Chapter 4

Calculation of Protonation and Redox Equilibria in Proteins

It is well-known that a protein can exist in multiple conformational states (Frauenfelder *et al.*, 1988). Beside these different conformational states, the proteins can also exist in different protonation states. If the protein is redox active, also different redox states can occur. Thus to identify a protein state, the specification of the protein conformation, the protonation state, and the redox state is required.

If a single titratable group has two possible protonation forms, the total number of protonation states of a protein with N titratable groups is 2^N. In general, however, each titratable group can have multiple tautomeric forms and a protein can have multiple conformational states. Sometimes also multiple redox states of prosthetic groups may occur. A protein, which has N titratable groups each with n_i possible protonation forms (protonation and tautomeric forms), K redox-active groups each with k_j possible redox forms and L conformational states, possesses a total number of P states (eq 4.1).

$$P = L \prod_{i=1}^{N} n_i \prod_{j=1}^{K} k_j \tag{4.1}$$

If each titratable (redox) group would have only two protonation (redox) forms, then the state of the protein could be described by a number specifying the conformational state and a *(N+K)*-dimensional vector specifying the protonation and redox state of the protein. The components x_μ of that vector adopt either the value 1 or 0 depending on whether the group μ is protonated (reduced) or deprotonated (oxidized). If each group has more than two forms, the protein can be described by an integer identifying the conformational state, a *(N+K)*-dimensional vector specifying the protonation and redox state of the protein (elements ranging from 1 to n_i or k_i), and a *(N+K)*-dimensional vector specifying whether form n_i of group μ is protonated (reduced) or not (the elements specify either the number of bound protons or electrons). Although, a treatment allowing more than two protonation (redox) forms for each group would in principle be possible, it is not commonly used. Therefore, I refer in the following to the commonly used treatment considering two possible protonation (redox) forms for each group. Furthermore, I assume, that each conformational state can be represented by a single protein structure.

The protonation probability $\langle x_\mu \rangle$ of the site μ is given by a thermodynamic average over all

possible protonation, redox, and conformational states given by eq 4.2,

$$\langle x_\mu \rangle = \frac{\sum_{i=1}^{P} x_\mu \exp(-G_i/RT)}{\sum_{i=1}^{P} \exp(-G_i/RT)} \tag{4.2}$$

where x_μ is one or zero depending whether the site μ is protonated or not, G_i is the free energy of the state $i = \{p,l\}$ ($p = \{n,k\}$ specifies the protonation and redox state, l specifies the conformational state), R is the universal gas constant and T is the temperature. Analogous to eq 4.2, the probability of a particular redox state or a particular conformation can be obtained. From this probability $\langle x_\mu \rangle$, it is possible to calculate the energetic parameters, such as redox potentials as a function of environmental parameters, as for instance the pH value.

4.1 Protonation Equilibrium of a Single Titratable Group

The protonation equilibrium of a single titratable group can be described by eq 4.3.

$$\mathrm{HA} \rightleftharpoons \mathrm{A^-} + \mathrm{H^+} \tag{4.3}$$

The equilibrium constant is given by eq 4.4.

$$K_a = \frac{[\mathrm{A^-}][\mathrm{H^+}]}{[\mathrm{HA}]} \tag{4.4}$$

The pH of the solution and the pK_a of an acid are defined as the negative of the decadic logarithm of the hydrogen ion concentration (eq 4.5) and the K_a value (eq 4.6), respectively.

$$\mathrm{pH} = -\lg[\mathrm{H^+}] \tag{4.5}$$

$$\mathrm{p}K_a = -\lg K_a \tag{4.6}$$

By taking the logarithm of eq 4.4, multiplying the so obtained equation by -1 and replacing $-\lg[\mathrm{H^+}]$ and $-\lg K_a$ by pH and pK_a, respectively, we obtain after some additional transformations

$$\mathrm{pH} = \mathrm{p}K_a + \lg\frac{[\mathrm{A^-}]}{[\mathrm{HA}]} \tag{4.7}$$

Eq 4.7 is the so-called Henderson-Hasselbalch equation. The relation between pK_a and the standard reaction free energy G_a° is given by eq 4.8.

$$\mathrm{p}K_a = -\frac{G_a^\circ}{RT\ln 10} \tag{4.8}$$

The pH dependence of the protonation probability of an acid can be described as follows. The probability $\langle x \rangle$ that the acid AH is protonated, is given by $\langle x \rangle = [\mathrm{HA}]/([\mathrm{HA}]+[\mathrm{A^-}])$. With these definitions and eq 4.4 one obtains eq 4.9 and eq 4.10.

$$[\mathrm{HA}] = \frac{\langle x \rangle [\mathrm{A^-}]}{1-\langle x \rangle} \tag{4.9}$$

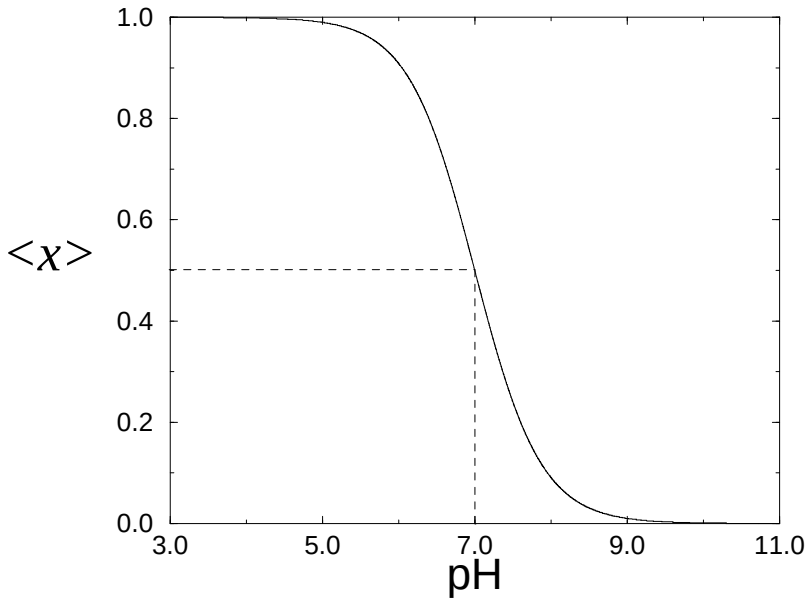

Figure 4.1: pH dependence of the protonation probability $\langle x \rangle$ of a titratable group. This curve was calculated with eq 4.12 using a pK_a of 7.0.

$$K_a = \frac{(1 - \langle x \rangle)[\mathrm{H}^+]}{\langle x \rangle} \tag{4.10}$$

With eqs 4.5 and 4.6, eq 4.10 becomes

$$\mathrm{pH} - \mathrm{p}K_a = -\lg \frac{\langle x \rangle}{(1 - \langle x \rangle)} \tag{4.11}$$

and finally after some transformations

$$\langle x \rangle = \frac{10^{(\mathrm{p}K_a - \mathrm{pH})}}{1 + 10^{(\mathrm{p}K_a - \mathrm{pH})}} \tag{4.12}$$

The function given by eq 4.12 is depicted in Figure 4.1. From eq 4.11 one can see that the pK_a of an acid is equal to the pH at which fifty percent of the acid molecules are protonated.

Because of eq 4.11 and because of the fundamental thermodynamic relation in eq 4.13,

$$G = -RT \ln \frac{\langle x \rangle}{(1 - \langle x \rangle)} \tag{4.13}$$

that the difference between the pH and the pK_a relates to the free energy required to protonate an acidic group at a given pH by eq 4.14.

$$G = RT \ln 10 (\mathrm{pH} - \mathrm{p}K_a) \tag{4.14}$$

4.2 Redox Equilibrium of a Single Redox-Active Group

The description of redox equilibria is in principle very similar to the description of protonation equilibria presented in Section 4.1. The equilibrium between the redox couple $\mathrm{A_{ox}}/\mathrm{A^-_{red}}$ (eq 4.15)

$$\mathrm{A_{ox}} + \mathrm{e}^- \rightleftharpoons \mathrm{A^-_{red}} \tag{4.15}$$

is defined in eq 4.16.

$$K_{ET} = \frac{[\mathrm{A_{red}}]}{[\mathrm{A_{ox}^-}][\mathrm{e^-}]} \tag{4.16}$$

By taking the natural logarithm, eq 4.16 transforms to eq 4.17.

$$-\ln[\mathrm{e^-}] = \ln K_{ET} - \ln\frac{[\mathrm{A_{red}}]}{[\mathrm{A_{ox}^-}]} \tag{4.17}$$

Analogously to the derivation of the Henderson-Hasselbalch equation (eq 4.7), two variables are defined (eqs 4.18 and 4.19).

$$E = -\frac{RT}{F}\ln[\mathrm{e^-}] \tag{4.18}$$

$$E^{\mathrm{o}} = \frac{RT}{F}\ln K_{ET} \tag{4.19}$$

Here, however, the natural logarithm rather than the decadic logarithm is used. A factor $-RT/F$, where F the Faraday constant, R the gas constant, and T is the thermodynamic temperature, is needed to obtain E and E^{o} in Volt. E^{o} is the standard redox potential, E is the eletromotive force or the redox potential. Multiplying eq 4.17 by RT/F and using eqs 4.18 and 4.19, eq 4.17 is transformed to the Nernst equation (eq 4.20).

$$E = E^{\mathrm{o}} + \frac{RT}{F}\ln\frac{[\mathrm{A_{ox}}]}{[\mathrm{A_{red}^-}]} \tag{4.20}$$

The relation between E^{o} and the standard reaction free energy G_{et}^{o} is given by eq 4.21.

$$E^{\mathrm{o}} = \frac{G_{et}^{\mathrm{o}}}{F} \tag{4.21}$$

The probability that a redox-active group is oxidized can be described similarly to the dependence of the protonation probability of an acid on pH (Section 4.1). The probability $\langle z\rangle$ that a redox-active group is in its oxidized state is defined as $\langle z\rangle = [\mathrm{A_{ox}}]/([\mathrm{A_{red}}]+[\mathrm{A_{ox}}])$. After an derivation analogous to the one presented in Section 4.1, the probability $\langle z\rangle$ that a redox-active group is reduced is finally obtained as

$$\langle \mathrm{z}\rangle = \frac{\exp\left(\frac{F}{RT}(E-E^{\mathrm{o}})\right)}{1+\exp\left(\frac{F}{RT}(E-E^{\mathrm{o}})\right)} \tag{4.22}$$

In analogy to eq 4.14, the free energy required to oxidize a redox-active group at a given redox potential of the solution is given by eq 4.23

$$G = -F(E-E^{\mathrm{o}}) \tag{4.23}$$

4.3 Titration Curves for a Protein in a Single Conformational State

The formalism to calculate the protonation probability of titratable groups in proteins for a single protein conformation is well established (Bashford & Karplus, 1990; Yang *et al.*, 1993;

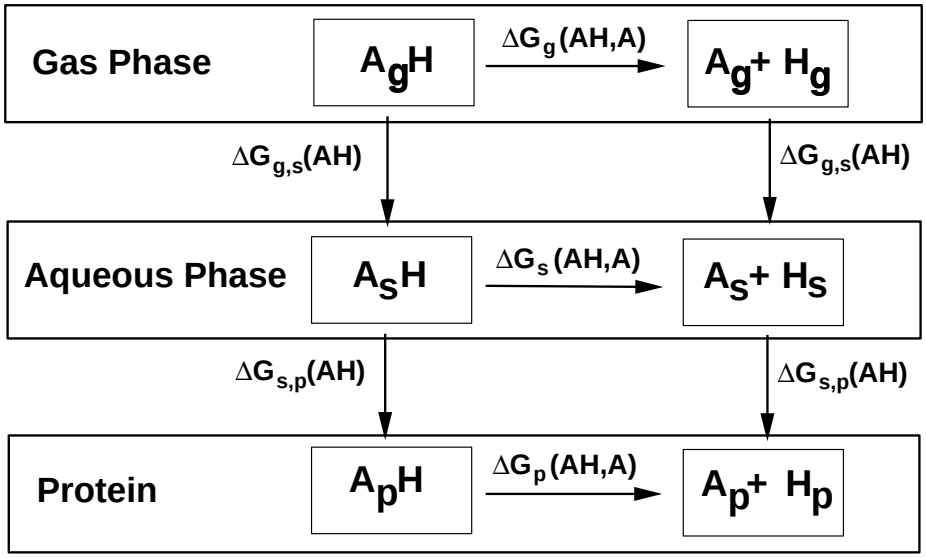

Figure 4.2: Thermodynamic cycle to calculate the protonation energy from gas phase properties. In the first step the molecule is transferred from the gas phase (g) to a solution (s) and in a second step the molecule is transferred from the solution into the protein (p).

Beroza *et al.*, 1991; You & Bashford, 1995; Bashford & Karplus, 1991; Antosiewicz *et al.*, 1994; Beroza & Fredkin, 1996; Sham *et al.*, 1997). The calculation of protein titration curves dissects in two parts. In the first part, the intrinsic pK_a values and the interaction energies between titratable groups are computed. In the second part, these energy terms are used to calculate the protonation probability of titratable groups in a protein.

4.3.1 Expressions for the Protonation State Energy

The pK_a value of a titratable group within a protein can be considerably shifted compared to the pK_a value of the same titratable group in aqueous solution. The shift is caused by interactions between the titratable group and other charges in the protein and also by changes in the dielectric environment of the titratable group when the group is transferred from aqueous solution into the protein.

If the protonation of only a single titratable group in the protein depends on pH, the pK_a shift is related to the difference between protonation energy of this group in the protein and the protonation energy of a proper model compound. The pK_a shift can be obtained via the thermodynamic cycles shown in Figure 4.2. The pK_a of model compounds in aqueous solution are usually determined experimentally (Tanford, 1962). Sometimes, however, experimental pK_a values are not easily measurable, because no appropriate model compounds exist outside proteins. In these cases, the upper (gas phase) thermodynamic cycle in Figure 4.2 can be used to estimate the solvent pK_a of a model compound that reconstitutes the properties of the group inside the protein. The energy difference between the protonated and unprotonated form in vacuum (upper horizontal transition) can be obtained from quantum mechanical calculations (Lim *et al.*, 1991; Potter *et al.*, 1994; Li *et al.*, 1996; Richardson *et al.*, 1997). The energy change due to the transition from vacuum to aqueous solution is accessible either from molecular dynamics calculations or from electrostatic calculations using the numerical solution of the Poisson-Boltzmann-Equation. The upper thermodynamic cycle of Figure 4.2 describes the calculation of a pK_a value in aqueous solution using the information obtained from a gas phase

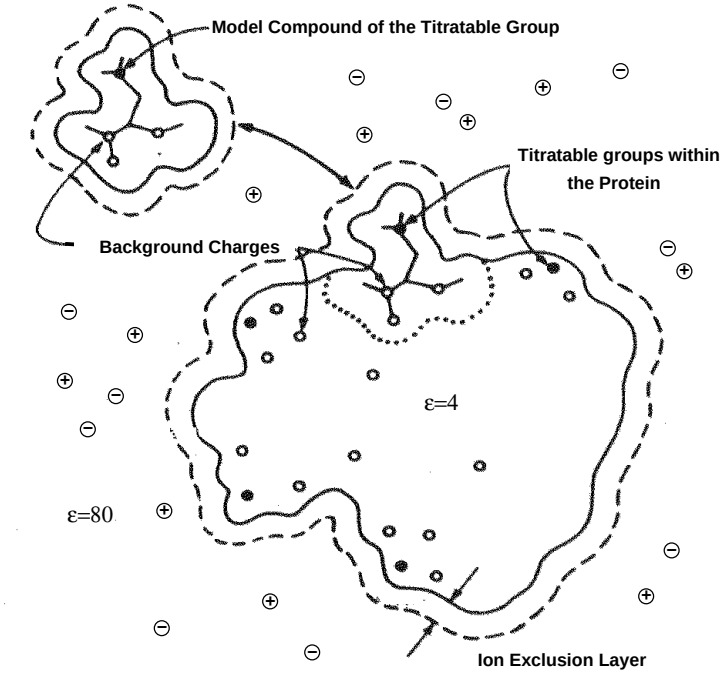

Figure 4.3: Titratable Groups and Model Compound an Ionic Solution. The model compound needs to have the same conformation as the amino acid in the protein in order to maintain the cancellation of the artificial grid energy by using the thermodynamic cycle technique.

quantum mechanical calculation. In the lower thermodynamic cycle of Figure 4.2 illustrates the calculation of the shift of the pK_a-value due to the transition of the titratable group from aqueous solution to the protein environment.

In general, the protonation of more than one titratable group in a protein depends on pH. Thus the interaction between these titratable groups is also pH dependent. Due to the interactions between these groups, titration curves of amino acids in proteins can deviate considerably from standard Henderson-Hasselbalch curves (Figure 4.1, eq 4.2). Because of this deviations, it is sometimes impossible to assign pK_a values to specific groups. Therefore, the pH value, at which the protonation probability of the titratable group is 0.5, the so-called p$K_{1/2}$ value, is used instead to describe the titration behavior. However, only a curve showing the pH dependence of the protonation probability of a particular group can describe the titration behavior correctly.

If only electrostatic contributions cause the difference between the protonation energies of a titratable group in a protein and in aqueous solution, the Poisson-Boltzmann equation provides a reasonable approximation of this energy difference. One energy contribution is the pK_a value of a proper model compound of the titratable group μ in aqueous solution, p$K_{a,\mu}^{\text{model}}$. Transferring the titratable group μ into a protein, in which all other titratable groups are in their uncharged protonation form, causes an energy shift. This energy shift can be separated into two contributions. The first energy contribution $\Delta\Delta G_{Born}$ is a Born-energy-like term (eq 4.24), which arises from the interaction of the charges of the titratable group with its reaction field. The second energy contribution $\Delta\Delta G_{back}$ arises from the interaction of the charges of the titratable group with non-titrating background charges and with the charges of the uncharged form of all other

titratable groups (eq 4.25).

$$\Delta\Delta G_{\mathrm{Born},\mu} = \frac{1}{2}\sum_{i=1}^{N_{Q,\mu}} Q^h_{i,\mu}[\phi_p(\mathbf{r}_i;Q^h_\mu) - \phi_m(\mathbf{r}_i;Q^h_\mu)] - \frac{1}{2}\sum_{i=1}^{N_{Q,\mu}} Q^d_{i,\mu}[\phi_p(\mathbf{r}_i;Q^d_\mu) - \phi_m(\mathbf{r}_i;Q^d_\mu)] \quad (4.24)$$

$$\Delta\Delta G_{\mathrm{back},\mu} = \sum_{i=1}^{N_p} q_i[\phi_p(\mathbf{r}_i;Q^h_\mu) - \phi_p(\mathbf{r}_i;Q^d_\mu)] - \sum_{i=1}^{N_m} q_i[\phi_m(\mathbf{r}_i;Q^h_\mu) - \phi_m(\mathbf{r}_i;Q^d_\mu)] \quad (4.25)$$

The summations in eq 4.24 run over the $N_{Q,\mu}$ atoms of group μ that have different charges in the protonated (h) ($Q^h_{i,\mu}$) and in the deprotonated (d) ($Q^d_{i,\mu}$) form. The first summation in eq 4.25 runs over the N_p charges of the protein that belong to atoms in non-titratable groups or to atoms of titratable groups (not to μ) in their uncharged protonation form. The second summation in eq 4.25 runs over the N_m charges of atoms of the model compound that do not have different charges in the different protonation forms. The terms $\phi_m(\mathbf{r}_i,Q^h_\mu)$, $\phi_m(\mathbf{r}_i,Q^d_\mu)$, $\phi_p(\mathbf{r}_i,Q^h_\mu)$, and $\phi_p(\mathbf{r}_i,Q^d_\mu)$ denote the values of the numerical solutions of the Poisson-Boltzmann equation at the position $\mathbf{r}$ of the atom i. The numerical solution of the Poisson-Boltzmann equation was obtained using the shape of either the protein (subscript p) or the model compound (subscript m) as dielectric boundary and assigning the charges of the titratable group μ in either the protonated (Q^h_μ) or the deprotonated (Q^d_μ) form to the respective atoms. These two energy contributions and the pK_a value of the model compound are combined to the so-called intrinsic pK_a value p$K_{intr,\mu}$ (eq 4.26).

$$\mathrm{p}K_{intr,\mu} = \mathrm{p}K^{\mathrm{model}}_{a,\mu} - \frac{1}{RT\ln 10}(\Delta\Delta G_{\mathrm{Born},\mu} + \Delta\Delta G_{\mathrm{back},\mu}) \quad (4.26)$$

The intrinsic pK_a value is the pK_a value that this site would have, if all other titratable groups are in their reference protonation form. I use the uncharged protonation form as reference form here. The interaction $W_{\mu\nu}$ between the two sites μ and ν in their charged form is defined in eq 4.27.

$$W_{\mu\nu} = \sum_{i=1}^{N_{Q,\mu}} [Q^h_{\mu,i} - Q^d_{\mu,i}][\phi_p(\mathbf{r}_i,Q^h_\nu) - \phi_p(\mathbf{r}_i,Q^d_\nu)] \quad (4.27)$$

The energy of a protonation state $\vec{x}_n$ of the protein, which is characterized by the protonation state vector $\vec{x}_n = (x^n_1, x^n_2, \ldots, x^n_N)$, is given by eq 4.28

$$G_n = \sum_{\mu=1}^{N}\left(x^n_\mu RT\ln 10(\mathrm{pH} - \mathrm{p}K_{intr,\mu})\right) + \frac{1}{2}\sum_{\mu=1}^{N}\sum_{\nu=1}^{N}\left(W_{\mu\nu}(x^n_\mu + q^o_\mu)(x^n_\nu + q^o_\nu)\right) \quad (4.28)$$

where x^n_μ is 1 or 0 depending whether the site μ is protonated or not, q^o_μ is the formal charge of the deprotonated form of group μ, i. e. -1 for acids and 0 for bases. The sums run over all N titratable groups. Here and in all subsequent equations it is assumed that $W_{\mu\mu} = 0$.

The problem of calculating the energy of 2^N protonation states is scaled down to the problem of calculating N intrinsic pK_a values and $(N \times (N-1))/2$ interaction energies. Instead of solving the Poisson-Boltzmann equation for the protein 2^N times numerically, i. e., once for each protonation state, only $2 \times N$ numerical calculations for the protein are needed. The $2 \times N$ additional calculations to obtain the electrostatic energies of the model compounds are needed

anyway. The calculation of the energy of a protonation state of the protein is given by the sum in eq 4.28. This energy can in principle be used in eq 4.2 to obtain titration curves as the thermodynamic average of the protonation of group i. But this calculation is often too time consuming for larger proteins. For this reason, approximation methods have been developed to reduce the computational burden (Tanford & Roxby, 1972; Bashford & Karplus, 1991; Yang *et al.*, 1993; Gilson, 1993).

4.3.2 Calculation of the Protonation Probability of Titratable Groups in a Protein

Several strategies to calculate protonation probabilities of proteins or other biomolecules exist in the literature. Here I describe those that are most commonly used.

Tanford-Roxby Approximation

One of the earliest attempts to calculate pK_a values of individual amino acids in proteins was made by Tanford & Roxby, 1972. Later, Bashford & Karplus, 1991 showed that this approach is a mean field approximation to the exact theory. The Tanford-Roxby approximation works well for weakly interacting groups, but fails if the protonation of two groups with similar pK_a is strongly coupled. It is computationally much less expensive than the exact treatment, because the full summation over all possible states in eq 4.2 is avoided. As already pointed out by Tanford & Roxby, 1972, the pK_a calculated by this approach depends on pH.

The Tanford-Roxby approximation assumes that the average charge of a titratable residue depends on the average charge of all other titratable groups. The pK_a inside the protein is calculated in an iterative manner by the algorithm introduced by Tanford & Roxby, 1972. In the i-th iteration, the pK_a of site μ is given by eq 4.29.

$$\mathrm{p}K_{a,\mu}(i) = \mathrm{p}K_{intr,\mu} - \frac{1}{RT\ln 10}\sum_{\nu=1}^{N}\left(W_{\mu\nu}(\langle x_\mu(i)\rangle + q_\mu^o)(\langle x_\nu(i)\rangle + q_\nu^o)\right) \tag{4.29}$$

where q_μ^o is the formal charge of the group μ of the deprotonated form, i. e., -1 for acids and 0 for bases. This pK_a is used in eq 4.30 in order to obtain the protonation probability for the $(i+1)$-th iteration by eq 4.30.

$$\langle x_\mu(i+1)\rangle = \frac{10^{(\mathrm{p}K_{a,\mu}(i)-\mathrm{pH})}}{1+10^{(\mathrm{p}K_{a,\mu}(i)-\mathrm{pH})}} \tag{4.30}$$

The iteration proceeds until self-consistency is reached. In the initial step, the pK_a values inside the protein are assumed to be identical with the intrinsic pK_a values (eq 4.26). Note that eq 4.29 uses p$K_{intr,\mu}$ rather than p$K_{a,\mu}(i)$. Otherwise the interaction energy would accumulate during iteration.

Hybrid Statistical Mechanical/Tanford-Roxby Algorithm

Yang *et al.*, 1993 developed an algorithm which combines an exact statistical mechanical treatment with the Tanford-Roxby approximation. All titratable residues within a certain cut-off distance of a titratable group are treated with an exact statistical mechanical algorithm, while

the titratable residues outside this cut-off radius are treated with the Tanford-Roxby approximation. A cut-off distance that gave reasonable results was 7 Å (Yang *et al.*, 1993). Alternatively, also an energy cut-off for the interaction energy $W_{\mu\nu}$ can be used.

In this approach, the titration curve of a protein with N titratable groups is calculated as follows; j_s titratable groups are within a certain cut-off distance of a particular group s, the remaining $N - j_s$ groups are outside this cut-off distance. The full statistical mechanical treatment is only done for the j_s groups inside the cut-off radius, the other groups are considered by a Tanford-Roxby treatment. The energy of the protonation substate n_s of the j_s residues in the i-th iteration is given by eq 4.31

$$\begin{aligned} G_{n_s}(i) &= \sum_{\mu=1}^{j_s} x_\mu^{n_s} RT \ln 10(\mathrm{pH} - \mathrm{p}K_{intr,\mu}) + \frac{1}{2}\sum_{\mu=1}^{j_s}\sum_{\nu=1}^{j_s}\left(W_{\mu\nu}(x_\mu^{n_s}+q_\mu^o)(x_\nu^{n_s}+q_\nu^o)\right) \\ &+ \sum_{\mu=1}^{j_s}\sum_{\nu=j_s+1}^{N}\left(W_{\mu\nu}(x_\mu^{n_s}+q_\mu^o)(\langle x_\nu(i)\rangle + q_\nu^o)\right) \end{aligned} \tag{4.31}$$

The average protonation $\langle x_s(i+1)\rangle$ of group s in the $(i+1)$-th iteration is calculated from eq 4.32.

$$\langle x_s(i+1)\rangle = \frac{\sum_{n_s=1}^{2^{j_s}} x_s \exp(-G_{n_s}(i)/RT)}{\sum_{n_s=1}^{2^{j_s}} \exp(-G_{n_s}(i)/RT)} \tag{4.32}$$

The same procedure is applied to all N titratable groups and iterated until self-consistency is reached. The calculation must be repeated at each pH in order to obtain a titration curve. In the first iteration step, the pK_a are assumed to be identical with the intrinsic pK_a values. A cut-off distance of 0 Å leads to the Tanford-Roxby approximation, while a cut-off distance of infinity leads to the exact statistical mechanical treatment. Note that for each group a different set of titratable sites will be treated exactly.

A similar but not identical treatment was used by Gilson, 1993 (Cluster Method). In this approach, the protein is divided into clusters of coupled residues. The criterion for the clustering is the coupling between the titratable groups. The interaction within a cluster is calculated exactly, while the interaction between different clusters is treated by the Tanford-Roxby approximation. Also in this approach, the calculation of the protonation probability iterates until self-consistency is reached.

Reduced Site Approximation

In order to avoid unnecessary calculations of energies of protonation states which are unlikely to occur, Bashford & Karplus, 1991 developed a method that considers only states that have a considerable contribution to the summation in eq 4.2. If the pH is far away from the pH at which the particular site titrates, the site will be considered to have a fixed protonation state. This reduces the number of protonation states from 2^N to 2^{N-M} if M sites have fixed protonations.

The standard free energy for adding a proton to site μ will be maximal, when all other sites are protonated, thus the maximum free energy is calculated with eq 4.33.

$$G_{max,\mu} = G_{intr,\mu} + \sum_{\nu=1}^{N} W_{\mu\nu}(q_\mu^o + 1.0) \tag{4.33}$$

The free energy will be minimal, when all other titratable groups are deprotonated, i. e.,

$$G_{min,\mu} = G_{intr,\mu} + \sum_{\nu=1}^{N} W_{\mu\nu} q_{\mu}^{o} \tag{4.34}$$

At a given pH, the maximum (minimum) protonation possible for site μ is obtained from eq 4.35.

$$\Theta_{max(min),\mu} = \frac{\exp(-\frac{\Delta G_{max(min),\mu} - RT \ln 10 \text{ pH}}{RT})}{1 + \exp(-\frac{\Delta G_{max(min),\mu} - RT \ln 10 \text{ pH}}{RT})} \tag{4.35}$$

A site μ is fixed in its protonated form if $\Theta_{min,\mu} > 1.0 - \varepsilon$ and in its unprotonated form if $\Theta_{max,\mu} > \varepsilon$, ε is an adjustable, albeit small parameter and should be set to 0.05 or less. However, the protonation of the fixed titratable groups will influence the pK_{intr} value of the unfixed titratable group μ. The correction is given by eq 4.36.

$$\Delta G_{intr,\mu} = \sum_{\nu=1}^{N_{fixed}} W_{\mu\nu} (q_{\nu}^{o} + f_{\nu}) \tag{4.36}$$

i.e.,

$$\text{p}K_{intr,\mu}^{correct} = \text{p}K_{intr,\mu} - \frac{1}{RT \ln 10} \sum_{\nu=1}^{N_{fixed}} W_{\mu\nu} (q_{\mu}^{o} + f_{\mu}) \tag{4.37}$$

where f_i is either one or zero according to if the site j is fixed in the unprotonated or protonated form.

Monte Carlo Titration

Beroza *et al.*, 1991 developed a Monte Carlo (MC) method to calculate titration curves of amino acids within proteins. This MC method is described below. With this MC method, states which are unlikely to occur are not considered. The average protonation $\langle x_{\mu} \rangle$ of site μ is obtained by averaging x_{μ} over the sampled states. The MC titration method can also be combined with the reduced site approximation (Bashford & Karplus, 1991) described above.

Standard Treatment. The initial vector $\vec{x}$ describing the protonation state of the protein is generated randomly. In a Monte Carlo move the protonation of the randomly chosen site μ is changed. The corresponding change in free energy ΔG is obtained from eq 4.38,

$$\Delta G = \Delta x_{\mu} \left(\ln 10 RT (\text{pH} - \text{p}K_{intr,\mu}) + \sum_{\nu=1}^{N} W_{\mu\nu} (x_{\mu} + q_{\mu}^{o})(x_{\nu} + q_{\nu}^{o}) \right) \tag{4.38}$$

where $\Delta x_{\mu} = \pm 1$ is the change in the protonation of site μ. The new protonation state is accepted according to the Metropolis criterion, if $\Delta G \leq 0$, the protonation of site μ is always changed, if $\Delta G > 0$, the protonation of site μ is changed with probability $\exp(-\Delta G/RT)$. A Monte Carlo scan is finished after N moves, i. e., after each of the N titratable groups has been tried to changed the protonation state once in average. After some equilibration scans, typically a few hundred, the protonation states of each scan are used to evaluate the average protonation.

Treatment of Strongly-Coupled Sites. If two sites are strongly coupled, the intermediate state between two protonation states may have a high energy and so the transition from one

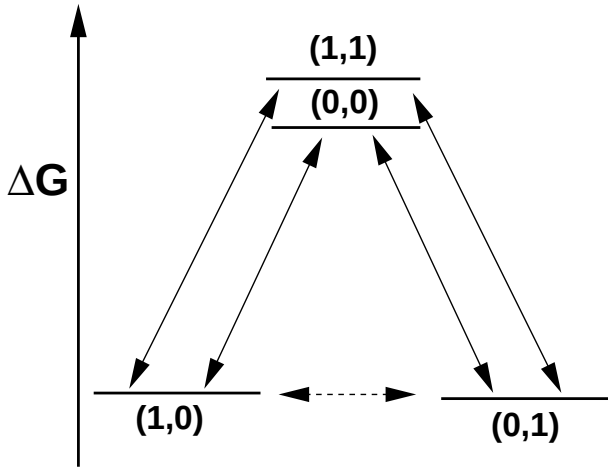

Figure 4.4: Treatment of two strongly-coupled sites in Monte Carlo titration calculation. If the energy barrier for the transition from state (1,0) to state (0,1) via the states (1,1) or (0,0) is too large (sold arrows), a Monte Carlo step is performed that simultaneously switches the protonation of both sites (dashed arrow).

state to an other might be unlikely as depicted in Figure 4.4. The problem can be avoided when the protonation of the two sites are switched simultaneously, which can correspond to a direct proton exchange between these two sites. The double switch was accepted according to the Metropolis criterion similar to the single site titration. At very low dielectric constants ($\varepsilon \leq 2$), the even three groups can be coupled strongly. Then the introduction of triple-moves can help to prevent sampling problems (Rabenstein *et al.*, 1998a).

Estimation of the Statistical Uncertainty. To estimate the statistical uncertainty of the MC calculation, it is necessary to calculate the number of independent data sets in the sample. The correlation function for the protonation of site μ determines the correlation time τ_μ^{corr} between approximately independent measurements. It is given by eq 4.39,

$$C_\mu(\tau) = \frac{1}{T-\tau} \sum_{t=0}^{T-\tau-1} x_\mu(t+\tau)x_\mu(t) - \langle x_\mu \rangle^2 \tag{4.39}$$

where t is the time in units of MC scans, T is the total number of scans (or the maximum time), and τ is the time variable of the correlation function. The correlation time τ_μ^{corr} is the time for which $C_\mu(\tau)$ becomes negligible (for instance $| C_\mu(\tau) | < C_\mu(0)/10$). The number of independent measurements is T/τ_μ^{corr}. The variance of one measurement is $C_\mu(0)$. The use of the average of T/τ_μ^{corr} independent measurements provides the standard deviation σ_μ given in eq 4.40.

$$\sigma_\mu = \sqrt{\frac{C_\mu(0)}{T/\tau_\mu^{corr}}} \tag{4.40}$$

4.4 Titration Curves for a Protein with Multiple Conformational States

4.4.1 Consideration of the Conformational Energy

If the protein can adopt multiple conformational states, eq 4.28 needs an additional term that accounts for the difference in the energy between the conformations in their respective reference states. As long as this additional term is additive, the calculation of titration curves with multiple conformation can be done in the same frame work as the calculation of titration curves with a single conformation. The energy of the conformational state l in the protonation state n is given by eq 4.41

$$G_n^l = \sum_{\mu=1}^{N} \left(x_\mu^{n,l} RT \ln 10 (\mathrm{pH} - \mathrm{p}K_{intr,\mu}^l) \right) + \frac{1}{2} \sum_{\mu=1}^{N} \sum_{\nu=1}^{N} \left(W_{\mu\nu}^l (x_\mu^{n,l} + q_\mu^o)(x_\nu^{n,l} + q_\nu^o) \right) + \Delta G_{conf}^l \tag{4.41}$$

where ΔG_{conf}^l is the energy difference between an arbitrarily chosen reference conformation r and conformation l. The protein is in its reference protonation state in both conformations r and l, i. e., the all titratable groups are in their uncharged protonation form.

The energy difference between two conformations arises from electrostatic and non-electrostatic interactions. Electrostatic contribution can be calculated from the numerical solution of the Poisson-Boltzmann equation. However, this numerical solution requires to assign the charges discretely on the grid. This results in an artifact termed grid energy. This grid energy precludes direct calculation of the binding energy by comparing the different conformations, along the thick horizontal line in Figure 4.5. This artifact can be avoided by calculating the other three steps in the thermodynamic cycle (Ullmann *et al.*, 1997b). The energies required to transfer the reference conformation r and conformation l of the protein from a medium with the dielectric constant of the protein to a medium with the dielectric constant of water (the vertical arrows) is calculated from the numerical solution of the linearized Poisson-Boltzmann equation as given in eq 4.42.

$$\Delta G_R = \frac{1}{2} \sum_{i=1}^{N_p} q_{i,p} \left(\phi_p^{\mathrm{inhom}}(\mathbf{r}_i, q_p) - \phi_p^{\mathrm{hom}}(\mathbf{r}_i, q_p) \right) \tag{4.42}$$

The term $\phi_p^{\mathrm{inhom}}(\mathbf{r}_i, q_p)$ denotes the value of the numerical solution Poisson Boltzmann equation at position $\mathbf{r}_i$ obtained for an inhomogeneous medium. This means, the dielectric constant is low and the ionic strength is zero within the protein, while dielectric constant is high outside the protein and the ionic strength must not vanish outside the protein. The term $\phi_p^{\mathrm{hom}}(\mathbf{r}_i, q_p)$ denotes the value of the numerical solution Poisson Boltzmann equation at position $\mathbf{r}_i$ obtained for an homogeneous medium. This means, the dielectric constant is low and the ionic strength is zero everywhere. Both electrostatic fields, $\phi_p^{\mathrm{hom}}(\mathbf{r}_i, q_p)$ and $\phi_p^{\mathrm{inhom}}(\mathbf{r}_i, q_p)$, were obtained from all charges q_p of the protein in the reference protonation state, in which all titratable groups are in their uncharged protonation form. The sum runs over all N_p charges $q_{i,p}$ of the protein. The grid energies cancel in ΔG_R. The difference in solvation energy between the reference conformation r and conformation l is obtained by eq 4.43.

$$\Delta\Delta G_R^l = \Delta G_R^r - \Delta G_R^l \tag{4.43}$$

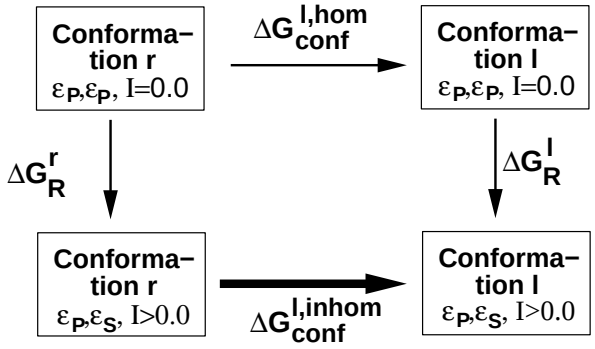

Figure 4.5: Thermodynamic cycle to calculate the difference in free energy between two conformations of the same molecule.

The energy difference along the upper arrow in Figure 4.5 between the reference conformation r and conformation l in a homogeneous medium is calculated analytically. Since the dielectric constants of the solute and the medium are equal, this energy change ΔG_{FF} is calculated with molecular mechanics force field using the dielectric constant of the protein ε_P (Brooks & Case, 1993; Vásquez *et al.*, 1994).

$$\begin{aligned} G_{FF} \;=\; & \sum_{i=1}^{N_b} \frac{1}{2} k_{b,i}(r_i - r_{o,i})^2 + \sum_{i=1}^{N_\theta} \frac{1}{2} k_{\theta,i}(\theta_i - \theta_{o,i})^2 + \sum_{i=1}^{N_\omega} \frac{1}{2} k_{\omega,i}(\omega_i - \omega_{o,i})^2 \\ & + \sum_{i=1}^{N_\phi} k_{\phi,i}(1 - cos(n_i\phi_i - \delta_i)) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} \frac{1}{4\pi\varepsilon_o\varepsilon_P}\frac{q_i q_j}{r_{ij}} \\ & + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1,j\neq i}^{N} 4\xi_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \end{aligned} \tag{4.44}$$

All symbols in eq 4.44 have its usual meaning, i. e., r_i, θ_i, ω_i, and ϕ_i denote the i-th bond length, the i-th bond angle, the i-th improper torsion angle, and the i-th torsion angle, respectively. The respective force constants are $k_{b,i}$, $k_{\theta,i}$, $k_{\omega,i}$, and $k_{\phi,i}$; $r_{o,i}$, $\theta_{o,i}$, and $\omega_{o,i}$ are the minima of the harmonic potentials; N, N_b, N_θ, N_ω, and N_ϕ are the number of atoms, bonds, bond angles, improper torsion angles, and torsion angles, respectively; $r_{i,j}$ is the distance between the atoms i and j having the charges q_i and q_i, ξ_{ij} is the depth of the Lennard-Jones potential, and σ_{ij} the distance at which the Lennard-Jones potential between two atoms is zero. The summation in the Coulomb term runs over all atomic charges in the reference protonation state. No Coulomb and Lennard-Jones energies are calculated for atoms that are connected by a covalent bond or a bond angle. It may be advantageous to consider only the electrostatic energy and to neglect all other contributions (Ullmann *et al.*, 1996).

The non-electrostatic solvation energy G_{NE}, which arises from non-electrostatic interactions of the molecule with its surrounding, is believed to be proportional to the solvent accessible surface as shown in eq 4.45,

$$\Delta G^l_{NE} = \gamma(A^r - A^l) \tag{4.45}$$

where A^l and A^r are the solvent accessible surfaces of the reference conformation r and conformation l respectively; γ is an empirically obtained parameter (Sitkoff *et al.*, 1994).

The total energy difference between conformation l and the reference conformation r is then

given by eq 4.46.

$$\Delta G^{l}_{conf} = \Delta\Delta G^{l}_{R} + \Delta G^{l}_{FF} + \Delta G^{l}_{NE} \tag{4.46}$$

This energy is used to calculate the total energy of the protein in conformation l and protonation state n in eq 4.41.

4.4.2 Modified Reduced Site Approximation to Calculate the Protonation Probabilities of Titratable Groups

The number of possible states of a protein is usually too large to evaluate eq 4.2 explicitly. In single conformer titration, the MC titration algorithms can help to overcome the sampling problem (Beroza *et al.*, 1991). The MC moves in multiple conformer MC titration, would be i) changes in protonation states of titratable groups and ii) changes in the protein conformation. This adapted MC algorithm, however, can cause serious problems. If the most probable protonation of one conformational state is very different from the most probable protonation of another conformational state, a transition between these conformational states would be a rare event in a simulation. Such a transition will most likely be rejected, because the energy difference between the present conformational state in its most probable protonation state and the new conformational state in the same protonation state will be rather large. Such a constellation causes sampling problems and a bias towards the initial conformational state. Furthermore, it is not quite clear, how to define the statistical error of the adapted MC procedure and so how to determine the quality of the simulation. Although, these problems can surely be solved within the MC frame work, I avoided these pitfalls with an approach, which is detailed below.

The approach I have developed is based on the reduced site approximation first introduced by Bashford & Karplus, 1991. In this approach, all sites that have a minimum possible protonation of more than 1-ε or a maximum possible protonation less than 0+ε are fixed in their respective protonation states, where ε is an adjustable cut-off parameter. The intrinsic pK_a values need corrections because of the interaction with the fixed charges. In contrast to the original reduced site approximation (Bashford & Karplus, 1991), which were used for a single conformation only, more energy parameters need corrections to maintain the various conformational states properly weighted. The corrected conformational energy $G^{corr,l}_{conf}$ of conformational state l includes the energy change due to the protonation of site μ, if site μ is protonated, and also interaction energy between the sites with non-variable protonation.

$$\begin{aligned}\Delta G^{corr,l}_{conf} &= \Delta G^{l}_{conf} + \sum_{\mu=1}^{N}\left(\delta^{l}_{\mu} x^{n,l}_{\mu} RT \ln 10(\mathrm{pH} - \mathrm{p}K^{l}_{intr,\mu})\right) \\ &\quad + \frac{1}{2}\sum_{\mu=1}^{N}\sum_{\nu=1}^{N}\left(\delta^{l}_{\mu}\delta^{l}_{\nu} W^{l}_{\mu\nu}(x^{n,l}_{\mu} + q^{o}_{\mu})(x^{n,l}_{\nu} + q^{o}_{\nu})\right)\end{aligned} \tag{4.47}$$

where δ^{l}_{μ} is 1 if the protonation of site μ in conformation l is variable, otherwise it is 0. The interaction between the sites with non-variable and with variable protonation needs corrections in two ways. If the site ν with variable protonation is in its protonated (h) state, the correction is given in eq 4.48; if the site ν with variable protonation is in its deprotonated (d) state, the correction is given in eq 4.49.

$$\Delta G^{h,l}_{\nu} = \sum_{\mu=1}^{N}\left(\delta^{l}_{\mu} W^{l}_{\mu\nu}(x^{n,l}_{\mu} + q^{o}_{\mu})(1 + q^{o}_{\nu})\right) \tag{4.48}$$

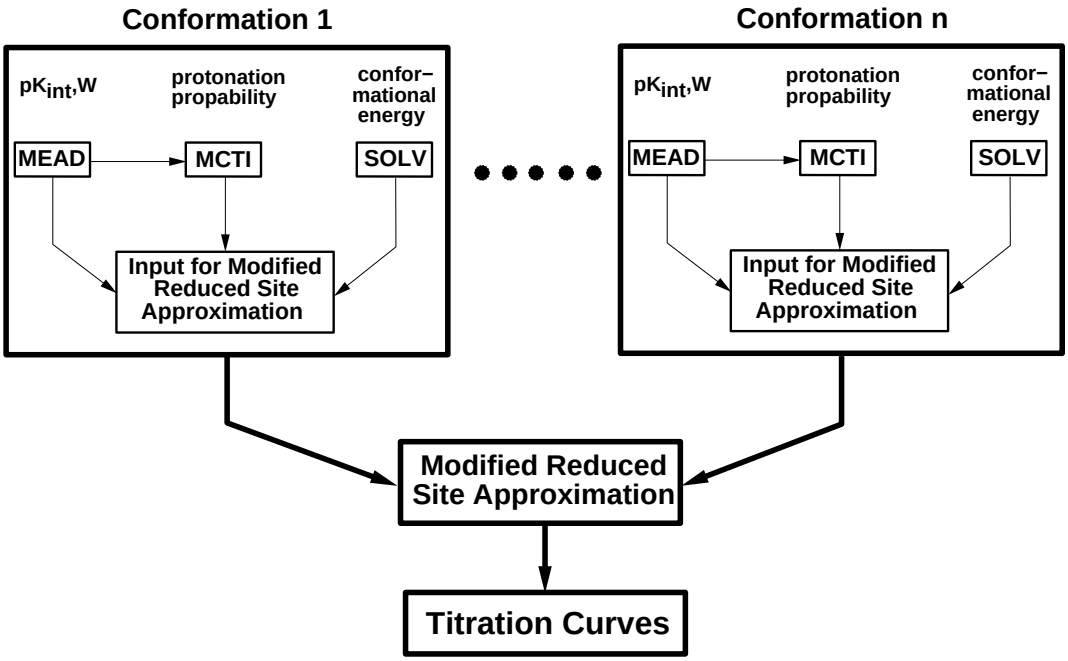

Figure 4.6: Flow chart for the calculation of titration curves with multiple conformations using the modified reduced site approximation. The program MEAD calculates the intrinsic pK_a values and the interaction energies, the program MCTI performs the Monte Carlo titration, the program SOLV calculates the conformational energy of the reference protonation state. A more detailed explanation is given in the text.

$$\Delta G_{\nu}^{d,l} = \sum_{\mu=1}^{N} \left(\delta_{\mu}^{l} W_{\mu\nu}^{l} (x_{\mu}^{n,l} + q_{\mu}^{o})(0 + q_{\nu}^{o}) \right) \tag{4.49}$$

The total energy of conformation l in protonation state n is given by eq 4.50.

$$\begin{aligned} G_{n}^{l} \;=\; & \sum_{\mu=1}^{N} \left((1-\delta_{\mu}^{l}) x_{\mu}^{n,l} RT \ln 10 (\mathrm{pH} - \mathrm{p}K_{intr,\mu}^{l}) \right) \\ & + \sum_{\mu=1}^{N} \left((1-\delta_{\mu}^{l}) \left(x_{\mu}^{n,l} \Delta G_{\mu}^{h,l} + (1-x_{\mu}^{n,l}) \Delta G_{\mu}^{d,l} \right) \right) \\ & + \frac{1}{2} \sum_{\mu=1}^{N} \sum_{\nu=1}^{N} \left((1-\delta_{\mu}^{l})(1-\delta_{\nu}^{l}) W_{\mu\nu}^{l} (x_{\mu}^{n,l} + q_{\mu}^{o})(x_{\nu}^{n,l} + q_{\nu}^{o}) \right) + \Delta G_{conf}^{corr,l} \end{aligned} \tag{4.50}$$

I determined the sites with fixed protonation from a MC titration of each conformational state, which precedes the actual calculation of titration curves. If the protonation of a site differs less than a given cut-off value ε from fully protonated or fully deprotonated, their protonation is fixed. This drastically reduces the number of variable titratable groups and the titration curves even of large proteins can be calculated in this way with only moderate computational effort (Rabenstein *et al.*, 1998a; Rabenstein *et al.*, 1998b).

Figure 4.6 shows a scheme representing the relation between the calculations carried out to obtain titration curves. The intrinsic pK_a and the interaction energies of the titratable groups are calculated first by the program MEAD. These are used as input for a conventional MC titration procedure using the program MCTI. Parallel the energies of the considered conformations are

calculated by the program SOLV. The intrinsic pK_a and the interaction energies of the titratable groups, the protonation probabilities, and the conformational energies of the conformational states are used as input for the actual calculation of titration curves. The entire computation scheme and also many single computation steps are highly parallelizable in a trivial way.

4.5 pH Dependent Processes Involving Proteins

4.5.1 Conformational Transitions

The pH-dependence of conformational transitions can in principle be calculated very similar to the calculation of protonation probabilities. The probability of conformation l_i can be obtained by the thermodynamic average over all possible states (eq 4.51),

$$p(l_i) = \frac{\sum_{n=1}^{2^N} \exp(-G_n^l/RT)}{\sum_{l=1}^{L} \sum_{n=1}^{2^N} \exp(-G_n^l/RT)} \tag{4.51}$$

where G_n^l is defined in eq 4.41. All symbols have the same meaning as in previous sections. If there are two possible conformations A and B, the free energy change for the conformational transition is given by eq 4.52

$$\Delta G(\mathrm{A} \rightarrow \mathrm{B}) = -RT \ln \frac{p(l_\mathrm{A})}{(1 - p(l_\mathrm{A}))} \tag{4.52}$$

4.5.2 Redox Reactions

Because the derivation of the pH dependence of redox processes is basically similar to the derivation of the pH dependence of protonation reactions and conformational transitions (Sections 4.3, 4.6, and 4.5.1), I describe a more complex scenario here. Given is a system with N titratable groups and K redox-active groups. Therefore the charge of the protein does not only depend on pH but also on the redox potential of the solution. I present the description of the model only for a protein with a single conformational state. An extension to multiple conformational states is, however, straight forward.

For the energy calculation of a particular protonation/redox state of the protein, additional energy terms are needed. The energy terms are the standard redox potential of a proper model compound, the difference between the redox potential of the model compound and the redox potential of the redox-active center in the reference state of the protein, the interaction between the redox-active sites, and the interaction between the redox-active sites and the titratable groups.

Like the pK_a of a model compound needed for the calculation of the protonation state energy, the standard redox potential of a model compound can be obtained either from experiment or from quantum chemical calculations. Sometimes only quantum chemical calculations are able to obtain the standard redox potentials of a model compounds for prosthetic groups in proteins, because no appropriate model compounds exist experimentally.

Similar to the definition of an intrinsic pK_a value pK_{intr} in Section 4.3.1, I define an intrinsic standard redox potential $E^{\mathrm{o}}_{intr,\eta}$ for the redox-active site η. It is the standard redox potential the

redox-active site would have, if all other titratable and redox-active groups are in their respective reference form.

$$E^{\mathrm{o}}_{intr,\eta} = E^{\mathrm{o}}_{model,\eta} + \frac{1}{F}(\Delta\Delta G^{\mathrm{redox}}_{\mathrm{Born},\eta} + \Delta\Delta G^{\mathrm{redox}}_{\mathrm{back},\eta}) \tag{4.53}$$

The term $\Delta\Delta G^{\mathrm{redox}}_{Born}$ arises from the interaction of the charges of the redox-active group η with its reaction field. The term $\Delta\Delta G^{\mathrm{redox}}_{\mathrm{back},\eta}$ arises from the interaction of the charges of the redox-active group η with invariant background charges and with the charges of the reference form of all other titratable and redox-active groups.

$$\Delta\Delta G^{\mathrm{redox}}_{\mathrm{Born},\eta} = \frac{1}{2}\sum_{i=1}^{N_{Q,\eta}} Q^{\mathrm{ox}}_{i,\eta}[\phi_p(\mathbf{r}_i;Q^{\mathrm{ox}}_\eta) - \phi_m(\mathbf{r}_i;Q^{\mathrm{ox}}_\eta)] - \frac{1}{2}\sum_{i=1}^{N_{Q,\eta}} Q^{\mathrm{red}}_{i,\eta}[\phi_p(\mathbf{r}_i;Q^{\mathrm{red}}_\eta) - \phi_m(\mathbf{r}_i;Q^{\mathrm{red}}_\eta)] \tag{4.54}$$

$$\Delta\Delta G^{\mathrm{redox}}_{\mathrm{back},\eta} = \sum_{i=1}^{N_p} q_i[\phi_p(\mathbf{r}_i;Q^{\mathrm{ox}}_\eta) - \phi_p(\mathbf{r}_i;Q^{\mathrm{red}}_\eta)] - \sum_{i=1}^{N_m} q_i[\phi_m(\mathbf{r}_i;Q^{\mathrm{ox}}_\eta) - \phi_m(\mathbf{r}_i;Q^{\mathrm{red}}_\eta)] \tag{4.55}$$

The meaning of the symbols in eqs 4.54 and 4.55 is analogous to meaning in eqs 4.24 and 4.25.

The interaction $U_{\eta,\chi}$ between the redox-active groups η and χ is given by eq 4.56.

$$U_{\eta\chi} = \sum_{i=1}^{N_{Q,\eta}} [Q^{\mathrm{ox}}_{\eta,i} - Q^{\mathrm{red}}_{\eta,i}][\phi_p(\mathbf{r}_i,Q^{\mathrm{ox}}_\chi) - \phi_p(\mathbf{r}_i,Q^{\mathrm{red}}_\chi)] \tag{4.56}$$

The interaction $V_{\eta\mu}$ between the redox active groups η and titratable group μ is given by eq 4.57.

$$\begin{aligned} V_{\eta\mu} &= \sum_{i=1}^{N_{Q,\eta}} [Q^{\mathrm{ox}}_{\eta,i} - Q^{\mathrm{red}}_{\eta,i}][\phi_p(\mathbf{r}_i,Q^h_\mu) - \phi_p(\mathbf{r}_i,Q^d_\mu)] \\ &= \sum_{i=1}^{N_{Q,\mu}} [Q^h_{\mu,i} - Q^d_{\mu,i}][\phi_p(\mathbf{r}_i,Q^{\mathrm{ox}}_\eta) - \phi_p(\mathbf{r}_i,Q^{\mathrm{red}}_\eta)] \end{aligned} \tag{4.57}$$

The energy of a particular protonation/redox state n of a protein depends on the pH and the redox potential E of the solution.

$$\begin{aligned} G_n &= \sum_{\mu=1}^{N}\left(x^n_\mu RT\ln 10(\mathrm{pH} - \mathrm{p}K_{intr,\mu})\right) + \frac{1}{2}\sum_{\mu=1}^{N}\sum_{\nu=1}^{N}\left(W_{\mu\nu}(x^n_\mu + q^o_\mu)(x^n_\nu + q^o_\nu)\right) \\ &\quad - \sum_{\eta=1}^{K}\left(x^n_\eta F(E - E^{\mathrm{o}}_{intr,\eta})\right) + \frac{1}{2}\sum_{\eta=1}^{K}\sum_{\chi=1}^{K}\left(U_{\eta\chi}(x^n_\eta + q^o_\eta)(x^n_\chi + q^o_\chi)\right) \\ &\quad + \sum_{\eta=1}^{K}\sum_{\mu=1}^{N}\left(V_{\eta\mu}(x^n_\eta + q^o_\eta)(x^n_\mu + q^o_\mu)\right) \end{aligned} \tag{4.58}$$

The actual oxidation or protonation probabilities can be calculated for instance by the MC titration method, the (modified) reduced site approximation, or the cluster method.

If there are only two redox sites in a protein or a protein complex, the reaction free energy of the electron transfer taking place between them depends on the pH of the solution. It is,

however, independent of the redox potential of the solution, since the solution redox potential cancels in the energy difference. Only the standard redox potential difference between the two sites and their electrostatic interaction energy needs to be considered and eq 4.58 reduces to eq 4.59,

$$\begin{aligned} G_n &= \sum_{\mu=1}^{N} \left(x_\mu^n RT \ln 10 (\mathrm{pH} - \mathrm{p}K_{intr,\mu}) \right) + \frac{1}{2} \sum_{\mu=1}^{N} \sum_{\nu=1}^{N} \left(W_{\mu\nu} (x_\mu^n + q_\mu^o)(x_\nu^n + q_\nu^o) \right) \\ &\quad - \delta_{et} F \Delta E^{\circ}_{intr}(\eta,\chi) + \sum_{\eta=1}^{2} \sum_{\mu=1}^{N} \left(V_{\eta\mu} (x_\eta^n + q_\eta^o)(x_\mu^n + q_\mu^o) \right) \end{aligned} \tag{4.59}$$

where $\Delta E^{\circ}_{intr}(\eta,\chi)$ is the energy difference between the redox state $\{\eta,\chi\} = \{1,0\}$ and the redox state $\{\eta,\chi\} = \{0,1\}$ if the protein is in its reference state, i. e., all titratable groups are in its reference protonation form; δ_{et} is one if the electron transfer took place, otherwise it is 0.

Together with Björn Rabenstein and Ernst-Walter Knapp, I applied eq 4.59 to calculate the redox potential difference between the quinones in the bacterial photosynthetic reaction center (bRC) (Rabenstein *et al.*, 1998b; Rabenstein *et al.*, 1998a). In our calculations, we used the crystal structure of the bRC from *Rps. viridis* with a resolution of 2.3 Å (Deisenhofer *et al.*, 1995; PDB entry 1prc). Since the cytochrome *c* subunit is more than 25 Å away from the quinone binding sites, we neglected this subunit in our calculations.

All water molecules, sulfate ions, and detergent molecules were removed. The influence of water was considered exclusively by a dielectric constant of 80 in cavities and outside of the protein. In some recent applications, selected water molecules were explicitly included in pK_a-calculations (Cometta-Morini *et al.*, 1993; Sampogna & Honig, 1994; Gibas & Subramaniam, 1996; Ullmann *et al.*, 1996). These calculations yielded sometimes results different from those obtained without explicit water molecules. Different selection schemes for the water molecules were applied. Only a few crystal water molecules, all crystal water molecules, or even additional solvent molecules were included. For hen egg lysozyme, the agreement between calculated and measured pK_a-values was better without explicit water molecules (Gibas & Subramaniam, 1996). Calculations without explicit water may agree better with experiments because the orientation of water molecules needed for these calculations is not known from X-ray crystallography. Since the orientation of water molecules is uncertain, we decided to remove all water molecules. We used an extended atom representation for most non-polar hydrogen atoms, except for the quinones, the chlorophylls, and the pheophytins, for which all hydrogens were treated explicitly. Coordinates of explicitly treated hydrogen atoms were generated with CHARMM (Brooks *et al.*, 1983). The positions of hydrogen atoms were energetically optimized, while the heavy atom positions were fixed.

In the used crystal structure, the Q_B binding pocket is occupied to 30 % only (Deisenhofer *et al.*, 1995). A structure with improved occupancy and resolution at the Q_B site has been solved but was not freely available when we did the calculations (Lancaster & Michel, 1996; Lancaster *et al.*, 1995). Therefore, we manipulated the bRC structure (PDB entry 1prc) at the Q_B site according to informations from Lancaster *et al.* (Lancaster & Michel, 1996; Lancaster *et al.*, 1995). We rotated Q_B around the axis through the methyl group at the quinone ring that is perpendicular to the ring plane, so that the methoxy group opposite to the methyl group shifted by 0.75 Å towards Ser L223. Furthermore the carboxyl group of Glu L212 was rotated by 90° around the bond between C_γ and C_δ.

Atomic partial charges of the amino acids, including the protonated and deprotonated state of titratable amino acids, were adopted from the CHARMM 21.3 parameter set. The acidic hy-

drogen atom of protonated glutamate and aspartate was not represented explicitly. Instead, appropriate charges at the two carboxyl oxygen atoms were assigned symmetrically. The atomic partial charges that are not included in the CHARMM 21.3 parameter set were calculated quantum-chemically with the program Spartan 4.0. Using the CHELPG-like method (Breneman & Wiberg, 1990) implemented in Spartan, we adjusted the atomic partial charges to represent the electrostatic potential calculated from the molecular wave function faithfully . The atomic partial charges of chlorophylls and pheophytins were calculated semiempirically at the PM3 level, those of the quinones (in all possible redox and protonation states) and of the deprotonated cysteine were calculated by an *ab-initio* method with a 6-31G** basis set. The atomic partial charges of high-spin non-heme iron (Kartha *et al.*, 1991) and its ligands were calculated by a density functional method (LSDA/VWN) implemented in Spartan using the DN** basis. The neurosporen and the isoprene tails of Q_A and Q_B, chlorophylls and pheophytins were not considered in the quantum chemical calculations. The atomic partial charges of these apolar groups were set to zero in the calculations of the electrostatic energy. The experimental pK_a-values were taken from Bashford *et al.*, 1993. For cysteine, which is not present in the protein studied by Bashford *et al.*, 1993, we used a pK_a-value of 8.5 for the model compound.

We performed focusing (Klapper *et al.*, 1986) in three steps. Initially, we used a 250 Å-cube with a 2.5 Å lattice spacing centered at the protein, followed by a 60 Å-cube with a 1.0 Å lattice spacing and finally a 15 Å-cube with 0.25 Å lattice spacing, both centered at the considered titratable group. We used an ionic strength of 100 mM, an ion exclusion layer of 2 Å, and a solvent probe radius of 1.4 Å. The dielectric constant in the protein was set to $\varepsilon_p = 4$, the dielectric constant of the solvent was set to $\varepsilon_s = 80$. These parameter values are similar to those used by Bashford, 1991 and also to those used in earlier calculations investigating the protonation of the bRC (Beroza *et al.*, 1995; Lancaster *et al.*, 1996). We neglected the influence of the membrane, since calculations on the membrane protein bacteriorhodopsin with (Bashford & Gerwert, 1992) and without (Sampogna & Honig, 1994) a membrane model yielded basically the same results.

The thermodynamic average over all possible protonation states can not be calculated exactly since the number of possible protonation states of the bRC (2^n with $n \approx 200$) is far too large. Instead we used a Metropolis MC method, implemented in the program MCTI (Beroza *et al.*, 1991). The statistical uncertainty of this method can be estimated by evaluating the protonation correlation function for each individual titratable group. To improve sampling efficiency, titratable groups coupled stronger than 1.5 pK_a-units changed their protonation state simultaneously in one MC move (Beroza *et al.*, 1991). Such simultaneous MC moves were done in addition to simple MC moves. In the unprotonated state of histidine two tautomers are possible. Both were considered in our calculations (Bashford *et al.*, 1993). Hence we could calculate the fraction of ε- and δ-tautomers of histidines.

The protonation states of all titratable groups of the bRC were calculated separately for each of the ten possible protonation and redox states of the quinone pair. Arginine, aspartate, cysteine, glutamate, histidine, lysine, tyrosine, and the C- and N-terminus, if not formylated, were considered as titratable groups. The histidines coordinating the magnesium ions of the chlorophylls and also the glutamate and the histidines coordinating the non-heme iron were not considered as titratable groups. Explicit water molecules were not included and hence were not considered as titratable groups. Their pK_a-values in aqueous solution are with -1.7 and 15.7 rather extreme. We did not expect that the protein is able to stabilize H_3O^+ or OH^- ions. Although ionized water molecules may participate as transient states in proton-transfer reactions, their presence in equilibrium states, which we investigated, is unlikely. In total 194

titratable groups were included in the calculations.

The MC sampling was done at pH 7.5 and 300 K. One MC scan comprises as many random individual protonation changes (MC moves) as there are titratable groups considered. We did at first 3000 MC scans. Then we fixed the protonation of all groups whose protonation probability differed from unity or zero by less than 10^{-6} in their respective protonation state and excluded these groups from further MC sampling. With this reduced set of titratable groups we performed another 7000 MC scans. This more efficient MC method is also implemented in the program MCTI. The sampling was sufficient to reach a standard deviation of less then 0.01 protons at each individual titratable group. In general, the standard deviation of a single group was much smaller than 0.01. The sum of the standard deviations of all protonation probabilities was for each state about 0.02 protons.

Continuum electrostatic methods are capable to yield reliable values merely for energy differences in different electrostatic environments. For that reason, our calculation required pK_a-values and redox potentials of menaquinone (MQ) and ubiquinone (UQ) in aqueous solution as reference for computing the reaction energy of electron-transfer and protonation reactions in the bRC. MQ and UQ, however, are not soluble in pure water. Swallow (Swallow, 1982) and Morrison *et al.* (Morrison *et al.*, 1982) extrapolated the pK_a-values of the quinones from pK_a-values of water soluble quinone derivatives taking into account the effect of different quinone ring substituents on the pK_a-value. The resulting pK_a-values are 4.9 for $UQ^{\cdot -}/UQ^{\cdot}H$ and 11.7 for UQH^-/UQH_2.

Redox potentials of quinones can not be measured in a protic solvent, since a reduced quinone will inevitably take up a proton. Redox potentials in aprotic solvents, however, are known; $MQ/MQ^{\cdot -}$ in DMF (dimethylformamide): -709 mV, $UQ/UQ^{\cdot -}$ in DMF: -602 mV (Prince *et al.*, 1983), and $UQ^{\cdot -}/UQ^{2-}$ in acetonitrile: -1450 mV (Morrison *et al.*, 1982).

We corrected the redox potentials of the quinones obtained in non-aqueous solutions to get redox potentials in aqueous solutions by accounting for the different solvation energies in the respective solvents. For that purpose, we calculated the energy for dissolving the quinones in their different redox states in water, acetonitrile, and DMF by the continuum electrostatic method (Sitkoff *et al.*, 1994). In the calculations, the following solvent parameters were used: The dielectric constants are $\varepsilon = 80$ for water, $\varepsilon = 38$ for acetonitrile, and $\varepsilon = 37$ for DMF. The solvent radii are 1.4 Å for water, 2,0 Å for acetonitrile, and 2.8 Å for DMF. Finally we obtained the following redox potentials of the quinones in aqueous solution: -699 mV for $MQ/MQ^{\cdot -}$, -592 mV for $UQ/UQ^{\cdot -}$, and -1420 mV for $UQ^{\cdot -}/UQ^{2-}$.

Titratable groups that have a protonation probability of less than 0.05 or more than 0.95 after MC titration in all considered bRC states are assumed to contribute to the thermodynamic average only in their totally deprotonated or protonated state, respectively. Therefore, we fixed those titratable groups in the totally deprotonated or protonated state and did not consider them as titratable groups in subsequent calculations. Only five titratable groups had variable protonations at pH 7.5 according to this criterion. These residues are aspartate M182 and the glutamates H45, H177, H234, and M76. To test whether fixing the protonation state of nearly completely protonated or unprotonated residues is justifiable, we repeated the calculation of protonation patterns state with this reduced set of unfixed titratable groups. The difference between the protonation probability in this calculation and the calculation with the full set of titratable groups was less than 0.05 for each of the five residues with variable protonations.

We calculated the energy of the protonation reactions of Q_B by the continuum electrostatic method described in Section 4.3. In addition to the five titratable groups with variable protonations, Q_B was also considered as titratable group. Since the number of variable titratable

group was small in this application. No MC sampling was necessary and the thermodynamic average could be calculated exactly. The exact evaluation of the thermodynamic average prevents sampling problems, which may occur with MC sampling if the protonation energy of Q_B is too large. The protonation energy was obtained from the protonation probability Q_B by using eq 4.13. The reaction energy of electron transfer from Q_A to Q_B was obtained by using eqs 4.59, 4.2, and 4.13. The results of the calculations on the bRC quinones are summarized in Section 5.2.

4.5.3 Protein-Protein Association Equilibrium

Measurements on proton uptake or release upon association or reduction of proteins have frequently been used to determine the pH dependence of association constants (Wyman, 1964; Lebowitz & Laskowski, 1962; Mauk *et al.*, 1991; Mauk *et al.*, 1994; van Vlijmen *et al.*, 1998), of redox potentials (McPherson *et al.*, 1988; Beroza *et al.*, 1995), and of unfolding free energies (Tanford, 1970; Yang & Honig, 1993; Schaefer *et al.*, 1997). The method used in these studies is based on the so-called *proton-linkage model* that relates the pH dependence of an equilibrium constant to the proton release upon this reaction (Laskowski & Finkenstadt, 1972). The proton linkage model can also be used to determine the pH dependence of the free energy change of conformational transitions and the pH dependence of redox reactions. The free energy change upon a reaction can however not be calculated by the proton linkage model.

The association between two molecules is described by eq 4.60, where K_A represents the association constant.

$$\mathrm{A} + \mathrm{B} \rightleftharpoons \mathrm{AB}; \qquad K_A = \frac{[\mathrm{AB}]}{[\mathrm{A}][\mathrm{B}]} \tag{4.60}$$

The molecules A, B, and AB involved in that reaction can exist in L_A, L_B, and L_AB protonation states, respectively. The total concentration of each species is then given by the sum over all possible protonation states. Therefore, eq 4.60 was rewritten and resulted in eq 4.61 (Laskowski & Finkenstadt, 1972),

$$K_A = \frac{\sum_{i=0}^{L_\mathrm{AB}} [\mathrm{AB}_i]}{\sum_{i=0}^{L_\mathrm{A}} [\mathrm{A}_i] \sum_{i=0}^{L_\mathrm{B}} [\mathrm{B}_i]} \tag{4.61}$$

where $[\mathrm{A}_i]$, $[\mathrm{B}_i]$, and $[\mathrm{AB}_i]$ denote the concentration of A, B, and AB in the respective protonation state i. Because the concentrations $[\mathrm{A}_i]$, $[\mathrm{B}_i]$, and $[\mathrm{AB}_i]$ depend on pH, K_A depends on pH. This approach leads finally to eq 4.62, where a represents the number of protons released upon association.

$$\lg K_A(\mathrm{pH}_2) = \lg K_A(\mathrm{pH}_1) + \int_{\mathrm{pH}_1}^{\mathrm{pH}_2} a \ \mathrm{dpH} \tag{4.62}$$

This equation states that if the association constant $K_A(\mathrm{pH}_1)$ at pH_1 is known, the association constant $K_A(\mathrm{pH}_2)$ at pH_2 can be obtained simply by integrating over the number of protons released upon association. A detailed derivation of eq 4.62 is given in Section A.

Alternatively to the proton linkage model, the absolute pH dependent free energy change upon association can be calculated by an approach similar to that outlined in Section 4.5.1. This approach would however also require the computation of the free energy of association in a defined reference state. Such a computation is, however, fairly difficult.