0273-2289/01/9601-3/0093/$14.00

Free-Energy Analysis of Enzyme-Inhibitor Binding

Aspartic Proteinase–Pepstatin Complexes

Parul Kalra, Achintya Das, and B. Jayaram*

Department of Chemistry, Indian Institute of Technology, Hauz Khas, New Delhi 110 016, India, E-mail: bjayaram@chemistry.iitd.ernet.in

Abstract

Expeditious *in silico* determinations of the free energies of binding of a series of inhibitors to an enzyme are of immense practical value in structure-based drug design efforts. Some recent advances in the field of computational chemistry have rendered a rigorous thermodynamic treatment of biologic molecules feasible, starting from a molecular description of the biomolecule, solvent, and salt. Pursuing the goal of developing and making available a software for assessing binding affinities, we present here a computationally rapid, albeit elaborate, methodology to estimate and analyze the molecular thermodynamics of enzyme-inhibitor binding with crystal structures as the point of departure. The complexes of aspartic proteinases with seven inhibitors have been adopted for this study. The standard free energy of complexation is considered in terms of a thermodynamic cycle of six distinct steps decomposed into a total of 18 well-defined components. The model we employed involves explicit all-atom accounts of the energetics of electrostatic interactions, solvent screening effects, van der Waals components, and cavitation effects of solvation combined with a Debye-Huckel treatment of salt effects. The magnitudes and signs of the various components are estimated using the AMBER parm94 force field, generalized Born theory, and solvent accessibility measures. Estimates of translational and rotational entropy losses on complexation as well as corresponding changes in the vibrational and configurational entropy are also included. The calculated standard free energies of binding at this stage are within an order of magnitude of the observed inhibition constants and necessitate further improvements in the computational protocols to enable quantitative predictions. Some areas such as inclusion of structural adaptation effects, incorporation of site-dependent amino acid pK_a shifts, consideration of the dynamics of the active site for fine-tuning the methodology are easily envisioned. The present series of studies, nonetheless, creates potentially useful qualitative information for design purposes on what factors favor protein-drug binding.

*Author to whom all correspondence and reprint requests should be addressed.

The net binding free energies are a result of several competing contributions with 6 of the 18 terms favoring complexation. The nonelectrostatic contributions (i.e., the net van der Waals interactions) and the differential cavitation effects favor binding. Electrostatic contributions show considerable diversity and turn out to be favorable in a consensus view for the seven aspartic proteinase-inhibitor complexes examined here. Implications of these observations to drug design are discussed.

Index Entries: Protein–ligand interactions; binding free energy; computer modeling.

Introduction

The concept of specificity is central to the action of enzymes, which are the functional units of cell metabolism. The recognition of a specific substrate by an enzyme from a myriad of molecular species leads to the proper conduct of physiologic activities and thus the sustenance of life. Genetic disorders can occur owing to a deficiency or a total absence of one or more enzymes in the tissues. In other abnormal conditions, the excessive activity of a specific enzyme can sometimes be controlled by a drug designed to inhibit its catalytic activity *(1)*. In such cases, estimation of the binding free energy of the drug with its receptor is a prerequisite to determine the potency of the drug. Recent advances that have been made in the description of intermolecular interactions using empirical force fields *(2,3)* and the development of a new methodology for obtaining estimates of the free energy of solvation simply but accurately using the "generalized Born-solvent accessibility" model *(4–9)* now enable us to determine the various energetic components involved in noncovalent associations.

In this article, we present a theoretical determination of the diverse free-energy contributors to binding in the complexes of aspartic proteinases with the natural inhibitor pepstatin and pepstatin-like transition state analogs *(10–14)* (Fig. 1A–E, Table 1). In addition, we highlight the potential utility of the methodology and results thus obtained for drug design efforts.

The aspartic proteinases form one of the major classes of proteolytic enzymes together with the serine, thiol, and metallo proteinases. Enzymes of this class are widely distributed, having been found in microorganisms, plants, and mammals. They have been isolated from stomach (pepsin, chymosin, and gastriesin), lysozome (cathepsin D), kidney and submaxillary gland (renin), and so on. A number of fungal enzymes, such as those from *Rhizopus chinensis*, *Endothia parasitica*, and *Penicillium janthinellum*, have also been isolated. All these enzymes are characterized by having two aspartyl residues in their catalytic apparatus and are inhibited by diazonorleucine methyl ester, epoxy (*p*-nitrophenoxy)-propane, and pepstatin, a hexapeptide from streptomyces. All these enzymes have molecular weights in the region of 35–42 kDa and pH optima for catalytic action in the range of 1.5–5.0 with the exception of renin, which operates at a slightly higher

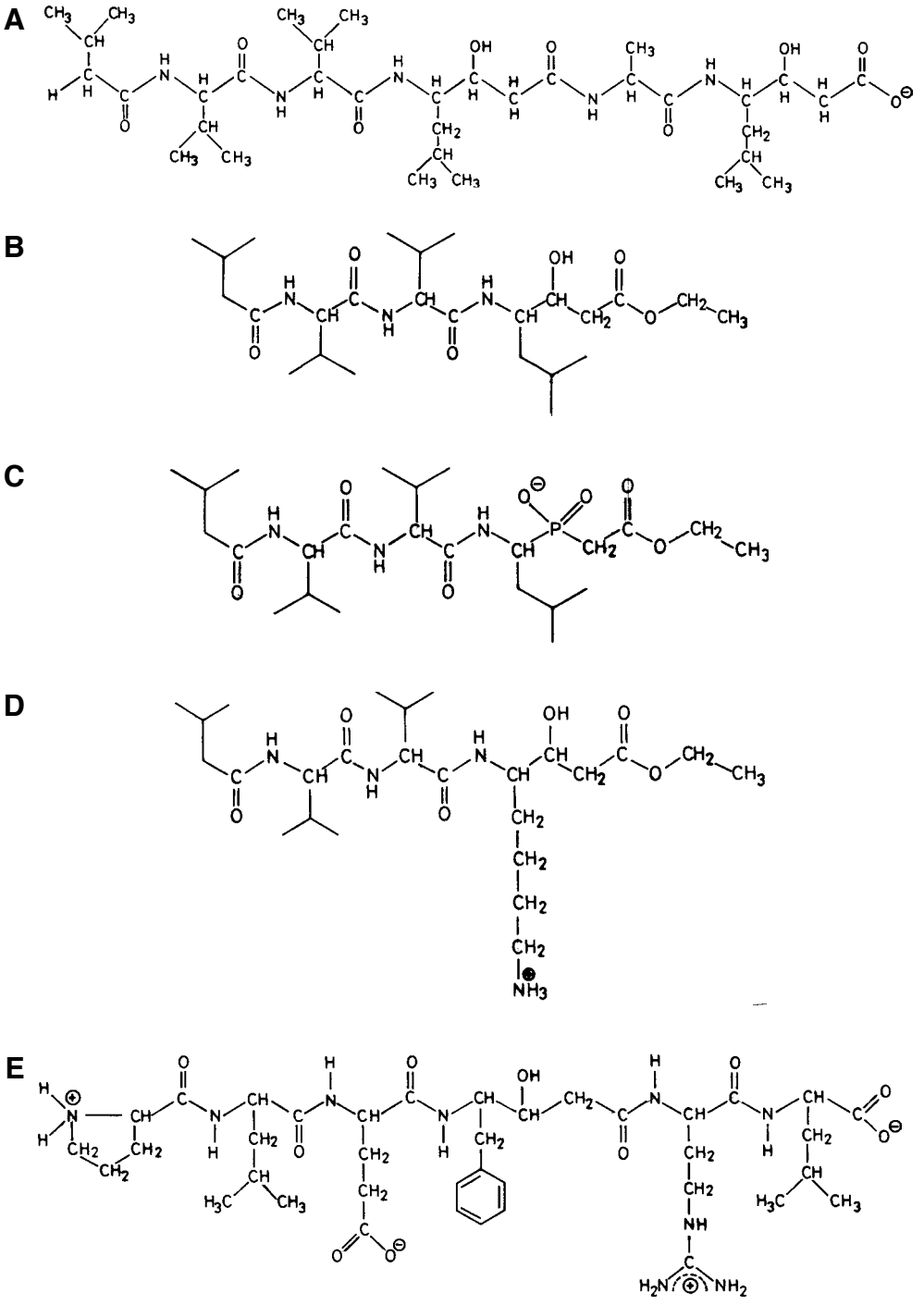

Fig. 1. Molecular structural formulae of the inhibitors of aspartic proteinases investigated: **(A)** Iva-Val-Val-Sta-Ala-Sta; **(B)** Iva-Val-Val-Sta-Oet; **(C)** Iva-Val-Val-Sta^{P}-Oet; **(D)** Iva-Val-Val-Lysta-Oet; **(E)** Pro-Leu-Glu-Psa-Arg-Leu.

Table 1
Aspartic Proteinase-Inhibitor Complexes Investigated and Their PDB Codes

PDB code	Name of complex
1APU	Penicillopepsin complexed with Iva-Val-Val-Sta-Oet
1APT	Penicillopepsin complexed with Iva-Val-Val-Lysta-Oet
1PPK	Penicillopepsin complexed with Iva-Val-Val-Sta^{P}-Oet
4ER2	Endothiapepsin complexed with pepstatin (Iva-Val-Val-Sta-Ala-Sta)
1PSO	Pepsin 3A complexed with pepstatin (Iva-Val-Val-Sta-Ala-Sta)
6APR	Rhizopuspepsin complexed with pepstatin (Iva-Val-Val-Sta-Ala-Sta)
1EPL	Endothiapepsin complexed with PS1 (Pro-Leu-Glu-Psa-Arg-Leu)

pH. They all possess essentially similar substrate specificity, preferring hydrophobic residues on both sides of the scissile bond. In addition, their catalytic action appears to involve extended subsite interactions with several amino acid side chains of the substrate.

X-ray diffraction analyses are in progress for several of the acid proteases, and three-dimensional structures have been determined for porcine pepsin *(15)*, pepsinogen *(16)*, chymosin *(17)*, cathepsin D *(18)*, renin *(19)*, and the three fungal proteinases, penicillopepsin *(20)*, endothiapepsin *(21)*, and rhizopuspepsin *(22)*. The aspartic proteinases are pharmacologically important in that they are associated with the onset of several pathologic conditions such as hypertension (renin), gastric ulcers (pepsin), molecular dystrophy, and neoplastic diseases (cathepsins D and E); thus, their inhibition is of considerable clinical interest. Recent discovery that the retroviral proteinases are also related to the pepsin-like aspartic proteinases has stimulated further interest, especially in view of the possibility that their inhibition could retard viral replication. They are also attractive targets for inhibitors as a consequence of their role in human pathogens such as malarial parasites and the fungus *Candida albicans (23–25)*.

No consensus exists on the mechanism of action of the carboxyl proteinases, despite the fact that they have been widely studied. Pepstatin, which is a natural inhibitor of acid proteinases, contains the unusual amino acid statine, and it has been proposed that this statyl residue is responsible for the unusual inhibitory capability of pepstatin. Studies on the mechanism of inhibition of this family of enzymes by pepstatin has led to the elucidation of the mechanism of catalysis. It is presumed that the manner in which pepstatin binds to the enzyme leads naturally to a model in which the substrate binds as an extended chain. Marciniszyn et al. *(26)* suggested that statine-4 residue of pepstatin could be an analog of the transition state for catalysis for the aspartic proteinases. The observed binding of pepstatin places the hydroxyl group on the tetrahedral C-3 atom of statine-4 in direct contact with the two catalytic aspartate carboxylates, Asp-35 and 220 (in rhizopuspepsin), with its side chains bound in a well-defined hydrophobic pocket. During catalysis *(10)*, the carboxylate group of Asp-220 donates its

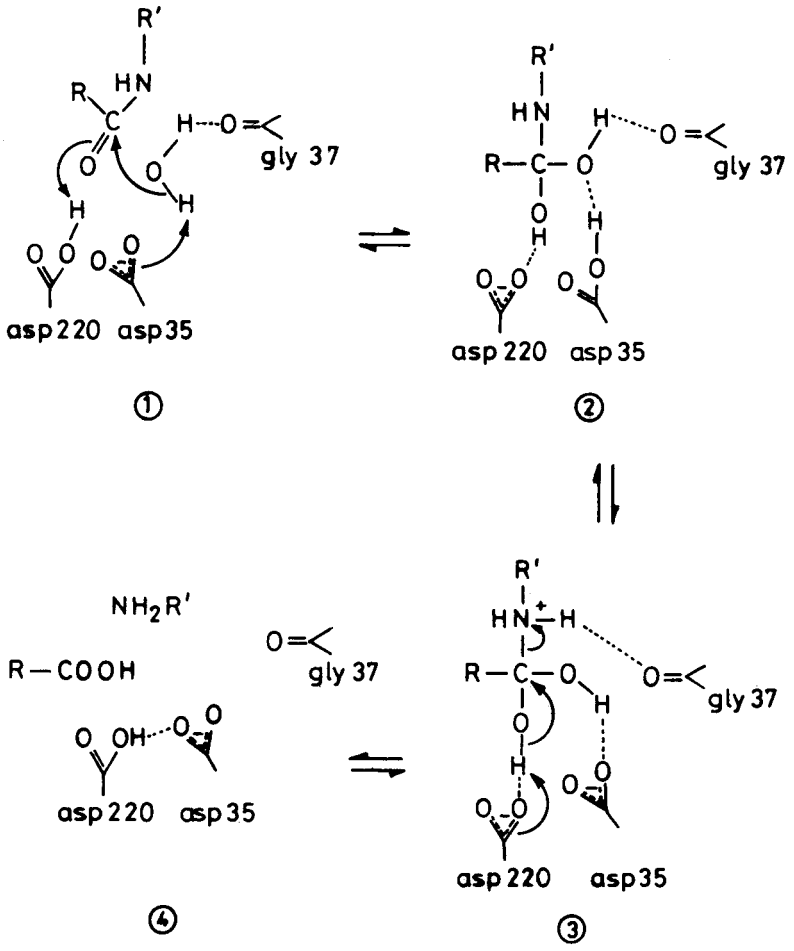

Fig. 2. Mechanism of catalysis of aspartic proteinases. Rhizopuspepsin is taken as an illustrative case. (Adapted from ref. *10*.)

proton to the carbonyl oxygen of the substrate while a water molecule loses a proton to Asp-35; when this proton is lost to the carbonyl oxygen of the substrate, the carboxyl of Asp-35 becomes more electronegative and more capable of abstracting a proton from a water molecule and vice versa. The resulting hydroxide ion undergoes a nucleophilic attack on the carbonyl carbon of the substrate, giving the first tetrahedral intermediate. This intermediate would be in equilibrium with a second intermediate in which the amide nitrogen is protonated. The transfer of a proton to the amide nitrogen could be mediated by the carbonyl oxygen of Gly-37, which is within hydrogen-bonding distance of the second hydroxyl oxygen and the amide nitrogen. The tetrahedral intermediate is unstable and the products are created by a single elimination step. The aforementioned mechanism, shown schematically in Fig. 2, is a general acid-base catalysis with no covalent intermediates.

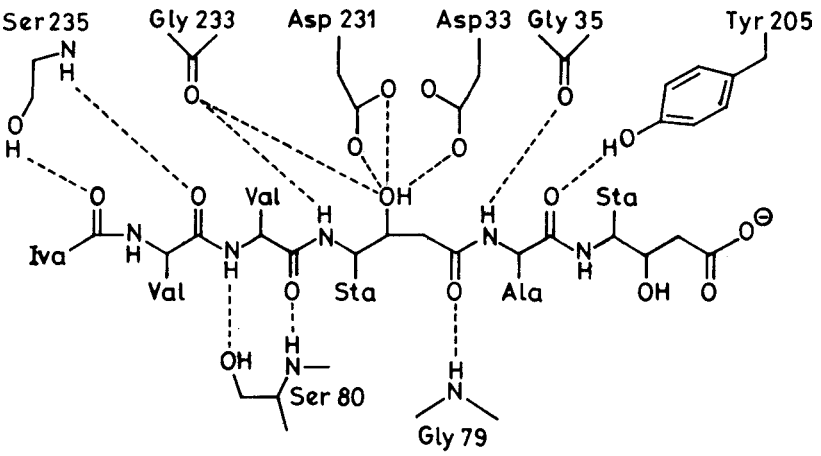

Fig. 3. Schematic representation of the residues in the active site of cathepsin D and the residues reported to interact with pepstatin. (Adapted from ref. *27*.)

The binding of pepstatin to acid proteases has been studied in order to predict the catalytic mechanism and to design new inhibitors. The complex of pepstatin with cathepsin D *(27)* is stabilized by numerous hydrogen bonds between backbone atoms of the inhibitor and both main chain and side chain atoms of the enzyme (Fig. 3). The central statine hydroxyl group occupies the position of a water molecule that interacts with the two active-site aspartate residues, and it is this (3S)-hydroxyl group on the P1 statine residue that contributes substantially to the binding affinity of cathepsin D for pepstatin. It has been found that the deoxy and R-hydroxy analogs are much weaker inhibitors. The (S)-OH configuration permits the formation of favorable hydrogen bonds to both Asp-33 and Asp-231. The P3' statine residue of the inhibitor makes no obvious hydrogen bond with the enzyme despite the presence of several donor and acceptor substituents. His-77 interacts electrostatically with the C-terminus of the inhibitor. The OH group of Tyr-205 is hydrogen bonded to the P2' carbonyl oxygen, and Ser-235 OG forms a hydrogen bond to the P4 carbonyl oxygen. The inhibitor side chains make extensive van der Waals contacts with the residues such as His-77, Tyr-78, Phe-126, Phe-131, Glu-14, Leu-303, Ile-220, Ile-311, Ile-229, Met-307, Tyr-78, and Glu-260 in the enzyme subsites S3, S2, S1, and S2'. The S1' and S3' subsites also interact with the C-terminal statine residue. The central statine residue of pepstatin can be considered to be a P1-P1' dipeptide isostere of Leu-Gly for which there is no P1' side chain substituent.

The crystal structures of the complexes of pepstatin and pepstatin fragments with other aspartic proteinases have also been reported *(10–14)*. In these structures, pepstatin adopts a very similar conformation with the central statine residue making the largest number of contacts and straddles between the plane of the two catalytic aspartate residues. Minor differences occur in hydrogen-bonding patterns owing to variation in the amino acid sequence. For example, if we compare the binding of pepstatin to cathepsin

D and rhizopuspepsin, Tyr-205 is replaced by Trp-194 in rhizopuspepsin, which points toward the P2' carbonyl group but does not form a hydrogen bond with the inhibitor. In the S3' subsite, the presence of His-77 in cathepsin D affords more contacts with pepstatin. There is a serine at this position in rhizopuspepsin. In cathepsin D, the presence of the smaller Ser-80 side chain permits the flap to move closer to the inhibitor and results in an increase in the number of van der Waals contacts between enzyme and inhibitor in the S2 subsite.

A study of the mode of binding of various inhibitors to an enzyme can be used for designing new inhibitors. The protease family of enzymes generally involve an acidic residue in their catalytic action. A study of the inhibitors of one class of enzymes can be exploited in terms of their efficiency of binding for another class. For example, the mode of binding of phosphonamidate and phosphonate inhibitors to carboxypeptidase A *(28)* and thermolysin *(29)*, respectively, led to the development of these inhibitors for penicillopepsin. A free-energy component analysis can then be carried out on these new inhibitors, and the factors that dominate the binding of a ligand can be elucidated. It is hoped that such studies will lead to the emergence of suggestions on new inhibitors that are of use in the pharmaceutical industry.

Theory and Methodology

In this study, the net standard free energy of binding is treated as a sum of a near-comprehensive set of individual contributions *(30,31)*. With the assumption of additivity *(32)* and an arbitrary although rational selection of terms, component analysis is not theoretically rigorous. One can expect at best only a semiquantitative account *(33,34)*, and hence, expectations must be framed accordingly *(35)*. However, for complex processes such as protein-inhibitor binding, no viable alternative currently exists, and simple identification of the important terms, estimates of their relative magnitudes, and determination of whether they make favorable or unfavorable contributions to the free energy of complexation provides potentially useful new knowledge.

The thermodynamic cycle for protein-inhibitor binding in solution used in this study is presented in Fig. 4. Here, the net binding process is decomposed into six steps. Step I is the process of converting the uncomplexed protein denoted P, to the form P^* in which the protein has adapted its structure to that of the inhibitor bound form. The free energy of this step is given as

$$\Delta G_{\mathrm{I}}^{0} = \Delta G_{1}^{adpt,P} \tag{1}$$

Step II is the corresponding structural adaptation of the inhibitor required to convert the uncomplexed form I to the complexed form I^* in solution. The free energy is thus

$$\Delta G_{\mathrm{II}}^{0} = \Delta G_{2}^{adpt,I} \tag{2}$$

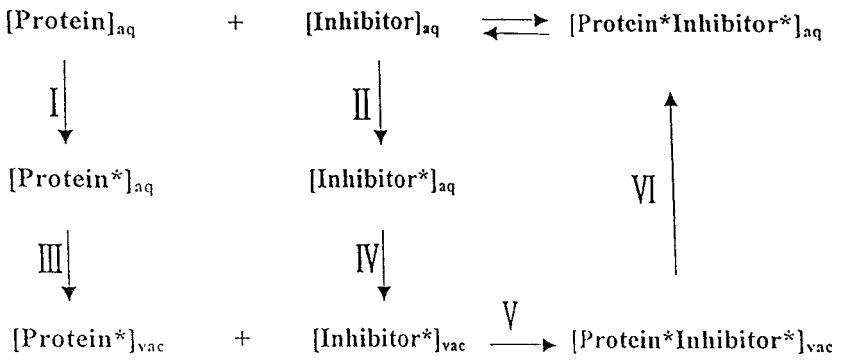

Fig. 4. Thermodynamic cycle used for a componentwise estimation and analysis of the binding free energies of enzyme-inhibitor complexes.

The next two steps (III and IV) involve desolvation of *P** a..nd *I** from aqueous medium to vacuum. The free energy of each of these steps is written as a sum of four components:

$$\Delta G^0_{\mathrm{III}} = \Delta G_3^{el,P} + \Delta G_4^{vdw,P} + \Delta G_5^{cav,P} + \Delta G_6^{DH,P} \tag{3}$$

$$\Delta G^0_{\mathrm{IV}} = \Delta G_7^{el,I} + \Delta G_8^{vdw,I} + \Delta G_9^{cav,I} + \Delta G_{10}^{DH,I} \tag{4}$$

This involves contributions from electrostatic effects of desolvating the macromolecule (*el*), the van der Waals interactions with the solvent (*vdw*), elimination of the solvent cavity (*cav*) in which the molecule is accommodated and the change in added salt effects (*DH*). The transfer from aqueous medium to vacuum in steps III and IV involves the loss of favorable electrostatics and van der Waals interactions with the solvent and a gain from the cavity term, the latter being, of course, the reverse of the free energy of cavity formation. The free energy of interaction with added salt is also lost on desolvation.

In step V, the protein and the inhibitor associate as a noncovalently bound complex. The thermodynamics of this step can be described as

$$\Delta G^0_{\mathrm{V}} = \Delta H_{11}^{el,C} + \Delta H_{12}^{vdw,C} - T\Delta S_{13}^{tr+rot} - T\Delta S_{14}^{vib+conf} \tag{5}$$

Complexation involves introducing the electrostatic and van der Waals interactions between the protein and the inhibitor in vacuo. A change in external entropy owing to loss of translational and rotational degrees of freedom occurs, which always disfavors complexation. The lost external modes are converted into low-frequency internal vibrational and configurational degrees of freedom in the complex and are reflected along with motional changes occurring as a consequence of the burial of amino acid side chains on complexation in the corresponding change in vibrational and configurational entropy *(36)*.

In step VI, the complex is transferred from vacuum back to aqueous solution, and the free energy change is owing to solvation of the complex:

$$\Delta G^{0}_{VI} = \Delta G^{el,C}_{15} + \Delta G^{vdw,C}_{16} + \Delta G^{cav,C}_{17} + \Delta G^{DH,C}_{18} \quad (6)$$

Here, again, an electrostatic component, a van der Waals component, a cavity formation component, and added salt effects are involved. Whereas the cavitation term is unfavorable to complexation, all the other terms are favorable.

In summary, the binding process in solution as considered here consists of six well-defined thermodynamic steps each of which can be decomposed into physically meaningful thermodynamic components. The total number of individual contributions to the free energy of binding in this model is 18. Further details on the evaluation of each component as well as the full theory and methodology for obtaining thermodynamic indices of macromolecular complexation are described in refs. *30* and *31*.

The atomic coordinates of rhizopuspepsin, endothiapepsin, pepsin 3A in complex with pepstatin, and the complexes of penicillopepsin with pepstatin fragments Iva-Val-Val-Sta-Oet, Iva-Val-Val-Sta^{P}-Oet, Iva-Val-Val-Lysta-Oet and that of endothiapepsin with PS1 were retrieved from the RCSB Protein Data Bank (Table 1). Our calculations are based on all-atom models for the enzyme and the inhibitors in which hydrogens are added explicitly to the crystal structure. AMBER partial charges are assigned to the enzyme atoms. 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 *(37)* and then fitting them with the RESP module of the AMBER 4.1 molecular modeling package *(2)*. In relation to our earlier study on carboxypeptidase-inhibitor complexes *(31)*, this is a new feature here. The ionization states of the charged residues in the inhibitor were determined on the basis of their isoelectric points. The N-terminal residue was taken to be positively charged and the C-terminus negatively charged. A more appropriate method would have been to determine the pK_a shifts of the charged residues of both the inhibitor and the protein. We defer this issue to a subsequent study. Energy minimization of the enzyme-inhibitor complex was then performed using the Sander module of AMBER in order to relieve any unfavorable clashes in the crystal structure. Here, 500 steps of minimization were carried out restraining heavy atoms (50 steps of steepest descent, *SD*, followed by 450 steps of conjugate gradient, *CG*), followed by a further 500 steps (50 *SD* + 450 *CG*) of free minimization using a sigmoidal distance-dependent dielectric function *(38,39)*. The electrostatic contribution to solvation was calculated via the generalized Born model using the effective radii parameters derived by Jayaram et al. *(8)* based on AMBER charges and sizes. The nonelectrostatic contribution to solvation, which involves molecular surface area calculations, was performed using the ACCESS

program based on the algorithm of Lee and Richards *(40)*. The added salt concentration employed in the Debye-Huckel term was 0.18 M, mimicking physiologic conditions. Results on each of the components enumerated as contributing to binding (Fig. 4) for each system are presented next.

Results

The calculated contributions to the standard free energy of binding for the various aspartic proteinase-inhibitor complexes studied are given in Table 2. As per standard conventions, negative values are favorable and positive values unfavorable to binding. The net free energy of binding is a result of various compensatory effects, mainly between (1) the internal and solvation electrostatics, (2) the direct van der Waals interactions between protein and inhibitor and loss in van der Waals interactions with the solvent, and (3) the cavitation effects and entropic losses of the enzyme and inhibitor molecules on complexation.

The net contribution of electrostatics, shape complementarity, hydrophobic effects, structural adaptation, and so on to the binding can be obtained from a combination of the values associated with the primary terms in Table 2. Any such combination, however, is nonunique. With this caveat, the contribution of structural adaptation to free energy includes terms 1 and 2 (Table 2). Similarly, the contribution of electrostatics (excluding the small ion effects) to the free-energy result can be expressed as the sum of terms 3, 7, 11, and 15. The van der Waals interactions, effectively the net energetics of shape complementarity, involve the terms 4, 8, 12, and 16. The total contribution of cavitation effects to the binding is the sum of terms 5, 9, and 17. The entropy change on complexation is described by the combination of terms 13 and 14. Small ion effects on free energy, owing to added salt in the model are obtained via the terms 6, 10, and 18. The sum of all these terms equals the net standard free energy of binding. An analysis of the results of Table 2 based on the net contributions from electrostatics, van der Waals, cavitation, entropy, and added salt effects is presented schematically in Fig. 5 as averages for all seven complexes.

Discussion

An exhaustive quantification of free-energy contributors to the binding process of aspartic proteinases with seven inhibitors was undertaken with a view to identifying forces favorable to complexation. Overall, the binding free energies indicate that the nonelectrostatic contributions (i.e., the van der Waals interactions) and the differential cavitation effects (i.e., hydrophobic forces arising owing to both nonpolar and polar atoms) are favorable for complexation (Table 3). The electrostatic contribution is favorable for some systems while unfavorable for others. The result that the net electrostatics is unfavorable for complexation for some complexes does not imply that electrostatics is unimportant. In considering the relative binding process of a series of molecular or macromolecular ligands,

Table 2
Calculated Values of Various Contributions to Standard Free Energy of Binding (kcal/mol) for Aspartic Proteinase-Inhibitor Complexes

Term	Component	1APU	1APT	1PPK	4ER2	1PSO	6APR	1EPL
	Step I: Structural adaptation of enzyme							
$\Delta G_1^{adpt,P}$	Free-energy change for the process $P \rightarrow P^*$	—	—	—	—	—	—	—
	Step II: Structural adaptation of inhibitor							
$\Delta G_2^{adpt,I}$	Free-energy change for the process $I \rightarrow I^*$	—	—	—	—	—	—	—
	Step III: Desolvation of enzyme							
$\Delta G_3^{el,P}$	Electrostatic component of P^* desolvation	4760.5	4704.5	4337.9	4860.1	9192.0	3754.8	4896.4
$\Delta G_4^{vdw,P}$	*vdw* component of P^* desolvation	505.4	505.3	509.6	533.3	530.4	529.8	537.0
$\Delta G_5^{cav,P}$	Cavity component of P^* desolvation	–596.8	–596.7	–601.8	–629.7	–626.4	–625.6	–634.1
$\Delta G_6^{DH,P}$	Loss of added salt interactions	46.1	45.7	42.7	39.2	83.8	32.4	39.2
	Step IV: Desolvation of inhibitor							
$\Delta G_7^{el,I}$	Electrostatic component of I^* desolvation	36.8	83.4	81.2	97.3	95.3	95.9	207.9
$\Delta G_8^{vdw,I}$	*vdw* component of I^* desolvation	33.4	33.3	33.2	41.6	42.3	41.3	46.7
$\Delta G_9^{cav,I}$	Cavity component of I^* desolvation	–39.4	–39.4	–39.2	–49.1	–50.0	–48.7	–55.2
$\Delta G_{10}^{DH,I}$	Loss of added salt interactions	0.3	0.6	0.6	0.7	0.7	0.7	1.3
	Step V: Complex formation in vacuo							
$\Delta H_{11}^{el,C}$	Electrostatic interactions between of P^* and I^*	–60.6	–534.1	245.1	175.5	349.0	100.4	–200.9
$\Delta H_{12}^{vdw,C}$	*vdw* interactions between P^* and I^*	–65.3	–64.7	–65.0	–78.8	–86.0	–83.1	–99.9
$-T\Delta S_{13}^{tr+rot}$	Rotational, translational entropy change	25.4	25.5	25.5	26.3	26.3	26.3	26.9
$-T\Delta S_{14}^{vib+conf}$	Vibrational, configurational entropy change	11.8	9.8	9.2	12.5	13.1	13.1	15.7
	Step VI: Solvation of complex							
$\Delta G_{15}^{el,C}$	Electrostatic component of complex solvation	–4741.2	–4304.3	–4654.0	–5146.8	–9639.1	–3950.4	–4894.2
$\Delta G_{16}^{vdw,C}$	*vdw* component of complex solvation	–492.4	–491.3	–496.3	–521.2	–516.0	–517.1	–521.7
$\Delta G_{17}^{cav,C}$	Cavity component of complex solvation	581.5	580.2	586.1	615.5	609.3	610.7	616.1
ΔG_{18}^{DH}	Added salt interactions with complex	–46.2	–42.9	–45.7	–41.8	–88.2	–34.3	–39.6
ΔG_{NET}^{0}	Net free energy of binding	–40.7	–85.1	–30.9	–65.4	–63.5	–53.8	–58.4

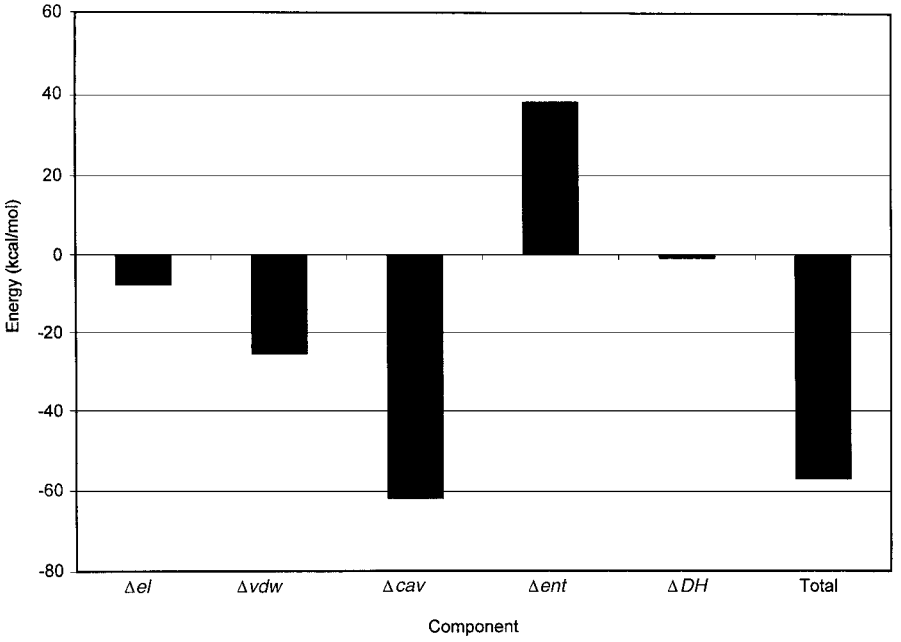

Fig. 5. Histogram of the calculated contributions to the binding free energy of the aspartic proteinase complexes.

Table 3
Conventional Combination of Computed Primary Terms Contributing to Net Free Energies of Binding (kcal/mol) of Aspartic Proteinase-Inhibitor Complexes

	1APU	1APT	1PPK	4ER2	1PSO	6APR	1EPL	Average
ΔG^0_{el}	–4.5	–50.5	10.2	–13.9	–2.8	0.7	9.2	–7.4
ΔG^0_{vdw}	–18.9	–17.4	–18.5	–25.1	–29.3	–29.1	–37.9	–25.2
ΔG^0_{cav}	–54.7	–55.9	–54.9	–63.3	–67.1	–63.6	–73.2	–61.8
$\Delta G^0_{entropy}$	37.2	35.3	34.7	38.8	39.4	39.4	42.6	38.2
ΔG^0_{DH}	0.2	3.4	–2.4	–1.9	–3.7	–1.2	0.9	–0.7
ΔG^0_{tot}	–40.7	–85.1	–30.9	–65.4	–63.5	–53.8	–58.4	–56.8
ΔG^0_{expt}	–12.1	–14.8	–10.5	—	—	—	—	—

differential effects of electrostatics may still be critical in the result. Salt effects are favorable for complexation in cases in which the inhibitor molecule is negatively charged (e.g., as with pepstatin) and unfavorable in cases in which the inhibitor is positively charged owing to charge screening. Note that the enzymes carry a net negative charge in all the systems considered. The effect of the net entropy change is to disfavor complexation in all seven complexes, which is expected since on formation of the complex

there occurs a motional restriction on the amino acid residues of the protein and the inhibitor.

A common feature among the three complexes of penicillopepsin-pepstatin fragments inferred from X-ray crystallography is that the hydroxyl group of the statine residue on the inhibitor makes hydrogen bonds to the side chain oxygen atoms of the catalytic aspartates and also to the Gly-76 and the Gly-215 main chain atoms of the enzyme. The valine residue at the second position on the inhibitor (see Fig. 1B–D) makes hydrogen bonds by means of its main chain NH and O atoms to Thr-217 and valine in the third position to Gly-76 and Asp-77. The net binding free energy thus is expected to be similar in these three cases. This, however, is not supported by either the experimental binding constant *(13,14,41)* data or the theoretical estimates carried out here. Additional ion-pair interactions in 1APT between the lysine side chain (substituting for statine side chain) and Glu-16 as well as Asp-115 lead to highly favorable electrostatics between the inhibitor and the enzyme (net electrostatics: –50.1 kcal/mol; Table 3). Between 1PPK and 1APU the phosphorous analog is experimentally observed to be a weaker inhibitor, and this also is reflected in our results on the net binding free energies.

In the complexes of pepstatin (4ER2, 1PSO, 6APR in Table 1 and Fig. 1A), a similar binding pattern for the central statine residue and the two valine residues is observed. The net van der Waals contributions are higher than those for the pencillopepsin complexes (Table 3) because the inhibitor is larger, leading to a tighter fit in the enzyme subsites. This is reflected in the additional interactions of the alanine residue with Gly-37 and the terminal statine with serine or threonine depending on the enzyme sequence. The cavitation effects parallel the van der Waals contributions in favoring complexation. In contrast to the other enzymes, in pepsin (1PSO, Table 3), it is threonine instead of aspartate that interacts with the valine main chain atoms of the inhibitor, leading to weaker hydrogen bonding interactions. The subtle amino acid sequence–dependent effects on the energetics are thus discernible.

In the carboxypeptidase A–inhibitor complexes *(31)* as well as in the protein-DNA complexes *(30)* examined earlier, the net van der Waals and the cavitation effects were favorable for binding and the entropy effects were unfavorable as seen here with the aspartic proteinases. The net electrostatics and the added salt effects in all cases showed considerable variation in both sign and magnitude. These observations suggest a molecular mechanism for macromolecular binding. Given the favorable van der Waals and hydrophobic components, which are typically size dependent, optimal binding (with desired affinity) could be accomplished via fine-tuning of the electrostatics, which includes both direct interactions and desolvation components.

Regarding the quantitative predictions, the calculated binding constants are an order of magnitude higher than the experimental values, pointing to the necessity for further work on improving the computational

protocols. By adhering strictly to the policy of deriving ab initio estimates of free energies based on a uniform atomic-level treatment of the systems (whether protein-DNA or enzyme-ligand) without any system-dependent parameterization, we envision the following refinements. The structural adaptation expenses (i.e., components 1 and 2 in Table 2) are not included in this study, because of the nonavailability of the native protein structure. The deformation energy in principle could be estimated via conducting molecular dynamics studies on the uncomplexed and complexed protein/inhibitor with explicit solvent molecules and then noting the difference in the energies of the two forms. Further improvements in the results are expected via forming averages of the individual energy components over the molecular dynamics trajectories. The molecular dynamics simulations offer the additional advantage of making the configurational entropies amenable to a more rigorous treatment. Also, estimation of the ionization states on the basis of isoelectric points could be improved on as the neighboring amino acids in the protein have a marked effect on the pK_a values. These can be predicted by carrying out an analysis of the pK_a shifts of the charged amino acid residues in the protein environment. Furthermore, the complete sequence of the sugar residues attached to the amino acids of the proteins such as penicillopepsin (xylose attached to Thr-7 in 1APU and 1APT) is missing in the crystal structure. The problem can be overcome by constructing these residues using computer modeling and then carrying out molecular dynamics on the system. The inclusion of sugar residues should improve the results for these two systems because there would be a restriction on the conformations that the protein could adopt. Work is in progress for a hierarchical elimination of the possible sources of error we have discussed.

The results we have described must be analyzed in the context of the expected uncertainties. Nevertheless, numerous new features can be learned from these studies, such as ideas about relative magnitudes and signs of various contributions, considerations of both initial and final states in estimating thermodynamic components, and, above all, the rapidity with which such binding free-energy calculations can be performed incorporating environmental effects. In this study, the objective was to demonstrate the potential of the methodology to conduct a reliable and computationally expeditious analysis of the energetics of protein-inhibitor complexation based on crystal structure data.

Conclusion

The net binding free energies for seven aspartic proteinase-inhibitor complexes were estimated using an all-atom treatment of the reactants and a methodology that is transferable across diverse systems. The net van der Waals and the electrostatic effects were found to be favorable with the contributions from cavitation effects paralleling the trend in van der Waals contributions. Salt effects were favorable whereas entropy effects were

unfavorable for complexation. The detailed enumeration of the free-energy contributors enables identification of forces—derived as compounded subsets of the primary contributors—favorable for complexation, thus aiding drug design initiatives.

Acknowledgment

We gratefully acknowledge funding from the Council of Scientific and Industrial Research, India.

References

1. Lehninger, A. L. (1990), *Principles of Biochemistry*, CBS Publishers and Distributors, New Delhi.
2. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Kenneth, M. M., Ferguson, D. M., Spellmeyer D. C., Fox, T., Caldwell J. W., and Kollman, P. A. (1995), *J. Am. Chem. Soc.* **117,** 5179–5197.
3. Pearlman, D. A., Case, D. A., Caldwell, J. W., Cheathnam, T. E., Debolt, S., Ferugson, D., Seibel, G., and Kollman, P. A. (1995), *Comput. Phys. Commun.* **91,** 1–41.
4. Still, W. C., Tempczyk, A., Hawley, R. C., and Hendricksen, T. J. (1990), *J. Am. Chem. Soc.* **112,** 6127–6129.
5. Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996), *J. Phys. Chem.* **100,** 19,824–19,839.
6. Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1995), *Chem. Phys. Lett.* **246,** 122–129.
7. Jayaram, B., Liu, Y., and Beveridge, D. L. (1998), *J. Chem. Phys.* **109,** 1465–1471.
8. Jayaram, B., Sprous, D., and Beveridge, D. L. (1998), *J. Phys. Chem. B* **102,** 9571–9576.
9. Reddy, M. R., Erion, M. D., Agarwal, A., Vishwanadhan, V. N., McDonald, Q., and Still, W. C. (1998), *J. Comput. Chem.* **19,** 769–780.
10. Bott, R., Subramanian, E., and Davies, D. R. (1982), *Biochemistry* **21,** 6956–6962.
11. Bailey, D. and Cooper, J. B. (1994), *Protein Sci.* **3(11),** 2129–2143.
12. Fujinaga, M., Chernaia, M. M., Tarasova, N. I., Mosimann, S. C., and James, M. N. (1995), *Protein Sci.* **4(5),** 960–972.
13. Fraser, M. E., Strynadka, N. C. J., Bartlett, P. A., Hanson, J. E., and James, M. N. G. (1992), *Biochemistry* **31,** 5201–5214.
14. James, M. N. G., Sielecki, A., Salituro, F., Rich, D. H., and Hofmann, T. (1982), *Proc. Natl. Acad. Sci. USA* **79,** 6137–6141.
15. Sielecki, A. R., Fedorov, A. A., Boodhoo, A., Andreeva, N. S., and James, M. N. G. (1990), *J. Mol. Biol.* **214,** 143–170.
16. James, M. N. G. and Sielecki, A. R. (1986), *Nature* **319,** 33–38.
17. Newman, M., Safro, M., Frazao, C., Khan, G., Zdanov, A., Tickle, I. J., Blundell, T. L., and Andreeva, N. (1991), *J. Mol. Biol.* **221,** 1295–1309.
18. Gulnik, S., Baldwin, E. T., Tarasova, N., and Erickson, J. (1992), *J. Mol. Biol.* **227,** 265–270.
19. Sielecki, A. R., Hayakawa, K., Fujinaga, M., Murphy, M. E. P., Bhasin, M., Mull, A. K., Carilla, C. T., Lewicki, J. A., Baxter, J. D., and James, M. N. G. (1989), *Science* **243,** 1346–1351.
20. James, M. N. G. and Sielecki, A. R. (1983), *J. Mol. Biol.* **163,** 299–361.
21. Blundell, T. L., Jenkins, J. A., Sewell, B. T., Pearl, L. H., Cooper, J. B., Tickle, I. J., Veerapandian, B., and Wood, S. P. (1990), *J. Mol. Biol.* **211,** 919–941.
22. Sugana, K., Bott, R. R., Padian, E. A., Subramanian, E., Sheriff, S., Cohen, G. H., and Davies, D. R. (1987), *J. Mol. Biol.* **196,** 877–900.
23. Wlodawer, A. and Erickson, J. W. (1993), *Annu. Rev. Biochem.* **62,** 543–585.
24. Cutfield, S. M., Dodsen, E. J., Anderson, B. F., Moody, D. C., Marshall, C. J., Sullivan, P. A., and Cutfield, J. F. (1995), *Structure* **3,** 1261–1271.

25. Francis, S. E., Banerjee, R., and Goldberg, D. E. (1997), *J. Biol. Chem.* **272,** 14,961–14,968.
26. Marciniszyn, J., Hartsuck, J. A., and Tang, J. (1976), *J. Biol. Chem.* **251,** 7088–7094.
27. Baldwin, E. T., Narayana, T. B., Sergei, G., Madhusoodan, V. H., Raymond, C. S., Cachau, R. E., Collins, J., Silva, A. M., and Erickson, J. W. (1993), *Proc. Natl. Acad. Sci. USA* **90,** 6796–6800.
28. Hidong, K. and Lipscomb, W. N. (1990), *Biochemistry* **29,** 5546–5555.
29. Holden, H. M., Tronrud, D. E., Monzingo A. F., Weaver, L. H., and Matthews, B. W. (1987), *Biochemistry* **26,** 8542–8553.
30. Jayaram, B., Mc Connell, K. J., Dixit, S. B., and Beveridge, D. L. (1999), *J. Comp. Phys.* **151,** 333–357.
31. Kalra, P., Das, A., Dixit, S. B., and Jayaram, B. (2000), *Ind. J. Chem. Sect. A*, **39A,** 262–273.
32. Dill, K. A. (1997), *J. Biol. Chem.* **272,** 701–704.
33. Janin, J. (1995), *Proteins* **21,** 30–39.
34. Janin, J. (1995), *Prog. Biophys. Mol. Biol.* **64,** 145–156.
35. Novotny, J., Bruccoleri, R. E., Davis, M., and Sharp, K. A. (1997), *J. Mol. Biol.* **268,** 401–411.
36. Gilson, M. K., Given, J. A., Bush, B. L., and McCammon, J. A. (1997), *Biophys. J.* **72,** 1047–1069.
37. 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., and Montgomery, J. A. (1993), *J. Comput. Chem.* **14,** 1347–1363.
38. Arora, N. and Jayaram, B. (1998), *J. Phys. Chem.* **102,** 6139–6144.
39. Arora, N. and Jayaram, B. (1997), *J. Comput. Chem.* **18,** 1245–1252.
40. Lee, K. and Richards, F. M. (1971), *J. Mol. Biol.* **55,** 379–406.
41. James, M. N. G. and Sielecki, A. R. (1985), *Biochemistry* **24,** 3701–3713.