The evolution of the HIV-1 protease folding stability

David Ferreiro,[1,2] Ruqaiya Khalil,[1,2] María J. Gallego,[1,2] Nuno S. Osorio,[3,4] and Miguel Arenas[1,2,5,*,†]

[1]CINBIO, Universidade de Vigo, Vigo 36310, Spain, [2]Departamento de Bioquímica, Genética e Inmunología, Universidade de Vigo, Vigo 36310, Spain, [3]Life and Health Sciences Research Institute, School of Medicine, University of Minho, Braga 4710-057, Portugal, [4]ICVS/3Bs—PT Government Associate Laboratory, Guimarães 4806-909, Portugal and [5]Galicia Sur Health Research Institute (IIS Galicia Sur), Vigo 36310, Spain
[†]https://orcid.org/0000-0002-0516-2717
*Corresponding author: E-mail: marenas@uvigo.es

Abstract

The evolution of structural proteins is generally constrained by the folding stability. However, little is known about the particular capacity of viral proteins to accommodate mutations that can potentially affect the protein stability and, in general, the evolution of the protein stability over time. As an illustrative model case, here, we investigated the evolution of the stability of the human immunodeficiency virus (HIV-1) protease (PR), which is a common HIV-1 drug target, under diverse evolutionary scenarios that include (1) intra-host virus evolution in a cohort of seventy-five patients sampled over time, (2) intra-host virus evolution sampled before and after specific PR-based treatments, and (3) inter-host evolution considering extant and ancestral (reconstructed) PR sequences from diverse HIV-1 subtypes. We also investigated the specific influence of currently known HIV-1 PR resistance mutations on the PR folding stability. We found that the HIV-1 PR stability fluctuated over time within a constant and wide range in any studied evolutionary scenario, accommodating multiple mutations that partially affected the stability while maintaining activity. We did not identify relationships between change of PR stability and diverse clinical parameters such as viral load, CD4[+] T-cell counts, and a surrogate of time from infection. Counterintuitively, we predicted that nearly half of the studied HIV-1 PR resistance mutations do not significantly decrease stability, which, together with compensatory mutations, would allow the protein to adapt without requiring dramatic stability changes. We conclude that the HIV-1 PR presents a wide structural plasticity to acquire molecular adaptations without affecting the overall evolution of stability.

Key words: protein evolution; HIV-1 protease; resistance mutations; protein folding stability; protease inhibitors

1. Introduction

After 40 years since the detection of the first human immunodeficiency virus (HIV) case, the HIV pandemic is far from over with around 1.5 million people acquiring HIV and 680,000 - deaths related with the acquired immunodeficiency syndrome (AIDS) in 2020 (UNAIDS 2021). Despite the global efforts, the development of an effective vaccine against HIV remains elusive (Cohen 2020; Ng'uni, Chasara, and Ndhlovu 2020), mainly due to the rapid HIV evolution (Cuevas et al. 2015) and tissue reservoirs of HIV-infected T-cells (Lorenzo-Redondo et al. 2016; Fromentin and Chomont 2020). Hence, the currently available antiretroviral therapy against this virus is based on a diverse repertoire of drugs that include chemokine receptor antagonisms, inhibitors of the viral fusion with the cellular membrane, and inhibitors of the viral replication machinery such as the reverse transcriptase, integrase, and protease (PR), among others (see, for a review, Arts and Hazuda 2012). Concerning the latter, protease inhibitors (PIs) constitute a well-established therapy with active research to design more effective and durable drugs (Arenas, Villaverde, and Sussman 2009; Ghosh, Osswald, and Prato 2016; Weber, Wang, and Harrison 2021). Still, drug resistance mutations (DRMs) can emerge (Fig. S1, Supplementary Data) (Wang and Kollman 2001; Obasa et al. 2020),

allowing the resistant viral populations to escape from the therapy (Arenas 2015; Pennings, Kryazhimskiy, and Wakeley 2014; Shah et al. 2020). Resistance mutations can also cause costs to the protein fitness by altering the protein stability (Laville et al. 2020) and the activity derived (Birolo et al. 2021). However, protein stability could be restored by the acquisition of compensatory or accessory mutations (Chang and Torbett 2011; Weikl and Hemmateenejad 2020; Moyano et al. 2022). Altogether, the evolution of stability of viral proteins is driven by contrasted evolutionary forces that include adaptation to changing environments (i.e. immune system and antiretroviral therapy) and selection on the protein function required for the virus replication. Then, for HIV-1 PR, the stability of the protein over time could remain constant (i.e. due to a strong selection pressure for maintaining the protein function, among others), increase (i.e. by structural hydrophobic adjustments fixed by selection), or decrease (i.e. by certain resistance mutations required as a consequence of an antiretroviral therapy). Olabode et al. (2017) found that the HIV-1 PR generally decreased its stability over time as a consequence of acquiring resistance mutations against therapies and suggested that maintaining an optimal protein structure is a major constraint factor in the evolution of this protein. This study provided a relevant

progress in the field but made assumptions that could affect the major conclusions. In particular, the study predicted the stability of ancestral protein sequences reconstructed under the general empirical substitution model WAG (Whelan and Goldman 2001), instead of applying an empirical substitution model based on HIV-1 proteins such as HIVb and HIVw (Nickle et al. 2007) that, as expected, usually better fit with the observed HIV-1 protein sequences (Nickle et al. 2007; Del Amparo and Arenas 2022a, 2022b). Moreover, it is well known that the ancestral sequences reconstructed under empirical substitution models of protein evolution are unrealistically unstable due to ignoring constraints from the protein structure and assuming that all the protein sites evolve under the same exchangeability matrix and amino acid frequencies (Arenas et al. 2017; Arenas and Bastolla 2019). Consequently, stability-constrained substitution (SCS) models, which outperform the traditional empirical substitution models by producing much higher phylogenetic likelihoods in terms of fitting with empirical data and more realistic folding stability (Bordner and Mittelmann 2014; Arenas, Sánchez-Cobos, and Bastolla 2015), should be considered to properly study protein evolution (see the review by Liberles et al. 2012). Additional technical limitations in the study by Olabode et al. (2017) are the prediction of the protein folding stability ignoring other structural conformations (i.e. an energy minimization) after acquiring mutations and neglecting the error of protein stability predictions. Finally, to evaluate the evolution of the protein stability over time, Olabode et al. (2017) used ancestral sequences reconstructed under certain substitution models of protein evolution. In this concern, an analysis based on real ancestral data could avoid biases caused by the modeling of protein evolution.

Here, we investigated the evolution of the HIV-1 PR folding stability under diverse evolutionary scenarios, including its possible relationships with clinical parameters of virus population dynamics. We analyzed both real data monitored over time and ancestral data reconstructed under SCS models and applied accurate protein stability prediction methods that consider flexibility of folding conformations and provide prediction errors. We found fluctuations of the HIV-1 PR stability over time but falling within a constant and wide interval, demonstrating that this protein is highly robust to accommodate a variety of mutations while maintaining viral fitness.

2. Material and methods

First, we investigated the intra-host evolution of HIV PR stability in a large number of patients monitored over time. We also evaluated HIV PR stability in sequences collected from the same patient before and after the treatment with diverse PIs. Next, we investigated long-term inter-host evolution of the HIV PR stability through phylogenetic reconstruction and the analysis of ancestral sequences. Finally, we explored the particular effect of observed HIV-1 PR DRMs, and their observed combinations, on the protein stability. The evolutionary analyses performed in this study are illustrated in Fig. 1.

2.1 Study data and evolutionary analyses

The analysis of intra-host HIV PR stability evolution was based on the data obtained from the records on HIV-1–infected individuals followed by the Specialized Assistance Services on Sexually Transmitted Diseases and HIV/AIDS in Brazil between 1 January 2008 and 30 April 2017 (Santos-Pereira et al. 2021; Souto et al. 2021). A total of seventy-five HIV-1–infected individuals (Supplementary Table S1) were selected for this study according

to the following inclusion criteria: (1) be under antiretroviral therapy and (2) present complete HIV PR sequences obtained at least at four different sampling times in the clinical follow-up. As a result, a total of 311 HIV PR sequences (including DRMs and other variants) obtained as part of the routine clinical testing were included in the study. For every sampling point, we also collected viral loads and CD4+ T-cell counts (Supplementary Table S1), which were considered for exploring their possible relationships with the PR stability by a correlation analysis. The collection of the data was performed anonymously after the approval by the Brazilian National Ethics Committee through the protocol Identification of the ethical approval (CAAE) 53757016.0.0000.5504. The sequences were aligned with the well-established framework *MUSCLE* (Edgar 2004) using settings by default. We also calculated the proportion of ambiguous sites (PAS, the observed number of ambiguous sites normalized by the sequence length), which is related to the time since the infection (Kouyos et al. 2011; Andersson et al. 2013), and evaluated its evolution over time and with respect to the PR stability. We applied a Wilcoxon signed-rank test (without and with *P* values corrected for multiple tests with *Hochberg* and *Bonferroni* methods) to evaluate the variation of the PR stability at different PAS levels.

We studied the variation of the HIV PR stability before and after the treatment with a particular PI (or combination of PIs) in data publicly available from the Stanford HIV-1 Drug Resistance Database (HIVdb) (Shafer 2006). Specifically, we analyzed a total of 135 patients with available HIV-1 PR sequences collected before and after the treatment with the PIs atazanavir, indinavir, lopinavir, nelfinavir, ritonavir (RTV), saquinavir (SQV) and RTV + SQV (Table S2; Supplementary Data). Again, we used the Wilcoxon signed-rank test to evaluate the variation of the PR stability as a consequence of applying every specific treatment.

The analysis of the long-term inter-host evolution of the HIV PR stability was also based on protein sequences available from the HIVdb. In particular, we built a dataset of fifty-three PR sequences belonging to different HIV-1 subtypes (A, B, C, D, F, G, and J) and sixteen PR HIV-2 sequences (Table S3; Supplementary Data). Of note, these sequences included DRMs and other variants. The sequences were aligned with *MUSCLE* under settings by default. Next, we identified the best-fitting substitution model of protein evolution under the Bayesian Information Criterion with *ProtTest3* (Darriba et al. 2011), and the selected substitution model was the empirical model *HIVb* (Nickle et al. 2007), considering the substitution rate variation among sites according to a gamma distribution +*G*. We inferred a maximum-likelihood phylogenetic tree under the selected substitution model with *RAxML-NG* (Kozlov et al. 2019). We rooted the phylogenetic tree considering the HIV-2 sequences as an outgroup. Next, we inferred ancestral sequences under the *mean-field* SCS model (Arenas, Sánchez-Cobos, and Bastolla 2015) implemented in the ancestral sequence reconstruction (ASR) framework *ProtASR2* (Arenas and Bastolla 2019). This ASR requires the multiple sequence alignment, the previously reconstructed phylogenetic tree, a representative protein structure (which was the selected structural template indicated in the next section), and the parameters of the structurally constrained substitution model that were specified by default following the developers' recommendations (Arenas and Bastolla 2019) (for further details, see the material about ASR provided in Data Availability). Finally, we also applied the Wilcoxon signed-rank test to evaluate the evolution of the PR stability over time in the studied evolutionary history.

Finally, we analyzed the particular effect of all the DRMs, and their combinations, observed in sequences from the HIVdb on the
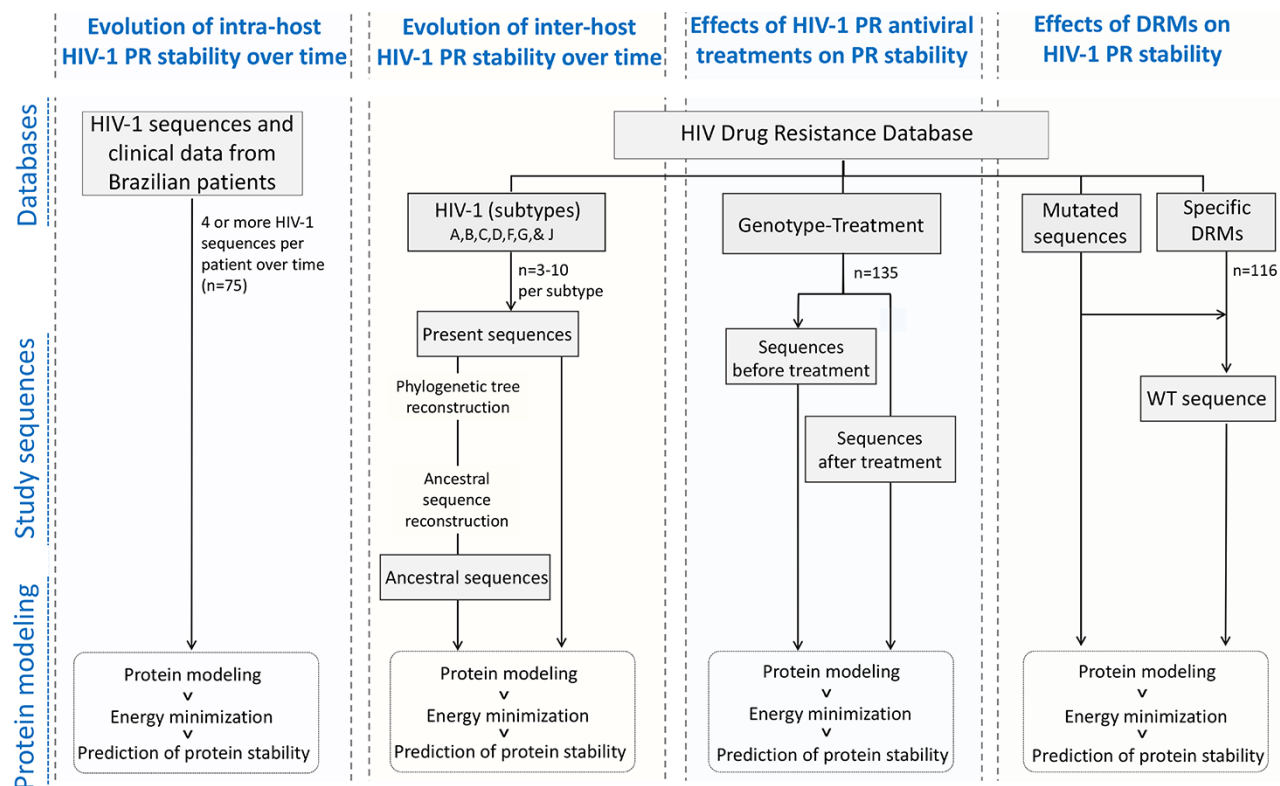
Figure 1. Evolutionary scenarios and pipeline of methodologies applied in the present study. From top to bottom, it shows the databases, the evolutionary methods, and the protein modeling.

HIV PR folding stability. As a reference, we considered a consensus sequence (also available from the HIVdb) that does not include any known DRM (hereafter wild-type [WT] sequence). Next, the observed DRMs were incorporated into the WT sequence to obtain resistance sequences that differ from the WT sequence by one or more DRMs observed in the HIVdb, which allows evaluating the specific consequences of every DRM or combination of DRMs. We analyzed a total of 116 single and combined DRMs (Table S4; Supplementary Data).

2.2 Protein modeling and prediction of protein folding stability

For every HIV PR sequence presented in the previous section, we obtained a protein structure and predicted its folding stability (Supplementary Fig. S1).

First, we obtained protein structures from protein sequences by homology modeling (Sali and Blundell 1993; Muhammed and Aki-Yalcin 2019), taking advantage that a large number of HIV PR structures are deposited in the Protein Data Bank (PDB). We identified the best-fitting structural template [without ligand to avoid possible biases from ligand-induced structural conformations (Huang et al. 2012; Sartori et al. 2019)] for the WT sequence with *SWISS-MODEL* (Arnold et al. 2006), selecting the structure with PDB code 3IXO (Robbins et al. 2010). Conveniently, this template was used to predict the structure in all the PR sequences, avoiding possible stability prediction biases that could be caused by specifying different templates and also to replace possible indels that cannot be treated by the predictor of protein folding stability (see later). Next, protein structures (models) were obtained with *MODELLER* using the standard homology modeling (Sali and Blundell 1993), and we selected the model with the highest fitting score (Eramian et al. 2008). For the specific

analysis of PR stability with every observed DRM (or combination of DRMs; Supplementary Table S4), we obtained a PR structural model for every studied DRM (or combination of DRMs) from the WT structure using *FoldX* (Guerois, Nielsen, and Serrano 2002).

Next, we predicted the protein folding stability for every structure. In order to remove steric clashes and constrained bonding conditions that could bias predictions of protein stability, the modeled structures were subjected to an energy minimization with *GROMACS v2021.4* following the standard protocol (Scott et al. 1999). For each modeled structure, the protein topology was generated with the '*pdb2gmx*' function under the *AMBER99* force field. We defined simulation cubic boxes (unit cells with size $10 \times 10 \times 10$ Å) and filled the system with water. The system was maintained with periodic boundary conditions ensuring a minimum of 10 Å distance between any protein atom and boundaries. The box was then solvated using the simple point charge (SPC)216 water model (Berendsen et al. 1981). Note that the SPC model defines a water molecule as a rigid three-site molecule, and each atom is assigned a charge and *Lennard–Jones* parameter (Berendsen et al. 1981). Next, we added counter ions to neutralize the charge of the system. Finally, we minimized the energy of the system with the well-established steepest descent algorithm (Abrudan, Eriksson, and Koivunen 2008). The applied number of iterations was based on reaching convergence among three independent runs, converging when all of them presented a maximum force below 1,000 kJ/mol/nm. The difference in the atomic coordinates by the root-mean-square deviation (RMSD) was also used to evaluate convergence among runs (RMSD <0.5 Å among the resulting structures was considered as converged).

Finally, we applied *FoldX* (Guerois, Nielsen, and Serrano 2002) to predict the thermodynamic folding stability (Gibbs free energy, ΔG) of the resulting HIV PR structures. Note that the lower the

Gibbs free energy, the higher is the stability. This framework predicts the folding stability (ΔG) with the *FOLDEF* empirical energy function that considers *van der Waals*, solvation, hydrogen bond, electrostatic, and entropy contributions, among other parameters (Guerois, Nielsen, and Serrano 2002), and it is frequently used to study protein folding stability (i.e. Olabode et al. 2017).

3. Results

3.1 Temporal fluctuations of intra-host HIV-1 PR folding stability within a constant range

The evolution of the HIV-1 PR stability in the HIV-1–infected individuals followed in Brazil showed fluctuations over time (Fig. 2A). The observed evolutionary trajectories were both stabilizing and destabilizing, and, in general, they fell within a wide range statistically constant over time (Fig. 2B; Wilcoxon signed-rank with P always >0.05, indicating a lack of significant variation of protein stability over time). Most of the PR sequences presented stability (ΔG) within 45–105 kcal/mol (thus falling within an interval of 60–75 kcal/mol) with a mean of 74.82 ± 1.36 kcal/mol that was overall maintained over time (Fig. 2B). Next, we identified a lack of correlation of PR stability with CD4$^+$ T-cell counts (*Pearson correlation $r = 0.001$ with $P = 0.988$*) and with viral load (*Pearson correlation $r = 0.02$ with $P = 0.392$*) (Fig. S2; Supplementary Data). In addition, the predicted PR stability did not present a significant variation with the observed PAS levels (Fig. S3; Supplementary Data; Wilcoxon signed-rank with $P > 0.05$).

3.2 The HIV PR folding stability can be affected toward any direction by PIs but within a constant range

For the studied patients, the stability of HIV PR variants collected before and after the treatment with a particular PI (or combination of PIs) was maintained (20.74 per cent of patients), decreased (43.70 per cent of patients), or increased (35.56 per cent of patients), and thus, all the trends were observed (Fig. 3A). Moreover, for any particular treatment, we found all these possible variations of stability (Fig. 3A). However, again these stability changes occurred within a constant stability range (Fig. 3A). Actually, if the data are grouped by patients receiving every treatment, we did not find statistical relationships between PR stability before and after receiving every treatment (Wilcoxon signed-rank; $P > 0.05$) (Fig. 3B).

3.3 Temporal fluctuations of inter-host HIV-1 PR folding stability within a constant range

We analyzed the evolution of the HIV-1 PR stability over time along a phylogenetic tree inferred from a sample of PR sequences from all the HIV-1 subtypes (rooted with HIV-2 sequences), considering ancestral sequences reconstructed under realistic SCS models of protein evolution. The results also showed temporal fluctuations of the PR folding stability over time (Fig. 4A), including stabilizing and destabilizing evolutionary trajectories, within a wide range. Indeed, in agreement with the previous results for intra-host virus evolution, the overall protein stability was constant over time (Fig. 4B). In particular, protein sequences belonging to the present and to the past (reconstructed) produced a ΔG mean of 68.22 ± 1.96 kcal/mol (95 per cent confidence interval), while only ancestral proteins or only contemporary proteins showed a ΔG mean of 67.5 ± 2.68 kcal/mol and 68.94 ± 2.87 kcal/mol, respectively (note that these estimates overlap).

3.4 Diverse consequences of observed resistance mutations (and their combinations) on the HIV-1 PR folding stability

In the previous sections, the results showed fluctuations of the HIV-1 PR stability over time under any studied evolutionary scenarios, but also those fluctuations fell within a range that was maintained over time even if PIs are applied. In order to understand the causes of the evolution of the HIV-1 PR stability delimited within a constant range, we investigated the influence of every main DRM (and combination of main DRMs) observed in HIV PR sequences of the HIVdb. Protein variants with DRMs showed a folding stability significantly similar (41.74 per cent), lower (more unstable, 56.52 per cent), and higher (more stable, 1.74 per cent) than the stability of the WT (Figs. 5 and S4 Supplementary Data; Table S4). These results also indicated that a single DRM can maintain, stabilize, or destabilize the HIV-1 PR. In other words, a single DRM does not necessarily produce a low effect on stability. Also, multiple DRMs do not necessarily cause a high effect on stability. Instead, the influence of DRMs on the PR stability depended on the nature of the involved mutations rather than on the number of DRMs. For example, an interesting comparison between single DRMs (from the perspective of derived PR stability) involves I54V and I54L, note that both affect the same position. In comparison with the stability of the WT, the former mutation stabilizes the PR, while the latter destabilizes it. We speculate that this destabilization could be caused by the disruption of the hydrophobic packing upon P79 mediated by I54 but not by L54 (Fig. S5; Supplementary Data), thus producing a reduced hydrophobic sliding that is essential for flap closing due to a reduced number of *van der Waals* interactions (Henes et al. 2019; Leidner, Kurt Yilmaz, and Schiffer 2021).

4. Discussion

Protein evolution is fundamental to understand the past and the future of organisms (Pál, Papp, and Lercher 2006; von Heijne 2018). Concerning antiretroviral therapies, taking into account the evolution of viral proteins that are used as molecular drug targets, can be crucial to design efficient therapies (Zhang et al. 2010; Rhee et al. 2015). Protein evolution involves not only sequence evolution but also the evolution of the protein structure (Liberles et al. 2012; Wilke 2012). In structural proteins, most of the possible (random) mutations can destabilize the protein structure (Tokuriki et al. 2007). However, it is also known that the protein structure is more conserved than the protein sequence (Illergård, Ardell, and Elofsson 2009; Pascual-García et al. 2010; Pascual-García, Arenas, and Bastolla 2019); actually, this is the basis of well-established approaches like homology modeling (Sali and Blundell 1993; Muhammed and Aki-Yalcin 2019). Therefore, although a number of mutations can destabilize the protein structure and affect the protein function, they are not usually observed due to selection. Still, many mutations could be neutral (i.e. involving small structural changes) and frequently observed. In this concern, an interesting case is the evolution of protein drug targets of pathogens. Proteins from successful pathogens must acquire resistance mutations to escape from the host immune system and therapies. However, the evolved protein variants should display a sufficient stability to maintain the protein function required for pathogen replication (Stella-Ascariz et al. 2017; Wensing et al. 2019). The acquired DRMs allow the pathogen to adapt and survive to the new environment (Arenas 2015; Poon et al. 2007), but despite maintaining fundamental functionalities, DRMs could
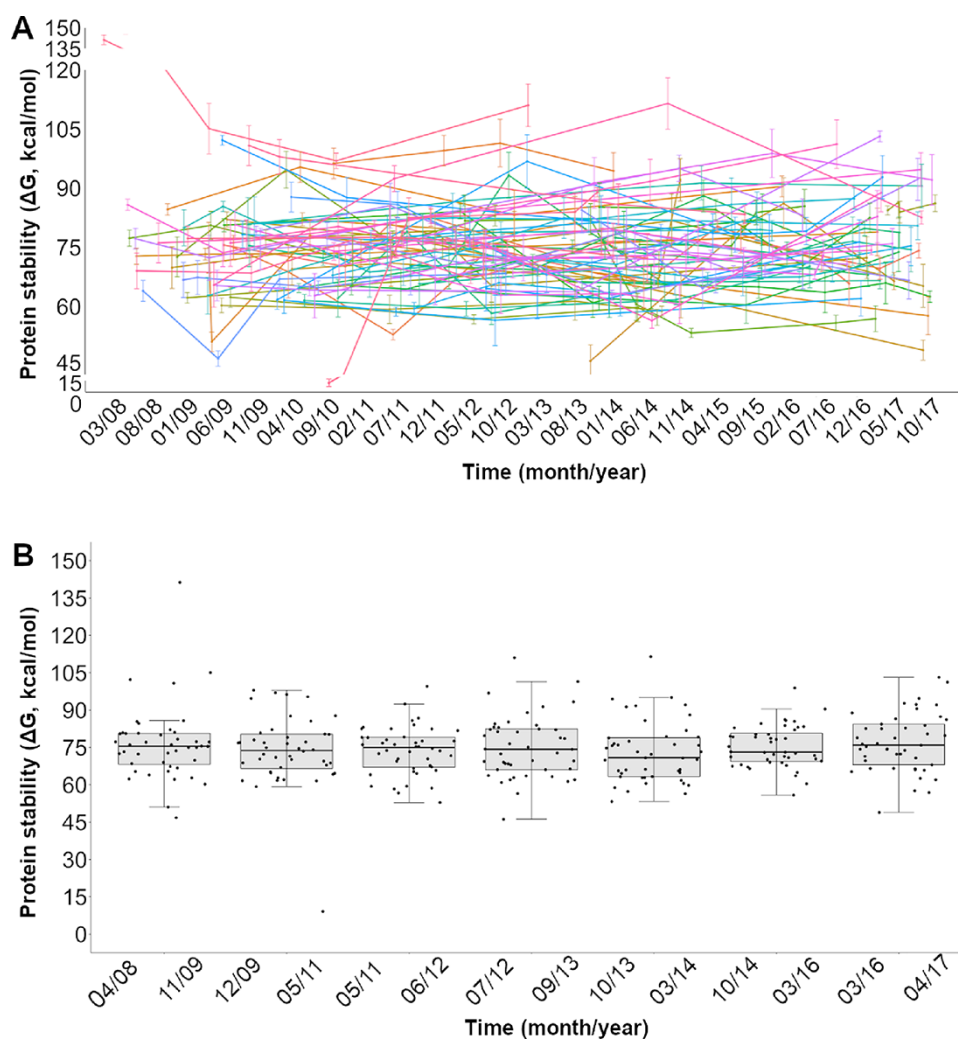
Figure 2. Evolution of intra-host HIV-1 PR folding stability over time. Evolution of the HIV-1 PR folding stability (ΔG) over time in the cohort of Brazilian patients. (A) Every patient is shown with a color and every dot belongs to a sample. Every sample includes the ΔG mean predicted from independent runs (see Section 2), and the error bars indicate the 95 per cent confident interval of the mean. (B) Evolution of the predicted folding stabilities over time distributed in several bins (boxplots) with same sample size (the data are shown with dots).

reduce the fitness of the protein (i.e. by reducing the viral replication capacity) (Melnyk, Wong, and Kassen 2015; Trompet et al. 2020). These possible negative effects of DRMs on the protein fitness could partially be removed by the incorporation of compensatory (accessory) mutations (González-Ortega et al. 2011; Wood et al. 2013; Weikl and Hemmateenejad 2020). However, the specific consequences of DRMs (and combination of DRMs) on the stability of multiple protein drug targets are not yet totally clear. Moreover, although the sequence evolution of protein drug targets was widely investigated (Yang et al. 2000; Castro-Nallar et al. 2012; Del Amparo and Arenas 2022a), little is known about the evolution of the stability of protein drug targets. For example, could the administration of antiretroviral therapies produce a recent decrease of stability of the HIV protein drug targets in the long-term evolutionary history of these proteins? To our knowledge, only Olabode et al. (2017) comprehensively studied this aspect by exploring the evolution of the stability in some HIV protein drug targets. Interestingly, they predicted fluctuations of the stability over time, which are in agreement with our results (discussed later). They also found an overall decrease of stability over time and proposed that it could be caused by the use of antiretroviral therapies. However, these conclusions could be influenced by methodological

limitations. First, that study analyzed the evolution of protein stability, considering reconstructed ancestral proteins (to compare the stability of contemporary and ancestral proteins) and did not evaluate ancient real data (i.e. serial samples derived from monitoring virus evolution over time), and consequently, the results are highly influenced by the realism of the used ASR. In particular, they applied an empirical substitution model of protein evolution to reconstruct ancestral sequences, and it is known that empirical substitution models can be biased toward producing proteins with unrealistic stabilities (Arenas et al. 2017; Arenas and Bastolla 2019). In addition, the study predicted protein stability without relaxing the protein conformation (i.e. by an energy minimization or molecular dynamics analysis) and did not consider stability prediction errors, leading to a lack of statistical evaluation of predictions. Thus, to advance in this topic, here, we investigated the evolution of the HIV-1 PR folding stability under diverse evolutionary scenarios, including (1) serial real data derived from patients monitored over time, (2) real data collected before and after the treatment with diverse PIs, (3) the evolution of stability along the PR HIV phylogeny based on sequences from different subtypes and sequences reconstructed considering substitution models of evolution that account for the selection on the protein
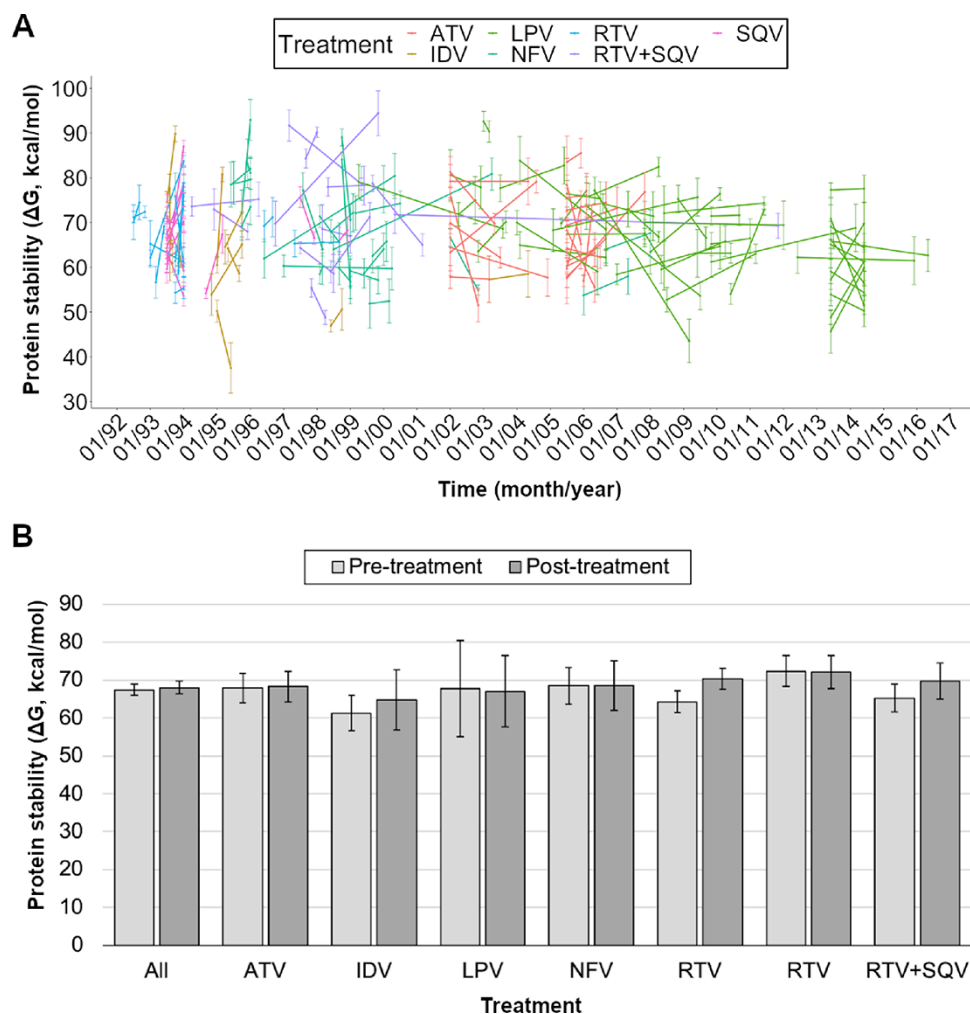
Figure 3. Influence of PR inhibitors on the HIV-1 PR stability. The plots show the PR folding stability (ΔG) of extant protein variants collected before and after a particular PI treatment (Supplementary Table S2). (A) Every patient is represented with a line with two samples (before and after treatment), and every treatment is represented with a color. Every sample includes the ΔG mean predicted from independent runs (see Section 2), and the error bars indicate the 95 per cent confident interval of the mean. (B) The predicted folding stability by groups of samples collected before and after every particular treatment. Error bars indicate 95 per cent confident interval of the mean of predictions from the protein variants of the corresponding group.

structure and that outperformed empirical substitution models (Bordner and Mittelmann 2014; Arenas, Sánchez-Cobos, and Bastolla 2015; Arenas and Bastolla 2019), and finally, (4) the specific consequences on stability of the DRMs (including their combinations) observed in data from the HIVdb. In addition, we predicted the protein folding stability allowing flexibility of the structural conformation to realistically allocate mutations, and we included stability prediction errors. We believe that accounting for these aspects can provide a more realistic and accurate view of the evolution of the HIV PR, which is one of the most important molecular targets of HIV antiretroviral therapies.

We found fluctuations of HIV-1 PR stability over time, which are in agreement with Olabode et al. (2017). However, in contrast to Olabode et al. (2017) about the proposal of an overall decrease of HIV-1 PR stability over time, we found an overall maintenance of the PR stability over time. We also found that a large proportion of DRMs are not dramatic in terms of decreasing stability, and we believe that this prediction, together with the acquisition of compensatory mutations (González-Ortega et al. 2011; Wood et al. 2013; Weikl and Hemmateenejad 2020), could explain the overall constant evolution of the HIV PR stability over time. Next, we discuss the results derived from every analysis.

First, we investigated the intra-host evolution of the PR folding stability in seventy-five HIV-1–infected individuals with at least four samples collected at different times during the clinical follow-up. The results showed fluctuations of the PR folding stability over time, in agreement with Oladobe et al. (2017). These fluctuations fell within a range of stability, which suggests limits of stability to accept mutations (Zeldovich, Chen, and Shakhnovich 2007; Serohijos and Shakhnovich 2014; Shah et al. 2020). Thus, the HIV-1 PR presented a large flexibility to accommodate the wide variety of observed mutations, suggesting a marginal stability behavior of this protein (Taverna and Goldstein 2002). Next, mutations producing variants that fall out of that range of marginal stability could be selected against. The large capacity of the HIV-1 PR to accommodate mutations was also observed in other studies. For example, Wu et al. (2003) observed mutations at sixty-two of the ninety-nine amino acid positions of the HIV-1 PR, and even mutations in all the HIV-1 PR positions were observed (Zhang et al. 2010). Here, we showed that the folding stability of the accepted PR variants, including those with DRMs (discussed later in detail), falls within a particular range. If all the patients are considered together, which is convenient to identify an inter-patient trend, we found a significant maintenance of the PR folding stability
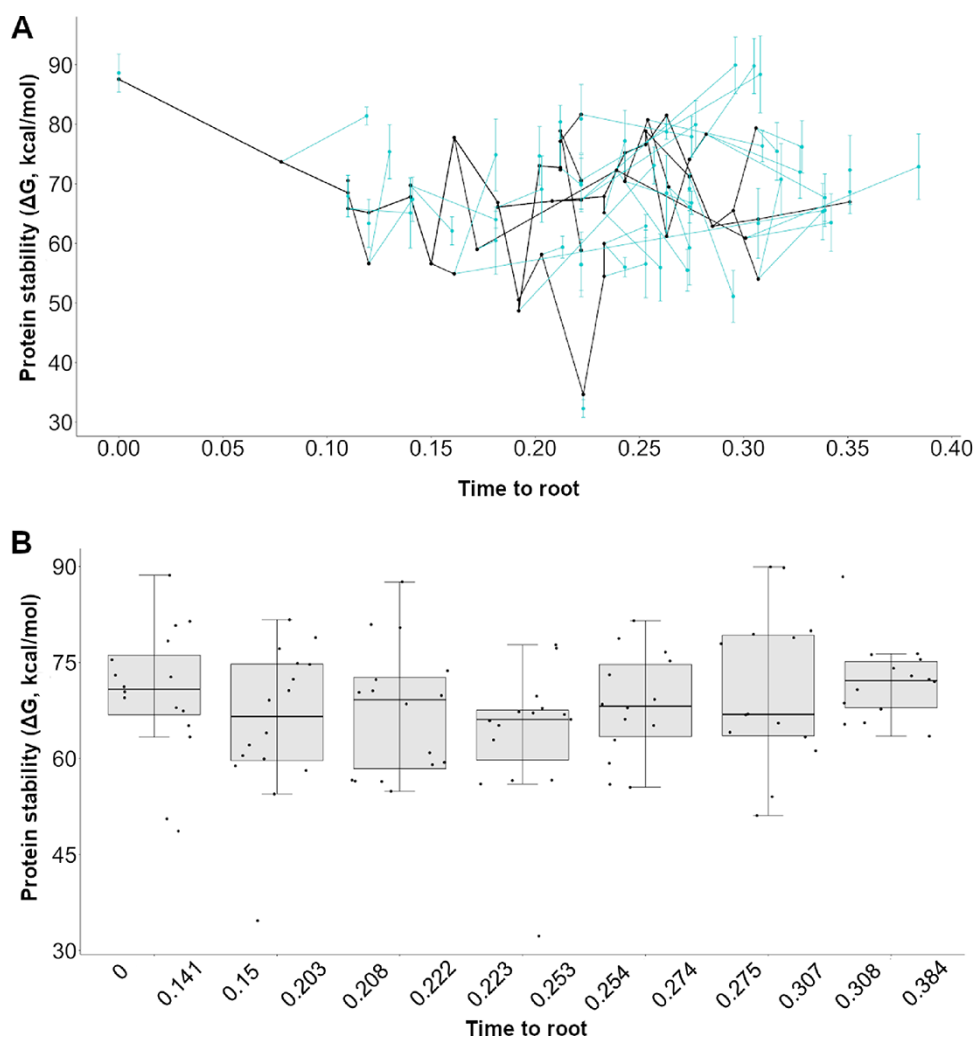
Figure 4. Evolution of inter-host HIV-1 PR folding stability over time. Evolution of the folding stability (ΔG) over time, from the root to the present, on a phylogenetic tree based on sequences from all the HIV-1 subtypes (Supplementary Table S3). (A) Black lines correspond to the internal branches of the phylogenetic tree, while clear lines correspond to the terminal branches. Every sample includes the ΔG mean predicted from independent runs (see Section 2), and the error bars indicate the 95 per cent confident interval of the mean. (B) Evolution of the predicted folding stabilities over time distributed in several bins (boxplots) with same sample size (the data are shown with dots).

over time. This finding, together with other results from our study (discussed later), suggests that the HIV-1 PR tends to maintain stability over time. In this concern, Olabode et al. (2017) proposed that the stability of HIV PR decreases over time due to the fixation of DRMs. We believe that the finding could be caused by applying an unrealistic empirical substitution model of protein evolution in the ASR (Arenas et al. 2017; Arenas and Bastolla 2019). Note that empirical substitution models of protein evolution assume a same exchangeability matrix for all the protein sites and ignore structural constraints, leading to ancestral protein sequences with unrealistic folding instabilities (Arenas et al. 2017; Arenas and Bastolla 2019). Finally, we did not find relationships between the time from the infection and the predicted protein folding stability, which additionally supports our previous findings about the overall maintenance of the HIV-1 PR stability over time with fluctuations within a range. We also found a lack of statistical relationships between the HIV-1 PR stability and parameters of the viral population dynamics, including the viral load and CD4+ T-cell count. These results suggest that HIV-1 PR variants with stability falling within the identified range can maintain activity and do not affect the global virus fitness.

PIs are the basis of multiple common therapies against HIV, but, unfortunately, a variety of DRMs can emerge and are associated with one or more PIs (Rhee et al. 2005; Wensing et al. 2019). It is important to note that PI-associated DRMs can differ among PIs (Shafer 2006). Considering that different PIs can promote different selection pressures (Arenas 2015; Poon et al. 2007), we compared the HIV-1 PR folding stability in samples from the same patient collected before and after the treatment with a particular PI (or combination of PIs). If one considers that the treatment favors the fixation of certain associated DRMs, then the sample collected after the treatment should present more DRMs than the sample collected before the treatment, and thus, if DRMs alter the folding stability, an effect on the folding stability should be observed as a consequence of the viral adaptation to the treatment. We found that the evolution of the HIV-1 PR stability differs among patients, observing patients with maintenance, decrease, and increase of stability over time. However, considering all the patients together for evaluating global patterns, PI administration did not result in folding stability changes, a finding that is in agreement with Olabode et al. (2017). We consider that these results also support the flexibility of the HIV PR to accommodate DRMs
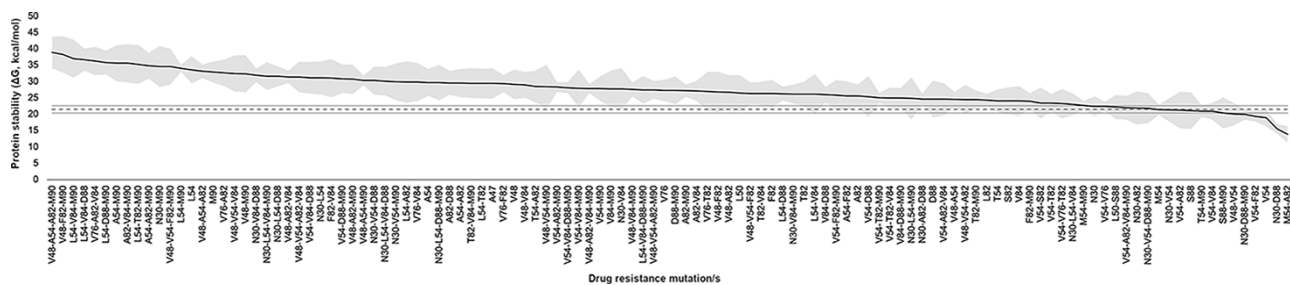
Figure 5. Influence of the observed DRMs on the HIV-1 PR folding stability. Mean of the predicted protein folding stability (ΔG) from independent runs for every main DRM (and combination of DRMs) observed in the HIVdb. DRMs are ordered from the highest to the lowest (most unstable to most stable) mean of stability. The light gray area indicates the 95 per cent confident interval of the mean from the independent runs. The dashed line corresponds to the WT protein variant (without any DRM), and its top and bottom dark gray lines indicate the 95 per cent confident interval of the mean from the independent runs. The predicted stabilities are also included in Supplementary Table S4. Supplementary Fig. S6 shows the same results but ordered by the minimum folding stability of the confidence interval.

within a 'safe' interval, leading to a global maintenance of HIV PR stability during its evolution.

We also investigated the evolution of the HIV PR stability at a longer timescale, considering the evolutionary history of protein variants from all the HIV-1 subtypes. The idea of this study was similar to that performed by Olabode et al. (2017), with the difference that we performed ASR based on a SCS model of protein evolution, which is much more realistic than any empirical substitution model in terms of likelihood (Bordner and Mittelmann 2014; Arenas, Sánchez-Cobos, and Bastolla 2015) and folding stability (Arenas et al. 2017; Arenas and Bastolla 2019). Again, we found fluctuations of the HIV PR folding stability over time that agree with the findings by Olabode et al. (2017). However, we found a global maintenance of HIV PR folding stability since the onset of the studied evolutionary history, in contrast to Olabode et al. (2017) that found a decrease of stability with time. As previously discussed, we believe that this discrepancy is derived from the substitution models applied in the ASR [Olabode et al. (2017) applied an empirical substitution model that is not directly related to HIV (Whelan and Goldman 2001), which can affect ASR (Del Amparo and Arenas 2022b), and that makes assumptions such as a same exchangeability matrix for all the protein sites that can also affect ASR (Arenas et al. 2017; Arenas and Bastolla 2019)]. By contrast, we applied a substitution model that considers structural constraints from the HIV-1 PR and exchangeability matrices differing among protein sites and that outperforms empirical substitution models (Arenas, Sánchez-Cobos, and Bastolla 2015; Arenas and Bastolla 2019). The results again suggest that the HIV-1 PR is highly flexible to acquire multiple mutations with stability fluctuating within a 'safe' range that was overall constant over time along the virus evolutionary history.

Since our previous results indicated temporal fluctuations of the HIV-1 PR stability within a wide and constant range, we aimed to understand the causes of these fluctuations, especially concerning the effects of DRMs as suggested by Olabode et al. (2017). Thus, we studied the HIV-1 PR folding stability under every DRM (and combination of DRMs) observed in the PR sequences of the HIVdb and compared them with the stability of the WT sequence as a reference. We found that more than 40 per cent of the studied DRMs contribute to maintain or increase the PR folding stability compared to the WT, while the remaining DRMs significantly destabilize the protein to a greater or lesser extent. Of course, some HIV-1 PR resistance variants are highly destabilizing, but they are a minority. However, in general, the findings suggest that the influence of DRMs on the HIV-1 PR folding stability does not seem as dramatic as we originally expected. We previously

found that the HIV-1 PR folding stability in virus populations of patients with therapy failure is not related to the global viral population dynamics, which could suggest that the observed DRMs in the HIV-1 PR can be accommodated by the large flexibility of this protein. Finally, we found that the influence of DRMs on the protein folding stability is highly affected by the amino acids involved in the mutation. This is hardly surprising that more than 50 years ago Zuckerkandl and Pauling (1965) indicated that 'it is the type rather than number of amino acid substitutions that is decisive', and we believe that this is also the case for HIV PR DRMs.

We conclude that the observed HIV PR stability can vary over time, displaying both stabilizing and destabilizing evolutionary trajectories within a stability range, probably established by the required protein activity. In our opinion, DRMs should not be considered as a kind of 'artificial' mutations associated with treatments. These mutations are spontaneous and can also occur in naive patients (Parisi et al. 2007). In the presence of a therapy, their frequency in the virus population can increase due to the selection imposed by the treatment. Our findings indicate that the HIV PR can fix a wide variety of DRMs by its large structural flexibility, with multiple DRMs falling within levels that maintain the protein activity. Here, we focused on observed protein variants. Other PR variants may not be observed due to producing dramatic stability and/or activity alterations and thus could be removed by natural selection (Tokuriki et al. 2007). Their study would require complex computer simulations accounting for the population genetics and evolution of the virus.

Data availability

The viral sequences collected from Brazilian patients were deposited in GenBank with accession codes from ON982820.1 to ON983130.1. The accession codes of the sequences collected from the HIVdb are provided in the Supplementary Tables S2 and S3. All the studied sequences (including those from the Brazilian patients), protein structures, and ASR files are available at Zenodo repository from the URL https://doi.org/10.5281/zenodo.7153503.

Supplementary data

Supplementary data are available at *Virus Evolution* online.

Funding

References

Abrudan, T. E., Eriksson, J. and Koivunen, V. (2008) 'Steepest Descent Algorithms for Optimization under Unitary Matrix Constraint', *IEEE Transactions on Signal Processing*, 56: 1134–47.

Andersson, E. et al. (2013) 'Evaluation of Sequence Ambiguities of the HIV-1 Pol Gene as a Method to Identify Recent HIV-1 Infection in Transmitted Drug Resistance Surveys', *Infection, Genetics and Evolution*, 18: 125–31.

Arenas, M. (2015) 'Genetic Consequences of Antiviral Therapy on HIV-1', *Computational and Mathematical Methods in Medicine*, 2015: 1–9.

Arenas, M. et al. (2017) 'ProtASR: An Evolutionary Framework for Ancestral Protein Reconstruction with Selection on Folding Stability', *Systematic Biology*, 66: 1054–64.

Arenas, M. and Bastolla, U. (2019) 'ProtASR2: Ancestral Reconstruction of Protein Sequences Accounting for Folding Stability', *Methods in Ecology and Evolution*, 11: 248–57.

Arenas, M., Sánchez-Cobos, A. and Bastolla, U. (2015) 'Maximum-Likelihood Phylogenetic Inference with Selection on Protein Folding Stability', *Molecular Biology and Evolution*, 32: 2195–207.

Arenas, M., Villaverde, M. C. and Sussman, F. (2009) 'Prediction and Analysis of Binding Affinities for Chemically Diverse HIV-1 PR Inhibitors by the Modified SAFE_p Approach', *Journal of Computational Chemistry*, 30: 1229–40.

Arnold, K. et al. (2006) 'The SWISS-MODEL Workspace: A Web-based Environment for Protein Structure Homology Modelling', *Bioinformatics*, 22: 195–201.

Arts, E. J. and Hazuda, D. J. (2012) 'HIV-1 Antiretroviral Drug Therapy', *Cold Spring Harbor Perspectives in Medicine*, 2: a007161.

Berendsen, H. J. C. et al. (1981) 'Interaction Models for Water in Relation to Protein Hydration'. In: Pullman, B. (ed.) *Intermolecular Forces*, 14, pp. 331–42. Dordrecht: Springer Netherlands.

Birolo, G. et al. (2021) 'Protein Stability Perturbation Contributes to the Loss of Function in Haploinsufficient Genes', *Frontiers in Molecular Biosciences*, 8: 620793.

Bordner, A. J. and Mittelmann, H. D. (2014) 'A New Formulation of Protein Evolutionary Models That Account for Structural Constraints', *Molecular Biology and Evolution*, 31: 736–49.

Castro-Nallar, E. et al. (2012) 'The Evolution of HIV: Inferences Using Phylogenetics', *Molecular Phylogenetics and Evolution*, 62: 777–92.

Chang, M. W. and Torbett, B. E. (2011) 'Accessory Mutations Maintain Stability in Drug-Resistant HIV-1 Protease', *Journal of Molecular Biology*, 410: 756–60.

Cohen, J. (2020) 'Combo of Two HIV Vaccines Fails Its Big Test', *Science*, 367: 611–2.

Cuevas, J. M. et al. (2015) 'Extremely High Mutation Rate of HIV-1 In Vivo', *PLoS Biology*, 13: e1002251.

Darriba, D. et al. (2011) 'ProtTest 3: Fast Selection of Best-Fit Models of Protein Evolution', *Bioinformatics*, 27: 1164–5.

Del Amparo, R. and Arenas, M. (2022a) 'HIV Protease and Integrase Empirical Substitution Models of Evolution: Protein-Specific Models Outperform Generalist Models', *Genes*, 13: 61.

—— (2022b) 'Consequences of Substitution Model Selection on Protein Ancestral Sequence Reconstruction', *Molecular Biology and Evolution*, 39: msac144.

Edgar, R. C. (2004) 'MUSCLE: Multiple Sequence Alignment with High Accuracy and High Throughput', *Nucleic Acids Research*, 32: 1792–7.

Eramian, D. et al. (2008) 'How Well Can the Accuracy of Comparative Protein Structure Models Be Predicted?', *Protein Science*, 17: 1881–93.

Fromentin, R. and Chomont, N. (2020) 'HIV Persistence in Subsets of CD4+ T Cells: 50 Shades of Reservoirs', *Seminars in Immunology*, 51: 101438.

Ghosh, A. K., Osswald, H. L. and Prato, G. (2016) 'Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS', *Journal of Medicinal Chemistry*, 59: 5172–208.

González-Ortega, E. et al. (2011) 'Compensatory Mutations Rescue the Virus Replicative Capacity of VIRIP-Resistant HIV-1', *Antiviral Research*, 92: 479–83.

Guerois, R., Nielsen, J. E., and Serrano, L. (2002) 'Predicting Changes in the Stability of Proteins and Protein Complexes: A Study of More Than 1000 Mutations', *Journal of Molecular Biology*, 320: 369–87.

Henes, M. et al. (2019) 'Picomolar to Micromolar: Elucidating the Role of Distal Mutations in HIV-1 Protease in Conferring Drug Resistance', *ACS Chemical Biology*, 14: 2441–52.

Huang, X. et al. (2012) 'Inhibitor-Induced Conformational Shifts and Ligand-Exchange Dynamics for HIV-1 Protease Measured by Pulsed EPR and NMR Spectroscopy', *The Journal of Physical Chemistry B*, 116: 14235–44.

Illergård, K., Ardell, D. H., and Elofsson, A. (2009) 'Structure Is Three to Ten Times More Conserved Than Sequence—A Study of Structural Response in Protein Cores', *Proteins: Structure, Function, and Bioinformatics*, 77: 499–508.

Kouyos, R. D. et al. (2011) 'Ambiguous Nucleotide Calls from Population-Based Sequencing of HIV-1 Are a Marker for Viral Diversity and the Age of Infection', *Clinical Infectious Diseases*, 52: 532–9.

Kozlov, A. M. et al. (2019) 'RAxML-NG: A Fast, Scalable and User-Friendly Tool for Maximum Likelihood Phylogenetic Inference', *Bioinformatics*, 35: 4453–5.

Laville, P. et al. (2020) 'Impacts of Drug Resistance Mutations on the Structural Asymmetry of the HIV-2 Protease', *BMC Molecular and Cell Biology*, 21: 46.

Leidner, F., Kurt Yilmaz, N. and Schiffer, C. A. (2021) 'Deciphering Complex Mechanisms of Resistance and Loss of Potency through Coupled Molecular Dynamics and Machine Learning', *Journal of Chemical Theory and Computation*, 17: 2054–64.

Liberles, D. A. et al. (2012) 'The Interface of Protein Structure, Protein Biophysics, and Molecular Evolution', *Protein Science*, 21: 769–85.

Lorenzo-Redondo, R. et al. (2016) 'Persistent HIV-1 Replication Maintains the Tissue Reservoir during Therapy', *Nature*, 530: 51–6.

Melnyk, A. H., Wong, A. and Kassen, R. (2015) 'The Fitness Costs of Antibiotic Resistance Mutations', *Evolutionary Applications*, 8: 273–83.

Moyano, A. et al. (2022) 'Immunoescape of HIV-1 in Env-EL9 CD8 + T Cell Response Restricted by HLA-B*14:02 in a Non Progressor Who Lost Twenty-Seven Years of HIV-1 Control', *Retrovirology*, 19: 6.

Muhammed, M. T. and Aki-Yalcin, E. (2019) 'Homology Modeling in Drug Discovery: Overview, Current Applications, and Future Perspectives', *Chemical Biology & Drug Design*, 93: 12–20.

Ng'uni, T., Chasara, C. and Ndhlovu, Z. M. (2020) 'Major Scientific Hurdles in HIV Vaccine Development: Historical Perspective and Future Directions', *Frontiers in Immunology*, 11: 590780.

Nickle, D. C. et al. (2007) 'HIV-Specific Probabilistic Models of Protein Evolution', *PLoS One*, 2: e503.

Obasa, A. E. et al. (2020) 'Drug Resistance Mutations against Protease, Reverse Transcriptase and Integrase Inhibitors in People Living with HIV-1 Receiving Boosted Protease Inhibitors in South Africa', *Frontiers in Microbiology*, 11: 438.

Olabode, A. S. et al. (2017) 'Adaptive HIV-1 Evolutionary Trajectories Are Constrained by Protein Stability', *Virus Evolution*, 3: vex019.

Pál, C., Papp, B. and Lercher, M. J. (2006) 'An Integrated View of Protein Evolution', *Nature Reviews Genetics*, 7: 337–48.

Parisi, S. G. et al. (2007) 'Both Human Immunodeficiency Virus Cellular DNA Sequencing and Plasma RNA Sequencing Are Useful for Detection of Drug Resistance Mutations in Blood Samples from Antiretroviral-Drug-Naive Patients', *Journal of Clinical Microbiology*, 45: 1783–8.

Pascual-García, A. et al. (2010) 'Quantifying the Evolutionary Divergence of Protein Structures: The Role of Function Change and Function Conservation', *Proteins: Structure, Function, and Bioinformatics*, 78: 181–96.

Pascual-García, A., Arenas, M., and Bastolla, U. (2019) 'The Molecular Clock in the Evolution of Protein Structures', *Systematic Biology*, 68: 987–1002.

Pennings, P. S., Kryazhimskiy, S. and Wakeley, J. (2014) 'Loss and Recovery of Genetic Diversity in Adapting Populations of HIV', *PLoS Genetics*, 10: e1004000.

Poon, A. F. Y. et al. (2007) 'Mapping Protease Inhibitor Resistance to Human Immunodeficiency Virus Type 1 Sequence Polymorphisms within Patients', *Journal of Virology*, 81: 13598–607.

Rhee, S. et al. (2005) 'HIV-1 Protease and Reverse-Transcriptase Mutations: Correlations with Antiretroviral Therapy in Subtype B Isolates and Implications for Drug-Resistance Surveillance', *The Journal of Infectious Diseases*, 192: 456–65.

Rhee, J. et al. (2015) 'HIV-1 Drug Resistance Mutations: Potential Applications for Point-of-Care Genotypic Resistance Testing', *PLoS One*, 10: e0145772.

Robbins, A. H. et al. (2010) 'Structure of the Unbound Form of HIV-1 Subtype A Protease: Comparison with Unbound Forms of Proteases from Other HIV Subtypes', *Acta Crystallographica Section D Biological Crystallography*, 66: 233–42.

Sali, A. and Blundell, T. (1993) 'Comparative Protein Modelling by Satisfaction of Spatial Restraints', *Journal of Molecular Biology*, 234: 779–815.

Santos-Pereira, A. et al. (2021) 'Nationwide Study of Drug Resistance Mutations in HIV-1 Infected Individuals under Antiretroviral Therapy in Brazil', *International Journal of Molecular Sciences*, 22: 5304.

Sartori, G. R. et al. (2019) 'Ligand-Induced Conformational Selection Predicts the Selectivity of Cysteine Protease Inhibitors', *PLoS One*, 14: e0222055.

Scott, W. R. P. et al. (1999) 'The GROMOS Biomolecular Simulation Program Package', *The Journal of Physical Chemistry A*, 103: 3596–607.

Serohijos, A. W. and Shakhnovich, E. I. (2014) 'Merging Molecular Mechanism and Evolution: Theory and Computation at the Interface of Biophysics and Evolutionary Population Genetics', *Current Opinion in Structural Biology*, 26: 84–91.

Shafer, R. W. (2006) 'Rationale and Uses of a Public HIV Drug-Resistance Database', *The Journal of Infectious Diseases*, 194: S51–S58.

Shah, D. et al. (2020) 'Evolution of Drug Resistance in HIV Protease', *BMC Bioinformatics*, 21: 497.

Souto, B. et al. (2021) 'Evolutionary Dynamics of HIV-1 Subtype C in Brazil', *Scientific Reports*, 11: 23060.

Stella-Ascariz, N. et al. (2017) 'The Role of HIV-1 Drug-Resistant Minority Variants in Treatment Failure', *The Journal of Infectious Diseases*, 216: S847–S850.

Taverna, D. M., and Goldstein, R. A. (2002) 'Why Are Proteins Marginally Stable?', *Proteins: Structure, Function, and Genetics*, 46: 105–9.

Tokuriki, N. et al. (2007) 'The Stability Effects of Protein Mutations Appear to Be Universally Distributed', *Journal of Molecular Biology*, 369: 1318–32.

Trompet, E. et al. (2020) 'Viral Fitness of MHV-68 Viruses Harboring Drug Resistance Mutations in the Protein Kinase or Thymidine Kinase', *Antiviral Research*, 182: 104901.

UNAIDS. (2021), *UNAIDS DATA 2021* (Geneva, Joint United Nations Programme on HIV/AIDS) <https://www.unaids.org/sites/default/files/media_asset/JC3032_AIDS_Data_book_2021_En.pdf> Accessed 23 May 2022.

von Heijne, G. (2018) 'Protein Evolution and Design', *Annual Review of Biochemistry*, 87: 101–3.

Wang, W. and Kollman, P. A. (2001) 'Computational Study of Protein Specificity: The Molecular Basis of HIV-1 Protease Drug Resistance', *Proceedings of the National Academy of Sciences of the United States of America*, 98: 14937–42.

Weber, I. T., Wang, Y.-F. and Harrison, R. W. (2021) 'HIV Protease: Historical Perspective and Current Research', *Viruses*, 13: 839.

Weikl, T. R., and Hemmateenejad, B. (2020) 'Accessory Mutations Balance the Marginal Stability of the HIV-1 Protease in Drug Resistance', *Proteins: Structure, Function, and Bioinformatics*, 88: 476–84.

Wensing, A. M. et al. (2019) '2019 Update of the Drug Resistance Mutations in HIV-1', *Topics in Antiviral Medicine*, 27: 111–21.

Whelan, S., and Goldman, N. (2001) 'A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach', *Molecular Biology and Evolution*, 18: 691–9.

Wilke, C. O. (2012) 'Bringing Molecules Back into Molecular Evolution', *PLoS Computational Biology*, 8: e1002572.

Wood, M. P. et al. (2013) 'A Compensatory Mutation Provides Resistance to Disparate HIV Fusion Inhibitor Peptides and Enhances Membrane Fusion', *PLoS One*, 8: e55478.

Wu, T. D. et al. (2003) 'Mutation Patterns and Structural Correlates in Human Immunodeficiency Virus Type 1 Protease Following Different Protease Inhibitor Treatments', *Journal of Virology*, 77: 4836–47.

Yang, C. et al. (2000) 'Sequence Note: Phylogenetic Analysis of Protease and Transmembrane Region of HIV Type 1 Group O', *AIDS Research and Human Retroviruses*, 16: 1075–81.

Zeldovich, K. B., Chen, P. and Shakhnovich, E. I. (2007) 'Protein Stability Imposes Limits on Organism Complexity and Speed of Molecular Evolution', *Proceedings of the National Academy of Sciences of the United States of America*, 104: 16152–7.

Zhang, J. et al. (2010) 'Detecting and Understanding Combinatorial Mutation Patterns Responsible for HIV Drug Resistance', *Proceedings of the National Academy of Sciences of the United States of America*, 107: 1321–6.

Zuckerkandl, E. and Pauling, L. (1965) 'Evolutionary Divergence and Convergence in Proteins'. In: Bryson, V. and Vogel, H.J. (eds) *Evolving Genes and Proteins*, pp. 97–166. New York: Academic Press.