

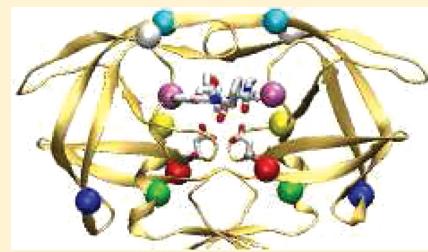
Resolution of Discordant HIV-1 Protease Resistance Rankings Using Molecular Dynamics Simulations

David W. Wright and Peter V. Coveney*

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

 Supporting Information

ABSTRACT: The emergence of drug resistance is a major challenge for the effective treatment of HIV. In this article, we explore the application of atomistic molecular dynamics simulations to quantify the level of resistance of a patient-derived HIV-1 protease sequence to the inhibitor lopinavir. A comparative drug ranking methodology was developed to compare drug resistance rankings produced by the Stanford HIVdb, ANRS, and RegaDB clinical decision support systems. The methodology was used to identify a patient sequence for which the three rival online tools produced differing resistance rankings. Mutations at only three positions (L10I, A71IV, and L90M) influenced the resistance level assigned to the sequence. We use ensemble molecular dynamics simulations to elucidate the origin of these discrepancies and the mechanism of resistance. By simulating not only the full patient sequences but also systems containing the constituent mutations, we gain insight into why resistance estimates vary and the interactions between the various mutations. In the same way, we also gain valuable knowledge of the mechanistic causes of resistance. In particular, we identify changes in the relative conformation of the two beta sheets that form the protease dimer interface which suggest an explanation of the relative frequency of different amino acids observed in patients at residue 71.



INTRODUCTION

In the three decades since its recognition in 1981, AIDS has killed more than 25 million people.^{1–3} World Health Organization (WHO) projections suggest it will remain one of the top three causes of human death until at least 2030.⁴ Despite the unravelling of almost all steps in the life cycle of HIV, the etiological agent that causes AIDS, and the approval of 24 distinct drugs targeting the virus, all efforts to eradicate the virus entirely within infected individuals, have proved unsuccessful. This does not mean that considerable successes have not been achieved; indeed for those with access to modern combination therapy HIV infection has been transformed from a “death sentence” to a controllable disease which requires early diagnosis and life-long treatment. While post diagnosis life expectancy under highly active antiretroviral treatment (HAART) can extend to 21.5 years,⁵ the rapid acquisition of mutations conferring resistance to particular drugs remains a significant cause of treatment failure.

Several studies have demonstrated that improved patient outcomes are correlated with treatment by clinicians with greater experience.^{6,7} However, even for expert clinicians it is frequently impossible to identify straightforward relationships between genotype and drug response. This has resulted in the production and use of computer-based clinical decision support systems (CDSS). These systems are designed to enable medical practitioners to easily keep up to date with the latest clinical advances and incorporate them directly in their decision-making processes. Traditionally the sole source used to inform these support systems has been data gathered from the published literature; the “scoring systems” almost universally give mutations individual scores which are combined in a purely additive fashion. Recently,

it has become recognized that nonlinear combinatorial epistatic effects between mutations play an important role in determining the level of resistance^{8,9} and that the inclusion of data from a broader range of sources, including computational modeling, in decision support software could lead to further improvements in treatment advice.¹⁰ Motivated by this belief, ViroLab (www.virolab.org) created a virtual laboratory consisting of complementary, multilevel computational tools aimed at comparing and enhancing the existing repertoire of decision support approaches.^{11–13} Here we detail the use of molecular dynamics (MD) simulations to investigate an instance where existing systems disagree on the resistance levels produced by a set of mutations (L10I, A71I, and L90M in the HIV-1 protease). The performance of these simulations requires the use of substantial computational resources as well as the production and management of large amounts of data, requirements which have prompted the development of an automated simulation pipeline called the Binding Affinity Calculator (BAC).¹⁴ The workflow from the creation of the system to be simulated to the final analysis can be executed across multiple distributed computational resources by making use of two further tools developed within the ViroLab project, the Application Hosting Environment (AHE)¹⁵ and Grid-Space Engine.¹¹ Molecular-level simulations are not only vastly cheaper than wet lab alternatives but potentially offer the ability to generate atomistic understanding of the causes of resistance. This represents the “lowest level” of the ViroLab philosophy of creating a holistic approach to decision support, from molecule to man.¹³

Received: July 7, 2011

Published: September 09, 2011

VIRTUAL PATIENT EXPERIMENT

The ViroLab virtual laboratory (VL) contains a wealth of tools for investigating the relationship between the HIV genomic sequence and the level of resistance to antiretroviral drugs.¹¹ The Virtual Patient Experiment (VPE) was designed to show the potential of integrating diverse systems such as traditional drug ranking systems, literature mining, patient data, and simulation into a single interface. The aim of the VPE was to take a patient sequence for which the ViroLab comparative drug ranking system (cDRS) provided discordant results for one of the available protease inhibitors and to use the other tools within the virtual laboratory to produce the sort of insight that could help a clinician facing a decision on how to treat this virtual patient. The cDRS allows the user to simultaneously obtain drug resistance rankings (susceptible, intermediate, or resistant) for an input sequence (or set of mutations) from three well-established drug ranking systems: Stanford HIVdb (hivdb.stanford.edu), ANRS (www.anrs.fr), and RegaDB (www.rega.kuleuven.be/cev/regadb/). A clinician treating our unfortunate hypothetical patient could also access a patient database containing anonymised information on thousands of patient treatment histories with associated viral load and CD4 counts and query it for cases in which the relevant mutations were encountered to see how patient and virus had responded to different treatment regimens. In addition to refining and integrating well-established decision support tools, the ViroLab environment seeks to enhance these by building connections to modeling techniques from different areas. The focus of this article is one of these modeling approaches, the evaluation of the impact of mutations within the HIV-1 protease upon drug binding using molecular dynamics (MD) simulations.

The ViroLab and EuResist (www.euresist.org) databases were queried to find patient sequences that produced discordant results from the different CDSS, resulting in the choice of a sequence containing the mutations L10I, I13V, K14KR, I15V, K20T, L63P, A71IV, V77IV, L90M, and I93L in combination with the drug lopinavir (LPV). [The following versions of the drug ranking systems rule sets were used in determining the sequence used in the VPE: HIVdb 5.1.2, ANRS 17, and Rega 8.0.1.] This sequence was deemed to be susceptible by HIVdb but displayed intermediate resistance according to ANRS and Rega. In instances such as this, the VL provides a tool which allows a clinician or researcher to investigate the cause of the discordance by inspecting the rules used to determine the ranking by each system. The only mutations within the sequence identified by the VL as influencing the ranking were L10I, A71IV, and L90M.

In the context of the Virolab VL, BAC is a tool which can be used not only to rank mutant sequences in terms of resistance but also to provide some level of insight into the way in which different mutations within the sequence combine. It is hoped that this will shed light on the origin of inconsistencies in data derived from other sources. With this in mind, it was decided that rather than investigating the L10I, A71IV, and L90M variants alone we would simulate each individual point mutation and all possible combinations thereof. In addition, the full patient sequence was simulated (using the non-wildtype residues at positions where polymorphisms were detected) with both valine and isoleucine present at position 71. For simplicity, the two full patient sequences shall be referred to as VPE-A71V and VPE-A71I, depending on the amino acid present at position 71.

Table 1. Percentage of Patient Sequences Containing the Mutations Identified As Potentially Resistance Causing in the VPE for Both PI Treatment Naive and Experienced Individuals in Data Sets from the HIVdb and the Work of Garriga et al.^{17,a}

mutation	HIVdb		Garriga et al. ¹⁷	
	treatment naive	PI treated	treatment naive	PI treated
L10I	7.3	41.0 (12.5)	7.7	27.0
A71I	0.0	3.3 (0.0)	0.0	1.1
A71V	4.6	38.0 (7.0)	5.4	20.1
L90M	0.0	43.0 (5.2)	2.9	24.3

^a Figures in brackets are the percentages for those undergoing LPV monotherapy.

The mutations identified for study in the VPE occur with differing frequencies in both naive and treated patients as shown in Table 1. L10I and A71V are infrequent polymorphisms in naive patients which are strongly selected for under protease inhibitor (PI) treatment. Neither A71I or L90M are observed in treatment naive patients but are selected for by drug therapy to very different degrees, with A71I only occurring in 3.3% of treated patients and L90M in 43.0% in the HIVdb data set¹⁶ (these figures are for treatment with either one or more PIs). The HIVdb only contains 57 sequences from patients undergoing LPV monotherapy, analysis of this limited data set suggests that selection for the resistance linked mutations in the VPE is less strong for lopinavir than when other PIs are used. However, a study of 1313 HIV infected individuals in Spain, conducted by Garriga et al.,¹⁷ investigated the effects of therapies containing LPV. This study found similar occurrence frequencies to those reported in HIVdb among treatment naive patients but higher levels of selection under LPV treatment for all mutations under consideration, albeit the level of selection remains less than that found in the HIVdb for general PI treatment.

Binding Affinity Calculator. In order to incorporate molecular models into a CDSS environment, we must identify a property calculable from microscopic simulations which provides information with the potential to predict patient outcomes. The clinical impact of mutations is determined by a number of factors, including the strength of drug binding but also of changes in enzymatic efficacy and interactions with other host or disease factors. In the case of HIV-1, there is evidence that genotype to phenotype mapping correlates strongly with clinically observed outcomes.¹⁸ The assays upon which these conclusions are founded concentrate upon drug binding alone and do not consider other potential confounding factors. While the experiments upon which these studies are based are too time-consuming and expensive to be applied routinely in the clinical context, the results suggest that measurement of the strength of drug binding would be a useful, predictive metric to obtain from simulations.

In order to quantify the strength of drug binding, it is necessary to consider the underlying physics of binding. The binding of reactants at constant temperature and pressure is driven by the minimization of the thermodynamic potential known as the Gibbs free energy, G . The strength of protein-ligand binding is characterized by the change in this potential, ΔG , (also known as the binding free energy) which is given by

$$\Delta G = \Delta H - T\Delta S \quad (1)$$

at absolute temperature T , where ΔH is the change in enthalpy and ΔS the change in entropy upon binding. The more negative

the ΔG value, the more tightly a drug binds to its target. Any attempt to evaluate the relative strength of drug binding equates to an estimate of the changes in ΔG . In this paper, we use the term “binding affinity” as synonymous with the binding free energy, ΔG ; however it is also widely used to refer to the equilibrium association constant for drug and protein, K_a . The two quantities are related via the van’t Hoff equation:

$$\Delta G = -RT \ln K_a \quad (2)$$

where R is the gas constant and T the temperature. A change in binding free energy of 1.4 kcal mol⁻¹ corresponds to a 10-fold change in K_a ; changes of this magnitude result in significantly reduced inhibitor efficacy.

Many approaches are available for calculating binding affinities from MD simulations ranging from the theoretically exact, such as thermodynamic integration (TI), to the largely empirical, such as the linear interaction energy (LIE) method (excellent reviews of the subject are available by Gilson and Zhou¹⁹ and Steinbrecher and Labahn²⁰). The computational requirements of these methods tend to increase considerably as more physical detail is included in the model. The CDSS context means that simulation results must be turned around on a time scale of only a few days. In order to fulfill this requirement, we employ the approximate molecular mechanics Poisson–Boltzmann solvent accessible surface area (MMPBSA) methodology^{21,22} which provides a trade-off between rapidity and accuracy of calculation. This method possesses several limitations for computing absolute binding free energies. It does not implicitly account for free energy differences that arise due to conformational changes upon binding, possible variations in key protonation states, and changes due to explicit water-mediated binding between protein and ligand, all of which can provide significant contributions to the binding free energy.^{23,24} Despite these limitations, our previous work demonstrates that changes in binding affinity of less than 1 kcal mol⁻¹ between HIV-1 protease mutants can be distinguished.²⁵ Closer agreement with experimental binding affinity values can be achieved by incorporating a normal mode (NMODE)²⁶ estimation of the entropic component of the binding free energy.

Both MMPBSA and NMODE computations are applied to configuration snapshots generated over the course of MD simulations. The absolute free energy difference of binding, ΔG_{theor} , calculated using this methodology is given by

$$\Delta G_{\text{theor}} = \langle \Delta G_{\text{MMPBSA}} \rangle_M - \langle T \Delta S_{\text{conf}} \rangle_N \quad (3)$$

Here, $\langle \Delta G_{\text{MMPBSA}} \rangle_M$ denotes the average of the enthalpically dominated MMPBSA calculation over M snapshots, while $\langle \Delta S_{\text{conf}} \rangle_N$ denotes the average change in configurational entropy resulting from NMODE calculations across N snapshots.

The enthalpic value of each snapshot is given by

$$\Delta G_{\text{MMPBSA}} = \Delta G_{\text{vdW}}^{\text{MM}} + \Delta G_{\text{ele}}^{\text{MM}} + \Delta G_{\text{pol}}^{\text{sol}} + \Delta G_{\text{nonpol}}^{\text{sol}} \quad (4)$$

where $\Delta G_{\text{vdW}}^{\text{MM}}$ and $\Delta G_{\text{ele}}^{\text{MM}}$ are the van der Waals and electrostatic contributions to the molecular mechanics free energy difference, respectively, and $\Delta G_{\text{pol}}^{\text{sol}}$ and $\Delta G_{\text{nonpol}}^{\text{sol}}$ are the polar and nonpolar solvation terms, respectively. Modules of the AMBER 9 package²⁷ were used in the evaluation of all components of the MMPBSA calculation. The SANDER module was employed to calculate both molecular mechanics terms ($\Delta G_{\text{vdW}}^{\text{MM}}$ and $\Delta G_{\text{ele}}^{\text{MM}}$), with no cutoff being applied to the nonbonded energies. The electrostatic free energy of solvation, $\Delta G_{\text{pol}}^{\text{sol}}$, was calculated by the PBSA

module. Internal and external dielectric constants of 1 and 80 were used, respectively. A thousand iterations of the linear Poisson–Boltzmann equation were performed on a cubic lattice grid with a spacing of 0.5 Å. The nonpolar solvation energy, $\Delta G_{\text{nonpol}}^{\text{sol}}$, was calculated from the solvent accessible solvent area (SASA) using the MSMS program²⁸ with a 1.4 Å radius probe. The surface tension (γ) and offset (β) were set to the standard values of 0.0052 and 0.92 kcal mol⁻¹, respectively.

A major challenge in the computational calculation of binding affinities is to gain sufficient sampling of the energy landscape to gain converged results. Recent studies in our group²⁹ and by others³⁰ have indicated that using ensembles of short simulations with subtly different initial conditions is a more efficient approach than the use of single long trajectories. The division of the ensemble into a set of small simulations (henceforth referred to as “replicas”) lends itself perfectly to the utilization of distributed resources such as those available on the US Teragrid (www.teragrid.org), UK National Grid Service (www.ngs.ac.uk), and EU DEISA (www.deisa.eu) grids. The execution of a large number of replicas in parallel also provides a significant improvement in terms of the turnaround time comparative to running a single longer simulation. The drawback of this approach is the need to manage the data for each replica individually. In response to this we have developed BAC which automates the creation, performance, and analysis of the system, in turn using AHE to transparently implement data transfer and access remote computational resources. BAC was used to coordinate the initial preparation, production molecular dynamics simulations, and postprocessed free energy calculations.¹⁴ The full details of the methodology are provided in the works of Sadiq et al.¹⁴ and Stoica et al.²⁵

Model Preparation. The 1MUI crystal structure³¹ was used as the basis of all the models of protease bound to lopinavir created in this study, with mutations inserted in silico. The mutant protease models were derived from the 1MUI structure using the mutational protocol of the VMD³² visualization package, which was also used to insert the hydrogen atoms not contained in the crystal structure. Each dimeric mutation corresponds to two amino acid substitutions, one on each monomer. In addition to the substitutions required to recreate the mutant sequences under study, the mutation S37N was incorporated into all of the models in order that the model sequence matches the canonical HXB2 subtype B sequence (Genbank accession number K03455) used by Ohtaka et al.³³ Inserting mutations is not expected to disrupt the protease structure as comparisons of crystal structures indicate that the tertiary structure of the enzyme is stable despite considerable variation in the amino acid sequence.³⁴ Missing hydrogens were inserted on drug coordinates using the PRODRG tool.³⁵ Gaussian 03³⁶ was used to perform geometric optimization of all inhibitors at the Hartree–Fock level with 6-31G** basis functions. The restrained electrostatic potential (RESP) procedure, which is part of the AMBER9 package,²⁷ was used to calculate the partial atomic charges. The Visual Molecular Dynamics (VMD) package³² was used to incorporate protease mutations, to insert all missing protease hydrogens, and to assign the protonation state of the catalytic dyad. The force field parameters for the inhibitors were completely described by the general AMBER force field (GAFF).³⁷ The standard AMBER force field for bioorganic systems (ff03)³⁸ was used to describe the protein parameters as well as those for the natural substrates. It is well-established from the many crystal structures of ligand-bound HIV-1 protease complexes that a water molecule mediates flap–ligand interactions.³⁹ This water molecule

is not resolved in the 1MUI crystal structure; consequently, a water molecule was inserted in the correct tetrahedrally coordinated geometry between lopinavir and the protease flaps. The LEAP module in the AMBER 9 software package was then used to electrically neutralize each ligand-bound system and to solvate using atomistic TIP3P water⁴⁰ in a cubic box with a minimum buffering distance of 14 Å in all three orthogonal dimensions, resulting in a fully atomistic system of approximately 40 000 atoms for each system.

Simulation Protocol. All simulations presented here were performed in the molecular dynamics package NAMD2⁴¹ using the protocol incorporated into the BAC software (based on that originally employed by Perryman et al.⁴²) which has previously been successfully used to calculate binding free energies for a number of inhibitors bound to various HIV protease sequences.^{25,29} The system is first minimized with all protein and ligand heavy atoms constrained to their positions in the initial structure. Each system is virtually heated from 50 to 300 K over 50 ps and then maintained at a temperature of 300 K. Once the system reaches the correct temperature in all subsequent simulation steps the pressure is maintained at 1 bar. This results in the system sampling an isothermal isobaric (*NPT*) ensemble. The simulation proceeds for 200 ps before a mutation relaxation protocol is enacted. The relaxation protocol consists of the sequential release of constraints on each mutated residue (together with any residue within 5 Å) for 50 ps (constraints are maintained on the rest of the protein structure). This allows the residues to reorientate into more favorable conformations if necessary. After the 50 ps relaxation period, the restraints are reapplied to each region. The final equilibration stage is the gradual reduction of the restraining force on the complex from 4 to 0 kcal mol⁻¹ Å⁻² during a 350 ps period. Following this, the systems are allowed to evolve freely (a more detailed description of the equilibration protocol is given in the Supporting Information). The entire equilibration stage is designed to take 2 ns for all systems, meaning that this final stage varies in length according to the number of mutations that require relaxation in the previous stages. After the equilibration is complete, structures are output for analysis every picosecond. Each output snapshot is postprocessed using MMPBSA, meaning that a hundred sets of coordinates are analyzed for each nanosecond of simulation. The more computationally expensive NMODE analysis is performed on every 20 snapshots, producing 5 entropy estimates per nanosecond of simulation. All free energy values reported in the following studies were obtained from ensembles of 50 replica simulations, generating 4 ns of production simulation, varying from one another only in the velocities initially assigned to the atoms in the simulation (these are obtained from a randomly seeded Maxwellian velocity distribution).

Infrastructure. A variety of machines on both the US Tera-Grid and EU DEISA network were used to run our simulations. In particular, we used the Ranger machine (with 62 976 cores) at the Texas Advanced Computing Center (TACC), the Kraken machine (112 896 cores) at the National Institute for Computational Sciences (NICS) at the University of Tennessee, HLRB-II (9728 cores) at the Leibniz-Rechenzentrum (LRZ) in Germany, and the Huygens machine (with 3328 cores) at the SARA supercomputing center in The Netherlands for the production simulations. The optimally scaled rate of computation was approximately 4 h ns⁻¹ on 64 cores replica⁻¹. The postprocessing MMPBSA and normal-mode analysis steps took 3 and 10 h ns⁻¹, respectively, but were run in parallel task farm fashion for each nanosecond using the Leeds node (256 cores) of the UK National Grid Service.

Table 2. Two-Letter Codes and Sequence Composition for the Protease Sequences of the Multidrug Resistant (MDR) Mutants Used to Assess the Simulation Protocol

code	description	mutations
WT	wildtype	HXB2
HM	MDR hexa mutant	L10I, M46I, I54V, V82A, I84V, L90M
QM	MDR quattro mutant	M46I, I54V, V82A, I84V
AS	active site mutant	V82A, I84V
FL	flap mutant	M46I, I54V
DM	dimer interface mutant	L10I, L90M

BENCHMARK VALIDATION STUDY

In order to establish the efficacy of our simulation and analysis protocol, we conducted a study in which we attempted to replicate the experimental results achieved by Ohtaka et al.³³ on LPV bound to a series of multidrug resistant (MDR) protease mutants. Alongside the HXB2 wildtype, there were five mutants comprising subsets of a complete set of six MDR mutations, namely L10I, M46I, I54V, V82A, I84V, and L90M, with varying degrees of resistance. The subsets of mutations have been labeled with two letter codes which are shown in Table 2, and this nomenclature will be used for the remainder of the present paper. HIV-1 protease is a homodimer with C₂-symmetry, and hence, a single dimeric mutation corresponds to two amino acid mutations, one at each identical position along the monomer (see Figure 1). Full details of the results presented here are available in a recently published paper by Sadiq et al.²⁹

Using this set of MDR mutants, we refined our earlier single simulation protocol²⁵ and determined that ensembles of 50 replicas each containing 4 ns of production simulation produced well-converged results. As can be observed in Figure 2, we achieved excellent agreement with the experimental result for both MMPBSA alone, ΔG_{MMPBSA} , and the absolute binding free energy (including a normal mode derived estimate of the entropic contribution), ΔG_{theor} , exhibiting correlation coefficients with the experimental values of 0.98 and 0.89, respectively. We repeated the simulations for the two systems with the most divergent binding affinities (WT and HM) and found our results to be reproducible to within a root-mean-square (rms) error of 0.7 kcal mol⁻¹ for ΔG_{MMPBSA} and 0.8 kcal mol⁻¹ for ΔG_{theor} . It is clear from Figure 2 that there is a systematic error between the computed absolute binding energies and the experimental values; minimization of the deviations of each system suggests that our results are 4.92 kcal mol⁻¹ less attractive than experiment. This value is in line with the observation that our analysis does not explicitly include the binding of a water molecule along with lopinavir, a feature common to all clinically approved protease inhibitors, which has been estimated to contribute between 3 and 6 kcal mol⁻¹ to the binding affinity.²⁹

Despite the closer correspondence between the magnitude of the observed ΔG_{theor} values and experiment the ΔG_{MMPBSA} values show greater discrimination between systems, in particular for the more resistant mutants AS, QM, and HM. In this study, we use comparisons with the WT and HM sequences to provide benchmarks for susceptible and highly resistant sequences, respectively, and the AS mutant as an example of a system of intermediate resistance.

BINDING ASSESSMENT OF MUTANT SEQUENCES

Figure 3 shows comparisons of both the computed binding enthalpy, ΔG_{MMPBSA} , and absolute binding affinities, ΔG_{theor} , for

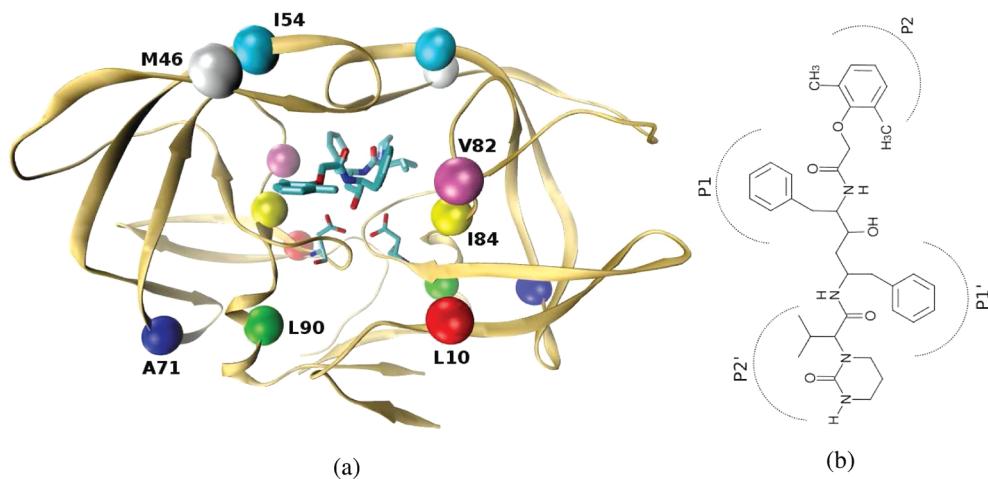


Figure 1. (a) HIV-1 protease (backbone shown in ribbon representation) bound to the inhibitor lopinavir (shown in chemical structure representation along with the catalytic dyad at position 25 of each protease monomer). The locations of the mutations found in the multidrug resistant (MDR) mutants (described in Table 2) used for the benchmark simulations and residue A71 are highlighted. Protease is a homodimer, and the location of each mutation is given the same color on both monomers. (b) Chemical structures of the HIV-1 protease inhibitor lopinavir (arc segments surrounding sections of the drug indicate the moieties which interact with specific regions known as subsites within the protease).

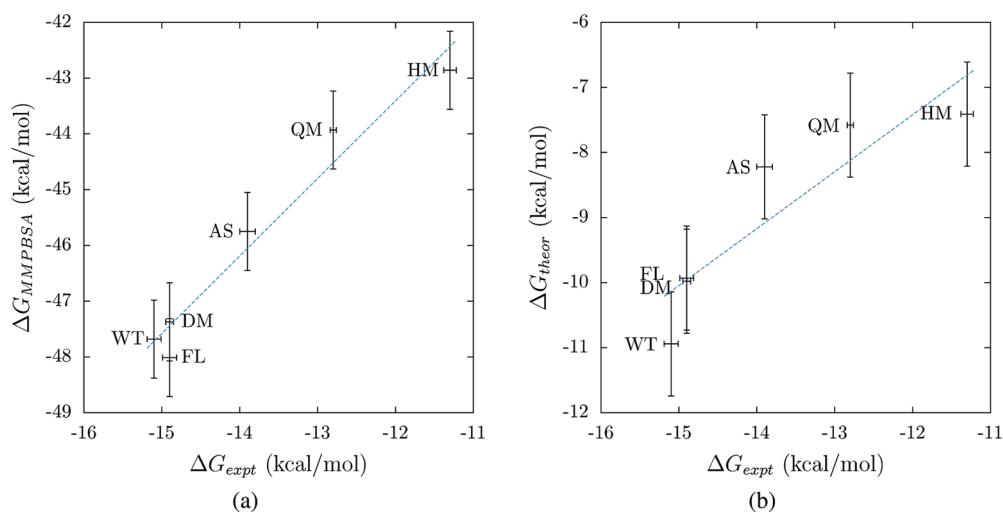


Figure 2. Comparison of average computed binding affinities with those derived experimentally by Ohtaka et al.³³ (a) ΔG_{MMPBSA} , the enthalpically dominated binding affinity from MMPBSA calculation alone; (b) ΔG_{theor} , the absolute binding affinity with the entropic component calculated from normal-mode analysis. The blue line represents a linear regression performed on both data sets which exhibit correlation coefficients between the computed and experimental values of 0.98 and 0.89, respectively. The error bars of 0.7 kcal mol⁻¹ for ΔG_{MMPBSA} and 0.8 kcal mol⁻¹ for ΔG_{theor} were derived from a reproducibility study of the WT and HM systems.

all systems simulated in the VPE along with the WT and HM benchmark systems. The sequences have been separated, for clarity, into subsets containing alanine or valine, and alanine or isoleucine at position 71. In both subsets it is clear that none of the sequences incorporating mutations at position 71 induce any level of resistance using either metric, with the exception of the full patient sequences VPE-A71I and VPE-A71V. A particularly striking difference is exhibited between the triple mutant L10I-A71V-L90M and the same mutations incorporated in the full patient sequence. The VPE-A71V sequence is more resistant than the triple mutant by 3.06 kcal mol⁻¹ using ΔG_{theor} and 2.47 kcal mol⁻¹ according to ΔG_{MMPBSA} .

Both VPE-A71I and VPE-A71V are considerably less resistant than the benchmark HM value using either metric. VPE-A71V

has relative binding affinities with reference to wild type, ΔG_{MMPBSA} and ΔG_{theor} , of 2.07 and 2.24 kcal mol⁻¹, respectively, which are comparable to those for the AS mutant from the benchmark study of 1.93 and 2.72 kcal mol⁻¹. The AS mutant is known to exhibit clinically relevant levels of resistance. The VPE-A71I sequence exhibits a lower level of resistance using both metrics with a $\Delta \Delta G_{MMPBSA}$ value of 0.96 kcal mol⁻¹ and a $\Delta \Delta G_{theor}$ of 1.98 kcal mol⁻¹. These values would lead us to rank both VPE-A71I and VPE-A71V as producing intermediate resistance to lopinavir.

The decomposition of ΔG_{MMPBSA} shown in Table 3 indicates that the origin of the decrease in attraction of VPE-A71V is the polar solvation term, ΔG_{pol}^{sol} , which is 3 kcal mol⁻¹ more repulsive than in the WT case, with both the van der Waals and electrostatic

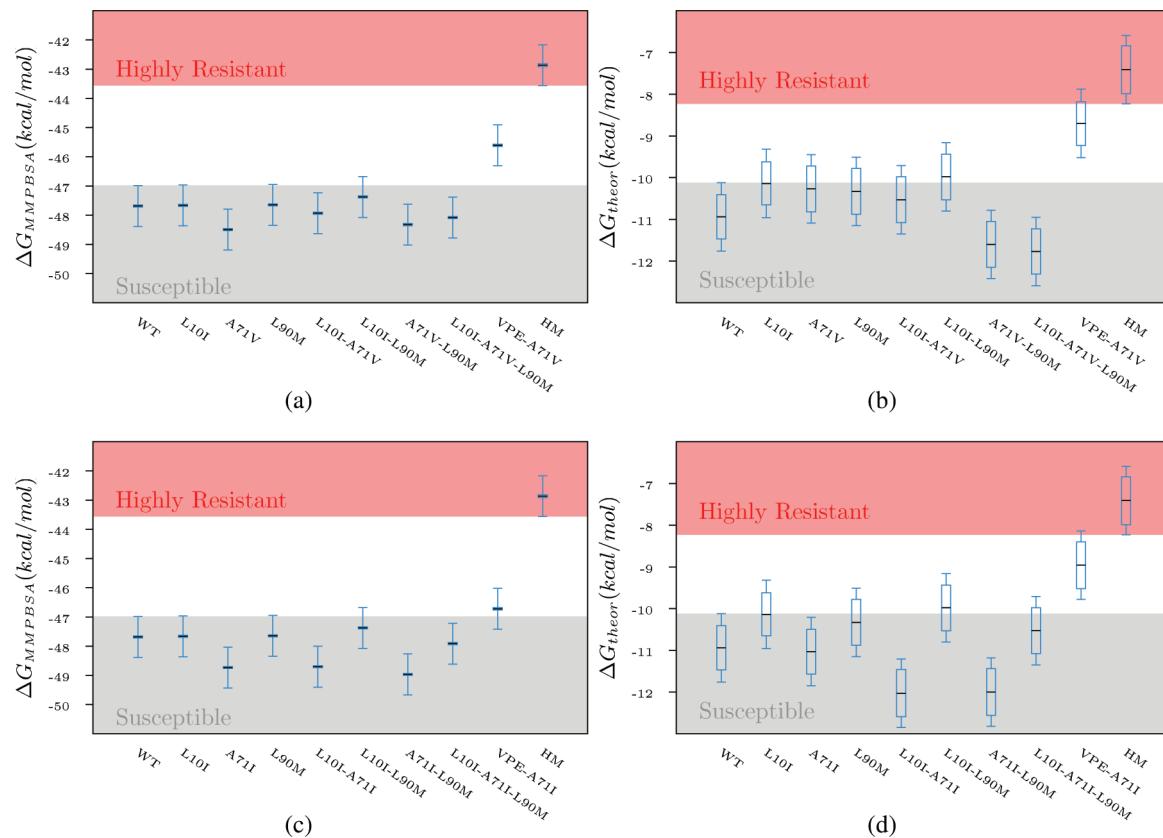


Figure 3. Comparison of the binding free energies computed for all mutants studied in the virtual patient experiment bound to LPV with those of the known susceptible WT sequence and known resistant HM sequence. The enthalpically dominated ΔG_{MMPBSA} and the absolute binding free energy change, ΔG_{theor} are shown for the mutants containing either alanine or valine at position 71 (a and b) and containing either alanine or isoleucine at position 71 (c and d). The black lines show the mean, the candle stick, the standard error, and the whiskers, the error based on the WT and HM reproducibility for each system. The gray and red shaded region show the range of values deemed susceptible and resistant defined using the WT and HM benchmark values.

contributions ($\Delta G_{\text{vdw}}^{\text{MM}}$ and $\Delta G_{\text{ele}}^{\text{MM}}$, respectively) being marginally more attractive. This is a different pattern to that seen in the AS and HM sequences where small changes in polar solvation energy are augmented by considerable reduction in the attractiveness of the van der Waals interactions (directly caused by the mutation of the active site residues 82 and 84). Further differentiation of the resistant VPE-A71V sequence from the resistant mutants in the benchmark study comes from the observation that it (along with all other sequences containing A71V) maintains hydrogen bonds between the hydroxyethylene moiety (labeled O3, nitrogen, and oxygen atoms are separately numbered from the left of the schematic shown in Figure 1b) and the catalytic ASP 125 more frequently than the WT (see Table 4 for a list of hydrogen bond frequencies between LPV and the protease systems under investigation). The resistance of VPE-A71I, like that of VPE-A71V, is primarily caused by changes in the electrostatic components of binding. However, where the majority of the change in the A71V containing variant is in $\Delta G_{\text{sol}}^{\text{pol}}$ in this case $\Delta G_{\text{ele}}^{\text{MM}}$ is more significant (with changes of 0.25 and 0.58 kcal mol⁻¹ compared to WT, respectively). This is reflected in slightly different changes in the active site hydrogen bonding networks between the two patient sequence-derived systems. Unlike VPE-A71V, the bond between catalytic ASP 125 and the lopinavir hydroxyethylene moiety is reduced in occupation compared to WT in VPE-A71I. Both patient derived sequences, however, exhibit reduced frequency of bonding with residues 27 and 29 in a similar fashion to the known resistant mutants from the benchmark study.

Again, in common with the AS and HM systems, both VPE-A71I and VPE-A71V sequences have substantial hydrogen bonds between the oxygen O1 and the backbone nitrogen of residue 30 in the P2' subsite. The increase in bonding frequency with residue 30 is particularly pronounced in VPE-A71I with occupation levels over 60% (compared to less than 30% in VPE-A71V). Together with the binding affinity results, these observations indicate that in order to adequately assess both the level of resistance and the binding mode of a particular sequence and drug combination it is necessary to consider the interactions of mutations throughout the protein structure. This suggests that the discordances between CDSS identified by the cDRS may be a result of nonlinear relationships between larger sets of mutations being poorly handled by the statistical models employed.

Figure 3a indicates that according to the ΔG_{MMPBSA} metric three sequences (A71V, A71V-L90M, and L10I-A71V-L90M) may bind more strongly than the wild type. The largest change in ΔG_{MMPBSA} (of 0.81 kcal mol⁻¹) appears in the A71V single mutant, but this increase in attraction is counteracted by a large change in entropic contribution ($-T\Delta S_{\text{conf}}$) which results in an absolute binding affinity, ΔG_{theor} , value less negative than that computed for the WT (see Figure 3a and Table 3). The increase in binding affinity for the A71V-L90M and L10I-A71V-L90M persists even when entropy is accounted for in ΔG_{theor} , indicating that at least these two mutants containing A71V may be hypersusceptible to LPV. The phenomenon of hypersusceptibility to LPV

Table 3. Decomposed Contributions to the Free Energy of Binding of Lopinavir to Wildtype Protease and Selected Mutant Sequences Studied in the Virtual Patient Experiment^a

sequence	$\Delta G_{\text{vdw}}^{\text{MM}}$	$\Delta G_{\text{ele}}^{\text{MM}}$	$\Delta G_{\text{pol}}^{\text{sol}}$	$\Delta G_{\text{nonpol}}^{\text{sol}}$	ΔG_{MMPBSA}	$-T\Delta S_{\text{conf}}$
WT	-72.82 (0.03)	-51.29 (0.05)	84.56 (0.05)	-8.13 (0.00)	-47.68 (0.03)	36.74 (0.49)
A71I	-73.33 (0.03)	-50.07 (0.05)	82.79 (0.04)	-8.13 (0.00)	-48.73 (0.03)	37.70 (0.50)
A71V	-72.97 (0.03)	-51.04 (0.05)	83.65 (0.05)	-8.13 (0.00)	-48.49 (0.03)	38.23 (0.52)
L10I-A71I-L90M	-72.55 (0.03)	-50.70 (0.05)	83.44 (0.04)	-8.10 (0.00)	-47.92 (0.04)	37.39 (0.51)
L10I-A71V-L90M	-73.31 (0.03)	-52.45 (0.06)	85.81 (0.05)	-8.14 (0.00)	-48.08 (0.04)	36.31 (0.50)
VPE-A71I	-72.63 (0.03)	-50.71 (0.06)	84.81 (0.05)	-8.19 (0.00)	-46.72 (0.04)	37.76 (0.52)
VPE-A71V	-73.32 (0.03)	-51.65 (0.06)	87.56 (0.06)	-8.20 (0.00)	-45.61 (0.04)	36.90 (0.48)
AS	-71.02 (0.03)	-51.97 (0.07)	85.38 (0.05)	-8.13 (0.00)	-45.75 (0.04)	35.57 (0.54)
HM	-70.43 (0.04)	-50.51 (0.08)	86.17 (0.06)	-8.10 (0.00)	-42.86 (0.05)	35.45 (0.52)

^a Details of all systems simulated are provided in the Supporting Information. Mean energies are in kilocalories per mole. Standard errors are shown in parentheses.

Table 4. Frequency of Occupation of Putative Hydrogen Bonds between the Inhibitor Lopinavir (LPV) and the HIV-1 Protease Sequences in Selected Sequences from the Virtual Patient Experiment Calculated over the Production MD Simulations (Consisting of 50 Replicas Producing 4 ns Trajectories) for Each System^a

sequence	donor	LPV O3	LPV N1	LPV N3	LPV N4	LPV O3	LPV N1	29 N	30 N
sequence	acceptor	125 OD2	29 OD1	27 O	125 OD2	125 OD1	29 OD2	LPV O1	LPV O1
WT		0.945	0.772	0.496	0.205	0.179	0.134	0.999	0.051
A71I		0.986	0.832	0.536	0.155	0.134	0.065	0.999	0.025
A71V		0.991	0.811	0.534	0.175	0.137	0.108	0.999	0.044
L10I-A71I-L90M		0.968	0.704	0.510	0.243	0.172	0.114	0.992	0.019
L10I-A71V-L90M		0.987	0.709	0.407	0.347	0.193	0.174	0.991	0.057
VPE-A71I		0.939	0.676	0.414	0.291	0.191	0.122	0.986	0.697
VPE-A71V		0.977	0.543	0.331	0.382	0.161	0.127	0.996	0.209
AS		0.931	0.418	0.234	0.535	0.212	0.235	0.986	0.160
HM		0.819	0.258	0.198	0.408	0.247	0.199	0.960	0.121

^a Details of all systems simulated are provided in the Supporting Information. A hydrogen bond is identified as being formed when a donor–acceptor pair is separated by less than 3.5 Å and the donor–hydrogen–acceptor angle is less than 120°.

has been observed experimentally in a range of protease sequences (particularly those of subtype C viruses) although the clinical impact remains uncertain.^{43–45} A similar pattern is observed in the A71I, L10I-A71I, and A71I-L90M systems (see Figure 3c and d), which have relative binding enthalpy, $\Delta\Delta G_{\text{MMPBSA}}$, values of -1.05, -1.02, and -1.29 kcal mol⁻¹, respectively. While these values are close to the resolution limit of our method, they are greater than the reproducibility variability seen in the WT and HM benchmark systems. This increase in the strength of binding is conserved for the L10I-A71I and A71I-L90M systems when the entropic contribution is included in the results (they have $\Delta\Delta G_{\text{theor}}$ values of -1.09 and -1.06 kcal mol⁻¹, respectively) but not for the A71I single mutant (which is almost indistinguishable from WT with a $\Delta\Delta G_{\text{theor}}$ of -0.09 kcal mol⁻¹).

STRUCTURE AND DYNAMICS

The differences in binding affinity observed between systems reflects changes in the conformations explored. The root-mean-square fluctuations can be used to investigate alterations in flexibility while principal component analysis provides a useful tool for gaining insight into the most significant structural changes. Here we use both techniques to investigate the changes in protein structure and dynamics produced by the mutations under study.

Structural Flexibility. An assessment of the flexibility of a protein bound to LPV during a simulation can be made using the root-mean-square fluctuation (RMSF) of the structures explored during each ensemble relative to the average structure for that sequence. In order to describe the structure of the protease, we adopt the terminology set out by Perryman et al.⁴² According to this naming scheme, the various sections of the structure are labeled according to the supposed resemblance of the protease backbone to a bulldog's face (see Figure 4). The active site of the protease is situated in a cleft which is covered by the region known as the “flaps” (residues 43–58). The flaps are connected to the “ears” (residues 33–42), the catalytic ASP containing “eyes” (residues 23–32, which form the base of the active site), the “cheek sheet” (residues 59–78), the “wall turn” (residues 79–85), the “fulcrum” (residues 11–22), the “nose” (residues 6–10), and the “whiskers” (residues 1–5 and 95–99). In order to distinguish the residues of the two identical monomers that form the HIV-1 protease homodimer, we henceforth number the residues of the two chains from 1 to 99 and 101 to 199, respectively. The monomer with the lower indices, labeled chain A in the 1MUI PDB, contains the P1 and P2' subsites, while the monomer containing the higher indexed residues contains P2' and P1 and is labeled chain B in the PDB.

Figure 5a shows the RMSF of each residue within the WT structure and the differences (ΔRMSF) for all positions in each

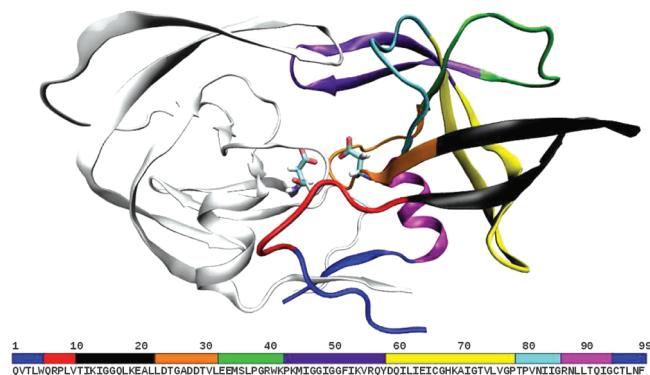


Figure 4. Structure of the HIV-1 protease dimer shown in cartoon representation. The structural elements identified in the “bulldog’s face” description of the structure are shown highlighted in the right-hand monomer with the whiskers in blue, nose in red, fulcrum in black, eyes in brown, ears in green, flaps in purple, cheek sheet in yellow, wall turn in cyan, and the final helix in pink. The catalytic dyad is depicted in chemical structure. The HXB2 wildtype protease sequence is also shown with the positions within it of each of the structural elements indicated using a color bar.

of the sequences in the A71V related subset of the VPE relative to this. In the WT there are two very stable regions in each monomer: the first around the catalytic ASP (25/125) coinciding with the eyes, Ey, region, and the other is the helix formed by residues 86/186 and 94/194. The most flexible regions occur in similar areas in both monomers: the fulcrum (F), residues 41 and 141, and parts of the cheek sheet (C), preceding residue 70/170. With the exception of the A71V and VPE-A71V sequences, all of the systems exhibit broadly similar flexibility to WT; some additional flexibility is shown in the helix and whiskers particularly around residues 87/187, 90/190, and 93/193. In line with the increased entropic barrier to LPV binding, the A71V single mutant exhibits reduced flexibility across the entire structure with an average ΔRMSF of -0.2 \AA . Notable exceptions to this pattern are residues 29 and 108. The loss of mobility is particularly large in the area of the fulcrum of both chains and the flaps of the second chain where losses of up to 0.8 \AA are observed. A noticeable reduction is also seen in the ears (E) and whiskers (W), particularly between residues 95 and 99 in the first monomer and 195 and 199 in the second which are involved in the dimer β sheet. Many of the regions seen to be stabilized in the A71V mutant are more flexible than WT in the VPE-A71V system. It is worth noting that for both of these systems, changes observed in one monomer are almost invariably present also in the other. The region with the most prominently increased fluctuations is the fulcrum where ΔRMSF values of 0.75 \AA are observed, but the flaps, ears, and wall turn (T) also gain flexibility. The wall turn contains key hydrophobic residues involved in interactions with the drug such as V82 and I84. The added flexibility in this region may allow the protein to gain more favorable interactions with LPV, partially explaining the more attractive $\Delta G_{\text{vdw}}^{\text{MM}}$ component of the binding affinity. The region of the cheek sheet around residue 69, in contrast to much of the rest of the structure, is seen to strongly stabilize with a ΔRMSF of under 0.5 \AA . The observed overall gain in flexibility of the VPE-A71V system is not reflected in a higher entropic barrier to binding, suggesting that this flexibility must also be present in the free enzyme as well as the drug bound complex. With the

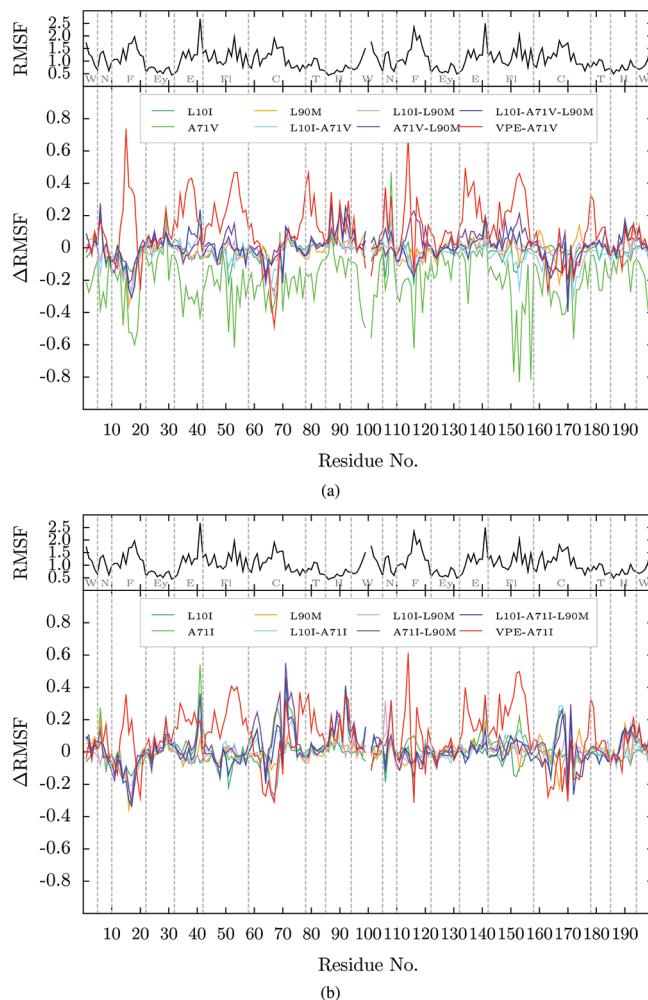


Figure 5. Root-mean-square fluctuation, RMSF, relative to the average structure is shown for each residue, in number order, of the WT sequence. Beneath this is the per residue difference in RMSF, ΔRMSF , exhibited by each of the sequences. (a) Values for all sequences studied as part of the virtual patient experiment for which position 71 is occupied by either alanine or valine. (b) Those with alanine or isoleucine. A positive ΔRMSF value indicates that the fluctuations observed at that position in the mutant system are greater than those at the corresponding location in WT.

exception of the A71I single mutant, the changes in RMSF relative to that seen in WT observed for all systems containing isoleucine at position 71 follow similar patterns to their valine-containing counterparts. Only minor deviations from wildtype flexibility are observed for any of the systems except for VPE-A71I, where substantial increases in flexibility are seen in the fulcrum, flaps, ears, and wall turn. The gain in flexibility in these regions is shared with the VPE-A71V system, but the magnitude of the change is reduced.

Active Site Water Entry. In the benchmark study,²⁹ the loss of bonding with residues 27 and other changes in the hydrogen-bonding network (involving residues 29 and 30), which are similar to those observed in the patient derived sequences studies here, was correlated with both resistance and the entry of water molecules into the catalytic site. The population density of water molecules within 3 \AA of the catalytic dyad (ρ) for each sequence under consideration was calculated using VMD.³² The values for

each system are shown in Table 5. Only the virtual patient sequences, VPE-A71I and VPE-A71V, show substantial water occupation in this area, with population densities of 0.324 and 0.253, respectively, compared to 0.057 for WT. This is between the levels observed for the known resistant AS and HM mutants. The A71V single mutant shows the opposite change in active site accessibility with water ingress exhibited by less than one percent of snapshots (a fifth of that seen for the WT sequence). This change is not replicated in the A71I single mutant.

Principal Component Analysis. Principal component analysis (PCA) is a dimensional reduction technique that allows the isolation of the most significant conformational differences between a set of structures. Here the structures are provided by snapshots from the molecular dynamics trajectory. The correlation matrix is calculated from an aligned molecular dynamics trajectory and then diagonalized. This provides an orthogonal set of eigenvectors representing linearly independent modes of conformational change

Table 5. Population Density, ρ , of Water within 3 Å of the Catalytic Dyad Exhibited by Each of the Sequences Studied As Part of the Virtual Patient Experiment

sequence	ρ
WT	0.057
L10I	0.042
A71I	0.025
A71V	0.009
L10I-A71I	0.037
L10I-A71V	0.053
A71V-L90M	0.084
L90M	0.035
L10I-L90M	0.039
A71I-L90M	0.058
L10I-A71V-L90M	0.031
L10I-A71I-L90M	0.064
VPE-A71I	0.253
VPE-A71V	0.324
AS	0.167
HM	0.484

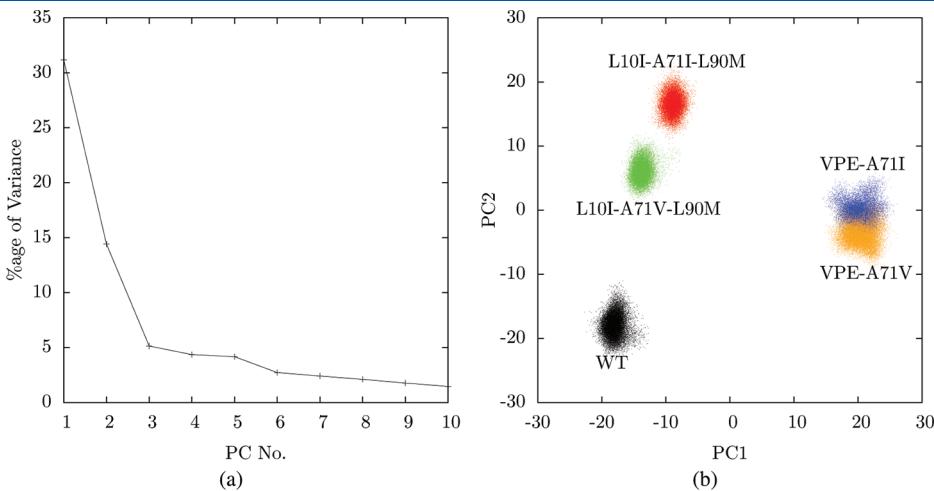


Figure 6. (a) Percentage of the variation observed over the concatenated trajectory of WT, L10I-A71I-L90M, L10I-A71V-L90M, VPE-A71I, and VPE-A71V which is captured by each principal component. The first two PCs account for approximately 46% of all variance observed. (b) Projections for each snapshot of the concatenated trajectory along PCs 1 and 2. All five systems can easily be distinguished using PCs 1 and 2.

called principal components. The eigenvalues associated with each principal component are a measure of the variance in the original data set described by that component. The principal component analysis presented here was performed on the backbone coordinates of a concatenated trajectory of all mutant sequences under investigation along with the WT and HM benchmark sequences. All structures used for MMPBSA calculations were included in the PCA. The trajectory concatenation and side chain atom stripping were performed using VMD,³² and the PCA was conducted using the PTTRAJ module of AMBER9.²⁷ The key questions we wish to answer are (i) which structural changes are associated with resistance and (ii) which are associated with deformation caused by accommodation of mutated amino acids and the general flexibility of the protease. In order to focus on the resistance associated changes, a combined trajectory of the backbone of the WT, L10I-A71I-L90M, L10I-A71V-L90M, VPE-A71I, and VPE-A71V systems was created and analyzed to produce principal components (PCs) which capture the most significant variance between snapshots. The PCs can then be used to identify changes that distinguish the systems which are susceptible and resistant to LPV according to the binding affinity calculations presented above.

Figure 6a shows the level of variation captured by each of the first ten PCs. Only the first three describe greater than 5% of the observed differences in structure between snapshots. The first two components account for 46% of the total variation, and the focus of this section will primarily be on what they can tell us about the differences between systems (the variation described by PC3 does not allow the differentiation of the different sequences but is, nonetheless, detailed in the Supporting Information). In the context of protein conformational changes, the assumption of PCA that the observed data set (in this case the coordinates of each protein atom) is best expressed as a linear combination of certain basis vectors is significant. It is likely that the real conformational changes are not optimally described using such an assumption, and hence, PCA is better employed as an investigative tool to identify parts of the structure and general trends that can be explored further. Here the results of PCA are used to identify metrics which help describe conformational changes associated with resistance. Once such measurements have been identified they are applied to the larger data set of all the sequences studied within the VPE to ensure that the

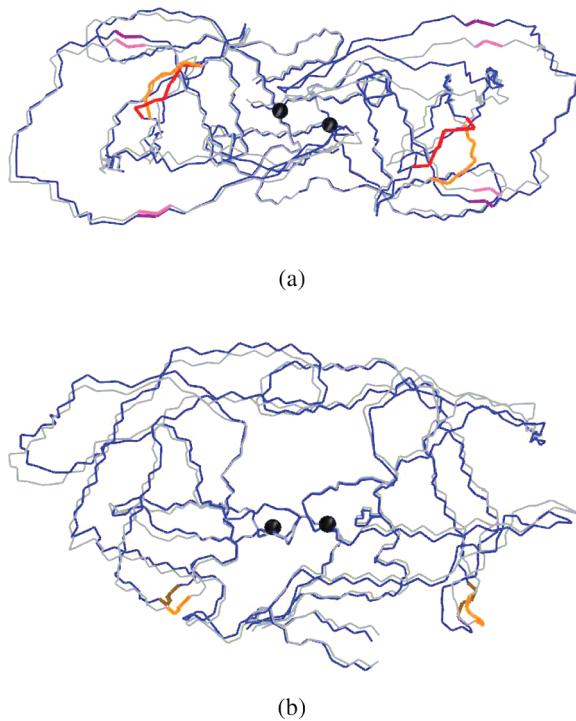


Figure 7. Structural variation of the protease described by PC1. The most negative projection observed is shown in blue, and the most positive, in gray. (a) View down from the PR flaps and (b) along the active site cavity. Residues which undergo significant changes along PC1 are highlighted, with darker shades used for the positions at the negative extrema, lighter for the positive. In part a red highlights the 77–79 and 177–179 loops and purple highlights residues 35, 45, 135, and 145, and in part b residues 68–69 and 168–169 are depicted in brown. In both views, the position of the catalytic dyad (residues 25 and 125) is indicated by black beads.

differences can still be used to distinguish the resistant systems. If this is the case, we can have confidence that we have identified structural changes that are linked to the changes in binding affinity.

The changes described by PC1 across the combined trajectory are represented in Figure 7. The structure of the most negative projection from the trajectory is seen in blue, that of the most positive in gray (this convention will be used in all other structural figures from PCA). The most pronounced change occurs as a global expansion of the structure along the axis of the active site cavity. This change is particularly visible in the change in separation of residues 35/135 (in the ears) and 45/145 (in the flaps) highlighted in purple in Figure 7a. The region around residues 79/179, between the cheek sheet and wall turn, however, is seen to change conformation and move toward the active site (this change is highlighted in red in Figure 7a). The separation between residues 35 and 45, 135 and 145, 25 and 79, and 125 and 179 were chosen as metrics to investigate in the full selection of sequences. Figure 6b shows that the projections corresponding to the WT and two triple mutant systems have negative values while the VPE sequences have positive values. This gives us an expectation of structural expansion, including increased separation between residues 35/135 and 45/145 in the VPE structures and a reduced distance between 79/179 and the catalytic dyad relative to the WT. Figure 8 shows the averages of the distance between the C_{α} atoms of residues 35 and 45 and residues 135 and 145. In all cases except the

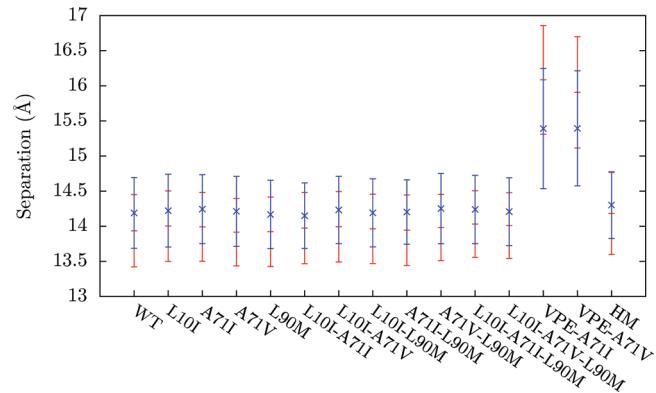


Figure 8. Separation of residues 35 and 45 and residues 135 and 145, for all systems under study, shown in red and blue, respectively. Distances are measured between the C_{α} atoms of both residues. The error bars indicate the standard deviation.

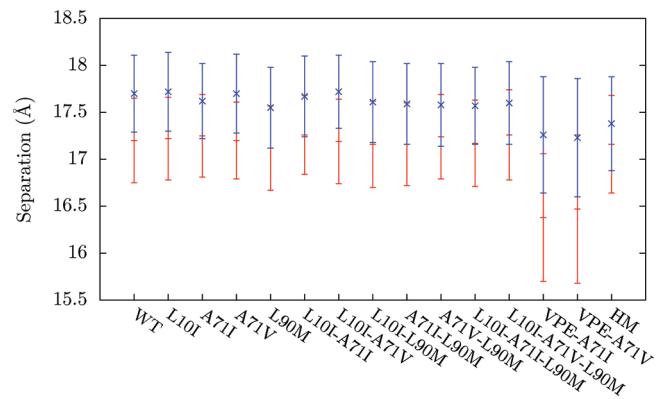


Figure 9. Separation of residues 25 and 79, and residues 125 and 179, for all systems under study, shown in red and blue respectively. Distances are measured between the C_{α} atoms of both residues. The error bars indicate the standard deviation.

two patient-derived sequences, the average distance between the two pairs of residues were around 14 and 14.2 Å (the averages are slightly higher for HM but the standard deviations on the measurements make it hard to evaluate the significance of this change). The VPE systems, however, have separations of 16 and 15.4 Å for the first and second monomers, respectively. This indicates that both systems do indeed enlarge in this dimension and also that the change is asymmetric with the change in the first monomer being around 2.0 Å and only 1.2 Å in the second. The predicted movement of residues 79 and 179 toward the active site ASPs in the VPE sequences is shown in Figure 9; the changes are relatively small, however, with a difference of only 0.8 and 0.4 Å for the first and second monomer, respectively. The second monomer of HM undergoes a similar change to that seen in the patient sequence-based systems, but all other measurements in the figure exhibit no change from those of WT. The fact that neither the enlargement of the structure nor the alteration of the position of residues 79 and 179 seen in the VPE systems is replicated in HM indicates that a different mechanism is causing the resistance in the two cases.

Figure 10 shows the structural variations described by PC2. The ears, fulcrum, and cheek sheet of the second monomer move as a unit. These sections of the protease have previously been observed to move as a rigid unit and are believed to facilitate flap

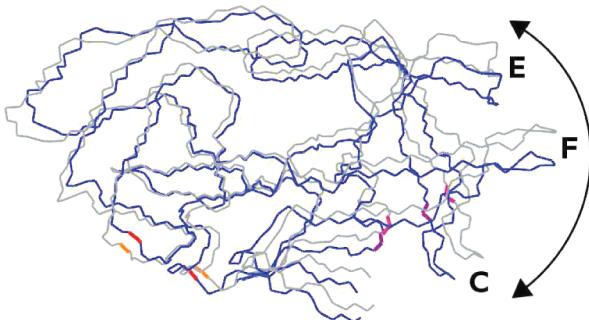


Figure 10. The structural variation of the protease described by PC2. The most negative projection observed is shown in blue, and the most positive in gray. The (E)ars, (F)ulcrum, and (C)heek sheet of the second monomer move as a unit. Red and purple are used to highlight residues 71 and 93, and 171 and 193, respectively, which undergo significant shifts along PC2 (darker shades are used for the positions at the negative extrema, and lighter ones, for the positive).

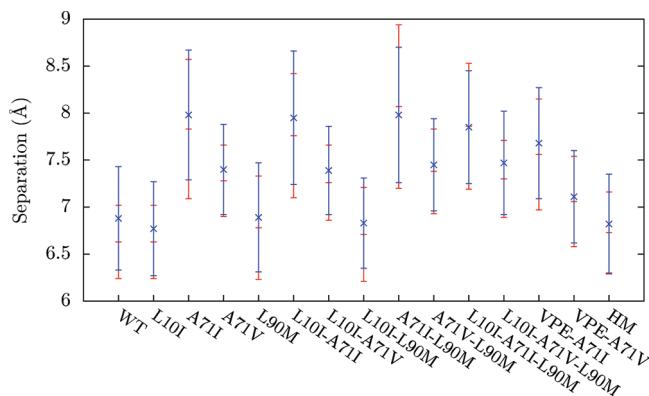


Figure 11. Separation of residues 71 and 93, and residues 171 and 193, for all systems under study, shown in red and blue, respectively. Distances are measured between the C_α atoms of both residues. The error bars indicate the standard deviation.

opening and impact upon substrate specificity.^{46,47} Another potentially significant shift occurs in the separation between the residues around 71 (in the cheek sheet) and those around 93 (in the loop between the helix and whiskers), and those around 171 and 193. This rearrangement is of particular interest as it involves the area surrounding the mutated residue in position 71. The separation between 71 and 93 increases along with the value of the projection. This means that according to Figure 6b the two triple mutants (L10I-A71I-L90M and L10I-A71V-L90M) should have the largest value, followed by the patient sequences and finally the WT. In both cases when the sequences vary only by a different residue at position 71, the system containing isoleucine has a higher projection value than that containing valine. This observation is in line with these changes being caused by the structure accommodating the mutation at position 71 as isoleucine is bulkier than valine. Measurements of the separation of residues 71 and 93, and residues 171 and 193, are shown in Figure 11 for all systems under study. These data show the expected pattern of differing separation between 71 and 93 depending on the size of the residue at position 71 (the wildtype contains alanine which is smaller than either of the mutant residues). In both cases, the VPE sequences have lower separation than the other mutants containing the same residue at position 71 (in the first

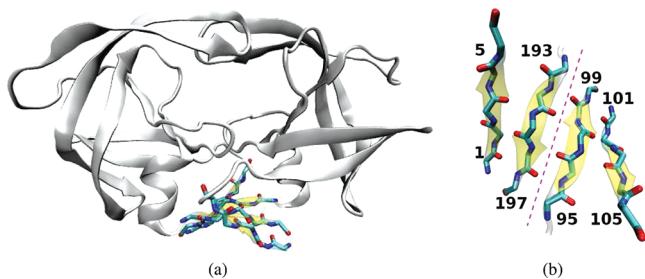


Figure 12. Dimer β sheet location and conformation within the HIV-1 protease. (a) Dimer β sheet in the context of the overall structure of the HIV-1 protease. (b) Strands of the sheet with the first and last residues of each labeled. The sections on either side bend about the dashed line.

Table 6. Average of the Angle, θ , between the Planes Formed by Strands of the Dimer β Sheet on Either Side of the Dashed Line Indicated in Figure 12b^a

sequence	$\langle \theta \rangle$	$\Delta \langle \theta \rangle$
WT	146.83 (7.04)	
L10I	147.69 (6.99)	0.85
A71I	145.77 (7.53)	-1.06
A71V	147.28 (6.87)	0.44
L90M	146.11 (7.66)	-0.72
L10I-A71I	146.09 (6.96)	-0.74
L10I-A71V	147.04 (7.28)	0.20
L10I-L90M	146.87 (7.87)	0.03
A71I-L90M	145.28 (7.84)	-1.55
A71V-L90M	147.15 (7.67)	0.31
L10I-A71I-L90M	145.32 (8.16)	-1.51
L10I-A71V-L90M	147.15 (8.47)	0.32
VPE-A71I	149.19 (7.75)	2.36
VPE-A71V	150.18 (7.55)	3.35
HM	146.38 (9.63)	-0.45

^aThe standard deviation is given in brackets, and the difference between each system and the WT value is also shown.

monomer the differences are 0.3 Å for both sequences; in the second, they are 0.2 and 0.4 Å for VPE-A71I and VPE-A71V, respectively).

Dimer Interface Conformation. On the basis of crystal structure evidence, Skalova et al.⁴⁸ suggested that one effect of the A71V mutation is the alteration of the conformation of the dimer interface which is then communicated to the ears, fulcrum, and cheek sheet regions. These regions exhibit conformational changes correlated with deformations around residues 71 and 171 in PC2 of the principal component analysis presented here. Along with the evidence of changes induced in static structures, NMR evidence suggests that the four-stranded dimer β sheet is divided into two sections allowing bending about the center line (shown in Figure 12).⁴⁹ Alterations in the relative orientation of these sheets are hard to detect in the PCA analysis, so direct analysis was applied to see if such changes are apparent in the simulations presented here.

The cross product of the vectors running between the first and last residue of each strand, as identified in Figure 12b, was used to define the normal of a plane representing each half of the β sheet. The cross product of the two normals was then used to calculate

the angle, θ , between the two planes. The average angle of each system under investigation is shown in Table 6. The two VPE mutants increase θ and exhibit the largest deviation from the angle observed in WT. The change produced in VPE-A71V is 1° larger than that in the VPE-A71I sequence; however, the behavior in the latter system is particularly notable as all other A71I containing systems decrease θ . The change in the angle is not reproduced in the resistant HM system, which reinforces the conclusions of our PCA which suggested that the mechanism of resistance encountered in the VPE sequences is of a different character to that of systems where direct active site mutations are involved.

The difference in conformation observed in the VPE structures is suggestive of an effect being mediated by the mutation of the nearby residue 93 (from isoleucine to leucine). Mutations at position 93 are strongly associated with substrate recognition.⁵⁰ If this conjecture is correct, then a plausible hypothesis to explain the relative rarity of the A71I at position 71 is that, without other mutations in the sequence, it distorts the dimer interface in a way that diminishes the ability of the enzyme to discriminate natural substrates, reducing viral fitness. The VPE sequences must be viable as they are derived from patient data; this would suggest that the positive changes in θ which they exhibit do not hinder recognition of natural substrate. The structural changes induced in the A71V containing variants also result in positive changes in θ , in contrast to the negative changes observed in the A71I containing sequences (except VPE-A71I). It is plausible that this difference (allied to the smaller distortion of the region measured by the separation between residues 71 and 93, and 171 and 193 presented in Figure 11) explains the lower fitness penalty apparently associated with the introduction of A71V compared to A71I, in the absence of compensatory mutations. Some credence is lent to this idea by in vitro experiments that show HIV-1 sequences containing protease with A71V have increased replicative capacity.⁵¹

■ DISCUSSION

We have used an ensemble molecular dynamics approach to investigate the impact of a set of mutations identified as producing ambiguous results from existing clinical decision support systems as part of the ViroLab virtual patient experiment (VPE). The list of mutations (L10I, A71IV, and L90M) were simulated in the context of the HXB2 wildtype and in all possible combinations along with the full patient sequence (with both isoleucine and valine in position 71). Our results suggest that other than the two full patient sequences (VPE-A71I and VPE-A71V), all mutants investigated should be assessed as susceptible to lopinavir because have a binding free energy at least as attractive the HXB2 wild type within the resolution of our calculations. [This is now the consensus result of the systems with their current rule sets, updated since the instigation of this study. The rules of all CDSS referred to in this study are constantly revised as new patient data and literature becomes available. The results of this study were not considered by any of the CDSS.] The full patient sequences, however, would be ranked as having intermediate levels of resistance. Using the absolute binding energy, ΔG_{theor} , VPE-A71I and VPE-A71V have changes in binding affinity comparative to WT of 1.98 and 2.07 kcal mol⁻¹, respectively. These values are close to that shown by the known resistance mutant AS, of 2.7 kcal mol⁻¹, in the MDR test set used to validate the ability of our simulation and free energy calculation protocol to reproduce in vitro experimental results. Excluding the configurational entropy component of the

calculation (which is known to have less reliable convergence properties compared to the MMPBSA part of the free energy calculation), both systems remain resistant, but the VPE-A71I system has a $\Delta\Delta G_{\text{MMPBSA}}$ value of only 0.98 kcal mol⁻¹ while the VPE-A71V value remains high at 2.07 kcal mol⁻¹. This ranking provides evidence that in many circumstances the *entire* sequence may need to be considered in order to gain an accurate assessment of the drug resistance of the virus infecting a particular patient. In the case of the specific sequence considered here (containing L10I, I13V, K14KR, I15V, K20T, L63P, A71IV, V77IV, L90M, and I93L relative to HXB2 wildtype), the binding affinity calculations presented here are not the only evidence available as a correlational study of resistance to patient genotype has identified a pattern of mutations at positions 10, 63, 71, 90, and 93 as being associated with protease inhibitor resistance (albeit including few patients treated with LPV involved in the cohort).⁵²

In addition to the binding affinity results, molecular simulation also allows us to gain mechanistic insight into patterns of resistance. Here we have shown that the full patient sequences both containing A71I and A71V adopt substantially different conformations to those of other resistant mutants such as the hexamutant (HM) in the MDR test set. In the VPE mutants, the overall enzyme conformation is expanded compared to WT (and all other mutants investigated in this study) with the distances between residues 35 and 45 on both monomers increased by 2.0 Å for the first and 1.2 Å for the second. Contrary to this overall movement, residue 79 on both monomers bends in toward the active site. The structural deformation, caused by replacing alanine with the bulkier valine or isoleucine, observed in the patient derived sequences is much lower than in the other structures simulated. Some similarities in the mode of resistance compared to the multiple drug resistance active site (AS) and hexamutant (HM) variants do exist. For example, the active site hydrogen bond network is perturbed and water molecules are observed to occupy the catalytic cavity more frequently than in wild type for all resistant protease mutants we have simulated.

Further structural changes can be observed in the dimer β sheet, the four strands of which form two subsheets which can bend relative to one another. The angle between the two pairs of strands is substantially increased in the resistant VPE mutants compared to WT. This is notable in the case of VPE-A71I as all other variants studied containing the A71I substitution have a decreased angle. It is possible that such deformations may impact the enzymatic fitness of the protease sequence and hence at least partially explain the rarity of the A71I mutation compared to A71V which is known to increase the replicative capacity of HIV-1 in vitro.⁵¹

Using the BAC tool¹⁴ to automate the system setup, data transfer, simulation, and analysis allowed us to easily run a very large number of molecular dynamics simulations across a variety of resources on both the US TeraGrid (Ranger and Kraken) and the EU DEISA network (HECToR, SARA, and LRZ). To simplify the analysis of the theoretical minimum turn around time of BAC orchestrated free energy calculations consider the case when only the Ranger machine (with 62 976 cores) at the Texas Advanced Computing Center (TACC) is used for the production simulations. The optimally scaled rate of computation was approximately 4 h ns⁻¹ on 64 Opteron cores replica⁻¹. The theoretical minimum turn around time for the 13 sequences investigated for this study (using 50 replicas each producing 4 ns trajectories) using 41 600 cores (simulating 650 replicas simultaneously) could be as short as 16 h. In practice, peak performance was around 300 ns day⁻¹ and

generally far lower. MMPBSA and normal-mode postprocessing took 3 and 20 h ns⁻¹, respectively but were run in parallel for each nanosecond using the Leeds node (256 cores) of the U.K. National Grid Service and our local Mavrino cluster (96 cores). The theoretical minimum total turn around time (simulation + free energy calculation) using this approach was thus approximately a week. This would allow such simulations to be provided on a clinically relevant time scale assuming suitable levels of resources were available. Many challenges remain before free energies from molecular dynamics simulations can be routinely used as part of CDSS, but we have demonstrated the potential of such an approach to accurately rank drug binding affinities on clinically relevant time scales (of 2–3 days).

■ ASSOCIATED CONTENT

Supporting Information. Details of the protocol used to perform the simulations presented here and description of conformational changes associated with PC3 from the principal component analysis. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: p.v.coveney@ucl.ac.uk.

■ ACKNOWLEDGMENT

This work has been supported by the EU FP6 ViroLab (IST-027446) and the EU FP7 CHAIN (HEALTH-2007-2.3.2-7) projects. Access to US TeraGrid was made available by the National Science Foundation under NRAC grant MCA04N014 and by the DEISA Consortium (cofunded by the EU, FP7 project 222919) to supercomputers in Europe via the Virtual Physiological Human Virtual Community managed by the VPH Network of Excellence (IST-223920; www.vph-noe.eu). We are grateful to EPSRC and CoMPLEX for funding the Ph.D. studentship of D.W.W. We also wish to acknowledge the UK NGS for providing access to their resources and their support for our work.

■ REFERENCES

- (1) Kuehn, B. M. UNAIDS report: AIDS epidemic slowing, but huge challenges remain. *J. Am. Med. Assoc.* **2006**, *296*, 29–30.
- (2) Haub, C. World Population Highlights: Key Findings From PRB's 2007 World Population Data Sheet. *Population Bull.* **2007**, *62*, 1–12.
- (3) UNAIDS, *AIDS Epidemic Update 2009*, 2009.
- (4) Mathers, C. D.; Loncar, D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med.* **2006**, *3*, e442.
- (5) Fang, C. T.; Chang, Y. Y.; Hsu, H. M.; Twu, S. J.; Chen, K. T.; Lin, C. C.; Huang, L. Y. L.; Chen, M. Y.; Hwang, J. S.; Wang, J. D.; Chuang, C. Y. Life expectancy of patients with newly-diagnosed HIV infection in the era of highly active antiretroviral therapy. *QJM* **2007**, *100*, 97–105.
- (6) Kitahata, M. M.; Rompaey, S. E. V.; Shields, A. W. Physician experience in the care of HIV-infected persons is associated with earlier adoption of new antiretroviral therapy. *J. Acquir. Immune. Defic. Syndr.* **2000**, *24*, 106–114.
- (7) Delgado, J.; Heath, K. V.; Yip, B.; Marion, S.; Alfonso, V.; Montaner, J. S. G.; O'Shaughnessy, M. V.; Hogg, R. S. Highly active antiretroviral therapy: physician experience and enhanced adherence to prescription refill. *Antivir. Ther.* **2003**, *8*, 471–478.
- (8) Haq, O.; Levy, R. M.; Morozov, A. V.; Andrec, M. Pairwise and higher-order correlations among drug-resistance mutations in HIV-1 subtype B protease. *BMC Bioinformatics* **2009**, *10* (Suppl 8), S10.
- (9) Zhang, J.; Hou, T.; Wang, W.; Liu, J. S. Detecting and understanding combinatorial mutation patterns responsible for HIV drug resistance. *Proc. Natl. Acad. Sci. USA* **2010**, *107*, 1321–1326.
- (10) Lengauer, T.; Sing, T. Bioinformatics-assisted anti-HIV therapy. *Nat. Rev. Microbiol.* **2006**, *4*, 790–797.
- (11) Bubak, M. et al. *Virtual Laboratory for Development and Execution of Biomedical Collaborative Applications*, 2008.
- (12) Assel, M.; van de Vijver, D.; Libin, P.; Theys, K.; Nuallain, D. H. B. O.; Nowakowski, P.; Bubak, M.; Vandamme, A.; Imbrechts, S.; Sangeda, R.; Jiang, T.; Frentz, D.; Sloot, P. A collaborative environment allowing clinical investigations on integrated biomedical databases. *Stud. Health Technol. Inform.* **2009**, *147*, S1–S1.
- (13) Sloot, P. M. A.; Coveney, P. V.; Ertaylan, G.; Müller, V.; Boucher, C. A.; Bubak, M. HIV decision support: from molecule to man. *Philos. Trans. R. Soc. A* **2009**, *367*, 2691–2703.
- (14) Sadiq, S. K.; Wright, D.; Watson, S. J.; Zasada, S. J.; Stoica, I.; Coveney, P. Automated Molecular Simulation Based Binding Affinity Calculator for Ligand-Bound HIV-1 Proteases. *J. Chem. Inf. Model.* **2008**, *48*, 1909–1919.
- (15) Zasada, S.; Coveney, P. Virtualizing access to scientific applications with the Application Hosting Environment. *Comput. Phys. Commun.* **2009**, *180*, 2513–2525.
- (16) Rhee, S.-Y.; Gonzales, M. J.; Kantor, R.; Betts, B. J.; Ravela, J.; Shafer, R. W. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* **2003**, *31*, 298–303.
- (17) Garriga, C.; Pérez-Elías, M. J.; Delgado, R.; Ruiz, L.; Nájera, R.; Pumarola, T.; del Mar Alonso-Socas, M.; García-Bujalance, S.; Menéndez-Arias, L. Mutational patterns and correlated amino acid substitutions in the HIV-1 protease after virological failure to nelfinavir- and lopinavir/ritonavir-based treatments. *J. Med. Virol.* **2007**, *79*, 1617–1628.
- (18) Gallego, O.; Martín-Carbonero, L.; Aguero, J.; de Mendoza, C.; Corral, A.; Soriano, V. Correlation between rules-based interpretation and virtual phenotype interpretation of HIV-1 genotypes for predicting drug resistance in HIV-infected individuals. *J. Virol. Methods* **2004**, *121*, 115–118.
- (19) Gilson, M. K.; Zhou, H. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.
- (20) Steinbrecher, T.; Labahn, A. Towards Accurate Free Energy Calculations in Ligand Protein-Binding Studies. *Curr. Med. Chem.* **2010**, *17*, 767–785.
- (21) Massova, I.; Kollman, P. Computational Alanine Scanning To Probe Protein-Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *J. Am. Chem. Soc.* **1999**, *121*, 8133–8143.
- (22) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889–897.
- (23) Mobley, D. L.; Dill, K. A. Binding of small-molecule ligands to proteins: “what you see” is not always “what you get”. *Structure* **2009**, *17*, 489–498.
- (24) Wittayanarakul, K.; Hannongbua, S.; Feig, M. Accurate prediction of protonation state as a prerequisite for reliable MM-PB(GB)SA binding free energy calculations of HIV-1 protease inhibitors. *J. Comput. Chem.* **2008**, *29*, 673–685.
- (25) Stoica, I.; Sadiq, S.; Coveney, P. Rapid and Accurate Prediction of Binding Free Energies for Saquinavir-Bound HIV-1 Proteases. *J. Am. Chem. Soc.* **2008**, *130*, 2639–2648.
- (26) Brooks, B. R.; Janezic, D.; Karplus, M. Harmonic analysis of large systems. I. Methodology. *J. Comput. Chem.* **1995**, *16*, 1522–1542.
- (27) Case, D. A.; Cheatham, T. E.; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M.; Onufriev, A.; Simmerling, C.; Wang, B.; Woods, R. J. The Amber biomolecular simulation programs. *J. Comput. Chem.* **2005**, *26*, 1668–1688.
- (28) Sanner, M. F.; Olson, A. J.; Spehner, J. C. Reduced surface: an efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38*, 305–320.
- (29) Sadiq, S. K.; Wright, D. W.; Kenway, O. A.; Coveney, P. V. Accurate ensemble molecular dynamics binding free energy ranking of

- multidrug-resistant HIV-1 proteases. *J. Chem. Inf. Model.* **2010**, *50*, 890–905.
- (30) Genheden, S.; Ryde, U. How to obtain statistically converged MM/GBSA results. *J. Comput. Chem.* **2010**, *31*, 837–846.
- (31) Stoll, V.; Qin, W.; Stewart, K. D.; Jakob, C.; Park, C.; Walter, K.; Simmer, R. L.; Helfrich, R.; Bussiere, D.; Kao, J.; Kempf, D.; Sham, H. L.; Norbeck, D. W. X-ray crystallographic structure of ABT-378 (lopinavir) bound to HIV-1 protease. *Bioorg. Med. Chem.* **2002**, *10*, 2803–2806.
- (32) Humphrey, W.; Dalke, A.; Schulten, K. VMD - Visual Molecular Dynamics. *J. Mol. Graphics* **1996**, *14*, 33–38.
- (33) Ohtaka, H.; Schön, A.; Freire, E. Multidrug resistance to HIV-1 protease inhibition requires cooperative coupling between distal mutations. *Biochemistry* **2003**, *42*, 13659–13666.
- (34) Zoete, V.; Michielin, O.; Karplus, M. Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for the analysis of protein flexibility. *J. Mol. Biol.* **2002**, *315*, 21–52.
- (35) Schüttelkopf, A. W.; van Aalten, D. M. F. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr. Sect. D* **2004**, *60*, 1355–1363.
- (36) Frisch, M. J. et al. *Gaussian 03*, revision C.02. 2004; Gaussian, Inc., Wallingford, CT, 2004.
- (37) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (38) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (39) Wlodawer, A.; Vondrasek, J. Inhibitors of HIV-1 protease: a major success of structure-assisted drug design. *Annu. Rev. Biophys. Biomol. Struct.* **1998**, *27*, 249–284.
- (40) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (41) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (42) Perryman, A. L.; Lin, J.; McCammon, J. A. HIV-1 protease molecular dynamics of a wildtype and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci.* **2004**, *13*, 1108–1123.
- (43) Gonzalez, L. M. F.; Santos, A. F.; Abecasis, A. B.; Van Laethem, K.; Soares, E. A.; Deforche, K.; Tanuri, A.; Camacho, R.; Vandamme, A.; Soares, M. A. Impact of HIV-1 protease mutations A71V/T and T74S on M89I/V-mediated protease inhibitor resistance in subtype G isolates. *J. Antimicrob. Chemother.* **2008**, *61*, 1201–1204.
- (44) Martinez-Picado, J.; Wrin, T.; Frost, S. D. W.; Clotet, B.; Ruiz, L.; Brown, A. J.; Petropoulos, C. J.; Parkin, N. T. Phenotypic hypersusceptibility to multiple protease inhibitors and low replicative capacity in patients who are chronically infected with human immunodeficiency virus type 1. *J. Virol.* **2005**, *79*, 5907–5913.
- (45) Yanchunas, J.; Langley, D. R.; Tao, L.; Rose, R. E.; Friberg, J.; Colonna, R. J.; Doyle, M. L. Molecular basis for increased susceptibility of isolates with atazanavir resistance-conferring substitution I50L to other protease inhibitors. *Antimicrob. Agents Chemother.* **2005**, *49*, 3825–3832.
- (46) Tozzini, V.; Trylska, J.; Chang, C. A.; McCammon, J. A. Flap opening dynamics in HIV-1 protease explored with a coarse-grained model. *J. Struct. Biol.* **2007**, *157*, 606–615.
- (47) Trylska, J.; Tozzini, V.; Chang, C. A.; McCammon, J. A. HIV-1 protease substrate binding and product release pathways explored with coarse-grained molecular dynamics. *Biophys. J.* **2007**, *92*, 4179–4187.
- (48) Skalova, T.; Dohnalek, J.; Duskova, J.; Petrokova, H.; Hradilek, M.; Soucek, M.; Konvalinka, J.; Hasek, J. HIV-1 protease mutations and inhibitor modifications monitored on a series of complexes. Structural basis for the effect of the A71V mutation on the active site. *J. Med. Chem.* **2006**, *49*, 5777–5784.
- (49) Ishima, R.; Freedberg, D. I.; Wang, Y. X.; Louis, J. M.; Torchia, D. A. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Structure* **1999**, *7*, 1047–1055.
- (50) Kontijevskis, A.; Prusis, P.; Petrovska, R.; Yahorava, S.; Mutulis, F.; Mutule, I.; Komorowski, J.; Wikberg, J. E. S. A look inside HIV resistance through retroviral protease interaction maps. *PLoS Comput. Biol.* **2007**, *3*, e48.
- (51) Nijhuis, M.; Schuurman, R.; de Jong, D.; Erickson, J.; Gustchina, E.; Albert, J.; Schipper, P.; Gulnik, S.; Boucher, C. A. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS* **1999**, *13*, 2349–2359.
- (52) Wu, T. D.; Schiffer, C. A.; Gonzales, M. J.; Taylor, J.; Kantor, R.; Chou, S.; Israelski, D.; Zolopa, A. R.; Fessel, W. J.; Shafer, R. W. Mutation patterns and structural correlates in human immunodeficiency virus type 1 protease following different protease inhibitor treatments. *J. Virol.* **2003**, *77*, 4836–4847.