

Published in final edited form as:

*Proteins*. 2009 October ; 77(1): 52–61. doi:10.1002/prot.22415.

## Automated Site Preparation in Physics-Based Rescoring of Receptor Ligand Complexes

Chaya S. Rapp<sup>1,a</sup>, Cheryl Schonbrun<sup>1</sup>, Matthew P. Jacobson<sup>2</sup>, Chakrapani Kalyanaraman<sup>2</sup>, and Niu Huang<sup>3</sup>

<sup>1</sup> Department of Chemistry, Stern College for Women, Yeshiva University, New York NY

<sup>2</sup> Department of Pharmaceutical Chemistry, University of California San Francisco, San Francisco, CA

<sup>3</sup> Department of Computational Chemistry, National Institute of Biological Sciences, Beijing, China

### Abstract

Hydrogen atoms are not typically observable in xray crystal structures but inferring their locations is often important in structure-based drug design. In addition, protonation states of the protein can change in response to ligand binding, as can the orientations of OH groups, a subtle form of “induced fit”. We implement and evaluate an automated procedure for optimizing polar hydrogens in protein binding sites in complex with ligands. Specifically, we apply the previously described ICDA algorithm (*Proteins* **66**: 824–837), which assigns the ionization states of titratable residues, the amide orientations of Asn/Gln side chains, the imidazole ring orientation in His, and the orientations of OH/SH groups, in a unified algorithm. We test the utility of this method for identifying native-like ligand poses using 247 protein-ligand complexes from an established database of docked decoys. Pose selection is performed with a physics-based scoring function based on a molecular mechanics energy function and a Generalized Born implicit solvent model. The use of the ICDA receptor preparation protocol, implemented with no knowledge of the native ligand pose, increases the accuracy of pose selection significantly, with the average RMSD over all complexes decreasing from 2.7 to 1.5 Å when applying ICDA. Large improvements are seen for specific classes of binding sites with titratable groups, such as aspartyl proteases.

### Keywords

pKa prediction; force field based scoring function; rescoring docked complexes; decoys; molecular mechanics; MM-GBSA; ICDA

## I. Introduction

The two major challenges in virtual ligand screening for drug design are sampling—enumerating possible conformations of ligands in the receptor—and scoring, i.e., estimating the relative binding affinities among different poses of the same ligand and among different potential ligands [1–18]. In general, current docking programs are effective at generating native like binding modes but are less effective at identifying the correct binding mode among alternatives and rank-ordering ligands according to predicted binding affinity [3,6,8,10,18]. A great deal of effort has been devoted to developing and improving scoring functions for these purposes, using a variety of strategies, often categorized as knowledge

<sup>a</sup>Correspondence to: Chaya S. Rapp, Stern College for Women, Yeshiva University, 245 Lexington Avenue, New York 10016. phone: 212-340-7744. fax: 212-340-7788. email: rappc@yu.edu.

based [11,19,20], empirical [4,7,15,21,22], consensus [23–26], or force field based [12,14,16,27–34].

Thus far, knowledge-based and empirical scoring functions have been much more extensively used than force-field based methods. This has been due in part to computational expense associated with computing force field based energies, especially when treating the effects of water using implicit solvent models. However, recently there have been a number of studies evaluating the accuracy of these types of scoring functions using different metrics of success [12,14,16,27–34]. In our own studies, we have noted that the results of using MM-GBSA scoring functions for protein-ligand complexes can be exquisitely sensitive to the treatment of hydrogen atoms in the binding site, a conclusion also reached by other authors [31]. Receptor preparation, including the treatment of hydrogen atoms, is frequently accomplished by manual inspection, but here we investigate whether it is possible to perform this task in an automated way, increasing the robustness and reproducibility of physics-based scoring of protein-ligand complexes.

The problem of assigning hydrogen atoms to crystal structures is an old one and we build upon much prior work. For most experimental structures, the positions of hydrogen atoms cannot be resolved and thus protonation states of titratable residues, the tautomeric states of His, and the orientations of hydroxyl or thiol groups are undefined, and may in fact vary depending on what ligand is bound. The orientations of the side chains of Asn, Gln and His may be similarly undefined because of the difficulty in differentiating between N, C, and O atoms from electron densities obtained from crystallography. A number of algorithms have been described for generating all-atom protein models from heavy-atom-only crystal structures, which deal with some or all of these ambiguities, e.g., REDUCE [35], WHATIF [36,37], and various methods for predicting protonation states/pKa's based on molecular mechanics and implicit solvent [38–42]. Recently, a new method called the Independent Cluster Decomposition Algorithm (ICDA) [43] has been described that assigns the ionization states of titratable residues, the amide orientations in Asn/Gln, the imidazole ring orientation in His, and the orientations of OH/SH groups in a unified algorithm that utilizes an all atom protein force field (OPLS) and a Generalized Born continuum solvation model for scoring different possibilities.

Here we use the ICDA algorithm in the context of docking by applying it to protein-ligand binding sites. The atomic level details predicted by ICDA can significantly impact scoring, e.g., errors can result if hydrogen bond donors are incorrectly treated as acceptors, charged side chains are treated as neutral, etc. A common practice in docking and scoring protocols has been to assign protonation and tautomeric states by visual inspection [10,14,16], e.g., protonating one of two neighboring aspartic acid side chains in the active site of HIV-protease [44]. Others have modified titratable residues based on their dominant states at neutral pH [1,5,45] or based on the pH of the experimental crystal structure or binding assay [44–46]. The drawback of these approaches is that the pKa of a given group depends on its local environment (electrostatic potential, solvent exposure) which may change in response to ligand binding [47–51]. Similarly, the orientations of OH groups in Ser, Thr, and Tyr, and amide groups in Asn and Gln, can also change in response to ligand binding (a subtle form of “induced fit”) [52].

In this study we examine whether ICDA can be used to improve the identification of native-like binding poses. We use ICDA to determine positions of all atoms undefined by the crystal structure in an automated protocol which does not involve user intervention. Pose selection is performed with a physics-based scoring function based on a molecular mechanics energy function and a Generalized Born implicit solvent model [53–55], the same energy function used in the ICDA procedure. This type of scoring function has been shown

to be effective in identifying binders in virtual screening [12,16,28] and rescoring conformational ensembles to predict the native ligand pose [14]. Similar Molecular Mechanics/Poisson Boltzmann approaches have been used for the calculation of binding affinities [56–59].

It is important to note that there are several alternative approaches to “physics-based” scoring of protein-ligand complexes, including the problem of assigning protonation states, other than the approaches pursued here. Notably, the PDL/D/S-LRA method, which uses the Protein Dipoles Langevin Dipoles approach to treat solvated electrostatics rather than the more approximate Generalized Born implicit solvent models used here, has been shown to accurately reproduce binding affinities in a variety of protein-ligand systems [32–34]. The PDL/D approach has also been used to accurately predict pKa's of groups in proteins [41,42] suggesting that it could provide a self-consistent, physics-based approach to both assigning protonation states and estimating binding affinities.

Our focus is more modest, and we judge success of the ICDA method for preparing the binding site by the ability to subsequently identify near-native conformations of ligands. To focus on the problem of scoring, as distinct from ligand sampling, we test our scoring protocol using an established database of docked decoys [14,60]. The ability to select native-like poses is tested for a diverse set of 247 protein-ligand complexes taken from the Ligand Protein Database (LPDB) of Brooks et al. [61]. We perform two types of tests. First, we apply the ICDA algorithm to the crystal structure of the protein-ligand complex, and then use that prepared receptor to score the decoys. The native complex would not be known in any realistic application, but this test provides a benchmark for how much improvement may be possible by applying ICDA. We refer to this test below as “biased” (by knowledge of the native structure). Second, we apply ICDA in an “unbiased” way by applying it to every decoy for all 247 cases in the test set (~12,000 protein-ligand structures in total). We compare the results for both of these tests to a “default” receptor preparation procedure that does not optimize protonation states or other optimizations performed by ICDA. We demonstrate that the use of the receptor preparation protocol, in either the “biased” or “unbiased” manner, increases the accuracy of pose selection significantly, with large improvements seen for specific classes of binding sites with titratable groups, such as aspartyl proteases. Overall, our results highlight the importance of optimizing protonation states and the positions of polar hydrogens in binding sites, and present a promising automated method for addressing this challenge as part of a virtual screening protocol.

## II. Materials and Methods

### Test Set

Our test set is taken from the Ligand Protein Database (LPDB) of Brooks et al. [61]. That database contains 262 receptor ligand complexes, drawn from the Protein Databank (PDB) [62], and includes 76 unique proteins and 240 unique ligands. For each complex, the LPDB contains 10–120 (50 on average) ligand poses in the receptor binding site, referred to as “active site” decoys. The following complexes are excluded from our study: 1c2t, 1cbs, 1dy9, 1jld, 1mcb, and 1hsh because the ligand file (1mcb), the decoy files (1c2t, 1cbs, 1hsh) or all files (1dy9, 1jld) are missing from the LPDB; 1ac0 and 2acs because we could not create parameter files for the ligand glc (1ac0) and the cofactor nap (2acs); 1bid because the optimization of the native receptor failed; and 1cvu, 1hge, 1hgh, 1hgj, 4hmg, and 5cna because application of the ICDA algorithm failed due to the large size of the receptor (> 900 residues). In total, 247 receptor ligand complexes are included in our study. These complexes include a range of receptor types, including 85 aspartyl proteases, 38 oxoreductases, 31 serine peptidases, 10 immunoglobulins, 14 metallo peptidases, 8 transferases, 6 isomerases, and several others.

## Rescoring Protocol

All calculations are run in PLOP (free academic version 8.1; commercial version marketed as Prime by Schrodinger Inc.) [63–65] using the all-atom OPLS force field OPLS-AA [53,54] and the Surface Generalized Born implicit solvent model [55] described in earlier works. Three sets of calculations were performed:

1. The “default protocol” is run without application of the ICDA algorithm. The CHARMM [66] minimized receptor structure is taken from the LPDB and subjected to energy minimization in which all hydrogen atoms are free to move. Each decoy ligand is then inserted into the optimized receptor and subject to energy minimization [67].
2. The “biased protocol” is used as a benchmark of how well we could expect to do using automated site preparation. In this protocol, the native receptor is prepared and all protein residues, regardless of their distance from the ligand, are optimized using ICDA in the presence of the ligand in its crystallographically observed pose. Each decoy is then minimized in this prepared receptor.
3. Same as the “biased” protocol except that ICDA is applied to every ligand pose. This is referred to as the “unbiased protocol”. For this, we performed ICDA optimization on a reduced set of residues, for computational efficiency. Specifically, we compiled a list of residues within 3.6 Å of any decoy pose.

The ICDA algorithm is described in an earlier work [43] and determines positions of all atoms undefined by the crystal structure, including protonation states for all ionizable residues (His, Glu, Asp, Lys, Arg, Cys, Tyr), ambiguous side chain orientations (His, Asn, Gln) and ambiguous OH/SH group orientations (Tyr, Cys, Ser, Thr). The program assigns polar states in clusters based on neighboring polar residues. The assignments of each cluster are combined and all residues are subject to energy minimization [43]. A pH value of 7.0 is assumed in all calculations.

The unbiased protocol is significantly more computationally intensive than the default or biased protocols, as it requires application of the ICDA algorithm to every ligand pose. For example, for the complex 1a07 with a set of 60 decoys, the total computational expense for the default, “biased”, and “unbiased” protocols were 29 cpu-minutes, 48 cpu-minutes, and approximately 17.5 cpu-hours, respectively. In practical application, we expect that ICDA would not be applied to as many poses per compound as we have done here. Our primary goal here is to provide a proof-of-concept, and we did not focus on algorithmic speed. We expect that it will be possible to improve the speed of the ICDA algorithm as well.

## Energies and RMSD

Energies and RMSDs are calculated with the lowest energy decoy being ranked as the top scoring pose. The native pose is not included in the ranked structures.

In the case of the “unbiased” protocol, the protonation states for different decoy poses can vary. To account for this when comparing energies for different poses, we used a standard thermodynamic cycle where the reference state is amino acid dipeptides in solution. For example, a correction factor of  $-60.65$  kcal, the energy difference of charged and neutral glutamate dipeptide, is added for every protonated Glu residue. The pH is accounted for using the term  $2.3RT \cdot (\text{pH} - \text{pK}_a)$ . Similar correction factors are applied for all changes in protonation state. The relevant parameters are described in the ICDA paper [43].

RMSD values are calculated in PLOP and are reported for all heavy atoms in the ligand with respect to the CHARMM minimized ligand taken from the LPDB.

### III. Results and Discussion

#### Effect of Receptor Preparation

The results of decoy pose selection for the 247 complexes are summarized in Table 1, which reports the average RMSD of the selected, ie: lowest energy scoring, decoy pose for each protocol, as well as the fraction of cases where the selected pose is  $\leq 2$  Å RMSD, a commonly used cutoff, and  $\leq 1$  Å RMSD, a stringent cutoff. The application of the ICDA algorithm clearly results in substantial improvements in these overall results, and applying it in the “unbiased” protocol results in only a small degradation of the results relative to the “biased” results. This is important, because it suggests that 1) it is not necessary to know the native pose, only to sample conformations near it, and 2) it is not necessary to apply ICDA to all residues in the protein, as we did in the “biased” results, only to the residues close to the ligand, as we did in the “unbiased” results, greatly reducing the computational expense. Our results when using the ICDA optimization are comparable to the best reported in the literature for this dataset; an assessment of scoring functions by Ferrara et al. [31] reports a success rate ( $\leq 2$  Å RMSD) of 80–90% for a subset of 189 LPDB complexes. Our results also highlight the utility of an MM-GBSA energy function for pose selection, as also demonstrated in other work [14,16,28].

Figure 1 reports the breakdown of RMSD for all protocols. The effect of the ICDA optimization is particularly evident in the significant increase in cases with RMSD  $\leq 1$  Å and the decrease in failed predictions with RMSD  $>4$  Å. Table 2 lists 41 cases for which ICDA optimization makes a large difference in pose prediction, defined as a difference in RMSD between the default and unbiased protocols of greater than 2 Å. For each case, the columns list the RMSD of the selected pose, ie: the lowest energy scoring decoy pose, for the “unbiased” ICDA and default protocols. To examine why receptor preparation makes a difference in scoring accuracy in these cases, the table lists the changes in protonation states (from defaults) in the binding site predicted by the ICDA algorithm. Out of 95 changes, 66 changes relate to the protonation of an Asp side chain, while 21 changes relate to a His residue, and 9 changes relate to protonation of a Glu side chain. The exceptionally high RMSD obtained for 1ivp in the default protocol is due to a decoy positioned outside of the protein, far from the binding site.

Table 3 lists the difference in energy between the native and selected pose for all cases listed in Table 2. Energy differences are calculated as  $E_{\text{decoy}} - E_{\text{native}}$ , with the native referring to the native ligand pose, subject to the same protocol as the set of decoys. The energy differences obtained from both the unbiased and default protocols show that the lowest energy decoys poses, in most cases, achieve energies lower than the optimized native. This indicates that the decoy sets provide an effective search of conformational space. The ICDA receptor preparation method also dramatically reduces the energy gaps between the native and selected poses in many cases.

Table 4 lists instances of changes in ambiguous Asn/Gln/His side chain orientations, i.e., “flipped” residues, within 3.6 Å of the binding site among the 41 complexes in Table 2. However, the flipped residues do not interact directly with the ligand, and as such likely do not contribute to the accuracy of pose selection.

#### Aspartyl Proteases

Table 2 shows that, among cases in which ICDA optimization makes a difference, three receptors dominate: HIV protease (11 complexes), endothiapepsin (9 complexes), and penicillopepsin (8 complexes). These receptors are aspartyl proteases characterized by a catalytic dyad of aspartic acid residues in the active site: Asp 25 and Asp 25' in HIV protease, Asp 32 and 215 in endothiapepsin, and Asp 33 and 213 in penicillopepsin. In



general, experimental and theoretical studies of these complexes [68–72] have shown protonation of one of the Asp residues on the receptor, with the exception of complexes of phosphinate and phosphonate inhibitors in which the PO<sub>2</sub> group is stabilized by protonation of both Asp residues [73].

Our results for HIV protease are consistent with those reported in the literature and show that in general the receptor is protonated on either Asp 25 or Asp 25'; for example complex 1aaq shown in Figure 2a. The only HIV protease complex in Table 2 containing a phosphinate inhibitor is 1hos shown in Figure 2b, and our results are consistent with studies showing that complexes of these inhibitors are protonated on both Asp residues [73].

For endothiapepsin and penicillopepsin, Table 2 shows the protonation of one dyad Asp residue (Asp 32 in endothiapepsin and Asp 33 in penicillopepsin) for the complexes that do not contain phosphonate or phosphinate inhibitors (1eed, 1epp, 2er7, 3er5, 4er1, 4er2, 1apu, 1apv, 1apw). For the complexes that contain phosphinate and phosphonate inhibitors (1ent, 2wea, 2web, 1ppk, 1ppl, 1ppm), both dyad Asp residues (Asp 32 and 215 in endothiapepsin, and Asp 33 and 213 in penicillopepsin) are protonated, for example 1ppk in Figure 2c. Our results show that the endothiapepsin complexes 3er3 and 5er2 are diprotonated. These complexes do not contain phosphonate or phosphinate inhibitors, or any other type of negatively charged ligand, and as such we have no simple explanation for these results. The other protonation state changes listed for the endothiapepsin and penicillopepsin receptors are for residues that do not interact directly with the ligand, with the exception of Asp 12 which is in close proximity to an acceptor oxygen on the ligands in complexes 2er7 and 3er5.

### Effect of Cofactors and Ions

Table 5 reports average RMSD for complexes that contain a cofactor or ion within 3.6 Å of the ligand. These results indicate successful parameterization of the relevant cofactors in the all-atom OPLS-AA force field [53–54]. Some previous studies have noted challenges in pose selection when a metal cation is coordinated to the ligand in the binding site [20]. While our results do not show higher RMSD when a manganese, magnesium or iron cation is in proximity to the ligand, the presence of a zinc ion does result, on average, in higher RMSD.

### Effect of Structure Resolution

The resolution of the crystal complexes in the LPDB ranges from 1.0 to 3.2 Å; 130 complexes have a resolution of below 2 Å. The resolution of the crystal complex is shown to correlate slightly with the RMSD of pose selection. Using the “unbiased” protocol, native-like poses (RMSD ≤ 2 Å) are selected for 87% of complexes with resolution below or equal to 2 Å, and 82% of complexes with resolution poorer than 2 Å.

### Effect of pH

Our protocol assumes a pH of 7 in all calculations. We made this choice because pH information is inconsistently reported in PDB files. In reality, however, many complexes were obtained at acidic or basic pH values. This inconsistency may account for some prediction failures. Using the “biased” protocol, a total of 19 complexes result in RMSD > 4 Å. Among these failures, six are for complexes with pH values below 6.5 (1a30, 1afl, 1bai, 1cpi, 1eed, 1epo) and another six (1a07, 1avn, 1ba8, 1cnw, 1cnx, 1cny) are for complexes with pH values above 7.5.

For some complexes in the LPDB [74], the database contains alternative sets of decoys generated with a modified protonation state on either the receptor or ligand. Running the

protocol using these modified structures gives insight into the effect of pH on pose selection. For example, for the protease 1bai which was crystallized at a pH of 6.2, an alternative decoy set was generated using a modified ligand, protonated on a carboxylic acid group. Using the modified ligand and alternative decoy set results in an RMSD of 2.0 Å, in contrast to the unmodified set which results in an RMSD of 16.5 Å.

### Other Causes of Failed Pose Selection

Several of the failed predictions are for ligands that are significantly solvent exposed, for example several carbonic anhydrase II complexes (1avn, 1cnw, 1cnx, 1cny); see Figure 3 in which the 1cny ligand protrudes out of the receptor structure. In some cases, such as 1a07, the protruding ligand interacts with a second unit of a dimer structure [75], which is not present in the LPDB complex [74]. In other cases the ligand may be interacting with neighboring molecules within the crystal unit cell.

### Comparison with GLIDE

To compare our results with a state of the art ligand docking program, we ran the Glide docking program (standard precision, version 50207) [76,77] on our test set using the score in place option. Of the 247 complexes we ran through our protocol, we were successful in running 227 complexes in Glide. Twelve cases failed to run because of the large size of the ligand, for example 1fkn (125 atoms), 1hhk (156 atoms), and 3er5 (185 atoms), due to a limit of 35 rotatable bonds in the ligands that can be docked by Glide. Another two cases (1b38, 1b39) failed due to the presence of a magnesium ion. The complexes eliminated from the original dataset include six histocompatibility antigen receptors, four endotheapepepsin receptors, two cell division protein kinase receptors and eight other receptor types.

Table 6 shows results for the 227 cases, for Glide and the “unbiased ICDA” and default PLOP protocols. Overall, the results show that the MM-GBSA scoring procedure is only competitive with Glide scoring when the ICDA receptor preparation procedure is applied. Docking and scoring functions such as those used in Glide have been extensively parameterized and validated in a variety of tests, including the ability to reproduce ligand conformations in crystal structures. Many of the proteins in our test set (106 out of the 227 complexes described here) were used in Glide parameterization [78]. In light of this, the fact that the force field based scoring works nearly as well, with no parameterization at all, is an achievement, and lends further evidence to support the potential utility of MM/GBSA scoring functions for inhibitor discovery. Nonetheless, the gap in performance probably does point to some limitations of the physics-based scoring function, most likely the implicit solvent model, for example in treating solvation in cavities [78]. It is also clear that obtaining correct protonation states and hydrogen atom placements is absolutely critical for good performance with the MM/GBSA scoring function. Finally, we note that MM/GBSA scoring is slower than Glide and other empirical scoring functions. We do not anticipate it being used for very large scale *in silico* screening (except as a secondary “rescoring” protocol [16]), but the speed is appropriate for, e.g., congeneric series of hundreds or thousands of compounds.

## IV. Conclusion

We have demonstrated that optimizing protonation states and the positions of polar hydrogens in binding sites can play a significant role in selecting native-like ligand poses. We used a scoring function based on a molecular mechanics force field (OPLS) and continuum solvent (GB) for selecting poses, lending further evidence that physics-based scoring functions of this type can be usefully employed in docking protocols [28–29]. We believe that it will also be possible to incorporate the optimization of protonation states into

protocols for “induced fit” docking that treat conformational changes in the protein due to ligand binding [44,79–81]. Receptor preparation can in principle also be conducted in conjunction with treatment of the ligand protonation states using a program such as Epik [82] which uses empirical pKa prediction to generate possible protonation states of ligands under specified conditions. Finally, future studies will examine the application of the protocol to virtual screening and ranking ligands in congeneric series [83].

## Acknowledgments

We thank Dr. Kai Zhu for help with the ICDA algorithm. This work was supported in part by the Sandler Program in Basic Sciences, and NIH grants GM071790 and AI035707 (to MPJ). MPJ is a consultant to Schrodinger LLC.

## References

1. Chen H, Lyne PD, Giordanetto P, Lovell T, Li J. On Evaluating Molecular-Docking Methods for Pose Prediction and Enrichment Factors. *J Chem Inf Model*. 2006; 46(1):401–415. [PubMed: 16426074]
2. Carlsson J, Boukharta L, Aqvist J. Combining docking, molecular dynamics and the linear interaction energy method to predict binding modes and affinities for non-nucleoside inhibitors to HIV-1 reverse transcriptase. *J Med Chem*. 2008; 51(9):2648–2656. [PubMed: 18410085]
3. Enyedy IJ, Egan WJ. Can we use docking and scoring for hit-to-lead optimization? *J Comput Aided Mol Des*. 2008; 22:161–168. [PubMed: 18183356]
4. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes. *J Med Chem*. 2006; 49(21):6177–6196. [PubMed: 17034125]
5. Kolb P, Huang D, Dey F, Caflisch A. Discovery of kinase inhibitors by high-throughput docking and scoring based on a transferable linear interaction energy model. *J Med Chem*. 2008; 51(5): 1179–1188. [PubMed: 18271520]
6. Leach AR, Shoichet BK, Peishoff CE. Prediction of Protein-Ligand Interactions. Docking and Scoring: Successes and Gaps. *J Med Chem*. 2006; 49(20):5851–5855. [PubMed: 17004700]
7. Pham TA, Jain AM. Parameter Estimation for Scoring Protein-Ligand Interactions Using Negative Training Data. *J Med Chem*. 2006; 49(20):5856–5868. [PubMed: 17004701]
8. Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *British J Pharm*. 2008; 53:S7–S26.
9. Sousa SF, Fernandes PA, Ramos MJ. Protein-ligand docking: Current status and future challenges. *Proteins*. 2006; 65(1):15–26. [PubMed: 16862531]
10. Warren GL, Webster AC, Capelli AM, Clarke B, LaLonde J, Lamber MH, Peishoff CE, Head MS. A Critical Assessment of docking programs and scoring functions. *J Med Chem*. 2006; 49(20): 5912–5931. [PubMed: 17004707]
11. Amini A, Shrimpton PJ, Muggleton SH, Sternberg MJE. A general approach for developing system-specific functions to score protein-ligand docked complexes using support vector inductive logic programming. *Proteins*. 2007; 69(4):823–831. [PubMed: 17910057]
12. Graves AP, Shivakumar DM, Boyce SE. Rescoring docking hit lists for model cavity sites: Predictions and experimental testing. *J Mol Biol*. 2008; 377(3):914–934. [PubMed: 18280498]
13. Lee J, Seok CA. A statistical rescoring scheme for protein-ligand docking: Consideration of entropic effect. *Proteins*. 2008; 70(3):1074–1083. [PubMed: 18076034]
14. Lee MR, Sun Y. Improving docking accuracy through molecular mechanics Generalized Born optimization and scoring. *J Chem Theory and Comput*. 2007; 3:1106–1119.
15. Martin O, Schomburg D. Efficient comprehensive scoring of docked protein complexes using probabilistic support vector machines. *Proteins*. 2008; 70(4):1367–1378. [PubMed: 17894343]

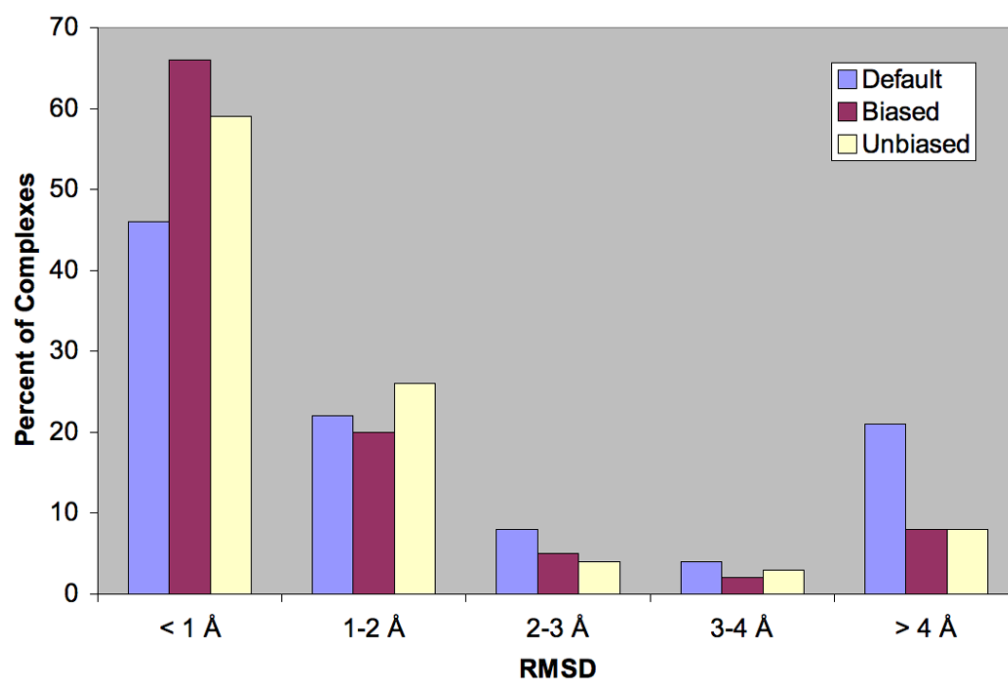


16. Huang N, Kalyanaraman C, Irwin JJ, Jacobson MP. Physics-Based Scoring of Protein-Ligand Complexes: Enrichment of Known Inhibitors in Large-Scale Virtual Screening. *J Chem Inf Model*. 2006; 46(1):243–253. [PubMed: 16426060]
17. Ruvinsky A. Role of Binding Entropy in the Refinement of Protein-Ligand Docking Predictions: Analysis Based on the Use of 11 Scoring Functions. *J Comput Chem*. 2007; 28:1364–1372. [PubMed: 17342720]
18. Spyraakis F, Amadasi A, Fornabai M, Abraham DJ, Mozzarelli A, Kellogg GE, Cozzini P. The consequences of scoring docked ligand conformations using free energy correlations. *Eur J Med Chem*. 2007; 42:921–933. [PubMed: 17346861]
19. Huang SY, Zou X. An Iterative Knowledge-Based Scoring Function to Predict Protein-Ligand Interactions: II. Validation of the Scoring Function. *J Comput Chem*. 2006; 27:1876–1882. [PubMed: 16983671]
20. Seebeck B, Reulecke I, Kamper A, Rarey M. Modeling of metal interaction geometries for protein-ligand docking. *Proteins*. 2008; 71(3):1237–1254. [PubMed: 18041759]
21. Gorelik B, Goldblum A. High quality binding modes in docking ligands to proteins. *Proteins*. 2008; 71(3):1373–1386. [PubMed: 18058908]
22. Artemenko M. Distance Dependent Scoring Function for Describing Protein-Ligand Intermolecular Interactions. *J Chem Inf Model*. 2008; 48(3):569–574. [PubMed: 18290639]
23. Oda A, Tsuchida K, Takakura T, Yamaotsu N, Hirono S. Comparison of Consensus Scoring Strategies for Evaluating Computational Models of Protein-Ligand Complexes. *J Chem Inf Model*. 2006; 46(1):380–391. [PubMed: 16426072]
24. Baber JC, Shirley WA, Gao Y, Feher M. The Use of Consensus Scoring in Ligand-Based Virtual Screening. *J Chem Inf Model*. 2006; 46(1):277–288. [PubMed: 16426063]
25. Renner S, Derksen S, Radestock S, Mörch F. Maximum Common Binding Modes (MCBM): Consensus Docking Scoring Using Multiple Ligand Information and Interaction Fingerprints. *J Chem Inf Model*. 2008; 48(2):319–332. [PubMed: 18211051]
26. Teramoto R, Fukunishi H, Teramoto R. Consensus Scoring with Feature Selection for Structure-Based Virtual Screening. *J Chem Inf Model*. 2008; 48(2):288–295. [PubMed: 18229906]
27. Fischer B, Fukuzawa K, Wenzel W. Receptor-specific scoring functions derived from quantum chemical models improve affinity estimates for in-silico drug discovery. *Proteins*. 2008; 70(4):1264–1273. [PubMed: 17876816]
28. Huang N, Kalyanaraman C, Bernacki K, Jacobson MP. Molecular mechanics methods for predicting protein-ligand binding. *Phys Chem Chem Phys*. 2006; 8(44):5166–5177. [PubMed: 17203140]
29. Huang N, Jacobson MP. Physics-Based Methods for Studying Protein-Ligand Interactions. *Current Opinion in Drug Discovery and Development*. 2007; 10:325–331.
30. Bottegoni G, Kufareva I, Totrov M, Abagyan R. A new method for ligand docking to flexible receptors by dual alanine scanning and refinement (SCARE). *J Comput Aided Mol Des*. 2008; 22:311–325. [PubMed: 18273556]
31. Ferrara P, Gohlke G, Price DJ, Klebe G, Brooks CL. Assessing Scoring Functions for Protein-Ligand Interactions. *J Med Chem*. 2004; 47(12):3032–3047. [PubMed: 15163185]
32. Sham YY, Chu ZT, Tao H, Warshel A. Examining methods for calculations of binding free energies: LRA, LIE, PDL=LR, and PDL/S\_LRA calculations of ligands binding to an HIV protease. *Proteins*. 2000; 39(4):393–407. [PubMed: 10813821]
33. Lee FS, Chu ZT, Bolger MB, Warshel A. Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. *Prot Eng*. 1992; 5(3):215–228.
34. Muegge I, Tao H, Warshel A. A fast estimate of electrostatic group contributions to the free energy of protein-inhibitor binding. *Prot Eng*. 1997; 10(12):1363–1372.
35. Word J, Lovell S, Richardson J, Richardson D. Asparagine and Glutamine: Using Hydrogen Atom Contacts in the Choice of Side-chain Amide Orientation. *J Mol Biol*. 1999; 285(4):1735–1747. [PubMed: 9917408]
36. Hooft R, Sander C, Vriend G. Positioning Hydrogen Atoms by Optimizing Hydrogen Bond Networks in Protein Structures. *Proteins*. 1996; 26(4):363–376. [PubMed: 8990493]

37. Vriend G. WHAT IF: A molecular modeling and drug design program. *J Mol Graph.* 1990; 8:52–56. [PubMed: 2268628]
38. Yang A, Honig B. Structural Origins of pH and Ionic Strength Effects on Protein Stability: Acid Denaturation of Sperm Whale Apomyoglobin. *J Mol Biol.* 1994; 237(5):602–614. [PubMed: 8158640]
39. Bashford D, Karplus M. pKa's of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model. *Biochemistry.* 1990; 29:10219–10225. [PubMed: 2271649]
40. Georgescu RE, Alexov EG, Gunner MR. Combining conformational flexibility and continuum electrostatics for calculating pKa's in proteins. *Biophys J.* 2002; 83:1731–1748. [PubMed: 12324397]
41. Warshel A. Calculations of enzymatic reactions: calculations of pKa, proton transfer reactions, and general acid catalysis reactions in enzymes. *Biochemistry.* 1981; 20(11):3167–3177. [PubMed: 7248277]
42. Sham YY, Chu ZT, Warshel A. Consistent Calculations of pKa's of Ionizable Residues in Proteins: Semi-microscopic and Microscopic Approaches. *J Phys Chem A.* 1997; 101(22):4458–4472.
43. Li X, Jacobson MP, Zhu K, Zhao S, Friesner RA. Assignment of Polar States for Protein Amino Acid residues using an Interaction Cluster decomposition Algorithm and its application to High Resolution Protein Structure Modeling. *Proteins.* 2007; 66(4):824–837. [PubMed: 17154422]
44. Naim M, Bhat S, Rankin KN, Dennis S, Chowdhury SF, Siddiqi I, Drabik P, Sulea T, Bayly CI, Jakalian A, Purisima EO. Solvated Interaction Energy (SIE) for Scoring Protein-Ligand Binding Affinities. 1. Exploring the Parameter Space. *J Chem Inf Model.* 2007; 47(1):122–133. [PubMed: 17238257]
45. Fischer B, Basili S, Merlitz H, Wenzel W. Accuracy of Binding Mode Prediction With a Cascadic Stochastic Tunneling Method. *Proteins.* 2007; 68(1):195–204. [PubMed: 17427957]
46. Jain AN. Bias, reporting, and sharing: computational evaluations of docking methods. *J Comput Aided Mol Des.* 2008; 22:201–212. [PubMed: 18075713]
47. Czodrowski P, Sotriffer CA, Klebe G. Protonation Changes upon Ligand Binding to Trypsin and Thrombin: Structural Interpretation Based on pK<sub>a</sub> Calculations and ITC Experiments. *J Mol Biol.* 2007; 367(5):1347–1356. [PubMed: 17316681]
48. Mason AC, Jensen JH. Protein–protein binding is often associated with changes in protonation state. *Proteins.* 2008; 71(1):81–91. [PubMed: 17932920]
49. Ohno K, Sakurai M. Linear-scaling molecular orbital calculations for the pKa values of ionizable residues in proteins. *J Comp Chem.* 2006; 27(7):906–916. [PubMed: 16550537]
50. Parthasarathi R, Padmanabhan J, Elango M, Chitra K, Subramanian V, Chattaraj PK. pK<sub>a</sub> Prediction Using Group Philicity. *J Phys Chem A.* 2006; 110(20):6540–6544. [PubMed: 16706412]
51. Tynan-Connolly BM, Nielsen JE. Redesigning protein pK(a) values. *Protein Sci.* 2007; 16(2):239–249. [PubMed: 17189477]
52. Nabuurs SB, Wagener M, De Vlieg J. A flexible approach to induced fit docking. *J Med Chem.* 2007; 50(26):6507–6518. [PubMed: 18031000]
53. Jorgensen WL, Maxwell DS, Tirado-Rives J. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J Am Chem Soc.* 1996; 118:11225–11236.
54. Kaminski GA, Friesner RA, Tirado-Rives J, Jorgensen WL. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison with Accurate Quantum Chemical Calculations on Peptides. *J Phys Chem B.* 2001; 105:6474–6487.
55. Ghosh A, Rapp CS, Friesner RA. Generalized Born model based on a surface integral formulation. *J Phys Chem B.* 1998; 102:10983–10990.
56. Zhou Z, Bates M, Madura JD. Structure Modeling, Ligand Binding, and Binding Affinity Calculation (LR-MM-PBSA) of Human Heparanase for Inhibition and Drug Design. *Proteins.* 2006; 65(3):580–592. [PubMed: 16972282]
57. Ferrari AM, Degliesposti G, Sgobba M, Rastelli G. Validation of an automated procedure for the prediction of relative free energies of binding on a set of aldose reductase inhibitors Theoretical

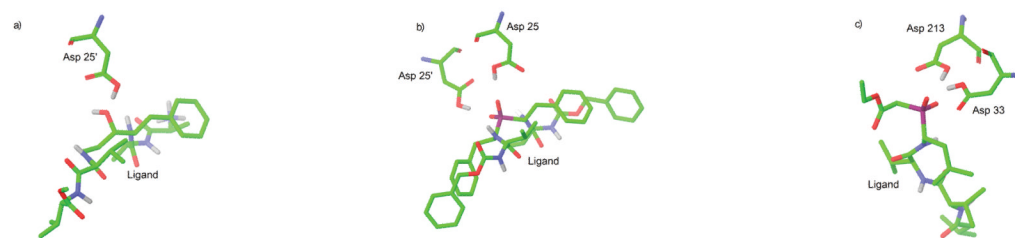
- calculation of the binding free energies for pyruvate dehydrogenase E1 binding with ligands. *Bioorg Med Chem*. 2007; 15:7865–7877. [PubMed: 17870536]
58. Xiong Y, Li Y, He H, Zhan CG. Theoretical calculation of the binding free energies for pyruvate dehydrogenase E1 binding with ligands. *Bioorg Med Chem Lett*. 2007; 17:5186–5190. [PubMed: 17644334]
  59. Liu HY, Zou X. Electrostatics of Ligand Binding: Parameterization of the Generalized Born Model and Comparison with the Poisson-Boltzmann Approach. *J Phys Chem B*. 2006; 110:9304–9313. [PubMed: 16671749]
  60. Graves AP, Brenk R, Shoichet BK. Decoys for Docking. *J Med Chem*. 2005; 48(11):3714–3728. [PubMed: 15916423]
  61. Roche O, Kiyama R, Brooks CL. Ligand-Protein DataBase: Linking Protein-Ligand Complex Structures to Binding Data. *J Med Chem*. 2001; 44(22):3592–3598. [PubMed: 11606123]
  62. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Databank. *Nucleic Acids Res*. 2000; 28:235–242. [PubMed: 10592235]
  63. Jacobson MP, Kaminski GA, Friesner RA, Rapp CS. Force field validation using protein side chain prediction. *J Phys Chem B*. 2002; 106:11673–11680.
  64. Jacobson MP, Pincus DL, Rapp CS, Day TJ, Honig B, Shaw DE, Friesner RA. A hierarchical approach to all-atom protein loop prediction. *Proteins*. 2004; 55(2):351–367. [PubMed: 15048827]
  65. Li X, Jacobson MP, Friesner RA. High-resolution prediction of protein helix positions and orientations. *Proteins*. 2004; 55(2):368–382. [PubMed: 15048828]
  66. Momany FA, Rone R. Validation of the General Purpose Quant 3.2/CHARMm force field. *J Comput Chem*. 1992; 13:888–900.
  67. Zhu K, Shirts MR, Friesner RA, Jacobson MP. Multiscale optimization of a truncated Newton minimizer and applications to proteins and protein-ligand complexes. *J Chem Theory and Comput*. 2007; 3:640–648.
  68. Wittayanarakul K, Hannongbua S, Feig M. Accurate Prediction of Protonation State as a Prerequisite for Reliable MM-PB(GB)SA Binding Free Energy Calculations of HIV-1 Protease Inhibitors. *J Comput Chem*. 2008; 29:673–685. [PubMed: 17849388]
  69. Coates L, Tuan HF, Tomanicek S, Kovalevsky A, Mustyakimov M, Erskine P, Cooper J. The Catalytic Mechanism of an Aspartic Proteinase Explored with Neutron and X-ray Diffraction. *J Am Chem Soc*. 2008; 130:7235–7237. [PubMed: 18479128]
  70. Kovalevsky AY, Chumanevich AA, Liu F, Luis JM, Weber IT. Caught in the Act: The 1.5 Å Resolution Crystal Structures of the HIV-1 Protease and the I54V Mutant Reveal a Tetrahedral Reaction Intermediate. *Biochemistry*. 2008; 46:14854–14864. [PubMed: 18052235]
  71. Bas DC, Rogers DM, Jensen JH. Very fast prediction and rationalization of pK<sub>a</sub> values for protein-ligand complexes. *Epub Proteins*. May 22.2008
  72. Velazquez-Campoy A, Luque I, Todd MJ, Milutinovich M, Kiso Y, Freire E. Thermodynamic dissection of the binding energetics of KNI-272, a potent HIV-1 protease inhibitor. *Protein Sci*. 2000; 9:1801–1809. [PubMed: 11045625]
  73. Vidossich P, Carloni P. Binding of Phosphinate and Phosphonate Inhibitors to Aspartic Proteases: A First-Principles Study. *J Phys Chem B*. 2006; 110:1437–1442. [PubMed: 16471695]
  74. <http://www.lpdb.chem.lsa.umich.edu>
  75. <http://www.pdb.org>
  76. Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, Repasky MP, Knoll EH, Shelley M, Perry JK, Shaw DE, Francis P, Shenkin PS. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy. *J Med Chem*. 2004; 47(7):1739–1749. [PubMed: 15027865]
  77. Halgren TA, Murphy RB, Friesner RA, Beard HG, Frye LL, Pollard WT, Banks JL. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *J Med Chem*. 2004; 47(7):1750–1759. [PubMed: 15027866]
  78. Friesner RA, Murphy RB, Repasky MP, Frye LL, Greenwood JR, Halgren TA, Sanschagrin PC, Mainz DT. Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein Ligand Complexes. *J Med Chem*. 2006; 49(21):6177–6196. [PubMed: 17034125]

79. Wong CF. Flexible ligand-flexible protein docking in protein kinase systems. *Biochem Biophys Acta*. 2008; 1784(1):244–251. [PubMed: 17996742]
80. Zhao Y, Sanner MF. FLIPDock: Docking flexible ligands into flexible receptors. *Proteins*. 2007; 68(3):726–737. [PubMed: 17523154]
81. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R. Novel procedure for modeling ligand/receptor induced fit effects. *J Med Chem*. 2006; 49(2):534–553. [PubMed: 16420040]
82. Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: a software program for pKa prediction and protonation state generation for drug-like molecules. *J Comput Aided Mol Des*. 2007; 21:681–691. [PubMed: 17899391]
83. Lyne PD, Lamb ML, Saeh JC. Accurate Prediction of the Relative Potencies of Members of a Series of Kinase Inhibitors Using Molecular Docking and MM-GBSA Scoring. *J Med Chem*. 2006; 49(16):4805–4808. [PubMed: 16884290]



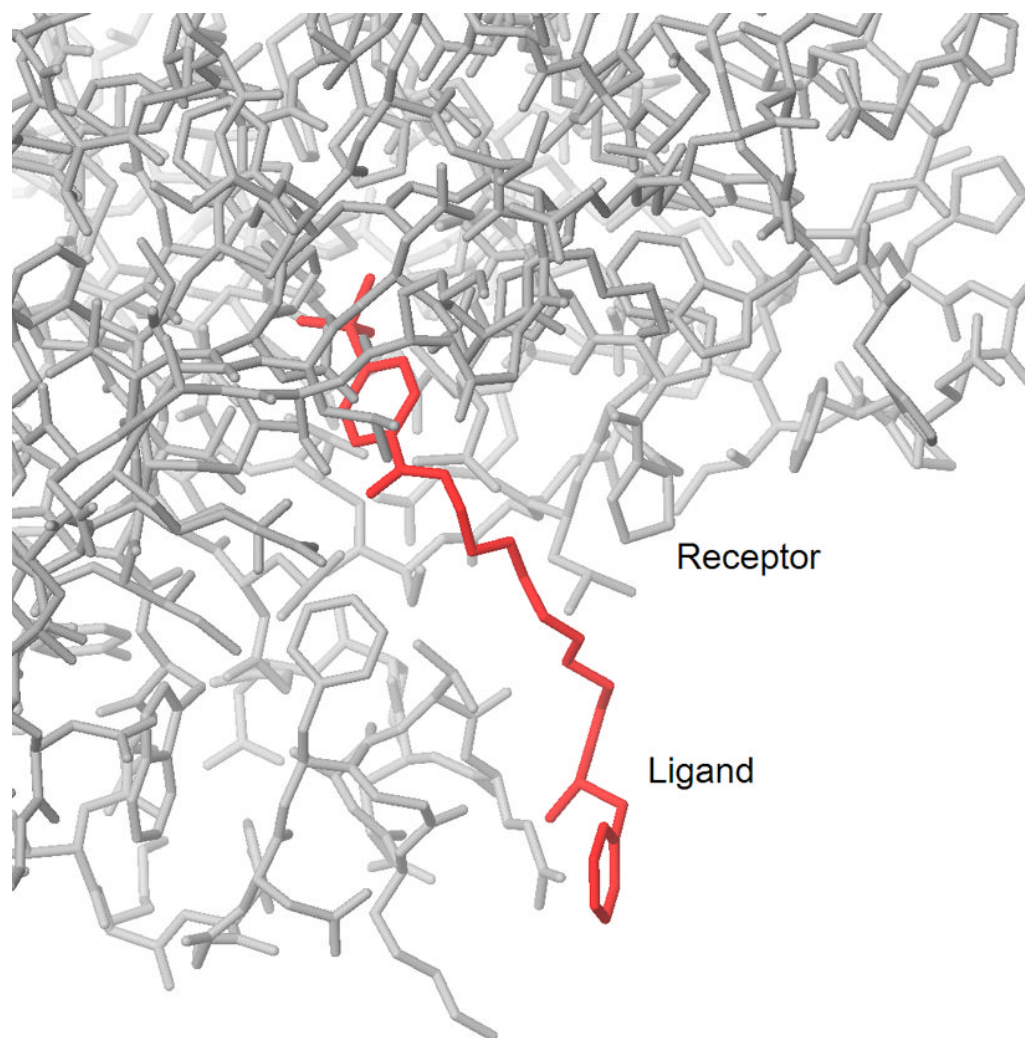
**Figure 1.** Percent of complexes with given RMSD for selected pose using default, “biased” ICDA, and “unbiased” ICDA protocols.





**Figure 2.**

Ligand surrounded by receptor residues with ICDA-predicted changes in protonation state for aspartyl proteases a) 1aaq (HIV protease), in which only Asp 25' is protonated; b) 1hos (HIV protease) in which both Asp 25 and Asp 25' are protonated; and c) penicillopepsin, 1ppk, showing a diprotonated state characteristic of phosphinate and phosphonate inhibitors. CPK color scheme is used with green carbon atoms and grey polar hydrogen atoms.



**Figure 3.**  
In PDB structure 1cny, the ligand (red) protrudes out of the receptor (grey) into solvent.

**Table 1**

Summary of results of using ICDA receptor preparation on pose selection, compared to a default protocol that does not perform the preparation. Columns list average RMSD over all complexes, % of complexes with  $\text{RMSD} \leq 2 \text{ \AA}$ , and % complexes with  $\text{RMSD} \leq 1 \text{ \AA}$  for the default, “biased” ICDA and “unbiased” ICDA protocols.

Protocol	Default MM-GBSA	“Biased” ICDA	“Unbiased” ICDA
Average RMSD ( $\text{\AA}$ )	2.7	1.4	1.5
% $\leq 2 \text{ \AA}$	68%	86%	85%
% $\leq 1 \text{ \AA}$	46%	66%	59%

**Table 2**

Cases in which application of ICDA affects pose selection. Columns list PDB id, protein name, RMSD of selected pose using the default protocol, RMSD of selected pose using the “unbiased” ICDA protocol, and protonation state changes predicted by ICDA. The positively charged form of His is designated as “+”, while the two tautomeric neutral forms are designated by which nitrogen is protonated,  $\delta$  or  $\epsilon$ .

PDB	Protein	RMSD of selected pose using ...		Titratable groups changing protonation state using ICDA		
		Default protocol	“Unbiased” ICDA	Protonated Asp side chains	Protonated Glu side chains	His states
1A9M	HIV-1 Protease	9.5	1.1	25		48 (+) 48' (+)
1AAQ	HIV-1 Protease	8.5	0.5	25'		
1BDR	HIV-1 Protease	12.0	0.3	25		
1G35	HIV-1 Protease	13.1	1.2	25'		
1HBV	HIV-1 Protease	3.3	0.7	25		
1HII	HIV-2 Protease	13.9	0.4	25		
1HOS	HIV-1 Protease	11.3	0.8	25 25'		
1HPX	HIV-1 Protease	2.9	0.5	25		
1HSG	HIV-1 Protease	2.9	0.4	25'		
1HVJ	HIV-1 Protease	2.3	0.3	25'		
1IVP	HIV-2 Protease	22.5	0.4	25		
1EED	Endothiapepsin	10.0	1.1	12 30 32 37		
1EPP	Endothiapepsin	8.5	0.9	12 30 32		
2ER7	Endothiapepsin	5.9	1.3	12 30 32 37		
3ER3	Endothiapepsin	6.4	3.0	12 30 32 215		
3ER5	Endothiapepsin	10.6	1.9	12 30 32 37		
4ER1	Endothiapepsin	10.1	0.7	12 30 32 37		
4ER2	Endothiapepsin	8.5	2.5	12 30 32		
5ER2	Endothiapepsin	4.0	1.1	12 30 32 215		
1ENT	Endothiapepsin	9.6	0.6	12 30 32 37 215		
1APU	Penicillopepsin	7.8	0.8	33	16	
1APV	Penicillopepsin	8.7	0.3	33	16	
1APW	Penicillopepsin	8.9	0.7	33		
2WEA	Penicillopepsin	5.5	0.8	33 213	15 16	

PDB	Protein	RMSD of selected pose using ...		Titratable groups changing protonation state using ICDA		
		Default protocol	"Unbiased" ICDA	Protonated Asp side chains	Protonated Glu side chains	His states
2WEB	Penicillopepsin	9.4	1.0	33 115 213 300		
1PPK	Penicillopepsin	8.6	0.3	33 213	16	
1PPL	Penicillopepsin	8.1	0.4	33 213	16	
1PPM	Penicillopepsin	9.1	2.3	33 115 213		
1AFL	Ribonuclease A	11.2	1.3			12 (+) 119 (+)
1AFK	Ribonuclease A	10.7	1.3			12 (+) 119 (+)
1AET	Cytochrome C Peroxidase	2.9	0.6	135		175 (δ)
1B38	Cell Division Protein Kinase 2	12.0	0.9	145		84 (δ) 125 (+)
1B39	Cell Division Protein Kinase 2	10.5	0.4			84 (δ) 161 (δ) 125 (+)
6CPA	Carboxypeptidase A	3.6	0.5		270	69 (δ) 196 (δ)
1ATL	Atrolysin C	4.0	1.0		143	146 (δ) 152 (δ) 129 (ε) 142 (+)
1SRE	Streptavidin	7.6	0.6		none	
1STP	Streptavidin	3.9	0.5		none	
1BIL	Renin	12.1	0.4	215		287 (+)
1ETR	Epsilon-Thrombin	5.6	1.2			57 (+)
1FKN	Menapsin 2	11.5	1.2	38		
1PGP	6-Phosphogluconate Dehydrogenase	5.3	1.5		190	186 (+)



**Table 3**

Energy difference between selected pose and native pose, for cases from Table 2, using “unbiased” ICDA and default protocols. Columns list PDB id, RMSD of native pose, RMSD of selected pose, and energy difference (selected pose – native), for the two protocols. Units of energy are kcal/mol.

PDB	“Unbiased ICDA”			Default		
	Native RMSD	Selected Pose RMSD	$\Delta$ Energy	Native RMSD	Selected Pose RMSD	$\Delta$ Energy
1A9M	0.7	1.1	9.3	0.6	9.5	-14.7
1AAQ	0.5	0.5	-0.1	0.5	8.5	-23.4
1BDR	0.3	0.3	-0.4	0.3	12.0	-33
1G35	0.6	1.2	-2.6	0.5	13.1	-4.8
1HBV	0.4	0.7	-19.7	0.9	3.3	-18.6
1HII	0.4	0.4	0	0.4	13.9	-24
1HOS	0.9	0.8	1.9	1.8	11.3	-43.4
1HPX	0.5	0.5	-0.5	0.6	2.9	-8.7
1HSG	0.2	0.4	-20.3	0.7	2.9	-19.1
1HVJ	0.3	0.3	-0.1	0.3	2.3	-12
1IVP	0.6	0.4	11.6	0.5	22.5	-19.2
1EED	0.8	1.1	-2.4	0.6	10.0	-51.6
1EPP	1.1	0.9	-47	8.5	1.6	36.1
2ER7	0.8	1.3	18.9	1.0	5.0	-6.7
3ER3	0.5	3.0	-16.3	0.6	6.4	-102.5
3ER5	1.9	1.9	-9.2	1.9	10.6	-47.7
4ER1	0.5	0.7	-8.6	0.4	10.1	-69
4ER2	0.5	2.5	1.3	0.5	8.5	-29.9
5ER2	1.6	1.1	-6	1.5	4.0	-64.1
1ENT	0.5	0.6	2.3	2.7	9.6	-90.9
1APU	0.6	0.8	0	0.5	8.0	-14
1APV	0.2	0.3	-0.5	0.2	8.7	-9.1
1APW	0.3	0.7	-1.9	0.3	8.9	-5
2WEA	0.6	0.8	-1.1	0.7	5.5	-83.9
2WEB	0.7	1.0	-1.6	0.5	9.4	-86.8

PDB	“Unbiased ICDA”			Default		
	Native RMSD	Selected Pose RMSD	$\Delta$ Energy	Native RMSD	Selected Pose RMSD	$\Delta$ Energy
1PPK	0.4	0.3	-0.5	0.2	8.6	-134.2
1PPL	0.4	0.4	0.4	0.5	8.1	-114.8
1PPM	0.5	2.3	5.7	1.0	9.1	-58.6
1AFL	0.7	1.3	-87.3	0.6	11.2	-46.5
1AFK	0.5	1.3	-3.4	0.5	10.7	-35.1
1AET	0.6	0.6	-0.8	0.9	2.9	-16.4
1B38	0.2	0.9	-21	0.4	12.1	-33
1B39	0.4	0.4	-51	0.3	10.5	-35

**Table 4**

Cases from Table 2 in which side chain orientations are modified by ICDA. Columns list PDB id and residues with changed orientations.

PDB	Residues with changes in side chain orientation using ICDA
1A9M	Gln 58' His 48'
3ER5	Asn 129
4ER2	Gln 187
3ER3	Asn 129
5ER2	Asn 129

**Table 5**

Effect of ions and cofactors on RMSD. Columns list cofactor or ion, number of complexes that include this ion or cofactor within 3.6 Å of the ligand, and average RMSD for these complexes. A complex may include more than one ion or cofactor. RMSD values are for “unbiased” ICDA.

Cofactor/Ion	Number of Complexes	Average RMSD (Å)
none	181	1.5
Zn	31	1.8
Fe	13	1.4
Mg or Mn	6	1.5
HEME	10	0.8
NAD, NADP	7	0.7
Dimethylformamide (DMF)	3	0.5

**Table 6**

Comparison of results using “unbiased” ICDA MM-GBSA protocol, default MM-GBSA protocol, and standard precision Glide for pose selection. The ICDA and default MM-GBSA results differ slightly from those in Table 1 because the results exclude 20 cases for which Glide obtained no result. Columns list average RMSD over all complexes, % of complexes with  $\text{RMSD} \leq 2 \text{ \AA}$ , and % complexes with  $\text{RMSD} \leq 1 \text{ \AA}$  for “unbiased” ICDA MM-GBSA, default MM-GBSA, and Glide protocols.

Protocol	“Unbiased ICDA”	Default MM-GBSA	Glide
Average RMSD ( $\text{\AA}$ )	1.5	2.5	1.2
% $\leq 2 \text{ \AA}$	85%	68%	87%
% $\leq 1 \text{ \AA}$	62%	47%	59%