

Some Critical Aspects of Molecular Interactions Between Drugs and Receptors

María J. R. Yunta*

Departamento de Química Orgánica I, Facultad de Química, Universidad Complutense, Madrid, Spain

*Corresponding author: mjryun@ucm.es

Received September 26, 2014; Revised October 15, 2014; Accepted October 19, 2014

Abstract Protonation states are sometimes crucial for free energy calculations to predict binding ability of molecules to receptors, a main subject for drug design. This mini-review seeks to identify the importance of knowing the influence of pH in such studies to better achieve correct predictions for drug candidates. Protonation states, for both proteins and drugs, need to be considered in docking studies although they have been usually neglected.

Keywords: *molecular modeling, protonation state, computer aided drug design*

Cite This Article: María J. R. Yunta, "Some Critical Aspects of Molecular Interactions Between Drugs and Receptors." *American Journal of Modeling and Optimization*, vol. 2, no. 4 (2014): 84-102. doi: 10.12691/ajmo-2-4-1.

1. Introduction

The prediction of the reactivity and the selectivity of a chemical process is a critical question. Thus, different theories have been proposed to explain chemical reactivity and/or selectivity. Each theory has proposed a set of global and/or local indexes aiming to describe the chemical processes, and global indexes such as electronegativity [1] and hardness [2], for example, have been introduced to characterize the reactivity of a compound, and local indexes like Fukui's frontier molecular orbital [3] or local softness [4] have been introduced to explain the regio and/or stereo selectivity of a reaction.

A special kind of chemical process is molecular recognition, a problem of fundamental importance in biology. Understanding its principles would lead to a more efficient application of medicinal chemistry. In particular, it is of great important for drug discovery to capture the physical principles responsible for the recognition of a drug by its protein target. The ever growing amount of structural information makes computer-aided structure-based ligand design methods or, generally speaking, molecular modeling, a useful alternative strategy to experimental high-throughput screening to find novel leads in a drug development program.

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule relative to a second one when bound together to form a stable complex [5], and, in turn, the knowledge about the preferred orientation may be used to predict the strength of association or binding affinity between two molecules using one of the several developed scoring functions. Typically, the goals of molecular docking are the identification of a ligand that binds to a specific receptor binding site and the identification of its preferred,

energetically most favorable, binding pose, where the term "binding pose" considers the orientation of a ligand relative to its receptor as well as the ligand's conformation.

In order to accomplish this task, molecular docking tools will generate a set of different ligand binding poses and use a scoring function to estimate binding affinities for the generated ligand poses in order to determine the best binding mode. This method is frequently used to predict the binding orientation of small molecule drug candidates in relation to their protein targets in the aim of being able to predict the affinity and activity of the small molecules, and hence it plays an important role in the rational design of drugs [6].

2. Interactions of Molecules with Proteins

To effectively predict free energy of binding for protein-ligand associations, given the biological and pharmaceutical significance of molecular docking, considerable efforts have been directed towards improving the methods used to predict docking, being a critical goal of computational chemistry/biology. In fact, this goal was described as the "Holy Grail" by Pearlman and Kollman a number of years ago [7], or by Gohlke and Klebe [8],

Computer-aided structure-based methods aimed at predicting the binding mode of a ligand in the binding site of a protein or any molecular target and at obtaining an estimate of the binding affinity involve two computational steps: docking and scoring. In the docking step, multiple protein-ligand configurations, called *poses*, are generated. This is typically carried out by performing a number of trials and keeping those poses that are energetically best. It involves finding the correct orientation and, as most ligand molecules are flexible, the correct conformation of the docked molecule. This implies that the degrees of

freedom to be searched include translational and rotational degrees of freedom of the ligand as a whole, as well as its internal degrees of freedom, i.e., predominantly the rotatable bonds. The search stops once a certain number of trials have been carried out and/or a sufficient number of poses have been found for a molecule. The decision to keep a trial pose is based on the computed ligand–receptor interaction energy (score) of that pose.

Several current docking programs have the ability to generate poses close to the native structure (usually an rmsd of ≤ 2 Å is accepted as close) in many cases [9-14]. Then, a scoring function is used to calculate the affinity between the receptor and the ligand for each pose. In addition, if multiple ligands are docked, their binding free energies need to be ranked accurately.

Scoring functions can be grouped into three classes: force-field-based, knowledge-based, and empirical scoring functions (see refs 8 and 15 for a review). For example, ab initio energy calculation methods have been used to improve threading, especially for the weakly/distant homologous templates of proteins [16], or in calculations of the potential grid used for docking [17], while semiempirical methods have been used in the MOZYME methodology for estimation of the target-ligand binding energies and its comparison with available experimental data [18]. Force field-based scoring functions apply classical molecular mechanics energy functions. They approximate the binding free energy of protein-ligand complexes by a sum of van der Waals and electrostatic interactions.

Protein-ligand complexation involves the formation of inter-molecular interactions such as hydrogen-bonds, aromatic stacking, salt bridges and hydrophobic complementarity. A large variety of different computational approaches exist for modelling these processes, ranging from ab initio, semi-empirical, classical mechanics to empirical methods. These methods are capable of computing individual aspects of the binding process; e.g., calculating the free energy of binding, simulating the reaction path for a substrate, examining the conformational flexibility, investigating the role of water, to mention only a few. One effect that these methods usually neglect concerns the protonation change upon ligand binding. Possible explanations for this deficiency are the high computational burden and the lack of experimental data.

The calculation of hydrogen positions is a common preprocessing step when working with crystal structures of protein-ligand complexes. An explicit description of hydrogen atoms is generally needed in order to analyze the binding mode of particular ligands or to calculate the associated binding energies. Due to the large number of degrees of freedom resulting from different chemical moieties and the high degree of mutual dependence this problem is anything but trivial. In addition to an efficient algorithm to take care of the complexity resulting from complicated hydrogen bonding networks, a robust chemical model is needed to describe effects such as tautomerism and ionization consistently. Bietz et al. [19] presented a method for the placement of hydrogen coordinates in protein-ligand complexes taking tautomers and protonation states of both protein and ligand into account. This method generates the most probable hydrogen positions on the basis of an optimal hydrogen

bonding network using an empirical scoring function. The high quality of its results could be verified by comparison to the manually adjusted Astex diverse set and a remarkably low rate of undesirable hydrogen contacts compared to other tools [19].

It is important to point out that electrostatic interactions are among the most important factors to be considered when analyzing the function of biological molecules. Since the pioneering work of Linderstrøm-Lang [20], significant progress has been made in the qualitative understanding and quantitative evaluation of electrostatic interactions. It is now generally recognized that one must analyze the electrostatic forces in a biomolecule to properly understand its function. Structure-based calculations of the electrostatic potential in and around a biomolecule, therefore, play a large role in virtually any protein-related research effort, as witnessed by the omnipresence of graphical representations of electrostatic potential surface maps in publications and oral presentations.

The majority of chemical reactions occur in acid medium. Hence, it is expected that the substrates in acid medium will attach with hydrogen ion, the proton, through the process of chemical and physical interactions as well. The physico-chemical processes of protonation are ubiquitous in almost all the branches of chemistry and biochemistry [21-25]. The resulting protonated molecule is frequently a pivotal intermediate that guides the succeeding steps of the entire chemico-physical process. The chemical process of protonation is also a fundamental step involving many chemical rearrangements and enzymatic reactions [24]. The knowledge of the intrinsic basicity and the site of protonation of a molecule is the key information for the understanding of the mechanism of chemical reactions and designing synthesis. The proton affinity is defined as the negative of the enthalpy change of a protonation reaction at the standard conditions. The gas-phase proton affinities are a quantitative measure of the intrinsic basicity of molecules [26]. The study of thermochemistry of the proton transfer reaction in the gas phase is well-known experiment of acid-base reactions [27]. The dynamics of proton transfer is also important for ionization processes in mass spectroscopy [28].

When a proton dynamically approaches a molecule in the physico-chemical process of protonation from a long distance, it is equally attracted by the electron cloud of the molecule. Thus a proton, acting as an electrophile, starts soaking the electron density from the entire skeleton of the nucleophile [29] and as a result, the electron density of the molecule is reorganized and redistributed. Ultimately the proton fixes at a site of lone pair – the site of protonation-of the molecule. However, if there is no lone pair in the structure of the molecule, the proton remains electrostatically attached to the sphere of the charge cloud of the molecule. During the process of protonation, the proton induces concomitant with the charge density reorganization, a physical process of structural and energetic changes in the molecule. The structural and energetic changes induced by the polarizing power of the proton are expected to be at its maximum at the gas phase of the molecule.

Understanding protein–ligand interactions is a major focus for modern molecular biophysics and structural bioinformatics research. On the practical side, application

of drug design techniques requires the availability of fast and reliable docking methods that can account for all major aspects of molecule interaction physics. Despite the progress in prediction via *in silico* methods, intricacies in protein–ligand interactions are still beyond our reach [30]–[32].

Prediction of protein–protein and protein–ligand interactions via docking methods is at the focus of intense research [33–43]. An essential step of any docking workflow is to find a list of ranked mutual orientations based on a scoring measure for shape complementarity and long-range interactions (electrostatics). The methods implementing rigid-body dock borrow ideas from protein–protein docking approaches such as the popular ZDOCK [33], Hex [34], PIPER [35] and GRAMM-X [36]. The first rigid body docking program based on fast Fourier transformation is the pioneering DOT application [37].

A subsequent step is aimed at refinement of rigid docking results by taking into account short-range interactions. A precise treatment requires account for backbone and side chain flexibility [38] e.g. Rosetta Dock [39] and HadDock [40]. Specific popular applications for protein–ligand docking that dominate the field are AutoDock [13] and SwissDock [34]. An alternative idea for docking is the search for analogy in known protein–ligand interfaces reminiscent of the protein–protein docking as implemented in PRISM [43]. However, all these methods do not face two issues quantum effects and the self-consistency of electrostatic interactions (including the mutual influence of docking partners on their protonation states through interdependent perturbation of pKa values).

A variety of computational methods and algorithms for studying proteins and their interactions with other molecules has been constructed. Methods such as molecular dynamics (MD) simulations, Brownian dynamics simulations, protein pKa calculations, protein–design algorithms and protein–drug and protein–protein docking algorithms are widely used in modern biological research. All these algorithms use 3D structures of protein molecules to predict and analyze protein characteristics, such as catalytic activity, folding pathway, stability, solubility and ligand/drug binding specificity.

3. Electrostatics of the Interactions

Electrostatic energies play a large role in determining all of the aforementioned characteristics, and the full potential of methods for predicting physical/biochemical characteristics will, therefore, only be realized once we are able to accurately calculate electrostatic energies and forces in and around proteins. The calculation of electrostatic interactions in biomolecules arguably presents one of the largest obstacles in improving the accuracy and usefulness of structure-based energy-calculation algorithms.

Protonation changes are omnipresent in the field of enzyme catalysis, because many catalyzed processes involve an acid-base reaction. Hence, a thorough understanding of such processes is of utmost importance being confronted with aspects of computational drug design such as ligand optimization, docking or molecular dynamics simulations. Protein X-ray crystallography

represents an experimental method heavily applied in all stages of drug discovery, and we will use many such structures as reference points for our pKa calculations, but most often their resolution is not sufficient to draw conclusions about protonation states [44].

Knowledge-based scoring functions represent the binding affinity as a sum of protein–ligand atom pair interactions. These potentials are derived from the protein–ligand complexes with known structures, where probability distributions of interatomic distances are converted into distance-dependent interaction free energies of protein–ligand atom pairs using the “inverse” Boltzmann law [45]. However, the structures deposited in the Protein Data Bank (PDB) do not provide a thermodynamic ensemble at equilibrium, and therefore, a knowledge-based potential should be considered as a statistical preference rather than a potential of mean force. A key ingredient of a knowledge-based potential is the reference state, which determines the weights between the various probability distributions. Recently, several approaches to derive these potentials have been proposed [46,47,48,49]. They differ in their definition of the reference state, the protein and ligand atom types, and the list of protein–ligand complexes from which they were extracted.

The experimental conditions, such as pH or salt concentration under which crystallization and/or the binding assay are performed, and the packing in the crystal form, can have a profound impact on the binding mode and the affinity of the ligand. This has been recently revealed in a study of trypsin crystals [50], which shows the occurrence of protonation and crystal form dependent binding modes. For a series of aliphatic cyclic ureas bound to HIV-1 protease, the binding energies computed by the Poisson equation significantly depend on the protonation state of the two active site aspartic acids [51]. Therefore, it may be anticipated that it is important to correctly model the protonation state of the ligand and the receptor, at least for the scoring functions that make use of partial charges. It is worth noting that the pH of crystallization of the complexes deposited in the LPDB is as low as 3.0 and as high as 8.5.

Unfortunately, electrostatic phenomena in biomolecular systems are very complex and much research is needed in this field to understand the factors that are important for determining their properties. The comparatively long-range nature of electrostatic forces presents a significant obstacle to progress, since one must consider interactions between a large number of solvent and solute atoms in order to solve the problem correctly. The many charged groups in proteins and the protein-induced variation in amino acid pKa values further complicate the calculation, since the protein charge distribution, which itself is dependent on the electrostatics in the protein, has to be calculated. The heterogeneous dielectric properties of the protein interior complicates the evaluation of electrostatic interactions even further by requiring energy-calculation methods to model protein and solvent reorganization in response to the variations in electrostatic field.

The importance of predicting protein characteristics, such as catalytic activity, stability and ligand/drug binding affinities, from protein 3D structures has led to the development of a wide range of electrostatic models in the last few decades. These models have been divided by

Warshel and co-workers into three groups: microscopic all-atom models, simplified microscopic models and continuum (macroscopic) models [52].

The binding affinity of a ligand to a protein controls the ligand concentration needed to achieve a high saturation of the protein. In the case of drug design, it is essential to achieve a tight association (high association constant K_A) in order to minimize the drug concentration needed during treatment. A lower drug concentration during treatment lowers the risk of side effects caused by nonspecific binding or binding to no target proteins, and is generally beneficial for issues related to administration, toxicology, metabolism and cost, for example. The prediction of the binding affinity of a drug typically relies on structure-based energy calculations performed according to the following thermodynamic cycle:

$$\Delta G_{bind} = \Delta G_{PL} - \Delta G_P - \Delta G_L = RT \ln K_A \quad (1)$$

where K_A represents the association constant, R is the gas constant and T refers to the absolute temperature. Equation 1 defines the free energy of binding ΔG_{bind} as the difference between the free energy of the solvated protein–ligand complex ΔG_{PL} and the free energy of the solvated protein ΔG_P and solvated ligand ΔG_L individually. Evaluating the ΔG_{bind} value with acceptable reliability and accuracy is crucial for ranking candidate ligands in Computer Aided Drug Design. Unfortunately, it is very difficult to predict ΔG_{bind} accurately using current structure-based energy calculation methods, since issues such as conformational changes, ligand and water entropy and protein electrostatics are poorly understood and only modeled approximately using current methods.

The electrostatic free energy ΔG_{el} of binding represents an integral part of ΔG_{bind} , and the accurate evaluation of this term is of utmost importance for drug discovery. Computational models in ligand-binding applications usually decompose ΔG_{el} into a favorable charge–charge interaction energy (often called the Coulombic energy) and an unfavorable desolvation penalty:

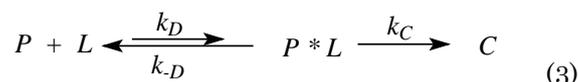
$$\Delta G_{el} = \Delta G_{coul} + \Delta G_{desolv} \quad (2)$$

where ΔG_{coul} refers to free energy changes in charge–charge interactions upon binding, and ΔG_{desolv} represents the energy penalty associated with moving the ligand from bulk water into the binding site of the protein.

The prediction of ΔG_{el} is very sensitive to the particular computational model used and evaluation of ΔG_{el} has been the subject of many recent computational studies [53,54,55,56]. The computational efficiency represents a bottleneck of microscopic models in screening the potential drug candidates (i.e., these models are much too slow), whereas the problematic concept of the protein dielectric constant defines the major problem for the accuracy of macroscopic models (see previous discussion).

Apart from the role of electrostatics in determining the binding specificity of a biological macromolecule, long-range electrostatic interactions also play a prominent role in steering charged ligands into the protein binding site.

The electrostatic enhancement of binding association rate has been studied extensively in recent years by Zhou, Schreiber and co-workers [54,57,58,59,60]. In order to better understand the role of long-range electrostatics in protein–ligand association kinetics, they used the hypothetical two-step kinetic scheme:



where k_D and k_{-D} represent the diffusion-controlled association and disassociation rate constants and k_C is the reaction-controlled rate constant. The loose complex ($P * L$) is referred to as an ‘encounter complex’ and is formed before the ligand is docked into its final position in the binding pocket. In comparison with the final complex C , which is stabilized by favorable interactions of many forces (hydrophobic interactions, hydrogen bonds, aromatic stacking, electrostatic and Van der Waals interactions), the encounter complex $P * L$ is predominantly stabilized by long-range electrostatic interactions between the protein and the ligand. Favorable electrostatic interactions are produced by complementary charge distribution between binding partners, and the steering force enhances the overall association rate according to the equation:

$$k_{on} = k_{on}^0 e^{\frac{\Delta G_{el}}{RT}} \quad (4)$$

with k_{on} and k_{on}^0 being the overall association constants in the presence and absence of electrostatic forces, and ΔG_{el} being the electrostatic free energy of the encounter complex. The accurate evaluation of the electrostatic energies represented in the ΔG_{el} term in Equation 4 is necessary to accurately predict the effect of long-range electrostatics in increasing protein–ligand association rates.

4. Influence of pH

The pH dependence of protein–ligand binding arises from changes in the pKa values of ionizable groups upon complex formation. It is well known that the pKa values of some ligand-titratable groups are perturbed from their solution values when the ligand binds to its receptor. The change in pKa values causes protons to be released or consumed upon binding [61-67]. The changes in pKa values of the ligand and the protein upon complexation are, therefore, of importance to the affinity of the drug [68], and pKa calculations of protein and ligand-titratable groups should thus be integrated into docking and screening algorithms [69,70].

The pH-dependent correction to a binding free energy $\Delta G_{bind,pH}$ is usually predicted using the following equation for various values of pH:

$$\Delta G_{bind,pH} = \int_{pH_{ref}}^{pH} \ln(10) kT \Delta Q_{bind} \quad (5)$$

where pH_{ref} represents the reference pH value when $\Delta G_{bind,pH}$ is assumed to be zero, k is the Boltzmann constant and T is the absolute temperature. The value ΔQ_{bind} refers to the amount of charges released into solution or consumed from it upon the binding ($\Delta Q_{bind} = Q_{PL} - Q_P - Q_L$). For more information on the pH dependence of the ligand-binding process, readers should consult the recent review by Jensen [71] and the references therein.

Some results showed considerable net changes in protonation states of complexes ΔQ_{bind} relative to free forms and, furthermore, the reported pH-dependent corrections to a binding free energy $\Delta Q_{bind,pH}$ were able to

induce the changes of the binding constant as high as three orders of magnitude [72].

The binding free energy can be affected significantly by both the ionic strength and the pH of the solution. The pH determines the charges of the ionizable residues, which in turn modulates the electrostatic interactions that contribute to protein-protein binding. Furthermore, these charges can change due to binding (giving rise to a pH-dependent binding constant), which affects the binding free energy. Thus, even if one is interested in the protein binding at pH 7 only it is still important to compute the pH-dependent part of the binding energy.

One of the most important findings was that the optimum pH values (pH of maximal stability) of the complex tend to be the same as the optimum pH values of the complex components. [71] This finding can be used in the homology-based prediction of the 3D structures of protein complexes, especially for evaluating and ranking putative models. It is more likely for a model to be correct if both components of the model complex and the entire complex have the same or at least similar values of the optimum pH. However, there is now very strong indication that pH effects due to binding induced changes in protonation states make a non-negligible contribution to the binding energy of most protein-protein complexes. This observation, together with more efficient pKa prediction methods (e.g. [73]) and the emergence of constant pH molecular dynamics simulations [74-80] to model the protonation-dependent structural changes will spark more experimental and theoretical work in pH effects on protein-protein binding.

Being protein-protein association a pH-dependent process, the binding affinity depends on the local pH. *In vivo* the association occurs in a particular cellular compartment, where the individual monomers are supposed to meet and form a complex. Since the monomers and the complex exist in the same micro environment, it is plausible that they coevolved toward its properties, in particular, toward the characteristic subcellular pH. It has been shown that the pH at which the monomers are most stable (pH-optimum) or the pH at which stability is almost pH-independent (pH-flat) of monomers are correlated with the pH-optimum of maximal affinity (pH-optimum of binding) or pH interval at which affinity is almost pH-independent (pH-flat of binding) of the complexes made of the corresponding monomers [81]. The analysis of interfacial properties of protein complexes demonstrates that pH-dependent properties can be roughly estimated using the interface charge alone. In addition, we introduce a parameter beta, proportional to the square root of the absolute product of the net charges of monomers, and show that protein complexes characterized with small or very large beta tend to have neutral pH-optimum. Furthermore, protein complexes made of monomers carrying the same polarity net charge at neutral pH have either very low or very high pH-optimum of binding. These findings are used to propose empirical rule for predicting pH-optimum of binding provided that the amino acid compositions of the corresponding monomers are available.

The change of the pKa's of ionizable groups upon the binding results in proton uptake/release and in turn in the pH dependence of the binding energy [71]. Similarly, the change of pKa's upon protein folding results in pH

dependence of the folding energy [82]. The magnitude of the energy change, either binding or folding energy, upon pH variations could be significant and in some cases could alter the minimal value by more than 50%, even at physiological pH [83-87]. Obviously at pH different from physiological pH, acidic/basic denaturation may occur and protein-protein binding may be abolished as well [88,89,90].

However, the physiological pH varies within cellular compartments (an excellent collection of characteristic subcellular pHs can be found in the works of Warwicker and coworkers [91,92]). For example, pH is nearly neutral in the cytoplasm, in the endoplasmic reticulum, and in mitochondria; it is more acidic in vacuoles, lysosomes (as low as pH 5), and the Golgi; and it is more basic in the nucleus and peroxisomes (as high as pH 8).

Protonation or deprotonation of titratable groups can cause changes in binding affinities, enzymatic activities, and structural properties of proteins and often represents a key event in enzymatic reactions. The problem of the determination of ionization state under physiological conditions is of great value to computational chemists, and several approaches have been proposed for pKa prediction, determination of protonation and redox equilibria in proteins, prediction of ionization states of titratable residues in proteins, and ionization/tautomerization of small molecules at binding sites [93-107].

One of the intriguing characteristics of tumors is the acidic microenvironment (i.e., an extracellular pH as low as 5.8), compared with normal tissues [108,109].

It is well known that pH plays an important role in modulating the functional activity of proteins. This phenomenon also applies to many transporters. Various mechanisms are responsible for the pH-sensitive transport activity, including proton-coupled transport, membrane potential-dependent transport, and the pH-dependent ionization of key amino acid residues at active binding sites. In particular, it has been documented that histidine plays an important role in pH-sensitive transporters such as proton-coupled folate transporter and peptide transporter 1 because histidine can interact differently with substrate molecules depending on its protonation state at various pH values [110,112,113].

Initial interest in the effect of residue ionization state on free energy predictions was piqued by a series of data reported by Bartlett et al. [114] that (a) showed K_i as a function of pH for a complex between penicillopepsin and a phosphonate-containing inhibitor and (b) proposed active site models, including protonation of two aspartates, to explain the pH vs K_i observation.

To investigate protein-protein interactions and the pH-optimum/pH-flat of binding, many global and interfacial properties need to be calculated, and attempted to correlate them with pH-best of binding. In many cases there is no correlation of the pH-dependence and isoelectric point (pI), neither with the mass, size, number of amino acids, and global charge of the complex or individual monomers. [81,115] However, Mitra et al. found that the pH-best of binding showed a significant tendency related to the macroscopic charges of individual monomers, and introduced the parameter β [81] as:

$$\beta = \sqrt{\text{abs}[Q(A)*Q(B)]} \quad (6)$$

The finding that the pH-dependence of the binding free energy can be related to the net charges of the monomers gives an opportunity of providing a clue of the pH-dependence of binding when the charges of the monomers which form the complex are known. It is important to emphasize that 3D structures of the monomers are not required to make prediction of the pH-optimum. Rather just the amino acid sequences of the monomers known to form a complex are sufficient. If the interfacial charge is known, either from existing 3D structure or other experiments, then a simple polynomial formula can be used to predict the pH-best of binding.

The binding process can be considered to be made of three important steps: (1) recognition: the receptor and the ligand must sense each other in the crowded cellular environment; (2) orientation: the receptor and the ligand must be properly oriented in such a way that their interfaces point toward each other; and (3) physical binding: the receptor and the ligand form 3D complex (oligomer).

During each of the above mentioned steps, the receptor and ligand will interact, and various forces will make contributions to the binding. Of these forces, the electrostatics in particular is expected to be the dominant force during steps 1 and 2; this is because there is no direct contact between the receptor and ligand in these steps, and all other forces, being short ranged, will have negligible effect. The electrostatic force is capable of causing conformational and ionization changes in both the receptor and the ligand and can therefore alter the protonation states. These changes may result in changes in the charge distribution, which will affect the electrostatic interactions. An even more complicated balance is expected to occur during the final stage (the physical binding), since more forces will be involved and both the receptor and the ligand interfaces will be shielded from the water phase. It is very likely that the dramatic changes in the interface environment and the newly formed interactions across the interface will induce conformational and protonation changes. In this review we will outline the importance of correctly assigning protonation states for each of the three above mentioned steps.

Therefore, in order to correctly model native binding, one should know what the physiological pH of the reaction is and be able to predict correctly the protonation states at this particular pH.

5. Influence of Protonation States

Molecular mechanics-based methods involving docking studies have been used to study the binding orientations and affinities for the complexation of glutamate by several receptors. As the experimental data showed that at physiological pH 7.4, those receptors exist in different protonation states, a theoretical study of the protonation of these ligands was done, including all of the species shown in 5% or more abundance in the potentiometric measurements [116]. In each case, the more favored positions of protons were calculated for mono-, di-, tri-, and tetraprotonated species, and molecular dynamics studies were performed to find the minimum energy conformations with simulated solvent effects. The energy

of complexation was calculated for each compound after minimization of the selected trajectory frames, as each species, being the contribution of each species to the total energy a function of their relative abundance, and the total energy for complexation of each receptor with glutamate the sum of all of these partial contributions. Relative values of complexation energies calculated in this way showed excellent agreement with experimental data.

To understand the role of protonation states in the initial steering, one must be able to predict the pK_a's of the ionizable groups of both the receptor and the ligand in their unbound state. Protonation states must be predicted on a theoretical basis because the use of x-ray crystallography cannot determine the position of hydrogen atoms [44]. Note that at this stage, the interactions are supposed to be quite weak and not to affect the energetics of individual molecules, including the pK_a's of ionizable groups. Thus, if the structures of unbound monomers are available, one could apply standard pK_a calculations [117,118] and be able to estimate the protonation states at the desired pH. Unfortunately, it is quite likely that there will be no a priori information about the physiological pH at which the complex forms, but it is necessary to predict it by either using the information about sub-cellular location provided in the Protein Data bank (PDB) [119] or by carrying out additional investigations to search the literature for homologous annotated protein(s). Obviously, the problem disappears in cases of predicting protonation states for a particular experiment carried out with well-defined pH.

Once the receptor and the ligand are within close proximity to each other, they are expected to sense each other's interfaces, and restrict their rotational and translational motions in order to orient their interfaces properly. Several works have discussed that the binding of proteins depends on the folding of proteins themselves [120], and physicochemical properties of the interface [121]. The recent work of Honig et al. [122] summarized that the complementarity between surfaces of binding proteins are based on the shape [123] and electrostatic properties of binding sites [124,125]. This is the reason why before two proteins associate to form a complex, they have to find the proper relative orientation by translational and rotational diffusion [126]. The protein association can be rate limited either by the diffusional approach of the subunits to form a transient complex, or by the subsequent conformational rearrangement to form the native complex [127].

During the final stage of the binding (the physical binding), after the partners are colocalized in space and time, and are properly oriented to each other, they should come together to form a complex. The final stage of binding is a complicated event, which may induce further conformational and ionization changes. In cases where no further conformational changes occur upon the physical binding, the protonation states of titratable residues situated at the interface can be affected by both the loss of interactions with the water phase and by newly formed interactions across the interface. The presence of the partner can also affect the pK_a of distant (away from interface) groups via long range electrostatic interactions as well. In cases of further conformational changes induced by the physical binding, especially if they are propagated throughout the entire structures, the pK_a

changes can be scattered everywhere, including the back of the molecules far away from the binding site.

Receptor-ligand binding is driven by the difference between the free energy of the complex relative to unbound states [128,129]. As binding free energy consists of several energy terms: including non-bonded terms such as electrostatic, van der Waals, and bonded terms such as the mechanical energy of the system, and the change of the entropy [130], and all of these energies are coupled with the protonation states and cannot be calculated separately, this coupling of various energy terms and effects is the main difficulty in modeling proton uptake/release upon the binding. Many experimental and computational studies were carried out in the past to reveal the relative role of the energy components, conformational, and ionization changes induced by the binding. Some investigations have focused on conformational flexibility demonstrating that it may disfavor or favor the association [131]. It was shown that different sites of the proteins undergo conformational changes but at different amplitudes [132]. Analysis of the disordered and ordered proteins complexes showed that although disordered proteins adopt particular conformations in their bound form, they are still much less rigid than ordered proteins [133].

The interaction between a pharmacologic ligand and a macromolecular receptor complex is commonly quantified by a measure of affinity (e.g., the K_d or K_i) and often characterized as the net attraction resulting from the making and breaking of multiple weak intermolecular bonds between ligand and receptor. However, all such interactions occur in some solvent/buffer environment, whether in vitro or in vivo, and include the rearrangement of bonds between the interacting species and buffer molecules. This affects the measure of affinity, as seen from the relationship between affinity (reciprocal of the equilibrium constant, K_{eq}) and the free energy change of the interaction,

$$\Delta G = -RT \ln K_{eq} = \Delta H - T\Delta S \quad (7)$$

in which the enthalpic (ΔH) and entropic (ΔS) terms include contributions of the buffer interactions. Binding occurs only if the total free energy decreases, thus solvent effects are as important as ligand-receptor bonds to the energy balance [134]. To the extent that the measure of affinity ignores the contribution of buffer, it is a poorer estimate of the affinity of ligand and macromolecule.

A common interaction with buffer is exchange of protons (protonation). Modern isothermal titration calorimetry (ITC) devices [135] measure an observed enthalpy change, ΔH_{obs} , as a composite of binding enthalpy (ΔH_{bind}) and enthalpy of ionization of the buffer (ΔH_{buffer}) according to [136]:

$$\Delta H_{obs} = \Delta H_{bind} + n_{H^+} \Delta H_{buffer} \quad (8)$$

from which the binding enthalpy (and, hence, affinity) can be determined by repeating the experiment at the same pH in buffers of different ΔH_{ion} (tabulated by Fukada and Takahashi, [137]). In this way, an affinity up to 3-fold higher than that obtained without correction for protonation, can be calculated [138].

These results demonstrate: (1) the importance of the protonation effect on ligand affinity, (2) the ability of

isothermal titration microcalorimetry to measure and correct for the effect and (3) the value of measuring the thermodynamic parameters that underlie the affinity of pharmacologic interactions [139].

Since the interfacial area is an important characteristic of the binding, its properties were statistically assessed, and indicated that the average interface has the same non-polar character as the protein surface as a whole, but carries fewer charged groups [140]. The fewer charged group observation is attributed to the fact that charged or polar protein residues that were initially exposed to the water phase tend to be partly or fully dehydrated due to being buried inside the protein complex [120,141]. However, the energy penalty due to the reduction of the proteins' flexibility and residues' desolvation can be compensated by newly formed interactions across the interface [122,142,143,144]. Correct assignment of protonation states would be crucial for accurate calculations of the balance of these opposing effects.

Methods for modeling receptor-ligand binding free energy vary from structure based virtual screening methods to more sophisticated and more computationally costly free energy calculation methods. Docking methods utilize these methods or use a scoring function to predict the most stable conformation of the ligand-receptor complex using the structures of the unbound receptor and ligand. In both cases, the free energy calculations and the docking results depend on various details and assumptions, such as the choice of protonation sites on the receptor and the ligand [145,146].

Proteins binding energy can be presented by equation (1) [147], where these free energies include various energy contributions, and in the case of changes of protonation state from unfolded to folded state, must include the energy of ionization [67].

Many factors can influence the protonation states of the receptor and the ligand, such as their concentration and the concentration of other molecules, temperature, and mostly the pH. Protomers differ in shape, functional groups, surface, and ability to form hydrogen bonds and therefore will experience different effects due to abovementioned factors. Only in very few cases does experimental data exist of the protonation states before and after the binding. In the vast majority of the cases, these protonation states (and changes) must be computationally predicted. However, the other frequently unknown factor is the pH of the binding. At some pH values, the binding may not involve protonation changes, while at other pH it may [148,149]. In summary, computing the binding free energy or predicting the binding mode via ab-initio docking while taking into account protonation and conformational changes and having little knowledge of the physiological pH, is a challenge [150].

In terms of predicting protonation states of proteins and receptor-ligand complexes, one can utilize standard pKa calculations if the structures of both the unbound and bound entities are available [117,151]. The problem becomes more complicated if several groups change protonation state [71]. Nielsen pointed out that that this might be accomplished by calculating the titration curves of all ionized groups in a range of pH values, while taking into account the effect of the interactions between them [152]. There are now online servers that can predict the best combination of protonation states for each ionizable

residue as well as the Gibbs free energy of binding for the ionization-optimized protein–ligand complex using coordinate information for the protein [153]. In terms of *in silico* screening, it was shown that the optimization of protonation states can play a significant role on selecting native-like ligand structures [154].

Different protonation was shown to result in alternative binding modes in docking simulations [155]. Another work showed the influence of variations of ligand protonation on the binding energy landscape of protein–ligand complexes using a structural consensus that was derived from multiple docking simulations [145]. Similarly, it was pointed out that the ionization state of ligand and receptor functional groups strongly affects their binding energy [156]. Cozzini et al. [157] found that both the ionization state and the associated local pH play significant roles in binding free energy and are potentially valuable in protein-protein complex formation. The importance of taking protonation into account in molecular docking has been highlighted in several other articles [147,158,159]. Another work pointed out that one of the difficulties in the implementation of residues' protonation state in binding energy calculations is that the ionized state for each defined residue cannot be well-defined, because protons are not static, they transfer between molecules [156]. A new computation method called “computational titration” that implements the pH dependence of proteins binding and allows scoring the protonation-dependent models was reported [156]. Another group developed the combinatorial method called SPORES (structure protonation and recognition system) to generate protonation states [44].

Some published results [160] suggested that different modeling of the protonation state of the ligand and the receptor does not influence the recognition of well-docked configurations. On the other hand, a virtual screening experiment against dihydrofolate reductase has shown that the protonation state of methotrexate and pteridine affects significantly their binding energies computed by the DOCK force field [161]. Furthermore, these two ligands obtained a high rank only in their protonated form, which is in agreement with experiment.

In other cases it has been found very similar correlation values between the standard and modified protonation states for all the scoring functions, but the results depend strongly on the solvation model used with the CHARMM force field [160], suggesting that using more realistic protonation states in conjunction with a reasonably accurate treatment of solvation effects may lead to better agreement with experimental binding affinities. Nevertheless, neither the GB model nor the Poisson equation ranks significantly better binding affinities than a distance dependent dielectric function.

6. Influence of pH and Protonation States in Docking Studies

Protein-ligand docking, one key method in structure-based drug design, is often used for the prediction of the interaction patterns of small molecules in the active site of a protein as well as in virtual screening of large databases to identify new and hopefully highly active lead structures. A large variety of different approaches for a solution of

this problem has been applied [6,8,162,163,164,165] ranging from fragment-based (e.g., DOCK [166,169], FlexX [6,8,11,170], SLIDE [171], SURFLEX [172], and GLIDE [173,174,175]) over stochastic optimization methods for finding the global minimum (e.g., GOLD [176,177,178], AutoDock [179,180], PLANTS [181,182,183], ICM [184], QXP [185], and PRO_LEADS [186,187]) to multiconformer docking (e.g., FRED [188]).

In any case, the computational methods should be able to model receptor-ligand binding at any pH, including pH at which the binding induces proton transfer and is pH-dependent. However, it was demonstrated that predicting the absolute binding energy, calculating the pH dependence and scoring small molecule binding to the target protein taking into consideration plausible protonation changes is still a challenge.

Despite the good results of some approaches, a universal docking tool (algorithm and scoring function) that outperforms all others on every system is not available at the moment. Besides insufficient conformational sampling, the scoring functions were identified as the main reason for these failures [162,163,189,190]. Additionally, the importance of the preparation of the protein and especially the ligand structure and its influence on the docking results was stressed (once again) in these studies. Because the position of hydrogen atoms cannot be determined experimentally by X-ray crystallography, protonation and tautomeric states must be predicted on a theoretical basis, and the assignment of atomic hybridization and bond orders are not always straightforward.

For scoring functions, which take actual hydrogen positions into account for the calculation of hydrogen-bonding strengths, the correct placement of these atoms is often the key to identifying the correct structure with a docking approach. The results from the docking experiments with protonation states of the ligands from a standardized test set show that the ligand protonation has a significant influence on the accuracy of docking results. Although the manually revised protonation still outperforms the automatically assigned standard protonation, it is clear that some of the advantages are based on knowledge about the binding pocket and binding pose. For example, SPORES [44] intentionally handles ligands and proteins separately, which is an advantage in the preparation of data sets for virtual screening, where, on one hand, the number of structures prevents manual preparation and, on the other, an equal treatment of active and inactive structures is needed to avoid artificial enrichments. Additional protonation states lead to a reduction of the success rate on this well prepared test set. These problems are mainly caused by the generation method, which does not discriminate between favorable and unfavorable protonation states. Together with the scoring functions this leads in some cases to over favoring of highly charged ligands because of the possibility of forming charged interactions to the protein thus creating false positive docking results. When comparing two docking programs, PLANTS and GOLD, it is noticeable that PLANTS gives an overall better performance, but the influence of the protonation states is about the same for both programs despite the different approaches in terms of scoring function [176,181]. But the inclusion of additional protonation states may prove useful when less information

is available about the targets or when the performance of the standard protonation is poor for the specific target. The main drawback, the creation of unreasonable protonation states, can be improved by either a fast incremental pKa calculation or by a rule based approach to avoid too highly charged ligand structures. Until such a procedure is implemented, it is recommended to use only the standard protonation in docking and especially in virtual screening experiments. Only in the case that for a specific target the experimental results cannot be reproduced in redocking experiments, additional protonation states should be produced and docked after removing the implausible ones.

The structural picture that has emerged from all the above mentioned studies reveals multiple direct and water-mediated interactions between drug amino groups and the host enzyme. It is likely that the nature and strength of these interactions will be influenced by the protonation state of the amino groups.

Recently, it has been shown that the binding of various 4,5-disubstituted 2-DOS aminoglycosides to the A site model of an RNA oligomer is coupled to drug protonation in a pH-dependent manner, [191] and, in a similar way, at pH 5.5, the binding of paromomycin, neomycin, and lividomycin sulfate to the A site of the same oligonucleotide is independent of drug protonation. In the absence of drug protonation effects, paromomycin binding to the A site RNA oligomer is primarily entropy driven. The RNA binding of paromomycin and neomycin sulfate at pH 9.0, as well as the binding of lividomycin sulfate at pH 8.8, is linked to the uptake of protons [192].

There has been a great deal of theoretical work done to calculate acidic properties of numerous organic compounds, but most efforts have focused on gas phase basicities, proton affinities, or relative pKa values in aqueous solution [193,194]. It was not until recently that the effort has been made to calculate absolute pKa values in aqueous solution accurately [195-200]. Recently, Jang et al. have used first principles QM methods (density functional theory (DFT), B3LYP, in combination with the Poisson-Boltzmann (PB) continuum-solvation model) to calculate pKa values for DNA bases in aqueous solution [200,201]. The agreement between calculated and experimental values for a series of pyrimidine derivatives is excellent (within 0.7 pKa units), having been demonstrated that these methods can be used to predict the site of ionization where multiple, potential sites exist. These studies suggested to us that ionization and tautomerization might be linked, in that ionization might provoke a shift to an alternative tautomeric form.

The continuous development of implicit solvent models (also known as continuum solvent models) during the last decades together with their simplicity and low computational costs in front of the more complex explicit solvent simulations, made possible the theoretical prediction of physicochemical properties of a broad variety of chemical species in different solution media [202-207]. One of the remaining subjects in continuum solvent calculations is the accurate calculation of acidity constants (Ka) or pKa in its negative logarithmic form [208].

A calculation method with moderate computational demands is given by the Poisson-Boltzmann (PB) theory, which is a continuum electrostatics procedure. In general, inter and intra-molecular electrostatic interactions are

among the key factors determining the functional properties and structural stability of proteins [209] and these interactions are directly related to charges of titratable groups, which in turn can have modulated protonation states. There is strong evidence for this, e.g., the pH-dependence of enzyme activity. The ubiquitously present water molecules in proteins and protein-ligand complexes are simulated as a dielectric sphere, and this is the continuum part in the PB model.

The explicit treatment of water molecules together with the continuum solvent model has been demonstrated to improve the pKa predictions with thermodynamic cycles by including the real quantum hydrogen bond interactions and by breaking the artificial boundary delimited by the dielectric continuum and therefore reducing the influence of the solute-continuum electrostatic interaction [210-218]. Since the artificial interaction between the dielectric continuum and the solute also limits the performance of the aqueous isodesmic reaction, a similar behavior is expected when explicit water molecules are included. According to this, a water molecule was added to both pyridine and pyridinium molecules acting as hydrogen bond donor/acceptor, respectively. In the case of carbon acids, all the acidic carbons are placed in a-position respect to a carbonyl group. So, a single water molecule was placed (in both acids and conjugate bases) donating a hydrogen bond to the carbonyl oxygen to describe the enolate resonance structures according to Ho and Coote's criterion [210].

Classical Hansch analysis for most series of ionizable compounds cannot identify the active species as being neutral or ionized. The $\log P^N$ and $\log P^I$ of these series are too highly correlated. Scherrer and Leo [219] found a simple solution to this problem by just combining data obtained at more than one pH (multi-pH QSAR). A bonus is that the statistical power of a series is increased one fold for each pH level added. Datasets of simple homologous series serve as proofs of concept. Two of these reveal hidden correlations involving both the neutral and ionized species. These "true" correlations would be impossible to find at a single pH. An example is the inhibition of monoamine oxidase by a series of primary n-alkylamines. Multi-pH QSAR indicates that the protonated amine enhances inhibition and the neutral species diminishes it, illustrating the first substantial use in QSAR of the species-specific terms, $\log D^N$ (neutral) and $\log D^I$ (ionized). Multi-pH QSAR can give direction to lead development and facilitate the In-Silico modeling of ADMET processes by associating rate-limiting steps with the correct ionization state.

In a typical QSAR analysis, $\log P^N$ and $\log P^I$ often individually produce identical correlation coefficients for a series of compounds. In these cases, it is not possible to combine these parameters in the same equation. They are too highly correlated with each other. Multi-pH QSAR solves the problem by breaking this correlation.

There is a second requirement for this analysis. We need property terms to represent the concentration of neutral and ionized species at a site at any pH. This is provided by species-specific log D terms, $\log D^N$ and $\log D^I$. The latter are obtained by factoring the octanol/water distribution coefficient into its neutral and ionized components (Equations 9-13) [220,221,222]. These examples provide the first substantial applications

involving the use of $\log D^N$ and $\log D^I$. They have such compelling advantages, they will undoubtedly push $\log D$ into the “Old-School” category as more data on $\log P^I$ becomes available [220]. $\log D^I$ is essentially $\log P^I$ adjusted for the fraction ionized at a given pH ($P^I \times F_{i(w)}$). The calculations of $\log D^N$ and $\log D^I$, are described below.

$$D = \left\{ \begin{array}{l} [\text{Neutral species}]_{\text{oct}} \\ + [\text{Ion pair+non-paired ion species}]_{\text{oct}} \end{array} \right\} \quad (9)$$

$$/ [\text{All species}]_w = D^N + D^I$$

(For acids)

$$\log D^N = \log P^N - \log \left(1 + 10^{(\text{pH}-\text{pKa})} \right) \quad (10)$$

$$\log D^I = \log \left(P^I \times 10^{(\text{pH}-\text{pKa})} \right) - \log \left(1 + 10^{(\text{pKa}-\text{pH})} \right) \quad (11)$$

(For bases)

$$\log D^N = \log P^N - \log \left(1 + 10^{(\text{pH}-\text{pKa})} \right) \quad (12)$$

$$\log D^I = \log \left(P^I \times 10^{(\text{pKa}-\text{pH})} \right) - \log \left(1 + 10^{(\text{pKa}-\text{pH})} \right) \quad (13)$$

The thermodynamic cycle of partitioning between octanol and water and ionization in water and ionization in octanol allows one to equate ($\log P^N - \log P^I$) to ($\text{pK}_{a,\text{oct}} - \text{pK}_a$) for acids, or to ($\text{pK}_a - \text{pK}_{a,\text{oct}}$) for bases, knowing that:

a) P^I is the ratio of the concentration of an ionized compound in the octanol phase, to the concentration of ionized compound in the aqueous phase. It is dependent on the concentration of counter ion in the aqueous phase. This is a non-classical definition because it does not differentiate the ion-pair from the dissociated ion forms in each phase.

b) D^I is the ratio of the concentration of the ion pair (plus any dissociated-ion form) of a compound in the octanol phase to the total concentration of all forms of the compound in the aqueous phase at a given pH. D^I is basically P^I adjusted for the fraction ionized. It is, likewise, dependent on the counter-ion concentration.

c) D^N is the ratio of the concentration of the unionized portion of a compound in the octanol phase, to the total concentration of all forms of the compound in the aqueous phase at a given pH.

d) The distribution coefficient, D , is the ratio of the total concentration of the neutral and ionized forms of a compound in the octanol phase, at a given pH, to the total of all forms of the compound in the aqueous phase.

e) $\text{pK}_{a,\text{oct}}$ is equal to the pH of a physiological saline phase when the concentrations of the ionized and neutral species in the octanol phase are equal.

Generally, drugs are weakly ionizable molecules, which can efficiently penetrate the plasma membrane in their unionized form [223]. In this context, the absorbed drug which may be unionized at low pH (stomach, $\text{pH} = 2$) can easily pass across the membrane but when it reaches the blood ($\text{pH} = 7 \times 4$) reconverts to ionized form and cannot pass back across the membrane. Hence, both the pH and the ionization normally play important roles in pH

partition of many drugs. Moreover, drugs with pKa close to physiological pH can be considered for interesting pharmacological properties [223].

Although the protonation of this drug might take place at more than one basic site, an appropriate approach is required in changing the pKa of a drug. However, the prediction of absolute pKa is extremely difficult. In order to rationalize the most basic site, the quantitative estimation of proton affinities at various positions is necessary.

Amino and amido functionalized organic compounds are ubiquitous in nature. Their biological importance [224] has led to extensive studies of their structural and physicochemical properties. In addition to their relevance in biochemistry and pharmaceutical chemistry, amines and amides are attractive building blocks in supramolecular chemistry [225,226,227]. Depending on solution pH, amine-based ligands can act as both cation and anion chelators [228,229]. Amide-containing receptors also exhibit dual cation/anion binding properties. They have emerged as attractive building blocks for a variety of anion receptors due to their relatively strong hydrogen bond donor N-H groups [228,229]. In addition, they contain oxygen and nitrogen heteroatoms that can coordinate with metal ions [230,231].

The use of solution-phase optimized geometries gives calculated pKa values in excellent agreement with experimental measurements.

Molecular docking results from Paz et al. [232] provided an interaction model of LTA4 with LTA4H in excellent correlation with the experimental data: the substrate fits in the active site of the enzyme arranging its functional groups according to the reaction mechanism proposed by several authors [233,234,235,236]. Both species of LTA4, protonated and deprotonated, bind to the same site with proper orientation of the key groups for the enzymatic reaction. The substrate is held in the active site by multiple weak attractive interactions along its surface and a hydrogen bond to Arg563

In other study, quinoxalines were docked considering two different possible protonation states: the neutral structure and one monoprotinated form (obtained by adding a hydrogen atom to the N4 of the molecules). In this way the possibility of binding with protonated structures has been explored, showing no good results with the protonated form, as the predicted binding energies were higher than those obtained from the neutral structures and they did not correlate with the experimental data [237].

Patronov et al. investigated the influence of pH on the predictive ability, deriving different correlations at two pH values: 5.0 and 7.0. The only amino acid sensitive to pH in the range 5.0 to 7.0 is Histidine. The pKa of the His imidazole is 6.0, thus making His protonated and very hydrophilic at pH 5.0 and unprotonated and less hydrophilic at pH 7.0. The influence of pH on the affinity of peptide binding to HLA-DP proteins has two potential aspects: influence on peptide protonation/deprotonation and influence on protein binding site protonation/deprotonation [238].

It is worth to point out that despite the extensive investigation of ionization states in proteins, the effect of ligand protonation remained unexplored. Analyzing the active site of them, however, a significant impact of ligand

protonation on screening efficacy can be expected. Quite recently Knox et al. [147] showed on alphaestrogen receptor that ligand preparations tautomerism, protonation, stereochemical, and conformational states had not affected the outcome of virtual screening studies. However, their findings cannot be generalized and should be explored with several targets, scoring functions and docking algorithms. Warren and co-workers screened their ionized ligand databases against 8 different proteins with 10 docking programs and 37 scoring functions. Although the docking programs reproduced experimental binding modes, none of the scoring functions were able to discriminate between highly potent and inactive compounds [239]. So, Polgar et al. [70], attempted to investigate the impact of ligand protonation on the results of virtual screening. In the study they showed that ligand protonation can affect the results of virtual screening against BACE1; however, it is much more dependent on the characteristics of the scoring functions used. Dock, Gold, ChemScores, and GOLDScore were not able to select good poses from all the protomers, while the SurfexScore was able to find a reliable pose resulting in acceptable enrichment factors. In all cases neutral molecules or compounds protonated at pH 6.6 were ranked well.

Based on these results obtained for BACE1 they suggested the investigation of the effect of ligand protonation in conjunction with docking algorithms and scoring functions for each different target. Dealing with all possible protomers does not obviously improve screening efficacy but definitely increases computational resources allocated to the screening campaign.

The protonation of the ligand molecule and the protein binding site has a significant influence on the results obtained by protein-ligand docking. Due to the inability of X-ray crystallography to resolve the hydrogen atom in protein and protein complex structures, the correct protonation for the protein and the ligand has to be assigned on a theoretical basis before the structures can be used. Because of the local environment inside the binding site and because of the influence of the ligand and the protein onto each other, the ligand protonation can differ from the protonation one would expect for the ligand in solution under physiological conditions. Hence for protein-ligand docking different proton states of the ligand have to be taken into account.

Ten Brink et al. presented a new approach in which the ligand atoms considered for the combinatorial method not chosen from predefined functional groups but according to the calculated pKa values which leads to a wider variety but with a smaller overall number of protonation states [240].

It has also been proposed that electrostatic interactions play a relatively minor role in determining binding affinity since thermodynamic parameters obtained at pH 5 are essentially unchanged from those obtained at pH 9. It is not clear, however, whether this last point is to be reassessed in light of the Aubard group's subsequent finding that the pKa of N2 is substantially shifted on DNA binding [241].

The prediction of pKa values of protein-ligand complexes is of great practical importance to molecular modeling in rational drug design because the strength of ligand binding (i.e. the binding free energy) is dependent

on the protonation states of the ionizable residues and functional groups in the active site. Moreover, upon ligand binding, the pKa values of the ionizable groups may change, resulting in the uptake or release of protons. This leads to a pH-dependent correction to the binding free energy that can be computed using the pKa values. Additionally, studies of enzyme reaction mechanisms will benefit from the correct assignment of the protonation states of the ionizable groups in the active site and those of the substrate.

There are several ways to compute pKa values for proteins [78,94,96,98,100,101,102,105,242-256], including utilizing free energies from molecular dynamics simulations [78,102,242,253,255,256], and from numerical solutions of the linearized Poisson-Boltzmann equation (LPBE) [94,96,98,100,101,105,243-255]. In the LPBE approaches the protein is described by the partial charges from a classical force field and embedded within a dielectric continuum with dielectric constants (ϵ) of 80 for the solvent and between 4 and 20 for the interior of the protein. pKa shifts are obtained by computing the difference in electrostatic energy of a charged residue with that of its neutral form; the resulting shift is added to a model pKa value. Typically these models [100,101,243,244,245] have root-mean-square errors with respect to experiment of less than 1 pH unit.

Difficulties arise when extending PB approaches to incorporate a ligand bound to a protein. In addition to pKmodel values for the ionizable groups in the ligand, charges and other parameters that describe the ligand are required. Consequently, only a small number of computational studies on protein-ligand complexes have been performed.

PROPKA 2.0 is an open source code that is freely distributed under the BSD style license [257,258], with the aim of being a useful tool for the interpretation of pH-dependent effects of ligand binding, thereby generating more experimental data - badly needed for a thorough validation of new methods for pKa prediction of protein-ligand complexes.

Several software tools to generate ligand tautomers and protomers have also been published. An overview of software packages dealing with various problems regarding tautomers and their representation was given by Warr [259]. Despite the mentioned publications about protonation and tautomers in drug design, remarkably few publications deal with the problem of protomeric and tautomeric states in docking. Pospisil et al. [158] gave a short review about tautomers in drug design in which they concluded, that the inclusion of tautomers can be beneficial for virtual screening as the coverage of chemical space is improved and false negative hits are avoided. Brenk et al. [260] conducted a comparison of virtual screening and high throughput screening on the *E. coli* dihydrofolate reductase (DHFR) target, in which they emphasize the need of tautomeric states and protomers to avoid false negative predictions by overlooking likely protonation states or to create false positive predictions by including unrealistic protonation states. As recently shown, the enumeration of all possible protonation states and the use of all these in the docking is contra-productive. Very often unreasonably highly-charged structures with wrong poses are generated nevertheless obtaining high scores due

to the functional forms of today's scoring functions [44,155].

The structure preparation tool SPORES [44] was introduced by ten Brink et al. along with a study on the influence of the introduction of ligand protonation states into the docking process. A so-called standard protonation was defined based on a set of rules designed to give, at least in most cases, the protomer corresponding to physiological conditions. With this standard protonation, good docking results were obtained but it could not fully compete with the manually derived ligands.

Their studies demonstrated that unreasonable states should be removed prior to docking or must be additionally penalized. This can be done by calculating pKa values and add an energy penalty for the protomers which are not the most favorable at physiological conditions as done in GLIDE [173,174,175] or in an approximatively way as in the earlier version of SPORES [44].

An interesting finding is the performance of the most populated protonation state of the ligands predicted for pH = 7.4. Even without the penalty term which presses the ranking toward this structure, 80% of the correctly docked ligands and 78% of all ligands have the best ranked structure in this protonation state. This shows that for the majority of the proteins the ligand protonation has not to be adapted to the binding site and explains the good performance of the standard protonation. For automated ligand preparation tools, which do not take the binding site into account like SPORES, this leads to the conclusion that a good representation of the ligand protonation at physiological pH can result in docking success rates comparable to manually revised structures at least for standard protein targets. Only metal-ion-containing binding sites have been identified here to be in need of a special treatment since the ions have a great influence on the ligand protonation. Thus, it is beneficial to include some information on the binding site e.g. by modifying the pH value for which the microspecies distribution is calculated. Overall, the metalloproteins show a reduced docking success rate compared to the non-metalloproteins from the data set in all approaches. This is especially the case for the standard protonation as well as high filter settings (pKa based approach), in which the correct protomer is very often not among the ones used for docking. But if the protonation can adopt more flexibly to the binding site (ASTEX and microspecies approach), results between metalloproteins and non-metalloproteins are much more comparable.

In addition to global pH effects, the local effective "pH" in specific loci can be perturbed enough to protonate or deprotonate acidic/basic residues, or, in other words, the pKa of ionizable residues can be affected by environment. This is a significant issue in many drug design projects because virtual screening experiments of databases against protein targets are almost exclusively carried out with no or very limited attempts to optimize the ionization state of residues surrounding the binding site and/or functional groups on the ligands themselves. It is not difficult to envision cases where the protonation of a single acid residue or ligand functional group would make the difference between a favorable and unfavorable binding event. "Getting it right" argues for consideration of pH effects in virtual screening protocols. Kellogg et al.

have developed a new modeling tool, called computational titration, which includes a careful modeling of the ionization states and resonance forms for the ligands and protein residues at the binding site and the evaluation of the free energy of binding for modeled complexes by their HINT force field [261].

One of the consequences of this approach is that it forces us to recognize that there is not a *single* global model for each protein–ligand system, i.e., a well-determined ionized state for each defined residue, because protons are not static and the ionization state of residues is a group function. In other words, there may be multiple energetically accessible states for each complex. Proton transfer between molecules and, in particular, proton migration across hydrogen bonds, has been identified as one of the fundamental mechanisms for biological processes [262]. Thus, modeling protein–ligand complexes of this type as an ensemble of multiple ionization models is a more biologically reasonable approach.

7. Conclusions

The association of two molecular entities in the biological environment is a process governed by free energy, that is to say that entropy is important. Especially significant is the role of solvent, i.e., water. The displacement of water molecules when two biological molecules associate is clearly a major source of entropy. This is manifested by the hydrophobic effect and desolvation of functional groups buried by the association. Solvent effects also modulate the ionization states of acidic and basic functional groups on the protein side chains (or ligand functional groups) and thereby biomolecular associations. These properties have been little studied and are poorly understood but must be computationally accounted for in any realistic model. It is well known that "local" pH at specific sites within proteins can be considerably different than the global solution pH, and this, in turn, affects the ionization state of these residues [93,94,97,102,105,107]. Computational approaches to target the above factors (entropy, hydrophobicity, solvation/desolvation, and pH) in the biological environment are currently at the cutting edge of simulations of molecular interactions. There are, loosely, three tactics applied in modeling complex systems such as those involving protein–ligand interactions. The first is based on extensive molecular dynamics simulations of the complete system in a box including explicit water molecules with boundary conditions. By configuration of these simulations for thermodynamic integration (TI) or free energy perturbation (FEP), the free energy for specific events, e.g., ligand binding, can be estimated [263,264,265,266]. Alternatively, the linear response method analyzes the states of multiple dynamics simulations to obtain a semi-empirical estimation of free energy [267,268]. These methods are computationally expensive to perform and are tied to the quality of the underlying Newtonian molecular mechanics (MM) force fields. Thus, only interaction types specifically programmed into the force field will be modeled and observed. In particular, MM force fields do not include specific terms for the hydrophobic effect and hydrophobic–hydrophobic interactions are usually

indicated to be energetically unfavorable. The second class of tactic for simulating the biological environment is based on the electrostatic properties of the molecules and their response to the dielectric variations at the molecule/solvent interface. Several implementations of solutions to the Poisson–Boltzmann equation (PBE) have been used in the biological environment [93,107,269]. While PBE approaches are very robust in terms of representing electrostatic effects, particularly at molecular surfaces, they do not have terms to represent other non-covalent forces so can be of limited utility in cases that are not dominated by electrostatic effects. For this reason, some of the more recent simulations have utilized methods combining PBE (or generalized Born, a faster, less robust electrostatic method) with FEP or TI calculations [270,271,272]. The third tactic for understanding the biological environment is the application of empirical free energy scoring algorithms. These tools have often been developed for feedback in docking or de novo molecule design programs, and thus have been optimized for speed. In general, simple metrics such as partial charge and solvent accessible surface area, combined with structural features as extracted from experimental measurements, produce a “score” that can often be correlated with free energy [48,269,273]. For example, the hydrophobic effect is usually represented by a nonpolar surface contact area. While useful within their native context, i.e., docking or de novo building, these algorithms are often invalid outside their training and validation environment. The purpose of this article is to consider the more general problem of the applicability of 3D-molecular structures in a given environment, the energetic penalties involved with using a non-equilibrium form and possible changes in structure between different environments. This fundamental concern for 3D-molecular design encompasses questions such as solvation, acidity and basicity and conformational flexibility. Above it, it requires that we develop a sensible strategy to consider all these factors in drug design. We are accustomed to considering the receptor as a moving target (but then generally fix its structure), but often ignore the chameleon-like character of the ligand completely.

We have neglected some aspects of tautomerism for too long, but taking it into account will force us to think about other aspects of 3D drug design. These all relate to using the correct structure (tautomer, protonation state and conformation) and being able to relate the free energy of the structure used to that of the system (both ligand and receptor) under physiological conditions. This is doable, but requires that we rethink the physical models behind the methods that we use and use the power of modern hardware to improve them, rather than just to calculate more molecules.

Competing Interests

The author declare no competing financial interest.

Acknowledgement

The autor thanks the MCINN Project Consolider Ingenio CSD2010-00065 for financial support.

References

- [1] Pauling, L. The nature of the chemical bond. IV. The energy of single bonds and the relative electronegativity of atoms. *J. Am. Chem. Soc.*, 1932, *54*, 3570-3582.
- [2] Pearson, R.G. Chemical hardness, in: Sen, K. D. (Ed.), *Structure and Bonding*, vol. 80, Springer, Berlin, 1993.
- [3] Fukui, K.; Yonesawa, Y.; Shingu, H. A molecular orbital theory of reactivity in aromatic hydrocarbons. *J. Chem. Phys.*, 1952, *20*, 722-725.
- [4] Yang, W.; Parr, R.G. Hardness, softness, and the Fukui function in the electronic theory of metals and catalysis, *Proc. Natl. Acad. Sci. USA*, 1985, *82*, 6723-6726.
- [5] Lengauer, T.; Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol.*, 1996, *6*, 402-406.
- [6] Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: Methods and applications. *Nat. Drug Discovery*, 2004, *3*, 935-949.
- [7] Pearlman, D. A.; Kollman, P. A. The calculated free energy effects of 5-methyl cytosine on the B to Z transition in DNA, *Biopolymers*, 1990, *29*, 1193-1209.
- [8] Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors, *Angew. Chem. Int. Ed. Engl.*, 2002, *41*, 2644-2676.
- [9] Abagyan, R.; Totrov, M.; Kuznetsov, D. ICM-A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J. Comput. Chem.*, 1994, *15*, 488-506.
- [10] Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.*, 1995, *245*, 43-53.
- [11] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, 1996, *261*, 470-489.
- [12] Ewing, T. J. A.; Kuntz, I. D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.*, 1997, *18*, 1175-1189.
- [13] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, 1998, *19*, 1639-1662.
- [14] Wu, G.; Robertson, D. H.; Brooks, C. L., III; Vieth, M. Detailed analysis of grid-based molecular docking: A case study of CDOCKERA CHARMM-based MD docking algorithm. *J. Comput. Chem.*, 2003, *24*, 1549-1562.
- [15] Bühm, H. J.; Stahl, M. The use of scoring functions in drug discovery applications. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: New York, 2002; Vol. 18, pp 41-87.
- [16] Shao, M.; Wang, S.; Wang, C.; Yuan, X.; Li, S. C.; Zheng, W.; Bu, D. Incorporating Ab Initio energy into threading approaches for protein structure prediction. *BMC Bioinformatics*. 2011, *12* (Suppl 1), 54-67.
- [17] de Jongel, M. R.; Vinkers, M.; van Lenthe, J. H.; Daeyaert, F.; Bush, I. J.; van Dam, H. J. J.; Sherwood, P.; Guest, M. F. Ab Initio potential grid based docking: From High Performance Computing to In Silico Screening. *COMPLIFE*, 2007, *940*, 168-178.
- [18] V. Vasilyev, V.; Bliznyuk, A. Application of semiempirical quantum chemical methods as a scoring function in docking. *Theor. Chem. Acc.*, 2004, *112*, 313-317.
- [19] Bietz, S.; Urbaczek, S.; Schulz, B.; Rarey, M. Protoss: a holistic approach to predict tautomers and protonation states in protein-ligand complexes. *J. Cheminf.*, 2014, *6*, 1-12.
- [20] Linderstrøm-Lang, K. On the ionization of proteins. *Cr. Trav. Lab. Carlsberg*, 1924, *15*, 1-29.
- [21] Stewart, R. *The proton: appellation to organic chemistry*. Academic Press, New York, 1985.
- [22] Carrol, F.A. *Perspectives on structure and mechanism in organic chemistry*. Brooks-Cole, New York, 1998.
- [23] Zhao, J.; Zhang, R. Proton transfer reaction rate constants between hydronium ion (H₃O⁺) and volatile organic compounds. *Atmos. Environ.*, 2004, *38*, 2177-2185.
- [24] Kennedy, R. A.; Mayhew, Ch. A.; Thomas, R.; Watts, P. Reactions of H₃O⁺ with a number of bromine containing fully and partially halogenated hydrocarbons. *Int. J. Mass Spectrom.*, 2003, *223*, 627-637.

- [25] Bouchoux, G. Gas-phase basicities of polyfunctional molecules. Part 1: Theory and methods. *Mass Spectrom. Rev.*, 2007, 26, 775-835.
- [26] Deakyne, C.A. Proton affinities and gas-phase basicities: theoretical methods and structural effects. *Int. J. Mass. Spectrom.*, 2003, 227, 601-616.
- [27] Lias, S. G.; Liebman, J. F.; Levine, R. D. Evaluated gas phase basicities and proton affinities of molecules; heats of formation of protonated molecules. *J. Phys. Chem. Ref. Data*, 1984, 13, 695-808.
- [28] Rajak, S. J.; Ghosha, D. C. Correlating the site selectivity of protonation in some ambidentate molecules in terms of the dual descriptor. *Eur. Phys. J. D.*, 2012, 66, 6 pp.
- [29] Ghosh, D. C. *A theoretical study of some selected molecules and their protonation by the application of CNDO method* (Premch and Roych and Research Studentship Award, University of Calcutta, 1976).
- [30] Halperin, I.; Ma, B.; Wolfson, H.; Nussinov, R. Principles of docking: an overview of search algorithms and a guide to scoring functions. *Proteins: Struct. Funct. Bioinf.*, 2002, 47, 409-443.
- [31] Smith, G. R.; Sternberg, M. J. E. Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct. Biol.*, 2002, 12, 28-35.
- [32] Ritchie, D. Recent progress and future directions in protein-protein docking. *Curr. Prot. Pept. Sci.*, 2008, 9, 1-15.
- [33] Chen, R.; Li, L.; Weng, Z. ZDOCK: an initial-stage protein-docking algorithm. *Proteins: Struct. Funct. Bioinf.*, 2003, 52, 80-87.
- [34] Macindoe, G.; Mavridis, L.; Venkatraman, V.; Devignes, M.; Ritchie, D. HexServer: an FFT-based protein docking server powered by graphics processors. *Nucleic Acids Res.*, 2010, 38, W445-W449.
- [35] Kozakov, D.; Brenke, R.; Comeau, S.; Vajda, S. PIPER: an FFT-based protein docking program with pairwise potentials. *Proteins: Struct. Funct. Bioinf.*, 2006, 65, 392-406.
- [36] Tovchigrechko, A.; Vakser, I. A. GRAMM-X public web server for protein-protein docking. *Nucleic Acids Res.*, 2006, 34, W310-W314.
- [37] Mandell, J. G.; Roberts, V. A.; Pique, M. E.; Kotlovyy, V.; Mitchell, J. C.; Nelson, E.; Tsigelny, I.; ten Eyck, L. F. Protein docking using continuum electrostatics and geometric fit. *Protein Eng.*, 2001, 14, 105-113.
- [38] Andrusier, N.; Mashlach, E.; Nussinov, R.; Wolfson, H.J. Principles of flexible protein-protein docking. *Proteins: Struct. Funct. Bioinf.*, 2008, 73, 271-289.
- [39] Lyskov, S.; Gray, J. J. The Rosetta Dock server for local protein-protein docking. *Nucleic Acids Res.*, 2008, 36, W233-W238.
- [40] Dominguez, C.; Boelens, R.; Bonvin, A. M. J. J. HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J. Am. Chem. Soc.*, 2003, 125, 1731-1737.
- [41] Ogmen, U.; Keskin, O.; Aytunas, S.; Nussinov, R.; Gursoy, A. PRISM: protein interactions by structural matching. *Nucleic Acids Res.*, 2005, 134, W331-W336.
- [42] Grosdidier, A.; Zoete, V.; Michielin, O. Swiss Dock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.*, 2011, 39, W270-W277.
- [43] Ritchie, D.; Kozakov, D.; Vajda, S. Accelerating and focusing protein-protein docking correlations using multi-dimensional rotational FFT generating functions. *Bioinformatics*, 2008, 24, 1865-1873.
- [44] ten Brink, T.; Exner, T. E. Influence of protonation, tautomeric, and stereoisomeric states on protein-ligand docking results. *J. Chem. Inf. Model.*, 2009, 49, 1535-1546.
- [45] Sippl, M. J. Calculation of conformational ensembles from potentials of mean force. An approach to the knowledge-based prediction of local structures in globular proteins. *J. Mol. Biol.*, 1990, 213, 859-883.
- [46] DeWitte, R.; Shakhnovich, E. SMOG: De novo design method based on simple, fast, and accurate free energy estimates. 1. Methodology and supporting evidence. *J. Am. Chem. Soc.*, 1996, 118, 11733-11744.
- [47] Mitchell, J. B. O.; Laskowski, R. A.; Alexander, A.; Thornton, J. M. BLEEP-Potential of mean force describing protein-lig and interactions: I. Generating potential. *J. Comput. Chem.*, 1999, 20, 1165-1176.
- [48] Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *J. Med. Chem.*, 1999, 42, 791-804.
- [49] Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein-ligand interactions. *J. Mol. Biol.*, 2000, 295, 337-356.
- [50] Stubbs, M. T.; Reyda, S.; Dullweber, F.; Möller, M.; Klebe, G.; Dorsch, D.; Mederski, W.; Wurzigler, H. pH-dependent binding modes observed in trypsin crystals: Lessons for structure-based drug design. *Chem Bio Chem*, 2002, 3, 246-249.
- [51] Mardis, K. L.; Luo, R.; Gilson, M. K. Interpreting trends in the binding of cyclic ureas to HIV-1 protease. *J. Mol. Biol.*, 2001, 309, 507-517.
- [52] Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W. Modeling electrostatic effects in proteins. *Biochim. Biophys. Acta*, 2006, 1764, 1647-1676.
- [53] Brock, K.; Talley, K.; Coley, K.; Kundrotas, P.; Alexov, E. Optimization of electrostatic interactions in protein-protein complexes. *Biophys. J.*, 2007, 83, 3340-3352.
- [54] Tjong, H.; Zhou, H-X. Accurate calculations of binding, folding, and transfer free energies by a scaled generalized born method. *J. Chem. Theory Comput.*, 2008, 4, 1733-1744.
- [55] Bertoniati, C.; Honig, B.; Alexov, E. Poisson-Boltzmann calculations of nonspecific salt effects on protein-protein binding free energies. *Biophys. J.*, 2007, 92, 1891-1899.
- [56] Schmidt, M.; Lopes, A.; Amara, N.; Bathelt, C.; Simonson, T. Testing the coulomb/accessible surface area solvent model for protein stability, ligand binding, and protein design. *BMC Bioinformatics*, 2008, 9, 148.
- [57] Schreiber, G.; Haran, G.; Zhou, H-X. Fundamental aspects of protein-protein association kinetics. *Chem. Rev.*, 2009, 109, 839-860.
- [58] Shaul, Y.; Schreiber, G. Exploring the charge space of protein-protein association: a proteomic study. *Proteins: Struct. Funct. Bioinf.*, 2005, 60, 341-352.
- [59] Alsallaq, R.; Zhou, H-X. Electrostatic rate enhancement and transient complex of protein-protein association. *Proteins: Struct. Funct. Bioinf.*, 2008, 71, 320-335.
- [60] Dong, F.; Zhou, H-X. Electrostatic contribution to the binding stability of protein-protein complexes. *Proteins: Struct. Funct. Bioinf.*, 2006, 65, 87-102.
- [61] Czodrowski, P.; Sottriffer, C. A.; Klebe, G. Protonation changes upon ligand binding to trypsin and thrombin. Structural interpretation based on pK_a calculations and ITC experiments. *J. Mol. Biol.*, 2007, 367, 1347-1356.
- [62] Djurdjevic-Pahl, A.; Hewage, C.; Malthouse, J. P. G. Ionisations within a subtilisin-glyoxal inhibitor complex. *Biochim. Biophys. Acta*, 2005, 1749, 33-41.
- [63] 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. Determining the molecular basis for the pH-dependent interaction between the link module of human TSG-6 and hyaluronan. *J. Biol. Chem.*, 2007, 282, 12976-12988.
- [64] Lu, Y.; Yang, C-Y.; Wang, S. Binding free energy contributions of interfacial waters in HIV-1 protease/inhibitor complexes. *J. Am. Chem. Soc.*, 2006, 128, 11830-11839.
- [65] Mason, A. C.; Jensen, J. H. Protein-protein binding is often associated with changes in protonation state. *Proteins: Struct. Funct. Bioinf.*, 2008, 71, 81-91.
- [66] Archontis, G.; Simonson, T. Proton binding to proteins. A free-energy component analysis using a dielectric continuum model. *Biophys. J.*, 2005, 88, 3888-3904.
- [67] Alexov, E. G. Calculating proton uptake/release and binding free energy taking into account ionization and conformation changes induced by protein-inhibitor association. Application to plasmepepsin, cathepsin D and endothiapepsin-pepstatin complexes. *Proteins: Struct. Funct. Bioinf.*, 2004, 56, 572-584.
- [68] Mitra, R.; Shyam, R.; Mitra, I.; Miteva, M.; Alexov, E. Calculating the protonation states of proteins and small molecules. Implications to ligand-receptor interactions. *Curr. Comput.-Aided Drug Des.*, 2008, 4, 169-179.
- [69] Polgar, T.; Keseru, G. M. Virtual screening for β -secretase (BACE1) inhibitors reveals the importance of protonation states at Asp32 and Asp228. *J. Med. Chem.*, 2005, 48, 3749-3755.
- [70] Polgar, T.; Magyar, C.; Simon, I.; Keseru, G. M. Impact of ligand protonation on virtual screening against β -secretase (BACE1). *J. Chem. Inf. Mod.*, 2007, 47, 2366-2373.
- [71] Jensen, J. H. Calculating pH and salt dependence of protein-protein binding. *Curr. Pharm. Biotechnol.*, 2008, 9, 96-102.
- [72] Kukić, P.; Nielsen, J. E. Electrostatics in proteins and protein-ligand complexes. *Future Med. Chem.*, 2010, 2, 647-666.

- [73] Li, H.; Robertson, A. D.; Jensen, J. H. Very fast empirical prediction and rationalization of protein pK_a values. *Proteins: Struct. Funct. Bioinf.*, 2005, *61*, 704-721.
- [74] Khandogin, J.; Brooks, C. L. Linking folding with aggregation in Alzheimer's β -amyloid peptides. *Proc. Natl. Acad. Sci. USA*, 2007, *104*, 16880-16885.
- [75] Chen, X.; Deng, Y. F. Long-time molecular dynamics simulations of botulinum toxin type-A at different pH values and temperatures. *J. Mol. Model.*, 2007, *13*, 559-572.
- [76] Khandogin, J.; Raleigh, D. P.; Brooks, C. L. Folding intermediate in the villin headpiece domain arises from disruption of a N-terminal hydrogen-bonded network. *J. Am. Chem. Soc.*, 2007, *129*, 3056-3057.
- [77] Srivastava, J.; Barber, D. L.; Jacobson, M. P. Intracellular pH sensors: design principles and functional significance. *Physiology*, 2007, *22*, 30-39.
- [78] Khandogin, J.; Brooks, C. L. Toward the accurate first-principles prediction of ionization equilibria in proteins. *Biochemistry*, 2006, *45*, 9363-9373.
- [79] Dlugosz, M.; Antosiewicz, J. M. Effects of solute-solvent proton exchange on polypeptide chain dynamics: A constant-pH molecular dynamics study. *J. Phys. Chem. B*, 2005, *109*, 13777-13784.
- [80] Mongan, J.; Case, D. A. Biomolecular simulations at constant pH. *Curr. Opin. Struct. Biol.*, 2005, *15*, 157-163.
- [81] Mitra, R. C.; Zhang, Z.; Alexov, E. In silico modeling of pH optimum of protein-protein binding. *Proteins: Struct. Funct. Bioinf.*, 2011, *79*, 925-936.
- [82] Yang, A.-S.; Honig, B. On the pH dependence of protein stability. *J. Mol. Biol.*, 1993, *231*, 459-474.
- [83] Whitten, S.; Garcia-Moreno, B. pH Dependence of stability of staphylococcal nuclease: Evidence of substantial electrostatic interactions in the denaturated state. *Biochemistry*, 2000, *39*, 14292-14304.
- [84] Tollinger, M.; Crowhurst, K.; Kay, L.; Forman-Kay, J. Site-specific contributions to the pH dependence of protein stability. *Proc. Natl. Acad. Sci. USA*, 2003, *100*, 4545-4550.
- [85] Bauman, A. T.; Jaron, S.; Yukl, E. T.; Burchfiel, J. R.; Blackburn, N. J. pH Dependence of peptidylglycine monooxygenase. Mechanistic implications of Cu-methionine binding dynamics. *Biochemistry*, 2006, *45*, 11140-11150.
- [86] Bidwai, A. K.; Ok, E. Y.; Erman, J. E. 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*, 2008, *47*, 10458-10470.
- [87] Matthew, J. B.; Gurd, F. R. N.; Garcia-Moreno, B.; Flanagan, M. A.; March, K. L.; Shire, S. J. pH-Dependent processes in proteins. *CRC Crit. Rev. Biochem.*, 1985, *18*, 91-197.
- [88] Anderson, D. E.; Becktel, W. J.; Dahlquist, F. W. pH-Induced denaturation of proteins: A single salt bridge contributes 3-5 kcal/mol to the free energy of folding of T4-lysozyme. *Biochem.*, 1990, *29*, 2403-2408.
- [89] Pace, C. N.; Laurents, D. V.; Erickson, R. E. Urea denaturation of barnase: pH dependence and characterization of the unfolded state. *Biochemistry*, 1992, *31*, 2728-2734.
- [90] Khurana, R.; Hate, A.; Nath, U.; Udgaonkar, B. pH dependence of the stability of barnase to chemical and thermal denaturation. *Prot. Sci.*, 1995, *4*, 1133-1144.
- [91] Chan, P.; Lovric, J.; Warwicker, J. Subcellular pH and predicted pH-dependent features of proteins. *Proteomics*, 2006, *6*, 3494-3501.
- [92] Chan, P.; Warwicker, J. Evidence for the adaptation of protein pH-dependence to subcellular pH. *BMC Biol.*, 2009, *7*, 69, 10 pp.
- [93] Gilson, M. K.; Honig, B. H. Calculation of electrostatic potentials in an enzyme active site. *Nature*, 1987, *330*, 84-86.
- [94] Bashford, D.; Karplus, M. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry*, 1990, *29*, 10219-10225.
- [95] Gilson, M. K. Multiple-site titration and molecular modeling: two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins: Struct. Funct. Bioinf.*, 1993, *15*, 266-282.
- [96] Yang, A.; Gunner, M. R.; Sampogna, R.; Sharp, K.; Honig, B. On the calculation of pK_a in proteins. *Proteins: Struct. Funct. Bioinf.*, 1993, *15*, 252-265.
- [97] Dixon, S. J.; Jurs, P. C. Estimation of pK_a for organic oxyacids using calculated atomic charges. *J. Comput. Chem.* 1993, *14*, 1460-1467.
- [98] Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. Prediction of pH-dependent properties of proteins. *J. Mol. Biol.*, 1994, *238*, 415-436.
- [99] Honig, B.; Nicholls, A. Classical electrostatics in biology and chemistry. *Science*, 1995, *268*, 1144-1149.
- [100] Antosiewicz, J.; Briggs, J. M.; Elcock, A. H.; Gilson, M. K.; McCammon, J. A. Computing ionization states of proteins with a detailed charged model. *J. Comput. Chem.*, 1996, *17*, 1633-1644.
- [101] Antosiewicz, J.; McCammon, J. A.; Gilson, M. K. The determinants of pK_as in proteins. *Biochemistry*, 1996, *35*, 7819-7833.
- [102] Sham, Y. Y.; Chu, Z. T.; Warshel, A. Consistent calculations of pK_a's of ionizable residues in proteins: Semi-microscopic and microscopic approaches. *J. Phys. Chem. B*, 1997, *101*, 4458-4472.
- [103] Duarte, H. A.; Carvalho, S.; Paniago, E. B.; Simas, A. M. Importance of tautomers in the chemical behavior of tetracyclines. *J. Pharm. Sci.*, 1999, *88*, 111-120.
- [104] Briggs, J. M.; Antosiewicz, J. Simulation of pH-dependent properties of proteins using mesoscopic models. *Rev. Comput. Chem.*, 1999, *13*, 249-311 and references therein.
- [105] Ullmann, G. M.; Knapp, E. W. Electrostatic models for computing protonation and redox equilibria in proteins. *Eur. Biophys. J.*, 1999, *28*, 533-551.
- [106] Crnogorac, M. M.; Ullmann, G. M.; Kostic, N. M. Effects of pH on protein association: modification of the proton-linkage model and experimental verification of the modified model in the case of cytochrome *c* and plastocyanin. *J. Am. Chem. Soc.*, 2001, *123* (44), 10789-10798.
- [107] Nielsen, J. E.; McCammon, J. A. On the evaluation and optimization of protein X-ray structures for pK_a calculations. *Protein Sci.*, 2003, *12*, 313-326.
- [108] Tannock, I. F.; Rotin, D. Acid pH in tumors and its potential for therapeutic exploitation. *Cancer Res.*, 1989, *49*, 4373-4384.
- [109] Ojugo, A. S.; McSheehy, P. M.; McIntyre, D. J.; McCoy, C.; Stubbs, M.; Leach, M. O.; Judson, I. R.; Griffiths, J. R. Measurement of the extracellular pH of solid tumours in mice by magnetic resonance spectroscopy: a comparison of exogenous ¹⁹F and ³¹P probes. *NMR Biomed.*, 1999, *12*, 495-504.
- [110] Said, H. M.; Mohammadkhani, R. Folate transport in intestinal brush border membrane: involvement of essential histidine residue(s). *Biochem. J.*, 1993, *290*, 237-240.
- [111] Fei, Y. J.; Liu, W.; Prasad, P. D.; Kekuda, R.; Oblak, T. G.; Ganapathy, V.; Leibach, F. H. Identification of the histidyl residue obligatory for the catalytic activity of the human H⁺/peptide cotransporters PEPT1 and PEPT2. *Biochemistry*, 1997, *36*, 452-460.
- [112] Metzner, L.; Natho, K.; Zebisch, K.; Dorn, M.; Bosse-Doenecke, E.; Ganapathy, V.; Brandsch, M. Mutational analysis of histidine residues in the human proton-coupled amino acid transporter PAT1. *Biochim. Biophys. Acta*, 2008, *1778*, 1042-1050.
- [113] Unal, E. S.; Zhao, R.; Chang, M. H.; Fiser, A.; Romero, M. F.; Goldman, I. D. The functional roles of the His 247 and His 281 residues in folate and proton translocation mediated by the human proton-coupled folate transporter SLC46A1. *J. Biol. Chem.*, 2009, *284*, 17846-17857.
- [114] Bartlett, P. A.; Hanson, J. E.; Giannousis, P. P. Potent Inhibition of Pepsin and Penicillopepsin by Phosphorus-Containing Peptide Analogues. *J. Org. Chem.*, 1990, *55*, 6268-6274.
- [115] Alexov, E. Numerical calculations of the pH of maximal protein stability. The effect of the sequence composition and 3D structure. *Eur. J. Biochem.*, 2004, *271*, 173-185.
- [116] Miranda, C.; Escartí, F.; Lamarque, L.; Yunta, M. J. R.; Navarro, P.; García-España, E.; Jimeno, M. L. New 1H-pyrazole-containing polyamine receptors able to complex L-glutamate in water at physiological pH values. *J. Am. Chem. Soc.*, 2004, *126*, 823-833.
- [117] 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. Progress in the prediction of pK_a values in proteins. *Proteins: Struct. Funct. Bioinf.*, 2011, *79*, 3260-3275.
- [118] Nielsen, J. E.; Gunner, M. R.; Garcia-Moreno, B. E. The pK_a cooperative: a collaborative effort to advance structure-based calculations of pK_a values and electrostatic effects in proteins. *Proteins: Struct. Funct. Bioinf.*, 2011, *79*, 3249-3259.

- [119] Rose, P. W.; Beran, B.; Bi, C.; Bluhm, W. F.; Dimitropoulos, D.; Goodsell, D. S.; Prlic, A.; Quesada, M.; Quinn, G. B.; Westbrook, J. D.; Young, J.; Yukich, B.; Zardecki, C.; Berman, H. M.; Bourne, P. E. The RCSB protein data bank: redesigned web site and web services. *Nucleic Acids Res.*, 2011, *39*, D392-401.
- [120] Janin, J.; Bahadur, R. P.; Chakrabarti, P. Protein-protein interaction and quaternary structure. *Q. Rev. Biophys.*, 2008, *41*, 133-80.
- [121] Ozbabacan, S. E.; Engin, H. B.; Gursoy, A.; Keskin, O. Transient protein-protein interactions. *Protein Eng. Des. Sel.*, 2011, *24*, 635-648.
- [122] Sheinerman, F. B.; Honig, B. On the role of electrostatic interactions in the design of protein-protein interfaces. *J. Mol. Biol.*, 2002, *318*, 161-77.
- [123] Lawrence, M. C.; Colman, P. M. Shape complementarity at protein/protein interfaces. *J. Mol. Biol.*, 1993; *234*, 946-50.
- [124] Vakser, I. A.; Aflalo, C. Hydrophobic docking: a proposed enhancement to molecular recognition techniques. *Proteins: Struct. Funct. Bioinf.*, 1994, *20*, 320-9.
- [125] McCoy, A. J.; Epa, V. C.; Colman, P. M. Electrostatic complementarity at protein/protein interfaces. *J. Mol. Biol.*, 1997, *268*, 570-584.
- [126] Vijayakumar, M.; Wong, K. Y.; Schreiber, G.; Fersht, A. R.; Szabo, A.; Zhou, H-X. Electrostatic enhancement of diffusion-controlled protein-protein association: comparison of theory and experiment on barnase and barstar. *J. Mol. Biol.*, 1998, *278*, 1015-24.
- [127] Qin, S.; Pang, X.; Zhou, H. X. Automated prediction of protein association rate constants. *Structure*, 2011, *19*, 1744-1751.
- [128] Keskin, O.; Gursoy, A.; Ma, B.; Nussinov, R. Principles of protein-protein interactions: What are the preferred ways for proteins to interact? *Chem. Rev.*, 2008, *108*, 1225-1244.
- [129] Prada-Gracia, D.; Gomez-Gardenes, J.; Echenique, P.; Falo, F. Exploring the free energy landscape: from dynamics to networks and back. *PLoS Comput. Biol.*, 2009, *5*, e1000415, 9 pp.
- [130] Jackson, R. M.; Sternberg, M. J. A continuum model for protein-protein interactions: application to the docking problem. *J. Mol. Biol.*, 1995, *250*, 258-75.
- [131] Frederick, K. K.; Marlow, M. S.; Valentine, K. G.; Wand, A. J. Conformational entropy in molecular recognition by proteins. *Nature*, 2007; *448*, 325-9.
- [132] Caffrey, D. R.; Somaroo, S.; Hughes, J. D.; Mintseris, J.; Huang, E. S. Are protein-protein interfaces more conserved in sequence than the rest of the protein surface? *Protein Science*, 2004, *13*, 190-202.
- [133] Meszaros, B.; Simon, I.; Dosztanyi, Z. The expanding view of protein-protein interactions: complexes involving intrinsically disordered proteins. *Phys. Biol.*, 2011, *8*, 035003, 10 pp.
- [134] Collins, K. D. Charge density-dependent strength of hydration and biological structure. *Biophys. J.*, 1997, *72*, 65-76.
- [135] Wadsö, I. Isothermal microcalorimetry for the characterization of interactions between drugs and biological materials. *Thermochim. Acta*, 1995, *267*, 45-59.
- [136] Jelesarov, I.; Bosshard, H. R. Isothermal titration calorimetry and differential scanning calorimetry as complementary tools to investigate the energetics of bimolecular recognition. *J. Molec. Recog.*, 1999, *12*, 3-18.
- [137] Fukada, H.; Takahashi, K. Enthalpy and heat capacity changes for the proton dissociation of various buffer components in 0.1 M potassium chloride. *Proteins: Struct. Funct. Bioinf.*, 1998, *33*, 159-166.
- [138] Raffaa, R. B.; Stagliano, G. W.; Spencera, S. D. Protonation effect on drug affinity. *Eur. J. Pharmacol.*, 2004, *483*, 323-324.
- [139] Raffa, R. B. (Ed.), *Drug-Receptor Thermodynamics: Introduction and Applications*. John Wiley and Sons, Chichester, 2001, pp. 1-781.
- [140] Lo Conte, L.; Chothia, C.; Janin, J. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.*, 1999, *285*, 2177-98.
- [141] Harms, M. J.; Castaneda, C. A.; Schlessman, J. L.; Sue, G. R.; Isom, D. G.; Cannon, B. R.; Garcia-Moreno, E. B. The pK_a values of acidic and basic residues buried at the same internal location in a protein are governed by different factors. *J. Mol. Biol.*, 2009, *389*, 34-47.
- [142] Gerstein, M.; Lesk, A. M.; Chothia, C. Structural mechanisms for domain movements in proteins. *Biochemistry*, 1994, *33*, 6739-49.
- [143] Wilson, I. A.; Stanfield, R. L. Antibody-antigen interactions: New structures and new conformational changes. *Curr. Opin. Struct. Biol.*, 1994, *4*, 857-67.
- [144] DuBay, K. H.; Geissler, P. L. Calculation of proteins' total side-chain torsional entropy and its influence on protein-ligand interactions. *J. Mol. Biol.*, 2009; *391*, 484-97.
- [145] Todorov, N. P.; Monthoux, P. H.; Alberts, I. L. The influence of variations of ligand protonation and tautomerism on protein-ligand recognition and binding energy landscape. *J. Chem. Inf. Mod.*, 2006, *46*, 1134-1142.
- [146] Knox, A. J. S.; Meegan, M. J.; Carta, G.; Lloyd, D. G. Considerations in compound database preparation: "Hidden" impact on virtual screening results. *J. Chem. Inf. Mod.*, 2005, *45*, 1908-1919.
- [147] Brooijmans N, Kuntz ID. Molecular recognition and docking algorithms. *Ann. Rev. Biophys. Biomol. Struct.*, 2003, *32*, 335-373.
- [148] Kundrotas, P. J.; Alexov, E. Electrostatic properties of protein-protein complexes. *Biophys. J.*, 2006, *91*, 1724-36.
- [149] Mitra, R. C.; Zhang, Z.; Alexov, E. In silico modeling of pH-optimism of protein-protein binding. *Proteins: Struct. Funct. Bioinf.*, 2011, *79*, 925-936.
- [150] Gilson, M. K.; Zhou, H. X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.*, 2007, *36*, 21-42.
- [151] Witham, S.; Talley, K.; Wang, L.; Zhang, Z.; Sarkar, S.; Gao, D.; Yang, W.; Alexov, E. Developing hybrid approaches to predict pK_a values of ionizable groups. *Proteins: Struct. Funct. Bioinf.*, 2011, *79*, 3389-3399.
- [152] Jones, G.; Willett, P.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, 1997, *267*, 727-748.
- [153] Bayden, A. S.; Fornabaio, M.; Scarsdale, J. N.; Kellogg, G. E. Web application for studying the free energy of binding and protonation states of protein-ligand complexes based on HINT. *J. Comput. Aided Mol. Des.*, 2009, *23*, 621-32.
- [154] Rapp, C. S.; Schonbrun, C.; Jacobson, M. P.; Kalyanaraman, C.; Huang, N. Automated site preparation in physics-based rescoring of receptor ligand complexes. *Proteins: Struct. Funct. Bioinf.*, 2009, *77*, 52-61.
- [155] Kalliokoski, T.; Salo, H. S.; Lahtela-Kakkonen, M.; Poso, A. The effect of ligand-based tautomer and protomer prediction on structure-based virtual screening. *J. Chem. Inf. Mod.*, 2009, *49*, 2742-2748.
- [156] Fornabaio, M.; Cozzini, P.; Mozzarelli, A.; Abraham, D. J.; Kellogg, G. E. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 2. Computational titration and pH effects in molecular models of neuraminidase-inhibitor complexes. *J. Med. Chem.*, 2003, *46*, 4487-4500.
- [157] Cozzini, P.; Fornabaio, M.; Marabotti, A.; Abraham, D. J.; Kellogg, G. E.; Mozzarelli, A. Simple, intuitive calculations of free energy of binding for protein-ligand complexes. 1. Models without explicit constrained water. *J. Med. Chem.*, 2002; *45*, 2469-2483.
- [158] Pospisil, P.; Ballmer, P.; Scapozza, L.; Folkers, G. Tautomerism in computer-aided drug design. *J. Recept. Signal Transduct. Res.*, 2003, *23*, 361-371.
- [159] Kirchmair, J.; Markt, P.; Distinto, S.; Wolber, G.; Langer, T. Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection--what can we learn from earlier mistakes? *J. Comput. Aided Mol. Des.*, 2008, *22*, 213-28.
- [160] Ferrara, P.; Gohlke, H.; Price, D.J.; Klebe, G.; Brooks C. L. III. Assessing scoring functions for protein-ligand interactions. *J. Med. Chem.*, 2004, *47*, 3032-3047.
- [161] Zou, X.; Sun, Y.; Kuntz, I. D. Inclusion of solvation in ligand binding free energy calculations using the generalized-Born model. *J. Am. Chem. Soc.*, 1999, *121*, 8033-8043.
- [162] Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative evaluation of eight docking tools for docking and virtual screening accuracy. *Proteins: Struct. Funct. Bioinf.*, 2004, *57*, 225-242.
- [163] Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins: Struct. Funct. Bioinf.*, 2004, *56*, 235-249.
- [164] Krovat, E. M.; Steindl, T.; Langer, T. Recent advances in docking and scoring. *Curr. Comput.-Aided Drug Des.*, 2005, *1*, 93-102.
- [165] Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein small molecule docking methods. *J. Comput.-Aided Mol. Des.*, 2002, *16*, 151-166.
- [166] Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, Th, E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.*, 1982, *161*, 269-288.

- [167] DesJarlais, R. L.; Sheridan, R. P.; Seibel, G. L.; Dixon, J. S.; Kuntz, I. D.; Venkataraghavan, R. Using shape complementarity as an initial screen in designing ligands for a receptor binding site of known three-dimensional structure. *J. Med. Chem.*, 1988, *31*, 722-729.
- [168] Gschwend, D. A.; Kuntz, I. D. Orientational sampling and rigid-body minimization in molecular docking revisited: On-the-fly optimization and degeneracy removal. *J. Comput.-Aided Mol. Des.*, 1996, *10*, 123-132.
- [169] Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: Search strategies for automated molecular docking of flexible molecule databases. *J. Comput.-Aided Mol. Des.*, 2001, *15*, 411-428.
- [170] Rarey, M.; Wefing, S.; Lengauer, T. Placement of medium-sized molecular fragments into active sites of proteins. *J. Comput.-Aided Mol. Des.*, 1996, *10*, 41-54.
- [171] Zawodszky, M. I.; Sanschagrin, P. C.; Korde, R. S.; Kuhn, L. A. Distilling the essential features of a protein surface for improving protein-ligand docking, scoring, and virtual screening. *J. Comput.-Aided Mol. Des.*, 2002, *16*, 883-902.
- [172] Jain, A. N. Surflex: Fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem.*, 2003, *46*, 499-511.
- [173] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, 2004, *47*, 1739-1749.
- [174] Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.*, 2004, *47*, 1750-1759.
- [175] Friesner, R. A.; Murphy, R. B.; Repasky, M. P.; Frye, L. L.; Greenwood, J. R.; Halgren, T. A.; Sanschagrin, P. C.; Mainz, D. T. Extra precision Glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.*, 2006, *49*, 6177-6196.
- [176] Verdonk, M. L.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins: Struct. Funct. Bioinf.*, 2003, *52*, 609-623.
- [177] Jones, G.; Willett, P.; Leach, A. R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, 1997, *267*, 727-748.
- [178] Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using a genetic algorithm with a description of desolvation. *J. Mol. Biol.*, 1995, *245*, 43-53.
- [179] Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins: Struct. Funct. Bioinf.*, 1990, *8*, 195-202.
- [180] Morris, G. M.; Goodsell, D. S.; Halliday, R. S.; Huey, R.; Hart, W. E.; Belew, R. K.; Olson, A. J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, 1998, *19*, 1639-1662.
- [181] Korb, O.; Stützel, T.; Exner, T. E. PLANTS: Application of ant colony optimization to structure-based drug design. In *Ant Colony Optimization and Swarm Intelligence, 5th International Workshop, ANTS 2006, LNCS 4150*; Dorigo, M.; Gambardella, L. M.; Birattari, M.; Martinoli, A.; Poli, R.; Stützel, T., Eds. pp 247-258.
- [182] Korb, O.; Stützel, T.; Exner, T. E. An ant colony optimization approach to flexible protein-ligand docking. *Swarm Intell.*, 2007, *2*, 115-134.
- [183] Korb, O.; Stützel, T.; Exner, T. E. Empirical scoring functions for advanced protein-ligand docking with PLANTS. *J. Chem. Inf. Model.*, 2009, *49*, 84-96.
- [184] Totrov, M.; Abagyan, R. Flexible protein ligand docking by global energy optimization in internal coordinates. *Proteins: Struct. Funct. Bioinf.*, 1997, *1*, 215-220.
- [185] McMartin, C.; Bohacek, R. S. QXP: Powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.*, 1997, *11*, 333-344.
- [186] Westhead, D. R.; Clark, D. E.; Murray, C. W. A comparison of heuristic search algorithms for molecular docking. *J. Comput.-Aided Mol. Des.*, 1997, *11*, 209-228.
- [187] Baxter, C. A.; Murray, C. W.; Clark, D. E.; Westhead, D. R.; Eldridge, M. D. Flexible docking using tabu search and an empirical estimate of binding affinity. *Proteins: Struct. Funct. Bioinf.*, 1998, *33*, 367-382.
- [188] McGann, M.; Almond, H.; Nicholls, A.; Grant, J. A.; Brown, F. Gaussian docking functions. *Biopolymers*, 2003, *68*, 76-90.
- [189] Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: Comparative data on docking algorithms. *J. Med. Chem.*, 2004, *47*, 558-565.
- [190] Venkatachalam, C. M.; Jiang, X.; Oldfield, T.; Waldman, M. Ligand Fit: A novel method for the shape-directed rapid docking of ligands to protein active sites. *J. Mol. Graphics Model.*, 2003, *21*, 289-307.
- [191] Kaul, M.; Pilch, D. S. Thermodynamics of aminoglycoside-rRNA recognition: the binding of neomycin-class aminoglycosides to the A site of 16 S rRNA. *Biochemistry*, 2002, *41*, 7695-7706.
- [192] Kaul, M.; Barbieri, C. M.; Kerrigan, J. E.; Pilch, D. S. Coupling of drug protonation to the specific binding of aminoglycosides to the A site of 16 S rRNA: Elucidation of the number of drug amino groups involved and their identities. *J. Mol. Biol.*, 2003, *326*, 1373-1387.
- [193] Schüürmann, G.; Cossi, M.; Barone, V.; Tomasi, J. Prediction of the pK_a of carboxylic acids using the ab initio continuum-solvation model PCM-UAHF. *J. Phys. Chem. A*, 1998, *102*, 6706-6712.
- [194] Toth, A. M.; Liptak, M. D.; Phillips, D. L.; Shields, G. C. Accurate relative pK_a calculations for carboxylic acids using complete basis set and Gaussian-n models combined with continuum solvation methods. *J. Chem. Phys.*, 2001, *114*, 4595-4606.
- [195] Topol, I. A.; Tawa, G. J.; Burt, S. K.; Rashin, A. A. Calculation of absolute and relative acidities of substituted imidazoles in aqueous solvent. *J. Phys. Chem. A*, 1997, *101*, 10075-10081.
- [196] da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. Ab initio calculations of absolute pK_a values in aqueous solution, I. Carboxylic acids. *J. Phys. Chem. A*, 1999, *103*, 11194-11199.
- [197] Topol, I. A.; Tawa, G. J.; Caldwell, R. A.; Eissenstat, M. A.; Burt, S. K. Acidity of organic molecules in the gas phase and in aqueous solvent. *J. Phys. Chem. A*, 2000, *104*, 9619-9624.
- [198] da Silva, C. O.; da Silva, E. C.; Nascimento, M. A. C. Ab initio calculations of absolute pK_a values in aqueous solution, II. Aliphatic alcohols, thiols, and halogenated carboxylic acids. *J. Phys. Chem. A*, 2000, *104*, 2402-2409.
- [199] Liptak, M. D.; Shields, G. C. Accurate pK_a calculations for carboxylic acids using complete basis set and Gaussian-n models combined with CPCM continuum solvation methods. *J. Am. Chem. Soc.*, 2001, *123*, 7314-7319.
- [200] Jang, Y. H.; Sowers, L. C.; Cagin, T.; Goddard, W. A., III. First principles calculation of pK_a values for 5-substituted uracils. *J. Phys. Chem. A*, 2001, *105*, 274-280.
- [201] La Francois, C. J.; Jang, Y. H.; Cagin, T.; Goddard, W. A., III; Sowers, L. C. Conformation and proton configuration of pyrimidine deoxynucleoside oxidation damage products in water. *Chem. Res. Toxicol.*, 2000, *13*, 462-470.
- [202] Tomasi, J.; Persico, M. Molecular interactions in solution: An overview of methods based on continuous distributions of the solvent. *Chem. Rev.*, 1994, *94*, 2027-2094.
- [203] Cramer, C. J.; Truhlar, D. G. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.*, 1999, *99*, 2161-2200.
- [204] Orozco, M.; Luque, F. J. Theoretical methods for the description of the solvent effect in biomolecular systems. *Chem. Rev.*, 2000, *100*, 4187-4225.
- [205] Tomasi, J. Thirty years of continuum solvation chemistry: a review, and prospects for the near future. *Theor. Chem. Acc.*, 2004, *112*, 184-203.
- [206] Tomasi, J. Quantum mechanical continuum solvation models. *Chem. Rev.*, 2005, *105*, 2999-3094.
- [207] Cramer, C. J.; Truhlar, D. G. A universal approach to solvation modeling. *Acc. Chem. Res.*, 2008, *41*, 760-768.
- [208] Ho, J.; Coote, M. L. A universal approach for continuum solvent pK_a calculations: are we there yet? *Theor. Chem. Acc.*, 2010, *125*, 3-21.
- [209] Czodrowski, P. Prediction of protonation states in ligand-protein complexes upon ligand binding Dissertation zur Erlangung des Doktorgrades der Naturwissenschaften (Dr. rer. nat.) dem Fachbereich Pharmazie der Philipps-Universität Marburg Marburg/Lahn, 2006.
- [210] Ho, J.; Coote, M. L. pK_a calculation of some biologically important carbon acids-An assessment of contemporary theoretical procedures. *J. Chem. Theory Comput.*, 2009, *5*, 295-306.
- [211] Casanovas, R.; Frau, J.; Ortega-Castro, J.; Salvà, A.; Donoso, J.; Muñoz, F. Absolute and relative pK_a calculations of mono and

- diprotic pyridines by quantum methods. *J. Mol. Struct. Theochem.*, 2009, 912, 5-12.
- [212] Pliego, J. R. Jr; Riveros, J. M. Theoretical Calculation of pK_a using the cluster-continuum model. *J. Phys. Chem. A*, 2002, 106, 7434-7439.
- [213] Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. SM6: A density functional theory continuum solvation model for calculating aqueous solvation free energies of neutrals, ions, and solute-water clusters. *J. Chem. Theory Comput.*, 2005, 1, 1133-1152.
- [214] Kelly, C. P.; Cramer, C. J.; Truhlar, D. G. Adding explicit solvent molecules to continuum solvent calculations for the calculation of aqueous acid dissociation constants. *J. Phys. Chem. A*, 2006, 110, 2493-2499.
- [215] Sánchez-Marcos, E.; Terryn, B.; Rivail, J. L. Protonation of nitrogen-containing bases in solution: continuum vs. discrete-continuum models for aqueous solutions. *J. Phys. Chem.*, 1985, 89, 4695-4700.
- [216] Claverie, P.; Daudey, J. P.; Langlet, J.; Pullman, B.; Plazzola, D.; Huron, M. J. Studies of solvent effects. I. Discrete, continuum, and discrete-continuum models and their comparison for some simple cases: ammonium (1+) ion, methanol, and substituted ammonium (1+) ion. *J. Phys. Chem.*, 1978, 82, 405-418.
- [217] Vyacheslav, S.; Bryantsev, S.; Diallo, M. S.; Goddard, W. A. III. Calculation of solvation free energies of charged solutes using mixed cluster/continuum models. *J. Phys. Chem. B*, 2008, 112, 9709-9719.
- [218] Pliego, J. R. Jr; Riveros, J. M. The cluster-continuum model for the calculation of the solvation free energy of ionic species. *J. Phys. Chem. A*, 2001, 105, 7241-7247.
- [219] Scherrer, R. A.; Leo, A.J. Multi-pH QSAR: A method to differentiate the activity of neutral and ionized species and obtain true correlations when both species are involved. *Mol. Inf.*, 2010, 29, 687-693.
- [220] Scherrer, R. A.; Donovan, S. F. Automated potentiometric titrations in KCl/water-saturated octanol: Method for quantifying factors influencing ion-pair partitioning. *Anal. Chem.*, 2009, 81, 2768-2778.
- [221] Caron, G.; Ermondi, G.; Scherrer, R. A. Lipophilicity, polarity, and hydrophobicity, in comprehensive medicinal chemistry II, (Eds: B. Testa, H. van de Waterbeemd), Elsevier, Oxford, 2007, ch. 5.18, pp. 425-452.
- [222] Scherrer, R. A.; Donovan, S. F. Species-specific log D in drug design and ADMET analyses. Presented at the 29th National Medicinal Chemistry Symposium, Madison, WI, June 2004, 70 (a copy is available on request).
- [223] Parajuli, R.; Medhi, C. Basicities of some 9-substituted acridine-4-carboxamides: A density functional theory (DFT) calculation. *J. Chem. Sci.*, 2004, 116, 235-241.
- [224] Lehninger, A. L.; Nelson, D. L.; Cox, M. M. *Principles of biochemistry*, 2nd ed.; Worth Publishers: New York, 1993.
- [225] Steed, J.; Atwood, J. *Supramolecular chemistry*; John Wiley and Sons, LTD: Chichester, 2000.
- [226] Schneider, H.-J.; Yatsimirsky, A. *Principles and methods in supramolecular chemistry*; John Wiley and Sons, LTD: Chichester, 2000.
- [227] Fitzmaurice, R. J.; Kyne, G. M.; Douheret, D.; Kilburn, J. D. Synthetic receptors for carboxylic acids and carboxylates. *J. Chem. Soc., Perkin Trans.*, 2002, 1, 841-864.
- [228] Gale, P. A. Anion coordination and anion-directed assembly: highlights from 1997 and 1998. *Coord. Chem. Rev.*, 2000, 199, 181-233.
- [229] Gale, P. A. Anion receptor chemistry: highlights from 1999. *Coord. Chem. Rev.*, 2001, 213, 79-128.
- [230] Sigel, H.; Martin, R. B. Coordinating properties of the amide bond. Stability and structure of metal ion complexes of peptides and related ligands. *Chem. Rev.*, 1982, 82, 385-426.
- [231] Clement, O.; Rapko, B. M.; Hay, B. P. Structural aspects of metal-amide complexes. *Coord. Chem. Rev.*, 1998, 170, 203-243.
- [232] Paz, P. B.; Vega-Hissi, E. G.; Estrada, M. R.; Garro-Martinez, J. C. *In silico* modeling of the molecular structure and binding of leukotriene A4 into leukotriene A4 hydrolase. *Chem. Biol. Drug Des.*, 2012, 80, 902-908.
- [233] Thunnissen, M. M. G. M.; Nordlund, P.; Haeggstrom, J. Z. Crystal structure of human leukotriene A4 hydrolase, a bifunctional enzyme in inflammation. *Nat. Struct. Biol.*, 2001, 8, 131-135.
- [234] Haeggstrom, J. Z. Leukotriene A4 hydrolase/aminopeptidase, the gatekeeper of chemotactic leukotriene B4 biosynthesis. *J. Biol. Chem.*, 2004, 279, 50639-50642.
- [235] Kirkland, T. A.; Adler, M.; Bauman, J. G.; Chen, M.; Haeggstrom, J. Z.; King, B.; Kochanny, M. J.; Liang, A. M.; Mendoza, L.; Phillips, G. B.; Thunnissen, M.; Trinh, L.; Whitlow, M.; Ye, B.; Ye, H.; Parkinson, J.; Guilford, W. J. Synthesis of glutamic acid analogs as potent inhibitors of leukotriene A4 hydrolase. *Bioorgan. Med. Chem.*, 2008, 16, 4963-4983.
- [236] Thunnissen, M. M. G. M.; Andersson, B.; Samuelsson, B.; Wong, C. H.; Haeggstrom, J. Z. Crystal structures of leukotriene A4 hydrolase in complex with captopril and two competitive tight-binding inhibitors. *FASEB J.*, 2002, 16, 1648-1650.
- [237] Gavernet, L.; Palestro, P. H.; Bruno-Blanch, L. Docking applied to the study of inhibitors of c-met kinase. *ISRN Phys. Chem.*, 2012, Article ID 391897, 5 pages.
- [238] Patronov, A.; Dimitrov, I.; Flower, D. R.; Doytchinova, I. Peptide binding to HLA-DP proteins at pH 5.0 and pH 7.0: a quantitative molecular docking study. *BMC Struct. Biol.*, 2012, 12, 20-34.
- [239] 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, E. C.; Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem.*, 2006, 49, 5912-5931.
- [240] ten Brink, T.; Exner, T. E. A combined combinatorial and pK_a-based approach to ligand protonation states. *J. Cheminf.*, 2010, 2 (Suppl 1), P43.
- [241] Aubard, J.; Schwaller, M. A.; Pantigny, J.; Marsault, J. P.; Levi, G. Surface-enhanced Raman spectroscopy of ellipticine, 2-N-methyllellipticinium and their complexes with DNAJ. *Raman. Spectrosc.*, 1992, 23, 373-377.
- [242] Warshel, A. Electrostatic basis of structure-function correlation in proteins. *Accounts Chem. Res.*, 1981, 14, 284-290.
- [243] Demchuk, E.; Wade, R. C. Improving the continuum dielectric approach to calculating pK_as of ionizable groups in proteins. *J. Phys. Chem.*, 1996, 100, 17373-17387.
- [244] Nielsen, J. E.; Vriend, G. Optimizing the hydrogen-bond network in Poisson-Boltzmann equation-based pK_a calculations. *Proteins: Struct. Funct. Bioinf.*, 2001, 43, 403-412.
- [245] Karshikoff, A. A simple algorithm for the calculation of multiplesite titration curves. *Protein Eng.*, 1995, 8, 243-248.
- [246] Mehler, E. L.; Guarnieri, F. A self-consistent, microenvironment modulated screened Coulomb potential approximation to calculate pH-dependent electrostatic effects in proteins. *Biophys. J.*, 1999, 77, 3-22.
- [247] Wisz, M. S.; Hellinga, H. W. An empirical model for electrostatic interactions in proteins incorporating multiple geometry-dependent dielectric constants. *Proteins: Struct. Funct. Bioinf.*, 2003, 51, 360-377.
- [248] Havranek, J. J.; Harbury, P. B. Tanford-Kirkwood electrostatics for protein modeling. *Proc. Natl. Acad. Sci. USA*, 1999, 96, 11145-11150.
- [249] Nielsen, J. Calculating pK_a values in enzyme active sites. *Protein Sci.*, 2003, 12, 1894-1901.
- [250] Barth, P.; Alber, T.; Harbury, P. B. Accurate, conformation-dependent predictions of solvent effects on protein ionization constants. *Proc. Natl. Acad. Sci. USA*, 2007, 104, 4898-4903.
- [251] Godoy-Ruiz, R.; Perez-Jimenez, R.; Garcia-Mira, M. M.; del Pino, I. M. P.; Sanchez-Ruiz, J. M. Empirical parametrization of pK values for carboxylic acids in proteins using a genetic algorithm. *Biophys. Chem.*, 2005, 115, 263-266.
- [252] He, Y.; Xu, J.; Pan, X. M. A statistical approach to the prediction of pK_a values in proteins. *Proteins: Struct. Funct. Bioinf.*, 2007, 69, 75-82.
- [253] Simonson, T.; Carlsson, J.; Case, D. A. Proton binding to proteins: pK_a calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.*, 2004, 126, 4167-4180.
- [254] Tynan-Connolly, B. M.; Nielsen, J. E. Redesigning protein pK_a values. *Protein Sci.*, 2007, 16, 239-249.
- [255] Del Buono, G. S.; Figueirido, F. E.; Levy, R. M. Intrinsic pK_as of ionizable residues in proteins: an explicit solvent calculation for lysozyme. *Proteins: Struct. Funct. Genet.*, 1994, 20, 85-97.
- [256] Warshel, A.; Sussman, F.; King, G. Free energy of charges in solvated proteins: microscopic calculations using a reversible charging process. *Biochemistry*, 1986, 25, 8368-8372.
- [257] <http://www.opensource.org/licenses/BSD-2-Clause> (accessed 2014/9/3).
- [258] <http://propka.ki.ku.dk/>. (accessed 2014/9/3).
- [259] Warr, W. Tautomerism in chemical information management systems. *J. Comput.-Aided Mol. Des.*, 2010, 24, 497-520.

- [260] Brenk, R.; Irwin, J. J.; Shoichet, B. K. Here be dragons: docking and screening in an uncharted region of chemical space. *J. Biomol. Screen*, 2005, 10, 667-674.
- [261] Kellogg, G. E.; Fornabaio, M.; Spyraakis, F.; Lodola, A.; Cozzini, P.; Mozzarelli, A.; Abraham, D. J. Getting it right: modeling of pH, solvent and "nearly" everything else in virtual screening of biological targets. *J. Mol. Graph. Mod.*, 2004, 22, 479-486.
- [262] Bountis T. (Ed.), Proton transfer in hydrogen-bonded systems, NATO ASI Series B: Physics, vol. 291, Plenum Press, New York, 1992, pp. 1-355.
- [263] Kollman, P. A. Free energy calculations: applications to chemical and biochemical phenomena, *Chem. Rev.*, 1993, 93, 2395-2417.
- [264] Reddy, M. R.; Erion, M. D.; Agarwal, A. Free energy calculations: use and limitations in predicting binding affinities, *Rev. Comput. Chem.*, 2000, 16, 217-304.
- [265] Straatsma, T. P. Free energy by molecular simulation, *Rev. Comput. Chem.*, 1996, 9, 81-127.
- [266] Ota, N.; Stroupe, C.; Ferreira da Silva, J. M.; Shah, S. A.; Mares-Guia, M.; Brunger, A. T. Non-Boltzmann thermodynamic integration (NBTI) for macromolecular systems: Relative free energy of binding of trypsin to benzamidine and benxylamine, *Proteins: Struct. Funct. Genet.*, 1999, 37, 641-653.
- [267] Åqvist, J.; Luzhkov, V. B.; Brandsal, B. O. Ligand binding affinities from MD simulations, *Acc. Chem. Res.*, 2002, 35, 358-365.
- [268] Jorgensen, W. L. Free energy calculations: a breakthrough for modeling organic chemistry in solution, *Acc. Chem. Res.*, 1989, 22, 184-189.
- [269] Grant, J.A.; Pickup, B.T.; Nicholls, A. A smooth permittivity function for Poisson-Boltzmann solvation methods, *J. Comput. Chem.*, 2001, 22, 608-640.
- [270] Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham III, T.E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.*, 2000, 33, 889-897.
- [271] Fogolari, F.; Brigo, A.; Molinari, H. Protocol for MM/PBSA molecular dynamics simulations of proteins, *Biophys. J.*, 2003, 85, 159-166.
- [272] Gohlke, H.; Kiel, C.; Case, D.A. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RaIGDS complexes, *J. Mol. Biol.*, 2003, 330, 891-913.
- [273] Böhm, H.-J.; Stahl, M. Rapid empirical scoring functions in virtual screening applications, *Med. Chem. Res.*, 1999, 9, 445-462.