# Free Energy Component Analysis for Drug Design: A Case Study of HIV-1 Protease−Inhibitor Binding

Parul Kalra,[†] T. Vasisht Reddy,[‡] and B. Jayaram*,[†]

*Department of Chemistry and Department of Biochemical Engineering and Biotechnology, Indian Institute of Technolgy, Hauz Khas, New Delhi 110 016, India*

*Received April 20, 2001*

A theoretically rigorous and computationally tractable methodology for the prediction of the free energies of binding of protein−ligand complexes is presented. The method formulated involves developing molecular dynamics trajectories of the enzyme, the inhibitor, and the complex, followed by a free energy component analysis that conveys information on the physicochemical forces driving the protein−ligand complex formation and enables an elucidation of drug design principles for a given receptor from a thermodynamic perspective. The complexes of HIV-1 protease with two peptidomimetic inhibitors were taken as illustrative cases. Four-nanosecond-level all-atom molecular dynamics simulations using explicit solvent without any restraints were carried out on the protease−inhibitor complexes and the free proteases, and the trajectories were analyzed via a thermodynamic cycle to calculate the binding free energies. The computed free energies were seen to be in good accord with the reported data. It was noted that the net van der Waals and hydrophobic contributions were favorable to binding while the net electrostatics, entropies, and adaptation expense were unfavorable in these protease−inhibitor complexes. The hydrogen bond between the $CH_2OH$ group of the inhibitor at the scissile position and the catalytic aspartate was found to be favorable to binding. Various implicit solvent models were also considered and their shortcomings discussed. In addition, some plausible modifications to the inhibitor residues were attempted, which led to better binding affinities. The generality of the method and the transferability of the protocol with essentially no changes to any other protein−ligand system are emphasized.

## I. Introduction

The ability to accurately predict binding affinities from structural considerations is an essential prerequisite to developing a thermodynamically sound strategy for drug design. Molecular simulations employing free energy perturbation[1−5] and thermodynamic integration[6−8] methods developed almost a decade and a half ago within the framework of statistical mechanics constitute some of the best options from the standpoint of theoretical rigor for evaluating receptor−ligand binding free energies. However, computational demands and issues related to sampling and convergence arising in these methods prevent them from being widely and routinely used in structure-based drug design.[9−11] Furthermore, the statistical noise associated with a series of simulations required to elicit the all-important single quantity, namely, the binding free energy, in these methods diminishes somewhat their discriminatory power in making a judicious choice among alternative candidate molecules. Thus, the quest for computationally expeditious yet reliable methods (to be used in structure-based drug design) for obtaining free energy estimates from a single simulation on the receptor−ligand complex or directly from the crystal structure of the complex continues. In this context, the master equation approach formulated by Ajay and Murcko,[12] which assumes additivity of the free energy components and the hierarchical improvements feasible within the free energy component structure,[13,14] offers an attractive alternative. An implicit further expectation from theory beyond accurate determinations of binding free energies is that some atom/functional group based general principles of protein−drug recognition could be catalogued eventually at least for a specific protein target if not in the general case. It is here that the free energy component analysis could prove to be powerful, limited only by the accuracy of the force field parameters used in computing energies. Recent improvements in the description of intermolecular interactions using the second-generation molecular mechanics (MM) force fields[15,16] and the development of a new methodology for obtaining the free energies of solvation accurately using the "generalized Born solvent accessibility" (GBSA) model[17−22] together provide a state of the art method (MMGBSA) for rapid free energy evaluations for binding processes involving noncovalent macromolecular associations. Furthermore, any limitations arising due to additivity approximation of the free energy components are overcome to a large extent by an appropriate choice of the solvation model compatible with the force field. In this article, we illustrate this methodology for calculating the free energies of protein−drug binding using HIV-1 protease−inhibitor complexes as a test case.[23,24] HIV-1 protease has been chosen because of the importance of developing effective drugs against HIV and also because of the wealth of activity data and structural information on this system. Active research is being done on the design of newer and better inhibitors for HIV-1 protease to combat AIDS.[25−29] We

* To whom correspondence should be addressed. Phone: +91-11-6591505. Fax: +91-11-6862037. E-mail: bjayaram@chemistry.iitd.ac.in.
[†] Department of Chemistry.
[‡] Department of Biochemical Engineering and Biotechnology.

10.1021/jm010175z CCC: $20.00 © 2001 American Chemical Society
Published on Web 11/10/2001

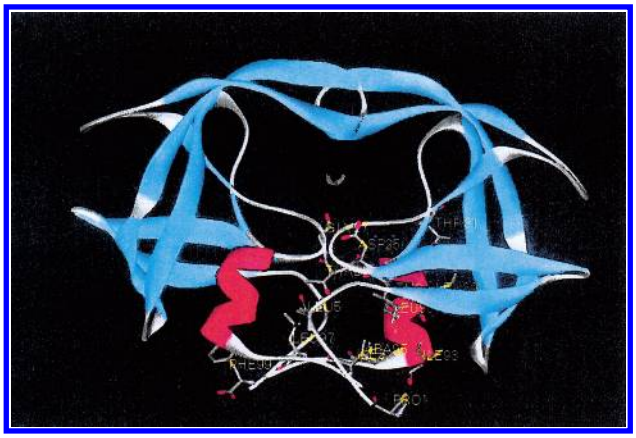

**Figure 1.** Schematic of HIV-1 protease indicating some important residues from one monomer.

describe herein a theoretical "component analysis" of the standard free energy of binding for some HIV-1 protease-inhibitor complexes, paying particular attention to the merits and limitations of the binding affinity estimates and directions to take in further methodology improvements aimed at drug design.

The HIV-1 protease (PR) is a member of the well-characterized aspartic protease family, which also includes mammalian enzymes such as renin, pepsin, and chymosin. The function of HIV-1 protease is to cleave the translated viral gag-pol polyprotein into discrete components. Without protease activity, the viral particles remain noninfective, and this makes the protease an attractive candidate for therapeutic drug design against AIDS. The HIV-1 protease structure corresponds closely to the arrangement of the completely symmetrical dimer that has been proposed as the ancestor of the aspartic proteases.[30,31] The two PR subunits interact to form a symmetric dimer with an active site that is similar to the highly conserved active sites of monomeric cellular proteases. The conserved active-site residues Asp25, Thr26, Gly27 and Asp25′, Thr26′, Gly27′ form a symmetrical and highly hydrogen-bonded arrangement virtually identical to that described for pepsins (Figure 1). The two threonines are inaccessible to solvent and are hydrogen-bonded so that the $O^{\gamma}$ atoms bond to the main-chain N−H functional groups of the other subunit. The two aspartates lie roughly coplanar, with their inner carboxylate oxygens hydrogen-bonded to the N−H group of Gly-27 and Gly27′, respectively. An extensive hydrophobic core extends through the dimeric interface. Pro-1, Ile-3, and Leu-5 of the N-terminal domain, and Cys-95, Leu-97, and Phe-99 of the C-terminal domain, form one side of the central four-stranded $\beta$-pleated sheet and pack onto Leu-24 and Thr-26, adjacent to the catalytic residues and the hydrophobic residues of the helix (viz. Leu-89, Leu-90, and Ile-93). These residues of the core tend to be highly conserved among the retroviral proteases. The only truly buried polar residue apart from Thr-26 is Thr-31, which is also hydrogen-bonded to main-chain N−H and C═O functional groups in a fashion that is often important to protein structure.[30]

HIV-1 protease−ligand complexes share a number of structural features preserving a common ligand−protein interaction pattern. Hydrophobic amino acids are seen to be preferred at the P1−P1′ residues that flank the scissile peptide bond, aliphatic and Glu/Gln residues are often found at P2′, aromatic residues are almost never found at P3′, and small residues are preferred at P2. Several sequences contain an aromatic residue at P1 followed by a Pro at P1′ (Table 1 and Figure 2).[31] Studies with peptides corresponding to a target in the HIV-1 gag precursor have shown that a heptapeptide is cleaved by the HIV-1 PR.[28] Thus, the minimal target sequence for retroviral proteases should have four amino acids (P4−P1) corresponding to the side that is to become the new carboxyl terminus and three amino acids (P1′−P3′) corresponding to that which is to become the new amino terminus of the cleaved product. Binding of an HIV-1 peptide substrate,[33] Ser-Gln-Asn-Tyr-Pro-Ile-Val, which is cleaved efficiently between Tyr and Pro, shows that the two catalytic residues Asp25 and Asp25′ from both subunits in the dimer are near the scissile bond. It has been proposed that several potential hydrogen bond interactions occur between the main-chain atoms of the substrate and the main-chain atoms of Gly-51′, Asp-29′, Phe-53′, Gly-27′, Asp-25′, Asp-25, Phe-53, Gly-27, Asp-29, and Gly-51 (Figure 2). The position of the two flexible flaps gets altered in the presence of the substrate. This change probably enhances binding of the inhibitor by facilitating several hydrogen bond interactions; it also causes exclusion of water from the catalytic site.

**Table 1.** Some Substrates for HIV-1 Protease

PROCESSING SITES FOR HIV-1 PROTEASE

| Site | P4 | P3 | P2 | P1 | P1′ | P2′ | P3′ | P4′ |
|---|---|---|---|---|---|---|---|---|
| p1/p6 | —Pro | —Gly | —Asn | —Phe/ | Leu | —Gln | —Ser | — Arg — |
| PR/RT | —Thr | — Leu | —Asn | —Phe/ | Pro | —Ile | —Ser | — Pro — |

There are seven subsites known to be present in HIV-1 protease, and the corresponding residues present in these subsites[28] are shown in Table 2. The subsite S1 has space for a large hydrophobic residue. The Phe residue occurs at P1 in several HIV-1 cleavage sites, and this interacts favorably with Phe-53. A smaller hydrophobic side chain is preferred at S1′, and Pro is often present in this location. The substrate residue P2′ tends to be polar because of a flanking Asp-30. A smaller residue is predicted to bind in subsite S4, and there is a potential to bind polar residues at substrate residue P3 because of the presence of conserved Asp and Arg at positions corresponding to residues 29 and 8 of the HIV-1 protease. Further from the cleaved peptide, that is, beyond subsites S4 and S3′, the substrate lies near the surface of the dimer where there are several charged residues, so polar amino acids could be preferred at distal locations in the substrate.

The structural information summarized above necessitates interpretation in energetic terms and further delineation of free energetically favorable interactions between the enzyme and the ligand. We present here a free energy component analysis of the binding of HIV-1 protease with two inhibitors, focusing on the forces stabilizing these complexes. On the basis of the results here and our previous free energy studies on the

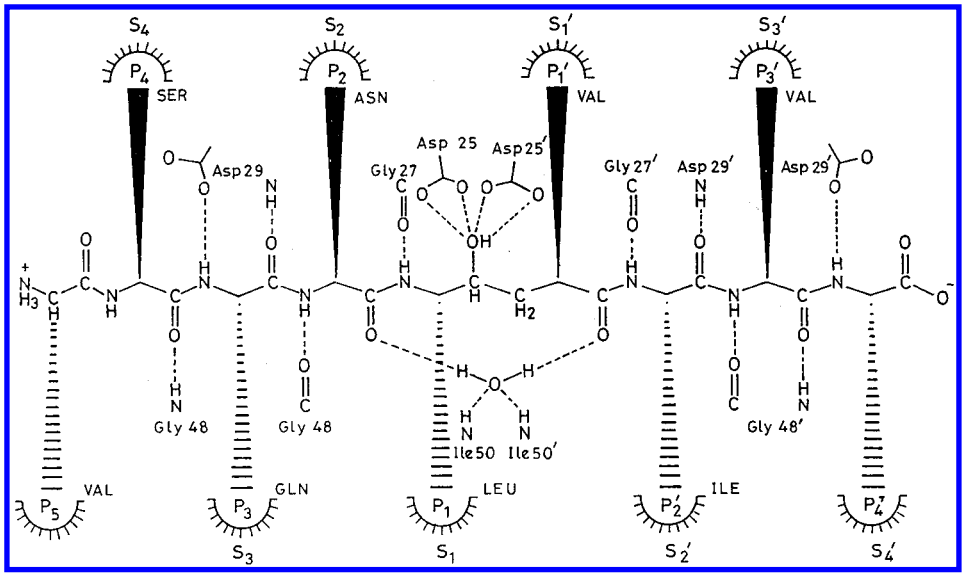


**Figure 2.** Diagrammatic representation of the modes of interaction of HIV-1 protease with an inhibitor (layout is adapted from ref 31).

**Table 2.** Residues and Atoms of the Protease That Interact with the Substrate Atoms at Various Subsites

| subsite | substrate atom | HIV-1 PR atom | HIV-1 PR |
|---|---|---|---|
| S4 | P4 NH | | $Asp^{30'}Ile^{50'}Ile^{54'}$ |
| | P4 C=O | O $Asp^{29'}$ | $Arg^{8}$ $Asp^{29'}Leu^{23}$ |
| | P3 C=O | NH $Asp^{29'}$ | $Val^{82}Arg^{87'}$ |
| S2 | P2 NH | C=O $Gly^{51'}$ | $Ala^{28'}Val^{32}$ $Phe^{53'}$ |
| | P2 C=O | NH $Phe^{53'}$ | $Ile^{54'}$ $Leu^{76'}Thr^{80'}Ile^{84'}$ |
| S1 | P1 NH | C=O $Gly^{27'}$ | $Leu^{23}$ $Asp^{25}$ $Phe^{53}$ |
| | P1 C=O | O $Asp^{25}$ | $Pro^{81}$ $Val^{82}$ $Ile^{84}$ |
| S1′ | P1′NH | O $Asp^{25'}$ | $Leu^{23'}Asp^{25*'}Phe^{53'}$ |
| | P1′ C=O | NH $Phe^{53}$ | $Ile^{84}$ |
| S2′ | P2′ NH | C=O $Gly^{27}$ | $Ala^{28}$ $Asp^{30}$ $Val^{32}$ |
| | P2′ C=O | NH $Asp^{29}$ | $Phe^{53}$ $Ile^{54}$ $Leu^{76}$ $Ile^{84}$ |
| S3′ | P3′ NH | C=O $Gly^{51}$ | $Arg^{8'}$ $Asp^{29}$ $Ile^{50}$ |
| | P3′ C=O | NH $Gly^{51}$ | $Pro^{81'}$ $Arg^{87}$ |

aspartic protease−pepstatin complexes reported previously, we attempt to draw some general inferences of relevance to inhibitor design.

## II. Theory and Methodology

A number of theoretical methods for analyzing the energetics of protein−ligand interactions are presented in the literature ranging from the free energy perturbation[1−5] methods at one end to the analysis of energy-minimized structures at the other.[34] The master equation approach proposed by Ajay and Murcko (please see Appendix for a detailed discussion) offers a middle ground between these two extremes and allows for the development of computationally rapid methods.

$$\Delta G^\circ = \Delta G^\circ_{tr} + \Delta G^\circ_{rot} + \Delta G^\circ_{intra} + \Delta G^\circ_{solv} + \Delta G^\circ_{salt}$$

Some of the approximations inherent in this equation such as additivity of the components, particularly that of intrasolute and solvation contributions, and validity of gas-phase statistical mechanics for translational and rotational contributions, etc. can be alleviated by working with molecular dynamics trajectories developed with explicit waters and then carrying out a *post facto* free energy analysis with a force-field-compatible solvation model via a suitable thermodynamic cycle.

Several system-specific methods have been used successfully in the development of potent enzyme inhibitors. The simplest method[34] calculates only the enzyme−inhibitor interaction at the minimum-energy configuration, without including the solvent molecules. This method is very fast but could be oversimplified if the desolvation energy and entropic contributions are important.

Aqvist et al.[35] derived an equation for the series expansion of the free energy difference, and their method requires the inclusion of coefficients in addition to force field parameters. In this form, the original ensemble average for the exponential of potential energies reduces to the calculation of the ensemble average of simple potential energies. The series expansion, in principle, involves an infinite number of terms. In the linear interaction energy (LIE) approximation, Hansson and Aqvist[36] assumed that the second and higher order terms can be neglected, and LIE thus becomes a first-order approximation of the binding free energy.

$$\Delta G^\circ = \Delta G^\circ_{el} + \Delta G^\circ_{vdw} \approx \alpha(V^{el}_{bound} - V^{el}_{free}) + \beta(V^{vdw}_{bound} - V^{vdw}_{free})$$

where $V^{el}_{bound}$ and $V^{vdw}_{bound}$ are the electrostatic and van der Waals interaction energies between the ligand and the solvated protein from an MD trajectory with the ligand bound to the protein and $V^{el}_{free}$ and $V^{vdw}_{free}$ are the corresponding interaction energies between the ligand and water from an MD trajectory with the ligand in water. $\alpha$ and $\beta$ are two empirical parameters.

Lee et al.[37] determined the binding free energies at different levels of approximation for HIV PR, but they found the need for including the solvent molecules explicitly. Wang et al.[38] determined the binding free energies of PR−inhibitor complexes accurate up to the second decimal place by optimizing the $\beta$ coefficient in the LIE method. The coefficient, however, is system-specific. Dominy and Brooks[39] developed a protocol to generate ensembles of protein−ligand complexes and then to carry out the energy analysis. Verkhivker et al.[40] carried out an empirical free energy calculation on the complexes of HIV-1 protease, and reported good correlation with the experimental data. Quantitative struc-

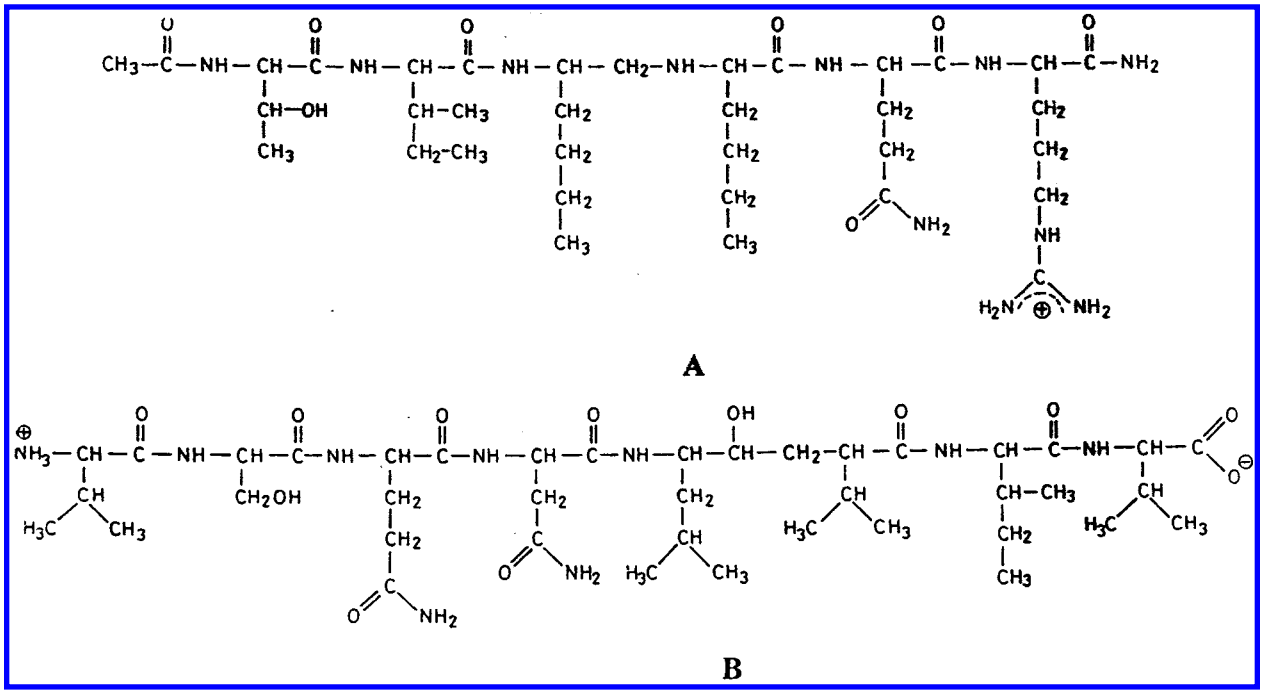


**Figure 3.** Molecular structural formulas of inhibitors studied: (A) 4hvp; (B) 8hvp.

ture−activity relationship studies have proven to be useful in the context of drug design efforts for computing the relative binding affinities of closely related inhibitors.[41−45] These methods seek to correlate an activity such as binding affinity with some function of the inhibitor properties. Although useful for a particular case in which large amounts of binding data are available, these methods are not generally applicable when little experimental data are available. In addition, it is difficult to rationalize the resulting function in terms of physicochemical interactions.

The MMGBSA (molecular mechanics−generalized Born solvent accessiblity) model illustrated here in this contribution constitutes the current state of the art to estimate binding free energies short of full free energy simulations, the only other alternative at this level of theoretical rigor being the MMPBSA (molecular mechanics−Poisson−Boltzmann surface area) model.[46] Kollman and co-workers[47] have recently derived the binding free energies of 12 TIBO-like HIV-1 RT inhibitors using the MMPBSA methodology. The MMPBSA methodology as reported by Kollman and co-workers and the MMGBSA as illustrated and applied in this contribution are similar in spirit. Whereas the former utilizes the Poisson−Boltzmann treatment for including solvation effects, the latter adopts generalized Born methodology for the same. Both methods are expected to give similar results, with the latter expected to be faster computationally. The method reported in this article assumes additivity[14] wherein the net free energy change is treated as a sum of a comprehensive set of individual contributions. Each component is estimated in a force-field-compatible manner. Proper partial atomic charges derived from ab initio quantum calculations are assigned to all atoms of the protease and the inhibitor. This is followed by nanosecond-long molecular dynamics simulations with explicit solvent on the bound and the unbound enzymes. No distance-specific or residue-specific restraints are applied in the molecular dynamics (MD) simulation. Structures thus generated from simulations are then utilized to form ensemble averages for each free energy component to quantify their relative magnitudes and to determine whether they make favorable or unfavorable contributions to the free energy of complexation, providing potentially useful new knowledge on inhibitor design.

## III. Calculations

The atomic coordinates of the complexes of HIV-1 protease with inhibitor molecules were obtained from the RCSB Protein Data Bank[48] (PDB code: 4hvp and 8hvp). A diagrammatic representation of the inhibitor molecules is shown in Figure-3. Our calculations are based on all-atom models for the enzyme and the inhibitors in which hydrogen atoms are added explicitly to the crystal structure. AMBER[15,16] partial charge assignment for the enzyme atoms proceeds in a straightforward manner. After a series of preliminary investigations with all permutations of ionization states of the catalytic aspartates Asp25 and Asp25′, we have decided on a protonated Asp25′ and deprotonated Asp25.[24] Partial atomic charges for the inhibitor atoms were derived consistent with the AMBER protocol. The charges on the atoms of a residue were determined along with its flanking residues by generating the electrostatic potentials with the 6-31G* basis set using GAMESS[49] and then fitting them with the RESP[50] module of the AMBER molecular modeling package. The enzyme−inhibitor complex was surrounded by a box of water molecules (∼4500 waters), and periodic boundary conditions were applied to the system. Energy minimization was then performed using the Sander module of AMBER to relieve any unfavorable clashes in the crystal structure. Here, 1000 steps of minimization were carried out first for waters only (500 steps of steepest descent, SD, followed by 500 steps of conjugate gradient, CG), followed by a further 500 steps (250 SD + 250 CG) of hydrogen atoms on the complex and water molecules. After this, 150 steps (100 SD + 50 CG) of free minimization was carried out. The minimized structure was

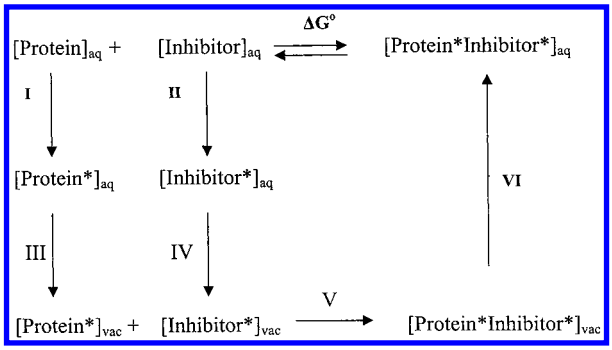


**Figure 4.** Thermodynamic cycle adopted to construct the standard free energies of enzyme−inhibitor binding from molecular dynamics simulations.

then used as input to the molecular dynamics study. We used a 1 fs ($10^{-15}$ s) time step for integrating the equations of motion in all our MD studies. Here, 20 ps (picosecond) of heating phase was first carried out in which the system was heated from 100 to 300 K using a constraint of 25 kcal on all the atoms except waters. This was followed by a 50 ps period of equilibration in which all constraints were removed, and finally a 1 ns data collection phase was carried out. This protocol for developing a nanosecond molecular dynamics trajectory on the protease−inhibitor complex with a box of waters is applied to each of the two complexes (viz. 4hvp and 8hvp). MD was also carried out on the uncomplexed proteins (separate MD simulations were performed on the native structures for both the complexes because of the differences in resolution of the crystal structures) in order to assess the free energy expense of structural adaptation of the protease in forming the complex. Because in these systems the inhibitors were small and bind to the protease in an extended form, MD simulations on the free inhibitors were not deemed required.[24,51] In cases where the inhibitor is flexible, MD simulations can be performed on the free inhibitor as well to calculate its adaptation expense. The calculations were performed on a SGI Origin 200 four-processor machine.

A *post facto* analysis of the molecular dynamics trajectories on each of the complexes is then carried out to construct a thermodynamic cycle for the protein−inhibitor binding in aqueous solution (Figure 4). About 100 structures at intervals of 10 ps are culled from each trajectory followed by a computation of the averages for each component in the cycle listed below. The intramolecular energetics is based on the AMBER force field. The electrostatic contribution to solvation was calculated via the AMBER compatible modified generalized Born model.[17−22] The nonelectrostatic contribution to solvation, which involves molecular surface area calculations, were performed using the ACCESS program based on the algorithm of Lee and Richards[52] but with AMBER van der Waals radii. The added salt concentration employed in the Debye−Huckel term was 0.18 M. Entropies are computed using statistical mechanical expressions. Further methodological details on the evaluation of each component are provided previously.[53,54] Results for each of the components enumerated as contributing to the binding for each system are presented below.

**Table 3.** Conventional Combination of the Computed Primary Terms Contributing to the Net Free Energies of Binding (kcal/mol) for the HIV-1 Protease−Inhibitor Complexes

| | 4HVP | 8HVP |
|---|---|---|
| $\Delta G°_{el}$ | 38.95 | 24.37 |
| $\Delta G°_{vdw}$ | −16.73 | −13.85 |
| $\Delta G°_{cav}$ | −74.42 | −76.44 |
| $\Delta G°_{entropy}$ | 37.47 | 39.32 |
| $\Delta G°_{ions}$ | −0.23 | 0.73 |
| $\Delta G°_{adp}$ | 5.23 | 8.14 |
| $\Delta G°_{tot}$ | −9.73 | −17.73 |
| $\Delta G°_{expt}$ | −8.40 | −12.30 |

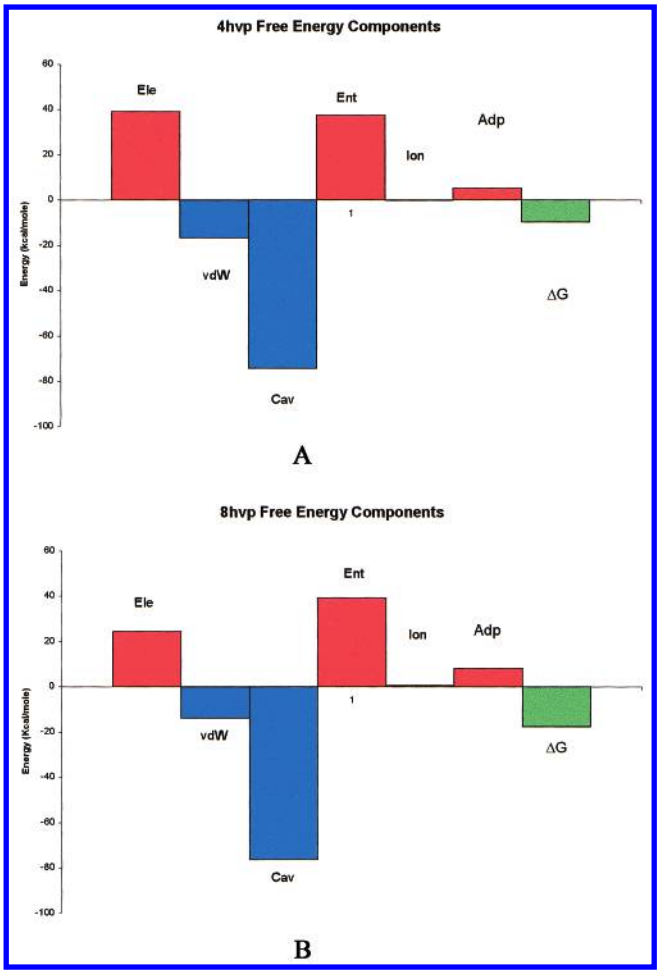


**Figure 5.** Contribution of the free energy components to the net binding energy: (A) 4hvp; (B) 8hvp.

## IV. Results

An important aspect in a conventional free energy analysis study is the net contribution of electrostatics (including hydrogen bonds), shape complementarity, hydrophobic effects, structural adaptation, etc. to binding, the values of which are shown in Table 3. The sign conventions are defined in such a way that negative values are favorable and positive values unfavorable to binding. An analysis of the results of Table 3 based on the net contributions from electrostatics, van der Waals, cavitation, entropy, and ion effects are presented schematically in Figure 5 for the two complexes. Here, the differential effects of direct van der Waals interactions between the protein and the drug and cavity formation upon complexation are seen to be favorable to binding, while electrostatics, adaptation, and entropy losses are unfavorable to binding. Ion effects were found to be

**Table 4.** Binding Free Energies (in kcal/mol) for HIV-1 Protease−Inhibitor Complexes: A Comparison of Various Protocols

| | 4hvp | 8hvp |
|---|---|---|
| I. Sigmoidal Dielectric Function | | |
| a. minimization (500/150)[a] | −13.70 | −12.40 |
| b. molecular dynamics (1 ns) | 8.30 | 5.60 |
| II. 4*r* Dielectric Function | | |
| a. minimization (500/150) | −12.00 | −6.80 |
| b. molecular dynamics (1 ns) | 7.70 | 5.90 |
| III. Explicit Water (Solvent Shell, ∼1500 Waters) | | |
| a. minimization (1000/500/150) | −30.10 | −25.30 |
| b. molecular dynamics (1 ns) | −27.90 | −13.10 |
| IV. Explicit Water (Solvent Box, ∼4500 Waters) | | |
| a. minimization (1000/500/150) | −25.23 | −22.44 |
| b. molecular dynamics (1.2 ns) | −9.73 | −17.73 |
| experimental values | −8.40 | −12.30 |

[a] See text.

**Table 5.** Hydrogen Bond Distances in the Crystal Structure and the Molecular Dynamics Simulation Averages

| inhibitor | protein | cryst (Å) | MD (Å) |
|---|---|---|---|
| 8hvp | | | |
| N (201) | O (46) | 3.7 | 4.9 |
| O (202) | N (48) | 3.1 | 3.2 |
| OG (202) | OD2 (30) | 2.6 | 2.7 |
| N (203) | OD2 (29) | 3.9 | 3.9 |
| O (203) | N (29) | 3.2 | 3.7 |
| N (204) | O (48) | 2.7 | 3.0 |
| ND2 (204) | N (29) | 3.2 | 4.4 |
| ND2 (204) | N (30) | 3.6 | 4.3 |
| ND2 (204) | O (30) | 3.4 | 3.0 |
| N (205) | O (27) | 3.4 | 3.0 |
| OS (205) | OD1 (25) | 2.4 | 3.2 |
| OS (205) | OD2 (25) | 2.9 | 2.8 |
| OS (205) | OD1 (125) | 3.1 | 3.4 |
| OS (205) | OD2 (125) | 2.7 | 3.0 |
| N (207) | O (127) | 3.4 | 5.3 |
| O (207) | N (129) | 3.4 | 5.2 |
| N (208) | O (148) | 2.9 | 2.9 |
| O (208) | N (148) | 3.4 | 3.3 |
| 4hvp | | | |
| ACE (O) | H (48) | 2.9 | 2.3 |
| O (201) | H (29) | 3.9 | 5.8 |
| N (202) | O (48) | 2.6 | 2.1 |
| N (203) | O (27) | 2.9 | 3.6 |
| H (205) | O (148) | 3.0 | 2.1 |
| H (amide) | OD2 (129) | 3.0 | 3.3 |

system-specific, being favorable for 4hvp and unfavorable for 8hvp. The ion effects were treated with a continuum model using Debye−Huckel theory, and hence, concerns regarding the dependence of ion locations on MD run lengths are circumvented. Overall, the computed net binding free energies with both inhibitors are in good accord with experimental results.

Promotion of hydrogen bond interactions and van der Waals packing between the receptor and the drug are two recurring chemical features in drug design attempts. Structural studies on the complexes under investigation have implicated several hydrogen bond interactions. The MD simulations can bring out the dynamical character of these hydrogen bonds and whether the interface is tight. Using the snapshots from the MD simulation trajectories on the complexes, we monitored the hydrogen bond distances between the catalytic aspartates and the hydroxyl group at the scissile bond. Table 5 shows some of the important distances resulting from the MD simulations on the complexes along with the values reported in crystal studies. Most of the contacts reported in the crystal structures remained intact during the course of the simulations. This merely indicates that the interactions are favorable in a final state analysis, i.e., in the complex. The consequence of the contacts to the net binding free energy is a different matter and has to be analyzed in terms of initial (reactants infinitely apart in aqueous medium) versus final state energetics. As mentioned in the previous paragraph, Figure 5 represents these differential effects wherein van der Waals is favorable in both complexes, i.e., protease with 4hvp and 8hvp, implying that van der Waals contacts between the inhibitors and the protease promote binding. The net electrostatics is unfavorable in both the complexes. The net electrostatics is a composite of direct electrostatics, i.e., the Coulombic interactions and the desolvation expense. The direct electrostatics shows case specificity, being unfavorable with 4hvp and favorable with 8hvp. The presence of a hydrogen bond or a salt bridge in the final state does not necessarily imply a more favorable net binding free energy for association, an expectation especially relevant in aqueous medium.[55]

From a drug design perspective, it is of interest to pin down residues or functional groups on the inhibitor that can be modified for optimal binding. The MD simulation trajectories on the complexes can be used to identify the inhibitor residues that interact favorably or unfavorably with the enzyme. Taking this interaction energy as a guide, a modification is attempted and its consequence tested on the net binding free energy. For estimation of the effect of inhibitor modification on the net binding free energies, select mutations are carried out on the snapshots of the MD trajectories of the unmodified complex and the thermocycle is followed through as a preliminary indicator. If the modifications are successful, i.e., if the net binding free energies are more favorable than with the original inhibitor, then more accurate estimates of the binding free energies for the modified inhibitors can be developed via a quantum derivation of charges of the modified inhibitors followed by nanosecond MD simulations on the enzyme-modified inhibitor complexes. We carried out a residue-wise analysis of direct electrostatics and van der Waals interactions for each unit of the inhibitor and compared the results with that of the entire protein (Figure 6) to provide insights into some plausible modifications. We tested our ideas by first shortening and neutralizing the side chain of the terminal arginine residue of the inhibitor of 4hvp because the residue was showing unfavorable interactions with some of the protein residues. Essentially, the arginine side chain was converted into an ethyl group. In the second mutation, the isoleucine of the inhibitor was converted into valine to see how reducing the side chain of the residue causes a difference in packing. In the first mutation, it was observed that the binding free energy became more favorable and also the direct electrostatic interactions with the mutated arginine and the protein became less unfavorable. The second mutation, however, did not significantly affect the binding free energy. In case of 8hvp, several mutation studies were carried out. To check whether a hydrogen bonding group at the scissile bond is desirable, the binding occurring in aqueous

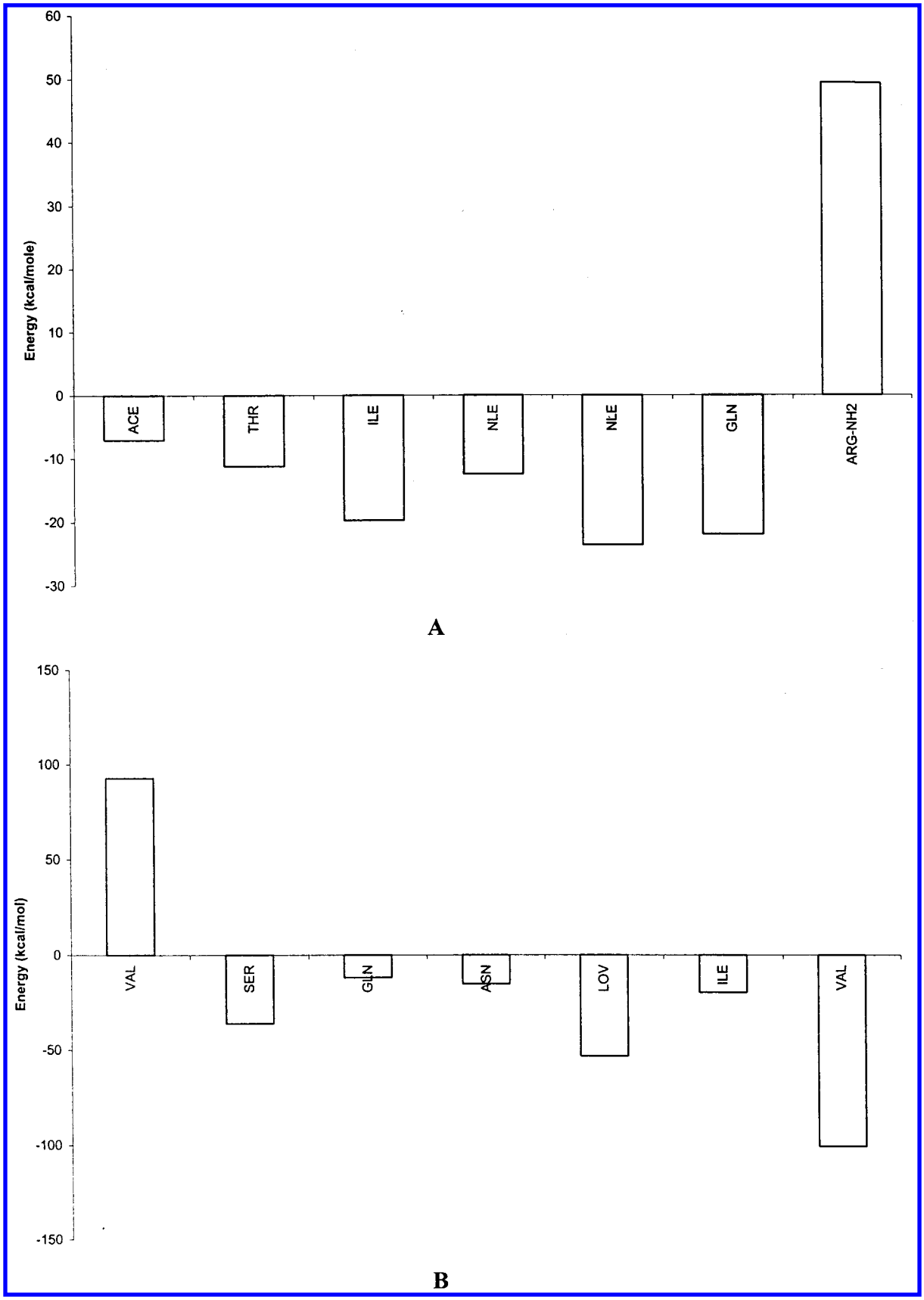


**Figure 6.** Interaction (electrostatics + van der Waals) energies of each residue of the inhibitor with the protease: (A) 4hvp; (B) 8hvp.

medium, we converted the CH(OH) group to $CH_2$ and found that the magnitude of the binding free energy was reduced drastically. We also monitored the effect of discharging the residues at the termini and found that neutralizing the N-terminus did not cause any change in the free energy value, whereas discharging the C-terminus causes the magnitude to decrease precipitously. The isoleucine residue at the P2′ side was

converted to valine, and this caused a decrease in the magnitude of free energy, hence causing less efficient packing. These mutational studies, in a nutshell, represent interaction and binding views of the energetics of enzyme−inhibitor association.

## V. Discussion

The application of the MMGBSA method for evaluating the standard free energies of protein−drug association is described. The statistical mechanical basis for the MMGBSA model and the scope for hierarchical improvements are presented in the Appendix. A free energy component analysis of the complexes of HIV-1−4hvp and HIV-1−8hvp enabled identification of molecular level forces favorable to the binding of inhibitors with HIV-1 protease. We note from the analysis that net van der Waals (i.e., van der Waals interactions between the protein and the ligand minus van der Waals interactions of the protein and ligand with the solvent) and cavitation effects (loss in solvent accessible surface area of both polar and nonpolar atoms upon complexation) are favorable, whereas net electrostatics, entropy, and adaptation effects are unfavorable to binding. We would expect the entropies to be unfavorable because the translational and rotational degrees of freedom are decreased and the motion of the side chains at the interface gets restricted when association occurs. Also, the natural form of the protein is known to be the most stable one and on binding there occurs some strain in the protein because it has to adapt itself to the structure of the inhibitor. This fact is reflected in the positive value of the adaptation energy expense. In our analysis, the direct electrostatic interaction between the inhibitor and the protein was found to be favorable for 8hvp and unfavorable for 4hvp, but combined together with the desolvation expense, the net electrostatics component turned out to be unfavorable to binding (Table 3). Packing effects (differential van der Waals energetics) are found to be favorable, and this is consistent with the theory of induced fit. Cavitation effects are always favorable to binding in aqueous medium.

In addition to the free energy component analysis, we also performed a residue-wise analysis on the two complexes. On the basis of this study, we identified residues whose modification could lead to a potential increase in the magnitude of the free energy of binding. In the case of 4hvp, we found that mutating the terminal arginine by shortening its side chain to an ethyl group gave better binding affinity. In the case of HIV-1 protease, a hydrogen bonding group at the cleavage site is a desirable feature because it can interact with the catalytic aspartates and contribute favorably to binding. This was reflected in our residue-wise interactions and the mutational studies on 8hvp. We also considered the effect of discharging the terminal residues in the case of 8hvp. Discharging the N-terminus does not have a significant effect on the binding free energy. On the other hand, discharging the C-terminus causes a precipitous drop in the free energy of binding. So a negatively charged residue at the C-terminus is favorable to binding. The residue-wise interactions show that the N-terminus is interacting unfavorably with the protein residues (Figure 6), but discharging it does not affect the free energy of binding. This points to a subtle balance between the interaction and solvation energies.

The methodology reported here has been arrived at after a critical evaluation of numerous protocols for developing binding free energies. One popular theoretical alternative is to cut down on the degrees of freedom and hence computational time by working with implicit solvent (dielectric continuum models) during minimization and molecular dynamics stages. Some inherent limitations of the continuum solvent models rendered it impossible to obtain reliable estimates of binding free energies. The various protocols and the values obtained are listed in Table 4. The free energy values obtained after minimization show close correlation to the experimental values, but this does not include the adaptation expense, and the closeness to the experimental values is merely fortuitous and may not be observed with other systems.[51,52] We observed that with an implicit solvent (sigmoidal and 4*r* dielectric models) the electrostatics tends to get overestimated and the system tries to maximize its electrostatics at the expense of van der Waals interactions, further leading to large deviations from the starting structure that may or may not be realistic. Minimization with explicit solvent is one option. Because the number of variables in the minimization becomes extremely large, locating minima on the free energy landscape is difficult. This makes a molecular dynamics study imperative. However, in the MD studies with a solvent shell, i.e., with a limited number of waters and without periodic boundary conditions, the water molecules were observed to fly off because there were no restrictions on the density of the system. We thus converged to nanosecond-long molecular dynamics simulations with ∼4500 waters and periodic boundary conditions. This, however, is not to undermine the importance of the continuum solvent methods, which can work extremely well in specific cases, particularly when the interface is tight and no significant structural changes occur upon complexation.

Eventually, the theory should be able to predict the binding free energies from first principles, where no structural data are available on the native protein and no experimental thermodynamic data are available for verifying the predictions. It is here that adaptation could prove to be extremely difficult to handle. Adaptation involves the partial folding/restructuring of the protein from its native state to the complexed state. To calculate the adaptation energy, we took the coordinates of the adapted protein from the complex and subjected the system to nanosecond-level molecular dynamics simulations with explicit solvent. Figure 7 shows the superposition of the average structure of the complexed and the native form of the protein for each system. The average structure of the protein emerging from MD simulations differs from that in the complex. The implication of this to adaptation expense is obvious, and in such situations MD simulations may be indispensable.

We further investigated the sensitivity of the free energy estimates to the simulation run length. We checked the convergence of the binding free energy results for 50, 70, and 100 points (structures culled from the MD trajectories), and the results did not vary much.

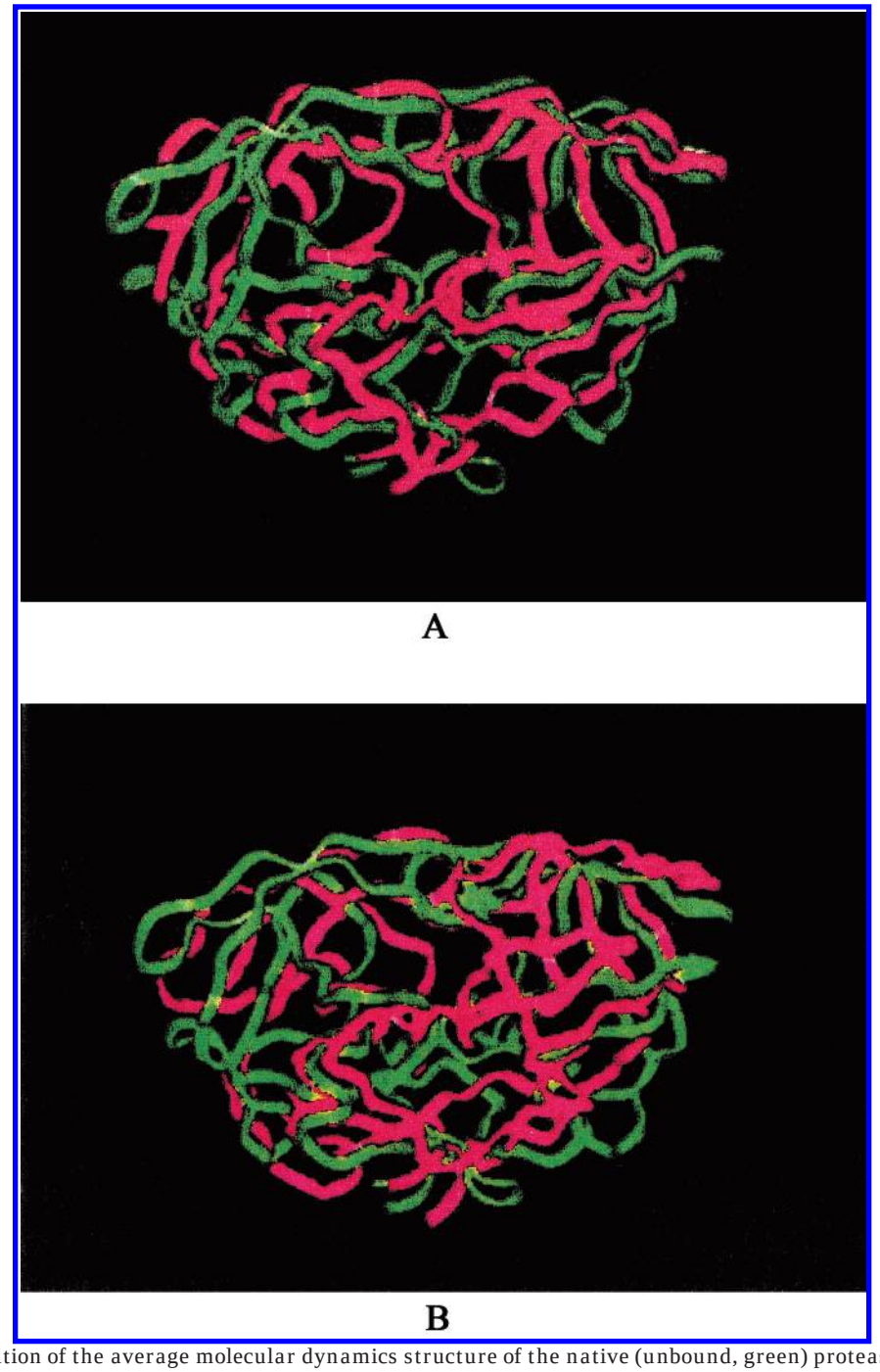


**Figure 7.** Superposition of the average molecular dynamics structure of the native (unbound, green) protease with the complexed (red) form: (A) 4hvp; (B) 8hvp.

We extended the 1 ns MD trajectories on each of the systems by an additional 200 ps and found the binding free energies to be stable. The variation of the net free energy of binding for both systems is given as a function of time in Figure 8. For the HIV-1 protease case investigated here, nanosecond-level MD appears satisfactory for the results to converge. If partial folding is involved or allosteric changes occur, much longer simulations on the unbound protein may be necessary to obtain reliable values of adaptation and hence the net binding free energies.

A detailed energy analysis of the various inhibitors bound to the enzyme yielded suggestions on better inhibitor design. Though the present study was conducted on the HIV-1 protease system, the methodology adopted is general and the drug design principles for any other system can be elicited using the same strategy.

In summary, the MMGBSA method for estimating the binding free energies is different from the other approaches described extensively in the literature. Once the force field and a solvation energy model compatible with the force field are fixed, the method does not need to be calibrated from system to system and does not require the use of any empirical coefficients, the limitation being the computational demands. Even this will soon not be a handicap with increasing processor speed and parallel architecture on which the simulations can be conducted. Very soon, drug design algorithms may have to start directly from genomic information with little structural data. This is an emerging computational challenge. The first step in this challenge is the receptor

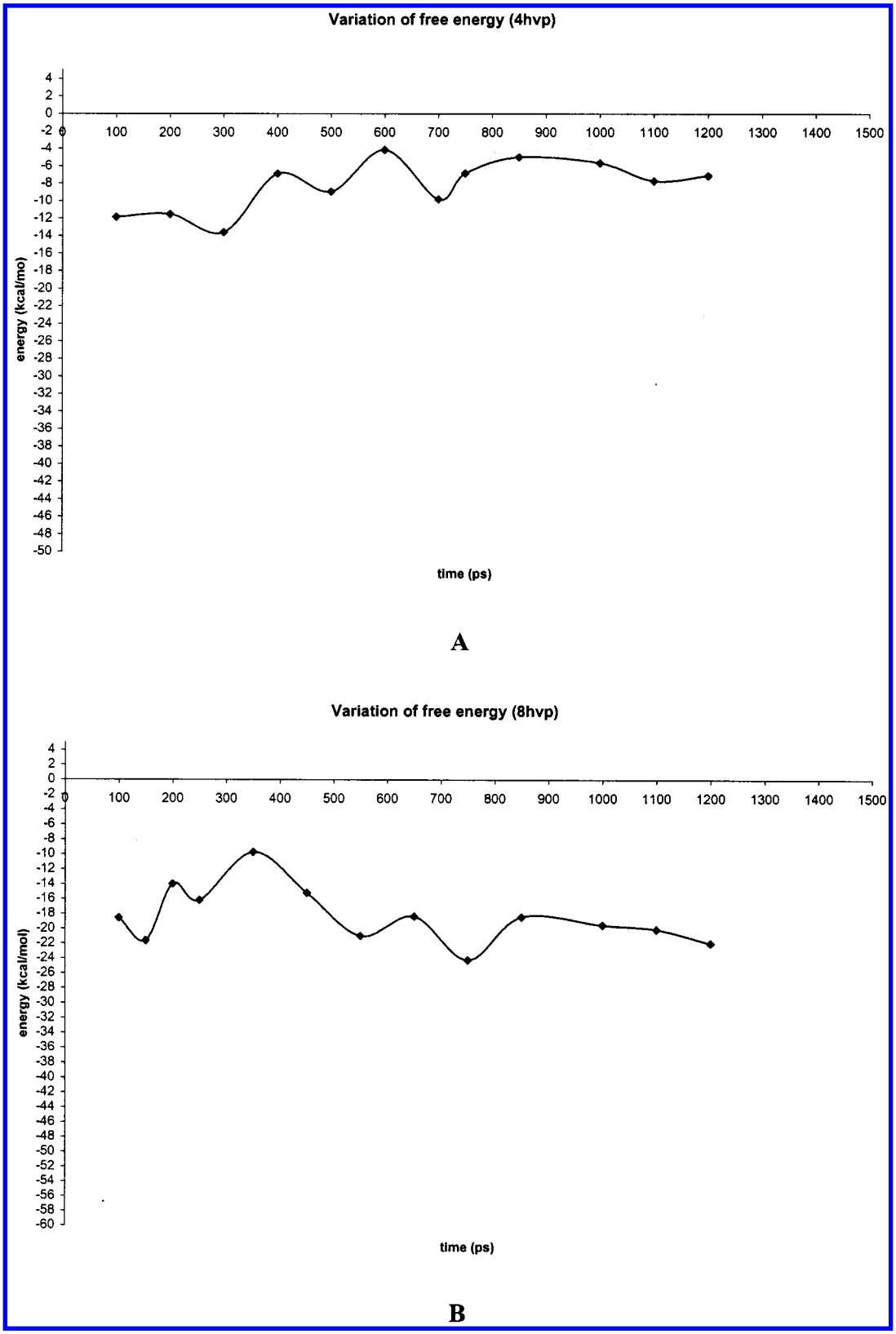


**Figure 8.** Variation in binding free energy estimates as a function of molecular dynamics simulation run length: (A) 4hvp; (B) 8hvp.

structure prediction, the second step being the design of lead compounds to match the receptor site. The third step is docking with the attendant kinetic issues. The fourth step is the evaluation of binding affinity. Steps 2−4 may have to be repeated in several cycles with computational protocols of increasing levels of confidence to obtain viable suggestions on drug design. The methodology presented here is robust enough to address the issue of binding affinity.

## VI. Conclusions

A free energy methodology (MMGBSA) formulated within the framework of statistical mechanics has been

reported and applied to two HIV-1 protease−inhibitor complexes. Several other computationally simpler protocols have been tried before arriving at this methodology, and some of their drawbacks are highlighted. The strategy involves *post facto* analyses of the molecular dynamics simulations, and this is seen to yield reliable free energy estimates. Also, several of the crystal contacts are confirmed during the simulations. The free energy component analysis presented here lends itself to automation for eliciting suggestions on drug design, and this is illustrated for the two HIV-1 protease complexes. It is found that a less extended residue with a neutral side chain at the P3′ site and a smaller residue at the P2 site is preferred. In addition to this, the presence of a hydrogen bonding group at the scissile position and a negatively charged C-terminus are desirable features for HIV-1 protease inhibitors. The generality of the method is emphasized in that the same protocol can be used unchanged to calculate binding free energies of any other protein−ligand system.

**Acknowledgment.** Funding from the Indo-French Centre for the Promotion of Advanced Research (IFCPAR) is gratefully acknowledged. Ms. Parul Kalra thanks the Council of Scientific and Industrial Research (CSIR) for the award of Senior Research Fellowship. The authors thank Mr. Achintya Das for helpful comments and Compaq Testdrive for machine time.

## Appendix

**Theory and Methodology for Enzyme−Inhibitor Binding in Aqueous Media.** Taking off from the discussions on the statistical thermodynamics of binding put forth by Ben Naim,[56] Gilson et al.,[57] Janin,[58,59] Ajay and Murcko,[12] and Atkins,[60] we present here, within the statistical mechanical framework, a hierarchy of methods accessible to theory of varying levels of rigor and computational requirements.

Let E and I be the reactants and E*I*, the product of binding in aqueous medium.

$$[\mathrm{E}]_{aq} + [\mathrm{I}]_{aq} = [\mathrm{E^*I^*}]_{aq} \tag{1}$$

At equilibrium,

$$\mu_{\mathrm{E.aq}} + \mu_{\mathrm{I.aq}} = \mu_{\mathrm{E^*I^*.aq}} \tag{2}$$

$\mu_{\mathrm{E.aq}}$ is the chemical potential of species E in the solvent medium (partial molar Gibbs free energy) and $\mu^\circ_{\mathrm{E.aq}}$ is its standard chemical potential, i.e., under conditions of 1 bar in the gaseous state and 1 M (designated as $C^\circ$) in liquid state.

$$\mu^\circ_{\mathrm{E.aq}} + RT\ln(a_{\mathrm{E}}) + \mu^\circ_{\mathrm{I.aq}} + RT\ln(a_{\mathrm{I}}) = \mu^\circ_{\mathrm{E^*I^*.aq}} + RT\ln(a_{\mathrm{E^*I^*}}) \tag{3}$$

where $a_{\mathrm{E}}$ ($=\gamma_{\mathrm{E}}C_{\mathrm{E}}/C^\circ$) is the activity of E, $\gamma_{\mathrm{E}}$ is the activity coefficient of species E, and $C_{\mathrm{E}}$ its concentration. The standard molar Gibbs free energy of the reaction (standard absolute molar Gibbs free energy of binding) is

$$\Delta G^\circ_{aq} = \mu^\circ_{\mathrm{E^*I^*.aq}} - (\mu^\circ_{\mathrm{E.aq}} + \mu^\circ_{\mathrm{I.aq}}) = -RT\ln[a_{\mathrm{E^*I^*}}/(a_{\mathrm{E}}a_{\mathrm{I}})] = -RT\ln K_{\mathrm{eq.aq}} \tag{4}$$

In terms of canonical partition functions ($Q$),

$$\Delta G^\circ_{aq} = \Delta A^\circ_{aq} + P\Delta V^\circ_{aq} = -RT\ln K_{\mathrm{eq.aq}} = -RT\ln\left[\frac{\dfrac{Q_{\mathrm{E^*I^*.aq}}}{N_{\mathrm{A}}Q_{\mathrm{w}}}}{\dfrac{Q_{\mathrm{E.aq}}}{N_{\mathrm{A}}Q_{\mathrm{w}}}\dfrac{Q_{\mathrm{I.aq}}}{N_{\mathrm{A}}Q_{\mathrm{w}}}}\right] + P\Delta V^\circ_{aq} \tag{5}$$

Equation 5 is an exact expression for noncovalent associations in aqueous medium. The assumption that translations and rotations are separable from intrasolute degrees of freedom as well as those of solvent (a common practice is gas-phase statistical mechanics) then leads to

$$\Delta G^\circ_{aq} = -RT\ln\left[\frac{Q^{\mathrm{tr}}_{\mathrm{E^*I^*}}Q^{\mathrm{rot}}_{\mathrm{E^*I^*}}Z^{\mathrm{int}}_{\mathrm{E^*I^*.aq}}Q^{\mathrm{el}}_{\mathrm{E^*I^*}}N_{\mathrm{A}}Q_{\mathrm{w}}}{(Q^{\mathrm{tr}}_{\mathrm{E}}Q^{\mathrm{rot}}_{\mathrm{E}}Z^{\mathrm{int}}_{\mathrm{E.aq}}Q^{\mathrm{el}}_{\mathrm{E}})(Q^{\mathrm{tr}}_{\mathrm{I}}Q^{\mathrm{rot}}_{\mathrm{I}}Z^{\mathrm{int}}_{\mathrm{I.aq}}Q^{\mathrm{el}}_{\mathrm{I}})}\right] + P\Delta V^\circ_{aq} \tag{6}$$

$\Delta A^\circ$ is the standard Helmholtz free energy of the reaction. The Avogadro number $N_{\mathrm{A}}$ in the above equation originates from expressing partition functions $Q$ as molar partition functions (following the notation of Atkins[6]), and $P\Delta V^\circ_{aq}$ is the pressure−volume correction to the Helmholtz free energy in the solvent medium. $Q_{\mathrm{w}}$ denotes the partition function for pure solvent (water). $Z^{\mathrm{int}}$ is the configurational partition function. It includes contributions from vibrations and internal motions as well as solvation (hydration) effects. The translational and rotational terms have been separated; i.e., momentum-dependent terms have been integrated from an integral of the following type:

$$Z^{\mathrm{int}}_{\mathrm{E.aq}} = \int\!\cdots\!\int \exp\left\{\frac{-E(\mathbf{X}^{\mathrm{N}}_{\mathrm{E}},\mathbf{X}^{\mathrm{M}}_{\mathrm{W}})}{k_{\mathrm{B}}T}\right\} \mathrm{d}\mathbf{X}^{\mathrm{N}}_{\mathrm{E}}\,\mathrm{d}\mathbf{X}^{\mathrm{M}}_{\mathrm{W}} = \left\langle \exp\left(\frac{E(\mathbf{X}^{\mathrm{N}}_{\mathrm{E}},\mathbf{X}^{\mathrm{M}}_{\mathrm{W}})}{k_{\mathrm{B}}T}\right)\right\rangle \tag{7}$$

$\mathbf{X}^{\mathrm{N}}_{\mathrm{E}}$ and $\mathbf{X}^{\mathrm{M}}_{\mathrm{W}}$ represent the configurational space accessible to the solute E and solvent W, respectively, in the presence of each other. $E(\mathbf{X}^{\mathrm{N}}_{\mathrm{E}},\mathbf{X}^{\mathrm{M}}_{\mathrm{W}})$ denotes the total potential energy of the system describing nonidealities. It includes intramolecular interactions within the solute E and solvent W as well as intermolecular interactions between the solute and the solvent. $k_{\mathrm{B}}T$ is the product of the Boltzmann constant and temperature (in kelvin).

$$Q^{\mathrm{el}}_{\mathrm{E}} \approx 1 \quad \text{(assumed for noncovalent associations)} \tag{8}$$

$$\Delta G^\circ = -RT\ln\left[\frac{Q^{\mathrm{tr}}_{\mathrm{E^*I^*}}N_{\mathrm{A}}}{Q^{\mathrm{tr}}_{\mathrm{E}}Q^{\mathrm{tr}}_{\mathrm{I}}}\right] - RT\ln\left[\frac{Q^{\mathrm{rot}}_{\mathrm{E^*I^*}}}{Q^{\mathrm{rot}}_{\mathrm{E}}Q^{\mathrm{rot}}_{\mathrm{I}}}\right] - RT\ln\left[\frac{Z^{\mathrm{int}}_{\mathrm{E^*I^*.aq}}Q_{\mathrm{w}}}{Z^{\mathrm{int}}_{\mathrm{E.aq}}Z^{\mathrm{int}}_{\mathrm{I.aq}}}\right] + P\Delta V^\circ_{aq} \tag{9}$$

Equation 9 is an exact expression for noncovalent associations in aqueous media and is subject only to the approximation of the separability of translations and

rotations from the remaining degrees of freedom in the system including those of solvent. The third term in eq 9 is accessible to free energy molecular simulations[10] configured in the canonical ensemble, albeit they are computationally expensive. The corresponding expression for associations in the gas phase is given as

$$\Delta G^\circ_g = -RT \ln\left[\frac{Q^{tr}_{E^*I^*} N_A}{Q^{tr}_E Q^{tr}_I}\right] - RT \ln\left[\frac{Q^{rot}_{E^*I^*}}{Q^{rot}_E Q^{rot}_I}\right] - RT \ln\left[\frac{Z^{int}_{E^*I^*}}{Z^{int}_E Z^{int}_I}\right] + P\Delta V^\circ_g \quad (10)$$

In the following we consider some approximations commonly employed to bring into the feasibility domain the binding free energy computations via evaluation of the right-hand-side expression in eq 9.

**Approximation 1. Formal Separation of External Degrees of Freedom**. The molecular translational partition function of E is

$$q^{tr}_E = \frac{V}{\Lambda^3_E} = \frac{V}{\left(\frac{h^2}{2\pi m_E k_B T}\right)^{3/2}} \quad (11)$$

The molar partition function of E is $Q^{tr}_E = (q^{tr}_E)^{N_A}$.

Note that the volume $V$ has been included in the translational part, consistent with ideal gas statistical mechanics. This requires that the $Z^{int}$ be divided by $V$ to quantify nonidealities (excess free energies). The translational part of the free energy in eq 9 is now given by the Sackur−Tetrode equivalent as

$$\Delta G^\circ_{tr} = -RT \ln\left[\left(\frac{N_A}{V}\right)\left(\frac{\Lambda^3_E \Lambda^3_I}{\Lambda^3_{E^*I^*}}\right)\right] = -RT \ln\left[\left(\frac{N_A}{V}\right)\left(\frac{h^2}{2\pi k_B T}\right)^{3/2}\left(\frac{m_{E^*I^*}}{m_E m_I}\right)^{3/2}\right] \quad (12)$$

The expression in the brackets in eq 12 is dimensionless. $N_A/V$ may be replaced by a concentration term. Note that the form of this expression is the same whether for the gas phase or the liquid phase provided the translational and rotational motions of the solute are unaffected by the solvent. Mathematically this implies that the integrations in the momentum space can be carried out separately for the solute and solvent. This will be true only in a continuum, frictionless solvent influencing the position-dependent potential energy but not the velocity-dependent kinetic energy of the solute. Hence, in a transfer process (an experiment involving transfer of species E from one phase to another phase such as from the gas phase to the liquid phase or octanol to water, etc.), this term cancels out. In binding processes, however, no such cancellation occurs. Also, if E, I, and E*I* are each treated as a collection of nonbonded monoatomic particles, then again the translational partition function for each species is written as a product of the individual partition functions of the constituent atoms, and because the number of atoms is conserved during binding, these terms cancel out. Again, this is not so for polyatomic species where the mass in the translational partition function $m_E$ ($=\sum_i m_i$) is evaluated as a sum of the masses of the constituent atoms.

Similar arguments apply to the rotational partition functions. Separating the rotational part from internal motions implies working under the rigid rotor approximation.

$$\Delta G^\circ_{rot} = -RT \ln\left[\left(\frac{\sigma_E \sigma_I}{\sigma_{E^*I^*}}\right)\left(\frac{1}{8\pi^2}\right)\left(\frac{h^2}{2\pi k_B T}\right)^{3/2}\left(\frac{I^a_{E^*I^*} I^b_{E^*I^*} I^c_{E^*I^*}}{I^a_E I^b_E I^c_E I^a_I I^b_I I^c_I}\right)^{1/2}\right] \quad (13)$$

$I^a_E$, $I^b_E$, and $I^c_E$ are the components of the moments of inertia of species E along the principal axes, and $\sigma_E$ is its symmetry number. Contributions from external degrees of freedom having been accounted for by eqs 12 and 13, the net binding free energy is expressed as

$$\Delta G^\circ = \Delta G^\circ_{tr} + \Delta G^\circ_{rot} - RT \ln\left[\frac{Z^{int}_{E^*I^*.aq} Q_w V}{Z^{int}_{E.aq} Z^{int}_{I.aq}}\right] + P\Delta V^\circ_{aq} \quad (14)$$

"For the process of bringing E and I from fixed positions and orientations at infinite separation to a final fixed position and orientation of the complex E*I*, i.e., freezing the translational and rotational degrees of freedom of the reactants and products ",[56] both $\Delta G^\circ_{tr}$ and $\Delta G^\circ_{rot}$ are individually zero.

**Approximation 2**. **Formal Separation of Solvent Effects from Internal Motions of the Solute**.

$$Z^{int}_{E.aq} = Z_E^{vib.conf} Z_E^{solvn}$$

$$Z^{int}_{E.aq} = \int\!\!\dots\!\int \exp\left(\frac{-E(\mathbf{X}^N_E, \mathbf{X}^M_W)}{k_B T}\right) d\mathbf{X}^N_E \, d\mathbf{X}^M_W \simeq \int\!\!\dots\!\int \exp\left[\frac{-\{(E(\mathbf{X}^N_E) + E(\mathbf{X}_E^{Nfixed}, \mathbf{X}^M_W)\}}{k_B T}\right] d\mathbf{X}^N_E \, d\mathbf{X}^M_W \quad (15)$$

$$\simeq \int\!\!\dots\!\left\{\int \exp\left[\frac{-E(\mathbf{X}^N_E)}{k_B T}\right] d\mathbf{X}^N_E\right\}\left\{\int\!\!\dots\!\int \exp\left[\frac{-E(\mathbf{X}_E^{Nfixed}, \mathbf{X}^M_W)}{k_B T}\right] d\mathbf{X}^M_W\right\} \quad (16)$$

Equations similar to eqs 15 and 16 can be written for I and E*I* and converted to excess free energies. Such a separation allows

$$\Delta G^\circ = \Delta G^\circ_{tr} + \Delta G^\circ_{rot} + \Delta G^\circ_{intra} + \Delta G^\circ_{solvn} \quad (17)$$

Equation 17 forms the basis for "master equation"[12] methods. Further, if the internal degrees of freedom of the solutes E, I, and E*I* are frozen in addition to translational and rotational degrees of freedom, the binding free energy can be written as

$$\Delta G^\circ = \Delta E^\circ_g + \Delta G^\circ_{solvn} \quad (18)$$

an expression that is commonly employed in some

earlier theoretical work. The $P\Delta V°_{aq}$ term in eq 9 is often neglected in liquid-state work. Equations 9 and 18 constitute two extreme theoretical models for binding, with eq 17 falling between. Some prescriptions to alleviate approximations in eq 17 and to make the results correspond as closely as possible to eq 9 are the following: (i) compute the translational and rotational contributions in gas phase by forming a suitable thermocycle while adopting static structures such as from modeling, X-ray crystallography, or NMR; (ii) generate an ensemble of structures of the reactants and products separately in the solvent medium with structural inputs from either modeling or experiment and then apply eq 17.[61,62] This corresponds to a *post facto* analysis of the molecular dynamics trajectories. The former circumvents the objections concerning the extension of the Sackur−Tetrode equation for reactions in the liquid phase. The latter (a) helps convert $\Delta E°_g$ into $\Delta H°_g$ in eq 18 and (b) avoids decoupling internal motions of the solute from those of solvent because the simulation incorporates solvent explicitly in the development of the ensemble of structures. The problem then shifts to solvation energy estimates and whether an ensemble of structures consistent with the solution phase was considered in arriving at solvation model parameters. This problem may be less severe than it appears if the calibration of solvation parameters is performed against experimental results.

Some areas for further refinement of the binding free energy theory are (i) the vibrational and configurational entropies that contribute to the $T\Delta S°_{intra}$ terms in eq 17, (ii) the environmental effects such as those due to counterion association and release, particularly in binding equilibria involving nucleic acids,[63,64] (iii) the "bound" water/water-mediated interactions,[65] (iv) complexes involving metal ions exhibiting charge transfer, etc.

## References


(1) Straatsma, T. P.; McCammon, J. A.; Computational Alchemy. *Annu. Rev. Phys. Chem*. **1992**, *43*, 407−435.

(2) Reddy, M. R.; Vishwanadhan, V. N.; Weinstein, J. N. Relative Differences in the Binding Free Energies of Human Immunodeficiency Virus-1 Protease Inhibitors: A Thermodynamic Cycle−Perturbation Approach. *Proc. Natl. Acad. Sci. U.S.A*. **1991**, *88*, 10287−10291.

(3) Ferguson, D. M.; Radmer, R. J.; Kollman, P. A. Determination of the Relative Binding Free Energies of Peptide Inhibitors to the HIV-1 Protease. *J. Med. Chem*. **1991**, *34*, 2654−2659.

(4) Cieplak, P.; Kollman, P. A. Peptide Mimetics as Enzyme Inhibitors: Use of Free Energy Perturbation Calculations to Evaluate Isosteric Replacement for Amide Bonds in a Potent HIV Protease Inhibitor. *J. Comput. Aided. Mol. Des.* **1993**, 7291−7304.

(5) Rao, B. G.; Murcko, M. A. Free Energy Perturbation Studies on Binding of A74704 and Its Diester Analog to HIV-1 Protease. *Protein Eng.* **1996**, *9*, 767−771.

(6) Warde, R. C.; McCammon, J. A. Binding of an antiviral agent to a sensitive and a resistant human rhinovirus. Computer Simulation Studies with Sampling of Amino Acid Side Chain Conformation. II. Calculation of Free Energy Differences by Thermodynamic Integration. *J. Mol. Biol.* **1992**, *225*, 697−712.

(7) Ota, N.; Stroupe, C.; Ferreira-da-Silva, J. M.; Shah, S. A.; Mares-Guia, M.; Brunger, A. T. Non Boltzmann Thermodynamic Integration for Macromolecular Systems: Relative Free Energy of Binding of Trypsin to Benzamidine and Benzylamine. *Proteins* **1999**, *37*, 641−653.

(8) Kamath, S.; Coutinko, E.; Desai, P. Calculation of Relative Binding Free Energy Difference of DHFR Inhibitors by a Finite Difference Thermodynamic Integration (FDTI) approach. *J. Biomol. Struct. Dyn*. **1999**, *16*, 1239−1244.

(9) Kollman, P. A. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395−2417.

(10) Beveridge, D. L.; Dicapua, F. M. Free Energy via Molecular Simulations: Applications to Chemical and Biochemical Systems. *Annu. Rev. Biophys. Biophys. Chem*. **1989**, *18*, 431−492.

(11) Van Gunsteren, W. F. Methods for Calculations of Free Energies and Binding Constants: Successes and Problems. In *Computer Simulations of Biomolecular Systems;* Van Gunsteren, W. F., Weiner, P. K., Eds.; ESCOM: Leiden, The Netherlands, 1989; pp 27−59.

(12) Ajay; Murcko, M. A. Computational Methods to Predict Binding Free Energy in Ligand Receptor Complexes. *J. Med. Chem*. **1995**, *38*, 4953−4967.

(13) Jayaram, B.; McConnell, K. J.; Dixit, S. B.; Beveridge, D. L. Free Energy Analysis of Protein−DNA Binding: The EcoRI Endonuclease−DNA Complex. *J. Comput. Phys*. **1999**, *151*, 333−357.

(14) Dill, K. A. Addivity Principles in Biochemistry. *J. Biol. Chem*. **1997**, *272*, 701−704.

(15) Cornell, W. D.; Cieplak, P.; Balyl, C. F.; Gould, I. R.; Kenneth, M. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc*. **1995**, *117*, 5179−5197.

(16) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Cheatham, T. E., III; Debolt, S.; Ferguson, D. M.; Seibel, G.; Kollman P. A. *Comput. Phys. Commun*. **1995**, *91*, 1

(17) Still, W. C.; Tempczyk, A.; Hawley, R. C.; Hendricksen, T. J. *J. Am. Chem. Soc*. **1990**, *112*, 6127.

(18) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem*. **1996**, *100*, 19824.

(19) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *Chem. Phys. Lett*. **1995**, *246*, 246.

(20) Jayaram, B.; Liu, Y.; Beveridge, D. L. A Modification of the Generalized Born Theory for Improved Estimates of Solvation Energies and p*K* Shifts. *J. Chem. Phys*. **1998**, *109*, 1465−1471.

(21) Jayaram, B.; Sprous, D.; Beveridge, D. L. Solvation Free Energy of Biomolecules: Parameters for a Modified Generalized Born Model Consistent with the AMBER Force Field. *J. Phys. Chem. B* **1998**, *102*, 9571−9576.

(22) Reddy, M. R.; Erion, M. D.; Agarwal, A.; Vishwanadhan, V. N.; McDonald, Q.; Still, W. C. Solvation Free Energies Calculated Using the GB/SA Model: Sensitivity of Results on Charge Sets, Protocols and Force Fields. *J. Comput. Chem*. **1998**, *19*, 769−780.

(23) Miller, M.; Schneider, J.; Sathyanarayana, B. K.; Toth, M. V.; Marshall, G. R.; Clawson, L.; Seth, L.; Kent. S. B. H.; Wlodawer, A. Structure of Complex of Synthetic HIV-1 Protease with a Substrate Based Inhibitor at 2.3 Å Resolution. *Science* **1989**, *246*, 1149−1152.

(24) Jaskolski, M.; Tomaselli, A. G.; Sawyer, T. K.; Staples, D. G.; Heinrihson, R. L.; Schneider, J.; Kent, S. B. H.; Wlodawer, A. Structure at 2.5 Å Resolution of Chemically Synthesized Human Immunodeficiency Virus Type-1 Protease Complexed with a Hydroxyethylene Based Inhibitor. *Biochemistry* **1991**, *30*, 1600−1609.

(25) Wlodawer, A.; Erickson, J. W. Structure Based Inhibitors of HIV-1 Protease. *Annu. Rev. Biochem*. **1993**, *62*, 543−585.

(26) Bartlett. J. G.; Moore, R. D. Improving HIV Therapy. *Sci. Am*. **1998**, July, 84−87.

(27) Navia, M. A.; Fitzgerald, P. M. D.; McKeever, B. M.; Leu, C. T.; Heimbach, C.; Herber, W. K.; Sigal, I. S.; Darke, P. L.; Springer, J. P. 3D Structure of Aspartyl Protease from Human Immunodeficiency Virus HIV-1. *Nature* **1989**, *337*, 615−620.

(28) Weber, I. T.; Miller, M.; Jaskolski, M.; Leis, J.; Skalka, A. M.; Wlodawer, A. Molecular Modelling of HIV-1 Protease and Its Substarte Binding Site. *Science* **1989**, *243*, 928−931.

(29) McKeever, B. M.; Navia, M. A.; Fitzgerald, P. M. D.; Springer, J. P.; Leu, C. T.; Heinibach, J. C.; Herber, W. K.; Sigal, I. S.; Darke, P. L. Crystallization of the Aspartyl Protease from the Human Immunodeficiency Virus HIV-1. *J. Biol. Chem*. **1989**, *264*, 1919−1921.

(30) Lapatto, R.; Blundell, T.; Hemmings, A.; Overington, J.; Wilderspin, A.; Wood, S.; Messon, J. R.; Whittle, P. J.; Dankey, D. E.; Geogheyan, K. F.; Hawrylik, S. J.; Lee, S. E.; Scheld, K. G.; Hobart, P. M. X-ray Analysis of HIV-1 Proteinase at 2.7 Å Resolution Confirms Structural Homology among Retroviral Enzymes. *Nature* **1989**, *342*, 299−302.

(31) Erickson, J. W.; Eissenstal, M. A. HIV-1 Protease as a Target for the Design of Antiviral Agents for AIDS. In *Proteases of Infective Agents*; Academic Press: New York, 1999; pp 1−59.

(32) Miller, M.; Jaskolski, M.; Mohana Roy, J. K.; Leis, J.; Wlodawer, A. Crystal Structure of a Retroviral Protease Proves Relationship to Aspartic Protease Family. *Nature* **1989**, *37*, 576−579.

(33) Silva, A. M.; Cachan, R. E.; Sham, H. L.; Erickson, J. W. Inhibition and Catalytic Mechanism of HIV-1 Aspartic Protease. *J. Mol. Biol*. **1996**, *255*, 321−346.

(34) Holloway, M. K.; Wai, J. M.; Halgren, T. K.; Fitzgerald, P. M.; Vacca, J. P.; Dorsey, B. D.; Levin, R. B.; Thompson, W. J.; Chen, L. J.; deSolms, S. J. A priori prediction of activity for HIV-1 protease inhibitors employing energy minimization in the active site. *J. Med. Chem.* **1995**, *38*, 305−317.
(35) Aqvist, J.; Medican, C.; Samuelsson, J. E. A New Method for Predicting Binding Affinity in Computer Aided Drug Design. *Protein Eng.* **1994**, *7*, 385−391.
(36) Hansson, T.; Aqvist, J. Estimation of Binding Free Energies for HIV Proteinase Inhibitors by Molecular Dynamics Simulations. *Protein Eng.* **1995**, *8*, 1137−1144.
(37) Lee, C. Y.; Yang, P. K.; Tzou, W. S.; Hwang, M. J. Estimates of Relative Binding Free Energies for HIV Protease Inhibitors Using Different Levels of Approximation. *Protein Eng.* **1998**, *11*, 429−437.
(38) Wang, W.; Wang, J.; Kollman, P. A. What Determines the van der Waals Coefficients $\beta$ in the LIE (Linear Interaction Energy) Method To Estimate Binding Free Energies Using Molecular Dynamics Simulations. *Proteins: Struct., Funct., Genet.* **1999**, *34*, 395−402.
(39) Dominy, B. N.; Brooks, C. L., III. Methodology for Protein−Ligand Binding Studies: Application to a Model for Drug Resistance, the HIV/FIV Protease System. *Proteins: Struct., Funct., Genet.* **1999**, *36*, 318−331.
(40) Verkhivker, G.; Appelt, K.; Fruer, S. T.; Villafranca, J. E. Empirical Free Energy Calculations of Ligand Protein Crystallographic Complexes. I. Knowledge Based Ligand Protein Interaction Potentials Applied to the Prediction of Human Immunodeficiency Virus-1 Protease Binding Affinity. *Protein Eng.* **1995**, *8*, 677−691.
(41) Kubinyi, H. QSAR and 3D-QSAR in Drug Design, Part 2: Applications and Problems. *Drug Des. Today.* **1997**, *2*, 538−546.
(42) Hirst, J. D. Nonlinear Quantitative Structure Activity Relationship for the Inhibition of Dihydrofolate Reductase. *J. Med. Chem.* **1996**, *39*, 3526−3532.
(43) Selassie, C. D.; Gan, W.; Kallander, L. S.; Klein, T. E. Quantitative Structure Activity Relationships of 2,4-Diamino-5-(2-X-Benzyl) Pyrimidines versus Bacterial and Aveian Dihydrofolate Reductase. *J. Med. Chem.* **1998**, *41*, 4261−4272.
(44) Clare, B. W. The Frontier Orbital Phase Diagrams: Novel QSAR Descriptors for Benzene Derivatives Applied to Phenylalkylamine Hallucinogens. *J. Med. Chem.* **1998**, *41*, 3845−3856.
(45) Altomare, C.; Cellamare, S.; Summo, L.; et al. Inhibition of Monoamine Oxidase B by Condensed Pyridazines and Pyrimidines: Effects of Lipophilicity and Structure Activity Relationships. *J. Med. Chem.* **1998**, *41*, 3812−3820.
(46) Massova, I.; Kollman, P. A. Computational Alanine Scanning To Probe Protein−Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.* **1999**, *121*, 8133−8143.
(47) Wang, J.; Morin, P.; Wang, W.; Kollman, P. A. Use of MM-PBSA in Reproducing the Binding Free Energies to HIV-1 RT of TIBO Derivatives and Predicting the Binding Mode to HIV-1 RT of Efavirenz by Docking and MM-PBSA. *J. Am. Chem. Soc.* **2001**, 123, 5221−5230.
(48) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(49) Schmidt, M. W.; Baldridge, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S. J.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. *J. Comput. Chem.* **1993**, *14*, 1347−1363.
(50) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. A Well Behaved Electrostatic Potential Based Method Using Charge Restraints for Determining Atom Centered Charges: The RESP Model. *J. Phys. Chem.* **1993**, *97*, 10269.
(51) Rao, B. G.; Baker, C. T.; Court, J. T.; Deininger, D. D.; Griffith, J. P.; Kim, E. E.; Kim, J. L.; Li, B.; Pazhanisamy, S.; Salituro, F. G.; Schairer, W. C.; Tung, R. D. Structure Based Design of Novel Conformationally Restricted HIV Protease Inhibitors. In *Rational Drug Design: Novel Methodology and Practical Applications*; ACS Symposium Series 719; American Chemical Society: Washington, DC, 1999.
(52) Lee, K.; Richards, F. M. The Interpretation of Protein Structures: Estimation of Static Accessibility. *J. Mol. Biol.* **1971**, *55*, 379−400.
(53) Kalra, P.; Das, A.; Dixit, S. B.; Jayaram, B. Free Energy Analysis of Protein−Drug Complexation: The Carboxypeptidase−Inhibitor Complexes. *Indian J. Chem.* **2000**, *39A*, 262−273.
(54) Kalra, P.; Das, A.; Jayaram, B. Free Energy Analysis of Protein−Drug Complexation: The Aspartic Proteinase−Inhibitor Complexes. *Appl. Biochem. Biotechnol.*, in press.
(55) Misra, V. K.; Honig, B. On the Magnitude of the Electrostatic Contribution to Ligand−DNA Interactions. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 4691−4695.
(56) Ben-Naim, A. In *Statistical Thermodynamics for Chemists and Biochemists*; Plenum: New York, 1992.
(57) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047−1069.
(58) Janin, J. Protein−Protein Recognition. *Prog. Biophys. Mol. Biol.* **1995**, *64*, 145−166.
(59) Janin, J. Elusive Affinities. *Proteins* **1995**, *21*, 30−39.
(60) Atkins, P. W. In *Physical Chemistry*; W. H. Freeman and Company: New York, 1990.
(61) Jayaram, B.; Sprous, D.; Young, M. A.; Beveridge, D. L. Free Energy Analysis of the Confomational Preferences of A and B Forms of DNA in Solution. *J. Am. Chem. Soc.* **1998**, *120*, 10629−10633.
(62) Kombo, D. C.; Jayaram, B.; McConnell, K. J.; Beveridge, D. L. Calculation of the Affinity of the $\lambda$ Repressor−Operator Complex Based on Free Energy Component Analysis. *Mol. Simul.*, in press.
(63) Manning, G. S. The Molecular Theory of Polyelectrolyte Solutions with Applications to the Electrostatic Properties of Polynucleotides. *Q. Rev. Biophys.* **1978**, *11*, 179−246.
(64) Jayaram, B.; Beveridge, D. L. Modelling DNA in Aqueous Solution: Theoretical and Computer Simulation Studies on the Ion Atmosphere of DNA. *Annu. Rev. Biophys. Biomol. Struct.* **1996**, *25*, 367−394.
(65) Otwinowski, Z.; Schevitz, R. W.; Zhang, R. G.; et al. Crystal Structure of Trp Repressor/Operator Complex at Atomic Resolution. *Nature* **1988**, *335*, 321.

JM010175Z