

SCIENTIFIC REPORTS



OPEN

Characterizing early drug resistance-related events using geometric ensembles from HIV protease dynamics

Olivier Sheik Amamuddy¹, Nigel T. Bishop² & Özlem Tastan Bishop¹ 

The use of antiretrovirals (ARVs) has drastically improved the life quality and expectancy of HIV patients since their introduction in health care. Several millions are still afflicted worldwide by HIV and ARV resistance is a constant concern for both healthcare practitioners and patients, as while treatment options are finite, the virus constantly adapts via complex mutation patterns to select for resistant strains under the pressure of drug treatment. The HIV protease is a crucial enzyme for viral maturation and has been a game changing drug target since the first application. Due to similarities in protease inhibitor designs, drug cross-resistance is not uncommon across ARVs of the same class. It is known that resistance against protease inhibitors is associated with a wider active site, but results from our large scale molecular dynamics simulations combined with statistical tests and network analysis further show, for the first time, that there are regions of local expansions and compactions associated with high levels of resistance conserved across eight different protease inhibitors visible in their complexed form within closed receptor conformations. The observed conserved expansion sites may provide an alternative drug-targeting site. Further, the method developed here is novel, supplementary to methods of variation analysis at sequence level, and should be applicable in analysing the structural consequences of mutations in other contexts using molecular ensembles.

Antiretroviral (ARV) drug resistance still persists despite recent improvements in antiretroviral therapy¹. As the viral genome continues to accumulate mutations under the selective pressures of therapy², surviving viral populations inevitably become less sensitive to one or more drugs over time. HIV reservoirs and its existence as a quasispecies³ means that an ARV should ideally inhibit a pool of slightly different conformations of receptor targets. The decreasing efficacy of drug binding over time means that patients may have to switch to more difficult treatment regimens with the possibility of experiencing more severe side-effects if no better-tolerated alternative exists. At the same time, ARVs are a finite resource which should be used with proper timing failing which resistance develops sooner. In order to design more robust ARVs and/or improve onto existing resistance prediction methods, additional knowledge of the motions associated with resistance may be helpful. However this is not straight-forward, as patterns of resistance mutations in HIV are complex⁴, and may require special consideration in order to extract deeply-engrained behaviour. In this manuscript, we focus on HIV protease, which is a crucial enzyme for viral maturation, and is a well-established HIV drug target^{5,6}. There are minor differences as to how various functional segments of HIV protease are defined in literature (possibly due to its high variability) for which we show some of the structural features in Fig. 1. Protease inhibitors (PIs) competitively inhibit the enzyme⁷, which under normal circumstances process the viral polyproteins Gag and Gag-Pol⁸. Multi-drug resistance within members of the PI class is not uncommon due to their long period of use and their three-dimensional and electrostatic similarities^{9,10}.

In resource-available settings drug efficacy can be inferred by monitoring viral load or CD4 cell counts, yet when available, knowledge of genotypic information can improve the choice of therapy to be used¹¹. However, drug resistance mechanisms in HIV are not fully-understood¹². Previous research has evaluated various computational modelling approaches over the years in order to predict or understand drug resistance mechanisms in HIV

¹Research Unit in Bioinformatics (RUBi), Department of Biochemistry and Microbiology, Rhodes University, Grahamstown, 6140, South Africa. ²Department of Mathematics (Pure & Applied), Rhodes University, Grahamstown, 6140, South Africa. Correspondence and requests for materials should be addressed to Ö.T.B. (email: o.tastanbishop@ru.ac.za)

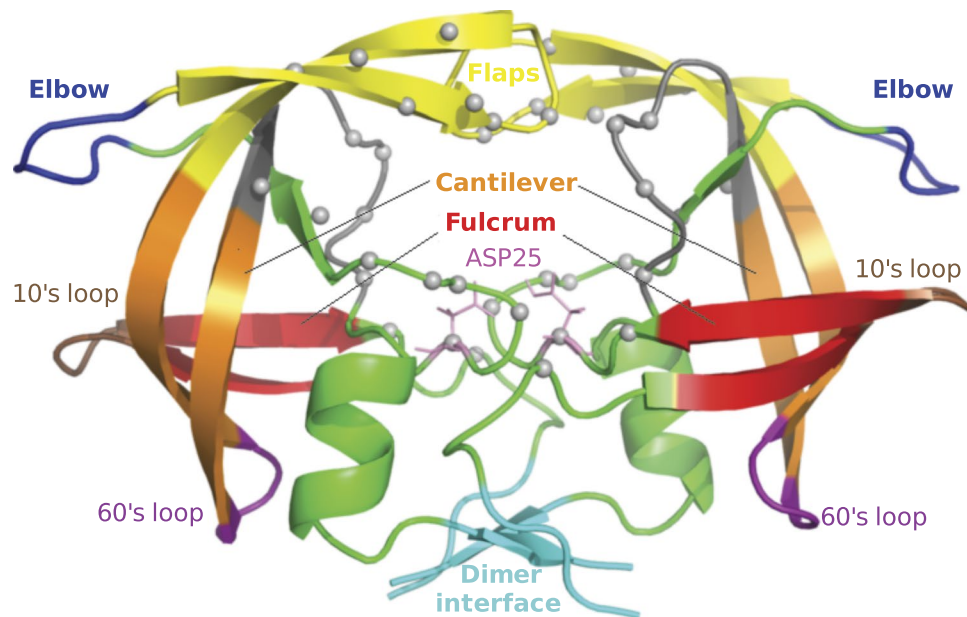


Figure 1. Functional regions within HIV protease. Grey spheres depict the residues constituting the binding cavity.

protease, ranging from the use of delaunay triangulations from static protease structures¹³, binding energies from molecular docking^{14,15} to elastic network modelling applied to coarse-grained models using a uniform spring stiffness¹⁶, 100 fs molecular dynamics (MD)^{15,17} and many more, as reviewed by Cao and co-workers¹⁸. Here we adopt a structural approach using MD applied to 100 highly-resistant and 100 hyper-susceptible HIV sequences against eight docked protease inhibitors using the labelled sequence data available from the Stanford HIVdb¹⁹. We focus on the majority subtype (B), with the sequences containing rare residues removed, as described in our previous work²⁰. 3D structures of the 200 protease sequences are built using homology modelling²¹ with a common drug-bound template for each target. Ligand docking with the eight ARVs, namely atazanavir (ATV), darunavir (DRV), fosamprenavir (FPV), indinavir (IDV), lopinavir (LPV), nelfinavir (NFV), saquinavir (SQV) and tipranavir (TPV) then gives 1,600 3D structures of drug-bound protease complexes. FPV is the amprenavir prodrug which is released in its active form upon hydrolysis²².

Proteins are in constant motion²³ and drug-binding alters their dynamics²⁴ so each case requires its own MD run. Allowing for replications, a total of 3,200 MD runs are performed. Each run is about 2 ns, so that in total the MD simulations amount to about 6,400 ns. Such a design was necessary in order to cover as much resistance-related complexity from as large a number of independent observations such that highly-conserved patterns would emerge within the noise from protein dynamics in order to minimize bias while maintaining biological variance, for a reasonable amount of CPU hours. The observation of conserved resistance-related dynamics across this high number of independent short simulations of PI-bound receptor complexes shows that highly drug-resistant sequences may be structurally-detectable in a short amount of simulated time. Considerable amounts of conformational sampling are typically required to observe motions that are of large amplitude^{25,26} or rare²⁷. Same applies for increasing the accuracy of binding free-energy estimations (for instance between a ligand and a receptor), which comes with increased computational costs^{28,29}. We circumvent these issues in this context, by describing two short and specific motions that are detectable very early in all-atom dynamics simulations of the retroviral protease, using the idea of preferential attachment applied to local residue motion. This concept stems from the tendency of initially highly-connected nodes to attract new connections, leading to a global behaviour referred to as being scale-free, in non-random graph topologies³⁰. We use this idea to reinforce the detection of significantly different (smaller and larger) pairwise residues distances derived from statistical tests of averages on the premise that a residue will most likely be at a given distance within an ensemble if there is larger number of other residue pairs supporting the observed difference.

The *in silico* methods used to simulate MD of drug-receptor complexes are partially stochastic^{31–33} - we therefore mitigate chance events by calculating statistical properties of each ensemble and applying the network degree centrality (connectivity) measure. Networks are an intuitive way of representing relational data using nodes and edges³⁴, with much of the underlying ideas having emerged from insights made by representing social networks as graphs³⁵. Network analysis has thus evolved into an ideal tool for examining inherently complex biological contexts such as single nucleotide variation analysis^{36,37}, protein-protein interactions^{38,39}, gene co-expression data^{34,40}, intra-protein networks^{37,41–44} and allosteric modulation analysis⁴⁵. A tool that uses network analysis over MD simulations is given in⁴⁴. Network graphs are composed of nodes and edges, where each node represents a particular object while an edge is drawn between any node pair to represent a shared property. Edges can either be directed, in which case a relationship does not entail reciprocation or conversely be undirected whereby connections are mutual. Additionally, edges can be weighted or binary. While the former preserves information continuity, the

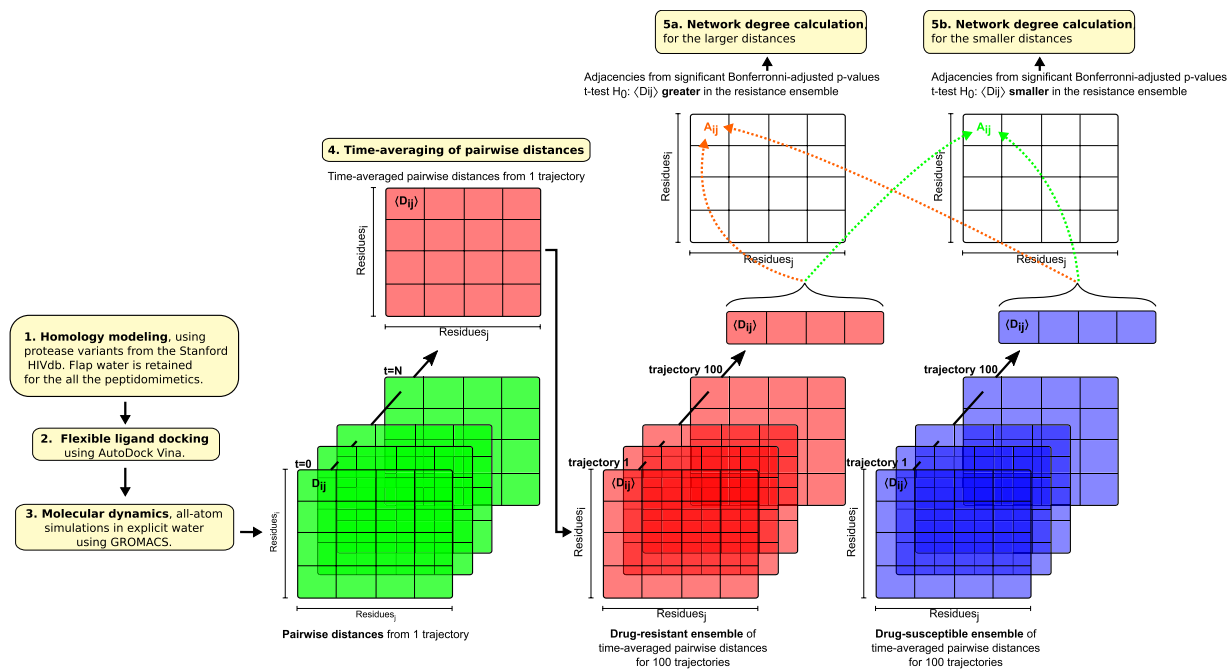


Figure 2. Experimental work-flow for the distance-based network construction and analysis from MD trajectories.

latter only denotes the presence or absence of a connection³⁴. Undirected binary edges are used in our case to independently represent a significantly larger or a smaller distance on separate graphs. In our analysis, we used the degree centrality, which is simply the number of neighbours adjacent to a given node. Adjacencies can be represented as a square matrix. The degree centrality then is the row (or column) sum from the matrix.

By calculating degree centrality within graphs generated from statistically-inferred edges between C_α atoms, we remarkably find structural features of HIV protease that differ between susceptible and resistant sequences, with further conservation occurring across all 8 ARVs. The results can form the basis for more robust ARV design and better prediction of drug resistance. Further, the combination of molecular dynamics, network centrality and statistical analyses used here provides an alternative way of analysing the effects of non-synonymous mutations, and should be applicable to other diseases whereby protein variants can be independently simulated as two ensembles of variants, ideally comprising over 30 samples in each case, to represent cases and controls to benefit from normalizing properties of the central limit theorem⁴⁶ for the t-test. Given our performance at predicting conserved motion using this method, such an approach could be very insightful if similar motions were to be picked up in non-B subtypes of varying residue composition as our method highlights the most definitively distinct dominant motions prevailing between drug resistance and susceptibility within the confines of their ensembles. Schematics illustrating the experimental workflows are shown in Fig. 2 for the distance-based approach and in (Supplementary Fig. S1) for the angle-based method.

Results and Discussion

In this study, two replicas of 1600 MD simulations were performed totalling 6400 ns, over 100 highly-resistant and 100 hyper-susceptible HIV protease structures complexed with eight docked protease inhibitors. As a quality control for all the MD simulations, C_α root mean square deviation (RMSD) values were first computed to exclude any error in periodic boundary corrections. A condensed representation of the mean and the standard deviations of RMSD values for each ARV is depicted in Supplementary Fig. S2. The runs were found to display slightly higher variation (in red) for the first 100 ps before stabilizing (yellow to white) thereafter in each case. We then begin the experiment with a global assessment of the distributions of protein compactness using the radius of gyration (R_g) across drug ensembles, as shown in Fig. 3 to more local evaluations, namely pairwise residue distances and C_α angles from receptors (Figs 4, 5, 6 and 7).

Global assessment via radius of gyration. Previous work described distinct mechanisms associated with ARV drug resistance that all point to active site expansion, namely (1) impaired hydrophobic sliding shown in the G48T/L89M double mutant with saquinavir⁴⁷, (2) reduced dimer stability in L24I, I50V and F53L mutants⁹ and (3) single or co-operative distal mutations^{48,49}. Additional research on multi-drug resistant HIV protease further described an expanded active site pocket^{50,51} and cavity expansion due to atomic volume for the HIV protease mutants V82A and I84V⁵. As R_g measures compaction by calculating the RMSD of each atomic centre-of-mass with respect to that of the whole molecule, we generally expected that higher values would be associated with resistance ensembles. However, our simulations showed no global tendency towards a less compact (larger R_g) series of conformations in the resistant ensemble compared to the drug-susceptible state, as shown in Fig. 3. Taken individually, only DRV, NFV and SQV display larger R_g values in their resistant ensemble. The reverse is

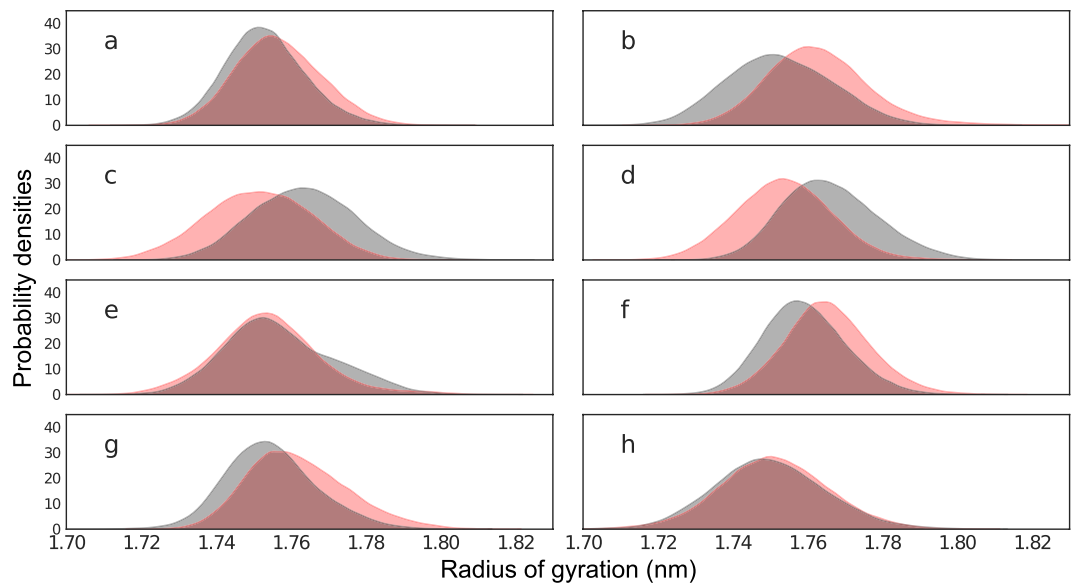


Figure 3. Distributions of R_g values for protease inhibitor complexes containing ATV (a), DRV (b), FPV (c), IDV (d), LPV (e), NFV (f), SQV (g) and TPV (h). Resistant ensembles are shaded in red while susceptible ensembles are in grey.

actually observed in FPV and IDV. No appreciable shifts were observed for ATV, LPV and even less for TPV. Of notable interest are the slight shifts in the means of the R_g distributions across all drug ensembles, which are suspected to either be inherited from the template used for modelling or the docked drugs themselves, which could be propagating a different set of local receptor-ligand signals towards various parts of the receptor. Similar trends were observed upon replication (Supplementary Fig. S3).

We hypothesize that compactations may instead be observable at a residue level and that these might be masked by more chaotic motions happening globally. Therefore we further investigated each protein-ligand complex by increasing experimental sensitivity by calculating time-averaged one-tailed t-tests from pairwise residue distances obtained from aggregations of independent MD simulations before summarizing and analysing the results with network analysis.

Local evaluation via statistical tests coupled with network analysis. Results from Bonferroni-moderated t-tests are transformed into a network graph as explained in the Materials and Methods section for the distances, and are represented as normalized degree centrality plots onto which several architectural features of HIV protease are mapped using the coloring scheme from Fig. 1. Time averages of the distances for each residue pair are combined across proteins within an ensemble, instead of using the actual distances, mainly for computational efficiency, but also made the distributions of the variables more normal, as per the central limit theorem. For all MD simulations, an initial region of higher RMSD fluctuation (100 ps) was discarded to reduce residual effects coming from prior equilibration. We further filter out stochastic variations by basing ourselves on the network concept of preferential attachment, which is the tendency of scale-free networks to attach new nodes to those that already have high connectivities⁵². The five top-ranked residue positions are subsequently prioritized on the basis of their connectivities showing statistically-significant larger or smaller distances across the ensembles.

It is remarkable to note that despite the stochasticity of different sections of the experiment compounded with protein variations, our results show features common to all drug complexes. The base of the cantilever (very close to the 60's loop) is drawn closer to the catalytic core for all drug complexes (Figs 4 and 5) in the resistant state. All the mutations present in each ARV's resistance ensemble are shown in Supplementary Table S4, where known accessory and major drug resistance mutations (DRMs)⁵³ are shown in bold black and red fonts respectively. From our simulations, a similarly conserved behavior is not immediately apparent from the degree centralities of those residues with larger distances in the susceptible ensemble, but can be seen from their mapping onto protease 3D structures. A lateral widening involving the elbow region and/ or the 10's loop of the fulcrum is generally observed across all PIs. Part of this motion is described by Hornak and co-workers⁵⁴ as events leading to flap opening, involving a concerted downward motion of the cantilever, fulcrum and flap elbow with an upward motion of the catalytic aspartate from the floor of the binding cavity. Further, we observed that residues of the receptor cavity do not appear amongst any of the top-ranked residues for each drug and ensemble, and behave in a quite opposite manner, with very low degree centralities. Occasional spikes did manifest themselves for some cavity residues, but these can be ignored as they may be chance events that would not be ranked similarly upon replication of the experiment, as seen in Supplementary Fig. S5. Low degree centralities in both ensembles (i.e. neither larger in the resistant nor in the susceptible ensembles) would point to the fact there is no consistent motion within the binding cavity that would define the state of drug-resistance or susceptibility, at least not within the time limit and conformational landscapes explored. This hints at a receptor pocket that is very malleable

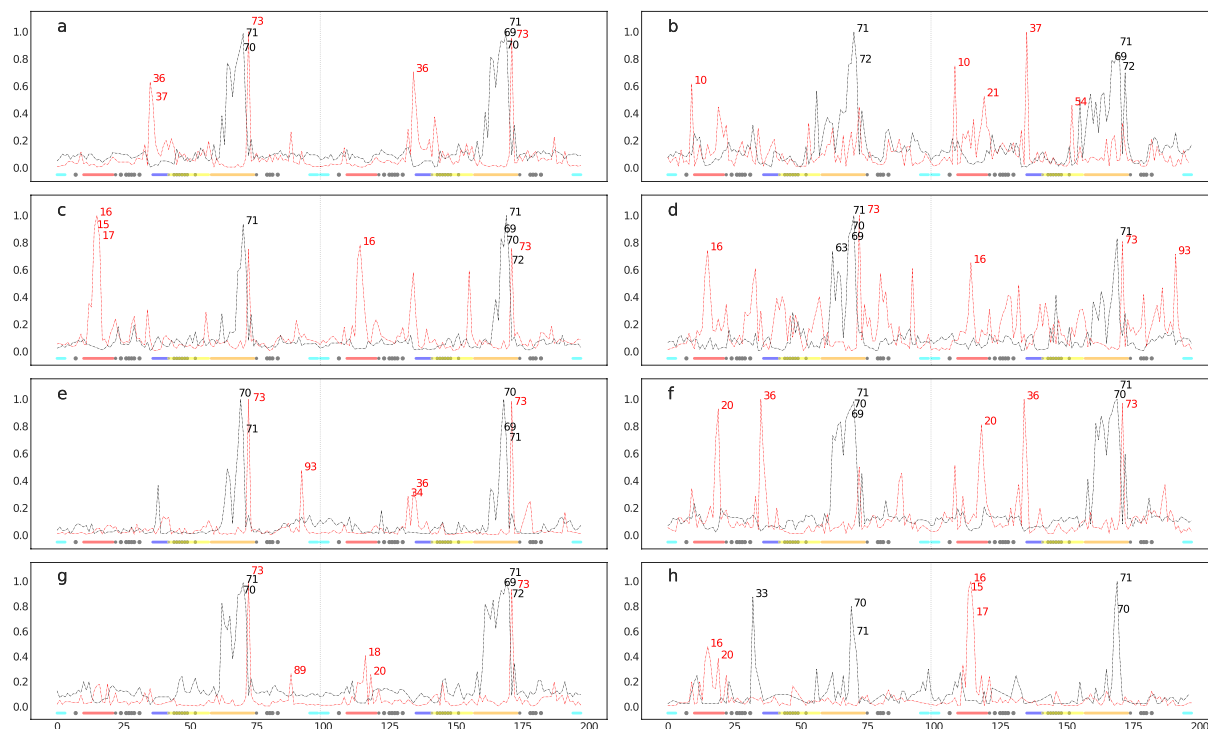


Figure 4. Normalized degree centralities of significantly larger (red lines) and smaller (black lines) distances observed in resistant ensembles for 8 FDA-approved protease inhibitor complexes, namely ATV (**a**), DRV (**b**), FPV (**c**), IDV (**d**), LPV (**e**), NFV (**f**), SQV (**g**) and TPV (**h**). The top 5 residue positions with the highest connectivities are labelled at the peaks in each graph. Inserted underneath are the functional protease residues depicted as colored dots, namely the fulcrum (red), the elbow (blue), the flap (yellow), the cantilever (orange), the interface (cyan) and the binding cavity residues (grey).

with multiple internal cavity dynamics that can lead to similar states, both within and between PI drug classes. A second scenario that could result in such low connectivities would be that cavity residues move in a coordinated manner across ensembles irrespective of drug exposure, which is unlikely.

ARV-specific results. In Fig. 4a for ATV residues showing smaller distances in the resistant ensemble include positions 70, 71 (on chains A and B) and position 69–71 (chain B), while larger distances are at positions 36, 37, 73 (chain A) and at positions 36, 73 (chain B). Mapping these positions onto protein structures (Fig. 5a) shows that regions predicted to be larger in the resistant ensemble move in a lateral outward direction, favoring a wider conformation, while regions predicted to be smaller in the resistance ensemble (or larger in the susceptible ensemble) show an upward motion with respect to the flaps. The residues involved in widening and shortening around the binding cavity show a high level of symmetry between each monomer of the protein. A very similar profile was obtained upon replication, with residue 36 peaking from within chain A instead of chain B. Such conservation in behavior may find direct application in drug resistance prediction or for feature augmentation for improving machine learning prediction of resistance. Mutations present in the ATV resistance ensemble include both accessory DRMs 10IVF, 32I, 33FV, 34Q, 46LI, 48V, 53L, 54LVM, 60E, 62V, 64VM, 71IV, 73STA, 90M, 93LM and a major DRM 84V, in addition to multiple other variations. In the case of DRV (Figs 4b and 5b), residues at positions 71 and 72 (chain A) and positions 69, 71, 72 (chain B) move closer to the catalytic wall in the resistance ensemble in a symmetric fashion. Larger distances are at positions 10 (chain A) and 10, 21, 37, 54 (chain B). The elbow movement is not mirrored in chain A, however position 10 in chains A and B move away from the plane spanning the surface of the page, showing another way of active site expansion in addition to elbow flaring associated with the resistance ensemble. Residue 54 (chain B) is also seen to move away from the the binding cavity, but same does not occur under the replication (Supplementary Fig. S5). DRV's resistance ensemble includes amongst other variations, the accessory DRMs 11I, 32I, 33F, 89V and the major DRMs 47V, 54LM, 84V.

For FPV (Figs 4c and 5c), smaller distances in the resistance ensemble are at positions 71 (chain A) and 69–72 (chain B), while larger distances in the resistance ensemble are at positions 15–17 (chain A) and 16, 73 (chain B). As in DRV lateral expansion is observed, but mainly involves the 10's region in addition to the surface residue 73 in both replicates. The constriction behavior is also reproduced very closely in the replicate (Supplementary Fig. S5). The major DRM includes 84V while accessory DRMs include 10IVF, 32I, 46LI, 47V, 54LVM, 73S, 76V, 82TA, 90M in addition to other variations.

IDV (Figs 4d and 5d) displays smaller distances in its resistance ensemble at positions 63, 69–71 (chain A) and 71 (chain B). Larger distances are observed for the same ensemble at positions 16, 73 (chain A) and 16, 73, 93 (chain B). Upon replication, same residues were found to be involved in expansion, while only chain A

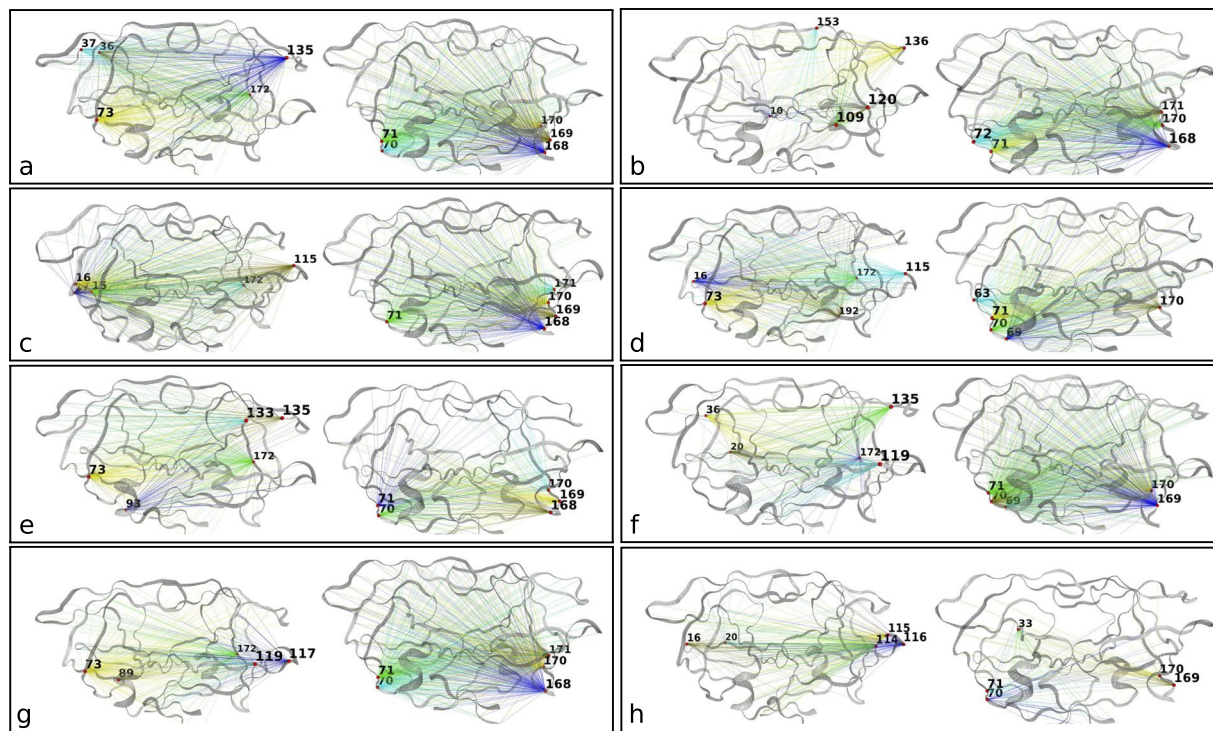


Figure 5. Mapping of the edges for top-ranked degree centralities onto HIV 3D protease structures for the significantly larger (left) and smaller (right) distances observed in resistant ensembles for 8 FDA-approved protease inhibitor complexes, namely ATV (a), DRV (b), FPV (c), IDV (d), LPV (e), NFV (f), SQV (g) and TPV (h).

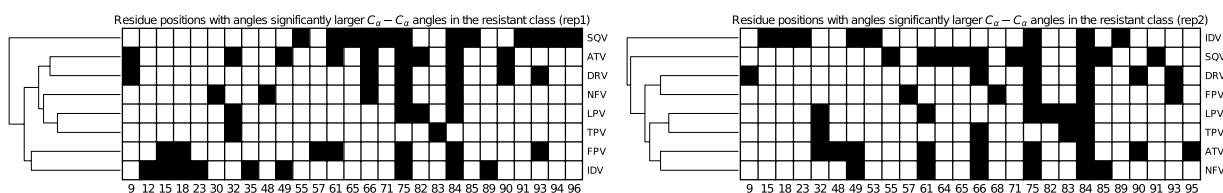


Figure 6. Heat map of residue positions with significantly larger C_{α} angles in the resistant ensemble for each PI. The hierarchical cluster tree is displayed on the left. The first replicate is at the left and the second replicate is at the right.

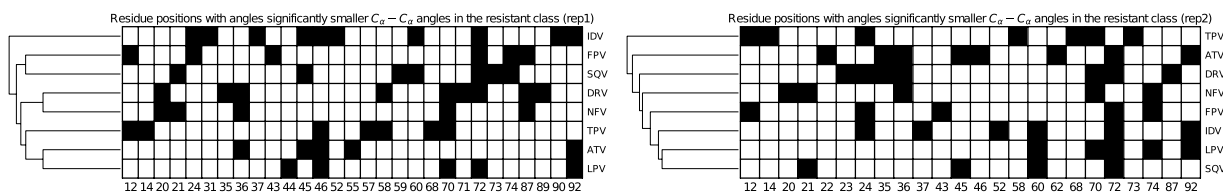


Figure 7. Heat map of residue positions with significantly smaller C_{α} angles in the resistant ensemble for each PI. The hierarchical cluster tree is displayed on the left. The first replicate is at the left and the second replicate is at the right.

showed the cantilever loop compaction towards the active site. Once more, the cantilever residue 73 is found to contribute to lateral widening in both chains. Replication identified identical residues involved in expansion at the 10's loop from both chains, while those involved in compaction included residues 69–71 only from chain A (Supplementary Fig. S5). Major DRMs in the IDV resistance ensemble include 46LI, 82FTA, 84V and the accessory mutations 10IV, 20R, 32I, 36I, 54V, 71TV, 73SA, 76V, 77I, 90M.

Resistance in LPV (Figs 4e and 5e) was associated with smaller distances in the resistance ensemble at positions 70, 71 (chain A) and position 69–71 (chain B), while larger distances were located at positions 73, 93 (chain A) and positions 34, 36, 73 (chain B). Replicate runs are very concordant for residues involved in expansion with

the exception of residues 36 and 81 in chain B, which rank differently despite displaying similar trends. Those involved in compaction again point to the cantilever residues of both chains, whereby residue 71 is replaced by 69 in the replicate. The major DRMs of the resistance ensemble consist of 32I, 47VA, 76V, 82SFTA while accessory DRMs comprise 10IFV, 20RM, 24I, 33F, 46LI, 50V, 53L, 54LTVMS, 63P, 71VT, 73S, 84V, 90M.

In the case of NFV (Figs 4f and 5f), shorter distances for the resistance ensemble were at positions 69–71 (chain A) and positions 70, 71 (chain B), while larger distances for the same ensemble were at positions 20, 36 (chain A) and at positions 20, 36, 73 (chain B). In the replicate run, a very similar profile is observed, however it would appear that residue 20 (part of the fulcrum) and 36 (close to elbow) are moving in concert during expansion. Contraction is observed as for other ARVs, close to the cantilever loop region. Major DRMs of the resistance ensemble consist of 30N, 90M and accessory DRMs consist of 10IE, 36I, 46LI, 71TV, 77I, 82FA, 84V, 88D.

SQV (Figs 4g and 5g) displays smaller distances in the resistance ensemble at positions 70, 71 (chain A) and positions 69, 71, 72 (chain B). Larger distances for the same ensemble are observed at positions 73, 89 (chain A) and positions 18, 20, 73 (chain B). Very similar symmetric compaction is observed at the fulcrum region as seen for other ARVs on both chains and the widening peak positions are also very similar despite a slightly changed degree ranking. DRMs for the resistance ensemble include the majors 48V, 90M and the accessory mutations 10I, 54LV, 62V, 71TV, 73S, 77I, 82A, 84V.

In the case of TPV (Figs 4h and 5h), smaller distances in the resistances are at positions 33, 60, 71 (chain A) and positions 70, 71 (chain B), while larger distances are at positions 16, 20 (chain A) and positions 15–17 (chain B). Replication reproduced compactions once more, close to the cantilever loop but also included the buried residue at position 33 on chain A, surrounded by the 80's loop, the cantilever and the elbow regions. According to our simulation conditions, compaction at this region appears to be specific to TPV. Lateral expansions, though not identical, are also closely reproduced around the 10's loop region. The resistance ensemble includes the major DRMs 47V, 58E, 74P, 82LT, 83D, 84V and the accessory DRMs 10V, 33F, 36IV, 43T, 46L, 54VM, 89VM.

Local evaluation via statistical tests coupled with angular distributions. We further investigated receptor backbone movement by comparing angle distributions occurring at protein C_{α} atoms. Absolute conservation was observed at residue position 84 in only one of the replicates in Fig. 6. Positions 75 and 84 however displayed strongly conserved larger angles in the resistance ensemble, including ATV, DRV, FPV, IDV, LPV, NFV and SQV. At 99% confidence, one-tailed t-tests did not detect any strong conservation of global angular behavior - both for the same drug replicated and across all drugs as shown by the non-reproducible clustering patterns in Figs 6 and 7. This supports the fact that the enzyme is very malleable, even in the closed conformation complexed with the drug and points to the fact that multiple residue arrangements along the backbone can lead to the same effect.

In conclusion, HIV protease inhibitors are used to delay the symptoms associated with late stages of the infection, however resistance is unrelenting due to the virus's resilience to the current drug designs. Drug resistance patterns are complex. Nevertheless, our large scale simulations show that despite various DRMs and additional variations, lateral expansion and fulcrum compaction are conserved in the drug resistance state, both within and between different types of PIs. The observation of conserved lateral expansion provides additional support for investigating alternative drug-targeting sites rather than the active site, as done in⁵⁵. The results may be hinting at (1) conserved mechanistic ripple effects emanating from certain similarities in PI drug design, which possibly hints at how crucial these preliminary early movements are in leading to a less-favorable drug positioning within the active site, or (2) a well-conserved pair of local motions associated with drug resistance lying underneath the complexity of DRM patterns.

Analysis of the backbone motions hinted that there is no single angular trajectory leading to resistance, even for the same sequence. Knowledge of characteristic motions around similar energy wells may be an interesting and inexpensive route for supplementing extant drug resistance prediction approaches in HIV subtype B. Given phenotypically-labelled protease sequences from other subtypes, a similar experimental design may prove to be quite useful in extracting conserved local motions. Additionally, this approach could theoretically extend to proteases harbouring indel mutations by selecting homologous residues after simulation, shedding more light on sequences that are more divergent from the consensus B subtype.

We have used MD simulations coupled with network centrality measures to identify common structural features in drug-resistant mutations of HIV protease. As opposed to the conventional ways of constructing residue contact networks using distance cut-offs, we used statistical tests, thus mitigating the known effect of edge discontinuity³⁴ which may arise when pairwise distances are very close to, but not bound by, the chosen cut-off distance. To our knowledge, this method is novel, although elastic network models were used to determine the functional effects of variants in other proteins⁵⁶. While the Anisotropic (ANM) and Gaussian Network Models (GNM) are based on the application of Hooke's potential on a single structure with a uniform spring constant, our method is based on the more thorough Newtonian mechanical simulation employing an all-atom forcefield to analyse a large number of independent observations. We expect that our method will be highly useful in other cases for analysing protein structural variations. For instance, one could use a subset of validated antimalarial drug targets from artemisinin-resistant variants and another batch of sequences for artemisinin-susceptible variants and extract subtle motions hidden within the protein dynamics.

Methods

Dataset preparation. HIV subtype B protease sequence variants labelled with fold drug resistance ratios were obtained from the Stanford HIVdb unfiltered dataset¹⁹. These were reconstituted and filtered as explained in²⁰. After ranking the sequences based on decreasing average distance for each of the 8 PIs, 100 highly-resistant and 100 hyper-susceptible sequences were short-listed, using cut-offs defined in⁵⁷. These two classes of sequences

are henceforth referred as to the resistant and the susceptible ensembles respectively. Sequences are provided in Supplementary Dataset S6. Pandas 0.21.0⁵⁸ was used for dataset storage and manipulation. Seaborn 0.7.1 and matplotlib 2.1.0⁵⁹ were used for plotting.

Homology modelling. Modeller (version 9.16) was used to model each of the protein sequences in their closed receptor conformation. The main criteria for choosing the templates, in order of selection were (1) presence of the PI complexed within the a closed conformation receptor active and (2) high resolution of the crystal structures. These two characteristics were determined to be important in giving a good starting point for observing comparable changes within short dynamics simulations. High resolution (<1.55 Å) crystal structure templates were thus retrieved for each of the 8 available PIs from the HIVdb dataset (PDB accessions: 3NU3⁶⁰, 3EL9⁶¹, 2HS1⁶², 2AVO⁶³, 2O4S⁶⁴, 3EL5⁶¹, 2NMZ⁶⁵ and 3SPK⁶⁶). Very slow refinement was used, with a random seed set at -10000 , while model quality was assessed using z-DOPE scores. As a preparation for molecular docking, template crystal structures were systematically preprocessed to only retain high-occupancy side-chain rotamers. The last rotamers for each concerned residue were kept in cases of equal occupancies. Interfacial water was retained from each template crystal structure by choosing any water molecule shared between the ligand and receptor flaps (ILE50 from chains A and B), at an intersecting distance of 3 Å, except for the case of TPV, which does not require such for stability.

Ligand docking. Flexible ligand docking was performed using AutoDock Vina (version 1.1.2)³³ to place each PI in its respective receptor variant. The docking center (20.147, 29.716, 16.093) was picked from a saquinavir atom from template 2NMZ subsequently used as reference to align the totality of the homology models using ProDy⁶⁷. Receptors were protonated to pH7 using PDB2PQR (version 2.1.0)⁶⁸ using the PROPKA method before merging non-polar hydrogen atoms and assigning Gasteiger partial charges using the *prepare_receptor4.py* tool from AutoDockTools (ADT)⁶⁹, whilst having interfacial water present. Ligands were fully-protonated using ADT's *prepare_ligand4.py* tool. A grid box size of dimensions $20 \times 26 \times 20$ Å³ and an exhaustiveness value of 16 was chosen for ligand docking at the designated grid center.

Molecular dynamics. The previously-protonated receptors were used, while parameters for the docked ligand poses were determined using ACPYPE⁷⁰ after full protonation using VEGA (version 3.1.1)⁷¹. All 8×200 complexes were prepared for molecular dynamics using GROMACS (version 2016.1)⁷². The AMBER03 forcefield was used with a short-range non-bonded interaction cut-off distance of 1.2 nm. Long-range electrostatics were handled using the smooth Particle Mesh Ewald algorithm. Energy-minimization was performed using the method of steepest descent after neutralizing charges using 0.15 M sodium chloride in SPC-modeled water within a triclinic periodic box. A 50 ps temperature equilibration (at 310 K) was followed by 50 ps of pressure equilibration (1 atm) with time steps of 2 fs and finally a 2 ns production MD was performed at the same temperature, pressure and time step. All MD runs were distributed over a 2400-core queue with 24 cores per job using GNU Parallel (version 20160422)⁷³, managed by the PBS Professional scheduler over the lengau cluster (Centre for High Performance Computing (CHPC)).

Trajectory analysis. After generating MD trajectories, the proteins were centered and rotations/ translations were removed using the *trjconv* command in GROMACS. RMSD values were first evaluated to detect any failure in correcting periodic boundary conditions. These plots identified an initial period of fluctuation spanning the first 100 ps, which were dropped from any subsequent analysis. R_g values were calculated to have an overview of the levels of compaction observed in the resistance ensemble compared to susceptible ensemble for each drug investigated. Thereafter, local analyses were performed: (a) Welsch t-tests were evaluated over pairwise residue distances across the ensembles. To do so, pairwise C_β (and C_α for glycine) atom distances from each trajectory were time-averaged within each ensemble. For each drug, each pairwise residue distance was aggregated into separate two-dimensional arrays - one for each ensemble. The t-tests were then performed between each analogous array at a 99% confidence level. (b) Similarly, the time-averaged angles between C_α residue triplets were computed for each complex within an ensemble and compared against the analogous array of time-averaged angles in the other ensemble using t-tests. Only those angles corresponding to the negative logarithm (base 10) p-value being above 2.5 standard deviations were retained for either of the larger or smaller angles in the resistance ensembles. Bonferroni correction was applied in both approaches to correct for multiple testing and reduce chances of false positives. Finally, the angles found to be significant for each drug were clustered by average linkage from the matrix of pairwise Euclidean distances. The MDTraj library (version 1.9.1)⁷⁴ was used in Python 3.5 for trajectory distance and angle calculations. Numpy 1.13.3 and scipy 1.0.0⁷⁵ were used for general computations and statistical tests respectively.

Network analysis. Network graphs were built from nodes corresponding to C_β (or glycine C_α) atoms. Edges were obtained from significant p-values obtained from independent t-tests performed on arrays of time-averaged pairwise residue distances. In other words, each time-averaged pairwise distance $\langle D_{ij} \rangle$ for a given protein concatenated to those of other proteins within the ensemble. Each array of $\langle D_{ij} \rangle$ values is then compared to its corresponding position in the other ensemble of $\langle D'_{ij} \rangle$ values using 2 sample t-tests. In order to expose more information, one-tailed tests were performed to determine whether distances are larger or smaller between the resistance ensembles. Same method was applied for all drugs. Finally the node degree centralities were calculated and the top 5 most central nodes for both higher and lower distances were shown as text labels for each drug. Network construction and analysis were performed using the NetworkX library (version 1.11)⁷⁶. Edge mappings onto protein structures were generated using the NGLview library (version 1.0)⁷⁷.

Data Availability

All data presented is either in the manuscript or in the Supplementary information.

References

- Riemenschneider, M. & Heider, D. Current Approaches in Computational Drug Resistance Prediction in HIV. *Current HIV research* 1–9 (2016).
- Cai, Y. *et al.* Drug Resistance Mutations Alter Dynamics of Inhibitor-Bound HIV-1 Protease. *Journal of chemical theory and computation* **10**, 3438–3448, <https://doi.org/10.1021/ct4010454> (2014).
- Doekes, H. M., Fraser, C. & Lythgoe, K. A. Effect of the Latent Reservoir on the Evolution of HIV at the Within- and Between-Host Levels. *PLoS Computational Biology* **13**, e1005228, <https://doi.org/10.1371/journal.pcbi.1005228> (2017).
- Liu, T. F. & Shafer, R. W. Web resources for HIV type 1 genotypic-resistance test interpretation. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* **42**, 1608–1618, <https://doi.org/10.1086/503914> (2006).
- Weber, I. T., Kneller, D. W. & Wong-Sam, A. Highly resistant HIV-1 proteases and strategies for their inhibition. *Future medicinal chemistry* **7**, 1023–38, <https://doi.org/10.4155/fmc.15.44> (2015).
- Drag, M. & Salvesen, G. S. Emerging principles in protease-based drug discovery. *Nature Reviews Drug Discovery* **9**, 690–701, <https://doi.org/10.1038/nrd3053> (2010).
- Prashar, V., Bihani, S. C., Ferrer, J. L. & Hosur, M. V. Structural basis of why nelfinavir-resistant D30N mutant of HIV-1 protease remains susceptible to saquinavir. *Chemical Biology and Drug Design* **86**, 302–308, <https://doi.org/10.1111/cbdd.12494> (2015).
- Fun, A., Wensing, A. M. J., Verheyen, J. & Nijhuis, M. Human Immunodeficiency Virus gag and protease: Partners in resistance. *Retrovirology* **9**, 63, <https://doi.org/10.1186/1742-4690-9-63> (2012).
- Weber, I. T. & Agniswamy, J. HIV-1 Protease: Structural Perspectives on Drug Resistance. *Viruses* **1**, 1110–1136, <https://doi.org/10.3390/v1031110> (2009).
- Nalam, M. N. L. & Schiffer, C. A. New approaches to HIV protease inhibitor drug design II: Testing the substrate envelope hypothesis to avoid drug resistance and discover robust inhibitors, <https://doi.org/10.1097/COH.0b013e3283136cee> (2008).
- Rhee, S.-Y. *et al.* HIV-1 Drug Resistance Mutations: Potential Applications for Point-of-Care Genotypic Resistance Testing. *PLoS One* **10**, 1–17, <https://doi.org/10.1371/journal.pone.0145772> (2015).
- Wallis, C. L. *et al.* Drug susceptibility and resistance mutations after first-line failure in resource limited settings. *Clinical Infectious Diseases* **59**, 706–715, <https://doi.org/10.1093/cid/ciu314> (2014).
- Yu, X., Weber, I. T. & Harrison, R. W. Sparse Representation for Prediction of HIV-1 Protease Drug Resistance. *Proceedings of the 2013 SIAM International Conference on Data Mining. SIAM International Conference on Data Mining* **2013**, 342–349, <https://doi.org/10.1137/1.9781611972832.38> (2013).
- Toor, J. S. *et al.* Prediction of drug-resistance in HIV-1 subtype C based on protease sequences from ART naive and first-line treatment failures in North India using genotypic and docking analysis. *Antiviral Research* **92**, 213–218, <https://doi.org/10.1016/j.antiviral.2011.08.005> (2011).
- Jenwitheesuk, E. & Samudrala, R. Prediction of HIV-1 protease inhibitor resistance using a protein – inhibitor flexible docking approach. *Antiviral Therapy* **10**, 157–166 (2005).
- Mao, Y. Dynamical basis for drug resistance of HIV-1 protease. *BMC structural biology* **11**, 31, <https://doi.org/10.1186/1472-6807-11-31> (2011).
- Antunes, D. A. *et al.* New insights into the in silico prediction of HIV protease resistance to nelfinavir. *PLoS One* **9**, <https://doi.org/10.1371/journal.pone.0087520> (2014).
- Cao, Z. W. *et al.* Computer prediction of drug resistance mutations in proteins. *Drug Discovery Today* **10**, 521–529, [https://doi.org/10.1016/S1359-6446\(05\)03377-5](https://doi.org/10.1016/S1359-6446(05)03377-5) (2005).
- Stanford HIVdb. Genotype-Phenotype Datasets (2014).
- Sheik Amamuddy, O., Bishop, N. T. & Tastan Bishop, Ö. Improving fold resistance prediction of HIV-1 against protease and reverse transcriptase inhibitors using artificial neural networks. *BMC bioinformatics* **18**, 369, <https://doi.org/10.1186/s12859-017-1782-x> (2017).
- Šali, A. Modelling mutations and homologous proteins. *Current Opinion in Biotechnology* **6**, 437–451, [https://doi.org/10.1016/0958-1669\(95\)80074-3](https://doi.org/10.1016/0958-1669(95)80074-3) (1995).
- US Food and Drug Administration. LEXIVA[®] (fosamprenavir calcium) Tablets and Oral Suspension (2009).
- Özen, A., Haliloglu, T. & Schiffer, C. A. Dynamics of preferential substrate recognition in HIV-1 protease: Redefining the substrate envelope. *Journal of Molecular Biology* **410**, 726–744, <https://doi.org/10.1016/j.jmb.2011.03.053> (2011).
- Liu, Z. *et al.* Effects of Hinge-region Natural Polymorphisms on Human Immunodeficiency Virus-Type 1 Protease Structure, Dynamics, and Drug Pressure Evolution. *The Journal of biological chemistry* **291**, 22741–22756, <https://doi.org/10.1074/jbc.M116.747568> (2016).
- Bernardi, R. C., Cann, I. & Schulten, K. Molecular dynamics study of enhanced Man5B enzymatic activity. *Biotechnology for Biofuels* **7**, 83, <https://doi.org/10.1186/1754-6834-7-83> (2014).
- Batista, P. R. *et al.* Free Energy Profiles along Consensus Normal Modes Provide Insight into HIV-1 Protease Flap Opening. *Journal of Chemical Theory and Computation* **7**, 2348–2352, <https://doi.org/10.1021/ct200237u> (2011).
- Chipot, C. Frontiers in free-energy calculations of biological systems. *Wiley Interdisciplinary Reviews: Computational Molecular Science* **4**, 71–89, <https://doi.org/10.1002/wcms.1157> (2014).
- Cournia, Z., Allen, B. & Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *Journal of Chemical Information and Modeling* **57**, 2911–2937, <https://doi.org/10.1021/acs.jcim.7b00564> (2017).
- Zhang, H. *et al.* Accurate estimation of the standard binding free energy of netropsin with DNA. *Molecules* **23**, 1–15, <https://doi.org/10.3390/molecules23020228> (2018).
- Barabási, A.-L. & Pósfai, M. Network science. In *Network Science*, chap. Chapter 5 (Cambridge University Press, Cambridge, 2016).
- Van Der Spoel, D. *et al.* GROMACS: Fast, flexible, and free, <https://doi.org/10.1002/jcc.20291> (2005).
- Feyfant, E., Sali, A. & Fiser, A. Modeling mutations in protein structures. *Protein science: a publication of the Protein Society* **16**, 2030–41, <https://doi.org/10.1110/ps.072855507> (2007).
- Trott, O. & Olson, A. J. AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry* **31**, 455–461, <https://doi.org/10.1002/jcc.21334> (2009).
- Horvath, S. *Weighted Network Analysis: Applications in Genomics and Systems Biology*, <https://doi.org/10.1007/978-1-4419-8819-5> (Springer, 2011).
- Freeman, L. C. & White, D. R. Using Galois Lattices to Represent Network Data. *Sociological Methodology* 127–146 (1993).
- Brown, D. K. & Tastan Bishop, Ö. Role of Structural Bioinformatics in Drug Discovery by Computational SNP Analysis: Analyzing Variation at the Protein Level. *Global Heart* **12**, 151–161, <https://doi.org/10.1016/j.gheart.2017.01.009> (2017).
- Brown, D. K., Sheik Amamuddy, O. & Tastan Bishop, Ö. Structure-Based Analysis of Single Nucleotide Variants in the Renin-Angiotensinogen Complex. *Global Heart*, <https://doi.org/10.1016/j.gheart.2017.01.006> (2017).
- Petsko, G. A. & Yates, J. R. Analyzing molecular interactions. *Current Protocols in Bioinformatics*, <https://doi.org/10.1002/0471250953.bi0801s36> (2011).

39. Hou, T., Li, N., Li, Y. & Wang, W. Characterization of Domain–Peptide Interaction Interface: Prediction of SH3 Domain-Mediated Protein–Protein Interaction Network in Yeast by Generic Structure-Based Models. *Journal of Proteome Research* **11**, 2982–2995, <https://doi.org/10.1021/pr3000688> (2012).
40. Chatterjee, P., Roy, D., Bhattacharyya, M. & Bandyopadhyay, S. Biological networks in Parkinson's disease: An insight into the epigenetic mechanisms associated with this disease. *BMC Genomics* **18**, 721, <https://doi.org/10.1186/s12864-017-4098-3> (2017).
41. Hu, Z. *et al.* Ligand binding and circular permutation modify residue interaction network in DHFR. *PLoS Computational Biology* **3**, 1097–1107, <https://doi.org/10.1371/journal.pcbi.0030117> (2007).
42. Xue, W., Jiao, P., Liu, H. & Yao, X. Molecular modeling and residue interaction network studies on the mechanism of binding and resistance of the HCV NS5B polymerase mutants to VX-222 and ANA598. *Antiviral Research* **104**, 40–51, <https://doi.org/10.1016/j.antiviral.2014.01.006> (2014).
43. Piovesan, D., Minervini, G. & Tosatto, S. C. The RING 2.0 web server for high quality residue interaction networks. *Nucleic Acids Research* **gkw315**, <https://doi.org/10.1093/nar/gkw315> (2016).
44. Brown, D. K. *et al.* MD-TASK: a software suite for analyzing molecular dynamics trajectories. *Bioinformatics* **33**, 2768–2771, <https://doi.org/10.1093/bioinformatics/btx349> (2017).
45. Penkler, D. L., Atilgan, C. & Tastan Bishop, Ö. Allosteric Modulation of Human Hsp90 α Conformational Dynamics. *Journal of Chemical Information and Modeling* **58**, 383–404, <https://doi.org/10.1021/acs.jcim.7b00630> (2017).
46. Kwak, S. G. & Kim, J. H. Central limit theorem: the cornerstone of modern statistics. *Korean Journal of Anesthesiology* **70**, 144, <https://doi.org/10.4097/kjae.2017.70.2.144> (2017).
47. Goldfarb, N. E. *et al.* Defective Hydrophobic Sliding Mechanism and Active Site Expansion in HIV-1 Protease Drug Resistant Variant Gly48Thr/Leu89Met: Mechanisms for the Loss of Saquinavir Binding Potency. *Biochemistry* **54**, 422–433, <https://doi.org/10.1021/bi501088e> (2015).
48. Ohtaka, H., Schön, A. & Freire, E. Multidrug Resistance to HIV-1 Protease Inhibition Requires Cooperative Coupling between Distal Mutations. *Biochemistry* **42**, 13659–13666, <https://doi.org/10.1021/bi0350405> (2003).
49. Louis, J. M. *et al.* The L76V drug resistance mutation decreases the dimer stability and rate of autoprocessing of HIV-1 protease by reducing internal hydrophobic contacts. *Biochemistry* **50**, 4786–4795, <https://doi.org/10.1021/bi200033z> (2011).
50. Logsdon, B. C. *et al.* Crystal Structures of a Multidrug-Resistant Human Immunodeficiency Virus Type 1 Protease Reveal an Expanded Active-Site Cavity. *Journal of Virology* **78**, 3123–3132, <https://doi.org/10.1128/JVI.78.6.3123-3132.2004> (2004).
51. Martin, M. *et al.* “Wide-Open” 1.3 Å Structure of a Multidrug-Resistant HIV-1 Protease as a Drug Target. *Structure* **13**, 1887–1895, <https://doi.org/10.1016/j.str.2005.11.005> (2005).
52. Barabasi, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, <https://doi.org/10.1126/science.286.5439.509> (1999).
53. Wensing, A. *et al.* 2017 Update of the Drug Resistance Mutations in HIV-1. *Top Antivir Med* **24**, 132–133 (2017).
54. Hornak, V., Okur, A., Rizzo, R. C. & Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **103**, 915–920, <https://doi.org/10.1073/pnas.0508452103> (2006).
55. Meng, X. M., Hu, W. J., Mu, Y. G. & Sheng, X. H. Effect of allosteric molecules on structure and drug affinity of HIV-1 protease by molecular dynamics simulations. *Journal of Molecular Graphics and Modelling* **70**, 153–162, <https://doi.org/10.1016/j.jmkgm.2016.09.015> (2016).
56. Ponzoni, L. & Bahar, I. Structural dynamics is a determinant of the functional significance of missense variants. *Proceedings of the National Academy of Sciences* **115**, 4164–4169, <https://doi.org/10.1073/pnas.1715896115> (2018).
57. Hedlin, H. Genotype-Phenotype Datasets: DRMcV (2014).
58. McKinney, W. Data Structures for Statistical Computing in Python. In *Proceedings of the 9th Python in Science Conference*, vol. 445, 51–56 (2010).
59. Hunter, J. D. Matplotlib: A 2D graphics environment. *Computing in Science and Engineering* **9**, 99–104, <https://doi.org/10.1109/MCSE.2007.55> (2007).
60. Shen, C.-H., Wang, Y.-F., Kovalevsky, A. Y., Harrison, R. W. & Weber, I. T. Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters. *The FEBS journal* **277**, 3699–714, <https://doi.org/10.1111/j.1742-4658.2010.07771.x> (2010).
61. King, N. M. *et al.* Extreme Entropy–Enthalpy Compensation in a Drug-Resistant Variant of HIV-1 Protease. *ACS Chemical Biology* **7**, 1536–1546, <https://doi.org/10.1021/cb300191k> (2012).
62. Kovalevsky, A. Y. *et al.* Ultra-high resolution crystal structure of HIV-1 protease mutant reveals two binding sites for clinical inhibitor TMC114. *Journal of molecular biology* **363**, 161–73, <https://doi.org/10.1016/j.jmb.2006.08.007> (2006).
63. Liu, F. *et al.* Kinetic, stability, and structural changes in high-resolution crystal structures of HIV-1 protease with drug-resistant mutations L24I, I50V, and G73S. *Journal of molecular biology* **354**, 789–800, <https://doi.org/10.1016/j.jmb.2005.09.095> (2005).
64. Muzammil, S. *et al.* Unique Thermodynamic Response of Tipranavir to Human Immunodeficiency Virus Type 1 Protease Drug Resistance Mutations. *Journal of Virology* **81**, 5144–5154, <https://doi.org/10.1128/JVI.02706-06> (2007).
65. Tie, Y. *et al.* Atomic resolution crystal structures of HIV-1 protease and mutants V82A and I84V with saquinavir. *Proteins: Structure, Function, and Bioinformatics* **67**, 232–242, <https://doi.org/10.1002/prot.21304> (2007).
66. Wang, Y. *et al.* The higher barrier of darunavir and tipranavir resistance for HIV-1 protease. *Biochemical and Biophysical Research Communications* **412**, 737–742, <https://doi.org/10.1016/j.bbrc.2011.08.045> (2011).
67. Bakan, A., Meireles, L. M. & Bahar, I. ProDy: Protein dynamics inferred from theory and experiments. *Bioinformatics* **27**, 1575–1577, <https://doi.org/10.1093/bioinformatics/btr168> (2011).
68. Dolinsky, T. J., Nielsen, J. E., McCammon, J. A. & Baker, N. A. PDB2PQR: An automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Research* **32**, <https://doi.org/10.1093/nar/gkh381> (2004).
69. Morris, G. M. *et al.* AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **30**, 2785–2791, <https://doi.org/10.1002/jcc.21256> (2009).
70. Sousa da Silva, A. W. & Vranken, W. F. ACPYPE - AnteChamber PYthon Parser interfacE. *BMC research notes* **5**, 367, <https://doi.org/10.1186/1756-0500-5-367> (2012).
71. Pedretti, A., Villa, L. & Vistoli, G. Atom-type description language: A universal language to recognize atom types implemented in the VEGA program. *Theoretical Chemistry Accounts* **109**, 229–232, <https://doi.org/10.1007/s00214-002-0402-6> (2003).
72. Abraham, M. J. *et al.* Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1–2**, 19–25, <https://doi.org/10.1016/j.softx.2015.06.001> (2015).
73. Tange, O. GNU Parallel: the command-line power tool. *login: The USENIX Magazine* **36**, 42–47, <https://doi.org/10.5281/zenodo.16303> (2011).
74. McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **109**, 1528–1532, <https://doi.org/10.1016/j.bpj.2015.08.015> (2015).
75. Van Der Walt, S., Colbert, S. C. & Varoquaux, G. The NumPy array: A structure for efficient numerical computation. *Computing in Science and Engineering* **13**, 22–30, <https://doi.org/10.1109/MCSE.2011.37> (2011).
76. Hagberg, A. A., Schult, D. A. & Swart, P. J. Exploring network structure, dynamics, and function using NetworkX. In Varoquaux, G., Vaught, T. & Millman, J. (eds) *Proceedings of the 7th Python in Science Conference (SciPy2008)*, vol. 836, 11–15 (Pasadena, CA USA, 2008).

77. Nguyen, H., Case, D. A. & Rose, A. S. NGLview—interactive molecular graphics for Jupyter notebooks. *Bioinformatics*, <https://doi.org/10.1093/bioinformatics/btx789> (2017).

Acknowledgements

We thank the Centre for High Performance Computing (CHPC), South Africa for computational resources, and the National Research Foundation (NRF), South Africa for funding under grant 93690. The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of the funder.

Author Contributions

Author contributions: O.S.A. designed and performed the research and wrote the first draft of the manuscript; Ö.T.B. and N.T.B. supervised the research and edited the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-36041-8>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018