Review

Intrinsic protein electric fields: basic non-covalent interactions and relationship to protein-induced Stark effects

Monique Laberge *

Johnson Research Foundation, Department of Biochemistry and Biophysics, School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

Received 15 February 1998; revised 20 March 1998; accepted 28 March 1998

Abstract

Knowledge of the interactions involving charged, polar and polarizable groups in proteins is fundamental, not only because they are important determinants for gaining insight into biophysical molecular recognition and assembly processes, but also for understanding how the matrix of a protein can be viewed as an electric field capable of inducing Stark perturbations on the spectral properties of biological optical centers. This review describes the essential features of non-covalent interactions in protein systems and discusses the concept of the dielectric constant of a protein in the context of different microscopic and macroscopic modeling approaches. It also provides an account of a specific type of high resolution vibrational and optical Stark spectroscopy attempting to correlate the observed spectral properties of biological optical centers to the intrinsic protein fields induced by the matrix in which they reside. © 1998 Elsevier Science B.V. All rights reserved.

Keywords: Electrostatic interaction; Protein electric field; Stark effect

Contents

1. Scope of review .......................................................... 306

2. The protein viewed as a self-modulating, non-uniform electric field .............................................. 306
   2.1. Basic non-bond interactions in proteins .............................................. 307
   2.2. The dielectric constant ................................................................. 316

3. Electrostatic modeling in proteins ................................. 319
   3.1. Macroscopic methods ........................................................... 320
   3.2. Microscopic methods .................................................................. 321
   3.3. Combined approaches .............................................................. 322

Abbreviations: FDPB, finite difference Poisson–Boltzmann method; KF, Kirkwood–Fröhlich formalism; PDLD, protein dipoles Langevin dipoles method; SE, Stark effect; TK, Tanford–Kirkwood method; VSE, vibrational Stark effect

* Fax: +1 (215) 573-2042; E-mail: labergem@mail.med.upenn.edu

0167-4838/98/$19.00 © 1998 Elsevier Science B.V. All rights reserved.
PII: S0167-4838(98)00100-9
1. Scope of review

This review is intended to describe not so much the recent advances in understanding electrostatic interactions and the methods used to model them in proteins, but rather the background required to understand how protein matrix field effects can affect the vibrational and optical spectral properties of biological optical centers. No attempt will be made to systematically review the field of non-covalent interactions; excellent monographs and reviews, especially in the context of protein folding and molecular recognition, are available [1–12]; rather, after briefly outlining the essential features of the basic non-covalent interactions, the review will follow with a discussion of some of the microscopic and macroscopic treatments used to model electrostatics in protein systems, referring the reader to the more fundamental papers for a more thorough theoretical assessment. In order to make it reasonably self-contained, the range of recent applications of these modeling methods will not be discussed; the focus will be instead on accurately describing the basic issues raised by protein intrinsic field interactions in a specific context of interest, i.e. consideration of the available electrostatic models in the interpretation of Stark-perturbed vibrational and electronic spectral properties.

During the past decade, high resolution laser spectroscopy has significantly contributed to the realistic description of the macroscopic behavior of proteins and the reader is referred to the companion reviews on spectral hole burning [13,14] and fluorescence line narrowing [15] in this issue. However, the interpretation of the spectral behavior of biological optical centers, as perturbed by electric field effects, remains challenged by the fact that some of the forces involved in generating the internal field of a protein are not yet fully understood as well as by the exceptional complexity of the interactions. From this perspective, this review should accordingly first and foremost be considered as an attempt to clearly state the general problem of describing the electrostatics of complex biophysical systems. Second, it will draw attention to a comparative type of Stark spectroscopy, in which the optical or vibrational spectral properties of protein optical centers are compared in the presence of slightly different electric fields induced by charge redistribution in their protein matrices, either as a result of conformational change or of substrate binding. Hopefully, it will generate more insight than confusion and encourage the development of the combined computational–spectroscopic approaches still required to monitor accurately structure–function problems.

2. The protein viewed as a self-modulating, non-uniform electric field

A protein consists of an assembly of amino acids held together by strong covalent bonds along its backbone and a multitude of much weaker, non-covalent cross-connections. These amino acid assemblies create a distribution of charge with a given set of spatial coordinates which generates an electric field in the overall matrix and thus, proteins can be viewed as simple distributions of charge (cf. Fig. 1). This spatial charge distribution, or protein internal field, however, is highly asymmetric: for example, some regions are strongly polar (peptide bonds) while others are highly charged (wherever charged residues are present). Since the protein is constantly undergoing fluctuations, its backbone and side chain
coordinates are in constant motion and the non-covalent interactions that they specify are also being constantly modified along with the solvent that has to reorganize. It is believed that the cooperative effects of all these fluctuating interactions, in spatial and in field quantities, determine the biologically useful function of a protein. As such, the field generated by a protein is self-modulating, in that it responds in an electrostatically significant way to stimuli which are understood to be random (fluctuations) as well as directed (e.g. ligand binding, resulting for example in the ionization of some residues); and this field is also non-uniform, in that charge is not evenly distributed in the protein matrix and in that it also undergoes significant charge restructuring in interacting with a functional partner or in response to conformational change. This was recently demonstrated in the molecular dynamics study of Wendoloski et al. [16], who explored the conformational space of tuna cytochrome c and showed that conformers significantly removed from the average structure also undergo notable electrostatic rearrangement. Fig. 1 shows that charge redistribution also occurs in horse heart cytochrome c as it samples different conformations.

The study of the non-covalent interactions that define the local electric field of a protein (used here to specify the intrinsic field as opposed to an externally applied field) in the absence of time-dependent variations is the subject of electrostatics in which three broad types of interactions are generally recognized: a first class, which includes all force-type interactions of a fundamentally electrostatic nature; a second class, consisting of interactions which require quantum mechanical treatment or crude approximations; and lastly, hydrophobic interactions, which are not the result of the application of a force per se, but rather the result of entropy changes. We summarize them in the next section.

2.1. Basic non-bond interactions in proteins

2.1.1. Electrostatic forces

Non-covalent interactions are usually defined as interactions which are basically electrostatic in nature, i.e. are the result of forces that electrical charges exert on one another. This discussion accordingly starts with a brief description of these interactions in proteins, following the multipole approach of Burley [2] which is of interest not only because it presents the basic electrostatic forces as a succinct
'multipolar' whole but also because it describes well the lesser investigated interactions, such as quadrupole and weakly polar.

In the absence of orbital overlap, the electrostatic potential energy resulting from a charge distribution \( \rho(r) \) placed in an electric field \( \Phi(r) \) induced by the presence of a second nearby charge distribution \( \rho'(r) \) can be described as a multipole representation [2]:

\[
W = \int \rho(r) \Phi(r) \, dr
\]  

(1)

where \( r \) stands for the position vector over the volume of \( \rho(r) \) or \( \rho'(r) \) and with the electric potential function usually expressed as:

\[
\Phi(r) = \int \frac{\rho'(r')}{(|r-r'|)^4} \, dr'
\]

(2)

In proteins, the charge distribution is associated with discrete atoms and can accordingly be expressed as:

\[
\rho(r) = \sum_a \rho_a(r)
\]  

(i.e. as a function of \( \rho_a(r) \)), the atomic charge density function for atom \( a \). Eqs. 1 and 2 can then be reformulated as:

\[
W = \sum_a \int \rho_a(r) \Phi(r) \, dr
\]

(4)

Table 1
H-bonding R-groups

<table>
<thead>
<tr>
<th>Donors</th>
<th>Acceptors</th>
</tr>
</thead>
<tbody>
<tr>
<td>arg</td>
<td>his, thr =N</td>
</tr>
<tr>
<td></td>
<td>H</td>
</tr>
<tr>
<td>arg</td>
<td>glu, ser O=C—O</td>
</tr>
<tr>
<td></td>
<td>asn, gln C=O</td>
</tr>
<tr>
<td></td>
<td>tyr, thr H</td>
</tr>
<tr>
<td>lys</td>
<td>met —S</td>
</tr>
<tr>
<td></td>
<td>H</td>
</tr>
<tr>
<td>trp</td>
<td>tyr, thr ring</td>
</tr>
<tr>
<td></td>
<td>H</td>
</tr>
<tr>
<td>gln</td>
<td>—N—H</td>
</tr>
<tr>
<td>asn</td>
<td>—N—H</td>
</tr>
<tr>
<td></td>
<td>H</td>
</tr>
<tr>
<td>tyr</td>
<td>—C—O—H</td>
</tr>
<tr>
<td>thr</td>
<td></td>
</tr>
<tr>
<td>ser</td>
<td></td>
</tr>
<tr>
<td>cys</td>
<td>S—H</td>
</tr>
</tbody>
</table>

After [29], p. 185.

and

\[
\Phi = \sum_b \Phi_b = \sum_b \int \rho'_b(r')/(|r-r'|) \, dr'
\]

(5)

\( \Phi(r) \) can be expanded in series to yield the following expression for the electrostatic energy resulting from the interaction of two molecules \( i \) and \( j \):

\[
W = \sum_{ij} q_i q_j/|r_{ij}| + 1/2 \sum_{ij} (q_i q_j) \, d^2 \Phi(0)/dr_i dr_j + ...
\]

(6)

with \( q \) representing the net charge and \( \mu \) the dipole moment. Further, the potential function \( \Phi(r) \) can be written as a summation of atomic components:

\[
\Phi(r) = \sum_i q_i |r|^{-1} + (\mu_i \cdot r) |r|^{-3} + (r \vec{q}_d \cdot r) |r|^{-5}...
\]

(7)

which leads to the following expression for the electrostatic energy summing the interactions between the multipoles of two molecules:

\[
W = \sum_i \sum_j W_{ij}
\]

(8)

The first nine interaction terms of the expansion are:

(i) \( W_{00} = q_1 q_2/|r| \) (monopole–monopole): describing the energy between interacting groups that have a net charge; the force of interaction is non-directional and only shows a dependence on the distance between charges \( q_1 \) and \( q_2 \). As such, this type of ‘charge–charge’ interaction has the longest range among the non-covalent interactions and the potential energy of interaction is usually expressed as:

\[
V_{c–c} = q_1 q_2/|r|_{ij}
\]

(9)

The typical extent of these energies is well illustrated by the 0.3–0.5 kcal mol\(^{-1}\) calculated at \( \sim 12 \) Å for the interaction between single charges on the surfaces of proteins (e.g. Glu, Asp, Lys, Arg) which gradually fall off to \( \sim 0.05–0.3 \) kcal mol\(^{-1}\) at 20 Å [17].

(ii) \( W_{01} = q_1 (\mu_2 \cdot r)/|r|^{3} \) (monopole–dipole): describing the energy of a charge \( q_1 \) in the field of a molecule that has no net charge but a permanent electric dipole moment \( \mu_2 \). arising from the asymmetric displacement of electron charge density along covalent bonds; thus, the dipole moment of such polar molecules strongly depends on the environment [7]. Accordingly, the charge–dipole interaction is directional and depends on the orientation of the dipole (cf. Fig. 2a); it is short-range and proportional to \( 1/|r|^3 \):

\[
V_{c–d} = q_1 \mu_2 \cos \theta_{ij}/|r|_{ij}
\]

(10)
Fig. 2. (a) Charge-dipole interaction. At long distances, the energy of interaction varies as $1/r^2$. This type of interaction is directional as it is strongly dependent on the respective orientations of the interacting dipole and charge. (b) Dipole-dipole interaction. Much shorter range interaction than the charge-dipole. (c) The interaction becomes repulsive when the dipoles are in a parallel orientation. (d) The maximum interaction occurs when the two dipoles are spatially sequential to each other with the positive end of the first dipole abutting the negative end of the other (cf. [2] pp. 137–138).
A classic example would be the interaction of a positively charged \(-\text{N}^+\text{H}_3\) group with a polar water molecule, which has a dipole moment of 1.85 D or with a glycine (16.7 D). The distribution of partial charge among the constituent amino acids of an \(\alpha\)-helix can also generate a dipole moment, the magnitude of which is considerable enough to be equivalent to placing 0.5–0.7 positive unit charge at the N-terminus and 0.5–0.7 negative unit charge at the C-terminus [18,19]. On a larger scale, proteins as a whole also generate a dipole moment, the magnitude of which is considerable enough to be equivalent to an inverse sixth power distance dependence. A glycine (16.7 D). The distribution of partial charge grouped as interactions with electronic quadrupoles: the spatial distribution of their constituent amino acid charges and their fluctuations generates an overall dipole which can represent a significant contribution in molecular recognition processes [1,20].

(iii) \(W_{10} = q_2(\mu_1 \cdot r)/|r|^3\) (dipole–monopole).

(iv) \(W_{11} = (\mu_1 \cdot \mu_2)/|r|^{-3} - 3(\mu_1 \cdot r)(\mu_2 \cdot r)/|r|^{-5}\) (dipole–dipole): a permanent dipole can also interact with another permanent dipole (Fig. 2b). The interaction of antiparallel dipoles is weaker and it becomes repulsive when the dipoles are in a parallel orientation (Fig. 2c). The strongest interaction occurs when the two dipoles are following each other in line with the positive end of the first dipole abutting the negative end of the other (Fig. 2d). The energy of interaction is not as strong as that of the charge–dipole and is described for two point dipoles by:

\[
V_{\text{dip–dip}} = -\mu_1 \cdot \mu_2 (2 \cos \theta \cos \phi - \sin \theta \sin \phi)/r^3
\]  
(11)

Two dipoles of equal magnitude of 1 D and sequentially aligned as in Fig. 2d thus have an interaction energy in vacuo approximately equal to \(kT\) at \(r = 0.36\) nm and when overlapping in a parallel arrangement as in Fig. 2c, \(kT\) is achieved at \(r = 0.29\) nm. This amounts to stating that dipolar interactions can only bind very polar molecules. At larger separations, Eq. 11 must include Boltzmann averaging of the interaction energy which introduces a weighting factor for those dipole orientations that have a lower energy [7]. The net effect is that we thus obtain the ‘orientation’ or Keesom interaction energy which has an inverse sixth power distance dependence.

(v–ix) The last four terms of the expansion are grouped as interactions with electronic quadrupoles:

\[
W_{02} = q_1 (r q_2 \cdot r)/|r|^{-5}\) (monopole–quadrupole)

\[
W_{20} = q_2 (r q_1 \cdot r)/|r|^{-5}\) (quadrupole–monopole)

\[
W_{12} = 2(\mu_1 \cdot q_2 \cdot r)/|r|^{-5} - 5(\mu_1 \cdot r)(q_2 \cdot r)/|r|^{-7}
\]  
(dipole–quadrupole)

\[
W_{12} = -2(\mu_2 \cdot q_1 \cdot r)/|r|^{-5} + [5(\mu_1 \cdot r)(q_1 \cdot r)/|r|
\]  
(quadrupole–dipole)

\[
W_{22} = (1/3)[\text{trace of } q_1 \cdot q_2]/|r|^{-5} - (20/3)[r q_2 \cdot q_1 \cdot r]/|r|^{-7}
\]  
\[+(35/2)(r q_1 \cdot r)(r q_2 \cdot r)/|r|^{-9}\]  
(quadrupole–quadrupole)

(12)

Fig. 3. (a) Schematic view of different electric multipoles; examples would be \(\text{Na}^+\) and \(\text{F}^-\) (monopoles), the \(\text{HCl}\) molecule (dipole) and \(\text{CO}_2\) (quadrupole). \(\text{CO}_2\) has no dipole moment (because it is symmetrical about the C atom), but the polarity of the C–O bonds results in a charge distribution that corresponds to a quadrupole. (b) The distribution of partial electronic charge in benzene, treated as discrete charges located at each atom. The molecule then has no net charge, no net electronic dipole moment and consists of three electronic quadrupoles [2,119].
The electronic quadrupole moments of aromatic residues can be quite significant in proteins. They give rise to quadrupolar interactions which mostly occur between aromatic side chains and a selective group of other atoms. They result from the partial charge distribution occurring in, for example, aromatic groups and involve interactions between electronic moments, such as quadrupole–quadrupole, monopole–quadrupole and dipole–quadrupole (Fig. 3a). They usually involve amino acids such as Trp, Phe and Tyr, which feature a localization of partial electronic charge as sometimes modeled in benzene (Fig. 3b). Briefly, this treatment distributes the $\delta^+\pi$-electron density cloud as discrete partial charges in the plane of the molecule and assigns partial positive charges to the $\delta^+$ hydrogens bound to the ring carbons. The interaction is thus strongly dependent on the spatial arrangement of the atoms.

2.1.2. Induced polarization

Molecular polarization can be induced in any atom or molecule by the electric fields produced by nearby molecules and the electronic polarizability, $\alpha_o$, is then defined by the strength of the induced dipole moment which they acquire when subjected to such fields [7]. $\alpha_o$ can be understood simply as the capacity of a given distribution of electrons to be spatially distorted; as such, it is inversely proportional to the electronegativity of the atom and directly proportional to the atomic radius. In the case of polar molecules, other effects will of course contribute to the polarizability. But, for non-polar molecules, which have no permanent dipole moment, and which thus undergo induced polarization in an externally applied field, the polarizability essentially arises as a result of the polarization of their electronic charge distribution and is thus viewed as the readjustment of their internal molecular charge distribution, i.e. of the negative electron cloud relative to the nucleus, when they are subjected to a static external potential [21]. Their polarizability is thus related to the induced dipole moment:

$$\mu_{\text{ind}} = \alpha_o E$$

In the case of polar molecules, which can be considered as freely rotating dipoles with a time-averaged $\mu_{\text{avg}} = 0$, an added contribution, i.e. that of the orientational polarizability ($\alpha_{\text{orient}}$), due to the effect of the field on the Boltzmann-averaged orientations of the rotating dipole, must be taken into account [7]. The total polarizability ($\alpha$) of a polar molecule is then no longer simply defined as $\mu_{\text{ind}}/E$ but rather as:

$$\alpha = \alpha_o + \alpha_{\text{orient}}$$

(14)
As a result of being placed in an electric field, the center of gravity of the electron density shifts because the induced dipole moment lowers the energy of the system. The bond dipoles also readjust to align favorably with respect to the field. The magnitude of such electronic polarizability is generally of the order of \( r^3 \) as can be seen with water which has \( \alpha_o/4\varepsilon_o = 0.114 \ \text{nm}^3 \) \( (r = 0.135 \ \text{nm}) \) [7].

In Fig. 4, the overall effect of polarization can be seen by comparing calculations performed on a water molecule placed in a uniform electric field of \( 5.14 \times 10^7 \ \text{V/cm} \) and under zero field conditions [2]. The magnitude of the field is representative of the field experienced by molecules located at van der Waal distance from polar molecules. The results show that the charge densities are not significantly affected by the external field, i.e. that the conformation and size of water remains constant and that the charge density shifts in the direction of the field. However, the electrostatic potential arising from the charge distribution of water varies under applied field by some 10% [21].

In proteins, two major induced polarization cases occur: the interaction between a monopole and a polarizable group and the interaction between a dipole and a polarizable group. The energy of interaction between a charged molecule and an isotropically polarized one can be described as follows:

\[
E_{c-\text{inddip}} = -Ze\mu_{\text{ind}}/\varepsilon r^2 = -Ze^2\alpha/2\varepsilon^2 r^4 \tag{15}
\]

where \( Ze \) is the charge on the charged molecule, \( \mu_{\text{ind}} \) is the induced dipole on the interacting molecule and \( \alpha \) is the isotropic polarizability. In the case of dipole-induced dipole interactions, the polarizing field originates from a permanent dipole instead of from a charge and the interaction energy is referred to as the Debye or induction energy:

\[
E_{\text{dip-\text{inddip}}} = -\mu_{\text{ind}}/\varepsilon r^2 = -\mu^2/\varepsilon^2 r^6 \tag{16}
\]

### 2.1.3. Weakly polar interactions

This subject is treated in depth by Burley and Petsko [22] and the reader is referred to their detailed discussion. The importance of these weak interactions is being increasingly recognized in proteins and this realization followed, not only from ab initio calculations, but also from biophysical observations and structural studies.
calculations performed on model compounds, but also from the fact that there are statistically significant preferred three-dimensional stereochemistries between some specific atoms or chemical groups and aromatic residues, such as between aromatic amino acids themselves, between oxygen and aromatic residues and between sulfur atoms and aromatic side chains. Fig. 5 shows such an aromatic cluster in cod parvalbumin.

It can be seen that aromatic residues, Phe and Trp, are arranged ‘edge-to-face’, a geometry known to promote favorable interaction between two quadrupoles (cf. Fig. 3 and Section 2.1.1, v–ix) and it has been suggested that such interactions stabilize the tertiary structure of calcium-binding proteins while defining their core functional stereochemistry. Another illustration is shown in Fig. 6, which compares the structures of horseradish peroxidase in the presence and absence of substrate within a 6 Å-probe radius from the heme. The right-hand side view clearly shows Phe A68 also positioned ‘edge-to-face’ with respect to the benzhydroxamic acid substrate and thus no doubt contributing to its stabilization.

2.1.4. van der Waals and repulsive forces

The van der Waals interaction occurs between all molecules including those that have no net charge and that are not polarizable. It is due to the motions of the electrons as they approach one another, and to their tendency to avoid one another. As such, it is quantum mechanical in nature and a very crude approximation is given by:

$$E_{\text{disp}} = -\alpha^2 I / r^6$$

where $I$ is the mean ionization energy. Dispersion forces are characterized as follows [7]: (i) they can be effective at long range ($>10$ nm) and become very significant at very short-range (0.2 nm); (ii) they can be repulsive or attractive and usually do not follow simple power law derivations; (iii) they also orient and align molecules and as such, their effect is weak; (iv) the dispersion interaction between two molecules is also non-additive, i.e. is affected by the presence of other nearby molecules.

Between polar molecules, the long range van der Waal interaction sums three predominant inverse sixth power contributions: the induction force, the

---

Fig. 6. Stereoview of the horseradish peroxidase heme pocket environment showing the residues located at 6 Å from the heme (red). Left: the enzyme without substrate. Right: HRP in the presence of the benzhydroxamic acid substrate (yellow), stabilized by a weakly polar interaction with Phe A68 (purple). HRP coordinates, pdb1atj.ent and pdb2atj.ent [98], from the Brookhaven Protein Data Bank [117].
orientation force and the dispersion force. The first is the Debye contribution, the second is simply the Keesom interaction described in the previous section (cf. supra, Section 2.1.1, iv) and the last is usually the dominant force in the overall net van der Waals force, except for small highly polar molecules. Dispersion energies are in the range of 0.2 to 2.4 kcal mol\(^{-1}\) between small molecules that approach one another to their van der Waals radii. However, these energies are higher for \(\pi\)-electron systems with low-lying electronic excitations (\(\sim 6\)–8 kcal mol\(^{-1}\)) [23].

Between macromolecules, low-frequency density fluctuations give rise to strong, long-range dispersion interactions [24]. In protein systems, the structural complexity has led to the use of idealized models with microscopic treatments attempting to relate dispersion forces to molecular properties (such as the ionization potential) and macroscopic approaches attempting to overcome the oversimplification of the former by incorporating rigorous spatial and geometric considerations [25–27]. Recent calculations performed on actual molecular shapes show that the magnitude of dispersion interactions differ significantly from the results obtained using idealized (spherical) models. This has an important consequence in that the actual irregular shape of a protein points to a significant steric contribution, i.e. molecules of complementary shape for example are found to exhibit very strong attractive interaction terms, which is not at all predicted with sphere models [27].

Repulsive interactions act between molecules that are not covalently bound as they approach one another at distances close enough to allow overlap of their outer electron orbitals and to give rise to a strong, very short range, repulsive force. As such, it is also quantum mechanical and usually described by a number of various empirical approaches such as the hard and soft sphere repulsion potentials [7].

The most widely used potential is the Lennard-Jones ‘6–12’ potential which combines an inverse sixth power potential for the van der Waals interaction and an inverse twelfth power repulsive potential:

\[
E_{ij} = \varepsilon_{ij} \left( \frac{r_{ij}^6}{r_{ij}^{12}} - 2 \left( \frac{r_{ij}^6}{r_{ij}^{12}} \right)^{\frac{1}{2}} \right)
\]

where \(\varepsilon_{ij}\) is the potential well depth and \(r_{ij}\) is the interatomic distance at which the minimum occurs. Polarizability will of course significantly affect the attractive term. The repulsive interactions between polar and apolar atoms are believed to play a significant role in the stability of a protein matrix. In cytochrome \(c\), a large number of carbon-backbone nitrogen pairs have interatomic distances some 0.4 nm shorter than the sum of their van der Waals radii. Such pairs would be expected to be subject to repulsive interactions that would increase their mutual separation. That the attractive term predominates is quite surprising; it was shown that the repulsive interaction is overcome by a combination of strong dipole-induced dipole, H-bonding and hydrophobic effects associated with \(\alpha\)-helical packing processes and that the net van der Waals interaction (i.e. attractive) thus destabilizes the structure by some \(\sim 55\) kcal mol\(^{-1}\) [28].

2.1.5. Special interactions: H-bonding, hydrophobic and hydrophilic

Protein fields also include special intermolecular interactions, which all have in common the involvement of water, that most unusual and complex substance. The first special interaction is H-bonding, and it can be viewed as a special case of combined dipole-dipole and charge interaction [2]. The small size and orientation of the large dipole moment of water molecules makes it a strong attractive force, due to the fact that, when an hydrogen atom is covalently bound to O, N or other electronegative atom A, the electron density is partly shifted on the non-H atom; thus, the hydrogen atom is partially unshielded and has partial positive charge. As such, A-H represents the H-bond donor and will accordingly attract H-bond acceptors possessing lone-pair electrons or \(\pi\)-electrons (cf. Table 1). H-bonds belong to three

<table>
<thead>
<tr>
<th>Hydrophobic</th>
<th>Hydrophilic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Molecules</td>
<td>oil, alkanes</td>
</tr>
<tr>
<td>Amino acids</td>
<td>Phe, Trp, Tyr, Ile, Leu, Met, Val</td>
</tr>
<tr>
<td>Molecules</td>
<td>alcohols, sugars, soluble proteins</td>
</tr>
<tr>
<td>Amino acids</td>
<td>Asp, Glu, Lys, Arg, His</td>
</tr>
<tr>
<td>Anionic groups</td>
<td>–COO(^-), –SO(_3)^-</td>
</tr>
<tr>
<td>Cationic groups</td>
<td>–N(^+)(CH(_3))(_3), –N(^+)H(_3)</td>
</tr>
<tr>
<td>Polar groups</td>
<td>–NH(_2), –OH</td>
</tr>
</tbody>
</table>

Compiled from a longer list given by Israelachvili ([7], p. 107).
classes: (i) strong H-bonds, i.e. ‘ionic’ with bond energies varying between 15- and 40 kcal mol\(^{-1}\) and occurring between groups in which there is a deficiency of electron density in the donor group or an excess of electron density in the acceptor group \[N^+...H.....N\]; (ii) moderate H-bonds, i.e. ‘neutral’ with energies ranging between 4 and 14 kcal mol\(^{-1}\) and in which the donor atoms A are electronegative with respect to hydrogen with the acceptor B atoms having lone-pair unshared electrons \[-O-H...O-H-\]; and (iii) weak H-bonds, with energies in the 1-4 kcal mol\(^{-1}\) range which occur when hydrogen is covalently bound to an atom more electroneutral than itself (C–H, Si–H) [29]. Two linearly H-bonded water molecules have a maximum stabilization energy of some \(\sim 4.8\) kcal mol\(^{-1}\) when the oxygens are 0.28 nm apart from each other and the H-bond interaction cancels out for oxygen separations greater than 0.4 nm [30].

In proteins, H-bonds are usually classified into three groups, i.e. main chain-main chain H-bonds between the N-H and O\(\equiv\)C bonds of peptide groups, main chain-side chain H-bonds between the peptide N-H and O\(\equiv\)C groups and R-groups and side chain-side chain H-bonds. It is interesting to note that H-bonding patterns differ significantly in proteins as compared to the types of H-bonding observed in simple amino acid chains and peptides, as discussed by Vinogradov [31] who distinguishes the following: (i) proteins exhibit more peptide-peptide H-bonds than peptide-side chain H-bonds; (ii) they also form significantly more peptide C=O...H-O bonds with side chains than amino acids do; (iii) there is a tendency for the internal waters of a protein to H-bond with peptide carbonyls and side chain hydroxyl groups.

Thus, in the modulation of protein matrix fields, H-bonding must play a significant role: it is recognized that a protein folds so as to avoid leaving any unmatched H-bond donor or acceptor in its interior, consistent with the fact that isolated charged groups are also never found in the hydrophobic core regions of the matrix [1,32].

The hydrophobic interaction (cf. Table 2) is not the result of the application of a force; rather, it is due to hydrophobic side chains assembling tightly together to avoid contact with solvent water. This surprisingly strong attraction between non-polar hydrophobic groups in water is in direct contradiction with van der Waals dispersion theory and arises from the rearrangement of water molecules in the overlapping solvation zones. The interaction still represents an excessively difficult theoretical problem and a sound model has yet to emerge [7,33]. The interaction is known to decrease exponentially with distance in the 0–10 nm range for small solute molecules and it has been proposed that the free energy of hydrophobic interaction is directly proportional to the solute diameter [7]:

\[
\Delta G_{\text{hydr. int.}} = -4.8 \sigma \text{ kcal mol}^{-1}
\]  

The importance of the hydrophobic interaction in the context of protein structure and folding has long been recognized [34]. It is generally accepted that the most stable protein conformations are achieved when the hydrophobic residues are deeply buried in the protein core without contact with solvent water molecules. However, recent work has demonstrated that some globular proteins exhibit a form of distributional disorder leading to the possibility of finding some 40% of the hydrophobic residues at the surface of the protein matrix [2,35]. In the context of protein folding, hydrophobic interactions are believed to drive the protein towards a more condensed structure by decreasing the unfavorable contacts between hydrophobic residues and solvent. In various \(c\)-type cytochromes, the hydrophobic character of specific
residues near the heme is conserved as well as one Trp residue, H-bonded to a heme propionate. It is thought that this reflects the capacity of the protein to protect its redox center by providing it with a hydrophobic environment and similarly to favor its solubility by orienting hydrophilic residues towards its surface [36].

There is no specific ‘hydrophilic effect’ as such, but the interaction is being increasingly recognized as significant in protein systems. It refers to the tendency of given groups and molecules to be water soluble and to strongly repel one another in aqueous solution. Thus it is a repulsive type of interaction since such groups prefer to be in contact with water than with themselves. A good example is provided by a recent study on the hydrophilic bridges present in protein complexes and their importance with respect to both folding and binding processes. They were recognized as very important factors of protein-protein association stability [37]. The hydrophobic and hydrophilic interactions, unlike the van der Waals dispersion interaction, are not additive because they are both solvation phenomena and as such, very dependent on the water arrangement. They are long range interactions and in the presence of amphiphilic groups, as is the case in amino acids and lipid bilayers, the net interaction is never additive. This of course contributes to the extreme complexity of the interaction [7].

The distance dependence of non-bond interactions (cf. Table 3) can provide a measure of their relative significance as long as the equal importance of the orientation factor is fully recognized for those types of interaction that include it. For example, the dipole-dipole interaction shows the same distance dependence as monopole-quadrupole interactions, but the extent of the orientation contribution is significantly different. Also, the importance of previously underestimated interactions, such as the weakly polar, is increasingly being recognized and it has become apparent in several types of studies that we can no longer calculate reliable energies by considering only the more conventional non-bond interactions.

2.2. The dielectric constant

The Coulomb force between two charges \( q_1 \) and \( q_2 \) separated by a distance \( r \) can also be written as:

\[
F = \frac{q_1 q_2}{4\pi\varepsilon_0 r^2} = \frac{z_1 z_2 e^2}{4\pi\varepsilon_0 r^2}
\]  

(20)

where \( \varepsilon_0 \) is the permittivity of free space and \( \varepsilon \) that of the medium. With charges of the same polarity, \( F \) is positive and repulsive and with charges of opposite polarity, \( F \) is negative and attractive. If such a coulombic force acts on a charge \( q_1 \) at a distance \( r \), an electric field is generated such that:

\[
E_1 = \frac{q_1}{4\pi\varepsilon_0 r^2} \text{V m}^{-1}
\]

(21)

and when this field \( E_1 \) acts on another charge \( q_2 \) at a distance \( r \), it again gives rise to a Coulomb force:

\[
F = q_2 E_1 = \frac{q_1 q_2}{4\pi\varepsilon_0 r^2} = \frac{z_1 z_2 e^2}{4\pi\varepsilon_0 r^2}
\]

(22)

In a vacuum, \( F \) depends only on the distance separating the two charged particles and their charge. If some medium is introduced between the charges, the force per unit charge will decrease as a result of the permittivity of the medium, i.e. of its dielectric constant \( \varepsilon \), which is simply a measure of the capacity of the medium to decrease the coulombic force. Thus, Eq. 9 can be rewritten as:

\[
V_{c-c} \cong F = \frac{q_1 q_2}{4\pi\varepsilon r^2}
\]

(23)

and the dielectric constant is likewise incorporated into the other electrostatic force interactions as well.

At this point, it becomes relevant to discuss the dielectric constant of a protein. Not only is it the object of on-going controversy, but it is of definite pertinence for the concepts introduced in the next section (cf. Harvey for a comprehensive discussion [38]). A description of macroscopic and microscopic protein dielectric properties is also given by Simonson and Perahia [39].

2.2.1. Definitions of the dielectric of a protein

It is generally agreed that \( \varepsilon_{prot} \) is rather low, with measurements ranging between 2 and 4 for protein powders and in solution [40–42] and values of 2–4 also calculated with dielectric models incorporating electronic polarizabilities in the range of 1–2 Å. In models based on the Kirkwood–Frohlich (KF) formalism [43], the interaction energy is correlated to the spatial relationship of charged sites to account for protein–solvent interface inhomogeneities [16]. We refer to Gilson and Honig for a full theoretical description of the KF model and derivation of
$\varepsilon \approx 2.5-4$ for a folded protein [44]. Other authors assign much higher values to the dielectric, e.g. 15–20 [20,45] but what should be noted here is that the uncertainty in the value of the dielectric constant of a protein depends first and foremost on the property used to define it [46]. Briefly, various dielectric models account differently for the screening of monopole–monopole interactions and charge solvation inside a protein [38]. For example, the KF model as implemented by Gilson and Sharp [47,48] assumes that the major contributions to the dielectric constant of a polar liquid are the electronic and atomic polarizations (≈ induced polarization) and the dipolar rotation. The first yields dielectric constants of the order of $\sim 1.5$–2.5, the second is a fraction of the refractive index (0.05–0.30) and the dipolar rotation (i.e. the capacity of dipolar groups to align with an electric field) leads to large $\varepsilon$-values for several polar liquids [44]. The various dielectric definitions can thus be grouped under the parameter that they take into account [46]:

(i) Local polarity of the medium: The basis of this definition is solvation, occurring either in polar solvent such as water (large $\varepsilon$), which are very good at stabilizing charged groups, or non-polar solvents such as hydrocarbons (small $\varepsilon$). This treatment yields large $\varepsilon$ values for proteins in the vicinity of charges.

(ii) Macroscopic bulk: This definition considers the protein matrix as a uniform medium and the average value of the dielectric is evaluated by applying weak electric fields to the protein sample, usually in powder form [40]. These measurements yield dielectric constants of the order of $\sim 2$, but as Warshel et al. point out [46], this average protein dielectric is of no use to account for charges located in the protein interior and says nothing about the microscopic charge stabilization processes while predicting that no charges could be stable in the protein interior.

(iii) Microscopic treatments: These approaches take into account the charge heterogeneity of the protein matrix and $\varepsilon$ will depend on the protein region used to define it. The dielectric is no longer considered a constant but rather a spatially varying function which approximates $\approx 40$ in the regions where most ionizable groups are found (e.g. the protein surface) with significant monopole–monopole interactions and $\sim 2$–4 in the protein interior (cf. Fig. 7).

The advantage of this type of ‘dielectric’ definition is that electrostatic effects can be accurately described. For example, by considering the interaction between a monopole and surrounding dipoles, such as N–H groups in ion binding sites, the effect of a single dipole can be evaluated. For the amide N–H bond, $\mu \approx 0.3$ and using Eq. 10, modified to include $\varepsilon$ [46]:

$$
\varepsilon_{\text{eff}}(r) = -332Q\mu\cos \theta/r^2 \Delta G(r)
$$

(24)

a value of $\sim 4$ is obtained for the dielectric which implies that the forces on the protein are weaker than those involved in the monopole–monopole interactions. Similarly, it can be verified that dipole–dipole interactions are even weaker since they yield dielectric values lower than 4. Thus, the weaker and the more effective the force at short range, the lower the effective dielectric. Table 4 lists various dielectric constant definitions and applications for proteins.

<table>
<thead>
<tr>
<th>Table 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>The dielectric constant of a protein as a function of its definition/intended application</td>
</tr>
<tr>
<td>Definition/use of $\varepsilon$</td>
</tr>
<tr>
<td>Polar = large $\varepsilon$; measuring relaxation properties</td>
</tr>
<tr>
<td>Non-polar = small $\varepsilon$; measuring equilibrium properties</td>
</tr>
<tr>
<td>$\varepsilon_{\text{eff}} = 332 \frac{Q_i Q_j}{r^2} \Delta G$</td>
</tr>
<tr>
<td>Large $\varepsilon$ for monopole–monopole interactions</td>
</tr>
<tr>
<td>$1 - 1/\varepsilon_{\text{eff}} = -\Delta G/166Q^2$</td>
</tr>
<tr>
<td>$\varepsilon_{\text{eff}} = -332 \frac{Q_i \mu \cos \theta}{r^2} \Delta G$</td>
</tr>
<tr>
<td>$\varepsilon = 4\varepsilon_0(P + C)/\langle C / \xi \rangle$</td>
</tr>
<tr>
<td>After [6] (p. 280) and [50].</td>
</tr>
</tbody>
</table>

After [6] (p. 280) and [50].
2.2.2. Dependence on equilibrium and relaxation properties

Another important consideration is that the dielectric constant of a medium, just as the polarizability of an atom, is frequency-dependent [49]. For example, water has a substantial dipole moment and is as such very polarizable ($\varepsilon \approx 80$ at 20°C). In a slowly oscillating electric field, the dipoles tend to align with the field but thermal motion counters this effect by causing random reorientation of the molecules. In this process of rotational diffusion, the electric field only exerts a small effect; thus, in a high frequency oscillating field, the individual molecules are unable to reorient fast enough to align with the field. In this case, the high frequency $\varepsilon$ of water is 4 at 20°C (for $E > 200$ GHz) which is significantly different than the low frequency value of 80 (for $E$ up to $\sim 2$ GHz) [38]. For water, there is a broad dispersion at 16 GHz and this must be taken into account in simulations on the ps time scale since the rotational relaxation time of water is ca. 10 ps. In this context, see also the discussion by Israelachvili [7].

The relationship that exists between equilibrium and relaxation properties is also of importance and it is accordingly summarized in this section, essentially following Simonson and colleagues [50,51] who use the dielectric constant in a continuum approach to calculate the equilibrium distribution of charge, field and potential and to determine the relaxation of the protein as it reacts to an external perturbation, such as the presence of a redox electron or of an ion. The determination of equilibrium and relaxation properties usually requires different values for the dielectric constant and the calculation of mixed equilibrium/relaxation properties accordingly necessitate a third approach altogether. Thus, for a protein undergoing relaxation as a result of a pertur-
bation, the free energy of perturbation $A$ can be defined as follows:

$$ A = A_{\text{static}} + A_{\text{relaxation}} $$

$A_{\text{static}}$ is defined by the equilibrium field in and around the protein; it effectively constrains the system and prevents relaxation while $A_{\text{relaxation}}$ is determined by the polarizability of the protein and solvent and thus represents the free energy released upon removal of the constraints. We see that the definition per se of the dielectric constant does not change; it is rather that some types of applications are more dependent on equilibrium properties, such as the calculation of the equilibrium field of an unperturbed protein, well-defined by a force field and starting coordinates, as opposed to other applications more sensitive to the relaxation properties of the protein system as, for example, after an electron transfer or ion binding event which results in charge reorganization in the matrix interior as well as in the solvent (cf. Table 4).

This discussion on the dielectric constant of a protein was intended as a caveat in that the application context of $\varepsilon$ should be recognized as crucial in protein modeling. In practice, this also translates as follows: a high dielectric should be selected for modeling small protein systems when using an implicit solvent approximation. The low dielectric should be restricted to use with explicit solvent representation, provided that the protein is fully solvated (as opposed to using minimal layer solvation). (cf. infra Section 3.1.1).

### 3. Electrostatic modeling in proteins

It follows from the previous sections that the magnitude of the problem associated with modeling the internal electric field of a protein represents one of the most laborious challenges in the field of biophysics with the corollary being no less exacting, i.e. how can the complex energetic contributions of these different interactions be taken into account to model this fluctuating protein internal electric field with reasonable accuracy? In reviewing the treatment of electrostatic effects, Harvey [38] states the general problem associated with macromolecular modeling: if it is to succeed to accurately simulate the physical properties of protein systems, it must incorporate potential energy functions that describe how the potential energy of the protein assembly depends on the position of the constituent atoms while realistically account-

---

Fig. 8. Macroscopic (left) and microscopic (right) protein–solvent models. In the macroscopic case, the protein is surrounded by a continuum solvent region with $\varepsilon=80$ with Glu/Asp shown in red and Arg/Lys in blue. In the microscopic case, the protein is surrounding by explicit water molecules. Cytochrome $b_5$ coordinates (pdb3b5c.ent) [123] from the Brookhaven Protein Data Bank [117].
ing for electrostatic interactions. Briefly, macroscopic approaches use macroscopic electrostatic equations to derive macroscopic properties and the protein is considered as a continuous distribution of charge with interfacial regions of discontinuities [52]. On the other hand, microscopic models explicitly treat the electrostatic interactions and derive macroscopic properties using some form of statistical averaging [53]. We refer to existing reviews for a comprehensive description of these two fundamentally different formalisms and their respective applications [6,8,9,38,46,52,54-60]. In this section, we briefly describe them and summarize two approaches used to conceptualize and represent the macroscopic and microscopic properties of proteins (Fig. 8) and selected simply because of their wide acceptance by the macromolecular modeling community and also because the required programs are available for distribution.

3.1. Macroscopic methods

In macromolecular modeling, methods based on a macroscopic continuum formalism have gained wide acceptance because the macroscopic formulations are readily available and also because they are not very expensive from a computational point of view. The drawback is well-known from a statistical mechanics point of view and it is the familiar one of using a macroscopic scale to describe properties on a microscopic scale. They also require a significant number of totally arbitrary parameters that have to be fixed before carrying out the calculations. More importantly, they involve the use of algorithms which smooth out the essentially heterogeneous nature of the dielectric regions of a protein and ignore the discreteness of the solvent. And so, unless they incorporate some form of microscopic treatment, for example, the explicit incorporation of the effect of local protein polarity or the explicit representation of dipoles such as H-bonds when assigning a low dielectric to the protein [46], they are not successful at realistically describing the electrostatic energies of proteins. However, Sharp and Honig [61] addressed the question of the validity of classical electrostatics for the accurate description of microscopic protein properties and demonstrated that its application to appropriate problems yielded a rigorous approach to a wide range of microscopic phenomena. This view has been corroborated in recent years by the success of macroscopic formalism at treating solute-solvent interactions, such as the structural origins of pH effects on protein stability [62,63] and substrate binding [64], the effect of solvation on protein stability and binding [65] and the directionality of electron transfer [66], to mention but a few applications.

In general, macroscopic treatments use two different approaches for calculating electrostatics and they are based on the use of different models to represent the charge distribution within the protein. The most widely applied are the Tanford–Kirkwood (TK) and modified TK methods [67,68], in which the protein is represented as a sphere of low dielectric; the modified TK formalism corresponds to a polar protein interior around charged groups, consistent with the experimental finding of ions inside proteins, and adjusts monopole–monopole interactions by the corresponding solvent accessibilities. The validity and applicability of these models is discussed by Warshel et al. [52]. The other widely used macroscopic approach is the Finite Difference Poisson–Boltzmann (FDPB) method, which uses a Cartesian grid system to map the protein and treats ionic strength effects as well [69].

3.1.1. Finite difference methods

Finite Difference methods [47,52,55,57,70,71] calculate electrostatic interactions with the method of finite differences applied to a continuum model for the medium surrounding the protein rather than a collection of atoms (Fig. 8). They are attractive because they are rigorously derived from classical continuum electrostatic theory. First applied to protein systems by Warwicker and Watson [72], finite difference methods describe the solvated protein as a discretized continuum, mapping the whole space, i.e. protein and surrounding solvent continuum, onto a three-dimensional grid with local dielectric values (εprot and εsolv, respectively) assigned to each grid point. This is because, in the presence of an electric field, three charge-screening processes must be taken into account, namely: the electronic polarizability, which describes the reorientation of the electron cloud around a nucleus upon application of the field, the reorientation of the permanent dipoles and the subsequent redistribution of charge. Electronic polarizability and dipole reorientation are taken into ac-
count by using a position-dependent $\epsilon$ and a finite difference method is then used to solve the Poisson equation combined to the Coulomb equation by numerical approximation methods [73]:

$$\nabla^2 \phi(r) = -4\pi \rho(r)/\epsilon(r)$$  \hspace{1cm} (26)

in which $\phi(r)$ is the electrostatic potential with position $r$ and $\rho(r)$ is the charge density distribution. Since proteins are often solvated in the presence of mobile ions, consideration of ionic strength represents an added modeling advantage and this leads to the Finite Difference Poisson–Boltzmann (FDPB) equation [3,61], fully implemented as Delphi in the InsightII software package [69]:

$$\nabla [\epsilon(r) \phi(r)] - \epsilon(r) \kappa^2(r)^2 \sinh[\phi(r)] + 4\pi \rho^f(r)/kT = 0$$  \hspace{1cm} (27)

in which $\rho^f$ is the fixed charge density in proton charge units, $k$ the Boltzmann constant, $T$ the absolute temperature, $\phi(r)$ the dimensionless electrostatic potential expressed in units of $kT/q$, $q$ being the charge on a proton and $\kappa^2 = \lambda^2 = 8\pi q^2 Ile kT$, where $\lambda$ is the Debye length and $I$ is the ionic strength of the solvent. $\phi, \epsilon, \kappa, \rho^f$ are all functions of the position vector $r$. As with the finite difference method, a local dielectric is assigned to each of the lines connecting the grid points which are also given appropriate Debye–Hückel values. The choice of dielectric depends on the properties being studied (cf. supra Section 2.2). Briefly, a high protein $\epsilon$ takes into account the polarizability of the medium and accordingly, it is adequate for calculating relaxation properties, such as relaxation free energies in response to a perturbing charge. If, however, the focus of interest is on the equilibrium field and potential of the protein matrix, a low $\epsilon$ would describe more accurately the equilibrium charge distribution.

The trade-off between fine grids (more accurate) and coarse ones (fast computing speed) can be balanced by using focusing methods that implement finer grids in protein regions of interest [47]. Thus, electrostatic models based on the FDPB formalism can incorporate detailed structural information, along with charge and dielectric properties, solvent and ionic strength effects. As such, they provide an essentially complete description of the electrostatics of a given conformation or structure of a protein. The FDPB approach provides a good trade-off between realistic description of the electrostatics with computational efficiency, and has been used successfully to model many protein electrostatic properties, including redox potentials, pKa values, association rates, binding constants and catalysis (cf. Sharp and Honig [61] for a full review of applications).

In our context of interest, applications of the FDPB method include the study of protein polar contributions to the frequency shifts of small ligands [74], the capacity of the charge distribution in a protein matrix to affect the CO ligand/heme geometry [75], the study of the effect of the $\alpha$-helix electrostatic field on electronic spectra [76] and the correlation of changes in the population distribution of the 0,0 transition frequencies of mesoporphyrin in cytochrome $c$ peroxidase with changes in electrostatic field at the heme due to protonation of the axial histidines [77]. This work showed that protonation of the axial histidines produces a large change in the potential at the heme, as well as changes in the electric field. When the pH is lowered, protonating the histidine, a new species appears in the population distribution of the 0,0 transition frequencies of the peroxidase. The prediction that a change in the spectral energy distributions correlates with the electric field is thus qualitatively borne out. These results provide strong support to the idea that similar calculations can be applied quantitatively to study the contribution of the protein matrix to the electric field in more complex protein systems.

3.2. Microscopic methods

The development of these methods for application to protein systems can be traced back to the early work of Warshel [6] motivated by the inadequacy of macroscopic approaches to solve the dielectric problem and by the desire to consider the discrete nature of the solvated protein. The obvious advantage is theoretical and it is that they require no arbitrary parameters (besides an effective $\epsilon$) because they treat the system at the atomic level. The modeled effects can also be directly correlated with microscopic properties and the only requirement is to account for all relevant energy contributions. However, their computational complexity prevents them from widespread application since an explicit representation of
all atoms in the protein and the solvent is required as well as the description of the orientation of all dipoles. Thus, a thing of beauty, but that requires hard work.

3.2.1. The protein dipole Langevin dipole method

The protein dipoles Langevin dipoles method (PDLD), developed by Warshel and co-workers [6,46], describes the average energy of a protein system using the electrostatic energy derived from its average structure, usually its crystallographic structure, itself a time-averaged structure. Most modeling procedures re-minimize the starting structure for the different force fields used to describe the charge distributions. Treatment of the solvent polarization is explicit, but can be simplified through use of a 3-D grid model in which the water molecules are represented by point dipoles that can also penetrate protein cavities. The polarization of a solvent dipole in the local electric field is described by a Langevin approximation [78]:

\[ \mu = e_i \mu_0 (\coth \chi_i - 1/\chi_i) \]  

where:

\[ \chi_i = C \mu_0 (\zeta_i - \zeta_i^e) \zeta_i^o / \zeta_i^o k_b T \]  

in which \( e_i \) is a unit vector in the direction of the local electric field \( \zeta_i \) on the \( i \)th dipole, \( \mu_0 = 1.8 \) D, \( C \) is a solvation energy parameter, \( \zeta_i^o \) is the electric field from the charge distribution on the \( i \)th dipole, and \( \zeta_i^e \) is the field from its nearest neighbors. The solvent has a uniform polarizability, or dielectric response, represented by discrete polarizabilities that follow a polarization law, which corresponds to a thermal average of the orientations and spatial distribution of the permanent and induced dipoles and which is obtained by calibrating the parameter \( C \) and the van der Waals distances according to atom type with the object of reproducing realistic solvation energies for groups with different charges and radii. The effect of the protein’s electronic polarization is also explicitly treated by assigning induced dipole moments to all atoms. An iterative procedure is then used to treat charge-dipole and dipole-dipole interactions. A drawback is that the procedure iterates on the magnitude of the Langevin dipole, but not on its direction and also that ionic strength effects are not treated [38].

The PDLD formalism is available for distribution (as Polaris in the Molaris software package [79]) and has been applied to the study of monopole-monopole interactions in proteins [6], and has tackled the problem of ‘self-energies’ i.e. the solvation of charge in the protein matrix. This self-energy is important because it provides a measure of the capacity of H-bonds to stabilize a charged group. It also addresses such questions as how the polarity of the protein environment can contribute to the control of heme protein redox potentials [80], and how the electrostatic control of photosynthetic processes occurs. For example, it is known that one of two symmetric electron pathways is blocked in the reaction center and the PDLD method was used to correlate this effect with the different polarities of the sites associated with the accessory chlorophyll present in the two branches of the electron pathway [81].

3.3. Combined approaches

Electrostatic modeling approaches are being constantly refined as a result of the ever-increasing computational capacity and as a consequence of theoretical and conceptual advances. Progress towards achieving the goal of the accurate representation of the internal field of a protein, including all electrostatic factors, is being achieved by the increasing sophistication of both macroscopic and microscopic approaches. For example, a semi-microscopic approach was recently described [46,82–84] which scales the PDLD method to calculate the energy of the charges in a solvated protein: an energy term is obtained from the interaction of the protein with its charged groups and with the surrounding Langevin dipoles by scaling the dielectric factors. The model yields very accurate results for surface groups using \( e_m \sim 6 \). As for the continuum formalism, Nakamura [85] describes a combination of the approach with molecular dynamics in which only the pure solvent region is represented as a continuum, the protein being described by molecular theory. Sharp [48] has also developed a computationally efficient means of incorporating continuum electrostatic forces in molecular dynamics calculations, in which he also treats the protein with molecular theory. See Nakamura [1] and Gilson et al. [86] for a detailed discussion.
Stark spectroscopy is usually understood to refer to the application of an external electric field intended as a perturbation of the spectral characteristics of an optical center [87,88]. In the optical domain, the Stark effect (SE) has a long history of being used for the identification of degenerate states, the determination of molecular orientations, or of the number of molecules per primitive unit cell in X-ray crystallography and for distinguishing between polar and non-polar isomers in solids [89]. As such, the SE-induced spectral changes can be related to significant parameters of a given charge distribution in the molecular system under investigation. Relevant to our context are the SE applications reporting on the measurement of microscopic electric fields in condensed phases induced by polarization or the heterogeneity of charge distributions [90].

In protein systems, Stark spectroscopy is used to probe the interaction of chromophores with the protein matrix in which they reside and we refer to the companion reviews in this special issue for a full discussion of these interactions involving the application of external electric fields [13,14] as well as to a recent review on Stark spectroscopy methods and applications [91]. In this section, we briefly define the Stark effect and proceed with discussing recent electronic and vibrational experimental studies focused on the SE observed, not as a result of the application of an external field, but rather attributed to changes in the internal field of a protein matrix.

4. Stark spectroscopy

Stark spectroscopy is usually understood to refer to the application of an external electric field intended as a perturbation of the spectral characteristics of an optical center [87,88]. In the optical domain, the Stark effect (SE) has a long history of being used for the identification of degenerate states, the determination of molecular orientations, or of the number of molecules per primitive unit cell in X-ray crystallography and for distinguishing between polar and non-polar isomers in solids [89]. As such, the SE-induced spectral changes can be related to significant parameters of a given charge distribution in the molecular system under investigation. Relevant to our context are the SE applications reporting on the measurement of microscopic electric fields in condensed phases induced by polarization or the heterogeneity of charge distributions [90].

In protein systems, Stark spectroscopy is used to probe the interaction of chromophores with the protein matrix in which they reside and we refer to the companion reviews in this special issue for a full discussion of these interactions involving the application of external electric fields [13,14] as well as to a recent review on Stark spectroscopy methods and applications [91]. In this section, we briefly define the Stark effect and proceed with discussing recent electronic and vibrational experimental studies focused on the SE observed, not as a result of the application of an external field, but rather attributed to changes in the internal field of a protein matrix.

4.1. The electronic and vibrational Stark effects

The effect of an electric field on molecular spectra, or Stark effect, can be taken into account simply by adding the following term to the optical center’s Hamiltonian [90]:

$$H = -\mu \cdot E$$

(30)

where $\mu$ is the dipole moment and $E$ the electric field applied. In cases where the field is non-uniform, additional multipole moments must be considered [92]. Briefly, SE’s are observed in optical spectra as line shifts, splittings or broadenings (i.e. unresolved splittings) due to the interaction of the electric field with the dipole moment of the optical center (first order SE) or with the field-induced dipole moments (second order SE). Intensity changes are also observed as a result of field-induced mixing of states with different transition moments. Estimates of the magnitude of electric fields capable of inducing Stark shifts in electronic absorption spectra range between $4 \times 10^6$ V/cm in the case of a simple helical peptide and $5 \times 10^6$ V/cm for a transient dipole induced by electron transfer in photosynthetic reaction centers [93].

In protein systems, Stark spectroscopy is used to probe the interaction of chromophores with the protein matrix in which they reside and we refer to the companion reviews in this special issue for a full discussion of these interactions involving the application of external electric fields [13,14] as well as to a recent review on Stark spectroscopy methods and applications [91]. In this section, we briefly define the Stark effect and proceed with discussing recent electronic and vibrational experimental studies focused on the SE observed, not as a result of the application of an external field, but rather attributed to changes in the internal field of a protein matrix.

4.1. The electronic and vibrational Stark effects

The effect of an electric field on molecular spectra, or Stark effect, can be taken into account simply by adding the following term to the optical center’s Hamiltonian [90]:

$$H = -\mu \cdot E$$

(30)

where $\mu$ is the dipole moment and $E$ the electric field applied. In cases where the field is non-uniform, additional multipole moments must be considered [92]. Briefly, SE’s are observed in optical spectra as line shifts, splittings or broadenings (i.e. unresolved splittings) due to the interaction of the electric field with the dipole moment of the optical center (first order SE) or with the field-induced dipole moments (second order SE). Intensity changes are also observed as a result of field-induced mixing of states with different transition moments. Estimates of the magnitude of electric fields capable of inducing Stark shifts in electronic absorption spectra range between $4 \times 10^6$ V/cm in the case of a simple helical peptide and $5 \times 10^6$ V/cm for a transient dipole induced by electron transfer in photosynthetic reaction centers [93].

Similarly, among the perturbations that can affect the vibrational frequency of an oscillator, the vibrational Stark effect (VSE) (Fig. 9), resulting from the application of an external electric field is perhaps the best known. The VSE can be attributed to a perturbation of the vibrational Hamiltonian through its projection on the dipole vector of the vibrational mode which can thus significantly alter both the frequency and oscillator strength of the transition. The perturbed Hamiltonian can be expressed as follows [74]:

$$H = (p^2/2m) - \mu(Q) \cdot E + \Phi(Q)$$

(31)

where $Q$ is the generalized coordinate of the normal mode, $p$ the conjugate momentum, $m$ the reduced mass, $\mu$ the dipole moment, $E$ the applied electric field and $\Phi(Q)$ the potential in the absence of electric field.

The energy of an ideal dipole $\mu$ in a uniform electric field is given by the following dot product:
If the field is applied during a vibrational transition, the energies of the initial and final vibrational states will be shifted by $-\Delta \mu$, i.e. the difference between the dipole moments of both states. At low applied field, the VSE is linear and proportional to the Stark tuning rate, defined as [94]:

$$\delta v_E = \frac{d\hbar v_{01}}{dE}$$

where $d\hbar v_{01}$ is the difference in energy between the $v=0$ and $v=1$ vibrational states and $\hbar$ is Planck’s constant. Different VSE perturbation mechanisms have been proposed but whether classical or quantum mechanical [95], these approaches have all shown that the VSE can account for changes which are, as in the case of the electronic SE, also experimentally observed as line shifts, band splittings and intensity variations.

In a biological context, very little vibrational Stark work has been reported: a VSE was computed at the semi-empirical level (AM1) for the retinal molecule [95] and both the frequency and oscillator strength were shown to be significantly sensitive to fields of the order of 0.005 a.u. In a study performed on the reaction center of a photosynthetic pigment, intensity
increases of the amide stretch were attributed to a VSE and semi-empirical calculations showed that vibrational Stark shifts of the order of 2 cm$^{-1}$ could be expected [96]. Testing the role of electrostatic interactions in determining the CO-stretch frequency in carbonmonoxymyoglobin, Decatur and Boxer [93] calculated that a shift in the CO-stretch frequency of 10 cm$^{-1}$ would require an electric field of some $2 \times 10^7$ V/cm.

4.2. Experimental survey

The first study on the effect of a protein-induced electric field on the optical properties of a biological chromophore reported on the binding of mesoporphyrin-substituted cytochrome $c$ peroxidase (MP-CcP) and cytochrome $c$ [77]. The optical center of interest was a mesoporphyrin and its emission was measured by fluorescence line narrowing in the bound and unbound states as a function of pH and ionic strength (cf. Section 3.1.1 in the companion paper [15]). A FDPB approach was selected to calculate the electric potential and field at the pyrrole nitrogens of the porphyrin macrocycle, also as a function of ionic strength. The rationale for performing the FDPB calculation at the nitrogens was that any observed spectral change in the spectriscopically monitored Q-band region would stem from the perturbation of porphyrin transitions known to be restricted to the delocalized $\pi$ system of the tetrapyrrole [97]. Accordingly, the spectral changes could be correlated to the potentials measured at this location. Fig. 10 shows the distribution of electronic energies extracted from the FLN spectra.

Inspection of the IDF$s$ clearly shows that the 0,0 bands are different at neutral and alkaline pH, with three and two components respectively. This translates into a spectral shift of 60–100 cm$^{-1}$ among the three components that widen to 145 cm$^{-1}$ with pH. The disappearance of one of the components at higher pH was interpreted to be a result of the protonation of His$^{52}$, which was supported by the electrostatic calculations, yielding increased potentials at the porphyrin nitrogens by as much as 200 mV/A when the calculations was performed with this distal histidine charged, especially along the normal to the heme, i.e. along the $\pi \rightarrow \pi^*$ transition axis. The difference in the overall calculated heme electrostatic potentials, as a result of charging the histidine or not, is illustrated in Fig. 11.

Thus, the spectral changes shown in Fig. 10 can be correlated to the different fields experienced at the heme as a result of charge redistribution in the protein matrix. The effect of dispersion forces were not taken into account in the calculations, but they are known to significantly affect electric field fluctuations and thus, must contribute to the width of the determined IDF components. The authors also compared the same chromophore (i.e. mesoporphyrin) in two enzymes, i.e. CeP and horseradish peroxidase, reporting that the first excited singlet vibrational frequencies differed by some 6 cm$^{-1}$ (cf. Table 1 in [77]). The recent publication of the HRP X-ray structure [98] should allow further testing of the protein field/spectral property correlation in the near future.

In the VSE context, the study of various carbonmonoxy heme proteins is confirming the importance of the protein matrix field with respect to significantly affecting vibrational stretch frequencies. That charge and polar interactions in the heme ligand pocket are critical factors modulating the binding of CO has been recognized for some time, both experimentally, in IR studies of mutant proteins [99–101], and theoretically [75,102–105]. In the case of myoglobin, the use of distal mutants as local modulators of pocket field interactions has resulted in observable spectral shifts suggested to arise from internal Stark effects due to the interaction of the electric field of the protein and the change in the CO dipole moment [93]. Electrostatic potential and electric field calculations were also recently performed at the CO-ligand of the heme in both horse heart and yeast cytochrome $c$ using the FDPB formalism [75,106]. The calculations supported recent experimental and theoretical evidence suggesting that polar interactions could significantly affect the ligand’s vibrational frequency. The work showed that the ionizable amino acid residues and polar contributions of the protein matrix could induce a Stark effect on the CO-stretch frequency of the carbonmonoxy cytochromes. The observed CO-stretches, at 1965.9 and 1960.1 cm$^{-1}$ for horse and yeast cytochrome, respectively, show that the experimental shift is of the order of 6 cm$^{-1}$. This is in good agreement with a calculated value of 8 cm$^{-1}$ ($\pm 1$ cm$^{-1}$) for a VSE shift due to
the different charge distributions in both cytochromes.

The effect of the orientation of CO bound to the heme of two \(c\)-type cytochromes on the CO-stretch frequency \(v_{\text{co}}\) was also investigated using molecular mechanics and \(\text{FDPB}\) calculations [75]. Results showed that modifying the Fe–C–O bending angle did not change the CO-stretch frequency within a range of 100–175°, equivalent to a bending motion away from the propionic acid chains. The calculated Stark shifts ranged from 5.4 to 8.6 cm\(^{-1}\) and were in good agreement with the experimentally observed shift (6 cm\(^{-1}\)). However, motion of the CO towards the propionics was found to exert significant influence on the calculated shifts (−4.5 to −1.8 cm\(^{-1}\)) which were then in total disagreement with experiment, not only in magnitude, but also in predicting the wrong direction for the shift. The Stark shifts calculated for the tilt angle showed that it had no significant effect on \(v_{\text{co}}\) within a 15° distortion range. With respect to the proximal histidine displacement angle, only a complete (and unlikely) 90° rotation about the Fe–N bond was effective in significantly affecting the Stark shift and a major contribution to the CO distortion from the coordination geometry of the proximal histidine was accordingly ruled out. Overall, the calculations showed that the CO ligand thus enjoys a significant amount of flexibility in the heme pocket, as required to approach and leave the heme group.

More recent work [107,108] reports on the vibrational spectra of the asymmetric carboxylate region of three \(c\)-type cytochromes combined to electrostatic calculations selected to target specific protein matrix regions liable to yield pertinent protein matrix charge information that would correlate with the observed spectroscopy. Deconvolution of the IR carboxylate stretch frequency yielded different \(v_{\text{coo}}\) distributions among the three proteins investigated and the frequency distribution of the calculated electrostatic potentials were found to correlate with the experimental observations.

5. Concluding remarks

The purpose of this review was to outline the non-covalent interactions that are so important for protein conformational rearrangements and to survey the few available studies attempting to correlate the observed vibrational and electronic spectroscopy of biological chromophores as perturbed by charge redistribution effects in protein matrices. To achieve reasonably accurate qualitative correlations, two additional considerations need to be addressed.

First, in modeling the protein system, the molecular mechanics approach needs an adequate representation of the non-covalent charge interactions. Also, the very nature of geometry optimization, i.e. the fact that it is an iterative process, means that the minima achieved are necessarily questionable, i.e. extracting the lowest energy conformations from the conformational space sampled requires normal mode coordinate analysis, a not so simple task for complex biological systems.

The second consideration is of course the problem of force field parametization: most force fields have been derived for organic molecules and have to be re-parametized especially in the case of metalloprotein systems. The pitfalls are challenging and far from trivial. To list but a few: the derivation of force constants from IR spectra is difficult for metal–ligand vibrations because most of them occur in the low frequency domain which is also the hardest region to work in from an experimental point of view; reliable partial charges for metal ions and for the nearest neighbors are difficult to obtain and also, there is the question of transferability of parameters which becomes a serious scale problem, in that parameters derived quantum mechanically are usually not applicable to classical force fields. For a comprehensive discussion of these issues, we refer to the available reviews [21,109–114].

Last but not least, there is the question of the accurate semi-empirical methods required to analyze the electronic properties of biological optical centers. It is a complex undertaking: protein systems are enormous from a computational efficiency point of view, for example, the average porphyrin optical center consists of several hundred atoms, and reproducing the major features of experimentally acquired electronic spectra at an acceptable level of accuracy requires treatment of the protein embedding the optical center as a polarizable medium. Encouraging results have recently been reported for the photosynthetic reaction center using a semi-empirical ap-
proach [115] and it is to be hoped that the approach will soon be verified and applied to other biological chromophores.

Acknowledgements

Supported by GM-PO1-48130, GM-RO1-55004 and NSF INT95-13315.

References

[36] S.C. Harvey, Treatment of electrostatic effects in macromo-
[75] M. Laberge, K.A. Sharp, J.M. Vanderkooi, Protein electric...
field effects on the CO-stretch frequency of carbonmonoxycytochromes c as a function of carbonyl tilting and bending investigated with a continuum electrostatic approach, J. Phys.


