22.3. Electrostatic interactions in proteins

BY K. A. SHARP

22.3.1. Introduction

Electrostatic interactions play a key role in determining the structure, stability, binding affinity, chemical properties, and hence the biological reactivity, of proteins and nucleic acids. Interactions where electrostatics play an important role include:

(1) Ligand/substrate association. Long-range electrostatic forces can considerably enhance association rates by facilitating translational and rotational diffusion or by reduction in the dimensionality of the diffusion space.

(2) Binding affinity. Tight specific binding is often a prerequisite for biological activity, and electrostatics make important contributions to desolvation and formation of chemically complementary interactions during binding.

(3) Modification of chemical and physical properties of functional groups such as cofactors (haems, metal ions *etc.*), alteration of the ionization energy (pK_a) of side chains and shifting of redox midpoints.

(4) The creation of potentials or fields in the active sites to stabilize functionally important charged or dipolar intermediates in processes such as catalysis.

In this chapter I will discuss, within the framework of classical electrostatics, how such effects can be modelled starting from the structural information provided by X-ray crystallography. Nevertheless, many of the concepts of classical electrostatics can be used in combination with molecular dynamics (MD), quantum mechanics (QM) and other computational methods to study a wider range of macromolecular properties, for example specific protein motions, the breaking or forming of bonds, determination of intrinsic pK_a's, determination of electronic energy levels *etc*.

The central aim in studying the electrostatic properties of macromolecules is to take the structural information provided by crystallography (typically the atomic coordinates, although *B*-factor information may also be of use) and obtain a realistic description of the electrostatic potential distribution $\varphi(\mathbf{r})$. The electrostatic potential distribution can then be used in a variety of ways: (i) graphical analysis may reveal deeper aspects of the structure and help identify functionally important regions or active sites; (ii) the potentials may be used to calculate energies and forces, which can then be used to calculate equilibrium or kinetic properties; and (iii) the electrostatic potentials may be used in conjunction with other computational methods such as QM and MD.

Three problems must be solved to obtain the electrostatic potential distribution. The first is to model the macromolecular charge distribution, usually by specifying the location and charge of all its atoms. Although the coordinates of the molecule are determined by crystallographic methods, the charge distribution is not. A number of atomic charge distributions have been developed for proteins and nucleic acids using quantum mechanical methods and/or parameterization to different experimental data. The second problem is that the positions of the water molecules and solvent ions are generally not known. (Water molecules and ions seen in even the best crystal structures usually constitute a small fraction of the total important in solvating the molecule. Moreover, the orientation of the crystallographic water molecules, crucial in determining the electrostatic potential, is rarely known.) Within the framework of classical electrostatics, inclusion of the *effect* of the solvating water molecules and ions is handled not by treating them explicitly, but implicitly in terms of an 'electrostatic response' to the field created by the molecular charge distribution. The third problem is that incorporation of the available structural information at atomic resolution results in a complicated spatial distribution of charge, dielectric response *etc*. Numerical methods for rapidly and accurately solving the electrostatic equations that determine the potential are therefore essential.

22.3.2. Theory

22.3.2.1. *The response of the system to electrostatic fields*

The response to the electrostatic field arising from the molecular charge distribution arises from three physical processes: electronic polarization, reorientation of permanent dipolar groups and redistribution of mobile ions in the solvent. Movement of ionized side chains, if significant, is sometimes viewed as part of the dielectric response of the protein, and sometimes explicitly as a conformational change of the molecule.

Electronic polarizability can be represented either by point inducible dipoles (Warshel & Åqvist, 1991) or by a dielectric constant. The latter approach relates the electrostatic polarization, $\mathbf{P}(\mathbf{r})$ (the mean dipole moment induced in some small volume V) to the Maxwell (total) field, $\mathbf{E}(\mathbf{r})$, and the local dielectric constant representing the response of that volume, $\varepsilon(\mathbf{r})$, according to

$$\mathbf{P}(\mathbf{r}) = [\varepsilon(\mathbf{r}) - 1]\mathbf{E}(\mathbf{r})/4\pi. \qquad (22.3.2.1)$$

The contribution of electronic polarizability to the dielectric constant of most organic material and water is fairly similar. It can be evaluated by high-frequency dielectric measurements or the refractive index, and it is in the range 2–2.5.

The reorientation of groups such as the peptide bond or surrounding water molecules which have large permanent dipoles is an important part of the overall response. This response too may be treated using a dielectric constant, *i.e.* using equation (22.3.2.1) with a larger value of the dielectric constant that incorporates the additional polarization from dipole reorientation. An alternative approach to equation (22.3.2.1) for treating the dipole reorientation contribution of water surrounding the macromolecules is the Langevin dipole model (Lee *et al.*, 1993; Warshel & Åqvist, 1991; Warshel & Russell, 1984). Four factors determine the degree of response from permanent dipoles: (i) the dipole-moment magnitude; (ii) the density of such groups in the protein or solvent; (iii) the freedom of such groups to reorient; and (iv) the degree of cooperativity between dipole motions. Thus, water has a high dielectric constant ($\varepsilon = 78.6$ at 25 °C). For electrostatic models based on dielectric theory, the experimental solvent dielectric constant, reflecting the contribution of electronic polarizability and dipole reorientation, is usually used. From consideration of the four factors that determine the dielectric response, macromolecules would be expected to have a much lower dielectric constant than the solvent. Indeed, theoretical studies of the dielectric behaviour of amorphous protein solids (Gilson & Honig, 1986; Nakamura *et al.*, 1988) and the interior of proteins in solution (Simonson & Brooks, 1996; Simonson & Perahia, 1995; Smith *et al.*, 1993), and experimental measurements (Takashima & Schwan, 1965) provide an estimate of $\varepsilon = 2.5$–4 for the contribution of dipolar groups to the protein dielectric.

The Langevin model can account for the saturation of the response at high fields that occurs if the dipoles become highly aligned with the field. The dielectric model can also be extended to incorporate saturation effects (Warwicker, 1994), although there is a compensating effect of electrostriction, which increases the local dipole density (Jayaram, Fine *et al.*, 1989). While the importance of saturation effects would vary from case to case, linear solvent dielectric models have proven sufficiently accurate for most protein applications to date.

references