

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

A

**COMPUTATIONAL STUDIES OF STRUCTURE-FUNCTION
RELATIONSHIPS IN THE BACTERIOPHAGE LAMBDA
REPRESSOR**

by

DAVID CALEB KOMBO

**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**

1996

UMI Number: 9707117

UMI Microform 9707117
Copyright 1996, by UMI Company. All rights reserved.

**This microform edition is protected against unauthorized
copying under Title 17, United States Code.**

UMI
300 North Zeeb Road
Ann Arbor, MI 48103

**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.**

April 25, 1996
Date

Craig T. Benham
Chair of Examining Committee

April 25, 1996
Date

[Signature]
Executive Officer

George Némethy

Shalom Rackovsky

Harold A. Scheraga

Supervisory Committee

The City University of New York

***Dedicated to
the late professor George Némethy;
to his wife, Judith Némethy; to their children,
Maria and Kinga; and to his brother,
professor Peter Némethy***

Acknowledgements

I wish to thank the late professor George Némethy for his supervision, friendship and tremendous support, even during the difficult time of his long illness. I also wish to acknowledge his collaborators, Professors Shalom Rackovsky and Harold A. Scheraga of Cornell University, who accepted to assure the supervision of my thesis after the decease of George. The three professors have made invaluable contributions to my training and the completion of the work described herein. I also wish to thank the following professors; Harel Weinstein, J.B.Alexander Ross, Massimo Sassaroli , Istvan Sugar and Ronald Kohanski who have made particularly helpful suggestions and comments both in the design and maturation of this project.

I am grateful to the Dean of Graduate School, professor Terry Ann Krulwich, and the acting Chairman of the Biomathematical Sciences Department, professor Craig Benham, for their moral support throughout my graduate training. Special thanks are due to my lifetime friend Severin Kamosi Mabola for his unconditional friendship, support, encouragement and assistance during my High School and College years. I am deeply indebted to USIA (The United States Information Agency) for financial support through a predoctoral Fulbright Scholarship.

TABLE OF CONTENTS

<u>General Introduction</u>	1
1. Aims and specific objectives	2
2. Methods	
a. Overview of conformational energy computations	4
b. Related computational studies on mutants of the λ repressor system	12
1. Neglect of the DNA operator	12
2. Inclusion of the DNA operator	18
<u>Chapter 1: Computer-aided discrimination between active and inactive mutants of the bacteriophage lambda repressor</u>	22
Abstract	23
1. Introduction	24
2. Results	30
a. Computational study of the wild-type helix-5-helix-5' complex	31
b. Conformational energy of mutants of the helix-5...helix-5' complex	32
c. Number of pairs of atoms in close contacts	33
d. Interhelical geometry	34
e. Root-mean-square deviation between mutants and wild-type protein	39
f. Computational results for the whole N-terminal domain dimer	40
1. Residues involved in inter-monomer contacts	41
2. Relative geometry of recognition helices	43
3. Discussion	44
4. Materials and Methods	50
a. Fitting of X-ray derived polypeptide chains to standard-geometry models of the N-terminal domain	50
b. Energy minimization and geometrical analysis	52
c. Starting conformations	54
d. Choice of mutants	56

Chapter 2: Effects on protein structure and function of replacing tryptophan with 5-hydroxytryptophan: Single tryptophan mutants of the N-terminal domain of the bacteriophage λ repressor **57**

Abstract	58
1. Introduction	59
2. Methods	62
a. Determination of a standard-geometry model of 5OH-Trp	62
b. Computation of conformational energy of model polypeptides	63
c. Packing of helix-5 and helix-5' of variants of the N-terminal domain of the λ repressor	65
3. Results	67
a. Conformational states of 5-hydroxy Trp	67
b. Conformational constraints of 5-hydroxy-Trp side chain in a right-handed poly(L-Alanine) α -helix	69
c. Computational study of the helix-5-helix-5' complex of variants of the N-terminal domain of the λ repressor with 5OH-Trp substituted for Trp.	70
4. Discussion	72

Chapter 3: Computational analysis of variants of the operator binding domain of the bacteriophage λ repressor **78**

Abstract	79
1. Introduction	80
2. Methods	84
a. Generation of standard-geometry models and energy-minimization	84
b. Determination of geometrical parameters	86
c. Choice of mutants	87
3. Results	88
a. Helical hydrogen-bonding pattern of the fifth helices	88
b. Helicoidal parameters of the fifth helices	92
c. Inter-monomer van der waals	93
1. Involving aromatic rings of Tyr-85 and Tyr-85' in the	

isolated helix-5-helix-5' complexes	93
2. Between the whole monomers of the N-terminal domain of the wild-type repressor protein	95
4. Discussion	98
a. Helical distortions at the dimer interface	98
b. Tolerance for amino acid substitution in the fifth helices	100
Conclusions	104
Tables and Figures	107
a. Chapter 1	108
b. Chapter 2	134
c. Chapter 3	146
References	157

List of Tables

Chapter 1

Table 1. Parameters characterizing the packing structures found after minimizing the energy of the wild-type helix-5 –helix-5' complex of the λ repressor DNA operator binding domain. 109

Tabl 2. Parameters characterizing the packing structures found after minimizing the energy of mutants of the helix-5 –helix-5' complex of the λ repressor DNA operator binding domain. 111

Table 3. Trends in the variations of the intermolecular energy and the experimentally determined dimer dissociation constants. 118

Table 4. Parameters characterizing the packing structures of helices 3 and 3' found after minimizing the energy of mutants of the N-terminal domain of the λ repressor. 120

Table 5. Summary of scores in percentage of criteria for discrimination

between active and inactive mutants of the helix-5-helix-5' complex.	122
--	-----

Chapter 2

Table 1. Minimum-energy conformations of the N-acetyl-N'-methanamide of 5-hydroxytryptophan.	135
--	-----

Table 2. Minimum-energy side-chain conformation in an α -helix.	140
--	-----

Table 3. Parameters characterizing the packing states found after minimizing the energy of mutants of the helix-5-helix-5' complex of the λ repressor DNA operator binding domain.	142
--	-----

Chapter 3

Table 1. Parameters of helices.	147
---------------------------------	-----

Table 2. Van der Waals contacts involving Tyr-85 and Tyr-85'.	149
---	-----

Legends for figures

Chapter 1

Fig.1. Representation of the X-ray crystal structure of the dimeric N-terminal domain of the λ repressor, illustrating the packing of helix-5 (in green) against helix-5' (in yellow). Helices 4 and 4' and the loop connecting them to helices 5 and 5', respectively, are in black. The two helix-turn-helix portions that bind to DNA are illustrated in blue. The rest of the protein is illustrated in magenta. The two strands of the DNA are illustrated in red and orange, respectively. 124

Fig.2. a. X-ray crystal structure of one monomer of the N-terminal domain of the λ repressor. Helix-5 (in green) protrudes outward from the globular domain (in red). 125

Fig.2. b. X-ray crystal structure of the other monomer of the N-terminal domain of the λ repressor. Helix-5' (in yellow) protrudes outward from the globular domain (in red). 126

Fig.3 Intermolecular energy for active mutants vs r.m.s deviation ($\text{\AA}$).

● Active mutants with Ala at position 88. ◇ Active mutants with Ala at position 91. × Other active mutants. 127

Fig.4. Intermolecular energy for active and inactive mutants vs r.m.s. deviation ($\text{\AA}$). ● Active mutants with Ala at position 88. ◇ Active mutants with Ala at position 91. × Other active mutants. × Inactive mutants. 128

Fig.5. Number of atom pairs in close contact for active mutants vs r.m.s. deviation ($\text{\AA}$). ● Active mutants with Ala at position 88. ◇ Active mutants with Ala at position 91. × Other active mutants. 129

Fig.6 Number of atom pairs in close contact for active and inactive mutants vs r.m.s. deviation ($\text{\AA}$). ● Active mutants with Ala at position 88. ◇ Active mutants with Ala at position 91. × Other active mutants. × Inactive mutants. 130

Fig.7 Absolute values of the projected interhelical torsional angle for active and inactive mutants vs r.m.s. deviation ($\text{\AA}$). ◇ Active mutants with Ala at position 88. × Other active mutants. ● Inactive mutants. 131

Fig.8 a). r.m.s. deviation ($\text{\AA}$) of low-energy conformations with respect to the lowest minimum-energy wild-type standard geometry model vs r.m.s. deviation ($\text{\AA}$) with respect to the X-ray crystal conformation of the DNA-bound protein. × Other active mutants. ◇ Active mutants with Ala at position 88. ● Inactive mutants. 132

Fig.8 b) r.m.s. deviation ($\text{\AA}$) of low-energy conformations with respect to the X-ray crystal structure conformation of the DNA-bound protein vs r.m.s. deviation ($\text{\AA}$) of the same conformations with respect to the X-ray crystal structure conformation of the free protein. × Other active mutants. ◇ Active mutants with Ala at position 88. ● Inactive mutants. 133

Chapter 2

Fig.1. Structure of the lowest-energy conformation of the 5-hydroxytryptophan amino acid with acetyl and methyl end groups. The intensity of the black color used in shading the atoms is decreasing in the following order: Nitrogen, oxygen, carbon and hydrogen atoms. 144

Fig.2. Boltzman probability P_i for backbone conformational states ($J=A, C, D, E, F, G, A'$) of the low-energy local minima of the terminally-blocked amino acid residue Trp (Vásquez et al., 1983) and 5-hydroxy-Trp. 145

Chapter 3

Fig.1. Total number of helical hydrogen-bonds in the fifth helices of the N-terminal domain of the λ repressor. ● Active mutants. ■ Inactive mutants. 150

Fig.2. $C_{85}O_{85}N_{89}$ angles against the $O_{85}...N_{89}$ distances for the twelve helices of the active and twelve helices of the inactive mutants. ● Active mutants with hydrogen bond. ○ Active mutants without hydrogen bonds. ■ Inactive mutants with hydrogen bonds. □ Inactive mutants without hydrogen bond. 151

Fig.3. Torsion angles ψ_{85} and ϕ_{86} for the twenty-four helices defined in the legend to Fig.2. 152

Fig.4. Rise between Tyr-85 and the residue at position 86 plotted against the corresponding tilt for the twenty-four helices defined in the legend to Fig.2. 153

Fig.5. Axial inclination (A_{in}) between Tyr-85 and the residue at position 86 plotted against the corresponding Axial tip (A_{tp}) for the twenty-four helices defined in the legend to Fig.2. 154

Fig.6. Number of pairs of atoms in close inter-monomer contact involving residue i of the fifth helices vs. residue position. 155

Fig.7. Number of pairs of atoms in close inter-monomer contact involving residue i of the fifth helices (excluding the first-two) vs position $i-2$. 156

General Introduction

1. Aims and specific objectives

I have used the computer package ECEPP/3 (Empirical Conformational Energy for Peptides) (Momany et al., 1975; Némethy et al.;1983,1992) to discriminate between experimentally characterized active and inactive mutants of the N-terminal domain of the λ repressor protein bearing specific amino acid substitutions at the dimer interface (Reidhaar-Olson & Sauer, 1989, 1990).

The specific questions I have addressed are as follows:

1. What are the energetics of the interactions of the helix-5-helix-5', helix-4-loop-helix-5-helix-4'-loop-helix-5', and the full length protein complexes in the wild-type N-terminal domain of the λ repressor?
2. How do these inter-monomer interactions in mutants with DNA-binding activity differ from those occurring in the wild-type and biologically inactive mutants?
3. Are the relative orientation, position and minimal distance of separation between fifth helices in the wild-type conserved in the biologically active mutants?
4. Are the relative orientation, position and minimal distance of separation

between the recognition helices in the wild type conserved in the biologically active mutants?

5. Is there interdomain communication between the dimer interface and the protein core? For example, how does a mutation in helix-5 affect the relative orientation and distance of the recognition helices helix-3 and helix-3'?

6. 5-OH Trp is a fluorophore probe with demonstrated applications in studies on protein-DNA interactions. Studies have shown that a spectrally enhanced protein could be obtained by metabolically incorporating 5-OH Trp instead of Trp (Hogue et al., 1992; Ross et al., 1992; Laue et al., 1993; Hogue & Szabo, 1994). However, little is known about the structural and functional changes that may result from such a replacement. We have addressed the following question: How would the incorporation of 5-OH Trp at Trp-tolerating positions of helix-5 affect the inter-monomer packing interactions and indirectly DNA-binding activity?

2. Methods

(a) Overview of conformational energy computations

The working hypothesis on which any computation of conformational energy in molecular biology is based is that the most stable equilibrium state has the lowest free energy. The concept of the thermodynamic hypothesis was proposed following Anfinsen and coworkers's experiment which showed that denatured proteins could spontaneously regain their native conformations in vitro (Epstein et al., 1963; Anfinsen, 1973). This experiment provided the first evidence that the tertiary fold of a protein arises primarily from the physico-chemical properties of its amino acid side-chains and the surrounding solvent medium. In principle, it is possible to achieve a theoretical prediction of protein structure from its primary sequence and a potential function describing intramolecular and intermolecular interactions occurring in the system (For review see Reeke, 1988; Némethy & Scheraga, 1990). The ultimate goal within the context of calculations consists of finding the minimum

of the potential energy function.

However, some other evidence has brought about the idea of chaperone-assisted protein folding. Such examples include the case of the mitochondrial inner matrix proteins that do not fold properly in the absence of hsp60, the function of rubisco-binding protein in chloroplasts and that of BIP in the endoplasmic reticulum (For review see Lodish, 1988; Horwich, 1990).

Pioneering studies on molecular mechanic computations of polypeptides were carried out by Ramachandran, Liquori, Scheraga & Némethy, and Flory. The latter development was then carried out by Scheraga & Némethy, Kollman & Weiner, Karplus, Hagler, Hermans and van Gunsteren (Momany et al., 1975, Brooks et al., 1983; Weiner et al., 1984, 1986; Dauber-Osguthorpe et al., 1988). The force fields commonly used in conformational energy computations include the following contributions: nonbonded, electrostatic, hydrogen bonding interactions and intrinsic torsional potential.

These non-covalent interactions are mathematically formulated as follows:

$$1. \text{ Nonbonded energy ; } U_{NB} = \sum_{i,j>1} (a/r_{ij}^{12} - c/r_{ij}^6) \quad (1)$$

The first term of this Lennard-Jones 12-6 potential represents the repulsive energy between atoms i and j separated by the distance r_{ij} . The second term

represents the attractive dispersion energy. The parameters a and c are related to the potential well as well as the corresponding internuclear distance, and thus depend on the chemical nature of the atoms i and j . The parameter a is determined experimentally by the deviation from ideality of collisions in gases. The London parameter c in the repulsive term is experimentally determined by means of spectroscopy and fitting of crystal structures.

$$2. \text{ Hydrogen bond energy ; } U_{\text{HB}} = \sum (a'/r_{ij}^{12} - c'/r_{ij}^{10}) \quad (2)$$

where $i=\text{H}$ and $j=\text{O}$ or N . The parameters a' and c' depend on the nature of the hydrogen bond donor-acceptor pair and are also related to the depth of the energy minimum and the corresponding internuclear distance. Both parameters are derived from crystal geometry or thermodynamic measurements.

$$3. \text{ Electrostatic energy ; } U_{\text{ES}} = \sum q_i q_j / D r_{ij} \quad (3)$$

Electrostatic interactions are thus calculated by using the well-known Coulomb's law. q_i and q_j designate partial atomic charges on atoms i and j , r_{ij} is the internuclear distance and D is the dielectric constant of the medium. The charge on an atom in a molecule cannot be measured experimentally and neither can it be defined uniquely. Instead, the partial atomic charges used in molecular mechanics computer packages are usually obtained from

quantum mechanical calculations using Mulliken population analysis. Values near 80 and 2-3-4 are commonly used for the dielectric constants of water-exposed and buried protein regions, respectively.

$$4. \text{ Torsional energy ; } U_{\text{TOR}} = \sum U_0/2 (1 + \cos n\theta) \quad (4)$$

where U_0 is the difference (in Kcal/mol) between the experimentally determined barrier height and the one calculated from the first three above-mentioned contributions. θ and n are the dihedral angle and the symmetry of the barrier, respectively. Spectroscopic measurements (IR, Raman, microwave, NMR) constitute the main experimental sources for the rotational barrier.

In any computation of conformational energy, the input consists of the specific amino acid sequence, the specific starting values of the dihedral angles specifying the starting conformation, the molecular geometry of the amino acid residues (bond lengths and bond angles) and the parameters of the potential function as described above. The residue geometry is derived from X-ray crystal structures of amino acids and small peptides. The computational procedure consists of the following steps. First, the atomic coordinates are computed to generate the starting conformation. Second, the interatomic distances are determined. Then, the potential energy E is

computed. Finally, the function E is minimized with respect to the dihedral angles of internal rotation, i.e. backbone and/or side-chain dihedral angles as variables.

The most popular macromolecular computer packages used for computations of conformational energy are AMBER (Weiner et al., 1984, 1986), CHARMM (Brooks et al., 1983), DISCOVER (Dauber-Osguthorpe, 1988), and ECEPP (Momany et al., 1975, Némethy et al., 1983, 1992). The main difference between these algorithms is that the first three use flexible geometry by including in the potential energy function terms that take into account molecular distortions in the following manner :

$$1. \text{ Bond Stretching ; } U_{\text{str}} = \sum k_{\text{str}}(b-b_0)^2 \quad (5)$$

$$2. \text{ Bond angle bending ; } U_{\text{bend}} = \sum k_{\text{bend}}(\tau - \tau_0)^2 \quad (6)$$

where b_0 and τ_0 designate equilibrium values of bond length b and bond angle τ , while k_{str} and k_{bend} designate force constants. Such a formalism assumes that the atoms of a macromolecule behave as though they were joined by mutually independent harmonic oscillators. These springs obey Hooke's law. A Morse potential function is therefore applicable. The experimental source for numerical values of the distortion parameters is vibrational spectroscopy.

By contrast, ECEPP, the computer package that I have used in this

work, uses rigid geometry. Although molecules are in principle flexible, the use of rigid geometry simplifies significantly the computational procedure by decreasing the number of variables by about a factor of three. Thus, energy minimization is done with respect to only dihedral angles of internal rotation about single bonds. It relieves the steric overlaps between atoms which mainly result from long-range interactions in a folded protein. Since unfavorable atomic contacts arise from short-range interactions only in small organic molecules, it is not necessary to vary bond lengths and bond angles in studying proteins. Moreover, the use of flexible geometry may bring about some numerical uncertainties due to the lack of knowledge about the anharmonic terms which are neglected in expressions 5 and 6.

ECEPP uses a semi-empirical molecular orbital method namely CNDO/2 (Complete Neglect of Differential Overlap) to determine partial atomic charges. Is it really the best theoretical description of atomic charges? Although it is extremely difficult to determine which theoretical charge distribution is the most realistic, one could make a judgement based on the calculated dipole moments. Ample evidence exists that ab initio quantum mechanical calculations based on Hartree-Fock approximation and methods beyond (e.g. Møller-Plesset perturbation theory) give dipole moments in

better agreement with experiments. More precisely, the split-valence and polarization basis set constitute the minimal level which provide better descriptions of dipole moments (Hehre et al., 1986).

Another potential limitation of the ECEPP algorithm is the neglect of implementation with molecular dynamic simulation. Numerous studies have shown that the biological function of a macromolecule results not only from its static structure, but also its dynamic behavior (for review see McCammon & Harvey, 1987).

A comparison of the CHARMM, AMBER and ECEPP potential functions has been carried out in two ways by Roterman et al. (1989 a-b). First, using the same starting conformations for the tandemly repeated peptide (Asn-Ala-Asn-Pro), the conformations obtained after energy minimization using any of these potential functions were compared. The results suggested that low-energy conformations obtained with any two of these potential could resemble each other or not. Moreover, the ordering and the energy difference of these low-energy conformations were different. Second, the ϕ - ψ maps obtained from computations were used to test the ability of these potential functions to predict experimental results, namely the characteristic ratio for poly-L-alanine. Although none of the three algorithms was found to reproduce

completely the experimental findings, ECEPP and AMBER using fixed geometry with $\epsilon=4.0$, demonstrated best agreement.

The two most general drawbacks inherent to any computation of conformational energy, regardless of the algorithm in use, are the inaccuracy in modelling the interactions with the solvent and the multiple-minima problem (For review see Gibson & Scheraga, 1988). The free energy, which is the thermodynamic state function used to assess protein stability, takes into account both the energetic and entropic contributions. The latter result mainly from solvent interactions and molecular vibrations. Neglecting the entropic contributions is nearly equivalent to modelling an isolated and rigid protein molecule in vacuo. Even if the free energy function were accurately described, the incredible number of internal variables would still render the finding of the global minimum very challenging. In fact, even with the use of the most efficient minimizer algorithms and rigid geometry, it is a local minimum of the multidimensional energy surface which is actually determined. This local minimum is usually located near the starting point.

(b) Related computational studies on mutants of the λ repressor system

1. Neglect of the DNA operator

Computational studies of the structure, stability and macromolecular recognition properties of mutants of the N-terminal domain of the λ repressor along with their purification and physico-chemical characterization are desired in order to learn more about structure-function relationships in this system. The first step in this direction was made by Pabo and Suchanek (1986). Focussing on shape complementarity, they used model-building strategies, implemented with a computer program named Proteus, to find positions in the N-terminal domain of the λ repressor where a cysteine residue could be introduced that would result in a disulfide bond between the two monomers.

They assumed that if any two residues in the repressor have their backbone atoms spatially arranged in the same way as the backbone atoms of two disulfide-linked Cys residues in their data base, then these positions are likely to tolerate disulfide-linked Cys residues. Their database of disulfide-bond conformations was derived from the Protein Data Bank

(Bernstein et al., 1977) and was supplemented with 243 conformations related to the left-handed spiral and 243 conformations related to the right-handed hook (Richardson, 1981; Thornton, 1981). They identified position 88 and their prediction was experimentally verified by Sauer et al. (1986). The Cys-88 disulfide-linked homodimer was found more stable than the wild-type protein and showed increased DNA-binding activity. As a control, the computational study also predicted that replacing Tyr-85 by Cys would result in an inactive disulfide-linked homodimer. Strikingly, this prediction was also experimentally verified by Sauer et al. (1986).

In their study, Pabo and Suchanek used the 3.2Å resolution X-ray crystal structure of the free protein which was composed of only C α coordinates (Pabo & Lewis, 1982). They used early versions of the ECEPP standard geometry for amino acid residues (Momany et al., 1975) and the CHARMM potential function (Brooks et al., 1983) for model-building and energy minimization, respectively. They did not account for the fact that the relative position of two Cys residues in a disulfide bond can also result from a stereochemical rearrangement necessary to accommodate the disulfide bond. The solvent, DNA and counterions were not included in the modeling procedure. Even so, the results illustrate an excellent example of accurate

prediction of mutational effect which has pioneered computational studies of the λ repressor system.

A second attempt based on shape complementarity to predict stability and biological activity in the N-terminal domain of the λ repressor, was made by Lee and Levitt (1991), using the DNA-bound conformation (Jordan & Pabo, 1988). They applied a method developed by Lee and Subbiah (1991) for predicting side chain conformations by packing optimization, using a simulated-annealing approach (Kirkpatrick et al., 1983), to study mutants of the N-terminal domain of the λ repressor with amino acid substitution in the protein core. The calculated energy correlated successfully with the measured DNA binding activity, determined by Lim and Sauer (1991). Active and inactive mutants could be discriminated reliably with a score of 92%.

However, the potential function that was used, included only two terms: van der Waals interaction and torsional potential. Such a simplification gave nevertheless very good results, presumably because the mutations were done in the protein core where specific hydrophobic residues were replacing other hydrophobic residues.

The method of Lee and Levitt (1991) may not be generalized to site-directed mutagenesis on a solvent-exposed protein region. Electrostatic

interactions as well as hydrogen bonding interactions were neglected in their potential function, and these play a dominant role in solvent-exposed regions of a protein. Even in a hydrophobic protein core, hydrogen bonds between buried residues still may occur. Moreover, the energy was minimized with respect to only side-chain dihedral angles of residues near the site of mutation. The protein backbone was kept rigid. Lim et al. (1994), however, have recently solved the crystal structure of one of the mutants studied by Lee & Levitt. Surprisingly, a backbone shift of helix-4 relative to helix-3, has been observed in this inactive mutant. This result clearly demonstrates the importance of varying the backbone dihedral angles during energy-minimization.

Using simple models based on structural considerations alone (amino acid side-chain volume, shape and flexibility) to predict the activity of the same mutants studied by Lee and Levitt (1991), van Gunsteren and Mark (1992) have achieved comparable agreement with experimental results. Although these authors sharply criticized the study of Lee and Levitt, we have to acknowledge, however, that the empirical parameters used in their study are just another way of expressing the terms of the potential function of Lee and Levitt, namely van der Waals and torsional interactions. Furthermore, the

method of Lee and Levitt has been able to accurately and rapidly predict protein side-chain conformations of nine different protein cores of known structures (Lee and Subbiah, 1991); providing an independent demonstration of the efficiency of the method.

Focussing again on mutations of hydrophobic core residues of the N-terminal domain of the repressor, Lee has improved his, Subbiah's and Levitt's method by using a self-consistent ensemble optimization approach (Lee, 1994). The terms in the potential function are the same as in the previous study and no explicit interactions with the solvent or DNA are considered. The method predicts the most likely ensemble for the native state at a given temperature instead of predicting a static minimum-energy structure. The calculated difference energies correlated very well with the experimentally-determined thermal stability and measured free energies of unfolding.

A very promising computational method developed by Hellinga and Richards (1994), has recently been applied on mutants of the λ repressor with substitution in the protein core. In this approach, which has been named "Protein Simulated Evolution" (PROSE), both sequence and conformational spaces are explored. A molecular mechanic potential function and a simulated

spaces are explored. A molecular mechanic potential function and a simulated annealing approach are used to evaluate conformations and eliminate sequences of high energies, respectively. The force-field includes solvation, entropy and non-bonded interactions terms. PROSE has successfully simulated the random mutagenesis experiments carried out by Lim & Sauer (1989, 1991).

Although this algorithm exhibits the advantage of explicitly exploring sequence space and including solvation and entropic contributions, some limitations can be noticed. First, the protein backbone was kept fixed and energy-minimization was done with respect to only side-chain dihedral angles of residues at the sites of mutation. Second, only van der Waals interactions are evaluated. Electrostatic and hydrogen bonding are ignored. Third, only one monomer of the repressor was used as molecular model. The energy computed for one monomer, is compared to the biological activity of the whole system, neglecting that dimerization is coupled to operator binding.

In the present study, we have used a docking approach based on the ECEPP/3 algorithm (Momany et al., 1975; Némethy et al., 1983, 1992; Gibson & Scheraga, 1994) to discriminate between active and inactive mutants of the N-terminal domain of the λ repressor with substitution in the dimerization

interface (Reidhaar-Olson & Sauer; 1988, 1990; Reidhaar-Olson et al., 1990; Weiss et al., 1987a). The DNA-bound conformation of the N-terminal domain of the λ repressor, elucidated at 1.8 Å resolution by Beamer and Pabo (1992), has been used in this study. Energy minimization was done with respect to all internal variables (backbone and side-chain dihedral angles) and all external variables (Euler angles and translational displacements). A geometric analysis of the low-energy conformations was then carried out. The results are discussed in the subsequent chapters.

2. Inclusion of the DNA operator

A flexible docking method followed by a Poisson-Boltzman treatment was used by Zacharias et al. (1994) to predict low-energy conformations resulting from a C→T mutation at position 6 of the consensus half-site of the OL1 DNA operator. The first step of their approach involved a Monte Carlo conformational search in a dihedral angle space with bond lengths and angles kept fixed. The conformational search was restricted to the region neighboring the site of mutation. The potential function used included torsional, Lennard-Jones, hydrophobic and electrostatic terms. The latter involved

involved a calculation of the electrostatic interaction energy by means of Finite-difference Poisson-Boltzman equation. Finally, solvent-accessible surface area calculations were carried out in order to determine hydrophobic contributions to the interaction energy. The protein-DNA model system included only the consensus monomer and the corresponding half-site DNA operator region. Moreover, the conformational search did not account for the possibility of structural rearrangement of the protein or DNA in the regions relatively far from the site of mutation. Nevertheless, the study predicted low-energy side-chain conformations of Lys at position 4, which were close to the X-ray structure. In addition, a 1.1 kcal/mol binding energy difference was observed for the C.G→T.A mutation in comparison to the wild-type operator. This result is in reasonable qualitative agreement with the experimentally-determined value of 3.0 kcal/mol (Sarai & Takeda, 1989).

In an attempt to assess the effect of amino acid substitutions on the DNA-binding affinity of the N-terminal domain of the λ repressor, Wang and Sarai (1994) have developed a method based on shape complementarity and energy-minimization of a potential function including torsional, van der Waals, hydrogen-bond and electrostatic interactions. In their study, the macromolecular model was composed of both the protein and DNA. The

potential energy of the protein-DNA complex (E_{pd}) was given as

$$E_{pd} = E_p + E_d + E_{p \cdot d} \quad (7)$$

where E_p , E_d and $E_{p \cdot d}$ represent the potential energy of the protein, the DNA and protein-DNA intermolecular interaction. The Tripos force field (Clark et al., 1989) and Amber partial atomic charges (Weiner et al., 1984) were used in the calculations. Energy-minimization was limited to the residues close to the site of mutation.

A binding affinity index, a quantity proportional to the energy difference between the mutant and the wild type protein, was defined and used to discriminate between active and inactive mutants of the N-terminal domain of the λ repressor with substitution in the flexible N-terminal arm (Clarke et al., 1991). Although hydrophobic interactions, conformational entropy and hydration effects were ignored in the calculations, the results obtained were in good qualitative agreement with experiment.

Most recently, Jayaram & Beveridge (1996) have carried out a computational analysis of the energetic contributions in the N-terminal domain of the λ repressor-operator complex. To decipher the protein-DNA

intermolecular interactions, they have proposed a potential function including three terms: electrostatic, van der Waals and desolvation for the non polar groups. Their study has shown that van der Waals interactions involving the recognition helix and its flanking residues are predominant in the process of DNA recognition. This result justifies a posteriori the success obtained by the computational studies solely based on shape complementarity. The study has also revealed that the direct code, which involves interactions between the DNA bases and amino acid side chains, contributes to the fourth of the total interaction energy. However, the study used the crystal structure of the λ repressor-operator complex (Beamer & Pabo, 1992) as such, without any further energy-minimization and/or molecular dynamic refinement.

**Chapter 1: Computer-aided discrimination
between active and inactive mutants of the
N-terminal domain of the bacteriophage λ
repressor**

Abstract

23

Binding of the N-terminal domain of the λ repressor to DNA is coupled to dimerization. Hydrophobic interactions between helix-5 and helix-5' drive the packing at the dimer interface. We have carried out computations of the conformational energy of packing of the fifth helices (and of the helix-4-loop-helix-5 portions) of variants of the λ repressor operator binding domain, using an ECEPP/3-based packing algorithm. Here, we report the results for twenty-six mutants chosen among those that have been characterized experimentally by Weiss et al.(1987), Reidhaar-Olson & Sauer (1988,1990) and Reidhaar-Olson et al. (1990). We find that the relative orientation of the fifth helices for active mutants is very similar to the wild-type. The fifth helices of the inactive mutants have a significantly different relative orientation. This result illustrates that a unique specific orientation pattern of helix-5 relative to helix-5' is required for dimerization-coupled DNA binding activity. This finding is further supported by computational studies of the whole N-terminal domain of ten variants which showed that all active mutants, including the wild-type protein, have similar values of the number of contacts between the two monomers in the dimer, involving two amino acid residues of the fifth helices (Positions 84 and 87 in each monomer). A

decrease in the number of such contacts abolishes DNA-binding activity. Furthermore, all active mutants have their "DNA-recognition helices", numbers 3 and 3' positionned so that they can fit in the DNA operator like those of the wild-type protein, while some inactive mutants exhibit a substantial change in the relative orientation of their recognition helices.

1. Introduction

Given the importance of being able to change at will the stability and/or macromolecular recognition properties of proteins, general methods to guide such tasks are needed. Previous computational attempts have made use of perturbation methods and free energy calculations (Bash et al.,1987; Dang et al.,1989; Tidor & Karplus, 1991; McCammon, 1991). Reasonable agreement with experimental results has been observed. However, these methods consume a large amount of computer time.

In an attempt to develop rapid and efficient alternative strategies for protein design and modification, we propose a modelling procedure that is based on an analysis of the geometrical arrangement of the partners in a complex, implemented with energy minimization. Our strategy is based on previous studies of the energetics of interactions of

regular structural elements in proteins (Chou et al., 1983, 1984, 1990; Gerritsen et al., 1985), which showed that some overall aspects of the assembly of regular structural elements result from local interaction energies. To test the potential of this strategy for predicting and accounting for protein-protein interactions, we have used the operator binding domain of the cl λ repressor as a model system.

The λ repressor, the cl gene product of the bacteriophage lambda, is both a positive and a negative gene regulator (Meyer et al., 1975, 1980; Meyer and Ptashne, 1980; Maurer et al., 1980). Enzymatic digestion of the λ repressor with papain generates two fragments: the N-terminal domain, which corresponds to residues 1-92 (Fig.1), and the C-terminal domain, which corresponds to residues 93-236 (Pabo et al., 1979). The N-terminal domain is the DNA operator binding portion. The C-terminal domain (not shown in Fig.1) mediates dimerization of the full length protein. Dimerization of the N-terminal domain alone in solution is also required for its DNA-binding activity. The crystal structure of the N-terminal domain of the λ repressor is known (Pabo & Lewis, 1982), as is the crystal structure of its complex with a DNA operator site (Fig.1) (Jordan & Pabo, 1988; Beamer & Pabo, 1992). Each monomer is made of five α -helices interspaced by loops. The first four helices form the protein core. The

second and the third helices are part of the helix-turn-helix DNA recognition motif. The fifth helix provides the dimer interface through hydrophobic interaction with the fifth helix of the other monomer.

The geometrical basis of our computational methodology and its application to the N-terminal domain of the λ repressor stems from the following experimental and theoretical observations: 1). The DNA operator sequence has a pseudo two-fold symmetry which requires that each monomer of the repressor binds to a half-site operator sequence in a similar manner (Maniatis et al., 1975, Meyer et al., 1975, Humayun et al., 1977). 2). The third helices in both monomers (second helix of the helix-turn-helix unit, in blue in Fig.1), termed the recognition helices, fit into the major groove and are separated from each other by a distance of 34 Å, which is the length of a B-DNA turn (Jordan & Pabo, 1988; Beamer & Pabo, 1992). 3). The appropriate distance and correct relative orientation between the two recognition helices is provided by the dimerization process, mediated by interactions between the fifth helices, which protrude from the globular domain of each monomer (Fig. 1, 2a-b) (Pabo & Lewis, 1982; Weiss et al., 1987). 4). Dimerization, per se, is not sufficient for a repressor mutant to bind to DNA (Sauer et al., 1986). 5). The relative orientation and position of secondary structural elements is a

consequence of the interaction energy between them. For example, it has been shown that electrostatic interactions tend to favor an antiparallel relative orientation of two α -helices that are located side by side (Chou et al., 1983). 6). The arrangement of helix dipoles and the overall shape in the helix-turn-helix unit seems to play a major role in DNA recognition by means of favorable electrostatic interactions between the negative charges of DNA and the positive pole of the N-terminus of each helix (Sauer et al., 1990). Therefore, it seems reasonable to expect that alteration of the orientation and position of these helices through mutation might disrupt the close fit of the helix-turn-helix portion into the major groove. Experimental evidence for such a situation has recently been provided by the crystal structure of the *inactive* mutant VL36ML40VI47 of the repressor, which has revealed a rigid-body shift of α -helices relative to the wild type (Lim et al., 1994).

The free energy of protein-DNA association results from the local and global shape complementarity mediated through direct protein-DNA interactions, the release of water molecules, and the release of counterions. In this particular system, however, we have focussed only on shape complementarity and used solely the DNA-binding conformation of the protein as a criterion of good binding, assuming that this property

makes the largest contribution for the following reasons: a) Binding of the N-terminus of the λ repressor to DNA occurs primarily by a lock and key complementarity involving the helix-turn-helix unit (Pabo & Lewis, 1982; Jordan & Pabo; 1988), b) The binding does not significantly promote structural reorganization of either the protein or the DNA (Pabo & Lewis, 1982; Jordan & Pabo; Beamer & Pabo, 1992) which, otherwise, would be dominated by electrostatics and desolvation, c) studies of salt effects on λ repressor - DNA interactions (Misra et al., 1994) have shown that geometry not only plays a role in hydrophobic, hydrogen-bond and ion-pair interactions, but also in determining the free energy of macromolecular interaction with solvent and ions, d) A computational analysis of the energetics of the λ repressor - operator complex, using a potential function developed for studying protein-DNA interactions and including electrostatics, van der Waals and desolvation terms has revealed the predominant role of steric complementarity (Jayaram, 1995). This finding has justified the prior success of numerous studies of this system, which focussed on the shape complementarity. In fact, previous computational studies on mutants of the N-terminal domain of the λ repressor have successfully predicted the DNA-binding activity using energy minimization of a potential function for the protein, excluding DNA and/or

water molecules (Pabo & Suchanek, 1986; Lee & Levitt, 1991; Hellinga & Richards, 1994; Lee, 1994). Another study based on shape complementarity but which included DNA has also been reported (Wang & Sarai, 1994).

We have hypothesized that: a) The dimerization which is required for DNA-binding activity, is achieved when the mutant structure has both the wild-type geometric patterns of helix-5 relative to helix-5' at the dimer interface, and of helix-3 relative to helix-3' in the protein core. This will place the protein core of each monomer in proper register for binding to a half operator site of the two-fold symmetric operator region. b) Inactive mutants should present any geometric pattern of helix-5 relative to helix-5', which differs from that of the wild type, irrespective of the geometric pattern of helix-3 relative to helix-3'. Clearly, these statements do not exclude or contradict the possibility that a mutant with impaired dimerization will be inactive regardless of the relative orientation of its DNA recognition helices.

To test these hypotheses, we have carried out computations of portions and of the whole protein dimer, using an ECEPP/3-based packing program. We find that active and inactive mutants can be discriminated in terms of the following geometrical parameters of the energy-minimized structures: Interhelical orientational torsion angle, interhelical projected

torsion angle, interhelical closest distance of approach, number of contacts, and root-mean-square deviation.

2. Results

Examination of the crystal structure of the λ repressor N-terminal domain - operator DNA complex reveals that the fifth helix of each monomer protrudes outward from the globular core and packs against the fifth helix of the other monomer (Fig.1 and 2). In an attempt to depict the packing interactions between monomers, we first explored the interhelical interactions between helix-5 and helix-5' in an isolated context, i.e., excluding the rest of the protein. Then, we explored the packing interactions between helix-4-loop-helix-5 and helix-4'-loop-helix-5', also in an isolated context. Consideration of the *whole* N-terminal domain is then presented.

(a) *Computational study of the wild-type helix-5...helix-5' complex*

First, we generated a standard-geometry model of the helix-5...helix-5' complex by using the procedure described in the Methods section. The lowest-energy standard-geometry structure of the wild-type protein obtained after the constrained minimization has a total ECEPP energy equal to -260.3 kcal/mol, an intermolecular energy of -35.1 kcal/mol and an r.m.s. deviation of 1.11 Å with respect to the X-ray structure. As a test, we determined the r.m.s deviation between the C α atoms of the helix-5...helix-5' complexes of the isolated protein (only C α atomic coordinates are known; Pabo & Lewis, 1982) and the protein in the DNA complex (Beamer & Pabo, 1992), and found it also to be 1.11 Å. The minimum-energy standard-geometry structure was used as a starting point for further minimization without any constraints. A further computational test of the reasonableness of our use of the wild-type standard geometry with 1.11 Å deviation from the crystal structure was carried out by determining the residues which are in close contact (see Methods section) in the lowest-energy wild-type conformation. We found

that Ile-84, Tyr-88, Met-87, Val-91, Ala-81 and Tyr-85, respectively, contribute to close contacts, with the following numbers of pairs of atoms (i,j) in contacts: 123, 112, 66, 54, 23, and 11 (where i belongs to helix-5 and j to helix-5'). This result is in agreement with previous observations (Pabo & Lewis, 1982; Weiss et al, 1987), namely, that the helix-helix contact at the dimer interface of the N-terminal domain of the λ repressor is hydrophobic in nature and involves the side chains of Ile-84, Tyr-85, Met-87, and Tyr-88.

The results shown in Table 1 indicate that the lowest-energy conformation of the wild-type complex has an intermolecular ECEPP energy of -34.5 kcal/mol, 234 contacts, an orientation torsional angle (Ω_o) of -129.2°, a projected torsional angle (Ω_p) of -129.5°, and r.m.s. deviations with respect to the free and the DNA-bound protein of 0.7 and 1.0 Å, respectively. The energies and geometrical parameters of the other wild-type low-energy conformations were used as a background range. To allow for fluctuations, half the length of the interval was added and subtracted to the maximum and minimum values for each parameter unless otherwise indicated. We consider a parameter of a given mutant conformation to be significantly different from the wild-type value and thus of inactive mutant type when it falls outside of the background

ranges given in the last-two lines of Table 1.

b) Conformational energy of mutants of the helix-5...helix-5' complex

Since the mutants under study have different amino acid sequences, their intramolecular energies and total energies cannot be compared. The energy parameter that *can* serve as a criterion for comparison of these structures is the intermolecular term, which is in fact a packing energy. Since packing energies are referred to the infinitely-separated partners at zero energy, the packing energy is a valid quantity to decide in which complex the partners bind more strongly than in another pair of monomers. The results shown in Table 2 indicate that most of the active mutants have intermolecular energies varying between -35.4 and -31.9 kcal/mol, regardless of the position at which the amino acid substitution was made. However, exceptions are mutants containing Ala at position 88. The latter active mutants have higher values of the intermolecular ECEPP energy than wild-type, within the range -37.2 to -24.7 kcal/mol (Fig.3). Strikingly, the lowest-energy conformations of the inactive mutants with triple substitutions, namely IA⁸⁴MV⁸⁷YA⁸⁸ and IT⁸⁴MT⁸⁷YV⁸⁸ as well as the

90-residue fragment have higher intermolecular energy: -21.7, -25.7 and -27.2 kcal/mol, respectively. By contrast, the low-energy conformations of the inactive mutant IS⁸⁴ have intermolecular energies within the active range [-35.4,-31.9] kcal/mol. However, the low-energy conformations for the helix-4-loop-helix-5...helix-4'-loop-helix-5' complexes of the wild-type and active mutants have intermolecular energy values varying between -98.0 and -94.0 kcal/mol, while the low-energy conformations of IS⁸⁴ have intermolecular energies in the range of -87.0 to -89.0 kcal/mol (Table 3). Thus, including a fragment of the protein core in the simplest model involving only the fifth helices has provided a better distinction between the wild-type and this mutant in terms of intermolecular energies. Similarly, the lowest-energy conformations for the helix-4-loop-helix-5...helix-4'-loop-helix-5' complexes of the inactive mutants IA⁸⁴MV⁸⁷YA⁸⁸ and IT⁸⁴MT⁸⁷YV⁸⁸ have higher intermolecular energies than the wild-type and active mutants, around -70.0 and -78.0 kcal/mol, respectively. The lowest-energy conformations for the helix-5...helix-5' complex of the single-site inactive mutants VY⁹¹, MF⁸⁷, IM⁸⁴ and IF⁸⁴ have lower intermolecular energies, varying from -34.2 to -45.7 kcal/mol. Figure 4 shows the high variability of intermolecular energy values for the various inactive mutants. These results indicate that the

intermolecular energy of the dimer, in some cases but not in others, is a good parameter to discriminate between active and inactive mutants of the λ repressor N-terminal domain (Table 2 and Figs.3-4), although good agreement between variation in the intermolecular energy and the variation in the experimentally-determined dimer dissociation constants has been observed (Table 3).

On the basis of their intermolecular energy values, the mutants could be divided into four classes: active class 1, active class 2, inactive class 1, and inactive class 2 (see Table 2). Mutants of active class 1, which included the wild-type sequence, generally had intermolecular energies within the wild-type range, i.e. [-35.4,-31.9] kcal/mol, while mutants of active class 2 had higher intermolecular energies. Mutants of inactive class 1 had intermolecular energies higher than the wild type background range whereas inactive mutants of class 2 had intermolecular energies within or lower than the wild-type background range.

(c) Number of pairs of atoms in close contacts

As with the intermolecular packing energy, the number of pairs of atoms in close contact, likewise, in most cases discriminates between

active and inactive mutants (Table 2, Figs. 5-6). The wild-type and active mutant structures have similar numbers of pairs of atoms in close contacts, within the range 207 to 243. However, the active mutants with Ala at position 88 have markedly fewer pairs of atoms in close contact, within the range 156-201, whereas the active mutants with Ala at position 91 have intermediate values, within the range 197-224 (Fig.5). Most inactive mutants have a markedly reduced number of pairs of atoms in close contact (Fig.6 and Table 2). The number of pairs of atoms in close contact for the inactive mutant IS⁸⁴ are altogether outside of the active range [207-243], and higher than the number of contacts for the two triple-site inactive mutants; thus, the number of contacts differentiates this inactive mutant from the wild-type while the intermolecular packing energy failed to provide a clear-cut answer. The mutant MT⁸⁷ can be classified as a member of inactive class 2. In fact, although having intermolecular energies close or equal to -32.0 kcal/mol, the low-energy conformations of this mutant have very low numbers of pairs of atoms in contact, specifically in the range of 200-206, which is completely outside of the wild-type background range [207,243]. Likewise, three out of five low-energy conformations of the mutant IV⁸⁴ have number of contacts outside the wild-type background range, suggesting that this mutant can

be classified as inactive. computational studies on the full-length N-terminal domain of the λ repressor provide a further criterion which identifies the mutant IV⁸⁴ as an inactive one (see section f). It is still difficult, however, to identify the mutant MV⁸⁷ as inactive even with this additional criterion.

(d) *Interhelical geometry*

Compared to the wild type, the other active mutants have nearly identical values for both torsional angles, specifically within the range of -128.4° to -132.8° . The low-energy conformations of the active mutants with Ala at position 88 have interhelical angles within almost the same range (-127.7 to -133.3°). Two out of ten low-energy conformations had values of the projected torsional angle outside of the allowed range for the active mutants (-146.2° and -141.4°). This may be understood as though the population of conformations of this mutant is heterogenous. Of the five low-energy conformations of YW⁸⁵YA⁸⁸, three exhibit an interhelical orientation like the active mutants whereas the remaining two conformations behave like conformations of inactive mutants. This is not surprising because replacing Tyr, whose side chain contributes to

inter-monomer contacts and has χ_1 predominantly in the trans conformation (MacGregor et al., 1987; Piela et al., 1987), by a smaller side chain such as Ala (with χ_1 in the g⁺ conformation) may significantly affect the packing of the fifth helices in some conformations.

By contrast, the inactive mutants have interhelical angles with values mostly significantly outside of the wild-type background range (Table 2 and Fig.7). For example, the low-energy conformations of some of the inactive mutants IA⁸⁴MV⁸⁷YA⁸⁸, IT⁸⁴MT⁸⁷YV⁸⁸, IF⁸⁴ and IM⁸⁴ have absolute values of their interhelical angle of 149°, 136°, 107° and 158°, respectively. Taken together, these results illustrate that the active mutants, including the one with Ala at position 88, can be discriminated from the inactive mutants by the projected torsional angle and orientation torsional angle. However, the inactive mutants MT⁸⁷, MV⁸⁷ and IV⁸⁴ have values of their interhelical angles in the range allowed for the active mutants. The interhelical angle of these mutants constitute a false positive for our discrimination criteria.

It appears that active mutants, including those with higher intermolecular energy and lower number of close contacts than wild type, can be distinguished from the inactive mutants by the projected torsional angle and the orientational torsional angle (Table 2 and Fig. 7).

(e) *Root-mean-square deviation between mutants and wild-type protein*

As shown in Table 2 and Figures 4-8, the r.m.s deviations between the mutant and wild-type protein also provides an additional criterion for distinguishing between active and inactive mutants. With the exception of active mutants containing Ala at position 88, all the active mutants exhibit low r.m.s deviations with respect to the lowest-energy conformation of the wild-type, between 0.12 Å and 0.70 Å. All inactive mutants have higher values of r.m.s. deviation. As for the interhelical geometry parameters, the inactive mutants MT⁸⁷, MV⁸⁷ and IV⁸⁴ are characterized by r.m.s deviations with respect to the lowest-energy wild-type standard-geometry model which are similar to those of the active mutants. Consideration of the r.m.s deviations with respect to the X-ray crystal structure of the bound and the free protein, however, clearly shows that the mutants MV⁸⁷ and MT⁸⁷ can be discriminated as inactive (Table 2) whereas the interhelical angle criterion failed to provide a clear-cut answer. As a control, the backbone r.m.s. deviation between the minimum-energy structures of mutants and the lowest-energy wild-type

structure was plotted vs. the r.m.s. deviation between the same structures and the DNA-binding conformation of the protein (Fig.8a and Table 2). The latter parameter in turn was plotted vs. the r.m.s. deviation with respect to the free form of the protein (Fig.8b and Table 2). The results indicate that a large r.m.s. deviation (e.g. 1.5 Å) from the lowest-energy wild-type model could not place a mutant structure closer to the DNA-bound protein conformation compared to the wild-type model.

(f) Computational results for the whole N-terminal domain dimer

To investigate further the packing at the dimer interface as well as the structural disruptions occurring in the protein core (which contains the DNA-binding unit) as a consequence of mutations at the dimer interface, we have applied the same computational methodology to the whole N-terminal domain dimer. We have selected at least one mutant from each class. The results obtained are summarized below.

1) *Residues involved in inter-monomer close contacts*

Determination of the residues which are involved in inter-monomer contacts in the lowest-energy conformation of the whole wild-type dimer indicated that Ile-84, Met-87 and Val-91 contribute the most in inter-monomer contacts. To analyze further the effect of mutation on the packing at the dimer interface, we determined the number of contacts made by residues at positions 84, 87 and 91, in the lowest-energy conformations of mutants of the N-terminal domain of the λ repressor. The results are shown in Table 4. We observe that all active mutants including YA⁸⁸, are characterized by numbers of pairs of atoms in contact at positions 84 and 87 which are similar to the corresponding values for the wild-type protein. By contrast, the inactive mutants show significantly reduced values at both positions or only at position 84. For example, the single-site inactive mutant VY⁹¹, despite having the same residues at positions 84 and 87 as the wild-type protein, shows a considerable decrease in the number of close contacts at these positions (78 and 109 vs. 129 and 127 for the wild type). Similarly, the triple-site mutant IT⁸⁴MT⁸⁷YV⁸⁸, although characterized by a lower inter-monomer energy and higher total number of contacts than wild type, still has very low

contributions at positions 84 and 87, as could be predicted from considering the size of Thr. Likewise, the inactive mutants MI⁸⁷ and IL⁸⁴ have a very high total number of contacts (635 and 612, respectively) and a very low inter-monomer energy (-113.9 and -122.7 kcal/mol, respectively), but also show a decrease in the number of contacts involving Ile-84 (118 and 93, respectively). With regard to contributions at position 91, the active mutants show higher than, or similar values to, wild type. By contrast, the inactive mutants MI⁸⁷, IA⁸⁴MV⁸⁷YA⁸⁸, IT⁸⁴MT⁸⁷YV⁸⁸, and wild type (1-90) show a significantly reduced number of contacts at position 91 (115, 113, 80 and 0, respectively), while the inactive mutants IV⁸⁴, VY⁹¹ and IL⁸⁴ behave like active mutants. We conclude that the close inter-monomer contacts involving residues at positions 84 and 87, particularly at position 84, are very important in determining the relative helical geometry which is required for DNA-binding activity. This result is consistent with the work of Reidhaar-Olson & Sauer (1988) which showed that Ile is the only acceptable residue at position 84, and Met can be substituted only by Leu at position 87.

2) *Relative geometry of recognition helices*

The relative orientation and position of "DNA recognition" helices (second α -helix of the DNA recognition helix-turn-helix unit) for various mutants of the N-terminal domain of the λ repressor have been determined. The results are summarized in Table 4. We find that the active mutants of both classes have similar values of the projected torsional angle Ω_p , the orientational torsional angle Ω_0 and closest distance of approach D. The single-site inactive mutants VY⁹¹ and IV⁸⁴, also show similar values for these parameters. This result suggests that the interhelical geometry of the recognition helices per se, is not sufficient for a mutant to be active. Strikingly, a significant change in the interhelical angle is observed for the inactive mutants, wild type (1-90), IT⁸⁴MT⁸⁷YV⁸⁸, IL⁸⁴ and MI⁸⁷. In the last three mutants, this change is accompanied by a decrease of 4 to 8 Å in the distance of closest approach. This result indicates that substantial changes in the interhelical geometry of the recognition helices can abolish DNA-binding activity. The inactive mutant IA⁸⁴MV⁸⁷YA⁸⁸ has intermediate values of the geometrical parameters between the active mutants and these inactive mutants with remarkable structural changes.

3. Discussion

Conformational energy computations of the helix-5...helix-5' complex followed by a geometrical analysis of the energy-minimized structures has been used to discriminate between active and inactive mutants of the N-terminal domain of the λ repressor. The successes and failures of each criterion and combination of various parameters are summarized in Tables 5A and 5B, respectively. The results in Table 5A indicate that the proper dimerization which is required for DNA-binding activity is reflected primarily by the interhelical angle at the dimer interface. Although the results in Table 5A also indicate that active mutants are better predicted than inactive ones, Table 5B shows that combination of interhelical angles with other criteria substantially increases the score for predicting an inactive mutant as inactive. Therefore, a good discrimination scheme between active and inactive mutants would start with the interhelical angles. The false positive will then be further discriminated by considering simultaneously the interhelical angles, the r.m.s. deviations with respect to the X-ray crystal structure of the DNA-bound form and the number of contacts. Using such a

scheme would accurately discriminate the inactive mutant MV⁸⁷, MT⁸⁷ and IV⁸⁴ as inactive, whereas the interhelical angle alone would have classified these mutants as false positive.

Sauer et al. (1986) have shown that dimerization per se is not sufficient for DNA-binding activity. Our results are in agreement with this experimental observation since some inactive mutants (e.g. VY⁹¹, IF⁸⁴ and IM⁸⁴) have lower intermolecular energies than wild type. Moreover, our results demonstrate that the wild-type orientation of helix-5 vs. helix-5' is responsible for promoting a dimer stability with a specific geometry capable of maintaining the protein core in the correct DNA-binding conformation. This conclusion can be tested experimentally by purifying mutants with low energy but incorrect geometry, and those with high energy but correct geometry, and comparing their dimerization dissociation constants with the affinity for DNA. Although our data have more to do with protein-protein interactions, they show reasonable consistence in predicting DNA-binding activity, a further indirect proof that DNA-binding is coupled to dimerization.

One of the very few studies of the dimerization of the operator binding domain of the λ repressor was carried out by Weiss et al. (1987) who used ¹H NMR and gel filtration in order to study structure-function

relationships in this system. The results obtained for the mutants IS⁸⁴, the wild-type 1-92 residue-fragment and the 1-90 residue-fragment are summarized in Table 2, together with our computational results. It can be seen that there is good agreement between the trend of variation in the intermolecular energies and the trend of variation in the dimer dissociation constants for the three proteins. However, we acknowledge that our computational procedure did not take the solvent into account, and we lack information about the unfolded state. Also, our intermolecular energy for the folded state is not a free energy. The small structural perturbations observed for the inactive mutants IS⁸⁴ and WT(1-90) compared to the wild type are also in agreement with experimental results. In fact, NMR studies have shown that the solution structure of these two mutants are very similar to the wild-type structure (Weiss et al., 1987).

The number of inter-monomer contacts mediated by residues at positions 84 and 87 seems to be very important for dimerization-coupled DNA-binding activity (Table 4). A significant decrease in these parameters results in an inactive mutant. Interestingly, active mutants of class 2, although characterized by a lower number of contacts than wild type and a higher inter-monomer energy have their residues at positions 84 and 87

contributing to close inter-monomer contacts to the same extent as in the wild-type protein. Conversely, the inactive mutants of class 2, despite being characterized by a higher number of contacts and a lower inter-monomer energy than wild type, have a decreased relative contribution of inter-monomer contacts at positions 84 and 87. The reduced number of contacts at positions 84 and 87 for the mutant IV⁸⁴ allows us to classify it now as an inactive mutant. Our studies which used portions of the protein core failed to discriminate the biological activity of this mutant.

We have shown that active mutants of both classes do indeed have their recognition helices properly positioned to fit into DNA. A change in the interhelical angle (of about 6° or more) and/or in the closest distance of approach (of about 4 Å or more) of the recognition helices results in loss of DNA-binding activity. The inactive mutants, VY⁹¹ and IV⁸⁴, however, have wild-type-like interhelical parameters (Table 4). The loss of DNA-binding activity for these mutants is probably due to weak association of the monomers to form the dimer. Since dimerization is coupled to operator binding, impairment of dimerization will affect DNA-binding affinity, whether or not an appropriate "lock and key" fit of the recognition helices into DNA may occur. In addition to the helix-turn-helix region, the flexible N-terminal (λ) arm (residues 1 to 5) and residues

which make contacts with the sugar-phosphate backbone are also involved in the protein-DNA interface (Jordan and Pabo, 1988; Clark et al., 1991; Beamer and Pabo, 1992). Consequently, another possibility which could account for the inactivity of the mutants with wild-type-like interhelical geometry of the recognition helices may be that, in these mutants, a conformational change in the *flexible* N-terminal arm (only partially resolved in the crystal structure) could affect the DNA-binding activity. In fact, Zacharias et al. (1990) have shown that a slight rotation of the λ arm significantly affects both sterically and electrostatically the association of the repressor with its DNA operator. Our studies, however, have used the crystal structure of Beamer & Pabo (1992) in which the non consensus N-terminal arm could not be observed. Moreover, our studies have not investigated the dynamics of the protein-DNA association to which the N-terminal arm probably makes a significant contribution. Finally, given the complexity of this system and the qualitative nature of the in-vivo assay used to determine the biological activity of these mutants, one can't expect perfect rules. It is important to recall that a mutant is said to be active in vivo if it has at least 10 % of the in-vivo DNA-binding activity of the wild-type protein (Reidhaar-Olson & Sauer; 1988,1990). It may be very difficult to discriminate computationally

borderline mutants, for example, inactive mutants with 7 to 9 % of wild-type activity and active mutants with 10 to 13 % of wild-type activity.

It is interesting that the inactive mutants IT⁸⁴MT⁸⁷YV⁸⁸, MI⁸⁷ and IL⁸⁴, which are characterized by lower inter-monomer energy and higher number of contacts than wild type, exhibit a decrease in the closest distance of approach between their recognition helices, suggesting stronger packing of the protein monomers. By contrast, the inactive mutant VY⁹¹, although also characterized by lower intermolecular energy and higher number of contacts than wild type, has almost the same value of D as the wild type (Table 4). This result suggests that the packing between monomers is weaker in this mutant compared to wild-type, supporting the hypothesis of dimer instability.

The computational procedure that we have presented here may find application in the design of proteins with enhanced biological function. Various DNA-binding proteins such as λ Cro, 434 repressor and restriction enzymes, which specifically recognize palindromic oligonucleotide sequences, are suitable for the application of such a modeling strategy in which geometry is analyzed after carrying out energy minimization. The availability of amino acid sequences which are known to result in biologically active or inactive proteins also opens new avenues for the

inverse folding problem. We have shown that, in order for a λ repressor mutant to be biologically active, its tertiary and quaternary structures must be very similar to those of the wild-type. Consequently, selecting for biologically active mutants is equivalent to finding which amino acid sequences adopt the wild-type fold. An understanding of the physico-chemical properties that are common to the amino acid sequences which have passed the selection test is an important step towards formulating the rules that govern protein folding. Such knowledge could be exploited in the development of algorithms for homology-based modeling of proteins.

4. Materials and methods

(a) Fitting of X-ray derived polypeptide chains to standard-geometry models of the N-terminal domain.

X-ray coordinate sets of the N-terminal domain of the λ repressor were obtained from the Brookhaven protein data bank [file 1lmb.pdb (Beamer & Pabo,1992)]. We examined the structure using the Polygen software package Quanta. All water molecules and the DNA molecule were removed from the structure, and the remaining structure is shown in Fig.1.

The X-ray coordinates of helix-5 and helix-5', helix-4-loop-helix-5 and helix-4'-loop-helix-5' as well as those of the whole N-terminal domain of the λ repressor were fitted to the ECEPP (Empirical Conformational Energy Program for Peptides) standard geometry (Momany et al., 1975; Némethy et al., 1983; Némethy et al., 1992). For the latter model, residues 1 to 5 of the non-consensus monomer were not observed in the crystal structure, and were omitted in our computations. The fitting procedure was carried out in the following manner; 1). The dihedral angles calculated from the X-ray coordinates were used to generate the atomic coordinates of standard-geometry chains using the geometrical parameters supplied in the ECEPP standard residue library. 2). The X-ray structure of the polypeptide was then rotated and translated to achieve an optimal superposition with the standard-geometry conformation generated in the previous step. The constrained energy of the standard-geometry polypeptide chain, consisting of the ECEPP/3 energy terms and harmonic terms constraining all the heavy atoms of the standard-geometry chain to their respective corresponding atoms in the X-ray structure, was minimized in an iterative manner. The relative weight of the harmonic constraint was decreased from 1.0 to 0.0 in steps during the iterative procedure. The final structure, that was obtained when the weight

assigned to the harmonic distance constraint was zero, was selected as the standard-geometry structure to use for further computations. The r.m.s. deviations of the heavy atoms of the final standard-geometry structures with respect to the corresponding atoms of the X-ray crystal structure were 1.11, 0.76 and 2.0 Å for the helix-5...helix-5' complex, the helix-4-loop-helix-5...helix-4'-loop-helix-5' complex and the full length N-terminal domain, respectively.

(b) *Energy minimization and geometrical analysis*

An ECEPP/3-based packing algorithm was then applied to determine the minimum intra- and interhelical energy of the wild-type protein and of mutants thereof; this algorithm has been described elsewhere (Fossey et al., 1991; Gibson & Scheraga, 1994). The algorithm uses the secant unconstrained minimization solver (SUMSL) of Gay (1983) to minimize the energy with respect to *all* backbone and side-chain dihedral angles and rigid-body variables.

Conformational energy computations with the ECEPP/3 packing algorithm have successfully predicted structures for the Poly(L-Alanine-Glycine) forms of the crystalline domains of *Bombyx mori* silk fibroin

(Fossey et al., 1991; Gibson & Scheraga, 1994). The results were in agreement with X-ray diffraction, NMR and IR spectroscopic data. Excellent agreement between the predicted and the observed single crystal structure has also been obtained in computations of the structure and assembly of triple-stranded models of collagen by using earlier versions of this algorithm (Némethy & Scheraga, 1989). These experimental results constitute an independent demonstration of the validity of this computational method (see also Vásquez et al., 1994).

Since the calculations involve the packing of α -helices in a crystal, no hydration effects were included. The coordinates of the $C\alpha$ atoms of the minimum-energy conformations were used for determining the following parameters for helices 5 and 5': the orientation torsional angle (Ω_0), and the projected torsion angle (Ω_p) and the distance of closest approach (D) (Chou et al.; 1983,1984). To analyze the packing of the fifth helices further, we determined the extent of intermolecular contacts between atoms. The number of pairs of atoms in close contact for each minimum-energy structure of these protein fragments was determined by using equation (35) from Chou et al. (1983). We adopted the definition of close contacts which was proposed by these authors; thus, an atom i of helix-5 is said to be in close contact with an atom j of helix-5' if and only if the

distance r_{ij} of separation of their centers is less than d^* , i.e.,

$$r_{ij} < d^* = (2 - 2^{-1/6}) (r_i^0 + r_j^0) \quad (1)$$

where r_i^0 and r_j^0 are the respective van der Waals radii of these atoms as used in the ECEPP/3 algorithm, and d^* is a cutoff distance which takes into account fluctuations in r_{ij} above and below the distance at which the van der Waals interaction energy between atoms i and j is minimum. Two residues belonging to different polypeptide chains are said to be in close contact if at least two of their respective atoms are in close contact. To determine the r.m.s. deviations between mutants and the wild-type protein, the backbone atoms C^α , C' , N , O and C^β of the low-energy conformations of the mutant helices were superposed on to the corresponding atoms of the lowest-energy conformation of the wild-type structure using a method based on the singular value decomposition of a certain matrix (Golub and Van Loan, 1983).

(c) *Starting conformations*

The mutations were generated in both monomers. The starting

standard-geometry backbone conformation for all mutants was the same as that of the energy-minimized wild-type protein. The starting side-chain conformations for the wild-type and mutant sequences were obtained by exploring combinatorially the most preferred conformations of amino acid side chains in an α -helix (Piela et al., 1987). Residues whose side-chain dihedral angles were varied in order to generate various starting conformations for the wild-type structure were Ile-84, Tyr-85, Met-87, Tyr-88 and Val-91. These residues were chosen because they contribute to close contacts between the fifth helices (see results section). Glu-86, Glu-89, and Ala-90 contribute to close contacts only when helix-4 or the whole protein core is included. Side-chain dihedral angles at these three positions as well as all the other sites of mutation were explored combinatorially to generate the starting side-chain conformations for the relevant mutant proteins. Consequently, the starting conformations of the mutants differ from those of the wild-type structure only in the side-chain dihedral angles of the residue at the site of mutation. The external variables of all the starting conformations (Euler angles and translational displacements) had the same values, the ones which resulted from the standard-geometry model generation procedure.

(d) Choice of mutants

Since our aim is to identify the different possible packing arrangements of the fifth helices that may provide DNA-binding activity to the protein core, the mutants under study were selected as follows; 1) Active mutants in which a residue that contributes to close intermolecular contacts was replaced either by a non-conservative residue (e.g. Tyr by Ala at position 88) or by a smaller side chain (Val by Ala at position 91, Tyr by Ala). 2) Conservative substitution at positions 88 and 91 were selected as controls (e.g. VT⁹¹, YF⁸⁸). 3) Some mutants were chosen because of the availability of additional experimental data (e.g. IS⁸⁴, wild-type (1-90) and Pro-78 mutants). 4) Substitutions at position 84 and 87 were particularly selected to probe the importance of these positions at which Ile and Met, respectively, are highly conserved (Reidhaar-Olson & Sauer, 1988). The local environment within helix-5, including the sites of mutation under study, has been described extensively in the literature (Pabo & Lewis, 1982; Jordan & Pabo, 1988; Weiss et al., 1987; Reidhaar-Olson & Sauer, 1988; 1990; Reidhaar-Olson et al., 1990; Sauer et al., 1990).

Chapter 2: Effects on protein structure and function of replacing tryptophan with 5-hydroxytryptophan: Single tryptophan mutants of the N-terminal domain of the bacteriophage λ repressor.

Abstract

Conformational energy computations have been carried out on the N-Acetyl N'-Methylamide of 5-hydroxytryptophan (5OH-Trp) using ECEPP/3. As observed with tryptophan (Trp), the most preferred conformation about the C α -C β bond of the side chain is g⁺ or t. This preference is reduced to only the t conformational state when 5-hydroxyTrp is in the middle of a right-handed poly(L-Alanine) α -helix. A similar result has been obtained with Trp (Piela et al., 1987). These results suggest that replacement of Trp by its analog 5-hydroxyTrp may be tolerated in an α -helix. To test this hypothesis, we have replaced Trp by 5OH-Trp in the fifth helices of two functionally active mutants of the N-terminal domain of the bacteriophage λ repressor. Computations on the packing of these helices have shown that no significant structural changes result from the replacement of Trp by 5OH-Trp. The DNA-binding activity of these mutants, as assessed indirectly through geometrical parameters, is also unaltered.

1. Introduction

The use of Trp absorption and fluorescence to study protein-DNA interactions is hampered by the fact that DNA absorption overlaps with that of Trp. This has been partially overcome by the use of extrinsic spectroscopic probes on either the protein or DNA. However, the latter approach, which depends on chemical treatment, is accompanied by a potential risk of altering not only the structure but also the functional characteristics of either macromolecule.

An alternative procedure for studying protein-DNA interactions involves the use of Trp analogs as intrinsic probes for absorption and fluorescence spectroscopy. To be used for such a purpose, an indole derivative must fulfill the following two conditions:

1. It must have spectral properties that are markedly different from those of Trp.
2. It must be readily incorporated biosynthetically in the protein of interest without altering its structure and its function.

5OH-Trp fulfills these criteria and has been used successfully in recent years as an intrinsic fluorophore probe (Hogue et al., 1992; Ross et al., 1992; Laue et al., 1993; Hogue & Szabo, 1993; Laws et al., 1995). For

example, Ross et al. (1992), have shown that 5OH-Trp is readily incorporated into the full-length bacteriophage λ cl repressor protein. The modified protein thus obtained still maintains wild-type properties. The 5OH-Trp repressor monomer self-associates to form dimers, and the dimers bind to DNA in a manner indistinguishable from the wild-type protein as shown by the dimerization equilibrium constants and DNA binding free energies. The fluorescence of the 5OH-Trp repressor can also be observed selectively in the presence of exogenous Trp or DNA. An important caution in generalizing the success of this approach, however, is the fact that the targeted Trp residues all belong to the carboxy terminal domain whose three-dimensional structure is unknown. One may argue that, had the Trp residues been part of a specific motif important for protein stability and/or function, their replacement by the analog might have been disruptive.

In an attempt to assess the conformational and functional changes that could result from replacement of Trp by its analog 5OH-Trp in a protein, we have used the N-terminal domain of the bacteriophage λ repressor as a model system. The wild-type protein does not contain any Trp residues, but several functionally active mutants which contain Trp in the fifth helices, have been identified (Reidhaar-Olson & Sauer, 1988). We

have chosen to investigate two Trp variants, namely YW⁸⁵YA⁸⁸ and AW⁹⁰VA⁹¹ that have been studied by Kombo et al. (1995).

Examination of the protein environment within helix-5 does not suggest that 5-OHTrp would not be tolerated at these positions. However, given the numerous exceptions to tolerance of amino acid substitution in helix-5 that have been observed experimentally (Reidhaar-Olson & Sauer, 1988;1990), one cannot make predictions based only on the graphic examination of the protein environment. For example, at position 85, the aromatic ring of Tyr participates in stacking interactions with Tyr-88 which in turn stacks against Tyr-88'. Although Phe could provide comparable interactions at position 85, it is not tolerated. By contrast, Tyr-88 tolerates substitution by Ala.

We report here conformational energy computations on the packing of the fifth helices of two active Trp-containing variants in which 5OH-Trp replaces Trp. In the present paper, we see that the replacement of Trp by its analog 5OH-Trp in these variants affects neither the key structural features of the mutant proteins nor their DNA-binding activity, as inferred here from the computed geometry of the helix-5...helix-5' dimer.

2. Methods

a) Determination of a standard-geometry model of 5OH-Trp

Bond lengths, bond angles and dihedral angles of 5OH-Trp were obtained from the X-ray crystal structure of this compound elucidated by Wakahara et al. (1973). The ECEPP/3 computer package was then used to determine the cartesian coordinates of each atom of the molecule N-Acetyl-N'-methyl-5OH-Tryptophanamide. The overlap normalized partial atomic charges for every atom of this terminally-blocked single residue were determined by the complete neglect of differential overlap (CNDO/2) molecular orbital method (Pople & Segal, 1966; Momany et al., 1971). Nine different backbone and side-chain conformations of this molecule were investigated, and the resulting partial atomic charges were averaged. The standard residue data (geometric and potential energy parameters) for the 5OH-Trp residue were then determined as described in the ECEPP computer package (Momany et al., 1975; Némethy et al., 1983, 1992). Since an experimental value of the torsional barrier height for the side-chain

dihedral angle about the C ζ -O η bond of 5OH-Trp is not available, we used instead the corresponding value for tyrosine (Tyr), which is equal to 3.5 Kcal/mol (Lowe, 1968; Momany et al., 1975). The geometric and potential energy parameters for 5OH-Trp thus obtained were incorporated in the standard residue library used by the ECEPP/3 algorithm. These data were used for the generation of the atomic coordinates of a 5OH-Trp-containing polypeptide and the computation of its conformational energy.

b) Computation of conformational energy of model polypeptides

Conformational energy computations on the molecule N-Acetyl N'-Methyl 5OH-Tryptophanamide and the polypeptide CH₃CO-(L-Ala)₉-5OH-Trp-(L-Ala)₉-NHCH₃ were carried out with ECEPP/3, the latest version of the ECEPP computer package. The energies of various starting conformations were minimized with the Secant Unconstrained Minimization Solver algorithm (Gay, 1983). Two options were available:

1. The SUMSL version, which requires an analytical form of the derivative of the energy.
2. The SMSNO routine, which carries out a numerical computation of the derivatives.

In some computations, the two methods were used one after the other, in order to attempt further minimization. The dihedral angles used in the starting conformations were selected as described by Zimmerman et al. (1977). During minimization of the empirical energy of the poly-Alanine peptide, the side-chain dihedral angles of 5OH-Trp were the only ones allowed to vary (unless otherwise indicated), and the backbone was kept fixed with : $\phi = -68.1^\circ$, $\psi = -38.3^\circ$, $\omega = 180.0^\circ$, as described previously by Piela et al. (1987).

The normalized Boltzman factor v_i for each conformation was computed by the following equation;

$$v_i = Q^{-1} \exp (-\Delta E_i / RT) \quad (1)$$

where $Q = \sum_i \exp (-\Delta E_i / RT) \quad \text{for } i=1, n \quad (2)$

n being the total number of minima and ΔE_i the relative conformational energy at the i^{th} minimum. The Boltzman probability P_j was calculated as follows,

$$P_j = \sum v_j \quad \text{for } j \in J \quad (3)$$

where J is a group of low-energy conformations having a specific backbone or side-chain conformation. The resulting side-chain dihedral angles of 5OH-Trp in the low-energy conformations of the poly-Alanine polypeptide were used as the starting conformations for the minimization of the energy of the helix-5...helix-5' complexes, as described below. The low-energy conformations were chosen within a cutoff of 5 kcal/mol from the lowest energy.

c) Packing of helix-5 and helix-5' of variants of the N-terminal domain of the λ repressor

The computational procedure for the packing of the fifth helices of the 5OH-Trp-containing variants of the N-terminal domain of the λ repressor was carried out as described by Kombo et al. (1995). The atomic coordinates of the fifth helices of the N-terminal domain of the λ repressor were taken from the refined 1.8 Å crystal structure of its complex with the DNA operator (Beamer & Pabo, 1992). We removed the DNA and all water molecules, and then fitted the atomic coordinates of the fifth helices to the ECEPP (Empirical Conformational Energy Program

for Peptides) standard geometry (Momany et al., 1975; Némethy et al., 1983, 1992). The fitting procedure has been described in detail by Kombo et al. (1995). The r.m.s. deviation of the heavy atoms of the final standard-geometry structure obtained after constrained minimization with respect to the corresponding atoms of the X-ray crystal structure was 1.11Å. This structure was used for further minimization, using an ECEPP/3-based packing program (Gibson & Scheraga, 1994). However, some minor changes in the program were made in order to accommodate the non-naturally occurring amino acid residue 5OH-Trp. The structures of 5OH-Trp-containing mutants were obtained by substituting 5OH-Trp for Tyr-85 and Ala-90, respectively, in *both* monomers of the standard-geometry wild-type structure, and carrying out energy minimization with respect to *all* backbone and side-chain dihedral angles and external variables (Euler angles and translational displacements). The definition and procedure to determine the interhelical parameters [orientation torsional angle (Ω_0), projected torsional angle (Ω_p) and close contacts] were those described by Chou et al. (1983,1984).

3. RESULTS

(a) Conformational states of 5-hydroxy Trp.

The computed minimum-energy conformations of the N-Acetyl-N'-Methylamide of 5-hydroxy-Trp are listed in Table 1. The structure of the lowest-energy obtained is shown in Fig.1. From the data in Table 1, we determined the relative populations of various backbone conformational states and those of the χ_1 side-chain rotamer, and compared them with the corresponding results for Trp (Vásquez et al., 1983). As shown in Fig.2, on the whole both Trp and its analog exhibit the same trends in their preference of backbone conformations. In both cases, the extended backbone conformation (E) has the highest relative probability of occurrence while the left-handed α -helical conformation (A^*) has the lowest one. A slightly higher probability of the E conformation is observed for 5OH-Trp, which is counterbalanced by a slightly lower probability of the A and C conformational states.

The trend of conformational preferences of the side-chain rotamer

χ_1 is the same for both amino acid residues, i.e. $g^+ > t > g^-$.

The relative Boltzman probabilities P_j for the g^+ , t and g^- rotamers around the $C\alpha-C\beta$ bond for Trp, determined by Vásquez et al. (1983), are equal to 0.433, 0.354 and 0.213, respectively. The corresponding computed values for 5OH-Trp are 0.424, 0.320 and 0.256, respectively. Moreover, in both Trp and its analog, the first three lowest-energies are found for the E conformation, with χ_1 being in either the g^+ or the t conformational state. Similar to Tyr, the side-chain dihedral angle about the $C\zeta-O\eta$ bond of 5OH-Trp is either near 0° or 180° , corresponding to the cis or the trans conformation, respectively.

To search for possible hydrogen bonds, we specified the maximal distance between the hydrogen atom and the acceptor atom as 2.5 Å, and the minimal angle at the hydrogen atom as 90° . Our choice of these parameters was based on the results of the analysis of hydrogen bonding in globular proteins by Baker and Hubbard (1984). We have not observed any hydrogen bonds between backbone and side-chain atoms. In accord with these observations, conformational energy computations carried out on the N-Acetyl-N'-Methylamide of Tyr and Trp (Vásquez et al., 1983) also did not show the presence of any hydrogen-bond formation between backbone and side-chain atoms. Moreover, no hydrogen bond between

backbone atoms was observed by us for the N-Acetyl N'-Methylamide of 5-OH-Trp and by Vásquez for the N-Acetyl N'-Methylamide of Tyr. However, in the case of the terminally-blocked amino acid residue Trp, a backbone-to-backbone hydrogen bond was found in all the E and C conformational states that had a total energy within 1.2 kcal/mol above the lowest-energy state (Vásquez et al., 1983).

(b) Conformational constraints of 5-hydroxy-Trp side chain in a right-handed poly(L-Alanine) α -helix.

The above results arise from short-range interactions which occur between the atoms of 5-hydroxy-Trp and those of the immediate neighboring peptide units, as represented by the two terminal blocking groups. Longer-range interactions, such as side chain-side chain interactions in a protein are not present in this small model compound. To assess the restriction on the conformational freedom of the 5OH-Trp side chain that might result from long-range interactions, we initially incorporated 5OH-Trp into the middle of poly(L-Alanine) folded as a regular α -helix. The results presented in Table 2 indicate that the t rotamer is the only one to survive the conformational constraints. The

reduction in the relative probabilities of the g^+ and g^- conformations is due to steric hindrance, as shown by higher values of the nonbonded energy term. A similar computation for Trp also showed a preference for the t rotamer (Piela et al., 1987). Taken together, these results suggest that substitution of Trp by 5OH-Trp in an α -helix will not result in a significant change in conformational freedom of the aromatic side chain. However, our use of a poly-Alanine polypeptide as an α -helical model does not include the general effects of backbone-side chain interactions as well as side chain-side chain interactions involving amino acid residues larger than Alanine.

(c) Computational study of the helix-5...helix-5' complex of variants of the N-terminal domain of the λ repressor with 5OH-Trp substituted for Trp.

To analyze further the changes in the conformational freedom accompanying substitution of 5OH-Trp for Trp in a protein, we used as a more realistic sequence a Trp-containing variant of the fifth helix of the N-terminal domain of the λ repressor. This α -helix is a suitable model for

a more extended and thorough investigation for the following reasons:

1. It is an amphipathic α -helix containing amino acid residues that are larger than Alanine and have diverse chemical properties (polar, aromatic, hydrophobic and hydrophilic).
2. It packs against a symmetry-related α -helix, leading to dimerization of the N-terminal domain monomer.
3. A functional assay for DNA-binding activity is computationally feasible by means of a geometry-based modelling strategy following energy minimization (Kombo et al., 1995).

The parameters characterizing the packing states of the low-energy conformations of the helix-5...helix-5' complex for two mutants in which Trp is replaced by 5OH-Trp are shown in Table 3. The low-energy conformations of the mutant A(5OH-W)⁹⁰VA⁹¹ generally have interhelical angles of about -130° and an rms deviation with respect to the wild-type protein, in the range of 0.49 - 0.54 Å. The number of atom pairs in close contact is in the range of 199 - 215, which is nearly identical to the corresponding range for its counterpart active mutant AW⁹⁰VA⁹¹ (Kombo et al., 1995). Similar to its counterpart active mutant YW⁸⁵YA⁸⁸, the 5OH-Trp-containing mutant Y(5OH-W)⁸⁵YA⁸⁸ is characterized by very low numbers of pairs of atoms in close contacts,

specifically in the range of 155-199, and high rms deviations. The projected torsional angle Ω_p and the orientational torsion angle Ω_o also have numerical values that are strikingly similar to the ones for the mutant YW⁸⁵YA⁸⁸. For example, two of the low-energy conformations have exceptionally high absolute values of 155° and 142° for the interhelical angles in both the Trp- and 5OH-Trp-containing compounds. Taken together, all of the above results predict that expression of the two mutants A(5OH-W)⁹⁰VA⁹¹ and Y(5OH-W)⁸⁵YA⁸⁸ should result in proteins that are competent in DNA binding, in a manner similar to their Trp-containing counterparts.

4. Discussion

5OH-Trp is a spectroscopic probe with demonstrated applications in studies of protein-DNA interactions. Experimental studies by Szabo & coworkers (Hogue et al., 1992) and Ross & coworkers (Ross et al., 1992, Laue et al., 1993) have shown that a spectrally enhanced protein could be obtained by metabolically incorporating 5OH-Trp instead of Trp in both oncomodulin and the full-length λ repressor protein. However, little is known about possible structural and functional changes that may result

from such a replacement. In this paper, we have addressed the following question: How would the incorporation of 5OH-Trp at Trp-tolerating positions of helix-5 affect the intermolecular packing interactions and indirectly the DNA-binding activity? We have shown that the probabilities of occurrence of backbone and side-chain conformational states of the terminally-blocked amino acid 5OH-Trp are very similar to those of Trp. For both protein building blocks, the E conformational state is the most preferred whereas the A' conformation is least preferred. The most predominant side-chain rotamers for both single amino acid residues are g⁺ and t. The computed minimum-energy conformations of 5OH-Trp can now be used as starting points in any minimization study of the conformational energy of polypeptides and proteins containing this Trp analog. The side-chain conformation that occurs with the highest Boltzman probability when either Trp or its analog 5OH-Trp is in the middle of a right-handed poly(Alanine) α -helix is the trans state. This result suggests that the replacement of Trp by its analog in an α -helix should not impair the structural and functional properties.

This hypothesis has been confirmed by the calculations showing that variants of the fifth helices of the N-terminal domain of 5OH-Trp-containing λ repressor pack in a similar fashion as the fifth helices of

their Trp-containing counterparts. The very small values obtained for the root-mean-square deviations of the C α , C', N, O, and C β atoms of the three 5OH-Trp-containing mutants compared to those of their Trp-containing counterparts (Table 3, fifth column, right), suggest that no significant structural change capable of abolishing the DNA-binding activity has taken place. On the basis of the intermolecular energy values discussed by Kombo et al. (1995), the mutant A(5OH-W)⁹⁰VA⁹¹ can be classified as member of active mutants class 1, a group in which its counterpart Trp-containing mutant AW⁹⁰VA⁹¹ belong. Likewise, the mutant Y(5OH-W)⁸⁵YA⁸⁸ is an active mutant of class 2.

It is interesting to note that replacement of Trp by 5OH-Trp in the mutant YW⁸⁵YA⁸⁸ has affected the relative probabilities of different conformations that have total energies within 5 kcal/mol of the lowest-energy conformation. In the mutant YW⁸⁵YA⁸⁸, the lowest-energy conformation has a projected torsional angle of -146.2° whereas in the 5OH-Trp-containing counterpart, the low-energy conformation with a similar value of the projected torsional angle (-143.0°) is the fourth-lowest energy conformation. Consequently, the Boltzman probability of the conformations with the interhelical angle required for DNA-binding activity is higher in the mutant Y(5OH-Trp)⁸⁵YA⁸⁸ than in its Trp-

containing counterpart. Furthermore, the low-energy conformations of the mutant Y(5OH-W)⁸⁵YA⁸⁸ exhibit on the whole lower values of the r.m.s deviations with respect to the wild type (Table 3) in comparison to the corresponding values obtained for its Trp-containing counterpart (Table 1 of Kombo et al., 1995). Taken together with the aforementioned observation on the change in the relative probabilities of different conformations upon replacing Trp by 5OH-Trp (see discussion above), these results suggest that the structure of the mutant Y(5OH-W)⁸⁵YA⁸⁸ is very likely closer to the wild-type structure than is the mutant YW⁸⁵YA⁸⁸. Interestingly, this finding demonstrates that replacing Trp by 5OH-Trp in a mutant protein could result in a lesser perturbed structure than the Trp-containing counterpart.

Another possible significant conformational change that may result from replacing Trp by 5OH-Trp in an α -helix is the change in the backbone screw sense of the α -helical conformation. An example of such an effect is illustrated by the computational and experimental conformational analysis of single-stranded o-, m- and p-Chlorobenzyl esters of poly-L-Aspartic acid (Yan et al., 1970; Erenrich et al., 1970). These studies showed that the chlorine atom on the benzene ring affects the handedness of the helical backbone. The ortho and meta isomers form left-handed α -

helices while the para isomer forms a right-handed one. It should be noted that the side-chain bond C ζ -O η of 5OH-Trp has a dipole moment as does the C-Cl bond of the chlorobenzyl esters of poly-L-Aspartic acid. In our study, we have not observed any change in the α -helical screw sense when Trp is replaced by its analog 5OH-Trp. Starting from both the left-handed and right-handed α -helical polypeptides CH₃CO-(L-Ala)₉-Trp-(L-Ala)₉-NCH₃ and CH₃CO-(L-Ala)₉-5OH-Trp-(L-Ala)₉-NCH₃, we have minimized the conformational energy with respect to all backbone and side-chain dihedral angles. We have observed that replacing Trp by its analog 5OH-Trp, does not affect the helical handedness, as indicated by the values of the dihedral angles Φ and Ψ of the low-energy conformations within 5.0 kcal/mol of the lowest-energy minima (data not shown). Likewise, the Boltzman probability of the A* (left-handed α -helix) conformational state is almost the same in both model compounds N-Acetyl-N'-Methyltryptophanamide and N-Acetyl-N'-Methyl 5-hydroxy Tryptophanamide. A is preferred over A* for both Trp and 5OH-Trp even in the model compounds. The slight decrease in the Boltzman probability of the A and C conformational states observed with 5-hydroxy-Trp is counterbalanced by an increase in the population of the E conformational state, rather than that of the A* conformational state.

Engineering of spectrally-enhanced proteins by substitution of 5OH-Trp or other analogs of interest for Trp can be assisted by computational analysis if a three-dimension structural model of the protein of interest is available. First, it is important to make sure that the targeted Trp residue is not buried in the interior of the protein where mutations are less likely to be tolerated (Reidhaar-Olson & Sauer; 1988,1990). Further investigations would involve conformational energy computations on the 5OH-Trp-containing protein, using the low-energy side-chain conformations of 5OH-Trp that we have determined here, as starting points for the minimization procedure. Finally, it will be important to determine whether the replacement may affect the structure at the active site or in a functional domain of the protein. This could be accomplished by a geometrical analysis following energy minimization (Kombo et al., 1995) or by using more general methods for comparing conformations, such as differential geometry (Rackovsky & Scheraga, 1978, 1980,1981, 1984), optimal superposition (Rossmann & Argos, 1976), and distance maps (Levitt, 1976; Kuntz et al., 1976).

**Chapter 3: Computational analysis of
variants of the operator binding domain of
the bacteriophage λ repressor.**

Abstract

In a continuing effort to understand the molecular basis of the dimerization-coupled DNA-binding activity of the bacteriophage λ repressor protein, we have analyzed conformational energy computations results on the wild type and its mutants with substitution in the dimer interface. We find that the hydrogen bonds made by the peptide carbonyl oxygen of Tyr-85 to the peptide NH group of the residue at position 89 in active mutants are longer and less linear than their corresponding ones in inactive mutants. This behavior is due to the outward tilting of the carbonyl oxygen of Tyr-85 from the helix axis in active mutants, which in many cases results in disruption of the i to $i+4$ hydrogen bond involving this residue. The helical and hydrogen-bond parameters computed for both active and inactive mutants have been compared with the results obtained by others in the analysis of X-ray-derived crystal structures of α -helices in globular proteins. We find that the packing of α -helices at the dimer interface of active and inactive mutants results in a hydrogen bonding geometry for the carbonyl group of Tyr-85 that is similar to

that found in α -helices with and without CO groups hydrogen bonded to a solvent molecule, respectively. A rough correlation is also observed between the tolerance for amino acid substitution at position i of helix-5 with the number of close inter-monomer contacts involving the residues at position i and $i-2$. On the basis of these findings, we discuss the mutability of amino acid residues at the dimer interface .

1. Introduction

There is an increasing need to understand macromolecular recognition properties that govern physiologically important processes such as drug-receptor, enzyme-substrate, protein-protein and protein-DNA interactions. An understanding of such interactions, coupled with a knowledge of the structural features necessary for biological activity, is expected to lead to the design of more effective ligands.

The N-terminal domain of the bacteriophage λ repressor binds to dyad symmetric DNA motifs as a symmetry-related dimer, and regulates gene expression (Maniatis et al., 1975; Meyer et al., 1975; Ptashne et al., 1976; Humayun et al., 1977; Pabo et al., 1979). Each monomer consists of five

the symmetry-related fifth helix of the other monomer (See Figs. 1 and 2 of Chap. 1 and Kombo et al., 1996). The packing of these fifth helices, usually referred to as helix-5 and helix-5', mediates dimerization of the N-terminal domain. The second and the third helix in each monomer, and the loop between them, constitute the DNA-binding helix-turn-helix unit. The third helix of each monomer is termed the recognition helix. The DNA-operator half-site which is conserved in almost all known operators is termed the consensus site (Sauer et al., 1990). The other half-site differs by two to four base-pairs from the consensus sequence and is termed the non-consensus half-site. Consequently, the protein monomer which binds specifically to the consensus site is called the consensus monomer. Likewise, the protein monomer which binds specifically to the non-consensus half-site is called the non-consensus monomer.

Kombo et al. (1996; see also chap. 1) have recently investigated protein-protein interactions involved in the λ repressor-DNA complex. They have used energy-minimization with respect to all backbone and side-chain dihedral angles and rigid-body variables, followed by a geometrical analysis of the energy-minimized structures, for such a system in which a pseudo two-fold symmetry of the dimer is required for binding to nearly two-fold

dihedral angles and rigid-body variables, followed by a geometrical analysis of the energy-minimized structures, for such a system in which a pseudo two-fold symmetry of the dimer is required for binding to nearly two-fold symmetric DNA operator sites. They have been able to discriminate between active and inactive mutants of the N-terminal domain of the λ repressor on the basis of differences in geometrical properties. In particular, they have shown that DNA-binding activity correlates well with the interhelical angles for the fifth helices and the recognition helices, as well as the number of inter-monomer contacts involving buried amino acid residues at positions 84 and 87. It is important to note that their theoretical predictions about the trend in the variation of the dimer dissociation constants of some mutants have been experimentally verified by Senear and coworkers (personal communication from Dr. Donald Senear).

It should be noted, however, that the relative orientation between the fifth helices, as determined by Kombo et al. (1996; see also chap.1), was based solely on C α positions. Such an analysis does not provide any information about the local distortions that might result from mutating a specific residue. In fact, Némethy et al. (1967) observed that, in some α -helices, rotation of a single peptide group can keep the position of C α atoms

unchanged, while affecting the dihedral angles ϕ and ψ or the hydrogen-bonding pattern. Similarly, a mutation can affect either the hydrogen-bonding pattern or dihedral angles ϕ and ψ in helices and thereby alter a biologically important interaction without changing the positions of C^α atoms very much. It may then be difficult to discriminate between active and inactive mutants by means of parameters determined solely from C^α atomic positions. Consequently, the study of Kombo et al. (1996; see also chap. 1) did not address a number of important questions such as: 1) Is the rigid-body shift of helix-5 relative to helix-5' observed in inactive mutants accompanied by helical distortions? 2) How does a mutation at position i of helix-5 affect position j , situated farther along the sequence? 3) What is the rationale for the tolerance of amino acid substitution in the fifth helices of the N-terminal domain of the λ repressor?

In this chapter, we present an extension of their investigation together with a detailed computational analysis, aimed at addressing the aforementioned questions. We find that the dimer interface of the N-terminal domain of the λ repressor is disrupted as a result of distorting the helical hydrogen-bonding pattern. Active and inactive mutants could be further discriminated by means of helicoidal parameters. The tolerance of amino acid

substitution can easily be rationalized by means of some geometrical properties. Tyr-85, a *solvent-exposed residue*, seems to play a critical role in maintaining the dimer interface in the appropriate geometry required for dimerization-coupled DNA-binding activity.

2. Methods

a) Generation of standard-geometry models and energy-minimization

The atomic coordinates of the N-terminal domain of the λ repressor alone (excluding the DNA and water molecules) were taken from the refined 1.8 Å crystal structure of its complex with the DNA operator (Beamer & Pabo, 1992), and were then fitted to the ECEPP (Empirical Conformational Energy Program for Peptides) standard geometry (Momany et al., 1975; Némethy et al., 1983, 1992). The generation of the standard geometry models of the helix-5-helix-5' complex and the full-length N-terminal domain of the λ repressor are described elsewhere (Chap. 1 and Kombo et al., 1996). The models obtained in that study are the same ones used here.

Basically, the backbone conformation was kept in the wild-type form,

whereas side-chain conformations of the replacing amino acid residue were explored combinatorially using the rotamer library of Vázquez et al. (1983) and Piela et al. (1987). Minimization of the conformational energy was carried out with respect to all backbone and side-chain dihedral angles and rigid-body variables. Use was made of an ECEPP/3-based packing algorithm (Gibson & Scheraga, 1994) which utilizes the secant unconstrained minimization solver (SUMSL) of Gay (1983). Independent demonstration of the applicability of this method in predicting experimental results has been provided by various examples such as structural predictions of the Poly(L-Alanine-Glycine) silk-I and silk-II forms of the crystalline domains of *Bombyx mori* silk fibroin and Poly(L-Alanine-Glycine) silk-II (Gibson & Scheraga, 1994; Fossey et al., 1991). Excellent agreement with X-ray diffraction data, NMR and IR spectroscopic data were found. Similar successes were obtained in the study of structure and assembly of triple-stranded models of collagen by means of earlier versions of this computer package (Némethy & Scheraga, 1989). Additional studies have been described in detail by Vázquez et al. (1994).

b) Determination of geometrical parameters

To determine the inter-monomer contacts, we applied to the lowest-energy conformations, the definitions and procedures described by Chou et al. (1983, 1984). The helicoidal parameters [axial tip (ATP), axial inclination (AIN), rise, tilt and roll] were determined as described by Sklenar et al. (1989) and Swaminathan et al. (1990). The possibility for hydrogen-bonding was assessed by the criteria used in the STRIPS algorithm (Ravishanker et al., 1994). Specifically, the distance between the acceptor atom and the hydrogen atom was set at $2.25 - 0.75 \text{ \AA}$. The angle CO-H at the acceptor atom was set at $150 \pm 30^\circ$, while the electrostatic interaction involving the N-H and C=O groups was set at $-4 \pm 3 \text{ kcal/mol}$. The solvent accessible surface area of atoms was determined by means of the algorithm of Lee and Richards (1971). To compute the surface area, a probe of radius $1.4 \text{ \AA}$ was rolled on the surface of each atom.

c) Choice of mutants

The fifth helices of the N-terminal domain of the wild-type λ repressor protein start from Pro-78 and end at Ser-92. Their amino acid sequence is as follows: -P-S-I-A-R-E-I-Y-E-M-Y-E-A-V-S-COOH (Sauer & Anderregg, 1978; Pabo & Lewis, 1982). The mutants under investigation were chosen from those that were characterized experimentally by Weiss et al. (1987), and by Reidhaar-Olson & Sauer (1988, 1990). Our selection was based on the following criteria: 1). Ile-84, Met-87, Tyr-88 and Val-91 are residues of helix-5 which contribute the most to the fifth helix-mediated intersubunit close contacts in the N-terminal domain of the λ repressor (Pabo & Lewis, 1982; Weiss et al., 1987; Kombo et al., 1996; see also Chap.1). Therefore, we have focussed primarily on some active and inactive mutants with substitution at least at one of these key role-playing positions. Specifically, We have chosen the active mutants (VT⁹¹, YF⁸⁸ and ML⁸⁷) and inactive ones (IS⁸⁴, MF⁸⁷, VY⁹¹, IA⁸⁴MV⁸⁷YA⁸⁸, IT⁸⁴MT⁸⁷YV⁸⁸ and IF⁸⁴) which were studied previously by Kombo et al. (1996; see also Chap. 1). 2) As a control, we also chose mutations involving Glu-86 and Glu-89, two highly solvent-exposed residues which are not involved in helix-5-helix-5' contacts.

3) Results

a) Helical hydrogen-bonding pattern of the fifth helices

To investigate the helical distortions that occur at the dimer interface of variants of the N-terminal domain of the λ repressor, we applied the geometric and energetic criteria for hydrogen-bonding (see the Methods section) to the twenty-four helices. The total number of hydrogen-bonds involving the CO group of residue i and the NH group of residue $i+4$ is plotted against the CO group position, for both classes of mutants (see Fig.1). The results indicate that the helical hydrogen-bonds are disrupted in inactive mutants throughout the whole sequence of helix-5 and helix-5', except at position 85, where the CO group of inactive mutants makes hydrogen-bonds in eleven out of twelve helices, while its counterpart of active mutants is involved in only three corresponding hydrogen-bonds.

Further evidence for helical distortions at the dimer interface of these mutants is shown in Fig.2, where the C85O85N89 angles are plotted against the $O_i...N_{i+4}$ distances. It can be seen that the mean $O_{85}...N_{89}$ distances for

hydrogen-bonds of active and inactive mutants are 3.05 ± 0.03 and $2.89 \pm 0.11 \text{ \AA}$, respectively. The corresponding means for the $C_{85}O_{85}N_{89}$ angles for the two classes are $144^\circ \pm 1^\circ$ and $150^\circ \pm 4^\circ$, respectively. We conclude that the main-chain-to-main-chain hydrogen-bonds involving the carbonyl group of Tyr-85 and the backbone amide group of the residue at position 89 are longer and less linear in active mutants than in the inactive ones. This effect is presumably due to the tilting of the carbonyl at position 85 further outward from the helix axis in active mutants as indicated by the plot of the torsion angles ψ_{85} and ϕ_{86} which is shown in Fig.3. For active mutants with $O_{85} \dots H-N_{89}$ hydrogen bonds, ψ_{85} has a mean value of $-36 \pm 2^\circ$, whereas the corresponding mean value for inactive mutants is $-40 \pm 4^\circ$. The mean values of ϕ_{86} for the active and inactive mutants are $-69 \pm 1^\circ$ and $-66 \pm 3^\circ$, respectively.

The results presented so far are summarized in Table 1, together with the hydrogen-bonding geometrical parameters in X-ray-derived crystal structures of α -helices (Blundell et al., 1983; Baker & Hubbard; 1984; Barlow & Thornton; 1988). The results in Table 1 indicate that the overall trend in the variation of the following helical parameters ϕ , ψ , $O_i \dots N_{i+4}$ and $C_iO_iN_{i+4}$ involving Tyr-85 and the residue at position 89 for active vs. inactive mutants is strikingly similar to the one observed between "hydrophobic" vs. "buried"

hydrogen-bonds (Blundell et al., 1983; Barlow & Thornton; 1988), respectively. This finding is further supported by comparing the trend of mean values of the distances $O_i \dots N_{i+4}$, $O_i \dots H_{i+4}$ and angles $N_{i+4}H_{i+4}O_i$ for active vs. inactive mutants with the values obtained by Baker & Hubbard (1984) on α -helices with CO groups making additional hydrogen-bonds with water or side-chain vs. those without such bonds. The helical hydrogen-bonding parameters for the inactive mutants are, however, surprisingly very similar to those of a putative "hydrophilic" α -helix, computer-generated by assigning to all residues the dihedral angles values $[(\phi, \psi) = (-66^\circ, -41^\circ)]$ of residues in which the carbonyl is hydrogen-bonded to a solvent molecule or side chain (Blundell et al., 1983).

The mean values of solvent accessibility surface area of the carbonyl oxygen atom at position 85 for active and inactive mutants are 1.44 and 0.64 $\text{\AA}^2$, respectively, while the corresponding values for the backbone nitrogen atom of the residue at position 89 are 0.072 and 0.344 $\text{\AA}^2$, respectively. This result further suggests that the backbone carbonyl group of Tyr-85 is more tilted outward of the helix axis and thus more solvent-exposed in active mutants compared to their inactive counterparts. This phenomenon is accompanied by an inward tilt of the backbone nitrogen atom of the residue

at position 89, which is more pronounced in active mutants compared to their inactive counterparts.

Since only three out of twelve helices show hydrogen-bonds between Tyr-85 and the residue at position 89 in the class of active mutants , while eleven out of the twelve helices show corresponding hydrogen bonds in the inactive class, the comparison of the mean values of helical parameters between the two classes seems biased. To demonstrate the reasonableness of the use of the mean values for such a small sample of active mutants with hydrogen bonds, we determined the mean values of the same parameters for all twelve helices of active mutants. The results shown in parentheses in Table 1 indicate that the trend in variations of these parameters for all the twelve helices of active mutants is very consistent with the one obtained considering only the limited number of active mutants with hydrogen bonds. This suggests that little would probably be gained if one increases the sample size for active mutants with helical hydrogen bonds involving position 85 and 89.

b) Helicoidal parameters of the fifth helices

Further evidence of helical distortions at position 85 of the N-terminal domain of the λ repressor is provided by the helicoidal parameters which are defined by the P-curve algorithm of Sklenar et al. (1989). The values of these parameters, which characterize the successive peptide groups at positions 85 and 86 for active and inactive mutants, are shown in Figs. 4 and 5. The results indicate that active mutants generally have higher tilt but lower rise than their counterpart inactive mutants. Interestingly, the plots of the rise against the tilt (Fig.4) and the axial inclination (AIN) against the axial tip (ATP) (Fig.5) are very similar in shape to the ones observed when the angle CON was plotted against the distance N...O (Fig.2) and the dihedral angle ψ against ϕ (Fig.3), respectively, providing further ways for discriminating between active and inactive mutants.

c) Inter-monomer van der waals contacts

1) involving aromatic rings of Tyr-85 and Tyr-85' in the isolated helix-5-helix-5' complexes.

Our study so far has focussed on geometrical parameters determined exclusively from the backbone atoms and dihedral angles. It is known, however, that the helix-helix contact at the dimer interface of the N-terminal domain of the λ repressor involves the side chains of Tyr-85, in addition to those of Ile-84, Met-87 and Tyr-88 (Pabo & Lewis, 1982; Weiss et al., 1987; Kombo et al., 1995). Basically, the aromatic ring of Tyr-85 of one monomer participates in stacking interactions with the aromatic ring of Tyr-88 of the other monomer. To investigate the structural changes that are occurring in the side chain of Tyr-85 as a result of the outward tilt of its backbone carbonyl group observed in active mutants, we determined the number of inter-monomer contacts involving the aromatic ring atoms of Tyr-85 (85') interacting with Tyr-88'(88). The results, shown in Table 3, indicate that in all active mutants with Tyr at position 85, the percentage of the number of such contacts with respect to the total number of contacts made by Tyr-85(85')

varies between 50 and 60%. The active mutant EA⁸⁶YA⁸⁸, in which Tyr-88 is replaced by a smaller side chain, is an exception to this observations.

By contrast, inactive mutants generally show significantly decreased corresponding values (less than 40%). In the inactive mutant IS⁸⁴, however, 79% of the contacts made by Tyr-85 involved its aromatic ring atoms.

In addition to the number of contacts, the results shown in Table 3 also indicate that the side-chain dihedral χ^2 of Tyr-85 and Tyr-85' in active mutants is generally close to 90° while the corresponding values in inactive mutants are significantly decreased and close to 60°. We have also observed that the χ^1 dihedral angle of the residue at position 89 and 89' is always close to -60° for all active mutants (data not shown). Some inactive mutants, by contrast, show values in the trans conformation (180°). It may be that a very well-defined contribution of aromatic ring atoms of Tyr-85, specifically determined by the rotation about the C ^{β} -C ^{γ} bond, is also required for achieving the dimerization required for DNA-binding activity.

2) between the whole monomers of the N-terminal domain of the wild-type repressor protein

The residues which are involved in inter-monomer contacts in the lowest-energy conformation of the whole wild-type N-terminal domain dimer are listed below in decreasing order of their respective number of pairs of atoms in contact, given in parentheses: Ile-84 (129), Met-87 (127), Val-91 (126), Tyr-88 (116), Ala-90 (76), Lys-67 (58), Ala-63 (55), Val-73(47), Tyr-85 (33), Ala-81 (24), Glu-86 (22), Ala-59 (22), Ser-72 (20), Glu-89 (19), Tyr-60 (16), Glu-74 (15), Ala-66 (9) and Glu-83 (7).

The partners in the pairs of inter-monomer contacts for residues of the fifth helices have been determined. Three classes of residues are found: 1) Residues which exclusively contact the fifth helix of the other monomer (Ala-81, Ile-84, Tyr-85 and Tyr-88); 2) residues which make contacts exclusively with the protein core of the other monomer (Glu-83, Glu-86, Glu-89 and Ala-90); and 3) residues which contact both the fifth helix and the protein core of the other monomer (Met-87 and Val-91). This result is consistent with the amphipathic nature of helix-5. The side chains of the

charged residues Glu-83, Glu-86 and Glu-89 in each monomer are solvent-exposed and contact only residues of the protein core of the other monomer, while the hydrophobic residues such as Ala-81, Ile-84, Met-87 exclusively contact residues of the fifth helix of the other monomer.

The numbers of pairs of atoms in inter-monomer contact for residues of the fifth helices are shown in Fig. 6 and 7. A rough correlation appears to exist between this parameter and the tolerance for amino acid substitution at different positions in the region of helix-5 [Reidhaar-Olson & Sauer (1988, 1990)]. The higher the number of atom pairs in close contact for residue *i*, the lower the tolerance for amino acid substitution at that position (Fig. 6). For example, Ile-84, Met-87 and Val-91 contribute to the highest number of contacts and are among the most conserved. Ser-79, Arg-82, Glu-83, Tyr-85, Glu-86 and Glu-89, which have none or a very low number of contacts, tolerate a wide range of amino acid substitutions. Tyr-88 and Ala-90, show intermediate values of the numbers of inter-monomer contacts and also tolerate a medium range of substitutions between the two extremes. Exceptions are noted, however, at positions 78, 80 and 81 which contribute to none or a very low number of inter-monomer contacts, but Pro, Ile and Ala are still very conserved at those positions.

We have also observed that the higher the number of pairs of atoms in close contact for residue i , the higher the tolerance for amino acid substitution at position $i-2$ (Fig.7). For example, the first four highest ordinates in Fig.2 are found at positions 82, 85, 86 and 89. As shown in Fig.6, Ile-84, Met-87, Tyr-88 and Val-91 contribute to very high numbers of inter-monomer contacts. Experimentally, it has been demonstrated by Reidhaar-Olson and Sauer (1988, 1990) that positions 79, 82, 85, 86 and 89 show the highest number of allowed functional substitutions (see Fig.3 of Reidhaar-Olson and Sauer; 1988, 1990; or Fig.16 of Sauer et al., 1990). Likewise, the residues at positions 80, 82, 83, 86 and 89 contribute a very low number of inter-monomer close contacts (Fig.6). Experimentally, it has been demonstrated that residues at positions 78, 80, 81, 84 and 87 are very restricted in tolerating mutations. Exceptions are now found, however, at positions 79, 83 and 90, where the rough correlation breaks down.

4. Discussion

a) Helical distortions at the dimer interface

The results presented in this paper show that the fifth helices of active variants of the N-terminal domain of the λ repressor have their carbonyl groups at position 85 and 85' tilted further outwards from the helix axis than those of inactive mutants. This presumably allows these carbonyls group to accept hydrogen bonds from the solvent. The corresponding carbonyl groups of inactive mutants are relatively buried and make more shorter and linear helical hydrogen bonds than those of the active mutants. These observations are similar to those of Blundell et al. (1983), Baker and Hubbard (1984) and Barlow & Thornton (1988) in their analyses of different samples of X-ray derived crystal structures of helices with carbonyl groups hydrogen-bonded to water or lacking such bonds. Furthermore, the average values of the five pairs of the ϕ , ψ angles in the fifth helices of active mutants were : (i) -69° , -36° ; (i+1) -70° , -45° ; (i+2) -62° , -40° , (i+3) -60° , -40° ; and (i+4) -69° , -34° . The trend in the variations of these values is similar to the one observed in the analysis of 35 protein structures by Sundraralingam and Sekharudu

(1989) for the three-centered hydrogen bonds involving the backbone carbonyl group of residue i , the amide group of residue $i+4$ and a water molecule. In particular, the tendency for distortion in the direction of 3_{10} helix ($\phi=-71^\circ$, $\psi=-18^\circ$) has also been observed.

The large difference in interhelical angles between active and inactive mutants observed by Kombo et al. (1996) can also be compared to the case of troponin C, where hydrogen bonding to water molecules causes a significant difference in interhelical geometry. In fact, in the apo Ca^{+2} binding motifs of troponin C, the interhelical angles of the pairs of helices A/B and C/D have a 10° difference due to the fact that the B helix has inserted water molecules, while the interhelical angles of the holo calcium-binding motifs (E/F and G/H) are the same (Satyshur et al., 1988; Sundralingam & Sekharudu; 1989).

We acknowledge, however, that our current computational study has not included solvent molecules explicitly. *The results presented in this chapter are presumably due to helix packing restraints that mimic the helical distortions caused by solvent.* One of the future directions of this work involves investigations of the same system that include water molecules, in a study similar to the one carried out by DiCapua et al. (1991).

b) Tolerance for amino acid substitution in the fifth helices

It is generally believed that buried residues contribute significantly in determining the structure and function of a protein since they usually do not tolerate mutations very well [Reidhaar-Olson & Sauer (1988, 1990)]. Our previous study (Chap. 1 and Kombo et al., 1996) consistently demonstrated the structural and functional importance of the buried residues Ile-84 and Met-87. Our present results have shown, however, that active and inactive mutants can be discriminated by geometrical parameters around Tyr-85 and Glu-86, two residues that are highly solvent-exposed in the fifth helices of the repressor protein and tolerate a wide range of functional amino acid substitutions.

The requirement that the dihedral angle χ^2 of Tyr-85 and χ^2 of Tyr-85' be close to $+90^\circ$ with the ring lying flat against the main-chain may explain the biological inactivity of the disulfide-linked homodimer mutant YC⁸⁵ which was characterized experimentally by Sauer et al. (1986). Analysis of disulfide bonds has shown that a left-handed spiral conformation and a right-handed hook conformation are especially stable and favorable (Richardson, 1981;

Thornton, 1981). These conformations are characterized by $\chi^2/\chi^{2'}$ values of $-90/-90^\circ$ and $120/-50^\circ$, respectively; which are significantly different from the required $+90^\circ$. Consequently, it seems reasonable to assume that replacing Tyr by Cys at position 85 will result in an active mutant. We do not, however, exclude the possibility that this inactivity may be due to steric constraints by -S-S- bonds.

Reidhaar-Olson & Sauer (1988, 1990) have shown that position 89 shows considerable variability in tolerating functional substitutions. An exception, however, was found with the aromatic residues which were not functionally acceptable at this position. Our analysis of the lowest-energy conformations of active and inactive mutants has revealed that, at position 89, around the $C^\alpha-C^\beta$ bond, the g^- conformation is preferentially favored. This is presumably needed to allow the main-chain N atom to contract at most a weak hydrogen bond with the carbonyl oxygen atom of Tyr-85. Interestingly, an analysis of side-chain conformations within α -proteins, which was carried out by Summers et al. (1987), showed an absence of the g^- conformation for the $C^\alpha-C^\beta$ bond of aromatic residues. This observation was further supported by the study of McGregor et al. (1987) who showed, for example, that the percentage of Trp, Tyr and Phe with the $C^\alpha-C^\beta$ bond in the g^-

conformation at the Carboxy-terminus, is as low as 0, 0 and 4%, respectively. This finding and our observation outlined above may explain why aromatic residues are not acceptable at position 89 of the repressor protein. To test our hypothesis, we have replaced Glu-89 by Tyr and applied the same computational procedure to this inactive mutant structure. We found that, in the lowest-energy conformation obtained, χ^1 of Tyr-89 is around 180° while χ^2 of Tyr-85 is close to -90° in both helices, further supporting our computational predictions.

Reidhaar-Olson & Sauer (1988, 1990) have found a rough correlation between the tolerance for amino acid substitution in helix-5 and the solvent accessibility of the side chains. Pro-78, a highly surface-exposed residue, however, was found to be conserved and Ala-90, a buried residue, was found to tolerate many functional substitutions. We have observed two similar rough correlations between the mutability of a residue of helix-5 and its number of inter-monomer contacts or that of its near neighbor situated two residues apart. Our first correlation (number of close inter-monomer contacts of residue i versus position i' (see Fig.6) correctly predicts that Ala-90 should tolerate a medium number of substitutions. Our second correlation (number of close inter-monomer contacts of residue i versus position $i-2$, see

Fig.7), which actually is a consequence of both the first correlation model and the helical periodicity between exposed and buried residues, correctly predicts that Pro-78 should be conserved. It may be that calculations of solvent-accessibility and inter-monomer van der Waals contact, taken together provide better guidance for predicting the range of functional substitutions which could be tolerated at a given residue position.

Conclusions

Computational studies of the λ repressor-operator complex have enhanced our understanding of structural and functional consequences resulting from amino acid. However, many questions remain and need to be addressed.

First, it is necessary to extend comparative computational studies to other DNA-binding proteins for which extensive experimental data are available. This may reveal the common and system-specific themes underlying the molecular mechanism of sequence-specific recognition.

Second, a real test of how well a macromolecular computer package can predict the three-dimension structure of a λ repressor mutant-operator complex has been made possible by the availability of the X-ray atomic coordinates of the λ repressor triple mutant VL36ML40VI47 (Lim et al., 1994).

Third, as proposed by Sauer et al. (1990), it is important to know to which extent protein-DNA interactions mediated by the flexible N-terminal arm are coupled to those mediated by the helix-turn-helix unit, and likewise, how interactions mediated by the DNA phosphate backbone are coupled to those mediated by bases. It is incumbent to theoreticians to develop appropriate

Methods to address these questions.

Fourth, the applicability of the predictions based on computational studies of the N-terminal domain alone to the entire protein might be problematic. In fact, adding the C-terminal domain increases DNA-binding affinity by several orders of magnitude itself.

Fifth, some mutations at position 88 affect OR1 vs. OR3 discrimination, but not operator vs. Non-operator DNA (personal communication from Dr Donald Senear, University of California at Irvine). It is important to understand the molecular and structural basis of this phenomenon.

Finally, we have seen that computational methods based solely on shape complementarity are able to discriminate between active and inactive mutants of the lambda repressor. However, the complexity of protein-DNA association arises not only from electronic and shape complementarity, but also from DNA conformational deformability, which is mainly due to hydration and salt effect. Therefore, that the steric effect

cannot be compensated by other energetic components of the association such as long range protein-DNA electrostatics and salt effects remains difficult to demonstrate. This yet unresolved question explains why these studies have been limited to predicting the effect of charge-conserving mutations. Likewise, studies based on the ion atmosphere contributions to protein-DNA association are mainly concerned with electrostatics and salt effects, and are therefore more suitable to study the effects due to charge-changing mutations. Furthermore, neglecting the solvent effects in the protein-DNA computer model might also result in significant problems since water molecules can contribute to the steric components of protein-DNA association. A computational method which can include shape complementarity, hydration effects, long range electrostatics and salt dependent effects is needed to understand the full picture of protein-DNA association. Such knowledge may significantly improve our understanding of numerous cellular processes such as DNA replication, transcription, transposition, recombination and DNA packaging in which proteins interact with DNA.

Tables and Figures

A. Chapter 1

TABLE 1. Parameters characterizing the packing structures found after minimizing the energy of the wild-type helix-5...helix-5' complex of the λ repressor DNA operator binding domain.

	ETOT ^a	EINTER ^a	NUMBERS OF CONTACTS	R.M.S. Deviation ^b	R.M.S. Deviation ^c	R.M.S. Deviation ^d	Ω_0^e	Ω_P^e
	-267.1	-34.5	234	0.00	0.70	1.00	-129.2	-129.5
	-266.4	-34.1	224	0.20	0.75	0.96	-131.2	-131.3
	-265.2	-34.5	231	0.19	0.74	0.96	-131.0	-131.2
	-264.7	-33.0	219	0.50	0.74	1.01	-130.6	-130.6
	-264.5	-33.7	216	0.22	0.78	0.94	-131.4	-131.6
	-264.5	-33.9	222	0.18	0.74	0.94	-130.6	-130.8
	-264.4	-32.8	216	0.56	0.86	1.04	-130.9	-130.9
Lower limit ^f	-267.1	-35.4	207	0	0	0	-132.5	-132.7
Upper limit ^f	-264.4	-31.9	243	0.84	0.94	1.05	-128.1	-128.4

Footnotes to Table 1

^a ETOT is the total energy of the system in kcal/mol, consisting of the total intramolecular energy of each helix and the intermolecular energy, EINTER. All conformations with $\Delta\text{ETOT} < 5.0$ kcal/mol are listed, with $\Delta\text{ETOT} = \text{ETOT} - \text{ETOT}_0$, where ETOT_0 is the energy of the lowest-energy minimum for each sequence.

^b (In angstrom units) between the given conformation and the lowest-energy conformation of the standard geometry model of the wild-type protein (first row).

^c (In angstrom units) between the given conformation and the X-ray conformation of the free protein conformation. The lower limit has been extended from 0.62 to 0 to include conformations with the lowest deviation.

^d (In angstrom units) between the given conformation and the X-ray conformation of the DNA-bound protein conformation. The lower limit has been extended from 0.91 to 0 to include conformations with the lowest deviation.

^e In degrees.

^f See text for explanation of calculation of these limits.

TABLE 2. Parameters characterizing the packing structures found after minimizing the energy of mutants of the helix-5...helix-5' complex of the λ repressor DNA operator binding domain.

	ETOT ^a	EINTER ^a	NUMBERS OF CONTACTS	R.M.S. Deviation ^b	R.M.S. Deviation ^c	R.M.S. Deviation ^d	Ω_0^e	Ω_P^e
1. ACTIVE (a) class 1								
YF88	-259.9	-33.6	234	0.34	0.76	0.86	-129.2	-129.5
VT91	-282.8	-33.4	223	0.47	0.61	0.91	-130.2	-130.5
	-278.2	-33.1	220	0.48	0.62	0.88	-128.4	-128.5
ALA-91								
AW90VA91	-273.6	-33.8	212	0.68	0.84	0.82	-132.5	-132.6
	-271.2	-35.9	239	0.55	0.95	0.85	-130.3	-130.8
	-270.6	-32.0	197	0.61	0.77	0.85	-131.4	-131.6
	-270.6	-32.6	211	0.56	0.74	0.88	-130.9	-131.2
	-270.2	-32.1	198	0.50	0.77	0.84	-131.3	-131.4

Table 2. (Continued)

	-269.1	-34.4	224	0.54	0.92	0.95	-129.0	-129.7
	-268.8	-34.3	224	0.53	0.92	0.94	-128.9	-129.6
AV90VA91	-266.5	-32.6	210	0.41	0.85	0.89	-130.1	-130.6
POSITION 78								
PA78	-229.0	-35.0	241	0.34	0.79	0.88	-130.6	-130.9
PF78	-239.3	-35.1	240	0.34	0.79	0.88	-130.6	-130.9
PY78	-239.4	-35.0	240	0.34	0.78	0.89	-130.5	-130.8
	-239.4	-35.1	240	0.34	0.78	0.89	-130.5	-130.8
	-239.1	-35.1	240	0.34	0.78	0.89	-130.5	-130.8
	-239.0	-35.2	240	0.34	0.78	0.89	-130.6	-130.8
PT78	-239.4	-35.0	241	0.34	0.78	0.88	-130.5	-130.8
	-239.3	-35.1	241	0.34	0.78	0.88	-130.5	-130.8
	-239.1	-35.1	241	0.34	0.78	0.88	-130.5	-130.8
	-239.0	-35.2	241	0.34	0.79	0.88	-130.6	-130.8
POSITION 87								
ML87	-262.3	-31.9	230	0.58	0.96	1.10	-132.8	-133.1
	-261.7	-32.9	220	0.50	0.91	1.09	-130.4	-131.2
	-260.4	-32.9	223	0.43	0.91	1.15	-130.2	-131.1

Table 2. (Continued)

	-259.5	-32.8	226	0.51	0.90	1.02	-131.1	-131.1
	-259.4	-33.3	224	0.12	0.71	0.98	-130.0	-130.0
	-258.6	-32.5	220	0.34	0.72	0.87	-129.6	-129.9
	-257.2	-31.4	210	0.63	0.98	1.08	-130.5	-131.7
b) Class 2								
ML87YF88	-261.4	-30.6	223	0.57	0.96	1.09	-132.7	-133.0
	-259.0	-31.6	218	0.70	0.90	1.00	-130.9	-130.9
	-258.7	-31.8	218	0.49	0.91	1.06	-129.8	-130.8
	-258.4	-31.8	218	0.35	0.70	0.86	-129.3	-129.6
ALA-88								
YA88	-242.6	-24.7	156	0.83	1.10	1.27	-128.6	-128.9
	-242.0	-25.7	168	0.84	1.10	1.25	-128.2	-128.4
	-241.4	-26.6	171	0.87	1.09	1.25	-127.5	-127.7
	-238.3	-29.0	189	0.92	1.23	1.27	-131.5	-131.8
YW85YA88	-251.2	-37.2	201	1.29	1.76	1.49	-145.8	-146.2
	-249.5	-29.9	195	0.80	1.06	1.18	-130.3	-130.5
	-249.5	-33.7	185	1.02	1.47	1.35	-141.2	-141.4
	-248.4	-27.6	176	0.81	1.09	1.20	-130.0	-130.2

Table 2. (Continued)

	-248.3	-26.4	165	0.76	1.11	1.17	-133.1	-133.3
2. INACTIVE (a) class 1								
IA84MV87YA88	-235.0	-21.7	116	1.41	1.84	1.60	-148.9	-149.1
IT84MT87YV88I	-251.2	-25.7	164	2.73	2.68	2.41	-136.3	-136.5
	-242.8	-17.9	139	1.89	1.73	2.04	-137.0	-138.5
WT (1-90)9	-248.5	-27.2	179	0.60	0.52	0.76	-126.3	-126.3
	-247.0	-27.2	178	0.59	0.51	0.76	-126.2	-126.6
	-245.5	-27.2	183	0.59	0.50	0.76	-126.2	-126.6
	-245.5	-27.2	182	0.59	0.51	0.76	-126.2	-126.6
	-245.5	-27.2	178	0.59	0.51	0.75	-126.4	-126.8
	-245.5	-27.2	182	0.59	0.50	0.77	-126.1	-126.5
	-243.9	-30.9	204	0.62	0.55	0.78	-126.5	-126.8
MV87	-250.7	-31.5	211	0.58	1.03	1.12	-130.9	-131.7
	-249.7	-31.8	217	0.55	0.96	1.12	-129.4	-130.3
	-248.2	-31.8	227	0.69	0.95	1.07	-131.4	-131.4
	-247.5	-35.7	238	0.82	1.03	1.09	-130.1	-130.4
	-245.8	-31.8	216	0.56	1.00	1.09	-130.6	-131.4
	-245.7	-36.6	246	0.82	1.18	1.10	-133.2	-133.4

Table 2. (Continued)

MI87	-248.8	-31.4	209	0.63	1.14	1.09	-134.0	-134.9
	-247.9	-29.4	209	0.50	0.96	1.11	-133.3	-133.6
	-246.6	-31.4	211	0.60	1.07	1.12	-131.9	-132.9
	-246.1	-31.5	206	0.59	1.08	1.08	-133.0	-133.7
(b) class 2								
IS84	-268.2	-32.9	185	0.79	0.89	1.20	-126.4	-126.5
	-266.6	-33.6	194	0.88	0.92	1.25	-125.2	-125.2
	-266.6	-33.6	190	0.84	0.91	1.25	-125.3	-125.4
	-266.0	-33.8	193	0.90	0.94	1.30	-125.1	-125.1
	-265.7	-33.4	197	0.82	0.90	1.22	-126.3	-126.3
	-265.6	-34.1	206	0.89	0.93	1.25	-125.3	-125.3
	-265.6	-34.2	196	0.95	0.95	1.30	-123.3	-123.4
	-265.6	-33.7	197	0.83	0.91	1.24	-125.2	-125.7
	-265.6	-34.1	202	0.86	0.92	1.26	-125.2	-125.3
IL84	-266.8	-33.1	215	0.89	1.00	1.30	-127.7	-127.6
	-264.9	-33.3	209	0.61	0.96	1.23	-127.1	-128.5
	-264.4	-33.1	211	0.35	0.93	0.95	-134.0	-134.2
	-262.2	-36.1	239	0.92	1.20	0.73	-135.4	-135.8

Table 2. (Continued)

	-261.2	-30.5	200	0.81	0.84	1.32	-124.0	-124.1
IM84	-273.3	-39.6	247	1.12	1.27	1.32	-131.9	-131.9
	-272.5	-40.2	202	2.04	2.34	2.28	-155.0	-155.1
	-271.8	-41.5	211	2.18	2.54	2.38	-158.5	-159.2
	-271.4	-38.6	216	1.90	2.27	2.23	-151.0	-151.0
	-271.2	-34.2	185	1.06	1.47	1.51	-135.8	-138.3
	-269.7	-38.3	215	1.16	1.45	1.17	-143.1	-143.1
IT84	-271.3	-34.1	220	0.70	0.87	1.12	-126.2	-126.3
	-266.9	-34.1	221	0.66	0.87	1.06	-127.0	-127.1
	-266.7	-33.0	203	0.49	0.83	0.96	-129.6	-129.8
MT87	-265.0	-31.9	203	0.52	0.93	1.08	-129.3	-130.0
	-260.9	-32.0	206	0.52	0.93	1.09	-129.1	-129.9
	-260.8	-32.0	200	0.51	0.95	1.05	-130.4	-131.1
	-256.7	-32.2	200	0.51	0.95	1.07	-130.0	-130.7
VY911	-268.3	-41.8	241	1.50	1.44	1.20	-134.4	-134.4
	-261.1	-38.2	233	1.80	1.34	1.81	-122.9	-123.1
MF87	-233.0	-45.7	291	2.46	2.74	2.32	-107.0	-107.0
IF84	-259.3	-42.2	212	3.05	2.77	3.01	-106.4	-107.6

Table 2. (Continued)

IV84	-259.1	-33.7	224	0.43	0.84	0.84	-131.0	-131.3
	-255.5	-32.6	204	0.39	0.76	0.97	-129.8	-130.3
	-255.4	-33.9	216	0.48	0.89	0.95	-131.0	-131.2
	-254.1	-35.5	249	0.56	0.86	0.82	-130.6	-130.7
	-246.1	-31.5	206	0.59	0.86	0.82	-133.0	-133.

FOOTNOTES TO TABLE 2

^a ETOT is the total energy of the system in kcal/mol, consisting of the total intramolecular energy of each helix and the intermolecular energy, EINTER. All conformations with $\Delta ETOT < 5.0$ kcal/mol are listed, with $\Delta ETOT = ETOT - ETOT_0$, where $ETOT_0$ is the energy of the lowest-energy minimum for each sequence.

^b (In angstrom units) between the given conformation and the lowest-energy conformation of the standard geometry model of the wild-type protein (first row data of Table 1).

^c (In angstrom units) between the given conformation and the X-ray conformation of the free protein.

^d (In angstrom units) between the given conformation and the the X-ray conformation of the DNA-bound protein.

^e In degrees.

^f For the inactive mutants IT84MT87YV88 and VY91, conformations with $\Delta ETOT < 10.0$ kcal/mol are listed.

^g WT (1-90) represents the inactive mutant 1-90 fragment, which results from the removal of Val-91 and Ser-92 from the wild-type protein by partial proteolysis (Weiss et al., 1987).

TABLE 3. Trends in the variations of the intermolecular energy and the experimentally determined dimer dissociation constants^a.

	EINTER ^b	EINTER ^c	K _d ^d
WT (1-90)	-27.2	-78.1	> 1 x 10 ⁻¹
	-27.2	-78.9	
		-78.8	
		-79.1	
IS84	-32.9	-87.9	> 2 x 10 ⁻²
	-33.6	-88.5	
	-33.6	-88.6	
Wild-type	-34.5	-96.6	6 x 10 ⁻⁴
		-97.0	

FOOTNOTES TO TABLE 3

- ^a All conformations with $\Delta E_{TOT} < 2.0$ kcal/mol are listed, with $\Delta E_{TOT} = E_{TOT} - E_{TOT_0}$, where E_{TOT_0} is the energy of the lowest-energy minimum for each sequence.
- ^b Intermolecular energy of the h5-h5' complex, in kcal/mol.
- ^c Intermolecular energy of the h4-loop-h5...h4'-loop-h5' complex, in kcal/mol.
- ^d Dimer dissociation constant, in moles/liter (Weiss et al., 1987).

Table 4. Parameters characterizing the packing structures of helices 3 and 3' found after minimizing the energy of mutants of the N-terminal domain of the λ repressor.

	ETOT ^a	EINTER ^a	Total numbers of contacts ^b	Numbers of contacts made by residues at positions 84, 87 and 91, respectively.			Ω_0^c	Ω_P^c	D ^d
1. ACTIVE a) class 1									
1. Wild type	-1455.8	-90.7	493	129	127	126	-119.2	-119.2	30.4
2. YF ⁸⁸	-1455.2	-89.5	483	132	128	128	-119.2	-119.3	30.4
3. VT ⁹¹	-1472.9	-96.0	518	133	134	145	-116.4	-116.4	29.7
b) class 2									
1. YA ⁸⁸	-1431.2	-78.7	445	124	132	142	-118.8	-118.9	30.6
2. INACTIVE a) class 1									
1. IA ⁸⁴ MV ⁸⁷ YA ⁸⁸	-1412.1	-71.1	384	54	90	113	-123.1	-123.2	28.9
2. wild type (1-90) ^e	-1431.2	-75.5	379	79	130	0	-113.6	-113.6	29.9

Table 4. (Continued)

3. IV ^{84I}	-1450.6	-89.5	467	99	105	125	-118.8	-118.9	30.6
b) class 2									
1. IT ⁸⁴ MT ⁸⁷ YV ⁸⁸	-1455.9	-99.4	538	76	99	80	-126.1	-126.5	26.7
2. MI ⁸⁷	-1468.0	-113.9	635	118	128	115	-109.4	-109.4	22.1
3. IL ⁸⁴	-1482.0	-122.7	612	93	91	120	-109.4	-109.4	22.1
4. VY ⁹¹	-1485.4	-100.9	513	78	109	147	-117.7	-117.8	31.0

Footnotes to Table 4

^a ETOT is the total energy, in Kcal/mol, consisting of the total intramolecular energy of each monomer and the intermolecular energy, EINTER. Only the lowest-energy conformation is listed, for each sequence, since no significant difference in energy and geometrical parameters is observed for conformations with ETOT within 5 kcal/mol above the lowest-energy minimum.

^b Van der Waals close inter-monomer contacts.

^c Interhelical angle between the recognition helices, in degrees.

^d Closest distance of approach between the recognition helices (numbers 3 and 3'), in angstrom units.

^e Wild type (1-90) represents the inactive mutant 1-90 fragment, which results from the removal of Val-91 and Ser-92 from the wild-type protein by partial proteolysis (Weiss et al., 1987).

^f The starting side-chain conformations for Val were taken from Piela et al. (1987) and from Vásquez et al. (1983) for the α -helical backbone conformation.

Table 5. Summary of scores of criteria for discrimination between active and inactive mutants of the helix-5...helix-5' complexes.

A. Single criterion

	R.M.S Deviation ^a	Ω_0	Ω_p	Numbers of Contacts	R.M.S Deviation ^b	R.M.S Deviation ^c	Einter
Active ^d	90	88	88	75	75	70	68
Inactive ^e	36	70	71	61	55	50	59

B. Combined criteria^f

Ω_0 and Ω_p	R.M.S Deviation ^a	R.M.S Deviation ^a and numbers of contacts	Numbers of Contacts	R.M.S Deviation ^a and R.M.S Deviation ^c	R.M.S Deviation ^c	R.M.S Deviation ^c and numbers of contacts
Active ^d	85	70	68	65	65	60
Inactive ^e	75	86	88	89	91	96

Footnotes to Table 5

^a Same as in footnotes of Tables 1-2.

^b Same as in footnotes of Tables 1-2.

^c Same as in footnotes of Tables 1-2.

^d There are in total 40 minimum-energy conformations for all the 12 active mutants (Table 2). The raw data are the percentage of the number of positif scores, i.e. a conformation of an active mutant is discriminated as active. Parameters in A) are given from left to right in the *decreasing* order of their scores for success in discriminating *active* mutants.

^e There are in total 56 minimum-energy conformations for all the 14 inactive mutants (Table 2). The raw data are the percentage of the number of positif scores, i.e. a conformation of an inactive is discriminated as inactive. Parameters in B) are given from left to right in the *increasing* order of their scores for success in discriminating *inactive* mutants.

^f The criteria for discrimination are a combination of the interhelical angles Ω_0 and Ω_p , together with the parameter (or combination of parameters) given in the first row of B).



Fig.1.

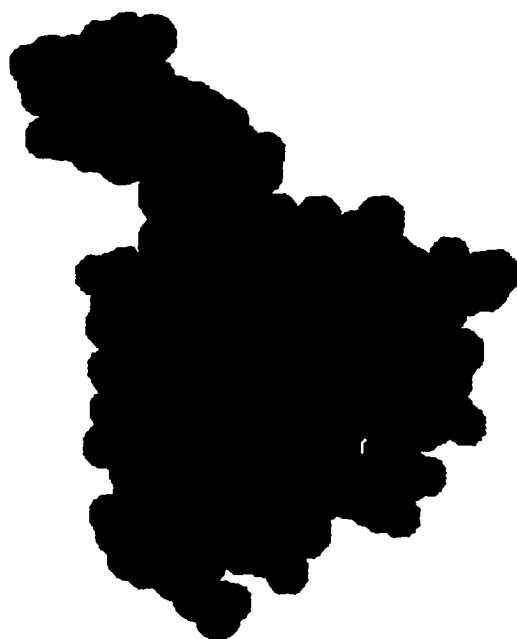


Fig.2. a.

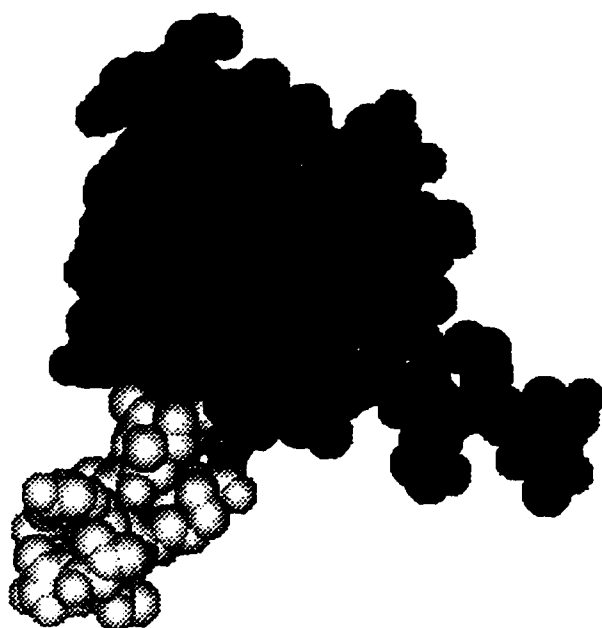


Fig.2. b.

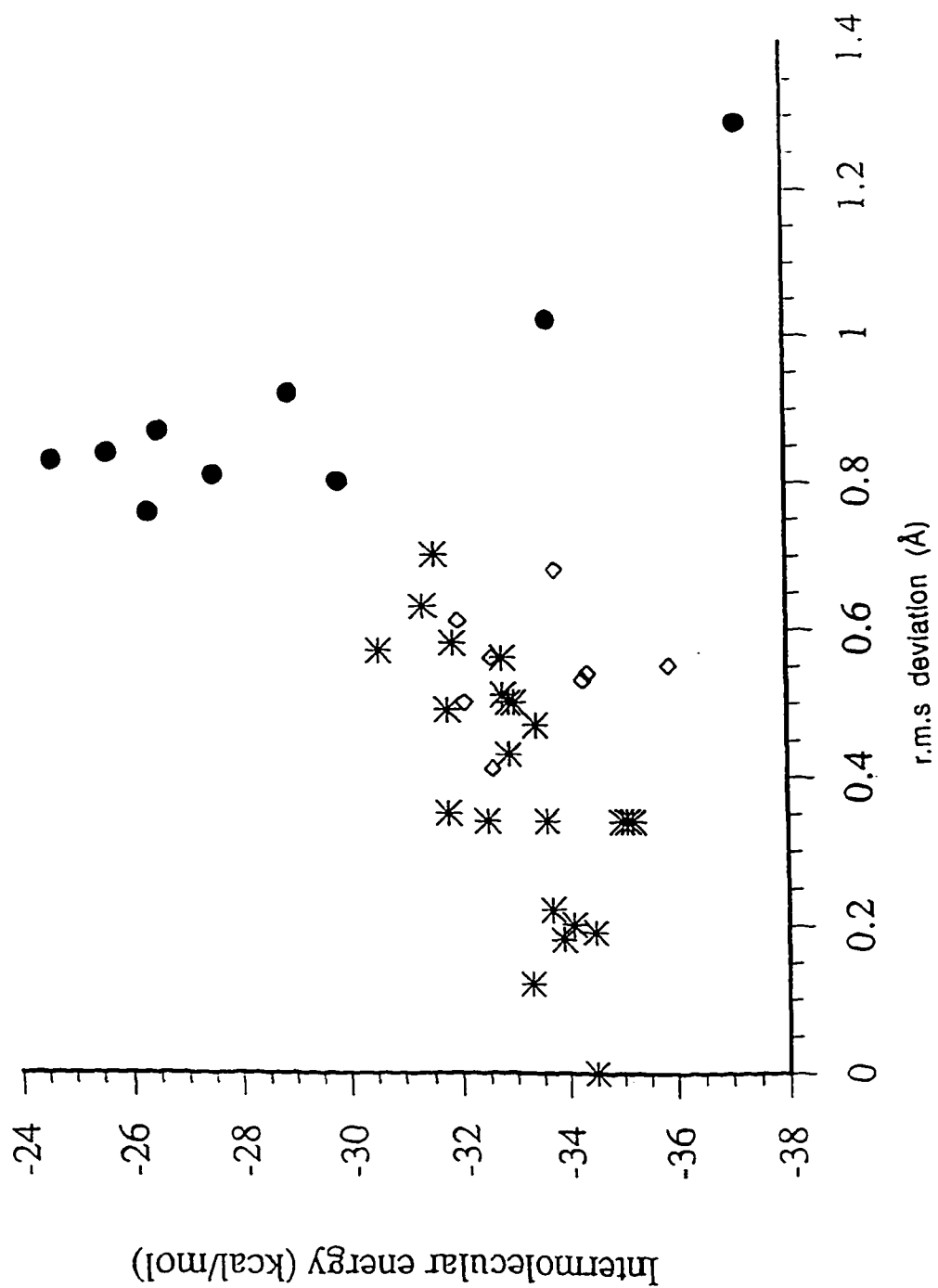


Fig.3.

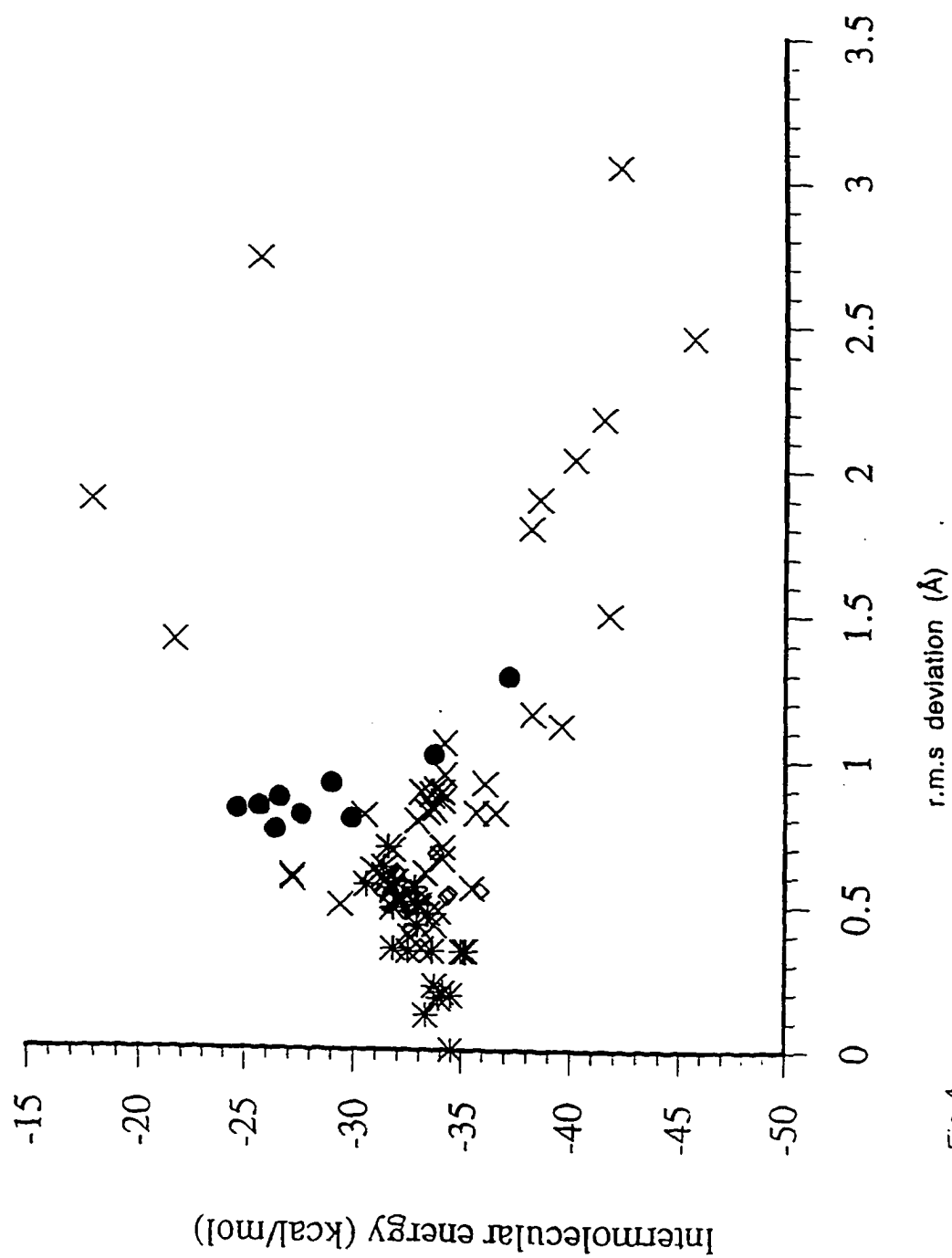


Fig.4.

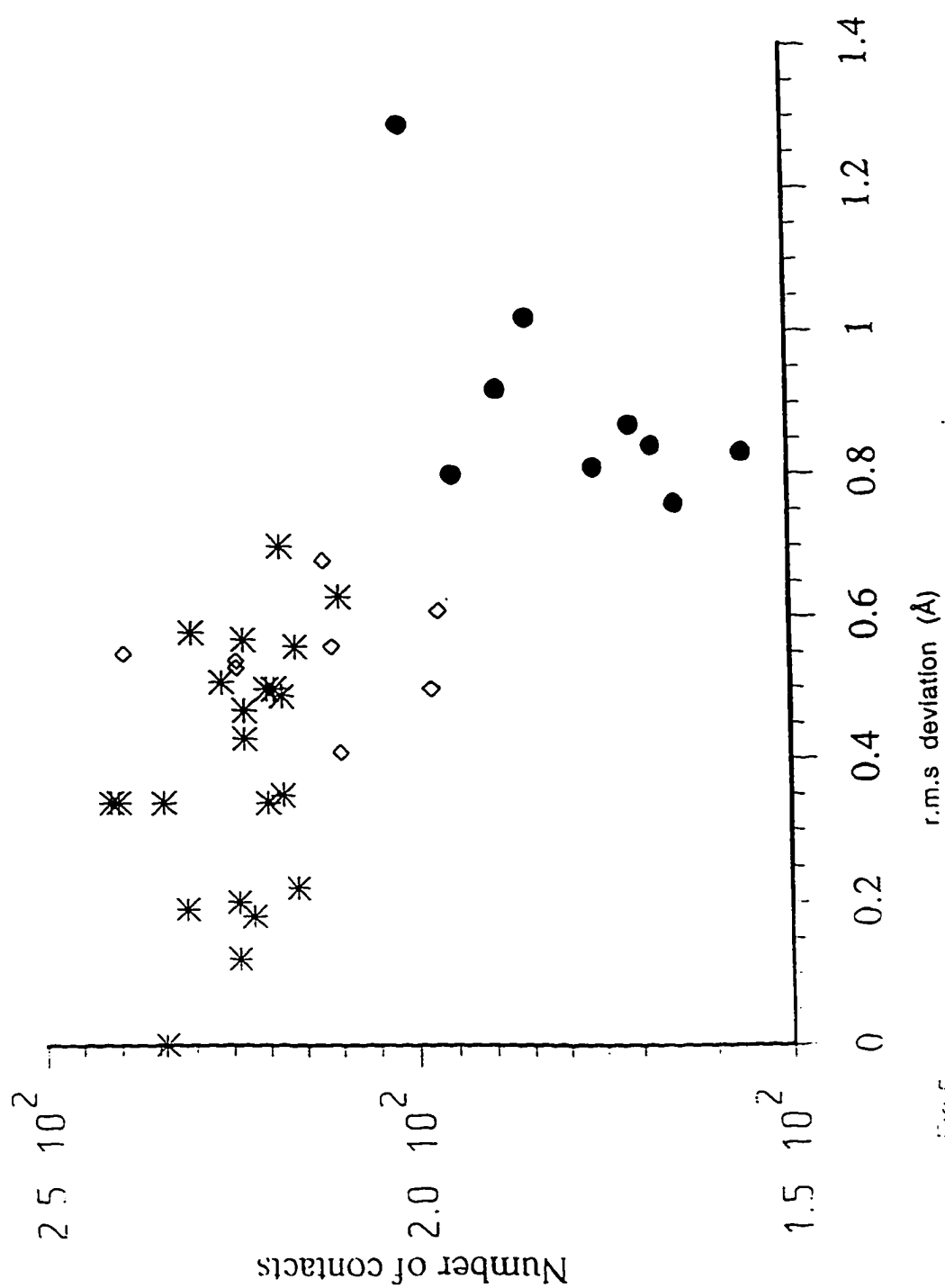


Fig 5

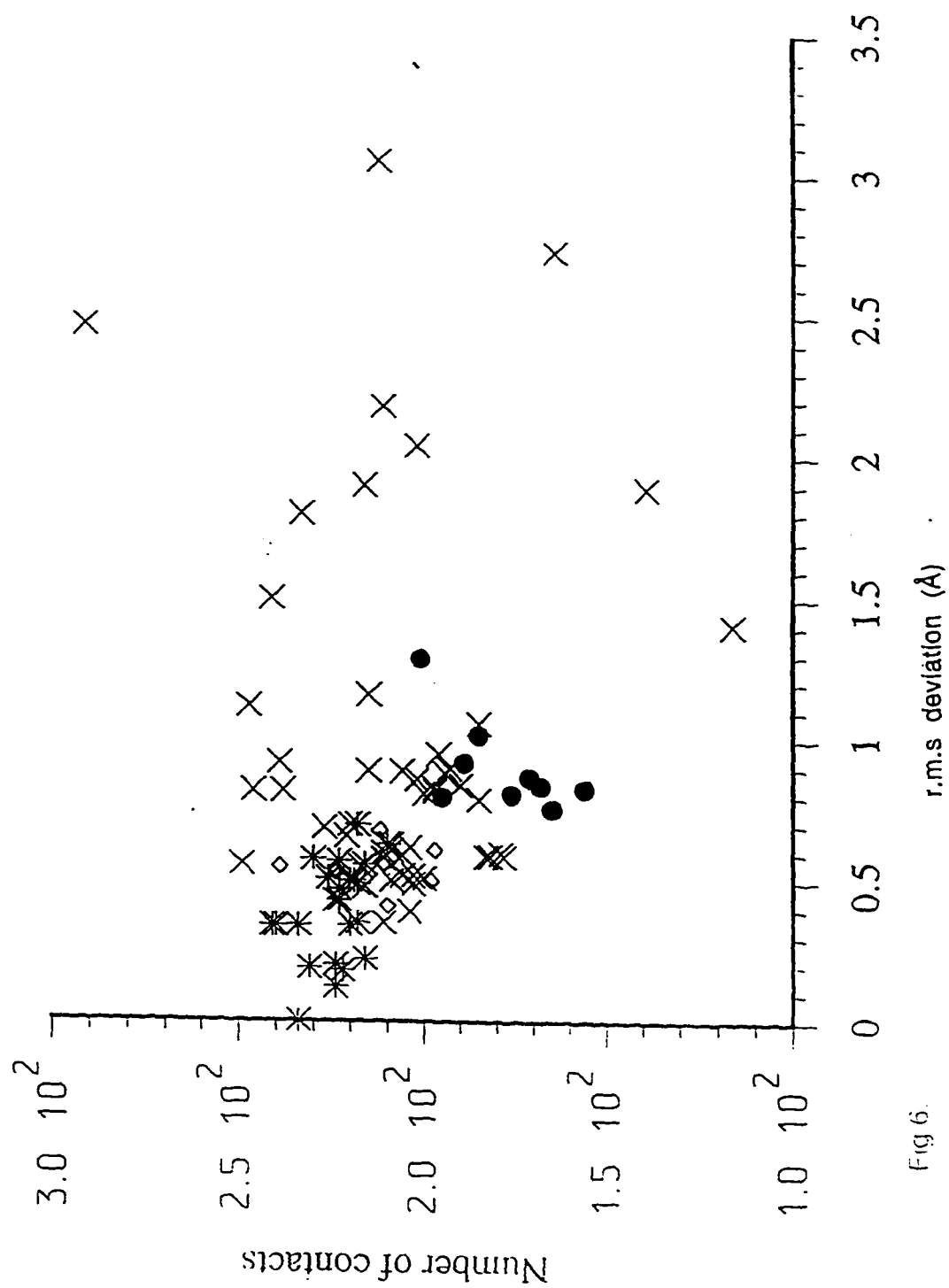


Fig 6.

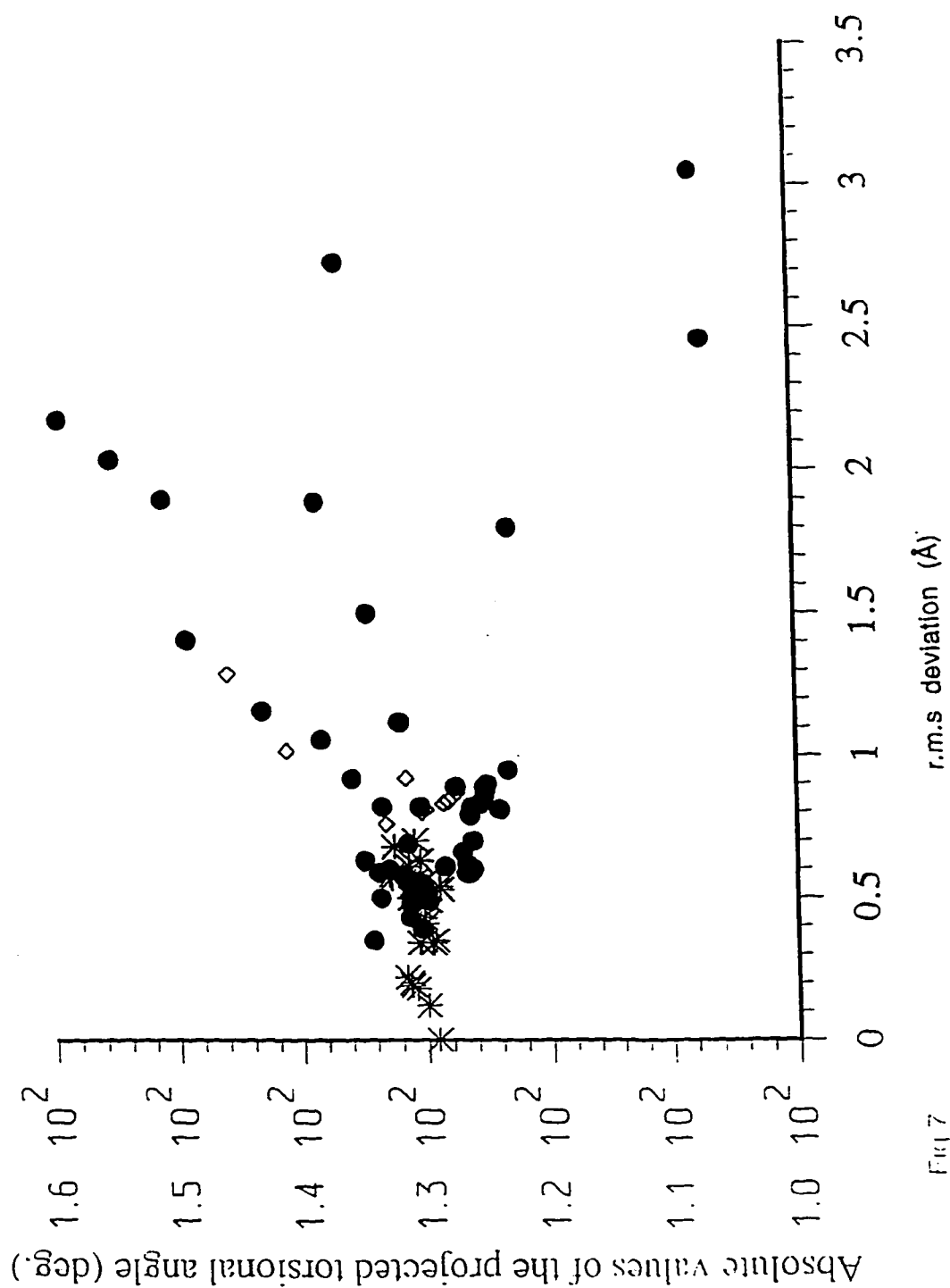


Fig 17

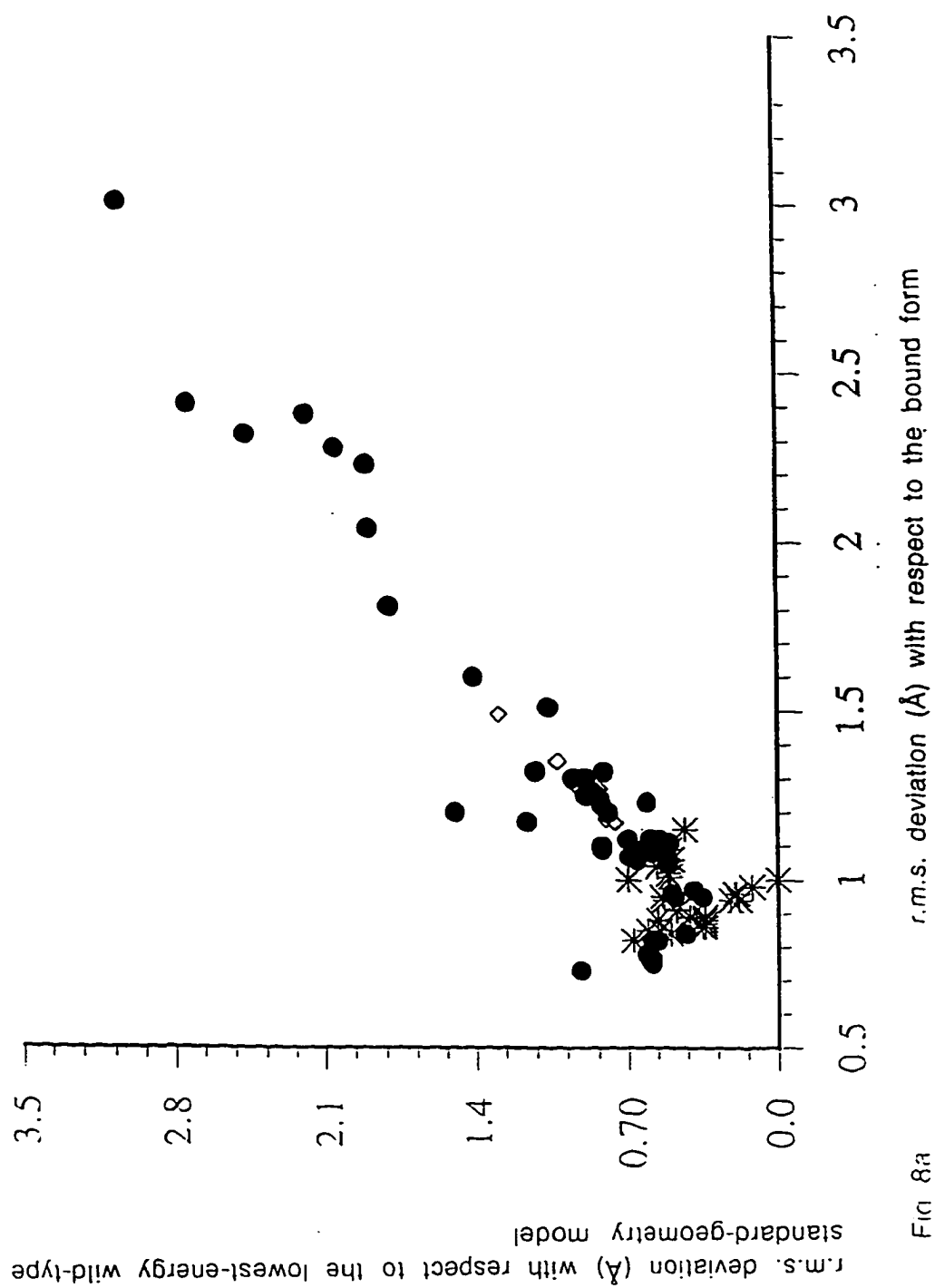


Fig 8a

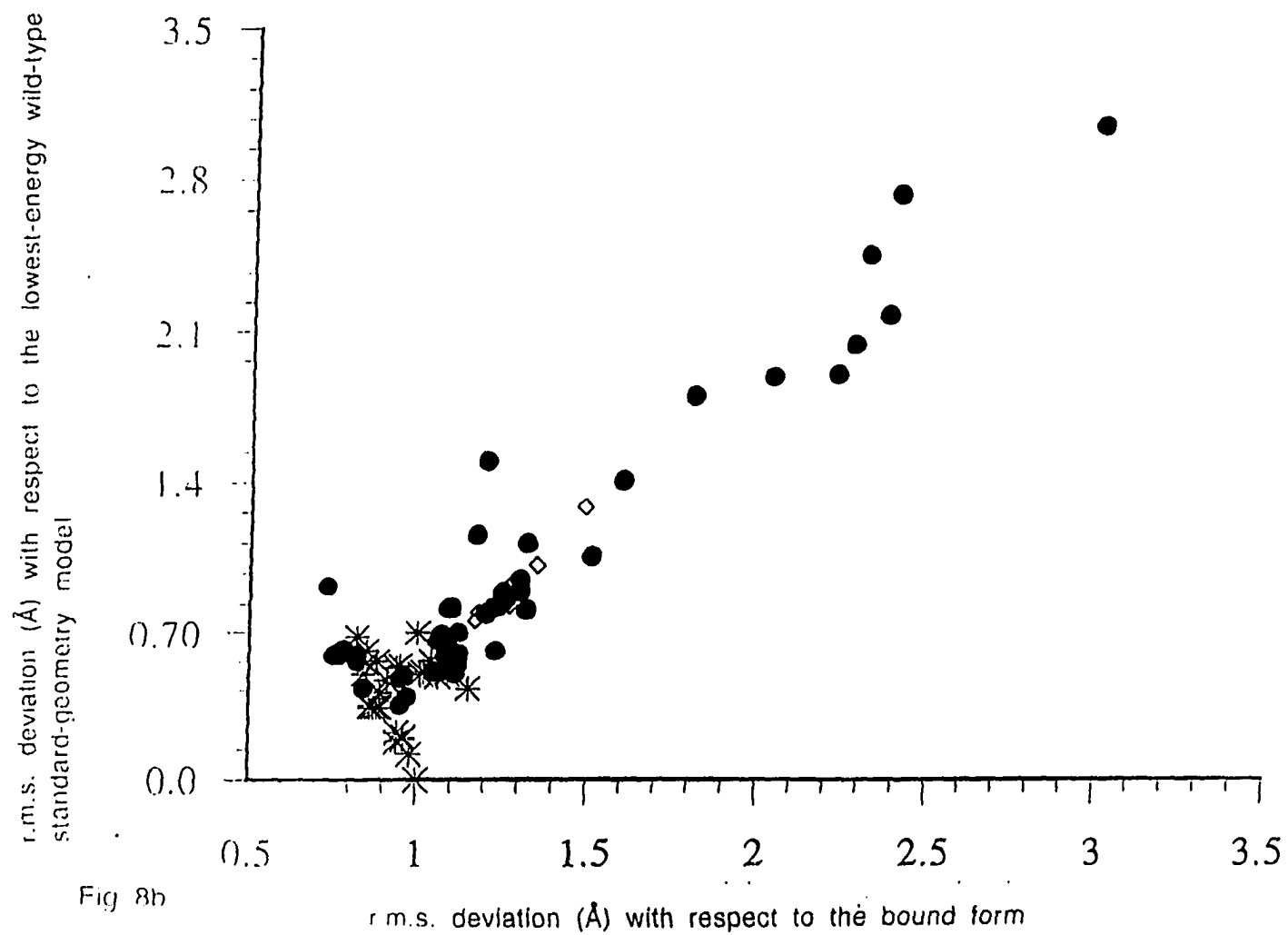


Fig 8b

B. Chapter 2

Table 1. Minimum-energy conformations of the N-acetyl-N'-methylanide of 5-hydroxytryptophan

Code ^a	ΔE^b	v^c	ΔE_{es}^d	ΔE_{nb}^d	ΔE_{tar}^d	ϕ^e	ψ^e	χ^{1e}	χ^{2e}	χ^{3e}
Eg ⁺	0.000	0.135	0.000	0.000	0.000	-155	156	60	93	0
Eg ⁺	0.079	0.118	0.003	0.067	0.008	-154	157	58	-96	1
Et	0.363	0.073	-0.096	0.453	0.006	-153	150	-180	-91	0
Eg ⁻	0.551	0.053	0.066	0.483	0.001	-144	153	-61	93	-1
Eg ⁺	0.708	0.041	0.187	0.895	0.000	-155	157	60	93	-180
Ft	0.760	0.038	0.204	0.538	0.016	-78	145	-177	-91	0
At	0.834	0.033	0.537	0.294	0.000	-74	-36	-180	-93	0
Et	0.850	0.032	-0.058	0.909	0.000	-153	150	180	77	0
Ag ⁺	0.874	0.031	0.405	0.206	0.142	-74	-36	-180	-93	0
Dg ⁻	0.935	0.028	0.232	0.702	0.001	-141	42	-59	92	0
Ct	0.947	0.027	-0.869	1.774	0.431	-80	80	-177	76	0
Cg ⁻	0.978	0.026	-0.737	1.677	0.037	-83	75	-60	103	0
Eg ⁻	0.988	0.026	-0.737	1.677	0.037	-83	75	-60	103	0
Ct	1.010	0.025	-0.795	1.770	0.035	-80	81	-178	-112	1
Cg ⁻	1.057	0.023	-0.718	1.744	0.031	-83	76	-60	-78	0

Table 1. (Continued)

Eg ⁺	1.110	0.021	0.146	0.959	0.004	-153	157	58	-96	179
Dg ⁺	1.138	0.020	0.141	0.936	0.060	-144	154	-60	-78	0
Et	1.345	0.014	0.023	1.302	0.020	-153	157	58	-96	179
Eg ⁻	1.385	0.013	0.023	1.342	0.008	-144	154	-61	93	-177
Dg ⁻	1.437	0.012	0.213	1.202	0.021	-141	41	-56	-77	1
Fl	1.441	0.012	0.233	1.208	-0.001	-78	142	-180	77	0
Fg ⁺	1.490	0.011	0.362	0.922	0.205	-76	146	-71	89	0
At	1.528	0.010	0.344	1.176	0.008	-74	-37	-180	-93	178
Ag ⁺	1.573	0.009	0.464	1.027	0.081	-79	-27	67	-100	-1
Al	1.593	0.009	0.559	1.032	0.001	-75	-36	-179	80	-1
Dg ⁻	1.687	0.008	0.122	1.561	0.003	-141	40	-58	92	-18
Eg ⁻	1.716	0.007	-0.068	1.762	0.022	-144	154	-61	-77	175
Ag ⁻	1.731	0.007	0.561	1.164	0.005	-80	-33	-59	104	1
Gg ⁻	1.769	0.007	0.719	0.989	0.060	-145	-58	-65	94	-1
Fl	1.774	0.007	0.366	1.382	0.025	-79	146	-177	-91	-178
Cg ⁻	1.782	0.006	-0.777	2.520	0.039	-84	75	-60	103	179
Cl	1.803	0.006	0.587	1.177	0.055	-157	-57	174	-103	-1
Gl	1.820	0.006	0.587	1.177	0.055	-157	-57	174	-103	-1

Table 1. (Continued)

Ag ⁻	1.859	0.006	0.595	1.261	0.002	-79	-33	-60	-78	1
E1	1.894	0.005	0.138	1.741	0.014	-153	149	178	77	177
Cg ⁻	1.895	0.005	-0.717	2.580	0.032	-82	76	-60	-78	179
Dg ⁺	1.912	0.005	0.257	1.643	0.011	-150	37	61	108	-1
Ag ⁺	1.957	0.005	0.715	1.102	0.139	-74	-33	69	88	179
Fg ⁻	1.962	0.005	0.226	1.716	0.018	-90	142	-58	102	0
C1	1.992	0.004	-0.661	2.612	0.042	-80	80	-178	-113	177
Fg ⁻	1.995	0.004	0.282	1.704	0.008	-85	142	-59	-78	1
Fg ⁺	2.087	0.004	0.382	1.631	0.072	-85	153	66	-100	0
Dg ⁺	2.147	0.003	0.239	1.854	0.053	-142	32	56	-91	-178
Fg ⁺	2.212	0.003	0.186	1.813	0.212	-76	146	71	89	-178
G1	2.230	0.003	0.615	1.558	0.056	-157	-58	174	79	0
F1	2.306	0.002	0.255	2.052	-0.001	-77	142	-180	78	180
Dg ⁻	2.311	0.002	0.242	2.036	0.032	-141	42	-60	-77	176
Ag ⁺	2.370	0.002	0.410	1.867	0.093	-79	-27	67	-100	-178
Gg ⁻	2.395	0.002	0.702	1.649	0.044	-146	-58	-64	-77	1
Ag	2.407	0.002	0.397	2.003	0.006	-80	-34	-59	104	180
Gg ⁺	2.436	0.002	0.573	1.767	0.095	-163	-49	53	98	0

Table 1. (Continued)

Gg-	2.448	0.002	0.527	1.865	0.055	-145	-58	-65	94	-177
A1	2.456	0.002	0.577	1.860	0.018	-74	-36	-179	81	-176
Gl	2.484	0.002	0.376	2.050	0.057	-157	-57	175	-103	-178
A*g-	2.703	0.001	0.470	2.106	0.125	54	46	-53	106	0
Fg-	2.753	0.001	0.209	2.530	0.013	-85	143	-59	-79	177
Fg-	2.786	0.001	0.216	2.552	0.018	-90	143	-58	101	-179
Ag-	2.812	0.001	0.710	2.091	0.010	-79	-33	-60	-78	177
Fg+	2.887	0.001	0.348	2.469	0.074	-83	152	66	-100	179
Dg+	2.950	0.000	0.457	2.479	0.013	-150	38	61	109	-178
A*g-	3.211	0.000	0.462	2.586	0.162	56	45	-51	-79	-1
Gl	3.269	0.000	0.812	2.377	0.079	-157	-58	174	79	-178
Gg-	3.305	0.000	0.769	2.462	0.072	-146	-58	-65	-77	174
Gg+	3.353	0.000	0.626	2.605	0.121	-163	-50	53	98	176
A*g-	3.835	0.000	0.732	2.989	0.114	54	46	-53	106	179
A*1	3.875	0.000	0.516	3.178	0.180	54	49	-170	-114	1
Gg+	3.879	0.000	0.661	3.108	0.110	-156	-47	61	-91	0
A*1	3.946	0.000	0.414	3.250	0.282	54	49	-167	74	0
A*g-	4.027	0.000	0.442	3.405	0.179	56	45	-56	-78	-177

Table 1. (Continued)

A*t	4.691	0.000	0.319	4.090	0.281	54	49	-168	74	178
Gg*	4.938	0.000	0.798	4.011	0.129	-155	-47	62	-91	-177
A*t	4.996	0.000	0.779	4.022	0.194	54	48	-170	-115	176

Footnotes to Table 1

- ^a Conformational letter code (Zimmerman et al., 1977), referring to the backbone and χ^1
- ^b All minimum-energy conformations with $\Delta E < 5.0$ kcal/mol are listed; $\Delta E = E - E_0$, where $E_0 = -10.2$ kcal/mol, is the energy of the lowest-energy minimum.
- ^c Computed with equation 1.
- ^d These are the electrostatic, nonbonded, and torsional energy components, respectively, in kcal/mol.
- ^e In degrees. The backbone dihedral angle ω was kept fixed at a value of 180° , and ϕ , ψ , χ^1 , χ^2 and χ^3 were varied.

Table 2. Minimum-energy side-chain conformation in an α -helix^a

Residue	Code ^b	ΔE^c	ν^d	χ^{1e}	χ^{2e}	χ^{3e}
Trp ^l	t	0.000	0.9290	174	90	
	t	1.554	0.0690	174	-109	
	g-	4.304	0.0007	-86	120	
5-OH-Trp	t	0.000	0.8010	176	-91	1
	t	1.013	0.1480	176	-92	175
	t	1.801	0.0400	176	74	-1
	t	2.620	0.0100	176	74	-177

Footnotes to Table 2

^a Sequence used : CH₃CO-L-ALA₉-(5OH-Trp)-L-ALA₉-NHCH₃

^b Conformation letter code referring to χ^1

^c All minimum-energy conformations with $\Delta E < 5.0$ kcal/mol are listed; $\Delta E = E - E_0$, where $E_0 = -53.5$ kcal/mol, is the energy of the lowest-energy minimum. All side-chain starting conformations were chosen from those listed in Table 1 and correspond to the A region, with $\Delta E < 3.0$ kcal/mol.

^d Computed with equation 1.

^e In degrees. The backbone was kept fixed with : $\phi = -68.1^\circ$, $\psi = -38.3^\circ$, $\omega = 180.0^\circ$.

^f The results for Trp were taken from the work of Vásquez et al. (1983) and were reproduced by us as a control, using ECEPP/3, the latest version of the ECEPP algorithm.

Table 3. Parameters characterizing the packing states found after minimizing the energy of mutants of the h5...H5' complex of the λ repressor DNA operator binding domain.

	ETOT ^a	EINTER ^a	NUMBER OF CONTACTS ^b	R.M.S. DEVIATION ^c	Ω_0^d	Ω_p^d
1. A(5OH-W) ⁹⁰ VA ⁹¹	-284.6	-33.7	208	0.53 0.19	-131.3	-131.5
	-283.9	-33.9	211	0.52 0.19	-131.3	-131.5
	-283.9	-34.0	211	0.54 0.18	-131.4	-131.6
	-283.2	-34.2	215	0.53 0.18	-131.5	-131.7
	-281.3	-32.0	202	0.49 0.23	-130.8	-131.1
	-281.3	-32.0	199	0.52 0.23	-130.8	-130.0
	-280.6	-32.3	207	0.49 0.24	-130.9	-131.2
	-280.6	-32.3	200	0.53 0.22	-131.0	-131.2
	-280.4	-32.0	202	0.49 0.23	-130.8	-131.1
	-280.4	-31.9	199	0.52 0.23	-130.8	-131.0
	-279.7	-32.2	207	0.49 0.24	-130.9	-131.2
	-279.6	-32.2	200	0.53 0.22	-131.0	-131.2
2. Y(5OH-W) ⁸⁵ YA ⁸⁸	-256.6	-36.6	183	0.74 0.84	-133.6	-133.8

Table 3. (Continued)

	-254.7	-36.1	199	0.68	1.05	-129.7	-129.9
	-253.5	-30.5	183	0.74	1.04	-130.1	-130.3
	-253.5	-30.9	155	1.11	0.32	-142.8	-143.0
	-253.4	-30.4	182	0.74	1.02	-130.4	-130.6
	-253.0	-30.8	154	0.33	0.79	-143.0	-143.2
	-253.0	-33.5	172	0.78	0.29	-134.7	-134.9

Footnotes to Table 3

^a ETOT is the total energy of the system in kcal/mol, consisting of the total intramolecular energy of each helix and the intermolecular energy, EINTER. All conformations with $\Delta E_{tot} < 5.0$ kcal/mol are listed; $\Delta E_{TOT} = E_{TOT} - E_{TOT_0}$, where E_{TOT_0} is the energy of the lowest-energy minimum for each sequence, as given in the table.

^b See Chou et al. (1983,1984) and Kombo et al. (1995) for definition.

^c In angstrom units. The left column is the r.m.s. deviation with respect to the lowest-energy conformation of the wild-type protein, as in Kombo et al (1995). The right column is the r.m.s. deviation with respect to the lowest-energy conformation of the corresponding Trp-containing mutant.

^d In degrees [See Chou et al. (1983,1984) and Kombo et al. (1995) for definitions].

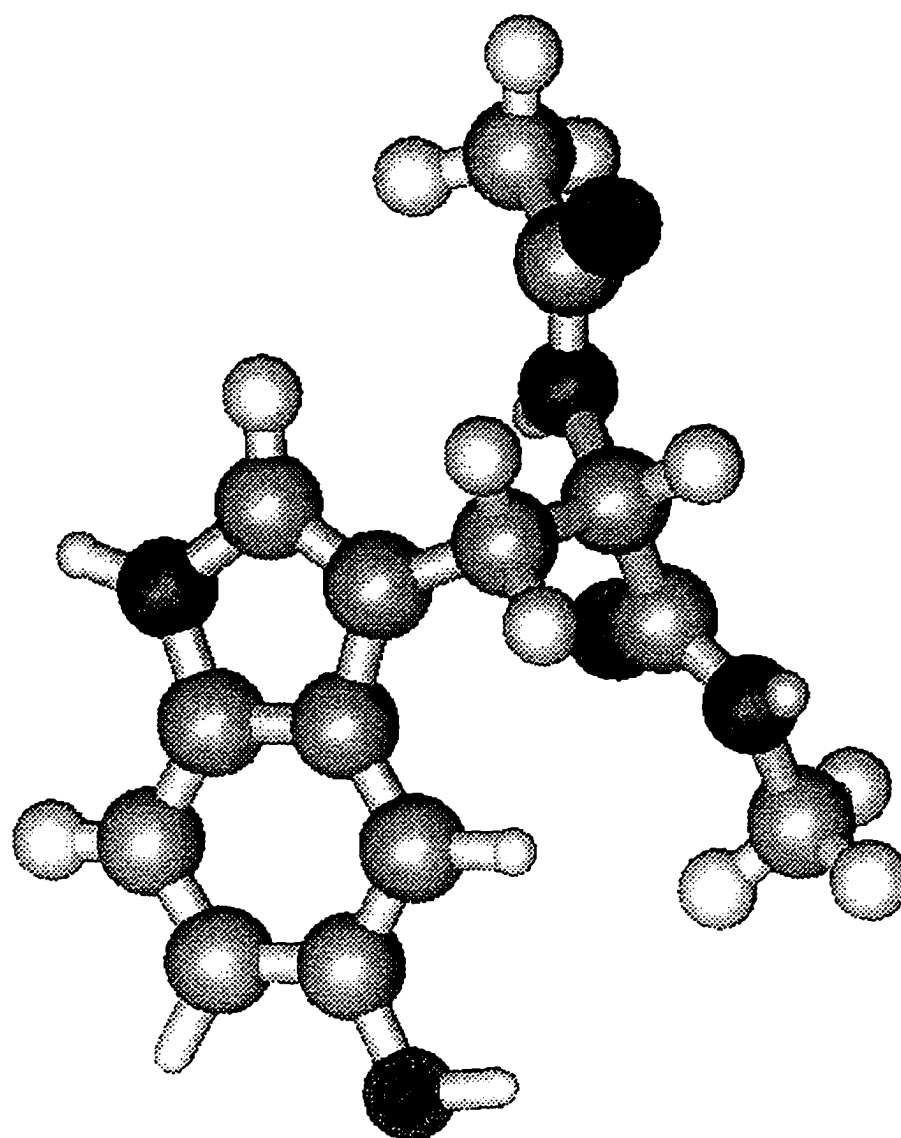


Fig.1.

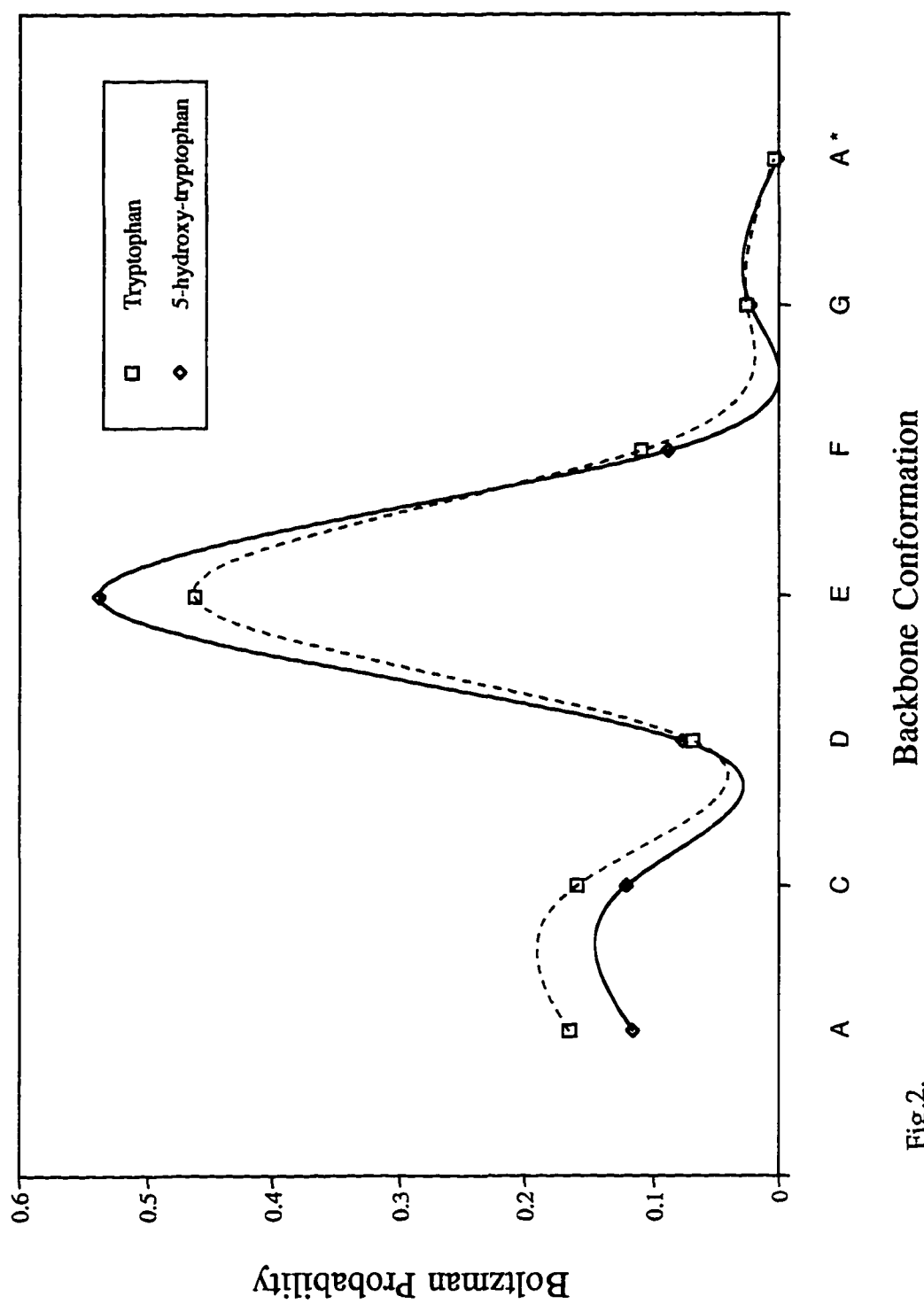


Fig.2.

Backbone Conformation

□ Tryptophan
◇ 5-hydroxy-tryptophan

Boltzmann Probability

C. Chapter 3

Table 1. Parameters of helices

Various Helices	ϕ^a	ψ^a	rise ^b	$O_i-N_{i+4}^b$	$C_iO_iN_{i+4}^a$	$O_i...H_{i+4}^b$	$N_{i+4}H_{i+4}O_i$
Active mutants	-69.1±1.0 (-70.0±2.0)	-36.2±2.0 (-36.0±3.0)	1.36±0.02 (1.34±0.04)	3.05±0.03 (3.27±0.21)	144.0±1.0 (139.0±4.0)	2.14±0.04 (2.38±0.24)	149.2±0.8 146.5±3.5
Inactive mutants	-66.0±3.0	-40.0±4.0	1.47±0.06	2.89±0.11	150.0±4.0	1.97±0.01	152.1±9.0
Hydrophilic environment ^c	-66.0	-41.0	1.47	2.88	147.4	ND	ND
Hydrophobic environment ^c	-59.0	-44.0	1.54	3.04	153.2	ND	ND
Hydrophilic C=O group ^d	-66.0±5.0	-41.0±6.0	ND	3.09±0.13	148.0±6.0	ND	ND
Hydrophobic C=O group ^d	-59.0±6.0	-44.0±6.0	ND	2.91±0.06	157.0±5.0	ND	ND
Hydrophilic C=O group ^e	ND ^f	ND	ND	3.05±0.13	ND	2.12±0.15	154.0±10.0

Table 1. (Continued)

Hydrophobic	ND	ND	ND	2.99±0.12	ND	2.05±0.15	157.0±9.0
C=O group ^e							

Footnotes to Table 1

^a Angle in degrees

^b Distance in Å

^c Computer-generated α -helix model (Blundell et al., 1983) obtained by assigning to all backbone dihedral angles ϕ/ψ the corresponding mean values given in (d) and defined below.

^d Parameters from Barlow and Thornton (1988) for residues whose C=O groups is hydrogen bonded to a solvent molecule or side chain (hydrophilic) or do not form such bonds (hydrophobic)

^e Parameters from Baker and Hubbard (1984), defined as in (d).

^f Not determined.

Table 3. Vdw contacts involving Tyr-85 and Tyr-85'

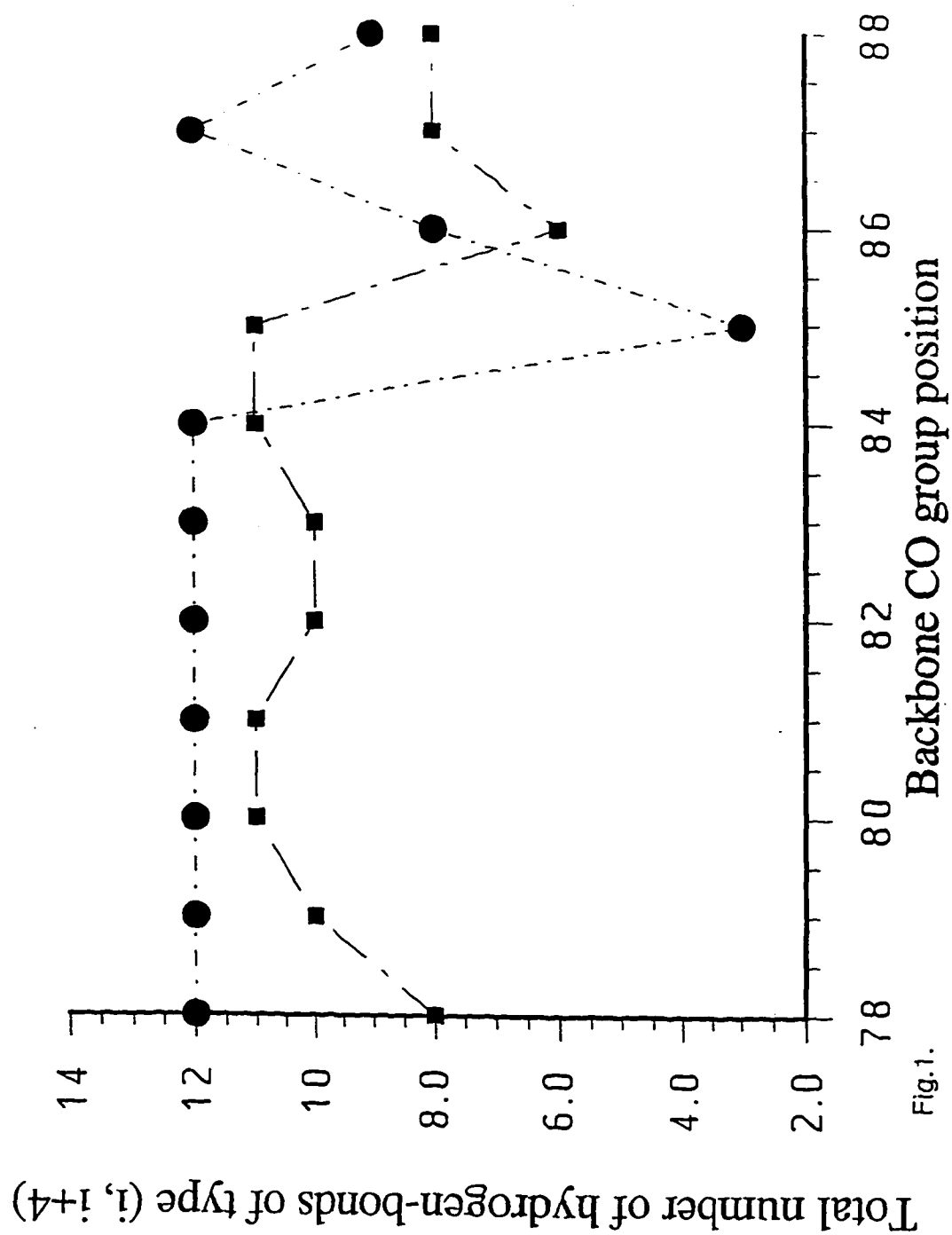
	Total	Y-85 -Y(X)-88 ^a	%	χ_2^b	χ_2^c
A. Active mutants					
1. WT	23	13	57	84	87
2. V91T	16	9	56	81	79
3. Y88F	17	10	59	83	83
4. E89L	28	16	57	89	88
5. M87L	34	16	52	90	87
6. E86A					
Y88A	42	10	24	75	86
B. Inactive mutants					
7. I84S	28	22	79	59	57
8. M87F	17	1	5	76	54
9. V91Y	48	9	19	62	54
10. I84A					
M87V					
Y88A	22	4	18	70	69
11. I84T					
M87T					
Y88V	0	0	-	62	54
12. I84F	46	18	39	53	54

Footnotes to Table 3

^a Number of pairs of atoms in close contact involving exclusively the aromatic ring atoms of Tyr-85 and Tyr-85' with any atom of the residue at position 88 and 88'.

^b Torsion angle (in degrees) of rotation about the C β -C γ bond of Tyr-85.

^c Torsion angle (in degrees) of rotation about the C β -C γ bond of Tyr-85'.



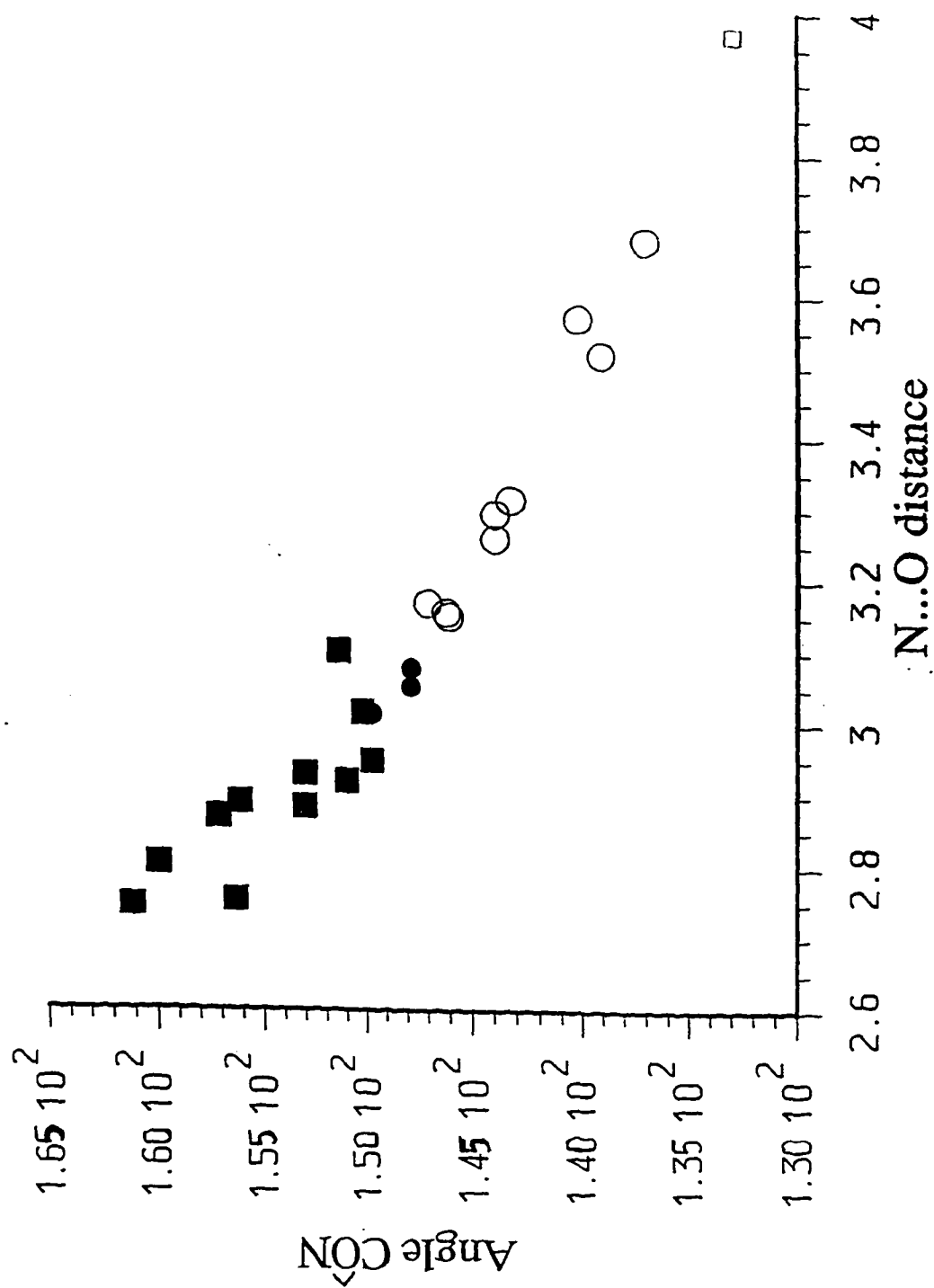


Fig.2.

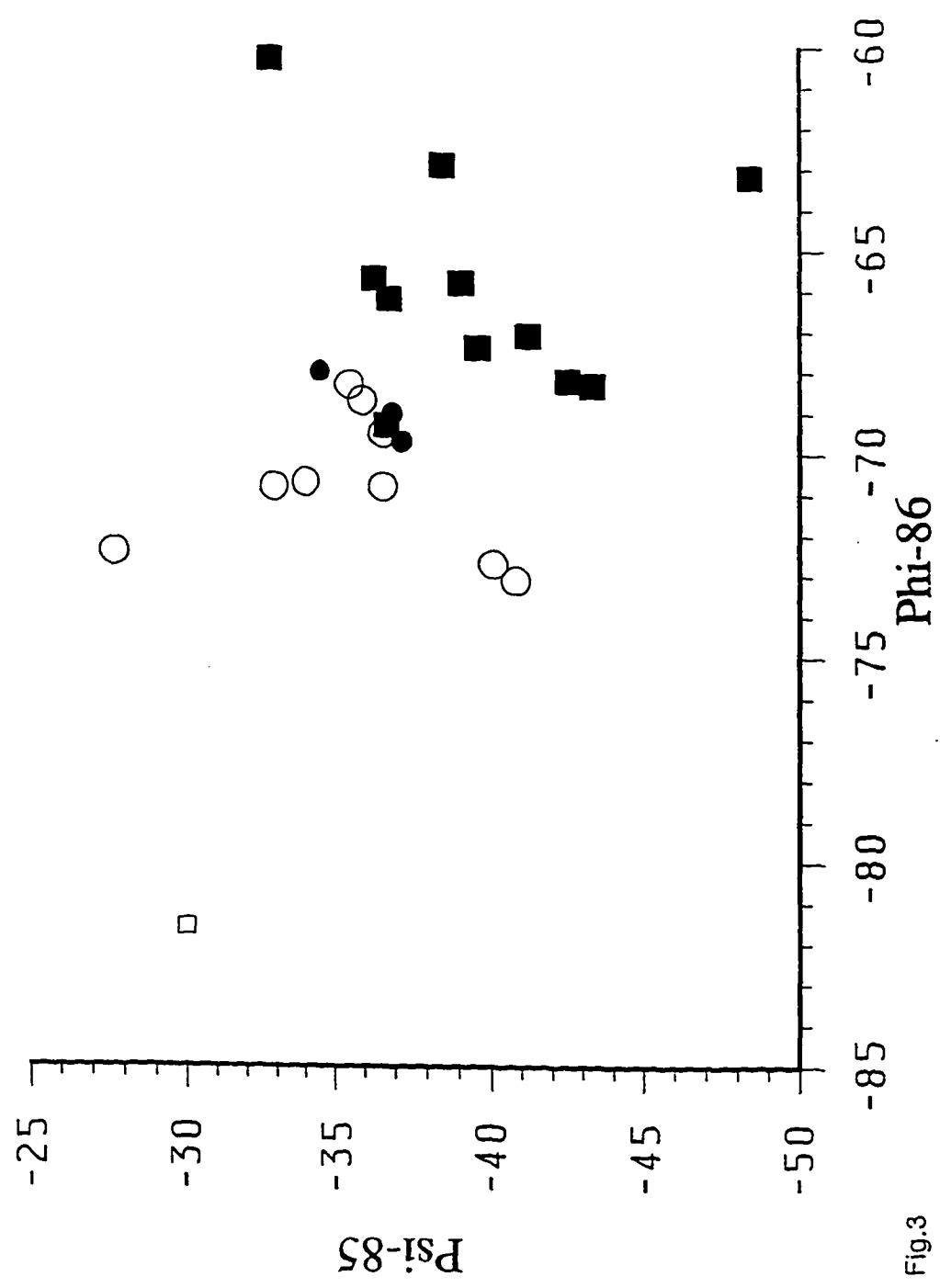


Fig.3

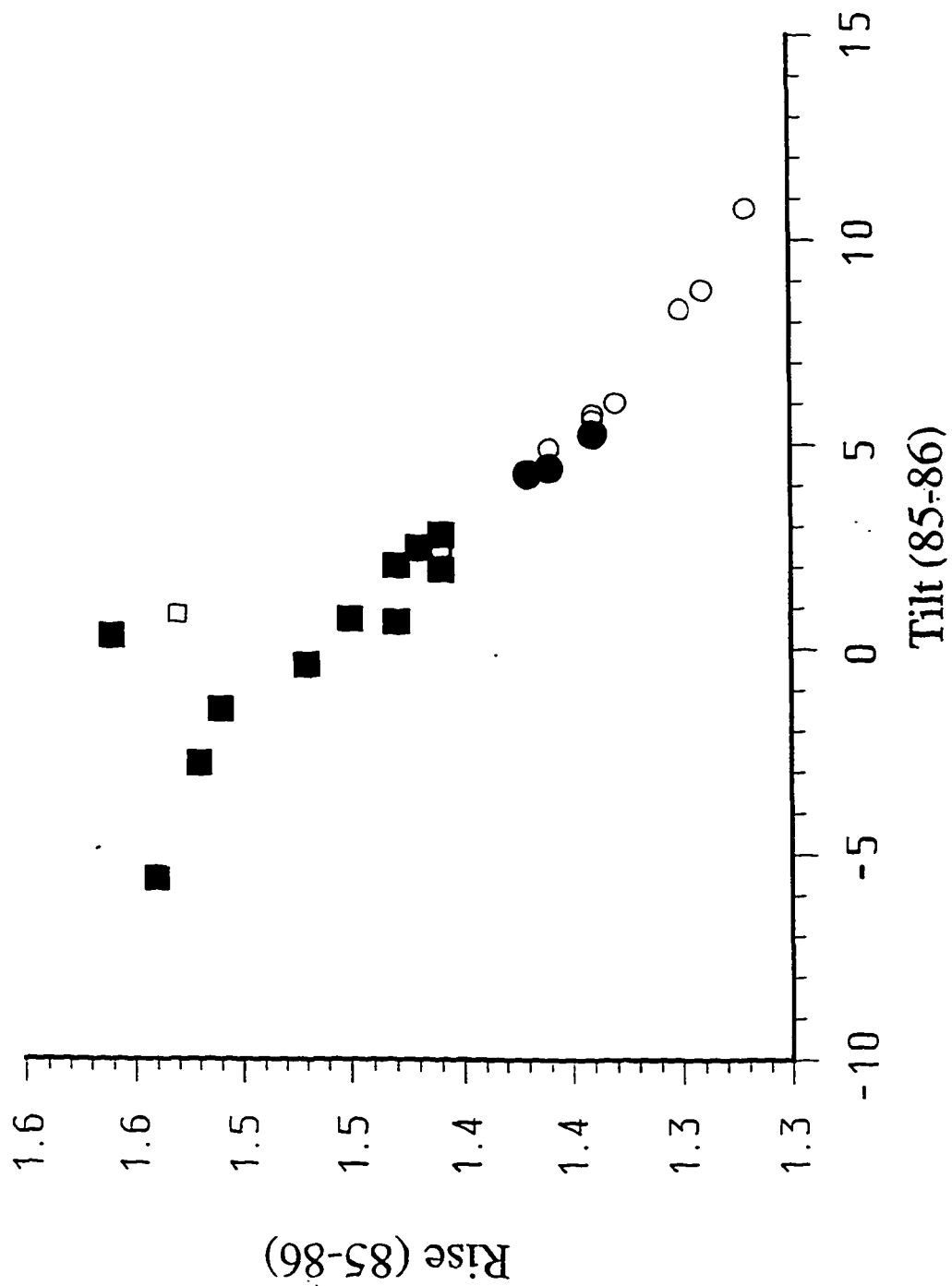


Fig.4.

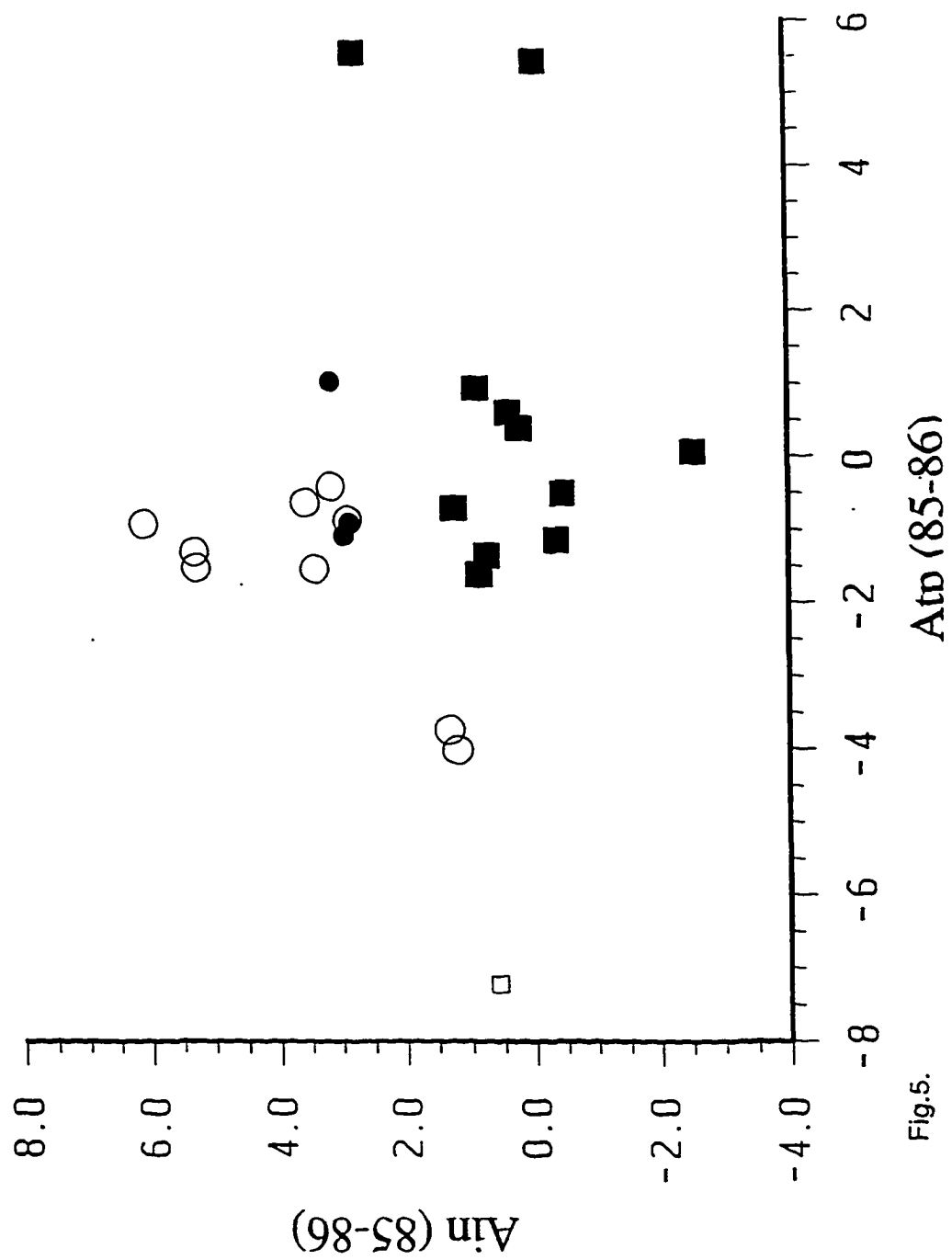


Fig.5.

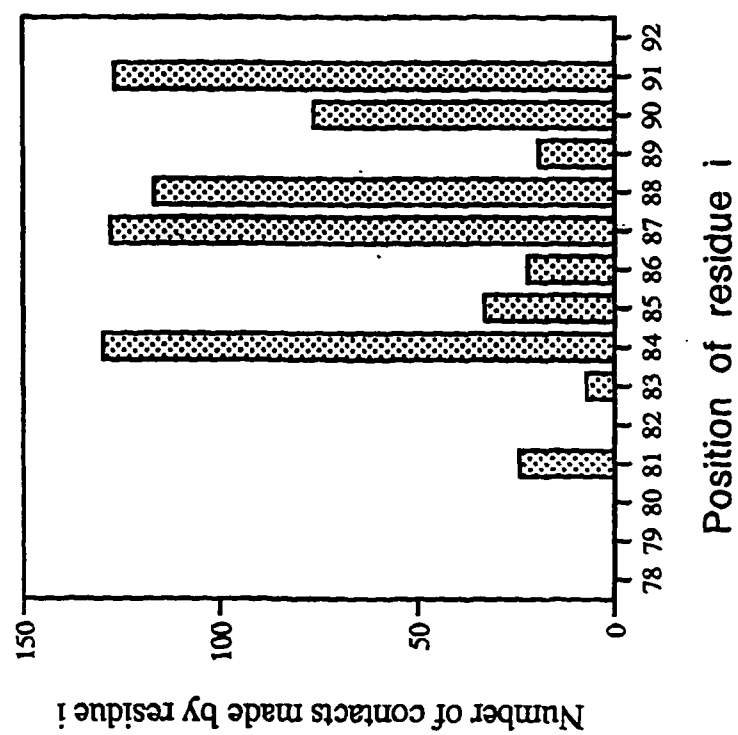


Fig.6

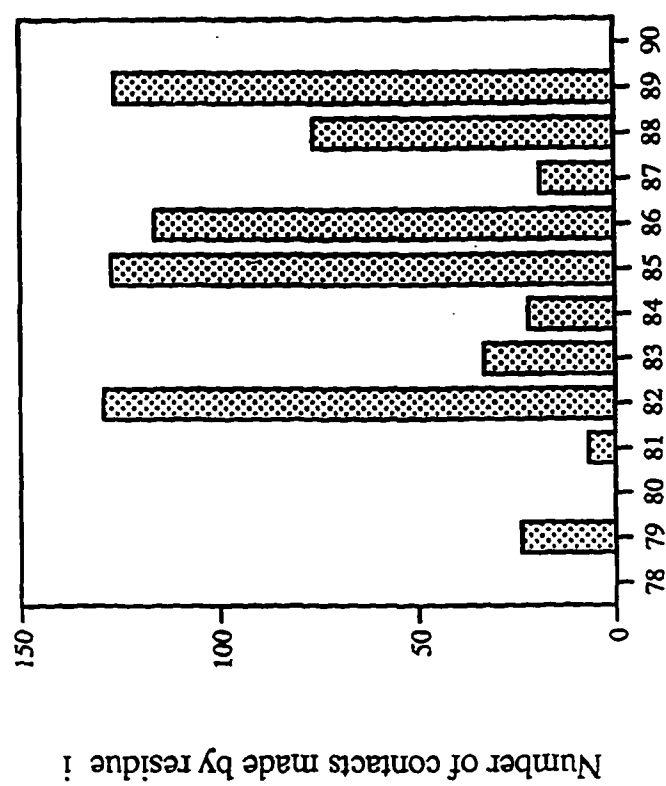


Fig.7

References

- Anfinsen, C.B. (1973). *Science* **181**, 223-230.
- Baker, E.N. & Hubbard, R.E. (1984). *Progress in Biophysics and Molecular Biology*, **44**, 97-179.
- Barlow, D.J & Thornton, J.M. (1988). *J.Mol.Biol.* **110**, 601-619.
- Bash, P.A., Singh, U.C., Langridge, R. & Kollman, P.A. (1987). *Science* **236**, 564-568.
- Beamer, L.J. & Pabo, C.O. (1992). *J.Mol.Biol.*, **227**, 177-196.
- Bernstein, F.C., Koetzle, T.F., Williams, G.J.B., Meyer, E.F., Brice, M.D., Rodgers, J.R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J.Mol.Biol.* **112**, 535-542.
- Blundell, T., Barlow, D., Borkakoti, N. & Thornton, J. (1983). *Nature* **306**, 281-283.
- Brooks, B., Bruccoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., & Karplus, M. (1983). *J.Comput.Chem.* **4**, 187-217.
- Chothia, C. (1984). *Ann. Rev. Biochem.*, **53**, 537-572.
- Chou, K.C., Némethy, G. & Scheraga, H.A. (1983). *J. Phys.Chem.*, **87**, 2869-2881.
- Chou, K.C., Némethy, G. & Scheraga, H.A. (1984) . *J.Am. Chem. Soc.*, **106**, 3161-3170.
- Chou, K.C., Némethy, G. & Scheraga, H.A. (1990) .*Acc. Chem. Res.*, **23**, 134-141.
- Clark, M., Cramer, R.D., III & Opdenbosch, N.V. (1989). *J.Comput.Chem.* **10**, 982-1012.

- Clarke, N.D., Beamer, L.J., Goldberg, H.R., Berkower, C. & Pabo, C.O. (1991). *Science* **254**, 267-270.
- Dang, L.X., Merz, K.M.Jr & Kollman, P.A. (1989). *J.Am. Chem. Soc.* **111**, 8505-8508.
- Dauber-Osguthorpe, P., Roberts, V.A., Osguthorpe, D.J., Wolff, J., Genest, M., & Hagler, A.T. (1988). *Proteins Struct. Funct. and Genet.* **4**, 31-47.
- DiCapua, F.M., S.waminathan & Beveridge, D.L. (1991) *J.Am.Chem.Soc.*, **113**, 6145-6155.
- Epstein, C.J., Goldberg, R.F., & Anfinsen, C.B. (1963). *Cold Spring Harbor Symp. Quant. Biol.* **28**, 439-449.
- Erenrich, E.H., Andreato, R.H. & Scheraga, H.A. (1970). *J.Am.Chem.Soc.*, **92**, 1116-1119
- Fossey, S.A., Némethy, G., Gibson, K.D. & Scheraga, H.A. (1991). *Biopolymers* **31**, 1529-1541.
- Gay, D.M. (1983). *ACM Trans. Math. Software* **9**, 503-524.
- Gerritsen, M., Chou, K.C., Némethy, G. & Scheraga, H.A. (1985). *Biopolymers* **24**, 1271-1291.
- Gibson, K.D. & Scheraga, H.A. (1994). *J.Comp.Chem.*, **15**, 1403-1413.
- Gibson, K.D. & Scheraga, H.A. (1988). *From Proteins to ribosomes*, **1**, 67-94, Adennine Press.
- Golub, G.H. & Van Loan, C.F. (1983). *Matrix Computations*, Johns Hopkins University Press, Baltimore, MD, pp. 425-426.
- Hehre, W.J., Radom, L., Schleyer, P.v.R. & Pople, J.A. (1986). *Ab Initio Molecular Orbital Theory*, John Wiley & Sons, Inc., New York.

- Hellingsa, H.W. & Richards, F.M. (1994). *Proc. Natl. Acad. Sci. USA* **91**, 5803-5807.
- Hogue, C.W.V., Rasquinha, I., Szabo, A.G. & MacManus, J.P. (1992). *Febs lett.* **310**, 3, 269-272.
- Hogue, C.W.V. & Szabo, A.G. (1993). *Biophys. Chem.* **48**, 159-169.
- Horwich, A. (1990). *Curr. Opin. Cell. Biol.* **2**, 625-633.
- Humayun, Z., Kleid, D. & Ptashne, M. (1977). *Nucleic Acids Res.* **4**, 1595-1607.
- Jayaram, B. (1996). *Ninth Conversation in the discipline of Biomolecular Structure and dynamics*, Albany, NY, 1995.
- Jordan, S.R., & Pabo, C.O. (1988). *Science* **242**, 893-899.
- Kirkpatrick, S., Gelatt, C.D. & Vecchi, M.P. (1983). *Science* **220**, 671-680.
- Kombo, D.C., Némethy, G., Gibson, K.D., Rackovsky, S. & Scheraga, H.A. (1996). *J.Mol.Biol.*, **256**, 517-532.
- Kuntz, I.D., Crippen, G.M., Kollman, P.A. & Kilmelman, D. (1976). *J.Mol.Biol.* **106**, 983-994.
- Laue, T.M., Senear, D.F., Eaton, S. & Ross, J.B.A. (1993). *Biochemistry*, **32**, 2469-2472.
- Laws, W.R., Schwartz, G.P., Rusinova, E., Burke, G.T., Chu, Y.C., Katsoyannis, P.G. & Ross, J.B.A. (1995). *J. Protein Chem.* **9**, 235-246.
- Lee, C. (1994). *J.Mol.Biol.* **236**, 918-939.
- Lee, C. & Subbiah, S. (1991). *J.Mol.Biol.* **217**, 373-388.
- Lee, C. & Levitt, M. (1991). *Nature* **352**, 448-451.

- Lee, B. & Richards, F.M. (1971). *J.Mol.Biol.* **55**, 379-400.
- Levitt, M. (1976) *J.Mol.Biol.*, **104**, 59-107.
- Lim, W.A., Hodel, A., Sauer, R.T. & Richards, F.M. (1994). *Proc. Natl. Acad. Sci. USA* **91**, 423-427.
- Lim, W.A. & Sauer, R.T. (1989). *Nature* **339**, 31-36.
- Lim, W.A. & Sauer, R.T. (1991). *J.Mol.Biol.* **219**, 359-376.
- Lodish, H.E. (1988). *J.Biol.Chem.*, **263**, 2107-2110.
- Lowe, J.P. (1968). *Prog. Phys. org. Chem.*, **6**, 1-80.
- MacGregor, M.J., Islam, S.A. & Sternberg, M.J.E. (1987). *J.Mol.Biol.* **198**, 295-310.
- Maniatis, T., Ptashne, M., Backman, K., Kleid, D., Flashman, S., Jeffrey, A. & Maurer, R. (1975). *Cell* **5**, 109-113.
- Maurer, R., Meyer, B.J., & Ptashne, M. (1980). *J.Mol.Biol.* **139**, 147-161.
- McCammon, J.A. (1991). *Curr. Opin. Struct. Biol.* **1**, 196-200.
- McCammon, J.A. & Harvey, S.C. (1987). *Dynamic of proteins and nucleic acids*, Cambridge University Press.
- Meyer, B.J., Kleid, D.G. & Ptashne, M. (1975). *Proc. Natl. Acad. Sci. U S A* **72**, 4785-4789.
- Meyer, B.J. & Ptashne, M. (1980). *J.Mol.Biol.* **139**, 195-206.
- Meyer, B.J., Maurer, R. & Ptashne, M. (1980). *J.Mol.Biol.* **139**, 163-194.
- Misra, V.K., Hecht, J.L., Sharp, K.A., Friedman, R.A., & Honig, B. (1994). *J.Mol.Biol.* **238**, 264-280.

- Momany, F.A., McGuire, R.F., Yan, J.F. & Scheraga, H.A. (1971). *J.Phys.Chem.*, **75**, 2286-2297.
- Momany, F.A., McGuire, R.F., Burgess, A.W. & Scheraga, H.A. (1975). *J.Phys.Chem.* **79**, 2361-2381.
- Némethy, G., Phillips, D.C., Leach, S.J., Scheraga, H.A. (1967).. *Nature*, **214**, 363-365.
- Némethy, G., Pottle, M.S., Scheraga, H.A. (1983). *J.Phys.Chem.*, **87**, 1883-1887.
- Némethy, G., Gibson, K.D., Palmer, K.A., Yoon, C.N., Paterlini, G., Zagari, A., Rumsey, S. & Scheraga, H.A. (1992). *J.Phys.Chem.* **96**, 6472-6483.
- Némethy, G. & Scheraga, H.A. (1989). *Bull. Inst. Chem. Res., Kyoto Univ.* **66**, 398-408.
- Némethy, G. & Scheraga, H.A. (1990). *Faseb J.* **4**, 3189-3197.
- Pabo, C.O., Sauer, R.T., Sturtevant, J., & Ptashne, M. (1979). *Proc. Natl. Acad. Sci. USA* **76**, 1608-1612.
- Pabo, C.O. & Lewis, M. (1982). *Nature* **298**, 443-447.
- Pabo, C.O. & Suchanek, E.G. (1986). *Biochemistry* **25**, 5987-5991.
- Piela, L., Némethy, G. & Scheraga, H.A. (1987). *Biopolymers* **26**, 1273-1286.
- Pople, J.A. & Segal, G.A. (1966). *J.Chem. Phys.* **44**, 3289-3296.
- Ptashne, M., Backman, K., Humayun, M.Z., Jeffrey, A., Maurer, R., Meyer, B. & Sauer, R.T. (1976). *Science* **194**, 156-161.
- Rackovsky, S. & Scheraga, H.A. (1978). *Macromolecules*, **11**, 1168-1174.
- Rackovsky, S. & Scheraga, H.A. (1980). *Macromolecules*, **13**, 1440-1453.

- Rackovsky, S. & Scheraga, H.A. (1978). *Macromolecules*, **14**, 1259-1269.
- Rackovsky, S. & Scheraga, H.A. (1978). *Acc. Chem. Res.*, **17**, 209-214.
- Ravishanker, G., Vijakumar, S. & Beveridge, D.L. (1994). *ACS Symposium Series* **569**, 210-219.
- Reidhaar-Olson, J.F. & Sauer, R.T. (1988). *Science* **241**, 53-57.
- Reidhaar-Olson, J.F. & Sauer, R.T. (1990). *Proteins struct. Funct. Genet.* **7**, 306-316.
- Reidhaar-Olson, J.F., Parsell, D.A. & Sauer, R.T. (1990). *Biochemistry* **29**, 7563-7571.
- Reeke, G. (1988). *Ann. Rev. Comput. Sci.* **3**, 59-84.
- Richardson, J. S. (1981). *Adv. Protein Chem.* **34**, 167-339.
- Richardson, J.S. & Richardson, D.C. (1988). *Science* **240**, 1648-1652.
- Ross, J.B.A., Senear, D.F., Waxman, E., Kombo, B.B., Rusinova, E., Huang, Y.T., Laws, W.R., Hasselbacher, C.A. (1992). *proc. Natl. Acad. Sc.*, **89**, 12023-12027.
- Rossmann, M.G. & Argos, P. (1976). *J.Mol.Biol.* **105**, 75-95.
- Roterman, I.K., Gibson, K.D., & Scheraga, H.A. (1989a). *Journ. Biomolec. Struct. Dyn.*, **7**, 391-419.
- Roterman, I.K., Gibson, K.D., & Scheraga, H.A. (1989b). *Journ. Biomolec. Struct. Dyn.*, **7**, 421-453.
- Sarai, A. & Takeda, Y. (1989). *Proc. Natl. Acad. Sci. USA* **86**, 6513-6517.
- Satyshur, K.A., Rao, S.T., Pyzalska, D., Drendel, W., Greaser, M. & Sundaralingam, M. (1988). *J.Biol.Chem.* **263**, 1628-1647.

- Sauer, R.T. & Anderegg, R. (1978). *Biochemistry* **17**, 1092-1100.
- Sauer, R.T., Hehir, K., Stearman, R.S., Weiss, M.A., Jeitler-Nilson, A., Suchanek, E.G. & Pabo, C.O. (1986). *Biochemistry* **25**, 5992-5998.
- Sauer, R.T., Jordan, S.R. & Pabo, C.O. (1990). *Advan. Protein Chem.* **40**, 1-61.
- Sklenar, H., Etchebest, C. & Lavery, R. (1989). *Proteins: Struct. Funct. Genet.* **6**: 46-60.
- Summers, N.L., Carlson, W.D., & Karplus, M. (1987). *J.Mol.Biol.* **196**, 175-198.
- Sundaralingam, M. & Sekharudu, Y.C. (1989). *Science* **244**, 1333-1337.
- Swaminathan, S., Ravishanker, G., Beveridge, D.L., Lavery, R., Etchebest, C. & Sklenar, H. (1990). *Proteins: Struct. Funct. Genet.* **8**: 179-193.
- Thornton, J.M. (1981). *J.Mol.Biol.* **151**, 261-287.
- Tidor, B. & Karplus, M. (1991). *Biochemistry* **30**, 3217-3228.
- Vásquez, M., Némethy, G. & Scheraga, H.A. (1983). *Macromolecules* **16**, 1043-1049.
- Vásquez, M., Némethy, G. & Scheraga, H.A (1994). *Chem. Rev.*, **94**, 2183-2239.
- Van Gunsteren, W.F. & Mark, A.E. (1992). *J.Mol.Biol.* **227**, 389-395.
- Wakahara, A., Kido, M., Fujiwara, T., Tomita, K.I. (1973). *Bull. Chem. Soc. Jpn*, **46**, 2475-2480.
- Wang, Y. & Sarai, A. (1994). *Prot. Eng.*, **7**, 1083-1087.

Weiner, S.J., Kollman, P.A., Case, D. A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S.Jr., & Weiner, P. (1984). *J.Amer.Chem.Soc.* **106**, 765-784.

Weiner, S.J., Kollman, P.A., Nguyen, D.T. & Case, D.A. (1986). *J.Comp.Chem.* **7**, 230-252.

Weiss, M.A., Pabo, C.O., Karplus, M. & Sauer, R.T. (1987a). *Biochemistry* **26**, 897-904.

Weiss, M.A., Pabo, C.O., Karplus, M. & Sauer, R.T. (1987b). *J.Biomolec.Str.Dyn.*, **5**, 539-556.

Yan, J.F., Momany, F.A. & Scheraga, H.A. (1970). *J. Am.Chem.Soc.*, **92**, 1109-1115.

Zacharias, M., Luty, B.A., Davis, M.E. & McCammon, J.A. (1990). *Biophys. J.*, **63**, 1280-1285.

Zimmerman, S.S., Pottle, M.S., Némethy, G. & Scheraga, H.A. (1977). *Macromolecules*, **10**, 1-9.