Quarterly Reviews of Biophysics **46**, 2 (2013), pp. 181–209. © Cambridge University Press 2013
doi:10.1017/S0033583513000024 Printed in the United States of America

Protonation and pK changes in protein–ligand binding

Alexey V. Onufriev[1]* and Emil Alexov[2]*

[1] Departments of Computer Science and Physics, 2050 Torgersen Hall, Virginia Tech, Blacksburg, VA 24061, USA
[2] Department of Physics and Astronomy, Clemson University, Clemson, SC 29634, USA

Abstract. Formation of protein–ligand complexes causes various changes in both the receptor and the ligand. This review focuses on changes in pK and protonation states of ionizable groups that accompany protein–ligand binding. Physical origins of these effects are outlined, followed by a brief overview of the computational methods to predict them and the associated corrections to receptor–ligand binding affinities. Statistical prevalence, magnitude and spatial distribution of the pK and protonation state changes in protein–ligand binding are discussed in detail, based on both experimental and theoretical studies. While there is no doubt that these changes occur, they do not occur all the time; the estimated prevalence varies, both between individual complexes and by method. The changes occur not only in the immediate vicinity of the interface but also sometimes far away. When receptor–ligand binding is associated with protonation state change at particular pH, the binding becomes pH dependent: we review the interplay between sub-cellular characteristic pH and optimum pH of receptor–ligand binding. It is pointed out that there is a tendency for protonation state changes upon binding to be minimal at physiologically relevant pH for each complex (no net proton uptake/release), suggesting that native receptor–ligand interactions have evolved to reduce the energy cost associated with ionization changes. As a result, previously reported statistical prevalence of these changes – typically computed at the same pH for all complexes – may be higher than what may be expected at optimum pH specific to each complex. We also discuss whether proper account of protonation state changes appears to improve practical docking and scoring outcomes relevant to structure-based drug design. An overview of some of the existing challenges in the field is provided in conclusion.

1. Introduction 182

2. Origins of protonation state changes in receptor–ligand binding and related corrections to binding affinity 183

3. Magnitude and prevalence of protonation and p*K* changes in receptor proteins upon ligand binding 187

3.1. What is known from experiment? 187
3.2. What do calculations tell us? 190
3.2.1. Earlier studies 190
3.2.2. Large scale studies 190
3.2.3. Relative prevalence of various physical causes of p*K* shifts in ligand binding 193
3.2.4. Long-range coupling between ligand binding and protonation 193

* Authors for correspondence: Alexey V. Onufriev, Departments of Computer Science and Physics, 2050 Torgersen Hall, Virginia Tech, Blacksburg, VA 24061, USA. Email: alexey@cs.vt.edu; Emil Alexov, Department of Physics and Astronomy, Clemson University, Clemson, SC 29634, USA. Email: ealexov@clemson.edu

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

4. Interplay among subcellular pH, pH-optimum of ligand binding, and pH of the corresponding reaction 195

5. Improving docking and scoring outcomes 197

6. Outlook 201

7. Acknowledgements 203

8. References 203

1. Introduction

Protein–ligand binding is one of the most ubiquitous and diverse processes in biology. Proteins bind to each other to form vital complexes such as hemoglobin; DNA in higher organisms is found mostly in complex with the histones and other proteins; and when the ligand is a small molecule, it can be the substrate that binds to an enzyme in the course of a biochemical reaction. Binding of small molecules to proteins is of interest for yet another important reason: early stages of structure-based drug discovery (Gilson & Zhou, 2007; Gohlke & Klebe, 2002; Jorgensen, 2004) often involve identifying a ligand that binds to the target protein with high affinity.

Regardless of the nature of receptor–ligand binding, understanding the mechanisms involved requires detailed knowledge of the nature and origins of changes in physical states of the participating proteins and ligands. It is well known that structural complementarity plays a critical role in the binding process, and so it is not surprising that structural re-arrangements that can accompany protein–ligand binding have been extensively explored (Goh *et al.* 2004; Holmes *et al.* 2003; Boehr & Wright, 2008), including variations in ligand-binding mode to the same receptor (Mobley & Dill, 2009). The 'induced fit' mechanism, whereby the protein adjusts its shape to better accommodate the ligand, has long been in textbooks (Voet & Voet, 1995). Structure–energy relationships in the binding process have also been systematically investigated (Bordner & Abagyan, 2005; Elcock *et al.* 2001; Ofran & Rost, 2003; Sheinerman & Honig, 2002).

In contrast, relatively little was known until recently about the magnitude, prevalence, and detailed origins of changes in charge states of receptors and ligands that accompany the binding. These changes are directly related to binding-induced changes in pK[1] values of ionizable groups in proteins and their ligands. In part, the relative underdevelopment of this field compared to its close 'relative' – studies of structural changes in protein–ligand binding – can be attributed to the difficulties associated with experimental determination of pK and protonation state changes in a complex formation. Even today, only a handful of relevant data points exist. Moreover, even though rigorous physics-based methods able to compute titratable group pKs based on atomic resolution protein structures have existed for at least two decades (Alexov *et al.* 2011; Bashford & Karplus, 1990; Gilson, 1993; Yang *et al.* 1993), no 'large scale' studies of pK changes in protein–ligand appeared until recently.

It is important to realize that energetic consequences of binding-induced protonation state changes may be as large as those caused by partial protein unfolding or destabilizing mutations. For example, at pH = 6.5, the cost of changing the ionization state of a single group with pK = 4.5 is larger than 2 kcal/mol – a realistic scenario in protein–ligand binding. This relatively

[1] At the level of this review, we do not make a distinction between pK_a and p$K_{1/2}$ (mid-point of titration curve), and denote both by pK, see e.g. Ref. (Onufriev *et al.* 2001) for an in-depth discussion of this issue.

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

large energy should not be surprising if one draws an analogy between ligand binding and protein folding: both processes often involve burial (desolvation) of large parts of the protein. It was well known for a long time that energetic costs of several kcal/mol are often associated with burial of certain ionizable groups inside folded proteins. In fact, it is this effect that leads to the observed pH dependence of protein stability (Tanford, 1970; Yang & Honig, 1992). Not surprisingly, the change in the protonation state of the complex relative to that of the separated receptor and its ligand is also a prerequisite for the observed (Blundell *et al.* 2007; Djurdjevic-Pahl *et al.* 2005) pH dependence of receptor–ligand binding (Bidwai *et al.* 2008; Hansen *et al.* 2011; Maiti & Drohat, 2011; Sackett *et al.* 2011). The change of protonation states of titratable groups during particular reaction is the cause of pH-dependence of most enzymatic reactions (Voet & Voet, 1995) as well. Since evolution cares about survival and reproduction only, which is related to biological activity and networks on the cellular level, one can ask whether the pH-dependence of activity of biological macromolecules is somehow related to their pH-dependence of stability and interactions (Talley & Alexov, 2010). In addition, how the pH-dependent properties are related to the subcellular and tissue pH? (Garcia-Moreno, 2009; Kundrotas & Alexov, 2006; Mitra *et al.* 2011). Is it possible that macromolecular interactions have evolved toward specific pH, and can tolerate small deviations around it. Alternatively, they may have evolved to be strongly pH-dependent to sense or regulate the local pH (Mitra *et al.* 2011). These issues must be considered when discussing protonation state changes in ligand binding and their possible effects on practical computations.

Until recently, many questions related to the magnitude, prevalence, detailed physical origins, and practical implications of changes in the charge state of receptor proteins (and ligands) on complex formation were not explored on a statistically large scale. Even though earlier case studies suggested that accounting for possible changes in the charge state on binding may be important for structure-based drug design, practical, automated ligand docking, and affinity ordering (scoring) methods (Taylor *et al.* 2002) routinely ignored the possibility of ionization state changes on the binding. Occasional manual intervention, e.g., based on 'visual inspection' of polar hydrogens was reported (Warren *et al.* 2005), but it is still a common practice to perform a docking calculations with a single protonation state of both the ligand and receptor. Useful prediction of ligand-binding affinities is still challenging (Ge & Roux, 2010; Gilson & Zhou, 2007; Lee & Olson, 2008; Rodinger *et al.* 2008; Singh & Warshel, 2010; Wang *et al.* 2006; Warren *et al.* 2005; Woo & Roux, 2005; Woo, 2008; Zhou & Gilson, 2009).

The situation is now changing rapidly. Reduced computational costs and improved accuracy of practical p*K* calculations (see recent reviews (Alexov *et al.* 2011; Nielsen *et al.* 2011)), along with availability of such methods as fast web-based tools (Anandakrishnan *et al.* 2012; Bas *et al.* 2008; Dolinsky *et al.* 2007), have facilitated large-scale analysis of ionization state changes in protein–ligand binding. Promising methods begin to appear that consistently take variability of receptor/ligand ionization (protonation) states into account (Park *et al.* 2011). In this brief review, we will attempt to summarize what is known, and what challenges lie ahead.

2. Origins of protonation state changes in receptor–ligand binding and related corrections to binding affinity

Any ionizable (titratable) group in a protein is involved in a number of electrostatic interactions, both with the other groups and with the solvent. Ultimately, it is the change of these interactions

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

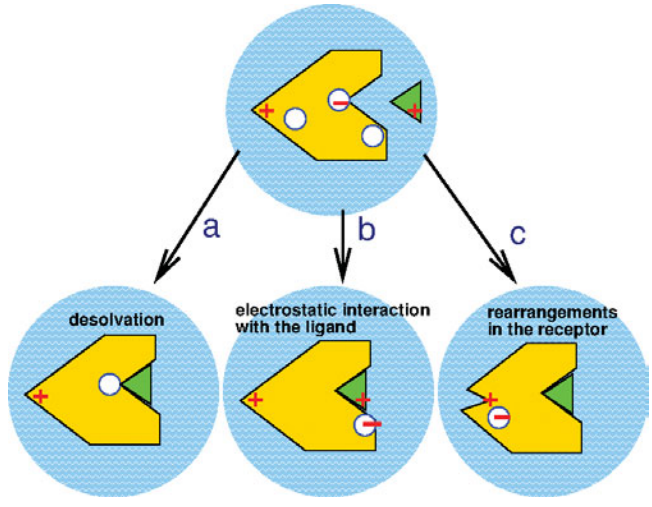


Fig. 1. Three main mechanisms behind changes in pK (and state of protonation) of an ionizable group in receptor protein upon ligand binding. (a) Desolvation of the group upon binding of the ligand. (b) Direct electrostatic interaction with the ligand. (c) Structural re-arrangements in the receptor protein caused by the binding.

on ligand binding that is responsible[2] for the change in the group's pK and charge (protonation) state. The general idea of electrostatic origin of observed pK shifts of ionizable groups in proteins, relative to their values in solution, is at the core of most methods for pK prediction in proteins, both historic and modern (Antosiewicz *et al.* 1994; Bas *et al.* 2008; Bashford & Karplus, 1990; Beroza *et al.* 1991; DelBuono *et al.* 1994; Demchuk & Wade, 1996; Georgescu *et al.* 2002; Khandogin & Brooks, 2006; Kieseritzky &Knapp, 2008a; Mongan *et al.* 2004; Nielson & Vriend, 2001; Sham *et al.* 1997; Simonson *et al.* 2004; Song *et al.* 2009; Spassov & Yan, 2008; Takahashi *et al.* 1992; Tanford & Kirkwood, 1957; Tanford & Roxby, 1972; Ullmann & Knapp, 1999; Yang *et al.* 1993).

As a reasonable first approximation, assume that ionizable groups that change their charge states upon ligand binding do not interact with each other; arguments in favor of the approximation are given in (Mason & Jensen, 2008). Within this simplification, the pK change of a given ionizable group in the receptor protein can be decomposed into two distinct contributions (Bashford & Karplus, 1990),

$$\Delta \mathrm{p}K = \Delta \mathrm{p}K^{\mathrm{Born}} + \Delta \mathrm{p}K^{\mathrm{Back}}, \quad (1)$$

where the first term is the so-called 'Born' contribution or desolvation penalty, and the second is the 'background' contribution. The Born term describes the energetic costs associated with changing the degree of solvation of the group. Typically, these are unfavorable costs of desolvating the group by replacing the solvent–water and ions–in the binding pocket with the ligand, Fig. 1*a*. Structural re-arrangements caused in the receptor by the ligand binding, Fig. 1*c*, may also contribute to $\Delta \mathrm{p}K^{\mathrm{Born}}$ by virtue of changing the solvation pattern of the group. For example, a solvent exposed glutamic acid in the binding pocket may have a pK near its model

[2] With the caveat that hydrogen bonds, which contribute significantly to pKs , are not entirely electrostatic (Li *et al.* 2011) in nature. This caveat does not affect the conceptual picture presented in this section.

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

compound value in solution (about 4.5), thus carrying a net charge of −1 at neutral pH = 7.0. Upon complexation, it may find itself buried inside a low dielectric medium (protein + ligand). If it were to stay charged, the energetic penalty would be large. Assuming the proton charge in this case is smeared over a sphere of radius $R = 3$ Å, the unfavorable cost of transferring it from water $\varepsilon_w = 80$ to protein interior $\varepsilon_p = 4$, is $\Delta G \approx 166[\text{kcal/mol}](1/\varepsilon_p - 1/\varepsilon_w)e^2/R \approx 12$ kcal/mol. On the other hand, if the group picks up a proton from the solvent at this pH, the cost is only $2.3\, k_B T(\text{pH} - \text{p}K) \approx 3.4$ kcal/mol, and there is no desolvation penalty to bury the now neutral group. This energetically more favorable scenario is realized, and the group becomes neutralized, Fig. 1*a*. Another way of thinking about the process in this example is that the group's pK is shifted up by (12 kcal/mol)/(2.3 $k_B T$) ≈ 8.8 pK units upon ligand binding. Its new p$K \approx 13.3$ is now larger than the pH of the solvent, and the group picks up a proton.

The $\Delta\text{p}K^{\text{Back}}$ term in Eq. (1) sums up the effects of altering electrostatic interactions between the group in question and the solute 'background' charges surrounding it. These effects can be further broken down into two contributions: interactions of the group with the ligand partial charges, Fig. 1*b*, and altered interactions with partial charges within the receptor protein following conformational rearrangements in the receptor induced by formation of the complex, (Fig. 1*c*). These re-arranegemnts may propagate towards a distant ionizable group, altering its local electrostatic environment (e.g., by changing the hydrogen bonding network) thereby causing the pK shift. Note that alterations of hydrogen-bonding networks, interactions with molecular quadrupoles, etc. are included in $\Delta\text{p}K^{\text{Back}}$. For example, a glutamic acid originally partially buried in the binding pocket of the receptor may already have its pK shifted up due to partial desolvation, so that the group is protonated and is net neutral at pH = 7.0. On complexation, however, the extra cost of interaction of the group's proton with a nearby positive charge on the ligand may become larger than the cost of releasing the proton back to the solvent. The group will release the proton to become negatively charged, (Fig. 1*b*).

As we shall see below, all of the three scenarios shown in Fig. 1 are realized in real protein–ligand complexes.

Corrections to the binding affinity: Regardless of the specific mechanism, ionization state changes upon ligand binding cause the binding free energy to become pH-dependent. One of several equivalent expressions (Jensen, 2008) for ΔG (pH) in this case is

$$\Delta G(\text{pH}) = \Delta G_{\text{ref}} - k_B T \ln\left(\sum_{i=\text{Asp, Glu, His, CT}} \frac{1+10^{pK_i^C - \text{pH}}}{1+10^{pK_i^F - \text{pH}}}\right) - k_B T \ln\left(\sum_{i=\text{Arg, Lys, Cys, Tyr, NT}} \frac{1+10^{\text{pH} - pK_i^C}}{1+10^{\text{pH} - pK_i^F}}\right), \quad (2)$$

where indices 'C' and 'F' denote the complexed and the free states of the receptor and ligand, respectively (Mason & Jensen, 2008). The first term in the above equation is the binding free energy in the reference state of protonation that corresponds to all Asp, Glu, His (no distinction between tautomers is made), and C-terminal (CT) residues held deprotonated, while all Cys, Tyr, Lys, Arg, and N-terminal (NT) residues are protonated. This particular state of protonation is a common, albeit not perfect, assumption in many molecular modeling studies. Note that this state does not necessarily occur naturally at any given pH. The remaining terms in Eq. (2) are the pH-dependent correction to the reference binding free energy. In a way, the correction can be taken as a measure of the (pH-dependent) error associated with the use of the standard

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

protonation state. As we shall see below, both experiment and computation have demonstrated that corrections to the binding free energy that arise from protonation state changes can be as large as the binding free energy estimated without taking the protonation state change into account. Examples of calculations based on Eq. (2) can be found in (Mason & Jensen, 2008; Palpant *et al.* 2010).

The mechanisms of binding-induced pK changes shown in Fig. 1 remain qualitatively the same even if the ionizable groups can no longer be approximated as non-interacting, but a more complex formalism than that based on Eq. 1 or 2 is needed for quantitative prediction of pK shifts. Namely, for a receptor and ligand with N and n ionizable groups respectively, all $2^{(N+n)}$ possible protonation micro-states of the protein–ligand system needed to be considered (Bashford & Karplus, 1990; Kieseritzky & Knapp, 2008b; Misra *et al.* 1998; Ullmann & Knapp, 1999), and proper Boltzmann averages computed. To handle the combinatorial complexity of the multiple protonation problem, various approximations are used, e.g. Monte-Carlo-based sampling (Beroza *et al.* 1991), or approaches that single out small clusters of strongly interacting groups (Gilson, 1993; Myers *et al.* 2006). These methods are now available in many packages, including web servers (Gordon *et al.* 2005; Kieseritzky & Knapp, 2008b). Once titration curves of the protein and ligand have been obtained, a general thermodynamic linkage relationship (Tanford, 1970) can be used to estimate the pH-dependent correction to the binding energy:

$$\Delta G(\mathrm{pH}) = \Delta G(\mathrm{pH}_{\mathrm{ref}}) + k_{\mathrm{B}} T 2.3 \int_{\mathrm{pH}_{\mathrm{ref}}}^{\mathrm{pH}} \Delta Q(\mathrm{pH})\, d\mathrm{pH}, \tag{3}$$

where ΔQ(pH) is the difference between the net charge of the complex and the corresponding free protein and the ligand at the given pH, $\Delta Q(\mathrm{pH}) = Q_{\mathrm{complex}}(\mathrm{pH}) - Q_{\mathrm{protein}}(\mathrm{pH}) - Q_{\mathrm{ligand}}(\mathrm{pH})$. The reference $\mathrm{pH}_{\mathrm{ref}}$ can be chosen arbitrarily: often, the choice is dictated by computational convenience. For example, at extreme pH values the protonation states of both the complex and the unbound receptor and ligand are obvious. Note that this type of reference state (definite pH) is different from that used in Eq. (2) (definite state of protonation). It is worth mentioning that the same approach is valid for investigating pH-dependence of protein stability, if one replaces bound and unbound states with folded and unfolded states. The approach dates back to Tanford (Tanford, 1970) and Linderstrom-Lang (Linderstrom-Lang, 1924). Examples of calculations based on Eq. (3) can be found in (Alexov, 2004; Misra *et al.* 1998; Mitra *et al.* 2011; Yang & Honig, 1993). An alternative approach not based on Eq. (3) or (2) was recently developed and explored in the context of protein–ligand docking (Park *et al.* 2011). Note that in general, protein–ligand binding involves complex interplay between multiple protonation, receptor and ligand conformation states, considered on the same footing, without any assumptions of weak coupling. A Monte-Carlo simulation package developed to handle this very general case is now available (Ullmann & Ullmann, 2012). The approach is based on a full microstate description of the receptor and an average description of its surroundings in terms of chemical potentials. The receptor can be modeled in great detail, including conformational flexibility and many binding sites that can bind different ligand types. It should be mentioned that titration curves of strongly interacting groups can sometimes deviate so much from the standard sigmoidal (Henderson–Hasselbalch) shape, that the usual notion of pK as mid-point of the curve may no longer be meaningful (Bombarda & Ullmann, 2010; Onufriev *et al.* 2001). Coupling to ligand binding may produce non-trivial cooperative effects (Onufriev & Ullmann, 2004), even without any conformational change in the receptor protein (Spassov & Bashford, 1998). In fact,

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

the latter possibility was not included in Fig. 1: proton transfer to a distant group can occur via a concerted change of protonation states in a strongly coupled cluster (Onufriev *et al.* 2003). However, based on available data presented in the following sections, this scenario, if it occurs, is definitely not among the common ones in protein–ligand binding.

3. Magnitude and prevalence of protonation and p*K* changes in receptor proteins upon ligand binding

3.1 What is known from experiment?

Methods that can characterize changes in protonation states of the protein and its ligand upon binding can be divided into two broad classes: 'macroscopic' and 'atomistic'. Below we briefly outline each of them.

The first (macroscopic) class is represented almost exclusively (Bruylants *et al.* 2007) by isothermal titration calorimetry (ITC) (Baker & Murphy, 1996; Perozzo *et al.* 2004). In this technique, the overall heat of the complex formation is recorded as the ligand is being added to the buffered receptor solution at constant temperature. To detect whether proton transfer from the buffer to the complex has occurred upon binding, the heat of complex formation is measured under different buffer conditions. The stoichiometry of the protonation reaction can also be obtained. Based on thermodynamics, the approach has a clear advantage of being general; it gives direct access to binding affinities. However, by itself, the method can only tell whether a *net* proton uptake/release has occurred upon binding, but it can not determine which group(s) became protonated or deprotonated. Specific assignments have to be made based on additional information such as computation (Steuber *et al.* 2007), mutagenesis (Qasim *et al.* 1995), ultra-high resolution X-ray structures (Steuber *et al.* 2007), or NMR data (Blundell *et al.* 2007; Bruylants *et al.* 2007). Relatively small number of ITC studies of proton transfer in complex formation have been performed to date, owing to their experimental complexity (Czodrowski *et al.* 2007). For small molecule ligands, readily available literature references are limited to very few cases in complexes with enzymes or antibodies (Czodrowski *et al.* 2007; Gómez & Freire, 1995; Murphy *et al.* 1993; Steuber *et al.* 2007; Xie *et al.* 1997). In some of these cases, binding-induced protonation state change in the receptor protein was reported. For example, for serine proteases trypsin and thrombin, binding of a series of small molecule ligands at pH = 7.8 was investigated (Czodrowski *et al.* 2007): formation of 3 out of 12 complexes was associated with a net proton uptake from the buffer. For protein ligands, a summary of experimental data for five different protein–protein complexes involved in electron transfer is given in (Kresheck *et al.* 1995): at near neutral pH, full proton release was observed in one case, and in two cases a partial ($\sim$50%) release was reported. The relative error of the measured net proton uptake can be expected to range from 7 to 20% (Czodrowski *et al.* 2007; Kresheck *et al.* 1995). In a separate study, Horn *et al.* (Horn *et al.* 2003) investigated ovomucoid third domain binding to subtilisin Carlsberg. It was concluded that one residue changes p*K* appreciably upon binding, from 8.5 to 6.7. At a pH in the middle of the range, a significant deprotonation of the residue occurs. In another work, a very large p*K* shift, from 4.4 to 8.7, was reported for a Glu residue in Streptomyces griseus proteinase upon complexation with with turkey ovomucoid third domain protein (Qasim *et al.* 1995). In fact, one can apply the general Eq. (3) to calculate the corresponding net proton uptake/release practically from any experimentally documented pH-dependent binding data. Table 1 provides an example of such experimental data, including the

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

Table 1. *Cases of receptor–ligand complexes experimentally shown to have pH-dependent binding, which was used to calculate net proton uptake/release (the ΔQ (Eq. (3))) at specific pH. Data are taken from* (*Buckle* et al. *1994; Buczek* et al. *2002; Krowarsch* et al. *1999; McPhalen* et al. *1985; Schreiber & Fersht, 1993; Schreiber & Fersht, 1995*). *Where the experimental 3D structure is unavailable, the closest homolog is listed.*

Name	PDB complex	Receptor	Ligand	ΔQ (at pH)
beta-trypsin-BPTI	2PTC	1TGN	6PTI	−0.68(pH = 5.5)
Chymotrypsin-BPTI	1CBW	1YPH	6PTI	−0.48(pH = 5.5)
Salmontrypsin-BPTI	1BZX	1BIT	6PTI	−0.82(pH = 5.5)
Barnase-Barstar	1BRS	1A2P	1A19	−0.53(pH = 6.0)
Subtilisin Carlsberg-OMTKY3	1ROR 1YU6	1SBC	1OMU	−0.5(pH = 7.0)
Porcine pancreatic elastase-OMTKY3	1PPF	3EST	1OMU	−0.61(pH = 5.5)

corresponding PDB ID of the structures of the complex, the receptor and the ligand. It should be mentioned that with respect to the proton uptake/release, the collected data set is particularly interesting, since the protonation changes are non integers, and therefore cannot be modeled by manually changing the ionization state (neutral or charged) of any single amino acid. Instead, the modeling would require an ensemble representation where the proton uptake/release may be distributed over several amino acids, and their protonation states may vary within the ensemble. In the simplest case scenario, the ensemble can be represented by a single structure, but different protonation states, or in a more realistic scenario, one should generate an ensemble with both structural and protonation variants. Such ensembles are generated naturally within constant pH Molecular Dynamics (Khandogin & Brooks, 2006; Lee *et al.* 2004; Mongan *et al.* 2004; Swails *et al.* 2009). For example, in the approach developed by Mongan *et al.* (Mongan *et al.* 2004), protonation states are accepted or rejected on the fly, according to a Metropolis criterion, during the course of the simulation.

For nucleic acid ligands, proton uptake by receptor protein upon complexation was reported too (Kozlov & Lohman, 2000; Mishra *et al.* 2009). In the case of a single-stranded DNA binding to an *Escherichia-coli* protein (Kozlov & Lohman, 2000), the proton uptake was observed over a wide pH range (5.0–10.0). Importantly, most of the calorimetry studies that reported thermodynamic parameters of complex formation, such as ΔH or ΔC_p, also included corrections from protonation-binding linkage. In some cases (Kozlov & Lohman, 2000), these corrections were as much as half of the total value of the measured parameter.

'Atomistic' techniques such as X-ray crystallography, neutron-scattering, and NMR can, in principle, pinpoint positions of individual protons in both the bound and unbound states of the complex. In practice, each of the methods has severe limitations when applied to the study of protonation and pK changes upon complex formation–only very few relevant data points exit to-date. The resolution of the absolute majority of currently available X-ray structures is still too low to determine hydrogen positions reliably. While some ultra-high (better than 1 Å) resolution structures are now available, one needs both the complex and the unligated protein structures to tell if a protonation state has indeed occurred upon binding. Our recent (Aguilar *et al.* 2010) search of public databases for pairs of protein–ligand complexes along with their matching unligated proteins identified such pairs only at 1.2 Å or lower resolution, which is still not enough for hydrogen assignment from the electron density. While protons are readily available in neutron-scattering structures, these are so few that we were not able to find any (complex +

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

Table 2. *Experimental pK changes in several proteins upon ligand binding. For each group, ΔpK = pK (complex) – pK (unligated receptor). Experimental references are given in ( ): (a) (Smith* et al. *1996; Wang* et al. *1996), (b) (Xie* et al. *1997), (c) (Yamazaki* et al. *1994), (d) (Cassidy* et al. *1997), (e) (Joshi* et al. *2001), (f) (Stranzl* et al. *2004)*

Receptor	Ligand	Group	ΔpK	PDB w/o ligand	PDB of complex
HIV-1 protease	Pepstatin	ASP25	> +2.2 (a)	3HVP	1HPX
HIV-1 protease	Pepstatin	ASP25′	< −1.5 (a)	3HVP	1HPX
Plasmepsin II	Pepstatin	ASP34/214	+0.1 (b)	1PFZ	1SME
Plasmepsin II	Pepstatin	HIS164	+1.5 (b)	1PFZ	1SME
HIV-1 protease	DMP-32	ASP25/25′	+2.19 (c)	1HHP	1QBS
Chymotrypsin	N-Acetyl-DL-Phe-CF3	HIS57	+3.3 (d)	6GCH	6GCH
Xylanase	2FXb	GLU172	−2.5 (e)	1BVV	1BVV
Hydroxynitrile lyase	Thiocyanate	HIS235	+5.5 (f)	2YAS	2YAS

unligated protein) pair. NMR thus remains the key source of residue-specific p*K* in proteins, with a few hundred of experimental data points now available (Grimsley *et al.* 2009), often used to benchmark computational methods (Antosiewicz *et al.* 1994; Georgescu *et al.* 2002; Spassov & Yan, 2008) that predict p*K*. However, only a handful of these data points correspond to pairs of protein–ligand complexes and their matching unligated receptors. Protein–protein complexes continue to present challenge to present day NMR, most likely due to their relatively large size. Several recent computational studies reported relevant compilations from experimental literature (Aguilar *et al.* 2010; Bas *et al.* 2008), see a representative sample in Table 2 from (Aguilar *et al.* 2010). The only selection criterion used in that compilation was availability of reasonable quality X-ray structures for both the complex and the unligated protein, including the requirement for no missing residues in the middle of the amino-acid sequence. Most of the p*K* s listed in Table 2 were determined by NMR.

The immediate conclusion from Table 2 is that appreciable, and sometimes very large p*K* shifts do occur upon binding. We stress that a p*K* shift, even a large one, does not necessarily entail a protonation state change of the group in question, unless the ambient pH is in-between the group's p*K* values before and after binding. However, in the case of the groups listed in Table 2, most of their measured p*K* values fall into the biologically relevant pH range (4–8), which means that a protonation state change can occur upon the complex formation under biologically relevant conditions. For example, HIS235 of hydroxynitrile lyase is unprotonated at pH = 7.0 (its p*K* = 2.5), while complexation with thiocyanate shifts its p*K* to 8.0, resulting in protonation of the group (Stranzl *et al.* 2004). Clearly, the energy cost of mis-assigning the ionization state of just this one group at pH = 7.0 would already be well above the typical error margin of experiments that measure ligand-binding affinity (Gilson & Zhou, 2007). However, one can immediately notice that the diversity of both the receptor proteins and their ligands in Table 2 is very limited, reducing generality of any conclusion that can be drawn from the data.

To summarize, various experimental techniques point unambiguously to large p*K* shifts and protonation state changes that can occur in proteins upon protein–ligand binding. The phenomenon is observed for different classes of biologically relevant ligands such as small molecules, other proteins, or nucleic acids. Corrections to thermodynamic parameters of the complex formation associated with these changes were large enough to warrant consideration

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

and inclusion. However, the scarcity of the experimental data makes it hard to estimate the overall prevalence of these effects within a reasonable statistical margin. For the same reason, it is impossible to extract, from the currently available experimental data alone, trends in how the binding-induced protonation changes may be distributed in the receptor and the ligand. These questions are important both for fundamental understanding of protein–ligand-binding process and for the development of computational methods that utilize atomistic protein structures to make predictions.

3.2 What do calculations tell us?

Given the scarcity of the experimental data, computational methods become particularly valuable in studying p*K* and ionization state changes in protein–ligand binding.

3.2.1 Earlier studies

Earlier studies, see (Brandsdal *et al.* 2006; Elcock *et al.* 2001; McDonald *et al.* 1995; Misra *et al.* 1998; Sims *et al.* 2005; Trylska *et al.* 1999) for a representative sample, made progress by detailed analysis of a small number of specific proteins, ligands, and ionizable groups. These works often reported large p*K* shifts and protonation changes in protein–ligand binding, and pointed to the need to take the associated energetic costs into account for accurate prediction of the binding properties. For example, McDonald *et al.* (McDonald *et al.* 1995), used continuum electrostatic calculations to investigate the titration behavior of the ionizable residues of the HyHEL-5-hen egg lysozyme complex and its individual components. Several residues of HyHEL-5 had p*K* values shifted upon binding by more than three pH units. Brandsdal *et al.* (Brandsdal *et al.* 2006) systematically investigated (using free energy calculations) the binding of BPTI with glutamic acid at the primary binding position to trypsin. The predicted p*K* shift of the key Glu residue upon binding was as large 9.8 units, yielding a p*K* of 14.3 in the complex. The resulting unusual neutral charge state of the Glu residue was found to be essential to obtain quantitative agreement with the experimental binding data. In another study, Spassov and Bashford (Spassov & Bashford, 1998) proposed an alternative mechanism for the cooperative binding of charged ligands to proteins based on electrostatic coupling to ionizable groups. The mechanism was investigated on a monomeric protein (calbindin D9k) containing two Ca^{2+}-binding sites and 30 ionizable groups. Alexov (Alexov, 2004) investigated the role of individual residues in proton uptake/release experimentally measured for three complexes involving aspartic proteases and pepstatin. It was demonstrated that the conformational changes induced by the binding contribute significantly to the calculated p*K* changes. In addition, it was pointed out that for correct calculations of the binding energy, one should model ionization and conformational changes within the same algorithm that calculates the energy.

3.2.2 Large scale studies

However, it was not until very recently that computational works based on statistically large sets–thousands of computed p*K* s–began to appear (Aguilar *et al.* 2010; Mason & Jensen, 2008; Kundrotas & Alexov, 2006). These recent studies have explored and quantified statistical trends in protein–ligand binding in addition to analyzing individual cases in detail. Perhaps the most intriguing finding that has emerged is that changes in p*K* and ionization states of titratable amino

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

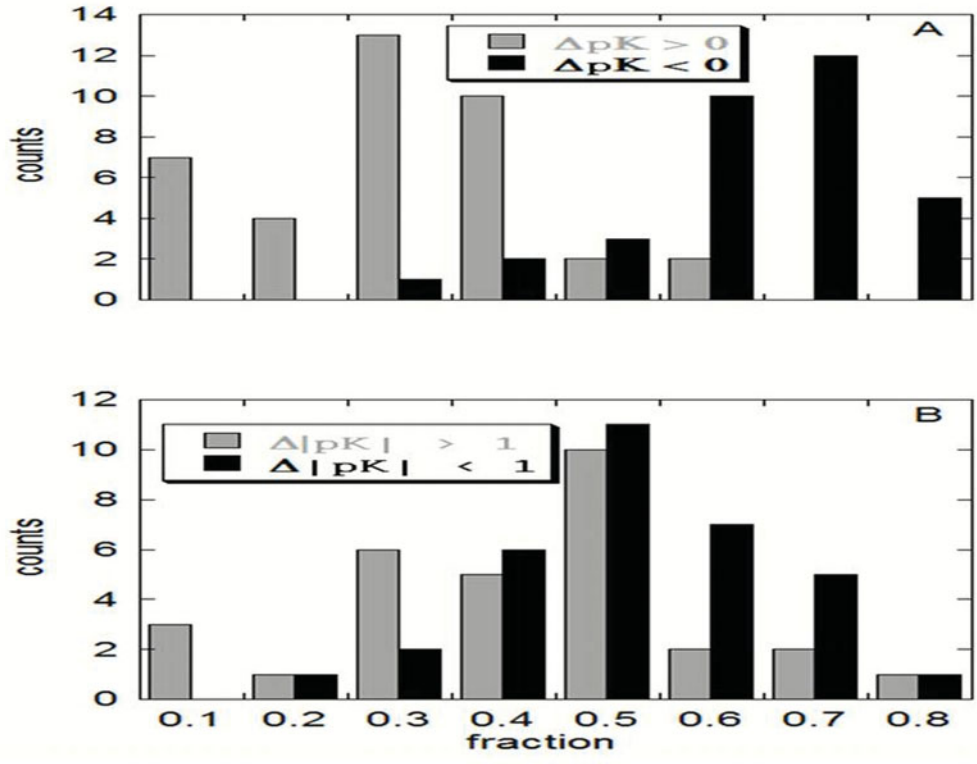


Fig. 2. The distribution of p*K* changes induced by protein–protein binding. (*A*) all p*K* changes, (*B*) large absolute p*K* changes *versus* small absolute p*K* changes. The horizontal axis shows the fraction of interfacial groups satisfying the corresponding p*K* shift, while the vertical axis indicates in how many (out of 37) protein complexes this shift occurred. Adopted from Kundrotas *et al.* (Kundrotas & Alexov, 2006).

acids occur quite commonly in proteins upon ligand binding. The exact statistics, however, varies appreciably, most likely due to variations in the methodology and data sets used.

Kundrotas *et al.* (Kundrotas & Alexov, 2006) investigated statistical and electrostatic properties of the interfaces for a set of 37 complexes with diverse functions extracted from a database of protein–protein complexes previously developed by the authors. Multi-conformer, continuum electrostatic p*K* calculations (Georgescu *et al.* 2002) were performed for all 37 selected complexes for the entire complex and for each of the monomers. The structures of the monomers were kept as they were in the complex, i.e., no conformational changes during the complex formation were considered in this study. It was found that 50% of interfacial acidic groups have their p*K* shifted substantially, that is by more than one pH unit relative to the unbound state. The distribution of the p*K* shifts is presented in Fig. 2; panel (A) shows that the vast majority (about 80%) of the interfacial acidic groups have a net stabilizing contribution to the association of the complex (dark bars). However, still about 20% of the groups have a destabilizing contribution (white bars). Panel (B) of the same figure shows the distribution of the large p*K* shift, which peaks at 50%, whereas the distribution of small p*K* shift results in a broad maximum ranging from 20 to 50%. The finding that 50% of interfacial acidic groups have their p*K*s shifted by more than one pH unit indicates that individual charged groups, in principle, can contribute notably into the protein binding. Interestingly, the acidic groups were predicted to behave

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

differently from histidines (note that pK shifts associated with Arginines and Lysines were not considered to be important for proton uptake/release, because the corresponding pK s are typically outside the experimental/physiological range of pH). It was demonstrated that the complex formation tends to favor acidic pK s by making them lower in the bound state compared with the unbound state. In contrast, histidine pK s tend to be lower in the complex, indicating that the bound state does not provide enough favorable interactions.

In a separate work, Mason *et al.* (Mason & Jensen, 2008) used fast empirical methodology (PROPKA (Bas *et al.* 2008), web-based) to compute pK values of ionizable residues in 75 protein–protein complexes and their corresponding free forms. Two main types of computations were described. In the first type, the individual (unligated) protein structures were taken directly from the structure of the complex just like in the earlier work by Kundrotas *et al.* (Kundrotas & Alexov, 2006). However, the other set of computations was based on separate X-ray structures determined for the proteins in the absence of the other proteins. This second type of calculation included effects of binding-induced structural rearrangement on the pK values. With the inclusion of these effects, at least one residue per protein was found to completely change its protonation state in 28% of the complexes. In the computation of the first type (no binding-induced rearrangements), the corresponding number was 12%. In both types of calculations, protonation state change prevalence increased 2-to 3-fold if partial protonation events (probability $\geqslant 0.5$) were also counted. Change in *net* protonation state of the complex relative to the uncomplexed proteins, that is proton uptake/release from solution, occurred in about 30% of the cases. The distinction may be important, as protonation state changes in each of the proteins may occur, e.g., through proton transfer between the proteins, without the net proton uptake/release from the buffer solution. Over-all, changes in protonation states upon complex formation were found to be most pronounced for the formation of complexes of large and conformationally flexible proteins; in some of these complexes multiple protonation state changes were observed. In the same work (Mason & Jensen, 2008), protonation–conformation coupling was included as a correction to the predicted binding energy for all 75 complexes at pH7 relative to the standard protonation state, Eq. (2). With binding-induced rearrangements considered, the correction was at least 1.4 kcal/mol (corresponding to one order of magnitude change in the binding constant at room temperature of 298 K) in 45% of the complexes. Without binding-induced rearrangements, the corresponding fraction was 23% of the complexes. In both cases, some of the corrections were larger than 4.2 kcal/mol by absolute value, affecting the binding constant by more than three orders of magnitude.

In another recent work, Aguilar *et al.* (Aguilar *et al.* 2010) examined, on an equal footing, pK and protonation state changes in protein–ligand binding for three types of complexes: protein–protein, protein–small molecule, and protein–nucleic acid. Standard continuum electrostatic pK calculations (Bashford & Karplus, 1990) (H++ (Gordon *et al.* 2005) web-based) were used, followed by a careful examination of the associated uncertainties–the 'noise' level in pK calculations of this type. The study analyzed pK changes upon ligand binding for a total of almost 6000 ionizable residues in 20 protein–protein, 20 protein–small molecule, and 20 protein–nucleic acid complexes, selected on the basis of reasonably high structure quality. For each protein–ligand complex used in the analysis, the experimentally determined X-ray structure of the corresponding unligated protein was also available. The results of the calculation, which included effects of binding-induced structural rearrangement on the pK values, are as follows. In 60% of all protein–small molecule, 90% of all protein–protein, and 85% of all protein–nucleic acid complexes, at least one ionizable residue was found to completely change its charge state upon

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

Table 3. *Percentages of ionizable residues in receptor proteins that exhibit substantial pK change ($|\Delta pK| > 1$) upon ligand binding in the biologically relevant pH range from 4 to 8. Adopted from Aguilar* et al. (*Aguilar* et al. *2010*)

Complex type	Minimum (%)	Average (%)	Maximum (%)
Protein–protein (20)	1.5	8.7	16.7
Protein–small molecule (20)	2.1	5.7	12.5
Protein–nucleic acid (20)	5.6	12.3	24.4

ligand binding at neutral pH = 7.0. Over-all, 9% of all ionizable residues (averaged over all complex types) presented a substantial p*K* change ($|\Delta pK| > 1.0$) due to ligand binding in the biologically relevant pH range from 4 to 8. However, averages do not tell the whole story in this case, as the fraction of ionizable sites with substantial p*K* changes upon binding differs greatly between receptor proteins (Table 3). Overall, substantial p*K* changes due to binding were found to occur most often in protein–nucleic acid complexes and least often in complexes of proteins with small molecules. Not surprisingly, for all ligand types the effect is the strongest in the immediate binding interface, (Fig. 3).

3.2.3 Relative prevalence of various physical causes of p*K* shifts in ligand binding

Computationally, one can trace (Aguilar *et al.* 2010) the origins of the p*K* changes to the three electrostatic mechanisms discussed above, (Fig. 1). In the binding interface region, direct electrostatic mechanism – which includes desolvation and direct electrostatic interaction with the ligand–is predominant in effecting p*K* changes in protein–nucleic acid complexes, (Fig. 3). In protein–(small molecule) and protein–protein complexes, both the binding-induced structural rearrangements ('induced-fit') and direct electrostatic effects play equally important roles in the interface region. The effect of direct electrostatic perturbation is generally 'short ranged', only causing significant p*K* changes in residues located within $\sim$6 Å away from the ligand in the majority of protein–protein and protein–small molecule complexes, and weakly extending to 12 Å from the ligand in the case of binding of highly charged nucleic acids. At the same time, p*K* perturbations due to structural re-arrangements propagate well beyond the immediate interface region (Aguilar *et al.* 2010; Bas *et al.* 2008).

3.2.4 Long-range coupling between ligand binding and protonation

As seen in Fig. 3, a significant number of substantial p*K* changes apparently occurs well outside the binding interface, sometimes as far as 24 Å away from the ligand, and even further (Aguilar *et al.* 2010). For protein–small molecule complexes a similar observation was also made by Bas *et al.* (Bas *et al.* 2008), on roughly the same sample size. A natural question arises if these very distant changes are real, or artifacts of the methodology and inevitable structural imperfections? Note that even though the average systematic error of current ΔpK estimates is probably no larger than the threshold ($|\Delta pK| > 1.0$) typically used to filter meaningful p*K* changes, there still exists a non-zero probability for any such p*K* change to be a statistical false positive. This is because the underlying methods are far from perfect: the p*K* values computed from atomic-resolution input structures are known to be sensitive to structural details (Nielsen & McCammon, 2003). Uncertainties in the input X-ray structure will be propagated into errors

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

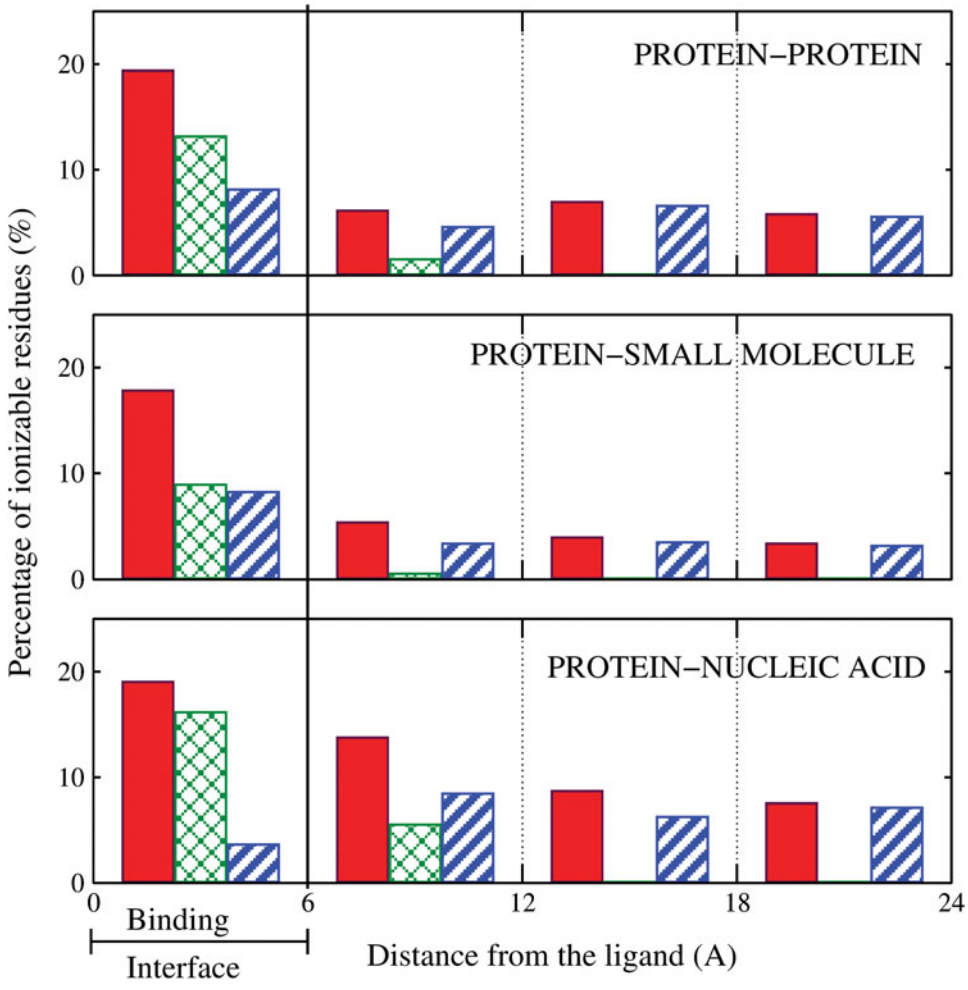


Fig. 3. Distance distribution of ionizable residues with substantial p*K* shifts upon ligand binding in biologically relevant pH range, by the specific mechanism causing the shift. Distance is to the bound ligand. The three mechanisms, labeled a, b, and c, are described in Fig. 1. Solid red bars: all three mechanisms combined (a + b + c). Cross-hatched green bars: direct electrostatics (a + b). Striped blue bars: binding-induced structural rearrangements in the receptor(c only). The percentage reported for each region is relative to the total number of ionizable residues located in that region. Adopted from Aguilar *et al.* (Aguilar *et al.* 2010).

in the computed p*K*s. To estimate this type of error, Aguilar *et al.* analyzed the variation in computed p*K* values that result from structural deviations between several available X-ray structures of the same protein. In an ideal world, these differences should be zero, but in reality they are sometimes larger than 1 p*K* unit–events that should be counted as 'noise' in distributions such as that in Fig. 3. As it turns out (Aguilar *et al.* 2010), the 'noise level' in Fig. 3 is lower than the signal everywhere except for protein–small molecule complexes in the region far from the binding interface. A more detailed follow-up statistical analysis could neither confirm nor reject, for the sample size used in Aguilar *et al.,* the possibility of statistically substantial p*K* changes outside the binding interface in protein–small molecule complexes.

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

Table 4. *Subcellular compartments and their characteristic pH.*

Subcellular compartment	Characteristic pH
Nucleus	7.7
Endoplasmic reticulum,	7.1
Golgi apparatus	6.6
Mitochondrion	7.5
Vacuole	5.3
Cytoplasm	7.3
Lysozome	4.8
Peroxisome	8.2

One may also wonder if all residues with substantial p*K* change in regions far from the ligand belong to a small group of complexes that are 'unique' in some way. That is not the case (Aguilar *et al.* 2010): the occurrence of substantial p*K* changes far from the binding interface appears to be a general property of the ligand binding process. Moreover, there appears to be no single dominant type of ligand-induced structural perturbation that alters local electrostatic interactions ultimately responsible for the 'distant' p*K* changes. Those identified in Aguilar *et al.* were described previously in different contexts (Betts & Sternberg, 1999; Dan *et al.* 2009; Gerstein *et al.* 1994), and by no means exhaust the list of possible structural perturbations associated with ligand-induced p*K* changes (Bas *et al.* 2008; Maler *et al.* 2000).

In summary, the 'long-range p*K* coupling' effect is real in protein–protein and protein-DNA complexes, but the results are still inconclusive for protein-(small molecule) binding where the effect, if it exists, is the weakest. The long-range effect might play a role in allosteric regulation: subtle changes on one site of the receptor protein may affect ligand binding properties on another, distant site. For example, strong long-range site-site interactions were experimentally demonstrated to be important in a family of Ca^{2+}-binding proteins (Maler *et al.* 2000).

4. Interplay among subcellular pH, pH-optimum of ligand binding, and pH of the corresponding reaction

The (negative log of) concentration of hydrogen ions, typically referred to as pH, is an essential characteristics of the cell and living organisms. At the cell level, different cell compartments (organelles) have different characteristic pH as outlined in (Chan *et al.* 2006; Kulichikhin *et al.* 2009). A short sample list of characteristic subcellular pH values is shown in Table 4. In summary, at the cellular level, macromolecules and their complexes can be exposed to pH ranging from value as low as 4.8 and as high as 8.2. Similar wide range of pH can be found at the level of the entire organism, where the pH differs greatly between vital organs and systems. In some organs, the pH is maintained with high precision, while it may vary greatly within others. For example, in human, the normal blood pH is about 7.3; even a small deviation from this optimal value, by as little as fraction of pH unit, can be fatal. For example, if the blood pH drops below 7.0 (acidosis), the effect is depression of the central nervous system resulting in coma. In the opposite case, if the pH increases above 7.8 (alkalosis), the effect is overexcitability of the nervous system that results in convulsions. In contrast, in other parts of the human body, the pH can vary greatly, as for example in the digestive track where it can range from 1.5 up to 7.0 (Davenport, 1996). Broad range of pH, from 4.5 up to 8.0, occurs in kidney (Guyton, 1976). All these

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

Table 5. *Examples of various pH-optima of enzyme activity.*

Enzyme	pH-optimum
Lipase (pancreas)	8.0
Lipase (stomach)	4.0–5.0
Lipase (castor oil)	4.7
Pepsin	1.5–1.6
Urease	7.0
Invertase	4.5
Maltase	6.1–6.8
Amylase (pancreas)	6.7–7.0
Trypsin	7.8–8.7

examples indicate that biological macromolecules and the corresponding complexes function at very different pH conditions, and depending on their function in terms of subcellular or tissue localization and trafficking, they should have evolved to either adapt to the corresponding pH or sense it. (see, e.g., review by Garcia-Moreno (Garcia-Moreno, 2009)).

From the evolutionary standpoint, the most important factor is the optimization of biochemical reactions to provide better chances for survival and reproduction. Such evolutionary pressure does not necessary have to be linked with optimized stability of individual macromolecules and affinity of their complexes. However, indirect link may exist since in most cases the optimal biochemical reaction may invoke optimal macromolecular interactions or optimized structural integrity of individual biomolecules. When discussing the optimum pH we will follow the existing literature and term the pH of optimum of reaction, stability or binding, as the pH-optimum (Mitra *et al.* 2011; Talley & Alexov, 2010; Yan & Wu, 2011; Yang *et al.* 2008). In terms of pH-optimum of activity, the pH-optimum varies for different enzymes and biological reactions (Chang *et al.* 2009; Commet *et al.* 2012; Reeves Gibbs *et al.* 2010). Several examples of various pH-optima of the corresponding reaction are shown in Table 5 to indicate the broad range of pH conditions in biology. These biochemical characteristics were conveniently collected into databases, as for example BRENDA database (Chang *et al.* 2009; Schomburg *et al.* 2004), which makes it possible to address questions about plausible correlations among pH-optimum of binding, activity, and stability in a statistical manner.

Many different approaches exist to predict pH-optimum of binding, but arguably the most convenient is to use Eq. (3) introduced above. One can simply calculate the protonation states of the complex and the unbound monomers as a function of pH, and their difference will represent the net pH-dependent proton uptake/release (ΔQ) induced by the binding. Assuming that the optimum (minimum) of the binding free energy occurs at some point within the pH interval where the receptor and the ligand are stable, $\partial \Delta G\,(\mathrm{pH})/\partial \mathrm{pH} = 0$ at the pH-optimum[3]. The corresponding net proton uptake/release must then equal zero at this pH according to Eq. (3). The above strategy was used in several cases to explore the idea of the similarity of biophysical properties of proteins known to bind to each other (Kundrotas & Alexov, 2006). In order to interact, two proteins must co-occur spatially and temporally. Recently, the possibility of co-evolution of pH-optimum of protein stability and pH-optimum of protein interactions was

[3] With the caveat that there may be cases for which the condition $\partial \Delta G\,(\mathrm{pH})/\partial \mathrm{pH} = 0$ is not satisfied because $\Delta G\,(\mathrm{pH})$ does not reach an extremum inside the relevant pH interval. For example, $\Delta G\,(\mathrm{pH})$ could be a monotonic function within the interval.

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

studied and it was demonstrated that they are indeed correlated (Mitra *et al.* 2011). The idea was used by other researchers to match entries in two databases: the protein–protein interactions and subcellular location and substantial overlap was reported (Shin *et al.* 2009).

These observations suggest that, perhaps, native (or sometimes referred to as physiological) macromolecular interactions should not involve proton uptake/release. Indeed, if the pH – optimum of binding and the pH of the environment where the complex is formed are the same or quite similar, then as discussed above, the net charge difference between the complex and the unbound molecules should be zero at pH-optimum. This does not, however, mean that protonation states are not important for the binding or that they do not change upon it. The only quantity that does not change in physiological (native) interactions is the net charge difference between the complex and unbound monomers. Similar observation was reached in another study which modeled the conformational change observed upon ligand binding and phosphorylation for the cAMP-dependent protein kinase (Bjarnadottir & Nielsen, 2010). It was conjectured that the binding of ligands requires the smallest proton release/uptake around pH 7.0–8.0. Physiological pH for receptor–ligand binding is therefore the pH value where the energy difference between the bound and unbound states reaches its minimum, that is at the pH-optimum of binding. Moreover, a recent work implied that proton transfer between histidines across the interface of the complex changes the protonation states of the bound monomers, but results in zero net proton uptake/release, and thus no pH-independence of the binding (Wang *et al.* 2012).[4] These suggestions may appear to be at odds with the conclusion from the experimental data and large-scale computational studies reviewed above. The latter indicate that net proton uptake/release due to binding is not an uncommon phenomena in protein–ligand binding, while the above discussion seem to suggest that perhaps the receptor–ligand complexes have evolved to avoid proton uptake/release. There is, however, no immediate controversy. The statement about the pH-optimum suggests that each receptor–ligand complex has evolved toward particular pH-optimum, which is related to its function and subcellular and tissue localization. However, experiments discussed above are typically either performed at neutral pH, or over a range of pH. Obviously, if the experiment is not done precisely at pH-optimum, binding-induced proton uptake/release may be observed. Similarly, the large-scale computational studies we have discussed are typically done on diverse sets of receptor–ligand complexes (presumably each having its own distinctive pH-optimum), and the protonation state changes are reported at a particular pH, e.g., pH = 7. The implication is that had these studies been performed at pH-optimum for each complex, the reported prevalence of protonation states changes may have been lower.

5. Improving docking and scoring outcomes

One of the most recognized approaches to predict 3D structures of receptor–ligand complexes is *ab-initio* docking (in case of macromolecular binding) or virtual screening/scoring (in case of

[4] The reader is reminded that ΔQ in Eq. (3) is the difference of the net charge of the complex and the unbound receptor and the ligand. Thus, any protonation state change that causes only internal proton reorganization/exchange within the complex or unbound receptor and ligand will result in no change of the net charge of the system, and thus no pH-dependence of the binding. Likewise, an unusual protonation state, as for example of a given ionizable group that may have unusual *pK*, and is neutral in both the unbound state and in the complex (Alexov, 2004), will result in altered net charges of both the complex and the unbound receptor or ligand, but the ΔQ will be zero.

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

small molecule binding to a macromolecule). Namely, given the structures of unbound receptor and ligand, one utilizes computational methods to bring them together to predict the 3D structure of the complex. It is outside the scope of this work to review docking and screening methods, but some representative examples can be found in (Carter *et al.* 2005; Schneidman-Duhovny *et al.* 2005; Schueler-Furman *et al.* 2005; Spitzer & Jain, 2012; Trott & Olson, 2010). Although docking and screening simulations focus on determining the 3D structure of the complex without modeling the binding pathway, the alternative approach known as binding simulations (McCammon, 1998) reveals all association details including the binding constant of the 3D complex structure (Gohlke *et al.* 2003). Further enlargement of the concept of the binding addresses the issue of modeling supramolecular assemblages (McGuffee & Elcock, 2006). All these approaches differ in their methodology, force fields, and assumptions made. However, despite the differences between the above-mentioned docking algorithms, two major problems are common: (a) how to treat conformational changes induced by the binding and (b) how to model protonation changes caused by the complex formation? Obviously, these two issues are coupled and, ideally, have to be modeled within the same numerical protocol. While the importance of conformational changes induced by the complex formation is very well recognized, and significant efforts have been invested to address it, the possibility that binding may cause protonation changes is rarely incorporated into the docking and the screening models. As discussed in the previous sections, it is quite likely that burial of an ionizable group at the binding interface of the complex can cause its protonation/deprotonation, as demonstrated both experimentally and theoretically. If this possibility is not considered, then the most probable candidate complexes causing such ionization changes are likely to be rejected by the scoring function because of unfavorable interactions. The side-chain orientations will also be incorrect, even after energy minimization, if the ionization states are not assigned according to the new environment at the interface. However, as detailed in literature (Georgescu *et al.* 2002), side-chain orientation and calculations of the ionizable states should ideally be done simultaneously in the same protocol since they are interdependent. The ionized and neutral form of amino acids differ by a single proton, so from geometric perspective their side chains are almost the same and if the ionization/de-ionization does not cause significant rearrangement, the effect of protonation states on the overall shape can be neglected. However, the charge states affect the scoring function. Typically, the scoring function is derived either from statistical analysis of known receptor–ligand complexes, or it is developed based on some set of empirical rules. Many scoring functions of the first type are based on statistical potentials that depend on amino acid composition of the interfaces and pairing frequencies across the interfaces (Jones & Thornton, 1997; Tsai *et al.* 1997). Such scoring functions can also factor in the significant electrostatic and hydrogen-bonding complementarity observed at the interfaces of hetero complexes (McCoy *et al.* 1997). However, standard protonation states for all ionizable groups are typically assumed. Obviously, the resulting scoring function will not favor unusual protonation states, and wild-type receptor–ligand complexes involving non-standard ionization states will be scored unfavorably. Fully empirical scoring functions, on the other hand, should be able to handle cases with unusual protonation states, at least in principle. However, they too will most likely favor ionized states of titratable groups, if they do not properly account for the presence of the solvent, that if no desolvation penalty (Fig. 1) is included.

We illustrate the importance of taking into account protonation states for correct prediction of the binding mode using one of the most popular software packages for *ab-initio* docking, ZDOCK (Pierce *et al.* 2007). Two cases are considered: a case of hetero- and a case

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

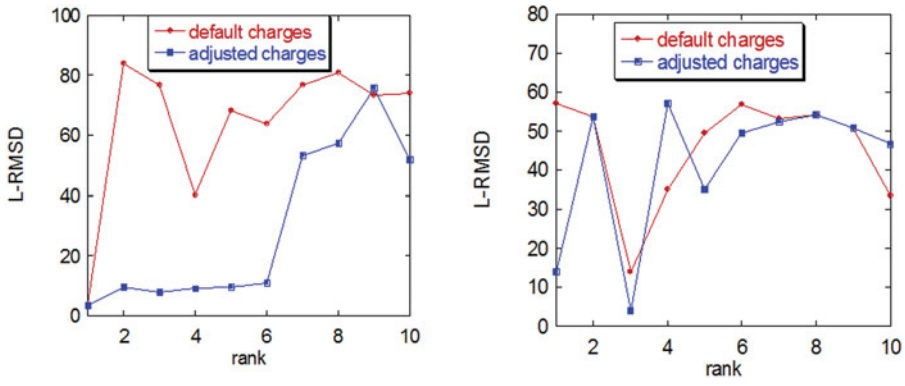


Figure 4. L-RMSD of the ten best predictions obtained with ZDOCK with standard (default) protonation states and with adjusted protonation states (see text for details). Left panel: homo-complex, PDB ID 1sji. Right panel: hetero-complex, PDB ID 1gx2.

homo-complex, namely PDB codes 1sji and 1xg2 (Fig. 4). For 1sji, Lys 49, Glu 117, and Glu 256 were predicted to be neutral at pH = 7.0 in the complex according to p*K* calculations performed with MCCE package (Georgescu *et al.* 2002). In the case of 1xg2, Asp B 130 was predicted to be neutral, whereas His A 137 was found to be ionized. The charges of these groups were adjusted accordingly[5], and docking predictions are made (Fig. 4). For assessment of the predictions, we use the L-RMSD (Mendez *et al.* 2005), which is the ligand RMSD as defined in the CAPRI experiment (Lensink *et al.* 2007).

It can be seen that the charge adjustment dramatically changes the quality of the predictions for the homo-complex (1sji), with the adjusted charges (presumably the correct protonation state) resulting in better predictions compared to the default charges corresponding to the standard protonation state (Fig. 4, left panel). For the hetero-complex, the results are shown in the right panel of Fig. 4. Again, it can be seen that the adjustment of the charges, although they are only two in this case, resulted in different predictions that were slightly better than for the complex with the default charges. The above results were obtained by simply turning on/off the charges of the residues predicted to have unusual charge states in the complex. No extra protons were added or removed. Although this simplification has a negligible effect on the shape of the interface, it has a significant effect on the energy and thus on the quality of the predictions. The effect was found to be very pronounced for the homo-complex, while it was very moderate for the hereto-complex. The reason for this difference could be that more protonation states were adjusted for 1sji (three titratable groups), while only two ionizable groups were adjusted for 1gx2. In addition, the adjustments were made assuming that for both cases the physiological pH is about 7.0, and this may favor 1sji predictions over 1gx2.

Although the above example clearly illustrates the importance of correct assignment of protonation states for *ab-initio* docking, it does not offer a practical solution to real case scenarios where the structure of the corresponding complex is unknown. The other unknown is the relevant pH of the complex formation. As it was discussed in the previous section, some complexes are formed at low physiological pH, whereas others exist at high physiological pH.

[5] The charges were manipulated by modifying the charge column in the 'marked' file output by 'mark_sur' program within the ZDOCK distribution. The uncharged Asp, Glu, and Lys were modeled by turning off the partial charges of the side chain terminal atoms, while the ionized His+ was modeled by distributing a +1 charge over the ND1 and NE2 atoms.

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

The critical question with respect to docking is this: does the ligand-binding mode (the binding site and pose of the ligand) changes as ambient pH deviates from the corresponding physiological value? In the vast majority of cases, the answer is 'no': pH changes are shown to affect the affinity (binding free energy) but not the binding mode (see example cases listed in Table 1). However, opposite examples do exist as well (Sakurai *et al.* 2001). Thus, in cases where pH variations and, therefore, protonation states, do not affect the binding mode, the protonation states are not critical for predicting the binding mode. However, the correct assignment of protonation states will still be crucial for ranking and scoring of the candidate complexes. A wrong assignment of protonation states may cause the native ligand or the best candidate small molecule not to be ranked at the top of the list, and thus not to be predicted. Even more severe consequences of not treating the possibility of protonation state changes may be expected for calculations of the absolute binding energy. One way to deal with this problem is to allow all ionizable groups (or restricted number of groups located at the interface, assuming that the interfaces of both the receptor and the ligand have been identified first) to sample different protonation states during the docking, scoring, or energy calculations (Park *et al.* 2011). Of course, in this approach one should know the relevant physiological pH for the investigated complex(es). In cases for which the binding mode changes with pH, the problem is much more complicated. The receptor–ligand interface and the binding mode/pose is different at different pH, so one should define the relevant pH first, as mentioned above. It is quite unlikely that the predictions of alternative interfaces will be very successful, so one should let all ionizable (or at minimum all surface groups) to sample different protonation states during the docking, scoring, or energy calculations. This strategy may not be feasible at present, especially in virtual screening, where one has to screen millions of small molecule candidates. A simple consistency check – whether the assumed, e.g., standard, ionization state stays unchanged upon docking – may be less expensive, and still a better strategy than completely ignoring the possibility of ionization state change.

Individual case studies appear to support the importance of considering the possibility of binding-induced changes in the charge state of the receptor. It was shown both experimentally (Clackson & Wells, 1995; Lowman *et al.* 1991) and theoretically (Massova & Kollman, 1999) that altering the charge state of the binding interface via specific mutations can affect protein–ligand binding affinity, and can even be used to design complexes with higher affinity (Green & Tidor, 2005; Kangas & Tidor, 1999). Quantum-mechanical calculations have demonstrated that docking accuracy (Cho *et al.* 2005) and binding affinity prediction (Donnini *et al.* 2009) improve when the energy model accounts for the redistribution of ligand charges upon binding. In another recent study (Wittayanarakul *et al.* 2008), it was shown that accurate prediction of ionization states is a prerequisite for the accurate prediction of binding affinities between HIV protease and some inhibitors. At the same time, results of 'large-scale' docking studies are not uniform with respect to conclusions regarding the value of accounting for protonation state changes in binding, in part because of the use of different scoring and energy function. (Krieger *et al.* 2012; Labute, 2009; Li *et al.* 2007; Milletti *et al.* 2009; Milletti & Vulpetti, 2010; Rapp *et al.* 2009) On the one hand, it was shown that proper assignment of the protonation states results in better predictions, especially if the receptor is aspartyl protease (Rapp *et al.* 2009). This result is consistent with the above discussion, since aspartyl proteases typically function at lower pH and therefore some acidic groups are expected to have unusual protonation states. The effect of protonation states on the binding results was also studied on a set of three proteins complexed with ligands having different tautomeric forms (Todorov *et al.* 2006). It was demonstrated that the majority

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

of the protonation and tautomeric states of the individual groups within the three proteins can be recovered by the docking procedure. However, another recent study (Ten Brink & Exner, 2009) reported a similar analysis in which the protonation, tautomeric, and stereoisometric states of the ligands were varied; two popular docking programs were employed. It was found that the addition of different ligand protomers sometimes improved the docking results and sometimes did not. This type of improvement could also be receptor-dependent (Park *et al.* 2011). One possible explanation for the apparent case-dependence of the docking outcomes that consider protonation state changes is as follows. Perhaps the dependence has the same origin as outlined above, namely that different complexes are formed under different pH conditions. Thus, in some cases, standard protonation/tautomer states are already optimal for docking, whereas in other cases for which the binding occurs at lower or higher pH, adding alternative protonation states in the docking protocol should offer an advantage. This case dependence illustrates one intrinsic difficulty of the docking and virtual screening approaches – the fact that the true physiological pH is unknown for the candidate complexes. This is particularly important when inhibitors are designed to compete with natural substrates (Zapata-Torres *et al.* 2012). Recent work showed that the sensitivity of the inhibitor and substrate differs with respect to the protonation states (Zapata-Torres *et al.* 2012). Case-dependence of the docking outcomes may also stem from variations in the scoring functions used: highly detailed, physics-based scoring functions can be expected to be more sensitive to the proper protonation assignments than knowledge-based functions.

There is little doubt that a truly complete description of protein–ligand binding must consider the possibility of protonation state changes on the same footing with other hard-to-treat effects such conformational changes or water displacement upon binding. Various approaches have recently been developed to account for correct protonation states in *ab-initio* docking and virtual screening (Krieger *et al.* 2012; Labute, 2009; Li *et al.* 2007; Milletti *et al.* 2009; Milletti & Vulpetti, 2010; Rapp *et al.* 2009). An undeniable advantage of methods accounting for protonation changes upon binding is that they reveal the correct physics and provide the details of the binding reaction. Still, it remains to be seen to what extend today's fast, practical virtual screening methods will benefit from careful treatment of protonation state changes. First, there is the possibility that improving just one approximation out of many that these methods must make to achieve robustness and high computational performance may reduce their accuracy, e.g., by destroying some cancelation of errors that the less detailed descriptions may benefit from. Indeed, many practical scoring functions and docking algorithms do not consider protonation states at all, but perform reasonably well (Gohlke *et al.* 2000; Wang *et al.* 2003). Also, as we have seen, while protonation state changes often occur upon ligand binding, they by no means occur all of the time – neglecting this effect entirely may still be acceptable since the goal of virtual screening is often to find some good drug candidates, not all of them.

6. Outlook

We have presented a review of one aspect of protein–ligand binding process: changes in ionization (protonation, p*K*) that may occur in the protein when the complex forms. Just like many other binding-related effects, notably conformational changes, changes in the ionization state are complex, and are closely related to other phenomena, such as pH-dependence of binding affinity. Consistent account for these effects is important for both fundamental understanding of one of the most important processes in biology as well as for a very practical reason: early stages of

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

modern drug-discovery often involve computational screening of protein–ligand complexes. Considerable progress has been made, especially recently, in characterizing and understanding binding-induced protonation state changes, as well as incorporating them into computational protocols. At the same time, significant challenges remain.

(1) Experimental characterization of pH and/or protonation state changes upon complex formation remains challenging; to the best of our knowledge, no 'gold standard' set of well-characterized structures is available to serve as a reference and test set for computational studies. Such a gold standard set would have included a good number of very high-resolution complexes along with the corresponding unbound proteins and their ligands. All protonation state changes would have been assigned (experimentally) to specific groups, and binding affinities measured at least at some biologically relevant pH.

(2) Experimental difficulties are also likely to hinder verification of some of the intriguing computational predictions (and their implications), such as the existence of the 'long-range' effect of binding on p*K* of distant groups.

In the computational arena, the availability of fast and reasonably accurate methods for prediction of p*K* and protonation states in atomistic protein structures has facilitated dramatically the inclusion of p*K* and protonation state changes in studies of protein–ligand complexes and their diverse biological properties. Such computational studies are no longer a prerogative of a few expert groups, but are becoming accessible to a much broader class of researchers, especially through web-based tools that have recently come on-line. This, combined with ever growing computational power, facilitates integration of protonation changes into modeling protocols. Yet, many challenges remain as well. Among them, one important class is related to correct modeling of protonation states in studies of protein–ligand binding, which in turn is important to improving docking, scoring, screening, and energy calculations. We see the following challenges in this area:

(3) Accuracy of methods for energy calculations and scoring, as compared with the magnitude of the binding free energy and the binding free energy difference among competing ligands. Frequently, the binding free energies of different ligands (including small molecules) to the same receptor differ by less than 1 kcal/mol. In order to benefit from the inclusion of protonation states into docking, scoring, screening, or energy calculation protocols, the accuracy of the computation should be better than the free energy differences between alternative ligands.

(4) Accuracy of p*K* predictions. Success of docking, screening, and free energy calculations that take into account protonation states of the complex, receptor, and ligand, critically depends on the accuracy of the p*K* predictions (although in some limited cases the necessary experimental data may be available as well). There are many existing methods to compute p*K*, a large set of them was recently evaluated via pKa-cooperative, an initiative to bring together researchers working in the field of p*K* calculations (see reviews (Alexov *et al.* 2011; Nielsen *et al.* 2011)). The outcome of the last pKa-cooperative meetings indicates that significant progress still needs to be made to achieve better quality p*K* predictions, especially for reaching accuracy of 1 p*K* unit for the p*K* calculations involving buried groups.

(5) The nature of the scoring function in docking and virtual screening. To the best of our knowledge, most docking and virtual screening scoring functions are based on calculating or estimating the interaction energy between the receptor and the ligand in the complex. In this case, the protonation states of the unbound receptor and the ligand become irrelevant

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

for the scoring (although they may be physically very important). The only thing that matters are the correct protonation states in the complex. Docking methods that utilize some form of a free energy calculation, by taking the difference between the free energy of the complex and unbound receptor and ligand, are very few, although such an approach clearly represents the right physics. However, to the best of our knowledge, even these advanced approaches do not take into account the pH-dependent term. Typically they assume that the default pH is neutral pH, and assign the protonation states according to the pK predictions at such pH.

(6) The interplay between protonation states and the pH of complex formation. Even in case that the structures of both the bound and unbound receptor and ligand are known, and one can perform p*K* calculations, the relevant protonation states cannot be assigned if the pH of the corresponding biochemical reaction of the complex formation is unknown. Simply assuming that all receptor–ligand complexes are formed at neutral pH = 7.0 is a great oversimplification. Perhaps the time is ripe for the computational community to add these considerations about the relevant pH when preparing the set of complexes to be instigated, or in the computational protocols that model the docking or binding free energy.

(7) Treatment of non-integer protonation states. At a particular pH of binding, the given titratable group may be partially ionized, or the net proton uptake may be spread over several titratable group resulting in partial protonation changes (non-integer number of protons). As it was outlined many times in the literature, it is not correct to simply assign partial net charges to the appropriate ionizable groups; instead, the macroscopically calculated partial charge must be properly recovered within an ensemble (Alexov, 2004; Park *et al.* 2011).

(8) A related challenge is consistent treatment of binding-induced protonation state changes on an equal footing with conformational changes. Constant pH Molecular Dynamics is a natural, theoretically well-grounded tool for the job, that also automatically takes care of the partial protonation states issue. However, available implementations are still extremely demanding computationally.

7. Acknowledgements

A. V. O. and E. A. acknowledge support from the NIH grant numbers (R01 GM076121 and R01 GM093937, respectively). The authors thank G. Matthias Ullmann for useful comments, Holger Gohlke for reading the draft of the manuscript and providing helpful feedback, and Boris Aguilar for help with some of the figures.

8. References

Aguilar, B., Anandakrishnan, R., Ruscio, J. Z. & Onufriev, A. V. (2010). Statistics and physical origins of pK and ionization state changes upon protein–ligand binding. *Biophysics Journal* **98**, 872–880.

Alexov, E. (2004). Calculating proton uptake/release and binding free energy taking into account ionization and conformation changes induced by protein–inhibitor association: application to plasmepsin, cathepsin D and endothiapepsin–pepstatin complexes. *Proteins* **56**, 572–584.

Alexov, E., Mehler, E. L., Baker, N., Baptista, A. M., Huang, Y., Milletti, F., Nielsen, J. E., Farrell, D., Carstensen, T., Olsson, M. H., Shen, J. K., Warwicker, J., Williams, S. & Word, J. M. (2011). Progress in the prediction of pka values in proteins. *Proteins* **79**, 3260–75.

Anandakrishnan, R., Aguilar, B. & Onufriev, A. V. (2012). H++ 3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Research* **40** (Web Server issue), 537–541.

Antosiewicz, J., McCammon, J. A. & Gilson, M. K. (1994). Prediction of pH-dependent properties of proteins. *Journal of Molecnlor Biology* **238**, 415–436.

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

Baker, B. M. & Murphy, K. P. (1996). Evaluation of linked protonation effects in protein binding reactions using isothermal titration calorimetry. *Biophysical Journal* **71**, 2049–2055.

Bas, D. C., Rogers, D. M. & Jensen, J. H. (2008). Very fast prediction and rationalization of pKa values for protein–ligand complexes. *Proteins* **73**, 765–783.

Bashford, D. & Karplus, M. (1990). pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* **29**, 10219–10225.

Beroza, P., Fredkin, D. R., Okamura, M. Y. & Feher, G. (1991). Protonation of interacting residues in a protein by Monte Carlo method. *Proceeding of the National Academy of Sciences of the United States of America* **88**, 5804–5808.

Beroza, P., Fredkin, D. R., Okamura, M. Y. & Feher, G. (1991). Protonation of interacting residues in a protein by a Monte Carlo method: application to lysozyme and the photosynthetic reaction center of rhodobacter sphaeroides. *Proceeding of the National Academy of Sciences of the United States of America* **88**, 5804–5808.

Betts, M. J. & Sternberg, M. J. E. (1999). An analysis of conformational changes on protein–protein association: implications for predictive docking. *Protein Engineering* **12**, 271–283.

Bidwai, A. K., Ok, E. Y. & Erman, J. E. (2008). pH dependence of cyanide binding to the ferric heme domain of the direct oxygen sensor from *Escherichia coli* and the effect of alkaline denaturation. *Biochemistry* **47**, 10458–10470.

Bjarnadottir, U. & Nielsen, J. E. E. (2010). Calculating pKa values in the cAMP-dependent protein kinase: the effect of conformational change and ligand binding. *Protein Science: a Publication of the Protein Society* **19**, 2485–2497.

Blundell, C. D., Mahoney, D. J., Cordell, M. R., Almond, A., Kahmann, J. D., Perczel, A., Taylor, J. D., Campbell, I. D. & Day, A. J. (2007). Determining the molecular basis for the pH-dependent interaction between the link module of human TSG-6 and hyaluronan. *Journal of Biological Chemistry* **282**, 12976–12988.

Boehr, D. D. & Wright, P. E. (2008). Biochemistry: how do proteins interact? *Science* **320**, 1429–1430.

Bombarda, E. & Ullmann, G. M. (2010). pH-dependent pKa values in proteins: a theoretical analysis of protonation energies with practical consequences for enzymatic reactions. *The Journal of Physical Chemistry B* **114**, 1994–2003.

Bordner, A. J. & Abagyan, R. (2005). Statistical analysis and prediction of protein–protein interfaces. *Proteins* **60**, 353–366.

Brandsdal, B. O., Smalås, A. O. & Åqvist, J. (2006). Free energy calculations show that acidic p1 variants undergo large pKa shifts upon binding to trypsin. *Proteins* **64**, 740–748.

Bruylants, G., Wintjens, R., Looze, Y., Redfield, C. & Bartik, K. (2007). Protonation linked equilibria and apparent affinity constants: the thermodynamic profile of the alpha–chymotrypsin–proflavin interaction. *European Biophysics Journal* **37**, 11–18.

Buckle, A. M., Schreiber, G. & Fersht, A. R. (1994). Protein–protein recognition: crystal structural analysis of a barnase–barstar complex at 2.0-a resolution. *Biochemistry* **33**, 8878–89.

Buczek, O., Koscielska-Kasprzak, K., Krowarsch, D., Dadlez, M. & Otlewski, J. (2002). Analysis of serine proteinase–inhibitor interaction by alanine shaving. *Protein Sciences* **11**, 806–819.

Carter, P., Lesk, V., Islam, S. & Sternberg, M. (2005). Protein–protein docking using 3d-dock in rounds 3,4 of capri. *Proteins* **60**, 281–288.

Cassidy, C. S., Lin, J. & Frey, P. A. (1997). A new concept for the mechanism of action of chymotrypsin: the role of the low-barrier hydrogen bond. *Biochemistry* **36**, 4576–4584.

Chan, P., Lovric, J. & Warwicker, J. (2006). Subcellular ph and predicted pH-dependent features of proteins. *Proteomics,* **6**, 3494–3501.

Chang, A., Scheer, M., Grote, A., Schomburg, I. & Schomburg, D. (2009). Brenda, amenda and frenda the enzyme information system: new content and tools in 2009. *Nucleic Acids Research* **35**, D588–D592.

Cho, A. E., Guallar, V., Berne, B. J. & Friesner, R. (2005). Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach. *Journal of Computational Chemistry* **26**, 915–931.

Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone–receptor interface. *Science* **267**, 383–386.

Commet, A., Boswell, N., Yocum, C. F. & Popelka, H. (2012). pH optimum of the photosystem ii h(2)o oxidation reaction: effects of psbo, the manganese-stabilizing protein, cl(-) retention, and deprotonation of a component required for o(2) evolution activity. *Biochemistry* **51**, 3808–3818.

Czodrowski, P., Sotriffer, C. A. & Klebe, G. (2007). Protonation changes upon ligand binding to trypsin and thrombin: structural interpretation based on pKa calculations and ITC experiments. *Journal of Molecular Biology* **367**, 1347–1356.

Dan, A., Ofran, Y. & Kliger, Y. (2009). Large-scale analysis of secondary structure changes in proteins suggests a role for disorder-to-order transitions in nucleotide binding proteins. *Proteins* **78**, 236–248, 1 February 2010.

Davenport, H. (1996). *Physiology of the Digestive Tract.* Chicago: Year Book Medical Publishers, Inc.

DelBuono, G. S., Figueiridо, F. E. & Levy, R. M. (1994). Intrinsic pK_a s of ionizable residues in proteins:

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

an explicit solvent calculation for lysozyme. *Proteins* **20**, 85–97.

Demchuk, E. & Wade, R. C. (1996). Improving the continuum dielectric approach to calculating pK_as of ionizable groups in proteins. *The Journal of Physical Chemistry* **100**, 17373–17387.

Djurdjevic-Pahl, A., Hewage, C. & Malthouse, J. P. (2005). Ionisations within a subtilisinglyoxal inhibitor complex. *Biochimica et Biophysica Acta* (*BBA*) – *Proteins & Proteomics* **1749**, 33–41.

Dolinsky, T. J., Czodrowski, P., Li, H., Nielsen, J. E., Jensen, J. H., Klebe, G. & Baker, N. A. (2007). PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Research* **35** 2007 July; (Web Server issue), W522–W525.

Donnini, S., Villa, A., Groenhof, G., Mark, A. E., Wierenga, R. K. & Juffer, A. H. (2009). Inclusion of ionization states of ligands in affinity calculations. *Proteins* **76**, 138–150.

Elcock, A. H., Sept, D. & Mccammon, J. A. (2001). Computer simulation of protein–protein interactions. *Journal of Physical Chemistry B* **105**, 1504–1518.

Garcia-Moreno, B. (2009). Adaptations of proteins to cellular and subcellular pH. *Journal of Biology* **8**, 96.

Ge, X. & Roux, B. (2010). Absolute binding free energy calculations of sparsomycin analogs to the bacterial ribosome. *Journal of. Physical Chemistry B* **114**, 9525–9539.

Georgescu, R., Alexov, E. & Gunner, M. (2002). Combining conformational flexibility and continuum electrostatics for calculating pK_as in proteins. *Biophysical. Journal* **83**, 1731–1748.

Gerstein, M., Lesk, A. M. & Chothia, C. (1994). Structural mechanisms for domain movements in proteins. *Biochemistry* **33**, 6739–6749.

Gilson, M. K. & Zhou, H. X. (2007). Calculation of protein–ligand binding affinities. *Annual Review of Biophyises and Biomolecular Structure* **36**, 21–42.

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

Goh, C-S., Milburn, D. & Gerstein, M. (2004). Conformational changes associated with protein–protein interactions. *Current Opinion Structure Biology* **14**, 104–109.

Gohlke, H., Hendlich, M. & Klebe, G. (2000). Knowledge-based scoring function to predict protein–ligand interactions. *Journal of Molecular Biology* **295**, 337–356.

Gohlke, H., Kiel, C. & Case, D. A. (2003). Insights into protein–protein binding by binding free energy calculation and free energy decomposition for the ras–raf and ras–ralgds complexes. *Journal of Molecular Biology* **330**, 891–913.

Gohlke, H. & Klebe, G. (2002). Approaches to the description and prediction of the binding affinity of Small-molecule ligands to macromolecular receptors. *Angewandte Chemie International Edition* **41**, 2644–2676.

Gómez, J. & Freire, E. (1995). Thermodynamic mapping of the inhibitor site of the aspartic protease endothiapepsin. *Journal of Molecular Biology* **252**, 337–350.

Gordon, J. C., Myers, J. B., Folta, T., Shoja, V., Heath, L. S. & Onufriev, A. (2005). H^{++}: a server for estimating pK_as and adding missing hydrogens to macromolecules. *Nucleic Acids Research*, **33** (Web Server issue) 368–371.

Green, D. F. & Tidor, B. (2005). Design of improved protein inhibitors of HIV-1 cell entry: optimization of electrostatic interactions at the binding interface. *Proteins* **60**, 644–657.

Grimsley, G. R., Scholtz, J. M. & Pace, C. N. (2009). A summary of the measured pK values of the ionizable groups in folded proteins. *Protein Sciences* **18**, 247–251.

Guyton, A. (1976). *Textbook of Medical Physiology*. Philadelphia: W.B. Saunders Co.

Hansen, M. J., Olsen, J. G., Bernichtein, S., O'Shea, C., Sigurskjold, B. W., Goffin, V. & Kragelund, B. B. (2011). Development of prolactin receptor antagonists with reduced pH-dependence of receptor binding. *Journal of Molecular Recognition* **24**, 533–547.

Holmes, K C., Jon, I. A., Jahn, W. & Schroder, R. R. (2003). Electron cryo-microscopy shows how strong binding of myosin to actin releases nucleotide. *Nature* **425**, 423–427.

Horn, J. R., Ramaswamy, S. & Murphy, K. P. (2003). Structure and energetics of protein–protein interactions: the role of conformational heterogeneity in OMTKY3 binding to serine proteases. *Journal of Molecular Biology* **331**, 497–508.

Jensen, J. H. (2008). Calculating pH and salt dependence of protein-protein binding. *Current Pharmaceutical Biotechnology* **9**, 96–102.

Jones, S. & Thornton, J. M. (1997). Principles of protein–protein interactions. **93**, 13–20.

Jorgensen, W. L. (2004). The many roles of computation in drug discovery. *Science* **303**, 1813–1818.

Joshi, M. D., Sidhu, G., Nielsen, J. E., Brayer, G. D., Withers, S. G. & Mcintosh, L. P. (2001). Dissecting the electrostatic interactions and pH-dependent activity of a family 11 glycosidase. *Biochemistry* **40**, 10115–10139.

Kangas, E. & Tidor, B. (1999). Charge optimization leads to favorable electrostatic binding free energy. *Physical Review E* **59**, 5958–5961.

Khandogin, J. & Brooks, C. L. (2006). Toward the accurate first-principles prediction of ionization equilibria in proteins. *Biochemistry* **45**, 9363–9373.

Kieseritzky, G. & Knapp, E. W. (2008a). Improved pKa prediction: combining empirical and semimicroscopic methods. *Journal of Computational Chemistry* **29**, 2575–2581.

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

Kieseritzky, G. & Knapp, E-W. (2008b). Optimizing pKA computation in proteins with pH adapted conformations. *Proteins* **71**, 1335–1348.

Kozlov, A. G. & Lohman, T. M. (2000). Large contributions of coupled protonation equilibria to the observed enthalpy and heat capacity changes for ssDNA binding to escherichia coli SSB protein. *Proteins* **Suppl 4**, 8–22.

Kresheck, G. C., Vitello, L. B. & Erman, J. E. (1995). Calorimetric studies on the interaction of horse ferricytochrome c and yeast cytochrome c peroxidase. *Biochemistry* **34**, 8398–8405.

Krieger, E., Dunbrack, Jr., R. L., Hooft, R. W. & Kriege, B. (2012). Assignment of protonation states in proteins and ligands: combining pKa prediction with hydrogen bonding network optimization. *Methods in Molecular Biology* **819**, 405–421.

Krowarsch, D., Dadlez, M., Buczek, O., Krokoszynska, I., Smalas, A. O. & Otlewski, J. (1999). Interscaffolding additivity: binding of p1 variants of bovine pancreatic trypsin inhibitor to four serine proteases. *Journal of Molecular Biology* **289**, 175–186.

Kulichikhin, K. Y., Greenway, H., Byrne, L. & Colmer, T. D. (2009). Regulation of intracellular pH during anoxia in rice coleoptiles in acidic and near neutral conditions. *Journal of Experimental Botany* **60**, 2119–2128.

Kundrotas, P. J. & Alexov, E. (2006). Electrostatic properties of protein-protein complexes. *Biophysies Journal* **91**, 1724–1736.

Labute, P. (2009). Protonate3d: assignment of ionization states and hydrogen coordinates to macromolecular structures. *Proteins* **75**, 187–205.

Lee, M. S., Salsbury, F. R. & Brooks, C. L. (2004). Constant-pH molecular dynamics using continuous titration coordinates. *Proteins* **56**, 738–752.

Lee, M. S. & Olson, M. A. (2008). Calculation of absolute ligand binding free energy to a ribosome-targeting protein as a function of solvent model. *Journal of Physical Chemistry* **112**, 13411–13417.

Lensink, M. F., Mendez, R. & Wodak, S. J. (2007). Docking and scoring protein complexes: Capri 3rd edition. *Proteins* **74**, 484–495.

Li, X., Jacobson, M. P., Zhu, K., Zhao, S. & Friesner, R. A. (2007). Assignment of polar states for protein amino acid residues using an interaction cluster decomposition algorithm and its application to high resolution protein structure modeling. *Proteins* **66**, 824–837.

Li, X-Z., Walker, B. & Michaelides, A. (2011). Quantum nature of the hydrogen bond. *Proceedings of the National Academy of Sciences* **108**, 6369–6373.

Linderstrom-Lang, K. (1924). On the ionization of proteins. *Competes Rendus des Travaux du Laboratoire Carlsberg* **15** (7), 1–29

Lowman, H. B., Cunningham, B. C. & Wells, J. A. (1991). Mutational analysis and protein engineering of receptor-binding determinants in human placental lactogen. *Journal of Biological Chemistry* **266**, 10982–10988.

Maiti, A. & Drohat, A. C. (2011). Dependence of substrate binding and catalysis on pH, ionic strength, and temperature for thymine dna glycosylase: insights into recognition and processing of g.t mispairs. *DNA Repair* **10**, 545–53.

Maler, L., Blankenship, J., Rance, M. & Chazin, W. J. (2000). Site–site communication in the EF-hand Ca^{2+} – binding protein calbindin d9k. *Nature Structure Molecular Biology* **7**, 245–250.

Mason, A. C. & Jensen, J. H. (2008). Protein–protein binding is often associated with changes in protonation state. *Proteins* **71**, 81–91.

Massova, I. & Kollman, P. A. (1999). Computational alanine scanning to probe protein–protein interactions: a novel approach to evaluate binding free energies. *Journal of the American Chemical Society* **121**, 8133–8143.

McCammon, J. A. (1998). Theory of biomolecular recognition. *Current Opinion in Structural Biology* **8**, 245–249.

McCoy, A. J., Epa, V. C. & Colman, P. M. (1997). Electrostatic complementarity at protein/protein interfaces. *Journal of Molecular Biology* **268**, 570–584.

McDonald, S. M., Willson, R. C. & McCammon, J. A. (1995). Determination of the pKa values of titratable groups of an antigen–antibody complex, HyHEL-5-hen egg lysozyme. *Protein Engineering* **8**, 915–924.

McGuffee, S. R. & Elcock, A. H. (2006). Atomically detailed simulations of concentrated protein solutions: the effects of salt, pH, point mutations, and protein concentration in simulations of 1000-molecule systems. *Journal of the American Chemical Society* **128**, 12098–12110.

McPhalen, C. A., Svendsen, I., Jonassen, I. & James, M. N. (1985). Crystal and molecular structure of chymotrypsin inhibitor 2 from barley seeds in complex with subtilisin novo. **82**, 7242–7246.

Mendez, R., Leplae, R., Lensink, M. & Wodak, S. (2005). Assessment of capri predictions in rounds 3–5 shows progress in docking procedures. *Proteins* **60**, 150–169.

Milletti, F., Storchi, L. & Cruciani, G. (2009). Predicting protein pK(a) by environment similarity. *Proteins* **76**, 484–495.

Milletti, F. & Vulpetti, A. (2010). Tautomer preference in PDB complexes and its impact on structure-based drug discovery. *Journal of Chemical Information Modeling* **50**, 1062–1074.

Mishra, S. H., Spring, A. M. & Germann, M. W. (2009). Thermodynamic profiling of HIV RREIIB RNAzinc finger interactions. *Journal of Molecular Biology* **393**, 369–382.

Misra, V. K., Hecht, J. L., Yang, A-S. & Honig, B. (1998). Electrostatic contributions to the binding free energy of the Â»cI repressor to DNA. *Biophysical Journal* **75**, 2262–2273.

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

Mitra, R. C., Zhang, Z. & Alexov, E. (2011). *In silico* modeling of pH-optimum of protein–protein binding. *Proteins* **79**, 925–936.

Mobley, D. L. & Dill, K. A. (2009). Binding of small-molecule ligands to proteins: what you see is not always what you get. *Structure* **17**, 489–498.

Mongan, J., Case, D. A. & McCammon, J. A. (2004). Constant pH molecular dynamics in generalized born implicit solvent. *Journal of. Computational Chemistry* **25**, 2038–2048.

Murphy, K. P., Xie, D., Garcia, K. C., Amzel, L. M. & Freire, E. (1993). Structural energetics of peptide recognition: angiotensin II/antibody binding. *Proteins* **15**, 113–120.

Myers, J., Grothaus, G., Narayanan, S. & Onufriev, A. (2006). A simple clustering algorithm can be accurate enough for use in calculations of pKs in macromolecules. *Proteins* **63**, 928–938.

Nielsen, J. E., Gunner, M. R. & Garcia-Moreno, B. E. (2011). The pKa cooperative: a collaborative effort to advance structure-based calculations of pKa values and electrostatic effects in proteins. *Proteins* **79**, 3249–3259.

Nielsen, J. E. & McCammon, J. A. (2003). On the evaluation and optimization of protein X-ray structures for pKa calculations. *Protein Science: a Publication of the Protein Society* **12**, 313–326.

Nielson, J. E. & Vriend, G. (2001). Optimizing the hydrogen-bond network in Poisson–Boltzmann equation-based pK_a calculations. *Proteins* **43**, 403–412.

Ofran, Y. & Rost, B. (2003). Predicted protein–protein interaction sites from local sequence information. *FEBS Letters* **544**, 236–239.

Onufriev, A., Case, D. A. & Ullmann, G. M. (2001). A novel view of pH titration in biomolecules. *Biochemistry* **40**, 3413–3419.

Onufriev, A., Smondyrev, A. & Bashford. D. (2003). Proton affinity changes during unidirectional proton transport in the bacteriorhodopsin photocycle. *Journal of Molecular Biology* **332**, 1183–1193.

Onufriev, A. & Ullmann, G. M. (2004). Decomposing complex cooperative ligand binding into simple components: connections between microscopic and macroscopic models. *Journal of Physical Chemistry B* **108**, 11157–11169.

Palpant, N. J., Houang, E. M., Delport, W., Hastings, K. E., Onufriev, A. V., Sham, Y. Y. & Metzger, J. M. (2010). Pathogenic peptide deviations support a model of adaptive evolution of chordate cardiac performance by troponin mutations. *Physiological Genomics* **42**, 287–299.

Park, M-S., Gao, C. & Stern, H. A. (2011). Estimating binding affinities by docking/scoring methods using variable protonation states. *Proteins* **79**, 304–314.

Perozzo, R., Folkers, G. & Scapozza, L. (2004). Thermodynamics of protein–ligand interactions: history, presence, and future aspects. *Journal of Receptor and Signal Transduction Research* **24**, 1–52.

Pierce, B. G., Hourai, Y. & Weng, Z. (2007). Accelerating protein docking in zdock using an advanced 3d convolution library. *PLoS One* **6**, e24657.

Qasim, M. A., Ranjbar, M. R., Wynn, R., Anderson, S. & Laskowski, M. (1995). Ionizable p1 residues in serine proteinase inhibitors undergo large pK shifts on complex formation. *Journal of Biological Chemistry* **270**, 27419–27422.

Rapp, C. S., Schonbrun, C., Jacobson, M. P., Kalyanaraman, C. & Huang, N. (2009). Automated site preparation in physics-based rescoring of receptor ligand complexes. *Proteins* **77**, 52–61.

Reeves Gibbs, M. D. R. A., Choudhary, P. R. & Bergquist, P. L. (2010). Alteration of the pH optimum of a family 11 xylanase, xynb6 of dictyoglomus thermophilum. *New Biotechnology* **27**, 803–9.

Rodinger, T., Howell, P. L. & Pomes, R. (2008). Calculation of absolute protein–ligand binding free energy using distributed replica sampling. *Journal of Chemical Physics* **129**, 155102.

Sackett, K., TerBush, A. & Weliky, D. P. (2011). Hiv gp41 six-helix bundle constructs induce rapid vesicle fusion at Ph 3.5 and little fusion at pH 7.0: understanding pH dependence of protein aggregation, membrane binding, and electrostatics, and implications for hiv-host cell fusion. *European Biophysics Journal* **40**, 489–502.

Sakurai, K., Oobatake, M. & Goto, Y. (2001). Salt-dependent monomerdimer equilibrium of bovine b-lactoglobulin at pH 3. *Protein Sciences* **10**, 2325–2335.

Schneidman-Duhovny, D., Inbar, Y., Nussinov, R. & Wolfson, R. (2005). Geometry-based flexible and symmetric protein docking. *Proteins* **60**, 224–231.

Schomburg, I., Chang, A., Ebeling, C., Gremse, M., Heldt, C., Huhn, G., & Schomburg, D. (2004). Brenda, the enzyme database: updates and major new developments. *Nucleic Acids Research* **32**, D431–D43.

Schreiber, G. & Fersht, A. R. (1993). Interaction of barnase with its popypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **32**, 5145–5150.

Schreiber, G. & Fersht, A. R. (1995). Energetics of protein–protein interactions: analysis of the barnase–barstar interface by single mutations and double mutant cycles. *Journal of Molecular Biology* **248**, 478–486.

Schueler-Furman, O., Wang, C. & Bake, D. (2005). Progress in protein–protein docking: atomic resolution in the predictions in the Capri experiment using rosattadock with an improved treatment of the side-chain flexibility. *Proteins* **60**, 187–194.

Sham, Y. Y., Chu, Z. T. & Warshel, A. (1997). Consistent calculations of pKa pK_a's of ionizable residues in proteins: semi-microscopic and microscopic approaches. *Journal of Physical Chemistry* **101**, 4458–4472.

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

Sheinerman, F. B. & Honig, B. (2002). On the role of electrostatic interactions in the design of protein–protein interfaces. *Journal of Molecular Biology* **318**, 161–177.

Shin, C. J., Wong, S., Davis, M. J. & Ragan, M. A. (2009). Protein–protein interaction as a predictor of subcellular location. *BMC Systems Biology* **3**, 28.

Simonson, T., Carlsson, J. & Case, D. A. (2004). Proton binding to proteins: pKa calculations with explicit and implicit solvent models. *Jorunal of American Chemistry Society* **126**, 4167–4180.

Sims, P. A., Wong, C. F., Vuga, D., Mccammon, A. J. & Sefton, B. M. (2005). Relative contributions of desolvation, inter- and intramolecular interactions to binding affinity in protein kinase systems. *Journal of Computational Chemistry* **26**, 668–681.

Singh, N. & Warshel, A. (2010). Absolute binding free energy calculations: on the accuracy of computational scoring of protein–ligand interactions. *Proteins* **78**, 1705–1723.

Smith, R., Brereton, I. M., Chai, R. Y. & Kent, S. B. (1996). Ionization states of the catalytic residues in HIV-1 protease. *Nature Structural and Molecular Biology* **3**, 946–950.

Song, Y., Mao, J. & Gunner, M. R. (2009). MCCE2: Improving protein pKa calculations with extensive side chain rotamer sampling. *Journal of Computational Chemistry* **30**, 2231–2247.

Spassov, V. & Bashford, D. (1998). Electrostatic coupling to pH-titrating sites as a source of cooperativity in protein–ligand binding. *Protein Science : a Publication of the Protein Society* **7**, 2012–2025.

Spassov, V. Z. & Yan, L. (2008). A fast and accurate computational approach to protein ionization. *Protein. Sciences* **17**, 1955–1970.

Spitzer, R. & Jain, A. N. (2012). Surflex-dock: docking benchmarks and real-world application. *Journal of Computer Aided Molcenlar Design* **26** (**6**), 687–699.

Steuber, H., Czodrowski, P., Sotriffer, C. A. & Klebe, G. (2007). Tracing changes in protonation: a prerequisite to factorize thermodynamic data of inhibitor binding to aldose reductase. *Journal of Molecular Biology* **373**, 1305–1320.

Stranzl, G. R. Gruber, K., Steinkellner, G., Zangger, K., Schwab, H. & Kratky, C. (2004). Observation of a short, strong hydrogen bond in the active site of hydroxynitrile lyase from hevea brasiliensis explains a large pK_a shift of the catalytic base induced by the reaction intermediate. *Journal of Biological Chemistry* **279**, 3699–3707.

Swails, J. M., Meng, Y., Walker, A. F., Marti, M. A., Estrin, D. A. & Roitberg, A. E. (2009). pH-dependent mechanism of nitric oxide release in nitrophorins 2 and 4. *Journal of Physical Chemistry B* **113**, 1192–1201.

Takahashi, T., Nakamura, H. & Walda, A. (1992). Electrostatic forces in two lysozymes: calculations and measurements of histidine pK_a values. *Biopolymers* **32**, 897–909.

Talley, K. & Alexov, E. (2010). On the pH-optimum of activity and stability of proteins. *Proteins* **78**, 2699–2706.

Tanford, C. (1970). Protein denaturation. c. theoretical models for the mechanism of denaturation. *Advances in Protein Chemistry* **24**, 1–95.

Tanford, C. & Kirkwood, J. (1957). Theory of protein titration curves. *Journal of the American Chemical Society* **79**, 5333–5339.

Tanford, C. & Roxby, R. (1972). Interpretation of protein titration curves. *Biochemistry* **11**, 2192–2198.

Taylor, R. D., Jewsbury, P. J., & Essex, J. W. (2002). A review of protein–small molecule docking methods. *Journal of Computer-Aided Molecular Design* **16**, 151–166.

Ten Brink, T. & Exner, T. E. (2009). Influence of protonation, tautomeric, and stereoisomeric states on protein–ligand docking results. *Journal of Chemical Information and Modelling* **49**, 1535–1546.

Todorov, N. P., Monthoux, P. H. & Alberts, I. L. (2006). The influence of variations of ligand protonation and tautomerism on protein–ligand recognition and binding energy landscape. *Journal of Chemical Information and Modelling* **46**, 1134–1142.

Trott, O. & Olson, A. J. (2010). Autodock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**, 455–61.

Trylska, J., Antosiewicz, J., Geller, M., Hodge, C. N., Klabe, R. M., Head, M. S., & Gilson, M. K. (1999). Thermodynamic linkage between the binding of protons and inhibitors to HIV-1 protease. *Protein Science* **8**, 180–195.

Tsai, C., Lin, S., Wolfson, H. & Nussinov, R. (1997). Studies of protein–protein interfaces: a statistical analysis of the hydrophobic effect. *Protein Science* **6**, 53–64.

Ullmann, G. M. & Knapp, E-W. (1999). Electrostatic models for computing protonation and redox equilibria in proteins. *European Biophysical Journal* **28**, 533–551.

Ullmann, R. T. & Ullmann, G. M. (2012). GMCT: a Monte Carlo simulation package for macromolecular receptors. *Journal of Computational Chemistry,* **33**, 887–900.

Voet, D. & Voet, J. G. (1995). *Biochemistry*. John Wiley & Sons, 2 ed. New York.

Wang, J., Deng, Y. & Roux, B. (2006). Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophysical Journal* **91**, 2798–2814.

Wang, L., Witham, S., Zhang, Z., Li, L., Hodsdon, M. & Alexov, E. (2013). In silico investigation of pH-dependence of prolactin and human growth hormone binding to human prolactin receptor. *Communications in Computation and Physics* **13**, 207–222.

Wang, R., Lu, Y. & Wang, S. (2003). Comparative evaluation of 11 scoring functions for molecular docking. *Journal of Medicinal Chemistry* **46**, 2287–2303.

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press

Wang, Y. X., Freedberg, D. I., Yamazaki, T., Wingfield, P. T., Stahl, S. J., Kaufman, J. D., Kiso, Y. & Torchia, D. A. (1996). Solution NMR evidence that the HIV-1 protease catalytic aspartyl groups have different ionization states in the complex formed with the asymmetric drug KNI-272. *Biochemistry* **35**, 9945–9950.

Warren, G. L., Andrews, C. W., Capelli, A.-M., Clarke, B., LaLonde, J., Lambert, M. H., Lindvall, M., Nevins, N., Semus, S. F., Senger, S., Tedesco, G., Wall, I. D., Woolven, J. M., Peishoff, C. E., & Head, M. S. (2005). A critical assessment of docking programs and scoring functions. *Journal of Medicinal Chemistry* **49**, 5912–5931.

Wittayanarakul, K., Hannongbua, S. & Feig, M. (2008). Accurate prediction of protonation state as a prerequisite for reliable MM-PB(GB)SA binding free energy calculations of HIV-1 protease inhibitors. *Journal of Computational Chemistry* **29**, 673–685.

Woo, H. J. (2008). Calculation of absolute protein–ligand binding constants with the molecular dynamics free energy perturbation method. *Methods in Molecular Biology* **443**, 109–120.

Woo, H. J. & Roux, B. (2005). Calculation of absolute protein–ligand binding free energy from computer simulations. *Proc Natl Acad Sci USA* **102**, 6825–6830.

Xie, D., Gulnik, S., Collins, L., Gustchina, E., Suvorov, L., & Erickson, J. W. (1997). Dissection of the pH dependence of inhibitor binding energetics for an aspartic protease: direct measurement of the protonation states of the catalytic aspartic acid residues. *Biochemistry,* **36**, 16166–16172.

Yamazaki, T., Nicholson, L. K., Wingfield, P., Stahl, S. J., Kaufman, J. D., Eyermann, C. J., Hodge, N. C., Lam, P. Y. & Torchia, D. A. (1994). NMR and X-ray evidence that the HIV protease catalytic aspartyl groups are protonated in the complex formed by the protease and a non-peptide cyclic urea-based inhibitor. *Journal of the American Chemical Society* **116**, 10791–10792.

Yan, S. & Wu, G. (2011). Searching of predictors to predict pH optimum of cellulases. *Applied Biochemistry and Biotechnology* **165**, 856–69.

Yang, A. S. & Honig, B. (1993). On the pH dependence of protein stability. *Journal of Molecular Biology* **231**, 459–474.

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

Yang, A. S. & Honig, B. (1992). Electrostatic effects on protein stability. *Current Opinion in Structural Biology* **2**, 40–45.

Yang, J. H., Park, J. Y., Kim, S. H. & Yoo, Y. J. (2008). Shifting pH optimum of bacillus circulans xylanase based on molecular modeling. *Journal of Biotechnology* **133**, 294–300.

Zapata-Torres, G., Fierro, A., Miranda-Rojas, S., Guajardo, C., Salgado, J. C., Saez-Briones, P. & Celis-Barros, C. (2012). Influence of protonation on substrate and inhibitor interactions at the active site of human monoamine oxidase-a. *Journal of Chemical Information and Modelling* **52**, 1213–1221.

Zhou, H. X. & Gilson, M. K. (2009). Theory of free energy and entropy in noncovalent binding. *Chemical Reviews,* **109**, 4092–107.

https://doi.org/10.1017/S0033583513000024 Published online by Cambridge University Press