

Web application for studying the free energy of binding and protonation states of protein–ligand complexes based on HINT

Alexander S. Bayden · Micaela Fornabaio ·
J. Neel Scarsdale · Glen E. Kellogg

Received: 13 May 2008 / Accepted: 22 January 2009 / Published online: 25 June 2009
© Springer Science+Business Media B.V. 2009

Abstract A public web server performing computational titration at the active site in a protein–ligand complex has been implemented. This calculation is based on the Hydropathic interaction noncovalent force field. From 3D coordinate data for the protein, ligand and bridging waters (if available), the server predicts the best combination of protonation states for each ionizable residue and/or ligand functional group as well as the Gibbs free energy of binding for the ionization-optimized protein–ligand complex. The 3D structure for the modified molecules is available as output. In addition, a graph depicting how this energy changes with acidity, i.e., as a function of added protons, can be obtained. This data may prove to be of use in preparing models for virtual screening and molecular docking. A few illustrative examples are presented. In β secretase (2va7) computational titration flipped the amide groups of Gln12 and Asn37 and protonated a ligand amine yielding an improvement of 6.37 kcal mol⁻¹ in the protein–ligand binding score. Protonation of Glu139 in mutant HIV-1 reverse transcriptase (2opq) allows a water bridge between the protein and inhibitor that increases the protein–ligand interaction score by 0.16 kcal mol⁻¹. In human sialidase NEU2 complexed with an isobutyl ether mimetic inhibitor (2f11) computational titration suggested that protonating Glu218, deprotonating Arg237, flipping the

amide bond on Tyr334, and optimizing the positions of several other polar protons would increase the protein–ligand interaction score by 0.71 kcal mol⁻¹.

Keywords Crystallography · Computational titration · Web application · Gibbs free energy · Protonation · Proteins · HINT

Introduction

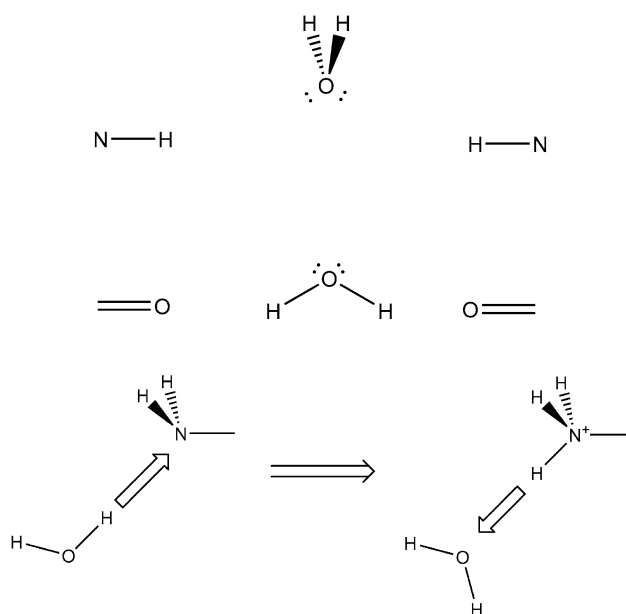
Even if one has an “atomic resolution” crystal structure of a protein–ligand complex, quantitatively modeling the Gibbs free energy of binding for the ligand can be challenging. There are several reasons for this difficulty, many of which are associated with interpretation of the crystal structure data. For example, the asparagine, glutamine and histidine functional groups may be rotated incorrectly, because in a typical (1.8 Å resolution or poorer) crystal structure hydrogen atoms are not visible and it is hard, if not impossible, to distinguish between N and O of the amide group (Asn or Gln) and C and N atoms in the His imidazole ring unless the interactions these functional groups make with neighboring residues are carefully considered. In fact, Weichenberger et al. [1] argue that the average rate of Asp and Gln rotation errors found in the current Protein Data Bank (PDB) is as high as 20% [2–6].

An even greater challenge is the protonation state of the system. As noted above, most crystal structures do not contain information on the positions of hydrogen atoms. This means that for some (ionizable) groups, both on the protein and potentially on the ligand, it is hard to say which protons are present, which ones are absent, and if they are present, to define their orientation. These groups are by no means isolated from each other, but influence each other's

A. S. Bayden · M. Fornabaio · G. E. Kellogg (✉)
Department of Medicinal Chemistry & Institute for Structural
Biology and Drug Discovery, Virginia Commonwealth
University, Richmond, VA 23298-0540, USA
e-mail: glen.kellogg@vcu.edu

J. N. Scarsdale
Department of Biochemistry and Molecular Biology & Institute
for Structural Biology and Drug Discovery, Virginia
Commonwealth University, Richmond, VA 23298-0614, USA

states, such that their geometries and protonation states can not be evaluated independently of each other. The number of model possibilities in this ensemble grows exponentially with the number of ionizable groups in the active site. An active site can have several protonation states that exist at equilibrium with each other that produce many energetically accessible models. Water molecules present in the active site further complicate the problem because hydrogen atoms on them can point in various directions. Water can mediate hydrogen bonds by acting both as a Lewis base and/or as a Lewis acid (below, left) to convert a weakly repulsive polar interaction into a strongly favorable interaction. Thus, a water molecule can “buffer” the active site by rotating and changing its character from a donor to an acceptor when an interacting functional group is protonated or deprotonated (below, right).



Scoring Functions for Protein–Ligand Associations. For a protein ligand complex where the geometry and the ionization state are known, calculating the Gibbs free energy is still non-trivial [7–11]. This is commonly referred to as the “scoring function problem” and is the subject of intense research in computational chemistry. Simply put, when most modeling packages report energy, they are reporting enthalpy, not Gibbs free energy. One conventional approach to predicting the Gibbs free energy of protein–ligand binding is using simplistic scoring functions calibrated against crystal structures of protein–ligand complexes. These scoring systems are obtained by considering a set of protein–ligand crystal structures for which the experimental dissociation constant is known. protein–ligand interactions are classified and counted for each structure and other surface and flexibility-related properties can also be determined for

each case. By assigning relative contributions to these interaction and energy components, the sum of effects for each complex results in a free energy “score” that should correlate with the energetics of protein–ligand binding as encoded in the dissociation constant for the complex. Examples of such scoring systems are SCORE1 [12], SCORE2 [13] and ChemScore [14]. There are obviously a large number of assumptions inherent in this approach, including additivity of contributions [15–17], the radically different experimental conditions between a low-temperature crystallographic experiment and room temperature solutions where association and dissociation measurements are made [18]. But also, these scoring functions are based on fairly small sets of data, usually on the order of a hundred protein–ligand complexes, so they can be thrown off by interactions that occur in the training set, but are rare in the real world, or vice versa. Most importantly, these scoring systems often are quite poor for compounds that are very different from those in the training set [11]. Knowledge-based score systems, such as DrugScore [19], focus instead on frequency of interaction types in known crystal structures using the assumption that the more favorable the interaction is, the more frequently it will appear. Sophisticated, and more computationally expensive, ways to find the free energy of binding in protein–ligand complexes include the free-energy-perturbation (FEP) [20, 21] and linear response [22, 23] methods that rely on statistical mechanics analysis of molecular dynamics or Monte Carlo simulations. These methods can be somewhat compromised by errors from a variety of sources [24].

Placing Protons and Optimizing Ionization States. The problem of placing and optimizing polar protons has been of interest for some time and a variety of methods have been used to correct ambiguous atom placement and assign protonation states and hydrogen orientations to ionizable residues in proteins. First, in terms of correcting X-ray protein crystal structures for problems with Asn, Gln, His, etc., there are several web applications available [25–27]. For example, MolProbity [26] can be used both for the placement of hydrogen atoms and correction of errors in protein structures, while others, like NQ-Flipper [25], ignore hydrogens altogether. Solutions to the more complex problems associated with ionization state evaluation have previously not been made available in a web application, but are available in a number of diverse tools. Quantum mechanics [28] and quantum mechanics–molecular mechanics [29, 30] approaches and methods based on molecular dynamics (MD) simulations [31–33] are available. However, most methods rely on solving the Poisson–Boltzmann equation, e.g., DelPhi [34], to evaluate possible protonation states and hydrogen positions [35–41]. More recently, solving the Poisson–Boltzmann equation has been combined with MD simulation [42–46]. Alternatively, Mehler and Guarnieri

quantitatively characterize the hydrophilicity or hydrophobicity of the microenvironment around each titratable group instead of obtaining grid-based solutions to the Poisson–Boltzmann equation [47]. However, these methods pursue a different goal than the application described in this paper in determining the protonation state and hydrogen geometry most commonly encountered for a protein, such that the results are most often described as pK_a s for protein residues. Also, these approaches do not generally focus on the key residues/functional groups leading to ligand binding. Our web application is pursuing that goal—finding the protonation state and hydrogen geometry for which the protein–ligand interaction is the strongest.

HINT. While clearly an empirical scoring function, with many of the inherent limitations described above, the HINT (Hydropathic INTeractions) forcefield and scoring tool has a number of advantages over other algorithms, particularly with respect to accounting for hydrophobic and desolvation contributions to binding energy that are somewhat entropic in nature and otherwise difficult to estimate. The HINT model calculates the free energies of association for non-covalently bound organic and biological molecules through a summation of hydrophobic atom constants that are derived from a large set of experimental partition coefficients of small molecules between 1-octanol and water phases. Because these partition coefficients are experimentally derived thermodynamic quantities, they directly correlate with the free energies of interaction and contain information about all non-covalent interactions that are likely to be encountered in biological environments. Solvation/desolvation energies are inherently included in the parameterization of HINT, because molecules moving between 1-octanol and water are at the heart of the partition coefficient, which is the free energy for solvent transfer. HINT determines the Gibbs free energy of non-covalent interactions from partial contributions of individual atoms to the partition coefficient and the distance between these atoms. The interaction between two molecules is determined using the equation:

$$H_{\text{TOT}} = \sum \sum b_{ij} = \sum \sum (a_i S_i a_j S_j R_{ij} T_{ij} + r_{ij}) \quad (1)$$

where b_{ij} represents the specific hydropathic (hydrophobic and polar) interaction between atoms i and j . If $b_{ij} > 0$, the interaction is favorable. If $b_{ij} < 0$, the interaction is unfavorable. a is a hydrophobic atom constant derived from partition coefficients, calculated in a manner similar to the CLOG-P method of Hansch and Leo [48]. S is the solvent-accessible surface area. It represents the fact that the deeper the atom is buried inside a group, the harder it is for it to interact with atoms from other groups and molecules. T_{ij} is a function that differentiates between favorable and unfavorable polar-polar interactions. Its value is +1 for acid–base/hydrogen bonds and –1 for

base–base and acid–acid interactions. R_{ij} and r_{ij} are functions of the distance between atoms i and j . The function r_{ij} represents repulsive and attractive van der Waals interactions. The former interactions are perhaps most important as they dominate when the interacting molecules are too close to each other. H_{TOT} , the HINT interaction score, a cumulative sum of all atom–atom interactions, has been shown to have a linear correlation with the Gibbs free energy of interaction [49].

$$\Delta G = -0.00195 H_{\text{TOT}} - 5.543 \quad (2)$$

It should be noted that partition coefficients not only implicitly encode information about non-bonded intermolecular interactions, but also about the difficulty of reaching those states because the magnitude of partition coefficients correlates with formal charge (and hydrophobic atom constants correlate with partial atomic charges).

Hydropathic interaction has been successfully applied to many problems: in a study of protein–protein interactions in native and mutant hemoglobins, very good correlations were found between HINT scores and the free energies of dimer–dimer association [50–52]; interactions between proteins and small molecules showed a rather good correlation between HINT score and the $\Delta G_{\text{binding}}$ for 53 protein–ligand complexes [49]; studies have also indicated that HINT analysis is useful for understanding DNA–small molecule [53–56] and DNA–protein [57, 58] interactions; recently HINT was also used to study the effect of tyrosine nitration of I κ B α on NF- κ B activation [59].

Computational Titration. One key advantage of HINT is its speed. It can be used for evaluating large sets of data from docking, virtual screening or other sources. This is a major advantage for examining and evaluating the multitude of different protonation states and possible geometries in a protein–ligand complex even when referencing a single high quality crystal structure of a protein–ligand complex. We refer to the collection of molecular models that would fit within the experimental electron density of a crystal structure, but differ in terms of proton placement or special rotations (Asn, Gln, His, etc.), as being “isocrystallographic”. The speed of HINT thus allows optimization of these ill-defined molecular features in a reasonable time frame. This process encompasses the computational titration algorithm for which some aspects have been described previously [60–63]. In this report, we are describing an online version of the HINT-based computational titration method, i.e., a free web service for studying protonation states and Gibbs free energies of binding for protein–ligand complexes. This user-friendly web service can be used to solve quite a few potential problems in protein–ligand structural models; namely: questionable group rotations, optimized rotations for H-donor protons, and multiple

interacting protonation states. The algorithm calculates a HINT score-based free energy of binding for the examined protein–ligand complex models. The server can be accessed at <http://hinttools.isbdd.vcu.edu/CT>.

This service joins an ever expanding field of on-line tools for computational chemistry and drug discovery. It is beyond the scope of this article to fully review this phenomenon, but a few related servers of note are: (1) The Virtual Computational Chemistry Laboratory (<http://www.vcclab.org>) that calculates logP using a variety of algorithms and pK_a [64]; (2) The quantum mechanics-based pK_a (protonation states) prediction tool of Quantum Pharmaceuticals (http://www.q-pharm.com/home/contents/drug_d/order_form/online_services/pka_prediction) [65]; (3) H^{++} (<http://biophysics.cs.vt.edu/H++/>) that computes pK values of ionizable groups in macromolecules and adds missing H atoms according to a specified pH [66]; and (4) PHEPS (pH-dependent Protein Electrostatics Server) (<http://pheps.orgchm.bas.bg/home.html>) that performs local and global pH-dependent analysis of the electrostatics in proteins [67].

Results and discussion

The newly available computational titration web service is based on the computational titration algorithm [62] implemented in the HINT program. While a flow chart of this algorithm has previously been published [60], Fig. 1

contextualizes it within the web server. Computational titration was used successfully in modeling binding affinities of neuraminidase–inhibitor complexes where it confirmed the experimental observations that the binding in these complexes depends on acidity, being the strongest at neutral pH [63]. Computational titration has also been applied to studies of the HIV-1 protease active site. First, the complex between the peptide Glu–Asp–Leu and HIV-1 protease was examined [68] and the experimental dependence of binding on pH was computationally reproduced. In a second study, computational titration experiments for HIV-1 protease with six small molecule inhibitors of this enzyme were performed to optimize protonation state models. Binding energies for five ligands out of six were within an average error of $2.5 \text{ kcal mol}^{-1}$ [62]. Modeling these interactions with other methods would have been very challenging, because these systems can assume many thousands of different ionization state combinations [69]. In many such cases no single ionization state dominates; rather hundreds of different ionization state ensembles are at equilibrium, structurally similar [49] and well within the expected 1 kcal mol^{-1} experimental error for binding energy measurement. Thus, the “optimum” ionization state reported by computational titration may not be the single best state, but one likely close to it in free energy.

To date, the only practical way to use the computational titration algorithm was through a Sybyl [70, 71] interface to HINT available in our local development platform.

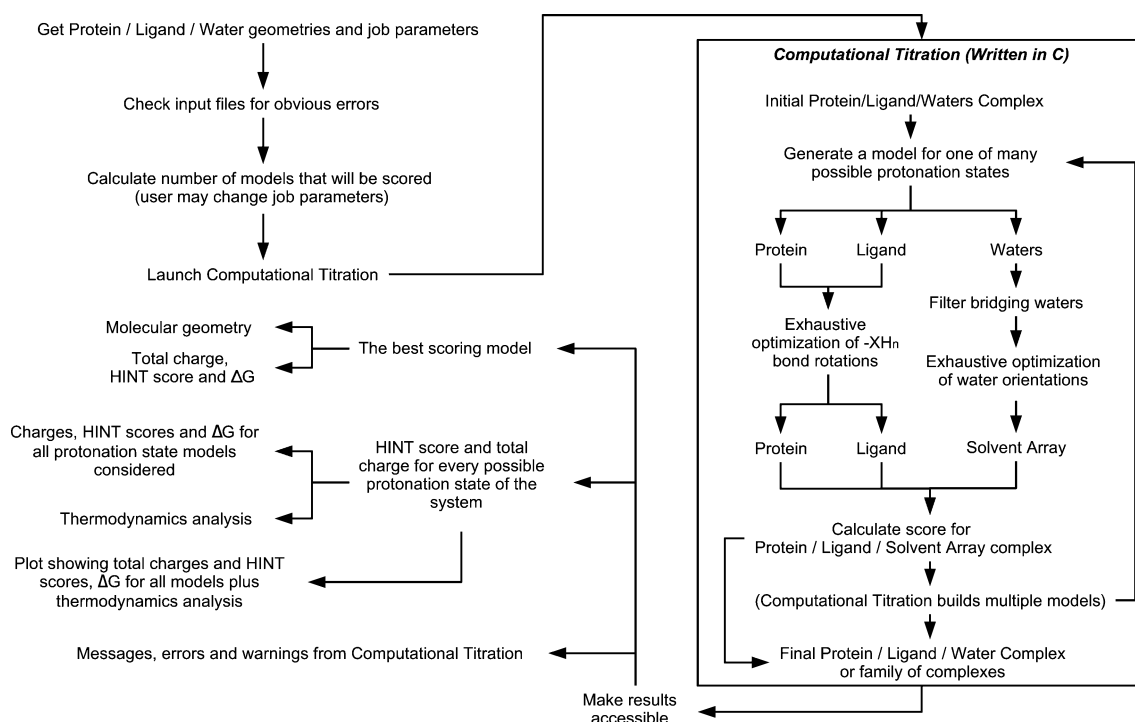


Fig. 1 Flow chart for the computational titration algorithm. Steps performed by C program are enclosed in *box*; remainder is the performed by Python script/web server

While Sybyl is limited to only two platforms, Irix and Linux, we have been developing the underlying HINT toolkit across a much broader set of platforms [72]. However, providing HINT-based tools in a usable form outside the Sybyl infrastructure and user interface has been a limiting factor. In effect, very few users have had access to computational titration. Thus, we have chosen to make it available online to a broader audience in the computational molecular design community.

Implementation and Functional Usage. The computational titration service is implemented in several layers. The entire application is controlled through a web server written in Python [73] that can display static web pages, like the front page, or pages generated by Python-based CGI scripts. These scripts serve several purposes. They provide an easy-to-use HTML interface for the HINT program. The computational titration algorithm is implemented as a binary program written in C and linked to the HINT toolkit. This program is the heart of the web application. In addition to this application, for successful runs Python scripts also interact with Gnuplot [74] to make plots of the computational titration results (*vide infra*). Python CGI scripts are also used to catch a variety of errors in input files and to provide an intuitive interface that helps users to monitor their jobs.

The computational titration server protocol operates in several steps. In the first step the user uploads the protein in PDB format, the ligand in Sybyl mol2 format, and optionally waters in PDB format. The user also selects what types of protonation scenarios should be considered. For example: within what distance from the ligand should the protonation states of protein amino acids be varied, should the input water set be focused to the most relevant [75] waters, and whether one should consider ionization states where phenols are potentially deprotonated on the ligand. A complete set of the current user titration options can be found in Table 1 for protein residues and in Table 2 for the functional groups of the small molecule ligand. This information is also available at <http://hinttools.isbdd.vcu.edu/CT/help.html>. Next, the server enumerates the ensemble of all protonation state models to be considered and presents the user with a choice of whether to go forward with the calculation, modify the options, proceed by sampling the ensemble, or stop. At times, depending on the complexity of the active site environment, the number of protonation states can be unreasonably high. Following this step, the calculation exhaustively evaluates all (or randomly samples) the protonation state models found by the previous step. Each ensemble model has a specific and unique list of atoms to be protonated or deprotonated. After these changes are made to each model, the relevant hydrogen coordinates are *exhaustively* [61] optimized on the protein, ligand and waters, if present, to get the geometry with the best possible

interaction between the protein and the ligand. A key assumption of CT is that differences in these interactions are the primary source of free energy differences between the ensemble models, i.e., other, largely intramolecular, contributions are not considered in the current algorithm. We also need to point out that the exhaustive optimization is a particularly significant feature as it circumvents the possibility of rotatable polar hydrogens, e.g., as in R–OH, R–NH₂ or water, being trapped in local minima and not achieving optimal hydrogen bonding geometry. The Gibbs free energy of binding for each such geometry is calculated from the HINT score, a measure of non-bonding interactions between the molecules that is linearly correlated with their Gibbs free energy of binding (see Eq. 2). Finally, the results of the computational titration are output in several forms: (1) PDB files of the protein and water set and a Sybyl mol2 file of the ligand, representing the molecular ionization state ensemble and geometry with the best interaction between the protein and the ligand; (2) the HINT score and the Gibbs free energy of the protein–ligand interaction for this best model; and (3) a table of normal average, Boltzmann averages, maximum and minimum values of Gibbs free energies of all protonation states at different charges of the active site; and (4) a plot of HINT score and Gibbs free energy plotted as a function of the charge in the active site charge (see Fig. 5). All evaluated protonation ensembles are plotted on this graph, and for each charge, the normal and the Boltzmann averages of Gibbs free energies of all protonation state ensembles associated with that charge are plotted. This plot can help the user to understand how the protein–ligand binding changes with pH or acidity. It also convincingly illustrates how little energy difference exists between many of the models in the protonation ensemble and that there is not *one* true ionization state ensemble, but a family of coexisting closely related states through the exchange of fairly fluid protons.

It should be noted that the computational titration algorithm optimizes the local environment of affected residues and ligand functional groups, but does not energy minimize the active site as a whole. Also, the tool that protonates amines and phosphines can assign unrealistic bond angles for the hydrogens it adds to these groups if the nitrogen or phosphorus is not or poorly puckered. The user should examine the geometry of the generated models and (generally) subject them to hydrogen-only molecular mechanics minimization to finalize the model. Further revisions of the program will likely include some molecular mechanics cleanup of the output models.

The computational titration server is currently asgi Altix 350 system with multiple processor queuing and can thus execute many jobs at the same time. The queuing system is intuitive and allows users to view the progress of their jobs.

Table 1 Protein residue-specific optimization actions available in computational titration


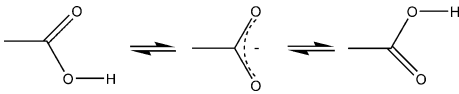
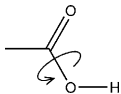
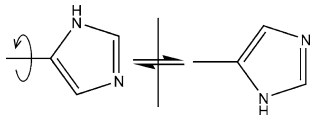
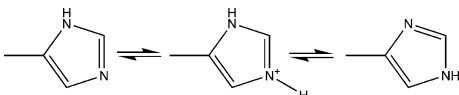
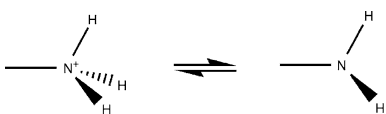

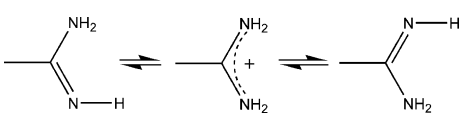
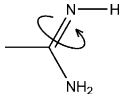

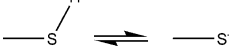
Residue	Action		Details
Asn, Gln	Check amide orientation		Check if amide O and NH ₂ atoms are correctly assigned in the structure
Asp, Glu	Titrate acid		Find the best ionization state for acid residues
	Optimize R-OH rotation		If residue in acid form, exhaustively optimize rotation of -OH
His	Check ring orientation		Check if CD1, NE1, ND2 and CE2 atoms are correctly assigned in the structure
	Titrate base		Find the best ionization state for imidazole ring of His residue
Lys	Titrate base		Find the best ionization state for amine of Lys residue
	Optimize R-NH ₂ rotation		Exhaustively optimize the rotation of -NH ₂ in neutral Lys residues
Arg	Titrate base		Find the best ionization state for guanidine of Arg residue
	Optimize R-NH rotation		If Arg is deprotonated, optimize the rotation angle of imine NH
Tyr	Titrate phenol		Find the best ionization state for Tyr residues
Cys	Titrate thiol		Find the best ionization state for Cys residues

Table 1 continued

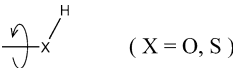
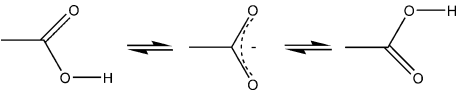
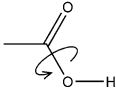


Residue	Action		Details
Cys, Ser, Thr, Tyr	Optimize R–XH (X=O,S) rotation		Exhaustively optimize the rotation of –XH for Cys, Ser, Thr and Tyr residues
C-terminus	Titrate C-term acid		Find the best ionization state for the C-terminus acid
	Optimize R–OH rotation		If C-terminus in acid form, exhaustively optimize rotation of –OH
N-terminus	Titrate base		Find the best ionization state for the protein N-terminus amine
	Optimize R–NH2 rotation		Exhaustively optimize the rotation of –NH2 in neutral N-termini

Table 2 Ligand-specific optimization actions available in computational titration

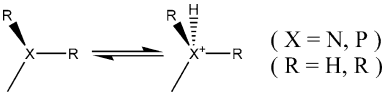
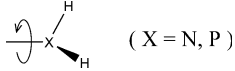

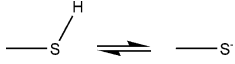
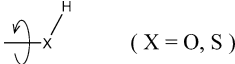
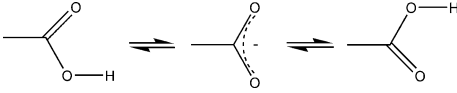
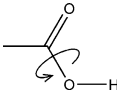
Functional group	Action		Details
Amine, phosphine	Titrate base		Find the best ionization state for ligand amine
	Optimize R–XH2 rotation		Exhaustively optimize the rotation of –NH2 in primary amine
Phenol	Titrate phenol		Find the best ionization state for aromatic alcohol in ligand
Thiol	Titrate thiol		Find the best ionization state for ligand thiol
Hydroxyl, thiol	Optimize R–XH rotation		Exhaustively optimize the rotation of –XH in ligand
Carboxylic acid	Titrate acid		Find the best ionization state for carboxylic acid in ligand
	Optimize R–OH rotation		If in acid form, exhaustively optimize rotation of –OH

Fig. 2 Structural elements of the X-ray crystal structure of liganded β secretase (BACE-1, PDB code 2va7) as modified by computational titration. All residues with ionizable functional groups within 6 Å of the ligand were included in the titratable pool. Water molecules within 4 Å of atoms in both the ligand and protein were selected for optimization. **a** The amide group of Asn37 is flipped 180°—improving its contact with backbone atoms of Arg128. **b** Flipping the amide of Gln12 improves its interaction with Lys9. **c** Protonating the ligand's amine improves its interactions with both Asp32 and Asp228

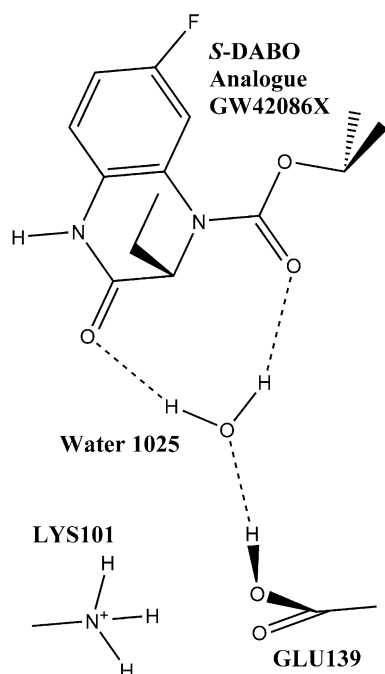
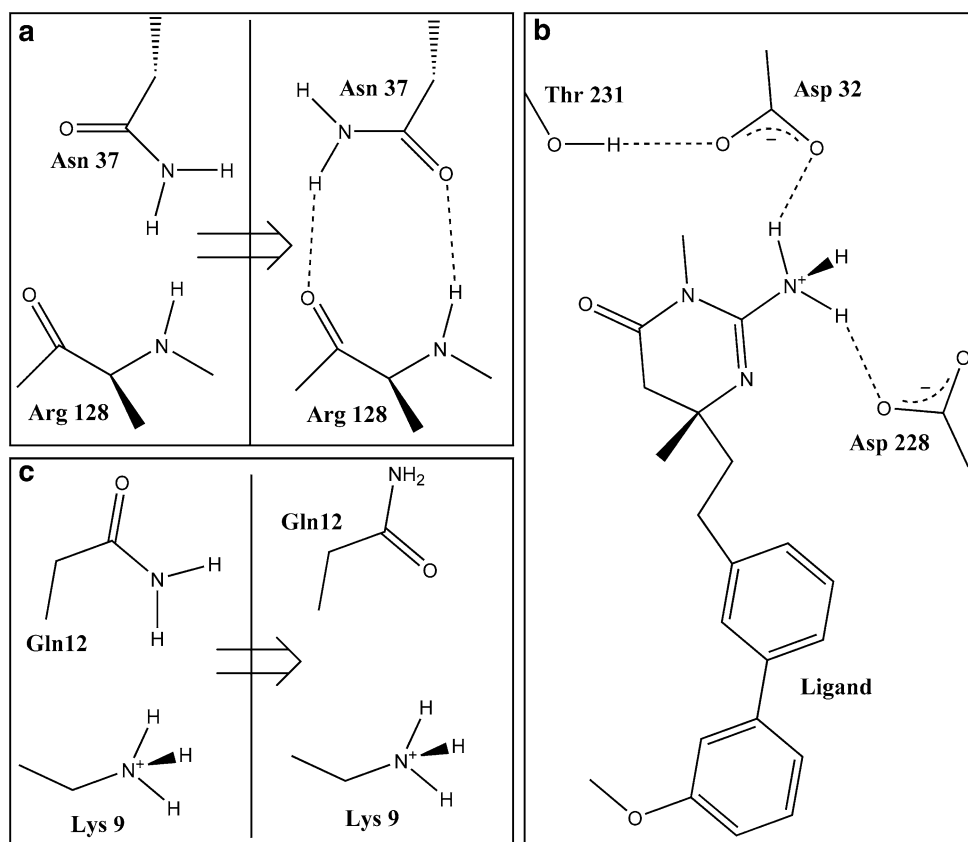


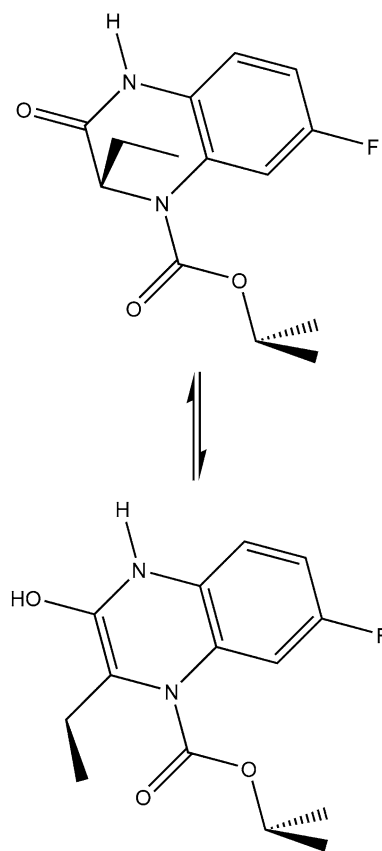
Fig. 3 Active site optimized through computational titration for 2opq. All residues with ionizable functional groups within 6 Å of the ligand were included in the titratable pool. Water molecules within 4 Å of atoms in both the ligand and protein were selected for optimization. Notably, Glu139 is protonated to the acid form and the proton is turned towards water1025 forming a water mediated hydrogen bond with the ligand

Computational Titration Case Studies. One of the difficulties encountered in virtual screening is that often there is a mismatch between the ionization state(s) of the active site residues and proposed incoming ligand if both molecules are modeled as “pH 7”. For example, if a putative ligand placed a carboxylate next to a deprotonated Asp or Glu residue, it would be rejected because of highly unfavorable interactions. Protonating either the ligand carboxylate or the acid residue could in many cases produce a strong hydrogen bond between the species. Thus, it would seem that performing virtual screening based on a crystal structure of a known protein–ligand complex, successfully identifying all potential leads would be enhanced if the hydrogen atoms in the active site are assigned correctly both in terms of protonation state and orientation. In this section we describe a few examples where either assignment of hydrogen atom configuration is not trivial, or the orientation of Asn, Gln or His residues may have been assigned incorrectly. We applied the automated computational titration analysis described above to these structures and report the results here. It should be noted that our application *simultaneously* varies protonation states on the ligand and in the protein active site. Many of the results we obtained show protonation states different from those typically encountered at pH 7, anomalies caused by the very specific nanoenvironments at the protein active sites.

The first example is an X-ray crystal structure of β secretase (BACE-1) complexed with the inhibitor (6*R*)-2-amino-6-[2-(3'-methoxybiphenyl-3-yl)ethyl]-3,6-dimethyl-5,6-dihydropyrimidin-4(3*H*)-one (see Fig. 2) [76]. This structure is stored in the Protein Databank [1] with the code 2va7. Computational titration makes several changes to the structure of this complex. First, it rotates two amide groups (Asn37 and Gln12). In the original structure the side chain of Asn37 was engaged in highly unfavorable acid–acid and base–base interactions with the backbone of Arg128. Our application flipped the amide group on the side chain of Asn37 and transformed these interactions into hydrogen bonds (Fig. 2a). Also, in the original structure the side chain amide group of Gln12 made unfavorable contacts with the amine of Lys9, which was repaired by flipping the Gln12 amide group (Fig. 2b). The program also protonated the primary amine on the ligand and increased the strength of favorable interactions between the ligand and protein residues Asp32 and Asp228 by adding a charge–charge interaction to the existing hydrogen bonds (Fig. 2c). The positions of hydrogens on many waters and alcohols were also adjusted. For example the alcohol group on Thr231 was rotated to form a hydrogen bond with the carboxyl of Asp32 (Fig. 2c). The HINT score for the nominal “pH 7” model, even after molecular mechanics energy minimization, was $-2,236$ (-1.18 kcal mol $^{-1}$), while that of the optimized (titrated) and energy minimized model is $1,031$ (-7.55 kcal mol $^{-1}$), if the free energy is calculated using Eq. 2.

The second example is the crystal structure of L100I mutant HIV-1 reverse transcriptase in complex with the *S*-DABO analogue GW420867X inhibitor (PDB code 2opq) (Fig. 3). For this structure the computational titration application suggests that the ligand binds best when Glu139 is protonated. Most often when a carboxylate group is protonated to an acid, the hydrogen points forward as it is stabilized by interacting with the other (carbonyl) oxygen of the same group. However, in this case the hydrogen of the protonated carboxylate on Glu139 is turned away, towards water1025, making a new hydrogen bond. The usual orientation for this hydrogen would have produced highly unfavorable interactions with the amine group of Lys101. In this model, water1025 serves as a bridge between Glu139 and the ligand's two carbonyl groups by accepting a proton from Glu139 and donating its protons to the ligand. For this case, the HINT score for the nominal, energy minimized, “pH 7” model was 229 (-5.99 kcal mol $^{-1}$), while for the optimized and minimized model the HINT score is 310 (-6.15 kcal mol $^{-1}$). While this is not a large score difference, understanding the role of the bridging water could open up possibilities for displacing it through ligand design [75].

We should point out that a prototropic (keto-enol) tautomeric change in the ligand was not considered for this protein–ligand complex because the resulting enol would not have been isocrystallographic with the ketone. (If this tautomerization occurs, the formed double bond would force the ethyl into the plane of the ring system, which would be observed in a relatively high resolution X-ray structure).



Computational titration produced dramatic results for the crystal structure of the human sialidase NEU2 in complex with the isobutyl ether mimetic inhibitor (*R*)-5-acetylamino-(*S*)-4-hydroxy-(*R*)-6-isobutoxy-5,6-dihydro-4*H*-pyran-2-carboxylic acid (below) (PDB code 2f11). This structure is a refinement of 1snt [77]. If all ionizable residues within 6 Å of the ligand were chosen to define the active set, 944,784 configuration models would have to be analyzed, too many for completion in a reasonable amount of time. A number of options for reducing the scale of this problem are available, including random sampling and calculation for a fraction of models, eliminating functional groups and/or residue types from the titratable set, and reducing the volume of the titratable active set. In this case, we have adjusted the settings by reducing the contact cutoff distance from 6 to 4 Å, and eliminated the possibility of tyrosine and cysteine ionizations, i.e., these residues were not titrated. This reduced the

number of models in the set to 2,187. (In fact, no cysteine residues were present in or around the active site.)

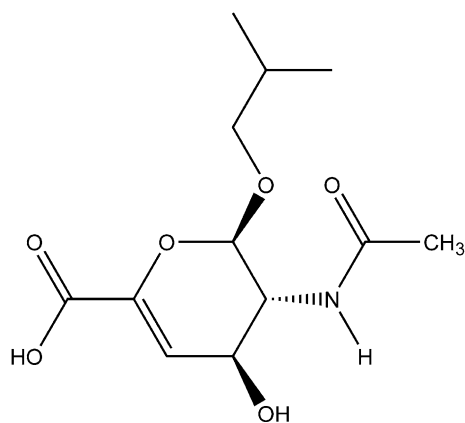


Figure 4 illustrates the active site environment for 2f11. Although the overall resulting change in HINT scores was modest, 1,879–2,245 (-9.21 to -9.92 kcal mol $^{-1}$), a number of significant (balanced between favorable and unfavorable) adjustments in the active site, highlighted in yellow and purple, were proposed by the computational titration algorithm that are instructive for understanding the ligand binding. The first, somewhat unexpected, result was that in the best model Glu218 was protonated. This protonation serves two purposes: first, it reduces the unfavorable interactions between Glu218 and the hydrophobic part of the ligand; second, this protonation influences the geometry of the neighboring residue, Tyr334. In the absence of the ligand, Glu218 would likely have been deprotonated, forming a hydrogen bond with Tyr334; in the presence of the ligand, the hydrogen of the tyrosine hydroxyl group can interact favorably with the ligand's carboxyl group (H...O, 2.52 Å; H...C, 2.40 Å), the heterocyclic oxygen (H...O, 2.62 Å) and the double bond of the ligand (H...C = C, 2.32 Å). The hydrogen donated by the protonated Glu218 helps keep the hydroxyl group of Tyr334 oriented properly for these favorable interactions with the ligand (see Fig. 4a, b).

While it is debatable whether such a change would occur because of the rather unfavorable pK_b for guanidinium,¹ to avoid unfavorable interactions with the hydrophobic part of the ligand, this model suggests that Arg237 is deprotonated, which in turn leads to flipping the amide group on the side chain of Gln270, so that the amide nitrogen of Gln270 can hydrogen bond donate to the (now) deprotonated nitrogen of Arg237 (H...N, 2.25 Å) (see Fig. 4c, d). On the other side of the Gln270 amide the oxygen forms a favorable interaction with one hydrogen of

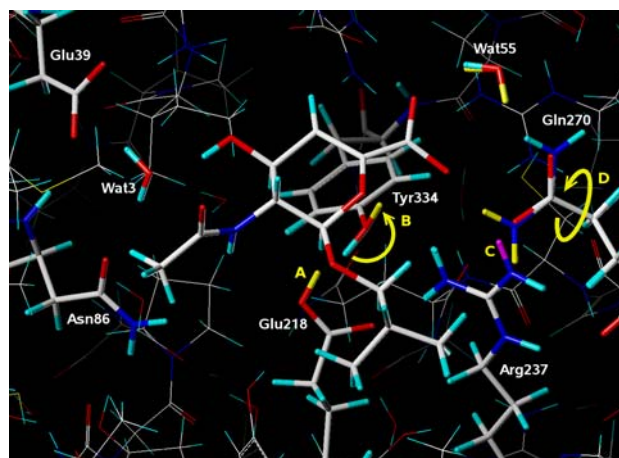


Fig. 4 A comparison between two models of the crystal structure of human sialidase NEU2 in complex with an isobutyl ether mimetic inhibitor (PDB 2f11). In one model hydrogen atoms for ionizable polar atoms were assigned conventionally (H atoms colored cyan and purple), using protonation states that are typically assumed for a neutral solution. In the other model both the protonation state for the ionizable polar groups and the resulting positions of ionizable hydrogens were optimized with computational titration. In both models the positions of hydrogens were energy minimized with molecular mechanics. Atoms colored yellow (added) and purple (deleted) indicate differences between the two models as labeled: **a** Glu218 is protonated; **b** the hydroxyl group on Tyr334 is rotated about 180°; **c** Arg237 is deprotonated; and **d** the amide group on the side chain of Gln270 is flipped

water55 (H...O, 2.50 Å). The other hydrogen of water55 forms a favorable interaction with the carboxylate group of the ligand (H...O, 2.07 Å). Thus, water55 serves as bridge between Gln270 and the carboxylate group of the ligand. Another bridging water molecule is water3. It forms one hydrogen bond by donating to the oxygen atom on the side chain of Asn86 and also hydrogen bonds by donating to Glu39. Water3 also forms a hydrogen bond by accepting from the ligand's hydroxyl group; thus, it significantly increases the strength of favorable interactions between the protein and the ligand.

However, it should be made clear that the model described above and shown in Fig. 4 is by no means the dominant protonation model for this system. There are many other models that are energetically quite similar to this one. Fig. 5 is a graph of HINT score as a function of site charge for the 2,187 models examined by computational titration for this complex. Each red circle represents an individual model at calculated site charge and HINT score (left axis). The right axis indicates the free energy of binding for the ligand as calculated from the HINT score using Eq. 2. The configurations labeled “Best Model” and “pH 7 Model” are those illustrated in Fig. 4. This graph also indicates the average HINT score for each site charge and the Boltzmann weighted energy for each site charge. The Boltzmann term is calculated using Eq. 2 to convert

¹ Guanidinium and guanidine have surprisingly similar hydrophobic properties: all of the nitrogens in either case are both H-bond donors and acceptors; the unsaturation in the neutral species compensates for the loss of the H $^{+}$; and the formal charge is quite delocalized.

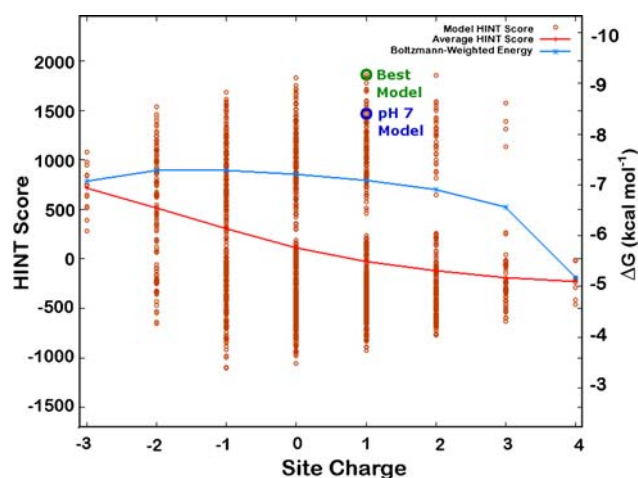


Fig. 5 Titration curve for 2f11. The best model and pH 7 models as shown in Fig. 4 are indicated. See text for more information

HINT score to free energy. It should be noted that Eq. 2 is a “generic” conversion scheme with an uncertainty of $\pm 2.5 \text{ kcal mol}^{-1}$ [49]. We have shown that scoring functions specifically calibrated to convert HINT score to free energy for particular protein targets are more robust and can have uncertainties of around $1.0 \text{ kcal mol}^{-1}$ or less [78, 79]. The Boltzmann-weighted average shows a maximum (highest binding energy) at a site charge of -2 . This value may be interpreted as an expectation value for experimental measurement of binding energy, where many energetically similar states may be occupied. Likewise, it is apparent from this graph that many models *are* energetically similar. In fact, the difference in energy between the pH 7 Model and the Best Model is only about $0.71 \text{ kcal mol}^{-1}$, and neither are of the same binding site charge as the maximum in the Boltzmann-weighted energy curve.

Conclusions

We have implemented the computational titration algorithm as a freely available web service at <http://hinttools.isbdd.vcu.edu/CT>. This web server is designed to be an intuitive tool that can help users improve their models of protein–ligand interactions as well as calculate the Gibbs free energies of binding for protein–ligand complexes at various acidities. The basis of free energy scoring for computational titration is the HINT program. While this scoring function is simplistic, it has been fruitfully used in a number of studies of protein–ligand interactions, and is rapid enough to make computational titration practical. It should be emphasized that absolute free energies of binding are difficult to calculate, but the relative energies and ordering between protonation models should be fairly reliable. The goal of this modeling tool, and in general

most modeling tools, is to facilitate visualization of complex phenomena. The myriad of protonation ensembles in a protein–ligand complex is particularly challenging, so we feel that the computational titration tool will be of benefit to the modeling community. We should note that, because the user selects which residue types are subject to the titration protocol, the user is controlling a crucial aspect of the algorithm. In particular this is, in effect, allowing the user to decide which ionizations he believes are germane to the system. If ionization of tyrosine *is* selected, some models *will* be created with tyrosinate. We are exploring automation of this aspect, i.e., incorporating modeled pK_a s for each ionizable functional group in the computational titration algorithm as an intramolecular contribution to free energy that may be significant in some cases, but at present this is not available.

We are currently developing additional functionality for computational titration and will make it available as it is coded and validated. In particular we are considering: (1) improved optimization of protonated amines and phosphines. The current algorithm for adding protons in such cases is purely geometric, whereas more reasonable conformations would be obtained by a molecular mechanics approach; (2) support for titrating additional functional groups such as phosphates and sulfates; (3) support for titrating nucleotide backbone phosphates and sugars; and (4) implementation of stochastic optimization algorithms to find the best protonation state in cases where the number of potential models is exceedingly large. Nonetheless, even in its current implementation, computational titration should be a useful tool for creating starting ionization state models for protein–ligand complexes.

Acknowledgments We gratefully acknowledge the support of the US N.I.H. Grant GM071894. In addition, the assistance of Dr. P. D. Mosier in configuring the server hardware and software is greatly appreciated.

References

1. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) Nucl Acids Res 28:235
2. Weichenberger CX, Sippl MJ (2006) Bioinformatics 22:1397
3. Weichenberger CX, Sippl MJ (2006) Structure 14:967
4. Word JM, Lovell SC, Richardson JS, Richardson DC (1999) J Mol Biol 285:1735
5. Hooft RW, Sander C, Vriend G (1996) Proteins 26:363
6. McDonald IK, Thornton JM (1995) Protein Eng 8:217
7. Cozzini P, Fornabaio M, Marabotti A, Abraham DJ, Kellogg GE, Mozzarelli A (2004) Curr Med Chem 11:3093
8. Gohlke H, Klebe G (2002) Angew Chem Int Ed 41:2644
9. Lazaridis T (2002) Curr Org Chem 6:1319
10. Wang W, Donini O, Reyes CM, Kollman PA (2001) Annu Rev Biophys Biomol Struct 30:211
11. Ajay, Murcko MA (1995) J Med Chem 38:4953

12. Böhm H-J (1994) *J Comput-Aided Mol Des* 8:243
13. Böhm H-J (1998) *J Comput-Aided Mol Des* 12:309
14. Eldridge MD, Murray CW, Auton TR, Paolini GV, Mee RP (1997) *J Comput Aided Mol Des* 11:425
15. Brady GP, Sharp KA (1997) *Curr Opin Struct Biol* 7:215
16. Ben-Naim A (1997) *J Chem Phys* 107:3698
17. Dill KA (1997) *J Biol Chem* 272:701
18. Chou JJ, Li S, Klee CB, Bax A (2001) *Nat Struct Biol* 8:990
19. Gohlke H, Hendlich M, Klebe G (2000) *J Mol Biol* 295:337
20. Reddy MR, Erion MD, Agarwal A (2000) *Rev Comput Chem* 16:217
21. Kollman PA (1993) *Chem Rev* 93:2395
22. Åqvist J, Luzhkov VB, Brandsal BO (2002) *Acc Chem Res* 35:358
23. Jorgensen WL (1989) *Acc Chem Res* 22:184
24. Reddy MR, Erion MD, Agarwal A (2000) In: Lipkowitz KB, Boyd DB (eds) *Reviews in computational chemistry*, vol 16. Wiley-VCH, New York, pp 217–304
25. Weichenberger CX, Sippl MJ (2007) *Nucl Acids Res* 35:W403
26. Davis IW, Murray LW, Richardson JS, Richardson DC (2004) *Nucl Acids Res* 32:W615
27. Hooft RW, Vriend G, Sander C, Abola EE (1996) *Nature* 381:272
28. Zhong W, Gallivan JP, Zhang Y, Li L, Lester HA, Dougherty DA (1998) *Proc Natl Acad Sci USA* 95:12088
29. Murphy RB, Philipp DM, Friesner RA (2000) *J Comput Chem* 21:1442
30. Gao J, Xia X (1992) *Science* 258:631
31. Kuhn B, Kollman PA, Stahl M (2004) *J Comput Chem* 25:1865
32. Simonson T, Carlsson J, Case DA (2004) *J Am Chem Soc* 126:4167
33. Mongan J, Case DA, McCammon JA (2004) *J Comput Chem* 25:2038
34. Nicholls A, Honig B (1991) *J Comput Chem* 12:435
35. Voges D, Karshikoff A (1998) *J Chem Phys* 108:2219
36. Krishtalik LI, Kuznetsov AM, Mertz EL (1997) *Proteins: Struct Funct Genet* 28:174
37. Demchuk E, Wade RC (1996) *J Phys Chem* 100:17373
38. Antosiewicz J, McCammon JA, Gilson MK (1994) *J Mol Biol* 238:415
39. Yang A-S, Gunner MR, Sampogna R, Sharp K, Honig B (1993) *Proteins: Struct Funct Genet* 15:252
40. Beroza P, Fredkin DR, Okamura MY, Feher G (1991) *Proc Natl Acad Sci USA* 88:5804
41. Bashford D, Karplus M (1990) *Biochemistry* 29:10219
42. Georgescu RE, Alexov EG, Gunner MR (2002) *Biophys J* 83:1731
43. Bartlett GJ, Porter CT, Borkakoti N, Thornton JM (2002) *J Mol Biol* 324:105
44. Koumanov A, Ruterjans H, Karshikoff A (2002) *Proteins: Struct Funct Genet* 46:85
45. Van Vlijmen HWT, Schaefer M, Karplus M (1998) *Proteins: Struct Funct Genet* 33:145
46. Nielsen JE, Vriend G (2001) *Proteins: Struct Funct Genet* 43:403
47. Mehler EL, Guarnieri F (1999) *Biophys J* 77:3
48. Hansch C, Leo AJ (1979) *Substituent constants for correlation analysis in chemistry and biology*. Wiley, New York
49. Cozzini P, Fornabaio M, Marabotti A, Abraham DJ, Kellogg GE, Mozzarelli A (2002) *J Med Chem* 45:2469
50. Burnett JC, Botti P, Abraham DJ, Kellogg GE (2001) *Proteins: Struct Funct Genet* 42:355
51. Burnett JC, Kellogg GE, Abraham DJ (2000) *Biochemistry* 39:1622
52. Abraham DJ, Kellogg GE, Holt JM, Ackers GK (1997) *J Mol Biol* 272:613
53. Cashman DJ, Kellogg GE (2004) *J Med Chem* 47:1360
54. Cashman DJ, Scarsdale JN, Kellogg GE (2003) *Nuc Acid Res* 31:4410
55. Kellogg GE, Scarsdale JN, Cashman DJ (1999) *Med Chem Res* 9:592
56. Kellogg GE, Scarsdale JN, Fornari FA (1998) *Nuc Acid Res* 26:4721
57. Spyraakis F, Cozzini P, Bertoli C, Marabotti A, Kellogg GE, Mozzarelli A (2007) *BMC Struct Biol* 7:4
58. Marabotti A, Spyraakis F, Facchiano A, Cozzini P, Alberti S, Kellogg GE, Mozzarelli A (2008) *J Comput Chem* 29:1955
59. Yakovlev VA, Barani IJ, Rabender CS, Black SM, Leach JK, Graves PR, Kellogg GE, Mikkelsen RB (2007) *Biochemistry* 46:11671
60. Kellogg GE, Fornabaio M, Chen DL, Abraham DJ, Spyraakis F, Cozzini P, Mozzarelli A (2006) *J Mol Graph. Model* 24:434
61. Kellogg GE, Chen DL (2004) *Chem Biodiver* 1:98
62. Tripathi A, Fornabaio M, Spyraakis F, Mozzarelli A, Cozzini P, Kellogg GE (2007) *Chem Biodiver* 4:2564
63. Fornabaio M, Cozzini P, Mozzarelli A, Abraham DJ, Kellogg GE (2003) *J Med Chem* 46:4487
64. Tetko IV, Gasteiger J, Todeschini R, Mauri A, Livingstone D, Ertl P, Palyulin VA, Radchenko EV, Zefirov NS, Makarenko AS, Tanchuk VY, Prokopenko VV (2005) *J Comput Aid Mol Des* 19:453
65. http://www.q-pharm.com/home/contents/drug_d/order_form/online_services/pka_prediction (accessed October 2008)
66. Gordon JC, Myers JB, Foltz T, Shoja V, Heath LS, Onufriev A (2005) *Nucl Acid Res* 33:W368
67. Kantardjiev AA, Atanasov BP (2006) *Nucl Acid Res* 34:W43
68. Spyraakis F, Fornabaio M, Cozzini P, Mozzarelli A, Abraham DJ, Kellogg GE (2004) *J Am Chem Soc* 126:11764
69. Navia MA, Fitzgerald PMD, McKeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL, Springer JP (1989) *Nature* 337:615
70. Sybyl, version 7.3 (2006) Tripos Associates Inc., 1699 S Hanley Rd., St. Louis, MO 63144, USA
71. <http://www.edusoft-lc.com/hint> (Accessed May 2008)
72. Kellogg GE, Fornabaio M, Chen DL, Abraham JD (2005) *Internet Electr J Mol Design* 4:194
73. <http://www.python.org> (Accessed May 2008)
74. <http://www.gnuplot.info> (Accessed May 2008)
75. Amadasi A, Surface JA, Spyraakis F, Cozzini P, Mozzarelli A, Kellogg GE (2008) *J Med Chem* 51:1063
76. Edwards PD, Albert JS, Sylvester M, Aharony D, Andisik D, Callaghan O, Campbell JB, Carr RA, Chessari G, Congreve M, Frederickson M, Folmer RH, Geschwindner S, Koether G, Kolmodin K, Krumrine J, Mauger RC, Murray CW, Olsson LL, Patel S, Spear N, Tian G (2007) *J Med Chem* 50:5912
77. Chavas LMG, Tringali C, Fusi P, Venerando B, Tettamanti G, Kato R, Monti E, Wakatsuki S (2005) *J Biol Chem* 280:469
78. Spyraakis F, Amadasi A, Fornabaio M, Abraham DJ, Mozzarelli A, Kellogg GE, Cozzini P (2007) *Eur J Med Chem* 42:921
79. Fornabaio M, Spyraakis F, Mozzarelli A, Cozzini P, Abraham DJ, Kellogg GE (2004) *J Med Chem* 47:4507