PROGRAM PROTEUS FOR ADDING HYDROGENS TO A PROTEIN STRUCTURE AND ELECTROSTATIC FIELD ACROSS CAROTENOIDS IN LIGHT HARVESTING COMPLEXES AND REACTION CENTERS FROM BACTERIAL SOURCES

by

SAMIR LIPOVACA

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

2001

Bell & Howell Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

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

9/25/00
Date

Chair of Examining Committee

10/2/00
Date

Executive Officer

R. Callender

J. Lombardi

T. Boyer

L. Themis

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

PROGRAM PROTEUS FOR ADDING HYDROGENS TO A PROTEIN STRUCTURE AND ELECTROSTATIC FIELD ACROSS CAROTENOIDS IN LIGHT HARVESTING COMPLEXES AND REACTION CENTERS FROM BACTERIAL SOURCES

by

SAMIR LIPOVACA

Advisor: Professor Marilyn R. Gunner

The hydrogen construction method presented in the program PROTEUS treats hydrogens depending on their torsional degrees of freedom. The positions of hydrogens with restricted torsional degrees of freedom are completely determined by the heavy atoms positions in the structure. The hydroxyl and water hydrogens are the only hydrogens that PROTEUS accepts as movable hydrogens (having rotational degrees of freedom). Their positions are determined by the interactions with neighboring atoms. PROTEUS interaction energy corresponds to a view that the hydrogen bond is affected, besides electrostatic effects and steric constraints of neighboring groups, by an inherent energy barrier that opposes free rotation of the hydroxyl hydrogen. For the water hydrogens that barrier is zero. The hydroxyl and water hydrogens are minimized within a short distance using the Threshold Accepting (TA) energy minimization method. PROTEUS can provide reasonable positions of movable hydrogens and a good initial protein structure for further investigations.

We applied the program PROTEUS to place hydrogens in several resolved three-dimensional crystal structures of light harvesting complexes (LHCs) and reaction centers (RCs) from bacterial sources. Using program DelPhi we calculated the local electrostatic field across carotenoid generated by the protein's charges. In each structure we identified amino acids responsible for the field. Much of the field is generated by the charged residues. There are different ways that a RC or LHC uses charged residues. A nearby dipole consisting of the charged residues which are ionized in the physiological pH range (like Arg-Asp), is often used. Clusters of charged residues or scattered isolated charged residues around the carotenoid molecule also contribute. The polarizable field is not necessarily along the carotenoid molecule principal axis. For soluble LHCs the contribution of polar residues to the field cannot be neglected. Our calculations indicate an important relation between the field and the protein's overall structure : the minimum value of the field arises from the backbone dipoles. In contrast to the charged residues, hydroxyl and water dipoles are responsible for a fine tuning of the field in a response to the protein's environment.

ACKNOWLEDGMENT

I am very grateful to my thesis advisor Professor Marylin R. Gunner for her guidance, encouragement and support; for giving me an opportunity to work in the field of biophysics which I like so much.

I am very grateful to James Bautista from Professor Harry Frank's Lab for providing me with valuable data.

I very much thank co-workers Jiali Li, Emil Alexov and Qing Xu for their help and support.

Finnaly, I would like to thank my wife, Besima Lipovaca, for her support and patience over the years.

TO MY WIFE AND DAUGHTER

TABLE OF CONTENTS

LIST OF TABLES

PART I

PART II

LIST OF FIGURES

PART I

PART II

PART I

I. INTRODUCTION

Hydrogen bonding is one of the key forces in protein folding and stability. The earliest work noting the importance of the hydrogen bonding was the prediction by Linus Pauling of the α and β secondary structure elements (helices and sheets) [1,2]. A significant survey of hydrogen bonding in proteins was a systematic analysis by Baker and Hubbard [3]. They showed that ninety percent of the hydrogen bonds were main - chain to main - chain N-H----O=C, eighty-five percent of these bonds were in α -helices or β -sheets, and a significant number (12% of N-H groups and 11% of C=O groups) were not involved in hydrogen bonding.

Protein folding is thought to be primarily driven by removal from solvent (water) of a hydrophobic surface of the unfolded chain. The stability of a folded protein over the denaturated state is around 5-15 kcal/mol, which is the total interaction energy of one to three hydrogen bonds [4]. So, if three independent hydrogen bonds cannot be formed in the folded protein, the folding stability might be lost (assuming that in the unfolded state all hydrogen bond donors and acceptors in the protein are hydrogen bonded to solvent molecules).

Location of the hydrogen atoms is essential to understanding the nature of the hydrogen bond. Protein data bank files (pdb files), deposited in Protein Data Bank of Brookhaven National Laboratory are three dimensional structures of proteins determined by X-ray diffraction. X-ray diffraction cannot resolve hydrogen positions in these structures. Only the heavy atoms (for example C, N, O, S) are located in a structure. A difference of one electron cannot be seen since the resolution distance for proteins has the order of magnitude of one angstrom. In contrast, neutron diffraction studies of crystallized proteins can locate the positions of hydrogen atoms. For neutron scattering, the 'visibility' of the atoms is not determined by the number of electrons they carry, but by the scattering length of the nucleus. The scattering length for hydrogen nuclei (proton) is of the same order of magnitude as for non hydrogen atoms, making hydrogens as

visible as other atoms. Because of this property the neutron scattering technique can in principle locate hydrogen atoms experimentally. A comparison of X-ray and Neutron Diffraction Single Crystal analysis is given in the Appendix A.

A correct treatment of hydrogen bonding requires determining the positions of donor hydrogen atoms. It has been shown that the fraction of unsatisfied hydrogen bond donors in protein structures is very low [5]. Hydrogens in protein structures can be classified as movable (hydrogens with degrees of freedom) or as static (hydrogens with no degrees of freedom). Hydrogens with degrees of freedom are important in stabilization of different ionization states of proteins. In particular, the polar hydrogens with dihedral angle rotational degrees of freedom (Ser, Tyr, Thr) play a critical role in stabilizing tertiary structure and protein packing [6].

Water molecules are tightly associated with proteins. There is a layer of ordered water molecules at the surface of the protein molecule seen in X-ray diffraction data. The high dielectric constant of solvent water, which arises from its mobility, serves to attenuate the attractive links between charged groups in proteins which form the strong hydrogen bonds known as salt bridges. Such strong intramolecular bonding could otherwise impose an unacceptable rigidity on the protein molecule. Water molecules are indicated in the bending and unfolding of helices and in the occurrence of a variety of turns [7]. So, locating the positions of polar and water hydrogens within the protein is necessary in order to determine their role in hydrogen bonding.

Since X-ray structural data for proteins at the present level of resolution do not generally locate hydrogen atoms it is necessary to have a computational procedure for constructing them from the heavy atom coordinates to obtain a starting structure for calculations. To validate the procedure and measure the possible benefits the predicted hydrogen positions may be compared with the hydrogen positions determined experimentally by neutron diffraction.

Currently, there are five distinct methods (as described bellow) to place hydrogens in their minimum energy positions in a protein structure. The first method is an arbitrary initial placement

with energy minimization that moves the hydrogens to their nearest local minima. This method does not adequately treat polar hydrogens since the energy minimization leads to the nearest, but not to the true minima. Brunger and Karplus method [8] finds the best torsion angle of each polar hydrogen independently, but it does not consider the simultaneous interactions of polar hydrogens in a case of two or more neighboring polar hydrogens. There is a limited planar search for positions of water hydrogens. The third method [6] (program NETWORK) maximizes the number of hydrogen bonds within a protein based on geometrical constraints (bond angles, bond lengths) of the hydrogen bonds. Since the method optimizes hydrogen positions for a single potential acceptor, it cannot find bifurcated hydrogen bonds (two or more acceptors). The fourth method [17] positions polar hydrogens in protein structures by optimizing the total hydrogen bond energy using an empirical hydrogen bond force field derived from small molecule crystal structures. In contrast, the neutron solvent difference map data have shown [21] that there is a strong influence of the local electrostatic environment on the positions of the polar hydrogens which, certainly, cannot be derived from small molecule crystal structures (without using a protein structure). The fifth method [22] uses a unique stochastic approach to overcome a combinatorial explosion of alternative combinations for all movable hydrogens based on all possible hydrogen bonds in which movable hydrogens could participate. It does not include in energy evaluations the torsion energy term for hydroxyl (Ser, Tyr, Thr) hydrogens that opposes free rotation of the hydroxyl hydrogen.

We developed a computational procedure (program PROTEUS) to place hydrogens in a protein structure that overcomes the weakness of the above described methods. PROTEUS interaction energy contains three terms: electrostatic interaction energy, torsion energy, and Van der Waals term. The hydroxyl and water protons are minimized within PROTEUS interaction energy from the atoms in the protein. For each hydroxyl or water oxygen only residues within a short distance of the hydroxyl or water oxygen are included in the analysis. If several hydroxyls or

water molecules are within the distance then all of their protons are minimized together in order to find their true minima. By increasing the distance long-range electrostatic effects on minimization can be probed. Positions of water hydrogens are evaluated without bias for a specific direction and bifurcated hydrogen bonds can be found.

We will compare the program PROTEUS with the above methods to validate the procedure. The comparison will be based on rms differences of the donor hydrogens between PROTEUS (or above methods) and four neutron diffraction structures (RNase - A (pdb file 5RSA), trypsin (file 1NTP), bovine pancreatic trypsin inhibitor (file 5PTI), and insulin (file 3INS)).

1.1 Classification of hydrogens in a protein structure

The protein or nucleic acid hydrogens can be classified according to their bonding environment. There are 5 possibilities that have to be considered in construction of hydrogen positions [8]:

1. Hydrogen bound to a donor with at least two heavy donor antecedents with known coordinates (e.g., peptide hydrogens, histidine imidiazole hydrogens, aliphatic CH_2 hydrogens).
2. Hydrogen bound to a donor with only one known donor antecedent and no other atoms bound to it (e.g., [η - H] group of a tyrosine residue).
3. Hydrogen bound to a donor with only one known donor antecedent and one additional hydrogen (e.g., [η - H_2] group of arginine).
4. Hydrogen bound to a donor with only one known donor antecedent and two additional hydrogens (e.g., [ε - H_3] group of lysine or a methyl group).
5. Water hydrogens can be placed anywhere over a sphere one angstrom from the center water oxygen (O).

The attribute "known" in the above usually refers to a heavy atom donor antecedent whose coordinates are determined by X-ray crystallography, but it may also refer to a particular

hydrogen whose position is determined by some other means, e.g., by neutron diffraction. In addition, the effect of hybridization has an important impact on hydrogen positions. For example, sp^2-hybridised oxygen (C=O group in proteins) is able to accept two hydrogen bonds by utilizing both lone pairs of electrons. Similarly for sp^3 hybridization (OH group of Ser and Thr) the two lone pairs on the sp^3-hybridized oxygen can accept two hydrogens. Hydrogens belonging to class 1 have a restricted dihedral (torsional) degree of freedom. Classes 2-4 have in common a torsional degree of freedom in placing the hydrogen. The "movable hydrogens" (hydrogens belonging to classes 2-5) influence functional properties of proteins (midpoint potentials, pK values, local electrostatic fields). The most important of the movable hydrogens are the hydrogens which can form hydrogen bonds (bonded to O or N), since these form specific interactions with other groups in a protein.

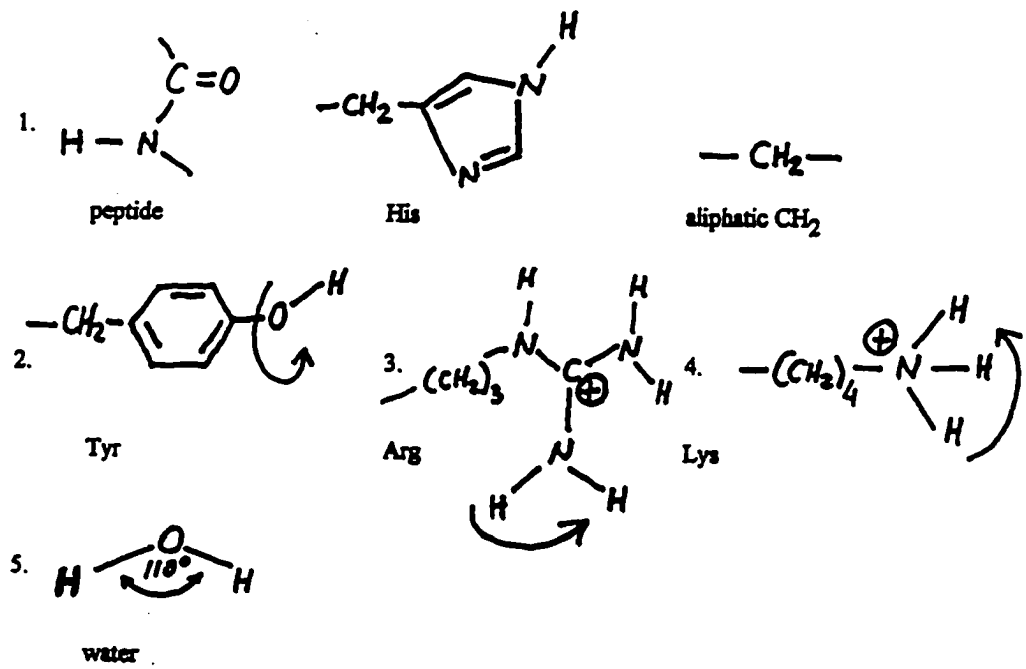


Fig. 1. Examples of possibilities that have to be considered in construction of hydrogen positions

1.2. Methods to place hydrogens in a protein structure

Computer-based methods for automatically, analytically adding hydrogen atoms to X-ray-determined structures exist for those hydrogens which have restricted torsional degrees of freedom. Their positions are determined by the heavy atoms positions in the structure using equilibrium bond lengths, bond angles, and dihedral angles. For movable hydrogens no unique method exists for finding the minimum energy location for these hydrogens.

The simplest method to place hydrogens in a protein structure is to use an extended atom model [9,10]. All hydrogen atoms were treated as part of the heavy atoms to which they are attached. The heavy atoms had their parameters adjusted to take account of the inclusion of the hydrogens (for example contributions of hydrogen bonds, bond angles, dihedral angles were incorporated into those parameters).Later, to better model hydrogen-bonding interactions in proteins, polar hydrogens can be specifically added (i.e., included in the calculations on the same level as the heavy atoms) while aliphatic hydrogens remained omitted [11,12]. Finally, methods were developed for placing hydrogens in a protein structure which include all hydrogen atoms explicitly [13,14]. Currently, there are five distinct methods to place hydrogens in a protein structure.

The first is to make an arbitrary initial placement and let energy minimization move the hydrogens to their nearest local minima. Due to the multiple-minima problem, [15,16] this method does not adequately treat polar hydrogens. Energy minimization leads to the nearest local minima: hydrogens will not cross any barriers to find its true minima.

Brunger and Karplus [8] used an iterative search-and-local-energy-minimization procedure to find the best torsion angle of each polar hydrogen independently. This method will find the true minimum for a particular hydrogen. They found that in most cases the dihedral angles of the hydrogen donor groups are placed near the torsional energy minima. But the method does not consider the simultaneous interactions of all rotatable polar hydrogens. The authors recognized

this problem and attempted to reduce the effect of this omission by performing the search twice: once forward through the list of polar hydrogens and once backward through the list.

The third method is to maximize the number of hydrogen bonds within a protein or model system (program NETWORK) based on conventional geometric rules [6]. The protein is divided into groups of interacting hydrogen bond donors and acceptors, called networks. Each network is independently evaluated to maximize the total number of hydrogen bonds. If two or more arrangement have the same number of hydrogen bonds, the arrangement with the shortest set of hydrogen bonds is selected. Since only a strictly limited set of hydrogen positions is considered for each of the rotational degrees of freedom of a donor, each one optimized for a single potential acceptor, bifurcated hydrogen bonds will not be found.

The fourth method positions polar hydrogens in protein structures by optimizing the total hydrogen bond energy [17]. For this goal, an empirical hydrogen bond force field was derived from small molecule crystal structures. Bifurcated hydrogen bonds are taken into account. Crystal structure symmetry is taken into account where appropriate. There is a high degree of similarity between that hydrogen bond force field and a molecular dynamics description of hydrogen bond [18-20]. But, the neutron solvent difference map data [21] have shown that orientation of hydroxyl (Tyr, Ser, Thr) hydrogens is not only determined by hydrogen bonding geometries. In fact there is a strong influence of the local electrostatic environment on the positions of the polar hydrogens. This reveals a strong inherent complementarity between the low-energy hydroxyl orientations and the local electrostatic environment.

The fifth method [22] adds initially, to a protein structure, non-movable hydrogens, such as peptide backbone hydrogens according to geometric considerations. The, water protons and polar side chain protons of Lys, Ser, Thr, Tyr, Asp, Glu and the C and N termini of a protein are added according to energy considerations. A unique stochastic approach overcomes a combinatorial explosion in the search for the lowest energy structure. First, the protein structure

is divided into ensembles. Each ensemble is treated separately. N conformations are sampled at random, and energies computed. Then common components of high - energy combinations are gathered on one hand, and low - energy combinations on the other. Components that yield only high - energy conformations and do not contribute to any low energies are excluded. This is reiterated while the total amount of combinations is decreased along the iterative process. When the total number of combinations is lower than a user defined threshold, all remaining combinations are evaluated by exhaustive search. Energy evaluations use only nonbonding energy terms.

1.2.1 The Multiple - Minima Problem and Its solution

Energy landscape of a protein structure consists of many local minima (multiple - minima) separated by energy barriers [23]. This feature of the protein emerges from the fact that there are many different arrangement of the atoms resulting in the same state of the protein. For example, there are alternative positions of flexible polar hydrogens, many of which could form hydrogen bonds. The problem is how to find the true minimum (the lowest one). Thus, the principal requirement for any optimization method is that it must be able to avoid entrapment in local minima and continue the search for a near - optimal solution regardless the initial conditions.

An arbitrary initial placement of hydrogens followed by energy minimization leads to their nearest local minima. Hydrogens can not cross any barrier to find their true minimum. The method is unable to continue the search after a local minimum is reached and therefore the solution is sensitive to the initial placement of the hydrogens.

There are several ways to overcome the multiple - minima problem. For example, Brunger and Karplus [8] overcomes the multiple - minima problem by an iterative search - and - local - energy - minimization procedure, which can be described as a sequence of locally confined energy minimizations by which any close contacts between hydrogens are avoided. The program

NETWORK overcomes the problem by maximizing the number of hydrogen bonds within a protein based on conventional geometric rules [6] and the method which optimizes the total hydrogen bond energy based on an empirical hydrogen bond force field that is derived from small molecule crystal structures [17] uses the Threshold Acceptance (TA) energy minimization method (Appendix B). The fifth method [22] overcomes the multiple - minima problem by dividing a protein into ensembles. An exhaustive search is done when the number of combinations is less then a user - defined threshold. If the number of combinations exceeds the user defined threshold a stochastic search is performed on an ensemble. It was shown that the stochastic search can find the global minimum out of a large number of possible combinations.

1.2.2. Brunger and Karplus method

Given the heavy atom coordinates, construction of all hydrogens of the protein is performed by a three - step iterative procedure [8]. In the first step, any crystallographically defined water molecules are represented by uncharged oxygen atoms. All hydrogens belonging to class 1 are placed by using equilibrium bond lengths, bond angles and dihedral angles.

In the next step the hydrogen positions characterized by classes 2 - 4 are constructed. All of these classes have in common that there is a dihedral degree of freedom in placing the hydrogen. To find energetically favorable conformation, the dihedral angle with respect to the antecedent - donor symmetry axis is modified in small steps. An angle grid of a 10^0 was found to be satisfactory. For each dihedral angle, the hydrogens of the donor are placed according to their equilibrium bonding geometry and the relative energy of the corresponding conformation is evaluated. The bonding geometry of the hydrogen donor for cases 2 and 3 is planar and for case 4 is tetrahedral. The energy is determined by using a hydrogen bond potential, a Van der Waals

term, an electrostatic term (Coulomb's law) and proper dihedral term as described in reference [24]. The conformation corresponding to the dihedral angle with the lowest total energy is taken and the hydrogens of the donor group are placed according to this conformation. This procedure is performed in the order given by the residue sequence of the protein. Hydrogens that have not yet been constructed are not included in energy evaluations, while those that have been constructed are included.

After construction of all hydrogens of the protein the water hydrogens are constructed. The water molecules are represented by the ST2 model [25], however the algorithm can be used with any explicit hydrogen water model. To avoid a dependence on the input sequence the water molecules are ordered with respect to the minimum distance of the water oxygen to any protein atom. Water molecules near the protein are treated first. The following possibilities are distinguished: a) water molecules able to form two hydrogen bonds to known acceptor atoms in the vicinity of the water molecule, b) water molecules able to form only one hydrogen bond to a known acceptor atom, c) water molecules not able to form any hydrogen bonds to known acceptor atoms. A known "acceptor atom" can be another water oxygen atom or any protein hydrogen bond acceptor atom, where the position is known from the X - ray structure or other considerations. Most water molecules correspond to class a, but classes b and c occur for situations where only very few water oxygen positions are known. In case a, a rotation of the water molecule in the plane defined by the two nearest acceptor atoms is performed, and the hydrogens of the water molecule are placed according to the minimum energy conformation. In case b one hydrogen is placed between the water oxygen and the acceptor atom. The water molecule is then rotated around this hydrogen bond axis, and the other hydrogen is placed in the minimum energy conformation. The energy expression used consists of the Van der Waals, electrostatic and water model - specific interactions [24,26]. Finally, the water molecules of case c are placed in a standard way (H_1 on x - axis, H_2 in the positive x, y quadrant).

After all hydrogen positions have been determined by the hydrogen construction, a second pass is performed in which the minimum energy conformation for each donor group and water molecule is reevaluated. This time all hydrogens are included in the interaction energy terms. No distinction is made between hydrogen and deuterium atoms.

The above sequential procedure has a built - in bias for the order in which hydrogens are placed and useful check is to start constructing the protein hydrogens at two different points, e. g., from the N - terminal and C - terminal ends.

1.2.3 The program NETWORK

In hydrogen placing method used by the program NETWORK a protein is divided into a series of interacting donors and acceptors called networks [6]. The word "donor" refers to a heavy atom which is covalently bound to a hydrogen which can participate in a hydrogen bond. The word "acceptor" refers to a noncovalently bound electronegative heavy atom which can interact with the electropositive hydrogen. A network is defined as a group of donors and acceptors which are linked together as potential partners. The term "DA" (donor/acceptor) refers to a heavy atom which can both donate and accept a hydrogen to form a hydrogen bond. In general, only donors which have hydrogens which are capable of significant rotational freedom are considered. This includes the hydroxyl hydrogens of Ser, Thr, and Tyr and the amino hydrogens of Lys and the amino - terminal residue of a protein (N - terminal). The polar hydrogens of the histidine side chain are treated as a special case since the hydrogen can be placed either on NE2 or ND1 nitrogens and the ionization state of the histidine can be adjusted to optimize the number of hydrogen bonds. Bifurcated hydrogen bonds are not considered by this program. At all times, the coordinates of the heavy atoms and excluded hydrogens (amide, guanidinyl, aliphatic, and aromatic hydrogens) remain fixed in space. Each network is rigorously searched for the optimum number of hydrogen bonds between the donors and acceptors. In the case where two or more

alternative arrangements with the same number of hydrogen bonds exists, the optimum network is selected by using the arrangement with the smallest total donor - acceptor distance. The structure of the algorithm is the following:

Read Donors and Acceptors

Find Partners

Find Networks

Optimize Networks

Assign Leftovers

Move Hydrogens

We will briefly describe each section of the algorithm.

Read Donors and Acceptors deals with a definition of donors and acceptors. Two auxiliary files contain the definition of all atoms to be read in from the data file (a PDB file) as hydrogen bond donors or acceptors. The first step in reading the donors and acceptors is to read the necessary information for each residue from the data file into an array. Each atom in the residue is compared to the definition files to identify it as a donor or acceptor and is stored in either the donor or acceptor list.

Find Partners is the next step in the algorithm and it defines all potential partners between donors and acceptors. This is done by searching the acceptor list once for each element in the donor list. If a donor and acceptor are less than 3.2 angstroms from each other, they are considered to be potential partners.

Find Networks defines networks. Because it is possible for an atom to be both a donor and an acceptor (for example, the oxygen in water) and because making one hydrogen bond in one place may preclude making a hydrogen bond elsewhere, it is necessary to define and optimize

such networks instead of arbitrarily making hydrogen bonds on a first come, first serve basis. There are three different levels for defining networks. In level 1, a network is defined as all nonwater donors and acceptors which are linked together as potential partners. Level 2 allows nonwater to nonwater as well as nonwater to water hydrogen bonds, but water to water hydrogen bonds are excluded from the network. Level 3 consider all possible bonds in determining a network. The potential hydrogen bonds within these networks will be optimized, while potential bonds not included in a network will be assigned arbitrarily in a step following optimization.

Optimize Networks step optimizes the number of hydrogen bonds within the network. This is done by checking every possible permutation of the hydrogen bonds within a network, and implementing the permutation that provides the most hydrogen bonds. If a different permutation has an equal number of hydrogen bonds but a smaller sum of all donor - acceptor distances, that permutation is regarded as optimal. Because each network is examined for the best possible set of hydrogen bonds, the number of comparisons scales with the factorial of the number of elements in the network.

The next step is **Assign Leftovers**. In some cases, due to the specified definition of a network, certain donor - acceptor pairs will not be included in the optimization of the network. For example, if level 2 is considered, only protein - protein or protein - water hydrogen bonds will be included in the networks. No water - water hydrogen bonds will be considered for optimization. These additional hydrogen bonds are assigned arbitrarily in this step.

The last step in the algorithm is **Move Hydrogens**. Once the bonding partners have been defined, the spatial locations of the donated hydrogens are changed. This is done by moving the hydrogen to the point closest to the acceptor that maintains the constraints of covalent bond length and valence angle for the hydrogen. If one hydrogen attached to a donor is moved, the rest of the hydrogens attached to the donor must be moved to maintain the approximate geometry of the donor hydrogens. In other words, NETWORK moves all hydrogens covalently bonded to the

donor by adjusting each of the hydrogens' dihedral angles. It does not check to see if an unmoved hydrogen overlaps a moved hydrogen. In the last step the authors used an energy minimization (an initial step of constrained energy minimization [27]) to resolve small distortions that may result from the torsion angles between the hydrogens and any bad nonbonded contacts.

1.2.4 The method that positions polar hydrogens by optimizing the total hydrogen bond energy

This method [19] positions polar hydrogen atoms in protein structures by optimizing the total hydrogen bond energy. To achieve this:

1. A hydrogen bond force field is derived specifically for this purpose from small molecule crystal structures. The energetics of hydrogen bonds in proteins are subject to the same physical rules as hydrogen bonds in small molecules. To derive a good hydrogen bond force field the Boltzmann's equation relating the ratio of the frequency of occurrence of two states to the energy difference between them is used. An analysis of hydrogen bonds of the Cambridge Structural Database [28] was used as a source of statistical data.

2. The orientation of His, Asn and Gln residues is taken into account, using an approach that considers the full hydrogen bond network at once. This would resolve ambiguities in a X - ray crystal structure of protein related to His, Asn and Gln side chains. Since a difference of one electron cannot be seen directly in a protein structure using X - ray diffraction data at less than atomic resolution, it is impossible to decide which of the two nitrogens in a His residue carries a hydrogen atom. For the same reason it is often not possible to see which of the two atoms is N or O in amide side chains. Similarly, the nitrogen and carbon atoms in the His side chain cannot be distinguished. From the direct environment of Asn, Gln, and His side chains, 75% [19] of the

orientations can be assigned unambiguously since the immediate surroundings leave only one possibility. In 25% of cases the approach that looks at each side chain one at a time in its surroundings will not work because it cannot take correlated configurations of two neighboring residues or water molecules into account. The approach that considers the full hydrogen bond network at once would resolve these cases.

3. Normal as well as bifurcated hydrogen bonds are scored, using an empirical hydrogen bond force field derived from small molecule crystal structures.

4. Crystal symmetry is used to locate hydrogen bonds between neighboring molecules in X - ray structures.

To calculate the total hydrogen bonding energy of any particular configuration, the contributions of the individual hydrogen bonds are added. All N and O atoms carrying H atoms are used as donors. S - H and C - H donors are not taken into account. All O - atoms are considered as acceptors, as are His side chain N atoms that do not carry a H atom, and Cys and Met S atoms. Aromatic π electron systems are not considered as acceptors. Energetically unfavorable contacts form exceptions to the rule:

a. Whenever a hydrogen atom comes within 2.0 angstroms of a hydrogen atom attached to its acceptor, the interaction is not taken into account. This is done to prevent situations where two O atoms donate their hydrogen atoms to each other.

b. Whenever a hydrogen atom comes within 2.5 angstroms of a positively charged atom, a penalty of 12 kcal/mol (just a large value) is subtracted from the total interaction energy. This makes it possible to describe the coordination of metal ions without explicit interaction terms.

The force field is meant for hydrogen bonds only. Due to the fact that it was derived from an analysis of observed hydrogen bonds, it contains all effects that can influence the preferred geometry of hydrogen bonds. Among these are the actual electrostatic interaction from the hydrogen bond, but also steric effects.

Three classes of degrees of freedom are considered: amino acid side chain flips, the presence or absence of hydrogen atoms in acidic and basic groups, and the rotational freedom in the placing of hydrogen atoms. All of these degrees of freedom are referred as ambiguities. Some restrictions of freedom are necessary to avoid physically unrealistic situations. The program uses the alternative of penalties: situations that are "undesirable" are given an energy penalty. An advantage of this approach is that the undesirability of the situation can be expressed by the size of the penalty.

For many of the hydrogen bond donors (backbone N, Arg side chain, etc.) accurate coordinates of the hydrogen atom can be calculated by using a simple geometric construction: hydrogen atoms are placed at ideal triangular or tetrahedral positions. Other hydrogen atoms (hydroxyl H, Lys NH_3, N-terminal(if not Pro) NH_3) are free to move on a circle. Hydrogen atoms connected to water molecules make a H - O - H angle of approximately 110^0, but the orientation of the molecule is free. In all cases hydrogen atoms are placed at a distance of 1.0 angstroms from the donor atom [3]. Both the amino acid side - chain flips and the presence or absence of hydrogen atoms have a discreet number of possibilities. To simplify the optimization procedure the rotations were also discretized:

1. NH_3 groups have 12 different angular orientations, each 10^0 apart.
2. OH groups have 36 different angular orientations, each 10^0 apart.
3. Water molecules have 366 different orientations, constructed using pairs of points from a 55 - point Fibonacci sphere [19] . This resulted in a homogeneous distribution of orientation without bias for a specific direction.

The optimization protocol starts with the construction of a list of all possible hydrogen bonds. All possible side - chain conformations of donor and acceptor and/or all possible positions of the hydrogen atom on the donor are taken into account. Acceptors in symmetry related molecules can be considered if desired; this will ensure that all intersymmetry hydrogen bonds will

be taken into account. From the resulting list all potential acceptors for each of the donors are extracted and stored.

After this, for each of the ambiguities a list is made of all affected donors. A donor is affected by an ambiguity if the value of the hydrogen bonds originating at that donor can be influenced by changing the status of the ambiguity. For an ambiguous donor the affected donors are the ones for which it is a potential acceptor. For an ambiguous residue the affected donors are all donors that have one or more of the side - chain atoms of the residue as potential acceptor. The cooperative effect (the effect that a hydrogen bond is stronger if the acceptor is donating another hydrogen itself) is not taken into account since this could slow down calculations.

Any ambiguity of Asn, His, or Gln side chains that cannot influence the total H - bond energy is fixed to the conformation of the side chain as given in the PDB file. The lists of affected donors and potential acceptors are searched to locate clusters of ambiguities in the hydrogen bond network that can form hydrogen bonds between them, but not to any other ambiguities. The hydrogen bond network is thus subdivided in subnetworks. Each of the subnetworks is considered independently.

In a subnetwork some ambiguities could be linked to only one other ambiguity. When this has been done, the ambiguity does not have to be considered any longer, and is removed from the subnetwork. This cutting procedure is repeated until no more ambiguities can be removed. If a water molecule is only linked to one other water molecule, the total number of evaluations needed for cutting is $366^2 = 133956$. This number of evaluations requires a considerable amount of time (10 seconds on a fast workstation) and cutting is not done. Cutting is only performed when it can be done using less than $10^{4.3} \sim 20000$ evaluations (a parameter of the program).

For the remaining part of the subnetwork an exhaustive evaluation is performed if the total number of evaluations needed is less than the exhaustive search limit. For subnetworks that are bigger than this, the Threshold Accepting method (Appendix B) is selected. The final energies of

the optimizations differ no more than a few kilocalories per mole, mostly due to different orientations of water molecules. The TA method was selected after genetic algorithms (GA) and the simulated annealing (SA) failed to give satisfying results.

1.2.5. Energy - Based Stochastic Method for positioning Polar Protons

The hydrogens and lone pairs, which are to be added are divided [22] into three categories:

1. Trivial hydrogens - those hydrogens that may be located by using coordinates and hybridization of heavy atoms such as aliphatic and aromatic hydrogens;
2. Nontrivial hydrogens - polar hydrogens, which have rotational degrees of freedom, such as Ser, Thr, and Tyr hydroxyls;
3. Nontrivial lone pairs - those with the same geometrical properties of nontrivial hydrogens such as the two lone pairs of the hydroxyl oxygen of Ser or Thr.

Trivial hydrogens are added first. Their coordinates are calculated by using the coordinates of the heavy atoms, the bond length and angles from the database as well as the standard dihedral angles.

Nontrivial hydrogens and lone pairs are divided into ensembles, and their coordinates are not yet calculated. An ensemble is defined as a group of nontrivial hydrogens or lone pairs that interact among themselves. The ensemble cutoff is user defined. It is measured from the coordinates of the heavy atom bonded to the nontrivial atom, because the nontrivial atom has not been located yet. Ensembles are composed of "segments". Each segment includes a rotation around a bond connecting two heavy atoms, one of which is bonded to a polar proton. Each segment may use various positions in space to fulfill H - bonding conditions. In addition to the ensemble cutoff, two other cutoff conditions are used: an energy cutoff and a cutoff that is used for locating hydrogen bonding partners around a rotatable segment. This is always more than or

equal 3 angstroms to allow inclusion of all close partners for H - bonding and to avoid the risk of missing solutions for a segment. However, increasing this cutoff over 4.5 angstroms creates many nonrealistic optional partners and extends the time for searching solutions.

Each ensemble is treated separately. A two - dimensional matrix is formed to calculate the coordinates of the nontrivial hydrogens and lone pairs. It is a list of all hydrogen bonds that may be formed between donors and acceptors. The larger the H - bonding cutoff, the more options for hydrogen bond connections will be formed, and the larger the two - dimensional matrix of alternative interactions will be. The matrix is refined, i.e., a location that yields a high - energy value is deleted. For this purpose, the energy threshold (a user defined cutoff) and nonbonding energy expressions are used.

By using the refined two - dimensional matrix, a three dimensional matrix is formed, where all combinations in an ensemble are uniquely defined, i.e., in any combination, there is only a single option for any nontrivial (rotatable) hydrogen and nontrivial lone pair. Each combination is evaluated, and the best combination is the result for the ensemble. In case of more than one ensemble, the process is repeated for each ensemble.

The energy criterion used to evaluate the quality of each combination is pairwise "nonbonding" energy function $E(r_{i,j})$:

$$E(r_{i,j}) = \Sigma_{i<j}(A_{i,j}/r_{i,j}^{12} - B_{i,j}/r_{i,j}^{6} + q_i q_j/\varepsilon r_{i,j})$$

Van der Waals (Lennard-Jones) interaction between the two atoms (i, j) with the corresponding repulsion parameter $A_{i,j}$ and attractive polarizability parameter $B_{i,j}$ and the electrostatic interaction (Coulomb's law). The atoms (i, j) have electric charges $q_{i,j}$, $r_{i,j}$ is the distance between them and ε is the dielectric constant chosen to be 4.

Energy calculations extend over the borders of each ensemble. The cutoff distance for calculating $E(r_{i,j})$ is user defined. It is recommended to avoid cutoffs so that long range electrostatic interactions may be accounted for. The main problem with this ensemble approach is to calculate interactions between nontrivial atoms in one ensemble and nontrivial atoms in another ensemble. In this case, the coordinates of the nontrivial hydrogens in a second ensemble, which have not been positioned yet, are assumed to coincide with the coordinates of the heavy atoms to which they are bonded. This is a "unified atom" approximation justified by the relatively long distance between the known position of one atom and the yet undetermined position of another. This approximation can be avoided if the system is handled as one huge ensemble, i.e., in this case, all nontrivial hydrogens and lone pairs are added simultaneously, with exact positions.

It is clear that in case of a large biological system constituting a single ensemble, a very large combinatorial problem results. To reduce the size of this problem, a stochastic approach is developed. The algorithm switches from exhaustive to stochastic calculation of an ensemble once the number of combinations exceeds a user - defined threshold. In the ensemble, the locations in some segments are unknown. For each nontrivial hydrogen or lone pair, there is usually more than one location, but only one would give the lowest energy. Nontrivial hydrogens and nontrivial lone pairs affect each other: if a lone pair is located, its location will dictate the location of a hydrogen bonded to the same heavy atom, and vice versa.

N conformations are sampled at random and their energies computed. Common components of high - energy combinations are gathered at the high - energy end, while low - energy combinations are gathered at the low - energy end. Components that yield only high - energy conformations and do not contribute to any low energies are excluded. This process is reiterated while the total number of combinations is decreased. When the total number of combinations is lower than the user defined threshold, all remaining combinations are evaluated by exhaustive search.

II. METHODS

We developed a program (PROTEUS) to place hydrogens in a protein structure. The program is written in C. It reads the protein data bank coordinate file (a PDB file). Two auxiliary ASCII files (fort.10 and param.lst) are used as parameter files. Fort.10 contains the connectivity of all atoms, their charges and what atoms have movable hydrogens. Param.lst contains A and B parameters for the Lennard-Jonnes function, the conversion factor for the electrostatic interaction that express energy in kT units (default is 4 for the value of the protein dielectric constant and this factor scales as (protein dielectric constant)$^{-1}$),various cutoffs and running options of the program. The running options determine what hydrogens to place in the protein structure. For example, place only hydroxyl hydrogens, or place only hydrogens on waters. The user may easily edit these files to add, delete, modify residue types, or change the running options.

The program has three modules. The first module performs lexical changes in the PDB file according to the fort.10. The second module deletes all water molecules (or specified residue type(s) in fort.10) that are not buried inside of the protein if the user wants to consider only water molecules that are inside of the protein (usually functionally important water molecules). For that purpose it calls executable routine "surfv" [29]. The third module is a proton placing module described bellow.

The hydrogens are divided into three categories:

(1) hydrogens that may be located by using coordinates and hybridization of heavy atoms such as aliphatic and aromatic hydrogens (hydrogens with restricted torsional degrees of freedom (non-movable hydrogens));

(2) polar hydrogens which have rotational degrees of freedom such as Ser, Thr, and Tyr hydroxyls;

(3) water hydrogens.

Lone pairs are not taken into account in the proton placing module as potential hydrogen acceptors.

Each movable hydrogen (hydroxyl or water hydrogen) has a discrete set of coordinates. A hydroxyl hydrogen has 36 different angular orientations, each 10^0 degrees apart and at the distance of 1 angstrom from the hydroxyl oxygen. There are two options for placing water hydrogens. The first option generates 20 positions for first proton homogeneously distributed over a sphere one angstrom from the center water oxygen . These positions form a dodecahedron. The task is to start with a maximum number of positions for first proton homogeneously distributed over a sphere. A dodecahedron is a Platonic solid with the maximum number of corners (20) with respect to other 4 Platonic solid. The second proton is placed on the sphere around the oxygen atom to get as compact a packing as possible. In total 30 water orientations are generated. In the second option the positions for first protons are generated as in the first option. Then for each first proton 36 different angular orientations, each 10^0 degrees apart, of the second water proton are created resulting in 720 water orientations. The discrimination between these two options is completely determined by the program user. Once chosen, the option will be applied on all water oxygens in the structure. The second option extends the time for searching hydrogen positions.

A flow chart for the hydrogen positioning algorithm is given in the Appendix C. Hydrogens of the first category are added first. Their coordinates are completely determined by the coordinates of the heavy atoms, the bond length, bond angles and dihedral angles. Then, for each hydroxyl and water oxygen a list of close neighboring atoms is formed. In other words, for each hydroxyl or water oxygen only atoms within a short cutoff (a distance ~3 angstroms) are included into the list to allow the inclusion of all close partners for hydrogen bonding. The distance is measured from the coordinates of the heavy atom bonded to movable hydrogen, because the movable hydrogen has not been located yet. Various cutoffs are allowed with respect

to inclusion of hydroxyls, waters or other atoms into a list. Hydroxyl - hydroxyl or water hydroxyl inclusion is treated by common cutoff. Water - water inclusion has its own cutoff as does the inclusion of acceptors for hydroxyl or water hydrogen bonds. Inclusion of all other atoms into a list is evaluated by the appropriate cutoff. Increasing any of these cutoffs over 4 angstroms creates many unrealistic proton positions due to the long-range electrostatic interactions.

Lists are combined into webs. We define a web as a group of hydroxyls and waters that interact among themselves. Each web is treated separately. The task is to find those positions of movable hydrogens that minimize the energy of a web. The energy of i-th web (E_{wcb}^{i}) contains three terms: electrostatic interaction energy (E_{el}^{i}), torsion energy (E_{T}^{i}), and Van der Waals (Lennard - Jones) term (E_{LJ}^{i}).

$$E_{web}^{i} = E_{el}^{i} + E_{T}^{i} + E_{LJ}^{i}$$

$$E_{el}^{i} = \Sigma_{k<l}\, C * q_k q_l / r_{k,l}$$

$$E_{T}^{i} = \Sigma\{ \text{ for TYR: } -2.97\cos(2*\text{torsion angle})$$
$$\text{for SER,THR: } -1.1\cos(3*(\text{torsion angle} - 60))\}$$
$$\text{(sum is over all Tyr,Ser, Thr in the i-th web)}$$

$$E_{LJ}^{i} = \Sigma_{k<l}\, (A/r_{k,l}^{12} - B/r_{k,l}^{6})$$

$q_{k,l}$ are electric charges of atoms k and l in the web, $r_{k,l}$ their distance, and A, B, and C are parameters that yield the energy in kT units (1kT = 0.59 kcal/mol; A=679.11, B=56.27, and C=140.0). The values for A and B corresponds to the minimum of Van der Waals (Lennard - Jones) energy for a distance of 1.7 angstroms between atoms. The value for C corresponds to the

dielectric constant of 4 inside of the protein. The expression for the torsion energy is taken from the program CHARMM [8]. Steric repulsion values [42] range from 1.3 kcal/mol for Ser and Thr to 3.5 kcal/mol for Tyr and this barrier opposes free rotation of the hydroxyl hydrogen.

This corresponds to a view that the hydrogen bond is affected, besides electrostatic effects and steric constraints of neighboring groups, by an inherent energy barrier reflecting the steric repulsion of adjacent bonded atoms (torsion energy) [21]. This barrier opposes free rotation of the hydroxyl hydrogen. E_{web}^{i} is minimized with respect to possible positions of hydroxyls and waters using the Threshold Acceptance Method (Appendix B). The energy of hydroxyls and waters that are not part of any web (isolated hydroxyls/waters) is minimized by an exhaustive search.

III. RESULTS

3.1. PDB files used to test the program PROTEUS

Four protein structures with hydrogen positions determined experimentally by neutron diffraction were used to validate and measure the possible benefit of the program PROTEUS. These are RNase - A (Brookhaven Protein Data Bank file 5RSA), trypsin (file 1NTP), bovine pancreatic trypsin inhibitor (file 5PTI), and insulin (file 3INS). In the case of trypsin, the monoisopropylphosphoryl group was removed for this validation. This structure did not contain any waters of crystallization. In the case of insulin, both dimers in the asymmetric unit were included. The insuline structure did not contain the positions of the water hydrogens.

3.2. Rms differences of the donor hydrogens between PROTEUS and neutron diffraction structures

Table 1 summarizes the rms differences of the donor hydrogens in the PROTEUS - optimized structure as compared with the positions determined by neutron diffraction. Angular differences between PROTEUS and the neutron diffraction hydroxyl positions for all four protein structures are in the range from 7^0 to 40^0. For comparison the results of Brunger and Karplus, the program NETWORK, and the energy - based stochastic method for positioning polar protons are included. Results for the procedure that locates polar hydrogen atoms in protein structures by optimizing the total hydrogen bond energy were not available. The view of the authors was that many differences were found due to problems in the experimentally determined structures. Therefore, they accepted that the comparisons with the positions determined by neutron diffraction are not useful as a way to access the accuracy of their method. Instead, they successfully performed molecular dynamics simulations to test the method.

Table 1. rms Differences of the Donor Hydrogens Between PROTEUS and Neutron Diffraction Structures

R.M.S. differences (A)

protein	method	SER	TYR	THR	HOH
trypsin(1NTP)	number of H	33	10	10	0*
	PROTEUS	0.61	0.83	0.6	-
	NETWORK	0.34	0.44	0.17	-
	B & K	0.89	0.64	0.34	-
	Stochastic	0.53	0.45	0.27	-
RNase(5RSA)	number of H	15	6	10	2*
	PROTEUS	0.68	0.58	0.52	0.41
	NETWORK	0.61	0.6	0.3	-
	B & K	0.98	0.6	1.12	1.2
	Stochastic	0.56	0.48	0.56	0.65
BPTI(5PTI)	number of H	1	4	3	4*
	PROTEUS	0.43	0.21	0.58	0.28
	NETWORK	0.96	0	0.07	-
	B & K	0.71	0.81	0.19	0.35
	Stochastic	0.30	0.58	0.41	0.67
Insulin(3INS)	number of H	6	8	4	1*
	PROTEUS	0.13	0.53	0.29	-
	NETWORK	0.63	0.35	0.81	-

B & K (Brunger and Karplus method of placing polar hydrogens in protein structures)

*number of internal water molecules

PROTEUS used the second option for placing of water hydrogens (720 positions)

3.3. NETWORK and PROTEUS

As we can see from the Table 1 for BPTI the program NETWORK did not move the tyrosine hydroxyl hydrogens from the neutron diffraction structure. The reasons for this were missing acceptors within the NETWORK cutoff distance (hydrogen-to-acceptor distance less than 2.5 angstroms) for two Tyr residues and other two Tyr residues (10 and 23) were optimized as acceptors to a water molecule. This finding is supported by the program PROTEUS. NETWORK found all the hydrogen bonds which were present in the neutron diffraction structure and an additional hydrogen bond between side chain of Ser 47 and its carbonyl oxygen [6]. The neutron diffraction structure does not indicate such hydrogen bond since the distance between hydroxyl hydrogen of Ser 47 and its carbonyl oxygen is 3.82 angstroms. PROTEUS placement of that hydrogen also does not indicate the hydrogen bond (the distance between hydrogen and oxygen is 3.91 angstroms).

For the insulin NETWORK found three additional hydrogen bonds: Thr A8 side chain to Glu A4 carbonyl oxygen and Ser 12 side chain to its carbonyl oxygen in both the A and C subunits. The neutron diffraction structure and PROTEUS do not indicate hydrogen bonds for Ser 12 in both the A and C subunits. The hydrogen-to-acceptor distance is around 4 angstroms. The hydrogen bond between Thr A8 hydroxyl hydrogen and Glu A4 carbonyl oxygen is questionable. For the neutron diffraction hydroxyl hydrogen of Thr A8 the hydrogen-to-acceptor distance is 3.61 Angstroms and for PROTEUS hydrogen the distance is 3.79 angstroms.

For RNase - A the PROTEUS optimized structure appears to be closest to the structure optimized by the program NETWORK. In the case of trypsin, PROTEUS rms differences are closer to those of Brunger and Karplus than to the NETWORK rms differences.

3.4. B & K method and PROTEUS

Proteus water positions are close to the neutron diffraction positions than the B & K water positions. This is due to the fact that the B & K hydrogen construction method performs only a limited planar search for a water molecule and does not include out of plane conformations. PROTEUS does search over a full conformational water space. Although the conformational search for PROTEUS first water algorithm (30 water orientations) is not limited in a plane as in the case of Brunger and Karplus water orientations, results are similar (not shown). It appears that increasing the number of degrees of freedom for water orientations leads to the reduction in the rms difference between constructed and observed positions of water hydrogens (Table 1).

Combining possible positions of hydroxyl hydrogens and water hydrogens results in many hydrogen conformations of the protein - water system. Brunger and Karplus found that the energy differences between those conformations appear to be small (under 0.5 kcal/mol) [8]. PROTEUS energy differences compare well with this (e.g., in the case of BPTI we found 0.06 kcal/mol for the average difference). In general, the PROTEUS estimate for the order of the magnitude of the average energy difference between two succesive conformations is 10^{-2} kcal/mol when the number of steps for the energy minimization is large (around 1000000) and the second water algorithm for water hydrogens is used.

For RNase PROTEUS' rms differences for threonines are half of the B & K rms differences. This illustrates PROTEUS' ability to search over all possible threonines conformations while the B & K algorithm cuts the search after two iterations. The authors noted that for most practical purposes, two iterations of the hydrogen construction will be sufficient. However, that is not the case for situations such as Thr residues in RNase, where several alternative conformations are possible and the algorithm needs more than two iterations in order to reach self - consistency.

3.5. The energy - based stochastic method and PROTEUS

It appears that PROTEUS' water positions are closer to the neutron diffraction positions than the water positions determined by the energy - based stochastic method. This is related to different approaches between these two methods for placing water hydrogens. PROTEUS uniformly places hydrogens over a sphere one angstrom from the water oxygen and then selects the minimum energy conformation. The stochastic method, like the program NETWORK, only considers possible hydrogen bonds that the water molecule can make with potential acceptors and between them selects the minimum energy conformation. This limited conformational search could miss some intermediate water conformations (between two hydrogen bonds).

There is a good agreement for hydroxyl hydrogen positions between these two methods, although they are based on different energy concepts. PROTEUS considers the torsion energy term for hydroxyls, in addition to the electrostatic and Van der Waals terms. The stochastic method does not consider the torsion energy. On the other hand, the stochastic method adds the lone pairs on the hydrogen bond acceptor to the structure and the location of the lone pair will dictate the location of a hydrogen bonded to the same heavy atom. PROTEUS does not consider the influence of the lone pairs.

Table 2. Average R.M.S. Differences of the Donor Hydroxyl Hydrogens Between PROTEUS and NETWORK, Brunger and Karplus, and Stochastic Structures

Average R.M.S. differences (A) for TYR, SER, THR

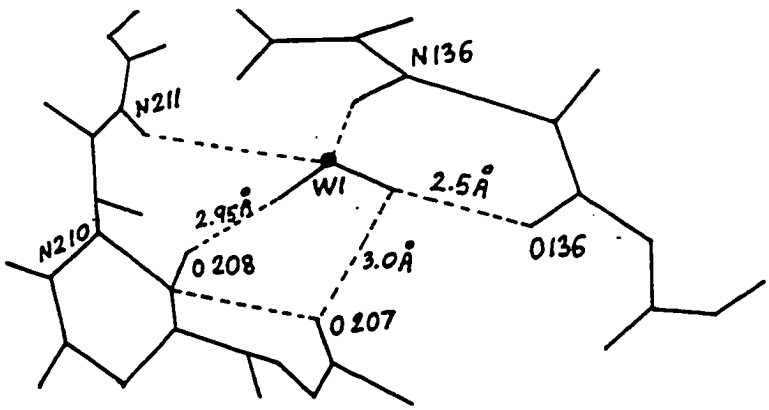
STOCHASTIC	NETWORK	B & K	PROTEUS	pdb file
0.42	0.32	0.62	0.68	1NTP
0.53	0.5	0.9	0.59	5RSA
0.43	0.34	0.57	0.41	5PTI
-	0.6	-	0.32	3INS

A predictive strength of the program PROTEUS is shown in the Table 2, where average rms differences for hydroxyl hydrogens are compared. The PROTEUS optimized protein structure and the same structure optimized by B & K method, the program NETWORK, or the stochastic method provide similar but not identical results.

3.6. The method that optimizes the total hydrogen bond energy and PROTEUS

Baker and Hubbard [3] gave a few interesting examples of clusters of buried water molecules in the actinidin structure (PDB file is 2ACT). For water molecule 1, surrounded by two donors and three acceptors (Fig. 2), the PROTEUS optimized hydrogen bond network has one bifurcated hydrogen bond in a complete agreement with the hydrogen bond network optimization based on optimization of the total hydrogen bond energy .

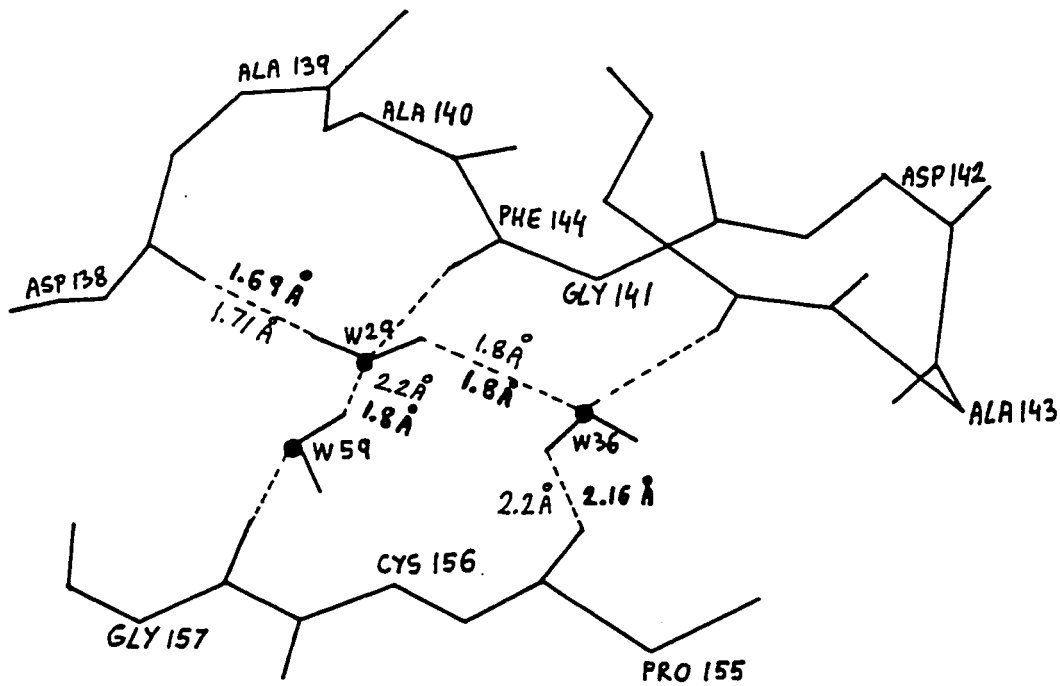
Fig 2. Water molecule 1 is within hydrogen bond distance of the three acceptors (all C=O groups) and two donors (NH groups)



Both options for treatment of water hydrogens in the program PROTEUS gave the same hydrogen bond lenghts output for the hydrogen bond network of water 1.

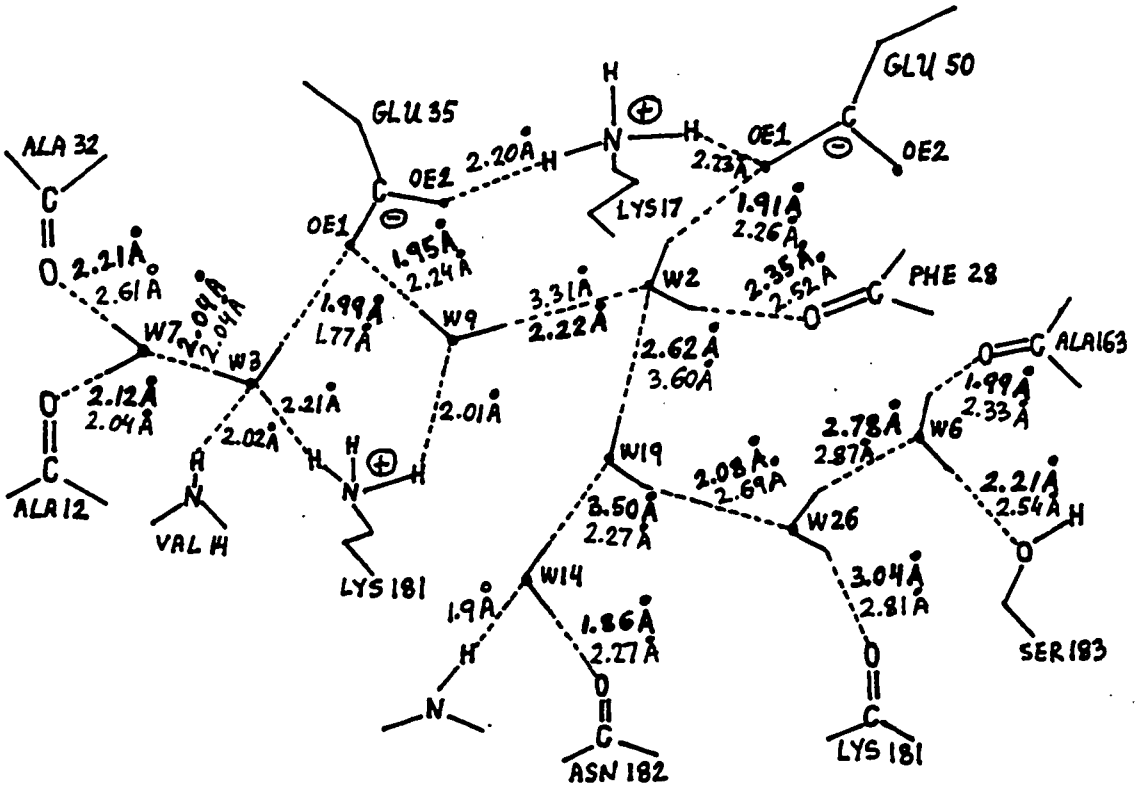
The pair of water molecules 29 and 36 fill a surface crevice in a complete agreement with the procedure that optimizes the total hydrogen bond energy (Fig 3.)

Fig 3. Pair of buried water molecules between two sections of polypeptide chain



The buried cluster of eight water molecules for which the hydrogen bonding scheme was completely hand-assigned in Baker and Hubbard is assigned in exactly the same way by the program PROTEUS and the procedure that optimizes the total hydrogen bond energy (Fig 4.).

Fig 4. A cluster of eight water molecules is shown, forming a network of hydrogen bonds in the interior of the actinidin structure



For Fig 3, 4. bold notation of hydrogen bond lengths represents using the second PROTEUS option for water hydrogens (each water molecule has 720 possible hydrogen orientations).

IV. DISCUSSION

PROTEUS is a program which automatically finds the optimal locations of hydrogens in a protein structure. The positions of hydrogen atoms which have restricted torsional degrees of freedom are completely determined by the heavy atoms positions in the structure using equilibrium bond lengths, bond angles, and dihedral angles. The hydroxyl and water hydrogens (movable protons) are the only atoms that do not have their positions completely determined by the heavy atoms in the structure. Their positions are influenced by the interactions with neighboring atoms. PROTEUS interaction energy contains three terms: electrostatic interaction energy, torsion energy, and a Van der Waals term. This corresponds to a view that the hydrogen bond is affected, by electrostatic effects and steric constraints of neighboring groups, as well as by an inherent energy barrier reflecting the steric repulsion of adjacent bonded atoms (torsion energy) [21]. This barrier opposes free rotation of the hydroxyl rotor.

4.1. PROTEUS and previous methods for placing hydrogens in a protein structure

It has been shown by the neutron solvent difference map data [21] that there is a strong inherent complementarity between the low-energy hydroxyl orientations and the local electrostatic environment. Our probe calculations for hydroxyl hydrogens (not shown here) indicated that PROTEUS interaction energy is mostly dominated by electrostatic interaction energy, with Van der Waals and the torsion energy terms in their minima.

4.1.1 Brunger and Karplus method

Brunger and Karplus have calculated the best torsion angle of each polar hydrogen based on steric and electrostatic factors. The energy is determined by using a hydrogen bond potential, a

Van der Waals term, an electrostatic term, and a proper dihedral term (torsion energy) as described in the program CHARMM [8]. They found that in the most cases the conformations of lowest intrinsic torsion energies are statistically favored for hydroxyl groups. But they did not consider the simultaneous interactions of all rotatable polar hydrogens and a limited planar search for hydrogen positions of water molecules was performed. The energy minimization was limited to the environment of the hydrogen atoms bound to a common heavy atom or to a single water molecule. PROTEUS does not discriminate between movable hydrogens. All movable hydrogens (hydroxyl and water hydrogens) within the distance are minimized together and each water molecule can have 720 different orientations. These orientations form a homogeneous distribution.

Brunger and Karplus discussed a specific example of the difference between the neutron diffraction and their hydrogen-constructed structure (Tyr 94 of trypsin). They found that the η hydrogen of Tyr 94 does not donate hydrogen bonds. The hydroxyl oxygen is accepting a hydrogen bond from water 56. In contrast, the neutron diffraction structure shows an approximately 0.4 kcal/mol more favorable [10] conformation where the hydroxyl hydrogen of Tyr 94 points toward a lone pair of water 56. This difference is a consequence of the B & K hydrogen construction method. During placing of the Tyr 94 hydroxyl hydrogen the water is only represented by an uncharged oxygen atom. This forces one of the hydrogens of water 56 to a position where it points toward the tyrosin hydroxyl oxygen. This conformation is stable enough not to be changed during further iterations or minimizations. In contrast, PROTEUS responds to the charge of water 56 and it places hydroxyl hydrogen of Tyr 94 almost at the neutron diffraction position (the distance between two positions is 0.07 angstroms).

4.1.2 NETWORK

The second method (program NETWORK) considers only a strictly limited set of hydrogen positions for each of the rotational degrees of freedom of a donor, each one optimized for a single potential acceptor. As a consequence of this, bifurcated hydrogen bonds will not be found. The possibility of bifurcated hydrogen bonds is built-in in PROTEUS, as a consequence of the rotational freedom for hydroxyls and waters and inclusion of all discrete hydrogen positions in the energy minimization search.

For BPTI NETWORK found an additional hydrogen bond between side chain of Ser 47 and its carbonyl oxygen [6] that is not indicated by the neutron diffraction structure and PROTEUS. It seems that this hydrogen bond is a consequence of maximizing the number of hydrogen bonds that can be formed for each group of interacting donors and acceptors. In other words, NETWORK always places a hydrogen to a position which allows the closest approach of the hydrogen to the acceptor respecting the hydrogen's bond length and angle.

4.1.3. The method that optimizes the total hydrogen bond energy

The fourth method places polar hydrogens in protein structures by optimizing the total hydrogen bond energy. An empirical hydrogen bond force field was derived from small molecule crystal structures . Electrostatic interactions are not explicitly included into the force field and bifurcated hydrogen bonds as well as crystal structure symmetry are taken into account. The authors compared predicted hydrogen positions with the experimentally determined positions for four neutron diffraction structures from the PDB. Many differences were found and the authors assigned these differences to the problems in the experimentally determined structures. For example, the temperature factors of all Ser residues of the trypsin neutron diffraction structure are quite large [8].For the ribonuclease neutron diffraction structure, hydrogen positions are less refined than in the other neutron diffraction structures. Therefore, they concluded that the

comparisons are not useful as a way to assess the accuracy of this method. In most cases molecular dynamics runs starting from their optimized hydrogen positions equilibrated quicker to a lower root-mean-square (RMS) value than molecular dynamics runs starting from the uncorrected coordinates with protons placed in standard positions by the molecular dynamics program [17]. The improvement is mostly the result of selecting the correct local minimum for a larger number of hydrogens. Molecular dynamics energy minimizations change these hydrogen positions only by a very small amount, indicating a high degree of similarity between their hydrogen bond force field and a molecular mechanics description of a hydrogen bond. But it appears that the orientation of hydroxyl hydrogens is not only determined by hydrogen bonding geometries, electrostatic and steric factors also contribute forming a strong relation between the low-energy hydroxyl orientations and the local electrostatic environment [21]. This influence of the local electrostatic environment on the positions of the polar hydrogens cannot be derived from small molecule crystal structures (without using a protein structure) and is not contained in the hydrogen bond force field of the above method.

4.1.4 The stochastic method

The stochastic method adds hydrogens as well as lone pairs to a protein structure. This is the first method that treats lone pairs at the same level as hydrogens. Despite the fact that the torsion energy is not included in the energy function of the stochastic method, there is a good agreement between hydroxyl hydrogen positions for PROTEUS and the stochastic method. Therefore, it appears that the torsion energy and lone pairs are both small effects in a sense that the hydrogen positions determined by the electrostatic and Van der Waals interactions only, are not significantly perturbed. This may be an expression of the protein's density in the vicinity of hydroxyl protons, i.e. a "dense" protein should increase the barriers for rotation of hydroxyl

hydrogen [22]. It appears that hydroxyl hydrogens are "embedded" in a less dense surrounding and that implies a relative "free" rotation of these hydrogens.

The discrepancy between water positions for both methods are related to PROTEUS' inclusion of intermediate water hydrogen conformations (conformations between all possible hydrogen bonds in which the water molecule can participate) into the energy minimization procedure. This reduces rms difference between the water hydrogen positions determined by PROTEUS and the neutron diffraction results.

V. CONCLUSIONS

The hydrogen construction method presented in the program PROTEUS treats hydrogens depending on their torsional degrees of freedom. The positions of hydrogens with restricted torsional degrees of freedom are completely determined by the heavy atoms positions in the structure. The hydroxyl and water hydrogens are the only hydrogens that PROTEUS accepts as movable hydrogens (having rotational degrees of freedom). Their positions are determined by the interactions with neighboring atoms. PROTEUS interaction energy corresponds to a view that the hydrogen bond is affected by, besides electrostatic effects and steric constraints of neighboring groups, an inherent energy barrier that opposes free rotation of the hydroxyl hydrogen. For the water hydrogens that barrier is zero. The hydroxyl and water hydrogens are minimized within a short distance using the Threshold Accepting (TA) energy minimization method. PROTEUS can provide reasonable positions of movable hydrogens and a good initial protein structure for further investigations. Especially, PROTEUS algorithm for placing water hydrogens gives smaller rms differences from experimental observations than Brunger & Karplus and the Stochastic method (Table 1).

Additional improvements could be made to PROTEUS. The intention is to treat the polar side chain hydrogens of Lys, the amino terminus of a protein and the methyl group hydrogens like

hydroxyl hydrogens. In addition, to treat the case whether the hydrogen is placed on the NE2 or ND1 nitrogens of a His residue. Since, the pKa of the His residue is near neutrality a number of charged residues, containing hydrogens on both nitrogens might be expected in a large protein even at neutral pH.

Additional improvement could be made to a treatment of biological water. The water molecules enclosed within the solvation shell present in the immediate vicinity of the protein (biomolecule) are termed "biological water". It is known that the hydration shell surrounding a protein molecule comprises different types of water [30-32]. Few water molecules remain rigidly bound to the protein for a very long time. In the immediate vicinity of the surface of the protein, there are water molecules that experience much faster rotational and translational diffusion rate than the water molecules directly bound to the protein. Thus, it is believed that biological water consists of two kinds of water molecules, referred to as "bound" and "free", depending on their momentary states of existence. There is dynamic exchange between these two species. There exists a quantitative theory of dielectric relaxation of biological water [33] based on a dynamic equilibrium between the free and bound water molecules. The dielectric relaxation depends on several parameters such as the rotational time constant of the protein molecule, the dimension of the hydration shell, the strength of the hydrogen bond, and the static dielectric constant of the water bound to the protein (biomolecule). The theory includes all these aspects in a consistent way. The current version of PROTEUS treats all crystallographic waters explicitly, assigning dielectric constant of 4 to the protein interior and 80 to the protein exterior.

APPENDIX A

Comparison of X-ray and Neutron Diffraction Single Crystal Analysis [34]

X - ray diffraction	Neutron diffraction
X-rays available on demand from laboratory instruments	Neutrons available only from national or international centers
Data collection time a day or less (for routine work)	Data collection time a few weeks
Temperatures below 120 K not easily available	Temperatures down to 10 K conveniently available
Hydrogen atoms poorly located, especially O - H. Accuracy ~0.1 angstroms	Hydrogen positional parameters comparable in accuracy to C, N, and O, ~0.001 angstroms
Cannot analyze anisotropic thermal motion or disorder of H atoms	Analysis comparable to that of C, N, O
Small crystals can be used (~0.01 mm^3, ~0.01 mg)	Large crystals required (~1 mm^3, 1-2 mg)

Can be difficult to distinguish between thermal motion and disorder even for nonhydrogen atoms because of fall-off in intensity with scattering angle	Fall-off in intensity with scattering angle only due to its thermal motion. Therefore easier distinguish from disorder
Number of variable parameters 9A+4B+1, where A is the number of nonhydrogen atoms, B the number of hydrogen atoms	For comparable observation-to-parameter ratio, needs more observation. Number of variables 9N+7, where N is the total number of atoms and six anisotropic extinction parameters are used. If the structure contains many hydrogens, deuterium substitution may be essential to reduce incoherent background.
Careful absorption corrections necessary for other than first-row atom molecules	Absorption negligible, except for crystals containing B, Cd, Sm, Li (corrections for H advisable for molecules with large H content)
Extinction generally not serious for organic compounds	Extinction serious and pervasive

Radiation damage can occur and must be monitored by repeating selected measurements	No radiation damage

Recently, it was shown [35] that some technical difficulties relating to the neutron diffraction (for example, the large incoherent scattering of hydrogens results in background scattering that greatly reduces the signal to noise of the experiment) could be overcomed by the use of deuterated protein (all hydrogens in the protein replaced by deuterium atoms). Such structure reveals much chemical information about the molecule, including the geometry of hydrogen bonding, states of protonation of histidines or the location and geometry of water molecules at the surface of the protein.

APPENDIX B

Proteins provide a natural frame for a large combinatorial problems. For example, in RNase A (5RSA) there are 1.76×10^{59} alternative combinations for all rotatable hydrogens based on all possible hydrogen bonds in which rotatable hydrogens could participate [22]. One wants to find a combination that minimizes or maximizes a certain cost function of the system (in case of a protein the cost function is the energy of the protein). The cost function depends on the detailed configuration of the many parts of the system. The number of variables involved may range up into the tens of thousands. A search trough all combinations is time consuming and it exceeds the computer capabilities. To reduce the size of the combinatorial problem various combinatorial optimization procedures are developed. Simulated annealing (SA), the Threshold acceptance method (TA), the genetic algorithm (GA), Taboo search (TS) and TRUST are examples of such optimization procedures.

SA [36] is inspired by statistical physics in an analogy with annealing in solids where, by slowly decreasing the temperature from the molten state, the system solidifies in a state of minimum energy. TA is very similar to SA and it is simpler structured than SA. It has been shown [37] that TA method is not so dependent on a cooling strategy and it yields better results than SA (at least for the traveling salesmen problem). GA [38] relies on the Darwinian principle of evolution: the algorithm cross-breeds trial solutions and allows only the "fittest" to survive after several iterations. TS [39] is based on concepts from artificial intelligence. It is a flexible framework of a variety of strategies of intelligent problem solving. TRUST [40] is founded on two innovative concepts, subenergy tunneling and non-Lipschitzian terminal repellers, to ensure escape from a local minima in a fast and computationally efficient manner.

This appendix we want to devote briefly to the Simulated annealing and Threshold acceptance methods. SA is a stochastic optimization algorithm, which borrows deep ideas from statistical physics. There is a deep and useful connection between statistical mechanics (the

behavior of systems with many degrees of freedom in thermal equilibrium at a finite temperature) and combinatorial optimization (finding the minimum of a given function depending on many parameters (cost function)). A fundamental question in statistical mechanics concerns what happens to the system in a limit of low temperature - for example, whether the atoms remain fluid or solidify, and if they solidify, whether they form a crystalline solid or a glass. Ground states and configurations close to them in energy are extremely rare among all the configurations of a macroscopic body, yet they dominate its properties at low temperatures because as temperature (T) is lowered the Boltzmann distribution collapses into the lowest energy state or states. But low temperature is not a sufficient condition for finding ground states of matter. Experiments that determine the low temperature state of a material (for example, by growing a single crystal from melt) are done by careful annealing, first melting the substance, then lowering the temperature slowly and spending a long time at temperatures in the vicinity of the freezing point. If this is not done, and the substance is allowed to get out of equilibrium, the resulting crystal will have many defects, or the substance may form a glass, with no crystalline order and only metastable, locally optimal structures. Finding the low temperature state of a system when a prescription for calculating its energy (cost function) is given is an optimization problem. However, the concept of the temperature of a physical system has no obvious equivalent in the systems being optimized. Therefore, an effective temperature (in the same units as the cost function) can be introduced and a simulated annealing process carried out in order to obtain better heuristic solutions to combinatorial optimization problems.

Iterative improvement is a commonly applied strategy to such problems. This strategy is much like the microscopic rearrangement processes modeled by statistical mechanics, with the cost function playing the role of energy [36]. One starts with the system in a known configuration. A standard rearrangement operation is applied to all parts of the system in turn, until a rearrangement configuration that improves the cost function is discovered. The rearranged

configuration then becomes the new configuration of the system, and the process is continued until no further improvements can be found. In other words iterative improvement searches in this coordinate space for rearrangement steps which lead downhill. Therefore, this search usually gets stuck in a local but not a global optimum and it is customary to carry out the process several times, starting from different randomly generated configurations, saving the best result.

Accepting only rearrangements that lower the cost function of the system is like extremely rapid quenching from high temperatures to T = 0, so it should not be surprising that resulting solutions are usually metastable. The Metropolis procedure from statistical mechanics provides a generalization of iterative improvement in which controlled uphill steps can also be incorporated in the search for a better solution. Metropolis et al. [41] introduced a simple algorithm that can be used to provide an efficient simulation of a collection of atoms in equilibrium at a given temperature. In each step of this algorithm, an atom is given a small random displacement and the resulting change, ΔE, in the energy of the system is computed. If ΔE is less or equal 0, the displacement is accepted, and the configuration with the displaced atom is used as the starting point of the next step. The case ΔE greater then 0 is treated probabilistically: the probability that the configuration is accepted is $P(\Delta E) = \exp(-\Delta E/k_B T)$. Random numbers uniformly distributed in the interval (0,1) are a convenient means of implementing the random part of the algorithm. One such number is selected and compared with $P(\Delta E)$. If it is less than $P(\Delta E)$, the new configuration is used to start the next step. By repeating the basic step many times, one simulates the thermal motion of atoms in thermal contact with a heat bath at temperature T. This choice of $P(\Delta E)$ has the consequence that system evolves into a Boltzmann distribution.

Using the cost function in place of the energy and defining configurations by a set of parameters (coordinates), it is straightforward with the Metropolis procedure to generate a population of configurations of a given optimization problem at some effective temperature. This temperature is simply a control parameter in the same units as the cost function. The Simulated

annealing (SA) is essentially the Metropolis algorithm applied on a wide class of combinatorial problems. SA consists of first "melting" the system being optimized at a high effective temperature, the lowering the temperature by slow stages until system "freezes" and no further changes occur. At each temperature, the simulation must proceed long enough for the system to reach a steady state. The sequence of temperatures and the number of rearrangements of parameters (coordinates) attempted to reach equilibrium at each temperature is an annealing schedule. Annealing, as implemented by the Metropolis procedure, differs from iterative improvement in that the procedure cannot get stuck since transitions out of a local optimum are always possible at nonzero temperature. Gross features of the eventual state of the system appear at higher temperatures; fine details develop at lower temperatures. A typical SA algorithm runs as follows:

SA ALGORITHM FOR MINIMIZATION

choose an initial configuration

choose an initial temperature $T > 0$

L1: choose a new configuration which is a stochastic small perturbation of the old configuration

compute ΔE = cost function (new configuration)-cost (old configuration)

IF $\Delta E < 0$

THEN old configuration = new configuration

ELSE with probability $\exp(\Delta E/T)$

old configuration = new configuration

IF a long time no decrease in quality or too many iterations

THEN lower temperature T

IF some time no change in cost function anymore

THEN stop

GOTO L1

If the temperature is high, very often unfavorable configurations are accepted. It is a kind of art to choose a successful annealing schedule (a rule for lowering the temperature in the algorithm). In most applications the success of the algorithm can be very sensitive to the choice of the annealing schedule.

As we see in SA the chance p_a that a new configuration will be accepted is $p_a = \exp(\Delta E/T)$ if $\Delta E > 0$; otherwise it is 1. In the Threshold acceptance method (TA) [37] this is replaced by $p_a = 0$ if $\Delta E > E_T$ where E_T is a "threshold energy"; otherwise 1. TA algorithm runs as follows:

TA ALGORITHM FOR MINIMIZATION

choose an initial configuration

choose an initial THRESHOLD T > 0

L1: choose a new configuration which is a stochastic small perturbation of the old configuration

compute ΔE = cost function (new configuration)-cost (old configuration)

IF ΔE < T

THEN old configuration = new configuration

IF a long time no decrease in cost function or too many iterations

THEN lower THRESHOLD T

IF some time no change in cost function anymore

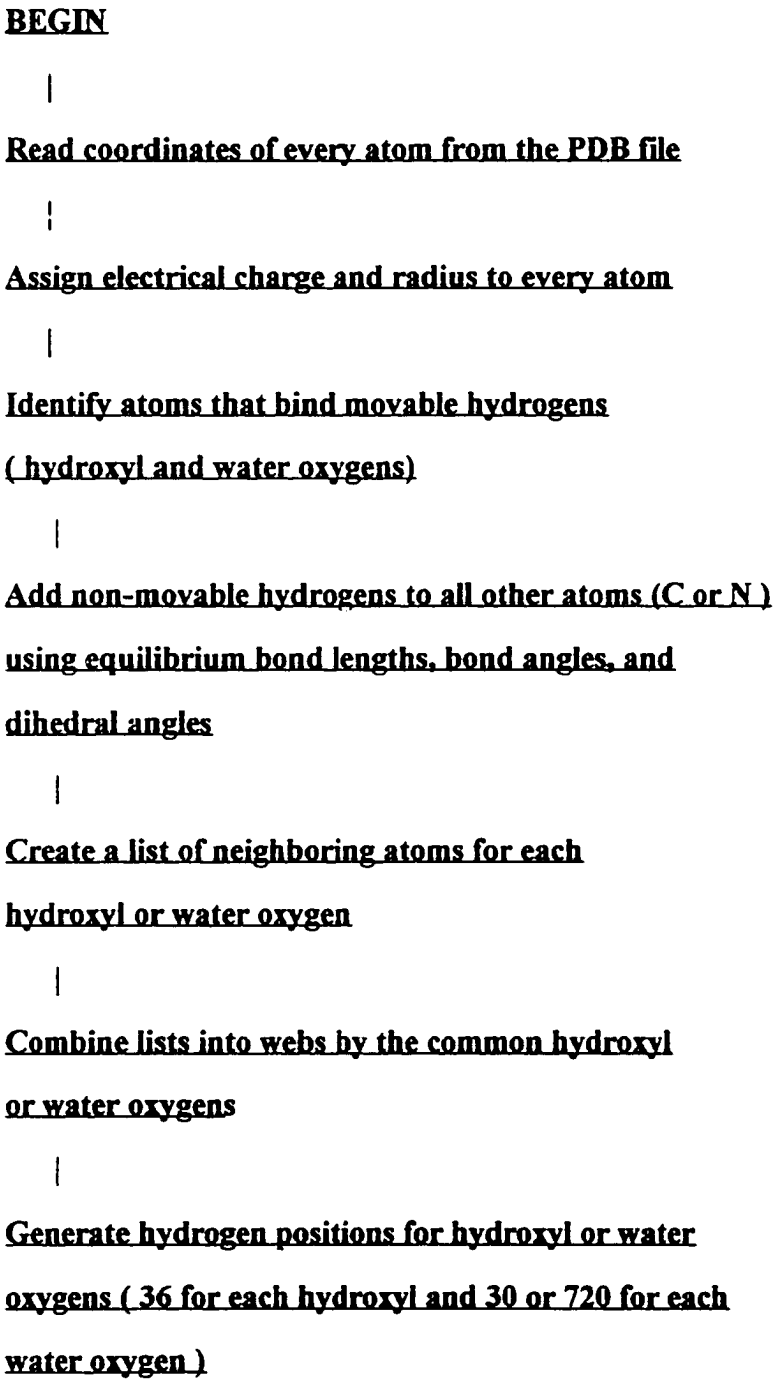
THEN stop

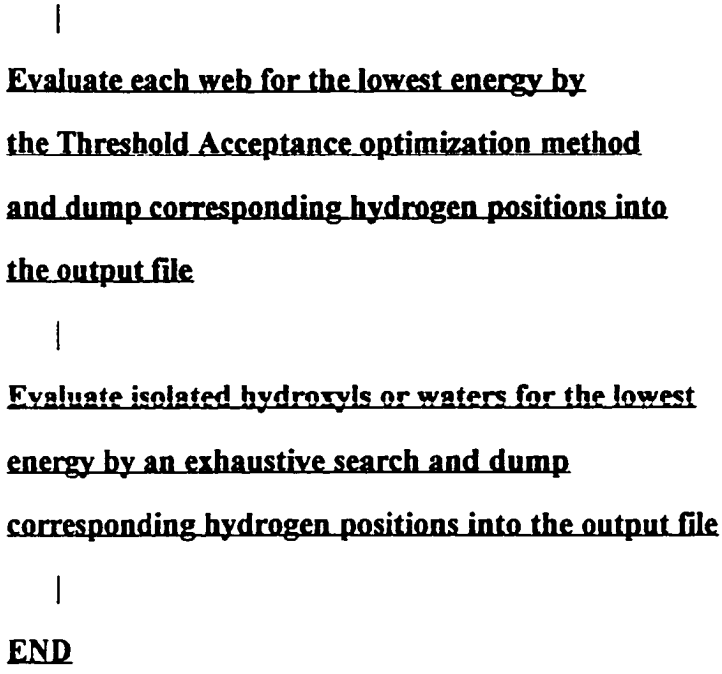
GOTO L1

As we can see, TA is a stochastic procedure like SA, in which the exponential threshold is replaced by a sharp cutoff. The essential difference between SA and TA consists of the different acceptance rules. TA accepts every new configuration which is not much worse than the old one while SA accepts worse solutions only with small probabilities.

APPENDIX C

Here we present a flow chart for the hydrogen positioning algorithm in the program PROTEUS.





References

[1] Pauling, L., Corey, R.B., Branson, H. R. The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U.S.A.* 37:205-211, **1951**.

[2] Pauling, L., Corey, R.B. Configurations of polypeptide chains with favored conformations around single bonds: Two new pleated sheets. *Proc. Natl. Acad. Sci. U.S.A.* 37:729-740, **1951**.

[3] Baker, E. N., Hubbard, R. E. Hydrogen bonding in globular proteins. *Prog. Biophys. Mol. Biol.* 44:97-179, **1984**.

[4] Branden, C., Tooze, J. "Introduction to Protein Structure." New York: Garland, **1991**.

[5] Mc Donald, I. K., Thornton, J. M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* 238:777-793, **1994**.

[6] Bass, M. B., Hopkins, D. F., Jaquysh, W. A. N., Ornstein, R. L. A method for determining the positions of polar hydrogens added to a protein structure that maximizes protein hydrogen bonding. *Proteins* 12:266-277, **1992**.

[7] Jeffrey, G. A. "An introduction to Hydrogen Bonding". New York: Oxford University Press, **1997**. p.204.

[8] Brunger, A. T., Karplus, M. Polar hydrogen positions in proteins: Empirical energy placement and neutron diffraction comparison. *Proteins* 4:148-156, **1988**.

[9] McCammon, J. A., Gelin, B. R., Karplus, M. Dynamics of folded proteins. *Nature* 267:585-590, **1977**.

[10] Brooks, B. R., Karplus, M. Harmonic dynamics of proteins: Normal modes and fluctuations in bovine pancreatic trypsin inhibitor. *Proc. Natl. Acad. Sci. U.S.A.* 80:6571-6575, **1983**.

[11] Brunger. A. T., Brooks, C. L., Karplus, M. Active site dynamics of ribonuclease A. *Proc. Natl. Acad. Sci. U.S.A.* 82:8458-8562. **1985**.

[12] Post, C. B., Brooks, B. R., Dobson, C. M., Artymuik, P., Cheetham, J., Phillips, D. C., Karplus, M. Molecular dynamics simulations of native and substrate-bound lysozome. *J. Mol. Biol.* 190:455-479, **1986**.

[13] Nillson, L., Karplus, M. Empirical energy functions for energy minimization and dynamics of nucleic acids. *J. Comp. Chem.* 7:591-616, **1986**.

[14] Weiner, S. J., Kollman, P., Nguyen, D. T., Case, D. An all atom force field for simulations of proteins and nucleic acids. *J. Comp. Chem.* 7:230-252, **1986**.

[15] Scheraga, H. A. Recent progress in the theoretical treatment of protein folding. *Biopolymers* 22:1-14, **1983**.

[16] McCammon, J. A., Harvey, S. C. "Dynamics of Proteins and Nucleic Acids." Cambridge: Cambridge University Press, **1987**.

[17] Hooft, R. W. W., Sander, C., Vriend, G. Positioning Hydrogen Atoms by Optimizing Hydrogen-Bond Networks in Protein Structures. *Proteins* 26:363-376, **1996**.

[18] Dauber-Osguthorpe, P., Roberts, V. A., Osguthorpe, D. J., Wolff, J., Genest, M., Hagler, A. T. Structure and energetics of ligand binding to proteins: *Escherichia coli* dihydrofolate reductase-trimethoprim, a drug-receptor system. *Proteins* 4:31-47, **1988**.

[19] Weiner, S. J., Kollman, P. A., Nguyen, D. T., Case, D. A. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* 7:230-252, **1986**.

[20] Nilsson, L., Karplus, M. Empirical energy functions for energy minimization and dynamics of nucleic acids. *J. Comput. Chem.* 7:591-616, **1986**.

[21] Kossiakoff, A. A., Shpungin, J., Sintchak, M. D. Hydroxyl hydrogen conformations in trypsin determined by the neutron diffraction solvent difference map method: Relative importance of steric and electrostatic factors in defining hydrogen-bonding geometries. *Proc. Natl. Acad. Sci. U.S.A.* 87:4468-4472, **1990**.

[22] Glick, M., Goldblum, A. A Novel-Based Stochastic Method for Positioning Polar Protons in Protein Structures From X-Rays. *Proteins:Structure, Function, and Genetics* 38:273-287, **2000**.

[23] Frauenfelder, H., Wolynes., P. G. Biomolecules: Where the Physics of Complexity and Simplicity Meet. *Phys. Today* Feb, 58-64, **1994**.

[24] Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M. CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4:187-217, **1983**.

[25] Stillinger, F. H., Rahman, A. Improved simulation of liquid water by molecular dynamics. *J. Chem. Phys.* 60:1545-1557, **1974**.

[26] Jorgensen, W. L., Chandrasekhar, J., Madura, J. D. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:927-935, **1983**.

[27] Paulsen, M. D., Ornstein, R. L. A 175 ps molecular dynamics simulation of camphor-bound cytochrome P-450_{cam}. *Proteins* 11:184-204, **1991**.

[28] Allen, F. H., Kennard, O., Taylor, R. Systematic analysis of structural data as research technique in organic chemistry. *Acc. Chem. Res.* 16:146-153, **1983**.

[29] Bharadwaj, R., Windemuth, A., Sridharan, S., Honig, B., Nicholls, A. *J. Comput. Chem.* **1995**, 16, 898.

[30] Pethig, R. *Annu. Rev. Phys. Chem.* **1992**, 43, 177.

[31] Rupley, J. A.; Careri, G. *Adv. Protein. Chem.* **1991**, 41, 37.

[32] Kuntz, I. D., Jr.; Kauzmann, W. *Adv. Protein. Chem.* **1974**, 28, 239.

[33] Nandi, N., Bagchi, B. Dielectric Relaxation of Biological Water. *J. Phys. Chem.* 101:10954-10961, **1997**.

[34] Jeffrey, G. A. "An introduction to Hydrogen Bonding". New York: Oxford University Press, **1997**. p.237.

[35] Shu, F., Ramakrishnan, V., Schoenborn, B. P. Enhanced visibility of hydrogens atoms by neutron crystallography on fully deuterated myoglobin. *Proc. Natl. Acad. Sci.* 97(8):3872-7, Apr 11, **2000**.

[36] Kirkpatrick, S., Gelatt, C. D., Vecchi, M. P. Optimization by Simulated Annealing. *Science* 220:671-680, **1983**.

[37] Dueck, G., Scheuer, T. Threshold Accepting: A General Purpose Optimization Algorithm Superior to Simulated Annealing. *J. Comp. Phys.* 90:161-175, **1990**.

[38] D. E. Goldberg, "Genetic Algorithms in Search, Optimization, and Machine Learning." Addison Wesley, Reading, MA, **1989**.

[39] Cvijovic, D., Klinowski J. Taboo Search: An Approach to the Multiple Minima Problem. *Science* 267:664-666, **1995**.

[40] Barhen, J., Protopopescu, V., Reister, D. TRUST: A Deterministic Algorithm for Global Optimization. *Science* 276:1094-1097, **1997**.

[41] N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, *J. Chem. Phys.* 21, 1087 (**1953**).

[42] Weiner, S. J., Kollman, P. A., Case, D., Singh, U. C., Ghio, C., Alagona, G., Profeta, P. S. & Weiner, P. *J. Am. Chem. Soc.* **1984**, 106, 765.

PART II

I. INTRODUCTION

1.1 General overview of photosynthesis

Photosynthesis is the most important biological proces on Earth. The capture of solar energy by photosynthetic organisms, including plants, algae and a variety of types of bacteria, and its conversion into the chemical energy is the primary energy source of almost all the living world. For example, in plants the light - driven reactions convert energy of light into a stable transmembrane potential. The transmembrane potential results from a difference in electrochemical activity of protons across the cell membrane. It has both electrical and chemical contributions. The first being the difference in electrostatic potential from all ions which affect the energy of the charged protons and the latter coming from difference in the proton concentration. Oxygen (O_2) is evolved and ATP and NADPH are formed during the light reactions. The light reactions are a series of electron transfer reactions occurring only during light illumination. In a series of dark reactions, which can occur in both light and darkness, ATP and NADPH are used to reduce CO_2 to form glucose and other vital organic products. Both the light and dark reactions take place in the chloroplasts. The light reactions [1-8] occur in two closely coupled pigment systems. Light energy is absorbed by a network of antenna pigment - proteins (light harvesting complexes (LHCs)). It is then very efficiently transported through energy transfer (Appendix A) to the photochemical reaction center (RC) where the energy is converted through a sequence of electron transfer reactions to the transmembrane potential. The light harvesting complexes and reaction centers are also present in photosynthetic bacteria.

Three types of pigments can be present in these systems which act to absorb light or serve as energy transfer sites: (1) bacteriochlorophyll or chlorophyll, (2) carotenoid and (3) bilin (open chain tetrapyrolle) depending on type of organism (a plant, an algae or a bacteria).

Carotenoids, which are the focus of this part of the dissertation, are especially important in photosynthetic systems where they have the dual functions of light harvesting and photoprotection [9,10]. They collect light at wavelengths where bacteriochlorophylls or chlorophylls are inefficient absorbers. In their photoprotective function they minimize the formation of singlet molecular oxygen which strongly reacts with lipids, proteins and nucleic acids (Figure 1). Cells exposed to high concentrations of singlet oxygen die rapidly.

One remarkable property of carotenoids, which has made them a valuable experimental tool, is their extreme sensitivity to the electric field. Their absorption spectrum is shifted instantaneously in response to an electric field. This has been widely used to measure the transmembrane potential [11]. The difference in the electrostatic potential across an active membrane produces an electric field of ~10^5 V/cm. A shift of the carotenoid absorption spectrum in vitro is proportional to the square of the applied electric field [12]. A quadratic response in the applied electric field is characteristic for molecules without dipole moments. Surprisingly, the shift of the absorption spectrum of carotenoids in vivo (both in plants and photosynthetic bacteria) has been found to be linear in the applied electric field [11]. This result has been explained by assuming that carotenoids in vivo are exposed to a local electric field that is much stronger than the applied electric field. The magnitude of that local electric field was estimated to be as 10^6-10^7 V/cm. The experiments in vitro used the applied field of ~10^5 V/cm. As described in detail below the effect of the local electric field is to induce dipole moments in the carotenoid that then give rise to a linear shift in the presence of an applied electric field.

If a local field explains the linear shift of the carotenoid absorption spectrum in vivo, what could be a source of the local field? In photosynthetic systems carotenoids are bound by proteins (LHCs or RCs). Proteins carry charges and it is well established that electrostatic interactions play a central role in a variety of functions associated with proteins [13-15]. For example, they govern the interactions between proteins and ligands in enzymatic reactions; they are important in

determining the stability of folded proteins. Charged amino acids are usually distributed over the surface of proteins to maintain solubility and to define docking sites. The charged amino acids of a light-harvesting complex or photochemical reaction center could generate the local field. The local field would induce dipole moments in the bound carotenoid molecule which will give rise to a linear shift of the carotenoid absorption spectrum in the presence of an applied electric field.

Continuum electrostatics provides a rational approach to the treatment of electrostatic interactions in proteins, since it incorporates the essential electrostatic features of the physical system under study: the dielectric constants of the solvent and protein, the solvent's ionic strength and the locations and magnitudes of the protein's charges. It has been shown that the computer program DelPhi [16-18] (Appendix B), which provides a numerical solution to the Poisson - Boltzmann equation, is a reliable means for calculating electrostatic potentials and electric fields in and around proteins.

What the reader should expect in this part of the dissertation? We will calculate the local electric field, generated by the protein's charges across carotenoids in resolved three-dimensional crystal structures of LHCs and RCs from bacterial sources using program DelPhi. We will identify for each structure charged amino acids responsible for the corresponding local field. Knowing these amino acids is useful information, since that will open an opportunity to mutate them and to gain an additional insight in function of carotenoids in photosynthetic systems from bacterial sources. Further, we will show contributions to the local field depending on the type of the residue (for example, backbone,charged residue, polar residue,...) including also contributions of water molecules. This will give us some light on general principles of the generation of the local field.

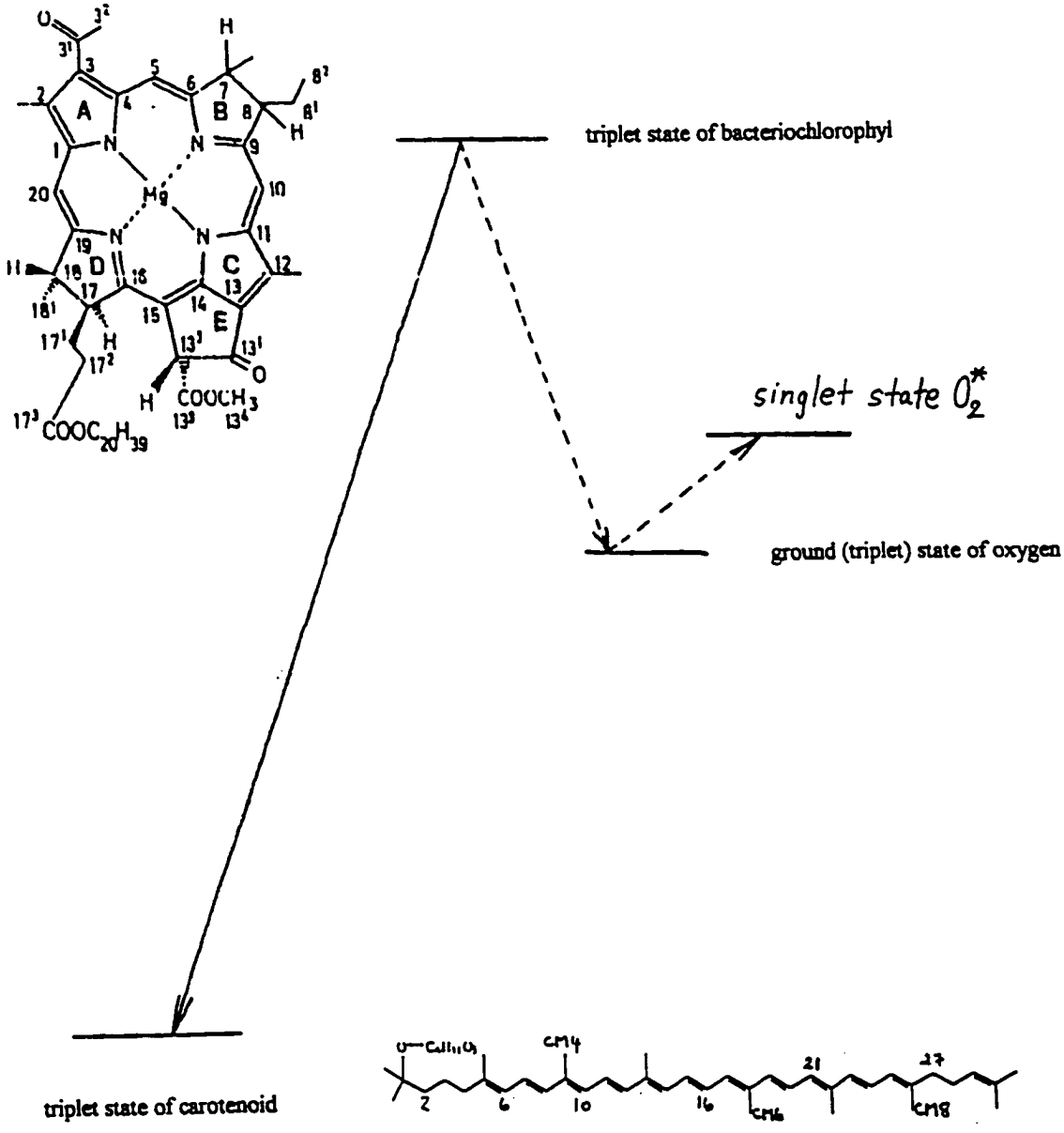


Figure 1.

1.2 Carotenoids

1.2.1 Structure of Carotenoids

The basic structure of carotenoids is a symmetrical tetraterpene skeleton formed by tail to tail linkage of two C_{20} units. This long system of conjugated π double bonds is responsible for carotenoids absorbing light in the visible region, giving them their strong coloration (yellow, orange and red) and is an essential requirement for carotenoids to participate in the energy transfer process.

1.2.2 Response to an electric field

The probability of the photon being absorbed by a molecule depends on the magnitude of the transition dipole moment between the two states (for example between the ground and an excited state). The transition dipole moment is defined by the integral

$$e < \psi_k | \sum_i z_i \vec{r}_i | \psi_n > \tag{1}$$

where $\mathbf{r}_i$ is the position vector of the ith particle (electron or proton), with a charge $z_i e$, in the molecule. This is a vector quantity, with magnitude and direction. If k=n then eq. 1 is just the dipole moment in the state ψ_k. The total dipole moment of the molecule in the state ψ_k in an applied electric field **E** is given by

$$\vec{\mu}_k^E = \vec{\mu}_k^0 + \alpha_k \vec{E} \tag{2}$$

where μ^E and μ^0 are the dipole moments in the presence and in the absence of the electric field **E** in the state ψ_k and α_k is the polarizability tensor in that state. The corresponding expression for the energy of the molecule is

$$U_k^E = U_k^0 - (\vec{\mu}_k^0 + \frac{1}{2}\alpha_k \vec{E})\vec{E} \tag{3}$$

where U's are the energies in the presence and in the absence of the electric field **E**. If the ground and an excited state of the molecule have different total dipole moments (differ in dipolar character), the energy gap between them will change in the electric field **E** (Figure 2). This change in energy is seen as a change in the frequency of absorption (Stark effect) such that

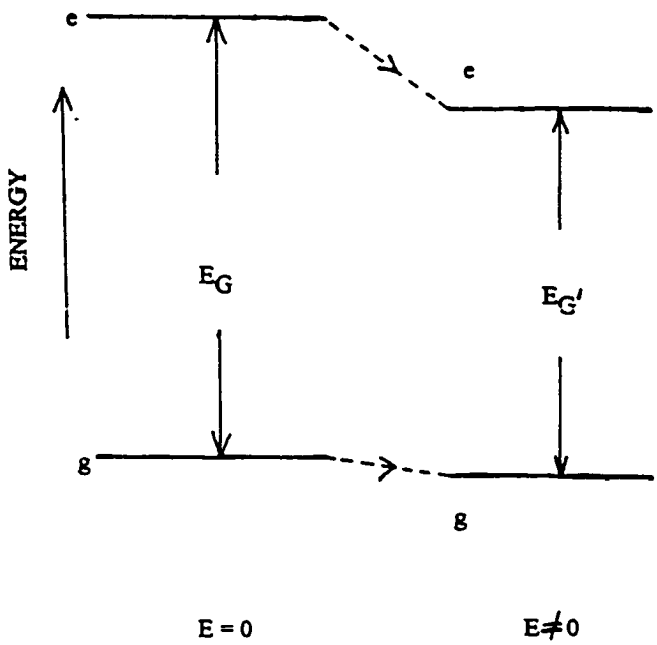
$$\Delta\nu = -\frac{1}{h}[(\vec{\mu}_e^0 - \vec{\mu}_g^0)\vec{E} + \frac{1}{2}(\alpha_e - \alpha_g)\vec{E}\cdot\vec{E}] \tag{4}$$

where $\Delta\nu$ and h are the frequency shift and Planck's constant respectively, and subscripts e ang g refer to excited and ground states.

Since carotenoids have a center of symmetry in the conjugated double-bond system, they should lack dipole moments in the ground and excited states in the absence of a field **E** . Thus, the frequency shift should be proportional to the square of the electric field **E** (second term in $\Delta\nu$). Experiments in vitro showed such response. But, in vivo (using a mutant of *Rb. sphaeroides*, G1C, containing only a single carotenoid component) found a linear response with $\Delta\nu$ of ~0.4 nm [19] to the electric field **E**. The magnitude of the applied field **E** in these experiments was 10^5 V/cm. Even using the large changes in polarizability observed for long, conjugated molecules like carotenoids, the second term predicted values of $\Delta\nu$ of 0.05 nm or less for the band shift. This is two orders of magnitude less than those experimentally reported. However, a linear response in the electric field **E** can be obtained if carotenoids in vivo are exposed to a local electric field [19] generated by the protein matrix into which the carotenoid molecule was embedded. This field polarizes the carotenoid leading to different dipole moments in the ground and excited states in the absence of the applied field **E**. Thus, when an external field is applied the first term of $\Delta\nu$ must be taken into consideration also. Using electrochromic theory [20-22], the shift, $\Delta\lambda_m$, of the peak wavelength λ_m, is then approximately

$$\Delta\lambda_m \approx \frac{\lambda_m^2}{2hc}\Delta\alpha(E_0 + E)E \qquad (5)$$

where $\Delta\alpha$ is the difference in polarizability between the excited and ground carotenoid's state, and E_0 is a local electric field. c is the speed of light and h is Planck's constant. Eq (5) for the local electric field of ~10^6 V/cm gave measured band shifts (~0.4 nm for a mutant of *Rp. sphaeroides*, G1C [11]). The applied field was varied two to threefold by a series of flashes. After the first flash, the isosbestic point was about 2.5 nm shifted from the absorbance peak in the dark, corresponding to a bandshift of 5nm. In subsequent flashes, although the field have been increasing, the isosbestic point moved only a further 0.4 nm. Since the magnitude of the local field was assumed to be at least 10 times larger than the applied field **E**, the local field produced a linear response in the presence of the applied field.



g = energy level of the ground state
e = energy level of an excited state
E_G = energy gap when the field E is not present
E_G'= energy gap when the field E is present

$E_G > E_G'$ ——> RED SHIFT

Figure 2.

1.2.3 Evidence of a strong local field in proteins and membranes

Although, the local field has been never directly measured, there is evidence of a strong local electric field:

- Stark effect measurements of the carotenoid band shift in an applied field are linear in the field [23-25];

- a large spectral shift of carotenoids between the peak wavelength in the protein in the absence of the applied field and the peak wavelength in the solution (~ 25-27 nm) [26];

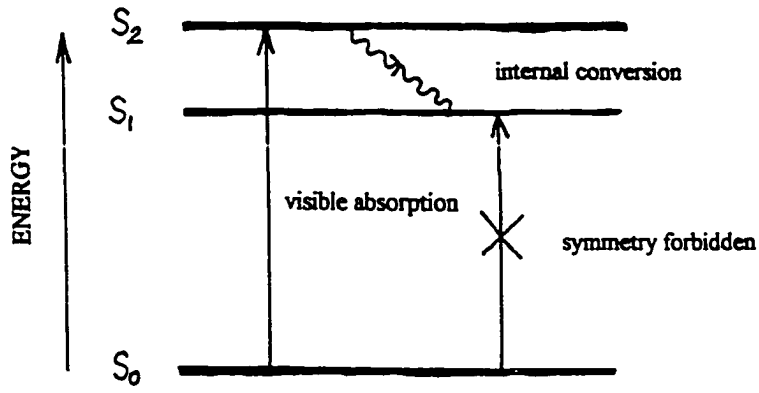
- other indications of large local fields in photosynthetic proteins (for example, the observed polarizability of the special pair, P, of *Rb. sphaeroides* suggested a local field greater than 1.2 x 10^6 V/cm at 1.5 K [27]; the calculated average field of 2.3 x 10^6 V/cm in RC from *Rps. viridis* is consistent with that value [28]; the behavior of spin labels at different depths in phospholipid bilayers indicated 10^6-10^8 V/cm short range dipole fields due to fixed charges on the membrane surface [29]).

1.2.4 Electronic states of Carotenoids

Carotenoids posses two low - lying electronic excited states: S_2 ($1B_u$) and S_1 ($2A_g^1$) states (Figure 3) [30]. Their strong visible absorption is thought to arise from a ground S_0 ($1A_g^1$) state to a S_2 electronic state transition. The S_0 to S_1 transition is symmetry forbidden. The S_0 to S_2 transition increases in λ_{max} (becomes lower in energy) with increasing π - electron conjugation. The singlet excitation energy of the carotenoids (when the carotenoid molecule is either in S_2 or S_1 state) is transferred by an energy transfer mechanism (Appendix A) to the ground state of the bacteriochlorophylls or chlorophylls by two pathways. The first pathway is from the carotenoid S_2 state to the ground state of the bacteriochlorophyll or chlorophyll. The second pathway is from the carotenoid S_1 state (after internal conversion from S_2 to S_1) to the ground state of the bacteriochlorophyll or chlorophyll. The final result is that the carotenoid

molecule is returned to its ground state and the bacteriochlorophyll or chlorophyll molecule is excited.

The triplet excited bacteriochlorophyll or chlorophyll sensitizes the formation of the singlet molecular oxygen (an excited state of molecular oxygen) . This is a very powerful oxidant which can react with lipids, proteins and nucleic acids so that cells exposed to high concentrations of singlet oxygen die rapidly. The photoprotective function of carotenoids utilizes triplet - triplet energy transfer from bacteriochlorophyll or chlorophyll to carotenoid which occurs much more readily than energy transfer from bacteriochlorophyll or chlorophyll to molecular oxygen (Figure 4). Carotenoids can also accept excitation energy from singlet molecular oxygen if any should be formed returning it to the stable triplet form of O_2(ground state).



Low-lying electronic states of a carotenoid (singlet states)

Figure 3.

1.2.5 Function of Carotenoids

Carotenoids enhance the light - capturing ability of bacteriochlorophylls or chlorophylls in LHCs of photosynthetic organisms by collecting light at wavelengths in the spectral region where bacteriochlorophylls or chlorophylls are inefficient absorbers. The singlet excitation energy of the carotenoids is then transferred to the ground state of the bacteriochlorophylls or chlorophylls after which bacteriochlorophylls or chlorophylls become excited. The energy transfer is usually discussed in terms of the Forster and / or Dexter energy transfer mechanisms (Appendix A). In their protective role carotenoids act as quenchers(competitors to molecular oxygen) of the triplet state of bacteriochlorophylls or chlorophylls. This, which is the most important function of the carotenoids in RC, is usually discussed in terms of Dexter energy transfer mechanism. A third, and often neglected function of carotenoids in antenna complexes is that carotenoids maintain internal structural stability of antenna complexes, especially in membrane antenna complexes, where carotenoids are embedded into a protein matrix making contacts with the matrix and bacteriochlorophylls or chlorophylls.

We will calculate the local electric field across carotenoids in resolved three-dimensional crystal structures of light-harvesting complexes and reaction centers from bacterial sources using program DelPhi. The resolved three-dimensional crystal structures are in a Brookhaven Protein Data Bank format file (pdb file).

1.2.6 Light-harvesting complexes (LHCs)

The first two resolved light-harvesting complexes (LHC) were not of the membrane-bound variety.In 1975 the structure of a soluble LHC (pdb file is 3BCL) was obtained, isolated from a species of green sulfur bacteria that lives at a depth of about 10 meters in lakes

[60].In 1985 a different crystal structure of another soluble LHC [31] was resolved, this one from the cyanobacterium *Mastigocladus laminosus* (1B33). In 1990 a similar structure (1CPC) from the cyanobacterium *Fremyella diplosiphon* [32] was obtained.

The sulfur bacterium's LHC which absorbs blue light at about 460 nm and in the near infrared, consists of three identical subunits, each containing seven bacteriochlorophylls. The cyanobacterium's LHC differs in the pigment used - phycobilins or phycocyanobilin (types of carotenoids), which absorbs in the 500 to 650 nm range (green, yellow or orange light). The cyanobacterium's LHC consists of tiny rodlike assemblies.

In 1995 two sorts of membrane-bound LHC found in purple, nonsulfur photosynthetic bacteria which live almost everywhere were described [33]: LH1 has one ring of bacteriochlorophylls plus protein, while LH2 (1KZU) [34] contains two rings of bacteriochlorophyll molecules, one close to the membrane surface with the second set in the middle of the membrane bilayer. This arrangement puts each ring of pigments in a different chemical environment, so they absorb at different wavelengths, extending the spectral range of light harvested.This is very helpful, since these bacteria often live in murky water, so they have to use whatever light is available.

In 1996 the crystal structure of membrane-bound LH2 from *Rhodospirillum Molischianum* (1LGH) [35] similar to 1KZU and the structure of a soluble LHC from *A. carterae* (the dinoflagelate) (1PPR) [36] were resolved. The structure of 1PPR is especially interesting. The peridinin (a type of carotenoid) and chlorophyll molecules are tightly packed within a vessel formed by the protein (only 3 to 4 angstroms apart) and oriented so that energy can easily be transferred from peridinins to the chlorophylls. This tightly packed arrangement probably favors the Dexter energy transfer mechanism between carotenoids and chlorophylls.

1.2.7 Photosynthetic reaction centers (RCs)

Photosynthetic reaction centers are proteins that store a photon's energy through a sequence of electron transfer reactions. All RCs isolated from wild-type strains of purple photosynthetic bacterial species contain one bound carotenoid molecule per RC pigment - protein complex [12]. An approximate twofold rotation symmetry relates all other bound chromophores (bacteriochlorophylls, bacteriopheophytins and quinones). The carotenoid molecule is the only cofactor in the RC that doesn't respect that symmetry, that is, an identical carotenoid does not appear on the opposite side of the protein. In all RC complexes, the sole carotenoid is located between the B and C helices of the M protein subunit near the monomeric accesory bacteriochlorophyll (B_B) which lies between the carotenoid and the primary donor (a pair of bacteriochlorophylls) (P).

Carotenoids in RCs are properly placed and energetically tuned to eliminate triplet states of the primary donor as quickly as they formed like in *Rb. sphaeroides*. However, this is not the case for 1,2-dihydroneurosporene from *Rps. viridis* (1PRC) [37,38], since the singlet state of the molecular oxygen is above the primary donor triplet state and arguably the primary donor in a triplet state cannot sensitizes the formation of the singlet molecular oxygen. In addition, the light-harvesting function of 1,2-dihydroneurosporene molecule must be minimal and probably 1,2-dihydroneurosporene molecule is a remnant from an ancestral form of *Rps. viridis* that required photoprotection.

II. THE OBJECTIVES OF THIS PART OF THE DISSERTATION

2.1 An average electrostatic field across corresponding carotenoid

Using the program DelPhi and atom coordinates in Brookhaven Protein Data Bank format: 1KZU,1CPC,1B33,3BCL,1PPR,1PRC and two structures of *Rb. sphaeroides* RC (Authors: H. Komiya, T. O. Yeates, A. J. Chirino, D. C. Ress, J. P. Allen, G. Feher and Urlich Ermler, Susan Buchanan, Guenter Fritzsch, Harmut Michel) we will calculate an average electrostatic field across carotenoids (chlorophylls for 3BCL) and identify residues functionally important for generating of the electrostatic field.

Table 1. Chromophores that will be used in the calculation of the local field

SCOP- criteria		protein	source	function	location	chromoph.
class	fold					
1	2	1KZU	bacteria	LHC(light harvesting)	membrane	rhodopin glucoside
3	4	1CPC	bacteria	LHC(light harvesting)	soluble	phycocyanobilin
3	4	1B33	bacteria	LHC(light harvesting)	soluble	phycocyanobilin
5	6	3BCL	bacteria	LHC(light harvesting)	soluble	bacteriochlorophyl
3	7	1PPR	algae	LHC(light harvesting)	soluble	peridinin
5	8	1PRC	bacteria	RC(photoprotective)	membrane	dihydro-neurosporene
5	8	rcs	bacteria	RC(photoprotective)	membrane	dihydro-neurosporene

rcs is notation for two structures of *Rb. sphaeroides* RC (Authors: H. Komiya, T. O. Yeates, A. J. Chirino, D. C. Ress, J. P. Allen, G. Feher and Urlich Ermler, Susan, Buchanan, Guenter Fritzsch, Harmut Michel)

SCOP (Structural Clasification of Proteins):

1 means membrane and cell surface proteins and peptides class; 2 means membrane all alpha fold

3 means all alpha proteins class; 4 means globin like fold

5 means all beta proteins class; 6 means bacteriochlorophyl A protein fold

7 means peridinin chlorophyl protein fold

8 means photosynthetic reaction center, M chain, cytoplasmic domain fold;

CAROTENOIDS:

[rhodopine glucoside ($C_{46}H_{66}O_6$)],[phycocyanobilin ($C_{33}H_{37}N_4O_6$)],

[peridinine ($C_{39}H_{50}O_7$)],[dihydro-neurosporene ($C_{40}H_{62}$)],

in 3BCL the chromophore is bacteriochlorophyl A ($C_{55}H_{74}N_4O_6Mg$);

2.2 Regularities in designing of the electrostatic field by a LHC or RC

We will evaluate a possible dependence of the electrostatic field on type of the protein matrix (LHC or RC). In other words, we will test whether there are any regularities in generating or designing of the electrostatic field by a LHC or RC. If so, what are the differences between them, since the light-harvesting is the main function of a carotenoid in a LHC, while in a RC it is the photoprotective function.

2.3 Dependence of the electrostatic field of LHC on type of the pigment and the location in the cell (membrane or solution)

We will evaluate for LHCs how the electrostatic field is influenced by the types of pigment (a carotenoid or bacteriochlorophyll (3BCL)) and by the fact that some LHCs are membrane-bound, while others are soluble.

For carotenoids, the difference between dipole moments in the ground and excited states is oriented approximately along the carotenoid C-C backbone. We will ask if this direction corresponds to the direction of the calculated field. Bacteriochlorophylls have transition dipoles which enable them to promote energy transfer across long distances (10-80 angstroms). The major transition dipoles, corresponding to the singlet-singlet energy transfer, are denoted Q_x and Q_y. These dipoles differ due to the asymmetry of the conjugated structure. We will compare the directions of these dipoles with the directions of the calculated electrostatic field to see if they are aligned with the field.

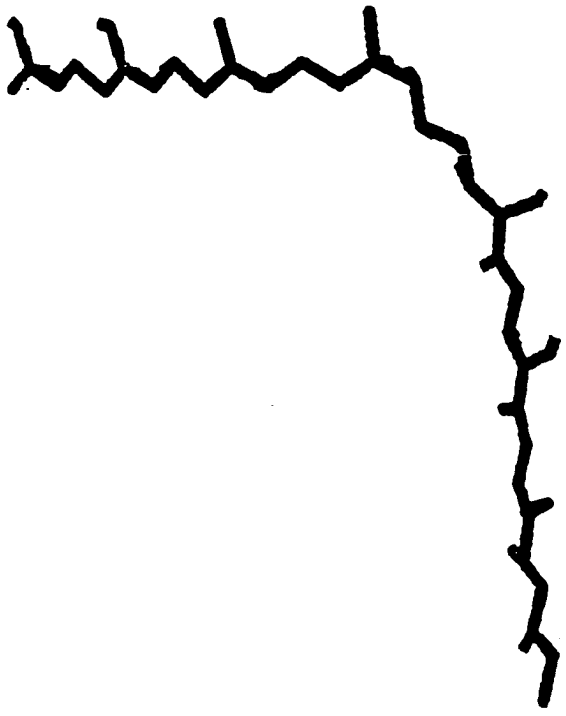
Cell membranes are low-dielectric media, while water is a high dielectric media. The transfer rate for Forster energy transfer depends on the refraction index of a media as (refraction index)$^{-4}$ (Appendix A). In other words, the transfer rate falls off as (dielectric constant)$^{-2}$.Therefore, we expect that Forster energy transfer would not be favorized in a high dielectric medium as water (dielectric constant ~80). We expect that Dexter energy transfer (Appendix A) for soluble LHCs is more appropriate, since Dexter energy transfer doesn't depend on the dielectric constant. However, for membrane-bound LHCs both mechanisms are possible. We expect that all these factors influence the design of the local field by a protein matrix.

2.4 Are the electrostatic field in RC from *Rps. viridis* and hydrophobic interactions with nearby residues strong enough to leave the carotenoid molecule essentially without the photoprotective function?

We will compare the data from 2.1 between 1PRC and *Rb. sphaeroides* RCs to associate differences with the fact that 1PRCs carotenoid does not do the photoprotective function; the aim is to investigate a possibility that the electrostatic field in the reaction center from *Rps. viridis* and hydrophobic interactions with nearby residues are strong enough to lower the primary donor triplet state below the ground state of oxygen (O_2) and raise the triplet state of

1,2-dihydroneurosporene above the primary donor triplet state leaving the carotenoid molecule essentially without the photoprotective function.

Experiments in solutions showed that 1,2-dihydroneurosporene tends to be in an extended (linear) conformation. In solutions there is no a local electric field.The 1,2-dihydroneurosporene in reaction center from *Rps. viridis* is not linear (Figure 4). Some work (isomerisation energy) is needed to bring 1,2-dihydroneurosporene from a linear conformation to the conformation in reaction center from *Rps. viridis*. We want to propose that the local electric field in the reaction center from *Rps. viridis* and hydrophobic interactions with nearby residues could do that work. They, working together, could "bend" 1,2-dihydroneurosporene molecule to the conformation which 1,2-dihydroneurosporene molecule has in reaction center from *Rps. viridis*. This "bending" would deform the electronic states of the 1,2-dihydroneurosporene molecule such that the photoprotective function of 1,2-dihydroneurosporene would be ruled out.



The conformation of 1.2-dihydroneurosporene in reaction center from *Rps. viridis*

Figure 4.

2.5 DelPhi can provide some insight into the magnitude of the protein's dielectric constant

The dielectric constant of a protein is a controversial concept. In physics,the dielectric constant of a medium describe it's response to an applied electric field averaged over some region. In the case of a protein the response arises from three physical processes: electronic polarization, reorientation of permanent dipolar groups (peptide bonds or water molecules), and movement of whole charges (monopoles).

The electronic polarization describes the motion of electrons in response to the applied field. The distortion of the electron cloud around an atom by the field results in an induced dipole. The reorientation of groups such as the peptide bond or water molecules (especially surrounding water molecules in case of a soluble protein) which have large permanent dipoles is an important part of the overall response. Four factors determine the degree of response from permanent dipoles: the dipole moment magnitude, the density of such groups in the protein or solvent (water), the freedom of such groups to reorient and the degree of cooperativity between dipole motions. The larger the fluctuations induced by thermal motions, the more easily permanent dipolar groups can reorient by an applied field. Charge rearrangement can occur where there are charged groups which have some degree of mobility. The most mobile are salt ions in the solvent. In addition, charged groups on proteins, especially the longer side chains on the surface of proteins (Arg, Lys, Glu, Asp) have the ability to change their conformation in response to electrostatic fields.

There are several conceptual issues related to the electrostatic response of a protein. The first one is the distinction between polarity and polarizability. Polarity describes the density of charged and dipolar groups in a particular region of the protein. Polarizability is the potential for reorganizing charges, orienting permanent dipoles, and inducing dipoles. The polarizability depends both on the polarity and the freedom to reorganize in response to an applied electric field.

The second issue is whether the dielectric constant describes well the average polarization produced in some region of the protein by an applied field. It describes electronic polarizability and polarizability of permanent dipolar groups. The response is averaged over some volume of the protein. If the volume is large on the atomic scale then the dielectric constant is used in a macroscopic sense, otherwise it is used in a microscopic sense and different dielectric constants may be used for different atoms or groups. In case of large protein dipole motions (for example protein conformational changes) it is inappropriate to treat the response as a dielectric constant. Such motions should be described explicitly in terms of changes in the protein's charge distribution.

The third issue is related to the use of "effective" dielectric constant to describe the strength of interaction between two protein charges. This is defined as the ratio between the electrostatic interaction energy when charges are in the protein and the electrostatic interaction energy between the same two charges in vacuum. If a protein was homogeneous in terms of its electrostatic response and involved no charge rearrangement, then the effective dielectric constant would describe the response of the protein to the applied electric field. But, this is never the case. The strength of interaction between two protein charges is determined by net contribution from protein, solvent and dissolved ions. Thus, the effective dielectric constant does not give any information about response of any particular region of the protein.

Concepts of polarity, polarizability, dielectric constant and effective dielectric constant raised a confusion about what is the dielectric constant of a protein. The confusion produced a range of answers ranging from high (greater than or equal to that of water [67]) to low (like organic solids [68-70]). Thus, this question is not a useful question and should be applied on a case-by-case basis. Instead, a question, what are the different contributions of a protein to the electrostatic response in an applied field and how to model them, should be raised. However, when the electrostatic response of a protein to an applied field involves little rearrangement (for

example many electron transfer proteins) the concept of the dielectric constant (electronic polarizability and polarizability of permanent dipoles (peptide bonds, water molecules) averaged over same region of the protein) seems appropriate.

Theoretical studies of the normal modes of random α - helices packed with the density found in a protein yield values for the dielectric constant of a protein between 2 and 4 [50] that are in a good agreement with the measured bulk dielectric constant of dry or partially hydrated protein [68]. A value of 4 has been quite successful in describing many functional properties of proteins (like midpoint potentials, pK values [71-72]).

We will show that DelPhi can provide a value of the local field in LH2 from *Rps. acidophila* (1KZU) which is consistent with the value of 4 for the dielectric constant of a protein.

III. METHODS

3.1 The calculation of the local field

Protons were placed on the structures using the program PROTEUS (see the first part of the disertation). The hydroxyl and water protons are the only atoms with partial charges that do not have their positions completely determined by the heavy atoms in the structure. PROTEUS orients the positions of protons on waters and hydroxyls to minimize electrostatic interaction energy, Van der Waals interaction energy, and torsion energy of the proton with residues within 3 angstroms of the hydroxyl or water oxygen.

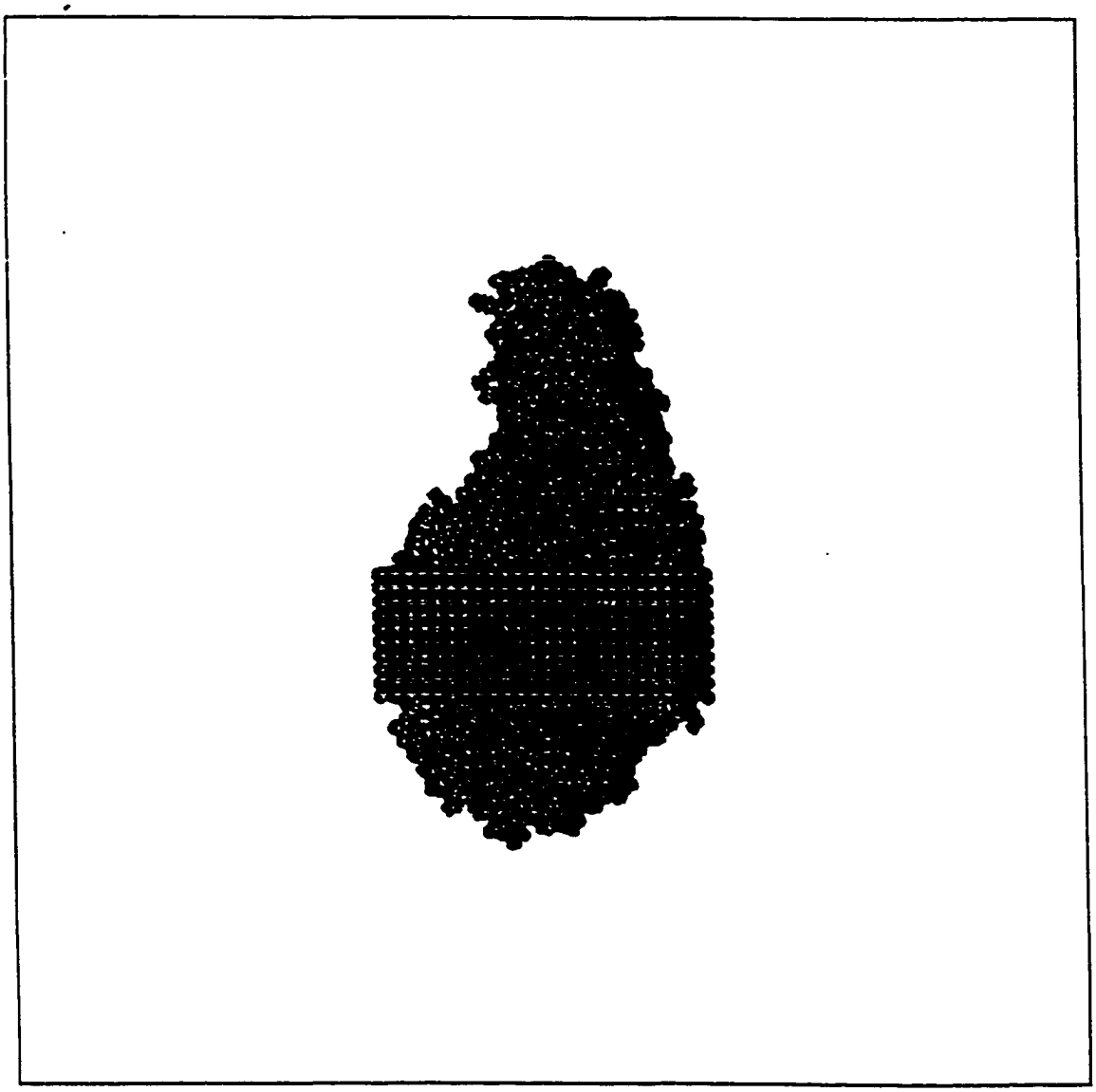
For each structure (Table 1) the local field was calculated using the DelPhi program. DelPhi (Appendix B) first maps the molecular system onto a 65^3 grid by defining charges and radii for all atoms. Charges are distributed to nearby grid points. The midgrid positions within the protein or membrane are assigned a low dielectric constant (2 or 4) while exterior points, or cavities with diameter greater than 2.8 angstroms are given a dielectric constant of 80

representing water. When membrane is included, all internal cavities are assigned the dielectric constant of the protein.

The influence of the membrane (in case of 1KZU, 1PRC, rcs_up) was estimated by including a cube of low dielectric material surrounding the protein (for example 72x72x33 A^3 for 1PRC and 93x93x30 A^3 for 1KZU; Figure 5 and 6). The short 33 angstroms transmembrane dimension is comparable to the detergent layer shown in the neutron diffraction picture of the reaction centers [47]. The following atomic radii: C 1.8, N 1.5, O 1.6, H 1.2, S 1.9, Fe and Mg 1.45 and 1.8 angstroms for membrane atoms were used with CHARMM partial charge set for DelPhi runs [28]. The charge on an amino acid (residue) is not uniformly distributed. For example polar residues (TYR, THR, SER) and the protein's backbone, though neutral overall, are treated as being made up of atoms with positive or negative partial (fractional) charges. The charges on charged (ionizable) residues or the charges on non-polar residues are also distributed among several atoms of the residue. All histidine residues and bacteriochlorophylls were assumed to be neutral. All crystallographic water molecules were treated as part of the protein. Ionic strength effects were not considered. However, mobile ions were found to make essentially no contribution to electrostatic potentials in the interior of RC [28] and probe calculations for LH2 from *Rps. acidophila* changing ionic strength showed no essential contribution to potentials in the interior of LH2.

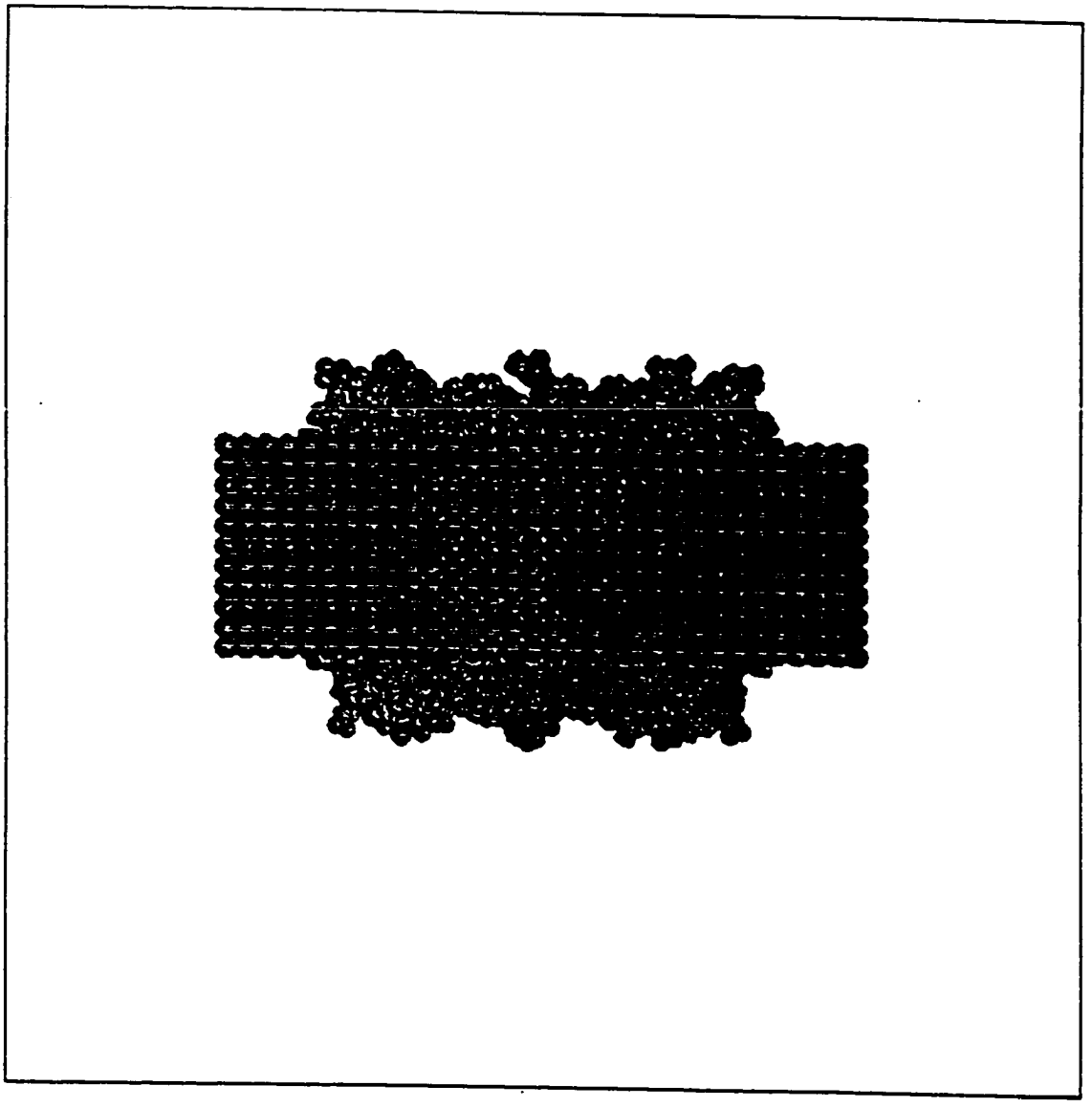
The electrostatic field on each carbon atom of carotenoids or chlorophylls was calculated. Then the average x,y, and z components of the field were obtained for the molecule by summing all x components, all y components and all z components of the electric field across the carbon atoms of the molecule and dividing by the number of carbon atoms. These averaged components were used in order to obtain an average electrostatic field across the molecule. Then the average electric field of each charged, polar and non-polar residue of the protein was calculated on each carbon atom of the carotenoid or chlorophyll molecule and residues generating an average electric

field across the molecule more than or equal to 10^6 V/cm were identified. The average electric fields from water molecules and the protein's backbone were also calculated.



Reaction center (RC) from *Rps. viridis*

Figure 5.



Light-harvesting complex (LH2) from *Rps. acidophila*

Figure 6.

3.2. LH2 from *Rps. Acidophila* (1KZU)

1KZU pdb file contains only one protomer unit. The protomer unit consists of two helices, one carotenoid, and three bacteriochlorophylls. The functional unit is a symmetrical ring of nine protomer units. To get the functional unit we performed eight successive 40 degrees rotations of the protomer unit about the z-axis.

The carotenoid molecule in LH2 from *Rps. acidophila* (rhodopine-glucoside) is nearly a linear molecule. Due to asymmetry (glycosyl ring) rhodopin-glucoside molecule there must be some difference between ground - and excited - state dipole moments even if the field due to the protein matrix is not present ($\Delta\mu_0$). The difference between ground - and excited - state dipole moments was measured for the absorption band of 528 nm [49]. These provided estimates of $\Delta\mu_0$ and the difference between ground - and excited - state dipole moments due to the local field ($\Delta\alpha E_0$). The theory of the frequency shift for a carotenoid molecule was used. For other carotenoids (see Table 1) such calculations were not possible due to a lack of experimental data for the difference between ground - and excited - state dipole moments given the nonlinearity of the carotenoid molecule.

Described below is the theory of the frequency shift for a carotenoid molecule used in case of rhodopine-glucoside.

3.3 Theory of the frequency shift for a carotenoid molecule

3.3.1 The frequency shift

It is known that the frequency shift (in cm^{-1}), $\Delta\nu$, associated with the electronic transition for the carotenoid molecule in the electric field **E** is given by

$$\Delta\nu = -\frac{1}{hc}[(\vec{\mu}_e - \vec{\mu}_g) + \frac{1}{2}(\alpha_e - \alpha_g)\cdot\vec{E}]\cdot\vec{E} \qquad (6)$$

where μ_g and μ_e are the ground and excited state dipole moment vectors in the absence of the field **E** and α_g and α_e are the corresponding polarizability tensors [19-21].

Molecular orbital calculations [39] on different carotenoids indicate that the dipole moments μ_g and μ_e are approximately collinear and lie along C-C backbone of the carotenoid molecule. Thus, the first term in (6) reduces to

$$(\vec{\mu}_e - \vec{\mu}_g)\cdot\vec{E} = (\mu_e - \mu_g)E\cos\theta = \Delta\mu_0 E\cos\theta \tag{7}$$

where θ is the angle between the vector **E** and the vector μ_e-μ_g. Since, the vector μ_e-μ_g is oriented along the carotenoid C-C backbone, θ is also the direction of the field vector relative to a vector along the carotenoid C-C backbone.

The analysis of the second term in eq (2) is complicated by the tensor nature of α_g and α_e. Fortunately, the dominant components of both α_g and α_e most likely lie along the C-C backbone of the carotenoid molecule [20,40]. Therefore,

$$\frac{1}{2}[(\alpha_e - \alpha_g)\cdot\vec{E}]\cdot\vec{E} = \frac{1}{2}(\alpha_e^{zz} - \alpha_g^{zz})E^2\cos^2\phi = \frac{1}{2}\Delta\alpha E^2\cos^2\phi \tag{8}$$

where ϕ is the angle between the electric field vector and the principal (zz) component of the polarizability of the carotenoid molecule defining the z-axis along the carotenoid C-C backbone. Carotenoids are similar to linear polyenes (chains of conjugated π double bonds) since they have the same conjugated structure. They differ from the linear polyenes in their end groups. In the case of long chain polyenes the angles θ and ϕ have approximately the same values. Thus, eq (6) can be approximated as

$$\Delta\nu = -\frac{1}{hc}[\Delta\mu_0 E\cos\theta + \frac{1}{2}\Delta\alpha E^2\cos^2\theta] \quad (9)$$

The asymmetry of the molecule produces an asymmetry of the electronic distribution along the C-C backbone leading to the difference between ground - and excited state dipole moments even if the field **E** is not present. For asymmetric polyenes the magnitude of $\Delta\mu_0$ is about 6-12D [39,20,40,41]. For symmetric polyenes $\Delta\mu_0$ is most likely zero. Since carotenoids have a center of symmetry in the conjugated double-bond system they should lack dipole moments in the ground and excited states if the field **E** is not present. Thus, like for symmetric polyenes $\Delta\mu_0$ is zero for carotenoids.But, if the symmetry in the conjugated double-bond system for carotenoids is violated, like for rhodopin-glucoside molecule in LH2 from *Rps. acidophila* (only one end of the molecule has a glucosil ring) $\Delta\mu_0$ may not be zero.

3.3.2 Electric fields acting on a carotenoid molecule

For a carotenoid molecule in vivo two electric fields act on it. The carotenoid molecule is bound by the protein matrix which produces a local electric field. The effect of the local field (at least ~10^6 V/cm) is to induce dipole moments in the ground and excited carotenoid's states. The protein matrix is embedded in a cell membrane. There is a difference in the electrostatic potential (one component of the transmembrane potential) across the membrane which produces another field (~10^5 V/cm). In addition to these two fields, experiments for band shift of the absorption carotenoid spectrum in vivo add the field generated by the light induced reactions (light-induced field) (~10^5 V/cm) as well as an (external) applied electric field (for example from an electric capacitor) (~10^5 V/cm).

The function of the light-induced field is to supply photons for the excitation of the carotenoid molecule (the carotenoid molecule absorbs a photon and jumps into an excited state).The applied electric field is an outside field controlled by experimentalists. This field is found to linearly shift the carotenoid absorption spectrum in vivo.

The difference between ground - and excited - state dipole moments due to the local field depends on the structure of the molecule. Its value is estimated as $\Delta\alpha E_0$ where E_0 is the value of the local field. For a symmetric carotenoid molecule in the protein there is only $\Delta\alpha E_0$, since the difference between dipole moments for the ground and excited states when there is no field ($\Delta\mu_0$) is zero. For an asymmetric carotenoid molecule in the protein there are essentially two contributions to the difference between ground - and excited - state dipole moments ($\Delta\mu$): $\Delta\mu_0$(the difference between dipole moments for the ground and excited states when there is no any field) and $\Delta\alpha E_0$ (the difference between ground - and excited - state dipole moments due to the local field).

3.3.3 Polarizability

The difference in polarizabilities between the ground - and excited - carotenoid's states ($\Delta\alpha$) must be known to calculate the contribution of the local field.The polarizability depends on the number of conjugated double carbon bonds [20,40,42]. Measurements of the change in polarizability on excitation of crocetin (9 conjugated bonds), lutein (10 conjugated bonds) and bixin (11 conjugated bonds) gave the values of (620 ± 160), 910 and (1245 ± 315) x 10^{-23} cm^3, respectively [14].

3.3.4 Correction factor

Interestingly, the electric field present at the carotenoid molecule due to an externally applied electric field is an unknown quantity because of the protein response (electronic

polarization, reorientation of permanent dipolar groups (peptide bonds or water molecules), and movement of whole charges) to the applied field. This response creates an additional field that also acts on the carotenoid molecule. Due to the irregular structure of the protein, this field cannot be calculated. Quantitatively, this has been described by an unknown correction factor f. For example,fx(the externally applied electric field) determines the electric field present at the carotenoid molecule due to the applied field. In general, f is a tensor quantity; however it is usually treated in a protein as a scalar for simplicity.

The two commonly used expressions for f involve either the Lorentz (f_L) or the spherical cavity (f_C) approximation [44,45,46] (in an elliptical approximation f is 1.0-1.1 [45]). Usually, in an experiment, PVA film is used to embed a protein (LHC) into it. Thus, the difference between the dielectric constants of the protein and PVA must be taken into account. Assuming a spherical protein, the expressions for f are

$$f_L = \left(\frac{3\varepsilon_{PVA}}{2\varepsilon_{PVA}+\varepsilon_P}\right)\left(\frac{\varepsilon_P+2}{3}\right) \tag{10}$$

and

$$f_C = \left(\frac{3\varepsilon_{PVA}}{2\varepsilon_{PVA}+\varepsilon_P}\right)\left(\frac{3\varepsilon_P}{2\varepsilon_P+1}\right) \tag{11}$$

where the first term corrects for the field inside the protein due to the field inside the PVA film, and the second term corrects for the field at the carotenoid molecule; (ε_{PVA}= 3.5) and ε_P are the dielectric constants of the PVA film and protein, respectively. The relation between $\Delta\mu$ and the measured (apparent) value of $\Delta\mu$ is:

$$\Delta\mu = \frac{\Delta\mu_{app}}{f}$$

(12)

3.3.5 The difference between ground - and excited - state dipole moments for rhodopine - glucoside

Because of the glucosyl ring, rhodopin-glucoside molecule must have some difference between ground - and excited - state the dipole moments even if the field due to the protein matrix is not present ($\Delta\mu_0$) . $\Delta\mu_0$ can be estimated with eq (9). For that task the frequency shift ($\Delta\nu$), the local field, and the polarizability ($\Delta\alpha$) are needed. The measured lowest peak in LH2 from *Rps. Acidophila* is 460 nm [49]. James A. Bautista calculated the corresponding peak in a solution to be at ~457 nm. Then, the frequency shift can be estimated as 109 cm^{-1}. Since rhodopin-glucoside has 11 conjugated bonds a reasonable estimate for $\Delta\alpha$ is 10^{-23} cm^3.

The local field polarizes the rhodopin-glucoside molecule contributing $\Delta\alpha E_0$ to the difference between ground - and excited - state dipole moments. Therefore, there are essentially two contributions to the difference between ground - and excited - state dipole moments ($\Delta\mu$): $\Delta\mu_0$(the difference between dipole moments for the ground and excited states in absence of the local field; this comes due to assymetry of the molecule) and $\Delta\alpha E_0$. The measured (apparent) value of $\Delta\mu$ [49] for rhodopin-glucoside molecule in LH2 for the band of 528 nm was 10.3 ± 2.4D. Then equation (12) yields for $\Delta\mu$

$$\frac{10.3D}{f}$$

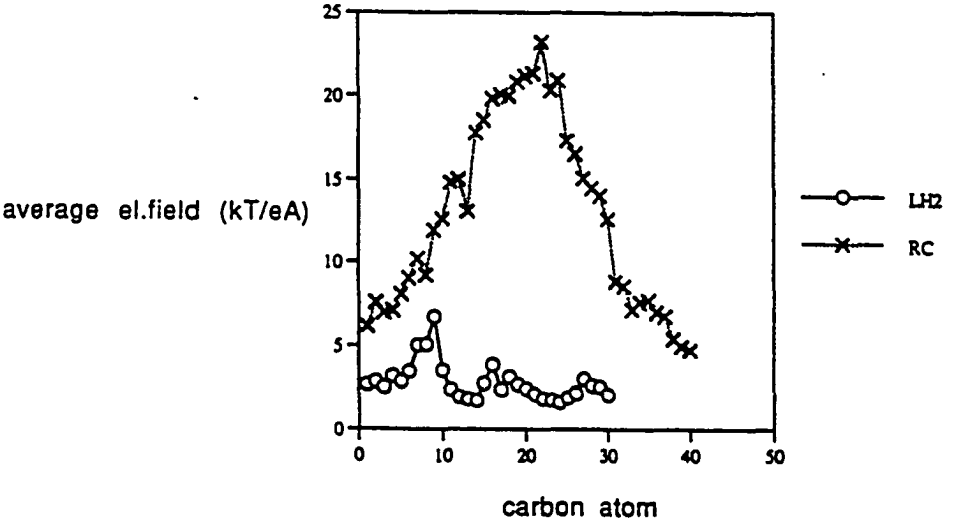
(13)

The above two calculated contributions to $\Delta\mu$ ($\Delta\mu_0$ and $\Delta\alpha E_0$) will be compared with eq (13) as described bellow.

4.1 LH2 from *Rps. acidophila* (1KZU)

4.1.1 The local field in LH2

The calculated average electric field across the carotenoid molecule (rhodopin-glucoside) in LH2 from *Rps. acidophila* with the dielectric constant of 2 inside of the protein was 1.8×10^6 V/cm (Table2). If the dielectric constant of 4 was assigned to the interior of the protein the calculated average field was 1.2×10^6 V/cm. The molecule is nearly linear. Figure 7 shows how the electrostatic field varies across rhodopin-glucoside molecule indicating its nearly constant value. An analyses using program GRASP [48] showed contributions from residues with charge approximately parallel to the C-C carotenoid backbone in the direction to the cytoplasmic side of the membrane, leading to the value of θ nearly zero.



average electrostatic field across carotenoids as a function of carbon atoms

LH2 from *Rps. acidophila*
RC from *Rps. viridis*

Figure 7.

4.1.2 Dependence of the local field on dielectric constant and membrane

Interestingly, there is only small decrease in the field value of 0.6 x 10^6 V/cm if the dielectric constant inside of the protein is increased by the factor of 2 due to the fact that the carotenoid molecule is not burried deeply inside of the membrane like the carotenoid molecule in RC from *Rps. viridis*. Since rhodopin-glucoside spans the depth of the membrane the contribution of the dielectric boundaries (the membrane edges) could not be neglected. However, placing the LH2 complex in a uniform dielectric enormously increases the average electric field across the carotenoid molecule leaving the pattern of the electrostatic field (Figure 7) variation across C-C carotenoid backbone essentially the same. If the dielectric constant of 2 was assigned to the interior and exterior of the protein the calculated average field was 1.1 x 10^7 V/cm and two times less (5.5 x 10^6 V/cm) if the dielectric constant of 4 was assigned. An increase in the membrane thickness would lead to a corresponding increase of the local field across the molecule. To test this we have increased the membrane thickness by 9 angstroms. The corresponding values for the average electric field were 2.2 x 10^6 V/cm and 1.4 x 10^6 V/cm for the dielectric constants of 2 and 4 assigned respectively to the protein's interior.

4.1.3. The difference between ground - and excited - state dipole moments for rhodopine - glucoside

$\varepsilon_{in} : \varepsilon_{out}$	E_0	$\Delta\mu_0$	$\Delta\alpha E_0$	$\Delta\mu_0 + \Delta\alpha E_0$	$\frac{10.3D}{f}$	approximation
4 : 80	1.2	6D	3.6D	9.6D	9.3D	elliptical
					5.4D	Lorentz
					8.1D	Spherical
2 : 80	1.8	6.1D	5.4D	11.5D	9.3D	elliptical
					6.6D	Lorentz
					7.3D	Spherical

The Debye unit, abbreviated as D, equals 10^{-18} esu multiplied by cm.
The esu is an electrostatic unit of charge equals 3.33×10^{-10} coulombs and
cm is centimetar.

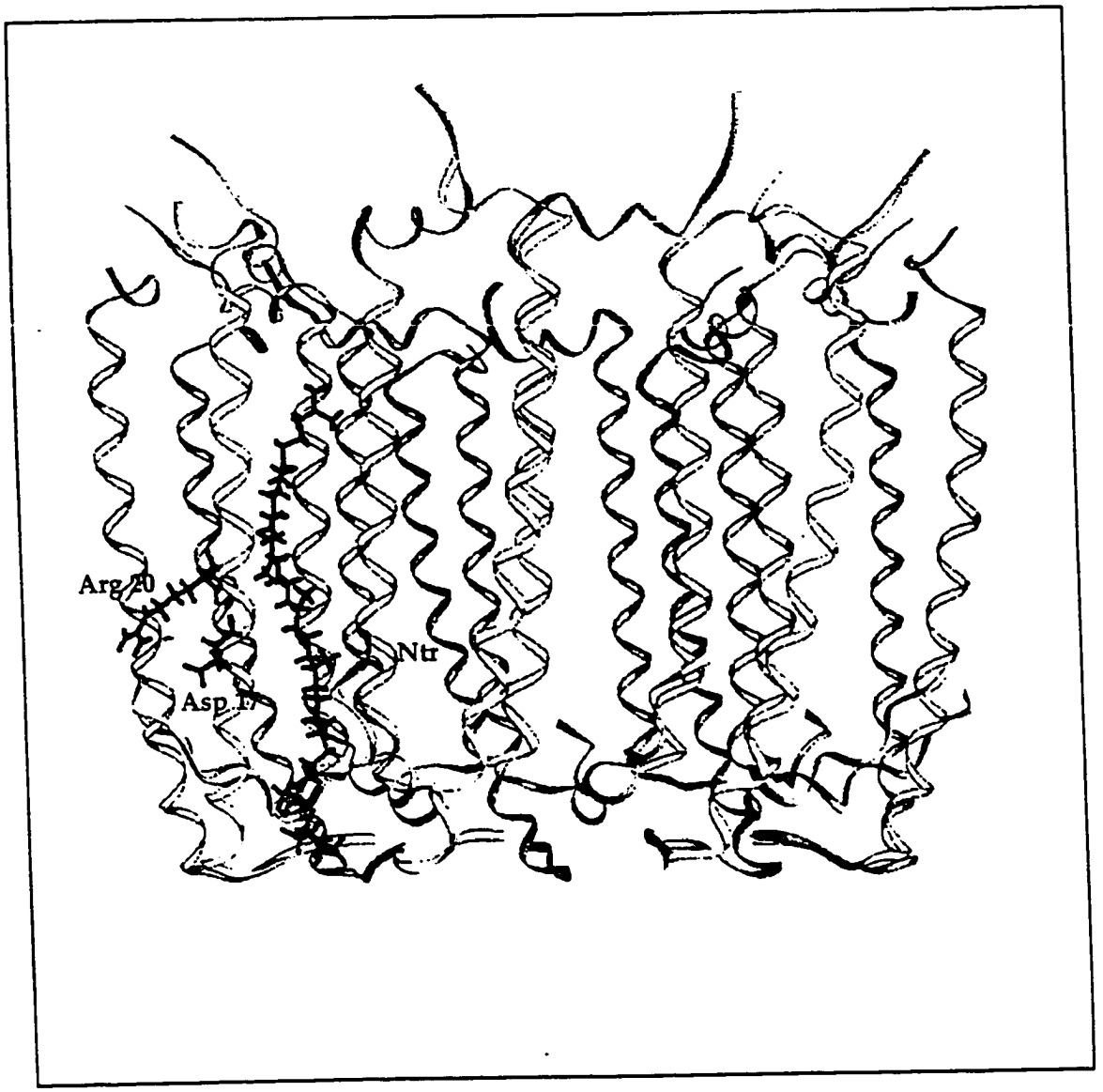
Table 2. Contributions to the difference between ground - and excited - state dipole moment vs. dielectric constant

As we can see from the Table 2 there is a good agreement between the value for $\Delta\mu$ in elliptical approximation and contributions $\Delta\mu_0$ and $\Delta\alpha E_0$ for the dielectric constant of 4 assigned to the protein's interior. The value of 6D for $\Delta\mu_0$ is in the range for asymmetric polyenes (6-12D [39,20,40,41]). In addition, James B. Bautista performed a molecular orbital calculation of $\Delta\mu_0$, utilizing a configuration interaction algorithm and single point calculation and received a value close to 6D (5.8D).

4.1.4 Specific residues

Although the electric field inside of a protein is a collective effect from all charged, dipolar and polarizable groups of the protein, there are residues with contributions more than 1.0×10^6 V/cm to the electric field due to the protein matrix.

For LH2, β Arg20, β Asp17 and 2 α N-terminals generate those local fields, where Arg20, Asp17 and one N-terminal belong to the same protomer unit as the corresponding carotenoid molecule, while the second N-terminal belongs to a nearby protomer (Figure 8). The largest polarizable effect is from Arg20-Asp17 dipole. Arg20 and Asp17 are located on the outer face of LH2.



Most important residues in LH2 from *Rps. acidophila*

Figure 8.

4.2. LHC from *Fremyella diplosiphon* (1CPC)

4.2.1 Introduction

In the antenna systems of cyanobacteria (blue green algae) (also in the antennae of eukaryotic red algae and cryptomonads), bilin chromophores are found and bound to the apoproteins. The pigment-protein complexes in these organisms are arranged in water-soluble complexes called phycobilisomes (hemidiscoidal or hemiellipsoidal aggregates forming regular arrays on the thylakoid membrane). Phycobilisomes are composed of phycobiliproteins and linker proteins. The phycobiliproteins are phycoerythrin (PE) and phycoerythrocyanin (PEC) which are located on the tips of the antenna substructure, phycocyanin (PC) in the middle of the antenna rod and an allophycocyanin (APC) core that is in close proximity to the charge separating photosynthetic reaction center. This spatial arrangement of the components makes the phycobilisome an ideal light transducer with an overal quantum efficiency of approximately 100% [52]. The phycobiliproteins extend the spectral range of the trapped sunlight and funnel energy to the reaction centers.

4.2.2 The structure of phycobiliproteins

All phycobiliproteins have very similar three-dimensional structures. The basic building block is composed of an (αβ)-monomer. Both α and β chains have a backbone conformation related to globins [53]. Eight helical segments (termed as X,Y,A,B,E,F,G and H) are the dominant secondary structure elements. In C-phycocyanins, the α-subunit carries one phycocyanobilin chromophore covalently bound to cysteine α-84, whereas the β-subunit carries two phycocyanobilins at cysteine β-84 and β-155, respectively. In PEC, the α-84 chromophore is a chemically different phycobiliviolin and in PE the chromophores are phycoerythro-bilins or phycourobilins. Three (αβ)-monomers are arranged around 3-fold symmetry axis and form an

$(\alpha\beta)_3$-trimer. Two trimers are associated face-to-face to form an $(\alpha\beta)_6$-hexamer, which is the functional unit in the native antenna [53].

4.2.3 Spectral characteristics of the phycobiliproteins

Some phycobiliprotein-containing organisms have the ability to alter the composition and the size of the phycobilisome antennae with light quality. This phenomenon, called complementary chromatin adaptation [54-55], enables the organism to adjust to the prevalent spectrum of the incident sunlight.

1CPC is a constutive phycocyanin (PC_c) that is always present, independent of the wavelength of the incident light in the phycobilisomes. It is the first phycocyanin of known three dimensional structure from a chromatically adapting organism.

The spectral characteristics of the phycobiliproteins and the nature of the energy transfer are dependent on the chemical nature of the chromophore and on the surrounding protein, especially polar and aromatic residues, which tune the spectrum of the chromophore. The neighboring amino acids keep the chromophores in a rigid conformation, which is essential for efficient energy transfer.

4.2.4 Configuration and stereochemistry of chromophores

The configurations of the α-84 and β-84 chromophores are Z, Z, Z(E for β-84) for carbon atoms C4, C10 and C15. The β-155 chromophores differs in configuration with Z, Z, Z. The stereochemistry of β-155 is different too: R, R, S for the chiral centers C2, C3 and C31, whereas chromophores α-84 and β-84 have R, R, R stereochemistry. The Z, Z, Z configuration of the β-155 chromophore is a consequence of the hexameric aggregation state of the protein whereby β-155 of subunit 1β comes in contact with helices Y and H of subunit 6α in the other trimer. Two hydrogen bonds stabilize the Z configuration of ring D: N24 β-155 to OD2 6α-147

(3.2 angstroms) and O19 β-155 to OE1 6α-33 (3.16 angstroms)[32]. The rigidification of ring D β-155 upon hexamer formation has a probably functional character. This may be a factor in explaining the increased ratio of visible to near ultraviolet absorbance and the decrease of fluorescence lifetimes [56].

In phycocyanins, the geometry of the chiral center at C31 β-155 seems to be independent of aggregation state and species. The difference from α-84 and β-84 might be an additional factor determining the spectral differences between the chromophores. Investigations by Arciero et al. [57-59] revealed that phycocyanobilins bind spontaneously to apophycocyanins. The stereochemistry at C31 may be therefore controlled by the different protein environments at the α-84, β-84, and β-155 attachment sites and/or by different enzymes catalyzing the chromophore attachment.

4.2.5 Energy transfer between chromophores

In phycobilisomes, energy is transferred among the chromophores. In the (αβ) monomer the separation and the geometry of the chromophores allows only weak couplings. If three (αβ) monomers are assembled in an $(\alpha\beta)_3$-trimer, the environment of the α-84 and β-84 chromophores changes drastically: their centers of mass come to lie at a distance of 20.6 angstroms and give rise to a strong interaction. Intra-hexamer energy transfer pathways are mainly found between α-84 chromophores, since the distance and orientation factors are favourable in this arrangement. Furthermore, the β-155 chromophores become more efficiently coupled. Assuming a distance of 60 angstroms between two distinct hexamers only the β-84 are potential candidates for efficient inter-hexamer energy transfer. It should be noted, that linker proteins, which are believed to be in the central cavity of the hexamer, play an important role in the energy transfer. This influence is observed as a red-shift of absorbance maximum when linkers are bound to the protein.

4.2.6 The local field and important residues

We calculated the average electrostatic field across the chromophores using 1CPC pdb file for C-phycocyanin from *Fremyella diplosiphon*. The pdb file is an asymmetric unit of the crystal cell consisting of two (αβ)-monomers related by a local dyad. It would be more appropriate to do the same calculation using three asymmetric units, since $(\alpha\beta)_6$-hexamer is the functional unit in the native antenna rod. In that case we could gain a better understanding of the energy transfer in terms of the electrostatic field and environment of the chromophores. For now, we will look only at one building block of the functional unit.

This LHC provides a good example illustrating of a possible correlation between the field across the chromophores and their properties. As can be seen from the Table 3, different spectral characteristics of β–155 and α,β-84 chromophores might be correlated with the average electrostatic field across the chromophores. Using program GRASP we found that the electric field at carbon atoms of carotenoids is not aligned along the axis of the carotenoids. The backbone contribution to the field is significantly larger for β–155 than for α,β-84 chromophores. We suggest that this contributes to the different stereochemistry at C31 for β-155 compared to α,β-84. The average electrostatic field across the chromophores is dominated by charged residues (Figure 9-14) (mainly by ASP and ARG residues for α,β-84 chromophores, while for β–155, in addition, by GLU residues). Polar and aromatic residues do fine tuning of the field at the chromophores, since their contributions are significantly smaller. Table 4 contains residues that contribute to the average electrostatic field more than 10^6 V/cm. We notice ARG-ASP dipole motif among these residues like in LH2 from *Rps. Acidophila*. Chromophores are surrounded by these residues, and that indicates, besides their polarization role, a role in keeping the chromophores in a rigid conformation, which is essential for efficient energy transfer.

carotenoid	cyca	cycb84	cycb155	cyck84	cycl84	cycl155
polar	1.16	0.78	0.37	0.97	0.75	0.62
aromatic	0.62	0.54	0.07	0.86	0.51	0.1
charged	5.93	4.84	5.22	5.35	4.7	5.54
backbone	1.42	0.57	2.58	1.74	0.43	2.7
average field	8.71	4.67	4.83	7.0	5.0	7.0

Table 3. Contributions in 10^6 V/cm to the average electrostatic field across chromophores (ε_{in}=4; ε_{out}=80)

carotenoid	important residues
cyc a	cys a 84,asn a 119,asp a 123 tyr a 129,lys a 83, ARG A 86 ASP A 87
cyc b 84	cys b 84,arg b 79,arg b 80 asn b 72, ARG B 86, ASP B 87
cyc b 155	gln k 33, asn b 35, arg k 30 asp b 33, glu a 23, cys b 155 asp k 147, lys b 36, glu k 161 arg b 43, ARG B 37, ASP B 39
cyc k 84	arg k 93, cys k 84, tyr k 129 lys k 83,ARG K 86, ASP K 87 asn k 119
cyc l 84	cys l 84, thr l 124, arg l 80 asn l 72, ARG L 86, ASP L 87 arg l 79, arg l 93
cyc l 155	thr l 151, gln a 33, asn l 35 arg a 30, asp l 33, cys l 155 glu k 23, lys l 36, glu a 161 arg l 43, asp a 147 ARG L 37, ASP L 39

Table 4. Residues with the contributions to the electrostatic field more than 10^6 V/cm for carotenoids in 1CPC

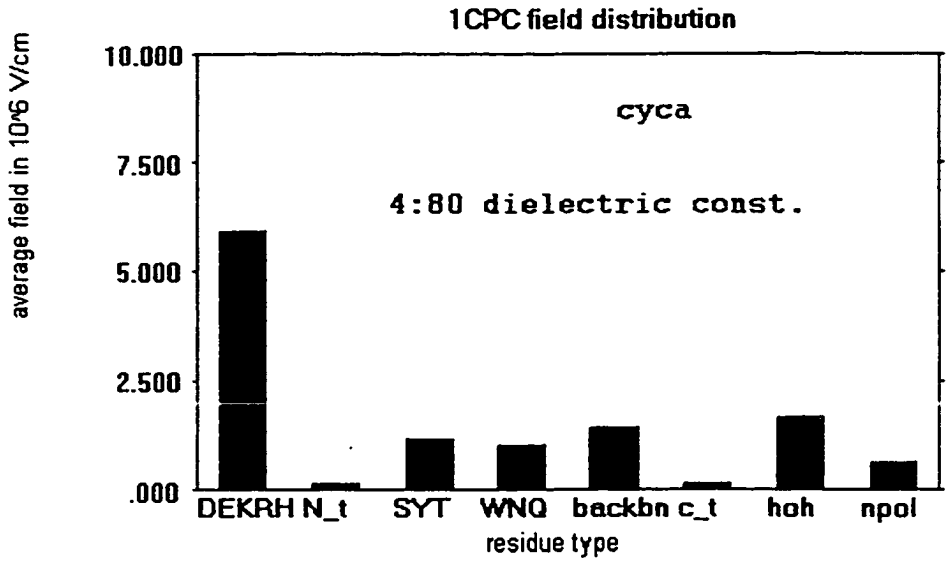


Fig. 9

DEKRH=asp,glu,lys,arg,his; **N_t**=Nterm; **SYT**=ser,tyr,thr; **WNQ**=trp,asn,gln,cys; **backbn**=backbone; **c_t**=Cterm; **hoh**=water; **npol**=nonpolar residues

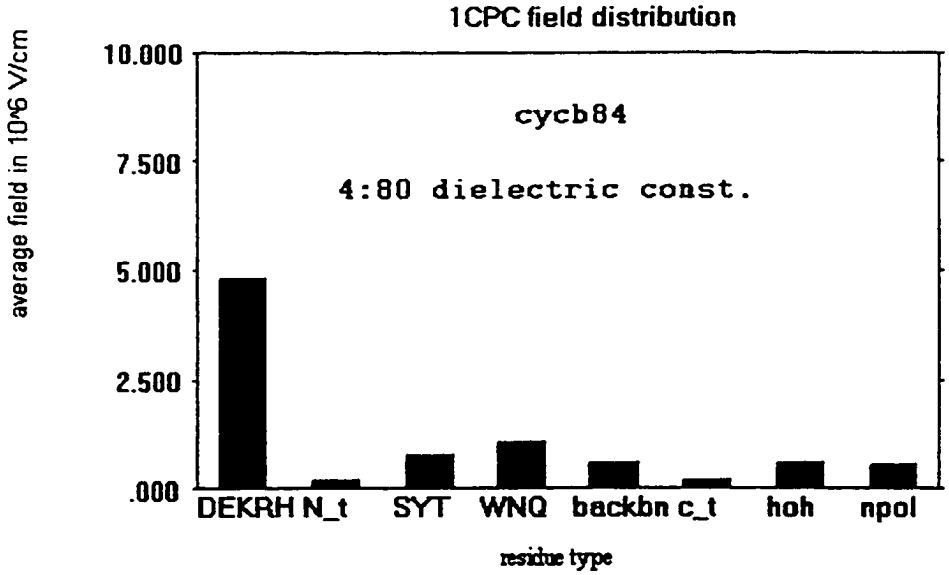


Fig. 10

DEKRH=asp,glu,lys,arg,his; **N_t**=Nterm; **SYT**=ser,tyr,thr; **WNQ**=trp,asn,gln,cys; **backbn**=backbone; **c_t**=Cterm; **hoh**=water; **npol**=nonpolar residues

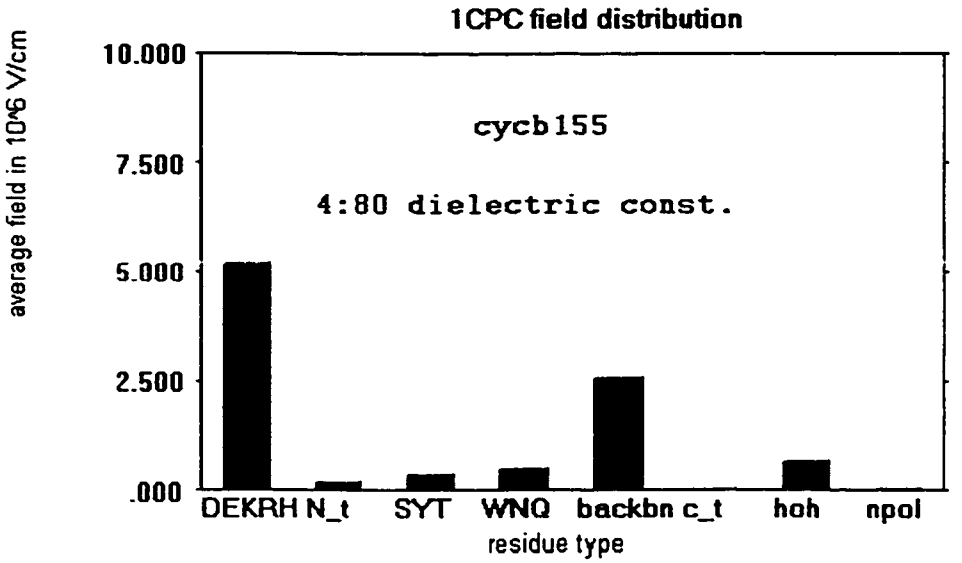


Fig. 11

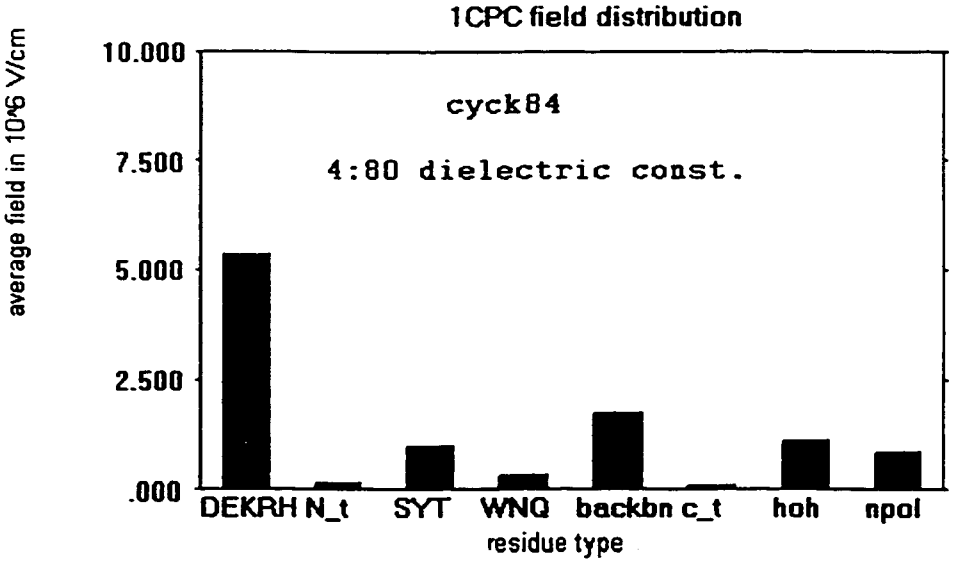


Fig. 12

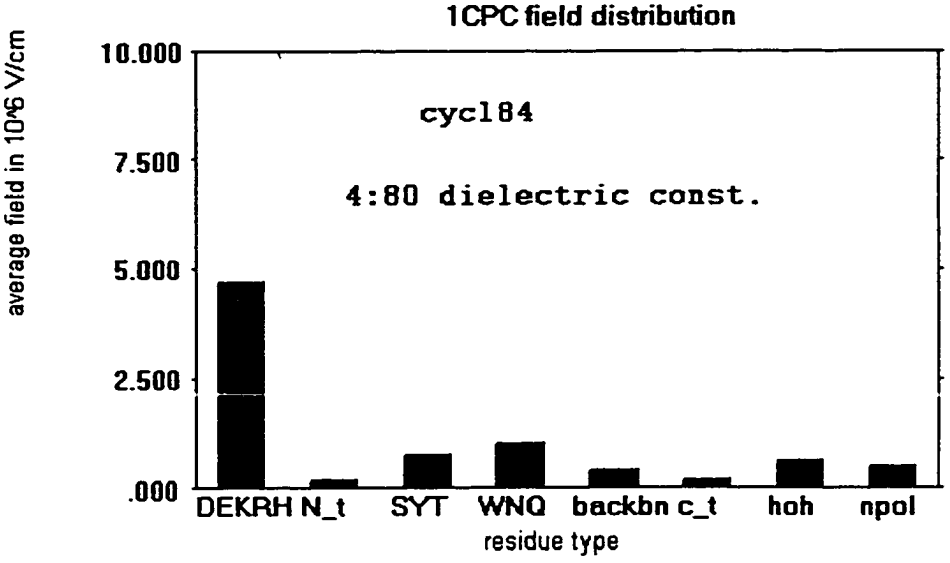


Fig. 13

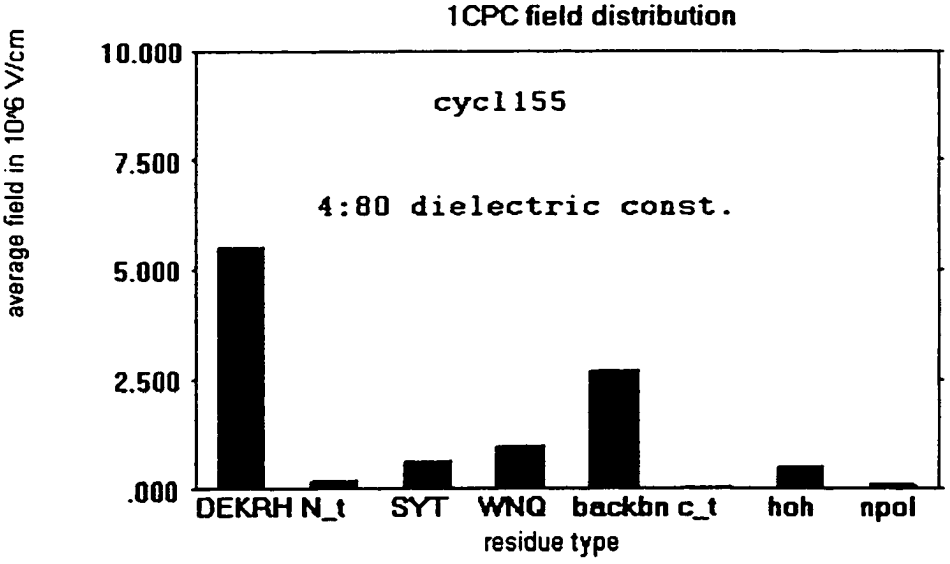


Fig. 14

DEKRH=asp,glu,lys,arg,his; **N_t**=Nterm; **SYT**=ser,tyr,thr; **WNQ**=trp,asn,gln,cys; **backbn**=backbone; **c_t**=Cterm; **hoh**=water; **npol**=nonpolar residues

(standard one-letter notation used for labeling; Appendix C)

4.3. LH2 from *Mastigocladus laminosus* (1B33)

4.3.1 The local field and imortant residues

LH2 from *Mastigocladus laminosus* [31] consists of two identical subunits (A and B), each containing one CYC carotenoid molecule (phycocyanobilin) (Figure 15).

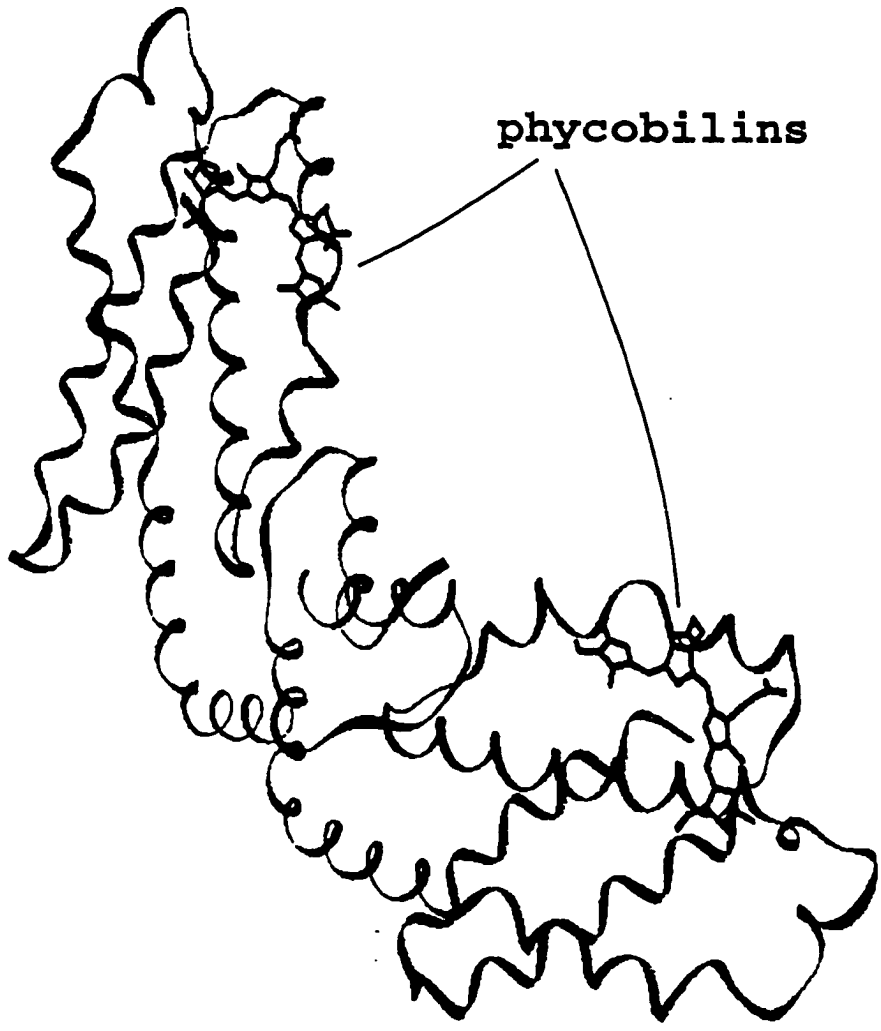


Figure 15.

The carotenoid type is the same as in LHC from *Fremyella diplosiphon*. The average electrostatic field across CYC A molecule is 4.0 x 10^6 V/cm and the corresponding field across CYC B molecule is 7.2 x 10^6 V/cm (Table 2) with a dielectric constant of 4 inside the protein. There are residues which contribute an average electrostatic field across the carotenoid molecule of more than 10^6 V/cm. For CYC A important residues are: CYS A 84, ARG A 93, ASN A 72, ARG A 86, and ASP A 87. For CYC B the following residues are important: ASP B 89, CYS B 84, THR B 129, ARG B 80, ASN B 72, ARG B 79, and ARG B 86. An ARG-ASP dipole appears among above groups of residues (ARG A 86-ASP A 87 for CYC A and ARG B 86-ASP B 89 for CYC B). This is the same motif we found in LH2 from *Rps. Acidophila*, and LHC from *Fremyella diplosiphon.* These residues surround the carotenoid molecule inside a hydrophobic/hydrophilic core. There is a good match between the residues contributing to the field and the residues listed by the program LIGPLOT [51] which automatically generates schematic diagrams of interactions in proteins. For CYC A LIGPLOT found the following residues that make a hydrogen bond with the carotenoid molecule: ASN A 72, CYS A 84, ASP A 87, and ARG A 86. The following residues make a hydrogen bond with CYC B according to the program LIGPLOT: ASP B 87, CYS B 84, ASN 72 B, ARG B 79 and ARG B 86. For both molecules CYS 84, ASN 72, ASP 87, and ARG 86 are present as hydrogen bond partners indicating their probable functional importance in inducing dipole moments in the corresponding carotenoid molecule. We found ARG B 79 as the most responsible residue for generating a field in excess of 3.0 x 10^6 V/cm across the CYC B molecule. The evidence that CYC B molecule experiences a larger polarizing electrostatic field than the CYC A molecule is probably functionally important. This could enhance an energy transfer between the two carotenoid molecules.

We found that different values of the dielectric constant inside the protein (2, 4, 80) preserves the larger electrostatic field across CYC B molecule with respect the field across CYC

A. The magnitude of the average electrostatic field across the molecule CYC B is nearly twice as large as the magnitude of the average field across CYC A molecule. This is also conserved in a uniform dielectric (the dielectric constant inside the protein is equal to the dielectric constant outside the protein) for different values of the dielectric constant (1, 4, 80). An analysis using program GRASP [48] revealed no preferred direction for the electrostatic field as a function of carbon atoms along the axis of the carotenoid molecules. The contributions to the local field depending on the type of the residue are shown on Figures 16 and 17. The largest contributions are for the charged residues.

carotenoid	average field	important residue
CYC A	4	CYS A 84,ARG A 93,ASN A 72,ARG A 86, ASP A 87
CYC B	7.2	ASP B 89, CYS B 84, THR B 129, ARG B 80, ASN B 72, ARG B 79, ARG B 86

Table 5. Residues with the contribution to the electrostatic field more than 10^6 V/cm for carotenoids in 1B33(the average field is in the same units)

DEKRH=asp,glu,lys,arg,his; **N_t**=Nterm; **SYT**=ser,tyr,thr; **WNQ**=trp,asn,gln,cys; **backbn**=backbone; **c_t**=Cterm; **hoh**=water; **npol**=nonpolar residues

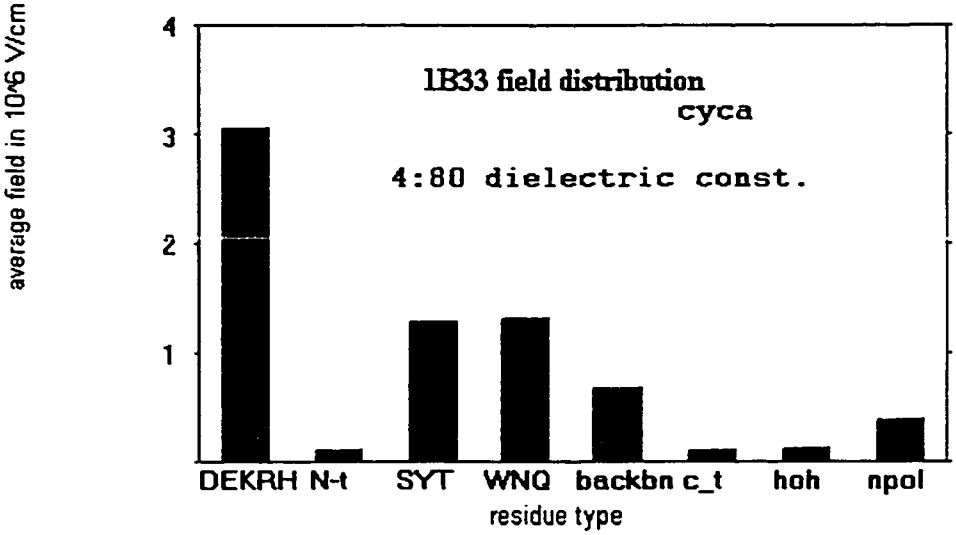


Fig. 16

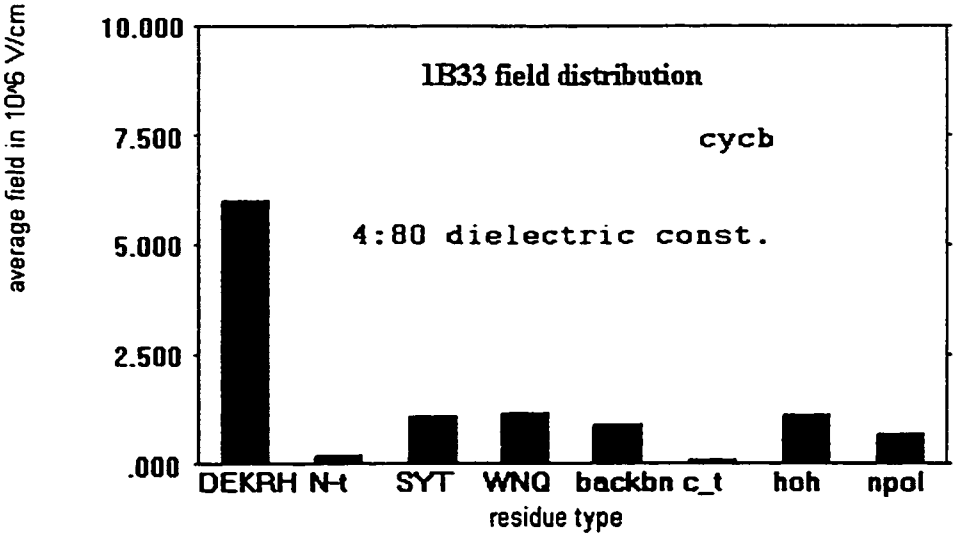


Fig. 17

4.5. LHC from *Prosthecochloris aestuari* (3BCL)

4.5.1 Introduction

The bacteriochlorophyll A protein is thought to transmit excitation energy to a special pair of chlorophyll molecules in the reaction center where the first step in energy transduction (charge separation) occurs. LHC consists of three identical subunits [60] related by 120 degrees rotations about a 3-fold axis of symmetry. In each subunit the polypeptide backbone forms a large, twisted β - sheet of 16 strands that forms the part of the protein molecule exposed to solvent and encloses a central core of seven bacteriochlorophyll (BChl) A molecules. The trimeric structure seen in the crystals occurs also in vivo and is not an artifact of crystallization.

4.5.2 Ligands

The ligands to the seven BChl magnesiums are, respectively, five histidines, a carbonyl oxygen from the polypeptide backbone of the protein, and a bound water molecule. One of the striking features is the distribution of the magnesium positions. In two cases (BChls 3 and 7) the metal ions are below the ring plane whereas in the other five cases they are above. There are no intermediate states. The displacement of each magnesium is always towards the side of the ring plane from which the metal is liganded to the protein. Therefore the displacement direction is not an intrinsic property of the BChl molecule itself. The protein environment determines the direction of the out-of-plane magnesium displacement.

4.5.3 The lipid interactions

The interactions between the phytyl chains of the BChl a molecules and the amino acide side chains provide a good model for lipid-protein interactions. Also, the packing of the BChl A molecules gives an insight into other protein-chlorophyll interactions.

Lipid conformations are well-defined but irregular. The phytyl chains of the BChls occupy well-defined positions in space with thermal factors comparable to the surrounding protein. In some cases the hydrocarbon chains are quite extended but they are not in idealized all-trans conformations. In this respect the BChl phytols differ markedly from crystal structures of isolated lipids in which very regular extended conformations are the rule [61-63]. Lipid tails prefer parallel interactions. In a membrane bilayer, the amphipathic nature of membrane lipids results in the well-known bilayer structure with polar head-groups at the surface and non-polar tails below the surface, interacting with each other. Because of these constraints the lipid tails are forced to lie more or less parallel to each other. In the BChl protein, where the hydrocarbon tails do interact they clearly prefer to do so in a parallel manner. The individual phytyl chains have unique conformations and often make sharp bends at substituted positions. Although there are extensive interactions between the phytyl chains of neighboring BChls, these interactions do not dominate over interactions between the phytols and the surrounding protein.

4.5.4 Difficulties in determining an amino acid sequence from crystallographic data

There were several difficulties in determining an amino acid sequence from crystallographic data. Several pairs of amino acids (Val, Thr), (Glu, Gln), and (Asp, Asn) are chemically different but isostructural so these pairs are impossible to differentiate in the electron density map. In addition, some amino acid side-chains, especially on the surface of the protein, are mobile and hard to identify. In extreme cases the backbone atoms also undergo large "termal" displacements and cannot be seen. The authors [60] adopted the conservative policy of deleting atoms from the model rather than having them present in doubtful positions. These difficulties clearly influence our knowledge of the protein's charge distribution (Asp(-1) <-> Asn(0)),

(Glu(-1) <-> Gln(0)) and set an important limitation on the calculation of the average electrostatic field across chlorophylls.

4.5.5. The local field across BChl's and important residues

Our calculation revealed an increasing trend in average electrostatic field across chlorophylls 1-7 (Table 6). The trend is probably related to an enhancement of the light-capturing ability of bacteriochlorophylls. Due to that, the intra-molecular energy transfer between BChls is quite possible. Since *Prosthecochloris aestuari* lives at a depth of about 10 meters in lakes its life really depends on every single photon. In other words, the bacteria has to use whatever light is available. Lack of carotenoids in LHC of *Prosthecochloris aestuari* might have driven the evolution of LHC in the direction of extending the spectral range of light harvested by BChls. Extending of the spectral range might result from changes in the wave functions of those electrons that are involved in the optical transitions. Two possible causes are the interactions of these electrons with neighboring Bchl molecules and distortions due to binding of the BChls to proteins and lipids in the cell membrane. In this light we need to view all interactions between the phytyl chains of neighboring BChls and the packing of the bacteriochlorophyll A molecules.

The amino acids that contribute to the average electrostatic field across BChls of more then 10^6 V/cm are liganding histidines (Table 7). Table 7 also shows that for each bacteriochlorophyl the average electrostatic field is influenced, besides by the corresponding ligand, by ligands of other BChls. In addition, analysis using program GRASP does not suggest any polarization effect of the electrostatic field along the Q_x and Q_y Bchl dipoles. Interestingly, Thr 219 and Arg 230 contribute more than 10^6 V/cm to the average electrostatic field across BChl 3 and BChl7 respectively. These residues might play a role in the displacement of Mg ion below the ring plane of the corresponding BChls. Figure 18 shows important contributions of

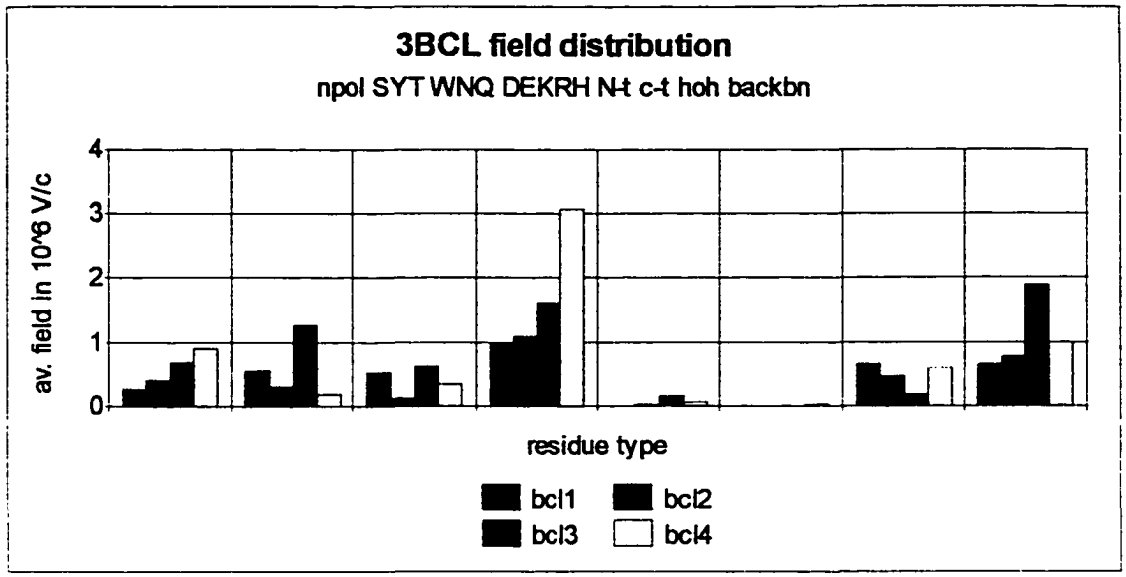
charged residues and the backbone. A better resolution crystal structure would give more details about the electrostatic features of this LHC.

Bchl	bcl1	bcl2	bcl3	bcl4	bcl5	bcl6	bcl7
polar	0.54	0.3	1.26	0.18	0.53	0.72	0.33
charged	1	1.1	1.6	3.06	4.4	3.49	2.24
backbone	0.64	0.77	1.89	0.98	1.68	0.71	0.92
av.field	1.2	1.61	2.26	3.41	3.53	3.78	4.59

Table 6. Contributions to the average electrostatic field across Bchls (in 10^6 V/cm; 4:80 dielectric constants in and out)

bacteriochlorophyll	important residues
bcl1	his 219,140,105
bcl2	his 290,219,289,140,105
bcl3	thr 293
	his 219,140,282,289,290
bcl4	his 219,289,230,352,290,282
bcl5	his 290,219,140,289,282 arg 230,90
bcl6	his 290,105,282,289,140,219 gln 192,138 arg 230,90
bcl7	his 105,140,282,219,290,289 arg 230

Table 7. Residues with the contribution to the electrostatic field more than 10^6 V/cm for Bchlls in 3BCL



residue types (left to right):

DEKRH=asp,glu,lys,arg,his; **N_t**=Nterm; **SYT**=ser,tyr,thr; **WNQ**=trp,asn,gln,cys;

backbn=backbone; **c_t**=Cterm; **hoh**=water; **npol**=nonpolar residues

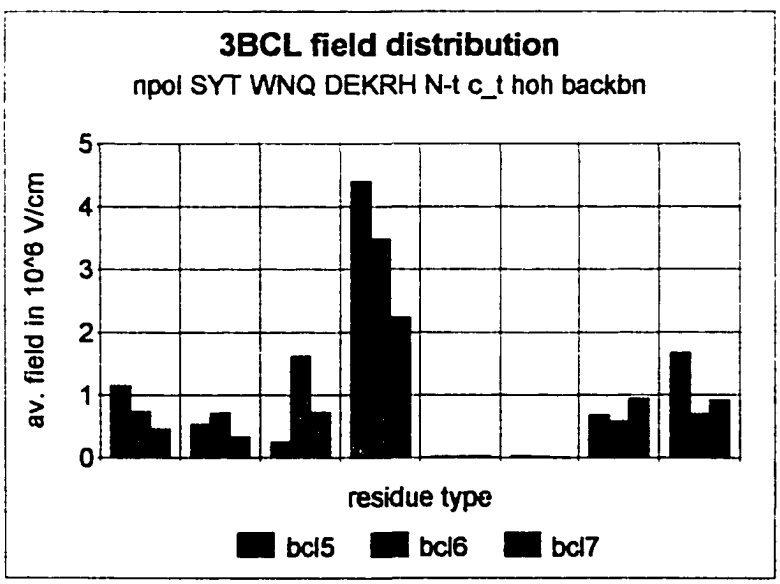


Fig 18. 3BCL field distribution as a function of a residue type

(ε_{in}=4; ε_{out}=80)

4.6. LHC from *Amphidinium carterae* (1PPR)

4.6.1. Introduction

Photosynthetic dinoflagellates, which make up much of the sea plankton and are the cause of red tides, use about equal amounts of carotenoids and chlorophylls for the efficient harvesting of light. Most dinoflagellates have peridinin as their predominant carotenoid, enabling them to capture solar energy in the blue-green range (470 to 550 nm), which is inaccessible by chlorophyll alone. In addition to a membrane-bound LHC, dinoflagellates have developed a soluble antenna with a high carotenoid:chlorophyll ratio, peridinin-chlorophyll-protein (1PPR), that has no sequence similarity with other known proteins.

4.6.2 The functional structure of 1PPR

The functional structure of this LHC is a trimer in which each polypeptide contains an unusual jellyroll fold of the α-helical amino and carboxyl-terminal domains [36] . These domains constitute a scaffold with pseudo-twofold symmetry surrounding a hydrophobic cavity filled by two lipid, eight peridinin and two chlorophyll a molecules. Each domain binds a cluster of one chlorophyll A and four peridinin molecules. Within each cluster, the efficiency of singlet energy transfer from peridinin to chlorophyll is close to unity. The lipids are an integral part of the structure. Their existence in 1PPR was unexpected and points to the complexity of the holoprotein assembly. The lipids were identified as digalactosyl diacyl glycerol (DGDG). DGDG binds tightly to the chlorophylls before complex formation.

4.6.3 The structural basis for efficient energy transfer

The structural basis for efficient excitonic energy transfer from peridinin to chlorophyll is found in the clustering of peridinins around the chlorophylls at Van der Waals distances. Circular

dichroism studies of 1PPR suggested two chlorophyll-peridinin clusters in which each chlorophyll is arranged between two pairs of mutually orthogonal peridinins. The crystal structure of 1PPR confirms the existence of a NH_2-terminal and a COOH-terminal pigment clusters, related by the same local twofold symmetry as is the apoprotein. Each cluster has two pairs of peridinins surrounding a chlorophyll. Within each pair, closest distances of the polyene chains are less than 4 angstroms. All peridinins are in an extended all-trans conformation and show small distortions from their average geometry, which mainly result from the interaction with the protein environment. The conjugated regions of all peridinins are in Van der Waals contact (3.3 to 3.8 angstroms) with the tetrapyrole ring of chlorophyll, allowing efficient excitonic energy transfer from peridinin to chlorophyll. The chlorophylls are completely buried in a hydrophobic environment. The closest protein-chlorophyll contacts occur through a stacking at Van der Waals distance of the imidiazole rings of two conserved His residues, His 66 and His 229, onto the tetrapyrole C ring of the NH_2 and COOH terminal chlorophylls, respectively. A water molecule, which is hydrogen bonded on one side to these residues, provides the fifth coordination site (distance 2.0 angstroms) of the center Mg atoms.

4.6.4 Energy transfer between chlorophylls

Unlike the situation in LH2 from *Rps. Acidophila* where excitonic energy transfer between adjacent chlorophylls takes place, the geometry in 1PPR allows only Forster energy transfer between chlorophylls of each monomer (distance 17.0 angstroms). Distances between chlorophylls of different subunits within a trimer are in the 40 to 55 angstroms range. The COOH-terminal chlorophylls appear to be oriented most suitably for Forster exchange of excitation energy between the subunits of a trimer as their Q_y transition moments are approximately parallel to the trimer axis (inclination 25 degrees) and their distance (44 angstroms) is well within the range of dipole-dipole interactions.

4.6.5 The local field and important residues

The unique structure of 1PPR (jellyroll fold) is reflected in the field generation by the protein matrix. Our calculation indicates that the local field is a collective effect from all charged, dipolar and polarizable groups of the protein with a few residues with a contribution more than 10^6 V/cm (Table 8). These residues probably tune and control an efficient excitonic energy transfer from each peridinin to chlorophyll. No direction along the carotenoids is found for the field. Since the conjugated regions of all peridinins are in Van der Waals contact with the tetrapyrole ring of chlorophyll these residues may also facillitate Dexter type of energy transfer. Dexter type of energy transfer is relevant when the pigments molecules are very close (less than 5 angstroms) so their orbitals overlap and electron exchange is enhanced. Most probably, both mechanisms are contributing to an energy transfer between peridinins and chlorophylls. This is an interesting situation where the protein matrix plays a double role: it polarizes peridinins influencing the excitonic transfer and it shapes the orbital overlap allowing electron exchange between peridinins and chlorophylls.

We assumed that Dexter energy transfer for soluble LHCs is more appropriate then Forster type of energy transfer, since Dexter energy transfer does not depend on the dielectric constant. In 1PPR antenna complex the folding of the polypeptide chain of 1PPR is such that chlorophylls are completely burried in a hydrophobic environment. In addition the lipids are an integral part of the structure bound tightly to the chlorophylls and might additionally lower the inner dielectric constant. All these allow an efficient Forster energy transfer between chlorophylls, since water molecules of the outside dielectric (bulk water) are completely excluded from the protein core.

carotenoid	average field	important residues	
pid m 611	0.84	Lys m 134	Ser m 203
pid m 612,613	2.31,1.94	-	
pid m 614	2.63	Glu m 101	Gly m 78 Thr m 80
pid m 621	2.68	-	
pid m 622	2.76	Asn m 38	
pid m 623	1.82	-	
pid m 624	1.74	Thr m 242	Gly m 240

Table 8.Residues with the contribution to the electrostatic field more than 10^6 V/cm for carotenoids of m subunit in 1CPC

(the average field is in the same units; (ε_{in}=4; ε_{out}=80)

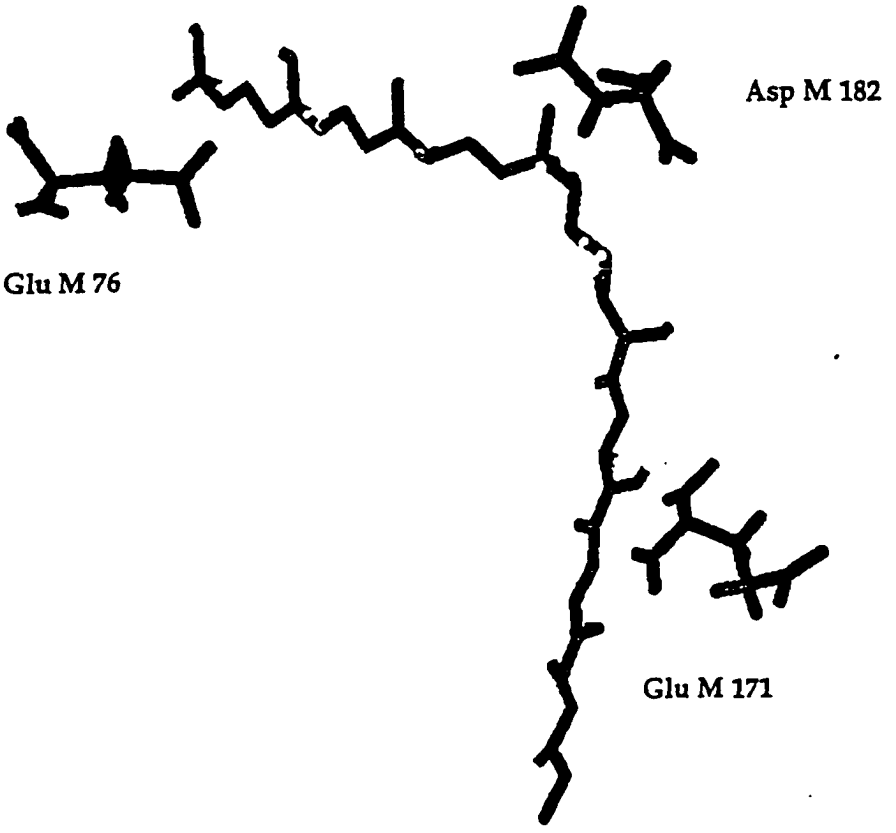
4.7. RC's from *Rps. viridis* and *Rb. sphaeroides*

4.7.1 Local fields

The carotenoid molecule (1,2-dihydroneurosporene) adopts a 13'-S-cis conformation with an additional out-of-plane twist at C-14'-C-15' bond in RC from *Rps. viridis*. This geometry produces a large variability of the electric field across the molecule (Figure 7) with the largest electric field values around 13', 14', 15' carbon atoms. The average electric field across 1,2-dihydroneurosporene molecule for the dielectric constant of 4 inside of RC and 80 outside was 2.0 x 10^6 V/cm. In *Rb. sphaeroides* reaction center 1,2-dihydroneurosporene adopts an extended (linear) conformation. The average field across the carotenoid molecule in *Rb. sphaeroides* is 4.9 x 10^6 V/cm (4:80 dielec. constants).

4.7.2. Important residues

For both RC's the direction of along the carotenoid axis is not a preferable direction for the local field.For 1,2-dihydroneurosporene molecule in RC from *Rps. viridis* GluM76, GluM171 and AspM182 generate contributions more than 1.0 x 10^6 V/cm to the local field (Figure 19). For 1,2-dihydroneurosporene in RC from *Rb. sphaeroides* contributions to the field of more than 10^6 V/cm are generated by histidines: His L 168 and His M 182.



* hydrophobic residues are surrounding the caroteroid (not shown)

Most important residues in RC from *Rps. viridis*

Figure 19.

4.7.3. Are the electrostatic field in RC from *Rps. viridis* and hydrophobic interactions with nearby residues strong enough to leave the carotenoid molecule essentially out of a function?

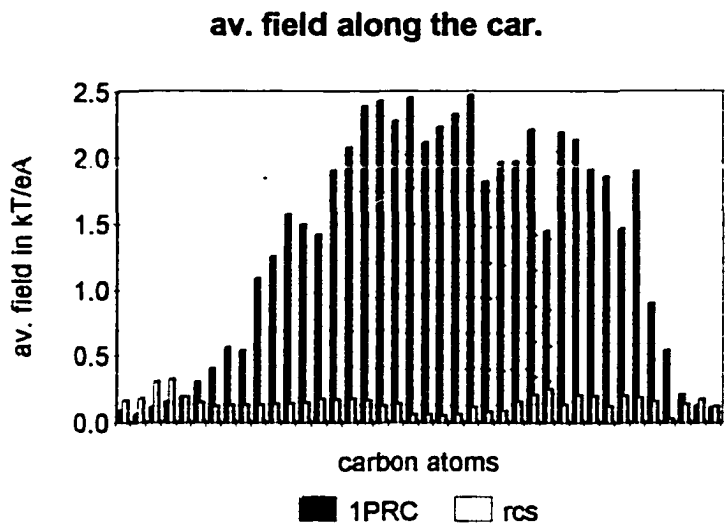
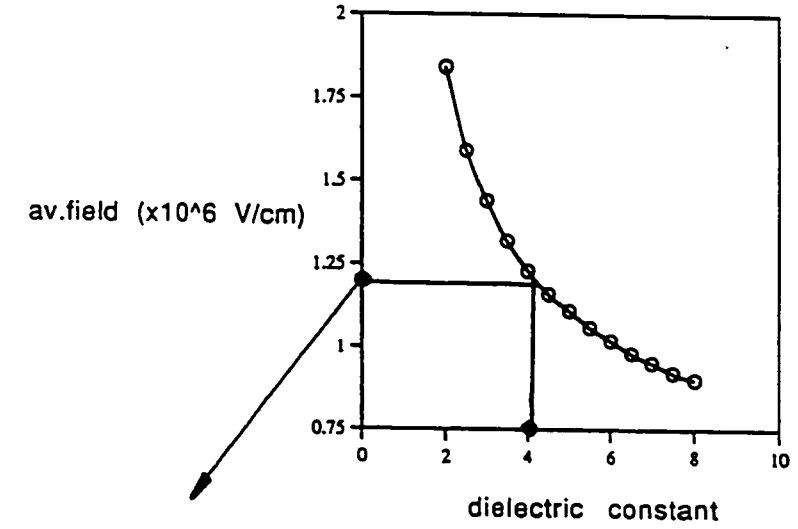


Fig 20. Electrostatic field across carbon atoms of 1,2-dihydroneurosporene in reaction centers of *Rps. virdis* (1PRC) and of *Rb. sphaeroides* (rcs) (the magnitude of the electrostatic field is in (kT/e(angstrom)))

It is clear that 1,2-dihydroneurosporene experiences much stronger electrostatic field along carbon atoms in *Rps. viridis* reaction center than in *Rb. sphaeroides* reaction center (Fig. 20). This gives more confidence to a possibility that the local field and hydrophobic interactions with nearby residues could "bend" 1,2-dihydroneurosporene molecule to the conformation which the molecule has in the reaction center from *Rps. viridis* (Fig. 19).

4.8 DelPhi can provide an reasonable input for the carotenoid spectroscopy

Strong relation between the local electrostatic electric field and the carotenoid band shift would provide an interesting tool for determining the dielectric constant of the protein's interior. Delphi can produce dependence of the local field on the dielectric constant (Figure 21).This can be compared with eq (6) with the experimentally determined band shift value. Interestingly, our ball park agreement for the local field in LH2 is with the dielectric constant of 4 inside of the protein. This makes a sense, since experimentally determined $\Delta\mu_{app}$ is performed in an applied field of ~10^5 V/cm. There is, probably, some movement of the helices of LH2 in a response to that field. It was shown that dielectric response of helices in proteins leads to the dielectric constant of 4 for the protein's interior [50].



Average electrostatic field in LH2 from *Rps. acidophila* as a function of dielectric constant

Figure 21.

4.9. How the local field is generated in a light harvesting complex or reaction center by the protein matrix

The field distributions (Figure 9-18) shed some light on general principles of the generation of the local field. There is an important contribution to the electrostatic field from the backbone dipoles. In other words, the minimum value of the field is determined by the protein's three dimensional structure. This indicates an important relation between the field and the protein's function. The rest of the field (the dominant part) is provided by the charged residues and a fine tuning is done, mainly, by hydroxyls and water molecules.

		LH			RC		
membrane			soluble				
1KZU		3BCL	1B33	1CPC	1PPR	1PRC	rcs
nonpolar					Gly		
polar		Thr,Gln	Thr,Asn, Cys	Cys,Thr Gln,Asn	Ser ,Thr , Asn		
charged	Arg,Asp	His,Arg	Arg,Asp	Arg,Glu, Asp,Lys	Lys,Glu	Glu, Asp	His
FUNCTION		light TYPE	harvesting OF	ENERGY	TRANS.	photo.	protective
Forster	+	+	+	+	+	-	+
Dexter	+	+			+	-	

Table 9. Summary data for light harvesting complexes and reaction centers showing the type of residues most important for generating the local field that polarizes carotenoids (BChls) and most probable type of the energy transfer

V. DISCUSSION

5.1. How the local field is generated

Table 9 suggests some regularities in the design of the electrostatic field by LHC or RC. For a membrane-bound LHC, charged residues determine the field, while in soluble LHCs polar residues may also play an important role. Histidines are most important for polarizing the bacteriochlorophyls in a LHC, since they are usually ligands of BChls. For carotenoids in a LHC, histidines are not of much importance. Interestingly, in RC of *Rb. sphaeroides* histidines polarize the 1,2-dihydroneurosporene molecule. This carotenoid has a photoprotective function and it seems that nature here implements the same polarization procedure for bacteriochlorophylls as in a LHC. Light-harvesting is the main function of BChls in a LHC. But, histidines are not of importance for 1,2-dihydroneurosporene molecule in RC of *Rps. viridis*. In this case the carotenoid does not do any photoprotective function and the carotenoid is polarized by charged residues in analogy to the membrane-bound LHC. Interestingly, those charged residues facilitate an efficient electron transfer in RC of *Rps. viridis*. For soluble LHCs Forster type of the energy transfer is often used between chromophores. This is probably related to an exclusion of water from the protein core (like in 1PPR) and burring chromophores in a hydrophobic environment. Membrane-bound LHCs use all possible ways of the energy transfer (like LH2 from *Rps. acidophila*).

5.2. Constraints on the local field

In case of carotenoids, our results show that most of the polarizing field is provided by charged residues. These amino acids tend to have relatively low mobility as estimated from the refined crystallographic termal factors. There are different ways that a RC or LHC can use charged residues to achieve this goal.

A dipole motif is often found as a means to generate a significant polarizing field. The dipole consists of the charged residue mostly ionized in a physiological pH range (like Arg-Asp in 1KZU, 1CPC, 1B33). Alternately clusters of charged residues around the carotenoid molecule can provide the needed field too (for example in 1CPC). The interactions among amino acids inside of these clusters are crucial to control the protein's response to the pH changes and to maintain the field. Scattered isolated charged residues around the carotenoid molecule provide another way (like in 1PRC). In this case the long-range electrostatic interactions play the important role in the maintenance of the field.

How the protein will generate the polarizing field depends also on the type of the chromophore and its function. For example, in the case of the bacteriochlorophyls in a LHC, most of the field is provided by liganding histidines (like in 3BCL) while histidines are not of much importance for carotenoids in a LHC.

The main function of carotenoids in a LHC is the light harvesting function. They enhance the light - capturing ability of bacteriochlorophylls or chlorophylls in LHCs of photosynthetic organisms by collecting light at wavelengths in the spectral region where bacteriochlorophylls or chlorophylls are inefficient absorbers. The singlet excitation energy of the carotenoids is then transferred to the ground state of the bacteriochlorophylls or chlorophylls after which bacteriochlorophylls or chlorophylls become excited (like in LH2 from *Rps. Acidophila*). In other words, the energy transfer is from the short to long wavelength absorbing chromophore and it strongly depends on the magnitudes and relative orientations of the transition dipoles between chromophores (Appendix A). This has direct impact on the magnitude and direction of the polarizing field which provides the difference between ground - and excited - state dipole moments of the carotenoid molecule.

VI. CONCLUSIONS

Carotenoids were the focus in the second part of the dissertation. We applied the program PROTEUS to place hydrogens in several resolved three-dimensional crystal structures of LHCs and RCs. Using program DelPhi we calculated the local electrostatic field across carotenoids (in one structure across bacteriochlorophylls) generated by the protein's charges. In each structure we identified amino acids responsible for the field. Much of the field is generated by the charged residues. There are different ways that a RC or LHC uses charged residues. A nearby dipole consisting of the charged residues which are ionized in the physiological pH range (like Arg-Asp), is often used. Clusters of charged residues or scattered isolated charged residues around the carotenoid molecule also contribute. The type of the chromophore (carotenoid or bacteriochlorophyl) and its function plays a significant role in the way the polarizing field is generated. The polarizable field is not necessarily along the carotenoid molecule principal axis. In the case of the bacteriochlorophyls, most of the field is provided by liganding histidines. Histidines are not of any importance for carotenoids in LHCs. For soluble LHCs the contribution of polar residues to the field cannot be neglected. Our calculations indicate an important relation between the field and the protein's overall structure : the minimum value of the field arises from the backbone dipoles. In contrast to the charged residues, hydroxyl and water dipoles are responsible for a fine tuning of the field in a response to the protein's environment.

APPENDIX A (Energy transfer mechanisms)

In general, the migration of energy between a donor and an acceptor is discussed in terms of the Forster and / or Dexter energy transfer mechanisms.

The Forster mechanism [64,65] of energy transfer is described by the electrostatic interaction between the donor and acceptor electronic transition dipoles, since the dipole-dipole interaction is the lowest-order nonvanishing term in the multipole expansion of the electrostatic interaction between the donor and acceptor if both of them are uncharged.The transfer rate, k, and efficiency, Φ, can be expressed as:

$$k = \left(\frac{R_0}{R}\right)^6 \cdot \frac{1}{\gamma} \tag{A1}$$

$$\Phi = \left(\frac{R_0}{R}\right)^6 \tag{A2}$$

$$R_0^6 = C\kappa^2\eta^{-4}\int f(\nu)\varepsilon(\nu)\nu^{-4}d\nu \tag{A3}$$

where R is the distance between the dipoles, R_0 is the constant of proportionality, γ the intrinsic lifetime of the excited state of the donor, η the refractive index of the medium, C a constant and κ is the geometric factor, reflecting the relative orientations of the transition dipoles.The functions f and ε are the fluorescence of the donor and the absorption of the acceptor, respectively. The

probability with which the energy transfer will occur strongly depends on the magnitudes and relative orientations of the transition dipoles.

In a case of a strong Coulombic coupling of the transition dipoles of closely interacting pigments, as seen in the B850 ring of LH2 from *Rps. acidophila*, one cannot speak of a particular chromophore being excited. Instead the excitation appears delocalized over the interacting pigments. This exciton energy of interaction is proportional to the energy transfer rate [34] and varies as the inverse cube of the distance between the dipoles.

In contrast, the Dexter (or exchange) mechanism [66] is described by the simultaneous exchange of two electrons between the donor and acceptor pair. The necessary condition for this process is close proximity between the donor and acceptor, which promotes orbital overlap.

APPENDIX B (Delphi Program)

DelPhi program [16-18] is a numerical method for calculating the electrostatic potential and electrostatic fields of molecules in solution using the Poisson-Boltzmann equation. The Poisson-Boltzman equation is

$$\nabla[\varepsilon(\vec{r})\nabla\Phi(\vec{r})] = -4\pi\rho(\vec{r}) + \kappa^2 \sinh(\Phi(\vec{r}))$$

(B1)

Here Φ is the electrostatic potential and is determined by ε, the spatial dielectric function, κ, a modified Debye-Huckel parameter and ρ, the charge distribution function. The modified Debye-Huckel parameter κ is equal to

$$\left(\frac{8\pi e^2 N_A I}{1000kT}\right)^{1/2}$$

(B2)

where I is the bulk salt concentration in moles per litre, N_A is Avogadro's number, k is Boltzmann's constant, T the temperature in Kelvin and e the charge on an electron in Coulombs. κ provides a measure of how salt attenuates the effect of a charge with distance. Most application use the linearized Poisson-Boltzmann equation

$$\nabla[\varepsilon(\vec{r})\nabla\Phi(\vec{r})] - \kappa^2\Phi(\vec{r}) = -4\pi\rho(\vec{r})$$

(B3)

which is an approximation that is valid when $e\Phi << kT$. Once when the electrostatic potential is calculated the electrostatic field can be simply obtained taking the gradient of the electrostatic potential.

The functions of DelPhi may be divided into:

(a) preprocessing - the assignment of values to the finite-difference grid before solving the equation; in other words, preprocessing reads in atom coordinates in Brookhaven Protein Data Bank format (pdb file), along with Van der Waals radii and set of atomic charges, and sets up the three-dimensional (65x65x65) arrays used in the iterative solution of the Poisson-Boltzmann equation; and

(b) iterative solution of the equation which employs a finite-difference metod similar to that of Klapper et al [39].

APPENDIX C (One letter amino-acid notation)

One letter amino-acid notation is used to write a protein's primary structure as a sequence of letters. One letter is assigned to each amino acid:

ALA A	CYS C
ASP D	GLU E
PHE F	GLY G
HIS H	ILE I
LYS K	LEU L
MET M	ASN N
PRO P	GLN Q
ARG R	SER S
THR T	VAL V
TRP W	TYR Y

Also one of the basic classification of amino acids is related to their interaction with water. If they like to interact with water they are called hydrophilic amino acids. If they don't like to interact with water, they are hydrophobic:

nonpolar (hydrophobic)

GLY, ALA, VAL, LEU, ILE

PHE, PRO, MET

polar, uncharged (hydrophilic)

SER, THR, TYR, TRP, ASN

GLN, CYS

polar, charged (hydrophilic)

ASP, GLU, LYS, ARG, HIS

References

[1] Borisov, A. Yu.; Freiberg, A. M.; Godik, V. I.; Rebane, K. K.; Timpmann, K. K. *Biochim. Biophys. Acta* **1985**, 807, 221.

[2] Sundstrom, V.; Van Grondelle, R.; Bergstrom, H; Akesson, E; Gillbro, T. *Biochim. Biophys.Acta* **1986**, 851, 431.

[3] Du, M.; Xie, X., Jia, Y.; Mets, L.; Fleming, G. R. *Chem. Phys. Lett.* **1993**, 210, 535.

[4] Hess, S.; Akesson, E.; Cogdell, R.; Pullerits, T.; Sundstrom, V. *Biophys. J.* **1995**, 69, 2211.

[5] Visser, H. M.; Somsen, O. J. G.; Van Mourik, F.; Lin, S.; Van Stokkum, I.; Van Grondelle, R. *Biophys. J.* **1995**, 69, 1083.

[6] Shreve, A. P.; Trautman, J. K.; Frank, H. A.; Oewns, T. G.; Albrecht, A. C. *Biochim. Biophys. Acta* **1991**, 1051, 280.

[7] Timpmann, K.; Zhang, F. G.; Freiberg, A.; Sundstrom, V. *Biochim. Biophys. Acta* **1993**, 1183, 185.

[8] Kleinherenbrink, A. M.; Deinum, G.; Otte, S.C. M.; Hoff, A. J.; Amesz, J. *Biochim. Biophys. Acta* **1992**, 1099, 175.

[9] Cogdell, R.J. & Frank, H.A. (**1987**). How carotenoids function in photosynthetic bacteria. *Biochim. Biophys. Acta* 895, 63-79.

[10] Siefermann-Harms, D. (**1985**). Carotenoids in photosynthesis. I. Location in photosynthetic membranes and light-harvesting function. *Biochim. Biophys. Acta* 811, 325-355.

[11] C. A. Wraight, R. J. Cogdell, B. Chance, in *The Photosynthetic Bacteria*, R. K. Clayton and W. R. Sistrom, Eds. (Plenum, New York, **1978**), pp. 471-511.

[12] T. Kakitani, B. Honig, A. R. Crofts, *Biophys. J.* **1982**, 39, 57.

[13] Koppenol, W. (**1981**) in *Oxygen and Oxy-Radicals in Chemistry and Biology* (Rogers, M., & Powers, E., Eds.) pp 691-674. Academic Press, New York.

[14] Sharp, K., Fine, R., & Honig, B. (**1987**) *Science* 236, 1460-1463.

[15] Northrup, S., Boles, J., & Reynolds, J. (**1988**) *Science* 241, 67-70.

[16] Klapper,I., Hagstrom, R., Fine, R., Sharp, K., Honig, B., *Proteins* **1986**, 1, 47.

[17] Gilson, M. K., Honig, B., *Proteins* **1988**, 3, 32.

[18] Nicholls, A., Honig, B., *J. Comput. Chem.* **1991**, 12, 435.

[19] Crofts, A. R., Prince, R. C., Holmes, N. G., and Crowther, D., **1974**, *Electrogenic electron transport and the carotenoid change in photosynthetic bacteria* , in: Proceedings of Third International Congress on Photosynthesis (M. Avron, ed.), pp. 1131-1146, Elsevier Scientific, Amsterdam.

[20] Labhart, J. (**1967**) in: *Advances in Chem. Phys.* (Prigogine, I. ed) vol.13 pp. 174-204, Interscience, New York.

[21] Liptay, W. (**1969**) *Angew. Chem.* (Internat. Edit) 8, 177-188.

[22] Reich, R. Scheerer, R. and Sewe, K.-U. (**1976**) *Bre. Bunsenges. Physik. Cem.* 80, 254-246.

[23] Gottfried, D. S., Steffen, M. A., Boxer, S. G.(**1991**). Large Protein-Induced Dipoles for a Symmetric Carotenoid in a Photosynthetic Antenna Complex. *Science* 251, 662-665.

[24] Holmes, N. G. and Crofts, A. R. (**1977**) *Biochim. Biophys. Acta* 459, 492-505.

[25] T. Kakitani, B. Honig, A. R. Crofts, *Biophys. J.* **1982**, 39, 57.

[26] Amsez, J. and Vrendenberg, W. J. (**1966**) in: *Currents in Photosynthesis* (Thomas, J.B. and Goedheer, J. C. eds), pp 75-83. Donker, Rotterdam.

[27] Middendorf, T. R., Mazzola, L. T., Law, K., Steffen, M. A., Boxer, S. G. *Biochim. Biophys. Acta* **1993**, 223, 1143.

[28] Gunner, M. R., Nicholls, A., Honig, B. *J. Phys. Chem.* **1996**, 100, 4277.

[29] Seeling, J., Limacher, L., and Bader, P., (**1972**). Molecular architecture of liquid crystalline bilayers, *J. Am. Chem. Soc.* 94: 6364-6371.

[30] Hudson, B. S. Kohler, B. E. & Shulten, K. (**1982**). Linear polyene electronic structure and potential surfaces. In *Excited States* (Lim, E. C., ed), vol. 6, pp. 22-95, Academic Press, New York.

[31] K. Brejc, R. Ficner, R. Huber, S. Steinbacher, *J. Mol. Biol.* **1995**, v. 249, 424.

[32] M. Duerring, G. B. Schmidt, R. Huber, *J. Mol. Biol.* **1991**, v. 217, 577.

[33] Isaacs, N. W., Cogdell, J. R., Freer, A. A., Prince, S. M., *Current Opinion in Structural Biology* **1995**, 5:794-797.

[34] McDermott, G., Prince, S. M., Freer, A. A., Hawthornthwaite-Lawless, A. M., Papiz, M. Z., Cogdell, R. J., Isaacs, N. W., *Nature* **1995**, 374:517-521.

[35] J. Koepke, X. Hu, C. Muenke, K. Schulten, H. Michel, *Structure* **1996**, v. 4, 581.

[36] Moffat, A. S., *Science* **1996**, v.272, 1743-1744.

[37] Deisenhofer, J., Michel, H., *EMBO J.* **1989**, 8, 2149.

[38] Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. F., Tasumi, M., *J. Mol. Biol.* **1977**, 112, 535.

[39] O'Neil, M. P.; Wasielewski, M. R.; Khaled, M. M.; Kispert, L. D. *J. Chem. Phys.* (**1991**), 95, 7212.

[40] Ponder, M.; Mathies, R. *J. Phys. Chem.* (**1983**), 87, 5090.

[41] Mathies, R.; Stryer, L. *Proc. Natl. Acad. Sci. U. S. A.* (**1976**), 73, 2169.

[42] Crofts, A. R.; Prince, R. C.; Holmes, N. G.; and Crowther, D., (**1974**), Electrogenic electron transport and the carotenoid change in photosynthetic bacteria, in: *Proceedings of the Third International Congress on Photosynthesis* (M. Avron, ed.), pp. 1131-1146, Elsevier Scientific, Amsterdam.

[43] Gosztola, D.; Yamada, H. and Wasielewski, M. R. *J. Am. Chem. Soc.* (**1995**), 117, 2041-2048.

[44] Lockhart, D. J. and Boxer, S. G. *Biochemistry* (**1987**), 26, 664-668.

[45] Bottcher, J. F. *Theory of Dielectric Polarization*, Elsevier: New York, (**1973**); Vol. 1.

[46] Liptay, W. Z. *Naturforsch.* (**1965**), 20a, 272-289.

[47] Roth, M.; Lewit-Bentley, A.; Michel, H.; Deisenhofer, J.; Huber, R.; Oesterhelt, D. *Nature* (**1989**), 340, 659.

[48] Nicholls, A.; Sharp, K.; Honig, B. *Proteins* (**1991**), 11, 281.

[49] Dietmar Bernlochner in *Electroabsorption und Magneto-Optische Differenz-Spektroskopie an den Singullet-und Triplett-Zustanden von Pigment-Protein-Komplexen der Photosynthese*, Stuttgarter Hochschulschriften, (**1995**), p.73.

[50] Honig, B.; Sharp, K.; Yang, A.-S. *J. Phys. Chem.* (**1993**), 97, 1101.

[51] A. C. Wallace, R. A. Laskowski, J. M. Thornton. *Protein Engineering* **1995**, v. 8, 127.

[52] Sauer, K. (**1975**). In *Bioenergetics of Photosynthesis* (Govindjee, ed.), p.115, Academic Press, San Francisco.

[53] Schirmer, T., Huber, R., Schneider, M., Bode, W., Miller, M. & Hackert, M. L. (**1986**). *J. Mol. Biol.* 188, 651-676.

[54] Bogorad, L. (**1975**). *Ann. Rev. Plant. Physiol.* 26, 369-401.

[55] Tandeu de Marsac, N. (**1983**). *Bull. Inst. Pasteur*. 81, 201-254.

[56] MacColl, R. & Berns, D. S. (**1981**). *Isr. J. Chem.* 21, 296-300.

[57] Arciero, D. M., Bryant, D. A. & Glazer, A. N. (**1988a**), *J. Biol. Chem.* 263, 18343-18349.

[58] Arciero, D. M., Dallas, J. L. & Glazer, A. N. (**1988b**), *J. Biol. Chem.* 263, 18350-18357.

[59] Arciero, D. M., Dallas, J. L. & Glazer, A. N. (**1988c**), *J. Biol. Chem.* 263, 18358-18363.

[60] Tronrud, D. E., Schmidt, M. F., & Matthews, B. W. *J. Mol. Biol.* (**1986**), 188, 443.

[61] Pascher, I. (**1976**). *Biochim. Biophys. Acta.* 455, 433-451.

[62] Pascher, I. & Sondell, S. (**1977**). *Chem. Phys. Lipids.* 20, 175-191.

[63] Pearson, R. H. & Pascher, I. (**1979**). *Nature (London).* 281, 499-501.

[64] Forster T (**1948**) Intermolecular energy transfer and fluorescence. *Ann. Phys.* (Leipzig) 2:55-75.

[65] Forster T (**1959**) Transfer mechanisms of electronic excitation. *Discuss. Faraday Soc.* 27:7-17.

[66] Dexter DL (**1953**) A theory of sensitized luminescence in solids. *J. Chem. Phys.* 21:836-850.

[67] Lee FS, Chu ZT, Warshel A (**1993**) *J. Comp. Chem.* 2-51.

[68] Takashima, S., & Schwan, H. P. (**1965**) *J. Phys. Chem.* 69, 4176-4182

[69] Pennock, B. D., & Schwan, H. P. (**1969**) *J. Phys. Chem.* 73, 2600-2610.

[70] Harvey, S. C., & Hoekstra, P. (**1972**) *J. Phys. Chem.* 76, 2987-2994.

[71] Gunner, M. R., & Honig, B. (**1991**) *Proc. Natl. Acad. Sci. USA* 88, 9151-9155.

[72] Bashford, D., & Karplus, M. (**1990**) *Biochemistry* 29, 10219-10225.