# Chapter 2
Titration curves of proteins

In this chapter, I will give the theoretical background and describe a method how the titration curves of the proteins can be calculated. After a brief introduction about acid-base and redox equilibria, the electrostatic calculation of the protonation and oxidation probabilities by solving the Poisson-Boltzmann equation numerically on a grid with a subsequent Monte Carlo titration will be explained. Among the 20 native amino acids in the proteins, about half of them contain side chains that are titratable groups. Also some cofactors are redox-active and have a role in electron transfer processes. Since protonation and redox reactions are coupled, their equilibria have to be considered simultaneously. It requires to redefine the standard definitions of the pK_a and E^0 values, which become dependent on pH and the solution redox potential. Only using these values, one can describe titration behavior and energetics in the proteins properly.

2.1 Introduction

2.1.1 Acid-base equilibrium of a single titratable group

Among amino acids, there are two types of titratable residues which can take part in the protonation equilibrium – acidic and basic amino acids. The protonation equilibrium of a single titratable group can be described by eq. 2.1 for acids and eq. 2.2 for bases,

$$HA \underset{K_b}{\overset{K_a}{\rightleftharpoons}} A^- + H^+ \tag{2.1}$$

$$BH^+ \underset{K_b}{\overset{K_a}{\rightleftharpoons}} B + H^+ \tag{2.2}$$
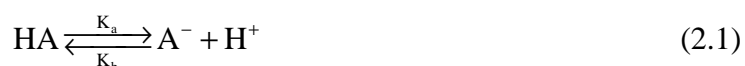
where the K_a equilibrium constant (acid constant) is defined by mass action law:

$$K_a = \frac{[A^-][H^+]}{[HA]} \tag{2.3}$$

While acids are in protonated state neutral, bases are positively charged. One can notice that the K_b base constant for opposite reaction is reciprocal value of the K_a acid constant.

$$K_b = \frac{[BH^+]}{[B][H^+]} = \frac{1}{K_a} \tag{2.4}$$

The pH of the solution and the pK_a of an acid are defined as the negative decadic logarithms of the hydrogen ion concentration and of the K_a value, respectively.

$$pH = -\log[H^+] \tag{2.5}$$

$$pK_a = -\log K_a \tag{2.6}$$

Using these definitions one obtains the Henderson-Hasselbach equation (eq. 2.7) from eq. 2.3.

$$pH = pK_a + \log \frac{[A^-]}{[HA]} \tag{2.7}$$

The pK_a value is related to the standard reaction (protonation) free energy G^0_{prot} of the acid-base equilibrium by equation 2.8, where R is the gas constant and T is the temperature in K.

$$G^0_{prot} = -RT \ln 10 pK_a \tag{2.8}$$

The probability $\langle x \rangle$ that the acid HA is protonated is given by $\langle x \rangle = [HA]/([HA]+[A^-])$. With this definition and equations 2.3 and 2.7 one obtains the protonation probability of an acid HA as:

$$\langle x \rangle = \frac{\exp(-\ln 10(pH - pK_a))}{1 + \exp(-\ln 10(pH - pK_a))} \tag{2.9}$$

This equation describes the pH dependence of the protonation probability of an acid and its graphical presentation is sigmoidal curve so called acid-base titration curve, where $\langle x \rangle$ is plotted against pH. At pH equal to pK_a, protonation probability becomes $\langle x \rangle = 0.5$. It means that an acid is in equilibrium where one half of all molecules are deprotonated. From eq. 2.9 one can see that the free energy required to protonate a titratable group at a given pH and temperature is:

$$G_{prot} = RT \ln 10(pH - pK_a) = -RT \ln \frac{\langle x \rangle}{1 - \langle x \rangle} \tag{2.10}$$

2.1.2   Redox equilibrium of a single redox-active group

The redox-active group and its redox equilibrium can be treated similarly. Redox-active groups one can find very often in protein systems and they are typically cofactors like heme, FAD and NAD, as well as some aromatic amino acid residues, which can take part in radical reactions and electron transfer through a protein.

If we are looking in a single, isolated redox active group, the equilibrium between the redox couple A_{oxd}/A_{red} is given as:

$$A^+_{oxd} + e^- \xrightleftharpoons{K_{ET}} A_{red} \tag{2.11}$$

and the K_{ET} equilibrium constant is defined with equation 2.12.

$$K_{ET} = \frac{[A_{red}]}{[A^+_{oxd}][e^-]} \tag{2.12}$$

In analogy to the derivation of the Henderson-Hasselbalch equation (eq. 2.7) one defines the solution redox potential (or the electromotive force) E_{sol} and the standard redox potential (or the midpoint potential) E^0 of the redox couple A_{oxd}/A_{red}, respectively as:

$$E_{sol} = -\frac{RT}{F}\ln[e^-] \tag{2.13}$$

$$E^0 = \frac{RT}{F}\ln K_{ET} \tag{2.14}$$

Here, the natural logarithm rather than the decadic logarithm is used. F is the Faraday constant, R is the universal gas constant and T is the absolute temperature. The quantities E_{sol} and E^0 are obtained in the units of Volt. Taking the logarithm of eq. 2.12, multiply by a factor RT/F and using the definition of E_{sol} and E^0 given in eq. 2.13 and 2.14, one obtains Nernst law defined by eq. 2.15. As one can see the similarity with the Henderson-Hasselbalch equation (eq. 2.7) is obvious.

$$E_{sol} = E^0 + \frac{RT}{F}\ln\frac{[A^+_{oxd}]}{[A_{red}]} \tag{2.15}$$

The E^0 value relates to the standard reaction free energy G^0_{redox} as given by eq. 2.16.

$$G^0_{redox} = FE^0 \tag{2.16}$$

The probability $\langle x \rangle$ that a redox-active group is in its oxidized state is defined by $\langle x \rangle = [A_{oxd}]/([A^+_{oxd}]+[A_{red}])$. While the protonation probability of an acid depends on the solution pH value (see eq. 2.9), the oxidation probability of an isolated redox couple depends on the solution redox potential.

$$\langle x \rangle = \frac{\exp\left(\frac{F}{RT}\left(E_{sol}-E^0\right)\right)}{1+\exp\left(\frac{F}{RT}\left(E_{sol}-E^0\right)\right)} \tag{2.17}$$

The graphical presentation of this equation gives the sigmoidal redox titration curve, where $\langle x \rangle$ is plotted against the solution redox potential E_{sol}. The concentration of reduced and oxidized form becomes equal at $E_{sol} = E^0 = E^0_{1/2}$.

From eq. 2.17 one can read that the free energy required to oxidize a redox-active group at a given solution redox potential and temperature is:

$$G_{redox} = -F(E_{sol}-E^0) = -RT\ln\frac{\langle x \rangle}{1-\langle x \rangle} \tag{2.18}$$

2.1.3   The model of a protein in solution

As we have already seen the titration behavior of a single acid or base (redox-active) group is completely determined by its pK_a (E^0) value. It exhibits standard sigmoidal curve that obeys Henderson-Hasselbach (or Nernst) law. It is enough to know the pK_a (E^0) value in water and we can describe titration behavior of a single titratable group in water. But it is well known that titration curves of native proteins differ substantially from the sums of unperturbed titrations of the constituent acidic and basic groups. The dominant factor responsible for the difference arises from electrostatic interactions between titratable groups (Tanford & Roxby 1972). If a titratable group is buried in a hydrophobic region, it can also show non-standard titration behavior due to the environment. Hydrogen bonds could affect pK_a values, too.

6

Thus, in proteins the local environment of titratable groups is much more complex due to the interactions between titratable groups and other partial charges of polar groups as well as ions in the surrounding solvent. Finally, the dielectric environment is changing when the group is transferred from aqueous solution into the protein. The figure 2.1 depicts a protein molecule in a heterogeneous dielectric medium and explains the model of continuum electrostatics that we used for electrostatic calculations in all applications. In this approach, the protein is described by a low dielectric constant (typically $\varepsilon = 2$ or 4), ionic strength $I = 0$ and with charges of titratable groups and background charges of polar groups. The interior of the protein is separated from the solvent by the solvent accessible surface (the Connolly surface), symbolized by a solid line while the border of the ion exclusion layer is marked by a dashed line. Within the ion exclusion layer the dielectric constant ε has a high value but the ionic strength I is still equal to 0. The solvent is not represented by explicit solvent molecules, but rather implicitly by a medium of high dielectric value (for water $\varepsilon = 80$) and ionic strength $I \neq 0$ (typically $I = 0.1$). Thus, the mobile ions are presented also implicitly through ionic strength. The solvent accessible surface is defined by rolling a sphere of radius, $r = 1.4$ Å (so called probe radius) along the protein surface (VAN DER WAALS surface). In that way all cavities and holes inside a protein, that can adopt (host) the sphere of probe radius, will be assigned a dielectric constant of $\varepsilon = 80$.
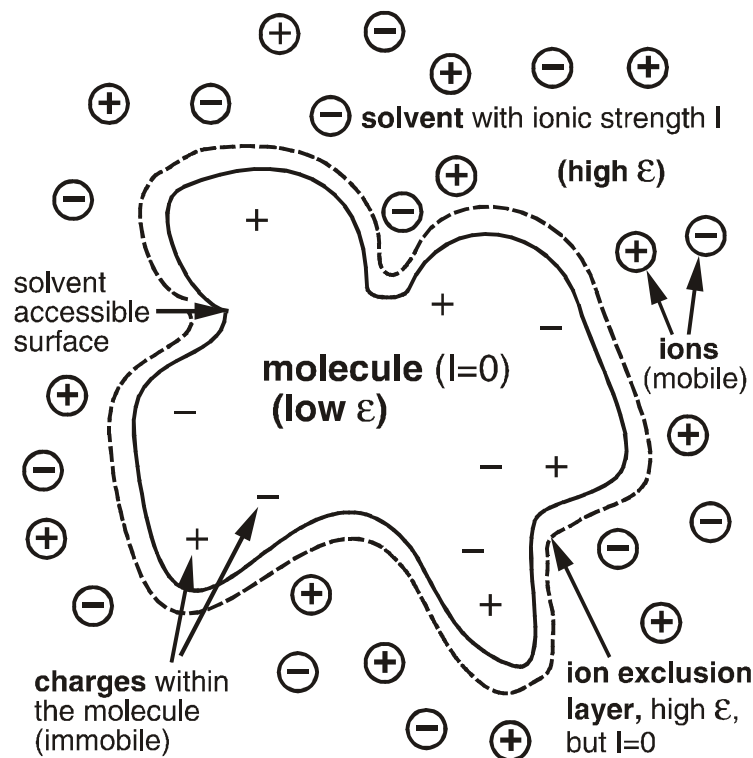

Figure 2.1: A protein molecule in heterogeneous dielectric medium, using the continuum electrostatic model for a solvent. A protein can be modeled as a low dielectric cavity with fixed charges in a high dielectric environment with mobile ions.

2.2 Continuum electrostatic calculations

The methodology to calculate the protonation probability of titratable groups in proteins for a single and later for multiple protein conformations has been developed during last decade (Bashford & Karplus, 1990; Bashford & Karplus, 1991; Beroza et al., 1991; Yang et al., 1993; Antosiewicz et al., 1994; You & Bashford, 1995; Beroza & Fredkin, 1996; Beroza & Case, 1996; Sham et al., 1997; Alexov & Gunner, 1997; Beroza & Case, 1998; Ullmann 1998). These calculations are performed in two steps. In the first step, the intrinsic pK_a values and interaction energies between charged titratable groups are computed, solving the linear Poisson-Boltzmann equation (LPBE) numerically on a grid for a protein in inhomogeneous dielectric medium and using a continuum electrostatic model for solvent. If we assume that electrostatic interactions are the predominant contributions responsible for the difference between the protonation energies of titratable groups in a protein and in aqueous solution, than using the LPBE (see Honig & Nicholls, 1995 for a review) it is possible to evaluate these energy differences with high accuracy and by using the additivity of the solutions of the LPBE it becomes possible to separate them into several independent energy contributions. For the matter of completeness, it should be mentioned that there are also applications where the non-linear Poisson-Boltzmann equation (NPBE) has been used (Oberoi & Allewell, 1993; Ripoll et al., 1996; Juffer et al., 1997; Vila et al., 1998), but the advantage of additivity makes the LPBE to be more useful.

In the second step, these energy terms (the intrinsic pK_a values and W matrix with interaction energies between titratable groups) are used to compute titration curves for all titratable sites within a protein. Here, there have been several approximation methods developed to compute protonation probabilities because exact summation for a protein that contains very often several tens to hundreds of titratable sites is not possible.

2.2.1 The Poisson-Boltzmann equation

Derivation

The electrostatic interaction of the protein and its solvent environment, as it is depicted in Figure 2.1, can be described by the Poisson-Boltzmann equation in terms of the electrostatic potential $\phi(\vec{r}, \rho)$ at the position $\vec{r}$, which arises from the charge distribution (density) $\rho(\vec{r}')$ at $\vec{r}'$. The Poisson equation is the basic equation of electrostatics and it is derived from the Coulomb potential. If the point charges q_i are distributed in a homogeneous system with dielectric constant ε, the electrostatic potential reads:

$$\phi(\vec{r}) = \sum_i \frac{q_i}{\varepsilon |r - r_i|} \qquad (2.19)$$

where the sum runs over all point charges q_i at position $\vec{r_i}$. This and all of the following equations are given in reduced electrostatic units, so the factor $4\pi\varepsilon_0$ is unity. The electrostatic energy of a such system is given by Coulomb law:

$$U = \sum_i \phi(\vec{r_i})\, q_i = \frac{1}{2} \sum_{i \neq j} \frac{q_i q_j}{\varepsilon |r_i - r_j|} \qquad (2.20)$$

Instead of the explicit point charge representation, one can use the charge density ρ, where the Coulomb potential (eq. 2.21) and the corresponding electrostatic energy (eq. 2.22) for the system with uniform dielectric constant ε is defined as:

$$\phi(\vec{r}) = \int_V \frac{\rho(\vec{r}\,')d\vec{r}\,'}{\varepsilon|\vec{r}-\vec{r}\,'|} \qquad (2.21)$$

$$G_{el} = \frac{1}{2}\int_V \phi(\vec{r})\,\rho(\vec{r})\,d\vec{r} = \frac{1}{2}\int_V\int_V \frac{\rho(\vec{r})\rho(\vec{r}\,')}{\varepsilon|\vec{r}-\vec{r}\,'|}d\vec{r}\,d\vec{r}\,' \qquad (2.22)$$

Applying the Laplace operator and few transformations, it yields the Poisson equation for a homogeneous (uniform) dielectric medium:

$$\nabla^2\phi(\vec{r}) = -4\pi\frac{\rho(\vec{r})}{\varepsilon} \qquad (2.23)$$

where the effect of the homogeneous dielectric medium on the electrostatic potential is taken into account through a constant ε value.

If ε is not constant but varies through space, the system becomes heterogeneous and the Poisson equation has to be modified. In a such situation, the Poisson equation for an inhomogeneous dielectric medium ensures the form:

$$\nabla\left[\varepsilon(\vec{r})\,\nabla\phi(\vec{r})\right] = -4\pi\rho(\vec{r}) \qquad (2.24)$$

Now, all quantities, the electrostatic potential $\phi(\vec{r})$, the charge density $\rho(\vec{r})$ inside a protein, as well as the dielectric constant $\varepsilon(\vec{r})$ depend on the spatial position $\vec{r}$. The simple Coulomb law is no longer valid for the situation of a spacially variable dielectric constant.

The Poisson-Boltzmann equation can be derived using additionally the Debye-Hückel theory of the ionic solutions. If the mobile ions in the solvent are also considered, they will arrange themselves in the electric field according to the Boltzmann distribution. Therefore, one additional term should be added to the Poisson equation to describe the charge density of all ions:

$$\rho_{ion}(\vec{r}) = \sum_s c_s^{bulk}(\vec{r})q_s \exp\left(-\beta q_s\phi(\vec{r})\right) \qquad (2.25)$$

The sum runs over all kind of ions s, where q_s and c_s^{bulk} are the charge and original concentration of ions s. β is equal $(k_BT)^{-1}$. Although it is not made explicit in the formula, note that the bulk ion concentrations depend upon position, possessing zero value inside the solute (protein) and having the bulk solution values outside. With that additional term the Poisson equation is transformed to the Poisson-Boltzmann equation (PBE):

$$\nabla\left[\varepsilon(\vec{r})\,\nabla\phi(\vec{r})\right] = -4\pi\left(\rho(\vec{r})+\rho_{ion}(\vec{r})\right)$$
$$\nabla\left[\varepsilon(\vec{r})\,\nabla\phi(\vec{r})\right] = -4\pi\left(\rho(\vec{r})+\sum_s c_s^{bulk}(\vec{r})q_s \exp\left(-\beta q_s\phi(\vec{r})\right)\right) \qquad (2.26)$$

If the solution contains only monovalent ions, the PBE becomes:

$$\nabla\left[\varepsilon(\vec{r})\,\nabla\phi(\vec{r})\right] = -4\pi\rho(\vec{r}) + 8\pi\,c_s^{bulk}(\vec{r})\sinh\left[\beta e\phi(\vec{r})\right] \qquad (2.27)$$

e is the unit charge and $\sinh(x) = 1/2(e^x - e^{-x})$. The PBE can be linearized by expanding the exponentials.

$$\sum_s c_s^{bulk}(\vec{r})q_s\exp\left(-\beta q_s\phi(\vec{r})\right) \approx \sum_s c_s^{bulk}(\vec{r})q_s - \beta\sum_s c_s^{bulk}(\vec{r})q_s^2\phi(\vec{r}) \qquad (2.28)$$

The first term vanishes because of overall electroneutrality of the ionic solution. Introducing the ionic strength $I(\vec{r})$ and the inverse Debye length $\kappa(\vec{r})$ will simplify the second term.

$$\sum_s c_s^{bulk}(\vec{r})q_s = 0$$

$$I(\vec{r}) = \frac{1}{2}\sum_s c_s^{bulk}(\vec{r})q_s^2 \qquad (2.29)$$

$$\kappa(\vec{r}) = \sqrt{8\pi\beta I(\vec{r})}$$

Finally, the linearized Poisson-Boltzmann equation (LPBE) gets the form:

$$\nabla\left[\varepsilon(\vec{r})\,\nabla\phi(\vec{r})\right] = -4\pi\rho(\vec{r}) + \kappa^2(\vec{r})\phi(\vec{r}) \qquad (2.30)$$

For a small value of the ionic strength, the error made by linearization is small and if $I(\vec{r}) = 0$, we get again the Poisson equation. The advantage of the LPBE is the additivity of electrostatic potentials and charge densities, what makes calculations faster and easier. A more detailed discussion about the PBE and the Debye-Hückel theory is given elsewhere, as for instance in chapter 15-1 (pp. 328-340) of McQuarrie (1976). This equation enables to compute the electrostatic potential of a macromolecule of known structure in solution. However, an analytical solution is possible only for very simple geometries, as for instance of a plane or sphere (Kirkwood, 1934; Daune, 1997). For proteins with very complex dielectric boundaries at the protein surfaces, it can be solved only numerically. The most common method, that is also applied in MEAD or MULTIFLEX program (Bashford & Gerwert, 1992; Bashford, 1997) and used in this doctoral work is a finite difference method (Warwicker & Watson, 1982; Nicholls & Honig, 1991; Honig & Nicholls, 1995). Alternatively but not so often, other numerical methods were applied as for instance boundary element methods (Sklenar et al., 1990; Zauhar & Varnek, 1996) or multigrid-based methods (Holst, et al., 1994; Holst & Saied, 1995).

Therefore, all relevant quantities, like partial charges, electrostatic potentials, dielectric constant and ionic strength are mapped on a cubic lattice with grid constant *l*, by linear interpolation. Thereby, the differential operators can be replaced by differences resulting in the finite difference form of the LPBE (Nicholls & Honig, 1991; Klapper, et al. 1986):

$$\phi_0 = \frac{\left(\sum_{i=1}^{6}\varepsilon_i\phi_i\right) + \frac{4\pi q_0}{1}}{\left(\sum_{i=1}^{6}\varepsilon_i\right) + 8\pi\beta I_0 l^2} \qquad (2.31)$$

Then, an iterative scheme is used to compute the electrostatic potential for each grid point. At the borders of the lattice, the grid points have less than six neighboring points. The problem

that appears is how to assign the values of the electrostatic potential to the edges of the grid. If the lattice is much larger then the molecule, it can be set to zero or evaluated according to the Debye-Hückel theory. However, than the resolution of the grid has to be poor due to the limitations of computer memory and CPU time. The solution to this problem is a focusing procedure, where the larger lattice of lower resolution is placed on the geometric center of the molecular system (and has to be at least two times larger then the molecule) and the smaller lattices with higher resolution are centered at the corresponding titratable groups (region of the interest). The smaller lattices are embedded in the larger lattice. On the boundaries of the finest lattice the electrostatic potential is taken by interpolation from the larger lattice. For very large macromolecules, the focusing can be done in several steps.

Artificial grid energy

The electrostatic energy G_{el} of a point charge q_i in its own electrostatic potential ϕ_i at the position of atom i is infinitely large, since the distance $\left|\vec{r} - \vec{r}\,'\right|$ in eq. 2.22 becomes infinitely small. This is the so called a self-energy. The divergence of the self-energy can be avoided, by solving the LPBE on a grid, since a point charge is smeared over the number of cubic grid parts. Nevertheless, the self-energy has then an arbitrary value, depending on the relative position and resolution of the grid. The remainder of the infinite self-energy, the so called grid energy artifact, has a finite but unknown value. One way to overcome the problem is to calculate electrostatic energy between the molecular systems where the grid energy is equal, so that the grid energy vanishes in a difference, rather than to calculate the absolute values of the electrostatic energy. Therefore, we computed the electrostatic energy difference between the investigated system with the dielectric constant $\varepsilon = 4$ and $\varepsilon = 80$. The difference is a grid energy free, since in a both calculations the charges, the grid resolution and position are unchanged.

Dielectric constant

In this approach, the solvation effects are modeled by a dielectric constant that depends on the position $\vec{r}$. So, the effect of the medium on the electrostatic interactions is described with an appropriate ε value. $\varepsilon(\vec{r})$ varies within the protein-solvent system and a choice of the dielectric constant value for a protein is a controversial topic, which depends very much on the experience (Warshel & Russel, 1984; Gilson & Honig, 1986; Harvey 1989; Warshel & Aqvist, 1991; Gilson 1995; Warshel & Papazyan, 1998).

Experimental and theoretical investigations suggests that if a "poor charge model" is used better results are obtained with a dielectric constant of 20 for a protein (Antosiewicz et al., 1994), but with a very "detailed charge model" a dielectric constant of 4 is more appropriate (Antosiewicz et al., 1996). In the first model, the unit charge is placed only on one atom of a charged titratable group, while in the second one the charge is distributed over all atoms of the titratable group. Also a large dielectric constant (ε=10-20) gives better agreement between calculation and experiments for the residues that are on the surface of a protein, while a small dielectric constant (ε= 2−4) explains better the behavior of the residues which are buried in a protein (Demchuck & Wade, 1996; Simonson et al., 1992; Simonson & Perahia, 1995a, b; Simonson & Brooks, 1996). That is not surprising since residues on the protein surface are partially exposed to or in direct contact with solvent molecules, while the residues inside a protein are everywhere surrounded by other protein residues. Further, a large

value of ε compensates for an unrealistic charge distribution used in a "poor model", but even with "detailed charge model" a dielectric constant of 2 or 4 has to be used instead of 1. The explanation is that a factor of 2 should compensate the effects of electronic polarization and another factor of 2 (or more) is used for the effects of the nuclear polarization, i.e. for the reorientation of permanent dipoles and displacement of atoms carrying atomic partial charges.

2.2.2   A single titratable group in a protein

The pK_a value of a titratable group in a protein can be quite different from the value in aqueous solution. To understand the shift of the pK_a value, one should look at the energetics of the protonation equilibrium of that group in different media. The thermodynamic cycle in Figure 2.2 shows the deprotonation of a titratable group in three media – gas phase, water and protein. In addition the solvation free energy of transferring the protonated or deprotonated species from one medium to another is also shown. Using this thermodynamic cycle, it is in principal possible to calculate the deprotonation free energy of a titratable group in a protein, if we know the experimental deprotonation free energy in water or the theoretically calculated value in the gas phase (Lim et al., 1991; Potter et al., 1994; Li et al., 1996; Richardson et al., 1997) and the energy needed to transfer the titratable residue from solution (gas phase) into the protein.
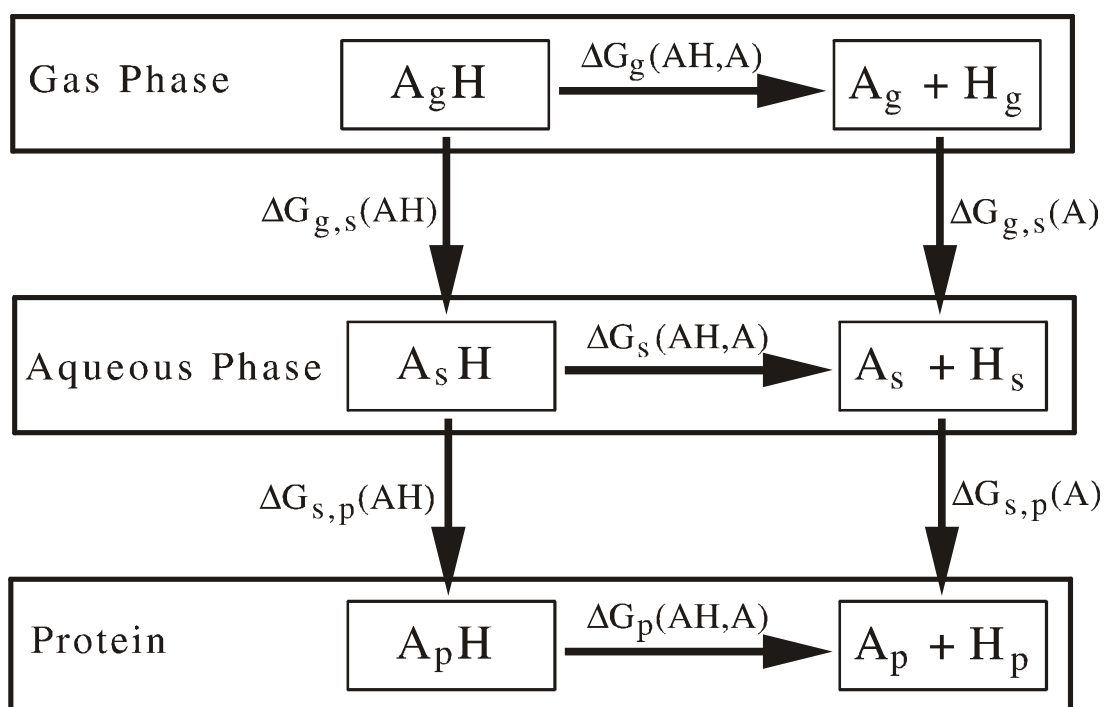

Figure 2.2: Thermodynamic cycle to calculate the protonation energy of titratable groups in solution and in protein from the gas phase properties. The deprotonation reaction and the corresponding deprotonation energy of an acidic group AH is given in three media: gas phase (g), aqueous solution (s) and protein (p). The solvation energies to transfer the titratable group AH from gas phase to solution (g,s) and further from solution into protein (s,p) are also shown.

If there is only a single titratable group in a protein, whose protonation is pH dependent, using the thermodynamic cycle, one can obtain the pK_a shift with respect to the pK_a value in aqueous solution. From eq. 2.10, one can deduce the deprotonation free energy in a protein or aqueous solution as:

$$\Delta G_x(AH, A) = G_x(A^-) - G_x(AH) = RT\ln 10(pH - pK_a^x)$$ (2.32)

where $x = p, s$ denotes protein or solvent. The pK_a shift is then:

$$\Delta pK_a = pK_a^p - pK_a^s = -\frac{1}{RT\ln 10}\left(\Delta G_p(AH, A) - \Delta G_s(AH, A)\right) =$$
$$= -\frac{1}{RT\ln 10}\left(\Delta G_{s,p}(A^-) - \Delta G_{s,p}(AH)\right) = -\frac{\Delta\Delta G}{RT\ln 10}$$ (2.33)

where double difference can be computed on two ways using the thermodynamic cycle.

$$\Delta\Delta G = \Delta G_p(AH, A) - \Delta G_s(AH, A) = \Delta G_{s,p}(A^-) - \Delta G_{s,p}(AH)$$ (2.34)

In this consideration, we assume that the pK_a shift in a protein is caused by electrostatic interactions between the titratable group and other charges in the protein and also by changing the dielectric medium. It is also assumed that the conformation of titratable group does not change during the proton transfer. The quantum-chemical contribution to the pK_a value does not depend on the actual dielectric medium. Therefore, it cancels in the double difference $\Delta\Delta G = \Delta G_{s,p}(A^-) - \Delta G_{s,p}(AH)$. Lacking the other titratable groups, which can simultaneously titrate and interact between themselves, makes expression 2.33 simpler.

The model compound

For electrostatic calculations of the protein, we need the pK_a values of the proper model compounds in water. Hence, we need to know the energy required to protonate a single titratable group in aqueous solution, as we have just seen. A model compound is the titratable group in solution (usually water) isolated from the rest of the molecule, but with the same conformation as in the protein. In the case of an amino acid, it is the N-formyl-N-methylamide derivative of that amino acid. It means that its N- and C-termini are blocked by derivatization, such that the only titratable site belongs to the side chain of that amino acid. The same corresponds to the situation of the amino acid within a protein, where amino- and carboxylate- groups are engaged in peptide bonds and only acidic or basic side chains are titratable. For amino acids, such model compounds can be synthesized and the corresponding pK_a values can be experimentally measured (Tanford, 1962), see Appendix D. However, sometimes it is not possible to synthesize the model compounds, as for example for some protein active centers with metals or there are no available experimental data. In these cases, the thermodynamic cycle could be also used to estimate the solution pK_a values, from the quantum mechanical computations of the protonation energies in vacuum (Richardson et al., 1997). Also some cofactors are not water soluble, therefore their pK_a values in water are not measurable. Then, one can use pK_a value of that compound in some other solvent, where it is soluble to obtain through the thermodynamic cycle the corresponding pK_a value in water (Kallies & Mitzner, 1997). The same is valid for the standard redox potential E^0 of a redox-

active group as model compound. Such experimental data are even less available and their water solubility is generally smaller. Typical examples are reduced quinones whose redox potentials are not measurable in water, since they would immediately take up a proton in aqueous solution, but there are measurements of their redox potentials in some aprotic solvents (acetonitrile, dimethylformamide). These informations can be used to evaluate their redox potentials in water (Rabenstein et al., 1998a,b). Using the quantum chemical calculations (Mouesca et al., 1994), the redox potentials of different Fe−S clusters were also estimated in aqueous solution.

2.2.3   Protonation state energy

If the difference between protonation energies of titratable groups in solution and in protein is only of electrostatic nature, then solving the LPBE is the way to evaluate these energy differences. Using the LPBE we are able to calculate the electrostatic potentials of each titratable group and from them to evaluate their protonation energies. Electrostatic potentials computed from the LPBE are additive as long as the dielectric boundaries of a macromolecule are the same. Thus, the conformation of titratable groups (or more general − the conformation of a protein) can not be changed during proton exchange. Due to the additivity of the potentials obtained from LPBE, it is possible to separate the individual contributions to the protonation energy of titratable groups and to treat them independently.

For each titratable group, the LPBE has to be solved four times. On that way, one gets the electrostatic potential caused by the charges of the protonated $\phi_m\left(\vec{r}_i; Q_\mu^h\right)$ and unprotonated $\phi_m\left(\vec{r}_i; Q_\mu^d\right)$ model compound μ (usually in water) and the equivalent protonated $\phi_p\left(\vec{r}_i; Q_\mu^h\right)$ and unprotonated $\phi_p\left(\vec{r}_i; Q_\mu^d\right)$ titratable group within the protein, where all other residues are in their reference charge state. As the reference state, one can choose the uncharged state of all titratable sites. These *4N* electrostatic potential terms (where *N* is the number of titratable sites) are used in the next steps to evaluate the protonation energies.

The protonation energy of a titratable group μ in a protein can be considered to consist of four contributions. These contributions for group μ are: the pK_a value of the model compound $pK_{a,\mu}^{model}$, the Born-solvation energy $\Delta\Delta G_\mu^{Born}$, the influence of background charges $\Delta\Delta G_\mu^{back}$ and the energy of coupling two charged sites $W_{\mu\nu}$. Now, I will explain in more details these energy terms.

As I have already shown for the protein with only a single titratable group, the pK_a value and the corresponding protonation energy shift can be obtained comparing the energetics of the protonation equilibrium in a protein with the one in solution. Similarly, a comparison can also be done for real proteins containing hundreds of titratable sites, where protonation depends on pH. Therefore, the first energy contribution is given by the pK_a value of a proper model compound of the titratable groups μ in aqueous solution, $pK_{a,\mu}^{model}$. Thus, this term defines the energy (see eq. 2.8) required to protonate or deprotonate an isolated titratable group μ in aqueous solution. The experimental $pK_{a,\mu}^{model}$ values needed for the titratable amino acids are given in Appendix D. Transferring the titratable group μ from aqueous solution into a protein in which all other titratable groups are in their uncharged reference state, will cause an energy shift, that can be further divided in two contributions. This is depicted in figure 2.3.

The Born energy term $\Delta\Delta G_\mu^{Born}$ arises from the interaction of the partial charges $Q_{i,\mu}$ of the titratable group μ with its reaction field, which is mediated by the dielectric medium. The sum runs over the $N_{Q,\mu}$ atoms of the group μ that have different charges in the protonated (h) $\left(Q_{i,\mu}^h\right)$ and in the deprotonated (d) $\left(Q_{i,\mu}^d\right)$ form. The terms $\phi_m\left(\vec{r_i};Q_\mu^h\right)$, $\phi_m\left(\vec{r_i};Q_\mu^d\right)$, $\phi_p\left(\vec{r_i};Q_\mu^h\right)$ and $\phi_p\left(\vec{r_i};Q_\mu^d\right)$ denote the electrostatic potentials at the position $\vec{r_i}$ of atom i in the protein (p) and in aqueous solution for the model compound (m). According to the Born theory (Born, 1920) each charged group in a polar solution enforces the solvent molecules to orient their dipoles in the electric field of the those charges. The influence of the resulting (induced) electric field of the solvent onto the charges is described as their reaction field. This energy term involves a double difference in electrostatic energy due to the change of the dielectric medium from aqueous solution to protein for the protonated and deprotonated form.

$$\Delta\Delta G_\mu^{Born} = \frac{1}{2}\sum_{i=1}^{N_{Q,\mu}} Q_{i,\mu}^h \left[\phi_p\left(\vec{r_i};Q_\mu^h\right) - \phi_m\left(\vec{r_i};Q_\mu^h\right)\right] - \frac{1}{2}\sum_{i=1}^{N_{Q,\mu}} Q_{i,\mu}^d \left[\phi_p\left(\vec{r_i};Q_\mu^d\right) - \phi_m\left(\vec{r_i};Q_\mu^d\right)\right] \qquad (2.35)$$
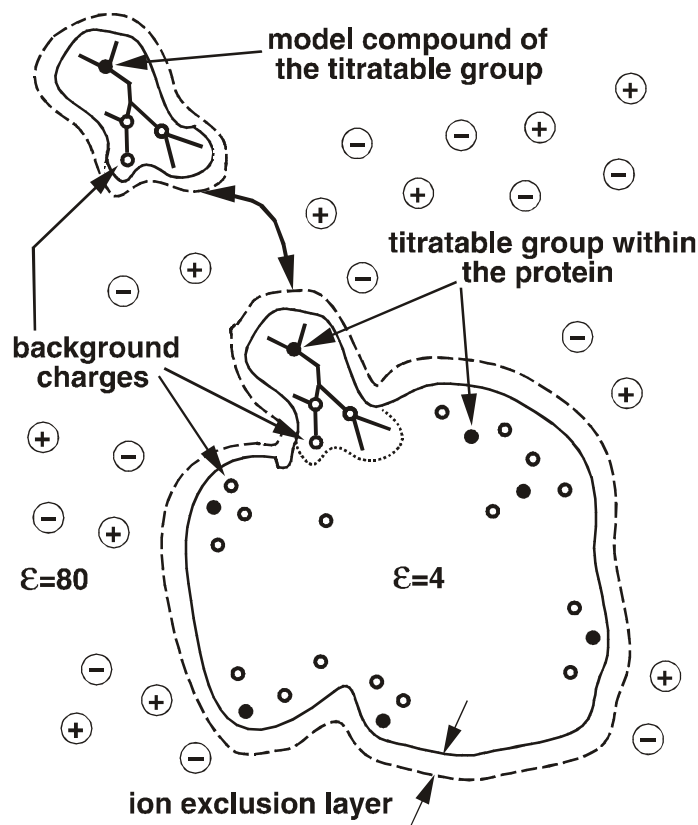

Figure 2.3: A titratable group in a protein and the corresponding model compound in ionic aqueous solution should have the same conformation. To avoid the grid energy artifact during the solvation of the LPBE, the thermodynamic cycle technique is applied. Transferring the titratable group from water into the protein the dielectric environment and the electrostatic interactions are changed. This causes a shift in the protonation energy, which can be divided in four contributions (see the text).

The $\Delta\Delta G_\mu^{back}$ energy term arises from the interaction of the charges of the titratable group μ with ''background charges'' – the charges of non-titratable residues and the charges of all other titratable residues ($\nu \neq \mu$) in their uncharged reference state. The first summation in eq. 2.36 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 state. The second summation runs over the N_m charges of atoms of the model compound μ that do not have different charges in the different protonation forms.

$$\Delta\Delta G_\mu^{back} = \sum_{i=1}^{N_p} q_i \left[\phi_p\left(\vec{r}_i; Q_\mu^h\right) - \phi_p\left(\vec{r}_i; Q_\mu^d\right)\right] - \sum_{i=1}^{N_m} q_i \left[\phi_m\left(\vec{r}_i; Q_\mu^h\right) - \phi_m\left(\vec{r}_i; Q_\mu^d\right)\right] \qquad (2.36)$$

These three energy contributions ($pK_{a,\mu}^{model}$, $\Delta\Delta G_\mu^{Born}$ and $\Delta\Delta G_\mu^{back}$) together yield the so-called intrinsic pK_a value (it is interesting to compare this equation with eq. 2.33):

$$pK_{a,\mu}^{intr} = pK_{a,\mu}^{model} - \frac{1}{RT\ln 10}\left(\Delta\Delta G_\mu^{Born} + \Delta\Delta G_\mu^{back}\right) \qquad (2.37)$$

The intrinsic pK_a value is the pK_a value that the titratable group μ in a protein would have, if all other titratable sites are in their uncharged reference protonation form. Therefore, the intrinsic pK_a values are only useful theoretical constructions rather then measurable quantities. Writing the eq. 2.37 in another form, where $\Delta G_\mu^{model} = RT\ln 10(pH - pK_{a,\mu}^{model})$:

$$RT\ln 10\left(pH - pK_{a,\mu}^{intr}\right) = \Delta G_\mu^{model} + \Delta\Delta G_\mu^{Born} + \Delta\Delta G_\mu^{back} \qquad (2.38)$$

and comparing with eq. 2.10, one can clearly see that transferring the titratable group from aqueous solution into a protein two additional contributions appear. They have been already discussed. Nevertheless, all three energy terms relate to the situation where the protonation of the titratable sites in a protein are considered independently of each other in the sense that all other sites are uncharged and neutral. However, the protonation behavior of the titratable sites depends strongly on neighboring charges. Specially, if titratable groups are close to each other, the coupling between them could be very strong. It means that we need one more energy term to describe the interactions between the titratable sites when they are charged. These energies are given in form of the W-matrix. The electrostatic interaction energy $W_{\mu\nu}$ of the titratable group μ and ν in their charged form is defined by equation 2.39. It is assumed that $W_{\mu\mu} = 0$.

$$W_{\mu\nu} = \sum_{i=1}^{N_{Q,\mu}} \left[Q_{\mu,i}^h - Q_{\mu,i}^d \right]\left[\phi_p\left(\vec{r}_i, Q_\nu^h\right) - \phi_p\left(\vec{r}_i, Q_\nu^d\right)\right] \qquad (2.39)$$

Finally, using the energy terms $pK_{a,\mu}^{intr}$ and $W_{\mu\nu}$, it is possible to compute the energy of each protonation state n of the protein molecule. The sum in eq. 2.40 runs over all N titratable groups.

$$G_n = \sum_{\mu=1}^{N} (x_\mu^n - x_\mu^0)\left(RT\ln 10\left(pH - pK_{a,\mu}^{intr}\right)\right) + \frac{1}{2}\sum_{\mu=1}^{N}\sum_{\nu=1}^{N}\left(W_{\mu\nu}\left(x_\mu^n + z_\mu^0\right)\left(x_\nu^n + z_\nu^0\right)\right) \qquad (2.40)$$

A particular protonation state is characterized by the state vector $\vec{x}^n = (x_1^n, x_2^n, x_3^n, ..., x_N^n)$, where x_μ^n is 1 or 0 depending on whether a group μ is protonated or unprotonated. z_ν^0 is the

unitless formal charge of the deprotonated form of the group μ, i.e. –1 for acids and 0 for bases. The term x_μ^0 in the first summation refers to the uncharged – reference state. Its value is 1 for acids and 0 for bases. The eq. 2.40, one can write in more compact way,

$$G_n = \sum_{\mu=1}^{N} (x_\mu^n - x_\mu^0) \left(RT \ln 10 \left(pH - pK_{a,\mu}^{intr} \right) + \sum_{\nu>\mu}^{N} \left(W_{\mu\nu}(x_\nu^n - x_\nu^0) \right) \right) \qquad (2.41)$$

where it becomes obvious that for the reference state is $G_n = 0$, which is accordingly the zero point of the system. The first factor $(x_\mu^n - x_\mu^0)$ is positive for protonated base μ and negative for deprotonated acid μ. It also takes care that the second term is negative for attractive interactions (charged acid-base pair) and positive for repulsion (two charged acids or two charged bases).

If *N* is a number of the titratable groups in a protein, and assuming that each group μ has only two protonation states, the total number of possible protonation states of a protein is 2^N. Using the equation 2.40, one can calculate the protonation energy G_n of any protein state, without to solve the Poisson-Boltzmann equation 2^N times. Instead of that, it is enough to solve the LPBE *4N* times as it was explained before. Now, it becomes clear how much savings of computer power is possible by applying the LPBE.

2.2.4   Redox reactions

Due to the similarity of protonation and redox reactions, similar expressions can be derived for redox-active groups. Let us consider a system with *N* acid-base titratable groups and *K* redox-active groups. Because the protonation state of acid-base groups depends on pH and the oxidation state of redox groups depends on the solution redox potential *E_{sol}*, the charge and the energy of the system will depend on both of them. It means for a complete description of such system additional terms are needed. For the energy calculation of a particular protonation and redox state of the protein, beside the already defined terms: the pK_a value of the proper model compounds ($pK_{a,\mu}^{model}$), the difference between the pK_a value of model compound and the pK_a value of the protonatable group in the reference charge state of the protein ($pK_{a,\mu}^{intr}$) and the interaction between the protonatable groups ($W_{\mu\nu}$), we also need: the standard redox potential of the proper model compounds ($E_{model,\eta}^0$) for the redox-active groups, the difference between the E^0 value of model compound and the E^0 value of the redox group in the reference state of the protein ($E_{intr,\eta}^0$), the interaction between the redox groups ($U_{\eta\chi}$) and the interaction between the redox-active groups and the protonatable groups ($V_{\eta\mu}$).

The intrinsic standard redox potential of the redox-active group η is defined similar to eq. 2.37. It is the standard redox potential that the redox-active group η would have if all other titratable groups are in their reference charge state.

$$E_{intr,\eta}^0 = E_{model,\eta}^0 + \frac{1}{F} \left(\Delta\Delta G_{redox,\eta}^{Born} + \Delta\Delta G_{redox,\eta}^{back} \right) \qquad (2.42)$$

The energy term $\Delta\Delta G_{redox,\eta}^{Born}$ describes the interaction of the charges of the redox-active group η with its reaction field.

$$\Delta\Delta G_{redox,\eta}^{Born} = \frac{1}{2}\sum_{i=1}^{N_{Q,\mu}} Q_{i,\eta}^{oxd}\left[\phi_p\left(\vec{r_i};Q_\eta^{oxd}\right)-\phi_m\left(\vec{r_i};Q_\eta^{oxd}\right)\right]-\frac{1}{2}\sum_{i=1}^{N_{Q,\mu}} Q_{i,\eta}^{red}\left[\phi_p\left(\vec{r_i};Q_\eta^{red}\right)-\phi_m\left(\vec{r_i};Q_\eta^{red}\right)\right]$$

(2.43)

The term $\Delta\Delta G_{redox,\eta}^{back}$ represents the interaction energy of the charges of the redox-active group η with background charges of all nontitratable residues and the charges of reference form of all other titratable groups.

$$\Delta\Delta G_{redox,\eta}^{back} = \sum_{i=1}^{N_p} q_i\left[\phi_p\left(\vec{r_i};Q_\eta^{oxd}\right)-\phi_p\left(\vec{r_i};Q_\eta^{red}\right)\right]-\sum_{i=1}^{N_m} q_i\left[\phi_m\left(\vec{r_i};Q_\eta^{oxd}\right)-\phi_m\left(\vec{r_i};Q_\eta^{red}\right)\right] \quad (2.44)$$

The meaning of the symbols in eqs. 2.45 and 2.46 is analogous to the meaning in eqs. 2.35 and 2.36.

The interaction $U_{\eta\chi}$ between the redox-active groups η and χ is defined by eq. 2.47:

$$U_{\eta\chi} = \sum_{i=1}^{N_{Q,\eta}}\left[Q_{\eta,i}^{oxd}-Q_{\eta,i}^{red}\right]\left[\phi_p\left(\vec{r_i},Q_\chi^{oxd}\right)-\phi_p\left(\vec{r_i},Q_\chi^{red}\right)\right]$$

(2.45)

The interaction $V_{\eta\mu}$ between the redox-active groups η and titratable group μ is defined as:

$$V_{\eta\mu} = \sum_{i=1}^{N_{Q,\eta}}\left[Q_{\eta,i}^{oxd}-Q_{\eta,i}^{red}\right]\left[\phi_p\left(\vec{r_i},Q_\mu^{h}\right)-\phi_p\left(\vec{r_i},Q_\mu^{d}\right)\right]$$

(2.46)

$$= \sum_{i=1}^{N_{Q,\mu}}\left[Q_{\mu,i}^{h}-Q_{\mu,i}^{d}\right]\left[\phi_p\left(\vec{r_i},Q_\eta^{oxd}\right)-\phi_p\left(\vec{r_i},Q_\eta^{red}\right)\right]$$

Now, we have to include these additional energy terms to describe the energy G_n of a particular protonation and redox state n of a protein. One can see that this energy depends on pH and redox potential of solution E_{sol}.

$$G_n = \sum_{\mu=1}^{N}(x_\mu^n - x_\mu^0)\left(RT\ln 10\left(pH-pK_{a,\mu}^{intr}\right)\right)+\frac{1}{2}\sum_{\mu=1}^{N}\sum_{\nu=1}^{N}\left(W_{\mu\nu}\left(x_\mu^n+z_\mu^0\right)\left(x_\nu^n+z_\nu^0\right)\right) \quad (2.47)$$

$$-\sum_{\eta=1}^{K}(x_\eta^n - x_\eta^0)\left(F\left(E_{sol}-E_{intr,\eta}^0\right)\right)+\frac{1}{2}\sum_{\eta=1}^{K}\sum_{\chi=1}^{K}\left(U_{\eta\chi}\left(x_\eta^n+z_\eta^0\right)\left(x_\chi^n+z_\chi^0\right)\right)$$

$$+\sum_{\eta=1}^{K}\sum_{\mu=1}^{N}\left(V_{\eta\mu}\left(x_\eta^n+z_\eta^0\right)\left(x_\mu^n+z_\mu^0\right)\right)$$

x_η^n is 1 or 0 depending on whether a group η is oxidized or reduced. $z_\eta^0 = 0$ is the unitless formal charge of the reduced form of the redox group η. As the reference state of redox-active groups, we took the reduced form, such that $x_\eta^0 = 0$. All other symbols have the same meaning as before.

2.2.5 *Protonation and oxidation probability*

For given a system with N protonatable and K redox-active groups, their protonation and oxidation probabilities can be calculated, as follows. Each protonation and oxidation state n of a protein can be described by $N+K$-component vector $\vec{x}^n = (x_1^n, x_2^n, ..., x_{N+K}^n)$ and

protonation/oxidation free energy G_n. The component x_μ^n represents the protonation or oxidation state of the group μ, that it has in the protein charge state n. Then, protonation (oxidation) probability of group μ is calculated as the thermodynamic average over all 2^{N+K} possible protonation and oxidation states of the protein:

$$\langle x_\mu \rangle = \frac{\displaystyle\sum_{n=1}^{2^{N+K}} x_\mu^n \exp\left(-\frac{G_n}{RT}\right)}{\displaystyle\sum_{n=1}^{2^{N+K}} \exp\left(-\frac{G_n}{RT}\right)} \tag{2.48}$$

x_μ^n is 1 or 0 depending on whether a group μ is protonated (oxidized) or unprotonated (reduced) in nth protonation and oxidation state of a protein.

Computing the G_n energies at different pH or solution redox potential and from them protonation (oxidation) probabilities $\langle x_\mu \rangle$, one can obtain titration curves as the thermodynamic average of the protonation (oxidation) of group μ. The titratation curves can be further used to obtain the $pK_{1/2}$ or $E_{1/2}^0$ values. Also the dependence of redox potential of a redox-active group on the solution redox potential or pH can be evaluated. All other energetic parameters are available, too. (For more discussion see next sections). For instance, from the probability $\langle x_\mu \rangle$, it is possible to calculate the free energy required to protonate or oxidize group μ at given pH, temperature and redox potential of the solvent E_{sol}:

$$G_\mu(pH, E_{sol}) = -RT \ln \frac{\langle x_\mu \rangle}{1 - \langle x_\mu \rangle} \tag{2.49}$$

G_μ represents the free energy of group μ thermodynamically averaged over all protonation and redox states of the protein.

2.2.6   *The coupling of protonation and redox reactions*

Intending to evaluate any of the relevant energetic parameters, the protonation and redox equilibria in a protein have to be considered simultaneously. The coupling between protonation and redox reactions plays a significant role in many biochemical important charge transfer processes. It is often caused by electrostatic interactions between protonatable and redox-active groups. Namely, the positive charge of a proton can neutralize the negative charge of an electron. Thus, the reduction of a redox-active group can go together with protonation of neighboring acid-base groups, increasing their pK_a values. Or the oxidation of a redox-active group can lead to the deprotonation of adjacent protonatable groups, decreasing their pK_a values. Thereby, the redox-active and the protonatable group may belong to the same or to the two different molecular groups. This phenomenon, the electrochemical coupling is also known as redox Bohr effect (Wyman, 1968; Kilmartin & Bernardi, 1973; Dutton & Wilson, 1974; Urban & Klingenberg, 1969; Soares et al., 1997; Papa et al., 1979).

The coupling between protonation and redox reaction can be investigated by this approach. As one can see, the first summation in eq. 2.47 runs over N protonatable groups and it is the only term where the pH value of the solution appears. The third summation relates to the K redox-active groups and only there the solution redox potential E_{sol} appears as variable. So, one can wrongly conclude that only protonation probabilities and titration behavior of

acidic and basic groups depend on pH and that only oxidation probabilities of redox-active groups depend on the solution redox potential E_{sol}. Thereby, increasing pH will cause deprotonation of protonatable groups, and increasing the solution redox potential E_{sol} will oxidize the redox-active groups. But due to the mutual coupling between protonatable and redox groups (see eq. 2.47), oxidation probabilities are also pH dependent and protonation probabilities can depend on the solution redox potential E_{sol}. Namely, increasing pH causes the deprotonation of nearby protonatable groups, the total charge of a protein becomes more negative, that will stabilize the oxidized form, and thus the standard redox potential of the redox-active group η decreases. The redox-active groups are usually surrounded by many protonatable groups in a protein and even when the mutual coupling is not very strong the total effect of changing the charges of adjacent protonatable groups with pH will cause changes of redox potential. On other side, a few redox-active groups, changing the oxidation states due to the change of solution redox potential can influence the protonation probabilities of the titratable groups μ, that are in their vicinity and couple strongly with them. Therefore, the number of protonatable groups that depend on the solution redox potential is usually not so large. Thereby, the pK_a values of the protonatable groups drop by increasing the solution potential E_{sol}. The more redox-active group is oxidized, the harder it can protonate nearby titratable groups.

Hence, the mutual coupling of protonatable and redox-active groups is responsible for pH dependence of the redox potentials, as well as for the redox Bohr effect, i.e. for the dependence of protonation probabilities (and pK_a) of titratable groups on the solution redox potential.

2.2.7 pH dependent pK_a values and the solution redox potential dependent E^0 values

The mutual coupling between titratable sites has an influence on the titration behavior and the shape of the titration curves. The protonation or oxidation probabilities of the titratable groups are pH or redox potential dependent. It causes, that their mutual interactions are also dependent on the pH and solution redox potential. Because of that the titration curves can deviate significantly from the standard sigmoidal shape of the Henderson-Hasselbach or Nernst titration curves of isolated protonatable or redox-active groups. Thus, beside protonation and oxidation probabilities, also pK_a and E^0 values depend on the pH or solution redox potential. In some cases it is not possible to assign a unique redox potential E^0 or pK_a value to a specific group. Instead of that, very often the $E^0_{1/2}$ or $pK_{1/2}$ values are used. These quantities correspond to the pH or solution redox potential E_{sol} values at which one half of variably charged group is neutralized (deprotonated or reduced). In the case of an isolated titratable group $pK_a = pK_{1/2}$ and $E^0 = E^0_{1/2}$ is valid, and knowing these values one can reproduce the complete titration curve as a function of *pH* or E_{sol} range. But in the protein molecules, due to the mutual interactions of titratable groups, the titration curves show deviations from the standard behavior and $pK_{1/2}$ or $E^0_{1/2}$ values are not directly related to the energy required to protonate or to oxidize a variably charged group (eq. 2.49).

These deviations require the definition of pK_a and E^0 values that depend on pH or the solution redox potential.

$$pK_{a,\mu} = pH + \frac{1}{\ln 10} \ln \frac{\langle x_\mu \rangle}{1 - \langle x_\mu \rangle} \tag{2.50}$$

$$E_\eta^0 = E_{sol} - \frac{RT}{F} \ln \frac{\langle x_\eta \rangle}{1 - \langle x_\eta \rangle} \tag{2.51}$$

The above expressions for $pK_{a,\mu}(pH, E_{sol})$ and the standard redox potential $E_\eta^0(pH, E_{sol})$ are appropriate, if one likes to characterize the energetics of protonation and redox reactions in proteins in all detail, throughout the whole range of pH and solution redox potential. For several mutually interacting variably charged groups, the $pK_{a,\mu}$ and E_η^0 quantities in equations (2.50) and (2.51) possess explicit and implicit pH and E_{sol} dependences. The implicit dependence comes though the probabilities $\langle x_\mu \rangle$, which are also pH and E_{sol} dependent. Note that for an isolated variably charged group μ or η, the expressions (2.50) and (2.51) are equivalent to the conventional definitions of the $pK_{a,\mu}$ value and the standard redox potential. There, the direct and indirect dependences cancel yielding pK_a and E^0 values that are constant. This means that only the whole titration curves or the curves that show the dependence $pK_{a,\mu} = f(pH)$ or $E^0 = f(E_{sol})$ can completely describe the energetics of a titratable group within a protein.

2.3   Monte Carlo titration

2.3.1   *The sampling method*

The sum in eq. 2.48 can be evaluated explicitly only for a very limited number of titratable groups. Already with 40 titratable sites, the exact summation requires too much CPU time. Many proteins contain several tens or hundreds of titratable groups, what means that a direct evaluation of the Boltzmann averaged sum of all 2^N protein states is unfeasible. The number of possible states, whose energies have to be computed and summed up, becomes very large if multiple conformations are considered. For all these reasons, several approximation methods have been developed: the Tanford-Roxby approximation, which is a mean field approximation to the exact theory (Tanford & Roxby, 1972), the reduced site approximation (Bashford & Karplus, 1991), a hybrid statistical mechanical/Tanford-Roxby approximation (Yang et al., 1993), the cluster method (Gilson, 1993) and the Monte Carlo (MC) method (Beroza, et al., 1991). A review of all these different methods can be found in Ullmann & Knapp, 1999. Here, I will explain in more detail, only the Monte Carlo (MC) method, which I used in this work for the calculation of titration curves.

The MC titration is a method to calculate the protonation pattern of a protein without to sum up all protonation states. Instead of that only a subset of all possible protonation states is sampled according to the Metropolis criterion (Metropolis, et al., 1953). This sampling procedure is valid, if all states, which are likely to occur are sampled and most of the states which are unlikely to appear are not considered. Thereby, the protonation states have to be sampled with the probability with which they occur (*importance sampling*). In this way, the ensemble of sampled states represents the whole ensemble correctly. The protonation probability $\langle x_\mu \rangle$ of group μ is then obtained by averaging x_μ over all sampled states.

In our calculations, we used the MC method, developed by Beroza, et al. 1991, which is implemented in the program KARLSBERG (Rabenstein, 1999). The protonation state vector $\vec{x}$ can be initialized randomly or on the basis of the intrinsic pK_a values (as in KARLSBERG). The latter should prevent to start the MC sampling with an unrealistic protonation pattern. An MC

move is an attempt to change the protonation state of one randomly chosen titratable group μ with subsequent evaluation of the Metropolis criterion. The corresponding change in free energy ΔG_μ before and after the MC move is computed as:

$$\Delta G_\mu = \Delta x_\mu \left(RT \ln 10 (pH - pK_{a,\mu}^{intr}) + \sum_{\nu=1}^{N} W(x_\nu + z_\nu^0) \right) \qquad (2.52)$$

where $\Delta x_\mu = x_\mu^{new} - x_\mu^{old}$ can adopt the value ± 1. Then, the Metropolis criterion is applied to decide whether this MC move is accepted. If the new energy is lower than the old one, $\Delta G_\mu \leq 0$, the move is accepted and the protonation state of group μ is changed. Otherwise $(\Delta G_\mu > 0)$, the new protonation state of site μ is accepted with probability $\exp(-\Delta G/RT)$. One MC scan is completed after *N* attempts (moves) to change the protonation state. Thus on the average one MC move per each titratable group is performed. Usually a few hundred of MC scans are needed to equilibrate the system. After that, the results from the MC scans are accumulated and after applying a large enough number of the MC scans and if the sampling efficiency is satisfactory, the Boltzmann weighted ensemble is generated. Then, the protonation states of all scans are used to evaluate the average protonation of each titratable site.

The program KARLSBERG (that we used) is based on the same method as the original MCTI program from Beroza et al. (1991) but has some additional features (Rabenstein & Knapp, 2001). It can treat the multiple conformations of a protein, has in addition double and triple MC moves included, can perform the energy-biased MC and a parallel tempering, uses the reduce site approximation to avoid in the Boltzmann sum the states that are always protonated or deprotonated by fixing the corresponding protonation states. This method was tested on many protein systems and it is able to calculate titration curves for very large molecules, which contain even more than 650 titratable sites. But, as in other similar methods, one has to be aware that the obtained results are strictly correct only for pH value at which the protein structure has been solved. Namely, the change of pH can implicate conformational changes of the structure, which are not included in these computations. A change of the conformation will cause changes of the electrostatic boundary that can have an influence on the pK_a values.

2.3.2   Additional features

Using reduced MC scans

A set of MC moves, where it is attempted to change the charge state on the average at all variably charged groups, is called an MC scan. Also it comprises all necessary double and triple moves. To reduce the CPU time and to avoid unnecessary calculations of energies for residues, which are always protonated or deprotonated, the MC titration is combined with the reduced site approximation. Usually for each MC titration, we performed first 1000 full MC scans. After the full MC titration has finished, all titratable groups, that did not change their protonation state, were fixed in their respective protonation state and excluded from further sampling. Then, 10000 reduced MC scans are performed without those sites. This MC titration procedure leads to a standard deviation of less than 0.01 protons at each titratable group.

Double and triple moves

Sometimes the protonation state of two or three titratable groups is not independent from each other. That situation appears if two (three) groups are strongly coupled. For instance, two acidic groups in close neighborhood, where one proton can be either on the one or on the other group, but where is energetically unfavorable that both of them are protonated or deprotonated. Thus, the changes of the protonation states at these two groups strongly depend on each other and applying the single MC move, where the protonation of only one of these two sites is changed, will almost always be rejected, based on the Metropolis criterion ($\Delta G_\mu > 0$). Thus, an exchange between these two energetically favorable states will rarely occur, because the intermediate state between the two low energy states has a high energy, rendering the transition from one to another state unlikely. The problem can be avoided when the protonations of the two groups are switched simultaneously, which correspond to direct proton exchange between these two groups. Hence, only double (triple) moves can treat the strongly coupled groups, correctly (see Figure 2.4).

To improve the sampling efficiency and to prevent sampling problems of the MC titration, also two (three) variably charged groups that couple stronger than 2.5 (5.0) pK_a units will change their charge state simultaneously in one MC move. Such double and triple MC moves were done in addition to simple moves and also were evaluated according to the Metrpolis criterion.
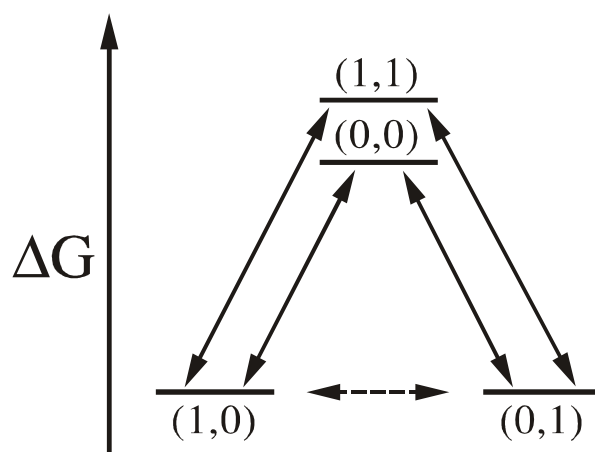

Figure 2.4: Treatment of two strongly coupled sites in the MC titration. The protonation states (1,0) and (0,1) are separated with a high energy barrier (states: (1,1) and (0,0)). A change of the protonation of only one group (solid arrows) would lead to a state with high energy. Only double MC move can simultaneously switch the protonation state of both sites (dashed arrow) and require little energy change.

Two protonation modes of histidine

While it is not so critical for aspartate, glutamate, lysine or arginine on which atom of titratable group is exactly located proton that will be released by deprotonation, in the case of histidine that is very important. For the calculations of the acidic and basic groups, a fraction of a proton charge is equally distributed at all equivalent positions (both carboxylate oxygens of Asp, Glu or all three amino-hydrogens of Lys). Hence, the symmetrical charge distributions for the protonated forms of aspartic and glutamic acids effectively average the

two symmetrical protonated forms. Such symmetric charge models for titratable groups are helpful to avoid biased hydrogen bonding and biased protonation patterns, which may occur with the usage of a single molecular conformation.

A more significant problem appears with histidine, since the two potential protonation sites are chemically distinct, and since steric interactions are more likely to prevent side chain rotations that would average the charge distributions. The protonated non-ligated histidine possesses two non-equivalent hydrogens ε-H and δ-H and both can be deprotonated. Even their pK_a values are not exactly the same (6.6 and 7.0). Thus, in the deprotonated-neutral state, histidine exists in two tautomeric forms. Including the double deprotonated state, which is however very unlikely, there are four possible protonation states for each histidine. The program MEAD (MULTIFLEX) (Bashford & Gerwert, 1992; Bashford, 1997) available to calculate the electrostatic energies of titratable groups in different protonation states cannot handle titratable groups with more than two protonation states. In the following, I will outline a way in which this tautomeric freedom can be incorporated into the usual framework of protein titration calculations.

Because the titratable $\delta-$ and $\varepsilon-$pseudo-groups of histidine have shared atoms, the electrostatic computations can not be done for both forms independently. It means that the W-matrix has to be extended to include the interactions with δ- and ε-tautomers. In our calculations, both histidine forms are taken into account by representing them as two coupled titratable sites, which cannot simultaneously be deprotonated. The pair of the deprotonated forms of histidine cannot be present at the same time, therefore we have to avoid it. The way to do it, is to set up coupling constant $W_{\delta\varepsilon}$ i.e. the interaction energy between two sites of the same histidine residue to a very high value and to adjust the pK_a values of the two forms accordingly. At the same time the pK_a values of all other titratable groups have to be corrected, because their W-interaction terms with the two protonation modes of histidines are also doubled. On that way, we account for the fact that other charges in the protein see only a single histidine, and not two sites. The corrected W-matrix and corrected $pK_{intr,\mu}$ values (so called global-values) can be used in the MC titration, to obtain protonation probabilities of all titratable groups. Then, the fraction of $\delta-$ and $\varepsilon-$tautomers are directly obtained from the protonation probabilities computed by the MC titration. The fraction of the double protonated histidine form, one calculates as a difference between one and the sum of probabilities of the two tautomers.

Hence, in order to avoid that the $\delta-$ and $\varepsilon-$tautomer of histidine appear at the same time, since it means that histidine is then double protonated, we have to treat the two equilibria on a special way. Applying a large value of the $W_{\delta\varepsilon}$ interaction between the $\delta-$ and $\varepsilon-$sites (when they are in their charged state) will hinder the appearance of the two pseudo-tautomers at the same time. However, we do not like that the energetically very unfavorable double deprotonated state occurs at all. That will be avoided by increasing the pK_a values of the corresponding protonation equilibria. Interestingly, figure 2.4 gives also the explanation of this problem. Applying a large $W_{\delta\varepsilon}$ interaction between two states, they become two strongly coupled groups and the energy barrier between them becomes larger. Increasing the pK_a values of the corresponding eqilibria, a double protonated state will be shifted toward a lower energy, but unlikely double deprotonated state will still stay up in energy. Finally, it means that only three charge states of histidine (an double protonated and the two neutral states) can be reached by a MC sampling. With the explained procedure, standard titration programs that expect two-state models can be used directly to handle this more complex behavior.

The same approach could be easily extended for other residues. To find out whether the protonation equilibrium of diphosphate acid (DPA) from FAD in the photolyase predominantly occupies the protonation states DPA^0/DPA^- or DPA^-/DPA^{-2}, we considered both equilibria separately and described the two different protonation states of DPA^- in the same way as we did it for histidine. The figure 2.5 shows several examples where we used

this technique and explains in which direction the pK_a value should be shifted, to prevent the appearance of energetically unfavorable protonation states.
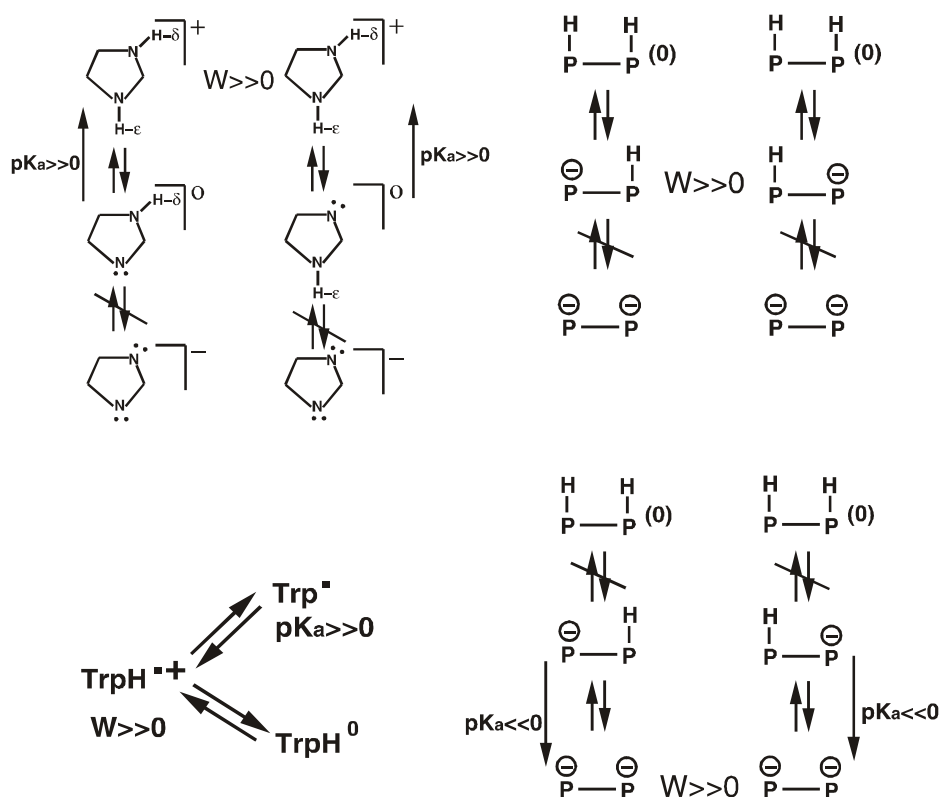

Figure 2.5: Some examples of a treatment for the residues with more than two possible protonation states: (a) histidine, (b) coupling the protonation and redox reaction for thryptophan, (c) diphosphate acid – DPA^0/DPA^- equilibrium, (d) diphosphate acid – DPA^-/DPA^{-2} equilibrium. Unwanted or energetically unfavorable states are avoided by applying a very large or a very small pK_a value ($pK_a \gg 0$ or $pK_a \ll 0$), and using a large $W_{\delta\varepsilon}$ interaction.

Energy biased MC

The protonation and oxidation probabilities calculated with MC sampling can be used to obtain the protonation and oxidation free energy (energy required to protonate/oxidate a titratable group) by eq. 2.49, or the $pK_{a,\mu}(pH, E_{sol})$ and standard redox potential $E_\eta^0 (pH, E_{sol})$ can be evaluated by eq. 2.51 and 2.52. In case that a super-residue has been used the eq. 2.49 yields the reaction or transition free energy, where $\langle x_\mu \rangle$ and $1 - \langle x_\mu \rangle$ are the average occupancies of the final and initial state, respectively. A super-residue is a technical solution to treat a protonation or ET reaction by grouping the atoms of the two participating residues in one new super-residue. It contains the atoms of the two residues involved in the reaction equilibrium, where proton or electron can be located on the one or the other residue. Hence, it describes the two possible variably charged states of such system of the two residues. Notice,

that state 1 and state 2 possess the same total charge, since it is summed over all atoms of both residues. In this case, the reference state could be any of them, but by convention the program chooses state 1 as reference state.

If the probability $\langle x \rangle$ is close to zero or one, the reaction equilibrium is shifted very much in one direction and the free energy difference ΔG would have a large positive or negative value. Then, even a small statistical error in occupation probability of the MC sampling would lead to a large statistical error of the ΔG, $pK_{a,\mu}$ and E_η^0 values. In these situations, we applied a bias energy to that specific site and repeated the MC sampling. The bias energy is added to the protonation energy of the titratable group to shift the equilibrium from one protonation state toward another. The statistical error of the calculated energy is minimal, when the bias is chosen such that the protonation probability is close to 0.5. Finally, the bias has to be subtracted from the calculated value to get the original result but with a strongly reduced statistical error.

Statistical error of the MC titration

The statistical uncertainty of the protonation probability x_μ calculated by the MC sampling is estimated by a correlation function (Beroza et al., 1991):

$$C_\mu(\tau) = \frac{1}{T-\tau} \sum_{t=0}^{T-\tau-1} x_\mu(t+\tau)x_\mu(t) - \left\langle x_\mu \right\rangle^2 \qquad (2.53)$$

where t is the time in units of the MC scans, T is the total number of scans and τ is the time variable for the correlation function. For each site μ, the correlation function $C(\tau)$ determines the correlation time τ_μ^{corr} between approximately independent MC moves. τ_μ^{corr} is the time for which $C_\mu(\tau)$ becomes negligible (here: $\left| C_\mu(\tau) \right| < 0.1 C_\mu(0)$). The variance of one measurement $C_\mu(0)$ divided with number of independent measurements gives the standard deviation:

$$\sigma_\mu = \sqrt{\frac{C_\mu(0)}{T/\tau_\mu^{corr}}} \qquad (2.54)$$