIA
HUMADRBRA       0 ----MGQPGNGSAFLLAPNGSHAPDHDVTQQRDEVWVVGMGIVMSLIVLAIVFGNVLVITAIA
HUMADRBR        0 ----MGQPGNGSAFLLAPNRSHAPDHDVTQQRDEVWVVGMGIVMSLIVLAIVFGNVLVITAIA
HUMBAR          0 ----MGQPGNGSAFLLAPNRSHAPDHDVTQQRDEVWVVGMGIVMSLIVLAIVFGNVLVITAIA
RATB2AR         0 ----MEPHGNDSDFLLAPNGSRAPGHDITQERDEAWVVGMAILMSVIVLAIVFGNVLVITAIA
RATADBC         0 ----MEPHGNDSDFLLAPNGSRAPGHDITQERDEAWVVGMAILMSVIVLAIVFGNVLVITAIA
```

```
HAMARBR       0 ----MGPPGNDSDFLLTTNGSHVPDHDVTEERDEAWVVG-AILMSVIVLAIVFGNVLVITAIA
TKYARBR       0 WLPPDCGPHNRSGGGGATAAPTGSRQVSAELLSQQWEAGMSLLMALVVLLIVAGNVLVIAAIG
RATA2BR       0 --------------------MSGPTMDHQEPYSVQATAAIASAITFLILFTIFGNALVILAVL
HUMA1AADR     0 RTPTYRSTEMVQRLRMEAVQHSTSTAAVGGLVVSAQGVGVGVFLAAFILMAVAGNLLVILSVA
RATADRA1A     0 TGSGEDNQSSTGEPGAAASGEVNGSAAVGGLVVSAQGVGVGVFLAAFILTAVAGNLLVILSVA
RATADRA1B     0 HNTSAPAHWGELKDDNFTGPNQTSSNSTLPQLDVTRAISVGLVLGAFILFAIVGNILVILSVA
RATA1B        0 HNTSAPAHWGELKDDNFTGPNQTSSNSTLPQLDVTRAISVGLVLGAFILFAIVGNILVILSVA
BOVANDRE      0 -----------MVFLSGNASDSSNCTHPPPPVNISKAILLGVILGGLILFGVLGNILVILSVA
HUMDD2        0 ---MDPLNLSWYDDDLERQNWSRPFNGSDGKADRPHYNYYATLLTLLIAVIVFGNVLVCMAVS
HUMD2A        0 ---MDPLNLSWYDDDLERQNWSRPFNGSDGKADRPHYNYYATLLTLLIAVIVFGNVLVCMAVS
HUMDRD2A      0 ---MDPLNLSWYDDDLERQNWSRPFNGSDGKADRPHYNYYATLLTLLIAVIVFGNVLVCMAVS
RATD2REC      0 ---MDPLNLSWYDDDLERQNWSRPFNGSEGKADRPHYNYYAMLLTLLIFIIVFGNVLVCMAVS
MUSD2AR       0 ---MDPLNLSWYDDDLERQNWSRPFNGSEGKADRPHYNYYAMLLTLLIFIIVFGNVLVCMAVS
HUMDOPD2      0 TALMDPLNLSWYDDDLERQNWSRPFNGSDGKADRPHYNYYATLLTLLIAVIVFGNVLVCMAVS
RATD2RA       0 ---MDPLNLSWYDDDLERQNWSRPFNGSEGKADRPHYNYYAMLLTLLIFIIVFGNVLVCMAVS
RATDOPRD2     0 ---MDPLNLSWYDDDLERQNWSRPFNGSEGKADRPHYNYYAMLLTLLIFIIVFGNVLVCMAVS
XELD2DOPAR    0 ---MDPQNLSMYNDDINNGT------NGTAVDQKPHYNYYAMLLTLLVFVIVFGNVLVCIAVS
RATDD3        0 -------MAPLSQISTHLNSTCGAENSTGVNRARPHA-YYALSYCALILAIIFGNGLVCAAVL
HUMDRD5A      0 PGSNGTAYPGQFALYQQLAQGNAVGGSAGAPPLGPSQVVTACLLTLLIIWTLLGNVLVCAAIV
HUMHD5DR      0 PGSNGTAYPGQFALYQQLAQGNAVGGSAGAPPLGPSQVVTACLLTLLIIWTLLGNVLVCAAIV
RATD1BR       0 LPPGRNRTAQPARLGLQRQLAQVDAPAGSATPLGPAQVVTAGLLTLLIVWTLLGNVLVCAAIV
HUMD1DO       0 --------------MRTLNTSAMDGTGLVVERDFSVRILTACFLSLLILSTLLGNTLVCAAVI
RATD1DRA      0 KNLRGKSPEVCSFWKMAPNTSTMDEAGLPAERDFSFRILTACFLSLLILSTLLGNTLVCAAVI

                1       2        2        2          3        3
                6       4        5        6          2        3
                0       0        0        0          0        0

                ____    ______________________             ______________
                   L    N   SLA  D        B P          W    C-       DB  T S
RATSR1CA      1 MEKKLHN-ATNYFLMSLAIADMLVGLLVMPLSLLAILYDYVWPLPRYLCPVWISLDVLFSTAS
RATSR5HT2     1 LEKKLQN-ATNYFLMSLAIADMLLGFLVMPVSMLTILYGYRWPLPSKLCAIWIYLDVLFSTAS
HUM5HT1A      0 LERSLQN-VANYLIGSLAVTDLMVSVLVLPMAALYQV-LNKWTLGQVTCDLFIALDVLCCTSS
HSHTRB        0 LERSLQN-VANYLIGSLAVTDLMVSVLVLPMAALYQV-LNKWTLGQVTCDLFIALDVLCCTSS
HUMSER1DRA    0 LTRKLHT-PANYLIGSLATTDLLVSILVMPISIAYTI-THTWNFGQILCDIWLSSDITCCTAS
HUM5HT1DA     0 LTRKLHT-PANYLIGSLATTDLLVSILVMPISIAYTI-THTWNFGQILCDIWLSSDITCCTAS
RAT5HT1D      0 LTKKLHT-PANYLIGSLATTDLLVSILVMPISIAYTT-TRTWNFGQILCDIWVSSDITCCTAS
HSHTRA        0 RTRKLHT-PANYLIASLAVTDLLVSILVMPISTMYTV-TGRWTLGQVVCDFWLSSDITCCTAS
HUM5HT1BSR    0 RTRKLHT-PANYLIASLAVTDLLVSILVMPISTMYTV-TGRWTLGQVVCDFWLSSDITCCTAS
HUMSER1DRB    0 RTRKLHT-PANYLIASLAVTDLLVSILVMPISTMYTV-TGRWTLGQVVCDFWLSSDITCCTAS
MM5HT1B       0 RTRKLHT-PANYLIASLAVTDLLVSILVMPISTMYTV-TGRWTLGQVVCDFWLSSDITCCTAS
RAT5HT1BR     0 RTRKLHT-PANYLIASLAVTDLLVSILVMPISTMYTV-TGRWTLGQVVCDFWLSSDITCCTAS
DRO5HTR       0 MVRKLRR-PCNYLLVSLALSDLCVALLVMPMALLYEV-LEKWNFGPLLCDIWVSFDVLCCTAS
DRO5HT2AR     0 LERNLQN-VANYLVASLAVADLFVACLVMPLGAVYEI-SQGWILGPELCDIWTSCDVLCCTAS
DRO5HT2BR     0 LERNLQN-VANYLVASLAVADLFVACLVMPLGAVYEI-SNGWILGPELCDIWTSCDVLCCTAS
HUMADRA       0 TSRALKA-PQNLFLVSLASADILVATLVIPFSLANEV-MGYWYFGKTWCEIYLALDVLFCTSS
HUMADRA2R     0 TSRALKA-PQNLFLVSLASADILVATLVIPFSLANEV-MGYWYFGKAWCEIYLALDVLFCTSS
PIGA2AR       0 TSRALKA-PQNLFLVSLASADILVATLVIPFSLANEV-MGYWYFGKAWCEIYLALDVLFCTSS
RATRG20       0 TSRALKA-PQNLFLVSLASADILVATLVIPFSLANEV-MGYWYFGKVWCEIYLALDVLFCTSS
RATRG10       0 TSRALRA-PQNLFLVSLASADILVATLVMPFSLANEL-MAYWYFGQVWCGVYLALDVLFCTSS
HUMADRA2C     0 TSRALRA-PQNLFLVSLASADILVATLVMPFSLANEL-MAYWYFGQVWCGVYLALDVLFCTSS
HUMADRA2RA    0 TSRSLRA-PQNLFLVSLAAADILVATLIIPFSLANEL-LGYWYFRRTWCEVYLALDVLFCTSS
RATB1ARA      0 KTPRLQT-LTNLFIMSLASADLVMGLLVVPFGATIVV-WGRWEYGSFFCELWTSVDVLCVTAS
RATB1AR       0 KTPRLQT-LTNLFIMSLASADLVMGLLVVPFGATIVV-WGRWEYGSFFCELWTSVDVLCVTAS
HUMADRB1      0 KTPRLQT-LTNLFIMSLASADLVMGLLVVPFGATIVV-WGRWEYGSFFCELWTSVDVLCVTAS
MUSB2ARG      0 KFERLQT-VTNYFIISLACADLVMGLAVVPFGASHTS-MKMWNFGNFWCEFWTSIDVLCVTAS
HUMBARR       0 KFERLQT-VTNYFITSLACADLVMGLAVVPFGAAHIL-MKMWTFGNFWCEFWTSIDVLCVTAS
```

```
HUMADRBRA     0 KFERLQT-VTNYFITSLACADLVMGLAVVPFGAAHIL-MKMWTFGNFWCEFWTSIDVLCVTAS
HUMADRBR      0 KFERLQT-VTNYFITSLACADLVMGLAVVPFGAAHIL-MKMWTFGNFWCEFWTSIDVLCVTAS
HUMBAR        0 KFERLQT-VTNYFITSLACADLVMGLAVVPFGAAHIL-MKMWTFGNFWCEFWTSIDVLCVTAS
RATB2AR       0 KFERLQT-VTNYFITSLACADLVMGLAVVPFGASHIL-MKMWNFGNFWCEFWTSIDVLCVTAS
RATADBC       0 KFERLQT-VTNYFITSLACADLVMGLAVVPFGASHIL-MKMWNFGNFWCEFWTSIDVLCVTAS
HAMARBR       0 KFERLQT-VTNYFITSLACADLVMGLAVVPFGASHIL-MKMWNFGNFWCEFWTSIDVLCVTAS
TKYARBR       0 RTQRLQT-LTNLFITSLACADLVMGLLVVPFGATLVV-RGTWLWGSFLCECWTSLDVLCVTAS
RATA2BR       0 TSRSLRA-PQNLFLVSLAAADILVATLIIPFSLANEL-LGYWYFWRAWCEVYLALDVLFCTSS
HUMA1AADR     0 CNRHLQT-VTNYFIVNLAVADLLLSATVLPFSATMEV-LGFWAFGRAFCDVWAAVDVLCCTAS
RATADRA1A     0 CNRHLQT-VTNYFIVNLAVADLLLSAAVLPFSATMEV-LGFWAFGRTFCDVWAAVDVLCCTAS
RATADRA1B     0 CNRHLRT-PTNYFIVNLAIADLLLSFTVLPFSATLEV-LGYWVLLSFFCDIWAAVDVLCCTAS
RATA1B        0 CNRHLRT-PTNYFIVNLAIADLLLSFTVLPFSATLEV-LGYWVLGRIFCDIWAAVDVLCCTAS
BOVANDRE      0 CHRHLHS-VTHYYIVNLAVADLLLTSTVLPFSAIFEI-LGYWAFGRVFCNVWAAVDVLCCTAS
HUMDD2        0 REKALQT-TTNYLIVSLAVADLLVATLVMPWVVYLEV-VGEWKFSRIHCDIFVTLDVMMCTAS
HUMD2A        0 REKALQT-TTNYLIVSLAVADLLVATLVMPWVVYLEV-VGEWKFSRIHCDIFVTLDVMMCTAS
HUMDRD2A      0 REKALQT-TTNYLIVSLAVADLLVATLVMPWVVYLEV-VGEWKFSRIHCDIFVTLDVMMCTAS
RATD2REC      0 REKALQT-TTNYLIVSLAVADLLVATLVMPWVVYLEV-VGEWKFSRIHCDIFVTLDVMMCTAS
MUSD2AR       0 REKALQT-TTNYLIVSLAVADLLVATLVMPWVVYLEV-VGEWKFSRIHCDIFVTLDVMMCTAS
HUMDOPD2      0 REKALQT-TTNYLIVSLAVADLLVATLVMPWVVYLEV-VGEWKFSRIHCDIFVTLDVMMCTAS
RATD2RA       0 REKALQT-TTNYLIVSLAVADLLVATLVMPWVVYLEV-VGEWKFSRIHCDIFVTLDVMMCTAS
RATDOPRD2     0 REKALQT-TTNYLIVSLAVADLLVATLVMPWVVYLEV-VGEWKFSRIHCDIFVTLDVMMCTAS
XELD2DOPAR    0 REKALQT-TTNYLIVSLAVADLLVATLVMPWAVYMEV-VGEWRFSRIHCDIFVTLDVMMCTAS
RATDD3        0 RERALQT-TTNYLVVSLAVADLLVATLVMPWVVYLEVTGGVWNFSRICCDVFVTLDVMMCTAS
HUMDRD5A      0 RSRHLRANMTNVFIVSLAVSDLFVALLVMPWKAVAEV-AGYWPFGA-FCDVWVAFDIMCSTAS
HUMHD5DR      0 RSRHLRANMTNVFIVSLAVSDLFVALLVMPWKAVAEV-AGYWPFGA-FCDVWVAFDIMCSTAS
RATD1BR       0 RSRHLRAKMTNIFIVSLAVSDLFVALLVMPWKAVAEV-AGYWPFGT-FCDIWVAFDIMCSTAS
HUMD1DO       0 RFRHLRSKVTNFFVISLAVSDLLVAVLVMPWKAVAEI-AGFWPFGS-FCNIWVAFDIMCSTAS
RATD1DRA      0 RFRHLRSKVTNFFVISLAVSDLLVAVLVMPWKAVAEI-AGFWPLGP-FCNIWVAFDIMCSTAS

                3         3                   4         4         4         4
                4         5                   4         5         6         7
                0         0                   0         0         0         0
                ---------------               ---------------------
                I  LC I  DRY  B     Y         T       B   W     BS

RATSR1CA      1 IMHLCAISLDRYVAIRNPIEHSRFN---SRTKAIMKIAIVWAISIGVSVPIPVIGLR------
RATSR5HT2     1 IMHLCAISLDRYVAIQNPIHHSRFN---SRTKAFLKIIAVWTISVGISMPIPVFGLQ------
HUM5HT1A      0 ILHLCAIALDRYWAITDPIDYVNKR---TPRP-RALISLTWLIGFLISIPPILGWRTP-----
HSHTRB        0 ILHLCAIALDRYWAITDPIDYVNKR---TPRRAAALISLTWLIGFLISIPPMLGWRTP-----
HUMSER1DRA    0 ILHLCVIALDRYWAITDALEYSKRR---TAGHAATMIAIVWAISICISIPP-LFWRQAK----
HUM5HT1DA     0 ILHLCVIALDRYWAITDALEYSKRR---TAGHAATMIAIVWAISICISIPP-LFWRQAK----
RAT5HT1D      0 ILHLCVIALDRYWAITDALEYSKRR---TAGHAAAMIAAVWAISICISIPP-LFWRQAT----
HSHTRA        0 ILHLCVIALDRYWAITDAVEYSAKR---TPKRAAVMIALVWVFSISISLPP-FFWRQAK----
HUM5HT1BSP    0 ILHLCVIALDRYWAITDAVEYSAKR---TPKRAAVMIALVWVFSISISLPP-FFWRQAK----
HUMSER1DRB    0 ILHLCVIALDRYWAITDAVEYSAKR---TPKRAAVMIALVWVFSISISLPP-FFWRQAK----
MM5HT1B       0 IMHLCVIALDRYWAITDAVEYSAKR---TPKRAAIMIVLVWVFSISISLPP-FFWRQAK----
RAT5HT1BR     0 IMHLCVIALDRYWAITDAVDYSAKR---TPKRAAIMIVLVWVFSISISLPP-FFWRQAK----
DRO5HTR       0 ILNLCAISVDRYLAITKPLEYGVKR---TPRRMMLCVGIVWLAAACISLPPLLILGNE-----
DRO5HT2AR     0 ILHLVAIAVDRYWAVTN-IDYIHSR---TSNRVFMMIFCVWTAAVIVSLAP-QFGWKDP----
DRO5HT2BR     0 ILHLVAIAADRYWTVTN-IDYNNLR---TPRRVFLMIFCVWFAALIVSLAP-QFGWKDP----
HUMADRA       0 IVHLCAISLDRYWSITQAIEYNLKR---TPRRIKAIIITCWVISAVISFPPLISIEKK-----
HUMADRA2R     0 IVHLCAISLDRYWSITQAIEYNLKR---TPRRIKAIIITVWVISAVISFPPLISIEKK-----
PIGA2AR       0 IVHLCAISLDRYWSITQAIEYNLKR---TPRRIKAIIVTVWVISAVISFPPLISIEKK-----
RATRG20       0 IVHLCAISLDRYWSITQAIEYNLKR---TPRRIKAIIVTVWVISAVISFPPLISIEKK-----
RATRG10       0 IVHLCAISLDRYWSVTQAVEYNLKR---TPRRVKATIVAVWLISAVISFPPLVSFYRR-----
HUMADRA2C     0 IVHLCAISLDRYWSVTQAVEYNLKR---TPRRVKATIVAVWLISAVISFPPLVSLYRQ-----
HUMADRA2RA    0 IVHLCAISLDRYWAVSRALEYNSKR---TPRRIKCIILTVWLIAAVISLPPLIYKGDQ-----
```

```
RATB1ARA      0 IETLCVIALDRYLAITSPFRYQSLL---TRARARALVCTVWAISALVSFLPILMHWWR-----
RATB1AR       0 IETLCVIALDRYLAITLPFRYQSLL---TRARARALVCTVWAISALVSFLPILMHWWR-----
HUMADRB1      0 IETLCVIALDRYLAITSPFRYQSLL---TRARARGLVCTVWAISALVSFLPILMHWWR-----
MUSB2ARG      0 IETLCVIAVDRYVAITSPFKYQSLL---TKNKARVVILMVWIVSGLTSFLPIQMHWYRATHK-
HUMBARR       0 IETLCVIAVDRYFAITSPFKSQSLL---TKNKARVIILMVWIVSGLTSFLPIQMHWYRATHQ-
HUMADRBRA     0 IETLCVIAVDRYFAITSPFKYQSLL---TKNKARVIILMVWIVSGLTSFLPIQMHWYRATHQ-
HUMADRBR      0 IETLCVIAVDRYFAITSPFKYQSLL---TKNKARVIILMVWIVSGLTSFLPIQMHWYRATHQ-
HUMBAR        0 IETLCVIAVDRYFAITSPFKYQSLL---TKNKARVIILMVWIVSGLTSFLPIQMHWYRATHQ-
RATB2AR       0 IETLCVIAVDRYVAITSPFKYQSLL---TKNKARVVILMVWIVSGLTSFLPIQMHWYRATHK-
RATADBC       0 IETLCVIAVDRYVAITSPFKYQSLL---TKNKARVVILMVWIVSGLTSFLPIQMHWYRATHK-
HAMARBR       0 IETLCVIAVDRYIAITSPFKYQSLL---TKNKARMVILMVWIVSGLTSFLPIQMHWYRATHQ-
TKYARBR       0 IETLCVIAIDRYLAITSPFRYQSLM---TRARAKVIICTVWAISALVSFLPIMMHWWRDEDP-
RATA2BR       0 IVHLCAISLDRYWAVSRALEYNSKR---TPCRIKCIILTVWLIAAVISLPPLIYKGDQRPDA-
HUMA1AADR     0 ILSLCTISVDRYVGVRHSLKYPAIM---TERKAAAILALLWVVALVVSVGPLLGWKEPVPPD-
RATADRA1A     0 ILSLCTISVDRYVGVRHSLKYPAIM---TERKAAAILALLWAVALVVSVGPLLGWKEPVPPD-
RATADRA1B     0 ILSLCAISIDRYIGVRYSLQYPTLV---TRRKAILALLSVWVLSTVISIGPLLGWKEPAPND-
RATA1B        0 ILSLCAISIDRYIGVRYSLQYPTLV---TRRKAILALLSVWVLSTVISIGPLLGWKEPAPND-
BOVANDRE      0 IMGLCIISIDRYIGVSYPLRYPTIV---TQKRGLMALLCVWALSLVISIGPLFGWRQPAPED-
HUMDD2        0 ILNLCAISIDRYTAVAMPMLYNTRYS--SKRRVTVMISIVWVLSFTISCPLLFGLNN------
HUMD2A        0 ILNLCAISIDRYTAVAMPMLYNTRYS--SKRRVTVMISIVWVLSFTISCPLLFGLNN------
HUMDRD2A      0 ILNLCAISIDRYTAVAMPMLYNTRYS--SKRRVTVMISIVWVLSFTISCPLLFGLNN------
RATD2REC      0 ILNLCAISIDRYTAVAMPMLYNTRYS--SKRRVTVMIAIVWVLSFTISCPLLFGLNN------
MUSD2AR       0 ILNLCAISIDRYTAVAMPMLYNTRYS--SKRRVTVMIAIVWVLSFTISCPLLFGLNN------
HUMDOPD2      0 ILNLCAISIDRYTAVAMPMLYNTRYS--SKRRVTVMISIVWVLSFTISCPLLFGLNN------
RATD2RA       0 ILNLCAISIDRYTAVAMPMLYNTRYS--SKRRVTVMIAIVWVLSFTISCPLLFGLNN------
RATDOPRD2     0 ILNLCAISIDRYTAVAMPMLYNTRYS--SKRRVTVMIAIVWVLSFTISCPLLFGLNN------
XELD2DOPAR    0 ILNLCAISIDRYTAVAMPMLYNTRYS--SKRRVTVMISVVWVLSFAISCPLLFGLNN------
RATDD3        0 ILNLCAISIDRYTAVVMPVHYQHGTGQSSCRRVALMITAVWVLAFAVSCPLLFGFNT------
HUMDRD5A      0 ILNLCVISVDRYWAISRPFRYKRKM---TQRMALVMVGLAWTLSILISFIPVQLNWHRDQAAS
HUMHD5DR      0 ILNLCVISVDRYWAISRPFRYKRKM---TQRMALVMVGLAWTLSILISFIPVQLNWHRDQAAS
RATD1BR       0 ILNLCIISVDRYWAISRPFRYERKM---TQRVALVMVGLAWTLSILISFIPVQLNWHRDKAGS
HUMD1DO       0 ILNLCVISVDRYWAISSPFRYERKM---TPKAAFILISVAWTLSVLISFIPVQLSWH------
RATD1DRA      0 ILNLCVISVDRYWAISSPFQYERKM---TPKAAFILISVAWTLSVLISFIPVQLSWH------

                  5         5         5         5         5         5         5
                  0         1         2         3         4         5         6
                  0         0         0         0         0         0         0

                                                C                   ST   F  P        Y +
RATSR1CA      1 --------------------DESKVFVNNTTCVL--NDPNFVLIGSFVAFFIPLTIMVITYFL
RATSR5HT2     1 --------------------DDSKVFK-EGSCLL--ADDNFVLIGSFVAFFIPLTIMVITYFL
HUM5HT1A      0 -----------------------EDRSDPDACTISKDH-GYTIYSTFGAFYIPLLLMLVLYGR
HSHTRB        0 -----------------------EDRSDPDACTISKDH-GYTIYSTFGAFYIPLLLMLVLYGR
HUMSER1DRA    0 ------------------------AQEEMSDCLVNTSQISYTIYSTCGAFYIPSVLLIILYGR
HUM5HT1DA     0 ------------------------AQEEMSDCLVNTSQISYTIYSTCGAFYIPSVLLIILYGR
RAT5HT1D      0 ------------------------AHEEMSDCLVNTSQISYTIYSTCGAFYIPSILLIILYGR
HSHTRA        0 ------------------------AEEEVSECVVNTDHILYTVYSTVGAFYFPTLLLIALYGR
HUM5HT1BSR    0 ------------------------AEEEVSECVVNTDHILYTVYSTVGAFYFPTLLLIALYGR
HUMSER1DRB    0 ------------------------AEEEVSECVVNTDHILYTVYSTVGAFYFPTLLLIALYGR
MM5HT1B       0 ------------------------AEEEMLDCFVNTDHVLYTVYSTVGAFYLPTLLLIALYGR
RAT5HT1BR     0 ------------------------AEEEVLDCFVNTDHVLYTVYSTVGAFYLPTLLLIALYGR
DRO5HTR       0 ----------------------HEDEEGQPICTVC-QNFAYQIYATLGSFYIPLSVMLFVYYQ
DRO5HT2AR     0 ---------------------DYLQRIEQQKCMVS-QDVSYQVFATCCTFYVPLMVILALYWK
DRO5HT2BR     0 ---------------------DYMKRIEEQHCMVS-QDVGYQIFATCCTFYVPLLVILFLYWK
HUMADRA       0 -------------------GGGGGPQPAEPRCEIN-DQKWYVISSCIGSFFAPCLIMILVYVR
HUMADRA2R     0 -------------------GGGGGPQPAEPRCEIN-DQKWYVISSCIGSFFAPCLIMILVYVR
```

```
PIGA2AR      0 ------------------AGGGGQQPAEPRCEIN-DQKWYVISSCIGSFFAPCLIMILVYVR
RATRG20      0 ------------------GAGGGQQPAEPSCKIN-DQKWYVISSSIGSFFAPCLIMILVYVR
RATRG10      0 ----------------------PDGAAYPQCGLN-DETWYILSSCIGSFFAPCLIMGLVYAR
HUMADRA2C    0 ----------------------PDGAAYPQCGLN-DETWYILSSCIGSFFAPCLIMGLVYAR
HUMADRA2RA   0 ---------------------GPQPRGRPQCKLN-QEAWYILASSIGSFFAPCLIMILVYLR
RATB1ARA     0 ----------------AESDEARRCYNDPKCCDFV-TNRAYAIASSVVSFYVPLCIMAFVYLR
RATB1AR      0 ----------------AESDEARRCYNDPKCCDFV-TNRAYAIASSVVSFYVPLCIMAFVYLR
HUMADRB1     0 ----------------AESDEARRCYNDPKCCDFV-TNRAYAIASSVVSFYVPLCIMAFVYLR
MUSB2ARG     0 --------------------KAIDCYTEETCCDFF-TNQAYAIASSIVSFYVPLVVMVFVYSR
HUMBARR      0 --------------------EAINCYANETCCDFF-TNQAYAIASSIVSFYVPLVIMVFVYSR
HUMADRBRA    0 --------------------EAINCYANETCCDFF-TNQAYAIASSIVSFYVPLVIMVFVYSR
HUMADRBR     0 --------------------EAINCYANETCCDFF-TNQAYAIASSIVSFYVPLVIMVFVYSR
HUMBAR       0 --------------------EAINCYANETCCDFF-TNQAYAIASSIVSFYVPLVIMVFVYSR
RATB2AR      0 --------------------QAIDCYAKETCCDFF-TNQAYAIASSIVSFYVPLVVMVFVYSR
RATADBC      0 --------------------QAIDCYAKETCCDFF-TNQAYAIASSIVSFYVPLVVMVFVYSR
HAMARBR      0 --------------------KAIDCYHKETCCDFF-TNQAYAIASSIVSFYVPLVVMVFVYSR
TKYARBR      0 --------------------QALKCYQDPGCCDFV-TNRAYAIASSIISFYIPLLIMIFVYLR
RATA2BR      0 -------------------------RGLPQCELN-QEAWYILASSIGSFFAPCLIMILVYLR
HUMA1AADR    0 ------------------------ERFCGIT-EEAGYAVFSSVCSFYLPMAVIVVMYCR
RATADRA1A    0 ------------------------ERFCGIT-EEVGYAIFSSVCSFYLPMAVIVVMYCR
RATADRA1B    0 ------------------------DKECGVT-EEPFYALFSSLGSFYIPLAVILVMYCR
RATA1B       0 ------------------------DKECGVT-EEPFCALFCSLGSFYIPLAVILVMYCR
BOVANDRE     0 ------------------------ETICQIN-EEPGYVLFSALGSFYVPLTIILVMYCR
HUMDD2       0 -----------------------ADQNECII--ANPAFVVYSSIVSFYVPFIVTLLVYIK
HUMD2A       0 -----------------------ADQNECII--ANPAFVVYSSIVSFYVPFIVTLLVYIK
HUMDRD2A     0 -----------------------ADQNECII--ANPAFVVYSSIVSFYVPFIVTLLVYIK
RATD2REC     0 -----------------------TDQNECII--ANPAFVVYSSIVSFYVPFIVTLLVYIK
MUSD2AR      0 -----------------------TDQNECII--ANPAFVVYSSIVSFYVPFIVTLLVYIK
HUMDOPD2     0 -----------------------ADQNECII--ANPAFVVYSSIVSFYVPFIVTLLVYIK
RATD2RA      0 -----------------------TDQNECII--ANPAFVVYSSIVSFYVPFIVTLLVYIK
RATDOPRD2    0 -----------------------TDQNECII--ANPAFVVYSSIVSFYVPFIVTLLVYIK
XELD2DOPAR   0 -----------------------TGSKVCII--DNPAFVIYSSIVSFYVPFIVTLLVYVQ
RATDD3       0 -----------------------TGDPSICSI--SNPDFVIYSSVVSFYVPFGVTVLVYAR
HUMDRD5A     0 --WGGLDLPNNLANWTPWEEDFWEPDVNAENCDSS-LNRTYAISSSLISFYIPVAIMIVTYTR
HUMHD5DR     0 --WGGLDLPNNLANWTPWEEDFWEPDVNAENCDSS-LNRTYAISSSLISFYIPVAIMIVTYTR
RATD1BR      0 --QGQEGLL---SNGTPWEE-GWELEGRTENCDSS-LNRTYAISSSLISFYIPVAIMIVTYTR
HUMD1DO      0 ---------KAKPTSPSDGNATSLAET-IDNCDSS-LSRTYAISSSVISFYIPVAIMIVTYTR
RATD1DRA     0 ---------KAKPTWPLDGNFTSLEDTEDDNCDTR-LSRTYAISSSLISFYIPVAIMIVTYTS

                       5         5         5
                       7         8         9
                       0         0         0
                      ____
                       B
RATSR1CA     1 TIYVLRRQTLMLLRGHTEEELANMSLNFLNCC------------------------------
RATSR5HT2    1 TIKSLQKEATLCVSDLSTRAKLASFSFLPQSS------------------------------
HUM5HT1A     0 IFRAARFRIRKTVKKVEKTGADTRHGASP--------------------------------
HSHTRB       0 IFRAARFRIRKTVKKVEKTGADTRHGASP--------------------------------
HUMSER1DRA   0 IYRAARNRILN-PPSLYGKRFTTAHLITGS--------------------------------
HUM5HT1DA    0 IYRAARNRILN-PPSLYGKRFTTAHLITGS--------------------------------
RAT5HT1D     0 IYVAARSRILN-PPSLYGKRFTTAQLITGS--------------------------------
HSHTRA       0 IYVEARSRILKQTPNRTGKRLTRAQLITDS--------------------------------
HUM5HT1BSR   0 IYVEARSRILKQTPNRTGKRLTRAQLITDS--------------------------------
HUMSER1DRB   0 IYVEARSRILKQTPNRTGKRLTRAQLITDS--------------------------------
MM5HT1B      0 IYVEARSRILKQTPNKTGKRLTRAQLITDS--------------------------------
RAT5HT1BR    0 IYVEARSRILKQTPNKTGKRLTRAQLITDS--------------------------------
```

```
HSHTRA          0 VNQVKVRVSDALLEKKKLMAARERKATKTLGIILGAFIVCWLPFFIISLVMPICKDAC-----
HUM5HT1BSR      0 VNQVKVRVSDALLEKKKLMAARERKATKTLGIILGAFIVCWLPFFIISLVMPICKDAC-----
HUMSER1DRB      0 VNQVKVRVSDALLEKKKLMAARERKATKTLGIILGAFIVCWLPFFIISLVMPICKDAC-----
MM5HT1B         0 VNQVKVRVSDALLEKKKLMAARERKATKTLGIILGAFIVCWLPFFIISLVMPICKDAC-----
RAT5HT1BR       0 VNQVKVRVSDALLEKKKLMAARERKATKTLGIILGAFIVCWLPFFIISLVMPICKDAC-----
DRO5HTR         0 VSGSTGLLGSPHHKKLRFQLAKEKKASTTLGIIMSAFTVCWLPFFILALIRPFETMH------
DRO5HT2AR       0 LSSIANPMQKVNKRKETLEAKRERKAAKTLAIITGAFVVCWLPFFVMALTMPLCAACQI----
DRO5HT2BR       0 LASIANPHQKLAKRRQLLEAKRERKAAQTLAIITGAFVICWLPFFVMALTMSLCKECEI----
HUMADRA         0 QGRGRSASGLPRRRAGAGGQNLEKRFTFVLAVVIGVFVVCWFPFFFTYTLTAVGCSVP-----
HUMADRA2R       0 QGRGRSASGLPRRRAGAGGQNREKRFTFVLAVVIGVFVVCWFPFFFTYTLTAVGCSVP-----
PIGA2AR         0 TGAGEERGGVAKASRWRGRQNREKRFTFVLAVVIGVFVVCWFPFFFTYTLTAVGCSVP-----
RATRG20         0 SGQGEERAGGAKASRWRGRQNREKRFTFVLAVVIGVFVVCWFPFFFTYTLIAVGCP-------
RATRG10         0 FFLSRRRRARSSVCRRKVAQAREKRFTFVLAVVMGVFVLCWFPFFFSYSLYGICREACQLP--
HUMADRA2C       0 FFLSRRRRARSSVCRRKVAQAREKRFTFVLAVVMGVFVLCWFPFFFIYSLYGICREACQVP--
HUMADRA2RA      0 LLGRGVGAIGGQWWRRRAHVTREKRFTFVLAVVIGVFVLCWFPFFFSYSLGAICPKHCKVP--
RATB1ARA        0 PADSLANGRSSKRRPSRLVALREQKALKTLGIIMGVFTLCWLPFFLANVVKAFHRDLVP----
RATB1AR         0 PADSLANGRSSKRRPSRLVALREQKALKTLGIIMGVFTLCWLPFFLANVVKAFHRDLVP----
HUMADRB1        0 ATAPLANGRAGKRRPSRLVALREQKALKTLGIIMGVFTLCWLPFFLANVVKAFHRELVP----
MUSB2ARG        0 SQVEQDGRTGHGLRRSSKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIRDNLIP----
HUMBARR         0 SQVEQDGRTGHGLRRSSNFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIR----
HUMADRBRA       0 SQVEQDGRTGHGLRRSSKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIR----
HUMADRBR        0 SQVEQDGRTGHGLRRSSKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIR----
HUMBAR          0 SQVEQDGRTGHGLRRSSKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIR----
RATB2AR         0 SQVEQDGRSGHGLRSSSKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIRANLIP----
RATADBC         0 SQVEQDGRSGHGLRSSSKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIRANLIP----
HAMARBR         0 GQVEQDGRSGHGLRRSSKFCLKEHKALKTLGIIMGTFTLCWLPFFIVNIVHVIQDNLIP----
TKYARBR         0 HQPILGNGRASKRKTSRVMAMREHKALKTLGIIMGVFTLCWLPFFLVNIVNVFNRDLVP----
RATA2BR         0 LLGKNVGVASGQWWRRRTQLSREKRFTFVLAVVIGVFVVCWFPFFFSYSLGAICPQHCK----
HUMA1AADR       0 MRSAKGHTFRSSLSVRLLKFSREKKAAKTLAIVVGVFVLCWFPFFFVLPLGSLFPQLKP----
RATADRA1A       0 TQSSKGHTLRSSLSVRLLKFSREKKAAKTLAIVVGVFVLCWFPFFFVLPLGSLFPQLKP----
RATADRA1B       0 STKAKGHNPRSSIAVKLFKFSREKKAAKTLGIVVGMFILCWLPFFIALPLGSLFSTLKP----
RATA1B          0 STKAKGHNPRSSIAVKLFKFSREKKAAKTLGIVVGMFILCWLPFFIALPLGSLFSTLKP----
BOVANDRE        0 GSGVTSAKNKTHFSVRLLKFSREKKAAKTLGIVVGCFVLCWLPFFLVMPIGSFFPDFRP----
HUMDD2          0 MPNGKTRTSLKTMSRRKLSQQKEKKATQMLAIVLGVFIICWLPFFITHILNIHCDCNIP----
HUMD2A          0 MPNGKTRTSLKTMSRRKLSQQKEKKATQMLAIVLGVFIICWLPFFITHILNIHCDCNIP----
HUMDRD2A        0 MPNGKTRTSLKTMSRRKLSQQKEKKATQMLAIVLGVFIICWLPFFITHILNIHCDCNIP----
RATD2REC        0 MPNGKTRTSLKTMSRRKLSQQKEKKATQMLAIVLGVFIICWLPFFITHILNIHCDCNIP----
MUSD2AR         0 MPNGKTRTSLKTMSRRKLSQQKEKKATQMLAIVLGVFIICWLPFFITHILNIHCDCNIP----
HUMDOPD2        0 MPNGKTRTSLKTMSRRKLSQQKEKKATQMLAIVLGVFIICWLPFFITHILNIHCDCNIP----
RATD2RA         0 MPNGKTRTSLKTMSRRKLSQQKEKKATQMLAIVLGVFIICWLPFFITHILNIHCDCNIP----
RATDOPRD2       0 MPNGKTRTSLKTMSRRKLSQQKEKKATQMLAIVLGVFIICWLPFFITHILNIHCDCNIP----
XELD2DOPAR      0 MPNGKTRTSIKTMSKKKLSQHKEKKATQMLAIVLGVFIICWLPFFIIHILNMHCNCNIP----
RATDD3          0 SNGRLSTSLRLGPLQPRGVPLREKKATQMVVIVLGAFIVCWLPFFLTHVLNTHCQACHVS---
HUMDRD5A        0 ----RSSAACAPDTSLRASIKKETKVLKTLSVIMGVFVCCWLPFFILNCMVPFCSGH-PEGPP
HUMHD5DR        0 ----RSSAACAPDTSLRASIKKETKVLKTLSVIMGVFVCCWLPFFILNCMVPFCSGH-PEGPP
RATD1BR         0 ----RSRGAYEPDPSLRASIKKETKVFKTLSMIMGVFVCCWLPFFILNCMVPFCSSGDAEGPK
HUMD1DO         0 TGNGKPVECSQPESSFKMSFKRETKVLKTLSVIMGVFVCCWLPFFILNCILPFCGSG-----E
RATD1DRA        0 AGNGNPVECAQSESSFKMSFKRETKVLKTLSVIMGVFVCCWLPFFISNCMVPFCGSE-----E

                    7       7       7       7       7       7
                    3       4       5       6       7       8
                    0       0       0       0       0       0
                          ------------------------------------
                          W GY NS  NP IY     -F+  F

RATSR1CA        1 --NQKLMEKLLNVFVWIGYVCSGINPLVYTLFNKIYRRAFSKYLRCDYKPDKKPPVRQIPRVA
RATSR5HT2       1 --NENVIGALLNVFVWIGYLSSAVNPLVYTLFNKTYRSAFSRYIQCQYKENRKPLQLILVNTI
```

```
HUM5HT1A      0 ----HMPTLLGAIINWLGYSNSLLNPVIYAYFNKDFQNAFKKIIKCNFCRQ
HSHTRB        0 ----HMPTLLGAIINWLGYSNSLLNPVIYAYFNKDFQNAFKKIIKCKFCRQ
HUMSER1DRA    0 ----WIHPALFDFFTWLGYLNSLINPIIYTVFNEEFRQAFQKIVPFRKAS
HUM5HT1DA     0 ----WIHPALFDFFTWLGYLNSLINPIIYTVFNEEFRQAFQKIVPFRKAS
RAT5HT1D      0 ----WIHPALFDFFTWLGYLNSLINPVIYTVFNEDFRQAFQRVVHFRKAS
HSHTRA        0 ----WFHLAIFDFFTWLGYLNSLINPIIYTMSNEDFKQAFHKLIRFKCTS
HUM5HT1BSR    0 ----WFHLAIFDFFTWLGYLNSLINPIIYTMSNEDFKQAFHKLIRFKCTS
HUMSER1DRB    0 ----WFHLAIFDFFTWLGYLNSLINPIIYTMSNEDFKQAFHKLIRFKCTS
MM5HT1B       0 ----WFHMAIFDFFNWLGYLNSLINPIIYTMSNEDFKQAFHKLIRFKCAG
RAT5HT1BR     0 ----WFHMAIFDFFNWLGYLNSLINPIIYTMSNEDFKQAFHKLIRFKCTG
DRO5HTR       0 -----VPASLSSLFLWLGYANSLLNPIIYATLNRDFRKPFQEILYFRCSSLNTMMRENYYQDQ
DRO5HT2AR     0 ------SDSVASLFLWLGYFNSTLNPVIYTIFSPEFRQAFKRILFGGHRPVHYRSGKL
DRO5HT2BR     0 ------HTAVASLFLWLGYFNSTLNPVIYTIFNPEFRRAFKRILFGRKAAARARSAKI
HUMADRA       0 -------RTLFKFFFWFGYCNSSLNPVIYTIFNHDFRRAFKKIL-CRGDRKRIV
HUMADRA2R     0 -------RTLFKFFFWFGYCNSSLNPVIYTIFNHDFRRAFKKIL-CRGDRKRIV
PIGA2AR       0 -------PTLFKFFFWFGYCNSSLNPVIYTIFNHDFRRAFKKIL-CRGDRKRIV
RATRG20       0 -----VPYQLFNFFFWFGYCNSSLNPVIYTIFNHDFRRAFKKIL-CRGDRKRIV
RATRG10       0 -------EPLFKFFFWIGYCNSSLNPVIYTVFNQDFRRSFKHILFRRRRRGFRQ
HUMADRA2C     0 -------GPLFKFFFWIGYCNSSLNPVIYTVFNQDFRPSFKHILFRRRRRGFRQ
HUMADRA2RA    0 -------HGLFQFFFWIGYCNSSLNPVIYTIFNQDFRRAFRRIL-CRPWTQTAW
RATB1ARA      0 -------DRLFVFFNWLGYANSAFNPIIYCRS-PDFRKAFQRLL-CCARRAACRRRAAHGDRP
RATB1AR       0 -------DRLFVFFNWLGYANSAFNPIIYCRS-PDFRKAFQRLL-CCARRAACRRRAAHGDRP
HUMADRB1      0 -------DRLFVFFNWLGYANSAFNPIIYCRS-PDFRKAFQGLL-CCARRAARRRHATHGDRP
MUSB2ARG      0 -------KEVYILLNWLGYVNSAFNPLIYCRS-PDFRIAFQELL-CLRRSSFETYGNGYSSNS
HUMBARR       0 -------KEVYILLNWIGYVNSGFNPLIYCRS-PDFRIAFQELL-CLRRSSLKAYGNGYSSNG
HUMADRBRA     0 -------KEVYILLNWIGYVNSGFNPLIYCRS-PDFRIAFQELL-CLRRSSLKAYGNGYSSNG
HUMADRBR      0 -------KEVYILLNWIGYVNSGFNPLIYCRS-PDFRIAFQELL-CLRRSSLKAYGNGYSSNG
HUMBAR        0 -------KEVYILLNWIGYVNSGFNPLIYCRS-PDFRIAFQELL-CLRRSSLKAYGNGYSSNG
RATB2AR       0 -------KEVYILLNWLGYVNSAFNPLIYCRS-PDFRIAFQELL-CLRRSSSKTYGNGYSSNS
RATADBC       0 -------KEVYILLNWLGYVNSAFNPLIYCRS-PDFRIAFQELL-CLRRSSSKTYGNGYSSNS
HAMARBR       0 -------KEVYILLNWLGYVNSAFNPLIYCRS-PDFRIAFQELL-CLRRSSSKAYGNGYSSNS
TKYARBR       0 -------DWLFVFFNWLGYANSAFNPIIYCRS-PDFRKAFKRLL-CFPRKADRRLHAGGQPAP
RATA2BR       0 -----VPHGLFQFFFWIGYCNSSLNPVIYTVFNQDFRRAFRRIL-CRPWTQTGW
HUMA1AADR     0 ------SEGVFKVIFWLGYFNSCVNPLIYPCSSREFKRAFLRLLRCQCRRRRRRRPLWRVYGH
RATADRA1A     0 ------SEGVFKVIFWLGYFNSCVNPLIYPCSSREFKRAFLRLLRCQCRRRRRRLWSLRPPLA
RATADPA1B     0 ------PDAVFKVVFWLGYFNSCLNPIIYPCSSKEFKRAFMRILGCQCRGGRRRRRRRRLGAC
RATA1B        0 ------PDAVFKVVFWLGYFNSCLNPIIYPCSSKEFKRAFMRILGCQCRGGRRRRRRRRLGAC
BOVANDRE      0 ------SETVFKIAFWLGYLNSCINPIIYPCSSQEFKKAFQNVLRIQCLRRKQSSKHTLGYTL
HUMDD2        0 -------PVLYSAFTWLGYVNSAVNPIIYTTFNIEFRKAFLKILHC
HUMD2A        0 -------PVLYSAFTWLGYVNSAVNPIIYTTFNIEFRKAFLKILHC
HUMDRD2A      0 -------PVLYSAFTWLGYVNSAVNPIIYTTFNIEFRKAFLKILHC
RATD2REC      0 -------PVLYSAFTWLGYVNSAVNPIIYTTFNIEFRKAFMKILHC
MUSD2AR       0 -------PVLYSAFTWLGYVNSAVNPIIYTTFNIEFRKAFMKILHC
HUMDOPD2      0 -------PVLYSAFTWLGYVNSAVNPIIYTTFNIEFRKAFLKILHC
RATD2RA       0 -------PVLYSAFTWLGYVNSAVNPIIYTTFNIEFRKAFMKILHC
RATDOPRD2     0 -------PVLYSAFTWLGYVNSAVNPIIYTTFNIEFRKAFMKILHC
XELD2DOPAR    0 -------QALYSAFTWLGYVNSAVNPIIYTTFNVEFRKAFIKILHC
RATDD3        0 -------PELYRATTWLGYVNSALNPVIYTTFNVEFRKAFLKILSC
HUMDRD5A      0 AGFPCVSETTFDVFVWFGWANSSLNPVIY-AFNADFQKVFAQLLGCSHFCSRTPVETVNISNE
HUMHD5DR      0 AGFPCVSETTFDVFVWFGWANSSLNPVIY-AFNADFQKVFAQLLGCSHFCSRTPVETVNISNE
RATD1BR       0 TGFPCVSETTFDIFVWFGWANSSLNPIIY-AFNADFRKVFAQLLGCSHFCFRTPVQTVNISNE
HUMD1DO       0 TQPFCIDSNTFDVFVWFGWANSSLNPIIY-AFNADFRKAFSTLLGCYRLCPATNNAIETVSIN
RATD1DRA      0 TQPFCIDSITFDVFVWFGWANSSLNPIIY-AFNADFQKAFSTLLGCYRLCPTTNNAIETVSIN
```

4.2.- Analysis of the conservation pattern from the Multiple Sequence Alignment (MSA), excluding the 5HT2 subclass.

The analysis has been restricted to putative TMH domains, and the residue identifier for each position within the MSA is marked by "NAA" in the figures and tables presented. The results are presented in two formats corresponding to 2 different categories, as described in Section 2.3.1. First, the results of the conservation of properties that have been quantified for each natural AA -hdp and volume- are presented together with the conservation indexes. Figure 4.1.a-f compares the relative conservation among the 3 calculated properties for each TMH individually, where the TMH boundaries are putative predictions until they are defined in Section 5. For ease of reference, the sequence of the 5HT2CR has been added to the AA numbers on the abscissa. The respective threshold values for each property (ΔVol=22, Δhdp=1.2) are indicated by a line through the plots. Positions identified as conserved by these criteria are marked with an asterisk. As an illustrative case see H5 residues $Tyr_{5.48}$ and $Ile_{5.49}$; 5.48 is conserved in volume but poorly conserved in hdp, while for 5.49 the situation is the exact opposite. In fact, 5.48 is either Tyr or Phe thus 100% conserved aromatic, while 5.49 is always very hydrophobic but variable, indicative of a lipid-exposure (see Section 2.3.2.1).

The results of the conservation of properties that have been quantified as % of AA type, e.g. aromaticity (see Section 2.3.1.2) are shown in Table 4.2 which presents each AA position and the conserved property that is > 70%

conserved at this locus. In these cases, the AA property code (see Section 2.3.1.2) and the % conserved is shown. In cases where an individual AA is 100% conserved the residue single letter code is also shown. The sequence of the 5HT2C is shown with the AA number for reference. Note that the above mentioned Tyr_{548} is 100% conserved in aromaticity (99% due to numerical roundoff.

Table 4.2 :

Conservation analysis in terms of AA properties quantified by % of AA type. Conserved residues according to this criterion (>70%) are shown by the AA number, by the AA most conserved at that position, and the 5HT2C sequence in parenthesis. Residues conserved 100% in the AA identity show this AA name using the single-letter code. The codes used for each property are shown below -see Section 2.3.1.2.

Description	Code
1) Positively charged groups: {R, K}.....	+
2) Negatively charged groups: {E, D}.....	-
3) Aromatic groups: {F, W, Y, H}.....	arom
4) Betaß-Branched hydrophobic: {I, V}.....	bbra
5) HBD/HBA at position g along the side-chain: {S, T, C}.....	hdag
These 3 are considered as good HBA as HBD because in an alpha-Hx environment they often H-bond back to the C=Oi-4, thus rendering the O or S lone pair available for H-bonding.	
6) HBD at position d along the side-chain: {N, H}.....	hbdd

7) HBD at position e along the side-chain: {Q, H, W, R}..... hbde

8) HBD at position z along the side-chain: {K}..... hbdz

9) HBD at position μ along the side-chain: {Y, R}..... hbdm

10) HBA at position a along the side-chain: {G, P}..... hbaa

Here I refer to the C=Oi-4 of these residues as HBA, since they render a O lone pair accessible to solvent- see table 5.

11) HBA at position d along the side-chain: {N, D, H}..... hbad

12) HBA at position e along the side-chain: {Q, E, H}..... hbae

AA		PROPERTY_ %conserved		
143-I(I)		bbra_ 76.0		
147-V(I)		bbra_ 80.0		
149-G(G)		hbaa_ 84.0		
150-N(N)	N	hbdd_100.0	hbad_100.0	
153-V(V)	V	bbra_100.0		
157-V(V)		bbra_ 99.0		
161-R(K)		+ 76.0		
163-L(L)	L			
164-Q(H)		hbde_ 90.0		
240-N(N)		hbdd_ 99.0	hbad_ 99.0	
241-Y(Y)		arom_ 70.0		
243-I(L)		bbra_ 82.0		
245-S(S)		hdag_ 90.0		
246-L(L)	L			
247-A(A)	A			
250-D(D)	D	- 100.0	hbad_100.0	
257-V(V)		bbra_ 99.0		
259-P(P)	P	hbaa_100.0		
260-F(L)		arom_ 75.0		
266-V(L)		bbra_ 75.0		
343-C(C)	C	hdag_100.0		
344-D(P)		- 89.0		
346-W(W)		arom_ 99.0		
350-D(D)	D	- 100.0	hbad_100.0	
351-V(V)		bbra_ 99.0		
354-C(S)		hdag_ 76.0		
355-T(T)	T	hdag_100.0		
357-S(S)	S	hdag_100.0		
358-I(I)	I	bbra_100.0		
361-L(L)	L			
362-C(C)		hdag_ 96.0		
364-I(I)	I	bbra_100.0		
367-D(D)	D	- 100.0	hbad_100.0	
368-R(R)	R	+ 100.0	hbde_100.0	hbdm_100.0
369-Y(Y)	Y	arom_100.0	hbdm_100.0	
372-I(I)		bbra_ 99.0		
373-T(R)		hdag_ 73.0		

```
438-T(S)          hdag_ 99.0
441-R(K)            +   87.0
446-I(I)          bbra_ 90.0
449-V(V)          bbra_ 81.0
450-W(W)    W     arom_100.0      hbde_100.0
453-S(S)          hdag_ 81.0
456-I(V)          bbra_ 83.0
457-S(S)    S     hdag_100.0
460-P(I)          hbaa_ 81.0
538-Y(F)          arom_ 97.0      hbdm_ 79.0
540-I(L)          bbra_ 86.0
542-S(G)          hdag_ 93.0
543-S(S)          hdag_ 97.0
544-I(F)          bbra_ 71.0
546-S(A)          hdag_ 80.0
547-F(F)    F     arom_100.0
548-Y(F)          arom_ 99.0      hbdm_ 84.0
549-V(I)          bbra_ 71.0
550-P(P)    P     hbaa_100.0
553-I(I)          bbra_ 80.0
558-Y(Y)    Y     arom_100.0      hbdm_100.0
560-R(L)            +   93.0      hbde_ 78.0    hbdm_ 75.0
561-I(T)          bbra_ 99.0
562-Y(I)          arom_ 99.0      hbdm_ 73.0
566-R(R)            +   84.0
568-R(Q)          hbde  84.0
571-K(M)            +   70.0
629-R(N)            +   97.0
630-E(E)    E       -  100.0      hdae_100.0
631-K(K)            +   83.0
632-K(K)            +   99.0
639-I(I)          bbra  97.0
640-I(V)          bbra  99.0
642-G(F)          hbaa_ 98.0
644-F(F)    F     arom_100.0
645-I(L)          bbra_ 74.0
647-C(M)    C     hdag_100.0
648-W(W)    W     arom_100.0      hbde_100.0
650-P(P)    P     hbaa_100.0
651-F(F)    F     arom_100.0
652-F(F)    F     arom_100.0
740-W(W)    W     arom_100.0      hbde_100.0
742-G(G)    G     hbaa_100.0
743-Y(Y)          arom_ 99.0      hbdm_ 90.0
745-N(C)    N     hbdd_100.0      hbad_100.0
746-S(S)    S     hdag_100.0
749-N(N)    N     hbdd_100.0      hbad_100.0
750-P(P)    P     hbaa_100.0
751-I(L)          bbra_ 80.0
752-I(V)    I     bbra_100.0
753-Y(Y)    Y     arom_100.0      hbdm_100.0
```

```
754-T(T)          hdag_ 74.0
759-D(I)            -   99.0
760-F(Y)     F    arom_100.0
761-R(R)            +   89.0      hbde_ 80.0      hbdm_ 71.0
764-F(F)     F    arom_100.0
```

4.3.--Identification of amino acid positions in the 5HT2C receptor sequence where there are significant deviations from the pattern of conserved properties explored above.

The identified positions where such deviations exist are listed below, with the particular property conserved indicated. The code for the properties is taken from Section 2.3.1.2 (see Table 4.2).

NAA	GPCR	5HT2C	cons. index	σ_{vol}	σ_{hdp}
3.26	D	P			1.34
3.35	C	F	0.59		0.37
4.62	F	V		18.05	
5.42	S	G	0.71	10.98	
5.61	I	T	0.18	13.50	0.15
5.62	Y	I	0.08	0.50	
5.65	A	L	0.04		1.00
6.29	R	N	0.63	15.37	
6.41	M	F		19.38	
6.42	G	F		16.50	0.20
6.47	C	M	0.00	0.00	0.00
6.49	L	C	0.10	17.50	0.50
7.45	N	C	0.00	0.00	0.00
7.59	D	I	0.11	15.00	0.00
7.67	I	Y	0.68	12.50	0.29

NAA	GPCR	5HT2C	PROPERTY_ %conserved
2.43	I	L	bbra_ 82.0
2.60	F	L	arom_ 75.0
2.66	V	L	bbra_ 75.0
3.26	D	P	- _ 89.0
3.55	T	R	hdag_ 73.0
4.60	P	I	hbaa_ 81.0

5.38	Y	F	arom_ 97.0	hbdm_ 79.0	
5.40	I	L	bbra_ 86.0		
5.42	S	G	hdag_ 93.0		
5.44	I	F	bbra_ 71.0		
5.46	S	A	hdag_ 80.0		
5.48	Y	F	arom_ 99.0	hbdm_ 84.0	
5.60	R	L	+ _ 93.0	hbde_ 78.0	hbdm_ 75.0
5.62	Y	I	arom_ 99.0	hbdm_ 73.0	
5.71	K	M	+ _ 70.0		
6.29	R	N	+ _ 97.0		
6.42	G	F	hbaa_ 98.0		
6.45	I	L	bbra_ 74.0		
6.47	C	M	hdag_100.0	C	
7.45	N	C	hbdd_100.0	hbad_100.0	N
7.51	I	L	bbra_ 80.0		
7.59	D	I	- _ 99.0		
7.60	F	Y	arom_100.0	F	

The most significant deviation of the 5HT2CR from the conservation profile derived from neurotransmitter GPCRs (which did not include the 5HT2Rs) is the substitution found at two 100% conserved positions: C6.47 becomes a Met, and N7.45 becomes a Cys. The coordinated pattern of substitution observed for these two sites through evolution would suggest an interrelated role and possible spatial proximity between these side chains (see review (Ballesteros 1995) in Section 2.1). However, there are many alternative explanations possible at this stage for the deviations in the conservation profile identified for the 5HT2CR sequence. Their significance will be discussed below whenever an evolutionary criterion is used to direct the modelling, since only then we will have a structural template on which to explore the significance of the differences presented above (e.g., see the proposed interrelated role between M6.47 and C7.45 in the context of

modeling the Pro-kink of H6 in Section 7.2.5). An illustrative case is that of the conserved Arg/Lys at cytoplasmic boundaries, that is never 100%; the length of these side-chains and the capability of fulfilling their role to interact with the phospholipid head-groups from a nearby position renders those changes in AA character non-critical. It is noteworthy that some of the AA sites identified as significant alterations of the consensus GPCR template are or will be proposed to belong to the binding site, e.g. 5.46-S(A) (see Section 8.2).

Figure 4.1.a

TMH1 conservation index, conservation of volume and hydrophobicity (hdp)

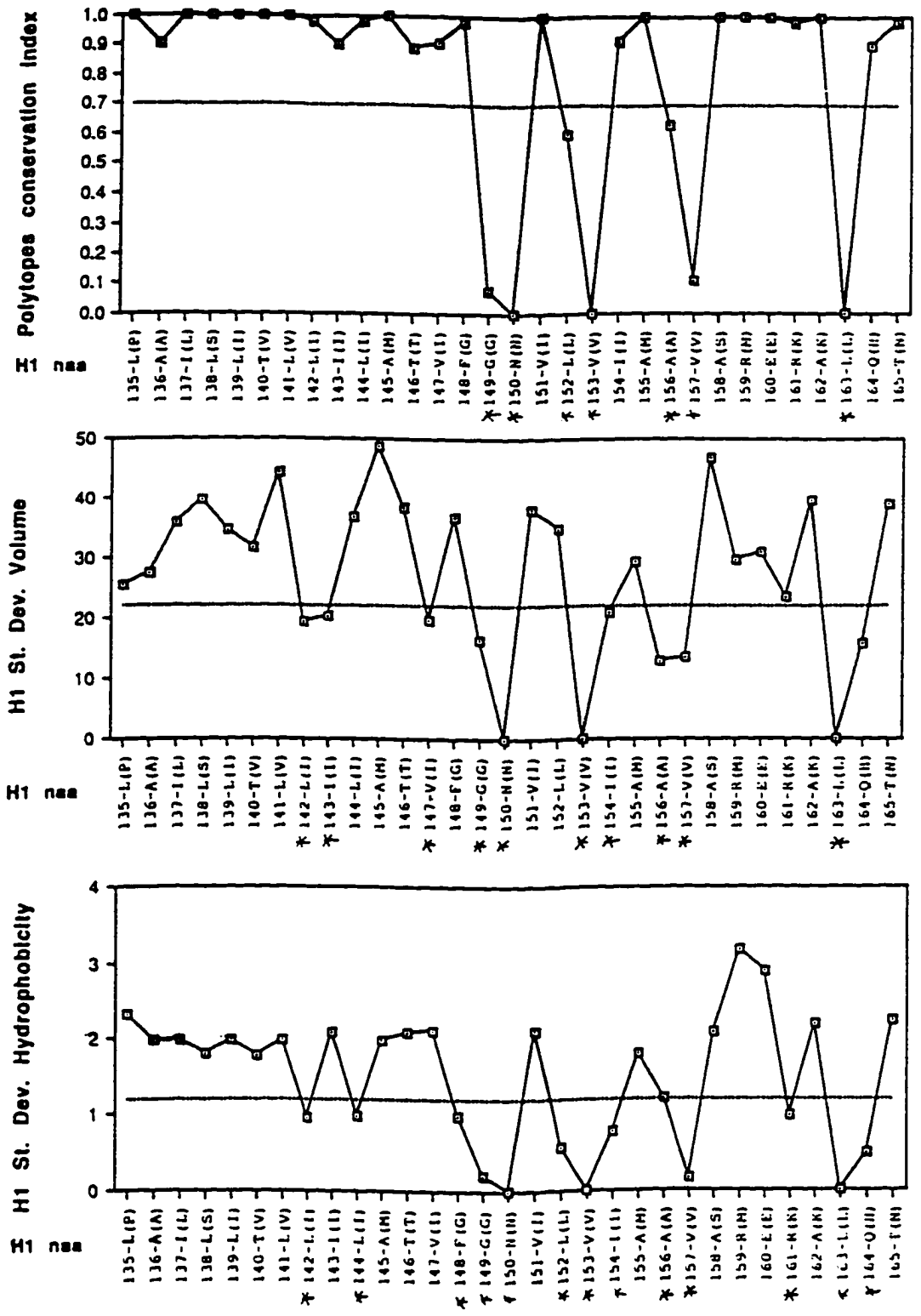


Figure 4.1.b

TMH2 conservation index, conservation of volume and hydrophobicity (hdp)

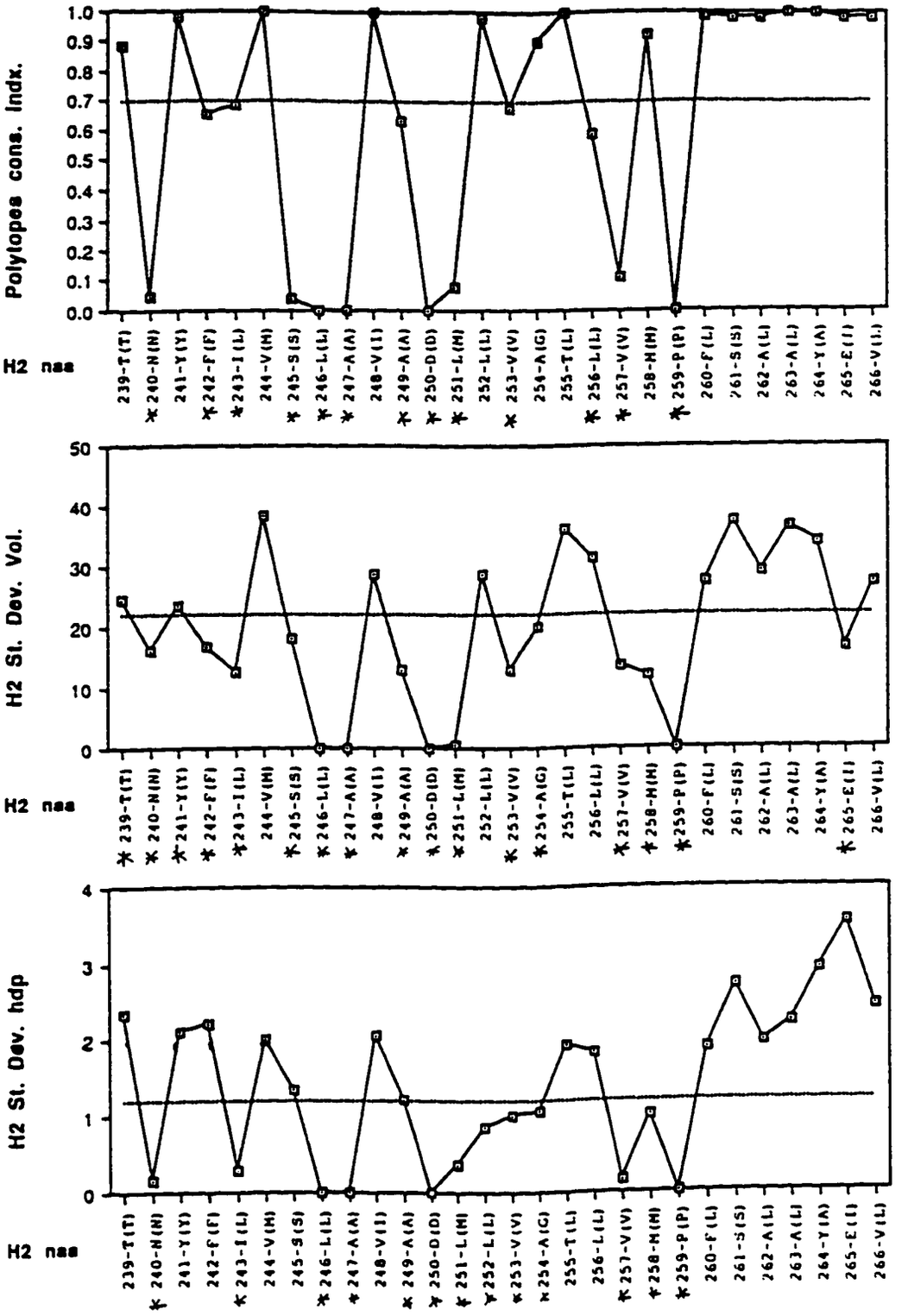


Figure 4.1.c

TMH3 conservation index, conservation of volume and hydrophobicity (hdp)

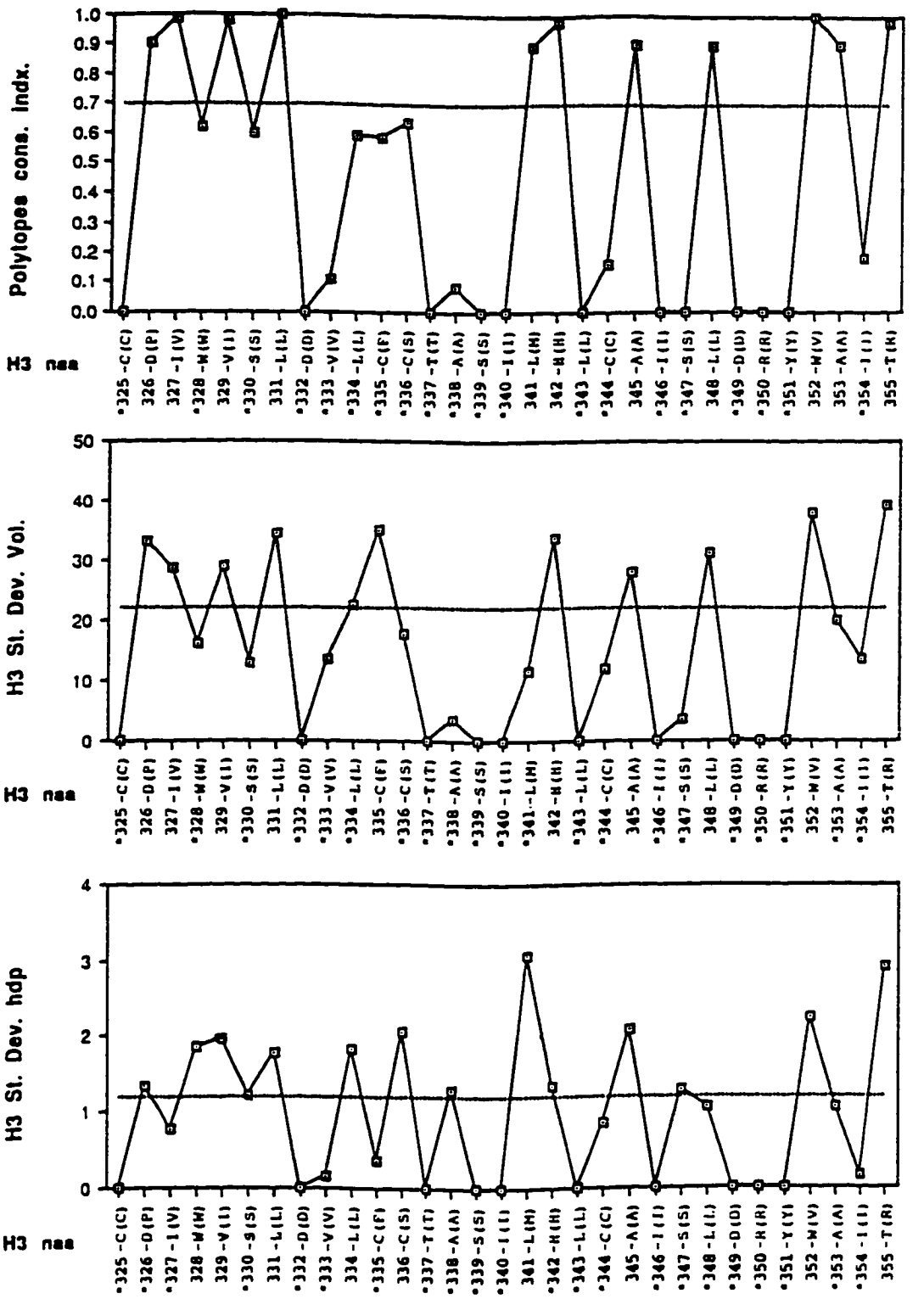


Figure 4.1.d

TMH4 conservation index, conservation of volume and hydrophobicity (hdp)

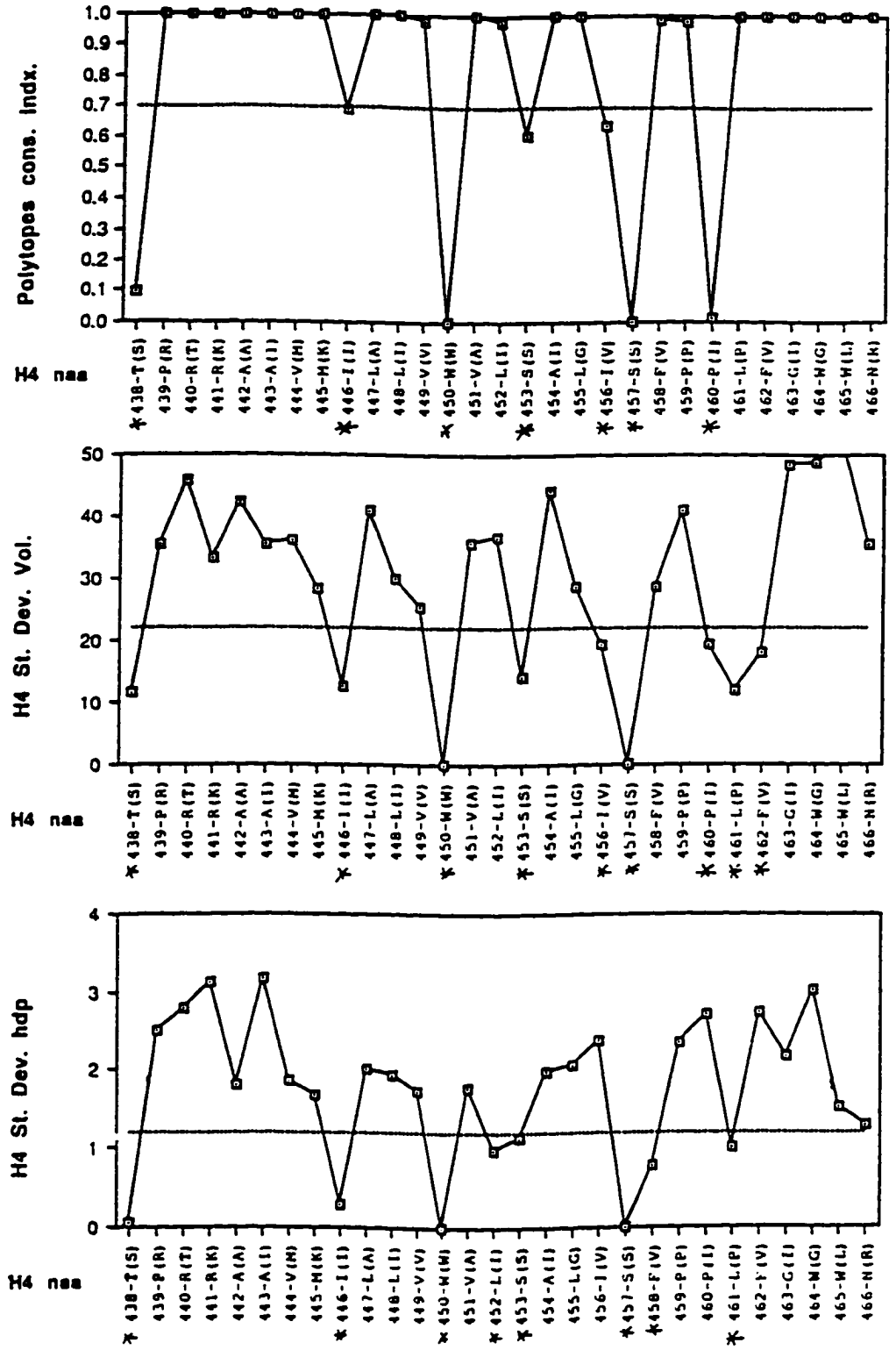


Figure 4.1.e

TMH5 conservation index, conservation of volume and hydrophobicity (hdp)

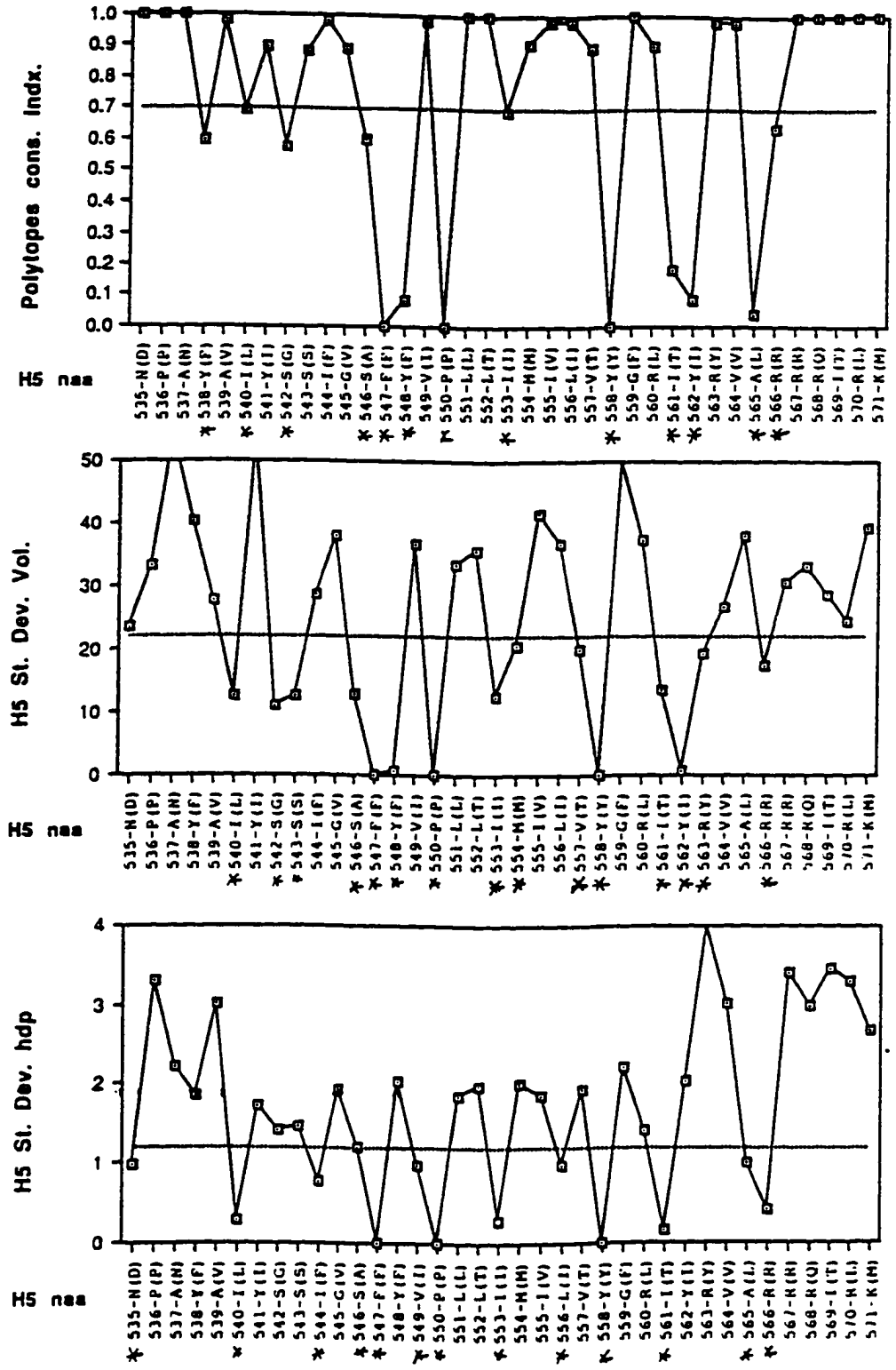


Figure 4.1.f

TMH6 conservation index, conservation of volume and hydrophobicity (hdp)

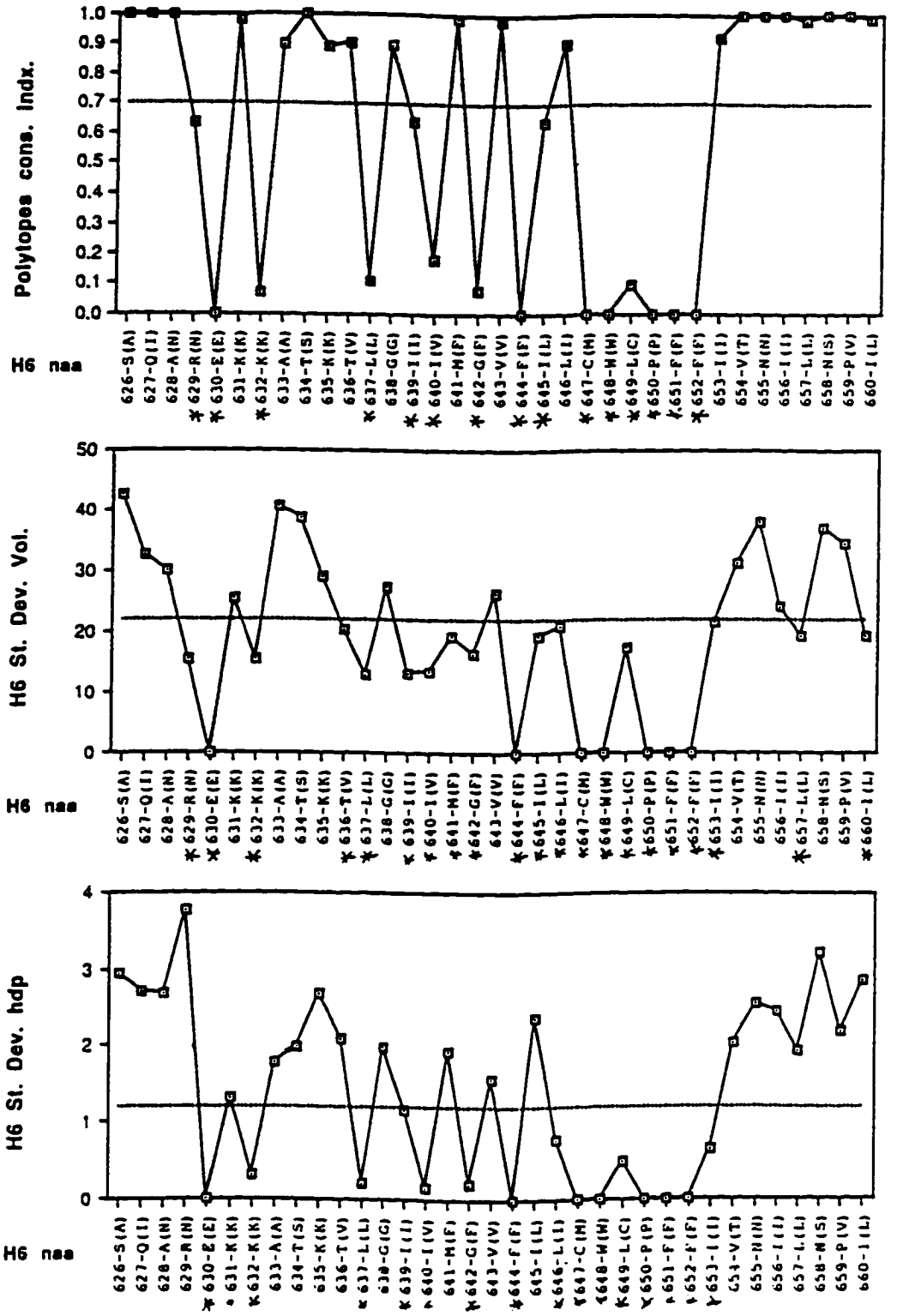
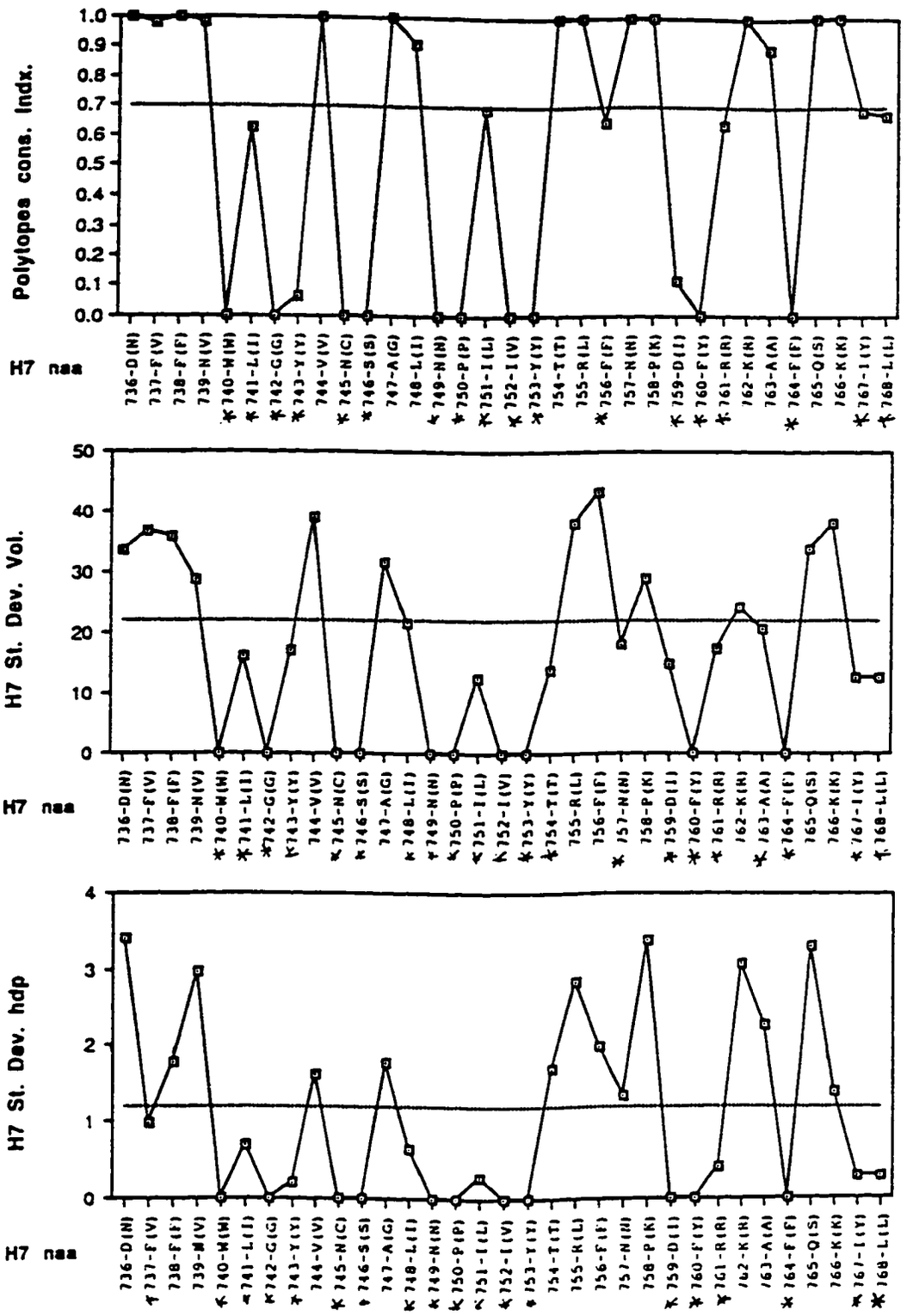


Figure 4.1.g

TMH7 conservation index, conservation of volume and hydrophobicity (hdp)



5.- SECONDARY STRUCTURE ANALYSIS: Refined prediction of the boundaries of the 7 TMHs in the 5HT2CR by the combined use of previously proposed procedures, pursuing convergence among their results. Test of these predictions with experimental data pertinent to 2º structure.

Derivation of these TMH boundaries entails detailed integration of the extensive conservation analysis presented in the previous section for each single AA position. The inferences derived from each criterion applied independently to each TMH are presented in Sections 5.1 to 5.6. The presentation of each criterion is done on helical nets where the resulting TMH ends are already included. Consistency among the diverse criteria applied can thus be directly analyzed visually in the graphical representations of the results from each individual criterion; inconsistencies become readily apparent in this representations and can be thus simply highlighted at each step. However, and as an illustrative case, I will discuss in a bit more detail the derivation of the TMH boundaries for H7. The derivation of the TMH ends for the other six TMHs will be briefly reviewed in Section 5.7.

The proposed TMH boundaries are compared with pertinent experimental data in Section 5.8, where the consistency between the theoretical and experimental results validates the methodological approaches applied. Interestingly, analysis of the experimental results in a structural context provided a reinterpretation of current techniques to determine TMH ends in GPCRs that was validated experimentally in a collaborative work with the laboratory of Dr. Jonathan Javitch (Fu 1996).

5.1- Initial definition of the 5HT2C TMH boundaries in terms of the hydrophobicity (hdp) profile, as presented in the literature (Julius 1988).

The predicted boundaries using this criterion (Julius 1988) are shown in Figure 5.1 on top of the prediction made in this thesis. Thus, the actual

TMH boundaries of this figure correspond to our predictions. Residues whose prediction based on the hdp profile differ from the prediction here proposed are shown shaded. The differences between the two sets of TMH ends are shown below, where NT and CT stands for N-terminus and C-terminus respectively. ΔCyt. and ΔExt. stands for differences in the cytoplasmic and extracellular TMH boundaries, respectively.

	This work		hdp profile		(this work - hdp)			
	NT	CT	NT	CT	ΔNT	ΔCT	ΔCyt.	ΔExt.
TMH1	P1.35	H1.64	Q1.32	V1.57	3	7	7	3
TMH2	N2.40	Y2.67	Y2.41	S2.61	-1	6	-1	6
TMH3	C3.25	R3.55	V3.27	L3.48	-2	7	7	-2
TMH4	S4.38	P4.59	A4.42	G4.64	-4	-5	-4	-5
TMH5	N5.37	R5.67	F5.38	F5.59	-1	8	8	-1
TMH6	E6.30	V6.59	V6.36	L6.57	-6	2	-6	2
TMH7	L7.35	C7.70	L7.34	F7.56	1	14	14	1

Significant differences can be observed in the extracellular boundaries of TMH2 and TMH4, and in the cytoplasmic boundaries of TMH1, TMH2, TMH4, TMH5, TMH6, TMH7. Therefore most of these differences appear at the cytoplasmic side of the membrane and are due to the Arg/Lys motif (TMH1, TMH4, TMH5, TMH6, TMH7), as shown before for bacteriorhodopsin (see review (Ballesteros 1995) in Section 2.1). The other significant cytoplasmic difference occurs at TMH3 and can be explained by the conservation profile as presented in Section 5.5. Significant differences occur at the extracellular end of TMH2 and TMH4. The difference in the TMH2 extracellular boundaries are not very important because for most other GPCR the TMH boundaries predicted based on the hdp profile agree with the prediction proposed here

(see (Probst 1992) for a review). The difference in TMH4 is due to the presence of Pro-Pro dipeptide as presented in Section 5.6. The case of TMH7 is special in that there is a 14 AA difference between the two predictions; this case will be discussed in some detail for every prediction criterion in that section as an illustrative example.

5.2- Analysis of the occurrence of insertions or deletions (indels) within putative TMH segments as indication of helix ends.

The positions where indels occur are shown in Figure 5.2 as shaded AA on a helical net representation of the 5HT2CR. Note the general consistency of the predictions with this criterion, i.e. indels occur outside the TMH boundaries or in the first or last positions of the TMH. Two exceptions can be found, discussed below:

Indel at A_{442} : This indel occurs only at the human 5HT1A GPCR, with code HUM5HT1A in Table 4.1. It represents a clear inconsistency with the proposed criterion, and it was considered a cloning error in the sequence, since the sequence of another human 5HT1a GPCR -code HSHTRB- is absolutely identical to the HUM5HT1A, including the very poorly conserved extracellular N-terminus, except for TMH4. Therefore, this indel was not considered relevant for the prediction of TMH boundaries. Since the time of this prediction, this indel has been shown to be, indeed, a cloning error, thus illustrating the power of this analysis in interpreting data derived from a

particular receptor sequence in the context of the remaining GPCRs sequences.

Indels at T_{754} and N_{757} : Both are single indels separated by 3 AA within TMH7 and are thus positioned next to each other in a TMH, as seen in Figure 5.2. Both indels can be consistently found in a number of GPCR (see Table 4.1) and are thus a strong prediction of non-helical character. This prediction will be contrasted with the results of the other prediction tools presented in that section in the integration of the results presented below (Section 5.7).

5.3.- Application of the hypothesis that Arg and Lys residues adjacent to TMH in the cytoplasmic side actually belong to the TMH and can be used to define the cytoplasmic end of these TMH.

The positions where those R/K occur within the GPCR alignment (Table 4.1) are shown in Figure 5.3 as shaded AA on a TMH net representation of the 5HT2CR. The AA sites where an Arg or Lys residue exists in the 5HT2C sequence are marked with a thicker circle for clarity. Note the general consistency of the predictions with this criterion. In particular, TMH7 displays a highly remarkable TMH periodicity on the evolutionary disposition of those R/K in the sequence; they always occur on a single face of TMH7 predicted to face the lipids (see Section 6). This is considered an strong indication of TMH character. The 5HT2C sequence is in itself an example of

that TMH periodicity of the R/K sites; note the TMH patch defined by K7.58, R7.61, R7.62, K7.66, and R7.69. Furthermore, the palmytolated Cys7.70 belongs to this same TMH patch predicted to be the TMH-Lipid interface for TMH7 (see Section 6). The palmytoyl moiety is strongly predicted to face the lipid milieu since its aliphatic component is chemically identical to the lipid chains. The breakdown of TMH periodicity in the distribution of R/K residues after C7.70 is sharp and clear (a continuous stretch of 5 AA sites all having R/K residues at some GPCR follows), thus defining the TMH7 cytoplasmic boundary. Those AA sites will not fulfill the criterion of evolutionary consistency for the prediction of TMH boundaries (Section 2.1).

The only exception of this hypothesis found in the GPCR is Arg3.50, predicted to lie in TMH-TMH interfaces. The basis of this prediction, presented in Section 6, is that this residue is 100% conserved not only in the GPCR alignment of homologous GPCR (Table 4.1) but also in all GPCRs belonging to the rhodopsin family. Furthermore, this residue belongs to the most conserved TMH patch of TMH3.

5.4- Prediction of TMH segments from α-helix periodicity of properties measured by Fourier transform analysis.

The plots of the quantification of TMH significance ψ versus the sliding window of 12 AA (ΔAA) were presented in Figure 2.8 of our review chapter (Ballesteros 1995) (see Section 2.1). The significance threshold of ψ=2

proposed in the literature is shown for each plot. It should be noted that those plots were done using a variability profile V(NAA) quantified by the number of different AA present at a given AA site within the alignment, as presented in the literature (Section 2.1). Although the conservation index defined in this thesis could be used for that purpose, a test performed on the PRC showed puzzling results, in that some TMHs were predicted with much higher accuracy while other were not predicted at all. Thus, the Fourier transform analysis was performed using the variability profile. Through the {ψ,ΔAA} plots presented in Figure 2.8 of our review chapter (see Section 2.1) the predicted TMH boundaries using the criterion of maximal relative drop is shown by vertical lines. Those predictions are listed below and compared with the final predictions (the notation is as explained above):

	final predictions		Fourier predictions		(final - Fourier)			
	NT	CT	NT	CT	ΔNT	ΔCT	ΔCyt.	ΔExt.
TMH1	P1.35	H1.64	N1.33	H1.64	2	0	0	2
TMH2	N2.40	Y2.67	N2.37	P2.59	3	8	3	8
TMH3	C3.25	R3.55	S3.28	R3.55	-5	0	0	-5
TMH4	S4.38	P4.59	S4.38	I4.63	0	-4	0	-4
TMH5	N5.37	R5.67	I5.41	Q5.68	-4	-1	-1	-4
TMH6	E6.30	V6.59	N6.29	N6.55	1	4	1	4
TMH7	L7.35	C7.70	F7.38	D7.71	-3	-1	-1	-3

From this table it can be seen that the differences in the predictions is minimal at the cytoplasmic side of the TMH. Interestingly, that implies a remarkable convergence between those predictions and the predictions using the Arg/Lys motif at cytoplasmic boundaries; this convergence is shown by

marking these Arg/Lys with a positive sign in the {ψ,ΔAA} plots presented in Figure 2.8 of our review chapter (see Section 2.1).

From this table it can also be seen that the differences in the predictions are concentrated at the extracellular side of the TMH. There are two reasons for that divergence, both already discussed in Section 2.1; the effect of the Pro-kink (PK) and the effect of functionally divergent TMH portions. I discuss both effects below.

First, the effect of the PK twisting the face of the TMH segment N-terminal from the Pro relative to the TMH segment C-terminal from the Pro (see review Section 2.1). The net effect is the presence of two different periodicities, one within each TMH segment. This effect would vitiate the TMH periodicity ψ as quantified here since a window of 12 AA would incorporate both TMH segments at the same time, thus significantly decreasing the calculated ψ value. Such an effect can be observed for TMH7 in Figure 2.8.f of our review chapter (see Section 2.1), where the position of Pro7.50 induces a significant drop in ψ. This effect is especially damaging when one of the two TMH segments (NT or CT from the Pro) is too short to induce a significant value of ψ for a window of 12 AA; i.e. consider calculating ψ for a TMH five residues long with a window of 12 AA. This seems to be the case for TMH2, TMH5, and TMH6.

Second, the value of ψ is heavily dependent on the particular subset of proteins chosen, varying from $\psi=4$ to $\psi<2$ for the same TMH segment -data

not shown. The subset dependence is observed whenever a TMH segment has sequence variability related to functional diversity in addition to structural constraints; such a subset would significantly decrease the ψ relative to the subset containing only functionally homologous proteins. For the case of the GPCR considered here, TMH portions forming the ligand binding pocket would have a significantly lower calculated ψ value. Those portions have been shown to lie in the extracellular half of the TMH domain, thus potentially affecting the predictions made for TMH2, TMH3, TMH4, TMH5, TMH6, and TMH7. An illustrative example is the extracellular segment of TMH7, where residue V7.39 has been shown to interact with ligands based on experimental data (see Section 2.2), thus predicted to face the binding site away from the lipid based on functional criteria. However, position 7.39 shows a very divergent conservation pattern which in terms of the structural criteria proposed would normally correspond to lipid-exposed residues. The net effect is that functionally divergent helical segments may show a more relaxed or interrupted helical periodicity in their conservation pattern, and thus may not be recognized by the Fourier transfer method.

5.5- Prediction of TMH segments from α-helix periodicity measured by surface patches on a helical net representation.

The conservation of physico-chemical properties as defined in Section 2.3.1.2, in addition of the conservation of specific AA residues, will be used to

define these sites on the 5HT2C receptor sequence.

The AA sites selected as conserved are shown in Figures 5.4 to 5.8 as shaded AA on a TMH net representation of the 5HT2CR. The different properties whose conservation profile have been quantified are presented in 5 figures, as listed below:

Figure 5.4 : AA sites conserved by the criterion of the conservation index.

Figure 5.5 : AA sites conserved by the criterion of volume.

Figure 5.6 : AA sites conserved by the criterion of hdp.

Figure 5.7 : AA sites conserved by the criterion of H-bonding and charge.

Figure 5.8 : AA sites conserved by the criterion of aromatic or β-branched character.

From these figures I would emphasize the general consistency of the predictions with the various criteria. I will discuss TMH7 in some detail as an illustrative case.

In Figure 5.4 AA sites with a conservation index less than 0.7 are shown shaded. Note that for TMH7 those sites are positioned within a common face of the TMH. Especially important is the prediction of the cytoplasmic TMH boundaries, since they differ by 14 AA from previous predictions based on the hdp profile (see Section 5.1). It can be observed that conserved AA sites by this criterion consistently support a continuous TMH

from the hydrophobic core into the cytoplasmic projection, up to the last predicted turn. This is considered as a strong prediction of the cytoplasmic TMH boundary at the palmytolated Cys7.70. The predicted extracellular boundaries exceed those supported by the conservation index. The reason may be the role of this portion of TMH7 in ligand binding, as suggested for position 7.39 based on mutagenesis experiments (see Section 2.1).

The criterion of volume conservation (Figure 5.5) does not improve the prediction of the TMH boundaries as discussed for the conservation index (Figure 5.4) for TMH7. Nonetheless, it provides additional support.

The criterion of conservation of the hdp character provides additional support for the predicted TMH boundaries at the extracellular side. Val7.37 in the last turn forms a continuous TMH patch with the highly conserved residues {W7.40,I7.41,Y7.43,S7.46, etc...} thus suggesting that it belongs to TMH7.

The other conservation criteria presented do not affect the predictions made for TMH7 already discussed, but support them. Of special interest may be the high conservation of aromatic character (Figure 5.8) for residues Y7.60 and F7.64, which belong to the controversial cytoplasmic extension of TMH7.

5.6- Analysis of the occurrence of non-conserved Pro residues within putative TMH segments as indication of helix ends.

The positions where Pro residues occur within the GPCR alignment

(Table 4.1) are shown in Figure 5.9 as shaded AA on a TMH net representation of the 5HT2C GPCR. As mentioned in the published review (Ballesteros 1995) (Section 2.1) Pro residues are allowed at the N-terminus and forbidden before the C-terminus, unless highly conserved within the TMH where they form a PK. Some incompatibilities with these predictions can be observed in Figure 5.9, and are discussed below:

K4.41 : A Pro residue appears here only for the HUM5HT1A GPCR, and as discussed in Section 5.2 in the presentation of the insertions and deletions (indels) this sequence was considered suspicious (later corrected as a cloning error) and should thus not be considered in that analysis.

T7.54, K7.58 and R7.62 : The presence of those Pro residues is not accidental, since there is a constant 4 AA increment between these three positions thus defining a single continuous patch on the TMH surface (see Figure 5.9), and they occur in several GPCRs. This suggests that their conformational effects on H7 are not just a breakdown of α-helical character and thus that their position bears significance for the prediction of TMH7 boundaries. Because the patch defined by prolines 7.54, 7.58, and 7.62 belongs to the Arg/Lys motif described above (see Figure 5.3), all three are predicted in TMH-Lipid interfaces. This implies that the direction of bending induced by these PKs would be consistently towards the TMH-TMH interfaces (see Section 7.2.1 for the rationale). This directionality suggests that the presence of these prolines are subjected to specific structural constraints that are pertinent to α-helical

character, rather than a loop structure.

The analysis of the Arg/Lys motif and the conservation pattern of TMH7 would support a continuous TMH7 extending into the cytoplasmic side until Cys7.70. In order to reconcile a continuous TMH7 until Cys7.70 with the presence of both the indels and the Pro residues at the same positions in TMH7, the turn 7.54-7.58 of TMH7 is predicted to be a flexible hinge; the cytoplasmic extension of TMH7 after this hinge is predicted to be helical, and could still form a continuous TMH with the core of TMH7 at some point during the activation process of the GPCR. Whether the continuous TMH is formed during the activated or resting state of the receptor can not be discerned at the present stage. However, this prediction would provide an explanation for the ambivalent inferences derived from the conservation analysis on TMH7, based on evolutionary criteria. There is experimental support for the hypothesis of a helical conformation of the cytoplasmic extension of H7 after the 7.54-7.58 hinge. NMR studies on this segment in isolation, in the presence of lipid vesicles, showed it to adopt a well-defined α-helix (Jung 1995). There is also experimental data supporting a functional role through significant conformational changes at this locus. First, the cytoplasmic extension of TMH7 has been implicated in G-protein coupling in the context of the full receptor (Lefkowitz 1988; Caron 1989; Kobilka 1992), and shown to activate the G-protein as an isolated peptide (Jung 1995). Second, the palmitoylation of Cys7.70 (ODowd 1989; Papac 1992) seems to be important for

GPCR-Gprotein interaction, since mutation of the palmytolated Cys7.70 to Gly produces an uncoupled β2-adrenergic GPCR (ODowd 1989), and DTT treatment which removes the palmitoyl leads to significant changes in the FTIR spectrum of this segment upon rhodopsin activation; the FTIR frequencies for the segment were assigned to helical conformations (Ganter 1992). Third, spin labeling experiments in rhodopsin have shown a significant movement between residues 1.60 and 7.63 upon rhodopsin activation (Yang 1996a).

5.7.- Integration of results from the various criteria.

The derivation of the N-terminal (NT) and C-terminal (CT) TMH boundaries for helical segments H1 to H6 is reviewed briefly below. Describing the integration of the results from of each criterion applied independently in the previous sections, would have required extensive and continuous reference to the figures presented above in this chapter. These references are omitted for clarity, because from the criterion utilized the figure of interest, they can be found in the corresponding section.

H1 NT: The N-terminus predicted by the hdp profile (1.32) and the Fourier transform (1.33) is shifted by 2-3 residues due to the presence of an indel at position 1.35, which becomes the selected H1 end. This selection is favored for the 5HT2CR because a proline at position 1.35 is expected to stabilize the helix

acting as an N-capping residue (see review in Section 2.1).

H1 CT & H2 NT: The criteria for the prediction of the CT of H1 and the NT of H2 overlap, due to the presence of a very short turn predicted between these helices, and are thus presented here together. Fourier transform (FT) predicts the CT of H1 at 1.64 and the NT of H2 at 1.66, indicating the presence of a very short loop. An insertion at 1.66 and a non conserved Pro at 1.67 repositions the NT of H2 at 1.67. This results in a loop of only 2 residues (1.65-1.66). This loop can not be much longer, because at either end there are 100% conserved residues whose periodicity agrees with their respective helices. At the N-terminus of this loop there is a 100 % conserved L1.63 that appears in the patch of other conserved residues within H1 (N1.50, V1.53, V1.57) and should thus belong to H1. The next residue, H1.64, is a 90.0 % conserved H-bond donor at position ε along the side chain although different amino acids at present at this locus, and belongs to a patch of residues facing inward (E1.60, V1.57, N1.50), thus predicted to belong to H1. At the C-terminus of this loop there is a 100 % conserved N2.40 (or N1.68) that appears on the same face as the conserved A2.47-D2.50 and is thus predicted to lie within H2. Thus the maximal H1-H2 loop that can be predicted and it is therefore selected based on this analysis is 1.65-1.66-1.67, with the CT of H1 at H1.64 and the NT of H2 at N2.40. The resulting H1 CT is consistent with the presence of Arg/Lys at positions 1.61 and 1.62 inside the helix, and a continuous stretch after the CT (1.64-1.65-1.66), thus extending the cytoplasmic TMH boundaries predicted by

the hdp profile.

H2 CT : There is a clear drop in the α-helix periodicity measured by FT after P2.59, which could be due to the PK induced by this proline residue. There is an indel at 2.67 thus determining the maximal CT end of H2. The pattern of hdp residues would support either a 2.59 or a 2.66/2.67 CT. There is no significantly conserved residue after P2.59, supportive of a CT at 2.59. I had excluded the β3-adrenergic receptor sequence from the subset of GPCR selected for the conservation analysis because among other significant differences with other GPCRs considered, it had a Pro at position 2.60. This proline results in a Pro-Pro motif at positions 2.59-2.60 that is suggestive of the end of α-helical conformation, but it occurs in only one receptor among all. Thus the theoretical analysis is inconclusive in predicting the CT of H2, as it could be either 2.59 or 2.67. There are experimental data indicating that residues 2.63 or 2.64 (it could not be determined) are the attachment sites for a covalently bound pindolol derivative in the β2-adrenergic receptor (Dohlman 1988), thus suggesting that at least one of these residues is facing the binding site, i.e. H3 at the level of D3.32. Residue 2.64 has been also used to engineer a Zn binding site between H2 and H3 (H2.64-H3.28) (Elling 1996), suggesting that it is facing H3. Because, as it will be shown in Section 6.2.2, adopting an α-helical conformation for the extracellular portion of H2 after P2.59 positions 2.64 facing H3, the CT of H2 was selected at position 2.67.

H3 NT : FT analysis detects α-helical periodicity for H3 starting at position

3.30, far from the maximal ends predicted by the indel at 3.23 or the non conserved prolines at 3.22. The residues between these two possible NT (3.30 or 3.22/3.23) are overall highly polar, but include significantly conserved residues. The C3.25 is 100 % conserved and is i-7 from the D3.32 thus on a common face of an α-helix. The next residue, P3.26, is at the locus of a 89 % conserved acidic residue. Position 3.28 is also significantly conserved, especially in aromatic character, and links positions 3.25 and 3.32 thus defining a continuous patch of conserved residues on an α-helical conformation. Thus C3.25 was selected as the NT of H3.

H3 CT : As the conservation pattern for H3 is analyzed from the NT (C3.25) onwards, it is found that this helix has a remarkable degree of conservation in a continuous stretch of residues spanning from D3.32 to I3.40 (see CI in Figure 5.4), thus inconsistent with α-helical periodicity. Nonetheless, the FT analysis can detect helicity in this region that extends from 3.30 to 3.55. Consistent with this prediction, an indel at 3.57, which is often (88 %) a Pro, indicates that H3 can not extend beyond 3.56. There are highly conserved residues up to I3.54 that define a continuous patch on an α-helical conformation, also consistent with the predicted CT by FT at 3.55, which is therefore selected.

H4 NT : The FT analysis suggest an NT at position 4.38 or 4.42, with a shallow curve between them. The criterion of the Arg/Lys motif at the cytoplasmic boundaries resolves this difficulty because from 4.38 onwards there is a high density of non-conserved Arg/Lys at positions 4.39-4.40-4.41-4.43-4.45 thus

consistent with a-helical periodicity, while before 4.38 there is a continuous stretch of 4 residues containing non-conserved Arg/Lys residues. This helical pattern after position 4.38 is consistent with the high conservation at 4.38, which belongs to a thin patch of residues at (i, i+4) that includes the 100 % conserved β-branched 4.46 and W4.50. There is a non-conserved Pro in the resulting first turn at position 4.40 that could be tolerated. Thus S4.38 is predicted the NT of H4. Note that the resulting H3-H4 loop is only 9 residues long.

H4 CT : The FT analysis shows a drop in helical character at 4.63. However, there are indels at positions 4.61 coupled with many PxP or PP motifs at 4.59-4.60-4.61 that define a maximal CT for H4 at P4.59. This selection is supported by the lack of any significant conservation observed after the 4.59-4.61 segment described above. It follows that the H4-H5 loop following the CT of H4 predicted at P4.59 is highly divergent and unstructured as inferred from its conservation pattern. It should be noted that for receptors containing a single Pro at position 4.59 (or 4.60), the criterion presented here is inconclusive in that it does not suggests a helical conformation after this Pro, but it doesn't preclude it either.

H5 NT : Note the overall low helical periodicity detected by the FT analysis, which nonetheless shows a shallow but significant drop in α-helical character at 5.41. This could be due to a short helical segment N-terminal to the PK at P5.50, as suggested by the maximal NT predicted by the presence of indels at

5.33, far from 5.41 predicted by the FT method. There are non-conserved Pro at positions 5.36 and 5.37, which would redefine an extended NT to 5.36 or 5.37. This is supported by the presence of a highly polar triad at 5.34-5.35-5.36, thus suggesting 5.37 as the NT of H5. The conservation pattern supports this prediction because there is a cluster of conserved polar side chains (Ser, Thr) at position 5.42-5.43-5.46 which can be extended in either end by the highly conserved aromatic character of F5.38 at the N-terminus of this fragment and F5.47-F5.48 at the C-terminus of the Ser/Thr cluster. Note that F5.48 is not in the same patch of conserved residues and would thus be predicted to face the lipid milieu, as will be discussed in Section 6.1. Thus the patch of conserved residues from 5.38 to 5.47 supports a definition of the NT of H5 at N5.37.

H5 CT : FT analysis detect α-helical periodicity up to residue 5.68. The Arg/Lys motif shows a continuous stretch of Arg/Lys from 5.66 to 5.72, of which 5.66 and 5.67 are consistent with α-helical periodicity because they appear in opposite faces than the conserved residues (T5.61-I5.62-L5.65). Thus, the CT is defined at residue 5.67.

H6 NT : Although the α-helical significance of the FT analysis is rather low, a sharp increase at 6.29 predicts this residue as the NT of H6. This is consistent with the indel present at 6.27. There is a high density of Arg/Lys residues at this locus, whose helical pattern is clear after the 100% conserved E6.30. Before E6.30 Arg/Lys residues appear at 6.29, 6.28, and continuously before 6.27, but position 6.29 is offset from the patch of Arg/Lys defined by K6.31-

K6.32-K6.35 and thus half way between the lipid-face and the interior face of H6. Position 6.29 is thus predicted outside the helix. The resulting NT of H6 at E6.30 is consistent with the patch of 100 % conserved residues E6.30-L6.37-V6.40. However, note that K6.32 is highly conserved in terms of the CI, volume, charge and other properties. K6.31 and N6.29 are also highly conserved charged residues. The implication of the high degree of conservation of these Arg/Lys positions will be discussed in the next chapter (Chapter 6) in the context of helix-lipid versus helix-helix interfaces.

H6 CT : The FT analysis suggests a CT at position 6.55, although the presence of a PK at the conserved P6.50 that this CT implies, is likely to hinder FT procedures. This hypothesis was confirmed by analyzing FT using a shorter AA window (7 AA), which shifted the predicted CT to 6.62. This is consistent with the maximal ends predicted by the presence of indels at 6.64-6.65. Non conserved Pro are present at 6.56 and 6.59, indicative of helix ends. However, the predicted CT at 6.62 would allow P6.56, but not P6.59, to form a stable PK, because three residues after P6.59 are insufficient to stabilize an α-helical conformation. Thus the H6 CT is redefined to 6.59.

5.8.- The 2° structure boundaries predicted using these criteria will be tested for their consistency with experimental data pertinent to 2° structure for GPCR: NMR, identified cleavage sites, SCAM and spin labeling experiments.

Direct structural determination of TMH boundaries is not available

yet for GPCRs, except for the NMR studies of an isolated peptide corresponding to the cytoplasmic extension of H7 7.56(345)-7.70(359) here predicted, which appeared as an α-helix in the presence of lipid vesicles (Jung 1996). The boundaries determined by this NMR experiments are in agreement with the controversial prediction developed in this thesis of a cytoplasmic extension of H7 comprising residues 7.57-7.70 (Jung 1996). NMR studies on the C-terminal domain of rhodopsin also suggested the proposed H7 boundary at the palmitoylation site (Yeagle 1995a). In the absence of further direct structural determination, more indirect experimental approaches pertinent to secondary structure are reviewed below, involving cleavage sites, Cys scanning (Javitch 1995a; Javitch 1995b; Fu 1996) and spin labeling studies (Farahbakhsh 1995; Altenbach 1996; Farrens 1996; Yang 1996a; Yang 1996b).

The identified protease cleavage sites in neurotransmitter GPCR (Dohlman 1987; Luxembourg 1991), and the differences with the predicted TMH boundaries, are listed below:

TMH	Cleavage sites	Predicted NT	Predicted CT	ΔAA outside TMH
TMH1	1.19-1.20	1.35		15
TMH5	5.70-5.71		5.67	3
TMH6	6.24-6.25	6.30		5
TMH7	7.63-7.64		7.70	-6

Those cleavage sites should not be within the predicted TMH boundaries since that would render them inaccessible to the protease. Note the consistency with the predictions for TMH1, TMH5 and TMH6. Although the cleavage site for TMH1 is far from the TMH boundaries, there is a remarkable agreement at the cytoplasmic boundaries of TMH5 and TMH6, thus supporting the Arg/Lys motif used in those predictions. For TMH7 there is an inconsistency of 6 AA; the predicted hinge at 7.57 can explain this case since the cleavage site would lie between the hinge and the predicted CT, i.e. 7.57<7.63-7.64<7.70.

The experimental techniques of the Substituted Cys Accessibility Method (SCAM) (Javitch 1995a; Javitch 1995b; Fu 1996) and Spin labeling (Farahbakhsh 1995; Altenbach 1996; Farrens 1996; Yang 1996a; Yang 1996b), described in the Methods Section 2.2.2, provide information about the relative accessibility of each mutated residue in the receptor structure. The accessibility measured by these techniques for each Cys-substituted residue in a continuous stretch of the receptor sequence, is conceptually equivalent to the theoretical methods used in this chapter of the thesis to predict the accessibility of each residue independently. Thus the extent of the α-helical periodicity observed in the experimentally measured accessibilities can be compared to the equivalent analysis made here on theoretical grounds, and thus derive the resulting TMH boundaries by similar procedures. Spin labeling studies on rhodopsin (Farahbakhsh 1995; Altenbach 1996; Yang 1996b)

have proposed the TMH cytoplasmic ends of H3 (3.55), H4 (4.40), H5 (5.66) and H6 (6.33). Although H5 and H6 were shown to protrude extensively into the cytoplasm (up to 5.72 and 6.23, respectively), the authors distinguish these aqueous boundaries from the membrane boundaries observed at 5.66 and 6.33 used here for the TMH ends. The proposed extensions of H5 and H6 into the aqueous cytoplasm in rhodopsin pertain to receptor segments for which I did not observe any conservation among the GPCRs analyzed, except the triad 6.29-6.31-6.32. Thus, the observation that these segments may be helical in rhodopsin does not imply a similar conformation for the equivalent segment of the 5HT2CR. SCAM analysis on the D2 receptor (Javitch 1995a; Javitch 1995b; Fu 1996) have proposed the extracellular boundaries for H3 (3.28), H5 (5.46), and H7 (7.39). The results of these two experimental techniques are shown on a helical net of the 5HT2CR in Figure 5.10 where the TMH ends are as defined in this work. The consistency of the predictions with the experimental results is summarized in the Table 5.1 below (the notation is as explained above, and the H7 CT corresponds to the NMR data described above):

Table 5.1 : Consistency of the predicted TMH boundaries with experimental data pertinent to secondary structure: NMR, SCAM, and Spin labeling.

	Theoretical predictions		NMR, SCAM, Spin labeling		(observed-predicted)			
	NT	CT	NT	CT	ΔNT	ΔCT	ΔCyt.	ΔExt.
TMH1	P1.35	H1.64						
TMH2	N2.40	Y2.67						
TMH3	C3.25	R3.55	3.28	3.55	-3	0	0	-3
TMH4	S4.38	P4.59	4.40		-2		-2	
TMH5	N5.37	R5.67	5.46	5.66	-9	1	1	-9
TMH6	E6.30	V6.59	6.33		-3		-3	
TMH7	L7.35	C7.70	7.39	7.70	-4	0	0	-4

Note the significant agreement with the predictions of the CT ends (average ΔCT=0.3), but the significant deviations for the predicted NT ends (average ΔNT=-4.2) which are consistently longer (i.e. negative sign) than the experimental findings. The largest deviation occur at the NT of H5, where several proposed ligand binding residues (5.42, 5.43) are predicted to lie outside H5. However, the authors of the SCAM study proposed an alternative explanation where the extracellular side of H5 would be highly flexible and still mostly helical. I would favor this hypothesis and thus remove the H5 data from this analysis, which improves significantly the agreement with experimental data (average ΔNT=-3). This implies that the NT predictions are consistently one turn longer than the NTs detected by these Cys-substitution techniques. This deviation can be rationalized based on the orientation of the C^{α}-C^{β} bond towards the N-terminal end of an α-helix, which would allow the side chains of all cysteines substituted for the residues in the first turn of an

α-helix to protrude into the extracellular environment where they could react with the Cys-specific reagents. Such an orientation-based reactivity would explain the lack of α-helical periodicity in the observed accessibility pattern at the predicted first turn of the helix, but would also imply that this effect should not be present at the CT ends, thus consistent with the remarkable agreement observed at the TMH CTs. The structural basis of this argument is shown in Figure 5.11 on a standard poly-Ala α-helix. The effect of the orientation of the C^{α}-C^{β} vector in the solvent accessibility for Cys residues at the C-terminus of an α-helix would be accentuated by the known tendency of Cys residues to H-bond back to the backbone carbonyl i-3 or i-4 from the preceding turn, orienting them further into the helix (i.e. the membrane) and farther away from the more solvent-exposed C-terminus.

It is worth pointing out that the TMH boundaries predicted for the rhodopsin GPCR based on a large body of biochemical data (Findlay 1984) are closer to the TMH ends proposed here than those proposed in the literature for any other GPCR; this is especially true for the cytoplasmic boundaries which include the Arg/Lys as proposed before, the short loop connecting TMH1 and TMH2, and the CT of TMH3, all of which are original predictions in this thesis.

Figure 5.1

Predicted Hx boundaries for the 5HT-1c GPCR
Deviations from previous boundaries, in excess or defect, are shown shaded.
Highly conserved AA are shown in bold letters

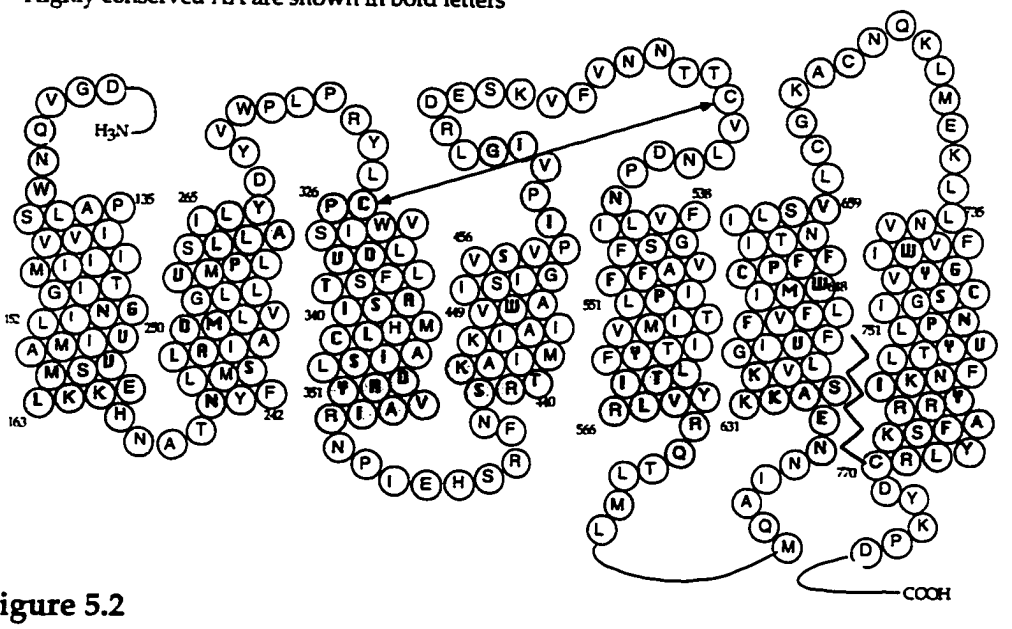


Figure 5.2

Prediction of Hx boundaries by the presence of insertions or deletions within the alignment (shaded AA). Highly conserved AA (cons. index < 0.3) are shown in bold letters.

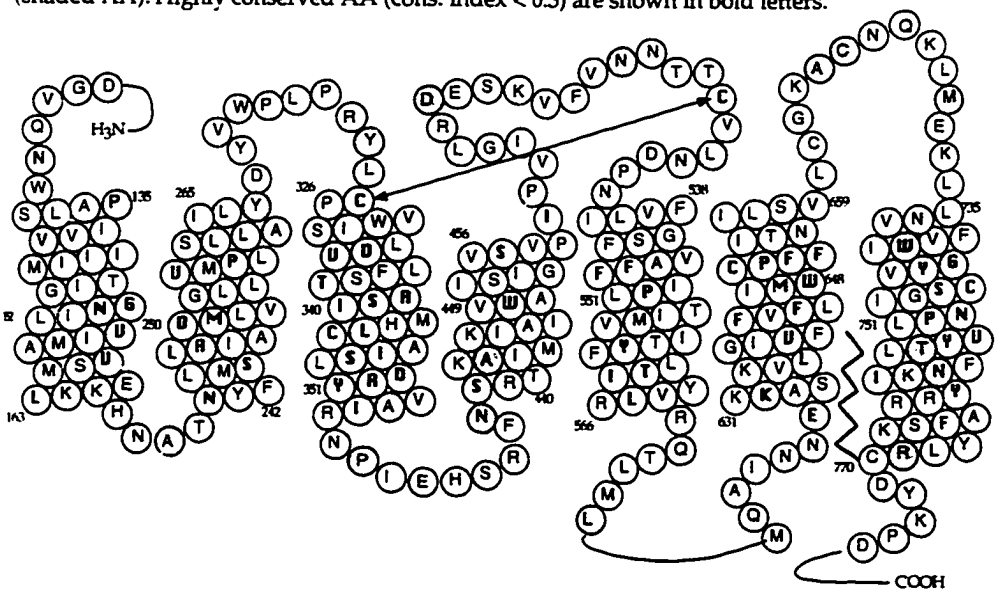


Figure 5.3

Prediction of Hx boundaries by the Arg/Lys cytoplasmic motif. Shaded AA represent position where there are R/K in any GPCR within the alignment. They are superimposed on to the 5HT1c Hx nets in order to look for evolutionary consistency of the predictions. The Arg/Lys residues present in the 5HT1c are marked with a thicker circle. Highly conserved AA (cons. index < 0.3) are shown in bold letters.

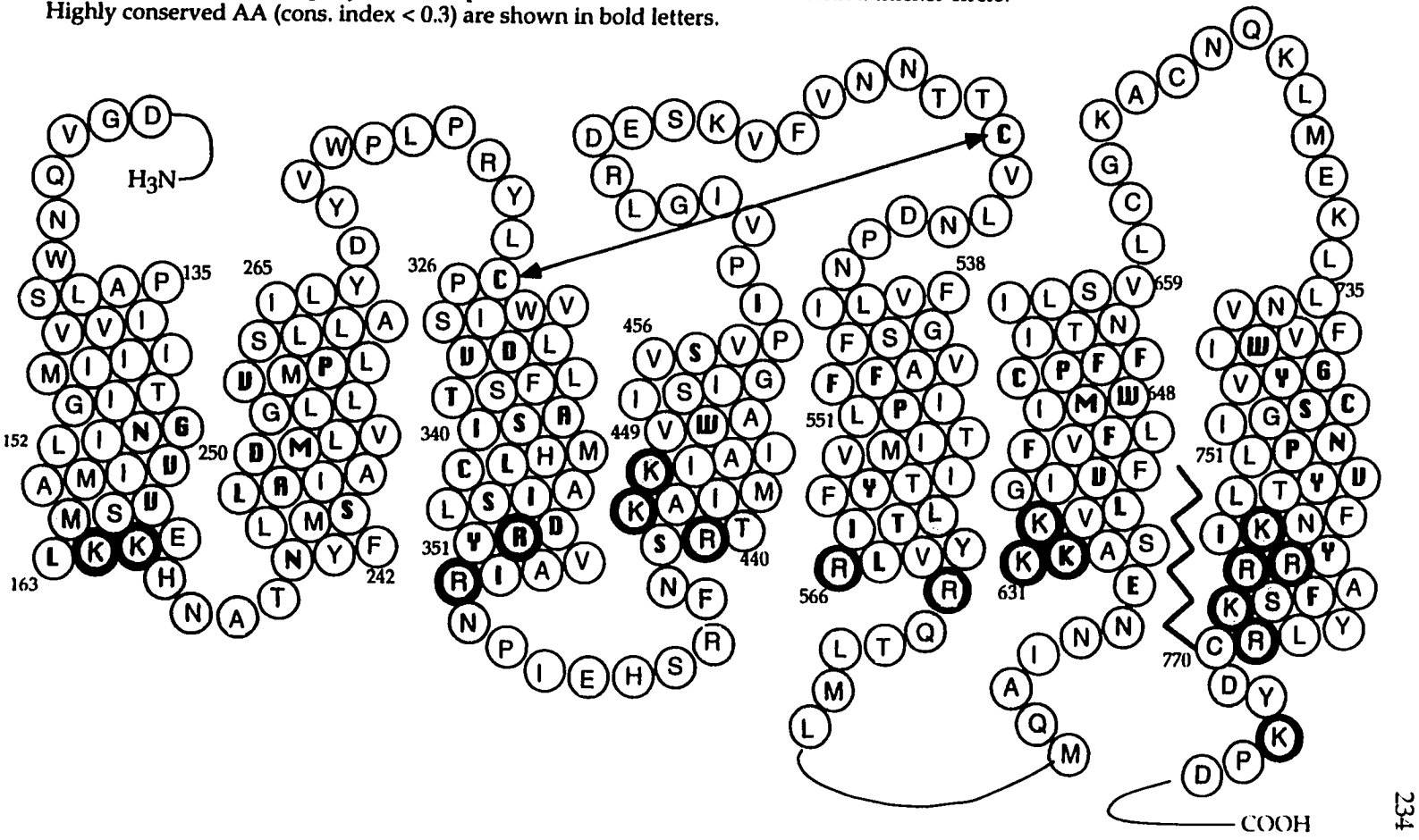


Figure 5.4

Hx nets representation of the conserved AA sites by the conservation index (shaded AA). Highly conserved AA (cons. index < 0.3) are shown in bold letters.

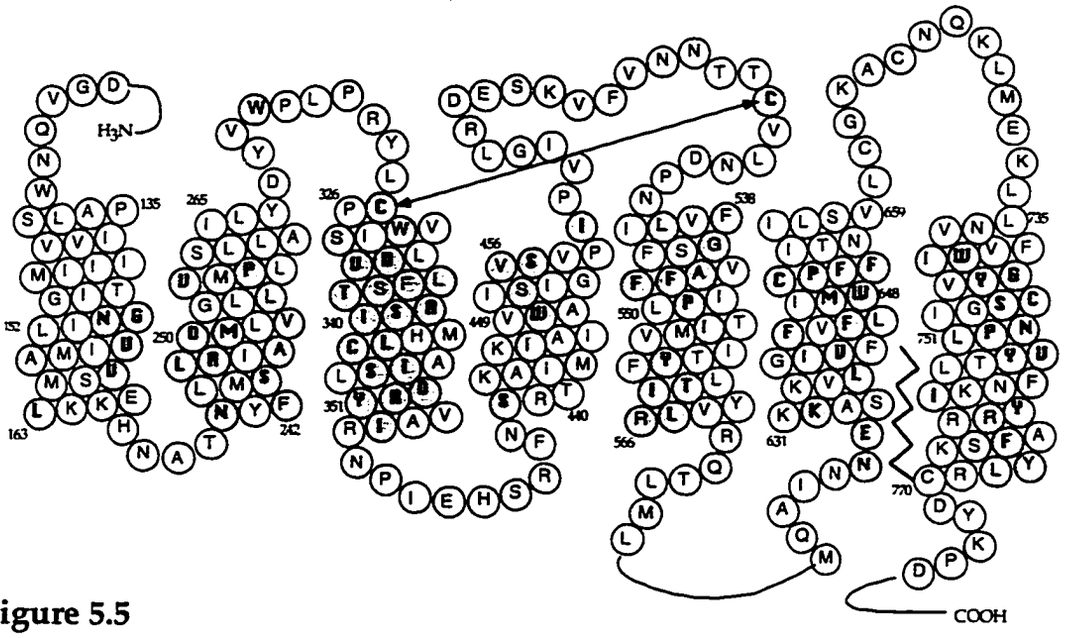


Figure 5.5

Hx nets representation of the conserved AA sites in terms of the volume (shaded AA). Highly conserved AA (cons. index < 0.3) are shown in bold letters.

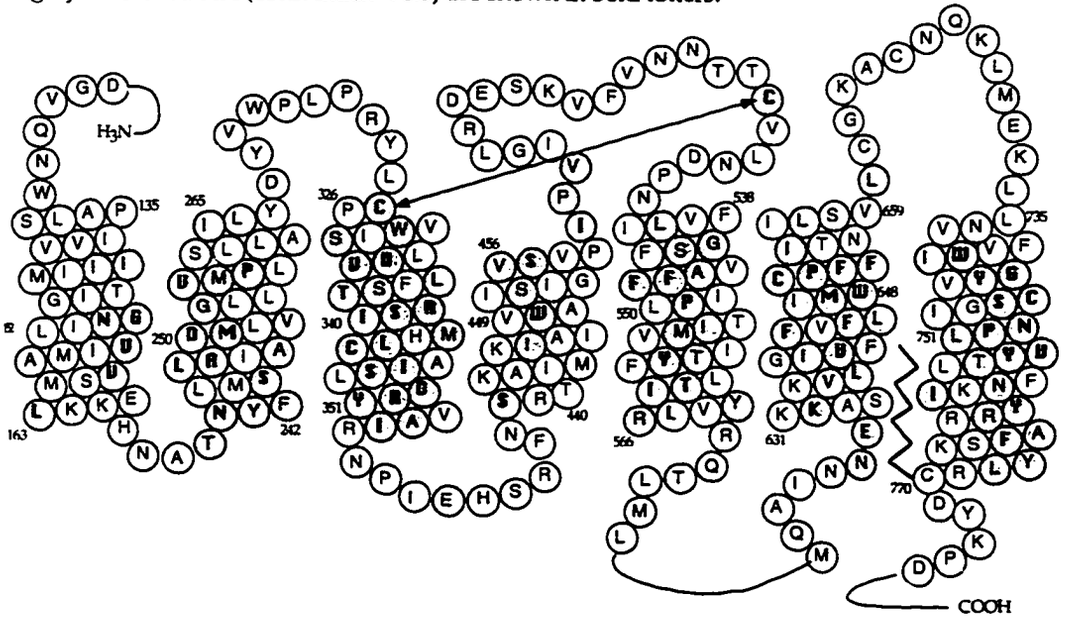


Figure 5.6

Hx nets representation of the conserved AA sites in terms of the hydrophobicity (shaded AA). Highly conserved AA (cons. index < 0.3) are shown in bold letters.

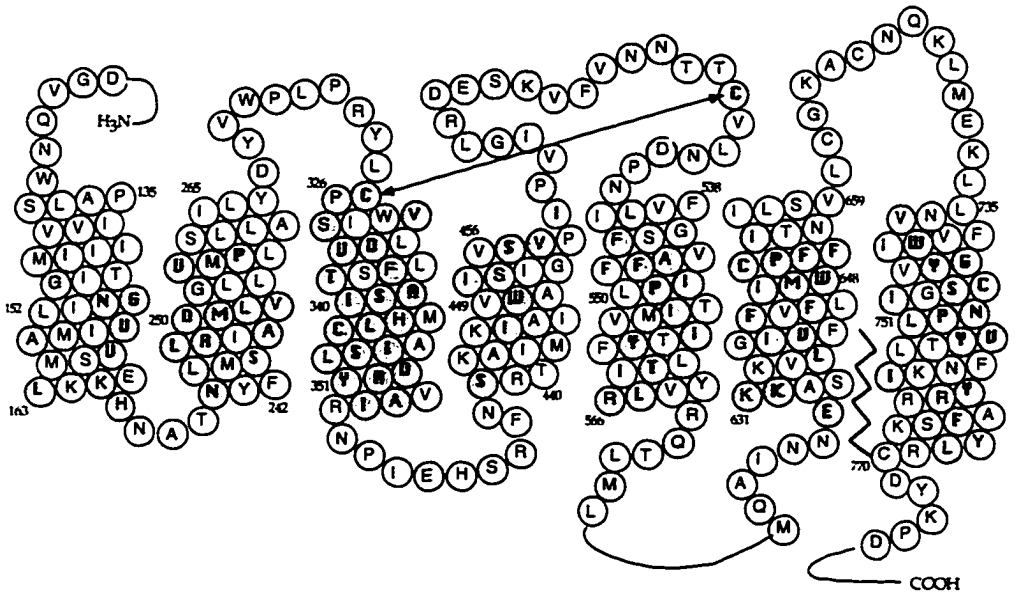


Figure 5.7

Hx nets representation of the conserved AA sites in terms of the H-bonding capabilities(shaded AA), including charged residues (shaded AA with thicker circle). Highly conserved AA (cons. index < 0.3) are shown in bold letters.

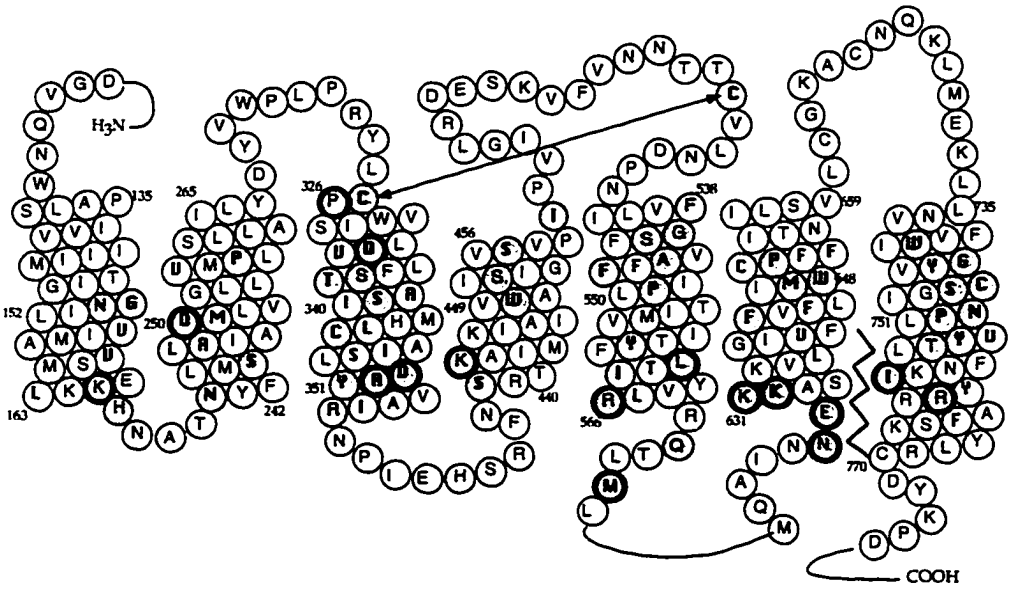


Figure 5.8

Hx nets representation of the conserved AA sites in terms of the aromatic character (shaded AA), or in terms of the β-branched character (thicker circle). Highly conserved AA (cons. index < 0.3) are shown in bold letters.

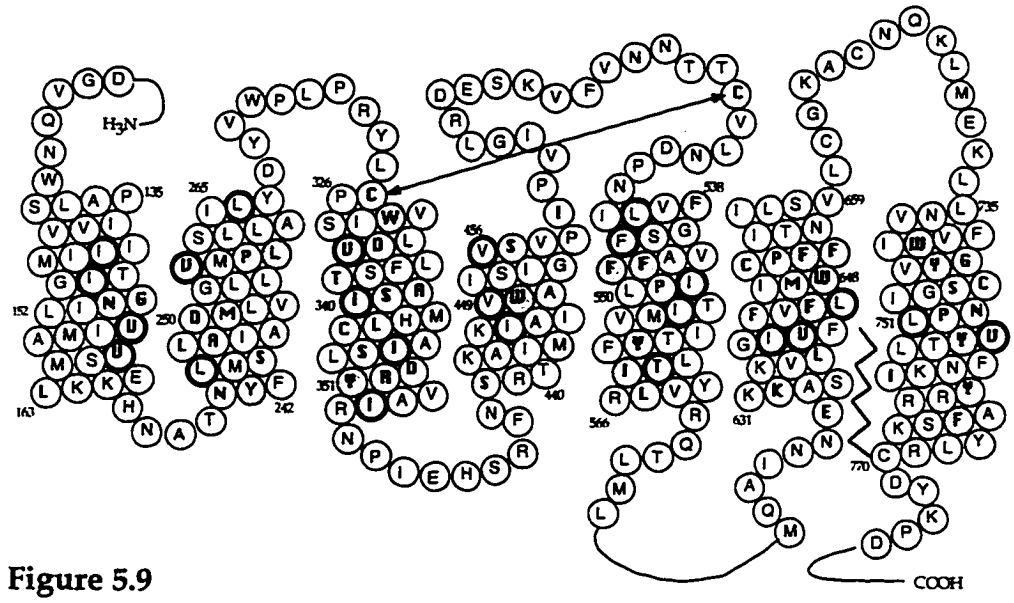
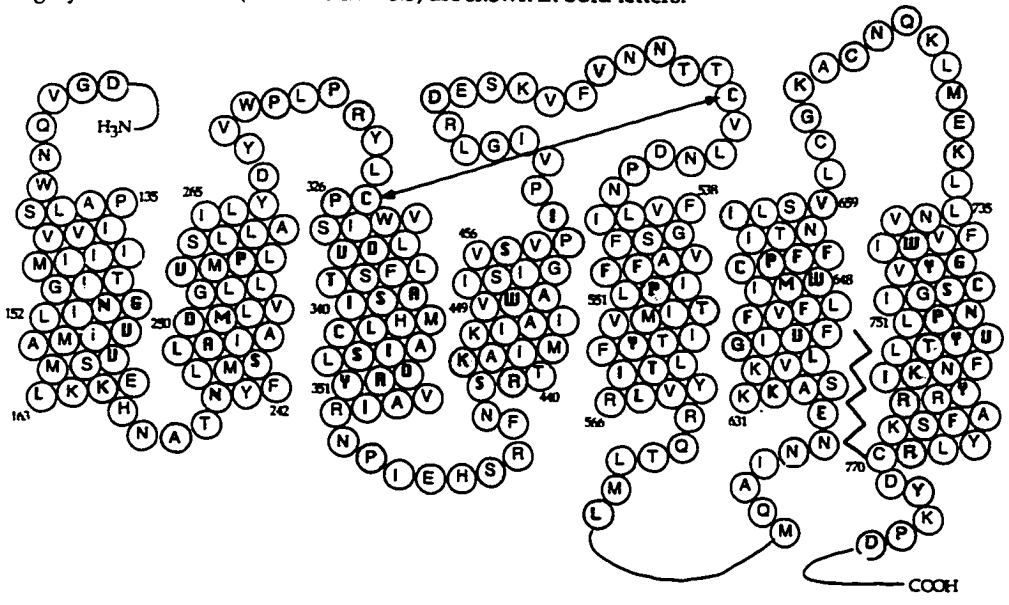


Figure 5.9

Prediction of Hx boundaries by the presence of Pro residues within the alignment (shaded AA). Highly conserved AA (cons. index < 0.3) are shown in bold letters.



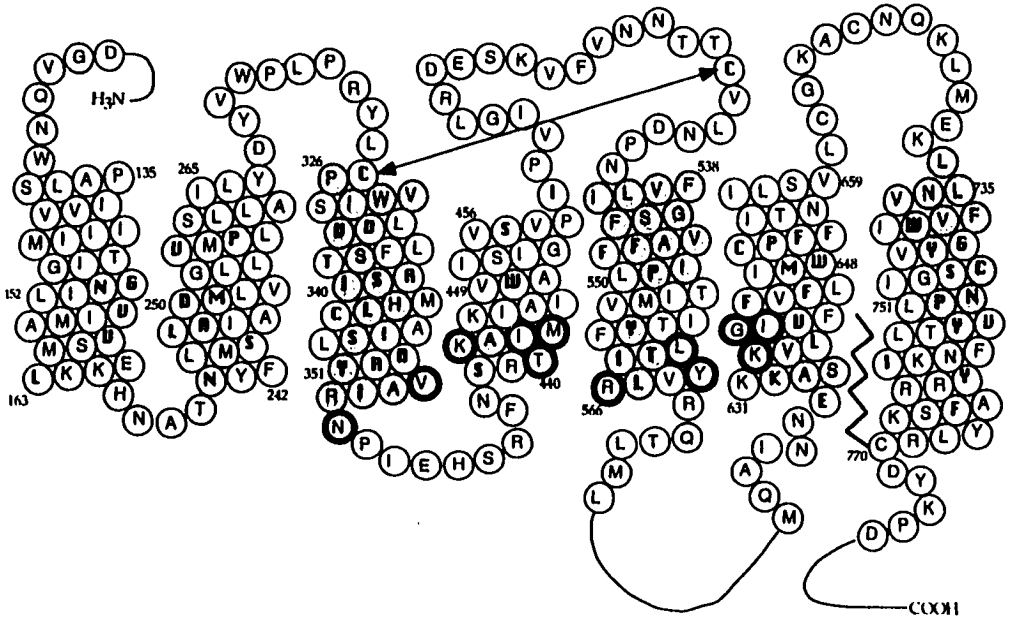


Figure 5.10

Consistency of the predicted TMH boundaries and orientations for the 5HT2CR with experimental data as follows:
Residues accessible from the binding site crevice by SCAM techniques (3.44-3.47, 5.38-5.58, 7.34-7.53) are shown shaded. Residues identified by Spin labeling studies at the cytoplasmic side (3.53-3.56, 4.40-4.44, 5.60-5.66, 6.33-6.39) to reside in TMH-TMH (shaded) or TMH-Lipid (thicker circle) interfaces. Highly conserved AA (cons. index < 0.3) are shown in bold letters.

Figure 5.11

The effect of the orientation of the Cα-Cβ vector towards the N-terminus of an α helix on the solvent accessibility for Cys at the first (top) or last (bottom) turn of the α-helix, shown on a standard poly-Ala α-helix.

6.- ORIENTATION OF THE 7 TMHs WITHIN THE TMH DOMAIN.

6.1.- Prediction of TMH-TMH versus TMH-Lipid interfaces: orientation for each TMHs.

The criteria used to predict whether an AA is oriented towards the TMH-TMH or the TMH-Lipid interfaces relies on the same principle of pursuing a consistent α-helical periodicity in the conservation pattern that was used above to predict the TMH boundaries. Thus, the same figures and common arguments are applied in this section; i.e. AA sites which are evolutionarily constrained are predicted to be at the TMH-TMH interfaces. Thus, the analysis performed in the previous section is now revised and considered together with novel approaches pertinent to tertiary structure analyses. I apply first criteria for the prediction of AAs in TMH-TMH interfaces (Section 6.1.1), followed by the equivalent approaches to predict AAs in TMH-Lipid interfaces (Section 6.1.2). However, there are residues whose properties preclude a specific prediction in terms of interior or lipid-exposed positions. Therefore, the predictions derived in Sections 6.1.1 and 6.1.2 are complemented in Section 6.1.3 by the analysis of the alternative positioning of {Pro, Ser, Thr, Cys, Gly} residues in TMH-TMH or TMH-Lipid interfaces. The predictions derived independently in these three sections are then integrated into a final prediction of the TMH-TMH and TMH-Lipid interfaces for the 5HT2CR in Section 6.1.4.

6.1.1.- Prediction of AA in TMH-TMH interfaces.

AA sites identified by the criteria of the conservation index, volume, H-bonding, and aromatic or β-branched character are predicted to lie in TMH-TMH environments (see Figures 5.4, 5.5, 5.7, 5.8). Conservation of hdp or charge could be predictive of TMH-TMH or TMH-Lipid environments (e.g. a highly conserved Leu, Arg, or Asp), and thus these criteria require additional considerations described below.

Sites conserving a negatively charged character have been shown to localize in TMH-TMH interfaces for TMH domains (Ballesteros 1992a; Bormann 1992). The AA sites which conserve a negatively charged character are listed below by their residue identifier, most conserved amino acid, and residue present at each equivalent position in the 5HT2CR (in parenthesis).

```
2.50-D(D)     D    - 100.0 %
3.26-D(P)          -  89.0 %
3.32-D(D)     D    - 100.0 %
3.49-D(D)     D    - 100.0 %
6.30-E(E)     E    - 100.0 %
7.59-D(I)          -  99.0 %
```

These residues are predicted to lie in TMH-TMH interfaces. Note that there are 6 highly conserved acidic groups, of which two are absent in the 5HT2CR (P3.26 and I7.59), which were not included in the conservation profile.

Sites conserving a positively charged character (Arg/Lys) at the cytoplasmic boundaries could be facing the phospholipid head-groups, as

mentioned before (Section 2.1), and will thus be considered below for the prediction of TMH-Lipid interfaces. However, positions preserving a very high positive charge character could be predicted in either interior or lipid-exposed environments by the conservation or Arg/Lys motif, respectively. The positions where the positive charge character are more than 90% conserved are listed below:

```
3.50-R(R)      + 100 % Arg
5.60-R(L)      +  93 % Arg/Lys
6.32-K(K)      +  99 % Arg/Lys
```

Arg3.50 is absolutely conserved in its amino acid character and thus predicted in TMH-TMH interfaces. The remaining residues interconvert between Arg and Lys, and although they could be facing the protein interior, I would not change their current prediction based on the Arg/Lys motif but I will take on these alternative orientations when their position is considered below in the structural context of the three-dimensional model of the TMH domain.

The conservation of the hdp character (low standard deviation hdp) could be due to highly hydrophobic residues -i.e. high hdp mean value- facing the lipid milieu. Those sites would be identified below for the prediction of TMH-Lipid interfaces. On the other hand, AA sites conserving the hdp character yet having a low hdp mean value -highly hydrophilic- would be predicted to be in TMH-TMH interfaces. The identification of those sites is complicated by two factors. First, the low mean hdp could be due to the

presence of Arg/Lys (hdp=-4.5,-3.9) predicted to face the phospholipid head-groups, thus vitiating the analysis. Second, Ser and Thr residues are hydrophilic (hdp=-0.8, -0.7) although they are not an indication of TMH-TMH environments as explained below (Section 2.1). Therefore, the residues invariably predicted to be in TMH-TMH interfaces are {His, Glu, Gln, Asp, Asn} whose hdp values range from -3.2 to -3.5. I have therefore selected the AA sites with σhdp<1.2 and mean hdp within the range (-3.0, -3.6), excluding Arg and/or Lys residues. Those selected AA sites, listed below, are predicted to lie in TMH-TMH interfaces. It should be noted that AA sites conserving an hydrophilic character have already been discussed above if they conserve the charge or the H-bonding capabilities.

naa	mean hdp	σ_{hdp}	cons. index.
1.50-N(N)	-3.50	0.00	0.00
2.40-N(N)	-3.35	0.15	0.04
2.50-D(D)	-3.50	0.00	0.00
3.32-D(D)	-3.50	0.00	0.00
3.49-D(D)	-3.50	0.00	0.00
5.35-N(D)	-3.00	0.99	1.00
6.30-E(E)	-3.50	0.00	0.00
7.45-N(C)	-3.50	0.00	0.00
7.49-N(N)	-3.50	0.00	0.00
7.59-D(I)	-3.50	0.00	0.11

The final selection of AA sites predicted to lie in a TMH-TMH environment is shown in Figure 6.1 as shaded circles. It contains all residues identified above using the various conservation criteria. The degree to which a given AA site is predicted to be in the TMH-TMH environment depends on

the degree of the conservation of the property(s) used in its selection. The relative strength of each prediction is not shown in the figures presented in this section, with exception of the conservation index. Instead, I refer to the corresponding figures in Section 5 for this purpose, where the degree of the conservation of each property is individually quantified for each property at each AA site.

The individual AA sites predicted to face TMH-TMH environments are expected to define the TMH-TMH interfaces of each TMH. Therefore, they should lie within a common face of each TMH defining a patch on the TMH surface that would be predicted to constitute a TMH-TMH interface. This hypothesis can be explored by analyzing the position of those AA sites in a helical net representation. Such representation is shown in Figure 6.1 where the predicted interior-facing AA sites are in fact within a common face of each TMH. However, there are cases where entire turns of a TMH are predicted to face the interior of the TMH bundle -e.g. TMH6. This would imply a completely buried TMH, i.e. a TMH without contact with the lipids, or a revision of the predictions, a decision that should await comparison with the predicted TMH-Lipid interfaces and the subsequent integration of these different criteria, described in Section 6.1.4.

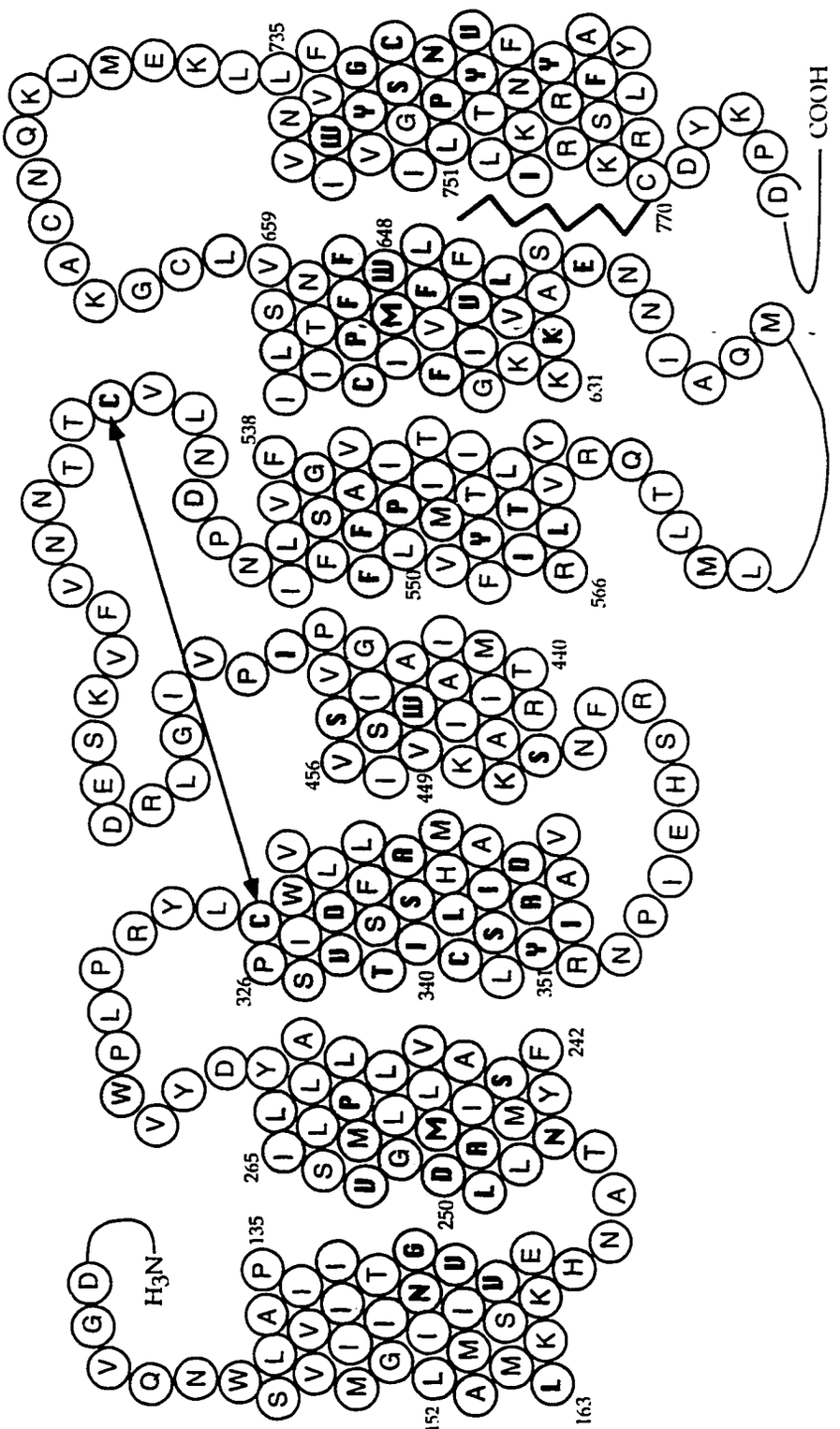


Figure 6.1 :

Prediction of AA sites in TMH-TMH interfaces (shaded AA). Highly conserved AA (cons. index < 0.3) are shown in bold letters.

6.1.2.- Prediction of AA in TMH-Lipid interfaces.

The lipid membrane contains two physico-chemically distinct phases, a very hydrophobic core composed by lipid side chains that spans ±40 Å, and the polar phospholipid headgroup region, highly polar in the extracellular side and mostly negatively charged in the interior monolayer, which spans about 15 Å (see Section 2.1). Thus, different criteria are applied for the prediction of AAs in TMH-Lipid interfaces for each one of these regions. Section 6.1.2.1 presents the prediction of AAs facing the hydrophobic core of the membrane, and Section 6.1.2.2 presents the prediction of AAs facing the phospholipid headgroups at the cytoplasmic boundaries (the Arg/Lys motif). These two criteria, applied independently, are integrated in Section 6.1.3 resulting in the predicted TMH-TMH and TMH-Lipid interfaces for the 5HT2CR. Note that there is no proposed criterion for the prediction of AAs facing the lipid headgroups at the extracellular side of the membrane, as discussed in Section 2.1.

6.1.2.1.- Prediction of AA facing the hydrophobic core (Lipid chains): Quantification by the probability function P(TMH-Lipid).

The energy-based argument that highly hydrophobic AA are likely to face the lipid milieu, was coupled with the low conservation profile of surface-exposed residues to predict AA sites facing the lipid chains. This criterion was quantified in Section 2.3.2 based on a probability function

P(TMH-Lipid), whose satisfactory application on the Photosynthetic Reaction center (PRC) suggested a threshold value of 6 for prediction purposes (see Figures 2.18 and 2.19). A similar analysis performed on the multiple sequence alignment of GPCR discussed in Section 5.1 is presented below in Figure 6.2.

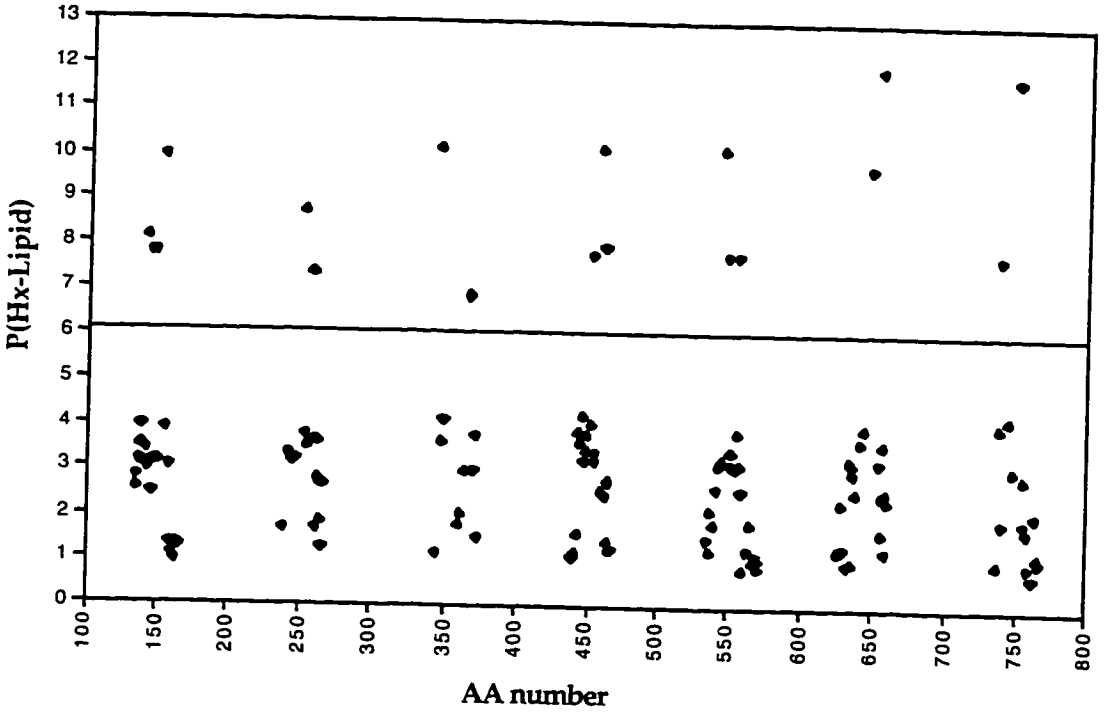


Figure 6.2:
The probability function P(TMH-Lipid) predicts as lipid-facing residues those above the threshold value of 6.

The AA sites predicted to lie in TMH-Lipid interfaces are thus defined as those with a probability value greater than 6. They are listed below:

# aa	mean hdp	σ_{hdp}	CI	P(TMH-Lipid)
1.42	3.44	0.96	0.98	8.13
1.44	3.42	0.99	0.98	7.83
1.48	3.42	0.99	0.98	7.84
1.54	3.75	0.76	0.92	9.94
2.52	3.04	0.84	0.98	8.74
2.58	3.60	1.01	0.92	7.38
3.27	3.56	0.78	0.98	10.15
3.48	3.58	1.05	0.90	6.90
4.52	3.42	0.99	0.98	7.83
4.58	3.56	0.78	0.98	10.17
4.61	3.60	1.01	1.00	8.00
5.44	3.56	0.78	0.98	10.14
5.49	3.42	0.99	0.98	7.83
5.56	3.42	0.99	0.98	7.83
6.46	3.75	0.76	0.90	9.76
6.53	3.83	0.64	0.92	11.98
7.37	3.42	0.99	0.98	7.84
7.48	3.83	0.64	0.91	11.80

The individual AA sites predicted to face the lipid milieu are expected to define the TMH-Lipid interface of each TMH. Therefore, they should lie within a common face of each TMH defining a patch on the TMH surface that would be predicted to be the TMH-Lipid interface. This hypothesis can be explored by analyzing the position of those AA sites in a helical net representation. Such a representation is shown in Figure 6.3 where the predicted lipid-facing AA sites are in fact within a common face of each TMH. This can be seen in TMH6 for I6.46 and I6.53, separated by 7 AA which is the maximal azimuthal equivalence in a TMH. (In the inspection of Figure 6.3, remember that whenever an AA site lies at the beginning or end of a given turn in an helical net, it can be repositioned at the end of the last turn or the beginning of the next turn respectively, due to the helical symmetry). Notably, there is a remarkable convergence between these predictions with

other criterion for the prediction of TMH-TMH versus TMH-Lipid interfaces, as discussed below in the integration of the results.

6.1.2.2.- Arg/Lys at cytoplasmic boundaries are proposed to face the phospholipid head-groups.

Arg and Lys residues found in AA sites within the GPCR alignment, with special attention to those present in the 5HT2CR, are shown in Figure 6.3 on an helical net as shaded circles; the R/K present in the 5HT2CR are marked with a thicker circle. Those sets of residues were discussed in Section 5.3 in connection to the prediction of the TMH boundaries. I shall note here the helical periodicity of their distribution. As mentioned above for lipid-chain facing AA, the individual AA sites predicted by the R/K motif to face the lipid head-groups should lie within the common face of each TMH that defines a patch predicted to be the TMH-Lipid interface. This can be seen in H6 for K6.31, K6.32 and K6.35, all clustered on one face of H6. Note also the high density of Arg/Lys residues present at the predicted cytoplasmic extension of H7, and the absence of any Arg/Lys in the entire alignment at the cytoplasmic boundaries of H2.

At this point we can integrate the results of the predicted TMH-Lipid AA sites, using the criteria of the R/K or the highly variable yet highly hdp sites. They should define essentially identical TMH-Lipid interfaces, and the degree of convergence of their predictions can serve as an estimate of their

reliability. Both sets of predictions are shown together in Figure 6.3, where a remarkable agreement between the prediction tools is observed. Particularly, in TMH6 I6.46 and I6.53 clearly belong to the same TMH patch defined by K6.31, K6.32 and K6.35.

6.1.3.- Alternative positioning of {Pro, Ser, Thr, Cys, Gly} residues in TMH-TMH or TMH-Lipid interfaces.

Conserved Pro residues are an exception of the criterion that conserved residues face the interior of the TMH bundle, as mentioned in Section 2.1, because they can also exert their structural role from a lipid-facing position. The polar character of {Ser, Thr, Cys, Gly} residues would be considered an indication of TMH-TMH interfaces. However, the ability of {Ser, Thr, Cys} to satisfy their H-bonding requirements by H-bonding back to the backbone carbonyls allows them to be exposed to the lipid environment (Gray 1984), as found for known membrane protein structures (Section 2.1); consequently, their presence in the lipid-facing environments should not be disallowed. Glycine residues have also been found to partition equally well into TMH-TMH and TMH-Lipid interfaces. Thus, I have explored the presence of conserved Pro, {S,T,C} and Gly residues in the TMH domain, with special attention to those AA sites located within the TMH patches predicted to be the TMH-Lipid interfaces. The results are presented below and integrated with the previous predictions of TMH-TMH and TMH-Lipid interfaces in

Section 6.1.4, below.

Presence of conserved Pro: Conserved proline residues appear at 2.59, 5.50, 6.50, and 7.50. Of these, P5.50 and P7.50 belong to helical patches of other highly conserved residues and are thus predicted in TMH-TMH interfaces. The conserved P2.59 belongs to the TMH-Lipid face of H2 defined by L2.53 and M2.58. The P6.50 belongs to a stretch of 6 highly conserved residues in H6 (M6.47-W6.48-C6.49-P6.50-F6.51-F6.52), thus all predicted to face the protein interior. However, P6.50 belongs to the surface patch defined by I6.46 and I6.53 predicted to face the lipids by the P(TMH-Lipid) function in Section 6.2.1. Thus, the initial prediction of P6.50 facing the protein interior is redefined, resulting in a continuous TMH-Lipid face at this segment of H6 defined by I6.46-P6.50-I6.53.

Presence of Gly: There are 4 AA sites in the TMH domain where a Gly is present in more than 50 % of the receptors: G1.49 (78 %), G7.42 (100 %), G6.38 (51 %) and G6.42 (76 %), shown on a helical net in Figure 6.4. The first three, G1.49, G7.42, and G6.42 are highly conserved and predicted in TMH-TMH interfaces. Note that residues G6.38 and 6.42 are (i,i+4) within the same helical turn in H6, thus belonging to the same face of H6 which suggests that G6.38 is also in TMH-TMH interfaces.

Presence of {Ser, Thr, Cys}: There are several AA sites in the TMH domain where a Ser, a Thr or a Cys amino acid is present in more than 50 % of the receptors, listed in Table 6.1 and shown on a helical net in Figure 6.4. Note that H3 shows an unusual concentration of Ser/Cys/Thr amino acids (9 residues), in sharp contrast to other TMHs which contain only 2-3 cases. The highly conserved residues (>80%) within the set of {Ser, Thr, Cys} would be predicted to face the interior of the TMH bundle; these residues, shown in bold in Table 6.1, are S2.45 in H2, C3.25-T3.37-S3.39-C3.44 in H3, 4.38-4.53-4.57 in H4, 5.42-5.43-546 in H5, C6.47 in H6, and S7.46 in H7. Note that 5.42-5.43-5.46 define a cluster on the H5 surface proposed as direct ligand binding sites, reviewed in Section 2.2. The conserved Ser/Thr/Cys residues in H3 appear on opposite faces of the helix, an inconsistency with helical periodicity that would become more apparent in the integration of all the previous predictions of TMH-TMH and TMH-Lipid interfaces presented in the next section.

Table 6.1 :

Occurrence of Ser/Thr/Cys residues within the TMH domain of neurotransmitter GPCRs. Positions were Ser/Thr/Cys residues are present in more than 50% of the receptors aligned (see Section 4.1) are shown for each TMH. The residue identifier (ID), the most commonly present amino acid (AA) at this loci, and the % population of these three amino acids combined is shown. Highly conserved residues (>80%) are shaded and shown in bold. Note that H3 is significantly overpopulated in these residues relative to other TMHs.

H1	H2	H3	H4	H5	H6	H7
ID AA %	ID AA %	ID AA %	ID AA %	ID AA %	ID AA %	ID AA %
1.40 T 51	2.45 S 90	3.25 C 100	4.38 T 99	5.42 S 93	6.36 T 62	7.46 S 100
		3.30 S 61	4.53 S 81	5.43 S 97	6.47 C 100	7.54 T 74
		3.35 C 60	4.57 S 100	5.46 S 80		7.70 C 73
		3.36 C 70				
		3.37 T 100				
		3.39 S 100				
		3.44 C 90				
		3.47 S 50				
		3.55 T 70				

6.1.4.- Integration of the predictions of TMH-TMH and TMH-Lipid environments.

The predictions resulting form the application of three sets of criteria described in the previous three sections, and shown in Figures 6.1, 6.3, and 6.4, are shown together in Figure 6.5 where they can now be integrated to derive the final predictions for AA sites defining the TMH-TMH and TMH-Lipid interfaces for the seven helices. Note the general consistency among these different criteria. However, some significant inconsistencies appear; most of those involve residues predicted to face both environments simultaneously. This pattern is certainly feasible on a three-dimensional structure where many residues are often partially exposed (see quantification of solvent accessibility for the model presented in Section 8), and has also

been observed experimentally in the photoactivated labeling studies in rhodopsin (Davison 1986a; Davison 1986b). However, for modeling purposes it is more helpful to restrain the predictions to only two environments, where residues are predicted either in TMH-TMH or TMH-Lipid interfaces. It should be noted that whenever these prediction methods based on sequence analyses have been tested on the known structure of the PRC ((Deisenhofer 1989) and Section 2.3.2), the aligned sequences were significantly more divergent than the set of GPCR selected for this work (compare alignments for PRCs in Table 2.5 and GPCRs in Table 4.1). This implies that relative to the test performed for these methodologies on the PRC, the significantly higher homology among the GPCRs analyzed overemphasizes the conservation pattern, thus favoring the overprediction of TMH-TMH interfaces and the misprediction of TMH-Lipid interfaces. This can be illustrated by considering either extreme; applying these analyses to an alignment of the rat and human 5HT2CRs would predict almost all residues buried, while analyzing 200 divergent GPCRs as Baldwin did (Baldwin 1993) would render the criteria for lipid-exposed residues the main tool to predict the orientation of these TMHs. Thus, whenever a residue is predicted in both environments, the prediction of a TMH-Lipid environment will prevail unless otherwise stated.

The most significant inconsistencies in Figure 6.5 pertain to cases where entire turns of a TMH are predicted to face the interior of the TMH bundle, thus eliminating α-helical periodicity -e.g. H3 or H6. This would

require a completely buried TMH turn, or a revision of the predictions. A careful analysis of these predictions based on the relative strength of each of them can identify TMH patches whose predictive strength is significantly weaker. Those TMH patches would thus become putative TMH-Lipid interfaces that can be compared with the TMH-Lipid interfaces identified in Section 6.1.2, pursuing convergence among the results to obtain a predicted continuous TMH-Lipid interface. At this stage the importance of quantifying the predictions described done above in Section 4, can be appreciated. This process is illustrated for H6, where a redefinition of predicted TMH-TMH AA sites results in a continuous TMH-Lipid interface with proposed functional implications, and for H3, where such a redefinition is not possible and thus H3 is predicted to be completely buried from 3.32 to 3.44.

Note from Figure 6.5 that H6 is predicted to be almost completely buried. Among the various criteria presented, the strongest is near 100% conservation of a given AA, that corresponds to a conservation index less than 0.3: these AA sites, shown in bold letters in Figure 6.5 are {E6.30, K6.32, L6.37, V6.40, F6.42, F6.44, M6.47, W6.48, C6.49, P6.50, F6.51, F6.52}. Within this AA subset there are two incompatibilities. First, there are 6 consecutive AA sites (M6.47 to F6.52) which span almost 2 turns of TMH6. However, as discussed in Section 6.1.3, P6.50 can be redefined as a lipid-facing residue where it can exert its structural effects; thus P6.50 forms a continuous TMH patch with I6.46 and I6.53 predicted to face the lipids (see Figure 6.6). This

prediction solves this first inconsistency and predicts that {M6.47, W6.48, C6.49, F6.51, F6.52} define the TMH-TMH interface. This is consistent with the TMH-TMH interface defined by {E6.30, L6.37, V6.40, F6.42, F6.44}. The second inconsistency is K6.32, which lies at the cytoplasmic boundaries and is thus predicted to face the phospholipid head-groups, as mentioned above (see Figure 6.3). Section 6.1.3 analyzed the presence of Gly in predicted TMH-Lipid interfaces (see G6.38 and F6.42 for the 5HT2CR in Figure 6.4). It was found that residues G6.38 and 6.42 are (i,i+4) within the same helical turn in H6, where they belong to the TMH-Lipid interface defined by residues K6.31-K6.32-K6.35-I6.46-I6.53 (see Figure 6.5). This orientation is quite surprising for the highly conserved (76 %) G6.42, thus predicted in TMH-TMH interfaces in Section 6.3 (see Figures 6.10 and 5.4). Groups of glycine residues have been observed on a common face of some TMH (Travers 1984; Bormann 1992; Cosson 1992; Lemmon 1992) and are likely to constitute a characteristic motif for TMH domains (Bormann 1992). The proposed role of such surface patches is to allow for very close TMH-TMH packing requiring a small interhelical angle, thus supporting TMH interactions (Travers 1984; Bormann 1992). This structure-function relationship for Gly-patches in TMH proteins has been recently supported by their observed colocalization with Ala residues, which facilitates close interhelical contacts by their small size (Travers 1984; Bormann 1992) (position 6.38 is 51 % Gly and 28 % Ala). In fact Gly-patches do occur in the

experimentally resolved TMH protein structures (Bacteriorhodopsin and PRC). However, it is clear in those figures that Gly-patches are exposed to the hydrophobic lipid phase, not involved in TMH-TMH interactions, at least in the monomeric state. This suggest that Gly-patches might code for oligomeric contacts between TMH domains. Thus, the inconsistency found for the prediction of G6.42 in a TMH-TMH interface based on its high conservation, yet positioned within a patch of residues oriented towards the lipids, could be resolved if G6.38 and G6.42 would belong to an oligomerization interface. Recently, this hypothesis has been supported experimentally for the β2-adrenergic receptor (Hebert 1996). A bulky Phe residues present at position 6.42 in the 5HT2CR is inconsistent with the previously characterized oligomerization domains described above. Thus, receptor oligomerization will not be addressed in this work for the 5HT2CR, but evidence about receptor dimerization in the adrenergic receptor can be used to explain the inconsistencies found for the orientation of position 6.42 between TMH-TMH and TMH-Lipid interfaces. The redefinition of the predicted TMH-TMH and TMH-Lipid interfaces for H6 that results from the analyses described above is presented in Figure 6.6.

The other TMH where several consecutive turns are predicted as buried is H3, from D3.32 to C3.44 (see Figure 6.5). Residue L3.34 can be redefined to a TMH-Lipid interface because is significantly less conserved than others and belongs to the lipid-facing patch of V3.27 and the unassigned

L3.31. However, the next turn of H3 is formed by 4 highly conserved residues, indicating that H3 at this level would be completely buried. The next residue in the sequence is position 3.41 where a Glu is present in β2-adrenergic receptors; because H3 at this level would be surrounded by the hydrophobic lipid chains, residue 3.41 should be oriented towards the interior of the TMH bundle. A similar orientation should be required for the highly polar H3.42 that follows M3.41 in the sequence. But M3.41-H3.42 are followed by two 100% conserved residues: L3.43 and C3.44. Thus the H3 segment between F3.35 and C3.44, comprising 10 continuous residues, is predicted to be completely buried within the TMH domain. The tertiary environment predicted for the remaining turns of H3 follows a consistent pattern of TMH TMH and TMH-Lipid interfaces, and the resulting prediction for the entire helix is shown in Figure 6.6.

The pattern of predicted TMH-TMH and TMH-Lipid interfaces for the remaining TMHs is much more consistent than the H3 and H6 cases described above, and their integration leads to the final predictions for the seven TMHs presented in Figure 6.6. Few positions present inconsistencies in their predicted pattern that deserve commentary. Only Phe and Tyr amino acids are present at F5.48, thus 100 % conserved in its aromatic properties, and yet this position is clearly offset from the patch of residues defining the TMH-TMH interface for H5. Thus, F5.48 is predicted in the TMH-Lipid interface, an inconsistency explained in Section 7.4 in the context of a

three-dimensional model of the 5HT2CR TMH domain, where the long and bulky Phe side chain can adopt a conformation from which it can reside mostly in the membrane phase yet participate in specific aromatic-aromatic interactions with other conserved residues. It is worth pointing out that several Arg/Lys predicted here to face the phospholipid headgroups can be highly conserved, and thus were also predicted in TMH-TMH interfaces (see Figure 6.5). Those with conservation profiles higher than 80% may have been incorrectly predicted towards the lipid face. However, based on the significant number of such cases (see Section 6.1.1 and Table 4.2) and the important electrostatic effect of positioning an Arg/Lys inside the TMH bundle, these conserved Arg/Lys positions, except the 100% conserved R3.50, are all predicted outside the protein interior at this stage and will be reconsidered later (Section 7.4) in the context of a detailed three-dimensional model. As an example, note that residues E6.30 in H6 and E1.60 in H1 of the 5HT2CR define a shrinking of the patch of interior facing residues at the cytoplasmic boundaries from two amino acids per turn to one single residue predicted inside the TMH bundle. However, these two acidic groups are next to positions whose Arg/Lys content is quite conserved (83 % for K6.31 and 76 % for K1.61). Although there is no basis at this structural level of analysis to predict that K6.31 and K1.61 are facing the interior of the TMH bundle, during the process of detailed helix packing at atomic level specific interactions within these charged pairs will be proposed that would redefine their lipid

exposure predicted here. This illustrates the interrelated role of the different levels of structural analyses applied in this work.

Figure 6.6 shows the final predictions of AA sites in TMH-TMH or TMH-Lipid interfaces for those residues for which a significant prediction could be made based on the conservation analysis presented above. This explains the presence of discontinuous patches of residues predicted in either environment in Figure 6.6. However, a more clear and thus useful pattern of interior versus lipid exposed residues can be achieved by expanding the set of AA sites for which a specific prediction exists, deriving also a prediction for the remaining unassigned residues in the 7 TMHs. The criteria for such a derivation are the same as used earlier, though now based on a lower statistical significance, and by the rule that unassigned AAs which lie within surface patches that have a predicted orientation towards TMH-TMH or TMH-Lipid interfaces, are assigned to such a patch. The continuous TMH-TMH and TMH-Lipid interfaces derived based on this criteria are shown in Figure 6.7. This representation will be used in the subsequent sections for illustrative and modeling purposes. Note that the predicted hinge in H7 at the turn T7.54-N7.57 is now incorporated to this representation, once the consistency in the relative orientation between the helical faces of H7 and its cytoplasmic extension has been documented by in the previous helical nets.

6.1.5- Comparison of predicted TMH orientations with experimental data: SCAM and Spin labeling studies.

The experimental techniques of SCAM (Javitch 1995a; Javitch 1995b; Fu 1996) (D2 receptor) and spin labeling (Farahbakhsh 1995; Altenbach 1996; Farrens 1996; Yang 1996a; Yang 1996b) (rhodopsin) of Cys mutants at consecutive residues, described in Section 5.8, have determined the accessibility and environment of several residues within H3, H4, H5, H6, and H7. These results were shown together in Figure 5.10 on helical nets representing the 5HT2CR with the TMH boundaries predicted here, where their observed localization in TMH-TMH or TMH-Lipid interfaces can be compared with the equivalent predictions shown in Figure 6.7. Note the remarkable agreement if the extracellular portion of H5 is excluded from the analysis, as discussed in Section 5.8. The agreement with the experimental data is especially significant in two segments. First, the turn 4.40 to 4.44 at the cytoplasmic side of H4 validates the prediction of a stretch of single residues at positions (i, i+4) facing the interior of the TMH bundle surrounded by residues exposed to the lipid milieu, a prediction based on the Arg/Lys motif discovered in this thesis. In fact, the agreement of these experimental results with predictions derived from the Arg/Lys motif, present in H4, H5, and H6, is exact for all these helices, thus validating this novel approach to the model TMH proteins. Second, note the irregular pattern of residues accessible to the ligand binding crevice in H7, which not only are in agreement with the

experimental results but were actually used to aid in the interpretation of this SCAM results in a collaborative effort with the laboratory of Dr. Jonathan Javitch (Fu 1996), as described in Section 7.2.2. Some inconsistencies appear, especially at the N-terminus of these TMHs where our predicted TMH boundaries and the ensuing α-helical periodicity in the accessibility pattern are consistently 2-3 residues larger than the experimental results. This deviation was rationalized and explained in Section 5.8 in terms of the directionality of the C^{α}-C^{β} bond towards the N-terminus of an α-helix. Finally, there is an inconsistency in the predictions of R3.55 and V3.52 at the last turn of H3, which may indicate a misspredict ion at these residues, facilitated by the lack of any significant prediction derived for V3.52 (see Figure 6.7).

It is significant that the residues in the D2 receptor that were accessible to Cys-specific reagents applied from the extracellular side on intact cells were initially thought to label residues exposed to the binding site crevice. But then are found to reach a much broader area of the receptor than the previously considered ligand binding sites (see review in Section 2.1). Some of these AA sites found to be accessible by SCAM techniques include residues proposed to participate in specific TMH-TMH interaction, such as the D7.49-D2.50 interaction we have proposed based on reciprocal mutations on the GnRH and 5HT2A receptors (Zhou 1994; Sealfon 1995) (see Section 7.3.1). This probably means that the chemical reagent can diffuse inside the TMH bundle

reaching amino acid sites within TMH-TMH interfaces, where the interior of the TMH bundle behaves as a solvent for the probe. This is indicative of significant dynamical fluctuations in the interior of the receptor structure. It could also be that these residues are more accessible to the binding site, and thus presumably to ligands as well, than was previously thought.

6.2.- A sequential and anticlockwise arrangement of the 7 TMHs based on the rhodopsin map.

Once the predicted orientations of the seven TMHs in terms of TMH-TMH and TMH-Lipid interfaces have been derived in the previous section for each helix independently, these results can now be integrated into specific packing arrangements that define the relative positions of the 7 helices within the TMH domain. The prediction of a tertiary environment (TMH-TMH versus TMH-Lipid) for AA positions in all seven helices can be integrated first within each TMH, defining the extent of solvent (lipid) accessibility for each helix (Section 6.2.1). The resulting packing arrangement of the TMH domain should satisfy the predicted extent of lipid exposure for each helix. This criterion would determine the shape of the TMH domain, if the relative positioning of the seven TMHs were known. The relative positions of these seven segments in the TMH domain have been determined by electron cryoscopy studies of three different rhodopsin species (Schertler 1993; Schertler 1995; Davies 1996). From the resulting electron density

projection map of rhodopsin, the authors proposed the relative positioning of seven helical features. The predicted degree of lipid exposure for each TMH derived in Section 6.2.1, combined with experimentally derived TMH-TMH interactions and the adjacency between sequential TMHs imposed by short loops, are used in Section 6.2.2 to assign the identified helical features in rhodopsin to the seven predicted TMHs for the 5HT2CR. This results in the derivation of a sequential and anticlockwise helix arrangement, viewed from the extracellular side. The proposed packing arrangement facilitates the classification of interior-facing residues into specific TMH-TMH interfaces (Section 6.2.3).

6.2.1.- The prediction of a tertiary environment (TMH-TMH versus TMH-Lipid) for AA positions is determining the shape of the TMH domain.

There is a direct correlation between those predictions, presented above in Section 6.1 and Figure 6.7 , and surface accessibility. AA sites predicted in TMH-TMH interfaces would be buried, while AA sites predicted to be at TMH-Lipid interfaces would be exposed. Therefore, the predictions of specific surface accessibilities constrain the possible packing arrangements for the 7 TMHs of the 5HT2CR. The degree of azimuthal surface predicted in either environment is the determinant for those constraints. The extent of azimuthal surface predicted to be exposed to the lipids in each TMH can be appreciated by projecting onto the membrane plane the tertiary environment

predictions developed in the previous section, as shown schematically in Figure 6.8. Note that H1 and H4 have a single TMH patch predicted to be buried, i.e. they have a large azimuthal surface predicted to face the lipids. The opposite is true for H3, which is predicted to be almost completely buried. The other helices have intermediate degrees of azimuthal surface predicted to be buried. The integration of the predicted azimuthal surface accessibilities for each TMH presented here can determine the shape of the TMH domain, as illustrated below where this criterion is used to assign the observed density features in the rhodopsin projection map to individual TMHs 1 to 7.

6.2.2- Derivation of a sequential and anticlockwise helix arrangement.

The electron density projection map of rhodopsin was interpreted by the authors as representing seven TMH peaks (Schertler 1995). However, electron cryo-microscopy alone could not identify the individual TMHs to the seven proposed densities. Therefore, additional information is necessary to assign each of the seven density features to the proposed seven TMHs. I have chosen a sequential and anticlockwise arrangement of the 7 TMHs based upon a set of inferences explained in detail in the published review chapter (see Section 2.1). There are three independent sets of inferences that can be used to assign individual TMHs on the rhodopsin map:

a) An increasingly large number of experimentally derived helix-helix interactions, reviewed in Section 2 and shown schematically in Figure 2.16.

b) The shortness of interhelical loops which dictates spatial proximity for the consecutive TMHs they link. I applied this criterion to H1-H2 (IC1 loop of 3 residues), H2-H3 (EC2 loop 10 residues), and H3-H4 (IC2 loop 9 residues).

c) Biophysical criteria for the prediction of buried and lipid-exposed residues for each TMH which provide an estimate of the extent to which such a TMH is surrounded by others in the TMH bundle, or is lipid-exposed, as derived in the previous section (see Figure 6.8).

These three independent criteria can be integrated as shown in the published review (Ballesteros 1995) (Section 2.1) to substantiate a sequential and anticlockwise arrangement for the TMHs in the bundle. Since the chapter was written, there have been many new experimentally derived TMH-TMH interactions, which allow a much stronger derivation of the sequential and anticlockwise packing arrangement. I present the new derivation in this section, using as reference Figure 2.14 of the review chapter. For clarity, I will refer to TMHs in Figure 2.14 of the chapter as A, B, C, D, E, F, G, and derive below criteria for their identification as helices 1 to 7.

The most salient feature of the analysis of the degree of lipid exposure for each TMH is that H3 is the most buried helix and H4 and H1 are the most exposed TMHs (see Figure 6.8). Because H3 and H4 are linked by a short loop (9 residues), they would be adjacent in space and thus in the rhodopsin map. This identifies H3 and H4 in the rhodopsin map as shown in Figure 2.14 of the chapter, where the rhodopsin helical segments are labelled from A to G,

and thus H3=C and H4=D. The Zn binding sites engineered between H3-H2 and H3-H5 at adjacent residues in H3 (3.28 and 3.29) (Elling 1995; Elling 1996) indicates that H5 can be either E or F, while H2 can be either B or G. The multiple Cys crosslinks between H5-H6 define the H5->H6 in a counterclockwise arrangement, so that if H5 could be either E or F, the counterclockwise H5->H6 connectivity necessarily implies that H5=E and H6=F. This is consistent with the accessibility pattern and with the Cys crosslinking experiments between H3-H6 at the cytoplasmic boundaries (Farrens 1996), as shown in Figure 2.16. In order to assign H1, H2, and H7 to A, B, G, I know that H2=B or G based on the 3.28-2.64 Zn binding site. The same residue 3.28 in rhodopsin is the counterion for the Schiff base on K7.43, indicating that H7= B or G. The identification of H2 and H7 with B and G is consistent the proximity between these helices based on the N7.49-D2.50 revertant mutant , and is also consistent with the sites of attachment of covalently bound probes, 2.64 (Dohlman 1988) and 7.40 (Wong 1988), which are "linked" to D3.32 in H3 through the ligand. At the extracellular end, residue 7.39 in H7 is a well characterized ligand binding residue (7.39), which together with the covalently attached ligand site W7.40 indicates that both residues would be facing towards H3, which together with the revertant mutant 1.39-7.36 (Liu 1995) requires an orientation for H7 as observed in helix G of rhodopsin, so that H1=A and H2=B. This conclusion is consistent with the degree of lipid exposure for H1-H2-H7, as shown in Figure 2.10 of the

chapter. Note that the accessibility pattern is well explained for all seven TMHs with the proposed assignment of the helices. Most proposed residue-to-residue contacts are consistent with this arrangement, although residue N7.49 in H7 is not quite facing D2.50 in H2. This, and other proposed contact sites will thus require a reevaluation as we progress towards an atomic-scale 3D model of the 5HT2CR, as detailed in subsequent sections below.

6.2.3- Classification of interior-facing residues into sequential TMH-TMH interfaces.

Assuming a sequential and anticlockwise packing arrangement of the 7 TMHs as viewed form the extracellular side as described above, the interior-facing residues predicted in Section 6.1 can now be associated with sequential TMH-TMH interfaces based on the rhodopsin template derived above. This derivation is illustrated below in Figure 6.9. In this figure, the TMH patches adjacent to the lipid-facing patches (blue) face their next-neighbor TMHs in the following order: The patch immediately to the right (red) of the lipid-facing patch faces the next TMH in a clockwise direction, while the patch immediately to the left (red) of the lipid-facing patch faces the next TMH in an anticlockwise direction. In this way, these TMH patches can be paired into specific TMH-TMH interfaces. This pairing will greatly facilitate the search for evolutionary revertant mutants (**erms**, see Section 2.1) presented in Chapter 7.

6.3.- Prediction of the direction and degree of helix tilt from the analysis of the predicted helix-helix versus helix-lipid interface patterns: comparison with the tilts observed for rhodopsin.

The helix packing arrangement described in the previous section according to the electron density projection map of rhodopsin defines the azimuthal orientation of the 7 TMHs on the plane of the membrane. However, significant tilting of these helices was observed in the three-dimensional reconstruction of the rhodopsin electron density profiles (Unger 1995). In this section, a new method derived in this thesis and described in Section 2.3.2.2 is applied to predict the direction and degree of helix tilt for the 7 TMHs of the 5HT2CR from the analysis of the TMH-TMH versus TMH-Lipid interface patterns. These patterns were shown for the 7 TMHs in Figure 6.7 from an extracellular point of view. Residues predicted to face the interior of the TMH bundle, shown shaded, define patches on these helical nets with characteristic shapes. These shapes were classified in Section 2.3.2.2 as corresponding to different tilting orientations of the TMHs (see Figure 2.20 and 2.21 for the definitions).

The patterns of the predicted TMH-TMH interfaces can be decomposed into the four distinctive shapes on helical nets that correspond to the four tilting components (inward/outward and right/left), as shown in panel A of Figure 6.10. Note that I have used a cytoplasmic perspective to display the results, in order to compare the predictions made for the 5HT2CR with the

observed tilts in rhodopsin (Unger 1995). It should be clarified that the shapes of these TMHs are not scaled according to the degree of tilting observed in the TMH-TMH pattern, e.g. note that the outward tilt predicted for H1 or H4 is more pronounced that the same tilt predicted for H7. The shapes and ensuing tilts predicted for H3 are not very significant because only 4 AA are predicted in TMH-Lipid interfaces. Note that these shapes and the ensuing helix tilts include local disruptions due to the presence of Pro-kinks in H2, H5, H6, and H7. Although most TMH-TMH interfaces have continuous tilts, H4 is remarkably different in that there is a sharp bend in the direction of right/left tilting, a pattern present also to some extent in H1. Because H4 does not contain any proline or other indications of a disruption in the direction of its α-helix, this sharp bend means that the tilt of adjacent TMHs changes significantly at the level of W4.50. Because H3, H5, and H6, in potential contact with H4 according to the packing arrangement derived above (see Figure 2.14 of the review chapter, Section 2.1), are predicted to have continuous tilts, the sharp bend in the TMH-TMH pattern of H4 suggests that the helices interacting with the H4 helical segments before and after W4.50, are different.

The predicted tilts for the 7 TMHs of the 5HT2CR can be compared with the tilts inferred for rhodopsin from three-dimensional reconstruction of the measured electron density profiles (see Figure 6 of Schertler et al. (Schertler 1995) (frog)). This comparison is shown in panels B and C of Figure 6.3 for the

TMH domain viewed from the cytoplasmic side. The predicted inward/outward tilts for the 5HT2CR agree with the observed tilts in rhodopsin, except for H2. In contrast, the predicted left/right tilts for H2, H5, and H6 in the 5HT2CR are in opposite directions than those estimated by rhodopsin. The reasons for these deviation can not be explained at this level of analysis. It is resolved by modeling helix-helix packing interactions at the atomic level where the assumptions made here of ideal, "cylinder-like" α-helices will emerge as oversimplifications. A clear example is the bend introduced in the TMH structure by the Pro-kinks (PK), whose effect is present in all helices where there is a deviation between the predicted and the observed helix tilts (H2, H5, H6). In fact, the conserved Pro in H2 of neurotransmitter GPCRs is absent in rhodopsin, and instead a Pro is present in H1 for rhodopsin that is absent in H1 of the neurotransmitter receptors. The effect of these different PKs in H1 and H2 in the resulting H1 and H2 tilts will be discussed in Section 7.4. Exploring the conformations of these PKs in Section 7.2 would also suggest significant structural differences between the 5HT2CR and rhodopsin at the level of the H5-H6 interface at the extracellular side of the receptor.

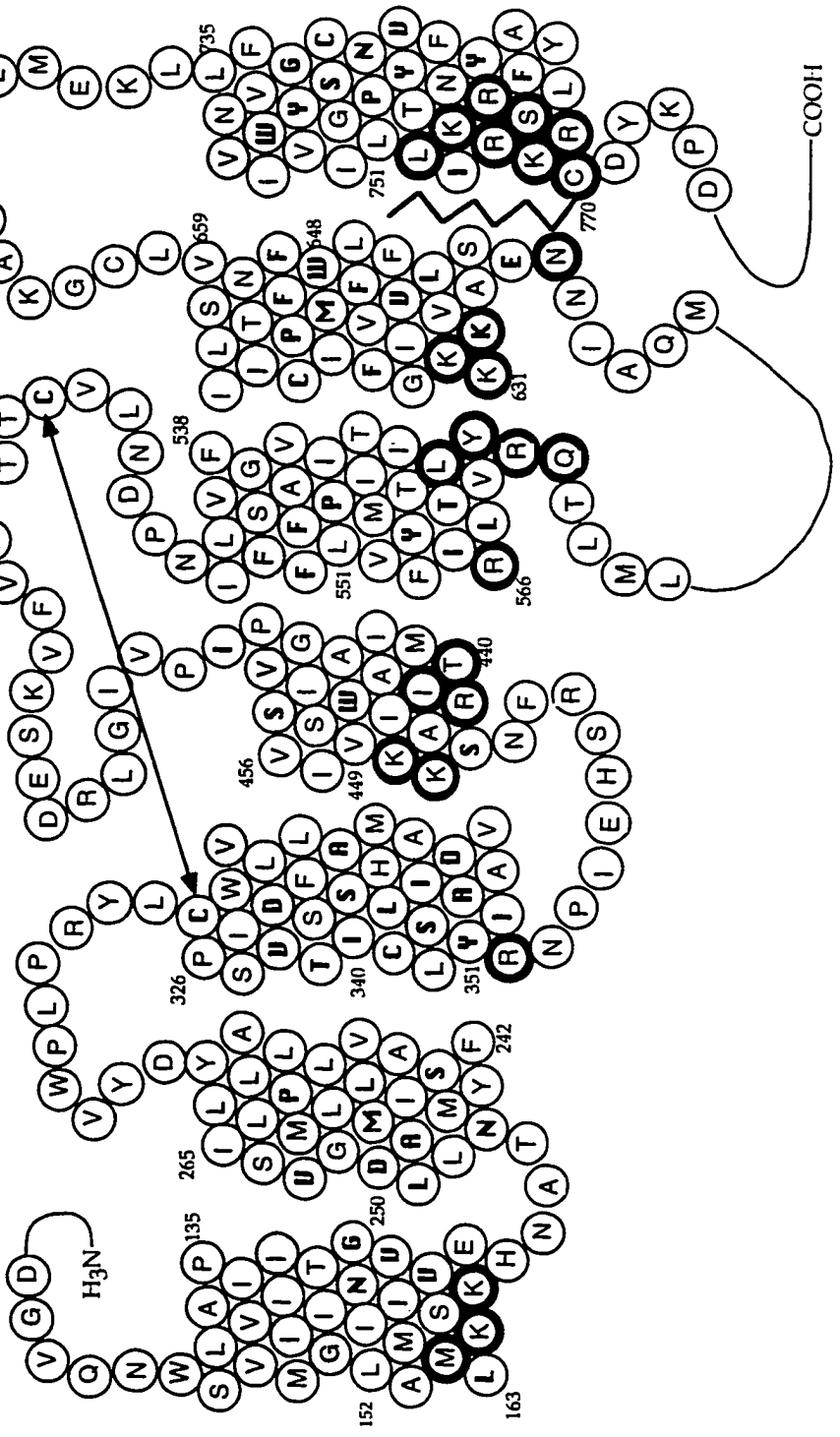
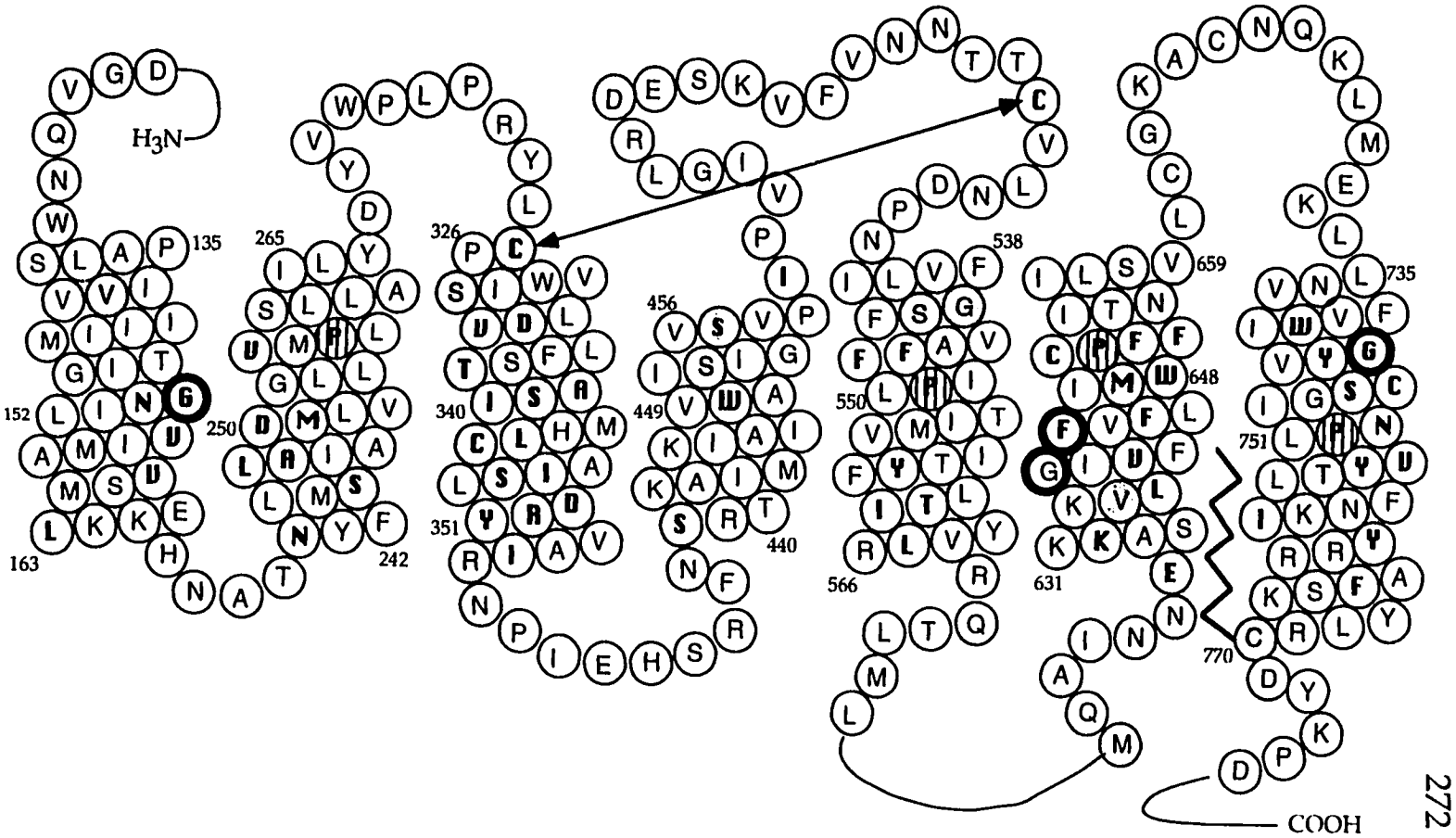


Figure 6.3
Predicted AA sites in TMH-Lipid interfaces. Shaded AA represent sites predicted using the criterion of highly variable sites but conserving a high hydrophobic character. Thicker circles represent sites predicted using the criterion of Arg/Lys at cytoplasmic boundaries -see Figure 5.3. Highly conserved AA (cons. index < 0.3) are shown in bold letters.

Figure 6.4 :
Occurrence of conserved Pro, Ser/Thr/Cys, and Gly residues within the TMH domain of neurotransmitter GPCRs.
Positions were conserved Pro (stripes), Ser/Thr/Cys (shaded) and Gly (thicker circle) residues are present in more than 50% of the receptors aligned (see Section 4.1) are shown for each TMH. Note that H3 is significantly overpopulated in Ser/Thr/Cys residues relative to other TMHs. Highly conserved AA (cons. index < 0.3) are shown in bold letters.



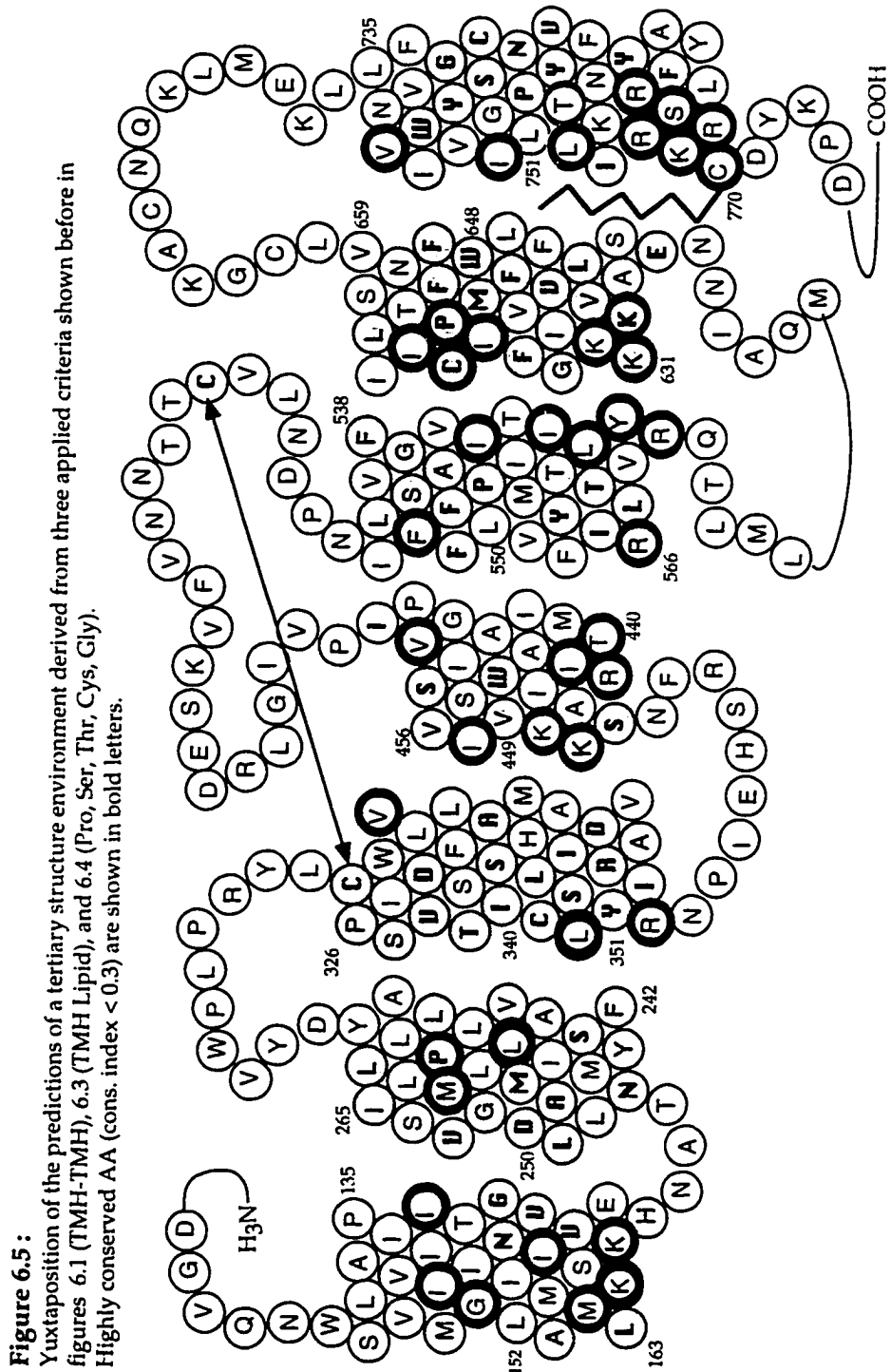


Figure 6.5 :
Yuxtaposition of the predictions of a tertiary structure environment derived from three applied criteria shown before in figures 6.1 (TMH-TMH), 6.3 (TMH Lipid), and 6.4 (Pro, Ser, Thr, Cys, Gly).
Highly conserved AA (cons. index < 0.3) are shown in bold letters.

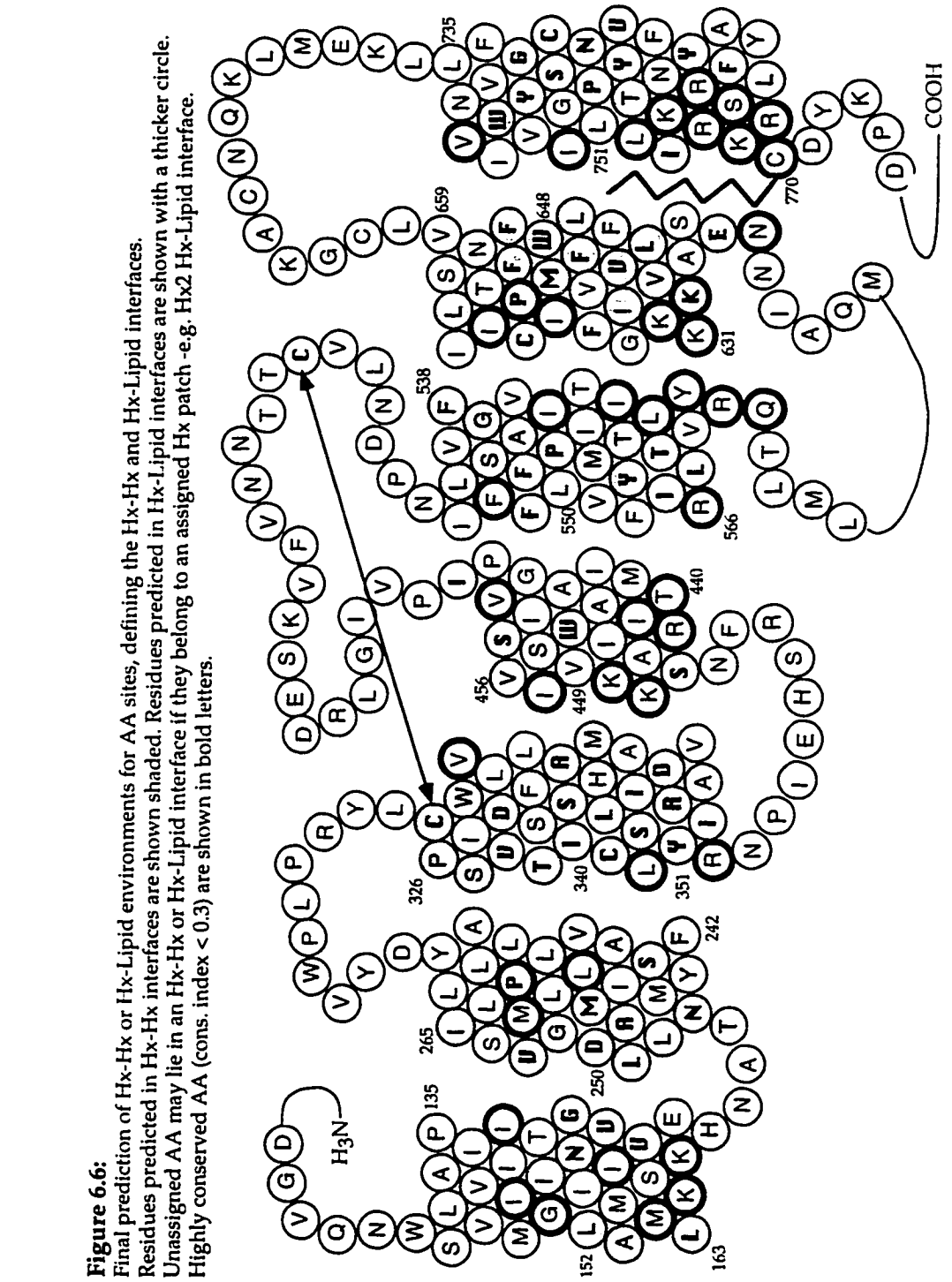


Figure 6.6:
Final prediction of Hx-Hx or Hx-Lipid environments for AA sites, defining the Hx-Hx and Hx-Lipid interfaces.
Residues predicted in Hx-Hx interfaces are shown shaded. Residues predicted in Hx-Lipid interfaces are shown with a thicker circle.
Unassigned AA may lie in an Hx-Hx or Hx-Lipid interface if they belong to an assigned Hx patch -e.g. Hx2 Hx-Lipid interface.
Highly conserved AA (cons. index < 0.3) are shown in bold letters.

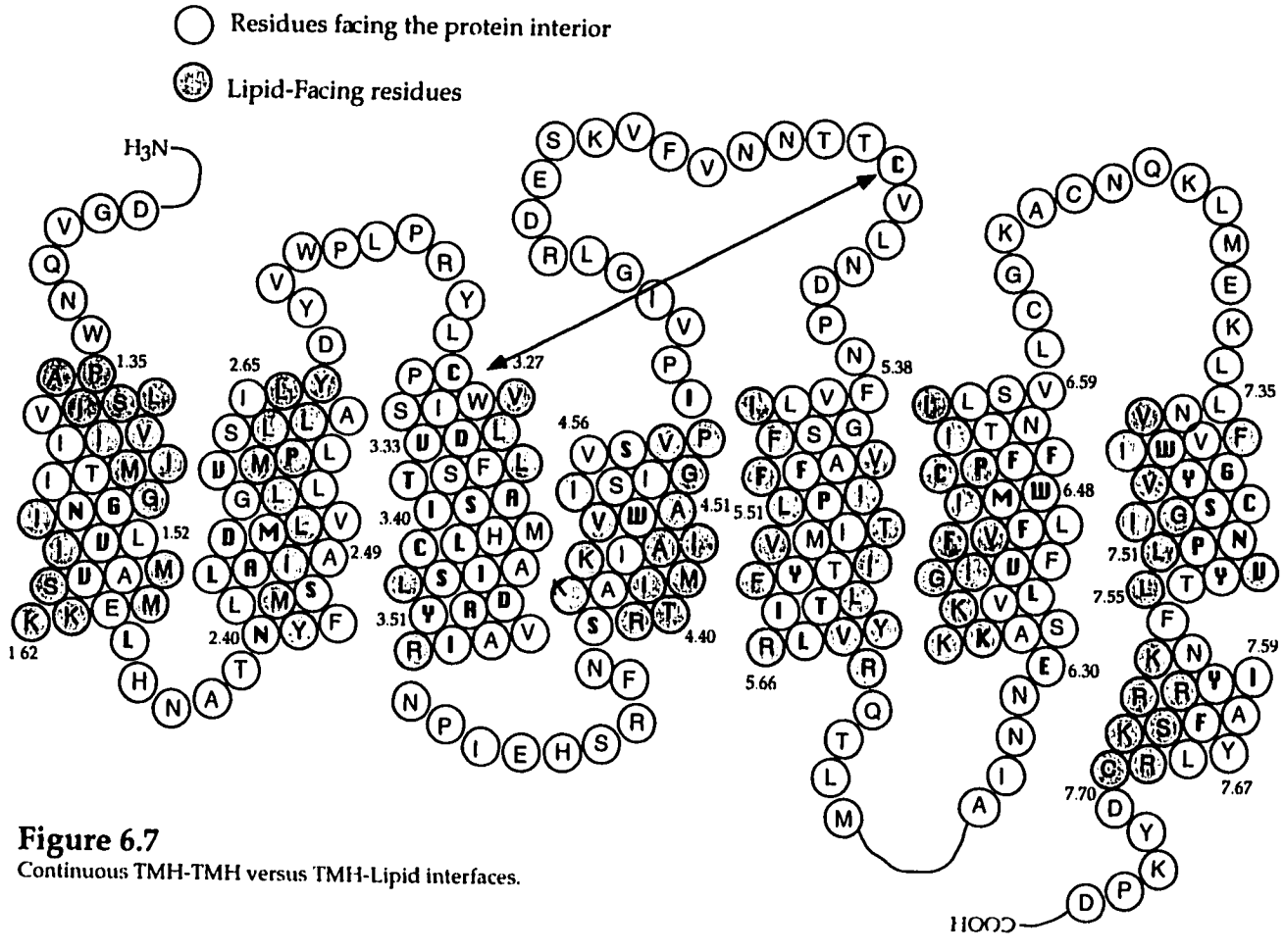


Figure 6.7
Continuous TMH-TMH versus TMH-Lipid interfaces.

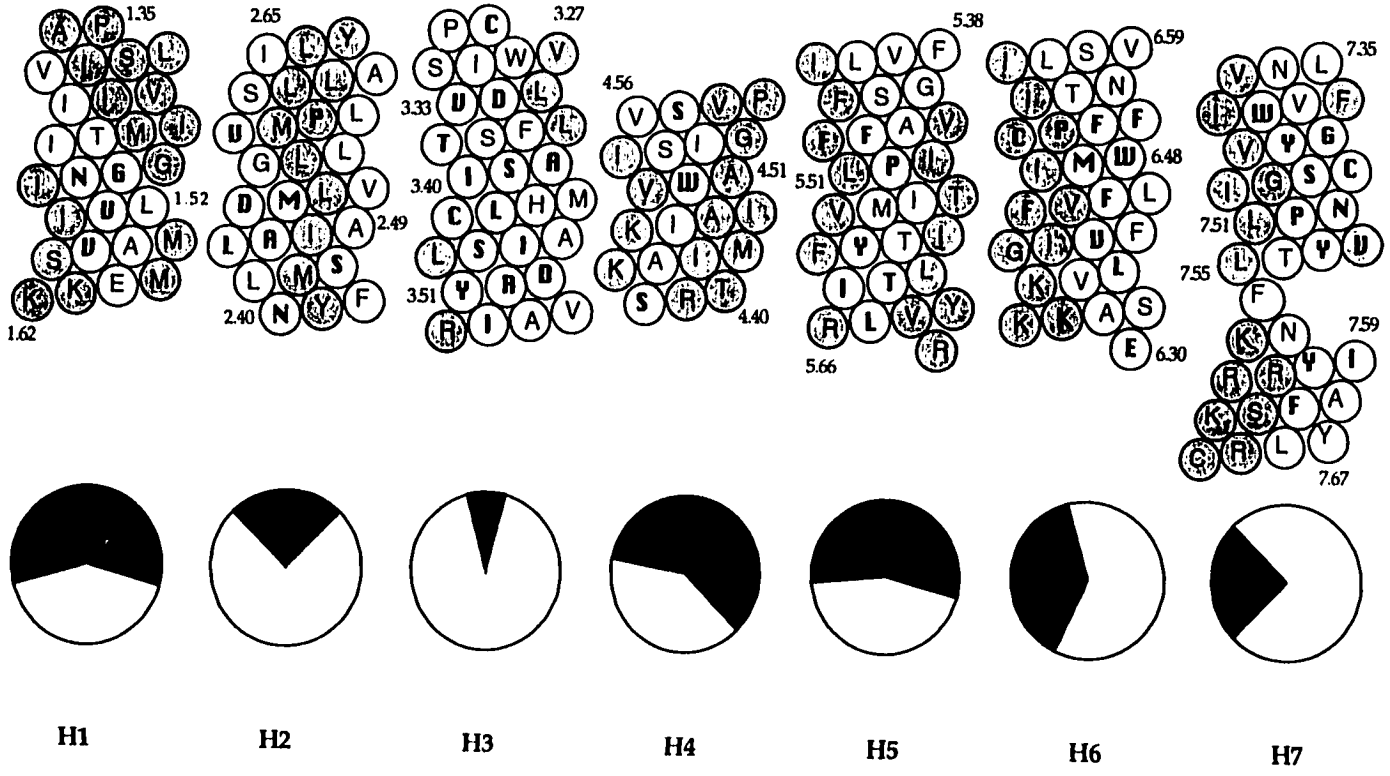


Figure 6.8

The extent of azimuthal surface predicted to be exposed to the lipids (shaded) in each TMH can be estimated by projecting onto the membrane plane the tertiary environment predictions developped before for each AA site.

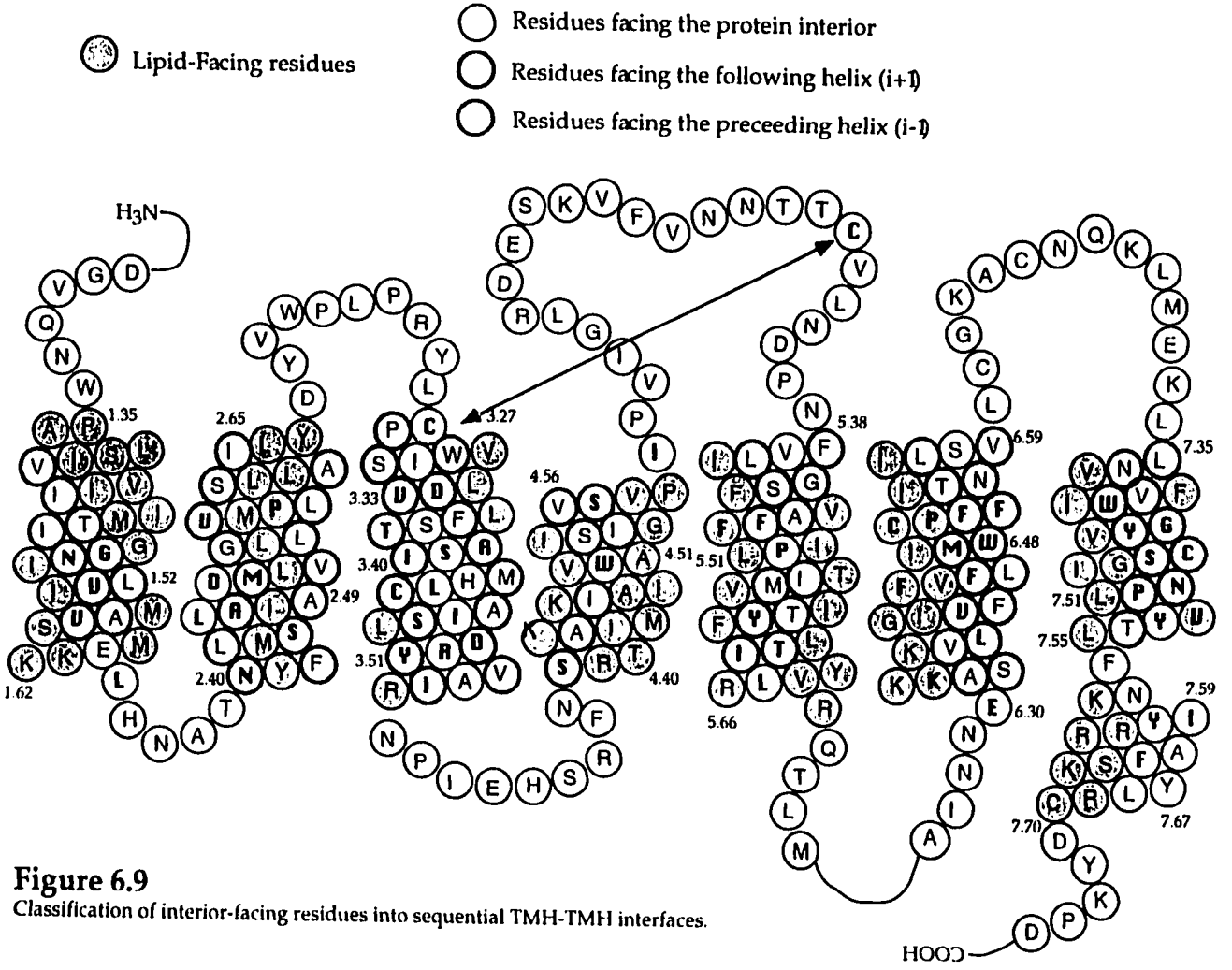


Figure 6.9
Classification of interior-facing residues into sequential TMH-TMH interfaces.

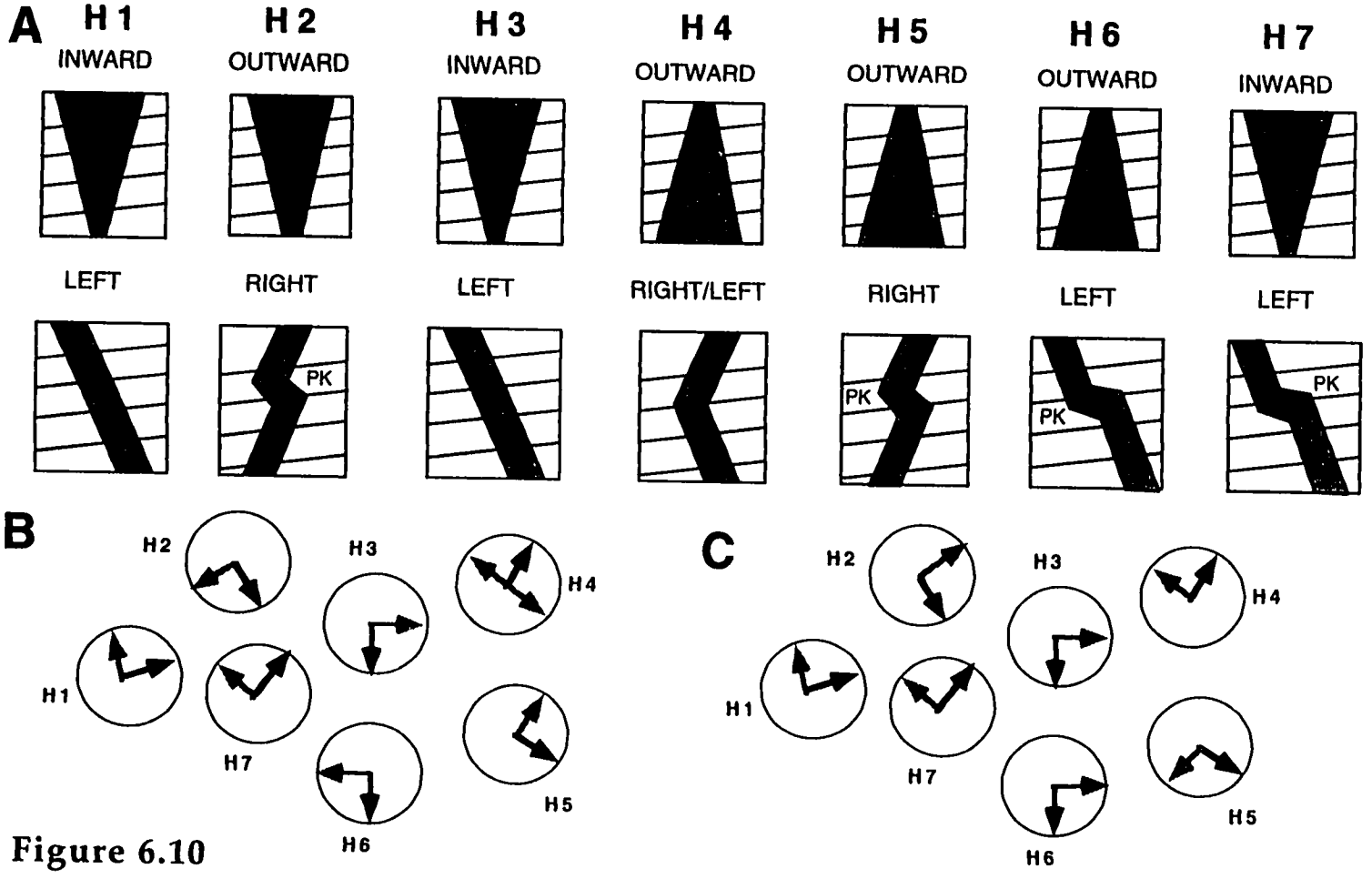


Figure 6.10

Predicting the TMH tilt for the 5HT2CR. A) Tilting components for the seven TMHs predicted by the shapes of the TMH patches. B) Predicted tilts for the 5HT2CR, and comparison with the observed tilts for rhodopsin (C).

7.- MODELING THE TERTIARY STRUCTURE OF THE THREE-DIMENSIONAL MODEL OF THE 5-HT_{2C} GPCR.

Following the guidelines described above, the model starts from an overall helix arrangement and progresses towards atomic resolution. The modeling process is described in the following sections as a step-by-step procedure that attempts to follow this logical progression. Initially, the seven TMHs are brought together based on the rhodopsin map and guided further by their predicted orientations and relative tilt (Section 7.1). This allows an exploration in 3D of the consistency among experimentally derived interactions. However, simple translations and rotations of the seven TMHs treated as rigid bodies can't explain several of the experimentally based criteria for the construction of the model. For this purpose it becomes necessary to invoke the flexibility provided by the PKs in H2, H5, H6 and H7 to overcome the inconsistencies of the initial model (Section 7.2). Once the overall backbone conformation has been resolved, modeling progresses towards the detailed side chain packing among the seven TMHs and the ligand. As the model is being refined to increased resolution, proposed interhelical interactions become increasingly interrelated and define specific structural motifs or receptor "microdomains". These microdomains are then used as "building blocks" whose assembly determines the TMH domain structure. I describe first in Section 7.3 the major organizing motifs or structure-

functional microdomains, which can be classified according to their functional correlates: three evolutionary conserved motifs proposed to participate in the receptor activation mechanism, and an evolutionary divergent motif, the ligand binding domain. First, three structurally conserved motifs involving the most conserved residues within the 7 TMHs and proposed to participate in the receptor activation mechanism are presented: A cluster of polar residues between H1-H2-H7 defined by the conserved N1.50-D2.50-N7.49 (Section 7.3.1); a cluster of charge residues at the cytoplasmic boundaries, we termed the "Arginine-cage" motif (Section 7.3.2); and a cluster of aromatic residues from H4-H5-H6 that functions as a "rotamer switch" in receptor activation centered on W6.48 (Section 7.3.3). The resulting positioning in three dimensions of proposed ligand-receptor interaction sites define the "Ligand binding domain" at the extracellular side of H3-H5-H6-H7, already shown to cluster in Figure 2.16.

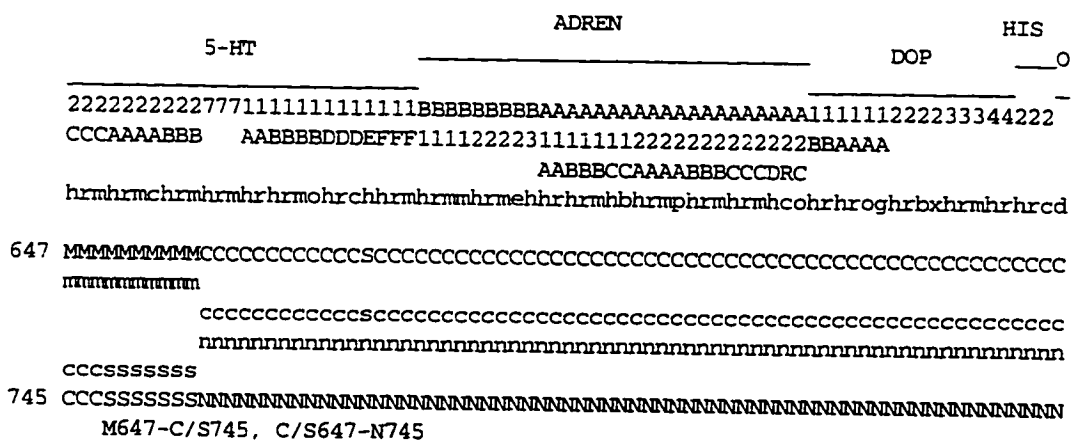
These organizing motifs, initially defined based on a few conserved or functionally relevant residues, are subsequently expanded through the process of optimizing helix-helix packing, described in Section 7.4. The result of this sequential modeling process outlined above is the proposed atomic scale three-dimensional model of the TMH domain of the 5-HT2C receptor. A summary of the proposed model, reviewing the experimentally derived constraints and the functional inferences developed in the modeling process is presented in Section 7.5. The evaluation of the resulting model with

energy related criteria (Molecular Dynamics simulations) and from functional considerations (the role or residue A/S5.46 in ligand selectivity between the 5HT2C and 5HT2A receptors) is described in the next chapter.

The methodological approaches applied through the modeling process outlined above changed significantly depending on the level of atomic resolution of the model. The initial steps in helix-helix packing are aided by inferences provided by experimentally derived constraints and by information about interactions among the most conserved residues. As the model evolves more towards a detailed atomic scale, other approaches becomes increasingly important, such as the inferences provided by evolutionary revertant mutants (**erms**) or the constraints imposed by the desired packing density. Known protein structures are remarkable in showing a high and consistent packing density of 72 % (72 % of the space inside the protein is occupied by protein atoms). It is difficult to obtain this degree of packing density through modeling, because it requires developing complementary Van deer Waals surfaces among interacting helices, but the experimental data can be used as criteria in the final stages of model refinement.

I described in our published review (Ballesteros 1995) (Section 2.1) the derivation of predicted tertiary structure contacts based on an analysis of the receptor sites whose mutations through evolution show a coordinated pattern, in which mutations at one site are accompanied by mutations at

another receptor site. Because this analysis is conceptually analog to experimentally found revertant mutants, and are interpreted in an equivalent manner inferring spatial proximity, I refer to these sites as "evolutionary revertant mutants" (**erm**). Due to the lack of systematic algorithms available for this analysis for large alignments, as done here for 60 GPCRs, the analysis has been performed by visual inspection. This task is facilitated by displaying the same alignment presented in Section 4 in a different format, where the rows and columns are transposed. An example is illustrated below:



Each line represents a single residue position, i.e. 6.47 or 7.45, and each column represents the amino acid present at this position for different receptors. Note that for this analysis I have used a more recent alignment of neurotransmitter GPCRs than the alignment presented and analyzed in Sections 4 and 5 to predict the TMH ends and relative orientations. This alignment was constructed by another student in our department, Jennifer Weinstein, following the same guidelines and criteria as described in this

thesis under Dr. Weinstein's supervision and my direct collaboration.

Inferences derived from experimental data or highly conserved residues were complemented with the analysis of **erms** in two different modes of application: First, clear, statistically significant **erm** patterns can guide the initial helix packing by assuming a spatial proximity among the evolutionary correlated residues. This methodology is especially important when other information is not available, e.g. see H4 packing below. As the TMHs are packed with increasing atomic detail, inferences from **erm** can be used in a somewhat different approach: assuming a common scaffold or backbone conformation for all the neurotransmitter GPCRs, the model accommodates the presence of whatever different side chains are present through evolution. This means that significant changes in a given buried locus are often accompanied by concomitant changes in nearby residues. By this approach I can identify at the atomic level concomitant changes in nearby residues that are otherwise poorly *pairwise* correlated through evolution. This corresponds to finding **erms** based more on a spatial arrangement (i.e., based on which residues are in contact with a residue at a given locus the side chain properties at this locus change significantly), rather than on the connectivity among residues provided by the primary sequence (i.e., which residues in the receptor changes when a given residue changes). This approach may provide specific structural hypotheses for correlations observed at the sequence level.

I discussed in Section 3 that there are several functional and conformational states of the receptor, and that I would minimally have to consider the ligand-bound receptor in its inactive and active forms. Most of the experimental data pertinent to structure for GPCRs is derived from what should be considered an inactive state. There is little experimental data pertinent to the structure of the active state of the receptor, and most is related to conformational changes occurring upon activation, thus relative to the more well defined inactive state. Consequently, the model of the TMH domain of the 5HT2CR that I have developed in atomic detail is meant to represent the inactive state of the receptor. Nonetheless, inferences have emerged and will continue to be made on the nature of the conformational changes leading to activation or to other available receptor states, at the level of specific structural motifs.

7.1- Consistency among experimentally derived interactions in an initial 3D template of the TMH domain: rigid-body motions of standard TMHs is insufficient to explain available constrains.

The seven TMHs were built selecting default backbone and dihedral angles as specified in the published review chapter (Section 2.1). This procedure was automated by Liisa Laakkonen in the form of a CHARMM script, which I used to build all TMHs. The three-dimensional models of the seven TMHs built individually are first brought together into a sequential

and anticlockwise packing arrangement following the rhodopsin map, as presented in Section 6.2. This initial arrangement is guided by the orientation (Section 6.1) and tilt (Section 6.3) predicted for each TMH. Packing the seven TMHs was performed starting from this template and searching for specific interactions among the seven TMHs based on the experimentally derived residue-to-residue contacts and the role of the highly conserved residues.

The experimentally derived interhelical interactions used in modeling the 5HT2CR were described in detail in Section 2.2.2.2 and summarized in Table 2.3, and illustrated in Figure 2.16 on 7 helical wheels. The helical wheels in this figure are actually ideal α-helices viewed from the extracellular side, represented by their C^{α}-C^{β} vectors. These 7 ideal TMHs are arranged according to the rhodopsin projection map (Schertler 1995), and oriented by positioning conserved residues (shown darker) towards the protein interior as described in Chapter 6. Thus, the consistency among these proposed TMH-TMH interaction sites in terms of their azimuthal orientation on the membrane plane, assuming ideal α-helices, can be explored in Figure 2.16. The orientation of H7 derived from these criteria is consistent with the double revertant mutant between positions 7.36 and 1.39 in muscarinic receptors (Liu 1995), and with residues 7.39 and 7.40 facing the binding site (yellow lines in Figure 2.16). Note that the proposed ionic bond in rhodopsin between 7.43 and 2.57 (Rao 1994) is also consistent, but cannot be used as criterion here because this interaction pertains to the active state of the receptor (see Section

2.2.2.2 and Table 2.3). The resulting H7 orientation positions N7.49 towards H6 instead of towards H2, where it is proposed to interact with D2.50 (Zhou 1994; Sealfon 1995). This inconsistency can not be overcomed by simple translations and rotations of ideal TMHs treated as rigid bodies. The flexibility provided by the Pro-kinks is necessary to overcome this and other inconsistencies, as described below (Section 7.2).

7.2- Variability in the initial positioning of the seven TMH template induced by the effect of the Pro-kinks (PK) in H2, H5, H6, and H7: criteria and selection of specific PK conformations.

The extent to which the flexibility provided by the presence of Pro residues can alter the orientation of the helical backbones for H2, H5, H6, and H7 is illustrated in Figure 7.1.a; 100 representative Pro-kink (PK) conformations of H6 and H7 obtained from MonteCarlo simulations described below are superimposed in the context of the initial 3D model of the TMH domain. Note the wide range of alternative helix-helix interactions that arises from the analysis of the available Pro-kink conformations for modeling purposes. The PKs present also in H2 and H5 will add further degrees of freedom for modeling the TMH domain. However, the range of energetically available PK configurations is subjected to certain conformational restrains, described in Section 7.3.1, that would limit the available choices. Still, specific criteria are necessary to select among these multiple, energetically available

PK conformations, described below for H7 (Section 7.2.2), H5 (Section 7.2.3), and H6 (Section 7.2.4).

7.2.1- Defining the extent of the conformational freedom induced by PK: Constraints on the direction of bending and face-twisting allowed by a PK.

A Pro residue in an α-helix environment would clash with the C=O i-4 of the preceding turn, thus inducing a kink in the helix turn comprising residues (i-4, i) that breaks the 3.6 amino acids per turn periodicity and generates up to three H-bond free backbone groups (Barlow 1988; Williams 1991; Ballesteros 1992a; Sankararamakrishnan 1992; Ballesteros 1995). The resulting orientation of the α-helical segments N-terminal (NTPK) and C-terminal (CTPK) from the PK can be described schematically by a bend and a face twist, as shown in Figure 2.10 of our published review (Ballesteros 1995) (see Section 2.1). Sankar et al. (Sankararamakrishnan 1992) have shown that these two parameters are correlated, a larger bend implies a larger face-twist, and vice versa.

The bend results from the need to alleviate the steric clash between the C^{δ} carbon of the Pro ring and the C=Oi-4, to which the α-helix reacts by opening the (i-4,I) turn increasing the distance between these two atoms. It follows that the orientation of the bend induced on the CTPK segment would always be in a direction opposite to the Pro residue. I have consistently found this restraint in the conformational flexibility of PKs in all known protein

structures and in multiple simulations on a variety of helical segments, including the GPCRs discussed here. This directionality can be observed in the 100 representative structures of the PKs in H6 and H7 shown in Figure 7.1.b; because P7.50 is facing inside the bundle (see Figure 6.9) H7 bends only outwards, while because P6.50 is facing the lipids (see Figure 6.9) H6 bends only inwards. Thus, based on the predicted orientation of P2.59 and P6.50 in TMH-Lipid interfaces, and P5.50 and P7.50 in TMH-TMH interfaces (see Figure 6.9), it can be concluded that H2 and H6 bend outwards while H5 and H7 bend inwards. Spin labeling experiments in rhodopsin (Farrens 1996) have confirmed the hypothesis that receptor activation involves a rigid-body motion of the TMHs moving away from each other at the cytoplasmic boundaries (Luo 1993; Zhang 1993). This movement was shown to be facilitated by the PKs in H5 and H6 based on molecular dynamics simulations of a model for the 5HT2AR (Luo 1993; Zhang 1993). Conformational changes at these PKs responsible for this outward movement upon receptor activation can be described by an increased or decreased degree of bending, but always in the same direction. Thus, according to the predicted environment for these Pro residues (see Figure 6.9), receptor activation could be facilitated by a adopting a larger kink in H5 and/or H7, and by adopting a smaller kink in H2 and/or H6.

The face twist induced by a PK arises due to the break in the 3.6 amino acids per turn periodicity that results when the backbone in the (i-4, i) turn

adopts dihedral angles that deviate from standard α-helical values. I proposed the novel concept of the face-twist induced by a PK from an analysis of known protein structures. I have observed such face-twist in all PK structures available in the PDB database of known structures. An analysis of the face twist present in PKs for all known protein structures indicates that the direction of the face twist is always right-handed, never left-handed. This observation can be rationalized by noting that the distortion of the helical (i-4, i) backbone dihedral angles is concentrated on *phi* and *psi* angles, energetically much more flexible than the pi-character omega angles. The rotation axis of *phi* and *psi* angles in an α-helix are consistently at an angle of 32±3° from the helix axis. Significant rotations on either of these *phi* and *psi* angles individually is energetically feasible in only one direction; *phi* angles can be decreased but not increased, and *psi* angles can be increased but not decreased. This is due to severe steric clashes with the backbone in the opposite directions. The net effect of this conformational constraint at the level of individual *phi* /*psi* angles is that only right-handed, but not left-handed, rotations are allowed, explaining the observed restriction of only right-handed face twists observed for PKs in known protein structures. Thus the right-handed character of an α-helix constrains the direction of the face twist induced by a PK to an accentuation of the inherent right-handedness. This conformational restriction would become a significant restraint in accommodating the experimentally derived interactions for H7 and H5, as

described below.

7.2.2- The conformation of the PK in H7: A significant face twist required by interhelical constraints and supported experimentally by SCAM analysis on the D2 receptor.

We have proposed an interaction between N7.49 and D2.50 in GPCRs, based on mutagenesis experiments on the GnRH and 5-HT2A receptors (Zhou 1994; Sealfon 1995), that supported a reciprocal an interchangeable role for these two residues, described in Section 7.3.1 below. However, the orientation of N7.49 that results from the initial positioning of the 7 TMHs into the rhodopsin template is inconsistent with this hypothesis, as shown above in Figure 2.16. In order to reorient N7.49 towards H2 maintaining the orientation of the extracellular segment of H7 to satisfy the other experimentally proposed interactions, a significant rotation is necessary in a right-handed direction (see Figure 2.16). This rotation can be accomplished first by modeling N7.49 with X1=g-, a rotamer conformation normally disallowed in α-helices due to steric clashes with the C=O i-3. However, the flexibility of the PK between S7.46 and P7.50 could be used to accommodate this rotamer conformation. Analysis of known PK in high resolution structures indicates that the X1-g- conformation for residues i-1 from the Pro never occurs, due to a steric clash with the Pro ring. This clash can be alleviated by altering the *psi* angle linking N7.49 to P7.50 towards

increasingly higher values, i.e., from -41° to -18°. The final value of this *psi* angle corresponds to a 3-10 helix. This conformation of N7.49 can be stabilized by H-bonding between the N7.49 side chain and the $C{=}O_{i-4}$. The energetic profile of this interaction and the ensuing conformation of the NP motif was explored by the MonteCarlo-based technique of Conformational Memories (Guarnieri 1996) on the long version of H7 (residues 7.35 to 7.70). The NTPK and CTPK segments were constrained α-helical and the exploration was centered on the PK allowing *phi* and *psi* angles within this region to vary ±50° during the simulation, all side chains were allow to rotate freely. The proposed conformation was not stable as an independent structural motif, i.e. was not significantly populated. Nonetheless, constraining the proposed H-bond in these simulations provided a set of structural templates of this NP motif for further analysis (see Section 7.4.8). A representative set of 100 structures obtained by this procedure was shown above in Figure 7.1. The PK conformations induced by the proposed H-bonding scheme were characterized by a highly kinked H7, as shown in Figure 7.2. It is noteworthy that the same conformation was derived independently for this NP motif by modeling the proposed interaction of R3.50 with D2.50 and N7.49 in the context of the Arg-cage motif described in Section 7.3.2 (see Figure 7.13). Interestingly, the *psi* angle of N7.49 is the backbone dihedral angle within the PK that most significantly modulates the degree of face twisting between the NTPK and CTPK helical segments. Notably, increasing N7.49 *psi* induces a

significant right-handed face twist as required to reorient N7.49 towards H2 (see Figure 7.4 below). The resulting reorientation of N7.49 allows the proposed interaction between N7.49 and D2.50 (Zhou 1994; Sealfon 1995), as described in Section 7.3.1.

The resulting conformation of the NP motif in H7 also brought into consistency the pattern of predicted interior versus lipid-facing residues, as was experimentally validated by a collaboration with the laboratory of Dr. Jonathan Javitch using the Cys Substituted Accessibility Method (SCAM) analysis on the D2 receptor. This study have been published (Fu 1996) and therefore I present here only the relevant figures describing the results as Figure 7.3 (helical net representations of the SCAM results illustrating the kink in the probe accessibility pattern) and Figure 7.4 (H7 modeled as described above is in agreement with the SCAM data).

7.2.3.- The conformation of the PK in H5: A dramatic face twist.

The patch of residues in H5 predicted to be in TMH-TMH interfaces exhibits an abrupt change in direction on the surface of this helix in the three-dimensional model at the level of the PK. This effect can be appreciated in Figure 7.5.a on a Van der Waals representation of H5 colored according to the predicted TMH-TMH (green) and TMH-Lipid (yellow) interfaces. Note the sudden alteration in the direction of the green patch at the level of the Pro residue, shown in purple. As it progresses towards the cytoplasm after P5.50,

the interior-facing residues swing around the helix surface describing an arc of 180° degrees as they reach the cytoplasmic boundary. The net result is a dramatic twist in the orientation of the residues at the cytoplasmic side, i.e., a "face-twist". This reorientation gives rise to an inconsistency between experimentally derived interactions, as shown in Figure 7.5.b, reflected in the resulting orientation of Y5.58 and C5.59 relative to A5.46. These residues at the cytoplasmic side of H5 were shown by Cys crosslinking experiments in rhodopsin to face H6 at the level of residue 6.37 (Kono 1996). We have proposed that residue 5.46 interacts with the indole nitrogen of ergolines and indolalkylamines in the 5HT2CR based on modeling and mutagenesis studies (Almaula 1996a) presented in Section 8.3.1. Therefore, these residues should be facing the interior of the TMH bundle. However, residues Y5.58 and C5.59 appear on the face opposite to A5.46, as can be seen more clearly in Figure 7.5.c where the solvent accessibility surface of H5 is shown colored according to the predicted exposure, and these three residues are shown in red. Note that A5.46 and C5.59 are barely visible on the fringes of the predicted TMH-TMH interface.

The remarkable face-twist observed in Figure 7.5.c where the predicted interior-facing residues swinging around H5 was not expected from the schematic representations of H5 presented before in this thesis. This is illustrated in Figure 7.5.d where the same interior-facing residues are shown on a helical wheel representation; note that they define a continuous and

straight TMH-TMH interface, contrary to the pattern observed in the 3D representation. The helical wheel presented in panel D is actually the same 3D structure of H5 shown in panels A, B, and C, represented by its Cα->Cβ vectors and viewed across its helix axis. The only difference between these two conformations of H5 is the backbone dihedral angles of the PK, which in the helical wheel representation adopt ideal α-helical values. Thus, the "face-twist" observed in H5 is induced by the PK at this helix, consistent with the observation that it is at the position of the Pro residue where the abrupt change in orientation takes place (see panel A above).

In order to relieve the inconsistency described above for the orientation of Y5.58 and C5.59 relative to A5.46, or to facilitate the ensuing TMH-TMH packing arrangements required to satisfy the experimentally derived interactions, I should select a PK conformation that reverts the face-twisting observed above. In this regard, I should emphasize that the 3D structure of H5 analyzed above was built using standard procedures as described in our review Chapter (Section 2.1). Thus, the PK conformation responsible for the face-twisting effect observed is an standard, minimal PK. This situation resembles the requirement for a significant face twist at the PK of H7 to bring N7.49 towards D2.50 discussed in Section 7.3.1. However, the direction of the desired face twist for H5 is opposite to the direction found necessary for H7, as illustrated in panel E where these two cases are compared: Looking from the extracellular, N-terminus of H5 and H7, the desired face twist for H5 is left-

handed while for H7 it is right-handed.

I described in Section 7 the conformational properties of a PK. When I discovered the novel concept of a face-twist induced by a PK (Ballesteros 1992a; Ballesteros 1992b), in addition to the previously reported bend (Barlow 1988), I also noted that the face twist induced by a Pro residue in an α-helix is unidirectional. That is, only right-handed PKs are found in known structures. I rationalized this finding based on the right-handedness of the α-helix itself, which prevented the PK from adopting *phi /psi* values corresponding to a left-handed twist because they would clash severely with the helix backbone. Thus, the PK in H5 of GPCRs is not only unable to provide the face twist necessary to solve the Y5.58-S5.46 conundrum, but actually further accentuates the problem relative to an ideal α-helix. An additional mechanism, different from the Pro-induced face twist proposed in the case of H7, is thus necessary to bring the Y5.58-S5.46 discrepancy into consistency with the model of the receptor. It follows that the starting conformation for the H5 PK should be selected as having the minimal face twist possible. There is a broad but consistently positive correlation between the bend angle and the degree of face-twisting, as quantified by Sankararamakrishnan et al (Sankararamakrishnan 1992) (note the different geometrical definition between a face-twist and the wobble-angle proposed by these authors). Therefore, I selected a PK conformation as minimally kinked as possible for modeling purposes.

The observation that the PK in H5 accentuates the difficulty mentioned above to satisfy all packing constraints is a bit perplexing, because such a situation might indicate shortcomings of the modeling process. In the case of H5, the final conformation of the PK that is selected to accommodate optimally all the criteria is a standard PK geometry as defined in the literature. Nonetheless, in the process of modeling the ligand inside the receptor binding site, as described later on Section 7.4.5, I did find it consistently difficult to keep the extracellular end of H5 sufficiently close to the extracellular end of H3. Although these helical segments are proposed to be in close contact based on the Zn binding site between 3.29 and 5.35 or 5.39 (Elling 1996), this constraint applies only to the unliganded receptor, not to the liganded receptor (see Section 7.4.5 for a details). Although I emphasize again that this could suggest a shortcoming of the model, the consistent presence of these difficulties observed for H5 could also reflect an intrinsic property of the receptor structure. In particular, the role of the highly conserved disulfide bridge between C3.25-C5.31 could be related to the requirements of an unattainable face twist required for H5 PK. The reason is that the packing arrangements that result from fitting the required interactions using a slightly kinked H5, especially upon ligand binding, positions the extracellular end of H5 towards the membrane, away from H3, at a resulting H3-H5 maximal distance close to that allowed by the loop between H3-H5 created by the disulfide between C3.25-C5.31. In these

instances, the disulfide bridge could be important in holding H5 relative to H3 and thus stabilizing a more packed TMH domain.

The hypothesis that the disulfide bridge between C3.25-C5.31 plays a crucial role in packing H5 towards the TMH domain at the extracellular side of the receptor would explain why DTT treatment, or mutation of the Cys residues in other GPCRs equivalent to C3.25 and/or C5.31, which disrupts this Cys-Cys bridge, induces receptor unfolding. This interpretation is further supported by the findings in the β2-adrenergic receptor that the destabilizing effect of disrupting the Cys-Cys by DDT can be partially reverted by an antagonist; antagonist binding prior to DTT treatment induces a much slower unfolding kinetics as measured by Trp fluorescence, indicating a stabilizing effect of the antagonist on the Cys-Cys disrupted conformation of the receptor. I interpret these data as suggesting that the bound antagonist, by interacting simultaneously with H3-H5-H6 at the EC side of the receptor, brings together H5 and H3 in a manner similar to the native Cys-Cys disulfide bridge, but with decreased strength.

It should be noted that the difficulties described above in satisfying the required orientation and interactions for H5, and the ensuing prediction of a little kinked H5 at P5.50, apply only to the inactive state of the receptor. Upon activation, Spin labeling techniques have shown that the cytoplasmic side of these TMHs become more exposed to the aqueous cytoplasm, with an increased flexibility in their side chains suggestive of a looser packing

(Farahbakhsh 1995; Altenbach 1996; Farrens 1996; Yang 1996a; Yang 1996b). This is expected because the last turns of the TMHs, comprising the Arg/Lys motif we predicted to be intrinsic to the TMHs, contain the regions most relevant for G-protein coupling and would thus be expected to become accessible to the G-protein upon activation. This is consistent with the proposal that the conformation of the H6 PK in the active state is bent more towards H7, whereas of the inactive state preferred H6 PK conformation, bent towards H5 (see Section 7.3.3). It follows that the cytoplasmic side of H6 moves away from H5 upon activation, thus further diminishing the constraints discussed in this section on the conformation of the H5 PK. These various inferences regarding the conformational changes following receptor activation indicate that in the active state, the PK of H5 would be freed from the constraints derived for its conformation in the inactive state described above. This would be expected to result in a more flexible H5 PK, adopting multiple conformations. This "active state" of the PK in H5 can be envisioned from the illustrated superposition of 100 representative H6 and H7 PK conformations shown in Figure 7.1.

7.2.4- The conformation of the PK in H6: Intrahelical modulation by H-bonding (C6.49) and aromatic (F6.44-W6.48-F6.52) interactions.

The modulation of the PK conformation by specific intrahelical interactions, such as the aromatic cluster (described in Section 7.3.3.1 below)

and H-bonding within the PK, together with the experimentally shown H3-H6 rearrangement upon activation, constitute independent and compatible criteria to model the PK in H6. Integration of structural inferences arising from these three independently derived criteria determines the range of PK conformations available for modeling the TMH bundle in atomic detail.

The conformation of the H6 PK was explored by means of MonteCarlo simulations on the isolated H6 segment, following procedural steps as described for H7 simulations in Section 7.2.2. A set of 100 representative conformations was shown above in Figure 7.1. The results show that C6.49 modulates the PK conformation by H-bonding to the C=O of L6.45 (97 % of the conformations analyzed) and inducing a bend of the cytoplasmic portion of H6 towards H5 (see Section 7.4.4 for a description of the directionality). It was also found that the PK conformation was modulated by the conformational state of the aromatic cluster, defined by F6.44, W6.48, and F6.52 and described in detail in Section 7.3.3.1 below. When W6.48 adopts the X1=g+ rotamer, the resulting configuration of the aromatic cluster induced a PK conformation where the cytoplasmic segment of H6 was preferentially bent towards H5. When W6.48 adopts the X1=t rotamer, the resulting configuration of the aromatic cluster induced a preferential bend towards H7. Spectroscopic experiments on rhodopsin (Lin 1996) suggest that the X1=g+ configuration of W6.48 is prevalent in the inactive state of the receptor (see Section 7.3.3.1). Therefore, the PK conformation of H6 was selected from the set of

representative structures of the MonteCarlo simulation on the g+ configuration of the aromatic cluster. Receptor activation involves a significant movement of H6 at the cytoplasmic side, as proposed by molecular dynamic simulations (Luo 1993; Zhang 1993) and Spin labeling studies on rhodopsin (Farrens 1996). I have proposed that the PK on H6 is likely to participate in this conformational rearrangement (see below and Section 7.3.3.1). This implies that upon activation the PK of H6 adopts a conformation that is significantly distinct from the PK conformation present in the inactive state. I have observed such a difference in the conformations of the H6 PK resulting from MonteCarlo simulations on the isolated H6 with a W6.48 in the conformations corresponding to the proposed "activated" (W6.48 X1=t) or the "inactivated" (W6.48 X1=g+) forms (see Section 7.3.3.1). The results can now be used to select the PK conformation for the inactive state of the 5HT2CR modeled here. Therefore, among the representative structures obtained from the MonteCarlo simulations on the g+ configuration of the aromatic cluster, I have discarded those PK structures which overlapped with PK conformations obtained from simulations on the *trans* configuration.

I derived an additional and independent criterion to select the H6 PK conformation based on the conformational changes observed between H6-H3 at the cytoplasmic side for rhodopsin based on spin labeling experiments (Farrens 1996), Cys-Cys crosslinking (Farrens 1996), or engineered Zn binding sites (Sheikh 1996) between residues 3.53-3.54 in H3 and 6.30-6.34 in H6.

Previous Spin labeling experiments substituting Cys only on individual helices (H3 or H6) at the same positions had indicated that the cytoplasmic side of H6 undergoes a significant movement upon activation becoming much more exposed to the aqueous cytoplasm (Yang 1996b), while H3 did not show a significant change in its degree of exposure (Farahbakhsh 1995; Altenbach 1996). Therefore, the authors proposed that the observed movement between H3-H6 upon activation was due to a rigid-body movement of H6, in the direction away from H3 and the remaining TMH domain. However, results from my simulations of H6 described in Section 7.2.4 suggest that the proposed movement of H6 in rhodopsin might be facilitated by the flexible hinge provided by the conserved Pro, as shown in Figure 7.1, where different PK conformations would result in H6 moving away or closer to H3. Note that because P6.50 is predicted on the lipid face of H6, a more pronounced kink brings H6 closer to H3, while a less pronounced kink moves H6 away from H3. Thus I propose that in the inactive state the PK of H6 is held in a significantly bent conformation bringing H3-H6 together at the cytoplasmic boundaries, and that upon activation the PK of H6 is either more flexible, adopting on average a less pronounced kink conformation, or is held in a less pronounced PK conformation. There is experimental support for the proposed role of the H6 PK in receptor activation from recent mutagenesis experiments in D1 receptors aimed at residues adjacent to P6.50 and P5.50 searching for effects of the substitutions on the conformational

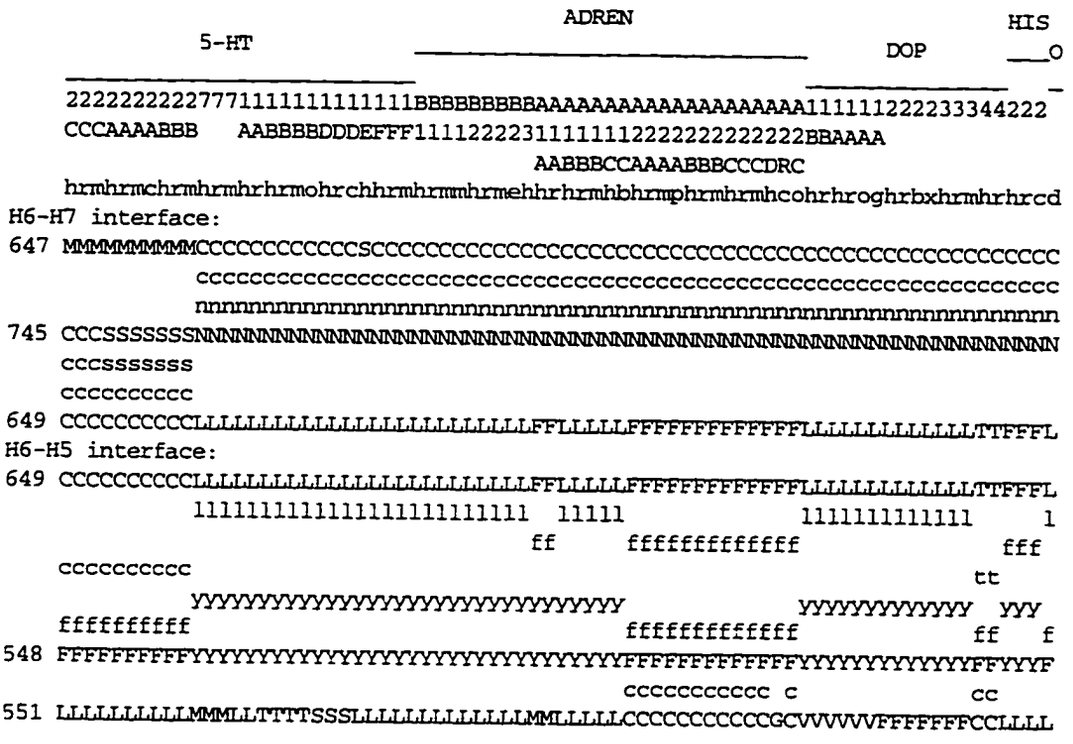
changes upon activation occurring at these PKs. The most drastic behavior was observed for the L6.49(286)->A mutation, which constitutively activated the receptor (Cho 1996).

The criteria presented above for selecting the preferred conformation of the H6 PK can now be integrated to model specific H6 conformations. Modeling the PK conformation of H6 means selecting among PK conformations obtained from MonteCarlo simulations on H6. According to the criteria derived from the observed modulation of the PK conformation by W6.48 adopting the X1=g+ or t rotamer (Section 7.3.3.1), I should select those that kink the cytoplasmic portion of H6 towards H5. This criterion is analogous and thus compatible with the effect of C6.49 in bending H6 towards H5. According to the proposed role of the PK in moving H6 away from H3 at the cytoplasmic boundaries, I should select those PK conformations that are significantly bent. Thus all three criteria are compatible in proposing a significantly bent PK conformation preferentially kinked towards H5 at the cytoplasmic side.

7.2.5- Modulation of the H5-H6 PKs in other GPCRs: A possible source of structural variability and hence a possible indirect mechanism for ligand specificity.

Given the observed flexibility of a PK, and the presence of PKs at H5 and H6 right at the juncture between the binding site and the cytoplasmic portion of the receptor, it is questionable whether a similar backbone conformation can be assumed by the distinct GPCRs, given the observed divergence in the ligand specificities for the various neurotransmitter GPCRs considered in this study. Distinct PK modulatory mechanisms similar to the ones described here for the 5HT2CR could exist in different GPCRs resulting in a different orientations of conserved residues in the binding site, thus providing a mechanism for engineering ligand specificity through evolution different than a direct substitution of binding site residues. Analysis of the conservation pattern for residues proposed to modulate the conformation of the PKs in H5 and H6 described above for the 5HT2CR suggests the presence of different modulatory mechanisms for different GPCRs. I have described the role of C6.49 in modulating the conformation of the PK in H6 based on MonteCarlo simulations of an isolated H6 (see Section 7.2.4). H-bonding of C6.49 to the backbone C=O of the preceding turn induces a bend in H6 towards H5. Table 7.1 below presents the **erms** found for C6.49 at positions within the TMH domain that are amenable to participate in specific H-bonding to the PK in H6.

Table 7.1 : Convergent and divergent H-bonding schemes modulating the PK conformation in H6, and hence the aromatic cluster configuration: Inferences from **erms**.



Thus, whenever the C6.49 present in the 5HT2 subclass is absent in other receptors, other residues change in a correlated manner that provide alternative and feasible PK H-bonding schemes. These C6.49 **erm** can be grouped into two subgroups, depending on whether they occur at the H6-H7 interface (C6.49->L/F, M6.47->C and C/S7.45->N), or at the H6-H5 interface (C6.49->LF, F5.48->Y, L5.51->C). The possible H-bonding schemes of these **erms** is discussed below.

A set of **erms** occurs in the H6-H7 interface: C6.49->L/F is correlated with M6.47->C and C/S7.45->N. Modeling these alternative amino acid patterns at the H6-H7 interface based on the preliminary 7 TMH template indicates that in the absence of the H-bond between C6.49 SH...O=C 6.45 for

non-5HT2 receptors the duo C6.47-N7.45, highly conserved among GPCRs, can participate in specific H-bonding of the carbonyl of residue 6.47, at position i-3 from the PK of H6. This interaction is likely to modulate the conformation of the PK of H6. MonteCarlo simulations on the isolated H6 of the β2-adrenergic receptor, methodologically analogous to the simulations presented for H6 of the 5HT2CR in Section 7.2.4, indicated that C6.47 was H-bonded to its own carbonyl, i-3 from the PK, in 75% of the resulting conformations. This H-bonding interaction required a *psi* angle of C6.47 (-23° on average) lower than the starting α-helical value (-41°), which induced a bending of H6 in the direction towards H5. The bending induced by C6.47 is thus similar in nature to that observed for C6.49 in simulations of the 5HT2CR H6, as shown in Figure 7.6 where the helix axis of 100 representative H6 conformations for both simulations are superimposed. It would appear that the role of C6.49 in the 5HT2R subclass, i.e., bending H6 towards H5, is performed in other GPCRs by the highly conserved C6.47, possibly aided in this function by N7.45. Note that residues M6.47 and C7.45 were highlighted at the level of the primary sequence analysis (Section 4.2) as the two residues where the 5HT2CR deviated more from the pattern of observed amino acids at equivalent positions in the remaining GPCRs aligned. Thus the **erm** presented here and the structural interpretation modulating the H6 PK has provided a specific hypothesis on the molecular basis for the most significant alterations of the conserved pattern present in the 5HT2C receptor. The

hypothesis of a similar role for C6.49 in the 5HT2R subclass and C6.47-N7.45 in other GPCRs is consistent with mutagenesis experiments on the 5HT2CR done in the laboratory of Dr. Stuart Sealfon designed to test this motif. Single mutations of M6.47->C or C7.45->N had no significant functional effect. The double mutant M6.47->C, C7.45->N, also had wild type behavior. This result was unexpected given the high conservation of both a Cys at 6.47 and an Asn at 7.45, and the significant functional effects of their mutation in other GPCRs (van Rhee 1996). These puzzling results (given the high conservation of C6.47 and N7.45 across GPCRs) could be explained by the proposal that C6.49 is fulfilling the same role as C6.47-N7.45. Thus, the mutations simply revert M6.47 and/or C7.45 to the observed residues in other receptors which can only stabilize further a conformation of the PK in H6 that is already stabilized by C6.49. However, this hypothesis would not be experimentally validated until mutation of C6.49->L is shown to have a deleterious effect on receptor function that can be reverted by the additional concomitant mutation of M6.47->C and C7.45->N. This inference was proposed for future testing (Section 11).

Another set of **erm** involving C6.49 is present at the H5-H6 interface: whenever C6.49 is not present in a receptor, this change is accompanied by either a F5.48->Y or a L5.51->C mutation. According to the TMH model discussed above, a Tyr at position 5.48 is oriented towards H6 at the level of the H6 PK, but it is not close enough to H-bond backbone carbonyls within the

H6 PK, in a manner similar to C6.49 in the 5HT2CR. A Cys residue in position 5.51 is best positioned to H-bond the backbone carbonyl of F5.47. Residue F5.47 is positioned i-3 within the PK of H5, and thus an H-bond between C5.51-SH...O=C-F5.47 is likely to affect the conformation of the PK in H5. This suggests that a C6.49, a Y5.48, or a C5.51 could provide different alternatives for modulating the PKs of H5 and H6. These do not result in similar H5-H6 orientations, contrary to the effect noted for the **erms** of C6.49 in the H6-H7 interface described above. It is not possible to know at present how these correlated residues modulate the conformation of the PKs in H5 and H6. Nonetheless, it is conceivable that the resulting PK conformations induced by these divergent interactions would result in different bends and twist on the extracellular segments of H5 and H6, thus inducing a different spatial positioning of the aromatic residues in the H5-H6 interface which have been proposed to interact directly with the ligands in several receptors (van Rhee 1996). This could have significant influence on the ligand-receptor interactions and thus illustrates a possible mechanism by which highly conserved residues may interact differently with the same ligands in different receptors.

This analysis illustrates how observed **erm** can not only be brought close to each other in space, but also how the derivation of specific hypotheses for their interactions can guide the detailed packing of the helix-helix interfaces.

An important consideration arising from this analysis is that the orientation observed for the H5-H6 interface in rhodopsin by electron crymicroscopy may be different to the orientation present in the 5HT2C receptor, depending upon the existence of similar or divergent specific interaction modulating the PK conformations. Rhodopsin contains the highly conserved C6.47, but the conserved N7.45, predicted to be its partner in modulating the PK of H6, is substituted by a Ser residue. In addition, rhodopsin lacks an aromatic residue at position 6.52, which belongs to the aromatic cluster on H6, defined as F6.44-W6.48-F6.52 in Section 7.3.3, where it was shown to affect the conformation of the PK on H6. Therefore, we might expect some structural deviation between rhodopsin and the neurotransmitter receptors in this region, consistent with the very different chemical nature of their respective ligands.

7.3.- Organizing motifs within the TMH domain:

Selection of specific PK conformations as described in the previous section results in an overall defined backbone conformation for modeling purposes, which can now progress towards the detailed side chain packing among the seven TMHs and the ligands. With the packing process, the interhelical interactions become increasingly interrelated and local interactions define specific structural motifs (or receptor "microdomains"). I describe here the major organizing motifs classified according to their

functional correlates: three evolutionary conserved motifs proposed to participate in the receptor activation mechanism (Sections 7.3.1 to 7.3.3), and an evolutionary divergent motif. Residues proposed to participate in direct interactions with neurotransmitter receptor ligands were shown in Figure 2.16 to cluster on the extracellular side of H3-H5-H6-H7: The ligand binding domain.

7.3.1.- The H1-H2-H7 cluster of polar residues: A proposed H2-H7 interaction was validated experimentally through reciprocal mutations on the GnRH and 5-HT_{2A} receptors.

A set of highly conserved polar residues from H1 (N1.50), H2 (D2.50), and H7 (N7.49) cluster together in space within the model defining a motif which has been shown to participate in the process of receptor activation in an interrelated way (Zhou 1994; Sealfon 1995; Perlman 1997).

Our most successful application of the **erms** analysis involved the observation that the highly conserved D2.50 and N7.49 in neurotransmitter receptors, were reciprocally interchanged in the GnRH receptor into N2.50-D7.49. This hypothesis was tested experimentally in the laboratory of Dr. Stuart Sealfon by single and double mutations at this locus which reverted the residues found in the GnRHR (N2.50/D7.49) to the pattern conserved in most other GPCRs (D2.50/N7.49) (Zhou 1994). A similar mutagenesis approach was conducted in the 5-HT_{2A} receptor which further confirmed the

interrelated functional role of D2.50 and N7.49 (Zhou 1994; Sealfon 1995). At the time (1992), these experiments provided the first residue-to-residue double revertant mutant within GPCRs. From the results it was concluded that an interaction between these two residues is required for receptor activation. Because both studies have been published (Zhou 1994; Sealfon 1995), I present here only the implementation of the proposed D2.50-N7.49 interaction in the model of the 5-HT2CR.

Residue N7.49 belongs to the conserved NP motif in H7. I described in Section 7.2.2 the requirements of a PK conformation at this locus with a significant right-handed face-twist. The selected conformation for the NP motif, whose resulting pattern of accessibility was validated experimentally by SCAM analysis of the D2 receptor (Fu 1996), was characterized by a H-bond between the N7.49 side chain and the $C{=}O_{i-4}$. As can be seen in Figure 7.7, the remaining NH group of N7.49 is oriented towards H2 where it can H-bond D2.50. Note that in this conformation, the conserved Y7.53 belonging to the NPxVY motif in H7 can H-bond D2.50.

7.3.2.- Charge-charge interactions at the cytoplasmic side.

The cytoplasmic side of the TMH domain is characterized by the presence of numerous charged and other highly polar residues (see helical nets of the 7 TMHs in Figure 6.9). Due to the energetic significance of the interactions whithin these highly polar residues in the TMH domain, they

are likely to determine the preferred conformations of the TMH domain at the cytoplasmic boundaries. The charged residues present at this loci of the reeptor can be classified for modeling purposses into the Arg/Lys residues predicted to face the phospholipid headgroups and the interior-facing residues, as predicted in Chapter 6 and shown in Figure 6.9. The proposed role of the Arg/Lys facing the phospholipid headgroups is to anchor the TMH domain to the membrane through ionic bonds with the negatively charged phosphate moieties. Therefore, packing interactions among the TMHs at the cytoplasmic side would be guided by interactions among the numerous highly polar residues facing the interior of the TMH bundle. This hypothesis is supported by the higher conservation profile of these interior-facing residues compared to the Arg/Lys side chains oriented towards the membrane. The high conservation pattern of these clusters of charged residues at the cytoplasmic side implies that the interactions among them would also be preserved among GPCRs. Thus, inferences on the structural and functional role of this motif derived from different receptors could be assumed for most GPCRs, as long as the particular side chains involved are preserved among them.

We have studied a subsegment of this motif, the conformation and interactions of the highly conserved Arg at the cytoplasmic boundary of H3, belonging to the conserved DRY motif, referred collectively as the "Arg-Cage" motif. These specific interactions are proposed to participate at different stages

of the receptor activation mechanism. This motif is presented first in the context of an isolated H3 (Section 7.3.2.1), based on computational and mutagenesis techniques applied synergistically to study this motif in the GnRH receptor. The mutagenesis studies were conducted in the laboratory of Drs. Stuart Sealfon and James Davidson. Their results are presented in Table 7.4 as they pertain to the hypothesis developped and this work is still unpublished. The resulting structural hypotheses are adopted for modeling the 5-HT2C receptor, because the main interactions identified within this motif are also present in neurotransmitter receptors.

7.3.2.1.- The Arg-cage motif in TMH 3: Studies on the GnRHR and implications for the 5-HT_{2C}R

The gonadotropin-releasing hormone (GnRH) receptor, which is a member of the rhodopsin-like G-protein coupled receptor (GPCR) family (Tsutsumi 1992; Chi 1993). Sequence alignment of these heptahelical proteins, which include the visual opsins and receptors for neurotransmitters, peptides and glycoproteins, shows that certain amino acids are highly conserved at corresponding positions within the putative transmembrane domains (Probst 1992). These conserved amino acids are found in GPCRs that are activated by diverse stimuli and that couple to various G-proteins. Most likely, the conserved residues contribute to three-dimensional spatial motifs that function as microdomains necessary for mechanisms subserved by these

receptors. One process shared by the GPCRs, that may require similar structural machinery, is the conformational rearrangement underlying receptor activation. This supposition is supported by the results of reciprocal mutation studies of the GnRH and serotonin receptors that demonstrate an interdependent role in receptor activation for conserved residues in transmembrane domains (TMH) 2 and 7 (Zhou 1994; Sealfon 1995).

A clear example of a structural element that is likely to be essential for functional mechanisms shared by the GPCRs is the most conserved residue in GPCRs, the Arg at the cytoplasmic side of TMH 3, that is commonly identified by the sequence pattern DRY. This Arg, referred to here as $R^{3.50(139)}$ (see methods), has been proposed to be involved in the activation mechanism of GPCRs by a number of groups based on results from mutagenesis studies (Min 1993; Scheer 1996) and molecular modeling (Oliveira 1994; Scheer 1996). An understanding of the molecular basis for the functional role of this Arg requires the identification of those residues whose specific interactions with the Arg determine its conformation and ensuing conformational changes upon activation. This task is complicated by the intrinsic flexibility of the Arg side chain, which can adopt more than 81 different rotamer orientations spanning two α-helical turns, thus capable of interacting with many different residues on several TMHs. This set of interacting residues constitutes a three-dimensional motif which functions as a structural microdomain that can confine the flexible Arg side chain to specific, functionally relevant,

conformational states. Thus, we refer to these residues collectively as the "Arg-cage".

In the absence of a known three dimensional structure of the receptor, the integration of these observations from sequence analysis into a mechanistic scheme for receptor function requires molecular modeling (Ballesteros 1995). Such molecular models can provide a rationalization of current experimental data into a structural framework in which the mechanism underlying the functional perturbations induced by mutations can be explored, and they provide the basis for specific and experimentally testable structural hypotheses to explain the observed experimental phenotypes. Using such approaches, several groups have proposed specific partners for R3.50 such as the Asp in TMH2 (D2.50) (Oliveira 1994; Scheer 1996), based on the rationale of a similar conservation pattern and the need to neutralize a positive charge in a low dielectric environment. However, several other suitable interacting residues are present within the 7 TMHs. In particular, analysis of the conservation pattern centered on $R^{3.50(139)}$ identified highly conserved residues that are likely to affect the conformation of $R^{3.50(139)}$. These residues form the consensus sequence (I/L)XXDRYXX(I/V) shown in Figure 7.8. $R^{3.50(139)}$, E or $D^{3.49(138)}$, and I or $V^{3.54(143)}$ are present in all cloned GPCRs belonging to the rhodopsin family. In the GnRH receptor, this conserved motif is represented by IXXDRSXXI (Figure 7.8). In an α-helical environment, the conservation pattern described above forms an envelope of

conserved residues consisting of $I^{3.46(135)}$, $D^{3.49(138)}$, $S^{3.51(140)}$ and $I^{3.54(143)}$, that surrounds the absolutely conserved $R^{3.50(139)}$, as shown in Figure 7.9.

A complete exploration of the vast conformational space of the Arg side chain in the context of a full receptor model is not computationally feasible. We have focused our modeling study on TMH3 alone where a complete conformational exploration by MonteCarlo simulations can be accomplished and where the proposed interactions with the Arg-cage would take place (Figure 7.9). The helical structure and the helix ends of TMH3 have been substantiated experimentally by Cys scanning (Javitch 1995a) and spin labeling experiments (Farahbakhsh 1995). Thus, spin labeling experiments on rhodopsin (Farahbakhsh 1995) showed that TMH3 was helical at least up to residue 3.55 which is the selected TMH3 end in our receptor model. The environment surrounding R3.50 in rhodopsin has been characterized as a interior protein up to residue 3.52, and as a water accessible environment after residue 3.53. For the exploration of the conformational space we have applied a novel biphasic solvent model that reproduces interface between the interior of a protein and a water environment above and below of residues 3.52-3.53, respectively (see Methods Section 2.3.3).

RESULTS

The GPCRs demonstrate a pattern of conservation for several residues that are in spatial proximity to $R^{3.50(139)}$ when the cytoplasmic side of TMH 3

is modeled as a regular α-helix (see Figure 7.9). These potential local sites of interaction for $R^{3.50}$ include the large hydrophobic residue (I/V/L) at 3.46, the acidic residue at 3.49, the Tyr (represented by Ser in the GnRH receptor) at 3.51 and the β-branched large hydrophobic residue (I/V) at 3.54.

The interaction patterns and rotamer positioning of $R^{3.50(139)}$ with respect to these neighboring TMH 3 residues were explored with MonteCarlo simulations for the wild-type helix and for various mutant receptors. Many conformations of the flexible Arg side chain were not attainable due to steric clashes with the helix backbone. For example, all Arg rotamers whose χ1=g- are unpopulated because of a clash between the Arg γ-methyl and the backbone carbonyl from the preceding turn of the helix (McGregor 1987).

The most striking observation to emerge from the simulations is the tendency for $R^{3.50(159)}$ to form an ionic bond with $D^{3.49(138)}$, as illustrated in Figure 7.9.B. Nearly half of the $R^{3.50(139)}$ rotamers observed were bound to this aspartic acid (Table 7.2). As shown in Figure 7.10, a variety of Arg conformations were identified that form the Arg-Asp interaction. Among the remaining residues in the helix, no significant interactions were found. Most other side chains remained static throughout the simulations, consistent with their rotamer populations in known α-helical structures: Ser and Thr residues were overwhelmingly (92-99%) in the χ1=g+ conformation due to H-bonding to the backbone carbonyl of the preceding turn (Gray 1984; McGregor 1987), and β-branched residues (Val, Ile, Thr) except $I^{3.54(143)}$ were

constrained to one single $\chi 1$ population (83-99%) due to clashes with the helix backbone (McGregor 1987).

Table 7.2 :

Preferred conformations of the Arg side chain based on rotamer populations for wild type (WT) and locus D3.49 modified by mutation (D3.49N) or protonation (D3.49DH). Only rotamers populated over 5 % are shown.

interaction	orientation		X1	X2	X3	X4	% WT	% D349N	% D349DH
3.49	3.49		g+	g+	g-	t	11.9	0.0	0.0
3.49	3.49		g+	g-	t	t	11.1	0.0	1.0
3.49	3.49		t	g-	g+	g-	6.3	2.8	0.0
3.49	3.49		t	g-	g-	g+	7.9	0.0	5.1
3.49	3.49		t	g-	g-	t	8.3	0.1	6.5
3.49 total %							45.5	2.9	12.6
Ser H-bond	3.47	3.51	t	t	g-	g+	0.5	7.0	5.9
Ser H-bond	3.51	3.47	t	t	t	g+	0.0	3.2	11.6
Ser H-bond	3.51	3.47	t	t	g-	t	0.0	12.9	10.3
Ser H-bond	3.51	3.47	t	t	g-	g-	0.0	14.0	13.8
Ser H-bond total %							0.5	37.1	41.6
water Cyt.	3.54	3.53	t	g-	t	g+	0.0	7.5	0.5
water Cyt.	3.53	3.54	t	g-	t	t	3.2	21.0	6.6
water Cyt.	3.53	3.54	t	g-	t	g-	0.9	0.9	5.8
water Cyt. total %							4.1	29.4	12.9

The structural effects of several mutations of the residues surrounding $R^{3.50(139)}$ studied are presented below ordered by mutation sites.

$D^{3.49(138)}$: The acidic character is absolutely conserved at this locus, where either Asp and Glu residues are present in GPCR sequences. MonteCarlo simulations were performed for the $D^{3.49(138)}N$ mutant, which would weaken significantly the proposed ionic bond between $R^{3.50(139)}$ and $D^{3.49(138)}$ by neutralizing the charge at this locus. This mutation has dramatic effects on the conformational preferences of the Arg side chain, shown in Table 7.2. In the $D^{3.49(138)}N$ mutant, the Arg side chain is no longer interacting with the 3.49 locus (3 %), and instead two new orientations appear populated, shown in Figure 7.11: 1) $R^{3.50(139)}$ is oriented towards positions 3.47-3.51 where it can H-bond to $S^{3.47(136)}$ and $S^{3.51(140)}$ (37.1 %). 2) $R^{3.50(139)}$ is oriented towards positions 3.53-3.54 where it can be solvated by water at the cytoplasmic boundaries. Because activation of rhodopsin has been shown to involve a proton uptake by E3.49(134) (Arnis 1994), simulations where also performed for the protonated form of the aspartic acid, termed $D^{3.49(138)}DH$. The results yielded a pattern of preferred conformations similar to that of $D^{3.49(138)}N$ (Table 7.2).

$I^{3.54(143)}$: Only Ile or Val residues appear in GPCR sequences at this position. Therefore, the properties absolutely conserved at this locus are β-branched character and bulky hydrophobic. In order to test the structural implications of these two properties, we substituted this residue by Val, Leu, and Ala. An Ala residue lacks both properties, Val is also β-branched and bulky hydrophobic, and Leu is also bulky hydrophobic but not β-branched. Analysis of all rotamers populated over 5 %, shown in Table 7.3, indicates that the prevailing interaction for wild type and the three mutants is still the ionic bond between $R^{3.50(139)}$ and $D^{3.49(138)}$ (65-80 %). Though maintaining the same interaction, the individual rotamer conformations preferred for this interaction vary among the mutants and with respect to wild type (Table 7.3). Notably, a novel orientation for the Arg side chain towards residues 3.53-3.54 appears significantly populated in the $I^{3.54(143)}$A mutant (14.3 %), but not in the wild type, $I^{3.54(143)}$V, or $I^{3.54(143)}$L mutants. This orientation positions the Arg side chain towards the aqueous cytoplasm, as illustrated in Figure 7.12 where a set of representative conformations for the $I^{3.54(143)}$A mutant are shown.

Table 7.3 :

Preferred conformations of the Arg side chain based on rotamer populations for wild type (WT) and I3.54 mutants to Ala, Leu and Val. Only rotamers populated over 5 % are shown.

interaction	orientation	X1	X2	X3	X4	%WT	%I354A	%I354L	%I354V
ionic bond	3.49	g+	g+	g+	t	2.9	6.8	4.1	6.3
ionic bond	3.49	g+	g+	t	g+	3.6	18.9	7.3	3.3
ionic bond	3.49	g+	g+	t	t	3.5	9.9	7.6	5.6
ionic bond	3.49	g+	g+	g-	t	11.9	2.8	4.7	3.2
ionic bond	3.49	g+	g-	g+	t	1.4	0.0	6.9	0.5
ionic bond	3.49	g+	g-	t	t	11.1	5.8	3.3	5.5
ionic bond	3.49	g+	g-	t	g-	0.2	0.5	5.2	1.2
ionic bond	3.49	t	g-	g+	g+	3.8	2.3	4.2	6.7
ionic bond	3.49	t	g-	g+	t	4.3	0.9	8.4	6.5
ionic bond	3.49	t	g-	g+	g-	6.3	3.1	5.5	4.5
ionic bond	3.49	t	g-	g-	g+	7.9	12.5	19.5	15.5
ionic bond	3.49	t	g-	g-	t	8.3	3.0	3.7	16.4
ionic bond total %						65.2	66.5	80.4	75.2
water Cyt.	3.54 3.53	t	g-	t	g+	0.0	12.7	0.0	0.0
water Cyt.	3.53 3.54	t	g-	t	t	3.2	1.6	3.1	5.3
water Cyt. total %						3.2	14.3	3.1	5.3

$I^{3.46(135)}$: The residue at position 3.46(135) is not as conserved as the other conserved residues surrounding the $R^{3.50(139)}$ in TMH3. Although Ile is most commonly found at this site, Leu or Met also occur in different receptors. We substituted $I^{3.46(135)}$ with Ala, Val and Leu to test: 1) the functional relevance of the wild type side chain (by substituting Ala), 2) the role of the β-branched character of the residue (by substituting Val), and 3) the effect of the large hydrophobic side chain (by substituting Leu). MonteCarlo simulations of all three mutant constructs showed that the conformational preferences for the

Arg side chain were similar to those found for the wild type receptor. Therefore, according to our simulations $I^{3.46(135)}$ does not seem to modulate the orientation of $R^{3.50(139)}$ through a direct interaction.

$S^{3.51(140)}$: A highly conserved Tyr residue normally occurs at this position in GPCRs. This residue is capable of H-bonding the Arg side chain in combination with $S^{3.47(136)}$, as shown in MonteCarlo simulations for the $D^{3.49(138)}N$ mutant (Figure 7.11). We tested the functional role of this residue with the mutant construct $S^{3.51(140)}A$. Hydrogen bonding between $R^{3.50(139)}$ and $S^{3.51(140)}$ was observed in the $D^{3.49(138)}N$ mutant receptor, but the mutation $S^{3.51(140)}A$ did not change significantly the orientation of the Arg side chain relative to the wild type construct. Evidently, this H-bond is not energetically competitive with the ionic bond $R^{3.50(139)}$-$D^{3.49(138)}$.

In order to correlate the predicted local structural roles of the conserved TMH3 residues with their effect on receptor function, constructs obtained from the mutation of $R^{3.50}$ and its surrounding residues were tested for their effects on ligand binding and inositol phosphate accumulation in the laboratory of Drs. Sealfon and Davidson. The relative efficiencies of the various receptor constructs obtained by these groups in signal transduction are summarized in Table 7.4, below. I should clarify that I did not participate in any of these experimental approaches, and the results are included because they are still unpublished.

Table 7.4 :

Binding and activation of wild-type and mutant GnRH receptors expressed in COS-1 cells. The K_i and B_{max} values are obtained from competition binding assays (mean ± S.E. from 3-5 experiments). The E_{max} and EC_{50} values are obtained from the PI-hydrolysis experiment (mean ± S.E. from 3-5 experiments). For the definition of efficiency see text. The $D^{3.49(138)}$->N and $R^{3.50(139)}$->Q constructs were performed in the mouse GnRH receptor (m-mouse), while the $I^{3.54(143)}$ and $I^{3.46(135)}$ constructs were performed in the human (h-human) receptor. Mutant construcs whose expression levels were too low to be characterized are marked nd (not determined).

constructs	competition binding GNRH-A		PI assay		efficiency
	$B_{max(\%wt)}$	$K_{i(nM)}$	$E_{max(\%wt)}$	$EC_{50(nM)}$	(% wild type)
h wild type	100	1.6±0.2	100	0.3±0.1	100
$I^{3.54(143)}A$	166±14	6.8±0.5	69±10	9.5±1.3	6
$I^{3.54(143)}L$	45±4	4.6±1.0	66±10	0.8±0.3	160
$I^{3.54(143)}V$	68±9	2.1±0.6	70±7	0.3±0.1	133
$I^{3.46(135)}A$	nd	nd	nd	nd	nd
$I^{3.46(135)}L$	23±1	3.5±1.2	96±22	1.2±0.1	226
$I^{3.46(135)}V$	nd	nd	nd	nd	nd
m wild type	100	0.4±0.07	100	0.45±0.2	100
$D^{3.49(138)}A$	nd	nd	nd	nd	nd
$D^{3.49(138)}N$	56±4.8	0.4±0.08	98±6	0.31±0.1	250
$R^{3.50(139)}Q$	559±112	0.8±0.1	68.3±7	65.2±8	0.3
$R^{3.50(139)}A$	nd	nd	nd	nd	nd
$R^{3.50(139)}K$	nd	nd	nd	nd	nd
$R^{3.50(139)}H$	nd	nd	nd	nd	nd

DISCUSSION

Because all mutants studied that displayed detectable binding had affinities for GnRH similar to wild type (less than 5-fold different), we assume that the overall structure of the GnRH receptor was not perturbed

significantly for any of these mutants. Mutations that did not show detectable binding did not show any detectable activation either, most likely indicating that the structure or the intracellular processing of the receptor was significantly damaged by the mutation.

Oliveira et al (Oliveira 1994) proposed a mechanism for receptor activation in which the change in orientation of this $R^{3.50(139)}$ constitutes an "Arg-switch" leading to activation of GPCRs, a hypothesis further substantiated and expanded recently by a combination of mutagenesis and computational simulations on adrenergic receptors (Scheer 1996; Scheer 1997). Mutation of the Arg residue in the GnRH receptor to Ala, Gln, His, and Lys abolished or severely impaired IP generation, supporting a proposed crucial role of this residue in receptor function.

Study of the interaction between the $R^{3.50(139)}$ and $D^{3.49(138)}$ site of the arginine cage motif - The preferred conformation of the Arg side chain in the wild type construct is determined by an ionic bond with $D^{3.49(138)}$. This constrained was relieved by the $D^{3.49(138)}N$ mutant, which experimentally enhanced receptor efficiency. These results suggests that the $R^{3.50(139)}$ side chain is held by $D^{3.49(138)}$, but not by the $D^{3.49(138)}N$ mutant, through an ionic bond in the inactive state of the receptor, and that receptor activation involves releasing this constraint on the $R^{3.50(139)}$ side chain conformation. This hypothesis is supported by similar mutations D3.49N in an α-adrenergic receptor and

E3.49Q in rhodopsin, which also enhanced receptor activation. For the adrenergic receptor, this mutation actually induced constitutive activation. Rhodopsin showed that receptor activation was accompanied with proton uptake. The authors identified the acidic residue being protonated as the E3.49, based on the lack of proton uptake by the E3.49->Q mutant, which can be considered as a protonated E3.49. We performed the computational experiment of protonating $D^{3.49(138)}$ in the GnRH receptor.

Rhodopsin activation is accompanied by uptake of 2 protons at the cytoplasmic side. Residue E3.49 in rho was identified as one of the residues becoming protonated upon activation, based on the lost of one of these two protonation sites for the E3.49Q mutant receptor, which can be considered as a protonated E3.49. Thus in these two receptors neutralization of the 3.49 locus by either protonation or mutation favors the activated form of the receptor. To rationalize these results in the structural context of the proposed R3.50-D3.49 interaction, we performed the computational experiment of protonating $D^{3.49(138)}$ in the GnRH receptor. Neutralization of the acidic group at position 3.49, which would weaken significantly this interaction, abolished (D3.49N) or dramatically decreased (D3.49DH) the orientation of the Arg side chain towards the 3.49 locus based on MonteCarlo simulations on TMH3 (Table 7.2). Based on these experimental and conformational findings, we hypothesize that the conserved Arg side chain is held by $D^{3.49(138)}$ through an ionic bond in the inactive state of the receptor, and that receptor activation

involves releasing this constraint on the $R^{3.50(139)}$ side chain conformation, most likely by protonation of the aspartic residue.

The consistency and implications of the proposed $R^{3.50(139)}$-$D^{3.49(138)}$ interaction derived from conformational analysis on TMH3 alone were analyzed in the context of a seven TMH model of the receptor. Because the strongest interaction of R3.50 would be with an ionic bond as the one proposed with D3.49, we explored whether an alternative "charge counterpart" (D/E) could be found within other TMHs. A charge counterpart for the Arg would be expected to share its high degree of conservation and to reside at the cytoplasmic side of the TMH. However, the only conserved acidic group in the TMH of the GnRH receptor is the $D^{3.49(138)}$ in TMH3 studied here. There are two non conserved acidic residues present in the TMHs of the GnRH receptor at positions $E^{2.53(90)}$ in TMH2 and $D^{7.49(319)}$ in TMH7. An interaction of $R^{3.50(139)}$ with $E^{2.53(90)}$ is inconsistent with a TMH2-TMH3 engineered Zn binding site reported recently on the NK-1 receptor (Elling 1996) and was therefore discarded. An interaction of $R^{3.50(139)}$ with the non-conserved $D^{7.49(319)}$ could be possible in the context of a seven TMH model of the GnRH receptor.

We have proposed earlier that residue $D^{7.49(319)}$ is a reciprocal mutation of $N^{2.50(87)}$ through evolution (Zhou 1994), meaning that an Asp acid highly conserved among GPCRs in TMH2 (D2.50) has been interchanged in the

GnRH receptor through the reciprocal mutation $D^{2.50(87)}N$ and $N^{7.49(319)}D$. We have substantiated this hypothesis by interchanging these two residues in the GnRH and 5-HT2A receptors (Zhou 1994; Sealfon 1995); the revertant mutant phenotype observed for these constructs supports a direct interaction between $N^{2.50(87)}$ and $D^{7.49(319)}$. Thus it could be that $D^{7.49(319)}$ in the GnRH receptor is substituting the conserved aspartic acid normally found at the $N^{2.50(87)}$ locus in other GPCRs in terms of the interaction with the conserved $Arg^{3.50(139)}$ in TMH3. This hypothesis would resolve the apparent lack of conservation of acidic character observed at the $D^{7.49(319)}$ locus in the GnRH receptor. Exploring this possibility in the context of a model of the TMH of the GnRH receptor supports an interchangeable role between $D^{7.49(319)}$ and $N^{2.50(87)}$ in the interaction with the Arg residue in TMH3, as shown in Figure 7.13.a. The Arg side chain is capable of extending from TMH3 towards TMH2-TMH7 where it can interact simultaneously with $D^{7.49(319)}$ and $N^{2.50(87)}$. This interaction is inconsistent with the $R^{3.50(139)}$-$D^{3.49(138)}$ interaction proposed here, due to the axial displacement away from the cytoplasmic side necessary for the Arg side chain to reach the $D^{7.49(319)}$-$N^{2.50(87)}$ loci, which was never observed in the simulations.

If $D^{7.49(319)}$ and $N^{2.50(87)}$ are indeed interchangeable in terms of their interaction with $R^{3.50(139)}$, then the same or an equivalent interaction should occur when these two locus are interchanged in other receptors, i.e. in the

$N^{7.49(319)}$ and $D^{2.50(87)}$ arrangement. Such an arrangement occurs in the 5-HT2A receptor where we have confirmed the related contributions of these two residues by mutagenesis studies (Sealfon 1995). The interaction between the Arg residue and the 2.50 and 7.49 locus proposed in Figure 7.13.a for the GnRH receptor is shown in Figure 7.13.B as it applies for the serotonin receptor. Because $R^{3.50(139)}$ interacts with the carbonyl groups of $D^{7.49(319)}$ and $N^{2.50(87)}$ (see Figure7.13.A), the same interaction is possible with $N^{7.49(319)}$ and $D^{2.50(87)}$, as shown in Figure7.13.B. Therefore, the set of interactions proposed here for the conserved Arg in TMH3 with the conserved polar residues in TMHs 2 and 7 can successfully rationalize the revertant mutant character we have observed earlier at the TMH2-TMH7 loci based on the symmetrical disposition of their interacting carbonyls. The proposed ability of the 2.50 and 7.49 residues to constrain the Arg residue into a specific conformation suggests that these residues form part of the Arg-cage motif.

Because the proposed interactions of $R^{3.50(139)}$ with either $D^{7.49(319)}$-$N^{2.50(87)}$ or $D^{3.49(138)}$ are mutually exclusive, yet feasible and anchored on our own experimental data in the same receptor system (GnRH), they should exist at different conformational states of the receptor. Based on the results presented here, the $R^{3.50(139)}$-$D^{3.49(138)}$ interaction holds the Arg side chain in the inactive state of the receptor. Based on our earlier mutagenesis data, an interaction between $D^{7.49(319)}$ and $N^{2.50(87)}$ is required for receptor activation.

Thus, both sets of experiments are consistent with the proposal that the $R^{3.50(139)}$ side chain is held in the inactive state of the receptor by the $R^{3.50(139)}$-$D^{3.49(138)}$ interaction, and stabilized in the activated state by the $R^{3.50(139)}$-$D^{7.49(319)}$-$N^{2.50(87)}$ interaction. It follows that the Arg residue may act as a conformational switch intimaly involved in the receptor activation mechanism. Note that both interactions involve ionic bonds and would therefore be highly sensitive to the charge states of the acidic residues involved. Thus neutralization of the acidic residue at position 3.49 would be expected to enhance receptor activation, as found here for the $D^{3.49(138)}$->N mutant, in rhodopsin by direct protonation and mutation (Arnis 1994), and in the α1b-adrenergic receptor by all mutants of $D^{3.49}$ (Scheer 1996; Scheer 1997). Similarly, neutralization of the acidic residue at position 7.49 or 2.50 would be expected to hinder receptor activation, as found in our previous studies on the GnRH ($D^{7.49(319)}N$) and 5-HT_{2A} ($D^{2.50}N$) receptors, as well as similar D2.50 mutations to Ala or Asn in many other GPCRs. Further support for this hypothesis is provided by the enhanced activating profile of the N7.49D mutant in the β2-adrenergic receptor, which would further stabilize the proposed $R^{3.50}$-$D^{7.49}$-$D^{2.50}$ interaction.

Recently, FTIR experiments on rhodopsin indicated that $D^{2.50}$ becomes more strongly H-bonded upon photoexcitation (Fahmy 1993). This would be consistent with our proposal of an ionic interaction between $R^{3.50}$ and $D^{2.50}$

(in rhodopsin) upon activation, and thus provide a molecular rationale for the observed phenotype. It is well known for many GPCRs that binding of cations such as sodium at the $D^{2.50}$ locus inactivates the receptor (Kobilka 1992; Surprenant 1992; Tomic 1993; Vernier 1993). It could be that Na^+ binding at this locus neutralizes the negative charge at the 2.50-7.49 locus and repels the positively charged R3.50 side chain, both effects concomitantly inducing the inactive conformation of the Arg side chain. It follows that residue 7.49, in addition to 2.50, may also be involved in cation modulation.

Oliveira et al (Oliveira 1994) proposed a mechanism in which the change in orientation of this $R^{3.50}$ constitutes an "Arg-switch" leading to activation of GPCRs, an hypothesis further substantiated and expanded recently by a combination of mutagenesis at the $D^{2.50}$ locus and computational simulations on adrenergic receptors (Scheer 1996; Scheer 1997). Scheer et al. (Scheer 1996; Scheer 1997) proposed that receptor activation involves a change in the orientation of the $R^{3.50(139)}$ from a pocket of polar residues in TMH1, 2 and 7, which includes the positions 2.50 and 7.49 discussed above, towards the aqueous cytoplasm. Our finding that mutation $D^{3.49(138)}N$ enhanced activation follow the same trend than the experimental results reported by Scheer et al. in that mutation of $D^{3.49}$ to any other amino acid induced constitutive activation of the receptor (Scheer 1997). We are also in accordance with their proposal of an interaction between $R^{3.50(139)}$ and the

$D^{7.49(319)}$-$N^{2.50(87)}$ locus. We disagree with Scheer et al. in that the interaction of $R^{3.50(139)}$ with $D^{7.49(319)}$-$N^{2.50(87)}$ would represent for us the activated, and for them the inactivated, state of the receptor. Additionally, we postulate a direct interaction between $R^{3.50(139)}$ and $D^{3.49(138)}$ while they proposed an indirect mechanism by which protonation or neutralization of $D^{3.49(138)}$ alters the interactions of the Arg residue. This illustrates the notion that although theoretical approaches to elucidate the structural basis of receptor function can converge in proposing a set of specific interactions, experimental data are necessary to assign these proposed interactions to specific functional states.

Study of the interaction between the $R^{3.50(139)}$ and $I^{3.54(143)}$ site of the arginine cage motif - This locus shows a 100% conservation profile of a β-branched, bulky hydrophobic residue (Ile or Val), that parallels the conservation profile of $R^{3.50(139)}$. In order to test these two structural parameters, $I^{3.54(143)}$ was mutated to Val, Leu, and Ala. Experimentally, the $I^{3.54(143)}$V and the $I^{3.54(143)}$L mutants behaved like wild type, while the $I^{3.54(143)}$A mutant had only 6 % of the wild type efficiency. Conformationally, the preferred orientations of the Arg side chain in the $I^{3.54(143)}$V and the $I^{3.54(143)}$L mutants were similar to wild type. However, in the $I^{3.54(143)}$A mutant a new orientation of the Arg side chain was significantly populated (14.3 %, Table 7.3). Analysis of these rotamer conformations populated in the $I^{3.54(143)}$A mutant but not in the

other $I^{3.54(143)}$ constructs provided a rationale for the observed phenotype: The bulky side chain of an Ile, Leu, or Val residue at this position would clash with the Arg side chain when this residue adopts the rotamer (t, g-, t, g+), as shown in Figure 7.14. An Ala amino acid at this position, lacking the side chain, would allow this conformation of the Arg residue, as can be seen in Figure 7.12. According to the membrane-water boundary determined experimentally for rhodopsin, located between residues 3.52 and 3.53, the (t, g-, t, g+) rotamer of the Arg would orient the charged guanidinium group towards the aqueous cytoplasm where it could get solvated. A solvated Arg would not be able to participate in specific interactions within the transmembrane domain required for function, such as the interaction of $R^{3.50(139)}$ with the $D^{7.49(319)}$-$N^{2.50(87)}$ loci proposed above. Thus, the role of a conserved bulky residue, preferably β-branched, at position 3.54 could be to prevent the Arg side chain from reaching the aqueous cytoplasm where it would be solvated. The local character of the steric hindrance imposed by $I^{3.54(143)}$ on the $R^{3.50(139)}$ side chain, which depends only on assuming an α-helical conformation at the $R^{3.50(139)}$-$I^{3.54(143)}$ turn, is expected to prevail in the context of the full receptor. Thus $I^{3.54(143)}$ is proposed to form an intrinsic part of the Arg-cage motif.

Study of the interaction between the $R^{3.50(139)}$ and $I^{3.46(135)}$ site - Simulations

of wild type and $I^{3.46(135)}$ mutants to Ala, Val, and Leu showed similar orientations of the Arg side chain and the surrounding residues. This suggests that the conservation pattern observed at this locus is better explained by interhelical interactions of $I^{3.46(135)}$ not considered in this study. Mutagenesis experiments show that mutation of $I^{3.46(135)}$ to either Ala or Val is not tolerated, while substitution by Leu shows wild type behavior. This pattern of tolerance in the amino acid substitution suggests that a methyl at position delta along the chain is required for the functional integrity of the receptor. Based on these results, $I^{3.46(135)}$ does not seem to form a part of the Arg-cage motif.

Study of the interaction between the $R^{3.50(139)}$ and $S^{3.51(140)}$ site of the arginine cage motif - Although there is a possible H-bonding interaction between $R^{3.50(139)}$ and $S^{3.51(140)}$, the interaction energy is not competitive with the strong $R^{3.50(139)}$-$D^{3.49(138)}$ interaction which predominates in the wild type receptor. Removal of the H-bonding group of $S^{3.51(140)}$ by substitution to Ala gave no detectable alteration of the functional properties of the GnRH receptor, suggesting that this interaction never occurs, or if it occurs its energetic contribution to the receptor activation mechanism is not significant.

In conclusion, the orientation of the highly conserved $R^{3.50(139)}$ side chain is constrained by interactions with the neighboring conserved residues

$D^{3.49(138)}$ and $I^{3.54(143)}$, which are thus predicted to form part of the Arg-cage motif. Based on the results presented here, we propose that the $R^{3.50(139)}$ side chain is held by $D^{3.49(138)}$ through an ionic bond in the inactive state of the receptor, and that receptor activation involves releasing this constraint on the $R^{3.50(139)}$ side chain conformation. Once the Arg side chain is freed from the $D^{3.49(138)}$ interaction, residue $I^{3.54(143)}$, through a steric clash, prevents it from adopting a conformation where it could reach the aqueous cytoplasm and become solvated. Instead, $R^{3.50(139)}$ is attracted by the negative charge of the $D^{7.49(319)}$-$N^{2.50(87)}$ loci. The resulting $R^{3.50(139)}$-$D^{7.49(319)}$-$N^{2.50(87)}$ interaction stabilizes the active state of the receptor, as supported by an increasing wealth of experimental data, and can explain our previously observed revertant mutant phenotype at the 7.49-2.50 locus of the GnRH and 5-HT_{2A} receptors. The role of the Arg-cage motif is to constrain the inherent flexibility of the absolutely conserved Arg side chain into specific conformations suitable for function.

7.3.3.- The aromatic cluster motif on H6-H5-H4.

A salient feature of the TMH bundle is the spatial clustering of several highly conserved aromatic residues from H6 (F6.44-W6.48-F6.52), H5 (F5.47-F5.48), and H4 (W4.50) within the receptor transmembrane core, underneath the ligand binding domain (see below) and above the G-protein coupling

domain (see below). Its position makes it likely that this aromatic cluster would be involved in transducing the signal triggered by agonist binding into the G-protein coupling domain at the cytoplasmic boundaries, a hypothesis supported by multiple mutagenesis studies in several GPCRs at this locus which consistently affect receptor activation (van Rhee 1996), and by simulations and spectroscopic data described below for specific residues. In modeling the conformation of these side chains and the ensuing aromatic-aromatic interactions, specific criteria were developed pertinent to intrahelical interactions relating to the aromatic residues in H6 and H4, as described below in Sections 7.3.3.1 and 7.3.3.2, respectively.

7.3.3.1- The aromatic rotamer switch in H6. A cluster of aromatic residues in TMH 6 of the GnRH receptor and many GPCRs functions as a "rotamer switch" that contributes to receptor activation.

H6 in neurotransmitter GPCRs has a 100% conserved cluster of aromatic residues at positions F6.44, W6.48, F6.51, F6.52, which define a continuous aromatic patch on the same face of the helix surface (see Figure 6.9). The high degree of conservation of this aromatic cluster makes it plausible that these side chains constitute a structural domain of functional importance.

Each aromatic side chain in an α-helix can, in theory, exist in either of two rotamer conformations in terms of their X1 angles, *trans* or *gauche+*. (See Section 2.1 and (McGregor 1987)). MonteCarlo simulations following

procedural steps as described in Section 2.3 were carried out on isolated H6 segments, exploring alternative rotamer configurations of these aromatic side chains. Data from analyzing these computational simulations and known protein structures demonstrate that the conformation of the side chains of F6.44, W6.48, and F6.52 in the aromatic cluster are interdependent. When the W6.48 side-chain adopts the *trans* configuration, it forces F6.52 to adopt a *trans* configuration as well. When the W6.48 side-chain adopts the *gauche* configuration, it forces F6.44 to adopt a *gauche* configuration as well. (W was not found in the *gauche+* configuration in the PDB search). This interdependence is supported by results from a search in all known protein structures (PDB) for the aromatic side chains spaced at (i, i+4), and by the results of MonteCarlo simulations of an isolated H6 in which I determined the correlation of the rotamer positioning, as shown in Table 7.5, below.

Table 7.5: Effect of W6.48 $\chi 1$ configuration (trans or *gauche+*) on the $\chi 1$ configuration of F6.44 and F6.52.

	From PDB search		From MC Simulations	
	W6.48 trans	W6.48 *gauche+*	W6.48 trans	W6.48 *gauche+*
F6.44 trans:*gauche+*	5:0	0:0	53%:47%	18%:82%
F6.52 trans:*gauche+*	6:0	3:2	99%:1%	42%:58%

Thus W6.48 acts as a conformational switch which directs the conformation of the conserved aromatic cluster in H6, between a *trans* and a *gauche+* configuration. This aromatic cluster surrounds the conserved PK in H6, which was proposed in Section 7.2.4 above to modulate the conformational changes responsible for activation. In fact, simulations of H6 with the aromatic cluster in either of the two configurations described above show that the configuration of the aromatic cluster modulates the conformation of the PK in H6. Simulations of H6 with the aromatic cluster adopting the *gauche+* configuration resulted in PK conformations preferentially bent towards H5, as compared with the analogous simulations with the aromatic cluster adopting the *trans* configuration, which induced PK conformations preferentially bent towards H7, as shown in Figure 7.15.

These observations raise the possibility that the linked repositioning of these aromatic side chains between *trans* and *gauche+* rotamer positioning represents a **toggle switch** that accompanies receptor rearrangement between inactive and active conformers. This hypothesis has been supported recently by elegant work combining mutagenesis and spectroscopic techniques on rhodopsin (Lin 1996). It was found that the indole plane of W6.48 changes its orientation upon activation by tilting towards the membrane plane. This changes resembles the proposed conformational change in the aromatic cluster observed in H6 simulations from a *gauche+* to *trans* configurations, as shown in Figure 7.15; Note that the orientation of the indole plane with

respect to the membrane plane is nearly perpendicular when W6.48 adopts the X1=g+ rotamer (Figure 7.15.a), and that it is parallel when W6.48 adopts the X1=t rotamer (Figure 7.15.b). It follows that the inactive state of the receptor is likely to be characterized by a *gauche+* configuration of the aromatic cluster in H6 (W6.48 X1=g+, F6.44 X1=g+), while the activated state of the receptor would then be characterized by a *trans* configuration of the aromatic cluster in H6 (W6.48 X1=t, F6.52 X1=t).

7.3.3.2.- A conformationally constrained Trp in H4: role of the conserved β-branched residue at position 4.46.

The side chain conformation of the 100% conserved W4.50 is modulated by I4.46, a residue whose β-branched character (Ile, Val) is 96 % conserved (see Table 4.2) through evolution in neurotransmitter GPCRs. The special conformational restrictions of β-branched residues in α-helices where they are constrained to one single rotamer population, and the ensuing capability to modulate the conformation or nearby residues have been described in detail in the context of the Arg-cage motif (Section 7.3.2). Side chain configurations of nearby residues which would clash with the β-branched residue are best relieved energetically by altering the nearby residue and not the β-branched amino acid. The aromatic moiety of W4.50 would clash with I4.56 if it adopts the X1=g+ rotamer, suggesting that the role of I4.56 is to prevent this orientation of W4.50 and thus constraining the aromatic side

chain to a single rotamer, defined by X1=t. This hypothesis has been substantiated by analysis of known protein structures at high resolution (2.5 A), which shows that when a β-branched residue occurs in an α-helix in the fourth residue position (i-4) preceding a Trp, in 8/11 cases the Trp adopts the X1=t rotamer. Furthermore, in 7 out of these 8 cases, the X2=-90 suggesting that the conformational restriction of a β-branched residue at i-4 may extend to the X2 dihedral angle of Trp. Thus I have modeled W4.50 using X1=t and X2=-90 rotamers. Because this conformational restriction is of local (intrahelical) character, it is expected to prevail in the context of the full receptor structure.

The analysis presented above suggests that the conformation of W4.50 would be constant throughout the conformational changes the receptor may undergo, including those pertaining to the receptor activation mechanism. This hypothesis is supported by a combination of spectroscopic and mutagenesis studies on rhodopsin which found that the orientation of the aromatic side chain of W4.50, judged by its transition dipoles, was unchanged upon photoexcitation (Lin 1996).

7.4- Optimization of helix-helix packing for the TMH bundle, including serotonin and LSD as prototype ligands: specific interactions define receptor microdomains, which are then used as building blocks whose assembly determines the TMH domain structure modeled.

The organizing motifs defined in the previous section, initially defined based on few conserved or functionally relevant residues, are subsequently expanded through the process of optimizing helix-helix packing, described in this section 7.4 and organized as follows: I start by modeling the proposed interaction between H3 and H6 at the cytoplasmic side (Section 7.4.1). The cytoplasmic segment of H5 can then be added to the H3-H6 template (Section 7.4.2). The resulting positions and orientations of the extracellular segments of H3-H5-H6, determined by the available PK conformations for H5 and H6, are then refined (Section 7.4.3). H4 can then be incorporated into the modeled H3-H5-H6 template resulting in the agonist binding site proposed in the literature (Section 7.4.4). Serotonin and LSD are used as prototype ligands and included as intrinsic components of the receptor modeling process (Section 7.4.5). The proposed ligand interaction sites can be completed by the incorporation of the extracellular segment of H7, comprising residues 7.35-7.46 (Section 7.4.6). H2 is then added to the H3-H4-H5-H6-H7ec template (Section 7.4.7). The segment corresponding to the controversial NP motif in H7, comprised by residues 7.46-7.53 and proposed to interact with D2.50, can now be refined in detail within the context of the H3-H4-H5-H6-H7ec-H2 template already modeled (Section 7.4.8). H1 is then incorporated into the remaining TMH domain already modeled (Section 7.4.9). Finally, the cytoplasmic extension of H7, residues 7.54-7.70, is incorporated into the model (Section 7.4.10).

7.4.1- Packing H3-H6 at the cytoplasmic boundaries

Experiments on bovine rhodopsin involving Cys crosslinking and Spin labeling (Farrens 1996), and engineered Zn binding sites (Sheikh 1996), have all shown a proximity between H3 and H6 at the cytoplasmic boundaries in the inactive state of the receptor, as reviewed in Section 2.2. Taken together, these experiments point to interactions between residues 3.53-3.54 in H3 and 6.30-6.31-6.32-6.33-6.34 in H6. Although the observation that Cys-Cys crosslinking can occur between C3.54 and a Cys at any position between 6.30 to 6.34 appears to contradict a helical conformation of the segment between 6.30 and 6.34, work by the same authors supports my prediction of an α-helical conformation at this segment from Spin labeling experiments (Yang 1996b). Therefore, the model maintains the predicted α-helical character for H6 and the experimental data from Cys crosslinking are interpreted to indicate a proximity between H3 and H6 at these loci. It should be noted that the above mentioned crosslinked rhodopsins showed near wild type absorbance, an indication that the retinal chromophore was not significantly perturbed. This is not inconsistent with my interpretation of a loss of α-helical character induced by some of these crosslinks at the cytoplasmic boundaries of H6, because rhodopsin has been shown to maintain near wild type absorbance under conditions of significant structural perturbations (see Section 2.2 and references (Sikora 1994; Strassburger 1997)). Examples include a 25 % loss in

α-helical structure upon solubilization and reconstitution (Sikora 1994), as well as thermal stability in the absorbance spectra from 4° up to 90° or more (Shnyrov 1988), depending on the specific conditions. Thus, in interpreting biophysical experiments on rhodopsin such as the spin labeling considered above, as well as other spectroscopic measurements that require solubilization and reconstitution of rhodopsin for sample preparation, we should expect some degree of structural perturbation in regions far from the retinal binding site, such as the cytoplasmic boundaries considered here.

Modeling the H3-H6 interaction between residues 3.53-3.54 and 6.30-6.34 in the 5-HT2CR brings the conserved R3.50 in close proximity to the conserved E6.30. Modeling supports the formation of an ionic bond between these two residues. The rotamer conformation for R3.50 used for this interaction is the one that emerged among the most populated rotamers in the MonteCarlo simulations on the Arg-cage motif presented in Section 7.3.2. This Arg conformation supports the R3.50-D3.49 ionic bond proposed earlier in the study of the Arg-cage motif (see Section 7.3.2). Thus, I propose that the 100% conserved R3.50 interacts in the inactive state of the receptor with two acidic groups, E6.30 and D3.49, as shown in Figure 7.16 in the context of an expanded Arg-cage motif. This hypothesis is supported by a variety of experiments in other GPCRs reviewed below:

a) Mutations at the 6.30 locus: All available mutations of the acidic residue at position 6.30 constitutively activates the receptor, i.e. D6.30G for the

LH/CG and D6.30G for the TSH GPCRs (van Rhee 1996).

b) Deletion of E6.30: For the β2-adrenergic receptor, deletion of the cytoplasmic end of H6 including the E6.30 completely converted the receptor to its high affinity form, though it did not couple to G-proteins.

c) It has been shown that rhodopsin activation involves protonation of two acidic groups, one of which was identified as a Glu at position 3.49. I propose here that R3.50 is held in the inactive state by two acidic groups, D3.49 and E6.30, and that activation involves freeing R3.50 from these two acidic groups. The experimental observation would be consistent with my proposal if the second acidic group protonated upon rhodopsin activation would correspond to E6.30, which is also present in rhodopsin.

d) Mutation of K6.31H in the α1b-adrenergic receptor leads to a constitutively activated receptor. In the E6.30-R3.50 interaction modeled above, K6.31, one residue away from E6.30, can interact with E6.30 through an ionic interaction, thus further stabilizing the modeled E6.30-R3.50-D3.49 network. Notably, this additional interaction would provide a molecular basis for the activating effect observed for the K6.31H mutation, because the side chain of the substituting His is too short to reach and stabilize E6.30 in an α-helical conformation as required for the E6.30-R3.50 interaction.

The K6.31 is predicted to face the phospholipid headgroups in Section 6.1, and this same residue is now predicted to participate with interior-facing residues in a role significantly related to the activation mechanism. This

indicates that the application of the Arg/Lys motif for the prediction of interior facing residues in the initial stage was overstated, as supported by the high conservation of positive charge character (Arg, Lys) at the 6.31 locus (83 %) and the other indications that K6.31 should be considered as an interior facing residue. Note that for the 5HT2CR, the incorporation of the equivalent R6.31 as an interior facing residue results in a more coherent face of H6 facing inside the TMH bundle; before E6.30 was left alone on a helical face that otherwise showed two residues per turn predicted to point inside the protein.

The implications of the misprediction of R6.31 as facing the phospholipid headgroups prompts some reconsideration of the use of the Arg/Lys motif at cytoplasmic boundaries to predict the TMH end and orientation, a novel criterion discovered during this work. The presence of Arg and Lys, together with other amino acid residues, at position 6.31 through evolution fits the conservation pattern observed in known membrane protein structures for residues facing the phospholipid headgroups. However, it has a relatively high conservation of positive charge character that may suggest otherwise. There is a discrepancy between the criteria of conservation of amino acid character (Arg, Lys, His, so not conserved) and the criteria of conservation of a specific physico-chemical property (83% positive charge). I conclude that sequence analysis alone, in the absence of a structural template to analyze the hypothesis, is not sufficient for accurate prediction of the orientation in these cases. Therefore, given the significant effect that a Lys or Arg residue would

have facing the interior of the TMH bundle, the initial prediction should be maintained unless proven otherwise. However, special care should be taken with Arg/Lys positions at the cytoplasmic side that are highly conserved in their charged character, for it is an indication that they could be participating in important interactions with other residues, possibly facing the protein interior.

The conservation pattern of E6.30 in neurotransmitter receptors parallels the 100% conservation of R3.50 and D3.49, thus supportive of the proposed interactions. However, a Glu residue at position 6.30 is not conserved among non-neurotransmitter GPCRs, while R3.50 and D3.49 are, which suggests that the R3.50-D3.49 interaction is more generalizable to other receptors than the E6.30-R3.50 interaction. It should be noted that the D3.49-R3.50 interaction maintains R3.50 in a similar orientation than the D3.49-R3.50-E6.30 interaction. This lack of conservation at the 6.30 locus was specially significant in my studies on the "Arg-cage" motif described in Section 7.3.2, because the theoretical predictions were combined with mutagenesis experiments to study the GnRH receptor, which lacks an acidic group at position 6.30. Because E6.30 is 100% conserved among rhodopsin and the neurotransmitter receptors, and due to the wealth of experimental data supporting the E6.30-R3.50 interaction, I interpret the lack of an acidic residue at this locus for more divergent receptors as indicative that different molecular interactions may be used in different receptor subfamilies to modulate an otherwise

preserved activation mechanism. An example could be the GnRHR where a Thr residue at position 6.49, i-1 from the P6.50, may play a similar role by inducing a bend on H6 that brings the cytoplasmic end of H6 closer to H3 (see Section 7.2.4), thus achieving a similar structural effect than the E6.30-R3.50 interaction in neurotransmitter receptors (see below). It is also conceivable that for more divergent receptors a different, but as yet undefined molecular mechanism is used to maintain the receptor in the inactive state.

It is interesting that according to the H3-H6 interactions at the cytoplasmic side modeled as described above, the ionic interaction between E6.30 and R3.50 would be responsible for holding H6 close to H3 at the cytoplasmic boundaries in the inactive state, as observed experimentally for rhodopsin (Farrens 1996; Sheikh 1996). I proposed that the observed movement of H6 away form H3 at the cytoplasmic boundaries was facilitated by the flexible hinge provided by the conserved P6.50, as discussed in Section 7.3.3 and shown in Figure 7.1, where different PK conformations would result in H6 moving away or closer to H3. It would seem that releasing the E6.30-R3.50 interaction upon activation would remove the "ionic lock" between H3-H6 and thus result in a less constrained, more flexible H6 PK, explaining the observed increased in solvent accessibility and side chain mobility at the H6 cytoplasmic boundaries for rhodopsin upon activation (Altenbach 1996).

It is noteworthy that the H3-H6 interaction at the cytoplasmic side modeled above positions the proposed ligand binding residues in H3 (D3.32,

S3.36) and H6 (F652, N6.55) in spatial proximity, i.e. at a similar height. The resulting H3-H6 contacts are consistent with the pattern of buried versus exposed residues I predicted above (see Section 6.1), as can be seen in Figure 7.16 where residues predicted to face the lipid are colored in yellow.

7.4.2- Packing H5 into the H3-H6 cytoplasmic template.

Incorporation of H5 into the H3-H6 cytoplasmic template described above is guided by results from Cys cross-linking experiments in rhodopsin, which linked residues L6.37C-Y5.58C and L6.37C-F5.59C. Modeling the 100% conserved Y5.58 close to L6.37 positions the Tyr side chain on top of the R3.50-E6.30 interaction described above. In this conformation, Y5.58 could stabilize the R3.50-E6.30 interaction by pi-pi interactions, a role frequently observed for Tyr residues in known protein structures (Flocco 1994). Modeling a stacking interaction between the phenyl ring of Tyr and the guanidinium group of Arg indicates that the Y5.58-OH is positioned to H-bond a carboxylic oxygen of D3.49, as illustrated in Figure 7.16. This proposed interaction of Y5.58 with D3.49 and R3.50 requires a X1=g+ rotamer conformation for the Y5.58 side chain, because a X1=*t* rotamer would orient H5 parallel to the membrane plane and thus would be incapable of fulfilling other observed H5-H6 crosslinking at the extracellular side (Yu 1995; Kono 1996) (it would alter as well the pattern of lipid accessibility predicted and confirmed experimentally for H5 at the cytoplasmic boundaries). Modeling Y5.58 as adopting a X1=g+

rotamer and participating in the specific interactions with D3.49-R3.50 described above, defines the orientation of nearby residues within the H3-H5-H6 cytoplasmic interface. An important example is T5.61, because its 100% conserved β-branched character suggests an important structural and/or functional role. Modeling the Y5.58-D3.49-R3.50 interaction described above positions T5.61 facing H3 between the side chains of residues R3.50 and I3.54. The rotamer conformation modeled for R3.50 as described above adopts an arc-shape as seen from the T5.61 position, creating an empty space underneath the R3.50 side chain that is flanked by the I3.54 side chain, and the dimensions of this empty space between R3.50 and I3.54 are enough to accommodate the T5.61 side chain. Modeling indicates that the T5.61 side chain can be easily docked into the space comprised between the R3.50 and the I3.54 side chain, as illustrated in Figure 7.16, maintaining the Y5.58-D3.49-R3.50 interaction proposed above. This suggests that the role of the conserved T5.62 is to pack against the R3.50 side chain so as to: 1) induce a conformation of the Arg side chain that favors the interactions of R3.50 with D3.49 and E6.30 simultaneously as described above, and 2) prevent the Arg side chain from adopting a rotamer conformation oriented towards T5.61 and I3.54. I described in Section 7.3.2 that an orientation of the R3.50 side chain towards I3.54 would allow the R3.50 side chain to become solvated. This occurrence would be deleterious to the functional role of R3.50 because it would incapacitate this residue, once solvated, from participating in specific

interhelical interactions required for function, such as those proposed here. Consequently, the proposed role for T5.61 would explain the conservation of the beta-branched character at this locus, because the inherent lack of conformational flexibility of these residues may endow them with the role of modulating the conformation of nearby residues, as described in detail in Section 7.3.2. This role of T5.61 in "caging" the R3.50 conformation is reminiscent of the role of I3.54 proposed in Section 7.3.2 in terms of the "Arg-cage" motif. As shown in Figure 7.16, T5.61 is positioned between the R3.50 side chain and I3.54, thus fulfilling a role similar to that proposed for I3.54 in the context of H3 alone (see Section 7.3.2). This set of proposed interactions contains all the 100% conserved residues present at the cytoplasmic H3-H5-H6 interface, and can thus be considered as forming part of a common structural motif, to which we refer as the extended "Arg-cage" motif for its role in constraining the R3.50 side chain conformation (see Section 7.3.2). The interactions proposed above are consistent with the predicted accessibility pattern, as shown in Figure 7.16 where residues predicted to face the lipid are colored in yellow.

The "Arg-cage" motif presented above could provide a molecular basis for several constitutively activating mutations (CAM) described in the literature. I described in Section 7.3.2 how the D3.49-R3.50 interaction could explain the observed CAM resulting from substituting D3.49 in several receptors, most notably in the α1b-adrenergic receptor where all 19 amino acid

substitutions show a CAM phenotype (Scheer 1996; Scheer 1997). Mutations of A6.34 in the α1b-adrenergic receptor to all 19 amino acid substitutions also induced a CAM phenotype (Kjelsberg 1992). An equivalent L6.34A mutation in the β2-adrenergic receptor resulted also in a CAM phenotype (Samama 1993). Residue 6.34, which also belongs to the "Arg-cage" motif according to the considerations given above is directly interacting with Y5.58 as it stacks against R3.50-E6.30. In the process of modeling this motif the most difficult part was to avoid steric clashes between S6.34 in the 5HT2CR and Y5.58 as it packed towards the R3.50-E6.30 interaction. It is conceivable that mutations of residue 6.34 in these receptors perturbs the Y5.58-R3.50-E6.30 interaction which holds H3-H6-H5 together in the inactive state of the receptor, thus promoting the reported CAM phenotype. Note that according to this interpretation the set of mutations at positions 3.49, 6.31, and 6.34 inducing a CAM phenotype could all be related to a destabilization of the Arg-cage motif. The idea that these mutations represent a region that constrains the receptor in the inactive conformation was already presented by the authors of these experiments, e.g. see (Kjelsberg 1992), and it receives here a molecular rationale.

7.4.3- Modeling the extracellular segments of H3-H5-H6:

The relative positioning of H3-H5-H6 in the axial direction that result from the above mentioned interactions at the cytoplasmic side is consistent

with the required interhelical interactions at the extracellular side, as expressed in the relative positions of the ligand binding residues, and the results of Cys crosslinking (Yu 1995; Kono 1996) and Zn binding sites (Elling 1995; Elling 1996; Sheikh 1996) (see Figure 2.16 and Table 2.3). Thus, optimization of the packing of the extracellular portions of H3-H5-H6 is performed starting from the backbone orientations at this region that result from modeling the Arg-cage motif at the cytoplasmic side as described above. The most significant conformational degrees of freedom available to modify the direction of the backbone for satisfying the H3-H5-H6 packing requirements are the Pro-kink (PK) conformations of H5 and H6. In the initial stages of modeling the 7 TMHs packing according to experimentally derived interactions (Section 7.3), I developed criteria that guides the selection of specific PK conformations in H5 and H6. The PK of H6 was predicted significantly bent and oriented towards H5. In contrast, H5 was predicted minimally kinked by P5.50. These guidelines determine the range of PK conformations available in incorporating the extracellular segments of H5-H6 on the template defined by the cytoplasmic packing of H3-H5-H6.

The modeling of H5-H6 interactions at the extracellular side, above the PK in H5 and H6, is guided by the extensive Cys-crosslinking studies on rhodopsin (Yu 1995; Kono 1996) and the Zn binding sites engineered on the NK1 receptor (Elling 1995). Interhelical contacts between these segments are dominated by the presence of bulky aromatic side chains highly conserved in

both H5 and H6, which form an aromatic cluster that expands the previously characterized aromatic cluster on H6 (see Section 7.3.3). Aromatic clusters are found in known protein structures where they pack against each other in a T-shape interaction instead of stacking geometries, and are highly conserved through evolution, serving an important functional role (Burley 1985). The aromatic cluster in GPCRs is certainly highly conserved, and I have proposed for it an important functional role (Section 7.3.3); its packing is modeled in the above mentioned T-form.

The aromatic residues in H6 oriented towards H5, as defined in Section 6.2.2, are W6.48 and F6.52. Because I have selected the g+ configuration for these residues within H6 (Section 7.3.3), the X1 side chain rotamer of W6.48 is fixed but F6.52 can adopt either the *trans* or the g+ rotamer. The accessibility pattern and the ensuing H6 orientation relative to H5 and the remaining helices indicates that only in the g+ configuration can F6.52 face towards the binding site to interact with the ligand, as proposed for the 5HT2AR based on mutagenesis experiments. Therefore, F6.52 is modeled adopting the g+ rotamer. The conformation of the aromatic residues in H5, F5.47 and F5.48 are defined as follows: F5.48 is exposed to the lipid milieu although it is 100% conserved. This apparent dichotomy can be explained by noting that when F5.48 adopts the g+, but not the X1=t conformation it can participate in specific aromatic-aromatic interactions with F5.47 and F6.52, although it is overwhelmingly exposed to the lipid milieu. This would explain the high

conservation observed for F5.48 and I therefore select the X1=g+ for this residue. F5.47 can adopt either the X1=t or g+ conformation, but only in the t configuration can it interact with F5.48 which otherwise would have no partner. F5.47 is likely to be involved in ligand binding. In the X1=t F5.47 would be able to interact in a T-shape geometry with the aromatic moiety of the ligand, assuming the proposed H-bonding interaction between the ligand and S5.46 and described in Section 7.4.5. In the g+ conformation, there would be a significant clash with the ligand if the kink at P5.50 is very small as selected for the inactive conformation. However, there are feasible *phi /psi* angles around F5.47 that would allow a favorable stacking interaction with the ligand, but this PK conformation would be inconsistent with the remaining constraint already imposed on the model, so I selected the X1=t for F5.47.

The resulting H5-H6 packing at the extracellular ends is shown in Figure 7.17. The aromatic cluster fills the space in between these two helices interacting in a T-shape geometry. A significant consequence of the presence of bulky aromatic side chains defining the H5-H6 interface is that the distance between H5 H6 in this region is significantly larger that the 10 A normally found between adjacent helices. This could explain the rhodopsin density map in this region, where the distance between H6 and H4 is too large to be filled by a single helix (H5) (Schertler 1995). The aromatic side chains provide a significantly expanded helix-helix interaction surface. The predicted

accessibility pattern is consistent with this H5-H6 interface, as can be seen in Figure 7.17 where residues predicted to face the lipid are colored in yellow.

Significant constraints for modeling the interaction of H3 with the extracellular porions of H5-H6 modeled above arise from the maximum length of the loop between the N-terminus of H5 and the Cys-Cys crosslinked normally present in these GPCRs, described already in Section 7.2.4 above. The relation of the H5-H6 packing to the H3 segment at the extracellular side is guided by the requirements of ligand binding, and will therefore be described in a subsequent section as part of modeling the ligand binding sites on the crevice defined by H3-H5-H6. Therefore, the incorporation of H4 into the H3-H5-H6 template is presented first to complete the binding site pocket formed by H3-H4-H5-H6. Subsequently, incorporation of the prototype ligands 5-HT and LSD allows refinement of the H3-H4-H5-H6 interactions within the model. Modeling the interaction of H3-H5-H6 at the extracellular side in the absence of the ligand would have resulted in a model of the unliganded receptor, which is not the aim of this thesis. Nonetheless, inferences would be made in the modeling process as to the nature of the conformational changes that would result in the conversion between the ligand-bound and the unliganded forms of the receptor.

7.4.4- Packing H4 on to the H3-H5-H6 template.

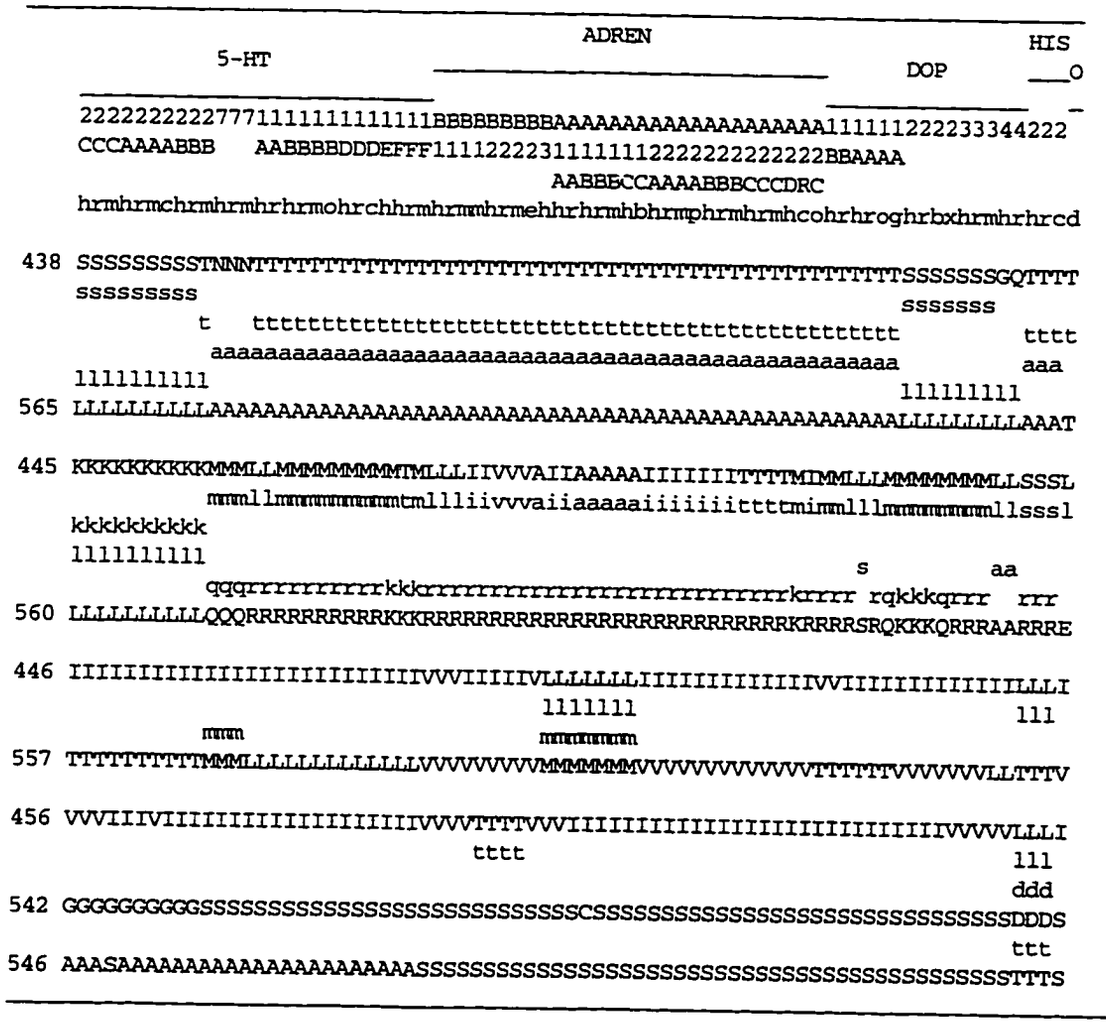
The most salient feature of H4 guiding the packing into the transmembrane bundle is its irregular surface pattern, shown in Figure 7.18. Note the presence of a very narrow stretch of residues predicted to face the interior of the TMH bundle from the cytoplasmic end up to the 100 % conserved W4.50: these are residues S4.38-A4.42-I4.46, followed by a sharp turn and a widening of the patch of interior facing residues towards the extracellular boundaries, defined by S4.53-I4.54-V4.56-S4.57. The bend at W4.50 in the direction of the interior facing residues becomes a guiding principle for packing H4 into the TMH domain. I described in Section 7.3.3.2 the role of the conserved β-branched I4.46 in constraining the conformation of W4.50 into the X1=t rotamer, whose final conformation would be defined by X2=+90 or X2=-90. Modeling W4.50 in the X1=t rotamer orients its indole ring perpendicular to the H4 axis and thus parallel to the membrane plane. This results in a highly irregular Van deer Waals surface pattern for the interior facing residues of H4 where the aromatic moiety of W4.50 protrudes from the H4 surface, as shown in Figure 7.18; the requirement to optimize packing density within the TMH domain would make this Van der Waals pattern the guiding principle in modeling helix-helix interactions for H4. The C-terminus of H4 contains a PXP motif that terminates the helix (Section 5.4), which is a highly conserved motif adopting different forms (PxP, PP, P, see Section 5.4). I have modeled this motif by taking the backbone conformation

observed for a similar motif present in dehalogenase (code 1EDE in the PDB), and have included only the first Pro4.59 in subsequent modeling.

The resulting H4 was then added to the template form by H3-H5-H6 at the cytoplasmic side. There aren't any experimentally derived interhelical interactions proposed for H4. This helix is significantly shorter than any other helix (22 residues), and thus may lack interhelical interactions at either end of the membrane. The high concentration of non-conserved Arg and Lys residues at the cytoplasmic side of H4 (see Figure 5.3), predicted to face the phospholipid headgroups, would force the cytoplasmic portion of H4 towards the cytoplasmic boundary of the membrane, at the level where other TMHs have Arg/Lys residues proposed to participate in analogous interactions. This criterion is exemplified in the 5HT2C receptor where a Lys at position 4.45 facing the lipids, just 5 residues before the conserved W4.50, would need to reach the phospholipid headgroups. I have inferred a more precise axial displacement of H4 relative to H3-H5-H6 from the presence of self-consistent evolutionary revertant mutants (**erms**) between H4 and H5 at the cytoplasmic side, shown below in Table 7.6.

Table 7.6

Erms between H4 and H5 used to incorporate H4 to the H3-H5-H6 template.



Orienting H4 into the H3-H5-H6 template based on these **erms** positions the 100% conserved W450 facing a groove in the H3-H5 interface formed by residues P5.50, I3.40, L3.43 and C3.44, also 100% conserved. Packing of W4.50 into the TMH bundle is a hard task because the conserved beta-branched residues at position 4.46 constrains W4.50 to a X1=t, where the

indole side chain is parallel to the membrane plane and thus requires a significant opening at the H3-H5 interface, which I have modeled very closely packed at the cytoplasmic boundaries as described above. This necessary and rather abrupt widening of the H3-H5 interface can be accomplished by the kink induced in H5 by P5.50, by tilting H3 towards H2, and by the role of S3.39 and C3.44 in stabilizing a curved H3, as can be seen in Figure 7.21. Because W4.50 is oriented towards Pro5.50, the kink induced in H5 by this Pro residue necessarily bends H5 away from the W4.50 (see Section 7.3). Significant tilting of H3 towards H2 is actually one of the most salient features of the rhodopsin structure (Schertler 1995). Thus according to this model, the role of W4.50 could be, in part, to induce the observed large tilt of H3. I propose a novel role for the multiple Ser, Cys, and Thr residues present in H3, in stabilizing bends or curvatures on α-helices, as described below.

Ser and Cys residues in α-helices are special in that they can H-bond to the carbonyl i-3 or i-4 of the preceding turn, as observed in known protein structures (see Section 2.1 and references (Gray 1984; McGregor 1987)). This H-bonding can open (or stabilize an opening of) the helical turn, due to bifurcated H-bonding at the i-3 or i-4 carbonyl, much in the same way that water molecules open helical turns when they H-bond to the backbone carbonyl (Blundell 1983). H3 shows an unusual concentration of Ser/Cys/Thr residues, as shown in Table 6.1 where Ser/Cys/Thr present in over 50% of the receptors at any given position are listed; note that H3 contains 9 such

positions while the other TMHs contain only 2-3 cases. The equivalent residues in the 5-HT2CR are S3.36, T3.37, S3.39, C3.44, and S3.47. Of these residues, T3.37, S3.39, and C3.44 are 100% conserved. In MonteCarlo simulations on the isolated H3 of the GnRH receptor (see Section 7.3.2) I have observed this interaction for the Ser/Cys/Thr residues present in over 90% of the population, meaning that such conformation is energetically preferred. The net effect was to induce a continuous but not dramatic bend on H3, i.e. a curved helix. Energy minimization of an isolated H3 of the 5HT2C confirmed this effect. The bend induced or stabilized by a Ser/Cys residue is significantly larger when they adopt the X1=g- due a closer distance between the gamma OH/SH and the carbonyl oxygen than in the X1=g+ conformation. This is consistent with the observation that side chains other than Cys and Ser rarely occur at this rotamer due to steric clashes with the C=O i-3, which Ser and Cys can H-bond (see Section 2.1). This proposed structural role for Ser/Cys was confirmed by searching the PDB database for α-helices containing a Cys or Ser residue whose X1=g-; 125 structures were found supporting the feasibility of this conformation. The extent of the bending induced varied widely from a small curvature to a sharp bend, an example of each case is illustrated in Figure 7.19. Interestingly, the α-helical segment shown in panel B was extracted from the cystal structure of the Photosynthetic Reaction Center (PRC), where it spans the membrane bilayer. Note that three out of the five {Ser, Thr, Cys} residues present in TMH5 of the PRC adopt the X1=g-

conformation, where they H-bond the backbone carbonyl i-3 bending significantly this TMH. Thus, analysis of the PRC structure indicates that the role of multiple Ser and Cys residues in a TMH (such as H3 of GPCRs) could be to bend significantly the α-helix by adopting the X1=g- rotamer.

A similar structural role could be expected for the multiple Ser and Cys residues in H3 of GPCRs. Because the X1=g- conformation can be attained by Ser and Cys but not Thr (Gray 1984; McGregor 1987), an AA site where Ser and Cys are conserved but a Thr is never allowed would become a candidate for bending the TMH adopting the X1=g- rotamer. The conserved residues S3.39 and C3.44 fit this conservation pattern and thus their ability to bend H3 by adopting the X1=g- conformation was explored by energy minimization. The results are shown in Figure 7.20. Note from panel A that the resulting H3 backbone conformation can be bent either towards H2 or towards H4, inducing large displacements for D3.32 (the binding site for protonated amines) when the cytoplasmic portion (residues R3.50-D3.49) is maintained constant. The direction of the bend was consistent throughout the analyzed structures, as illustrated in Figure 7.20.b from a side view. This directionality was also observed in known structures, as shown in Figure 7.19.a. Thus the presence of S3.39 and C3.44, as well as other Ser/Cys/Thr residues in H3 of GPCRs could allow for significant structural deviations among them at the level of the binding site.

Modeling S3.39 as adopting X1=g- shows that it can H-bond both to the

C=O i-3 (3.36) and also to the NH of W4.50. Selecting S3.39 X1=g- H-bonding to C=O 3.36 followed by energy minimization induces an opening of the S3.39/I3.40-L3.43 groove. This widening facilitates docking W4.50 into the turn formed by the 100% conserved residues S3.39/I3.40-L3.43. However, this opening actually bends the extracellular portion of H3 far from H5 and the NH+ of LSD/5HT. The helix bending effect of S3.39 adopting the X1=g- conformation can be reverted by C3.44 adopting the X1=g- as well : Because S3.39 and C3.44 are on opposite sites of H3, both with an X1=g- they maintain the overall orientation of H3 yet allowing an enlargement of the H3 groove between S3.39-I3.40 and L3.43. This facilitates the docking of W4.50 into this groove, shown in Figure 7.21. Thr 3.37 is also facing the membrane and is also 100% conserved; if required, it could play a similar role in bending or stabilizing a bend or a curvature in H3. This hypothesis provides a structural role for C3.44 and T3.37 and thus explains their high conservation pattern despite their facing the lipids in the model. It also explains the lack of helical character in the conservation pattern for the entire H3 turn 3.37-3.38-3.39-3.40, which is nearly 100% conserved (see Section 5.7 and Figure 6.7).

The role of S3.39 and/or C3.44 in bending or stabilizing a bend on H3 when they adopt X1=g- may play a role in the different conformational states adopted by the receptor, especially between the unliganded versus the liganded receptor forms. In modeling the unliganded receptor, H5 at the extracellular side needs to get close to the extracellular side of H3, as suggested

above from the Zn binding site 3.29-5.39 (Elling 1996) and the erm N6.55-S5.43. However the PK of H5, modeled in a low kink angle, can not bend further towards H3 because a Pro induces precisely the opposite bend, i.e., a bend towards the opposite face of the Pro so as to open up the helical turn and relieve the steric clash between the Pro ring and the C=Oi-4 (see Section 7.2.3). This analysis indicates a possible mechanisms for opening and closing the binding site of the receptor (for the conversion between a liganded and an unliganded receptor) that could be accomplished not only through modulating the PK conformations on H5 and H6, but also by inducing or stabilizing a bend in H3 by T3.37, S3.39, and C3.44. Thus, S3.39 adopting the X1=g- would induce an opening while C3.44 X1=g- would induce an occlusion of the binding site crevice. In conclusion, for the liganded form of the receptor I have chosen C3.44 X1=g- and S3.39 X1=g-.

The hypothesis that W4.50 is H-bonded to S3.39 is contrary to the experimental observation in rhodopsin that W4.50 is not H-bonded (Lin 1996). However, this discrepancy can be explained by analysis of the equivalent residues to S3.39 in rhodopsin, shown below, which indicates that rhodopsin lacks an H-bonding group at this locus.

```
                  3    3 3       3         3                4         4         4
                  2    3 3       4         5                4         5         5
                  5    0 2       0         0                0         0         9
OPSD_BOVIN        GCNLEGFFATLGGEIALWSLVVLAIERYVVVCKPMSN-FRFGENHAIMGVAFTWVMALACAAP
                   * .    .  * .  ..  *  ....*** . .*. .    .   *** .*. * .... . *
RAT 5HT2C         LCPVWISLDVLFSTASIMHLCAISLDRYVAIRNPIEHSRFNSRTKAIMKIAIVWAISIGVSVP
                   __________Hx3_____________        _______Hx4_________
```

Residue S3.39 in rhodopsin is an Ala, and 3.40 is a Leu, so there is no equivalent to S3.39 in rhodopsin which supports the finding that W4.50 is not H-bonded in rhodopsin but it may well be in 5HT2CR.

When W4.50 is fitted H-bonding its NH to S339 X1=g-, it is surrounded by the ring of P5.50 and the side chains of L3.43 and I3.40; thus the 100% conserved W4.50 is surrounded and closely packed by residues which are also 100% conserved. Optimization of adjacent packing interactions positions the highly conserved S4.53 to H-bond the C=O 5.46. Because this C=O is in position i-4 from P5.50, this H-bonding could modulate the conformation of the PK in H5 and explains the high conservation of a Ser residue at position 4.53. Ser is never substituted by a Thr, because the proposed H-bond requires S4.53 to adopt the X1=t rotamer, which is energetically unfavorable for a Thr residues. Other, less conserved residues can be packed so as to alleviate existing steric clashes until the resulting packing of H4 into the H3-H5-H6 template, show in Figure 7.21, is considered energetically favorable.

Incorporating H4 into the receptor model as described above implies that the extracellular end of H4, defined by the PXP motif centered at 4.59, is positioned at the level of residues D3.32-V3.33 in H3 and S5.46 in H5. Because these residues are proposed ligand interaction sites, it follows that the "loop" from P4.59 onwards may contain crucial ligand-receptor interaction sites, as shown in the identification of a covalent attachment site for a nitrene-activated clozapine derivative, an agonist, to a 11-residue fragment of

the alpha-adrenergic receptor in this region (Matsui 1989). A 100% conserved beta-branched residue at position 3.33, one residue after the D3.32 and thus oriented towards H4 at the level of the PxP motif, is positioned so as to block the access of this loop towards the D3.32, a role combined with the rigidity imposed by the PP or PXP motif right at this locus. This structural role would explain the high conservation observed for both V3.33 and a Pro at position 4.59 or 4.60.

A cascade of interrelated and specific interactions lead to the formation of a cavity between H3-H4-H5-H6, referred to in the literature as the "ligand binding crevice" (Javitch 1995a). These sets of interactions are initiated by the steric requirements of docking W4.50. The NH of W4.50 can H-bond S3.39, which adopts the X1=g- conformation where it can also H-bond the backbone C=O of S3.36. Modeling S3.39 (and C3.44) adopting the X1=g- has two effects on H3. Locally, an opening of the groove between S3.39-I3.40-L3.43 is induced, which facilitates docking of W4.50. This local distortion bends H3 towards H2 (see Figure 7.20). This H3 orientation is synergistically induced by the same requirement to accomodate the bulky W4.50, which can be fullfilled by tilting H3 towards H2 as shown here. The synergistic reorientation of H3 caused by bending and tilting results in an opening at the binding site between H3 and H5, as shown in Figure 7.21. The extent of this opening is limited by the disulfide bridge between C3.25 and C5.31 conecting the extracellular ends of H3 and H5 (blue strip in Figure 7.21). The effect of this opening is to create a

cavity lined by H3 (right side), H5 (left side), the bulky W4.50 (bottom), and the H3-H5 loop segment at the top created by the disulfide bridge C3.25-C5.31. This cavity encompases several proposed binding residues (D3.32, S5.46, not shown), and has therefore a functional correlate: the "binding site crevice", shown shaded in Figure 7.21.

An alternative solution to prevent the resulting cavity of the binding site is to model the extracellular portions of H3 and H5 so as to close the opening described above. Previous modeling of the allowed conformations of these H3-H5 regions suggested that H3 could be bent towards H5 utilizing the proposed role of C3.44 and T3.37 described above. The ability of the extracellular segment of H5 to bend towards H3 was discussed in terms of the available PK conformations of H5 in Section 7.3.3; it was concluded that the bend and twist here from the H5 PK was of an opposite direction than the normal effect of a PK. The presence of P5.50 in this context, accentuates the observed opening of the binding site crevice at the H3-H5 interface. The inability of the H3 and H5 extracellular segments to prevent the formation of a cavity at the binding site described above suggests that the resulting cavity is an intrinsic property of the model constructed.

7.4.5.- Incorporating the ligands in the H3-H5-H6 template: criteria for helix packing from interactions with docked ligands.

The ligands 5-hydroxy-tryptamine (5-HT) and Lysergic acid (LSD) were

selected for modeling purposes due to the presence of specific ligand-receptor interactions proposed in the literature for these compounds. We proposed an interaction between the indole nitrogen of both LSD and 5-HT with a Ser residue substituted at position 5.46 in the 5HT2CR, normally present in the human 5HT2AR (Almaula 1996a). These studies provided a molecular basis for the observed selectivity of some ergolines and alkylindolines between these two receptors, and is presented in Section 8. The pharmacological profile of the 5HT2CR is very similar to that of 5HT2AR, and previous structure-activity relationships (SAR) did not distinguish well among them (Glennon 1991) thus referring to the 5HT2 subclass as a group. This means that proposed interaction sites for the 5HT2A subtype can be assumed for the 5HT2C subtype modeled here unless otherwise stated. This is useful in assuming three additional ligand-receptor interacting sites: The protonated amine of neurotransmitter receptors has been proposed to bind to the Asp in position 3.32, conserved among this subclass of GPCRs. This interaction has been supported experimentally for the 5HT2AR (van Rhee 1996), and is thus assumed in this studies. Another researcher in this laboratory, Dr. Daqun Zhang, proposed an interaction between the protonated amine of 5-HT and S3.36 on the 5HT2AR, an hypothesis substantiated by substitution of S3.36 with Cys and Ala (Almaula 1996b). Because S3.36 is conserved among the 2A and the 2C receptors, I have assumed a similar interaction. The last proposed interaction pertinent to the 5HT2CR is a series of extensive mutagenesis

studies on the 5HT2AR which suggested a direct interaction between 5-HT and ergolines with of F6.52, although the hypothesis of a direct interaction in this case is significantly weaker than with D3.32, S3.36, or A/S5.46.

The requirements for ligand binding derived from specific interactions proposed above with D3.32-S3.36-A/S5.46-F6.52 can be expanded by considering the proposed physico-chemical properties of the receptor binding site or "pharmacophore" for the 5HT2 subclass, derived from extensive SAR analysis (Glennon 1991). The proposed "pharmacophore" for the 5HT2CR is shown in Figure 7.22 together with the numbering of the indole and ergoline moieties that are used below to rationalize the observed pharmacological effect of 5-HT and LSD analogs. Adopting the chemical conventions to number the atoms of the indole moiety would have resulted in two different numbering schemes for the same moiety in both ligands. Therefore, in order to facilitate the discussion of their respective interactions I have adopted a common numbering scheme for the indole moiety of 5HT and LSD, corresponding to the serotonin numbering. This figure is complemented by a schematic representation of the direct ligand-receptor interactions proposed in this section (panel C), which identifies the specific amino acids in the 5HT2CR responsible for the observed pharmacological profile.

The sites where the proposed pharmacophore is significantly different from the pharmacophore for similar receptors suggests that amino acids are present at the corresponding locus in the 5HT2 subclass that are significantly

different from the amino acids present at the same locus in other receptors. It its simplest form, this means that we can draw a parallel between the differences among pharmacophores for different receptor subtypes and **erms** among the same subtypes. Thus, a correlation between these **erms** and the differences in the pharmacophores can be interpreted to propose specific residues responsible for the observed pharmacological profile. One example is the observed preference within the 5HT2 receptors for small hydrophobic substituents at positions 6 and 7 of the indole ring (see Figure 7.22 for the numbering used): an OH at position 6 significantly decreases the affinity, whereas a methyl but not a large substituent at position 7 is well tolerated. This suggests that positions 6 and 7 of 5-HT are facing a region of the binding site of 5HT2Rs that is characterized by higher hydrophobicity and is sterically limited. Residues F5.44 and L5.40 could be responsible for this role, as suggested by the **erm** within the 5HT subtypes shown below:

```
                                      ADREN                          HIS
              5-HT                                            DOP    ___O

    22222222227771111111111111BBBBBBBBBAAAAAAAAAAAAAAAAAAA111111222233344222
    CCCAAAABBB   AABBBBBDDDEFFF1111222231111111222222222222222BBAAAAA
                                      AABBBCCAAAABBBCCCDRC
    hrmhrmchrmhrmhrhrmohrchhrmhrmmhrmehhrhrmhbhrmphrmhrmhcohrhroghrbxhrmhrhrcd
544 FFFFFFFLLLAAAFFVVVVCCCLFFFVVVIIIIISVVLLLLLIIIIIIIIIIILILLVLLLIIIIVVVVILLLL
    fffffff       ff           fff
    111111111                             1 111111     111111111             111
540 LLLLLLLLLVIIIIIVVVVIIIIIIIIIIIIIIILVILLLLLIIIILLLLLLLLLLIIIIIIVVVIIIIVVLLLI
```

Modeling the indole moiety of 5-HT and LSD with the indole NH H-bonding a substituted Ser in position 5.46 is consistent with an orientation of positions

6 and 7 of the indole ring towards F5.44 and L5.40, as shown in Figure 7.23. Note that F5.44 adopting the X1=g+ rotamer can interact with positions 6-7 of the aromatic ring of 5-HT and LSD in a T-shape packing, and L5.40 is also facing these positions of the docked ligand. The refined docking of 5-HT and LSD with H5 using this criterion served to position 5-HT and LSD into the H3-H4-H5-H6 template already modeled.

Packing the cytoplasmic side of H3-H4-H5-H6 as described above positions the ligand binding residues of H3-H5-H6 in a configuration suitable to satisfy the required specific ligand-receptor interactions for the 5HT2R described above. The resulting ligand-receptor interactions at the binding site defined by H3-H5-H6 is shown in Figure 7.23 for 5-HT and LSD. The protonated amine of all three ligands is involved in an ionic bond with the D3.32. For 5-HT but not LSD, the protonated amine also participates in H-bonding to S3.36, which acts as an H-bond acceptor because its OH is H-bonded to the backbone carbonyl of the fourth preceding residue, as found commonly in protein structures (Gray 1984). The indole-NH of 5-HT and LSD is shown as H-bonded to S5.46 (in the 5HT2C receptor this locus is an Ala residue but the interaction is shown here because it is part of the proposed ligand interaction site as discussed in our publication (Almaula 1996a)). This is consistent with the proposal of a similar orientation for both 5-HT and LSD at this locus between the 5HT2C and 5HT2A receptors (see Section 8.2 below). The LSD geometry is as found in the crystal structure (Baker 1972), and in the 5-HT the

ethylamine side chain adopts a fully extended conformation with ideal dihedral values X1=90 and X2=180. The 6 and 7 position of the indole ring of 5-HT and LSD are interacting with F6.52, as suggested by mutagenesis experiments on the 5HT2A receptor (Choudhary 1992; Choudhary 1995). Finally, the indole ring of 5-HT and LSD is interacting with the cluster of aromatic residues in H6, composed of F6.52 and W6.48. W6.48 is interacting in a T-form with the aromatic core of all three ligands, while F6.52 participates in lateral contacts.

Thus the proposed model at the level of the H3-H4-H5-H6 binding site can satisfy all direct ligand-receptor interactions required, as defined in Section 2.2.

Note that in this proposed arrangement the cluster of aromatic residues from H5 and H6 (W6.48,F6.52, F5.47) are interacting with the indole ring but not very extensively, as judged by the small percentage of the indole surface area in direct contact with these aromatic residues. There are alternative aromatic conformations that would allow for a much more extensive Van der Waals contact between residues belonging to the aromatic cluster motif and the indole moiety of 5-HT. The low percentage of ligand-aromatic surface contact is a consequence of the T-shape mode of aromatic-aromatic interactions modeled, which now include the indole ring. The T-mode packing of aromatic-aromatic side chains was originally induced by W6.48 adopting the g+ configuration for the inactive state of the receptor,

which orients the Trp side chain perpendicular to the membrane plane as suggested by fluorescence experiments ((Lin 1996) and see Section 7.3.3). These same experiments and the ensuing structural interpretation indicate that receptor activation involves a significant rearrangement of the aromatic residues where W6.48 and F6.52 adopt the X1=t configuration. Because these two aromatic residues are modeled in direct contact with the ligand, the model suggests that receptor activation entails a significant rearrangement of the aromatic ligand-receptor interactions.

Other significant ligand-receptor interactions take place in the model. One is the H-bond between the OH of 5-HT and the N6.55 side chain that acts as a donor, shown in Figure 7.23. This is consistent with previous SAR analysis suggesting the importance of an H-bond acceptor site at position 5 of the indole ring (see Figure 7.22 and (Glennon 1991)). It is noteworthy that the hypothesis of H-bonding to N6.55 has been implicated before in ligand binding for other GPCRs (Wieland 1996). In particular, mutagenesis experiments in the β2-adrenergic receptor have suggested that N6.55 acts as an H-bonding donor to the ethylamine OH of epinephrine (EPI). An interaction with the site on the equivalent position of the ergoline core of LSD has also been shown to be important for ligand binding to 5HT2 receptors (see Figure 7.22 and (Glennon 1991)). LSD lacks an alkoxy site, but the double bond between C12-C13 has been proposed to act as an H-bond acceptor based on results from quantum mechanical calculations (Weinstein 1978).

This hypothesis suggests a superposition of the LSD and 5-HT molecules such that the minimum of the electrostatic potential (MEP) for the double bond in LSD, 1.5 A over the midpoint of the double bond, is superimposed on the MEP of the 5-OH of serotonin. The position of the MEP of the serotonin 5-OH depends on the directionality of the O-H bond. However, no specific interactions for the hydroxyl hydrogen atom are found, consistent with an acceptor site at that position suggested by SAR. Therefore this criterion would be fulfilled by different LSD to 5HT superpositions. The spatial relationship between 5-HT and LSD bound to the 5-HT2CR can be further defined by the interaction of these two molecules with the 5.46 locus in the receptor. We and others laboratories have proposed an H-bond between the indole NH of 5HT and LSD and a Ser residue at position 5.46 in the human 5-HT2AR and the mutant A5.46->S 5HT2CR, based on mutagenesis and pharmacological experiments (Almaula 1996a). Thus the indole NH group of LSD and 5-HT occupy similar positions in the bound receptor. This necessarily implies that the indole moieties of both ligands are not superimposable, a conclusion consistent with the observed divergent pattern of tolerance between 5HT and LSD at position 2: LSD can tolerate a Bromo or Iodo substitution at this position, while 2-methyl 5-HT significantly reduces the affinity for the 5-HT2C receptor. (Note that the 2-methyl-5-HT binding data was derived from the rat receptor, not the human receptor modeled here).

Another ligand interaction site not proposed before involves position 2 of the indole ring of 5-HT, which is facing 4.54 in the model (4 A, see Figure 7.24), a position where a β-branched residue (I,V) is conserved among most 5HT receptors except 5HT7 (Ala) and 5HT1A (Phe), as shown below:

```
                5-HT
              _____________________________
    222222222277711111111111111
    CCCAAAABBB   AABBBBDDDEFFF
    hrmhrmchrmhrmhrhrmohrchhrm
454 IIIVVVVIIIAAAFFIIIVIIIIVVV
              aaa
    iiivvvviii     iiiviiiivvv
                  ff
```

It has been shown that 2-methyl substitutions on 5-HT lower the affinity significantly for all 5HT1 and 5HT2 subtypes. This is consistent with the proposed proximity of position 2 of 5-HT to position 4.54 in the receptor, because β-branched residues are conformationally restrained in α-helices and thus best suited to induce steric constraints on nearby residues or as in this case, with ligand moieties. The bulky Phe residue present in 5HT1ARs could not also fulfill this role, but actually explain the high affinity of 5HT1A ligands which lack the indoline 5-membered ring that occupies a region in space in close proximity to F4.54, such as 8-OH-DPAT. The 5HT7R subtype, however, has an Ala at position 4.54. These considerations lead to the conclusion that substitutions at position 2 of serotonin could be selective for 5HT7 receptors versus 5HT2 and 5HT1 receptors, a prediction already confirmed in the literature (Carter 1995). The proposed role of I4.54 as

responsible for the steric hindrance upon substitutions at position 2 of the indole ring of 5-HT, can now be used to rationalize the opposite pattern of tolerance observed for Bromo and Iodo substitutions at position 2 of the indole ring of LSD (Ismaiel 1990; Glennon 1991; Burris 1992) described above. This rationalization defines the relative orientations of the indole rings of 5-HT and LSD which, combined with the criteria described above on the spatial relationship between 5-HT and LSD within the binding site modeled, determine the relative orientation between 5-HT and LSD, as shown in Figure 7.24.

A general structural consideration from the modeling of ligand binding is that introducing the ligand into the receptor model requires a significant distance between the N-termini of H3 and H5. This appears be inconsistent with an engineered Zn-binding site between 3.29 and 5.35 (5.39 to a lesser extent) (Elling 1996). Although 5.35 is already in the H4-H5 loop in the present model of the receptor, the distances are still too large to explain the Zn binding site, which normally implies that the respective C^{α} atoms are separated by less than 13 A. It could be that the receptor in the absence of the ligand allows H3 and H5 to become closer to each other and to satisfy the Zn binding constraint. This would explain the effect of this engineered Zn binding acting as a competitive antagonist inhibiting substance P binding (Elling 1996), by stabilizing or inducing a close proximity between H3-H5 that would close the binding site for other, larger, ligands.

Multiple possible packing interactions as described here are feasible for several regions of the receptor model. Selecting a particular choice in the modeling process, even in the presence of compelling inferences, does not exclude a role for these other, alternative packing arrangements. Instead, it may indicate either receptor flexibility and/or a role for these alternative packing arrangements in the different conformational states that the receptor may adopt. Interestingly, in addition to the flexibility provided by the Pro-kinks in H5 and H6, I have identified above S3.39 and C3.44 in H3, and the loop following P4.59 in H4, as potential sources of structural heterogeneity. These "conformational hinges" in H3, H4, H5, and H6 define the spatial relationship between the conserved inner core of the TMH domain, defined above by the aromatic cluster (see Figure 7.17), and the more divergent "binding site crevice" towards the extracellular side (see Figure 7.21). Note the correlation between the partition of the receptor based on the "conformational hinges" identified in this work, and similar partitions of the TMH domain based on conservation and/or functional criteria. This correlation suggests that the sources for structural heterogeneity identified in this work, i.e. the proposed "conformational hinges", could play a role in the interconversions among the "conformational states" defined in Section 3.1, i.e. the different inactivated and activated forms of the receptor. An example is the proposed conformational change of the PK in H6 upon activation

(Section 7.2.4). Alternatively, these same "conformational hinges" may have been utilized through evolution as a source of functional specificity, as proposed in Section 7.2.5.

7.4.6- Incorporation of the extracellular segment of H7, comprising residues 7.35-7.46, into the H3-H4-H5-H6 template.

The ligand interaction sites proposed in the previous section can be completed by the incorporation of the extracellular segment of H7, comprising residues 7.35-7.46. The residue at position 7.39 has been implicated in antagonist binding for several neurotransmitter GPCRs; T7.39 in the human 5HT1B receptor is a N7.39 in the rodent 5HT1B receptor and the substitution T7.39N can explain the species-differences in their pharmacological profile. An Asn at position 7.39 is also present in the 5HT1AR where it has been proposed to interact directly with pindolol and analogs. A F/N7.39 difference in the α2-β2-adrenergic subtypes, respectively, has also been shown to modulate the affinities of pindolol-like ligands between these two subtypes. The conserved W7.40 residue, one residue after the 7.39 site discussed above, has been identified as the site of covalent attachment of a nitrene-derivated analog of cyanopindolol to the α-adrenergic receptor (Wong 1988). It follows that positions 7.39 and 7.40 are facing the binding site of GPCRs. This is consistent with the projection map of rhodopsin, which shows H7 positioned adjacent to H3 and H6 where other

ligand binding site residues are located.

From the set of ligand-receptor interactions described above, the most useful and pertinent for modeling H7 contacts in the 5HT2C receptor model is the T7.39N mutation performed on the 5HT1BR (Guan 1992; Metcalf 1992; Oksenberg 1992; Adham 1994; Glennon 1996) , because several ergolines tested (Methylsergide and Metergoline) had significant changes in affinity upon this mutation. Assuming a similar binding mode among ergolines at serotonin receptors, as supported by the apparent high affinity of several ergolines like LSD for many different serotonin receptor subtypes, the already modeled LSD-receptor interactions can be used to orient the incorporation of H7: thus, V7.39 in the 5HT2CR model, isoesteric with the T7.39 present in the 5HT1B receptor, is oriented towards the divergent moiety of these ergolines (see Figure 7.22). This orientation is consistent with W7.40 also facing the binding site, as shown in Figure 7.25. I should note that the constraint derived from rhodopsin by the placing of retinal between the Schiff base on K7.43(296) and E3.28(113), proposed in the literature, has been seriously challenged by NMR results suggesting that the interaction between these two groups is mediated by water and C12 within the retinal chain. If the latter is correct, then this criterion becomes sterically too fuzzy to serve as a restrictive criterion for modeling helix-helix packing.

Other constraints for the positioning of H7 in the H3-H4-H5-H6 template can be satisfied. The side chain of F6.51 is modeled in the X1=g+

rotamer in which it is packed against the conserved G7.42, thus explaining the requirement of a Gly residue at position 7.42. This same Y6.51, if it were to adopt the X1=t rotamer, would be capable of fulfilling an **erm** found with 7.35 shown below:

```
735 LLLLLLLLLLEEEGGFFFFFFFASSSFFFFYYYYFFFFFFFFFFFFFFFFFFFFFFFFFFFFFYYYYYYYYYVVEEEK
              eee                                                                eee
                                                                                 yyy
651 FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFYYYF
```

This **erm** could be explained by a Tyr in position 7.39 (F7.39->Y) that would H-bond the correlated Glu in 7.39 (V7.39->E).

An **erm** discussed in Section 7.2.5 was proposed to modulate the PK of H6 by specific H-bonding interactions, involving residues C7.45-M6.47, which are also seen in close contact in the model shown in Figure 7.25. This positions N7.36 in a conformation suitable for H-bonding to W3.28, an interaction supported by the **erm** character between these two residues shown below:

```
328 WWWWWWWWWWFFFFFWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWWYYYYYYYYYYYYYYWWWWWWFFFFFFFLLYYYW
    wwwwwwwwww      wwwwwwwwwwwwwwwwwwwwwwwwwwwww                 wwwwww            w
    nnnnnnneqe      ddddddddnnn                  nn qq          dddddd            n
736 NNNNNNNEQERRRAADDDDDDDDNNNVVVVIIIILKKKKKKKKNNKQQQKKKKKKDDDDDDSSSSSRRSSAGAN
```

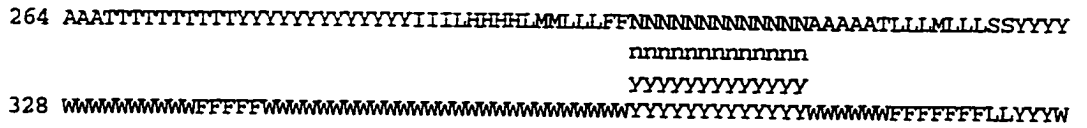
This N7.36-W3.28 positions W3.28 above of D3.32 where it could stabilize the ligand-receptor interactions, and favors aT-shape packing of W3.28 with W7.40, which may explain the high conservation of aromatic character found at these residues.

The resulting packing of the H7 extracellular segment into the H3-H4-H5-H6 template is shown in Figure 7.25 where the interactions described above are marked by dashed lines. Note that the resulting packing is consistent with the accessibility pattern predicted in Section 6 (predicted lipid-facing residues are shown in yellow) and experimentally validated by SCAM analysis of the D2 receptor (Fu 1996).

7.4.7- Incorporation of H2 into the H3-H4-H5-H6-H7ec template.

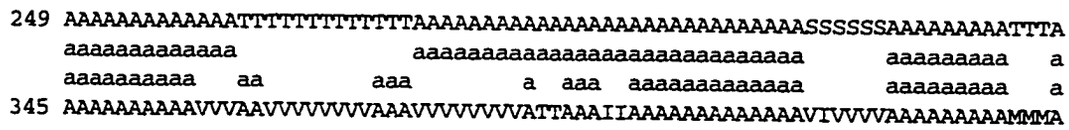
The rhodopsin map indicates that the interaction of H2 with the H3-H4-H5-H6 template is restricted to H2-H3 contacts, H4-H5-H6 being farther away. This is not a handicap for modeling H2 into this template because a variety of criteria exist that defines sufficiently the interaction between H2-H3 so as to guide incorporation of H2 into the preexisting template.

There is a Zn binding site engineered between 3.28-2.64 in the NK1 GPCR (Elling 1996) that determine the relative axial displacement between H2-H3. However, the Zn data on the NK1 receptor is not directly transferable to the 5HT2C receptor, because the NK1 receptor differs from the neurotransmitter receptors here considered in that it lacks the Pro at 2.59 and it has an extra Pro at 3.32, which would alter the relative position of 2.64-3.28 with respect to the cytoplasmic portions of H2-H3 among these receptors. Nonetheless, an **erm** exist between these two sites shown below:



This **erm** could be interpreted as suggesting an H-bond between N2.64 and Y3.28 whose geometrical requirements parallels the Zn binding site H2.64-H3.28 in NK1 receptors (Elling 1996) and was thus used to position H2 relative to H3. The inferences from the Zn binding site and this **erm** are consistent with the proximity between 2.64 and D3.32 inferred from the identification of 2.63 and/or 2.64 as the attachment sites in the β2-adrenergic receptor for a covalently bound ligand (Dohlman 1988), a derivative of cyano-pindolol which would be predicted to interact simultaneously with D3.32.

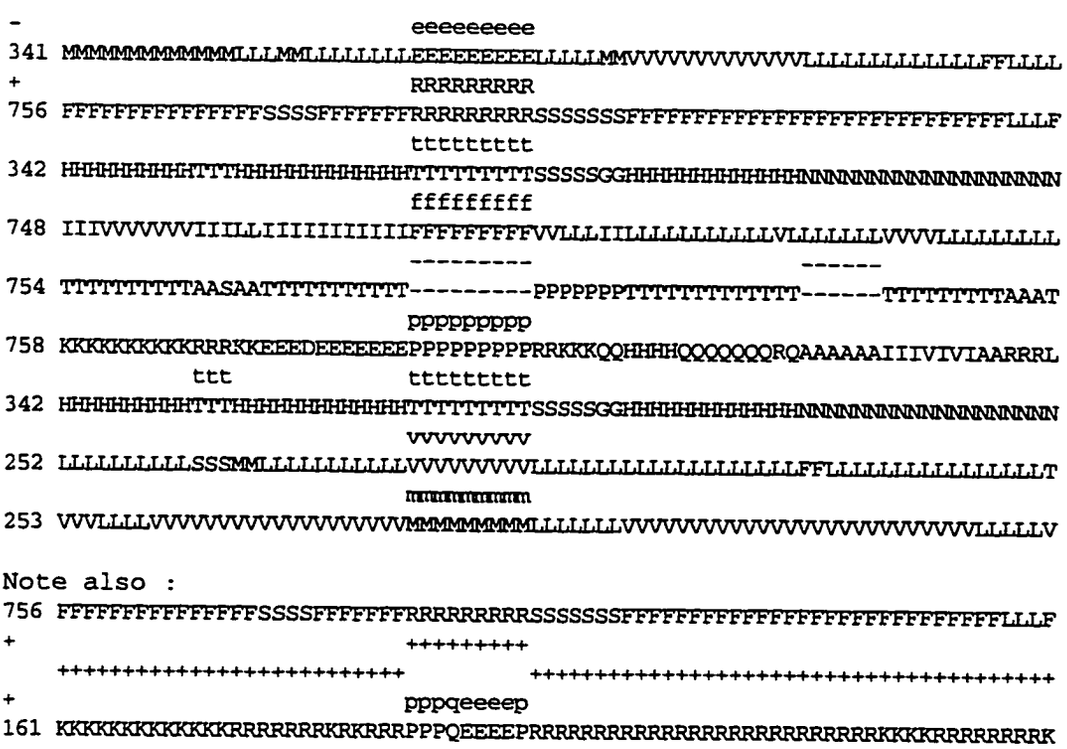
The resulting H2-H3 interface is consistent with a large set of inferences. The conserved D2.50 is a second major criterion in packing H2, and a conserved His/Asn residue at 3.42 is capable of stabilizing D2.50 through an H-bond. The conservation pattern at 3.42 (Asn/His) suggest that an H-bond donor at position δ along the side chain is the conserved property at this locus. Thus a H-bond between H3.42-$^{\delta}$NH...OCO-D2.50 was used to orient the H2-H3 interface. Modeling the D2.50-H3.42 H-bond as mentioned above positions the highly conserved V2.57 in contact with the highly conserved A3.38. The H2-H3 interface becomes tighter at the level of A249-A345, both conserved in their small volume and correlated through evolution as shown below:



There is a large set of **erm** that arises from the presence of a Glu residue at position 3.41 within the β-adrenergic receptor subtype that relates to the H3-H2-H7 interface, shown in Table 7.7 below.

Table 7.7

Erm patterns related to the presence of a Glu residue at position 3.41 in the β2-adrenergic receptors.



The conflicting presence of this Glu residue in a position facing the lipid membrane was mentioned by Lybrand in his modeling studies on the

β2-adrenergic receptor. This Glu residue should face the interior of the TMH bundle, and in fact the **erm** sites correlated with E3.41, shown schematically in Figure 7.26 on a projection map, indicates that it is oriented towards H2-H7. Note that this figure represents the packing arrangement at the cytoplasmic boundaries. The only change in this projection map relative to Figure 2.16 is a rotation of H7 according to criterion presented in Section 7.2.2, i.e. the face twist induced by the NP motif in H7, indicated by an arrow. The residues corresponding to the cytoplasmic extension of H7 have been incorporated since they pertain to this **erm** set. Figure 7.26 illustrates the consistency of the **erms** related to E3.41 with α-helical character in the this portion of H7.

The correlated M3.41E **erms** H3.42T-L252V-V253M indicated that in the β2-adrenergic receptor E3.41 is oriented towards the H2-H3 interface at the level of the conserved D2.50. Because there is no positive countercharge proposed for D2.50 within the TMH domain, the inference that two acidic residues (D2.50-E3.41) would be in close proximity at this locus of the receptor seems energetically highly unreasonable. However, among the **erms** of E3.41 is an Arg residue at a position of a conserved Phe, i.e. F7.56R. Because we have proposed a D2.50-N7.49 interaction (Zhou 1994; Sealfon 1995) (see Section 7.3.1), and residue 7.56 is only two turns away from N7.49 and facing H2 as N7.49 (note that they are just 7 residues away), modeling suggest that an interaction of R7.56 with E3.41-D2.50 is feasible, and would explain satisfactorily the role of E3.41 and the observed **erms** at these loci. The

direction of R7.56 upwards and towards E3.41-D2.50 is supported by the **erm** F7.48 which can stabilize R7.56 in this orientation. A conserved β-branched residue at position 2.53 would prevent E3.41 from coming into the interior of the TMH bundle, and thus becomes a Met. The conserved Leu at 2.52 becomes a Val which would reorient E3.41 towards the protein interior. Interestingly, H3.42 becomes a Thr, consistent with the proposed role of the conserved H/N3.42 in stabilizing D2.50, which in the β2 receptor is already stabilized by R7.56. The proximity of these **erm** sites in the model is shown in Figure 7.27. The modeling described above for the **erms** related to E3.41 can now help to explain the surprising observation that a stretch of 12 residues in H3, from F3.35 to S3.47, was predicted to face the protein interior in Section 6. I. This observation was rationalized in Section 7.4.4 based on the role of the 100% conserved T3.37 and C3.44. I have now explained how the acidic residue at 3.41 and the highly polar character at H3.42 can be integrated into a three-dimensional model where some of these residues do face the lipid membrane exerting an important structural role (T3.37, C3.44), or can be reoriented into the protein interior due to their side chain length (E3.41) as supported by related **erms**. The consistency among all these **erm** within the H3-H2 interface modeled supports and determines the incorporation of H2 into the H3-H4-H5-H6 template, shown in Figure 7.27.

It is interesting to consider the consistency of the H2-H3 interface here modeled (for the inactive state of the receptor) with the proposed interaction

between R3.50...D2.50 in the activated state of the receptor presented in the context of the Arg-cage motif (Section 7.3.2). The axial displacement between H2-H3 required for the R3.50-D2.50 interaction is inconsistent with the modeled H2-H3 interface. A reference interaction to compare both cases is the D2.50-H3.42 interaction, which is inconsistent with the axial displacement required for the R3.50-D2.50 interaction by about 3-4 Å. Because H3.42 is a primary candidate to H-bond D2.50 in the inactive state and all other **erms** suggests a similar height as modeled an described above, I would interpret this finding as suggesting that activation involves a significant axial displacement (3-4 Å) between H3 and H2. This is supported by the Zn binding site 2.64-3.28 (Elling 1996), which is at the same height in the inactive state and one turn off in the active state of the receptor. This could explain the inactivating effect of Zn binding in the NK1R mutant H2.64-H3.28 (Elling 1996), because it would prevent the R3.50-D2.50 interaction thus destabilizing the active state and stabilizing the inactive state.

7.4.8- Modeling the NPxVY motif in H7 into the H1-H2-H3-H4-H5-H6-H7$^{(7.35-7.46)}$ template.

The segment corresponding to the controversial NP motif in H7, residues 7.46-7.53, can now be refined in detail within the context of the receptor TMH domain already modeled. I have proposed in Section 7.2.2 a conformation of the NP motif that satisfied the criterion of a significant

reorientation of N7.49 towards H2 at the level of D2.50. The proposed interaction between N7.49 and D2.50 (Zhou 1994; Sealfon 1995) was described in Section 7.3.1 and shown in Figure 7.7 where N7.49 acts as H-bond donor to D2.50 and to the C=O7.46, position i-4 within the PK. The same figure shows how Y7.53 can further stabilize D2.50 by H-bonding. This H-bonding scheme is extended by the H-bond of H3.42 to D2.50 described in the last Section (see Figure 7.27).

The significant face twist obtained for the modeled H7 PK described in Section 7.2.2 induced a concomitant large bend at the NP motif, as shown in Figure 7.28. This large bend is important in facilitating the packing of H7 against H6, as explained below, and provides a structural link between the aromatic cluster motif discussed in Section 7.3.3 and the NPxVY or H2-H7 motif discussed here.

A salient feature of the face of H6 facing H7 is how the phenyl ring of F6.44 is bulging out towards H7, shown in Figure 7.28. F6.44 is 100% conserved and a member of the aromatic cluster discussed earlier in section 7.3.3, where I explored the conformational properties of the aromatic residues present in H6. Simulation of an isolated H6 indicated that the conformation of the F6.44 side chain is severely constrained by residues W6.48 and V6.40. It was found that when W6.48 adopts the X1=g+ rotamer it forces F6.44 to adopt also the X1=g+ conformation, i.e. $0 < X1 < -120$. However, the range of X1 angles that F6.44 can adopt is further restrained by the presence of a β-

branched residue four residues preceding F6.44, i.e. V6.40. Both MonteCarlo simulations and a search of the Protein data bank indicate that a β-branched residue at position i-4 does not prevent a Phe from adopting the X1=g+ rotamer, as it does for a Trp residue (see Section 7.3.3.2). But the β-branched residue does prevent X1 angles lower than 80°, further restraining the conformational space available for the F6.44 side chain. Therefore, F6.44 is significantly more rigidly held that any other aromatic side chain within the receptor model, and in terms of the Van der Waals requirements, a big protuberance from H6 towards H7. Only Val or Ile occur at position 6.40, thus maximally conserved in its β-branched character responsible for the conformational restraint on F6.44 described here. Therefore, the 100% conservation profile of V/I6.40, F6.44, and W6.48 suggest that this structural motif is preserved in all neurotransmitter GPCRs and is probably an important determinant of the structure and/or function of the receptor.

The "protuberance" on H6 at the H6-H7 interface created by V6.40-F6.44 would normally prevent close proximity between H6 and H7, as required to attain a packing density in accordance with known protein structures. Instead, the "bulge" created by V6.40-F6.44 represents the necessary complementary Van der Waals surface on H6 to the large bend present in H7 as a result of the NP motif conformation modeled. This is illustrated in Figure 7.28: Note that the large bend present in H7 at the level of the NP motif has a shape that is complementary to the bulge created in H6 by V6.40-F6.44, allowing for close

contacts at the H6-H7 interface. Note that this interaction involves some of the most conserved residues within GPCRs: V6.40-F6.44 in H6 with the NPxVY motif in H7. Notably, this interaction provides a direct link between two organizing motifs or structure-functional microdomains of the receptor structure, the aromatic cluster and the H7-H2 motif, both of which have been implicated in the receptor activation mechanism.

Packing the NPxVY motif into the H2-H3-H4-H5-H6-H7ec template to reach the final conformation presented above was performed starting from the conformation described in Section 7.2.2 and our published paper (Fu 1996), where it was substantiated experimentally by SCAM technique. This conformation was characterized by *phi* and *psi* values that would correspond to a 3-10 helix for the NP dipeptide. Following the principle of helix-helix packing based on ridges-into-grooves observed in known protein structures (Chothia 1981), detailed packing of F6.44 and V6.40 as they protrude from H6 into the H6-H7 interface at the level of the NPxVY motif indicated the need to enlarge the initial α-helical groove of the NPxVY turn to accommodate these bulky side chains. The enlargement of the α-helical turn was accomplished by decreasing *phi* and increasing *psi* values, until the typical values observed for 3-10 helices are reached and constitute the next stable energy minimum conformation of the helical backbone (note that the opposite direction also available for modifying *phi* /*psi* angles was discarded because it would tend towards a pi-helix whose groove is thinner relative to an α-helix). Once the

side chains of V6.40-F6.44 were fitted within the groove of the NPxVY turn, the resulting conformation contained 3-10 helical values for the step NP and the step VY. To refine the resulting backbone geometry a search was performed on known protein structures for segments containing the desired structural elements. Searching for a 12-residue fragment, equivalent to 7.44-7.55, the criteria included a Pro residue at the position equivalent to 7.50, where 7.44-7.45-7.46 were in an a-helical conformation (H) and the rest in an overall helical conformation (h, any kind of helical) except for 7.52-7.53 being in a 3-10 helical conformation (G). The result was 12 known structural motifs, as shown below:

```
Query:
                777777777777
                444444555555
                456789012345
   sequence   = ??????P?????
   Structure = HHHhhhhhGGhh
        Resolution = 0-3 Å

Results: Secondary structure conformations shown by H,h,G (see text)

                           777777777777
 PDB code/                 444444555555
 residue id    sequence    456789012345  resolution
1HGE/A0068- DALLGDPHCDVF HHHHhgGGGGGG      2.6
1FCS/-0031- RLFKSHPETLEK HHHHHhGGGGGG      1.6
1ACK/-0207- GMHILSPGSRDL HHHHHhGGGGGG      2.8
1ACK/-0445- EFVFGLPLVKEL HHHhgGGGgGGG      2.8
1CPT/-0012- ARTVILPQGYAD HHHHHhgGGGGh      2.3
1HNC/A0165- RNQVFEPSCLDA HHHHHhGGGGGG      3.0
1ITH/A0032- KYLTAYPGDLAF HHHHHhGGGGGG      2.5
1MBA/-0031- ALFEKFPDSANF HHHHHhGGGGGG      1.6
1BAB/A0032- RMFLSFPTTKTY HHHHHhGGGGGG      1.5
1BAB/B0030- RLLVVYPWTQRF HHHHHhGGGGGG      1.5
1LH6/-0032- LVLEIAPAAKDL HHHHHhGGGGGG      2.0
2LHB/-0040- KFFTSTPAAQEF HHHHHhGGGGGG      2.0
```

Notably, all known structural motifs that satisfy the structural requirements imposed on H7 by V6.40-F6.44, i.e. where an α-helix is followed by a Pro and two residues (i+2,i+3 from Pro) in the 3-10 helical conformation, result in a continuous stretch of 6 residues (7.49/7.50-7.54/7.55) adopting the 3-10 helical conformation. It follows that the packing requirements that result from the special properties of V6.40-F6.44, protruding from H6 and conformationally severely restricted, induce a continuous stretch of 3-10 helix on H7 from the NP motif onwards.

The prediction resulting from the analysis described above of a 3-10 helical segment following the NP motif in H7, was contrasted with available experimental data on GPCRs. The presence of a 3-10 segment on a GPCR would have been determined by current FTIR techniques. A detailed FTIR study on bovine rhodopsin by Garcia-Quintana et al. (Garcia-Quintana 1993) characterized a band at 1639 cm-1 as evidence for a significant amount of 3-10 helix (15-17 % total) present in the rhodopsin structure. This is consistent with the prediction of a 3-10 helical segment following the NP, but these FTIR experiments could not assign the observed bands to particular sequence motifs. This assignment was performed by analogous FTIR studies on rhodopsin combined with protease treatment, known to remove from rhodopsin the so-called "I4-loop" at the level of the H7 cytoplasmic extension proposed earlier. The band at 1639 was missing or significantly altered by protease treatment, which lead the authors to conclude that this (an other)

bands could be assigned to the "I4-loop", consistent with the prediction of a 3-10 helix at least from P7.50 to L7.55. It should be noted that the proteases used (papain and protease K) also cleave other loops, which together with possible indirect conformational effects imply that this data are supportive but not conclusive for the prediction of a 3-10 helix at the H7 segment following P7.50. This same work also showed that DTT treatment, which removes the palmitate from Cys 7.69 and 7.70 in rhodopsin, had a deleterious effect on the band at 1639 cm-1 assigned before to a 3-10 helix. Furthermore, the effect of DTT treatment on the 3-10 helical band was much more pronounced upon rhodopsin activation, suggesting significant changes at this locus, an observation confirmed by further studies of Garcia-Quintana et al. (Garcia-Quintana 1995) on the changes upon activation of the 3-10 helix band characterized previously (Garcia-Quintana 1993).

Modeling a 3-10 helix from N7.49 to L7.55 facilitated a conformation of Y7.53 where the OH group can H-bond D2.50 as described above. The H7 turn following T7.54 was characterized in Section 5 as a transition between two predicted helical segments, 7.50-7.54 and 7.59-7.70. The conservation pattern was consistent with a continuous helix. However, the presence of some insertions and non conserved Pro residues were indicative of a breakdown of regular secondary structure (see Section 5.6). In the model presented here, I have included the stretch of residues between 7.54 and 7.57 as a continuation of the 3-10 helix modeled above for the NPxVY motif. This assumption

provided a continuous packing at the H6-H7 interface and facilitated the orientation of the conserved F7.56 towards the protein interior, which was used above in the context of packing the H2-H3 interface to model an ionic bond between R7.56-E3.41 present only within β2-adrenergic receptors (see Figure 7.27). However, the conformation of this turn would be strongly dependent of the desired positioning of the cytoplasmic extension of H7, packing against the remaining of the TMH domain at the cytoplasmic boundaries. This packing is not presented at this stage because there are very few experimental constraints available and it is not clear whether it should be treated as a continuous component of the TMH domain. The possible structural and functional implications of incorporating the H7 cytoplasmic extension (7.59-7.70) into the resulting model of the TMH domain are discussed below in a section dedicated to inferences for future work, where specific experiments are designed that would provide necessary data to guide modeling of this cytoplasmic segment.

7.4.9- Incorporation of H1 into the H2-H3-H4-H5-H6-H7 template.

The model of the 7 TMH domain is completed by the incorporation of H1 to the remaining 6 TMHs already modeled. The conservation pattern is defined by N1.50, V1.53, and L1.63 being 100% conserved, with G1.49 and V/I1.57 being also highly conserved within neurotransmitter receptors. H1 is characterized by a more buried cytoplasmic area and an almost completely

exposed extracellular side (see Section 6.3 and Figure 6.9). The rhodopsin map suggests that H1 is significantly tilted towards the interior at the cytoplasmic side (Unger 1995), consistent with my prediction that at the cytoplasm H1 is much more buried than at the extracellular side (see Section 6.3). The predicted pattern of lipid exposure has a remarkable kink in the directionality of the interior facing residues, located around residues L152-V153 (see Figure 6.9). H1 is facing H2 and H7 in the receptor model. The observed kink in the predicted accessibility pattern has a directionality in 3D that follows the path of the modeled kink at the PK of H7, and can thus be used to guide packing of H1 onto the H2-H7 template, as shown in Figure 7.29.

There are no experimental constraints on helix-helix interactions between H1 and H2-H3-H4-H5-H6. The only experimentally derived residue interactions on H1 are a double revertant mutant between 1.39 and 7.36 on H7 for muscarinic receptors (Liu 1995). The misfolding of the receptor induced by a H7.36(423)->T mutation was reverted by a second mutation T1.39->A on H1, suggesting that the simultaneous presence of T7.36 and T1.39 resulted in a steric conflict within the native receptor structure. The axial displacement suggested by these experiments between H1-H7 is consistent with the experimentally proposed interrelated role of N1.50 with N7.49 (Perlman 1997). This criterion positions the cytoplasmic ends of H1-H2 at a similar height, consistent with the prediction of a very short loop (3-4 residues) between these segments (see Figure 6.9).

The completely conserved N1.50 is facing H7 at the level of the NP motif, and can be modeled to H-bond the only H-bond free backbone carbonyl of the PK in H7, thus stabilizing the modeled NP conformation. This positions V1.53 to pack against the Pro ring, thus defining an interaction between the 100% conserved residues in H1 (N1.50-V1.53) and the conserved NP motif in H7. The H1-H2 interface at this level is characterized by the close packing of the 100% conserved A2.47 within the groove defined by V1.53-V1.57 and N1.50, completing the interaction among 100% conserved residues between H1-H2-H7 at this level. This set of interactions is shown in Figure 7.29.

The resulting packing of H1 with H2-H7 at the extracellular side is characterized by the presence of W7.40 which imposes an outward tilting on H1 and H2 at this level, as shown in Figure 7.30. However, it is not possible to bring T1.39-T7.36 in direct contact as suggested by the double revertant mutant character (Liu 1995). Instead, I would rationalize the experimental data by proposing that a Thr residue at position 7.36, i-4 from W7.40, would force W7.40 to adopt the X1=t rotamer (see the I4.56-W4.50 case in Section 7.3.3.2 for a detailed description), which would clash with T1.39.

The interactions of W7.40 with H1-H2-H3 are likely to be important for receptor stability and indicate a source of structural heterogeneity between rhodopsin and the 5HT2CR. W7.40 needs to adopt the X1=g+ rotamer conformation to face the binding site, as indicated by covalent attachment of a

cyano-pindolol derivative in α-adrenergic receptors (Wong 1988). W7.40 adopting the X1=t configuration would orient the side chain towards the lipids, which is inconsistent with its absolute conservation observed within neurotransmitter receptors. The effect of the bulky W7.40 adopting X1=g+ is to create significant steric hindrance with H1 and H2 that can be resolved by tilting H1 and H2 outwards at the extracellular side of the receptor. This represents a molecular basis for the predicted tilt of H1 and H2 in Section 6.3, in agreement with the tilts proposed for these helices in the rhodopsin electron density structure (Unger 1995). However, rhodopsin lacks the conserved Trp at position 7.40. This absence can be correlated with the different patterns of Pro residues in H1 and H2 of rhodopsin, characterized by the presence of a Pro at position 1.48 and the absence of the Pro at position 2.59, relative to neurotransmitter GPCRs. The net effect of this divergent pattern of prolines, based on the PKs expected at the corresponding positions, is structurally equivalent to the effect induced by W7.40. It follows that the pattern of helix tilting and packing observed in the rhodopsin structure may differ from the equivalent receptor region in the 5HT2CR and other neurotransmitter GPCRs. This illustrates the caution necessary in transferring structural features observed in rhodopsin to models of other GPCRs. Finally, W7.40 modeled as described above is a member of a set of highly conserved aromatic residues, consisting of W7.40-W3.28-Y7.43 and the 100% conserved W2.71. The latter is three residues away from the extracellular end of H2 and

is thus not included in the present model. Taken together, these considerations suggest a structural role for the extracellular domain comprised between H1-H2-H7, and identify an important structural role for W7.40 which would explain its 100% conservation pattern.

The cytoplasmic side of H1 after the kink in the accessibility pattern observed at L1.52-V1.53 is more buried that its extracellular counterpart and is packed against H2 and H7. There is a non conserved Glu at position 1.60 which is likely to have significant structural effects. This position shows an **erm** with position 7.59 in that an acidic residue is always present in either of them, as shown below:

```
160 EEEEEEEEEEVVVEETTTTTTTTTTTTTTTFFFFTNNNNNHHSSSSSSSSSSSSSSSFFFFEEEEEEEEENNNY
              vvv  ttttttttttttttttfffftnnnnnhhsssssssssssssssffff          nnny
    eeeeeeeeee   ee                                                 eeeeeeeee
              ddddddddededddddddddddddeeeeeeeedddddddddddddddddddeeeeeeeeeeddddd
    iiitttttttt
759 IIITTTTTTTTDDDDDDDDDEDEDDDDDDDDDDDDDEEEEEEEDDDDDDDDDDDDDDDDDDEEEEEEEEEDDDD
```

This implies that E1.60 within 5HT2 receptors could fulfill the same role that an acidic residue at position 7.59 conserved for all other neurotransmitter receptors. This is supported by spin labeling experiments in rhodopsin which found C7.63, one turn away form 7.59, in proximity to C1.60 (closer than 10 Å) in the inactive state (Yang 1996a). Interestingly, the distance between residues 1.60-7.63 increased significantly upon activation, suggesting a role of these loci in conformational changes involved in the activation mechanism. A highly conserved acidic group (E1.60-E/D7.59 combined) facing the protein interior

should be stabilized by specific interactions. This role can be fulfilled in the model by N2.40, another highly conserved residue, H1.64 (90 % conserved H-bonding donor at position ε, see Table 4.2), and by the adjacent K1.61. The latter is also a highly conserved R/K, 100% conserved except for the β-adrenergic receptors, where an Arg at position 7.56 substitutes for R/K161, as shown below:

```
756 FFFFFFFFFFFFFFFSSSSFFFFFFFRRRRRRRRRSSSSSSSFFFFFFFFFFFFFFFFFFFFFFFFFFFFLLLF
                              +++++++++
    ++++++++++++++++++++++++++         ++++++++++++++++++++++++++++++++++++++++
                              pppqeeeep
161 KKKKKKKKKKKKKRRRRRRRRKRKRRRPPPQEEEEPRRRRRRRRRRRRRRRRRRRRRRRRRRKKKKRRRRRRRRK
```

I described the possible role of this R7.56 within β-adrenergic receptors related to other **erm** characterized by the E3.41-R7.56 pattern and shown in Table 7.7. This suggests a Lys at position 1.61 could fulfill a similar role as an Arg at position 7.56. It is clear that an E1.60-K1.61 ionic interaction is feasible for the 5HT2CR. For the β-adrenergic receptors it also clear that R7.56 could interact with E7.59. It follows that the combination of the observed **erms** at this locus suggests an interaction between E1.60-E/D7.59 and K/R1.61-R7.56 that is preserved among all GPCRs analyzed, and that includes N2.40 and H1.64 as an intrinsic component (thus defining another microdomain within the receptor model). I have modeled the E1.60-K1.61-N2.40-H1.64 interactions as described above in the 5HT2CR model, as shown in Figure 7.29.

It is noteworthy that there are clear indication of further interactions

between this polar/charged microdomain and other previously described polar/charged motifs. For example, R7.56 in β-adrenergic receptors could interact simultaneously or at different states of the receptor with E/D7.59 described above, and with the D2.50-E3.41-N7.49 motif as discussed above and shown already in Figure 7.25. On the other hand, this microdomain is adjacent in space to the Arg-cage motif, and the **erm** between ED7.59-E1.60 could be extended towards the E6.30-R3.50 proposed interaction (see Section 7.4.1) as suggested by the following **erm** :

```
629 NNNNNNNNNNRRRRRRRRRRRRRRRRRRRRRKKKKRRRRRRRRRRRRRRRRRRRRRRRKKRRRRKKKKRRRRRRRRGK
            rrrrrrrrrrrrrrrrrrrrrkkkkrrrrrrrrrrrrrrrrrrrrrrrkkrrrrkkkrrrrrrr k
    nnnnnnnnnn
    iiitttttttt
            ddddddddededddddddddddddeeeeeeeedddddddddddddddddddeeeeeeeeeddd
759 IIITTTTTTTTDDDDDDDDDEDEDDDDDDDDDDDDDEEEEEEEDDDDDDDDDDDDDDDDDDDEEEEEEEEEDDDD
    eeeeeeeeee   ee                                                  eeeeeeeee
            vvv  ttttttttttttttttfffftnnnnnhhsssssssssssssssffff          nnny
160 EEEEEEEEEEVVVEETTTTTTTTTTTTTTTTFFFFTNNNNNHHSSSSSSSSSSSSSSSSFFFFEEEEEEEEENNNY
```

An interaction between the cluster of charged/polar residues at the cytoplasmic interface of H1-H2-H7 and the cluster of charged/polar residues at the interface of H3-H6 (the Arg-cage) would be highly favorable energetically, but is prevented in the developed model by the proximity between H7-H2 at the cytoplasmic side. However, several inferences on the conformational changes upon receptor activation developed in previous sections suggest an interaction between these two microdomains in the activated state of the receptor. Spin labeling of rhodopsin Cys residues at 1.60-7.63 indicates that the cytoplasmic portion of H7 (not modeled here) moves away from H1 upon activation (Yang 1996a). I have described in Section 7.3.2 (the "Arg-cage"

manuscript) how R3.50 would be predicted to change its orientation from H6 (E6.30-R3.50) towards H2 (R3.50-D2.50) upon activation. I have also described how the modulation of the PK in H6 by the aromatic cluster would orient the cytoplasmic segment of H6 towards H5 in the inactive state (where E6.30-R3.50), and towards H7 in the activate state (E6.30 towards K1.61). Note from Figure 7.13 of the Arg-cage manuscript in Section 7.3.2 that the proposed interaction between R3.50 with N7.49 and D2.50 upon activation bends the cytoplasmic segment of H7 following the PK outwards, away from H2-H3 and towards the membrane. Such a movement of the H7 segment following P7.50, which is preventing the microdomain defined by E1.60-K1.61 from reaching R3.50-E6.30-D3.49 (the Arg-cage motif), would allow the energetically favorable interaction between these two microdomains in the activated state of the receptor, an interaction favored by the concomitant movement of H6 (E6.30) and H3 (R3.50-D3.49) towards H1-H2 upon activation. It follows that the role of the microdomain surrounding E1.60-K1.61 at the cytoplasmic boundaries would be to stabilize the active state of the receptor.

Residue K1.61 was initially predicted to face the phospholipid headgroups (see Figure 6.9). However, the same figure illustrates that K1.61 lies within the patch of residues predicted to face the protein interior, and thus the initial prediction at the level of helical nets representations (Section 6) was modified and K1.61 considered to face the interior of the TMH bundle.

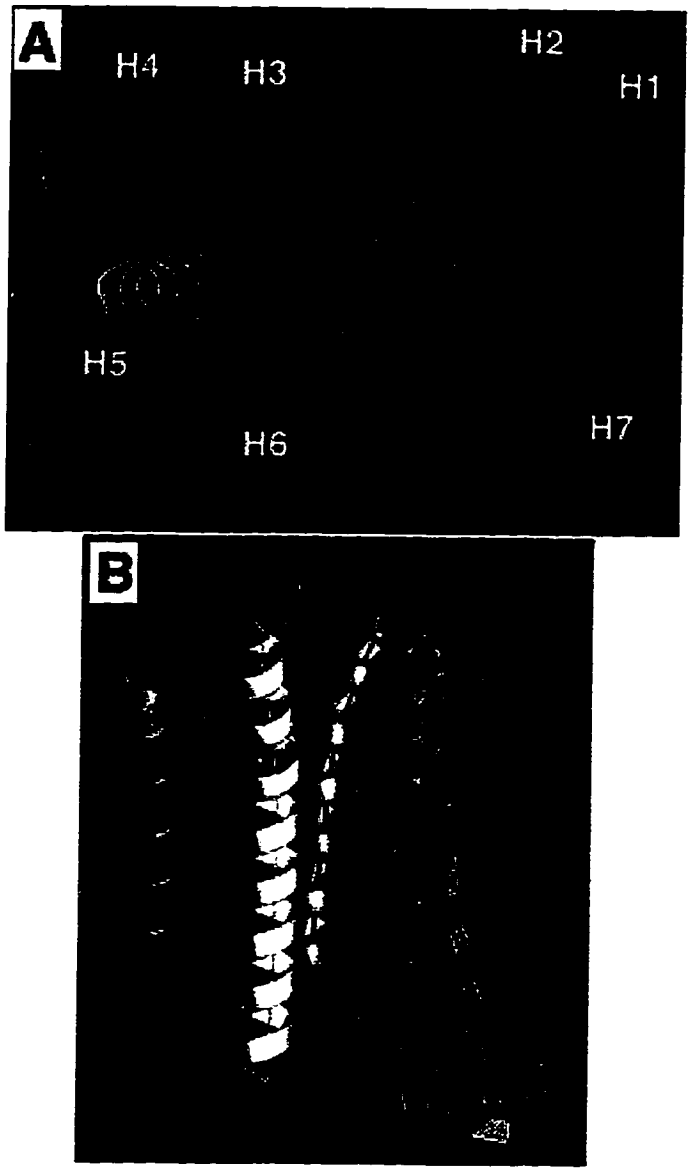


Figure 7.1

100 representative conformations of the Pro-kinks for H6 and H7 obtained by Monte-Carlo simulations are superimposed for each of these TMHs in the context of the initial 3D model of the TMH domain. A) Note the degree of conformational freedom induced by P6.50 and P7.50. This illustrates the wide range of alternative helix-helix interactions that arise from the presence of Pro kinks in H2, H5, H6, and H7 for modeling purposses. B) Note the directionality in the direction of bending relative to a standard helix; H6 bends only inwards, while H7 bends only outwards.

Figure 7.2:

Conformation of the PK in H7 comprising the conserved NP motif. N7.49 is modeled adopting X1=g- and psi=3-10 helix values. This conformation of N7.49 can be stabilized by H-bonding between the N7.49 side chain and the C=Oi-4, as shown, resulting in a significantly kinked H7.

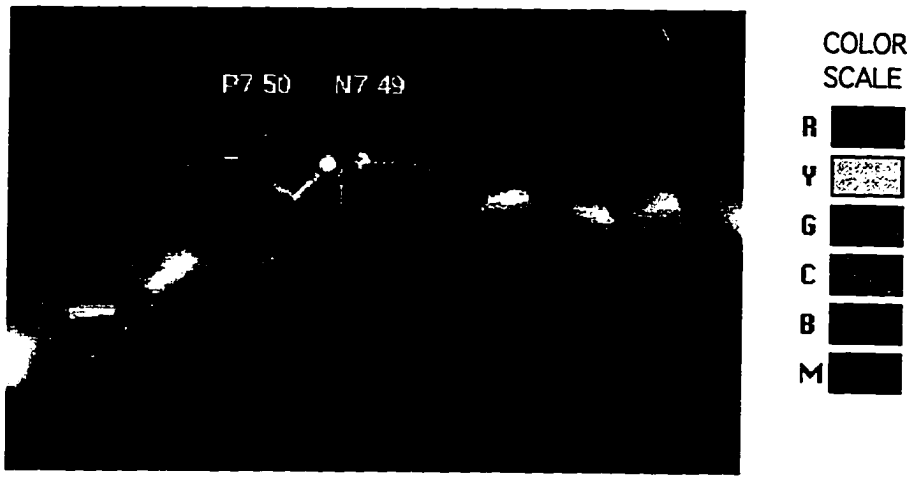
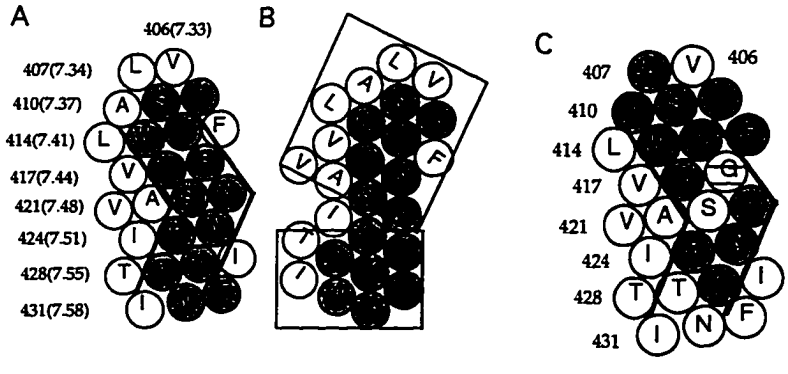


Figure 7.3:

Helical net comparison of our predictions of TMH-TMH interfaces (A,B,) and the experimentally observed accessibility to MTSEA by SCAM techniques (C) on the D2 receptor. In (A) and (B), H7 residues predicted in TMH-TMH interfaces are shown shaded. Note the apparent kink in (A) marked by bold lines, discussed in Section 6.3. (B) the segment is bent and twisted to illustrate how interior-facing residues may lie on a single face of a kinked -helix. In (C), the pattern of residues accessible to MTSEA (shaded) is shown to follow the predicted kink in H7, shown by bold lines.



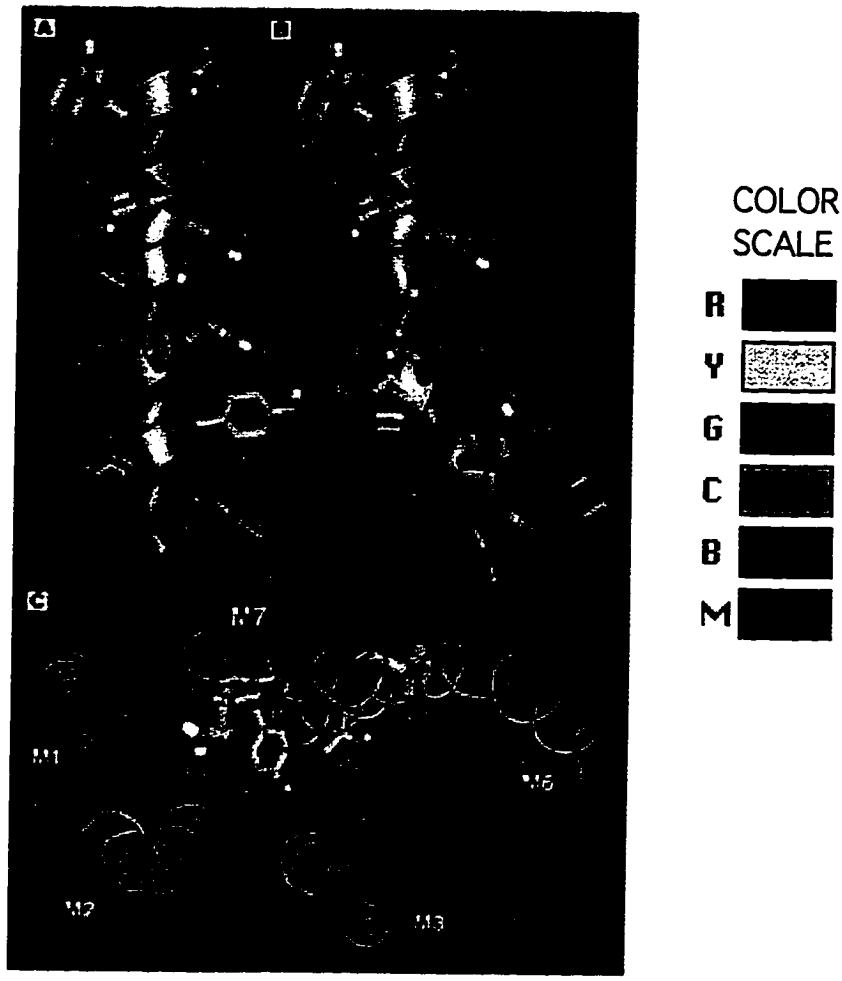


Figure 7.4

Modeled H7 is in agreement with the SCAM data. The residues from Thr412(7.39) to Ile431(7.58) are shown (A) in an ideal -helix and (B) and (C) in a kinked -helix. The residues accessible to MTSEA are shown in yellow; Gly415(7.42) and Ser419(7.46) are shown in an olive color to emphasize their location on a continuous surface patch of accessible residu MTSEA. Note that in (A) the accessible residues do not lie in one face. In (B) and (C), all accessible residues now lie in a common face of the kinked -helix. In (C), because of the twisting inducedby the Pro-kink, Asn422(7.49) is now oriented towards H2 and Phe411(7.38) is oriented towards H6 and H3, in agreement with experimental data (see legend to Figure 4). Note that Thr412(7.39), Trp413(7.40), and Tyr416(7.43) all face the binding-site crevice.

Figure 7.5

Face-twist observed for the PK in H5 from an analysis of the predicted TMH-TMH (green) and TMH-Lipid (yellow) interfaces. A) Van der Waals representation (Pro residue in purple). B) H5 Cα trace showing the resulting orientation of A5.46, Y5.58 and C5.59 (red, Van der Waals). C) Solvent accessibility surface colored as panel B. C) Cα-Cβ projection of an ideal α-helical conformation of H5. Note the absence of the face twist observed in prior panels. E) Handedness of the face twist required on H5 and H7 to satisfy the experimentally derived interactions shown.

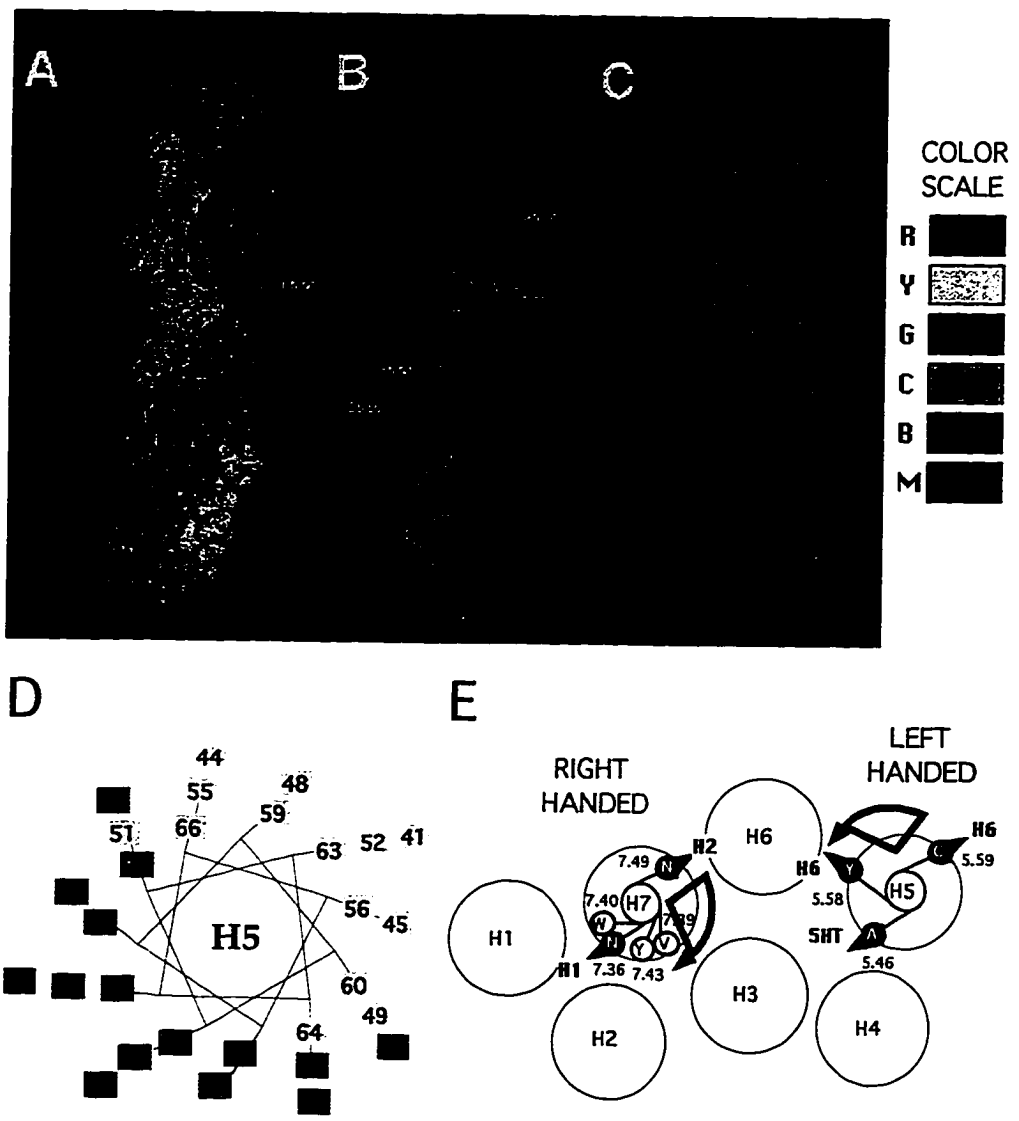


Figure 7.6 :

Comparison of the bending induced in the PK of H6 by the presence of C6.47 in the β2-adrenergic receptor and C6.49 in the 5HT2CR. The helix axis of 100 representative H6 conformations for each receptor, obtained by MonteCarlo simulations of isolated helices, are shown superimposed at the N-terminus. Note the similarity in the conformational space available observed for both constructs.

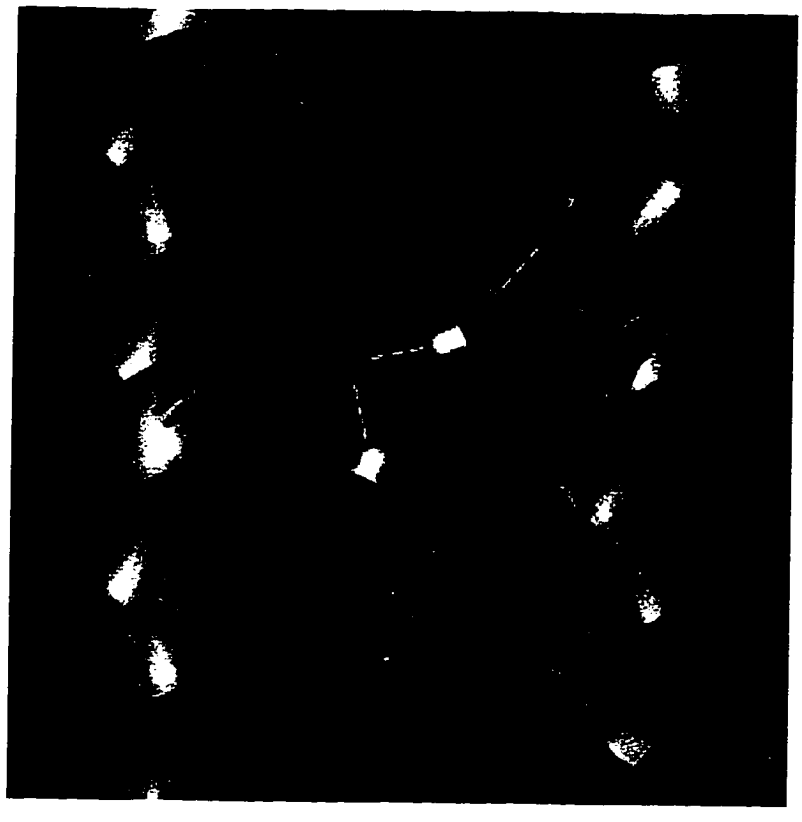


Figure 7.7

Proposed interactions between the NPxVY motif in H7 and D2.50 in H2. N7.49 H bonds the C=Oi-4 from the H7 PK and D2.50. In this conformation, the conserved Y7.53 can H-bond D2.50.

Figure 7.8 : The conservation pattern of this motif is presented in an alignment of several GPCRs.

GnRH-human	ISLDRSLAI
GPR/bombesin-mouse	LSADRYKAV
thrombin-mouse	ISIDRFLAV
interleukin 8-human	ISVDRYLAI
5HT1a-human	IALDRYWAI
5HT2a-human	ISLDRYVAI
α1-adrenergic-human	ISVDRYVGV
α2-adrenergic-human	ISLDRYWSI
β1-adrenergic-human	IALDRYLAI
β2-adrenergic-human	IAVDRYFAI
dopamine D2-human	ISIDRYTAV
opsin-drosophila	ISLDRYQVI
rhodopsin-chicken	LAVERYVVV

(I/L)xxDRYxx(I/V)

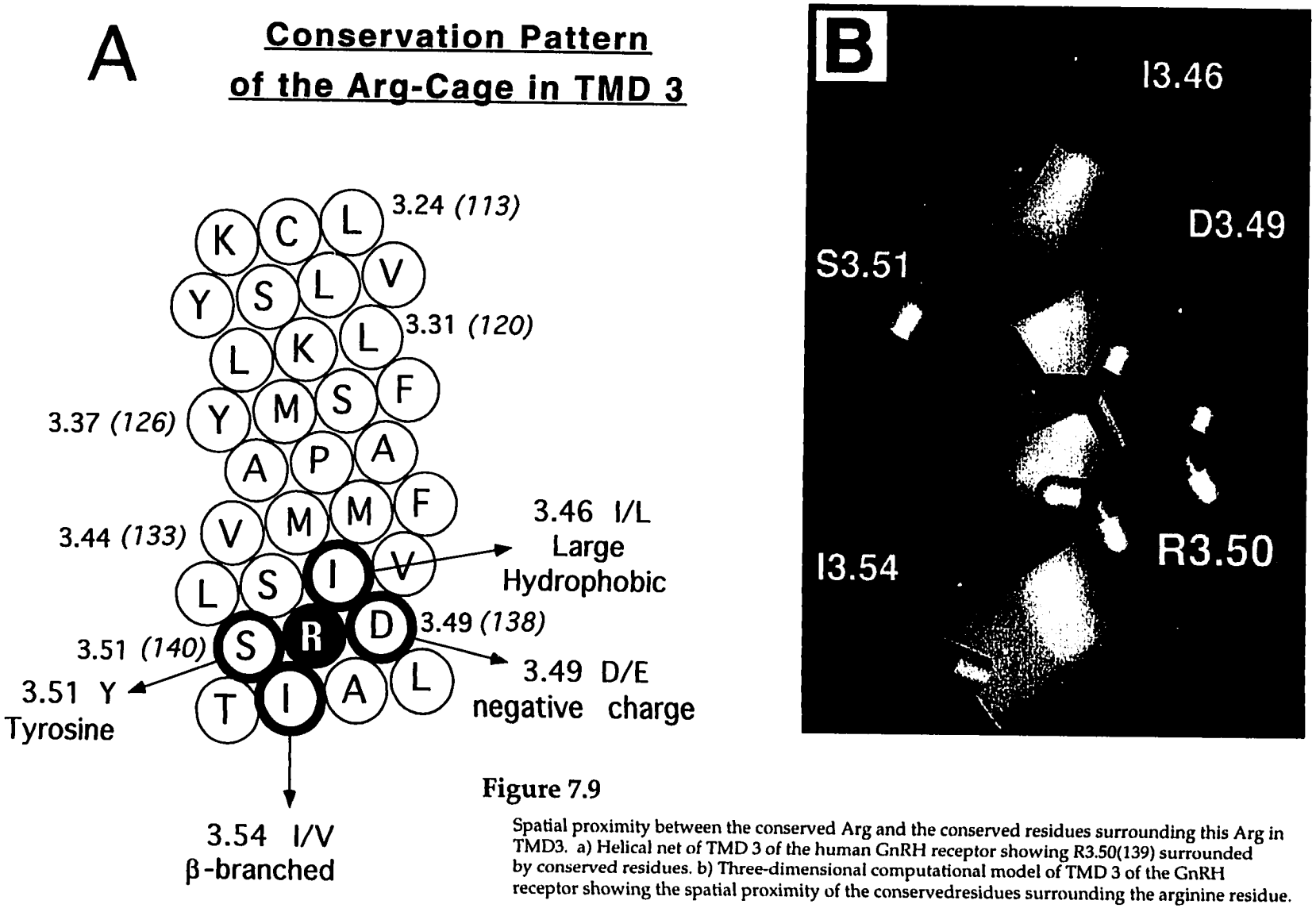


Figure 7.9

Spatial proximity between the conserved Arg and the conserved residues surrounding this Arg in TMD3. a) Helical net of TMD 3 of the human GnRH receptor showing R3.50(139) surrounded by conserved residues. b) Three-dimensional computational model of TMD 3 of the GnRH receptor showing the spatial proximity of the conservedresidues surrounding the arginine residue.

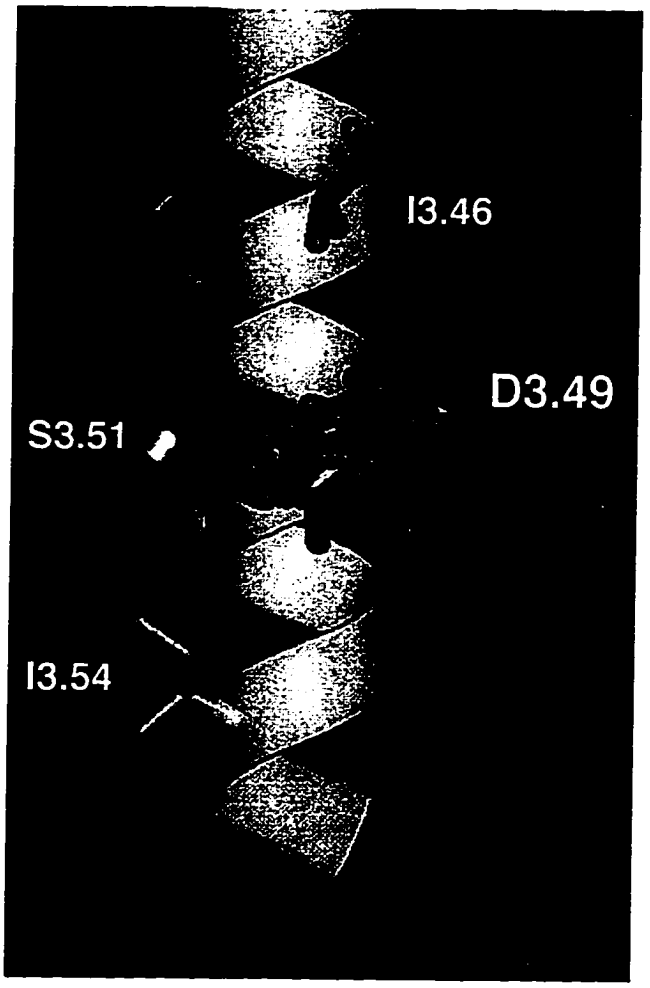


Figure 7.10. Wild type $R^{3.50(139)}$ orientations: Three-dimensional model of TMD 3 of the GnRH receptor illustrating the preferred conformations of the Arg side chain (purple) in the wild type construct. Conserved residues are highlited by thicker bonds. Note that all populated rotamer conformations are oriented towards $D^{3.49(138)}$, driven by an ionic bond between $R^{3.50(139)}$ and $D^{3.49(138)}$.

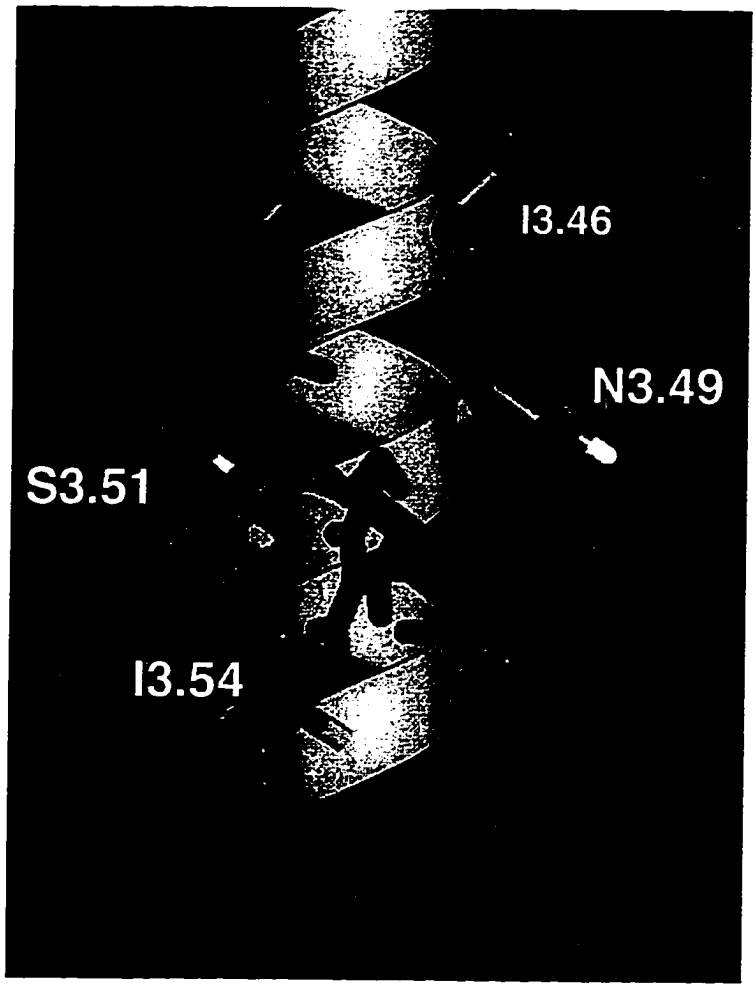


Figure 7.11. $R^{3.50(139)}$ orientations in the $D^{3.49(138)}N$ mutant: Three-dimensional model of TMD 3 of the GnRH receptor illustrating the preferred conformations of the Arg side chain (purple) in the $D^{3.49(138)}N$ mutant construct. Conserved residues are highlited by thicker bonds. Note that all the populated rotamer conformations are oriented away from $N^{3.49(138)}$. Instead, $R^{3.50(139)}$ is oriented either towards $S^{3.51(140)}$-$S^{3.47(136)}$ (37.1%) where it cah H-bond to the Ser residues, or towards the cytoplasm at the level of $I^{3.54(143)}$-$A^{3.53(142)}$ (29.4%) where it can be solvated.

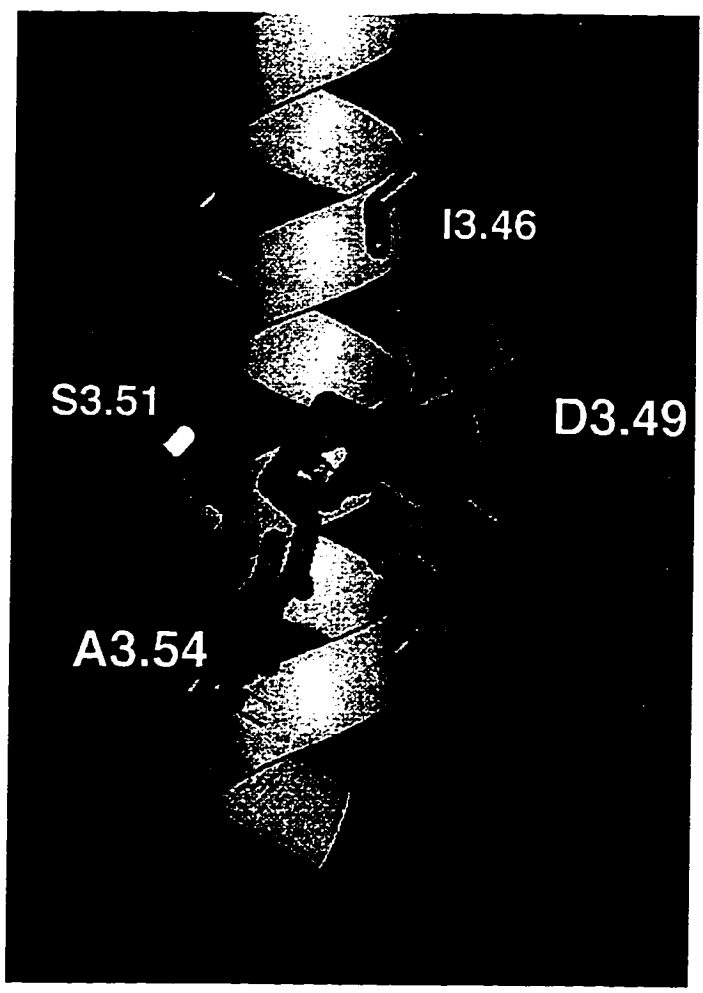


Figure 7.12. $R^{3.50(139)}$ orientations in the $I^{3.54(143)}A$ mutant: Three-dimensional model of TMD 3 of the GnRH receptor illustrating the preferred conformations of the Arg side chain (purple) in the $I^{3.54(143)}A$ mutant construct. Conserved residues are highlited by thicker bonds. Most of the populated rotamer conformations are oriented towards $D^{3.49(138)}$ as observed for the wild type (Figure 4), driven by an ionic bond between $R^{3.50(139)}$ and $D^{3.49(138)}$. However, in this construc $R^{3.50(139)}$ is also significantly oriented towards $A^{3.54(143)}$ (12.7%), where it could be solvated by the aequous cytoplasm.

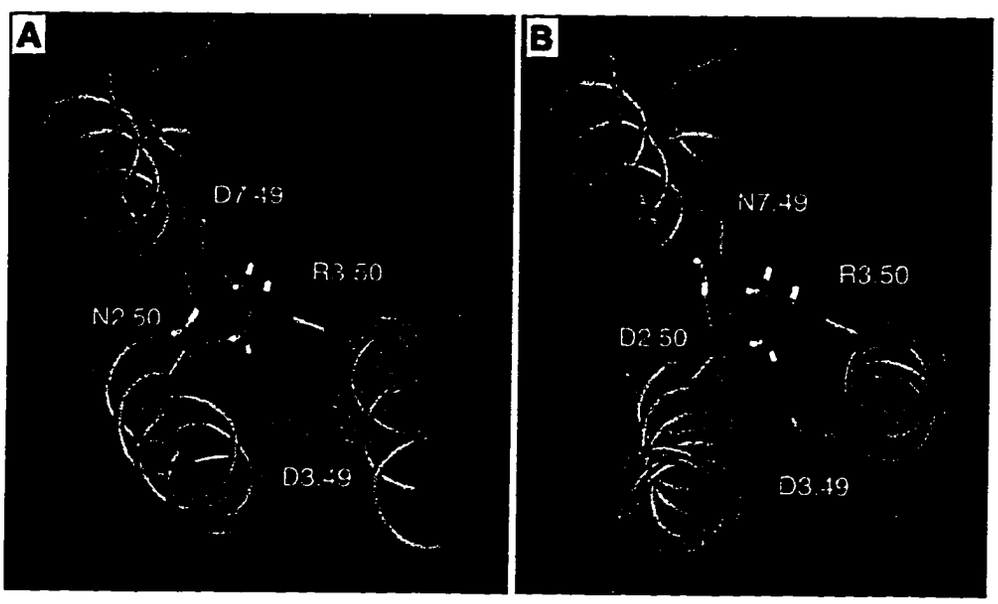


Figure 7.13. Three-dimensional model of TMD2, 3, and 7 of the GnRH receptor showing proposed interactions between $R^{3.50(139)}$ and the 3.49-2.50-7.49 loci. A) GnRH receptor: $R^{3.50}$ interacts with $D^{3.49}$ and the carbonyls of $D^{7.49}$ and $N^{2.50}$. B) 5-HT_{2A} receptor: $R^{3.50}$ interacts with $D^{3.49}$ and the carbonyls of $N^{7.49}$ and $D^{2.50}$. Note that the set of interactions between the Arg residue and the TMD2-TMD7 loci in the GnRH receptor can be mimicked in the 5-HT2A receptor despite the reciprocal arrangement of Asn and Asp at the 2.50 amd 7.49 loci.

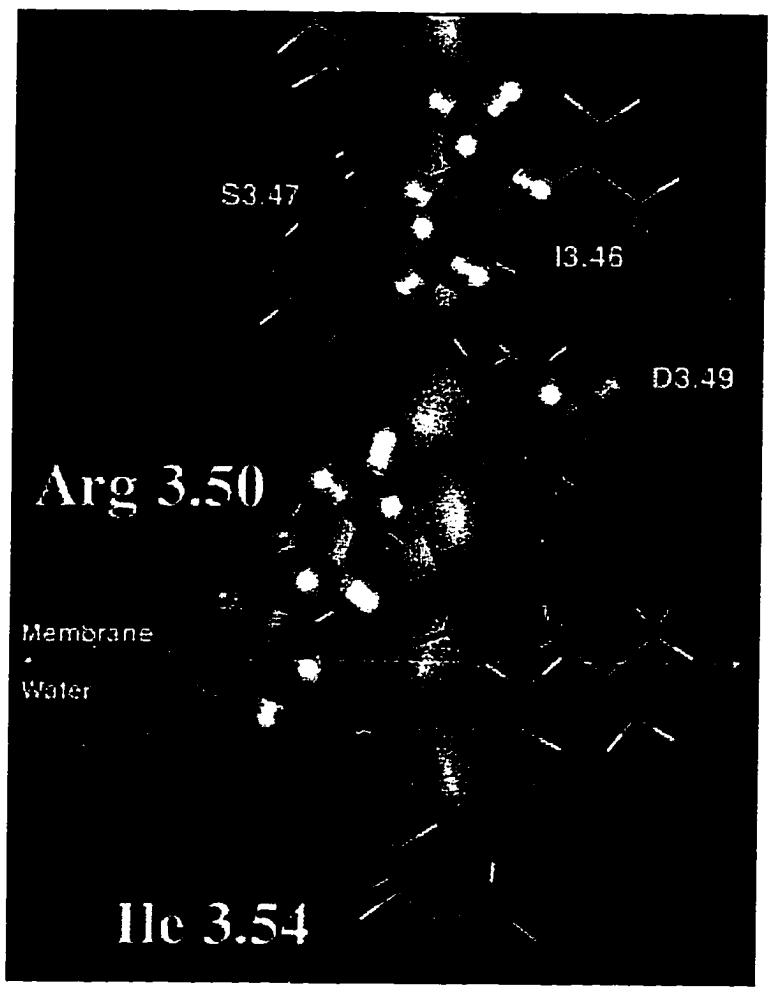


Figure 7.14. Proposed role of $I^{3.54(143)}$ in "caging" $R^{3.50(139)}$ through a steric clash: Three-dimensional model of TMD 3 of the GnRH receptor where $R^{3.50(139)}$ attempts to adopt the (t, g-, t, g+) rotamer configuration, oriented towards the C-terminus or cytoplasmic boundary of TMD 3, but clashes with $I^{3.54(143)}$. Note that in that conformation $R^{3.50(139)}$ reaches the membrane-cytoplasm interface (white line), as determined for rhodopsin (10), and could thus become solvated.

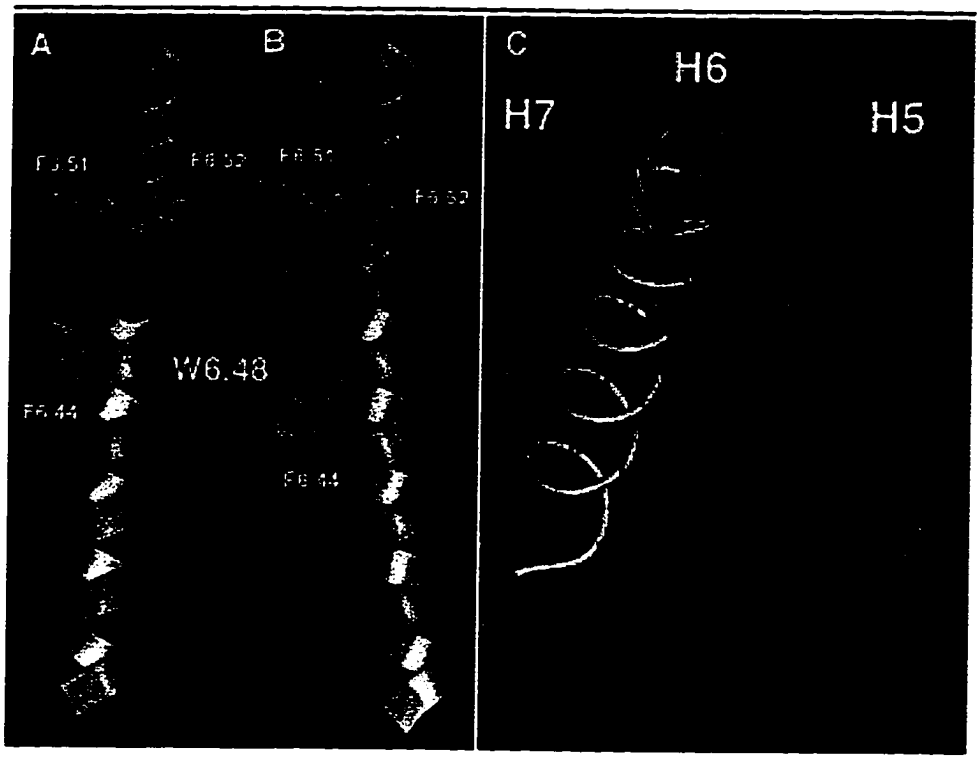


Figure 7.15 :
Alternative conformations of the aromatic cluster in H6 characterized by W6.48 (purple) adopting X1=trans (A) or X1=gauche+ rotamer. Note that the orientation of the indole plane with respect to the membrane plane is nearly perpendicular when W6.48 adopts the X1=g+ rotamer (A), and that it is parallel when W6.48 adopts the X1=trans rotamer (B). (C) Extracellular view of 100 representative conformations of the PK in H6 obtained by MonteCarlo simulations with W6.48 adopting the X1=gauche+ (red) or X1=trans (yellow) rotamer configuration, superimposed at the extracellular segment. Note that the gauche+ configuration (red) is preferentially bent towards H5, while the trans configuration (yellow) is preferentially bent towards H7.

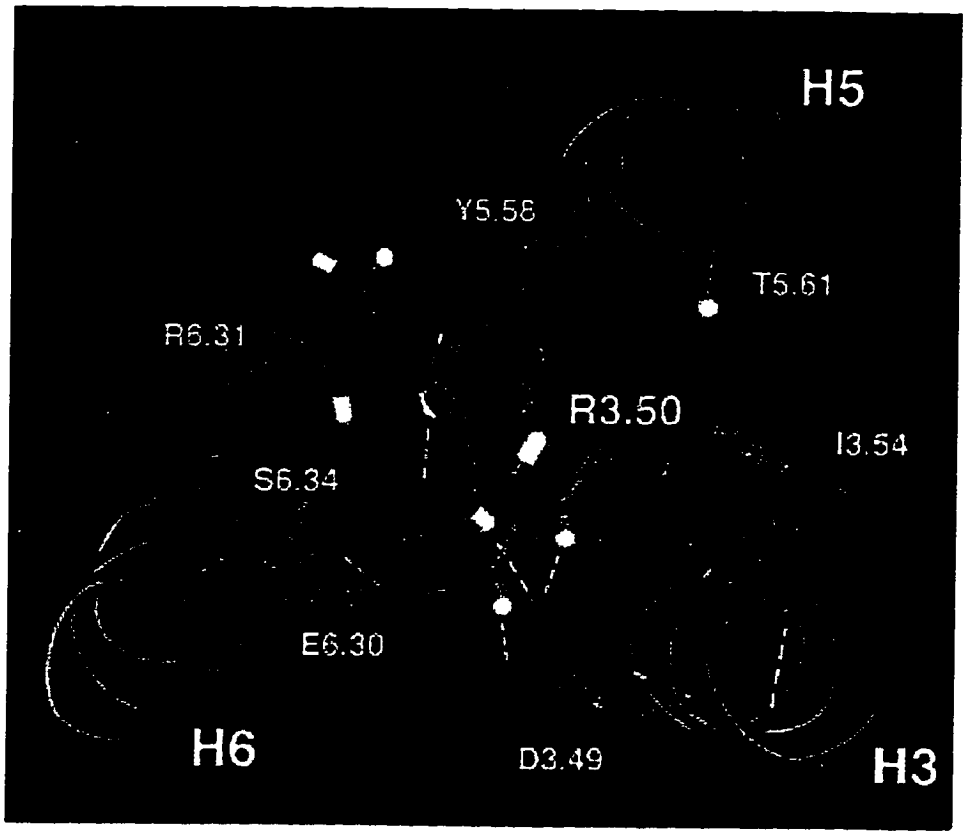


Figure 7.16:
The Arg-cage motif at the cytoplasmic boundaries of H3-H5-H6, comprising residues R3.50, D3.49, E6.30, S6.34, R6.31, Y5.58, T5.61, and I3.54. H-bonding interactions within this motif are shown by white dashed lines. Note the planar stacking between the phenyl group of Y5.58 and Arg3.50. The conserved beta-branched residues T5.61 and I3.54 are oriented towards the R3.50 side chain, constraining its conformation as explained in the text. Note the consistency of the proposed interactions with the predicted TMH-Lipid interface (yellow).

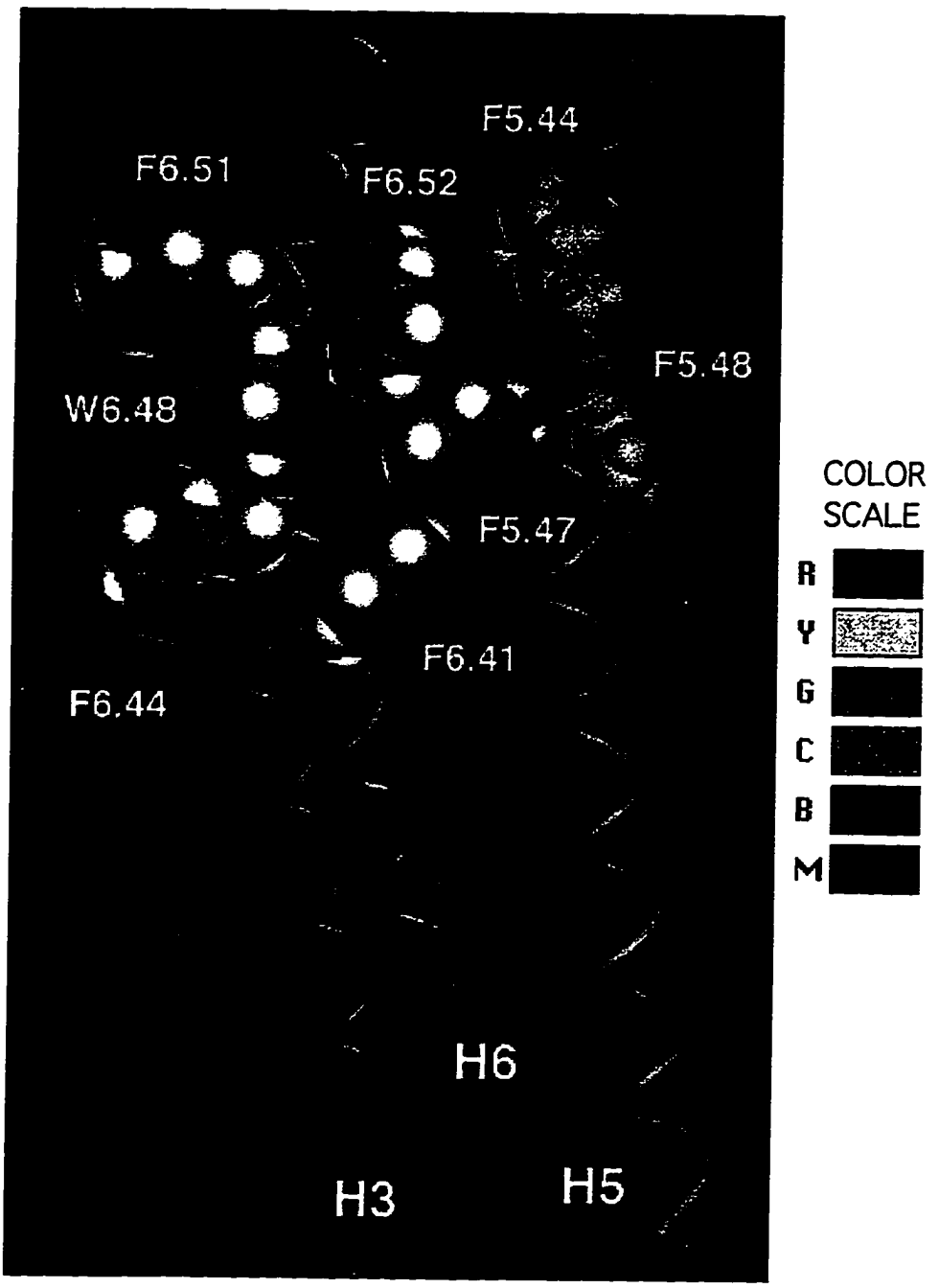


Figure 7.17

H3-H5-H6 packing at the extracellular side. The aromatic cluster fills the space in between H5-H6 interacting in a T-shape geometry. Note the consistency of the proposed interactions with the predicted TMH-Lipid interface (yellow).

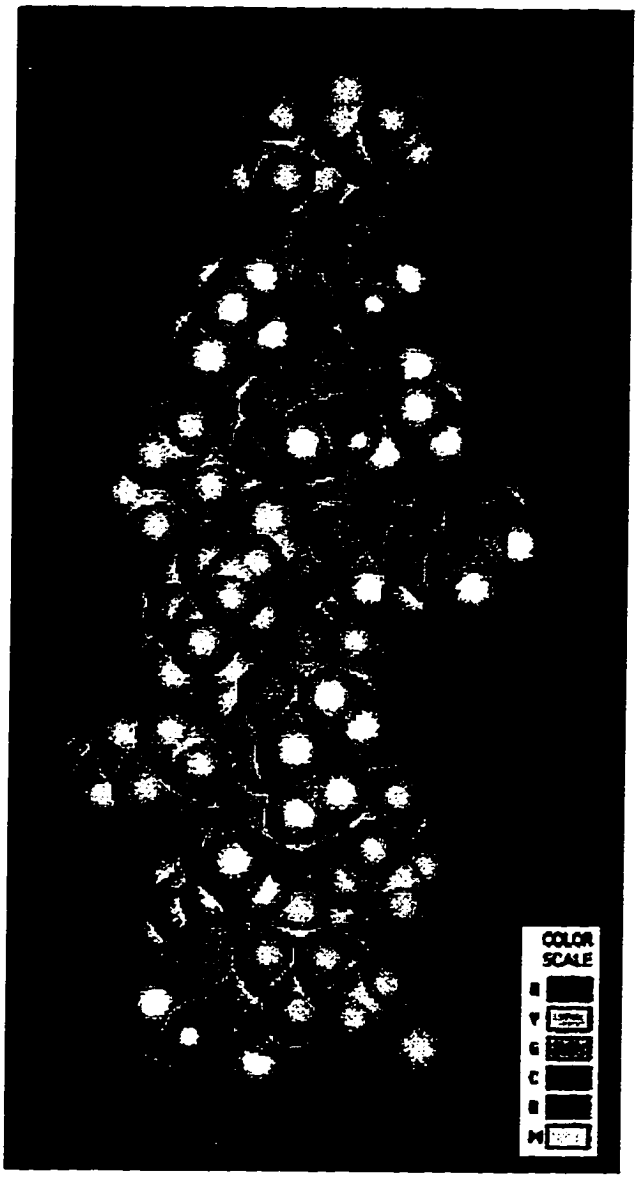


Figure 7.18:
The most salient features of interior facing residues in H4 guiding the helix-helix packing process are due to the Van deer Waals surface pattern of H4 interior facing residues. Predicted interior-facing residues in H4 define a significant bend on the helix surface. Note how the aromatic moiety of W4.50 protudes significantly from the H4 surface.

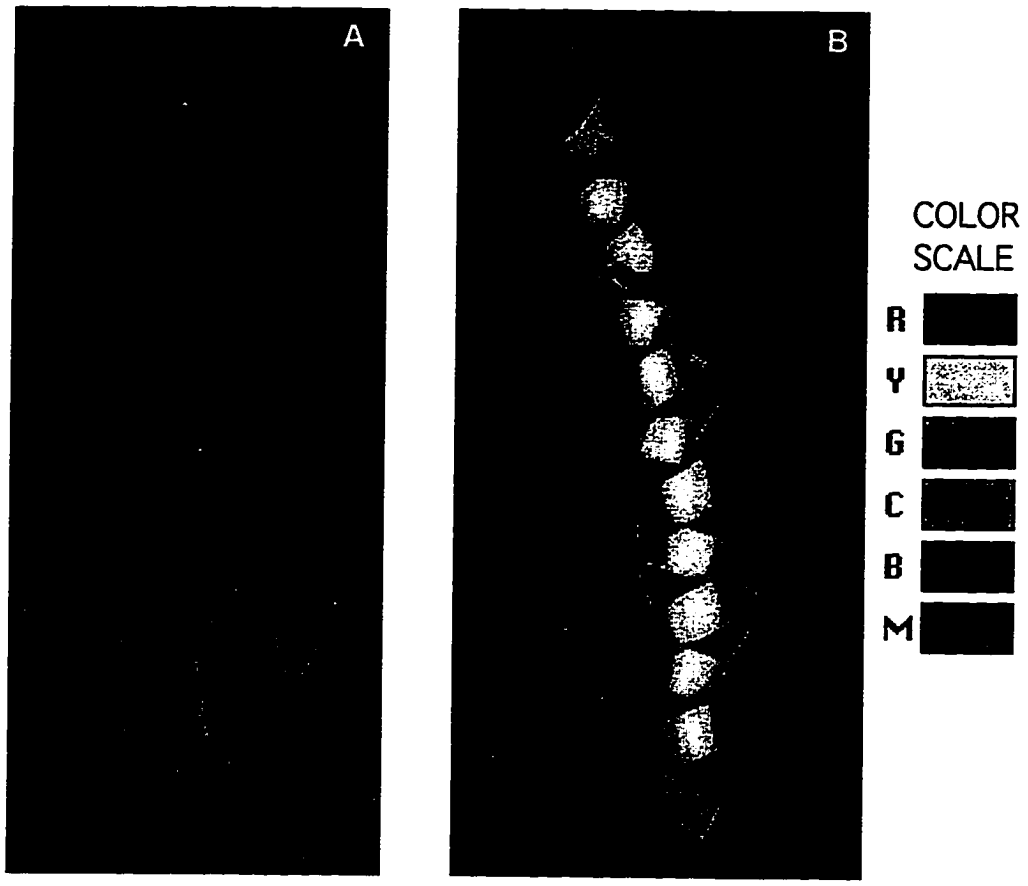


Figure 7.19

Bend in a Hx induced by SC X1=g- observed in PDB structures.
A) small bend = curvature (1PBE, 3SDH, 2HBG).
B) sharp bend found in the photosynthetic reaction center (1PSS).
Side chains shown pertain the Ser/Thr/Cys residues.

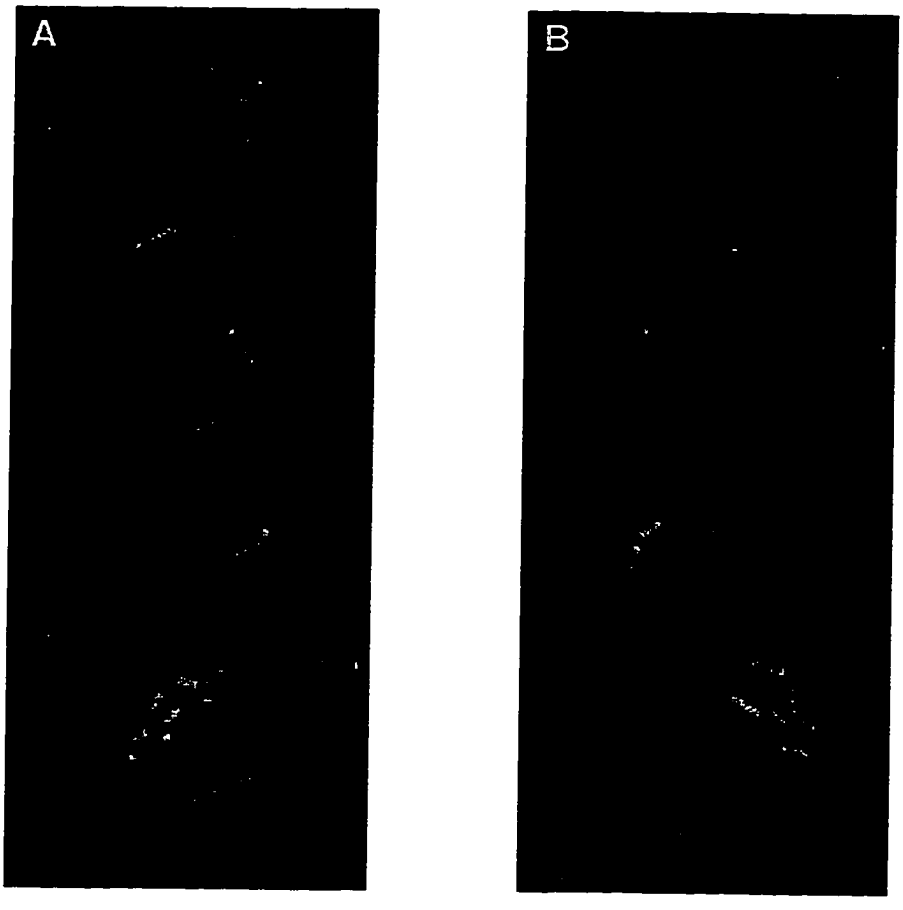


Figure 7.20

Superposition of some alternative H3 orientations induced by different X1=g+ or g- rotamers of Ser/Thr/Cys residues within H3. H3 was built originally with uniform phi/psi and X1=g+ as described earlier (See review, Section 2.1). In this conformation, all S/T/C H-bond back to the C=O i-4 of the preceding turn, inducing the helical orientation seen in red. The orientation of H3 can be modulated by S/C adopting different rotamers. C3.44 can reorient H3 adopting X1=g-, bending H3 towards H4 (green). The effect of C3.44 X1=g- is quite different if all other Ser residues (S336, S339, S347) adopt the X1=t rotamer where they do not interact with the helix backbone; the net effect is that only C344 X1=g- and T337 X1=t H-bond to the backbone inducing a bend in H3 towards H2 (yellow). This trend can be accentuated by S3.39 adopting X1=g- (blue), where all remaining Ser (S336, S347) can be modeled back to X1=g+ where they H-bond their C=O i-4; this is the selected structure for modeling the 5HT2C GPCR. Note that the resulting orientations of H3 all are identical at their C-termini, illustrated by the R350-D340 side chain in yellow on the figure. Thus these different S/C side chain conformations alter only the extracellular portion of H3, resulting in a different spatial orientation for D3.32 (see Figure). Interestingly, the different orientations observed for different S/C conformations are mostly due to a bending towards H2 or H4 (A), with little effect in the inward versus outward tilting of H3, as shown in (B) on a side view. Note on this view that T337 and C344, involved in the modulation of the H3 orientation, are facing the lipid ; their structural role proposed here which can be exerted from the lipid-exposed side of H3 would explain its high degree of conservation within neurotransmitter receptors.

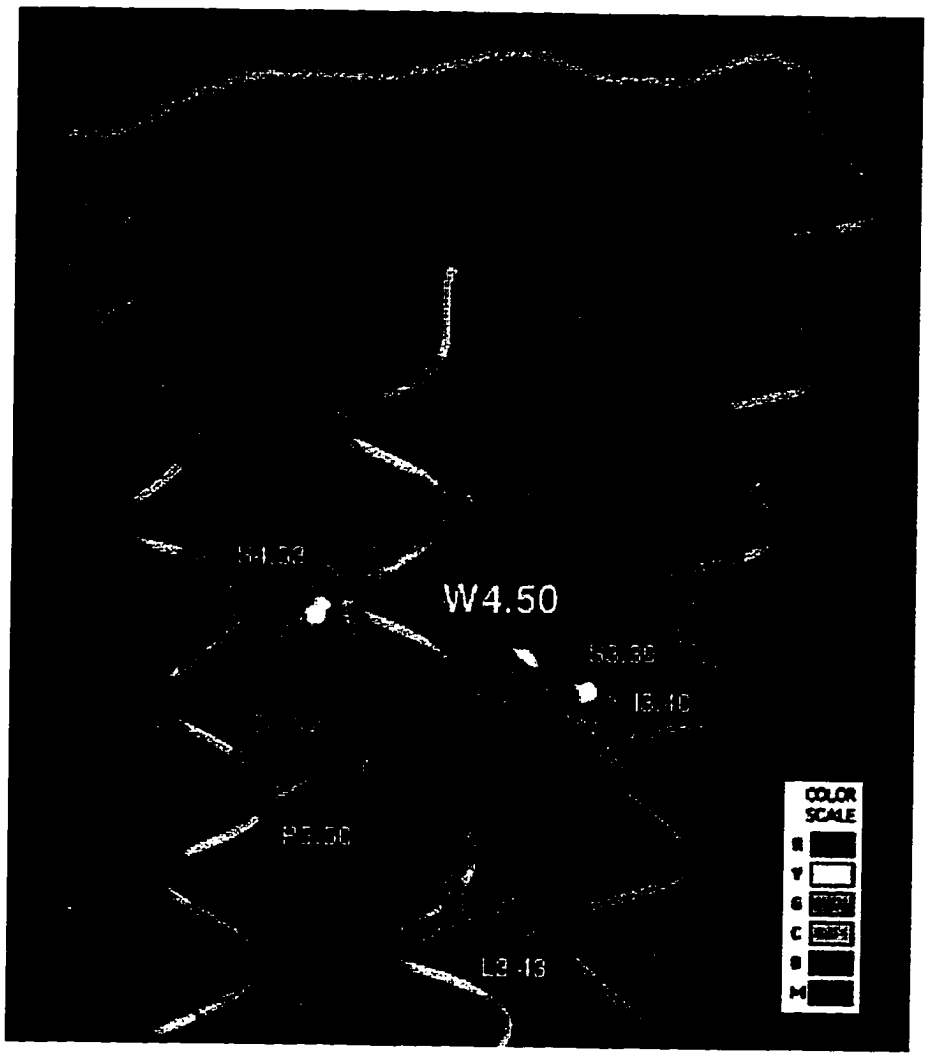


Figure 7.21:
Resulting packing of H4 into the H3-H5-H6 template. The 100% conserved W4.50, shown before to protrude from H4 (Figure 7.4.4.I), is inserted into the H3-H5 interface at the level of the 100% conserved residues S3.39-I3.40-L3.43-P5.50. The NH of W4.50 can H-bond S3.39, which adopts the X1=g- conformation where it can also H-bond the backbone C=O of S3.36. Modeling S3.39 (and C3.44) adopting the X1=g- has two effects on H3. Locally, an opening of the groove between S3.39-I3.40-L3.43 is induced, which facilitates docking of W4.50. This local distortion bends H3 towards H2 (see Figure 7.4.4.III). This H3 orientation is synergistically induced by the same requirement to accomodate the bulky W4.50, which can be fullfilled by tilting H3 towards H2 as shown here. The synergistic reorientation of H3 caused by bending and tilting results in an opening at the binding site between H3 and H5. The extent of this opening is limited by the disulfide bridge between C3.25 and C5.31 conecting the extracellular ends of H3 and H5 (blue strip). The effect of this opening is to create a cavity lined by H3 (right side), H5 (left side), the bulky W4.50 (bottom), and the H3-H5 loop segment at the top created by the disulfide bridge C3.25-C5.31. This cavity encompases several proposed binding residues (D3.32, S5.46, not shown), and has therefore a functional correlate: the "binding crevice", shown shaded.

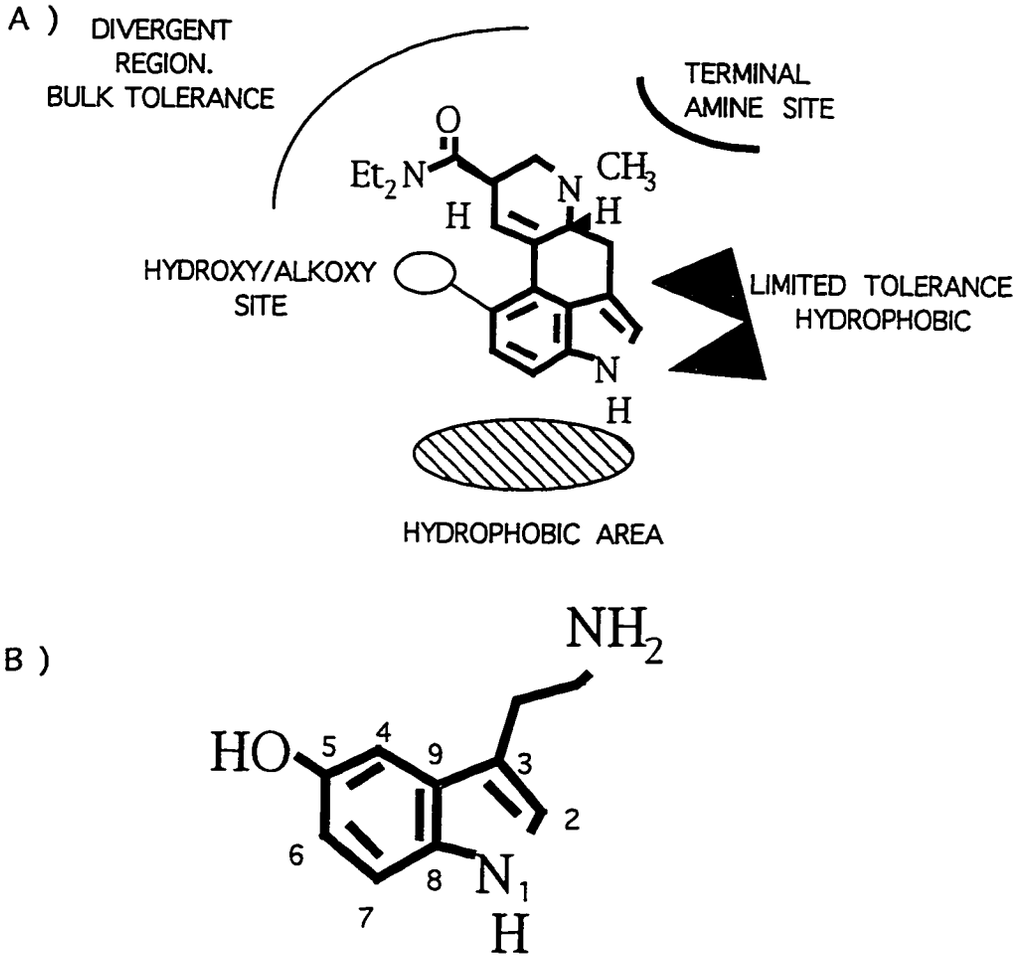


Figure 7.22

A) Proposed "pharmacophore" for the 5HT2CR, derived from reference (Glennon 1991), based on the LSD core.
B) Numbering scheme for the indole moiety of 5HT. The same numbering scheme was selected for the indole moiety of LSD for clarity, although the chemical conventions are different. A schematic representation of the proposed interactions of 5HT with the 5HT2CR according to the model is shown in Figure 7.24.

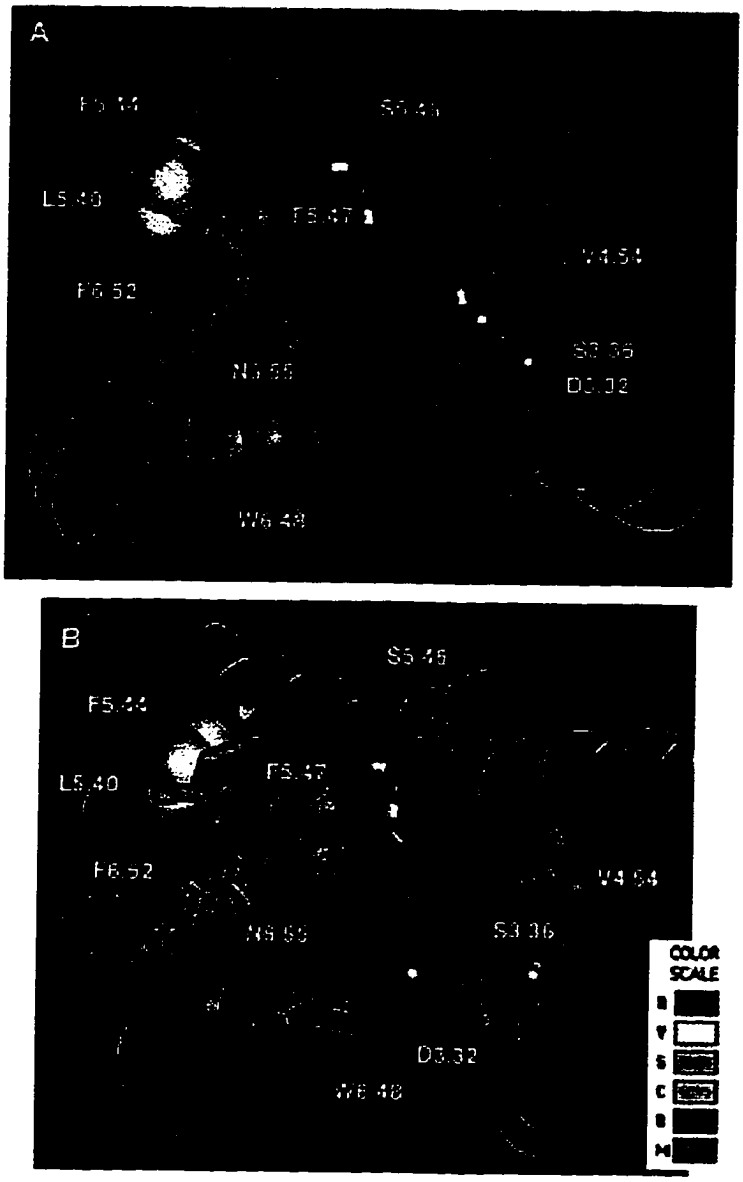


Figure 7.23:
The proposed ligand-receptor interactions at the agonist binding site defined by H3-H4-H5-H6 are shown for 5-HT (A) and LSD (B). These interactions were also shown schematically in Figure 7.4.5.I.
show D3.32-S3.36-S5.46-F6.52-W6.48, and F5.44-L5.40, and N6.55, and I4.54

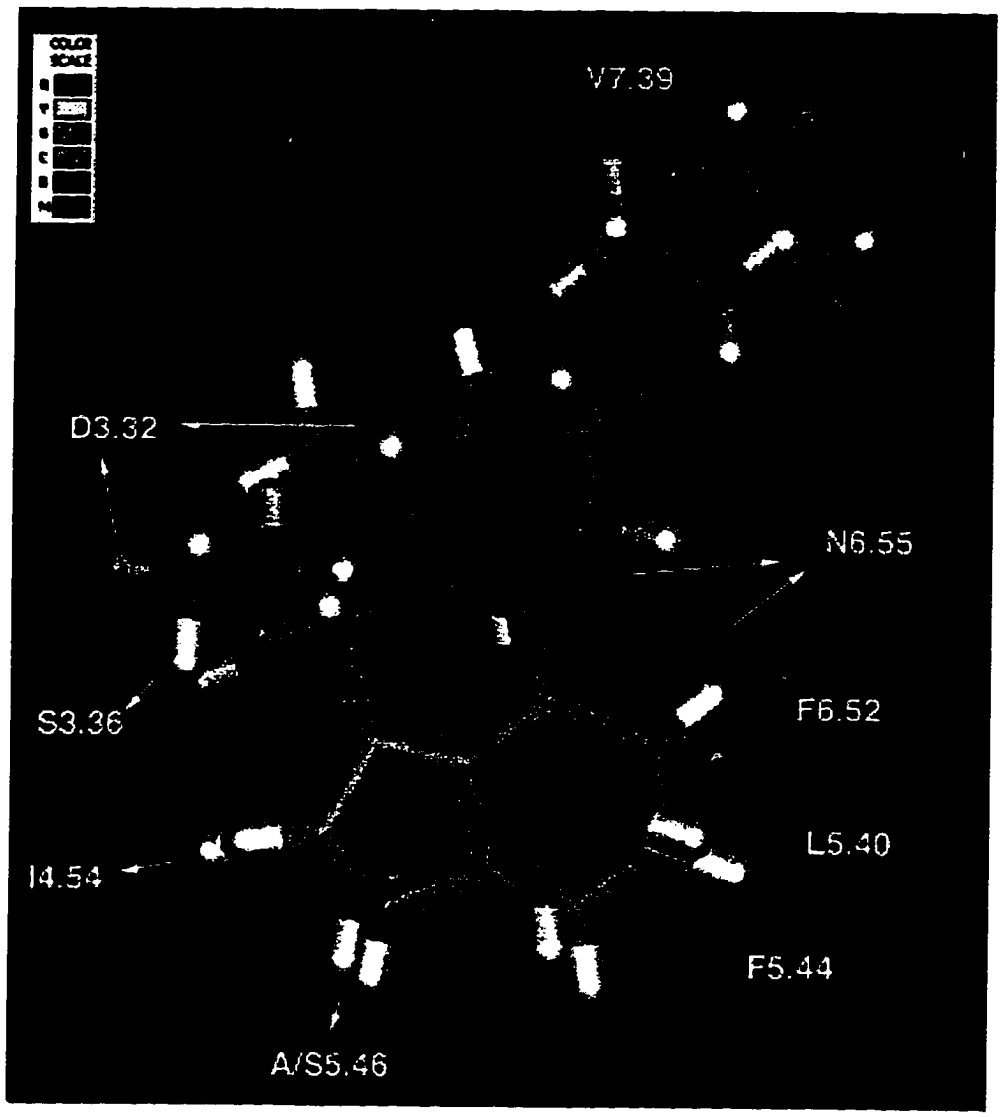


Figure 7.24:
Proposed spatial relationship between 5-HT and LSD within the binding site of the 5HT2CR modeled. Criteria for this superposition are described in the text.

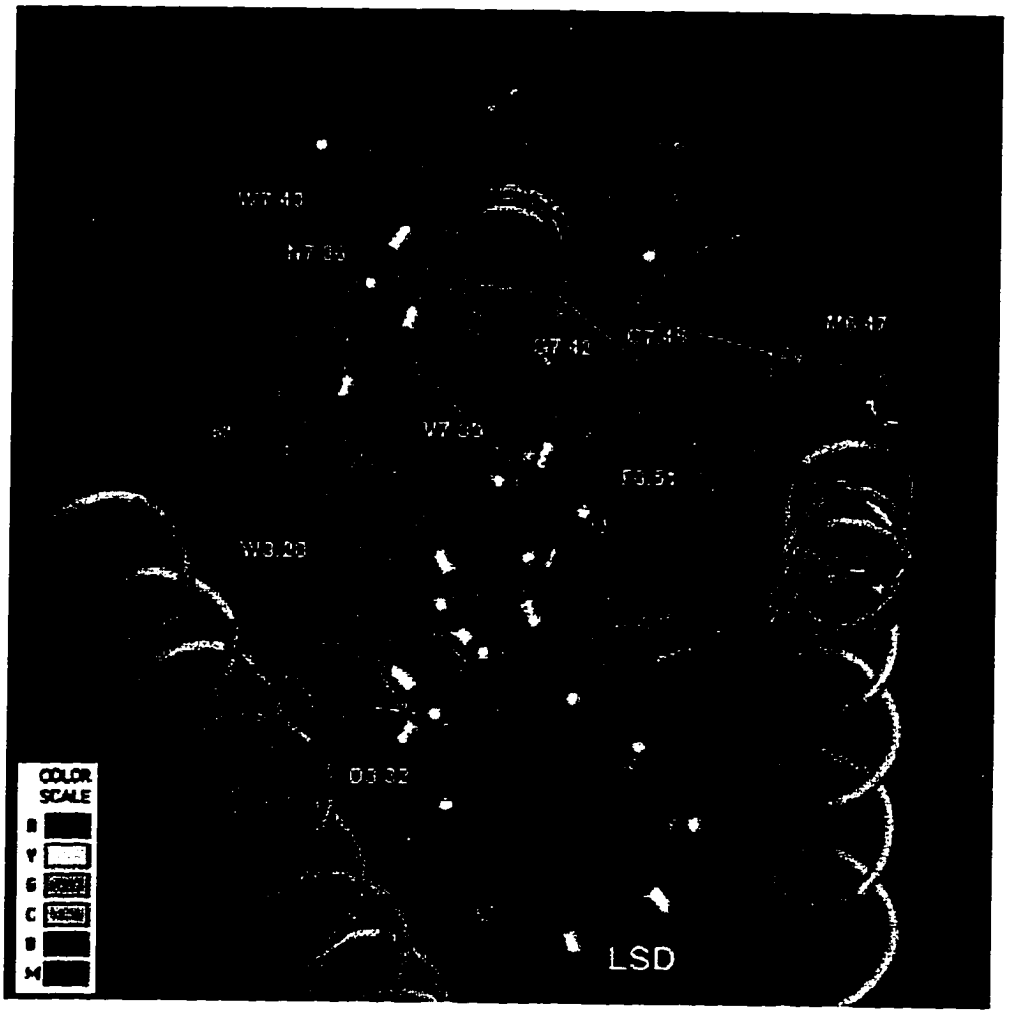


Figure 7.25:
Packing the extracellular segment of H7, comprising residues 7.35-7.46, into the H3-H4-H5-H6 template. Proposed interactions involving H7 residues that guided the packing are shown by dashed lines and explained in the text. Note the orientation of V7.39 and W7.40 towards the binding site, the later interacting with W3.28.

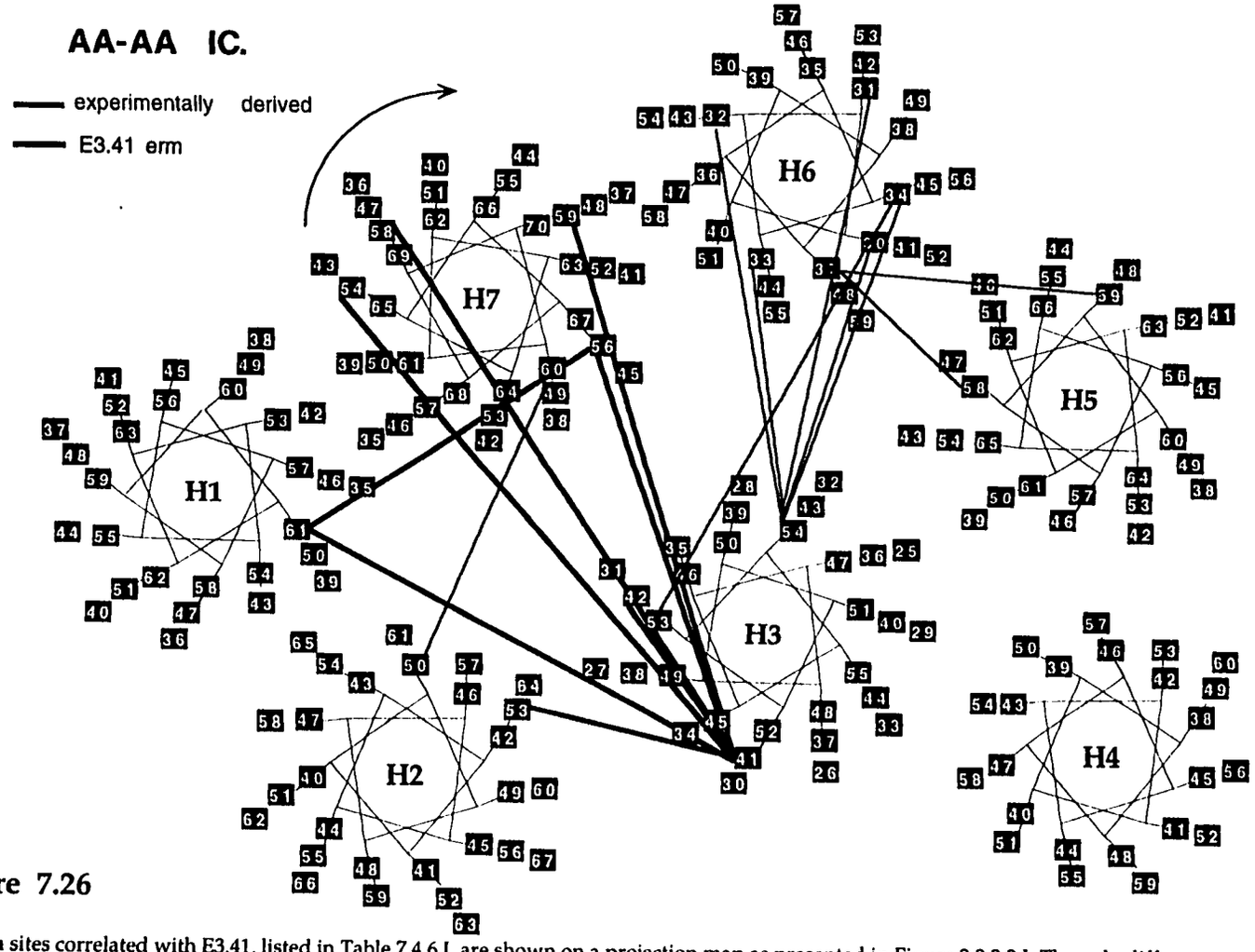


Figure 7.26

The erm sites correlated with E3.41, listed in Table 7.4.6.I, are shown on a projection map as presented in Figure 2.2.2.2.I. The only difference between these two projection maps is that H7 has been rotated according to the proposed face twist induced by the NP motif, discussed in Section 7.2.2. Therefore, this is a representations of the cytoplasmic region. Note the consistency with experimentally derived interactions (in red).

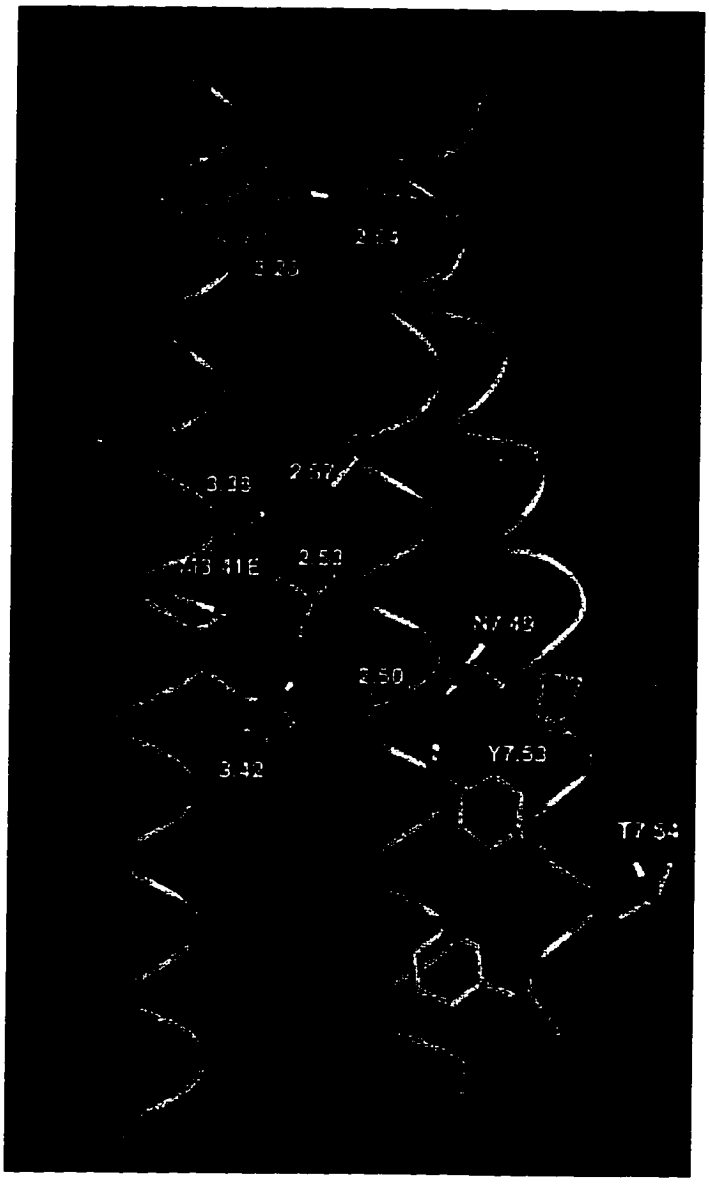


Figure 7.27:
H2 packing into the H3-H4-H5-H6-H7(7.35-7.46) template. Note the role of H3.42 H-bonding D2.50. Shaded lines indicates the **erm** sites correlated with E3.41.

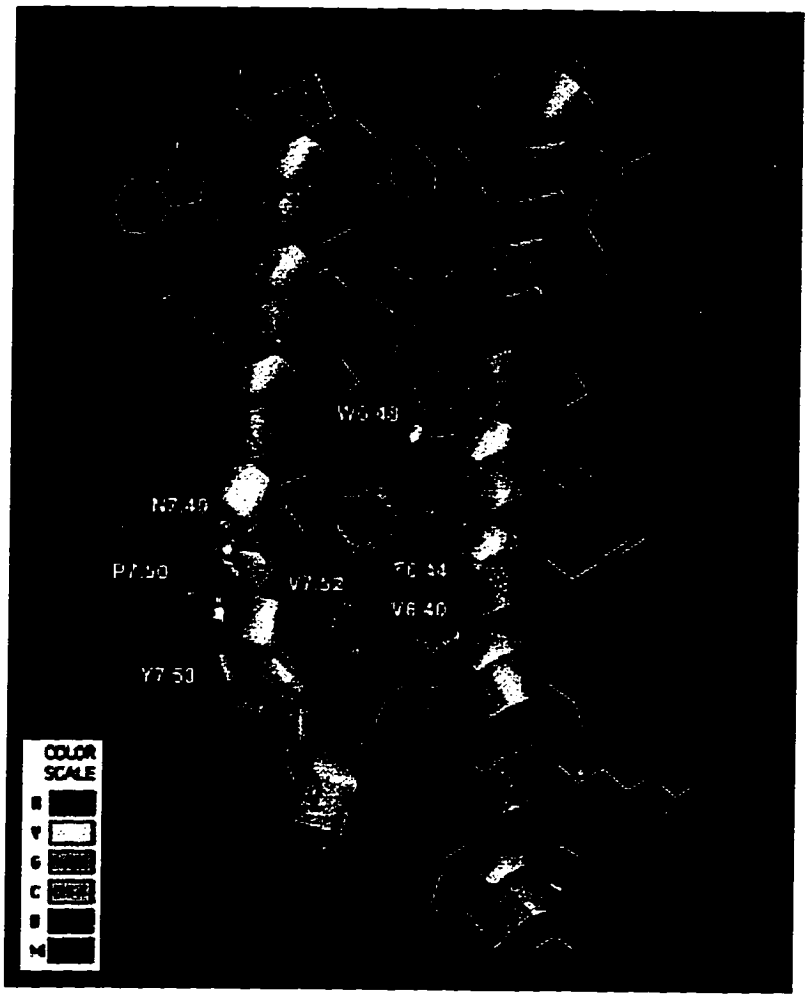


Figure 7.28:
Modeling the NPxVY motif in H7 into the H1-H2-H3-H4-H5-H6-H7(7.35-7.46) template.Note that the large bend present in H7 at the level of the NP motif has a shape that is complementary to the bulge created in H6 by V6.40-F6.44 .

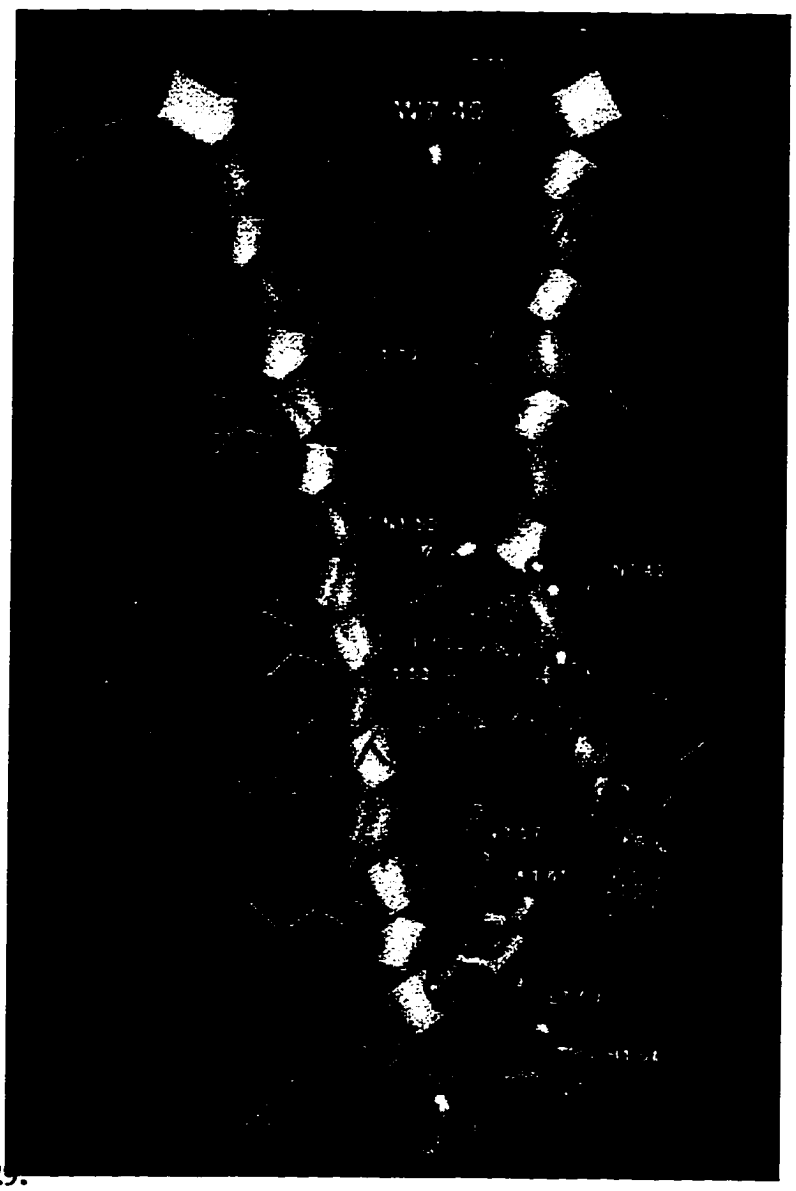


Figure 7.29.
Incorporation of H1 into the H2-H3-H4-H5-H6-H7 template. Note how the kink in the predicted accessibility pattern has a directionality in 3D that follows the path of the modeled kink at the PK of H7. There are three clusters of interacting residues described in the text and shown color coded. The aromatic side chains surrounding W7.40 are marked in red, residues N1.50-V1.53-V1.57 with thier interacing residues are shown in blue, and the cluster of polar residues at the cytoplasmic side is colored purplpe.

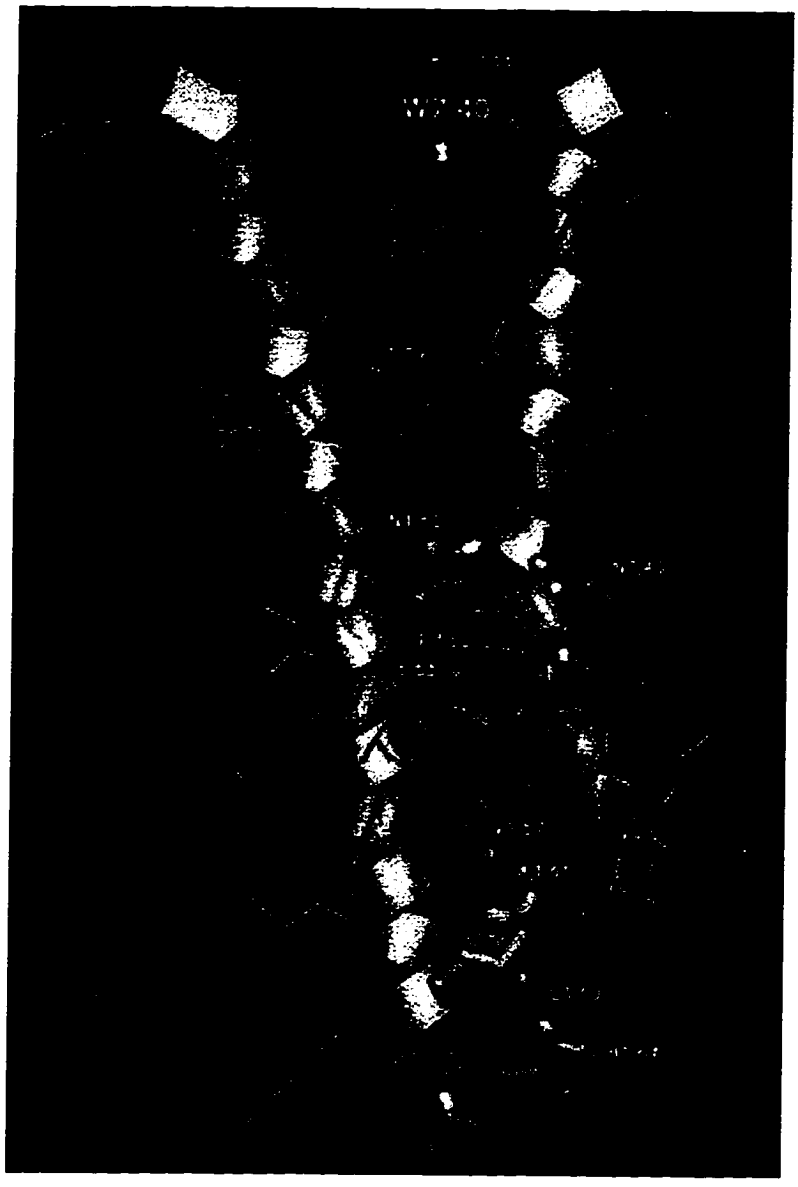


Figure 7.30:
The resulting packing of H1 with H2-H7 at the extracellular side is characterized by the presence of W7.40 which imposes an outward tilting on H1 and H2 at this level, as shown in

8.- EVALUATION OF THE 3D MODEL OF THE TMH DOMAIN OF THE 5HT2C RECEPTOR BY ENERGETIC, EXPERIMENTAL DATA AND FUNCTIONAL CRITERIA.

The result of the modeling process outlined above is an atomic three-dimensional model of the TMH domain of the 5-HT2C receptor at atomic resolution. Energetic, experimental data and functional criteria can be used for its evaluation. The proposed model of the serotonin-5HT2C receptor complex was subjected to Molecular Dynamics simulations (Section 8.1). The conformation of the modeled ligand-receptor complex was stable through 1 nanosecond at 300 K. The consistency of the energetically refined model with available data is illustrated by a summary of experimentally validated criteria used in the construction of the model (Section 8.2). In Section 8.3, the receptor model is evaluated by functional criteria in terms of ligand binding, the G-protein coupling domain, and inferences on the mechanism of receptor activation.

8.1- Evaluation of the proposed 5HT2CR model by energy-based criteria: Molecular dynamics simulations of the serotonin-5HT2C receptor model complex.

The model derived in Chapter 7 was subjected to Molecular Dynamics simulations (MD) following the procedural steps described in Section 2.4.

Some modifications of this model were necessary because the environment of the TMH domain was represented by a distance dependent dielectric without considering any explicit solvent molecules. First, the numerous Arg and Lys residues at the cytoplasmic boundaries facing the phospholipid headgroups were mutated to Ala to prevent artifacts by the absence of their charged counterparts. Second, due to the absence of water, charged groups facing the interior of the TMH bundle at the cytoplasmic boundaries (D3.49, R3.50, E6.30, R6.31, E1.60, K1.61) were neutralized, and the proposed specific interactions among them emulated by distance constraints. Third, D2.50 was kept charged but H-bonding N7.49 by a distance constraint. Fourth, the remaining charged groups in the simulation, D3.32 and the protonated amine of serotonin, were kept charged but their interaction was constrained together with other constraints on serotonin as described in Section 2.4.

After heating to 300 K and equilibrating the 5HT-receptor model through 200 picoseconds (ps) constraining *phi* and *psi* angles to ±10° of their initial values, these constraints were softened at the Pro-kink regions (150 ps). The constraints on the 5HT-receptor interactions not validated experimentally were then relieved and the resulting complexed receptor model was simulated for 900 picoseconds. Through this process, the model underwent minor alterations in its packing arrangement (rms=3-3.5 Å) and reached promptly (200 ps) a stable conformation maintained for the last 1 nanosecond of the simulation, as shown in Figure 8.1 by the root mean

square (rms) deviation of the C^{α}s from the initial coordinates. Therefore, the timeframe comprised between 200 and 1250 picoseconds was selected for analyses of structural features of the receptor (i.e. time average properties such as average H-bonding distances (<d>)) unless otherwise stated. To obtain a representative conformation, an average structure of the last 100 ps was subjected to energy minimization. The resulting conformation is the proposed three-dimensional model of the transmembrane domain of the 5HT2C receptor complexed with serotonin.

The changes in the orientation of the TMHs that result from the MD simulation can be appreciated from Figure 8.2 in a superposition of the initial model (yellow) and the MD model (red). The most significant changes occur in H1 and H4, the most lipid-exposed helices and thus with less specific interactions constraining their movements. Note that H3 and H6 at the extracellular side are brought together by the MD simulation. As will be discussed below in the description of the resulting binding site, this is due to the electrostatic predominance of the charged Asp3.32 attracting Asn6.55, an effect probably exaggerated in these simulations by the absence of explicit water molecules. The consistency of the derived model with the predicted pattern of TMH-TMH versus TMH-Lipid interfaces can be estimated by calculating the solvent accessibility per residue, which in the absence of the lipids (and water) would be equivalent to the predicted TMH-Lipid interface. The solvent accessibility per residue is calculated as a fraction of the average

accessibility found for each amino acid in known protein structures. This comparison is presented in Figure 8.3 on a helical net, where the calculated degree of solvent accessibility per residue is indicated by the degree of shading, and the previously predicted TMH-Lipid interfaces are shown by thicker circles. Note the consistency between the predicted and observed accessibilities inside TMHs, and the discrepancy at some TMH boundaries, a natural consequence of the absence of loop structures in the model. Helix 3 is more exposed than originally predicted, a consequence of the structural role played by the conserved T3.37 and C3.44 from a lipid-facing orientation, as observed in Figure 8.3 and discussed in detail in Section 7.4.4. Residues K1.61 and R6.31, originally predicted to face the phospholipid headgroups based on the Arg/Lys motif, are buried inside the TMH bundle participating in specific interaction as described in Chapter 7 and reviewed below.

The binding site for serotonin was characterized by a set of H-bonding interactions in the construction of the model. A more extensive network of ligand-receptor H-bonding interactions results from the MD simulations, supporting a well defined geometry that satisfies all the H-bonding potential of serotonin. Note that the simulations were initiated with distance constraints imposed on the interaction of 5HT with D3.32, S3.36, A5.46, and N6.55 (see methods Section 2.4). After the first 350 ps, the constraints for the interaction with S3.36 and N6.55, not validated experimentally in the 5HT2CR, were freed. This produced a disruption of the interaction with N6.55

which preferred to H-bond the charged Asp3.32. The set of H-bonding interactions that result from the simulation is shown in Figure 8.4. Note that the protonated amine fulfills its H-bonding potential by interacting simultaneously with D3.32 (<d>=2.5±0.5), S3.36 (<d>=2.8±0.2), and S4.57 (<d>=2.8±0.3). The interaction with S4.57 was not included in the original model and appeared in the first 200 ps after which it remained stable. This is shown in Figure 8.5 where the distances between the H-bonding donors and acceptors are plotted throughout the simulation trajectory. Note that the distances plotted for most interactions overlap and become barely distinguishable in the 2.5-3.5 A range that corresponds to H-bonding distances. The important information from this plot is when a given distance is not within this range, meaning that the H-bond was lost. The interaction of 5HT with D3.32 is not restricted to the protonated amine, but the 5-OH is also H-bonding the other carboxylic oxygen of D3.32 (<d>=2.5±0.1). This second carboxylic group of Asp3.32 is further stabilized by an H-bond with N6.55 (<d>=3.1±0.2), which also H-bonds to the H3 backbone (see Figure 8.4). The proposed H-bond between the 5-OH and N6.55 is lost immediately after the corresponding constraint is relieved at 350 ps (see Figure 8.5). However, the average distance between 5-OH and the NH_2 of N6.55 is only 3.7±0.5 and thus remains quite close to an H-bonding distance. Taken together, this set of extensive H-bonding interactions probably reflects the energetic predominance of the electrostatic effects induced by the presence of two

charged species in a simulation without explicit waters, which would be likely to mediate some of this H-bonding interactions and to screen the charge of these groups. This would explain the observed proximity of the extracellular sides of H3 and H6 described above (see Figure 8.1). Nonetheless, this simulation indicates the feasibility of the H-bonding contacts observed, especially 5-OH with D3.32 and S4.57 which were not considered before, although their energetic relevance is likely to be exaggerated by the inappropriate representation of the environment. Note from Figure 8.4 that the proposed H-bond between 5-OH and N6.55 could be easily accommodated. This would allow 5-OH to act as an acceptor and not only as a donor as the energy refined model proposes, and thus comply with the observed pharmacological profile at this serotonergic moiety (see Section 7.4.5). Thus, my interpretation of the results from this simulation maintains the proposal of the H-bonding interaction between 5HT and N6.55 described in Section 7.4.5. The set of observed H-bonding contacts between serotonin and the receptor is completed by an H-bond of the indole NH with the backbone carbonyl of S5.43, whose significance will be discussed in the next section. The proposed interaction between 5HT and I4.54 was not maintained in the simulations; it was lost concomitantly with the change in the orientation of H4 that accompanied formation of the H-bond between the protonated amine of 5HT and S4.57. It follows that these two interactions are not compatible with each other, and the model can not be used to discriminate between them

due to the absence of explicit water in these calculations which puts into question the S4.57 interaction.

In addition to the polar serotonin-receptor contacts described above, the energy refined model maintains the principal structure-functional motifs or microdomains defined throughout its construction This is shown in Figure 8.6 where the relative spatial distribution of these microdomains in the context of the seven TMH bundle can be appreciated. Because these motifs involve the most conserved residues, they are expected to define the structural and functional integrity of the receptor. Thus, modeling studies propose that the organizing principles of neurotransmitter GPCRs comprise a polar ligand binding site at the extracellular side, followed underneath by a cluster of aromatic residues that span the core of the TMH domain until they reach the NPxVY motif in H7 (the H2-H7 microdomain), and two sets of significantly charged clusters of polar residues at the cytoplasmic boundaries: the Arg-cage and the H1-H2-H7 cytoplasmic motifs (see Figure 8.6). The most significant interactions within these microdomains are described below, and the connections among them by specific interactions are discussed in Section 8.4 in the context of the receptor activation mechanism.

The interaction between 5HT and polar residues at the extracellular side of H3-H4-H5-H6 described above is complemented by direct aromatic-aromatic interactions with W4.50, F5.47 and F6.52, whose side chain rotamer conformations introduced in the model (Section 7.4) were not changed by the

MD simulations. The proposed interaction of 5HT with F5.44 (Section 7.4.5) is lost, due to the tilting of the extracellular sides of H6 and H3 towards each other, driven by the D3.32-N6.55 interaction described above. These aromatic residues belong to the aromatic cluster motif which spans the core of the TMH domain involving residues {F3.35, W4.50, F5.44, F5.47, F5.48, F6.41, F6.44, W6.48, F6.51, F6.52}. The dense network of interactions among these residues could explain why significant effects on binding affinity where observed following mutations at several of these residues (van Rhee 1996), in spite of the fact that only 3 residues participate in direct aromatic-aromatic interactions with the ligand. The conserved F6.44 of the aromatic cluster maintains the interaction proposed in Section 7.4.8 with the NPxVY motif (see Section 7.4.8), which forms part of the H2-H7 microdomain defined by the cluster of polar residues {D2.50, N7,49, P7.50, Y7.53, N1.50}. Well defined and stable H-bonding interactions were observed between the charged Asp2.50 and N7.49 (<d>=2.9±0.1), Y7.53 (<d>=2.5±0.1), and H3.42 (<d>=2.9±0.4). Note that N1.50 is probably too far removed from D2.50 (5 Å). This motif was proposed to interact with the conserved Arg3.50 in the active state of the receptor (and thus not in this model, see Section 7.3.2), belonging to the Arg-cage motif shown in yellow in Figure 8.6. In the Arg-cage manuscript presented in Section 7.3.2 we proposed that R3.50 is held in the inactive conformation by ionic bond interactions with D3.49 ((<d>=3.0±0.1) and E6.30 (<d>=3.1±0.1). Further stable H-bonding interactions within this

microdomain are E6.30-R6.31 (<d>=3.0±0.1), Y5.58-S6.34 (<d>=3.3±0.5), R3.50-S6.34 (<d>=3.1±0.2), and D3.49-S2.45 (<d>=3.4±0.4). Note that S6.34 and S2.45 play important roles not present in the initial model, and that the proposed D3.49-Y5.58 interaction (Section 7.4.2) is lost (<d>=6.1±1.2) in favor of the Y5.58-S6.34 H-bond. This significantly charged microdomain, the Arg-cage, is adjacent in space to another significantly charged cluster of polar residues at the cytoplasmic boundaries of H1-H2-H7 (see Figure 8.6). However, there are no significant interaction between these two microdomains due to the presence of aromatic residues from H2 (F2.42) and H7 (F7.56) interspersed between them. This last microdomain, formed by {E1.60, K1.61, H1.64, N2.40, Y2.41, N7.57} and described in Section 7.4.9, becomes stable for the last 250 ps of the simulation as indicated by the presence of the following well defined H-bonding interactions: E1.60-K1.61 (<d>=3.1±0.1), N7.57-N2.40 (<d>=2.9±0.2), H1.64-Y2.41 (<d>=2.8±0.2), and E1.60-N7.57 (<d>=3.5±0.6).

Thus, the energy-refined model that results from the MD simulations maintains the essential structural features of the models and supports the functional inferences. The consistency of the energy-refined model with available experimental data is presented below.

8.2.- Summary of experimentally validated criteria used in the construction of the model.

The consistency of the energy-refined model with experimentally

observed residue accessibilities can be analyzed by comparing Figures 8.3 and 5.10 were the respective data are shown on helical nets. The few residues where there is a disagreement parallel those discussed in Section 5.8 in the comparison between the predicted and observed accessibility pattern, except for the TMH ends whose solvent accessibility in the energy-refined model are exaggerated due to the absence of any loop structures.

The experimentally derived interhelical interactions were listed in Table 2.3. Considered for this analysis are the interactions involving Cys crosslinking (Yu 1995; Farrens 1996; Kono 1996) (Cys-Cys), His-engineered Zn binding sites (His-Zn-His) (Elling 1995; Elling 1996; Sheikh 1996), and double revertant mutations (Rao 1994; Zhou 1994; Liu 1995; Sealfon 1995; Han 1996). The consistency between the model and these data is interpreted in terms of distances observed between residues proposed to be in spatial proximity, taking the C^{β} carbon as reference since different amino acid side chains need to be compared. The proposed interactions are classified according to the experimental technique applied, since each one has different structural and distance requirements. The C^{β}-C^{β} distances obtained from the MD simulations for the proposed interactions are shown in Table 8.1 as means and standard deviations.

Table 8.1 :

Average distances (mean) and standard deviation (σ) observed in the Molecular Dynamics simulation between the Cβ atoms of residues proposed to be in spatial proximity based on Cys-crosslinking experiments (Cys-Cys), His-engineered Zn binding sites (His-Zn-His), and double revertant mutant phenotypes (2rm). The minimum and maximum distances are also shown for Cys-Cys data, indicative of the possible collision leading to covalent bond formation. Shaded interactions are considered in agreement with experimental data. *the D2.50-N7.49 was a constraint in the simulation.

Cys-Cys					His-Zn-His			2rm		
AA-AA	Mean	σ	Min	Max	AA-AA	Mean	σ	AA-AA	Mean	σ
3.54-6.30	9.9	0.5	8.8	11.5	**2.64-3.28**	9.6	0.3	1.39-7.36	11.5	0.7
3.54-6.31	14.0	0.5	12.4	15.6	5.39-3.29	13.5	1.0	3.36-6.44	13.9	1.0
3.54-6.32	16.9	0.4	15.0	18.3	**3.53-6.34**	8.2	0.4	2.50-7.49*	2.9	0.1
3.54-6.33	13.3	0.5	11.6	15.0	**5.39-6.59**	9.2	0.9			
3.54-6.34	12.2	0.5	10.5	13.7						
5.39-6.55	12.3	0.7	10.3	14.6						
5.39-6.56	7.4	0.7	5.4	9.8						
5.39-6.59	9.2	0.9	6.6	12.0						
5.40-6.55	11.2	0.6	9.1	12.8						
5.40-6.56	6.3	0.7	4.5	9.5						
5.40-6.59	10.4	0.7	6.8	13.8						
5.43-6.59	9.6	0.5	7.9	115						
6.37-5.58	9.0	0.4	7.38	10.3						
6.37-5.59	12.8	0.5	11.4	15.1						

Cys crosslinking (Cys-Cys): Analysis of disulfide patterns in known protein structures (PDB) indicates that the distances between the C^{β} atoms of two Cys residues bound by a disulfide bridge can vary between 3 and 5 Å. Inspection of Table 8.1 shows that none of the interresidue distances satisfy

this criterion, 5.40-6.56 (6.3 Å) and 5.39-6.56 (7.3 Å) being the closest average distances. It could be argued that formation of the Cys-Cys covalent bond is dependent on the probability of a collision between the respective sulfur groups more than on the average distance. Thus, the minimum and maximum distances observed between these residues is also shown in Table 8.1, where the distances between 5.40-6.59 (4.5 Å) and 5.39-6.56 (5.4 Å) would satisfy this criterion. Nonetheless, the agreement with the experimental data judged by these distance criteria is not satisfactory. The very low standard deviation observed for these distances, between 0.4 and 0.9, indicates that the dynamical fluctuations in these distances is surprisingly small. This could be due to strong electrostatic interactions present between these domains, and these are overemphasized in the simulations *in vacuum* ; the Arg-cage motif (R3.50-E6.30) would rigidify the H3-H6 cytoplasmic interface at the level of the Cys-Cys observed (Farrens 1996), and the ionic and H-bonding interactions at the ligand binding site would rigidify the H5-H6 extracellular interface at the level of the Cys-Cys observed (Yu 1995; Kono 1996). It should be noted that the Cys-Cys at the H3-H6 interface are inconsistent with an α-helical conformation between 6.30 and 6.34, as discussed in Section 7.4.1, and that the extracellular H5-H6 interface was predicted to be structurally different between rhodopsin and the 5HT2CR based on the analysis of the the Pro-kinks in H5 and H6 (see Section 7.2). It could also be that the distance criterion reported above is too stringent, so that average distances within 10 Å (3.54-

6.30, 5.39-6.56, 5.39-6.59, 5.40-6.56, 5.40-6.59, 5.43-6.59, 6.37-5.58), shaded in Table 8.1, can be proposed to be consistent with the Cys-Cys studies in rhodopsin (Yu 1995; Farrens 1996; Kono 1996). Thus, only one Cys-Cys is selected for the H3-H6 cytoplasmic interface, while all extracellular H5-H6 sites are selected except for those involving N6.55 (probably due to the strong interaction with D3.32;, one of the two cytoplasmic H5-H6 sites is also selected.

Engineered His-Zn-His binding sites: The geometric criterion for Zn binding to two His residues has been estimated to be of a 13 Å distance between the respective C^{α} atoms (Elling 1995; Sheikh 1996). The observed C^{β} distances below 10 Å for the proposed interactions between 2.64-3.28, 3.53-6.34, and 5.39-6.59 are thus consistent with the experimental results (Elling 1995; Elling 1996; Sheikh 1996). The proposed Zn binding site between 5.39-3.29 (Elling 1996) would be inconsistent with the model, a discrepancy discussed in Section 7.4.5 as likely reflecting the relation between the unliganded and liganded forms of the receptor, an explanation that would be consistent with the competitive antagonist profile of this Zn site (Elling 1996).

Double revertant mutants : The 11.5 Å distance between residues 1.39 and 7.36 is inconsistent with a steric clash between T1.39-T7.36 being the reason for the observed structural perturbation in muscarinic receptors (Liu 1995). Instead, a proposed indirect structural incompatibility between these two residues involving W7.40 was proposed in Section 7.4.9. The revertant character of the G3.36L-F6.44A double mutation in rhodopsin (Han 1996)

involves large side chains which may be related to the large distance (14 Å) between the corresponding C^{β} groups in the 5HT2CR model. The large distance between 3.36 and 6.44 is a direct consequence of the axial displacement between H3 and H6 suggested by the Cys-Cys (Farrens 1996) and His-Zn-His data (Sheikh 1996) for the interface between these two helices; note that a close 3.36-6.44 distance would be inconsistent with a close distance between 3.53-6.34 or 3.54-6.30, i.e. these two sets of interactions are separated by 17-18 residues in H3 and 10-14 residues in H6. Therefore, this revertant mutation does not invalidate the positioning of H3 and H6 in the proposed model. The reciprocal mutations between D2.50 and N7.49 (Zhou 1994; Sealfon 1995) cannot be used as validation because it was included as a constraint in the simulation. Two other revertant mutants in the literature were not included in Table 8.1 because the proposed interactions involve a single residue with entire TMH segments. Thus, the deleterious effects of an Asn or Phe at position 7.39 in adrenergic receptors could be reverted by chimeras involving H1 and H2, but not H2 alone, implying that 7.39 interacts with H1 (Suryanarayana 1992; Mizobe 1996). However, the same authors proposed that this same Asn or Phe difference at 7.39 was involved in direct ligand-receptor interactions (Suryanarayana 1993), and hypothesis supported by mutagenesis studies in other GPCRs (Suryanarayana 1991b; Suryanarayana 1991a; Guan 1992; Metcalf 1992; Oksenberg 1992; Adham 1994; Glennon 1996). Because W7.40 should also face the binding site (Wong 1988), I consider it

impossible for position 7.39 to be facing the binding site and H1 simultaneously (see Figure 7.4 for an illustration of this incompatibility). Thus, I conclude that the observed revertant phenotype results from an indirect, rather than a direct interaction. This same group found another revertant mutant between the same β2 and α2 adrenergic receptors, involving a Phe or Leu at position 7.38, which was proposed to face H3 and/or H6 based on the revertant phenotype (Mizobe 1996). A Phe is present also in the 5HT2CR at position 7.38, and in the model it is packed against H6 at the level of the conserved P6.50, thus consistent with these experiments (Mizobe 1996).

In conclusion, the energy-refined model that results from the MD simulations incorporates essential structural features and experimentally derived interactions that were used as a criteria for its construction, and is consistent with a large set of functional inferences.

8.3- Functional evaluation of the proposed 5HT2CR model: Ligand binding. G-protein coupling domain, and the activation mechanism.

The energy-refined model is evaluated by functional criteria at three different levels: Ligand binding (Section 8.3.1), the G-protein coupling domain (Section 8.3.2), and the receptor activation mechanism.

8.3.1.- Ligand binding.

In a collaborative effort with the laboratory of Dr. Stuart Sealfon, modeling approaches were combined with mutagenesis techniques to study the proposed interaction of alkylindolines and ergolines with the Ala or Ser residue at position 5.46 in TMH5 on the 5HT2C and 5HT2A receptors. These studies are already published (Almaula 1996a); the results offer a molecular basis for the observed selectivity of these ligands between the human 5HT2C and 2A receptors, based on the ability of the substitutent of the indole nitrogen to interact with the 5.46 locus. Thus, alkylindolines and ergolines containing a free NH interacted favorably with a Ser at position 5.46, while ligands of the same structural class containing a more bulky methylene substituent would clash with a Ser residue at this locus, but interact more favorably with an Ala residue as present in the 5HT2CR at this locus. This structural scheme is shown in Figure 8.7 in a comparison between LSD and Mesulergine, taken directly from the article (Almaula 1996a). Note that in this article mutagenesis was used to convert a locus of the 5HT2CR into the 5HT2AR, and the same locus on the 5HT2AR was converted into the 5HT2CR. These reciprocal interchanges, whose logic is parallel to the interpretation of double revertant phenotypes, were meant to provide a discrimination between a direct and an indirect ligand-receptor interaction.

Based on the results of this article we could rationalize the observed changes in affinities based on the proposed steric clash or H-bond with a Ser

residue in position 5.46. However, we did not understand which alternative contacts were available for the indole NH of 5HT, LSD and others in the human 5HT2C, having an Ala residue at this locus. The observation of the interchangeable role of this residue between these two receptors in terms of the measured ligand affinities indicated that the positioning of the indole NH in both the 2C and the 2A subtypes was likely to be similar. The MD simulations presented here suggest that in the 5HT2CR the indole NH is H-bonding the carbonyl of 5.43 (see Section 8.1). This interaction is consistent with a similar spatial orientation of the NH in the two receptors, because as described in the construction of the model, S5.46 was modeled adopting the X1=g- conformation instead of the more common X1=g+ rotamer. In the g- configuration, the S5.46 OH is H-bonding the carbonyl of S5.43. Thus, the OH of S5.46, modeled in very close spatial proximity to the carbonyl of S5.43, could explain the similarities observed in the positioning of the indole NH between the 5HT2CR and the 5HT2AR.

8.3.2.- The G-protein coupling domain.

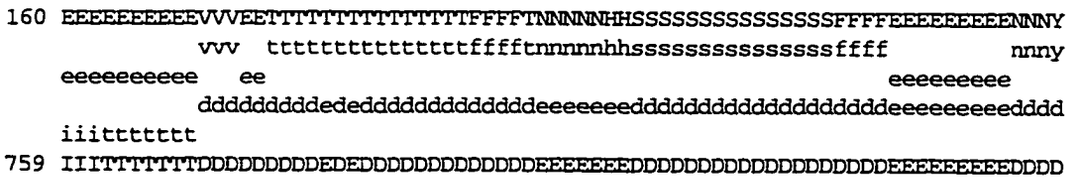
One of the novelties of the proposed model relative to previously reported modeling studies is the incorporation of most of the G-protein coupling domains proposed in the literature (van Rhee 1996) as intrinsic components of the TMH domain. This is due to the redefinition of the TMHs that results from the application of the Arg/Lys motif at the cytoplasmic

boundaries discovered in this thesis (see Section 2.1). According to the organizing principles that arise from the model based on specific microdomains, the identified G-protein coupling domains present in the model are localized within the two clusters of polar residues present at the cytoplasmic boundaries: the Arg-cage motif and the H1-H2-H7 cytoplasmic motif (see Figure 8.6 for definition). The cytoplasmic localization and accessibility of these two motifs is shown in Figure 8.8 by a Van der Waals representation of the model viewed from the cytoplasmic side. Note that both motifs are highly exposed to the aqueous cytoplasm, which is inconsistent with two criteria: First, the ionic bonds present within these motifs have been proposed to hold the receptor in the inactive conformation through specific interhelical interactions (e.g., see Arg-Cage manuscript in Section 7.3.2). This proposed function could not accomplished if the groups were exposed to water, which would solvate the charges and weaken the interhelical interactions. Second, because this model represents the inactive state of the receptor, the coupling domains are not expected to be accessible for G-protein binding. Thus, some part of the cytoplasmic segments of the receptor not considered in this model are expected to shield these G-protein coupling domains from the aqueous cytoplasm.

The best candidate to fulfill this role is the predicted cytoplasmic extension of H7, comprising residues 7.59-7.70. This segment shares with the Arg cage and the H1-H2-H7 cytoplasmic motifs the same high degree of

conservation. Furthermore, it has been shown to activate G-proteins by itself (Jung 1996), and thus should also be inaccessible to the G-protein in the inactivated state modeled. This inaccessibility could be achieved best by the segment shielding itself, the Arg-cage and the H1-H2-H7 cytoplasmic motifs simultaneously through direct interactions among them. Because the Arg-cage and the H1-H2-H7 motif can not interact directly with each other (see Section 7.4.9), the cytoplasmic extension of H7 should be responsible for shielding both motifs from water exposure. The feasibility of this hypothesis is illustrated in Figure 8.9, where the cytoplasmic extension of H7 is shown docked on the proposed model. Note that this segment, modeled as a standard α-helix as supported by NMR experiments (Jung 1996), is capable of extending from the H1-H2-H7 cytoplasmic motif, across the Arg-cage motif, towards the H5-H6 boundaries where its C-terminus protrudes out from the TMH domain into the membrane-cytosol interface, i.e. the phospholipids headgroups. As required, this packing arrangement shields the G-protein coupling domains present in the model from solvent exposure, as can be seen from Figure 8.9 were the cytoplasmic extension of H7 is colored according to the accessibility predictions developed before (Section 6.1).

Docking of the H7 cytoplasmic extension into the TMH domain as shown above was performed according to the following criteria: Spin labeling studies proposed a proximity between 1.60 and 7.63 (Yang 1996a) (< 10 Å, see Table 2.3). This proximity was consistent with the following **erm:**



Thus, the conserved acidic group at 7.59 which is present in most neurotransmitter GPCRs but is missing in the 5HT2 subclass, is present at position 1.60. The path followed for packing $H7^{(7.59\text{-}7.70)}$ with the cytoplasmic face of the TMH domain as it interacts with residues at the extreme of H3-H6 first, and H5-H6 afterwards, was guided by the **erms:**

The proposed packing for the segment $H7^{(7.59\text{-}7.70)}$ positions the Palmitoylation site C7.70, at the C-terminus of this segment, beyond the

H5-H6 boundary thus exposed to the lipid headgroups, consistent with the requirement of a lipid-exposed palmitoyl chain (Moench 1994) (see Figure 8.9). Thus, incorporation of the cytoplasmic portion of H7 can satisfy the structural and functional requirements relative to the G-protein coupling domains present in the 5HT2CR model.

8.3.3.- The receptor activation mechanism: Inferences from the model.

A putative mechanism of ligand-induced activation of the receptor arises from consideration of the receptor and receptor-ligand conformational properties discovered through the modeling process. The details of this proposed mechanism are consistent with the conformational changes involved in receptor activation that were suggested by experimental studies (see Section 2.2). Although many serotonergic ligands would share a similar binding site formed by H3-H4-H5-H6, different ligands would utilize different interaction sites to optimize favorable ligand-receptor interactions. Most ligands would utilize the H-bonding sites available on H3-H4-H5-H6 (D3.32, S3.36, S4.57, S5.43, S5.46, N6.55), as well as the cluster of aromatic residues formed by H4-H5-H6 (W4.50, F5.47, W6.48, F6.51, F6.52). Modeling ligand-receptor interactions to the inactive receptor based on two ligand prototypes, 5HT and LSD, suggests that agonists orient their aromatic moiety towards the H5-H6 interface at the level of F6.52-F5.47, through H-bonding interactions with polar side chains in H3-H4-H5-H6. The ability of agonists, but not

antagonists, to orient their aromatic moiety on the H5-H6 interface distal from D3.32, may be important in inducing or selecting a different orientation of the aromatic cluster than the conformation present in the inactive receptor. We have proposed different specific conformations of the aromatic cluster in the inactivated and activated forms of the receptor. Thus, I propose that agonists, by reaching to the H5-H6 interface close to the lipid milieu, could catalyze the proposed conformational change in the aromatic cluster. This activation-related conformational change is characterized by a change in the W6.48 X1 from *gauche+ (g+)* to *trans (t)*, which necessarily implies the conversion of F652 X1 from *g+* to *t* and probably also F5.47 from *t* to *g+*. Once the agonist has induced (or selected) the *trans*-configuration of the aromatic cluster, shown to span the core of the TMH domain in Figure 8.6, this would have significant effects on the conformation of the cytoplasmic portion of the receptor where ultimately G-proteins bind.

MonteCarlo simulations of the aromatic cluster in an isolated H6 indicated that the conformation of the aromatic cluster modulates the preferred conformation of the Pro-kink at P6.50, and in so doing alters the positioning of the cytoplasmic segment of H6. I distinguish two configurations of the aromatic cluster, defined by an orientation of W6.48 in the *g+* or the *trans* rotamer. The main difference I observed in the Pro-kink conformation between the aromatic cluster in the *g+* versus *trans* configuration is a redirection of the cytoplasmic portion of H6 towards H5 in

the *g+* configuration, and towards H7 in the *trans* configuration. Thus, if agonists induce a transition of the aromatic cluster from a *g+* to a *trans* configuration, this would imply a concomitant movement of H6 away from H5 and towards H7 at the cytoplasmic boundaries. The adoption of a *trans* configuration for the aromatic cluster in H6 would free F6.44 of its high conformational restraints induced by W6.48 adopting X1=*g+* and V6.40, so that F6.44 would now be able to adopt the *trans* rotamer as well. Because F6.44 in the *g+* conformation was shown to determine and severely constrain the conformation of the NPxVY motif in H7, the increased freedom of F6.44 upon the *g+* to *trans* conformational change at the aromatic cluster would imply relieving a significant constraint on H7 at the cytoplasmic side, so that now it can react to the torque from H6 by moving away from H6. As H7 moves away from H6, the palmitoylation site (C7.70) forces the C-terminus of the cytoplasmic extension of H7 to remain in the lipid milieu: this implies that $H7^{(7.59\text{-}7.70)}$ could only move away from H6 by moving towards the lipids, and not towards the protein interior. So the torque induced by H6 on H7 coupled with the relaxation on the constraints on the NPxVY motif both induce and allow a movement of $H7^{(7.59\text{-}7.70)}$ where 7.50-7.54 moves towards the lipids away from the TMH bundle as the $H7^{(7.59\text{-}7.70)}$ follows it with a rotation away from H3-H5 and H6 always maintaining the Palmitoylation site into the membrane.

The movement of $H7^{(7.59\text{-}7.70)}$ and H6 described above would affect the

H3-H6 interaction at the cytoplasmic side in two ways: first, it would destabilize the E6.30-R3.50 interaction by moving away E6.30. Second, the $H7^{(7.59\text{-}7.70)}$ movement towards the membrane would leave the E6.30-R3.50-D3.49 interaction within the Arg-cage motif accessible to the aqueous cytoplasm , where these charged residues can be easily solvated and thus the "lock" holding H3-H6 at the cytoplasmic boundaries would be severely weakened. The movement of H7 away towards the lipids at the cytoplasmic side allows also H1-H2 to move closer to the H6-H3 interface. This would be important because R3.50, now relieved from the interaction with E6.30 and with its interaction with D3.49 weakened by solvation and/or protonation, can rotate to interact with D2.50. The interaction of R3.50 with D2.50 positions the Arg side chain so as to also interact with the carbonyl group of N7.49, which in turn induces a reorientation of the H7 segment that follows the Pro7.50, towards the membrane lipids.

The activated form of the receptor would be achieved as a result of these interrelated conformational changes. The specific TMH-TMH interactions holding these helices together at the cytoplasmic side in the active state would be driven by the same set of conserved charged and polar residues ascribed to the Arg-cage and the H1-H2-H7 cytoplasmic motif in the inactive state. However, as opposed to the inactivated form of the receptor, in the active form these ionic and H-bonding interactions are highly solvated and thus much weaker than in the inactive form. It follows that these TMH

at the cytoplasmic boundaries would be much more flexible in the activated state, as observed by spin labeling experiments (Farahbakhsh 1995; Altenbach 1996; Farrens 1996; Yang 1996a; Yang 1996b). Both the increased solvent accessibility and the increased flexibility at the G-protein coupling domain would facilitate G-protein recognition and thus binding. The same inferences, but to a larger extent, apply for the conformation of the $H7^{(7.59\text{-}7.70)}$ segment in the active state of the receptor. The cytoplasmic extension of H7 would be now extended from the TMH domain towards the lipid headgroups, driven by the interaction of the extensive Arg/Lys patches in this fragment with the negatively charged phospholipid headgroups. The lack of significant tertiary structure interactions would increased the flexibility of the 7.59-7.70 segment, and would result in a significant increase in the exposure of the conserved residues in $H7^{(7.59\text{-}7.70)}$. i.e. {7.59, Y7.60, A7.63, F7.64} that were buried against the Arg-cage and the H1-H2-H7 cytoplasmic motif in the inactive state. The accentuated effect of the increased flexibility and solvent accessibility for the cytoplasmic extension of H7 makes this fragment a primary site for G-protein interactions, consistent with a more conserved pattern at these loci than in other coupling domains. The larger motion predicted for the 7.59-7.70 fragment relative to most other TMHs has been validated by spin labeling experiments between residues C1.60 and C7.63 in rhodopsin (Yang 1996a).

We have proposed an energetic imbalance between the inactive and the active state of the receptor where the active conformation was a transient,

high energy intermediate in the absence of G-protein. The conformational changes discussed above may shed some light into the possible molecular determinants of the proposed energetic imbalance. The cytoplasmic extension of $H7^{(7.59-7.70)}$ is a very amphiphilic helix; one face is conserved and presents two aromatic residues (Y7.60 and F7.64), an Ala at position 7.64, and the palmytoil chain at C7.70. On the opposite face, the length and concentration of Arg/Lys residues through the aligned receptors is unparalleled by any other TMH (see Figure 5.3). In the inactive state, the hydrophobic face is buried against other receptor domains and the Arg/Lys patch is solvated by the aqueous cytoplasm and by interactions with the interhelical connecting regions that were not modeled here. In the active state, the hydrophobic face would be highly exposed and the Arg/Lys patch would participate in extensive electrostatic interactions with the phospholipids headgroups. Energetically, this active state would be stabilized by the ionic interactions of the Arg/Lys lining the phospholipids headgroups, and destabilized by the resulting exposure of the hydrophobic residues Y7.60 (Phe in most receptors), A7.63, and F7.64. In a solvated medium, the hydrophobic effect might prevail over the ionic effect. Therefore, the conformation of the cytoplasmic extension of H7 in the active state would be significantly higher in energy than its inactive state counterpart, thus driving the active form rapidly back into its inactive conformation, in the absence of a stabilizing effect from interaction with the G-protein.

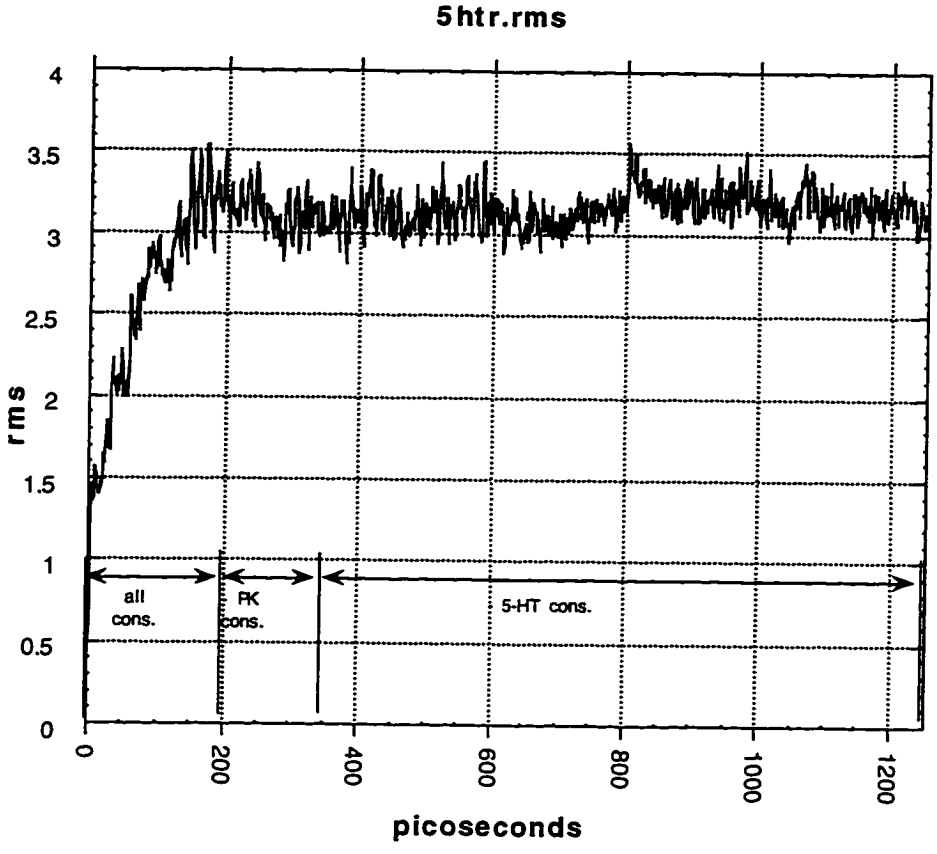


Figure 8.1

Root mean square (rms) deviation of the Cα coordinates from the initial coordinates observed in the MD simulation. The constraints used at each stage are indicated schematically as described in the text. Note that system becomes stable after the intial 200 ps. Therefore, the timeframe comprised between 200 and 1250 picoseconds was selected for analyses of structural features of the receptor

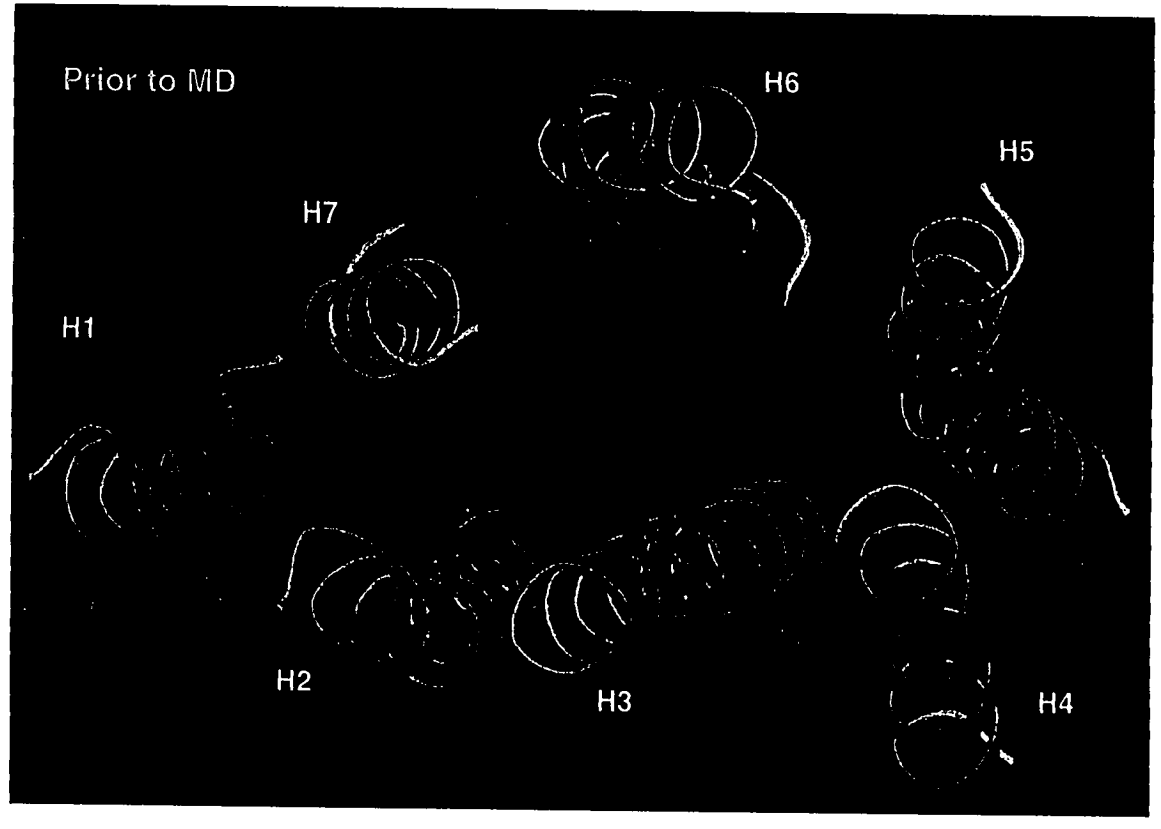


Figure 8.2

The changes in the orientation of the TMHs that result from the MD simulation can be appreciated in a superposition of the initial model (yellow) and the MD model (red). The most significant changes occur in H1 and H4, the most lipid exposed helices and thus with less specific interactions constraining their movements. Note that H3 and H6 at the extracellular side are brought together by the MD simulation.

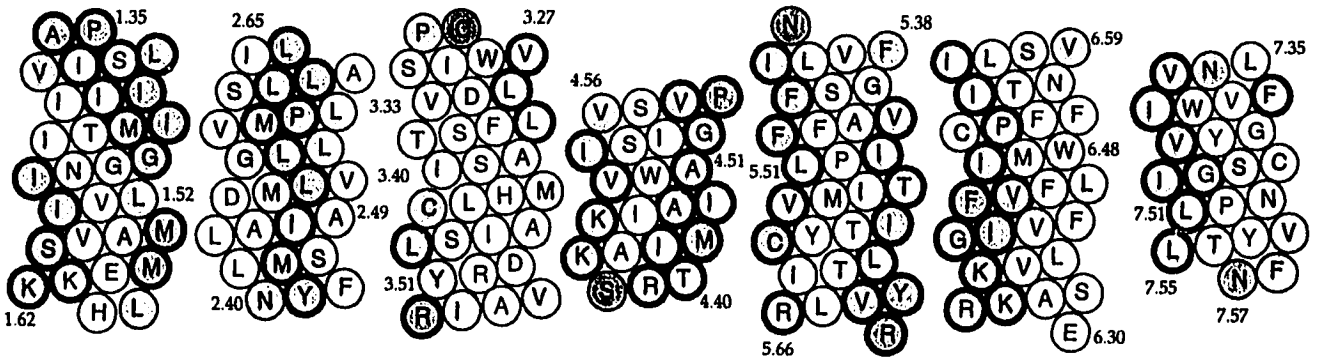


Figure 8.3 :

Solvent accessibility per residue calculated for the energy-refined model. The degree of solvent exposure is indicated by the degree of shading. Previously predicted lipid-exposed residues are indicated by a thicker circle. Note the consistency between predicted and observed accessibilities inside the TMHs, and the discrepancy at some TMH ends, probably due to the absence of loop structures.

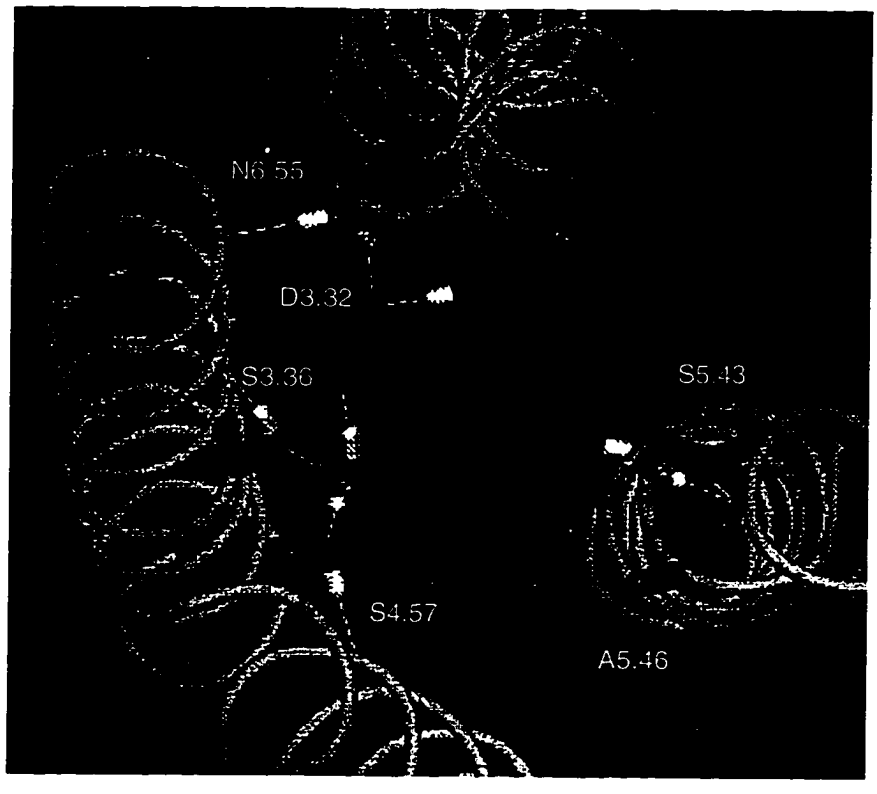


Figure 8.4 :
5HT-receptor H-bonding interactions that result from the simulation (dashed lines). Note the proximity between N6.55 and the 5-OH of serotonin.

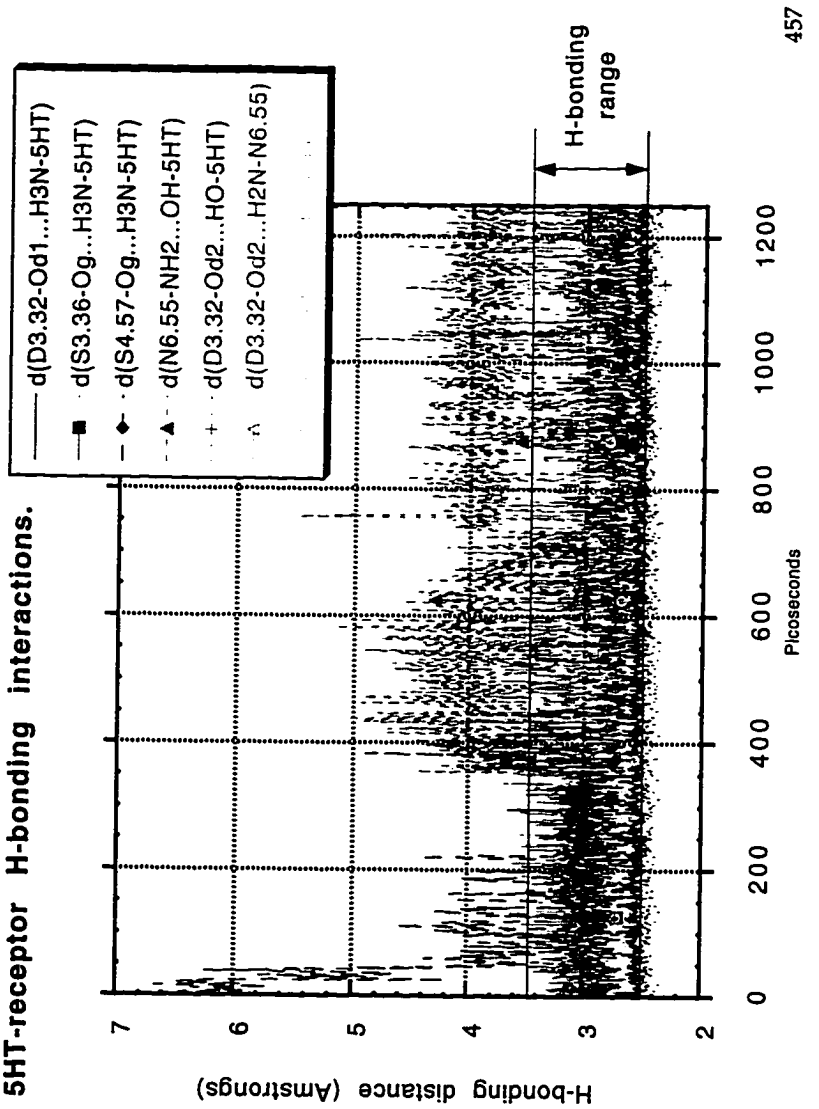


Figure 8.5: Dynamic evolution of 5HT-receptor H-bonding interactions.

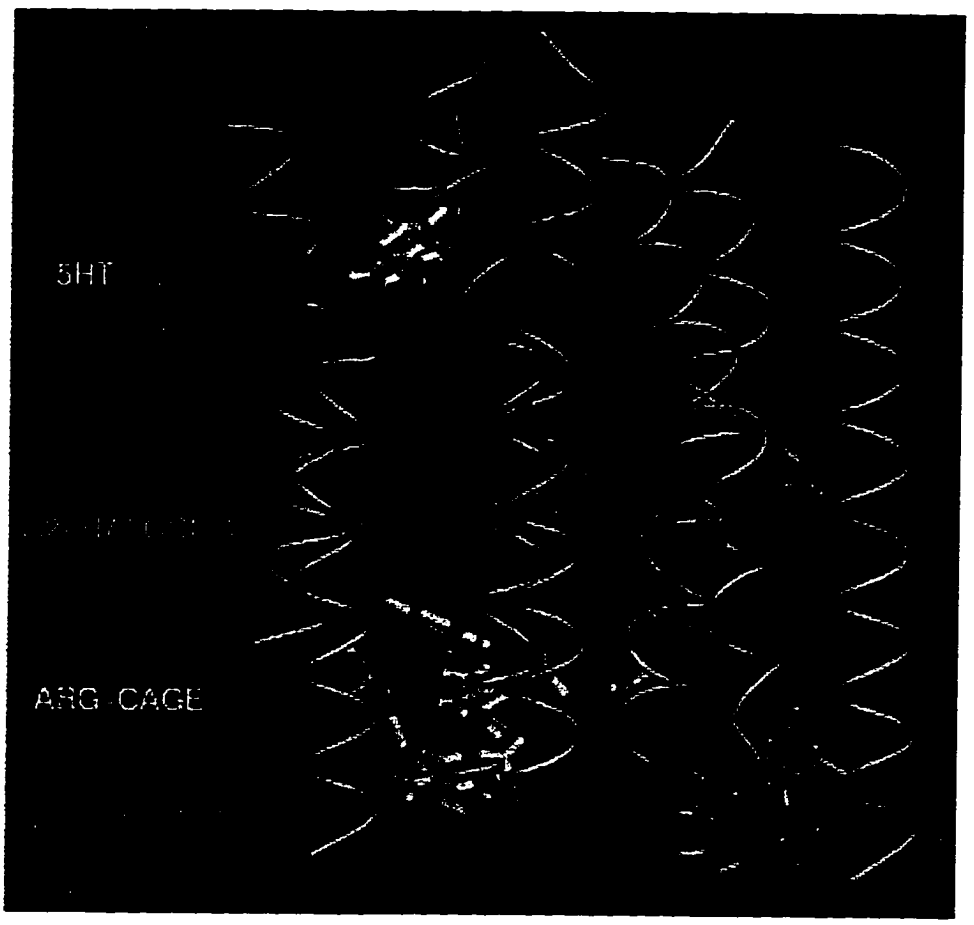


Figure 8.6 :
Organizing principles of the 5HT2C GPCR TMH domain model, color coded.

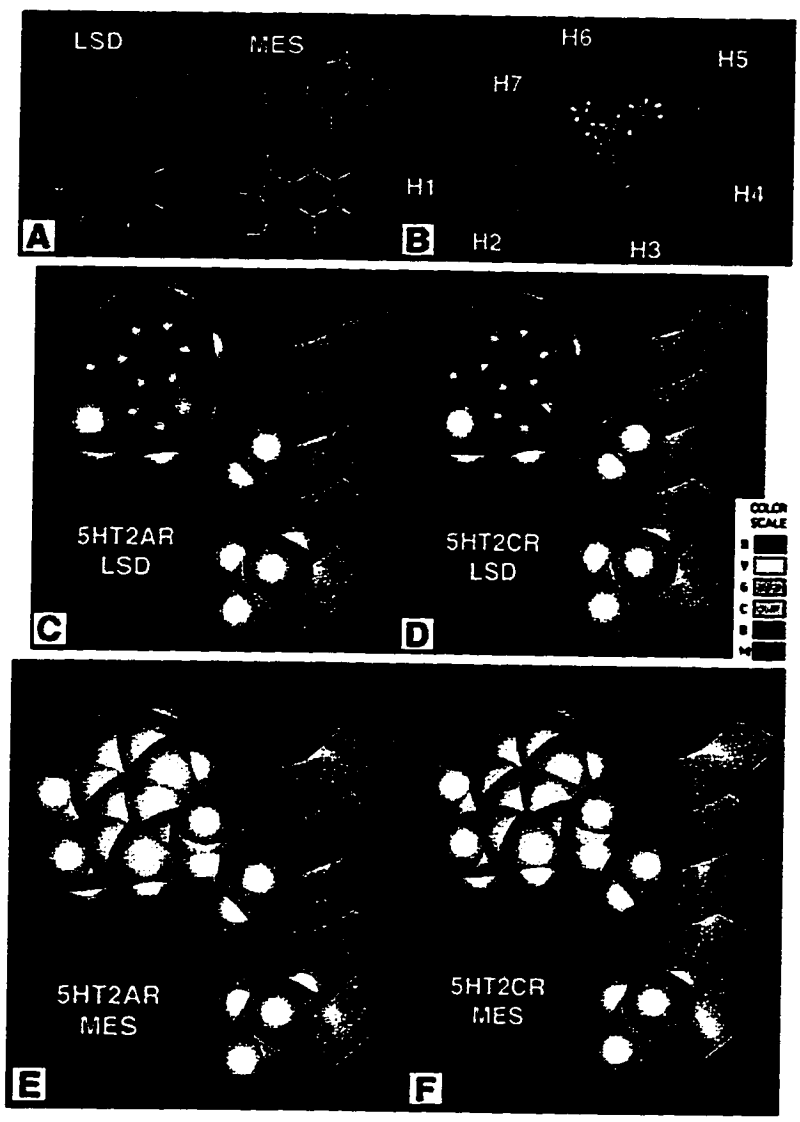


Figure 8.7 :

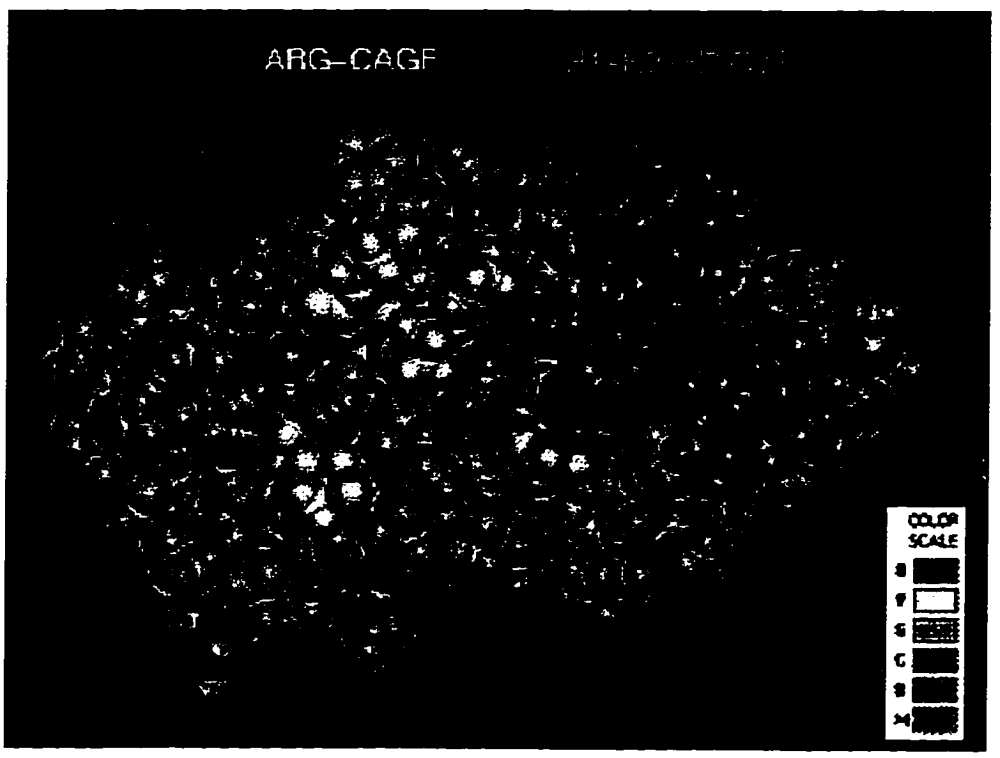


Figure 8.8 :
G-protein coupling domains present in the model are localized within the two clusters of polar residues present at the cytoplasmic boundaries: the Arg-cage motif (yellow) and the H1-H2-H7 cytoplasmic motif (blue). The accessibility of these two motifs is shown by a Van der Waals representation of the model viewed from the cytoplasmic side.

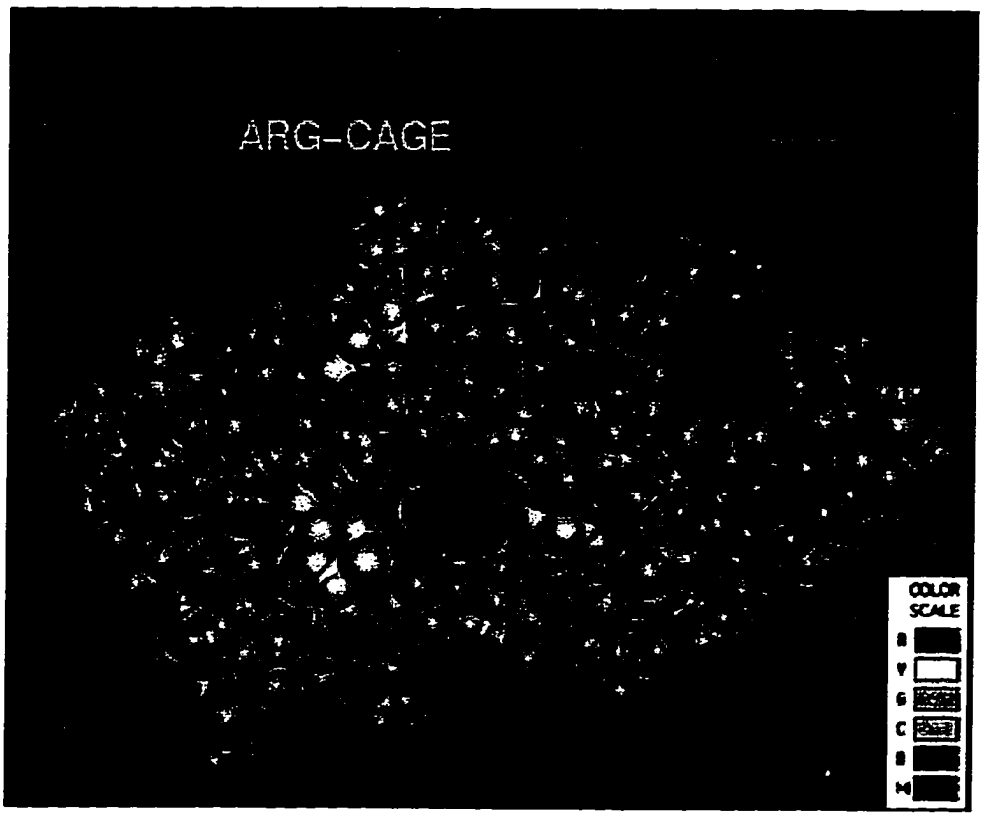


Figure 8.9 :

Van der Waals representation of the model, viewed from the cytoplasmic side, illustrating the hypothetical role of the cytoplasmic extension of H7 (K7.58-C7.70, purple) in shielding the G-protein coupling domain from aquoeus exposure. The coupling domain is illustrated by the Arg-cage motif (yellow) and the H1-H2-H7 cytoplasmic motif (blue).

9. INFERENCES FROM THE MODEL

Analysis of the model at various stages yielded a series of structural and functional inferences that could guide future testing of the GPCRs by experiments designed to probe the models and to produce new mechanistic insight. Some of the specific suggestions inspired by these inferences are presented below.

9.1- Modulation of the conformation of the aromatic cluster in H6 by incorporation of β-branched residues at positions i-4, i+4.

Based on MonteCarlo simulations of the isolated H6, W6.48 was proposed to act like a conformational switch directing the conformation of the conserved aromatic cluster in H6 between a *trans* and a *gauche+* configuration (see Section 7.3.3.1). These two configurations of the aromatic cluster were shown to modulate the conserved PK in H6 (Section 7.2.4), and this modulation has been proposed to be an intrinsic component of the conformational changes responsible for receptor activation (Section 8.4). Spectroscopic data from rhodopsin were interpreted as supporting a *gauche+* conformation of W6.48 in the inactive state and a *trans* configuration for the activated state (see Section 7.3.3.1).

To test this hypothesis, specific receptor mutations have been designed to selectively orient the rotamer positions of the clustered aromatic side

chains into either *trans* or *gauche+* configurations. The observed effect of β-branched amino acids acting as modulators of the conformation of nearby residues, especially Trp, as illustrated by the modulation of W4.50 by I4.56, proposed in section 7.3.3.2., can be used for this purpose. A search of known protein structures indicates that a Val/Ile at position 6.52 would prevent W6.48 from adopting the trans rotamer, thus locking the aromatic cluster in the *gauche+* configuration; a PDB search for aromatic clusters in α-helices showed that 7/8 of such cases had W chi1 in the *gauche+* configuration. Alternatively, a Val/Ile at position 6.44 would hinder W6.48 from adopting the *gauche+* rotamer, thus favoring the *trans* configuration for the aromatic cluster; a PDB search showed that 8/11 of such cases had W *chi1* in the *trans* configuration. Thus, mutation of F6.52->I and F6.44->I would selectively favor either the gauche+ or trans configuration of the aromatic cluster, respectively. I have tested this hypothesis computationally by analyzing MonteCarlo simulations in isolated H6 segments for these mutants, using the same procedures applied in Section 7.2.3 for the wild type H6 segments. In accordance with the hypothesis, it is anticipated that locking these side chains into each of the two possible rotamer configurations will cause the receptor to show increased or decreased efficiency of activation, respectively. If the hypothesis is correct, the receptor will show constitutive activity or a left-shift of the concentration-response curve with aromatic replacement by Ile in one position, while Ile in the other position will create a receptor with decreased

maximal stimulation and/or right-shift of the concentration-response curve. As a control mutation, Ala could also be introduced at each of the two loci studied.

9.2- Modulation of the H6 PK conformation by 5HT2 subtype specific H-bonds.

The proposed role of the PK in H6 in the receptor activation mechanism (see Sections 7.2.4 and 8.4) could be tested by mutations that would selectively affect the PK conformation. Such an interaction was proposed based on MonteCarlo simulations of an isolated H6 for the intrahelical H-bond between C6.49 and the backbone carbonyl of L6.45 (Section 7.2.3). This interaction was maintained in the MD simulations of the model for the TMH domain (C6.49-SH...OC-L6.45 (<d>=3.3±0.3)).

In Section 7.2.4, I derived the hypothesis that C6.49 in the 5HT2CR modulates the conformation of the H6 PK in a manner similar to the pair N7.45-C6.47 present in most GPCR but absent in the 5HT2CR. This hypothesis stems from **erm** among these three residues which have the following distribution:

5HT2R :	C6.49 —> S/C7.45, M6.47
other neuro. GPCRs :	LF6.49 <— N7.45, C6.47

The hypothesis of a similar net structural effect of these divergent H-bonding schemes can be tested by the following mutagenesis scheme, leading to a proposed revertant mutant character: (Note that I am using previously

characterized mutants)

Mutation	
Activation	
M6.47->C mutant (done)	wt
C7.45->N mutant (done)	wt
M6.47->C , C7.45->N double mutant (done)	wt
C6.49-> L mutant	altered
C6.49-> L , M6.47->C , C7.45->N	triple mutant revertant

These experiments could be coupled with computational simulations of the equivalent constructs. Note that there are two possible outcomes of these experiments with important implications: First, agreement with these predictions would strongly support a crucial role for the PK in H6; Second, potential changes in ligand affinity would support an indirect mechanism for ligand specificity between the 5HT2 subclass and other neurotransmitter GPCRs involving modulation of Pro-kinks.

9.3- Arg-cage expanded: Role of E6.30 holding R3.50 in the inactive form.

I have proposed in Sections 7.4.1 and 8.4 that E6.30 is holding R3.50 in the inactive conformation through an ionic interaction. This role resembles the role proposed for D3.49 in the Arg-Cage manuscript (Section 7.3.2), and thus E6.30 was proposed to be an intrinsic component of the Arg-cage motif (see Figure 8.6). This interaction could be responsible for the experimentally observed proximity between H3 and H6 at the cytoplasmic boundaries in the inactive state of rhodopsin (Farrens 1996; Sheikh 1996), which moves away from each other upon activation (Farrens 1996; Sheikh 1996). Thus, I

proposed that the E6.30-R3.50 interaction could be the "lock" holding H3-H6 together in the inactive state of the receptor. Several experimental data could support this hypothesis (see Section 7.4.1).

To test this hypothesis, E6.30 could be mutated to Asp, Gln and Leu, which would be expected to result in an enhanced activation of the 5HT2CR.

9.4- New ligand-receptor interactions proposed from the analysis of the receptor model:

N6.55 : The hypothesis that N6.55 H-bonds the hydroxy-alkoxy site at position 5 of indoline and ergolines, e.g. the 5-OH group of 5-HT, presented in Section 7.4.5, is consistent with the results from the MD simulations (see Section 8.1). This hypothesis could be tested by mutation of N6.55 to Ala, His, and Thr. The N6.55A mutant should show a decreased affinity for 5HT and 5-MeO-tryptamine. No significant affinity change for these ligands would be expected for the N6.55H mutant, which maintains the H-bond donor at position delta along the side chain. The N6.55T mutant is designed to position an H-bond acceptor at this locus, due to the strong preference of Thr residues to H-bond back to the backbone carbonyl of the preceding turn. Thus, the exchange of a H-bond donor site (N6.55) by a H-bond acceptor site (N6.55T) would be expected to increase the affinity for 5-NH_2-HT (an H-bond donor) and to decrease the affinity for 5MeO-HT (an H-bond acceptor) relative to wild type 5HT2CR. Note that the affinity of 5HT for the N6.55T mutant can not be

predicted due to the dual role of the OH group as donor and acceptor. The set of ligands proposed above is meant to illustrate the criterion of selecting alkyl-indolines and ergolines with a well determined H-bonding donor or acceptor capabilities at the 5-OH position of 5HT, which could be extended to other ligands.

S4.57 : The observed H-bond between S4.57 and the protonated amine of 5HT could be studied by mutation of S4.57 to Ala. The selection of ligands for this study would be the same set selected for the study of 5HT with S3.36 for the 5HT2AR published by other members of this laboratory. The appropriated ligands for this study will be a set of N-substituted and NN-disubstituted congeneric compounds.

9.5- Engineering an all-activated receptor that would test the role of the cytoplasmic extension of H7 in maintaining the inactive state of the receptor.

The hypothetical role of the cytoplasmic extension of H7(7.58-7.70) in the receptor activation mechanism was presented in Section 8.4, where it was predicted that this segment is important in stabilizing the inactive conformation of the receptor by shielding the G-protein coupling motifs from water exposure (see Figure 8.9). The resulting orientation of the helical segment (7.58-7.70) as it packs againts the TMH domain is shown in Figure 9.1; this is the inactive conformation. In the active receptor, this segment is predicted to lack significant tertiary structure interactions which should

render it highly flexible. It is most likely that this segment would line the membrane-water interface where it can be stabilized by ionic interactions with the phospholipid headgroups. The larger movement and flexibility predicted for this segment relative to other helical domains is due to the flexible hinge observed at the turn T7.54-N7.57 (see Section 5.6). Rigidifying this turn in a predetermined direction could induce an activated or inactivated state of the receptor. In particular, the double mutation L7.55C and K7.58C would be expected to form a disulfide bridge between these two positions, inducing a bend in the T7.54-N7.57 hinge that would orient the cytoplasmic extension of H7 towards the lipid phase, similar as the active state conformation. This mutation could therefore result in a covalently stabilized activated stated of the receptor, i.e. not a just a constitutively but a completely active receptor.

However, CAM receptors are also consitutively phosphorylated and thus significantly desensitized. There are reason to suggest that the L7.55C, K7.58C activated receptor may not be desensitized under the same conditions. The strong amphiphilic character of $H7^{(7.59\text{-}7.70)}$ was proposed to guide the energetic balance between the inactive and active conformations, based on the penalty of exposing hydrophobic residues in the active state and the energy gained by moving the Arg/Lys rich face of this helix from a water medium towards ionic interactions with the membrane. In this context, phosphorylation would induce a density of negatively charged, phosphate

groups at the cytoplasmic loops. These phosphates could act in the same role as the phosphates in the phospholipid headgroup, i.e. they would develop ionic interactions with the Arg/Lys patch of $H7^{(7.59-7.00)}$. It follows that phosphorylation would stabilize the inactive conformation through ionic interactions with the Arg/Lys on the hydrophilic face of $H7^{(7.59-7.70)}$, and in so doing would prevent this segment from adopting the active conformation, which depends on ionic bonds between same Arg/Lys patch on $H7^{(7.59-7.70)}$ and the phospholipid headgroups for its stabilization.

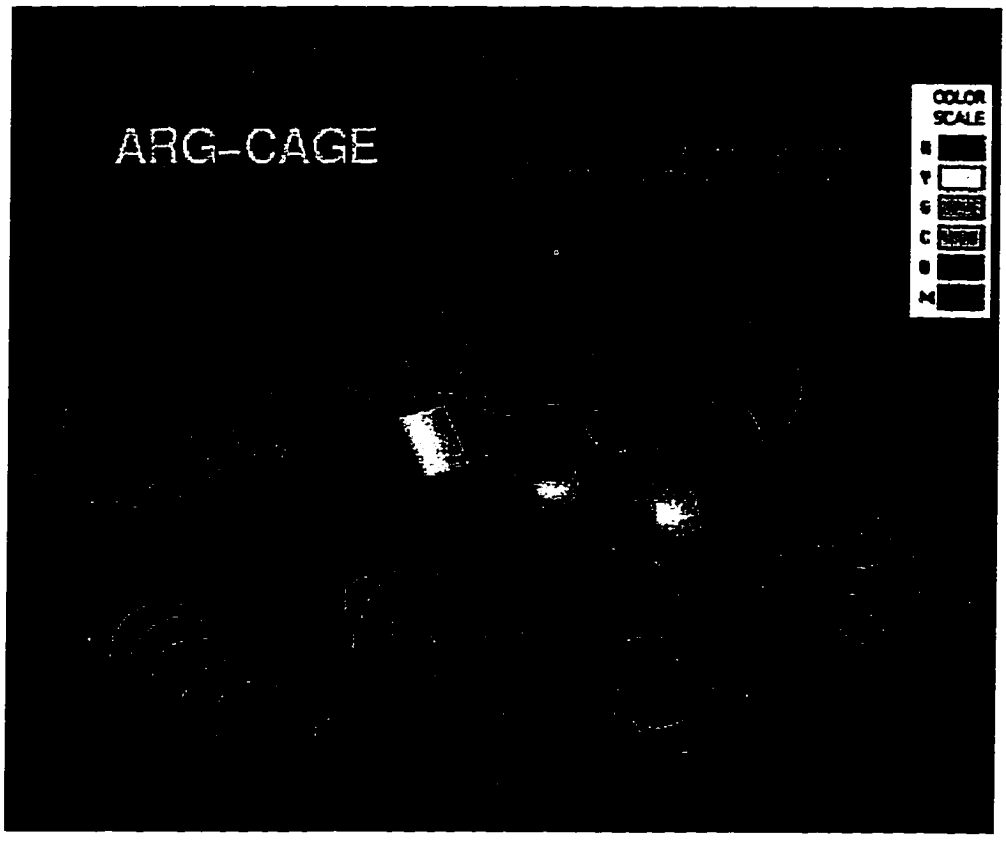


Figure 9.1 :
Inactive conformation of the cytoplasmic extension of H7 (K7.58-C7.70). See Figure 8.3.2.II for color coding.

10.- CONCLUSION

A model of the transmembrane domain of the serotonin 2C G-protein coupled receptor (GPCR) complexed with its natural ligand, serotonin (5HT), has been constructed and analyzed by energetic, experimental data, and functional criteria. To guide the molecular modeling approach, the ternary complex model of receptor activation is first interpreted in terms of energy levels for the principal states, thus providing a thermodynamic framework for receptor activation that suggests a metastable activated receptor in the absence of G-protein. The molecular modeling process was therefore focused on the inactive state of the receptor. The starting point was the analysis of conservation patterns and functional divergence, observed in a multiple sequence alignment of neurotransmitter GPCRs. The conserved properties at each locus are considered to be molecular determinants of the receptor structure, and the observed conservation patterns are used to derive specific predictions concerning the structures of individual transmembrane segments as well as the entire receptor domain.

The boundaries of the 7 transmembrane segments considered first to be α-helices (TMH) were predicted based on a hydrophobicity plot and were refined by the analysis of α-helix periodicity in the conservation pattern as measured by Fourier transform analysis or by surface patches on a helical net representation. The helical periodicity of non-conserved Arg/Lys occurring at

the cytoplasmic boundaries was identified as a novel predictive tool according to which such Arg/Lys are predicted to face the mostly negatively charged phospholipid headgroups at the cytoplasmic boundaries. This structural motif leads to a redefinition of the transmembrane segments. Integration of results from the various criteria results in TMH boundaries that include most proposed G-protein coupling domains, and are in agreement with results from NMR, SCAM and Spin labeling experiments.

The orientation of each TMH within the transmembrane domain of the receptor was then predicted based on the identification of amino acids (AA) in TMH-TMH or TMH-Lipid interfaces. Conserved positions are positioned at TMH-TMH interfaces, and a novel probability function was developed to predict AA facing the hydrophobic core (Lipid chains). The positioning of {Pro, Ser, Thr, Cys, Gly} residues in TMH-TMH or TMH-Lipid interfaces was considered in the light of experimental data regarding intrahelical hydrogen bonding to i-4 positions. Integration of the various criteria for the positioning of AA at various loci at TMH-TMH and TMH-Lipid interfaces yielded specific predictions that were shown to be in agreement with SCAM and Spin labeling experiments. The extent of lipid exposure for each TMH, estimated from these predictions, determines the shape of the TMH domain. This criterion, in combination with experimentally derived TMH-TMH interactions and the adjacencies proposed by short loops, is used to derive a sequential and anticlockwise arrangement

of the 7 TMHs based on the rhodopsin map. This allows the classification of interior-facing residues into sequential TMH-TMH interfaces. A new method was developed to predict the direction and degree of helix tilt within the transmembrane domain, from the analysis of the predicted helix-helix versus helix-lipid interface patterns, compared to the tilts observed experimentally for rhodopsin.

The TMH packing arrangement, orientations and tilts predicted above were used to position the transmembrane segments in a three-dimensional model of the 5-HT2C GPCR. Experimentally derived information about TMH-TMH interactions is used to select specific Pro-kink (PK) conformations in the TMHs, using a definition of the extent of the conformational freedom induced by PKs and the direction of bending and face-twisting it allows.

Several organizing motifs are recognized within the transmembrane domain of the receptor. The H2-H7-H1 cluster of polar residues is defined by involving D2.50 and N7.49 double revertant mutant constructs in the GnRH and 5HT2C receptors (Zhou 1994; Sealfon 1995). Charge-charge interactions involving the conserved Arg at the cytoplasmic boundary of H3 define another motif: The "Arg-cage". An "aromatic cluster" motif is identified among residues in transmembrane segments H4, H5, and H6. This motif is characterized by a proposed "aromatic rotamer switch" in H6 which relates to receptor activation, and by a Trp in H4 that is conformationally constrained due to the conserved β-branched residue at position 4.46. Residues proposed

to participate in direct interactions with neurotransmitter receptor ligands cluster on the extracellular side of H3-H5-H6-H7 and form the ligand binding motif.

The resulting three-dimensional model of the TMH domain of the 5-HT2C receptor (5HT2CR) is subjected to probing by criteria based on considerations of energetics, experimental data, and functional implications. Molecular Dynamics simulations of the 5HT2CR complex with serotonin produce an energy-refined model that satisfies the essential criteria used for its construction. A summary of experimentally validated criteria used in the modeling process is presented in the context of the energy-refined model. Functional evaluation of the 5HT2CR model is divided into predictions pertinent to ligand binding, the G-protein coupling domain, and the activation mechanism. The model is shown to be fully consistent with available data and to have yielded specific guides for experimental verifications that validate it.

REFERENCES

Adham, N., J. A. Tamm, J. A. Salon, P. J. Vaysse, R. L. Weinshank, and T. A. Branchek (1994). "A single point mutation increases the affinity of serotonin 5-HT1D alpha, 5-HT1D beta, 5-HT1E and 5-HT1F receptors for beta-adrenergic antagonists" Neuropharmacology **33**(3-4): 387-91 Issn: 0028-3908.

Almaula, N., B. J. Ebersole, J. A. Ballesteros, H. Weinstein, and S. C. Sealfon (1996a). "Contribution of a helix 5 locus to selectivity of hallucinogenic and nonhallucinogenic ligands for the human 5-hydroxytryptamine2A and 5-hydroxytryptamine2C receptors: direct and indirect effects on ligand affinity mediated by the same locus" Mol Pharmacol **50**(1): 34-42.

Almaula, N., B. J. Ebersole, D. Zhang, H. Weinstein, and S. C. Sealfon (1996b). "Mapping the binding site pocket of the serotonin 5-Hydroxytryptamine2A receptor. Ser3.36(159) provides a second interaction site for the protonated amine of serotonin but not of lysergic acid diethylamide or bufotenin" J Biol Chem **271**(25): 14672-5.

Altenbach, C., K. Yang, D. L. Farrens, Z. T. Farahbakhsh, H. G. Khorana, and W. L. Hubbell (1996). "Structural features and light-dependent changes in the cytoplasmic interhelical E-F loop region of rhodopsin: a site-directed spin-labeling study" Biochemistry **35**(38): 12470-8.

Altschuh, D., A. M. Lesk, A. C. Bloomer, and A. Klug (1987). "Correlation of co-ordinated amino acid substitutions with function in viruses related to tobacco mosaic virus" J Mol Biol **193**(4): 693-707.

Altschuh, D., T. Vernet, P. Berti, D. Moras, and K. Nagai (1988). "Coordinated amino acid changes in homologous protein families" Protein Eng **2**(3): 193-9.
Altschul, S. F. (1991). "Amino acid substitution matrices from an information theoretic perspective" J Mol Biol **219**(3): 555-65.

Arnis, S., K. Fahmy, K. P. Hofmann, and T. P. Sakmar (1994). "A conserved carboxylic acid group mediates light-dependent proton uptake and signaling by rhodopsin" J-Biol-Chem **269**(39): 23879-81.

Baker, R. W., C. Chothia, P. Pauling, and H. P. Weber (1972). "Molecular structure of LSD" Science **178**(61): 614-5.

Baldwin, J. M. (1993). "The probable arrangement of the helices in G protein-coupled receptors" Embo J **12**(4): 1693-703.

Ballesteros, J. A., S. Kitanovic, L. Chi, V. Rodic, P. D. Davis, B. J. Fromme, K. Konvicka, J. S. Davidson, H. Weinstein, and S. C. Sealfon (1996). "The orientation of the conserved Arg in helix 3 required for efficient signaling of the Gonadotropin Releasing Hormone receptor (GnRH) is maintained by a cage of surrounding residues: role of beta-branched aminoacids" Manuscript in preparation.

Ballesteros, J. A., and H. Weinstein (1992a). "Analysis and refinement of criteria for predicting the structure and relative orientations of transmembranal helical domains" Biophys J **62**(1): 107-9.

Ballesteros, J. A., and H. Weinstein (1992b). "The role of Pro/Hyp-kinks in determining the transmembrane helix length and gating mechanism of a [Leu]zervamicin channel" Biophys J **62**(1): 110-1.

Ballesteros, J. A., and H. Weinstein (1995). "Integrated Methods for Modeling G-Protein Coupled Receptors" Methods Neurosci. **25**:366-428.

Barlow, D. J., and J. M. Thornton (1988). "Helix geometry in proteins" J Mol Biol **201**(3): 601-19.

Bazzo, R., M. J. Tappin, A. Pastore, T. S. Harvey, J. A. Carver, and I. D. Campbell (1988). "The structure of melittin. A 1H-NMR study in methanol" Eur J Biochem **173**(1): 139-46.

Benner, S. A., and D. Gerloff (1991). "Patterns of divergence in homologous proteins as indicators of secondary and tertiary structure: a prediction of the structure of the catalytic domain of protein kinases" Adv Enzyme Regul **31**:121-81.

Berlose, J. P., O. Convert, A. Brunissen, G. Chassaing, and S. Lavielle (1994). "Three-dimensional structure of the highly conserved seventh transmembrane domain of G-protein coupled receptors" Eur. J. Biochem. **225**:827-843.

Blumenthal, L. M., 1953, Theory and applications of distance geometry, Oxford, New York, Oxford University Press, reprinted by Chelsea publishing company, p. 97-99.

Blundell, T., D. Barlow, N. Borkakoti, and J. Thornton (1983). "Solvent-induced distortions and the curvature of alpha-helices" Nature **306**(5940): 281-3.

Bordo, D., and P. Argos (1990). "Evolution of protein cores. Constraints in point mutations as observed in globin tertiary structures" J Mol Biol **211**(4): 975-88.

Bormann, B. J., and D. M. Engelman (1992). "Intramembrane helix-helix association in oligomerization and transmembrane signaling" Annu Rev Biophys Biomol Struct **21**:223-42.

Bouvier, M., P. Chidiac, T. E. Hebert, T. P. Loisel, S. Moffett, and B. Mouillac (1995). "Dynamic palmitoylation of G-protein-coupled receptors in eukaryotic cells" Methods Enzymol **250**:300-14.

Bramblett, R. D., A. M. Panu, J. A. Ballesteros, and P. H. Reggio (1995). "Construction of a 3D model of the cannabinoid CB1 receptor: determination of helix ends and helix orientation" Life-Sci **56**(23-24): 3205.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus (1983). "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations" J. Comp. Chem. **4**(2): 187-217.

Brooks, C. L., M. Karplus, and B. M. Pettitt, 1988, Proteins: A Theoretical Perspective of Dynamics, Structure, and their Thermodynamics: New York, John Wiley & Sons.

Burley, S. K., and G. A. Petsko (1985). "Aromatic-aromatic interaction: a mechanism of protein structure stabilization" Science **229**(4708): 23-8.

Burris, K. D., and E. Sanders-Bush (1992). "Unsurmountable antagonism of brain 5-hydroxytryptamine2 receptors by (+)-lysergic acid diethylamide and bromo-lysergic acid diethylamide" Mol Pharmacol **42**(5): 826-30.

Caron, M. G. (1989). "The guanine nucleotide regulatory protein-coupled receptors for nucleosides, nucleotides, amino acids and amine neurotransmitters" Curr Opin Cell Biol **1**(2): 159-66.

Carter, D., M. Champney, B. Hwang, and R. M. Eglen (1995). "Characterization of a postjunctional 5-HT receptor mediating relaxation of guinea-pig isolated ileum" Eur J Pharmacol **280**(3): 243-50.

Chan, S. C., A. K. Wong, and D. K. Chiu (1992). "A survey of multiple sequence comparison methods" Bull Math Biol **54**(4): 563-98.

Cheung, A. H., R. R. Huang, and C. D. Strader (1992). "Involvement of specific hydrophobic, but not hydrophilic, amino acids in the third intracellular loop

of the beta-adrenergic receptor in the activation of Gs" Mol Pharmacol **41**(6): 1061-5.

Chi, L., W. Zhou, A. Prikhozhan, C. Flanagan, J. S. Davidson, M. Golembo, N. Illing, R. P. Millar, and S. C. Sealfon (1993). "Cloning and characterization of the human GnRH receptor" Mol. Cell. Endo. **91**:R1-R6.

Cho, W., L. P. Taylor, and H. Akil (1996). "Mutagenesis of residues adjacent to transmembrane prolines alters D1 dopamine receptor binding and signal transduction" Mol Pharmacol **50**(5): 1338-45.

Chothia, C. (1984). "Principles that determine the structure of proteins" Annu Rev Biochem **53**:537-72.

Chothia, C., and A. M. Lesk (1986). "The relation between the divergence of sequence and structure in proteins" Embo J **5**(4): 823-6.

Chothia, C., M. Levitt, and D. Richardson (1981). "Helix to helix packing in proteins" J Mol Biol **145**(1): 215-50.

Choudhary, M. S., S. Craigo, and B. L. Roth (1992). "Identification of receptor domains that modify ligand binding to 5-hydroxytryptamine2 and 5-hydroxytryptamine1c serotonin receptors" Mol Pharmacol **42**(4): 627-33.

Choudhary, M. S., N. Sachs, A. Uluer, R. A. Glennon, R. B. Westkaemper, and B. L. Roth (1995). "Differential ergoline and ergopeptine binding to 5-hydroxytryptamine2A receptors: ergolines require an aromatic residue at position 340 for high affinity binding [published erratum appears in Mol Pharmacol 1995 Sep;48(3):568]" Mol-Pharmacol **47**(3): 450-7.

Clementi, E., 1991, Modern Techniques in Computional Chemistry: MOTECC-91., Leiden, ESCOM.

Colquhoun, D., and M. Farrant (1993). "The binding issue" Nature **366**:510-511.

Cosson, P., and J. S. Bonifacino (1992). "Role of transmembrane domain interactions in the assembly of class II MHC molecules" Science **258**(5082): 659-62.

Costa, T., Y. Ogino, P. J. Munson, H. O. Onaran, and D. Rodbard (1992). "Drug efficacy at guanine nucleotide-binding regulatory protein-linked receptors: thermodynamic interpretation of negative antagonism and of receptor activity in the absence of ligand" Mol Pharmacol **41**(3): 549-60.

Cowan, S. W., T. Schirmer, G. Rummel, M. Steiert, R. Ghosh, R. A. Pauptit, J. N. Jansonius, and J. P. Rosenbusch (1992). "Crystal structures explain functional properties of two E. coli porins" Nature **358**(6389): 727-33.

Cronet, P., C. Sander, and G. Vriend (1993). "Modeling of transmembrane seven helix bundles" Protein Eng **6**(1): 59-64.

Dahl, S. G., O. Edvardsen, and I. Sylte (1991). "Molecular dynamics of dopamine at the D2 receptor" Proc Natl Acad Sci U S A **88**(18): 8111-5.

Davies, A., G. F. Schertler, B. E. Gowen, and H. R. Saibil (1996). "Projection structure of an invertebrate rhodopsin" J Struct Biol **117**(1): 36-44.

Davison, M. D., and J. B. Findlay (1986a). "Identification of the sites in opsin modified by photoactivated azido[125I]iodobenzene" Biochem J **236**(2): 389-95.

Davison, M. D., and J. B. Findlay (1986b). "Modification of ovine opsin with the photosensitive hydrophobic probe 1-azido-4-[125I]iodobenzene. Labelling of the chromophore-attachment domain" Biochem J **234**(2): 413-20.

Dayhoff, M. O., W. C. Barker, and L. T. Hunt (1983). "Establishing homologies in protein sequences" Methods Enzymol **91**:524-45.

Deber, C. M., C. J. Brandl, R. B. Deber, L. C. Hsu, and X. K. Young (1986a). "Amino acid composition of the membrane and aqueous domains of integral membrane proteins" Arch Biochem Biophys **251**(1): 68-76.

Deber, C. M., D. W. Hughes, P. E. Fraser, A. B. Pawagi, and M. A. Moscarello (1986b). "Binding of human normal and multiple sclerosis-derived myelin basic protein to phospholipid vesicles: effects on membrane head group and bilayer regions" Arch Biochem Biophys **245**(2): 455-63.

Deber, C. M., Z. Li, C. Joensson, M. Glibowicka, and G. Y. Xu (1992). "Transmembrane region of wild-type and mutant M13 coat proteins. Conformational role of beta-branched residues" J Biol Chem **267**(8): 5296-300.

Degli Esposti, M., M. Crimi, and G. Venturoli (1990). "A critical evaluation of the hydropathy profile of membrane proteins" Eur J Biochem **190**(1): 207-19.

Deisenhofer, J., and H. Michel (1989). "Nobel lecture. The photosynthetic reaction centre from the purple bacterium Rhodopseudomonas viridis" Embo J **8**(8): 2149-70.

Deisenhofer, J., H. Michel, T. O. Yeates, H. Komiya, D. C. Rees, J. P. Allen, and G. Feher (1987). "Structure of the reaction center from Rhodobacter sphaeroides R-26: membrane-protein interactions" Proc Natl Acad Sci U S A **84**(18): 6438-42.

Dixon, R. A., I. S. Sigal, and C. D. Strader (1988). "Structure-function analysis of the beta-adrenergic receptor" Cold Spring Harb Symp Quant Biol **1**:487-97.

Dohlman, H. G., M. Bouvier, J. L. Benovic, M. G. Caron, and R. J. Lefkowitz (1987). "The multiple membrane spanning topography of the beta 2-adrenergic receptor. Localization of the sites of binding, glycosylation, and regulatory phosphorylation by limited proteolysis" J Biol Chem **262**(29): 14282-8.

Dohlman, H. G., M. G. Caron, C. D. Strader, N. Amlaiky, and R. J. Lefkowitz (1988). "Identification and sequence of a binding site peptide of the beta 2-adrenergic receptor" Biochemistry **27**(6): 1813-7.

Donnelly, D., and R. J. Cogdell (1993). "Predicting the point at which transmembrane helices protrude from the bilayer: a model of the antenna complexes from photosynthetic bacteria" Protein Eng **6**(6): 629-635.

Donnelly, D., J. B. C. Findlay, A. M. MacLeod, and T. L. Blundell (1994). "The evolution and structure of aminergic G protein-coupled receptors" submitted

Donnelly, D., M. S. Johnson, T. L. Blundell, and J. Saunders (1989). "An analysis of the periodicity of conserved residues in sequence alignments of G-protein coupled receptors. Implications for the three-dimensional structure" Febs Lett **251**(1-2): 109-16.

Donnelly, D., J. P. Overington, S. V. Ruffle, J. H. A. Nugent, and T. L. Blundell (1993). "Modelling alpha-helical transmembrane domains: The calculation and use of substitution tables for lipid-facing residues." Protein Sci. **2**:55-70.

Dunbrack, R. L. J., and M. Karplus (1993). "Backbone-dependent rotamer library for proteins. Application to side-chain prediction" J Mol Biol **230**(2): 543-74.

Edelman, J. (1993). "Quadratic minimization of predictors for protein secondary structure. Application to transmembrane alpha-helices" J Mol Biol **232**(1): 165-91.

Edelman, J., and S. H. White (1989). "Linear optimization of predictors for secondary structure. Application to transbilayer segments of membrane

proteins" J Mol Biol **210**(1): 195-209.

Edvardsen, O., I. Sylte, and S. G. Dahl (1992). "Molecular dynamics of serotonin and ritanserin interacting with the 5-HT2 receptor" Brain Res Mol Brain Res **14**(3): 166-78.

Eisenberg, D., E. Schwarz, M. Komaromy, and R. Wall (1984a). "Analysis of membrane and surface protein sequences with the hydrophobic moment plot" J Mol Biol **179**(1): 125-42.

Eisenberg, D., R. M. Weiss, and T. C. Terwilliger (1984b). "The hydrophobic moment detects periodicity in protein hydrophobicity" Proc Natl Acad Sci U S A **81**(1): 140-4.

Elling, C. E., S. M. Nielsen, and T. W. Schwartz (1995). "Conversion of antagonist-binding site to metal-ion site in the tachykinin NK-1 receptor" Nature **374**(6517): 74-7.

Elling, C. E., and T. W. Schwartz (1996). "Connectivity and orientation of the seven helical bundle in the tachykinin NK-1 receptor probed by zinc site engineering" Embo J **15**(22): 6213-9.

Fahmy, K., F. Jager, M. Beck, T. A. Zvyaga, T. P. Sakmar, and F. Siebert (1993). "Protonation states of membrane-embedded carboxylic acid groups in rhodopsin and metarhodopsin II: a Fourier-transform infrared spectroscopy study of site-directed mutants" Proc Natl Acad Sci U S A **90**(21): 10206-10.

Farahbakhsh, Z. T., K. D. Ridge, H. G. Khorana, and W. L. Hubbell (1995). "Mapping light-dependent structural changes in the cytoplasmic loop connecting helices C and D in rhodopsin: a site-directed spin labeling study" Biochemistry **34**(27): 8812-9.

Farrens, D. L., C. Altenbach, K. Yang, W. L. Hubbell, and H. G. Khorana (1996). "Requirement of rigid-body motion of transmembrane helices for light activation of rhodopsin" Science **274**:768-770.

Fasman, G. D., 1989, Prediction of Protein Structure and the Principles of Protein Conformation, New York, Plenum Press, p. 798.

Findlay, J., and E. Eliopoulos (1990). "Three-dimensional modelling of G protein-linked receptors [published erratum appears in Trends Pharmacol Sci 1991 Mar;12(3):81]" Trends Pharmacol Sci **11**(12): 492-9.

Findlay, J. B., P. L. Barclay, M. Brett, M. Davison, D. J. Pappin, and P.

Thompson (1984). "The structure of mammalian rod opsins" Vision Res **24**(11): 1501-8.

Findlay, J. B. C., and D. Donnelly, 1994, The superfamily: Molecular modeling, *in* B. Dickey, and L. Birnbaumer, eds., GTPases in biology, Heidelberg, Springer-Verlag.
Flocco, M. M., and S. L. Mowbray (1994). "Planar stacking interactions of arginine and aromatic side-chains in proteins" J Mol Biol **235**(2): 709-17.

Fong, T. M., R. R. Huang, and C. D. Strader (1992). "Localization of agonist and antagonist binding domains of the human neurokinin-1 receptor" J Biol Chem **267**(36): 25664-7.

Frauenfelder, H., F. Parak, and R. D. Young (1988). "Conformational substates in proteins" Annu Rev Biophys Biophys Chem **17**:451-79.

Fu, D., J. A. Ballesteros, H. Weinstein, J. Chen, and J. A. Javitch (1996). "Residues in the seventh membrane-spanning segment of the dopamine D2 receptor accessible in the binding-site crevice" Biochemistry **35**(35): 11278-85.

Gallusser, A., and A. Kuhn (1990). "Initial steps in protein membrane insertion. Bacteriophage M13 procoat protein binds to the membrane surface by electrostatic interaction" Embo J **9**(9): 2723-9.

Ganter, U. M., T. Charitopoulos, N. Virmaux, and F. Siebert (1992). "Conformational changes of cytosolic loops of bovine rhodopsin during the transition to metarhodopsin-II: an investigation by Fourier transform infrared
difference spectroscopy." Photochem Photobiol **56**:57-62.

Gantz, I., J. DelValle, L. D. Wang, T. Tashiro, G. Munzert, Y. J. Guo, Y. Konda, and T. Yamada (1992). "Molecular basis for the interaction of histamine with the histamine H2 receptor" J Biol Chem **267**(29): 20840-3.

Garcia-Quintana, D., A. Francesch, P. Garriga, A. R. de Lera, E. Padros, and J. Manyosa (1995). "Fourier transform infrared spectroscopy indicates a major conformational rearrangement in the activation of rhodopsin" Biophys J **69**(3): 1077-82.

Garcia-Quintana, D., P. Garriga, and J. Manyosa (1993). "Quantitative characterization of the structure of rhodopsin in disc membrane by means of Fourier transform infrared spectroscopy" J Biol Chem **268**(4): 2403-9.

Gether, U., J. A. Ballesteros, R. Seifert, E. Sanders-Bush, H. Weinstein, and B.

K. Kobilka (1997). "Structural instability of a constitutively active G protein coupled receptor: Agonist-independent activation due to conformational flexibility." J. Biol. Chem. **272**:2587-2590.

Gether, U., S. Lin, and B. K. Kobilka (1995). "Fluorescent labeling of purified ß2 adrenergic receptor: evidence for ligand-specific conformational changes" J. Biol. Chem. **270**:28268-28275.

Glennon, R. A., M. Dukat, R. B. Westkaemper, A. Ismaiel, D. G. Izzarelli, and E. Parker (1996). "The Binding of Propanolol at 5-Hydroxytryptamine1Db T355N Mutant Receptors May Involve Formation of Two Hydrogen Bonds to Asparagine" Mol. Pharmacol. **49**:198-206.

Glennon, R. A., R. B. Wetkaemper, and P. Bartyzel, 1991, Medicinal Chemistry of Serotonergic Agents, *in* S. J. Peroutka, ed., Serotonin recptor subtypes: basic and clinical aspects: Receptor biochemistry and methodology, New York, Wiley-Liss, Inc., p. 19-64.

Granatir, C. A., L. A. Parodi, and G. M. Maggiora (1994). "A consensus procedure for predicting the location of α-helical transmembrane segments: application to G-protein coupled receptors" Prot. Science **Submitted**:

Gray, T. M., and B. W. Matthews (1984). "Intrahelical hydrogen bonding of serine, threonine and cysteine residues within alpha-helices and its relevance to membrane-bound proteins" J Mol Biol **175**(1): 75-81.

Gregoret, L. M., and F. E. Cohen (1990). "Novel method for the rapid evaluation of packing in protein structures" J Mol Biol **211**(4): 959-74.

Guan, X. M., S. J. Peroutka, and B. K. Kobilka (1992). "Identification of a single amino acid residue responsible for the binding of a class of beta-adrenergic receptor antagonists to 5-hydroxytryptamine1A receptors" Mol Pharmacol **41**(4): 695-8.

Guarnieri, F. (1997). "Montecarlo simulations on the conformation of GnRH studied in a mixed water-membrane solvent model" manuscript in preparation

Guarnieri, F., and H. Weinstein (1996). "Conformational memories and the exploration of biologically relevant peptide conformations: An illustration for the gonadotropin-releasing hormone" J. Amer. Chem. Soc. **118**:5580-5589.

Han, M., B. S. DeDecker, and S. O. Smith (1993). "Localization of the retinal protonated Schiff base counterion in rhodopsin" Biophys J **65**(2): 899-906.

Han, M., S. W. Lin, M. Minkova, S. O. Smith, and T. P. Sakmar (1996). "Functional interaction of transmembrane helices 3 and 6 in rhodopsin. Replacement of phenylalanine 261 by alanine causes reversion of phenotype of a glycine 121 replacement mutant" J Biol Chem **271**(50): 32337-42.

Hebert, T. E., S. Moffett, J. P. Morello, T. P. Loisel, D. G. Bichet, C. Barret, and M. Bouvier (1996). "A peptide derived from a beta2-adrenergic receptor transmembrane domain inhibits both receptor dimerization and activation" J-Biol-Chem **271**(27): 16384-92.

Henderson, R., J. M. Baldwin, T. A. Ceska, F. Zemlin, E. Beckmann, and K. H. Downing (1990). "Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy" J Mol Biol **213**(4): 899-929.

Henderson, R., and G. F. Schertler (1990). "The structure of bacteriorhodopsin and its relevance to the visual opsins and other seven-helix G-protein coupled receptors" Philos Trans R Soc Lond [Biol] **326**(1236): 379-89.

Henry, G. D., and B. D. Sykes (1990). "Structure and dynamics of detergent-solubilized M13 coat protein (an integral membrane protein) determined by 13C and 15N nuclear magnetic resonance spectroscopy" Biochem Cell Biol **68**(1): 318-29.

Herzyk, P., and R. E. Hubbard (1995). "Automated method for modeling seven-helix transmembrane receptors from experimental data [see comments]" Biophys J **69**(6): 2419-42.

Hibert, M. F., S. Trumpp-Kallmeyer, A. Bruinvels, and J. Hoflack (1991). "Three-dimensional models of neurotransmitter G-binding protein-coupled receptors" Mol Pharmacol **40**(1): 8-15.

Hibert, M. F., S. Trumpp-Kallmeyer, J. Hoflack, and A. Bruinvels (1993). "This is not a G protein-coupled receptor" Trends Pharmacol Sci **14**(1): 7-12.

Hilbert, M., G. Bohm, and R. Jaenicke (1993). "Structural relationships of homologous proteins as a fundamental principle in homology modeling" Proteins: Struct. Fun. Gen. **17**:138-151.

Hirst, D. M., 1990, A computational approach to chemistry: London, Blackwell Scientific Publications.

Ho, B. Y., A. Karschin, T. Branchek, N. Davidson, and H. A. Lester (1992). "The role of conserved aspartate and serine residues in ligand binding and in

function of the 5-HT1A receptor: a site-directed mutation study" Febs Lett **312**(2-3): 259-62.

Hoflack, J., M. F. Hibert, T.-K. S., and J. M. Bidart (1993a). "Three-dimensional models of gonado-thyrotropin hormone receptor transmembrane domain" Drug Des Discov **10**(2): 157-71.

Hoflack, J., S. Trumpp-Kallmeyer, and M. Hibert, 1993b, Molecular modeling of G-protein coupled receptors, *in* H. Kubinyi, ed., 3D QSAR and drug design: theory, methods, and applications, New York, ESCOM.

Hoflack, J., S. Trumpp-Kallmeyer, and M. Hibert (1994). "Re-evaluation of bacteriorhodopsin as a model for G protein-coupled receptors" Trends Pharmacol Sci **15**:7-9.

Huggins, J. P., S. Trumpp-Kallmeyer, M. F. Hibert, J. M. Hoflack, B. O. Fanger, and C. R. Jones (1993). "Modelling and modification of the binding site of endothelin and other receptors" Eur J Pharmacol **245**(3): 203-14.

Humblet, C., and T. Mirzadegan, 1992, Three-dimensional models of G-protein coupled receptors, *in* M. C. Venuti, and J. A. Bristol, eds., Ann. Rep. Med. Chem., San Diego, Academic Press, p. 291-300.

Hwa, J., R. Gaivin, J. E. Porter, and D. M. Perez (1997). "Synergism of constitutive activity in alpha 1-adrenergic receptor activation" Biochemistry **36**(3): 633-9.

IJzerman, A. P., G. P. J. Van, and K. A. Jacobson (1992). "Molecular modeling of adenosine receptors. I. The ligand binding site on the A1 receptor" Drug Des Discov **9**(1): 49-67.

Ismaiel, A. M., M. Titeler, K. J. Miller, T. S. Smith, and R. A. Glennon (1990). "5-HT1 and 5-HT2 binding profiles of the serotonergic agents alpha-methylserotonin and 2-methylserotonin" J Med Chem **33**(2): 755-8.

Jahnig, F., and O. Edholm (1990). "Can the structure of proteins be calculated?" Z. Phys. B - Condensed Matter **78**:137-143.

Jahnig, F., and O. Edholm (1992). "Modeling of the structure of bacteriorhodopsin. A molecular dynamics study" J Mol Biol **226**(3): 837-50.

Javitch, J. A., D. Fu, J. Chen, and A. Karlin (1995a). "Mapping the binding-site crevice of the dopamine D2 receptor by the substituted-cysteine accessiblity method" Neuron **14**:825-831.

Javitch, J. A., D. Y. Fu, and J. Y. Chen (1995b). "Residues in the fifth membrane-spanning segment of the dopamine D2 receptor exposed in the binding-site crevice" Biochemistry **34**(50): 16433-16439.

Julius, D., A. B. MacDermott, R. Axel, and T. M. Jessell (1988). "Molecular characterization of a functional cDNA encoding the serotonin 1c receptor" Science **241**(4865): 558-64.

Jung, H., R. Windhaber, D. Palm, and K. D. Schnackerz (1995). "NMR and circular dichroism studies of synthetic peptides derived from the third intracellular loop of the beta-adrenoceptor" FEBS Lett **358**(2): 133-6.

Jung, H., R. Windhaber, D. Palm, and K. D. Schnackerz (1996). "Conformation of a beta-adrenoceptor-derived signal transducing peptide as inferred by circular dichroism and 1H NMR spectroscopy" Biochemistry **35**(20): 6399-405.

Juretic, D., B. Lee, N. Trinajstic, and R. W. Williams (1993). "Conformational preference functions for predicting helices in membrane proteins" Biopolymers **33**(2): 255-73.

Kao, H. T., N. Adham, M. A. Olsen, R. L. Weinshank, T. A. Branchek, and P. R. Hartig (1992). "Site-directed mutagenesis of a single residue changes the binding properties of the serotonin 5-HT2 receptor from a human to a rat pharmacology" Febs Lett **307**(3): 324-8.

Kjelsberg, M. A., S. Cotecchia, J. Ostrowski, M. G. Caron, and R. J. Lefkowitz (1992). "Constitutive activation of the alpha 1B-adrenergic receptor by all amino acid substitutions at a single site. Evidence for a region which constrains receptor activation" J Biol Chem **267**(3): 1430-3.

Kobilka, B. (1992). "Adrenergic receptors as models for G protein-coupled receptors" Annu Rev Neurosci **15**:87-114.

Komiya, H., T. O. Yeates, D. C. Rees, J. P. Allen, and G. Feher (1988). "Structure of the reaction center from Rhodobacter sphaeroides R-26 and 2.4.1: symmetry relations and sequence comparisons between different species" Proc Natl Acad Sci U S A **85**(23): 9012-6.

Kono, M., H. Yu, and D. D. Oprian (1996). "Interactions between helix 5 and helix 6 of rhodopsin using Cysteine-scanning mutagenesis of split receptor mutants" Biopys. J. **70**(2): 395.

Kontoyianni, M., and T. P. Lybrand (1993a). "Computer modeling studies of G

protein coupled receptors" Med. Chem. Res. 3:407-418.

Kontoyianni, M., and T. P. Lybrand (1993b). "Three-Dimensional models for integral membrane proteins: possibilities and pitfalls" Persp. Drug Disc. Design 1:291-300.

Konvicka, K., J. A. Ballesteros, and H. Weinstein, 1995, A Structural and Functional Role for the Conserrved NP/DP Motif in Helix 7 of GPCR Inferred from a 3D Model of the GNRH Receptor: Biophysical Society Meeting, p. TH-461.

Kyte, J., and R. F. Doolittle (1982). "A simple method for displaying the hydropathic character of a protein" J Mol Biol **157**(1): 105-32.

Lefkowitz, R. J., and M. G. Caron (1988). "Adrenergic receptors. Models for the study of receptors coupled to guanine nucleotide regulatory proteins" J Biol Chem **263**(11): 4993-6.

Lefkowitz, R. J., S. Cotecchia, P. Samama, and T. Costa (1993). "Constitutive activity of receptors coupled to guanine nucleotide regulatory proteins" Trends Pharmacol Sci **14**(8): 303-7.

Lemmon, M. A., J. M. Flanagan, H. R. Treutlein, J. Zhang, and D. M. Engelman (1992). "Sequence specificity in the dimerization of transmembrane alpha-helices" Biochemistry **31**(51): 12719-25.

Lesk, A. M., and D. R. Boswell (1992). "Homology modelling: inferences from tables of aligned sequences" Curr. Opin. Struc. Biol. **2**:242-247.

Lewell, X. Q. (1992). "A model of the adrenergic beta-2 receptor and binding sites for agonist and antagonist" Drug Des Discov **9**(1): 29-48.

Lin, S. W., and T. P. Sakmar (1996). "Specific tryptophan UV-absorbance changes are probes of the transition of rhodopsin to its active state" Biochemistry **35**(34): 11149-59.

Link, R., D. Daunt, G. Barsh, A. Chruscinski, and B. Kobilka (1992). "Cloning of two mouse genes encoding alpha 2-adrenergic receptor subtypes and identification of a single amino acid in the mouse alpha 2-C10 homolog responsible for an interspecies variation in antagonist binding" Mol Pharmacol **42**(1): 16-27.

Liu, J., T. Schoneberg, M. van Rhee, and J. Wess (1995). "Mutational analysis of the relative orientation of transmembrane helices I and VII in G protein-

coupled receptors" J Biol Chem **270**(33): 19532-9.

Livingstone, C. D., P. G. Strange, and L. H. Naylor (1992). "Molecular modelling of D2-like dopamine receptors" Biochem J **287**:277-82.

Loisel, T. P., L. Adam, T. E. Hebert, and M. Bouvier (1996). "Agonist stimulation increases the turnover rate of beta 2AR-bound palmitate and promotes receptor depalmitoylation" Biochemistry **35**(49): 15923-32.

Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall (1951). "Protein measurement with the Folin phenol reagent" J. Biol. Chem. **193**:265-275.

Luo, X., D. Zhang, and H. Weinstein (1993). "Rigid domain motion of proteins: A new approach to the analysis of protein dynamics" Prot. Sci. **2,S1**:158.

Luxembourg, A., M. Hekman, and E. M. Ross (1991). "Immunologic mapping of the amino- and carboxy-termini of the turkey erythrocyte beta-adrenergic receptor: selective proteolysis of both domains" Febs Lett **283**(1): 155-8.

MacArthur, M. W., and J. M. Thornton (1991). "Influence of proline residues on protein conformation" J Mol Biol **218**(2): 397-412.

MacKerell Jr, A. D., J. Wiorkiewicz-Kuczera, and M. Karplus (1995). "An all-atom empirical energy function for the simulation of nucleic acids" J. Am. Chem. Soc. **117**:11946-75.

MaloneyHuss, K., and T. P. Lybrand (1992). "Three-dimensional structure for the beta 2 adrenergic receptor protein based on computer modeling studies" J Mol Biol **225**(3): 859-71.

Mansour, A., F. Meng, W. J. H. Meador, L. P. Taylor, O. Civelli, and H. Akil (1992). "Site-directed mutagenesis of the human dopamine D2 receptor" Eur J Pharmacol **227**(2): 205-14.

Matsui, H., R. J. Lefkowitz, M. G. Caron, and J. W. Regan (1989). "Localization of the fourth membrane spanning domain as a ligand binding site in the human platelet alpha 2-adrenergic receptor" Biochemistry **28**(9): 4125-30.

McCammon, J. A. (1987). "Computer-aided molecular design" Science **238**:486-491.

McGregor, M. J., S. A. Islam, and M. J. Sternberg (1987). "Analysis of the relationship between side-chain conformation and secondary structure in globular proteins" J Mol Biol **198**(2): 295-310.

Metcalf, M. A., R. W. McGuffin, and M. W. Hamblin (1992). "Conversion of the human 5-HT1D beta serotonin receptor to the rat 5-HT1B ligand-binding phenotype by Thr355Asn site directed mutagenesis" Biochem Pharmacol **44**(10): 1917-20.

Min, K. C., T. A. Zvyaga, A. M. Cypess, and T. P. Sakmar (1993). "Characterization of mutant rhodopsins responsible for autosomal dominant retinitis pigmentosa. Mutations on the cytoplasmic surface affect transducin activation" J Biol Chem **268**(13): 9400-4.

Mizobe, T., M. Mazw, V. Lam, S. Suryanarayana, and B. K. Kobilka (1996). "Arrangement of transmembrane domains in adrenergic receptors. Similarity to bacteriorhodopsin" J. Biol. Chem. **271**:2387-2389.

Moench, S. J., J. Moreland, D. H. Stewart, and T. G. Dewey (1994). "Fluorescence studies of the location and membrane accessibility of the palmitoylation sites of rhodopsin" Biochemistry **33**(19): 5791-6.

Munson, P. J., and D. Rodbard (1980). "Ligand: a versatile computerized approach for characterization of ligand-binding systems" Anal Biochem **107**(1): 220-39.

Nordvall, G., and U. Hacksell (1993). "Binding-site modeling of the muscarinic m1 receptor: a combination of homology-based and indirect approaches" J Med Chem **36**(8): 967-76.

O'Dowd, B. F., M. Hnatowich, M. G. Caron, R. J. Lefkowitz, and M. Bouvier (1989). "Palmitoylation of the human beta 2-adrenergic receptor. Mutation of Cys341 in the carboxyl tail leads to an uncoupled nonpalmitoylated form of the receptor" J Biol Chem **264**(13): 7564-9.

Oksenberg, D., S. A. Marsters, B. F. O'Dowd, S. Havlik, S. J. Peroutka, and A. Ashkenazi (1992). "A single amino-acid difference confers major pharmacological variation between human and rodent 5-HT1B receptors" Nature **360**:161-163.

Oliveira, L., A. C. M. Paiva, C. Sander, and G. Vriend (1994). "A common step for signal transduction in G-protein coupled receptors" TIPS **15**:170-172.

Onaran, H. O., T. Costa, and D. Rodbard (1993). "Beta gamma subunits of guanine nucleotide-binding proteins and regulation of spontaneous receptor activity: thermodynamic model for the interaction between receptors and guanine nucleotide-binding protein subunits" Mol Pharmacol **43**(2): 245-56.

Oprian, D. D. (1992). "The ligand-binding domain of rhodopsin and other G protein-linked receptors" J Bioenerg Biomembr **24**(2): 211-7.

Overington, J., D. Donnelly, M. S. Johnson, A. Sali, and T. L. Blundell (1992). "Environment-specific amino acid substitution tables: Tertiary templates and prediction of protein folds" Protein Sci. **1**:216-226.

Overington, J., M. S. Johnson, A. Sali, and T. L. Blundell (1990). "Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction" Proc R Soc Lond [Biol] **241**(1301): 132-45.

Papac, D. I., K. R. Thornburg, E. E. Bullesbach, R. K. Crouch, and D. R. Knapp (1992). "Palmitylation of a G-protein coupled receptor. Direct analysis by tandem mass spectrometry" J Biol Chem **267**(24): 16889-94.

Pardo, L., J. A. Ballesteros, R. Osman, and H. Weinstein (1992). "On the use of the transmembrane domain of bacteriorhodopsin as a template for modeling the three-dimensional structure of guanine nucleotide-binding regulatory protein-coupled receptors" Proc Natl Acad Sci U S A **89**(9): 4009-12.

Pascarella, S., and P. Argos (1992). "Analysis of insertions/deletions in protein structures" J Mol Biol **224**(2): 461-71.

Pastor, N., D. Pinero, A. M. Valdes, and X. Soberon (1990). "Molecular evolution of class A beta-lactamases: phylogeny and patterns of sequence conservation" Mol Microbiol **4**(11): 1957-65.

Pastore, A., T. S. Harvey, C. E. Dempsey, and I. D. Campbell (1989). "The dynamic properties of melittin in solution. Investigations by NMR and molecular dynamics" Eur Biophys J **16**(6): 363-7.

Pei, G., P. Samama, M. Lohse, M. Wang, J. Codina, and R. J. Lefkowitz (1994). "A constitutively active mutant beta 2-adrenergic receptor is constitutively desensitized and phosphorylated" Proc Natl Acad Sci U S A **91**(7): 2699-702.

Perez, D. M., J. Hwa, R. Gaivin, M. Mathur, F. Brown, and R. M. Graham (1996). "Constitutive activation of a single effector pathway: evidence for multiple activation states of a G protein-coupled receptor" Mol-Pharmacol **49**(1): 112-22.

Perlman, J. H., A. O. Colson, W. Wang, K. Bence, R. Osman, and M. C. Gershengorn (1997). "Interactions between conserved residues in transmembrane helices 1, 2, and 7 of the thyrotropin-releasing hormone

receptor" J Biol Chem **272**(18): 11937-42.

Piela, L., G. Nemethy, and H. A. Scheraga (1987a). "Conformational constraints of amino acid side chains in alpha-helices" Biopolymers **26**(8): 1273-86.
Piela, L., G. Nemethy, and H. A. Scheraga (1987b). "Proline-induced constraints in alpha-helices" Biopolymers **26**(9): 1587-600.

Pittel, Z., and J. Wess (1994). "Intramolecular interactions in muscarinic acetylcholine receptors studied with chimeric m2/m5 receptors" Mol Pharmacol **45**(1): 61-4.

Pogozheva, I. D., A. L. Lomize, and H. I. Mosberg (1997). "The transmembrane 7-alpha-bundle of rhodopsin: distance geometry calculations with hydrogen bonding constraints" Biophys J **72**(5): 1963-85.

Pollock, N. J., A. M. Manelli, C. W. Hutchins, M. E. Steffey, R. G. MacKenzie, and D. E. Frail (1992). "Serine mutations in transmembrane V of the dopamine D1 receptor affect ligand interactions and receptor activation" J Biol Chem **267**(25): 17780-6.

Ponder, J. W., and F. M. Richards (1987). "Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes" J Mol Biol **193**(4): 775-91.

Probst, W. C., L. A. Snyder, D. I. Schuster, J. Brosius, and S. C. Sealfon (1992). "Sequence alignment of the G-protein coupled receptor superfamily" Dna Cell Biol **11**(1): 1-20.

Rao, V. R., G. B. Cohen, and D. D. Oprian (1994). "Rhodopsin mutation G90D and a molecular mechanism for congenital night blindness" Nature **367**:639-642.

Rath, P., L. L. DeCaluwe, P. H. Bovee-Geurts, W. J. DeGrip, and K. J. Rothschild (1993). "Fourier transform infrared difference spectroscopy of rhodopsin mutants: light activation of rhodopsin causes hydrogen-bonding change in residue aspartic acid-83 during meta II formation" Biochemistry **32**(39): 10277-82.

Raymond, J. R., M. Hnatowich, R. J. Lefkowitz, and M. G. Caron (1990). "Adrenergic receptors. Models for regulation of signal transduction processes" Hypertension **15**(2): 119-31.

Rees, D. C., L. DeAntonio, and D. Eisenberg (1989a). "Hydrophobic

organization of membrane proteins" Science **245**(4917): 510-3.

Rees, D. C., H. Komiya, T. O. Yeates, J. P. Allen, and G. Feher (1989b). "The bacterial photosynthetic reaction center as a model for membrane proteins" Annu Rev Biochem **58**:607-33.

Rost, B., and C. Sander (1993). "Prediction of protein secondary structure at better than 70% accuracy" J Mol Biol **232**(2): 584-99.

Roux, M., J. M. Neumann, R. S. Hodges, P. F. Devaux, and M. Bloom (1989). "Conformational changes of phospholipid headgroups induced by a cationic integral membrane peptide as seen by deuterium magnetic resonance" Biochemistry **28**(5): 2313-21.

Samama, P., S. Cotecchia, T. Costa, and R. J. Lefkowitz (1993). "A mutation-induced activated state of the beta 2-adrenergic receptor. Extending the ternary complex model" J Biol Chem **268**(7): 4625-36.

Sankararamakrishnan, R., and S. Vishveshwara (1990). "Conformational studies on peptides with proline in the right-handed alpha-helical region" Biopolymers **30**(3-4): 287-98.

Sankararamakrishnan, R., and S. Vishveshwara (1992). "Geometry of proline-containing alpha-helices in proteins" Int J Pept Protein Res **39**(4): 356-63.

Sankararamakrishnan, R., and S. Vishveshwara (1993). "Characterization of Proline-containing alpha-helix (helix F model of bacteriorhodopsin) by molecular dynamics studies" Proteins: Structure, Function, and Genetics **15**:26-41.

Sansom, M. S. (1992). "Proline residues in transmembrane helices of channel and transport proteins: a molecular modelling study" Protein Eng **5**(1): 53-60.

Savarese, T. M., and C. M. Fraser (1992). "In vitro mutagenesis and the search for structure-function relationships among G protein-coupled receptors" Biochem J **283**:1-19.

Savarese, T. M., C. D. Wang, and C. M. Fraser (1992). "Site-directed mutagenesis of the rat m1 muscarinic acetylcholine receptor. Role of conserved cysteines in receptor function" J Biol Chem **267**(16): 11439-48.

Scheer, A., F. Fanelli, T. Costa, P. G. De Benedetti, and S. Cotecchia (1996). "Constitutively active mutants of the alpha 1B-adrenergic receptor: role of highly conserved polar amino acids in receptor activation" EMBO-J **15**(14):

3566-78.

Scheer, A., F. Fanelli, T. Costa, P. G. De Benedetti, and S. Cotecchia (1997). "The activation process of the alpha-1B adrenergic receptor: Potential role of protonation and hydrophobicity of a highly conserved aspartate." submitted

Schertler, G. F., and P. A. Hargrave (1995). "Projection structure of frog rhodopsin in two crystal forms" Proc Natl Acad Sci U S A **92**(25): 11578-82.

Schertler, G. F., C. Villa, and R. Henderson (1993). "Projection structure of rhodopsin" Nature **362**(6422): 770-2.

Sealfon, S. C., L. Chi, B. J. Ebersole, V. Rodic, D. Zhang, J. A. Ballesteros, and H. Weinstein (1995). "Related contributions of specific helix 2 and helix 7 residues to conformational activation of the serotonin 5-HT_{2A} receptor" J. Biol. Chem. **270**:16683-8.

Sealfon, S. C., W. Zhou, N. Almaula, and V. Rodic (1996). "Cloning and site directed mutagenesis studies of gonadotropin-releasing hormone receptor" Methods in Neurosciences **29**:143-196.

Sheikh, S. P., T. A. Zvyaga, O. Lichtarge, T. P. Sakmar, and H. R. Bourne (1996). "Rhodopsin activation blocked by metal-ion-binding sites linking transmembrane helices C and F" Nature **383**(6598): 347-50.

Shnyrov, V. L., and A. L. Berman (1988). "Calorimetric study of thermal denaturation of vertebrate visual pigments" Biomed Biochim Acta **47**(4-5): 355-62.

Sikora, S., A. S. Little, and T. G. Dewey (1994). "Room temperature trapping of rhodopsin photointermediates" Biochemistry **33**(15): 4454-9.

Smolyar, A., and R. Osman (1993). "Role of threonine 342 in helix 7 of the 5-hydroxytryptamine type 1D receptor in ligand binding: an indirect mechanism for receptor selectivity" Mol Pharmacol **44**(4): 882-5.

Strader, C. D., M. R. Candelore, W. S. Hill, I. S. Sigal, and R. A. Dixon (1989). "Identification of two serine residues involved in agonist activation of the beta-adrenergic receptor" J Biol Chem **264**(23): 13572-8.

Strassburger, J. M., W. Gartner, and S. E. Braslavsky (1997). "Volume and enthalpy changes after photoexcitation of bovine rhodopsin: laser-induced optoacoustic studies" Biophys J **72**(5): 2294-303.

Strehlow, K. G., A. D. Robertson, and R. L. Baldwin (1991). "Proline for alanine substitutions in the C-peptide helix of ribonuclease A" Biochemistry **30**(23): 5810-4.

Summers, N. L., and M. Karplus (1991). "Modeling of side chains, loops, and insertions in proteins" Methods Enzymol **202**:156-204.

Surprenant, A., D. A. Horstman, H. Akbarali, and L. E. Limbird (1992). "A point mutation of the alpha 2-adrenoceptor that blocks coupling to potassium but not calcium currents" Science **257**(5072): 977-80.

Suryanarayana, S., D. A. Daunt, Z. M. Von, and B. K. Kobilka (1991a). "A point mutation in the seventh hydrophobic domain of the alpha 2 adrenergic receptor increases its affinity for a family of beta receptor antagonists" J Biol Chem **266**(23): 15488-92.

Suryanarayana, S., D. A. Daunt, Z. M. Von, and B. K. Kobilka (1991b). "A single point mutation in the seventh hydrophobic domain of the alpha 2-adrenergic receptor changes antagonist binding specificity to that of a beta receptor" Trans Assoc Am Physicians **104**:62-8.

Suryanarayana, S., and B. K. Kobilka (1993). "Amino acid substitutions at position 312 in the seventh hydrophobic segment of the beta 2-adrenergic receptor modify ligand-binding specificity" Mol Pharmacol **44**(1): 111-4.

Suryanarayana, S., M. von Zastrow, and B. K. Kobilka (1992). "Identification of intramolecular interactions in adrenergic receptors" J Biol Chem **267**(31): 21991-4.

Svensson, B., I. Vass, and S. Styring (1991). "Sequence analysis of the D1 and D2 reaction center proteins of photosystem II" Z Naturforsch [c] **46**(9-10): 765-76.

Taylor, E. W., and A. Agarwal (1993). "Sequence homology between bacteriorhodopsin and G-protein coupled receptors: exon shuffling or evolution by duplication?" Febs Lett **325**(3): 161-6.

Thornton, J. M., T. P. Flores, D. T. Jones, and M. B. Swindells (1991). "Protein structure. Prediction of progress at last [news]" Nature **354**(6349): 105-6.

Tomic, M., P. Seeman, S. R. George, and B. F. ODowd (1993). "Dopamine D1 receptor mutagenesis: role of amino acids in agonist and antagonist binding" Biochem Biophys Res Commun **191**(3): 1020-7.

Tota, M. R., and C. D. Strader (1990). "Characterization of the binding domain of the beta-adrenergic receptor with the fluorescent antagonist carazolol. Evidence for a buried ligand binding site" J Biol Chem **265**(28): 16891-7.

Travers, P., T. L. Blundell, M. J. Sternberg, and W. F. Bodmer (1984). "Structural and evolutionary analysis of HLA-D-region products" Nature **310**(5974): 235-8.

Traxler, K. W., and T. G. Dewey (1994). "Effects of depalmitoylation on physicochemical properties of rhodopsin" Biochemistry **33**(7): 1718-23.

Trumpp-Kallmeyer, S., J. Hoflack, A. Bruinvels, and M. Hibert (1992). "Modeling of G-protein-coupled receptors: application to dopamine, adrenaline, serotonin, acetylcholine, and mammalian opsin receptors" J Med Chem **35**(19): 3448-62.

Trumpp-Kallmeyer, S., J. Hoflack, and M. Hibert, 1993, Modeling of G-protein coupled receptors: application to the NK1 receptor, *in* S. H. Buck, ed., The Tachykinin Receptors, New York, Humana Press.

Tsutsumi, M., W. Zhou, R. P. Millar, P. L. Mellon, J. L. Roberts, C. A. Flanagan, K. Dong, B. Gillo, and S. C. Sealfon (1992). "Cloning and functional expression of a mouse gonadotropin-releasing hormone receptor" Mol Endocrinol **6**(7): 1163-9.

Turcatti, G., K. Nemeth, M. D. Edgerton, U. Meseth, F. Talabot, M. Peitsch, J. Knowles, H. Vogel, and A. Chollet (1996). "Probing the structure and function of the tachykinin neurokinin-2 receptor through biosynthetic incorporation of fluorescent amino acids at specific sites" J Biol Chem **271**(33): 19991-8.

Unger, V. M., and G. F. Schertler (1995). "Low resolution structure of bovine rhodopsin determined by electron cryo-microscopy" Biophys J **68**(5): 1776-86.

van Gunsteren, W. F., R. M. Brunne, A. E. Mark, and S. P. van Helden, 1992, Computer simulation of biomolecules: Comparison with experimental data, *in* J. Bertran, ed., Molecular Aspects of Biotechnology: Computational Models and Theories, Dordrecht, The Netherlands, Kluwer Academic Publishers, p. 105-122.

van Gunsteren, W. F., and P. K. Weiner, 1989, Computer Simulation of Biomolecular Systems, Leiden, The Netherlands, ESCOM, p. 224.

van Rhee, A. M., and K. A. Jacobson (1996). "Molecular Architecture of G Protein-Coupled Receptors" Drug Devel. Res **37**:1-38.

Vernier, P., H. Philippe, P. Samama, and J. Mallet (1993). "Bioamine receptors: evolutionary and functional variations of a structural leitmotiv" Exs 63:297-337.

Ward, W. H., D. Timms, and A. R. Fersht (1990). "Protein engineering and the study of structure--function relationships in receptors" Trends Pharmacol Sci 11(7): 280-4.

Weinstein, H., 1992, Computational simulations of molecular structure, dynamics and signal transduction in biological systems: Mechanistic implications for ecological physical chemistry: Second International Workshop on Ecological Physical Chemistry, p. 1-16.

Weinstein, H., and J. P. Green, 1981, Quantum Chemistry in Biomedical Sciences, Ann. N.Y. Acad. Sci., p. 552.

Weinstein, H., J. P. Green, R. Osman, and W. D. Edwards, 1978, Recognition and activation mechanisms on the LSD/serotonin receptor: the molecular basis of structure activity relationships, *in* G. Barnett, M. Trsic, and R. E. Willette, eds., QuaSAR, Quantitative Structure Activity Relationships of Analgesics, Narcotic Antagonists, and Hallucinogens, NIDA Monograph 22, U.S. Government Printing Office, p. 333-358.

Weinstein, H., and E. L. Mehler, 1992, Structural specificity in the engineering of biological function: Insights from the dynamics of calmodulin, *in* J. Bertran, ed., Molecular Aspects of Biotechnology: Computational Models and Theories, Dordrecht, The Netherlands, Kluwer Academic Publishers, p. 153-173.

Weiss, M. S., U. Abele, J. Weckesser, W. Welte, E. Schiltz, and G. E. Schulz (1991). "Molecular architecture and electrostatic properties of a bacterial porin" Science 254(5038): 1627-30.

Westkaemper, R. B., and R. A. Glennon (1991). "Approaches to molecular modeling studies and specific application to serotonin ligands and receptors" Pharmacol Biochem Behav 40(4): 1019-31.

Wieland, K., H. M. Zuurmond, C. Krasel, A. P. Ijzerman, and M. J. Lohse (1996). "Involvement of Asn-293 in stereospecific agonist recognition and in activation of the beta 2-adrenergic receptor" Proc-Natl-Acad-Sci-U-S-A 93(17): 9276-81.

Williams, K. A., and C. M. Deber (1991). "Proline residues in transmembrane

helices: structural or dynamic role?" Biochemistry **30**(37): 8919-23.

Wong, S. K., C. Slaughter, A. E. Ruoho, and E. M. Ross (1988). "The catecholamine binding site of the beta-adrenergic receptor is formed by juxtaposed membrane-spanning domains" J Biol Chem **263**(17): 7925-8.

Woolfson, D. N., S. R. J. Mortishire, and D. H. Williams (1991). "Conserved positioning of proline residues in membrane-spanning helices of ion-channel proteins" Biochem Biophys Res Commun **175**(3): 733-7.

Woolfson, D. N., and D. H. Williams (1990). "The influence of proline residues on alpha-helical structure" Febs Lett **277**(1-2): 185-8.

Yamamoto, Y., K. Kamiya, and S. Terao (1993). "Modeling of human thromboxane A2 receptor and analysis of the receptor-ligand interaction" J Med Chem **36**(7): 820-5.

Yang, K., D. L. Farrens, C. Altenbach, Z. T. Farahbakhsh, W. L. Hubbell, and H. G. Khorana (1996a). "Structure and function in rhodopsin. Cysteines 65 and 316 are in proximity in a rhodopsin mutant as indicated by disulfide formation and interactions between attached spin labels" Biochemistry **35**(45): 14040-6.

Yang, K., D. L. Farrens, W. L. Hubbell, and H. G. Khorana (1996b). "Structure and function in rhodopsin. Single cysteine substitution mutants in the cytoplasmic interhelical E-F loop region show position-specific effects in transducin activation" Biochemistry **35**(38): 12464-9.

Yeagle, P. L. (1982). "31P nuclear magnetic resonance studies of the phospholipid-protein interface in cell membranes" Biophys J **37**(1): 227-39.

Yeagle, P. L., J. L. Alderfer, and A. D. Albert (1995a). "Structure of the carboxy-terminal domain of bovine rhodopsin [letter]" Nat-Struct-Biol **2**(10): 832-4.

Yeagle, P. L., J. L. Alderfer, and A. D. Albert (1995b). "Structure of the third cytoplasmic loop of bovine rhodopsin" Biochemistry **34**(45): 14621-5.

Yeagle, P. L., J. L. Alderfer, A. C. Salloum, L. Ali, and A. D. Albert (1997). "The first and second cytoplasmic loops of the G-protein receptor, rhodopsin, independently form beta-turns" Biochemistry **36**(13): 3864-9.

Yu, H., M. Kono, T. D. McKee, and D. D. Oprian (1995). "A general method for mapping tertiary contacts between amino acid residues in membrane-embedded proteins" Biochemistry **34**(46): 14963-9.

Yun, R. H., A. Anderson, and J. Hermans (1991). "Proline in alpha-helix: stability and conformation studied by dynamics simulation" Proteins **10**(3): 219-28.

Zhang, D., and H. Weinstein (1993a). "Ligand selectivity and the molecular properties of the 5-HT2 receptor: computational simulations reveal a major role for transmembrane helix 7" Med. Chem. Res. **3**:357-369.

Zhang, D., and H. Weinstein (1993b). "Signal transduction by a 5-HT2 receptor: a mechanistic hypothesis from molecular dynamics simulations of the three-dimensional model of the receptor complexed to ligands" J. med. Chem. **36**:934.

Zhang, D., and H. Weinstein (1994). "Polarity conserved positions in transmembrane domains of G-protein coupled receptors and bacteriorhodopsin" Febs Lett. **337**:207-212.

Zhou, W., C. Flanagan, J. A. Ballesteros, K. Konvicka, J. S. Davidson, H. Weinstein, R. P. Millar, and S. C. Sealfon (1994). "A reciprocal mutation supports helix 2 and helix 7 proximity in the Gonadotropin-releasing hormone receptor" Mol. Pharm. **45**:165-170.

Zhou, W., V. Rodic, S. Kitanovic, C. A. Flanagan , L. Chi, H. Weinstein, S. Maayani, R. P. Millar, and S. C. Sealfon (1995). "A locus of the gonadotropin-releasing hormone receptor that differentiates agonist and antagonist binding sites" J. Biol Chem. **370**:18853-18857.