Exploration of the effects of sequence variations between HIV-1 and HIV-2 proteases on their three-dimensional structures

Dhoha Triki[1], Maxime Kermarrec[2], Benoît Visseaux[3], Diane Descamps[3], Delphine Flatters[2], Anne-Claude Camproux[2] and Leslie Regad[2*]

[1]Department of Chemistry, Bioinformatics Research Center, College of Science, North Carolina State University, Raleigh, NC, 27695, USA.

[2]Université de Paris, Biologie Fonctionnelle et Adaptative, UMR 8251, CNRS, ERL U1133, INSERM, Computational Modeling of Protein Ligand Interactions, F-75013 Paris, France.

[3]Université de Paris, IAME, UMR 1137, INSERM, AP-HP, Laboratoire de Virologie, Hôpital Bichat-Claude Bernard, F-75018, Paris, France.

*Corresponding author: leslie.regad@univ-paris-diderot.fr

Abstract

HIV protease inhibitors (PIs) approved by the FDA (US Food and Drug Administration) are a major class of antiretroviral. HIV-2 protease (PR2) is naturally resistant to most of them as PIs were designed for HIV-1 protease (PR1). In this study, we explored the impact of amino-acid substitutions between PR1 and PR2 on the structure of protease (PR) by comparing the structural variability of 13 PR regions using 24 PR1 and PR2 structures complexed with diverse ligands. Our analyses confirmed structural rigidity of the catalytic region and highlighted the important role of three regions in the conservation of the catalytic region conformation. Surprisingly, we showed that the flap region, corresponding to a flexible region, exhibits similar conformations in PR1 and PR2. Furthermore, we identified regions exhibiting different conformations in PR1 and PR2, which could be explained by the intrinsic flexibility of these regions, by crystal packing, or by PR1 and PR2 substitutions. Some substitutions induce structural changes in the R2 and R4 regions that could have an impact on the properties of the PI-binding site and could thus modify PI binding mode. Substitutions involved in structural changes in the elbow region could alter the flexibility of the PR2 flap regions relative to PR1, and thus play a role in the transition from the semi-open form to the closed form, and have an impact on ligand binding. These results improve the understanding of the impact of sequence variations between PR1 and PR2 on the natural resistance of HIV-2 to commercially available PIs.

Keywords: HIV-1 protease; HIV-2 protease; protease inhibitors; drug resistance; protein structure comparison

1- Introduction

The protease (PR) is an important target in the treatment of human immunodeficiency virus (HIV) type 1 and type 2 (HIV-1 and HIV-2) infections due to its essential role in virus maturation. PR hydrolyses the viral Gag and Gag-Pol precursor polyproteins during infectious viral particle maturation. Currently, nine Food and Drug Administration (FDA)-approved drugs targeting HIV-1 PR have been developed: saquinavir (SQV), ritonavir, indinavir, nelfinavir, amprenavir (APV), lopinavir (LPV), atazanavir, tipranavir, and darunavir (DRV). HIV-2 is naturally resistant to most of these drugs and only three drugs (LPV, SQV, and DRV) are currently recommended for the treatment of HIV-2 infection (Brower, Bacha, Kawasaki, & Freire, 2008; Desbois et al., 2008; Menéndez-Arias, 2013). HIV-2 is also naturally resistant to some other drugs available for HIV-1 therapy, i.e., all non-nucleoside inhibitors of reverse transcriptase or fusion inhibitors. Therefore, novel and effective therapeutic agents for HIV-2 infection are urgently needed.

To understand the natural resistance of HIV-2 to protease inhibitors (PIs), several studies have compared the structures of HIV-1 PR and HIV-2 PR, hereafter referred to as PR1 and PR2, respectively. These two proteins are aspartic proteases consisting of a homodimer with 99 residues in each monomer. PR1 and PR2 share approximately 50% sequence identity and exhibit a similar global fold (Tong et al., 1993; Chen et al., 1994; Tong et al., 1995; Kovalevsky, Louis, Aniana, Ghosh, & Weber, 2008). Even though the two chains of each PR present exactly the same amino acid sequences, these two targets exhibit a certain structural asymmetry, i.e., their two chains exhibit different conformations (Prabu-Jeyabalan, Nalivaika, & Schiffer, 2000; Triki et al., 2018a). By comparing conformations of the two PR1 monomers complexed with different

substrates, Prabu-Jeyabalan, Nalivaika, & Schiffer, (2000) showed that the most asymmetric PR1 regions were regions 12-19, 36-42, 49-53, and 79-82. We recently characterized the local structural asymmetry of PR2 by comparing residue local conformations in both chains of the 19 PR2 structures available in the Protein Data Bank (PDB, Berman et al., 2000). We showed that PR2 exhibits asymmetry in 31% of its positions all along its structure, particularly in the elbow (37-42) and flap (43-58) regions (Triki et al., 2018a). Twelve of these asymmetric residues are located in the PI-binding site, and the structural asymmetry of these residues seems to be induced by ligand binding, suggesting that these residues are important for the adaptation of PR2 to its ligands. Comparison of the structural asymmetry observed in PR1 and PR2 showed that regions 12-19 (centre of fulcrum), 36-42 (elbow), 49-53 (end of the flap), and 79-82 (wall) are structurally asymmetric in both PR1 and PR2, while catalytic regions are more asymmetric in PR2 than in PR1 (Triki et al., 2018a). Priestle et al. (1995) showed that the PR2 structure complexed with the inhibitor CGP 53820 seems to better fulfil the two-fold symmetry conditions than PR1 complexed with the same PI. They also revealed that region 36-42 is the most asymmetric region in both PR1 and PR2 (Priestle et al. 1995). Some substitutions between PR1 and PR2 also appear to modify the PR asymmetry, which is important for recognition of and adaptation to the ligand (Triki et al., 2018a).

Other studies have focused on the impact of substitutions observed between the PR1 and PR2 pockets on interactions with PIs by comparing the structures of PR1 and PR2 complexed with PIs. Each of these studies used a single FDA-approved drug, and three concluded that amino acid changes at positions 32, 47, 76, and 82 play a partial role in drug resistance (Gustchina & Weber, 1991; Sardana et al., 1994; Hoog et al. 1995). In addition, Tie et al. (2012) compared the structures of PR1 mutants containing three mutations (V32I, I47V and V82I) to wild-type

PR1 and PR2 structures and examined their global biochemical properties and interactions with APV (Tie et al. 2012). These authors showed that, despite the similarity between the PR1 wild-type and mutant structures, the biochemical properties of the PR1 mutant and its interactions with APV were similar to those of PR2. The importance of these three amino acid changes was also confirmed by Raugi et al. (2016), who incorporated four HIV-1-like mutations (I32V, V47I, L76M, and I82V) into wild-type PR2. This PR2 mutant exhibited susceptibility to all PIs, highly similar to PR1 (Raugi, Smith, Gottlieb, & The University of Washington-Dakar HIV-2 Study Group 2016; 2016). It has also been shown that these substitutions induced small rearrangements, modifying the internal interactions between PR2 and DRV compared to the PR1-DRV complex (Kovalevsky, Louis, Aniana, Ghosh, & Weber, 2008; Tie et al., 2012; Raugi, Smith, Gottlieb, & The University of Washington-Dakar HIV-2 Study Group 2016). By comparing the physicochemical and geometrical properties of ligand-binding pockets from 15 PR1 and 13 PR2 structures, we have previously shown that these small structural rearrangements induced by the PR1-PR2 substitutions increase the hydrophobicity of the PR2 pocket relative to that of the PR1 pocket, along with a few other property modifications. These modifications could affect PI binding and PR flexibility (Triki et al., 2018b). Thus, all these studies have highlighted the effects of some amino acid changes located with PR pocket, particularly at residues establishing interactions with drugs. However, these studies have provided little information regarding structural changes induced in PRs and the potential consequences of these changes on sensitivity to PIs.

In this study, we explored the structural impacts of PR1 and PR2 substitutions by comparing wild-type PR1 and PR2 structures. We used a set of 12 PR1 and 12 PR2 crystallographic structures complexed with diverse ligands previously characterized regarding local asymmetries

and biochemical properties (Triki et al., 2018b). This PR set allows consideration of ligand diversity, different crystallographic experimental conditions, and PR flexibility. Several studies showed that a structural analysis and comparison of multiple target conformations sets associated with a target allow to investigate structural variability and to capture information about its flexibility (Zoete et al., 2002; van Westen et al., 2010; Monzon et al., 2013; Gaillard et al., 2013; Scott et al., 2015; Monzon et al., 2016; Gaillard et al.,; 2016; Regad et al., 2017). We quantified the intra- and inter-structural variability of thirteen PR functional and structural regions described by Sadiq & de Fabritiis (2010) using the root mean square deviation (RMSD) parameter, which is a classic and useful parameter that is used to compare global and local protein structures (Kufareva & Abagyan, 2012). To highlight the putative reasons underlying the inter-structural variability of the PR1 and PR2 regions, we studied the differences in structural variability in terms of different factors, such as crystal packing, PR intrinsic flexibility, and sequence variations between the two PRs. Our results suggested that the structural variability of some PR regions could be induced by variations between the PR1 and PR2 sequences and could impact PI binding and PR flexibility. These results provide an improved understanding of the effects of sequence variations between PR1 and PR2 on the structural variability of some PR regions, which could play a role in the resistance of PR2 to PIs.

2- Material & Methods

2.1- Dataset preparation

2.1.1- Selection of the PR1 and PR2 structures

For this study, we used the PR1 and PR2 structure sets previously developed and characterized to compare PR1 and PR2 binding pocket structure and properties (Triki et al., 2018b). These sets contain 12 PR1 and 10 PR2 crystallographic structures extracted from the PDB (Berman et al., 2000). All these structures have good resolution (≤ 3 Å). Two PR2 PDB files (PDB codes: 2MIP and 1HSH) contain two dimers (2MIP_AB and 2MIP_CD; 1HSH_AB and 1HSH_CD). Because the structural variability between the two dimers of the same structure was equivalent to the structural variability between two dimers extracted from two different PR2 structures (Figure S1), we decided to conserve these four dimers and study them separately, resulting in a set of 12 PR2 structures.

The 24 selected structures were complexed with 12 ligands: eight of the ligands were small molecules in complex with both PR1 and PR2 structures, and the remaining ligands corresponded to different peptides (Table S1). The 24 PR structures were placed into the same 3D referential by superimposing them onto a same reference structure. We tested the 24 complexed PR structures plus the PR2 structure in free form (used as reference in Triki et al., 2018b) as reference structure. For each reference, we superimposed all complexed PR structures onto it defining a set of superimposed structures and we computed the average backbone global RMSD between all superimposed PR1 and PR2 structures (Figure S3). From these results, we chose to superimpose the 24 complexed PR structures onto the PR2 in free form (PDB code 1HSI; Chen et al., 1994). This step resulted in a set of 24 superimposed PR structures.

2.1.2- PR decomposition using 13 regions

Each PR monomer was split into thirteen structural regions (Figure 1). Nine regions (the Nter (1-4), fulcrum (10-23), catalytic (24-30), elbow (37-42), flap (43-58), cantilever (59-75), wall

(80-83), α-helix (87-95), and Cter (96-99) regions) were extracted using the definition presented in Sadiq & de Fabritiis (2010). The four remaining regions, named R1 (5-9), R2 (31-36), R3 (75-79), and R4 (84-86), correspond to four inter-regions between these nine established regions.

2.2- Comparison of the PR1 and PR2 global folds

To compare the global folds of the PR1 and PR2 structures, RMSD values between 24x24 PR structure pairs were computed based on the main-chain atoms of the 198 residues (99x2 chains). These RMSDs were determined using the superimposed PR structures.

2.3- Quantification of the intra- and inter-structural variability of the 13 regions

For each of the 13 regions, the corresponding fragments were extracted from the 24 superimposed PR structures. Main-chain RMSD values between all the PR1 and PR2 fragments (24x24 pairs) were computed for each region. These RMSD values were stored in an RMSD matrix with 24 columns and rows.

2.3.1- Quantification and comparison of intra-structural variability of the 13 regions

The intra-structural variability of each region in the PR1 set and the PR2 set was determined by quantifying the corresponding conformational diversity of the region in the PR1 and PR2 sets. The intra-structural variability of a region in the PR1 set (resp. the PR2 set) was quantified by extracting all RMSD values computed between each pair of fragment structures of this region in the PR1 set (resp. the PR2 set) from the RMSD matrix. The average main-chain RMSD values in the PR1 and PR2 sets were then computed and denoted $RMSD_{PR1}$ and $RMSD_{PR2}$, respectively. An average $RMSD_{PR1}$ (resp. $RMSD_{PR2}$) value less than 0.5 Å indicated a structurally conserved region in the PR1 set (resp. the PR2 set), while an average $RMSD_{PR1}$ (resp. $RMSD_{PR2}$) value higher than 0.5 Å was characteristic of a structurally variable region in the PR1 set (resp. the PR2 set). A value of 0.5 Å was chosen to consider crystallographic errors in the set of structures with a resolution of less than 3 Å (Weber, Kovalevsky, & Harrison, 2007). We assessed whether each region exhibits similar structural variability in the PR1 and PR2 sets by comparing the average $RMSD_{PR1}$ and $RMSD_{PR2}$ values using t-tests.

2.3.2- Quantification of the inter-structural variability of the 13 regions

The inter-structural variability of a region corresponds to its conformational diversity between the PR1 and PR2 structures and reflects the different conformations of a region sampled in the PR1 and PR2 sets. The inter-structural variability of each region between the two PR sets was quantified using all RMSD values computed between the corresponding PR1 and PR2 fragments extracted from the RMSD matrix. The corresponding average main-chain RMSD between the two

PR sets, denoted $RMSD_{PR1-PR2}$, was then computed for each region. The comparison between the average $RMSD_{PR1-PR2}$ value of a region with its average $RMSD_{PR1}$ and $RMSD_{PR2}$ values was performed using t-tests.

2.4- Multidimensional scaling analysis

We used multidimensional scaling (MDS) analysis, a method for visual exploration of object similarity, to compare the PR1 and PR2 conformations of the thirteen regions (Cox & Cox, 1994). The MDS algorithm starts with the RMSD matrix computed for a given region for placement of its different PR1 and PR2 conformations in a low-dimensional space, where the distances between fragment conformations were preserved as much as possible (Cox & Cox, 1994). We performed MDS for each region and generated its MDS map corresponding to the region fragment projection onto the first two MDS components. The MDS was computed using the *cmdscale()* function of the *MASS* package in R (Venables & Ripley, 2002). The obtained MDS map presents the conformational space of a region sampled by its PR1 and PR2 fragments. In the MDS map of a given region, clear separation between the PR1 and PR2 fragments indicates that the studied region has different conformations in the PR1 and PR2 structures.

2.5- Study of structural variability between PR1 and PR2 and crystal packing

For each PR structure, residues putatively involved in crystal packing were detected using the protocol presented in Carugo & Djinović-Carugo, 2012 and Ollitrault et al., 2018. Hereafter, the

molecule contained in the asymmetric unit of this structure was referred to as the "reference molecule". For a given structure, the "symmetry mate molecules" of the reference molecule were generated by performing symmetry operations with a 4.5 Å distance cut-off using PyMol software. This step is based on the CRYST1 record of the PDB file that presents the unit cell parameters, space group, and Z value. Atoms of the given structure involved in crystal packing were defined as reference atoms establishing intermolecular contacts with at least one symmetric-mate atom, i.e., situated at less than 4.5 Å from at least one symmetric mate. This cut-off of 4.5 Å was set to be longer than typical interactions: hydrogen bonds and electrostatic interactions. The extraction of atoms involved in crystal packing was performed using PyMoL software. A residue was considered to be involved in crystal packing and termed a crystal packing residue if at least one of its atoms was involved in crystal packing. This protocol was applied to the 24 PR structures to obtain a list of crystal packing residues for each PR structure.

3- Results & Discussion

3.1- PR1 and PR2 set presentation

Figure 1A presents the multiple alignment of amino acid sequences extracted from the 12 PR1 and 12 PR2 structures. There are nine mutated positions within the PR1 sequence set. Five of those PR1 mutations (Q7K, L33I, L63I, C67A, and C95A) correspond to experimental mutations introduced to minimize autoproteolysis and to prevent cysteine-thiol oxidation of PR1 (Rosé, Salto, & Craikl, 1993). The three remaining PR1 mutations (K14R, S37N, and R41K) correspond to natural polymorphism mutations that were observed in HIV-1 strains isolated from treatment-naïve and PI-experienced patients (Shaw et al., 2016; Descamps et al., 2009; Azam, Malik,

Rizvi, & Rai, 2014; Amiel et al., 2011). Finally, another mutation (K57L) was detected within the PR2 sequence set. This position has been reported as polymorphic with the mutation K57R (Damond et al., 2005). In addition, we counted 50 substitutions between the PR1 and PR2 sequence sets. Variations between the PR1 and PR2 sequences are located along the sequence with a relatively high representation in the elbow, cantilever, R2, and R3 regions (Figure 1A and 1B). The superimposition of the 24 PR structures shows that all the PR structures have a similar fold (Figure 1C), resulting in a weak average main-chain RMSD (0.6 ± 0.2 Å, Figure S1), consistent with (Gustchina & Weber, 1991; Tong et al., 1993; Kovalevsky, Louis, Aniana, Ghosh, & Weber, 2008; Tie et al., 2012). The largest structural differences between the PRs occur at residues in the outer loops regardless of the co-crystallized ligand types (small molecules or peptides, Figure S2), particularly in the elbow region (Figure 1C). This result is consistent with Tong et al., 1993 and Kovalevsky, Louis, Aniana, Ghosh, & Weber, 2008.

3.2- Determination of the intra-structural variability of each region in the PR1 and PR2 sets

The intra-structural variability of the thirteen PR regions in the PR1 and PR2 sets separately was determined by computing the average $RMSD_{PR1}$ and $RMSD_{PR2}$ values. The intra-variability of regions extracted from the PR1 and PR2 sets was quantified after the superimposition of complexed structures onto the same reference structure. To analyze the impact of the reference structure in the structure superimposition, we tested several structures (the 24 complexed PR

structures plus the PR2 structure in free form used in Triki et al. (2018b)) as reference structure. Figure S3 presents the quantification of the structural variability of the PR1 and PR2 sets composed of complexed PR structures superimposed on each tested reference structure. Similarly, figure S4 presents the quantification of the structural variability relatively to the 13 regions extracted from the 24 superimposed complexed structures). We observed that the choice of the reference structure has only a weak impact on the variability of the PR set as each reference structure leads to similar variability. Thus, we chose to work with the 24 complexed PR structures superimposed onto the PR2 structure in free form (PDB code 1HSI) as in Triki et al., (2018b). Figures 2A and S5 present the resulting intra-structural variability of the thirteen regions extracted from chains A and B for the PR1 and PR2 sets. We noted that chains A and B exhibited the same behaviour in terms of intra-variability for the thirteen regions. We first based our analyses on chain A and compared the structural variability of chains A and B in section 3.6. Figure 2A shows that there was no link between the size and the intra-structural variability of chain A regions: the longest regions (fulcrum and cantilever) were not the most variable. In addition, the intra-structural variability of regions is not linked to structural motifs. Indeed, two regions can exhibit the same structural motifs but have different intra-variability. For example, the catalytic or wall regions exhibit low intra-variability, while the elbow region is the most variable region in the two PR sets (Figure 2A), despite these three regions having a loop conformation (Figure 1B). We noted the same result for regions exhibiting a β-sheet conformation, such as the flap and fulcrum regions (Figure 2A).

The catalytic region was structurally conserved in both the PR1 and PR2 sets ($RMSD_{PR1}$ and $RMSD_{PR2}$ < 0.5 Å), consistent with Zoete, Michielin, & Karplus, 2002; Tong et al., 1995; Kar &

Knecht, 2012; Chen et al., 2014. The observed rigidity of the catalytic region was expected because its function requires a well-defined stable 3D structure. Our results also revealed structural conservation of the R1, R4, and α-helix regions, suggesting that these four regions play a role in the PR function or are important for the PR structure. Figure 3 shows that the regions R1, R4, and α-helix are located near the catalytic region, forming H-bonds between the catalytic region residue 29 and residues 8 (R1 region) and 85 (α-helix region), indicating that these three conserved regions (R1, R4, and α-helix regions) are important for the structure of the catalytic region and its conservation in both PRs. These results are consistent with our previous results, which showed that some R4 (84 and 85) and α-helix (87, 89, 94, 95) residues exhibit the same conformations in the 19 available PR2 structures (Triki et al., 2018c). These results support the interest in mining several structures associated with a target to gain valuable insight into target structural and functional regions (Zoete, Michielin, & Karplus, 2002; van Westen, Wegner, Bender, Ijzerman, & van Vlijmen, H.W., 2010; Regad et al., 2017). In contrast, nine regions (the Nter, fulcrum, R2, elbow, flap, cantilever, R3, wall, and Cter regions) are structurally variable in the PR1 and PR2 sets ($RMSD_{PR1}$ or $RMSD_{PR2}$ > 0.5 Å, Figure 2A). The structural variability of the wall region (loop 76-82) was previously reported in PR1 and was shown to be important for the recognition of different ligands (Prabu-Jeyabalan, Nalivaika, & Schiffer, 2000). The elbow region is the most variable in both the PR1 and PR2 sets, consistent with the high variability of this region observed after comparison of several PR1 and PR2 complexes (Zoete, Michielin, & Karplus, 2002; Tong et al., 1995; Kar & Knecht, 2012; Chen et al., 2014; Triki et al., 2018c). Based on the compositions of the PR1 and PR2 sets, the intra-structural variability in the two sets was explained by the intrinsic flexibility of PRs, binding of diverse ligands, and differences in experimental parameters.

All the regions exhibit greater variability in the PR2 set than in the PR1 set, i.e., the average $RMSD_{PR2}$ value was higher than the average $RMSD_{PR1}$ value (Figure 2B). This finding could be explained by substitutions between the two PRs and by the higher diversity of the PR2 set. Indeed, the PR1 set contains three structures complexed with APV, three complexed with DRV, and one complexed with a peptide, while the PR2 set contains only one structure complexed with each ligand and three structures complexed with three different peptides. In addition, the observed high variability of the flap region (residues 43-58) in the PR2 set relative to the PR1 set is not consistent with the larger flexibility observed in the region around residue 50 in the PR1-DRV complex compared to the PR2-DRV complex during molecular dynamics simulation (Chen et al., 2014). These differences could be explained by the use, in our work, of PR1 and PR2 structures complexed with 12 different ligands, whereas Chen et al. (2014) focused only on the PR1 and PR2 structures complexed with a single FDA-approved drug.

3.3- Comparison between PR1 and PR2 conformations

To assess whether the 13 regions exhibit a similar conformation in the PR1 and PR2 sets, their inter-structural variability between the PR1 and PR2 regions was quantified by computing the $RMSD_{PR1-PR2}$ value of each region (Figure 2A). Globally, the most variable regions between the PR1 and PR2 sets, i.e., those exhibiting high $RMSD_{PR1-PR2}$ values, have the highest intra-variability (Pearson correlation coefficient between $RMSD_{PR1}$ and $RMSD_{PR1-PR2}$=0.90 and Pearson correlation coefficient between $RMSD_{PR2}$ and $RMSD_{PR1-PR2}$=0.87). For example, the conserved catalytic region has a very weak $RMSD_{PR1-PR2}$ (close to 0.5 Å), implying that the PR1 and PR2 fragments corresponding to this region exhibit similar conformations. In contrast, the variable

elbow region has a high $RMSD_{PR1-PR2}$ value (> 2.5 Å), indicating that this region sampled different conformations in both sets.

The MDS maps of regions illustrate the diversity of conformations sampled by each region in the two sets (Figure 4). The four conserved regions (the R1, catalytic, R4, and α-helix regions) sampled a small space in the corresponding MDS maps, while the nine variable regions sampled a relatively large space. Regarding the separation between the PR1 and PR2 conformations, the MDS maps highlight two types of regions. Three regions (the catalytic, R3, wall, and flap regions) did not present clear separation between the PR1 and PR2 conformations, indicating that certain PR1 and PR2 fragments for these regions exhibit the same conformation in the PR1 and PR2 sets. These similar conformations of both PRs are expected for the catalytic region and confirm the importance of a conserved structure for PR activity. This result is surprising for the large variable flap and R3 regions, as these regions are highly substituted between PR1 and PR2 (Figure 1A). These two regions exhibit similar conformations in PR1 and PR2, except fragments extracted from PR2 complexed with APV and its derivatives, including DRV. Thus, our results suggest that the substitutions occurring in these regions do not induce structural changes in these regions and that the inter-structural variability is explained by the different experimental parameters used to solve the structures, the ligand diversity in the PR sets, and by the intrinsic flexibility of these regions. By comparing the dynamic behaviour of PR1 and PR2 structures in complex with DRV using molecular dynamics simulations, Chen et al. (2014) have suggested that the high flexibility of the flap region in the PR1-DRV complex relative to that in the PR2-DRV complex is induced by the I47V substitution (Chen et al., 2014). In PR1, I47 residue, located in

the binding pocket, is important for the ligand binding as it established interactions with clinical IPs (Ghosh et al., 2009; Koh et al., 2009). It is also involved in flap stability by forming hydrophobic interactions with its neighbouring residues (V32, K45, I50, I54, V56, and L76, Ghosh et al., 2009; Shen, Wang, Kovalevsky, Harrison, &Weber, 2010). In HIV-1, the single mutation I47V is responsible of resistance to several clinical IPs (APV, DRV, LPV, and TPV, Johnson et al., 2010; Rhee et al., 2010; Chang et al., 2012). The analysis of the structure and dynamics of the wild-type PR1 and the I47V-mutant PR1 showed that mutation I47V induces a loss of contacts with neighboring residues, which would decrease the stability of the flap region (Bandyopadhyay & Meher, 2006; Chang et al., 2012). Thus, we hypothesize that either the PR1- and PR2-DRV complexes have particular behaviours relative to other complexes or that the substitutions observed between the PR1 and PR2 sequences do not directly modify the PR1 and PR2 flap conformations but impact the intrinsic flexibility of these regions and, thus, the flap movement during the transition from the semi-open to the closed forms of the PR, as observed by Chen et al., 2014.

The second region type highlighted by MDS maps corresponds to regions exhibiting clearly different conformations in the PR1 and PR2 sets, as illustrated by a strong separation between the PR1 and PR2 conformations in the MDS maps. Such regions correspond to the R1, R4 and α-helix regions, despite their low structural variability, as well as the Nter, Cter, wall, fulcrum, R2, cantilever, and elbow regions. Thus, all these regions adopt several conformations that are clearly different between the two PR sets, which could result from substitutions between the PR1 and PR2 sequences but also from difference in experimental parameters, ligand diversity, and the intrinsic flexibility of these regions.

3.4- Link between structural deformation between PR1 and PR2 and crystal packing

One possible reason for the structural changes between the PR1 and PR2 regions is crystal packing in structures. Residues involved in crystal packing in the 24 PR structures were determined using the protocol presented in Carugo & Djinović-Carugo, 2012 and Ollitrault et al., 2018 (see Material and Methods section). Figure 5 presents the chain A residues involved in crystal packing. We noted that crystal packing involved residues across the entire PR structure but particularly those in the fulcrum, flap, cantilever and α-helix regions. Some residues are involved in crystal packing in most PR structures, such as 4A, 6A, 7A, and 53A. Some other residues appeared to be involved in crystal packing in only PR1 or PR2. For example, crystal packing at residues 18A (fulcrum) and 92A (α-helix) was observed in all the PR1 structures but in less than 20% of the PR2 structures. In contrast, crystal packing was detected at residues 42A (elbow) and 99A (Cter) in a few PR1 structures (25% and none, respectively) but in most PR2 structures (92% and 75%, respectively). Thus, the structural changes observed between PR1 and PR2 could be explained by differences in crystal packing between the two PRs. As illustrated in Figure 1A, three (42, 92, and 99) of these residues are substituted between PR1 and PR2. Thus, these amino acid changes seem to be responsible for the difference in crystal packing between PR1 and PR2 and thus for the structural differences between these residues.

3.5- Link between structural changes and sequence variations between PR1 and PR2

To highlight structural changes putatively induced by sequence variations between PR1 and PR2, we compared the intra-variability ($RMSD_{PR2}$) and inter-variability ($RMSD_{PR1-PR2}$) for regions exhibiting structural differences between the PR1 and PR2 structures (the Nter, Cter, wall, fulcrum, R2, cantilever, and elbow regions, see Figure 2B). According to these comparisons, we distinguished two types of regions: (i) regions with an average $RMSD_{PR1-PR2}$ value lower or similar to the corresponding average $RMSD_{PR2}$ value and (ii) regions with an average $RMSD_{PR1-PR2}$ value higher than the corresponding average $RMSD_{PR2}$ value. Figure 2B shows that the R1, Cter, and wall regions have similar $RMSD_{PR1-PR2}$ and $RMSD_{PR2}$ values, indicating the same variability in the PR2 set and between the PR1 and PR2 conformations. Thus, even though these regions exhibited different conformations in the PR1 and PR2 sets, the inter-structural variability of these regions was not higher than the intra-structural variability. The different conformations of these regions observed in the PR1 and PR2 structures were not induced by PR1-PR2 substitutions but seem to be explained by the high intrinsic flexibility of these regions in the PR2 structures, which can be explained by differences in experimental parameters, ligand diversity and the intrinsic flexibility of these regions.

In contrast, the R4, α-helix, Nter, fulcrum, R2, cantilever, and elbow regions have higher average $RMSD_{PR1-PR2}$ values than the average $RMSD_{PR2}$ value, indicating that these regions are more variable between the PR1 and PR2 structure sets than within the PR2 structures (Figure 2B). These results indicate that the reasons for the variability of these regions within the PR2 set do not entirely explain the conformational differences between the PR1 and PR2 structure

sets, suggesting a role of PR1-PR2 substitutions in these seven regions. Structural differences between the PR1 and PR2 fulcrum, elbow and cantilever conformations were previously reported (Tong et al., 1993). These authors suggested that the structural differences at residues 15-20 (part of fulcrum) between the PR1 and PR2 structures complexed with peptides may be an artefact of crystal packing and the differences observed in region 60-73 (in the cantilever region) could result from the C67L and A71V substitutions (Tong et al., 1993). The induced structural changes in the elbow region between PR1 and PR2 could have an impact on flap movement, which could result in different flap transitions from semi-open to closed forms in PR1 and PR2 complexes with APV and DRV, as reported by Chen et al., 2014. These different flap movements could have an impact on PI binding and could thus be involved in the resistance of PR2 to PIs. Moreover, it has been shown that the R2 and R4 regions are important for the conformation of the PI-binding pockets (Triki et al., 2018c). Indeed, two R2 residues are located within the PI-binding pocket, and two R4 residues (84 and 85) are involved in the hydrogen bond network with binding pocket residues (Triki et al., 2018c). Thus, structural changes involved by substitutions at the R2 and R4 regions could modify the conformation and properties of the PR2 binding pocket and could alter PI binding in PR2 relative to PR1. This finding is consistent with the fact that the PR2 binding pocket was characterized as being smaller and more hydrophobic than the PR1 pocket as observed after comparing the pocket flexibility using molecular dynamics simulations (Chen et al., 2014) or the properties of 28 PR1 and PR2 PI-binding pockets (Triki et al., 2018b).

3.6- Structural variability of these 13 regions of chain B

We quantified the structural intra-variability of the 13 chain B regions in the PR1 and PR2 sets by computing their $RMSD_{PR1}$ and $RMSD_{PR2}$ values in the two sets (Figure S5). In chain B, the R4 and R1 regions are structurally conserved within both the PR1 and PR2 sets. The α-helix region is conserved within only the PR1 set, while the catalytic region is conserved within only the PR2 set. Eight chain B regions presented a significantly higher variability within PR2 than within PR1, as in chain A. The fulcrum and R4 regions of the chain B exhibited the same intra-variability in both PR sets. The catalytic region presented a lower variability within PR2 than within PR1 (Table S2). Comparison of the intra- and inter-variability of these chain B regions showed similar behaviour to chain A regions. The catalytic, flap, and R3 regions of chain B sampled similar conformations in both PR1 and PR2 (Figure S6). Substitutions occurring at positions 4B (Nter), 71B (cantilever), and 92B (α-helix) between PR1 and PR2 seemed to be responsible for differences in crystal packing and could explain the structural differences in these regions (Figure S7). Similar to the chain A regions, sequence variations in PR1 and PR2 seemed not to be directly responsible for the different conformations observed in the R1, wall, and Cter regions. In contrast, the different conformations in the Nter, fulcrum, R2, elbow, cantilever, R4, and α-helix regions of chain B in PR1 and PR2 resulted at least partially from substitutions occurring between PR1 and PR2.

3.7- Structural asymmetry of the 13 regions in the two PRs

We compared the intra-structural variability of each region in chains A and B in the two PR sets (Table S2). In both PRs, the elbow region exhibited the same high intra-variability between the

two chains. In contrast, other regions appeared more variable in chain B than in chain A, i.e., these regions exhibited asymmetric behaviour in terms of differences in variability, in both the PR1 and PR2 sets (the catalytic, R2, cantilever, R3, and α-helix regions) or in only the PR1 set (the Nter, R1, fulcrum, flap, wall, R4, and Cter regions; see Table S2). Some of these regions have been previously detected as being structurally asymmetric in terms of conformational changes within the two chains for both PR1 and PR2, such as the catalytic, fulcrum and flap regions (Prabu-Jeyabalan, Nalivaika, & Schiffer, 2000; Triki et al., 2018a). For example, Triki et al. (2018) showed that structural asymmetry of the catalytic region in PR2 is ligand specific and could be important for ligand interaction and binding, whereas the structural asymmetry of the R2 and cantilever regions could be linked to the intrinsic flexibility of PR2. Our results reveal that the Nter, R1, fulcrum, flap, wall, R4, and Cter regions exhibit different asymmetric behaviours in terms of structural intra-variability in PR1 and PR2. Thus, we hypothesize that substitutions between PR1 and PR2 sequences could modify the flexibility of these proteins. Notably, the difference in the asymmetric behaviour of the wall region between PR1 and PR2 was previously observed in terms of conformational changes (Triki et al. 2018a). In PR2, residues 81 and 82 exhibit the same local conformations in both PR2 chains (Triki et al. 2018a), while Prabu-Jeyabalan, Nalivaika, & Schiffer, 2000 revealed that the 79-82 loop region exhibits different conformations in the two chains, which is important for ligand recognition and adaptation of the pocket to the substrate shape (Prabu-Jeyabalan, Nalivaika, & Schiffer, 2000). Triki et al. (2018a) suggested that this difference in the structural asymmetry of residues 81 and 82 could be induced by substitutions occurring at position 82 (V82I) between the PR1 and PR2 sequences, a substitution associated with the multi-PI resistance of HIV-1 (Kovalevsky, Louis, Aniana, Ghosh, & Weber, 2008). Thus, these results suggest that this substitution could modify the

flexibility of residue 82, which has an impact on the capacity of this residue to move and adjust the PR2 pocket to the ligand shape.

4- Conclusion

In this study, we explored the impact of sequence variations between the PR1 and PR2 structures on the backbone variability and conformations of 13 PR regions. We compared the intra- and inter-structural variability and conformations of these 13 regions extracted from two PR sets composed of 12 PR1 and 12 PR2 structures using main-chain RMSD values. The originality of our approach was in the use of several PR1 and PR2 structures, allowing easy and precise extraction of information regarding the structural variability of each region in both sets. In addition, the diversity of the PR1 and PR2 structures allowed identification of the structural variability resulting from both the intrinsic flexibility of PRs, induced by ligand binding, and from variations in experimental conditions.

Our analyses confirmed the structural rigidity of the catalytic region in both PR1 and PR2 and showed that the R1, R4, and α-helix regions seem important for conservation of the structure of the catalytic region. Comparison of the PR1 and PR2 region conformations showed that the R3, catalytic, and flap regions share similar conformations in the two PR structures, suggesting that sequence variations observed between the PR1 and PR2 structures do not impact their conformations. In addition, we observed that other regions exhibit different backbone conformations between the PR1 and PR2 sets. This structural variability between the two PRs could be explained by several factors: crystal packing in the crystallographic structures, intrinsic flexibility of PRs, and sequence variations between the two PRs. Our results suggest that the

different conformations observed between PR1 and PR2 fragments in the R1, wall, and Cter regions can be explained by the intrinsic flexibility of PRs but not by substitutions between the two PRs. In contrast, we showed that the different conformations observed in the Nter, R4, α-helix, fulcrum, R2, cantilever, and elbow regions between the PR1 and PR2 structures could be partially explained by amino acid substitutions. The structural changes observed in the R2 and R4 regions could modify the conformation and properties of the PI-binding pocket, which could directly modify the binding of PIs with PR2. In addition, the structural changes observed in the elbow region could have an impact on the PR dynamics by modifying the flap movement and, thus, the PI interactions. However, these results should be interpreted with caution, as differences in crystal packing were detected in the elbow and α-helix regions of PR1 and PR2. In addition, our results indicated that the substitutions could impact the flexibility of some regions, particularly the flap and wall regions. The former region is important for the transition from the semi-open to the closed forms during ligand binding, and the latter region is involved in ligand recognition. These modifications of the structural variability and conformations of these regions, putatively induced by sequence changes between PR1 and PR2, could play a role in natural resistance of PR2 to PIs. These results provide some new insights into the role of the substitutions between PR1 and PR2 in the low sensitivity of PR2 to commercially available PIs. The results highlight the importance of some PR2 regions (the Nter, R4, α-helix, fulcrum, R2, cantilever, and elbow regions) in the resistance of PR2, which could be linked to the modification of the flap flexibility of PR2 putatively induced by structural changes in the elbow region.

The study presented here, using an effective and rapid method based on available structures to understand PR2 resistance to PI, is a first step in the comparison of PR1 and PR2 structures associated with different ligands. Our results provided new data and hypothesis that have now to

be confirmed in a next step using the dynamics simulation methods, time consuming methods that have been previously used to explore the PI-binding on PR1 structure and impact of drug resistance mutations in PR1 (Hou, McLaughlin & Wang, 2008; Hu et al., 2010; Zhang, Hou, Wang & Liu, 2010; Makatini et al., 2011; King et al., 2012; Tzoupis, Leonis, Mavromoustakos & Papadopoulos, 2013; Hu, Ma, Dou, Zhao & Wang, 2016; Paulsen, Leidner, Ragland, Kurt Yilmaz & Schiffer, 2017; Karnati & Wang, 2019; Pawar et al., 2019) and to compare PR1 and PR2 dynamics to understand the resistance of PR2 against the clinical PIs (Kar et al., 2012; Chen et al., 2014). In the current work, we suggest that substitutions at positions 3-4, 31-36, 40-42, and 92-93 modify the conformations of the Nter, R2, elbow, and α-helix regions in PR2, respectively. Thus, modelling PR2 structures containing these specific substitutions could allow analyzing and confirming the effect of substitutions between PR1 and PR2 on the structural differences in the Nter, R4, α-helix, fulcrum, R2, cantilever, and elbow regions. More specifically, the comparison of these modelled PR2 mutant structures with the wild-type PR1 and PR2 structures, complexed or not with ligands, during molecular dynamics simulations could provide information regarding the structural deformation induced by these substitutions and facilitate detailed understanding of the role of these substitutions. The choice of such PR2 mutants and the main regions of interest will be orientated according to our results in this study.

Funding

This work was supported by an ANRS Grant to B.V., D.D., D.F., A.C.C. and L.R. D.T. was supported by an ANRS fellowship.

Declaration of interest statement

The authors declare no conflict of interest.

References

Amiel, C., Charpentier, C., Désiré, N., Bonnard, P., Lebrette, M. G., Weiss, L., Pialoux, G., & Schneider, V. (2011). Long-term follow-up of 11 protease inhibitor (PI)-naïve and PI-treated HIV-infected patients harbouring virus with insertions in the HIV-1 protease gene. *HIV Medicine*, 12(3), 138-44. https://doi: 10.1111/j.1468-1293.2010.00862.x

Azam, M., Malik, A., Rizvi, M., & Rai, A. (2014) Zero prevalence of primary drug resistance-associated mutations to protease inhibitors in HIV-1 drug-naive patients in and around Aligarh, India. *Journal of Infection in Developing Countries*, 8(1), 79-85. https://doi: 10.3855/jidc.3480

Bandyopadhyay, P. & Meher B. R. (2006). Drug Resistance of HIV-1 Protease Against JE-2147: I47V Mutation Investigated by Molecular Dynamics Simulation. *Chemical Biology & Drug Design,* 67(2), 155-161. https://doi: 10.1111/j.1747-0285.2006.00348.x

Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., … Bourne, P. E. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235-242. https://doi.org/10.1093/nar/28.1.235

Brower, E. T., Bacha, U. M., Kawasaki, Y., & Freire, E. (2008). Inhibition of HIV-2 protease by HIV-1 protease inhibitors in clinical use. *Chemical Biology and Drug Design*, *71*(4), 298-305. https://doi.org/10.1111/j.1747-0285.2008.00647.x

Carugo, O., & Djinović-Carugo, K. (2012). How many packing contacts are observed in protein

crystals? *Journal of Structural Biology*, *180*(1), 96-100. https://doi.org/10.1016/j.jsb.2012.05.009

Chen, J., Liang, Z., Wang, W., Yi, C., Zhang, S., & Zhang, Q. (2014). Revealing origin of decrease in potency of darunavir and amprenavir against HIV-2 relative to HIV-1 protease by molecular dynamics simulations. *Scientific Reports*, *4*, 6872. https://doi.org/10.1038/srep06872

Chang, Y. C. E., Yu, X. Zhang, Y., Tie, Y., Wang. Y. F., Yashchuk5, S., … Weber, I. T. (2012). Potent antiviral HIV-1 protease inhibitor GRL-02031 adapts to the structures of drug resistant mutants with its P1′ -pyrrolidinone ring. *Journal of Medicinal Chemistry*, 55(7), 3387-3397. https://doirg/10.1021/jm300072d.

Chen, Z., Li, Y., Chen, E., Hall, D. L., Darke, P. L., Culberson, C., … Kuo, L. (1994). Crystal structure at 1.9-A resolution of human immunodeficiency virus (HIV) II protease complexed with L-735,524, an orally bioavailable inhibitor of the HIV proteases. *Journal of Biological Chemistry*, *269*(42), 26344-26348.

Cox, T., & Cox, M. (1994). *Multidimensional Scaling*, 315-347. *In* Chen, C.C, Härdle, W., eds Handbook of Data Visualization. Springer. https://doi.org/https://doi.org/10.1007/978-3-540-33037-0_14

Descamps, D., Lambert-Niclot, S., Marcelin, A. G., Peytavin, G., Roquebert, B., Katlama, C., … Brun-Vézinet, F. (2009). Mutations associated with virological response to darunavir/ritonavir in HIV-1-infected protease inhibitor-experienced patients. *Journal of Antimicrobial Chemotherapy*, 63(3), 585-92. https://doi.org/10.1093/jac/dkn544

Damond, F., Brun-Vezinet, F., Matheron, S., Peytavin, G., Campa, P., Pueyo, S., … Descamps, D. (2005). Polymorphism of the human immunodeficiency virus type 2 (HIV-2) protease gene and

selection of drug resistance mutations in HIV-2-infected patients treated with protease inhibitors. *Journal of Clinical Microbiology*, 43(1), 484-487. https://doi.org/10.1128/JCM.43.1.484-487.2005.

Desbois, D., Roquebert, B., Peytavin, G., Damond, F., Collin, G., Bénard, A., … Descamps, D. (2008). In vitro phenotypic susceptibility of human immunodeficiency virus type 2 clinical isolates to protease inhibitors. *Antimicrobial Agents and Chemotherapy*, *52*(4), 1545-1548. https://doi.org/10.1128/AAC.01284-07

Ghosh, A. K., Leshchenko-Yashchuk, S., Anderson, D. D., Baldridge, A., Noetzel, M., Miller, H. B., … Mitsuya, H. (2009). Design of HIV-1 protease inhibitors with pyrrolidinones and oxazolidinones as novel P1′ -ligands to enhance backbone-binding interactions with protease: synthesis, biological evaluation, and protein-ligand X-ray studies. *Journal of Medicinal Chemistry*, 52, 3902- 3914. https://doi.org/10.1021/jm900303m

Gustchina, A., & Weber, I. T. (1991). Comparative analysis of the sequences and structures of HIV-1 and HIV-2 proteases. *Proteins: Structure, Function, and Bioinformatics*, *10*(4), 325-339. https://doi.org/10.1002/prot.340100406

Hoog, S. S., Towler, E. M., Zhao, B., Doyle, M. L., Debouck, C., & Abdel-Meguid, S. S. (1996). Human immunodeficiency virus protease ligand specificity conferred by residues outside of the active site cavity. *Biochemistry*, *35*(32), 10279-10286. https://doi.org/10.1021/bi960179j

Hou, T., McLaughlin, W. A., Wang, W. (2008). Evaluating the potency of HIV-1 protease drugs to combat resistance. *Proteins: Structure, Function, and Bioinformatics*, 71(3):1163-1174. https://doi.org/ 10.1002/prot.21808

Hu, G., Ma, A., Dou, X., Zhao, L., Wang, J. (2016). Computational studies of a mechanism for binding and drug resistance in the wild type and four mutations of HIV-1 protease with a GRL-0519 inhibitor. *International Journal of Molecular Sciences*, 17(6):819. https://doi.org/10.3390/ijms17060819

Hu, G. D., Zhu, T., Zhang, S. L., Wang, D., Zhang, Q. G. (2010). Some insights into mechanism for binding and drug resistance of wild type and I50V V82A and I84V mutations in HIV-1 protease with GRL-98065 inhibitor from molecular dynamic simulations. *European Journal of Medicinal Chemistry*, 45(1), 227-35. https:// doi: 10.1016/j.ejmech.2009.09.048

Johnson, V. A, Brun-Vezinet, F., Clotet, B., Gunthard, H. F., Kuritzkes, D. R., Pillay, D., … Richman, D. D. (2010). Update of the drug resistance mutations in HIV-1: December 2010. *Topics in. HIV Medicine*. 18(5), 156-163.

Kar, P., & Knecht, V. (2012). Origin of decrease in potency of darunavir and two related antiviral inhibitors against HIV-2 compared to HIV-1 protease. *Journal of Physical Chemistry B*, *116*(8), 2605-2614. https://doi.org/10.1021/jp211768n

Karnati, K.R. & Wang Y. (2019). Structural and binding insights into HIV-1 protease and P2-ligand interactions through molecular dynamics simulations, binding free energy and principal component analysis. *Journal of Molecular Graphics and Modelling*, 92, 112-122. https://doi.org/10.1016/j.jmgm.2019.07.008.

King, N. M., Prabu-Jeyabalan, M., Bandaranayake, R. M., Nalam, M. N., Nalivaika, E. A., Özen, A., … Schiffer, C. A. (2012). Extreme entropy-enthalpy compensation in a drug-resistant variant of HIV-1 protease. *ACS Chemical Biology*, 7(9):1536-1546. https://doi.org/10.1021/cb300191k

Koh, Y., Das, D., Leschenko, S., Nakata, H., Ogata-Aoki, H., Amano, M., … Mitsuya, H.

(2009). GRL-02031, a novel nonpeptidic protease inhibitor (PI) containing a stereochemically defined fused cyclopentanyltetrahydrofuran potent against multi-PI-resistant human immunodeficiency virus type 1 in vitro. *Antimicrobial Agents and Chemotherapy*, 53(3), 997-1006. https://doi.org/10.1128/AAC.00689-08

Kovalevsky, A. Y., Louis, J. M., Aniana, A., Ghosh, A. K., & Weber, I. T. (2008). Structural Evidence for Effectiveness of Darunavir and Two Related Antiviral Inhibitors against HIV-2 Protease. *Journal of Molecular Biology*, *384*(1), 178-192. https://doi.org/10.1016/j.jmb.2008.09.031

Kufareva, I., & Abagyan, R. (2012). Methods of protein structure comparison. *Methods in Molecular Biology,* 857, 231-257. https://doi: 10.1007/978-1-61779-588-6_10

Menéndez-Arias, L. (2013). Molecular basis of human immunodeficiency virus type 1 drug resistance: Overview and recent developments. *Antiviral Research,* 98(1), 93-120. https://doi.org/10.1016/j.antiviral.2013.01.007

Makatini, M. M., Petzold, K., Sriharsha, S. N., Ndlovu, N., Soliman, M. E., Honarparvar, B., … Govender, T. (2011). Synthesis and structural studies of pentacycloundecane-based HIV-1 PR inhibitors: a hybrid 2D NMR and docking/QM/MM/MD approach. *European Journal of Medicinal Chemistry*, 46(9), 3976-3985. https://doi.org/10.1016/j.ejmech.2011.05.071

Ollitrault, G., Fartek, S., Descamps, D., Camproux, A. C., Visseaux, B., & Regad, L. (2018). Characterization of HIV-2 protease structure by studying its asymmetry at the different levels of protein description. *Symmetry*, *10*, 644. https://doi.org/10.3390/sym10110644

Paulsen, J. L., Leidner, F., Ragland, D. A., Kurt Yilmaz, N., Schiffer, C. A. (2017). Interdependence of inhibitor recognition in HIV-1 protease. *Journal of Chemical Theory and Computation,* 13(5), 2300-2309. https://doi.org/10.1021/acs.jctc.6b01262

Pawar, S., Wang, Y. F., Wong-Sam, A., Agniswamy, J., Ghosh, A. K., Harrison, R. W., Weber, I. T. (2019). Structural studies of antiviral inhibitor with HIV-1 protease bearing drug resistant substitutions of V32I, I47V and V82I. *Biochemical and Biophysical Research Communications,* 514(3), 974-978. https://doi: 10.1016/j.bbrc.2019.05.064.

Prabu-Jeyabalan, M., Nalivaika, E., & Schiffer, C. A. (2000). How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *Journal of Molecular Biology,* 301(5), 1207-1220. https://doi.org/10.1006/jmbi.2000.4018

Priestle, J. P., Fässler, A., Rösel, J., Tintelnot-Blomley, M., Strop, P., & Grütter, M. G. (1995). Comparative analysis of the X-ray structures of HIV-1 and HIV-2 proteases in complex with CGP 53820, a novel pseudosymmetric inhibitor. *Structure*, *3*(4), 381-389. https://doi.org/10.1016/S0969-2126(01)00169-1

Raugi, D. N., Smith, R. A., Gottlieb, G. S., for the University of Washington-Dakar HIV-2 Study Group. (2016). Four Amino Acid Changes in HIV-2 Protease Confer Class-Wide Sensitivity to Protease Inhibitors. *Journal of Virology*, *90*(2), 1062-1069. https://doi.org/10.1128/jvi.01772-15

Regad, L., Chéron, J.B., Triki, D., Senac, C., Flatters, D., & Camproux, A.C. (2017) Exploring the potential of a structural alphabet-based tool for mining multiple target conformations and target flexibility insight. *PLoS One,* 12(8), e0182972 (2017). https://doi:10.1371/journal.pone.0182972

Rhee, S. Y., Taylor, J., Fessel, W.J., Kaufman, D., Towner, W., Troia, P., … Shafer, R. W. (2010) HIV-1 protease mutations and protease inhibitor cross-resistance. *Antimicrobial Agents and Chemotherapy*, 54(10), 4253-4261. https://doi.org/10.1128/AAC.00574-10.

Rosé, J. R., Salto, R., & Craik, C. S. (1993). Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions. *Journal of Biological Chemistry*, 268(16), 11939-11945.

Sadiq, S. K., & de Fabritiis, G. (2010). Explicit solvent dynamics and energetics of HIV-1 protease flap opening and closing. *Proteins: Structure, Function and Bioinformatics*, *78*(14), 2873-2885. https://doi.org/10.1002/prot.22806

Sardana, V. V., Schlabach, A. J., Graham, P., Bush, B. L., Condra, J. H., Culberson, J. C., … Culberson, J. C. (1994). Human Immunodeficiency Virus Type 1 Protease Inhibitors: Evaluation of Resistance Engendered by Amino Acid Substitutions in the Enzyme's Substrate Binding Site. *Biochemistry*, *33*(8), 2004-2010. https://doi.org/10.1021/bi00174a005

Shaw, W. H., Lin, Q., Muhammad, Z. Z., Lee, J. J., Khong, W. X., Ng, O. T., Tan, E. L., & Li, P. (2016). Identification of HIV Mutation as Diagnostic Biomarker through Next Generation Sequencing. *Journal of Clinical and Diagnostic Research.* 10(7), DC04-DC08. https://doi.org/10.7860/JCDR/2016/19760.8140

Shen, C-H., Wang, Y-F., Kovalevsky, A. Y., Harrison, R. W., Weber, I. T. (2010). Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters. *FEBS Journal,* 277(18), 3699-3714. https://doi.org/10.1111/j.1742-4658.2010.07771.x.

Tie, Y., Wang, Y. F., Boross, P. I., Chiu, T. Y., Ghosh, A. K., Tozser, J., … Weber, I. T. (2012). Critical differences in HIV-1 and HIV-2 protease specificity for clinical inhibitors.

Protein Science, *21*(3), 339-350. https://doi.org/10.1002/pro.2019

Tong, L, Pav, S., Mui, S., Lamarre, D., Yoakim, C., Beaulieu, P., & Anderson, P. C. (1995). Crystal structures of HIV-2 protease in complex with inhibitors containing the hydroxyethylamine dipeptide isostere. *Structure (London, England : 1993)*, *3*(1), 33-40. https://doi.org/10.1016/S0969-2126(01)00133-2

Tong, L., Pav, S., Pargellis, C., Florence, D., Lamarre, D., & Anderson, P. C. (1993). Crystal structure of human immunodeficiency virus (HIV) type 2 protease in complex with a reduced amide inhibitor and comparison with HIV-1 protease structures. *Proceedings of National Academy of Sciences U S A*, *90*(18), 8387-8391. https://doi.org/10.1073/pnas.90.18.8387

The PyMOL Molecular Graphics System, Version 1.8 Schrödinger, LLC.

Triki, D., Billot, T., Visseaux, B., Descamps, D., Flatters, D., Camproux, A. C., & Regad, L. (2018). Exploration of the effect of sequence variations located inside the binding pocket of HIV-1 and HIV-2 proteases. *Scientific Reports*, *8*:5789. https://doi.org/10.1038/s41598-018-24124-5

Triki, D., Cano Contreras, M. E., Flatters, D., Visseaux, B., Descamps, D., Camproux, A. C., & Regad, L. (2018). Analysis of the HIV-2 protease's adaptation to various ligands: Characterization of backbone asymmetry using a structural alphabet. *Scientific Reports*, *8*:710. https://doi.org/10.1038/s41598-017-18941-3

Triki, D., Fartek, S., Visseaux, B., Descamps, D., Camproux, A. C., & Regad, L. (2018). Characterizing the structural variability of HIV-2 protease upon the binding of diverse ligands using a structural alphabet approach. *Journal of Biomolecular Structure and Dynamics*, 37(17), 4658-4670. https://doi.org/10.1080/07391102.2018.1562985

Tzoupis, H., Leonis, G., Mavromoustakos, T., Papadopoulos M.G. (2013). A Comparative Molecular Dynamics, MM-PBSA and Thermodynamic Integration Study of Saquinavir Complexes with Wild-Type HIV-1 PR and L10I, G48V, L63P, A71V, G73S, V82A and I84V Single Mutants. *Journal of Chemical Theory and Computation*, 9(3), 754-764. https://doi.org/10.1021/ct301063k.

Venables, W.N., & Ripley, B.D. (2002). Modern Applied Statistics with S, Springer. https://doi.org/10.1007/978-0-387-21706-2

van Westen, G.J., Wegner, J.K., Bender, A., Ijzerman, A.P., & van Vlijmen, H.W. (2010). Mining protein dynamics from sets of crystal structures using "consensus structures". *Protein Science*. 19, 742-752. https://doi.org/10.1002/pro.350

Weber, I.T., Kovalevsky, A.Y., & Harrison, R.W. (2007). Structures of HIV protease guide inhibitor design to overcome drug resistance. In Frontiers in Drug Design and Discovery (Caldwell, G.W., Attaur-Rahman, Player, M.R. & Choudhary, M.I., eds), vol. 3, pp. 45-62(18). Bentham Science Publishers Sharjah, UAR; Bussum, The Netherlands; OakPark,IL, USA; Karachi, Pakistan. https://doi.org/10.2174/978160805201110703010045

Zhang, J., Hou, T., Wang, W., Liu, J. S. (2010). Detecting and understanding combinatorial mutation patterns responsible for HIV drug resistance. *Proceedings of the National Academy of Sciences,* 107(4),1321-1326. https://doi.org/10.1073/pnas.0907304107

Zoete, V., Michielin, O., & Karplus, M. (2002). Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for the analysis of protein flexibility. *Journal of Molecular Biology*, *315*(1), 21-52. https://doi.org/10.1006/jmbi.2001.5173

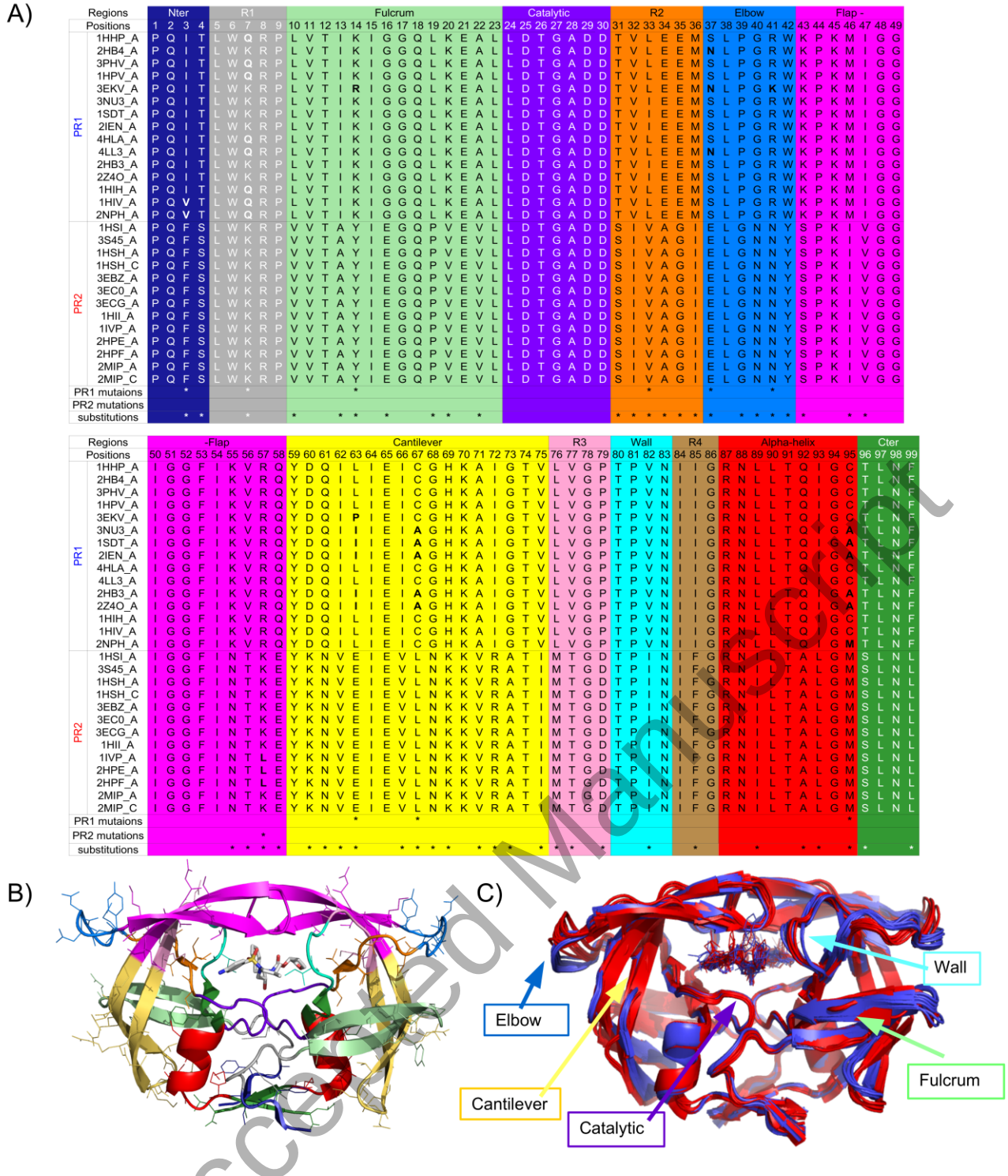
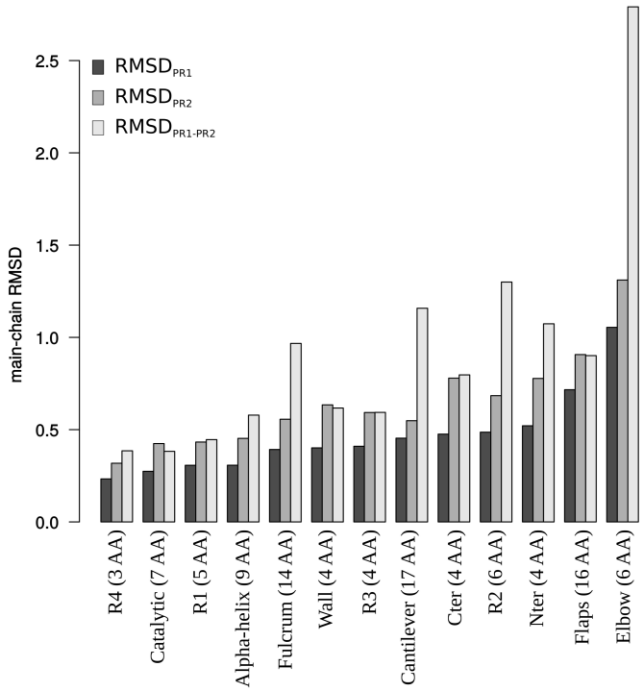


FIGURE 1. Presentation of the PR1 and PR2 structure sets. A) Multiple alignment of 24 PR1 and PR2 sequences. The alignment is coloured according to the 13 extracted PR regions. B) Visualization of the three-dimensional structure of PR2 complexed with APV (PDB code: 3S45). The PR2 structure is displayed in cartoon mode and coloured according to the 13 extracted PR regions. Substituted residues between PR1 and PR2 are displayed as sticks. The colour code for

each PR region is the same than used in Figure 1A. C) Superimposition of the 24 PR structures. PR structures are displayed as cartoons and coloured according to PR type and form. Bound PR1 and PR2 structures are coloured in blue and red, respectively. Bound ligands are displayed as lines.

A)



B)

	Structurally conserved regions in the PR1 and PR2 sets				Structurally weakly variable regions in the PR1 and PR2 sets								
Alternative hypothesis	**Catalytic**	**R1**	**R4**	**α-helix**	**R3**	**Flap**	**Cter**	**Wall**	**Cantilever**	**R2**	**Nter**	**Fulcrum**	**Elbow**
$RMSD_{PR1} < RMSD_{PR2}$	**5.53E-20**	**1.60E-13**	**5.69E-09**	**4.41E-12**	**1E-12**	**5.47E-7**	**4.94E-14**	**1.15E-14**	**6.75E-05**	**3.63E-011**	**2.76E-10**	**7.11E-11**	**6.83E-05**
$RMSD_{PR1-PR2} \neq RMSD_{PR2}$	**2.95E-03**	**0.33**	**1.83E-06**	**4.82E-11**	0.98	0.87	0.62	0.5	**3.80E-90**	**4.15E-62**	**5.93E-15**	**9.77E-47**	**1.42E-66**
$RMSD_{PR1-PR2} > RMSD_{PR2}$	0.99		**9.17E-07**	**2.41E-11**					**1.64E-90**	**2.07E-62**	**2.97E-15**	**4.89E-47**	**7.10E-67**
$RMSD_{PR1-PR2} < RMSD_{PR2}$	**1.44E-03**		> 0.99	> 0.99					> 0.99	> 0.99	> 0.99	> 0.99	> 0.99

FIGURE 2. Quantification and comparison of the structural variability of the 13 PR regions. (A) Distribution of the average $RMSD_{PR1}$, $RMSD_{PR2}$, and $RMSD_{PR1-PR2}$ values (in Å) for each region. The average $RMSD_{PR1}$ and $RMSD_{PR2}$ values quantify the intra-structural variability of each region in the PR1 and PR2 sets, respectively. $RMSD_{PR1-PR2}$ values quantify the inter-structural variability of each region between the PR1 and PR2 sets. The size of each region in terms of the number of amino acids (AA) is indicated in brackets. (B) Comparison of the intra- and inter-structural variability of each region. The table provides the p-value from the t-tests, allowing comparison of average $RMSD_{PR1}$, $RMSD_{PR2}$, and $RMSD_{PR1-PR2}$ values. For comparison of the average $RMSD_{PR1}$ and $RMSD_{PR2}$ values, we performed a t-test using the less alternative hypothesis ($RMSD_{PR1} < RMSD_{PR2}$, first line). For comparison of the average $RMSD_{PR1-PR2}$ and $RMSD_{PR1}$ values, two t-tests were performed using the two-sided alternative hypothesis

($RMSD_{PR1-PR2} \neq RMSD_{PR2}$, second line): greater alternative hypothesis ($RMSD_{PR1-PR2} > RMSD_{PR2}$, third line) and less alternative hypothesis ($RMSD_{PR1-PR2} < RMSD_{PR2}$, fourth line).

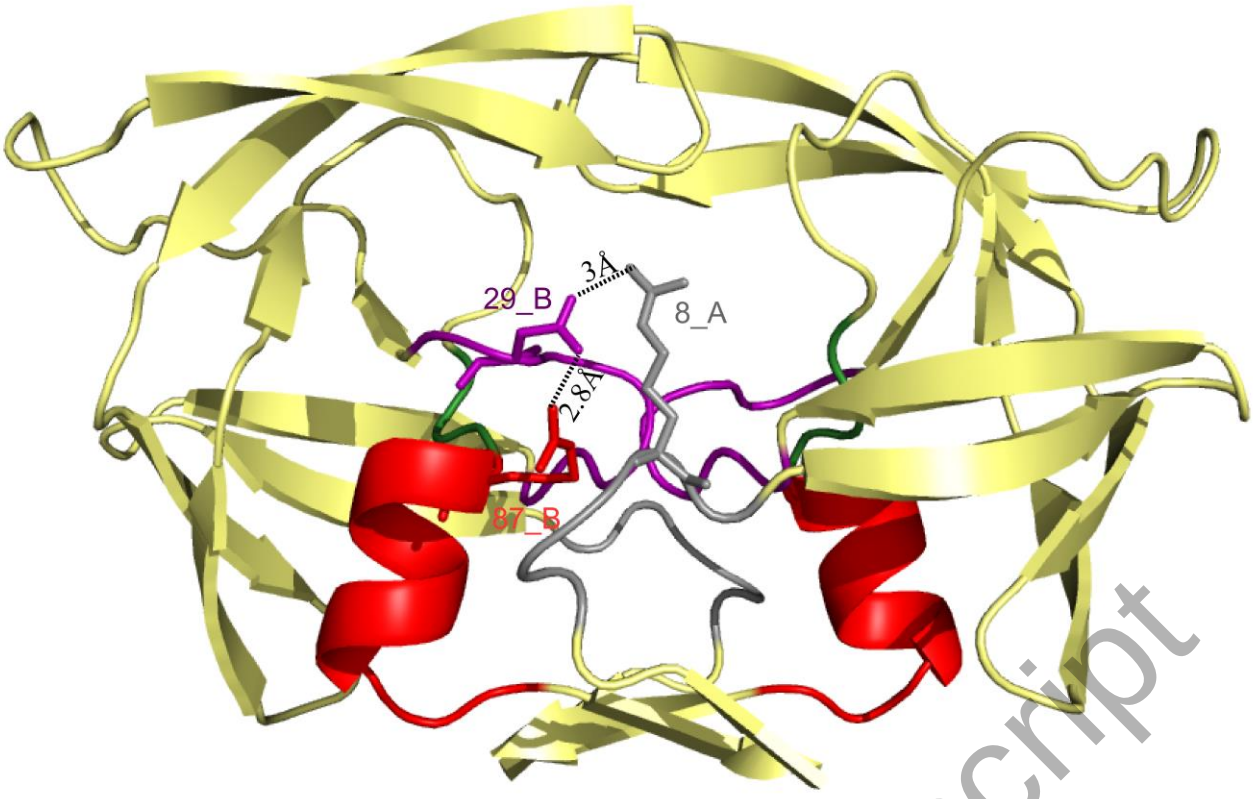


FIGURE 3. Visualization of the potential interactions between catalytic residues and residues of the α-helix and R1 regions. The PR2 structure corresponds to the PDB code 3S45 (PR2 in complex with APV). The PR2 structure is displayed in cartoons and coloured in yellow. The catalytic region is coloured in magenta; the R1 region is coloured in grey; the R4 region is coloured in green; and the α-helix region is coloured in red. Residues of these four regions, putatively interacting with each other, are displayed in sticks.

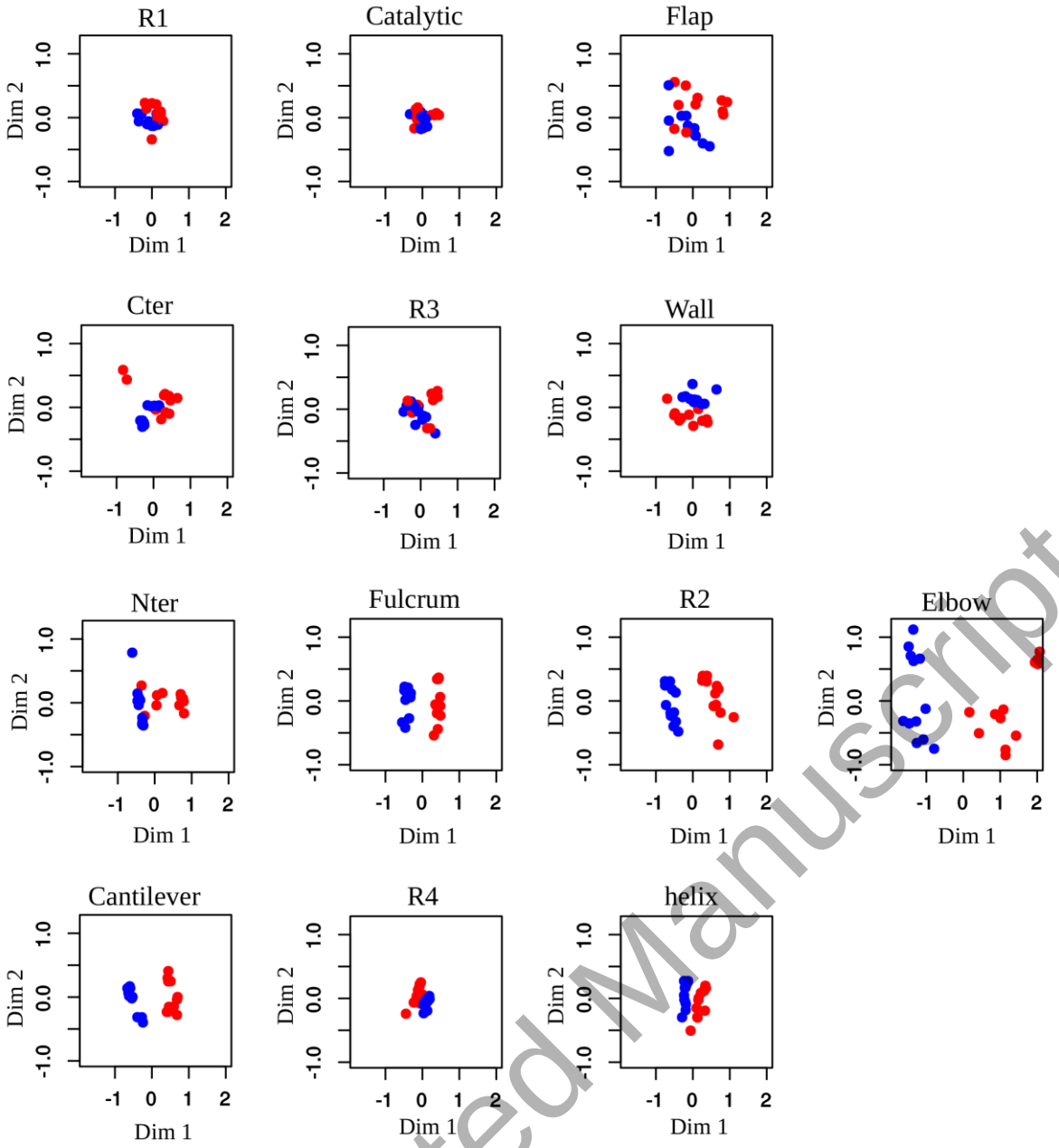


FIGURE 4. Multidimensional scaling (MDS) maps of conformations of the thirteen regions in the PR1 and PR2 sets. The distance in RMSD between each pair of conformations is projected onto two dimensions, retaining relative distance relationships so that two structurally similar conformations tend to be located near each other. The conformations extracted from PR1 structures are coloured in blue dots, and those extracted from PR2 are coloured in red dots.

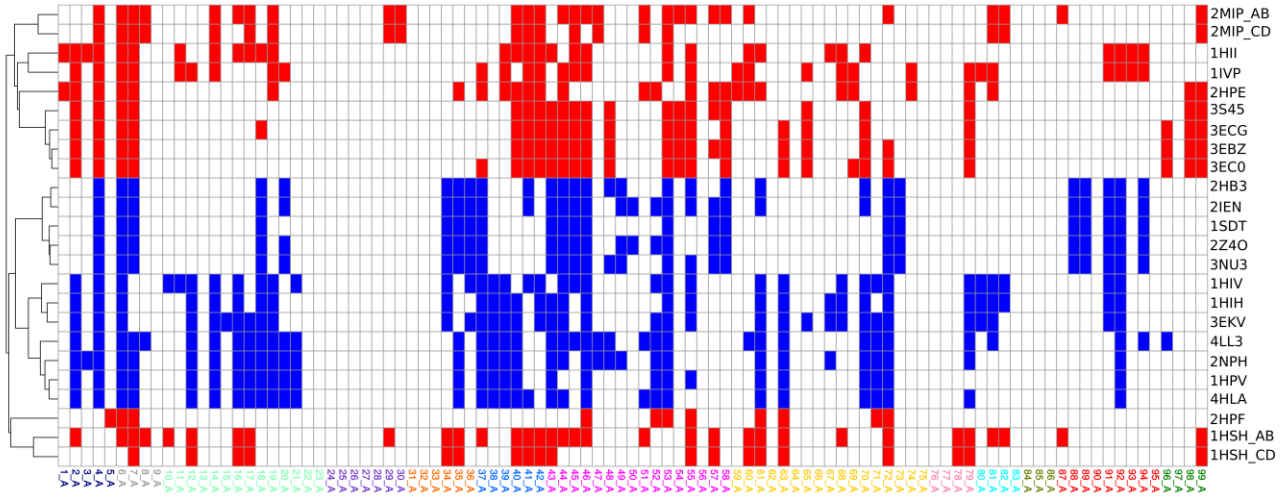


FIGURE 5. Residues of chain A involved in crystal packing in PR1 and PR2. PRs are presented in lines, and residues are presented in columns. Residues involved in crystal packing extracted from the PR1 and PR2 structures are coloured in blue and red, respectively. Residues are coloured according to the PR regions (see Figure 1 for the legend). Proteins were classified according to their residues involved in crystal packing using a binary distance and the complete aggregation method.