

Research article

Open Access

## Improved prediction of HIV-1 protease-inhibitor binding energies by molecular dynamics simulations

Ekachai Jenwitheesuk<sup>1,2</sup> and Ram Samudrala\*<sup>1</sup>

Address: <sup>1</sup>Department of Microbiology, University of Washington School of Medicine, Seattle, WA 98195, USA and <sup>2</sup>Department of Clinical Microbiology, Faculty of Medical Technology, Mahidol University, Bangkok, Thailand

Email: Ekachai Jenwitheesuk - ekachai@compbio.washington.edu; Ram Samudrala\* - ram@compbio.washington.edu

\* Corresponding author

Published: 1 April 2003

Received: 25 December 2002

*BMC Structural Biology* 2003, **3**:2

Accepted: 1 April 2003

This article is available from: <http://www.biomedcentral.com/1472-6807/3/2>

© 2003 Jenwitheesuk and Samudrala; licensee BioMed Central Ltd. This is an Open Access article: verbatim copying and redistribution of this article are permitted in all media for any purpose, provided this notice is preserved along with the article's original URL.

### Abstract

**Background:** The accurate prediction of enzyme-substrate interaction energies is one of the major challenges in computational biology. This study describes the improvement of protein-ligand binding energy prediction by incorporating protein flexibility through the use of molecular dynamics (MD) simulations.

**Results:** Docking experiments were undertaken using the program AutoDock for twenty-five HIV-1 protease-inhibitor complexes determined by x-ray crystallography. Protein-rigid docking without any dynamics produced a low correlation of 0.38 between the experimental and calculated binding energies. Correlations improved significantly for all time scales of MD simulations of the receptor-ligand complex. The highest correlation coefficient of 0.87 between the experimental and calculated energies was obtained after 0.1 picoseconds of dynamics simulation.

**Conclusion:** Our results indicate that relaxation of protein complexes by MD simulation is useful and necessary to obtain binding energies that are representative of the experimentally determined values.

### Background

The human immunodeficiency virus type 1 aspartic protease (HIV-1 PR) is an important enzyme due to its key role in viral maturation. Inactivation of the enzyme causes the production of immature, noninfectious viral particles. The enzyme therefore is an attractive target in anti-AIDS drug design, and the effect of binding various inhibitors on the protease structure is currently the focus of intensive research [1–3].

To obtain information about the position and energy of binding between an inhibitor and the corresponding protein, several automated docking programs have been developed [4–6]. Given recent improvements in search algorithms and energy functions, computational docking

methods have become a valuable tool to probe the interaction between an enzyme and its inhibitors. These methods can contribute significantly to the understanding of structural and energetic basis of enzyme-substrate interactions [7–9].

Protein-ligand docking methods aim to predict the binding energy of the protein-ligand complex given the atomic coordinates. In such calculations, both the protein and ligand can be treated as rigid bodies [10,11]; alternately, the ligand, the protein, or both molecules, can be completely or partially flexible [12,13].

One advantage of incorporating flexibility is that it enables a search without bias introduced by the initial model.

This bias normally influences both the orientation and conformation of the ligand in the active site, which usually represents a local minimum conformation [14]. More importantly, the lock and key concepts used to evaluate enzyme-substrate binding, in reality, refer to flexible locks and keys that are both in constant dynamic (thermal) motion [15]. For example, an analysis of conformational changes upon complex formation for a representative set of 39 pairs of ligand-free and ligand-bound structures showed that 50% of these proteins undergo substantial main-chain and side chain conformational changes when the ligand is bound [15].

Several techniques have been developed to predict the binding energies of HIV-1 protease-inhibitor complexes [16–21]. Head RD *et al.*, 1996 [16] used an approach that calculates physicochemical properties of the ligands and the receptor-ligand complexes to estimate the free energy of binding. The enthalpy of binding is calculated by molecular mechanics, while properties such as complementary hydrophobic surface area are used to estimate the entropy of binding through heuristics. Gohlke H *et al.*, 2000 [17] developed DrugScore, a knowledge-based scoring function, to discriminate between well-docked ligand binding modes and those largely deviating from the native structure.

Schapiro M *et al.*, 1999 [20] used the finite difference Poisson-Boltzmann implementation of the electrostatic term, in conjunction with appropriate surface and entropy terms to predict the binding energy of 13 HIV-1 protease-inhibitor complexes. The predicted binding energies had a correlation coefficient of 0.66 with the experimental data. Recently, Österberg F *et al.*, 2002 used AutoDock 3.0, a ligand flexible docking program, together with combining 21 protease structures into a single representative grid of interaction energies for incorporating the side chain motion into a docking simulation protocol [21]. The correlation coefficient between experimental and calculated binding energies produced by this technique was 0.79.

Since most current docking programs treat the receptor and/or ligand as rigid objects, this represents an important cause of failure to predict the correct binding enzyme-substrate energies [18]. We address this problem by performing Molecular Dynamic (MD) simulations on HIV-1 protease-inhibitor complexed and using the resulting structures to calculate the binding energies by AutoDock, a ligand flexible docking program.

## Results and discussion

### **Correlation between experimentally determined and calculated binding energy**

The primary objective of this study was to determine whether protein relaxation would improve prediction of

protein-ligand binding energies. Table 1 shows that the correlation coefficient of the experimentally determined and calculated binding energies from AutoDock for the twenty-five protease-inhibitor complexes after protein-rigid docking was 0.38. The correlations after 0.01, 0.1, 1, and 10 picoseconds (ps) of MD simulations were 0.53, 0.87, 0.79, and 0.74 respectively. These correlations are plotted in Figure 1. One outlier (from 1hvi) was noticed in Figure 1B; after it is eliminated, the correlation coefficient changes from 0.53 to 0.72 after 0.01 ps of simulation.

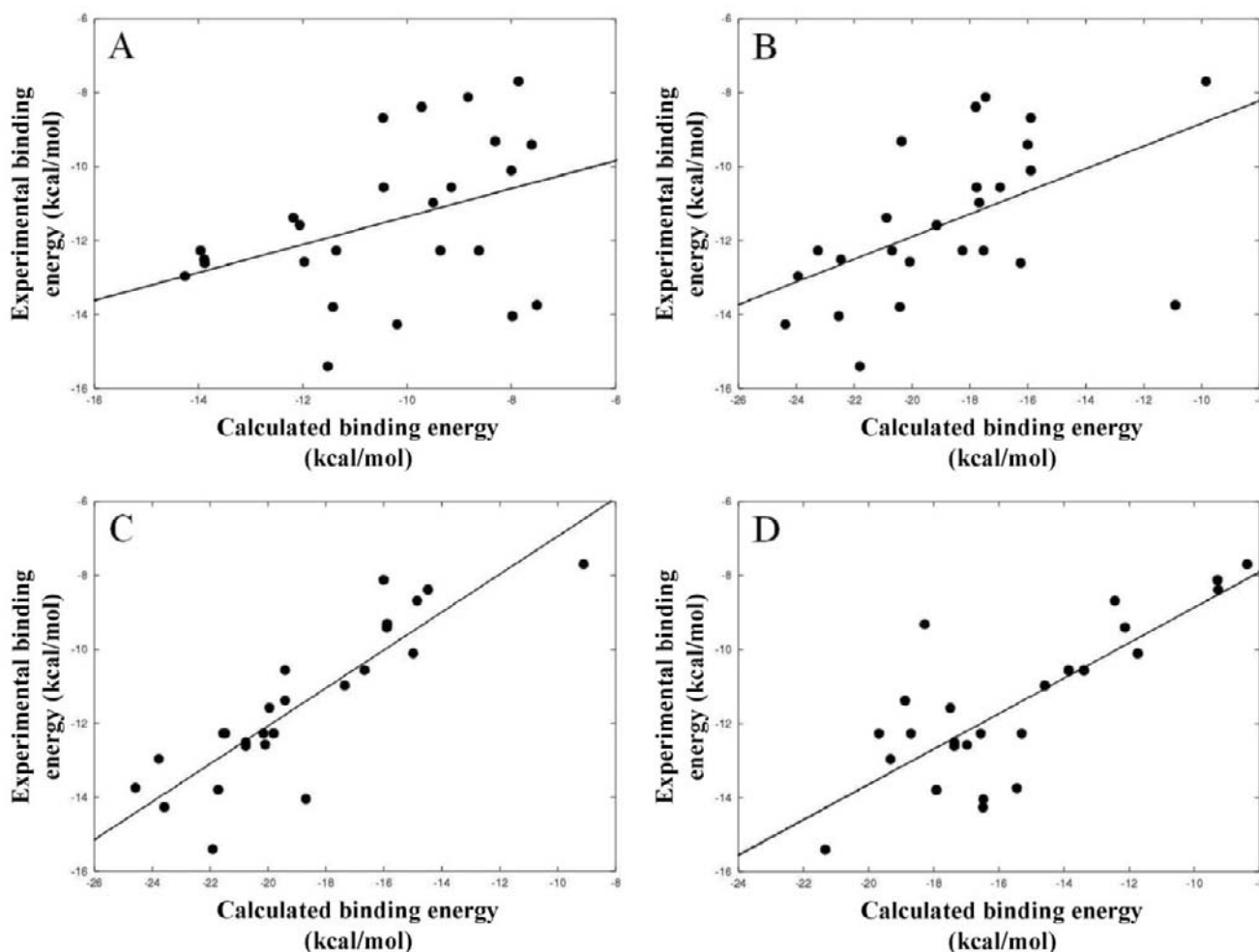
All these correlations represent a significant improvement over the protein-rigid docking results. The best results were obtained from the structures at the 0.1 ps MD simulation time point. These structures also have low ( $\cong 0.3$  Å) average all-atom root mean square deviations (RMSD) relative to the experimental results. Table 2 shows the all-atom RMSD between each simulated complex and the corresponding protease-inhibitor x-ray structure. The average all-atom RMSD for the complexes increases from 0.18 Å at 0.01 ps to 2.92 Å at the end of 10 ps of simulation time. Similar results are consistently observed regardless of the five starting seeds used in the MD simulations.

When a constant value of 6.5 is subtracted from the predicted energies after 0.1 ps MD, the binding energies of almost all of the predictions were within 2.0 kcal/mol of the experimental values. Three complexes, 1hvi, 1hvr, and 1hte had poor predicted energies, with an average error of 4.34, 4.31, and 5.09 kcal/mol, respectively. Among these, one (1hvi) had better predicted energies when compared to the results of protein-rigid docking.

### **Influence of the protease flap movement on calculated binding energy**

The beta-strand flap is the most flexible region in the HIV-1 protease. It is normally 7 Å RMSD from the active site and is in an open conformation in the native state [22,23]. The protease undergoes significant structural changes on binding to an inhibitor. The two flaps fold over the inhibitor to form a tunnel-shaped active site and are held in this close position by hydrogen bonding from Ile50 and Ile50' NH groups of the enzyme to a water molecule, which in turn is hydrogen bonded to the P2 and P1' CO groups of the inhibitor [24]. The bonding stabilizes the flaps in a closed position and inhibits the activities of the enzyme.

MD simulation has been used to study the movement of the flap region of HIV-1 protease with a ligand [25–30]. The flaps initially opened to an all-atom RMSD of 25 Å within 200 ps and became completely open at the end of a 10 ns simulation. In this study (Figure 2), the flaps opened up and moved away from the x-ray structure from 0.54 Å at 0.1 ps to 3.30 Å RMSD at 10 ps (the flap RMSD



**Figure 1**

Plots of experimentally-determined and calculated binding energies for the twenty-five HIV-1 protease-inhibitor complexes. The correlation coefficient improves from 0.38 without any MD simulation (A), to 0.53 at 0.01 ps of MD simulation (B), peaking at 0.87 after 0.1 ps (C), and dropping off to 0.79 at longer time scales of to 1 ps or more (D).

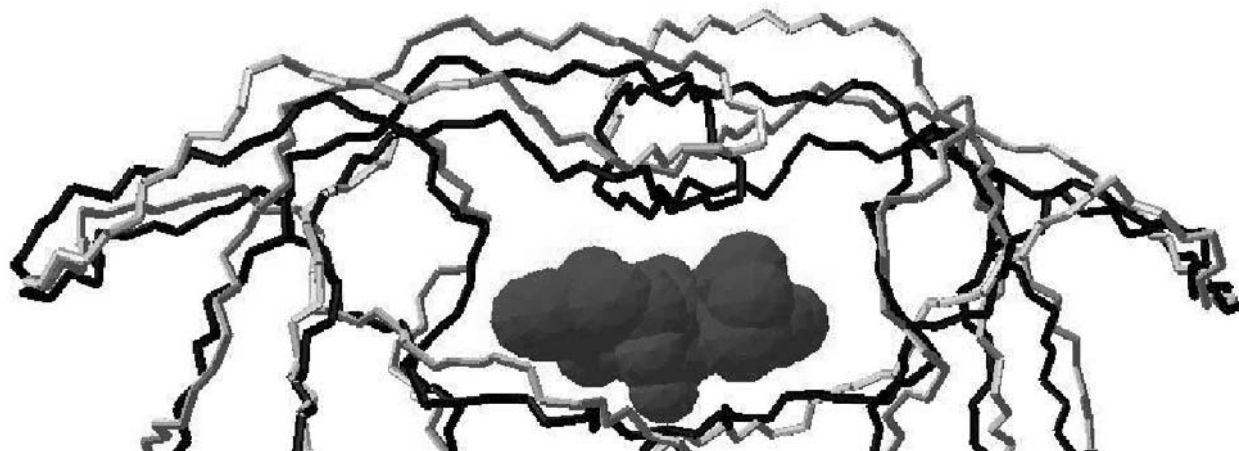
was calculated from residue 40 to 60 of each protein chain). These movements, after 0.1 ps of simulation, are inversely correlated with the quality of the binding energy prediction. As shown in Table 1, the correlation coefficient significantly decreased from 0.87 at 0.1 ps to 0.74 at 10 ps as the all-atom flap RMSD increased from 0.54 to 3.30 at 0.1 and 10 ps, respectively.

Complementarity between the ligand and the binding site is the basic concept behind ligand binding. This is manifest as steric complementarity, i.e. the shape of the ligand is mirrored in the shape of the binding site, allowing molecular interactions between two molecules [31]. MD simulations allow rearrangement of the protease side chain,

especially on the active site surface, which improves the interacting surface complementarities of the complex. As shown in Table 2, after 0.1 ps, the time scale that produced the best correlation coefficient, the average all-atom RMSD of the complex was only 0.35 Å, while the average all-atom RMSD of the flap region was 0.54 Å.

#### **Influence of MD simulation duration on ligand binding**

Ligand docking revealed a consistent set of recurring binding modes. For all MD time scales, well-clustered docking results could be obtained. Generally, the lowest binding energy clusters are associated with the lowest all-atom RMSDs of the ligands. The best results in terms of lowest binding energy are located in a similar position of the x-



**Figure 2**

Superposition of the  $C_{\alpha}$  traces of part of the HIV-1 protease x-ray structure 4phv before (dark line) and after (light line) 10 ps of MD simulation bound to the inhibitor (space-fill). The flap region (above the inhibitor) moved away from the x-ray structure during the simulation, with all-atom RMSDs of 0.54 Å at 0.1 ps, 0.95 Å at 1 ps and 3.30 Å at 10 ps, respectively. Generally, after 0.1 ps, as the all-atom RMSD increased, the correlation coefficient of the experimentally-determined and calculated binding energies decreased.

ray structure at the active site. Table 3 shows the number of docking solutions in a cluster ( $N$ ) along with the all-atom ligand RMSD for each MD time scale.

A small  $N$  value indicates strong specificity of binding, with all of the solutions resembling one of only a small number of binding conformations and orientations. On the other hand, if  $N$  is large, the experiment indicates a low specificity of binding, since the solutions are composed of many different binding conformations or orientations.

In this study, the clustering result of ligands had an average  $N$  of 10.72 for docking without any MD, 8.36, 4.12, 5.88, and 7.16 for docking with 0.01, 0.1, 1, and 10 ps MD simulation, respectively. On average, the lower number of docking solutions in the cluster for all MD simulation time scales indicates that the ligands bind to their binding pocket with high specificity. The docked ligands after 10 ps MD simulation exhibited a wide range of RMSDs (0.87 Å to 3.86 Å) with an average RMSD of 2.52 Å, which indicates the failure of the ligands to recognize and specifically bind to the binding site with the protease flaps open.

### Conclusions

In this study, we illustrate the importance of taking dynamics into account to predict the structure and energetics of protein-ligand interactions. It is clear that relaxation of

HIV-1 protease for 0.1 ps MD simulations is enough for rearrangement of the surface-binding pocket to produce good correlations between calculated and experimental binding energies. The binding energies of all protease molecules bound to different inhibitors were influenced by the movement of the flap regions with the correlation coefficient decreasing as the flaps moved away from the experimental structure. The differences in these correlations may reflect biological features of the dynamics of HIV-1 protease-inhibitor complexes.

Future work with larger data sets, different energy functions, different docking and binding energy evaluation methods, and more starting seeds, is necessary to determine the optimal parameters to robustly predict protein-substrate binding energies *in silico*.

### Methods

A set of twenty-five HIV-1 protease-inhibitor complexes with experimentally determined structures and binding energies chosen from the Protein Data Bank (PDB), were used in this study [32–50]. All complexes were determined by x-ray crystallography with resolution and R-factor less than 3.0 and 0.19, respectively. The PDB codes of the complexes along with the experimental binding energies are listed in Table 1.

**Table 1: Comparison of experimentally-determined and calculated binding energies for twenty-five HIV-1 protease-inhibitor complexes.**

PDB code	Resolution	R-factor	Experimental binding energy (kcal/mol)	Calculated binding energy (kcal/mol) (picosecond time scales)				
				no MD	0.01	0.10	1.00	10.00
ligno	2.30	0.17	-9.40	-7.61	-16.01	-15.90	-12.13	-10.30
lhbv	2.30	0.18	-8.68	-10.46	-15.90	-14.85	-12.44	-10.64
lhef	2.20	0.16	-12.27	-13.96	-18.26	-19.81	-15.30	-12.32
lheg	2.20	0.19	-10.56	-9.15	-17.77	-19.41	-13.39	-11.16
lhih	2.30	0.19	-10.97	-9.50	-17.68	-17.35	-14.59	-11.95
lhiv	2.00	0.17	-12.27	-11.36	-23.25	-21.54	-19.68	-17.31
lhps	2.20	0.14	-12.57	-11.97	-20.08	-20.10	-16.98	-16.60
lhpv	1.90	0.19	-12.60	-13.88	-16.25	-20.77	-17.37	-16.48
lhvi	1.80	0.18	-13.74	-7.51	-10.91	-24.58	-15.45	-14.72
lhvj	2.00	0.16	-14.26	-10.19	-24.37	-23.58	-16.49	-14.81
lhvk	1.80	0.18	-13.79	-11.42	-20.42	-21.72	-17.92	-14.98
lhvl	1.80	0.19	-12.27	-9.36	-17.53	-21.48	-16.56	-13.87
lhvr	1.80	0.19	-12.96	-14.26	-23.94	-23.77	-19.33	-16.22
lhvs	2.25	0.15	-14.04	-7.98	-22.53	-18.69	-16.48	-14.55
lhte	2.80	0.16	-7.69	-7.86	-9.85	-9.10	-8.38	-7.74
lhtf	2.20	0.19	-9.31	-8.31	-20.36	-15.89	-18.28	-18.87
lhtg	2.00	0.19	-11.58	-12.06	-19.15	-19.95	-17.49	-15.64
lpro	1.80	0.19	-15.40	-11.52	-21.80	-21.91	-21.33	-21.32
lsbg	2.30	0.19	-10.56	-10.45	-16.96	-16.67	-13.86	-11.91
2upj	3.00	0.14	-10.10	-8.00	-15.90	-14.99	-11.74	-11.18
4phv	2.10	0.18	-12.51	-13.89	-22.45	-20.77	-17.37	-16.48
4hvp	2.30	0.18	-8.38	-9.72	-17.80	-14.48	-9.27	-8.80
5hvp	2.00	0.18	-8.12	-8.83	-17.46	-16.01	-9.29	-8.09
8hvp	2.50	0.14	-12.27	-8.62	-20.69	-20.16	-18.70	-16.15
9hvp	2.80	0.18	-11.38	-12.18	-20.88	-19.41	-18.88	-15.55
Correlation coefficient				0.38	0.53	0.87	0.79	0.74

For each complex, the PDB code, resolution, R-factor, the experimental energy, the lowest binding energy as evaluated by AutoDock before any MD simulation, and the lowest binding energy after MD simulation for four time points (0.01, 0.1, 1, and 10 ps) are shown. The bottom row shows the correlation coefficient between the experimental and calculated energies. Protein-rigid docking produced poor correlation (of 0.38) between the experimental and calculated energies, while the best correlations (of 0.87) were obtained after 0.1 ps MD simulations.

The experimental binding energies used in this study were converted from the experimental inhibition constants ( $K_i$ ), using the equation:

$$\Delta G^0 = -RT \ln K_i$$

where  $\Delta G^0$  is the Gibbs free energy of binding (kcal/mol),  $R$  is the gas constant ( $1.987 \text{ cal K}^{-1} \text{ mol}^{-1}$ ) and  $T$  is the absolute temperature (room temperature, 300 K). Assuming that the  $K_i$  values of each complex were obtained under the quasi-equilibrium conditions of Michaelis-Menton kinetics [51,52], we expect that a plot of  $\Delta G^0$  against  $\ln K_i$  is a straight line, with a slope given by the thermal energy,  $RT = 0.6 \text{ kcal/mol}$  (at 300 K). Since the  $K_i$  values were determined by different experimental techniques, some lev-

el of noise might exist in the underlying data. We expect our method to behave robustly in this regard.

#### Molecular Dynamics simulations

MD simulations of all complexes were carried out with the NAMD software version 2.5b1 [53] using the X-PLOR force field [54]. Missing atoms in 1heg, 1hte and 5hvp were added by using the guesscoord command in NAMD. The van der Waals, bond, angle, dihedral, and improper dihedral parameters for all the ligands were adopted from the Hetero-compound Information Centre-Uppsala HIC-Up <http://xray.bmc.uu.se/hicup> [55]. The water molecule under the flaps present in all complexes was included in all steps throughout this study except for 1hvr, where the structural water was removed in the preparation of the protease to be docked with the ligand (the ligand of 1hvr,

**Table 2: All-atom root mean square deviation (RMSD) of the protein-ligand complexes relative to their corresponding x-ray structures.**

PDB code	Complex all-atom RMSD (Å) (picosecond time scales)			
	0.01	0.10	1.00	10.00
Igno	0.22	0.32	1.11	3.20
Ihbv	0.16	0.35	0.95	2.84
Ihef	0.16	0.34	0.90	2.71
Iheg	0.16	0.32	0.91	2.80
Ihih	0.18	0.35	0.90	2.93
Ihiv	0.20	0.36	0.85	2.98
Ihps	0.18	0.38	0.95	2.95
Ihvp	0.17	0.33	0.92	2.96
Ihvi	0.14	0.35	0.92	2.99
Ihvj	0.17	0.33	0.91	3.06
Ihvk	0.12	0.33	0.94	2.97
Ihvl	0.21	0.33	0.91	2.93
Ihvr	0.18	0.36	0.92	2.92
Ihvs	0.24	0.35	0.92	2.88
Ihte	0.20	0.37	0.90	2.89
Ihtf	0.10	0.32	0.93	3.06
Ihtg	0.14	0.31	0.76	2.86
Ipro	0.26	0.38	0.94	2.94
Isvg	0.20	0.36	0.93	2.86
2upj	0.23	0.41	0.90	2.88
4phv	0.14	0.35	0.90	2.93
4hvp	0.20	0.35	0.93	2.95
5hvp	0.18	0.31	0.94	2.93
8hvp	0.18	0.35	0.92	2.90
9hvp	0.19	0.35	0.92	2.90
Complex (average)	0.18	0.35	0.92	2.92
Flap (average)	0.14	0.54	0.95	3.30

The correlations between experimentally-determined and calculated binding energy significantly improved after MD simulation, and were inversely influenced by the all-atom RMSD of the complex as well as the protease flap region (i.e., as the all-atom RMSD increased with longer MD simulations, the correlations were reduced from their peak). The best correlation of 0.87 was observed at 0.1 ps. Similar results were observed regardless of the starting seed used.

which is a urea-based inhibitor, does not bind a water molecule in this position). The water molecules were added to the binding site of 1hvs because all the water molecules were absent in the experimental structure. Protein protonation states were modeled as in the related HIV-1 protease MD simulation study [56]. All protein residues were modeled in their charged state except for one of the two aspartic acid groups (Asp 25 and Asp 25') in the active site since previous studies [57–59] have shown that at least one of these two aspartic acids is protonated. We used a protonated Asp 25' and deprotonated Asp 25 for all HIV-1 protease-ligand complexes. The terminal residues of both monomers were also protonated (Pro 1, Pro 1', Phe 99 and Phe 99').

The complexes were immersed in a 20 Å radius sphere of TIP3-water using the program SOLVATE [60] to allow the protein-ligand complexes to relax in an aqueous environ-

ment. One hundred steps of energy minimization of the protein-ligand-water complex were initially performed, followed by 10 ps MD simulation at 300 K, with an atom-based shifted distance-dependent dielectric constant,  $\epsilon = 4r$ ; a switch function on van der Waals interaction, and a time step of 1 femtosecond (fs). The nonbonded interaction list was updated every 20 time steps. The van der Waals interactions were truncated at a distance of 12 Å. The simulations were repeated with five different starting seeds. The structures at 0.01 ps, 0.1 ps, 1 ps and 10 ps were recorded and processed in the docking step.

#### **Preparation of protease-inhibitor complexes**

To calculate the binding energy with AutoDock, we first prepared the protein complexes by separating each snapshot recorded from MD simulations into one file containing the protease and the water molecules, and one file containing only the ligand. Polar hydrogens were then

**Table 3: Ligand all-atom RMSD (Å) and the number of docking solutions (N) in the cluster from 100 Lamarckian genetic algorithm (LGA) docking runs of twenty-five protease-inhibitor complexes.**

PDB code	No MD		0.01 ps		0.1 ps		1 ps		10 ps	
	N	RMSD (Å)	N	RMSD (Å)	N	RMSD (Å)	N	RMSD (Å)	N	RMSD (Å)
lgn0	32	0.62	18	0.88	4	1.54	13	1.57	16	1.54
lhbv	4	2.15	8	1.18	6	1.30	6	2.05	9	3.18
lhef	8	1.46	6	1.24	4	1.13	12	2.14	9	2.23
lgeg	15	2.16	7	0.88	4	0.94	6	1.31	5*	1.20
lhih	14	1.21	6	0.58	4	0.75	7	1.81	4	3.86
lhiv	12	1.27	8	1.04	4	1.06	5	1.29	7	2.75
lhps	10	1.03	5	1.35	1	0.95	5	1.53	2	2.84
lhpy	3	1.33	9	1.32	5	0.73	5	2.92	7	2.59
lhvi	13	1.53	6	0.93	4	0.95	2	2.94	5	3.54
lhvj	3	1.33	3	0.75	2	0.95	2	2.35	4	2.86
lhvk	8	1.44	4	1.34	7	1.12	9	0.97	2	3.26
lhvl	12	1.72	2	1.61	2	1.54	5	2.49	3	2.92
lhvr	18	0.98	9	0.70	6	0.68	8	1.01	8	2.37
lhvs	13	1.73	7	0.87	1	1.03	3	0.99	3	1.24
lhve	3	2.26	9	1.21	3	1.88	7	2.72	8*	3.23
lhvf	13	1.11	8	1.03	6	0.93	13	2.11	10	2.9
lhvg	8	1.96	9	0.97	8	1.03	6	0.96	12	2.86
lhvh	18	0.76	15	0.78	7	0.77	9	0.63	9	0.87
lhvi	11	1.71	17	0.61	8	0.73	2	1.01	9	2.63
lhvj	15	1.75	17	1.74	4	1.79	2	2.07	9	3.54
lhvk	1	2.93	9	1.57	3	1.23	3	2.51	7	3.08
lhvl	3	1.97	11	0.98	2	1.37	5	1.91	7	2.16
lhvm	12	1.89	5	0.86	2	0.88	8	0.97	8*	1.01
lhvn	13	1.75	5	1.53	2	1.41	2	2.25	8	2.16
lhvo	6	2.66	6	1.48	2	1.54	2	2.51	8	2.12
Average	10.72	1.63	8.36	1.10	4.12	1.13	5.88	1.80	7.16	2.51

Docking solutions with ligand all-atom RMSDs within 1.0 Å of each other were clustered together and ranked by the corresponding lowest energy representative. The lowest energy solution of the lowest ligand RMSD cluster was accepted as the calculated binding energy. The lowest ligand RMSD cluster was usually ranked as the first cluster; the clusters that were ranked as a second cluster are marked with (\*). Average *N* values from docking results of all MD simulation time scales are lower than an average *N* value from protein-rigid docking, indicating that the ligands bind to the binding pocket of the MD simulated structures with higher specificity.

added to the protein coordinates with the PROTONATE utility from AMBER [61]. AMBER united atom force field charges were assigned, and solvation parameters were added using the ADDSOL utility. The 3D affinity grid fields were created using the auxiliary program AutoGrid. The center of protein mass was chosen as the grid center. In this stage, the protein was embedded in the 3D grid and a probe atom was placed at each grid point. The affinity and electrostatic potential grid was calculated for each type of atom in the protease molecule. The number of grid points in x, y, z-axis was 60 × 110 × 60 with grid points separated by 0.375 Å.

Ligands that had a peptide-like N- or C-terminal end were assigned a charge. Hydrogen atoms were added to fill all empty valences, and Kollman united-atom charges [62] were also assigned to the ligands. Rotatable dihedrals in

the ligands were assigned using the program AutoTors and were allowed to rotate freely. The nonpolar hydrogens were removed and the partial charges from these were added to the carbon that held the hydrogen. The atom type for the aromatic carbons was reassigned to be handled by the aromatic carbon grid map. These preparations were done for each ligand using the AutoTors module.

#### Automated docking

Docking calculations were carried out using AutoDock, version 3.0.5 [4]. Three binding energy terms were taken into account in the docking step: the van der Waals interaction represented as a Lennard-Jones 12–6 dispersion/repulsion term, the hydrogen bonding represented as a directional 12–10 term, and the Coulombic electrostatic potential.

Docking runs were performed using the Lamarckian genetic algorithm (LGA) [4] as previously described [63,64] with some modifications of the docking parameters. The LGA describes the relationship between the protein and the ligand by the translation, orientation, and conformation of the ligand. These "state variables" are the ligand's genotype, and the resulting atomic coordinates together with the interaction and intermolecular energies are the ligand's phenotype. The environmental adaptation of the ligand's phenotype was reverse transcribed into its genotype and became heritable traits.

Docking began with a population of random ligand conformations in random orientations and at random translations. Each docking experiment was derived from 100 different runs that was set to terminate after a maximum of 1,500,000 energy evaluations or 27,000 generations, yielding 100 docked conformations. The population size was set to 50. The elitism number, the rate of gene mutation and the rate of gene crossover were 1, 0.02 and 0.8 respectively. A pseudo-Solis and Wets local search was then used to minimize energy of the population. The probability that docking solution in the population would undergo a local search was set to 0.06 and the constraint was set to a maximum of 300 iterations per search. The maximum number of successes or failures before changing the size of local search space ( $\rho$ ) were both set to 4. The starting conformations of the ligand were set to random positions. Translations were set to have a maximum limit of 2 Å/step and the orientation step size for the angular component and the torsions had a maximum limit at 50 degrees/step.

At the end of a docking job with multiple runs, AutoDock performed cluster analysis. Docking solutions with ligand all-atom RMSDs within 1.0 Å of each other were clustered together and ranked by the lowest energy representative. The lowest-energy solution of the lowest ligand all-atom RMSD cluster was accepted as the calculated binding energy.

### Author's contributions

EJ performed Molecular Dynamics simulations, docking, evaluated the results, and drafted the manuscript. RS helped with evaluation of the results produced, refining the manuscript, and provided intellectual guidance and mentorship. All authors read and approved the final manuscript.

### Acknowledgments

This work was supported in part by a Searle Scholar Award to Ram Samudrala. We thank members of the Samudrala group for helpful comments.

### References

- Zhu J, Fan H, Liu H and Shi Y **Structure-based ligand design for flexible proteins: application of new F-DycoBlock** *J Comput Aided Mol Des* 2001, **15**:979-996
- Jayatilake PR, Nair AC, Zauhar R and Welsh WJ **Computational studies on HIV-1 protease inhibitors: influence of calculated inhibitor-enzyme binding affinities on the statistical quality of 3D-QSAR CoMFA models** *J Med Chem* 2000, **43**:4446-4451
- Sham YY, Chu ZT, Tao H and Warshel A **Examining methods for calculations of binding free energies: LRA, LIE, PDL-D-LRA, and PDL-D/S-LRA calculations of ligands binding to an HIV protease** *Proteins* 2000, **39**:393-407
- Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK and Olson AJ **Automated Docking Using a Lamarckian Genetic Algorithm and Empirical Binding Free Energy Function** *J Comput Chem* 1998, **19**:1639-1662
- Henry AG, Richard MJ and Michael JES **Modelling Protein Docking using Shape Complementarity, Electrostatics and Biochemical Information** *J Mol Biol* 1997, **272**:106-120
- Kramer B, Rarey M and Lengauer T **Evaluation of the FlexX incremental construction algorithm for protein-ligand docking proteins** *Structure Functions and Genetics* 1999, **37**:228-241
- Verkhivker GM, Rejto PA, Gehlhaar DK and Freer ST **Exploring the energy landscapes of molecular recognition by a genetic algorithm: 10 analysis of the requirements for robust docking of HIV-1 protease and FKBP-12 complexes** *Proteins* 1996, **25**:342-353
- Sham YY, Chu ZT, Tao H and Warshel A **Examining methods for calculations of binding free energies: LRA, LIE, PDL-D-LRA, and PDL-D/S-LRA calculations of ligands binding to an HIV protease** *Proteins* 2000, **39**:393-407
- Pegg SC, Haresco JJ and Kuntz ID **A genetic algorithm for structure-based de novo design** *J Comput Aided Mol Des* 2001, **15**:911-933
- Vakser IA **Evaluation of GRAMM low-resolution docking methodology on the hemagglutinin-antibody complex** *Proteins* 1997, **Suppl 1**:226-230
- Vakser IA, Matar OG and Lam CF **A systematic study of low-resolution recognition in protein - protein complexes** *Proc Natl Acad Sci USA* 1999, **96**:8477-8482
- Schaffer L and Verkhivker GM **Predicting structural effects in HIV-1 protease mutant complexes with flexible ligand docking and protein side chain optimization** *Proteins* 1998, **33**:295-310
- Verkhivker GM and Rejto PA **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* 1996, **93**:60-64
- Ota N and Agard DA **Binding mode prediction for a flexible ligand in a flexible pocket using multi-conformation simulated annealing pseudo crystallographic refinement** *J Mol Biol* 2001, **30**:607-617
- Betts MJ and Sternberg MJ **An analysis of conformational changes on protein-protein association: implications for predictive docking** *Protein Eng* 1999, **12**:271-283
- Head RD, Smythe ML, Oprea TI, Waller CL, Green SM and Marshall GR **VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands** *J Am Chem Soc* 1996, **118**:3959-3969
- Gohlke H, Hendlich M and Klebe G **Knowledge-based Scoring Function to Predict protein-ligand interactions** *J Mol Biol* 2000, **295**:337-356
- Akaho E, Morris GM, Goodsell DS, Wong D and Olson AJ **A study on docking mode of HIV protease and their inhibitors** *J Chem Software* 2001, **7**:103-114
- Klebe G **Recent developments in structure-based drug design** *J Mol Med* 2000, **78**:269-281
- Schapira M, Totrov M and Abagyan R **Prediction of the binding energy for small molecules, peptides and proteins** *J Mol Recognit* 1999, **12**:177-190
- Österberg F, Morris GM, Sanner MF, Olson AJ and Goodsell DS **Automated docking to multiple targets structures: Incorporation of protein mobility and structure water heterogeneity in AutoDock** *Proteins* 2002, **46**:34-40



22. Harte WE Jr, Swaminathan S, Mansuri MM, Martin JC, Rosenberg IE and Beveridge DL **Domain communication in the dynamical structure of human immunodeficiency virus I protease** *Proc Natl Acad Sci USA* 1990, **87**:8864-8868
23. Harte WE Jr, Swaminathan S and Beveridge DL **Molecular Dynamics of HIV-1 protease** *Protiens* 1992, **13**:175-194
24. York DM, Darden TA, Pedersen LG and Anderson MW **Molecular dynamics simulation of HIV-1 protease in a crystalline environment and in solution** *Biochemistry* 1993, **32**:1443-1453
25. Navia MA, Fitzgerald PM, McKeever BM, Leu CT, Heimbach JC, Herber WK, Sigal IS, Darke PL and Springer JP **Three dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1** *Nature* 1989, **337**:615-620
26. Wlodawer A, Miller M, Jaskolski M, Sathyanarayana BK, Baldwin E, Weber IT, Selk LM, Clawson L, Schneider J and Kent SB **Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease** *Science* 1989, **245**:616-621
27. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent SB and Wlodawer A **Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution** *Science* 1989, **246**:1149-1152
28. Scott WR and Schiffer CA **Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance** *Structure Fold Des* 2000, **8**:1259-1265
29. Collins JR, Burt SK and Erickson JW **Flap opening in HIV-1 protease simulated by 'activated' molecular dynamics** *Nat Struct Biol* 1995, **2**:334-338
30. Collins JR, Burt SK and Erickson JW **Activated dynamics of flap opening in HIV-1 protease** *Adv Exp Med Biol* 1995, **362**:455-460
31. Rastall RA **Protein, structure and function: Ligand binding** [<http://www.fst.rdg.ac.uk/courses/fs916/lect5/lect5.htm>]
32. PDB code **Crystal structures of complexes of a peptidic inhibitor with wild-type and two mutant HIV-1 proteases** *Biochemistry* 1996, **35**:10627-10633
33. PDB code **A check on rational drug design: crystal structure of a complex of human immunodeficiency virus type I protease with a novel gamma-turn mimetic inhibitor** *J Med Chem* 1995, **38**:3246-3252
34. PDB code **Comparative analysis of the X-ray structures of HIV-1 and HIV-2 proteases in complex with CGP 5 a novel pseudosymmetric inhibitor** *Structure* 3820, **3**:381-389
35. PDB code **Crystal structure of a complex of HIV-1 protease with a dihydroxyethylene-containing inhibitor: comparisons with molecular modeling** *Protein Sci* 1992, **1**:1061-1072
36. PDB code **Crystal-structure of HIV-1 protease in complex with VX-478, A potent and orally bioavailable inhibitor of the enzyme** *J Am Chem Soc* 1995, **117**:1181-1189
37. PDB code **X-ray crystallographic studies of a series of penicillin-derived asymmetric inhibitors of HIV-1 protease** *Biochemistry* 1994, **33**:8417-8427
38. PDB code: Ihvi, Ihvj, Ihvk, Ihvl, Hosur MV, Bhat TN, Kempf DJ, Baldwin ET, Liu BS, Gulnik S, Wideburg NE, Norbeck DW, Appelt K and Erickson JW **Influence of stereochemistry on activity and binding modes for C2 symmetry-based diol inhibitors of HIV-1 protease** *J Am Chem Soc* 1994, **116**:847
39. PDB code **Structural basis of drug resistance for the V82A mutant of HIV-1 proteinase** *Nat Struct Biol* 1995, **2**:244-249
40. PDB code **An orally bioavailable HIV-1 protease inhibitor containing an imidazole-derived peptide bond replacement: crystallographic and pharmacokinetic analysis** *Biochemistry* 1994, **33**:11671-11677
41. PDB code **A novel, picomolar inhibitor of human immunodeficiency virus type I protease** *J Med Chem* 1996, **39**:392-397
42. PDB code **Structure-based design of novel HIV protease inhibitors: carboxamide-containing 4-hydroxycoumarins and 4-hydroxy-2-pyrones as potent nonpeptidic inhibitors** *J Med Chem* 1995, **38**:3624-3637
43. PDB code **Rational design of potent, bioavailable, nonpeptide cyclic ureas as HIV protease inhibitors** *Science* 1994, **263**:380-384
44. PDB code: 4phv, Bone R, Vacca JP, Anderson PS and Holloway MK **X-Ray Crystal Structure of the HIV Protease Complex with L-700,417, an Inhibitor with Pseudo C2 Symmetry** *J Am Chem Soc* 1991, **113**:9382
45. PDB code **The crystal structures at 2.2-Å resolution of hydroxyethylene-based inhibitors bound to human immunodeficiency virus type I protease show that the inhibitors are present in two distinct orientations** *J Biol Chem* 1992, **267**:22770-22778
46. PDB code **Rational design, synthesis, and crystallographic analysis of a hydroxyethylene-based HIV-1 protease inhibitor containing a heterocyclic P1'-P2' amide bond isostere** *J Med Chem* 1994, **37**:3100-3107
47. PDB code **Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution** *Science* 1989, **246**:1149-1152
48. PDB code **Crystallographic Analysis of a Complex between Human Immunodeficiency Virus Type I Protease and Acetyl-Pepstatin at 2.0 Angstroms Resolution** *J Biol Chem* 1990, **265**:14209-14219
49. PDB code **Structure at 2.5-Å resolution of chemically synthesized human immunodeficiency virus type I protease complexed with a hydroxyethylene-based inhibitor** *Biochemistry* 1991, **30**:1600-1609
50. PDB code **Design, Activity and 2.8 Angstroms Crystal Structure of a C2 Symmetric Inhibitor Complexed to HIV-1 Protease** *Science* 1990, **249**:527-533
51. Stryer L *Biochemistry* WH Freeman and Co., New York 1999, 187-195
52. Creighton TE *Proteins: Structures and Molecular Properties* WH Freeman and Co., New York 1993, 386-389
53. Kale L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K and Schulten K **NAMD2: Greater scalability for parallel molecular dynamics** *J Comput Phys* 1999, **151**:283-312
54. Brunger AT **X-PLOR version 3.1, A system for X-ray crystallography and NMR** Yale University Press, New Haven, CT 1992,
55. Kleywegt GJ and Jones TA **Databases in protein crystallography** *Acta Cryst* 1998, **D54**:1119-1131
56. Geller M, Miller M, Swanson SM and Maizel J **Analysis of the structure of HIV-1 protease complexed with a hexapeptide inhibitor. Part II: Molecular dynamic studies of the active site region** *Proteins* 1997, **27**:195-203
57. Piana S and Carloni P **Conformational flexibility of the catalytic Asp dyad in HIV-1 protease: an ab initio study on the free enzyme** *Proteins: Struct Funct Genet* 2000, **39**:26-36
58. Hyland LJ, Tomaszek TA Jr and Meek TD **Human immunodeficiency virus-1 protease 2. Use of pH rate studies and solvent kinetic isotope effects to elucidate details of chemical mechanism** *Biochemistry* 1991, **30**:8454-8463
59. Yamazaki T, Nicholson LK, Torchia DA, Wingfield P, Stahl SJ, Kaufman JD, Eyer mann CJ, Hodge CN, Lam PYS, Ru Y, Jadhav PK, Chang CH and Weber PC **NMR and X-rays evidence that the HIV protease catalytic aspartyl groups are protonated in the complex formed by the protease and a non-peptide cyclic urea-based inhibitor** *J Am Chem Soc* 1994, **116**:10791-10792
60. Grubmuller H, Heymann B and Tavan P **Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force** *Science* 1996, **271**:997-999
61. Pearlman DA, Case DA, Caldwell JW, Ross WS, Cheatham TE, DeBolt S III, Ferguson D, Seibel G and Kollman P **AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules** *Comp Phys Commun* 1995, **91**:1-41
62. Weiner SJ, Kollman PA, Case DA, Singh UC, Ghio C, Alagona G, Profeta S and Weiner P **A new force field for molecular mechanical simulation of nucleic acids and proteins** *J Am Chem Soc* 1984, **106**:765-784
63. Huang X, Xu L, Luo X, Fan K., Ji R, Pei G, Chen K and Jiang H **Elucidating the inhibiting mode of AHPBA derivatives against HIV-1 protease and building predictive 3D-QSAR models** *J Med Chem* 2002, **45**:333-343
64. Rockey WM, Laederach A and Reilly PJ **Automated docking of  $\alpha$ -(1 $\rightarrow$ 4)- and  $\alpha$ -(1 $\rightarrow$ 6)-linked glucosyl trisaccharides and maltopentose 14 into the soybean  $\beta$ -amylase active site** *Protein* 2000, **40**:299-309