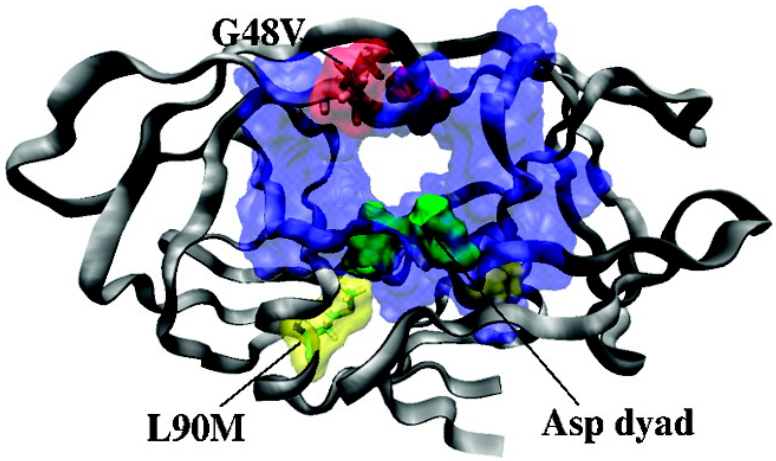
Article

Rapid and Accurate Prediction of Binding Free Energies for Saquinavir-Bound HIV-1 Proteases

Ileana Stoica, S. Kashif Sadiq, and Peter V. Coveney

J. Am. Chem. Soc., **2008**, 130 (8), 2639-2648 • DOI: 10.1021/ja0779250



More About This Article

Additional resources and features associated with this article are available within the HTML version:

- Supporting Information
- Links to the 4 articles that cite this article, as of the time of this article download
- Access to high resolution figures
- Links to articles and content related to this article
- Copyright permission to reproduce figures and/or text from this article

View the Full Text HTML

J|A|C|S
ARTICLES
Published on Web 01/29/2008

Rapid and Accurate Prediction of Binding Free Energies for Saquinavir-Bound HIV-1 Proteases

Ileana Stoica, S. Kashif Sadiq, and Peter V. Coveney*

Centre for Computational Science, Department of Chemistry, University College London, London WC1H 0AH, U.K.

Received October 16, 2007; E-mail: p.v.coveney@ucl.ac.uk

Abstract: To explain drug resistance by computer simulations at the molecular level, we first have to assess the accuracy of theoretical predictions. Herein we report an application of the molecular mechanics Poisson−Boltzmann surface area (MM/PBSA) technique to the ranking of binding affinities of the inhibitor saquinavir with the wild type (WT) and three resistant mutants of HIV-1 protease: L90M, G48V, and G48V/L90M. For each ligand−protein complex we report 10 ns of fully unrestrained molecular dynamics (MD) simulations with explicit solvent. We investigate convergence, internal consistency, and model dependency of MM/PBSA ligand binding energies. Converged enthalpy and entropy estimates produce ligand binding affinities within 1.5 kcal/mol of experimental values, with a remarkable level of correlation to the experimentally observed ranking of resistance levels. A detailed analysis of the enthalpic/entropic balance of drug−protease interactions explains resistance in L90M in terms of a higher vibrational entropy than in the WT complex, while G48V disrupts critical hydrogen bonds at the inhibitor's binding site and produces an altered, more unfavorable balance of Coulomb and polar desolvation energies.

1. Introduction

The development of drug resistance by the human immunodeficiency virus (HIV) continues to be a major problem in the treatment of AIDS. While several effective treatment regimens have been devised, involving inhibitors that target multiple viral proteins,[1] emergence of mutations in these proteins is a contributing factor to the eventual failure of treatment.

Given its crucial role in the proliferation of the virus, HIV-1 protease is an attractive target for HIV/AIDS therapy. The enzyme is responsible for proteolytic cleavage of the Gag and Gag-Pol polyprotein chains necessary for subsequent maturation of infectious virions. The protease is a C_2-symmetric dimer containing 99 amino acids in each chain; the dimerization interface forms the active site, with the catalytic aspartic acid dyad at the base and with a pair of flexible hairpin β-sheets known as "flaps" modulating access to the site.[2]

The potential of molecular simulations to enhance our understanding of drug resistance in HIV/AIDS relies ultimately on their capability to achieve an accurate ranking of drug binding affinities on clinically relevant time scales. Several computational approaches exist to estimate ligand binding affinities and selectivities, with various levels of accuracy and computational expense:[3,4] free energy perturbation (FEP), thermodynamic integration (TI), linear response (LR), and molecular mechanics Poisson−Boltzmann surface area (MM/PBSA).

MM/PBSA combines molecular mechanics energies with continuum solvent models to post-process a series of representative snapshots from molecular dynamics (MD) trajectories.[5] In contrast to more rigorous methods such as TI, the MM/PBSA approach is faster by several orders of magnitude while, unlike more approximate methods like LR, it does not require any experimental data or fitting to parameters.[6] On the other hand, the approximations inherent to MM/PBSA result in larger errors than those associated with some other methods (e.g., TI) and in occasional discrepancies with experimental results.[7] Several recent MM/PBSA studies have focused on ligand binding interactions and multi-drug resistance in HIV-1 protease.[8−13]

Here we investigate the potential of the MM/PBSA methodology to accurately rank the binding affinities of the inhibitor saquinavir to the wild type (WT) and resistant variants of HIV-1 protease: L90M, G48V, and G48V/L90M. We also use MM/PBSA to identify the thermodynamic determinants of binding, as a means to understand the mechanisms of emergence of the L90M and G48V mutations under chemotherapeutic pressure.

Saquinavir is a first-generation transition state analogue inhibitor of HIV-1 protease which blocks the maturation step of the HIV life cycle. L90M is a particularly important mutation,

(1) Alfano, M.; Poli, G. *Drug. Des. Rev.−Online* **2004**, *1*, 83−92.
(2) Wlodawer, A.; Miller, M.; Jaskolski, M.; Sathyanarayana, B. K.; Baldwin, E.; Weber, I. T.; Selk, L. M.; Clawson, L.; Schneider, J.; Kent, S. B. *Science* **1989**, *245*, 616−621.
(3) Wang, W.; Donini, O.; Reyes, C. M.; Kollman, P. A. *Annu. Rev. Biophys Biomol. Struct.* **2001**, *30*, 211−243.
(4) Wan, S.; Coveney, P. V.; Flower, D. R. *Philos. Trans. R. Soc. A* **2005**, *363*, 2037−2053.
(5) Kollman, P. A.; et al. *Acc. Chem. Res.* **2000**, *33*, 889−897.
(6) Rizzo, R. C.; Toba, S.; Kuntz. I. D. *J. Med. Chem.* **2004**, *47*, 3065−3074.
(7) Page, C. S.; Bates, P. A. *J. Comput. Chem.* **2006**, *27*, 1990−2007.
(8) Ode, H.; Neya, S.; Hata, M.; Sugiura, W.; Hoshino, T. *J. Am. Chem. Soc.* **2006**, *24*, 7887−7895.
(9) Ode, H.; Matsuyama, S.; Hata, M.; Hoshino, T.; Kakizawa, J.; Sugiura, W. *J. Med. Chem.* **2007**, *50*, 1768−1777.
(10) Hou, T.; Yu, R. *J. Med. Chem.* **2007**, *50*, 1177−1188.
(11) Wang, W.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14937−14942.
(12) Lepsik, M.; Kriz, Z.; Havlas, Z. *Proteins: Struct., Funct. Bioinf.* **2004**, *57*, 279−293.
(13) Kalra, P.; Reddy, T.; Jayaram, B. *J. Med. Chem.* **2001**, *44*, 4325−4338.

as it is clinically associated with resistance to nearly all approved inhibitors, yet it does not lie in the active site.[15] G48V resides in the flaps and is found predominantly in association with L90M in response to saquinavir therapy.[15] The two sets of experimentally determined inhibition constants for L90M, G48V, and G48V/L90M[16,17] indicate a modest increase in resistance in L90M, 20- and 3-fold respectively, an increase of 160- and 13.5-fold respectively for G48V, and an even more significant increase of 1000- and 400-fold respectively for G48V/L90M.

In this study, we attempt to quantify resistance in terms of a decrease in binding affinity of the inhibitor and, further, to compute changes in binding affinity upon mutation as changes in calculated ligand binding energies of protein−inhibitor complexes. In the following, the terms "binding affinity" and "free energy of binding" will be used interchangeably. "Relative affinity" is used to designate the change in binding energy in a mutant relative to the WT complex.

The paper is organized as follows: In the Methods section we describe the modeling of the protease−saquinavir complexes, the MD simulation protocol, and details of the MM/PBSA calculations. The Results and Discussion section starts with an assessment of the convergence and internal consistency of the binding enthalpy and entropy estimates, using the fully unrestrained, 10 ns long MD trajectories from these protease−inhibitor complexes. We compute demonstrably converged absolute and relative ligand binding energies and obtain remarkably high correlations to experimental data. The analysis is complemented by a free energy decomposition of the contributions to binding, which enables us to identify the thermodynamic determinants of drug resistance associated with the primary L90M and compensatory G48V mutations. We then provide an analysis of the structural constraints incurred by the L90M and G48V mutations and terminate the article with a Conclusions section.

2. Methods

2.1. Initial Preparation of Models. There are currently only two resolved crystal structures available for HIV-1 protease complexed with saquinavir: 1HXB[14] at 2.3 Å and 1FB7[15] at 2.6 Å resolution. The 1HXB structure is of the WT protease, while 1FB7 is of the G48V/L90M mutant. The 1HXB crystal structure contains two resolved rotationally symmetric sets of coordinates for saquinavir. 1HXB was used in previous studies of protease−saquinavir complexes.[8,18,20] To model the WT system, we also used 1HXB as the natural choice. For the double mutant G48V/L90M, we used the published crystal structure 1FB7 which, compared to 1HXB, presents several changes in the conformations adopted by the drug and the residues in the binding pocket, incurred by the L90M and G48V mutations.[15] For consistency, we also used the 1FB7 crystal structure to model the single mutants L90M and G48V.

The protonation state of the aspartic acid dyad was also considered. There are many uncertainties with regard to the appropriate protonation states of the protease under different conditions (see ref 19 and references therein). Exhaustive analysis of all possible combinations of protonation states for the Asp dyad is reported elsewhere[8,20,21] and is beyond the scope of the present study. Previous molecular simulations of the protease complexed with saquinavir have suggested a monoprotonated dyad, with D25 being thermodynamically favored in the WT system[8,20,21] and the G48V mutant[20] as well as in the L90M mutant.[8] On the basis of these results, we chose to model D25 as monoprotonated in all four protease−inhibitor complexes.

Gaussian 98[23] was used to perform geometric optimization of the inhibitor, with 6-31G** basis functions. The Restrained Electrostatic Potential (RESP) procedure, which is also part of the AMBER package, was used to calculate the partial atomic charges. The force field parameters for the inhibitor were described by the General Amber Force Field (GAFF).[24] The standard AMBER force field for bio-organic systems (ff03)[25] was used to describe the protein parameters.

Five Cl^- counter ions were added to electrically neutralize each inhibitor-bound system. Each system was then solvated using atomistic TIP3P water[26] in a cubic box with at least 14 Å distance around the complex. The size of each inhibitor-bound system was 45 314, 42 670, 42 680, and 42 676 atoms for the WT, L90M, G48V, and G48V/L90M systems, respectively.

2.2. MD Simulations. The molecular dynamics package NAMD2[27] was used throughout the production simulations as well as for the minimization and equilibration protocols. Minimization was conducted using the conjugate gradient and line search algorithms available in NAMD2 for 2000 iterations of each system with a force constant of 4 kcal mol^{-1} $Å^{-2}$ applied to all restrained atoms. Restrained atoms included all heavy atoms of HIV-1 protease and the ligand.

Long-range Coulombic interactions were handled using the particle mesh Ewald (PME) summation.[28] For the equilibration and subsequent production run, the SHAKE algorithm[29] was employed on all atoms covalently bonded to a hydrogen atom, allowing for an integration time step of 2 fs.

Each system was gently annealed from 50 to 300 K over a period of 50 ps. The systems were then maintained at a temperature of 300 K using a Langevin thermostat with a coupling coefficient of 5/ps for the rest of the equilibration and for all subsequent production runs. All subsequent stages were carried out in the isothermal isobaric (NPT) ensemble using a Berendsen barostat[30] with a target pressure of 1 bar and a pressure coupling constant of 0.1 ps. The systems were equilibrated for 200 ps while maintaining the force constants on the restrained atoms to allow for thorough solvation of the complex and to prevent premature flap collapse.[31]

This was followed by a mutation relaxation protocol to allow for optimal reorientation of all mutated amino acids. The heavy atoms of each mutated amino acid and those of amino acids within a 5 Å

(14) Krohn, A.; Redshaw, S.; Ritchie, J. C.; Graves, B. J.; Hatada, M. H. *J. Med. Chem.* **1991**, *34*, 3340−3342.
(15) Hong, L.; Zhang, X. C.; Hartsuck, J. A.; Tang, J. *Protein Sci.* **2004**, *9*, 1898−1904.
(16) Maschera, B.; Darby, G.; Palu, G.; Wright, L.; Tisdale, M.; Myers, R.; Blair, E. D.; Furfine, E. S. *J. Biol. Chem.* **1996**, *271*, 33231−33235.
(17) Ermolieff, J.; Lin, X.; Tang, J. *Biochemistry* **1997**, *36*, 12364−12370.
(18) Shenderovich, M. D.; Kagan, R. M.; Heseltine, P. N.; Ramnarayan, K. K. *Protein Sci.* **2004**, *12*, 1706−1718.
(19) Kovalskyy, D.; Dubyna, V.; Mark, A. E.; Korenelyuk, A. *Proteins: Struct., Funct. Bioinf.* **2005**, *58*, 450−458.
(20) Wittayanarakul, K.; Aruksakunwong, O.; Saen-oon, S.; Chantratita, W.; Parasuk, V.; Sompornpisut, P.; Hannongbua, S. *Biophys. J.* **2005**, *88*, 867−879.
(21) Wittayanarakul, K.; Hannongbua, S.; Feig, M. *J. Comput. Chem.* **2007**, DOI: 10.1002/jcc.20821 (published online Sept 11, 2007).
(22) Sadiq, S. K.; Wan, S.; Coveney, P. V. *Biochemistry* **2007**, *46*, 14865−14877.
(23) Frisch, M. J.; et al. *Gaussian 98*; Gaussian Inc.: Pittsburgh, PA, 2001.
(24) Wang, J.; Wolf, R. M.; Case, D. A.; Kollman, P. A. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(25) Wang, J. M.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.
(26) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926−935.
(27) Kale, L.; Skeel, R.; Bhandarkar, M.; Brunner, R.; Gursoy, A.; Krawetz, N.; Phillips, J.; Shinozaki, A.; Varadarajan, K.; Schulten, K. *J. Comput. Phys.* **1999**, *151*, 283−312.
(28) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T. *J. Chem. Phys.* **1995**, *103*, 8577−9593.
(29) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327−341.
(30) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak. J. R. *J. Chem. Phys.* **1984**, *81*, 3684−3690.
(31) Meagher, K. L.; Carlson, H. A. *Proteins: Struct., Funct. Bioinf.* **2005**, *58*, 119−125.

surrounding region of the mutation were relaxed sequentially for every mutation for a duration of 50 ps each. After each mutant region was relaxed for 50 ps, the heavy atoms of that region were again constrained by a force of 4 kcal mol^{-1} Å^{-2} before proceeding to the next step.

This procedure was followed by a gradual force reduction on the ligand from 4 to 0 kcal mol^{-1} Å^{-2} over a 200 ps period in equal stages of 1 kcal mol^{-1} Å^{-2} and then a similar force reduction on the protease from 4 to 1 kcal mol^{-1} Å^{-2} over a period of 150 ps. In the final stage of the equilibration, all constraints were removed from the protease and the system was allowed to evolve, completely unrestrained, up to a total duration of 2 ns. The length of this last stage therefore varied only according to the number of mutations that were incorporated within the system.

The production simulations for each system lasted 10 ns and were also continued in the isothermal−isobaric ensemble. Coordinate trajectories were recorded every 1 ps throughout all equilibration and production runs.

2.3. MM/PBSA Calculations. 2.3.1. MM/PBSA Methodology. The first step of the method is the generation of multiple snapshots from an MD trajectory of the ligand−protein complex, stripped of water molecules and counterions. For every snapshot, a free energy is calculated for each molecular species (complex, receptor, and ligand), and the ligand binding free energy is computed as the difference:

$$\Delta G_{\text{bind}} = G_{\text{complex}} - (G_{\text{receptor}} + G_{\text{ligand}}) \quad (1)$$

Snapshots, equally spaced at 10 ps intervals, were culled from the MD production runs, giving 1000 snapshots for 10 ns. A smaller time spacing between the snapshots makes it problematic to compute the variance of free energy averages because of the persistence of motional correlations on such time scales.[4] For the systems studied here, the correlation time for the decay of energy fluctuations was found to be about 4 ps.

The binding free energy contains an enthalpic and an entropic contribution:

$$\Delta G_{\text{bind}} = \Delta H - T\Delta S \quad (2)$$

The enthalpy of binding ΔH is composed of ΔG_{MM}, the change in the molecular mechanics free energy upon complex formation, and ΔG_{solv}, the solvated free energy contribution. The molecular mechanics energy ΔG_{MM} is further divided into a van der Waals and a Coulomb term,

$$\Delta G_{\text{MM}} = \Delta G_{\text{vdW}} + \Delta G_{\text{elec}} \quad (3)$$

and the solvation free energy is divided into a polar and a nonpolar part,

$$\Delta G_{\text{solv}} = \Delta G_{\text{solv-pol}} + \Delta G_{\text{solv-np}} \quad (4)$$

The PBSA module of the AMBER suite was used to evaluate the electrostatic energy of solvation $\Delta G_{\text{solv-pol}}$. A grid spacing of 0.5 Å was employed for the cubic lattice, and 1000 linear iterations were performed. The nonpolar solvation energy $\Delta G_{\text{solv-np}}$ was calculated from the solvent-accessible surface area (SASA) using the MSMS program,[32] with a probe radius of 1.4 Å, according to the equation

$$\Delta G_{\text{solv-np}} = \gamma \text{SASA} + \beta \quad (5)$$

where the surface tension γ and the offset β were set to the standard values of 0.00542 kcal mol^{-1} Å^{-2} and 0.92 kcal/mol, respectively.

Finally, the change in solute entropy during ligand association, $-T\Delta S$, was estimated by an all-atom normal-mode analysis performed with the AMBER NMODE module. Prior to the normal-mode calculations, the complex, receptor, and ligand were subjected to minimization with a distance-dependent dielectric constant $\epsilon = 4r$ and convergence tolerance tighter than rmsd = 10^{-4} kcal mol^{-1} Å^{-1}. Due to the high computational cost, snapshots equally spaced at 200 ps intervals were selected for entropy calculations, giving 20 and 50 snapshots for the 4 ns and 10 ns trajectories, respectively.

We calculated experimental binding energies from published inhibition constants K_{I} via

$$\Delta G_{\text{exp}} = -RT \ln K_{\text{I}} \quad (6)$$

Inhibition constants K_{I} were obtained experimentally by fitting the measured rate of enzymatic reaction upon inhibitor binding to the Michaelis−Menten equation, recast in terms of IC_{50} values.[16]

2.3.2. Further Discussion of the MM/PBSA Protocol. The choice of internal dielectric constant for evaluating solvation energies has been the subject of debate, with values of 1 typically used in simulations which sample conformational fluctuations and higher values (2, 4, or more) for "static" structures.[33] In previous MM/PBSA calculations of HIV-1 protease, values of 1,[11] 2,[9,10] and 4[20] have been used for the internal dielectric constant. We find that the binding enthalpies corresponding to an internal dielectric constant of 1, combined with the entropic contribution, produce overall energies closest to experimental affinities (Figure S1, Supporting Information). In the absence of entropy terms, computed binding enthalpies need to be scaled down to be brought within the range of experimental data, for instance via a high internal dielectric constant,[34] because of the enthalpy/entropy compensation.[35] Our results on the internal dielectric constant can therefore be interpreted as a consequence of explicitly incorporating conformational sampling, including entropic effects, in the evaluation of binding energies.

The next question we address is the choice of the continuum solvent model and its effect on the accuracy of MM/PBSA free energy ranking. The enthalpies obtained via the AMBER9 PBSA module were compared to those computed with the modified GB model developed by Onufriev et al.[36] and with the finite-difference Poisson−Boltzmann equation solver DelPhi.[37] For the DelPhi calculations, PARSE[38] atomic radii and Cornell et al. charges[39] were employed for all atoms. As expected, the GB approach is the most computationally efficient, while the DelPhi calculations are the most costly (by a factor of approximately 3). The differences in the polar solvation energies arise from different parametrizations of the atomic radii and effective Born radii.[7,40] By comparison with AMBER PBSA, the GB model under-estimates binding enthalpies by roughly 10 kcal/mol and DelPhi by 20−25 kcal/mol (Table S1, Supporting Information). When adding the entropic contributions, which for saquinavir are found to be in the 30 kcal/mol range, the closest match to experimental absolute binding energies is obtained with the AMBER9 PBSA solver. We also find that, for L90M and G48V/L90M, the ranking of binding enthalpies relative to the WT is preserved irrespective of the continuum solvent method (Table S1, Supporting Information), while for G48V the GB model gives the opposite trend from PBSA and DelPhi. We conclude that the PBSA solver is the most appropriate choice for the computation of absolute as well as relative binding energies, and we therefore use it in all subsequent analysis.

Finally, we examine the effect of retaining the structural water molecule on absolute MM/PBSA binding affinities and the ranking thereof. This conserved water molecule forms hydrogen bonds with

(32) Sanner, M. F.; Olson, A. J.; Spehner, J. C. *Biopolymers* **1996**, *38*, 305−320.
(33) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 265−284.
(34) Naim, M.; et al. *J. Chem. Inf. Modell.* **2007**, *47*, 122−133.
(35) Chen, W.; Chang, C. E.; Gilson, M. K. *Biophys. J.* **2004**, *87*, 3035−3049.
(36) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins: Struct., Funct. Bioinf.* **2004**, *55*, 383−394.
(37) Rocchia, W.; Sridharan, S.; Nicholls, A.; Alexov, E.; Chiabrera, A.; Honig, B. *J. Comput. Chem.* **2002**, *23*, 128−137.
(38) Sitkoff, D.; Sharp, K. A.; Honig, B. *J. Phys. Chem.* **1994**, *98*, 1978−1988.
(39) Cornell, W. D.; et al. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(40) Gohlke, H.; Case, D. A. *J. Comput. Chem.* **2004**, *25*, 238−250.

both the drug and the enzyme (at the Ile50 and Ile50′ residues) and plays a critical role in the interactions between HIV-1 protease and its inhibitors.[41] In MM/PBSA studies of HIV-1 protease, Lepsik et al.[12] and Ode et al.[8] made the choice of explicitly modeling the bound water. At the same time, Naim et al.[34] noted that the explicit inclusion of the structural water in HIV-1 protease did not have a noticeable effect on the fit of binding affinities to experimental data. Herein we find that the inclusion of the water molecule alters the binding affinities relative to the WT ($\Delta\Delta G^{WT}$) by no more than 0.5 kcal/mol for L90M and G48V, and by 0.7 kcal/mol for G48V/L90M (Table S2, Supporting Information). However, this is enough to overturn the delicate enthalpy/entropy balance and reverse the trend in binding affinities for G48V and G48V/L90M. The overall ranking of relative binding energies is therefore better without the structural water. We also find that the effect of the flap water on binding enthalpies is to shift them down by 4−9 kcal/mol, producing absolute energies farther from the experimental values (Table S2, Supporting Information).

2.4. Computational Requirements. Our study utilized elements of a grid-based binding affinity calculator (BAC). The BAC is a workflow tool for the binding free energy calculation of HIV-1 protease−ligand complexes; it automates the implementation of the model preparation, MD simulation, and post-production free energy calculation protocols reported herein. The BAC makes use of the Application Hosting Environment (http://www.omii.ac.uk/)[42,43] to submit distributed, workflow-controlled simulations through a single uniform and interoperable interface and to retrieve output data automatically once the simulations have finished. The MD simulations were performed using the UK National Grid Service (www.ngs.ac.uk) and the US TeraGrid (www.teragrid.org), with a wall-clock rate of approximately 6 h/ns. The free energy calculations were carried out using local resources. One MM/PBSA simulation with 400 snapshots required ∼12 h on one Opteron CPU. Entropy computations using a normal-mode treatment are expensive because the code is not parallelized, and each 20 snapshot calculations required ∼30 h on a single Opteron CPU. On the basis of the analysis of the last 4 ns of MD, the total wall-clock time to rank a mutant with respect to the WT was ∼100 h.

3. Results and Discussion

3.1. Time-Series Convergence and Internal Consistency. Previously reported MM/PBSA ranking studies of HIV proteases are based on nanosecond-level MD: Wittayanarakul et al. have performed 1 ns of production MD,[20] Ode et al. have carried out 3 ns[9] and 1.5 ns simulations,[8] and Hou et al. have carried out 1.5 ns MD simulations.[10] However, MM/PBSA energy estimates and ranking are reliable only if the average energies are converged, which is conditional upon adequate conformational sampling and may require longer MD trajectories.[40,44]

By performing 10 ns of unrestrained dynamics for each protease−inhibitor complex, we seek to address in more depth questions regarding stability, convergence, and internal consistency of MM/PBSA energy estimates and the ranking thereof. We also seek to understand how to make the best use of the MM/PBSA data in order to achieve the level of accuracy and discrimination required by a clinical decision system.

First, we present the time evolution of various free energy components for each of the four complexes (Figure 1). As can be seen in Figure 1, fluctuations of the enthalpic components between different snapshots are highest for Coulomb contributions (up to 15 kcal/mol), albeit they are compensated by corresponding fluctuations in the polar solvation contribution, and are lowest for the cavity energy. In the WT system, the time series of polar energy contributions (Coulomb and solvation energy) suggests that a different substate is sampled during the third and fourth nanoseconds of MD (Figure 1). Further inspection reveals that, in this conformation, the inhibitor is still coupled to G48 via a hydrogen bond but with O4 in an unfavorably close contact with the D30 side chain, which explains the rise in enthalpy. This alerts us to the fact that the sampling of different conformational substates will affect the MM/PBSA binding energy estimates and ranking thereof.

To assess the convergence of the enthalpic contribution to the free energy of binding, ΔH, we calculated the mean (μ) and standard error ($\sigma = \sigma_{sd}/N^{1/2}$) of ΔH as a function of both the forward (Δt_f) and reverse (Δt_r) time intervals from the beginning and end of the 10 ns trajectories, respectively. σ_{sd} is the standard deviation from N snapshots, where $N/\Delta t = 100$ ns^{-1}. Figure 2 shows the results of this analysis. The mean values of ΔH converge to within 0.5 kcal/mol at $\Delta t_f = 5$ ns for all systems (see Figure 2a), indicating that it is necessary to obtain a sample size that extends to at least 5 ns of the trajectory in the forward direction. Determining convergence properties in both temporal directions allows for differences between the earlier and later portions of each trajectory to be distinguished. In the reverse direction, ΔH converges to within 0.5 kcal/mol at $\Delta t_r = 4$ ns for all systems (see Figure 2c). The maximum variation between the last nanosecond and the first is approximately 1 kcal/mol. Interestingly, inclusion of the second nanosecond of the WT trajectory causes a deviation of ∼1 kcal/mol from the converged value, indicating that this portion of the trajectory is not optimal for the assimilation of binding data. This result will be further used in the assessment of internal consistency. The standard error for all systems decays, at the expected $1/N^{1/2}$ rate, with increased N.

The convergence of the entropic component of binding, $T\Delta S$, was determined in the same way as described for the enthalpy (see Figure 3), with the exception that the number of snapshots used was much smaller across the 10 ns trajectory ($N/\Delta t = 5$ ns^{-1}). The mean values of $T\Delta S$ converge to within 0.5 kcal/mol of each other at $\Delta t_f = 6$ ns for all systems (see Figure 3a). The standard error, σ, for the first 4 ns shows deviation from an expected decay rate (see Figure 3b), followed by reversion to normal decay with an increase in snapshot number.

In the temporally reversed direction, $T\Delta S$ also converges to within 0.5 kcal/mol at $\Delta t_r = 6$ ns for all systems (see Figure 3c), while the maximum variation between the last nanosecond and the first is approximately 2 kcal/mol. Incidentally, the value for the entropy after 4 ns of sampling is within the threshold for convergence. However, due to an ∼1 kcal/mol deviation exhibited after 5 ns, the data set does not converge until 6 ns into production. Similar deviations are observed for the decay of σ in the initial 3 ns, followed by normal decay with an increase in N (see Figure 3d). This may be a consequence of insufficient conformational sampling, coupled with the fact that the entropy exhibits large variations across the entire 10 ns.

To further investigate the sensitivity of enthalpy/entropy estimates to the selection of MD data and to conformational transitions therein, we next perform an assessment of internal

(41) Lu, Y.; Yang, C.-Y.; Wang, S. *J. Am. Chem. Soc.* **2005**, *128*, 11830−11839.
(42) Coveney, P. V.; Saksena, R. S.; Zasada, S. J.; McKeown, M.; Pickles, S. *Comput. Phys. Commun.* **2007**, *176*, 406−418.
(43) Sadiq, S. K.; Zasada, S. J.; Wright, D.; Stoica, I.; Coveney, P. V. A molecular simulation based automated binding affinity calculator for ligand-bound HIV-1 proteases. Preprint, 2008.
(44) Pearlman, D. A. *J. Med. Chem.* **2005**, *48*, 7796−7807.

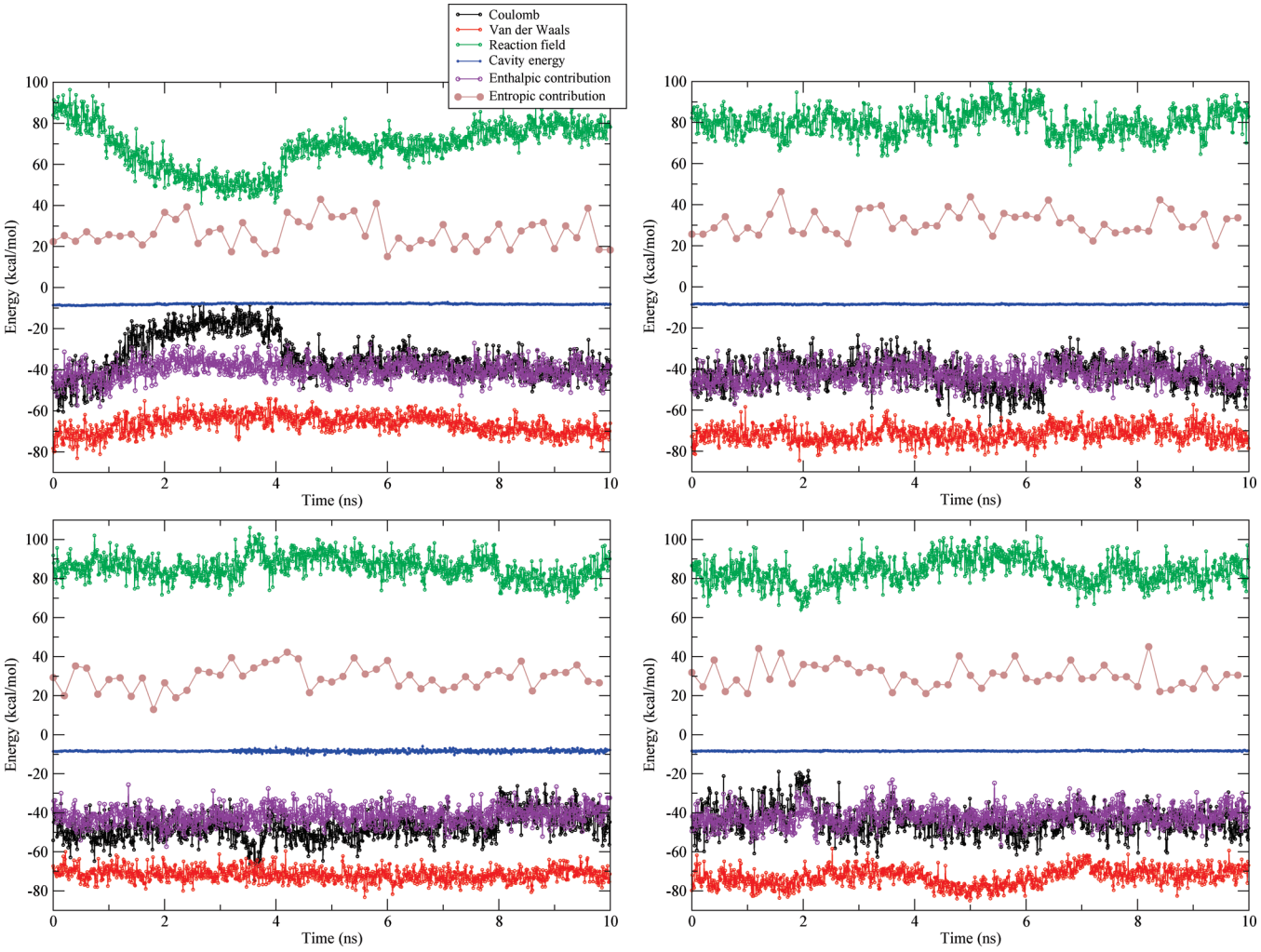


Figure 1. Time evolution of binding free energy components for the WT (upper left panel), L90M (upper right panel), G48V (lower left panel), and G48V/L90M (lower right panel). The entire 10 ns of MD, following unconstrained equilibration for 2 ns, was used for this analysis.

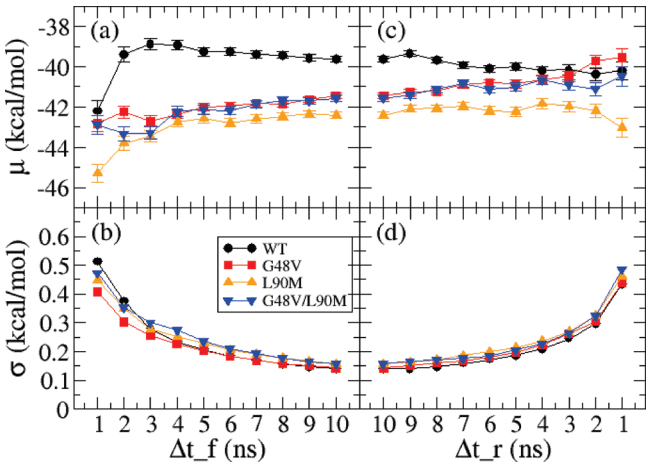


Figure 2. Convergence of the enthalpic component of binding, ΔH, assessed by the mean (μ) and the standard error ($\sigma = \sigma_{sd}/N^{1/2}$) of (a,b) the forward ($\Delta t_f$) and (c,d) the reverse (Δt_r) time intervals, across the 10 ns trajectories for each drug−protease system.

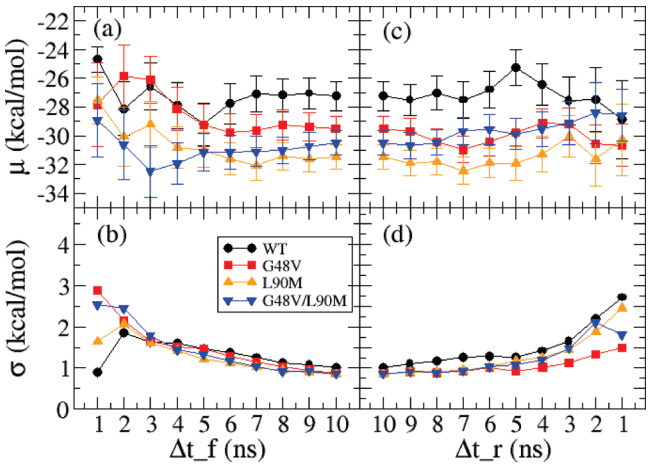


Figure 3. Convergence of the entropic component of binding, $T\Delta S$, assessed by the mean (μ) and the standard error ($\sigma = \sigma_{sd}/N^{1/2}$) of (a,b) the forward ($\Delta t_f$) and (c,d) the reverse (Δt_r) time intervals across the 10 ns trajectories for each drug−protease system.

consistency. In Table 1, we compare enthalpic and entropic contributions based on snapshots extracted from the first 4 ns and the last 4 ns of dynamics.

As can be seen in Table 1, in the mutant complexes the binding enthalpy averages over the first and the last 4 ns differ by 0.1 kcal/mol (L90M) up to 1.7 kcal/mol (G48V), which is well within the expected range for such applications.[40] For the WT system, a more appropriate choice to characterize the substate described previously (with O4 of saquinavir and the D30 carboxylate group in unfavorably close 3.5 Å range) is

Table 1. Impact of MD Data Selection on the Thermodynamics of Saquinavir Binding to HIV Proteases[a]

system		ΔG_{elec}	ΔG_{vdW}	$\Delta G_{solv-pol}$	ΔH	$-T\Delta S_{vib}$	$-T\Delta S_{tot}$
WT	3rd−4th ns	−18.59(0.29)	−63.11(0.23)	52.29(0.34)	−37.27(0.26)	0.87(1.82)	26.24(1.82)
	last 4 ns	−39.29(0.22)	−68.89(0.18)	76.05(0.30)	−40.20(0.21)	1.05(1.45)	26.44(1.45)
L90M	first 4 ns	−41.21(0.31)	−72.23(0.19)	79.13(0.27)	−42.74(0.25)	6.41(1.47)	31.82(1.47)
	last 4 ns	−42.03(0.34)	−70.53(0.21)	78.20(0.34)	−42.84(0.24)	5.88(1.29)	31.29(1.29)
G48V	first 4 ns	−49.30(0.32)	−70.75(0.17)	86.11(0.29)	−42.36(0.23)	2.73(1.55)	28.14(1.55)
	last 4 ns	−43.95(0.39)	−72.12(0.17)	83.78(0.30)	−40.66(0.22)	3.94(1.04)	29.35(1.04)
G48V/L90M	first 4 ns	−41.73(0.43)	−73.23(0.21)	81.17(0.30)	−42.21(0.28)	6.53(1.48)	31.92(1.48)
	last 4 ns	−44.55(0.28)	−70.66(0.19)	82.86(0.29)	−40.66(0.23)	4.61(1.23)	30.02(1.23)

[a] Mean energies are in kcal/mol, with corresponding standard errors in parentheses. Snapshots are taken every 10 ps for the enthalpy estimates, and every 50 ps for the entropy estimates.

the portion of the trajectory extending from 2 to 4 ns. For this selection of the trajectory, the Coulomb energies are substantially more repulsive, and the van der Waals interactions less favorable, than those over the last 4 ns of MD (Table 1). The overall difference in binding enthalpy averages is 3 kcal/mol, well beyond statistical errors and higher than for the rest of the proteases, giving a measure of variability due to the sampling of different local minima.

The major source of uncertainty in MM/PBSA calculations resides in the vibrational entropy component.[7] Large fluctuations in entropies computed via the normal-mode methodology may arise from a mismatch between the minimized geometries of the complex and of the receptor or ligand.[7] Moreover, the normal-mode methodology does not incorporate the configurational entropy associated with multiple energy minima and with transitions therein, leading to an incomplete representation of the entropic cost for binding.[45] Herein we find differences of up to 18 kcal/mol between successive snapshots (Figure 1) and standard deviations of about 1.5 kcal/mol (with 20 snapshots), which are typical for vibrational entropies computed by normal-mode analysis.[7,40] Oscillations in vibrational entropy are higher in the WT system (Figure 1), in the regions of the trajectory associated with conformational fluctuations (third and fourth nanosecond). The average entropy estimates are remarkably consistent, despite the large errors within the calculations. Differences in vibrational (and total) entropies averaged over the first and last 4 ns are less than 1 kcal/mol in the WT, L90M, and G48V systems and almost 2 kcal/mol for G48V/L90M (Table 1). The standard deviations are lower by up to 0.5 kcal/mol for the averages taken over the last 4 ns.

Overall, our analysis shows that sampling of at least 4 ns of the 10 ns trajectory, at a rate of $N/\Delta t = 100$ ns^{-1}, is necessary to obtain converged enthalpies of binding and that at least 6 ns of sampling, with $N/\Delta t = 5$ ns^{-1}, is necessary for the convergence of the entropy. Convergence analysis shows that, for both the enthalpy and the entropy, averaged energies derived from the latter portion of each trajectory are closer to the converged values than those derived from the earlier portions. This is likely due to further structural readjustments well into the post-equilibration phase (see section 3.5) and supports the notion that the entire 10 ns trajectory is not preferable to the latter several nanoseconds of each trajectory in the calculation of the free energy of binding.

To conclude, we find that the last 4 ns of dynamics exhibit structural stability and result in converged enthalpy and entropy estimates. We therefore use these binding energy values for subsequent ranking of affinities and for free energy decomposition. At the same time, the relatively high standard deviations (up to 1.5 kcal/mol if entropy is included) and the limited internal consistency of the MM/PBSA data make it challenging to accurately discriminate variants with only 1 order of magnitude difference in drug resistance relative to WT, such as L90M.

Table 2. Computed and Experimental Absolute Free Energies of Binding for Saquinavir-Bound WT, L90M, G48V, and G48V/L90M Proteases[a]

complex	ΔG_{theor}	ΔG_{exp1}	ΔG_{exp2}
WT	−13.76(1.46)	−14.30	−13.76
L90M	−11.55(1.31)	−12.51	−13.09
G48V	−11.31(1.06)	−11.28	−12.16
G48V/L90M	−10.64(1.25)	−10.21	−10.04

[a] Mean contributions are in kcal/mol, with corresponding standard deviations in parentheses.

3.2. Absolute and Relative Ligand Binding Energies. For all complexes, there are two sets of published inhibition constants.[16,17] Differences between the experimental binding affinities, for a given mutant, are within the "chemical" accuracy of 1 kcal/mol and may be attributed to different experimental conditions, including different pH values and temperatures. In Table 2 we present the absolute binding energies obtained from the MM/PBSA analysis, ΔG_{theor}, as well as those calculated from the inhibition constants published in refs 16 and 17, identified as sets 1 and 2, respectively. Several other K_I and isothermal titration calorimetry (ITC) measurements can be obtained for saquinavir by searching the Binding Database,[46] but no data are available for the mutants studied here.

As can be seen in Table 2, the MM/PBSA absolute binding energies exhibit a remarkable level of agreement of 0−1.5 kcal/mol to experimental binding energies. The correlation coefficients to experimental binding affinities are also high: 0.96 and 0.81, respectively.[16,17] The agreement is better for the data reported by Maschera et al.,[16] where the experiments were conducted at 298.15 K and pH = 6.5. Our measured average temperatures for all four MD simulations are 299.2 K, and our entropy calculations assume a temperature of 298.15 K, so the higher correlation to set 1 may be a result of similar conditions between experiment and simulation. The Ermolieff et al. data were derived at 310.15 K and pH = 5.0.[17] Our absolute binding affinities are all within 1 kcal/mol from the Maschera et al. data,[16] matching the level of "chemical" accuracy.

(45) Chang, C. E.; Chen, W.; Gilson, M. K. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 1534−1539.

(46) Liu, T.; Wen, Y.; Jorrisen, R. N.; Gilson, M. K. *Nucleic Acids Res.* **2006**, *35*, D198−D201.

Table 3. Computed and Experimental Relative Binding Affinities for Saquinavir-Bound L90M, G48V, and G48V/L90M Mutants[a]

complex	$\Delta\Delta H^{WT}_{theor}$	$-T\Delta\Delta S^{WT}_{theor}$	$\Delta\Delta G^{WT}_{theor}$	$\Delta\Delta G^{WT}_{exp1}$	$\Delta\Delta G^{WT}_{exp2}$
L90M	−2.64	4.85	2.21	1.79	0.67
G48V	−0.46	2.91	2.45	3.02	1.60
G48V/L90M	−0.46	3.58	3.12	4.09	3.72

[a] All values are in kcal/mol.

In Table 3 we present MM/PBSA and experimental relative binding affinities, together with their breakdown into enthalpic and entropic components. In the following, the term "relative" refers to differences in binding energy between a mutant and the WT complex and is marked by a "WT" superscript:

$$\Delta\Delta X^{WT}_{method} = \Delta X^{mut}_{method} - \Delta X^{WT}_{method} \quad (7)$$

where $X = H$, S, or G and "method" refers to theory or experiment.

L90M causes a marginal increase in resistance, with experimental relative binding energies $\Delta\Delta G^{WT}_{exp}$ smaller than 1.8 kcal/mol. From our previous discussion on the range of errors in MM/PBSA estimates, it follows that the level of L90M drug resistance is particularly challenging to discriminate. The MM/PBSA-calculated $\Delta\Delta G^{WT}_{theor}$ for L90M is higher than the experimental relative binding energies (Table 3), but it is outside the error bars and thus reliably predicts a more unfavorable binding energy for L90M relative to WT. For G48V, experiments reveal increases in binding energy of 3.0 and 1.6 kcal/mol, respectively,[16,17] relative to WT. As can be seen in Table 3, the MM/PBSA calculation successfully discriminates G48V against the WT complex (beyond the precision bounds). Moreover, the MM/PBSA relative binding energy for G48V is within the range defined by the two experimental values (Table 3). For G48V/L90M, the $\Delta\Delta G^{WT}_{theor}$ value also indicates good ranking against WT: 3.12 kcal/mol, within 1 kcal/mol of the experimental relative binding affinities, $\Delta\Delta G^{WT}_{exp}$ (Table 3).

To conclude, despite the relatively large errors inherent to the methodology, our converged free energy estimates reproduce the observed progression of the WT, L90M, G48V, and G48V/L90M levels of resistance, beyond the error bars. We also obtain an excellent accuracy of 1.5 kcal/mol for the absolute binding energies compared to experimental binding affinities.

3.3. Free Energy Decomposition. A summary of the thermodynamic contributions to binding in WT, L90M, G48V, and G48V/L90M systems is given in Figure 4. In all four protease−inhibitor complexes, the van der Waals interactions and the cavity energy resulting from the burial of saquinavir's hydrophobic groups are the basis for favorable binding free energies (Figure 4). The favorable Coulomb interactions within the inhibitor−protease complex are opposed by the unfavorable electrostatics of desolvation. The resulting balance of the two contributions, $\Delta G_{elec-tot}$, is unfavorable to binding in all four systems. $\Delta G_{elec-tot}$ is virtually indistinguishable (within the error bars) in L90M from the WT system, but it is more unfavorable by 1.5 kcal/mol in the double mutant and by as much as 3 kcal/mol in G48V. The polar energy term contributes to the reduced binding affinity of G48V and G48V/L90M relative to the WT system. The cause is likely to reside in a destabilization of electrostatic interactions at the ligand binding site, arising from a disruption of the hydrogen bonds which couple the inhibitor to the flap, as a result of the G48V mutation (see section 3.5).

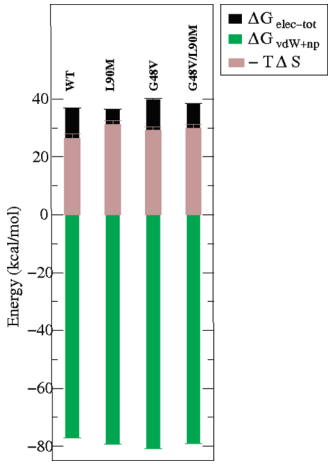


Figure 4. Thermodynamic decomposition of MM/PBSA binding energies for saquinavir with the four proteases under study, WT, L90M, G48V, and G48V/L90M, based on analysis of the last 4 ns of post-equilibration MD.

3.4. Enthalpy/Entropy Compensation and the Importance of the Entropic Contributions. To understand the energetic determinants of drug resistance at the molecular level, we need to consider the interplay between the enthalpic and entropic contributions to binding.[47,48] Enthalpic contributions provide a measure of the strength of the interactions between the inhibitor and the protein (hydrogen bonds, van der Waals interactions), relative to those with the solvent. Entropic contributions comprise the change in solvent entropy arising from the burial of hydrophobic groups upon binding and the loss of solute conformational degrees of freedom (translational, rotational, and vibrational).[48,49]

The first generation of HIV-1 protease inhibitors are structurally constrained, entropically optimized molecules which poorly accommodate the binding site variations due to mutations, leading to a high susceptibility to drug resistance.[48] Because of the major role played by entropy in the binding of such inhibitors, it is likely that entropic effects are also essential in defining the changes in ligand−protease interactions which lead to resistance upon mutation. An understanding of entropic effects is also essential for the design of future resistance-evading drugs, where the challenge is to optimize binding affinity while allowing for enough flexibility in the inhibitor to lower susceptibility to mutations.[45,47] Isothermal titration calorimetry has proven invaluable in dissecting the enthalpic and entropic components of binding for protease inhibitors and various protease mutants;[50] however, the changes in solute configurational entropy upon binding are not provided in isolation from the solvent entropy, and thus cannot be compared directly to simulations.[45] The incorporation of solute entropy in binding affinity calculations is a complex task which has been the focus of recent work.[45,51] It has been shown that the changes in configurational entropy are often substantial and that the omission of entropic effects leads to an overestimate of binding affinities[45] and degrades the correlation with experimental binding energies.

(47) Lafont, V.; Armstrong, A. A.; Ohtaka, H.; Kiso, Y.; Amzel, L. M.; Freire, E. *Chem. Biol. Drug Des.* **2007**, *69*, 413−422.
(48) Velazquez-Campoy, A.; Muzammil, S.; Ohtaka, H.; Schon, A.; Vega, A.; Freire, E. *Curr. Drug Targets Infect. Dis.* **2003**, *3*, 311−328.
(49) Gohlke, H.; Klebe, G. *Angew. Chem., Int. Ed.* **2002**, *41*, 2644−2676.
(50) Ohtaka, H.; Schon, A.; Freire, E. *Biochemistry* **2003**, *42*, 13659−13666.
(51) Gilson, M. K.; Zhou, H.-X. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21−42.

Here we find that the loss of configurational entropy upon binding of saquinavir is appreciable, opposing ligand association by 25−30 kcal/mol (Table 1), which is consistent with reported entropy estimates for HIV-1 protease inhibitors.[9,10,45] Despite the uncertainties in the computation of entropic contributions, a trend emerges from the normal-mode calculations: the vibrational entropies are more unfavorable by ∼3−5 kcal/mol in the mutants versus the WT complex. These differences are well beyond the error bars and are internally consistent (Table 1).

Only upon the inclusion of the entropic component are we able to obtain agreement with the experimental ranking of binding affinities. We find that the enthalpic and entropic contributions are anti-correlated, as previously shown by Gilson et al.,[35] and that the L90M, G48V, and G48V/L90M mutations present a mechanism for enthalpy/entropy compensation which consists of a (modest) enthalpic gain while sustaining a higher entropic penalty.

We note that, in an MM/PBSA study focusing on the L90M mutation, Ode et al.[8] also obtained a more favorable van der Waals contribution for the binding of saquinavir in L90M over the WT, but, because of the interplay of polar and nonpolar terms, their computed binding enthalpy is more unfavorable in L90M by 2.7 kcal/mol overall. At the same time, in a recent ITC study of HIV-1 protease mutations emerging from treatment with tipranavir, Muzammil et al.[52] identified a behavior similar to the one identified here: i.e., tipranavir-resistant mutations exhibit actual enthalpy gains which compensate for the increased entropy cost of binding.

As discussed previously, in G48V it is the balance of protein−ligand Coulomb interactions and polar solvation energies, $\Delta G_{\mathrm{elec-tot}}$, which underpins the drop in affinity relative to the WT complex. By compensation, in G48V the increase in configurational entropy relative to WT is more modest: less than 3 kcal/mol, compared to 4.85 kcal/mol in L90M. In G48V/L90M, the G48V compensatory mutation also alters the enthalpic signature of saquinavir by enlarging the gap between Coulomb and polar desolvation contributions (Figure 4), although by a smaller amount than in G48V alone. On the other hand, in G48V/L90M the vibrational entropy is larger by 3.6 kcal/mol than in the WT complex, more than that in G48V alone, but less than that in L90M alone (Table 3).

In all mutant proteases, susceptibility to mutations appears to develop by a destabilization of the enthalpy/entropy balance relative to the WT protease. Susceptibility to the L90M mutation is entropically rather than enthalpically driven. Susceptibility to G48V appears to be more enthalpically driven and is reflected in a more unfavorable electrostatics of binding. The structural analysis presented below offers further insight at the molecular level into mutation-induced changes in drug−protein interactions.

3.5. Structural Analysis. For ligands binding to HIV-1 protease, a full binding event comprises potential changes in drug−protein conformations and/or protonation states of the active catalytic dyad. The flap-opening event in the HIV-1 protease presents several different conformations, with the equilibrium in favor of a semi-open one.[53] Conversely, the flaps are more predominantly stabilized in the closed conformation upon ligand binding. Lepsik et al.[12] have addressed the free energy changes in altering flap conformations. Recent studies have also looked at the alteration of the protonation state upon binding. Kovalskyy et al.[19] have suggested that the equilibrium protonation changes from dianionic to mono-protonated upon or after ligand binding. Conversely, Wittayanarakul et al.[21] have recently modeled the free energy changes upon altering protonation for a range of inhibitors from D25 protonation in the apo-protease.

Furthermore, while a decomposition of binding free energies into constituent parts aids a thermodynamic understanding of binding differences between mutants, understanding the kinetic mechanisms underlying such binding events remains intractable by using only free energy methods. Measurements of the kinetic rates of association (k_{on}) and dissociation (k_{off}) provide more insight into kinetic mechanisms. Maschera et al.[16] showed that increased k_{off} in mutants is responsible for a reduction in binding affinity and suggested an increased rate of dissociation for the G48V mutant from a closed-bound to a semi-open-bound state. We have also proposed a simultaneously occurring mechanism for the increase in k_{off} as due to a lateral dissociation mechanism induced by the G48V mutation, when the catalytic dyad is in a dianionic protonation state.[22] As the mono-protonated state shows thermodynamic stability, permitting convergent calculations of free energies, it is likely that the protonation event stabilizes the binding. We therefore also analyzed structural properties of the mono-protonated systems presented in this study.

To better understand the deleterious effects of mutations on the drug's potency, we analyze positional fluctuations, hydrogen bonds, and steric clashes of the inhibitor in the WT and mutant proteases. During the 10 ns of production MD, all four protein−ligand trajectories exhibit low backbone root-mean-square deviation (rmsd) values (smaller than 1.5 Å at all times), which is indicative of stability (Figure S2, Supporting Information). Backbone fluctuations are highest in the WT protein, associated with conformational changes during the first half of the trajectory, and lowest in the G48V system. The thermal fluctuations of the aspartic acid dyad, of the S2 subsite containing D30, and of the S3 subsite containing G(V)48, measured by *B*-factors, are also lower in the mutants than in the WT protease (Figure S2, Supporting Information). Further rmsd data and visual inspection do not reveal substantial differences in flap dynamics as a result of the L90M or G48V mutations (data not shown).

The replacement of L90 with a methionine brings the sulfur atoms of M90 and C95 into unfavorably close contact (Figure S3, Supporting Information). In order to relieve bad contacts and optimize van der Waals interactions, M90 has to adopt a compromise position, which translates into a rearrangement of the binding site, with D30 being brought closer (within hydrogen bond range) to the P2 subsite of saquinavir. This rearrangement is likely to limit the capacity of the active site to accommodate saquinavir's rigid groups[8,15] and can explain the increased entropic penalty in L90M over the WT system.

Previous studies have found that, in the WT complex, the amide group of saquinavir's P2 subsite is hydrogen-bonded to the oxygen atom of G48 in the flap region, while in L90M it is hydrogen-bonded to the carboxylate group of D30.[8] The

(52) Muzammil, S.; Armstrong, A. A.; Kang, L. M.; Jakalian, A.; Bonneau, P. R.; Schmelmer, V.; Amzel, M. L.; Freire, E. *J. Virol.* **2007**, *81*, 5144−5154.
(53) Hornak, V.; Okur, A.; Rizzo, R. C.; Simmerling, C. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 915−920.

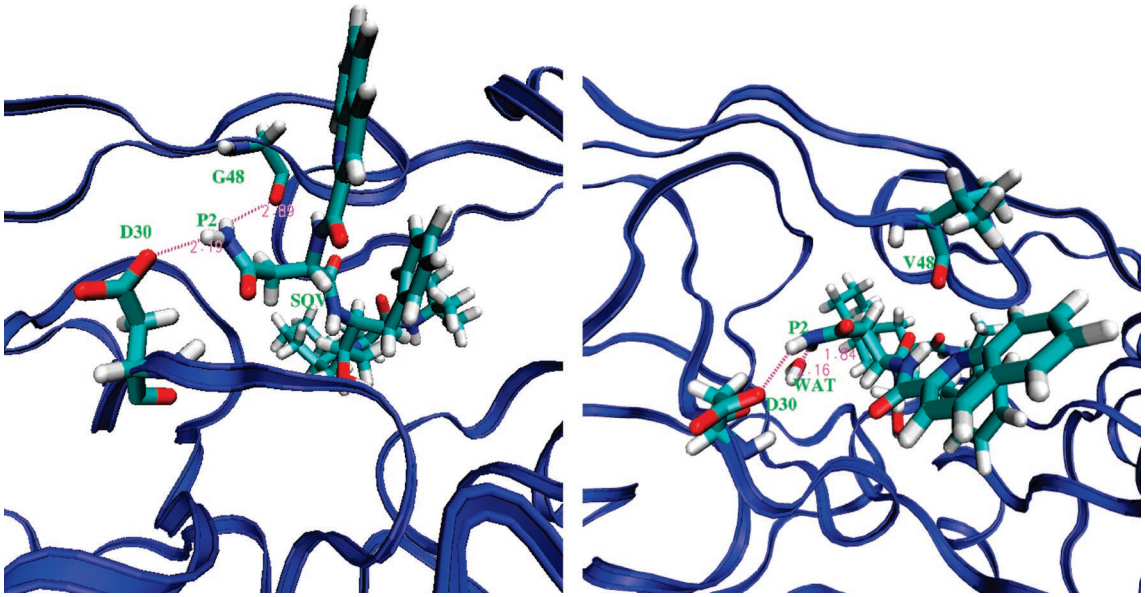


Figure 5. Representative snapshots of the L90M−saquinavir complex (left panel) and the G48V−saquinavir complex (right panel) after 4 ns of post-equilibration MD.

hydrogen bonds presented in Table S3 (Supporting Information) and computed from the first 1 ns of simulation are in agreement with these findings. However, by performing 10 ns of simulation for each protease−saquinavir complex, we are able to accumulate a larger degree of conformational sampling than in previously reported simulations,[8−10,20] and hence to obtain a more complete picture of drug−protein interactions. For instance, in both WT and L90M complexes, hydrogen bonds to both the flap (G48) and the S2 subsite of the enzyme at D30 are sampled over the course of 10 ns. However, since in L90M the D30 side chain is closer to the inhibitor, conformations coexist over the course of L90M dynamics where, unlike in the WT system, the P2 nitrogen is within hydrogen bond range of both G48 and D30 (Figure 5).

In the G48V and G48V/L90M proteases, the bulky valine at position 48 interposes between the quinoline ring of the ligand at the P3 subsite and the phenyl ring of F53, introducing conformational constraints.[15] Additionally, V48 has a stabilizing effect: in G48V, the motions of the P2 subsite and of the quinoline ring of the drug are confined around the starting structure. As a result, the P2 subsite is tightly positioned into the "rotated" conformation captured by the 1FB7 crystal structure over the entire course of the 10 ns of dynamics (Figure 5). Unlike the situation in G48V, rotations of the P2 side chain do occur in G48V/L90M over the course of the simulation which promote hydrogen bonds with the D30 carboxyl group. For instance, during the last 2 ns of MD, the drug remains hydrogen-bonded to D30. This is a consequence of the L90M substitution, which brings the D30 side chain in proximity to saquinavir's P2 subsite.

To summarize, we are able to qualitatively explain the interplay of the enthalpic and entropic components of binding in terms of structural constraints in the various protease mutants. Although distant to the inhibitor, the primary mutation L90M determines a repositioning of D30 relative to the drug and induces perturbations of the binding cavity, which translate into an entropic cost larger by 4.85 kcal/mol than in the WT protein (Table 3). The more favorable nonpolar interactions in L90M compensate in part for the higher entropic penalty, producing a total free energy only marginally more unfavorable than for the WT (Table 3). The experiments also predict marginal resistance levels of 20-fold[16] and 3-fold[17] for L90M.

The stabilizing effect of the G48V mutation may explain the smaller configurational entropy loss in G48V and G48V/L90M versus L90M. On the other hand, in G48V the drug is the most conformationally restricted; this leads to the adoption of substates with unfavorable electrostatic interactions. The accessibility of alternate conformations at saquinavir's P2 subsite, which we have identified as a result of the L90M mutation, will likely relieve some of the unfavorable interactions associated with the valine at position 48 in the G48V/L90M double mutant. Indeed, $\Delta G_{\mathrm{elec-tot}}$ in G48V/L90M is more unfavorable than in L90M but less unfavorable than in G48V (Figure 4). Thus, rather than manifesting a purely cumulative effect, in G48V/L90M the changes in polar energy, nonpolar energy, and entropy relative to those of WT are intermediate in magnitude between L90M and G48V (Figure 4). It is the balance of these components which makes L90M together with G48V a potent combination in terms of their effect on drug resistance.

4. Conclusions

In this study we perform 10 ns of fully unrestrained MD simulations for four saquinavir-bound HIV-1 proteases: WT, L90M, G48V, and G48V/L90M. For each system we identify conformational preferences of the inhibitor and binding site residues and characterize them structurally and energetically. Concerns regarding internal consistency, convergence, and sampling of multiple energy minima indicate that, for the systems studied here, several nanoseconds of MD simulation are necessary for an accurate ranking of binding affinities. Once the convergence criterion is satisfied, we are able to obtain excellent agreement with experimental binding energies, at a level of discrimination suitable for clinical applications. Furthermore, we provide a decomposed thermodynamic description for the basis of observed drug resistance and offer insight into the structure−affinity relationships at the molecular level. Our

results indicate that resistance in the three mutants arises from an interplay of enthalpic and entropic effects, being predominantly entropic in L90M, enthalpically driven in G48V, and intermediate but with a highly altered balance of the two in G48V/L90M. The present results strengthen the argument that, while more flexibility is desirable in future resistance-evading inhibitors in order to lower the entropy-based susceptibility to mutations, one needs to be aware of the enthalpic compensation that may also occur.[47]

While the present study focuses on protein−inhibitor interactions to understand the thermodynamic and structural basis of resistance, in vivo it is the overall viral fitness of a particular sequence that directs its persistence.[54] The emergence of drug-resistant mutations in response to chemotherapeutic pressure is indicative of an interplay between drug binding and the binding of the natural substrates.[55] The vitality metric describes the effect of mutations on the inhibition constant, K_I, as well as on the kinetic parameters for catalysis (k_{cat}/K_m).[55] A direct comparison of inhibitor and substrate binding affinities is a first step in assessing the viral fitness by means of computer simulations; as shown in the present study, the MM/PBSA method can achieve promising accuracy, conditional upon sufficient conformational sampling. A comparative MM/PBSA study of inhibitor and substrate binding should be facilitated by the recently crystallized structures of the naturally cleaved substrates bound to inactive protease.[56]

Acknowledgment. We thank Dr. Shunzhou Wan for helpful discussions. This research was partially funded by EPSRC through RealityGrid grant GR/R67699, by the EU-supported ViroLab project (IST-027446), and by the National Science Foundation under NRAC grant MCA04N014. We also wish to acknowledge the UK NGS (www.ngs.ac.uk) for providing access to their resources and support for this project.

Supporting Information Available: Complete refs 5, 23, 34, and 39; a figure depicting the effect of the internal dielectric constant on MM/PBSA binding enthalpies; two tables illustrating the impact of the continuum solvent method and of the structural water respectively on the MM/PBSA energies; plots of backbone rmsd values and *B*-factors for the four proteases; a table of hydrogen bonds formed by saquinavir during the first 1 ns of MD in all four systems; and a figure showing van der Waals contacts formed by the M90 residue during the first 1 ns of L90M dynamics. This information is available free of charge via the Internet at http://pubs.acs.org.

JA0779250

(54) Deveraux, H. L.; Emery, V. C.; Johnson, M. A.; Loveday, C. *J. Med. Virol.* **2001**, *65*, 218−224.
(55) Gulnik, S. V.; Suvorov, L. I.; Liu, B.; Yu, B.; Anderson, B.; Mitsuya, H.; Erickson, J. W. *Biochemistry* **1995**, *34*, 9282−9287.
(56) Prabu-Jeyabalan, M.; Nalivaika, E.; King, N. M.; Schiffer, C. A. *J. Virol.* **2005**, *78*, 12446−12454.