2. Theoretical Background

2.1 Overview of the computational procedures

The final goal in our computation is to obtain the pK_a for titratable groups or the redox potential E_m for redox-active groups in protein, and to elucidate redox-active reactions, ET or PT whose details still remain unknown. Our research often covers reaction mechanisms of physiologically importance.

2.1.1. Atomic coordinates

The basis of our computation is the solution of the linearized Poisson-Boltzmann (LPB) equation (see **2.3**) based on atomic coordinates of the protein crystal structures. The computed results are sensitively and faithfully determined by using the original atomic coordinates of the crystal structures. For this purpose, crystal structures at higher resolutions (preferentially, higher than ~3.0 Å) are suitable.

2.1.2. Generation of hydrogen atoms

Hydrogen atom positions are energetically optimized with CHARMM (Brooks et al., 1983). During this procedure the positions of all non-hydrogen atoms are fixed, and all titratable groups are kept in their standard protonation states i.e. acidic groups ionized and basic groups (including titratable histidines) protonated. Residues that are ligands of cofactors (e.g. His for heme/Chl*a* or Glu for the Fe-complex) are treated as non-titratable residues. Hereby, the redox state of the redox-active group should be treated with great care, because its different redox state often results in a different H-bond pattern for the H-bond partner or residues nearby (see **3.1**, **3.2**, **5.1**, and **10.1**).

2.1.3. Variable charges of titratable groups

Atomic partial charges of the amino acids were adopted from the all-atom CHARMM22 (MacKerell et al., 1998) parameter set. For carboxyl groups of Asp and Glu, the charges of the two oxygens are increased symmetrically by +0.5 unit charges to account implicitly for the presence of the proton, instead of attaching an explicit hydrogen atom on one of the oxygens in their protonated states. Similarly, upon the deprotonation of Arg and Lys, the charges of all protons at the corresponding basic group in their protonated states are diminished symmetrically by a total unit charge.

Except for a few limited cases, the atomic charges of (redox-active) cofactors or pigments are not available in the CHARMM22 parameter set. In such cases, they are determined by first calculating the electronic wave functions and then fitting the resulting electrostatic potential in the neighborhood of these molecules by the RESP procedure (Bayly et al., 1993). The electronic wave functions are often calculated with the DFT module in JAGUAR (Jaguar4.2, 1991-2000) using the B3LYP functional with LACVP basis set (6-31G with effective core potentials on heavy atoms).

2.1.4. Computation of protonation pattern, pK_a and E_m

In our computation, pK_a for titratable groups or E_m for redox-active groups in protein are obtained as the sum "the pK_a/E_m in the model system (i.e. reference value of pK_a/E_m) + the difference between the model system and protein ΔpK_a/ΔE_m".

The computation of the energetics of the protonation pattern is based on the electrostatic continuum model, in which the LPB equation is solved by the program MEAD from Bashford and Karplus (Bashford and Karplus, 1990). To sample the ensemble of protonation patterns by a Monte Carlo (MC) method, we use our own

program Karlsberg (Rabenstein, 1999).

The dielectric constant is set to $\varepsilon_P = 4$ inside the protein and $\varepsilon_W = 80$ for solvent and protein cavities corresponding to water. All computations are performed at 300 K, pH 7.0 and an ionic strength of 100 mM. The LPB equation is solved by a three-step grid-focusing procedure with a starting grid resolution of 2.5 Å, an intermediate grid resolution of 1.0 Å, and a final grid resolution of 0.3 Å.

The dielectric volume of a protein complex is the spatial region covered by molecular components of the protein that are polypeptide backbone, side chains and cofactors, but not water molecules. To facilitate direct comparison with our previous computational results, we uniformly use the same computational conditions and parameters such as atomic partial charges and dielectric constants. In general, all crystal waters are removed, because of lack of experimental information on hydrogen atom positions. Cavities resulting after removal of crystal water are uniformly filled with solvent dielectric of $\varepsilon_W = 80$. Accordingly, the effect of the removed water molecules is implicitly accounted for by the high value of the dielectric constant in these cavities.

2.2. Force field

The force field is generated by a set of energy function consisting of individual energy terms. In the CHARMM energy function (Brooks et al., 1983), the total energy of the system can be computed from a summation of the internal and external interaction terms.

$$E = E_{internal} + E_{external} \qquad \text{(Eq. 2-1)}$$

$$E_{\text{int}\,ernal} = \sum_{bonds} k_b (r - r_0)^2 + \sum_{angles} k_\theta (\theta - \theta_9)^2 + \sum_{dihedrals} k_\phi (1 + \cos(n\phi - \delta)) + \sum_{impropers} k_l (l - l_0)^2 \qquad \text{(Eq. 2-2)}$$

$$E_{external} = \sum_{electrostatic.i,j>i} \left(\frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \right) + \sum_{van\_der\_Waals.i,j>i} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right) \qquad \text{(Eq. 2-3)}$$

The internal energy term includes the bond, bond angle, dihedral angle and improper torsion potentials. The improper torsional potential has been designed both to maintain the chirality (e.g. asymmetric carbon such as α-carbon of the protein backbone) and to maintain planarity (e.g. sp^2 hybridized atoms such as double bond carbons). On the other hand, the external (non-bonded) energy term includes the electrostatic, van der Waals and H-bond interactions. The electrostatic potential is computed based on the atomic partial charges distributed to the atomic coordinates. Electrostatic interaction is effective over large distances. The van der Waals energy is often approximated by the Lennard-Jones potential (the 6-12 potential, the second term in the right side of eq. 2-3). A smoothing technique by a switching function that sets a cutoff distance is effective to reduce the number of pair interactions, without suffering from the discontinuity of energy conservation during the energy minimization (or molecular dynamics) that may arise from a mere usage of a cutoff distance.

In the present study, the force field parameter mainly contribute to the energy minimization process to determine the geometry of hydrogen atomic coordinates, which are generated based on the CHARMM22 parameter set (MacKerell et al., 1998).

2.3. Poisson-Boltzmann equation

The electrostatic potential of $\phi(\mathbf{r})$ generated by the charge density $\rho(\mathbf{r})$ is described by the Poisson equation for a homogeneous dielectric medium as

$$\nabla^2\phi(\mathbf{r}) = -4\pi \frac{\rho(\mathbf{r})}{\varepsilon} \qquad \text{(Eq. 2-4)}$$

where ε is the dielectric constant in the region ($\mathbf{r}$ denotes vector).
If ε is a function of r, Eq. 2-4 is written as

$$\nabla(\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})) = -4\pi\rho(\mathbf{r}) \qquad \text{(Eq. 2-5)}$$

On the other hand, if the system is solvated in a solvent with ions (i.e. forming the ionic solution), the charge density of the ions $\rho_{ion}(\mathbf{r})$ in the solvent can be described as a summation of the distribution of all kinds of ionic charges in the solution, according to the Debye-Hückel theory. Hereby, the ion distribution is subjected to the same potential $\phi(\mathbf{r})$:

$$\rho_{ion}(\mathbf{r}) = \Sigma\, c(\mathbf{r})\, q_i \exp(-\beta q_i \phi(\mathbf{r})) \qquad \text{(Eq. 2-6)}$$

where $c(\mathbf{r})$ is the concentration of an ionic species in solution, q is the charge of the ionic species; and $\beta = 1/(kT)$ with the Boltzmann constant k.

Since $\exp(-\beta q_i\phi(\mathbf{r})) \approx 1 - \beta q_i\, \phi(\mathbf{r})$ (based on the Taylor expansion), Eq. 2-6 can be approximated to

$$\rho_{ion}(\mathbf{r}) \approx \Sigma\, c(\mathbf{r})\, q_i - \beta \Sigma\, c(\mathbf{r})\, q_i^2 \approx -\beta \Sigma\, c(\mathbf{r})\, q_i^2 \qquad \text{(Eq. 2-7)}$$

The second part of the approximation in Eq. 2-7 is that the Debye-Hückel theory assumes charge neutrality in the solvent i.e., $\Sigma\, c(\mathbf{r})\, q_i = 0$.

Thus, if the charge density of the solvation is taken into account, Eq. 2-5 is extended to the linearized form of the Poisson-Boltzmann (LPB) equation (Eq. 2-8):

$$\nabla(\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})) - 4\pi\beta \Sigma\, c(\mathbf{r})\, q_i^2 = -4\pi\rho(\mathbf{r}) \qquad \text{(Eq. 2-8)}$$

Using the definition of the ionic strength

$$I(\mathbf{r}) = \frac{1}{2} \Sigma\, c(\mathbf{r})\, q_i^2 \qquad \text{(Eq. 2-9)}$$

Eq. 2-8 is also expressed as

$$\nabla(\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})) - 8\pi\beta I(\mathbf{r}) = -4\pi\rho(\mathbf{r}) \qquad \text{(Eq. 2-10)}$$

2.4. Protonation probability

In equilibrium a single acidic group (HA and A^- for protonated and deprotonated states, respectively), where K_a is the equilibrium constant, can be defined as

$$HA \rightleftharpoons A^- + H^+ \qquad \text{(Eq. 2-11)}$$

$$K_a = \frac{[A^-][H^+]}{[HA]} \quad \text{(Eq. 2-12)}$$

$$pK_a = -\log K_a \quad \text{(Eq. 2-13)}$$

Hereby, protonation probability of this residue $<x>$ is defined as

$$< x > = \frac{[HA]}{[HA] + [A^-]} \quad \text{(Eq. 2-14)}$$

(for a basic group, A is to be replaced with B^+).

Thus, one can calculate pK_a from the protonation probability of $<x>$ as

$$pK_a = pH + \frac{1}{\ln 10} \ln \frac{<x>}{1 - <x>} \quad \text{(Eq. 2-15)}$$

Alternatively, Eq. 2-15 is rewritten as

$$< x > = \frac{\exp(-\ln 10(pH - pK_a))}{1 + \exp(-\ln 10(pH - pK_a))} \quad \text{(Eq. 2-16)}$$

In general, the Gibbs energy shift ΔG corresponding to the shift ΔK_a (= pH – pK_a) of a titratable group is

$$\Delta G = -RT \ln K_a \left(= -\frac{RT}{\log e} \log K_a\right) = 2.303 kT\, pK_a \quad \text{(Eq. 2-17)}$$

In proteins there are often several titratable groups, which are interacting with each other. In this case, it is useful to introduce the intrinsic pK_a of a titratable site μ ($pK_{a\text{-int},\mu} = . \Delta G_{int,\mu}/(2.303\ kT)$): $pK_{a\text{-int},\mu}$ is obtained if all other titratable sites are electrically neutral (Bashford and Karplus, 1991). Then, protonation probability $<x_i>$ of a titratable group i is given by the thermodynamic average over all possible titratable sites in the protein.

$$< x_i > = \frac{\sum_q x_i \exp\left(-\beta \sum_\mu (x_\mu \Delta G_{int,\mu} + \frac{1}{2} \sum_{\nu \neq \mu} q_\mu q_\nu W_{\mu\nu}) \right)}{\sum_q \exp\left(-\beta \sum_\mu (x_\mu \Delta G_{int,\mu} + \frac{1}{2} \sum_{\nu \neq \mu} q_\mu q_\mu W_{\mu\nu}) \right)} \quad \text{(Eq. 2-18)}$$

where $\beta = 1/(kT)$. Vector q is the total charge of titratable group i. x_i is the unity/ zero if the titratable group is protonated/ deprotonated. $W_{\mu\nu}$ is the electrostatic interaction between the two titratable groups if they are in their charged states i.e. protonated or deprotonated states for basic or acidic groups, respectively (summarized in ref. (Rabenstein et al., 1998)).

2.5. Born energy and background charge

The intrinsic pK_a is divided into the following terms, using the pK_a for a titratable group of the model system (pK_a^{model}, for amino acid side-chains, preferentially the value

measured in aqueous solution).

$$pK_{a\text{-int}} = pK_a^{model} + \Delta pK_a^{Born} + \Delta pK_a^{background} \qquad \text{(Eq. 2-19)}$$

where ΔpK_a^{Born} and $\Delta pK_a^{background}$ are the pK_a shift originated from the difference of the Born energy and the background-charge energy between the model system and the protein ($\Delta\Delta G^{Born}$ and $\Delta\Delta G^{background}$), respectively. Here, the background charges of a titratable group refer to all the atomic charges that interact with this titratable group.

$\Delta\Delta G^{Born}$ is the protonation energy difference of the *intra*molecular interactions of the titratable group between model system and protein. This term is equal to the solvation energy difference of the titratable group between model system and protein. $\Delta\Delta G^{background}$ is the protonation energy difference of the *inter*molecular interactions of the titratable group between the model system and protein. For further details, see ref. (Ullmann and Knapp, 1999).

2.6. Solution to the protonation probability

An attempt for numerical solution of Eq. 2-18 often fails if the protein contains a number of titratable groups. If a protein possesses n titratable sites, the possible protonation patterns amounts to 2^n e.g. even 20 groups yields $2^{20} \approx 10^6$ different protonation patterns (Rabenstein et al., 1998). Note that a monomer unit of bRC, PSI and PSII proteins possess ~170, ~450 and ~530 titratable sites, respectively. Thus, to solve Eq. 2-18, we used a Metropolis Monte Carlo (MC) method with the program KARLSBERG (Rabenstein, 1999), which was developed based on the program MCTI (Beroza et al., 1991).

The MC method may be inefficient for sampling strongly interacting sites, where the change of protonation state is restricted to only one site. This may lead to a trap of MC trajectory in a local minimum due to the artifact of energy barriers, which reduces the sampling efficiency. To improve sampling efficiency, MCTI is able to change the protonation states of two strongly coupled titratable groups simultaneously in each set of MC move after the initial simple MC move (Beroza et al., 1991). However, such double move per MC scan may not be sufficiently effective for strongly coupled titratable sites of a protein where the relatively low value of 4 for its dielectric constant is used. KARLSBERG is further improved to perform triple moves for protonation states of the three titratable sites in each set of MC move. For further details see refs. (Rabenstein et al., 1998; Rabenstein, 1999).

2.7. Computation of pK_a of a titratable group in protein

There are a few different ways to compute pK_a. One should choose them depending on the system or referring to the experimental conditions (see refs. (Ishikita and Knapp, 2005d; Ishikita et al., 2006)).

In a straightforward approach a titratable residue is biased by an individual energy term to be 50% protonated, while the protonation states of the other titratable residues are fully relaxed at a fixed pH (for instance, pH 7). This bias energy can be used to define the pK_a of this residue [**Henderson-Hasselbalch pK_a**]. This pK_a describes how much energy is needed to change the protonation state of this residue in its protein environment where the protonation pattern changes locally by equilibration due to the charge change of this residue without involving changes of solvent pH. This pK_a definition is, for instance, useful to describe the energetics of adiabatic proton transfer

processes between different titratable groups. The protonation dependence of the considered titratable residue obeys the Henderson-Hasselbalch equation (equivalent to the Nernst equation for a redox-active group) as a function of the bias energy (see Eqs. 2-15 and 2-20).

The second approach to determine pK_a values is to calculate it from the protonation pattern of all titratable residues as a function of solvent pH for a large pH range. Here, the pK_a of the titratable residue under consideration can be defined as the pH value where this residue is to 50% protonated [**effective pK_a**]. This is a more common pK_a definition and corresponds to the conditions where pK_a values of titratable groups in proteins are determined experimentally. When the molecular system contains only a single or several non-interacting titratable groups, the same pK_a values are computed for both definitions.

The third approach is to apply the protonation probability of a titratable group calculated at a single pH to Eq. 2-15 directly. Although this method is useful for initial estimation, it is not able to determine the pK_a value when the protonation probability is nearly 0 or 1 (see the definition of Eq. 2-15). Furthermore, in protein regions containing clusters of titratable residues protonation probability of a titratable group versus pH (i.e. effective pK_a) is often not fitted to the sigmoid curve, while even in the same case Eq. 2-15 assumes the sigmoid curve. This gives rise to a significant discrepancy of the resulting values between the two approaches. Indeed, in contrast to Eq. 2-18, Eq. 2-16 does not contain the interaction term that considers their protonation probability simultaneously. Therefore, the precise application of Eq. 2-15 to obtain the pK_a holds true for the system that contains only a single titratable group.

In a region rich with titratable residues, it is often hard to give a specific single value for effective pK_a because of the moderate protonation change over a wider pH range. Thus, if otherwise not specified, we calculated pK_a values as Henderson-Hasselbalch pK_a at pH 7.0, while protonation probabilities are computed by coupling solvent pH at pH 7 to all titratable residues and not by using a bias energy term. Protonation probabilities obtained in this way qualitatively relate to the pK_a definition effective pK_a. But, there is no quantitative correspondence between these protonation probabilities and pK_a derived from the simple Henderson-Hasselbalch relation because of possible strong electrostatic coupling between different titratable residues. Note that we do not consider possible structural changes of a protein upon pH changes. Such structural changes may occur more prominently in protein regions containing clusters of titratable residues (summarized in refs. (Ishikita and Knapp, 2005d; Ishikita et al., 2006)).

2.8. Computation of the E_m of a redox-active group in protein

The shift of the E_m upon insertion of a redox-active group from model system to protein originates from the same interactions that are to be considered in the pK_a shift of a titratable group (see **2.4-2.5**). In the ideal model system for a redox-active group, for instance, where the redox-active group is solely solvated in aqueous solution, the potential of the system is a function of the reduced and oxidized states ([A_{ox}]/[A_{red}]), according to the Nernst equation (Eq. 2-20).

$$E = E^0 + \frac{RT}{nF} \ln\frac{[A_{ox}]}{[A_{red}]} \qquad \text{(Eq. 2-20)}$$

where n denotes the number of electron involved in the redox reaction: and E^0 denotes the standard potential of the group. When [A_{ox}] = [A_{red}], the E of the system is called the

(midpoint) redox potential E_m, which then equals to E^0.

In proteins, there often exist several titratable groups that interact with the redox-active group. As in the pK_a computation, the E_m computation for protein is required to consider protonation states of all titratable groups and redox states of all redox-active groups. Hereby, as an analogy to the protonation probability of titratable residues, the redox probability of redox-active groups is obtained from a solution of Eq. 2-18. Also in the protein, when $[A_{ox}] = [A_{red}]$, the potential E (i.e. solution potential) is identical to the E_m for the redox-active group.

2.9. Influence of ionic strength on E_m

We uniformly used the ionic strength of 100 mM for computations. The ionic strength has a tendency to screen the influence of atomic charges on the protein surface. Large ionic strength diminishes the influence of surface atomic charges. This indicates that the ionic strength stabilizes the charged state of the redox-active group e.g. reduced/oxidized states for anionic/cationic redox-active groups, leading to the up-shift/down-shift in their E_m. However, in actual computations, the changes of ionic strength often result in little effect on the E_m. Such resulting tiny changes in the E_m are pronounced especially when the redox-active group is located either in a region rich with titratable residues or a region highly shielded from bulk solvent. The former case is due to compensation effects by small changes in protonation of several residues in response to the change in ionic strength, implying a buffer effect of titratable residues. Therefore, unless the redox-active group is on the protein surface and simultaneously isolated from other titratable residues, its E_m remains essentially unchanged in the presence of a sufficient amount of titratable residues. For actual examples, see ref. (Ishikita and Knapp, 2005d).

2.10. Estimation of the ET rate

The rate of ET from an electron donor D to an electron acceptor A can be estimated based on the values of E_m(D) and E_m(A) by evaluating the following empirical rate expressions (Page et al., 1999). They describe ET processes at T = 300K, which are downhill in energy (exergonic)

$$k_{T=300K}^{exergonicET} = 10^{13-0.6(R-3.6)-3.1(-|\Delta G|+\lambda)^2/\lambda} \qquad \text{(Eq. 2-21)}$$

or uphill in energy (endergonic)

$$k_{T=300K}^{endergonicET} = 10^{13-0.6(R-3.6)-3.1(-|\Delta G|+\lambda)^2/\lambda-|\Delta G|/0.06} \qquad \text{(Eq. 2-22)}$$

where R (> 3.6 Å) is the edge-to-edge distance, ΔG the E_m difference of the participating electron donor and acceptor groups and λ the reorganization energy. In these rate expressions the energy parameters ΔG and λ are given in units of eV and R in units of Å.