PROTEINS: Structure, Function, and Genetics 33:295–310 (1998)

Predicting Structural Effects in HIV-1 Protease Mutant Complexes With Flexible Ligand Docking and Protein Side-Chain Optimization

Lana Schaffer and Gennady M. Verkhivker*
Agouron Pharmaceuticals, Inc., La Jolla, California

ABSTRACT **We present a computational approach for predicting structures of ligand-protein complexes and analyzing binding energy landscapes that combines Monte Carlo simulated annealing technique to determine the ligand bound conformation with the dead-end elimination algorithm for side-chain optimization of the protein active site residues. Flexible ligand docking and optimization of mobile protein side-chains have been performed to predict structural effects in the V32I/I47V/V82I HIV-1 protease mutant bound with the SB203386 ligand and in the V82A HIV-1 protease mutant bound with the A77003 ligand. The computational structure predictions are consistent with the crystal structures of these ligand-protein complexes. The emerging relationships between ligand docking and side-chain optimization of the active site residues are rationalized based on the analysis of the ligand-protein binding energy landscape. Proteins 33:295–310, 1998.** © 1998 Wiley-Liss, Inc.

Key words: molecular recognition; Monte Carlo docking; dead-end-elimination; rotamer library; correlated energy landscapes

INTRODUCTION

Theoretical insights into mechanisms of molecular recognition require determination of the lowest energy structure of a ligand-protein complex given the unbound conformations of the ligand and protein that is commonly known as the docking problem.[1–5] Molecular recognition models used in docking studies are based either on surface complementarity[6–13] or on atom-atom representations of the intermolecular interactions.[14–17] Adequate representation of the ligand-protein energetics imposes a thermodynamic requirement on the molecular recognition energy function, whereas the need to determine consistently the global free energy minimum on the binding energy landscape represents a kinetic aspect of the docking problem. The quality of the molecular recognition energy function used in docking simulations is usually evaluated exclusively by its thermodynamic properties, namely the location of the global minimum while the kinetic requirements of the docking problem are addressed by searching algorithms that explore the conformational space.

A variety of optimization techniques including Monte Carlo methods[18–21] and genetic algorithms[22–24] focused primarily on molecular docking of flexible ligands into proteins which are held fixed in a bound conformation, while the internal degrees of freedom of the ligand and its rigid body variables are optimized.[25] Combined flexible ligand docking and protein side-chain optimization techniques have been recently proposed in molecular recognition studies.[26–28] A variant of the dead-end elimination (DEE) algorithm has been used to avoid a combinatorial explosion by restricting both the ligand and the side-chains of the receptor residues to a limited number of discrete low-energy conformations.[27] However, the nonrelevance of the gas phase conformations of the isolated ligand to the favorable bound conformations in the native complex was the major drawback of that approach. The combinatorial problem in flexible peptide docking with major histocompatibility complexes receptors was also approached by using the DEE algorithm to optimize protein side-chains that adopt to the docked peptide conformations.[28] The DEE approach was successfully applied for protein side-chain placement in conjunction with Monte Carlo sequence optimization in a protein design automation cycle.[29]

There have been significant efforts in using side-chain prediction methods[30–42] to analyze structural and thermodynamic effects of mutations in proteins.[43–45] Side-chain placement approaches in homology modeling typically examine the effect of mutations on the protein structure by predicting side-chain conformations of the mutated residues starting from an approximation of the backbone framework of the wild-type protein. However, crystal structures of T4 lysozyme mutants have revealed a combination of main-chain and side-chains shifts with only marginal deviations from standard rotamer positions, whereas the computational prediction suggested a

*Correspondence to: Gennady M. Verkhivker, Agouron Pharmaceuticals, Inc., 3301 North Torrey Pines Court, La Jolla, CA 92037. E-mail: verk@agouron.com
Received 24 April 1998; Accepted 19 June 1998

more pronounced change in the side-chain torsional angles with the fixed backbone.[46,47] Modeling studies of some of the T4 lysozyme mutants suggested that the detailed torsional flexibility of side-chains not represented in rotamer libraries may alleviate uncertainties associated with the fixed backbone.[47]

Energy functions used in side-chain modeling studies may range from highly simplified coarse-grained models to complete molecular mechanics force fields.[39–42] Simplified energy functions are typically sufficient for predicting positions of buried hydrophobic groups, whereas the addition of electrostatic and hydrogen bond interactions may offer only a marginal improvement for predicting side-chain positions of polar residues. Structure prediction methods that used rotamer libraries of side-chains interacting only with the backbone appear to give results as accurate as when side-chain pairwise interactions are included.[33] This suggested that knowledge of only the local environment of the residue may be sufficient to predict the global energy minimum conformation of a given side-chain and thereby the combinatorial explosion may not be the bottleneck of the side-chain prediction methods. Furthermore, for a set of 11 proteins and 3 different rotamer libraries using the AMBER force field and the DEE search algorithm, Tanimura et al.[37] discovered that both side-chain-backbone and side-chain-side-chain interactions work concurrently to stabilize native protein conformations, which is consistent with the principle of minimal frustration.[48,49] The recent comparison of side-chain predictions for native proteins and randomly constructed protein systems analogous to spin glasses demonstrated that the reduction in rotamer combinations in the course of the DEE procedure is dramatically larger for the native proteins.[50] It was suggested that the correlated character of the energy landscape for the native proteins when the backbone is fixed in its native configuration is precisely what makes the DEE algorithm so effective.

In this work, we present a hierarchical computational approach for predicting structures of ligand-protein complexes that combines Monte Carlo simulated annealing technique to determine the ligand bound conformation with the DEE algorithm for side-chain optimization of the protein active site residues. Flexible ligand docking and optimization of mobile protein side-chains have been performed to predict structural effects in the V32I/I47V/V82I HIV-1 protease mutant bound with the SB203386 ligand and in the V82A HIV-1 protease mutant bound with the A77003 ligand. The effects of important drug resistance-conferring mutations in the active site of HIV-1 protease were recently studied on the basis of high-resolution crystal structures.[51,52] In particular, the crystal structures of both the wild-type HIV-1 protease complex and the V32I/I47V/V82I triple mutant complex were solved with the SB203386 inhibitor.[53] A structure comparison revealed a deviation of only 0.57 Å root mean square (rms) for all C_α atoms of the protein and 0.55 Å rms for the conformation of the bound inhibitor. The crystal structure of another mutant complex, V82A HIV-1 protease mutant with the A77003 ligand,[54] revealed significant deviations from the expected position of the inhibitor. Although the overall rms of the C_α atoms of the wild-type and mutant complexes is only 0.2 Å, significant main-chain and side-chains movements were observed for residues 81–84 and 181–184 in this crucial region of the active site of 0.6 Å and of 0.9 Å, respectively. Subtle changes not depicted in the initial modeling studies[54] were observed for the P1 and $P1'$ phenyl groups, a movement toward the residue 82 and away from the residue 182 by 0.4 Å.

A comprehensive understanding of the molecular basis of HIV-1 protease drug resistance implies not only the ability to predict structures of the ligand-protein complexes, but also to rank binding affinity of the ligands and subsequently design novel inhibitors that retain high affinity against mutant proteases. A simple estimate of the binding free energy differences for the V82A HIV-1 protease mutant with the A77003 ligand[54] based on the packing interaction energies for the crystal structures of the wild-type and mutant complexes has resulted in the asymmetric, single-sided reduction in van der Waals interactions of the $S1'$ subsite, consistent with the observed 4-fold reduction in binding affinity of the V82A mutant complex. The binding free energy changes on a simple hydrophobic mutation in inhibitor-protein complexes with HIV-1 protease may be evaluated within a model that includes the effects of solvation, cavity formation, conformational entropy, and mean field ligand-protein interactions.[55,56] However, the 6-fold reduction in the SB203386 binding affinity on the V32I/I47V/V82I mutation accompanied by subtle cumulative structural changes of both the ligand and the protein as revealed in the crystal structures of the wild-type and the mutant complexes would require a detailed binding free energy analysis that extends beyond the scope of this work and will be reported elsewhere.

METHODS

Molecular Recognition Energy Model in Ligand Docking

A family of molecular recognition energy functions that satisfy both thermodynamic and kinetic requirements in docking simulations by reducing frustration of the underlying binding energy landscape was recently designed.[57–63] Robust structure prediction of bound ligands given a fixed conformation of the native protein was achieved with this family of energy functions by generating binding energy landscapes with coexisting correlated, funnel-like[48,49] and uncorrelated, rugged features. The knowledge-

based energetic model includes intramolecular energy terms for the ligand, given by torsional and nonbonded functions,[63] and intermolecular ligand-protein steric and hydrogen bond interaction terms calculated from a piecewise linear potential summed over all protein and ligand heavy atoms.[23] The parameters of the pairwise potential depend on the four different atom types: hydrogen bond donor, hydrogen bond acceptor, both donor and acceptor, and nonpolar.[57–63] The ligand and protein atoms interact via steric, hydrogen bond, and repulsive potentials, which have the same functional form, with an additional three-body contribution to the hydrogen bond term (Fig. 1).

This function is designed primarily to predict structures of the ligand-protein complexes rather than provide precise binding energetics, and the parameters have been refined to yield the experimental crystallographic structure of a set of ligand-protein complexes as the global energy minimum.[57–63] The resulting energy landscapes are an intermediate representation between the highly simplified description accorded by minimalist lattice models and the more realistic yet computationally demanding all-atom force fields.

No explicit tuning for HIV-1 protease or for any of the ligands analyzed in this study have been performed, and no assumptions regarding either favorable ligand conformations or any specific ligand-protein interactions have been made. Monte Carlo simulated annealing docking simulations are performed for the wild-type HIV-1 protease and its mutants, with the studied inhibitors in a parallelepiped that encompasses the binding site obtained from the structure of the respective crystallographic ligand-protein complex with a 2 Å cushion added to every side of this box; the intermolecular potential is precalculated on a 0.2 Å grid that covers the protein binding site. A constant energy penalty of 200.0 energy units is added to every ligand atom outside the box.

The allowed rotatable bonds include bonds linking sp^3 hybridized atoms to either sp^3 or sp^2 hybridized atoms and single bonds linking two sp^2 hybridized atoms. The ligand bond distances, bond angles, and the torsional angles of the unrotated bonds were obtained from the crystal structures of the bound ligand-protein complexes. Crystallographic water molecules are included in the simulations as part of the protein structure.

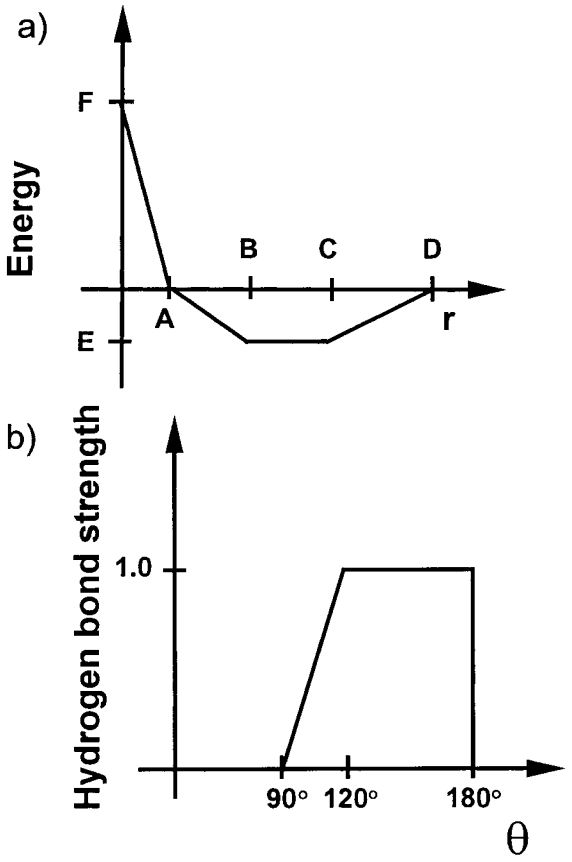

Fig. 1. (**a**) The functional form of the ligand-protein interaction energy. For steric interactions, A = 0.93B, C = 1.25B; D = 1.5B, E = −0.4, F = 15.0, and B = $r_l + r_p$ is the sum of the atomic radii for the ligand and protein atoms. The atomic radius for carbon, oxygen, nitrogen atoms is 1.8 Å and for sulfur is 2.2 Å. For hydrogen bond interactions, A = 2.3, B = 2.6, C = 3.1, D = 3.4, E = −2.0, F = 15.0, and for repulsive interactions, A = 3.4, E = 0.0, F = 15.0, and B, C, D are not relevant. A multiplicative desolvation penalty of 0.5 is applied to the attractive portion of the interaction between nonpolar and polar atoms. The units of A, B, C, and D are Å; for E and F the units are kcal/mole. (**b**) The hydrogen bond interaction energy is multiplied by the hydrogen bond strength term, which is a function of the angle θ determined by the relative orientation of the protein and ligand atoms. θ is defined to be the angle between two vectors, one of which points from the protein atom to the ligand atom. For protein atoms with a single heavy atom neighbor, the second vector connects the protein atom with its heavy atom neighbor, whereas for protein atoms with two heavy atom neighbors, it is the bisector of the vectors connecting the protein atom with its two neighbors.

Monte Carlo Ligand Docking Protocol

In ligand-protein docking simulations, the protein is fixed in its bound conformation while rigid body degrees of freedom and torsional angles of the ligand are chosen as independent variables. The conformational space of the ligand is explored by a series of uniform random displacements for each degree of freedom. To ensure efficient sampling in this study, we used the dynamically optimized acceptance ratio method.[64] In this method, the maximum step sizes at each temperature are dynamically chosen to optimize the acceptance ratio, which is the ratio of accepted conformations to the total number of trial conformations. In this approach, the simulations are arranged in cycles, and after a given cycle i, where the average acceptance ratio for each degree of freedom j is $\langle P_j \rangle^i$, the step sizes σ_j^i for each degree of freedom are updated for cycle $i + 1$ according to the

formula[64]

$$\sigma_j^{i+1} = \sigma_j^i \frac{\ln [a\langle P_{ideal}\rangle + b]}{\ln [a\langle P_j\rangle^i + b]},$$

where $\langle P_{ideal}\rangle$ is the desired acceptance ratio, chosen to be 0.5. The parameters a and b are used to ensure that the step sizes remain well-behaved when the acceptance ratio approaches 0 or 1. They are assigned so that the ratio σ^{i+1}/σ^i is scaled up by a constant value s for $\langle P_j\rangle^i = 0$, and down by the same constant for $\langle P_j\rangle^i = 1$. Solving the equations

$$s^{-1} = \frac{\ln [a\langle P_{ideal}\rangle + b]}{\ln [b]}$$

$$s = \frac{\ln [a\langle P_{ideal}\rangle + b]}{\ln [a + b]}$$

with $s = 3$ yields a = 0.673 and b = 0.065.

Monte Carlo docking simulations have been performed at 10 different temperatures, T = 100,000, 52,442, 27,501, 14,422, 7,563, 3,966, 2,079, 1,090, 571, and 299 K. At all temperatures each simulation consists of 5,000 sweeps, where a sweep is defined as a single trial move for each degree of freedom of the system and the maximum step sizes are updated every cycle of 50 sweeps.

DEE Protocol

There are three components of the DEE protocol: a rotamer library of side-chains,[65,66] a variant of the DEE algorithm,[67–72] and an energy model for evaluation of the solutions. The original Ponder and Richards rotamer library contained 84 rotamer states distributed over 18 amino acids.[65] Absent rotamers have been supplemented with standard *gauche* and *trans* combinations[67] and populated with more χ_2 angles for Asp, N, Glu, Gln,[69] Phe, Tyr, Trp, and His.[72] A final library consisting of 859 rotamer states[72] is used in this study. The DEE algorithm is based on a criterion that allows mathematically rigorous elimination of so-called "dead-end" rotamers that are not consistent with the global minimum energy conformation (GMEC) and thereby leads to reduction of the conformational search space.[67]

The original application of the algorithm was extended to rotamer pairs and was then modified by the "fuzzy-end elimination" theorem,[68] which correctly describes the usage of the dead-end rotamer pairs (DEP) to reduce the conformational search space. The logical pairs theorem allows elimination of any rotamer $\mathbf{i_r}$ that forms a DEP with each of the rotamers of some residue **j**. The "enhanced dead-end elimination" represents a more effective method to reduce the search complexity and allows the exploitation of a more detailed rotamer library.[69]

The potential energy function that uses the AMBER/OPLS force field[73] is represented as follows:

$$E_{total} = E_{template} + \sum_i E(\mathbf{i_r}) + \sum_i \sum_j E(\mathbf{i_r j_s}), \quad i \neq j,$$

where $E_{template}$ is the self-energy of the template, $E(\mathbf{i_r})$ is the energy of the side-chain rotamer $\mathbf{i_r}$ of the active site residue **i** including its self-energy and the interaction with the template, and $E(\mathbf{i_r j_s})$ is the nonbonded pairwise interaction energy between rotamers $\mathbf{i_r}$ and $\mathbf{j_s}$ of the active site residues **i** and **j**. The template used in this study is the union of the protein main-chain atoms, the C_β atoms, the side-chain atoms of the protein residues that are not among the active site residues, and the ligand atoms in their corresponding docked conformation.

Before the DEE protocol, the rotamers that have an interaction energy with the template greater than a defined threshold value of 30 kcal/mol are eliminated. If the minimum interaction energy of a rotamer with all rotamers of some other residue is greater than 30 kcal/mol, this rotamer is eliminated. Finally, rotamer pairs are flagged as dead-ending (DEP) if their side-chain interaction energy is greater than the threshold value.

The DEE protocol is implemented by iterating the following 3 steps until no further DE rotamer or DE rotamer pair is identified.

Step 1. The enhanced DEE criterion[50,69] for a single rotamer $\mathbf{i_r}$ expresses the inequality

$$E(\mathbf{i_r}) - E(\mathbf{i_u}) + \sum_j \mathbf{min}_{\text{no DEP with } \mathbf{i_r}} [\mathbf{E(i_r j_s)} - \mathbf{E(i_u j_s)}] > \mathbf{0}, \quad \mathbf{i} \neq \mathbf{j}.$$

This states that rotamer $\mathbf{i_r}$ is dead-ending if its best total interaction energy is larger than the worst interaction energy of another rotamer $\mathbf{i_u}$.

Step 2. The DE rotamer pairs are identified and flagged according to the original DEE theorem.[67] The rotamer pair energy $\epsilon(\mathbf{i_r j_s})$ is defined as the sum of the self-rotamer energy and their interaction energy.

$$\epsilon(\mathbf{i_r j_s}) = E(\mathbf{i_r}) + E(\mathbf{j_s}) + E(\mathbf{i_r j_s})$$

The rotamer pair $(\mathbf{i_r j_s})$ is dead-ending if this inequality is satisfied:

$$\epsilon(\mathbf{i_r j_s}) + \sum_k \mathbf{min}_t [\epsilon(\mathbf{i_r k_t}) + \epsilon(\mathbf{j_s k_t})] > \epsilon[(\mathbf{i_u j_v})] + \sum_k \mathbf{max}_t [\epsilon(\mathbf{i_u k_t}) + \epsilon(\mathbf{j_v k_t})], \quad \mathbf{i, j} \neq \mathbf{k}.$$

Step 3. The logical pairs theorem is applied to eliminate rotamers $\mathbf{i_r}$, which form DEPs with all rotamers of some residue *j*. Moreover, for a residue with a single rotamer $\mathbf{i_r}$, all rotamers forming a DEP

with $\mathbf{i_r}$ are declared as dead-ending. Typically, the number of solutions left at the end of this step is between 1 and 3,000.

Summary of the Combined Computational Protocol

We propose a hierarchical computational approach that combines flexible ligand docking, optimization of mobile protein side-chains, local minimizations, and energy evaluations of the resulting solutions for predicting structural effects in the HIV-1 protease mutant complexes. An initial structure of the mutated protein is constructed from the crystal coordinates of the corresponding wild-type complex with mutated residues substituted using the DEE algorithm. Then 25–50 independent Monte Carlo simulated annealing simulations are performed to dock the ligand into a mutated protein generated at the first stage of the protocol. Monte Carlo simulated annealing method is used for ligand docking with a simplified energy function to efficiently explore the binding energy landscape. The ligand binding modes are searched in the active site of the mutated protein held fixed in the native backbone conformation of the wild-type complex. Each of the docked ligand conformations, which represents a manifold of accessible ligand binding modes, is used to generate the templates for a subsequent step of protein side-chain optimization with the DEE procedure using a detailed energy force field.

Local minimizations and energy evaluations of all the generated DEE solutions are performed at the final stage to adjust unfavorable interatomic contacts in the active site caused by the backbone position and the rotamer library. The energy of each ligand-protein complex is determined by calculating the energy of the ligand and protein residues whose side-chains reside within 5 Å proximity from at least one of the mutated residues. All the side-chains within this 5 Å radius that have not been optimized by the DEE procedure are placed in their corresponding crystallographic positions. The energy of each ligand-protein complex is then subjected to 25 steps of the Polak-Ribiere conjugate gradient minimization as implemented in the MacroModel molecular modeling software package.[74] The ligand-protein complex with the lowest energy, as measured in the AMBER force field[75] with the generalized Born and solvent-accessible surface area (GB/SA) solvation model,[76] is considered the final solution and determines the structure prediction for a given ligand-protein system.

We have implemented this combined computational approach with flexible ligand docking and subsequent optimization of mobile protein side-chains to predict and rationalize structural effects in two ligand-protein systems, the V32I/I47V/V82I HIV-1 protease mutant complex bound with the SB203386 ligand (Fig. 2a) and the V82A mutant complex bound with the A77003 ligand (Fig. 2b). Side-chain conformations of 20 protein active site residues are optimized in each DEE simulation. For all the complexes in this study, the residues that are optimized for both chains of HIV-1 protease are enumerated: R8/108, L23/123, V32/132, I47/147, I54/154, L76/176, T80/180, V82/182, N83/183, and I84/184. Only 20 residues are optimized to determine the accuracy of the procedure. The V32/132, I47/147, and V82/182 residues are mutated to the I32/132, V47/147, I82/182 residues in the mutant complex with SB203386 and the V82/182 residues are mutated to A82/182 in the mutant complex with A77003.

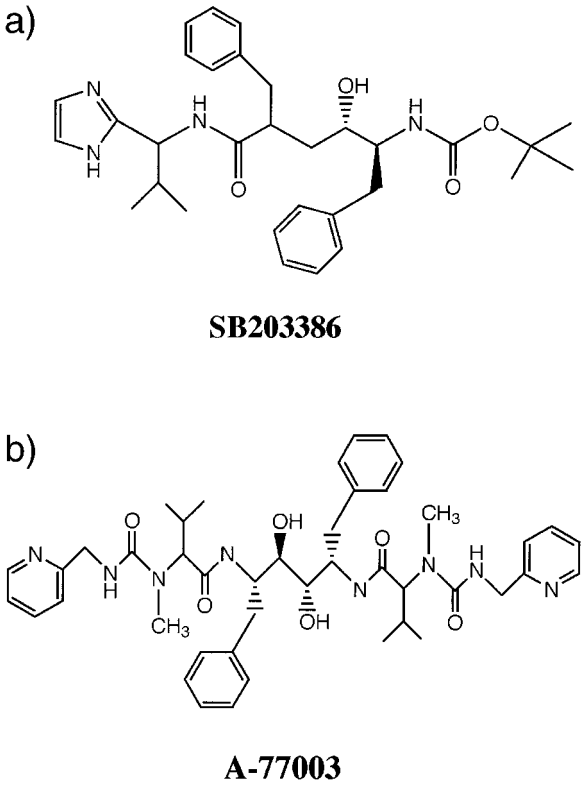

Fig. 2. Two-dimensional chemical structure representations of the SB203386 (**a**) and A77003 inhibitors (**b**).

RESULTS

The goal of this work focuses on predicting structural effects of mutations in the active site residues of HIV-1 protease on ligand binding. In this work, we address a number of issues that arose from the experimental and early modeling studies of HIV-1 protease recognition. Can structural effects of protein mutations be accurately assessed by ligand docking to the mutated protein starting from the wild-type protein? What is the accuracy of the approach in predicting structural effects of the HIV-1 protease mutant complexes when the backbone differences are less than 1 Å? What is the nature of the binding energy landscape for native complexes? Is

there a relationship between energies and structures of the ligand-protein complexes? How do protein side-chain positions respond to structurally different low-energy binding modes of the ligand? Are the side-chain conformations primarily determined by the field of the protein backbone or by the field of the ligand docked conformations, or both?

In the first set of computational experiments to validate the previous hypothesis, structural assessments of the wild-type HIV-1 protease complexes with SB203386 and A77003 ligands were performed. The lowest energy complexes of the docked ligand structure has an rms deviation of 1.04 Å and 0.92 Å from the crystallographic ligand conformations of SB203386 (13 rotatable bonds) and A77003 (17 rotatable bonds), respectively (Fig. 3a,c). Two different criteria were used to compare the modeled and crystal structures of the side-chains: the side-chain volume overlap[72] and a window of +/− 40° for χ_1 and χ_2 angles. We find that the protein side-chains are correctly positioned as observed by visual inspection of the crystallographic and predicted structures if the volume overlap is greater than the threshold of 65%. Using this criterion, the predicted side-chain conformations of the V32/132, T80/180, L23/123, and L176 residues for the wild-type HIV-1 protease complex with SB203386 do not agree with the crystal structure (Table I, Fig. 3a), and the side-chain volume overlap is only 65%. In the wild-type complex with A77003, side-chain conformations of residues V32, I47, T80, V82, R108, and T180 differ from the crystal structure of the complex. The rest of the active site residues show excellent agreement with the crystallographic positions (Table II, Fig. 3c), resulting in 70% correctly predicted residues.

The computational protocol was then used to predict structural effects for the HIV-1 protease mutants. In the lowest energy complex of the HIV-1 protease triple mutant with SB203386, the docked ligand conformation deviates by 0.99 Å from the ligand crystallographic position (Table III, Fig. 3b). Only 4 residues, L23/123 and V47/147, were mispredicted in this ligand-protein system, with the side-chain conformations of only one of the mutated residues, V47/147, not determined with a high degree of accuracy. The remaining 16 residues, including mutated I32/132 and I82/182, have an excellent level of prediction accuracy, resulting in 80% correctly predicted residues. The same quality of prediction was achieved for the V82A mutant complex with A77003, where an asymmetric pattern of misplaced V32, I47, and T180 residues was observed (Table IV, Fig. 3d).

Although the DEE procedure localizes the energy minima on the binding energy surface, side-chain conformations can be further refined by a local minimization to correct for errors caused by the limitations of the rotamer library approximation. We observe that there is practically no change in the fraction of the correctly predicted side-chain conformations as judged by inspection of the χ_1 and χ_2 angles for the active site residues in both wild-type (Tables V, VI) and mutant complexes (Tables VII, VIII), whereas the actual accuracy of prediction improves after refinement.

The results reveal a correlation between the energies of ligand-protein complexes and deviation of the corresponding ligand conformation in the native complexes (Tables I, II; Figs. 4, 5). We observed the phenomenon of a decrease in the energy of the complex when the unsymmetrical SB203386 ligand approaches 9 Å rms deviation from the crystal structure. This corresponds to a ligand binding mode that is the mirror image of the native conformation. This observation is not necessarily a flaw of the method but can be explained as a reflection of the dimeric nature of HIV-1 protease complexes. The crystallographic binding mode corresponds to the complex with the lowest energy and can also be distinguished from its "mirrored" alternative by its location in a wider energy funnel (Fig. 4).

We investigated the effect of the protein backbone and the ligand binding modes on the side-chain rotamer prediction. The results reveal that the protein side-chain conformations are not very sensitive to the orientations of the docked ligand (Tables I–IV). In the wild-type and mutant complexes with SB203386 there are no appreciable changes in the volume overlap of the predicted side-chain conformations with the crystal structure when the ligand docked conformation spans the range between 1 Å and 5 Å rms from its crystallographic position (Tables I, III). There is only one pair of residues, I82/182, whose volume overlap undergoes a significant change (from 50% to 85%) as the ligand docked conformation approaches the crystal structure (Table III). In the wild-type complex with A77003, L23 undergoes a change in the volume overlap from 55% to 80%, and the V132 residue alters its volume overlap from 55% to 87% as the protein accommodates to the docked conformation of the ligand to its crystallographic position (Table II). In the mutant complex with A77003, the L23/123 residues have improved the side-chain overlap from 65% to 80% (Table IV). Therefore, we observe accommodations of the residue side-chains directly involved in the ligand-protein contacts.

Comparison of the crystal structures between wild-type and mutant complexes reveals only minor changes in the crystallographic side-chain conformations (Tables V–VIII). In SB203386 complexes on the V32I/I47V/V82I mutation, for L23 the χ_2 angle changes from 197° to 70° and for L123 the χ_2 angle changes from 200° to 60°. Correspondingly, the same deviation of L23 side-chain as a result of mutation has also been observed in the crystal structures of the V82D and V82N of HIV-1 protease mutant complexes with the U-89360E inhibitor.[77] All other

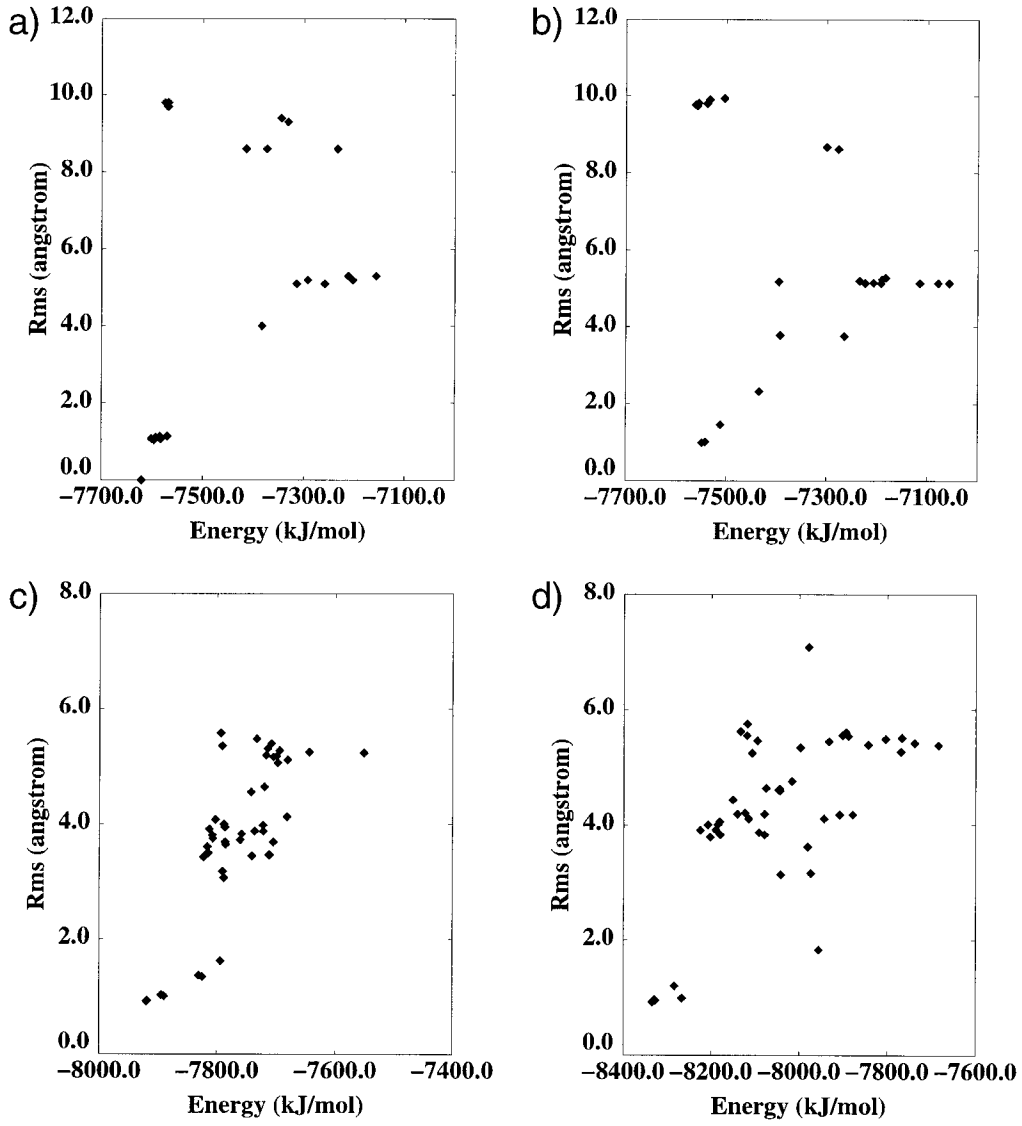

Fig. 3. Root mean square (rms) deviation of the predicted ligand conformation relative to the crystal structure versus energy for SB203386 in complexes with the wild-type protein (**a**) and the V32I/I47V/V82I mutant (**b**). Rms deviation of the predicted ligand conformation relative to the crystal structure versus energy for A77003 in complexes with the wild-type protein (**c**) and the V82A mutant (**d**).

significant side-chain deviations in the V82D and V82N mutant complexes are on the protein surface and have no contacts with the inhibitor. The predicted conformation for the L23/123 residues in complexes with SB203386 converges on an intermediate χ_2 angle of 120° that suffices for both wild-type and mutant structures but does not reflect the change observed in the crystal structures.

Prediction accuracy can be affected by the nature of crystallographic data, and there are cases in which a rotamer can be found in more than one conformation. According to the crystal structure of the wild-type HIV-1 protease complex with A77003, I84 can exist in either of two allowed conformations with $\chi_2 = -70°$ or 159°, whereas I184 is densely packed with the $\chi_2 = 159°$ conformer.[54] In the mutant complex with A77003, these residues have the same favorable rotamer with $\chi_2 = 158°$ for I84 and $\chi_2 = 176°$ for I184. In the predicted structures of the wild-type and mutant complexes with A77003, the I84/184 residues exhibit both $\chi_2 = 169°$ and $\chi_2 = -70°$ rotamer conformations (Tables VI, VIII). The predicted $\chi_2 = 169°$ for I184 in the wild-type complex agrees with the crystal structure where the tightly packed space allows only one conformer. However, the predicted $\chi_2 = -70°$ for I84/184 would indicate that both the experimentally observed conformer with $\chi_2 = 170°$ and computationally determined conformer with $\chi_2 = -70°$ may be allowed. Another assessment error may be introduced when the nitro-

TABLE I. Summary of the Structure Prediction Results for the Wild-Type HIV-1 Protease-SB203386 Complex†

Residue	Volume overlap (%)					
R8/108	75.8/76.2	75.8/76.2	75.8/76.2	75.8/76.2	75.8/76.2	56.0/76.2
L23/123	62.4/64.4	61.3/64.4	61.3/64.4	61.3/64.4	62.4/64.4	61.3/64.4
V32/132	51.6/55.3	51.6/55.3	51.6/89.6	51.6/55.3	51.6/55.3	51.6/55.3
I47/147	91.4/91.4	91.4/91.4	91.4/91.4	91.4/48.7	91.4/91.4	91.4/91.4
I54/154	87.5/93.4	87.5/93.4	87.5/93.4	87.5/93.4	87.5/93.4	87.5/93.4
L76/176	82.3/66.9	82.3/66.9	82.3/66.9	82.3/66.9	82.3/66.9	82.3/66.9
T80/180	53.1/53.8	53.1/53.8	53.1/53.8	53.1/53.8	53.1/53.8	53.1/53.8
V82/182	89.3/95.4	89.3/95.4	89.3/95.4	89.3/95.4	89.3/95.4	89.3/95.4
N83/183	80.5/90.6	80.5/90.6	80.5/90.6	80.5/90.6	80.5/90.6	80.5/90.6
I84/184	92.2/82.5	92.2/69.8	92.2/82.5	92.2/82.5	92.2/82.5	92.2/82.5
Ligand rms	1.04	1.14	4.0	5.1	5.3	9.8
Energy	−7595.0	−7568.8	−7383.0	−7258.5	−7212.0	−7574.3

†Volume overlap of the predicted protein side-chains for the active site residues and rms of the predicted ligand bound conformation relative to the crystal structure for a representative sample of 6 low-energy ligand binding modes.

TABLE II. Summary of the Structure Prediction Results for the Wild-Type HIV-1 Protease-A77003 Complex†

Residue	Volume overlap (%)						
R8/108	84.4/57.0	84.4/57.0	84.4/57.0	84.4/57.0	55.1/57.0	84.4/57.0	84.4/74.9
L23/123	79.3/82.6	79.3/82.6	79.3/82.6	54.8/82.6	79.3/82.6	71.0/61.0	54.8/61.0
V32/132	52.7/87.1	52.7/87.1	52.7/87.1	52.7/54.5	52.7/87.1	52.7/87.1	78.2/58.5
I47/147	53.2/72.7	53.2/72.7	53.2/72.7	71.2/53.3	71.2/72.7	71.2/72.7	71.2/72.7
I54/154	91.6/91.8	91.6/91.8	91.6/91.8	91.6/91.8	91.6/91.8	91.6/91.8	91.6/91.8
L76/176	77.6/82.5	77.6/82.5	77.6/82.5	77.6/82.5	77.6/82.5	77.6/82.5	77.6/82.5
T80/180	52.3/53.5	52.3/53.5	52.3/53.5	52.3/53.5	52.3/53.5	52.3/53.5	83.3/53.5
V82/182	58.5/55.1	58.5/86.8	88.9/55.1	54.8/55.1	88.9/55.1	88.9/55.1	88.9/58.9
N83/183	86.7/95.0	86.7/95.0	86.7/95.0	86.7/95.0	86.7/95.0	86.7/95.0	86.7/95.0
I84/184	75.0/72.5	75.0/92.9	75.0/72.5	88.7/92.9	75.0/92.9	75.0/92.9	75.0/92.9
Ligand rms	0.92	0.93	1.37	1.62	3.07	3.92	5.58
Energy	−7920.0	−7918.6	−7831.4	−7794.5	−7789.0	−7813.8	−7794.3

†Volume overlap of the predicted protein side-chains for the active site residues and rms of the predicted ligand bound conformation relative to the crystal structure for a representative sample of 7 low-energy ligand binding modes.

TABLE III. Summary of the Structure Prediction Results for the V32I/I47V/V82I Triple Mutant HIV-1 Protease-SB203386 Complex†

Residue	Volume overlap (%)					
R8/108	79.7/79.7	79.7/79.7	79.7/71.1	79.7/71.1	79.7/71.1	79.7/71.1
L23/123	61.9/61.9	59.4/59.4	59.4/60.8	61.9/60.8	61.9/60.8	59.4/60.8
I32/132	78.8/78.8	78.8/78.8	78.8/66.1	78.8/66.1	78.8/66.1	78.8/66.1
V47/147	53.3/53.3	64.0/64.0	64.0/50.6	53.3/50.6	53.3/50.6	53.3/50.6
I54/154	92.7/92.7	92.7/92.7	92.7/85.0	92.7/85.0	92.7/85.0	92.7/85.0
L76/176	82.3/82.3	82.3/82.3	82.3/67.3	82.3/67.3	82.3/67.3	82.3/67.3
T80/180	82.5/82.5	82.5/82.5	82.5/67.4	82.5/67.4	82.5/67.4	82.5/67.4
I82/182	85.6/85.6	50.2/50.2	50.2/52.6	85.6/49.7	53.5/49.7	53.5/49.7
N83/183	81.1/81.1	81.1/81.1	81.1/75.1	81.1/75.1	81.1/75.1	81.1/75.1
I84/184	90.8/90.8	90.8/90.8	90.8/84.4	90.8/84.4	90.8/84.4	90.8/84.4
Ligand rms	0.99	2.32	3.78	5.14	5.28	9.94
Energy	−7547.3	−7433.0	−7392.0	−7223.0	−7182.2	−7503.5

†Volume overlap of the predicted protein side-chains for the active site residues and rms of the predicted ligand bound conformation relative to the crystal structure for a representative sample of 6 low-energy ligand binding modes.

gen and oxygen positions for Asn side-chains are switched. This is the case of N83 in the V82A mutant complex with A77003, where the $\chi_2 = 159°$ is placed incorrectly and would not be predicted.

The observed similarity in the side-chain conformations of the active site residues for both wild-type and mutant HIV-1 protease complexes may be a result of the conservative nature of the mutations. However, with different ligands, there are side-chain conformational changes of the protein residues involved in binding of HIV-1 protease. Between the wild-type protease complexes with SB203386 and

TABLE IV. Summary of the Structure Prediction Results for the V82A Mutant HIV-1 Protease-A77003 Complex[†]

Residue	Volume overlap (%)						
R8/108	80.4/66.2	80.4/56.5	80.4/66.2	80.4/66.2	80.4/25.3	80.4/66.2	80.4/66.2
L23/123	84.9/79.3	84.9/79.3	69.7/62.5	64.9/79.3	64.9/59.5	55.3/59.5	55.3/79.3
V32/132	53.0/82.8	53.0/82.8	53.0/82.8	53.0/82.8	53.0/58.9	53.0/52.2	80.9/52.2
I47/147	50.6/79.1	50.6/79.1	76.0/79.1	76.0/79.1	76.0/79.1	76.0/79.1	76.0/79.1
I54/154	89.9/88.1	89.9/88.1	89.9/88.1	89.9/88.1	89.9/88.1	89.9/88.1	89.9/88.1
L76/176	75.2/81.0	75.2/81.0	75.2/81.0	75.2/81.0	75.2/81.0	75.2/81.0	75.2/81.0
T80/180	57.1/58.2	57.1/58.2	57.1/58.2	57.1/58.2	57.1/58.2	57.1/58.2	57.1/58.2
A82/182	—	—	—	—	—	—	—
N83/183	72.7/86.4	72.7/86.4	72.7/86.4	72.7/86.4	72.7/86.4	72.7/86.4	72.7/86.4
I84/184	73.6/70.2	83.3/91.2	83.3/91.2	83.3/70.2	73.6/91.2	83.3/91.2	83.3/91.2
Ligand rms	0.92	1.2	3.79	3.91	4.44	5.45	5.54
Energy	−8334.1	−8283.3	−8202.1	−8224.7	−8150.5	−7932.8	−7889.0

[†]Volume overlap of the predicted protein side-chains for the active site residues and rms of the predicted ligand bound conformation relative to the crystal structure for a representative sample of 7 low-energy ligand binding modes.

TABLE V. Side-Chain Conformations for the Active Site Protein Residues in the Wild-Type HIV-1 Protease-SB203386 Complex

Wild-type protein		DEE solution		Minimized DEE solution		Crystal structure	
Residue	Number	χ_1	χ_2	χ_1	χ_2	χ_1	χ_2
R	8	291.0	177.4	295.0	180.6	284.8	190.6
L	23	179.3	119.7	179.1	132.0	220.2	197.6
V	32	293.3	53.3	287.7	50.1	192.8	316.7
I	47	61.8	163.3	58.3	170.5	64.3	183.0
I	54	58.8	164.8	60.2	170.2	53.4	169.2
L	76	291.5	177.0	299.2	171.7	300.9	196.6
T	80	59.6	178.7	47.8	140.5	173.3	180.0
V	82	296.4	56.6	292.5	56.0	311.3	67.1
N	83	286.2	301.0	284.5	312.6	302.1	308.2
I	84	300.6	168.9	296.4	164.3	292.8	162.5
R	108	289.2	177.4	296.3	178.3	278.0	190.1
L	123	242.7	119.8	237.1	141.4	236.2	200.2
V	132	294.9	55.2	292.9	55.1	59.7	178.9
I	147	60.9	164.1	59.5	169.0	55.3	169.7
I	154	58.9	164.4	57.8	171.0	60.6	154.3
L	176	292.3	121.7	304.5	107.6	305.9	183.2
T	180	299.9	60.4	313.0	48.1	168.4	180.0
V	182	294.3	54.7	289.5	53.2	192.3	50.4
N	183	285.2	301.1	285.6	317.0	292.9	307.1
I	184	304.3	295.6	300.2	301.8	282.3	153.6

A77003, 55% of the residues within 5 Å radius from the ligand and 30% of the residues with 8 Å radius exhibit rotamer differences. Between the wild-type protease complexes with more similar ligands, SB203386 and SB203238,[78] 15% of the side-chains within 8 Å radius show rotamer differences. Combinations of subtle protein main-chain and side-chain movements serve to optimize the ligand-protein interactions and at the same time preserve the overall structural features of the system.

The predicted structure for the wild-type HIV-1 protease complex with SB203386 is in agreement with the crystal structure in so far as both $C_{\gamma 1}$ and $C_{\gamma 2}$ methyl groups of V182 make favorable van der Waals contacts with the *P*1′ benzyl group of the inhibitor (Fig. 6). Concurrent with the crystal structure of the triple mutant, these contacts in the mutant complex are reduced to only one favorable contact between the $C_{\gamma 2}$ methyl group of I182, while $C_{\gamma 1}$ and $C_{\sigma 1}$ methyl groups no longer maintain van der Waals contacts with the benzyl ring of SB203386. In the proximity of the mutated I82 residue, orientation of the C_{γ} methyl groups coupled with a movement of the P1 benzyl group of the ligand allows a single favorable contact with the larger I82 C_{γ} methyl group (Fig. 6) to be maintained.

In the crystal structure of the wild-type complex with A77003, the *P*1/*P*1′ benzyl groups of the inhibi-

TABLE VI. Side-Chain Conformations for the Active Site Protein Residues in the Wild-Type HIV-1 Protease-A77003 Complex

Wild-type protein		DEE solution		Minimized DEE solution		Crystal structure	
Residue	Number	χ_1	χ_2	χ_1	χ_2	χ_1	χ_2
R	8	293.9	177.0	294.7	181.8	293.0	184.7
L	23	184.4	63.6	182.3	70.8	183.5	74.8
V	32	298.5	58.3	288.3	50.8	199.5	319.3
I	47	186.2	72.6	188.6	62.5	34.2	98.4
I	54	63.3	163.7	59.5	169.8	59.9	162.5
L	76	295.2	176.7	299.1	170.1	312.9	170.7
T	80	64.8	179.2	53.4	138.2	179.4	211.8
V	82	297.3	57.7	282.4	46.5	58.4	181.4
N	83	292.9	298.9	285.9	314.6	300.1	314.4
I	84	298.7	168.9	293.6	164.0	299.7	289.8
R	108	292.3	177.2	300.6	175.9	299.3	183.9
L	123	183.3	62.8	183.6	73.5	186.0	75.7
V	132	175.5	296.0	189.8	310.0	189.4	310.2
I	147	65.0	164.0	54.2	173.6	41.4	96.7
I	154	60.3	164.0	59.6	170.6	58.5	172.3
L	176	295.3	176.2	298.5	167.0	309.6	179.3
T	180	66.5	179.2	44.2	143.9	183.6	214.1
V	182	72.2	191.2	63.4	186.6	59.5	186.6
N	183	292.7	299.2	286.4	308.2	294.7	305.4
I	184	299.5	168.9	302.8	164.4	293.3	158.7

TABLE VII. Side-Chain Conformations for the Active Site Protein Residues in the V32I/I47V/V82I Triple Mutant HIV-1 Protease-SB203386 Complex

Mutant protein		DEE solution		Minimized DEE solution		Crystal structure	
Residue	Number	χ_1	χ_2	χ_1	χ_2	χ_1	χ_2
R	8	291.0	177.4	295.0	180.9	274.2	188.9
L	23	179.3	119.7	179.1	131.6	188.1	70.2
I	32	295.7	168.8	301.0	172.3	317.4	187.8
V	47	174.2	294.3	181.2	301.9	111.3	235.8
I	54	58.8	164.8	60.4	169.7	53.4	172.3
L	76	291.5	177.0	299.3	172.1	309.2	183.0
T	80	187.1	179.4	183.3	215.1	177.3	180.0
I	82	193.5	166.3	187.0	169.6	194.8	175.0
N	83	286.2	301.0	185.5	312.8	289.3	321.8
I	84	300.6	168.9	299.3	167.1	288.1	165.7
R	108	289.2	177.4	296.3	178.3	271.9	183.7
L	123	242.7	119.8	223.2	145.9	183.6	60.3
I	132	297.9	168.9	291.2	171.3	337.2	262.2
V	147	296.7	56.5	296.2	58.5	183.9	292.1
I	154	58.9	164.4	58.0	170.8	48.8	151.7
L	176	293.3	121.7	304.6	108.6	305.7	172.7
T	180	190.1	180.6	186.1	212.3	175.5	190.0
I	182	298.2	296.3	198.7	199.3	177.8	176.0
N	183	285.2	301.1	290.0	314.2	290.3	314.0
I	184	303.0	168.4	296.6	163.6	283.6	159.7

tor make favorable interactions with the $C_{\gamma 1}$ methyl groups of the V82/182 residues of the protein. In contrast to the initial modeling studies of the mutant complex,[54] the structural repositioning of the main-chain near the mutated A82 and a movement of the P1 group of the ligand result in maintaining the favorable interactions. The predicted conformations of the ligand and side-chain show van der Waals

TABLE VIII. Side-Chain Conformations for the Active Site Protein Residues in the V82A Mutant HIV-1 Protease-A77003 Complex

Mutant protein		DEE solution		Minimized DEE solution		Crystal structure	
Residue	Number	χ_1	χ_2	χ_1	χ_2	χ_1	χ_2
R	8	293.9	177.0	294.8	181.0	295.2	192.2
L	23	184.4	63.6	183.8	73.1	187.4	69.5
V	32	298.5	58.3	289.9	52.4	192.5	314.4
I	47	194.5	166.3	193.4	167.9	41.1	108.8
I	54	63.3	163.7	59.8	170.1	71.3	174.1
L	76	295.2	176.7	298.9	170.4	307.8	177.1
T	80	64.8	179.2	51.8	145.2	178.3	180.0
A	82						
N	83	291.9	298.9	286.2	312.7	294.6	153.6
I	84	300.0	296.1	297.5	302.0	279.3	158.9
R	108	292.3	177.2	297.8	174.8	289.7	180.3
L	123	183.3	62.8	183.5	73.1	190.6	62.5
V	132	175.5	296.0	189.4	310.0	191.2	309.6
I	147	65.0	164.0	56.6	171.1	52.1	119.0
I	154	60.3	164.0	59.9	170.3	55.4	161.6
L	176	295.3	176.2	298.4	167.7	309.7	166.9
T	180	66.5	179.2	44.7	148.5	183.0	180.0
A	182						
N	183	292.7	299.2	287.5	307.7	291.7	318.1
I	184	300.8	296.1	301.3	300.9	293.2	176.2

contacts between A82 and a reorientation toward this residue by the P1 benzyl group of A77003 (Fig. 7). A retraction of the *P*1′ phenyl group of A77003 by approximately 0.4 Å from the *S*1′ subsite detected in the crystal structure of the mutant is correctly reproduced in the computational prediction (Fig. 7). The V82A mutation in HIV-1 protease was originally expected to result in a significant affinity change caused by the deletion of two methyl groups and a potential void on both sides of the active site suggested by the molecular modeling results. However, in the predicted structure of the V82A mutant complex with A77003, the interface of the P1 and S1 groups is well packed and indistinguishable from the wild-type complex, whereas a bigger void is formed in the *S*1′ subsite, which is consistent with the crystal structure.

DISCUSSION

In this work, we examine structural effects in the wild-type and mutant HIV-1 protease complexes where the ligand-protein interactions are primarily determined by favorable steric contacts between the ligand and hydrophobic side-chains in the active site of the protein. The results suggest that the binding energy landscapes for these ligand-protein complexes may have a correlated, funnel-like character. In this model, the low-energy binding modes of the ligand would infer structural rearrangements of the side-chains of the complexes that steadily propagate toward the bottom of the dominant funnel to the crystal structure as the energies of these complexes decrease. The nature of a dominant correlated funnel would result from a balance between interactions that stabilize the ligand conformations docked into a rigid protein and dominant hydrophobic interactions that stabilize both the ligand and protein side-chain conformations. This is reminiscent of the inherent structural harmony in minimally frustrated proteins in which local interactions that stabilize secondary structures must be consistent with tertiary interactions that stabilize complete protein folds. Consequently, once the vicinity of the dominant funnel is determined on a correlated energy landscape, the nature of the hypersurface would direct the search toward the native structure. This implies that the closer the ligand and the protein side-chains conform to the crystal structure, the lower the energy.

A more realistic representation of ligand-protein interactions may be necessary in ligand-protein systems where molecular recognition is primarily determined by specific hydrogen bonds and salt bridges which would require the selection of specific rotamer states. These energy landscapes are characterized by a multitude of ligand binding modes that would be accompanied by energetically compensating yet structurally different protein side-chain arrangements of the active site residues. The binding energy landscape of these ligand-protein systems may be uncorrelated with the structure of the ligand-protein complex, and energy evaluation would be subject to errors in this model because minor deviations from

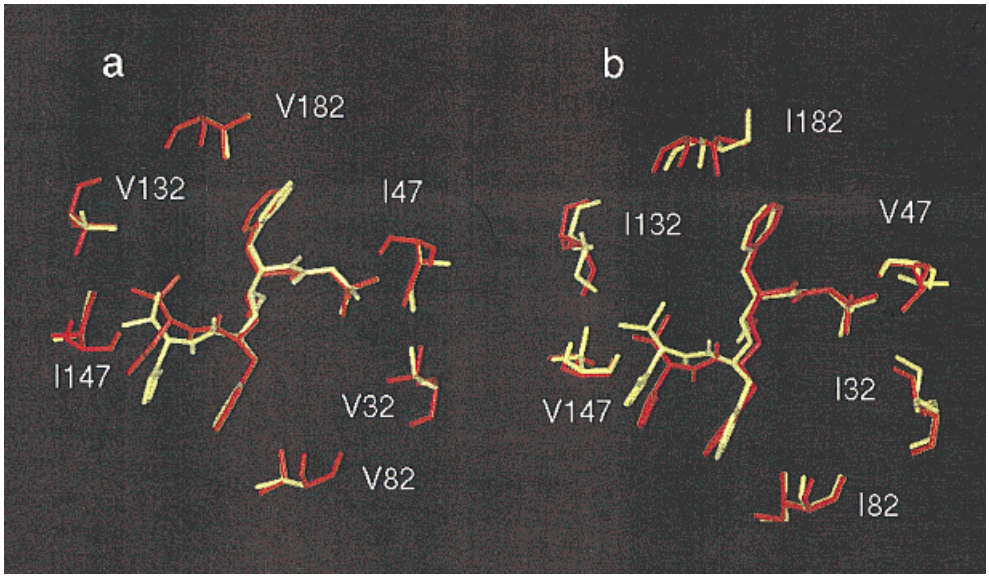

Fig. 4. Superposition of the crystal structure (yellow) and the predicted structure (red) of both the bound SB203386 ligand and the active site residues for the wild-type HIV-1 protease (**a**) and for the V32I/I47V/V82I triple HIV-1 protease mutant (**b**).

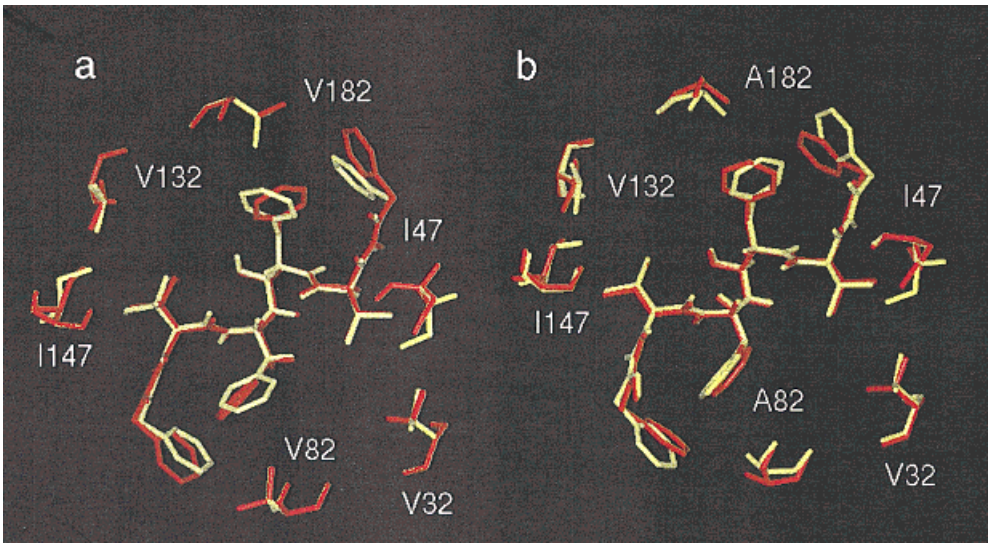

Fig. 5. Superposition of the crystal structure (yellow) and the predicted structure (red) of both the bound A77003 ligand and the active site residues for the wild-type HIV-1 protease (**a**) and for the V82/A HIV-1 protease mutant (b).

the global energy minimum would cause significant structural deviations from the native structure. A sensitive and thermodynamically accurate energetic model would then be necessary to distinguish between "correct" and "incorrect" solutions and locate the global energy minimum on the binding energy landscape where alternative structures may be ranked with energies similar to those of native complexes.

A complete description of molecular recognition includes a delicate balance of intermolecular ligand-protein interactions, solvation effects, and conformational entropy contributions of the ligand and protein side-chains, none of which can be computed accurately. The inherent complexity of the complete force field that describes the binding process and conflicts between intermolecular ligand-protein interactions and intramolecular ligand and protein constraints, which are simultaneously optimized when binding occurs, may cause a frustrated character of the underlying binding energy landscape. The multitude of low-energy metastable states of a thermody-

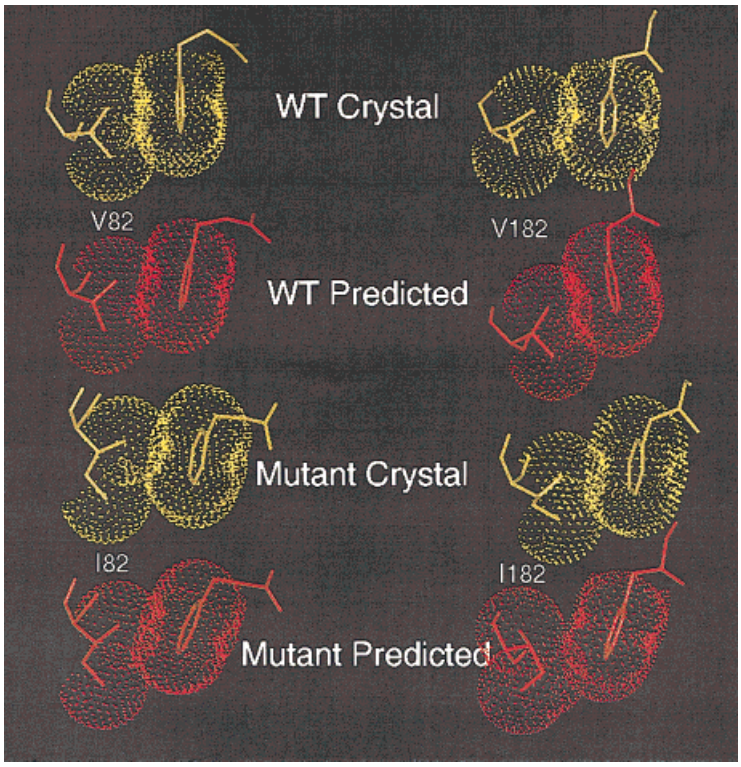

Fig. 6. Close-up of the van der Waals contacts between the *P*1/*P*1′ benzyl groups of SB203386 and the *S*1/*S*1′ subsite V82/182 residues in the wild-type crystal structure and wild-type predicted structure as well as the *S*/*S*1′ subsite I82/182 residues in the V32I/I47V/V82I triple mutant crystal structure and the V32I/I47V/V82I triple mutant predicted structure, respectively. Van der Waals surfaces are shown as dot surfaces.

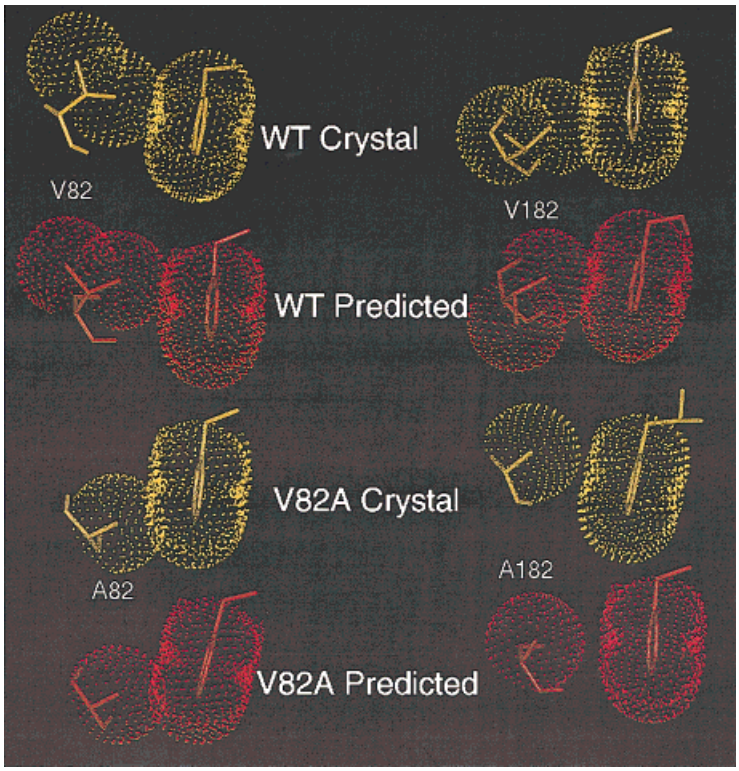

Fig. 7. Close-up of the van der Waals contacts between the *P*1/*P*1′ benzyl groups of A77003 and the *S*1/*S*1′ subsite V82/182 residues in the wild-type crystal structure and wild-type predicted structure as well as the *S*1/*S*1′ subsite A82/182 residues in the V82A mutant crystal structure and the V82A mutant predicted structure, respectively. Van der Waals surfaces are shown as dot surfaces.

namically complete and accurate energy function suitable for rigorous predictions of binding affinity may not automatically solve the kinetic aspect of the docking problem. A complicated and rough topology of the underlying binding energy landscape with low-energy valleys, high barriers, and kinetic traps

may lead to severe kinetic bottlenecks during docking simulations.

Ideally, the desired energy model used for predicting structural effects in docking simulations would also rank binding affinity of the ligand-protein complexes.[79–81] However, the goal in binding affinity analysis is to determine the binding free energy difference between bound and unbound states of the system, while in docking simulations attaining the free energy minimum only in the bound state is sufficient to predict the structure of the ligand-protein complex. Consequently, an effective computational protocol would use a hierarchical approach. A conventional scoring function, such as surface complementarity or a simplified atom-atom representation of the intermolecular interactions, is used to generate a number of possible solutions followed by a protein side-chain search and empirical free energy evaluations that estimate solvation and conformational entropy contributions.[10,13] The Poisson–Boltzmann method may be used to compute the electrostatic free energy of the complexed structures, whereas nonpolar solvation free energy contribution may be approximated by a term proportional to the solvent-accessible surface area.[10,13,21,26] The structures of the ligand-protein complexes can be refined by Monte Carlo minimization procedure that was recently expanded to include solvation contributions and flexible binding sites[21] as well as a more accurate evaluation of the electrostatic free energy and side-chain entropy.[26] A more elaborate and rigorous energy function would be required when the flexibility of the receptor side-chains is included to better discriminate between the native and nonnative solutions generated in docking simulations and to rank binding affinity of the ligand-protein complexes.

CONCLUSIONS

Structural effects of the active site mutations in HIV-1 protease complexes have been accurately assessed based on ligand docking to the wild-type protein backbone with the mutated residues followed by protein side-chain optimization with the rotamer library approximation. We demonstrate that an accurate prediction for the mutant complexes can be made given the protein backbone differences between the wild-type and the mutant protein within 0.3 Å rms and side-chains of nonactive residues within 1.2 Å rms. The ligand bound conformations are predicted within 1.0 Å rms from the crystal structure, and the optimized side-chains are predicted within 1.1 Å to 1.4 Å rms from the crystallographic conformations. The prediction accuracy of the side-chains measured by the volume overlap with the crystal structure is 60% to 75% for the wild-type complexes and 80% to 90% for the mutants. The energetic analysis of the ligand-protein complexes enables the crystal structure of the ligand to be distinguished from the "mirrored" alternative and the asymmetric nature of complexes with a symmetric protein active site becomes evident.

The rotamer library used in this study in conjunction with local refinement can adequately describe rotamers of the active site residues. Although the rotamer approximation significantly decreases the number of conformational possibilities for each amino acid, the number of combinations is still large enough to require the use of powerful searching techniques. The results of this study suggest that predicting side-chain conformations of the active site protein residues is determined by dual effect of both the protein backbone and the docked ligand conformations. Given the structure of the backbone, a number of low-energy side-chain conformations are allowed and the exact rotamer is chosen based on the most favorable interactions with the ligand. The side-chain conformations are tolerant within a window of at least $\pm 40°$ to energetically favorable ligand binding modes. The presence of the native backbone and the conservative nature of mutations makes the binding energy landscape for the side-chains less frustrated, resulting in consistent predictions of the native side-chain conformations.

The presented protocol can be extended to the study of protein-bound conformations in complexes with different ligands. There are more significant side-chain conformational changes in the protein residues involved in binding of HIV-1 protease with different ligands. The application of the described methodology to this more challenging problem may provide further insights into understanding the nature of ligand and protein side-chain interactions and the ability of computational methods to predict them.

ACKNOWLEDGMENTS

The authors are grateful to Dr. P. Rose, Dr. S. Freer, Dr. D. Bouzida, Dr. P. Rejto, D. Gehlhaar, A. Colson, V. Larson, and S. Arthurs, the members of the Agouron software development group, for developing the software used in this study. The authors also thank Dr. P. Rose for careful reading the manuscript and useful comments and Dr. Sherin S. Abdel-Meguid (Department of Macromolecular Sciences, SmithKline Beecham Pharmaceuticals) for providing the crystal structure coordinates of the V32I/I47V/V82I HIV-1 protease mutant complex with the SB203386 ligand.

REFERENCES

1. Kuntz, I.D. Structure-based strategies for drug design and discovery. Science 257:1078–1082, 1992.
2. Cherfils, J., Janin, J. Protein docking algorithms: Simulating molecular recognition. Curr. Opin. Struct. Biol. 3:265–269, 1993.

3. Rosenfeld, R., Vajda, S., DeLisi, C. Flexible docking and design. Annu. Rev. Biophys. Biomol. Struct. 24:677–700, 1995.
4. Jones, G., Willett, P. Docking small-molecule ligands into active sites. Curr. Opin. Biotechnol. 6:652–656, 1995.
5. Gschwend, D.A., Good, A.C., Kuntz, I.D. Molecular docking towards drug discovery. J. Mol. Recognit. 9:175–186, 1996.
6. Jiang, F., Kim, S.H. Soft docking: Matching of molecular surface cubes. J. Mol. Biol. 219:79–102, 1991.
7. Walls, P.H., Sternberg, M.J.E. New algorithm to model protein-protein recognition based on surface complementarity. Applications to antibody-antigen docking. J. Mol. Biol. 228:277–297, 1992.
8. Desjarlais, R.L., Dixon, J.S. A shape and chemistry-based docking method and its use in the design of HIV-1 protease inhibitors. J. Comput. Aided Mol. Design 8:231–242, 1994.
9. Vakser I.A., Aflalo, C. Hydrophobic docking: A proposed enhancement to molecular recognition techniques. Proteins 20:320–329, 1994.
10. Jackson, R.M., Sternberg, M.J.E. A continuum model for protein-protein interactions: Application to the docking problem. J. Mol. Biol. 250:258–275, 1995.
11. Fisher, D., Lin, S.L., Wolfson, H.J. Nussinov, R. A geometry-based suite of molecular docking processes. J. Mol. Biol. 248:459–477, 1995.
12. Wallqvist, A., Covell, D.G. Docking enzyme-inhibitor complexes using a preference-based free energy surface. Proteins 25:403–419, 1996.
13. Weng, Z., Vajda, S., DeLisi, C. Prediction of protein complexes using empirical free energy functions. Protein Sci. 5:614–626, 1996.
14. Shoichet, B.K., Kuntz, I.D. Protein docking and complementarity. J. Mol. Biol. 221:327–346, 1991.
15. Luty, B.A., Wasserman, Z.R., Stouten, P.F.W., C.N. Hodge, Zacharias, M., McCammon, J.A. A molecular mechanics-grid method for evaluation of ligand-receptor interactions. J. Comput. Chem. 16:454–464, 1995.
16. Rarey, M., Kramer, B., Lengauer, T., Klebe, G. A fast flexible docking method using an incremental construction algorithm. J. Mol. Biol. 261:470–489, 1996.
17. Welch, W., Ruppert, J., Jain, A.N. Hammerhead: Fast, fully automated docking of flexible ligands to protein binding sites. Chem. Biol. 3:449–462, 1996.
18. Goodsell, A.S., Olson, A.J. Automated docking of substrates to proteins by simulated annealing. Proteins 8:195–202, 1990.
19. Caflish, A., Niederer, P., Anliker, M. Monte Carlo docking of oligopeptides to proteins. Proteins 13:223–230, 1992.
20. Hart, T.N., Read, R.J. A multiple-start Monte Carlo docking method. Proteins 13:206–222, 1992.
21. Apostolakis, J., Pluckthun, A., Caflish, A. Docking small molecules in flexible binding sites. J. Comput. Chem. 19:21–37, 1998.
22. Oshiro, C.M., Kuntz, I.D., Dixon, J.S. Flexible ligand docking using a genetic algorithm. J. Comput. Aided Mol. Design 9:113–130, 1995.
23. Gehlhaar, D.K., Verkhivker, G.M., Rejto, P.A., et al. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: Conformationally flexible docking by evolutionary programming. Chem. Biol. 2:317–324, 1995.
24. Jones, G., Willett, P., Glen, R.C., Leach, A.R., Taylor, R. Development and validation of a genetic algorithm for flexible docking. J. Mol. Biol. 267:727–748, 1997.
25. Westhead, D.R., Clark, D.E., Murray, C.W. A comparison of heuristic search algorithms for molecular docking. J. Comput. Aided Mol. Design 11:209–228, 1997.
26. Totrov, M., Abagyan, R. Flexible protein-ligand docking by global energy optimization in internal coordinates. Proteins Suppl. 1:215–220, 1997.
27. Leach, A.R. Ligand docking to proteins with discrete side-chain flexibility. J. Mol. Biol. 235:345–356, 1994.
28. Desmet, J., Wilson, I.A., Joniau, M., De Mayer, M., Lasters, I. Computation of the binding of fully flexible peptides to proteins with flexible side-chains. FASEB J. 11:164–172, 1997.
29. Daniyat, B.I., Mayo, S.L. De novo protein design: Fully automated sequence selection. Science 278:82–87, 1997.
30. Lee, C., Subbiah, S. Prediction of protein side-chain conformation by packing optimization. J. Mol. Biol. 217:373–388, 1991.
31. Holm, L., Sander, C. Fast and simple Monte Carlo algorithm for side chain optimization in proteins: application to model building by homology. Proteins 14:213–223, 1992.
32. Tuffery, P., Etchebest, C., Hazout, S., Lavery, R. A critical comparison of search algorithms applied to the optimization of protein side chain conformations. J. Comput. Chem. 14:790–798, 1993.
33. Eisenmenger, F., Argos, P., Abagyan, R. A method to configure protein side chains from the main-chain trace in homology modeling. J. Mol. Biol. 231:849–860, 1993.
34. Laughton, C.A. Prediction of protein side-chain conformations from local three-dimensional homology relationships. J. Mol. Biol. 235:1088–1097, 1994.
35. Kono, H., Doi, J. Energy minimization method using automata network for sequence and side-chain conformation prediction from given backbone geometry. Proteins 19:244–255, 1994.
36. Koehl, P., Delarue, M. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. J. Mol. Biol. 239: 249–275, 1994.
37. Tanimura, R., Kidera, A., Nakamura, H. Determinants of protein side-chain packing. Protein Sci. 3:2358–2365, 1994.
38. Koehl, P., Delarue, M. A self-consistent mean field approach to simultaneous gap closure and side-chain positioning in homology modeling. Nat. Struct. Biol. 2:163–170, 1995.
39. Vasquez, M. An evaluation of discrete and continuum search techniques for conformational analysis of side chains in proteins. Biopolymers 36:53–70, 1995.
40. Shenkin, P.S., Farid, H., Fetrow, J.S. Prediction and evaluation of side-chain conformations for protein backbone structures. Proteins 26:323–352, 1996.
41. Sali, A. Modeling mutations and homologous proteins. Curr. Opin. Biotechnol. 6:437–451, 1995.
42. Vasquez, M. Modeling side-chain conformation. Curr. Opin. Struct. Biol. 6:217–221, 1996.
43. Lee, C., Levitt, M. Accurate prediction of the stability and activity effects of site-directed mutagenesis on a protein core. Nature 352:448–451, 1991.
44. Lee, C. Predicting protein mutant energetics by self-consistent ensemble optimization. J. Mol. Biol. 236:918–939, 1994.
45. Lee, C. Testing homology modeling on mutant proteins: Predicting structural and thermodynamic effects in the Ala98-Val mutants of T4 lysozyme. Fold. Des. 1:1–14, 1996.
46. Baldwin, E.P., Hajiseyedjavadi, O., Baase, W.A., Matthews, B.W. The role of backbone flexibility in the accommodation of variants that repack the core of T4 lysozyme. Science 262:1715–1718, 1993.
47. Lim, W.A., Hodel, A., Sauer, R., Richards, F.M. The crystal structure of a mutant protein with altered but improved hydrophobic core packing. Proc. Natl. Acad. Sci. USA 91:423–427, 1994.
48. Bryngelson, J.D., Onuchic, J.N., Socci, N.D., Wolynes, P.G. Funnels, pathways, and the energy landscape of protein folding: A synthesis. Proteins 21:167–195, 1995.
49. Dill, K.A., Chan, H.S. From Levinthal to pathways to funnels. Nat. Struct. Biol. 4:10–19, 1997.
50. Goldstein, R.F. Efficient rotamer elimination applied to protein side-chains and related spin glasses. Biophys. J. 66:1335–1340, 1994.
51. Wlodawer, A., Erickson, J.W. Structure-based inhibitors of HIV-1 protease. Annu. Rev. Biochem. 62:543–585, 1993.
52. Erikson, J.W. The not-so-great escape. Nat. Struct. Biol. 2:523–529, 1995.

53. Hoog, S.S., Towler, E.M., Zhao, B., Doyle, M.L., Debouck, C., Abdel-Meguid, S.S. Human immunodeficiency virus protease ligand specificity conferred by residues outside of the active site cavity. Biochemistry 35:10279–10286, 1996.
54. Baldwin, E.T., Bhat, T.N., Liu, B., Pattabiraman N., Erickson, J.W. Structural basis of drug resistance for the V82A mutant of HIV-1 proteinase. Nat. Struct. Biol. 2:244–249, 1995.
55. Verkhivker, G., Appelt, K., Freer, 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 HIV-1 protease binding affinity. Protein Eng. 8:677–691, 1995.
56. Verkhivker, G. Empirical free energy calculations of human immunodeficiency virus type 1 protease crystallographic complexes. II. Knowledge-based ligand-protein interaction potentials applied to thermodynamic analysis of hydrophobic mutations. In: "Pacific Symposium on Biocomputing-96." Hunter, L., Klein, T.E. (eds.). Singapore: World Scientific, 1996:638–652.
57. Verkhivker, G.M., Rejto, P.A. A mean field model of ligand-protein interactions: Implications for the structural assessment of human immunodeficiency virus type 1 protease complexes and receptor-specific binding. Proc. Natl. Acad. Sci. USA 93:60–64, 1996.
58. Rejto, P.A., Verkhivker, G.M. Unraveling principles of lead discovery: From unfrustrated energy landscapes to novel molecular anchors Proc. Natl. Acad. Sci. USA 93:8945–8950, 1996.
59. Verkhivker, G.M., Rejto, P.A., Gehlhaar, D.K., Freer, S.T. Exploring energy landscapes of molecular recognition by a genetic algorithm: Analysis of the requirements for robust docking of HIV-1 protease and FKBP-12 complexes. Proteins 25:342–353, 1996.
60. Rejto, P.A., Verkhivker, G.M., Gehlhaar, D.K., Freer, S.T. New trends in computational structure prediction of ligand-protein complexes for receptor-based drug design. In: "Computational Simulation of Biomolecular Systems." van Gunsteren, W., Weiner, P., Wilkinson A.J. (eds.). Leiden: ESCOM, 1997:451–465.
61. Verkhivker, G.M., Rejto, P.A. Mean field analysis of FKBP-12 complexes with FK-506 and rapamycin: Implications for a role of crystallographic water molecules in molecular recognition and specificity. Proteins 28:313–324, 1997.
62. Shah, N., Rejto, P.A., Verkhivker, G.M. Structural consensus in ligand-protein docking identifies recognition peptide motifs that bind streptavidin. Proteins 28:421–433, 1997.
63. Mayo, S.L., Olafson, B.D., Goddard, W.A. III. DREIDING: A generic force field for molecular simulation. J. Phys. Chem. 94:8897–8909, 1990.
64. Bouzida, D., Kumar, S., Swendsen, R.H. Efficient Monte Carlo methods for the computer simulation of biological molecules. Phys. Rev. A 45:8894–8901, 1992.
65. Ponder, J.W., Richards, F.M. Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J. Mol. Biol. 193:775–791, 1987.
66. Dunbrack, R.L. Jr., Karplus, M. Backbone-dependent rotamer library for proteins. Application to side-chain prediction. J. Mol. Biol. 230:543–574, 1993.
67. Desmet, J., De Maeyer, M., Hazes, B., Lasters, I. The dead-end elimination theorem and its use in protein side-chain positioning. Nature 356:539–542, 1992.
68. Lasters, I., Desmet, J. The fuzzy-end elimination theorem: Correctly implementing the side-chain placement algorithm based on the dead-end elimination theorem. Protein Eng. 6:717–722, 1993.
69. Lasters, I., De Maeyer, M., Desmet, J. Enhanced dead-elimination in the search for the global minimum energy conformation of a collection of protein side chains. Protein Eng. 8:815–822, 1995.
70. Keller, D.A., Shibata, M., Marcus, E., Ornstein, R.L., Rein, R. Finding the global minimum: A fuzzy and elimination implementation. Protein Eng. 8:893–904, 1995.
71. Desmet, J., De Maeyer, M., Lasters, I. The 'dead-end elimination' theorem: A new approach to the side-chain packing problem. In: "The Protein Folding Problem and Tertiary Structure Prediction." Merz, K.M., Jr., Le Grands, S.M. (eds.). Boston: Birkhauser, 1994:307–337.
72. De Maeyer, M., Desmet, J., Lasters, I. All in one: A highly detailed rotamer library improves both accuracy and speed in the modeling of side chains by dead-end elimination. Fold. Des. 2:53–66, 1997.
73. Jorgensen, W.L., Tirado-Rives, J. The OPLS potential functions for proteins. Energy minimizations for crystals of cyclic peptides and crambin. J. Am. Chem. Soc. 110:1657–1666, 1988.
74. Mohamadi, F., Richards, N.G.J., Guida, W.C., et al. MacroModel—an integrated software system for modeling organic and bioorganic molecules using molecular mechanics. J. Comput. Chem. 11:440–467, 1990.
75. Weiner, S.J., Kollman, P.A., Case, D.A., et al. A new force field for molecular mechanical simulation of nucleic acids and proteins. J. Am. Chem. Soc. 106:765–784, 1984.
76. Still, W.C., Tempczyk, A., Hawley, R.C., Hendrickson, T., Semianalytical treatment of solvation for molecular mechanics and dynamics. J. Am. Chem. Soc. 112:6127–6129, 1990.
77. Hong, L., Treharne, A., Hartsuck, J.A., Foundling, S., Tang, J. Crystal structures of complexes of a peptidic inhibitor with wild-type and two mutant HIV-1 proteases. Biochemistry 35:10627–10633, 1996.
78. Hoog, S.S., Zhao, B., Winborne, E., et al. A check on rotational drug design: Crystal structure of a complex of human immunodeficiency virus type 1 protease with a novel gamma-turn mimetic inhibitor. J. Med. Chem. 38: 3246–3252, 1995.
79. Cummings, M.D., Hart, T.N., Read, R.J. Atomic solvation parameters in the analysis of protein-protein docking results. Protein Sci. 4:2087–2099, 1995.
80. Janin, A.N. Scoring noncovalent protein-ligand interactions: A continuous differentiable function tuned to compute binding affinities. J. Comput. Aided Mol. Design 10:427–440, 1996.
81. Eldridge, M.D., Murray, C.W., Auton, T.R., Paolini, G.V., Mee, R.P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate binding affinity of ligands in receptor complexes. J. Comput. Aided Mol. Design 11:425–445, 1997.