Article

Characterization of HIV-2 Protease Structure by Studying Its Asymmetry at the Different Levels of Protein Description

Guillaume Ollitrault [1], Sandrine Fartek [1], Diane Descamps [2], Anne-Claude Camproux [1], Benoît Visseaux [2] and Leslie Regad [1,*]

[1] Molécules thérapeutiques in silico (MTi), UMR-S973, Université Paris Diderot, Sorbonne Paris Cité, INSERM, F-75013 Paris, France; guillaume.ollitrault@gmail.com (G.O.); fartek.sandrine@gmail.com (S.F.); anne-claude.camproux@univ-paris-diderot.fr (A.-C.C.)

[2] Infection, Antimicrobiens, Modélisation, Evolution (IAME), UMR 1137, Université Paris Diderot, Sorbonne Paris Cité; INSERM; AP-HP, Hôpital Bichat, Virologie, 75018 Paris, France; diane.descamps@aphp.fr (D.D.); benoit.visseaux@aphp.fr (B.V.)

* Correspondence: leslie.regad@univ-paris-diderot.fr; Tel.: +33-157-278-272

check for updates

Abstract: HIV-2 protease (PR2) is a homodimer, which is an important target in the treatment of the HIV-2 infection. In this study, we developed an in silico protocol to analyze and characterize the asymmetry of the unbound PR2 structure using three levels of protein description by comparing the conformation, accessibility, and flexibility of each residue in the two PR2 chains. Our results showed that 65% of PR2 residues have at least one of the three studied asymmetries (structural, accessibility, or flexibility) with 10 positions presenting the three asymmetries in the same time. In addition, we noted that structural and flexibility asymmetries are linked indicating that the structural asymmetry of some positions result from their large flexibility. By comparing the structural asymmetry of the crystallographic and energetically minimized structures of the unbound PR2, we confirmed that the structural asymmetry of unbound PR2 is an intrinsic property of this protein with an important role for the PR2 deformation upon ligand binding. This analysis also allowed locating asymmetries corresponding to crystallization artefacts. This study provides insight that will help to better understand the structural deformations of PR2 and to identify key positions for ligand binding.

Keywords: protein asymmetry; HIV-2 structural asymmetry; residue accessibility; protein flexibility; structural alphabet

1. Introduction

The human immunodeficiency virus of type 1 (HIV-1) and type 2 (HIV-2) are two etiological causative agents of AIDS (acquired immune deficiency syndrome). HIV-1 is observed worldwide, while HIV-2 is more restricted to West Africa and infects 1 to 3 million of people. The therapeutic arsenal is developed for HIV-1 and targets various viral proteins: integrase, reverse transcriptase, membrane fusion proteins, and protease (PR). However, a third of available drugs have no action against HIV-2. Indeed, HIV-2 is naturally resistant to all non-nucleoside reverse transcriptase inhibitors and fusion inhibitors. It also demonstrated reduced susceptibility to PR inhibitors (PIs) [1–6]. Thus, there is still a strong need today for developing new molecules specifically designed for HIV-2. To do so, it is important to better characterize PR2 structural particularities, as this protein has not been as deeply studied as HIV-1 protease (PR1).

PR2 is an aspartic protease hydrolyzing the viral precursor polyproteins (Gag and the Gag-Pol) during the maturation of viral particles. Like all aspartic proteases, PR2 contains the catalytic triplet Asp-Thr-Gly. PR2 is a C2-symmetric homodimer with 99 residues in each monomer. The PR2 is capable of recognizing diverse ligands, such as various non-homologous substrates and chemically dissimilar inhibitors [7]. Substrates and inhibitors bind the PR2 at the interface of the two monomers. The binding of these ligands, that are often asymmetric, induce large structural deformation corresponding to the transition from a semi-open form, allowing the ligand entry, to a closed form, allowing the catalytic action [7].

It has been previously shown that a dimer could modify the conformation of its side-chain or main-chain atoms in one monomer to recognize diverse ligands, resulting in structural asymmetries in the dimer. Those structural asymmetries, defined as differences in side-chain or backbone conformations of residues in the two dimer chains, should play a role in the adaptive recognition of ligands [8,9], in the differentiation of high and low-affinity binding sites [10], or of active from non-active binding sites [11]. For example, it has been demonstrated that the binding of six substrates on the HIV-1 protease breaks the target symmetry. This allows the PR2 to adapt its conformation to recognize non-homologous substrates [12]. Moreover, the specificity of this target for its substrates seems to link to the recognition of an asymmetric shape rather than the recognition of a particular substrate amino-acid sequence [12,13]. The link between PR2 specificity for PIs and its structural asymmetry is, to date, imperfectly characterized. Structural asymmetry has been previously detected in PR2, particularly in the tail, elbow and flap regions [14–18]. In a recent work, we detected structural asymmetry in the 18 available crystallographic structures of the PR2 dimer in complex with diverse ligands [18]. We located positions having different backbone conformations between the two chains in the 18 PR2 dimer structures using the HMM-SA structural alphabet (Hidden Markov Model—Structural Alphabet) [19]. This analysis enabled us to distinguish the structural asymmetry conserved across most PR2 dimers from the structural asymmetry specific to some complexes. In addition, we located structural asymmetry linked to the PR2 flexibility or putatively induced by ligand binding. This latter asymmetry allows the adaptation of PR2 binding site to bind diverse ligands and thus is important for the recognition of various ligands.

In this study, we focused on deeper analysis of asymmetry of the unbound PR2, i.e., not complexed with a ligand, by taking into account three levels of the characterization of PR2 structure: local conformations, accessible surface area, and flexibility. Firstly, we analyzed the structural asymmetry in the crystallographic structure of the unbound PR2 by detecting residues presenting different local conformations using the HMM-SA-based approach that we previously developed [18]. Secondly, we analyzed PR2 asymmetry by focusing on residue accessible surface area (ASA) that is used to define surface exposed residues often crucial for interactions with other proteins and buried residues contributing in the stability of the tertiary structure [20]. We detected residues having different ASA values in the two PR2 chains to identify the PR2 accessibility asymmetry. Thirdly, we compared the flexibility of each residue in both PR2 chains quantified by B-factor values [16,21] to highlight residues exhibiting flexibility asymmetry. The potential links between these three asymmetries (conformational, accessibility, and flexibility asymmetries) were then explored using a multivariate analysis method. Subsequently, we detected asymmetric positions that could correspond to crystallization artefacts resulting from crystal packing contacts [14,16,17]. This step consisted of crossing the detected structural asymmetric positions with the packing crystal positions defined after generating PR2 symmetric mates. Then, we compared the structural asymmetry detected in the original Protein Data Bank (PDB) structure of the unbound PR2 with the one extracted from minimized structures of the unbound PR2 generated to eliminate non-biological relevant contacts [22]. Our results cannot only provide a precise description of the structure of the unbound PR2 through its asymmetry, but also reveals insights about its flexibility and mechanisms induced in ligand recognition. This information is crucial for designing potent inhibitors of PR2.

2. Results & Discussion

2.1. Localization of Asymmetry in the PR2 Structure

2.1.1. Structural Asymmetry in PR2

Using two approaches, we analyzed the structural asymmetry of the unbound PR2 structure, i.e., the differences between its chain A and B conformations. Firstly, we compared the global conformation of the two optimally superimposed chains of the unbound PR2 structure (PDB code: 1HSI) using the root mean square deviation computed between their carbons α (Cα-RMSD). The two PR2 chains have a Cα-RMSD of 0.39 Å with larger deviations observed in the Nter region, the region near residue 50, and the region near residue 39 (Figure 1). This agrees with structural asymmetry previously detected in PR2 complexed with various ligands [14,16,17].
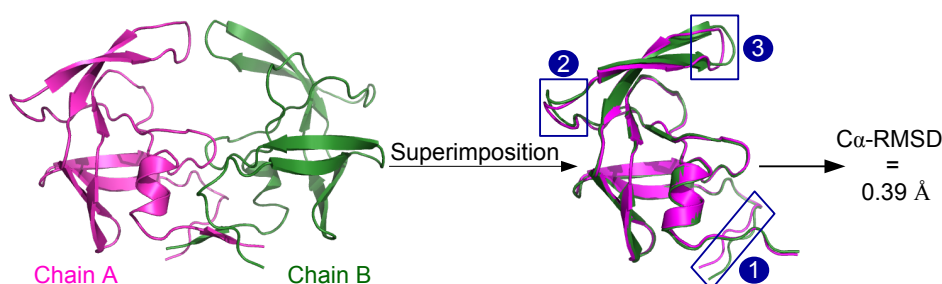

Figure 1. Characterization of the global structural asymmetry of the PR2 dimer corresponding to the PDB code 1HSI [23]. Protein is displayed in cartoon mode. Chains A and B are colored in magenta and green, respectively. Rectangles highlight regions with larger deviations between both chains: region 1 corresponds to Nter region, region 2 corresponds to region near residue 39, and region 3 corresponds to region near residue 50.

The unbound PR2 chains have a smaller Cα-RMSD than those of bound PR2, as it was evaluated at 0.53 Å with the same parameter for PR2 in complex with CGP 53820 (PDB code: 1HIH) [17]. However, the large deviations observed between both unbound PR2 chains in the Nter region, the region near residue 50, and region near residue 39 were all also observed in bound PR2 structures. This indicates that, although the two chains of PR2 have the same sequence, they do not exhibit the same backbone conformations even in an unbound conformation and this structural asymmetry is increased when the PR2 is in complex with ligands.

Secondly, we detected structural asymmetric positions, i.e., positions exhibiting different conformations in both chains using the HMM-SA-based method that we previously developed [18]. This method, presented in Figure 2, is based on the structural alphabet HMM-SA [19,24]. It compares the local conformations of each residue in chains A and B after simplifying the 3D structures of the two dimer chains into two sequences of structural letters, each structural letter describing the geometry of a 4-Cα fragment. From these two structural-letter sequences, structural asymmetric positions were located by identifying positions presenting different structural letters between chains A and B (Figure 2). Among the 96 positions of the unbound PR2 where a structural letter was determined by HMM-SA, 62 (65%) exhibit the same local conformation in the two PR2 chains, named $noAsym^{Struct}$ positions. The remaining 34 positions have different local conformations in the two chains and are herein named asymmetric positions and noted $asym^{Struct}$. These $asym^{Struct}$ positions represent 35% of the PR2 positions, confirming that the unbound PR2 presents backbone asymmetry. They are observed in regular secondary structures as well as in loop regions. These $asym^{Struct}$ positions are located all along the sequence, particularly in the elbow, flap, and cantilever regions (Figure 2). Several studies have shown that the region near residue 50 (flap) presents the largest structural asymmetry in PR2 complexed with inhibitors [15,16]. In their study, Mulichak et al. (1993) suggested that structural

asymmetry occurring at residues 42–53 is explained by ligand binding [14]. However, HMM-SA detected structural asymmetry in 8 out of these 12 residues in the unbound PR2 (Figure 2). Thus, the structural asymmetry of region 42–53 already exists in the unbound PR2 but could be modified upon ligand binding.
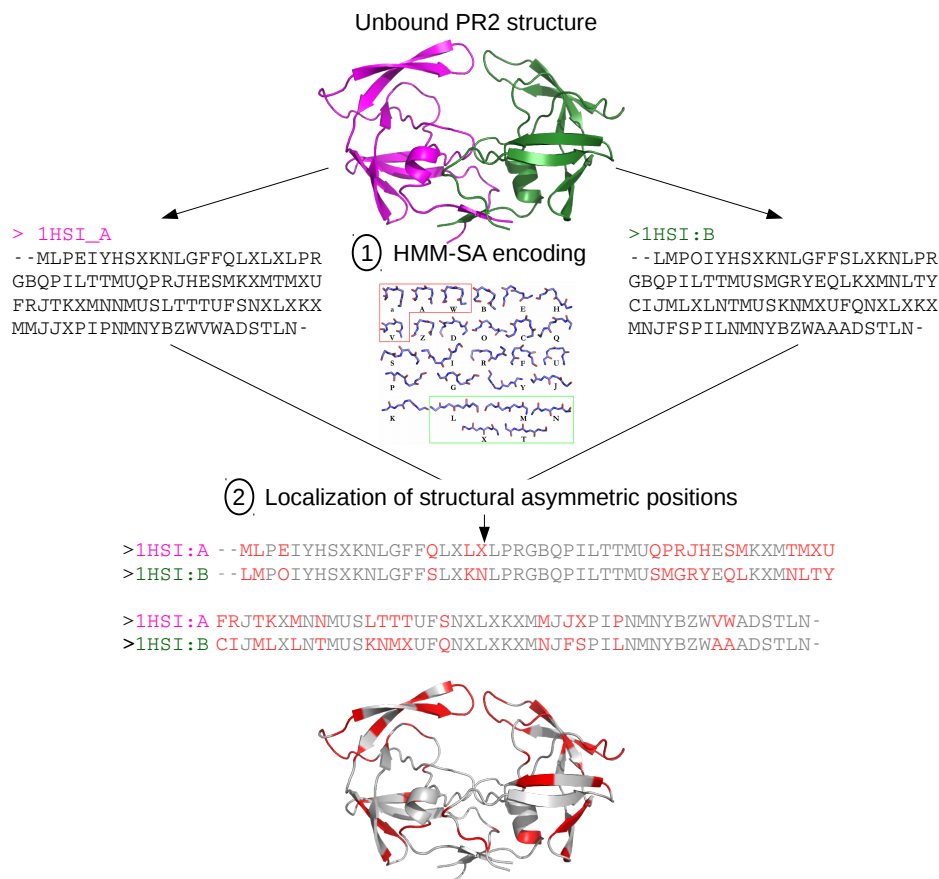

Figure 2. Identification of local structural asymmetry in the unbound PR2 (PDB code: 1HSI [23]) using the HMM-SA-based approach. HMM-SA is a library of 27 prototypes of protein 4-Cα fragments, named structural letters, classified according to their geometry. It simplifies the 3D (tri-dimensional) structure of the two chains of the dimer into two sequences of structural letters, where each structural letter corresponds to the geometry of a 4-Cα fragments. The two obtained structural-letter sequences are compared to locate asymmetric positions defined as positions exhibiting different local conformations, i.e., structural letters. These positions were colored in red in the structural-letter sequences and in the 3D structure of the unbound PR2.

Then, we explored if some structural letters are specific to $asym^{Struct}$ positions. Figure 3a presents the distribution of structural letters at $asym^{Struct}$ and $noAsym^{Struct}$ positions and shows that most structural letters are observed both at $asym^{Struct}$ and $noAsym^{Struct}$ positions. By considering the occurrences of each structural letter, we showed that the loop structural letter Q and the β-strand structural letter T are more observed at $asym^{Struct}$ positions than $noAsym^{Struct}$ positions. In addition, loop structural letters Z, B, and D are identified only at $noAsym^{Struct}$ positions (Figure 3a). In contrast, loop structural letters C, O, and α-helix structural letter V are specific to $asym^{Struct}$ positions. This result suggests that some $asym^{Struct}$ positions could exhibit particular conformations not observed at $noAsym^{Struct}$ positions, that could indicate particular deformation process. This hypothesis must be confirmed with more data.
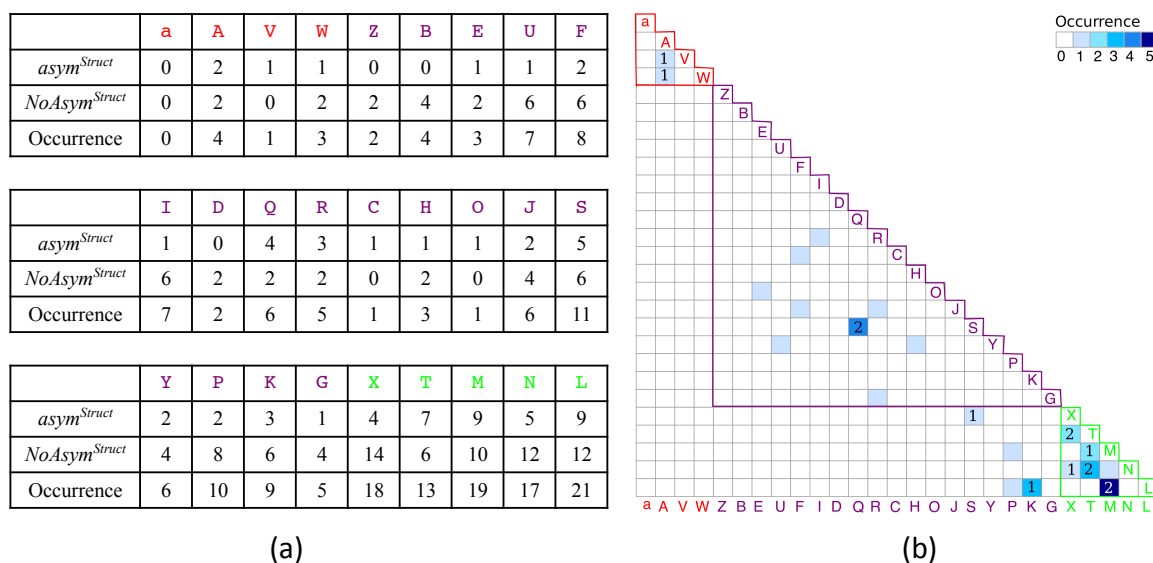
	a	A	V	W	Z	B	E	U	F
$asym^{Struct}$	0	2	1	1	0	0	1	1	2
$NoAsym^{Struct}$	0	2	0	2	2	4	2	6	6
Occurrence	0	4	1	3	2	4	3	7	8

	I	D	Q	R	C	H	O	J	S
$asym^{Struct}$	1	0	4	3	1	1	1	2	5
$NoAsym^{Struct}$	6	2	2	2	0	2	0	4	6
Occurrence	7	2	6	5	1	3	1	6	11

	Y	P	K	G	X	T	M	N	L
$asym^{Struct}$	2	2	3	1	4	7	9	5	9
$NoAsym^{Struct}$	4	8	6	4	14	6	10	12	12
Occurrence	6	10	9	5	18	13	19	17	21

(a)

(b)

Figure 3. Dependence between structural letters and structural asymmetry. (**a**) Occurrences of the 27 structural letters in $asym^{Struct}$ and $noAsym^{Struct}$ of the unbound PR2 structure. (**b**) Occurrences of the couples of structural letters extracted for a position in chains A and B. The color of each square indicates the occurrences of each couple. The darker is the blue, the more the couple is observed. Structural letters were colored according to the secondary structures they described: α-helix structural letters are colored in red, β-strand structural letters in green, and loop structural letters in purple. Numbers, when indicated, indicate the occurrences of structural-letter couples at $asym^{Struct}$ positions putatively induced by crystallization.

Finally, we analyzed if some deformations are recurrent in the unbound PR2. To do so, we counted the couple of structural letters observed in chains A and B at each $asym^{Struct}$ position. From the 34 $asym^{Struct}$ positions, we extracted 21 different couples of structural letters (Figure 3b). Most of them (71%) are observed at only one $asym^{Struct}$ position and six are recurrent. For example, the structural couple L-M were found at five $asym^{Struct}$ positions, indicating that five positions have the structural letter L in a chain and the structural letter M in the other (Figure 3b). In addition, most structural changes between chains A and B occur between close structural letters without secondary-structure change; only three couples correspond to structural letters specific to different secondary structures: L-P (1), M-P (1), L-K (2). This indicates that most structural deformations between the two chains in the unbound PR2 are of weak magnitude.

2.1.2. Asymmetry in Terms of Accessibility in PR2

Detecting surface exposed and buried residues based on their ASA values helps to better understand the role of residues in the protein structural integrity [25,26]. Exposed residues are often involved in interactions with partners, while buried residues is important for the structural stability [20]. We compared the accessibility of each residue in both unbound PR2 chains by comparing ASA values of each Cα in the two chains. Cα ASA values were computed using NACCESS program and presented in Figure A1. The two PR2 chains have the same average ASA value (*t*-test *p*-value = 0.95) that indicates that the unbound PR2 structure does not exhibit global accessibility asymmetry.

To analyze the local accessibility asymmetry, we computed the difference between the ASA value of each residue in chains A and B, noted $\delta(ASA)^{A-B}$ (Figure 4a). A total of 56 positions present $\delta(ASA)^{A-B}$ values different to 0 Å^2. We defined accessibility asymmetric positions as positions having a $|\delta(ASA)^{A-B}|$ higher than 1 Å^2. Using this definition, 31 positions were characterized as accessibility asymmetric positions and noted $asym^{ASA}$ positions, and 68 positions as non-accessibility asymmetric, noted $noAsym^{ASA}$. Seven positions present a large $|\delta(ASA)^{A-B}|$, higher than 5 Å^2 (Figure 4). The 31

$asym^{ASA}$ positions are located all along the sequence (Figure 4b). Like the structural asymmetry, accessibility asymmetry occurs in loop regions as well as in regular secondary structures (Figure 4b). In addition, we noted that $asym^{ASA}$ positions have significant higher ASA values than $noAsym^{ASA}$ positions (*t*-test *p*-value = $9e^{-10}$), indicating that $asym^{ASA}$ positions correspond to exposed residues.
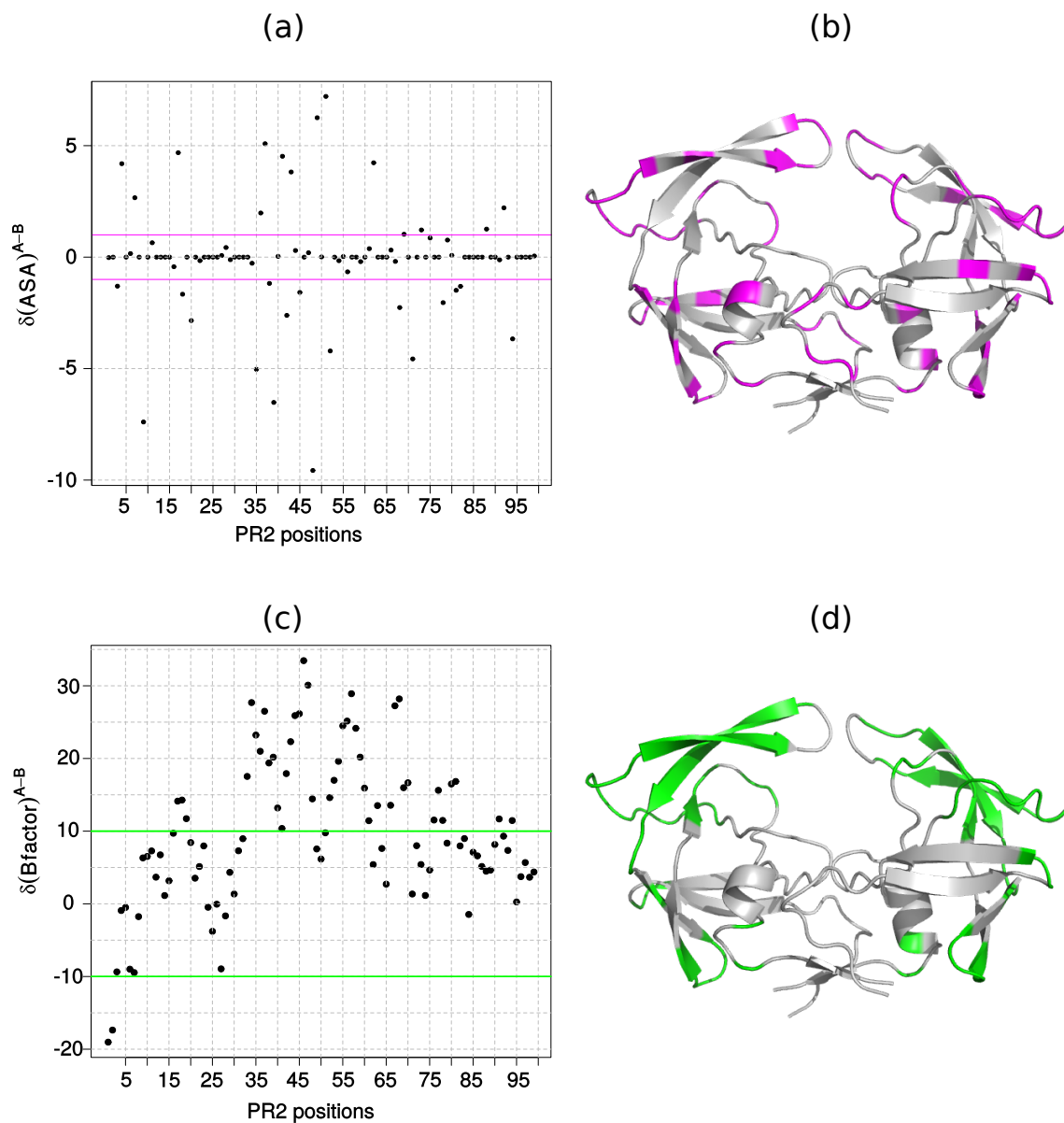

Figure 4. Localization of accessibility and flexibility asymmetry in PR2. (**a**,**b**) Localization of $asym^{ASA}$ positions. (**a**) $\delta(ASA)^{A-B}$ values for each PR2 position. Magenta lines correspond to the $\delta(ASA)^{A-B}$ cut-off values used to define $asym^{ASA}$ positions. (**b**) Representation of the structure of the unbound PR2 in cartoon mode. $asym^{ASA}$ positions are colored in magenta. (**c**,**d**) Localization of $asym^{Bfactor}$ positions. (**c**) $\delta(Bfactor)^{A-B}$ values for each PR2 position. Green lines correspond to the $\delta(Bfactor)^{A-B}$ cut-off values used to define $asym^{Bfactor}$ positions. (**d**) Representation of the structure of the unbound PR2 in cartoon mode. $asym^{Bfactor}$ positions are colored in green.

2.1.3. Asymmetry in Terms of Flexibility in PR2

To better understand the flexible mechanisms of PR2, we analyzed the flexibility asymmetry of the unbound PR2. Indeed, it has been shown that the intrinsic flexibility of PR2 is crucial for ligand binding (substrates and PIs) and its activity [22,27] and could have a role in ligand specificity [22]. In addition, previous studies showed that local asymmetry intrinsically influence the target structure to adopt different conformations [28,29], particularly in PR1 [12]. The flexibility of each residue in the two chains of the unbound PR2 structure was measured using the B-factor values of all Cα atoms (Figure A2). Chain A presents a higher average value of B-factor than chain B (*t*-test *p*-value $< 10^{-4}$), indicating that chain A of unbound PR2 is more flexible than chain B. This agrees with the higher flexibility of chain A previously detected in the PR2 structure complexed with inhibitors [16].

Then, we compared the flexibility of each residue in both chains by computing the difference of B-factor value between the two chains, noted $\delta(Bfactor)^{A-B}$, for all Cα atoms (Figure 4c). All positions have different B-factor values in the two chains, i.e., $|\delta(Bfactor)^{A-B}| \neq 0$ Å^2. A total of 44 positions have a $|\delta(Bfactor)^{A-B}|$ value higher than 10 Å^2 and were defined as flexibility asymmetric positions, noted $asym^{Bfactor}$, while the 55 remaining positions were characterized as non-flexibility asymmetric and noted $noAsym^{Bfactor}$ (Figure 4c). $asym^{Bfactor}$ positions are mainly located in the elbow region, in the two β-strands of the flap region, and in the loop of the cantilever and fulcrum regions (Figure 4d). In addition, $asym^{Bfactor}$ positions correspond to higher flexible positions than $noAsym^{Bfactor}$ positions, i.e., $asym^{Bfactor}$ positions exhibit significant larger B-factor values than $noAsym^{Bfactor}$ positions (*t*-test *p*-value $< 10^{-09}$).

2.2. Link between the Three Asymmetry Types

Although the two PR2 chains present the same amino-acid sequence, we identified asymmetric positions in the unbound structure that exhibit different local conformations, ASA, and flexibility in the two PR2 chains. To explore the link between these three asymmetries, we studied these three asymmetries together using a multiple correspondence analysis (MCA). MCA allowed analyzing the pattern of relationships among the positions according to their asymmetries, see Material & Methods. Figure 5a presents the projection of the 96 PR2 positions and the six asymmetry values onto the two first MCA components. The proportion of total variance accounted for the two first components is at 75%. The first MCA component, explaining the larger part of variability, separates positions according to their asymmetric status and the second component separates positions according to their accessibility asymmetry (Figure 5a). The MCA map highlights a link between the structural and flexibility asymmetries. This is confirmed by the fact that $asym^{Struct}$ positions have a higher $\delta(B-factor)^{A-B}$ than $noAsym^{Struct}$ positions (*t*-test *p* value < 0.008). We also showed that the $asym^{Struct}$ positions that are also flexibility asymmetric are overrepresented (χ^2 *p*-value = 0.02). In contrast, no link between accessibility and structural asymmetries or between accessibility and flexibility asymmetries was observed. Structurally $asym^{Struct}$ and $noAsym^{Struct}$ positions exhibit similar $\delta(ASA)^{A-B}$ ASA value (*t*-test *p*-value = 0.54) as well as $asym^{Bfactor}$ and $noAsym^{Bfactor}$ positions (*t*-test *p*-value = 0.11).

The MCA map highlights eight position clusters that group PR2 positions exhibiting same asymmetries. The composition of each group and their asymmetric profile are presented in Figure 5b. Cluster 3, the largest cluster, groups 34 positions that exhibit no asymmetry. These positions are in the fulcrum, catalytic, wall, α-helix, and Cter regions and in the end of the cantilever region. 32% of these residues (11) are in the PR2 pocket, suggesting that the conserved conformations of these residues are important for ligand binding and PR activity. In our previous study of the PR2 asymmetry using 18 PR2-ligand complexes, we demonstrated that residues 14, 15, 75, 86, 93, 96 are structurally asymmetric in most PR2 structures [18]. Thus, these positions are not structurally asymmetric in the unbound PR2, but they become structurally asymmetric when the PR2 is complexed with a ligand. This highlights that ligand binding induces structural backbone deformation at these positions that causes the structural asymmetry observed at these positions in PR2 complexes. However, these residues are not located in

the PR2 ligand-binding pocket (data not shown). Thus, the structural deformation observed at these positions could result from long-range effects of ligand binding.
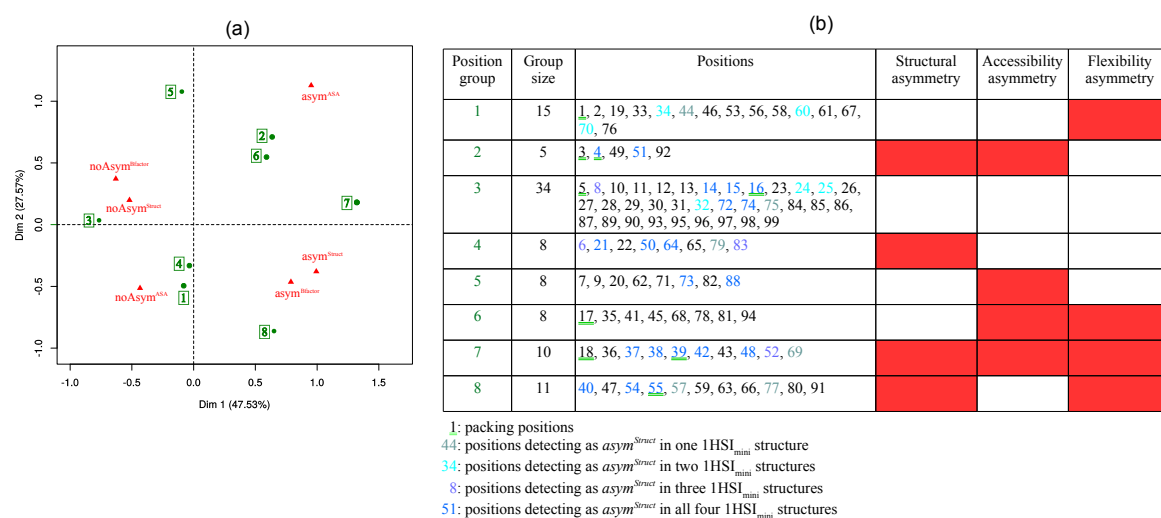
(a)　　　　　　　　　　　　　　　　　　　　(b)

Position group	Group size	Positions	Structural asymmetry	Accessibility asymmetry	Flexibility asymmetry
1	15	1, 2, 19, 33, 34, 44, 46, 53, 56, 58, 60, 61, 67, 70, 76			■
2	5	3, 4, 49, 51, 92	■	■	
3	34	5, 8, 10, 11, 12, 13, 14, 15, 16, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 72, 74, 75, 84, 85, 86, 87, 89, 90, 93, 95, 96, 97, 98, 99			
4	8	6, 21, 22, 50, 64, 65, 79, 83	■		
5	8	7, 9, 20, 62, 71, 73, 82, 88		■	
6	8	17, 35, 41, 45, 68, 78, 81, 94		■	■
7	10	18, 36, 37, 38, 39, 42, 43, 48, 52, 69	■	■	■
8	11	40, 47, 54, 55, 57, 59, 63, 66, 77, 80, 91	■		■

1: packing positions
44: positions detecting as $asym^{Struct}$ in one 1HSI$_{mini}$ structure
34: positions detecting as $asym^{Struct}$ in two 1HSI$_{mini}$ structures
8: positions detecting as $asym^{Struct}$ in three 1HSI$_{mini}$ structures
51: positions detecting as $asym^{Struct}$ in all four 1HSI$_{mini}$ structures

Figure 5. Relationship between the three asymmetry types. (**a**) Projection of the PR2 positions and the values of the three variables describing PR2 asymmetries into the first MCA plane explaining 75% of the data variability. Asymmetry values are symbolized in red triangle. Green points represent a set of PR2 positions exhibiting the same asymmetric behavior, named position clusters. (**b**) Composition and description of the eight position clusters deducted from the MCA first plane. Each position was also described according to its involvement in crystal packing and to the number of 1HSI$_{mini}$ structures where it was detected as structurally asymmetric.

Cluster 7 groups positions that present the three asymmetries and cluster 8 contains positions that are structurally and flexibility asymmetric in the same time. These positions correspond to flexible positions (high B-factor values, Figure A2) and are mainly located in the elbow and flap regions. This result suggests that the large PR2 flexibility could induce the structural asymmetry of these positions.

Clusters 1 and 6 residues correspond to residues that are flexibility asymmetric but not structurally asymmetric. These positions are mainly located in the flap and cantilever regions. It is not expected that flexibility changes will not be accompanied with structural change. Thus, one hypothesis is that HMM-SA tool could not be precise enough to capture these weak structural changes.

Clusters 4 and 2 group positions having structural asymmetry but not flexibility asymmetric. These positions are in the dimer, flap, fulcrum, and cantilever regions. Except the three positions located in flap, these positions are weakly flexible (B-factor value < 30, Figure A2). Thus, the structural variability observed at these positions do not seem to be linked to the intrinsic flexibility but could be mostly explained by PR2 dimerization or crystal packing.

2.3. Link between Asymmetry and Crystal Packing

As the 1HSI structure is a crystallographic structure, a part of the detected structural asymmetry could be induced by crystal packing and correspond to artefacts of crystallization as suggested by several studies focusing on the analysis of PR2 in complex with inhibitors [14,16,17]. To locate such non-biologically relevant structural asymmetry, we identified Cα atoms putatively involved in crystal packing contacts, i.e., Cα atoms located at less than 4.5 Å of the generated symmetric mates, see Material & Methods. In the unbound PR2 structure, nine positions have their Cα atoms putatively involved in crystal packing, named packing positions (Figure 5b). Five of these positions (3, 4, 18, 39, and 55) correspond to $asym^{Struct}$ and $asym^{Bfactor}$ positions, suggesting that the crystal packing could

induce structural changes accompanied with flexibility changes. This result confirms the suggested link between crystal packing and asymmetry of Nter region and region 37–41 detected in the PR2 in complex with U92163 and U75875 molecules, respectively [14].

Three packing positions (5, 16, and 17) exhibit no structural asymmetry indicating that the crystal packing occurring at these positions has non-impact on their local conformation, ASA, and flexibility of these residues. These results must be taken with caution as the structural letter observed in the two chains at positions 16 and 17 is F. This structural letter, specific of loop region, is associated with a large variability (RMSD between 4-Cα fragments corresponding to this letter is of 0.91 Å [19]) and could correspond to different conformations.

To pursue the analysis, we located structural asymmetric positions in four minimized structures of 1HSI, noted 1HSI$_{mini}$ structures, (Figures 5b and A3). The energetic minimization enabled removing crystal packing and contacts with no biological relevance in the PR2 dimer and was performed using GROMACS [30], see Material & Methods. The four minimized structures exhibit a larger global structural asymmetry with an average Cα-RMSD of 0.43 Å than the 1HSI structure (Table 1). However, in terms of local structural asymmetry, the four 1HSI$_{mini}$ structures exhibit less *asym^{Struct}* positions than 1HSI structure (Table 1). This suggests that there are less structural changes between the two chains in the 1HSI$_{mini}$ structures than in the 1HSI structure but they have higher magnitude.

Table 1. Quantification of the structural asymmetry in 1HSI and in the four 1HSI$_{mini}$ structures. Cα-RMSD corresponds to the RMSD computed between the two monomers of one dimer.

Structures	1HSI	1HSI$_{mini1}$	1HSI$_{mini2}$	1HSI$_{mini3}$	1HSI$_{mini4}$
Cα-RMSD (Å)	0.386	0.429	0.454	0.431	0.416
Number of *asym^{Struct}* positions	34	25	30	27	28

The comparison of localization of *asym^{Struct}* positions in 1HSI and 1HSI$_{mini}$ structures highlighted 14 positions that are *asym^{Struct}* in 1HSI structure and *noAsym^{Struct}* in all four 1HSI$_{mini}$ structures (Figure 5). Only two of these positions (3 and 18) are previously highlighted as *asym^{Struct}* positions corresponding to packing positions. This result reinforces that the structural asymmetry observed at positions 3 and 18 correspond to crystallographic artefacts. In contrast, the link between structural asymmetry at residues 4, 39, and 55 has not been confirmed by these results on the unbound PR2 structure, while it was observed in PR2 complexed with ligands [14]. Thus, either the structural asymmetry of these positions is not entirely involved by crystal packing or the method used for the energetic minimization was not adapted to remove all irrelevant contacts.

The 12 remaining positions being *asym^{Struct}* in 1HSI structure and *noAsym^{Struct}* in all four 1HSI$_{mini}$ structures do not correspond to packing positions (Figure 5). This suggests that their structural asymmetry is not directly induced by interactions with symmetric mates. Surprisingly, most of these positions are located at the beginning or the end of regular secondary structures (Figure A4). Their structural asymmetry could result from indirect effect of crystal packing. For example, asymmetric residues 59, 63, and 36 are located at less 5 Å of packing residues 39, 16, and 18, respectively. Thus, we suppose the structural asymmetry of these residues is induced by crystallization through direct and indirect effects. It is interesting to explore if such structural deformations observed at these *asym^{Struct}* putatively induced by crystallization are different to other deformations. To do so, we analyzed the occurrences of structural-letter couples observed in both chains at *asym^{Struct}* positions putatively induced by crystal packing (Figure 3b). Four structural-letter couples (X-T, X-N, X-S, and V-A) are observed only at *asym^{Struct}* positions putatively induced by crystallization that suggests that crystal packing is responsible for these particular deformations. This indicates that the crystallization method could produce particular local conformations as it was previously observed for nuclear magnetic resonance method [31]. However, as these couples were weakly observed, this result must be taken with caution.

Surprisingly, the comparison between structural asymmetry in 1HSI and 1HSI$_{mini}$ structures shows that seven positions are *noAsym^{Struct}* in 1HSI structure and *asym^{Struct}* in the four 1HSI$_{mini}$ structures. This suggests that this asymmetry is specific to minimized structures. Six of these positions defined two regions of the cantilever and fulcrum regions, that corresponds to the end of a β-strand and the beginning of a loop (Figure A4). Five of them belong to position cluster 3 grouping positions exhibiting no asymmetry and the two others belong to cluster 5 grouping residues exhibiting only accessibility asymmetry (Figure 5). This suggests that the non-structural asymmetry of these positions in the 1HSI structure can be a crystallization artefact.

Interestingly, thirteen positions are very specific; they are *asym^{Struct}* in the 1HSI structure and in all four 1HSI$_{mini}$ structures. Seven of these positions (37, 40, 42, 48, 50, 51, and 64) were detected as overrepresented structurally asymmetric positions in a set of 19 PR2 structures [18], highlighting a conservation of this structural asymmetry across the 19 PR2 structures and the 1HSI$_{mini}$ structures. Thus, the structural asymmetry of these thirteen positions should not be linked to crystal packing but to intrinsic properties of the 1HSI structure. Three of them (residues 4, 50, and 51) belong to the PR2 interface [18], suggesting that their structural asymmetry could be explained by the dimerization of the PR2. Eight of these structural asymmetric positions are in the elbow and flap regions (positions of clusters 7 and 8) that are important for the conformational transition from the open form to the closed form. Thus, we conclude that the structural asymmetry of these positions is important for the large deformations upon ligand binding.

3. Materials and Methods

3.1. Presentation of the Data

3.1.1. Crystallographic Structure of the Unbound PR2

In this study, we analyzed the asymmetry of PR2 target using the PDB file of the only one available PR2 structure in unbound form (PDB code: 1HSI [23]). This structure corresponds to a homodimer of 99 amino acids for each chain. No metal or crystallographic agent was contained in this PDB file. To study the PR2 structure, eight structural and functional regions were considered as previously described in [32]: the dimerization (1-5 + 96-99), fulcrum (10–23), catalytic (24–30), elbow (37–42), flap (43–58), cantilever (59–75), wall (80–83), and α-helix (87–95) regions.

3.1.2. Generation of the Energetic-Minimized Structures of the Unbound PR2

To remove bad contacts in the 1HSI structure we performed an energetic minimization of the 1HSI structure. The monoprotonated state was assigned to the oxygen atom OD2 of Asp25' in chain B of 1HSI structure using the PROPKA software [33]. 1HSI protein and water molecules were described using the force field AMBER ff99SB21 [34]. The system was solvated with a TIP3P water model with a truncated octahedron box with a 12.0 Å distance from the box edge. The energy minimization was carried out using GROMACS [30] and a conjugate gradient minimization of 2000 steps and a maximal force of 100 kJ mol^{-1}. The Particle Mesh Ewald (PME) method was adopted to consider the long-range electrostatic interactions [35,36]. A cut-off distance of 10.0 Å was used to treat the long-range electrostatic and van der Waals interactions. This procedure was run four times and results in four minimized structures of 1HSI, noted 1HSI$_{mini}$ structures.

3.2. Detection of Asymmetry in PR2 Structures

In this study, we focused on three asymmetries in PR2: (i) the structural asymmetry, (ii) the accessibility asymmetry, and (iii) the flexibility asymmetry. All these asymmetries are defined below.

3.2.1. Structural Asymmetry

Two approaches were used to quantify the structural asymmetry in the unbound PR2 structure. Firstly, we used the classical approach consisting of comparing the global conformation of both chains of the dimer by computing the Cα-RMSD between the two superimposed PR2 chains [14,16,17]. The superimposition of the two PR2 chains of 1HSI structure and the computation of the Cα-RMSD were performed using PyMoL software [37].

The second approach corresponds to the method that we have recently elaborated to study the link between the structural asymmetry in a set of 19 PR2 structures and the PR2 capacity to bind various ligands [18]. This method compares the local conformations of each position in the two chains of a dimer using the structural alphabet HMM-SA [19]. HMM-SA is a library of 27 structural prototypes of four residues, called structural letters. It was obtained after a classification of overlapping four-Cα fragments using the hidden Markov model based on the fragment geometry [19,24]. Among the 27 structural letters of HMM-SA, 4 describe helices, 5 describe β-strands, and 18 describe loops. Two loop structural letters have a large variability (RMSD between all fragments corresponding to these letters is higher than 0.5 Å [19]) that indicates that they are less accurate than others. In this approach, HMM-SA was used to simplify the 3D structure of the two chains of the dimer into two sequences of structural letters. Each structural letter describes the local geometry of each four-Cα fragment ($i - 2$, $i - 1$, i, and $i + 1$) and is assigned to the third residue (i) of the four-Cα fragment. Asymmetric positions were located after comparing the structural-letter sequences of the two dimer chains. They were defined as positions exhibiting different structural letters—i.e., local structures—between the two monomers A and B. The advantage of HMM-SA-based method is that it does not require angle or distance computation and monomer superposition but consists in a simple comparison of letter sequences of two monomers. We used this approach to extract structurally asymmetric positions in the crystallographic structure (1HSI) and in the four minimized structures (1HSI$_{mini}$ of the unbound PR2). From this data, we defined a categorical variable, named $Struct^{asym}$, that defines for each PR2 residue its structural asymmetry behavior. It takes two values: $asym^{Struct}$ for structurally asymmetric positions and $noAsym^{Struct}$ for non-structurally asymmetric positions.

3.2.2. Accessibility Asymmetry

Accessibility asymmetric positions correspond to positions exhibiting different ASA values in the two PR2 chains. The ASA value is the area of the surface swept out by the center of a probe sphere rolling over a molecule (atoms are spheres of varying radii) [38]. The ASA values of all atoms of 1HSI structure were computed using NACCESS software and a probe radius of 1.4 Å [38]. For each position of 1HSI, we computed the difference between the ASA values of the corresponding Cα atom in chains A and B, noted $\delta(ASA)^{A-B}$ value. Accessibility asymmetric positions were defined as positions having a $|\delta(ASA)^{A-B}|$ value higher than 1 Å^2. These data defined a categorical variable, named ASA^{asym}, which contains $asym^{ASA}$ for accessibility asymmetric positions and $noAsym^{ASA}$ for non-accessibility asymmetric positions.

3.2.3. Flexibility Asymmetry

Flexibility asymmetric positions were defined as positions having different flexibility in the two 1HSI chains. The flexibility of each position was quantified using the B-factor value of the corresponding Cα atoms. The B-factor value of an atom indicates the degree of isotropic smearing of its electron density around its center. Thus, the atom B-factor values reflect the relative vibrational motion of the different parts of the structure [21,39]. Atoms with low B-factors are in well-ordered and rigid part of the structure, while atoms with large B-factors generally are in to highly flexible part. The B-factor values of PR2 atoms of 1HSI structures were extracted from the PDB file. For each position, we computed the difference between B-factor values of corresponding Cα atoms in chains A

and B, noted $\delta(Bfactor)^{A-B}$ values. Flexibility asymmetric positions were defined as positions having a $|\delta(Bfactor)^{A-B}|$ value higher than 10 Å². These data defined a categorical variable, named $Bfactor^{asym}$, which contains $asym^{Bfactor}$ for flexibility asymmetric positions and $noAsym^{Bfactor}$ for non-flexibility asymmetric positions.

3.3. Analysis of the Link between the Three Asymmetry

Each position was characterized using three categorical variables (i) $Struct^{asym}$, taken two values ($asym^{Struct}$ and $noAsym^{Struct}$), (ii) ASA^{asym}, taken two values ($asym^{ASA}$ and $noAsym^{ASA}$), and (iii) $Bfactor^{asym}$ taken two values ($asym^{Bfactor}$ and $noAsym^{Bfactor}$). From the matrix containing the values of these three variables for the 96 positions (3–98), we computed an MCA. MCA is a descriptive technique dedicated to analyzing multiway tables in which a set of individuals (here PR2 positions) are described by a set of categorical variables (here the three asymmetry types). It allows analyzing the pattern of relationships among the categorical dependent variables using the inertia criterion with the chi-square metric [40]. MCA results was then plotted in a map where nearby points corresponded to PR2 positions that are similar regarding how they exhibit the same asymmetries. The MCA of the 196 positions described with the three binary variables were computed using the FactoMineR package of R software [41].

3.4. Detection of Positions Putatively Involved in Crystal Packing

PR2 residues putatively involved in crystal packing were determined from the crystal structure of the unbound PR2. Hereafter, the molecule contained in the asymmetric unit of this structure is referred to as "reference molecule". We then generated the "symmetry mate molecules" of the reference molecule by symmetry operations using 4 Å distance cut-off. This step is based on the CRYST1 record of the 1HSI PDB file that presents the unit cell parameters, space group, and Z value. Cα atoms involved in crystal packing were extracted using the protocol presented in [42] and PyMoL software [37]. They are defined as reference Cα-atoms being in intermolecular contacts with at least one symmetric mate atoms, i.e., situated at less than 4.0 Å to at least one symmetric mate. This cut-off of 4.0 Å was set to be a value longer than typical interactions: hydrogen bonding and electrostatic interactions.

4. Conclusions

In this work, we presented an in silico protocol to characterize the asymmetry of the unbound PR2, an important target for antiretroviral therapy of HIV-2 infection. This protocol is based on the comparison of the local conformations, ASA, and flexibility of each residue in both PR2 chains and on the comparison of asymmetry in crystallographic and minimized structures of the unbound PR2. Our results reveal that the unbound PR2 structure exhibits the three types of asymmetry (structural, accessibility, and flexibility) despite having two chains with identical same amino-acid sequence. The analysis of the structural asymmetry showed that some structurally asymmetric positions could correspond to crystallization artefacts. Our results also provide evidence that the non-artefactual asymmetry is an intrinsic property of the unbound PR2 and could be explained by the PR2 dimerization. This latter asymmetry could be important for the large deformation of the PR2 upon ligand binding. Our result also identified positions that are not asymmetric in the unbound PR2 structure but previously detected as asymmetric in PR2 complexed with various ligands [18]. These positions could be important for the PR2 adaptation during ligand recognition and binding. In addition, some of these positions are located outside of catalytic site. This suggests that the binding of a ligand in the catalytic site induce structural changes outside of this site. It is known that the substrate binding in the catalytic site of a semi-open conformation of the PR involves the flap closure onto the substrate that allows its catalytic hydrolyzing. Thus, avoiding the flap closure could be a pertinent approach to develop new inhibitor against PR. A new pocket, named "eye" pocket, was previously identified in flap region and corresponds to an allosteric site [43]. It has been shown that two ligands (5NI [44] and NIT (nitro-containing compound [45])) bind this eye pocket

and induce the inhibition of the PR1 with a mechanism other than competing for the active site [45]. Langevin dynamics simulations of the NIT-PR1 complex showed that the 5NI-binding in the eye pocket favors a semi-open conformation of PR1 that could directly modify the catalytic process. Thus, find new allosteric inhibitors of PR2 could be an alternative way to find new PIs effective against PR2. Previous results showed that PR1 and PR2 do not exhibit the same structural asymmetry [18] and that the binding of DRV and APV, two competitive inhibitors of the PR substrate, do not produce the same flap move in these two targets [22]. These results suggest that PR1 and PR2 could have different allosteric pathways. Thus, it would be interesting to study the allosteric inhibition mode for the PR2, particularly the fixation of inhibitors in the eye pocket of PR2.

To conclude our study is the first characterization of the asymmetry of the unbound PR2 with three levels of protein description. It provides a precise characterization of these asymmetries and new insights on the PR2 deformation upon ligand binding. In a next step, it would be interesting to analyze the PR2 asymmetry during molecular dynamics simulations of the PR2 complexed with competitive and non-competitive inhibitors to study more precisely the link between asymmetry and PR2 deformation upon ligand binding and the role of the asymmetry in ligand recognition and in allosteric pathway. All these data will provide a precise understanding of PR2 ligand binding, a necessary step before developing the new specific PR2 inhibitors required for infected patients treatments.

Abbreviations

The following abbreviations are used in this manuscript:

AIDS	acquired immunodeficiency syndrome
ASA	accessibility surface area
Cα-RMSD	root mean square deviation computed on Carbons α
HIV-1	human Immunodeficiency virus of type 1
HIV-2	human Immunodeficiency virus of type 2
HMM-SA	hidden Markov model - structural alphabet
PR	protease
PR1	HIV-1 protease
PR2	HIV-2 protease

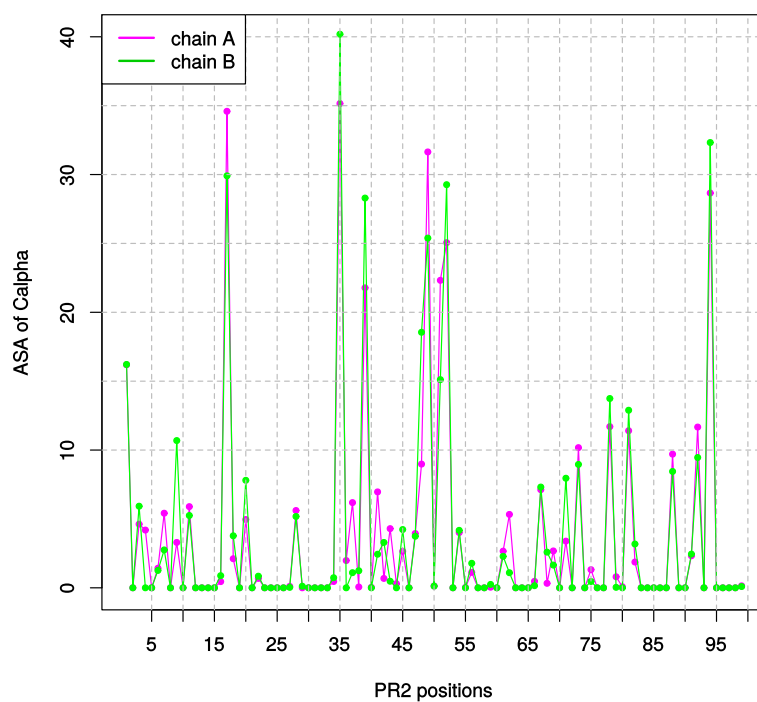## Appendix A. Accessibility of Each Residue

Figure A1. ASA value for each position in chains A and B of the unbound PR2.

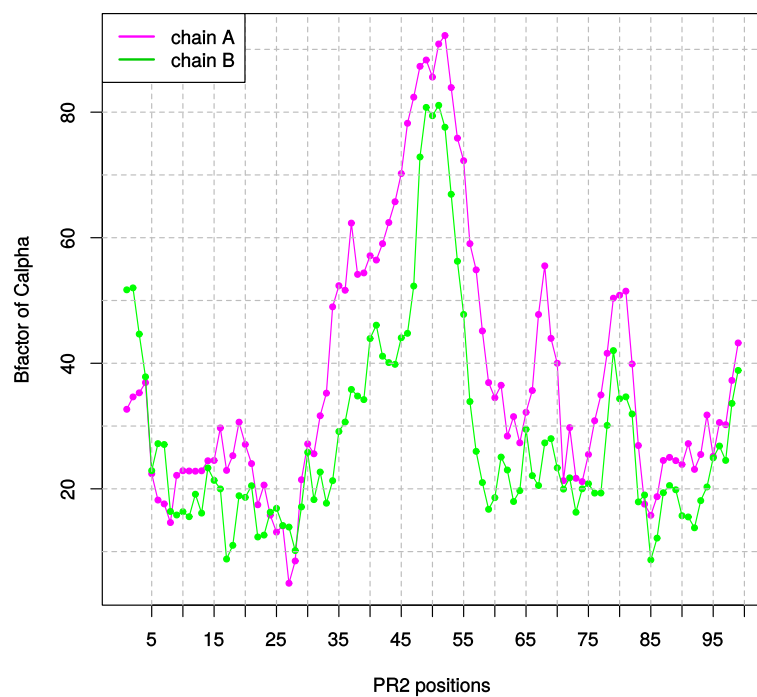## Appendix B. Flexibility of Each Residue

Figure A2. B-factor value for each position in chains A and B of the unbound PR2.

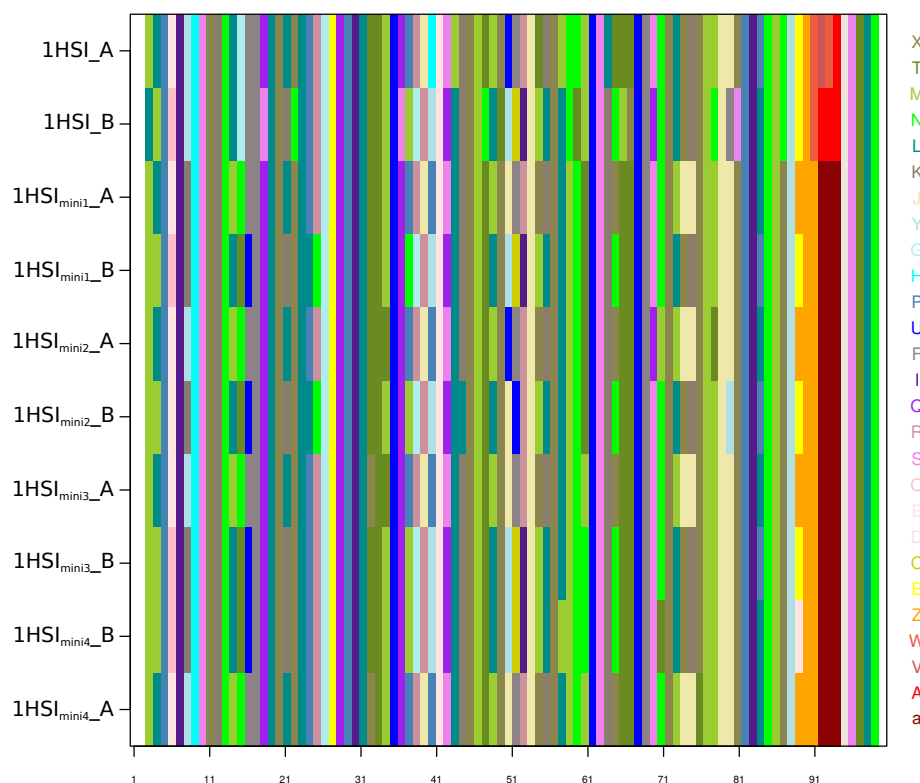## Appendix C. Structural-Letter Sequences of 1HSI and 1HSI$_{mini}$ Structures

Figure A3. The 8 chains are shown in rows and the 99 positions are shown in columns and colored according to the 27 structural letters [19]. [a, A, V, W] structural letters primarily found in the α-helix are colored in red, and [L, M, N, T, X] structural letters primarily found in the β-strand are colored in green, other structural letters correspond to loop.

Appendix D. Comparison of the Structural Asymmetry in 1HSI and 1HSI$_{mini}$ Structures
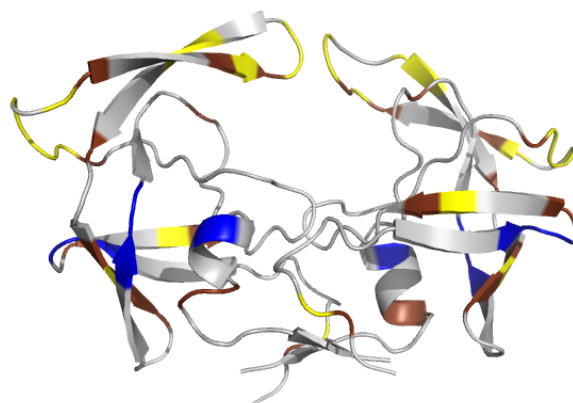
Figure A4. Structure of the unbound PR2 in cartoon mode. Residues colored in brown correspond to residue that are $asym^{Struct}$ in 1HSI structure and $noAsym^{Struct}$ in the four 1HSI$_{mini}$ structures. Residues colored in blue correspond to residue that are $noAsym^{Struct}$ in 1HSI structure and $asym^{Struct}$ in the four 1HSI$_{mini}$ structures. Residues colored in yellow correspond to residue that are $asym^{Struct}$ in 1HSI and in the four 1HSI$_{mini}$ structures.

References

1. Masse, S.; Lu, X.; Dekhtyar, T.; Lu, L.; Koev, G.; Gao, F.; Mo, H.; Kempf, D.; Bernstein, B.; Hanna, G.; et al. In vitro selection and characterization of human immunodeficiency virus type 2 with decreased susceptibility to lopinavir. *Antimicrob. Agents Chemother.* **2007**, *51*, 3075–3080. [CrossRef] [PubMed]

2. Rodés, B.; Sheldon, J.; Toro, C.; Jiménez, V.; Alvarez, M.; Soriano, V. Susceptibility to protease inhibitors in HIV-2 primary isolates from patients failing antiretroviral therapy. *Antimicrob. Agents Chemother.* **2006**, *57*, 709–713. [CrossRef] [PubMed]

3. Brower, E.; Bacha, U.M.; Kawasaki, Y.; Freire, E. Inhibition of HIV-2 protease by HIV-1 protease inhibitors in clinical use. *Chem. Biol. Drug Des.* **2008**, *71*, 298–305. [CrossRef] [PubMed]

4. Raugi, D.; Smith, R.; Ba, S.; Toure, M.; Traore, F.; Sall, F.; Pan, C.; Blankenship, L.; Montano, A.; Olson, J.; et al. Complex patterns of protease inhibitor resistance among antiretroviral treatment-experienced HIV-2 patients from senegal: Implications for second-line therapy. *Antimicrob. Agents Chemother.* **2013**, *57*, 2751–2760. [CrossRef] [PubMed]

5. Cavaco-Silva, J.; Aleixo, M.; Van Laethem, K.; Faria, D.; Valadas, E.; Gonçalves, M.D.; Gomes, P.; Vandamme, A.; Cunha, C.; Camacho, R.J. Mutations selected in HIV-2-infected patients failing a regimen including atazanavir. *Antimicrob. Agents Chemother.* **2013**, *68*, 190–192. [CrossRef] [PubMed]

6. Ntemgwa, M.; Toni, T.; Brenner, B.; Oliveira, M.; Asahchop, E.; Moisi, D.; Wainberg, M. Nucleoside and nucleotide analogs select in culture for different patterns of drug resistance in human immunodeficiency virus types 1 and 2. *Antimicrob. Agents Chemother.* **2009**, *53*, 708–715. [CrossRef] [PubMed]

7. Menéndez-Arias, L.; Alvarez, M. Antiretroviral therapy and drug resistance in human immunodeficiency virus type 2 infection. *Antiviral Res.* **2014**, *102*, 70–86. [CrossRef] [PubMed]

8. Xiao, B.; Singh, S.; Nanduri, B.; Awasthi, Y.; Zimniak, P.; Ji, X. Crystal structure of a murine glutathione S-transferase in complex with a glutathione conjugate of 4-hydroxynon, 2-enal in one subunit and glutathione in the other: Evidence of signaling across the dimer interface. *Biochemistry* **1999**, *38*, 11887–11894. [CrossRef] [PubMed]

9. Cha, H.; Kopetzki, E.; Hubert, R.; Lanzendorfer, M.; Brandsteller, H. Structural basis of the adaptive molecular recognition by MMP9. *J. Mol. Biol.* **2002**, *320*, 1065–1079. [CrossRef]

10. Jin, L.; Stec, B.; Lipscomb, W.N.; Kantrowitz, E.R. Insights into the mechanisms of catalysis and heterotropic regulation of Escherichia coli aspartate transcarbamoylase based upon a structure of the enzyme complexed with the bisubstrate analogue N-phosphonacetyl-L-aspartate at 2.1 AA. *Proteins Struct. Funct. Genet.* **1999**, *37*, 729–742. [CrossRef]

11. Renatus, M.; Stennicke, H.R.; Scott, F.L.; Liddington, R.C.; Salvesen, G.S. Dimer formation drives the activation of the cell death protease caspase 9. *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14250–14255. [CrossRef] [PubMed]

12. Prabu-jeyabalan, M.; Nalivaika, E.; King, N.; Schiffer, C. Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: Structural insights for better antiviral therapy. *Society* **2003**, *77*, 1306–1315. [CrossRef]

13. Prabu-Jeyabalan, M.; Nalivaika, E.; Schiffer, C.A. How does a symmetric dimer recognize an asymmetric substrate? A substrate complex of HIV-1 protease. *J. Mol. Biol.* **2000**, *301*, 1207–1220. [CrossRef] [PubMed]

14. Mulichak, A.M.; Hui, J.; Tomasselli, A.; Heinrikson, R.; Curry, K.; Tomich, C.; Thaisrivongs, S.; Sawyer, T.; Watenpaugh, K. The crystallographic structure of the protease from human immunodeficiency virus type 2 with two synthetic peptidic transition state analog inhibitors. *J. Biol. Chem.* **1993**, *268*, 13103–13109. [PubMed]

15. Tong, L.; Pav, S.; Pargellis, C.; Dô, F.; Lamarre, D.; Anderson, P. Crystal structure of human immunodeficiency virus (HIV) type 2 protease in complex with a reduced amide inhibitor and comparison with HIV-1 protease structures. *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 8387–8391. [CrossRef] [PubMed]

16. Tong, L.; Pav, S.; Mui, S.; Lamarre, D.; Yoakim, C.; Beaulieu, P.; Anderson, P. Crystal structures of HIV-2 protease in complex with inhibitors containing the hydroxyethylamine dipeptide isostere. *Structure* **1995**, *3*, 33–40. [CrossRef]

17. Priestle, J.; Fässler, A.; Rösel, J.; Tintelnot-Blomley, M.; Strop, P.; Grütter, M. Comparative analysis of the X-ray structures of HIV-1 and HIV-2 proteases in complex with CGP 53820, a novel pseudosymmetric inhibitor. *Structure* **1995**, *3*, 381–389. [CrossRef]

18. Triki, D.; Cano Contreras, M.; Flatters, D.; Visseaux, B.; Descamps, D.; Camproux, A.; Regad, L. Analysis of the HIV-2 protease's adaptation to various ligands: characterization of backbone asymmetry using a structural alphabet. *Sci. Rep.* **2018**, *8*, 710. [CrossRef] [PubMed]

19. Camproux, A.; Gautier, R.; Tufféry, P. A hidden Markov model derived structural alphabet for proteins. *J. Mol. Biol.* **2004**, *339*, 591–605. [CrossRef] [PubMed]

20. Chan, H.; Dill, K. Origins of structure in globular proteins. *Proc. Natl. Acad. Sci. USA* **1990**, *87*, 6388–6392. [CrossRef] [PubMed]

21. Karplus, P.; Schulz, G. Prediction of chain flexibility in proteins. *Naturwissenschaften* **1985**, *72*, 212–213. [CrossRef]

22. Chen, J.; Liang, Z.; Wang, W.; Yi, C.; Zhang, S.; Zhang, Q. Revealing origin of decrease in potency of darunavir and amprenavir against HIV-2 relative to HIV-1 protease by molecular dynamics simulations. *Sci. Rep.* **2014**, *4*, 6872. [CrossRef] [PubMed]

23. Chen, Z.; Li, Y.; Chen, E.; Hall, D.; Darke, P.; Culberson, C.; Shafer, J.; Kuo, L. 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. *J. Biol. Chem.* **1994**, *269*, 26344–26348. [PubMed]

24. Regad, L.; Guyon, F.; Maupetit, J.; Tufféry, P.; Camproux, A.C. A Hidden Markov Model applied to the protein 3D structure analysis. *Comput. Stat. Data Anal.* **2008**, *52*, 3198–3207. [CrossRef]

25. Rost, B.; Sander, C. Conservation and prediction of solvent accessibility in protein families. *Proteins* **1994**, *20*, 216–226. [CrossRef] [PubMed]

26. Pollastri, G.; Baldi, P.; Fariselli, P.; Casadio, R. Prediction of coordination number and relative solvent accessibility in proteins. *Proteins* **2002**, *47*, 142–153. [CrossRef] [PubMed]

27. Kar, P.; Knecht, V.J. Origin of decrease in potency of darunavir and two related antiviral inhibitors against HIV-2 compared to HIV-1 protease. *Phys. Chem. B* **2012**, *116*, 2605–2614. [CrossRef] [PubMed]

28. Goodsell, D.; Olson, A.J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 105. [CrossRef] [PubMed]

29. Brown, J.H. Breaking symmetry in protein dimers: Designs and functions. *Protein Sci.* **2006**, *15*, 1–13. [CrossRef] [PubMed]

30. Abraham, M.; Murtolad, T.; Schulz, R.; Pálla, S.; Smith, J.; Hess, B.; Lindahl, E. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **2015**, *1–2*, 19–25. [CrossRef]

31. Regad, L.; Chéron, J.; Triki, D.; Senac, C.; Flatters, D.; Camproux, A. Exploring the potential of a structural alphabet-based tool for mining multiple target conformations and target flexibility insight.. *PLoS ONE* **2017**, *12*, e0182972. [CrossRef] [PubMed]

32. Sadiq, S.K.; de Fabritiis, G. Explicit solvent dynamics and energetics of HIV-1 protease flap opening and closing. *Proteins Struct. Funct. Bioinform.* **2010**, *78*, 2873–2885. [CrossRef] [PubMed]

33. Li, H.; Robertson, A.; Jensen, J.H. Very fast empirical prediction and rationalization of protein pK a values. *Proteins Struct. Funct. Genet.* **2005**, *61*, 704–721. [CrossRef] [PubMed]

34. Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins* **2006**, *65*, 712–725. [CrossRef] [PubMed]

35. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092. [CrossRef]

36. Essmann, U.; Perera, L.; Berkowitz, M. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593. [CrossRef]

37. *The PyMOL Molecular Graphics System*, version 1.8; Schrödinger, LLC.: San Carlos, CA, USA, 2002.

38. Hubbard, S.; Thornton, J. *NACCESS*; Computer Program; Department of Biochemistry and Molecular Biology, University College London: London, UK, 1993.

39. Parthasarathy, S.; Murthy, M. Analysis of temperature factor distribution in high-resolution protein structures. *Protein Sci.* **1997**, *6*, 2561–2567. [CrossRef] [PubMed]

40. Abdi, H.; Valentin, D. Encyclopedia of measurement and statistics. In *Multiple Correspondence Analysis*; Sage: Thousand Oaks, CA, USA, 2007; pp. 651–657.

41. Lê, S.; Rennes, A.; Josse, J.; Husson, F. FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* **2008**, *25*, 18677. [CrossRef]

42. Carugo, O.; Djinovic-Carugo, K. How many packing contacts are observed in protein crystals? *J. Struct. Biol.* **2012**, *180*, 96–100. [CrossRef] [PubMed]

43. Damm, K.; Ung, P.; Quintero, J.; Gestwicki, J.; Carlson, H. A poke in the eye: Inhibiting HIV-1 protease through its flap-recognition pocket. *Biopolymers* **2008**, *89*, 643–652. [CrossRef] [PubMed]

44. Perryman, A.; Zhang, Q.; Soutter, H.; Rosenfeld, R.; McRee, D.; Olson, A.; Elder, J.; Stout, C. Analysis of temperature factor distribution in high-resolution protein structures. *Chem. Biol. Drug Des.* **2010**, *75*, 257–268. [CrossRef] [PubMed]

45. Ung, P.; Dunbar, J.; Gestwicki, J.; Carlson, H. An Allosteric Modulator of HIV-1 Protease Shows Equipotent Inhibition of Wild-Type and Drug-Resistant Proteases. *J. Med. Chem.* **2014**, *57*, 6468–6478. [CrossRef] [PubMed]