Biophys Rev (2017) 9:699–728
DOI 10.1007/s12551-017-0311-5

REVIEW

Development of constant-pH simulation methods in implicit solvent and applications in biomolecular systems

Fernando Luís Barroso da Silva[1,2,3] · Luis Gustavo Dias[4]

Received: 13 June 2017 / Accepted: 1 August 2017 / Published online: 18 September 2017
© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag GmbH Germany 2017

Abstract pH is a critical parameter for biological and technological systems directly related with electrical charges. It can give rise to peculiar electrostatic phenomena, which also makes them more challenging. Due to the quantum nature of the process, involving the forming and breaking of chemical bonds, quantum methods should ideally by employed. Nevertheless, due to the very large number of ionizable sites, different macromolecular conformations, salt conditions, and all other charged species, the CPU time cost simply becomes prohibitive for computer simulations, making this a quite complex problem. Simplified methods based on Monte Carlo sampling have been devised and will be reviewed here, highlighting the updated state-of-the-art of this field, advantages, and limitations of different theoretical protocols for biomolecular systems (proteins and nucleic acids). Following a historical perspective, the discussion will be associated with the applications to protein interactions with other proteins, polyelectrolytes, and nanoparticles.

Keywords Protein titration · RNA titration · Monte Carlo simulations · Tanford and Kirkwood model · Electrostatics interactions · pH effects

This article is part of a Special Issue on 'Latin America' edited by Pietro Ciancaglini and Rosangela Itri.

✉ Fernando Luís Barroso da Silva
flbarroso@usp.br; flbarros@ncsu.edu

1 Departamento de Física e Química, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Av. do café, s/no. – Universidade de São Paulo, BR-14040-903 Ribeirão Preto, SP, Brazil

2 UCD School of Physics, UCD Institute for Discovery, University College Dublin, Belfield, Dublin 4, Ireland

3 Department of Chemical and Biomolecular Engineering, North Carolina State University, Raleigh, NC, USA

4 Departamento de Química, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, Av. Bandeirantes, 3900 – Universidade de São Paulo, BR-14040-901 Ribeirão Preto SP, Brazil

Introduction

pH is a critical physical chemical parameter for biological, chemical, medical and technological systems. Being by a stoichiometric definition a measurement of the concentration of hydrogen ions ($[H^+]$) in the aqueous solution ($pH = -\log[H^+]$), it is intuitive to make a direct relation between H^+, an electrical charge and electrostatic interactions (Bell 1959). In fact, pH quantifies the availability of protons to go from the aqueous solution to titratable sites on the macromolecule (when there are H^+ available in the solution, i.e., at lower pH, the acid regime) or from the macromolecule to the solution (when the solution is lacking H^+, i.e., at higher pH, the basic regime). This does indicate that pH can control amino acids and nucleotides charges and as such the electrostatic interactions *in* and *between* biomolecules governing their conformation, stability, solubility, association, and function. Several works have given special emphasis to these features. For example, the Nobel laureate Perutz has started one of his classical papers (Perutz 1978) saying that *"electrostatic effects dominate many aspects of protein behavior"* exemplifying *"the decisive influence of electrostatic effects on the structure, assembly, and hydration of proteins, and on the catalytic power of enzyme"* (Perutz 1978).

Being well known that the magnitude of the electrostatics contributions depends on the charges that are given by pH, it becomes logical to relate them to the biological function. In reality, it is often illustrated in biochemical textbooks how the enzyme activity is controlled by pH (Garrett and Grisham 1999; Creighton 1983; Devlin 1997) and can affect clinical conditions (Piper and Fenton 1965). There is evidence that pH works as a signaling mechanism to regulate a number of cell processes (Schönichen et al. 2013; Jin et al. 2017). For instance, a small increase of 0.1 units in the intracellular pH can promote cell proliferation and cell cycle progression while a decrease can contribute to apoptosis. Some diseases like cancers and neurodegenerative disorders might be trigged by such changes in the intracellular pH as well (Schönichen et al. 2013). Studies have also reported pH-dependent ATPase activity (Jin et al. 2017). Other common examples of processes controlled by pH include protein/RNA stability and folding, (Stigter and Dill 1990; Garcia-Moreno 1995; Harano and Kinoshita 2006; Tang et al. 2007; Thaplyal and Bevilacqua 2014) regulation through conformational switches, (Lizatović et al. 2016) amyloid formation, (Enciso et al. 2013) protein–polyelectrolyte association, (Barroso da Silva and Jönsson 2009; Stoll 2014) protein–nanoparticle interactions (Chen et al. 2011; Barroso da Silva et al. 2014) protein–protein complexation (Sheinerman et al. 2000; Lund and Jönsson 2013; Delboni and Barroso da Silva 2016) and protein–RNA interactions (Ye et al. 2003; Koukiekolo et al. 2007; Barroso da Silva et al. 2017c). All these biomolecular systems are of importance too in many technological systems in food, brewing, pharma, bioseparations, and biomaterials in general (Chen et al. 2011; Steiner et al. 2011; Egan et al. 2014; Barroso da Silva et al. 2016; Wagoner et al. 2016).

Amino acids and nucleotides have ionizable groups that can be protonated or deprotonated depending on the solution pH (Nozaki and Tanford 1967; Thaplyal and Bevilacqua 2014). For instance, a single amino acid contains at least the amino group, whose net charge in elementary charge units (valency) can vary from +1 to 0, and a carboxyl, whose net charge can vary from 0 to −1, from the protonated to the deprotonated states in an aqueous solution. Side chains as α-carboxyl, aspartyl carboxyl, glutamyl carboxyl, imidazole, α-amino, thiol (when not involved in SS bridges), phenolic, amino and guanidyl groups in amino acids and imino nitrogens and phosphodiesters in nucleotides can also titrate (Nozaki and Tanford 1967; Thaplyal and Bevilacqua 2014). The proton-transfer mechanism can be direct from a donor to an acceptor (including proton sharing among two ionizable groups) or solvent mediated (via a proton conducting path between the donor and acceptor). Any change in the environment (e.g., ionic strength, molecular concentration, presence of other charged objects, hydration, etc.) can affect the ionization behavior of these titratable chemical groups. As suggested by Nozaki and Tanford (1967), the chemical structure of macromolecules such as a protein (or nucleic acid) might be seen as a (flexible) polymer chain with a number of ionizable side-chain groups attached on it that can all interfere with each other. At one end of this chain, there is always an amino group and a carboxyl group at the other end of the chain participating in this interplay that challenges experimental, theoretical, and computational approaches.

It seems natural to understand that at a given aqueous solution pH will control not only the amino acids, nucleotides, proteins, and nucleic acids electrical charges, but also all their multipolar moments. Consequently, their charge–charge, charge–dipole, dipole–dipole, etc. interactions with other molecules will be governed by pH. Perhaps less trivial is that these physical quantities can also fluctuate as a function of pH, which can result in *attractive* mesoscopic forces, forming an important driving force for macromolecular complexation (Kirkwood and Shumaker 1952). This peculiar electrostatic interaction, due to the fluctuations in proton charge as a function of pH, was analytically predicted by the Kirkwood–Shumaker (KS) theory (Kirkwood and Shumaker 1952). It is at the origin of the so-called "charge regulation mechanism", which is essential to explain macromolecular complexation particularly at low salt and at pH regimes closer to p*I* (the isoelectric point) (Barroso da Silva et al. 2006, 2014; Barroso da Silva and Jönsson 2009; Lund and Jönsson 2013; Barroso da Silva 2013). Since pH can also affect the macromolecular conformation, at pH regimes far from the p*I*, the macromolecule denaturation may happen as a result of the increase in the repulsive forces between its titratable groups that tend to repeal the monomers and expand the polymeric chain. This gives a tendency to expose hydrophobic residues leading them to stick on surfaces and/or aggregate as an indirect manner that pH can control inter macromolecular interactions.

Due to their unquestionable importance and rich physical chemical aspects, pH-related processes have been attracting scientific interest since ancient times. The search for the optimal pH condition at which an enzyme is most effective has probably motivated the initial investigations. It is not a surprise that various studies were carried out at Carlsberg Laboratory's Chemical Department: brewing requires to control the pH of the mash to optimize the effectiveness of the mash enzymes of the product stability and the yeast flocculation. It is also necessary to control the pH to prevent excessive tannin extraction (Lewis and Bamforth 2006; Steiner et al. 2011).

The modeling and theoretical computation of pH effects is far from trivial. Several species including the macromolecule(s), solvent, and mobile ions participate in a complex interplay of interactions and coupled mechanisms hardly to be well described by classical empirical force fields and even more difficult to be properly sampled by the present computer resources. Macromolecules are immersed in an electrolyte solution that is per se a quite complex system. Ion pairs and even larger clusters of ions can be formed due to correlations with molecular water in concentrated salt solutions (Tironi et al. 1995; Degrève and Barroso da Silva 1999a, b, 2000). Many aspects of this problem (e.g., the interconversion of ion pairs in solution) are still unclear since the complexity of the liquid state structure does not allow the easy development of exact liquid state theories. The modeling of the simple ionic species (Na, Cl, K, etc.) is another example of this field Achilles' heel. Simulation ion parameters are widely scattered, and their rational design might surprisingly reveal ambiguity (e.g., the same Lennard–Jones interaction strength parameter can be found for modeling different ionic species!) (Horinek et al. 2009). Ion specificity is another property to be taken into account (Medda et al. 2012; Becconi et al. 2017). Due to the high anisotropy of intermolecular forces involved in the solvation of these ionic species, the convergence of the required numerical sampling in molecular simulations is another complicating factor (Lyubartsev and Laaksonen 1996; Degrève and Barroso da Silva 1999a, b, 2000). Including a biomolecule in this medium increases drastically the difficulties as noted a long time ago (Northrup and McCammon 1980). Water modeling itself requires caution. A large diversity of molecular models are available for liquid water (Guillot 2002; Hess and van der Vegt 2006). Nevertheless, it seems difficult to obtain a model able to reproduce its static dielectric constant (78.43 $\pm$ 0.10 at room temperature) (Fernández et al. 1995). Only a quite few commonly used water molecular models have recently achieved a reasonable reproduction of its dielectric constant [e.g., Dill's SPC/DC model (78.3 $\pm$ 6), (Fennell et al. 2012) Barbosa's TIP4P/ϵ (78.3), (Fuentes-Azcatl and Barbosa 2016) Roux & MacKerell's polarizable water model (78.1) (Yu et al. 2013)] although the adequacy of their combination with available biomolecular force fields remains to be investigated. Up to this point, the presentation covered only nontitratable water models. The parametrization of good dissociative water potentials (and their combination with force fields parameters for the other molecular species) is another critical issue and open question (Mahadevan and Garofalini 2008). In the same line, it has not been proven yet that the classical force fields parameters obtained at a given fixed experimental pH and salt conditions and routinely used to study the biomolecular phenomena are valid to explore all other interesting physical chemical regimes different than the ones used in the calibration process.

Despite the complexity and difficulties, efforts to theoretically calculate and predict the proton binding started in remote times at Carlsberg before the modern computational era. Since then, the common philosophy has been to be able to capture the essential features of the real system in a simple tractable set of mathematical equations. As far as we are aware, the earliest theoretical–analytical attempts to study the ionization process in biomolecules is due to Linderstrøm-Lang (1924), after the experimental paper of Sorensen et al. (1917) (who had introduced before the pH definition (Sorensen 1909)) with the experimental titration of the egg albumin. Both works were conducted at Carlsberg. Later, Kirkwood's approach (Kirkwood 1934a, b) established the theoretical view of this problem. This model gave the mathematical basis of the work done subsequently by Hill, (1955, 1956a) and also the classical work of Tanford and Kirkwood (1957). The latter paper describes the famous Tanford–Kirkwood (TK) model, which is a well-known landmark in this field and will be commented out below in more details (see page 14). Tanford and Kirkwood have published several other important contributions in this area (Tanford and Kirkwood 1957; Kirkwood 1934b; Kirkwood and Westheimer 1938; Westheimer and Kirkwood 1938; Tanford 1957a, b; Roxby and Tanford 1971; Tanford and Roxby 1972). In common, these studies involve the implicit (or continuum) solvent description, i.e., the molecular water is replaced by a structureless continuum medium described by its bulk static dielectric constant. Therefore, the solvent effect only enters in these theoretical treatments by its averaged screening behavior. In contrast to explicit molecular water models, (Guillot 2002; Hess and van der Vegt 2006) in the implicit solvent description, water molecules coordinates and momenta are averaged over, losing their intrinsic molecular nature. This approximation is known as the McMillan–Mayer model level, (Friedman 1977, 1981) and it is also used in the notorious and successful DVLO theory for colloidal stability (Derjaguin and Landau 1941; Verwey and Overbeek 1948; Overbeek 1982). Another common characteristic of these models was the use of a rigid macromolecular structure without its internal degrees of freedom. This implies that any change in the protonation state does not affect the macromolecular conformation (and vice versa) that is assumed to be at the same conformation during all titration process. Well-known processes such as denaturation caused by changes in the solution pH (e.g., pepsin, the enzyme that breaks down protein in the stomach, becomes dysfunctional at high pH with gastrointestinal consequences because it changes its conformation to undergo an unfolding process (Piper and Fenton 1965)) can not be completely described by these theoretical approaches.

All these pioneering analytical works provided the solid theoretical background where the modern computational approaches, the core of our discussions here, are grounded. Different aspects of available computational methods to study acid-base processes in biomolecules have been reviewed in the literature (Mongan and Case 2005;

Chen et al. 2008, 2014; Wallace and Shen 2009; Alexov et al. 2011; Kim and McCammon 2016), evidencing their increasingly important role in biophysics, biochemistry, and biotechnological processes. Most of the reviews concentrated their discussions on atomistic level constant-pH (CpH) molecular dynamics (MD) techniques. Here, we will focus on aspects not covered before, particularly related with the fundamental ideas, simplified CpH Monte Carlo (MC) methods and the charge regulation mechanism (Kirkwood and Shumaker 1952; Barroso da Silva et al. 2006; Barroso da Silva and Jönsson 2009; Barroso da Silva 2013; Lund and Jönsson 2013). A short description of the main modern CpH methods and comparisons among them will also be partially described in this text highlighting the present state-of-art of this field, advantages, and limitations of different protocols for biomolecular systems. In the first part of this review article, we shall introduce the quantum and classical physical chemical approaches. A common set of basic thermodynamics concepts is initially introduced before the discussion of the computer models. In the central part of the text, the CpH MC methods and especially the fast proton titration scheme (FPTS) for proteins (Teixeira et al. 2010; Barroso da Silva and MacKernan 2017b) and nucleic acids (Barroso da Silva et al. 2017a) are critically presented. A following section is dedicated to some examples of the application of the CpH MC methods for protein complexation. Wherever possible, we complement the discussion providing alternative interpretations for some aspects of the problem and bringing other points of view. Future perspectives are outlined too. During our narrative, we shall also refer the reader to classical papers covering details not deeply considered here. This is the case, for instance, of the experimental approaches. A great diversity of experimental techniques has been employed to study pH effects in and between biomolecules. This is too broad of a research field and is impossible to be included in a computationally oriented review paper. Even specifically for ionization equilibria in biomolecules, a number of both macroscopic (e.g., potentiometric titrations) and spectroscopic methods [e.g., nuclear magnetic resonance (NMR), infrared, ultraviolet and visible spectroscopy] are routinely used. Here, we point the reader to specific texts (Schlichter 1980; Bartik et al. 1994; Legault and Pardi 1994; Borkovec et al. 2001; Harris and Turner 2002; Thurlkill et al. 2006).

A quantum mechanical treatment

Essentially, the proton transfer events between solute and solvent should be treated by quantum mechanics. Thus, the Schrödinger equation could be solve and after a long production time, observables values related to the proton transfer phenomenon would be evaluated. Such a scenario is not possible and many approximations are applied to simplify the quantum problem. It is worthy to cite the Born–Oppenheimer approximation, separating the motion of atomic nuclei and electrons in a molecule, and the orbital approximation, the overall wavefunction describing electrons is decomposed into antisymmetric product of monoelectronic functions, due to their importance in the electronic structure modern calculations (Szabo and Ostlund 1989). After the 1990s, density functional theory (DFT) came on the scene with low computational cost when compared to ab initio post-Hartree–Fock or even Hartree–Fock method (Capelle 2006).

Nowadays, protonation or deprotonation of molecules containing more than a dozen atoms immersed in a discrete and molecular solvent can be treated using DFT molecular dynamics or Monte Carlo simulations coupled to the rare event sampling technique. For a recent example, Tummanapelli and Vasudevan (Tummanapelli and Vasudevan 2015) have calculated the free energy profiles for proton dissociation of the 20 canonical alpha amino acids in water using Car-Parrinelo MD at HCTH-D2 level (Boese et al. 2000; Grimme 2006) and plane-wave basis set with metadynamics sampling (Laio and Parrinello 2002). Tummanapelli and Vasudevan (2015) have related a mean relative error of 0.2 pK_a units in their calculations.

Surely, the advantage of such an approach is an egalitarian treatment of the system, taking solvent and solute at same level of details. There is still room for development in sampling techniques, the choice of reaction progress coordinate (Meyer et al. 2016), new DFT methods for molecular interactions (Brémond et al. 2016; Taylor et al. 2016), and so on. Grand challenges are the DFT treatment for colloidal dimension systems as proteins and RNA/DNA or high-level quantum chemistry calculations for medium molecules with inclusion of the solvent sampling. Good ideas to solve these problems are divide-and-conquer, subsystem DFT, and fragment molecular-orbital approaches (Gordon et al. 2012; He and Jr 2010; Andermatt et al. 2016) combined with the special sampling techniques. As an example of these new approaches, Genova and co-workers have implemented and reported a speedup of 40x for simulation of (GLY)6 solvated by 395 water molecules (1230 atoms) using a frozen density embedding (FDE) formulation of subsystem DFT (1462 s for subsystem DFT against 56,405 s for conventional DFT, both carrying out ten steps of ab initio MD) (Genova et al. 2017).

At present, practical alternatives disregard the solvent and solute quantum mechanical nature, or better, only a few atoms are treated as quantum mechanics objects while the neighbors are approximated by classical ones. Thus, the reacting system can be described by a high-level ab initio

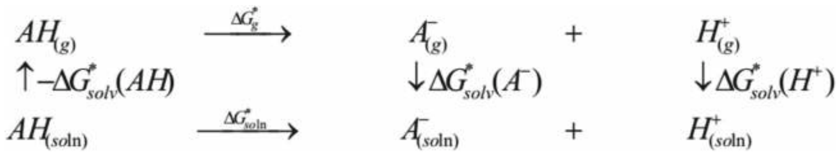

Fig. 1 Thermodynamic cycle for the pK_a calculation using the direct method. All quantities are standard-state (denoted by *) at 1 mol L^{-1}. The ΔG^*_g, ΔG^*_{soln}, $\Delta G^*_{solv}(AH)$, $\Delta G^*_{solv}(A^-)$ and $\Delta G^*_{solv}(H^+)$ are gas-phase acidity, Gibbs free energy of deprotonation in solution, Gibbs free energy of solvation for AH, A^- and proton species, respectively

method and the environment is represented by sites interacting following molecular mechanics force fields, (Kamerlin et al. 2009) or even simplified as a dielectric medium (Li et al. 2002; Freitas et al. 2007; Ho and Coote 2009b; Casasnovas et al. 2014).

Such calculations are all performed on thermodynamic cycles and, although the free-energy difference between initial and final states is not path dependent, the chosen cycle can determine the predicted pK_a accuracy (the physical meaning of this quantity and other thermodynamical ones are described in detail in the next section). As discussed by Ho and Coote (2009b), two methods are suitable for pK_a predictions: (i) the direct method, and (ii) the proton exchange method.

The direct method (scheme given in Fig. 1) combines gas-phase acidity experimental results or high level ab initio calculations with standard-state Gibbs free energy of solvation as calculated by either discrete or continuum solvent models (Marenich et al. 2009; Takano and Houk 2005; Shimizu et al. 2005; Florián and Warshel 1997; Klamt 1995).

From Fig. 1, it is possible to derive:

$$pK_a = -\log K_a = \frac{\Delta G^*_{soln}}{RT \ln 10} = \frac{\Delta G^*_g + \Delta G^*_{sol}(A^-) + \Delta G^*_{sol}(H^+) - \Delta G^*_{sol}(AH)}{RT \ln 10} \quad (1)$$

where R is the molar gas constant ($8.3144598\,J\,K^{-1}mol^{-1}$), and T is the absolute temperature.

Different thermodynamic cycles can also be derived by inclusion of solvent molecules, (Ho and Coote 2009b; Casasnovas et al. 2014) but they do not eliminate the sources of uncertainties. These uncertainties on the solvation-free energies are higher for ionic species when compared with neutral molecules (Ho and Ertem 2016). This implies that the pK_a predictions using the direct method can lead to very large errors. Also, the data scattering in the solvation free energy of proton could be an additional source of error. Actually, modern theoretical calculations for solvation-free energy of protons are very trustworthy. As an example, Rossini & Knapp have computed highly accurate proton solvation-free energies in acetonitrile, methanol, water, and dimethyl sulfoxide (Rossini and Knapp 2016). In water, they have obtained -266.3 kcal mol^{-1} against the consensus value of -265.9 kcal mol^{-1}.

The proton exchange method (scheme given in Fig. 2) is based on isodesmic reactions (i.e., reaction in which the type of chemical bonds broken in the reactant side appears in the product side).

From Fig. 2, it is simple to show that:

$$pK_a(A) - pK_a(B) = -\log\frac{K_a}{K_b} = \frac{\Delta G^*_{soln}}{RT \ln 10} = \frac{\Delta G^*_g + \Delta G^*_{sol}(A^-) + \Delta G^*_{sol}(BH) - \Delta G^*_{sol}(AH) - \Delta G^*_{sol}(B^-)}{RT \ln 10} \quad (2)$$

The uncertainties in the proton exchange method depend of the chemical similarities between the compounds in the cycle. A higher structural similarity between them must lead to higher cancellation of errors in the gas phase relative acidity calculation and solvation-free energies of ionic species. Thus, the proton exchange method must lead to more reliable pK_a estimates when compared to the direct method (Ho and Coote 2009a).

An alternative to these computationally quite expensive simulations is to invoke effective Hamiltonians that can capture at least the principal part of the real phenomenon (i.e., the free energy cost of a change in the net charge of the solute immersed in a medium). The next sections will address this question.

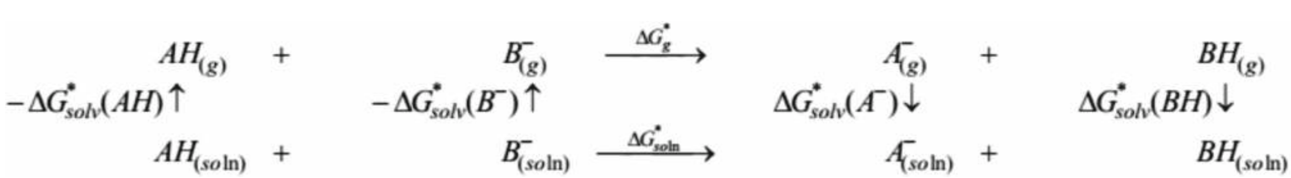

Fig. 2 Thermodynamic cycle for the pK_a calculation using the proton exchange method. The quantities have similar meaning as presented in Fig. 1

A classical physical chemical treatment

An alternative to the computationally expensive and often prohibitive quantum mechanical treatment is to invoke *effective* Hamiltonians that can capture at least the principal part of the real phenomenon. Diverse possibilities ranging from full atomistic to colloidal-like models for the macromolecules are available. Differences between the theoretical treatments can also be seen in the manner the solvent is described and/or the numerical method applied to solve a given model. Empirical methods such as PROPKA, (Li et al. 2005; Olsson et al. 2011) Vriend's method (Krieger et al. 2006) and the Burger & Ayers predictor (Burger and Ayers 2011) that have introduced plausible contributions in this field especially for high-throughput uses should also be mentioned here. Table 1 summarizes the main theoretical contributions to develop these classical effective CpH simulation methods that are also cited together with them as a function of the year of the published original works. The presentation here is not always complete, which means that the reader is expected to explore the list of references for details and other methods omitted in the text. We have picked up some main theoretical methods that we believe are key examples for an overview of the past, present, and future of this field. In the same line, Fig. 3 presents a scheme illustrating a selection of the main CpH simulation methods organized from their common features and using a historical point of view. This kind of conceptual map will be useful to relate the ideas further discussed. In the present article, we shall follow this historical perspective starting with models directly derived from the analytical ones already mentioned at the introduction section. They are represented in the first row of this picture classified as "rigid models in implicit solvent", beginning with the Linderstrøm-Lang colloidal-like protein model (Linderstrøm-Lang 1924) (at the top left) to the FPTS (at the top right) (Teixeira et al. 2010; Barroso da Silva et al. 2017a; Barroso da Silva and MacKernan 2017b). The other methods exemplified in this

Table 1 History of the main contributions to develop classical constant-pH simulation methods

Year	Author(s)	Contribution	Ref(s)
1924	Linderstrøm-Lang	Protein electrostatics described by a colloidal-like particle using the Debye-Hückel theory	(Linderstrøm-Lang 1924)
1934	Kirkwood	Mathematical basis of Tanford–Kirkwood model	(Kirkwood 1934a, b)
1957	Tanford & Kirkwood	Classical Tanford Kirkwood model	(Tanford and Kirkwood 1957)
1976	Warshel & Levitt	Proposed a microscopic dielectric model for proteins	(Warshel and Levitt 1976)
1981	Warshel	Introduced the empirical valence method to calculate pK_as	(Warshel 1981)
"	Berendsen, Postma, van Gunsteren & Hermans	Molecular dynamics simulation applied to study protein hydration	(Berendsen et al. 1981)
1982	Warwicker & Watson	Linear Poisson–Boltzmann applied for protein electrostatics (atoms represented at specific structural locations)	(Warwicker and Watson 1982)
1990	Svensson, Woodward & Jönsson	Monte Carlo simulation applied for protein electrostatics (with a uniform dielectric response)	(Svensson et al. 1990)
1996	Kong & Brooks	The "λ"'s dynamics method	(Kong and Brooks 1996)
1997	Baptista, Marte & Petersen	Coupling molecular dynamics with LPB	(Baptista et al. 1997)
2001	Barroso da Silva, Penfold & Jönsson	Monte Carlo simulation applied for protein electrostatics (with a dielectric interface)	(Barroso da Silva et al. 2001)
2004	Lee, Salsbury Jr & Brooks	Apply the "λ"'s dynamics method for CpH	(Lee et al. 2004)
2005	Li, Robertsen & Jensen	Empirical methods to predict pK_as	(Li et al. 2005)
2007	Tang, Alexov, Pyle & Honig	Use of PB to calculate pK_as in RNA	(Tang et al. 2007)
2010	Teixeira, Lund & Barroso da Silva	A fast proton titration scheme for proteins	(Teixeira et al. 2010)
2013	Chen, Wallace, Yue & Shen	Introduce a titratable water model for CpH simulations	(Chen et al. 2013)
2015	Chen & Roux	Hybrid nonequilibrium MD-MC for CpH	(Chen and Roux 2015)
2015	Carnal, Claviera & Stoll	Introduced a CG titratable flexible chain	(Carnal et al. 2015)
2016	Donnini, Ullmann, Groenhof & Grubmüller	Introduced the Parsimonious Proton Buffer	(Donnini et al. 2016)
2017	Barroso da Silva, Derreumaux & Pasquali	A fast coarse-grained model for RNA titration	(Barroso da Silva et al. 2017a)

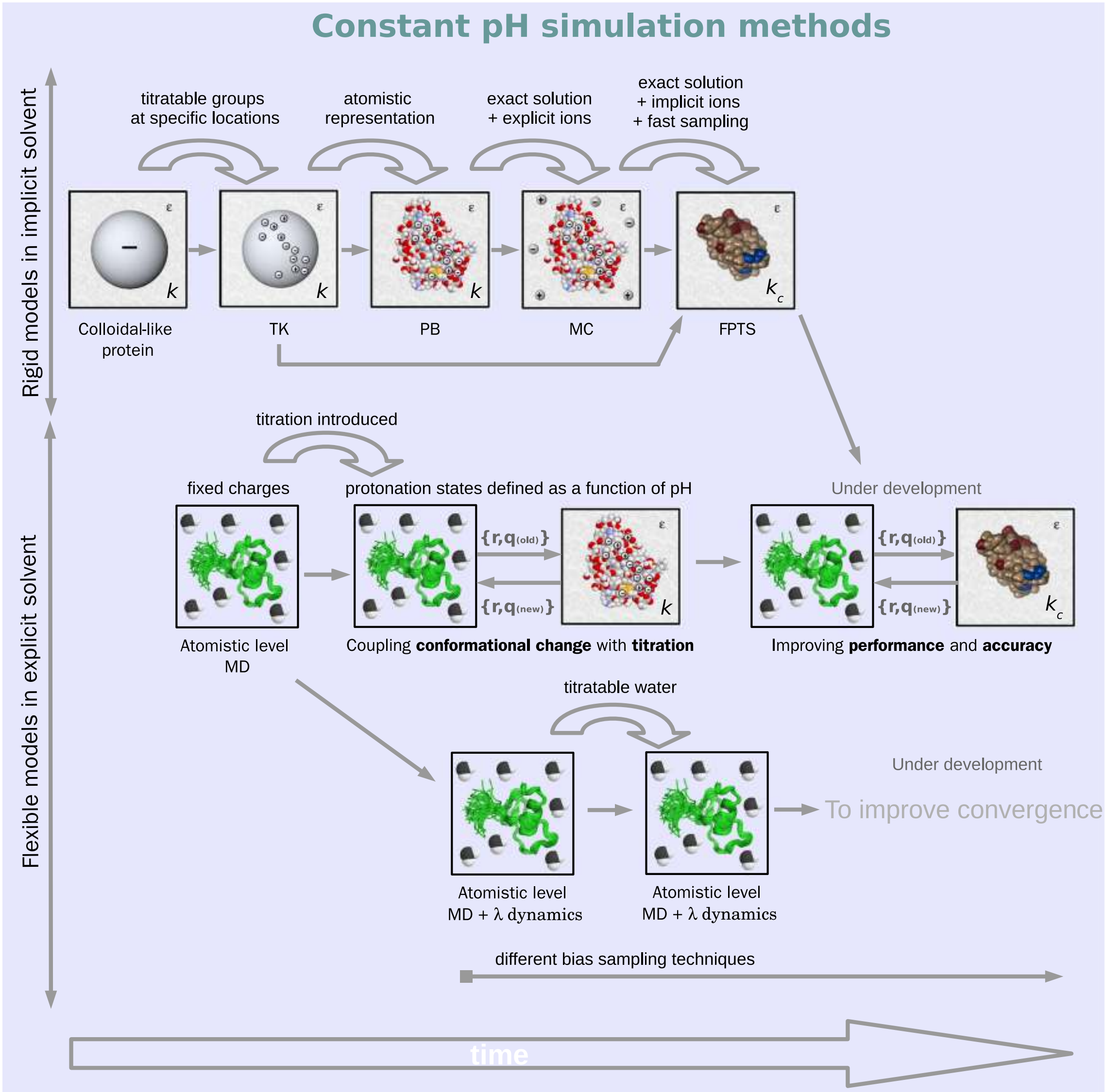

Fig. 3 Scheme illustrating a selection of the main classical constant-pH simulation methods organized based on their common features and using a historical point of view. The time arrow at the bottom indicates the chronological order. The small arrows between the models show how they evolved from the others. The double arrows (on the left) are used to classify models that allow conformational protein changes (flexible) or not (rigid). $\{r, q\}$ represents the protein atomic coordinates (r) and charges (q). κ is the inverse Debye-Hückel screening length, which is proportional to the square root of the salt concentration. This indicates the implicit ion models in this figure. Similarly, ϵ is used to label the implicit (or continuum) solvent models. Explicit (or molecular) solvent models are represented by the drawing of a water molecule. See the text for more details

figure will be introduced during our narrative after the MC schemes.

A Thermodynamical picture

Let us recall basic concepts that will be needed in the following sections. From a physical chemical perspective, the proton binding process is a chemical reaction. That is, a protonation can be written as

$$M + H^+ \overset{K_r}{\rightarrow} MH$$

According to thermodynamics, the binding constant relates to the Gibbs free energy change (ΔG_r) per mole for this reaction by: (Atkins 1995)

$$\Delta G_r = -RT \ln K_r \tag{3}$$

where K_r is the reaction constant (i.e., the binding constant). This constant is frequently expressed in concentration units, e.g., mol l^{-1}):

$$K_r = \frac{[MH]}{[M][H^+]} \tag{4}$$

Typically, K_r may be a number very small (10^{-14}) or very big (10^{+14}) (Devlin 1997). Therefore, it is much more convenient to describe it by the base-10 logarithm. That is, instead of K_r, one usually uses:

$$pK_r = -\log K_r \tag{5}$$

as is done for the pH definition.

This means that Eq. 3 can be rewritten as

$$\Delta G_r = 2.303\, RT\, pK_r \tag{6}$$

where the factor 2.303 (= ln 10) comes from the transformation between the base-10 and the natural logarithms.

However, *what is the point in calculating* ΔG_r? Well, the knowledge of ΔG_r provides a "feeling" for the reaction behavior, i.e., its driving force. For instance, it is often said that a process is spontaneous when $\Delta G_r < 0$. Conversely, if $\Delta G_r > 0$, the reverse process is spontaneous, and equilibrium is seen for $\Delta G_r = 0$. Therefore, comparing different ΔG_r's, we can have an idea of what is the effect of a specific change [e.g., mutation, variation in the ionic strength, macromolecular concentration, conformational change, presence of other (charged) molecules, etc.] on the system. It is particularly useful to quantify how specific micro-environments affect the proton binding reaction. This is really what we are after here, i.e., we are interested in calculating $\Delta\Delta G_r$, or equivalently, ΔpK_r. Particularly, we would like to know the effects that perturb and result in highly shifted pK_rs.

pK_rs for each individual titratable chemical group at a particular micro-environment (specified by the neighbor charges and ionic strength) are often obtained from a titration plot where net charge numbers (z) are given as a function of solution pH at this condition. It corresponds to the solution pH where this ionizable group i is half-protonated (e.g., $<z_i(pH)>= -0.5$, for ASP, $<z_i(pH)>= +0.5$, for LYS, and so on). Frequently, pK_r is called pK_a (or pK_{app} (Bashford and Karplus 1990)) and used to describe several physical events (e.g., protein stability, macromolecular assembly, binding of ligands, conformational changes, added salt effects, etc.) and their dependency with the environment (Garcia-Moreno 1995). This is a central physical quantity in most of the studies of biomolecular electrostatic interactions. From hereon, pK_a will be used here for the equilibrium constants when the ionizable group is at a particular physical chemistry condition specified by the temperature, salt solution, macromolecular concentration, and conformation.

For the deprotonation process, Eq. 4 can be rearranged into the so-called Henderson–Hasselbalch equation: (Devlin 1997)

$$pH = pK_a + \log \frac{[M]}{[MH]} \tag{7}$$

that can be conveniently re-written as

$$[MH] = \frac{[M]}{10^{pH-pK_a}} \tag{8}$$

Substituting this expression in the definition of the fraction of protonated molecules or degree of association (f_{MH}), one obtains:

$$f_{MH} = \frac{[MH]}{[MH]+[M]} = \frac{1}{10^{pH-pK_a}+1} \tag{9}$$

which yields the well-known sigmoid analytical titration curve for an ideal case, i.e., an isolated amino acid in the absence of any external field. Note that for pH equals to pK_a, f_{MH} becomes 0.5 as expected by the own definition of pK_a.

Alternatively, it follows that the absolute charge number for a base titratable group (e.g., LYS) is

$$z_i = \frac{1}{10^{pH-pK_a}+1} \tag{10}$$

being +1 at the very acid regime and 0 at the very basic regime. Analogously, for an acid ionizable group (e.g., GLU),

$$z_i = -\frac{1}{10^{pK_a-pH}+1} \tag{11}$$

varying from 0 (at low pH) to −1 (at high pH). Charge numbers are equivalent to valence in the chemical context, and both forms are used in biophysics texts. We shall next see how these physical quantities are computed by theoretical methods.

The Tanford–Kirkwood model Following a historical perspective, one of the earliest attempts to calculate the electrostatic ΔG_r and derived quantifies is due to Tanford and Kirkwood (1957). Employing the mathematical formalism deduced before by Kirkwood (1934a, b) they introduced discrete state variables for the enumeration of all possible protonation states of a polyprotic macromolecule. Therefore, for the first time, a model allowed the titratable chemical groups to be at specific locations, e.g., as given by the X-ray, NMR, or homology built model coordinates in the macromolecule (in the original TK model, a protein) rather than smeared out on the macromolecular surface.

The TK model belongs to the McMillan–Mayer model level (Friedman 1977, 1981). It is a dielectric continuum model (implicit solvent) that assumes that the protein (or any other macromolecule) may be modeled as a hard-sphere of radius R_p treated as a low dielectric permittivity (ϵ_p) body without internal degrees of freedom. This sphere is immersed in a medium with high dielectric permittivity (ϵ_s), the electrolyte solution. At this point, proposing a dielectric interface, they introduced a rather polemical and controversial model choice under intense debate in the literature

(Demchuk and Wade 1996; Penfold et al. 1998; Warshel and Åqvist 1991; King et al. 1991; Antonsiewicz et al. 1994, 1996; Simonson and Perahia 1995; Simonson and Brooks 1996; Löffler et al. 1997; Sham et al. 1997; Warwicker 1999; Barroso da Silva et al. 2001; Autreto et al. 2003; Schutz and Warshel 2001; Dudev and Lim 2000; Varma and Jakobsson 2004; Archontis and Simonson 2005; Ko et al. 2005; He et al. 2007; de Carvalho et al. 2008; Vicatos et al. 2009; Simonson 2013). This vast list of references represents only a small part of the papers and should be seen as an example of the scattered views of this issue.

A dielectric constant is a macroscopic parameter related to the movement of microscopic charges. It describes the reorientation of the electronic cloud around a nucleus and/or the reorientation of permanent dipoles in the presence of an electric field (Böttcher 1973). Being a macroscopic property, one might wonder how the dielectric constant can be applied on a microscopic object as a protein where atomic charges that can be at relatively short distances (< 5 Å) have to be handled. With two dielectric constants, ϵ_p and ϵ_s, a natural and necessary question is to define where one should place the dielectric interface, and also what value should be used for ϵ_p. There is a priori no prescription for locating the intervening dielectric boundaries, and a method for estimating the dielectric response of the macromolecule. Results are dramatically sensitive to these choices (Barroso da Silva et al. 2001). We shall return to this point in the next subsection (see Fig. 5). Although the physical and/or biological arguments may be criticized (Kukic et al. 2013), it seems an accepted idea in this set of strong divergent opinions that the dielectric constant of the biomolecule can be assumed to be an adjustable or empirical parameter (specific for a given model) whose choice is based on obtaining the best agreement between the predicted properties and the experimental results (Schutz and Warshel 2001; Autreto et al. 2003; de Carvalho et al. 2008).

With respect to the rigid macromolecular reference frame with fixed charged groups, the salt ions and other charged ligands are supposed to be in relative motion throughout the solvent medium. However, Tanford and Kirkwood avoided the formidable statistical thermodynamics problem that this motion implies by appealing to the construction of an effective interaction, eliminating explicit reference to the mobile particles and introducing the Debye–Hückel (DH) potential. This means that the salt ions and other charged ligands were not explicitly taken into account. They just "participate" through their mean-field contribution. All protein charges are assumed to be inside the low dielectric sphere (radius R_d). A schematic picture of the TK model is given in Fig. 4.

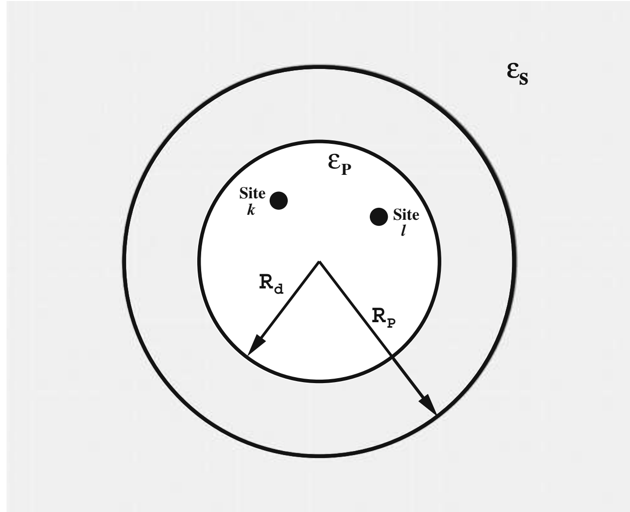

Fig. 4 The Tanford–Kirkwood model (adapted from ref. Tanford and Kirkwood (1957)). A spherical protein of radius R_P immersed in an electrolyte solution. The protein interior with a low dielectric permittivity (ϵ_P) is shown as a shaded region of radius $R_d < R_P$. Two protein titratable sites k and l are represented at a given specific structural locations. The solvent dielectric constant is ϵ_s. See text for more details

Basically, the TK model assumes that the electrostatic contributions to the free energy may be calculated by:

$$\Delta G_r = \frac{1}{2}\sum_{i=1}^{N} q_i \phi(\mathbf{r_i}) \tag{12}$$

where q_i is the net charge of site i and $\phi(\mathbf{r_i})$ is the electrostatic potential at position $\mathbf{r_i}$, which was obtained from the DH theory. The full analytical expressions of the model are given in their original work (Tanford and Kirkwood 1957) and critically analyzed in Barroso da Silva et al. (2001) by means of MC simulations. These simulations were carried out for a model including the dielectric discontinuity and the mobile species in the solution comprise salt particles as well as additional counter ions in order to maintain electroneutrality (Barroso da Silva et al. 2001).

The success of the TK model may be seen by the number of investigations where it has been invoked to study the interactions between charged ligands and proteins, membranes, and other macromolecules (e.g., Harvey (1989), Warwicker and Watson (1982), Warshel et al. (1984), Bashford et al. (1988), Havranek and Harbury (1999), and Teixeira et al. (2010)). Despite its evident success, the TK model suffers from two significant limitations:

1. closed form analytical solutions are only available in simple geometric configurations (usually spherical);
2. nonlinear thermal effects and explicit ion–ion interactions are ignored.

An additional simplification of the TK prescription, which may prove unrealistic at low ionic strength, lies in the assumption of infinitesimally small macromolecule concentration (Linse et al. 1995). In NMR studies of proteins, the typical protein concentration is of the order of 1 mM and the concentration of the accompanying counter ions could be an order of magnitude larger or more. Anomalous salt effects at moderate macromolecular concentrations have also been reported in the literature (Barroso da Silva et al. 2005).

These approximations were scrutinized in a set of numerical simulations (Barroso da Silva et al. 2001; de Carvalho et al. 2006, 2008). Unexpectedly, the findings show that the TK prescription is an excellent approximation for studies of the binding of charged ligands to macromolecules, especially at moderate- and high-salt concentrations. Macromolecules moderately charged (less than 10 units of elementary charge) give the best response of the model. For sufficiently highly charged systems, the limitations of the DH theory become apparent. The linearization in the DH equation is unjustified and one must invoke a more accurate theory (e.g., the nonlinear Poisson–Boltzmann equation). Ion-ion correlation effects are of minor importance for ion-binding measurements at symmetrical 1:1 electrolyte solutions. The TK predictions for the free energy shifts become less reliable at moderately macromolecular concentrations. This can be remedied by the replacement of the original DH screening length by its modified version that incorporates the counter ion concentration (Beresford-Smith and Chan 1983; Schmitz 1994).

The Poisson–Boltzmann equation With the advent of faster computers together with improvements in numerical algorithms, it became possible to numerically solve the TK model for arbitrary molecular shapes. It was the beginning of the Poisson–Boltzmann (PB) equation era in biophysics and biochemistry. The pioneering and landmark work of this "new" approach for macromolecules represented at atomistic level and nonuniform dielectrics (in implicit solvent) is due to Warwicker and Watson (1982), which was followed by many others (e.g., Davis and McCammon (1990), Holst (1993), Davis et al. (1991), Juffer et al. (1991), Juffer (1998), Honig and Nicholls (1995), Bashford et al. (1988), Bashford and Karplus (1990), Beroza et al. (1991), Warwicker (1999), Baker et al. (2001), Li et al. (2005), and Anandakrishnan et al. (2012)). This approach is indicated in Fig. 3 on the top row, third picture from left to right. The work of Warwicker & Watson came out about 1 year later after Warshel introduced the empirical valence method (EVM) for fast pK_as estimations (Warshel 1981). These two theoretical methods (PB and EVM) were proposed a few years later after the Nobel laureates Warshel & Levitt published in 1976 the microscopic dielectric model for proteins where the solvent and protein atoms were represented as an explicit grid of polarizable Langevin-type dipoles (Warshel and Levitt 1976). This Langevin dipolar (LD) model was combined with a quantum description in an approach that is now known as "multiscale modeling". Taking a developmental route independent from and in a way apart from the TK and the PB approaches, the LD model has evolved in parallel in many other (semi-)microscopic treatments resulting in the protein-dipoles–Langevin-dipoles (PDLD) model family (Warshel et al. 2006). In fact, Warshel made several contributions to protein electrostatics and modeling of the biological function that can be appreciated in a recent review written by himself (Warshel 2014).

The PB equation is obtained by the combination of fundamental electrostatic equations, the Poisson and the Boltzmann equations (a detailed and mathematically oriented derivation can be found elsewhere (Holst 1993)). The Poisson equation is used to calculate the three-dimensional electric potential (ϕ) generated by a macromolecule lying in an ionic solvent. This situation corresponds to a calculate ϕ for a local electric charge density (ρ_e) in a dielectric medium (assumed homogeneous and linear), (Reitz et al. 1986)

$$\nabla^2\phi = -\frac{\rho_e}{\epsilon_s\epsilon_0} \tag{13}$$

where ϵ_0 is the vacuum permittivity ($\epsilon_0 = 8.854 \times 10^{-12}$ C^2/Nm^2). To solve the Poisson's equation, one must know both ρ_e, which is supposed to be given by the Boltzmann's distribution and the boundary conditions. The Laplace operator, ∇^2, must be written in terms of an appropriate coordinate system (rectangular, spherical, cylindrical, etc) exploiting all the problem symmetries.

Let us consider the simpler case, where a positively charged surface is surrounded by cations and anions. Any of these ions apart from the surface experiences a potential ψ that is a result from the *average* force acting on one particular ion from the interactions of both the surface and all the other ions. Thus, ψ is a potential of mean force (Russel et al. 1989; McQuarrie 1976; Hill 1986) defined such that the Boltzmann distribution for the cations and anions is given by: (Shaw 1992)

$$n_k = n_{0k} \exp\left[\frac{-z_k\, e\, \psi_k}{k_B T}\right] \tag{14}$$

where k_B ($= 1.3807 \times 10^{-23}$ $J.mol^{-1}.K^{-1}$) is the Boltzmann constant, e is the elementary charge (e $= 1.602$ 10^{-19}C), z_k is the ion valency, n_k is in units of particles per volume (number density), and their density in the bulk solution is n_{0k}. Consequently, the charge density ρ_e can be, for a symmetric salt, written as:

$$\rho_e = z\,e\,(n_+ - n_-) \overset{Eq.14}{\overset{\downarrow}{=}} z\,e\,n_0\left(\exp\left[\frac{-z_+\,e\,\psi_+}{k_B T}\right] - \exp\left[\frac{-z_-\,e\,\psi_-}{k_B T}\right]\right) \quad (15)$$

where $z = z_+ = -z_-$. Since ρ_e in this case is the density of the *free* mobile charges at this point, it is more appropriate to call it ρ_f from now on.

Apparently, the expression above for ρ_f, together with the boundary conditions, is all that we need to write the PB equation. Nevertheless, the potential of mean force ψ (used in Eq. 14) is *not* the mean electrostatic potential $<\phi>$ (used in Eq. 13). The discrimination between these potentials is pedagogically presented by Lyklema (1991). As a first assumption, one can neglect ion–ion correlations, which gives $\psi = <\phi>$, using the type of approximation that is called a *mean-field approximation*. Therefore, substituting Eqs. 15 in 13, and writing $\psi = \phi$ results in the well-known PB equation for a two-component system at a charged surface:

$$\nabla^2\phi = -\frac{\rho_f}{\epsilon_0\,\epsilon_s} \overset{\psi=\phi}{\overset{\downarrow}{\approx}} \frac{2\,z\,e\,n_0}{\epsilon_0\,\epsilon_s}\sinh\left(\frac{z\,e\,\phi}{k_B T}\right) \quad (16)$$

This is a *nonlinear* equation partial differential of the second order and its mathematical solution (analytical or numerical) can be quite complex and tricky. Analytical solutions are available only for very simple cases. One of these special situations where an analytical solution is known is the infinite charged planar surface case. This example is given in detail in Refs. Russel et al. (1989), Evans and Wennerström (1994), and Usui (1984), and is usually described as the Gouy–Chapman case (Usui 1984). In many cases, including the applications to biomolecules, numerical techniques are required. Different methods are available. For example, the "finite element method", (Davis and McCammon 1990; Project 1995; Harvey 1989; Orttung 1977; Terán et al. 1989) the "finite difference method", (Davis and McCammon 1990; Project 1995; Harvey 1989; Warwicker and Watson 1982; Holst 1993; Davis et al. 1991; Sakalli and Knapp 2015) and the "boundary element method" (Davis and McCammon 1990; Project 1995; Harvey 1989; Juffer 1993, 1998; Juffer et al. 1991) are applied to solve the PB equation for biomolecular systems. There are also a number of generalized program packages available to study biomolecular phenomena (Warwicker and Watson 1982; Holst 1993; Davis et al. 1991; Bashford and Gerwert 1992; Honig and Nicholls 1995; Juffer 1992) and web-servers (Calixto 2010; Anandakrishnan et al. 2012; Smith et al. 2012; Wang et al. 2016). A quite recent new numerical implementation is the Gaussian-based dielectric function description for the nonlinear PB (NLPB) (Wang et al. 2015).

For multi-component systems with N ionic species, one should recognize that ρ_f is equal to the local excess of ionic charges: (Russel et al. 1989)

$$\rho_f = \sum_{k=1}^{N} e\,z_k\,n_k \quad (17)$$

where z_k and n_k are the valency and the number density of charges of species k. The expression for n_k is given by Eq. 14, where n_{0k} is the bulk density of species k. This results in

$$\nabla^2\phi = -\frac{1}{\epsilon_0\,\epsilon_s}\sum_{k=1}^{N} e\,z_k\,n_k = -\frac{1}{\epsilon_0\,\epsilon_s}\sum_{k=1}^{N} e\,z_k\,n_{0k}\,\exp\left(\frac{-ez_k\phi}{k_B T}\right) \quad (18)$$

which is a generalization of Eq. 16.

Due to the complexity of the non-linear equation, many studies are still done with the linear form, i.e., within the same statistical mechanical basis as the TK model involving the DH approximation. This is apparently the most popular approach today. In fact, great enthusiasm for this method is probably related to the fact that a dielectric interface can be relatively easily included in the model and aids to tune results to meet the experimental data. This turns out to be an effective way to remedy and introduce possible effects of conformational changes due to the variation of the protonation states, structural artifacts that might be induced from the crystal symmetry imposed by the crystallization process, low-quality homology built model, the lacking of ion–ion correlation in the mean-field description, and any other contribution missing in the model.

The superoxide dismutase (SOD), an enzyme that strongly interacts with the negatively charged superoxide radical (O_2^-) and linked to Gehrig's disease (Hurtley 2015), was probably the system that opened up the claims that a macromolecule should be treated as a body with low dielectric permittivity within the continuum model approach. Experimental data show an attraction between SOD and O_2^- under certain conditions. However, both molecules have paradoxically a negative net charge (SOD has a net charge number of -4.), which according to Coulomb's law should result in a repulsion between them. Calculations with a uniform dielectric constant in the PB approach failed to explain the experimental behavior, i.e., their attraction. Nevertheless, the assumption of a low dielectric constant for SOD ($\epsilon_p = 2$) completely changes the picture, revealing a region of positive electrostatic potential around the active site, where O_2^- should bind (Sharp et al. 1987). Conversely, this

"SOD paradox" has also been studied by means of MC simulations (Woodward and Svensson 1991; Bacquet et al. 1988; Barroso da Silva 1999). Contrary to previous PB studies, (Sharp et al. 1987) it was found that there was no clear need to consider a low dielectric permittivity to the enzyme, since the attraction found in these calculations was smaller than 0.1 k_BT units.

It is worth pointing out that the PB is likely to fail at high electrostatic coupling regimes. This can be found for example in the presence of multivalent ions, lowering the temperature or the dielectric constant, and/or increasing the charge of the macromolecular surface (Jönsson et al. 1996, 2007; Degrève et al. 1993). Thus, the validity of the PB approach can be questionable in some biomolecular conditions. Some authors claimed that the PB approach is valid for electrolyte solutions with concentrations that do not exceed 1 M and surface potentials less than 200 mV (Russel et al. 1989). However, Outhwaite and Bhuiyan argued that there is consensus only when the PB theory is applied to 1:1 electrolytes (Outhwaite and Bhuiyan 1991). Size ion–ion correlation can be included in the "modified Poisson–Boltzmann" version (Outhwaite and Bhuiyan 1991; Degrève et al. 1993). Other key assumptions and limitations in the PB approach are discussed elsewhere (Barroso da Silva 1999). There are also more refined statistical mechanical theories that can be used to replace the DH approximation (e.g., the hypernetted chain integral equation González-Tovar and Lozada-Cassou (1989) and Terán et al. (1989), the anisotropic reference hypernetted chain approximation (Greberg and Kjellander 1994), and the density functional theory (Lovett et al. 1976)). A more consistent alternative is to perform MC calculations that does provide an "exact" answer (within statistical errors) for a given physical model.

PB equation solvers are largely applied to predict pK_a for both proteins and nucleic acids systems (Sharp and Honig 1990; Wang et al. 2015; Tang et al. 2007). Standard descriptors to measure the quality of the PB equation results (or of any other theoretical method) in benchmark studies is done employing the maximum absolute deviation (MAX), the averaged absolute deviation (AAD), the root-mean-square deviation (RMSD) and the linear correlation coefficient (r) between the experimental and computed pK_as. The quality of the outcomes are often scrutinized by means of a comparison with the so-called "NULL model", where site–site interactions are altogether neglected (Schutz and Warshel 2001; Carstensen et al. 2011). This is equivalent to assume that the pK_a of a given titratable group at any experimental condition is identical to its model compounds given zero pK_a shifts ($\Delta pK_a = pK_a - pK_0 = 0$). Any good theoretical model should have a better prediction than the "NULL model", which is a task far from trivial to be achieved (Borkovec et al. 2001), making such comparison a real critical test in benchmark studies. Tests done with PB solvers for several proteins resulted in an overall RMSD of ca. 0.8 (Wang et al. 2015). Buried residues follows the typical trend to be predicted with more difficulties (RMSD of 1.1 (Wang et al. 2015)) than superficial groups. Nevertheless, as mentioned above, the choice of ϵ_p has a drastic effect on the outcomes. Figure 5 illustrates how these descriptors are sensitive to what value is assumed for ϵ_p. The data is for lysozyme (PDB id 2LTZ) at 100 mM of salt following the same simulation details as done by Bashford & Karplus with $\epsilon_p = 4$ (Bashford and Karplus 1990). It can also be observed in this figure that any $\epsilon_p > 10$ would result in better predictions than the "NULL model". Even a uniform dielectric response ($\epsilon_p = \epsilon_s$) would give similar results to the nonuniform dielectrics cases despite the fact that there seems to be an optimal ϵ_p between 20 and 25 that gives the best values for these descriptors (2.0–2.1, 0.6, 0.8 and 0.97–0.98 for MAX, AAD, RMSD, and r, respectively). It is worth mentioning that these results depend on the choice of the force-field parameters for the charges used to assign the protein and residues atomistic partial charges (Calixto 2010). Depending on the chosen force field, the optimal ϵ_p can be ϵ_s supporting the uniform dielectric response for this class of methods. In general, PB solvers using the finite differences method can improve the results repeating the calculations for rotated protein coordinates (e.g., $\pm 5°$) (Madura et al. 1994). Conversely, improving the molecular surface boundaries and using the finite element method, Sakalli & Knapp impressively obtained the smallest RMSD for lysozyme: 0.18. Multi approaches with a good description of proton isomerism have also been reported in the literature (Magalhães et al. 2017). In addition, a practical detailed information for the PB simulation protocols is given in this paper, which may be useful to the reader.

Titration schemes based on the Monte Carlo method

In terms of generic ion binding calculations by MC methods, the first paper that we are aware of is the work with calcium-binding proteins done by Svensson, Woodward & Jönsson (1990) in implicit solvent (Svensson et al. 1990), after they have developed the modified Widom's method for non-uniform electrolyte solutions (Svensson and Woodward 1988). Electrostatic free energies were obtained by the employment of this perturbation technique. A similar strategy was subsequently used by Barroso da Silva and coauthors to study the titration of fatty acids solubilized in cationic, nonionic, and anionic micelles (Barroso da Silva et al. 2002). In short, the model assumed a uniform dielectric response ($\epsilon_p = \epsilon_s = 78.7$ at 300 K), and removed the mean-field description for the electrolyte solution that was replaced by free explicit mobile ions (added salt and

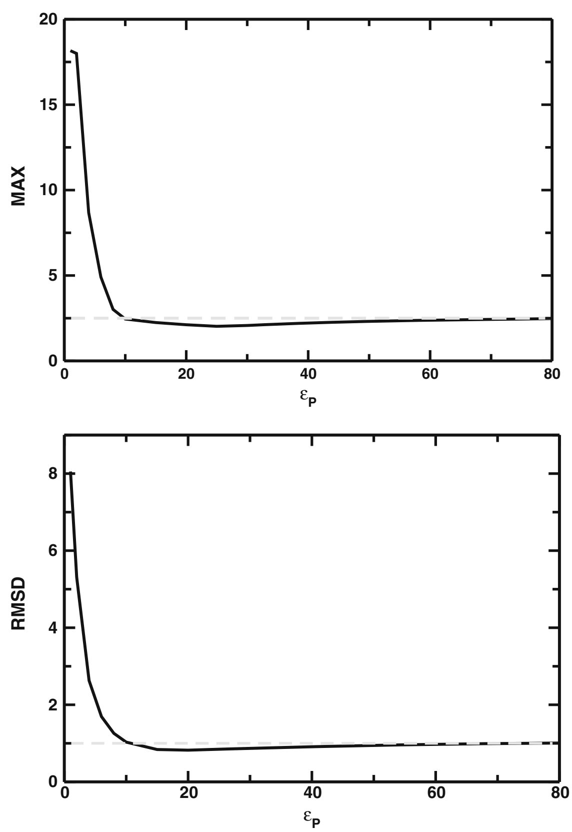

Fig. 5 Standard descriptors to measure the quality of calculated pK_a values by the PB equation as a function of the protein dielectric constant ϵ_p. The "NULL model" predictions are given by the dashed gray line. For these calculations, the pK_0s were taken from ref. Bashford and Karplus (1990) where simulation details are also described for $\epsilon_p = 4$. The PB calculations were carried out with the package MEAD 2.2.9. (Bashford 1997)

counter ions). These ions modeled by the restrictive primitive model (Levesque et al. 1986) were introduced in the simulation box instead of the DH/PB approximation. A rigid protein (or micelle) made of a collection of small charged hard-spheres of radius R_a (typically, $R_a = 2$ Å) and number charges (valences) z_a mimicking its atoms was placed and kept fixed at the center of an electroneutral spherical cell of radius R_{cell} (see the forth picture at the top row in Fig. 3). This corresponds to the *cell model* (cm) (Hill 1956b; Jönsson 1981). The entire system is confined in this closed spherical container. Particles only interact with other particles that are present in this cell. There is neither replicas nor nearest images. The particles cannot escape from this container. A full atomistic representation for the protein was followed with its atoms located according to the X-ray structures provided by the Protein Data Bank (Berman et al. 2000). No intramolecular degrees of freedom of the protein were included in the model. The valency of each atom was assigned based on their titratable characteristics but they were not allowed to change during the simulation run in these earliest works. Different charge schemes were tested (Svensson et al. 1990; Teleman et al. 1991). It was shown that partial charges on all protein atoms are not necessary in order to obtain good agreement with experiment. The crucial point is to have the appropriate net charges in the ionized residues.

In this model, the interaction between any two particles i and j is given by,

$$u(r_{ij}) = \begin{cases} \infty & , \ r_{ij} \leq 2R_a \\ \frac{q_i q_j}{4\pi \epsilon_0 \epsilon_s r_{ij}} & , \ \text{otherwise} \end{cases} \quad (19)$$

where $q_i = z_i e$ and $q_j = z_j e$ denote the charges on particles i and j, respectively, and r_{ij} their separation distance.

An external potential $v^{ex}(r_i)$ is used to impose a hard wall that defines the spherical cell,

$$v^{ex}(r_i) = \begin{cases} 0 & , \ r_i \leq R_{cell} \\ \infty & , \ \text{otherwise} \end{cases} . \quad (20)$$

The size of the cell is determined by the macroparticle concentration. That is the only way that macroparticle–macroparticle interactions enter in the model. Therefore, this potential permits the definition of a protein concentration for the system instead of the common periodic box conditions (pbc) frequently used in molecular simulation (Allen and Tildesley 1989). Obviously, one is partly neglecting correlations between macroparticles. This is not a crucial simplification, when the electrolyte solution is 1 : 1. That is because in this case there is often a strong repulsion between two identical macroparticles. For multivalent free ions, some care should be taken, since at some conditions these two identically charged macroparticles can *attract* each other (Linse and Lobaskin 1999). A comparison between the cm and the pbc applied to micellar systems has been performed by Linse and Jönsson (1983). They concluded that the cm accurately predicts thermodynamics properties at low micellar concentration (Linse and Jönsson 1983). However, it should be noted that for dilute concentrations of the macromolecule and relatively high concentration of added salt, the number of particles necessary in the simulation becomes prohibitively high. Therefore, one always tries to work with smaller systems (increasing the macromolecular concentration). In this case, runs with different numbers of particles and cell sizes within the computational available resources are carried out to check any inappropriate boundary effect. The cm has the additional benefit that all electrostatic interactions within the cell can be exactly treated in the sense that no cutoff scheme (like the Ewald summation (Hünenberger and McCammon 1999)) or potential truncation need to be included.

The total energy of the system for a given configuration is then,

$$U = \sum_{i=1}^{N_{mob}} v^{ex}(r_i) + \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} u(r_{ij}) \qquad (21)$$

$N_{mob} = N_c + N_s$ is the total number of mobile particles comprising N_c counterions and N_s added salt ions. $N = N_{mob} + N_p$ is the total number of particles including the N_p protein atoms.

Another intermediate model combining the TK's idea to display the ionizable charged groups within the protein with MC movements was proposed by Linse and co-authors and applied to investigate protein self-association, protein–polymer complexation and protein adsorption to charged surfaces (Carlsson et al. 2001a, b, 2004). Titratable sites could carry either a positive charge, a negative charge, or no charge based on their individual experimental pK_as (Carlsson et al. 2004). These charges were kept unmodified during the MC run.

Later, the fixed charge model was replaced by a proper proton titration scheme (Kesvatera et al. 1996, 1999, 2001) leading to the first CpH MC method in implicit solvent (with explicit mobile ions). Simulations were performed in a semi-grand canonical ensemble. The total number of particles is constant, but their charges can vary during the protonation process (i.e., they change their identities). A proton bath was coupled at the simulation cell in order to establish a constant pH in the system. After every tenth attempted move of the free mobile ions, an attempt was made to delete/insert protons on the ionizable groups. The acceptance/rejection of an attempt to change the protonation state of a residue was based on the trial energy,

$$\Delta U_{titra} = \Delta U_c \pm k_B T \ln 10(pH - pK_0) \qquad (22)$$

where ΔU_c is the corresponding change in Coulomb energy that gives the deviations from the ideal behavior (e.g. interaction with other charged amino acids, counter-ions, added salt, etc.), and pK_0 is the dissociation constant of the model compound. These values were taken from experiments performed by Nozaki and Tanford (1967). Corrections to this titration model were proposed later (Labbez and Jönsson 2007).

Other models in the same lines with improved features were developed by Stoll and collaborators (Stoll 2014; Carnal et al. 2015). An interesting characteristic of their CG model is the inclusion of some internal degrees of freedom of the protein allowing this computer model to explore the pH effects on the macromolecular (flexible) chain. Doing so, they could observe extended and folded conformations as a function of the solution pH (Stoll 2014; Carnal et al. 2015).

The use of the phenomenological description for the acid-base equilibrium by the second term $[k_B T \ln 10(pH - pK_0)]$ introduces pH as a simple input parameter in the calculation (covalent bonds cannot be broken in such models). This second term in Eq. 22 accounts for the (electrostatic) free energy change of the (de)protonation process for a titratable group, not affected by the presence of any other site nor by the interaction with any other mobile charged (added salt and counter-ions). Only the *differences* in free energy between the residue in the macromolecule chain and the corresponding reference protonation state for which the pK_a was originally obtained are taken into account. It is assumed that the electrostatic interactions are predominantly responsible for the shifts in ionization process. Other interactions as solvation effects are assumed to be the same in both microenvironments. This approximation, common in all numerical schemes that invoke a phenomenological approach, results in a drawback for this class of models. It is clear that the microenvironment of a ionizable group free in an aqueous solution is quite different from the situation where the same group is deeply buried in a macromolecular interior (even if no other charges are present!). This could already be seen when we discussed the PB's RMSD for buried amino acids above. As a matter of fact, a compilation of experimental pK_as values of the ionizable groups of proteins done by Pace and co-authors (Thurlkill et al. 2006) demonstrated that the pK_as are quite sensitive to the microenvironment conditions (temperature, ionic strength, and short peptide used to measure the pK_as of a given compounds). For example, the experimental pK_0 for the α-Carboxyl can vary from 3.0 to 4.3 depending on small variations of these conditions (Thurlkill et al. 2006). Nozaki & Tanford quoted 3.8 for this group (Nozaki and Tanford 1967). Fortunately, most of the titratable groups of biomolecules are close to the surface, and this approximation is expected to have a minor effect on the majority of interesting cases. Warshel and collaborators (Warshel et al. 1984; Warshel and Åqvist 1991; Warshel and Papazyan 1998; Schutz and Warshel 2001) proposed an elaborative alternative to the deal with buried ionizable groups by means of intermediate approaches such as the dipole-lattice model (Warshel et al. 2006; Warshel 2014) already mentioned above.

From an experimental point of view, for the interpretation of pK_0, we can also argue in a similar manner based on the colloidal literature. In truth, this discussion is similar to proton association/dissociation from/to a amphiphilic molecule of a micelle. The classical works of Mukerjee and Banerjee (1964) and Fernández and Fromherz (1977) suggested a relative simple equation to measure ΔpK based only on the electrostatic interactions and neglecting non-electrostatic interactions. They also left for debate a fundamental problem: the correct measurement of pK_0 (the reference dissociation constant). For micelles, the problem has sometimes been circumvented by assuming that pK_0 would be equivalent to the pK value obtained in pure water. However, there are studies that have indicated that this is not always the case. Instead, experiments with nonionic micelles are usually performed and believed to give better estimates of pK_0 which is unfortunately not available for proteins and nucleic acids. Even for micelles, the choice of the reference pK is still an unsolved problem (Barroso da Silva et al. 2002).

An explicit mobile ion description as followed by this CpH MC method offers the possibility to properly describe ion–ion correlation and anisotropic–salt interactions in titration studies. Nevertheless, it resulted in poor acceptance ratios when this titration model is applied to biomolecular interaction studies (e.g. protein–protein complexation). Very often a hard-core overlap with these charged particles happens when translating and/or rotating a macromolecule in a MC trial displacement movement. Moreover, since the CPU costs are roughly proportional to the square of the number of interacting sites, high ionic strength conditions resulted in prohibitive CPU costs. For the sake of convenience, theoretical studies of the macromolecular complexation have a tendency to be repeatedly carried out at very dilute salt conditions (Barroso da Silva et al. 2006; Jönsson et al. 2007; Barroso da Silva and Jönsson 2009). On the top of that, highly attractive or repulsive systems are naturally harder to sample due to their typical higher energetic barriers (Barroso da Silva et al. 2006).

The fast proton titration scheme—FPTS Aimed to developed a faster proton titration protocol for macromolecular systems with *multi* titrating objects each containing several ionizable sites with an implicit salt description, the TK model (Tanford and Kirkwood 1957; Barroso da Silva et al. 2001) was called out to inspire a new simplified titration scheme where salt is treated at the DH level. Following a coarse-grained (CG) description of the macromolecular system (Noid 2013) in implicit solvent and a phenomenological physical chemical approach, the FPTS for proteins was proposed (Teixeira et al. 2010) successfully reducing the computation time and also efficiently boosting sampling for applications in protein complexation studies (e.g., Teixeira et al. (2010), Persson et al. (2010), Kurut et al. (2015), Delboni and Barroso da Silva (2016), and Barroso da Silva et al. (2016)). This is clearly the main differential of this titration method that forward biomolecular applications on the large-scale scenario for protein–protein, protein–RNA/DNA, protein–polyelectrolyte and protein–nanoparticle interactions. Quite recently, the model was extended to nucleic acids (Barroso da Silva et al. 2017a).

Details of the new scheme are given in Refs. Teixeira et al. (2010), Barroso da Silva et al. (2017a), and Barroso da Silva and MacKernan (2017b) where the reader is refer to. In short, the main difference between the two MC titration models (with explicit mobile ions and the FPTS) is the replacement of Eq. 22 by

$$w_{TK} = \frac{e^2}{4\pi\epsilon_0\epsilon_s}\left[\sum_{i>j}^{N_p}\frac{z_i z_j}{r_{ij}} - \frac{Z_p^2\kappa_c}{2(1+\kappa_c b)}\right] +\lambda(pH - pKa)ln10 \tag{23}$$

where $Z_p = \sum_i^{N_p} z_i$, λ equals either -1 (deprotonation) or $+1$ (protonation), κ_c is the modified Debye parameter as suggested by Beresford-Smith and co-workers, (Beresford-Smith and Chan 1983) and b is assumed to be equal to the radius of a sphere that inscribes the macromolecule (Teixeira et al. 2010). The titration process as given by this equation was obtained from basic physical chemical arguments and is converted into a simple but efficient MC protocol (Teixeira et al. 2010). See Ref. (Barroso da Silva and MacKernan 2017b) for more details, and where its approximations and possible limitations are also discussed.

Invoking a DH treatment for the salt implies that such semi-empirical model assumed a mean-field approximation and neglected ion–ion correlation effects. However, despite this similarity with the PB approach and contrary to it, the FPTS is based on a MC process for protonation/deprotonation and incorporates aspects neglected by other methods (e.g., the chemical potential contribution between the two possible titratable states—see Ref. Barroso da Silva and MacKernan (2017b)). Such different aspects gave new features to the FPTS, widening its potential scope of application to include the modeling of systems with multiple ionizable in several experimental conditions (pH, salt, temperature, etc.) especially where charge fluctuations are important, at very lower CPU costs.

Surprisingly, despite all approximations, predicted pK_a values obtained by the FPTS are in general at least within the range of values given by different theoretical models (Stanton and Houk 2008; Chen et al. 2013; Barroso da Silva and MacKernan 2017b). A recent benchmark study demonstrated that in fact even atomistic-level molecular dynamics simulations at constant pH do not obtain better results than FPTS for some protein systems. In general, they are often poorer, and orders of magnitude more computationally expensive. In comparison with experimental measurements for proteins with a diversity of structural features, calculated pK_a values by the FPTS have the average, maximum absolute, and root-mean-square deviations of [0.4 − 0.9], [1.0 − 5.2], and [0.5 − 1.2] pH units, respectively, (Barroso da Silva and MacKernan 2017b). Recall that the overall RMSD for PB predictions is ca. 0.8. (Wang et al. 2015). For some protein systems, such as the binding domain of 2-oxoglutarate dehydrogenase multi-enzyme complex (PDB id 1W4H), the α-lactalbumin (PDB id 1F6S) and the turkey ovomucoid third domain (PDB id 1OMU), the predicted pK_a are closer to experimental results than any other modern theoretical methods. Similar or even better outcomes were observed for RNA systems at much lower computational costs (Barroso da Silva et al. 2017a). Predictions were in the large majority of the studied protein and RNA cases more accurate than the NULL model (Barroso da Silva et al. 2017a; Barroso da Silva and MacKernan 2017b).

Typical calculated titration plots by the FPTS for protein ionizable acid amino acids are shown in Fig. 6. Such graphics measure the degree of protonation, and are equivalent to the unprotonated fractions plots commonly reported in the literature (Wallace and Shen 2009). The data are for the turkey ovomucoid third domain (OMTKY3) at 10 mM. All these residues ASP-7, ASP-27, GLU-10, GLU-19, and GLU-43 behavior are well reproduced by the FPTS, as seen in Table 2. In this table, experimental and computed pK_a values by the hybrid nonequilibrium molecular dynamics–Monte Carlo (neMD−MC), (Chen and Roux 2015) PropKa, (Olsson et al. 2011) the NULL model and FPTS are shown. The smallest MAX (0.85), AAD (0.49), and RMSD (0.57) are obtained by the FPTS. The best linear correlation (r) between experimental and computed pKas are also given by this method ($r = 0.98$). In terms of RMSD, FPTS is followed by PropKa, which gives deviations at an intermediate level (RMSD = 0.92) in comparison with other theoretical schemes. These data confirm that FPTS is able to reproduce experimental pKa shifts even better than more sophisticated and expensive methods for some systems regardless of the model approximations adopted to speed up calculations. The higher RMSD is observed for the hybrid neMD−MC method (RMSD = 0.97). This is also the only theoretical method whose maximum absolute deviation is worse than the NULL model. From this result, apparently, the numerical convergence was probably not reached by the hybrid neMD−MC method for this specific protein system. Longer simulations might be necessary to properly sample the system due to its characteristically slow convergence. This indicates the extremely high CPU costs and infeasibility of such detailed schemes to be applied on the investigation of molecular complexation mechanisms.

For practical use, the general protonation trends are the results that matter the most, that is if the model suggests protonation for an amino acid that is actually found protonated in a given structure, and vice versa. This is indicated in the table by use of bold numbers for the cases where experimental and theoretical data have pKa shifts in opposite directions (e.g., $pK_{a,exp} - pK_0 > 0$ and $pK_{a,theoretical} - pK_0 < 0$, or the contrary). Comparing the theoretical methods, only one fault is noticed for FPTS and propKa while two are observed for the hybrid neMD−MC. This is useful to demonstrate that the FPTS is able to correctly predict the protonation states with chemical and biological significance.

Other protein systems were benchmarked before (Barroso da Silva and MacKernan 2017b) revealing that there is a slight tendency for the FPTS outcomes to be closer to the experimental measurements. From a comparison with 81 available points, (Barroso da Silva and MacKernan 2017b) the MAX, AAD, RMSD, and r are respectively, 3.4, 0.6, 0.9, and 0.68 for FPTS, and 4.9, 1.0, 1.4, and 0.51, for PropKa. Outcomes for RNA systems obtained by the FPTS are even slightly more accurate (Barroso da Silva et al. 2017a). Figure 7 shows calculated titration plots by the FPTS for RNA nucleotides. Data are for the domain 5 from *Azotobacter vinelandii* Intron 5 (Avd5) at 60 mM of salt. This is a more elongated molecule that increases the predictive difficulties of the FPTS. Nevertheless, the previous benchmark study confirmed that pK_a values calculated by FPTS give an AAA and MAX of 0.69 and 1.67 pH units in comparison to the experimental results (Pechlaner et al. 2015), which are virtually the same given by other theoretical methods. Moreover, the fast convergence properties of the FPTS is a real achievement that will make possible the studies of protein–RNA complexation mechanisms in several different experimental conditions due to its low CPU cost (without a significant loss of accuracy) (Barroso da Silva et al. 2017a).

Convergence properties of the FPTS is another specially positive feature of this titration scheme (Barroso da Silva et al. 2017a; Barroso da Silva and MacKernan 2017b). A typical production run for a single macromolecule at a given solution pH and salt concentration converges within 10^5 MC steps and takes ca. 10 s in a personal notebook (Intel i7-3630QM and 2.40 GHz – running ubuntu 12.04): (a) 1 s

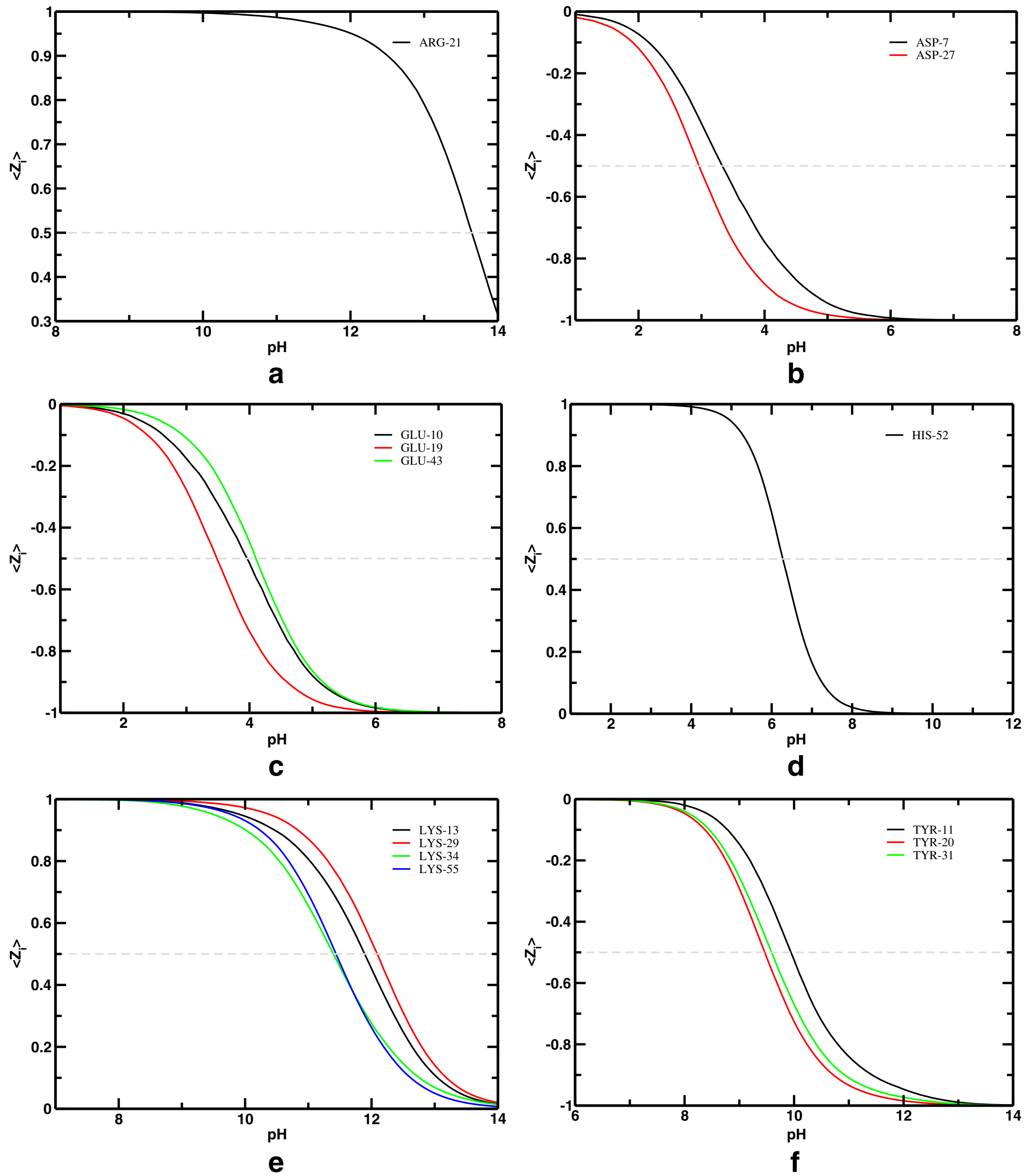

Fig. 6 Computed titration plots of the acid amino acid residues ARG **(a)**, ASP **(b)**, GLU **(c)**, HIS **(d)**, LYS **(e)**, and TYR **(f)** of turkey ovomucoid third domain at 10 mM salt concentration. The dashed gray lines indicate the half of the protonated states, which is used to predict the theoretical pK_a. Data are from the titration simulations with the FPTS (Teixeira et al. 2010). The intrinsic pK_0 values of the amino acid model compounds are 4.0, 4.4, 6.3, 9.6, 10.4, and 12.0, respectively, for ASP, GLU, HIS, TYR, LYS, and ARG (Nozaki and Tanford 1967). Simulation parameters are chosen as in Ref. Barroso da Silva and MacKernan (2017b).

for the lead-dependent ribozyme (PDB id 1LDZ), (b) 3 s for the thermostable actin binding 36-residue subdomain of chicken villin headpiece (PDB id 1VII), (c) 4 s for both the 45-residue binding domain of 2-oxoglutarate dehydrogenase multi-enzyme complex (PDB id 1W4H) and OMTKY3 (PDB id 1OMU), (d) 9 s for lysozymes (PDB ids 2LZT and 1AKI), (e) 10 s for the 124-residue ribonuclease A (PDB id 7RSA), (f) 12 s for the 122-residue α-lactalbumin (PDB id 1F6S), and (g) 13 s for the 135-residue staphylococcal nuclease (PDB ids 3D6C, 2RKS and 2SNM) (Barroso da Silva et al. 2017a; Barroso da Silva and MacKernan 2017b). For larger proteins, such as 6-phosphogluconate dehydrogenase (PDB id 2ZYG), the simulation time increases to 96 s. This performance is very fast when compared to other

Table 2 Calculated and experimental pK_a values of turkey ovomucoid third domain

Residue	Experiment[a]	hybrid neMD−MC[b]	PropKa	FPTS	NULL
Arg21			12.47	13.71(7)	12.0
His52			6.23	6.34(4)	6.3
Lys13			10.50	11.79(10)	10.4
Lys29			10.86	12.12(9)	10.4
Lys34			11.01	11.62(11)	10.4
Lys55			10.75	11.46(8)	10.4
Tyr11			10.13	10.12(13)	9.6
Tyr20			10.04	9.38(5)	9.6
Tyr31			10.81	9.57(7)	9.6
Asp7	2.7	3.43	3.43	3.34(7)	4.0
Asp27	2.3	**4.27**	3.69	3.15(10)	4.0
Glu10	4.1	4.04	**5.03**	3.98(6)	4.4
Glu19	3.2	3.53	4.14	3.38(9)	4.4
Glu43	4.8	**4.39**	4.64	**4.12(3)**	4.4
	MAX	1.97	1.35	0.85	1.7
	AAD	0.70	0.83	0.49	0.98
	RMSD	0.97	0.92	0.57	1.12
	r	0.44	0.86	0.98	0.82

Salt concentration is 10 mM. Simulation details for FPTS and PropKa are given in Ref. Barroso da Silva and MacKernan (2017b). The mean and standard deviations of the calculated FPTS pK_a values for turkey ovomucoid third domain were obtained from the results of all 50 NMR structures available in the PDB coordinates (PDB id 1OMU) as done in Ref. Tang et al. (2007). Only amino acids with available experimental data and predicted by the hybrid nonequilibrium molecular dynamics–Monte Carlo simulation method (Chen and Roux 2015) were used to calculate MAX, AAD, and RMSD[a] Experimental data from Ref. Schaller and Robertson (1995)[b] The theoretical data for the hybrid neMD−MC[b] was taken from Ref. Chen and Roux (2015) based on the averaged result for seven simulations

theoretical methods. For instance, using a PB solver, Wang and co-authors reported 9,185.2 s for the energy runtime in a single AMD Opteron 2356 processor (8 cores and 2.3 GHz) for the same system, (Wang et al. 2015) i.e., ca. 10^2 slower than the FPTS calculation.

The macromolecular flexibility in the simplified schemes To fully understand the function of biomolecules, it is necessary to consider both their structure and dynamics and their coupling with the protonation process. Macromolecular conformational changes are expected to happen due to

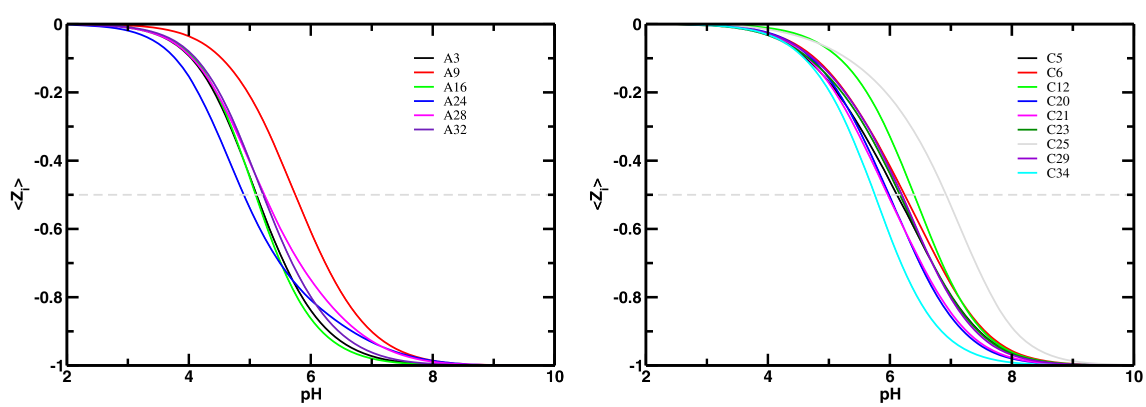

Fig. 7 Computed titration plots of the nucleotides adenosines (left panel) and cytosines (right panel) of the *A. vinelandii* domain 5 (AvD5) structure (PDB id 2m57) in 60 mM NaCl. The dashed gray lines indicate half of the protonated states, which is used to predict the theoretical pK_a. Data are from MC simulations with the FPTS. Simulation parameters are chosen as in Ref. Barroso da Silva et al. (2017a). The intrinsic pK_0 values of the nucleotides model compounds are 3.5 and 4.2, respectively, for A and C (Thaplyal and Bevilacqua 2014)

the protonation. In its turn, protonation is also affected by the change in the surrounded microenvironment after conformational changes. This dynamicaly and strongly coupled process is necessary for a complete understanding of the biomolecular processes. Nevertheless, as discussed so far, for computational reasons, it is often convenient to assume a static macromolecular structure.

The discussion on how conformational flexibility of proteins should be best accounted for started in the early days of the PB studies (You and Bashford 1995; Antonsiewicz et al. 1996; Alexov and Gunner 1997). A common agreement was that a single static conformation may be an inadequate representation of the strong coupling conformation-titration. Moreover, both the resolution and quality of the crystal structures, the main source of input for these simulations, could be questionable. Antosiewicz and co-authors pointed out that the average conformation of the protein in the crystal could be different from its solution behavior (You and Bashford 1995). Improvements in the computed pK_as were noticed when the conformational flexibility was used (You and Bashford 1995; Antonsiewicz et al. 1996; Alexov and Gunner 1997). Ensembles with all available low-energy NMR solution structures with equal weights have been used to overcome this problem, when possible (Tang et al. 2007). Alternatively, clustered structures obtained from classical MD trajectories (with fixed charges) are also used to *partially* explore the macromolecular conformational effects in rigid models (Barroso da Silva et al. 2016, 2017a). The standard protocol is to repeat the titration studies with the ensemble of clustered conformations and average out the results. These are all attempts to remedy the problem or at maximum to access the possible magnitude of the conformational effects on the ionization process. Any of these macromolecular ensembles *only* reflect the behavior of a single pH solution.

Other common constant pH simulation methods

A parallel route has being taken by other research groups after the introduction of MD simulations to simple liquids (Alder and Wainwright 1959; Verlet 1967) and later to biomolecules (Levitt and Lifson 1969; Karplus et al. 1977; Karplus and McCammon 1979; Berendsen et al. 1981). In Fig. 3, these theoretical methods that can explore macromolecular conformations were presented as "flexible models in explicit solvent". They started with conventional MD simulations where solution pH only entered the model at the initial setup when the user makes a choice between neutral or protonated amino acids (or nucleotides) to assign atomistic partial charges for the titratable sites as a function of pH. This approach is represented by the left picture at Fig. 3, middle row. Any of the previously cited rigid titration models presented above could be used to assign these charges being the Poisson–Boltzmann (PB) solvers (Bashford 1997; Baker et al. 2001; Anandakrishnan et al. 2012) or the empirical methods (Li et al. 2005; Olsson et al. 2011; Krieger et al. 2006) more frequently used for this purpose. During the simulation run, these atomistic partial charges are kept unchanged, ignoring the possible transformations in their microenvironment (exposure of the titratable side chains to water, interactions with other titratable groups or any charged species, salt and free counter-ions in the solution) that can induce alterations in their protonation states, and, reciprocally, could also cause the macromolecule chain to adopt a different conformation.

Such strong protonation–conformation coupling started to be better described only much later through the combination of classical MD simulations with protonation numerical schemes (Baptista et al. 1997). This was the beginning of a new era of CpH simulation methods still under intense development for biological systems (Baptista et al. 1997; Wallace and Shen 2012; Dashti et al. 2012; Goh et al. 2013a; Chen et al. 2013, 2014; Chen and Roux 2015; Socher and Stich 2016; Donnini et al. 2016). Several new simulation schemes that allow pH-coupled MD have emerged following either the Baptista's hybrid MD/PB approach (Baptista et al. 1997) or are rooted in the "λ"'s dynamics method (Kong and Brooks 1996). Very often there are merely subtle technical details to differ among the various available methods found in the literature.

The Baptista hybrid MD/PB CpH scheme is for the titration part based on a continuum modeling of the solvent and a mean-field description of the electrolyte solution given by the linear PB equation. Periodically, instantaneous macromolecular coordinates generated by the MD run (in an explicit solvent model) are passed to a PB solver (MEAD (Bashford 1988)) that will update the protonation states (in an implicit solvent model) and return the new partial atomistic charges to the MD engine. These new charged states should somehow access the changes in the local microenvironment. Although this periodic switch on/off in the protonation states may introduce discontinuities in forces, this pH-coupled MD scheme is by far the faster strategy. Later developments improved more the initial ideas (Baptista et al. 2002; Machuqueiro and Baptista 2007). The latest version of the method is implemented in the GROMACS simulation package (Machuqueiro and Baptista 2007). Benchmarking pK_as for lysozyme (PDB id 4LZT) resulted in good values for RMSDs [0.70 (for the GROMOS 43A1 force field) and 0.79 (for the GROMOS 53A6 force field)]. These calculations could also demonstrate the dependence of the force fields on the computed pK_as. As observed in many other calculations (Barroso da Silva and MacKernan 2017b), the proton donor amino acid Glu-35 in the catalytic site of lysozyme is a case that is very difficult to be properly described (Machuqueiro and Baptista 2011). The results obtained for their studies in a diversity of biomolecular systems reveal the vast possibilities of the method. Examples include investigation on the pH-dependent conformational states of the analgesic dipeptide kyotorphin (L-Tyr-L-Arg) (Machuqueiro and Baptista 2007), the possibility of a trigged pH action on the misfolding of the prion protein into a pathogenic β-rich form (Campos et al. 2010), the pH effects on the reversibility of prion misfolding (Vila-Viçosa et al. 2012), a pH titration of all constituent lipids of a 25% DMPA/DMPC bilayer membrane model (Santos et al. 2015), and the tight coupling between protonation and conformation for cytochrome c oxidase (Oliveira et al. 2016).

In 2004, Mongan, Case, and McCammon proposed a CpH MD in Generalized Born (GB) implicit solvent (Mongan et al. 2004). This is one method that was implemented with AMBER 9 (Case et al. 2006) and contributes to popularize the use of CpH methods among simulation users. Their pK_a predictions for lysozyme structures (PDB ids 1AKI, 1LSA, 3LZT, and 4LYT) resulted in a small RMSD of 0.82 (average for all the four structures) relative to experimental values. These different structures were chosen for maximum diversity of crystal characteristics to access the effect of the protein conformation on the calculated pK_as. The outcomes (RMSDs equals to 0.86, 0.77, 0.88, and 0.95, respectively, for PDB ids 1AKI, 1LSA, 3LZT, and 4LYT) confirm the dependence on the conformation and the need of a dynamics-based method (Mongan et al. 2004). Latest uses for the method comprehend the investigation of the conformational characteristics of the molten globule state of human α-lactalbumin (Bhattacharjee et al. 2013) and the study of HIV protease flaps dynamics in different pHs (Soares et al. 2016).

The "λ"'s dynamics started with the work developed by Brooks and co-authors (Kong and Brooks 1996; Lee et al. 2004). The idea is that the dynamics of an artificial titration coordinate "λ" should be given by forces between the protonated and deprotonated states. Various versions of this method are available including a tautomeric state titration model (tstm) that allows simultaneous titration at two competing titratable sites (Khandogin and Brooks 2005). The price to pay in this class of methods is the slow convergence. Attempts to deal with the high computational costs started with the replacement of the PB description by closely related analytical theories such as the generalized Born theory Lee et al. (2004). Nevertheless, the use of implicit solvent models did not help too much to overcome this difficulty. It takes ca. 500 ps to achieve pK_a convergence (Chen et al. 2014). Characteristics AADs are found around 0.6–1.0 pK_a units. For instance, computed pK_a for OMTKY3 and ribonuclease A – RNaseA (PDB id 7RSA) yielded AADs equal to 1.0 and 0.6, respectively (Khandogin and Brooks 2005). These numbers are not as good as the ones obtained by much simpler and faster models as the ones using static protein structures. PB data for the same systems and experimental conditions reported AADs of 0.6 (Forsyth et al. 1998) and 0.8, (Antonsiewicz et al. 1996) respectively, for OMTKY3 and RNaseA ($\epsilon_p = 20$). PropKa gives better predictions for OMTKY3 ($AAD_{OMTKY3} = 0.83$) (Barroso da Silva and MacKernan 2017b). Conversely, the best results for these two protein systems were obtained by the FPTS ($AAD_{OMTKY3} = 0.57$ and $AAD_{RNaseA} = 0.4$) (Barroso da Silva and MacKernan 2017b). Moreover, the tstm result is almost identical (and slightly inferior) to the value obtained by the NULL model ($AAD_{OMTKY3} = 0.98$ (Barroso da Silva and MacKernan 2017b)) that has virtually no CPU costs.

With the increase of computer power, it became doable to replace the implicit solvent model by full atomistic representations (Wallace and Shen 2012; Dashti et al. 2012; Chen et al. 2013, 2014; Goh et al. 2013a, b; Lee et al. 2015, 2016; Donnini et al. 2016). Brooks's lab was the first to extend the CpH MD to nucleic acids in explicit TIP3P water molecules (Goh et al. 2013a, b). From this point on, the hydrophobic effects and the dielectric response of the medium could in principle be better described by explicit water models. Another advantage was a more detailed understanding of proton translocation between the macromolecular titratable sites and the solvent. Processes such as the solvent-mediated proton transfer in close compartments could be fully described. Works carried out by Wallace and Shen (2011) and Swails et al. (2014) demonstrated that the RMSD can decrease, respectively, from 0.93 to 0.84 and 1.32 to 0.92 when replacing the implicit solvent by an explicit description. The study was done with hen egg white lysozyme (PDB ids were 2LZT and 3LZT, respectively). However, although this new class of models incorporated more realism, the outcomes do not show real significant progress. With more details in the model, the sampling difficulties increased. Much more computationally expensive simulations are necessary for convergence (10 ns as reported by Shen and collaborators, (Chen et al. 2013) or 40 ns for a simple dipeptide as quantified by Chen and Roux (2015)). Outcomes are not worse because the computers are faster and new advanced sampling techniques are employed. The combination of limited conformational and titration sampling may be why more empirical methods such as PROPKA (at negligible CPU time) or simplified rigid models such as FPTS (orders of magnitude faster) seem to obtain similar outcomes for pK_a predictions. This slow convergence might also indicate an insufficient modeling of the charge fluctuations, which would affect the proper description of all molecular mechanisms responsible for the macromolecular complexation.

Other research groups are working in this class of methods. For example, Grubmüller and collaborators have one version in explicit solvent implemented in GROMACS (Donnini et al. 2011). In a further study, a three-states model was suggested for an accurate description of chemically coupled titrating sites (Dobrev et al. 2017). These authors also introduced a kind of "hydronium ion" at the solvent. However, the accuracy of the method was tested only against titration curves of single amino acids and a dipeptide. A benchmark study with sets of proteins and nucleic acids remains to be done. Moreover, since the hydronium ion is created by introducing an extra charge on the conventional SPC water model (a water model whose parametrization is known to be quite sensitive to the assigned partial charge), solvent properties might become less reliable. The complexity of the recombination of hydronium and hydroxide ions in water can be seen in Ref. Hassanali et al. (2011).

A recent step to improve further the physical realism was the introduction of a titratable water model in the pH-coupled MD by Shen and collaborators (Chen et al. 2013). This type of model was applied to predict the proton titration in cationic micelle and bilayer environments (Eike et al. 2014). It is not clear how this approach affects the solvent structure and dynamics. Any artifact can also have an effect, perhaps in the wrong direction, on the macromolecular conformation, the diffusion of mobile charged species (added salt and counter-ions), and all their interplay. In terms of its predictions, results are similar to the ones obtained by other theoretical methods at much higher CPU costs.

Being an area of intense research activity, many laboratories have also contributed to the development of other coexisting methods. The differences between the ones already presented are often seen in small technical details. For the sake of completeness, we should cite the work with coarse-grained models done by Delle Site (for peptides) (Enciso et al. 2013), the Donnini's version (Donnini et al. 2016) of the "λ"'s dynamics for the MARTINI force field applied to study oleic acid aggregates (Bennett et al. 2013), and the initiatives with the empirical "λ" dynamics method of Börjesson & Hünenberger (for amines) (Börjesson and Hünenberger 2001; Baptista 2002), and the classical MD method coupled with quantum mechanically derived proton hopping (Q-HOP) method of Lill & Helms (applied on small molecules and protein) (Lill and Helms 2001; De Groot et al. 2003; Gu et al. 2007).

Comparison between the different theoretical methods

These different classical CpH techniques mostly differ in the way the macromolecule (atomistic level versus all possible coarse-grained descriptions), solvent (explicit or continuum solvent model) and salt particles (explicit or DH treatments) are modeled together with the method used to include and modify the protonation states. The choice of the ideal CpH method depends on the characteristics of the studied system together with the usual compromise between *(a) the property or quantity of interest, (b) the required accuracy, (c) the number of systems and/or different experimental conditions to be simultaneously investigated, and (d) the available computing power* (van Gunsteren and Berendsen 1990). As observed for a few examples given above, the inclusion of more details in the computer model does not guarantee better predictions and the CPU time can be prohibitive. In reality, a detailed model can result in poor predictions due to their slow convergence and poor sampling. In Table 3, we compared the main classes of theoretical methods available today. This might offer to the reader some updated practical guide to choose among the options based on the present discussions. An old and less detailed comparison was published before (Chen et al. 2014). Note that a special chapter for pK_a calculations is related with membrane proteins (not covered in this review).

In general, predictions by different techniques are relatively similar to the others as already pointed out by an early benchmark study with a large set of biomolecules (Stanton and Houk 2008). Xiao & Yu showed that even QM/MM methods can have comparable results with PropKa and PB solvers (Xiao and Yu 2016). There are some general trends that can be noticed. For pK_a predictions in the absence of any additional external potential (i.e., for only a single protein in an electrolyte solution), PropKa is the faster method. Results are in general even more precise than popular PB solvers (Davies et al. 2006). With the present computer power, CpH MD methods with λ dynamics seem suitable only for small molecules. This picture might be different in the near future due to the intense efforts to solve the sampling issues (Williams et al. 2010; Chen and Roux 2015; Radak and Roux 2016; Socher and Stich 2016; Donnini et al. 2016; Chen et al. 2016). Macromolecules with several titratable groups might be better simulated today with schemes like Baptista's approach (Baptista et al. 1997, 2002). This class provides slightly faster sampling and could better access the protonation–conformation coupling. Buried titratable groups might require an explicit solvent and hybrid approaches. The Stern's hybrid method that was applied as a proof of concept to an acetic acid in aqueous solution with an explicit representation of water molecules shows this trend (Stern 2007). For complexation studies where more than one macromolecule is present, all these more sophisticated techniques will suffer from the slow convergence (Chen et al. 2016). The interplay of so many titratable sites of several ionizable objects will slow down even more the already difficult sampling. The best

Table 3 Comparison between the different classes of theoretical methods

Method class	CPU costs	Pros	Cons
Empirical[a]	very low	• fast • easy to use • good accuracy	• it does not provide charge fluctuations • it cannot respond to external electrical fields • protein is treated as a rigid body
Poisson-Boltzmann[b]	low to medium	• simple to use • accuracy can be tuned by means of the use of nonuniform dielectrics	• it does not provide charge fluctuations • protein is treated as a rigid body • ion-ion correlations are neglected • high memory consuming
Monte Carlo schemes[c]	very low (for the FPTS) medium (in general)	• fast (for the FPTS) • good accuracy • ion-ion correlations are taken into account (for the explicit ions version) • very suitable for protein complexation applications	• protein is treated as a rigid body • it does not provide dynamical properties • not suitable for buried titratable groups
Molecular dynamics coupled with titration[d]	medium to high	• protonation-conformation coupling is included	• possible simulation time might be not enough to fully describe this coupling • the switch on/off in the protonation state my result in in conformational and energetic instabilities
Molecular dynamics with λ dynamics and other methods[e]	high	• protonation-conformation coupling is included • possibility to include titratable water models techniques	• slow convergence is an issue • still prohibitive for protein complexation applications • needs advanced sampling
Ab initio MD[f]	very high	• it does not need parametrization of the intermolecular and intramolecular potentials • the protonation-conformation coupling is naturally included	• very slow convergence • only suitable for amino acids and few water molecules (< 100) • requires large computational resources and advanced sampling techniques

[a]See Li et al. (2005), Krieger et al. (2006), Burger and Ayers (2011), and Olsson et al. (2011)
[b]See Bashford (1997), Baker et al. (2001), Anandakrishnan et al. (2012), Wang et al. (2015), and Sakalli and Knapp (2015)
[c]See Svensson et al. (1990), Kesvatera et al. (1996, 1999, 2001), Teixeira et al. (2010), Carnal et al. (2015), Barroso da Silva et al. (2017a), and Barroso da Silva and MacKernan (2017b)
[d]See Baptista et al. (1997, 2002), Baptista and Soares (2001), Machuqueiro and Baptista (2007), and Santos et al. (2015)
[e]See Lee et al. (2004), Wallace and Shen (2009), Dashti et al. (2012), Goh et al. (2013a), Chen et al. (2013, 2014, 2016), Chen and Roux (2015), Socher and Stich (2016), and Donnini et al. (2016)
[f]See Tummanapelli and Vasudevan (2015), Kamerlin et al. (2009), and Li et al. (2002)

alternative in this case is the MC titration schemes, particularly the FPTS. From this mesoscopic scheme, other intermediate models can also be derived in order to improve accuracy for specific tasks at higher CPU expenses.

Simplified models applications in biomolecular systems

Protein association introduces a next level of difficulty for constant-pH simulation methods due to the increase in the number of interacting titratable objects, the coupling between them, and the multi-macromolecular conformational changes. The driving force for macromolecular complexation is often charge–charge interactions, charge–dipole interactions, dipole–dipole interactions, and van der Waals interactions. Changes in the hydration may also play an important role. Less emphasized is the importance of mesoscopic electrostatic attraction forces resulting from proton fluctuations (Kirkwood and Shumaker 1952). This attraction is a result of the mutual rearrangements of the distributions of the charged groups due to the acid-base equilibrium as analytically predicted by the KS theory (Kirkwood and Shumaker 1952; Lund and Jönsson 2013; Barroso da Silva 2013). Such phenomena can only be properly described in a constant-pH simulation that has converged. This starts to place more constraints for the model choice for macromolecular complexation studies. On top of that, the need to explore a vast number of possible orientations and separation distances between the pairs of molecules to estimate the interaction free energy requires simplified models. Often, it is also necessary to repeat the calculations on a great number of different experimental conditions (e.g., different pHs, ionic strengths, macromolecular concentration, mutations, etc.).

This scenario is far from complicate for CpH MD approaches in explicit solvent. These methods can still not reach the desired scales to probe complexation mechanisms at so many conditions in computer simulations. The CpH MC schemes in implicit solvents discussed above meet well all these requirements. They have been intensively and successfully applied in several biomolecular systems: (a) protein–protein interactions (Lund and Jönsson 2003, 2005; Jönsson et al. 2007; Persson et al. 2010; Kurut et al. 2012, 2015; Delboni and Barroso da Silva 2016; Barroso da Silva et al. 2016), (b) protein–polyelectrolyte interactions (Barroso da Silva et al. 2006; Jönsson et al. 2007; Barroso da Silva and Jönsson 2009; Barroso da Silva 2013; Srivastava et al. 2017), (c) protein–peptide interactions (André et al. 2004; Jönsson et al. 2007), (d) protein–surface interactions (Nylander et al. 2017; Hyltegren and Skepö 2017), and (e) protein–nanoparticle interactions (Barroso da Silva et al. 2014; Carnal et al. 2015). There is also ongoing work on protein–RNA interactions at our laboratory together with Profs. Pasquali and Derreumaux as an application of the new RNA titration scheme (Barroso da Silva et al. 2017a, c).

Based on these studies, different driven forces were identified for biomolecular systems. For instance, Coulomb charge–charge interactions dominate the association of the whey proteins α-LA–β-LG, α-LA–LF and β-LG–LF (Delboni and Barroso da Silva 2016). Another example is the self-association of LF, which is also driven by a high charge complementarity across the contact surface of the proteins (Persson et al. 2010). Conversely, in a process trigged by pH, multipolar interactions drive the self-association of spidroins (Barroso da Silva et al. 2016).

Simplified CG models were used as well to demonstrate the significance of the charge regulation mechanism on complexation mechanisms (Barroso da Silva et al. 2006, 2017c; Jönsson et al. 2007; Barroso da Silva and Jönsson 2009; Lund and Jönsson 2013; Barroso da Silva 2013). An important lesson learned from these sets of biomolecular applications is that the simply use of fixed charges assigned at the beginning of the simulations as a function of pH does *not* let the complete description of all electrostatic mechanisms. The analysis of different criteria to assign partial charges for LYZ in a protein–polyelectrolyte complexation study without doubts indicates that charge fluctuations due to the acid-base equilibrium are a must to fully explore all physical mechanisms (Barroso da Silva et al. 2006). This can only be done by means of well-converged CpH simulations. Such an observation for LYZ might indicate a promising way to shed light on the understanding of the apparent paradox involved in the LYZ self-association (Shukla et al. 2008).

Conclusions

The development of constant-pH simulation methods is a formidable scientific problem, a quite hectic and key challenging research field. Different theoretical models are already available and routinely used to study the biomolecular phenomena. There is yet no perfect model for all desired applications. Good models are strongly dependent on the studied system, desired accuracy, number of different experimental conditions to be studied and compared, and accessible computational resources. The lack of use of a CpH technique can lead to an uncompleted physical description.

Toward accurate prediction of the protonation equilibrium of biomolecules, two directions are currently being largely explored. In the first one, efforts are dedicated to improve the accuracy of computed pKas by means of both more detailed models coupling ionization and conformational changes and developing new enhanced sampling techniques. In the other direction, assuming that the most important model feature is to let the biological process be rationalized (i.e., the proper prediction of the protonation state under a given set of experimental conditions), fast coarse-grained models that can well describe the pH effects on the large-scale scenario for systems with several macromolecules are a real need.

A new class of simplified Monte Carlo schemes has emerged during the last years. Surprisingly, the outcomes were equivalent or even better than more sophisticated methods that are more computationally costly. Such promising results indicate that schemes like the FPTS can contribute in both directions. On one side, the FPTS can be refined to improve pK_a predictions. On another side, the FPTS is robust enough to be applied on the multi-titrating objects, each containing several ionizable sites. Ongoing studies at our laboratory associated with other collaborators indicate that the FPTS can be both refined and coupled with MD engines making such scheme a powerful tool for studying molecular mechanisms that govern a wide variety of important biological processes.

Acknowledgements This work was supported in part by the *Fundação de Amparo à Pesquisa do Estado de São Paulo* [Fapesp 2015/16116-3 (FLBDS), Fapesp 2013/08166-5 and 2017/03204-7 (LGD)], the University Global Partnership Network (UGPN) Research Collaboration Fund (FLBDS) and the University College Dublin (UCD) through a visiting professors grant - Seed Funding (FLBDS). FLBDS also thanks the support of the University of São Paulo through the NAP-CatSinQ (Research Core in Catalysis and Chemical Synthesis), the computing hours at Rice University through the international collaboration program with USP and at The Swedish National Infrastructure for Computing (SNIC 2015/10-6), and the hospitality of the UCD School of Physics, UCD Institute for Discovery and CECAM/IRL. It is also a pleasure to acknowledge the collaboration in previous works and/or fruitful discussions with Bo Jönsson, Bo Svensson, Torbjörn Åkesson (in memoriam), Mikael Lund, Aatto Laaksonen, Erik Santiso, Samuela Pasquali, Philippe Derreumaux, Catherine Etchebest, and Donal MacKernan.

Compliance with ethical standards

Conflict of interests Fernando Luís Barroso da Silva declares that he has no conflicts of interest. Luis Gustavo Dias declares that he has no conflicts of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

Alder BJ, Wainwright TE (1959) Studies in molecular dynamics. I. General method. J Chem Phys 31(2):459–466

Alexov EG, Gunner MR (1997) Incorporating protein conformational flexibility into the calculation of pH-dependent protein properties. Biophys J 74:2075–2093

Alexov E, Mehler EL, Baker N, Baptista A, Huang Y, Milletti F, Nielsen JE, Farrell D, Carstensen T, Olsson MHM, Shen JK, Warwicker J, Williams S, Word JM (2011) Progress in the prediction of pKa values in proteins. Proteins 79(12):3260–3275

Allen MP, Tildesley DJ (1989) Computer Simulation of Liquids. Oxford University Press, Oxford

Anandakrishnan R, Aguilar B, Onufriev AV (2012) H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulation. Nucleic Acids Res 40(W1):E537—541

Andermatt S, Cha J, Schiffmann F, VandeVondele J (2016) Combining linear-scaling DFT with subsystem DFT in Born-Oppenheimer and Ehrenfest molecular dynamics simulations: From molecules to a virus in solution. J Chem Theory Comput 12:3214–3227

André I, Kesvatera T, Jönsson B, Åerfeldt KS, Linse S (2004) The role of electrostatic interactions in calmodulin-peptide complex formation. Biophys J 87:1929–1938

Antonsiewicz J, McCammon JA, Gilson MK (1994) Prediction of pH-dependence properties of proteins. J Mol Biol 238:415–436

Antonsiewicz J, McCammon JA, Gilson MK (1996) The determinants of pK_as in proteins. Biochemistry 35:7819–7833

Archontis G, Simonson T (2005) Proton binding to proteins: A free-energy component analysis using a dielectric continuum model. Biophys J 88:3888–3904

Atkins PW (1995) Physical Chemistry, 5th edn. Oxford University Press, London

Autreto PAS, Figueiredo FV, Nonato MC, Barroso da Silva FL (2003) Application of the Poisson–Boltzmann approach on structural biology: An initial study of the complex trypsin-BPTI. Braz J Pharm Sci Supl 2(39):203

Bacquet RJ, McCammon JA, Allison SA (1988) Ionic strength dependence of enzyme-substrate interactions. Monte Carlo and Poisson–Boltzmann results for the superoxide dismutase. J Phys Chem 92(25):7134–7141

Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. Proc Natl Acad Sci USA 98:10,037–10,041

Baptista AM (2002) Comment on ”explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. J Chem Phys 116(17):7766–7768. J Chem. Phys. 114, 9706 (2001)

Baptista AM, Soares CM (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

Baptista AM, Marte PJ, Petersen SB (1997) Simulation of protein conformational freedom as a function of pH: Constant-pH molecular dynamics using implicit titration. Proteins: Struc Func, and Genetics 27:523–544

Baptista AM, Teixeira VH, Soares CM (2002) Constant-pH molecular dynamics using stochastic titration. J Chem Phys 117(9):293–309

Barroso da Silva FL (1999) Statistical Mechanical Studies of Aqueous solutions and Biomolecular Systems. Reproenheten SLU Alnarp, Lund University, Sweden

Barroso da Silva FL (2013) Peculiaridades nos mecanismos moleculares de proteínas em solução aquosa: Exemplo da importância do equilíbrio ácido-base para aplicações em biotecnologia. Química 131(oct-dez):43–48

Barroso da Silva FL, Jönsson B (2009) Polyelectrolyte–protein complexation driven by charge regulation. Soft Matter 5(15):2862–2868

Barroso da Silva FL, Jönsson B, Penfold R (2001) A critical investigation of the Tanford-Kirkwood scheme by means of Monte Carlo simulations. Prot Sci 10:1415–1425

Barroso da Silva FL, Bogren D, Söderman O, Jönsson B (2002) Titration of fatty acids solubilized in cationic, nonionic and anionic micelles: Theory and experiment. J Phys Chem B 106:3515–3522

Barroso da Silva FL, Linse S, Jönsson B (2005) Binding of charged ligands to macromolecules. Anomalous salt dependence. J Phys Chem B 109:2007–2013

Barroso da Silva FL, Lund M, Jönsson B, Åkesson T (2006) On the complexation of proteins and polyelectrolytes. J Phys Chem B 110:4459–4464

Barroso da Silva FL, Boström M, Persson C (2014) Effect of charge regulation and ion–dipole interactions on the selectivity of protein–nanoparticle binding. Langmuir 30(14):4078–4083

Barroso da Silva FL, Pasquali S, Derreumaux P, Dias LG (2016) Electrostatics analysis of the mutational and pH effects of the n-terminal domain self-association of the major ampullate spidroin. Soft Matter 12:5600–5612

Barroso da Silva FL, Derreumaux P, Pasquali S (2017a) Fast coarse-grained model for RNA titration. J Chem Phys 146(3):035,101+

Barroso da Silva FL, MacKernan D (2017b) Benchmarking a fast proton titration scheme in implicit solvent for biomolecular simulations. J Chem Theory Comput 13(6):2915–2929

Barroso da Silva FL, Derreumaux P, Pasquali S (2017c) Protein-RNA complexation driven by the charge regulation mechanism, Biochem. Biophys. Res. Commun., in press

Bartik K, Redfield C, Dobson CM (1994) Measurement of the individual pKa values of acidic residues of hen and turkey lysozymes by two-dimensional 1H NMR. Biophys J 66(4):145

Bashford D (1988) An object-oriented programming suite for electrostatic effects in biological molecules. an experience report on the mead project. ISCOPE meeting

Bashford D (1997) An object-oriented programming suite for electrostatic effects in biological molecules. An experience report on the MEAD project. Springer Berlin Heidelberg, Berlin, pp 233–240

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

Bashford D, Karplus M (1990) pKa's of ionizable groups in proteins: Atomic detail from a continuum electrostatic model. Biochemistry 29:10,219–10,225

Bashford D, Karplus M, Canters GW (1988) Electrostatic effects of charge perturbations introduced by metal oxidation in proteins—a theoretical analysis. J Mol Biol 203:507–510

Becconi O, Ahlstrand E, Salis A, Friedman R (2017) Protein–ion interactions: Simulations of bovine serum albumin in physiological solutions of NaCl, KCl and LiCl. Isr J Chem 57(5):403–412

Bell RP (1959) The Proton in Chemistry. Cornell University Press, New York

Bennett WD, Chen AW, Donnini S, Groenhof G, Tieleman DP (2013) Constant pH simulations with the coarse-grained martini model—application to oleic acid aggregates. Can J Chem 91:839–846

Berendsen HJC, Postma JPM, van Gunsteren WF, Hermans J (1981) Interactions models for water in relation to protein hydration. In: Pullman B (ed) Intermolecular Forces. Reidel, Dordrecht, pp 331–342

Beresford-Smith B, Chan DYC (1983) Electrical double-layer interactions in concentrated colloidal systems. Faraday Disc Chem Soc 76:65–75

Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242

Beroza P, Fredkin DR, Okamura MY, 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 USA 88:5804–5808

Bhattacharjee N, Rani P, Biswas P (2013) Capturing molten globule state of α-lactalbumin through constant pH molecular dynamics simulations. J Chem Phys 138:095,101

Börjesson U, Hünenberger PH (2001) Explicit-solvent molecular dynamics simulation at constant pH: Methodology and application to small amines. J Chem Phys 114:9706–9719

Boese A, Doltsinis N, Handy N, Sprik M (2000) New generalized gradient approximation functionals. J Chem Phys 112:1670–1678

Borkovec M, Jönsson B, Koper G (2001) Ionization processes and proton binding in polyprotic systems: Small molecules, proteins, interfaces, and polyelectrolytes. In: Matijević E (ed) Surface and Colloid Science, Surface and Colloid Science, vol 16, Springer US, pp 99–339

Böttcher CJF (1973) Theory of Electric Polarization. Elsevier, Amsterdam

Brémond E, Ciofini I, Sancho-García J, Adamo C (2016) Nonempirical double-hybrid functionals: An effective tool for chemists. Acc Chem Res 49:1503–1513

Burger SK, Ayers PW (2011) A parameterized, continuum electrostatic model for predicting protein pKa values. Proteins 79:2044–2052

Calixto TMR (2010) Análises de propriedades eletrostáticas e estruturais de complexos de proteínas para o desenvolvimento de preditores de complexação em larga escala, Master thesis, University of So Paulo, Ribeirão Preto, SP

Campos SRR, Machuqueiro M, Baptista AM (2010) Constant-pH molecular dynamics simulations reveal a β-rich form of the human prion protein. J Phys Chem B 114(39):12,692–12,700

Capelle K (2006) A bird's-eye view of density-functional theory. Braz J Phys 36:6378–6396

Carlsson F, Linse P, Malmsten M (2001a) Monte Carlo simulations of polyelectrolyte–protein complexation. J Phys Chem B 105:9040–9049

Carlsson F, Malmsten M, Linse P (2001b) Monte Carlo simulations of lysozyme self-association in aqueous solution. J Phys Chem B 105:12,189–12,195

Carlsson F, Hyltner E, Arnebrant T, Malmsten M, Linse P (2004) Lysozyme adsorption to charged surfaces. A Monte Carlo study. J Phys Chem B 108:9871–9881

Carnal F, Claviera A, Stoll S (2015) Modelling the interaction processes between nanoparticles and biomacromolecules of variable hydrophobicity: Monte Carlo simulations. Environ Sci: Nano 2:327–339

Carstensen T, Farrell D, Huang Y, Baker NA, Nielsen JE (2011) On the development of protein pka calculation algorithms. Proteins 79(12):3287–3298

Casasnovas R, Ortega-Castro J, Frau J, Donoso J, Munoz F (2014) Theoretical pka calculations with continuum model solvents, alternative protocols to thermodynamic cycles. Int J Quantum Chem 114:1350–1363

Case D, Darden T, Cheatham T III, Simmerling C, Wang J, Duke R, Luo R, Merz K, Pearlman D, Crowley M, Walker R, Zhang W, Wang B, Hayik S, Roitberg A, Seabra G, Wong K, Paesani F, Wu X, Brozell S, Tsui V, Gohlke H, Yang L, Tan C, Mongan J, Hornak V, Cui G, Beroza P, Mathews D, Schafmeister C, Ross W, Kollman P (2006) Amber 9. University of California, San Francisco

Chen Y, Roux B (2015) Constant-pH hybrid nonequilibrium molecular dynamics Monte Carlo simulation method. J Chem Theory Comput 11:3919–3931

Chen J, Brooks CL III, Khandogin J (2008) Recent advances in implicit solvent-based methods for biomolecular simulations. Curr Opin Struct Biol 18:140–148

Chen K, Xu Y, Rana S, Miranda OR, Dubin PL, Rotello VM, Sun L, Guo X (2011) Electrostatic selectivity in protein–nanoparticle interactions. Biomacromolecules 12(7):2552–2561

Chen W, Wallace JA, Yue Z, Shen JK (2013) Introducing titratable water to all-atom molecular dynamics at constant pH. Biophys J 105:L15—L17

Chen W, Morrow BH, Shi C, Shen JK (2014) Recent development and application of constant pH molecular dynamics. Mol Sim 40:830–838

Chen W, Huang Y, Shen JK (2016) Conformational activation of a transmembrane proton channel from constant pH molecular dynamics. J Phys Chem Lett 7(19):3961–3966

Creighton TE (1983) Proteins—Structures and Molecular Principles. W. E. Freeman and Company, New York

Dashti DS, Meng Y, Roitberg AE (2012) pH-Replica exchange molecular dynamics in proteins using a discrete protonation method. J Phys Chem B 116:8805–8811

Davies MN, Toseland CP, Moss DS, Flower DR (2006) Benchmarking pKa prediction. BMC Biochemistry 7(1):1–12

Davis ME, McCammon JA (1990) Electrostatics in biomolecular structure and dynamics. Chem Rev 90:509–521

Davis ME, Madura JD, Luty BA, McCammon JA (1991) Electrostatics and diffusion of molecules in solution. Simulations with the University-of-Houston-Brownian dynamics program. Comp Phys Commun 62:187–197

de Carvalho SJ, Ghiotto RT, Barroso da Silva FL (2006) Monte Carlo and modified Tanford–Kirkwood results for macromolecular electrostatics calculations. J Phys Chem B 110:8832–8839

de Carvalho SJ, Fenley MO, Barroso da Silva FL (2008) Protein–ion binding process on finite macromolecular concentration. A Poisson–Boltzmann and Monte Carlo study. J Phys Chem B 112(51):16,766–16,776

Degrève L, Barroso da Silva FL (1999a) Large ionic clusters in concentrated aqueous NaCl solution. J Chem Phys 111:5150–5156

Degrève L, Barroso da Silva FL (1999b) Structure of concentrated aqueous NaCl solution: a Monte Carlo study. J Chem Phys 110(6):3070–3078

Degrève L, Barroso da Silva FL (2000) Detailed microscopic study of 1M aqueous NaCl solution by computer simulations. J Mol Liquids 87:217–232

Degrève L, Lozada-Cassou M, Sánchez E, González-Tovar E (1993) Monte Carlo simulation for a symmetrical electrolyte next to a charged spherical colloid particle. J Chem Phys 98(11):8905–8909

De Groot BL, Frigato T, Helms V, Grubmuller H (2003) The mechanism of proton exclusion in the aquaporin-1 water channel. J Mol Biology 333(2):279–293

Delboni L, Barroso da Silva FL (2016) On the complexation of whey proteins. Food Hydrocolloids 55:89–99

Demchuk E, Wade RC (1996) Improving the continuum dielectric approach to calculating pK_as of ionizable groups in proteins. J Phys Chem B 100:17,373–17,387

Derjaguin BV, Landau L (1941) Acta Phys Chim URSS 14:633

Devlin TM (ed) (1997) Textbook of Biochemistry with Clinical Correlations. Wiley-Liss, New York

Dobrev P, Donnini S, Groenhof G, Grubmüller H (2017) Accurate three states model for amino acids with two chemically coupled titrating sites in explicit solvent atomistic constant ph simulations and pKa calculations. J Chem Theory Comput 13:147–160

Donnini S, Tegeler F, Groenhof G, Grubmüller H (2011) Constant pH simulations with the coarse-grained martini model—application to oleic acid aggregates. J Chem Theory Comput 7:1962–1978

Donnini S, Ullmann RT, Groenhof G, Grubmüller H (2016) Charge-neutral constant pH molecular dynamics simulations using a parsimonious proton buffer. J Chem Theory Comput 12:1040–1051

Dudev T, Lim C (2000) Metal binding in proteins: The effect of the dielectric medium. J Phys Chem B 104:3692–3694

Egan T, O'Riordan D, O'Sullivan M, Jacquier JC (2014) Cold-set whey protein microgels as pH modulated immobilisation matrices for charged bioactives. Food Chem 156:197–203

Eike BHMDM, Murch BP, Koenig PH, Shen JK (2014) Predicting proton titration in cationic micelle and bilayer environments. J Chem Phys 141:084,714

Enciso M, Schutte C, Site LD (2013) A pH-dependent coarse-grained model for peptides. Soft Matter 9:6118–6127

Evans DF, Wennerström H (1994) The Colloidal Domain. VCH Publishers, New York

Fennell CJ, Li L, Dill KA (2012) Simple liquid models with corrected dielectric constants. J Phys Chem B 116(23):6936–6944

Fernández MS, Fromherz P (1977) Lipoid pH indicators as probes of electrical potential and polarity in micelles. J Phys Chem 81:1755–1761

Fernández DP, Mulev Y, Goodwin ARH, Levelt-Sengers JMH (1995) A database for the static dielectric constant of water and steam. J Phys Chem Ref Data 24(1):33–70

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

Forsyth WR, Gilson MK, Antonsiewicz J, Jaren OR, Robertson AD (1998) Theoretical and experimental analysis of ionization equilibria in ovomucoid third domain. Biochemistry 37:8643–8652

Freitas A, Shimizu K, Dias L, Quina F (2007) A computational study of substituted flavylium salts and their quinonoidal conjugate-bases: S0 ®s1 electronic transition, absolute pka and reduction potential calculations by DFT and semiempirical methods. J Braz Chem Soc 18:1537–1546

Friedman HL (1977) Introduction. Faraday Discuss of the Chem Soc 64:7–15

Friedman HL (1981) Electrolyte solutions at equilibrium. Ann Rev Phys Chem 32:179–204

Fuentes-Azcatl R, Barbosa MC (2016) Thermodynamic and dynamic anomalous behavior in the tip4p/ϵ water model. Physica A 444:86–94

Garcia-Moreno B (1995) Probing structural and physical basis of protein energetics linked to protons and salt. Methods in Enzymnology 259:512–538

Garrett R, Grisham C (1999) Biochemistry. Harcourt Brace & Company, EUA

Genova A, Ceresoli D, Kishtal A, Andreussi O, DiStasio RA Jr, Pavanello M (2017) eQE: An open-source density functional embedding theory code for the condensed phase. Int J Quantum Chem pp e25,401–n/a, e25401

Goh GB, Knight JL, Brooks CL (2013a) pH-dependent dynamics of complex RNA macromolecules. J Chem Theory Comput 9(2):935–943

Goh GB, Knight JL, Brooks CL (2013b) Toward accurate prediction of the protonation equilibrium of nucleic acids. J Phys Chem Lett 4(5):760–766

González-Tovar E, Lozada-Cassou M (1989) The spherical double layer: A hypernetted chain mean spherical approximation calculation for a model spherical colloid particle. J Phys Chem 93:3761–3768

Gordon M, Fedorov D, Pruitt S, Slipchenko L (2012) Fragmentation methods: A route to accurate calculations on large systems. Chem Rev 112:632–672

Greberg H, Kjellander R (1994) Electric double-layer properties calculated in the anisotropic reference hypernetted chain approximation. Mol Phys 83:789–801

Grimme S (2006) Semiempirical GGA-type density functional constructed with a long-range dispersion correction. J Comput Chem 27:1787–1799

Gu W, Frigato T, Straatsma TP, Helms V (2007) Dynamic protonation equilibrium of solvated acetic acid. Angew Chem Int Ed (English) 46(16):2939–2943

Guillot B (2002) A reappraisal of what we have learnt during three decades of computer simulations on water. J Mol Liquids 101(1-3):219–260

Harano Y, Kinoshita M (2006) On the physics of pressure denaturation of proteins. J Phys: Condens Matter 18:L107—L113

Harris TK, Turner GJ (2002) Structural basis of perturbed pK_a values of catalytic groups in enzyme active sites. Life 53:85–98

Harvey SC (1989) Treatment of electrostatic effects in macromolecular modeling. Proteins: Struc. Func and Genetics 5:78–92

Hassanali A, Prakash MK, Eshet H, Parrinello M (2011) On the recombination of hydronium and hydroxide ions in water. Proc Natl Acad Sci USA 108(51):20,410–20,415

Havranek JJ, Harbury PB (1999) Tanford–Kirkwood electrostatics for protein modeling. Proc Natl Acad Sci USA 96:11,145–11,150

He X, Jr KM (2010) Divide and conquer Hartree–Fock calculations on proteins. J Chem Theory Comput 6:405–411

He Y, Xu J, Pan XM (2007) A statistical approach to the prediction of pK_a values in proteins. Proteins: Struc Func, and Bioinformatics 69:75–82

Hess B, van der Vegt NFA (2006) Hydration thermodynamic properties of amino acid analogues: a systematic comparison of biomolecular force fields and water models. J Phys Chem B 110(35):17,616–17,626

Hill TL (1955) Approximate calculations of the electrostatic free energy of nucleic acids and other cylindrical macromolecules. Arch Biochem Biophys 57:229–239

Hill TL (1956a) Influence of electrolyte on effective dielectric constants in enzymes, proteins and other molecules. J Chem Phys 60:253–255

Hill TL (1956b) Statistical Mechanics. McGraw-Hill, New York

Hill TL (1986) An Introduction to Statistical Thermodynamics. Dover Publications Inc., New York

Ho J, Coote M (2009a) pka calculation of some biologically important carbon acids—an assessment of contemporary theoretical procedures. J Chem Theory Comput 5:295–306

Ho J, Coote M (2009b) A universal approach for continuum solvent pka calculations: are we there yet Theor Chem Acc 125:3–21

Ho J, Ertem M (2016) Calculating free energy changes in continuum solvation models. J Phys Chem B 120:1319–1329

Holst M (1993) Multilevel methods for the Poisson–Boltzmann equation. PhD thesis, Numerical Computing Group, Department of Computer Science. University of Illinois at Urbana-Champaign, USA

Honig B, Nicholls A (1995) Classical electrostatics in biology and chemistry. Science 268:1144–1149

Horinek D, Mamatkulov SI, Netz RR (2009) Rational design of ion force fields based on thermodynamic solvation properties. J Chem Phys 130:124,507

Hünenberger PH, McCammon JA (1999) Ewald artifacts in computer simulations of ionic solvation and ion–ion interactions: A continuum study. J Chem Phys 110:1856–1872

Hurtley SM (2015) New players in Lou Gehrig's disease. Science 347(6229):1432

Hyltegren K, Skepö M (2017) Adsorption of polyelectrolyte-like proteins to silica surfaces and the impact of pH on the response to ionic strength. A Monte Carlo simulation and ellipsometry study. J Colloid Interface Sci 494:266–273

Jin Y, Hoxie RS, Street TO (2017) Molecular mechanism of bacterial hsp90 ph-dependent ATPase activity. Protein Sci 26(6):1206–1213

Jönsson B (1981) The Thermodynamics of Ionic Amphiphilie–Water Systems—A Theoretical Analysis. PhD thesis. Lund University, Sweden

Jönsson B, Åkesson T, Woodward C (1996) Theory of interactions in charged colloids. In: Arora AK, Tata BVR (eds) Ordering and phase transitions in charged colloids, VCH, New York, pp 295–313

Jönsson B, Lund M, Barroso da Silva FL (2007) Electrostatics in macromolecular solution. In: Dickinson E, Leser ME (eds) Food Colloids: Self-Assembly and Material Science, Royal Society of Chemistry, Londres, pp 129–154

Juffer AH (1992) Melc—The Macromolecular Electrostatics Computer program. Laboratory of Physical Chemistry. University of Groningen, The Netherlands

Juffer AH (1993) On the modelling of solvent mean force potentials—from liquid argon to solvated macromolecules. PhD thesis, Rijkuniversiteit Groningen, The Netherlands

Juffer AH (1998) Theoretical calculations of acid-dissociation constants of proteins. Biochem Cell Biol 76:198–209

Juffer AH, Botta EFF, van Keulen BAM, van der Ploeg A, Berendsen HJC (1991) The electric potential of a macromolecule in a solvent: A fundamental approach. J Comp Phys 97:144–171

Kamerlin SL, Haranczyk M, Warshel A (2009) Progresses in ab initio QM/MM free energy simulations of electrostatic energies in proteins: Accelerated QM/MM studies of pKa, redox reactions and solvation free energies. J Phys Chem B 113:1253–1272

Karplus M, McCammon JA (1979) Protein structural fluctuations during a period of 100 ps. Nature 277:578

Karplus M, Gelin BR, McCammon JA (1977) Dynamics of folded proteins. Nature 267:585–590

Kesvatera T, Jönsson B, Thulin E, Linse S (1996) Measurement and modelling of sequence-specific pK_a values of calbindin D_{9k}. J Mol Biol 259:828

Kesvatera T, Jönsson B, Thulin E, Linse S (1999) Ionization behavior of acidic residues in calbindin D_{9k}. Proteins: Struc Func, and Genetics 37:106–115

Kesvatera T, Jönsson B, Thulin E, Linse S (2001) Focusing of the electrostatic potential at EF-hands of calbindin D_{9k}: Titration of acidic residues. Proteins: Struc Func, and Genetics 45:129–135

Khandogin J, Brooks CL III (2005) Constant pH molecular dynamics with proton tautomerism. Biophys J 89:141–157

Kim MO, McCammon JA (2016) Computation of pH-dependent binding free energies. Biopolymers 105:43–49

King G, Lee FS, Warshel A (1991) Microscopic simulations of macroscopic dielectric constants of solvated proteins. J Chem Phys 95:4366–4377

Kirkwood JG (1934a) Solutions containing zwitterions: Erratum. J Chem Phys 2:713

Kirkwood JG (1934b) Theory of solutions of molecules containing widely separated charges with special application to zwitterions. J Chem Phys 2:351–361

Kirkwood JG, Shumaker JB (1952) Forces between protein molecules in solution arising from fluctuations in proton charge and configuration. Proc Natl Acad Sci USA 38:863–871

Kirkwood JG, Westheimer FH (1938) The electrostatic influence of substituents on the dissociation constant of organic acids. I J Chem Phys 6:506–512

Klamt A (1995) Conductor-like screening model for real solvents: A new approach to the quantitative calculation of solvation phenomena. J Phys Chem 99:2224–2235

Ko J, Murga LF, Wei Y, Ondrechen MJ (2005) Prediction of active sites for protein structures from computed chemical properties. Bioinformatics 21(suppl 1):i258—i265

Kong X, Brooks CL III (1996) λ-dynamics: a new approach to free energy calculations. J Chem Phys 105:128–141

Koukiekolo R, Sagan SM, Pezacki JP (2007) Effects of pH and salt concentration on the siRNA binding activity of the RNA silencing suppressor protein p19. FEBS Lett 581:3051–3056

Krieger E, Nielsen JE, Spronk CAEM, Vriend G (2006) Fast empirical pKa prediction by Ewald summation. J Mol Graph Model 25:481–486

Kukic P, Farrell D, McIntosh LP, García-Moreno JKS, Toleikis Z, Teilum K, Nielsen JE (2013) Protein dielectric constants determined from NMR chemical shift perturbations. J Am Chem Soc 135(45):16,968–16,976

Kurut A, Persson BA, Åkesson T, Forsman J, Lund M (2012) Anisotropic interactions in protein mixtures: Self assembly and phase behavior in aqueous solution. J Phys Chem Lett 3(6):731–734

Kurut A, Dicko C, Lund M (2015) Dimerization of terminal domains in spiders silk proteins is controlled by electrostatic anisotropy and modulated by hydrophobic patches. ACS Biomater Sci Eng 1(6):363–371

Labbez C, Jönsson B (2007) A new Monte Carlo method for the titration of molecules and minerals. In: gström BK (ed) Lecture Notes in Computer Science. Springer-Verlag, Berlin, pp 66–72

Laio A, Parrinello M (2002) Escaping free-energy minima. Proc Natl Acad Sci USA 99:12,562–12,566

Lee MS, Jr FRS, Brooks CL (2004) Constant-pH molecular dynamics using continuous titration coordinates. Proteins 56:738–752

Lee J, Miller BT, Damjanovi A, Brooks BR (2015) Enhancing constant-pH simulation in explicit solvent with a two-dimensional replica exchange method. J Chem Theory Comput 11:2560–2574

Lee J, Miller BT, Brooks BR (2016) Computational scheme for pH-dependent binding free energy calculation with explicit solvent. Prot Sci 25:231–243

Legault P, Pardi A (1994) In situ probing of adenine protonation in RNA by 13C NMR. J Am Chem Soc 116(18):8390–8391

Levesque D, Weis JJ, Hansen JP (1986) Simulation of classical fluids. In: Binder K (ed) Monte Carlo Methods in Statistical Physics, vol 5. Springer-Verlag, Berlin, pp 47–119

Levitt M, Lifson S (1969) Refinement of protein conformations using a macromolecular energy minimization procedure. J Mol Biol 46:269

Lewis M, Bamforth C (2006) Essays in Brewing Science. Springer

Li H, Hains A, Everts J, Robertson A, Jensen J (2002) The prediction of protein pka's using qm/mm: The pka of lysine 55 in turkey ovomucoid third domain. J Phys Chem B 106:3486–3494

Li H, Robertson AD, Jensen JH (2005) Very fast empirical prediction and rationalization of protein pKa values. Proteins: Struct Funct Bioinf 61(4):704–721

Lill MA, Helms V (2001) Molecular dynamics simulation of proton transport with quantum mechanically derived proton hopping rates (Q-HOP MD). J Chem Phys 115(17):7993–8005

Linderstrøm-Lang K (1924) Om proteinstoffernes ionisation. C R Trav Lab Carlsberg [Meddelelser fra Carlsberg Lab] 15(7):1–28

Linse P, Jönsson B (1983) A Monte Carlo study of the electrostatic interaction between highly charged aggregates. A test of the cell model applied to micellar systems. J Chem Phys 78:3167–3176

Linse P, Lobaskin V (1999) Electrostatic attraction and phase separation in solutions of like-charge colloidal particles. submitted

Linse S, Jönsson B, Chazin WJ (1995) The effect of protein concentration on ion binding. Proc Natl Acad Sci USA 92:4748–4752

Lizatović R, Aurelius O, Stenström O, Drakenberg T, Akke M, Logan DT, André I (2016) A de novo designed coiled-coil peptide with a reversible ph-induced oligomerization switch. Structure 24(6):946–955

Löffler G, Screiber H, Steinhauser O (1997) Calculation of the dielectric properties of a protein and its solvent: Theory and a case study. J Mol Biol 270:520–534

Lovett RA, Mou CY, Buff FP (1976) The structure of the liquid–vapor interface. J Chem Phys 65:570–572

Lund M, Jönsson B (2003) A mesoscopic model for protein–protein interactions in solution. Biophys J 85:2940–2947

Lund M, Jönsson B (2005) On the charge regulation of proteins. Biochemistry 44(15):5722–5727

Lund M, Jönsson B (2013) Charge regulation in biomolecular solution. Q Rev Biophys 46:265–281

Lyklema J (1991) Fundamentals of Interface and Colloid Science. Academic Press, San Diego

Lyubartsev AP, Laaksonen A (1996) Concentration effects in aqueous NaCl solutions. A molecular dynamics simulation. J Phys Chem 100:16,410–16,418

Machuqueiro M, Baptista AM (2007) The pH-dependent conformational states of kyotorphin: A constant-pH molecular dynamics study. Biophys J 92:1836–1845

Machuqueiro M, Baptista AM (2011) Is the prediction of pka values by constant-pH molecular dynamics being hindered by inherited problems? Proteins: Struct Funct Bioinf 79(12):3437–3447

Madura JD et al (1994) Biological applications of electrostatic calculations and Brownian dynamics simulations. In: Lipkowitz KB, Boyd DB (eds) Reviews in Computational Chemistry, vol 5. VCH Publishers, Inc, New York, pp 229–267

Magalhães PR, Oliveira ASF, Campos SRR, Soares CM, Baptista AM (2017) Effect of a pH gradient on the protonation states of cytochrome c oxidase: A continuum electrostatics study. J Chem Inf Model 57(2):256–266

Mahadevan TS, Garofalini SH (2008) Dissociative water potential for molecular dynamics simulations. J Phys Chem B 111(30):8919–8927

Marenich A, Cramer C, Truhlar D (2009) Universal solvation model based on solute electron density and on a continuum model of the solvent defined by the bulk dielectric constant and atomic surface tensions. J Phys Chem B 113:6378–6396

McQuarrie DA (1976) Statistical Mechanics. Harper Collins, New York

Medda L, Barse B, Cugia F, Boström M, Parsons DF, Ninham BW, Monduzzi M, Salis A (2012) Hofmeister challenges: Ion binding and charge of the BSA protein as explicit examples. Langmuir 28:16,355–16,363

Meyer TD, Ensing B, Rogge S, Clerck KD, Meijer E, Speybroeck VV (2016) Acidity constant (pKa) calculation of large solvated dye molecules: Evaluation of two advanced molecular dynamics methods. ChemPhysChem 17:3447–3459

Mongan J, Case DA (2005) Biomolecular simulations at constant pH. Curr Opin Struct Biol 15:157–163

Mongan J, Case DA, McCammon JA (2004) Constant pH molecular dynamics in generalized born implicit solvent. J Comp Chem 25:2038–2048

Mukerjee P, Banerjee K (1964) A study of the surface pH of micelles using solubilized indicator dyes. J Phys Chem 68:3567–3574

Noid WG (2013) Perspective: Coarse-grained models for biomolecular systems. J Chem Phys 139:090,901

Northrup SH, McCammon JA (1980) Simulation methods for protein structure fluctuations. Biopolymers 19:1001–1016

Nozaki Y, Tanford C (1967) Examination of titratation behavior. Methods Enzymol 11:715–734

Nylander KHT, Lund M, Skepö M (2017) Adsorption of the intrinsically disordered saliva protein histatin 5 to silica surfaces. A Monte Carlo simulation and ellipsometry study. J Colloid Interface Sci 467:280–290

Oliveira ASF, Campos SRR, Baptista AM, Soares CM (2016) Coupling between protonation and conformation in cytochrome c oxidase: Insights from constant-pH MD simulations. Biochimica et Biophysica Acta 1857:759–771

Olsson MH, Sondergard CR, Rostkowski M, Jensen JH (2011) PROPKA3: Consistent treatment of internal and surface residues in empirical pKa predictions. J Chem Theory Comput 7(2):525–537

Orttung WH (1977) Direct solution of the Poisson equation for biomolecules of arbitrary shape, polarizability density and charge distribution. Ann N Y Acad Sci 303:22–37

Outhwaite C, Bhuiyan L (1991) A modified Poisson–Boltzmann analysis of the electric double layer around an isolated spherical macroion. Mol Phys 74(2):367–381

Overbeek JTG (1982) A fascinating subject: Introductory lecture in Colloidal Dispersion. The Royal Society of Chemistry. Ed. J. W. Goodwin, London

Pechlaner M, Donghi D, Zelenay V, Sigel RKO (2015) Protonation-dependent base flipping at neutral pH in the catalytic triad of a self-splicing bacterial group II intron. Angew Chem Int Ed Engl 54(33):9687–9690

Penfold R, Warwicker J, Jönsson B (1998) Electrostatic models for calcium binding proteins. J Phys Chem B 102:8599–8610

Persson B, Lund M, Forsman J, Chatterton DEW, Åkesson T (2010) Molecular evidence of stereo-specific lactoferrin dimers in solution. Biophys Chem 3(3):187–189

Perutz MF (1978) Electrostatic effects in proteins. Science 201:1187–1191

Piper DW, Fenton BH (1965) pH stability and activity curves of pepsin with special reference to their clinical importance. Gut 6:506–508

Project CSE (1995) Direct and Inverse Bioelectric Field Problems. http://csep1.phy.ornl.gov/bf/bf.html

Radak BK, Roux B (2016) Efficiency in nonequilibrium molecular dynamics Monte Carlo simulations. J Chem Phys 145:124,109

Reitz JR, Milford FJ, Christy RW (1986) Fundamentos da teoria eletromagnética. Rio de Janeiro, Editora Campus

Rossini E, Knapp EW (2016) Proton solvation in protic and aprotic solvents. J Comput Chem 37:1082–1091

Roxby R, Tanford C (1971) Hydrogen ion titration curve of lysozyme in 6 M guanidine hydrochloride. Biochemistry 10:3348–3352

Russel WB, Saville DA, Schowalter WR (1989) Colloidal Dispersions. Cambridge University Press, Cambridge

Sakalli I, Knapp EW (2015) pKa in proteins solving the Poisson–Boltzmann equation with finite elements. J Comput Chem 36(29):2147–2157

Santos HAF, Cosa DVV, Teixeira VH, Baptista AM, Machuqueiro M (2015) Constant-pH MD simulations of DMPA/DMPC lipid bilayers. J Chem Theory Comput 11(12):5973–5979

Schaller W, Robertson AD (1995) pH, ionic strength, and temperature dependences of ionization equilibria for the carboxyl groups in turkey ovomucoid third domain. Biochemistry 34(14):4714–4723

Schlichter CP (1980) Principles of Magnetic Resonance. Springer, Berlin

Schmitz KS (ed) (1994) Macro-ion Characterization: From Dilute Solutions to Complex Fluids. American Chemistry Society, Washington

Schönichen A, Webb BA, Jacobson MP, Barber DL (2013) Considering protonation as a posttranslational modification regulating protein structure and function. Annu Rev Biophys 42:289–314

Schutz CN, Warshel A (2001) What are the dielectric "constants" of proteins and how to validate electrostatic models? Proteins: Struc Func, and Genetics 44:400–417

Sham YY, Chu ZT, Warshel A (1997) Consistent calculations of pK_a's of ionizable residues in proteins: Semi-microscopic and microscopic approaches. J Phys Chem B 101:4458–4472

Sharp KA, Honig B (1990) Calculating total electrostatic energies with the nonlinear Poisson–Boltzmann equation. J Phys Chem 94:7684–7692

Sharp KA, Fine R, Honig B (1987) Computer simulations of the diffusion of a substrate to an active site of an enzyme. Science 236:1460–1463

Shaw DJ (1992) Introduction to Colloid and Surface Chemistry, 4th edn. Butterworths, London

Sheinerman FB, Norel R, Honig B (2000) Curr Opin Struct Biol 10:153–159

Shimizu K, Freitas AA, Farah JPS, Dias LG (2005) Predicting hydration free energies of neutral compounds by a parametrization of the polarizable continuum model. J Phys Chem A 109:11,322–11,327

Shukla A, Mylonas E, Di Cola E, Finet S, Timmins P, Narayanan T, Svergun DI (2008) Absence of equilibrium cluster phase in concentrated lysozyme solutions. Proc Natl Acad Sci USA 105(13):5075–5080

Simonson T (2013) What is the dielectric constant of a protein when its backbone is fixed J Chem Theory Comput 9(10):4603–4608

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

Simonson T, Brooks CL III (1996) Charge screening and the dielectric constant of proteins: Insights from molecular dynamics. J Am Chem Soc 118:8452–8458

Smith N, Witham S, Sarkar S, Zhang J, Li L, Li C, Alexov E (2012) Delphi web server v2: Incorporating atomic-style geometrical figures into the computational protocol. Bioinformatics 28(12):1655–1657

Soares RO, Torres PHM, da Silva ML, Pascutti PG (2016) Unraveling HIV protease flaps dynamics by constant pH molecular dynamics simulations. J Structural Biology 195:261–226

Socher E, Stich H (2016) Mimicking titration experiments with MD simulations: A protocol for the investigation of pH-dependent effects on proteins. Scientific Reports 22523:1–12

Sorensen SPL (1909) Biochem Z 21:131

Sorensen SPL, Hempel MHJ, Palitzsch S (1917) Studies on proteins. II. the capacity of egg-albumin to combine with acids or bases. C R Trav Lab Carlsberg [Meddelelser fra Carlsberg Lab] 12:68–163

Srivastava D, Santiso E, Gubbins KE, Barroso da Silva FL (2017) Computationally mapping pKa shifts due to the presence of a polyelectrolyte chain around whey proteins. https://doi.org/10.1021/acs.langmuir.7b02271

Stanton CL, Houk KN (2008) Benchmarking pKa prediction methods for residues in proteins. J Chem Theory Comput 4(6):951–966

Steiner E, Gastl M, Becker T (2011) Protein changes during malting and brewing with focus on haze and foam formation: a review. Eur Food Res Technol 232:191–204

Stern HA (2007) Molecular simulation with variable protonation states at constant pH. J Chem Phys 126:164,112

Stigter D, Dill KA (1990) Charge effects on folded and unfolded proteins. Biochemistry 29:1262–1271

Stoll S (2014) Computer simulations of soft nanoparticles and their interactions with DNA-like polyelectrolytes. In: Callejas-Fernandez J (ed) Soft Nanoparticles for Biomedical Applications. Royal Society of Chemistry, Londres, pp 342–371

Svensson B, Jönsson B, Woodward CE (1990) Electrostatic contributions of the binding of Ca^{2+} in calbindin mutants. A Monte Carlo study. Biophys Chem 38:179–183

Svensson BR, Woodward CE (1988) Widom's method for uniform and non-uniform electrolyte solutions. Mol Phys 64:247–259

Swails JM, York DM, Roitberg AE (2014) Constant pH replica exchange molecular dynamics in explicit solvent using discrete protonation states: Implementation, testing, and validation. J Chem Theory Comput 10:1341–1352

Szabo A, Ostlund NS (1989) Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory. McGraw-Hill, New York

Takano Y, Houk KN (2005) Benchmarking the conductor-like polarizable continuum model (CPCM) for aqueous solvation free energies of neutral and ionic organic molecules. J Chem Theory Comput 1:70–77

Tanford C (1957a) The location of electrostatic charges in Kirkwood's model of organic ions. J Am Chem Soc 79:5348–5352

Tanford C (1957b) Theory of protein titration curves II. Calculations for simple models at low ionic strength. J Am Chem Soc 79:5340–5347

Tanford C, Kirkwood JG (1957) Theory of protein titration curves I. General equations for impenetrable spheres. J Am Chem Soc 79:5333–5339

Tanford C, Roxby R (1972) Interpretation of protein titration curves. Application to lysozyme. Biochemistry 11:2192–2198

Tang CL, Alexov E, Pyle AM, Honig B (2007) Calculation of pKas in RNA: on the structural origins and functional roles of protonated nucleotides. J Mol Biol 366(5):1475–1496

Taylor D, Ángyán J, Galli G, Zhang C, Gygi F, Hirao K, Song J, Rahul K, von Lilienfeld A, Podeszwa R, Bulik I, Henderson T, Scuseria G, Toulouse J, Peverati R, Truhlar D, Szalewicz K (2016) Blind test of density-functional-based methods on intermolecular interaction energies. J Chem Phys 145:124,105

Teixeira AA, Lund M, Barroso da Silva FL (2010) Fast proton titration scheme for multiscale modeling of protein solutions. J Chem Theory Comput 6(10):3259–3266

Teleman O, Svensson B, Jönsson B (1991) Efficiency in statistical mechanical simulations of biomolecules—computer programs for molecular and continuum modelling. Comput Phys Commun 62(2-3):307–326

Terán LM, Dí az-Herrera E, Lozada-Cassou M, Saavedra-Barrera R (1989) A comparison of numerical methods for solving nonlinear integral equations found in liquid theories. J Comp Phys 84:326–342

Thaplyal P, Bevilacqua PC (2014) Experimental approaches for measuring pKa's in RNA and DNA. Methods Enzymol 549:189–219

Thurlkill RL, Grimsley GR, Scholtz JM, Pace CN (2006) pK values of the ionizable groups of proteins. Prot Sci 15:1214–1218

Tironi IG, Sperb R, Smith PE, van Gunsteren WF (1995) A generalized reaction field method for molecular dynamics simulations. J Chem Phys 102:5451–5459

Tummanapelli A, Vasudevan S (2015) Ab initio molecular dynamics simulations of amino acids in aqueous solutions: Estimating pka values from metadynamics sampling. J Phys Chem B 119:12,249–12,255

Usui S (1984) Electrical double layer. In: Kitahara A, Watanabe A (eds) Electrical Phenomena at Interfaces – Fundamentals, Measurements, and Applications. Marcel Dekker, Inc., New York, pp 15–46

van Gunsteren WF, Berendsen HJC (1990) Computer simulation of molecular dynamics: Methodology, applications, and perspective in chemistry. Angew Chem Int Ed Engl 29:992–1023

Varma S, Jakobsson E (2004) Ionization states of residues in OmpF and mutants: Effects of dielectric constant and interactions between residues. Biophys J 86(2):690–704

Verlet L (1967) Computer experiments on classical fluids. I. Thermodynamical properties of Lennard–Jones molecules. Phys Rev 159(1):98–103

Verwey EJW, Overbeek JTG (1948) Theory of the Stability of Lyophobic Colloids. Elsevier Publishing Company Inc., Amsterdam

Vicatos S, Roca M, Warshel A (2009) Effective approach for calculations of absolute stability of proteins using focused dielectric constants. Proteins 77(3):670–684

Vila-Viçosa DVV, Campos SRR, Baptista AM, Machuqueiro M (2012) Reversibility of prion misfolding: Insights from constant-pH molecular dynamics simulations. J Phys Chem B 116(30):8812–8821

Wagoner T, Vardhanabhuti B, Foegeding EA (2016) Designing whey protein–polysaccharide particles for colloidal stability. Annu Rev Food Sci Technol 7:93–116

Wallace JA, Shen JK (2009) Predicting pKa values with continuous constant pH molecular dynamics. Methods Enzymol 466:455–475

Wallace JA, Shen JK (2011) Continuous constant pH molecular dynamics in explicit solvent with pH-based replica exchange. J Chem Theory Comput 7:2617–2629

Wallace JA, Shen JK (2012) Unraveling a trap-and-trigger mechanism in the ph-sensitive self-assembly of spider silk proteins. J Phys Chem Lett 3(5):658–662

Wang L, Li L, Alexov E (2015) pKa predictions for proteins, RNAs, and DNAs with the Gaussian dielectric function using DelPhi pKa. Proteins 83:2186–2197

Wang L, Zhang M, Alexov E (2016) DelPhiPKa web server: predicting pKa of proteins, RNAs and DNAs. Bioinformatics 34(4):614–615

Warshel A (1981) Calculations of enzymatic reactions: Calculations of pKa, proton transfer reactions, and general acid catalysis reactions in enzymes. Biochemistry 20:3167–3177

Warshel A (2014) Multiscale modeling of biological functions: From enzymes to molecular machines (Nobel Lecture). Angew Chem Int Ed 53:10020–10031

Warshel A, Åqvist J (1991) Electrostatic energy and macromolecular function. Annu Rev Biophys Biophys Chem 20:267–298

Warshel A, Levitt M (1976) Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. J Mol Biology 103(2):227–249

Warshel A, Papazyan A (1998) Electrostatic effects in macromolecules: Fundamental concepts and practical modeling. Curr Opin Struct Biol 8:211–217

Warshel A, Russel ST, Churg AK (1984) Macroscopic models for studies of electrostatic interactions in proteins: Limitations and applicability. Proc Natl Acad Sci USA 81:4785–4789

Warshel A, Sharma PK, Kato M, Parson WW (2006) Modeling electrostatic effects in proteins. Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics 1764(11):1647–1676

Warwicker J (1999) Simplified methods for pK_a and acid pH-dependent stability estimation in proteins: Removing dielectric and counterion boundaries. Prot Sci 8:418–425

Warwicker J, Watson HC (1982) Calculation of the electric potential in the active site cleft due to α-helix dipoles. J Mol Biol 157:671–679

Westheimer FH, Kirkwood JG (1938) The electrostatic influence of substituents on the dissociation constant of organic acids. II. J Chem Phys 6:513–517

Williams SL, de Oliveira CAF, McCammon JA (2010) Coupling constant ph molecular dynamics with accelerated molecular dynamics. J Chem Theory Comput 6:560–568

Woodward CE, Svensson B (1991) Potentials of mean force in charged systems: Application to Superoxide Dismutase. J Phys Chem 95:7471–7477

Xiao K, Yu H (2016) Rationalising pKa shifts in *Bacillus circulans* xylanase with computational studies. Phys Chem Chem Phys 18:30,305–30,312

Ye K, Malinina L, Patel D (2003) Recognition of small interfering RNA by a viral suppressor of RNA silencing. Nature 426:874–878

You TJ, Bashford D (1995) Conformation and hydrogen ion titration of proteins: a continuum electrostatic model with conformational flexibility. Biophys J 69(5):1721–1733

Yu W, Lopes PEM, Roux B, MacKerell Jr AD (2013) Six-site polarizable model of water based on the classical drude oscillator. J Chem Phys 138:034,508