Center for Structural Biochemistry
Department of Biosciences at NOVUM
Karolinska Institutet, S-141 57 Huddinge, Sweden

Theoretical Prediction of Ionisation Properties of Proteins

Assen Koumanov

Stockholm 2003

Published and printed by Karolinska University Press
Box 200, SE-171 77 Stockholm, Sweden

ISBN 91-7349-535-2

Contents

Abstract

This work emphasises on elaboration of improved theoretical and computational methods for determination of protonation/deprotonation equilibria in proteins. It also aims to contribute for a better understanding on atomic level of ionisation properties of proteins and their role in structure-functional relationship. Presented here computations are based on a continuum model of protein-solvent system.

Limitations in the predictive power of current methods for pK_a calculations are mostly due to inadequate accounting of the structural flexibility, which is a key element of functional properties of proteins. This problem was approached in two ways. First, by combining calculations of ionisation equilibria with different techniques for generation of ensembles of protein structures – alternative conformers in X-ray structures, NMR models, molecular dynamics simulation. Second, the theoretical description of the pK_a calculations based on continuum electrostatic model was generalized in order to account explicitly for alternative hydrogen locations on titratable and polar non-titratable groups. This corresponds to an introduction of minor but very important structural flexibility. Such approach not only improves the accuracy of the pK_a calculations but also provides detail information that allows, for instance, prediction of pH-dependence of tautomerisation as well as of a reorganisation of H-bond networks. Entropic effects were pointed out.

Irregular (non-sigmoid) titration curves of ionisable groups were analysed theoretically. It was demonstrated that the pK_a values, extracted from multiple-step titration curves by means of fitting to a sum of Henderson-Hasselbalch equations, do not describe the ionisation equilibrium correctly, which may lead to irrelevant conclusions for the functional mechanisms. Conditions for appearance of irregular titration were derived analytically.

Continuum model was applied to simulate the non-equilibrium stationary process of steady-state ion flux through protein channels. The Poisson-Nernst-Planck (PNP) equations were modified to account for desolvation of mobile ions in the membrane pore and for effects related to ion sizes. A numerical algorithm was developed for 3D solution of PNP equations, applicable for arbitrary channel geometry and arbitrary protein charge distribution. Basic features of ion transport were illustrated by simulations on model channels.

Studied proteins: ribonuclease T_1 from the fungus *Aspergillus oryzae* (RNase T_1), alcohol dehydrogenase from *Drosophila lebanonensis* (DADH), xylanases from *Bacillus circulans* and *Bacillus agaradhaerens*, porin Omp32 from the bacterium *Delftia acidovorans*

Ionisation properties of individual titratable side chains of RNase T_1 were explored in detail both experimentally by NMR spectroscopy and theoretically by pK_a calculations. The study revealed a novel interpretation of the observed pH dependence of the chemical shifts of several residues. It was also shown that titration of Asp76 is coupled to dipole reorientation of a bounded water molecule, which suggested an interpretation of presumably contradicting experimental observations.

Theoretical investigations of DADH showed that ***(i)*** the protonation/deprotonation transition of the binary complex is related to the coupled ionisation of Tyr151 and Lys155 in the active site and ***(ii)*** the pH-dependence of the proton abstraction is correlated with a reorganization of the hydrogen bond network in the active site involving also the O2' ribose hydroxyl group from the co-enzyme, which acts as a switch.

pK_a calculations were combined with 1 ns MD simulations based on three different initial structures of xylanase. A tendency of improvement of predicted pK_a values was demonstrated, which agrees with the findings of other authors using similar computational protocol. The influence of length of MD simulation and correlations to the initial structure were discussed.

Main references

This thesis is based on the following papers, which will be referred to in the text by their Roman numerals:

I Spitzner, N., Löhr, F., Pfeiffer, S., Koumanov, A., Karshikoff, A., Rüterjans, H. (2001). Ionisation properties of titratable groups in ribonuclease T_1. I. pK_a values in the native state determined by two-dimensional heteronuclear NMR spectroscopy. *Eur. Biophys. J.* **30**(3), 186-197

II Koumanov, A., Spitzner, N., Rüterjans, H. , Karshikoff, A. (2001). Ionisation properties of titratable groups in ribonuclease T_1. II. Electrostatic analysis. Eur. Biophys. J. **30**(3), 198-206

III Koumanov, A., Rüterjans, H. , Karshikoff, A. (2002). Continuum electrostatic analysis of irregular ionization and proton allocation in proteins. *Proteins* **46**(1), 85-96

IV Koumanov, A., Karshikoff, A., Friis, E. P., Borchert, T. V. (2001). Conformational averaging in p*K* calculations: improvement and limitations in prediction of ionization properties of proteins. *J. Phys. Chem. B* **105**(38), 9339-9344

V Zachariae, U., Koumanov, A., Engelhardt, H., Karshikoff, A. (2002). Electrostatic properties of the anion selective porin Omp32 from *Delftia acidovorans* and of the arginine cluster of bacterial porins. *Protein Sci.* **11**(6), 1309-1319

VI Koumanov, A., Benach, J., Atrian, S., Gonzàlez-Duarte, R., Karshikoff, A., Ladenstein R. (2003). The catalytic mechanism of Drosophila alcohol dehydrogenase: Evidence for a proton relay modulated by the coupled ionization of the active site Lysine/Tyrosine pair and a NAD^+ ribose OH switch. *Proteins* **51**(2), 289-298

VII Koumanov, A., Zachariae, U., Engelhardt, H., Karshikoff, A. (2003). Improved 3D continuum calculations of ion flux through membrane channels. *(Manuscript)*

Abbreviations

3D	three dimensional
BD	Brownian dynamics
DADH	*Drosophila* alcohol dehydrogenase
FD	finite difference
FEP	free energy perturbation
GB	generalized Born
LPBE	linearised Poisson-Boltzmann equation
MC	Monte Carlo
MD	molecular dynamics
MM	molecular mechanics
NMR	nuclear magnetic resonance
NP	Nernst-Planck
PBE	Poisson-Boltzmann equation
PNP	Poisson-Nernst-Planck
QM	quantum mechanics
RNase	ribonuclease
SOR	successive over-relaxation

1 Introduction

Almost all proteins contain sites, such as some of the amino acid side chains and/or prosthetic groups, which are prone to release or bind protons or electrons. Such sites may alter their charge upon changes in their environment (solution pH or redox potential, ligand binding, etc.) and thus are usually referred to as ionisable sites. Vital biological processes as enzymatic catalysis, respiration and photosynthesis, are crucially dependent on the ability of particular protein groups to participate in acid-base and redox reactions. Those reactions are to a large extent regulated by factors, which are predominantly electrostatic in nature. In general, inter- and intra-molecular electrostatic interactions are among the key factors determining the functional properties and structural stability of proteins[1-8] and perhaps the most fundamental part of these interactions is directly related to the charges of titratable and redox groups. A strong evidence for this is, for instance, the pH-dependence of a number of physiologically important phenomena, such as: enzyme activity[9], protein substrate/inhibitor interactions[10], conductivity and selectivity of protein channels[11,12] and many others. Apparently, any thorough analysis of those phenomena requires a detail characterisation of ionisation properties of individual titratable and redox groups.

Despite of the availability of powerful experimental methods (NMR, mutation experiments and others) for analysis of electrostatic interactions and ionisation behaviour of biomolecules, the interpretation of the experimental observations often might be ambiguous. It is also well known that proton/electron binding affinities of titratable/redox groups in proteins may significantly differ from affinities of those groups (or similar to them model compounds) being free in solution. Therefore, an accurate theoretical prediction of the ionisation equilibrium constants or their equivalents, the pK_a and E^o values, of individual ionisable sites is of grate importance for the analysis and understanding of a variety of protein properties.

Good agreement between calculated and measured pK_a values has been reported in number of papers[13-18,I,II]. Computations of ionisation energies in proteins have shed light on essential details in: enzymatic mechanisms[19-25,VI] (for review see also[26,27]), redox reactions[28-30], proton transfer across membranes[29,31-38], ligand binding[39-43], electric field in porin channels[44,V], protein stability[45-50] and others. Nevertheless, theoretical approaches for investigation of protein ionisation and electrostatics are still under intensive development and are constantly improving their predictive power[51-58,III,IV].

This work focuses manly on application of classical continuum electrostatics in prediction of protonation/deprotonation equilibria in proteins with known structure. Other approaches and techniques are also reviewed. The last section and paper VII are devoted to application of continuum model to simulations of non-equilibrium, steady-state ion flux through membrane pores.

2 Basic definitions and concepts

Protonation/deprotonation equilibrium of a compound, which can bind or release only a single proton (one titratable site), in a dilute aqueous solution is fully characterised by the quantity $\mathrm{p}K_\mathrm{a} = -\,log\,K_\mathrm{a}$, where $K_\mathrm{a} = \frac{[\mathrm{A}^-][\mathrm{H}^+]}{[\mathrm{AH}]}$ is the equilibrium constant of the reaction $\mathrm{AH} \xleftrightarrow{K_\mathrm{a}} \mathrm{A}^- + \mathrm{H}^+$. Taking into account that the pH of the solution is defined as negative decimal logarithm of the concentration of hydrogen ions i.e. pH = – *log* [H^+], one obtains the Henderson-Hasselbalch equation:

$$\mathrm{pH} = \mathrm{p}K_\mathrm{a} + log\frac{[\mathrm{A}^-]}{[\mathrm{AH}]} = \mathrm{p}K_\mathrm{a} + log\frac{\theta}{1-\theta} \tag{1}$$

Here θ is the degree of deprotonation defined as $\theta = [\mathrm{A}^-] / ([\mathrm{A}^-] + [\mathrm{AH}])$ and can be expressed formally from Eq. 1 as:

$$\theta = \frac{10^{(\mathrm{pH}-\mathrm{p}K)}}{1+10^{(\mathrm{pH}-\mathrm{p}K)}} \tag{2}$$

Here and further "pK_a" is denoted as "pK" for sake of simplicity. For dilute solutions of a compound with single titratable site it is reasonable to accept pK as a constant i.e. its value depends only on the specificity of the compound but not on the external conditions such as pH, temperature and etc (as it will be discussed further, this assumption is not valid for titratable sites in protein).

In terms of statistical thermodynamics Eq. 2 can be written as:

$$\theta = \frac{exp(-\Delta\mathrm{G}/\mathrm{RT})}{1+exp(-\Delta\mathrm{G}/\mathrm{RT})} \tag{3}$$

where R and T are the gas constant and the temperature, respectively. Here, θ has the meaning of probability the compound to be deprotonated and the denominator represents the partition function of two-state system. The free energy change of the solute-solvent system upon deprotonation of the solute at given conditions (pH, temperature, etc.) is $\Delta\mathrm{G} = \Delta\mathrm{G}^\mathrm{o} - \mathrm{RT}ln10\,\mathrm{pH}$, where $\Delta\mathrm{G}^\mathrm{o} = \mathrm{RT}ln10\,\mathrm{p}K$ is the standard deprotonation free energy of the corresponding compound (the solute).

A redox reaction $A_{ox} + e^- \xleftrightarrow{K_{red}} A^-_{red}$ can be considered analogously to protonation/deprotonation equilibrium and the relation between the redox potential, E = -RT/F $ln[e^-]$, of the solution and the standard redox potential, E^o = -RT/F lnK_{red}, of the redox couple A_{ox}/A^-_{red} is given by:

$$E = E^\circ + \frac{RT}{F} ln \frac{[A_{ox}]}{[A^-_{red}]} = E^\circ + \frac{RT}{F} ln \frac{\langle A_{ox} \rangle}{1 - \langle A_{ox} \rangle} \quad . \tag{4}$$

Here F is the Faraday constant and the probability for the redox site to be oxidized, $\langle A_{ox} \rangle$, is equal to the right-hand side of Eq. 3, in which $\Delta G = -F(E - E^o)$ is the free energy for oxidation the site.

Ionisation equilibrium constants of model compounds (with single ionisable site), which resemble ionisable sites in proteins, usually can be experimentally measured or, in some special cases, obtained by quantum chemical calculations (Table 1). As seen from Table 1, there is no absolute agreement among the scientific society even on the seemingly simple issue of what pK (model compound) is appropriate to represent certain protein titratable group if it is considered free in solution. In proteins, the situation becomes even more complicated due to the fact that ionisation equilibria of individual sites are influenced by number of factors, the most important of which are listed in Table 2. In all theoretical studies, protein ionisation is described via evaluation of statistical averages such as θ and $\langle A_{ox} \rangle$. The statistical consideration allows equilibrium characteristics of the system to be obtained on the basis of free energy calculations and hence to analyse in details the role of different factors.

The similarity in the theoretical description of redox and acid-base reactions allows reduction/oxidation and protonation/deprotonation equilibria to be addressed by single generalized treatment[30,51]. The present work emphasise on the phenomenon relevant for majority of the proteins, namely protonation/deprotonation equilibrium of titratable sites. Inclusion of redox processes can be easily achieved by a trivial extension of the described below formalism.

Table 1. pK and standard reduction potential, E^o(at pH 7 and 25ºC), values of model compounds representing most common ionisable sites in proteins.

Titratable group	pK_{mod}
H_3O^+	-1.7 [a]
C-terminus	3.8 [c], 3.6 [d], 3.63 [e]
Asp	4.0 [c, d]
Glu	4.4 [c], 4.5 [d]
His	6.3 [c], 6.4 [d]
His (tautomers)	6.6 [h], 7.0 [i]
N-terminus	7.5 [c], 8.0 [d], 8.2 [j]
Cys	9.5 [c], 9.0 [d], 8.3 [k]
Tyr	9.6 [c], 10.0 [d]
Lys	10.4 [c, d]
His (2^{-nd} deprot.)	10.8 [d], 14.0 [l]
Arg	12.0 [c, d]
H_2O	15.7 [a]
Ser, Thr	16.0 [m]

Redox couple	E^o (mV)
$O_2/O_2\cdot^-$	-330 [b]
NAD^+/NADH	-320 [f]
Heme (His,His)	-220 [g]
Heme (Met,His)	-70 [g]
Cu^{2+}/Cu^+	153 [a]
Fe^{3+}/Fe^{2+}	770 [a]
$Tyr\cdot^+$/Tyr	930 [b]
OH/OH^-	2020 [a]

a) from Ref.[59]; b) from Ref.[60]; c) from Ref.[61]; d) from Ref.[5]; e) pK_a of N-acetyl glycine[59]; f) from Ref.[62]; g) from Ref.[63]; h) deprotonation of $N_{\delta 1}$ ($N_{\varepsilon 2}$ - methylated)[64]; i) deprotonation of $N_{\varepsilon 2}$ ($N_{\delta 1}$ - methylated)[64]; j) pK of glycine amide, http://www.Science.smith.edu/Biochem/Biochem_353/Common_Buffers.htm; k) from Ref.[20,65]; l) from Ref.[66]; m) see Table III in Ref.[67]

Table 2. Major factors determining the ionisation equilibria of individual groups in proteins.

	Factor	Comments
1	Charge-charge interactions between ionisable sites	These interactions are pH dependent and determine the co-operative character of the ionisation equilibria in proteins.
2	Electrostatic interaction of the titratable groups with the permanent protein charges	These are the interactions of a given ionisable group with charges not associated with ionisable sites (called also fixed or background charges; e.g. backbone dipoles).
3	Desolvation energy	This is the energy needed to transfer an ionisable group (model compound) from solvent to its location in the protein if the first two factors are neglected. It is primarily due to the difference in the response of solvent and protein environment to the electrostatic field created by the ionisable group. This energy is always positive (unfavourable) and is often called "desolvation penalty".
4	Conformational flexibility	At certain conditions proteins may exist in more then one conformation simultaneously. Also, upon ionisation of a given titratable group conformational changes may occur, which changes the factors 1 - 3.

3 Methods for calculation of ionisation free energies

The influence of protein environment on ionisation properties of individual groups is predominantly electrostatic in nature (see factors 1-3 in Table 2). That is why, the prediction of those properties is a fundamental objective and in the same time an ultimate test for the theories of electrostatic interactions in biomolecules.

3.1 Overview of the earliest approaches

First theoretical models relating electrostatics and pH titration of proteins appeared long before any knowledge about three-dimensional (3D) protein structure to be available. Linderstrøm-Lang[68] treated protein molecule as a sphere with a low permittivity, surrounded by a high permittivity continuum. Charges were considered to be uniformly spread over the surface of the sphere and p*K* shifts were assumed proportional to the average charge. This model is *a priori* incapable to distinguish titrations of individual residues and was used for simulating only the overall protein titration. Later on, the theory of Kirkwood and Tanford[69,70] became one of the fundamental tools for computation of protein electrostatic properties and had been actively used for a long period. This approach also employs spherical approximation of the protein molecule but in contrast to the Linderstrøm-Lang's model, charges of the titratable groups are represented as point charges on the surface of the sphere. This model allows analytical solution the linearised Poisson-Boltzmann equation (LPBE) and hence analytical expression for the charge-charge interaction energies, which are given in series of papers[69-72]. This theory had limited application for analysis of electrostatic interactions in proteins, however, it was successfully applied for the prediction of p*K* values of amino acids. It has two major weaknesses. The first one is that it does not account for the effect of desolvation of the titratable groups in the protein interior, the second one is that the assumed analytical (spherical) shape of the protein medium is a too crude approximation for the majority of the proteins. In order to account for the fact that different titratable groups are differently buried in the protein interior Gurd and co-workers introduced a correction proportional to the solvent accessibility of each group[73,74].

3.2 Micro- and macroscopic models

The progress in experimental techniques led, on one hand, to increase in the quantity and quality of the available structural data and, on the other hand, to ability of measuring p*K* values of individual titratable groups[75-87], which together with the advances in computer technology stimulated the development of various theoretical and computational methods in the field of protein ionisation and electrostatics. These

theoretical approaches differ from each other mainly in the representation of the investigated protein-solvent system and hence in the manner of calculating the ionisation free energies. Although there is no strict classification, usually methods employing macroscopic quantities (e.g. dielectric constant, ionic strength, etc.) for characterization of certain system properties are referred to as macroscopic or continuum models, while approaches entirely based on explicit representation of atomic details (e.g. electron densities, polarisabilities, etc.) are called microscopic. The continuum models benefit from their computational and conceptional easiness. In these models, the protein molecule is treated as dielectric material, the permittivity of which is, in fact, not well defined and often serves as a justifiable parameter. The microscopic approaches avoid this obstacle and provide more rigorous and detail description of the considered system, however, at much higher computational cost and by employing a large set of fitted and semi-empirical parameters. Different approaches are thoroughly described, analysed and compared in a number of reviews[88-93]. The choice of a method depends on the concrete task that has to be solved (i.e. on the concrete questions addressed to the examined protein) and in some cases it is difficult to prefer a particular approach among the broad spectrum of those currently available. Recently, there is an apparent trend of combining macroscopic and microscopic methods, so that parts of the molecule, which are of special interest, are treated microscopically, while their surrounding is represented by an appropriate continuum model.

The earliest models for description of protein ionisation and electrostatics are, in fact, continuum ones. Nowadays approaches based on continuum models will be discussed in more details in the next three parts of this work and the remaining of this part is a brief survey of microscopic methods for calculation of ionisation free energies.

3.3 Brief survey of microscopic approaches

As far as, the ionisation of a protein site involves either formation/fission of a covalent bond (in the case of protonation/deprotonation reaction) or radical changes in the structure of the electron shells (in redox processes), the most rigorous characterization of an ionisation event is provided by quantum mechanics (QM). Simulations of pH titration, in which energy calculations are taking explicitly into account changes in covalent bonding, are usually called absolute pK calculations. In principle, the solution of Schrödinger's equation completely specifies the physical-chemical properties of the studied system. However, *ab initio* calculations with high electron correlation, due to their extreme computational cost, are practically limited to systems containing up to 100-150 atoms, which makes the direct application of these methods to protein-solvent systems unfeasible. Results of QM calculations on small compounds are often used to elaborate parameters (atomic partial charges, torsion potentials, etc.), which are than employed in force fields for molecular modelling. Quantum-chemical calculations can also provide ionisation free energies (usually proton/electron binding affinities in vacuum) of model compounds, which resemble certain type of protein ionisable sites. Such computations are especially helpful if the

corresponding energies are experimentally unattainable. The necessity of extending QM methods over larger systems led to the development of various less rigorous approaches like semiempirical QM and hybrids of QM and classical models. Noodleman, Bashford and co-workers[25,67,94-97] elaborated an approach, in which density functional theory (DFT) is applied to a cluster of protein groups, whose environment is treated by continuum electrostatics. Recently, the mixed quantum mechanical/molecular mechanical (QM/MM) methodology became very popular. These methods retain the advantage of treating quantum-chemically only a small core of the system, while the remainder is modelled at the molecular mechanics level. The computational effort is comparable with that for a small model quantum-chemical computation, but considers the environment in a reasonably accurate fashion. QM methods are capable to evaluate directly the energetics of different pathways in enzymatic reactions and therefore p*K* values obtained in these simulations are seldom in the focus of the study[25,98-100]. Surveys on quantum-chemical approaches applied to biomolecules can be found in number of papers[88,92,101,102].

Approaches, merging statistical mechanical concepts with molecular dynamics (MD) or Monte Carlo (MC) dynamics in order to yield the free energy difference between two equilibrium states of a given system, are referred to as free energy perturbation (FEP) calculations. The FEP method is rather general and thus it is well developed and documented in details[103-107]. The thermodynamic basis of this method employs partition function formalism to relate the free energy difference $\Delta G = G_1 - G_0$ between two equilibrium states (designated by "0" and "1") to an ensemble average involving the difference between Halmiltonians of those states $\Delta H = H_1 - H_0$, namely $\Delta G = - RT\, ln\langle exp(-\Delta H / RT)\rangle_0$. Here the Boltzmann averaging is over the ensemble corresponding to state "0". Usually, it is assumed that the kinetic parts of H_1 and H_0 are equal and the Halmiltonian is replaced by an empirical potential energy function defined by the atomic coordinates and the molecular force field parameters used for the ensemble generation. Despite of the exactness of the above formula, practically it cannot be used straightforward because a reliable estimate of ΔG requires the two states to share large parts of their phase spaces, which is seldom the case. Therefore, the transition "0"→"1" is decomposed into a number of subsequent transitions "λ_i"→" λ_{i+1}", for which the corresponding $\Delta G_{\lambda i}$ can be accurately calculated from the above equation and, finally, ΔG is obtained as the sum of $\Delta G_{\lambda i}$. This procedure involves simulations of some intermediate states "λ" of the system. These states are treated as equilibrium ones although they are usually unrealistic i.e. the system is "alchemically" driven from state "0" to state "1". The FEP method also allows different contributions in ΔG (e.g. contributions arising from electrostatic, van der Waals, etc. interactions) to be evaluated by component analysis[108]. Normally, FEP calculations use an all-atom representation of the protein-solvent system, although in some cases the solvent can treated as a continuum. These calculations face the typical for MD and MC simulations problem of completeness of the collected ensembles with respect to the evaluation of the needed statistical averages. In addition, the cooperativeness of the protein ionisation makes FEP techniques computationally unattractive for complete characterisation of the ionisation properties of proteins an their application in the field is practically limited to

computations of solvation energies, intrinsic p*K*s[109] and p*K* values small clusters of titratable sites[110,111]. The FEP approach are used with a great success for investigation of energetics of binding between complex molecules (supposing single binding site per molecule) such as proteins, DNAs and drugs[105,112-114] as well as to study effects of point mutations[115,116].

An alternative microscopic model was proposed by Warshel and co-workers[4,117,118]. In this approach the protein molecule is considered on an atomic level by taking explicitly into account all atomic partial charges and polarisabilities and representing the surrounding solvent molecules as a lattice of Langevin dipoles. In this way, the usage of dielectric constants is avoided, however a screening function for the electrostatic field in the solvent part of the system is introduced along with the Langevin dipoles. pK values are calculated by reversible adiabatic charging of the groups of interest, which can be considered as a FEP technique.

4 Macroscopic treatment of protein electrostatics

In macroscopic approaches, both the protein molecule and the surrounding solvent are considered as continuum media characterised by macroscopic quantities, such as dielectric constant and ionic strength. The continuum dielectric models require relatively small number of parameters whit clear physical meaning to describe the protein-solvent system and to perform calculations. This computational and conceptional simplicity made the macroscopic models perhaps the most frequently used approaches for investigating the electrostatics in biomolecules.

The availability of high-resolution 3D protein structures and the modern computational facilities allow nowadays continuum models to employ a lot of atomic details in the representation of the protein molecule. Coordinates and van der Waals radii of protein atoms together with solvent probe radius are model parameters, which are taken from other experimental or theoretical studies. These parameters uniquely define the contact and the solvent accessible surfaces, either of which can serve as a dielectric boundary separating the protein moiety from the solvent. Other dielectric boundaries can be easily introduced if needed, for instance to represent a membrane or to treat the protein as a body with inhomogeneous permittivity. If solvent salt concentration is of interest, the free ions in the solvent are represented as a continuum matter with certain space distribution. The space accessible for solvent ions is delineated analogously to the solvent accessibility. The protein charge distribution is another essential parameter. Charge positions are determined by the protein structure. The charge values of the permanent charges, such as the partial charges determining the peptide bond dipoles, are available from other studies. The charge values of the titratable groups depend on their protonation/deprotonation equilibria in the concrete protein and on the solution characteristics, such as pH, ionic strength, temperature etc.

The studies presented in this dissertation are based on finite-difference solution of either the linearised Poisson-Boltzmann equation (papers II-VI) or the Poisson-Nernst-Planck equations (paper VII) and in both cases, focusing procedure was applied for refinement of the final results. Therefore, this methodology for computing the electrostatic potential in solvated biomolecules is given in detail, while other approaches are briefly commented.

4.1 Classical continuum model

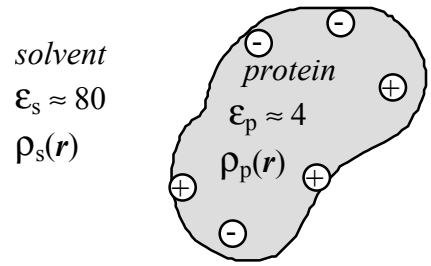

Figure 1. Continuum dielectric model for calculation of electrostatic interactions in proteins.

This model treats the protein as a rigid body with low dielectric constant ($\varepsilon_p = 2$ to 20, see below) and fixed charge distribution, $\rho_p(\boldsymbol{r})$, which is immersed in a high dielectric medium ($\varepsilon_s \approx 80$, assuming aqueous solution) with a distribution of free ions, $\rho_s(\boldsymbol{r})$, (Fig. 1). In such system, the electrostatic potential, $\varphi(\boldsymbol{r})$, at point $\boldsymbol{r}$ in the space can be obtained as a solution of the Poisson equation:

$$\nabla \cdot (\varepsilon(\boldsymbol{r})\nabla\varphi(\boldsymbol{r})) = -4\pi(\rho_p(\boldsymbol{r}) + \rho_s(\boldsymbol{r})). \tag{5}$$

The solvent ions, being mobile, have no fixed locations. Thus, their distribution is not preliminary defined and hence Eq. 5 cannot be directly solved. It is reasonable to assume that the mobile ions with charges $Z_i e$, where Z_i is the valence of the ***i***-th ion species and e is the value of the elementary charge, are influenced by the electrostatic field and in equilibrium obey the Boltzmann distribution, i.e.:

$$\rho_s(\boldsymbol{r}) = \sum_i C_i Z_i e \ exp\left(\frac{-Z_i e \ \varphi(\boldsymbol{r})}{kT}\right). \tag{6}$$

The summation is over all types of ions, C_i is the bulk concentration of ions of type *i* and k is the Boltzmann constant. A substitution of $\rho_s(\boldsymbol{r})$ from Eq. 6 in Eq. 5 leads to the Poisson-Boltzmann equation (PBE):

$$\nabla \cdot (\varepsilon(\boldsymbol{r})\nabla\varphi(\boldsymbol{r})) = -4\pi\rho_p(\boldsymbol{r}) + 4\pi\sum_i C_i Z_i e \ exp\left(\frac{-Z_i e \ \varphi(\boldsymbol{r})}{kT}\right). \tag{7}$$

The PBE (abbreviated often as NPBE) is a non-linear partial differential equation for the electrostatic potential. The non-linearity leads, first, to substantial difficulties in the solution of Eq. 7, and second, the resulting potential is a non-additive quantity, which appears to be a serious obstacle in many practical applications. The limited application[119-123] of the PBE is also due to the fact that, for the majority of proteins immersed in solvent with physiological or lower concentration of monovalent ions, the condition $Z_i e\varphi(\boldsymbol{r})/RT < 1$ is fulfilled in the ion accessible regions. This condition suggests that the exponent in Eq. 6 can be approximated by the linear part of its Maclaurin expansion. Further, taking into account the electroneutrality of the solution, i.e. $\Sigma C_i Z_i = 0$, Eq. 6 can be written as:

$$\rho_s(\boldsymbol{r}) = -\frac{e^2\varphi(\boldsymbol{r})}{kT}\sum_i C_i Z_i^2 . \tag{8}$$

This linearisation was first applied by Debye and Hückel, in 1923, in the derivation of their theory, which was initially implemented for calculation of the electrostatic free energy of small spherical ions in ionic solutions. An additional characteristic of the solution was introduced, namely, the Deby screening parameter, κ, which is defined as:

$$\kappa^2 = \frac{8\pi e^2 I}{\varepsilon_s kT} \tag{9}$$

where $I = 1/2 \sum C_i Z_i^2$ is the ionic strength of the bulk. Normally, ion concentrations, C_i, are given in molarities and k is expressed in $Å^{-1}$, in which case the right-hand side of Eq. 9 should be multiplied by a factor of $N_A \cdot 10^{-3}$ resulting from units' conversion.

After substitution of Eq. 8 and Eq. 9 in Eq. 5 one obtains the linearised Poisson-Boltzmann equation (LPBE):

$$\nabla \cdot (\varepsilon(\boldsymbol{r}) \nabla \varphi(\boldsymbol{r})) - \kappa^2 \varepsilon_s \varphi(\boldsymbol{r}) + 4\pi \rho_p(\boldsymbol{r}) = 0 . \tag{10}$$

The second term in the above equation turns to zero if point ***r*** is in an ion inaccessible region, where the ionic strength and hence κ and $\rho_s(\boldsymbol{r})$ are equal to zero. A detail derivation of Eq. 10 is given by Tanford[124]. It was shown that the linearised and the non-linear Poisson-Boltzmann equations give practically equal results[125] even for moderate, say physiological, ionic strength. Besides that it is computationally easier to solve Eq. 10 instead of Eq. 7, the electrostatic potential calculated by LPBE is additive i.e. the potential of a system of charges equals the sum of potentials created by each single charge of the system. As a consequence, different electrostatic factors, contributing to the ionisation equilibrium of individual groups (factors 1-3 in Table 1), can be considered additive without any additional postulations. This additivity facilitates both building of computational algorithms and the theoretical analysis.

It should be noted, that PBE and LPBE (Eqs. 7,10) have a unique solution in arbitrary volume if $\varepsilon(\boldsymbol{r})$, $\rho_p(\boldsymbol{r})$ and $\rho_s(\boldsymbol{r})$ (alternatively κ) are known for the entire volume and $\varphi(\boldsymbol{r})$ are given at boundary surface. The boundary conditions in most of the theoretical studies are set to $\varphi(\infty)=0$.

4.1.1 Finite-difference solution of linearised Poisson-Boltzmann equation

An analytical solution of the LPBE (Eq. 10) is possible only if the protein is approximated by a body with simple regular shape e.g. sphere[69]. In order to account the actual shape of the protein, as defined by its 3D structure, Eq. 10 has to be solved numerically by algorithm allowing arbitrary profile of the dielectric boundaries. The most popular and widely used routine is the finite-difference (FD) method, first proposed for proteins by Watson and Warwicker[126]. The FD solution of the LPBE (as well as of the PBE) was studied systematically by Honig and co-workers, who further developed and refined this method regarding its applications to biomolecules[119,127-130] (for review[131]). The FD technique for solving the LPBE and its application to p*K* calculations is excellently summarised by Beroza and Fredkin[132].

In the FD approach, the protein and part of the surrounding solvent are placed in a box with a 3D grid forming a cubic lattice as illustrated in Fig. 2. Quantities, such as permittivity, charge, and Debye parameter, characterising the protein and the solvent are attributed to the lattice nodes (grid points). This process of assignment is often called "mapping" of the system on the grid. The continuum task is transformed to a discrete one by replacing the system of interest by the set of grid points. The problem of finding $\varphi(\boldsymbol{r})$ is reformulated as a problem of calculating the electrostatic potential, φ_i, at each grid point.

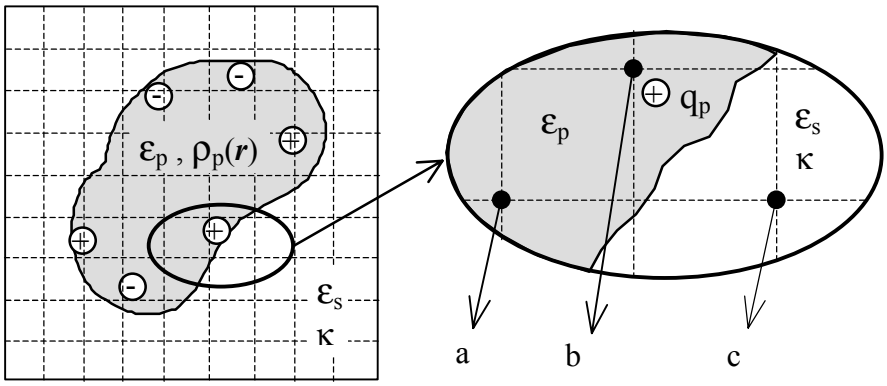

Figure 2. Lattice representation of the protein-solvent system (illustrated in 2D for simplicity). Permittivity, ε_i, charge, q_i, and Debye parameter value, κ_i, are assigned to each lattice nod ***i***. ***Nodes a***: $\varepsilon_a=\varepsilon_p$, $q_a=0$, $\kappa_a=0$. ***Nodes b***: $\varepsilon_b=\varepsilon_p$, $q_b=q_p$, $\kappa_b=0$. ***Nodes c***: $\varepsilon_c=\varepsilon_s$, $q_c=0$, $\kappa_c=\kappa$, $\kappa\neq 0$ at non-zero ionic strength.

The assignment of a dielectric constant and a value of the Debye parameter requires only a simple geometry check-up. However, the mapping of the protein charge distribution, $\rho_p(\boldsymbol{r})$, is not that straightforward as $\rho_p(\boldsymbol{r})$ is usually represented by set of point charges located at atom centres and it is very unlikely that any of these point charges would coincide with a grid point. There are different methods for assignment of charge values to the grid points. A simple and effective one is to distribute the charge over the 8 closest grid points by interpolation[127].

An algebraic expression for the electrostatic potential at arbitrary internal grid point (i.e. any grid point not lying on the walls of the computational box) can be derived by integrating Eq. 10 in a cubic volume, V, associated with the considered grid point (Fig. 3):

$$\int_V \nabla\cdot(\varepsilon(\boldsymbol{r})\nabla\varphi(\boldsymbol{r}))\mathrm{d}v - \int_V \kappa^2\varepsilon_s\varphi(\boldsymbol{r})\mathrm{d}v + 4\pi\int_V \rho_p(\boldsymbol{r})\mathrm{d}v = 0 \qquad (11)$$

The first integral in Eq. 11 is solved by using Gauss' theorem and the integration over the surface area, S, of the cube is than expressed in FD terms:

$$\int_V \nabla\cdot(\varepsilon(\boldsymbol{r})\nabla\varphi(\boldsymbol{r}))\mathrm{d}v = \oint_S \varepsilon(\boldsymbol{r})\nabla\varphi(\boldsymbol{r}).\mathrm{d}\boldsymbol{s} = h\sum\nolimits_j \varepsilon_{j0}(\varphi_j-\varphi_0).$$

Here and in the following derivation, "0" and "j" indicate the grid point where the integration cube is centred and the neighbouring points, respectively (Fig. 3). The

summation is over the six nearest neighbours. ε_{j0} is the dielectric constant at the cube wall located between points *0* and *j*. For practical reasons, ε_{j0} is usually calculated as an average of ε_j and ε_0. The parameter *h* is the lattice spacing, i.e. the distance (in Å) between neighbouring grid points.

The second integral in Eq. 11 givs:

$$\int_V \varepsilon_s \kappa^2 \varphi(\boldsymbol{r}) dv = \varepsilon_s \kappa_0^2 \varphi_0 h^3 = \begin{cases} 0 & \text{if point "0" is ion inaccessible} \\ \varepsilon_s \kappa^2 \varphi_0 h^3 & \text{if point "0" is ion inaccessible} \end{cases}$$

The third integral in Eq. 11 differs from zero only for the points where a charge value is assigned:

$$4\pi \int_V \rho(\boldsymbol{r}) dv = 4\pi q_0 \ .$$

Finally, Eq. 11 can be written in FD terms as:

$$h \sum_j \varepsilon_{j0} (\varphi_j - \varphi_0) - \varphi_0 \kappa_0^2 h^3 \varepsilon_s + 4\pi q_0 = 0 ,$$

which gives:

$$\varphi_0 = \frac{\sum_j \varepsilon_{j0} \varphi_j + 4\pi q_0 / h}{\sum_j \varepsilon_{j0} + \kappa_0^2 h^2 \varepsilon_s} \qquad (12)$$

This equation relates the electrostatic potential at certain lattice point with the potentials at the neighbouring points and holds for all grid points, which are internal for the computational box. The grid points, located at the walls of the box do not have neighbours in some directions, so that the electrostatic potential at these points has a preliminary defined value (boundary conditions). In this way, the FD technique converts the LPBE into a system of linear equations, which similarly to the LPBE has a unique solution at given boundary conditions.

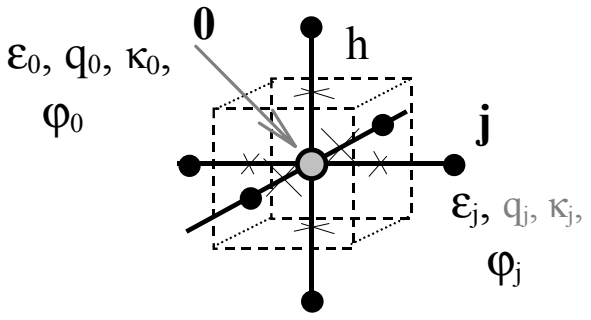

Figure 3. An internal grid point *0* (grey) and its six nearest neighbour grid points *j* (black) are given together with the parameters assigned to them. The cube (dashed lines), in which LPBE is integrated, is centred at point *0* and has edge lengths equal to the lattice spacing *h*.

For cubic computational box (see Fig. 2) with L+2 lattice nods per edge, the number of internal grid points is $N=L^3$, which is the number of equations in the linear system produced by FD. There are two main reasons for aiming at as large L as possible (limitations depend on the computational facilities). First, at $h \to 0$ the FD solution tends to the exact solution of LPBE. Second, there is some unavoidable inexactness in setting the boundary potential values. Thus, in order to reduce the influence of the boundary

conditions on the calculated potential in the region of interest (the protein), the walls of the computational box should be sufficiently distant from the protein, i.e. sufficiently large computational box is required. At present calculations are performed with 100-200 nods per edge, which results in huge systems of 10^6-10^7 equations.

4.1.1.1 successive over-relaxation

The successive over-relaxation (SOR) is a method for iterative solution of non-singular linear systems, which can be written in matrix form as:

$$\Phi = T\Phi + Q \,, \tag{13}$$

where $\Phi \in \Re^N$ is the vector that has to be found, while the vector $Q \in \Re^N$ and the matrix $T \in \Re^N$ are known. The SOR method allows the solution of Eq. 13, Φ, to be approximated with any desirable accuracy. For big systems, highly accurate SOR approximations are achieved much faster compare to methods providing the exact solution. Broad and well-ordered summary of the SOR theory is given by Hadjidimos[133]. Nicholls and Honig[129] wrote a very detail and comprehensive (despite of the typing mistakes) description of the application of this theory for FD solution of LPBE and PBE.

Eq. 13 yields the following iterative scheme:

$$\Phi^{(m+1)} = T\Phi^{(m)} + Q \,, \tag{14}$$

which converges if and only if $\lambda_T < 1$, where λ_T is the spectral radius of the matrix T (the largest absolute value of the eigenvalues of T). The spectral radius also determines the convergency rate. Using the formalism of eigenvectors and eigenvalues it can be shown that: (***i***) The absolute difference, $\| \Phi - \Phi^{(n)} \|$, between the exact solution, Φ, and the result of ***n***-th iteration step, $\Phi^{(n)}$, is approximately proportional to the ***n***-th power of λ_T , i.e. $\| \Phi - \Phi^{(n)} \| \propto \lambda^n_T$. This relation provides a useful estimate for the number of iterations needed to meet certain convergency criterion. (***ii***) The contribution of the initially chosen values, $\Phi^{(0)}$, in $\Phi^{(n)}$ is at most proportional to λ^n_T. It follows ($\lambda_T < 1$), that, with increasing the number of iterations, the iterative solution exponentially approaches the exact one and the convergency is faster for smaller λ_T. Apparently, the initial approximation is not essential (normally $\Phi^{(0)}=0$), although the number of iterations needed to meet a given convergency criterion, in general depends on the initial choice.

The matrix T and the vector Q can be constructed from Eq. 12 as $\{T\}_{ij} = (\Sigma_j \varepsilon_{ji}) / (\Sigma_j \varepsilon_{ji} + \kappa_j^2 h^2 \varepsilon_s)$,$i \neq j$, $\{T\}_{ii} = 0$ and $\{Q\}_i = f(q_i, \varphi^{boundary})$. The vector Φ is simply $\{\Phi\}_i = \varphi_i$. The boundary potential values, $\varphi^{boundary}$, are accounted in Q, because, being set preliminary, they play in Eq. 12 role similar to that of the known charge distribution represented by q_0. I was shown that the constructed in this way matrix T satisfies the condition $\lambda_T < 1$. Appling Eq. 14 with the above construct reproduces the Jacobi relaxation algorithm for Φ. Grid points can be enumerated so that "even" points have only "odd" points for neighbours and vice versa. This points can be ordered in Φ in such way that the first half of the vector to be occupied by even and the second half by odd points. Then, as seen from Eq. 12, the correspondingly ordered matrix, T_{ord}, will

have the structure $T_{ord} = \begin{pmatrix} 0 & T_1 \\ T_2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & T_1 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & 0 \\ T_2 & 0 \end{pmatrix}$, where the submatrices T_1 and T_2 update even entries with odd ones and odd with even, respectively. After updating the values for even points, these updates are used for calculating the new values for odd points and so on. One iterative step of this scheme (Gauss-Seidel) is, in fact, equivalent to two Jacobian steps. Formally, this can be reproduced by Eq. 14 if T and Q are substituted by $T'=T^2_{ord}$ and $Q'=(T_{ord} + E)Q_{ord}$, respectively; E is the identity matrix. Gauss-Seidel iterations are twice faster than the Jacobi procedure, which is also seen from the relation between their spectral radii $\lambda_{T'} = \lambda^2_T$ (or $\lambda_{GS} = \lambda^2_J$).

The SOR method modifies the Gauss-Seidel relaxation by introducing an over-relaxation parameter $\omega \in (0,2)$, which enables further improvement of the convergency rate. The introduction of ω can be done in the following way: $\Phi = \omega\Phi + (1-\omega)\Phi$, then Φ in the first term of the right-hand side is expressed from Eq. (13), where T and Q are substituted by T' and Q', which yields $\Phi = \omega(T'\Phi + Q') + (1-\omega)\Phi$. This suggests the following iterative procedure:

$$\Phi^{(m+1)} = \omega(T'\Phi^{(m)} + Q') + (1-\omega)\Phi^{(m)} \qquad (15)$$

The above equation is the basis of the SOR method. Eq. (14) becomes Eq. (15) after substitution of T and Q with $T'' = (\omega T' + (1-\omega)E)$ and $Q'' = \omega Q'$. Apparently, if $\omega=1$ SOR becomes Gauss-Seidel. It is possible to find a value, ω_b, for the over-relaxation parameter providing best convergence i.e. smallest spectral radius of T'' ($\lambda_{T''(\omega b)} \leq \lambda_{T''(\omega)}$). This value is given by:

$$\omega_b = \frac{2}{1 + \sqrt{1 - \lambda_{T'}}}. \qquad (16)$$

The corresponding spectral radius, $\lambda_{T''(\omega b)}$, is in this case $\lambda_{T''(\omega b)} = \lambda_{T'} / [1 + (1-\lambda_{T'})^{1/2}]$, which is apparently smaller than $\lambda_{T'}$. With the increase of the value $\lambda_{T'}$, $\lambda_{T''(\omega b)}$ increases as well, however the advantage of SOR over Gauss-Seidel becomes greater. In FD calculations of LPBE performed with $L \approx 100$, SOR relaxes normally 50- to 100-fold faster than Gauss-Seidel. The spectral radius $\lambda_{T'}$, determining ω_b, can be estimated by performing several (usually just one) Gauss-Seidel iterations with Q' set to zero and $\Phi^{(0)}$ selected to be a close approximation of the eigenvector of T' with the highest eigenvalue (i.e. $\lambda_{T'}$)[129].

4.1.1.2 boundary conditions

The proper setting of the boundary conditions is important for the precision of the calculations as far as they act as a source of potential and in that sense they are equivalent to a certain charge distribution. Most often diluted solutions of protein are considered and in those cases it is reasonable to approximate the electrostatic potential on the faces of the box according to the Debye-Hückel theory: $\varphi_k = \sum_j \frac{q_j e^{-\kappa r_{kj}}}{\varepsilon_s r_{kj}}$, where *k* and *j* enumerate the boundary and the charged grid points, respectively. This approximation, however, holds for distances $\mathbf{r}_{kj}$ significantly larger than the dimensions

of the protein molecule. Moreover, the distortion of calculated potential, due to inexactness of boundary values, is larger for the points close to the walls of the computational box than for the central ones[128]. Thus, the volume, in which LPBE is solved by FD, should be much larger than that of the protein and centred at the region of interest. On the other hand, the grid spacing should be considerably smaller than the atomic radii to properly describe the charge distribution and the overall shape of the molecule. Therefore, in order to achieve a satisfactory accuracy, the calculations should be performed in a large volume with fine grid spacing, which may be computationally prohibitive even for small proteins.

The focusing technique[128] allows to combine distant boundaries with a fine grid in the region of interest. This method consists in solving the LPBE consecutively in several, inserted one into another cubes containing the investigated area (Fig. 4). First, potentials are calculated in a big initial box with coarse grid and boundary conditions set, for instance, according to the Debye-Hückel theory. Then, calculations are repeated in a smaller box with finer grid located inside the previous box and centred at the region of interest and the boundary potentials are taken from the calculations in the previous coarser lattice. Depending on the task the focusing procedure might be repeated several times.

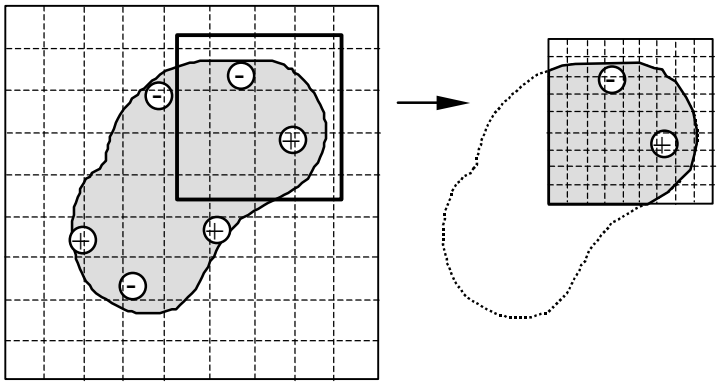

Figure 4. Focusing. The focused box (on right) is smaller than the preceding computational box (on left) and has a finer grid. Boundary values for the focused box are taken from the previous calculations in the larger box.

4.1.2 Other computational techniques

The electrostatic potential in the Poisson equation (Eq. 5) combines the potential created by the source charges with the reaction field potential, originating from the induced dielectric polarisation. A common technique to separate these potentials is to perform additional calculations with the same source charge in a homogeneous dielectric environment[134]. Alternatively, the electrostatic field in the protein can be considered as a field in an infinite homogeneous medium and the difference between the protein and solvent permittivities is accounted by explicit representation of the induced surface charge on the dielectric boundary. In this case, the total potential is a sum of Coulomb potentials (for the source charges) and an integral, over the boundary surface, of a Coulomb-like term (for the induced charge). This is, in fact, the theoretical ground of the boundary element method (boundary integral method), in which the integration is reduced to a summation by dividing the boundary surface in appropriate elements. Tomasi and Persico[88] reviewed the method in detail. The introduction of multiple

dielectric regions and salt effects create serious difficulties in the boundary element method, while in the FD solution of LPBE this is done without any special effort. Nevertheless, the boundary element method was successfully for electrostatic calculations in proteins[135-140].

The LPBE can be numerically solved by FD as well as by the finite element[141] method or by its more advanced variant the spectral element method[142,143]. In the finite element method, the computational space in finite domains (elements) and inside each element the potential is approximated by a polynomial expansions, which is used to discretize the differential equation.

In multi-grid techniques[120,144,145], more dense computational grid is used in some regions and more coarse in others, which is another implementation of the main idea behind the focusing procedure.

Instead of SOR, conjugate gradient methods[146] are employed in some studies.

The variety of numerical techniques for treating continuum electrostatics resulted in several widely used programs, such as DelPhi[147], UHBD[146], MEAD[148], ITPACT[149,150], Mainfold Code[151,152] and others. All calculations, in presented here studies[II-VII], were carried out with a software developed by Karshikoff and co-workers[44,153,154,III,VII].

4.1.3 Protein dielectric constant

The key parameter of the continuum dielectric model is the protein dielectric constant. While the dielectric constant of the bulk solvent can be measured, in the vicinity and inside the protein molecule its value can only be assumed. At the protein-solvent interface the dielectric constant may differ from that of the bulk. Lamm and Pack[155] have evaluated the dielectric constant in this region and have shown that it can be reduced to about 30. Theoretical studies[4] indicate that in presence of extremely strong electrostatic fields (e.g. around highly charged metal ions) the local permittivity of water may be even lower. The value of the protein dielectric constant, ε_p, is widely and controversially discussed[90,156,157]. The assignment of a macroscopic quantity (permittivity) to a microscopic object (single molecule or even parts of it) has been a subject of a principle criticism. In this regard it should be noted, that ε_p is a convenient parameter, which accounts for responses to an electric field that are not treated explicitly. Gilson and Honig[158] applied The Kirkwood-Fröhlich theory to a model of arbitrary protein and concluded that the best estimates of this parameter fall between 2.5 and 4. Theoretical studies, using protein dielectric constant value of 2 to 40, and experiments on dried proteins, showing ε_p values in the range from 2.5 to 3.5, are cited in the same paper. A value of 2, which reflects solely the electronic polarisability[158], is used in studies where electrostatics calculations are performed for a strictly fixed conformation and atomic motions are considered explicitly[159,160]. The most frequently used value of protein permittivity is 4. This value presumes small fluctuations in the positions of atomic nuclei and approximately accounts for the response of the dipolar groups of the polypeptide backbone[158]. Values between 10 and 20 were proposed on the basis of comparison between theoretical and experimental investigations of protein

ionisation properties[16,18,47,161,162]. These high values, in fact, implicitly account for the protein responses, like redistribution of atomic partial charges and/or local conformational changes, upon changes in the ionisation state. However, in approaches explicitly considering such responses, low dielectric constants are used[55,163-166,II-VI]. Demchuck and Wade[65] analysed the dependence of calculated pKa values on the protein permittivity and found that a high values of ε_p lead to better agreement with experiments for residues that are on the surface of the proteins, while for buried residues low values are more appropriate. This observation was practically confirmed by Simonson and co-workers, who investigated the protein dielectric relaxation from dynamic viewpoint[167-170] and estimated the dielectric constant to be about 10-20 at the more flexible surface of proteins and about 2-4 in the protein interior. Other authors have considered a representation of the protein molecule as a body with inhomogeneous permittivity as a possible solution of the problem. For instance, a high dielectric constant can be attributed to regions in proteins containing polar side chains[171]. Sharp et al.[172] have proposed calculation of the local dielectric constant based on Clausius-Mossotti equation. Other equations (Debye, Onsager, Kirkwood) that relate the microscopic properties, such as polarisability and dipole moment, to the macroscopic dielectric constant are also well known. However, they all treat homogenous matter, while in the case of proteins, it seems to be more relevant to consider inhomogeneous and anisotropic matter. Voges and Karshikoff[173] attempted to treat the protein molecule as an inhomogeneous dielectric medium and proposed a model, which leads to a position dependent dielectric constant in the protein moiety. The potential advantage of multi-permittivity models triggered some of the recent improvements in DelPhi[57].

4.2 Other dielectric models

The pursuit for reducing the computational expenses in electrostatic calculations led to the development of approximate methods, in which the potential is calculated from a simple analytical expression resembling the form of the Coulomb potential. Coulomb potential with distance dependent dielectric constant was applied for analysis of electrostatic interactions in DNA conformers[174] and for p*K* calculations[175]. The latter was sharply criticised[176] mainly for the negligence of the self-energy contribution to the ionisation equilibrium (factor 3 in Table 2). The idea of introducing a distance dependent permittivity originates from series of theoretical studies, where the screening of the electrostatic field was investigated on a microscopic level[117,171,177-180].

Another, recently very popular approach is the generalised Born (GB) approximation, nicely reviewed by Bashford and Case[93]. The molecule of interest (e.g. a protein) is considered as a material with uniform low dielectric constant, ε_p, immersed in a continuum solvent characterised by a high permittivity, ε_s, and eventually a Debye parameter, κ. Each atom in the molecule is represented as a sphere with a given radius R_i and a charge q_i at its center. The total electrostatic energy (the work needed to create the charge distribution in the given dielectric environment) is split in two parts: first, the energy of the charge constellation in a homogeneous environment with permittivity ε_p,

which is just sum of Coulomb energies, and second, the work for transferring the molecule from the homogeneous environment into the solvent. The second part, accounting for the effects of the polarisation charges at the solute-solvent dielectric boundary, is approximated (GB approximation) by a sum of Coulomb-like terms. The final expression for the energy allows a GB analogue of the electrostatic potential to be defined[52]. Compared to the GB model, the classical one, described by PBE, is more general not only because it treats polarisation effects in an exact fashion, but also because it dose not impose any restrictions on charge distribution and number of dielectric regions. However, in the majority of the applications both models use, in fact, identical representation of the protein-solvent system. The GB approach employs a set of adjustable parameters allowing energy estimates to be improved towards experimental data or FD solutions of PBE via careful parameterisation. In tasks, where an explicit consideration of molecular conformers is of interest, the minor errors produced by GB, compare to solutions of PBE, are paid-off by the enormous gain of computational time.

The GB approach was initially dedicated to estimate by a simple analytical function the polarisation part (G_{pol}) of the solvation free energy for small molecules with potential application to MM force fields for implicit solvent[181]. Assuming $\varepsilon_p=1$ and *n* atoms in the molecule, the GB approximation gives $G_{pol} = -\frac{1}{2}\left(1-\frac{1}{\varepsilon_s}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_i q_j}{f_{GB}(r_{ij},\alpha_i,\alpha_j)}$, where the effective distance parameter, f_{GB}, is a function of the distance between atoms *i* and *j*, r_{ij}, and the Born radii α_i and α_j of these atoms ($\alpha_i \neq R_i$). The function f_{GB} is not uniquely defined as it is only required to be simple and smooth and to reproduce the exact theoretical expressions for the electrostatic energy in the limiting cases of $r_{ij}\rightarrow\infty$ and $r_{ij}\rightarrow 0$. This function is the main subject of parameterisation in the GB theory. Still and co-authors[181] proposed $f_{GB} = (r_{ij}^2 + \alpha_i\alpha_j\, exp(-r_{ij}^2/4\alpha_i\alpha_j))^{1/2}$. Later, the GB theory was extensively parameterised over large set of compounds[182]. Jayaram and co-workers[183] found inconsistence between the representation of the charge-charge and that of the self-energy components of G_{pol}. Introduced in this way errors compensate in the total solvation energy but affect the intramolecular interactions and hence p*K* evaluations. The problem was solved by yet another refinement of the parameterisation of the Born radii. Salt effects were introduced in the GB theory on Debye-Hückel level by substituting $(1-1/\varepsilon_s)$ with $(1-\exp(-\kappa f_{GB})/\varepsilon_s)$, which was found to be a reasonably good approximation[184]. The GB theory was further modified to allow a more accurate treatment of macromolecules by abandoning $\varepsilon_p=1$ (substitution $(1-1/\varepsilon_s)\rightarrow (1/\varepsilon_p-1/\varepsilon_s)$ or $(1-\exp(-\kappa f_{GB})/\varepsilon_s)\rightarrow (1/\varepsilon_p-\exp(-\kappa f_{GB})/\varepsilon_s)$)[52] and by using integration over the volume of the solute to properly adjust the Born radii[52,54]. p*K* calculations performed on three different proteins showed, with some minor exceptions, that GB energy estimates and energies calculated from PBE lead to practically identical p*K* values.

5 p*K* calculations based on a single protein conformation

5.1 Basic concept

The difference between the protonation/deprotonation equilibrium of a given titratable group in the protein, in principle, differs from the ionisation behaviour of this group taken as a free amino acid or as a model compound. This difference can be analysed by the thermodynamic cycle shown in Fig. 5.

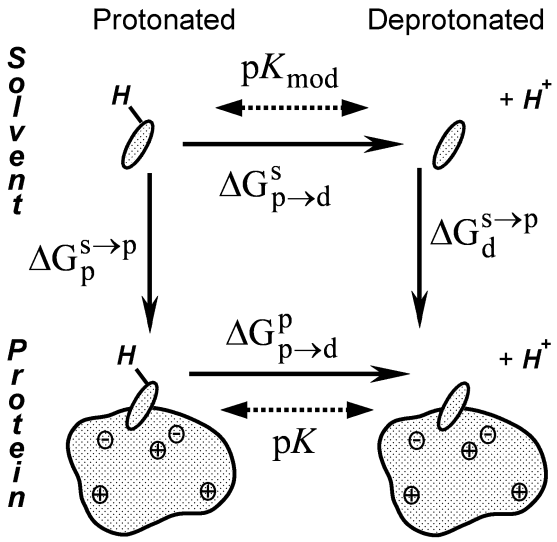

Figure 5. Thermodynamic cycle used for the calculation of the p*K* values of titratable groups in proteins.

In this cycle, each titratable group is considered as an appropriate model compound, which is transferred, in its protonated and deprotonated form, from solution to its place in the protein. This cycle relates the free energy of deprotonation, $\Delta G^{P}_{p\rightarrow d}$, of a given group in the protein to the deprotonation energy of the model compound, $\Delta G^{S}_{p\rightarrow d}$, as follows:

$$\Delta G^{P}_{p\rightarrow d} = \Delta G^{S}_{p\rightarrow d} + \Delta G_{d}^{S\rightarrow P} - \Delta G_{p}^{S\rightarrow P}, \qquad (17)$$

where $\Delta G_{d}^{S\rightarrow P}$ and $\Delta G_{p}^{S\rightarrow P}$ are the free energies of transfer of the titratable group (as a model compound) from solution to its location in the protein in deprotonated state and in protonated state, respectively. The p*K* value of the model compound, pK_{mod}, is usually taken from experimental results or, in special cases, from QM calculations. Therefore, as far as $\Delta G^{S}_{p\rightarrow d} = RT\ln 10(pK_{mod} - pH)$, $\Delta G^{P}_{p\rightarrow d}$ at given pH (or its equivalent, the pK of the group in the protein) can be obtained by calculating the difference in the corresponding transfer energies. These energies, however, depend on number of factors (see Table 2), a proper consideration of which is often computationally prohibitive. To reduce the complexity of the task, the following four

assumptions are usually made: ***(i)*** the protein can by represented by a single conformer, ***(ii)*** $\Delta G_d^{S\to P} - \Delta G_p^{S\to P}$ is entirely electrostatic in origin, ***(iii)*** the interaction between any couple of protein charges dose not depend on the presence of other protein charges, ***(iv)*** the statistical mechanical state of the protein is uniquely defined by the protonation state of each titratable group. These assumption constitute the ground for p*K* calculations based on continuum electrostatic models. The first assumption means that conformational changes that may occur upon ionisation are neglected or, at least, considered to be unessential for electrostatics, which can be true for a narrow pH interval. Calculated energies and hence of predicted p*K*s may be strongly dependent on the choice of conformer. Thus, the negligence of conformational diversity is considered to be the main origin of discrepancies with experimental data. On the other hand, p*K* calculations based on single conformers may be indicative for the role of flexibility in certain observable phenomenon[II]. Most of the approaches taking into account the conformational flexibility of proteins (see next part) are based on series of single structure calculations perhaps because it is the rigid body assumption, which permits to define uniquely dielectric boundaries and charge distributions, employed in macroscopic electrostatic calculations. The second assumption, which is quite reasonable in the frame of the first one, reduces the energy evaluation to electrostatic calculations. The third assumption, which (as discussed above) is automatically fulfilled if electrostatic calculations are based on the LPBE, allows different contributions to the electrostatic energy to be considered separately. In other words, the additivity of the electrostatic potential allows to represent the full electrostatic energy as a sum of independent energy terms that correspond to different sources of the electrostatic field.

With the above postulates, the difference in the transfer energies ($\Delta G_d^{S\to P}$ - $\Delta G_p^{S\to P}$ in Eq. 17) in can be represented as sum of three different electrostatic contributions (corresponding to factors 1-3 in Table 2). Hence, the free energy of deprotonation of *i*-th titratable site in the protein, ΔG_i ($\Delta G^P_{p\to d}$ in Eq. 17) can be written as:

$$\Delta G_i = RT ln10(pK_{i,mod} - pH) + \Delta G_{i,sol} + \Delta G_{i,pc} + \Delta G_{i,tc} , \qquad (18)$$

The last three terms in Eq 18 are in fact the subject of calculations. The contribution of the desolvation energy, $\Delta G_{i,sol}$, arises from the fact that the electrostatic energy (self energy or Born energy) of a charged body (*i*–th titratable site) depends not only on its charge distribution but also on the permittivity of the surrounding medium. It is calculated as:

$$\Delta G_{i,sol} = \frac{1}{2}\sum_{k\in\{i\}}\left[(\varphi^p(\rho_{id},k) - \varphi^s(\rho_{id},k))q_{id}(k) - (\varphi^p(\rho_{ip},k) - \varphi^s(\rho_{ip},k))q_{ip}(k)\right] \qquad (19)$$

where the summation is over all atoms *k* of group *i* with charges $q_{id}(k)$ and $q_{ip}(k)$ corresponding to charge distributions ρ_{id} and ρ_{ip} of this group in its deprotonated and protonated state. $\varphi(\rho,k)$ is the electrostatic potential at location *k* created by a charge distributions ρ. The superscripts *s* and *p* indicate solvent and protein environment of the model compound, respectively.

The contribution of protein permanent charges, $\Delta G_{i,pc}$ (Eq. 18), is given by:

$$\Delta G_{i,pc} = \sum_{k\in\{pc\}} (\varphi(\rho_{id}, k) - \varphi(\rho_{ip}, k)) q_{pc}(k) , \quad (20)$$

where *k* enumerates all atoms with permanent (non-titratable) protein charges $q_{pc}(k)$. Here and further upper indexes for the electrostatic potential are omitted as the calculations are relevant only if the model compound is placed in the protein.

It is also straightforward to define pair-wise charge-charge interaction, W_{ix_i,jx_j}, between *i*-th and *j*-th group in states x_i and x_j, respectively, as:

$$W_{ix_i, jx_j} = \sum_{k\in\{i\}} \varphi(\rho_{jx_j}, k) q_{ix_i}(k) , \quad (21)$$

where $i \neq j$ ($W=0$, if $i=j$) and each of the states x_i and x_j can be either protonated or deprotonated.

It is notable that the shifts in protonation/deprotonation equilibrium caused by the desolvation penalty and protein permanent charges depend only on the protein structure (i.e. on how the concrete group is situated in the protein) but not on the charge-charge interactions with other titratable sites. Hence, for fixed structure, it is convenient to introduce the quantity intrinsic pK value of site *i*, $pK_{i,int}$, as:

$$pK_{i,int} = pK_{i,mod} + (RT ln 10)^{-1} (\Delta G_{i,sol} + \Delta G_{i,pc}) , \quad (22)$$

which similarly to the model pK value, $pK_{i,mod}$, is pH independent. This, however, does not correspond to the Tanford definition of intrinsic pK[70], since the electrostatic interactions between *i-th* group and neutral states of the rest of the titratable sites are omitted. The introduction of these interactions in $pK_{i,int}$ makes it ambiguous if possible diversity of the neutral states is taken into account[51,III]. If each ionisable site is considered to have exactly one protonated and one deprotonated state, the Tanford definition can be formally fulfilled by some simple rearrangements of energy terms[II].

The last term, $\Delta G_{i,tc}$, in Eq. 18 accounts for the electrostatic interactions of *i*–th group with the rest of the titratable sites. This energy apparently depends on the protonation state of all titratable groups, which has to be determined. The correct solution of multiple-site titration problem is provided by statistical mechanics, first introduced for pK calculations by Bashford and Karplus[13]. The fourth assumption, stated in the beginning of this section, ensures complete statistical description of the protein molecule if the characteristics of each combination of protonation states of individual titratable sites (i.e. the characteristics of all possible protonation states of the protein) are available. Each unique protonation state of a protein with N ionisable sites is fully defined by an N–component vector, $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$, whose *i*-th component, x_i, has value of 0 or 1 depending on whether *i*-th group is protonated or deprotonated. At given state $\boldsymbol{x}$ of the protein and pH of the solvent, the energy of the protein-solvent system, $\Delta G(\boldsymbol{x}, pH)$, can be expressed as:

$$\Delta G(\boldsymbol{x}, pH) = RT ln 10 \sum_i x_i (pK_{i,int} - pH) + \frac{1}{2} \sum_i \sum_j W_{ix_i, jx_j} , \quad (23)$$

where indices *i* and *j* enumerate all titratable sites.

The Boltzmann statistical average of the variable x_i, as defined above, gives (analogously to Eq. 3) the degree of deprotonation, θ_i(pH), of group *i* at certain pH, i.e.:

$$\theta_i(\mathrm{pH}) = \langle x_i \rangle = \frac{\sum_{\{\boldsymbol{x}\}} x_i \, exp(-\Delta G(\boldsymbol{x}, \mathrm{pH})/RT)}{\sum_{\{\boldsymbol{x}\}} exp(-\Delta G(\boldsymbol{x}, \mathrm{pH})/RT)} \quad . \tag{24}$$

The summations in the above equation are over all 2^N possible protonation states $\{\mathbf{x}\}$. The calculation of θ_i(pH) is, in fact, equivalent to calculating the p*K* value (or the free energy of deprotonation) of *i*-th titratable site at the given conditions.

It is important to note, that the electrostatic potential resulting from FD solution of LPBE depends on how the investigated system is situated in the 3D grid, which leads to an energy erroneousness termed "grid energy". As far as p*K* calculations are based on energy differences, the grid energy practically cancels out if the four representations of each group (part of the protein or model compound, protonated or deprotonated – Fig. 5) are situated in the grid in identical way. Thus, in FD calculations of LPBE, model compounds have exactly the same conformation as the considered group in the protein. It follows, that energy calculations for ionisable sites of one and the same type employ different conformations of one model compound, however most often the titration of these different conformers is characterised by just one pK_{mod} value. This inconsistency can be resolved by considering energies of conformational transitions of the model compound, which is achievable by an additional thermodynamic cycle[164].

5.2 Allocation of polar hydrogens

The considerations in the previous section can be easily applied to an extended model of the protein molecule, which explicitly counts for possible alterations in the loci of titratable and polar hydrogens. Such alterations can be considered as introduction of conformational flexibility (within the limits of tautomers and rotamers), which does not violate the basic assumptions ensuring the successful application of continuum electrostatic theory in p*K* calculations. This is mainly due to the fact that reallocation of these hydrogens practically does not change the shape of the protein dielectric material. Hence, any variation in the hydrogen positions of site *i* does not affect electrostatic energies of the other sites, except for their pair-wise interactions with the *i*–th site. This approach allows not only prediction of pH-dependence of tautomerisation but also possible rearrangement H-bond network upon titration[III,VI].

The introduction of alternative proton positions within one group means that the individual sites (titratable or polar) may have more then two states. Generally, a given group can be fully described by a set of *n* microstates S_α (α = 0, 1, ..., n–1). Each microstate, α, is characterised by a certain number of titratable hydrogens, ν_α, and a specific charge distribution, ρ_α. As far as there is no rule in ordering the states, the choice of the reference state is in fact arbitrary and does not affect the final results or validity of the derived equations. In the further considerations the first state, S_0, of the

group will be used as reference state. The equilibrium of the microstates within a single group is determined by the microscopic equilibrium constants, K^{μ}_{α} , or their equivalent, pK^{μ}_{α} , defined as:

$$pK^{\mu}_{\alpha} = -logK^{\mu}_{\alpha} = -log\frac{[S_{\alpha}]}{[S_0]}[H^+]^{\Delta\nu_{\alpha}} = -log\frac{p_{\alpha}}{p_0} + \Delta\nu_{\alpha}\cdot pH \qquad (25)$$

where p_{α} is the population of state S_{α} ($\Sigma p_{\alpha} = 1$) and $\Delta\nu_{\alpha} = \nu_0 - \nu_{\alpha}$. In case of $\Delta\nu_{\alpha} = 1$, pK^{μ}_{α} coincides with the macroscopic pK.

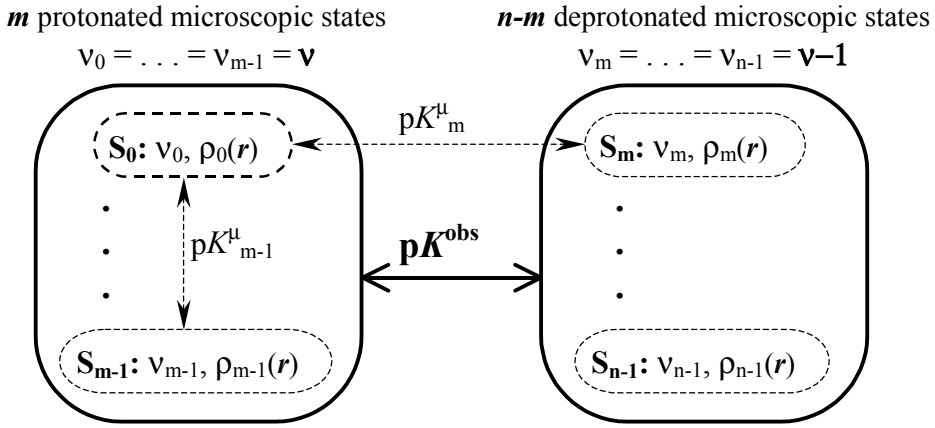

Figure 6. Schematic representation of the states of a titratable site with *n* microscopic and 2 (protonated and deprotonated) macroscopic states.

States with equal proton content are often experimentally indistinguishable and a single macroscopic pK, pK^{obs}, is observed. A useful relation between macroscopic and microscopic pK values can be derived if one considers a group with *m* protonated states, $\nu_0 = ... = \nu_{m-1} = \nu$, and the rest, *n-m*, deprotonated states, $\nu_m = ... = \nu_{n-1} = \nu - 1$, $(1 \le m \le n-1)$ (Fig. 6). If the ratios p_1/p_0 , ... , p_{m-1}/p_0 and p_{m+1}/p_m , ... , p_{n-1}/p_m are known, $pK^{\mu}_{\alpha,mod}$ can be calculated from:

$$pK^{\mu}_{\alpha} = \begin{cases} -log\dfrac{p_{\alpha}}{p_0}; & \alpha = 0,...,m-1 \\ pK^{obs} + \delta pK^{\mu}_{p} - \delta pK^{\mu}_{d} - log\dfrac{p_{\alpha}}{p_m}; & \alpha = m,...,n-1 \end{cases}, \qquad (26)$$

where $\delta pK^{\mu}_{p} = -log\sum_{\alpha=0}^{m-1} p_{\alpha}/p_0$ and $\delta pK^{\mu}_{d} = -log\sum_{\alpha=m}^{n-1} p_{\alpha}/p_m$

The terms δpK^{μ}_{p} and δpK^{μ}_{d} reflect the content of microscopic states in the protonated and deprotonated forms of a given titratable group, and hence have pure entropic meaning. It should be noted that, the charged form of a titratable group can be represented by a single microscopic state, as the hydrogen positions in this case are uniquely defined. However, a relevant description of a neutral site may require several microstates corresponding different hydrogen isomers, for instance tautomers of a deprotonated imidazole or rotamers of a protonated carboxyl/hydroxyl group. It follows that, for the acidic groups $\delta pK^{\mu}_{d} = 0$ and $pK^{\mu}_{\alpha} < pK^{obs}$ (α=m,…,n-1), while for the basic

sites $\delta pK^{\mu}_{p} = 0$ and $pK^{\mu}_{\alpha} > pK^{obs}$ (α=m,…,n-1). Eqs. 26 were applied to obtain the microscopic p*K* values for model compounds, $pK^{\mu}_{i\alpha,mod}$, which where than used in the calculations[III,V,VI]. All ionisable sites, except those of histidines, were considered to have equally populated neutral species in the model compounds. The ratio between the populations of histidine tautomers is set to correspond to a p*K* difference of 0.6[185].

The relation between microscopic equilibrium constants of the model compound and of the corresponding in the protein is obtained by a set of thermodynamic cycles shown in Fig. 7, which is simply an extension of the two state (protonated/deprotonated) model described in the previous section (Fig. 5, Eq. 17).

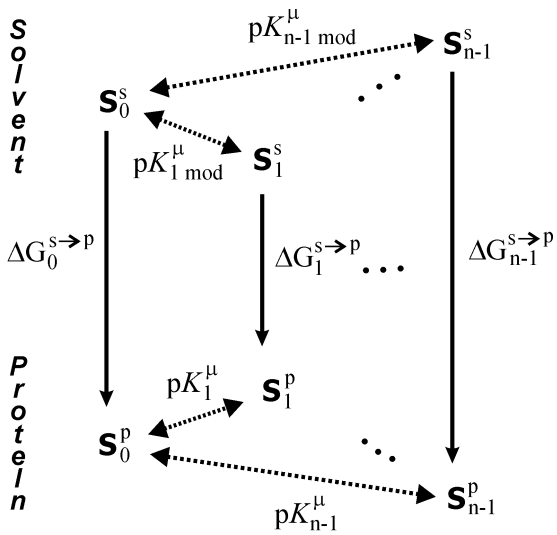

Figure 7. Thermodynamic cycles (*n-1*) used for the calculation of microscopic p*K*s in proteins, pK^{μ}, of a titratable or polar site with multiple (*n*) microscopic states. If the protonation (macroscopic) states of a titratable site are not degenerated, only one cycle (Fig. 5) is sufficient.

In this case a group *i* (titratable or polar) is transferred, in its reference state, S_{i0}, and in state, $S_{i\alpha}$, from solution into the protein. The free energy $\Delta G_{i\alpha}$ of the transition $S_{i0} \rightarrow S_{i\alpha}$ for *i-th* group in the protein, similarly to Eq. 18, can be written as:

$$\Delta G_{i\alpha} = RT\ln 10(pK^{\mu}_{i\alpha,mod} - \Delta\nu_{i\alpha}\cdot pH) + \Delta G_{i\alpha,sol} + \Delta G_{i\alpha,pc} + \Delta G_{i\alpha,tc} \quad (27)$$

The multi-state formalism can by developed by considering the transition $S_{i0} \rightarrow S_{i\alpha}$ instead of the transition protonated→deprotonated, applied in the two-state model. Thus, Eqs. 27-33 are the multi-state analogues of Eqs. 18-24.

$$\Delta G_{i\alpha,sol} = \frac{1}{2}\sum_{k\in\{i\}}\left[(\varphi^{p}(\rho_{i\alpha},k) - \varphi^{s}(\rho_{i\alpha},k))q_{i\alpha}(k) - (\varphi^{p}(\rho_{i0},k) - \varphi^{s}(\rho_{i0},k))q_{i0}(k)\right] \quad (28)$$

$$\Delta G_{i\alpha,pc} = \sum_{k\in\{pc\}}(\varphi(\rho_{i\alpha},k) - \varphi(\rho_{i0},k))q_{pc}(k) \quad (29)$$

$$W_{i\alpha,j\beta} = \begin{cases} \sum_{k\in\{i\}}\varphi(\rho_{j\beta},k)q_{i\alpha}(k) & i \neq j \\ 0 & i = j \end{cases} \quad (30)$$

$$pK^{\mu}_{i\alpha,int} = pK^{\mu}_{i\alpha,mod} + (RT\ln 10)^{-1}(\Delta G_{i\alpha,sol} + \Delta G_{i\alpha,pc}) \quad (31)$$

$$\Delta G(\boldsymbol{x}, pH) = RT ln10 \sum_i (pK^{\mu}_{ix_i,int} - \Delta\nu_{ix_i} \cdot pH) + \frac{1}{2}\sum_i \sum_j W_{ix_i, jx_j} \qquad (32)$$

$$p_{i\alpha}(pH) = \frac{\sum_{\{\boldsymbol{x}\}} \delta_{x_i,\alpha} \cdot exp(-\Delta G(\boldsymbol{x}, pH)/RT)}{\sum_{\{\boldsymbol{x}\}} exp(-\Delta G(\boldsymbol{x}, pH)/RT)} \; . \qquad (33)$$

Most of the notations have the meaning described previously. In Eqs. 32, 33 the vector $\boldsymbol{x}$, which uniquely defines the state of the protein molecule, contains M elements corresponding to the number of sites with variable number or spatial distribution of hydrogens. Each element $x_i = 0, 1, ..., n_i$, indicates the state of the *i*-th site. $\delta_{x_i,\alpha}$ is just the Kronecker delta ($\delta_{x_i,\alpha} = 1$ if $x_i = \alpha$ and $\delta_{x_i,\alpha} = 0$ if $x_i \neq \alpha$). Eq. 33 gives the population, $p_{i\alpha}$, of group *i* in microstate α. The degree of deprotonation of a given titratable group is just the sum of the populations of microstates representing the deprotonated form of this group. If one considers only titratable sites with two states each and values of 0 and 1 for protonated and deprotonated state of site *i*, respectively, are assigned to x_i, than $\delta_{x_i,1} = x_i$ and Eq. 33 reproduces Eq. 24 with $p_{i,1}(pH) = \theta_i(pH)$. In fact, the computation of microstate populations of ionisable and some of the polar groups provides essentially more detailed knowledge about protein properties than the calculation of only the degrees of deprotonation of titratable sites.

Different cases of tautomerisation/rotamerisation of ionisable sites were previously considered[15,186,187]. A statistical description of multiple sites with multiple states model was also shortly formulated by Schaefer and co-authors[47]. Entropic effects arising from non even number of neutral and ionised forms of given titratable groups were mentioned but, however, neglected by some authors[29,187]. Others have discussed these effects in the light of so-called "occupational entropy"[188].

The ionisation properties of individual titratable sites in *Aspergillus oryzae* ribonuclease T_1 (RNase T_1) were studied experimentally, by NMR spectroscopy[I], and theoretically, by two-state[II] and multi-state[III] approaches. For well-hydrated sites, which are not involved in strong electrostatic interactions, the two theoretical models gave practically identical results, which agreed very well with NMR observations. For the rest of the groups, the explicit consideration of variations in hydrogen loci resulted in some improvements in the theoretical prediction towards the experimental data. The most interesting outcome of the calculations was related to the ionisation of Asp76, for which seemingly contradictory experimental observations are available. On the basis of denaturation experiments on native and D76N mutant of RNase T_1, a pK value of 0.5 was evaluated for this residue[189]. On the other hand, the chemical shift of Asp76 $^{13}C^{\gamma}$ indicates that this residue is most likely protonated in the region of pH between 2 and 9[190]. Calculations without explicit consideration of hydrogen allocations[II] yielded a p*K* value of 6.5, which reflects the fact, that this residue is almost solvent inaccessible and is not salt-bridged, but it does not fit well to the experimental findings. The model used for multi-state calculations[III] included variable hydrogen positions for the ionisable and

some of the polar (Thr, Ser) residues as well as for one water molecule (Wat111 in the X-ray structure[191]), which is in the close vicinity of Asp76 and isolates the residue from the bulk solution. This approach reviled pK_{asp76}=2.0. This value is still higher than that evaluated by Giletto and Pace[189], which may be due to number of reason ranking from the quality of the evaluation to how representative, at pH below 2, is the structure used in the calculations (crystal was grown at pH 7). Calculations also showed that the ionisation of Asp76 is strongly correlated with dipole reorientation of the bounded water molecule so that at any pH Asp76 and Wat111 are H-bonded. Thus, a hydrogen atom is always present in the close vicinity of $O_{\delta 1}$ of Asp76. This result provides a possible interpretation of the fact that the chemical shift of $^{13}C_{\gamma}$ of Asp76 corresponds rather to the protonated carboxyl and its negligible pH dependence can be related to the synchronization between the ionisation of Asp76 and the reorientation of Wat111.

Drosophila lebanonensis alcohol dehydrogenase (DADH) catalyses the oxidation of primary and secondary alcohols to aldehydes and ketones, using NAD^+ as coenzyme. In the catalytic process, the active site (Tyr151, Lys155, Ser138) abstracts a proton from the hydroxyl group of the substrate (tyrosine-, lysine- and serine mechanisms have been proposed) and a hydride is transferred from the generated alcoholate anion to the NAD^+ molecule. Kinetic studies on the pH dependence of the different reaction steps led to the conclusion that one of the groups in the active site has a pK value of 7.1[192,193]. pK calculations including hydrogen multiple locations were performed on the crystal structure[194,195] of the binary complex E.NAD^+ of $DADH^{VI}$. Except for the protein residues, alternative hydrogen positions were allowed for a water molecule (W144 at H-bond distance from both Ser138 and Tyr151) and for the ribose hydroxyl group O2' of NAD^+. The calculations showed that: ***(i)*** Ser138 remains protonated even at pH 12, ***(ii)*** irregular and incomplete titration for both the residues Tyr151 and Lys155 and a midpoint at pH 7.2 for the overall titration of this couple, ***(iii)*** the orientation of the ribose hydroxyl is strongly correlated with the ionisation of the active site couple Tyr151-Lys155 and at pH higher than 8 acts as H-bond switch between these two groups, ***(iv)*** at high pH values a tiny amount of deprotonated W144 (i.e. OH^-) appears. These results suggest that couple Tyr151-Lys155 may act as a general base in catalytic reaction and thus is involved in the pH regulation of the proton abstraction, whereas Ser138 more likely plays a role in the correct orientation and binding of the substrate as it has been proposed earlier. The calculations also indicated that the O2' ribose hydroxyl group may play a key role in the process.

5.3 Computing statistical averages

Unfortunately, the direct application of the exact statistical mechanical calculations (Eq. 33) is quite limited because the CPU-time needed grows exponentially with the number of the considered sites. Nowadays computational facilities allow about 2^{30} states of the protein to be explored in a reasonable time. Apparently, the combinatorial task easily becomes unattractively time consuming even for small

proteins. There are methods, however, that can be used to overcome this problem by approximating the rigorous treatment.

According to the mean-field approach the pK shift of i-th group due to interactions with all other titratable groups is estimated as if i-th group interacts with the average charges of the rest of the groups. This concept was employed for pK calculations by Tanford and Roxby[196], who proposed an effective iterative procedure. At each iteration step, pK_i is calculated as a linear function of the average charges of other groups, which in turn are calculated from the pK values obtained in the previous iteration. Despite of the extremely low computational cost, the mean-field approaches were shown[14] to correspond to a non-correlated distribution function and thus as a rule are incorrect for sites participating in strong interactions. For those sites the iterations converge slowly or do not converge at all. This feature was used[153] to calculate pK values of weakly interacting groups and in the same time to separate them from those with co-operative ionisation, which are treated by more exact methods.

Groups that have pK_{int} far from the pH of interest do not undergo a protonation/ deprotonation transition. Such groups can be considered as being fixed in appropriate protonation state and can be excluded from further statistical calculations[14]. In the case of multi-state model, the irrelevant (improbable) for the given pH microstates are discarded. This stripping facilitates the calculations (especially at extreme pH values) but often dose not reduce the number of sites that should be treated statistically enough for direct application of Eq. 33.

Monte Carlo (MC) sampling of the states of the protein molecule was also successfully employed for pK calculations[35]. The MC sampling, in fact based on the Metropolis algorithm[197], is capable to provide reliable estimates of Boltzmann statistical averages even for big proteins. There is a variety of available techniques, all of them aiming to rapidly sample only the statistically "most important" states (states with lowest energy i.e. the most probable ones), which are only a small part of all possible states of the investigated system. These procedures can be interpreted to some extent as simulations of the thermal fluctuations of the considered system around states with energy minimum. In the standard implementation of the Metropolis sampling, the protonation state of a randomly (or sequentially) chosen site is altered and, with probability equal to the corresponding Boltzmann factor (i.e. depending on the energy difference between the new and the old state), the new state is selected or rejected. Each attempt for alteration of the protonation state is referred to as a MC step. The sequence of MC steps is equivalent to generation of a Markov chain for protein protonation states. The accuracy of MC methods depends on the length of the simulation and the specificity of the system. If the protein contains couples or clusters of strongly interacting groups with co-operative ionisation, the sampling is easily trapped in vicinity of a local energy minimum and a very long simulation is needed for achieving reliable estimates for the protonation states of those groups. This problem is avoided either by modifying the standard algorithm to allow also simultaneous (within one MC step) state alteration of two (or more) strongly interacting sites[35,198] or by producing a large number of relatively short Markov chains, which can be optimised by simulated annealing[154].

Computations can be essentially speeded up without significant violation of the accuracy of results by applying clustering methods[51,198,199]. In all these methods a set of strongly interacting (or closely situated) groups forms a cluster. The degree of deprotonation of these groups is calculated either by Boltzmann statistics or by Monte Carlo algorithm over the groups included in the cluster, while the influence of the rest of the groups is counted by mean-field approximation. Thus, the populations, $p_{i\alpha}$, of sites within one cluster are calculated according to Eq. 33, where $\{\boldsymbol{x}\}$, in this case, represents all possible combinations of states within the cluster. In order to count for the groups that do not belong to the cluster Eq. 19 is modified as follows:

$$\Delta G(\boldsymbol{x}, pH) = RT ln10 \sum_i (pK^{\mu}_{ix_i,int} - \Delta\nu_{ix_i} \cdot pH) + \frac{1}{2}\sum_i \sum_j W_{ix_i,jx_j} + \frac{1}{2}\sum_i \sum_k \sum_{\beta=0}^{n_k} p_{k\beta} W_{ix_i,jx_j} \quad (34)$$

where indices *i* and *j* enumerate all sites in the cluster, while *k* runs over the rest of the groups. The triple summation term (the third term of the right hand side of Eq. 34) represents the mean energy of interaction with the groups not included in the cluster. Such calculations are performed for all clusters (individual groups can be considered as clusters of one single group) and the whole procedure is iteratively repeated until a convergence criterion for the degree of deprotonation of all sites is met. Clustering methods differ from each other primarily in the way of defining clusters. The Tanford-Roxby procedure, in fact, coincides with the extreme case when all clusters consist of single sites only.

5.4 Titration curves

The titration curve, θ(pH) (i.e. the dependence of the degree of deprotonation on pH), of a titratable group can be reproduced by Eq. 2 if the quantity p*K* is known for any conditions (solution pH, salt concentration, temperature etc.). For dilute solutions of a compound with single titratable site (e.g. a model compound), p*K*=constant, at least at given salt concentration, temperature and pressure and assuming that the only reaction that may occur is the equilibrium proton exchange between the compound and the solvent. The value of this constant is a characteristic of the compound called p*K* value. The titration curve in such case is a simple sigmoid with an inflection point at θ=1/2 and pH=p$K_{1/2}$=p*K*. As it was already pointed out, owing to the interplay of different factors, the ionisation properties of an amino acid in a protein may essentially differ from those of a corresponding model compound. Using $\Delta G_i = RT ln10(pK_i - pH)$ and $\Delta G_{i,sol/pc/tc} = RT ln10\, \Delta pK_{i,sol/pc/tc}$, one can represent Eq. 18 in p*K* terns as:

$$pK_i = pK_{i,mod} + \Delta pK_{i,sol} + \Delta pK_{i,pc} + \Delta pK_{i,tc}(W_{ix_i,jx_j}, \theta_j(pH))$$

If the desolvation of *i*-th group in the protein interior and the electrostatic background created by the protein permanent charges remain unchanged in a certain pH

range (this is strictly fulfilled for fixed protein structure), both terms $\Delta pK_{i,sol}$ and $\Delta pK_{i,pc}$ are constants causing just a pH-independent shift in pK_i form the $pK_{i,mod}$ value. This results in a simple shift of the titration curve of the model compound along the pH scale (the value $pK_{i,mod}$ is substituted by the value $pK_{i,int} = pK_{i,mod} + \Delta pK_{i,sol} + \Delta pK_{i,pc}$). Thus the sigmoidal shape of the titration curve is preserved, however the change in the protonation/deprotonation equilibrium may be quite significant. Perhaps the most famous example for an extreme p*K* shift, is that of the hen egg white lysozyme active site Glu35, which has a p*K* value of 6.2[200]. Dramatic p*K* reduction due to burial of basic groups is also observed[161]. The charge-charge interactions between titratable groups depend on their degree of ionisation, which are function of the pH. It follows that, even in absences of pH-induced conformational changes, the equilibrium ionisation constants of the individual sites are pH-dependent ($pK_i(pH) \neq$constant due to the last term, $\Delta pK_{i,tc}$, in the above equation). This causes changes in the shape of the individual titration curves, which, depending on the co-operativity of the ionisation, may rank from slightly perturbed sigmoid to completely different shapes, called irregular[III] or complex[201] titrations. Commonly, most of the groups, in the pH range of their titration, experience only mild changes in the electrostatic field created by the rest of the protein multipole, which results in a broadening of their ionisation curves. This effect is often successfully mimicked by formally setting pK_i=constant but introducing in Eq. 2 an additional parameter. The simplest and most popular modifications of Eq. 2 are equivalent to assuming a linear dependence of the actual pK_i either on the pH (Hill coefficient) or on the rate of changing the total protein charge with pH (Linderstrøm-Lang model[68]). Such corrections in Eq. 2, however, completely fail to describe the cases of non-sigmoidal curves (irregular titration).

Two-step ionisation curves are often observed, for instance, by the pH dependence of the chemical shift of the individual titratable groups. Traditionally, when such an irregular titration curve is observed, two p*K* values are extracted by fitting to sum of two sigmoids and assigned to two different titratable groups[I]. However, in some special cases, such an interpretation may be misleading[II,III]. This can be illustrated by a simple example of two strongly interacting groups, say acidic one and a basic one, both of which manifest two-step titration if the following conditions are satisfied[III]: ***(i)*** if the interaction between these two groups is excluded, the p*K* value–approximately pK_{int} if the influence of the protein multipole is minor–of the acidic group should be higher than the one of the basic i.e. $\Delta pK_{non} \approx pK^{\mathbf{acid}}_{int} - pK^{\mathbf{base}}_{int} > 0$; ***(ii)*** the electrostatic interaction energy between their charged forms (assuming negligible interaction between the neutral forms), taken in p*K* units, should be close to ΔpK_{non}; ***(iii)*** the pair-wise interactions and ΔpK_{non} are approximately constant and in the entire pH interval where the titration of the couple takes place. These conditions may be fulfilled, for instance, for buried and immobilised salt-bridges. The peculiar ionisation behaviour of such acid-base couple is shown in Fig. 8. Titration curves of both groups have a two-step shape, which apparently cannot be properly reproduced by the Henderson-Hasselbalch equation (Eq. 1 or its equivalent Eq. 2) assuming that a protonation/deprotonation equilibrium is fully characterised by a single p*K* value. It is notable, that each curve can be decomposed into two sigmoidal segments with clearly distinguished inflection

points. Moreover, the inflection points of the titration of the acidic group coincide with those of the basic one. The overall titration of the pair (Fig. 8B) can be perfectly simulated by a sum of two Henderson-Hasselbalch equations, with p*K* values of 3.6 and 6.4, which reflect the positions of the inflection points. In between, around pH 5, the total charge of the couple indicates abstraction of one proton from the system. Thus, if such curve is experimentally observed one intuitively addresses the inflection points at low and high pH values to p*K* values of the acidic and the basic groups. This is, however, an example of a misleading interpretation. This should not be interpreted as a complete deprotonation of one group and the other one remaining fully protonated, but rather as an equal sharing of one proton between the two groups.

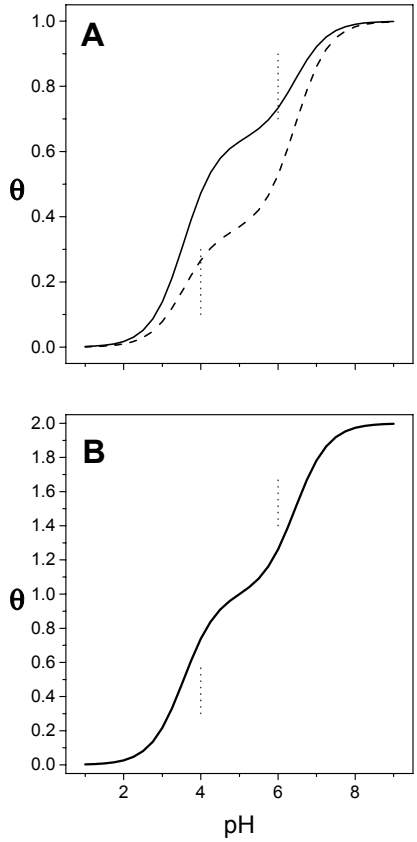

Figure 8. Co-operative ionisation of an acid-base pair. **A:** Irregular titration of the acidic (solid line) and the basic (dashed line) group. **B:** Overall titration of the pair. Vertical dotted lines indicate $pK^{\mathbf{base}}_{int} = 4$ (lower) and $pK^{\mathbf{acid}}_{int} = 6$ (upper). Influence of other titratable groups is neglected. Energy of electrostatic interaction between their charged forms is 3 kcal/mol (~ 2.2 p*K* units).

Clearly, in cases of strong co-operative titration the quantity $pK_{1/2}$, often derived from simple equilibrium functions, does not provide an appropriate characterisation of ionisation properties of the group. Statistical mechanical considerations provide a more realistic and detail interpretation of ionisation phenomena in proteins[201,III]. More complicated titrations can be observed in more complex systems e.g. large clusters of strongly interacting groups. In such a case, due to co-operativity of the system, a group may occur to be half protonated at several different pH values, which makes the determination of $pK_{1/2}$ uncertain. Irregular titrations also might be related to pH-dependent tautomerisation. Current theoretical approaches employ a lot of structural details in energy calculations and compute titration curves of individual sites on statistical thermodynamics basis and therefore are able to reliably predict and characterise irregular titrations[31,44,187,II,III,VI]. Thus, structure based p*K* calculations may

essentially contribute for the accuracy of the analysis of experimental ionisation curves. Conditions for appearance of irregular titrations in acid-acid, base-base and acid-base pairs were analytically derived and clearly formulated[III] (coupled ionisation in a theoretical model of two acidic groups (acid-acid pair) was also described earlier[198]). The achievement of those conditions is energetically demanding, which is most likely the reason complex ionisation to be observed relatively seldom compare to sigmoidal titration curves. One can further speculate that sites with irregular ionisation behaviour are most likely to have significant physiological importance (e.g. for enzymatic activity[VI]). Otherwise the particular residue constellation yielding irregular titration, being energetically unfavourable for the protein molecule itself, would have been perhaps discarded in the evolutionary process.

6 p*K* calculations accounting protein flexibility

Methods based on a single protein structure (usually determined by X-ray crystallography), as described in the first section of the previous part, characterise the protonation/deprotonation equilibrium correctly only in limited pH interval, for which the corresponding structure is relevant. Upon change of the charge state (change of pH), conformational changes may occur. One of the most prominent examples of pH-induced structural changes is the Bohr effect in hemoglobin. The sensitivity of the protonation/deprotonation equilibria of titratable groups to the conformational changes, which is a well-documented effect, appears to be one of the major problems for the accurate prediction of ionisation equilibria in proteins. Modern macroscopic approaches for p*K* calculations differ mostly in how the conformational heterogeneity is treated.

Protein flexibility may be implicitly accounted for by adjusting the protein dielectric constant (in that number, assigning site specific permittivities), as it was discussed above (section 4.1.3.). Although such methods generally improve the p*K* predictions, being still based on a single conformer, they are incapable to reveal any correlation between protonation and structural changes. Nevertheless, these seem to be still the methods of choice for large molecules due to the higher computational expenses required by other techniques.

In recent years a lot of efforts were devoted to developing computationally efficient algorithms for incorporating protein flexibility in continuum p*K* calculations in an explicit fashion. The milestones in these attempts are nicely reviewed[89,91]. A complete theoretical consideration of the problem has been elaborated by Spassov and Bashford[51]. These authors point out that the complexity of the task is prohibitively large and needs to be reduced. Thus, it is not surprising that all proposed ionisation equilibria calculations that explicitly account for conformational diversity employ additional simplifying assumptions. Most commonly, the conformational space of the protein is reduced to a limited number of conformers, defined either explicitly or by some combinatorics. This ensemble of pre-selected structures is assumed to adequately represent the protein conformational heterogeneity in solution in a certain pH range. Calculations of electrostatic free energies are performed as for single separate conformers, usually by FD solutions of LPBE, and the statistical considerations yielding titration curves (Eq. 24) are correspondingly modified (e.g. Eq. 33). Flexibility is introduced in titration simulations usually either by including local conformations i.e. small conformational changes in vicinity of ionisable sites (e.g. side chain conformers or their most restricted variant – polar hydrogen tautomers/rotamers) or by considering global conformations, in which structural changes involve the entire protein. A combination of global and local flexibility is, in principle, desirable if it is not computationally unfeasible[51]. As far as the ionisation free energies depend primarily on the local environment of the ionisable group, local flexibility can be considered as the main origin of p*K* changes. That is way, number of studies focus on a better representation of side-chain flexibility[15,58,164,186,187,202,III,VI], which improves the pK

prediction and provides very reliable results, especially when there are evidences for conserved backbone conformation in the pH interval of interest. However, such models are obviously incapable to predict changes in the protein fold with pH.

The ensemble of conformers used in p*K* calculations can be collected from available experimental structures or generated by computer simulations. Mixed approaches are also possible[29].

6.1 Ensembles of experimentally determined structures

High-resolution crystal structures, resolved by X-ray or neutron diffraction, often provide implicit (temperature factors) and explicit (alternative conformers) information about local flexibility, which can be used in p*K* calculations. The temperature factor was used for assignment of site-specific permittivity[203]. The influence of alternative side-chain conformers for ionisation properties of RNase T_1 was investigated and it was found that experimental observations for the titration of Glu28 and Asp29 can be reproduced only if Lys25 is considered flexible[II]. Knowledge about global flexibility can also be extracted from crystal structures, for instance, from structures obtained at different crystallization conditions or if the asymmetric unit contains more than one protein molecule (or analogously, in cases where the protein is a homo-dimmer/-trimer etc.) and local symmetries were not intensively used for structure refinement. When available, such information may be highly valuable for theoretical studies[29,204,205,IV,VI]. In the cases of *B. agaradhaerens* xylanase[IV] and DADH[VI] it was of specific interest to what extent the marginal structural differences affect the computed titration of individual sites.

NMR structure determination supplies, in fact, set of conformers presumed to give a good representation of protein structural diversity (global flexibility) in solution. Antosiewicz et al.[206] performed calculations on NMR models taken as single structures and obtained in this way p$K_{1/2}$ values were arithmetically averaged. Calculations were carried out for three proteins and, compare to single conformer calculations, an overall improvement towards experimental data was reported. The arithmetic averaging, however, supposes equal statistical weights of NMR conformers, which has no physical grounds. Further investigations[207] demonstrated that in the regions where NMR and X-ray structures differ significantly the p*K* values calculated on the basis of the X-ray structures are in better agreement with the experimental data. For solvent exposed residues, however, a better agreement with the experimental results has been obtained using the NMR structures. One can speculate that the crystal contacts are one of the main sources of discrepancy between the calculated and observed p*K* values in general. Although the averaging was over quite a limited and probably far form the representative set of structures, the result indicates that the consideration of more then a single conformation is a promising way to improve the theoretical prediction of ionisation equilibria in proteins. Calculations over 34 NMR models of RNase T_1 were carried out and the results confirmed the stated above general conclusions. For buried or strongly interacting groups (~30 of the groups), a large dispersion ($\max\Delta pK_{1/2} > 5$ pH

units) of the titration curves was observed and the further analysis did not suggest a straightforward way for retrieving any specific biologically significant information[II].

In general, calculations strictly based on one or more experimental structures tend to be biased towards reproducing the protonation state of the molecule corresponding to the conditions at which the structures were determined. Thus, results of such calculations are most reliable for pH close to that of the experimental conditions. In addition, side-chain conformations (often decisive for the outcome of titration simulations) given in X-ray and NMR structures are not always representative for the actual solution behaviour of the protein. For instance, it is easy to spot regions in X-ray structures where the local conformation may be artefact of the crystal packing, while in NMR models, due to lack of experimental data, side-chain conformations are usually result only of simple molecular modelling.

When ionisation equilibria are calculated for an ensemble of conformers (no matter how this ensemble is generated) a question arises how to properly estimate the statistical weight of each single conformer. A rigorous systematic approach to this problem was suggested by You and Bashford[164](see also[91,208]). This approach requires for each conformer the electrostatic energy to be calculated (FD LPBE) twice (first, with homogeneous dielectric with $\varepsilon=\varepsilon_p$ in the whole space and second, with heterogeneous permittivity as in the standard continuum model shown in Fig. 1) and combined with non-electrostatic energies, which are convenient to obtain from a MM force field. This method in fact evaluates the energy difference between arbitrary couple of conformers at given protonation state in solution through an additional thermodynamic cycle. This cycle is needed mainly to reduce the dependence of the energy difference estimates by cancellation of the grid energy arising from the numerical solution of the LPBE. Although this approach is general and has no principle limitations, its rigorous implementation is apparently computationally unattractive or even unfeasible if the full combinatorics of possible local conformers has to be explored. In particular, such technique can be applied to model compounds to remove the requirement that the model compound should always match the conformation of the concrete ionisable site under consideration (see 5.1.).

6.2 Computer generated ensembles of conformers

Various protocols combining calculations of ionisation equilibria based on continuum electrostatics with simulated protein flexibility has been elaborated. In different approaches, structural changes rank from involving only hydrogens[15,187,III,VI], through side-chains fluctuations[29,58,164,186,202], to accounting for global flexibility by applying MD[18,31,209-211,IV] or MC dynamics[212,213]. Different minimisation procedures or energy cut-off criteria are also often used to reduce in a reasonable way the sampling of the conformational space and the total statistical combinatorics[29,31,58,186,203,208].

You and Bashford systematically explored the conformational space defined by the side chain torsion angles[164], while Beroza and Case[186] used two conformations for each titratable side chain, one as determined by X-ray crystallography and the other

with maximized solvent accessibility (this can only be superficially reasoned). In both approaches as well as in all other studies considering combinatorics of local conformations, simplifying assumptions are made so that electrostatic calculations are carried out only for few instead for all possible conformers.

Van Vlijmen an co-workers[18] have used 100 ps MD simulations for generating ensembles of structures. An advantage of this approach is that, according to the ergodic hypothesis, conformers collected along one MD trajectory can be assumed equally populated without additional energy calculations. p*K*s were calculated by different procedures corresponding to arithmetical averaging of different quantities. The most relevant of these procedures, as pointed out also by the authors, is the averaging of titration curves, although all of them showed similar results and an overall improvement in p*K* prediction. However, the prediction of the p*K* values of some groups has not been improved. One reason for this, as the authors point out, might be the relatively short time of conformational sampling.

The influence of the initial structure and the simulation length was investigated by applying analogous computational protocol, however with 1 ns MD for three initial xylanase structures (*B. circulans* xylanase and two similar structures of *B. agaradhaerens* xylanase)[IV]. Results confirmed the general observation that calculations using an ensemble of structures describe titration properties of proteins more accurately than calculations based on a single structure, although for several sites (among which the active site residues), the improvements in p*K* values were quite insufficient. The minimum simulation time needed to achieve approximately constant average p*K* was 500 ps for the majority of titratable sites and more 1 ns for a couple of sites. The time evolution of p*K* values calculated on the base of slightly different initial structures of *B. agaradhaerens* xylanase demonstrated that the two MD runs most likely sampled different parts of the xylanase conformational space.

Nielsen and McCammon[205] investigated the quality of 41 X-ray structures of lysozyme with respect to continuum based p*K* calculations. The authors also tested MD, energy minimization and structure-perturbation methods as tools for structure optimisation. They concluded that optimisation techniques, in principle, do not improve the p*K*s of the active site residues.

Wlodek et al.[204] have used two different initial structures, each with two different charge states. In this way p*K* values from four 100 ps MD trajectories were calculated. They have shown that the differences in the calculated p*K* values are statistically insignificant for the majority of the groups (somewhat contradicting to more recent studies[205,IV]). It was demonstrated that the simulations correctly respond to the preliminary determined difference in the charge states i.e. MD runs with particular site (N-terminus) protonated yield structures stabilizing the protonated form of this residue and *vice versa*. This highlights the importance of the initial choice of the protein protonation state for MD simulations. It reveals another disadvantage (besides the one, that very long trajectory may be needed[IV]) of using MD generated structures for characterising the protein ionisation properties, namely that pK values are likely to be biased to the initial protonation choice. An attempt to overcome this problem was made by collecting conformers from two independent MD runs (in which case benefits from

the ergodic hypothesis are, in principle, lost): one with all titratable sites charged and one with all sites neutral[209,210]. Such approach may be, however, biased to two extreme unrealistic states. Another interesting approach emphasises on incorporating ionisation events in MD[214,215]. Such algorithm, in principle, can provide titration curves by separate runs corresponding to different pH values. The performance of this method for p*K* calculations is still not well established.

7 Continuum simulations of ion flux through protein channels

A large number of indispensable physiological processes involve transport of substances across biological membranes. The transport is nearly always realised and regulated by membrane proteins forming specific channels. Thus, such proteins, and in particular ion channels, have been subject of extensive experimental and theoretical investigations, which are nicely summarized in series of works[216-220]. Apparently, any thorough structure-based analysis of electrodiffusion properties of an ion channel requires, besides detail knowledge of the pore size and shape, a relevant description of the electrostatic field inside and in the vicinity of the channel, which, in general, invokes preliminary p*K* calculations[44,V]. The electrostatic field of porin Omp32 from the bacterium *Delftia acidovorans* was examined in detail[V]. The pH dependence of the field and the contribution of specific constellations of ionisable residues to electrodiffusion characteristics of this and closely related porins were discussed. The results of this study, however, are not directly comparable to the available patch-clamp measurements[221].

Different computational approaches are currently applied for prediction and analysis of electrophysiological properties of membrane channels. The earliest theories, based on kinetic rate models (well described by Hille[216]), are too formal and irrelevant if structure-function relationship is of interest[217,218]. The most rigorous direct simulations, like MD, BD and FEP (see reviews[219,220,222]), still have limited application because of the extreme computational demands for achieving reliable estimates of measurable currents[218]. Models based on Poisson-Nernst-Planck (PNP) theory of electrodiffusion are also widely used for calculation of ion conductivity and selectivity of channels[217,218]. These approaches utilise continuum representation of the protein-membrane-solvent system, where interactions determining the effective motion of solvent ions are approximated by mean-field, to reproduce a stationary ion fluxes (in the particular case of equilibrium, no ion fluxes, PNP equations are reduced to PBE). As a consequence, the specific interactions of mobile ions with the channel lining as well as ion-ion interactions are not considered quite accurately, which is partially masked by adjusting some macroscopic parameters (usually the diffusion coefficients). Protein flexibility is also neglected. The PNP models are often criticised because of the above weaknesses[219,222]. The omission of the desolvation of the mobile ions in the pore region was pointed out as a major problem of the traditional PNP approach, especially when it is applied to narrow channels[223-225]. Recently, Schuss et al.[226] derived a coupled system of PNP equations, where desolvation effects are correctly taken into account. However, as noted by the authors, a direct implementation of this theory is impossible because it relays on a conditional probability of ion distribution, which cannot be determined without further simplifications. In most of the applications of the electrodiffusion model, the real 3D task is reduced to 1D, which can be a reasonable approximation only

for narrow cylindrical channels. This artificial simplification was removed by Kurnikova and co-workers[227,228], who elaborated a self-consistent FD solution of the 3D PNP equations and applied this technique to the gramicidin A channel. The nature of the considered system and the simulated phenomenon[218] as well as the good agreement between the theoretical and experimental results, reported in many papers[143,227-230], suggest that the continuum model can provide, after all, a useful tool for analysis of steady-state ion fluxes though membrane channels. In order to benefit from the relative conceptual and computational easiness of the continuum model and in the same time to reduce the major disadvantages of the standard implementation of the PNP theory, a new algorithm for calculation of 3D PNP equations was developed[VII]. The approach could be considered as an extension of the one described by Kurnikova et al.[227,228], in which several essential improvements are introduced: ***(i)*** the desolvation of mobile ions in the membrane pore and effects related to ion sizes are taken into account, ***(ii)*** numerical obstacles, which in many cases prohibit the straightforward solution, are overcome, making the algorithm applicable for arbitrary channel geometry and arbitrary protein charge distribution, ***(iii)*** boundary conditions are refined by a focusing procedure.

7.1 Model and computational technique

The theoretical model and its numerical implementation are described and discussed in detail in the manuscript[VII]. Here only the main points will be outlined.

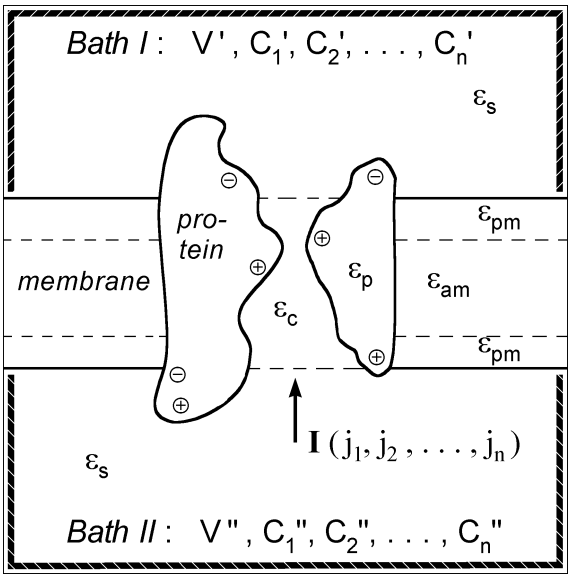

Figure 9. Schematic representation of the model used for calculations of ion fluxes through membrane channels.

The model used for calculations of the I-V characteristics of ion channels is illustrated in Fig. 9. The solvent is represented as a medium with a permittivity, ε_s, which contains *n* species of mobile ions treated as a continuum matter, hence described by the space distributions of their local concentration, $c_i(\boldsymbol{r})$, i = 1, …, n. Ion types are characterised by their charges, $Z_i e$, diffusion coefficients, $D_i(\boldsymbol{r})$, and radii, R_i. The latter define ion inaccessible regions around the membrane and the protein. The membrane is represented as a planar layer with a certain thickness in which the protein molecule is

immersed. Both the membrane and the protein are considered as a material with appropriate dielectric constants (Fig. 9), which is impenetrable for the solvent and the mobile ions. In this way, the solvent is separated by the membrane into two bath regions, which can exchange ions only through the protein channel. The shape and dimensions of the channel are defined by the 3D protein structure and the van der Waals radii of the protein atoms. The protein possesses also a charge distribution, $\rho_p(\boldsymbol{r})$, determined by its structure and ionisation state. The ion concentrations in the reservoirs on both sides of the membrane may be different, i.e. $C'_i \neq C''_i$, and an external voltage, V''-V', can be applied.

The algorithm is flexible and can produce a more complex mapping of permittivity and diffusion coefficient, thus allowing the system to be represented in more detail. For instance, similarly to the protein dielectric constant (see 4.1.3), the values of ε_s and D_i in the region of the solvent-membrane interface and in the channel interior are not well defined and, in general, are expected to differ from those in the bulk. On the other hand, these parameters are crucial for the final outcome of the simulation and can serve as justifiable parameters of the model.

In steady-state of the above described system, the dynamics of ions of type *i* (*i*=1…*n*) is given by the corresponding Nernst-Planck (NP) equation: $\nabla \cdot \left[D_i(\boldsymbol{r}) \left(\nabla c_i(\boldsymbol{r}) + (kT)^{-1} c_i(\boldsymbol{r}) \nabla U_i(\boldsymbol{r}) \right) \right] = 0$, where $U_i(\boldsymbol{r})$ is the potential energy of an ion of species ***i*** at point ***r*** and the term in the square brackets is the flux of this type of ions at the same point. Each NP equation yields a unique concentration distribution in any volume if the corresponding diffusion coefficient and potential energy values are known for the entire volume and the concentration at the boundary surface is given. These ion concentration distributions can be than used to calculate ion fluxes and the electric current through the channel. Clearly, as far as ions are charged particles, $U_i(\boldsymbol{r})$ contains electrostatic part, which can be assessed from the Poisson equation (Eq. 5, giving the men-field electrostatic potential of the entire system). In the standard PNP theory, it is assumed that the electrodiffusion process is fully described by a the Poisson equation together with *n* Nernst-Planck equations. This assumption is absolutely reasonable for solution volumes much larger than the ion sizes, while justification seems be needed for narrow channels with dimensions comparable to those of ions. The PNP theory was adapted[VII] by modifying the potential used in the NP equations in the following way. First, ion solvation energy was introduced to $U_i(\boldsymbol{r})$ as additional external potential considered to be independent on the charge-charge interactions and ion distributions in the system. Second, ions were locally treated as discrete particles (hard spheres) with non-negligible size (new atomic detail was introduced in the continuum model) and the mean-field electrostatic potential was accordingly corrected.

An FD based algorithm for self-consistent solution of the resulting system of equations was proposed. It was shown that the FD representation of arbitrary NP equation dose not lead automatically to a non-singular linear (at fixed potentials) system of equations, as is the case with PBE and LPBE. Singularity (and hence existence of solution) was ensured by artificial truncation of those extreme potential gradients that may cause non-singularity. The solution of the standard or the modified PNP equations

is achieved (or rather approached) by a self-consistent iterative procedure. Due to the non-linearity of the entire system, this procedure is prone to diverge, which was compensated by adjustable mixing of concentration maps produced in subsequent iterative steps. Similar numerical techniques were used earlier for solving the PNP equations[227,228,231] or their equilibrium analogue PBE[129]. In order to use fine grid with adequate boundary conditions, an appropriately modified focusing procedure was applied. Simple boundary conditions were used for the initial box. In the region of the focused box, however, the solvent polarization effects and the resulting Donnan potentials were satisfactorily reproduced. The focusing procedure and the entire algorithm were thoroughly tested on pore designed to resemble real protein channel. It was demonstrated that the computed electric current depends on how the pore is mapped on the grid, which in fact determines the precision of the numerical solution. The magnitude of this grid dependence is completely acceptable, although it is significantly larger than that in the p*K* calculations, where grid artefacts are strongly reduced within the applied thermodynamic cycle (see 5.1.). The results after focusing were practically identical to the usage of a big computational box with fine grid and it was also shown that this technique indeed improves reliability of the final results.

7.2 Applications to model channels

Simulations on model channels were used to illustrate the role of the pore shape and protein charge distribution in formation of basic electro-diffusion properties, such as channel conductivity and selectivity. The concentration distributions of mobile ions inside channels and close to the were also analysed. The importance of an adequate consideration of solvation effect was highlighted. The influence of the ionic strength in the bulk solution and of the externally applied electric field on channel properties were also discussed.

It was shown that the omission of ion solvation energy leads to an overestimation of the ion content in the channel (especially close to the channel walls) and thus to larger calculated currents. In charged channels, the overestimation of the total electric current tends to be reduced in parallel with a tendency of dramatic underestimation of the selectivity. Another interesting finding even completely neutral channels may manifest kind of selectivity due to the interplay between the external voltage and channel shape.

The modified PNP approach yields distribution profiles of mobile ions, which are in accord with the predictions based on the direct microscopic methods (MD or BD). Non-linearity and asymmetry of I-V curves, as well as ion selectivity and its dependence on channel architecture and ionic strength, are also adequately reflected. Unlike previous applications of the PNP theory, this method provides reasonable results even for extreme cases of narrow and/or highly charged channels.

Although the modified PNP approach does not overstep the limitations typical for continuum models, it allows the electro-diffusion properties of any protein channel with known structure to be scrutinized on an atomic level.

8 Acknowledgements

This entire work and every single piece of it are largely due to the expert guidance of my supervisor Assoc. Prof. Andrey Karshikoff who introduced me to the field of theoretical biophysics and to whom I am profoundly grateful. During the past four years, he has been not only an excellent tutor, always eager to provide utmost help on scientific and practical issues, but he also offered me his sincere friendship, which I highly appreciate.

I would like also to thank to:

Prof. Rudolf Ladenstein for his constant encouragement and support, for his inspiring enthusiasm and for maintaining truly academic and affable atmosphere in the group.

Xiaofeng Zhang for giving me practical advices on the usage of different software packages but mostly for being a friend, whose company I have been enjoying so often.

Dr. Petras Kundrotas for many stimulating discussions and his willingness to share his valuable knowledge and experience in various subjects.

other present and former members of the X-ray group: Dr. Jordi Benach, Dr. Winfried Meining, Dr. Bin Ren, Linda Arnfors, Dr. Heiko Bönisch, Gudrun Tibbelin, Mikael Karlström, as well as to Ana Caballero-Herrera, Dr. Nicolas Foloppe and Dr. Pekka Mark, colleagues from the MD group, for being kind and helpful in many ways.

Prof. Heinz Rüterjans from JWG University, Frankfurt, for the interesting discussions and for his marvellous hospitality

Normann Spitzner (JWG University, Frankfurt), Dr. Esben Friis (Novozymes, Denmark) and Ulrich Zachariae (Max-Planck-Institut, Martinsried), researchers with whom I had the pleasure to exchange knowledge and ideas in fruitful collaborations.

the personnel of the IT support unit and especially to Erik Lundgren for their efforts to keep the network and all computers properly running.

the administration staff for being always polite and of prompt assistance.

I am deeply grateful to my wife Silvia and my daughters Margaret and Katrin for their understanding, patience and love, for giving me their wholehearted support.

I revere my parents who gave me so much and who instilled in me the desire to explore and critically analyse everything in life or science.

I am also thankful to all Bulgarian friends who have not forgotten me during these four years.

This thesis was financially supported by the IV Biotechnology Program of the EC and the Department of Biosciences at NOVUM, Karolinska Institutet.

9 References

1. Tanford, C. (1970) Protein denaturation. *Adv Protein Chem* **25**: 1-95
2. Perutz, M.F. (1978) Electrostatic effects in proteins. *Science* **201**: 1187-1191
3. Warshel, A. (1981) Calculations of enzymatic reactions: calculations of pKa, proton transfer reactions and general acid catalysis in enzymes. *Biochemistry* **20**: 3167-3177
4. Warshel, A., Russell, S.T. (1984) Calculations of electrostatic interactions in biological systems and in solutions. *Q Rev Biophys* **17**: 283-422
5. Matthew, J.B. (1985) Electrostatic effects in proteins. *Annu Rev Biophys Biomol Struct* **14**: 387-417
6. Sharp, K., Fine, R., Schulten, K., Honig, B. (1987) Brownian dynamics simulation of diffusion to irregular bodies. *J Phys Chem* **91**: 3624-3631
7. Gilson, M., Honig, B. (1988) Calculation of the electrostatic energy of a macromolecular system: Solvation energies, binding energies and conformational analysis. *Proteins* **4**: 7-18
8. Warshel, A., Åqvist, J. (1991) Electrostatic energy and macromolecular function. *Annu Rev Biophys Biophys Chem* **20**: 267-298
9. Fersht, A. (1985) The pH dependance of enzyme catalysis. Enzyme structure and mechanism, 2 ed. W.H. Freeman & Co., New York 155-175
10. Szabo, A., Karplus, M. (1972) A mathematical model for structure-function relations in hemoglobin. *J Mol Biol* **72**: 163-172
11. Todt, J.C., Rocque, W.J., McGroarty, E.J. (1992) Effects of pH on bacterial porin function. *Biochemistry* **31**: 10471-10478
12. Todt, J.C., McGroarty, E.J. (1992) Involvement of histidine-21 in the pH-induced switch in porin channel size. *Biochemistry* **31**: 10479-10482
13. Bashford, D., Karplus, M. (1990) pK_a's of ionizable groups in proteins: Atomic detail from a continuum electrostatic model. *Biochemistry* **29**: 10219-10225
14. Bashford, D., Karplus, M. (1991) Multiple-site titration curves of proteins: An analysis of exact and approximate methods for their calculation. *J Phys Chem* **95**: 9556-9561
15. Bashford, D., Case, D.A., Dalvit, C., Tennant, L., Wright, P.E. (1993) Electrostatic calculations of side-chain pK_a values in myoglobin and comparison with NMR data for histidine. *Biochemistry* **32**: 8045-8056
16. Antosiewicz, J., McCammon, J.A., Gilson, M.K. (1994) Prediction of pH-dependent properties of proteins. *J Mol Biol* **238**: 415-436
17. Forsyth, W.R., Gilson, M.K., Antosiewicz, J., Jaren, O.R., Robertson, A.D. (1998) Theoretical and experimental analysis of ionization equilibria in ovomucoid third domain. *Biochemistry* **37**: 8643-8652
18. van Vlijmen, H.W.T., Schaefer, M., Karplus, M. (1998) Improving the accuracy of protein pK_a calculations - conformational averaging versus the average structure. *Proteins* **33**: 145-158

19. Warshel, A., Naray-Szabo, G., Sussman, F., Hwang, J.-K. (1989) How do serine proteases really work? *Biochemistry* **28**: 3629-3637

20. Raquet, X., Lounnas, V., Lamotte-Brasseur, J., Frère, J.M., Wade, R.C. (1997) pK_a calculations for class A α-lactamases: Methodological and mechanistic implications. *Biophys J* **73**: 2416-2426

21. Dillet, V., Dyson, H.J., Bashford, D. (1998) Calculations of electrostatic interactions and pK_as in the active site of *escherichia coli* thioredoxin. *Biochemistry* **37**: 10298-10306

22. Warshel, A. (1998) Electrostatic origin of the catalytic power of enzymes and the role of preorganized active sites. *J Biol Chem* **273**: 27035-27038

23. Lamotte-Brasseur, J., Lounnas, V., Raquet, X., Wade, R.C. (1999) pK_a calculations for class A β-lactamases: Influence of substrate binding. *Protein Sci* **8**: 404-409

24. Dillet, V., van Etten, R.L., Bashford, D. (2000) Stabilization of charges and protonation state in the active site of the protein tyrosine phosphatases: A computational study. *J Phys Chem B* **104**: 11321-11333

25. Asthagiri, D., Dillet, V., Liu, T.Q., Noodleman, L., Van Etten, R.L., Bashford, D. (2002) Density functional study of the mechanism of a tyrosine phosphatase: 1. Intermediate formation. *J. Am. Chem. Soc.* **124**: 10225-10235

26. Naray-Szabo, G. (2000) Enzyme mechanisms: interplay of theory and experiment. *Journal of Molecular Structure* **500**: 157-167

27. Warshel, A., Parson, W.W. (2001) Dynamics of biochemical and biophysical reactions: insight from computer simulations. *Q Rev Biophys* **34**: 563-679

28. Gunner, M.R., Honig, B. (1991) Electrostatic control of midpoint potentials in the cytochrome subunit of the *Rhodopseudomonas viridis* reaction center. *Proc Natl Acad Sci U S A* **88**: 9151-9155

29. Alexov, E.G., Gunner, M.R. (1999) Calculated protein and proton motion coupled to electron transfer: Electron transfer form $Q_A.Q_B$ to Q_B in bacterial photosynthetic reaction centers. *Biochemistry* **38**: 8253-8270

30. Baptista, A.M., Martel, P.J., Soares, C.M. (1999) Simulation of electron-proton coupling with a Monte Carlo method: Application to cytochrome c_3 using continuum electrostatics. *Biophys J* **76**: 2978-2998

31. Bashford, D., Gerwert, K. (1992) Electrostatic calculations of the pK_a values of ionizable groups in bacteriorhodopsin. *J Mol Biol* **224**: 473-486

32. Sampogna, R.V., Honig, B. (1994) Environmental effects on the states of active site residue in bacteriorhodopsin. *Biophys J* **66**: 1341-1352

33. Beroza, P., Fredkin, D.R., Okamura, M.Y., Feheer, G. (1995) Electrostatic calculations of acid titration and electron transfer, $Q^-_AQ_B$ - $Q_AQ^-_B$ in the reaction center. *Biophys J* **68**: 2233-2250

34. Sampogna, R.V., Honig, B. (1996) Electrostatic coupling between retinal Isomerization and ionization state of Glu-204: A general mechanism for proton release in bacteriorhodopsin. *Biophys J* **71**: 1165-1171

35. Beroza, P., Fredkin, M.Y., Okamura, M.Y., Feher, G. (1991) Protonation of interacting residues in a protein by a Monte Carlo method: Application to lysozyme and the photosynthetic reaction center of *Rhodobacter sphaeroides*. *Proc Natl Acad Sci U S A* **88**: 5804-5808

36. Sandberg, L., Edholm, O. (1997) pK_a calculations along bacteriothodopsin molecular dynamics trajectory. *Biophys Chem* **65**: 189-204

37. Kannt, A., Lancaster, C.R.D., Michel, H. (1998) The coupling of electron transfer and proton translocation - electrostatic calculations on paracoccus denitrificans cytochrome c oxidase. *Biophys J* **74**: 708-721

38. Spassov, V.Z., Luecke, H., Gerwert, K., Bashford, D. (2001) pK_a Calculations suggest storage of an excess proton in a hydrogen-bonded network in bacteriorhodopsin. *J Mol Biol* **321**: 203-219

39. Novothy, J., Sharp, K. (1992) Electrostatic fields in antibodies and antibody/antigen complexes. *Prog Biophys Mol Biol* **58**: 203-224

40. McDonald, S.M., Willson, R.C., McCammon, J.A. (1995) Determination of the pK_a values of titratable groups of an antigen-antibody complex, HyHEL-5-hen egg lysozyme. *Protein Eng* **8**: 915-995

41. Gibas, C.J., Subramaniam, S., McCammon, J.A., Braden, B.C., Poljak, R.J. (1997) pH dependence of antibody/lysozyme complexation. *Biochemistry* **36**: 15599-15614

42. Soriano, G.M., Cramer, W.A., Krishtalik, L.I. (1997) Electrostatic effects on electron-transfer kinetics in the cytochrom *f*-plastocyanin complex. *Biophys J* **73**: 3265-3276

43. Sheinerman, F.B., Norel, R., Honig, B. (2000) Electrostatic aspects of protein-protein interactions. *Curr. Opin. Struct. Biology* **10**: 153-159

44. Karshikoff, A., Spassov, V., Cowan, R.Q., Ladenstein, R., Schirmer, T. (1994) Electrostatic analysis of two porin channels from *E. coli*. *J Mol Biol* **240**: 372-384

45. Spassov, V.Z., Karshikoff, A.D., Ladenstein, R. (1994) The optimisation of the electrostatic interactions in proteins of different functional and folding type. *Protein Sci* **3**: 1556-1569

46. Spassov, V.Z., Karshikoff, A.D., Ladenstein, R. (1995) The optimization of protein solvent interactions. Thermostability and the role of hydrophobic and electrostatic interactions. *Protein Sci* **4**: 1516-1527

47. Schaefer, M., Sommer, M., Karplus, M. (1997) pH-dependence of protein stability: Absolute electrostattic free energy difference between conformations. *J Phys Chem B* **101**: 1663-1683

48. Xiao, L., Honig, B. (1999) Electrostatic contribution to the stability of hyperthermophilic proteins. *J Mol Biol* **289**: 1435-1444

49. Dong, F., Zhou, H.-X. (2002) Electrostatic contributions to T4 lysozyme stability: solvent-exposed charges versus semi-buried salt bridges. *Biophys J* **83**: 1341-1347

50. Kundrotas, P.J., Karshikoff, A. (2002) Modelling of denatured state for calculations of electrostatic contribution to protein stability. *Protein Sci* **11**: 1681-1686

51. Spassov, V.Z., Bashford, D. (1999) Multiple-site ligand binding to flexible macromolecules: Separation of global and local conformational change and an iterative mobile clustering approach. *J Comput Chem* **20**: 1091-1111

52. Onufriev, A., Bashford, D., Case, D.A. (2000) Modification of the generalized Born model suitable for macromolecules. *J Phys Chem B* **104**: 3712-3720

53. Arora, N., Bashford, D. (2001) Solvation energy density occlusion approximation for evaluation of desolvation penalties in biomolecular interactions. *Proteins* **43**: 12-27

54. Onufriev, A., Case, D.A., Bashford, D. (2002) Effective Born radii in the generalized Born approximation: The importance of being perfect. *J Comput Chem* **23**: 1297-1304

55. Gunner, M.R., Alexov, E. (2000) A pragmatic approach to structure based calculation of coupled proton and electron transfer in proteins. *Biochim Biophys Acta* **1458**: 63-87

56. Rocchia, W., Sridharan, S., Nicholls, A., Alexov, E., Chiabrera, A., Honig, B. (2002) Rapid grid-based construction of the molecular surface and the use of induced surface charge to calculate reaction field energies: Applications to the molecular systems and geometric objects. *J Comput Chem* **23**: 128-137

57. Rocchia, W., Alexov, E., Honig, B. (2001) Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. *J Phys Chem B* **105**: 6507-6514

58. Georgescu, R.E., Alexov, E., Gunner, M.R. (2002) Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins. *Biophys J* **83**: 1731-1748

59. Lide, D.R. (ed) (1999) CRC Handbook of Chemistry and Physics 80 ed. CRC Press

60. Cramer, W.A., Knaff, D.B. (1991) Energy transduction in biological membranes. Springer, New York

61. Nozaki, Y., Tanford, C. (1967) Examination of titration behaviour. *Methods Enzymol* **11**: 715-734

62. Stryer, L. (1988) Biochemistry 3ed. W.H. Freeman & Co., New Yourk

63. Wilson, G.S. (1983) Electrochemical studies of porphyrin redox reactions as cytochrome models. *Bioelectrochemistry and Bioenergetics* **1**: 172-179

64. Tanokura, M. (1983) 1H-NMR study on the tautomerism of the imidasole ring of histidine residues. 1. Microscopic pK values and molar ratios of tautomerism in histidine-containing peptides. *Biochim Biophys Acta* **742**: 576-585

65. Demchuk, E., Wade, R.C. (1996) Improving the continuum dielectric approach to calculating pK_as of ionizable groups in proteins. *J Phys Chem* **100**: 17373-17387

66. Kaim, W., Schwederski, B. (1995) Bioinorganic chemistry. Wiley, New York

67. Richardson, W.H., Peng, C., Bashford, D., Noodleman, L., Case, D.A. (1997) Incorporating solvation effects into density functional theory: calculations of absolute acidities. *Int J Quant Chem* **61**: 207-217

68. Linderstrøm-Lang, K. (1924) On the Ionisation of Proteins. *Comptes Rendus des Travaux du Laboratoire Carlsberg* **15**: 1-29

69. Kirkwood, J.G. (1934) Theory of solutions of molecules containing widely separated charges with special application to zwitterions. *J. Chem. Phys.* **2**: 351-361

70. Tanford, C., Kirkwood, J.G. (1957) Theory of titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* **79**: 5333-5339

71. Kirkwood, J.G., Westheimer, F.H. (1938) The electrostatic influence of substituents on the dissociation constants of organic acids. I. *J. Chem. Phys.* **6**: 506-512

72. Westheimer, F.H., Kirkwood, J.G. (1938) The electrostatic influence of substituents on the dissociation constants of organic acids. II. *J. Chem. Phys.* **6**: 513-517

73. Shire, S.J., Hanania, G.I.H., Gurd, F.R.N. (1974) Electrostatic effects in Myoglobin. Hydrogen ion equilibria in sperm whale ferrimyoglobin. *Biochemistry* **13**: 2967-2974

74. Matthew, J.B., Gurd, R.F.N. (1986) Calculation of electrostatic interactions in proteins. In: Hirs GHW, Timasheff SN (eds) Methods in enzymology. Enzyme structure, part K. Academic Press Inc., pp 413-453

75. Edsall, J.T., Martin, R.B., Hollingworth, B.R. (1958) Ionization of individual groups in dibasic acids, with application to the amino and hydroxyl groups of tyrosine. *Proc Natl Acad Sci U S A* **44**: 505-519

76. Cowgill, R.W. (1965) Fluorescence and the structure of proteins. V. Ionization of Tyrosyl residues. *Biochim Biophys Acta* **94**: 81–88.

77. Bradbury, J.H., Scheraga, H.A. (1966) Structural studies of ribonuclease. XXIV. The application of nuclear magnetic resonance spectroscopy to distinguish between the histidine residues of ribonuclease. *J. Am. Chem. Soc.* **88**: 4240-4246

78. Meadows, D.H., Jardetzky, O., Epand, R.M., Rüterjans, H.H., Scheraga, H.A. (1968) Assignment of the histidine peaks in the nuclear magnetic resonance spectrum of ribonuclease. *Proc Natl Acad Sci U S A* **60**: 766-772

79. Timasheff, S.N., Rupley, J.A. (1972) Infrared titration of lysozyme carboxyls. *Arch Biochem Biophys* **150**: 318–323.

80. Parsons, S.M., Raftery, M.A. (1972) Ionization behaviour of the catalytic carboxyls of lysozyme. Effects of ionic strength. *Biochemistry* **11**: 1623 –1629.

81. Parsons, S.M., Raftery, M.A. (1972) Ionization behaviour of the catalytic carboxyls of lysozyme. Effects of temperature. *Biochemistry* **11**: 1630 –1633.

82. Parsons, S.M., Raftery, M.A. (1972) Ionization behaviour of the cleft carboxyls in lysozyme-substrate complexes. *Biochemistry* **11**: 1633–1638.

83. Fersht, A. (1972) Conformational equilibria in α- and β-chymotrypsin. The energetics and importance of the salt bridge. *J Mol Biol* **64**: 497-509

84. Sternberg, M.J.E., Hayes, F.R.N., Russell, A.J., Thomas, P.G., Fersht, A.R. (1987) Prediction of electrostatic effects of engineering of protein charges. *Nature* **330**: 86-88

85. Russell, A.J., Thomas, P.G., Fersht, A.R. (1987) Electrostatic effects on modification of charged groups in the active site cleft of subtilisin by protein engineering. *J Mol Biol* **193**: 803-813

86. Oda, Y., Yamazaki, T., Nagayama, K., Kanaya, S., Kuroda, H., Nakamura, H. (1994) Individual ionization constants of all the carboxyl groups in ribonuclease HI from Escherichia coli determined by NMR. *Biochemistry* **33**: 5275-5284

87. Yamazaki, T., Nicholson, L.K., Torchia, D.A., Wingfield, P., Stahl, S.J., Kaufman, J.D., Eyermann, C.J., Hodge, C.N., Lam, P.Y.S., Ru, Y., Jadhav, P.K., Chang, C.H., Weber, P.C. (1994) NMR and X-ray evidence that the HIV protease catalytic aspartyl groups are protonated in the complex formed by the protease and a non-peptide cyclic urea-based inhibitor. *J. Am. Chem. Soc.* **116**: 10791-10792

88. Tomasi, J., Persico, M. (1994) Molecular interactions in solution: an overview of methods based on continuous distributions of the solvent. *Chem. Rev.* **94**: 2027-2094

89. Beroza, P., Case, D.A. (1998) Calculations of proton-binding thermodynamics in proteins. *Methods Enzymol* **295**: 170-189

90. Warshel, A., Papazyan, A. (1998) Electrostatic effects in macromolecules - fundamental concepts and practical modeling. *Curr. Opin. Struct. Biology* **8**: 211-217

91. Ullmann, G.M., Knapp, E.W. (1999) Electrostatic models for computing protonation and redox equilibria in proteins. *Eur Biophys J* **28**: 533-551

92. Náray-Szabó, G. (2000) Enzyme mechanisms: interplay of theory and experiment. *Theochem-Journal of Molecular Structure* **500**: 157-167

93. Bashford, D., Case, D.A. (2000) Generalized born models of macromolecular solvation effects. *Annu Rev Phys Chem* **51**: 129-152

94. Lim, C., Bashford, D., Karplus, M. (1991) Absolute pK_a calculations with continuum dielectric methods. *J Phys Chem* **95**: 5610-5620

95. Chen, J.L., Noodleman, L., Case, D., Basshford, D. (1994) Incorporating solvation effects into density functional electronic structure calculations. *J Phys Chem* **98**: 11059-11068

96. Li, J., Nelson, M.R., Peng, C.Y., Basshford, D., Noodleman, L. (1998) Incorporating protein environments in density functional theory: a self-consistent reaction field calculation of redox potentials of [2Fe2S] cluster in ferreoxin and pathalate dioxygenase reductase. *J Phys Chem A* **102**: 5610-5620

97. Konecny, R., Li, J., Fisher, C.L., Dillet, V., Bashford, D., Noodleman, L. (1999) CuZn superoxide dismutase geometry optimization, energetics, and redox potential calculations by density functional and electrostatic methods. *Inorg Chem* **38**: 940-950

98. Peräkylä, M. (1998) A model study of the enzyme-catalyzed cytosine methylation using ab Initio quantum mechanical and density functional theory calculations: pK_a of the cytosine N3 in the Intermediates and Transition States of the Reaction. *J. Am. Chem. Soc.* **120**: 12895-12902

99. Stanton, R.V., Perakyla, M., Bakowies, D., Kollman, P.A. (1998) Combined *ab initio* and free energy calculations to study reactions in enzymes and solution: Amide hydrolysis in trypsin and aqueous solution. *J. Am. Chem. Soc.* **120**: 3448-3457

100. Blomberg, M.R.A., Siegbahn, P.E.M. (2001) A quantum chemical approach to the study of reaction mechanisms of redox-active metalloenzymes. *J Phys Chem B* **105**: 9375-9386

101. Friesner, R.A., Beachy, M.D. (1998) Quantum mechanical calculations on biological systems. *Curr. Opin. Struct. Biology* **8**: 257-262

102. Bruice, T.C., Kahn, K. (2000) Computational enzymology. *Curr Opin Chem Biol* **4**: 540-544

103. Reynolds, C.A., King, P.M., Richards, W.Q. (1992) Free-energy calculations in molecular biophysics. *Molecular Physics* **76**: 251-275

104. Kollman, P. (1993) Free-energy calculations - applications to chemical and biochemical phenomena. *Chem. Rev.* **93**: 2395-2417

105. Zhang, L.Y., Gallicchio, E., Friesner, R.A., Levy, R.M. (2001) Solvent models for protein-ligand binding: Comparison of implicit solvent Poisson and surface generalized born models with explicit solvent simulations. *J Comput Chem* **22**: 591-607

106. Lu, N.D., Kofke, D.A. (2001) Accuracy of free-energy perturbation calculations in molecular simulation. I. Modeling. *J. Chem. Phys.* **114**: 7303-7311

107. Lu, N.D., Kofke, D.A. (2001) Accuracy of free-energy perturbation calculations in molecular simulation. II. Heuristics. *J. Chem. Phys.* **115**: 6866-6875

108. Boresch, S., Archontis, G., Karplus, M. (1994) Free energy simulations: The meaning of the individual contributions from a component analysis. *Proteins* **20**: 25-33

109. Delbuono, G.S., Figueirido, F.E., Levy, R.M. (1994) Intrinsic pK_as of ionizable residues in proteins - An explicit solvent calculation for lysozyme. *Proteins* **20**: 85-97

110. Åqvist, J. (1991) Free-energy perturbation study of metal-ion catalyzed proton transfer in water. *J Phys Chem* **95**: 4587-4590

111. Merz, K.M. (1991) Determination of pK_as of ionizable groups in proteins - the pK_a of Glu-7 and Glu-35 in hen egg-white lysozyme and Glu-106 in human carbonic anhydrase-II. *J. Am. Chem. Soc.* **113**: 3572-3575

112. Price, D.J., Jorgensen, W.L. (2001) Improved convergence of binding affinities with free energy perturbation: Application to nonpeptide ligands with pp60(src) SH2 domain. *J Comput Aided Mol Des* **15**: 681-695

113. Reddy, M.R., Erion, M.D. (2001) Calculation of relative binding free energy differences for fructose 1,6-bisphosphatase inhibitors using the thermodynamic cycle perturbation approach. *J. Am. Chem. Soc.* **123**: 6246-6252

114. Helms, V., Wade, R.C. (1998) Computational alchemy to calculate absolute protein-ligand binding free energy. *J. Am. Chem. Soc.* **120**: 2710-2713

115. Yamaotsu, N., Moriguchi, I., Hirono, S. (1993) Estimation of stabilities of staphylococcal nuclease mutants (met(32)-]ala and met(32)-]leu) using molecular-dynamics free-energy perturbation. *Biochim Biophys Acta* **1203**: 243-250

116. Florian, J., Goodman, M.F., Warshel, A. (2000) Free-energy perturbation calculations of DNA destabilization by base substitutions: The effect of neutral guanine thymine, adenine cytosine and adenine difluorotoluene mismatches. *J Phys Chem B* **104**: 10092-10099

117. Warshel, A., Russell, S.T., Churg, A.K. (1984) Macroscopic models for studies of electrostatic interactions in proteins: Limitations and applicability. *Proc Natl Acad Sci U S A* **81**: 4785-4789

118. Florián, J., Warshel, A. (1997) Langevin dipoles model for ab initio calculations of chemical processes in solution: Parametrization and application to hydration free energy of neutral and ionic solutes and conformational analysis in aqueous solution. *J Phys Chem B* **101**: 5583-5595

119. Sharp, K.A., Honig, B. (1990) Calculating total electrostatic energies with the nonlinear Poisson-Boltzmann equation. *J Phys Chem* **94**: 7684-7692

120. Oberoi, H., Allewell, N.M. (1993) Multigrid solution of the nonlinear Poisson-Boltzmann equation and calculation of titration curves. *Biophys J* **65**: 48-55

121. Holst, M., Saied, F. (1995) Numerical solution of the nonlinear Poisson-Boltzmann equation: Developing more robust and efficient methods. *J Comput Chem* **16**: 337-364

122. Forsten, K.E., Kozack, R.E., Lauffenburger, D.A., Subramaniam, S. (1994) Numerical solution of the nonlinear Poisson-Boltzmann equation for a membrane electrolyte system. *J Phys Chem* **98**: 5580-5586

123. Chen, S.W., Honig, B. (1997) Monovalent and divalent salt effect on electrostatic free energies defined by the nonlinear Poisson-Boltzmann equation: Application to DNA binding reaction. *J Phys Chem B* **101**: 9113-9118

124. Tanford, C. (1961) Physical chemistry of macromolecules. John Wiley & Sons, NY

125. Zhou, H.-X. (1994) Macromolecular electrostatic energy within the nonlinear Poisson-Boltzmann equation. *J. Chem. Phys.* **100**: 3152-3162

126. Warwicker, J., Watson, N.C. (1982) Calculation of the electric field potential in the active site cleft due to alpha-helix dipoles. *J Mol Biol* **157**: 671-679

127. Klapper, I., Hagstrom, R., Fine, R., Sharp, K., Honig, B. (1986) Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: Effects of ionic strength and amino-acid modification. *Proteins* **1**: 47-59

128. Gilson, M., Sharp, K.A., Honig, B.H. (1987) Calculating electrostatic potential of molecules in solution: Method and error assessment. *J Comput Chem* **9**: 327-335

129. Nicholls, A., Honig, B. (1991) A rapide finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J Comput Chem* **12**: 435-445

130. Honig, B., Sharp, K., Yang, A.-S. (1993) Macroscopic models of aqueous solution: Biological and chemical application. *J Phys Chem* **97**: 1101-1109

131. Honig, B., Nicholls, A. (1995) Clasical electrostatics in biology and chemistry. *Science* **268**: 1144-1149

132. Beroza, P., Fredkin, D.R. (1996) Calculation of amino acid pK_as in a protein from a continuum electrostatic model: Method and sensitivity analysis. *J Comput Chem* **17**: 1229-1244

133. Hadjidimos, A. (2000) Successive overrelaxation (SOR) and related methods. *J Comp Appl Math* **123**: 177-199

134. Mohan, V., Davis, M.E., Mccammon, J.A., Pettitt, B.M. (1992) Continuum model-calculations of solvation free-energies - accurate evaluation of electrostatic contributions. *J Phys Chem* **96**: 6428-6431

135. Zhou, H.-X. (1993) Boundary element solution of macromolecular electrostatics: interaction energy between two proteins. *Biophys J* **65**: 955-963

136. Purisima, E.O., Nilar, S.H. (1995) A simple yet accurate boundary element method for continuum dielectric calculations. *J Comput Chem* **16**: 681-689

137. Bharadeaj, R., Windemuth, A., Sridharan, S., Honig, B., Hichols, A. (1995) The fast multipole boundary element method for molecular electrostatics: An optimal approach for large systems. *J Comput Chem* **16**: 898-913

138. Zauhar, R.J., Varnek, A. (1996) A fast and space-efficient boundary elements method for computing electrostatic hydration effects in large molecule. *J Comput Chem* **7**: 864-877

139. Juffer, A.H., Argos, P., Vogel, H.J. (1997) Calculation acid-dissociation constant of proteins using boundary elements method. *J Phys Chem B* **101**: 7664-7673

140. Laing, J., Subramaniam, S. (1997) Computation of molecular electrostatics with boundary elements methods. *Biophys J* **73**: 1830-1841

141. You, T.J., Harvey, S.C. (1993) Finite element approach to the electrostatics of macromolecules with arbitrary geometries. *J Comput Chem* **14**: 484-501

142. Hollerbach, U., Eisenberg, R.S. (2002) Concentration-dependent shielding of electrostatic Potentials inside the gramicidin A channels. *Langmuir* **18**: 3626-3631

143. Hollerbach, U., Chen, D.P., Busath, D.D., Eisenberg, B. (2000) Predicting function from structure using the Poisson-Nernst-Planck equations: Sodium current in the gramicidin A channel. *Langmuir* **16**: 5509-5514

144. Holst, M., Kozack, R.E., Saied, F., Subramaniam, S. (1994) Treatment of electrostatic effects in proteins: Multigrid-based Newton iterative method for solution of the full nonlinear Poisson-Boltzmann equation. *Proteins* **18**: 231-245

145. Vorobjev, Y.N., Scheraga, H.A. (1997) A fast adaptive multigrid boundary element method for macromolecular electrostatic computations in a solvent. *J Comput Chem* **18**: 569-583

146. Davis, M.E., McCammon, J.A. (1989) Solving the finite-difference linearized Poisson-Boltzmann equation - a comparison of relaxation and conjugate-gradient methods. *J Comput Chem* **10**: 386-391

147. Nicholls, A., Sharp, K.A., Honig, B. (1990) DelPhi, 3.0 ed. Department of Biochemistry and Molecular Biophysics, Columbia University, New York

148. Bashford, D. (1997) An Object-Oriented Programming Suite for Electrostatic Effects in Biological Molecules. Springer, Berlin

149. Cortis, C.M., Friesner, R.A. (1997) An automatic three-dimensional finite element mesh generation system for the Poisson-Boltzmann equation. *J Comput Chem* **18**: 1570-1590

150. Cortis, C.M., Friesner, R.A. (1997) Numerical solution of the Poisson-Boltzmann equation using tetrahedral finite-element meshes. *J Comput Chem* **18**: 1591-1608

151. Holst, M., Baker, N., Wang, F. (2000) Adaptive multilevel finite element solution of the Poisson-Boltzmann equation I. Algorithms and examples. *J Comput Chem* **21**: 1319-1342

152. Baker, N., Holst, M., Wang, F. (2000) Adaptive multilevel finite element solution of the Poisson-Boltzmann equation II. Refinement at solvent-accessible surfaces in biomolecular systems. *J Comput Chem* **21**: 1343-1352

153. Karshikoff, A. (1995) A simple algorithm for calculation of multiple site titration curves. *Protein Eng* **8**: 243-248

154. Miteva, M., Demirev, P.A., Karshikoff, A.D. (1997) Multiply-protonated protein ions in the gas phase: Calculation of the electrostatic interactions between charged sites. *J Phys Chem B* **101**: 9645-9650

155. Lamm, G., Pack, G.R. (1997) Calculation of dielectric constant near polyelectrolyltes in solution. *J Phys Chem B* **101**: 959-965

156. Harvey, S.C. (1989) Treatment of electrostatic effects in macromolecular modelling. *Proteins* **5**: 78-92

157. Gilson, M.K. (1995) Theory of electrostatic interactions in macromolecules. *Curr. Opin. Struct. Biology* **5**: 216-223

158. Gilson, M., Honig, B. (1986) The dielectric constant of a folded protein. *Biopolymers* **25**: 2097-2191

159. Froloff, N., Windemuth, A., Honig, B. (1997) On the calculation of binding free energies using continuum methods: Application to MHC class I protein-peptide interactions. *Protein Sci* **6**: 1293-1301

160. Luo, R., David, L., Hung, H., Devaney, J., Gilson, M.K. (1999) Strength of solvent-exposed salt-bridges. *J Phys Chem B* **103**: 727-736

161. Dwyer, J.J., Gittis, A.G., Karp, D.A., Lattman, E.E., Spencer, D.S., Stites, W.E., Garcia-Moreno, B. (2000) High apparent dielectric constants in the interior of a protein reflect water penetration. *Biophys J* **79**: 1610-1620

162. Schwehm, J.M., Fitch, C.A., Dang, B.N., Garcia-Moreno, B., Stites, W.E. (2003) Changes in stability upon charge reversal and neutralization substitution in staphylococcal nuclease are dominated by favorable electrostatic effect. *Biochemistry* **42**: 1118-1128

163. Antosiewicz, J., Briggs, J.M., Elcock, A.H., Gilson, M.K., McCammon, J.A. (1996) Computing ionization states of proteins with a detailed charge model. *J Comput Chem* **17**: 1633-1644

164. You, T.J., Bashford, D. (1995) Conformation and hydrogen ion titration of proteins: A continuum electrostatic model with conformational flexibility. *Biophys J* **69**: 1721-1733

165. Alexov, E.G., Gunner, M.R. (1999) Calculated protein and protonmotion coupled to electron transfer: Electron transfer form Q_A-Q_B to Q_B in bacterial photosyntetic reaction centers. *Biochemistry* **38**: 8253-8270

166. Nielsen, J.E., Andersen, K.V., Honig, B., Hooft, R.W.W., Klebe, G., Vriend, G., Wade, R.C. (1999) Improving macromolecular electrostatics calculations. *Protein Eng* **12**: 657-662

167. Simonson, T., Perahia, D. (1995) Internal and interfatial dielectric properties of cytochrome *c* from molecular dynamics in aquous solution. *Proc Natl Acad Sci U S A* **92**: 1082-1086

168. Simonson, T., Perahia, D. (1995) Microscopic dielectric properties of cytochrome *c* from molecular dynamics simulations in aqueous solution. *J. Am. Chem. Soc.* **117**: 7987-8000

169. Simonson, T., Brooks III, C.L. (1996) Charge screening and the dielectric constant of proteins: Insights form molecular dynamics. *J. Am. Chem. Soc.* **118**: 8452-8458

170. Simonson, T. (1999) Dielectric relaxation in proteins: Microscopic and macroscopic models. *Int J Quant Chem* **75**: 331-331

171. King, G., Lee, F.S., Warshel, A. (1991) Microscopic simulations of macroscopic dielectric constant of solvated proteins. *J. Chem. Phys.* **95**: 4366-4377

172. Sharp, K., Jean-Charles, A., Honig, B. (1992) A local dielectric constant for model solvation free energies which accounts for solute polarizability. *J Phys Chem* **96**: 3822-3828

173. Voges, D., Karshikoff, A. (1998) A model for a local static dielectric constant in macromolecules. *J Phys Chem* **108**: 2219-2227

174. Mazur, J., Jernigan, R.L. (1991) Distance-dependent dielectric constants and their application to double-helical DNA. *Biopolymers* **31**: 1615-1629

175. Sandberg, L., Edholm, O. (1999) A fast and simple method to calculate protonation states in proteins. *Proteins* **36**: 474-483

176. Mehler, E.L., Warshel, A. (2000) Comment on "A fast and simple method to calculate protonation states in proteins". *Proteins* **40**: 1-3

177. Mehler, E.L. (1996) Self-consistent, free energy based approximation to calculate pH dependent electrostatic effect in proteins. *J Phys Chem* **100**: 16006-16018

178. Hassan, S.A., Guarnieri, F., Mehler, E.L. (2000) A general treatment of solvent effects based on screened Coulomb potentials. *J Phys Chem B* **104**: 6478-6489

179. Hassan, S.A., Mehler, E.L. (2002) A critical analysis of continuum electrostatics: The screened Coulomb potential-implicit solvent model and the study of the alanine dipeptide and discrimination of misfolded structures of proteins. *Proteins* **47**: 45-61

180. Mallik, B., Masunov, A., Lazaridis, T. (2002) Distance and exposure dependent effective dielectric function. *J Comput Chem* **23**: 1090-1099

181. Still, W.C., Tempczyk, A., Hawley, R.C., Hendrickson, T. (1990) Semianalitical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* **112**: 6127-6129

182. Hawkins, G.D., Cramer, C.J., Truhlar, D.G. (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreeneing of solute atomic charges from a dielectric medium. *J Phys Chem* **100**: 19824-19839

183. Jayaram, B., Liu, Y., Beveridge, D.L. (1998) A modification of the generalized Born theory for improved of solvation energies and pK shifts. *J. Chem. Phys.* **109**: 1465-1471

184. Srinivasan, J., Trevathan, M.W., Beroza, P., Case, D.A. (1999) Application of a pairwise generalized Born model to proteins and nucleic acids: inclusion of salt effects. *Theor Chem Acc* **101**: 426-434

185. Botelho, L., Friend, S., Matthew, J., Lehman, L., Hanania, G., Gurd, R. (1978) Proton nuclear magnetic resonance study of histidine ionizations in mioglobins of various species. Comparison of observed and computed pK values. *Biochemistry* **17**: 5197-5204

186. Beroza, P., Case, D.A. (1996) Including side chain flexibility in continuum electrosttatic calculations of protein titration. *J Phys Chem* **100**: 20156-20163

187. Alexov, E., Gunner, M.R. (1997) Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. *Biophys J* **72**: 2075-2093

188. Baptista, A.M., Soares, C.M. (2001) Some theoretical and computational aspects of the Inclusion of proton isomerism in the protonation equilibrium of proteins. *J Phys Chem B* **105**: 293-309

189. Giletto, A., Pace, C.N. (1999) Buried, charged, non-ion-paired aspartic acid 76 contributes favorably to the conformational stability of ribonuclease T_1. *Biochemistry* **38**: 13379-13384

190. Pfeiffer, S., Spitzner, N., Löhr, F., Rüterjans, H. (1998) Hydration water molecules of nucleotide-free Ribonuclease T1 studied by NMR spectroscopy in solution. *J Biomol NMR* **11**: 1-15

191. Martinez-Oyanedel, J., Choe, H.W., Heinemann, U., Saenger, W. (1991) Ribonuclease T1 with free recognition and catalytic site: Crystal structure analysis at 1.5 Å resolution. *J Mol Biol* **222**: 335-352

192. Brendskag, M.K., McKinley-McKee, J.S., Winberg, J.O. (1999) *Drosophila lebanonensis* alcohol dehydrogenase: pH dependence of the kinetic coefficients. *Biochim Biophys Acta* **1431**: 74-86

193. Winberg, J.O., Brendskag, M.K., Sylte, I., Lindstad, R.I., McKinley-McKee, J.S. (1999) The catalytic triad in *Drosophila* alcohol dehydrogenase: pH, temperature and molecular modelling studies. *J Mol Biol* **294**: 601-616

194. Benach, J., Atrian, S., Gonzalez-Duarte, R., Ladenstein, R. (1999) The Catalytic Reaction and Inhibition Mechanism of Drosophila Alcohol Dehydrogenase: Observation of an Enzyme-bound NAD-ketone Adduct at 1.4 Å Resolution by X-ray Crystallography. *J Mol Biol* **289**: 335-355

195. Benach, J., Atrian, S., Gonzalez-Duarte, R., Ladenstein, R. (1998) The Refined Crystal Structure of Drosophila lebanonensis Alcohol Dehydrogenase at 1.9 Å Resolution. *J Mol Biol* **282**: 383-399

196. Tanford, C., Roxby, R. (1972) Interpretation of protein titration curves. Application to Lysozyme. *Biochemistry* **11**: 2192-2198

197. Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. (1953) Equation of state calculations by fast computing machines. *J. Chem. Phys.* **21**: 1087-1092

198. Yang, A.-S., Gunner, M.R., Sampogna, R., Sharp, K., Honig, B. (1993) On the calculation of pK_as in proteins. *Proteins* **15**: 252-265

199. Gilson, M. (1993) Multiple-site titration and molecular modelling: Two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins* **15**: 266-282

200. Bartik, K., Redfield, C., Dobson, C.M. (1994) Measurement of the individual pK_a values of acidic residues of hen and turkey lysozymes by two-dimensional ^1H NMR. *Biophys J* **66**: 1180-1184

201. Onufriev, A., Case, D., Ullmann, G.M. (2001) A novel view of pH titration in biomolecules. *Biochemistry* **30**: 3413-3419

202. Alexov, E. (2003) Role of the protein side-chain fluctuations on the strength of pair-wise electrostatic interactions: comparing experimental with computed pK_as. *Proteins* **50**: 94-103

203. Nielsen, J.E., Vriend, V. (2001) Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pKa calculations. *Proteins* **43**: 403-412

204. Wlodek, S.T., Antosiewicz, J., McCammon, J.A. (1997) Prediction of titration properties of structures of a protein derived from molecular dynamics trajectories. *Protein Sci* **6**: 373-382

205. Nielsen, J.E., McCammon, J.A. (2003) On the evaluation and optimization of protein X-ray structures for pKa calculations. *Protein Sci* **12**: 313-326

206. Antosiewicz, J., McCammon, J.A., Gilson, M.K. (1996) The determination of pK_as in proteins. *Biochemistry* **35**: 7819-7833

207. Khare, D., Alexander, P., Antosiewich, J., Bryan, P., Gilson, M., Orban, J. (1997) pK_a measurements from nuclear magnetic resonance for B1 and B2 immunoglobin G-binding domain of protein G: comparison with calculated values for nuclear magnetic resonance and X-ray structures. *Biochemistry* **36**: 3580-3589

208. Rabenstein, B., Ullmann, G.M., Knapp, E.W. (1998) Calculation of protonation patterns in proteins with structural relaxation and molecular ensembles - application to the photosynthetic reaction center. *Eur Biophys J* **27**: 626-637

209. Yang, A.-S., Honig, B. (1993) On the pH dependence of protein stability. *J Mol Biol* **231**: 459-474

210. Yang, A.-S., Honig, B. (1994) Structural origin of pH and ionic strength effects on protein stability. Acid denaturation of sperm whale apomyoglobin. *J Mol Biol* **237**: 602-614

211. Gorfe, A.A., Ferrara, P., Caflisch, A., Marti, D.N., Bosshard, H.R., Jelesarov, I. (2002) Calculation of protein ionization equilibria with conformational sampling pK_a of a model leucine zipper, GCN4 and barnase. *Proteins* **46**: 41-60

212. Ripoll, D.R., Vorobjev, Y.N., Liwo, A., Vila, J.A., Scheraga, H.A. (1996) Coupling between folding and ionization equilibria: effects of pH on the conformational preferences of polypeptides. *J Mol Biol* **264**: 770-783

213. Vila, J.A., Ripoll, D.R., Vorobjev, Y.N., Scheraga, H.A. (1998) Computation of the structure-dependent pK_a shifts in a polypentapeptide of the poly[fv(IPGVG), fE(IPGEG)] family. *J Phys Chem B* **102**: 3065-33067

214. Baptista, A.M., Martel, P.J., Petersen, S.B. (1997) Simulation of protein conformational freedom as a function of pH: Constant-pH molecular dynamics using implicit titration. *Proteins* **27**: 523-544

215. Baptista, A.M., Teixeira, V.H., Soares, C.M. (2002) Constant-pH molecular dynamics using stochastic titration. *J. Chem. Phys.* **117**: 4184-4200

216. Hille, B. (1992) Ionic channels of excitable membranes, 2-nd ed. Sinauer Associates Inc., Sunderland, MA

217. Eisenberg, R.S. (1996) Cumputing the the field in proteins and membranes. *J Membr Biol* **150**: 1-25

218. Nonner, W., Chen, D.P., Eisenberg, B. (1999) Progress and prospects in permeation. *J Gen Physiol* **113**: 773-782

219. Tieleman, D.P., Biggin, P.C., Smith, G.R., Sansom, M.S.P. (2001) Simulation approaches to ion channel structure-function relationships. *Q Rev Biophys* **34**: 473-561

220. Roux, B. (2002) Theoretical and computational models of ion channels. *Curr. Opin. Struct. Biology* **12**: 182-189

221. Mathes, A., Engelhardt, H. (1998) Nonlinear and asymmetric open channel characteristics of an ion-selective porin in planar membranes. *Biophys J* **75**: 1255-1262

222. Levitt, D.G. (1999) Modeling of ion channeis. *J Gen Physiol* **113**: 789-794

223. Graf, P., Nitzan, A., Kurnikova, M.G., Coalson, R.D. (2000) A dynamic lattice Monte Carlo model of ion transport in inhomogeneous dielectric environments: Method and implementation. *J Phys Chem B* **104**: 12324-12338

224. Corry, B., Kuyucak, S., Chung, S.H. (2000) Tests of continuum theories as models of ion channels. II. Poisson-Nernst-Planck theory versus Brownian dynamics. *Biophys J* **78**: 2364-2381

225. Corry, B., Allen, T.W., Kuyucak, S., Chung, S.H. (2001) Mechanisms of permeation and selectivity in calcium channels. *Biophys J* **80**: 195-214

226. Schuss, Z., Nadler, B., Eisenberg, R.S. (2001) Derivation of Poisson and Nernst-Planck equations in a bath and channel from a molecular model. *Phys Rev E* **64**: art no 036116.

227. Kurnikova, M.G., Coalson, R.D., Graf, P., Nitzan, A. (1999) A lattice relaxation for three-dimensional Poisson-Nernst-Planck theory with application to ion transport through the gramicidin A channel. *Biophys J* **76**: 642-656

228. Cardenas, A.E., Coalson, R.D., Kurnikova, M.G. (2000) Three-dimensional Poisson-Nernst-Planck theory studies: Influence of membrane electrostatics on gramicidin A channel conductance. *Biophys J* **79**: 80-93

229. Nonner, W., Eisenberg, B. (1998) Ion permeation and glutamate residues linked by Poisson-Nernst-Planck theory in L-type calcium channels. *Biophys J* **75**: 1287-1305

230. Chen, D.P., Xu, L., Tripathy, A., Meissner, G., Eisenberg, B. (1999) Selectivity and permeation in calcium release channel of cardiac muscle: Alkali metal ions. *Biophys J* **76**: 1346-1366

231. Riveros, O., Croxton, T., Armstrong, W.M.D. (1989) Liquid junction potentials calculated from numerical solutions of the Nernst-Planck and Poisson equations. *J Theor Biol* **140**: 221-230