



HAL
open science

Étude des mécanismes de résistance du VIH-2 pour les inhibiteurs de protéase en utilisant des approches de bioinformatique structurale

Pierre Laville

► **To cite this version:**

Pierre Laville. Étude des mécanismes de résistance du VIH-2 pour les inhibiteurs de protéase en utilisant des approches de bioinformatique structurale. Bio-informatique [q-bio.QM]. Université Paris Cité, 2021. Français. NNT : 2021UNIP5245 . tel-04200205

HAL Id: tel-04200205

<https://theses.hal.science/tel-04200205>

Submitted on 8 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Paris

École doctorale Pierre Louis de santé publique – ED 393
Épidémiologie et sciences de l'information biomédicale

Laboratoire : Unité de Biologie Fonctionnelle & Adaptative (BFA), CNRS UMR 8251
Équipe 8 : *Computational Modeling of Protein-Ligand Interactions (CMPLI)*, Inserm U1133

Étude des mécanismes de résistance du VIH-2 pour les inhibiteurs de protéase en utilisant des approches de bioinformatique structurale

Par : Pierre LAVILLE

Thèse de doctorat de bioinformatique

Dirigée par D^r Michel PETITJEAN
et co-encadrée par D^r Leslie REGAD

Présentée et soutenue publiquement le 15 octobre 2021
à Université de Paris

Devant un jury composé de :

P ^r	Diane DESCAMPS	Université de Paris	Présidente
P ^r	Annick DEJAEGERE	Université de Strasbourg	Rapporteuse
P ^r	Bernard OFFMANN	Université de Nantes	Rapporteur
D ^r	Fabio PIETRUCCI	Sorbonne Université	Examineur
D ^r	Michel PETITJEAN	Université de Paris	Directeur de thèse
D ^r	Leslie REGAD	Université de Paris	Co-encadrante



Sauf mention contraire, le contenu de ce document est sous licence CC-BY-NC-ND :

<https://creativecommons.org/licenses/by-nc-nd/4.0/deed.fr>

Résumé

Les résistances aux médicaments antirétroviraux constituent un problème grave dans le traitement du virus de l'immunodéficience humaine (VIH), en particulier pour le type 2 (VIH-2), prédominant en Afrique de l'Ouest. La protéase du VIH est une cible habituelle pour la conception de nouveaux médicaments pour traiter l'infection au VIH. Des études précédentes ont montré que les structures cristallographiques de la protéase du VIH-2 (PR2), en forme liée ou non liée, présentent une asymétrie structurale importante pour la reconnaissance et la liaison au ligand. C'est pourquoi nous avons étudié les effets des mutations de résistance sur l'asymétrie structurale de la PR2. En raison de l'absence de données structurales sur les mutants de la PR2, nous avons modélisé leurs structures 3D avec des méthodes *in silico*.

La comparaison systématique de l'asymétrie de la structure sauvage avec celles d'un grand nombre de mutants a mis en évidence des acides aminés de grande importance pour la structure et la fonction de la PR2. Nos résultats ont révélé des changements structuraux induits par la flexibilité de la PR2 ou par des mutations de résistance. L'analyse de ces changements a montré que certaines mutations altèrent la stabilité protéique ou la liaison à l'inhibiteur. Cela a permis de suggérer trois mécanismes possibles de résistance pour la PR2, où les changements structuraux conduisent à des modifications de l'interface de dimérisation, de la reconnaissance du ligand ou de la liaison à l'inhibiteur.

Nous avons montré qu'une mutation pouvait induire de grands réarrangements structuraux dans les chaînes latérales et dans le squelette des acides aminés mutés, dans leur voisinage ou plus loin. Nous avons également mis en évidence que ces déformations modifient la conformation, le volume, et l'hydrophobicité de la poche de liaison et la composition et la taille de l'interface de la PR2. Ces résultats suggèrent que les mutations de résistance peuvent altérer la liaison au ligand par la modification des propriétés de la poche et de la stabilité de l'interface avec la PR2.

Nos résultats renforcent la compréhension des effets des mutations qui surviennent dans la PR2 et des différents mécanismes de résistance.

Mots-clés : Protéase du VIH-2; mutations de résistance aux médicaments; asymétrie structurale

Abstract

Study of HIV-2 resistance mechanisms against protease inhibitors by using structural bioinformatic approaches

Resistances to antiretroviral drugs is a severe problem in the human immunodeficiency virus (HIV) treatment, particularly for its type 2 (HIV-2), which is predominant in West Africa. HIV protease is a common target for the design of new drugs for treating HIV infection. Previous studies showed that the crystallographic structures of the HIV-2 protease (PR2) in bound and unbound forms exhibit structural asymmetry that is important for ligand recognition and binding. This is why we investigated the effects of resistance mutations on the structural asymmetry of PR2. Due to the lack of structural data on PR2 mutants, we modeled their 3D structures by *in silico* methods.

The systematic comparison of the asymmetry of the wild-type (WT) structure and a large number of mutants highlighted crucial residues for PR2 structure and function. Our results revealed structural changes induced by PR2 flexibility or resistance mutations. The analysis of these changes showed that some mutations alter protein stability or inhibitor binding. This suggested three possible resistance mechanisms of PR2, where structural changes induced by resistance mutations lead to modifications in the dimerization interface, ligand recognition or inhibitor binding.

We showed that one mutation could induce large structural rearrangements in side-chain and backbone of mutated residues, in its vicinity or further. We further showed that these deformations modify the conformation, volume, and hydrophobicity of the binding pocket, and the composition and size of the PR2 interface. These results suggest that resistance mutations could alter ligand binding by modifying pocket properties and PR2 stability, by impacting its interface.

Our results reinforce the understanding of the effects of mutations that occurred in PR2 and the different mechanisms of PR2 resistance.

Keywords: HIV-2 protease; drug-resistance mutations; structural asymmetry

À Anaëlle,

À mon frère,

À mes parents.

Remerciements

Merci à Leslie de m'avoir guidé et accompagné dans cette aventure qu'est la thèse. Ta persévérance, ta curiosité et ta soif de sciences resteront pour moi un modèle à l'avenir.

Merci à Michel pour ta rigueur, ton pragmatisme et ta sincérité dont tu as fait preuve dans mon encadrement.

Merci aux rapporteurs et aux membres du jury d'avoir accepté de consacrer du temps à l'appréciation de ce travail.

Merci à l'école doctorale de m'avoir jugé digne de confiance pour mener à bien ces travaux de recherche.

Merci à Delphine et Samuel, vous avez été de précieux conseils théoriques et pratiques en ce qui concerne la dynamique moléculaire.

Merci à mes collègues et anciens collègues de bureau pour votre bonne humeur, vos conseils et votre présence chaque jour : Baptiste, Bryan, Natacha, Mariem, Sarah et Vanille.

Merci à tous les autres membres du laboratoire pour votre accueil, les moments conviviaux et bien sûr les discussions scientifiques : Anne, Anne-Claude, Audrey, Gautier, Guillaume, Julien, Laurence, Olivier, Pierre, Rachel et Sjoerd.

Merci aux étudiants des masters ISDD (In Silico Drug Design) et BI (Bioinformatique) ainsi que les enseignants avec lesquels j'ai fait mes premiers pas de l'autre côté de la barrière.

Merci à toute ma famille, en particulier mes parents qui n'ont jamais cessé de me soutenir dans les échecs et les réussites.

Merci à tous mes amis proches et moins proches pour votre soutien et tous les moments partagés qui participent à mon épanouissement.

Merci à Anaëlle pour ta présence à mes côtés, ta patience et tout ce que tu m'apportes au quotidien.

Ces remerciements sont bien sûr non-exhaustifs et s'adressent à tous ceux qui se reconnaîtraient dans ces lignes. Ces remerciements, qui sont exposés avec la pudeur que m'impose le papier, ne traduisent en rien la gratitude que j'éprouve. Merci à vous tous.

Abréviations

ACP	Analyse en composantes principales
ATZ	Atazanavir
APV	Amprenavir
ADN	Acide désoxyribonucléique
ARN	Acide ribonucléique
DRV	Darunavir
FDA	<i>Food and Drug Administration</i>
HMM-SA	<i>Hidden Markov Model – Structural Alphabet</i>
IDV	Indinavir
IP	Inhibiteur de protéase
LPV	Lopinavir
MCS	<i>Mutant-Conserved Shifted atoms</i>
MDS	<i>Multidimensional scaling</i>
NFV	Nelfinavir
PDB	<i>Protein Data Bank</i>
PPIC	<i>Protein Protein Interface Computation</i> (logiciel)
PR1	Protéase du VIH-1
PR2	Protéase du VIH-2
RTV	Ritonavir
RMSD	<i>Root-mean-square deviation</i>
RMSF	<i>Root-mean-square fluctuation</i>
SAS	Surface accessible au solvant
SQV	Saquinavir
SIDA	Syndrome d'immunodéficience acquise
VIH-1	Virus de l'immunodéficience humaine de type 1
VIH-2	Virus de l'immunodéficience humaine de type 2
TPV	Tipranavir
WT	<i>Wild-type</i>

Sommaire

Sommaire	xiii
Liste des publications	xv
Liste des figures	xvi
Liste des tables	xviii
1 Introduction	1
1.1 Généralités	1
1.1.1 Présentation du VIH	1
1.1.2 Classification et origine	1
1.1.3 Structure du VIH	2
1.1.4 Cycle de réplication du VIH	3
1.1.5 Thérapeutique	5
1.2 Protéase du VIH-2	6
1.2.1 Structure de la protéase du VIH	6
1.2.2 Résistance de la PR2 face aux IP	10
Résistance innée	10
Résistance acquise	12
1.3 Projet de thèse	14
2 Impact des mutations sur l'asymétrie structurale de la PR2	17
2.1 Choix des mutations	17
2.2 Modélisation des mutants	17
2.3 Détection de l'asymétrie structurale	19
2.3.1 Présentation de l'alphabet structural HMM-SA	19
2.3.2 Détection et quantification de l'asymétrie structurale d'une structure	20
2.4 Détection de l'asymétrie structurale dans les structures sauvages	21
2.5 Asymétrie structurale des structures mutantes	21
2.6 Conclusion	25
2.7 1 ^{er} article	25
3 Impact des mutations sur la structure de la PR2	41
3.1 Choix des mutations	41
3.2 Modélisation des structures mutantes	42

3.3	Localisation et quantification des changements structuraux induits par les mutations	43
3.4	Impacts des mutations sur le site de liaison des IP	46
3.5	Comparaison des interfaces entre structures sauvages et mutantes de la PR2	47
3.6	Conclusion	48
3.7	2 ^e article	48
4	Étude de la dynamique des PR2 sauvage et mutantes	77
4.1	Choix des mutations	77
4.2	Choix de la structure de départ	78
4.2.1	Protocole de dynamique moléculaire	80
4.2.2	Étude de la conformation des flaps au cours des différentes simulations	80
	Étude de la simulation de la PR2 semi-ouverte	81
	Étude de la simulation de la PR2 fermée	82
	Conclusion sur la comparaison des deux simulations	84
4.3	Impact des mutations sur la structure de la PR2	86
	Étude de la stabilité et de la flexibilité de la PR2	86
	Exploration de l'espace conformationnel échantillonné par les structures	88
	Impacts des mutations sur la structure des flaps	91
	Conclusion	96
5	Conclusions et perspectives	99
	Annexe	105
	Bibliographie	111

Liste des publications

- Laville, P.; Petitjean, M.; Regad, L.
Structural impact of drug-resistance mutations appearing in HIV-2 protease.
Molecules **2021**, *26*[3], 611.
<https://doi.org/10.3390/molecules26030611>.
<https://hal.archives-ouvertes.fr/hal-03120812>.
- Laville, P.; Fartek, S.; Cerisier, N.; Flatters, D.; Petitjean, M.; Regad, L.
Impacts of drug-resistance mutations on the structural asymmetry of the HIV-2 protease.
BMC Mol. Cell Biol. **2020**, *21*, 46.
<https://doi.org/10.1186/s12860-020-00290-1>.
<https://hal.archives-ouvertes.fr/hal-02879916>.
- Laville, P.; Martin, J.; Launay, G.; Regad, L.; Camproux, A.-C.; de Vries, S.; Petitjean, M.
A non-parametric method to compute protein-protein and protein-ligands interfaces. Application to HIV-2 protease-inhibitors complexes.
bioRxiv **2018**, 498923.
<https://doi.org/10.1101/498923>.
<https://hal.archives-ouvertes.fr/hal-02323525>.

Liste des figures

1.1	Structure du génome du VIH et de ses neuf gènes	2
1.2	Arbre phylogénétique des VIH-1, VIH-2 et VIS	3
1.3	Cycle de réplication du VIH	4
1.4	Antirétroviraux et leurs associations approuvées en 2016 par la FDA	7
1.5	Découpage de la PR2 en 13 régions et formes semi-ouverte et fermée	9
1.6	Comparaison des structures et séquences de PR1 et PR2	11
1.7	Bilan des données de susceptibilité aux IP obtenues par des méthodes phénotypiques	14
2.1	Localisation des mutations de résistance sur la PR2	18
2.2	Protocole de modélisation des structures mutantes	19
2.3	Protocole de détection des positions asymétriques dans une structure	22
2.4	Nombre de positions asymétriques dans les cinq structures de chaque mutant	23
2.5	Changement d'asymétrie des six mutants qui présentent la mutation I82F	24
3.1	Localisation des mutations sur la PR2	42
3.2	Quantification des déplacements des atomes des résidus 27 et 50 entre la structure sauvage et une structure mutante	43
3.3	Localisation des réarrangements structuraux dans les 155 structures mutantes	45
3.4	Nombre d'atomes MCS par région de PR2	46
4.1	Variation de susceptibilité de médicaments pour différents mutants	78
4.2	Localisation des mutations I54M et I82F sur la structure de la PR2 et carte de contacts avec le DRV	79
4.3	Présentation de la distance $Ile50C^\alpha-Ile50'C^\alpha$ utilisée pour caractériser la conformation des flaps	81
4.4	Distribution de la distance $Ile50C^\alpha-Ile50'C^\alpha$ pour les structures extraites des trajectoires lancées sur la forme semi-ouverte et sur la forme fermée	82
4.5	Distribution des distances $Ile50C^\alpha-Ile50'C^\alpha$ pour les structures extraites des trajectoires des simulations des trois répliques de 100 ns lancées sur la forme semi-ouverte	83
4.6	Représentation des deux conformations fermées des flaps observées lors de la simulation de la PR2 sous la forme fermée	83
4.7	Distribution des distances $Ile50C^\alpha-Ile50'C^\alpha$ pour les structures extraites des trajectoires des simulations des trois répliques de 100 ns lancées sur la forme fermée	84
4.8	Évolution de la distance $Ile50C^\alpha-Ile50'C^\alpha$ au cours du temps lors de la simulation lancée sur la forme semi-ouverte	85

4.9	Évolution du RMSD au cours de la dynamique pour les trajectoires des structures sauvage et mutantes	87
4.10	Comparaison du RMSF des structures sauvage et mutantes sur les carbones α	89
4.11	Espace conformationnel échantillonné par les structures au cours des différentes simulations lancées sur les structures sauvage et mutantes	90
4.12	Espace conformationnel par région échantillonné par les structures au cours des différentes simulations lancées sur les structures sauvage et mutantes	92
4.13	Visualisation des régions présentant des conformations différentes lors de la simulation du double mutant I54M/I82F par rapport aux structures sauvages	93
4.14	Projection des structures sur le premier plan de l'ACP	94
4.15	Distributions des distances dont les différences sont significatives pour le WT et les mutants	97

Liste des tables

4.1	Description des distances utilisées dans la comparaison des structures extraites des simulations de la PR2 sauvage et mutantes	93
4.2	Comparaison des distances moyennes des structures sauvages et des structures mutantes	94

1 Introduction

1.1 Généralités

1.1.1 Présentation du VIH

Le sida (syndrome d'immunodéficience acquise) correspond à une immunodépression sévère qui rend l'organisme susceptible à des maladies opportunistes. Ce syndrome est causé par le virus de l'immunodéficience humaine (VIH). L'ONUSIDA ou Programme commun des Nations unies sur le VIH/sida a pour but la coordination des actions des actions de l'Organisation des Nations unies (ONU) dans la riposte contre le VIH/sida. Ce programme fait état de l'épidémie avec des statistiques dans sa fiche d'information de 2020 (ONUSIDA, 2020). En 2019, on dénombre 37,6 millions [30,2–45,0 millions]¹ de personnes vivant avec le VIH et 690 000 [480 000–1 million] décédées de maladies liées au sida. On compte 1,7 million [1,2–2,2 millions] de personnes nouvellement infectées la même année. Depuis le début de l'épidémie, ce sont 77,5 millions [54,6–110 millions] de personnes infectées et 34,7 millions [26,0–45,8 millions] de personnes décédées des suites de la maladie (chiffres fin 2019). Les nouvelles infections sont en recul, avec une réduction de 40 % depuis 1998. De même, les décès liés au sida ont atteint leur pic en 2004 et ont diminué de 60 % depuis.

Il existe deux types de VIH : le VIH de type 1 (VIH-1) et le VIH de type 2 (VIH-2). Le VIH-1 a été découvert par Barré-Sinoussi *et al.*, 1983 en 1983, alors que le VIH-2 a été découvert trois ans plus tard par Clavel *et al.*, 1986 en 1986. Plus de deux millions de personnes sont infectées par le VIH-2 (Visseaux *et al.*, 2016). Contrairement au VIH-1 qui est répandu à travers le monde, le VIH-2 se retrouve principalement en Afrique de l'Ouest, et en Europe (Damond *et al.*, 2005; Rodés *et al.*, 2006; Visseaux *et al.*, 2016; Peterson *et al.*, 2011). Dans les années 90, il était estimé que 1 à 2 millions de personnes étaient infectées par le VIH-2 en Afrique de l'Ouest (Visseaux *et al.*, 2016). Cette statistique restait d'actualité en 2018, tout en considérant que l'infection était plus rare ailleurs (van Tienen and van der Loeff, 2018). Les deux virus ont le même mode de transmission : par voies sanguine, sexuelle et de la mère à son enfant. Cependant, il existe certaines différences entre ces deux virus. Le VIH-2 est moins transmissible, avec une période plus longue avant de développer le sida, et est associé à une charge virale plus faible que le VIH-1 (Kanki *et al.*, 1994; Rodés *et al.*, 2006).

1.1.2 Classification et origine

Le VIH appartient à la famille des *Retroviridae* (rétrovirus) dans la classification du Comité international de taxonomie des virus (*ICTV*). Cette famille regroupe les virus

1. L'organisation précise que les nombres entre crochets représentent des fourchettes basées sur des estimations d'après les meilleures données disponibles (il est impossible de déterminer le nombre exact avec certitude), pouvant varier selon la taille de l'échantillon et les pays dans lesquels les enquêtes sont menées (ONUSIDA, 2017).

à ARN (acide ribonucléique) monocaténaire qui possèdent trois gènes : *Gag* (*group-specific antigen*), *Pol* (*polymerase gene*) et *Env* (*envelope glycoproteins*) encodant pour les protéines structurales et les enzymes permettant la rétro-transcription de leur ARN pour être intégré dans le génome de la cellule hôte (figure 1.1).

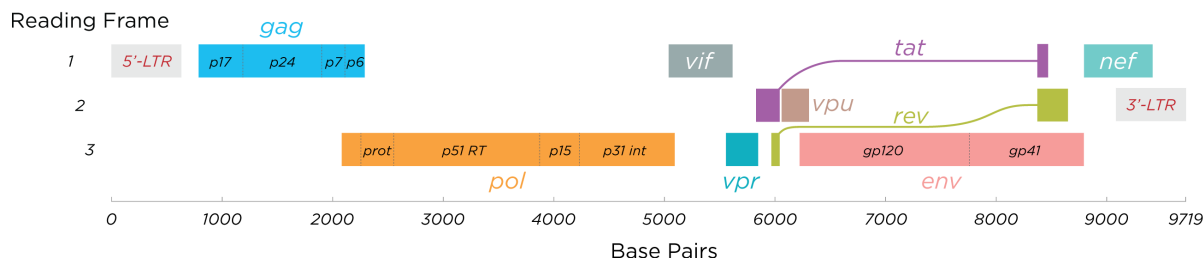


Figure 1.1 : Structure du génome du VIH-1 et de ses neuf gènes. Schéma par Thomas Spletstoeser sous licence CC-BY-SA 3.0 sur [Wikimedia Commons](#).

À l'intérieur de cette famille, le VIH appartient au genre des *Lentivirus* (du latin lentus, « lent ») qui se caractérisent par leur longue période d'incubation avant de manifester leur pouvoir pathogène sur les cellules. Les *Lentivirus* se divisent relativement aux hôtes auxquels ils sont associés. Il existe, à l'instar du VIH comme agent infectieux chez l'homme, des virus de l'immunodéficience bovine (VIB), féline (VIF), caprine (VAEC : virus de l'arthrite-encéphalite caprine), équine (VAIE : virus de l'anémie infectieuse des équidés) et simienne (VIS) ([Painter and Collins, 2019](#)).

L'origine du VIH a été retracée par des analyses phylogénétiques. Il partage un ancêtre commun avec le VIS, présent en Afrique chez des primates il y a au moins 32 000 ans ([Worobey et al., 2010](#)). Cette estimation est basée sur la présence de lignées du VIS endémiques sur l'île de Bioko (dans le golfe de Guinée) dont la formation géologique par montée des eaux a pu être datée. Le VIS se retrouve chez une trentaine d'espèces de primates africains ([Bibollet-Ruche et al., 2004](#)). Le VIH-1 et VIH-2 ont des origines génétiques différentes : alors que le VIH-1 serait issu du chimpanzé commun (*Pan troglodytes*) ([Keele et al., 2006](#)), le VIH-2 proviendrait du mangabey fuligineux (*Cercocebus atys*) ([Visseaux et al., 2016](#)) (figure 1.2)

1.1.3 Structure du VIH

Le VIH se compose d'une membrane lipidique qui contient un ARN simple brin et quelques protéines virales entourées d'une capsid. L'ARN du VIH comporte trois gènes Gag, Pol, et Env qui codent pour trois précurseurs polyprotéiques gag, gag-pol et env. Le précurseur gag sera clivé en protéines qui composeront la capsid du virus. Le précurseur gag-pol sera clivé en enzymes fonctionnelles indispensables à la réplication du VIH : l'intégrase, la transcriptase inverse et la protéase. Le précurseur env sera clivé en deux glycoprotéines gp120 et gp41 qui s'inséreront à la membrane lipidique du virus.

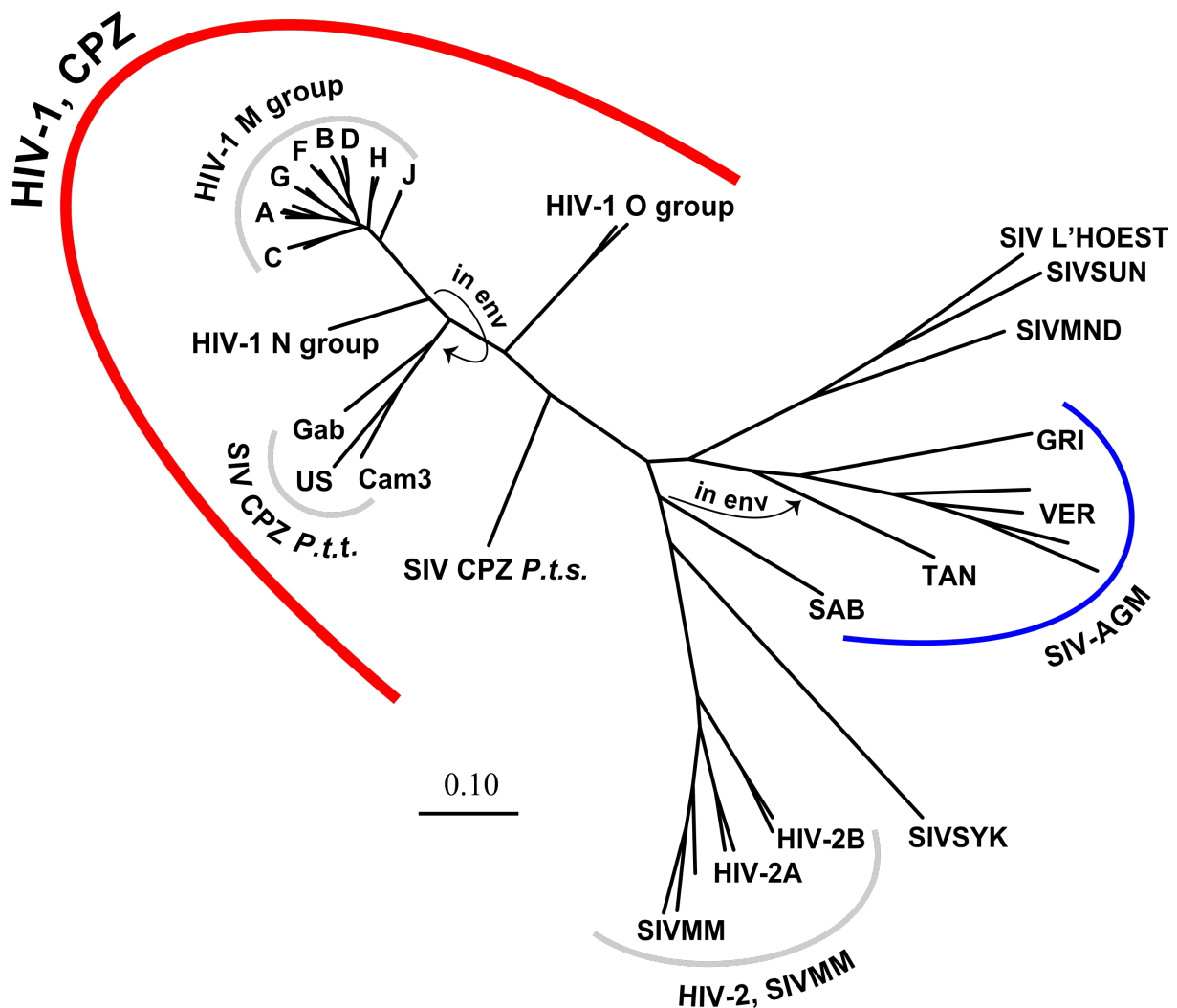


Figure 1.2 : Arbre phylogénétique des VIH-1, VIH-2 et VIS. Schéma par Kuiken, C., Foley, B., Hahn, B., Marx, P., McCutchan, F., Mellors, J. W., Mullins, J., Wolinsky, S. & Korber, B. (1999). A compilation and analysis of nucleic acid and amino acid sequences. In Human Retroviruses and AIDS. Los Alamos, New Mexico: Theoretical Biology and Biophysics Group, Los Alamos National Laboratory sur [Wikimedia Commons](#) (domaine public).

1.1.4 Cycle de réplication du VIH

Le VIH a besoin d'une cellule hôte pour effectuer sa réplication. Le cycle de réplication est présenté sur la figure 1.3; il se compose de différentes étapes commentées ci-dessous :

1. **Fixation et fusion** : les protéines gp120 à la surface du virion (forme extracellulaire d'un virus) se fixent au récepteur CD4 et au co-récepteur (CCR5 ou CXCR4) de la cellule hôte. Ces récepteurs sont principalement exprimés sur les lymphocytes de l'hôte, ce sont donc les cibles préférentielles de l'infection au VIH (les macrophages et cellules progénitrices hématopoïétiques dans une moindre

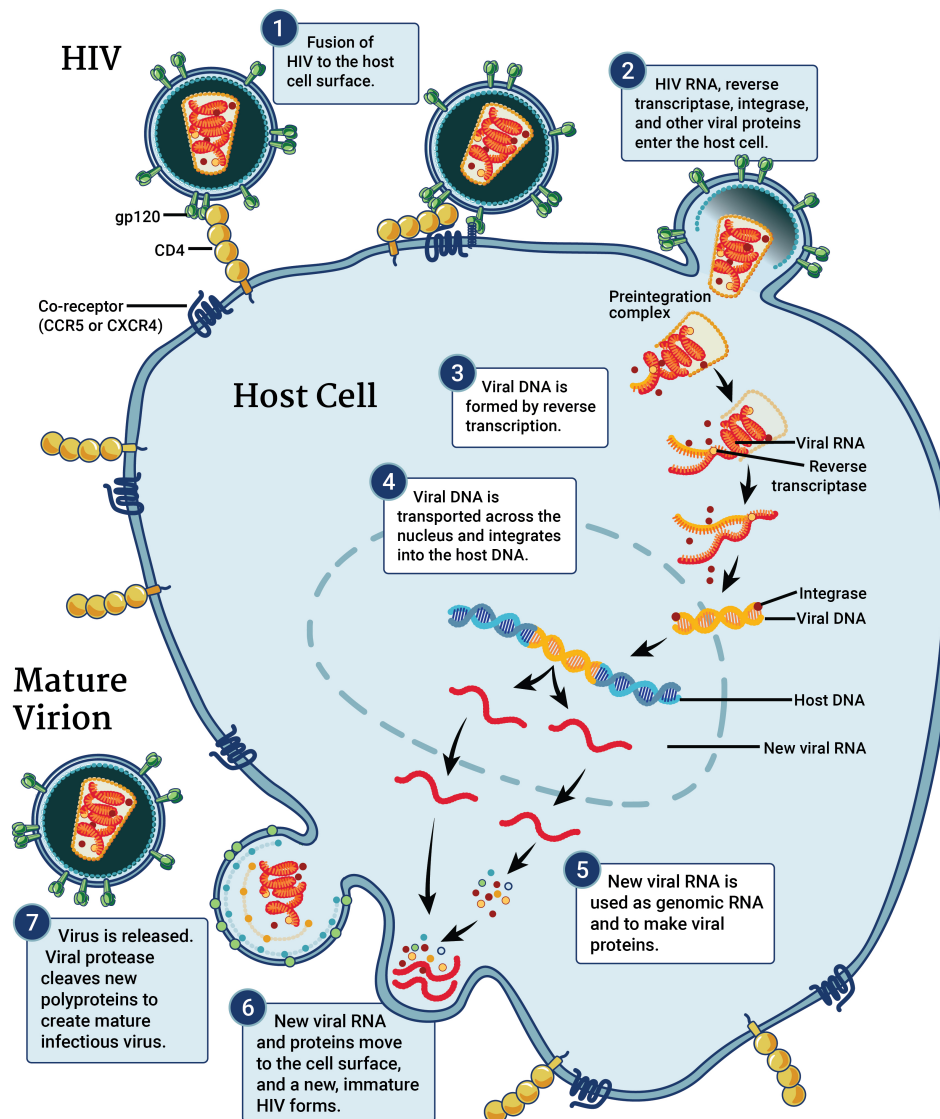


Figure 1.3 : Cycle de réplication du VIH. Schéma du National Institute of Allergy and Infectious Diseases sous licence CC-BY 2.0 (NIAID, 2018).

mesure également). La fusion à la membrane s'effectue grâce à un peptide de fusion (gp41) et permet l'entrée de la capsid virale dans le cytoplasme.

2. **Entrée** : après la fusion, les protéines et enzymes contenues dans la capsid sont libérées. On retrouve notamment l'ARN viral, la transcriptase inverse, l'intégrase, la protéase et la ribonucléase.
3. **Transcription inverse** : cette étape est la formation d'ADN viral (acide désoxyribonucléique) double brin à partir de l'ARN viral monocaténaire grâce à la transcriptase inverse. Cette étape est sujette à erreurs, ce qui est à l'origine de mutations potentiellement résistantes.

4. **Intégration** : le double brin d'ADN viral nouvellement formé vient pénétrer dans le noyau de la cellule hôte au travers d'un pore nucléaire puisque trop gros pour un transport passif à travers l'enveloppe nucléaire. C'est une des spécificités des lentivirus qui n'ont pas besoin d'attendre la rupture de cette enveloppe lorsque les cellules entrent en division. L'ADN viral est ensuite intégré dans le génome de la cellule cible à l'aide de l'intégrase virale. Le nouvel ARN viral peut désormais être transcrit par la machinerie cellulaire de la cellule hôte.
5. **Formation de protéines virales** : l'ARN viral sort du noyau et est traduit dans le cytoplasme par les ribosomes, engendrant la formation de nouvelles protéines virales non matures : gag, pol et env.
6. **Bourgeonnement** : les protéines virales sont transportées vers la surface de la cellule. Elles s'y assemblent en une structure globulaire qui aboutit à une particule virale non mature (virion) par bourgeonnement de la membrane de la cellule hôte.
7. **Maturation** : la particule virale relâchée est un virion immature qui nécessite le clivage de ses précurseurs polyprotéiques gag et gag-pol. C'est la protéase qui réalise cette étape et permet la formation d'un virion mature infectieux.

1.1.5 Thérapeutique

Les médicaments qui sont utilisés pour le traitement de l'infection au VIH-1 sont les antirétroviraux (Arts and Hazuda, 2012). Une combinaison de plusieurs médicaments issus d'au moins deux classes différentes est généralement utilisée dans le cadre d'une thérapie antirétrovirale. On parle de trithérapie lorsque ceux-ci sont au nombre de trois (Cihlar and Fordyce, 2016). Les antirétroviraux et leurs associations approuvées en 2016 par la FDA (*Food and Drug Administration*), sont présentés figure 1.4. Les traitements contre le VIH se divisent en quatre classes principales selon les cibles qu'ils visent :

- La **transcriptase inverse** : cette protéine est une ADN polymérase virale dont les inhibiteurs viennent bloquer l'activité de synthèse d'ADN viral. Ses inhibiteurs se divisent en plusieurs sous-classes : les inhibiteurs nucléosidiques, les inhibiteurs nucléotidiques et les inhibiteurs non nucléosidiques. Ils diffèrent par leur structure chimique, proche ou non de leurs analogues naturels nucléosides (base : adénine, guanine, cytosine ou thymine + sucre) et nucléotides (base + sucre + phosphate). Les antirétroviraux conçus pour traiter l'infection au VIH-1 ne sont pas tous efficaces contre le VIH-2 (Menéndez-Arias and Alvarez, 2014). En effet, les inhibiteurs non-nucléosidiques de la transcriptase inverse, qui constituent le traitement standard administré en premier lieu dans les pays à faibles revenus, sont sans effet sur le VIH-2 (Peterson *et al.*, 2011; Ba *et al.*, 2013).
- L'**intégrase** : l'inhibition de cette cible empêche le transport du double brin d'ARN viral dans le noyau de la cellule hôte. Des études phénotypiques ont montré que les trois inhibiteurs sur le marché (raltegravir, elvitegravir et dolutegravir) sont

efficaces contre VIH-2. Il existe néanmoins des mutations de résistance chez VIH-1 et VIH-2 pour ces trois médicaments (Desbois *et al.*, 2008; Charpentier *et al.*, 2010; Koh *et al.*, 2011; Smith *et al.*, 2011).

- Les **protéines impliquées dans l'entrée** du virus dans la cellule : les inhibiteurs ciblent soit les protéines de reconnaissance (par exemple le maraviroc pour CCR5, moins efficace contre VIH-2), soit les protéines de fusion (par exemple l'enfuvirtide pour gp41, naturellement résistant au VIH-2 (Borrego *et al.*, 2012; Menéndez-Arias and Alvarez, 2014)).
- La **protéase** : les inhibiteurs de protéase (IP) bloquent cette protéine qui intervient dans la maturation du virus. Il en existe deux types : de première et de deuxième génération, cette dernière étant conçue pour répondre à la résistance qui a émergé avec la première génération contre le VIH-1. Cela avait également pour but d'améliorer la biodisponibilité, la fréquence des doses et de minimiser les effets secondaires (Pokorna *et al.*, 2009). Il existe neuf inhibiteurs de protéase (IP) conçus pour traiter l'infection à VIH-1 : le saquinavir (SQV), le ritonavir (RTV), l'indinavir (IDV), le nelfinavir (NFV), l'amprenavir (APV), le lopinavir (LPV), l'atazanavir (ATZ), le tipranavir (TPV) et le darunavir (DRV). Mais seuls trois des neuf sont actifs contre le VIH-2 (DRV, SQV, LPV); les raisons de cette résistance intrinsèque aux IP sont inconnues (Raugi *et al.*, 2016).

La compréhension des mécanismes de résistance de la PR2 est donc enjeu important pour l'amélioration de la thérapie antirétrovirale, c'est pourquoi nous nous focaliserons sur cette cible dans l'ensemble de cette étude.

1.2 Protéase du VIH-2

La protéase est l'enzyme qui est responsable de la protéolyse des précurseurs viraux (gag, gag-pol) en protéines virales matures. La protéase du VIH est une protéase aspartique présentant deux aspartates (Asp) conservées en position 25 de chaque chaîne (figure 1.5 A), résidus situés dans la région catalytique en *sticks* verts).

1.2.1 Structure de la protéase du VIH

Il existe une grande disparité de structures cristallographiques disponibles dans la *PDB* (*Protein Data Bank*) entre PR1 (protéase du VIH-1) et PR2 : on dénombre 692 structures de PR1 publiées entre 1989 et 2021 contre seulement 19 structures de PR2, la première datant de 1993 et aucune publiée depuis 2012. De plus, aucune structure mutante n'est disponible pour PR2. Parmi ces structures, il y a une structure non liée (code PDB : 1HSI; Chen *et al.*, 1994) et 18 structures liées à divers ligands dont les trois médicaments approuvés par la FDA dans le traitement contre l'infection au VIH-2 : une avec le DRV (code PDB : 3EBZ; Kovalevsky *et al.*, 2008), une avec l'APV (code PDB : 3S45; Tie *et al.*, 2012), et la dernière avec l>IDV (code PDB : 1HSH; Chen *et al.*, 1994).

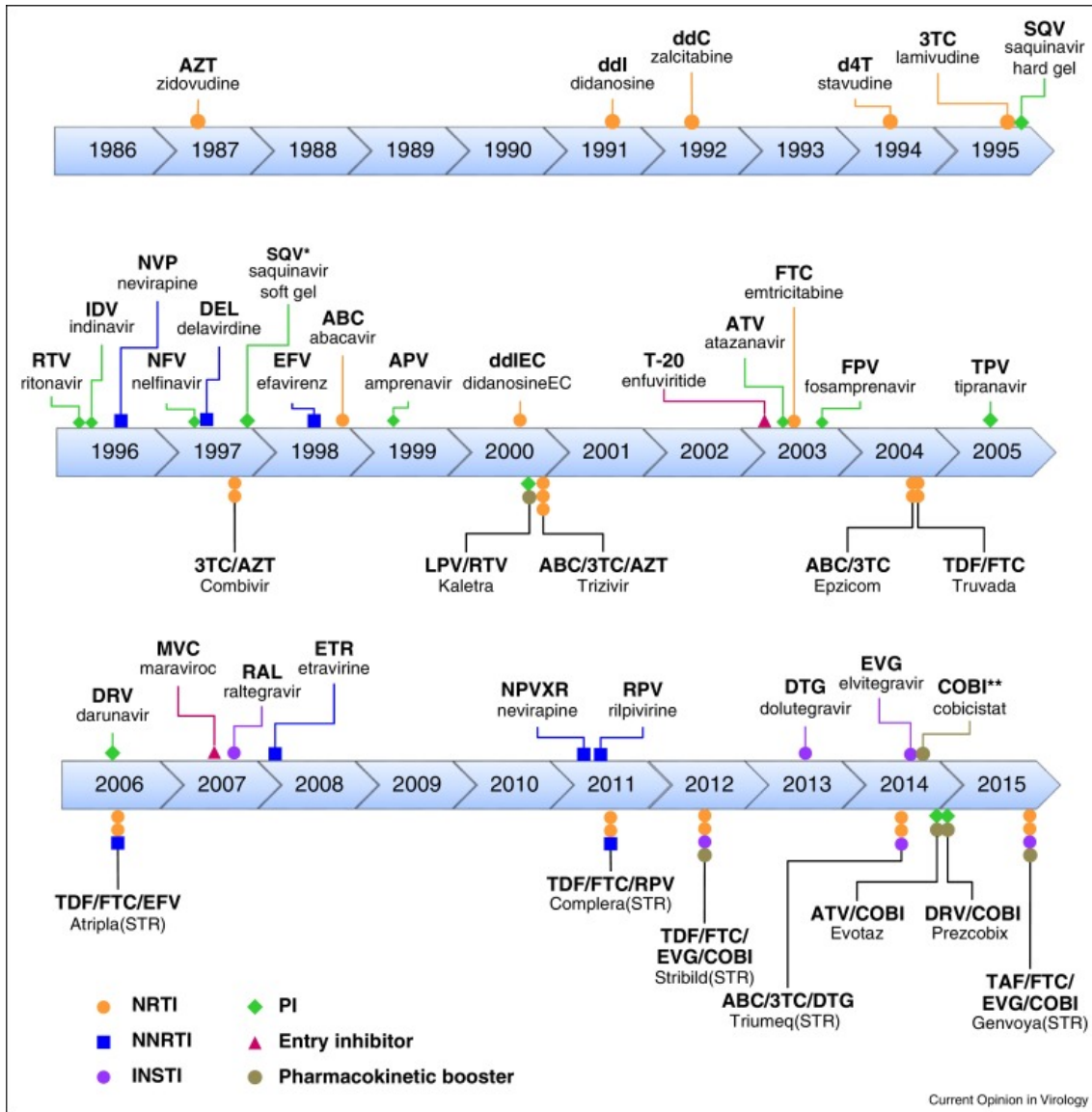


Figure 1.4 : Antirétroviraux et leurs associations approuvées en 2016 par la FDA. Schéma reproduit de Cihlar and Fordyce, 2016, sous licence CC-BY-NC-ND 4.0.

Enfin, on retrouve également des structures complexées avec des pseudo-peptides (codes PDB : 2MIP, 2HPF, 2HPE).

La protéase du VIH-2 (PR2) est un homodimère de 99 résidus pour chaque monomère. Chaque monomère est constitué de neuf brins β et d'une hélice α qui s'assemblent par des liaisons non covalentes. L'interaction entre les deux monomères se fait à travers les extrémités N-terminales (Nter) et C-terminale (Cter) et la région catalytique. La PR2 est classiquement découpée en 13 régions structurales et fonctionnelles définies par Sadiq and Fabritiis, 2010 (figure 1.5). La région catalytique se situe à l'interface entre les deux monomères et contient le site actif de la protéase. C'est dans cette zone que s'effectue la reconnaissance des sites de clivage des polyprotéines gag et gag-pol et la liaison avec les IP (Prabu-Jeyabalan *et al.*, 2000). Elle est bordée sur les côtés par deux

boucles (une sur chaque monomère, nommée *wall*, mur). Ces régions (catalytique et *wall*), constituent le site de liaison des substrats et des IP. Après fixation d'un ligand, le site de liaison se referme par l'intermédiaire des régions *flap* (rabat ; une région par monomère). Dans un monomère cette région est constituée de deux brins β formant un feuillet β antiparallèle. Dans le monomère, la région *flap* est bordée par les régions *elbow* (coude) et *cantilever* (porte-à-faux). La région *elbow* correspond à une boucle accessible au solvant. La comparaison des 19 structures cristallographiques de PR2 a montré que cette région est celle qui est le plus structurellement variable (Triki *et al.*, 2019). Le *cantilever*, tout comme la région du *fulcrum* (pivot) correspondent à des feuillets β composés de deux brins. La dernière région structurale correspond à l'hélice α . Les quatre autres régions, nommées R1, R2, R3 et R4 correspondent à des boucles reliant les régions structurales ou fonctionnelles.

Les analyses des structures cristallographiques des protéases ont montré que cette protéine adopte deux formes : la forme semi-ouverte et la forme fermée (figure 1.5). Lorsque la protéase est non complexée à un ligand, les flaps sont en conformation semi-ouverte définie par une distance entre les flaps minimale tout en maintenant une fermeture du site de liaison. Lorsque la protéase fixe un ligand, ces régions flaps se referment sur le site de liaison entraînant une fermeture totale du site de liaison. Dans cet état, les résidus 50 (isoleucines) des flaps qui se trouvent dans le coude du flaps établissent des interactions hydrophobes avec les ligands permettant leur bonne orientation dans le site catalytique (Gustchina and Weber, 1991; Boggetto and Reboud-Ravaux, 2001; Tóth and Borics, 2006; Hornak *et al.*, 2006). Dans une précédente étude du laboratoire, les auteurs ont montré que la fixation des IP entraîne des changements conformationnels au niveau des résidus de la poche de liaison mais aussi des mouvements coopératifs le long de la PR2, aboutissant à des déformations localisées dans l'hélice α , la fin du flap et le début du *fulcrum* (Triki *et al.*, 2019).

Au cours de leur étude qui a consisté à comparer les 19 structures de PR2, Triki *et al.*, 2019 se sont focalisés sur la caractérisation de la poche de liaison des IP. Pour cela, ils ont comparé la structure des résidus des poches extraites des 18 complexes de PR2 ainsi que les interactions établies entre la PR2 et les IP. Pour cela, chaque poche a été extraite en utilisant l'approche du ligand commun développé par Triki *et al.*, 2018a. Cette approche crée un « ligand commun » à partir de tous les ligands extraits des 18 complexes PR2-IP disponibles. Ce simple ligand commun est ensuite utilisé pour estimer une poche de liaison de chaque structure en extrayant tous les atomes de la PR2 à moins de 4,5 Å du ligand commun. L'avantage de cette approche est qu'elle permet de prendre en compte la diversité des ligands capables de fixer le site de liaison ainsi que leur flexibilité (Triki *et al.*, 2018a). Le site de liaison consensus est défini comme étant les 38 résidus qui contiennent au moins un atome dans une des poches des 18 complexes. L'analyse du site de liaison dans les 18 structures a montré qu'il est composé de trois types de résidus :

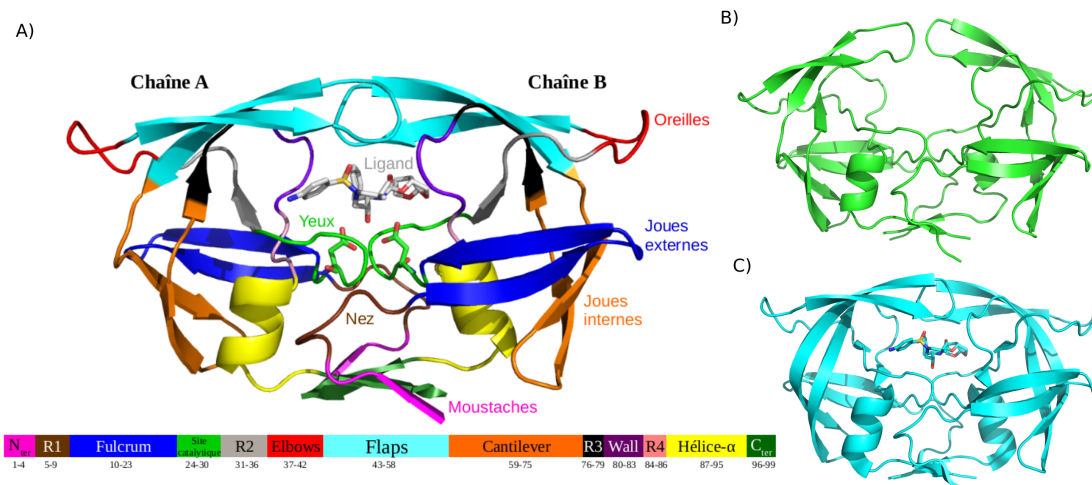


Figure 1.5 : A) Découpage de la PR2 en 13 régions (d'après [Sadiq and Fabritiis, 2010](#)). Figure A) extraite de [Triki, 2018](#) (code PDB : 3EBZ, complexe avec le DRV). B) Forme semi-ouverte en vert (code PDB : 1H5I) ; C) forme fermée en cyan, complexée au DRV en *sticks* (code PDB : 3EBZ).

- Les résidus qui présentent une conformation conservée dans les complexes et qui sont importants pour l'activité biologique de la PR2. Ces résidus sont principalement localisés au niveau du plancher de la poche. Cette région, qui contient le site catalytique, établit beaucoup d'interactions conservées avec les différents IP.
- Les résidus qui se déforment en réponse à la fixation des IP et qui sont donc importants pour la reconnaissance et la spécificité des IP. Ces résidus sont principalement localisés au niveau du côté gauche et de l'entrée du site. Ces résidus établissent des interactions spécifiques des différentes classes d'IP.
- Les résidus qui sont très flexibles et qui semblent importants pour l'ouverture et la fermeture du site de liaison. Ces résidus sont principalement localisés au niveau des flaps.

Par ailleurs, la comparaison des conformations locales des résidus des 18 structures de PR2 a montré que certains résidus de l'hélice α , le flap et le R4 présentent des conformations conservées importantes pour le maintien de la conformation de la poche de liaison ([Triki *et al.*, 2019](#)).

Il est classiquement considéré que la PR2 est un homodimère présentant un axe de symétrie C2 ([Menéndez-Arias and Alvarez, 2014](#)). Cependant, l'étude des structures cristallographiques disponibles de la PR2 a montré que la PR2 est une structure asymétrique, c'est-à-dire que les deux monomères ne présentent pas la même conformation ([Tong *et al.*, 1993](#); [Mulichak *et al.*, 1993](#); [Tong *et al.*, 1995](#); [Priestle *et al.*, 1995](#); [Triki *et al.*, 2018b](#); [Triki *et al.*, 2019](#)). Par exemple, il a été quantifié que les deux monomères présentent un *RMSD* (*root-mean-square deviation*) de 0,35 à 1,02 Å suivant la structure

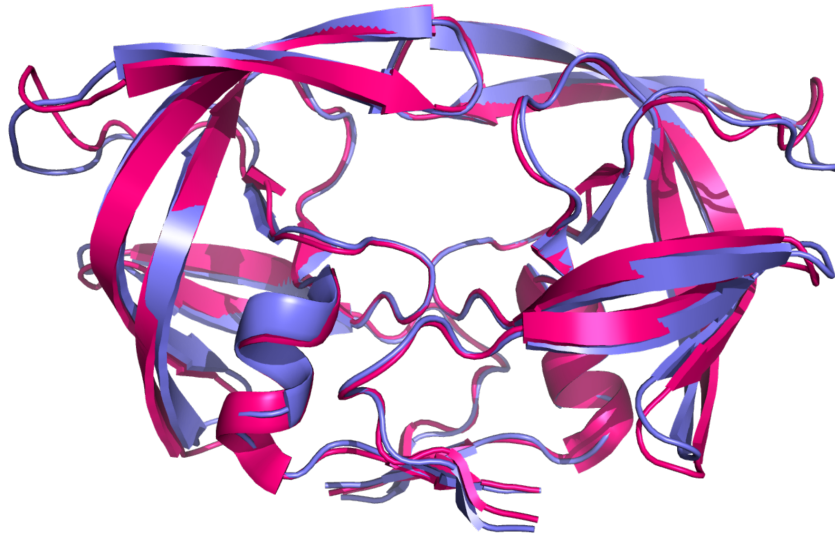
cristallographique étudiée (Tong *et al.*, 1993; Mulichak *et al.*, 1993; Tong *et al.*, 1995; Priestle *et al.*, 1995; Ollitrault *et al.*, 2018). De plus, une analyse des conformations locales des 19 structures de PR2 a montré que 31 % des positions de la PR2 sont asymétriques, c'est-à-dire qu'elles présentent des conformations différentes dans les deux chaînes des dimères (Triki *et al.*, 2018b). La localisation et la caractérisation des positions asymétriques identifiées dans les différentes structures a montré que cette asymétrie dans les structures cristallographiques résulte de différents facteurs : le *packing* (empilement) cristallin (Tong *et al.*, 1993; Mulichak *et al.*, 1993; Priestle *et al.*, 1995; Ollitrault *et al.*, 2018), la dimérisation (Triki *et al.*, 2018b), la fixation de ligands (Mulichak *et al.*, 1993; Tong *et al.*, 1995; Triki *et al.*, 2018b) et la flexibilité intrinsèque de la protéine. Cette asymétrie semble être impliquée dans différents mécanismes tels que la déformation structurale des dimères (Prabu-Jeyabalan *et al.*, 2003; Goodsell and Olson, 2000; Brown, 2006) et, en particulier, celle liée à la reconnaissance spécifique des ligands (Mulichak *et al.*, 1993; Tong *et al.*, 1995; Triki *et al.*, 2018b). De plus, l'étude de l'asymétrie des 19 structures de la PR2 de la PDB a permis d'identifier 10 résidus (dont cinq localisés dans le site de liaison) qui présentent une conservation de leur structure. Cela montre que la conformation de ces résidus est importante pour l'activité biologique ou la structure de PR2. À l'inverse, cette étude a mis en évidence que certains résidus des flaps et des elbow présentent de larges déformations qui peuvent s'expliquer par la fixation des ligands ou par leur flexibilité intrinsèque. Ces résultats montrent que l'étude et la caractérisation de l'asymétrie de la PR2 est une méthode efficace pour étudier les changements conformationnels situés au niveau du squelette peptidique (Triki *et al.*, 2018b).

Les études de la PR2 menées lors de la thèse de Triki, 2018, effectuée dans notre laboratoire, ont montré que la comparaison de plusieurs structures d'une même cible est une approche rapide qui permet d'extraire des informations structurales sur la cible, sur sa flexibilité et sa dynamique et permet d'explorer les changements structuraux observés et leur origines.

1.2.2 Résistance de la PR2 face aux IP

Résistance innée

Comme indiqué précédemment, la PR2 est naturellement résistante à six IP. Seulement trois sont recommandés pour le traitement du VIH-2 : DRV, SQV et LPV. Cette résistance naturelle s'explique par les substitutions entre les PR1 et PR2. La PR1 et la PR2 partagent environ 50 % d'identité de séquence, les substitutions se répartissant le long de la séquence (figure 1.6). Les deux protéines présentent le même repliement global quantifié par un faible RMSD proche de 1 Å (Tong *et al.*, 1993; Chen *et al.*, 1994; Kovalevsky *et al.*, 2008; Tong *et al.*, 1995). Les plus grandes différences structurales entre les deux protéases se localisent au niveau des résidus 15–20 (fulcrum), 30–40 (elbow), et 65–73 (cantilever) (Tong *et al.*, 1993; Kovalevsky *et al.*, 2008; Tong *et al.*, 1995).



```

4HLA_1|Chains PQITLWQRPLVTIKIGGQLKEALLDTGADDTVLEEMSLPGRWKPKMIGGIGGFIKVRQYD 60
3EBZ_1|Chains PQFSLWKRVPVTAYIEGQPVEVLLDTGADDSIVAGIELGNNYSPKIVGGIGGFINTKEYK 60
**::**:**:** * ** *.*****::: :.* ...**::*****:..:*.

4HLA_1|Chains QILIEICGHKAIGTVLVGPTPVNIIGRNLLTQIGCTLNF 99
3EBZ_1|Chains NVEIEVLNKKVRATIMTGDTPINIFGRNILTALGMSLNL 99
:: **: :.*. .*::.* **:**:**:** * * :**:
```

Figure 1.6 : Comparaison des structures et séquences de PR1 (violet, code PDB : 4HLA) et PR2 (rose, code PDB : 3EBZ). L'alignement des séquences montre (i) les résidus conservés : « * », (ii) les résidus qui ont des propriétés fortement similaires : « : » et (iii) les résidus qui ont des propriétés faiblement similaires : « . ».

Afin de comprendre la résistance innée du VIH-2 face aux IP, certaines études se sont focalisées sur l'étude de l'impact des substitutions localisées dans le site de liaison des deux protéases (Gustchina and Weber, 1991; Hoog *et al.*, 1996; Triki *et al.*, 2018a; Tie *et al.*, 2012; Raugi *et al.*, 2016). Ces études, qui comparent les structures cristallographiques de la PR2 complexée à différents inhibiteurs, ont montré que les substitutions en positions 32, 47, 76, et 82, sont fortement impliquées dans la résistance innée de PR2. Ces mutations induisent des réarrangements structuraux au niveau des résidus impliqués dans les interactions avec les IP (Kovalevsky *et al.*, 2008; Raugi *et al.*, 2016; Tie *et al.*, 2012). Une étude menée au laboratoire a porté sur la comparaison des propriétés physico-chimiques et géométriques des sites de liaison extraits de 28 structures cristallographiques de PR1 (15) et de PR2 (13) complexées au même IP (Triki *et al.*, 2018b). Cette étude a montré que les mutations localisées au niveau de la poche de liaison augmentent l'hydrophobicité de la poche comparée à celle de la PR1. Une autre étude (Triki *et al.*, 2019), a porté sur la comparaison de toute la structure de PR1 et PR2 : les auteurs ont montré que la majorité des régions de la PR2, exceptées les régions R3, catalytic, et flap, présentent des conformations différentes chez PR1 et PR2. Du fait du

protocole de comparaison, ces différences s'expliquent par plusieurs facteurs : le *packing* cristallin, la flexibilité intrinsèque des résidus et les substitutions. L'analyse de la diversité structurale de ces régions dans le jeu de structure a permis aux auteurs de montrer que les substitutions entre les deux protéases semblent avoir un impact sur la conformation des régions Nter, R4, hélice α , fulcrum, R2, cantilever, et elbow (Triki *et al.*, 2019). Il a été montré que les régions R4 et hélice α sont importantes pour le maintien de la conformation du site de liaison. Ainsi, il a été supposé que certaines substitutions entre PR1 et PR2 peuvent aussi modifier la fixation de l'IP en déformant la poche de liaison. La région elbow est, quant à elle, importante dans le mécanisme d'ouverture et de fermeture de la protéase. Dans leur étude, Triki, 2018 ont montré que les substitutions I32V et L76M localisées au niveau de la poche de liaison entraînent un grand réarrangement structural au niveau du résidu 45 de la chaîne B de la PR2, en comparaison à la PR1 Triki *et al.*, 2018a. Tout comme les substitutions M46I et V47I, ce changement conformationnel est localisé au niveau de la région flaps, région impliquée dans l'ouverture et la fermeture du site de liaison. Il est donc possible que ces substitutions aient aussi un impact sur le mécanisme d'ouverture et de fermeture de la protéase, et donc indirectement sur la liaison des IP. En utilisant des simulations de dynamique moléculaire sur la PR1 et PR2 complexées au DRV et l'APV, Chen *et al.*, 2014 ont montré que la PR2 présente des mouvements des flaps différents de ceux de la PR1. En effet, au cours de la simulation les auteurs ont observé que les flaps du complexe PR1-IP ont tendance à s'ouvrir vers l'extérieur. À l'inverse, les flaps de la PR2 ont, quant à eux, tendance à se refermer sur eux avec des mouvements différents suivant l'IP. Ainsi, il semble donc que les substitutions entre les deux protéases ont aussi un impact sur la transition entre les états ouvert et fermé des flaps.

Résistance acquise

Depuis plusieurs années, des résistances aux traitements antirétroviraux ont été observées chez des patients atteints du VIH-2 sous traitement. Ces résistances acquises s'ajoutent donc à la résistance innée de la PR2 face aux IP. Ces résistances s'expliquent par des mutations localisées sur les différentes protéines virales et particulièrement sur la PR2. Ces mutations ne modifient pas ou très peu l'activité catalytique de l'enzyme mais provoquent une perte de sensibilité aux médicaments. Ces mutations apparaissent suite aux erreurs de la transcriptase inverse lors de la synthèse de l'ADN viral à partir de l'ARN viral.

Concernant les mutations de résistance observées chez PR2, il existe peu de données. Les mutations ont été généralement identifiées après séquençage du virus extrait de patients infectés pour lesquels un échec thérapeutique a été observé (Masse *et al.*, 2007; Colson *et al.*, 2004; Damond *et al.*, 2005; Rodés *et al.*, 2006; Ntemgwa *et al.*, 2007; Desbois *et al.*, 2008; Jallow *et al.*, 2009; Raugi *et al.*, 2013; Larrouy *et al.*, 2013; Cavaco-Silva *et al.*, 2013; Storto *et al.*, 2018). Un test sanguin est réalisé chez ces patients. L'ARN

viral extrait est amplifié par *PCR* (*polymerase chain reaction*) en ADN qui est séquencé à l'aide de fluorophores pour marquer chaque base distinctement. Une comparaison avec la séquence sauvage permet de mettre en évidence les mutations et d'orienter la thérapeutique à mettre en place selon les résistances connues aux médicaments. Les mutations de résistances peuvent être ensuite confirmées par des tests de susceptibilité phénotypiques (Masse *et al.*, 2007; Rodés *et al.*, 2006; Ntemgwa *et al.*, 2007; Desbois *et al.*, 2008; Raugi *et al.*, 2013; Mahdi *et al.*, 2015). Cette approche permet de mesurer *in vitro* la capacité du virus à se développer dans différentes concentrations de médicaments antirétroviraux. Dans leur revue sur les mutations de résistance, Menéndez-Arias and Alvarez, 2014 comparent les mutations de résistance de la littérature par des approches phénotypiques selon différents IP (figure 1.7). Cette méthode aboutit à l'obtention d'une valeur de concentration inhibitrice médiane (nommée CI_{50} ou IC_{50} pour *half maximal inhibitory concentration*). Cette valeur indique la concentration requise d'un médicament pour inhiber la réplication du virus de 50 %. C'est parfois la concentration efficace médiane (nommée EC_{50}) qui représente la concentration d'un agoniste qui produit 50 % de l'effet maximal pour cet agoniste qui est quantifiée.

Ces mutations de résistance sont classiquement groupées en deux classes :

Les mutations majeures ou primaires : ces mutations sont les premières à être sélectionnées et sont souvent localisées au niveau du site de fixation des IP et altèrent la fixation des IP (Mahalingam *et al.*, 1999). Ces mutations ont souvent un impact délétère sur l'activité biologique de la protéase du virus résistant (Nijhuis *et al.*, 2001).

Les mutations mineures ou secondaires ou compensatrices : ces mutations apparaissent secondairement et sont localisées en dehors du site de liaison des IP. Ces mutations n'ont pas un effet direct sur la résistance mais elles la renforcent. De plus, elles permettent de compenser l'effet délétère sur l'activité de la protéase des mutations majeures.

Ces mutations peuvent entraîner des résistances à un ou plusieurs IP. Par exemple, la mutation V47A (localisée au niveau du site de liaison) entraîne une résistance au LPV avec une augmentation de l' EC_{50} d'un facteur 6,3 (par rapport à la PR2 sauvage) mais cause une hypersensibilité à SQV avec une diminution de l' EC_{50} d'un facteur 15 (Raugi *et al.*, 2013). La mutation I54M entraîne une résistance à la majorité des IP (IDV, NFV, APV, TPV, LPV, DRV) et une hypersensibilité pour le SQV (Ntemgwa *et al.*, 2007; Raugi *et al.*, 2013). La mutation I82F entraîne une résistance à l'IDV et une hypersensibilité au DRV et au SQV (Ntemgwa *et al.*, 2007; Raugi *et al.*, 2013). Ces mutations peuvent aussi se combiner, par exemple le double mutant I54M/I82F est résistant au LPV et hypersensible au SQV (Raugi *et al.*, 2013). Les mutations V62A et L99F confèrent une résistance à trois IP : IDV, NFV, et LPV (Ntemgwa *et al.*, 2007). Cependant, certaines études aboutissent à des résultats contradictoires. Par exemple, dans leur étude, Raugi *et al.*, 2013 identifient la mutation I82F comme sensible au LPV et la mutation V47A hypersensible au SQV.

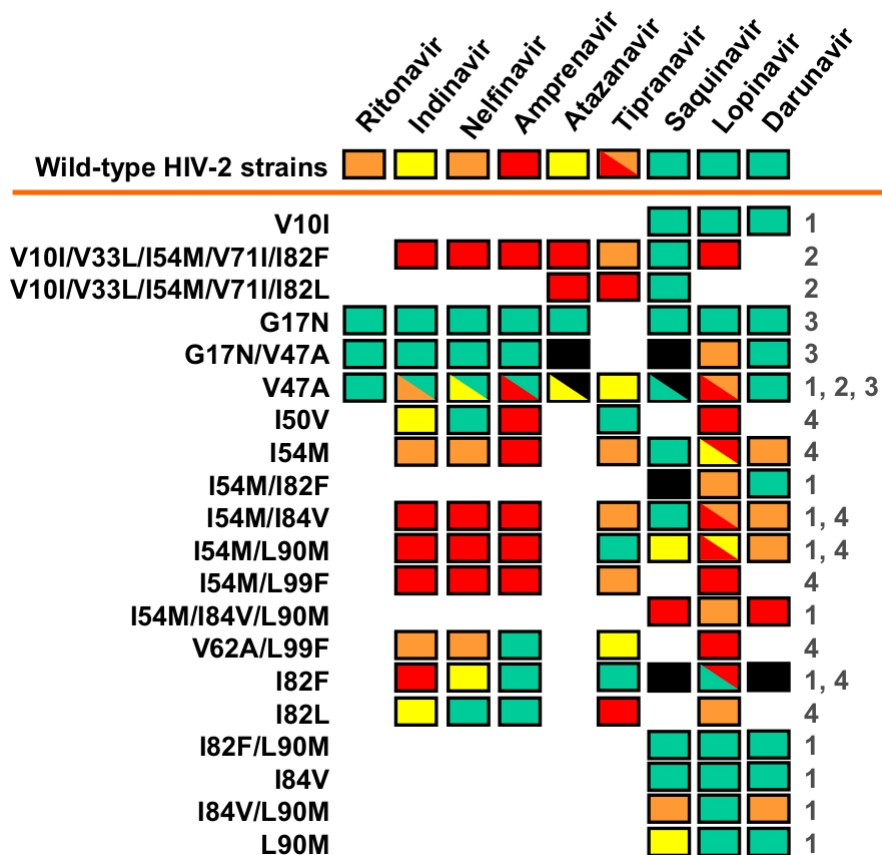


Figure 1.7 : Bilan des données de susceptibilité obtenues par des méthodes phénotypiques pour les structures *WT* et mutantes contre différents IP. Les couleurs renseignent sur le niveau de susceptibilité de manière quantitative : en rouge, une résistance forte (augmentation de l'IC₅₀ d'un facteur supérieur à 20) ; en orange, une résistance modérée (augmentation de l'IC₅₀ d'un facteur 5 à 20) ; en jaune, une résistance faible (augmentation de l'IC₅₀ d'un facteur 2,5 à 5) ; en vert les isolats susceptibles ; en noir, une hypersusceptibilité (diminution de l'IC₅₀ d'un facteur supérieur à 5). Figure extraite de [Menéndez-Arias and Alvarez, 2014](#) (fig. 8), réutilisation autorisée (n° licence : 5105421118386).

Ces résultats sont en contradiction avec ceux obtenus par [Ntemgwa et al., 2007](#) et [Masse et al., 2007](#), qui ont identifié la mutation I82F résistance au LPV ([Ntemgwa et al., 2007](#)) et la mutation V47A comme résistante au SQV ([Masse et al., 2007](#)).

1.3 Projet de thèse

Au moment où j'ai commencé ma thèse, aucune étude structurale des mutations de résistance chez PR2 n'avait été menée ; c'est dans cette optique que s'inscrit ce travail.

Dans un premier temps (chap. 2), nous avons collecté 30 mutants de résistance de la littérature en vue d'étudier l'impact de ces mutations. Faute de structures cristallographiques mutantes disponibles dans la PDB, un protocole de modélisation *in silico* automatisé de ces structures a été mis en place. À partir des modèles de structures mutantes, nous avons comparé leur l'asymétrie structurale à celle de la PR2 sauvage. En effet, plusieurs études ont montré que la PR2 est une protéine asymétrique, c'est-à-dire qui

présente des conformations différentes au niveau de ses deux chaînes malgré ses séquences identiques. L'étude de l'asymétrie structurale d'une cible est une approche efficace pour identifier les changements structuraux induits dans cette cible. La comparaison de l'asymétrie structurale entre les structures sauvages et mutantes nous a permis de mettre en évidence les réarrangements structuraux induits par les mutations étudiées sur le squelette peptidique des structures mutées modélisées en étudiant l'asymétrie entre les structures sauvages et mutantes.

Dans un second temps (chap. 3), nous avons mis à jour la précédente liste de mutants de résistance afin de prendre en considération une plus grande diversité de mutants. Le même protocole de modélisation a été utilisé pour la construction des structures mutantes. Afin d'aller plus loin, nous nous sommes intéressés aux impacts des mutations sur les chaînes latérales en plus du squelette peptidique. Nous avons caractérisé et quantifié les changements structuraux causés par les mutations et analysé leur impact notamment sur les propriétés de la poche de liaison aux IP.

Enfin, dans un troisième temps (chap. 4), nous avons débuté la prise en compte de l'aspect dynamique de la PR2 par des approches de dynamique moléculaire. Pour ce faire, des simulations de dynamique moléculaire d'1 μ s sur les structures sauvages et mutantes ont été lancées. La comparaison des dynamiques des formes semi-ouverte et forme fermée a mis en évidence l'impact sur la conformation des flaps selon la structure de départ. La comparaison avec les structures mutantes a été réalisée à partir de la forme fermée à l'aide d'une cartographie selon les 13 régions des structures échantillonnées pendant la dynamique et de l'étude de distances entre des positions spécifiques de la PR2. Cela a permis de déterminer les régions et les résidus sur lesquels les mutations ont une influence.

Face à ces deux types de VIH, la thérapeutique en place n'est pas adaptée au traitement de l'infection au VIH-2. En effet, le VIH-2 est résistant à un grand nombre d'IP parmi les molécules autorisées (six sur neuf). De plus, il existe des mutations de résistance qui réduisent l'efficacité à certains IP, d'où la nécessité de développer de nouveaux inhibiteurs pour traiter l'infection au VIH-2. Une des stratégies pour identifier de nouveaux IP est la compréhension des mécanismes moléculaires aboutissant à la résistance. Cela permettrait d'aider au développement de nouvelles molécules thérapeutiques qui ne sont pas affectées par des phénomènes de résistance.

2 Impact des mutations sur l'asymétrie structurale de la PR2

Un protocole a été mis en place au laboratoire pour étudier l'asymétrie structurale de la PR2 en comparant les structures locales des 19 structures cristallographiques (Triki *et al.*, 2018b; Ollitrault *et al.*, 2018). Ce protocole a permis de montrer que les deux chaînes de PR2 présentent des conformations différentes, et que cette propriété n'est pas due à des biais expérimentaux. Cette asymétrie est liée à plusieurs facteurs : la flexibilité naturelle de la PR2, les changements conformationnels dus à la fixation des ligands et à la dimérisation de la PR2. L'étude de l'asymétrie de la PR2 s'est montrée être une approche efficace pour explorer et localiser les changements structuraux induits au niveau du squelette peptidique.

Pour étudier les impacts des mutations de résistance observées chez PR2 sur sa structure, nous avons, dans un premier temps, choisi d'explorer si ces mutations modifient l'asymétrie structurale de la PR2. Pour cela, nous avons choisi de comparer l'asymétrie de la PR2 sauvage et celle de 30 mutants en utilisant le protocole développé par Triki *et al.*, 2018a. Ce protocole de modélisation des structures mutantes a été développé afin de pallier l'absence de structures mutantes de PR2 dans la PDB.

2.1 Choix des mutations

Un ensemble de 30 mutants de résistance à au moins un inhibiteur de protéase a été collecté à partir de la littérature. Ces mutants possèdent soit une mutation (simples mutants), deux mutations (doubles mutants) ou trois mutations (triples mutants). L'ensemble de ces mutants contient 14 mutations différentes (figure 2.1). La localisation de ces mutations sur la structure 3D de la PR2 montre que ces mutations sont réparties sur l'ensemble de la structure et ne se limitent pas à la poche de liaison des inhibiteurs (figure 2.1). Ceci suggère qu'il peut exister plusieurs mécanismes de résistance suivant les mutations.

2.2 Modélisation des mutants

Afin de localiser les changements conformationnels induits par les mutations, nous avons comparé l'asymétrie des structures sauvages et mutantes de la PR2. Faute de disposer de structures cristallographiques de mutants de la PR2 dans la PDB, il a été nécessaire de construire des modèles *in silico*. Avant d'appliquer le protocole de modélisation s'est posée la question du choix de la structure de départ pour la mutagenèse. Comme précisé précédemment, la PDB contient 19 structures cristallographiques de la PR2, uniquement sauvages. Une de ces structures correspond à la forme non complexée à un ligand, les 18 autres sont des complexes avec différents ligands. Ces structures présentent des résolutions convenables allant de 1,18 à 3 Å. Parmi les 18 complexes

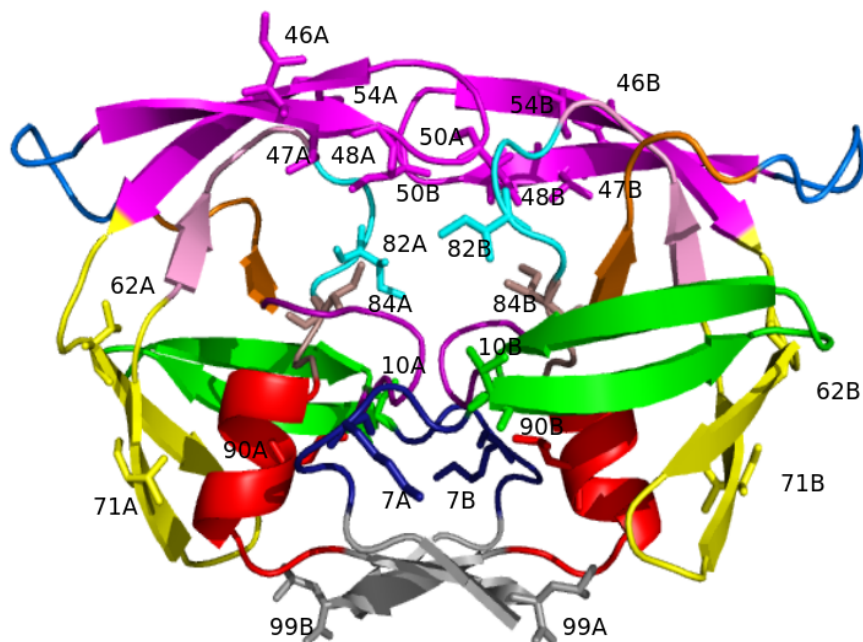


Figure 2.1 : Localisation des mutations de résistance sur la PR2. L'image est adaptée de [Laville et al., 2020](#). Les mutations sont en *sticks* et les couleurs représentent les 13 régions de la PR2.

disponibles, trois correspondent à la PR2 complexée à un médicament : DRV, APV et IDV. Nous avons choisi d'utiliser comme structure modèle la PR2 complexée au DRV (code PDB : 3EBZ ; [Kovalevsky et al., 2008](#)) car elle de très bonne résolution (1,20 Å) et est la seule structure disponible complexée avec un médicament recommandé pour traiter l'infection au VIH-2.

Le protocole de modélisation des structures mutantes que nous avons utilisé couple une étape de mutagénèse de la structure 3EBZ et une étape de minimisation de la structure mutante obtenue. La figure 2.2 présente le protocole de modélisation pour un mutant donné. La première étape de ce protocole a consisté à retirer le ligand du complexe PR2-DRV (code PDB : 3EBZ), ainsi que les molécules d'eau et les hétéroatomes. La structure obtenue a ensuite été optimisée en utilisant la commande `RepairPDB` du logiciel FoldX ([Guerois et al., 2002](#)). Cette commande est un pré-requis habituel de l'outil FoldX qui permet de corriger d'éventuels angles de torsion, de clashes de van der Waals ou d'énergie totale incorrects. La mutagénèse *in silico* (le changement d'acide aminé) de la structure optimisée a ensuite été réalisée à l'aide de la commande `BuildModel` du logiciel FoldX ([Guerois et al., 2002](#)). Cette commande remplace les acides aminés correspondant aux mutations à partir d'une bibliothèque de rotamères de chaînes latérales et optimise ensuite les chaînes latérales aux alentours des résidus mutés. Cette opération est répétée cinq fois pour chaque mutant pour prendre en compte les différents rotamères des acides aminés mutés. À la fin de l'étape de mutagénèse, cinq structures modèles sont obtenues par mutant. Une étape de minimisation énergétique de chaque modèle est ensuite effectuée

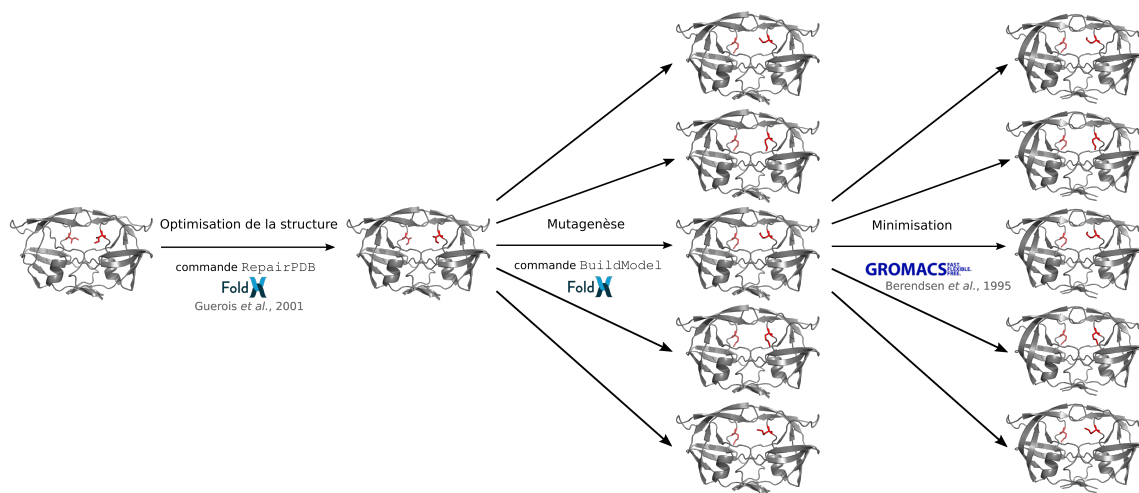


Figure 2.2 : Protocole de modélisation des structures mutantes.

pour relaxer les structures et prendre en compte les impacts des mutations sur le squelette peptidique et ceux à longue portée. Cette étape est réalisée avec le logiciel GROMACS (Berendsen *et al.*, 1995). Au préalable, l'Asp 25 de la chaîne B de chaque modèle a été protonée avec le logiciel PROPKA. Le système est solvaté dans une boîte d'eau (modèle TIP3P) octaédrique avec une distance de 12 Å aux bords de la boîte. Des ions chlore et sodium sont ajoutés pour neutraliser le système. Le champ de force utilisé est AMBER ff99SB (Maier *et al.*, 2015). La minimisation est une alternance de 1000 à 2000 itérations des algorithmes de la plus grande pente (*steepest descent*) et des gradients conjugués (*conjugated gradients*). Ce protocole de modélisation a été appliqué aux 30 mutants étudiés et a abouti à un jeu de 150 structures mutantes (30 mutants \times 5). Afin de prendre en compte les mouvements atomiques dus à la minimisation, nous avons appliqué le protocole de minimisation à la structure sauvage. Cette étape a abouti à la génération de trois structures sauvages minimisées.

2.3 Détection de l'asymétrie structurale

2.3.1 Présentation de l'alphabet structural HMM-SA

La détection et la quantification de l'asymétrie des structures sauvage et mutantes a été effectuée à l'aide du protocole développé par Triki *et al.*, 2018a. Ce protocole permet de détecter l'asymétrie locale d'un dimère sans nécessiter la superposition des deux chaînes comme dans d'autres protocoles (Mulichak *et al.*, 1993; Tong *et al.*, 1993; Priestle *et al.*, 1995). Il se base sur l'alphabet structural HMM-SA (*Hidden Markov Model - Structural Alphabet* : Camproux *et al.*, 2004; voir figure 2.3). HMM-SA est une librairie de 27 fragments de quatre carbones α obtenus après classification de l'ensemble de fragments

extraits de 1429 structures cristallographiques non redondantes. Cette classification a permis de regrouper les fragments en 27 classes suivant leur géométrie et leur transition et a été réalisée en utilisant les chaînes de Markov cachées. Les classes de fragments sont appelées lettres structurales et sont symbolisées par les lettres [A-Z, a]. À partir de HMM-SA, une structure 3D d'une protéine peut être encodée en une séquence de lettres structurales, où chaque lettre symbolise le repliement d'un fragment de quatre carbones α . Cet encodage consiste à découper la structure 3D en fragments de quatre carbones α . Ensuite, la lettre structurale la plus proche de chaque fragment en termes de géométrie et de probabilité de transition est recherchée parmi les 27 lettres de HMM-SA à l'aide de l'algorithme Viterbi (Rabiner, 1989). Il a été montré que HMM-SA permet de décrire précisément la structure des protéines avec quatre lettres décrivant les hélices, cinq lettres décrivant les feuillets et 18 décrivant les boucles. Ainsi, HMM-SA apporte une description précise des structures protéiques avec 27 états. De plus, il permet de simplifier la comparaison des structures protéiques. En effet, en utilisant HMM-SA, la comparaison de deux chaînes protéiques revient à comparer deux séquences de lettres structurales et ne nécessite aucune superposition des chaînes.

2.3.2 Détection et quantification de l'asymétrie structurale d'une structure

Pour détecter l'asymétrie d'un dimère, la première étape du protocole a consisté à simplifier la structure 3D des deux chaînes en deux séquences de lettres structurales. Pour chaque position, la lettre structurale correspondante des deux chaînes A et B sont comparées :

- si les lettres structurales sont identiques, alors la position présente la même conformation dans les deux chaînes et donc la position est symétrique,
- si les deux lettres structurales sont différentes, alors la position présente des conformations différentes dans les deux chaînes et elle sera qualifiée de position asymétrique (figure 2.3).

Une fois les positions asymétriques déterminées dans toutes les structures sauvages et mutantes, nous avons quantifié l'asymétrie globale de chaque structure en déterminant son nombre total de positions asymétriques. De plus, pour quantifier la conservation des positions asymétriques dans le jeu des structures mutantes, nous avons calculé, pour chaque position, le nombre de structures mutantes ayant la position comme asymétrique.

Pour étudier l'impact des mutations de résistance, nous avons ensuite localisé les changements d'asymétrie entre les structures sauvages et mutantes. Deux types de changements ont été définis :

- Les positions qui sont **symétriques** dans les structures sauvages et qui sont devenues asymétriques dans les structures mutantes.
- Les positions qui sont **asymétriques** dans les structures sauvages et qui sont devenues symétriques dans les structures mutantes.

Ces modifications d'asymétrie sont dues à des changements de conformations entre les résidus des deux chaînes. Ces changements structuraux peuvent s'expliquer par deux paramètres : (i) la flexibilité des résidus et/ou (ii) l'impact des mutations. Nous avons donc étudié plus en détail ces positions présentant des changements d'asymétrie entre les structures sauvages et mutantes afin d'identifier l'origine de ces changements d'asymétrie.

2.4 Détection de l'asymétrie structurale dans les structures sauvages

Nous avons appliqué le protocole de détection de l'asymétrie aux trois structures sauvages minimisées. Nous avons observé que la moitié (53 %) de leurs positions sont symétriques, c'est-à-dire qu'elles présentent la même conformation dans les deux chaînes. Ces positions symétriques sont réparties sur l'ensemble de la structure, en particulier dans les régions du fulcrum, des flaps et du cantilever. À l'inverse, les trois structures minimisées présentent entre 34 et 38 positions asymétriques par structure, c'est-à-dire qui présentent des conformations différentes dans les deux chaînes, notamment sur les régions de l'hélice alpha et du cantilever. Parmi ces positions asymétriques, 27 le sont dans les trois structures sauvages, montrant que cette asymétrie semble être importante pour la structure de la PR2 ou son activité biologique.

2.5 Asymétrie structurale des structures mutantes

Nous avons ensuite dénombré le nombre de positions symétriques et asymétriques dans chacune des 150 structures mutantes. Un total de 26 % des positions ont été détectées comme symétriques dans les 150 structures mutantes. Nous avons constaté que ces mêmes positions sont aussi symétriques dans les trois structures sauvages. Ces positions se retrouvent principalement dans les régions du fulcrum, des flaps et du cantilever et cinq d'entre elles au niveau du site actif. Parmi ces positions symétriques, quatre correspondent à des positions mutées (positions 7, 10, 71, 82). Ces résultats suggèrent que les mutations de résistance ont peu d'impact sur ces positions.

Concernant les positions asymétriques, une PR2 mutante contient en moyenne 32 ± 4 positions asymétriques. La figure 2.4 présente la distribution du nombre de positions asymétriques dans les structures mutantes. Cette figure montre qu'il existe une variabilité de l'asymétrie structurale dans les cinq structures d'un mutant donné. Nous remarquons que le mutant I50V (mutant m6) présente peu de variabilité (ses structures contiennent entre 32 et 35 positions asymétriques), alors que le nombre de positions asymétriques dans les cinq structures du mutant I46V (m3) varie de 20 à 36. Ceci montre l'importance de générer plusieurs structures mutantes par mutant. En moyenne, une position est asymétrique dans 67 structures mutantes (± 52).

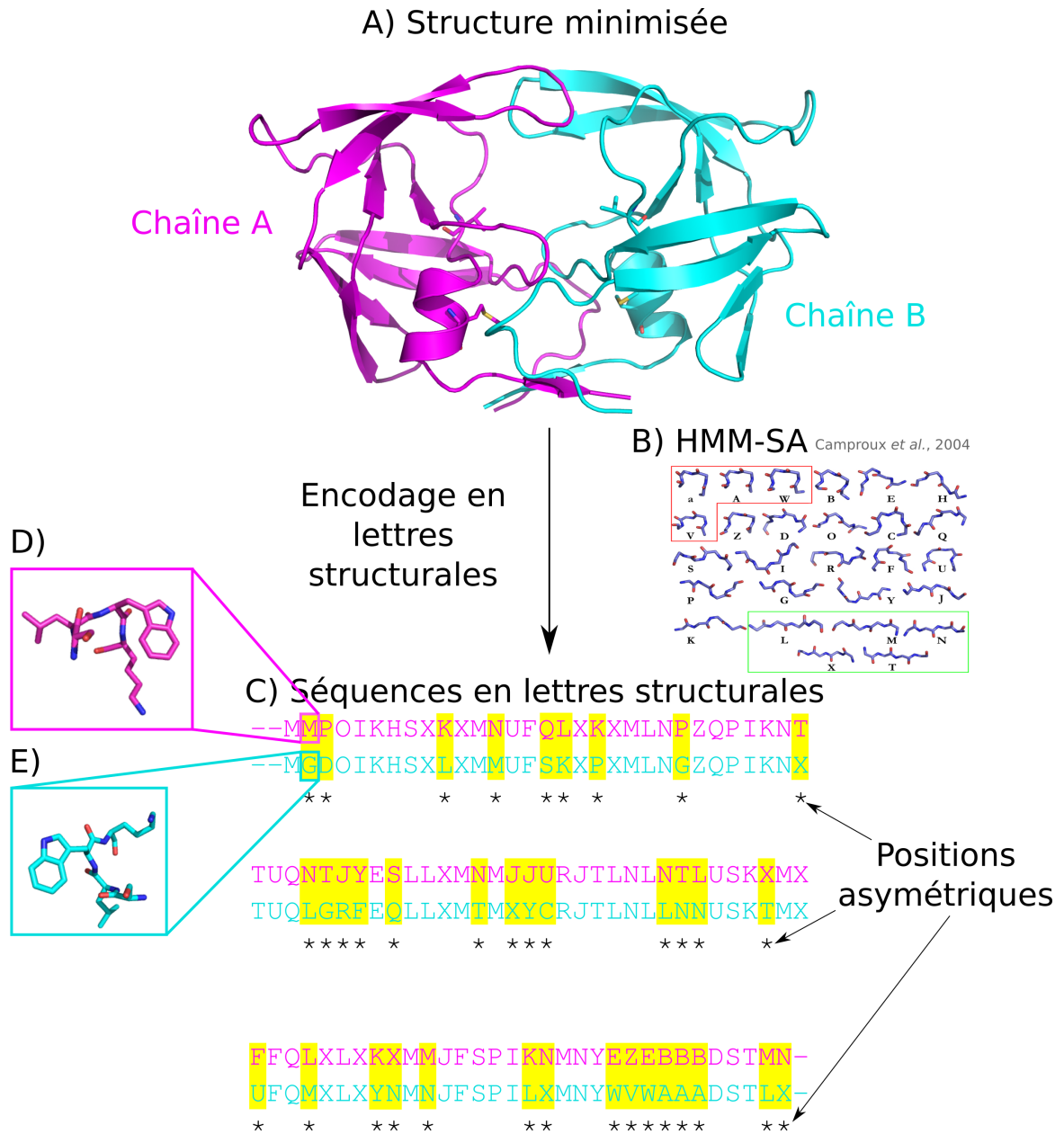


Figure 2.3 : Protocole de détection des positions asymétriques dans une structure. A) Représentation de la structure 3D de la PR2 en mode *cartoon* et colorée suivant ses deux chaînes. La chaîne A est en magenta et la chaîne B en cyan. B) Représentation de la géométrie des 27 lettres structurales de HMM-SA. C) Séquence de lettres structurales des chaînes A (magenta) et B (cyan) de la structure, obtenues après encodage par HMM-SA. D)-E) Représentation de la géométrie des lettres M et G extraites des séquences de lettres structurales. « * » : permet d'identifier les positions asymétriques, c'est-à-dire les positions qui présentent des conformations locales (lettres structurales) dans les deux chaînes.

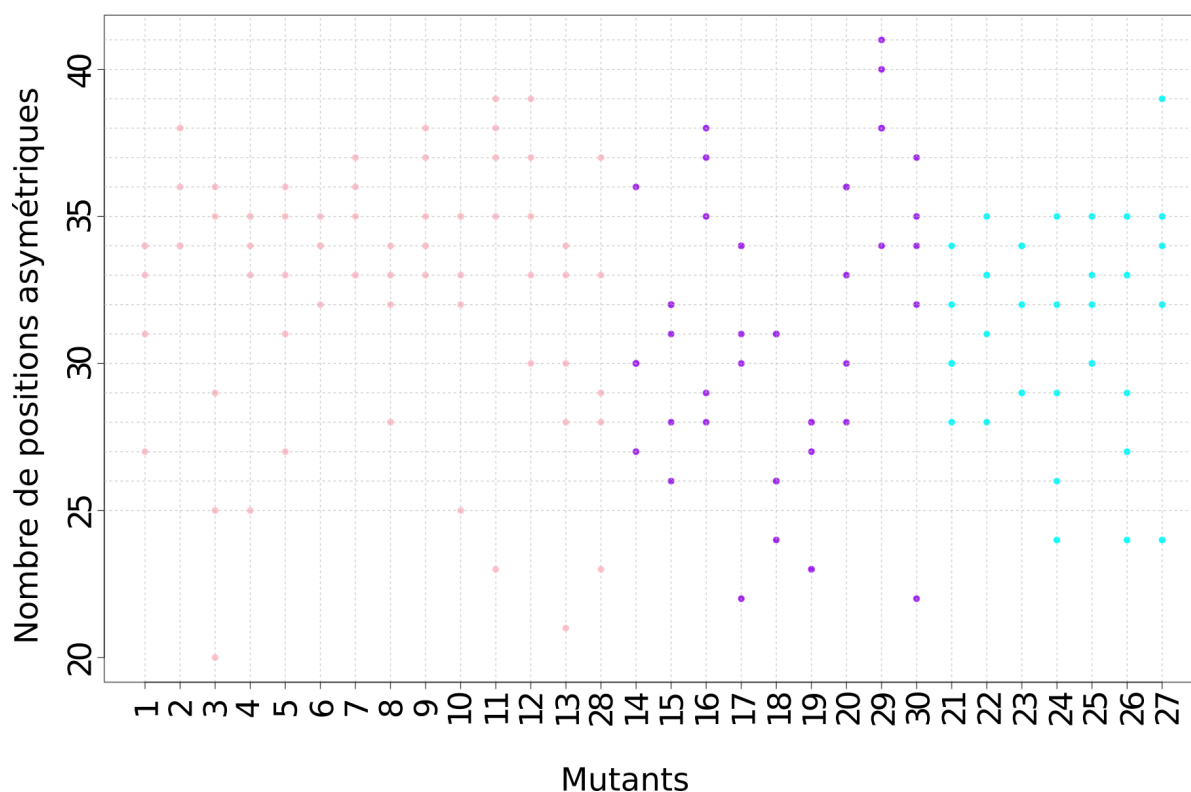


Figure 2.4 : Nombre de positions asymétriques dans les cinq structures de chaque mutant. Les structures mutantes sont colorées suivant le nombre de mutations qu'elles contiennent : les simples mutants sont colorés en rose, les doubles en violet et les triples en cyan.

Parmi les positions asymétriques de PR2, nous avons détecté 36 positions qui sont sur-représentées en termes d'asymétrie, c'est-à-dire asymétriques dans une majorité de structures. La caractérisation de ces positions a montré que ces positions ne sont pas les positions les plus flexibles (en termes de facteur B) ou les plus exposées, en termes de surface accessible au solvant (SAS). Ainsi, l'asymétrie de ces positions semble être une propriété conservée au sein des différents mutants qui s'explique par d'autres paramètres que la flexibilité intrinsèque de ces positions. Parmi ces positions sur-représentées en termes d'asymétrie, 11 sont symétriques dans les trois structures sauvages. De plus, deux positions (positions 40 et 41) sont asymétriques dans les structures sauvages et symétriques dans la majorité des mutants. Ainsi pour ces 13 positions, nous observons un changement structural au niveau du squelette peptidique entre les structures sauvages et mutantes. La localisation de ces 13 positions a montré que quatre d'entre elles correspondent à une position mutée ou sont proches d'un résidu muté. Ces résultats suggèrent que les mutations de résistance peuvent être responsables des changements structuraux observés.

Dans la suite de l'analyse, nous nous sommes focalisés sur l'étude des positions présentant des changements d'asymétrie dans au moins une structure mutante par rapport aux structures sauvages. Afin d'étudier les liens entre les mutations et les changements

d'asymétrie, un réseau reliant les mutants aux positions dans lesquelles un changement d'asymétrie était observé a été construit. Ce réseau nous permet de différencier les mutants qui présentent de nombreux changements d'asymétrie par rapport aux structures sauvages (par exemple le mutant K7R/I46V/V62A, mutant m25) et de ceux qui présentent peu de changements, par exemple le mutant V10I (mutant m2). De plus, nous observons que certains changements d'asymétrie en une position donnée sont communs à de nombreux mutants, alors que d'autres sont spécifiques à certains mutants comme le changement d'asymétrie en position 78 qui est observé uniquement chez le mutant 16.

Pour identifier plus précisément les mutations qui sont responsables des changements structuraux modifiant l'asymétrie de la PR2, nous avons étudié la localisation des changements d'asymétrie de toutes les structures mutantes contenant une mutation donnée. En fonction de la proximité entre la position mutée et le changement d'asymétrie, nous pouvons en déduire l'impact ou non de la mutation. Si la position mutée est proche d'une position asymétrique, cela laisse à penser à un effet de la mutation. À l'inverse, si la position mutée est éloignée, cela pourrait s'expliquer par la flexibilité intrinsèque de la protéine. La figure 2.5 présente ces résultats pour les structures mutantes contenant la mutation I84V. Certains des changements asymétriques se localisent loin du résidu muté : par exemple, les changements d'asymétrie en position 41 (changement identifié chez la majorité des mutants) et en position 98. Ces résultats suggèrent que cette mutation n'est pas responsable des changements structuraux observés sur cette position. De plus, des changements d'asymétrie en position 31, 33, 83 et 84, positions qui sont proches du résidu muté, ont été identifiés dans les structures contenant la mutation I84V (figure 2.5).

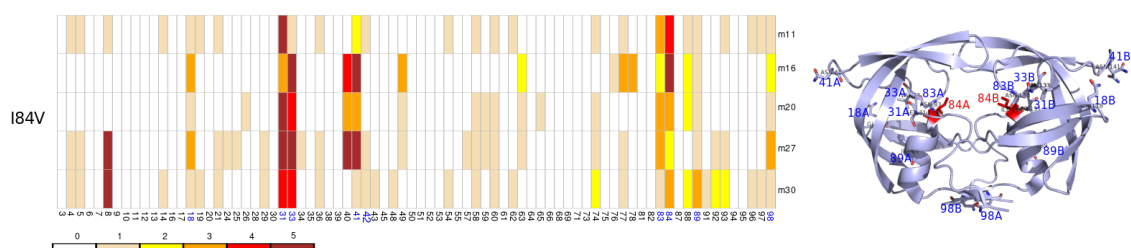


Figure 2.5 : Changement d'asymétrie des six mutants qui présentent la mutation I84V. La position mutée est en *sticks* rouges et les positions asymétriques en *sticks* bleus. L'image est adaptée de Laville *et al.*, 2021.

Ces résultats suggèrent que la mutation I84V est responsable de ces changements structuraux. De plus, les résidus 31 et 84 sont des résidus impliqués dans la poche de fixation des ligands, alors que les résidus 33 et 83 sont des résidus importants pour la conformation de ce site de liaison. Ainsi, il semble que la mutation I84V a un impact direct ou indirect sur la conformation du site de liaison. Cette étude par mutation a suggéré que certaines mutations, telles que la mutation L99F, induiraient des changements structuraux au niveau de l'interface des deux monomères.

2.6 Conclusion

Cette étude explore l'impact des mutations de résistance sur l'asymétrie, propriété impliquée dans la dimérisation, la reconnaissance et la liaison du ligand. Les résultats sont présentés dans l'article de [Laville *et al.*, 2020](#), reproduit dans les pages suivantes. La détection et la comparaison des positions asymétriques, qui ne présentent pas la même conformation dans les deux chaînes d'un dimère, des structures sauvages et des modèles mutants a permis d'identifier les positions capables de se déformer du fait de leur flexibilité ou des mutations. La localisation de ces réarrangements structuraux identifiés grâce à l'identification des changements d'asymétrie des structures sauvages et mutantes a permis d'identifier trois mécanismes de résistance.

Le premier correspond à des changements structuraux induits par les mutations qui modifient l'interface de PR2 et qui pourrait diminuer sa stabilité (exemple, la mutation L99F). Le deuxième et le troisième altèrent l'interaction de l'IP de manières directe et indirecte. Par exemple, nous avons montré que les mutations I46V et I84V induisent des changements conformationnels au niveau de résidus impliqués dans l'interaction avec les IP. De plus, nos résultats ont montré que certaines mutations ont aussi un impact sur le volume et/ou l'hydrophobicité du site de liaison en modifiant la conformation des résidus proches du site de liaison (exemple la mutation I84V). La principale limite de cette étude repose sur le fait que la détection de l'asymétrie par l'alphabet structural ne considère que le squelette de la protéine. Cela sera l'objet de l'étude suivante qui prend en compte les chaînes latérales en plus du squelette de la protéine.

2.7 1^{er} article

Ci-dessous est reproduit notre premier article *Impacts of drug resistance mutations on the structural asymmetry of the HIV-2 protease*, publié dans le journal *BMC Molecular and Cell Biology* le 23 juin 2020 ([Laville *et al.*, 2020](#)).

RESEARCH ARTICLE

Open Access

Impacts of drug resistance mutations on the structural asymmetry of the HIV-2 protease



Pierre Laville, Sandrine Fartek, Natacha Cerisier, Delphine Flatters, Michel Petitjean and Leslie Regad*

Abstract

Background: Drug resistance is a severe problem in HIV treatment. HIV protease is a common target for the design of new drugs for treating HIV infection. Previous studies have shown that the crystallographic structures of the HIV-2 protease (PR2) in bound and unbound forms exhibit structural asymmetry that is important for ligand recognition and binding. Here, we investigated the effects of resistance mutations on the structural asymmetry of PR2. Due to the lack of structural data on PR2 mutants, the 3D structures of 30 PR2 mutants of interest have been modeled using an in silico protocol. Structural asymmetry analysis was carried out with an in-house structural-alphabet-based approach.

Results: The systematic comparison of the asymmetry of the wild-type structure and a large number of mutants highlighted crucial residues for PR2 structure and function. In addition, our results revealed structural changes induced by PR2 flexibility or resistance mutations. The analysis of the highlighted structural changes showed that some mutations alter protein stability or inhibitor binding.

Conclusions: This work consists of a structural analysis of the impact of a large number of PR2 resistant mutants based on modeled structures. It suggests three possible resistance mechanisms of PR2, in which structural changes induced by resistance mutations lead to modifications in the dimerization interface, ligand recognition or inhibitor binding.

Keywords: HIV-2 protease, Drug-resistance mutations, Structural asymmetry, Structural alphabet

Background

Proteases (PRs) are an important therapeutic target in the treatment of HIV-1 and HIV-2 infections because of their indispensable role in Gag and Gag-Pol polyprotein hydrolysis during the viral maturation process [1]. Currently, nine drugs targeting PR for the treatment of HIV-1 have been approved for clinical use by the Food and Drug Administration (FDA), only three of which (darunavir (DRV), saquinavir (SQV), and lopinavir (LPV)) are effective in the treatment of HIV-2 infection [1, 2]. Some

studies have shown that the natural resistance of HIV-2 to protease inhibitors (PIs) can be partially explained by substitutions located within the PR2 pocket. These substitutions induce structural changes in pocket residues, modifying pocket properties and the internal interactions between PR2 and PIs, thus altering PI binding [3–10]. Other studies have suggested that substitutions between PR1 and PR2 could modify PI binding by altering the transition between the open and closed forms involved in ligand binding [11, 12].

In addition to the natural resistance of HIV-2 to most HIV-1 PIs, HIV-2 can evolve to achieve drug resistance through the accumulation of mutations on its PR. However, few studies conducted to date provide information

* Correspondence: leslie.regad@u-paris.fr
Université de Paris, BFA, UMR 8251, CNRS, ERL U1133, Inserm, F-75013 Paris, France



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

about mutations conferring resistance to PIs. Genome sequencing studies of HIV-2 from different HIV-2-infected patients have described the selection of some mutations that confer drug resistance in HIV-2 [7, 13–22]. These resistance mutations were subsequently confirmed using phenotypic susceptibility assays [7, 15–17, 19, 23]. Several studies have shown that a single mutation can confer resistance to one or several PIs, such as V47A, I50V, I54M, I82F, I84V, and L90M [7, 15, 17, 19, 24]. However, some studies have led to contradictory results. For example, the I82F mutation has been observed in the presence of ritonavir (RTV) and indinavir (IDV) in some studies [14, 16, 19], while the I54M mutation has appeared under treatment with amprenavir (APV) [16], nelfinavir (NFV) [14], and IDV [19]. Phenotypic susceptibility assays confirmed that the I82F mutation confers resistance to IDV as well as NFV and LPV [16]. In contrast, Raugi et al. [19] found that this mutation does not increase resistance to LPV but causes hypersusceptibility to both DRV and SQV. In addition, some mutations can appear together to confer high resistance to several PIs [7, 15, 17, 19, 24, 25]. For example, the I54M and I82F mutations confer cross-resistance to all PIs. The V62A and L99F mutations confer cross-resistance to three PIs (IDV, NFV, and LPV) [16]. As no tridimensional (3D) structure of the PR2 mutant is available in the Protein Data Bank (PDB [26]), no structural analysis was performed to study the mechanisms explaining HIV-2 resistance induced by these mutations. However, knowledge of these mechanisms leading to PR2 resistance is important for designing new PR2 inhibitors.

PR2 is a homodimer of 99-residue monomers. The interaction between the two monomers occurs through the Nter, catalytic, and Cter regions. Diverse ligands (peptide substrates with different amino acid sequences and structurally different inhibitors) bind to the PR2 central binding site. The comparison of the unbound and PI-bound PR2 structures showed that the flap region undergoes large structural changes upon ligand binding. In the unbound form, the flap region adopts an open conformation allowing ligand entry. Upon ligand binding, the flap region recloses over the central binding pocket [1].

The analyses of the crystallographic structures of PR2 have shown that the two monomers do not exhibit the same global and local conformations, indicating that in crystallographic structures, PR2 exhibits structural asymmetry [27–32]. This structural asymmetry is translated by slightly different orientation of its two monomers, quantified by a two-fold axis of 178.20° to 179.80° and a root mean square deviation (RMSD) of 0.35 to 1.02 Å [27–30, 33]. The largest deviations between the two monomers are located in the tail, elbow, and flap regions [27–31, 33]. Using a structural-alphabet-based approach, 31% of PR2 positions were identified as structurally asymmetric (i.e.,

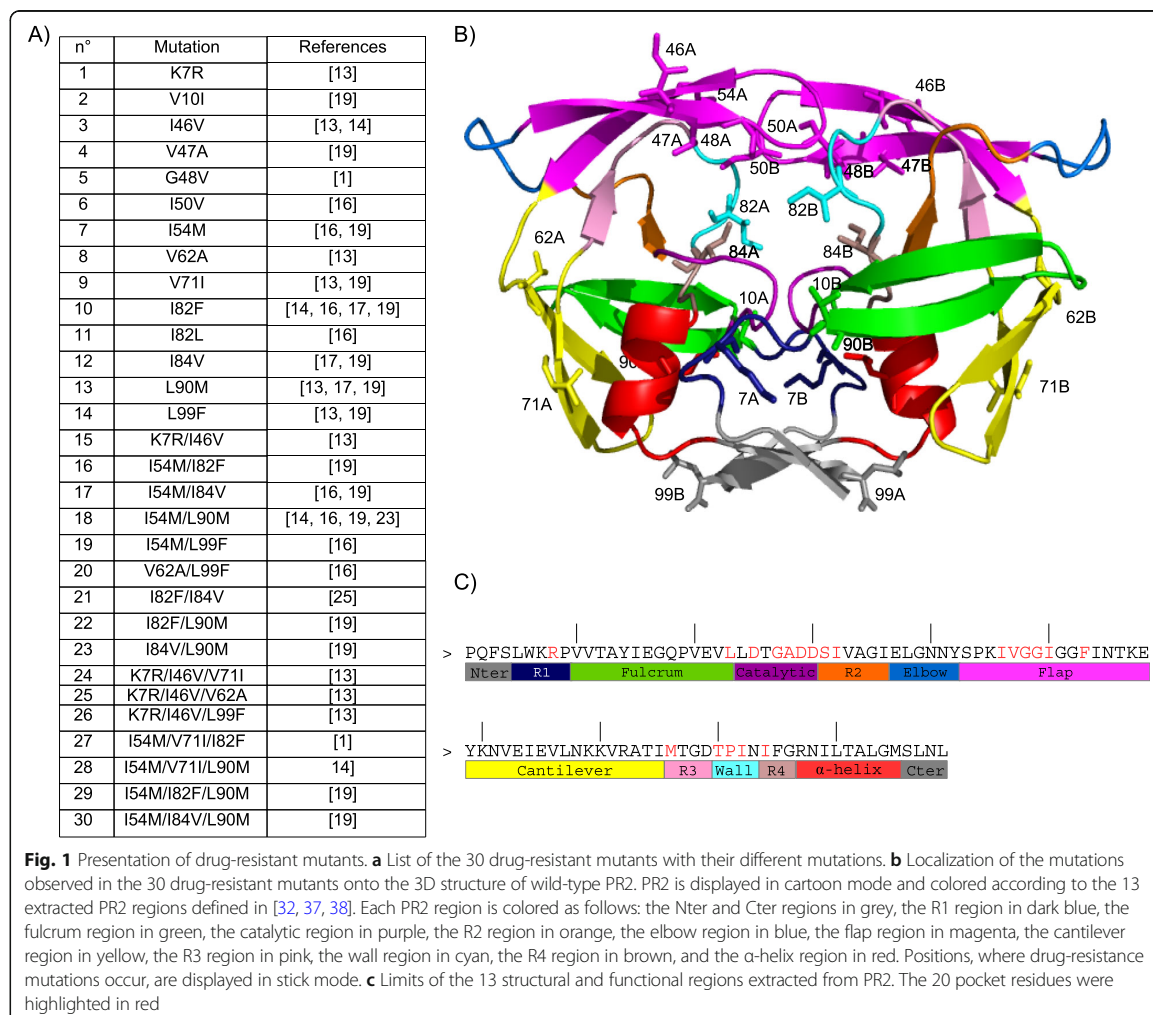
exhibiting different local conformations in their two chains), among which 75% were located outside of the PI-binding pocket [31, 33]. In a recent study, we explored the structural asymmetry of unbound PR [33]. Its two monomers exhibit a $C\alpha$ -RMSD value of 0.53 Å, which is smaller than that computed on the bound PR2 [27–30]. Our structural-alphabet-based approach highlighted that 35% of unbound PR2 positions are asymmetric. In the unbound and bound structures, the asymmetric positions are distributed throughout the structure, particularly in the interface region and in the flap, fulcrum, elbow, and α -helix regions and the binding site [31, 33]. Thus, the crystallographic PR2 structure exhibits structural asymmetry in its backbone, and this property is also found in the unbound structure. These results highlighted the asymmetric properties of the crystallographic structures of PR2, which are not caused by ligand binding alone. Indeed, proteins are dynamic objects that adjust the positions of their atoms to respond to different events, such as partner binding. In the case of the crystallographic PR2 structures, the structural asymmetry results from crystal packing [27, 28, 30, 33], protein dimerization [31], and ligand binding [28, 29, 31]. Different studies have differentiated the PR2 asymmetry induced by ligand binding that is important for ligand recognition and binding [28, 29, 31] to structural asymmetry corresponding to an intrinsic factor allowing the structural deformation of the target [34–36]. For example, Mulichak et al. [28] showed that the binding of a peptidic inhibitor in the PR2 specifically induces a move of the region 79–82 of chain B allowing inhibitor binding.

In this work, we explored the structural effects of some drug resistance mutations of PR2 by comparing the structural asymmetry of the wild-type and drug-resistant mutants of PR2. The studied PR2 drug-resistant mutants harbored one, two or three mutations. As no structural data are available for these PR2 mutants, we constructed their 3D structures using molecular modeling as in [10]. We then detected structural asymmetry (i.e., positions exhibiting different local conformations in the two PR2 chains) in the wild-type and mutant structures using our structural-alphabet-based approach [31]. The comparison of the structural asymmetry of wild-type and mutant structures highlighted three possible mechanisms that could explain PR2 resistance to PIs.

Results

Quantification of PR2 structural deformation induced by drug resistance mutations

We focused on a set of 30 drug-resistant mutants containing from one to three mutations (Fig. 1a). The 3D structure of each mutant was built using FoldX software [39] with five replications (see Methods),



resulting in a set of 150 mutant models. To compare the mutant and wild-type structures, we computed the all-atom RMSD, denoted as $RMSD_{aa}$, between the mutant structure and the three minimized wild-type structures ($3EBZ_{mini}$). The five structures of each mutant could exhibit large $RMSD_{aa}$ (i.e., different conformations) (Additional file 1: Figure S1). As expected, single mutants (i.e., those with only one mutation) exhibited a smaller average $RMSD_{aa}$ than the other mutants, indicating that introducing a single mutation induces less structural change than introducing several mutations (Additional file 1: Figure S1). In addition, there was a link between the number of mutations and structural diversity in terms of $RMSD_{aa}$. Indeed, the three types of mutants (single, double and triple) did not exhibit the same variability in terms of

$RMSD_{aa}$ (Bartlett-test p -value = 2.10^{-11}): triple mutants showed more conserved structures than the other types of mutants. Thus, compared to single or double mutants, the insertion of three mutations induced more structural diversity relative to the wild-type structure, but the five modeled mutant structures were more similar to each other.

Detection of structural asymmetry in the wild-type and mutant structures of PR2

To explore the structural effects of the studied drug resistance mutations, we compared the structural asymmetry in the three $3EBZ_{mini}$ structures and the 150 modeled mutant structures using an approach based on a structural alphabet [31].

Characterization of structural asymmetry in the three 3EBZ_{mini} structures

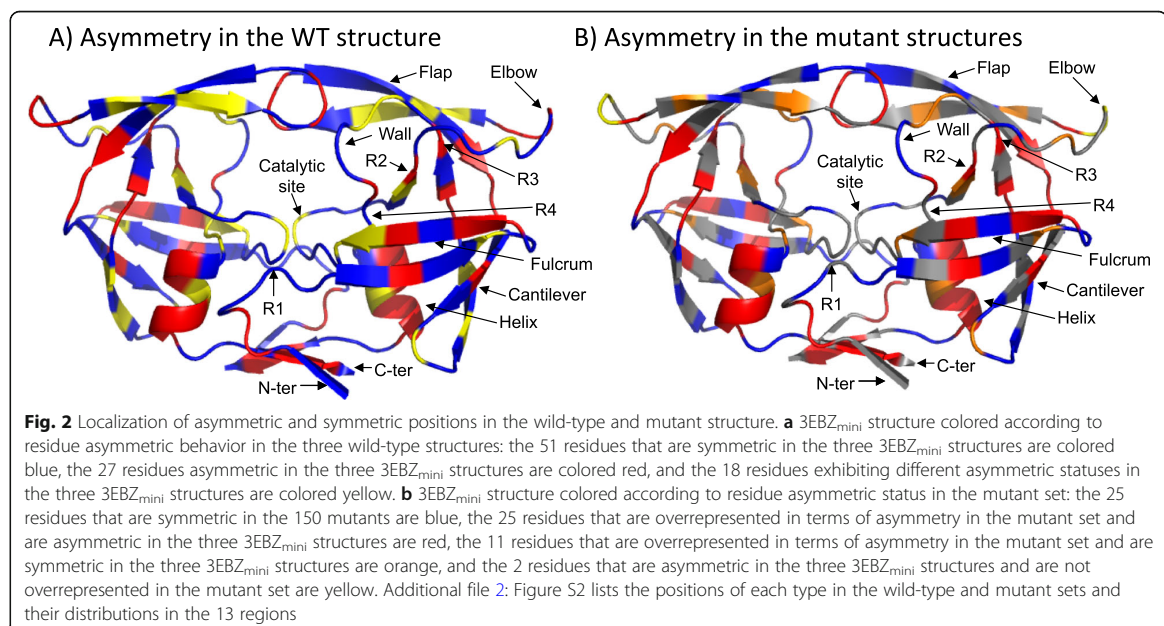
We determined the structural symmetric and asymmetric positions in the three 3EBZ_{mini} structures and their location in the 13 PR2 regions defined in [32, 37, 38] (Additional file 2: Figure S2). Half of the positions (53% of positions) were detected as symmetric (i.e., showing similar local conformations in the two chains in the three 3EBZ_{mini} structures). The entire PR2 structure was sampled, particularly the fulcrum, flap, and cantilever regions (Fig. 2a). Thus, the conserved local conformation in both monomers at these positions is important for PR2, particularly for pocket residues 28, 30, 53, 81 and 82, which could be important for ligand binding. The three 3EBZ_{mini} structures contained between 34 and 38 asymmetric positions, with 28% of the positions showing asymmetry in the three 3EBZ_{mini} structures. These positions were located throughout the structure, particularly in the α -helix and cantilever regions (Fig. 2a). This asymmetry conservation suggests an important role of these positions, particularly for the elbow and flap positions (positions 40–42, 50, 51, 58), which could be important for PR2 deformation.

A total of 18 positions did not exhibit the same asymmetry status in the three 3EBZ_{mini} structures (Fig. 2a). These 18 positions were not more flexible (in terms of B-factor value) or more accessible (in terms of accessible surface area (ASA) values) residues relative to other positions (Kruskal-Wallis test p -values = 0.25 and 0.46, Additional files 3 and 4: Figures S3 and S4). This suggests that the asymmetric variability of these positions

does not result from the intra-flexibility of PR2. However, we have previously shown that these 18 positions exhibit different conformations in the 18 available structures of PR2 complexed with different ligands [32]. Thus, these positions could modify their local conformation to adapt to different ligands.

Characterization of structural asymmetry in the PR2 mutant set

On average, a mutant structure contains 31.8 ± 4.4 asymmetric positions, which is close to the number of asymmetric positions in the three 3EBZ_{mini} structures. The five structures of a given mutant do not always have the same number of asymmetric positions and can exhibit few common asymmetric positions, such as the G48V mutation (m5); see Additional file 5: Figure S5. The variability of the number of asymmetric positions per mutant does not depend on the number of mutations (Bartlett test p -value = 0.14). As expected, a link was found between variability in terms of $RMSD_{aa}$ and both (i) the variability in terms of the number of asymmetric positions (Pearson correlation coefficient = 0.73) and (ii) the number of common asymmetric positions between the five mutant structures (Pearson coefficient between the standard deviation of the $RMSD_{aa}$ value for the 5 mutant conformations and the number of common asymmetric positions = -0.65). Thus, the mutants exhibiting greater diversity in terms of $RMSD_{aa}$ corresponded to the mutants showing five structures exhibiting different numbers of asymmetric positions with few common asymmetric positions.



To characterize the structural asymmetry in the mutant set, we then computed the asymmetry occurrence (AO) for each position (i.e., the number of mutant structures exhibiting asymmetry for a considered position). A total of 26% of the positions that were symmetric in all mutant structures were also symmetric in the three 3EBZ_{mini} structures. This indicates that the resistance mutations do not affect the structural symmetry of these positions, particularly the mutations occurring at some of these symmetric positions (K7R, V10I, V71I, I82F, and I82L). These positions were frequent in the fulcrum, flap, and cantilever regions and absent in the Nter, Cter, elbow, R3, and R4 regions (Fig. 2b). Five of them (28, 30, 53, 81, 82) were located in the binding site, confirming the important role of these positions in ligand binding.

In the mutant set, 36 positions correspond to overrepresented asymmetric positions and were denoted as *OR_{asym}* positions. Their asymmetry does not arise at random. These *OR_{asym}* positions are located throughout the structure except in the R1 and catalytic regions (Fig. 2b). These positions did not consist of more flexible (in terms of B-factor) or exposed (in terms of ASA) residues on average than other asymmetric positions (T-test *p*-values = 0.40 and 0.53, respectively, additional files 3 and 4; Figures S3 and S4). Seventy percent of the *OR_{asym}* positions were also asymmetric in the three 3EBZ_{mini} structures. These positions were particularly common in the α -helix, flap, cantilever, and fulcrum regions (Fig. 2b). Thus, the studied drug resistance mutations do not modify the structural asymmetry of these positions, reinforcing the important role of the structural asymmetry of these positions. These overrepresented asymmetric positions could be important for PR2 structure or activity, particularly the four residues belonging to the dimerization region (4, 5, 97, and 98) and the five pocket residues (23, 32, 47, 50, 80).

The remaining 11 *OR_{asym}* positions were symmetric in the three 3EBZ_{mini} structures. In addition, two positions (40 and 41) were asymmetric in the three 3EBZ_{mini} structures and were not overrepresented in the mutant set. Among these 13 positions, four were close (44 and 80) or corresponded (47 and 90) to mutated positions. Thus, these drug resistance mutations could be responsible for the changes in asymmetry at these mutated positions and their nearby residues, and they could modify the asymmetry of more distant residues.

Link between drug resistance mutations and changes in asymmetry occurring in mutants

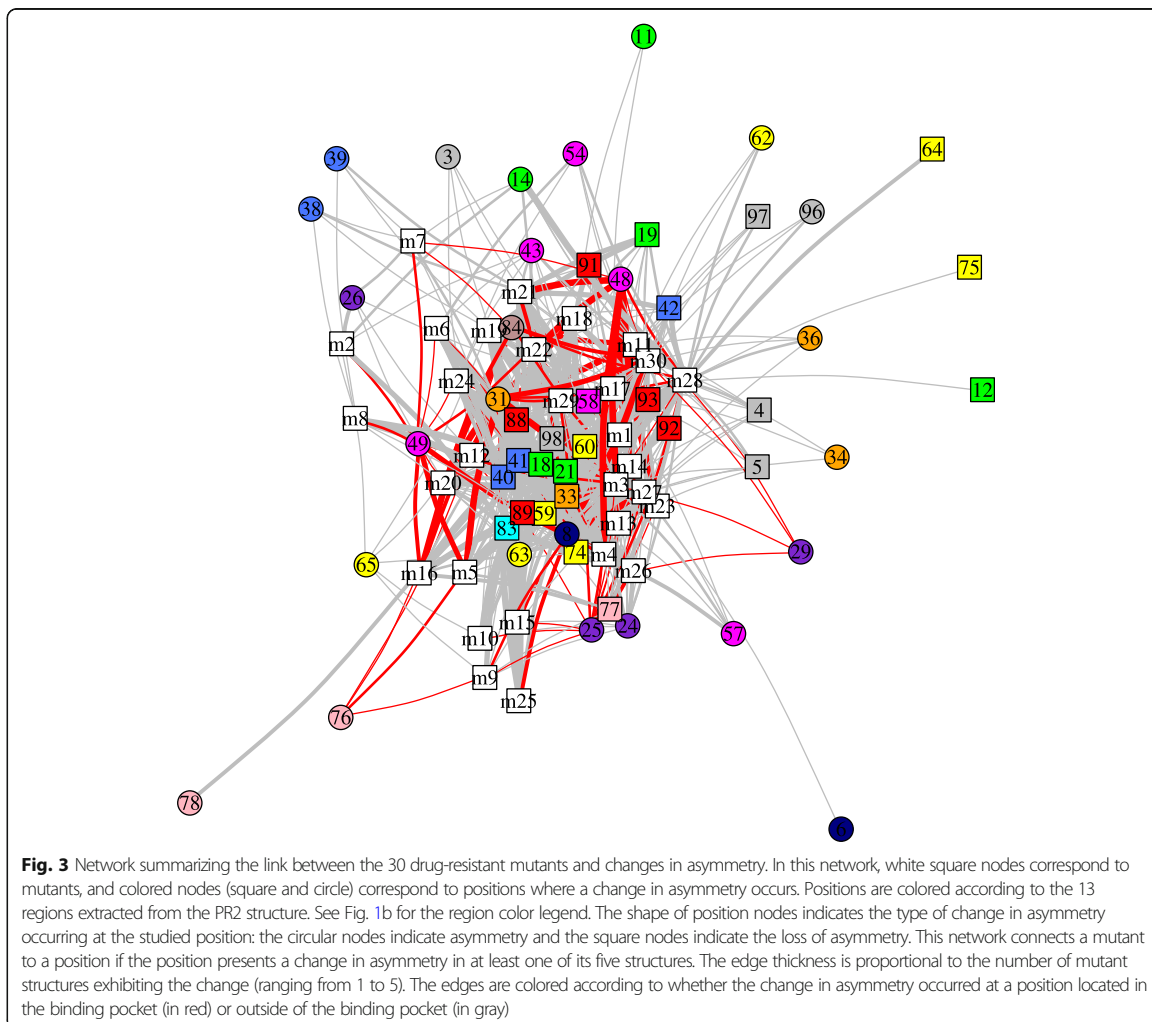
For each mutant, we determined how many of its five structures exhibited a change in asymmetry at each position relative to the three wild-type structures. The number of changes in asymmetry observed for each

mutant varied from 6 (mutant m2) to 36 (mutant m5) and did not depend on the number of mutations (*P*-value of the Kruskal-Wallis test = 0.55). Figure 3 presents the network connecting a mutant with its asymmetric positions. We observed that some changes in asymmetry occurred at positions connected with many mutations (which correspond to central nodes in the network), while others were connected to few mutations (which correspond to external nodes in the network, Fig. 3). For example, changes in asymmetry at positions 40, 41, 33, 18, and 98 were observed in more than 20 mutants (Fig. 3). These changes in asymmetry were not specific to certain mutations, suggesting that they were not induced by mutations. In contrast, structural backbone asymmetry at positions 6 and 78 was observed only in mutants I54M/I84V (m17) and K7R/I46V/L99F (m26), respectively, while such asymmetry at position 62 was observed in mutants I84V (m12), G48V (m5), and I84V/L90M (m23). The loss of structural asymmetry at positions 12, 64, and 75 was only observed in mutant G48V (m5), but the five structures of these mutants did not exhibit this loss.

To highlight the changes in asymmetry that are putatively induced by resistance mutations, we studied the conservation and location of structural changes in all structures containing at least one of the nine mutations observed in multiple mutants (K7R, I46V, I54M, V62A, V71I, I82F, I84V, L90M, and L99F); see Fig. 4. We noted that some changes in asymmetry occurred in all mutants exhibiting a given mutation. For example, a structural change at position 83 was observed in all mutants exhibiting the I82F or I84V mutation. The location of conserved structural changes in the PR2 structures allowed us to differentiate two types of changes in asymmetry: those occurring far from a mutated residue and those occurring close to a mutated residue, which were putatively induced by mutation.

Asymmetric changes occurring far from mutated residues and, thus, putatively not related to mutations

Figure 4 shows that the nonspecific changes in asymmetry occurred at positions 40 and 41 in most mutants, but they were located far from mutated residues. As previously assumed, the high frequency of these two changes in asymmetry suggests that they are not induced by a mutation. This was confirmed by the fact that they occurred at two exposed residues located among the elbow residues (Additional file 4: Figure S4). Thus, the loss of asymmetry observed at positions 40 and 41 could be induced by residue flexibility. In the mutants with K7R, I46V, I54M, I84V, or L99F mutations, changes in



asymmetry at positions 40 and 41 were accompanied by the loss of asymmetry at positions 18, 21, 58, 59, and 60 (Fig. 4). As these positions are located close to flexible residues 40 and 41, we suggest that these changes in asymmetry could be induced by the changes in asymmetry at positions 40 and 41. The nine mutants with the I54M mutation also presented a loss of asymmetry at position 33, which was also observed in all mutants exhibiting either the I84V or L90M mutation. Although this change in asymmetry was specific to certain mutations, it was not located close to the mutated residues (Fig. 4). The seven mutants containing the L90M mutation exhibited asymmetry at position 8, which was located far from the mutated residue 90 (Fig. 4). Thus, it is difficult to reach a conclusion regarding the link between the

changes in asymmetry that occurred at position 33 or 8 and mutations I54M and L90M.

Asymmetric changes close to mutated residues putatively related to mutations

Figure 4 shows that the three mutants with the V62A mutation exhibited asymmetry at positions 38 and 39 and loss of asymmetry at positions 40 and 41. Mutated residue 62 was close to residue 38, which was close to residues 39–41 (Fig. 4). Thus, the V62A mutation could be responsible for the changes in asymmetry at positions 38–41. However, the loss of asymmetry at positions 40 and 41 was highly recurrent in the mutant set, whereas the asymmetry at positions 38 and 39 was specific to V62A. Thus, it is difficult to conclude that the V62A mutation induces changes in asymmetry at positions 38 and 39.

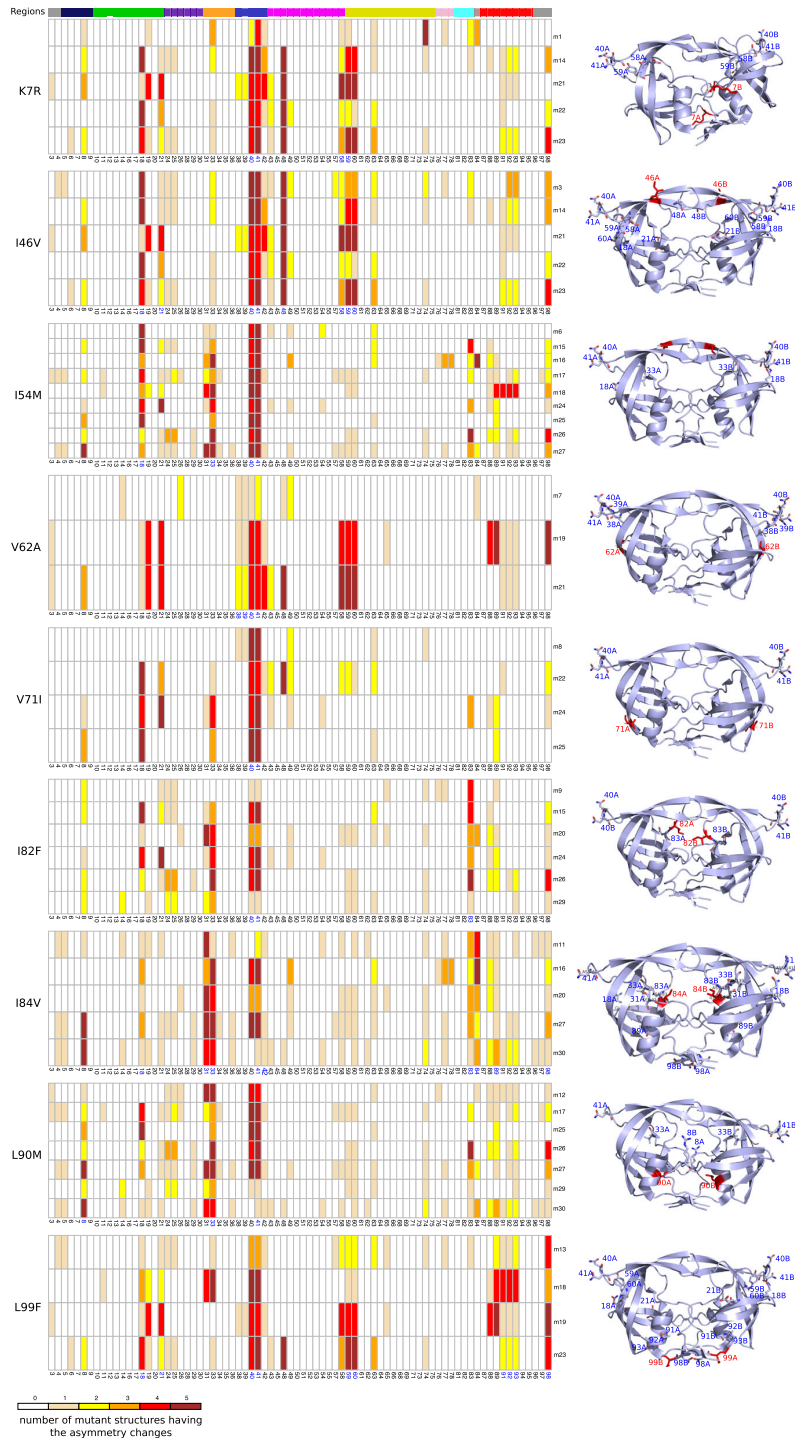


Fig. 4 (See legend on next page.)

(See figure on previous page.)

Fig. 4 Detection of changes in asymmetry in mutant structures. For this analysis, we selected all structures exhibiting the K7R, I46V, I54M, V62A, V71I, I82F, I84V, L90M, or L99F mutation. **a** presents matrices summarizing the positions exhibiting changes in asymmetry for the nine selected mutations. The matrix rows present mutant and matrix columns corresponding to PR2 positions. A matrix cell indicates the number of mutant structures (from 0 to 5) exhibiting changes in asymmetry at a given position. Positions in blue exhibit changes in asymmetry in all mutants. PR2 regions are presented at the top of the figure and are colored according to the color code of Fig. 1. **b** Localization of positions exhibiting changes in asymmetry in the PR2 structures. Proteins are displayed in cartoon mode and colored blue. Mutations are displayed with sticks and colored red. Positions exhibiting changes in asymmetry in the mutant structures relative to the wild-type structure are displayed in stick mode. Only positions exhibiting changes in asymmetry in at least one structure of all mutants harboring the studied mutations are indicated in the 3D structure of PR2

Some changes in asymmetry occurred at residues involved in the PI-binding pocket. For example, all mutants containing the I46V (or I84V) mutation exhibited asymmetry at pocket residue 48 (or 31 and 84). As residues 48 and 84 establish hydrophobic contacts with PIs (Additional file 6: Figure S6), we concluded that the I46V and I84V mutations could induce structural backbone changes at positions 31, 48, and 84 that could modify structural asymmetry and PI interactions.

Other changes in asymmetry occurred at residues that were located outside of the binding site but were important for its conformation, such as residues 33, 83, and 89 [10, 32]. Figure 4 shows that the structures of all mutants with I82F, I82L, or I84V mutations exhibited a loss of asymmetry at these important positions. Thus, these changes in asymmetry could be a consequence of structural changes induced by mutations at positions 82 and 84, which could modify the conformation of the binding pocket and, thus, indirectly alter PI binding.

Concerning the L99F mutation, we observed that the four mutants containing this mutation exhibited changes in asymmetry at positions 91, 92, 93, and 98. Residues 91, 92, 93, and 98 are involved in the interface between the two monomers: residues 91 and 92 establish non bonded contacts at the interface, and residue 98 establishes hydrogen bonds with residues 2 and 96 of the second monomer (Additional file 7: Figure S7). Thus, the L99F mutation could induce structural asymmetry at interface residues that could have an impact on the PR2 interface and modify the stability of the dimer.

Discussion

In this study, we explored the impact of resistance mutations arising in the PR2 target on its backbone asymmetry to obtain information about the structural effects of these mutations. Several studies have shown that although PR2 is a homodimer, its crystallographic structures exhibit structural asymmetry that is important in the mechanism of PI recognition and binding [10, 27–30, 33]. In this work, we compared the local structural asymmetry of a wild-type PR2 structure and 30 resistant mutants. As no structure of the PR2 mutant was available in the PDB, we modeled the structure of the 30 studied mutants using the protocol developed in our

previous study [10] based on FoldX software [39] and an energy minimization step. As FoldX software is based on a side-chain rotamer library, we built five structures per mutant to consider the different rotamers associated with each amino acid. The comparison between the wild-type and mutant structures confirmed the importance of generating several structures per mutant.

In the first step of this study, we analyzed the structural asymmetry in three minimized wild-type structures of PR2 and a set of 150 modeled mutant structures. Considering a large set of PR2 structures (wild-type and mutants) at the same time allowed us to extract information about the role of particular residues. Our results highlighted 25 residues that are symmetric in the wild-type and mutant structures i.e. presenting the same local conformation in the two monomers. Eight of these residues (17, 28, 30, 53, 68, 81, 82, 87) were previously detected as symmetric in a set of 19 structures of wild-type PR2 available in the PDB [31], in which they exhibited the same conformation [32]. This conservation of symmetric status highlights the important role of the symmetry of these residues for PR2. As residues 28, 30, 53, 81, and 82 are located in the PR2 pocket and residue 87 establishes interactions with pocket residues 28–29 [10, 30, 32], we concluded that the conserved conformation of these residues in the two chains is important for the binding site conformation and ligand binding. In contrast, our results identified 25 residues that were characterized as asymmetric in the three minimized wild-type structures and overrepresented in terms of asymmetry in the mutant structures. Twelve of these 25 asymmetric positions (18, 42, 50, 51, 59, 60, 64, 75, 77, 83, 92, and 93) were previously detected as overrepresented in terms of asymmetry in the 19 PR2 structures available in the PDB [31]. These results confirm that the backbone asymmetry of these residues is important for PR2 structure and function [31]. In addition, residues 18, 42, 59, and 75 are involved in the H-bond network with residues of the elbow, which is an important region for the flexibility of PR2 [32]. Thus, the structural asymmetry of these residues seems to be important for the transition between the PR2 open and closed forms. Residues 50 and 51 are involved in the interface of the two monomers, suggesting that the backbone asymmetry of these

residues is induced by dimerization. Residue 83 has been previously shown to be important for the binding site conformation [28, 30, 32]. Thus, the structural asymmetry of residue 83 could be important for PI recognition and binding. All these results confirm that backbone asymmetry is an intrinsic property of PR2 that is involved in its flexibility and ligand recognition and binding. They also confirm the interest in mining several structures associated with a target to offer valuable insight into target structure, flexibility and interaction mechanisms [40–42].

We showed that wild-type and mutant PR2 structures exhibit different asymmetric statuses at some positions. We distinguished changes in asymmetry between wild-type and mutant structures occurring at the same positions in many mutants. Some of these changes in asymmetry occurred at flexible residues and sites located far from mutated residues, such as residues at 18, 40, and 41. Thus, these changes in asymmetry were putatively induced by PR2 intra-flexibility. On the other hand, some changes in asymmetry occurred in some mutants at positions close to the mutated residues, suggesting that these changes in asymmetry resulted from structural changes induced by mutations. By analyzing the locations of these changes in backbone asymmetry, our results suggested different putative mechanisms of resistance, as observed in PR1 [43]. First, we proposed that resistance mutations could induce structural changes at the interface of the two monomers. For example, our results suggest that the L99F mutation might induce changes in asymmetry at interface positions 91, 92, 93, and 98, which could modify the PR2 interface and alter its stability. This deformation of the dimer interface induced by resistance mutation was previously observed in PR1 mutants with an L24I, F53L or I50V mutation [44, 45]. Liu et al. [44] showed that the L24I mutation induces structural changes in PR1 that modify the contacts between the two Cter regions formed by residues 95–99, which explains the reduced stability in urea and the increased dissociation of the dimer. Second, our results suggest that some resistance mutations could directly modify PI binding. For example, we showed that mutations I46V and I84V could induce structural changes at pocket residues 31, 48, and 84, where the last two residues interact with PIs through hydrophobic contacts. Thus, these structural changes located in the PI-binding pocket induced by mutations could directly alter PI binding. This resistance mechanism has been previously observed in some PR1 mutants. For example, Tie et al. [6] showed that mutation I84V induces structural changes in PR1 resulting in the loss of two van der Waals contacts between residue 84 and DRV. The third putative resistance mechanism corresponds to induced structural changes outside of the binding pocket but

within residues important for the binding site conformation. Indeed, we showed that the I84V mutation could induce changes in asymmetry at positions 33, 83, and 89 that are involved in the hydrogen-bond network with pocket residues [32]. Thus, the I84V mutation could also alter the PI-binding pocket by modifying pocket properties through structural changes at positions 33, 83, and 89.

One possible explanation for the observed structural asymmetry that we did not consider in this study is the error or packing in the crystallographic structure. We previously showed that 24 positions were involved in crystal packing in the PR2 structure complexed with DRV (PDB code 3EBZ) [37]. In addition, we previously highlighted some structural asymmetry linked to crystal packing in unbound PR2, for example, at positions 3 and 18 [33]. This suggests that the detected conservation of the asymmetry of some positions results from crystal packing. However, in this study, structural asymmetry was extracted from structures (crystallographic structures and models) that were energetically minimized. This process enables the removal of crystal packing and contact with no biological relevance in the PR2 dimer. Indeed, we have previously shown that minimized structures exhibit fewer structural asymmetric positions than crystallographic structures [33]. Thus, using energetically minimized structures decreases the impact of crystal errors and packing on structural asymmetry. One solution to remove this source of asymmetry could be to extract structural asymmetry from structure models generated during molecular dynamics simulations. In addition, the analysis of asymmetry during molecular dynamics simulations could provide information about the link between structural asymmetry and PR2 deformations induced by these substitutions and facilitate a detailed understanding of the role of these substitutions. One limitation of our study is that we focused on the structural changes occurring in the backbone of PR2 without considering side-chain deformations. Even though our results suggest the existence of some structural deformations induced by mutations that could lead to PR2 resistance to PIs, we may have missed some mechanisms. For example, we observed that all mutants with the L90M mutation exhibited a change in asymmetry at position 8. As positions 8 and 90 were distant from each other in the PR2 structure, we concluded that this structural change was not induced by this mutation. However, one possibility is that the L90M mutation could have induced structural changes in some side chains that were not detected by our approach. These structural changes could spread within the structure to lead to the deformation of the backbone of pocket residue 8. Thus, to better understand the resistance of PR2 induced by some mutations, it would be interesting to consider the deformations

occurring in side chains in the future. Another limitation of our study is that we considered only one template during the mutant structure-modeling step: wild-type PR2 complexed with DRV (PDB code 3EBZ [46]). However, not all studied mutants are resistant to DRV. For example, Raugi et al. [19] showed that the L90M mutant is resistant to SQV and exhibits only a weak decrease in susceptibility to DRV. Thus, the use of a single template complexed with only one PI could explain why, for some mutants, we did not identify significant structural changes. It would be interesting to consider the complete resistance profile of each mutant. This will require the construction of mutant structures complexed with all PIs and knowledge of the resistance profile of each mutant. However, the resistance profile is particularly difficult to obtain for all mutants because there are few studies that have tested the effect of PIs against HIV-2 mutants using enzymatic or phenotypic susceptibility assays [9, 16, 17, 19, 23] and some of these studies have led to opposing results [1].

Conclusions

In this paper, we studied the structural impact of resistance mutations occurring in PR2, an important therapeutic target in HIV-2 infection treatment. More specifically, we explored the effect of resistance mutations on structural backbone asymmetry, a property involved in dimerization, ligand recognition and binding. To do so, we detected the differences in terms of structural asymmetry between the wild-type and 30 modeled mutant structures. Studying a large set of mutant structures at the same time allowed us to confirm the functional and structural roles of some PR2 residues, such as residues 28, 30, 53, 81, 82, and 87, that have been identified as important for the binding site conformation and ligand binding.

In addition, the comparison between the structural asymmetry of wild-type and mutant structures allowed the detection of some structural changes that could be induced by either PR2 flexibility or the studied resistance mutations. The analysis of the latter type of structural changes revealed three possible resistance mechanisms of PR2 that could occur together. First, we observed that resistance mutations could modify the PR2 interface and its stability, such as mutation L99F that induces structural changes at interface positions 91, 92, 93 and 98. The second mechanism involves the alteration of the PI interaction directly induced by mutations, such as the I46V and I84V mutations that induce structural changes at pocket residues. The third resistance mechanism corresponds to indirect modifications of PI binding. We noted that the I84V mutation could modify the PI-binding site by inducing structural changes at residues important for the maintenance of the conformation of the PI-pocket binding site.

In conclusion, this study was a structural analysis of the impact of a large number of PR2 resistance mutations using modeled mutant structures. Our results provide a better understanding of the structural effects of mutations of PR2 and, thus, of PR2 resistance to PIs.

Methods

Presentation of the wild-type PR2 structure

To model the PR2 mutant structures, we used the crystallographic structure of PR2 in complex with DRV (PDB code: 3EBZ [46]). We chose this structure because it is the only available structure of PR2 complexed with one of the three FDA-recommended drugs for the treatment of HIV-2 infection (DRV, LPV, and SQV). This structure shows a good resolution of 1.20 Å.

Extraction of structural and functional regions of PR2

We considered thirteen regions extracted from each PR2 monomer as in [32, 37, 38] (Fig. 1b and c). The binding pocket (20 residues per chain) was determined using the limits presented in [32] (Fig. 1c).

Location of flexible and rigid residues of PR2

The flexibility of PR2 residues was quantified using B-factor values found in the PDB file that measure the atomic displacement factor of residues [47]. From the 3EBZ structure, we extracted the B-factor values of each atom. For each residue, the average B-factor value of its atoms was then calculated. The higher the average B-factor value of a residue is, the more flexible the residue is.

Location of accessible and buried residues of PR2

The ASA value of protein residues is usually used to differentiate surface-exposed, important for protein interactions, and buried residues, important for protein stability [48]. According to [49], the ASA value is defined as the area of the surface swept out by the center of a probe sphere rolling over a molecule. To compute the ASA value of PR2 residues, NACCESS software [49] was run using 3EBZ structure and a radius probe sphere of 1.4 Å. The higher the ASA value of a residue is, the more accessible the residue is.

Modeling of mutant PR2 structures

We selected 30 drug-resistant mutants of PR2 containing one (14 single mutants), two (9 double mutants), or three mutations (7 triple mutants, Fig. 1a) from the literature. Figure 1b indicates the location of the mutations in the selected PR2-drug resistant mutants according to the wild-type structure (PDB code: 3EBZ). Six of them (46, 47, 48, 50, 82, and 84) were located in the binding pocket.

As the 3D structure of these mutants was not available in the PDB, we modeled their 3D structure using an *in silico* protocol based on the FoldX suite [39], as we described previously [10]. To do so, we applied the following protocol to the PR2 crystallographic structure in complex with DRV (PDB code: 3EBZ). First, we prepared the protein structure by removing the DRV ligand, metal atoms, and water molecules. Second, we applied the *RepairPDB* command of the FoldX suite to reduce the energy content of the structure. Third, we performed *in silico* mutagenesis using the *BuildModel* command of the FoldX suite. This command first introduces one or several mutations in the two chains of the wild-type structure using a side-chain rotamer library. Second, it performs an optimization of side chain of amino acids in the vicinity of the mutated residue(s). Each model was generated with five replications, and other options were set to the defaults. At the end of the mutagenesis step, five structures were modeled per mutant. This resulted in a set of 150 mutant structures for the 30 selected drug-resistant mutants. Fourth, we applied an energetic minimization protocol to the 150 modeled mutant structures using the protocol developed in [12]. The mono-protonated state was assigned to the oxygen atom OD2 of Asp25 in chain B using PROPKA software [50]. The system was solvated in a truncated octahedron box of explicit solvent (TIP3P water model) with a 12.0 Å buffer in each dimension. An appropriate number of chloride ions were added to produce a neutral charge in the system. Protein and water molecules were described using the force field AMBER ff99SB [51]. A two-step energy minimization was carried out in GROMACS [52] using a combination of steepest descent and conjugate gradient algorithms of roughly 1000 and 2000 iterations. Water molecules and counterions were relaxed through a first step energy minimization, using a position harmonic restraining force of 100 kcal.mol⁻¹ Å⁻² on the heavy atoms of the protein. A second step of energy minimization was performed by removing restraints on protein atoms. The particle mesh Ewald (PME) method was adopted to treat the long-range electrostatic interactions [53, 54]. The cutoff distances for the long-range electrostatic and van der Waals interactions were set to 10.0 Å.

We also applied the minimization protocol to the 3EBZ crystallographic structure with three replications, resulting in three minimized structures of wild-type PR2 referred to as the 3EBZ_{mini1}, 3EBZ_{mini2}, and 3EBZ_{mini3} structures.

Comparison of wild-type and mutant structures

We computed the RMSD values between the 150 mutant structures and the three 3EBZ_{mini} structures. The mutant structures were optionally superposed onto the

3EBZ_{mini} structures using PyMOL software [55]. The all-atom RMSD values, denoted as $RMSD_{aa}$, were computed between the superimposed structures using PyMOL software [55].

To analyze the impact of the number of mutations introduced in the mutants on the structural variability between the mutants and the wild-type structures, we compared the average $RMSD_{aa}$ values of the single, double, and triple mutants using a Kruskal-Wallis test and pairwise Wilcoxon tests. We also compared the variability in terms of the $RMSD_{aa}$ values of single, double, and triple mutants using a Bartlett test and pairwise Fisher tests.

Analysis of the structural asymmetry of PR2

Detection of structural asymmetry in PR2 structures.

The detection of structural asymmetry in a dimer corresponded to the identification of positions that exhibit different backbone conformations in the two chains. In this study, we extracted structural asymmetry from the three minimized wild-type structures of PR2 and the 150 modeled mutant structures using the structural alphabet-based approach that we previously developed [31]. This protocol, presented in Fig. 5, is based on the HMM-SA structural alphabet (Hidden Markov Model - Structural Alphabet), which is a library of 27 prototypes of 4-C α residues classified according to their geometry, assigned structural letters and labeled [a, A-Z] [56]. First, HMM-SA was used to simplify the 3D structure of the two chains of p residues in two sequences of $(p-3)$ structural letters overlapping by 3 residues, where each structural letter corresponded to the geometry of a 4-C α fragment. Each structural letter was assigned to the third residue of the 4-residue fragment. As the fragments overlapped by 3 residues, no structural letter was assigned to the first, second, and last residues. Second, the structural letters of each position in both chains were compared. A position has two possible asymmetry profiles: either the position is asymmetric (i.e., exhibiting different local conformations) (=structural letters) in the two chains, or it is symmetric (i.e., exhibiting the same local conformation) (=structural letter) in the two chains.

Quantification of structural asymmetry

We first counted the number of asymmetric positions in the three minimized wild-type and 150 mutant structures, as described in [31]. To quantify the structural asymmetry in the mutant structure set, we determined the occurrence of asymmetry (AO) at each position, i (i.e., the number of mutant structures exhibiting asymmetry at position i). The statistical significance of the AO value of position i was determined using the overrepresentation p -value, p_{AO} . The p_{AO} of

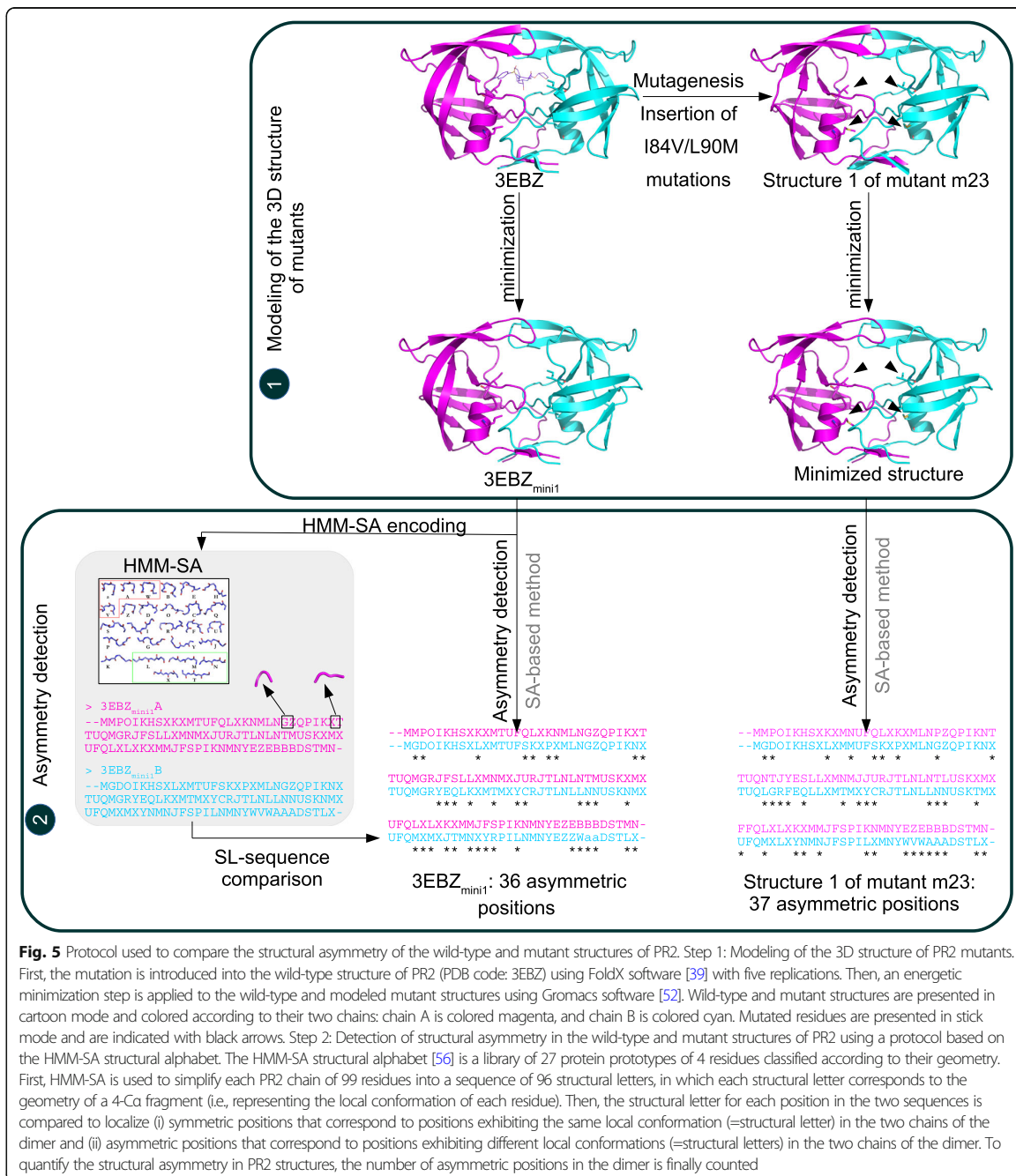


Fig. 5 Protocol used to compare the structural asymmetry of the wild-type and mutant structures of PR2. Step 1: Modeling of the 3D structure of PR2 mutants. First, the mutation is introduced into the wild-type structure of PR2 (PDB code: 3EBZ) using FoldX software [39] with five replications. Then, an energetic minimization step is applied to the wild-type and modeled mutant structures using Gromacs software [52]. Wild-type and mutant structures are presented in cartoon mode and colored according to their two chains: chain A is colored magenta, and chain B is colored cyan. Mutated residues are presented in stick mode and are indicated with black arrows. Step 2: Detection of structural asymmetry in the wild-type and mutant structures of PR2 using a protocol based on the HMM-SA structural alphabet. The HMM-SA structural alphabet [56] is a library of 27 protein prototypes of 4 residues classified according to their geometry. First, HMM-SA is used to simplify each PR2 chain of 99 residues into a sequence of 96 structural letters, in which each structural letter corresponds to the geometry of a 4-Ca fragment (i.e., representing the local conformation of each residue). Then, the structural letter for each position in the two sequences is compared to localize (i) symmetric positions that correspond to positions exhibiting the same local conformation (=structural letter) in the two chains of the dimer and (ii) asymmetric positions that correspond to positions exhibiting different local conformations (=structural letters) in the two chains of the dimer. To quantify the structural asymmetry in PR2 structures, the number of asymmetric positions in the dimer is finally counted

position i was computed by comparing its observed ($AO(i)$) and expected ($AO^{random}(i)$) occurrences. The AO^{random} occurrence of position i was computed in a random set generated with 150 random binary sequences of 96 positions, where each random sequence

was associated with a mutant structure. In each random binary sequence, the 0 and 1 values represent the symmetric and asymmetric positions, respectively. The number of asymmetric positions in the random sequence, j , corresponds to the number of asymmetric

positions in the j^{th} mutant structure. The AO^{random} value of position i corresponds to the number of random sequences in which that position is considered an asymmetric position (i.e., showing a value of 1 at the position). p_{AO} is estimated as the probability that AO^{random} is higher than $AO(i)$ using Eq. 1 and 2000 random sets.

$$p_{AO}(i) = \frac{\text{prob}[AO^{\text{random}}(i) > AO(i)]}{n_{\text{simu}}} \quad (\text{Eq.1})$$

where $n_{\{AO^{\text{random}}(i) > AO(i)\}}$ is the number of simulations in which $AO^{\text{random}}(i)$ is higher than $AO(i)$, and n_{simu} is the total number of simulations.

An asymmetric position was considered statistically overrepresented if its p_{AO} was smaller than a threshold of 0.0005, as determined using the Bonferroni adjustment to consider multiple tests (0.05/96 positions).

Study of the link between mutations and changes in asymmetry

Location of changes in asymmetry in mutant structures relative to wild-type structure

A change in asymmetry corresponds to a position that exhibits a difference in asymmetry status (asymmetric or symmetric) in both a given mutant structure and the three 3EBZ_{mini} structures. Two types of changes in asymmetry were defined: asymmetry (or loss of asymmetry) occurs at a given position of a mutant structure if the position is asymmetric (or symmetric) in the mutant structure, whereas it is symmetric (or asymmetric) in the 3EBZ_{mini} structures. To locate the positions exhibiting changes in asymmetry in each mutant structure, we compared the asymmetry status in the mutant and 3EBZ_{mini} structures. To do so, we focused on only the 78 positions exhibiting the same status (asymmetric or symmetric) in the three 3EBZ_{mini} structures.

Quantification of the changes in asymmetry per mutant

For a given mutant, we computed the number of its five structures exhibiting changes in asymmetry for the 78 selected positions. This number was ranked from 0 to 5. A value of 0 for a position means that the position exhibits the same asymmetry status in the three 3EBZ_{mini} structures and the five structures of the mutant. In contrast, a position with a value of 5 means that the position exhibits a different asymmetry status in the three 3EBZ_{mini} structures and the five structures of the mutant. For all mutants, these data were summarized in a network that connects mutants with a position when at least one structure of the mutant exhibits a change in asymmetry relative to the wild-type structure at the position. The

network was drawn using the *igraph* library of R software [57].

Supplementary information

Supplementary information accompanies this paper at <https://doi.org/10.1186/s12860-020-00290-1>.

Additional file 1. Structural comparison between mutant structures and the three 3EBZ_{mini} structures.

Additional file 2. Summary of structural asymmetry in the three wild-type structures 3EBZ_{mini} and the 150 mutant structures.

Additional file 3. Flexibility of PR2 residues quantified using B-factor values.

Additional file 4. Exposure of PR2 residues quantified using ASA values.

Additional file 5. Quantification of structural asymmetry in the 150 structures of the PR2 mutants.

Additional file 6. Interaction between PR2 and three drugs (APV, DRV, and IDV).

Additional file 7. Residues involved in the interface of the wild-type PR2 dimer.

Abbreviations

3D: Tridimensional; AO: Asymmetry occurrence; APV: Amprenavir; ASA: Accessible surface area; Cter: C-terminal end; DRV: Darunavir; FDA: Food and Drug Administration; HIV: Human immunodeficiency virus; HIV-1: Human immunodeficiency virus type 1; HIV-2: Human immunodeficiency virus type 2; HMM-SA: Hidden Markov Model - Structural Alphabet; IDV: Indinavir; LPV: Lopinavir; NFV: Nelfinavir; Nter: N-terminal end; OR_{asym}: Over-represented asymmetric; PDB: Protein Data Bank; PI: Protease inhibitor; PME: Particle Mesh Ewald; PR1: HIV-1 protease; PR2: HIV-2 protease; RMSD: Root mean square deviation; RMSDaa: RMSD computed using all common atoms; SQV: Saquinavir

Acknowledgements

Not applicable.

Authors' contributions

LR conceived and designed the work. MP and LR supervised the project. NC and LR collected data. SF and DF generated the 150 mutant models. PL and LR analyzed and interpreted the data. LR wrote the manuscript. All authors reviewed the manuscript. All authors approved the version of the manuscript to be published.

Funding

This work was supported by an ANRS Grant. NC and PL are supported by a fellowship from the Ministère de l'Éducation Nationale de la Recherche et de Technologie (MENRT).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 20 December 2019 Accepted: 16 June 2020

Published online: 23 June 2020

References

- Menéndez-Arias L, Alvarez M. Antiretroviral therapy and drug resistance in human immunodeficiency virus type 2 infection. *Antivir Res.* 2014;102:70–86. <https://doi.org/10.1016/j.antiviral.2013.12.001>.
- Brower ET, Bacha UM, 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. <https://doi.org/10.1111/j.1747-0285.2008.00647.x>.
- Gustchina A, Weber IT. Comparative analysis of the sequences and structures of hiv-1 and hiv-2 proteases. *Proteins: Structure, Function, and Bioinformatics.* 1991;10:325–39. <https://doi.org/10.1002/prot.340100406>.
- Sardana VV, Schlabach AJ, Graham P, Bush BL, Condra JH, Culbertson JC, Gottlieb L, Graham DJ, Kohl NE, et al. Human immunodeficiency virus type 1 protease inhibitors: evaluation of resistance engendered by amino acid substitutions in the enzyme's substrate binding site. *Biochemistry.* 1994;33:2004–10. <https://doi.org/10.1021/bi00174a005>.
- Hoog SS, Towler EM, Zhao B, Doyle ML, Debouck C, Abdel-Meguid SS. Human immunodeficiency virus protease ligand specificity conferred by residues outside of the active site cavity. *Biochemistry.* 1996;35:10279–86. <https://doi.org/10.1021/bi960179j>.
- Tie Y, Boross PI, Wang YF, Gaddis L, Hussain AK, Leshchenko S, Ghosh AK, Louis JM, Harrison RW, Weber IT. High resolution crystal structures of hiv-1 protease with a potent non-peptide inhibitor (uic-94017) active against multi-drug-resistant clinical strains. *J Mol Biol.* 2004;23:341–52. <https://doi.org/10.1016/j.jmb.2004.02.052>.
- Masse S, Lu X, Dekhtyar T, Lu L, Koev G, Gao F, Mo H, Kempf D, Bernstein B, Hanna GJ, Molla A. In vitro selection and characterization of human immunodeficiency virus type 2 with decreased susceptibility to lopinavir. *Antimicrob Agents Chemother.* 2007;51:3075–80. <https://doi.org/10.1128/AAC.00146-07>.
- Kovalevsky AY, Louis JM, Aniana A, Ghosh AK, Weber IT. Structural evidence for effectiveness of darunavir and two related antiviral inhibitors against hiv-2 protease. *J Mol Biol.* 2008;384:178–92. <https://doi.org/10.1016/j.jmb.2008.09.031>.
- Raugi DN, Smith RA, Gottlieb GS, The University of Washington-Dakar HIV-2 Study Group. Four amino acid changes in hiv-2 protease confer class-wide sensitivity to protease inhibitors. *J Virol.* 2016;90:1062–9. <https://doi.org/10.1128/JVI.01772-15>.
- Triki D, Billot T, Flatters D, Visseaux B, Descamps D, Camproux AC, Regad L. Exploration of the effect of sequence variations located inside the binding pocket of hiv-1 and hiv-2 proteases. *Sci Rep.* 2018;8:5789. <https://doi.org/10.1038/s41598-018-24124-5>.
- Kar P, Knecht VJ. Origin of decrease in potency of darunavir and two related antiviral inhibitors against hiv-2 compared to hiv-1 protease. *Phys Chem B.* 2012;116:2605–14. <https://doi.org/10.1021/JP211768N>.
- 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. <https://doi.org/10.1038/SREP06872>.
- Colson P, Henry M, Tourres C, Lozachmeur D, Gallais H, Gastaut JA, Moreau J, Tamalet C. Polymorphism and drug-selected mutations in the protease gene of human immunodeficiency virus type 2 from patients living in southern France. *J Clin Microbiol.* 2004;42:570–7. <https://doi.org/10.1128/jcm.42.2.570-577.2004>.
- Damond F, Brun-Vézinet F, Matheron S, Peytavin G, Campa P, Pueyo S, Mammano F, Lasteré S, Farfara I, Simon F, Chêne G, Descamps D. 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. *J Clin Microbiol.* 2005;43:484–7. <https://doi.org/10.1128/JCM.43.1.484-487.2005>.
- Rodés B, Sheldon J, Toro C, Jiménez V, Alvarez MA, Soriano V. Susceptibility to protease inhibitors in hiv-2 primary isolates from patients failing antiretroviral therapy. *Antimicrob Agents Chemother.* 2006;50:709–13. <https://doi.org/10.1093/jac/dkl034>.
- Ntemgwia M, Brenner BG, Oliveira M, Moisi D, Wainberg MA. Variations in reverse transcriptase and mase h domain mutations in human immunodeficiency virus type 1 clinical isolates are associated with divergent phenotypic resistance to zidovudine. *Antimicrob Agents Chemother.* 2007;51:604–10. <https://doi.org/10.1128/AAC.00646-07>.
- Desbois D, Roquebert B, Peytavin G, Damond F, Collin G, Bénard A, Campa P, Matheron S, Chêne G, Brun-Vézinet F, Descamps, French ANRS HIV-2 cohort (ANRS CO 05 VIH-2). In vitro phenotypic susceptibility of human immunodeficiency virus type 2 clinical isolates to protease inhibitors. *Antimicrob Agents Chemother.* 2008;52:1545–8. <https://doi.org/10.1128/AAC.01284-07>.
- Jallow S, Alabi A, Sarge-Njie R, Peterson K, Whittle H, Corrah T, Jaye A, Cotten M, Vanham G, McConkey SJ, Rowland-Jones S, Janssens W. Virological response to highly active antiretroviral therapy in patients infected with human immunodeficiency virus type 2 (hiv-2) and in patients dually infected with hiv-1 and hiv-2 in the Gambia and emergence of drug-resistant variants. *J Clin Microbiol.* 2009;47:2200–8. <https://doi.org/10.1128/JCM.01654-08>.
- Raugi DN, Smith RA, Ba S, Toure M, Traore F, Sall F, Pan C, Blankenship L, Montano A, Olson J, Dia Badiane NM, Mullins JJ, Kiviati NB, Hawes SE, Sow PS, Gottlieb GS, University of Washington-Dakar HIV-2 Study Group. Complex patterns of protease inhibitor resistance among antiretroviral treatment-experienced hiv-2 patients from senegal: Implications for second-line therapy. *Antimicrob Agents Chemother.* 2013;57:2751–60. <https://doi.org/10.1128/AAC.00405-13>.
- Larrouy L, Vivot A, Charpentier C, Bénard A, Visseaux B, Damond F, Matheron S, Chêne G, Brun-Vézinet F, Descamps D, ANRS CO5 HIV-2 Cohort. Impact of gag genetic determinants on virological outcome to boosted lopinavir-containing regimen in hiv-2-infected patients. *AIDS.* 2013;27:69–80. <https://doi.org/10.1097/QAD.0b013e32835a10d8>.
- Cavaco-Silva J, Aleixo MJ, Van Laethem K, Faria D, Valadas E, de Fátima Gonçalves M, Gomes P, Vandamme AM, Cunha C, Camacho RJ. Portuguese HIV-2 Resistance Study Group Mutations selected in hiv-2-infected patients failing a regimen including atazanavir. *Antimicrob Agents Chemother.* 2013; 68:190–2. <https://doi.org/10.1093/jac/dks363>.
- Storto A, Visseaux B, Bertine M, Le Hingrat Q, Collin G, Damond F, Khuong MA, Blum L, Tubiana R, Karmochkine M, Cazanave C, Matheron S, Descamps D, Charpentier C. ANRS HIV-2 CO5 cohort: minority resistant variants are also present in hiv-2-infected antiretroviral-naïve patients. *J Antimicrob Chemother.* 2018;73:1173–6. <https://doi.org/10.1093/jac/dkx530>.
- Mahdi M, Szojka Z, Motyan JA, Tózsér J. Inhibition profiling of retroviral protease inhibitors using an hiv-2 modular system. *Viruses.* 2015;7:16152–62. <https://doi.org/10.3390/v7122931>.
- Bénard A, Damond F, Campa P, Peytavin G, Descamps D, Lascoux-Combes C, Taieb A, Simon F, Autran B, Brun-Vézinet F, Chêne G, Matheron S. Good response to lopinavir/ritonavir-containing antiretroviral regimens in antiretroviral-naïve hiv-2-infected patients. *AIDS.* 2009;23:1171–83. <https://doi.org/10.1097/QAD.0b013e32832949f0>.
- Charpentier C, Visseaux B, Bénard A, Peytavin G, Damond F, Roy C, Taieb A, Chêne G, Matheron S, Brun-Vézinet F, Descamps D. ANRS CO5 HIV-2 cohort: transmitted drug resistance in french hiv-2-infected patients. *AIDS.* 2013;27: 1671–84. <https://doi.org/10.1097/QAD.0b013e32836207f3>.
- Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H. I.N., S., Bourne, P.E. the protein data bank. *Nucleic Acids Res.* 2000;28:235–42. <https://doi.org/10.1093/nar/28.1.235>.
- Tong L, Pav S, Pargellis C, Do⁺ F, Lamarre D, Anderson PC. 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 U S A.* 1993;90:8387–91. <https://doi.org/10.1073/pnas.90.18.8387>.
- Mulichak AM, Hui JO, Tomasselli AG, Heinrikson RL, Curry KA, Tomich CS, Thaisrivongs S, Sawyer TK, Watenpaugh KD. The crystallographic structure of the protease from human immunodeficiency virus type 2 with two synthetic peptidic transition state analog inhibitors. *J Biol Chem.* 1993;268:13103–9.
- Tong L, Pav S, Mui S, Lamarre D, Yoakim C, Beaulieu P, Anderson PC. Crystal structures of hiv-2 protease in complex with inhibitors containing the hydroxyethylamine dipeptide isostere. *Structure.* 1995;3:33–40. [https://doi.org/10.1016/S0969-2126\(01\)00133-2](https://doi.org/10.1016/S0969-2126(01)00133-2).
- Priestle JP, Fassler A, Rösel J, Tintelnot-Blomley M, Strop P, Grütter MG. 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;3:381–9. [https://doi.org/10.1016/S0969-2126\(01\)00169-1](https://doi.org/10.1016/S0969-2126(01)00169-1).
- Triki D, Cano Contreras ME, Flatters D, Visseaux B, Descamps D, Camproux AC, 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. <https://doi.org/10.1038/s41598-017-18941-3>.

32. Triki D, Fartek S, Visseaux B, Descamps D, Camproux AC, Regad L. Characterizing the structural variability of hiv-2 protease upon the binding of diverse ligands using a structural alphabet approach. *J Biomol Struct Dyn*. 2018;28:1–13. <https://doi.org/10.1080/07391102.2018.1562985>.
33. Ollitrault G, Fartek S, Descamps D, Camproux AC, Visseaux B, Regad L. Characterization of hiv-2 protease structure by studying its asymmetry at the different levels of protein description. *Symmetry*. 2018;10:644. <https://doi.org/10.3390/sym10110644>.
34. Prabu-jeyabalan M, Nalivaika EA, King NM, Schiffer CA. Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: structural insights for better antiviral therapy. *Society*. 2003;77:1306–15. <https://doi.org/10.1128/JVI.77.2.1306-1315.2003>.
35. Goodsell DS, Olson AJ. Structural symmetry and protein function. *Annu Rev Biophys Biomol Struct*. 2000;29:105. <https://doi.org/10.1146/annurev.biophys.29.1.105>.
36. Brown JH. Breaking symmetry in protein dimers: designs and functions. *Protein Sci*. 2006;15:1–13. <https://doi.org/10.1110/ps.051658406>.
37. Triki D, Kermarrec M, Visseaux B, Descamps D, Flatters D, Camproux AC, Regad L. Exploration of the effects of sequence variations between hiv-1 and hiv-2 proteases on their three-dimensional structures. *J Biomol Struct Dyn*. 2019 vol. 37. <https://doi.org/10.1080/07391102.2019.1704877>.
38. Sadiq SK, de Fabritiis G. Explicit solvent dynamics and energetics of hiv-1 protease flap opening and closing. *Proteins Struct Funct Bioinforma*. 2010; 78:2873–85.
39. Guerois R, Nielsen JE, Serrano L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol*. 2002;320:369–87. [https://doi.org/10.1016/S0022-2836\(02\)00442-4](https://doi.org/10.1016/S0022-2836(02)00442-4).
40. Zoete V, Michielin O, Karplus M. Relation between sequence and structure of hiv-1 protease inhibitor complexes: a model system for the analysis of protein flexibility. *J Mol Biol*. 2002;315:21–52. <https://doi.org/10.1006/jmbi.2001.5173>.
41. van Westen GJ, Wegner JK, Bender A, Ijzerman AP, van Vlijmen HW. Mining protein dynamics from sets of crystal structures using "consensus structures". *Protein Sci*. 2010;19:742–52. <https://doi.org/10.1002/pro.350>.
42. Regad L, Chéron JB, Triki D, Senac C, Flatters D, Camproux AC. Exploring the potential of a structural alphabet-based tool for mining multiple target conformations and target flexibility insight. *PLoS One*. 2017;12:0182972. <https://doi.org/10.1371/journal.pone.0182972>.
43. Weber IT, Agniswamy J. Hiv-1 protease: structural perspectives on drug resistance. *Viruses*. 2009;1:110–36. <https://doi.org/10.3390/v1031110>.
44. Liu F, Boross PI, Wang YF, Tözsér J, Louis JM, Harrison RW, Weber IT. Kinetic, stability, and structural changes in high-resolution crystal structures of hiv-1 protease with drug-resistant mutations I24i, i50v, and g73s. *J Mol Biol*. 2005; 354:789–800. <https://doi.org/10.1016/j.jmb.2005.09.095>.
45. Liu F, Kovalevsky AY, Tie Y, Ghosh AK, Harrison RW, Weber IT. Effect of flap mutations on structure of hiv-1 protease and inhibition by saquinavir and darunavir. *J Mol Biol*. 2008;381:102–15. <https://doi.org/10.1016/j.jmb.2008.05.062>.
46. Tie Y, Wang YF, Boross PI, Chiu TY, Ghosh AK, Tözsér J, Louis JM, Harrison RW, Weber IT. Critical differences in hiv-1 and hiv-2 protease specificity for clinical inhibitors. *Protein Sci*. 2012;21:339–50. <https://doi.org/10.1002/pro.2019>.
47. Drenth J. Principles of protein X-ray crystallography, pp. 189–198. New York: Springer Verlag; 1994.
48. Chan HS, Dill KA. Origins of structure in globular proteins. *Proc Natl Acad Sci U S A*. 1990;87:6388–92. <https://doi.org/10.1073/pnas.87.16.6388>.
49. Hubbard SJ, Thornton JM. Naccess. Computer Program, Department of Biochemistry and Molecular Biology University College London; 1993.
50. Li H, Robertson AD, Jensen JH. Very fast empirical prediction and rationalization of protein pK values. *Proteins*. 2005;61:704–21. <https://doi.org/10.1002/prot.20660>.
51. 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;65:712–25. <https://doi.org/10.1002/prot.21123>.
52. Abraham MJ, Murtolad T, Schulz R, Palla S, Smith JC, Hess B, Lindahl E. Gromacs: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*. 2015;1:2:19–25. <https://doi.org/10.1016/j.softx.2015.06.001>.
53. Darden T, York D, Pedersen L. Particle mesh ewald: an n log (n) method for ewald sums in large systems. *J Chem Phys*. 1993;98:10089–92. <https://doi.org/10.1063/1.464397>.
54. Essmann U, Perera L, Berkowitz M. A smooth particle mesh ewald method. *J Chem Phys*. 1995;103:8577–93. <https://doi.org/10.1063/1.470117>.
55. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. <https://pymol.org/2/support.html>.
56. Camproux AC, Gautier R, Tufféry P. A hidden Markov model derived structural alphabet for proteins. *J Mol Biol*. 2004;339:591–605. <https://doi.org/10.1016/j.jmb.2004.04.005>.
57. Csárdi G, Nepusz T. The igraph software package for complex network research. *InterJ Complex Syst*. 2006;1695:1–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions



3 Impact des mutations sur la structure de la PR2

Dans nos travaux précédents, nous avons analysé l’impact de 30 mutants de résistance de PR2 basé sur l’asymétrie de PR2, une propriété importante pour la fixation du ligand et la déformation de la cible. L’exploration de l’asymétrie dans les structures mutantes nous a permis d’identifier des déformations structurales du squelette peptidique induites par les mutations. La localisation de ces déformations nous a permis de suggérer trois mécanismes de résistance :

- Les mutations qui induisent des changements structuraux dans la poche de liaison qui pourraient altérer directement la fixation des IP.
- Les mutations qui ont un impact sur les propriétés et la conformation de la poche en engendrant des modifications structurales sur des résidus non impliqués dans celle-ci mais avec des résidus en contact avec la poche.
- Les mutations qui modifient l’interface de PR2 et sa stabilité par le biais de changement structuraux au niveau des résidus de l’interface.

Cependant, cette étude ne se base que sur l’étude des changements structuraux localisés au niveau du squelette peptidique de la PR2. De plus, l’étude précédente n’apporte aucune quantification de la déformation induite par ces mutations. C’est donc dans l’objectif de pallier ces deux limites que nous avons mené l’étude présentée dans ce chapitre. Cette étude vise à localiser et quantifier les déformations induites par ces mutations sur la PR2 en considérant cette fois-ci tous ses atomes. Pour cela, nous avons identifié dans un premier temps les déplacements atomiques induits par les mutations en comparant la position des atomes du squelette et des chaînes latérales dans la structure sauvage et dans les structures modélisées de 31 mutants. Ensuite, nous avons étudié plus en détails l’impact de ces changements structuraux induits par les mutations sur le site de liaison des inhibiteurs et sur la région d’interface des deux monomères de la PR2.

3.1 Choix des mutations

Pour cette nouvelle étude, nous avons affiné la liste des mutants de notre précédente étude (Laville *et al.*, 2020). Une sélection de 31 mutants résistants à au moins un médicament a été constituée dont 19 mutants communs à nos travaux précédents (Laville *et al.*, 2020). Ce nouveau jeu contient plus de mutants simples (20) et moins de mutants doubles (8) et triples (3). Sur l’ensemble des 31 mutants, certaines mutations sont présentes chez plusieurs mutants; nous comptons au total 22 mutations différentes (figure 3.1). Parmi ces 22 mutations, la moitié ne fait pas partie des 14 étudiées lors de nos travaux précédents. La figure 3.1 présente la localisation de ces 22 mutations sur la structure 3D de PR2 et montre que ces mutations échantillonnent l’ensemble de la structure de PR2.

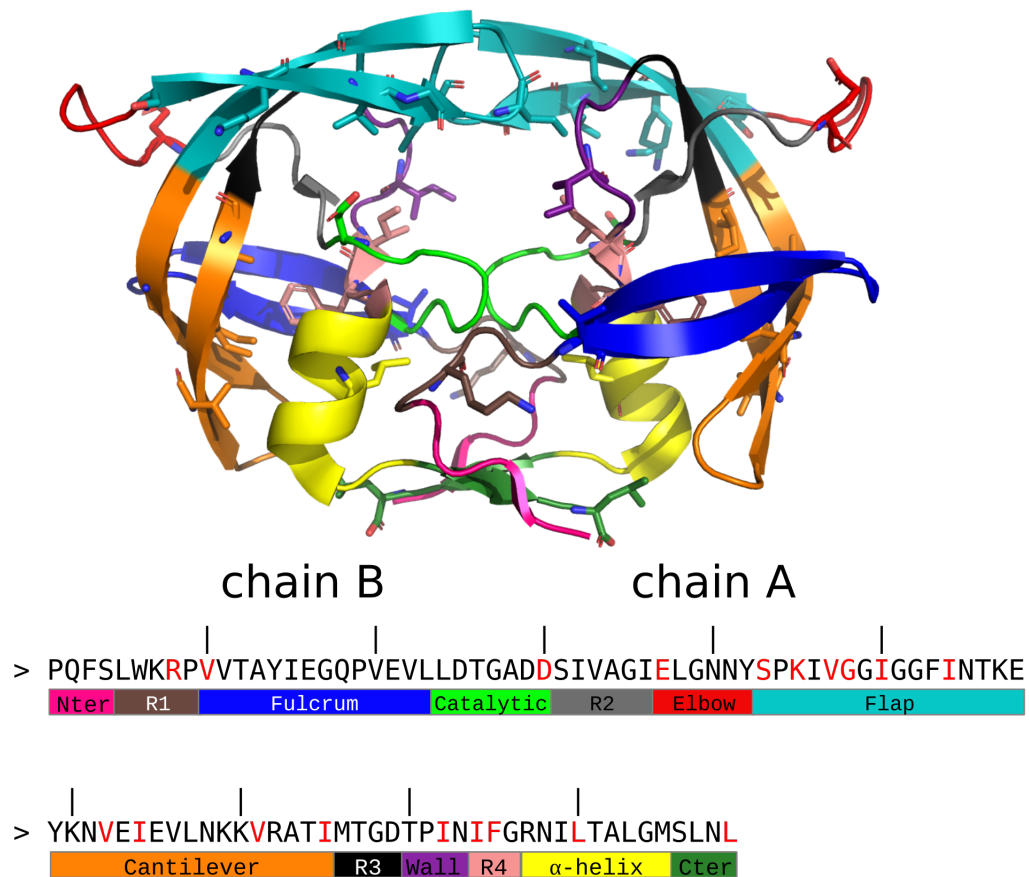


Figure 3.1 : Localisation des mutations sur la PR2. Celles-ci sont en *sticks* sur la structure 3D et en lettres rouges dans la séquence. Leur position indique leur appartenance à l'une des 13 régions qui colorent la structure 3D. L'image est adaptée de [Laville et al., 2021](#).

3.2 Modélisation des structures mutantes

Comme pour notre première étude, aucune structure mutante n'était disponible dans la PDB. Nous avons donc appliqué le protocole de modélisation développé précédemment aux 31 mutants. À partir de la structure de PR2 sauvage (code PDB : 3EBZ) nettoyée, les structures 3D des mutants sont modélisées à l'aide du logiciel FoldX ([Guerois et al., 2002](#)), puis une étape de minimisation énergétique a été appliquée. L'étape de minimisation a aussi été appliquée à la structure sauvage nettoyée pour prendre en considération les changements structuraux induits par la minimisation. Au final, un jeu de 155 structures mutantes est obtenu (cinq structures ont été modélisées par mutant pour tenir compte des différents rotamères disponibles pour chaque acide aminé) et une structure sauvage minimisée, d'où $31 \times 5 + 1 = 156$ structures au total.

3.3 Localisation et quantification des changements structuraux induits par les mutations

Dans la première partie de ce travail, nous avons cherché à localiser et quantifier les changements structuraux induits par les mutations en prenant en compte à la fois les atomes du squelette et des chaînes latérales. Pour cela, les réarrangements structuraux sont identifiés en comparant la position des atomes entre les structures mutantes et la structure sauvage. Pour identifier ces changements, nous avons utilisé le protocole en deux étapes développé dans l'étude de [Perrier *et al.*, 2019](#) (figure 3.2) :

1. Dans un premier temps, la structure mutante est superposée à la structure sauvage en considérant tous les atomes.
2. Pour chaque atome (hors hydrogènes) commun aux deux structures, la distance euclidienne est calculée entre ses positions dans les deux structures basées sur ses coordonnées 3D. Cette distance, notée $dist_{WT-mutant}$, quantifie le déplacement d'un atome dans la structure mutante par rapport à la structure sauvage.

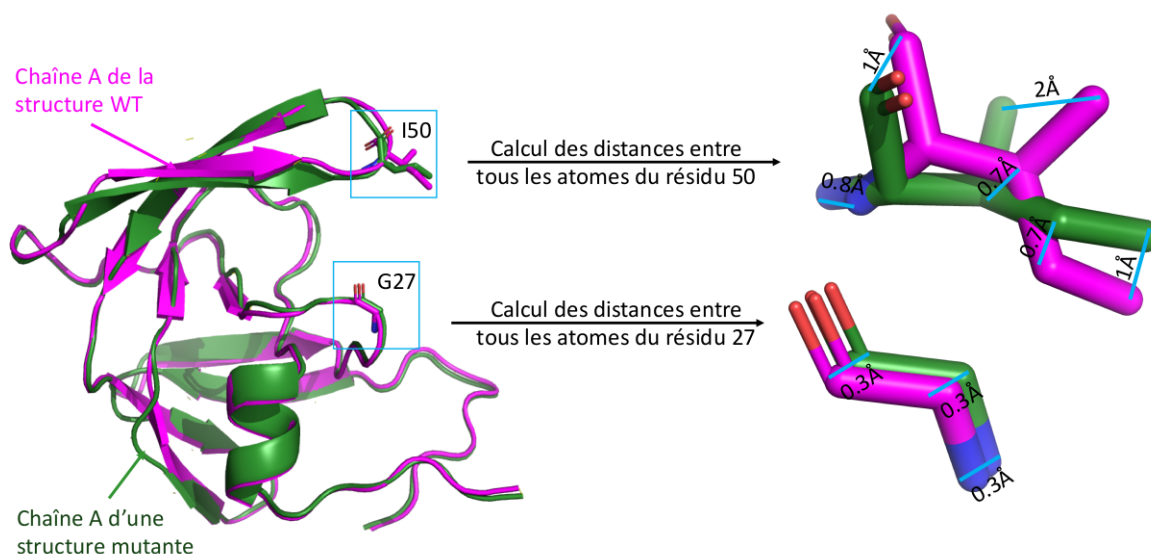


Figure 3.2 : Quantification des déplacements des atomes des résidus 27 et 50 entre la structure sauvage et une structure mutante. Les chaînes A des structures sauvage (en magenta) et mutantes (en vert) ont été superposées. Elles sont représentées en mode *cartoon*. Pour chaque atome de ces deux résidus, la distance euclidienne est calculée entre les positions des atomes dans les structures sauvage et mutantes.

En utilisant ce protocole, nous avons quantifié les réarrangements structuraux de tous les atomes dans les 155 structures mutantes. Dans le jeu des 155 mutants, les distances $dist_{WT-mutant}$ varient de 0 à 10,74 Å. La plupart (95 %) des atomes ont une distance $dist_{WT-mutant}$ inférieure à 0,14 Å, signifiant donc que la majorité des atomes sont peu déplacés entre les structures sauvages et mutantes. La figure 3.3 présente les atomes avec une distance $dist_{WT-mutant}$ supérieure à 0,3 Å dans les chaînes A et

B dans les 155 structures mutantes. Afin de distinguer les déplacements significatifs entre structures sauvage et mutantes, nous avons sélectionné les 2136 atomes avec une distance $dist_{WT-mutant}$ supérieure à 0,3 Å dans au moins trois structures d'un mutant donné. Parmi eux, 17 % présentent des déplacements importants ($dist_{WT-mutant} > 2$ Å). Nous avons appelé ces atomes « MCS » (*Mutant-Conserved Shifted Atoms*). Comme attendu, la majorité d'entre eux correspondent à des atomes des chaînes latérales. Ce sont aussi les déplacements de plus grande amplitude, c'est-à-dire présentant une plus grande distance $dist_{WT-mutant}$. Ces 2136 réarrangements structuraux identifiés touchent à la fois les résidus mutés (nommés déplacements directs) et les résidus non mutés (nommés déplacements indirects). La comparaison des amplitudes de déplacements directs et indirects montre que les déplacements directs sont de plus grande amplitude. Ces résultats suggèrent que les réarrangements structuraux identifiés entre structures mutantes et sauvages peuvent être dus soit à la flexibilité intrinsèque de la protéine, soit directement ou indirectement par la mutation.

Afin de distinguer les réarrangements induits par la flexibilité intrinsèque et ceux dus aux mutations, nous avons comparé les déplacements identifiés dans les structures mutantes avant et après l'étape de minimisation. Nous avons montré que 30 % des déplacements sont communs aux structures mutantes minimisées et non minimisées, montrant que ces déplacements sont induits par la mutation. Ils sont situés uniquement sur les chaînes latérales et aux alentours des résidus mutés, ce qui s'explique par le fait que le logiciel FoldX (logiciel utilisé pour la mutagenèse) n'optimise que les chaînes latérales proches de la zone de mutagenèse. Les déplacements non communs résultent soit de leur flexibilité intrinsèque soit des impacts des mutations.

Une fois les réarrangements structuraux potentiellement induits par les mutations détectés, nous avons étudié leur localisation sur la structure de PR2. La répartition des 2136 atomes présentant des réarrangements dans les 13 régions de la PR2 est hétérogène (figure 3.4). Néanmoins, une sur-représentation est observée au niveau de deux régions (flaps et cantilever) dans les deux chaînes. De plus, nous avons observé que les réarrangements ne sont pas répartis de manière identique sur les deux chaînes. Par exemple, nous avons détecté très peu de réarrangements structuraux dans les régions fulcrum et elbow de la chaîne A contrairement à la chaîne B. Il y a donc une asymétrie dans l'impact des mutations entre les deux chaînes malgré le fait que les deux chaînes possèdent les mêmes mutations : les mêmes mutations n'engendrent pas les mêmes effets d'une chaîne à l'autre.

Nous nous sommes ensuite demandés si le nombre de mutations par mutant avait une influence sur le nombre et l'amplitude des réarrangements structuraux. Nous n'avons détecté aucune corrélation entre nombre de mutations par mutant et nombre de réarrangements structuraux. Par exemple, le simple mutant K7R est celui qui contient le plus de réarrangements structuraux (> 140). De plus, le nombre de déplacements n'est pas corrélé non plus à leur amplitude. Sept mutants simples se sont révélés contenir des

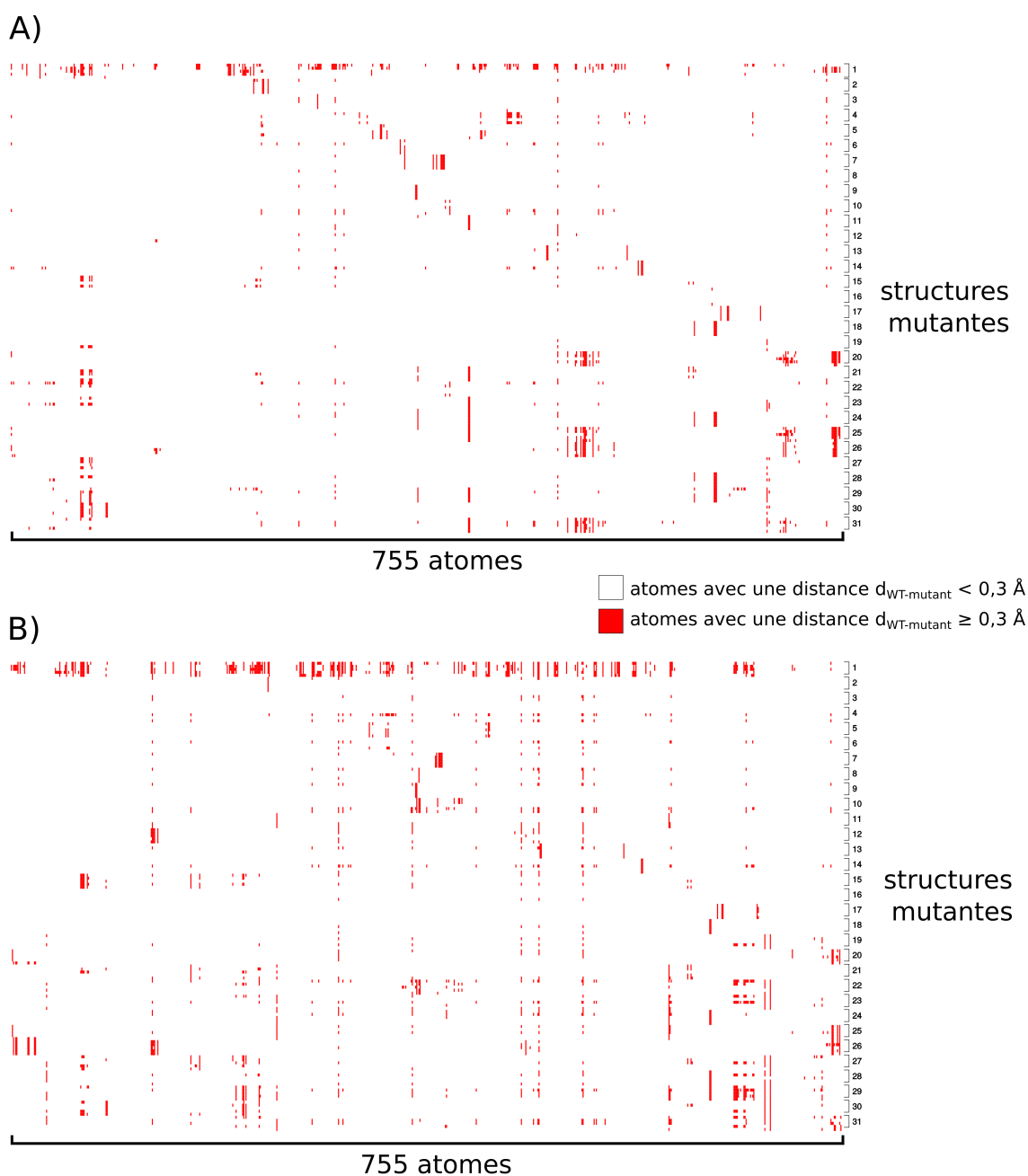


Figure 3.3 : Localisation des réarrangements structuraux dans les 155 structures mutantes. A) chaîne A, B) chaîne B. Les atomes sont représentés en colonnes et les 155 structures sont représentées en lignes et rangées suivant les mutants. En rouge sont indiqués les atomes ayant une distance $dist_{WT-mutant}$ supérieure ou égale à $0,3 \text{ \AA}$ et, en blanc, inférieure à $0,3 \text{ \AA}$.

déplacements de grande amplitude ($dist_{WT-mutant} \geq 2 \text{ \AA}$), ce qui montre qu'une seule mutation peut induire des déformations importantes.

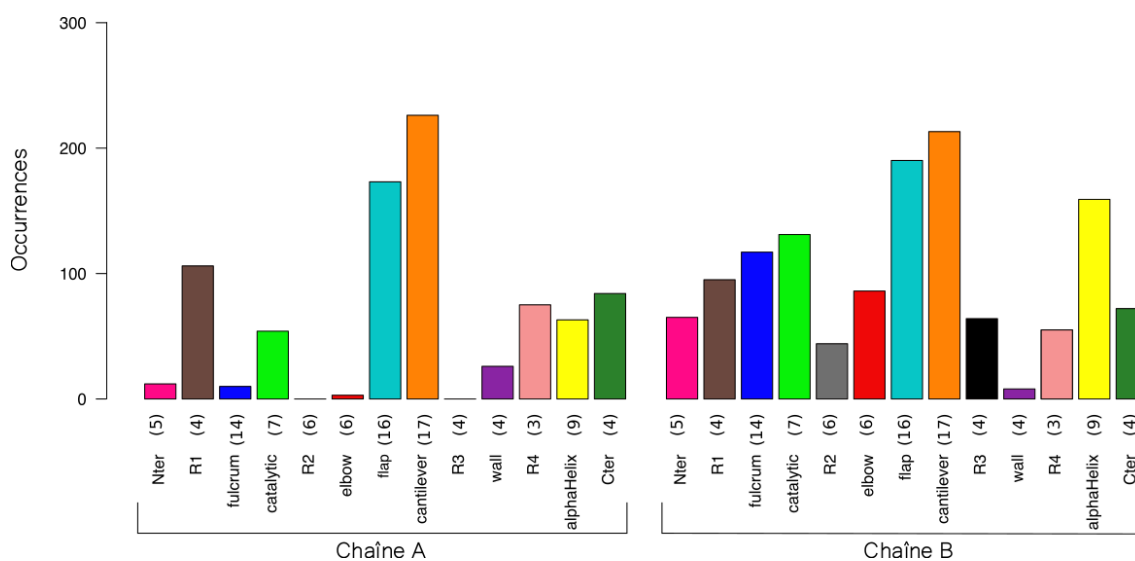


Figure 3.4 : Nombre d’atomes MCS par région de PR2. Les régions sont colorées suivant le code couleur de la figure 3.1. Figure adaptée de Laville *et al.*, 2021.

3.4 Impacts des mutations sur le site de liaison des IP

Pour mieux comprendre le lien entre les mécanismes de résistance et les changements structuraux induits par mutations, nous avons comparé le site de liaison des inhibiteurs des structures sauvage et mutantes. Sur les 31 mutants, 15 présentent au moins une mutation dans le site de liaison et tous présentent des réarrangements structuraux liés aux mutations localisées dans ce site de liaison (à l’exception du mutant I82M). De plus, des modifications structurales au sein de la poche ont aussi été remarquées pour cinq mutants avec des mutations hors du site de liaison (K7R, I54L, V62A, I54L/L90M et I54M/L90M). Ces résultats confirment que même des mutations localisés hors du site de liaison peuvent avoir des impacts sur la fixation des IP.

Afin d’étudier plus finement l’impact des mutations sur le site de liaison des inhibiteurs, nous avons comparé les sites de liaison extraits des structures sauvage et mutantes. Pour cela, les poches estimant le site de liaison ont été extraites de chaque structure en utilisant l’approche du « ligand commun » développée dans Triki *et al.*, 2018b. La poche d’une structure est définie comme l’ensemble des atomes situés à 4,5 Å d’un ligand virtuel (le « ligand commun »). Ce ligand est une superposition des 18 ligands extraits des complexes cristallographiques de PR2 disponibles dans la PDB.

Dans un premier temps, nous avons construit une classification hiérarchique des 156 poches sur la base de leur similarité structurale. Pour cela, le RMSD sur tous les atomes de la poche a été calculé entre chaque paire de poches. Cette classification des poches révèle que les mutants contenant les mutations K7R, I54M, I54L, I82F et I84V sont ceux qui présentent des poches les plus éloignées structurellement de la poche sauvage. Ceci indique

que ces mutations sont celles qui entraînent le plus de changements de conformation au niveau de la poche.

L'impact des mutations sur la poche a également été quantifié à l'aide de trois descripteurs qui quantifient l'hydrophobicité, le volume et la sphéricité de la poche. Ces descripteurs ont été calculés pour toutes les poches avec le logiciel PockDrug (Borrel *et al.*, 2015). La comparaison des valeurs de ces descripteurs pour toutes les poches a montré que les mutations K7R, K45R, V47A, G48V, I82F/M et L90M sont celles qui ont le plus modifié les propriétés du site de liaison, particulièrement leur volume ou leur hydrophobicité.

3.5 Comparaison des interfaces entre structures sauvages et mutantes de la PR2

Dans une troisième partie, nous avons cherché à caractériser plus spécifiquement les impacts des changements structuraux identifiés entre les structures sauvage et mutantes sur la région de l'interface des deux monomères de la PR2, région importante pour la stabilité de la cible. Pour cela, nous avons extrait et comparé les interfaces des 156 structures. L'extraction des interfaces des structures a été réalisée grâce au logiciel *PPIC Protein-Protein Interface Computation* (Laville *et al.*, 2021). Ce logiciel est présenté en annexe (page 105).

Les interfaces entre les monomères des structures sauvages et mutantes ont été déterminées avec le logiciel PPIC (Laville *et al.*, 2018). Parmi les 31 mutants, 13 d'entre eux présentent au moins une mutation dans l'interface. Exceptés les mutants I50L et I54M, tous présentent des déformations structurales au niveau de l'interface. De plus, la localisation des réarrangements structuraux par rapport à l'interface de chaque structure a permis de dénombrer cinq mutants ne contenant pas de mutation dans l'interface mais y présentant tout de même des déformations. Pour identifier l'impact des mutations sur l'interface, la taille de ces régions et leur composition en d'acides aminés ont été comparées. La taille des interfaces a été quantifiée avec le calcul de la surface accessible au solvant à l'aide du logiciel NACCESS (Hubbard and Thornton, 1993). Nous avons aussi déterminé une classification hiérarchique des 156 interfaces qui regroupe les interfaces en fonction de leur composition en acides aminés. Pour cela, nous avons calculé la distance de Hamming normalisée entre chaque paire d'interfaces qui se calcule à partir du nombre d'atomes total de chaque interface d'une paire et du nombre d'atomes communs entre les deux interfaces de la paire. Cette classification montre que les mutations K7R, I54L, L90M et L99F, seules ou en combinaison avec d'autres mutations, sont celles qui induisent le plus de réarrangements structuraux au niveau de l'interface et qui modifient le plus la composition et la taille de l'interface.

3.6 Conclusion

Dans cette étude, l'impact de 31 mutants de résistance sur la structure de la PR2 a été étudié. Les résultats sont présentés dans l'article de [Laville *et al.*, 2021](#), reproduit dans les pages suivantes. Nous avons, dans un premier temps, comparé les modèles des structures mutantes à la structure sauvage et plus spécifiquement les positions de leurs atomes. Cette étude a permis de localiser les réarrangements structuraux potentiellement induits par les mutations de résistances.

Nous avons montré qu'une mutation pouvait induire de larges déformations structurales sur le résidu muté, dans son voisinage ou même à distance de celui-ci. Dans un deuxième temps, nous nous sommes focalisés sur la comparaison de régions clés de la PR2 : la poche de fixation des inhibiteurs et son interface, extraites des structures sauvages et mutantes. Il a été observé que certaines mutations (K7R, V47A, G48V, I50L/V, I54L/M, V62A, I82F/M, I82M, I84V et L90M) induisent des réarrangements structuraux modifiant les propriétés de poche, ce qui pourrait avoir un impact de manière directe ou indirecte sur la fixation des IP. Certaines mutations ont par ailleurs une influence sur la conformation des régions elbow et flaps, régions impliquées dans la transition entre les formes ouverte et fermée lors de la fixation du ligand (K7R, E37D, S43T, K45R, V47A, G48V, I50V, I50L, I54L, I54M et V62A). Enfin, d'autres mutations engendrent des réarrangements au niveau de l'interface (K7R, G48V, I50L, I54L, L90M et L99F), pouvant modifier la stabilité de la PR2. Ces résultats montrent que les mutations de résistance étudiées ici peuvent altérer la fixation des IP en modifiant les propriétés et la flexibilité de la poche, l'interaction entre PR2 et un IP et/ou altérer la stabilité de PR2.




L'originalité de cette étude repose sur la prise en compte d'un grand nombre de mutants de résistance de PR2 par des approches structurales. Néanmoins, elle présente quelques limites. La première est l'établissement d'un seuil de distance afin de définir les réarrangements structuraux qui influence directement les déplacements induits comme tels ou non. Ensuite, une seule structure cristallographique a été utilisée comme modèle lors de la modélisation des structures mutantes. Il aurait été intéressant de comparer les réarrangements structuraux détectés sur différentes structures cristallographiques. Par ailleurs, ceux-ci sont dépendants en partie de notre protocole de modélisation bien qu'une grande partie (78 %) aient été détectés également par l'autre méthode de modélisation, Robetta ([Kim *et al.*, 2004](#)). Enfin, l'aspect dynamique de la protéine, capital dans la compréhension des mécanismes biologiques, n'a pas été pris en compte ici.

3.7 2^e article

Ci-dessous est reproduit notre second article *Structural Impacts of Drug-Resistance Mutations Appearing in HIV-2 Protease*, publié dans le journal *Molecules* le 25 janvier 2021 ([Laville *et al.*, 2021](#)).

Article

Structural Impacts of Drug-Resistance Mutations Appearing in HIV-2 Protease

Pierre Laville , Michel Petitjean  and Leslie Regad * 

Université de Paris, BFA, UMR 8251, CNRS, ERL U1133, Inserm, F-75013 Paris, France; pierre.laville@u-paris.fr (P.L.); petitjean.chiral@gmail.com (M.P.)

* Correspondence: leslie.regad@u-paris.fr; Tel.: +33-1-57-27-82-72

Abstract: The use of antiretroviral drugs is accompanied by the emergence of HIV-2 resistances. Thus, it is important to elucidate the mechanisms of resistance to antiretroviral drugs. Here, we propose a structural analysis of 31 drug-resistant mutants of HIV-2 protease (PR2) that is an important target against HIV-2 infection. First, we modeled the structures of each mutant. We then located structural shifts putatively induced by mutations. Finally, we compared wild-type and mutant inhibitor-binding pockets and interfaces to explore the impacts of these induced structural deformations on these two regions. Our results showed that one mutation could induce large structural rearrangements in side-chain and backbone atoms of mutated residue, in its vicinity or further. Structural deformations observed in side-chain atoms are frequent and of greater magnitude, that confirms that to fight drug resistance, interactions with backbone atoms should be favored. We showed that these observed structural deformations modify the conformation, volume, and hydrophobicity of the binding pocket and the composition and size of the PR2 interface. These results suggest that resistance mutations could alter ligand binding by modifying pocket properties and PR2 stability by impacting its interface. Our results reinforce the understanding of the effects of mutations that occurred in PR2 and the different mechanisms of PR2 resistance.

Keywords: drug-resistance mutations; HIV-2 protease; structural characterization; induced structural deformations



Citation: Laville, P.; Petitjean, M.; Regad, L. Structural Impacts of Drug-Resistance Mutations Appearing in HIV-2 Protease. *Molecules* **2021**, *26*, 611. <https://doi.org/10.3390/molecules26030611>

Academic Editors: Marco Tutone and Anna Maria Almerico
Received: 18 December 2020
Accepted: 19 January 2021
Published: 25 January 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The human immunodeficiency virus (HIV) affects humans and causes the acquired immunodeficiency syndrome (AIDS). The treatment against HIV-1 infection corresponds to the same molecules that target four HIV proteins: the fusion protein, protease (PR), integrase, and reverse transcriptase. The same molecules are used to fight HIV-2 infection but HIV-2 is naturally resistant to most of these inhibitors [1–9]. Thus, it is important to develop new molecules that inhibit the HIV-2 replication, particularly against the HIV-2 protease (PR2).

PR is a homodimer that plays a major role in the virus maturation process: it hydrolyzes the viral polyproteins Gag and Gag-Pol, causing the development of immature virions. There are currently nine protease inhibitors (PIs) clinically recommended for treating HIV-1 infection [10]. These drugs bind to the PR catalytic site in the interface of the two monomers. This binding induces structural changes in the entire PR2, particularly in the flap region allowing the closing of binding pocket [10–12]. PR2 is naturally resistant to most of these PIs and only three of them are recommended for the treatment of HIV-2 infection: darunavir (DRV), saquinavir (SQV), and lopinavir (LPV) [1,10]. The natural resistance of PR2 is explained by amino-acid changes between PR1 and PR2 that induce structural changes in the entire structure [13]. Some of these structural changes, located in the binding pocket, modify properties and conformation of the PI-binding pocket and the internal interactions between PR2 and PIs [3,5,14–20]. Other structural changes, occurring

in the elbow and flap regions, alter the transition between the open and closed forms involved in ligand binding [13,21,22].

In addition to its natural resistance, many acquired resistance mutations appear in PR2. The identification of these mutations have been performed using genome sequencing studies of HIV-2 virus extracted from infected patients [2,4–6,8,23–28]. For example, it has been shown that the V47A, I50V, I54M, I82F, I84V, and L90M mutations lead to several PIs resistance [2,4,5,7,8]. Phenotypic susceptibility assays were used to confirm the resistance of some mutants [2,4,5,8,26,29]. For example, genotypic and phenotypic analyses showed that the I82F mutation causes resistance to indinavir (IDV) [26]. Furthermore, this mutation has been identified as causing hypersusceptibility to both DRV and SQV using phenotypic assays [2]. In addition, combinations of several mutations confer high resistance to several PIs [2,4,5,7,8,30]. For example, the I54M and I82F mutations induce resistance to all PIs and the V62A/L99F mutant is resistant to nelfinavir, IDV, and LPV [26]. Few studies have focused on the structural analysis of the impacts of drug-resistance mutations reported in PR2 because no tridimensional (3D) structure of PR2 mutant have been solved. These structural studies could help to understand the atomistic mechanism of resistance mutations. In our previous work, we performed the first structural analysis of the impact of 30 drug-resistant mutants of PR2 based on modeled structures. More precisely, we explored the consequences of drug-resistance mutations on PR2 structural asymmetry, an important property for ligand-binding and the target deformation [31]. Our findings suggested three possible resistance mechanisms of PR2: (i) mutations that induce structural changes in the binding pocket that could directly alter PI-binding, (ii) mutations that could impact the properties and conformation of the binding pocket by inducing structural changes in residues outside of the binding pocket but involved in interaction with pocket residues, and (iii) mutations that could modify the PR2 interface and its stability through structural changes in interface residues. These results have been based on PR2 backbone analysis. However, a better characterization of the structural impacts of drug-resistance mutations on PR2 structure including side-chain atoms could help in a better understanding of different proposed mechanisms.

In this study, we structurally analyzed a set of 31 drug-resistant mutants of PR2 that was updated relative to [31]. The 3D structure of each mutant was modeled and its structure was compared to the wild-type PR2 to locate structural rearrangements induced by drug-resistance mutations at backbone and side-chain atoms. The study reported that drug-resistance mutations could impact the flexibility of PR2 and the closing binding pocket, conformation and properties of PI-pocket and the composition and size of the PR2 interface.

2. Results

We studied the impact of 22 drug-resistance mutations on PR2 structure. These mutations appeared alone or in combination with others (two or three mutations per mutant) resulting in a set of 31 mutants, Figure 1A. First, mutant structures were modeled using FoldX software [32] and an energetic minimization step using the crystallographic structure of the wild-type PR2 (PDB code: 3EBZ [33]). Five 3D structures were built for each mutant to consider the different possible rotamers per amino acids as illustrated by Figure A1. This mutagenesis process resulted in a set of 155 mutant structures. The crystallographic structure of the wild-type PR2 (PDB code: 3EBZ [33]) was also energetically minimized with the protocol used for mutant structures. In the following, the minimized wild-type structure was referred as the wild-type structure and its structure was compared to the minimized mutant structures.

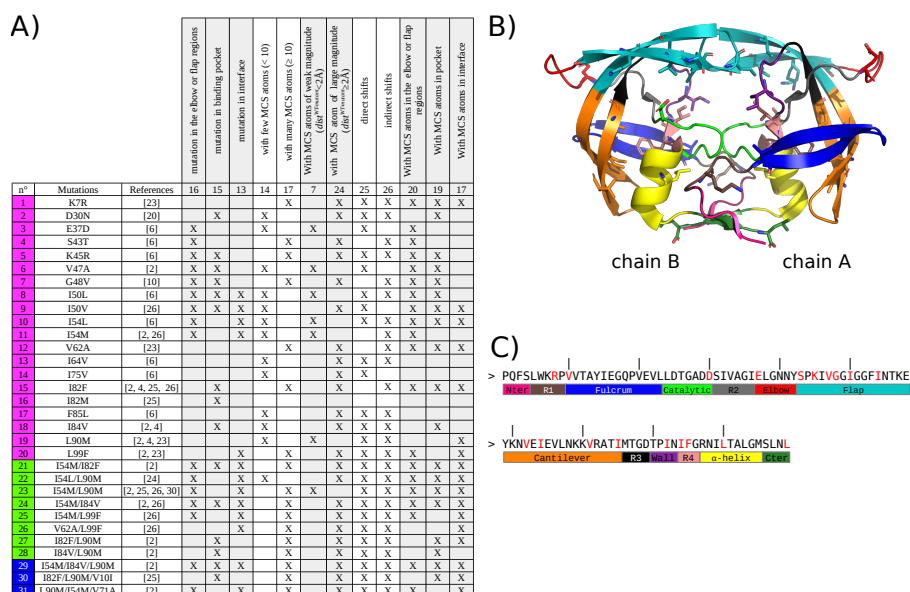


Figure 1. Description of the 31 drug-resistant mutants studied in this analysis. (A) Table listing the 31 drug-resistant mutants studied in this analysis. Single mutants are colored in magenta, double mutants in green and triple mutants in blue. (B) Location on PR2 structure of the 22 drug-resistance mutations included in the 31 mutant set. PR2 is represented in cartoon mode and colored according to the 13 regions defined in [34,35]. Mutations are represented in stick mode. (C) Amino acid sequence of PR2 presenting the limit of the 13 PR2 regions. All the mutated residues are colored in red in the sequence.

2.1. Identification of Atom Shifts in the Mutant-Structure Set

We first explored the structural deformations induced by mutations by locating shifted atoms in mutant structures relative to the wild-type structure. The shift of an atom was quantified by $dist^{WT-mutant}$, i.e., the distance between its positions in the wild-type and mutant structures. Figure A2 presents the distribution of $dist^{WT-mutant}$ of all atoms in the mutant structures set, except hydrogen atoms. These distances varied from 0.05 to 0.05, with 95% of atoms exhibited a $dist^{WT-mutant}$ smaller than 0.14 Å. These distances were summarized per structures by computing the RMSD between the wild-type and mutant structures (Figure A2). As expected, mutant structures exhibited close conformation than the wild-type, resulting in an average RMSD of 0.12 ± 0.09 Å. Only one mutant structure (one structure of the K7R mutant) exhibited a RMSD higher than 0.5 Å. Like in Liu et al., 2008 [36], we considered that a shift was significant if it had a magnitude higher than 0.3 Å because of uncertainties in the X-ray and mutant structures. Thus, to identify atom shifts induced by mutations, we retained the 2136 atoms with a $dist^{WT-mutant}$ higher than 0.3 Å in at least three structures of a given mutant. These selected atoms were denoted as “mutant-conserved shifted atoms” (MCS atoms) and the distribution of their $dist^{WT-mutant}$ is provided by Figure 2A. Most MCS atoms (70%) exhibited a shift of moderate magnitude (<1 Å). However, 17% of MCS atoms are of large magnitude (>2 Å), such as atom shifts detected in residue K69_A (Lysine 69 in chain A) in the L99F mutant that is illustrated in Figure 2B. Figure 2C presents another type of atom shift that corresponds to a flip of a ring of residue 3_B (residue 3 in chain B) in the five structures of the V62A/L99F mutant. This rearrangement type does not induce structural deformation and thus it does not seem to be linked to PI resistance.

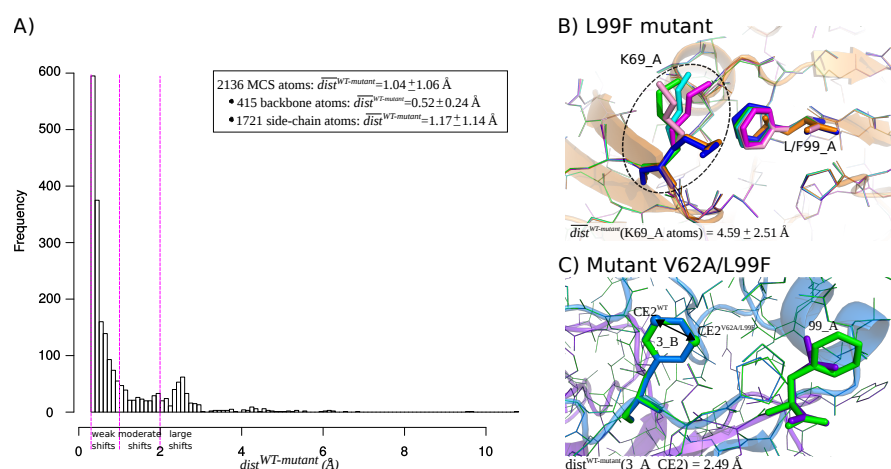


Figure 2. (A) Distribution of $dist^{WT-mutant}$ distances for MCS atoms extracted from the set of 155 mutant structures. Magenta lines corresponds to the cutoffs used to define a weak shift ($0.3 \text{ \AA} < dist^{WT-mutant} < 1 \text{ \AA}$), moderate shifts ($1 \text{ \AA} \leq dist^{WT-mutant} < 2 \text{ \AA}$), and large shifts ($dist^{WT-mutant} > 2 \text{ \AA}$). (B,C) Illustration of atom shifts in the L99F (A) and V62A/L99F (B) mutants. (B) Superimposition of the five structures of the L99F mutant and the wild-type structure. Wild-type structure is colored in orange and represented in line and cartoon modes. The five structure of the mutant are represented in line mode and colored in magenta, cyan, blue, green, and pink. The L99F mutated residue is represented in stick mode. (C) Illustration of structural shift occurring at residue 3_B in the V62A/L99F mutant. The wild-type structure is represented in cartoon and line modes and colored according to its two chains: chain A is colored in purple and chain B is colored in marien blue. The mutant V62A/L99F structure is represented in lines and colored in green. The mutated residue 99_A and shifted residue 3_A are represented in stick mode. The arrow represents the $dist^{WT-mutant}$ of the CE atom of residue 3_B computed between the wild-type and the first structure of mutant V62A/L99F. $CE2^{V62A/L99F}$ and $CE2^{WT}$ correspond to atom CE of residue 3 of chain B in the V62A/L99F and wild-type structures and are represented in sphere mode.

As expected, side-chain atoms were overrepresented in the MCS atom set (Pearson's Chi-squared Test p -value = 3×10^{-28}) and they had larger $dist^{WT-mutant}$ value than backbone atoms (Student's t -test p -value of 6×10^{-95}). This means that drug-resistance mutations have more impacts on side-chain atoms than on backbone atoms. From the 2136 MCS atoms, 543 (=25%) are atoms of mutated residues. The shift of these atoms, named direct shifts, was a direct consequence of mutations. In contrast, 1593 (=75%) MCS atoms corresponded to indirect shifts, i.e., they occurred in non mutated residues, and their shifts resulted either from the intrinsic flexibility of atoms or from indirect impacts induced by the mutation through contacts between these atoms and mutated residues. Direct shifts had larger magnitude than indirect shifts, i.e., they exhibited an average higher $dist^{WT-mutant}$ distances than indirect shifts (Student's t -test p -value = 3×10^{-13}). To distinguish structural shifts resulting from mutation from those induced by flexibility, we detected MCS atoms in non minimized structures, i.e., mutant structures corresponded to the output of FoldX software. From this set of non-minimized structures, we located 883 MCS atoms, with 646 that were also detected as MCS atoms in the minimized structures set. This means that the shift of these MCS atoms observed in minimized structures (30% of MCS atoms) were the consequence of the mutation. As expected, these shifts, observed in minimized structures, occurred only on side-chain atoms and in or close to the mutated residue because FoldX software optimizes the configuration of only side chains in the vicinity of the mutated residue. The remained detected shifts were explained by the mutation and intrinsic flexibility of atoms.

Figure 3A shows that the entire PR2 structure was sampled by MCS atoms. A total of eight regions had few MCS atoms, i.e., less than 50 MCS atoms were detected in the

155 mutant structures (Figure 3). In contrast, some regions exhibited many MCS atoms, such as the cantilever and flap regions of the two chains and the α -helix region of chain B, with more than 150 MCS atoms. In addition, an asymmetry between the two chains in terms of number of MCS atoms was observed. Indeed, chain B contained more MCS atoms than chain A (p -value of the Pearson's Chi-square test is of 5×10^{-24}). For example, the Nter, fulcrum, elbow, and R3 regions of chain A present few structural shifts, while they exhibited lots of deformations in chain B. Thus, even though mutations occurred in the two chains, they did not impact in the same way the two monomers.

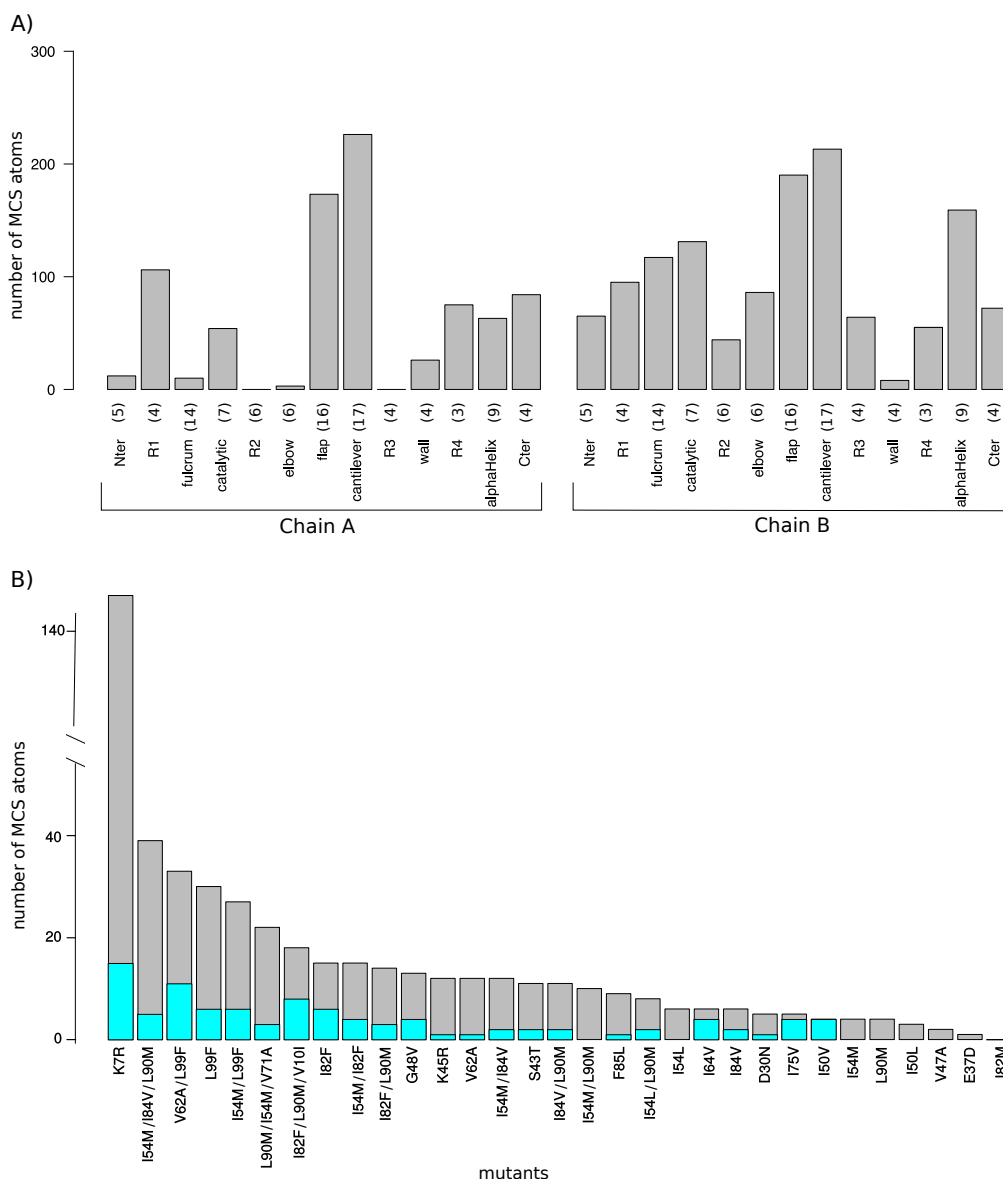


Figure 3. (A) Number of MCS atoms per mutant counted in the two PR2 chains. Each PR2 chain was divided in 13 regions according to [13]. Numbers in brackets indicate the size of each region in terms of amino acids. (B) Number of MCS atoms (grey) and MCS atoms with large magnitude ($dist^{WT-mutant} \geq 2 \text{ \AA}$, cyan) per mutants. Mutants were sorted according to their number of MCS atoms.

Figure 3B presents the number of MCS atoms per mutants. On average, a mutant contained a moderate number of MCS atoms (16.26 ± 26.09). The I82M mutant was particular because it had no MCS atoms, revealing that the I82M mutation induced few structural changes in PR2 structure. In contrast, the K7R mutant was the mutant with the most MCS atoms (147 MCS located in 53 different residues as illustrated in Figure 4). A total of 14 mutants had less than 10 MCS atoms, indicating that these mutations induced few impacts on PR2 structure (Figure 1A). Except the I54L/L90M, all these mutants were single mutants. Although these mutants exhibited few deformations, 57% of them had shifts with a large magnitude ($dist^{WT-mutant} \geq 2 \text{ \AA}$). For example, the D30N mutant induced 5 MCS atoms with one exhibiting large shift of 2.9 \AA , while the L90M mutant caused four small shifts at residues 90_A/B and 97_B with a magnitude varying from 0.35 to 0.48 \AA (Figures 3B and A3). In contrast, 17 mutants had many MCS atoms and all of these shifts had of large magnitude (Figure 3B). From these mutants, seven were single mutants, revealing that only one mutation could cause large deformations, such as those observed for the K7R and L99F mutants (Figures 4 and A3).

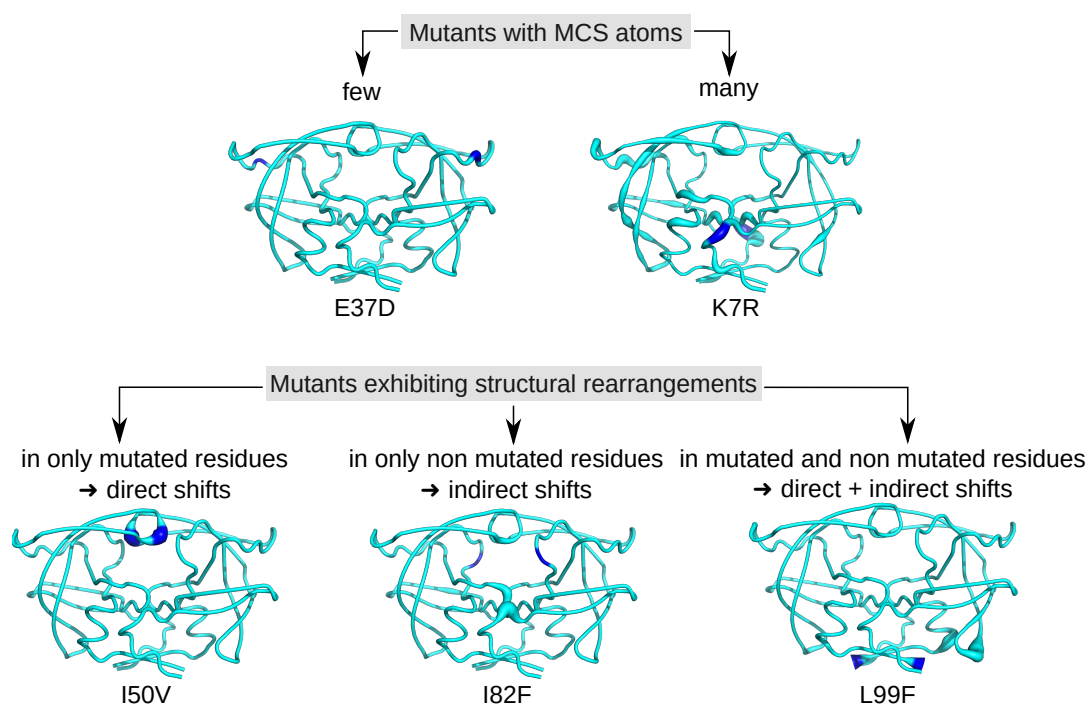


Figure 4. Location of the MCS atoms found in some mutants. PR2 is colored in cyan and represented in putty mode. The putty radius is relative to deformations induced by mutations: the higher the radius, the stronger the mutation-induced rearrangement. Mutated residues are colored in blue.

According to the location of MCS atoms, we differentiated three types of mutations (Figure 4). The first mutations corresponded to four mutations that induced structural rearrangements only in the mutated residues, i.e., having only direct impacts. For example, the I50V mutation caused structural changes in two atoms of residue 50 of the two chains, a residue involved in the binding pocket, the dimer interface and the flap region (Figure 4). The second type of mutations grouped five mutations impacting residues in their vicinity or further in structure, i.e., inducing indirect structural changes, such as the I82F mutations (Figure 1). Indeed, the I82F mutation induced many atoms rearrangements in five non mutated residues (8_A, 8_B, 21_B, 27_B, 49_B) with some of them of large magnitude. In this mutant structure, residue 8_A is located in the vicinity of mutated residue 82_A (located at less than 5.5 \AA), while residues 27_B and 49_B that are

located at more than 6 Å of the mutated residue 82_A (Figure A8). The last type of mutations corresponded to mutations that induced both direct and indirect rearrangements (Figures 1 and 4). This mutation type grouped most of mutations. Figure 4 showed that the L99F mutation produced large shifts in the mutated residues and also in its neighbor residues 68 and 69.

The location of MCS atoms (Figure A3) in mutant structures highlighted structural rearrangements located in important regions for PR2: in the elbow and flap regions that are implied in the PR2 deformations induced by ligand binding, in its pocket binding and in its interface. In the following, we explored the impacts of the studied resistance mutation in the PI-binding pocket and PR2 interface.

2.2. Impact of Mutations on the Properties of PI-Binding Pocket

From the 31 mutants, 15 had at least one mutation in the binding pocket (Figures 1A and A4). Except the I82M mutant, all these mutants presented MCS atoms in the pocket in the mutated or non mutated residue. Surprisingly, structural rearrangements in the binding pocket were also observed in five mutants without pocket mutations (K7R, I54L, V62A, I54L/L90M, I54M/L90M, Figure 1). A total of 36% of pocket atoms were deformed in at least one mutant, with an overrepresentation of side-chain atoms (Pearson's Chi-squared Test p -value = 1×10^{-3}), see Figure A4.

To explore impacts of these mutations on the conformation and properties of the PI-binding pocket, PI-binding pockets were extracted from the 156 structures (1 wild-type and the 155 mutant structures). These 156 pockets were then classified according to their structural similarity quantified by pairwise RMSD (Figures A5 and 5). In addition, their volume, sphericity, and hydrophobicity values were compared to those of the wild-type pocket (Figure 6). First, the five structures of a given mutant were not always bundled in the classification or presented some variability in terms of descriptor values. This highlighted a structural diversity of the five structures of mutants. This is explained by the minimization effects and the fact that several rotamers were possible for some amino acids during the mutagenesis process as illustrated by Figure A1 for the K45R mutation. Except structures of the I82M, I54M, L90M, I54M/L99F, and I54M/V71A/L90M mutants, most structures of mutants without MCS in pocket were close to the wild-type pocket in the hierarchical classification and presented similar descriptor values than the wild-type pocket (Figures 5 and 6). Pocket of mutants with the K7R, I54M, I54L, I82F, and I84V mutations were the farthest to the wild-type pocket in the classification, meaning these mutations had the most impact on the pocket structure (Figure 5). Figure 6 showed that the K7R and I82F mutations also strongly modified pocket properties, like the K45R, V47A, G48V, I82M/F, and L90M mutations. More precisely, the V47A, I82F, and I82M mutations strongly decreased the pocket hydrophobicity. The I50V, I50L, V62A, and I84V mutations also caused a reduction of pocket hydrophobicity but with a weaker magnitude. The I50L and I82F mutations were also responsible of an increase of the pocket volume, in contrast to the I82M and I84V mutations that caused a reduction of the pocket volume. The G48V and I54M mutations increased the hydrophobicity of the pocket that was accompanied with a modification of the pocket size: the G48V mutation led to a reduction of the pocket volume in contrast to the I54M. An increase of pocket volume was also observed in the K7R, I54L, and L90M mutants with different magnitudes and a decrease of the pocket volume in the D30N and V47A mutants. The volume modification of the pocket of the K7R and D30N mutants was accompanied with an increase of the sphericity of the pocket. The K45R mutant was distinct because its five structure presented large diversity in terms of descriptor values Figure 6. Two of these pockets were bigger and less hydrophobic than the wild-type pocket while the three others were smaller, more hydrophobic and more spheric than the wild-type pocket.

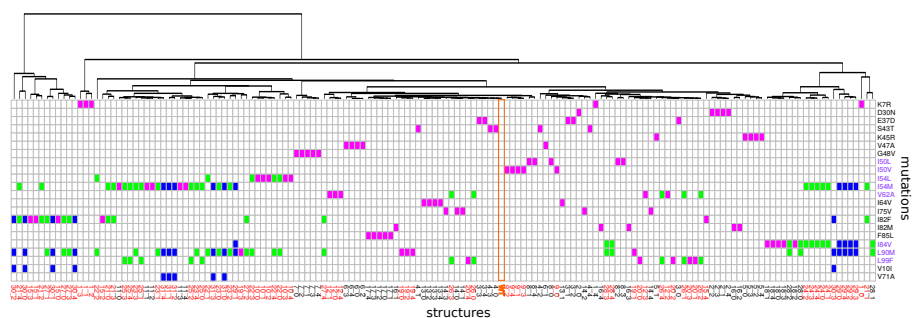


Figure 5. Hierarchical classification of wild-type and mutant pockets according to their conformational similarity quantified by pairwise RMSD computing using all pocket atoms. The table provides the description of each mutant structure in terms of mutations. Mutant structures are ranked according to their apparition in the classification. Single mutants are colored in magenta, double mutants in green and triple mutants in blue. The orange column locate the wild-type pocket. Mutations colored in purple correspond to mutations located in the binding pocket and mutant colored in red contain a MCS atom located in the binding pocket.

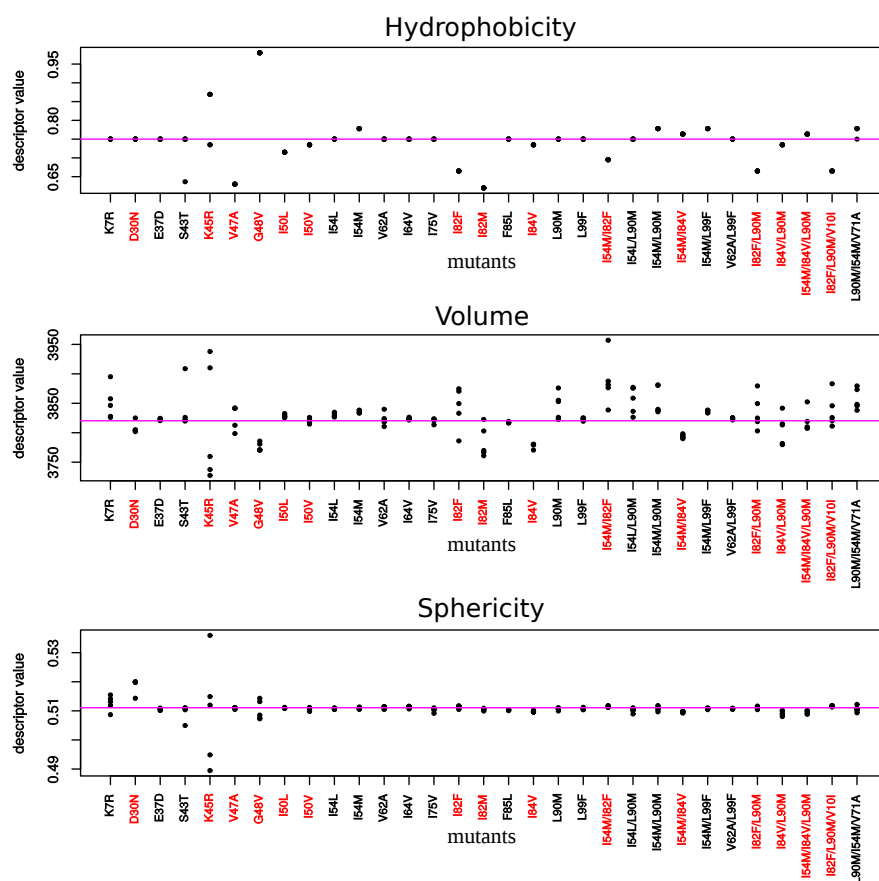


Figure 6. Hydrophobicity, volume, and sphericity of pockets extracted from each mutant structure. Red mutants correspond to mutants exhibiting at least one mutation in the binding. Each point corresponds to a mutant structure. For some mutants, less than five points appear meaning that several mutant structures exhibit same values for a given descriptor.

Figure A4 presents the MCS atoms occurring in the binding pocket in each mutant. We noted that some mutations induce structural rearrangements in residues important

for the PI binding. For example, the K7R, I50L/V and I54L mutations caused structural deformations in residues 25, 27, 30 that establish hydrogen bonds with PIs [18,33]. The D30N, K7R, I82F, I84V mutations led to atomic displacements in residues involved in van der Waals interactions with PI, such as in residues 23B, 27A, 28A, 30A, 49B, 48B, 82A and 84A.

Thus, the K7R, K45R, V47A, G48V, I50V/L, I54M, V62A, I82M/F, I84V, and L90M mutations could impact ligand binding by modifying pocket properties or the network of interactions with PIs.

2.3. Impact on Interface

From the 31 drug-resistant mutants, 13 had at least one mutation in the PR2 interface (Figure 1A). Except the I50L and I54M mutants, all these mutants contained MCS atoms in their interface. Five mutants without mutation in the interface presented structural deformations in their interface. To analyze the impact of these mutations on the PR2 interface, interfaces of the wild-type and mutant structures were extracted and compared in terms of amino acid composition and their size. To do so, a hierarchical classification of the 156 interfaces was computed according to their similarity in terms of interface composition (Figure 7). In addition, the Solvent Accessible Surface Area (SASA) value, measuring the interface size, of the two parts of the interface was computed for each mutant structure using NACCESS software [37] (Figure 8). Figure 7 showed that most structures without MCS atoms in interface were close to the wild-type interface in the classification, revealing that these mutations led weak changes in the PR2 interface. This was confirmed by the fact that these interfaces had similar size than the wild-type interface (Figure 8). Three mutants (I50L, I54M and I84V/L90M) without MCS atom in interface exhibited different interface composition relative to the wild-type. These differences in terms of interface composition led to an increase of the size of chain A interface in the I50L and I84V/L90M mutants. The G48V mutation was responsible of the presence in the interface of the two side-chain atoms of residue 48_A and the absence in the interface of one atom of residue 95_B and 99_B relative to the wild-type interface (Figure A6), resulting that the chain A interface of the mutant was larger and this of chain B was smaller than the wild-type interface (Figure 7). These differences in terms of interface observed in the I50L, I54M, and G48V mutants were explained by supplementary atoms in their interface induced by the mutation (Figure A6).

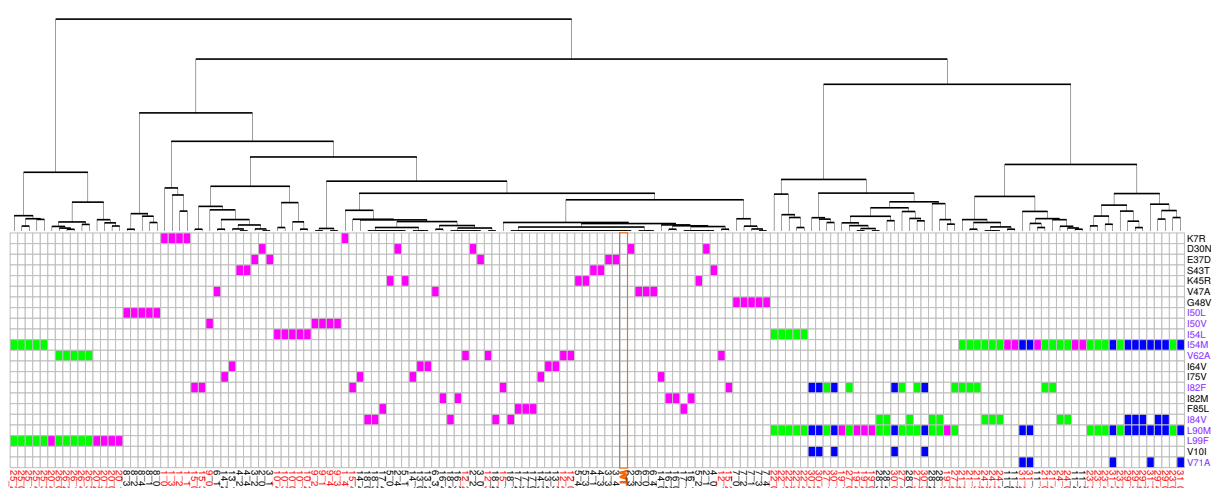


Figure 7. Classifications of the 156 structures according to their similitude of their interface. The table provides the description of each mutant structure in terms of mutations. Mutant structures are ranked according to their apparition in the classification. Single mutants are colored in magenta, double mutants in green and triple mutants in blue. The orange column locate the wild-type interface. Mutations colored in purple are involved in the interface. Residues colored in red correspond to MCS atoms in the interface.

mutations did not significantly increase the average number of MCS atoms per mutant (Student's *t*-test *p*-value = 0.59, Figure 3B). Comparison of MCS atoms in the single and multiple mutants showed that several multiple mutants exhibited specific MCS atoms relative to the corresponding single mutants. For example, we highlighted a displacement in atoms of residue 50_B in the I54L and I54L/L90M mutants but the shift was larger in the double ($dist^{WT-mutant}$ for CD_50_B and CG1_50_B atoms >2.5 Å) than in single mutants ($dist^{WT-mutant}$ for CD_50_B and CG1_50_B atoms was of 0.69 and 0.62 Å in the I54L mutant). Combining several mutations could induce apparition or loss of MCS atoms relative to the corresponding mutants. For example, the combination of the I54M and I82F mutations caused structural shifts in pocket residues 50_A, 81_A/B, and 82_A/B, while no shift at these residues were observed in the I54M and I82F mutants (Figure A4). In contrast, the I82F mutant contained a large shift at residue 8_B, a residue involved in the PR2 interface and pocket, those were not found in the I54M/I82F mutants (Figure A4). These structural changes in the double mutant relative to the I82F mutant induced a weak decrease of the interface size and an increase of the pocket volume (Figure 6).

2.5. Impact of Using Different Structure Modeling Software

In this section, we explored the impact of using another structure-modeling software in the detection of structural rearrangements induced by mutations. To do so, we modeled the structure of the 31 mutants using the webserver Robetta, see Appendix G. These mutant structures were denoted as $mutant^{Robetta}$. For a better clarity of this section, the mutant structures built using our initial protocol (based on FoldX software and an energetic minimization step) were denoted as $mutant^{FoldX+Mini}$. First, we compared mutant structures generated with the wild-type structure (crystallographic structure), Figure A7A,B. We noted that the protocol based on FoldX software plus a minimization step led to a set of structures exhibiting a larger diversity than Robetta webserver. Then, the two mutant sets were compared by computing RMSD between $mutant^{Robetta}$ structures and the five structures of $mutant^{FoldX+Mini}$, denoted as $RMSD^{WT*-mutant^{Robetta}}$. The two modeling protocols led to different structures with an average RMSD of 0.46 ± 0.04 between $mutant^{Robetta}$ and $mutant^{FoldX+Mini}$ structures (Figure A7C). We then compared the number of MCS atoms observed in each structure of the two mutant sets, see Appendix G.3. Figure 9 presents the number of MCS atoms detected in each mutant. We noted that $mutant^{Robetta}$ structures exhibited substantially more MCS atoms than $mutant^{FoldX-mini}$ structures. This was explained by two reasons. First, the determination of the MCS atoms in $mutant^{FoldX-mini}$ structures was based on the comparison of the mutant structures and the minimized wild-type structures, while MCS atoms in $mutant^{Robetta}$ structures were detected by comparing mutant structures with the non minimized wild-type structure. The second reason was that an atom was detected as a MCS atom in the $mutant^{FoldX-mini}$ if it had a $dist^{WT-mutant}$ larger than 0.3 Å in at least three structures of the mutant. Figure 9 presents the number of atoms detected as MCS in both mutant sets. From the 504 MCS atoms per mutants detected in the $mutant^{FoldX+mini}$ set, 78 % were detected as MCS atoms in the $mutant^{Robetta}$ set. Although, the protocol based on FoldX software minimized the detection of structural rearrangements, the extracted structural deformations using this protocol was mainly found by a protocol based on another modeling software.

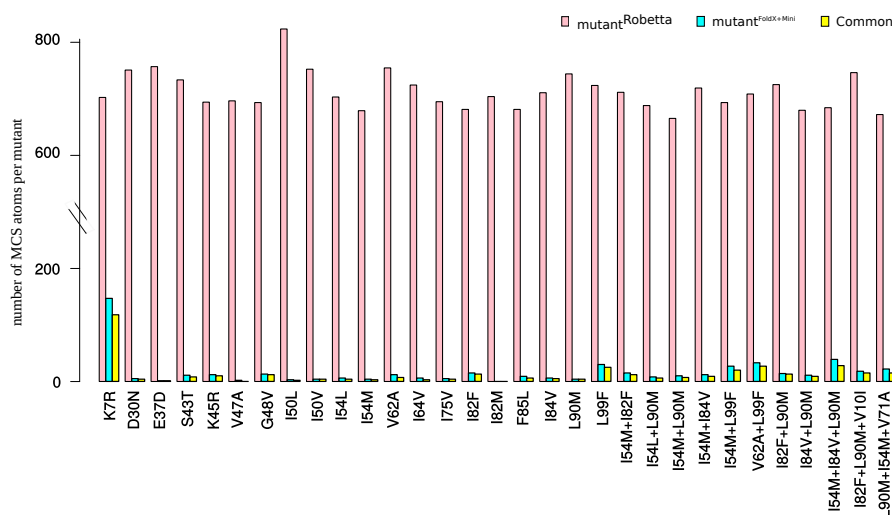


Figure 9. Number of MCS atoms in the *mutant^{Robetta}* set, in the *mutant^{FoldX+mini}* set, and observed both in the two mutant sets.

3. Discussion

In this paper, we proposed a quantification and location of structural deformations at backbone and side-chain atoms of PR2 occurred in an update set of 31 drug-resistant mutants based on their modeled structures. We detected a set of atoms presenting a shift in at least one mutant structure relative to the wild-type structure. To identify structural rearrangements resulted from the mutation, we then retained only MCS atoms, i.e., atoms with a distance between its position in the mutant and wild-type structure higher than 0.3 Å in at least three of structures of a given mutant. The distance cutoff was set up to 0.3 Å to select only significant shifts and considering uncertainties in the X-ray structure, like in Liu et al. [36]. This step allowed us to detect on average 16.26 ± 26.09 MCS atoms per mutants. However, these two cutoffs could lead to an under-estimation of detected structural shifts. This could explain the fact that no MCS atom was detected in the I82M mutant, while the binding pocket extracted from this mutant was smaller and less hydrophobic than the wild-type pocket. Indeed, using a distance cutoff of 0.1 Å, we counted 96 ± 107 MCS atoms per mutants and 17 MCS atoms in the I82M mutant.

The analysis of MCS atoms showed that they occurred in side-chain and backbone atoms in the mutated residues, in its vicinity or further in the structure. We noted that structural rearrangements in side-chain was more frequent and with larger magnitude than those observed in backbone atoms. Thus, drug-resistance mutations induced more deformations in side-chains than in backbone atoms. These results suggest that to combat against HIV drug resistance, it would be interesting to develop inhibitors that establish hydrogen interactions with backbone atoms. Favor backbone interactions between PI and PR is a strategy used for the design of DRV to avoid the detrimental effects of resistance mutations [38,39].

Our results showed that the studied drug-resistance mutations impacted all PR2 regions. However, even though mutations occurred in the two PR2 chains, we noted an asymmetry in the impact of mutations in the two chains. This asymmetric behavior of mutations is linked to the fact that PR2 is a structural asymmetric protein, i.e., its two chains exhibit different conformations in unbound and bound forms [11,12,35,40–43]. This structural asymmetry of PR2 could result from crystal packing, ligand binding, and intrinsic flexibility of PR2 [11], and may be involved in the structural changes of PR2, particularly upon ligand recognition and binding [11,12,41,42]. Our results are in agreement with previous findings that have showed that drug-resistance mutations could modify PR2 structural asymmetry [31]. Thus, as PR2 is an asymmetric protein, resistance mutations do not always have the same impact on the two chains.

The location onto PR2 structure of MCS atoms highlighted structural deformations that could be linked to resistance mechanisms. First, we observed, in 19 mutant structures, structural deformations in binding pocket residues as reported in Laville et al., 2020 [31]. Most of these mutations were located in the binding pocket, except the K7R, I54M and L90M mutations. Our findings revealed that the S43T, V47A, K45R, G48V, I82F, and I82M strongly modified pocket hydrophobicity. The two first decreased the hydrophobicity, while the four last increased it. The I50V, I50L, V62A, and I84V mutations induced also a modification of the pocket hydrophobicity but with a very weak magnitude. The K7R, I50L, I54M, I54L, I82F, and L90M mutations increased the pocket volume, while the D30N, V47A, G48V, I82M, and I84V mutations had the opposite effect. By comparing PR1 and PR2 binding pocket, we have previously observed that amino-acid changes occurring in pocket residues 31, 32, 46, 47, 76, and 82 increased the hydrophobicity of the binding pocket [19]. Chen et al. (2014) reported that these mutations have also an impact on the volume of the binding pocket [22]. In addition, the K7R, D30N, I50L/V, I54L, I82F, and I84V mutations seemed to have direct impact on the PI binding by causing structural rearrangements in pocket residues that establish hydrogen and van der Waals interactions with PIs. Secondly, our finding reported that mutations K7R, E37D, S43T, K45R, V47A, G48V, I50V, I50L, I54M, I54L, and V62A, occurring alone or in combination with others, induced directly or indirectly structural changes in the elbow and flap regions. These two regions are known to be important in the opening and closing mechanisms of the binding pocket during ligand binding [10]. Thus, these mutations could impact PI binding by modifying the flexibility and movement of the flap region upon PI binding. Thirdly, we reported that the K7R, I50L, I54L, G48V, L90M, and L99F mutations caused structural displacements that impacted the composition and size of the PR2 interface. This suggested that these mutations may alter the stability of PR2.

Several drug-resistant mutations were structurally studied in PR1 by comparing crystallographic structures of the wild-type and mutants in bound and unbound forms. These studies showed that drug-resistance mutations alter the conformation of flap residues and flap dynamics, modifying binding pocket properties and the interaction network with PI, and the PR2 stability. For example, it has been shown that some resistance mutations, such as the M46L, G48V, I50V, and I54V/M mutations, alter the conformation of flap residues and flap dynamics in PR1 [36,44–48]. The impact of resistance mutations on flap conformation had also observed in several multi-drug resistant mutants such as the PR20 (including 20 mutations), Flap+ (with L10I, G48V, I54V, and V82A mutations), and MDR-769 (with nine mutations) mutants having more opened flap region relative to the wild-type PR1 [49–52]. In addition, Shen et al., (2015) suggested that the E35D, M36I, and S37D mutations in the multi-drug resistant PRS17 mutant induce an increase of the flexibility of the flap region [53]. The impact of the amino-acid changes at residues 30, 50, 82, 84, and 90 on the pocket volume have been also observed in PR1 [48,49,52,54–57]. For example, the V82A and I84V mutations lead to an expansion of the active-site cavity [52,54], while the V82F and L90M mutations cause a volume reduction in the binding cavity [48,49,54,57]. The direct impact of mutations occurred in residues 30, 50, 82, and 84 on the network of PR-PI interactions were previously observed in PR1 [51,57–62]. The impact of the L90M mutation on the PR2 stability was previously observed in PR1 using urea denaturation experiment [63,64] or using sedimentation equilibrium analysis [65]. However, we noted several disagreements between findings obtained in PR1 and PR2. For example, Liu et al., (2008) observed structural deformations around the tip of the flap and 80 s loop (residues 78–82) in PR1 mutants G48V, I50V, I54V/M complexed with SVQ and DRV [36]. Our analysis of PR2 mutants detected displacements of atoms located around the flap tip in the I50V, I54M, and G48V mutants and in residue 53 in the G48V mutant. However, structural deformations in the 80 s loop is only found in the I54M mutant with a weak shift in backbone atom 79_B_0 ($dist^{WT-mutant}$ of 0.44 Å). An important structural rearrangement of the main-chain of residue 25 in the PR1 mutant L90M linking to the L90M resistance was observed in several studies [49,57]. Our result did not

highlight structural deformations at atoms of residue 25 in the L90M mutant of PR2; atoms of residue 25A/B exhibit an average $dist^{WT-mutant}$ of 0.07 ± 0.02 Å. In addition, we noted that the I84V and L90M mutations have not the same effect on the binding pocket in PR1 and PR2. Indeed, our findings showed that, in PR2, the I84V led to a reduction of pocket volume and the L90M mutation induced an augmentation of the pocket size in contrast in PR1 [48,49,52,54,57]. These disagreements in terms of impacts of drug-resistance mutations in PR1 and PR2 could be explained by the methodological differences of the two approaches. Indeed, our study was based on modeled structures, while PR1 studies used crystallographic structures. However, another reason that could explain these disagreements is the structural differences between PR1 and PR2 structures [13,18,19,40,42] leading to different PI-resistance profiles: PR2 is naturally resistant to six of the nine FDA (Food and Drug Administration)-approved PIs available for HIV-1 therapy [1,3,4]. More particularly, our previous work showed that the α -helix region (87–95), containing residue 90, presents different conformations in PR1 and PR2 and these structural differences could be partially explained by amino acid substitutions observed between the two PRs [13]. In addition, we previously showed that two PRs exhibit pockets with different properties: PR2 pockets are smaller and more hydrophobic than PR1 pockets [19,22]. These observations suggested that a same mutation could have different impacts on PR1 and PR2 structures.

In this study, we explored the impact of drug-resistance mutations on PR2 structure. To do so, we modeled 3D structures of mutant using as template the PR2 complexed with DRV (PDB code: 3EBZ [33]). However, it has been shown that drug-resistance mutations exhibit different sensitivities to the nine FDA-approved drugs [2,4,8,10,26]. For example, the I54M mutation leads to moderate resistance for indinavir, nelfinavir, tripanavir, DRV, and LPV, a high resistance for amprenavir, and is susceptible to SQV [26]. Thus, it would be interesting to consider different PIs and to cross information about the complete resistance profile of each mutant and the detected structural deformations. However, reliable resistance profiles for PIs are difficult to collect for all mutants as few studies are available [3,4,26,29] and some of these studies have led to opposing results [10].

4. Materials and Methods

4.1. Data

In this study, we started from the list of the 30 drug-resistant mutants of PR2 studied in our previous study [31]. This mutant list was updated and a list of 31 drug-resistant mutants of PR2 was selected. This mutant set contains 22 different mutations (Figure 1) and there are 20 single mutants (with one mutation), 8 double mutants (with two mutations), and 3 triple mutants (with three mutations). These mutations sample the entire PR2 (Figure 1).

4.2. Structure Modeling

We modeled the 3D structure of each mutant using the protocol used in Laville et al., 2020 [31] based on FoldX suite [32] and Gromacs software [66]. This two-step protocol was applied to the wild-type crystallographic structure of the PR2 complexed with DRV (PDB code: 3EBZ [33]). First, the DRV ligand, metal atoms, and water molecules were removed from the crystallographic structure of PR2. Then, the RepairPDB command of the FoldX suite [32] was applied to the PR2 structure to reduce its energy. The *in silico* mutagenesis was then performed using the BuildModel command of the FoldX suite based on a side-chain rotamer library [32]. This step was performed five times to consider the several rotamers available for each amino acid [32] and generated five structures per mutant. An energetic minimization was then applied to the five modeled structures of a mutant using the protocol developed in our previous study [31]. We applied PROPKA software [67] to monoprotonate the oxygen atom OD2 of Aspartate 25 in chain B. Then, the system was solvated in a truncated octahedron box of explicit solvent (TIP3P water model) with a 12.0 Å buffer in each dimension and its charge was neutralized using chloride ions. The minimization of the system was performed using a two-step protocol using the force field AMBER ff99SB in GROMACS [66] by applying a steepest descent algorithm

combined with a conjugate gradient algorithm. A first step energy minimization allowed relaxing water molecules and counterions using a position harmonic restraining force of $100 \text{ kcal}\cdot\text{mol}^{-1}\text{\AA}^{-2}$ on the heavy atoms of the protein. Then, restraints on protein atoms were removed using a second energy minimization step. The protocol used the particle mesh Ewald (PME) method to treat the long-range electrostatic interactions [68] and a cutoff distances of 10.0 \AA for the long-range electrostatic and van der Waals interactions. This protocol was applied to the 31 mutants generating 155 mutant structures (5 structures per mutant). The minimization protocol was also applied to the wild-type structure and the minimized wild-type structure was named wild-type structure.

4.3. Identification of Shifted Atoms Induced by Mutations.

First, all shifted atoms in each mutant structure were extracted by comparing the position of atoms in the wild-type and mutant structures. To do so, we applied the method used in Perrier et al., 2019 [69]. Each mutant structure was superimposed onto the wild-type structure (the minimized wild-type structure) using PyMOL software [70]. Superimposition was based on all atoms. Hydrogen atoms were removed. Euclidean distances between the position of each atom in the mutant and wild-type structure were computed. These distances were denoted as $dist^{WT-mutant}$. Higher a $dist^{WT-mutant}$ value of an atom is, more the atom is shifted in the mutated structure relative to the wild-type structure. According to the modelization protocol, detected atom shifts resulted from the mutagenesis and minimization. To retain only significant structural rearrangements induced by resistance mutation, only atoms with a $dist^{WT-mutant}$ value higher than 0.3 \AA were retained like in Liu et al., 2008 [36]. This distance cutoff of 0.3 \AA allowed selecting only significant structural displacements and removing uncertainties in the X-ray and mutant structures [36]. In addition, for each mutant, only shifted atoms observed in three of the five structures of the mutant were retained. These shifted atoms were named mutant-conserved shifted and noted MCS atoms.

4.4. Comparison of Wild-Type and Mutant Pockets

4.4.1. Pocket Estimation

From the 156 structures, we extracted the binding pocket. To consider the fact that the different known PI are structurally diverse and that resistance mutations induce resistance to one or several PI, we used the “common-ligand” approach to estimate pockets [11] that consisted to define the binding pocket as all atoms of PR2 capable to bind all co-crystallized ligands. To do so, a virtual ligand was built by superimposed all co-crystallized ligands extracted from a set of PRs. This virtual ligand was then placed in the query structure and its pocket was estimated as atoms located at least 4.5 \AA of the virtual ligand. This protocol was applied on the 156 wild-type and mutant structures using the virtual ligand built from the PR set used in Triki et al., 2018 [11]. The pocket extracted from the wild-type structure was denoted as wild-type and those extracted from mutant structures were denoted as mutant pockets.

4.4.2. Comparison of the Conformation of Wild-Type and Mutant Pockets

To compare the conformation of the 156 pockets, we computed the root mean square deviation (RMSD) between each pocket pair (156×156). The calculation of RMSD was performed using PyMOL software [70] based on all pocket atoms. From these RMSD values, we computed a hierarchical classification of the pockets using the Ward method aggregation.

4.4.3. Comparison of the Properties of Wild-Type and Mutant Pockets

Each pocket was characterized by two geometrical descriptors (VOLUME_HULL and PSI), and one physicochemical descriptor (hydrophobicity_kyte) [71]. The VOLUME_HULL descriptor provides an estimation of the volume of a pocket. PSI measures the sphericity of a

pocket. These two descriptors were computed using PCI software [72]. hydrophobicity_kyte descriptor quantifies the hydrophobicity of a pocket.

4.5. Interface Comparison

4.5.1. Interface Extraction

The PPIC (Protein-Protein Interface Computation) program [73,74] was used to determine atoms involved in the interface of a structure. This program is parameter-free. It takes in input the 3D structure of a complex with two molecules (molecules or macromolecules) A and B. It defines the interface between the two molecules in two parts: interface of A and interface of B. Each interface part corresponds to the non-redundant set of all nearest neighbor atoms in one molecule of the atoms of other molecules. The extraction of neighbor atoms is performed using a simpler method of the Voronoi tessellation method [75]. Contrary to the Voronoi tessellation method, PPIC approach does not generate neighbors at long distances in the interface.

We used PPIC program to extract interface from the wild-type structure, denoted as wild-type interface, and the interface from the mutant structures, denoted as mutant interface.

4.5.2. Comparison of the Interface Composition

We compared the composition of the 156 interfaces by computed the Hamming distance, noted $D_{interface}$ between each interface pairs (Equation (1)).

$$D_{interface}(I1, I2) = N_{I1} + N_{I2} - 2 \times N_{I12} \quad (1)$$

with $I1$ and $I2$ the interfaces extracted from two structures, N_{I1} and N_{I2} the number of atoms of the interfaces $I1$ and $I2$, respectively, and N_{I12} the number of common atoms of the two interfaces $I1$ and $I2$.

Higher $D_{interface}(I1, I2)$ is, more dissimilar the two interfaces $I1$ and $I2$ are. To facilitate the interpretation of these distances, each $D_{interface}(I1, I2)$ value was normalized by the maximum number of atoms in the two interfaces ($N_{I1} + N_{I2}$) using Equation (2), resulting in the computation of the $D_{interface}(I1, I2)^{norm}$ value for each interface pair.

$$D_{interface}(I1, I2)^{norm} = \frac{D_{interface}(I1, I2)}{N_{I1} + N_{I2}}. \quad (2)$$

The $D_{interface}(I1, I2)^{norm}$ is ranking from 0 to 1, with a $D_{interface}(I1, I2)^{norm}$ equal to 0 means that the composition of the two interfaces $I1$ and $I2$ is identical.

Using the $D_{interface}(I1, I2)^{norm}$ values of each interface pairs, we computed a hierarchical classification of the 156 structures allowing to group structures according to their similarity in terms of interface composition. This classification was computed using the Ward method aggregation.

4.5.3. Comparison of the SASA of Interface

The SASA of the interface of each structure was determined using NACCESS software [37]. First the two monomers of each structure were separated. Then, the accessible surface area of all atoms of the 312 monomers was computed using NACCESS software [37]. The SASA values of the two parts of the interface (chains A and B) of a structure were obtained by computed the sum of the SASA of each atom detected as involved in the interface. The SASA value reflects the size of a interface.

5. Conclusions

In this study, we explored the impact of drug-resistance mutations reported in PR2. We first compared the modeled structure of 31 mutants with the wild-type PR2 structure to locate structural rearrangements induced by mutations. Secondly, we studied the impact of these deformations on the conformation and properties of the binding pocket and

on the interface. Our findings showed that one mutation could induce large structural deformations located in the mutated residue, in its vicinity or far in the structure. These structural deformations occur in both side-chain and backbone atoms, with on average more impact in the former. However, we revealed that resistance mutations do not always have the same behavior in the two monomers of PR2, which is linked to the structural asymmetry of this target.

The analysis of the location of structural rearrangements induced by resistance mutations provided clues to better understand resistance mechanisms. First, some mutations have a direct or indirect impact on PI-binding. The K7R, V47A, G48V, I50L/V, I54L/M, V62A, I82F/M, I82M, I84V, and L90M induce structural rearrangements in the binding pocket that modify its conformation, volume and/or hydrophobicity. These changes in the binding pocket could have a negative effect on PI binding. In addition, some of these mutations (K7R, D30N, I50L/V, I54L, I82F, and I84V) have a direct impact on PI binding by causing structural displacements of residues establishing interactions with PI. Resistance mutations have also an impact on conformation of the elbow and flap regions, regions involved in the transition of the open and closed conformations of the PR2 upon ligand binding. Indeed, we reported that the K7R, E37D, S43T, K45R, V47A, G48V, I50V, I50L, I54M, I54L, and V62A led to atom shifts in the elbow and flap regions that could modify the flexibility and movement of the flap region important in the binding-pocket closing. Thirdly, the K7R, G48V, I50L, I54L, L90M, and L99F mutations induce structural rearrangements in the PR2 interface that modify its composition and size that may alter the stability of PR2. Finally, we highlighted several residues that were never deformed in mutant structures.

In conclusion, this study explored the impact of a large number of PR2 resistance mutations on PR2 structure, particularly on its pocket binding and interface. Our results showed that some structural rearrangements induced by resistance mutations are located in important regions of PR2: the elbow and flap regions inducing PR2 deformation upon ligand binding, in the PI-binding pocket and in the interface of the two monomers. Some of these deformations modify the properties of the binding pocket and the composition and size of the PR2 interface. These results suggest that the studied resistance mutations could alter PI binding by modifying the properties and flexibility of the pocket or the interaction between PR2 and PI and/or alter PR2 stability. Our results reinforce the resistance mechanisms proposed in our previous study [31] and lead to a better understanding of the effects of mutations that occurred in PR2 and the different mechanisms of PR2 resistance.

Author Contributions: L.R. conceived and designed the experiments; M.P. and L.R. supervised the project; P.L. and L.R. performed the experiments; P.L., M.P. and L.R. developed tools dedicated to analyses; P.L. and L.R. analyzed the data; L.R. wrote the paper. All authors reviewed the manuscript. All authors have read and agreed to the submitted version of the manuscript.

Funding: This work was supported by an ANRS Grant. P.L. is supported by a fellowship from the Ministère de l'Éducation Nationale de la Recherche et de Technologie (MENRT).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are openly available in https://figshare.com/articles/dataset/Data_of_Laville_et_al_2021/13634147.

Acknowledgments: We are grateful to D. Flatters for helpful discussions and O. Taboureau for proofreading the manuscript. L.R. thanks the University of Paris and l'UFR SdV.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

PR	Protease
PR2	HIV-2 Protease
PI	Protease inhibitor
MCS atoms	Mutant-Conserved Shifted atoms
RMSD	Root Mean Square Deviation
DRV	Darunavir
LPV	Lopinavir
SQV	Saquinavir
IDV	Indinavir
3D	Three-dimensional
SASA	Solvent Accessible Surface Area
PPIC	Protein-Protein Interface Computation

Appendix A

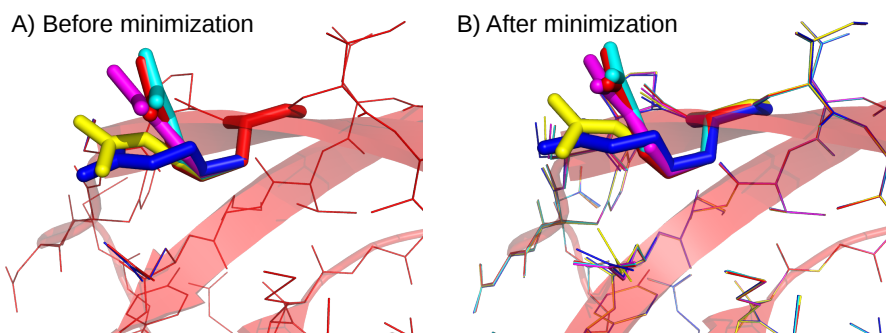


Figure A1. Illustration of the different rotamers for the arginine in position 45. Superimposition of the five structures of the K45R mutant before (A) and after the minimization process (B). Mutant structures are presented in line and cartoon mode. Residues R45 are presented in sticks.

Appendix B

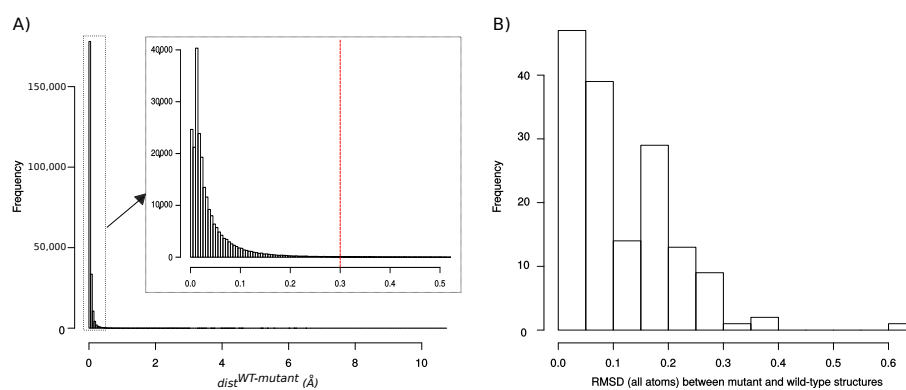


Figure A2. (A) Distribution of $dist^{WT-mutant}$ distances for each atom (hydrogen atoms were removed) in the set of 155 mutant structures. Red line corresponds the distance cutoff (0.3 Å) used to define an atom as shifted atom. (B) Distribution of RMSD (in Å) computed between wild-type and mutant structures. RMSD were computed using all atoms.

Appendix C

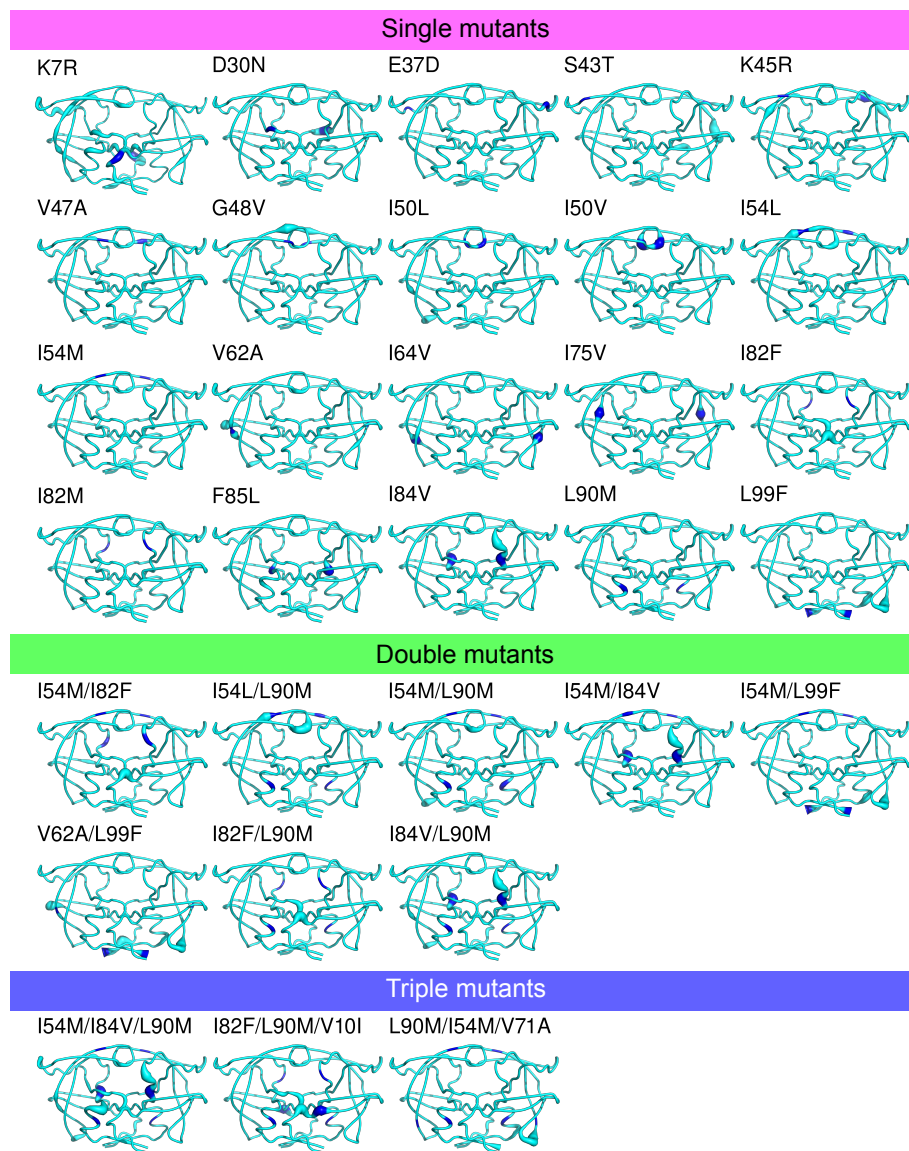


Figure A3. Location of the MCS atoms found in each mutant. PR2 is colored in cyan and represented in putty mode. The putty radius is relative to the deformation induced by mutations: the higher the radius, the stronger the mutation-induced rearrangement. Mutated residues are colored in blue.

Appendix D

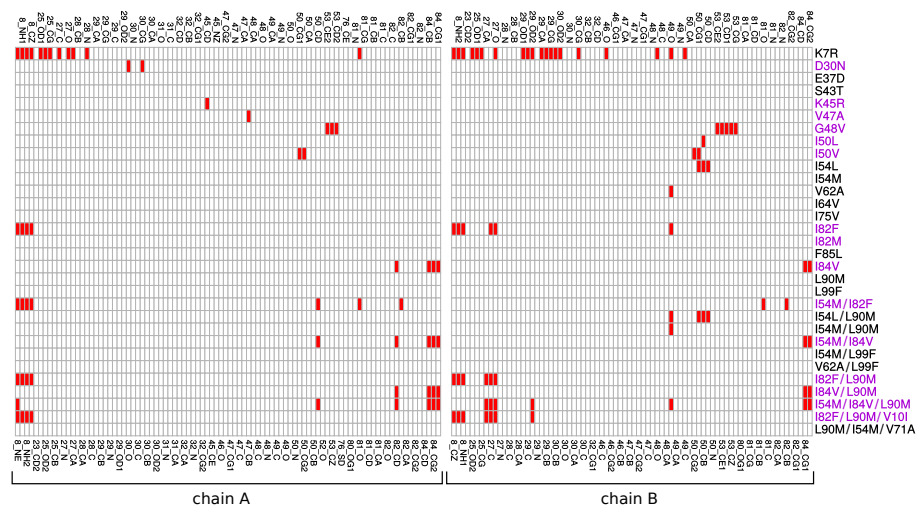


Figure A4. MCS atoms in the pocket. Mutants colored in purple present at least a mutation located in the binding pocket.

Appendix E

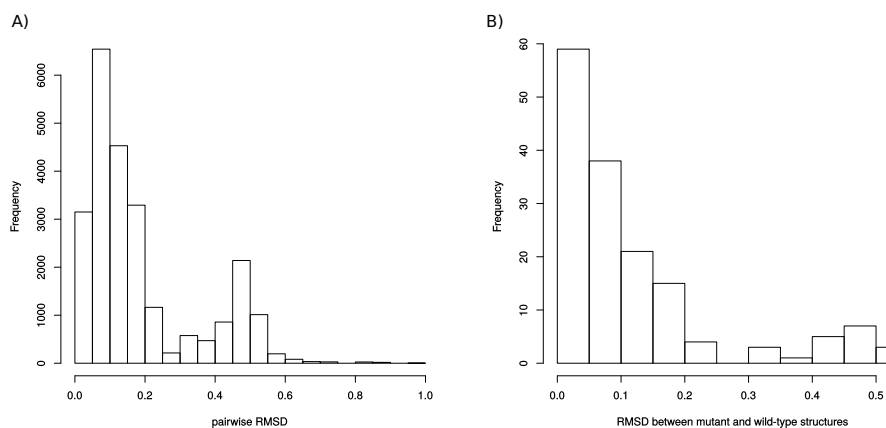


Figure A5. Comparison between pockets. (A) Distribution of RMSD computed between all pocket pairs. (B) Distribution of RMSD computed between the wild-type and mutant pockets.

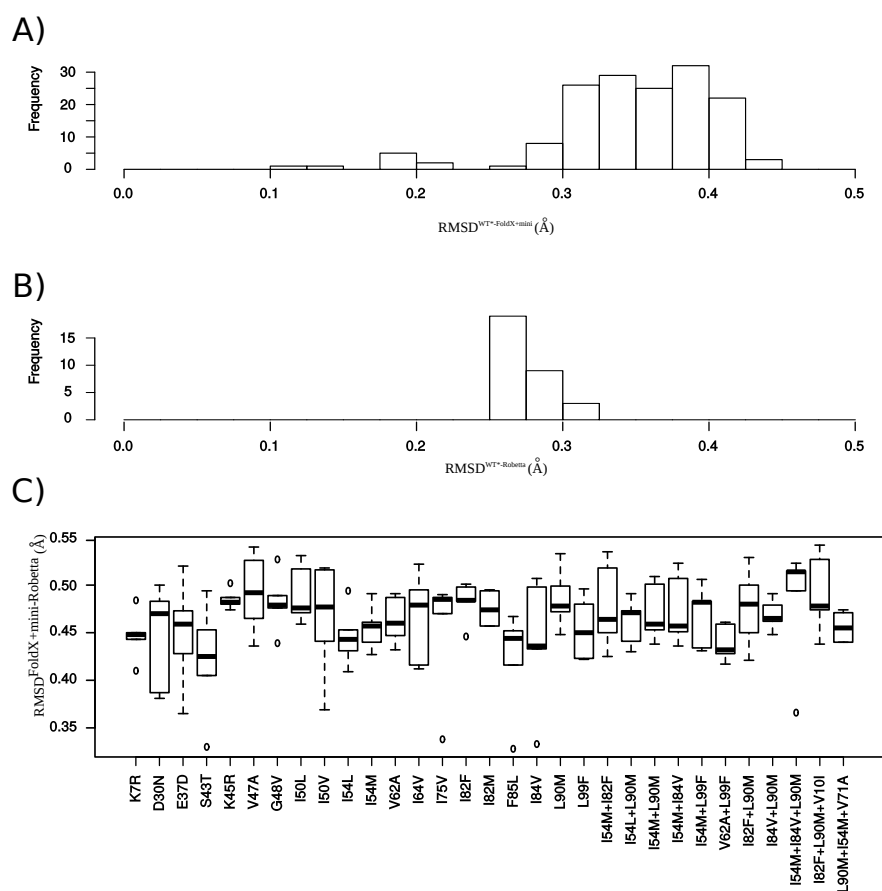


Figure A7. Comparison between $mutant^{FoldX+Mini}$ and $mutant^{Robetta}$ structures. (A) RMSD computed between the wild-type structure (not minimized, WT^* structure) and $mutant^{FoldX+Mini}$ structure. (B) RMSD computed between the wild-type structure (not minimized, WT^* structure) and $mutant^{Robetta}$ structure. (C) RMSD computed between each $mutant^{Robetta}$ structure and the five structure of $mutant^{FoldX+Mini}$.

Appendix G.2. Structural Comparison between the $Mutant^{FoldX+Mini}$ and $Mutant^{Robetta}$ Structures

First, we compared the two sets of PR2 mutants. For each mutant, its $mutant^{Robetta}$ structure was superimposed onto the five $mutant^{FoldX+Mini}$ mutants and the RMSD based on all atoms was computed using PyMol software [70]. These RMSD were denoted as $RMSD^{FoldX+Mini-Robetta}$.

Appendix G.3. Detection of Structural Rearrangements in the Set of $Mutant^{Robetta}$ Structures

First, each $mutant^{Robetta}$ structure were superimposed onto the 3EBZ (PDB code) structure using PyMol software [70]. Superimposition was based on all atoms. Euclidean distances between the position of each atom in the mutant and 3EBZ structure were computed. These distances were denoted as $dist^{WT^*-mutant^{Robetta}}$. An atom was considered as a MCS atom in a $mutant^{Robetta}$ structure if it had a $dist^{WT^*-mutant^{Robetta}}$ value higher than 0.3 Å.

Appendix H

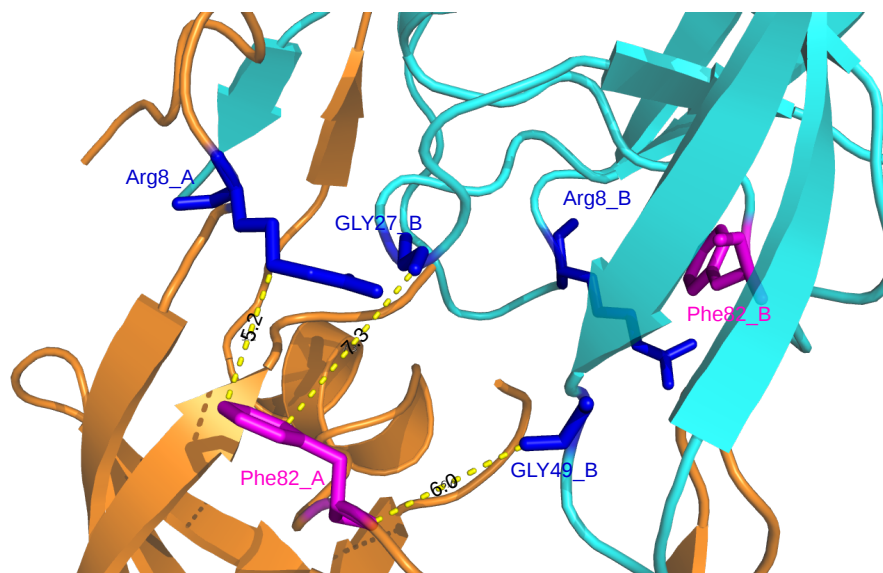


Figure A8. Location of residues exhibiting structural shifts in the I82F mutant. Chains A and B are presented in cartoon mode and colored in orange and blue, respectively. Mutated residues (82) are colored in magenta and presented in stick mode. Residues having shifted atoms are colored in blue and presented in stick mode.

References

1. 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. [\[CrossRef\]](#)
2. 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. [\[CrossRef\]](#) [\[PubMed\]](#)
3. Raugi, D.N.; Smith, R.A.; Gottlieb, G.S.; the University of Washington-Dakar HIV-2 Study Group. Four Amino Acid Changes in HIV-2 Protease Confer Class-Wide Sensitivity to Protease Inhibitors. *J. Virol.* **2016**, *90*, 1062–1069. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Desbois, D.; Roquebert, B.; Peytavin, G.; Damond, F.; Collin, G.; Bénard, A.; Campa, P.; Matheron, S.; Chêne, G.; Brun-Vézinet, F.; et al. In vitro phenotypic susceptibility of human immunodeficiency virus type 2 clinical isolates to protease inhibitors. *Antimicrob. Agents Chemother.* **2008**, *52*, 1545–1548. [\[CrossRef\]](#) [\[PubMed\]](#)
5. 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. [\[CrossRef\]](#)
6. Cavaco-Silva, J.; Aleixo, M.; Van Laethem, K.; Faria, D.; Valadas, E.; Gonçalves, M.D.F.; 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. [\[CrossRef\]](#)
7. Bénard, A.; Damond, F.; Campa, P.; Peytavin, G.; Descamps, D.; Lascoux-Combes, C.; Taieb, A.; Simon, F.; Autran, B.; Brun-Vézinet, F.; et al. Good response to lopinavir/ritonavir-containing antiretroviral regimens in antiretroviral-naïve HIV-2-infected patients. *AIDS* **2009**, *23*, 1171–1183. [\[CrossRef\]](#) [\[PubMed\]](#)
8. 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. [\[CrossRef\]](#)
9. 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. [\[CrossRef\]](#)
10. Menéndez-Arias, L.; Alvarez, M. Antiretroviral therapy and drug resistance in human immunodeficiency virus type 2 infection. *Antivir. Res.* **2014**, *102*, 70–86. [\[CrossRef\]](#)
11. 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. [\[CrossRef\]](#)

12. Ollitrault, G.; Fartek, S.; Descamps, D.; Camproux, A.; Visseaux, B.; Regad, L. Characterization of HIV-2 protease structure by studying its asymmetry at the different levels of protein description. *Symmetry* **2019**, *10*, 644. [[CrossRef](#)]
13. Triki, D.; Kermarrec, M.; Visseaux, B.; Descamps, D.; Flatters, D.; Camproux, A.; Regad, L. Exploration of the effects of sequence variations between HIV-1 and HIV-2 proteases on their three-dimensional structures. *J. Biomol. Struct. Dyn.* **2020**, *38*, 5014–5026. [[CrossRef](#)] [[PubMed](#)]
14. Gustchina, A.; Weber, I. Comparative analysis of the sequences and structures of HIV-1 and HIV-2 proteases. *Proteins Struct. Funct. Bioinform.* **1991**, *10*, 325–339. [[CrossRef](#)] [[PubMed](#)]
15. Sardana, V.; Schlabach, A.; Graham, P.; Bush, B.; Condra, J.; Culberson, J.; Gotlib, L.; Graham, D.E.A. Human Immunodeficiency Virus Type 1 Protease Inhibitors: Evaluation of Resistance Engendered by Amino Acid Substitutions in the Enzyme's Substrate Binding Site. *Biochemistry* **1994**, *33*, 2004–2010. [[CrossRef](#)]
16. Hoog, S.; Towler, E.; Zhao, B.; Doyle, M.; Debouck, C.; Abdel-Meguid, S. Human immunodeficiency virus protease ligand specificity conferred by residues outside of the active site cavity. *Biochemistry* **1996**, *35*, 10279–10286. [[CrossRef](#)]
17. Tie, Y.; Boross, P.; Wang, Y.; Gaddis, L.; Hussain, A.; Leshchenko, S.; Ghosh, A.; Louis, J.; Harrison, R.; Weber, I. High resolution crystal structures of HIV-1 protease with a potent non-peptide inhibitor (UIC-94017) active against multi-drug-resistant clinical strains. *J. Mol. Biol.* **2004**, *23*, 341–352. [[CrossRef](#)]
18. Kovalevsky, A.; Louis, J.; Aniana, A.; Ghosh, A.; Weber, I. Structural evidence for effectiveness of darunavir and two related antiviral inhibitors against HIV-2 protease. *J. Mol. Biol.* **2008**, *384*, 178–192. [[CrossRef](#)]
19. Triki, D.; Billot, T.; Flatters, D.; Visseaux, B.; Descamps, D.; Camproux, A.; Regad, L. Exploration of the effect of sequence variations located inside the binding pocket of HIV-1 and HIV-2 proteases. *Sci. Rep.* **2018**, *8*, 5789. [[CrossRef](#)]
20. Hightower, M.; Kallas, E.G. Diagnosis, antiretroviral therapy, and emergence of resistance to antiretroviral agents in HIV-2 infection: A review. *Braz. J. Infect. Dis.* **2003**, *7*, 7–15. [[CrossRef](#)]
21. 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](#)]
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](#)]
23. Colson, P.; Henry, M.; Tourres, C.; Lozachmeur, D.; Gallais, H.; Gastaut, J.; Moreau, J.; Tamalet, C. Polymorphism and drug-selected mutations in the protease gene of human immunodeficiency virus type 2 from patients living in southern France. *J. Clin. Microbiol.* **2004**, *42*, 570–577. [[CrossRef](#)] [[PubMed](#)]
24. Storto, A.; Visseaux, B.; Bertine, M.; Le Hingrat, Q.; Collin, G.; Damond, F.; Khuong, M.; Blum, L.; Tubiana, R.; Karmochkine, M.; et al. ANRS HIV-2 CO5 cohort. Minority resistant variants are also present in hiv-2-infected antiretroviral-naive patients. *J. Antimicrob. Chemother.* **2018**, *73*, 1173–1176. [[CrossRef](#)] [[PubMed](#)]
25. Damond, F.; Brun-Vezinet, F.; Matheron, S.; Peytavin, G.; Campa, P.; Pueyo, S.; Mammano, F.; Lastere, S.; Farfara, I.; Simon, F.; et al. 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. *J. Clin. Microbiol.* **2005**, *43*, 484–487. [[CrossRef](#)]
26. Ntemgwa, M.; Brenner, B.; Oliveira, M.; Moisi, D.; Wainberg, M. Variations in reverse transcriptase and RNase H domain mutations in human immunodeficiency virus type 1 clinical isolates are associated with divergent phenotypic resistance to zidovudine. *Antimicrob. Agents Chemother.* **2007**, *51*, 604–610. [[CrossRef](#)] [[PubMed](#)]
27. Larrouy, L.; Vivot, A.; Charpentier, C.; Bénard, A.; Visseaux, B.; Damond, F.; Matheron, S.; Chene, G.; Brun-Vezinet, F.; Descamps, D. ANRS CO5 HIV-2 Cohort. Impact of gag genetic determinants on virological outcome to boosted lopinavir-containing regimen in hiv-2-infected patients. *AIDS* **2013**, *27*, 69–80. [[CrossRef](#)]
28. Jallow, S.; Alabi, A.; Sarge-Njie, R.; Peterson, K.; Whittle, H.; Corrah, T.; Jaye, A.; Cotten, M.; Vanham, G.; McConkey, S.; et al. Virological response to highly active antiretroviral therapy in patients infected with human immunodeficiency virus type 2 (hiv-2) and in patients dually infected with hiv-1 and hiv-2 in the Gambia and emergence of drug-resistant variants. *J. Clin. Microbiol.* **2009**, *47*, 2200–2208. [[CrossRef](#)]
29. Mahdi, M.; Szojka, Z.; Mótyán, J.; Tözsér, J. Inhibition Profiling of Retroviral Protease Inhibitors Using an HIV-2 Modular System. *Viruses* **2015**, *71*, 6152–6162. [[CrossRef](#)]
30. Charpentier, C.; Visseaux, B.; Bénard, A.; Peytavin, G.; Damond, F.; Roy, C.; Taieb, A.; Chêne, G.; Matheron, S.; Brun-Vézinet, F.; et al. Transmitted drug resistance in French HIV-2-infected patients. *AIDS* **2013**, *27*, 1671–1684. [[CrossRef](#)]
31. Laville, P.; Fartek, S.; Cerisier, N.; Flatters, D.; Petitjean, M.; Regad, L. Impacts of drug resistance mutations on the structural asymmetry of the HIV-2 protease. *BMC Mol. Cell Biol.* **2020**, *51*, 46. [[CrossRef](#)] [[PubMed](#)]
32. Guerois, R.; Nielsen, J.E.; Serrano, L. Predicting changes in the stability of proteins and protein complexes: A study of more than 1000 mutations. *J. Mol. Biol.* **2002**, *320*, 369–387. [[CrossRef](#)]
33. Tie, Y.; Wang, Y.; Boross, P.; Chiu, T.; Ghosh, A.; Tozser, J.; Louis, J.; Harrison, R.; Weber, I. Critical differences in HIV-1 and HIV-2 protease specificity for clinical inhibitors. *Protein Sci.* **2012**, *21*, 339–350. [[CrossRef](#)] [[PubMed](#)]
34. 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](#)]
35. Triki, D.; Fartek, S.; Visseaux, B.; Descamps, D.; Camproux, A.; Regad, L. Characterizing the structural variability of HIV-2 protease upon the binding of diverse ligands using a structural alphabet approach. *J. Biomol. Struct. Dyn.* **2018**, *28*, 1–13. [[CrossRef](#)] [[PubMed](#)]

36. Liu, F.; Kovalevsky, A.; Tie, Y.; Ghosh, A.; Harrison, R.; Weber, I. Effect of flap mutations on structure of HIV-1 protease and inhibition by saquinavir and darunavir. *J. Mol. Biol.* **2008**, *381*, 102–115. [[CrossRef](#)]
37. Hubbard, S.; Thornton, J. NACCESS. In *Computer Program, Department of Biochemistry and Molecular Biology*; University College London: London, UK, 1993.
38. Koh, Y.; Nakata, H.; Maeda, K.; Ogata, H.; Bilcer, G.; Devasamudram, T.; Kincaid, J.; Boross, P.; Wang, Y.F.; Tie, Y.; et al. Novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (PI) UIC-94017 (TMC114) with potent activity against multi-PI-resistant human immunodeficiency virus in vitro. *Antimicrob. Agents Chemother.* **2003**, *47*, 3123–3129. [[CrossRef](#)]
39. Ghosh, A.K.; Dawson, Z.L.; Mitsuya, H. Darunavir, a conceptually new HIV-1 protease inhibitor for the treatment of drug-resistant HIV. *Bioorg. Med. Chem.* **2007**, *15*, 7576–7580. [[CrossRef](#)]
40. Tong, L.; Pav, S.; Pargellis, C.; Dó, 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](#)]
41. 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. [[CrossRef](#)]
42. 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](#)]
43. Priestle, J.; Fassler, 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](#)]
44. Kovalevsky, A.; Liu, F.; Leshchenko, S.; Ghosh, A.K.; Louis, J.; Harrison, R.; Weber, I. Ultra-high resolution crystal structure of HIV-1 protease mutant reveals two binding sites for clinical inhibitor TMC114. *J. Mol. Biol.* **2006**, *363*, 161–173. [[CrossRef](#)]
45. Zhang, H.; Wang, Y.; Shen, C.H.; Agniswamy, J.; Rao, K.; Xu, C.; Ghosh, A.; Harrison, R.; Weber, I. Novel P2 tris-tetrahydrofuran group in antiviral compound 1 (GRL-0519) fills the S2 binding pocket of selected mutants of HIV-1 protease. *J. Med. Chem.* **2013**, *56*, 1074–1083. [[CrossRef](#)] [[PubMed](#)]
46. Shen, C.; Wang, Y.; Kovalevsky, A.; Harrison, R.; Weber, I. Amprenavir complexes with HIV-1 protease and its drug-resistant mutants altering hydrophobic clusters. *FEBS* **2010**, *277*, 3699–3714. [[CrossRef](#)] [[PubMed](#)]
47. Scott, W.; Schiffer, C. Curling of flap tips in HIV-1 protease as a mechanism for substrate entry and tolerance of drug resistance. *Structure* **2000**, *8*, 1259–1265. [[CrossRef](#)]
48. Hong, L.; Zhang, X.; Hartsuck, J.; Tang, J. Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: Insights into the mechanisms of drug resistance. *Protein Sci.* **2000**, *9*, 1898–1904. [[CrossRef](#)]
49. Agniswamy, J.; Louis, J.; Roche, J.; Harrison, R.; Weber, I. Structural Studies of a Rationally Selected Multi-Drug Resistant HIV-1 Protease Reveal Synergistic Effect of Distal Mutations on Flap Dynamics. *PLoS ONE* **2016**, *11*, e0168616. [[CrossRef](#)] [[PubMed](#)]
50. Cai, Y.; Yilmaz, N.; Myint, W.; Ishima, R.; Schiffer, C. Differential Flap Dynamics in Wild-Type and a Drug Resistant Variant of HIV-1 Protease Revealed by Molecular Dynamics and NMR Relaxation. *J. Chem. Theory Comput.* **2012**, *8*, 3452–3462. [[CrossRef](#)]
51. Zhang, Y.; Chang, Y.C.E.; Louis, J.; Wang, Y.; Harrison, Y.; Weber, I. Structures of darunavir-resistant HIV-1 protease mutant reveal atypical binding of darunavir to wide open flaps. *ACS Chem. Biol.* **2014**, *9*, 1351–1358. [[CrossRef](#)]
52. Logsdon, B.; Vickrey, J.; Martin, P.; Proteasa, G.; Koepke, J.; Terlecky, S.R.; Wawrzak, Z.; Winters, M.A.; Merigan, T.C.; Kovari, L.C. Crystal structures of a multidrug-resistant human immunodeficiency virus type 1 protease reveal an expanded active-site cavity. *J. Virol.* **2004**, *78*, 3123–3132. [[CrossRef](#)] [[PubMed](#)]
53. Shen, C.; Chang, Y.; Agniswamy, J.; Harrison, R.; Weber, I. Conformational variation of an extreme drug resistant mutant of HIV protease. *J. Mol. Graph. Model.* **2015**, *62*, 87–96. [[CrossRef](#)] [[PubMed](#)]
54. Muzammil, S.; Armstrong, A.; Kang, L.; Jakalian, A.; Bonneau, P.; Schmelmer, V.; Amzel, L.; Freire, E. Unique thermodynamic response of tipranavir to human immunodeficiency virus type 1 protease drug resistance mutations. *J. Virol.* **2007**, *81*, 5144–5154. [[CrossRef](#)] [[PubMed](#)]
55. Yeddi, R.; Proteasa, G.; Martinez, J.; Vickrey, J.; Martin, P.; Wawrzak, Z.; Liu, Z.; Kovari, I.; Kovaria, L. Contribution of the 80's loop of HIV-1 protease to the multidrug-resistant mechanism: crystallographic study of MDR769 HIV-1 protease variants. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67*, 524–532. [[CrossRef](#)] [[PubMed](#)]
56. Martin, P.; Vickrey, J.; Proteasa, G.; Jimenez Y.L. Wawrzak, Z.; Winters, M.; Merigan, T.; Kovari, L. Wide-open 1.3 Å structure of a multidrug-resistant HIV-1 protease as a drug target. *Structure* **2005**, *13*, 1887–1895. [[CrossRef](#)] [[PubMed](#)]
57. Kovalevsky, A.; Tie, Y.; Liu, F.; Boross, P.; Wang, Y.; Leshchenko, S.; Ghosh, A.; Harrison, R.; Weber, I.T. Effectiveness of Nonpeptide Clinical Inhibitor TMC-114 on HIV-1 Protease with Highly Drug Resistant Mutations D30N, I50V, and L90M. *J. Med. Chem.* **2006**, *49*, 1379–1387. [[CrossRef](#)] [[PubMed](#)]
58. 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](#)] [[PubMed](#)]
59. Weber, I.; Agniswamy, J. HIV-1 Protease: Structural Perspectives on Drug Resistance. *Viruses* **2009**, *1*, 1110–1136. [[CrossRef](#)] [[PubMed](#)]
60. Meher, B.; Wang, Y. Interaction of I50V mutant and I50L/A71V double mutant HIV-protease with inhibitor TMC114 (darunavir): molecular dynamics simulation and binding free energy studies. *J. Phys. Chem.* **2012**, *116*, 1884–1900. [[CrossRef](#)] [[PubMed](#)]

61. Mittal, S.; Bandaranayake, R.; King, N.; Prabu-Jeyabalan, M.; Nalam, M.; Nalivaika, E.; Yilmaz, N.; Schiffer, C. Structural and thermodynamic basis of amprenavir/darunavir and atazanavir resistance in HIV-1 protease with mutations at residue 50. *J. Virol.* **2013**, *87*, 4176–4184. [[CrossRef](#)]
62. Lockbaum, G.; Leidner, F.; Rusere, L.; Henes, M.; Kosovrasti, K.; Nachum, G.; Nalivaika, E.; Ali, A.; Yilmaz, N.; Schiffer, C. Structural Adaptation of Darunavir Analogues against Primary Mutations in HIV-1 Protease. *ACS Infect. Dis.* **2019**, *5*, 316–325. [[CrossRef](#)] [[PubMed](#)]
63. Mahalingam, B.; Louis, J.; Reed, C.; Adomat, J.; Krouse, J.; Wang, Y.; Harrison, R.; Weber, I. Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. *Eur. J. Biochem.* **1999**, *263*, 238–245. [[CrossRef](#)] [[PubMed](#)]
64. Mahalingam, B.; Wang, Y.; Boross, P.; Tozser, J.; Louis, J.; Harrison, R.; Weber, I. Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site. *Eur. J. Biochem.* **2004**, *271*, 1516–1524. [[CrossRef](#)] [[PubMed](#)]
65. Xie, D.; Gulnik, S.; Gustchina, E.; Yu, B.; Shao, W.; Qoronfleh, W.; Nathan, A.; Erickson, J. Drug resistance mutations can effect dimer stability of HIV-1 protease at neutral pH. *Protein Sci.* **1999**, *8*, 1702–1707. [[CrossRef](#)]
66. Abraham, M.; Murtolad, T.; Schulz, R.; Palla, 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](#)]
67. Li, H.; Robertson, A.; Jensen, J.H. Very fast empirical prediction and rationalization of protein pK a values. *Proteins* **2005**, *61*, 704–721. [[CrossRef](#)]
68. 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](#)]
69. Perrier, M.; Castain, L.; Regad, L.; Todesco, E.; Landman, R.; Visseaux, B.; Yazdanpanah, Y.; Rodriguez, C.; Joly, V.; Calvez, V.; et al. HIV-1 protease, Gag and gp41 baseline substitutions associated with virological response to a PI-based regimen. *J. Antimicrob. Chemother.* **2019**, *74*, 1679–1692. [[CrossRef](#)]
70. The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC. Available online: <https://pymol.org/2/> (accessed on 20 January 2021).
71. Borrel, A.; Regad, L.; Xhaard, H.; Petitjean, M.; Camproux, A. PockDrug: A Model for Predicting Pocket Druggability That Overcomes Pocket Estimation Uncertainties. *J. Chem. Inf. Model.* **2015**, *55*, 882–895. [[CrossRef](#)]
72. Cerisier, N.; Regad, L.; Triki, D.; Camproux, A.; Petitjean, M. Cavity versus ligand shape descriptors: Application to urokinase binding pockets. *J. Comput. Biol.* **2017**, *24*, 1134–1137. [[CrossRef](#)]
73. Ozeel, V.; Perrier, A.; Vanet, A.; Petitjean, M. The Symmetric Difference Distance: A New Way to Evaluate the Evolution of Interfaces along Molecular Dynamics Trajectories; Application to Influenza Hemagglutinin. *Symmetry* **2019**, *11*, 662. [[CrossRef](#)]
74. Laville, P.; Martin, J.; Launay, G.; Regad, L.; Camproux, A.; de Vries, S.; Petitjean, M. A non-parametric method to compute protein–protein and protein–ligands interfaces. Application to HIV-2 protease–inhibitors complexes. *bioRxiv* **2018**. [[CrossRef](#)]
75. Eppstein, D.; Paterson, M.; Yao, F. On nearest-neighbor graphs. *Discret. Comput. Geom.* **1997**, *17*, 263–282. [[CrossRef](#)]
76. Song, Y.; DiMaio, F.; Wang, R.Y.R.; Kim, D.; Miles, C.; Brunette, T.; Thompson, J.; Baker, D. High-Resolution Comparative Modeling with RosettaCM. *Structure* **2013**, *21*, 1735–1742. [[CrossRef](#)]

4 Étude de la dynamique des PR2 sauvage et mutantes

Cette partie présente des résultats préliminaires non publiés sur l'étude des structures de la PR2 en formes semi-ouverte et fermée ainsi que des structures de la PR2 avec les mutations I54M, I82F et I54M/I82F.

4.1 Choix des mutations

Nos travaux précédents s'intéressent aux modifications structurales engendrées par les mutations sur des structures modélisées et minimisées. La prise en compte de la flexibilité de la protéine y est donc assez limitée. C'est pourquoi nous avons décidé d'approfondir ces analyses à l'aide de simulations de dynamique moléculaire. Cette méthode est néanmoins beaucoup plus coûteuse en temps de calcul que les approches précédentes. Ainsi, il a fallu réduire le nombre de mutants étudiés. Notre choix s'est porté sur les trois mutants suivants : I54M, I82F et I54M/I82F, pour différentes raisons. Ce sont des mutations fréquemment étudiées dans la littérature. La mutation I54M entraîne une résistance à tous les IP à l'exception du SQV pour lequel il est hypersensible (figure 4.1). À l'inverse, le mutant I82F est résistant à l'IDV, NFV, et LPV et hypersensible aux autres IP. Il est intéressant de noter que [Raugi et al., 2013](#) ont quantifié la susceptibilité de ces mutations face à des IP, notamment face au DRV. Ils montrent que le mutant I54M augmente la résistance au DRV. À l'inverse, le mutant I82F augmente la sensibilité au DRV. La combinaison de ces deux mutations I54M/I82F n'engendrant quant à elle ni susceptibilité ni résistance au DRV ([Raugi et al., 2013](#)).

De fait, la comparaison des trajectoires de dynamique moléculaire de ces trois mutants pourrait améliorer la compréhension des mécanismes de résistance mis en œuvre. La localisation de ces mutations sur la structure 3D de la PR2 est présentée sur la figure 4.2). La mutation I54M se trouve au niveau des flaps et la mutation I82F se trouve au niveau de la région wall. De plus, la mutation I82F se trouve au niveau de la poche de liaison des IP et le résidu I82 établit des interactions hydrophobes avec le DRV.

La localisation des mutations suggère que :

- la mutation I54M pourrait avoir un impact sur le mouvement des flaps et pourrait avoir une influence sur l'ouverture et la fermeture de la PR2 ;
- la mutation I82F pourrait altérer la fixation des IP.

Nos études précédentes ont montré que ces deux mutations ont un impact sur la poche et la zone de dimérisation. En effet, la mutation I54M augmente légèrement l'hydrophobicité et la taille de la poche de liaison des IP et modifie l'interface de dimérisation ([Laville et al., 2021](#)). La mutation I82F entraîne quant à elle une diminution du volume de la poche et des réarrangements structuraux au niveau du résidu 82 qui est impliqué dans l'interaction avec les IP. Ces résultats ont été obtenus en comparant les structures cristallographiques de la PR2 sauvages avec les structures modélisées des

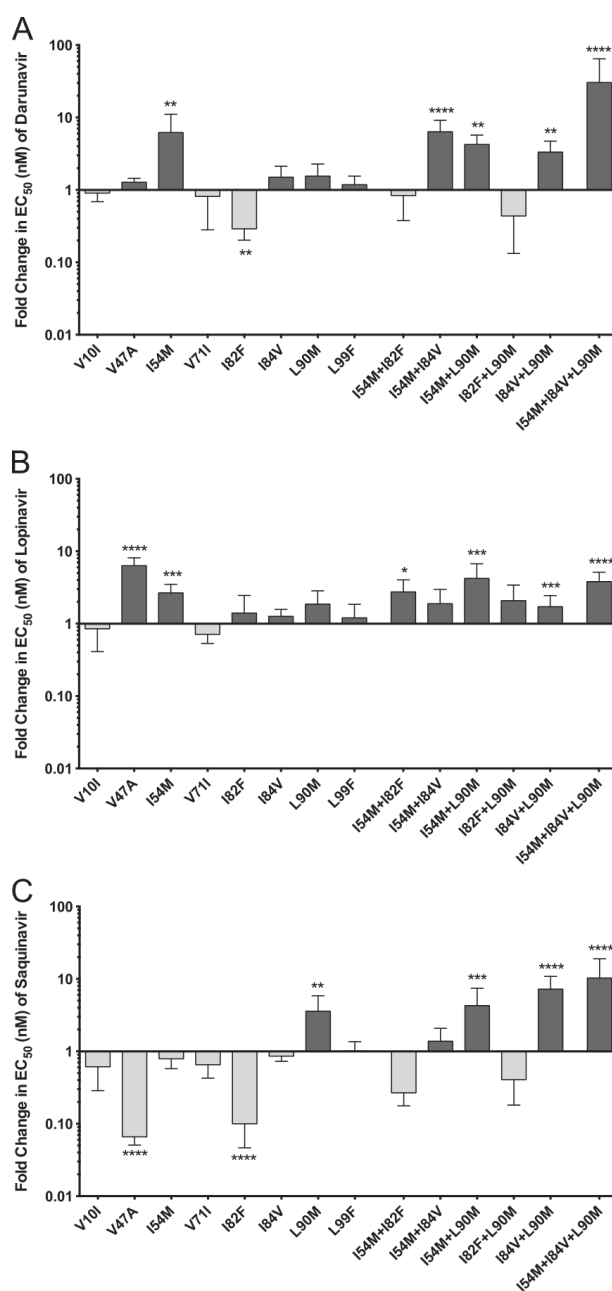


Figure 4.1 : Variation de susceptibilité des trois médicaments pour la PR2 : DRV (A), LPV (B) et SQV (C). Susceptibilité exprimée par le ratio $\frac{EC_{50\text{mutant}}}{EC_{50\text{sauvage}}}$ en nM. Figure extraite de Raugi *et al.*, 2013 reproduite avec la permission du journal *Antimicrobial Agents and Chemotherapy*.

mutants. Dans ces études, la flexibilité de la PR2 n'a été que faiblement prise en compte, et l'impact de ces mutations sur la dynamique de la PR2 n'a pas été étudiée.

4.2 Choix de la structure de départ

Dans la PDB, il existe 19 structures de la PR2 : 18 sont complexées à des inhibiteurs différents et seule une structure est une forme non complexée (code PDB : 1HSI ; Chen

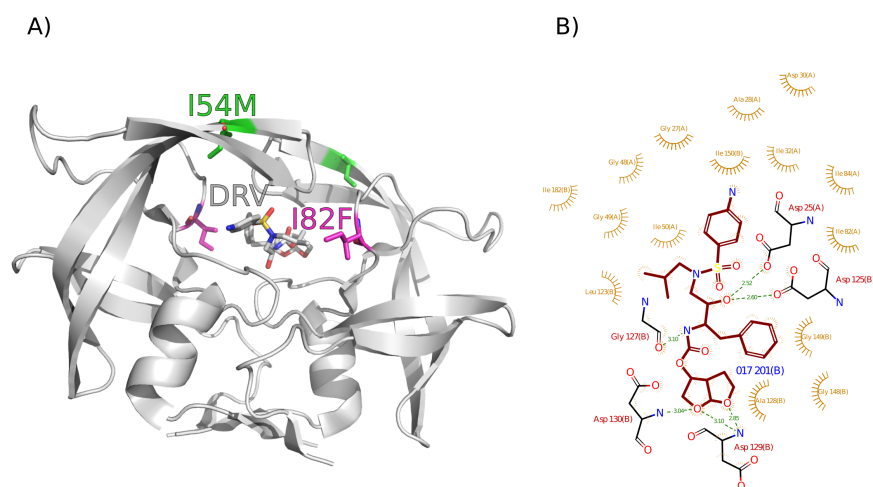


Figure 4.2 : A) Localisation des mutations I54M et I82F sur la structure de la PR2 (code PDB : 3EBZ). La PR2 est représentée en mode *cartoon* et colorée en gris. Les mutations sont représentées en *sticks* et colorées en vert (mutation I54M) et en rose (mutation I82F). Vue de $\frac{3}{4}$ face gauche afin de distinguer les résidus 82 et le DRV. B) Carte des interactions générée avec LIGPLOT de la structure de la PR2 (code PDB : 3EBZ) complexée au DRV (ligand 017). Les liaisons hydrogènes sont en pointillés verts et les interactions non liées (contacts hydrophobes) indiquées au niveau des résidus en jaune.

et al., 1994). La première question que nous nous sommes posée est la suivante : quelle structure cristallographique prendre comme structure de départ pour les simulations ? Parmi les 18 structures complexées à un ligand, trois sont complexées avec l'un des neuf médicaments (DRV, LPV, SQV). L'étude de ces structures cristallographiques a montré que la PR2 adopte deux conformations : une conformation semi-ouverte correspondant à la PR2 sans ligand et une forme fermée observée lorsque la PR2 est complexée à un ligand. La forme fermée se caractérise par une position des flaps abaissée sur le site de liaison permettant ainsi l'interaction entre les flaps et le ligand. À l'inverse, dans la forme semi-ouverte, les flaps sont relevés rendant accessible le site catalytique et permettant l'entrée des ligands (substrat ou inhibiteur). Nous nous sommes demandés si, après avoir supprimé le ligand de la PR2 complexée, cette PR2 était capable de retrouver la conformation de la PR2 sans ligand. Pour répondre à cette question, nous avons développé un protocole pour comparer l'état de la PR2 au cours de plusieurs simulations de dynamique moléculaire :

- une sur la structure cristallographique de la PR2 sans ligand (code PDB : 1HSI) ;
- une sur la structure complexée avec le DRV (code PDB : 3EBZ) où le DRV a été supprimé.

Nous avons ensuite étudié les structures obtenues au cours des deux simulations en comparant la conformation des flaps.

4.2.1 Protocole de dynamique moléculaire

Les simulations de dynamique moléculaire ont été réalisées avec le logiciel GROMACS (version 2018.3; Berendsen *et al.*, 1995). L'usage et l'automatisation ont été facilités par l'utilisation de Gromacs_py (Murail, 2018). Tous les calculs ont été réalisés sur le cluster du laboratoire. Les protéines et les molécules d'eau ont été décrites par le champ de force AMBER ff99SB-ILDN (Lindorff-Larsen *et al.*, 2010). Le système a été solvaté avec une boîte d'eau (modèle TIP3P) dodécaédrique avec une distance de 10 Å aux bords de la boîte. Des ions chlore et sodium ont été ajoutés pour neutraliser le système. Le système a été minimisé avant et après la solvataion et l'ionisation du système avec 5000 pas de *steepest descent* à chaque fois et une contrainte harmonique de $100 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{Å}^{-2}$. L'équilibration en trois étapes par défaut de Gromacs_py a été conservée. La première contraint les positions des atomes lourds, la deuxième celle des carbones α et la dernière impose une contrainte faible sur les carbones α . Les durées respectives de ces trois étapes sont de 1 ns, 5 ns et 10 ns. Le pas d'intégration est de 1 fs pour la première étape et de 2 fs pour les deux dernières. Il a été nécessaire de précéder cette étape d'une équilibration de 0,1 ns avec une contrainte sur les atomes lourds et un pas d'intégration de 0,1 fs. La simulation de dynamique moléculaire a été sans contrainte pendant la phase de production. Les liaisons covalentes concernant les atomes d'hydrogène ont été contraintes avec l'algorithme *LINCS* (*LINear Constraint Solver*). Les interactions électrostatiques à longue portée ont été traitées avec la méthode *PME* (*Particle Mesh Ewald*) et un *cutoff* de 10 Å pour les interactions électrostatiques à longue portée et de van der Waals. La température a été maintenue à 310 K avec le thermostat *v-rescale* et la pression contrôlée avec l'algorithme de Parrinello-Rahman.

Ce protocole de simulation a été lancé à partir des deux structures initiales (codes PDB : 1HSI et 3EBZ) sur un temps de 1 microseconde pour les structures sauvages et les trois structures mutantes. Trois répliques des simulations pour les structures sauvages ont été calculées sur un temps plus court de 100 ns. Cela représente un temps total de simulation de $2 \times 1 + 3 \times 1 + 3 \times 0,1 = 5,3 \mu\text{s}$. L'extraction des structures dans chaque trajectoire s'est faite avec un pas d'extraction de 1 ns.

4.2.2 Étude de la conformation des flaps au cours des différentes simulations

Dans la littérature, il existe plusieurs critères pour caractériser la conformation des flaps (Chen *et al.*, 2014; Tóth and Borics, 2006; Sadiq and Fabritiis, 2010; Hornak *et al.*, 2006). Par exemple, plusieurs études mesurent la distance entre les carbones α des résidus 50 (isoleucine) des deux chaînes (*Ile50C α -Ile50'C α*), qui mesure l'écartement entre les deux flaps présenté sur la figure 4.3 (Chen *et al.*, 2014; Tóth and Borics, 2006; Hornak *et al.*, 2006). Dans leur étude, Hornak *et al.*, 2006 ont caractérisé la conformation semi-ouverte de PR1 comme présentant une distance *Ile50C α -Ile50'C α* de 4,3 Å et la conformation fermée comme présentant une distance de 5,8 Å. Sadiq and Fabritiis, 2010

utilisent quant à eux le couple de distance entre les carbones α des résidus 50 et 81 pour la chaîne A et entre les résidus 49 et 80 pour la chaîne B. En plus de la distance ($Ile50C^\alpha - Ile50'C^\alpha$), [Chen et al., 2014](#) utilisent la distance entre les carbones α des résidus 25 et 50 dans les deux chaînes ainsi que le calcul de l'angle entre les carbones α des résidus 49, 50, et 51 de la chaîne A. Dans une étude récente, les auteurs caractérisent l'état des flaps en utilisant les distances ($Ile25C^\alpha - Ile50^\alpha$) dans les deux chaînes ainsi que l'angle entre les atomes carbones α des résidus 48, 49 et 50 dans les deux chaînes ([Sherry et al., 2021](#)).

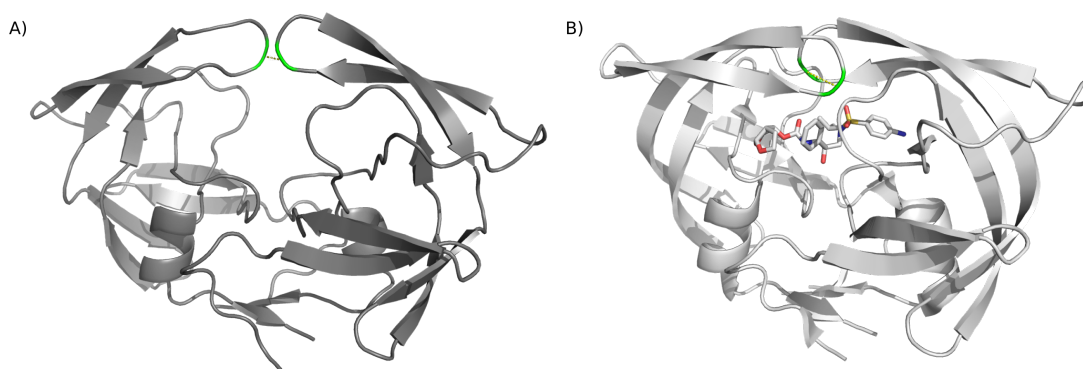


Figure 4.3 : Présentation de la distance $Ile50C^\alpha - Ile50'C^\alpha$ utilisée pour caractériser la conformation des flaps dans la forme A) semi-ouverte en gris foncé (code PDB : 1HSI), distance de 3,3 Å ; et B) fermée en gris clair (code PDB : 3EBZ, complexée avec le DRV), distance de 5,3 Å. Les résidus 50 de chaque chaîne sont colorés en vert et la distance en pointillés jaunes. Vue de $\frac{3}{4}$ face afin de distinguer la superposition des flaps sur B).

Parmi ces différentes mesures, nous avons choisi d'utiliser la distance $Ile50C^\alpha - Ile50'C^\alpha$ pour caractériser la conformation de la région des flaps. Nous avons calculé la distance $Ile50C^\alpha - Ile50'C^\alpha$ pour chacune des structures extraites des deux simulations. La figure 4.4 présente la distribution de cette distance pour les deux simulations. La première remarque que nous pouvons faire est que les deux simulations ne présentent pas le même profil de distance $Ile50C^\alpha - Ile50'C^\alpha$, ce qui montre que les régions flaps n'échantillonnent pas les mêmes conformations au cours des deux simulations. Par la suite, nous avons étudié plus en détails ces conformations pour les deux simulations.

Étude de la simulation de la PR2 semi-ouverte

Au cours de la simulation de la PR2 semi-ouverte, la distance $Ile50C^\alpha - Ile50'C^\alpha$ varie entre 3,99 Å et 8,69 Å avec un pic proche de 5 Å (figure 4.4). Nous remarquons que la majorité des structures (93 %) ont une distance comprise entre 4,3 et 5,8 Å (seuil définis par [Hornak et al., 2006](#) pour caractériser les conformations semi-ouverte et fermée des flaps chez PR1). Il semble donc qu'au cours de la simulation de la forme semi-ouverte, la PR2 adopte des conformations variant entre ces deux états. Étonnamment, au cours de cette simulation, il semble que la PR2 n'adopte pas de forme ouverte qui se caractérise par une distance $Ile50C^\alpha - Ile50'C^\alpha$ proche de 17 Å chez PR1 ([Hornak et al., 2006](#)). La figure 4.4

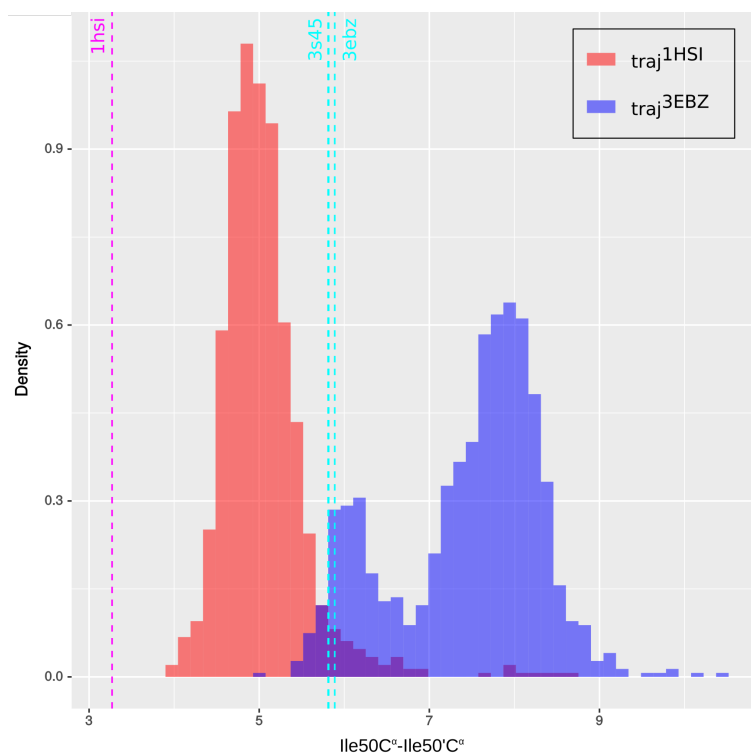


Figure 4.4 : Distribution de la distance $Ile50C^{\alpha}-Ile50'C^{\alpha}$ (en Å) pour les structures extraites des trajectoires lancées sur la forme semi-ouverte (rouge) et sur la forme fermée (bleue). Les lignes verticales pointillées représentent les distances $Ile50C^{\alpha}-Ile50'C^{\alpha}$ calculées sur les structures cristallographiques.

montre que la distance $Ile50C^{\alpha}-Ile50'C^{\alpha}$ calculée sur la structure cristallographique de la forme semi-ouverte (code PDB : 1HSI) n'est jamais retrouvée au cours des simulations. Ceci suggère que la structure cristallographique présente des flaps avec une conformation particulière qui pourrait être due au *packing* cristallin. Pour confirmer ces résultats, nous avons effectué cette même analyse sur les données des trois simulations des répliques de 100 ns. La figure 4.5 présente la distribution de la distance $Ile50C^{\alpha}-Ile50'C^{\alpha}$ pour chaque réplique. Nous remarquons que les distributions obtenues sur les trois répliques sont proches de celle obtenue au cours de la simulation longue.

Ainsi au cours de la simulation de la PR2 semi-ouverte, les flaps de la PR2 se déforment peu et évoluent principalement entre les formes semi-ouverte et fermée sans présenter d'ouverture totale.

Étude de la simulation de la PR2 fermée

La distribution de la distance $Ile50C^{\alpha}-Ile50'C^{\alpha}$ calculée au cours de la trajectoire de la forme fermée montre deux états majoritaires : le premier pic se trouve vers 6 Å et le second à 7,8 Å. D'après les seuils définis par [Hornak et al., 2006](#) (5,8 Å chez PR1), nous considérons que le premier état correspond à l'état fermé. Le second état ne fait référence à aucun état déjà défini dans la littérature. La visualisation des conformations de la région

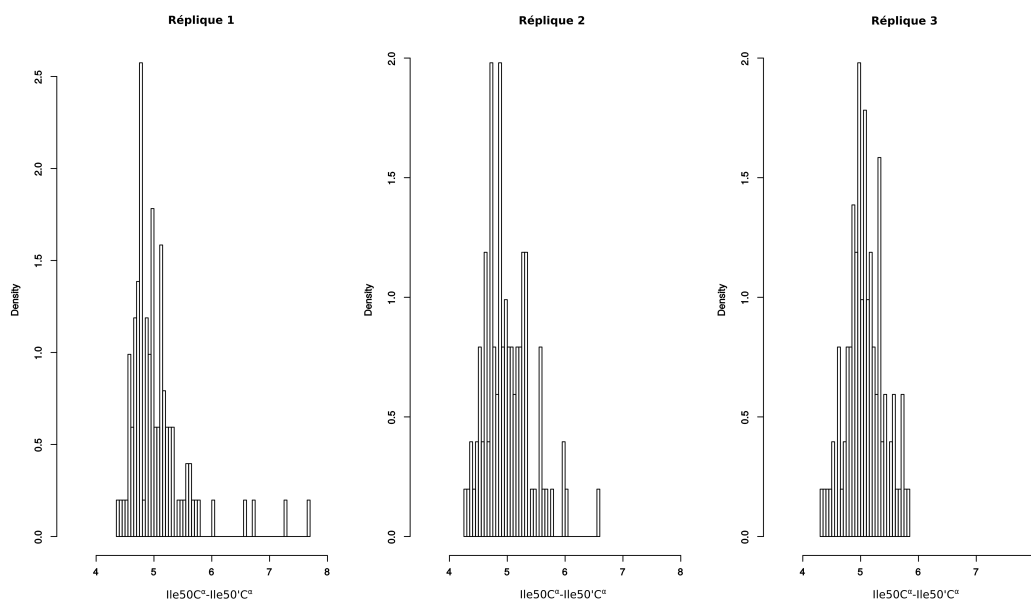


Figure 4.5 : Distribution des distances $Ile50C^{\alpha}-Ile50'C^{\alpha}$ pour les structures extraites des trajectoires des simulations des trois répliques de 100 ns lancées sur la forme semi-ouverte.

des flaps de ces structures montre que ces structures correspondent à des formes fermées mais avec un des flaps qui a pivoté sur lui-même, ce qui explique la distance $Ile50C^{\alpha}-Ile50'C^{\alpha}$ plus grande (figure 4.6). Nous appelons cet état l'état fermé incurvé. D'après la figure 4.5, nous observons que cet état est majoritaire par rapport à l'état fermé.

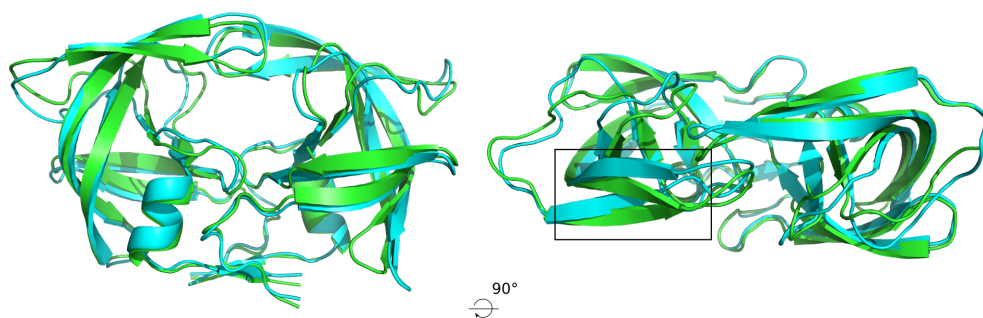


Figure 4.6 : Représentation des deux conformations fermées (fermée et fermée incurvée) des flaps observées lors de la simulation de la PR2 sous forme fermée. Les structures de la PR2 extraites aux temps 417 ns (vert) et 590 ns (cyan) sont représentées en mode *cartoon*. Les flaps de la première structure (417 ns) sont en conformation fermée et les flaps de la deuxième structure (590 ns) sont en conformation fermée incurvée.

Dans cette simulation, la structure extraite au temps 212 ns présente la plus petite distance $Ile50C^{\alpha}-Ile50'C^{\alpha}$ (distance $Ile50C^{\alpha}-Ile50'C^{\alpha} = 5,04 \text{ \AA}$). Ainsi, il semble qu'aucune structure extraite de simulation de la PR2 en forme fermée n'ait réussi à s'ouvrir pour atteindre l'état semi-ouvert (distance $Ile50C^{\alpha}-Ile50'C^{\alpha}$ proche de $4,3 \text{ \AA}$) ou ouvert (distance $Ile50C^{\alpha}-Ile50'C^{\alpha}$ proche de 17 \AA). De plus, la figure 4.4 montre

que les structures cristallographiques des complexes PR2-APV, PR2-IDV, et PR2-DRV présentent une distance $Ile50C^\alpha-Ile50'C^\alpha$ proche de 5,8 Å, qui correspond à la distance de l'état fermé. Ainsi, lors de la simulation de la PR2 sous forme fermée, la PR2 échantillonne des conformations proches de celles observées. Ceci peut s'expliquer par le fait que, dans la forme fermée, les flaps sont refermés sur le site catalytique et sûrement moins exposés au *packing* cristallin.

Pour étudier la robustesse de ces résultats, nous avons étudié la distribution de la distance $Ile50C^\alpha-Ile50'C^\alpha$ au cours de trois répliques de simulations de la forme fermée de PR2 (figure 4.7). Ces trois répliques confirment la présence de deux états majoritaires. Par contre, leur fréquence varie en fonction de la simulation. En effet, l'état fermé (pic proche de 6 Å) est plus fréquent dans la troisième réplique que dans les autres.

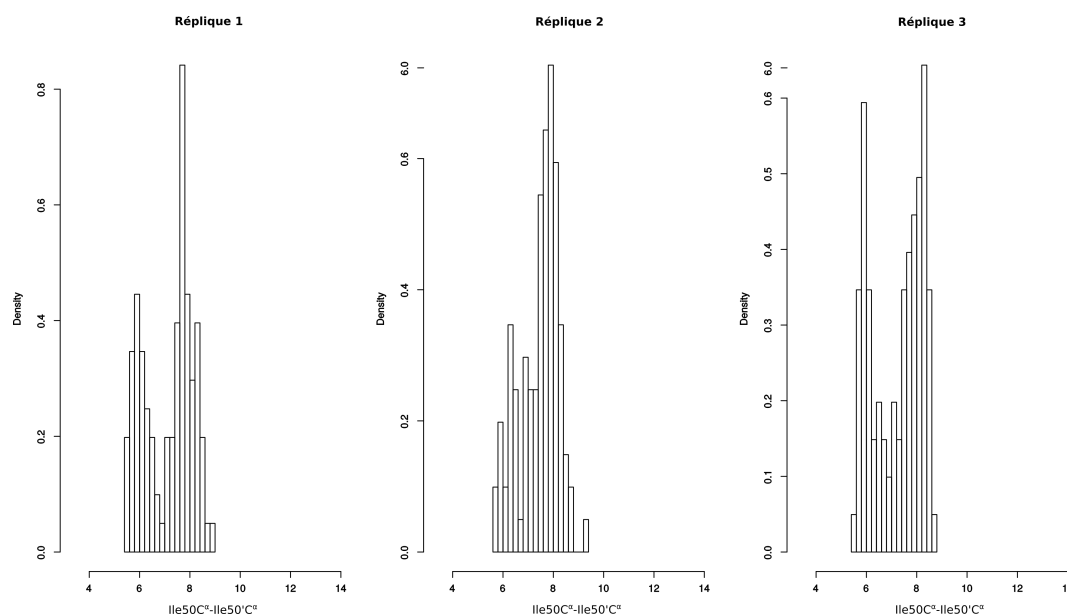


Figure 4.7 : Distribution des distances $Ile50C^\alpha-Ile50'C^\alpha$ pour les structures extraites des trajectoires des simulations des trois répliques de 100 ns lancées sur la forme fermée.

La figure 4.8 présente l'évolution de la distance $Ile50C^\alpha-Ile50'C^\alpha$ au cours du temps. Nous remarquons que la trajectoire correspond à une succession de passages d'un état fermé à un état fermé incurvé pour lequel nous observons des petits plateaux. Ceci suggère que ce second état fermé incurvé des flaps serait plus stable que l'état fermé.

Conclusion sur la comparaison des deux simulations

La comparaison de la distance $Ile50C^\alpha-Ile50'C^\alpha$ au cours des deux simulations sur la forme semi-ouverte et la forme fermée sans ligand montre que le fait d'enlever le ligand de la PR2 ne permet pas à la protéine d'avoir le même comportement que la structure semi-ouverte (sans ligand). En effet, nous avons remarqué que, lors de la simulation de la forme semi-ouverte, la PR2 échantillonne des conformations intermédiaires entre les conformations semi-ouverte et fermée. À l'inverse, au cours de la simulation de la PR2

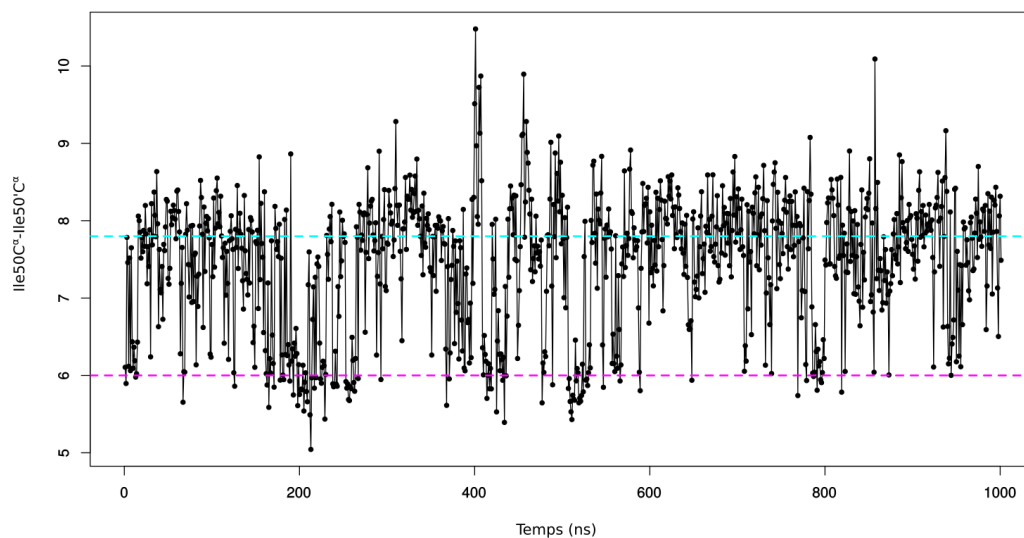


Figure 4.8 : Évolution de la distance $Ile50C^{\alpha}-Ile50'C^{\alpha}$ au cours du temps lors de la simulation lancée sur la forme semi-ouverte. Les lignes horizontales pointillées correspondent aux distances $Ile50C^{\alpha}-Ile50'C^{\alpha}$ du premier (magenta) et second (cyan) pics observés sur la distribution des distances $Ile50C^{\alpha}-Ile50'C^{\alpha}$.

fermée, la PR2 échantillonne différentes conformations mais en restant dans la forme fermée. Ainsi au cours des simulations réalisées, nous n'avons pas observé de larges mouvements des flaps aboutissant à l'ouverture ou la semi-ouverture de ces derniers. Ces résultats sont en accord avec ceux obtenus par [Chen et al., 2014](#) qui ont montré que PR1 et PR2 présentent des mouvements des flaps différents : chez PR1 les auteurs ont observé un éloignement des flaps aboutissant à l'ouverture de la PR2, alors que chez PR2 les flaps ont tendance à se rapprocher, favorisant la forme fermée de la PR2. De plus, lors de nos analyses, nous avons montré que la conformation des flaps observée dans la structure cristallographique de la PR2 en forme semi-ouverte n'a pas été échantillonnée au cours des simulations. Ce résultat suggère qu'il existe des erreurs dans cette structure pouvant s'expliquer par sa résolution moyenne (2,5 Å) et du *packing* cristallin. Ainsi pour étudier les impacts des mutations I54M et I82F, nous avons choisi de travailler avec les structures mutantes construites à partir de la PR2 complexée avec le DRV (code PDB : 3EBZ).

Au cours de notre analyse des conformations des flaps, nous nous sommes rendus compte que caractériser l'état des flaps en utilisant un seul descripteur (ici la distance $Ile50C^{\alpha}-Ile50'C^{\alpha}$) n'était pas assez précis. Comme nous l'avons dit précédemment, il existe un certain nombre de descripteurs qui ont déjà été utilisés pour définir la conformation des flaps. Actuellement, une caractérisation précise des différents états des flaps adoptés par la PR2 n'a jamais été apportée. Pour cela, il serait intéressant de

combiner les différents descripteurs en utilisant des approches multivariées pour apporter une classification des états adoptés par les flaps de la PR2.

4.3 Impact des mutations sur la structure de la PR2

Dans cette partie, nous avons étudié l’impact des mutations I54M et I82F sur la PR2 et sa dynamique. Pour cela, nous avons appliqué notre protocole de simulation de dynamique moléculaire sur les structures des trois mutants I54M, I82F et I54M/I82F que nous avons modélisées dans l’étude [Laville *et al.*, 2021](#) à partir de la PR2 en forme fermée (code PDB : 3EBZ ; sans le ligand). Les simulations ont été faites sur un temps d’1 μ s. Nous avons extrait 101 structures au cours de chaque simulation en utilisant un pas d’extraction de 10 ns (100×1 structure à chaque 10 ns + la structure de départ = 101 structures). Les résultats qui sont présentés dans cette partie sont des résultats préliminaires.

Étude de la stabilité et de la flexibilité de la PR2

Dans une première partie, nous avons exploré l’impact des mutations sur la stabilité et la flexibilité de la PR2. Tout d’abord, nous avons calculé le RMSD pour chaque trajectoire afin d’étudier la stabilité des structures. Le RMSD renseigne sur la stabilité globale de la structure au cours de la dynamique. Celui-ci a été calculé sur les carbones α pour les quatre dynamiques (figure 4.9). La valeur moyenne du RMSD pour la structure sauvage est de 1,21 Å. En comparaison à celle-ci, le RMSD moyen des simples mutants est inférieur : 1,19 Å et 1,14 Å respectivement pour les mutants I54M et I82F. Cela indique que ces mutations, notamment I82F, ont tendance à stabiliser la structure avec moins de mouvements au niveau du squelette. À l’inverse, le double mutant I54M/I82F présente un RMSD moyen supérieur à celui de la structure sauvage. La présence de ces deux mutations conduit donc à une augmentation de l’instabilité de la structure avec plus de mouvements au niveau des carbones α .

Par ailleurs, afin d’étudier l’impact des mutations sur la flexibilité de la PR2, nous avons mesuré pour chaque résidu le *RMSF* (*root-mean square fluctuation*) de chaque carbone α au cours de la simulation. Le RMSF est une mesure qui renseigne sur les mouvements par résidu pendant la simulation. Pour étudier l’effet des mutations sur le RMSF de chaque résidu, une comparaison a été effectuée entre les profils des RMSF obtenus pour la structure sauvage et les structures mutantes (figure 4.10). Cela permet de déterminer les résidus qui contribuent le plus aux mouvements au sein de la protéine.

Pour le mutant I54M, la plus grande différence de flexibilité observée se localise au niveau des résidus 48–50 sur les deux chaînes, correspondant à la région des flaps (différence de RMSF de + 0,6 Å). Nous observons une différence de l’ordre de + 0,4 Å aux positions 78–80 de la chaîne A (régions R3 et wall). Les régions flaps et wall sont importantes pour la structure du site de liaison des IP. Le fait que ces régions soient plus flexibles chez le mutant pourrait aboutir à des conformations altérées du site de liaison et

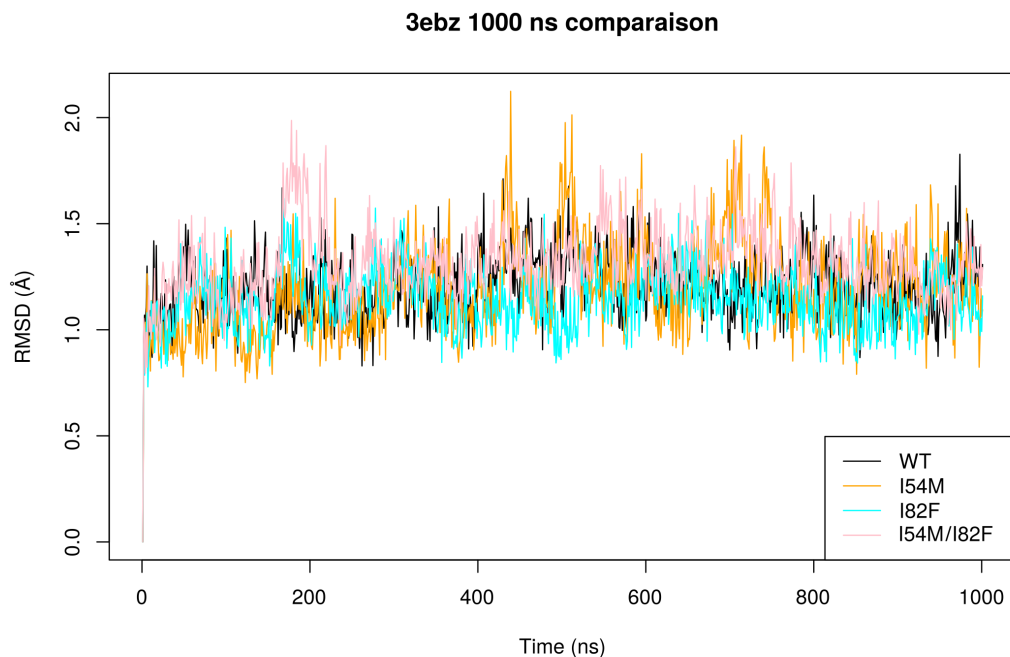


Figure 4.9 : Évolution du RMSD au cours de la dynamique pour les trajectoires des structures sauvage et mutantes.

donc une mauvaise fixation des IP. À l'inverse, nous notons que certaines positions sont moins flexibles chez le mutant I54M par rapport à la PR2 sauvage : les positions 34 et 67 de la chaîne A ($-0,2$ Å) et les positions 14 et 32 de la chaîne ($-0,3$ Å). La position 32 est une position importante dans la fixation des IP : elle établit des interactions hydrophobes avec le DRV qui présente une flexibilité importante pour la reconnaissance spécifique des ligands (Triki *et al.*, 2019).

Le mutant I82F présente quant à lui moins de différences de flexibilité par rapport à la structure sauvage comparé aux autres mutants. Seules les positions 78–80 de la chaîne A sont nettement plus flexibles ($+0,4$ Å) comparées à la structure sauvage. Comme pour le mutant I54M, l'augmentation de la flexibilité de la région wall de la chaîne A pourrait avoir un impact sur la conformation du site de liaison des IP et donc sur la fixation de ces derniers. Les positions moins flexibles sont les positions 15 sur les deux chaînes ($-0,2$ Å). Ces résidus sont situés dans la boucle au niveau du fulcrum. Le résidu 15 est un résidu éloigné du résidu muté : la distance entre les deux résidus est proche de 10 Å. Il semble donc que la mutation I82F ait des effets à longue distance.

C'est pour le double mutant I54M/I82F que les différences de flexibilité sont les plus importantes par rapport à la structure sauvage. Les positions 16 (région fulcrum) et 35 (région R2) de la chaîne A sont plus flexibles ($+0,4$ Å et $+1$ Å respectivement). Nous remarquons que les positions 48–50 (région flap, $+0,5$ Å), la position 67 (région cantilever, $+0,3$ Å) et 78–80 (régions R3 et wall, $+0,2$ Å) de la chaîne A et les positions

47–49 (région flap, + 0,2 Å) de la chaîne B sont aussi plus flexibles chez le double mutant que chez la PR2 sauvage. Inversement, les positions 14 (fulcrum, – 0,4 Å) et 32–33 de la chaîne B (région catalytique – 0,9 Å et – 0,4 Å) sont moins flexibles chez le mutant que chez la PR2 sauvage. Nous observons que certains de ces résidus sont éloignés des résidus mutés. Par exemple, le résidu 67 se trouve à plus de 14 Å des résidus 54 et 82.

Ces analyses montrent que la double mutation a un impact plus important sur la PR2 que les deux simples mutations. Elle augmente la flexibilité de plusieurs régions dont celle des flaps et des walls, ce qui peut entraîner des changements de conformations dans le site de liaison après la fixation des IP, altérant ainsi leur liaison. Comme nous l’avons observé chez le mutant I54M, la double mutation I54M/I82F semble diminuer la flexibilité des résidus 32 et 33, résidus flexibles directement impliqués dans l’interaction avec certains IP et qui jouent un rôle dans la reconnaissance spécifique de certains IP.

Exploration de l’espace conformationnel échantillonné par les structures

Dans un premier temps, nous avons cherché quels sont les impacts des mutations sur la conformation globale de la PR2 en comparant l’espace conformationnel échantillonné par les PR2 sauvage et mutantes au cours des quatre simulations. Pour cela, nous avons utilisé le protocole développé par [Triki *et al.*, 2018a](#) pour étudier l’impact des substitutions observées entre PR1 et PR2 sur la structure de PR2. Toutes les structures extraites des quatre simulations ont été superposées à la structure cristallographique correspondant au code PDB 3EBZ (PR2 complexé au darunavir) en utilisant le logiciel PyMOL ([PyMOL](#)) pour les placer dans le même référentiel.

Nous avons ensuite calculé le RMSD (sur les carbones α) entre toutes les paires de structures extraites des simulations de la PR2 sauvage et mutantes. À partir de ces valeurs de RMSD, nous avons calculé un positionnement multidimensionnel — ou *multidimensional scaling (MDS)* — pour générer la carte de MDS. Le MDS est une technique multivariée non supervisée de visualisation d’individus à partir de leur similarité [Cox and Cox, 2008](#). L’algorithme du MDS cherche l’espace à faible dimension où les distances entre les individus sont les plus conservées [Cox and Cox, 2008](#). Le MDS des structures mutantes et sauvages est calculé à partir de la matrice de RMSD permettant de quantifier la similarité structurale entre chaque structure.

À partir des données obtenues par le MDS, nous avons généré la carte de MDS qui correspond à la projection des individus sur le premier plan du MDS. Cette carte présente l’espace conformationnel échantillonné par les structures extraites des simulations de la PR2 sauvage et mutantes. Cette carte est représentée sur la figure [4.11](#). Sur cette carte, deux points proches correspondent à des structures avec une forte similarité structurale, alors que deux points éloignés correspondent à des structures différentes. La figure [4.11](#) montre que les structures des simples mutants et de la PR2 sauvage échantillonnent le même espace conformationnel. Ainsi, on peut conclure que ces deux mutations ont peu d’impact sur la conformation globale de la PR2. Cependant, nous remarquons que le

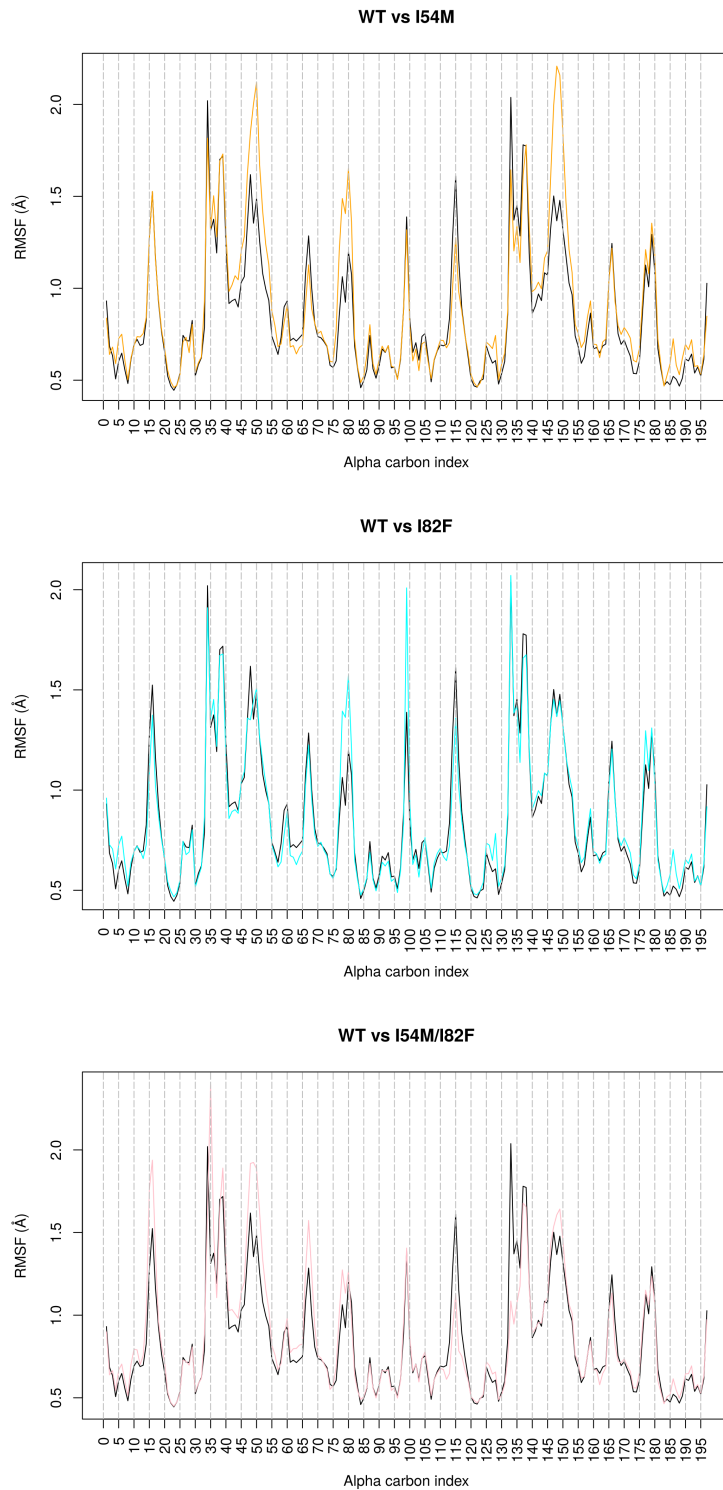


Figure 4.10 : Comparaison du RMSF des structures sauvage et mutantes sur les carbones α . Les index des carbones α sont numérotés de 1 à 198 : les index de la chaîne A vont de 1 à 99 et ceux de la chaîne B de 100 à 198 (pour les résidus 1 à 99 de la chaîne B).

mutant double (I54M/I82F) échantillonne un espace spécifique (coin en bas à droite, figure 4.11). Environ 75 % des structures du double mutant se trouvent dans cet espace.

Ceci montre que l'apparition de ces deux mutations ensemble modifie la conformation globale de la PR2.

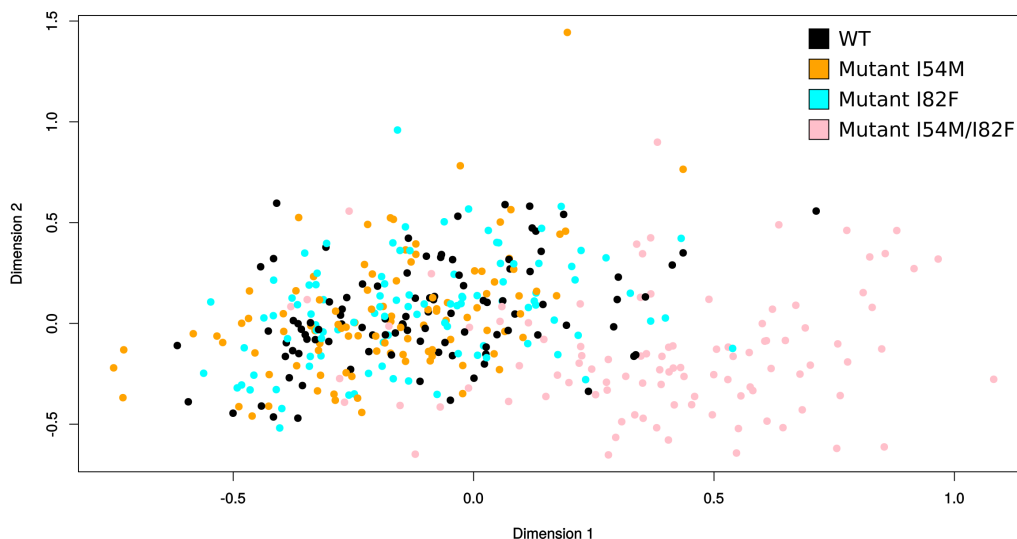


Figure 4.11 : Carte du MDS calculée à partir des RMSD entre les structures extraites des simulations lancées sur les structures sauvages et mutantes. Cette carte traduit l'espace conformationnel échantillonné par ces structures. Les points représentent les structures extraites de la simulation de la PR2 sauvages (noir), du mutant I54M (orange), du mutant I82F (cyan) et du double mutant I54M/I82F (rose).

Pour approfondir l'analyse, nous avons comparé les conformations des 13 régions de chaque structure extraite des simulations. Pour cela, nous avons calculé une carte de MDS pour chaque région des chaînes A et B après avoir calculé les RMSD entre fragments d'une région donnée extraits de toutes les structures extraites des quatre trajectoires. La figure 4.12 présente les cartes de MDS pour les 13 régions des chaînes A et B. Nous remarquons que, pour la majorité des régions, il existe peu de différences entre les structures mutantes et sauvages. Cependant, le double mutant I54M/I82F présente des conformations particulières pour les régions R2 et elbow de la chaîne A (environ 75 % des structures du double mutant) et les régions R3 et wall de la chaîne B (environ 92 % des structures du double mutant) comparé à la structure sauvage.

En étudiant le nombre des structures présentant une conformation particulière pour une région, nous constatons que la majorité des structures présentant une conformation du R2 de la chaîne A sont celles qui présentent une conformation particulière du R3 de la chaîne B. Ces différences structurales au niveau de ces deux régions sont capturées lors de la comparaison de la structure globale de structures sauvage et mutante (figure 4.11). Ces résultats montrent que les mutations I54M et I82F induisent des changements conformationnels au niveau du squelette de ces régions chez le double mutant.

La figure 4.13 présente la superposition d'une structure extraite de la simulation de la PR2 sauvage et une structure du double mutant (extraite au temps 750 ns) présentant

des conformations particulières pour ces régions. Cette figure permet de visualiser le large changement conformationnel qui existe entre les deux structures, particulièrement au niveau de la chaîne A. La région wall est une région impliquée dans la poche de liaison. Ainsi, les changements conformationnels dans cette région chez le double mutant sont en accord avec les résultats que nous avons obtenus lors de notre seconde étude (Laville *et al.*, 2021) où nous avons montré que ces deux mutations modifient la structure de la poche et induisent une augmentation de son volume et une diminution de son hydrophobicité chez le double mutant. La figure 4.11 montre aussi que quelques structures du mutant double présentent des conformations particulières pour la région cantilever de la chaîne A et pour la région Cter de la chaîne B et que quelques structures des deux mutants simples présentent des conformations particulières pour la région R3 de la chaîne A.

Cette analyse sur la comparaison des structures extraites des simulations de la PR2 sauvage et mutantes a permis d'identifier que l'apparition des mutations I54M et I82F ensemble modifie le squelette de plusieurs régions : R2 et elbow de la chaîne A et R3 et wall de la chaîne B. En ce qui concerne les mutations seules, notre analyse n'a pas permis de mettre en évidence de forts changements conformationnels. Pour approfondir l'analyse, il serait intéressant de comparer les changements conformationnels entre les structures mutantes et sauvage en utilisant une approche plus locale. Par exemple, nous pourrions utiliser l'outil SA-conf (Regad *et al.*, 2017), pour comparer la structure locale des 198 résidus de la PR2 dans chacune des structures extraites des quatre simulations. En parallèle, nous pourrions appliquer le protocole que nous avons mis en place dans l'étude Laville *et al.*, 2021 pour identifier précisément les changements conformationnels induits par les mutations au niveau du squelette peptidique et des chaînes latérales.

Impacts des mutations sur la structure des flaps

Pour continuer à étudier l'impact des mutations I54M et I82F, nous nous sommes focalisés sur la conformation des flaps. Pour cela, nous avons étudié la distance $Ile50C^\alpha-Ile50'C^\alpha$ au cours des quatre simulations. Pour approfondir l'étude des flaps, nous avons utilisé quatre autres distances ($Asp25C^\alpha-Ile50C^\alpha$, $Asp25'C^\alpha-Ile50'C^\alpha$, $Ile50C^\alpha-Ile82C^\alpha$ et $Ile50'C^\alpha-Ile82'C^\alpha$) en plus de la distance $Ile50C^\alpha-Ile50'C^\alpha$. De plus, nous avons ajouté six distances nous permettant de caractériser d'autres régions de la protéine : $Asp25C^\alpha-Asp25'C^\alpha$, $Leu97C^\alpha-Leu97'C^\alpha$, $Glu16C^\alpha-Leu38C^\alpha$, $Glu16'C^\alpha-Leu38'C^\alpha$, $Leu38C^\alpha-Val62C^\alpha$ et $Leu38'C^\alpha-Val62'C^\alpha$). Toutes ces distances sont explicitées dans le tableau 4.1.

Nous avons calculé les valeurs des 11 distances pour les 101 structures extraites des quatre simulations. À partir de ces distances, nous avons calculé une analyse en composantes principales (ACP) d'un jeu de données d'individus (ici les structures) décrites par un ensemble de variables (ici les six distances). L'ACP est une méthode de visualisation de données multivariées qui permet de calculer de nouvelles variables non corrélées, appelées composantes principales, à partir des données. Ces composantes principales

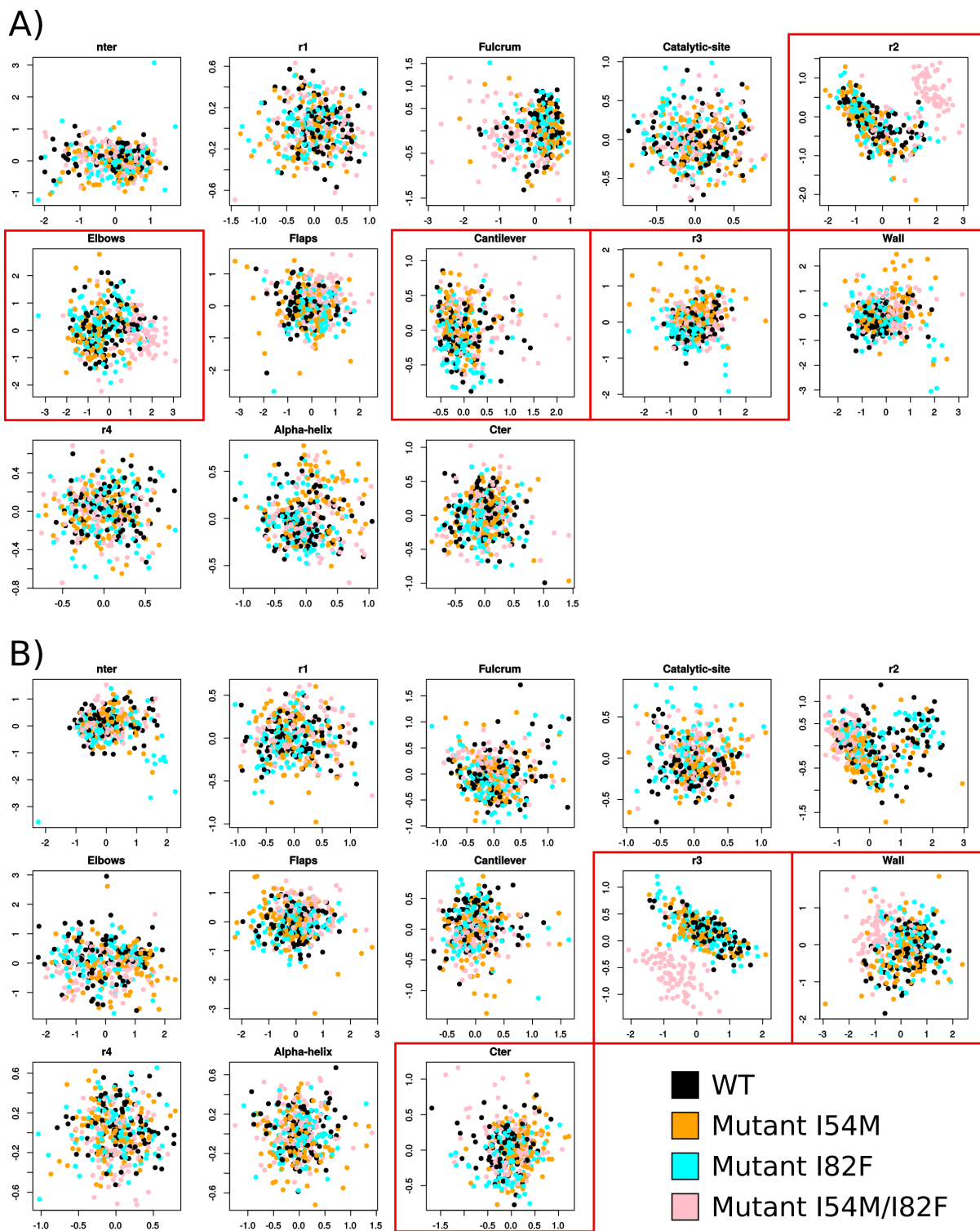


Figure 4.12 : Carte du MDS calculée à partir des RMSD entre les structures extraites des simulations lancées sur les structures sauvages et mutantes selon les 13 régions de la PR2. Les points représentent les structures extraites de la simulation de la PR2 sauvages (noir), du mutant I54M (orange), du mutant I82F (cyan) et du double mutant I54M/I82F (rose). Les régions encadrées en rouge sont les régions mentionnées dans le texte.

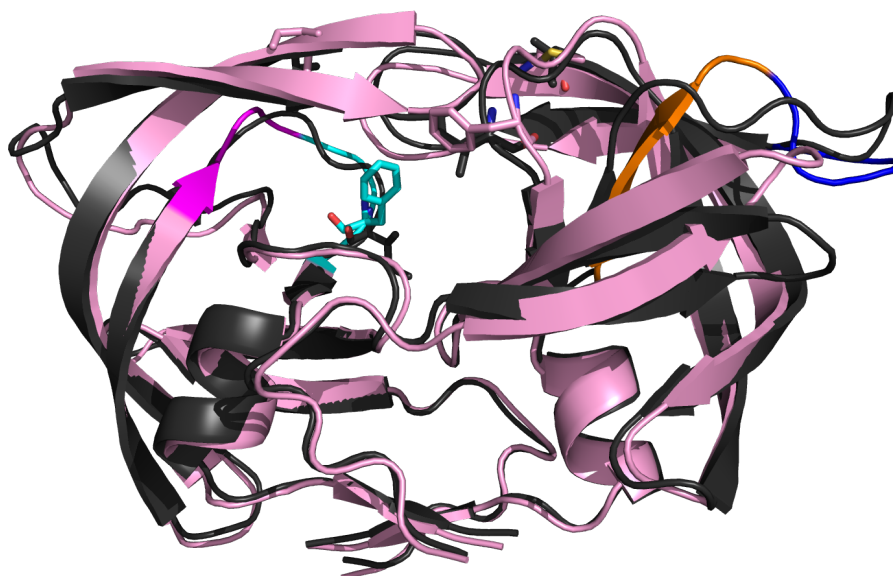


Figure 4.13 : La structure sauvage est en noir (structure extraite au temps 90 ns) et la structure du double mutant est en rose (extraite au temps 750 ns). Les régions R2 et elbow de la chaîne A sont colorées en orange et bleu, et les régions R3 et wall de la chaîne B sont colorées en magenta et cyan. Les résidus mutés en positions 54 et 82 sont représentés en *sticks*.

Distance	Définition
$Ile50C^{\alpha}-Ile50'C^{\alpha}$	distance entre les deux flaps
$Asp25C^{\alpha}-Ile50C^{\alpha}$	ouverture du flap de la chaîne A
$Asp25'C^{\alpha}-Ile50'C^{\alpha}$	ouverture du flap de la chaîne B
$Ile50C^{\alpha}-Ile82C^{\alpha}$	distance entre le flap et le wall de la chaîne A
$Ile50'C^{\alpha}-Ile82'C^{\alpha}$	distance entre le flap et le wall de la chaîne B
$Asp25C^{\alpha}-Asp25'C^{\alpha}$	distance entre les deux monomères au milieu de la PR2
$Leu97C^{\alpha}-Leu97'C^{\alpha}$	distance entre les deux monomères au niveau du bas de la PR2
$Glu16C^{\alpha}-Leu38C^{\alpha}$	distance entre le fulcrum et l'elbow de la chaîne A
$Glu16'C^{\alpha}-Leu38'C^{\alpha}$	distance entre le fulcrum et l'elbow de la chaîne B
$Leu38C^{\alpha}-Val62C^{\alpha}$	distance entre l'elbow et le cantilever de la chaîne A
$Leu38'C^{\alpha}-Val62'C^{\alpha}$	distance entre l'elbow et le cantilever de la chaîne B

Table 4.1 : Description des distances utilisées dans la comparaison des structures extraites des simulations de la PR2 sauvage et mutantes.

représentent des combinaisons linéaires des variables initiales et maximisent la variabilité des données. La figure 4.14 présente la projection des 404 structures sur le plan de l'ACP composé des deux premières composantes principales.

Ce premier plan représente une faible part de la variabilité : seulement 28 % de la variabilité totale. La figure 4.14 montre que les quatre types de structures (extraites des quatre simulations) échantillonnent globalement le même espace. Ainsi, l'utilisation des six distances n'a pas permis d'identifier des différences entre les trois structures mutantes et la structure sauvage.

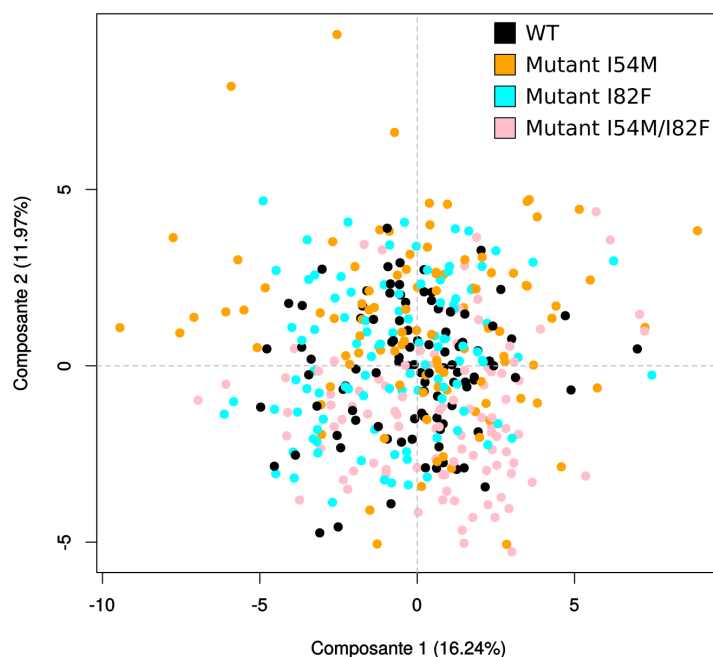


Figure 4.14 : Projection des structures sur le premier plan de l'ACP calculée à partir des six distances intra-protéiques. Les points représentent les structures extraites de la simulation de la PR2 sauvages (noir), du mutant I54M (orange), du mutant I82F (cyan) et du double mutant I54M/I82F (rose).

Distances	Kruskal-Wallis <i>p-value</i> *	Student <i>p-value</i> **		
		WT-I54M	WT-I82F	WT-I54M/I82F
<i>Ile50Cα</i> - <i>Ile50'Cα</i>	0,011	0,839	0,025	0,969
<i>Asp25Cα</i> - <i>Ile50Cα</i>	2×10^{-7}	0,849	0,002	0,001
<i>Asp25'Cα</i> - <i>Ile50'Cα</i>	0,254	-	-	-
<i>Ile50Cα</i> - <i>Ile82Cα</i>	0,037	0,005	0,330	0,494
<i>Ile50'Cα</i> - <i>Ile82'Cα</i>	2×10^{-6}	0,025	0,016	0,007
<i>Asp25Cα</i> - <i>Asp25'Cα</i>	0,049	0,44	0,194	0,09
<i>Leu97Cα</i> - <i>Leu97'Cα</i>	0,384	-	-	-
<i>Glu16Cα</i> - <i>Leu38Cα</i>	0,636	-	-	-
<i>Glu16'Cα</i> - <i>Leu38'Cα</i>	0,317	-	-	-
<i>Leu38Cα</i> - <i>Val62Cα</i>	2×10^{-7}	0,048	0,347	0,002
<i>Leu38'Cα</i> - <i>Val62'Cα</i>	4×10^{-8}	0,510	0,510	1×10^{-7}

Table 4.2 : Comparaison des distances moyennes des structures sauvages et des structures mutantes. « * » : *p-value* du test de Kruskal-Wallis permettant de tester si au moins un des quatre groupes présente une distance moyenne significativement différente. Les *p-values* ont été ajustées avec la méthode *false discovery rate*. « ** » : *p-values* des trois tests de Student permettant de comparer les distances moyennes entre les structures sauvages (WT) et les structures mutantes I54M, entre les structures sauvages (WT) et les structures mutantes I82F et entre les structures sauvages (WT) et les structures mutantes I54M/I82F. Les *p-values* ont également été ajustées avec la méthode *false discovery rate*. Les *p-values* significatives des différents tests sont en gras.

Pour continuer l'analyse, nous avons étudié séparément chacune des distances. Dans un premier temps, nous avons comparé les distances de chaque simulation. Pour chaque distance, nous avons comparé la distance moyenne de chaque simulation en utilisant le test de Kruskal-Wallis (test non paramétrique), table 4.2. Parmi les 11 distances, sept présentent une distance significative ($Ile50C^\alpha-Ile50'C^\alpha$, $Asp25C^\alpha-Ile50C^\alpha$, $Ile50C^\alpha-Ile82C^\alpha$, $Ile50C^\alpha-Ile82C^\alpha$, $Asp25C^\alpha-Asp25'C^\alpha$, $Leu38C^\alpha-Val62C^\alpha$ et $Leu38'C^\alpha-Val62'C^\alpha$), montrant qu'au moins une des quatre simulations présente une distance significativement différente. Pour ces sept distances, nous avons ensuite réalisé trois tests de Student pour comparer la distance moyenne entre les structures extraites de la simulation de la PR2 sauvage et (i) celles du mutant I54M, (ii) celles du mutant I82F, (iii) celles du mutant I54M/I82F (Table 4.2). Nous remarquons que les différences en termes de distance varient suivant le mutant.

Par exemple, nous observons des différences significatives pour les distances $Ile50C^\alpha-Ile82C^\alpha$, $Ile50C'^\alpha-Ile82C'^\alpha$ et $Leu38C^\alpha-Val62C^\alpha$ pour le simple mutant I54M. La figure 4.15 présente la distribution des distances pour les structures sauvages et mutantes. Dans la simulation de la structure mutante I54M, nous observons plus de structures présentant une plus grande distance $Ile50C^\alpha-Ile82^\alpha$ et inversement plus de structures présentant des plus petites distances $Ile50'C^\alpha-Ile82'^\alpha$. La mutation I54M semble donc avoir un impact sur la position des flaps et/ou wall dans les deux chaînes, et du elbow et/ou du cantilever dans la chaîne A.

Concernant le mutant simple I82F, nous observons des différences significatives au niveau des distances $Ile50C^\alpha-Ile50'C^\alpha$, $Asp25C^\alpha-Ile50C^\alpha$ et $Ile50C^\alpha-Ile82C^\alpha$. Ceci indique que la mutation I82F a un impact sur le mouvement de flaps, particulièrement celui de la chaîne A. D'après la figure 4.15 la mutation I82F semble modifier la fréquence des états fermé et fermé incurvé précédemment identifiés grâce à la distance $Ile50C^\alpha-Ile50C'^\alpha$: on observe plus d'états entre les deux formes. Cependant, au cours de l'analyse de la distance $Ile50C^\alpha-Ile50'C^\alpha$ pour les structures sauvages, nous avons remarqué que, chez les répliques, la fréquence de ces deux états était modifiée. Il serait donc intéressant d'approfondir cette analyse pour valider l'impact de la mutation I82F sur l'état fermé incurvé. Cette mutation semble aussi diminuer la distance $Ile50'C^\alpha-Ile82'C^\alpha$. D'après ces résultats, la mutation I82F semble donc avoir un impact sur la conformation des flaps.

La mutation I54M/I82F semble aussi avoir un impact sur le mouvement des flaps : les structures du double mutant présentent des distances $Asp25C^\alpha-Ile50C^\alpha$, $Ile50C^\alpha-Ile82C^\alpha$, $Leu38C^\alpha-Val62C^\alpha$ et $Leu38'C^\alpha-Val62'C^\alpha$ significativement différentes de celles des structures sauvages. D'après la figure 4.15, le double mutant présente des distances $Leu38C^\alpha-Val62C^\alpha$ et $Leu38'C^\alpha-Val62'C^\alpha$ plus faibles. Ces résultats montrent que ces deux mutations couplées ont un impact sur les régions elbow et/ou cantilever. Ces résultats confirment ceux obtenus lors de la comparaison des différentes régions où nous avons montré que la double mutation I54M/I82F modifie la conformation du elbow de la chaîne A et du wall de la chaîne B.

À travers la comparaison des valeurs de différentes distances des structures extraites de chaque simulation, nous avons mis en évidence que les deux mutations semblent avoir un impact sur le mouvement des flaps. De plus, cette analyse a permis de confirmer que la double mutation I54M/I82F a aussi un impact sur l'elbow de la chaîne A et le wall de la chaîne B. Pour continuer cette analyse, il serait intéressant de caractériser plus précisément le mouvement des flaps et les mouvements coopératifs détectés au cours des simulations comme l'ont fait [Chen *et al.*, 2014](#). Dans cette optique, l'analyse des simulations pourrait être couplée à la technique des modes normaux.

Conclusion

Pour étudier plus en détails les mutations I54M et I82F, nous avons effectué une comparaison avec la structure sauvage en utilisant une approche de dynamique moléculaire. Au cours de cette analyse, nous avons comparé la stabilité et flexibilité des PR2 sauvages et mutantes, leur structure globale et par région. Nos résultats ont montré que le fait de coupler les deux mutations tend à augmenter l'instabilité de la structure. Cela est confirmé par le fait que cette combinaison de mutations a plus d'impact sur la flexibilité et la structure de la PR2. De plus, nous avons montré que la mutation I54M, seule ou combinée avec la mutation I82F, augmente la flexibilité de certains résidus des flaps 48–49 de la chaîne A. De plus, chez les trois mutants, nous avons observé une augmentation de la flexibilité dans la région 78–80 (R3-wall) de la chaîne A. Le changement de flexibilité des régions R3 et wall s'accompagne par un changement de conformation de la région wall de la chaîne A chez le mutant I54M que l'on a détecté à l'aide de la distance $Ile50C^\alpha-Ile82C^\alpha$. Chez le double mutant, nous avons observé des changements de conformation des régions R2 et elbow de la chaîne A et des régions R3 et wall de la chaîne B.

Les régions flaps et wall sont importantes pour la structure du site de liaison des IP. Ainsi l'augmentation de la flexibilité des résidus de la région flap et du wall et leur changement de conformation peuvent entraîner des changements conformations dans le site de liaison après fixation des IP, altérant ainsi leur liaison.

La mutation I82F induit des changements conformationnels et semble modifier le mouvement des flaps (le mutant présente une distance $Ile50C^\alpha-Ile50'C^\alpha$ moyenne différente de la structure sauvage). En parallèle, nous avons observé que la mutation I82F, seule ou couplée à la mutation I54M, induit une diminution de la flexibilité de la position 32. Ce résidu est un résidu important dans la fixation des IP : il établit des interactions hydrophobes avec le DRV qui présente une flexibilité importante pour la reconnaissance spécifique des ligands ([Triki *et al.*, 2019](#)).

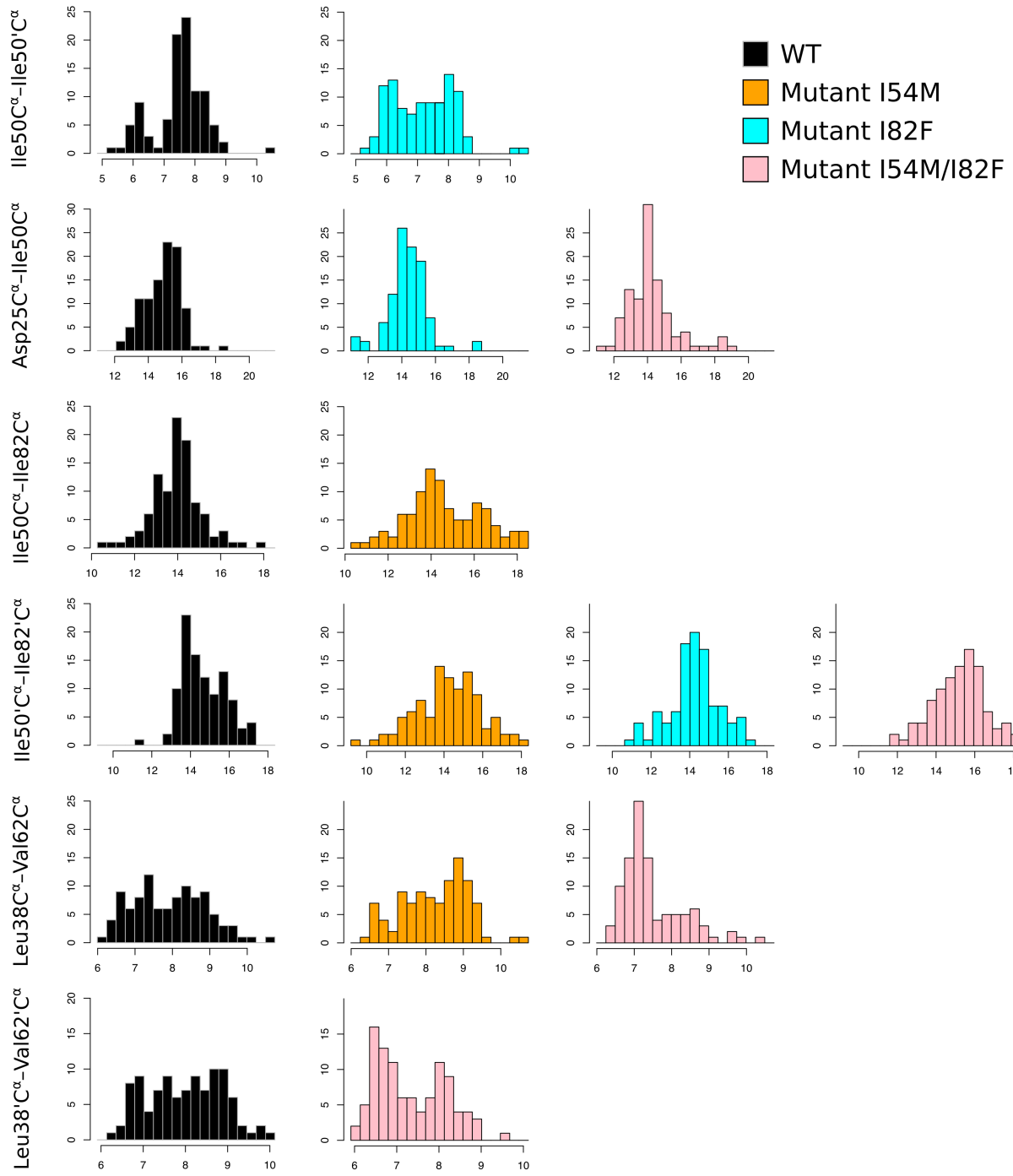


Figure 4.15 : Distributions des distances dont les différences sont significatives pour le WT et les mutants. En abscisse est présentée la distance en Å, en ordonnée le nombre de distances par classe. Le même code couleur est utilisé pour la représentation du WT et des mutants.

5 Conclusions et perspectives

Mon travail de thèse s'inscrit dans un projet qui vise à mieux comprendre les mécanismes expliquant la résistance acquise du VIH-2 contre les IP. Pour cela, nous avons axé nos études d'un point de vue structural. En effet, quand j'ai commencé ma thèse, aucune étude structurale n'avait été réalisée sur les mutants résistants de la PR2 du fait qu'aucune structure de PR2 mutante n'était disponible dans la PDB. Il nous a semblé qu'identifier, localiser et quantifier les modifications de la structure de PR2 induites par les mutations aiderait à mieux comprendre les mécanismes atomiques impliqués dans la résistance aux IP. Au cours de mon travail de thèse, j'ai effectué la première analyse structurale d'un grand nombre de mutants.

La stratégie que nous avons déployée au cours de ma thèse a été de comparer les structures mutantes avec la structure sauvage de la PR2. Comme indiqué précédemment, au moment où j'ai commencé ma thèse, aucune structure de la PR2 mutante n'était disponible dans la PDB, ce qui est toujours le cas aujourd'hui. Pour pallier ce manque de données, nous avons donc construit un protocole bioinformatique pour modéliser les structures des mutants étudiés. Comme notre but était d'étudier une liste de mutants, nous avons développé un protocole automatisable à plusieurs mutants en nous basant sur le protocole développé par [Triki *et al.*, 2018b](#) utilisant le logiciel FoldX ([Guerois *et al.*, 2002](#)). La limite de ce protocole est qu'il construit les structures mutantes uniquement en se basant sur les réarrangements structuraux induits par la ou les mutations au niveau des chaînes latérales localisées à proximité du ou des résidus mutés. Or, il est connu que les mutations peuvent entraîner des changements conformationnels au niveau du squelette peptidique et au niveau de résidus éloignés du résidu muté. C'est pourquoi, nous avons choisi d'ajouter une étape de minimisation au protocole déjà existant. La première partie de ma thèse a donc consisté à mettre au point ce protocole.

Une fois le protocole de modélisation prêt, nous avons sélectionné dans la littérature une première liste de 30 mutants à analyser. Ces mutants sont tous résistants à au moins un IP. Ces mutants ont tous été identifiés par des approches génomiques ([Masse *et al.*, 2007](#); [Colson *et al.*, 2004](#); [Damond *et al.*, 2005](#); [Rodés *et al.*, 2006](#); [Ntemgwa *et al.*, 2007](#); [Desbois *et al.*, 2008](#); [Jallow *et al.*, 2009](#); [Raugi *et al.*, 2013](#); [Larrouy *et al.*, 2013](#); [Cavaco-Silva *et al.*, 2013](#); [Storto *et al.*, 2018](#)), et certains d'entre eux ont été validés par des approches phénotypiques ([Masse *et al.*, 2007](#); [Rodés *et al.*, 2006](#); [Ntemgwa *et al.*, 2007](#); [Desbois *et al.*, 2008](#); [Raugi *et al.*, 2013](#); [Mahdi *et al.*, 2015](#)). Nous avons cherché à avoir un éventail de mutations qui se localisent dans différentes régions de la PR2 et pas seulement des mutations au niveau du site de liaison des IP. Nous avons donc sélectionné un ensemble de 30 mutants contenant de une à trois mutations et comptant au total 14 mutations différentes. Ce jeu de mutants a été l'objet de la première étude que nous avons menée (chap. 2). Pour notre seconde étude (chap. 3), cette liste de mutation a été retravaillée. Nous avons sélectionné une nouvelle liste de 31 mutants contenant plus de mutants simples

(20), et comportant plus de mutations localisées en dehors du site de liaison. Ce nouveau jeu de mutant rassemble 22 mutations différentes dont 11 sont communes à la première liste. Cette seconde liste de mutants a été traitée dans la seconde analyse effectuée dans ma thèse.

Au cours de notre première étude, nous avons donc étudié l'impact de ces 30 mutations sur la structure de la PR2. Afin d'étudier tous les effets des mutations et pas seulement ceux sur le site de liaison des IP, nous avons choisi de comparer la structure entière des PR2 sauvages et mutantes. Afin de simplifier un peu le problème, il nous a semblé intéressant dans un premier temps de nous focaliser sur l'exploration des impacts des mutations sur le squelette peptique. En effet, une approche qui permet d'étudier l'asymétrie de la PR2, c'est-à-dire les changements conformationnels des deux monomères d'un dimère, a été développée au laboratoire (Triki *et al.*, 2018b; Ollitrault *et al.*, 2018). Cette approche s'est montrée efficace pour localiser les changements structuraux localisés au niveau du squelette peptidique induits par soit la flexibilité intrinsèque des résidus soit par un facteur externe tel que les mutations ou la fixation des ligands (Triki *et al.*, 2018b). Au cours de cette étude nous avons donc recherché les impacts des 14 mutations de résistance sur l'asymétrie de la PR2, propriété importante dans les déformations de la PR2.

Lors de notre seconde étude, nous avons cette fois-ci étudié les impacts de 22 mutations de résistance sur l'ensemble de la PR2 en explorant les réarrangements structuraux localisés au niveau du squelette et des chaînes latérales chez 31 mutants. Après avoir localisé ces changements structuraux, nous avons ensuite étudié l'impact de ces réarrangements sur les propriétés du site de liaison des inhibiteurs et sur l'interface de ces 31 mutants. L'étude d'un large jeu de structures mutantes a permis d'apporter et de confirmer des connaissances structurales sur la PR2. Par exemple nos résultats ont montré que les résidus 8, 30, 53, 81, 82 et 87 ont une conformation fortement conservée qui est importante dans le maintien de la structure du site de liaison. De plus, nous avons identifié que les positions 40–42, 50, 51, 58 sont asymétriques dans l'ensemble des structures sauvages et mutantes. Cette conservation de l'asymétrie montre que c'est une propriété importante de ces résidus. Un des résultats intéressants que nos études ont mis en évidence est qu'une mutation n'a pas les mêmes conséquences sur les deux chaînes du dimère. En effet, nos études ont montré que les mutations induisent des changements structuraux qui sont différents dans les deux chaînes qui peuvent aboutir à des changements d'asymétrie. Ce résultat confirme que les deux chaînes de la PR2 n'ont pas le même comportement malgré leur composition en acides aminés identique, et que la PR2 présente bien une asymétrie.

Concernant l'impact des mutations, nos résultats ont montré que les mutations de résistance étudiées modifient la conformation de certains résidus du squelette peptidique ainsi que des chaînes latérales et qu'une seule mutation pouvait entraîner un grand nombre de réarrangements de forte amplitude, par exemple la mutation K7R. Certaines de ces

déformations induites touchent directement les résidus mutés, des résidus proches des résidus mutés, voire des résidus plus éloignés. Ces résultats mettent en évidence des mouvements coopératifs induits par les mutations qui aboutissent à des déformations structurales localisées loin des régions mutées.

La localisation des réarrangements structuraux induits directement ou indirectement par les mutations montrent que ces dernières ont un impact sur la conformation de régions importantes pour la stabilité de la protéine (l'interface de dimérisation) et pour la fixation des IP (les régions flaps et le site de liaison dont certains résidus impliqués dans l'interaction avec les IP). Notre seconde étude a montré que ces changements structuraux ont pour conséquence une modification de la structure, du volume et/ou de l'hydrophobicité des poches ainsi que la composition en acides aminés des interfaces. L'ensemble de ces résultats suggère trois mécanismes différents pouvant expliquer la résistance acquise du VIH-2 aux IP :

Mécanisme 1 : la modification de l'interface de dimérisation. Nos résultats ont montré que les changements structuraux induits par les mutations K7R, I50L, I54L, G48V, L90M, ou L99F modifient l'interface de dimérisation de la PR2, ce qui pourrait diminuer la stabilité de la PR2. Chez PR1, il a été démontré que la résistance du mutant L90M à l'IDV s'explique en partie par une diminution de la stabilité de la PR2 (Mahalingam *et al.*, 2004).

Mécanisme 2 : l'altération directe de l'interaction avec les IP. Dans notre analyse, les mutations K7R, D30N, I50L/V, I54L, I82F, et I84V ont été identifiées comme induisant des changements structuraux sur des résidus qui sont directement en contact avec les IP via des interactions hydrogènes ou de van der Waals. Ces changements structuraux pourraient directement altérer la fixation des IP. Chez PR1, il a été montré que la résistance du mutant V82A à l'IDV ou au RTV et celle du mutant I84V au DRV s'explique en partie par les changements dans l'interaction de l'IP avec la PR2 par rapport au complexe sauvage (Prabu-Jeyabalan *et al.*, 2003; Lockbaum *et al.*, 2019).

Mécanisme 3 : la modification des propriétés du site de liaison. Nos résultats ont montré que certaines mutations induisent des changements structuraux au niveau des résidus de la poche de liaison ou au niveau de résidus en dehors de la poche de liaison mais qui sont importants pour la conformation du site de liaison. Par exemple, nous avons identifié que les mutations S43T, V47A, K45R, G48V et I82F/M modifient l'hydrophobicité de la poche de liaison des IP. De plus, nous avons montré que les mutations K7R, D30N, V47A, G48V, I50V, I54M/L, I82F/M, I84V et L90M mutations modifient le volume de ce dernier. La modification des propriétés du site de liaison pourrait avoir un impact sur la liaison des IP, et donc altérer leur fixation. Ce mécanisme a été mis en évidence et caractérisé chez plusieurs mutants de PR1 avec les mutations D30N, I50V, V82A/F, I84V, et L90M

(Hong *et al.*, 2000; Agniswamy *et al.*, 2016; Logsdon *et al.*, 2004; Muzammil *et al.*, 2007; Yeddi *et al.*, 2011; Martin *et al.*, 2005; Kovalevsky *et al.*, 2006).

Un mutant peut faire intervenir plusieurs mécanismes. En effet, nos résultats montrent par exemple que la mutation L90M semble à la fois avoir un impact sur la région de dimérisation et les propriétés de la poche de liaison comme chez PR1 (Mahalingam *et al.*, 2004; Kovalevsky *et al.*, 2006).

En conclusion, nos deux études correspondent à une exploration des impacts d'un large nombre de mutations de résistance sur la structure de la PR2. Nos résultats apportent des clés pour mieux comprendre les mécanismes de résistance de ces mutations. Cependant nos études possèdent certaines limites. Dans nos deux études, nous avons mis en place un protocole systématique d'analyse d'une trentaine de mutants basé sur la modélisation des structures des mutantes. Cependant, les différents mutants ne présentent pas le même profil de résistance aux différents IP : I54M, I82F et L90M avec APV, NFV et IDV. Par exemple, le simple mutant I50V est résistant au SQV et au LPV alors qu'il est sensible au TPV et au NFV (Ntemgwa *et al.*, 2007). De plus, dans nos deux études, les modèles des structures mutantes sont construits à partir de la structure du complexe PR2-DRV. Or, il a été montré que le mutant L90M est résistant au saquinavir et présente une très faible résistance au DRV (Raugi *et al.*, 2013).

Ainsi, pour affiner les résultats obtenus et compléter l'analyse il était donc intéressant de prendre en compte les différents profils de résistance des différents inhibiteurs dans notre analyse. Pour cela, une possibilité serait d'utiliser des structures de PR2 complexées à différents médicaments lors de la modélisation. Cependant, il n'existe actuellement dans la PDB que trois structures de la PR2 complexée à un médicament : les complexes PR2-DRV, PR2-APV, et PR2-IDV. À part le DRV, les deux autres IP ne sont pas recommandés par la FDA pour le traitement du VIH-2. Pour étudier l'impact des autres IP, il serait nécessaire de modéliser les structures des complexes PR2-IP par des approches de *docking* (amarrage). Une autre possibilité serait de prendre en compte les profils de résistance de chaque mutant étudié dans l'analyse. Cependant ces données sont lacunaires : il y a peu d'études qui testent la résistance des neuf IP sur les différentes mutations par des approches génotypiques ou par des tests de susceptibilité phénotypique. De plus, certaines études aboutissent à des résultats contradictoires. Par exemple, Raugi *et al.*, 2013 ont montré que la mutation I82F entraîne peu de résistance contre le LPV, ce qui est en contradiction avec les résultats obtenus par Ntemgwa *et al.*, 2007, qui ont trouvé que la mutation I82F a provoqué une augmentation de l'IC₅₀ d'un facteur 36,3 pour le LPV.

Une deuxième limite à notre étude et que nous n'avons pas pris en compte ou très peu la flexibilité de la PR2. En effet, cette protéine est connue comme étant une protéine adoptant au moins deux états (la forme semi-ouverte et la forme fermée). Toutes nos analyses se sont basées sur la construction de modèle de structures mutantes à partir de la forme fermée de la PR2. C'est pourquoi nous avons étudié l'impact des mutations de résistance sur la PR2 en utilisant des simulations de dynamiques moléculaire. Nous nous

sommes limités à l'étude de trois mutants (I54M, I82F, et I54M/I82F) pour limiter les temps de calculs. Notre choix s'est porté sur ces trois mutants car nos études ont montré que ces mutations induisent des réarrangements structuraux au niveau de la poche, les régions flaps et la zone de dimérisation. [Raugi *et al.*, 2013](#) ont par ailleurs montré que la mutation I54M augmente la résistance au DRV. À l'inverse, la mutation I82F augmente la sensibilité au DRV. La combinaison de ces deux mutations I54M/I82F n'ayant quant à elle engendré ni susceptibilité ni résistance au DRV. De fait, la comparaison des trajectoires de dynamique moléculaire de ces trois mutants pourrait améliorer la compréhension des mécanismes de résistance mis en œuvre. Avant de lancer les simulations sur les mutants, nous nous sommes demandés si la structure cristallographique 3EBZ (PR2 sous forme fermée complexée au DRV) était le meilleur choix pour la modélisation des mutants, sachant que dans la PDB il existe deux formes : PR2 semi-ouverte et fermée. Pour répondre à cette question, nous avons comparé la conformation des flaps lors de simulations de dynamique moléculaire de la PR2 fermée à laquelle nous avons retiré le DRV et de la forme semi-ouverte. Pour cela, une simulation de 1 μ s a été lancée pour chacune des deux structures.

Nos résultats ont montré que la structure de départ de la simulation impactait fortement la conformation des flaps : lors de la simulation de la forme fermée (où le DRV a été retiré), la PR2 reste en forme fermée en échantillonnant deux conformations majoritaires : la forme fermée et la forme fermée incurvée. À l'inverse, lors de la simulation de la forme semi-ouverte, la PR2 échantillonne des conformations allant de la forme semi-ouverte à la forme fermée. De plus, nous avons observé que la conformation des flaps observée chez la structure cristallographique semi-ouverte (code PDB : 1HSI) n'est jamais retrouvée au cours des simulations. Ce résultat suggère que cette conformation contient des erreurs qui peuvent s'expliquer par sa résolution moyenne (2,5 Å) et du *packing* cristallin. Suite à ces résultats, nous avons donc choisi d'étudier les mutants I54M, I82F, et I54M/I82F par des simulations de dynamique moléculaire en utilisant les structures modélisées comme structures de départ à partir de la structure cristallographique 3EBZ à laquelle le ligand DRV a été retiré.

Pour chaque structure mutante, nous avons lancé des simulations de dynamique moléculaire de 1 μ s en utilisant le même protocole que celui utilisé pour les structures sauvages. Les analyses que nous avons réalisées correspondent à une exploration préliminaire. Nos résultats ont montré que la présence des deux mutations I54M et I82F individuellement renforcent la stabilité de la protéine alors que ces deux mutations combinées augmentent son instabilité en comparaison avec la structure sauvage. Nos analyses ont également montré que la mutation I54M, seule ou couplée à la mutation I82F, avait un impact sur la flexibilité des flaps et des régions R3 et wall de la chaîne A. Chez le mutant I82F, une augmentation de la flexibilité des régions R3 et wall de la chaîne A est aussi observée. De plus, nous avons aussi montré que la mutation I54M couplée à la mutation I82F a un impact sur le mouvement des flaps et sur la conformation de la région

R2 et elbow de la chaîne A et de la région R3 et wall de la chaîne B. Les régions flap, R2 et wall sont des régions impliquées dans la poche de fixation des IP. Ainsi l'augmentation de leur flexibilité et la modification de leur conformation peut induire une altération de la structure du site de liaison, ainsi que de la fixation des IP. Ces résultats sont en accord avec les résultats obtenus dans notre étude [Laville et al., 2021](#), où nous avons montré que ces deux mutations modifient la structure et les propriétés de la poche de liaison. De plus, nous avons observé que la mutation I82F, seule ou couplée à la mutation I54M, diminue la flexibilité du résidu 32, résidu important pour la fixation spécifique des IP. Ainsi, cette mutation peut altérer directement la fixation des IP. L'ensemble de ces résultats que nous avons présentés sont des résultats préliminaires.

Il serait intéressant de poursuivre l'analyse en comparant plus en détails les mouvements des flaps comme l'ont réalisé [Chen et al., 2014](#) lors de simulation de dynamique moléculaire et en utilisant la technique des modes normaux. De plus, une comparaison des mouvements coopératifs détectés au cours de chaque simulation permettrait de comprendre le lien entre les mutations I54M et I82F et les déformations observées dans les régions R2, elbow, wall et R3 chez le double mutant ainsi que d'étudier l'impact de ces déformations sur le mouvement des flaps. Nos études sur les modèles des mutants ont permis d'identifier que certaines mutations pourraient altérer la fixation des IP soit en induisant des changements conformationnels au niveau des résidus établissant des liaisons avec les IP ou en modifiant les propriétés des poches.

Pour affiner nos résultats, il serait intéressant d'étudier les propriétés du site de liaison au cours des différentes simulations et de comparer ces propriétés aux propriétés du site de liaison de la forme sauvage. Cette étude pourrait être couplée à une analyse du mode de liaison des IP aux différents mutants par des approches de *docking*. Pour cela, des modèles de la PR2 mutante complexée à différents IP pourraient être construits en utilisant le logiciel AutoDock ([Morris et al., 2009](#)) sur les structures des PR2 mutantes que nous avons construites dans ce projet. Ces expériences de *docking* permettraient de prédire la manière dont les IP se fixent sur les mutants de PR2. À partir de ces complexes PR2 mutant-IP, les modes de liaison des IP sur ces structures mutantes seront comparés aux modes de liaison des IP sur les structures sauvages. Là encore se pose la question de la structure de départ pour la modélisation des mutants. Est-ce judicieux de partir de la structure complexée au DRV pour faire des complexes de la PR2 mutante complexée à un autre IP ? Une solution qui pourrait être envisagée est d'utiliser plus de structures extraites de la trajectoire de la forme sauvage fermée que nous avons générée. Ces structures serviraient de point de départ pour les expériences de *docking*. Coupler les approches de dynamique moléculaire et de *docking* permettrait de prendre en compte la flexibilité de la PR2 dans l'analyse du mode de liaison.

Annexe

PPIC est un logiciel librement accessible¹ qui permet de détecter les atomes impliqués dans les interfaces protéine-protéine ou protéine-ligand. L'avantage de cette approche est que l'algorithme est sans paramètre, contrairement aux méthodes classiques comme la méthode du seuil de distance, la méthode basée sur les tessellations de Voronoi et celle basée sur la perte d'accessibilité au solvant lors de la liaison. À l'inverse de ces trois méthodes, PPIC ne nécessite pas de valeur seuil, facteur influençant grandement les résultats. Dans PPIC, l'estimation d'une interface à partir d'une structure 3D consiste à :

1. Calculer l'interface dans A comme l'ensemble non redondant des plus proches voisins de A dans les atomes de B.
2. Calcul de l'interface dans B comme l'ensemble non redondant des plus proches voisins de B dans les atomes de A.

Deux jeux de données ont été utilisés pour tester la méthode : un jeu de données de 1050 homodimères et hétérodimères (Martin *et al.*, 2008), et un jeu de 18 interfaces de PR2. Des *cutoffs* de 3,7 Å et 3,6 Å ont été observés respectivement pour ces deux jeux très différents. Cette valeur est en accord avec l'analyse statistique de McDonald and Thornton, 1994, qui suggère une distance donneur-accepteur inférieure à 3,9 Å et une distance hydrogène-accepteur inférieure à 2,5 Å. En ajoutant la longueur d'une liaison hydrogène à cette dernière valeur ($2,5 + 1,1 = 3,6$ Å), on retrouve notre valeur de 3,6 Å. Pour la méthode de tessellation de Voronoi, on obtient des *cutoffs* de 5,1 Å et 4,6 Å. Pour des atomes lourds, cette distance est trop importante, mais attendue puisque notre méthode de calcul est plus simple et ne tient pas compte des voisins à longue distance de l'interface. Concernant les complexes de PR2, 17 résidus ont été trouvés en contact avec les ligands, ce qui est en accord avec Triki *et al.*, 2018b, dont le calcul de la poche de liaison était basé sur une distance de *cutoff* de 4,5 Å.

L'article de Laville *et al.*, 2018, reproduit dans les pages suivantes, présente ces résultats.

1. Binaires et documentation sur <http://petitjeanmichel.free.fr/itoweb.petitjean.freeware.html>.

A non-parametric method to compute protein–protein and protein–ligands interfaces. Application to HIV-2 protease–inhibitors complexes

Pierre Laville¹, Juliette Martin², Guillaume Launay², Leslie Regad¹, Anne-Claude Camproux¹, Sjoerd de Vries¹ and Michel Petitjean^{1,3,*}

¹Université Paris Diderot, MTi, INSERM UMR-S 973, Paris, France.

²Université Lyon 1, IBCP, MMSB, CNRS UMR 5086, Lyon, France.

³Université Paris Diderot, E-pôle de génoinformatique, Institut Jacques Monod, CNRS UMR 7592, Paris, France.

*To whom correspondence should be addressed.

Motivation: The accurate description of interfaces is needed to identify which residues interact with another molecule or macromolecule. In addition, a data structure is required to compare interfaces within or between families of protein–protein or protein–ligands complexes. In order to avoid many unwanted comparisons, we looked for a parameter free computation of interfaces. This need appeared at the occasion of bioinformatics studies by our research team focusing on HIV-2 protease (PR2) resistance to its inhibitors.

Results: We designed the PPIC software (Protein Protein Interface Computation). It offers three methods of computation of interfaces: (1) our original parameter free method, (2) the Voronoi tessellation approach, and (3) the cutoff distance method. For the latter, we suggest on the basis of 1050 dimers protein–protein interfaces that the optimal cutoff distance is 3.7 Å, or 3.6 Å for a set of 18 PR2–ligand interfaces. We found at most 17 contact residues with PR2 ligands.

Availability: Free binaries and documentation are available through a software repository located at <http://petitjeanmichel.free.fr/itoweb.petitjean.freeware.html>

Contact: petitjean.chiral@gmail.com, michel.petitjean@univ-paris-diderot.fr

1 Introduction

As noticed by [Vangone and Bonvin \(2015\)](#), the number of connections between each pair of proteins is a strong predictor of how tightly the proteins connect to each other. According to [Janin *et al.* \(2008\)](#), [de Vries and Bonvin \(2008\)](#) and [Dequeker *et al.* \(2017\)](#), there are three main methods to compute interfaces between protein chains:

1. The cutoff method (implemented in PPIC).
2. The loss of ASA (accessible surface area) upon binding.
3. The Voronoi tessellation method (implemented in PPIC).

The cutoff method assumes that two residues are in contact when they have a pair of heavy atoms separated by a distance smaller than the cutoff distance. This cutoff distance may be set to 4.5 Å between non H atoms ([Gao and Skolnick, 2010](#); [Esque *et al.*, 2011](#); [Triki *et al.*, 2018](#)), to 5 Å from side-chain centers of mass ([Viloria *et al.*, 2017](#)), or to 8 Å between C_α atoms ([Esque *et al.*, 2011](#)). The choice

of the cutoff can significantly influence the result (de Vries and Bonvin, 2008). This method critically relies on the choice of atomic radii, which strongly depend on how they are defined (Bondi, 1964; Allinger *et al.*, 1994). The values recommended in the literature (Gavezzotti, 1983) induce the existence of intersections of 6 spheres or more. These intersections above 3 or 4 spheres are ignored by a number of published methods, inducing in turn large errors in surface calculations (Petitjean, 1994, 2013).

In its original variant, the Voronoi tessellation method is parameter-free (Cazals *et al.*, 2006; Bouvier *et al.*, 2009). It was implemented in PROVAT (Gore *et al.*, 2005). The full mathematical presentation of Voronoi diagrams and their calculation algorithms are out of the scope of this paper (see Edelsbrunner, 1987). We may just say that each atom lies inside a convex polyhedral cell having its polygonal faces located at mid distance from its neighbouring atoms. Thus, two atoms are neighbours if their Voronoi cells share a common face. Extensions taking in account atomic spheres were proposed (Esque *et al.*, 2011; Mahdavi *et al.*, 2012), but these latter are no more parameter-free.

2 Methods

Our original method, which is implemented in PPIC, is a variant of the one of Cerisier *et al.* (2017). The input is a complex with two partners (molecule or macromolecule), A and B. The algorithm has no parameter. It has two steps:

1. Generate the interface in A as the non-redundant set of all nearest neighbours in A of the atoms of B.
2. Generate the interface in B as the non-redundant set of all nearest neighbours in B of the atoms of A.

The interface has two parts, one in A and one in B. Each one is a subset of the interface that would be output by the Voronoi tessellation method (proof: see Eppstein *et al.*, 1997). However our method is much simpler than the latter, and it does not generate neighbours at long distances in the interface, while large Voronoi cells can induce such long distances. Furthermore, looking at interacting atom pairs among the nearest neighbours makes more sense than looking at interacting atom pairs at farther distances than the nearest neighbours.

3 Results and discussion

We considered the dimers set of Martin *et al.* (2008), containing 1050 protein homo- and heterodimers. We compared the 1050 couples of interfaces with the ones obtained with the cutoff distance method, for a cutoff ranging from 2 to 6 Å. For the full database, computing these interfaces took about 288 min on a workstation Ubuntu 18.04 (16 Intel® Xeon® CPUs, 3 GHz). This computing time was roughly the same for the three methods implemented in PPIC. The subsets of atoms defining the interfaces were compared with the symmetric difference metric (Deza and Deza, 2009). Similarly, we compared the 36 couples of interfaces for the set of 18 PR2-ligand interfaces from Triki *et al.* (2018). Figure 1 shows that the closest results for

the two methods were observed at a cutoff of 3.7 Å for the dimers set, and at 3.6 Å for the PR2 complexes set. This is a remarkable agreement since these two data sets are of different nature. It suggests an optimal cutoff value of 3.6–3.7 Å between heavy atoms. This optimal value is in agreement with the statistical analysis of McDonald and Thornton (1994): the distance donor–acceptor should be less than 3.9 Å and the distance hydrogen–acceptor should be less than 2.5 Å. So, assuming a covalent H bond length of 1.1 Å, the distance donor–acceptor should be in fact at most $2.5 + 1.1 = 3.6$ Å, which is our suggested value. For the Voronoi tessellation method, the closest results were observed at the respective cutoffs of 5.1 Å and 4.6 Å. In the case of distances between heavy atoms, these cutoff values are too large, as predicted in section 2.

For the PR2 complexes, we found that at most 17 residues are in contact with the ligands: Arg 8, Leu 23, Asp 25, Gly 27, Ala 28, Asp 29, Asp 30, Ile 32, Ile 46, Val 47, Gly 48, Gly 49, Ile 50, Phe 53, Pro 81, Ile 82 and Ile 84. All were found in the binding pocket computed by the consensus method of Triki *et al.* (2018) based on a 4.5 Å cutoff distance. This confirms the pertinence of our parameter-free method.

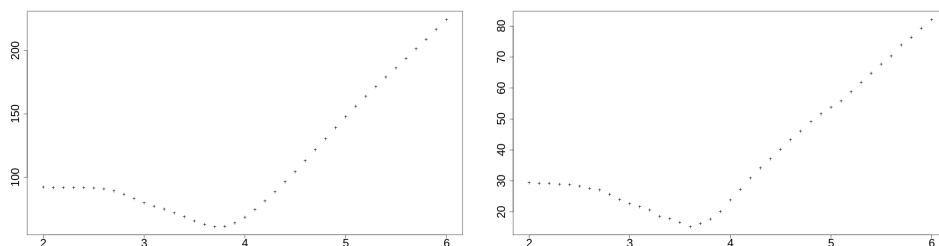


Figure 1: Mean symmetric difference distances in function of the cutoff value, for the 1050 dimers data set (on the left) and for the 18 PR2 complexes (on the right).

References

- Allinger,N., Zhou,X., Bergsma,J. (1994) Molecular mechanics parameters. *J. Mol. Struct.*, **312**, 69–83.
- Bondi,A. (1964) Van der Waals volumes and radii. *J. Phys. Chem.*, **68**, 441–451.
- Bouvier,B., Grünberg,R., Nilges,M., Cazals,F. (2009) Shelling the Voronoi interface of protein–protein complexes reveals patterns of residue conservation, dynamics, and composition. *Proteins*, **76**, 677–692.
- Cazals,F., Proust,F., Bahadur,R.P., Janin,J. (2006) Revisiting the Voronoi description of protein–protein interfaces. *Prot. Sci.*, **15**, 2082–2092.
- Cerisier,N., Regad,L., Triki,D., Camproux,A.-C., Petitjean,M. (2017) Cavity versus ligand shape descriptors: Application to urokinase binding pockets. *J. Comput. Biol.*, **24**, 1134–1137.

- Dequeker,C., Laine,E., Carbone,A. (2017) INTerface Builder: A fast protein–protein interface reconstruction tool. *J. Chem. Inf. Model.*, **57**, 2613–2617.
- de Vries,S.J., Bonvin,A.M.J.J. (2008) How proteins get in touch: Interface prediction in the study of bio-molecular complexes. *Curr. Prot. Peptide Sci.*, **9**, 394–406.
- Deza,M.M., Deza,E. (2009) *Encyclopedia of Distances*. Springer, New-York, p.46.
- Edelsbrunner,H. (1987) Voronoi Diagrams. In: Brauer,W., Rozenberg,G. and Salomaa,A. (eds), *Algorithms in Combinatorial Geometry*. Springer-Verlag, Berlin, chap. 13, pp. 293–334.
- Eppstein,D., Paterson,M.S., Yao,F.F. (1997) On nearest-neighbor graphs. *Discrete Comput. Geom.*, **17**, 263–282.
- Esque,J., Oguey,C., de Brevern,A.G. (2011) Comparative analysis of threshold and tessellation methods for determining protein contacts. *J. Chem. Inf. Model*, **51**, 493–507.
- Gao,M., Skolnick,J. (2010) iAlign: a method for the structural comparison of protein–protein interfaces. *Bioinformatics*, **26**, 2259–2265.
- Gavezzotti,A. (1983) The calculation of molecular volumes and the use of volume analysis in the investigation of structured media and of solid-state organic reactivity. *J. Am. Chem. Soc.*, **105**, 5220–5225.
- Gore,S.P., Burke,D.F., Blundell,T.L. (2005) PROVAT: a tool for Voronoi tessellation analysis of protein structures and complexes. *Bioinformatics*, **21**, 3316–3317.
- Janin,J., Bahadur, R.P., Chakrabarti,P. (2008) Protein-protein interaction and quaternary structure. *Quarterly Rev. Biophys.*, **41**, 133–180.
- Mahdavi,S., Mohades,A., Yazdi,A.S., Jahandideh,S., Masoudi-Nejad,A. (2012) Computational analysis of RNA-protein interaction interfaces via the Voronoi diagram. *J. Theor. Biol.*, **293**, 55–64.
- Martin,J., Regad,L., Etchebest,C., Camproux,A.-C. (2008) Taking advantage of local structure descriptors to analyze interresidue contacts in protein structures and protein complexes. *Proteins*, **73**, 672–689.
- Martin,J. (2014) Benchmarking protein–protein interface predictions: Why you should care about protein size. *Proteins*, **82**, 1444–1452.
- McDonald,I.K., Thornton,J.M. (1994) Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.*, **238**, 777–793.
- Petitjean,M. (1994) On the Analytical Calculation of van der Waals Surfaces and Volumes: Some Numerical Aspects. *J. Comput. Chem.*, **15**, 507–523.
- Petitjean,M. (2013) Spheres unions and intersections and some of their applications in molecular modeling. In: Mucherino,A., Lavor,C., Liberti,L., Maculan,N. (Eds.), *Distance Geometry: Theory, Methods, and Applications*, Springer, New-York, chap. 4, pp. 61–83.

- Triki,D., Cano Contreras,M.E., Flatters,D., Visseaux,B., Descamps,D., Camproux,A.-C., Regad,L. (2017) Analysis of the HIV-2 protease's adaptation to various ligands: characterization of backbone asymmetry using a structural alphabet. *Sci. Rep.*, **8**, 710.
- Vangone,A., Bonvin,A.M.J.J. (2015) Contacts-based prediction of binding affinity in protein-protein complexes. *eLIFE*, **4**, e07454.
- Viloria,J.S., Allega,M.F., Lambrugh,M., Papaleo,E. (2017) An optimal distance cutoff for contact-based protein structure networks using side-chain centers of mass. *Sci. Rep.*, **7**, 2838.

Bibliographie

- [Acevedo *et al.*, 2012] Acevedo, O.; Ambrose, Z.; Flaherty, P.T.; Aamer, H.; Jain, P. Sambasivarao, S.V. Identification of HIV inhibitors guided by free energy perturbation calculations. *Curr. Pharm. Des.* **2012**, *18*[9], 1199–1216.
<https://doi.org/10.2174/138161212799436421>.
- [Agniswamy *et al.*, 2016] Agniswamy, J.; Louis, J.M.; Roche, J.; Harrison, R.W.; Weber, I.T. Structural studies of a rationally selected multi-drug resistant HIV-1 protease reveal synergistic effect of distal mutations on flap dynamics. *PLoS One* **2016**, *11*[12].
<https://doi.org/10.1371/journal.pone.0168616>.
- [Arts and Hazuda, 2012] Arts, E.J.; Hazuda, D.J. HIV-1 Antiretroviral drug therapy. *Cold Spring Harb. Perspect. Med.* **2012**, *2*[4], 1–23.
<https://doi.org/10.1101/cshperspect.a007161>.
- [Ba *et al.*, 2013] Ba, S.; Dia-Badiane, N.M.; Hawes, S.E.; Degeuenovo, L.F.; Sall, F.; Ndour, C.T.; Faye, K.; Traoré, F.; Touré, M.; Sy, M.P.; Raugi, D.N.; Kiviat, N.B.; Smith, R.A.; Seydi, M.; Sow, P.S.; Gottlieb, G.S. Infection à VIH-2 au Sénégal: échecs virologiques et résistances aux antirétroviraux (ARV). *Pan. Afr. Med.* **2019**, *33*, 222.
<https://doi.org/10.11604/pamj.2019.33.222.15771>.
- [Barré-Sinoussi *et al.*, 1983] Barré-Sinoussi, F.; Chermann, J. C.; Rey, F.; Nugeyre, M. T.; Chamaret, S.; Gruest, J.; Dauguet, C.; Axler-Blin, C.; Vezinet-Brun, F.; Rouzioux, C.; Rozenbaum, W.; Montagnier, L. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science* **1983**, *220*[4599], 868–871.
<https://doi.org/10.1126/science.6189183>
- [Bénard *et al.*, 2009] Bénard, A.; Damond, F.; Campa, P.; Peytavin, G.; Descamps, D.; Lascoux-Combes, C.; Taieb, A.; Simon, F.; Autran, B.; Brun-Vézinet, F.; Chêne, G.; Matheron, S.; ANRS CO5 HIV-2 Cohort Study Group. Good response to lopinavir/ritonavir-containing antiretroviral regimens in antiretroviral-naïve HIV-2-infected patients. *AIDS* **2009**, *23*[9], 1171–1173.
<https://doi.org/10.1097/QAD.0b013e32832949f0>.
- [Berendsen *et al.*, 1995] Berendsen, H.J.C.; van der Spoel, D.; van Drunen, R. GROMACS: A message-passing parallel molecular dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*[1–3], 43–56.
[https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E).
- [Berman *et al.*, 2000] Berman, H.M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T.N.; Weissig, H.; Shindyalov, I.N.; Bourne, P.E. The protein data bank. *Nucleic Acids*

- Res.* **2000**, *28*[1], :235–242.
<https://doi.org/10.1093/nar/28.1.235>.
- [Bibollet-Ruche *et al.*, 2004] Bibollet-Ruche, F.; Bailes, E.; Gao, F.; Pourrut, X.; Barlow, K.L.; Clewley, J.P.; Mwenda, J.M.; Langat, D.K.; Chege, G.K.; McClure, H.M.; Mpoudi-Ngole, E.; Delaporte, E.; Peeters, M.; Shaw, G.M.; Sharp, P.M.; Hahn, B.H. New simian immunodeficiency virus infecting De Brazza’s Monkeys (*cercopithecus neglectus*): evidence for a *cercopithecus* monkey virus clade. *J. Virology*, **2004**, *78*[14], 7748–7762.
<https://doi.org/10.1128/JVI.78.14.7748-7762.2004>.
- [Boggetto and Reboud-Ravaux, 2001] Boggetto, N.; Reboud-Ravaux, M. HIV protease: structure and development of inhibitors. *Virologie* **2001**, *5*[6], 9–18.
https://www.jle.com/en/revues/vir/e-docs/la_protease_du_vih_et_le_developpement_d_inhibiteurs_260184/article.phtml.
- [Borrego *et al.*, 2012] Borrego, P.; Calado, R.; Marcelino, J.M.; Bártolo, I.; Rocha, C.; Cavaco-Silva, P.; Doroana, M.; Antunes, F.; Maltez, F.; Caixas, U.; Barroso, H.; Taveira, N. Baseline susceptibility of primary HIV-2 to entry inhibitors. *Antiviral Ther.* **2012**, *17*[3], 565–570.
<https://doi.org/10.3851/IMP1996>.
- [Borrel *et al.*, 2015] Borrel, A.; Regad, L.; Xhaard, H.; Petitjean, M.; Camproux, A.-C. PockDrug: a model for predicting pocket druggability that overcomes pocket estimation uncertainties. *J. Chem. Inf. Model* **2015**, *55*[4], 882–895.
<https://doi.org/10.1021/ci5006004>.
- [Brown, 2006] Brown, J.H. Breaking symmetry in protein dimers: designs and functions. *Protein Science*, **2006**, *15*[1], 1–13.
<https://doi.org/10.1110/ps.051658406>.
- [Camproux *et al.*, 2004] Camproux, A.C.; Gautier, R.; Tufféry, P. A hidden markov model derived structural alphabet for proteins. *J. Mol. Biol.* **2004**, *339*[3], 591–605.
<https://doi.org/10.1016/j.jmb.2004.04.005>.
- [Cavaco-Silva *et al.*, 2013] Cavaco-Silva, J.; Aleixo, M.J.; Van Laethem, K.; Faria, D.; Valadas, E.; de Fátima Gonçalves M.; Gomes, P.; Vandamme, A.M.; Cunha, C.; Camacho, R.J. Portuguese HIV-2 resistance study group mutations selected in hiv-2-infected patients failing a regimen including atazanavir. *Antimicrob. Agents Chemother.* **2013**, *68*[1], 190–192.
<https://doi.org/10.1093/jac/dks363>.
- [Cerisier *et al.*, 2017] Cerisier, N.; Regad, L.; Triki, D.; Camproux, A.-C.; Petitjean M. Cavity versus ligand shape descriptors: application to urokinase binding pockets. *J. Comput. Biol.* **2017**, *24*[11], 1134–1137.
<https://doi.org/10.1089/cmb.2017.0061>.

- [Charpentier *et al.*, 2010] Charpentier, C.; Larrouy, L.; Collin, G.; Damond, F.; Matheron, S.; Chêne, G.; Nie, T.; Schinazi, R.; Brun-Vézinet, F.; Descamps, D.; the French ANRS HIV-2 Cohort (ANRS CO 05 VIH-2). In-vitro phenotypic susceptibility of HIV-2 clinical isolates to the integrase inhibitor S/GSK1349572. *AIDS* **2010**, *24*[17], 2753–2755.
<https://doi.org/10.1097/QAD.0b013e32833f9e36>.
- [Charpentier *et al.*, 2011] Charpentier, C.; Roquebert, B.; Delelis, O.; Larrouy, L.; Matheron, S.; Tubiana, R.; Karmochkine, M.; Duval, X.; Chêne, G.; Storto, A.; Collin, G.; Bnard, A.; Damond, F.; Mouscadet, J.F.; Brun-Vézinet, F.; Descamps, D. Hot spots of integrase genotypic changes leading to HIV-2 resistance to raltegravir. *Antimicrob. Agents Chemother.* **2011**, *55*[3], 1293–1295.
<https://doi.org/10.1128/AAC.00942-10>.
- [Charpentier *et al.*, 2013a] Charpentier, C.; Camacho, R.; Ruelle, J.; Kaiser, R.; Eberle, J.; Gürtler, L.; Pironti, A.; Stürmer, M.; Brun-Vézinet, F.; Kaiser, R.; Descamps, D.; Obermeier, M. HIV-2EU: supporting standardized HIV-2 drug resistance interpretation in Europe. *Clin. Infect. Dis.* **2013**, *56*[11], 1654–1658.
<https://doi.org/10.1093/cid/civ572>.
- [Charpentier *et al.*, 2013b] Charpentier, C.; Visseaux, B.; Bénard, A.; Peytavin, G.; Damond, F.; Roy, C.; Taieb, A.; Chêne, G.; Matheron, S.; Brun-Vézinet, F.; Descamps, D.; ANRS CO5 HIV-2 Cohort. Transmitted drug resistance in French HIV-2-infected patients. *AIDS* **2013**, *27*[10], 1671–1674.
<https://doi.org/10.1097/QAD.0b013e32836207f3>.
- [Chen *et al.*, 1994] Chen, Z.; Li, Y.; Chen, E.; Hall, D.L.; Darke, P.L.; Culberson, C.; Shafer, J.A.; Kuo, L.C. Crystal structure at 1.9-Å 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*[42], 26344–26348.
[https://doi.org/10.1016/S0021-9258\(18\)47199-2](https://doi.org/10.1016/S0021-9258(18)47199-2).
- [Chen *et al.*, 2014] 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.
<http://doi.org/10.1038/srep06872>.
- [Cihlar and Fordyce, 2016] Cihlar, T.; Fordyce, M. Current status and prospects of HIV treatment. *Curr. Op. Virol.* **2016**, *18*, 50–56.
<http://doi.org/10.1016/j.coviro.2016.03.004>.
- [Clavel *et al.*, 1986] Clavel, F.; Guetard, D.; Brun-Vézinet, F.; Chamaret, S.; Rey, M.; Santos-Ferreira, M.; Laurent, A.; Dauguet, C.; Katlama, C.; Rouzioux, C.; et al. Isolation of a new human retrovirus from west african patients with aids. *Science*

1986, 233[4761], 343–346.

<https://doi.org/10.1126/science.2425430>

[Colson *et al.*, 2004] Colson, P.; Henry, M.; Tourres, C.; Lozachmeur, D.; Gallais, H.; Gastaut, J.A.; Moreau, J.; Tamalet, C. Polymorphism and drug-selected mutations in the protease gene of human immunodeficiency virus type 2 from patients living in southern France. *J. Clin. Microbiol.* **2004**, 42[2], 570–527.

<https://doi.org/10.1128/jcm.42.2.570-577.2004>.

[Cox and Cox, 2008] Cox, M.A.A.; Cox, T.F. Multidimensional Scaling. In: *Handbook of Data Visualization*. Springer, Berlin, Heidelberg, 2008.

https://doi.org/10.1007/978-3-540-33037-0_14.

[Damond *et al.*, 2005] Damond, F.; Brun-Vezinet, F.; Matheron, S.; Peytavin, G.; Campa, P.; Pueyo, S.; Mammano, F.; Lastere, S.; Farfara, I.; Simon, F.; Chene, G.; Descamps, D. 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. *J. Clin. Microbiol.* **2005**, 43[1], 484–487.

<https://doi.org/10.1128/JCM.43.1.484-487.2005>.

[Desbois *et al.*, 2008] Desbois, D.; Roquebert, B.; Peytavin, G.; Damond, F.; Collin, G.; Bénard, A.; Campa, P.; Matheron, S.; Chêne, G.; Brun-Vézinet, F.; Descamps, D.; French ANRS HIV-2 Cohort (ANRS CO 05 VIH-2). In vitro phenotypic susceptibility of human immunodeficiency virus type 2 clinical isolates to protease inhibitors. *Antimicrob. Agents Chemother.* **2008**, 52[4], 1545–1548.

<https://doi.org/10.1128/AAC.01284-07>.

[Ghosn *et al.*, 2018] Ghosn, J.; Taiwo, B.; Seedat, S.; Autran, B.; Katlama, C. HIV. *Lancet* **2018**, 392[10148], 685–697.

[https://doi.org/10.1016/S0140-6736\(18\)31311-4](https://doi.org/10.1016/S0140-6736(18)31311-4).

[Goodsell and Olson, 2000] Goodsell, D.S.; Olson, A.J. Structural symmetry and protein function. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, 29[1], 105–153.

<https://doi.org/10.1146/annurev.biophys.29.1.105>.

[Guerois *et al.*, 2002] Guerois, R.; Nielsen, J.E.; Serrano, L. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J. Mol. Biol.* **2002**, 320[2], 369–387.

[https://doi.org/10.1016/S0022-2836\(02\)00442-4](https://doi.org/10.1016/S0022-2836(02)00442-4).

[Gustchina and Weber, 1991] Gustchina, A.; Weber, I.T. Comparative analysis of the sequences and structures of HIV-1 and HIV-2 proteases. *Prot. Struct. Funct. Bioinf.* **1991**, 10[4], 325–339.

<https://doi.org/10.1002/prot.340100406>.

[Hong *et al.*, 2000] Hong, L.; Zhang, X.C.; Hartsuck, J.A.; Tang, J. Crystal structure of an in vivo HIV-1 protease mutant in complex with saquinavir: insights into the

- mechanisms of drug resistance. *Protein Sci.* **2000**[10], 9[10], 1898–1904.
<https://doi.org/10.1110/ps.9.10.1898>.
- [Hoog *et al.*, 1996] Hoog, S.S.; Towler, E.M.; Zhao, B.; Doyle, M.L.; Debouck, C.; Abdel-Meguid, S.S. Human immunodeficiency virus protease ligand specificity conferred by residues outside of the active site cavity. *Biochemistry* **1996**, 35[32], 10279–10286.
<https://doi.org/10.1021/bi960179j>.
- [Hornak *et al.*, 2006] Hornak, V.; Okur, A.; Rizo, R.C.; Simmerling, C. HIV-1 protease flaps spontaneously open and reclose in molecular dynamics simulations. *Proc. Natl. Acad. Sci. USA* **2006**, 103[4], 915–920.
<https://doi.org/10.1073/pnas.0508452103>.
- [Hubbard and Thornton, 1993] Hubbard, S.; Thornton, J.
NACCESS, Computer Program. *Department of Biochemistry Molecular Biology*, University College London, **1993**.
<https://sbgrid.org/software/titles/naccess>.
- [Jallow *et al.*, 2009] Jallow, S.; Alabi, A.; Sarge-Njie, R.; Peterson, K.; Whittle H.; Corrah, T.; Jaye, A.; Cotten, M.; Vanham, G.; McConkey, S.J.; Rowland-Jones, S.; Janssens W. Virological response to highly active antiretroviral therapy in patients infected with human immunodeficiency virus type 2 (hiv-2) and in patients dually infected with hiv-1 and hiv-2 in the Gambia and emergence of drug-resistant variants. *J. Clin. Microbiol.* **2009**, 47[7], 2200–2208.
<https://doi.org/10.1128/JCM.01654-08>.
- [Kanki *et al.*, 1994] Kanki, P.J.; Travers, K.U.; Marlink, R.G.; Essex, M.E.; MBoup, S.; Gueye-NDiaye, A.; Siby, T.; Thior, I.; Sankalé, J.-L.; Hsieh, C.C.; Hernandez-Avila, M.; NDiaye, I. Slower heterosexual spread of HIV-2 than HIV-1. *Lancet* **1994**, 343[8903], 943–946.
[https://doi.org/10.1016/S0140-6736\(94\)90065-5](https://doi.org/10.1016/S0140-6736(94)90065-5).
- [Keele *et al.*, 2006] Keele, B.F.; Van Heuverswyn, F.; Li, Y.; Bailes, E.; Takehisa, J.; Santiago, M.L.; Bibollet-Ruche, F.; Chen, Y.; Wain, L.V.; Liegeois, F.; Loul, S.; Ngole, E.M.; Bienvenue, Y.; Delaporte, E.; Brookfield, J.F.Y.; Sharp, P.M.; Shaw, G.M.; Peeters, M.; Hahn, B.H. Chimpanzee Reservoirs of Pandemic and Nonpandemic HIV-1. *Science* **2006**, 313[5786], 523–526.
<https://doi.org/10.1126/science.1126531>.
- [Kim *et al.*, 2004] Kim, D.E., Chivian, D.; Baker, D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Res.* **2004** 32[Web Server issue], W526–W531.
<https://doi.org/10.1093/nar/gkh468>.
- [Koh *et al.*, 2011] Koh, Y.; Matreyek, K.A.; Engelman, A. Differential sensitivities of retroviruses to integrase strand transfer inhibitors. *ASM J.* **2011**, 85[7], 3677–3682.
<https://doi.org/10.1128/JVI.02541-10>.

- [Kovalevsky *et al.*, 2006] Kovalevsky, A.Y.; Tie, Y.; Liu, F.; Boross, P.I.; Wang, Y.-F.; Leshchenko, S.; Ghosh, A.K.; Harrison, R.W.; Weber, I.T. Effectiveness of nonpeptide clinical inhibitor TMC-114 on HIV-1 protease with highly drug resistant mutations D30N, I50V, and L90M. *J. Med. Chem.* **2006**, *49*[4], 1379–1387.
<https://doi.org/10.1021/jm050943c>.
- [Kovalevsky *et al.*, 2008] Kovalevsky, A.Y.; Louis, J.M.; Aniana, A.; Ghosh, A.K.; Weber, I.T. Structural evidence for effectiveness of darunavir and two related antiviral inhibitors against HIV-2 Protease. *J. Mol. Biol.* **2008**, *384*[1], 178–192.
<https://doi.org/10.1016/j.jmb.2008.09.031>.
- [Larrouy *et al.*, 2013] Larrouy, L.; Vivot, A.; Charpentier, C.; Bénard, A.; Visseaux, B.; Damond, F.; Matheron, S.; Chene, G.; Brun-Vézinet, F.; Descamps, D, ANRS CO5 HIV-2 Cohort. Impact of gag genetic determinants on virological outcome to boosted lopinavir-containing regimen in hiv-2-infected patients. *AIDS* **2013**, *27*[1], 69–80.
<https://doi.org/10.1097/QAD.0b013e32835a10d8>.
- [Laville *et al.*, 2018] Laville, P.; Martin, J.; Launay, G.; Regad, L.; Camproux, A.-C.; de Vries, S.; Petitjean, M. A non-parametric method to compute protein-protein and protein-ligands interfaces. Application to HIV-2 protease-inhibitors complexes. *bioRxiv* **2018**, 498923.
<https://doi.org/10.1101/498923>.
- [Laville *et al.*, 2020] Laville, P.; Fartek, S.; Cerisier, N.; Flatters, D.; Petitjean, M.; Regad, L. Impacts of drug-resistance mutations on the structural asymmetry of the HIV-2 protease. *BMC Mol. Cell Biol.* **2020**, *21*, 46.
<https://doi.org/10.1186/s12860-020-00290-1>.
- [Laville *et al.*, 2021] Laville, P.; Petitjean, M.; Regad, L. Structural impact of drug-resistance mutations appearing in HIV-2 protease. *Molecules* **2021**, *26*[3], 611.
<https://doi.org/10.3390/molecules26030611>.
- [Larrouy *et al.*, 2013] Larrouy, L.; Vivot, A.; Charpentier, C.; Bénard, A.; Visseaux, B.; Damond, F.; Matheron, S.; Chene, G.; Brun-Vézinet, F.; Descamps, D, ANRS CO5 HIV-2 Cohort. Impact of gag genetic determinants on virological outcome to boosted lopinavir-containing regimen in hiv-2-infected patients. *AIDS* **2013**, *27*[1], 69–80.
<https://doi.org/10.1097/QAD.0b013e32835a10d8>.
- [Lindorff-Larsen *et al.*, 2010] Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J.L.; Dror, R.O.; Shaw, D.E. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Prot. Struct. Funct. Bioinf.* **2010**, *78*[8], 1950–1958.
<https://doi.org/10.1002/prot.22711>.
- [Lockbaum *et al.*, 2019] Lockbaum, G.J.; Leidner, F.; Rusere, L.N.; Henes, M.; Kosovrasti, K.; Nachum, G.S.; Nalivaika, E.A.; Ali, A.; Yilmaz, N.K.; Schiffer, C.A. Structural adaptation of darunavir analogues against primary

- mutations in HIV-1 protease. *ACS Infect. Dis.* **2019**, *5*[2], 316–325. <https://doi.org/10.1021/acsinfecdis.8b00336>.
- Correction: *ACS Infect. Dis.* **2019**, *5*[6], 1044.
<https://doi.org/10.1021/acsinfecdis.9b00098>.
- [Logsdon *et al.*, 2004] Logsdon, B.C.; Vickrey, J.F.; Martin, P.; Proteasa, G.; Koepke, J.I.; Terlecky, S.R.; Wawrzak, Z.; Winters, M.A.; Merigan, T.C.; Kovari, L.C. Crystal structures of a multidrug-resistant human immunodeficiency virus type 1 protease reveal an expanded active-site cavity. *J. Virol.* **2004**, *78*[6], 3123–3132.
<https://doi.org/10.1128/jvi.78.6.3123-3132.2004>.
- [Mahalingam *et al.*, 1999] Mahalingam, B.; Louis, J.M.; reed, C.C.; Adomat, J.M.; Krouse, J.; Wang, Y.-F.; Harrison, R.W.; Weber, I.T. Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. *Eur. J. Biochem.* **1999**, *263*[1], 238–244.
<https://doi.org/10.1046/j.1432-1327.1999.00514.x>.
- [Mahalingam *et al.*, 2004] Mahalingam, B.; Wang, Y.; Boross, P.; Tozser, J.; Louis, J.; Harrison, R.; Weber, I. Crystal structures of HIV protease V82A and L90M mutants reveal changes in the indinavir-binding site. *Eur. J. Biochem.* **2004**, *271*[8], 1516–1524.
<https://doi.org/10.1111/j.1432-1033.2004.04060.x>.
- [Mahdi *et al.*, 2015] Mahdi, M.; Szojka, Z.; Mótyán, J.A.; Tózsér J. Inhibition profiling of retroviral protease inhibitors using an hiv-2 modular system. *Viruses* **2015**, *7*[12], 6152–6162.
<https://doi.org/10.3390/v7122931>.
- [Maier *et al.*, 2015] Maier, J.A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K.E.; Simmerling, C. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*[8], 3696–3713.
<https://doi.org/10.1021/acs.jctc.5b00255>.
- [Martin *et al.*, 2005] Martin, P.; Vickrey, J.F.; Proteasa, G.; Jimenez Y.L.; Wawrzak, Z.; Winters, M.A.; Merigan, T.C.; Kovari, L.C. Wide-open 1.3 Å structure of a multidrug-resistant HIV-1 protease as a drug target. *Structure* **2005**, *13*[12], 1887–1895.
<https://doi.org/10.1016/j.str.2005.11.005>.
- [Martin *et al.*, 2008] Martin, J.; Regad, L.; Etchebest, C.; Camproux, A.-C. Taking advantage of local structure descriptors to analyze interresidue contacts in protein structures and protein complexes. *Prot. Struct. Funct. Bioinf.* **2008**, *73*[3], 672–689.
<https://doi.org/10.1002/prot.22091>.
- [Masse *et al.*, 2007] Masse, S.; Lu, X.; Dekhtyar, T.; Lu, L.; Koev, G.; Gao, F.; Mo, H.; Kempf, D.; Bernstein, B.; Hanna, G.J.; Molla, A. In vitro selection

- and characterization of human immunodeficiency virus type 2 with decreased susceptibility to lopinavir. *Antimicrob Agents Chemother.* **2007**, *51*[9], 3075–3080.
<https://doi.org/10.1128/AAC.00146-07>.
- [McDonald and Thornton, 1994] McDonald, I.K.; Thornton, J.M. Satisfying hydrogen bonding potential in proteins. *J. Mol. Biol.* **1994**, *238*[5], 777–793.
<https://doi.org/10.1006/jmbi.1994.1334>.
- [Menéndez-Arias and Alvarez, 2014] Menéndez-Arias, L.; Alvarez, M. Antiretroviral therapy and drug resistance in human immunodeficiency virus type 2 infection. *Antiviral Res.* **2014**, *102*, 70–86.
<https://doi.org/10.1016/j.antiviral.2013.12.001>.
- [Morris *et al.*, 2009] Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. Autodock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Computational Chemistry* **2009**, *30*[16], 2785–2791.
<https://doi.org/10.1002/jcc.21256>.
- [Mulichak *et al.*, 1993] Mulichak, A.M.; Hui, J.O.; Tomasselli, A.G.; Henrikson, R.L.; Curry, K.A.; Tomich, C.S.; Thaisrivongs, S.; Sawyer, T.K.; Watenpaugh, K.D. 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*[18], 13103–13109.
[https://doi.org/10.1016/S0021-9258\(19\)38625-9](https://doi.org/10.1016/S0021-9258(19)38625-9).
- [Murail, 2018] Murail, S. gromacs_py: a gromacs python wrapper. *ZENODO*, **2018**.
<https://doi.org/10.5281/zenodo.1455734>.
- [Muzammil *et al.*, 2007] Muzammil, S.; Armstrong, A.A.; Kang, L.W.; Jakalian, A.; Bonneau, P.R.; Schmelmer, V.; Amzel, L.M.; Freire, E. Unique thermodynamic response of tipranavir to human immunodeficiency virus type 1 protease drug resistance mutations. *J. Virol.* **2007**, *81*[10], 5144–5154.
<https://doi.org/10.1128/JVI.02706-06>.
- [NIAID, 2018] NIAID. HIV virus replication cycle, **2018**.
<https://www.niaid.nih.gov/diseases-conditions/hiv-replication-cycle>.
- [Nijhuis *et al.*, 2001] Nijhuis, M.; Deeks, S.; Boucher, C. Implications of antiretroviral resistance on viral fitness. **2001**, *14*[1], 23–28.
<https://doi.org/10.1097/00001432-200102000-00005>.
- [Ntemgwa *et al.*, 2007] Ntemgwa, M.; Brenner, B.G.; Oliveira, M.; Moisi, D.; Wainberg, M.A. Natural polymorphisms in the human immunodeficiency virus type 2 protease can accelerate time to development of resistance to protease inhibitors. *Antimicrob. Agents Chemother.* **2007**, *51*[2], 604–610.
<https://doi.org/10.1128/AAC.00870-06>.

- [Ollitrault *et al.*, 2018] Ollitrault, G.; Fartek, S.; Descamps, D.; Camproux, A.-C.; Visseaux, B.; Regad, L. Characterization of HIV-2 protease structure by studying its asymmetry at the different levels of protein description. *Symmetry* **2018**, *10*[11], 644.
<https://doi.org/10.3390/sym10110644>.
- [ONUSIDA, 2017] ONUSIDA. Des données de l'ONUSIDA précises et crédibles sur l'épidémie de VIH : la clé de voûte de la riposte au sida. 10 juillet **2017**.
https://www.unaids.org/fr/resources/featurestories_archive.
- [ONUSIDA, 2020] ONUSIDA. Fiche d'information 2020 – Dernières statistiques sur l'état de l'épidémie de sida. **2020**.
<https://www.unaids.org/fr/resources/fact-sheet>.
- [Painter and Collins, 2019] Painter, M.M.; Collins, K.L. HIV and retroviruses. In Schmidt, T.M., editor, *Encyclopedia of Microbiology* (4th edition), **2019**, pages 613–628. Academic Press, Oxford.
<https://doi.org/10.1016/b978-0-12-801238-3.66202-5>.
- [Perrier *et al.*, 2019] Perrier, M.; Castain, L.; Regad, L.; Todesco, E.; Landman, R.; Visseaux, B.; Yazdanpanah, Y.; Rodriguez, C.; Joly, V.; Calvez, V.; Marcelin, A.; Descamps, D.; Charpentier, C. HIV-1 protease, Gag and gp41 baseline substitutions associated with virological response to a PI-based regimen. *J. Antimicrob. Chemother.* **2019**, *74*[6], 1679–1692.
<https://doi.org/10.1093/jac/dkz043>.
- [Peterson *et al.*, 2011] Peterson, K.; Jallow, S.; Rowland-Jones, S.L.; de Silva, T.I. Antiretroviral therapy for HIV-2 infection: recommendations for management in low-Resource settings. *AIDS Res. Treat.* **2011**, *2011* 463704.
<https://doi.org/10.1155/2011/463704>.
- [Pokorna *et al.*, 2009] Pokorná, J.; Machala, L.; Řezáčová, P.; Konvalinka, J. Current and novel inhibitors of HIV protease. *Viruses* **2009**, *1*[3], 1209–1239.
<https://doi.org/10.3390/v1031209>.
- [Prabu-Jeyabalan *et al.*, 2000] 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*[5], 1207–1220.
<https://doi.org/10.1006/jmbi.2000.4018>.
- [Prabu-Jeyabalan *et al.*, 2003] Prabu-Jeyabalan, M.; Nalivaika, E.A.; King, N.M.; Schiffer, C.A. Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: structural insights for better antiviral therapy. *J. Virol.* **2003**, *77*[2], 1306–1315.
<https://doi.org/10.1128/jvi.77.2.1306-1315.2003>.
- [Priestle *et al.*, 1995] Priestle, J.P.; Fässler, A.; Rösel, J.; Tintelnot-Blomley, M.; Strop, P.; Grütter, M.G. 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*[4], 381–389.
[https://doi.org/10.1016/s0969-2126\(01\)00169-1](https://doi.org/10.1016/s0969-2126(01)00169-1).
- [PyMOL] The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
<https://pymol.org/2/support.html>.
- [Rabiner, 1989] Rabiner, L. A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE* **1989**, *77*[2], 257–286.
<https://doi.org/10.1109/5.18626>.
- [Raugi *et al.*, 2013] Raugi, D.N.; Smith, R.A.; Ba, S.; Toure, M.; Traore, F.; Sall, F.; Pan, C.; Blankenship, L.; Montano, A.; Olson, J.; Dia Badiane, N.M.; Mullins, J.I.; Kiviat, N.B.; Hawes, S.E.; Sow, P.S.; Gottlieb, G.S. 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*[6], 2751–2760.
<https://doi.org/10.1128/AAC.00405-13>.
- [Raugi *et al.*, 2016] Raugi, D.N.; Smith, R.A.; Gottlieb, G.S.; The University of Washington-Dakar HIV-2 Study group. Four amino acid changes in HIV-2 protease confer class-wide sensitivity to protease inhibitors. *J. Virol.* **2015**, *90*[2], 1062–1069.
<https://doi.org/10.1128/JVI.01772-15>.
- [Regad *et al.*, 2010] Regad, L.; Martin, J.; Nuel, G.; Camproux, A.C. Mining protein loops using a structural alphabet and statistical exceptionality. *BMC Bioinformatics* **2010**, *11*, 75.
<https://doi.org/10.1186/1471-2105-11-75>.
- [Regad *et al.*, 2017] Regad, L.; Chéron, J.B.; Triki, D.; Senac, C.; Flatters, D.; Camproux, A.-C. Exploring the potential of a structural alphabet-based tool for mining multiple target conformations and target flexibility insight. *PLOS ONE* **2017**, *12*[8], e0182972.
<https://doi.org/10.1371/journal.pone.0182972>.
- [Ren *et al.*, 2002] Ren, J.; Bird, L.E.; Chamberlain, P.P.; Stuart, D.I.; Stammers, D.K. Structure of HIV-2 reverse transcriptase at 2.35-Å resolution and the mechanism of resistance to non-nucleoside inhibitors. *Proc. Natl. Acad. Sci. USA* **2002**, *99*[22], 14410–14415.
<https://doi.org/10.1073/pnas.222366699>.
- [Rodés *et al.*, 2006] Rodés, B.; Sheldon, J.; Toro, C.; Jiménez, V.; Alvarez, M.A.; Soriano, V. Susceptibility to protease inhibitors in HIV-2 primary isolates from patients failing antiretroviral therapy. *J. Antimicrob. Chemother.* **2006**, *57*[4], 709–713.
<https://doi.org/10.1093/jac/dkl034>.
- [Sadiq and Fabritiis, 2010] Sadiq, S.K.; Fabritiis, G.D. Explicit solvent dynamics and energetics of HIV-1 protease flap opening and closing. *Prot. Struct. Funct. Bioinf.*

- 2010**, 78[14], 2873–2885.
<https://doi.org/10.1002/prot.22806>.
- [Sherry *et al.*, 2021] Sherry, D; Worth, R.; Ismail, Z.S.; Sayed, Y. Cantilever-centric mechanism of cooperative non-active site mutations in HIV protease: Implications for flap dynamics. *J. Mol. Graph. Model.* **2021** 106, 107931.
<https://doi.org/10.1016/j.jmglm.2021.107931>.
- [Smith *et al.*, 2011] Smith, R.A.; Raugi, D.N.; Kiviat, N.B.; Hawes, S.E.; Mullins, J.I.; Sow, P.S.; Gottlieb, G.S. Phenotypic susceptibility of HIV-2 to raltegravir: integrase mutations Q148R and N155H confer raltegravir resistance. *AIDS* **2011**, 25[18], 2235–2241.
<https://doi.org/10.1097/QAD.0b013e32834d8e52>.
- [Souquière *et al.*, 2001] Souquière, S.; Bibollet-Ruche, F.; Robertson, D.L.; Makuwa, M.; Apetrei, C.; Onanga, R.; Kornfeld, C.; Plantier, J.-C.; Gao, F.; Abernethy, K.; White, L.J.T.; Karesh, W.; Telfer, P.; Wickings, E.J.; Maucière, P.; Marx, P.A.; Barré-Sinoussi, F.; Hahn, B.H.; Müller-Trutwin, M.C.; Simon, F. Wild Mandrillus sphinx are carriers of two types of lentivirus. *J. Virol.* **2001** 75[15], 7086–7096.
<https://doi.org/10.1128/JVI.75.15.7086-7096.2001>.
- [Storto *et al.*, 2018] Storto, A.; Viseaux, B.; Bertine, M.; Le Hingrat, Q.; Collin, G.; Damond, F.; Khuong, M.A.; Blum, L.; Tubiana, R.; Karmochkine, M.; Cazanave, C.; Matheron, S.; Descamps, D.; Charpentier, C.; ANRS HIV-2 CO5 cohort. Minority resistant variants are also present in hiv-2-infected antiretroviral-naive patients. *J. Antimicrob. Chemother.* **2018**, 73[5], 1173–1176.
<https://doi.org/10.1093/jac/dkx530>.
- [Tie *et al.*, 2012] Tie, Y.; Wang, Y.-F.; Boross, P.I.; Chiu, T.-Y.; Ghosh, A.K.; Tozser, J.; Louis, J.M.; Harrison, R.W.; Weber, I.T. Critical differences in HIV-1 and HIV-2 protease specificity for clinical inhibitors. *Protein Sci.* **2012**, 21[3], 339–350.
<https://doi.org/10.1002/pro.2019>.
- [Tong *et al.*, 1993] Tong, L.; Pav, S.; Pargellis, C.; Dô, F.; Lamarre, D.; Anderson, P.C. 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[18], 8387–8391.
<https://doi.org/10.1073/pnas.90.18.8387>.
- [Tong *et al.*, 1995] Tong, L.; Pav, S.; Mui, S.; Lamarre, D.; Yoakim, C.; Beaulieu, P.; Anderson, P.C. Crystal structures of HIV-2 protease in complex with inhibitors containing the hydroxyethylamine dipeptide isostere. *Structure* **1995**, 3[1], 33–40.
[https://doi.org/10.1016/s0969-2126\(01\)00133-2](https://doi.org/10.1016/s0969-2126(01)00133-2).
- [Tóth and Borics, 2006] Toth, G.; Borics, A. Closing of the flaps of HIV-1 protease induced by substrate binding: a model of a flap closing mechanism in retroviral

- aspartic proteases. *Biochemistry* **2006**, *45*[21], 6606–6614.
<https://doi.org/10.1021/bi060188k>.
- [Triki, 2018] Triki D. Etude des mécanismes de résistance naturelle de la protéase du VIH-2 (PR2) contre les inhibiteurs de protéase par des approches de bioinformatique structurale pour la recherche d’inhibiteurs ciblés de PR2. *Thèse de doctorat*, Université Paris Diderot, 23 avril **2018**.
- [Triki *et al.*, 2018a] Triki, D.; Billot, T.; Visseaux, B.; Descamps, D.; Flatters, D.; Camproux, A.C.; Regad, L. Exploration of the effect of sequence variations located inside the binding pocket of HIV-1 and HIV-2 proteases. *Sci. Rep.* **2018**, *8*[1], 5789.
<https://doi.org/10.1038/s41598-018-24124-5>.
- [Triki *et al.*, 2018b] Triki, D.; Cano Contreras, M.E.; Flatters, D.; Visseaux, B.; Descamps, D.; Camproux, A.C.; 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*[1], 710.
<https://doi.org/10.1038/s41598-017-18941-3>.
- [Triki *et al.*, 2019] Triki, D.; Fartek, S.; Visseaux, B.; Descamps, D.; Camproux, A.C.; Regad, L. Characterization of the HIV-2 protease’s structural variability upon diverse ligand binding. *J. Biomol. Struct. Dyn.* **2019**, *37*[17], 4658–4670. <https://doi.org/10.1080/07391102.2018.1562985>.
- [van Tienen and van der Loeff, 2018] van Tienen C.; van der Loeff, M.S. (2018) Epidemiology of HIV-2 infection in West Africa. In: Hope T.J., Richman D.D., Stevenson M. (eds) *Encyclopedia of AIDS*. Springer, New York, **2018**.
https://doi.org/10.1007/978-1-4939-7101-5_34.
- [Visseaux *et al.*, 2016] Visseaux, B.; Damond, F.; Matheron, S.; Descamps, D.; Charpentier, C. Hiv-2 molecular epidemiology. *Infect. Genet. Evol.* **2016**, *46*, 233–240.
<https://doi.org/10.1016/j.meegid.2016.08.010>.
- [Worobey *et al.*, 2010] Worobey, M.; Telfer, P.; Souquière, S.; Hunter, M.; Coleman, C.A.; Metzger, M.J.; Reed, P.; Makuwa, M.; Hearn, G.; Honarvar, S.; Roques, P.; Apetrei, C.; Kazanji, M.; Marx, P.A. Island biogeography reveals the deep history of SIV. *Science* **2010**, *329*[5998], 1487–1487.
<https://doi.org/10.1126/science.1193550>.
- [Yeddi *et al.*, 2011] Yeddi, R.S.; Proteasa, G.; Martinez, J.L.; Vickrey, J.; Martin, P.D.; Wawrzak, Z.; Liu, Z.; Kovari, I.A.; Kovari, L.C. Contribution of the 80s loop of HIV-1 protease to the multidrug-resistant mechanism: crystallographic study of MDR769 HIV-1 protease variants. *Acta Crystallogr. D: Biol. Crystallogr.* **2011**, *67*[6], 524–532.
<https://doi.org/10.1107/S0907444911011541>.