



HAL
open science

Caractérisation de complexes d'inhibiteurs à visée thérapeutique avec des métalloprotéines : cas de l'intégrase du VIH-1

Perla El Darazi

► To cite this version:

Perla El Darazi. Caractérisation de complexes d'inhibiteurs à visée thérapeutique avec des métalloprotéines : cas de l'intégrase du VIH-1. Chimie théorique et/ou physique. Sorbonne Université; Université Saint-Joseph (Beyrouth), 2021. Français. NNT : 2021SORUS209 . tel-03519954

HAL Id: tel-03519954

<https://theses.hal.science/tel-03519954>

Submitted on 10 Jan 2022

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.

Sorbonne Université

Université Saint-Joseph de Beyrouth

Ecole Doctorale Chimie Physique et Chimie Analytique de Paris Centre (ED388)

Ecole Doctorale de Sciences, d'Ingénierie et de Technologies

Laboratoire de Chimie Théorique UMR7616

Laboratoire Structure et Interactions des Macromolécules

**Caractérisation de complexes d'inhibiteurs à visée
thérapeutique avec des métalloprotéines : Cas de
l'intégrase du VIH-1**

Par Perla EL DARAZI

Thèse de doctorat de Sciences de la Vie / Modélisation Moléculaire

Présentée et soutenue publiquement le 27 octobre 2021

Devant le jury composé de :

Rapporteurs : Professeur Latifa NEGADI
Docteur Thérèse MALLIAVIN

Examineurs : Professeur Vincent Maréchal
Professeur Mireille KALLASSY AWAD

Directeurs de thèse : Professeur Jean-Philip PIQUEMAL
Docteur Zeina HOBAIKA

Co-directeurs de thèse : Professeur Richard MAROUN
Docteur Nohad GRESH

Au Bon Dieu tout Puissant,

A mes parents,

A mes sœurs, Rita et Carla,

A Jamie et Steven,

A mon oncle Antoine.

Remerciements

Je tiens tout d'abord à remercier le doyen de la Faculté des Sciences de l'Université Saint-Joseph de Beyrouth ; Professeur Richard MAROUN et le directeur du laboratoire de Chimie Théorique ; Professeur Jean-Philip PIQUEMAL, pour m'avoir accueillie au sein de ces laboratoires.

Je remercie également les organismes financeurs de ce projet de thèse ; le conseil de la recherche de l'Université Saint-Joseph de Beyrouth, le Conseil National de la Recherche Scientifique au Liban (CNRS-L/USJ), l'institut français-campus France au Liban et le programme franco-libanais CEDRE.

Je remercie le Professeur Latifa NEGADI ainsi que Dr Thérèse MALLIAVIN qui ont accepté d'être les rapporteurs de cette thèse et m'ont fait l'honneur de juger ce travail. De même, je désire exprimer mes sincères remerciements aux Professeurs Vincent MARECHAL et Mireille KALLASSY AWAD qui m'ont gratifiée en acceptant d'examiner ce travail.

Je tiens à exprimer mes plus vifs remerciements aux directeurs de cette thèse ; Professeur Jean-Philip PIQUEMAL et Dr Zeina HOBAIKA, pour m'avoir donné l'opportunité d'effectuer des travaux de recherche combinant plusieurs disciplines, allant de la chimie théorique à la modélisation moléculaire, la virologie et la conception de médicaments. Malgré mes humbles connaissances dans ce domaine, vous m'avez fait confiance et vous m'avez bien formée.

Merci énormément Professeur PIQUEMAL pour votre soutien. C'était un honneur pour moi de travailler au sein de votre équipe. Vous m'avez transmis votre passion pour la recherche et vous m'avez toujours soutenue. J'ai pu profiter grâce à vous, de diverses opportunités professionnelles qui ont enrichi ma carrière scientifique. Je vous serais toujours reconnaissante. Merci aussi Dr HOBAIKA pour votre soutien et encadrement. Vous étiez toujours disponible et à l'écoute. Merci de m'avoir choisie pour continuer le travail sur ce projet de thèse. Votre confiance en moi me sera toujours précieuse. Merci également pour le temps que vous m'avez consacré malgré vos nombreuses responsabilités.

Je désire aussi remercier chaleureusement mes co-directeurs de thèse; le Professeur Richard MAROUN et le Dr Nohad GRESH. Merci énormément Dr GRESH pour votre aide et votre présence continue tout au long de cette thèse. Vous m'avez initiée aux notions de chimie quantique et de modélisation moléculaire ainsi qu'à l'utilisation du logiciel SIBFA, tout en me transmettant votre passion pour la recherche. Je n'oublierai jamais vos conseils qui m'ont guidée tout au long de mon travail. Merci Professeur MAROUN pour votre confiance et de m'avoir donné l'opportunité de travailler sur un sujet qui vous tient à cœur. Vos conseils scientifiques me seraient toujours précieux.

Finalement, j'adresse mes sincères remerciements aux membres de l'équipe du Laboratoire de Chimie Théorique, du Laboratoire de Structures et Interactions des Macromolécules de l'Université Saint-Joseph et du Laboratoire de Génie Biomédical à l'Université de Texas à Austin. Un grand merci au Professeur Pengyu REN pour avoir accepté de m'accueillir au sein de son Laboratoire à Austin pour un séjour de recherche, et pour m'avoir transmis ses connaissances diverses et enrichissantes. Merci beaucoup pour votre soutien et vos précieux conseils. Merci aux

Drs Léa EL KHOURY, Krystal EL HAGE et Sehr NASSEEM-KHAN, pour avoir été toujours prêtes à répondre à mes questions et à m'apprendre à manipuler les logiciels de calcul et de visualisation. Merci aux Drs Louis LAGARDERE et Frédéric CELERSE pour leur aide et soutien. Vous étiez toujours présents au cours de ces travaux. Merci aux Drs Chengwen LIU et Francis JING pour avoir toujours été à l'écoute et pour avoir enrichi mes connaissances dans le domaine de la modélisation moléculaire et de la chimie théoriques. Merci à mes collègues Dina EL AHDAB, Théo Jaffrelot-INIZAN et Brandon WALKER pour avoir toujours été à mes côtés et pour votre intérêt accordé à ce travail.

Table des matières

Remerciements	3
Tables des matières.....	5
Liste des abréviations.....	10
Symboles des acides aminés.....	14
Résumé en français de la thèse	15
Résumé en anglais de la thèse	16
Chapitre I : Introduction générale au projet	17
Chapitre II : Contexte Scientifique.....	24
Partie A : Conception <i>in silico</i> des médicaments	25
A.I.Introduction générale	26
A.II.Le processus de développement des médicaments	26
A.II.1. L'évaluation préclinique.....	27
A.II.2. L'évaluation clinique	31
A.III. La conception rationnelle de médicaments	32
A.III.1. La conception de médicaments basée sur la structure	34
Partie B : SIDA et VIH	38
B.I. La pandémie du SIDA	39
B.II. Le virus d'Immunodéficience Humaine de type 1 (VIH-1)	41
B.II.1. Structure virale	41
B.II.2. Génome du VIH-1	43
B.II.3. Cycle répliatif	44
B.III. Intégrase et mécanisme d'intégration	46
B.III.1. Structure de l'intégrase	47
B.III.1.1. Le domaine N-terminal.....	49
B.III.1.2. Le domaine central	50
B.III.1.3. Le domaine C-terminal	52

B.III.2. Mécanisme d'action de l'intégrase	53
B.III.2.1. Maturation en 3' de l'ADN viral	53
B.III.2.2. Etape de transfert de brin	53
B.III.3. Interactions de l'intégrase avec l'ADN viral.....	55
B.III.3.1. Régions de l'intégrase reconnues par l'ADN viral	55
B.III.3.2. Régions de l'ADN viral reconnues par l'intégrase	58
B.IV. Mécanisme d'inhibition de l'intégrase	59
B.IV.1. Thérapies anti-VIH-1.....	59
B.IV.1.1. Les inhibiteurs de l'entrée du VIH-1	60
B.IV.1.2. Les inhibiteurs de la fusion virus-cellule	61
B.IV.1.3. Les inhibiteurs de la transcription inverse	62
B.IV.1.4. Les inhibiteurs de l'activité protéolytique.....	63
B.IV.2. Inhibition de l'intégrase virale.....	64
B.IV.2.1. Les inhibiteurs dicétoniques (DKAs)	64
B.IV.2.1.1. Le mécanisme d'action des INSTIs.....	66
B.IV.2.1.2. Raltégravir	68
B.IV.2.1.3. Elvitégravir	70
B.IV.2.1.4. Dolutégravir	72
B.IV.2.2. Emergence des mutations de résistance contre les INSTIs.....	75
Conclusions du chapitre II	86
Approche de travail développée dans cette thèse : rationalisation par modélisation	
moléculaire	86
Chapitre III : Démarches et outils.....	91
Partie A : La Chimie Computationnelle : de la mécanique quantique à la mécanique	
moléculaire	92
A.I. Notions de base de chimie quantique	93
A.I.1. Théorie d'Hartree-Fock	93
A.I.2. Méthodes post-Hartree-Fock.....	95
A.I.3. Théorie de la Fonctionnelle de la Densité	96

A.I.4. Les fonctions de bases	97
A.I.5. L'erreur de superposition de base	97
A.II. Méthodes interprétatives quantiques.....	98
A.II.1. Méthodes de décomposition d'énergie	99
A.II.1.1. Décomposition d'énergie : Approche RVS.....	99
A.II.1.2. Décomposition d'énergie : Approche ALMOEDA	101
A.II.2. Méthodes d'analyse topologique.....	103
A.II.2.1. Contours des potentiels électrostatiques moléculaires	103
A.III. Mécanique moléculaire : approches classiques et polarisables.....	104
A.III.1. Champs de force classiques.....	105
A.III.2. Champs de force polarisables.....	109
A.III.2.1. Champ de force polarisable : SIBFA	111
A.III.2.1.1. Formulation du potentiel SIBFA.....	112
A.III.2.1.2. Calcul de l'énergie multipolaire	113
A.III.2.1.3. Calcul de l'énergie d'échange-répulsion.....	114
A.III.2.1.4. Calcul de l'énergie de polarisation.....	115
A.III.2.1.5. Calcul de l'énergie de transfert de charge.....	116
A.III.2.1.6. Calcul de l'énergie de dispersion	117
A.III.2.2. Champ de force polarisable : AMOEBA	117
A.III.2.2.1. Formulation du potentiel AMOEBA	118
A.III.2.2.2. Calcul de l'énergie de vdW.....	119
A.III.2.2.3. Calcul de l'énergie électrostatique	119
A.III.2.2.4. Calcul de l'énergie de polarisation.....	120
Partie B : La dynamique moléculaire conventionnelle	121
B.I. Principes de la dynamique moléculaire classique.....	123
B.I.1. Conditions de calcul	124
B.I.1.1. Echelles de temps	124
B.I.1.2. Les ensembles.....	125
B.I.1.3. Les conditions périodiques	126

B.II. Nos choix des systèmes biologiques simulés du VIH-1	127
B.III. Notre Protocole de dynamique moléculaire conventionnelle appliqué	130
Partie C : Calculs d'énergie libre	135
C.I. Notions théoriques générales sur le calcul d'énergie libre alchimique	136
C.II. Cycle thermodynamique alchimique	139
C.II.1. Cycle thermodynamique alchimique des énergies libres absolues.....	140
C.II.2. Types de contraintes.....	146
C.III. Méthodes d'estimation des énergies libres des simulations à l'équilibre.....	149
C.III.1. Approches de perturbation.....	150
C.III.1.1. Méthode FEP	150
C.III.1.2. Méthodes BAR et MBAR	151
Conclusions du chapitre III	154
Chapitre IV. Résultats et Discussion	162
Partie A : Conception et évaluation des nouveaux noyaux d'halobenzènes issus du DTG	163
A.I. Introduction générale de la partie A	164
A.II. Conception des nouveaux noyaux halogénés à partir du DTG	166
A.III. Evaluation de l'énergie d'interaction des nouveaux noyaux halogénés avec les bases terminales de l'ADN viral du VIH-1	170
A.III.1. Calculs mécanique et quantiques corrélés de l'énergie d'interaction totale du complexe C ₁₆ -G ₄ -nouveau noyau	171
A.III.2. Décomposition de l'énergie d'interaction totale des composés.....	180
A.IV. Evaluation qualitative de l'interaction des nouveaux noyaux halogénés avec les bases terminales de l'ADN viral du VIH-1	190
A.IV.1. Détermination des contours de potentiel électrostatique.....	190
A.V. Conclusions du chapitre IV partie A	193
Partie B : Simulations par dynamique moléculaire des complexes biologiques cibles du VIH-1	196
B.I. Introduction générale de la partie B.....	197
B.II. Présentation des systèmes à simuler	198
B.III. Analyse des simulations de dynamique moléculaire conventionnelle	203

B.III.1. Etude des interactions ADN viral - Intégrase	204
B.III.2. Etude des interactions ADN viral -Inhibiteurs	219
B.III.3. Etude des interactions Intégrase –Inhibiteurs	229
B.III.3.1. Evaluation de la distance entre des résidus clés de l’intégrase virale et les inhibiteurs	230
B.III.3.2. Evaluation de la distance entre les résidus causant des mutations de résistance primaires et secondaires chez l’intégrase du VIH-1 et nos inhibiteurs.....	240
B.IV.Conclusions du chapitre III partie B	252
Partie C : Calculs d’énergie libre alchimique absolue des interactions au sein des complexes simulés du VIH-1	256
C.I.Introduction générale de la partie C.....	257
C.II.Notre protocole pour le calcul des énergies libres alchimiques absolues	258
C.III.Résultats des calculs des énergies libres absolues alchimiques	266
C.III.1.Calculs des énergies libres absolues alchimiques des complexes ADN-Inhibiteurs	266
C.III.2.Calculs des énergies libres absolues alchimiques des complexes ADN-Intégrase- Inhibiteurs	269
C.III.3. Calculs des énergies libres absolues alchimiques des complexes ADN-DTG et Intégrase-DTG	272
C.IV. Conclusions du chapitre III partie C	276
Chapitre V. Conclusions et Perspectives	279
V.I. Conclusions des travaux de thèse	280
V.II. Perspectives des travaux de thèse	283
Chapitre VI. Annexes	285

Liste des abréviations

A	Adénine
ABFE	Absolute Binding Free Energy
ADN	Acide DésoxyriboNucléique
ADNv	Acide DésoxyriboseNucléique viral
AMBER	Assisted Model Building with Energy Refinement
AMOEBA	Atomic Multipole Optimized Energetics for Biomolecular simulation
ARN	Acide RiboNucléique
ARV	Antirétroviraux
BAR	Bennett Acceptance Ratio
BSSE	Basis Set Superposition Error
C	Cytosine
Cα	Carbone alpha
CCD	Catalytic Core Domain
cc-pVTZ	correlation-consistent polarized Valence-only Triple-Zeta
CCR5	C-C chemokine Receptor type 5
CD4⁺	Cluster of Differentiation 4
CHARMM	Chemistry at HARvard Molecular Mechanics
CPI	Complexe de Pré-Intégration
CTD	C-Terminal Domain
CXCR4	C-X-C chemokine Receptor type 4
DKA	DiKetoAcid
DFT	Density Functional Theory
DTG	Dolutégravir
EC₉₀	Concentration associated with 90% Efficacy
EMA	European Medicines Agency

Env	Envelope
EVG	Elvitégravir
FEP	Free Energy Perturbation
FDA	U.S. Food and Drug Administration
G	Guanine
Gag	Group-Specific Antigen
GAMESS	General Atomic and Molecular Electronic Structure System
gp120	glycoprotéine 120
gp41	glycoprotéine 41
GPU	Graphics Processing Unit
GROMOS	GRoningen MOlecular Simulation
GTO	Gaussian Type Orbitals
HAART	Highly Active AntiRetroviral Therapy
HF	Hartree-Fock
HTH	Hélice-Tour-Hélice
HTHi	Hélice-Tour- α inversé
IC₅₀	half maximal Inhibitory Concentration
IF	Inhibiteurs de Fusion
IN	Intégrase
INTI	Inhibiteurs Nucléosidiques de la Transcriptase Inverse
INNTI	Inhibiteurs Non-Nucléosidiques de la Transcriptase Inverse
INSTI	Integrase Strand Transfer Inhibitors
IP	Inhibiteur de Protéase
LCAO	Linear Combination of Atomic Orbitals
LTR	Long Terminal Repeats
MBAR	Multistate Bennett Acceptance Ratio
MD	Molecular Dynamics

MEP	Molecular Electrostatics Potential
MM	Molecular Mechanics
NTD	N-Terminal Domain
OA	Orbitales Atomiques
OM	Orbitales Moléculaires
PDB	Protein Data Bank
PFV	Prototype Foamy Virus
PIC	Pre-Integration Complex
Pol	Polymérase
PR	Protéase Virale
QC	Quantum Chemistry
QM	Quantum Mechanics
RAL	Raltégravir
RMN	Résonnance Magnétique Nucléaire
RT	Reverse Transcriptase
RVS	Reduced Variational Space
SAPT-DFT	Symmetry Adapted Perturbation Theory – Density Functional Theory
SCF	Self Consisted Field
SIBFA	SUM of Interaction Between Fragments Ab-initio
SIDA	Syndrome d'ImmunoDéficine Humaine Acquise
ST	Strand Transfer
STO	Slater Type Orbitals
T	Thymine
TI	Thermodynamic Integration
vdW	van der Waals
VIH-1	Virus d'Immunodéficine Humaine de type 1
VIH-2	Virus d'Immunodéficine Humaine de type 2

VIS Virus d'Immunodéficience Simienne
WT Wild Type

Symboles des acides aminés

A	Ala	Alanine
C	Cys	Cystéine
D	Asp	Acide Aspartique (ou Aspartate)
E	Glu	Acide Glutamique (ou Glutamate)
F	Phe	Phénylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Méthionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Sérine
T	Thr	Thréonine
V	Val	Valine
W	Trp	Tryptophane
Y	Tyr	Tyrosine

Résumé en français de la thèse

Le Virus de l'Immunodéficience Humaine de type 1 (VIH-1); agent causal du Syndrome d'Immunodéficience Acquise (SIDA) possède une enzyme virale essentielle à son cycle répliatif ; l'Intégrase (IN). Elle permet l'intégration du génome viral dans les cellules cibles humaines, via deux étapes ; le « 3'processing » suivi de l'étape de transfert de brin. Par conséquent, l'intégrase du VIH-1 constitue une cible thérapeutique majeure. De nos jours, les thérapies anti-IN ont recours à l'utilisation d'inhibiteurs appartenant à la famille des dicétoacides (DKAs) qui agissent à l'interface de l'intasome Intégrase-ADN viral, inhibant spécifiquement l'étape de transfert de brin, ce qui leur a valu le nom d'inhibiteurs de transfert de brin (INSTIs). L'INSTI le plus efficace à ce jour est le Dolutégravir (DTG), mais malgré son succès initial, ce médicament a causé le développement de mutations de résistance au niveau de l'intégrase chez les patients traités, compromettant l'efficacité à long terme des thérapies anti-VIH. Par conséquent, nous visons dans ces travaux de thèse, à renforcer les interactions des DKAs et notamment du DTG avec l'ADN viral, à l'interface IN-ADN afin de contourner le besoin d'interactions spécifiques avec la protéine et surmonter ainsi les mutations de résistance. Nous avons commencé par l'optimisation du noyau halogéné du DTG, étant donné que cette structure effectue des contacts spécifiques et directes avec les bases terminales de l'ADN viral ; C₁₆ et G₄. Cette optimisation s'est basée sur des substitutions d'halogènes et de groupements donneurs et attracteurs d'électrons. L'efficacité de nos nouvelles structures halogénées substituées, par rapport au DTG, a été ensuite évaluée par calcul de leur énergie d'interaction intermoléculaire (ΔE) avec les noyaux de C₁₆ et G₄, via des méthodes de calculs basés sur la mécanique quantique et la mécanique moléculaire polarisable. Une évaluation supplémentaire a été faite par détermination des contours de potentiel électrostatique pour chaque noyau afin d'évaluer l'effet de nos substitutions sur la répartition électronique et donc sur les interactions de ces derniers avec les bases terminales virales. Suite à la conception et l'évaluation des nouveaux dérivés halogénés à partir du DTG, nous avons extrapolé notre étude à une échelle plus grande, par le suivi du comportement de nos nouveaux inhibiteurs pris entièrement au sein de l'intasome viral en présence de Mg²⁺ et ceci via des simulations de dynamique moléculaire polarisable en solvant explicite. Ces simulations nous ont permis d'étudier en détails les interactions de nos nouvelles molécules avec les différents partenaires ; IN et ADN viral et de les comparer au mode d'action du DTG. De plus, nous avons simulé le DTG avec la protéine et l'ADN pris séparément, afin de comparer son interaction avec chacun des deux et de mettre en relief par la suite sa meilleure interaction avec l'ADN viral expliquant ainsi sa moindre sensibilité aux mutations de résistance de l'IN. La dernière partie de cette thèse repose sur la détermination des énergies libres de Gibbs (ΔG) pour les systèmes que nous avons simulés ; ADN-Intégrase-drogues halogénées, ADN-DTG et Intégrase-DTG, afin de confirmer tout d'abord l'interaction plus stable du DTG avec l'ADN viral par rapport à l'IN puis à démontrer que nos nouvelles substitutions ont renforcé les interactions des nouveaux inhibiteurs avec l'intasome viral et notamment avec l'ADN, permettant la génération de nouveaux inhibiteurs de transfert de brin capables de contourner le problème de résistance.

Mots-clés : VIH-1, SIDA, Intégrase, ADN viral, Dicétoacides, Inhibiteurs de transfert de brin, Dolutégravir, noyau halogéné, mécanique quantique, mécanique moléculaire polarisable, dynamique moléculaire, énergie d'interaction intermoléculaire, énergie libre de Gibbs.

Résumé en anglais de la thèse

The Human Immunodeficiency Virus type 1 (HIV-1); causative agent of the Acquired Immunodeficiency Syndrome (AIDS) has a viral enzyme that is essential to its replicative cycle; the Integrase (IN). It allows the integration of the viral genome into human target cells, via two steps; the 3' processing followed by the strand transfer step. Therefore, the HIV-1 integrase constitutes a major therapeutic target. Nowadays, anti-IN therapies resort to the usage of inhibitors belonging to the diketoacid family (DKAs) which act at the interface of the integrase-viral DNA intasome, specifically inhibiting the strand transfer step, earning thus the name of strand transfer inhibitors (INSTIs). The most effective INSTI to date is Dolutegravir (DTG), but despite its initial success, this drug has caused the development of resistance mutations at the integrase level in patients, compromising the long-term efficacy of anti-drug therapies. Therefore, we aim to strengthen the interactions of DKAs and in particular of DTG with the viral DNA, at the IN-DNA interface in order to bypass the need for specific interactions with the protein and thus overcome the resistance mutations. We started with the optimization of the halogenated ring of DTG, since this structure makes specific and direct contacts with the terminal bases of viral DNA; C₁₆ and G₄. This optimization was based on substitutions of halogens and of electron donor and attractor groups. The efficiency of our new halogenated substituted structures, compared to DTG, was then evaluated by calculating their intermolecular interaction energy (ΔE) with the C₁₆ and G₄ rings, via quantum mechanical and molecular mechanical methods based on polarisable potentials. An additional evaluation was made by determining their electrostatic potential contours in order to assess the effect of our substitutions on the electron distribution for each ring and therefore on the interactions of the latter with the viral terminal bases. Following the design and evaluation of our new halogenated derivatives from DTG, we extrapolated our study to a larger scale, by monitoring the behavior of our new inhibitors taken entirely, within the viral intasome in the presence of Mg²⁺, via molecular dynamics simulations in explicit solvent with a polarizable potential. These simulations allowed us to study in detail the interactions of our new molecules with the different partners; IN and viral DNA and compare them to the mode of action of DTG. In addition, we simulated DTG with each partner alone in order to compare its interaction with each one highlighting subsequently its better interaction with viral DNA, which allows it to be insensitive to IN resistance mutations. The last part of this thesis is based on the determination of Gibbs free energies (ΔG) for the simulated systems; DNA-Integrase-halogenated drugs, DNA-DTG and Integrase-DTG in order to first confirm the more stable interaction of DTG with viral DNA compared to IN and then to demonstrate that our new substitutions reinforced the interactions of the new inhibitors with the viral intasome and in particular with DNA, allowing the generation of new strand transfer inhibitors capable of circumventing the problem of resistance.

Keywords: HIV-1, AIDS, Integrase, viral DNA, Diketoacids, Strand transfer inhibitors, Dolutegravir, halogenated ring, quantum mechanics, polarizable molecular mechanics, molecular dynamics, intermolecular interaction energy, Gibbs free energy.

Chapitre I

Introduction générale au projet

Le Syndrome de l'Immunodéficience Acquisée (SIDA), causé par le Virus d'Immunodéficience Acquisée (VIH), fait partie des pandémies les plus graves et les plus dangereuses connues par notre monde, depuis les années 80 ^[1]. Afin de lutter contre le SIDA, diverses stratégies thérapeutiques ont été développées mettant en jeu plusieurs antirétroviraux (ARV). Le traitement actuel le plus courant est connu sous le nom de HAART (Highly Active Anti-Retroviral Therapy) ; il correspond à une thérapie antirétrovirale hautement active, se basant sur une combinaison d'ARV qui cible différentes étapes du cycle réplcatif du VIH. Cependant, cette thérapie reste coûteuse, exige de la discipline et engendre chez les patients traités, des mutations de résistance au sein des protéines virales ^[2]. D'où la nécessité de lutter contre le problème de résistance du VIH en développant de nouvelles drogues antirétrovirales plus potentielles.

Une des cibles thérapeutiques les plus prometteuses du VIH est l'intégrase virale (IN). En effet, cette enzyme clé joue un rôle important dans le cycle viral ; elle catalyse l'intégration du génome rétro-transcrit du virus dans les chromosomes des cellules infectées et ceci en deux étapes distinctes ; la maturation en 3' suivi de l'étape de transfert de brin (ST) ^{[3] [4]} (Figure 1), de plus, elle ne présente pas de contrepartie cellulaire. Par conséquent, le ciblage de l'intégrase et donc de l'étape d'intégration est une stratégie importante afin de bloquer le cycle réplcatif du VIH et éventuellement la production de nouvelles particules virales ^[5].

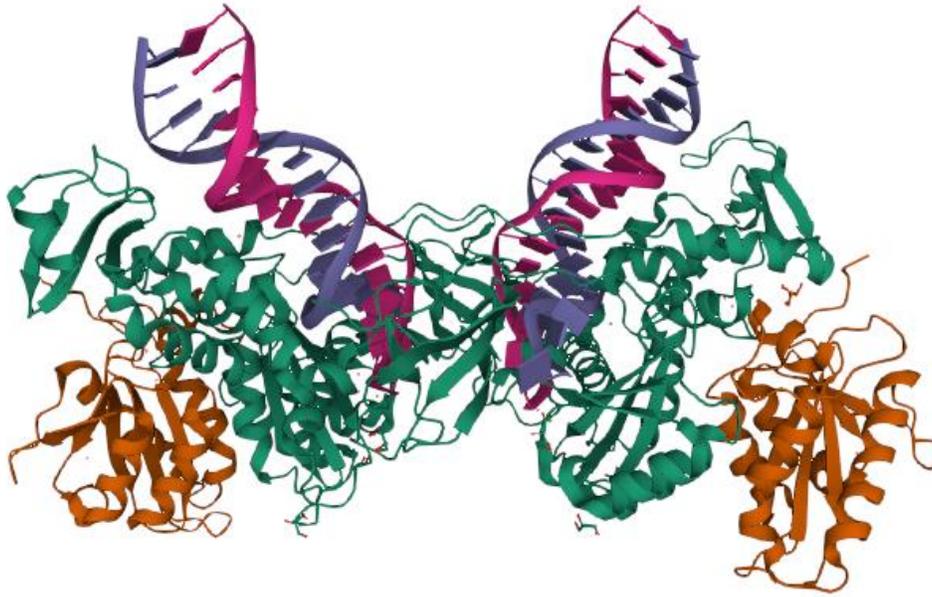


Figure 1 : Représentation tridimensionnelle de la structure cristalline du tétramère d'intégrase du Prototype Foamy Virus (PFV) en présence de l'ADN viral (code PDB : 3L2R) (*Hare et al., 2010. doi: 10.1038/nature08784*) Le double brin de chaque molécule d'ADN viral est coloré en magenta et violet, les sous-unités internes de l'intégrase sont colorées en vert, les sous-unités externes de l'intégrase sont colorées en orange.

En raison des analogies structurales et fonctionnelles entre les intégrases du VIH et du PFV, nous nous permettons d'utiliser ce dernier dans nos études dû à l'abondance des structures cristallines comprenant l'intégrase du PFV et l'insolubilité de l'intégrase du VIH-1.

Jusqu'à présent, trois inhibiteurs anti-IN ont été mis sur le marché, appartenant à la famille des inhibiteurs de transfert de brin (INSTIs) vu leur efficacité à inhiber spécifiquement l'étape de transfert de brins ^[6]. Il s'agit du Raltégravir (RAL), l'Elvitégravir (EVG) et le Dolutégravir (DTG) commercialisés en 2007 ^[7], 2012 ^[8] et 2013 ^[9], respectivement. Ils agissent à l'interface Intégrase – ADN viral (ADNv), ce qui leur permet d'effectuer des contacts directs avec les deux partenaires ^[10]. DTG présente la meilleure efficacité d'inhibition virale comparée à RAL et EVG, et une sensibilité moindre face aux mutations de résistance développées au sein de l'intégrase ^[11].

Des études effectuées précédemment au sein de notre équipe, combinant les calculs à l'expérience, ont mis en relief la spécificité de l'interaction respective

entre le RAL, l'EVG et le DTG avec l'ADN viral du VIH-1, impliquant les nucléotides terminaux de l'ADNv. De plus, nous avons démontré que le DTG affichait la meilleure interaction ce qui serait à l'origine de son efficacité accrue par rapport aux deux autres INSTIs ^[12]. Ajoutons qu'une meilleure affinité du DTG pour l'ADNv renforcerait son positionnement à l'interface ADNv-IN et compenserait son affinité envers l'enzyme virale. Il serait alors capable de contourner l'effet des mutations de résistance développées au sein de la protéine virale.

Concernant les sites d'interaction, nous avons réussi à démontrer antérieurement comment DTG et les INSTIs en général interagissaient avec les bases terminales de l'ADN par leur noyau halobenzénique qui opère un rôle d'ancrage essentiel à ce niveau. Par conséquent, notre stratégie d'optimisation des inhibiteurs de transfert de brin s'est basée sur la structure du DTG, au niveau de son noyau halobenzénique précisément. A ce stade, afin de renforcer les contacts INSTIs-ADNv et augmenter ainsi leur potentiel inhibiteur, il faudrait maintenant modifier la structure de l'halobenzène du DTG en effectuant des halogénations sélectives et des substitutions de groupements donneurs et attracteurs d'électrons : c'est l'un des buts principaux de cette thèse.

Dans un premier temps, nous avons travaillé sur la conception de nouvelles molécules inhibitrices, à partir du DTG. Les modifications sont basées sur une caractéristique des composés halogénés, qui est la présence d'un trou sigma « sigma-hole » au niveau de la liaison C-X (X = halogène) déterminant une déplétion de la densité électronique dans le prolongement de la liaison C-X et une accumulation de charges sur ses côtés ^[13]. Treize molécules ont été conçues et ont fait l'objet d'une étude quantitative de leur énergie d'interaction avec les bases terminales de l'ADNv, en ayant recours à des méthodes de calcul quantique et de

mécanique moléculaire polarisable utilisant un champ de force polarisable « SIBFA »^[14] développé au sein de notre laboratoire à Paris. Nous avons de même effectué une étude qualitative des nouveaux noyaux d'halobenzènes proposés par détermination de leur potentiel électrostatique moléculaire (MEP)^[15]. Ce dernier reflète la distribution des charges électroniques au sein des noyaux, nous permettant la compréhension des effets électroniques résultants des halogénations et substitutions ponctuelles et la validation des valeurs énergétiques obtenues.

La deuxième partie de notre travail a consisté en une étude plus approfondie du comportement du DTG ainsi que des nouveaux inhibiteurs au sein de l'intasome entier (la totalité du complexe IN-ADNv). Effectivement, nous sommes passés à une échelle d'étude plus élevée, en entamant une étude pionnière de dynamique moléculaire polarisable. Nous avons cherché à mieux comprendre le mécanisme de l'intégration, en suivant l'évolution de l'intasome du VIH-1 dans le temps. Ensuite, nous avons exploré les interactions du DTG au sein de l'intasome, avec l'ADNv entier ainsi qu'avec l'intégrase entière (code PDB : 3S3M) afin de mieux comprendre son positionnement vis-à-vis de ses deux cibles, éclaircissant ainsi son efficacité accrue, comparée à ses antécédents. De plus, nous avons choisi parmi les dérivés conçus à partir du DTG, ceux qui s'avéraient être les plus prometteurs et nous les avons intégrés dans la structure PDB 3S3M, remplaçant le DTG à l'interface ADNv-IN pour évaluer leur comportement au sein de l'intasome, et spécifiquement envers l'ADNv. Les résultats affichés ont alors confirmé ceux obtenus suite à nos calculs de mécanique quantique et de mécanique moléculaire.

Enfin, nous avons déterminé les énergies libres absolues des différentes interactions entre l'ensemble des partenaires au sein de l'intasome,

respectivement incluant nos nouveaux dérivés choisis. Cette méthode thermodynamique rigoureuse permet la prédiction de l'affinité d'une drogue à sa cible biologique et éclaire par la suite la stabilité des interactions des partenaires et leurs natures dominantes. Nous visons ainsi à valider nos résultats par une approche combinant la mécanique quantique, la mécanique moléculaire ainsi que la dynamique moléculaire.

A l'issue de ce travail, nous espérons proposer une nouvelle génération de molécules inhibitrices dotées d'une meilleure affinité pour l'ADN viral et donc d'un meilleur potentiel inhibiteur, dans le but de surmonter le problème de résistance que rencontrent les patients traités actuellement et de renforcer la lutte anti-VIH.

Références bibliographiques

- [1] F. Barre-Sinoussi *et al.*, "Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)," *Science*, vol. 220, no. 4599, pp. 868–871, May 1983, doi: 10.1126/science.6189183.
- [2] F. Clavel and A. J. Hance, "HIV Drug Resistance," *N. Engl. J. Med.*, vol. 350, no. 10, pp. 1023–1035, Mar. 2004, doi: 10.1056/NEJMra025195.
- [3] F. D. Bushman, T. Fujiwara, and R. Craigie, "Retroviral DNA integration directed by HIV integration protein in vitro," *Science*, vol. 249, no. 4976, pp. 1555–1558, Sep. 1990, doi: 10.1126/science.2171144.
- [4] E. Asante-Appiah and A. M. Skalka, "HIV-1 Integrase: Structural Organization, Conformational Changes, and Catalysis," in *Advances in Virus Research*, vol. 52, K. Rlaramorosch, F. A. Murphy, and A. J. Shawn, Eds. Academic Press, 1999, pp. 351–369.
- [5] J. Trivedi, D. Mahajan, R. J. Jaffe, A. Acharya, D. Mitra, and S. N. Byrareddy, "Recent advances in the development of integrase inhibitors for HIV treatment," *Curr. HIV/AIDS Rep.*, vol. 17, no. 1, pp. 63–75, Feb. 2020, doi: 10.1007/s11904-019-00480-3.
- [6] S. L. Karmon and M. Markowitz, "Next-generation integrase inhibitors : where to after raltegravir?," *Drugs*, vol. 73, no. 3, pp. 213–228, Mar. 2013, doi: 10.1007/s40265-013-0015-5.
- [7] B. Grinsztejn *et al.*, "Safety and efficacy of the HIV-1 integrase inhibitor raltegravir (MK-0518) in treatment-experienced patients with multidrug-resistant virus: a phase II randomised controlled trial," *Lancet Lond. Engl.*, vol. 369, no. 9569, pp. 1261–1269, Apr. 2007, doi: 10.1016/S0140-6736(07)60597-2.
- [8] M. Sato *et al.*, "Novel HIV-1 integrase inhibitors derived from quinolone antibiotics," *J. Med. Chem.*, vol. 49, no. 5, pp. 1506–1508, Mar. 2006, doi: 10.1021/jm0600139.
- [9] L. Vandekerckhove, "GSK-1349572, a novel integrase inhibitor for the treatment of HIV infection," *Curr. Opin. Investig. Drugs Lond. Engl.* 2000, vol. 11, no. 2, pp. 203–212, Feb. 2010.
- [10] S. Hare, S. S. Gupta, E. Valkov, A. Engelman, and P. Cherepanov, "Retroviral intasome assembly and inhibition of DNA strand transfer," *Nature*, vol. 464, no. 7286, pp. 232–236, Mar. 2010, doi: 10.1038/nature08784.
- [11] X. Z. Zhao *et al.*, "HIV-1 Integrase Strand Transfer Inhibitors with Reduced Susceptibility to Drug Resistant Mutant Integrases," *ACS Chem. Biol.*, vol. 11, no. 4, pp. 1074–1081, Apr. 2016, doi: 10.1021/acschembio.5b00948.
- [12] L. El Houry, J.-P. Piquemal, S. Fermandjian, R. G. Maroun, N. Gresh, and Z. Hobaiika, "The inhibition process of HIV-1 integrase by diketoacids molecules: Understanding the factors governing the better efficiency of dolutegravir," *Biochem. Biophys. Res. Commun.*, vol. 488, no. 3, pp. 433–438, Jul. 2017, doi: 10.1016/j.bbrc.2017.05.001.
- [13] T. Clark, M. Hennemann, J. S. Murray, and P. Politzer, "Halogen bonding: the σ -hole: Proceedings of 'Modeling interactions in biomolecules II', Prague, September 5th–9th, 2005," *J. Mol. Model.*, vol. 13, no. 2, pp. 291–296, Jan. 2007, doi: 10.1007/s00894-006-0130-2.
- [14] N. Gresh, P. Claverie, and A. Pullman, "Theoretical studies of molecular conformation. Derivation of an additive procedure for the computation of intramolecular interaction energies. Comparison with ab initio SCF computations," *Theor. Chim. Acta*, vol. 66, no. 1, pp. 1–20, Jan. 1984, doi: 10.1007/BF00577135.
- [15] P. Politzer, P. R. Laurence, and K. Jayasuriya, "Molecular electrostatic potentials: an effective tool for the elucidation of biochemical phenomena," *Environ. Health Perspect.*, vol. 61, pp. 191–202, Sep. 1985.

Chapitre II

Contexte Scientifique

Chapitre II
Contexte Scientifique

PARTIE A

Conception *in silico* des médicaments

A.I. Introduction générale

La conception de médicaments est une science axée essentiellement sur le patient, dans laquelle les laboratoires et les industries pharmaceutiques s'efforcent d'améliorer les médicaments déjà présents sur le marché ou d'inventer une entité chimique thérapeutique totalement nouvelle. Dans les deux cas, la vision finale est d'obtenir un nouveau médicament plus puissant, ayant un indice de chimiothérapie plus élevé que son antécédent dans la même catégorie. Cependant, le développement de nouvelles drogues à applications thérapeutiques potentielles reste un processus complexe et difficile ; des millions d'euros de capitaux et d'heures de travail y sont consacrés annuellement. De plus, divers aspects doivent être pris en considération au cours de ce développement étant donné qu'il s'effectue dans le cadre d'une approche pluridisciplinaire associant divers domaines scientifiques dont la biologie structurale, la chimie théorique, l'informatique, la pharmacologie et la synthèse organique ^[1] ^[2].

Les progrès technologiques récents, en particulier la fusion entre la biologie moléculaire et l'informatique, contribuent à accélérer, faciliter et améliorer le processus adopté par les sociétés pharmaceutiques pour découvrir de nouvelles cibles médicamenteuses et développer des inhibiteurs agissant à leur niveau ^[3].

A.II. Le processus de développement des médicaments

Le processus de développement de médicaments comprend deux étapes principales: l'étude préclinique suivie de l'étude clinique. L'étape préclinique se focalise globalement sur l'optimisation d'un candidat thérapeutique dans le but d'améliorer son potentiel et son efficacité envers sa cible alors que l'étape clinique consiste à évaluer le nouveau médicament chez l'homme avant sa mise sur le

marché. Nous allons détailler dans ce qui suit chaque étape du développement médicamenteux, en prenant comme référence la Figure 3 qui le schématise.

A.II.1. L'évaluation préclinique

L'étude préclinique est primordiale avant d'administrer la molécule thérapeutique chez les patients traités. Elle comprend plusieurs étapes successives ^[4], citées et détaillées ci-dessous :

- **Etape 1 : Identification de la cible :**

Le point de départ clé dans le pipeline de découverte de médicaments est l'identification du site de liaison à cibler correspondant à la pathologie en question. C'est une étape incontournable où s'effectue une analyse méticuleuse des caractéristiques structurales et chimiques de la cible, en se basant sur diverses références bibliographiques et sur des approches génomiques et protéomiques. La cible en question pourra alors être séquencée, clonée et purifiée en grandes quantités afin d'être testée.

- **Etape 2 : Construction de la cible :**

Après avoir identifié la cible thérapeutique, on passe à sa construction puis à son évaluation rigoureuse afin de démontrer que la modulation de la cible engendrera le résultat clinique souhaité.

- **Etape 3 : Identification de molécule « tête de série »**

L'étape la plus importante dans le processus de conception de médicaments est l'identification d'un inhibiteur « tête de série » (lead compound) à partir de petites molécules modulatrices de la fonction de la cible (Figure 2). Ces molécules peuvent être des produits naturels (sources animales, venins, toxines etc) ou des

médicaments préexistants. L'identification se fait par différentes approches telles que le criblage virtuel à haut débit (High Throughput Screening) ou la conception basée sur la structure (Structure-based Drug Design).

Les candidats « tête de série » sélectionnés à l'issue de ces approches possèdent une activité biologique optimale vis-à-vis de la cible, et d'une relation structure-activité liée à une efficacité thérapeutique. En d'autres termes, ce sont les molécules qui présentent un bon score de liaison évaluant le ratio affinité/spécificité et un bon équilibre entre hydrophilicité/lipophilicité, parmi les milliers de molécules testées.

Un autre critère de filtration des candidats médicamenteux se base sur la règle de 5 de Lipinski, qui comprend des critères physico-chimiques nécessaires à une bonne biodisponibilité et absorption des composés par voie orale ^[5] ^[6]. Les propriétés de Lipinski sont les suivantes :

1. Un poids moléculaire inférieur à 500 Da.
2. Log P (indice de lipophilie) inférieur à 5. C'est le coefficient de partage entre l'eau et le solvant organique, permettant d'évaluer la lipophilie d'un composé organique et donc d'estimer sa propriété à traverser ou non les membranes biologiques.
3. Le nombre de donneurs de liaisons hydrogènes est inférieur à 5.
4. Le nombre d'accepteurs de liaisons hydrogènes est inférieur à 10.

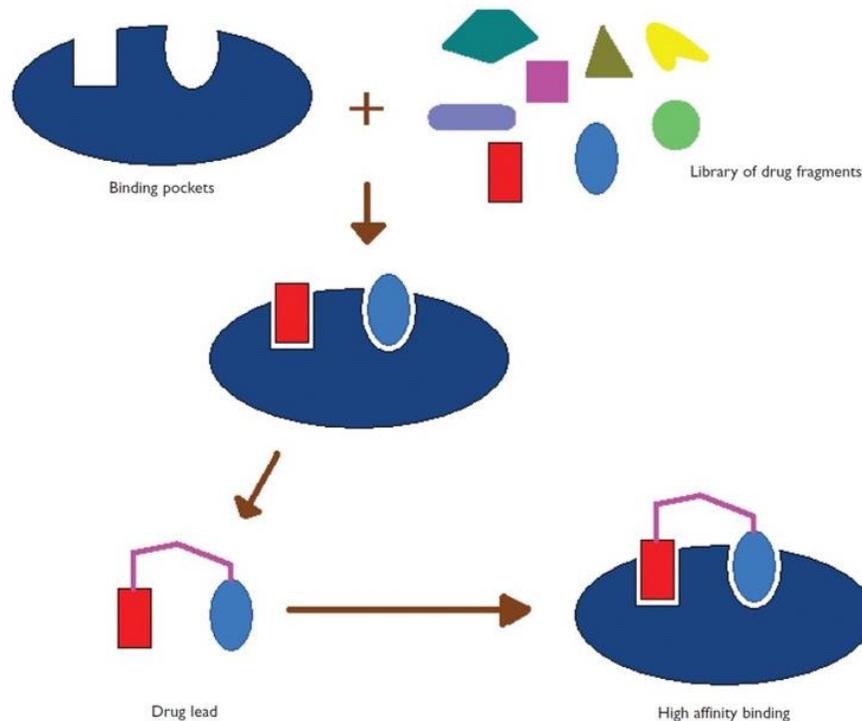


Figure 2 : Le processus d'identification des molécules « tête de série » par criblage

(*"Synthetic approaches to rational drug design – A changing landscape – Drug Discovery World (DDW)."* <https://www.ddw-online.com/synthetic-approaches-to-rational-drug-design-a-changing-landscape-925-200308/> accessed Mar. 25, 2021).

- Etape 4 : Optimisation de la molécule « tête de série »

Une fois choisies et identifiées, les molécules « leads » sont alors optimisées. Elles subissent un processus de raffinage de leur structure chimique, dans le but de construire le plus haut niveau possible de puissance, de sélectivité et d'affinité pour la cible d'intérêt. Le raffinage comprend précisément la modulation des paramètres pharmacocinétiques comme par exemple la relation dose-effet, la relation concentration-temps et la distribution du médicament-candidat.

- Etape 5 : Conception de formulaires et de schémas posologiques appropriés

Cet aspect de la conception diffère de la manipulation et de l'optimisation moléculaire par le fait qu'au cours de cette étape la composition du produit médicamenteux est modifiée et non la structure moléculaire du principe actif. Ainsi on cible obtenir un degré élevé de disponibilité biologique tout en conservant l'efficacité thérapeutique du médicament.

- Etape 6: Vérification de l'activité

Après avoir identifié une cible thérapeutique et filtré les molécules candidates (à peu près 250 composés arrivent à ce stade à partir des milliers testés), une série de tests est effectuée afin de valider l'activité de l'inhibiteur sur sa cible. La vérification se fait *in vitro* sur des modèles de laboratoire de complexité croissante, commençant par des modèles enzymatiques puis cellulaires puis tissulaires. Lors de cette phase, les propriétés ADME-Tox (Administration, Distribution, Métabolisme, Excrétion et Toxicité) seraient évalués ainsi que les propriétés mutagènes, cancérigènes et tératogènes les plus pertinents du médicament ^[7] ^[8]. Ces informations quantitatives sont nécessaires pour la distribution de la molécule du médicament en sa forme active dans les différents compartiments de l'organisme, en fonction de la dose et de l'intervalle entre les doses.

- Etape 7 : Tests *in vivo* chez les animaux

Après avoir testé plusieurs candidats thérapeutiques *in vitro*, on passe aux tests *in vivo* chez les animaux avant de débiter les essais cliniques chez l'homme. Cette dernière étape des études précliniques fournit des informations sur l'innocuité et l'efficacité du médicament sur un modèle animal présentant la maladie qu'on

désire traiter. Ainsi plusieurs données pourront être récupérées comprenant la toxicité aiguë, la cinétique et le métabolisme du médicament ainsi que la sensibilité des organes de l'animal testé. La dose de départ est évaluée de même afin qu'elle soit dans une marge de sécurité acceptable, sans danger aux humains.

A.II.2. L'évaluation clinique

Suite aux tests de la phase préclinique, moins que cinq composés à peu près arrivent à l'étape d'évaluation chez des êtres humains volontaires. C'est une étape cruciale avant la commercialisation du médicament, basée sur des données recueillies lors des essais cliniques, comme les effets toxicologiques de la drogue sur le corps humain, son niveau plasmatique, son absorption, son excrétion et la formation de ses métabolites^[9]. Notons qu'une demande d'approbation par la FDA doit être déposée pour pouvoir procéder à l'étude sur des patients humains.

L'étude clinique se fait en 3 étapes^{[10] [11]} :

- *Etape 1 : Etudes ADME-Tox chez les humains*

Une étude des propriétés ADME-Tox est menée à ce stade sur de petits groupes de volontaires sains (moins de 100 personnes) dans le but d'évaluer la sécurité du composé, notamment par la recherche de ses effets secondaires afin de corroborer la toxicité pré-évaluée chez l'animal.

- *Etape 2 : Evaluation pharmacologique*

Les essais de la phase 2 ont pour objectif de confirmer l'action thérapeutique du médicament-candidat et de vérifier sa dose optimale. Les tests sont alors menés sur des personnes malades, en commençant par un petit groupe de 100 personnes

avant de passer à plusieurs centaines de patients. On peut alors déterminer la dose à la fois efficace et bien tolérée de substance à administrer.

- Etape 3 : Etude en « double-aveugle »

Une étude comparative est exécutée, vis-à-vis du composé de référence et d'un placebo. Son but est la comparaison de l'efficacité thérapeutique du médicament par rapport au placebo. Elle contribue de même à évaluer le rapport bénéfice-risque. Ces essais sont menés sur de grandes populations de patients.

Un seul candidat thérapeutique sera choisi après son succès dans la phase clinique, il est alors prêt à être soumis aux autorités de santé pour sa commercialisation. C'est ce qu'on appelle l'Autorisation de Mise sur le Marché (AMM).

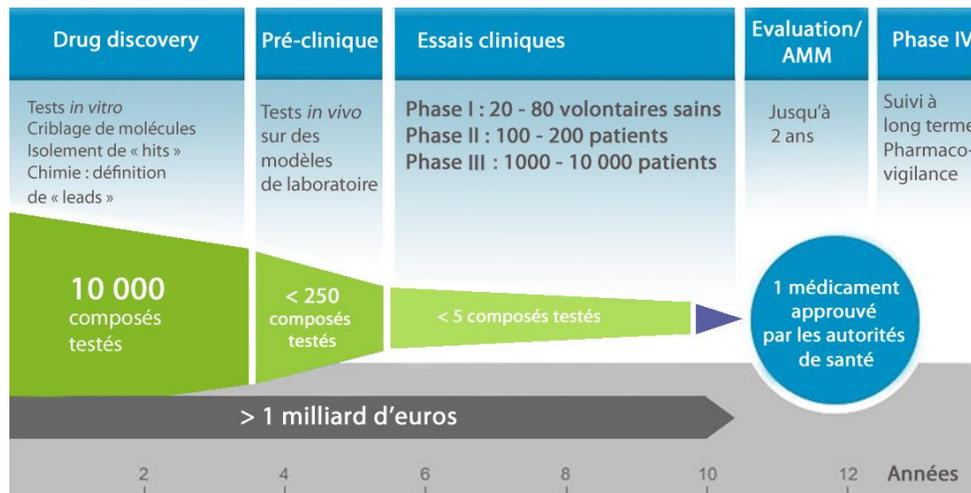


Figure 3: Schéma représentant les différentes étapes du processus de développement des médicaments (Le parcours du médicament, Fondation Synergie Lyon Cancer, <https://www.synergilyoncancer.fr/les-enjeux/le-parcours-du-medicament>)

A.III. La conception rationnelle de médicaments

En suivant les différentes étapes, nous constatons que le processus de développement des médicaments, présente différents défis. En fait, il est long (une

dizaine d'années) et coûteux (plusieurs essais sont entamés sur des milliers de molécules). Afin de relever ces défis, plusieurs approches multidisciplinaires sont nécessaires ; elles forment la base d'une conception rationnelle de médicaments.

La conception rationnelle de médicaments consiste à développer des composés thérapeutiques sur une base la plus rationnelle possible ; en les apparentant à des composés déjà connus, tout en les dotant d'une affinité augmentée. Par conséquent, une réduction au strict minimum du facteur d'essai et d'erreur dans la procédure sera établie, puisqu'en choisissant comme point de départ des composés préexistants, on diminuera le risque de faire des essais avec des produits chimiques inadaptés en tant qu'agents thérapeutiques, gagnant ainsi au niveau du coût et du temps ^[12]. Ajoutons que les entreprises pharmaceutiques sont en possession de « bibliothèques » de plusieurs centaines de millions de molécules chimiques disponibles pour effectuer des essais. Grâce à l'approche rationnelle, les chances d'identifier des composés « tête de série » parmi ces molécules sont augmentées ^[13]. C'est une approche pluridisciplinaire dont les progrès dépendent des contributions de divers domaines notamment la biologie structurale et moléculaire, la chimie organique et médicinale, la chimie théorique, la biophysique et la bio-informatique. Elle bénéficie particulièrement des progrès considérables de l'informatique et des avancées de la chimie théorique.

D'où l'apparition depuis quelques années des méthodes de calcul modernes se basant sur une conception médicamenteuse assistée par ordinateur (conception *in silico*) pour la sélection et l'optimisation des leads. Son importance ira en croissance étant donné qu'elle aide à guider la planification expérimentale et rend

potentiellement le processus de conception médicamenteuse plus efficace. De plus, elle bénéficie de l'accroissement des performances des ordinateurs.

Plusieurs méthodes *in silico* peuvent être utilisées en combinaison avec des preuves expérimentales afin d'aider à la compréhension de l'activité biologique et de l'affinité ligand-cible. Grâce à l'utilisation de l'approche *in silico*, il ne suffirait de synthétiser qu'un petit nombre de molécules candidates suite à la prédiction de leur activité.

Lorsque les structures tridimensionnelles des protéines cibles sont connues, on parle alors de « Conception de médicaments basée sur la structure » ou « Structure-based Drug Design). C'est une approche bien établie et largement appliquée par les industries pharmaceutiques, sur laquelle nous nous sommes basés dans nos travaux d'optimisation des inhibiteurs anti-VIH-1.

A.III.1. La conception de médicaments basée sur la structure

L'utilisation de la structure protéique tridimensionnelle (3D) dans le développement de nouveaux produits biologiquement actifs est appelée « Conception de médicaments basée sur la structure » (SBDD). C'est une stratégie réussie, attirante et très utilisée par la recherche académique et pharmaceutique. Elle permet l'étude de la structure cible-médicament et fournit par la suite des informations moléculaires détaillées sur le mécanisme d'action et d'inhibition de la cible ^[14] ^[15].

En effet, afin de réaliser de manière plus fructueuse des modifications de la structure principale de la molécule « lead », il est utile d'avoir une structure tridimensionnelle de la conformation bioactive du ligand thérapeutique lié à sa cible. De cette façon, différentes informations structurales du site actif (régions

hydrophobes, régions polaires, régions capables de faire des liaisons hydrogènes) peuvent être directement employées pour l'identification de nouveaux ligands qui pourraient avoir la taille, la forme et les propriétés électrostatiques adéquates afin de se lier plus spécifiquement à ce site. Ces connaissances structurales ont boosté la recherche de molécules tête de série, limitant le nombre de composés à évaluer expérimentalement ^[16].

La Figure 4 résume les étapes de la SBDD. Dans un premier temps, les structures des cibles (protéiques en général) apo (sans ligand) ou holo (avec ligand) sont déterminées expérimentalement par cristallographie aux rayons X ou par RMN. Ensuite, un criblage informatique des molécules « tête de série » a lieu. Les meilleures molécules seraient alors modifiées puis amarrées à nouveau au niveau du site ciblé, avant d'être étudiées rigoureusement par dynamique moléculaire, afin d'évaluer leur interaction au sein de ce dernier. Après l'identification théorique des meilleurs nouveaux composés, le passage aux tests in vitro et in vivo peut alors commencer. Finalement, le médicament choisi passe aux essais cliniques avant sa mise sur le marché.

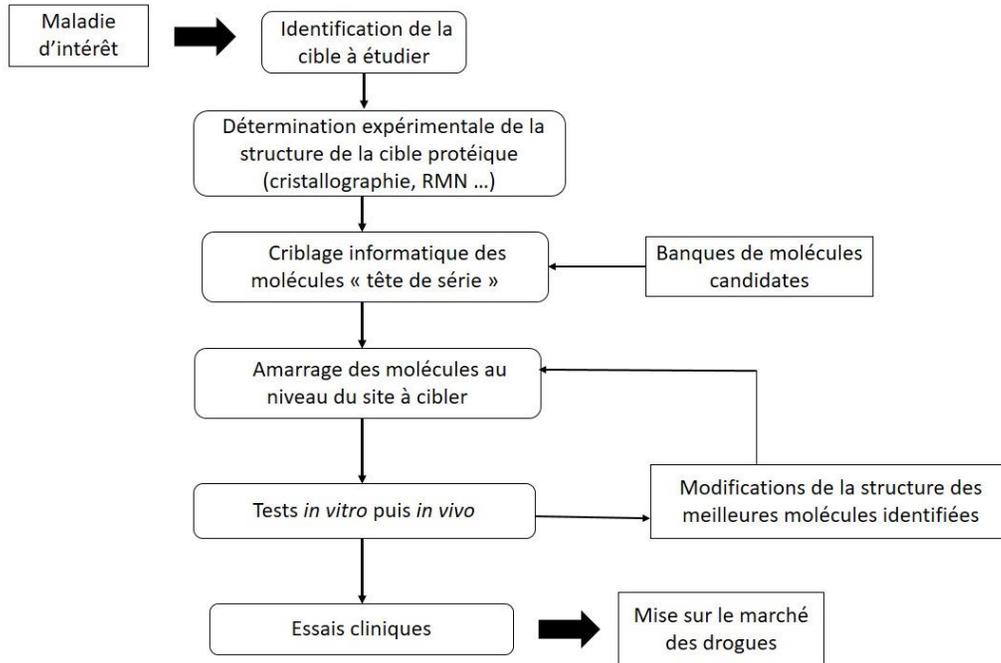


Figure 4 : Les étapes de la conception de médicaments basée sur la structure

Alternativement, lorsque la structure de la cible protéique n'est pas disponible, sa structure peut être dérivée à partir de modèles d'homologie. Notons que de nombreux médicaments approuvés par la FDA ont évolué grâce aux stratégies de conception basée sur la structure. Par exemple, en 2012, pas moins de 35 médicaments nouvellement approuvés ont émané de la conception *in silico* basée sur les structures cristallographiques ^[17].

Références bibliographiques

- [1] S.-F. Zhou and W.-Z. Zhong, "Drug Design and Discovery: Principles and Applications," *Mol. J. Synth. Chem. Nat. Prod. Chem.*, vol. 22, no. 2, Feb. 2017, doi: 10.3390/molecules22020279.
- [2] J. Kuhlmann, "Drug research: from the idea to the product," *Int. J. Clin. Pharmacol. Ther.*, vol. 35, no. 12, pp. 541–552, Dec. 1997
- [3] C. Stratowa and K. K. Wilgenbus, "Gene expression profiling in drug discovery and development," *Curr. Opin. Mol. Ther.*, vol. 1, no. 6, pp. 671–679, Dec. 1999.
- [4] C. Kumar, "An Insight to Drug Designing by In Silico approach in Biomedical Research," *J. Public Health Med. Res.*, vol. 1, pp. 63–65, Jan. 2013.
- [5] S. Mandal, M. Moudgil, and S. K. Mandal, "Rational drug design," *Eur. J. Pharmacol.*, vol. 625, no. 1–3, pp. 90–100, Dec. 2009, doi: 10.1016/j.ejphar.2009.06.065.
- [6] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, "Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings," *Adv. Drug Deliv. Rev.*, vol. 23, no. 1, pp. 3–25, Jan. 1997, doi: 10.1016/S0169-409X(96)00423-1.
- [7] R. Frank and R. Hargreaves, "Clinical biomarkers in drug discovery and development," *Nat. Rev. Drug Discov.*, vol. 2, no. 7, Art. no. 7, Jul. 2003, doi: 10.1038/nrd1130.
- [8] A. P. Li, "Screening for human ADME/Tox drug properties in drug discovery," *Drug Discov. Today*, vol. 6, no. 7, pp. 357–366, Apr. 2001, doi: 10.1016/s1359-6446(01)01712-3.
- [9] M. D. Rawlins, "Cutting the cost of drug development?," *Nat. Rev. Drug Discov.*, vol. 3, no. 4, pp. 360–364, Apr. 2004, doi: 10.1038/nrd1347.
- [10] K. B. Suresh, T. M. Gerald, G. Liang-Shang, W. Jing-Tao, and W. L. Frank, "Strategy of Utilizing In Vitro and In Vivo ADME Tools for Lead Optimization and Drug Candidate Selection," *Curr. Top. Med. Chem.*, vol. 5, no. 11, pp. 1033–1038, Aug. 2005.
- [11] W. C. Gary, Y. Zhengyin, T. Weimin, D. Malini, and H. Becki, "ADME Optimization and Toxicity Assessment in Early- and Late-Phase Drug Discovery," *Curr. Top. Med. Chem.*, vol. 9, no. 11, pp. 965–980, Jul. 2009.
- [12] W. N. Hunter, "Rational drug design: a multidisciplinary approach," *Mol. Med. Today*, vol. 1, no. 1, pp. 31–34, Apr. 1995, doi: 10.1016/1357-4310(95)80017-4.
- [13] N. K. Terrett, M. Gardner, D. W. Gordon, R. J. Kobylecki, and J. Steele, "Combinatorial synthesis — the design of compound libraries and their application to drug discovery," *Tetrahedron*, vol. 51, no. 30, pp. 8135–8173, Jul. 1995, doi: 10.1016/0040-4020(95)00467-M.
- [14] A. D. Andricopulo, L. B. Salum, and D. J. Abraham, "Structure-based drug design strategies in medicinal chemistry," *Curr. Top. Med. Chem.*, vol. 9, no. 9, pp. 771–790, 2009, doi: 10.2174/156802609789207127.
- [15] J. Greer, J. W. Erickson, J. J. Baldwin, and M. D. Varney, "Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design," *J. Med. Chem.*, vol. 37, no. 8, pp. 1035–1054, Apr. 1994, doi: 10.1021/jm00034a001.
- [16] S. Hirono, "[An introduction to the computer-aided structure-based drug design--applications of bioinformatics to drug discovery]," *Rinsho Byori*, vol. 50, no. 1, pp. 45–51, Jan. 2002.
- [17] A. K. Ghosh and S. Gemma, "Protein X-Ray Crystallography in Structure-Based Drug Design," in *Structure-Based Design of Drugs and Other Bioactive Molecules*, John Wiley & Sons, Ltd, 2014, pp. 183–198.

Chapitre II
Contexte Scientifique

PARTIE B
SIDA et VIH

B.I. La pandémie du SIDA

Au début des années 80, aux Etats-Unis, des cas de personnes malades présentant une baisse des défenses immunitaires ont été reportés ^[1]. Le SIDA (Syndrome d'Immunodéficience Acquis) a été alors abordé pour la première fois ^[2]. Cette maladie infectieuse attaque le système immunitaire humain et entraîne par conséquent une chute des défenses naturelles. Elle est considérée comme la maladie du vingtième siècle ayant connu l'expansion la plus rapide. Avec 1,7 millions de personnes nouvellement atteintes du SIDA en 2019 et plus de 38 millions de personnes vivant avec la maladie en 2019 ^[3] (Figure 5), cette épidémie est toujours d'actualité (Figure 6) et reste, de par sa gravité, parmi les principales causes mondiales de mortalité.

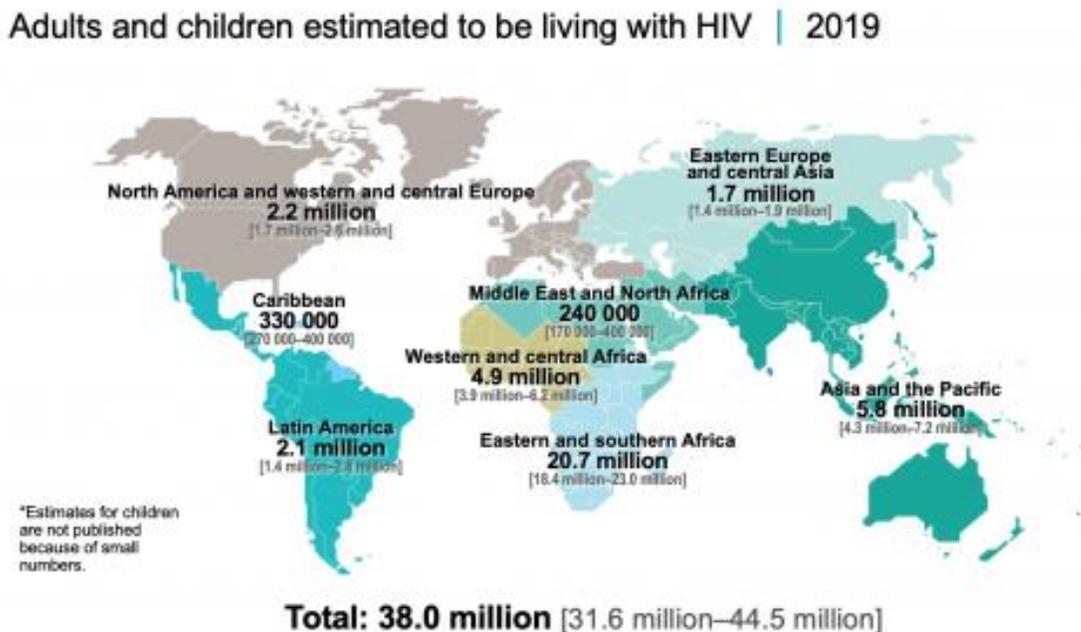


Figure 5 : Répartition des personnes (adultes et enfants) atteintes de SIDA dans le monde en 2019. La totalité des personnes infectées monte à 38 millions.

(B. Ryan, "Global HIV Treatment Rates Rise, but Progress Is Uneven," POZ, Jul. 06, 2020.)

L'agent causal du SIDA a été isolé et identifié comme le Virus d'Immunodéficience Humaine (VIH) en 1986 [4]. Il est présent sous deux types majeures ; le VIH-1 et le VIH-2 qui sont antigéniquement distincts mais partagent cependant des propriétés génétiques et biologiques (42% d'homologie au niveau de leur génome). Le VIH-1 a été isolé en premier et se trouve répandu dans le monde entier alors que le VIH-2 est localisé principalement en Afrique de l'Ouest et reste moins pathogène que le VIH-1 [5]. Depuis sa découverte, l'infection par le VIH se trouve encore répandue au sein des hommes, des femmes et des enfants de toutes les populations. Sa transmission se fait majoritairement par contact avec les liquides corporels de la personne infectée ; lors des rapports sexuels non protégés, lors des transfusions sanguines utilisant des seringues contaminées, par la transmission materno-fœtale ou via le lait maternel [6].

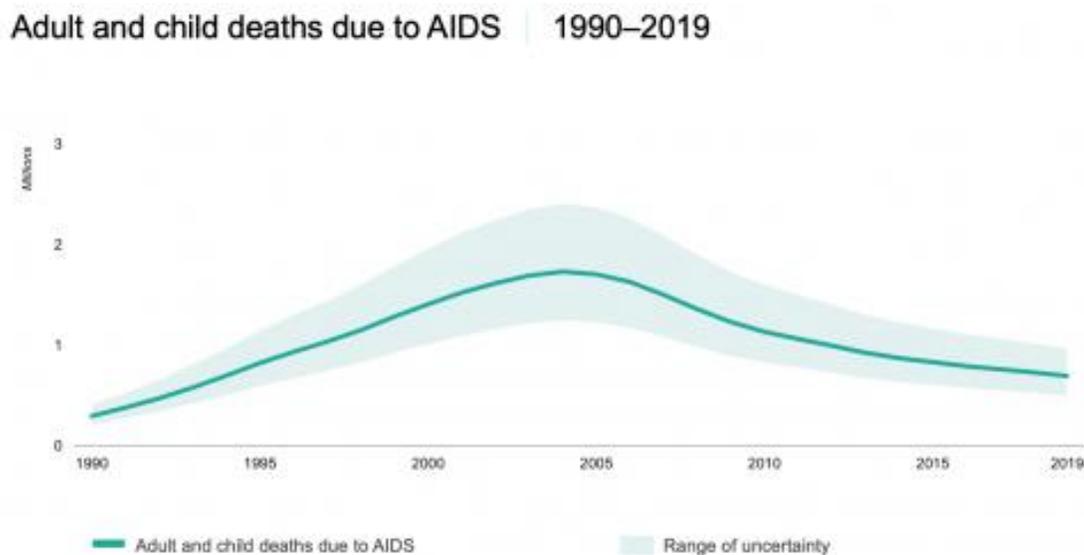


Figure 6 : Courbe croissante représentant l'évolution du nombre de personnes infectées (en millions) par le SIDA depuis 1990 à 2019.

(B. Ryan, "Global HIV Treatment Rates Rise, but Progress Is Uneven," POZ, Jul. 06, 2020.)

Aujourd'hui, et trente ans après l'identification du VIH, des progrès scientifiques et thérapeutiques ont eu lieu au niveau de la compréhension de la pathogénicité du SIDA, donnant lieu à diverses multithérapies. Ceci a considérablement augmenté le taux de survie des patients infectés. Cependant, le VIH est doté d'un taux élevé de variabilité génétique, attribué notamment à son mode de réplication [7]. De ce fait, des mutations de résistance sont apparues au sein des structures du VIH chez les personnes traitées, diminuant alors l'efficacité du traitement. Outre que l'émergence de mutations, le succès des thérapies anti-VIH reste aussi limité par l'accès difficile des pays en voie de développement à ces traitements coûteux. C'est pourquoi la recherche de nouveaux médicaments antirétroviraux continue et reste d'une importance majeure dans la lutte contre le SIDA.

B.II. Le virus d'Immunodéficience Humaine de type 1 (VIH-1)

Dans la suite de notre étude, nous allons nous focaliser sur l'étude du VIH-1 uniquement comme type du VIH, étant donné son abondance globale et sa fréquence d'infection par rapport au VIH-2.

B.II.1. Structure virale

Le Virus d'Immunodéficience Humaine de type 1 (VIH-1) appartient à la famille des Retroviridae ; c'est donc un rétrovirus. Cette famille est caractérisée par un génome viral constitué de deux fragments d'ARN homologues simple brin. Une transcription inverse de cet ARN en ADN double brin est nécessaire pour intégrer le matériel génomique viral dans les chromosomes des cellules cibles, permettant alors la réplication et l'expression du virus.

En fonction de leurs morphologies et de leurs caractéristiques biologiques et biochimiques, les rétrovirus sont classés en différents groupes [8]. Le VIH-1

appartient au groupe des lentivirus, il est donc responsable de l'apparition d'infections virales à évolution lente. Sa structure a été établie par microscopie électronique ^[9] où il apparaît sous forme globalement sphérique avec un diamètre d'environ 110 nm (Figure 7).

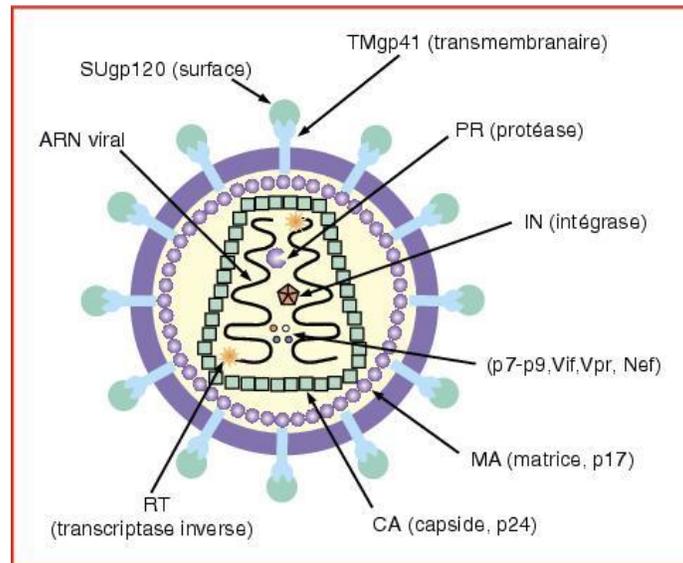


Figure 7 : Structure du VIH-1
(M. Pancera, *Virologie*, vol. 9, no. 6, pp. 457–472, Nov. 2005.)

Ses constituants structuraux majeurs sont les suivants : une enveloppe externe, une matrice et une nucléocapside. L'enveloppe externe, connue aussi sous le nom de coque, est formée d'une bicouche lipidique issue de la cellule hôte, dans laquelle sont ancrées deux types de glycoprotéines formant les spicules ; la SUgp120 (glycoprotéine de surface) et la TMgp41 (glycoprotéine transmembranaire). Cette enveloppe est reliée à la membrane interne ou matrice qui comprend la protéine virale p17. À l'intérieur de la matrice se trouve la capsid virale formée par une auto association des protéines de capsid p24. Cette capsid est associée aux protéines de la nucléocapsid p7. La nucléocapsid contient le matériel génétique viral (les deux brins d'ARN) ainsi que les trois enzymes virales indispensables au

cycle répliatif du VIH ; la transcriptase inverse (RT), la protéase (PR) et l'intégrase (IN). On y trouve de même d'autres protéines virales accessoires (Nef « Negative Expression Factor »^[10], Vpr « Viral protein R »^[11], Vif « Viral infectivity factor »^[12]) ainsi que des protéines cellulaires en provenance de la cellule hôte^[13].

B.II.2 Génome du VIH-1

Les deux brins d'ARN monocaténares homologues contiennent l'information génétique nécessaire à la réplication et la survie du VIH-1. Chaque brin d'ARN est encadré au niveau des extrémités 5' et 3' de séquences répétées (LTR) constituées de trois régions ; U5, R et U3 (Figure 8). Les extrémités LTR jouent un rôle dans l'intégration et la transcription du génome viral au sein de la cellule hôte^{[14] [15] [16]}.

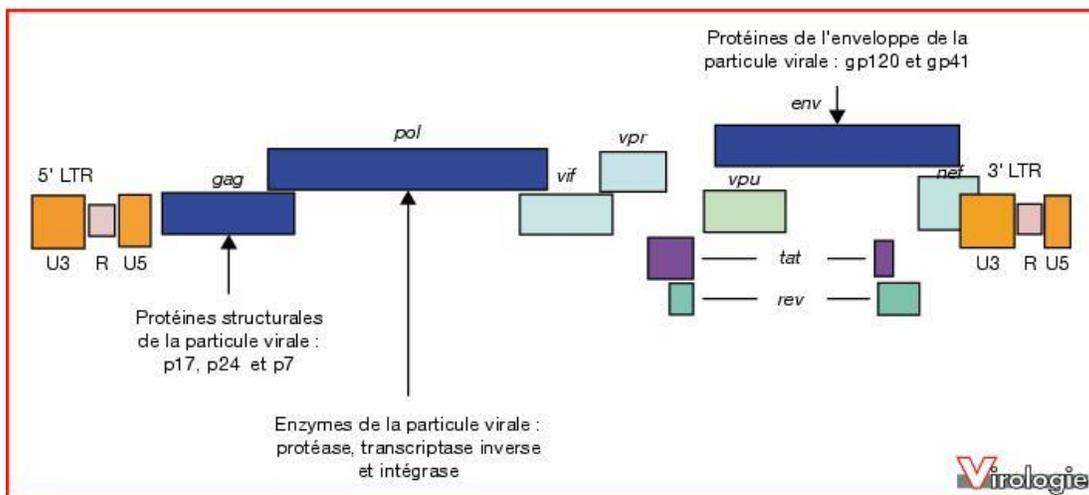


Figure 8: Schéma représentant l'organisation génomique du VIH-1

(L. Larrouy, F. Brun-Vézinet, and D. Descamps, *Virologie*, vol. 14, no. 2, pp. 119–128, Mar. 2010, doi: 10.1684/vir.2010.0289.)

En plus des extrémités LTR, l'ADN viral comprend trois gènes principaux : gag, pol et env, communs à tous les lentivirus, qui codent pour des précurseurs polyprotéiques et des gènes de régulation. Gag (group antigen) et pol (polmyerase)

codent pour un précurseur polypeptidique gag-pol dont le clivage auto-catalytique libère la protéase virale. PR à son tour sera responsable de la maturation d'autres précurseurs afin de libérer les protéines structurales de la matrice (p17), de la capsid (p27) et de la nucléocapsid (p7) ^[17]. La protéase permet également la libération des protéines enzymatiques ; l'intégrase et la transcriptase inverse. De son côté, le gène env génère les glycoprotéines de la coque, soient gp120 et gp 41 ^{[18] [19]}. De même, ce génome renferme les gènes vif, vpr et nef codant pour les protéines accessoires Vif, Vpr et Nef respectivement (Figure 8).

B.II.3. Cycle répliatif

La cible principale du VIH-1 est toute cellule humaine porteuse du récepteur CD4⁺, notamment les lymphocytes T Helper (T_H) et les macrophages. Ce virus cible donc les principaux acteurs de l'immunité innée et spécifique chez l'homme ^[2]. Le cycle répliatif, comme représenté dans la Figure 9, débute tout d'abord par une étape de reconnaissance qui met en jeu l'interaction des protéines de l'enveloppe externe (gp120 et gp41) avec les récepteurs CD4⁺. Cette reconnaissance est renforcée par la fixation du virus sur les corécepteurs CCR5 et CXCR4. Ceci aboutit à un changement de conformation donnant lieu à une fusion entre les membranes cellulaire et virale, par le biais de gp41. Le virus pénètre du coup à l'intérieur de la cellule hôte et décharge sa nucléocapsid virale dans le cytoplasme cellulaire ^{[21] [22]} ^[23]. La nucléocapsid libère par la suite les deux brins d'ARN viraux et les trois enzymes essentielles à la multiplication virale. La transcription inverse de l'ARNv en ADN double brin a alors lieu, orchestrée par la transcriptase inverse. Le produit de cette réaction (l'ADNv) sera par la suite traité par l'intégrase virale. En fait, les deux partenaires ainsi que des protéines cellulaires et virales s'organisent en un

complexe nucléoprotéique de pré-intégration (PIC) qui assurera un import nucléaire de l'ADNv et de l'intégrase. Une fois dans le génome de la cellule cible, l'intégrase catalyse l'intégration de l'ADN du VIH-1 dans le génome de la cellule cible, permettant sa transcription en ARN messager puis sa réplication en ARN génomique ainsi que sa traduction en précurseurs polyprotéiques. La protéase virale assure par la suite leur maturation en protéines virales (essentielle et accessoires). Le cycle viral se boucle par l'assemblage des protéines virales matures suivi du relargage des nouvelles particules virales infectieuses par bourgeonnement [24] [25] [26].

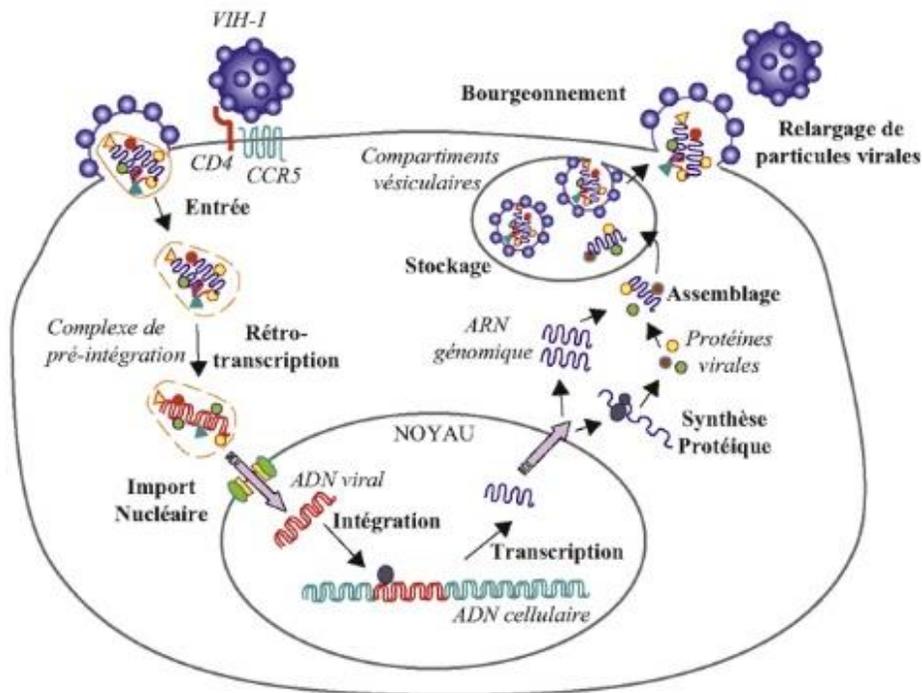


Figure 9 : Les différentes étapes du cycle réplcatif du VIH-1

(A. Bergamaschi, A. David, and G. Pancino, *Virologie*, vol. 15, no. 2, pp. 87–99, May 2011, doi: 10.1684/vir.2011.0400.)

Afin de contrôler la propagation du VIH-1, cibler les étapes du cycle répliatif est une stratégie primordiale à adopter. En fait, de nombreuses thérapies anti-VIH-1 existent déjà, visant les étapes répliatives suivantes ; l'entrée du virus/fusion membranaire, la transcription inverse et la maturation. Parmi elles, une trithérapie dénommée traitement antirétroviral hautement actif (HAART), représente jusqu'à ce jour la norme pour le traitement. Elle consiste en une combinaison d'inhibiteurs de la réverse transcriptase (NRTIs et NNRTIs) avec un inhibiteur de la protéase (PI) et un inhibiteur de la fusion et de l'entrée du virus. Suite à son adoption dans les années 90, cette multithérapie a pu réduire significativement le taux de mortalité des patients traités [27]. Notons toutefois qu'il existe des limitations à son utilisation, notamment son prix coûteux, l'exigence et la discipline du traitement et l'émergence de mutations de résistance au sein des souches virales traitées. En fait, le VIH-1 présente une fréquence de mutation élevée (10^{-4} à 10^{-5} mutations/nuéotide) ainsi qu'une fréquence de recombinaison élevée, ce qui contrebalance l'efficacité et l'intégrité de HAART [28][29]. Notons que c'est notamment la transcriptase inverse qui est à l'origine des variantes génétiques au niveau des glycoprotéines de la membrane du virus, puisqu'elle effectue des erreurs lors de la transcription inverse des deux brins d'ARN viral (une erreur sur toutes les 10000 bases) [30]. Par conséquent, l'adoption de nouvelles approches thérapeutiques s'avère être nécessaire.

B.III. Intégrase et mécanisme d'intégration

L'intégrase a été identifiée récemment comme une cible thérapeutique privilégiée pour le traitement de la maladie du SIDA. Tout d'abord, elle occupe un rôle primordial au niveau du cycle répliatif du VIH-1 étant donné qu'elle orchestre

l'insertion de l'ADN viral dans le génome des cellules à récepteurs CD4⁺. Elle permet donc l'expression du matériel génétique du VIH-1 et la réplication de ce dernier. De plus, en absence d'homologue humain pour l'intégrase, son inhibition chez les patients sera par la suite spécifique.

B.III.1. Structure de l'intégrase

L'intégrase est une protéine virale constituée de 288 acides aminés, avec une masse totale de 32kDa. Elle est caractérisée par une activité endonucléasique puisqu'elle fixe l'ADN et y effectue des clivages. Elle est codée par l'extrémité du gène viral pol, et fait donc partie du précurseur Gag-pol dont le clivage par la protéase virale lors de la maturation du virus libère alors l'intégrase ^{[31] [32]}.

Des études de protéolyse partielle ont révélé que l'intégrase du VIH-1 comprend trois domaines fonctionnels structuralement distincts (Figure 10) :

- Le domaine N-terminal (NTD) constitué des résidus de 1 à 50.
- Le domaine central (CCD) constitué des résidus de 51 à 212.
- Le domaine C-terminal (CTD) constitué des résidus de 213 à 288.

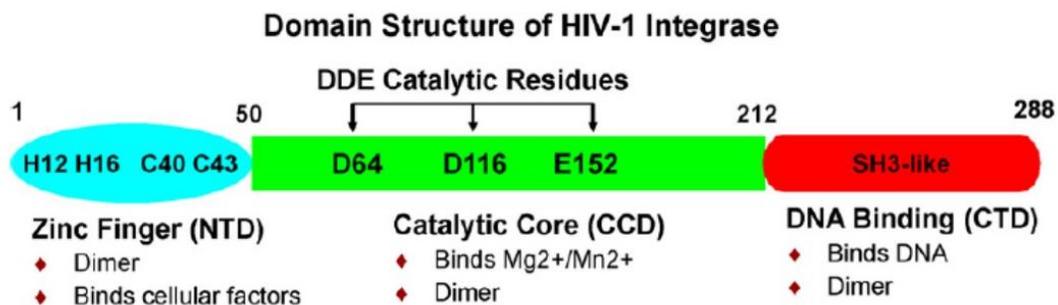


Figure 10 : Schéma représentant les domaines structuraux de l'intégrase du VIH-1

(D.J. McColl and X. Chen, Elsevier, vol. 85, no. 1, pp. 101–118, Jan 2010, doi: 10.1016/j.antiviral.2009.11.004)

Chaque domaine a pu être cristallisé ou étudié par RMN séparément ^[33] ^[34] ^[35]. Les bidomaines NTD-CCD et CCD-CTD ont eux aussi été cristallisés ^[36] ^[37]. Ceci a permis la modélisation de la forme globale de l'intégrase. Cependant, malgré son importance et son rôle dans les thérapies antirétrovirales, la structure complète de l'intégrase n'a pu être obtenue et ce malgré des efforts considérables, vu qu'elle possède une faible solubilité. Toutefois, la structure du tétramère d'intégrase du VIH-1 en complexe avec l'ADN viral et l'ADN cellulaire a été déterminée en 2017 par cryomicroscopie électronique grâce aux travaux de Passos *et al* ^[38].

Dans nos études, nous nous sommes basés sur un rétrovirus voisin du VIH-1 ; le Prototype Foamy Virus (PFV) qui présente une homologie élevée structurale et fonctionnelle avec le VIH-1 (Figure 11). En effet, plusieurs structures cristallisées de l'intégrase du PFV avec l'ADN viral ou avec divers inhibiteurs anti-IN existent ^[39] ^[40] ce qui nous a permis de prendre ces modèles comme références pour étudier scrupuleusement la structure et le mode d'action de l'intégrase du VIH-1 ainsi que son mode d'inhibition par les médicaments antirétroviraux.

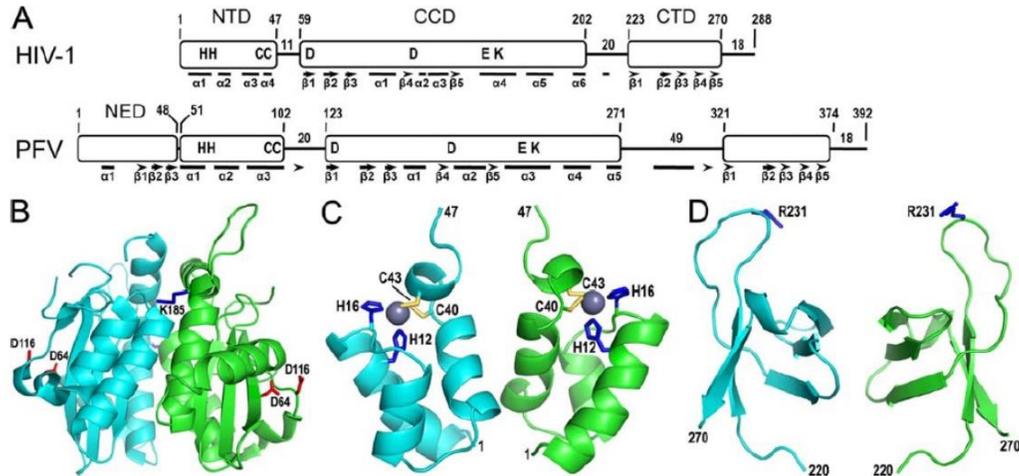


Figure 11 : Organisation des domaines des intégrases rétrovirales du Prototypé Foamy Virus (PFV) et du Virus de l'Immuno-déficience Humaine type 1 (VIH-1). **A.** Alignement des domaines structuraux (NTD, CCD et CTD) des intégrases du VIH-1 et du PFV, avec les résidus limites de chaque domaine ainsi que la longueur des régions de liaison inter domaine (NTD-CCD, CCD-CTD et la queue C-terminale). Les résidus conservés au niveau de toutes les intégrases rétrovirales sont indiqués en code à une seule lettre. Les barres et les flèches indiquent les structures secondaires en hélice alpha et feuillet bêta respectivement. **B.** Structure en cristal du domaine CCD de l'intégrase du VIH-1 (code PDB : 1ITG). Les résidus actifs sont indiqués et représentés en bâtons. **C.** Structure RMN du domaine NTD de l'intégrase du VIH-1 (code PDB : 1WJC) mettant en évidence les résidus H12, H16, C40 et C43 coordonnant un atome de zinc. **D.** Structure du domaine CTD de l'intégrase du VIH-1 déterminée par RMN (code PDB : 1IHV) où est indiqué le résidu R231 impliqué dans la liaison à l'ADN cellulaire

(A.Engelman, P. Cherepanov, pp. 1011–1033, November 2014, doi: 10.1128/microbiolspec.MDNA3-0024-2014)

B.III.1.1. Le domaine N-terminal

Le domaine N-terminal de l'intégrase du VIH-1, comprenant les 50 premiers résidus, est structuré en quatre hélices alpha, avec un motif Hélice-Tour-Hélice (HTH) au niveau des hélices $\alpha 2$ et $\alpha 3$. Ce motif possède un doigt de zinc HHCC, formé par quatre acides aminés conservés et essentiels (His12, His16, Cys40, Cys43) qui fixent un cation Zn^{2+} [41] (Figure 11). La fixation du zinc assure un repliement correct du NTD, et favorise la multimérisation de la protéine entière nécessaire à son activité catalytique [42] [43] [44]. Ce qui explique la conservation de la région HHCC au sein des intégrase rétrovirales, alors que le reste du domaine N-terminal du VIH-1 présente des homologies limitées par rapport à ces dernières[45]. Sa structure résolue par

RMN hétéronucléaire multidimensionnelle ^[33], est sous forme dimérique en solution (Figure 12) avec une interface composée majoritairement de résidus hydrophobes. Ajoutons que ce domaine est de même impliqué dans la reconnaissance d'extrémités spécifiques d'ADN viral ^[46].

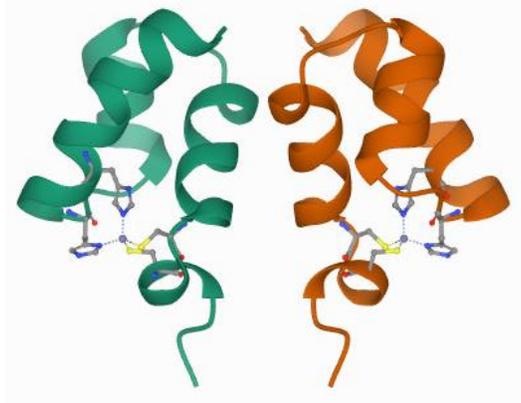


Figure 12: Structure RMN en solution du domaine N-terminal de l'intégrase du VIH-1. Les deux monomères de couleur orange et vert sont représentés par leurs hélices alpha. Les résidus HHCC du doigt de zinc sont représentés en bâtons (Code PDB : 1WJC)

(M. Cai, R. Zheng, M. Caffrey, R. Craigie, G. M. Clore, and A. M. Gronenborn,, *Nat. Struct. Biol.*, vol. 4, no. 7, pp. 567–577, Jul. 1997, doi: 10.1038/nsb0797-567.)

B.III.1.2. Le domaine central

Le domaine central, appelé aussi cœur catalytique comprend les résidus d'acides aminés 51 à 212. C'est le domaine le plus conservé au sein des intégrases rétrovirales. Il est caractérisé par un motif de trois résidus d'acides aminés (Asp 64, Asp 116 et Glu 152) qui constituent la triade catalytique ^[47]. Ces résidus hautement conservés sont très importants à l'activité catalytique de l'intégrase puisqu'une mutation au niveau de l'un d'entre eux suffit à supprimer l'infectivité du virus ^[47]. En fait, leur rôle majeur est la fixation de cations divalents, notamment le manganèse (Mn^{2+}) ou le magnésium (Mg^{2+}) ^[48]. A l'issue de cette fixation, l'intégrase adopte alors une conformation active ^[49]. Même si ce domaine contient les résidus requis pour l'activité catalytique, la totalité des trois domaines est nécessaire à

l'activité enzymatique complète de l'intégrase. Ajoutons que les régions à proximité du site actif sont dotées d'une flexibilité élevée qui leur permet d'induire un changement conformationnel au sein de l'intégrase suite à la liaison de l'ADN viral. En effet, la liaison du substrat d'ADN par le biais d'ions Mg^{2+} ou Mn^{2+} fixés au sein de la triade catalytique est nécessaire à l'adoption du domaine central de sa configuration active ^[50]. Les structures cristallographiques du domaine central de l'intégrase du VIH-1 ^[37] ont révélé une conformation en dimère globalement sphérique où chaque monomère comprend six hélices α et cinq feuillets β (Figure 13), avec une interface de dimérisation impliquant des interactions croisées entre les hélices $\alpha 1$ et $\alpha 5$ ^[51]. Il présente de plus un motif HTHi (Hélice-Tour-Hélice inversé) formé par les hélices $\alpha 4$ et $\alpha 5$. Cet épitope très réactif, identifié au sein du cœur catalytique de l'intégrase virale intervient dans la reconnaissance spécifique de l'ADN viral et non spécifique de l'ADN cellulaire ^[52].

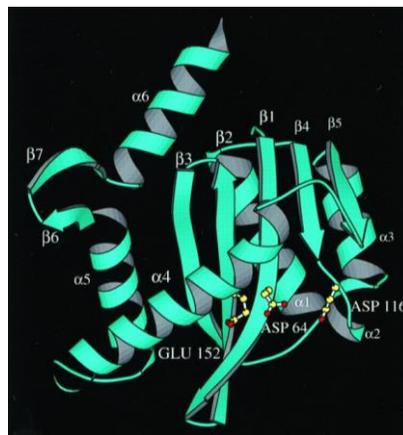


Figure 13 : Représentation en rubans de la structure secondaire du domaine central catalytique de l'intégrase du VIH-1. Les résidus de la triade catalytique (Asp64, Asp116 et Glu152) sont représentés en bâtons

(Y. Goldgur, F. Dyda, A.B. Hickman, T.M. Jenkins, R. Craigie, and D.R. Davis, *PNAS.*, vol. 95, no. 16, pp. 9150–9154, Aug. 1998, doi: 10.1073/pnas.95.16.9150)

B.III.1.3. Le domaine C-terminal

Le domaine C-terminal, constitué des résidus d'acides aminés de 213 à 288, est le moins conservé parmi les trois domaines de l'intégrase virale [31]. Il est principalement impliqué dans la liaison non spécifique à l'ADN cellulaire, favorisant ainsi la stabilisation du complexe ADN-IN [53]. Il interagit de même avec l'ADNv. En effet, des études par pontage covalent proposent une interaction du domaine C-terminal avec la région subterminale de l'extrémité de l'ADN viral [51].

En solution, le domaine CTD est présent sous forme dimérique ; sa structure déterminée par RMN exhibe deux protomères orientés parallèlement l'un à l'autre. Chaque unité monomérique est constituée de cinq brins β antiparallèles ayant une structure en tonneau [48] et un repliement de type SH3 (Src-Homology 3) [54] (Figure 14). Ajoutons que la surface de dimérisation est riche en acides aminés basiques (Lysine et Arginine) [55]. Ces résidus chargés positivement contribuent à l'accommodation de l'ADN par l'intégrase virale et à son interaction avec cette dernière. Ce domaine se trouve de même impliqué dans l'étape de maturation de l'ADNv et son transfert vers le cytoplasme cellulaire [28].

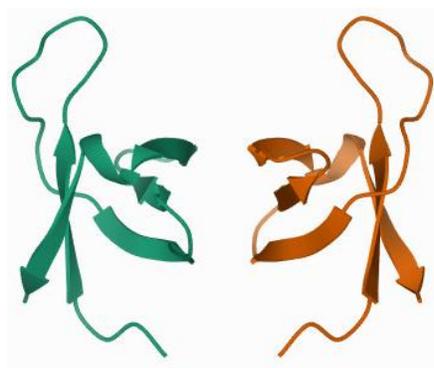


Figure 14 : Structure RMN en solution du domaine C-terminal de l'intégrase du VIH-1. Les deux monomères de couleur orange et vert sont représentés par leurs feuillets β antiparallèles en tonneau. (Code PDB : 1IHV)

(P. J. Lodi et al., *Biochemistry*, vol. 34, no. 31, pp. 9826–9833, Aug. 1995, doi: 10.1021/bi00031a002.)

B.III.2. Mécanisme d'action de l'intégrase

L'intégrase est l'enzyme rétrovirale qui catalyse une étape primordiale du cycle réplicatif du VIH-1 : l'insertion de l'ADN viral rétrotranscrit au niveau de l'ADN chromosomique de la cellule hôte. En fait, après l'étape de la transcription inverse, l'ADNv obtenu se trouve avec des extrémités franches correspondantes aux séquences LTR. Sous cette forme, il sera ensuite inséré dans le génome cellulaire lors de deux étapes principales; la maturation de l'ADN viral en 3' suivie de l'étape de transfert de brins ^[56]. La Figure 13 illustre l'ensemble de ces deux réactions.

B.III.2.1 Maturation en 3' de l'ADN viral

L'étape de maturation de l'ADN viral, connue aussi sous le nom de « 3' Processing » se déroule dans le cytoplasme de la cellule cible ^{[57][58]}. Le « 3' Processing » débute par la reconnaissance de l'intégrase des extrémités LTR au niveau 3' de l'ADNv, particulièrement une séquence conservée 5' CAGT 3' (Figure 15). L'intégrase effectue ensuite une attaque nucléophile à ce niveau où elle clive la liaison phosphodiester entre l'adénine et la guanine terminales, libérant un dinucléotide GT et exposant un dinucléotide CA qui présente alors une extrémité saillante 3'-OH ^{[59] [60] [61]}. Des travaux menés au sein de notre laboratoire ainsi que par d'autres équipes de recherche ont souligné l'importance des dix dernières paires de base pour la reconnaissance spécifique de l'ADNv par l'intégrase du VIH-1 ^{[62] [63]}. Des études supplémentaires sur ce processus ont révélé la présence de molécules d'eau intervenant au cours de cette réaction ^[64].

B.III.2.2 Etape de transfert de brin

Après le clivage en 3' de l'ADN viral par l'intégrase, cette dernière reste fixée à son substrat formant un complexe multimérique, connu sous le nom de complexe de

pré-intégration (PIC). Le PIC, qui comprend l'intégrase et l'ADN viral auxquels sont associées des protéines virales et cellulaires, est ensuite transporté au niveau du noyau de la cellule infectée ^[65]. Suite à cet import nucléaire, et après fixation du PIC à l'ADN cellulaire, l'intégrase catalyse une attaque nucléophile de l'extrémité 3'-OH réactive de l'ADN viral sur une liaison phosphodiester de l'ADN cellulaire. Ceci résulte en une liaison covalente qui se forme entre l'extrémité réactive CA-3'OH et le phosphate en 5' de l'ADN cellulaire, permettant ainsi l'insertion de l'ADN viral dans l'ADN cellulaire ^[66] (Figure 15). Cette intégration est finalisée par une étape de réparation du décalage de cinq paires de bases, issu de l'insertion de l'ADN viral dans l'ADN chromosomique. Ce sont des facteurs protéiques cellulaires qui sont impliqués dans cette réparation ^[67] ^[68].

Une fois intégré, l'ADN du VIH-1 peut alors utiliser la machinerie cellulaire afin de se répliquer et générer de nouvelles particules infectieuses, ce qui permet alors la propagation du virus dans l'organisme des personnes infectées.

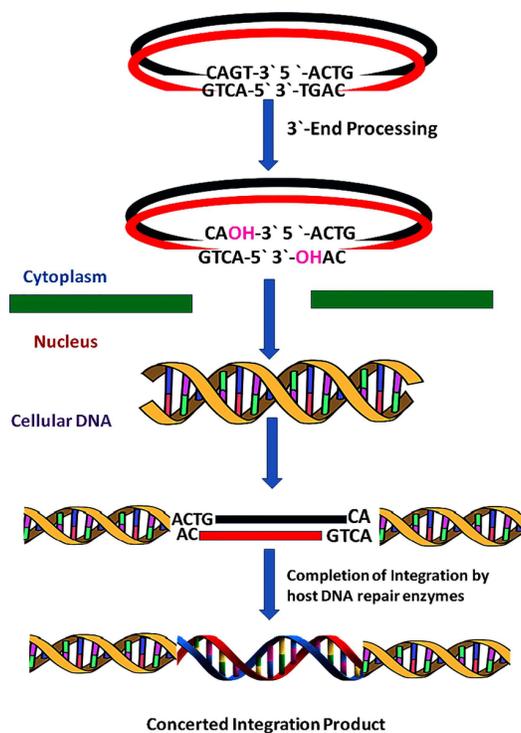


Figure 15: Schéma représentant les deux étapes du mécanisme d'action de l'intégrase du VIH-1 ; la maturation de l'ADN viral (3'-End Processing) suivi de l'étape de transfert de brin.

(P. Wadhwa, and P. Jain, *Current Drug Discovery Technologies*, no. 15, pp. 2-19, May 2017, doi: 10.2174/1570163814666170531115452)

B.III.3. Interactions de l'intégrase avec l'ADN viral

B.III.3.1 Régions de l'intégrase reconnues par l'ADN viral

C'est à partir de protéines chimériques qu'on a identifié la région de l'intégrase interagissant spécifiquement avec l'ADN viral ^[69]. Cette région, située à la surface du cœur catalytique, correspond à l'hélice $\alpha 4$ qui comprend les résidus 147 à 166. De même, on a caractérisé la boucle 140, en amont de l'hélice $\alpha 4$ (résidus 139 à 146), comme étant impliquée aussi dans les interactions IN-ADNv ^[63]. Ces deux entités sont hautement conservées au sein des intégrases rétrovirales (Figure 16).

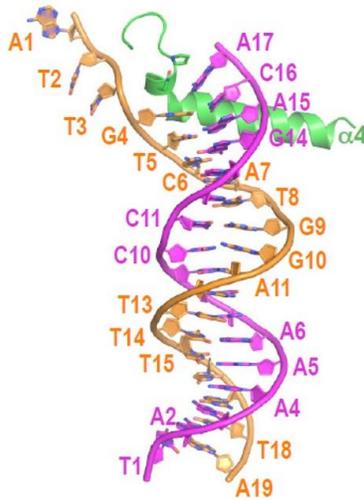


Figure 16 : Représentation schématique du double brin d'ADN viral du PFV, hautement homologue à celui du VIH-1. Le brin clivé mature (couleur mauve) comprend 17 résidus avec le dinucléotide réactif C₁₆-A₁₇ à l'extrémité 3'. Le brin non clivé (couleur orange) comprend 19 résidus avec une extrémité 5' saillante A1T2. L'hélice α 4 du domaine central ainsi que la boucle 140 en amont sont représentées en vert, montrant des contacts étroits avec les extrémités de l'ADN viral.

(S. Hare, S. Shree Gupta, E. Valkov, A. Engelman and P. Cherepanov. *Nature*, no. 464, pp. 232-236, March 2010, doi: 10.1038/nature08784)

Le motif d'hélice α 4 assure plusieurs fonctions au sein de l'intégrase, et ce à cause de son caractère amphiphile ; sa face hydrophobe apolaire, orientée vers l'intérieur de la protéine, est impliquée dans l'oligomérisation de l'IN alors que sa face hydrophile polaire, exposée vers l'extérieur, interagit avec l'ADN cellulaire et viral [35]. Au sein de cette face polaire existe en effet une région riche en lysine ; K156 – E157- L158 – K159 - K160, impliquée dans la fixation de l'enzyme à l'ADN viral [62] [63] (Figure 17). Les deux lysines K156 et K159 sont les piliers de cette fixation.

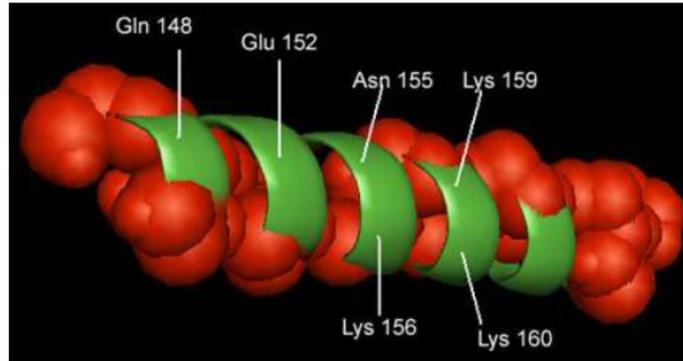


Figure 17 : Représentation de l'hélice amphiphile $\alpha 4$ par le programme « Pymol ». La surface en rouge simule la face hydrophobe et la surface en vert représente la face hydrophile de l'hélice. Les résidus principaux de la face hydrophile, interagissant avec l'ADN viral sont indiqués.

(Thèse de doctorat Zeina Hobaika, *Intégrase de VIH-1 : l'hélice $\alpha 4$, site d'interaction de l'ADN viral et d'inhibiteurs*, 2009)

En fait, des expériences de photo-pontage ont prouvé l'implication de K156 dans la maturation de l'ADNv et le transfert de brins ainsi que l'implication de K159 dans le positionnement de la liaison phosphate à cliver ($5'CA\downarrow GT3'$) lors de l'étape de maturation ^[70]. Ce positionnement est en effet favorisé par des contacts effectués entre K159 et l'adénine A_{17} (Figure 16) de la séquence terminale clivée en 3' ($5'C_{16}A_{17}GT3'$). Ces contacts sont les suivants: liaison K159 – atome N7 de l'adénine, liaison K159 – groupement NH_2 de l'adénine et liaison K159 – phosphate en 5' du phosphate clivé ^[71] ^[72]. La troisième lysine, K160, associée aux deux autres, est aussi cruciale à l'activité catalytique de l'intégrase et semble être importante pour sa dimérisation/tétramérisation ^[36].

Le dinucléotide en 5' du brin non clivé (A_1T_2 selon la Figure 16) interagit avec le cœur catalytique de l'IN, au niveau ou à proximité de la boucle flexible, puisque nous trouvons des interactions des résidus Y143 (de l'hélice $\alpha 4$) et Q148 (de la boucle 140) avec ce dinucléotide rendu flottant suite à la maturation de l'ADN viral ^[36] ^[64]. La mutation du résidu Q148 entraîne une baisse importante de l'infectivité du VIH-1, prouvant ainsi son rôle essentiel dans les interactions de l'intégrase avec

l'ADNv ^[73]. L'acide glutamique E152, de la triade catalytique de l'intégrase du VIH-1, interagit lui aussi avec l'ADN viral, et en particulier avec le dinucléotide C₁₆A₁₇ conservé en 3' ^[74] (Figure 18). Notons que la flexibilité de l'hélice α 4 et de sa boucle 140 permettent l'orientation de ce résidu vers les deux autres résidus de la triade catalytique : D64 et D116.

Ces rôles importants que jouent l'hélice α 4 et la boucle 140 ont été mis en relief et confirmés par des expériences de mutagenèse dirigée ^[75]. Par exemple, les mutations induites au niveau des résidus Lys156 (K156E) et Lys159 (K159E) engendrent une forte inhibition de la maturation en 3' et du transfert de brin ^[70]. De même, les mutations menées sur les résidus Val151 et Ser153 de la boucle 140 sont délétères surtout quand elles sont combinées à une mutation du résidu catalytique E152 ^[74].

B.III.3.2 Régions de l'ADN viral reconnues par l'intégrase

Comme nous l'avons déjà mentionné dans la partie B.III.2.1, l'intégrase doit reconnaître approximativement les dix derniers nucléotides à partir des extrémités 3' afin de catalyser la maturation et le transfert de brins de l'ADNv. Elle doit particulièrement identifier les dinucléotides terminaux CA et GT pour effectuer le clivage après l'adénine A₁₇ (Figure 16) libérant par conséquent la guanine et la thymine ^[68].

Des mutations au niveau des 6 premiers résidus à partir de l'extrémité 3' de l'ADN affectent considérablement l'activité de maturation du brin, ce qui souligne leurs contacts proches avec les résidus actifs de l'intégrase (Figure 18). Ces hypothèses ont été testées et validées *in vitro* ^[63] ainsi qu'*in vivo* ^[76].

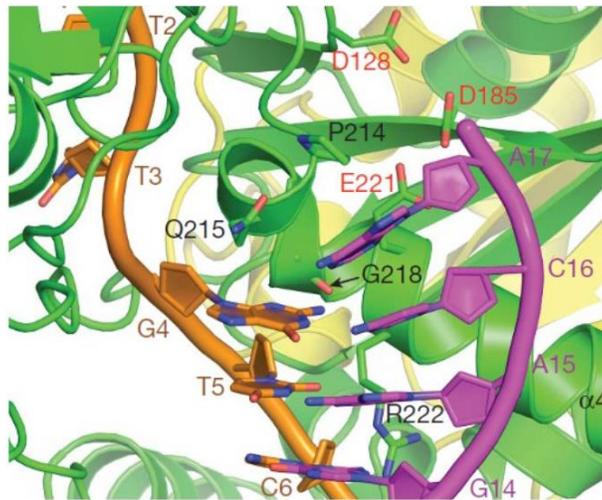


Figure 18: Représentation schématique des interactions entre l'intégrase du PFV et l'ADN viral. Le double brin correspond à l'ADN viral, montrant les nucléotides terminaux conservés au niveau du brin clivé (couleur mauve) : G14-A15-C16-A17 et montrant les cinq premiers résidus en 5' du brin non clivé (couleur orange). L'hélice $\alpha 4$ (couleur verte) est observée à proximité de l'extrémité réactive virale. Les résidus de la triade catalytique D128 (D64 pour VIH-1), D185 (D116 pour VIH-1) et E221 (E152 pour VIH-1) ainsi que les résidus P214 (P145 chez le VIH-1), Q215 (Q146 chez le VIH-1), G218 (G149 chez le VIH-1), R222 (S153) interagissant avec ces extrémités virales sont indiqués et représentés en bâtons.

(S. Hare, S. Shree Gupta, E. Valkov, A. Engelman and P. Cherepanov. *Nature*, no. 464, pp. 232-236, March 2010, doi: 10.1038/nature08784)

B.IV. Mécanisme d'inhibition de l'intégrase

B.IV.1. Thérapies anti-VIH-1

Avec l'expansion de la maladie du SIDA et le nombre élevé de personnes atteintes du VIH-1 annuellement, et en absence de vaccin, le développement des thérapies anti-SIDA se révèle être d'une importance majeure. D'où la conception, depuis l'apparition du virus, d'une panoplie de médicaments ciblant les différentes étapes du cycle répliatif viral. Ces antirétroviraux sont combinés dans une même multithérapie connue sous le nom de HAART qui consiste en une association d'au moins trois inhibiteurs à mécanisme d'action différents. Elle vise à diminuer la charge virale (nombre de particules virales dans l'organisme) et à stabiliser le nombre de lymphocytes T CD4⁺ (cellules cibles principales). De ce fait, elle a contribué à la diminution du taux de mortalité des personnes atteintes avec une

restauration du déficit immunitaire et de leur qualité de vie ^[77]. Les différentes molécules antirétrovirales sont classées en cinq groupes pharmacologiques, selon l'étape qu'elles ciblent au niveau du cycle viral comme la fusion, la transcription inverse ou la maturation protéolytique. Nous citons alors les inhibiteurs d'entrée du virus, les inhibiteurs de fusion, les inhibiteurs de la transcriptase inverse, les inhibiteurs de la protéase et finalement les inhibiteurs de l'intégrase.

Dans les parties qui suivent, nous allons présenter chaque classe d'inhibiteurs en détaillant précisément les inhibiteurs d'intégrase qui concernent notre travail.

B.IV.1.1. Les inhibiteurs de l'entrée du VIH-1

Les acteurs principaux impliqués dans cette première étape du cycle répliatif sont, au niveau cellulaire les récepteurs CD4⁺ et les corécepteurs CXCR4 et CCR5 situés principalement à la surface des cellules cibles immunitaires, et au niveau viral, c'est la glycoprotéine gp120. Ainsi, afin de bloquer l'entrée du virus, des inhibiteurs ciblant les trois acteurs cités ci-dessus sont utilisés dans les thérapies antirétrovirales. Parmi eux, nous trouvons le cyclotriazodisulfonamide connu aussi sous le nom de CADA, qui régule l'expression de CD4⁺ en la diminuant au niveau post traductionnel ^[78]. Il présente de même un effet synergétique une fois combiné à d'autres inhibiteurs du virus ^[79].

Pour les corécepteurs CXCR4 et CCR5, interagissant avec les glycoprotéines virales et les chimiokines, leur inhibition s'effectue par des molécules antagonistes chargées positivement et capables de réduire l'activité infectieuse du virus ^[80]. En particulier, nous trouvons la molécule AMD11070 qui a été le premier antagoniste de CXCR4 disponible par voie orale à entrer en phase clinique ^[81] et le Maraviroc qui a été découvert comme antagoniste pour CCR5. C'est la seule molécule

inhibitrice ciblant le récepteur CCR5 qui a été approuvée par l'US FDA et par l'Agence Européenne des médicaments (EMA) ^[82].

Concernant la glycoprotéine de surface gp120, son blocage est effectué par les lectines de plantes dérivant de « *Galanthus nivalis* » et de « *Hippeastrum hybrid* ». Elles sont spécifiques du mannose et agissent en provoquant des changements au niveau des sites de N-glycosylation pour les résidus Sérine et Thréonine au niveau de la gp120. Ceci inhibe la reconnaissance de gp120 par les récepteurs et les corécepteurs cellulaires, et obstruent donc l'entrée du virus ^[83]. Des études plus récentes ont identifié deux nouvelles classes d'inhibiteurs anti-gp120, qui agissent spécifiquement au niveau des interactions gp120-CD4⁺ et les inhibent. Ce sont des composés protéomimétiques qui se lient à la surface extérieure de gp120 et modulent les interactions protéine-protéine médiée par la glycoprotéine de surface ^[84], et des composés 2-Aminothiazolones qui inhibent les interactions gp120-CD4⁺. Ces derniers ont été identifiés comme non toxiques avec un spectre d'action large ^[85].

B.IV.1.2. Les inhibiteurs de la fusion virus-cellule

Le seul inhibiteur de la fusion de la membrane du VIH-1 avec la membrane cellulaire cible est l'Enfuvirtide (T-20). Il s'agit d'un peptide synthétique de 36 acides aminés mimant la séquence 127-162 trouvée chez la glycoprotéine transmembranaire gp41. Il inhibe la dernière étape de l'entrée du virus par sa liaison compétitive aux domaines hydrophobes de gp41, démasqués après la liaison du virus aux récepteurs CD4⁺ et aux corécepteurs CXCR4 et CCR5. De ce fait, il bloque la formation de la structure virale post-fusion, ralentissant alors l'infection virale ^[86] ^[87]. Cependant, suite à une activité antivirale relativement faible de T-20 et à son induction de résistance chez les particules virales, la nécessité de développer de

nouvelles générations de cet inhibiteur est devenue importante. Du coup, depuis les trois dernières années, diverses équipes travaillent sur la conception de lipopeptides basés sur T-20 et inhibant la fusion du VIH-1 avec un meilleur rendement et une meilleure efficacité. On cite parmi ces molécules la LP-40 (Lipopeptide 40), LP-50 (Lipopeptide 50) et LP-51 (Lipopeptide 51), issues du raffinement de la structure et de la fonction de T-20 ^[88] ^[89] ^[90].

B.IV.1.3. Les inhibiteurs de la transcription inverse

On reconnaît deux classes d'inhibiteurs de la transcription inverse ; les inhibiteurs nucléosidiques/nucléotidiques ou NRTIs (Nucleoside/Nucleotide Reverse Transcriptase Inhibitors) et les inhibiteurs non-nucléosidiques/non-nucléotidiques ou NNRTIs (Non-Nucleoside/Non-Nucleotide Reverse Transcriptase Inhibitors). Aujourd'hui, ils sont utilisés dans le cadre de la trithérapie HAART où on administre deux NRTIs combinés à un NNRTIs ou à un inhibiteur de la protéase.

Les premiers agents approuvés comme inhibiteurs de la transcriptase inverse (RT) étaient des inhibiteurs NRTIs, avec l'Azidothymidine (AZT) comme premier anti-RT mis sur le marché ^[91]. Les NRTIs sont des analogues nucléosidiques ou nucléotidiques dépourvus de leur groupement hydroxyle (-OH) en 3' du désoxyribose. Ils interfèrent avec la RT en inhibant l'élongation de l'ADN du VIH-1 issu de la transcription de l'ARN viral. Les exemples d'inhibiteurs nucléosidiques/nucléotidiques de la RT incluent : la Zidonedine (ZDV), l'Abacaltir, la Lamivudine et le Tenofovir. Suite à leur administration, ces inhibiteurs sont convertis par triphosphorylation en leur forme biologique active ^[92] ^[93]. Quant aux inhibiteurs non nucléosidiques/non nucléotidiques de la RT, leur mode d'action consiste à modifier la configuration de l'enzyme par fixation à un domaine à

l'extérieur du site catalytique, résultant d'une désactivation de la RT ^[94]. Ce sont donc des inhibiteurs non compétitifs de la reverse transcriptase et qui, contrairement aux NRTIs, n'exigent pas de phosphorylation pour être activés. On reconnaît trois molécules de 1^{ère} génération présentes sur le marché ; Efavirenz (EFV), Nevirapine (NVP) et Delavirdine (DLV), qui après avoir manifesté une faible barrière génétique ainsi qu'un faible profil de résistance, ont été remplacées par des NNRTIs de 2^{ème} génération ; Etravirine (ETR) et Rilpivirine (RPV) ^{[95] [96]}.

B.IV.1.4. Les inhibiteurs de l'activité protéolytique

Le rôle indispensable que joue la protéase virale (PR) dans le cycle répliatif du VIH-1 en fait une cible populaire des médicaments antirétroviraux. Comme déjà mentionné dans le paragraphe B.II.3, cette enzyme a pour rôle de cliver la polyprotéine issue de la traduction de l'ADN viral. Ce clivage est nécessaire afin de générer les différentes protéines matures du VIH-1.

Les inhibiteurs conçus contre la PR agissent en empêchant le clivage de la polyprotéine au niveau du précurseur gag-pol, ce qui rend les virions libérés immatures et non infectieux ^[97]. Il existe dix inhibiteurs de protéase approuvés par la FDA dont les plus connus sont le Saquinavir, le Ritonavir, l'Indinavir et le Lopinavir, qui partagent des similitudes au niveau de leur structure et de leur mode de liaison. En effet, ils effectuent des liaisons hydrogènes via leur groupement hydroxyle (-OH) avec le groupement carboxyle (-COOH) des résidus du site actif de la protéase ^[98]. Cependant, la majorité de ces inhibiteurs ont présenté des effets secondaires dans les traitements à long terme.

B.IV.2. Inhibition de l'intégrase virale

Comme déjà mentionné dans le paragraphe B.I, les progrès observés suite à l'application de la thérapie HAART et donc à l'utilisation de divers médicaments anti-VIH-1, ont été contrebalancés par l'émergence de mutations au niveau des protéines virales, causant une résistance aux drogues administrées. De plus, ces stratégies thérapeutiques sont coûteuses et nécessitent une discipline stricte, d'où la nécessité d'adopter de nouvelles approches se focalisant sur une nouvelle cible thérapeutique cruciale ; l'intégrase virale. Nous rappelons que l'importance de cette enzyme vient du fait qu'elle joue un rôle clé dans l'insertion et l'expression du matériel génétique du VIH-1 au niveau de sa cible et du fait qu'elle ne présente pas de contrepartie cellulaire donnant alors lieu à une inhibition spécifique (voir paragraphe B.III.)

B.IV.2.1 Les inhibiteurs dicétoniques (DKAs)

Grand nombre d'inhibiteurs de l'intégrase à été découvert depuis les années 90 par des tests de criblage *in vitro* utilisant l'intégrase recombinante et de petits oligonucléotides mimant les extrémités de l'ADN viral ^[99] ^[100]. On classe ces inhibiteurs en trois familles selon les critères suivants : leur capacité antivirale, leur cytotoxicité, leur activité *in vivo* et leur spécificité. Certains présentent des activités spécifiques contre l'intégrase virale, d'autres possèdent des activités non spécifiques ^[101]. Ces familles sont les phenoldipyrimides (PLP), les stryrylquinolines (SQL) et la famille des dicéto-acides (DKA). Ces derniers se sont avérés être les meilleurs inhibiteurs, possédant l'activité anti-intégrase la plus spécifique et la plus efficace par rapport aux deux autres familles. Ils ciblent en fait sélectivement l'intégrase dans les cellules infectées.

Les dicéto-acides, développés par les laboratoires Shionogi et Merck, interfèrent majoritairement avec la deuxième réaction catalysée par l'IN virale ; l'étape de transfert de brin, d'où leur nom d'inhibiteurs de transfert de brin (INSTIs) ^[102] ^[103]. Ils agissent à des concentrations du nanomolaire et agissent au sein du complexe IN-ADN viral-ADN cellulaire, inhibant l'intégration de l'ADNv au sein de l'ADN de la cellule hôte. Leur structure comprend les parties communes principales citées ci-dessous et indiquées à la Figure 19.

- Un groupement d'hétéroatomes coplanaires (atomes d'oxygène ou d'azote) présentant une fonction dicétone. Ce groupement est connu pour son aptitude à chélater les cations métalliques du site actif (Mg^{2+} ou Mn^{2+}), tout en s'empilant contre la base virale A_{17} au niveau 3' de l'ADN viral traité ^[104] ^[105]. Une fonction d'alcool juxtaposant la dicétone favorise l'éjection du 3'-OH réactif loin du site actif de l'intégrase.
- Une fonction aromatique qui peut porter une variété de substituants tels que des halogènes. Elle intervient directement dans l'inhibition de l'étape sélective du transfert de brin suite à son interaction avec les paires de bases terminales C_{16} et G_4 situés aux extrémités des brins traité et non traité respectivement ^[106] ^[107].
- Un « linker » liant les deux fonctions (dicétone et aromatique) dont la taille diffère selon l'inhibiteur

Ces différentes structures mettent en évidence le rôle crucial des fonctions des DKAs dans l'inhibition métallodépendante de l'intégrase du VIH-1 ^[103].

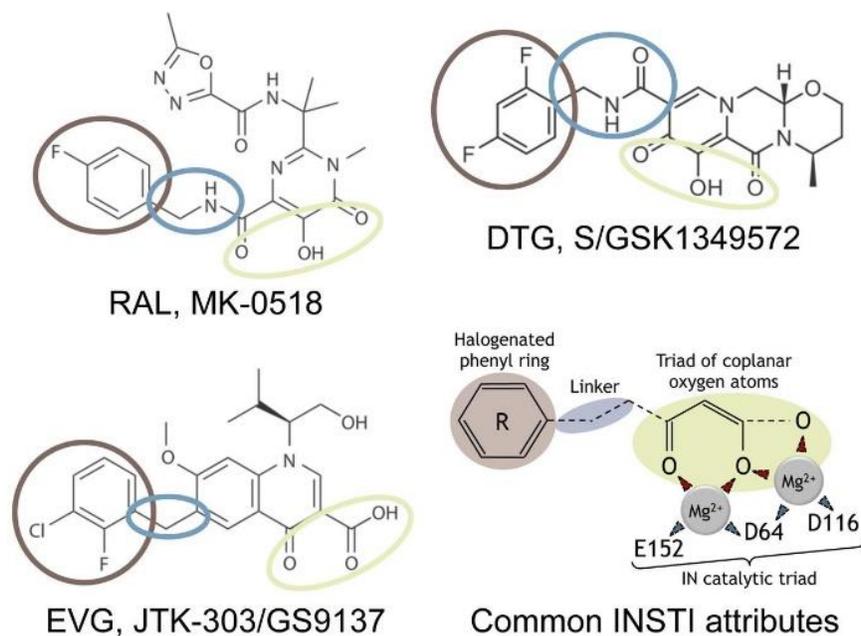


Figure 19 : Structure chimique des trois inhibiteurs de transfert de brin approuvés par la FDA ; RAL (MK-0518), EVG (JTK-303/GS9137) et DTG (S/GSK1349572). Les fonctions communes sont encerclées et indiquées dans le schéma en bas à droite. En beige est encerclée la triade d'atomes d'oxygène coplanaires qui chélatent deux ions Mg^{2+} . En bleu est encerclé le « linker » de longueur et de flexibilité variables. En marron est encerclé le noyau aromatique présentant des substitutions d'halogènes. Les résidus de la triade catalytique et les deux cations divalents de magnésium sont de même représentés.

(E. Thierry, E. Deprez, O. Delelis, *Frontiers in Microbiology*, no. 7, January 2017, doi: 10.3389/fmicb.2016.02165)

B.IV.2.1.1 Le mécanisme d'action des INSTIs

Les INSTIs inhibent particulièrement le mécanisme de transfert de brin, en agissant à l'interface ADN-IN. Ils présentent par la suite aussi bien des interactions avec l'intégrase virale qu'avec l'ADN du VIH-1 (Figure 20).

Au niveau de l'intégrase, ce sont les résidus interagissant avec l'ADNv qui sont de même impliqués dans les interactions avec les dicéto-acides. Ils sont donc localisés au niveau du cœur catalytique et plus précisément au niveau de l'hélice $\alpha 4$ et de la boucle139. En fait, des mutations de résistance caractérisées au niveau de ces résidus cibles ont diminué l'affinité entre la molécule inhibitrice et le complexe préformé IN-ADNv, ce qui confirme leur implication dans les interactions IN-DKAs

[102] [108]. On cite en particulier les résidus de l'hélice $\alpha 4$: Val151, Glu 152, Ser153 et Asn155, Lys156 et Lys159 [109]. De même, une mutation au niveau du résidu M154 (M154I) donne lieu à une forte résistance de l'intégrase aux dicéto-acides [110]. Le résidu Gln148 semble également être engagé dans l'interaction avec les INSTIs [111]. Au sein de la boucle 140, le résidu Y143 interagit aussi avec l'inhibiteur de transfert de brin [112]. L'implication de l'ensemble de ces résidus a été de même mise en évidence par une étude RMN menée au sein de notre équipe [62].

Au niveau de l'ADN viral, l'interaction des inhibiteurs dicéto-acides se fait en majorité avec C₁₆ et A₁₇ de l'extrémité 3' réactive et avec G4 de l'extrémité 5' non réactive [53] [113].

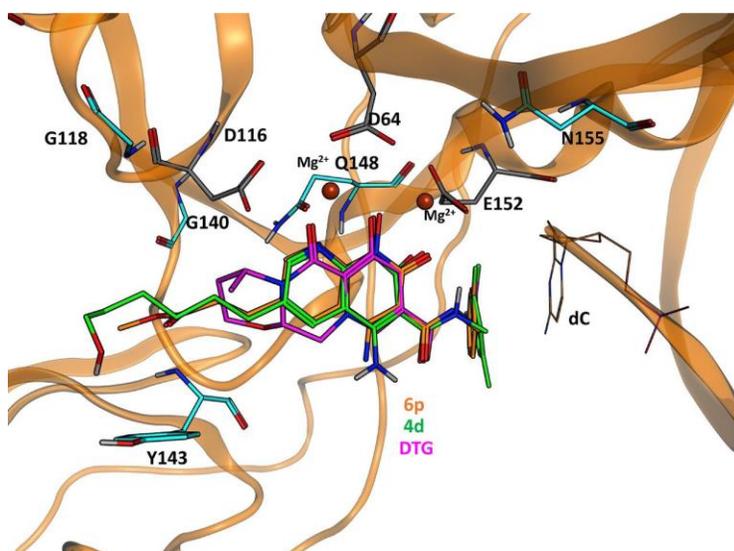


Figure 20 : Modélisation de trois inhibiteurs de transfert de brin (6p en orange, 4d en vert et DTG en magenta) au sein de l'intasome du VIH-1 (intégrase virale et ADN viral). L'intégrase est représentée en orange avec une mise en relief des résidus de la triade catalytique (D64, D116 et E152) ainsi que les résidus qui entrent en interaction avec les trois molécules inhibitrices. Les deux cations divalents de Mg²⁺, chélatés par la fonction carboxylate des inhibiteurs, sont représentés en sphères marron. La cytosine terminale correspondant à C₁₆ est indiquée sous le nom de dC. Elle interagit par empilement avec le noyau aromatique des inhibiteurs.

(S. J. Smith, X. Z. Zhao, T. R. Burke Jr, S. H. Hughes, *Antimicrobial Agents and Chemotherapy*, vol. 62, no. 9, September 2018, doi: 10.1128/AAC.01035-18)

Toutes ces données ont abouti à l'élaboration d'un modèle de mécanisme d'action des DKAs afin d'expliquer leur sélectivité pour la réaction de transfert de brin. Dans ce modèle (Figure 21), les DKAs se fixeraient à l'interface des trois partenaires ; IN-ADNv-Mg²⁺ après la maturation de l'ADN en 3'. Ils figeraient par la suite le complexe intermédiaire issu de la maturation et empêcheraient son transfert vers le noyau cellulaire [114].

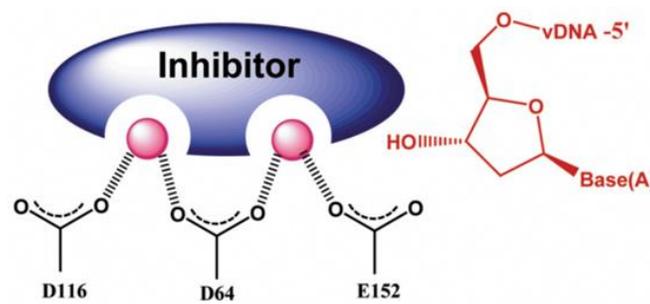


Figure 21 : Schéma simplifié représentant le mécanisme d'action des inhibiteurs de transfert de brin du VIH-1. L'inhibiteur (en bleu) est observé situé entre le résidu terminal (Adénine) de l'ADN viral, à proximité du groupement –OH saillant, et la triade catalytique de l'intégrase (résidus D64, D116 et E152). Il chélate les deux ions Mg²⁺ (sphères roses) nécessaires à l'activité du site actif.

(J. Mouscadet and L. Tchertanov, *European Journal of Medical Research*, vol. 14, no. 5, November 2009, doi: 10.1186/2047-783X-14-S3-5)

Dans les parties qui suivent, nous allons détailler la structure et le mode d'action des trois inhibiteurs de transfert de brin mis sur le marché et faisant partie de la multithérapie, le Raltégravir (RAL), l'Elvitégravir (EVG) et le Dolutégravir (DTG).

B.IV.2.1.2 Raltégravir

Le Raltégravir (RAL), connu également sous le nom de MK-0518 (Figure 19), est le premier inhibiteur de transfert de brin mis sur le marché en 2007, après approbation par la FDA. Développé par Merck et Co, il montre un potentiel antiviral efficace et rapide chez les patients traités, avec une concentration inhibitrice à 95% (IC95) de 33nM [115][116][117]. Sa structure comprend trois groupements chimiques :

l'isopropyl, le méthyloxadiazole et le fluorobenzène (Figure 19). Comme signalé auparavant, en absence de données cristallographiques sur l'intégrase entière en présence des inhibiteurs de transfert de brin (à cause de la faible solubilité de l'intégrase du VIH-1), nous nous sommes basés sur la structure de l'intégrase du PFV comme référence (Prototype Foamy Virus) puisqu'elle présente une homologie de haut niveau avec l'intégrase du VIH-1. La structure cristallographique de l'intégrase du PFV en présence du RAL et du magnésium (code PDB : 3OYA) a permis d'élucider son mode d'action à l'interface intégrase-ADN viral. On observe en effet au sein de l'intasome viral suite à la présence du Raltégravir, deux interactions essentielles via deux motifs conservés de ce dernier ; une chélation des deux ions Mg^{2+} par ses atomes d'oxygène coplanaires et un empilement du noyau aromatique fluoré de RAL contre la cytosine C_{16} au niveau 3' de l'ADNv réactif (Figure 22). L'empilement des deux noyaux est favorisé par une distorsion de 300 degrés de la boucle 139 en amont de l'hélice α_4 , provoquée par RAL, qui résulte en un éloignement du 3'-OH saillant d' A_{17} du site catalytique. Ceci crée alors une poche hydrophobe qui sera occupée par le noyau aromatique halogéné du RAL, qui en plus de son empilement avec C_{16} , interagira par des réactions électrostatiques avec la guanine G_4 complémentaire à C_{16} . Ces observations sont consistantes avec la preuve que la flexibilité de l'adénosine terminale réactive est importante pour la liaison efficace des INSTIs ^[118], vu que son déplacement pourrait expliquer le mécanisme d'inhibition du transfert de brin ^[119]. Outre ses interactions avec l'ADN viral, le RAL révèle aussi des interactions avec l'intégrase étant donné qu'il est situé à l'interface ADNv-Intégrase. C'est avec la proline en position 214 (qui correspond à la proline 145 pour le VIH-1) et la glutamine en position 215 (qui correspond à la glutamine 146 pour le VIH-1) que des interactions ont été observées avec le RAL,

au niveau de son groupement isopropyl. Elles contribuent à isoler l'extrémité 3'OH du brin d'ADN viral traité du site actif de l'intégrase. On a de même identifié des empilements de type π - π entre le noyau oxadiazole du RAL et celui de la tyrosine 212 (qui correspond à la tyrosine 143 pour le VIH-1) [40] (Figure 19).

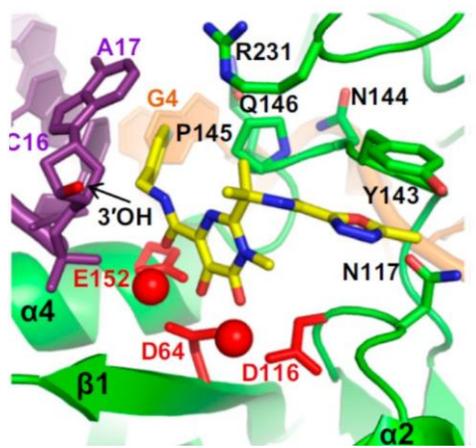


Figure 22 : Mode de liaison du Raltégravir au sein de l'intasome du VIH-1. Les deux bases terminales réactives C₁₆ et A₁₇ de l'ADN viral sont représentées en mauve avec l'extrémité réactive 3'-OH d'A₁₇ mise en relief en rouge. L'intégrase virale est en vert avec une indication des résidus impliqués dans les contacts directs avec la drogue. La triade catalytique (D64, D116 et E152) est représentée en bâtons rouges et les deux cations divalents en sphère rouge. Le Raltégravir est représenté en jaune (atomes de carbone), bleu (atomes d'azote) et rouge (atomes d'oxygène). Il fait des empilements d'une part avec Y143 via son noyau d'isopropyl et d'autre part avec C₁₆ via son noyau d'halobenzène. Il contribue à éloigner le 3'OH du site catalytique.

(L. Krishnan, X. Li, H. L. Naraharisetty, S. Hare, P. Cherepanov and A. Engelman, *PNAS*, vol. 107, no. 36, pp. 15910-15915, September 2010, doi: 10.1073/pnas.1002346107)

B.IV.2.1.3 Elvitégravir

L'efficacité du Raltégravir dans les traitements anti-intégrase a malheureusement été limitée par sa sensibilité aux mutations de résistance (voir le paragraphe B.IV.2.2). Ces limitations ont incité le développement d'agents présentant une barrière plus robuste face à ces mutations, d'où le développement de l'Elvitégravir (EVG, JTK-303 ou GS-9137) (Figure 19) par Gilead Sciences et sa mise sur le marché en 2012.

L'Elvitégravir est légèrement plus puissant que le Raltégravir, ayant montré *in vitro* une activité anti-intégrase importante avec une concentration efficace à 90% (EC90) de 1.7nM [120] [121] [122]. Sa structure reste toutefois similaire à celle du RAL mais avec un chlore et un fluor comme halogènes au niveau de son noyau aromatique et il est dépourvu de noyau oxadiazole (Figure 19). Pour cela, son mode d'interaction au sein de l'intasome du VIH-1 présente des points communs avec le Raltégravir. En fait, la structure cristalline d'EVG en présence du complexe binaire IN-ADNv du PFV et du magnésium (code PDB : 3L2U) indique un mode de fixation de l'INSTI à l'interface IN-ADNv. Tout comme le RAL, l'EVG interagit avec les bases terminales réactives des extrémités 3' (C₁₆ et A₁₇) et 5' (G₄) de l'ADN viral, favorisant un déplacement de l'adénine loin du site actif. En ce qui concerne son interaction avec l'intégrase, l'EVG interagit plus étroitement avec la glutamine 215 (glutamine 145 du VIH-1). N'ayant pas de noyau oxadiazole, il ne fait pas des empilements avec le noyau de la tyrosine 213 (tyrosine 143 du VIH-1), ce qui lui permet d'échapper à la mutation de résistance au niveau de ce résidu, confirmant sa moindre sensibilité comparé à RAL [39]. La Figure 23 illustre l'ensemble de ces interactions.

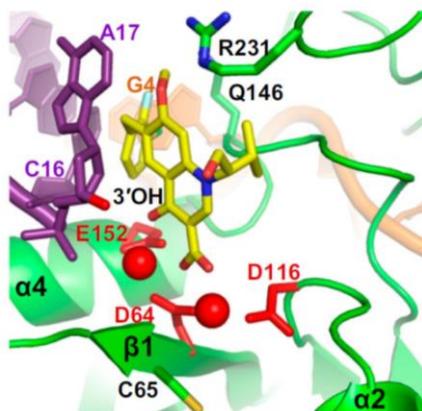


Figure 23 : Mode de liaison de l'Elvitégravir au sein de l'intasome du VIH-1. Les deux bases terminales réactives C₁₆ et A₁₇ de l'ADN viral sont représentées en mauve avec l'extrémité réactive 3'-OH d'A₁₇ mise en relief en rouge. L'intégrase virale est en vert avec une indication des résidus impliqués dans les contacts directs avec la drogue. La triade catalytique (D64, D116 et E152) est représentée en bâtons rouges et les deux cations divalents en sphères rouges. L'Elvitégravir est représenté en jaune (atomes de carbone), bleu (atomes d'azote) et rouge (atomes d'oxygène). Il fait des empilements avec C₁₆ via son noyau d'halobenzène, contribuant à éloigner le 3'OH du site catalytique. Notons l'absence de représentation de la tyrosine 143 puisqu'elle n'interagit pas avec l'Elvitégravir dépourvu d'isopropyl.

(L. Krishnan, X. Li, H. L. Naraharisetty, S. Hare, P. Cherepanov and A. Engelman, *PNAS*, vol. 107, no. 36, pp. 15910-15915, September 2010, doi: 10.1073/pnas.1002346107)

Toutefois, l'EVG a affiché une résistance croisée avec le RAL (paragraphe B.IV.2.2), et les travaux de recherche se sont poursuivis afin de développer une nouvelle génération d'inhibiteurs de transfert de brin plus efficaces, pouvant échapper aux mutations de résistance développées contre le RAL et l'EVG.

B.IV.2.1.4 Dolutégravir

Le Dolutégravir (DTG ou SGK) est l'inhibiteur de transfert de brin de 2^{ème} génération le plus avancé. Il a été introduit sur le marché en 2013 par Shionogi-GlaxoSmithKline présentant une efficacité *in vivo* et *in vitro* et doté d'une concentration inhibitrice à 50% (IC₅₀) dans le domaine des nanomolaires. De structure similaire à ses antécédents, il possède cependant deux atomes de fluor au niveau de son noyau aromatique et un « linker » plus long (Figure 19). Sa dissociation au niveau du complexe ADN viral – IN est plus lente par rapport aux

deux autres INSTIs et il présente une plus faible sensibilité aux mutations de résistance émergentes de l'intégrase. Il reste donc jusqu'à aujourd'hui l'INSTI le plus potentiel à être utilisé dans les thérapies anti-intégrase ^[123] ^[124] ^[125].

La structure cristallographique du DTG en complexe avec l'intasome du PFV (ADN viral + Intégrase) et le magnésium, établie par l'équipe de Hare en 2011, propose un mode d'inhibition du mécanisme de transfert de brin du Dolutégravir similaire à celui du RAL ou d'EVG, impliquant tout d'abord une chélation des cations divalents à l'intérieur du site actif par ses trois atomes d'oxygène coplanaires. Sa fonction aromatique de difluorophényle occupe la poche libérée par l'adénine terminale virale de l'extrémité 3', et effectue des contacts étroits avec la cytosine C₁₆. En fait, le « linker » du DTG est plus long et plus flexible en comparaison aux « linkers » des autres INSTIs ; six liaisons connectent le noyau halogéné du Dolutégravir à son cœur chélateur (alors que le RAL par exemple ne possède que quatre liaisons au sein de son « linker ») (Figure 24). Ceci permet alors au DTG de s'enfoncer plus profondément dans la poche libre et de réaliser des contacts plus proches avec la cytosine, confirmant ainsi son efficacité accrue par rapport aux inhibiteurs de 1^{ère} génération.

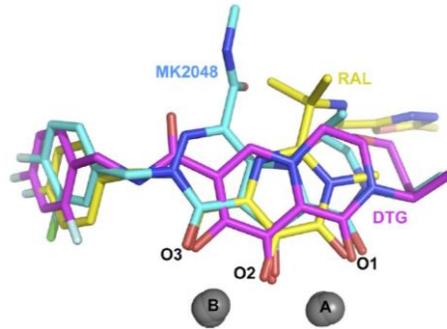


Figure 24 : Superposition de trois inhibiteurs de transfert de brin ; RAL (en jaune), DTG (en magenta) et MK-2048 (en bleu), avec leurs trois atomes d'oxygènes coplanaires indiqués par O1, O2 et O3, chélatant deux ions de magnésium représentés par deux sphères grises nommées A et B. La longueur de leurs « linkers », reliant le noyau halogéné au cœur chélateur de l'inhibiteur, diffère entre les trois, avec celui de DTG étant le plus long.

(S.Hare et al, *Mol.Pharmacol*, vol. 80, no. 4, pp. 565-572, June 2011, doi: 10.1124/mol.111.073189)

D'autre part, les deux substitutions de fluor du noyau aromatique du DTG établissent des interactions au sein de l'intasome ; le 2-fluoro réalise des interactions de vdW avec le résidu catalytique E221 (qui correspond à E152 pour le VIH-1) alors que le 4-fluoro apparaît très proche de la guanine G₄ de l'extrémité 5' du brin non traité, via des interactions électrostatiques. Des empilements sont également observés entre le noyau chélateur de la drogue et celui de l'adénine A₁₇ qui présente des instabilités au niveau de sa position suite à son déplacement. De plus, la fixation du Dolutégravir à l'interface ADN-IN affecte la structure de l'enzyme; une déstabilisation de la chaîne latérale de la glutamine 215 (qui correspond à la glutamine 146 pour le VIH-1) est observée, ainsi que des interactions sont affichées entre le DTG et la glycine 187 (qui correspond à la glycine 118 du VIH-1) à proximité du site actif, puisqu'il s'étend vers elle au sein de l'intasome. Les interactions du DTG sont faibles avec la tyrosine 212 (qui correspond à la tyrosine 143 du VIH-1), puisque contrairement au RAL, il ne possède pas de noyau oxadiazole donc ne s'empile pas avec le noyau de la tyrosine. La Figure 25 illustre l'ensemble de ces interactions.

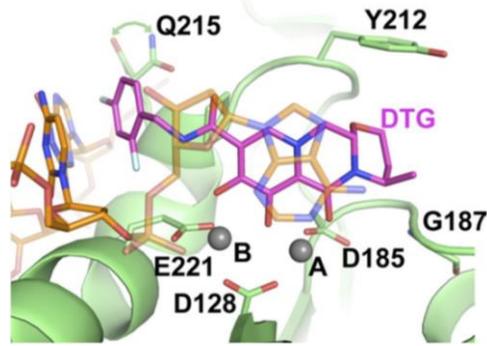


Figure 25 : Fixation du DTG au niveau du site actif de l'intasome du PFV. Les atomes de carbone du DTG sont en magenta, les atomes d'azote sont en bleu foncé, les atomes d'oxygène sont en rouges et les atomes de fluor sont en bleu clair. L'intégrase est représentée en vert et les bases de l'ADN viral en orange. Les atomes de Mg^{2+} chélatés par les oxygènes du DTG sont schématisés en sphères grises nommées A et B. Les résidus de la triade catalytique de l'intégrase ainsi que ceux impliqués dans les interactions avec le DTG sont représentés en bâtons. Pour plus de clarté, les bases terminales de l'ADN viral sont en mode semi-transparent. Les flèches vertes pâles indiquent la déstabilisation du résidu Q215.

(S.Hare et al, *Mol.Pharmacol*, vol. 80, no. 4, pp. 565-572, June 2011, doi: 10.1124/mol.111.073189)

B.IV.2.2 Emergence des mutations de résistances contre les INSTIs

Malgré les succès observés depuis 2007 suite à l'adoption des inhibiteurs de transfert de brin dans les thérapies antivirales, des mutations de résistance sont quand même apparues dans la région codante du gène pol, donnant ainsi lieu à une intégrase virale présentant des modifications et qui fait face aux drogues inhibitrices ce qui réduit la sensibilité du virus aux inhibiteurs et diminue par la suite l'efficacité des traitements cliniques. Nous parlons de mutations primaires celles qui sont fortement associées à une résistance accrue de l'intégrase du VIH-1 aux antirétroviraux. Alors que les mutations secondaires sont celles qui, combinées à d'autres mutations, confèrent à la protéine virale une résistance accrue à un médicament.

Dans le cas du Raltégravir, les mutations primaires apparues suite à son contact avec l'intégrase, sont associées aux résidus Y143R/C/H, N155H et Q148H/R/K. Les deux dernières sont aussi des mutations primaires de résistance croisées avec

l'Elvitégravir. Elles aboutissent à un échec des traitements ; la mutation N115H réduit les interactions de vdW entre ce résidu et les deux résidus catalytiques D64 et E152, elle altère alors la géométrie de l'interaction autour des ions Mg^{2+} diminuant ainsi la flexibilité de l'IN face à RAL et EVG. Quant à la mutation Q148H/R/K, elle est généralement accompagnée d'une mutation secondaire G140S/A/C qui évolue avec elle. Ce double mutant semble dissocier un réseau de molécules d'eau établi à l'interface drogue-protéine et qui est d'une grande importance pour l'interaction de l'inhibiteur avec l'IN, notamment dans le cas du RAL ^[113] ^[126] ^[127]. La mutation Y143R supprime l'interaction par empilement π de ce résidu avec le noyau oxadiazole du RAL, elle n'affecte donc que cette drogue vu que l'EVG et le DTG sont dépourvus de ce noyau dans leur structure. C'est pourquoi ces deux derniers restent efficaces contre l'intégrase modifiée au niveau Y143 ^[128]. Des mutations de résistance primaire sont apparues exclusivement dans le cas d'EVG. Elles sont au niveau des résidus T66I, S147G et E92Q. Il existe de même des mutations de résistances secondaires apparues dans l'intégrase du VIH-1 exclusivement suite à l'administration de RAL ; E138K et D232N et celles apparues suite à l'administration d'EVG ; L68I/N et S119 R/G, et des mutations secondaires croisées entre les deux INSTIS. A titre d'exemple, nous citons : H51Y, T66A/K, K160N, R166S, E170A, S230R et R263K ^[129] ^[130]. Le Dolutégravir, quant à lui, a montré cliniquement une meilleure efficacité contre les intégrases ayant les mutations primaires Y143R et N155H mais a présenté une susceptibilité moyenne envers la double mutation Q148H/G140S ^[113]. Il n'a cependant pas montré des résistances croisées avec les autres mutations primaires et secondaires de RAL et d'EVG ; sa liaison avec ces intégrases porteuses de mutations est en fait plus prolongée ^[131] ^[132] ^[124]. Le DTG est donc une drogue prometteuse pouvant contourner l'obstacle de la

résistance, et contribuer ainsi à inhibition plus efficace de la réplication VIH-1 sauvage ou muté.

Notons qu'un nouvel inhibiteur de transfert de brin, le Bictégravir (BIC) a été très récemment approuvé par la FDA (en 2018) pour des utilisations en combinaison avec d'autres inhibiteurs dans les thérapies anti-VIH. De même génération que le DTG et de structure similaire (Figure 26), son profil de résistance *in vitro* est comparable à celui de DTG et est nettement amélioré par rapport à RAL et EVG.

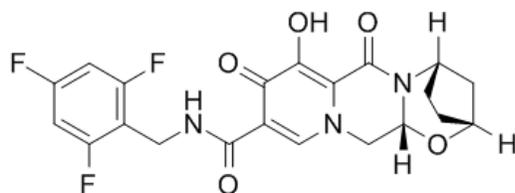


Figure 26: Structure du Bictégravir

(S. J. Smith, X. Z. Zhao, T. R. Burke and S. H. Huges, *Retrovirology*, vol. 15, no. 37, 2018, doi: 10.1186/s12977-018-0420-7)

Autre que son profil de résistance, il partage d'autres caractéristiques avec le DTG, telles qu'une longue demi-vie, une bonne tolérance et des interactions médicamenteuses minimales. Les études sur ce composé sont encore en cours notamment au niveau de ses effets cliniques indésirables ^[133] ^[134] ^[135].

Les résultats d'interaction des trois inhibiteurs présentés ci-dessus s'avèrent être d'un grand intérêt vu que les résidus impliqués dans les interactions de l'intégrase et de l'ADN avec les DKAs, sont les mêmes engagés dans les contacts ADN-Intégrase. Ceci souligne l'importance de concevoir des inhibiteurs dérivant des préexistants ayant un mode d'action renforcé à l'interface ADN-IN.

Références bibliographiques

- [1] M. S. Gottlieb *et al.*, “Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency,” *N. Engl. J. Med.*, vol. 305, no. 24, pp. 1425–1431, Dec. 1981, doi: 10.1056/NEJM198112103052401.
- [2] H. Masur *et al.*, “An outbreak of community-acquired Pneumocystis carinii pneumonia: initial manifestation of cellular immune dysfunction,” *N. Engl. J. Med.*, vol. 305, no. 24, pp. 1431–1438, Dec. 1981, doi: 10.1056/NEJM198112103052402.
- [3] “Fiche d’information 2020 — Dernières statistiques sur l’état de l’épidémie de sida.” <https://www.unaids.org/fr/resources/fact-sheet> (accessed Mar. 03, 2021).
- [4] J. Coffin *et al.*, “What to call the AIDS virus?,” *Nature*, vol. 321, no. 6065, p. 10, May 1986, doi: 10.1038/321010a0.
- [5] Z. J. da Silva *et al.*, “Changes in prevalence and incidence of HIV-1, HIV-2 and dual infections in urban areas of Bissau, Guinea-Bissau: is HIV-2 disappearing?,” *AIDS Lond. Engl.*, vol. 22, no. 10, pp. 1195–1202, Jun. 2008, doi: 10.1097/QAD.0b013e328300a33d.
- [6] NIH, “HIV/AIDS: The Basics”, AIDS INFO, September 24, 2020.
- [7] J. M. Cuevas, R. Geller, R. Garijo, J. López-Aldeguer, and R. Sanjuán, “Extremely High Mutation Rate of HIV-1 In Vivo,” *PLoS Biol.*, vol. 13, no. 9, Sep. 2015, doi: 10.1371/journal.pbio.1002251.
- [8] J. M. Coffin, S. H. Hughes, and H. E. Varmus, Eds., *Retroviruses*. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press, 1997.
- [9] H. R. Gelderblom, M. Özel, E. H. S. Hausmann, T. Winkel, G. Pauli, and M. A. Koch, “Fine structure of human immunodeficiency virus (HIV), immunolocalization of structural proteins and virus-cell relation,” *Micron Microsc. Acta*, vol. 19, no. 1, pp. 41–60, Jan. 1988, doi: 10.1016/0739-6260(88)90039-1.
- [10] M. W. Pandori, N. J. Fitch, H. M. Craig, D. D. Richman, C. A. Spina, and J. C. Guatelli, “Producer-cell modification of human immunodeficiency virus type 1: Nef is a virion protein.,” *J. Virol.*, vol. 70, no. 7, pp. 4283–4290, Jul. 1996.
- [11] Y. L. Lu, P. Spearman, and L. Ratner, “Human immunodeficiency virus type 1 viral protein R localization in infected cells and virions,” *J. Virol.*, vol. 67, no. 11, pp. 6542–6550, Nov. 1993, doi: 10.1128/JVI.67.11.6542-6550.1993.
- [12] M. Lu, S. C. Blacklow, and P. S. Kim, “A trimeric structural domain of the HIV-1 transmembrane glycoprotein,” *Nat. Struct. Biol.*, vol. 2, no. 12, Art. no. 12, Dec. 1995, doi: 10.1038/nsb1295-1075.
- [13] D. E. Ott *et al.*, “Cytoskeletal proteins inside human immunodeficiency virus type 1 virions.,” *J. Virol.*, vol. 70, no. 11, pp. 7734–7743, Nov. 1996.
- [14] E. O. Freed, “HIV-1 replication,” *Somat. Cell Mol. Genet.*, vol. 26, no. 1–6, pp. 13–33, Nov. 2001, doi: 10.1023/a:1021070512287.
- [15] W. C. Greene and B. M. Peterlin, “Charting HIV’s remarkable voyage through the cell: Basic science as a passport to future therapy,” *Nat. Med.*, vol. 8, no. 7, pp. 673–680, Jul. 2002, doi: 10.1038/nm0702-673.
- [16] L. A. Pereira, K. Bentley, A. Peeters, M. J. Churchill, and N. J. Deacon, “A compilation of cellular transcription factor interactions with the HIV-1 LTR promoter,” *Nucleic Acids Res.*, vol. 28, no. 3, pp. 663–668, Feb. 2000, doi: 10.1093/nar/28.3.663.
- [17] J. Garrido, N. Morellet, Y. Mély, and H. Rocquigny, “Properties, Functions, and Drug Targeting of the Multifunctional Nucleocapsid Protein of the Human Immunodeficiency Virus,” *Adv. Pharmacol. San Diego Calif*, vol. 55, pp. 299–346, Feb. 2007, doi: 10.1016/S1054-3589(07)55009-X.

- [18] W. G. Robey *et al.*, "Characterization of envelope and core structural gene products of HTLV-III with sera from AIDS patients," *Science*, vol. 228, no. 4699, pp. 593–595, May 1985, doi: 10.1126/science.2984774.
- [19] E. O. Freed, "HIV-1 assembly, release and maturation," *Nat. Rev. Microbiol.*, vol. 13, no. 8, pp. 484–496, Aug. 2015, doi: 10.1038/nrmicro3490.
- [20] J. S. McDougal, M. S. Kennedy, J. M. Sligh, S. P. Cort, A. Mawle, and J. K. Nicholson, "Binding of HTLV-III/LAV to T4+ T cells by a complex of the 110K viral protein and the T4 molecule," *Science*, vol. 231, no. 4736, pp. 382–385, Jan. 1986, doi: 10.1126/science.3001934.
- [21] P. Prabakaran, A. S. Dimitrov, T. R. Fouts, and D. S. Dimitrov, "Structure and Function of the HIV Envelope Glycoprotein as Entry Mediator, Vaccine Immunogen, and Target for Inhibitors," in *Advances in Pharmacology*, vol. 55, Academic Press, 2007, pp. 33–97.
- [22] D. S. Dimitrov, "Fusin – a place for HIV–1 and T4 cells to meet," *Nat. Med.*, vol. 2, no. 6, Art. no. 6, Jun. 1996, doi: 10.1038/nm0696-640.
- [23] J. Strizki, "Targeting HIV attachment and entry for therapy," *Adv. Pharmacol. San Diego Calif*, vol. 56, pp. 93–120, 2008, doi: 10.1016/S1054-3589(07)56004-7.
- [24] F. Turlure, E. Devroe, P. A. Silver, and A. Engelman, "Human cell proteins and human immunodeficiency virus DNA integration," *Front. Biosci. J. Virtual Libr.*, vol. 9, pp. 3187–3208, Sep. 2004, doi: 10.2741/1472.
- [25] V. D'Souza and M. F. Summers, "How retroviruses select their genomes," *Nat. Rev. Microbiol.*, vol. 3, no. 8, pp. 643–655, Aug. 2005, doi: 10.1038/nrmicro1210.
- [26] C. S. Adamson and E. O. Freed, "Human Immunodeficiency Virus Type 1 Assembly, Release, and Maturation," in *Advances in Pharmacology*, vol. 55, Academic Press, 2007, pp. 347–387.
- [27] D. D. Richman, "HIV chemotherapy," *Nature*, vol. 410, no. 6831, pp. 995–1001, Apr. 2001, doi: 10.1038/35073673.
- [28] J. A. Bartlett and J. F. Shao, "Successes, challenges, and limitations of current antiretroviral therapy in low-income and middle-income countries," *Lancet Infect. Dis.*, vol. 9, no. 10, pp. 637–649, Oct. 2009, doi: 10.1016/S1473-3099(09)70227-0.
- [29] A. Mtambo *et al.*, "Treatment limitations imposed by antiretroviral drug resistance mutations: implication for choices of first line regimens in resource-limited settings," *HIV Med.*, vol. 13, no. 3, pp. 141–147, Mar. 2012, doi: 10.1111/j.1468-1293.2011.00950.x.
- [30] J. D. Roberts, K. Bebenek, and T. A. Kunkel, "The accuracy of reverse transcriptase from HIV-1," *Science*, vol. 242, no. 4882, pp. 1171–1173, Nov. 1988, doi: 10.1126/science.2460925.
- [31] D. Esposito and R. Craigie, "HIV Integrase Structure and Function," in *Advances in Virus Research*, vol. 52, K. Rlaramorosch, F. A. Murphy, and A. J. Shaw, Eds. Academic Press, 1999, pp. 319–333
- [32] T. K. Chiu and D. R. Davies, "Structure and function of HIV-1 integrase," *Curr. Top. Med. Chem.*, vol. 4, no. 9, pp. 965–977, 2004, doi: 10.2174/1568026043388547.
- [33] M. Cai, R. Zheng, M. Caffrey, R. Craigie, G. M. Clore, and A. M. Gronenborn, "Solution structure of the N-terminal zinc binding domain of HIV-1 integrase," *Nat. Struct. Biol.*, vol. 4, no. 7, pp. 567–577, Jul. 1997, doi: 10.1038/nsb0797-567.
- [34] P. J. Lodi *et al.*, "Solution structure of the DNA binding domain of HIV-1 integrase," *Biochemistry*, vol. 34, no. 31, pp. 9826–9833, Aug. 1995, doi: 10.1021/bi00031a002.
- [35] Y. Goldgur, F. Dyda, A. B. Hickman, T. M. Jenkins, R. Craigie, and D. R. Davies, "Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 16, pp. 9150–9154, Aug. 1998, doi: 10.1073/pnas.95.16.9150.
- [36] J. Y. Wang, H. Ling, W. Yang, and R. Craigie, "Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein," *EMBO J.*, vol. 20, no. 24, pp. 7333–7343, Dec. 2001, doi: 10.1093/emboj/20.24.7333.

- [37] J. C. Chen *et al.*, "Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 15, pp. 8233–8238, Jul. 2000, doi: 10.1073/pnas.150220297.
- [38] D. O. Passos *et al.*, "CryoEM Structures and Atomic Model of the HIV-1 Strand Transfer Complex Intasome," *Science*, vol. 355, no. 6320, pp. 89–92, Jan. 2017, doi: 10.1126/science.aah5163.
- [39] S. Hare, S. S. Gupta, E. Valkov, A. Engelman, and P. Cherepanov, "Retroviral intasome assembly and inhibition of DNA strand transfer," *Nature*, vol. 464, no. 7286, pp. 232–236, Mar. 2010, doi: 10.1038/nature08784.
- [40] S. Hare, A. M. Vos, R. F. Clayton, J. W. Thuring, M. D. Cummings, and P. Cherepanov, "Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 46, pp. 20057–20062, Nov. 2010, doi: 10.1073/pnas.1010246107.
- [41] R. Zheng, T. M. Jenkins, and R. Craigie, "Zinc folds the N-terminal domain of HIV-1 integrase, promotes multimerization, and enhances catalytic activity," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 93, no. 24, pp. 13659–13664, Nov. 1996.
- [42] C. J. Burke *et al.*, "Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase," *J. Biol. Chem.*, vol. 267, no. 14, pp. 9639–9644, May 1992.
- [43] S. P. Lee, J. Xiao, J. R. Knutson, M. S. Lewis, and M. K. Han, "Zn²⁺ promotes the self-association of human immunodeficiency virus type-1 integrase in vitro," *Biochemistry*, vol. 36, no. 1, pp. 173–180, Jan. 1997, doi: 10.1021/bi961849o.
- [44] G. Eilers *et al.*, "Influence of the amino-terminal sequence on the structure and function of HIV integrase," *Retrovirology*, vol. 17, no. 1, p. 28, Aug. 2020, doi: 10.1186/s12977-020-00537-x.
- [45] V. Ellison, J. Gerton, K. A. Vincent, and P. O. Brown, "An Essential Interaction between Distinct Domains of HIV-1 Integrase Mediates Assembly of the Active Multimer (*)," *J. Biol. Chem.*, vol. 270, no. 7, pp. 3320–3326, Feb. 1995, doi: 10.1074/jbc.270.7.3320.
- [46] F. M. I. van den Ent, A. Vos, and R. H. A. Plasterk, "Dissecting the Role of the N-Terminal Domain of Human Immunodeficiency Virus Integrase by trans-Complementation Analysis," *J. Virol.*, vol. 73, no. 4, pp. 3176–3183, Apr. 1999.
- [47] A. Engelman and R. Craigie, "Identification of conserved amino acid residues critical for human immunodeficiency virus type 1 integrase function in vitro," *J. Virol.*, vol. 66, no. 11, pp. 6361–6369, Nov. 1992, doi: 10.1128/JVI.66.11.6361-6369.1992.
- [48] J. Kulkosky, K. S. Jones, R. A. Katz, J. P. Mack, and A. M. Skalka, "Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases," *Mol. Cell. Biol.*, vol. 12, no. 5, pp. 2331–2338, May 1992, doi: 10.1128/mcb.12.5.2331.
- [49] A. D. Leavitt, L. Shiue, and H. E. Varmus, "Site-directed mutagenesis of HIV-1 integrase demonstrates differential effects on integrase functions in vitro," *J. Biol. Chem.*, vol. 268, no. 3, pp. 2113–2119, Jan. 1993.
- [50] T. L. Diamond and F. D. Bushman, "Role of metal ions in catalysis by HIV integrase analyzed using a quantitative PCR disintegration assay," *Nucleic Acids Res.*, vol. 34, no. 21, pp. 6116–6125, Nov. 2006, doi: 10.1093/nar/gkl862.
- [51] R. Craigie, "HIV Integrase, a Brief Overview from Chemistry to Therapeutics*," *J. Biol. Chem.*, vol. 276, no. 26, pp. 23213–23216, Jun. 2001, doi: 10.1074/jbc.R100027200.
- [52] R. G. Maroun *et al.*, "Peptide inhibitors of HIV-1 integrase dissociate the enzyme oligomers," *Biochemistry*, vol. 40, no. 46, pp. 13840–13848, Nov. 2001, doi: 10.1021/bi011328n.
- [53] A. Engelman, A. B. Hickman, and R. Craigie, "The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding," *J. Virol.*, vol. 68, no. 9, pp. 5911–5917, Sep. 1994, doi: 10.1128/JVI.68.9.5911-5917.1994.

- [54] A. P. Eijkelenboom, R. A. Lutzke, R. Boelens, R. H. Plasterk, R. Kaptein, and K. Hård, "The DNA-binding domain of HIV-1 integrase has an SH3-like fold," *Nat. Struct. Biol.*, vol. 2, no. 9, pp. 807–810, Sep. 1995, doi: 10.1038/nsb0995-807.
- [55] Y. Pommier, A. A. Johnson, and C. Marchand, "Integrase inhibitors to treat HIV/AIDS," *Nat. Rev. Drug Discov.*, vol. 4, no. 3, pp. 236–248, Mar. 2005, doi: 10.1038/nrd1660.
- [56] S. Hare, G. N. Maertens, and P. Cherepanov, "3'-processing and strand transfer catalysed by retroviral integrase in crystallo," *EMBO J.*, vol. 31, no. 13, pp. 3020–3028, Jun. 2012, doi: 10.1038/emboj.2012.118.
- [57] B. Bowerman, P. O. Brown, J. M. Bishop, and H. E. Varmus, "A nucleoprotein complex mediates the integration of retroviral DNA," *Genes Dev.*, vol. 3, no. 4, pp. 469–478, Apr. 1989, doi: 10.1101/gad.3.4.469.
- [58] M. J. Roth, P. L. Schwartzberg, and S. P. Goff, "Structure of the termini of DNA intermediates in the integration of retroviral DNA: dependence on IN function and terminal DNA sequence," *Cell*, vol. 58, no. 1, pp. 47–54, Jul. 1989, doi: 10.1016/0092-8674(89)90401-7.
- [59] P. O. Brown, B. Bowerman, H. E. Varmus, and J. M. Bishop, "Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 86, no. 8, pp. 2525–2529, Apr. 1989, doi: 10.1073/pnas.86.8.2525.
- [60] C. Vink, D. C. van Gent, Y. Elgersma, and R. H. Plasterk, "Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage.," *J. Virol.*, vol. 65, no. 9, pp. 4636–4644, Sep. 1991.
- [61] F. Bushman *et al.*, "Genome-wide analysis of retroviral DNA integration," *Nat. Rev. Microbiol.*, vol. 3, no. 11, pp. 848–858, Nov. 2005, doi: 10.1038/nrmicro1263.
- [62] Z. Hobaika, L. Zargarian, Y. Boulard, R. Maroun, O. Mauffret, and S. Fermandjian, "Specificity of LTR DNA recognition by a peptide mimicking the HIV-1 integrase 4 helix," *Nucleic Acids Res.*, vol. 37, pp. 7691–7700, Oct. 2009, doi: 10.1093/nar/gkp824.
- [63] L. Zargarian *et al.*, "Strategy to discriminate between high and low affinity bindings of human immunodeficiency virus, type 1 integrase to viral DNA," *J. Biol. Chem.*, vol. 278, no. 22, pp. 19966–19973, May 2003, doi: 10.1074/jbc.M211711200.
- [64] A. A. Johnson *et al.*, "Integration Requires a Specific Interaction of the Donor DNA Terminal 5'-Cytosine with Glutamine 148 of the HIV-1 Integrase Flexible Loop *," *J. Biol. Chem.*, vol. 281, no. 1, pp. 461–467, Jan. 2006, doi: 10.1074/jbc.M511348200.
- [65] N. K. Raghavendra, N. Shkriabai, R. L. Graham, S. Hess, M. Kvaratskhelia, and L. Wu, "Identification of host proteins associated with HIV-1 preintegration complexes isolated from infected CD4⁺ cells," *Retrovirology*, vol. 7, no. 1, p. 66, Aug. 2010, doi: 10.1186/1742-4690-7-66.
- [66] C. Vink, M. Groenink, Y. Elgersma, R. A. Fouchier, M. Tersmette, and R. H. Plasterk, "Analysis of the junctions between human immunodeficiency virus type 1 proviral DNA and human DNA.," *J. Virol.*, vol. 64, no. 11, pp. 5626–5627, Nov. 1990.
- [67] R. Daniel *et al.*, "Evidence that stable retroviral transduction and cell survival following DNA integration depend on components of the nonhomologous end joining repair pathway," *J. Virol.*, vol. 78, no. 16, pp. 8573–8581, Aug. 2004, doi: 10.1128/JVI.78.16.8573-8581.2004.
- [68] A. M. Skalka and R. A. Katz, "Retroviral DNA integration and the DNA damage response," *Cell Death Differ.*, vol. 12, no. 1, Art. no. 1, Aug. 2005, doi: 10.1038/sj.cdd.4401573.
- [69] M. Katzman and R. A. Katz, "Substrate recognition by retroviral integrases," *Adv. Virus Res.*, vol. 52, pp. 371–395, 1999, doi: 10.1016/s0065-3527(08)60307-3.
- [70] T. M. Jenkins, D. Esposito, A. Engelman, and R. Craigie, "Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking," *EMBO J.*, vol. 16, no. 22, pp. 6849–6859, Nov. 1997, doi: 10.1093/emboj/16.22.6849.

- [71] D. J. C. A. W. It, H. Rw, and L. J, "Defining the DNA substrate binding sites on HIV-1 integrase.," *J. Mol. Biol.*, vol. 385, no. 2, pp. 568–579, Nov. 2008, doi: 10.1016/j.jmb.2008.10.083.
- [72] J. J. Kessl, C. J. McKee, J. O. Eidahl, N. Shkriabai, A. Katz, and M. Kvaratskhelia, "HIV-1 Integrase-DNA Recognition Mechanisms," *Viruses*, vol. 1, no. 3, pp. 713–736, Nov. 2009, doi: 10.3390/v1030713.
- [73] J. L. Gerton, S. Ohgi, M. Olsen, J. DeRisi, and P. O. Brown, "Effects of mutations in residues near the active site of human immunodeficiency virus type 1 integrase on specific enzyme-substrate interactions," *J. Virol.*, vol. 72, no. 6, pp. 5046–5055, Jun. 1998, doi: 10.1128/JVI.72.6.5046-5055.1998.
- [74] C. Calmels *et al.*, "Biochemical and random mutagenesis analysis of the region carrying the catalytic E152 amino acid of HIV-1 integrase," *Nucleic Acids Res.*, vol. 32, no. 4, pp. 1527–1538, 2004, doi: 10.1093/nar/gkh298.
- [75] R. Lu, A. Limón, E. Devroe, P. A. Silver, P. Cherepanov, and A. Engelman, "Class II Integrase Mutants with Changes in Putative Nuclear Localization Signals Are Primarily Blocked at a Postnuclear Entry Step of Human Immunodeficiency Virus Type 1 Replication," *J. Virol.*, vol. 78, no. 23, pp. 12735–12746, Dec. 2004, doi: 10.1128/JVI.78.23.12735-12746.2004.
- [76] R. Lu, A. Limón, H. Z. Ghory, and A. Engelman, "Genetic analyses of DNA-binding mutants in the catalytic core domain of human immunodeficiency virus type 1 integrase," *J. Virol.*, vol. 79, no. 4, pp. 2493–2505, Feb. 2005, doi: 10.1128/JVI.79.4.2493-2505.2005.
- [77] J. S. Eggleton and S. Nagalli, "Highly Active Antiretroviral Therapy (HAART)," in *StatPearls*, Treasure Island (FL): StatPearls Publishing, 2021.
- [78] K. Vermeire *et al.*, "CADA, a potential anti-HIV microbicide that specifically targets the cellular CD4 receptor," *Curr. HIV Res.*, vol. 6, no. 3, pp. 246–256, May 2008, doi: 10.2174/157016208784324958.
- [79] K. Vermeire *et al.*, "CADA, a novel CD4-targeted HIV inhibitor, is synergistic with various anti-HIV drugs in vitro," *AIDS Lond. Engl.*, vol. 18, no. 16, pp. 2115–2125, Nov. 2004, doi: 10.1097/00002030-200411050-00003.
- [80] F. Grande *et al.*, "CCR5/CXCR4 Dual Antagonism for the Improvement of HIV Infection Therapy," *Molecules*, vol. 24, no. 3, Feb. 2019, doi: 10.3390/molecules24030550.
- [81] R. M. Mosi *et al.*, "The molecular pharmacology of AMD11070: an orally bioavailable CXCR4 HIV entry inhibitor," *Biochem. Pharmacol.*, vol. 83, no. 4, pp. 472–479, Feb. 2012, doi: 10.1016/j.bcp.2011.11.020.
- [82] G. G. Xu, J. Guo, and Y. Wu, "Chemokine Receptor CCR5 Antagonist Maraviroc: Medicinal Chemistry and Clinical Applications," *Curr. Top. Med. Chem.*, vol. 14, no. 13, pp. 1504–1514, 2014.
- [83] J. Balzarini *et al.*, "Mannose-specific plant lectins from the Amaryllidaceae family qualify as efficient microbicides for prevention of human immunodeficiency virus infection," *Antimicrob. Agents Chemother.*, vol. 48, no. 10, pp. 3858–3870, Oct. 2004, doi: 10.1128/AAC.48.10.3858-3870.2004.
- [84] L. K. Tsou, C. Chen, G. E. Dutschman, Y. Cheng and A. D. Hamilton., "Blocking HIV-1 entry by a gp120 surface binding inhibitor," *Elsevier.*, vol. 22, pp. 3358–3361, Feb. 2012, doi: 10.1016/j.bmcl.2012.02.079.
- [85] M. Tiberi *et al.*, "2-Aminothiazolones as Anti-HIV Agents That Act as gp120-CD4 Inhibitors," *Antimicrob. Agents Chemother.*, vol. 58, no. 6, pp. 3043–3052, Jun. 2014, doi: 10.1128/AAC.02739-13.
- [86] J. P. Lalezari *et al.*, "Enfuvirtide, an HIV-1 fusion inhibitor, for drug-resistant HIV infection in North and South America," *N. Engl. J. Med.*, vol. 348, no. 22, pp. 2175–2185, May 2003, doi: 10.1056/NEJMoa035026.
- [87] D. Eggink, B. Berkhout, and R. W. Sanders, "Inhibition of HIV-1 by fusion inhibitors," *Curr. Pharm. Des.*, vol. 16, no. 33, pp. 3716–3728, 2010, doi: 10.2174/138161210794079218.
- [88] X. Ding *et al.*, "Enfuvirtide (T20)-Based Lipopeptide Is a Potent HIV-1 Cell Fusion Inhibitor: Implications for Viral Entry and Inhibition," *J. Virol.*, vol. 91, no. 18, Sep. 2017, doi: 10.1128/JVI.00831-17.

- [89] H. Chong *et al.*, "Design of Novel HIV-1/2 Fusion Inhibitors with High Therapeutic Efficacy in Rhesus Monkey Models," *J. Virol.*, vol. 92, no. 16, Aug. 2018, doi: 10.1128/JVI.00775-18.
- [90] T. Xiao *et al.*, "HIV-1 fusion inhibitors targeting the membrane-proximal external region of Env spikes," *Nat. Chem. Biol.*, vol. 16, no. 5, Art. no. 5, May 2020, doi: 10.1038/s41589-020-0496-y.
- [91] R. Yarchoan *et al.*, "ADMINISTRATION OF 3'-AZIDO-3'-DEOXYTHYMIDINE, AN INHIBITOR OF HTLV-III/LAV REPLICATION, TO PATIENTS WITH AIDS OR AIDS-RELATED COMPLEX," *The Lancet*, vol. 327, no. 8481, pp. 575–580, Mar. 1986, doi: 10.1016/S0140-6736(86)92808-4.
- [92] E.-M. Murphy, H. R. Jimenez, and S. M. Smith, "Current clinical treatments of AIDS," *Adv. Pharmacol. San Diego Calif*, vol. 56, pp. 27–73, 2008, doi: 10.1016/S1054-3589(07)56002-3.
- [93] A. D. Holec, S. Mandal, P. K. Prathipati, and C. J. Destache, "Nucleotide Reverse Transcriptase Inhibitors: A Thorough Review, Present Status and Future Perspective as HIV Therapeutics," *Curr. HIV Res.*, vol. 15, no. 6, pp. 411–421, 2017, doi: 10.2174/1570162X15666171120110145.
- [94] E. D. Clercq, "Antiviral therapy for human immunodeficiency virus infections.," *Clin. Microbiol. Rev.*, vol. 8, no. 2, pp. 200–239, Apr. 1995, doi: 10.1128/CMR.8.2.200.
- [95] I. Usach, V. Melis, and J.-E. Peris, "Non-nucleoside reverse transcriptase inhibitors: a review on pharmacokinetics, pharmacodynamics, safety and tolerability," *J. Int. AIDS Soc.*, vol. 16, no. 1, p. 18567, 2013, doi: <https://doi.org/10.7448/IAS.16.1.18567>.
- [96] V. Namasivayam *et al.*, "The Journey of HIV-1 Non-Nucleoside Reverse Transcriptase Inhibitors (NNRTIs) from Lab to Clinic," *J. Med. Chem.*, Dec. 2018, doi: 10.1021/acs.jmedchem.8b00843.
- [97] C. K. McDonald and D. R. Kuritzkes, "Human immunodeficiency virus type 1 protease inhibitors," *Arch. Intern. Med.*, vol. 157, no. 9, pp. 951–959, May 1997.
- [98] Z. Lv, Y. Chu, and Y. Wang, "HIV protease inhibitors: a review of molecular selectivity and toxicity," *HIV/AIDS Auckl. NZ*, vol. 7, pp. 95–104, Apr. 2015, doi: 10.2147/HIV.S79956.
- [99] R. Craigie, K. Mizuuchi, F. D. Bushman, and A. Engelman, "A rapid in vitro assay for HIV DNA integration," *Nucleic Acids Res.*, vol. 19, no. 10, pp. 2729–2734, May 1991.
- [100] D. J. Hazuda, J. C. Hastings, A. L. Wolfe, and E. A. Emini, "A novel assay for the DNA strand-transfer reaction of HIV-1 integrase.," *Nucleic Acids Res.*, vol. 22, no. 6, pp. 1121–1122, Mar. 1994.
- [101] Y. Pommier and C. Marchand, "Interfacial inhibitors of protein-nucleic acid interactions," *Curr. Med. Chem. Anti-Cancer Agents*, vol. 5, no. 4, pp. 421–429, Jul. 2005, doi: 10.2174/1568011054222337.
- [102] D. J. Hazuda *et al.*, "Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells," *Science*, vol. 287, no. 5453, pp. 646–650, Jan. 2000, doi: 10.1126/science.287.5453.646.
- [103] G. C. G. Pais *et al.*, "Structure activity of 3-aryl-1,3-diketo-containing compounds as HIV-1 integrase inhibitors," *J. Med. Chem.*, vol. 45, no. 15, pp. 3184–3194, Jul. 2002, doi: 10.1021/jm020037p.
- [104] J. A. Grobler *et al.*, "Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 10, pp. 6661–6666, May 2002, doi: 10.1073/pnas.092056199.
- [105] X. Li and R. Vince, "Conformationally restrained carbazolone-containing alpha,gamma-diketo acids as inhibitors of HIV integrase," *Bioorg. Med. Chem.*, vol. 14, no. 9, pp. 2942–2955, May 2006, doi: 10.1016/j.bmc.2005.12.013.
- [106] C. Marchand *et al.*, "Structural Determinants for HIV-1 Integrase Inhibition by β -Diketo Acids*," *J. Biol. Chem.*, vol. 277, no. 15, pp. 12596–12603, Apr. 2002, doi: 10.1074/jbc.M110758200.
- [107] M. Sechi *et al.*, "From Ligand to Complexes: Inhibition of Human Immunodeficiency Virus Type 1 Integrase by β -Diketo Acid Metal Complexes," *J. Med. Chem.*, vol. 49, no. 14, pp. 4248–4260, Jul. 2006, doi: 10.1021/jm060193m.
- [108] V. Fikkert *et al.*, "Development of resistance against diketo derivatives of human immunodeficiency virus type 1 by progressive accumulation of integrase mutations," *J. Virol.*, vol. 77, no. 21, pp. 11459–11470, Nov. 2003, doi: 10.1128/jvi.77.21.11459-11470.2003.

- [109] R. Reinke, D. J. Lee, and W. E. Robinson, "Inhibition of human immunodeficiency virus type 1 isolates by the integrase inhibitor L-731,988, a diketo Acid," *Antimicrob. Agents Chemother.*, vol. 46, no. 10, pp. 3301–3303, Oct. 2002, doi: 10.1128/aac.46.10.3301-3303.2002.
- [110] I. B. Dicker *et al.*, "Biochemical Analysis of HIV-1 Integrase Variants Resistant to Strand Transfer Inhibitors," *J. Biol. Chem.*, vol. 283, no. 35, pp. 23599–23609, Aug. 2008, doi: 10.1074/jbc.M804213200.
- [111] J.-F. Mouscadet *et al.*, "HIV-1 IN alternative molecular recognition of DNA induced by raltegravir resistance mutations," *J. Mol. Recognit.*, vol. 22, no. 6, pp. 480–494, 2009, doi: 10.1002/jmr.970.
- [112] O. Delelis *et al.*, "Impact of Y143 HIV-1 Integrase Mutations on Resistance to Raltegravir In Vitro and In Vivo," *Antimicrob. Agents Chemother.*, vol. 54, no. 1, pp. 491–501, Jan. 2010, doi: 10.1128/AAC.01075-09
- [113] S. Hare *et al.*, "Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572)," *Mol. Pharmacol.*, vol. 80, no. 4, pp. 565–572, Oct. 2011, doi: 10.1124/mol.111.073189.
- [114] K. K. Pandey and D. P. Grandgenett, "HIV-1 Integrase Strand Transfer Inhibitors: Novel Insights into their Mechanism of Action," *Retrovirology Res. Treat.*, vol. 2, pp. 11–16, Nov. 2008, doi: 10.4137/rrt.s1081.
- [115] B. Grinsztejn *et al.*, "Safety and efficacy of the HIV-1 integrase inhibitor raltegravir (MK-0518) in treatment-experienced patients with multidrug-resistant virus: a phase II randomised controlled trial," *Lancet Lond. Engl.*, vol. 369, no. 9569, pp. 1261–1269, Apr. 2007, doi: 10.1016/S0140-6736(07)60597-2.
- [116] M. Anker and R. B. Corales, "Raltegravir (MK-0518): a novel integrase inhibitor for the treatment of HIV infection," *Expert Opin. Investig. Drugs*, vol. 17, no. 1, pp. 97–103, Jan. 2008, doi: 10.1517/13543784.17.1.97.
- [117] V. Summa *et al.*, "Discovery of Raltegravir, a Potent, Selective Orally Bioavailable HIV-Integrase Inhibitor for the Treatment of HIV-AIDS Infection," *J. Med. Chem.*, vol. 51, no. 18, pp. 5843–5855, Sep. 2008, doi: 10.1021/jm800245z.
- [118] D. Langley, H. Samanta, Z. Lin, M. Walker, M. Krystal, and I. Dicker, "The Terminal (Catalytic) Adenosine of the HIV LTR Controls the Kinetics of Binding and Dissociation of HIV Integrase Strand Transfer Inhibitors," *Biochemistry*, vol. 47, pp. 13481–8, Dec. 2008, doi: 10.1021/bi801372d.
- [119] K. Ramkumar and N. Neamati, "Raltegravir: The evidence of its therapeutic value in HIV-1 infection," *Core Evid.*, vol. 4, pp. 131–147, Jun. 2010, doi: 10.2147/ce.s6004.
- [120] H. W. Lampiris, "Elvitegravir: a once-daily, boosted, HIV-1 integrase inhibitor," *Expert Rev. Anti Infect. Ther.*, vol. 10, no. 1, pp. 13–20, Jan. 2012, doi: 10.1586/eri.11.157.
- [121] K. Shimura and E. N. Kodama, "Elvitegravir: A New HIV Integrase Inhibitor," *Antivir. Chem. Chemother.*, vol. 20, no. 2, pp. 79–85, Oct. 2009, doi: 10.3851/IMP1397.
- [122] D. da Silva *et al.*, "HIV-1 resistance patterns to integrase inhibitors in antiretroviral-experienced patients with virological failure on raltegravir-containing regimens," *J. Antimicrob. Chemother.*, vol. 65, no. 6, pp. 1262–1269, Jun. 2010, doi: 10.1093/jac/dkq099.
- [123] R. C. Rathbun, S. M. Lockhart, M. M. Miller, and M. D. Liedtke, "Dolutegravir, a second-generation integrase inhibitor for the treatment of HIV-1 infection," *Ann. Pharmacother.*, vol. 48, no. 3, pp. 395–403, Mar. 2014, doi: 10.1177/1060028013513558.
- [124] K. E. Hightower *et al.*, "Dolutegravir (S/GSK1349572) Exhibits Significantly Slower Dissociation than Raltegravir and Elvitegravir from Wild-Type and Integrase Inhibitor-Resistant HIV-1 Integrase-DNA

- Complexes," *Antimicrob. Agents Chemother.*, vol. 55, no. 10, pp. 4552–4559, Oct. 2011, doi: 10.1128/AAC.00157-11.
- [125] D. A. Osterholzer and M. Goldman, "Dolutegravir: A Next-Generation Integrase Inhibitor for Treatment of HIV Infection," *Clin. Infect. Dis.*, vol. 59, no. 2, pp. 265–271, Jul. 2014, doi: 10.1093/cid/ciu221.
- [126] S. Reigadas *et al.*, "Structure-Analysis of the HIV-1 Integrase Y143C/R Raltegravir Resistance Mutation in Association with the Secondary Mutation T97A," *Antimicrob. Agents Chemother.*, vol. 55, pp. 3187–94, Jul. 2011, doi: 10.1128/AAC.00071-11.
- [127] J. Marinello, C. Marchand, B. T. Mott, A. Bain, C. J. Thomas, and Y. Pommier, "Comparison of raltegravir and elvitegravir on HIV-1 integrase catalytic reactions and on a series of drug-resistant integrase mutants," *Biochemistry*, vol. 47, no. 36, pp. 9345–9354, Sep. 2008, doi: 10.1021/bi800791q.
- [128] M. Métifiot *et al.*, "Elvitegravir overcomes resistance to raltegravir induced by integrase mutation Y143," *AIDS*, vol. 25, no. 9, pp. 1175–1178, Jun. 2011, doi: 10.1097/QAD.0b013e3283473599.
- [129] M. Métifiot, C. Marchand, K. Maddali, and Y. Pommier, "Resistance to Integrase Inhibitors," *Viruses*, vol. 2, no. 7, pp. 1347–1366, Jun. 2010, doi: 10.3390/v2071347.
- [130] A. Low *et al.*, "Natural Polymorphisms of Human Immunodeficiency Virus Type 1 Integrase and Inherent Susceptibilities to a Panel of Integrase Inhibitors," *Antimicrob. Agents Chemother.*, vol. 53, pp. 4275–82, Sep. 2009, doi: 10.1128/AAC.00397-09.
- [131] M. A. Wainberg and Y.-S. Han, "HIV-1 resistance to dolutegravir: update and new insights," *J. Virus Erad.*, vol. 1, no. 1, pp. 13–16, Jan. 2015, doi: 10.1016/S2055-6640(20)31150-X.
- [132] A. M. Geretti, D. Armenia, and F. Ceccherini-Silberstein, "Emerging patterns and implications of HIV-1 integrase inhibitor resistance," *Curr. Opin. Infect. Dis.*, vol. 25, no. 6, pp. 677–686, Dec. 2012, doi: 10.1097/QCO.0b013e32835a1de7.
- [133] J. Zeuli, S. Rizza, R. Bhatia, and Z. Temesgen, "Bictegravir, a novel integrase inhibitor for the treatment of HIV infection," *Drugs Today Barc. Spain 1998*, vol. 55, no. 11, pp. 669–682, Nov. 2019, doi: 10.1358/dot.2019.55.11.3068796.
- [134] M. Tsiang *et al.*, "Antiviral Activity of Bictegravir (GS-9883), a Novel Potent HIV-1 Integrase Strand Transfer Inhibitor with an Improved Resistance Profile," *Antimicrob. Agents Chemother.*, vol. 60, no. 12, pp. 7086–7097, Nov. 2016, doi: 10.1128/AAC.01474-16.
- [135] B. Max, "Update on HIV integrase inhibitors for the treatment of HIV-1 infection," *Future Virol.*, vol. 14, no. 10, pp. 693–709, Oct. 2019, doi: 10.2217/fvl-2019-0077.

Conclusions du chapitre II

Le SIDA, une pandémie majeure causée par le virus VIH-1, continue à infecter des millions de personnes annuellement. L'un des composants majeurs de ce virus est l'intégrase virale (IN), enzyme responsable de l'intégration de l'ADN viral (ADNv) au sein de l'ADN cellulaire via deux étapes ; la maturation en 3' de l'ADNv suivie du transfert de brin. Vu son rôle primordial dans le cycle répliatif viral et en absence d'homologue cellulaire de l'intégrase, cette dernière est devenue la cible essentielle des inhibiteurs anti-SIDA. Parmi eux, les inhibiteurs de transfert de brin (INSTIs) présentent l'activité anti-intégrase la plus spécifique, agissant à l'interface ADN-IN, après la maturation de l'ADN viral, de manière à inhiber son transfert dans la cellule cible. Le Dolutégravir (DTG) est l'INSTI le plus efficace à inhiber l'IN tout en échappant à la majorité des mutations de résistance qu'elle développe. Cependant, l'apparition de résistances et de toxicité persiste, ce qui rend indispensable la recherche et l'identification de nouveaux médicaments optimisés capables de surmonter l'obstacle des mutations de résistance.

Néanmoins, le processus de développement de nouveaux médicaments présente des limitations (durée longue, prix élevé). Par conséquent, les approches *in silico* s'imposent telles que la modélisation des interactions médicament candidat - cible. En effet, les méthodes de conception guidée par la structure permettent une étude rigoureuse des interactions du ligand potentiel avec sa cible permettant un développement rapide et précis de nouveaux inhibiteurs.

Approche de travail développée dans cette thèse : rationalisation par modélisation moléculaire

L'ultime but de notre projet est le développement d'une nouvelle série d'inhibiteurs anti-intégrase plus puissants et plus spécifiques que le Dolutégravir.

En fait, bien que les INSTIs soient des médicaments antirétroviraux révolutionnaires et puissants, leur efficacité clinique reste compromise par l'apparition de mutations de résistance au niveau de l'intégrase virale chez les patients traités. Ce qui affecte négativement l'efficacité des thérapies anti-VIH-1.

Puisque le DTG, par son mode d'action à l'interface ADN-IN, interagit avec les nucléotides terminaux de l'ADNv ainsi qu'avec les résidus catalytiques de l'intégrase, il paraît nécessaire, afin de surmonter le problème de la résistance de l'intégrase face aux INSTIS, d'affaiblir les interactions de cette dernière avec le DTG en renforçant son affinité avec l'ADN viral, surtout que la fréquence de mutation de l'ADNv est très faible par rapport à celle de l'enzyme virale. Ce renforcement commence par l'identification des groupes fonctionnels du DTG impliqués dans les interactions intimes avec l'ADN du VIH-1, et qui pourraient potentiellement être modifiés afin d'optimiser l'activité thérapeutique du DTG.

Notre équipe a réussi à démontrer antérieurement que la meilleure efficacité ainsi que la moindre sensibilité aux mutations de résistance de DTG par rapport à RAL et EVG, sont dues à ses interactions plus fortes avec l'ADNv ^[1]. De plus, et comme en l'absence de structure cristallographique complète de l'intasome du VIH-1 (ADNv + IN) complexé au DTG, nous nous sommes basés dans nos démarches d'optimisation sur la structure cristalline de l'intasome du PFV avec le DTG, en présence de Mg²⁺, le cristal établi par Hare et ses collaborateurs en 2011 (code PDB : 3S3M) ^[2] identifie le noyau benzylique fluoré du DTG comme le groupement le plus impliqué dans les contacts directs avec les extrémités réactives de l'ADNv. Nos travaux précédents ont confirmé de même que l'halobenzène est le site d'interaction principal avec l'ADNv ^{[3] [4]}. En effet, il occupe la position de la base terminale A₁₇, éjectant

son fragment réactif 3'-OH loin du site actif, déstabilisant ainsi l'étape de transfert de brin. Il interagit aussi avec l'avant-dernière paire de base C₁₆-G₄, particulièrement par des empilements de type π avec la cytosine et des interactions électrostatiques via son fluor avec la guanine. Par conséquent, dans la première partie de notre travail, nous avons effectué des modifications au niveau du noyau d'halobenzène du DTG (par des substitutions d'halogènes et de groupements donneurs et attracteurs d'électrons). Nous avons cherché à renforcer les liaisons de l'halobenzène avec les bases virales afin de compenser de potentielles interactions de ces dérivés avec l'IN, ce qui diminuerait leur sensibilité aux mutations de résistance développées au niveau de la protéine virale. Nous avons ensuite évalué nos nouvelles molécules conçues dans le cadre d'un système simplifié comprenant uniquement les entités impliquées directement, à savoir le nouveau noyau halogéné avec le noyau de C₁₆ et le noyau de G₄. L'évaluation de ce mini-complexe ternaire s'est faite par des calculs énergétiques employant des méthodes de chimie quantique réalisées dans le cadre de la fonctionnelle de la densité (DFT) ainsi que des méthodes de mécanique moléculaire employant le champ de force polarisable SIBFA développé au sein de notre laboratoire. Nous avons de même déterminé les contours de potentiel électrostatique moléculaire (MEP) pour chacun des noyaux conçus afin d'évaluer qualitativement l'impact de nos substitutions sur la distribution de la population électronique des noyaux et donc sur leurs interactions intermoléculaires avec les bases terminales virales.

Après la conception des nouveaux dérivés et leur étude énergétique au sein des petits complexes, la deuxième partie de notre travail a consisté en une extrapolation de l'étude à une échelle plus grande. Nous avons suivi le comportement des nouveaux inhibiteurs pris entièrement au sein de l'intasome

(ADN viral entier + Intégrase entière + Mg^{2+}) par des simulations de dynamique moléculaire polarisable en solvant explicite. Nous avons de même exploré les interactions de nos nouvelles molécules avec l'ADNv seul afin d'analyser l'effet de nos substitutions sur leur spécificité d'action au niveau des extrémités de l'ADN viral. En parallèle, des simulations du DTG au sein de l'intasome viral (structure de Hare et al) ainsi que des simulations de DTG avec chaque partenaire seul (DTG-ADNv et DTG-IN) ont été effectuées pour comparer premièrement l'interaction de DTG avec chaque partenaire, et mettre en relief sa meilleure interaction avec l'ADNv, expliquant ainsi sa moindre sensibilité aux mutations de résistance de l'IN. De plus, nous visons à comparer le comportement de DTG à celui des nouveaux inhibiteurs, afin de mieux comprendre la relation structure-interactions entre l'halobenzène et les bases C_{16} et G_4 . Nous avons établi enfin la dynamique de l'intasome du PFV en absence du DTG, structure fournie aussi par l'équipe de Hare en 2010 (code PDB : 3L2R) ^[5] afin de compléter l'étude du mécanisme d'inhibition de l'intégration par le DTG et potentiellement les nouveaux inhibiteurs qui en dérivent.

La troisième et dernière partie de notre travail comprend les calculs d'énergie libre de Gibbs pour chacun des systèmes simulés. Les énergies déterminées servent à confirmer l'interaction plus stable du DTG avec l'ADNv par rapport à l'intégrase puis à démontrer que les substitutions effectuées ont en effet contribué à renforcer les interactions des nouveaux inhibiteurs au sein de l'intasome entier, particulièrement avec l'ADNv, générant ainsi de nouveaux inhibiteurs de transfert de brin capables de contourner le problème de résistance.

Références bibliographiques

- [1] L. El Khoury, J.-P. Piquemal, S. Femandjian, R. G. Maroun, N. Gresh, and Z. Hobaika, "The inhibition process of HIV-1 integrase by diketoacids molecules: Understanding the factors governing the better efficiency of dolutegravir," *Biochem. Biophys. Res. Commun.*, vol. 488, no. 3, pp. 433–438, Jul. 2017, doi: 10.1016/j.bbrc.2017.05.001.
- [2] S. Hare *et al.*, "Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572)," *Mol. Pharmacol.*, vol. 80, no. 4, pp. 565–572, Oct. 2011, doi: 10.1124/mol.111.073189.
- [3] L. E. Khoury *et al.*, "Spectrometric and computational studies of the binding of HIV-1 integrase inhibitors to viral DNA extremities," *PeerJ Phys. Chem.*, vol. 1, p. e6, Dec. 2019, doi: 10.7717/peerj-pchem.6.
- [4] K. El Hage, J.-P. Piquemal, Z. Hobaika, R. G. Maroun, and N. Gresh, "Substituent-Modulated Affinities of Halobenzene Derivatives to the HIV-1 Integrase Recognition Site. Analyses of the Interaction Energies by Parallel Quantum Chemical and Polarizable Molecular Mechanics," *J. Phys. Chem. A*, vol. 118, no. 41, pp. 9772–9782, Oct. 2014, doi: 10.1021/jp5079899.
- [5] S. Hare, S. S. Gupta, E. Valkov, A. Engelman, and P. Cherepanov, "Retroviral intasome assembly and inhibition of DNA strand transfer," *Nature*, vol. 464, no. 7286, pp. 232–236, Mar. 2010, doi: 10.1038/nature08784.

Chapitre III

Démarches et outils

Chapitre III

Démarches et outils

Partie A : La Chimie Computationnelle : **de la mécanique quantique à la** **mécanique moléculaire**

Nous nous sommes basés au cours de nos travaux de thèse sur des méthodes de calcul de chimie quantique afin d'évaluer les énergies d'interactions intermoléculaires de nos entités biologiques étudiées reliées au VIH-1. Les méthodes de calcul choisies sont basées sur la théorie de la fonctionnelle de la densité ainsi que sur des méthodes de décomposition d'énergie et des méthodes d'analyses topologiques.

A.I. Notions de base de la chimie quantique

La chimie quantique est l'application de la mécanique quantique aux systèmes moléculaires afin de décrire le comportement électronique des atomes et des molécules. Elle permet donc d'expliquer les réactivités et les propriétés chimiques des systèmes, et ceci à partir de l'équation de Schrödinger ^[1]; équation fondamentale de la chimie quantique. Cette équation indépendante du temps est présente sous la forme suivante :

$$H |\varphi\rangle = E |\varphi\rangle$$

Avec E correspondant à l'énergie de la molécule, φ représente la fonction d'onde et H est l'opérateur Hamiltonien qui agit sur la fonction d'onde. Cependant, cette équation devient rapidement trop complexe pour être résolue exactement, d'où la nécessité d'en faire une résolution approchée via des méthodes d'approximation.

A.I.1. Théorie d'Hartree-Fock

La méthode Hartree-Fock (HF) est une méthode utilisée pour la résolution approchée de l'équation de Schrödinger. Connue aussi sous le nom de la méthode du champ auto-cohérent (Self Consistent Field ; SCF), elle constitue le point de

départ de résolution (méthode *ab-initio*), qui exige que le champ final calculé à partir de la distribution de charges soit « auto-cohérent » avec le champ initial.

Cette méthode permet la détermination de la fonction d'onde et de l'énergie d'un système quantique à plusieurs corps dans un état stationnaire, où elle suppose que la fonction d'onde du système peut être approximativement écrite sous la forme d'un déterminant de Slater ^[2] (équation représentée ci-dessous)

$$\psi_S(x_1, x_2, \dots, x_N) = |\varphi_i(x_1) \varphi_j(x_2) \dots \varphi_k(x_N)\rangle$$

et l'énergie de l'état stationnaire du système à plusieurs corps sera alors définie par l'équation ci-après

$$E_{HF} = \langle \psi_{HF} | H | \psi_{HF} \rangle$$

La méthode HF permet le calcul de la densité électronique mais cette approximation ne permet pas d'inclure les contributions de type corrélations électroniques associées aux systèmes à plusieurs corps en interaction. Elle considère donc que le mouvement d'un électron est indépendant du mouvement de ses voisins ^[3]. Afin d'inclure efficacement la corrélation électronique dans des systèmes d'intérêt pour nos études (i.e. des modèles de systèmes biologiques de taille non négligeable), nous pourrions nous baser sur la théorie de la fonctionnelle de la densité (Density Functional Theory, DFT) qui sera détaillée dans le paragraphe A.1.3.

L'approximation d'Hartree-Fock reste donc insuffisante pour calculer précisément les propriétés chimiques d'un système, d'où la nécessité de la corriger ou de la compléter. Diverses méthodes existent comme solution, connues sous le nom de méthodes post-HF qui seraient présentées dans le paragraphe A.1.2.

A.1.2. Méthodes post-Hartree-Fock

Les méthodes post HF sont des méthodes *ab-initio* développées afin d'améliorer la méthode Hartree-Fock. Elles utilisent la fonction d'onde HF comme approximation de départ pour déterminer d'une manière améliorée la fonction d'onde et l'énergie, et ceci en incluant la corrélation électronique entre les mouvements des électrons permettant ainsi d'éviter les erreurs flagrantes dans les calculs d'énergies des systèmes étudiés. L'introduction du facteur de correction se fait par les orbitales moléculaires qui prennent en considération les interactions instantanées des électrons (les forces individuelles répulsives électron-électron). Ce facteur est représenté par un opérateur exponentiel ^[4] ^[5].

Parmi les méthodes post-HF, nous citons :

- L'interaction de Configuration (IC) où la fonction d'onde est représentée comme une somme des déterminants de Slater qui représentent les excitations électroniques ^[6].
- Théorie de la perturbation de Møller-Plesset aux différents ordres MP2, MP3 et MP4 où l'hamiltonien dépend d'un paramètre λ ; une petite excitation ajoutée à ce dernier ^[7].

En dépit du fait que les méthodes post-HF génèrent de meilleurs résultats que les méthodes HF, elles nécessitent un coût de calcul supplémentaire qui s'avère le plus souvent important pour traiter de gros systèmes. Une alternative à ces méthodes sera la Théorie de la Fonctionnelle de la Densité (DFT) qui, au coût d'un calcul HF, permet le traitement de la corrélation de systèmes à plusieurs corps d'une manière précise et efficace.

A.1.3. Théorie de la Fonctionnelle de la Densité

La théorie de la Fonctionnelle de la Densité (DFT) est une méthode de calcul quantique qui offre la possibilité de calculer les propriétés énergétiques d'un système à plusieurs électrons en se basant sur la densité électronique, sans avoir recours à la fonction d'onde. L'énergie calculée est connectée à la densité grâce à une fonctionnelle (fonction dépendante d'une fonction). Cette énergie est donnée par la formule suivante :

$$E_{DFT} = E_{NN} + E_T + E_v + E_{coul} + E_{exch} + E_{corr}$$

Avec E_{NN} la répulsion noyau-noyau, E_T l'énergie cinétique de l'électron, E_v l'attraction noyau-électron, E_{coul} la répulsion électron-électron (répulsion de Coulomb), E_{exch} l'énergie d'échange électron-électron et E_{corr} l'énergie corrélée du mouvement des électrons à spin différent. E_{NN} , E_v et E_{coul} sont les mêmes énergies que celles utilisées dans la théorie HF alors qu' E_T et E_{exch} sont différentes que celles en HF. Le dernier terme, E_{corr} n'est pas pris en compte dans la théorie HF. Divers fonctionnelles peuvent être utilisées, dont les plus communes utilisées actuellement sont les fonctionnelles B97D3 et B3LYP ; ce sont des fonctionnelles hybrides, utilisant à la fois les énergies d'échange DFT et HF et donnant par la suite des résultats précis et relativement fiables.

L'avantage d'utiliser les méthodes DFT est leur dépendance de la densité électronique à 4 variables uniquement ; les trois coordonnées de l'espace et la coordonnée de spin, en comparaison à la dépendance de $4N$ variables de la fonction d'onde (avec N = nombre d'électrons). De plus, la DFT est indépendante de la taille du système. Par la suite, cette méthode permet une modélisation

précise et fiable des interactions inter-électroniques (notamment à courte portée) tout en nécessitant un faible coût du temps de calcul ^[8] ^[9].

A.I.4. Les fonctions de base

En chimie théorique et computationnelle, les fonctions de base (basis set) sont des ensembles de fonctions qui représentent la fonction d'onde utilisée dans les méthodes HF, post-HF et DFT, dans le but de transformer les équations aux dérivées partielles du modèle étudié en équation algébriques pouvant être implémentées et utilisées dans les calculs *in silico*. Elles apportent des améliorations à la description des orbitales moléculaires ^[10].

Dans le cadre de nos calculs de chimie théorique, nous avons eu recours aux bases cc-pVTZ ; ce sont des bases relativement étendues de type polarisation corrélacion-consistantes (cc-p), triple zeta ζ (TZ) pour les orbitales de valence V où les couches internes sont représentées avec une seule fonction par sous-couche et les couches de valence avec trois fonctions par sous-couche ^[11]. Notre choix de cette fonction de base est motivé par sa précision dans la représentation des orbitales atomiques et sa capacité d'autoriser des calculs sur nos complexes d'intérêt.

A.I.5 L'erreur de superposition de base

L'erreur de superposition de base (BSSE, Basis Set Superposition Error) est due au fait que les bases utilisées dans les calculs de chimie quantique comportent un nombre fini de fonctions. Or, pour une évaluation exacte de l'énergie, un nombre infini de fonctions est requis.

Par exemple, dans le cas d'un complexe AB, formé par deux molécules A et B où A possède une base a et B possède une base b, l'énergie d'interaction intermoléculaire du complexe AB (ΔE) est calculée de la manière suivante :

$$\Delta E_{AB} = E (AB)_{ab} - E (A)_a - E (B)_b$$

D'après la formule ci-dessus, la population électronique de A présente dans le complexe AB sera décrite par les fonctions composant la base a et par celles composant la base b, alors que pour les molécules isolées, leur population électronique sera décrite par leur propre base uniquement. Ce déséquilibre ainsi créé donnera lieu à un abaissement artificiel de l'énergie et sera connu sous le nom d'erreur de superposition de base ^[12].

Les calculs ab-initio présentés dans ce travail de thèse ont été effectués avec les logiciels GAUSSIAN09 ^[13] et GAMESS ^[14] en utilisant les fonctions de bases cc-pVTZ (-f) et en prenant en compte l'erreur BSSE.

A.II. Méthodes interprétatives quantiques

Après avoir calculé les énergies d'interactions entre les différents partenaires des systèmes complexes, une analyse des rendements obtenus est nécessaire afin d'interpréter les résultats. Différentes méthodes existent permettant l'analyse des interactions des molécules en termes physiques (énergie électrostatique, énergie de polarisation, énergie de répulsion ...) ainsi qu'en termes chimiques (liaisons chimiques, répartition de la population électronique ...)

Pour les méthodes d'analyse en termes physiques, elles se basent sur la décomposition de l'énergie d'interaction totale en plusieurs composantes ayant chacune un sens physique bien déterminé. Ces méthodes de décomposition

énergétique peuvent être basées sur la théorie de la perturbation de l'énergie d'interaction (méthode SAPT) ou bien sur un principe variationnel (méthodes RVS « Reduced Variational Space »^[15], KM « Kitaura-Morokuma »^[16] ou CSOV « Constrained Space Orbital Variation »)^[17].

Pour les méthodes d'analyses topologiques, elles se basent sur la densité électronique afin de fournir une représentation de l'état énergétique des molécules étudiées. On cite parmi ces méthodes, la méthode NCI (Non Covalent Interactions)^[18], la méthode ELF (Electron Pair Localization)^[19] et la méthode MEP (Molecular Electrostatic Potential)^[20].

Je me contenterais dans cette partie de ce chapitre d'expliquer plus particulièrement la méthode RVS et la méthode ALMOEDA comme démarches de décomposition énergétique ainsi que la méthode MEP comme outil d'analyse topologique.

A.II.1. Méthodes de décomposition d'énergie

A.II.1.1 Décomposition d'énergie : Approche RVS

L'évaluation et l'analyse de l'énergie d'interaction d'un ligand avec sa protéine cible sont primordiales pour la compréhension du mécanisme de leur interaction et pour l'amélioration de la spécificité de cette dernière. La méthode RVS (Reduced Variational Space), développée par Walter Stevens et William Fink, permet ces démarches en décomposant l'énergie d'interaction totale (ΔE) en composantes énergétiques distinctes ayant chacune un sens physique propre^[21].

$$\Delta E = E_S + E_X + E_{pol} + E_{ct} + E_{BSSE}$$

D'après la formule ci-dessus et la Figure 27, ΔE est en effet répartie en quatre contributions du premier et du second ordre, dans le but d'élucider l'énergie qui contribue essentiellement à l'interaction intermoléculaire d'un système. Le premier ordre correspond à un terme d'interaction électrostatique (E_s) ainsi qu'un terme d'échange répulsion (E_x) qui représente la répulsion entre les nuages électroniques occupés des monomères. Au premier ordre, l'énergie d'interaction se fait entre des monomères gelés ; aucune réorganisation du nuage électronique n'est possible dans les orbitales de chaque monomère. En fait, l'énergie du fragment A du complexe AB est obtenue en gelant les fonctions d'ondes du fragment B et inversement.

Le second ordre, qui est un terme d'induction, est lui-même décomposé en un terme de polarisation (E_{pol}) et un terme de transfert de charge (E_{ct}). Les deux énergies du second ordre résultent des transferts partiels des électrons des orbitales occupées d'un monomère vers ses propres orbitales vacantes (E_{pol}) ou vers les orbitales vacantes de l'ensemble des autres molécules (E_{ct}). Il ne faut pas oublier l'énergie due à l'erreur de superposition de base puisque le calcul est réalisé en base cc-pVTZ (-f) finie. Notons que cette analyse par décomposition se fait au niveau Hartree-Fock, donc en absence de la corrélation électronique, du coup le terme de dispersion n'est pas pris en compte.

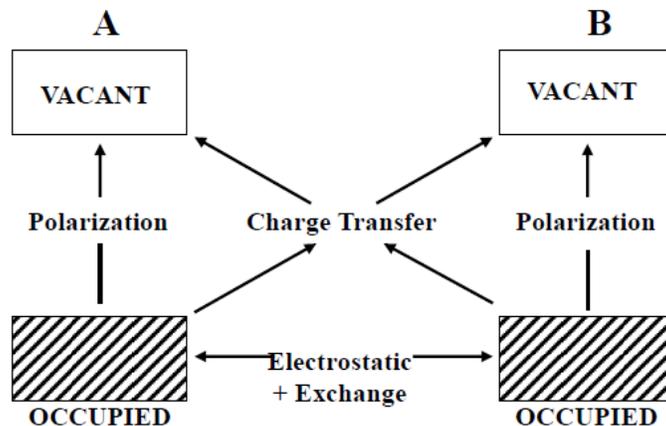


Figure 27 : Schéma de décomposition de l'énergie d'interaction en différentes contributions énergétiques selon la méthode RVS

(W.J Stevens and W.H. Fink, *Chemical Physics Letters*, vol. 139, no. 1, pp. 15-22, August 1987, doi: 10.1016/0009-2614(87)80143-4)

L'approche RVS permet l'évaluation de chaque contribution énergétique et donc sa reproduction d'une manière individuelle et indépendante des autres. Ceci la rend une méthode efficace pour la paramétrisation des champs de force polarisables.

A.II.1.2. Décomposition d'énergie : Approche ALMOEDA

ALMOEDA (Absolutely Localized Molecular Orbitals Energy Decomposition Analysis) est une technique de décomposition d'énergie qui partitionne cette dernière en divers composantes physiquement pertinentes. Elle permet donc la compréhension des aspects fondamentaux des interactions intermoléculaires [22].

Dans cette méthode computationnelle, l'énergie d'interaction totale est décomposée en trois termes ; un terme de polarisation (POL), un terme de transfert de charge (CT) ainsi qu'une composante appelée « densité gelée » (FRZ, Frozen Density), sous la forme suivante :

$$\Delta E = E_{FRZ} + E_{Pol} + E_{ct}$$

La composante de « densité gelée » représente la contribution des densités de monomères congelés en interaction. Elle correspond au changement d'énergie accompagnant le rapprochement de fragments infiniment séparés sans relaxation de leurs orbitales moléculaires. Le terme de polarisation représente quant à lui l'abaissement d'énergie suite à la relaxation intra-fragment des orbitales moléculaires occupées et gelées sur les fragments. Finalement, le terme d'énergie de transfert de charge est défini comme une relaxation supplémentaire entre les fragments des orbitales moléculaires. Il correspond à une baisse d'énergie suite au transfert d'électrons des orbitales occupées d'une molécule vers les orbitales virtuelles d'une autre molécule.

Notons que la méthode ALMOEDA, conceptuellement similaire aux autres méthodes de décomposition d'énergie, comprend toutefois plusieurs nouvelles fonctionnalités importantes. La première est un traitement auto-cohérent de la baisse d'énergie suite à la polarisation. En effet, cette diminution est évaluée par un calcul de champ auto-cohérent dans lequel les coefficients d'orbitales moléculaires sont absolument localisés dans les molécules en interaction afin de bloquer le transfert de charge. La deuxième fonctionnalité est l'utilisation d'une approximation perturbatrice au niveau du terme de transfert de charge ^[21].

Cependant, malgré l'utilité de la méthode ALMOEDA à analyser la nature des interactions intermoléculaires, elle présente deux inconvénients. Tout d'abord, le terme de polarisation ne présente pas une limite significative et définie de fonctions de bases (basis set) puis, le terme FRZ présente un caractère monolithique ; une décomposition plus poussée de ce terme en contributions électrostatiques permanentes, répulsion de Pauli et dispersion est souhaitée. Ces

obstacles ont été résolus récemment, définissant la seconde génération de la méthode ALMOEDA ; ALMOEDA2 ^[23].

A.II.2. Méthodes d'analyse topologique

Outre les méthodes de décomposition de l'énergie d'interaction qui quantifient la force de cette interaction, ils existent des méthodes topologiques qui permettent la localisation de cette dernière dans l'espace. En effet, ces méthodes proposent une étude de la densité ou des fonctions de la densité afin de rationaliser les structures (liaisons, population électronique ...)

Nous allons présenter dans cette partie la méthode de visualisation des contours de potentiels électrostatiques (MEP, Molecular Electrostatic Potential) que nous avons adopté dans nos travaux de thèse afin de visualiser la répartition des électrons au sein de nos inhibiteurs. Ceci permet alors d'expliquer leur comportement au sein des complexes biologiques du VIH-1.

A.II.2.1. Contours des potentiels électrostatiques moléculaires

Le potentiel électrostatique, créé notamment par les électrons dans l'espace autour de la molécule étudiée, est une propriété très utile pour l'analyse et la prédiction du comportement réactif de la molécule envers les espèces chargées se trouvant dans son environnement. Ce potentiel peut être déterminé expérimentalement par des techniques de diffraction ou bien théoriquement par calcul des contours des potentiels électrostatiques moléculaires. *In silico*, afin de générer ces contours, on commence par calculer l'énergie de potentiel électrostatique à une distance bien définie du noyau de la molécule cible. Cette énergie calculée est en fait une mesure de la force des charges voisines à une position particulière.

Puis, pour analyser avec précision la distribution de charges d'une molécule, les données d'énergie potentielle électrostatique générées sont représentées visuellement sous forme de contours. En fait, un spectre de couleur est donné avec des intensités différentes selon les valeurs d'énergie potentielle électrostatique ; les couleurs jaune/rouge correspondent aux zones de densité électronique élevée (δ^-) et les couleurs bleu/vert aux zones de faible densité électronique (δ^+) (Figure 28). Cette méthode, largement utilisée dans l'étude des systèmes et des processus biologiques ^[20] a été adoptée dans notre cas pour étudier l'effet des substitutions faites sur la molécule de DTG sur ses interactions avec l'intasome du VIH-1.

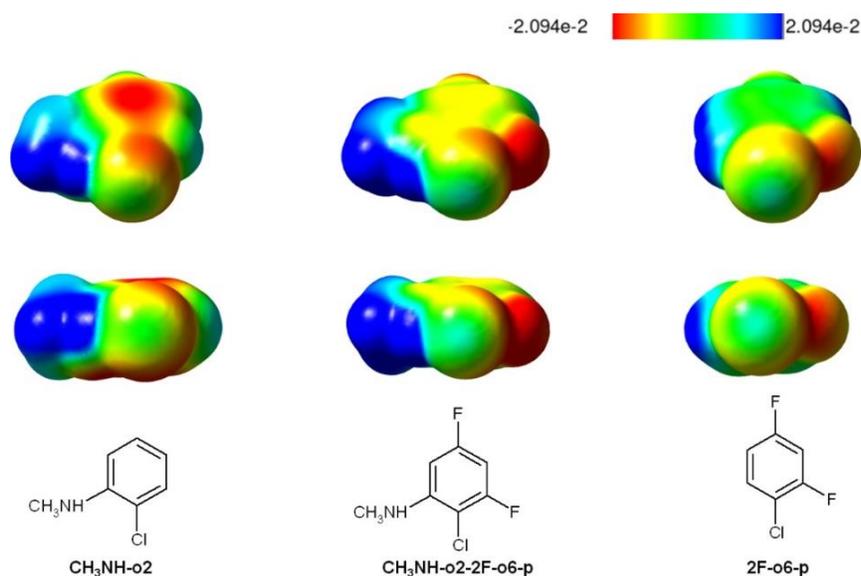


Figure 28: Représentation des contours de potentiels électrostatiques de trois ligands halobenzéniques de l'intégrase du VIH-1 à 0,0004 électrons Bohr⁻³ d'isodensité de surface.

(El Hage et al, *Journal of Computational Chemistry*, vol. 36, no. 4, pp. 210-221, February 2015, doi: 10.1002/jcc.23786)

A.III. Mécanique moléculaire : approches classiques et polarisables

La mécanique moléculaire (MM) est un ensemble de méthodologie appartenant à la discipline de la modélisation moléculaire, puisqu'elle se base sur les lois newtoniennes afin de modéliser des systèmes moléculaires. La MM joue donc un

rôle primordial dans la description des propriétés des molécules ; la prédiction de leurs structures, de leurs propriétés physico-chimiques ainsi que des forces qui s'exercent entre elles. Elle permet de même, par le biais d'un champ de force adéquat, de calculer l'énergie d'un complexe étudié. En fait, même si la chimie quantique reste l'approche idéale pour le calcul des interactions moléculaires, elle présente néanmoins quelques inconvénients majeurs pour la modélisation de systèmes biologiques ; elle est coûteuse en temps de calcul puisqu'elle prend en compte tous les électrons du système étudié, elle ne prend pas en considération la corrélation électronique entre les atomes et les forces de dispersion à longue distance et elle reste limitée à des complexes de petites tailles. D'où l'avantage de se baser sur la mécanique moléculaire pour les calculs et les analyses d'énergie d'interaction, puisqu'elle permet, en consommant peu de temps de calcul (elle considère les interactions entre atomes et non les atomes seuls), de modéliser et d'étudier des systèmes de tailles importantes.

Dans les parties suivantes, nous décrirons les champs de force les plus utilisés par les méthodes de mécanique moléculaire pour les études de systèmes biologiques et chimiques.

A.III.1. Champs de force classiques

Dans le cadre de la mécanique moléculaire, un champ de force est défini comme un ensemble de potentiels et de paramètres qui décrivent la structure atomique de l'énergie potentielle d'un système donné, sans avoir de recours explicite à des méthodes quantiques. Ces paramètres sont obtenus en étant calibrés sur des calculs de chimie quantique afin de reproduire des propriétés clés trouvées

expérimentalement. Un champ de force classique doit rendre compte des trois éléments suivants ^[24] :

- L'énergie résultant des interactions électrostatiques entre les molécules ; forces de Coulomb, forces de van der Waals et les liaisons hydrogènes.
- La modulation de cette énergie potentielle résultant de la polarisabilité des molécules.
- L'énergie à l'intérieur de la molécule, reliée à sa conformation (liaisons et angles).

D'où l'expression générale de l'énergie potentielle associée à un champ de force classique est donnée comme la somme des deux composantes intramoléculaires et intermoléculaires sous la forme suivante :

$$E_{\text{potentielle totale}} = E_{\text{intramoléculaire}} + E_{\text{intermoléculaire}}$$

Avec l'énergie intramoléculaire qui comprend les énergies au niveau des atomes liés par des liaisons covalentes (Energie d'élongation, énergie de flexion et énergie de torsion) et l'énergie intermoléculaire qui regroupe les énergies au niveau des atomes non liés par des liaisons covalentes (Energie électrostatique et Energie de van der Waals). L'expression générale de l'énergie potentielle devient alors :

$$E_{\text{potentielle totale}} = E_{\text{atomes liés coalement}} + E_{\text{atomes non liés coalement}}$$

$$E_{\text{potentielle totale}} = E_{\text{élongation}} + E_{\text{flexion}} + E_{\text{torsion}} + E_{\text{électrostatique}} + E_{\text{vdW}}$$

Ces différentes énergies prises en compte par les champs de force classiques sont représentées ci-dessous par la Figure 29.

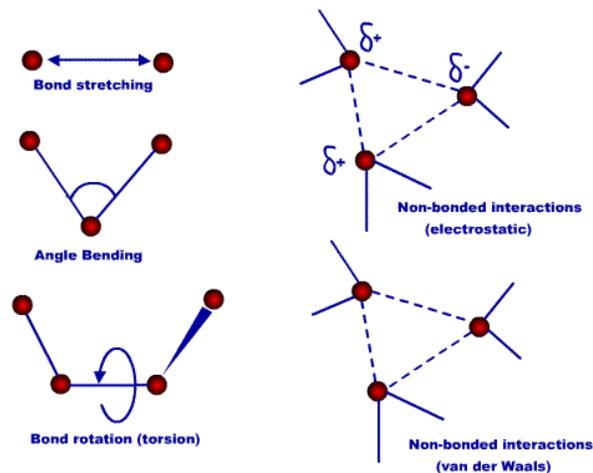


Figure 29 : Décomposition énergétique classique d'un champ de force de mécanique moléculaire. Les atomes sont représentés par des sphères rouges, les liaisons covalentes (à gauche) par des traits pleins bleus et les liaisons non covalentes (à droite) par des traits en pointillés.

(*Mécanique et modélisation moléculaire, biochimie.j.univ-angers.fr*)

L'énergie d'élongation correspond à l'énergie d'étirement de la liaison entre deux atomes voisins, l'énergie de flexion correspond à la courbure de l'angle entre trois atomes consécutifs et l'énergie de torsion est calculée au niveau de l'angle dièdre formé par quatre atomes consécutifs. Quant aux énergies intermoléculaires, nous trouvons un terme électrostatique calculé en attribuant aux atomes des charges ponctuelles partielles obtenues par calcul *ab initio* et un terme de vdW obtenu par un potentiel de type Lennard-Jones qui décrit les interactions attractives à longue portée résultant des mouvements corrélés des électrons et les interactions répulsives à courte portée entre deux particules séparées d'une distance r ($1/r^6$ pour les interactions attractives et $1/r^{12}$ pour les interactions répulsives).

Notons que les champs de force classiques ne diffèrent généralement que peu les uns des autres dans l'expression de leurs différentes composantes énergétiques et ne se distinguent que par les valeurs de leurs paramètres ^[25].

On cite comme exemple de champs de force classiques :

- AMBER (Assisted Model Building and Energy Refinement) ^[26] ; très utilisé pour la modélisation des protéines, ADN et ARN.
- CHARMM (Chemistry at HARvard Molecular Mechanics) ^[27] ^[28]; très utilisé pour les petites molécules ou les macromolécules.
- GROMOS ^[29] : champ de force classique utilisé notamment pour les simulations de dynamique moléculaire de systèmes biologiques en phase gazeuse ou en présence de molécules d'eau.

En dépit de leur vaste utilisation dans le domaine de la modélisation des complexes protéines/ligands, il existe toutefois quelques limitations à l'emploi des champs de force classiques, notamment dans l'étude de systèmes présentant des résidus chargés ou des cations métalliques. En effet, l'approche classique ne permet pas de traiter les effets de polarisation au niveau d'une molécule suite à l'approche d'une autre au sein du système biologique étudié. Ainsi le raffinement le plus recherché des champs de force classiques est l'addition explicite d'une contribution de l'énergie de polarisation (E_{pol}) qui donnera l'apparition des champs de force polarisables dits de seconde génération. Ces champs de force polarisables prennent en considération les effets de polarisation à N-corps, alors qu'avec les champs de force non polarisables, l'énergie n'est approximée qu'en utilisant des potentiels à deux corps (i.e. ce sont des potentiels de paires).

Le second raffinement des champs de force classiques s'est fait au niveau de la contribution électrostatique. En fait, ces derniers surestiment généralement les interactions électrostatiques dans le but de compenser l'absence de polarisation. De plus, ils sous-estiment l'anisotropie de la densité électronique puisqu'ils se basent sur des charges ponctuelles atomiques pour décrire la densité de charge.

Une solution à ces problèmes sera alors le développement de multipôles distribués d'ordre supérieur (charges, dipôles, quadrupôles) permettant ainsi de correctement rendre compte de l'anisotropie de la densité électronique et de mieux décrire la distribution de charges. Ces multipôles sont obtenus par des calculs de chimie quantique *ab initio* ^[30].

Comme exemple de champs de force polarisables de seconde génération, on peut citer :

- SIBFA (Sum of Interactions Between Fragments *Ab initio* computed) ^[31] ; champs de force pour les petites molécules et les protéines flexibles, formulé et calibré sur la base de calculs *ab initio*. Il a été développé par notre équipe, au sein du Laboratoire de Chimie Théorique, par Nohad Gresh et Jean-Philip Piquemal.
- AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) ^{[32] [33]} ; champs de force polarisable pour les biomolécules, développé notamment par Pengyu Ren et Jay Ponder ainsi qu'au sein de notre laboratoire.

Nous nous sommes intéressés aux deux champs de force polarisables mentionnés ci-dessus dans nos calculs énergétiques effectués sur les complexes du VIH-1. Je détaillerais donc dans la suite de cette partie ces deux champs de force.

A.III.2. Champs de force polarisables

Un des buts ultimes des travaux accomplis au cours de la thèse est d'ordonner les énergies d'interactions d'un nombre de ligands inhibiteurs ciblant l'intégrase du VIH-1 dans le but de prévoir leur affinité relative réelle pour l'IN. Ceci dit, il devient

critique de gérer de très faibles différences d'énergie ainsi que de réduire les marges d'erreur sur le calcul de l'énergie potentielle. D'où l'importance d'utiliser des champs de force de seconde génération, notamment dans les applications aux systèmes très chargés, puisqu'ils comprennent des raffinements au niveau du calcul de l'énergie potentielle, notamment l'inclusion explicite de la contribution de l'énergie de polarisation ^[34]. Cette dernière correspond à la réorganisation de la distribution de charges d'une molécule suite à sa soumission à une perturbation électrostatique externe et engendre une non-additivité de l'énergie totale qui n'est parfois pas négligeable dans l'étude des systèmes biologiques. Les champs de force polarisables prennent en compte cette non-additivité alors que, par définition, les champs de force non polarisables ne le font pas.

De nombreux modèles de polarisation explicite existent ^{[35] [36]} dont trois sont les plus utilisés dans les applications sur des systèmes d'intérêt biologiques. Nous trouvons le modèle de charges fluctuantes, le modèle de l'oscillateur de Drude et le modèle des multipôles ponctuels. C'est sur ce dernier modèle que sont basés les champs de force SIBFA et AMOEBA, où chaque centre polarisable présentera une polarisabilité dérivée d'un modèle empirique ou de la chimie quantique. Ils permettent ainsi une représentation de la distribution de charge incluant des dipôles et des quadripôles localisés sur les atomes et sur les liaisons chimiques ce qui permet de retranscrire une plus grande anisotropie des interactions électrostatiques.

Pour qu'un champ de force polarisable puisse être utilisé pour les calculs d'énergie, et reproduire avec précision les résultats obtenus par chimie quantique, il doit posséder les propriétés suivantes ^[37] :

- Séparabilité : ΔE totale d'interaction doit pouvoir s'exprimer sous la forme de contributions énergétiques distinctes dont chacune doit être formulée et calibrée sur les bases de la chimie quantique afin de bien reproduire sa contrepartie quantique résultant d'une analyse de décomposition d'énergie.
- Anisotropie : ΔE totale d'interaction et ses composantes doivent pouvoir reproduire les variations énergétiques en fonction de l'orientation respective des fragments.
- Non-additivité : ΔE totale d'interaction diffère de la somme des énergies de chaque monomère du complexe en interaction. Cette différence peut être significative. On parle alors d'interactions coopératives quand ΔE totale est plus grande en valeur absolue que la somme énergétique de chaque monomère et d'interactions anti coopératives quand ΔE totale est plus faible en valeur absolue.
- Transférabilité : Les paramètres calibrés sur de petits systèmes modèles doivent être appliqués à une pluralité de complexes de tailles arbitrairement grandes, sans avoir à changer la calibration initiale.

Les champs de force SIBFA et AMOEBA répondent à beaucoup des critères mentionnés ci-dessous et paraissent adaptés à l'étude des interactions ligands-macromolécules.

A.III.2.1 Champs de force polarisable : SIBFA

Introduit dans les années 80 par Nohad Gresh et ses collaborateurs ^[31], SIBFA (Sum of Interaction Between Fragments *Ab initio* Computed) est un champ de force polarisable de deuxième génération formulé et calibré sur la base de calculs *ab initio*, dans le but de reproduire les contributions énergétiques quantiques issues

des démarches de décomposition d'énergie au niveau Hartree-Fock. Ce potentiel comprend un traitement complet des effets à N-corps et décrit avec grande précision les effets de pénétration électrostatique et le transfert de charge. De plus, il a recours à des multipôles et des polarisabilités distribuées pour calculer les contributions électrostatiques et adopte une représentation explicite des doublets d'électrons libres pour le calcul des contributions à courte portée. SIBFA permet donc d'effectuer des calculs fiables sur les énergies intermoléculaires induisant les propriétés de liaison des molécules à intérêt biologique et pharmaceutique. Il est ainsi adéquat à être utilisé pour l'étude de métalloprotéines comportant un ou plusieurs sites de complexation des cations métalliques. C'est le cas de l'intégrase du VIH-1 présentant un site de fixation du Mg^{2+} au niveau de son domaine catalytique, d'où notre choix d'adopter SIBFA comme champs de force polarisable au cours des études de notre cible protéique.

A.III.2.1.1 Formulation du potentiel SIBFA

D'après le potentiel SIBFA, l'énergie d'interaction totale ΔE est égale à la somme de cinq contributions différentes dont chacune est calibrée sur la base de sa contribution correspondante issue d'un calcul quantique *ab initio*. Nous pourrions alors écrire la formule de $\Delta E_{\text{intermoléculaire totale}}$ de la forme suivante :

$$\Delta E_{\text{intermoléculaire totale}} = E_{\text{MTP}} + E_{\text{rep}} + E_{\text{pol}} + E_{\text{ct}} + E_{\text{disp}}$$

Elles représentent respectivement la contribution électrostatique multipolaire (E_{MTP}) issue des multipôles *ab initio*, la répulsion E_{rep} à courte portée destinée à reproduire la contribution d'échange-répulsion de Pauli entre les électrons des orbitales moléculaires des molécules en interaction. E_{pol} correspond à l'énergie de polarisation due aux transferts d'électrons partiels des orbitales moléculaires

occupées d'un monomère vers ses propres OM virtuelles, E_{ct} à l'énergie de transfert de charge des OM occupées d'un monomère vers les OM virtuelles de l'ensemble des autres molécules. Finalement E_{disp} représente l'énergie de dispersion qui est la stabilisation due aux interactions des dipôles instantanés de transition des molécules en interaction. A noter qu' E_{MTP} et E_{rep} font partie de la contribution énergétique du premier ordre (E_1) alors qu' E_{pol} et E_{ct} font partie de la contribution énergétique du deuxième ordre (E_2).

Chaque fragment moléculaire étudié en SIBFA sera défini par sa géométrie interne, sa topologie et ses multipôles et polarisabilités distribuées. Il sera ensuite stocké dans la bibliothèque de fragments SIBFA ^[31].

Nous allons détailler dans les parties qui suivent chaque contribution énergétique calculée par SIBFA.

A.III.2.1.2 Calcul de l'énergie multipolaire

La contribution électrostatique multipolaire (E_{MTP}) est calculée à partir de multipôles distribués sur les barycentres des atomes et des liaisons. Ces multipôles ; monopôle, dipôle et quadrupôle sont obtenus à partir des OM des molécules en question suite à une démarche développée par Vigné-Maeder et Claverie ^[38]. Des raffinements ont été apportés au calcul de cette contribution par ajout d'un terme de pénétration de charge E_{pen} ^{[39] [40]} qui traduit le recouvrement des nuages électroniques de deux atomes proches et permet par la suite une meilleure reproduction de la contribution Coulombienne quantique à courte distance.

E_{MTP} s'exprime comme la somme des six contributions électrostatiques suivantes :

$$E_{MTP} = E_{\text{mono-mono}} + E_{\text{mono-dip}} + E_{\text{mono-quad}} + E_{\text{dip-dip}} + E_{\text{dip-quad}} + E_{\text{quad-quad}}$$

Avec $E_{\text{mono-mono}}$ l'interaction monopôle-monopôle à courte distance, $E_{\text{mono-dip}}$ l'interaction monopôle-dipôle, $E_{\text{mono-quad}}$ l'interaction monopôle-quadropôle et $E_{\text{dip-dip}}$ l'interaction entre deux dipôles. Quant au terme $E_{\text{dip-quad}}$, il désigne l'interaction entre un dipôle et un quadropôle alors qu' $E_{\text{quad-quad}}$ correspond à l'interaction quadropôle-quadropôle.

A.III.2.1.3 Calcul de l'énergie d'échange-répulsion

L'énergie d'échange-répulsion (E_{rep}) est calculée comme la somme des interactions liaison-liaison, liaison-doublet libre et doublet libre – doublet libre. Elle représente donc la répulsion qui a lieu entre les distributions électroniques des fragments en interaction.

$$\begin{aligned}
 E_{\text{rep}} = & \sum_{\substack{p-q \\ \text{liaison de A}}} \left[\sum_{r-s} \text{rep}(p-q, r-s) \right] \\
 & + \sum_{\substack{p-q \\ \text{liaison de A}}} \left[\sum_{L_\beta} \text{rep}(p-q, L_\beta) \right] \\
 & + \sum_{\substack{L \\ \text{paire libre de A}}} \left[\sum_{r-s} \text{rep}(L_\alpha, r-s) \right] \\
 & + \sum_{\substack{L \\ \text{paire libre de A}}} \left[\sum_{L_\beta} \text{rep}(L_\alpha, L_\beta) \right]
 \end{aligned}$$

Dans l'équation ci-dessus, p-q et r-s représentent les liaisons entre deux atomes voisins des fragments A et B respectivement alors que L_α et L_β correspondent aux paires libres des fragments A et B respectivement.

Comme le cas d' E_{MTP} , E_{rep} a été raffinée par l'inclusion d'un terme S^2/R ^[41] où S représente le recouvrement des orbitales localisées et R la distance entre les centroïdes des orbitales impliquées.

A.III.2.1.4 Calcul de l'énergie de polarisation

L'énergie de polarisation (E_{pol}) est équivalente à la déformation des distributions de charges suite à l'interaction d'une molécule avec le champ électrique généré par ses partenaires environnants. Elle dépend donc du champ électrique et des polarisabilités des molécules. Ce champ polarisable sera calculé avec les mêmes multipôles permanents qu' E_{MTP} . Il est écranté par une fonction gaussienne afin de corriger les effets à courte-portée comme la polarisation d'échange ^[42]. Dans SIBFA, les polarisabilités sont distribuées sur les barycentres des liaisons et sur les doublets non-liants des hétéroatomes, selon une méthode proposée dans les années 90 par Garmer et Stevens permettant d'obtenir les multipôles distribués et les polarisabilités à partir d'un seul calcul ab initio effectué sur la molécule étudiée ^[43]. Ils peuvent être extraits de calcul avec le logiciel GAMESS ^[14] au niveau HF ou au niveau corrélé par la méthode DFT.

E_{pol} est calculée comme le produit scalaire d'un champ électrique avec un dipôle. Le champ décrit dans la formule ci-dessous sera égal à la somme du champ généré par tous les multipôles permanents du système et le champ généré par les dipôles induits ^[44].

$$\Delta\mu_p(i) = \sum_j^{x,y,z} \alpha_p(i,j) * E_p(j) \quad i \in x, y, z$$

Avec ΔU le dipôle induit crée suite à l'interaction des polarisabilités anisotropes avec le champ électrique situé au point p d'un fragment A. E_p est le champ électrique généré par les multipôles permanents et induits de tous les points q appartenant à tous les fragments B, différents du fragment A, et calculé au point p .

A.III.2.1.5 Calcul de l'énergie de transfert de charge

Le calcul de l'énergie de transfert de charge (E_{ct}) dérive des travaux de Murrell et ses collaborateurs [45]. E_{ct} est calculée entre tous les doublets-libre d'un fragment A donneur d'électrons et les orbitales virtuelles d'un fragment B accepteur d'électrons. Elle dérive d'une théorie de la perturbation et dépend du recouvrement des orbitales localisées des fragments donneurs et accepteurs d'électrons. L'équation ci-après permet l'expression de l'énergie de transfert de charge sous la forme suivante :

$$E_{CT} = -2C \sum_{l_\alpha \in A} \left(\sum_{\beta^* \in B} N_{occ}(l_\alpha) \frac{(I_{l_\alpha \beta^*})^2}{\Delta E_{l_\alpha \beta^*}} \right)$$

Où :

- l_α correspond à une paire d'électrons libres du fragment A.
- β^* à une orbitale virtuelle du fragment B.
- $N_{occ}(l_\alpha)$ est le nombre d'occupation de l'orbitale l_α .
- C est une constante ajustable

- $(I_{l_{\alpha\beta^*}})$ est une fonction du recouvrement entre les orbitales moléculaires du donneur et de l'accepteur
- $\Delta E_{l_{\alpha\beta^*}}$ est une fonction de la différence entre les potentiels d'ionisation du fragment A et de l'affinité électronique du fragment B.

A.III.2.1.6 Calcul de l'énergie de dispersion

L'énergie de dispersion (E_{disp})^[31] correspond aux interactions attractives à longue distance des dipôles instantanés. Elle est calculée entre deux atomes i et j comme une somme des interactions atome-atome, atome-paire libre et paire libre –paire libre, sous la forme :

$$Z = \frac{r_{ij}}{\sqrt{R_i * R_j}}$$

Dans la formule d' E_{disp} représentée ci-dessus, r_{ij} est équivalente à la distance entre les atomes i et j alors que R_i et R_j sont équivalents aux rayons effectifs des atomes i et j respectivement.

Comme ses contributions E_{MTP} et E_{rep} , E_{disp} a été aussi améliorée par l'ajout d'un terme d'échange-dispersion.

A.III.2.2 Champs de force polarisable : AMOEBA

AMOEBA (Atomic Multipole Optimized Energetics for Biomolecular Applications) est un champ de force polarisable de deuxième génération développé en 2004 par Jay Ponder et ses collaborateurs^{[46] [47]}, avec comme objectif d'atteindre une précision chimique dans les propriétés thermodynamiques des protéines en

solution tout en s'appuyant sur une évaluation des énergies des modèles de mécanique quantique. Comme SIBFA, il a donc été paramétré et calibré sur les bases de la chimie quantique avant d'être testé sur des cations monovalent tels que Na^+ et divalents comme Mg^{2+} , Ca^{2+} et Zn^{2+} . Il est donc adéquat à être utilisé pour l'étude de systèmes biologiques complexant des cations métalliques grâce à sa distribution multipolaire et sa prise en compte des effets de polarisation.

Nous avons eu recours dans notre cas à ce potentiel intégré dans un logiciel de dynamique moléculaire, Tinker-HP^[48], pour simuler les complexes de l'intasome du VIH-1 en présence du DTG et des nouveaux inhibiteurs et pour en extraire les valeurs d'énergie libre d'interaction des ligands avec les éléments de l'intasome viral (ADNv et IN).

A.III.2.2.1 Formulation du potentiel AMOEBA

L'énergie d'interaction intermoléculaire totale calculée par AMOEBA est présentée comme une somme de trois contributions. Elle s'écrit alors sous la forme qui suit :

$$\Delta E_{\text{intermoléculaire totale}} = E_{\text{vdW}} + E_{\text{elec}} + E_{\text{pol}}$$

Avec E_{vdW} , comprenant les contributions de dispersion et d'échange-répulsion, calculée avec un potentiel 7-14 proposé par Thomas Halgren^[55], E_{elec} correspondant à l'énergie électrostatique et E_{pol} désignant l'énergie de polarisation.

Nous détaillerons aussi les contributions énergétiques d'AMOEBA dans les parties qui suivent.

A.III.2.2.2. Calcul de l'énergie de vdW

E_{vdW} est calculée grâce à un potentiel « buffered 14-7 » à la place du potentiel de Lennard-Jones 12-6 standard ^[49]. Le nouveau potentiel s'est montré plus efficace pour reproduire les énergies d'interactions, notamment dans le cas des gaz rares.

La formule suivante est utilisée pour le calcul de E_{vdW} pour deux atomes i et j :

$$E_{vdw}(ij) = \epsilon_{ij} \left(\frac{1.07}{\rho_{ij} + 0.07} \right)^7 \left(\frac{1.12}{\rho_{ij}^7 + 0.12} - 2 \right)$$

Ce potentiel dépend de la distance entre i et j au travers du terme ρ_{ij} qui est égale à R_{ij} / R_{ij}^0 avec R_{ij}^0 la distance pour laquelle l'énergie est minimale. ϵ_{ij} correspond quant à elle à la profondeur du puits.

A.III.2.2.3. Calcul de l'énergie électrostatique

Dans AMOEBA, les multipôles sont localisés sur les centres des atomes et sont définis dans des repères locaux pour chaque site, construits à partir des atomes voisins. Ceci permet aux multipôles de rester constants dans leur cadre local défini au cours des simulations de dynamique moléculaire.

Ces multipôles sont extraits par la méthode DMA (Distributed Multipoles Analysis) ^[50] implémentée dans le logiciel GDMA de Stone ^[51]. La démarche d'extraction consiste en une répartition de la densité de charge sur les sites atomiques, dans le but de déterminer les moments dipolaires.

L'énergie d'interaction électrostatique sera calculée par la relation suivante :

$$E_{elec}^{perm} = \sum_{i,j \neq i \text{ atomes}} (M_i)^T T_{ij} M_j$$

Avec M_i l'ensemble des multipôles sur l'atome i , M_j l'ensemble des multipôles sur l'atome j et T_{ij} la matrice d'interaction des multipôles de i avec les multipôles de j .

A.III.2.2.4. Calcul de l'énergie de polarisation

L'énergie de polarisation est calculée dans AMOEBA suite à l'interaction du champ électrique d'un atome i avec les polarisabilités créant par la suite un dipôle induit. Elle prend en considération les interactions entre les moments dipolaires permanents et les interactions entre les moments induits.

Notons qu'un terme d'écrantage a été ajouté au calcul de l'énergie de polarisation pour diminuer la distribution de charge à courte distance, qui peut causer des problèmes quand les sites étudiés dans un système sont très proches. Le modèle de Thole est capable de reproduire les tenseurs de polarisabilité moléculaire de nombreuses petites molécules avec une précision raisonnable en utilisant uniquement la polarisabilité atomique ^[52].

Chapitre III

Démarches et outils

Partie B : La dynamique moléculaire conventionnelle

La modélisation moléculaire regroupe différentes méthodes permettant la simulation du comportement d'un système dont la taille peut aller d'une simple molécule diatomique à des macromolécules biologiques de plusieurs millions d'atomes. La modélisation moléculaire peut ainsi décrire le comportement des atomes en fonction du temps, pour tenter d'expliquer leurs interactions. Certains cas d'applications de la modélisation moléculaire sont dédiés à la compréhension du processus de repliement d'une protéine ou encore à l'explication de l'importance de certains acides aminés au niveau d'un site catalytique enzymatique.

Les méthodes de modélisation moléculaire reposent sur différentes approches de calculs et de simulation numérique qui expriment l'énergie d'interaction des atomes du système moléculaire en fonction de leur position relative. La méthode à laquelle nous nous sommes intéressés au cours de la thèse est la dynamique moléculaire conventionnelle. Notre choix pour cette méthode se base sur le fait qu'elle permet un suivi rigoureux de l'évolution de la structure d'un système donné et des interactions qu'il établit au sein de son environnement.

En effet, la dynamique conventionnelle a été mise au point dans le but de résoudre numériquement les problèmes à N-corps de la mécanique classique, en attribuant des équations du mouvement pour chaque particule. De plus, elle donne accès aux trajectoires dans l'espace. Ce qui fait de la dynamique conventionnelle une méthode spécifiquement adaptée au calcul des propriétés dynamiques.

Dans cette partie, nous allons présenter la méthode de dynamique conventionnelle dans ses grandes lignes puis nous passerons à son adaptation à nos études des systèmes de l'intasome du VIH-1 complexé avec les inhibiteurs de transfert de brin.

B.I. Principe de la dynamique moléculaire classique

La dynamique moléculaire classique est une technique de simulation numérique qui, en se basant sur les équations de Newton (notamment la seconde loi), permet de suivre l'évolution dans le temps d'un ensemble de particules en interaction. Suite à l'application de cette méthode, des trajectoires successives, désignant un ensemble de conformations successives d'un même système moléculaire, seraient générées. Elles décrivent la variation des coordonnées des atomes en fonction du temps. Des bilans énergétiques évaluant les paramètres thermodynamiques des interactions peuvent être tirés de ces trajectoires obtenues.

En connaissant l'énergie potentielle du système à un temps t , il est alors possible d'associer un vecteur vitesse à chacune de ses particules. La seconde loi de Newton est ainsi exprimée en dynamique moléculaire par l'équation suivante :

$$\vec{F}_i = m_i \vec{a}_i = m_i \frac{\partial^2 \vec{r}_i}{\partial t^2}$$

Avec F_i la force exercée sur la particule i de masse m_i et de position r_i (x, y, z). a_i représente l'accélération exercée sur la particule i .

La formule ci-dessus permet donc la détermination des positions des atomes via leurs coordonnées cartésiennes ainsi que leurs vitesses respectives. En fait, au temps $t = 0$, l'état initial du système étudié est spécifié dans une région finie de l'espace (comme la boîte de simulation) avant de commencer le suivi temporel de l'évolution du système où à chaque instant $t' = t + \Delta t$ (avec Δt désignant le pas de temps élémentaire), une nouvelle position des atomes est déterminée. Ceci produit

une quantité énorme de données nous permettant d'étudier et d'analyser des systèmes biologiques complexes dont la manipulation expérimentale peut s'avérer être compliquée. Plusieurs notions sont indispensables aux simulations de dynamique moléculaire conventionnelle. Nous les exposerons dans les parties qui suivent.

B.I.1. Conditions de calcul

Afin d'assurer le bon déroulement d'une simulation de dynamique moléculaire, plusieurs conditions et paramètres doivent être appliquées. Je commencerais par présenter ces conditions puis passer à l'explication du protocole de dynamique moléculaire que nous avons adopté.

B.I.1.1. Echelles de temps

Le défi majeur des simulations de biomolécules reste la modélisation des phénomènes lents tout en tenant compte des mouvements rapides. Une simulation d'un système cible donné ne peut pas se faire sur un temps infini. En effet, la dynamique moléculaire reste limitée par la vitesse et les capacités de calculs des ordinateurs.

Les échelles de temps peuvent varier de quelques nanosecondes à une dizaine de microsecondes. Il convient de choisir le pas de temps utilisé avec soin ; un pas de temps court n'est pas conseillé puisque la simulation ne peut pas converger ni explorer correctement l'espace des phases. Réciproquement, un pas de temps trop grand affecte la stabilité du système ^[53].

Les échelles de temps respectées au cours des simulations moléculaires sont les suivantes :

- Pour une élongation d'une liaison : 10^{-14} s.
- Pour la rotation des chaînes latérales de surface : $10^{-11} - 10^{-10}$ s.
- Pour la rotation des chaînes latérales intérieures : $10^{-4} - 1$ s.
- Pour le repliement des protéines : $10^{-6} - 10^{-2}$ s.

B.I.1.2. Les ensembles

Une simulation de dynamique moléculaire est caractérisée par plusieurs grandeurs telles que le nombre total de particules (N), le volume (V), la pression (P), la température (T) et l'énergie (E). Quelques-unes de ces grandeurs seraient fixées au cours des simulations alors que les autres variables non fixées seraient susceptibles de fluctuer. Ceci permet alors la définition des ensembles statistiques qui contrôlent les propriétés physiques des systèmes et permettent de reproduire des conditions physiques expérimentales.

Trois ensembles statistiques principaux sont majoritairement utilisés dans les simulations de dynamique conventionnelle ^[54]. Ils correspondent aux situations suivantes :

- L'ensemble NVE (ensemble microcanonique) : Le système n'échange pas ni énergie ni particules avec le milieu extérieur. On parle alors de système isolé thermodynamiquement. Cet ensemble est caractérisé par une énergie totale conservée, un nombre total de particules fixes ainsi qu'un volume constant. Le système peut évoluer sans limites n'étant pas soumis à d'autres restrictions.

- L'ensemble NVT (ensemble canonique) : cet ensemble est défini par un volume et une température constants où le système est en contact avec un thermostat. Dans ce cas, son énergie (E) n'est pas fixe en raison des échanges avec le thermostat. Cet ensemble mime les conditions expérimentales les plus fréquentes.
- L'ensemble NPT (ensemble isobare-isotherme) : dans cet ensemble, le système physique est en contact avec un barostat qui fixe la pression. De plus, son nombre de particules et sa température sont fixes. NPT permet donc d'étudier des transitions de phases puisque le volume du système varie en raison de ses échanges avec le barostat.

B.I.1.3. Les conditions périodiques

Au cours de nos simulations de dynamique moléculaire conventionnelle, des conditions périodiques sont appliquées afin de conserver les propriétés biophysiques du système. Ceci est fait puisque ce dernier est placé au centre d'une boîte avec d'autres molécules (eau, ions ...) dans le but de reproduire au plus près son environnement réel. Cependant, une grande portion des particules du système se trouvent près de la surface de la boîte, elles ne sont pas alors dans les mêmes conditions que les particules au centre de la boîte. Pour s'affranchir de cette difficulté, on utilise des conditions aux limites périodiques où la boîte est répliquée selon les trois directions de l'espace formant alors un système infini. Chaque atome appartient à une boîte entourée d'un ensemble de boîtes identiques de façon que tout atome qui sortirait par un côté de la boîte arriverait directement par le côté opposé avec la même vitesse (Figure 30). Ceci permettra à tous les atomes d'être entourés de la même façon par les molécules du solvant, même ceux qui sont situés

au niveau des bords de la boîte, ce qui garde une densité moyenne constante du système ^[55] ^[56]. Tout calcul effectué pour les interactions non liées tient bien évidemment compte des conditions périodiques et fait intervenir la version la plus proche de chacune des particules impliquées.

L'avantage des simulations dans les conditions périodiques est qu'il n'est pas nécessaire de conserver les coordonnées de toutes les images de la simulation mais seulement celles de la boîte centrale.

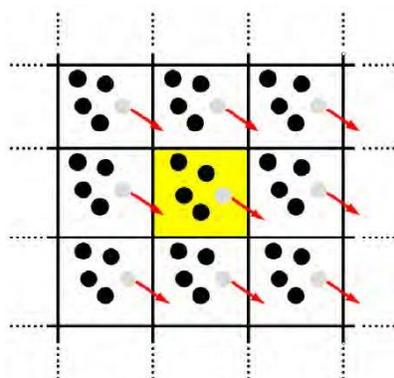


Figure 30 : Représentation bidimensionnel des conditions aux limites périodiques. Les trajectoires des particules dans la boîte de simulation centrale sont copiées dans toutes les directions.

(Steinhauser, *Molecular Dynamics – Studies of Synthetic and Biological Macromolecules*, April 2012, doi: 10.5772/36289)

B.II. Nos choix des systèmes biologiques simulés du VIH-1

La structure initiale de départ utilisée pour nos simulations de dynamique moléculaire est le cristal de l'intasome du Prototype Foamy Virus (Intégrase virale + ADN viral) en présence du Dolutégravir et des cations divalents de Mg^{2+} . Elle est déposée dans la banque de données protéiques (PDB) sous le code 3S3M. La Figure 31 schématise notre complexe initial sur lequel s'effectueront nos simulations.

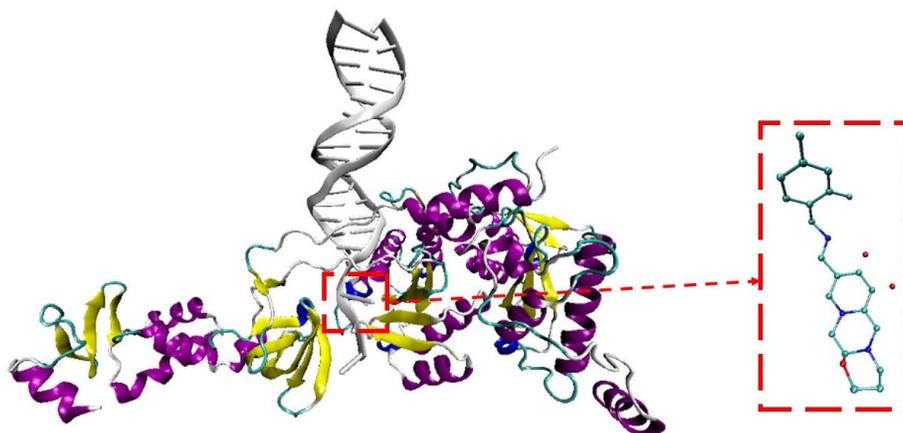


Figure 31 : Représentation tridimensionnelle de l'intasome du PFV en complexe avec le Dolutégravir (DTG) et Mg^{2+} . L'ADN viral double brin (couleur grise) est intégré au niveau de son extrémité 3' au sein de l'intégrase virale (colorée selon ses structures secondaires ; hélices α en violet et les feuillets β en jaune). Le DTG ainsi que Mg^{2+} (deux sphères rouges) se trouvent situés au niveau de l'extrémité 3' à l'interface ADN viral – Intégrase (ils sont mis en relief dans le cadre rouge à droite). Cette structure est déposée dans la banque de données « PDB » sous le code 3S3M.

Nous avons utilisé pour nos simulations le potentiel polarisable AMOEBA, implémenté dans le logiciel massivement parallèle Tinker-HP ^[48] développé dans notre laboratoire, puisque ce potentiel permet une représentation précise et correcte des effets de la polarisation et donc du site actif de nos complexes où se trouvent les Mg^{2+} chélatés. AMOEBA comprend les paramètres des acides nucléiques (ADN et ARN) et des acides aminés des protéines. Nous avons par conséquent eu besoin de générer les paramètres de nos nouveaux ligands (les ligands seront présentés dans les résultats du chapitre 4 partie B) pour pouvoir les simuler correctement. Ceci a été fait via le programme poltype ^[57], développé par l'équipe de Pengyu Ren à l'Université de Texas à Austin. Ce programme fournit, à partir d'une structure initiale (les coordonnées PDB dans notre cas) tous les paramètres liés et non liés d'un ligand, selon les étapes suivantes :

- 1- Optimisation de la géométrie de la structure initiale par calcul de chimie quantique au niveau MP2 (Møller-Plesset niveau 2) avec la fonction de bases 6-311G.
- 2- Analyse des multipôles distribués pour générer un ensemble de multipôles atomiques.
- 3- Affinement de ces multipôles en les adaptant au potentiel électrostatique moléculaire au niveau MP2 avec la fonction de base Aug-cc-pVTZ.
- 4- Attribution des paramètres de liaison, des angles, des dièdres et de vdW.
- 5- Affinement des paramètres aux forces intramoléculaires dérivées de la quantique.
- 6- Exécution d'une simulation de dynamique moléculaire afin de valider les paramètres et les ré-affiner si nécessaire.

Un récapitulatif de la démarche de paramétrisation est schématisé dans la Figure 32.

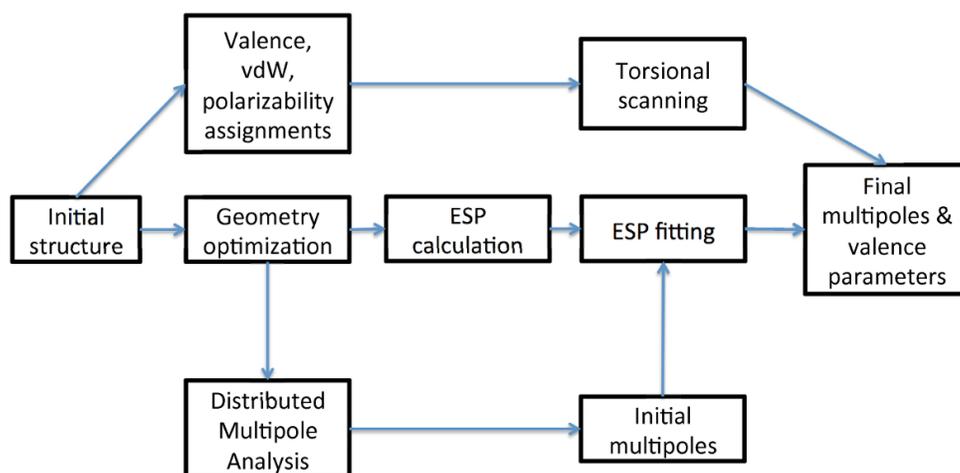


Figure 32 : Le processus global de paramétrisation des ligands avec le champ de force AMOEBA.

(P. Ren, C. Wu and J. W. Ponder, *The Journal of Chemical Theory and Computation*, vol 7, no. 10, August 2011, doi:10.1021/ct200304d)

B.III. Notre Protocole de dynamique moléculaire conventionnelle appliqué

Après avoir choisi nos ligands et généré leurs paramètres, nous sommes passés à la préparation de nos systèmes pour les simulations de dynamique moléculaire. Tout d'abord, il faut commencer par définir la boîte de simulation ainsi que l'ensemble des positions et des vitesses initiales des particules. Cette définition des positions se fait soit à partir de zéro, soit à partir des données d'une simulation précédente. Dans notre cas, les positions initiales des partenaires au sein de l'intasome sont issues de la structure PDB 3S3M. Il faut rappeler qu'il existe une homologie structurale importante entre l'intégrase du VIH-1 et celle du PFV. De plus, tous les résidus actifs de la triade catalytique et ceux en contact avec les molécules d'INSTIs sont invariants entre les deux intégrases virales (Figure 33). Ces arguments sont en faveur de l'utilisation de l'intégrase du PFV comme système de substitution fiable pour nos études structurales concernant l'intégrase du VIH-1 ^[58], en l'absence d'un cristal de l'intégrase entière de ce dernier.

Une fois la structure PDB récupérée, nous devons effectuer quelques modifications avant sa solvataion dans la boîte d'eau de simulation. Ceci se fait par ajout des atomes d'hydrogènes manquants à la structure cristallographique suivi d'une protonation des résidus d'acides aminés protéiques ce qui nous permet d'avoir la structure du complexe à un pH physiologique cellulaire (pH=7).

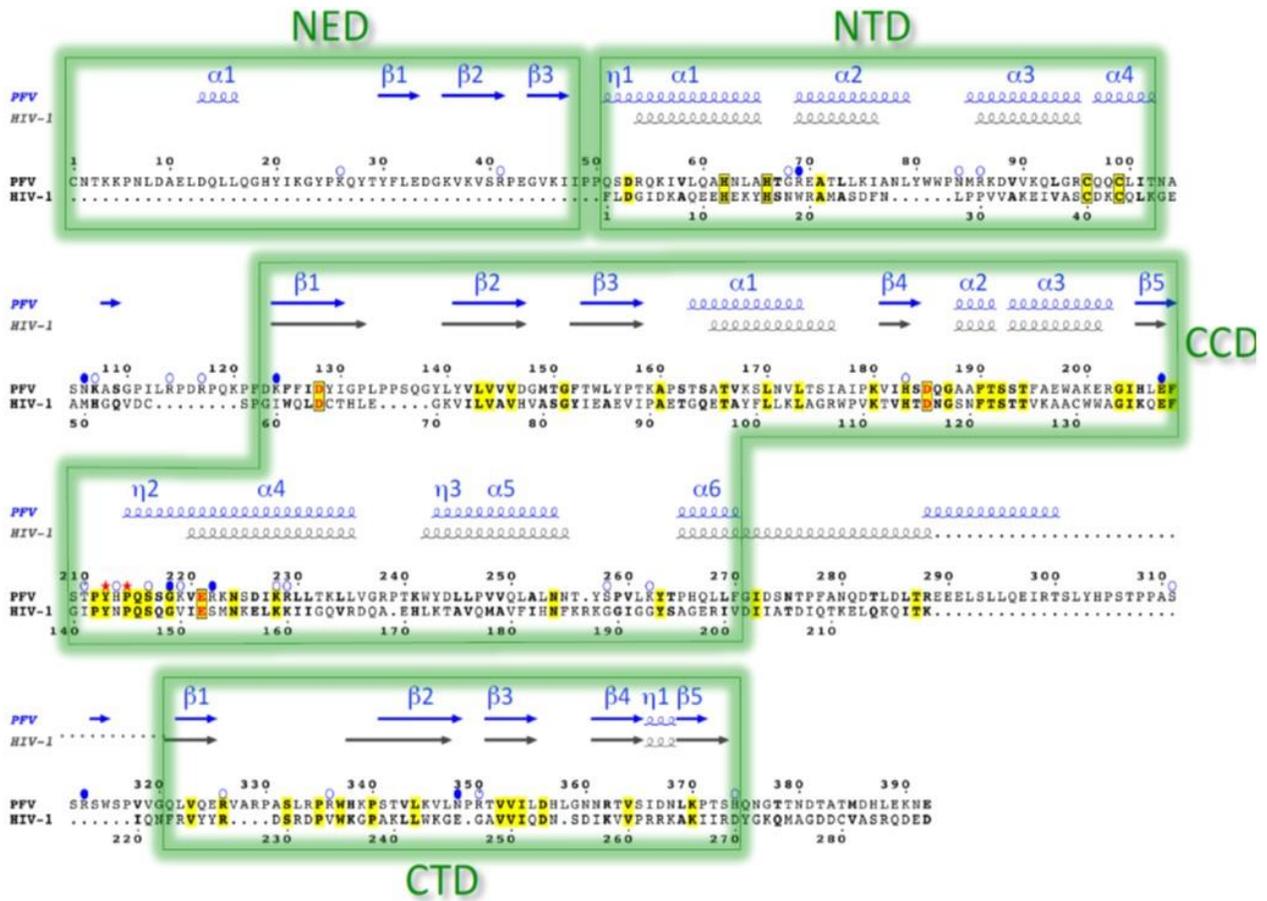


Figure 33 : Alignement des résidus de l'intégrase du PFV et ceux du VIH-1. Les résidus impliqués dans les interactions directes avec le squelette phosphodiester de l'ADN ou les bases virales présentent des cercles vides ou colorés en bleu respectivement. La Tyrosine 212 (Y212) et la Proline 214 (P214) impliquées dans les interactions avec les INSTIs sont mises en relief par une étoile rouge. Les structures secondaires ; hélice α , feuillet β et hélice 3_{10} η sont indiquées ainsi que les trois domaines N-terminal (NTD), C-terminal (CTD), le cœur catalytique (CCD) et le domaine d'extension du NTD (NED) chez l'intégrase du PFV. Les résidus conservés entre les deux virus sont marqués en gras et jaune et ceux appartenant à la triade catalytique sont encadrés et marqués en rouge.

Toutes nos simulations ont été réalisées dans des conditions de solvant explicite ; nos systèmes se trouvent entourés de molécules d'eau explicites formant une boîte cubique, comme représenté dans la Figure 34, avec une distance minimale entre le soluté et le bord de boîte le plus proche de 10 Å. Outre les molécules d'eau, nous avons ajouté des contre ions de positions de départ aléatoires (KCl de concentration de 150 mM) afin de neutraliser la charge totale du système. Le KCl a été choisi car il est le sel le plus abondant en milieu intracellulaire, nous permettant

de mimer le milieu physiologique intracellulaire au cours de nos simulations et de se rapprocher ainsi des conditions réelles. KCl sert de même à stabiliser la structure de l'ADNv pendant nos simulations.

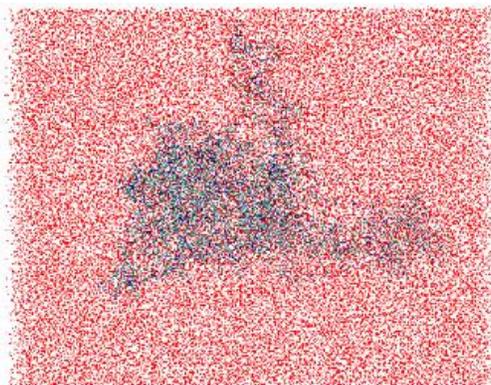


Figure 34 : Une de nos boîtes d'eau cubiques contenant l'intasome du PFV en présence du Dolutégravir et de Mg^{2+} , utilisée au cours de nos simulations.

Après solvatation de nos systèmes dans des boîtes d'eau de charge nulle, nous avons effectué une minimisation d'énergie respective, afin de générer pour chacun la conformation la plus stable, à partir de laquelle nous pourrions débiter les longues simulations de dynamique moléculaire. Au cours de la procédure de minimisation, seules les molécules d'eau et les contre ions sont relâchés alors que les partenaires étudiés sont figés, dans le but de permettre une organisation vraisemblable de l'eau autour des biomolécules. Vient ensuite l'étape d'équilibration, pendant laquelle le système est chauffé progressivement de 50K à 300K, par palier de 50 Kelvin, en ensemble NVT, tout en gardant le complexe fixe et les molécules d'eau et les ions KCl relâchés. Ce chauffage progressif permet aux systèmes d'atteindre une température stable de 300K (température ambiante). La durée de chaque étape de chauffage est de 2 ns, avec un enregistrement des trajectoires de dynamique moléculaire toutes les 100 ps. Une fois à 300K, une simulation supplémentaire est réalisée dans les mêmes conditions (durée totale de

2 ns, enregistrement toutes les 100 ps et système figé) en ensemble NPT afin de stabiliser la densité et le volume de la boîte d'eau. Ensuite, nos systèmes sont relaxés et des dynamiques moléculaires supplémentaires sont alors effectuées sans contraintes, pour une durée de 2 ns chacune, en commençant dans des conditions NVT puis NPT, afin de s'assurer de la stabilité du système libre dans la boîte d'eau. Une fois que nos systèmes sont stables, neutres et dans des conditions similaires au milieu physiologique cellulaire, nous pouvons débiter l'exécution de longues simulations de dynamique moléculaire ; c'est la phase de production. Les paramètres utilisés au cours de cette étape sont les suivants : un pas d'intégration de 10 fs (Il correspond à la durée d'une pause d'une simulation de dynamique moléculaire) avec l'intégrateur BAOABRESPA1^[59]. Ce dernier conçu au sein de notre laboratoire, est un intégrateur à multi-pas de temps (multi-timestep) nous permettant d'intégrer numériquement les équations du mouvement de Newton, tout en accélérant la dynamique moléculaire des dipôles ponctuels polarisables. BAOABRESPA1 réduit aussi l'écart de performance entre les champs de force polarisables et classiques tout en préservant les propriétés statiques et dynamiques du système simulé.

Puisque nos simulations sont réalisées dans des ensembles NVT et NPT, elles requièrent la présence d'un thermostat et d'un barostat pour assurer un contrôle de la température et de la pression respectivement^[60]. Tout d'abord, la température étant reliée à l'énergie cinétique du système, il est possible de la contrôler en ajustant les vitesses des atomes du système. Les thermostats les plus utilisés consistent à ajouter des variables pour imiter les propriétés d'un bain thermique qui contrôlerait la température. Plusieurs options de thermostats existent ; nous avons choisi le thermostat de Bussi^[61] où les vitesses sont ajustées

par un facteur λ défini de manière à assurer une évolution temporelle de l'énergie cinétique physique du système, engendrant des simulations canoniques et continues. La pression du système quant à elle est reliée à son volume. Plusieurs barostats existent pour assurer le contrôle de la pression au cours des simulations de dynamique moléculaire. Nous avons choisi le barostat de Berendsen ^[62] qui utilise un algorithme reliant les fluctuations de pression au volume, et maintient ce dernier constant en ajustant les coordonnées atomiques de la boîte de simulation. Un schéma récapitulatif de l'ensemble de notre protocole de dynamique moléculaire est représenté à la Figure 35.

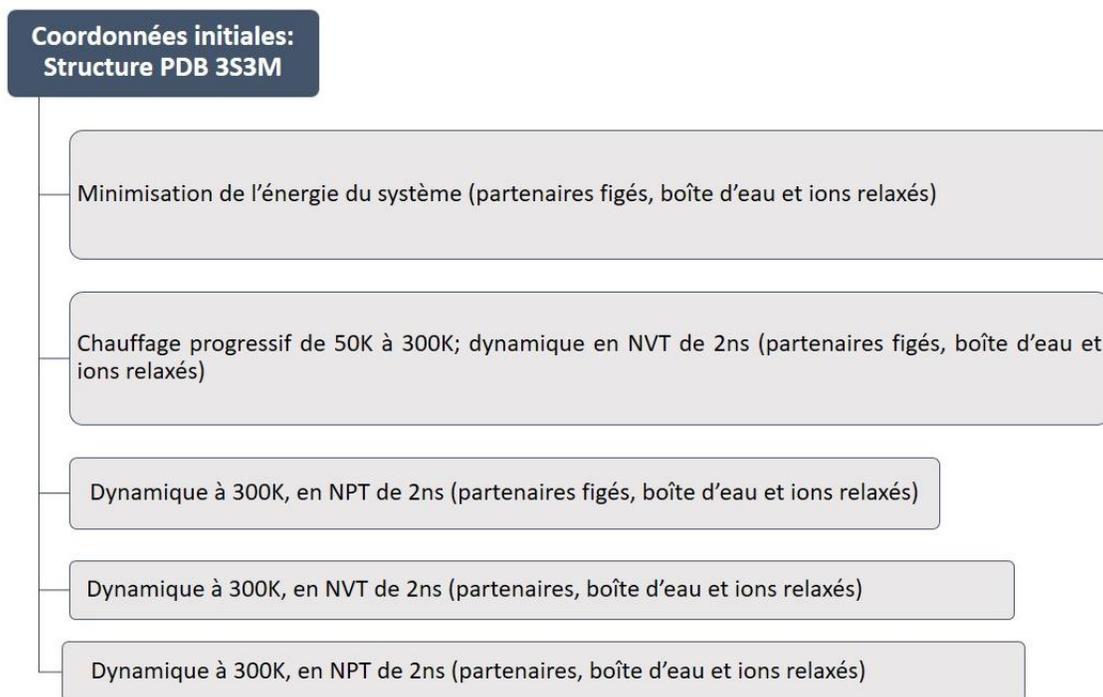


Figure 35 : Diagramme récapitulatif de notre protocole de dynamique moléculaire

Chapitre III

Démarches et outils

Partie C : Calculs d'énergie libre

Nous nous sommes appuyés sur la conception de médicaments, basée sur la structure, afin de concevoir à partir du Dolutégravir, de nouvelles molécules inhibitrices ciblant l'étape de transfert de brin orchestrée par l'intégrase. Afin d'évaluer le rendement énergétique des nouvelles drogues ainsi que leur efficacité thérapeutique, de nombreuses approches informatiques sont disponibles.

Nous avons choisi d'analyser les nouvelles molécules conçues par calcul de leur énergie libre d'interaction (ΔG) avec l'intasome viral. En fait, les calculs d'énergie libre alchimique basés sur des simulations de dynamique moléculaire en solvant explicite constituent une approche à haut niveau de rigueur théorique, calculant les différences d'énergie libre à partir de principes mécaniques incluant naturellement les effets du solvant ^[63]. Pour cela, cette méthode rigoureuse a été choisie malgré son coût en temps de calcul.

Ce chapitre sera organisé de la façon suivante : tout d'abord, nous commencerons par un aperçu des aspects théoriques de l'énergie libre de liaison y compris sa définition ainsi que celle du cycle thermodynamique. Nous passerons ensuite à la présentation des méthodes d'estimation des énergies libres des simulations équilibrées avant de terminer par notre protocole de calcul d'énergie libre absolue alchimique, utilisé pour effectuer nos calculs.

C.I. Notions théoriques générales sur le calcul d'énergie libre alchimique

Les calculs d'énergie libre alchimique se basent sur un cycle thermodynamique comprenant une série d'états intermédiaires alchimiques (non physiques) à partir desquels la différence d'énergie libre entre un état initial et un état final peut être calculée comme la somme des différences d'énergie libre entre toutes les paires d'éléments alchimiques des états intermédiaires ^[64]. Une partie du système, en

général un ligand, sera « modifiée » au cours des simulations effectuées créant ainsi les différents états alchimiques. Nous nous concentrerons en particulier sur les calculs d'énergie libre absolue de liaison (approche ABFE ; Absolute Binding Free Energy) où cours de leur cycle thermodynamique, le ligand « disparaît » non physiquement de la boîte de simulation et est « inséré » ensuite dans son site de liaison.

Malgré son coût élevé en temps de calculs, plusieurs avantages existent à l'utilisation de la méthode ABFE pour calculer les énergies d'affinité de nos nouvelles molécules conçues ; le premier est son efficacité à générer une valeur d'énergie libre exacte pour le modèle utilisé. Le deuxième avantage est la généralité de la méthode puisqu'elle peut être appliquée à n'importe quel complexe cible protéique-ligand. De plus, une solution au problème de la consommation importante en temps de calcul, est l'accès aux GPUs (Graphics Processing Unit) qui permettent d'effectuer des calculs à des échelles de temps de la microseconde, rendant les calculs alchimiques sur des complexes importants raisonnables à effectuer ^[65].

Des exemples d'applications potentielles de cette approche sont : la prédiction de la sélectivité et de l'affinité de ligands à échafaudages différents envers une cible protéique précise, le transfert d'une petite molécule d'une phase aqueuse à une phase apolaire et les effets des mutations de la chaîne latérale d'une protéine sur les affinités de liaison ^[66].

La liaison réversible d'un ligand (L) à une protéine (P) pour former un complexe (PL) peut être décrite par la formule suivante :



A l'équilibre, la constante de liaison K_b° est définie comme le ratio de la concentration du complexe et celles de chaque partenaire en solution ;

$$K_b^\circ = c^\circ \frac{[PL]}{[L][P]}$$

Avec c° la concentration de l'état standard équivalente à 1 mol/L.

Notons que l'état standard est défini par les paramètres suivants : une pression constante de 1 atm, une concentration des éléments du système de 1 mol/L, dans un volume standard V^0 de 1660 Å³ (volume occupé par une 1 mol/L).

L'énergie libre de liaison à l'état standard est par la suite définie par la formule suivante :

$$\Delta G_b^\circ = -k_B T \ln K_b^\circ$$

Où k_B est la constante de Boltzmann et T la température en Kelvin. Notons que dans les conditions standard, pour une mole du ligand, la constante des gaz parfaits ($R = 8,314 \text{ J mol}^{-1} \text{ K}^{-1}$) est alors utilisée à la place de la constante de Boltzmann.

Comme on peut le remarquer, la constante d'équilibre K_b° dépend de la concentration standard de référence. Il est nécessaire d'avoir la même concentration à l'état standard de 1 mol/L pour comparer des valeurs de ΔG entre elles. Cependant, au cours des calculs de ΔG , les simulations de dynamiques moléculaires de nos systèmes ne se font pas à la concentration standard de 1 mol/L, du coup une correction est nécessaire pour avoir les valeurs standard d'énergie libre de liaison.

Pour ΔG_b correspondant à l'énergie libre de liaison calculée dans une boîte de volume V , l'énergie libre standard ΔG_b° sera égale à la formule ci-après :

$$\Delta G_b^\circ = \Delta G_b - k_B T \ln \frac{V}{V^\circ}$$

Avec V° le volume standard de 1660 \AA^3 correspondant à la concentration standard c° de 1 mol/L .

C.II. Cycle thermodynamique alchimique

Dans les calculs d'énergie libre, le cycle thermodynamique alchimique comprend une série d'états intermédiaires alchimiques (non physiques) permettant de lier l'état initial du système à son état final. Par exemple, l'état initial peut correspondre au ligand libre en solution et l'état final peut représenter le même ligand lié au sein de son site d'interaction dans la protéine cible. En simulant uniquement ces deux états, il n'est pas possible d'obtenir une estimation fiable de l'énergie libre de liaison, d'où l'importance d'ajouter plusieurs étapes intermédiaires reliant les deux états, ce qui permet un passage régulier et correcte de la situation initiale à la situation finale.

Ce cycle thermodynamique doit être défini avant de commencer les calculs énergétiques. Il dépendra uniquement des états initial et final, quel que soit le nombre ou la nature des états intermédiaires choisis reliant les deux. En effet, la valeur de l'énergie libre de liaison (ΔG_b) ne doit pas varier en fonction des états intermédiaires utilisés. Par conséquent, ΔG_b est définie comme une fonction d'état (state function).

La position de départ du ligand, sa position finale ainsi que ses positions intermédiaires seraient simulées indépendamment les unes des autres, par une série de simulations équilibrées de dynamique moléculaire. Ensuite, la différence d'énergie libre entre un état et son successeur sera calculée. L'énergie libre totale de liaison sera alors récupérée comme la somme de toutes ces différences d'énergie.

Soit « i » un état intermédiaire simulé et « i+1 » son état successeur et N le nombre totale d'états constituant le cycle thermodynamique, l'énergie libre de liaison sera alors calculée de la manière suivante :

$$\Delta G_b = \sum_{i=0}^{i=N-2} \Delta G_{i,i+1}$$

Le cycle thermodynamique alchimique dans le cas de calcul d'énergie libre absolue de liaison sera détaillé dans la partie qui suit.

C.II.1. Cycle thermodynamique alchimique des énergies libres absolues

Dans le cas de calcul d'énergie libre absolue de liaison, le cycle thermodynamique correspondant comprend en général six étapes ; deux étapes physiquement significatives (état initial et état final) ainsi que quatre parties intermédiaires alchimiques, n'existant pas réellement. Comme mentionné précédemment, le rôle de ces dernières sera le passage de l'état de départ à l'état final d'interaction. La Figure 36 illustre en détails les différentes étapes du cycle thermodynamique dans le cas des énergies libres absolues de liaison. Il considère le cas où l'état initial est un ligand libre en solution avec un état final où le ligand est complexé à sa cible protéique. Les détails d'exécution des étapes divers de ce cycle seraient présentés et expliqués ultérieurement dans ce chapitre.

Tout d'abord, le cycle débute par le ligand (L) seul et libre en solution, non lié à la protéine cible (P) qui elle aussi se trouve libre dans un autre environnement (état A dans la Figure 36). Le ligand interagit avec les molécules d'eau via des interactions intermoléculaires électrostatiques ainsi que des interactions de vdW ($L_{\text{elec+vdW}}$). Ensuite, il sera découplé de son environnement de simulation ; ses interactions intermoléculaires (électrostatiques et vdW) avec le solvant seraient supprimées progressivement, tout en gardant ses interactions intramoléculaires. Ce découplage est quantifié par la valeur $\Delta G^{\text{solv}}_{\text{elec+vdw}}$ et est représenté par l'étape B. Notons qu'on parle d'annihilation quand on élimine à la fois les interactions intermoléculaires et intramoléculaires du ligand avec son environnement. L'étape C du cycle thermodynamique comprend l'ajout de contraintes au ligand découplé (L_{rest}) afin de limiter son volume d'échantillonnage accessible. Ceci donnera la valeur de $\Delta G^{\text{solv}}_{\text{rest}}$. L'ajout de ces contraintes favorise la convergence des calculs d'énergie libre. En fait, si le ligand n'est pas restreint lors de son découplage, il pourra s'éloigner de la protéine et donc de son site d'interaction et flotter dans la totalité de la boîte de simulation. Cela causera un problème une fois ses interactions avec le site de liaison protéique seraient réactivées, puisqu'il aura à passer par un processus de liaison physique afin de retrouver sa position dans la protéine. La durée des états simulés seraient alors très longues afin de ramener le ligand à sa position initiale. Ensuite, le ligand découplé et restreint sera transféré dans une boîte de simulation contenant la protéine avec laquelle il interagit. Ce transfert n'a pas de coût énergétique ($\Delta G = 0$) puisque la protéine et le ligand restreint n'interagissent pas encore entre eux. L'étape D de la Figure 36 illustre cette situation (PL_{rest}). L'étape suivante (étape E) sera de réactiver les interactions électrostatiques et de vdW du ligand avec la protéine, tout en gardant ses

contraintes. Ceci permet de retenir le ligand au sein de son site d'interaction afin de générer des valeurs correctes des énergies d'interaction ($\Delta G^{\text{prot}}_{\text{elec+vdW}}$). Finalement, après rétablissement des interactions intermoléculaires entre la protéine et le ligand, les contraintes fixant ce dernier sont éliminées. Cette étape F engendre alors la valeur $\Delta G^{\text{prot}}_{\text{rest}}$. Le ligand libre se trouve en interaction totale au sein de sa cible ($PL_{\text{elec+vdw}}$) et le cycle thermodynamique alchimique est alors bouclé. Ainsi, ΔG^0_b qui correspond à l'énergie libre standard de liaison du passage de l'état initial A à l'état final F est obtenue via plusieurs étapes intermédiaires n'existant pas réellement (étapes B, C, D et E).

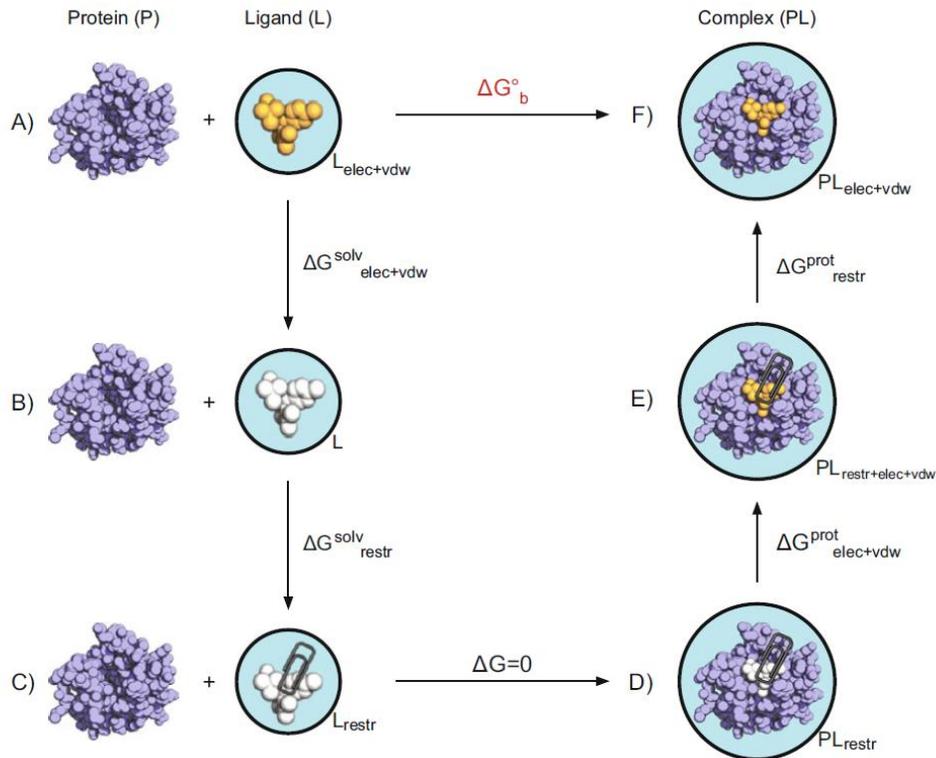


Figure 36 : Cycle thermodynamique pour le calcul d'énergie libre absolue. A) Etat initial avec le ligand (L) libre en interaction totale avec la solution et en absence de la protéine. B) Le ligand est découplé de la solution par élimination de ses interactions intermoléculaires électrostatiques et de vdW au cours d'une série de simulations équilibrées, fournissant le terme $\Delta G^{\text{solv}}_{\text{elec+vdw}}$. C) Le ligand découplé est restreint avec une valeur énergétique $\Delta G^{\text{solv}}_{\text{restr}}$. D) Le ligand restreint est transféré dans une solution contenant sa cible protéique. E) Les interactions intermoléculaires électrostatiques et de vdW sont régénérées tout en gardant les contraintes au niveau du ligand situé dans le site actif de la protéine, ce qui donne la valeur de $\Delta G^{\text{prot}}_{\text{elec+vdw}}$. F) Les contraintes du ligand sont éliminées et le cycle thermodynamique est ainsi bouclé avec comme étape finale le ligand en interaction totale avec la protéine cible.

(Aldeghi et al, *Chemical Science, Royal Society of Chemistry* vol. 7, no. 207, September 2015, doi:10.1039/c5sc02678d)

L'énergie libre de liaison à l'état standard sera alors égale à la somme des énergies libres calculées au niveau des étapes intermédiaires et mentionnées précédemment, sous la forme suivante :

$$\Delta G_b^{\circ} = \Delta G_{\text{elec+vdw}}^{\text{solv}} + \Delta G_{\text{restr}}^{\text{solv}} + \Delta G_{\text{elec+vdw}}^{\text{prot}} + \Delta G_{\text{restr}}^{\text{prot}} - k_B T \ln \frac{V}{V^0}$$

Avec $\Delta G^{\text{solv}}_{\text{elec+vdw}}$ l'énergie libre de découplage du ligand dans le solvant, correspondant à son passage de la phase aqueuse à la phase gazeuse (passage de l'état A à l'état B). $\Delta G^{\text{solv}}_{\text{rest}}$ est l'énergie libre d'ajout des contraintes au ligand non interagissant avec son environnement (passage de l'état B à l'état C). $\Delta G^{\text{prot}}_{\text{elec+vdw}}$ est l'énergie libre de la régénération des interactions intermoléculaires du ligand restreint au sein de son site protéique (passage de l'état D à l'état E) et finalement $\Delta G^{\text{prot}}_{\text{rest}}$ est l'énergie correspondant à l'élimination des contraintes qui ont gardé le ligand fixe quand il n'interagissait pas avec son environnement de simulation. A

noter que le dernier terme, $-k_B T \ln \frac{V}{V^0}$ représente la correction de l'état standard puisque les simulations de dynamique moléculaire ne se font pas à une concentration de 1 mol/L.

Notons qu'il existe trois règles qui doivent toujours être suivies lors de la définition des états intermédiaires et l'exécution de leur découplage.

- L'utilisation d'un potentiel « soft-core » lors du découplage des interactions de vdW, à la place du potentiel de Lennard-Jones. En effet, ce dernier exprime par sa formule ci-dessous les interactions de vdW attractives et répulsives, avec r la distance entre deux atomes, ϵ l'énergie de dispersion et σ la distance pour laquelle le potentiel d'énergie particule-particule V est nul.

$$V(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right]$$

Par conséquent, l'utilisation du potentiel de Lennard-Jones pour les variations des énergies de vdW entraîne un chevauchement inégal de l'espace de phase entre les états intermédiaires ainsi que des instabilités numériques suite à une singularité à

$r = 0$. Le potentiel « soft-core » modifie l'équation de Lennard-Jones de manière que l'énergie d'interaction soit finie pour toute configuration, via l'équation ci-dessous :

$$V_{ij} = 4\varepsilon_{ij}(1 - \lambda)^t \left([\alpha_{LJ}\lambda^s + (r_{ij}/\sigma_{ij})^n]^{-\frac{12}{n}} - [\alpha_{LJ}\lambda^s + (r_{ij}/\sigma_{ij})^n]^{-\frac{6}{n}} \right)$$

Où ε_{ij} et σ_{ij} sont les paramètres identiques à ceux exprimés dans la formule de Lennard-Jones, entre deux atome i et j . r_{ij} est aussi la distance interatomique. α_{LJ} est un paramètre ajouté pour ajuster le potentiel, dont une valeur de 0,5 donne la meilleure estimation de l'énergie libre de liaison. t , s et n sont des paramètres aux valeurs respectives de 1, 2 et 6. Le potentiel « soft-core » résolve donc les instabilités observées lors du découplage des paramètres de vdW [67] [68].

- Le découplage des interactions électrostatiques et de vdW doit se faire séparément, et ceci suite à la définition d'un paramètre de découplage ; λ qui peut prendre n'importe quelle valeur entre 0 et 1 et détermine ainsi l'état intermédiaire du système le long de la voie alchimique. En fait, pour $\lambda = 0$, le ligand est totalement découplé ; il y a absence des interactions intermoléculaires entre lui et ses partenaires. Alors que $\lambda = 1$ caractérise l'état où le ligand effectue des interactions intermoléculaires avec l'ADN viral. λ varie donc de 1 à 0 pour éliminer les interactions intermoléculaires et de 0 à 1 pour les engendrer au sein du complexe simulé. Pour une situation de départ où le ligand n'interagit pas avec son environnement, λ_{vdW} et $\lambda_{\text{électrostatique}}$ sont égales à 0. On commence par varier λ_{vdW} de 0 à 1 tout en gardant la valeur $\lambda_{\text{électrostatique}}$ fixée à zero. Puis une fois λ_{vdW} atteint la valeur de 1, les interactions électrostatiques peuvent alors apparaître avec une

variation de $\lambda_{\text{électrostatique}}$ de 0 à 1. Pour une situation de départ où le ligand est en interaction totale avec le milieu simulé, une valeur de 1 est attribuée à λ_{vdW} et $\lambda_{\text{électrostatique}}$. Dans ce cas, on commence par découpler les interactions électrostatiques ($\lambda_{\text{électrostatique}}$ passe de 1 à 0) en gardant λ_{vdW} égale à 1. Une fois $\lambda_{\text{électrostatique}}$ arrive à 0, le découplage des interactions de VdW peut débuter, λ_{vdW} varie alors de 1 à 0. En gros, les variations des interactions de vdW ne peuvent avoir lieu qu'en absence des interactions électrostatiques. Ceci est dû au fait qu'en présence de charges, si on varie les interactions de vdW selon le potentiel «soft-core », l'énergie potentielle de l'interaction risque de tendre vers l'infini dans la limite $r_{ij} = 0$. Il en résulte des forces importantes qui causeraient un crash des simulations. Pour cela, on doit s'assurer que les interactions de vdW sont présentes avant les interactions électrostatiques ou quand ces dernières sont découplées, les paramètres de vdW sont présents ^[69].

- Garder la charge nette du système constante et nulle tout au long du chemin alchimique puisque la plupart des simulations de dynamique moléculaire se base sur la sommation d'Ewald pour traiter périodiquement les contributions électrostatiques. Or cette méthode nécessite que la boîte de simulation soit neutre, pour cela des contre ions sont ajoutés au système simulé afin de le neutraliser ^[70].

C.II.2. Types de contraintes

Les contraintes appliquées au ligand jouent un rôle important dans le maintien de la position de ce dernier au niveau de son site de liaison protéique au cours des simulations de dynamique moléculaire visant à calculer son énergie libre de liaison.

Comme déjà expliqué, en éliminant progressivement les interactions intermoléculaires du ligand avec son environnement, il risque de quitter la poche de liaison et d'échantillonner tout le volume de la boîte. Ceci engendrera un grand espace de phase de configurations disponibles et entraînera des problèmes de convergence. On entend par espace de phase l'ensemble de toutes les configurations possibles que peut avoir un système simulé. L'ajout des contraintes est par la suite primordial pour s'assurer que la conformation échantillonnée lors des simulations correspond à un état lié bien défini et pour permettre un bon chevauchement de l'espace de phase entre deux états ce qui engendre une convergence plus rapide ^[71].

Plusieurs types de contraintes existent et peuvent être appliquées. En théorie, elles peuvent toutes être utilisées tant que le ligand est fixé dans sa pose de liaison et sa contribution d'énergie libre est correctement prise en compte. Les principales contraintes utilisées sont les suivantes ^[72]:

- Contraintes de distance harmonique entre le ligand et la protéine : ce type de contraintes maintient une distance fixe entre des atomes du ligand et des atomes de la protéine bien déterminés, tout au long des simulations de dynamiques moléculaire. Ils permettent au ligand d'explorer plusieurs modes de liaison au niveau d'un même site.
- Contraintes à fond plat (flat-bottomed) : ce type de contraintes agit d'une manière similaire aux contraintes de distance harmonique mais n'exercent pas leurs forces que lorsque le ligand dépasse une distance spécifique qu'on détermine. Par conséquent, ils permettent au ligand d'explorer plusieurs sites de liaison.

- Contraintes de Boresch : ce type de contraintes harmoniques est établi par Boresch et ses collaborateurs [73] où les six degrés de liberté régissant l'orientation du ligand par rapport à sa cible sont retenus. En fait, les contraintes de Boresch fixent une distance, deux angles et trois dièdres suite à la détermination de trois atomes du récepteur protéique et de trois atomes du ligand (Figure 37). Du coup, il restreint le ligand à un seul mode de liaison et donc à une orientation spécifique et ne lui permet pas d'explorer des possibilités de liaisons alternatives.

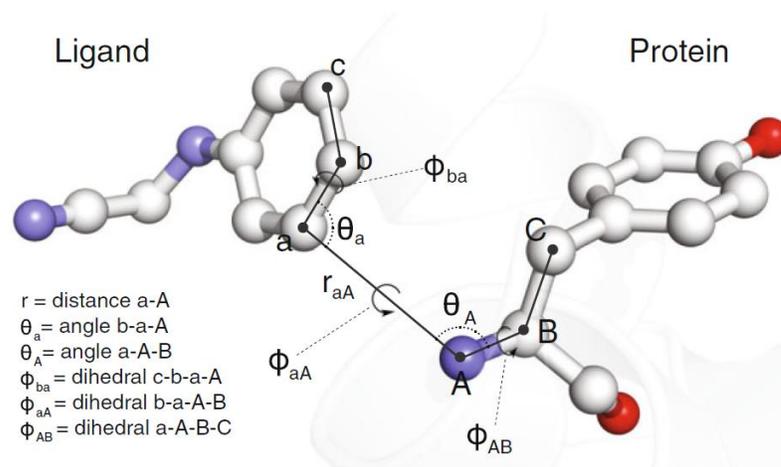


Figure 37 : Ensemble de contraintes proposé par Boresch et al pour les calculs d'énergie libre de liaison. Six atomes y sont impliqués ; atomes a, b et c au niveau du ligand et les atomes A, B et C appartenant à la protéine.

(Boresch et al, *The Journal of Physical Chemistry B*, vol 107, no. 35, August 2003, doi:10.1021/jp0217839)

De plus, l'utilisation des contraintes de Boresch fournit une solution analytique à la valeur de l'énergie libre de liaison, incluant la correction à l'état standard. Cette dernière sera alors représentée de la manière suivante :

$$k_B T \ln \left[\frac{8\pi^2 V^\circ}{r_{aA}^2 \cdot \sin \theta_a \cdot \sin \theta_A} \frac{(k_{r_{aA}} k_{\theta_a} k_{\theta_A} k_{\phi_{ba}} k_{\phi_{aA}} k_{\phi_{AB}})^{1/2}}{(2\pi k_B T)^2} \right]$$

Avec V^0 le volume standard de 1660 \AA^3 , r_{aA} est la valeur de référence choisie pour la contrainte de distance entre les deux atomes a et A , θ_a et θ_A sont les valeurs de référence pour les contraintes des angles de centre a et A respectivement et ϕ_{ba} , ϕ_{aA} et ϕ_{AB} sont les valeurs de référence pour les angles dièdres de centre ba , aA et AB . k présente la constante de force choisie pour les six contraintes harmoniques définies. k_B est la constante de Boltzmann qui, comme mentionné auparavant peut être remplacé par la constante des gaz parfaits R . Notons que Boresch et ses collaborateurs ont également démontré que la valeur de l'énergie libre totale ne dépend pas de la valeur de la constante de force choisie.

Le choix des atomes de référence pour n'importe quel type de contraintes est important. Ceux-ci doivent se trouver dans une partie rigide du ligand qui définit bien son orientation plutôt que dans une partie flexible [74].

En principe, quel que soit le type de contraintes appliquées, il ne doit pas affecter la valeur calculée de l'énergie libre tant que l'échantillonnage est adéquat et si l'effet de l'état standard est correctement géré.

C.III. Méthodes d'estimation des énergies libres des simulations à l'équilibre

Après avoir effectué plusieurs simulations de dynamique moléculaire, la prochaine étape serait d'analyser les données issues de ces simulations afin d'estimer la différence de l'énergie libre entre les divers états simulés. Différentes méthodes existent qui génèrent la valeur de l'énergie libre à partir de simulations à l'équilibre. On entend par simulations à l'équilibre, celles où le système subit une équilibration pour une durée déterminée suite à sa préparation et avant l'étape de production. Ils existent aussi d'autres méthodes d'estimation de ΔG à partir de simulations non

équilibrées dont la plus populaire est la méthode d'intégration thermodynamique (TI). Nous allons aborder et utiliser uniquement les méthodes à l'équilibre puisqu'elles sont plus pertinentes pour les voies alchimiques.

C.III.1. Approches de perturbation

C.III.1.1. Méthode FEP

La première approche d'estimation de la différence de l'énergie libre de liaison absolue est basée sur la théorie de perturbation introduite par Zwanzig en 1954 ^[75].

La méthode est appelée moyenne exponentielle (EXP, exponential averaging) ou plus communément perturbation d'énergie libre (FEP, Free Energy Perturbation).

Elle estime la valeur de ΔG entre deux étapes intermédiaires nommées ici « 0 » et « 1 » sous la forme suivante :

$$\Delta G_{0,1} = -k_B T \ln \left\langle e^{-\beta(U_1(x) - U_0(x))} \right\rangle_0$$

Selon FEP, la différence d'énergie libre est donc calculée comme le logarithme (ln) de la moyenne d'ensemble de l'exponentielle (e) de la différence d'énergie potentielle (U_1 et U_0) entre les deux états. La formule ci-dessus s'applique pour une « transformation en avant » impliquant le passage de l'état 0 à l'état 1. Il est possible de calculer la différence d'énergie libre pour « la transformation en arrière », impliquant le passage inverse de l'état 1 à l'état 0. On parle alors respectivement de « Forward FEP » et de « Backward FEP ». Notons que, selon la formule de FEP, l'échantillonnage à l'équilibre est effectué pour un état ; les simulations de dynamique se font pour l'état 0 et l'état 1 d'une manière indépendante mais les énergies seraient calculées pour les deux états sur les

mêmes configurations. La différence d'énergie potentielle est évaluée quant à elle pour le même ensemble de référence. De plus, même si les expressions de « Forward FEP » et « Backward FEP » seraient équivalentes, leurs propriétés de convergence peuvent différer, on aura alors une valeur de $\Delta G_{0 \rightarrow 1}$ différente de la valeur $\Delta G_{1 \rightarrow 0}$. Cette situation est connue sous le nom d'hystérésis qui engendre une mauvaise estimation de la valeur de l'énergie libre.

Bien que l'équation ci-dessus soit exacte, il reste quelques inconvénients à l'utilisation de cette méthode. Tout d'abord, FEP ne converge que lentement avec la quantité de données collectées et parfois même si les simulations convergent, un mauvais chevauchement entre les espaces de phase de deux états consécutifs peut avoir lieu ^[76]. D'où l'apparition de méthodes de perturbation alternatives à la méthode FEP pour l'analyse des données énergétiques.

C.III.1.2. Méthodes BAR et MBAR

Une autre méthode d'estimation de la différence d'énergie libre de liaison à partir de simulations à l'équilibre est la méthode BAR (Bennett Acceptance Ratio) établi par Bennett comme amélioration à la méthode FEP ^[77]. BAR est en effet plus robuste et plus optimale à estimer la valeur de ΔG à partir de deux états, puisqu'elle requiert moins de chevauchement entre les espaces de phase de deux états consécutifs pour converger les simulations et fournir une valeur exacte et correcte de ΔG ^[78]. Cependant, contrairement à la méthode FEP, BAR requiert un échantillonnage et une évaluation énergétique des configurations des deux états successifs afin d'estimer leur différence d'énergie libre. En effet, comme le chevauchement de l'espace des phases des simulations consécutives affecte la fiabilité de l'estimation, les différences d'énergie libre sont calculées dans le cas de

la méthode BAR en simulant plusieurs états intermédiaires autres que les deux états initial et final, dans le but d'augmenter le chevauchement entre les états voisins. La troisième méthode de perturbation qui estime la valeur de ΔG est la méthode MBAR (Multistate Bennett Acceptance Ratio) proposé par Shirts et ses collaborateurs ^[79]. MBAR est une extension de BAR, utilisée quand plusieurs états sont considérés en même temps. Dans cette méthode, une série de fonctions de pondération (weighting functions) sont dérivées. Ceci minimise les incertitudes dans les différences d'énergie libre entre tous les états considérés simultanément. En d'autres termes, MBAR se réduit à BAR lorsqu'uniquement deux états sont pris en compte.

Pour récapituler, chaque méthode de perturbation qui analyse les données des simulations de dynamique moléculaire à l'équilibre nécessite un type de données afin d'estimer correctement les différences d'énergie libres de liaison alchimique.

La méthode FEP nécessite $\Delta u_{k,k+1}(\vec{q})$ ou $\Delta u_{k,k-1}(\vec{q})$ qui représentent les fonctions des énergies potentielles entre deux états k/k+1 et k/k-1 respectivement, avec la première formule correspondant au « forward FEP » et la seconde au « backward FEP ». Tandis que la méthode BAR nécessite $\Delta u_{k,k+1}(\vec{q})$ et $\Delta u_{k,k-1}(\vec{q})$ qui correspondent aux fonctions d'énergie potentielles entre deux états voisins k/k+1 et k/k-1.

Il existe plusieurs points à prendre en considération afin de s'assurer de la validité de nos données. Tout d'abord, les estimations données par la méthode FEP pour les valeurs de « Forward » et « Backward » doivent être identiques et de signes opposés. Si les valeurs sont éloignées, un mauvais chevauchement de l'espace des

phases entre les états voisins existe alors, aboutissant à une estimation inexacte de l'énergie libre de liaison ce qui augmente le taux d'incertitude ^[80]. Une solution au problème de chevauchement serait d'augmenter le nombre d'états intermédiaires afin de faciliter le passage de l'état initial à l'état final, et ceci en ajoutant des valeurs supplémentaires de lambda. Par exemple, considérons les deux états caractérisés par les valeurs lambda suivantes : Etat 1 : $\lambda_{\text{vdW}} = 1$, $\lambda_{\text{électrostatique}} = 1$ et Etat 2 : $\lambda_{\text{vdW}} = 1$ et $\lambda_{\text{électrostatique}} = 0,9$. L'ajout d'un état supplémentaire afin d'améliorer le chevauchement entre les états 1 et 2 serait sous la forme suivante : Etat intermédiaire 1-2 : $\lambda_{\text{vdW}} = 1$, $\lambda_{\text{électrostatique}} = 0,95$.

Conclusions du chapitre III

Les principales méthodes que nous allons adopter dans nos travaux afin d'évaluer les nouvelles drogues inhibitrices issues de la structure du DTG sont les suivantes : tout d'abord on trouve les méthodes de DFT et d'ALMOEDA pour calculer les énergies d'interaction (ΔE) des drogues avec l'ADN viral. Ces techniques de calcul quantique post-HF se basent sur la structure électronique et sur les orbitales moléculaires respectivement, afin de permettre l'évaluation énergétique des molécules nouvellement conçues de manière exacte et précise. Une fois l'énergie d'interaction totale obtenue, nous allons procéder à sa décomposition en plusieurs composantes afin d'élucider celle qui contribue le plus au rendement énergétique observé. Cette démarche se fait via des méthodes interprétatives quantiques ; les méthodes RVS et ALMOEDA. Nous pourrons alors expliquer le comportement des nouvelles molécules vis-à-vis de leur cible. En plus des méthodes quantitatives, des méthodes d'analyse topologique comme la méthode MEP, générant les contours des potentiels électrostatiques seront adoptées. Cette dernière nous permettra de représenter la répartition de la population électronique au niveau des noyaux substitués du DTG et des nouvelles molécules par des régions colorées. Elle permet ainsi l'identification des régions de densité électronique faible et élevée en fonction des substituants ajoutés. Grâce à la méthode MEP, nous pourrons évaluer les effets des substitutions effectuées sur la répartition des électrons au niveau de nos molécules ainsi que sur les interactions qu'elles effectuent. Nous avons eu recours de même à des méthodes de mécanique moléculaire pour étudier nos systèmes d'intérêt. On nomme SIBFA comme champ de force polarisable utilisé dans nos travaux pour calculer l'énergie d'interaction des drogues avec l'ADN viral et pour effectuer aussi la décomposition de l'énergie d'interaction totale en plusieurs

contributions. Nous rechercherons alors un accord entre la démarche quantique et mécanique pour confirmer nos résultats énergétiques.

Des simulations de dynamique moléculaires ont été effectuées aussi sur les systèmes du VIH-1 afin d'observer leur évolution en fonction du temps et d'obtenir différentes propriétés macroscopiques du système à partir du comportement collectif de ses partenaires individuels. Ces simulations nous permettent de même de calculer l'énergie libre absolue de liaison (ΔG) de nos drogues entières à leur site d'action au niveau du complexe IN-ADNv et de comparer leur effet thérapeutique et leur efficacité par rapport au DTG. C'est par un cycle thermodynamique que les calculs de ΔG sont faits. Ce dernier comprend dans notre cas deux parties ; la première où le ligand se trouvant seul dans la boîte de simulation commence à perdre ses interactions intermoléculaires avec le solvant aqueux puis, une fois restreint par des contraintes notamment celles de Boresch, il commence à se situer au niveau de son site de liaison à l'interface de l'intasome du VIH-1 et à interagir avec lui. A noter que les contraintes de Boresch, contraintes les plus couramment utilisées, agissent en fixant une distance, deux angles et trois dièdres entre trois atomes du DTG et trois atomes de l'intasome viral. La perte ou le gain des interactions intermoléculaires au cours de nos calculs seraient évaluées ensuite par les méthodes d'estimation énergétique FEP et BAR qui est un estimateur plus efficace et moins biaisé que FEP. Ainsi, nous aurons exploité la majorité des voies convenables à l'évaluation de nos drogues inhibitrices, aux niveaux quantitatif et qualitatif.

Références bibliographiques

- [1] E. Schrödinger, "An Undulatory Theory of the Mechanics of Atoms and Molecules," *Phys. Rev.*, vol. 28, no. 6, pp. 1049–1070, Dec. 1926, doi: 10.1103/PhysRev.28.1049.
- [2] C. Froese Fischer, "General Hartree-Fock program," *Comput. Phys. Commun.*, vol. 43, no. 3, pp. 355–365, Feb. 1987, doi: 10.1016/0010-4655(87)90053-1.
- [3] P. Echenique and J. L. Alonso, "A mathematical and computational review of Hartree-Fock SCF methods in quantum chemistry," *Mol. Phys.*, vol. 105, no. 23–24, pp. 3057–3098, Dec. 2007, doi: 10.1080/00268970701757875.
- [4] "Post-Hartree-Fock Methods," in *Methods of Molecular Quantum Mechanics*, John Wiley & Sons, Ltd, 2009, pp. 133–139. doi: 10.1002/9780470684559.ch8.
- [5] "Introduction to Computational Chemistry, 2nd Edition | Wiley," *Wiley.com*. <https://www.wiley.com/en-us/Introduction+to+Computational+Chemistry%2C+2nd+Edition-p-9780470058046> (accessed Jul. 01, 2021).
- [6] C. David Sherrill and H. F. Schaefer, "The Configuration Interaction Method: Advances in Highly Correlated Approaches," in *Advances in Quantum Chemistry*, vol. 34, P.-O. Löwdin, J. R. Sabin, M. C. Zerner, and E. Brändas, Eds. Academic Press, 1999, pp. 143–269. doi: 10.1016/S0065-3276(08)60532-8.
- [7] Chr. Møller and M. S. Plesset, "Note on an Approximation Treatment for Many-Electron Systems," *Phys. Rev.*, vol. 46, no. 7, pp. 618–622, Oct. 1934, doi: 10.1103/PhysRev.46.618.
- [8] W. Kohn, A. D. Becke, and R. G. Parr, "Density Functional Theory of Electronic Structure," *J. Phys. Chem.*, vol. 100, no. 31, pp. 12974–12980, Jan. 1996, doi: 10.1021/jp960669l.
- [9] K. Burke, "Perspective on density functional theory," *J. Chem. Phys.*, vol. 136, no. 15, p. 150901, Apr. 2012, doi: 10.1063/1.4704546.
- [10] R. A. Kendall, T. H. Dunning, and R. J. Harrison, "Electron affinities of the first-row atoms revisited. Systematic basis sets and wave functions," *J. Chem. Phys.*, vol. 96, no. 9, pp. 6796–6806, May 1992, doi: 10.1063/1.462569.
- [11] S. Akbudak, "Selection of an Appropriate Basis Set for Accurate Description of Binding Energy: A First Principles Study," *Süleyman Demirel Üniversitesi Fen Bilim. Enstitüsü Derg.*, vol. 21, p. 10, Nov. 2017, doi: 10.19113/sdufbed.76392.
- [12] T. Van Mourik, A. K. Wilson, K. A. Peterson, D. E. Woon, and T. H. Dunning, "The Effect of Basis Set Superposition Error (BSSE) on the Convergence of Molecular Properties Calculated with the Correlation Consistent Basis Sets," in *Advances in Quantum Chemistry*, vol. 31, J. R. Sabin, M. C. Zerner, E. Brändas, S. Wilson, J. Maruani, Y. G. Smeyers, P. J. Grout, and R. McWeeny, Eds. Academic Press, 1998, pp. 105–135. doi: 10.1016/S0065-3276(08)60185-9.
- [13] Gaussian 09, Revision A.02, M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, Gaussian, Inc., Wallingford CT, 2016.

- [14] Schmidt, M. W.; Baldrige, K. K.; Boatz, J. A.; Elbert, S. T.; Gordon, M. S.; Jensen, J. H.; Koseki, S.; Matsunaga, N.; Nguyen, K. A.; Su, S.; Windus, T. L.; Dupuis, M.; Montgomery, J. A. General Atomic and Molecular Electronic Structure System *J. Comput. Chem.* 1993, 14: 1347- 1363.
- [15] K. Morokuma and K. Kitaura, "Energy Decomposition Analysis of Molecular Interactions," in *Chemical Applications of Atomic and Molecular Electrostatic Potentials: Reactivity, Structure, Scattering, and Energetics of Organic, Inorganic, and Biological Systems*, P. Politzer and D. G. Truhlar, Eds. Boston, MA: Springer US, 1981, pp. 215–242. doi: 10.1007/978-1-4757-9634-6_10.
- [16] K. Kitaura and K. Morokuma, "A new energy decomposition scheme for molecular interactions within the Hartree-Fock approximation," *Int. J. Quantum Chem.*, vol. 10, no. 2, pp. 325–340, 1976, doi: 10.1002/qua.560100211.
- [17] N. C. Hernández, C. Zicovich-Wilson, and J. Sanz, "The constrained space orbital variation analysis for periodic ab initio calculations.," *J. Chem. Phys.*, 2006, doi: 10.1063/1.2198528
- [18] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, and W. Yang, "Revealing Noncovalent Interactions," *J. Am. Chem. Soc.*, vol. 132, no. 18, pp. 6498–6506, May 2010, doi: 10.1021/ja100936w.
- [19] P. Fuentealba, E. Chamorro, and J. C. Santos, "Chapter 5 Understanding and using the electron localization function," in *Theoretical and Computational Chemistry*, vol. 19, A. Toro-Labbé, Ed. Elsevier, 2007, pp. 57–85. doi: 10.1016/S1380-7323(07)80006-9.
- [20] P. Politzer, P. R. Laurence, and K. Jayasuriya, "Molecular electrostatic potentials: an effective tool for the elucidation of biochemical phenomena.," *Environ. Health Perspect.*, vol. 61, pp. 191–202, Sep. 1985.
- [21] M. J. S. Phipps, T. Fox, C. S. Tautermann, and C.-K. Skylaris, "Energy decomposition analysis approaches and their evaluation on prototypical protein–drug interaction patterns," *Chem. Soc. Rev.*, vol. 44, no. 10, pp. 3177–3211, 2015, doi: 10.1039/C4CS00375F.
- [22] R. Z. Khaliullin, E. A. Cobar, R. C. Lochan, A. T. Bell, and M. Head-Gordon, "Unravelling the origin of intermolecular interactions using absolutely localized molecular orbitals," *J. Phys. Chem. A*, vol. 111, no. 36, pp. 8753–8765, Sep. 2007, doi: 10.1021/jp073685z.
- [23] M. J. S. Phipps, T. Fox, C. S. Tautermann, and C.-K. Skylaris, "Energy Decomposition Analysis Based on Absolutely Localized Molecular Orbitals for Large-Scale Density Functional Theory Calculations in Drug Design," *J. Chem. Theory Comput.*, vol. 12, no. 7, pp. 3135–3148, Jul. 2016, doi: 10.1021/acs.jctc.6b00272.
- [24] O. Andreussi, I. G. Prandi, M. Campetella, G. Prampolini, and B. Mennucci, "Classical Force Fields Tailored for QM Applications: Is It Really a Feasible Strategy?," *J. Chem. Theory Comput.*, vol. 13, no. 10, pp. 4636–4648, Oct. 2017, doi: 10.1021/acs.jctc.7b00777.
- [25] L. Kalé *et al.*, "NAMD2: Greater Scalability for Parallel Molecular Dynamics," *J. Comput. Phys.*, vol. 151, no. 1, pp. 283–312, May 1999, doi: 10.1006/jcph.1999.6201.
- [26] D. A. Case *et al.*, "The Amber biomolecular simulation programs," *J. Comput. Chem.*, vol. 26, no. 16, pp. 1668–1688, Dec. 2005, doi: 10.1002/jcc.20290.
- [27] B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations," *J. Comput. Chem.*, vol. 4, no. 2, pp. 187–217, 1983, doi: 10.1002/jcc.540040211.
- [28] A. D. MacKerell *et al.*, "CHARMM: The Energy Function and Its Parameterization," in *Encyclopedia of Computational Chemistry*, American Cancer Society, 2002. doi: 10.1002/0470845015.cfa007.
- [29] W. R. P. Scott *et al.*, "The GROMOS Biomolecular Simulation Program Package," *J. Phys. Chem. A*, vol. 103, no. 19, pp. 3596–3607, May 1999, doi: 10.1021/jp984217f.

- [30] N. Gresh, P. Claverie, and A. Pullman, "Theoretical studies of molecular conformation. Derivation of an additive procedure for the computation of intramolecular interaction energies. Comparison with ab initio SCF computations," *Theor. Chim. Acta*, vol. 66, no. 1, pp. 1–20, Jan. 1984, doi: 10.1007/BF00577135.
- [31] N. Gresh, G. A. Cisneros, T. A. Darden, and J.-P. Piquemal, "Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand–Macromolecule Complexes. A Bottom-Up Strategy," *J. Chem. Theory Comput.*, vol. 3, no. 6, pp. 1960–1986, Nov. 2007, doi: 10.1021/ct700134r.
- [32] J. W. Ponder *et al.*, "Current Status of the AMOEBA Polarizable Force Field," *J. Phys. Chem. B*, vol. 114, no. 8, pp. 2549–2564, Mar. 2010, doi: 10.1021/jp910674d.
- [33] J. C. Wu, G. Chattree, and P. Ren, "Automation of AMOEBA polarizable force field parameterization for small molecules," *Theor. Chem. Acc.*, vol. 131, no. 3, p. 1138, Feb. 2012, doi: 10.1007/s00214-012-1138-6.
- [34] W. L. Jorgensen, "Special Issue on Polarization," *J. Chem. Theory Comput.*, vol. 3, no. 6, pp. 1877–1877, Nov. 2007, doi: 10.1021/ct700252g.
- [35] E. Harder, B. Kim, R. A. Friesner, and B. J. Berne, "Efficient Simulation Method for Polarizable Protein Force Fields: Application to the Simulation of BPTI in Liquid Water," *J. Chem. Theory Comput.*, vol. 1, no. 1, pp. 169–180, Jan. 2005, doi: 10.1021/ct049914s.
- [36] F. Archambault, C. Chipot, I. Soteras, F. J. Luque, K. Schulten, and F. Dehez, "Polarizable Intermolecular Potentials for Water and Benzene Interacting with Halide and Metal Ions," *J. Chem. Theory Comput.*, vol. 5, no. 11, pp. 3022–3031, Nov. 2009, doi: 10.1021/ct9004189.
- [37] N. Gresh, "Development, validation, and applications of anisotropic polarizable molecular mechanics to study ligand and drug-receptor interactions," *Curr. Pharm. Des.*, vol. 12, no. 17, pp. 2121–2158, 2006, doi: 10.2174/138161206777585256.
- [38] F. Vigné-Maeder and P. Claverie, "The exact multicenter multipolar part of a molecular charge distribution and its simplified representations," *J. Chem. Phys.*, vol. 88, no. 8, pp. 4934–4948, Apr. 1988, doi: 10.1063/1.454705.
- [39] J.-P. Piquemal, N. Gresh, and C. Giessner-Prettre, "Improved Formulas for the Calculation of the Electrostatic Contribution to the Intermolecular Interaction Energy from Multipolar Expansion of the Electronic Distribution," *J. Phys. Chem. A*, vol. 107, no. 48, pp. 10353–10359, Dec. 2003, doi: 10.1021/jp035748t.
- [40] J.-P. Piquemal, H. Chevreaux, and N. Gresh, "Toward a Separate Reproduction of the Contributions to the Hartree–Fock and DFT Intermolecular Interaction Energies by Polarizable Molecular Mechanics with the SIBFA Potential," *J. Chem. Theory Comput.*, vol. 3, no. 3, pp. 824–837, May 2007, doi: 10.1021/ct7000182.
- [41] S. Creuzet, J. Langlet, and N. Gresh, "Adjustment of the SIBFA Method for Potential Maps to Study Hydrogen Bonding Vibrational Frequencies," *AIP Conf. Proc.*, vol. 239, no. 1, pp. 114–114, Oct. 1991, doi: 10.1063/1.41343.
- [42] N. Gresh, M. Leboeuf, and D. Salahub, "Energetics and Structure in Model Neutral, Anionic, and Cationic Hydrogen-Bonded Complexes" *Modeling the Hydrogen Bond.*, chap 6, pp. 82–112, May 5 1994, doi: 10.1021/bk-1994-0569.ch006.
- [43] J.-P. Piquemal, R. Chelli, P. Procacci, and N. Gresh, "Key Role of the Polarization Anisotropy of Water in Modeling Classical Polarizable Force Fields," *J. Phys. Chem. A*, vol. 111, no. 33, pp. 8170–8176, Aug. 2007, doi: 10.1021/jp072687g.
- [44] D. R. Garmer and W. J. Stevens, "Transferability of molecular distributed polarizabilities from a simple localized orbital based method," *J. Phys. Chem.*, vol. 93, no. 25, pp. 8263–8270, Dec. 1989, doi: 10.1021/j100362a023.

- [45] J. N. Murrell, M. Randić, D. R. Williams, and H. C. Longuet-Higgins, "The theory of intermolecular forces in the region of small orbital overlap," *Proc. R. Soc. Lond. Ser. Math. Phys. Sci.*, vol. 284, no. 1399, pp. 566–581, Mar. 1965, doi: 10.1098/rspa.1965.0081.
- [46] P. Ren and J. W. Ponder, "Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations," *J. Comput. Chem.*, vol. 23, no. 16, pp. 1497–1506, 2002, doi: 10.1002/jcc.10127.
- [47] P. Ren and J. W. Ponder, "Temperature and Pressure Dependence of the AMOEBA Water Model," *J. Phys. Chem. B*, vol. 108, no. 35, pp. 13427–13437, Sep. 2004, doi: 10.1021/jp0484332.
- [48] L. Lagardère *et al.*, "Tinker-HP: a Massively Parallel Molecular Dynamics Package for Multiscale Simulations of Large Complex Systems with Advanced Point Dipole Polarizable Force Fields," *Chem. Sci.*, Nov. 2017, doi: 10.1039/C7SC04531J.
- [49] T. A. Halgren, "The representation of van der Waals (vdW) interactions in molecular mechanics force fields: potential form, combination rules, and vdW parameters," *J. Am. Chem. Soc.*, vol. 114, no. 20, pp. 7827–7843, Sep. 1992, doi: 10.1021/ja00046a032.
- [50] A. J. Stone, "Distributed multipole analysis, or how to describe a molecular charge distribution," *Chem. Phys. Lett.*, vol. 83, no. 2, pp. 233–239, Oct. 1981, doi: 10.1016/0009-2614(81)85452-8.
- [51] A. J. Stone, "Distributed Multipole Analysis: Stability for Large Basis Sets," *J. Chem. Theory Comput.*, vol. 1, no. 6, pp. 1128–1132, Nov. 2005, doi: 10.1021/ct050190+.
- [52] B. T. Thole, "Molecular polarizabilities calculated with a modified dipole interaction," *Chem. Phys.*, vol. 59, no. 3, pp. 341–350, Aug. 1981, doi: 10.1016/0301-0104(81)85176-2.
- [53] J. A. McCammon and S. C. Harvey, *Dynamics of Proteins and Nucleic Acids*. Cambridge: Cambridge University Press, 1987. doi: 10.1017/CBO9781139167864.
- [54] F. Lado, "Some topics in the molecular dynamics ensemble," *J. Chem. Phys.*, vol. 75, no. 11, p. 5461, Jun. 1998, doi: 10.1063/1.441948.
- [55] F. Radjai, "Multi-periodic boundary conditions and the Contact Dynamics method," *Comptes Rendus Mécanique*, vol. 346, no. 3, pp. 263–277, Mar. 2018, doi: 10.1016/j.crme.2017.12.007.
- [56] D. C. Rapaport and R. Rapaport Dennis C., *The Art of Molecular Dynamics Simulation*. Cambridge University Press, 2004.
- [57] J. C. Wu, G. Chattree, and P. Ren, "Automation of AMOEBA polarizable force field parameterization for small molecules," *Theor. Chem. Acc.*, vol. 131, no. 3, p. 1138, Feb. 2012, doi: 10.1007/s00214-012-1138-6.
- [58] L. Lagardère, F. Aviat, and J.-P. Piquemal, "Pushing the Limits of Multiple-Time-Step Strategies for Polarizable Point Dipole Molecular Dynamics," *J. Phys. Chem. Lett.*, vol. 10, no. 10, pp. 2593–2599, May 2019, doi: 10.1021/acs.jpcllett.9b00901.
- [59] H. C. Andersen, "Molecular dynamics simulations at constant pressure and/or temperature," *J. Chem. Phys.*, vol. 72, no. 4, pp. 2384–2393, Feb. 1980, doi: 10.1063/1.439486.
- [60] G. Bussi, D. Donadio, and M. Parrinello, "Canonical sampling through velocity rescaling," *J. Chem. Phys.*, vol. 126, no. 1, p. 014101, Jan. 2007, doi: 10.1063/1.2408420.
- [61] Y. Lin, D. Pan, J. Li, L. Zhang, and X. Shao, "Application of Berendsen barostat in dissipative particle dynamics for nonequilibrium dynamic simulation," *J. Chem. Phys.*, vol. 146, no. 12, p. 124108, Mar. 2017, doi: 10.1063/1.4978807.
- [62] S. Hare *et al.*, "Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572)," *Mol. Pharmacol.*, vol. 80, no. 4, pp. 565–572, Oct. 2011, doi: 10.1124/mol.111.073189.
- [63] C. Chipot, "Frontiers in free-energy calculations of biological systems," *WIREs Comput. Mol. Sci.*, vol. 4, no. 1, pp. 71–89, 2014, doi: 10.1002/wcms.1157.

- [64] M. R. Shirts, D. L. Mobley and J. D. Chodera “Chapter 4 Alchemical Free Energy Calculations: Ready for Prime Time?” *Elsevier.*, vol. 3, pp. 41–59, 2007, doi: 10.1016/S1574-1400(07)03004-6.
- [65] R. Salomon-Ferrer, A. W. Götz, D. Poole, S. Le Grand, and R. C. Walker, “Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald,” *J. Chem. Theory Comput.*, vol. 9, no. 9, pp. 3878–3888, Sep. 2013, doi: 10.1021/ct400314y.
- [66] D. L. Mobley and K. A. Dill, “Binding of Small-Molecule Ligands to Proteins: ‘What You See’ Is Not Always ‘What You Get,’” *Struct. Lond. Engl.* 1993, vol. 17, no. 4, pp. 489–498, Apr. 2009, doi: 10.1016/j.str.2009.02.010.
- [67] V. Gapsys, D. Seeliger, and B. L. de Groot, “New Soft-Core Potential Function for Molecular Dynamics Based Alchemical Free Energy Calculations,” *J. Chem. Theory Comput.*, vol. 8, no. 7, pp. 2373–2382, Jul. 2012, doi: 10.1021/ct300220p.
- [68] T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, “Avoiding singularities and numerical instabilities in free energy calculations based on molecular simulations,” *Chem. Phys. Lett.*, vol. 222, no. 6, pp. 529–539, Jun. 1994, doi: 10.1016/0009-2614(94)00397-1.
- [69] J. W. Pitera and W. F. van Gunsteren, “A Comparison of Non-Bonded Scaling Approaches for Free Energy Calculations,” *Mol. Simul.*, vol. 28, no. 1–2, pp. 45–65, Jan. 2002, doi: 10.1080/08927020211973.
- [70] M. R. Shirts, “Best practices in free energy calculations for drug design,” *Methods Mol. Biol. Clifton NJ*, vol. 819, pp. 425–467, 2012, doi: 10.1007/978-1-61779-465-0_26.
- [71] M. Aldeghi, A. Heifetz, M. J. Bodkin, S. Knapp, and P. C. Biggin, “Predictions of Ligand Selectivity from Absolute Binding Free Energy Calculations,” *J. Am. Chem. Soc.*, vol. 139, no. 2, pp. 946–957, Jan. 2017, doi: 10.1021/jacs.6b11467.
- [72] D. L. Mobley, J. D. Chodera, and K. A. Dill, “On the use of orientational restraints and symmetry corrections in alchemical free energy calculations,” *J. Chem. Phys.*, vol. 125, no. 8, p. 084902, Aug. 2006, doi: 10.1063/1.2221683.
- [73] S. Boresch, and F. Tettinger, M. Leitgeb, and M. Karplus, “Absolute Binding Free Energies: A Quantitative Approach for Their Calculation,” 2003, doi: 10.1021/JP0217839.
- [74] A. S. J. S. Mey *et al.*, “Best Practices for Alchemical Free Energy Calculations,” *Living J. Comput. Mol. Sci.*, vol. 2, no. 1, 2020, doi: 10.33011/livecoms.2.1.18378.
- [75] N. Lu, J. K. Singh, and D. A. Kofke, “Appropriate methods to combine forward and reverse free-energy perturbation averages,” *J. Chem. Phys.*, vol. 118, no. 7, pp. 2977–2984, Feb. 2003, doi: 10.1063/1.1537241.
- [76] M. R. Shirts, D. L. Mobley, and S. P. Brown, “Free-energy calculations in structure-based drug design,” in *Drug Design: Structure- and Ligand-Based Approaches*, C. H. Reynolds, D. Ringe, and J. Merz Kenneth M., Eds. Cambridge: Cambridge University Press, 2010, pp. 61–86. doi: 10.1017/CBO9780511730412.007.
- [77] M. R. Shirts and V. S. Pande, “Comparison of efficiency and bias of free energies computed by exponential averaging, the Bennett acceptance ratio, and thermodynamic integration,” *J. Chem. Phys.*, vol. 122, no. 14, p. 144107, Apr. 2005, doi: 10.1063/1.1873592.
- [78] H. Paliwal and M. R. Shirts, “A Benchmark Test Set for Alchemical Free Energy Transformations and Its Use to Quantify Error in Common Free Energy Methods,” *J. Chem. Theory Comput.*, vol. 7, no. 12, pp. 4115–4134, Dec. 2011, doi: 10.1021/ct2003995.
- [79] M. R. Shirts and J. D. Chodera, “Statistically optimal analysis of samples from multiple equilibrium states,” *J. Chem. Phys.*, vol. 129, no. 12, p. 124105, Sep. 2008, doi: 10.1063/1.2978177.

[80] P. V. Klimovich, M. R. Shirts, and D. L. Mobley, “Guidelines for the analysis of free energy calculations,” *J. Comput. Aided Mol. Des.*, vol. 29, no. 5, pp. 397–411, May 2015, doi: 10.1007/s10822-015-9840-9.

Chapitre IV

Résultats et Discussion

Chapitre IV

Résultats et Discussion

Partie A : Conception et évaluation des nouveaux noyaux d'halobenzènes issus du DTG

A.I. Introduction générale de la partie A

Le premier but que nous désirons atteindre dans nos travaux de thèse est la conception d'inhibiteurs anti-intégrase plus puissants et plus spécifiques que les INSTIs préexistants sur le marché puisque ces derniers confrontent le problème des mutations de résistance développées au sein de l'intégrase. Ces inhibiteurs, agissent à l'interface ADN viral – Intégrase ; ils effectuent alors des interactions avec ces deux partenaires, principalement avec les résidus du site actif de l'intégrase ainsi qu'avec les nucléotides terminaux de l'extrémité 3' réactive de l'ADN du VIH-1. Pour cela il paraît nécessaire, afin de surmonter le problème des mutations de résistance associées à l'administration des INSTIs, d'affaiblir les interactions de ces derniers avec l'intégrase tout en augmentant leur affinité à l'ADN viral.

Notre stratégie de travail s'est basée sur l'optimisation du DTG, inhibiteur le moins sensible à la résistance, en se focalisant spécifiquement sur son noyau halogéné (Figure 38) qui opère un rôle d'ancrage essentiel au niveau de l'extrémité de l'ADN viral. En effet, comme mentionné précédemment, les cycles aromatiques halogénés des inhibiteurs de transfert de brin s'empilent sous la cytosine C₁₆ de l'ADN viral processé effectuant avec elle des interactions π - π , et interagissent simultanément avec les noyaux aromatiques riches en électrons de la guanine G₄ du brin opposé. Le noyau halogéné du DTG constitue donc un candidat idéal pour y effectuer des substitutions afin d'améliorer son efficacité thérapeutique.

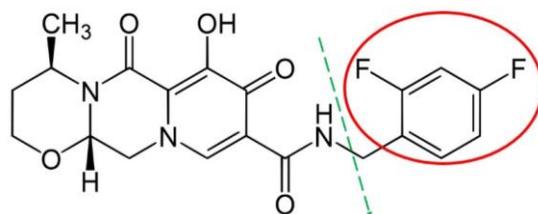


Figure 38 : Représentation de la structure du Dolutégravir, avec entouré en rouge le noyau d'halobenzène à modifier

Nous aborderons cette partie par la présentation des nouveaux composés halogénés conçus à partir du noyau du DTG, suite aux substitutions que nous avons effectuées. Puis nous passerons à l'évaluation de l'interaction de ces nouvelles molécules avec les bases terminales C₁₆ et G₄ de l'ADN viral par des calculs de l'énergie d'interaction (ΔE). Ces calculs se basent sur des méthodes de chimie théorique et de mécanique moléculaire ; les premières calculent l'énergie d'interaction dans le cadre de la théorie de la fonctionnelle de la densité (DFT) et des orbitales moléculaires localisées (ALMOEDA) alors que les secondes ont recours au champ de force polarisable SIBFA, formulé sur la base de la chimie quantique *ab-initio*. Ces ΔE calculées ont été ensuite décomposées par les méthodes RVS et SIBFA afin d'évaluer les contributions des différentes composantes énergétiques à la valeur de ΔE . Finalement, une évaluation qualitative de l'effet de nos substitutions a été entamée par la génération des contours de potentiel électrostatique pour chaque noyau, permettant la visualisation de la répartition de ses électrons. L'ensemble de ce travail a fait l'objet d'une publication parue en juin 2020 dans le journal « Frontiers in Chemistry », que nous présenterons en annexe à la page 283.

A.II. Conception de nouveaux noyaux halogénés à partir du DTG

Nous avons effectué des substitutions au niveau du noyau halogéné du DTG dans le but de renforcer ses interactions avec l'ADN viral du VIH-1. Les coordonnées du noyau halogéné du DTG sont issues du cristal de l'intasome du PFV en présence du DTG et des ions magnésium (code PDB : 3S3M).

Comme le montre la Figure 39, DTG possède deux atomes de fluor (F) en guise d'halogènes au niveau de son noyau. Les deux bases terminales interagissant fortement avec le noyau du DTG sont les suivantes (Figure 39) : la cytosine en position 16 en 3' du brin d'ADNv clivé, qui effectue des empilements π - π avec le noyau halogéné du DTG. Pour rappel, C₁₆ est reconnue par l'intégrase afin qu'il y ait élimination du dinucléotide GT situé en aval. La deuxième base est la guanine en position 4 complémentaire à C₁₆ et située en 5' du brin non clivé, vers laquelle l'atome de fluor du DTG pointe et interagit par des interactions électrostatiques avec son noyau riche en électrons (Figure 39). En effet, cette interaction est possible dû à la présence d'un trou sigma au niveau du prolongement de la liaison C-F. Ce trou est une caractéristique des liaisons C-X (X = halogène). Il se manifeste par une déplétion électronique le long de la liaison à cause de l'effet électroattracteur de l'halogène et par une région riche en électrons autour de l'halogène ^[1]. En fait, ce dernier (le fluor dans le cas du DTG) présente une électronégativité plus élevée par rapport au carbone. Du coup, il attire les électrons vers lui dans le cadre de sa liaison avec ce dernier, entraînant une déplétion de charges le long de la liaison et créant une région de densité électronique élevée autour de lui. C'est donc le prolongement de la liaison C-F du DTG, pauvre en

électrons et pointant vers le noyau riche en électrons de G₄, qui favorise l'interaction électrostatique entre la drogue et la base virale du VIH-1.

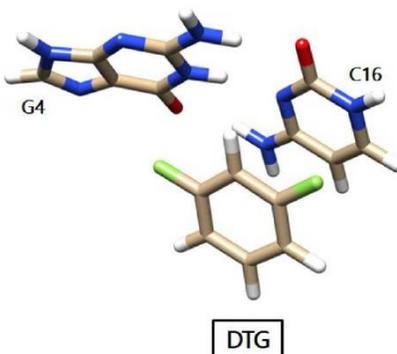


Figure 39 : Représentation des positions du noyau halogéné du Dolutégravir (DTG) et des noyaux des bases terminales virales C₁₆ et G₄, selon la structure cristallographique de l'intasome du PFV en présence du DTG et des ions Mg²⁺ (code PDB : 3S3M). Le noyau DTG est en interaction avec ceux de C₁₆ et G₄, via des empilements π - π et des interactions électrostatiques respectivement.

Code couleur : beige = atome de carbone / blanc = atome d'hydrogène / bleu = atome d'azote / rouge = atome d'oxygène / vert clair = atome de fluor

Ainsi, nous avons cherché à optimiser ces deux types d'interaction. Afin de renforcer l'empilement π - π entre le noyau d'halobenzène du DTG et C₁₆, nous avons remplacé le fluor concerné par d'autres halogènes tels que le brome (Br) et le chlore (Cl) ou par des groupements attracteurs d'électrons tels que le trifluorométhyle (CF₃) ou le cyanure (C triple liaison N). Nous avons choisi ces substitutions puisqu'elles ont le potentiel de modifier la répartition électronique du noyau par leur attraction des électrons et d'influencer par la suite les interactions des noyaux halogénés avec C₁₆. Quant au renforcement de l'interaction des noyaux avec G₄, nous avons trouvé plus instructif de remplacer le fluor qui interagit avec la guanine via son trou sigma, par des groupements donneurs d'électrons. Ces derniers vont conserver la présence d'une région pauvre en électrons à proximité de G₄, remplaçant le trou sigma du fluor et favorisant les interactions électrostatiques de ces groupements avec le noyau riche en électrons

de G₄. Ces groupements électropositifs sont NHCH₃ et NH₂. Les substitutions ont été effectuées *in silico* avec le logiciel Chimera 1.11.2 [2]. Nous avons obtenu alors treize nouveaux noyaux substitués que nous avons répartis en quatre groupes selon le type et le nombre de substitutions qui les caractérisent. Notons que la position des substitutions (ortho, méta et para) est par référence à l'hydrogène lié au reste de la drogue. Il est entouré en rouge dans la structure du DTG de la Figure 40. Les quatre groupes présentant nos treize composés sont représentés à la Figure 40. Ils sont les suivants :

- Groupe 1 : Les noyaux halogénés de ce groupe présentent deux halogènes en positions ortho et para. Nous trouvons le Dolutégravir avec 2 fluors (F) et le composé A1 avec un fluor en ortho et un chlore (Cl) en para
- Groupe 2 : Les noyaux halogénés de ce groupe possèdent un fluor en ortho, un groupement NHCH₃ en para et un troisième groupement électroattracteur. A2, B2, C2 et E2 possèdent ce groupement au niveau de la deuxième position ortho, située de l'autre côté (à gauche) de l'hydrogène pris comme référence. Nous trouvons le brome (Br) pour A2, le chlore pour B2, le CF₃ pour C2 et le cyanure pour E2. Quant au composé D2, il possède un chlore comme groupement électroattracteur, en méta par rapport à l'hydrogène de référence.
- Groupe 3 : Les noyaux halogénés de ce groupe possèdent comme substitutions un groupement NHCH₃ en para ainsi qu'un halogène. Il est en position ortho pour les composés A3, B3 et C3 et correspond respectivement au chlore, brome et fluor. Alors que cet halogène est en position méta pour le composé D3 et correspond au fluor.

- Groupe 4 : Les noyaux halogénés que nous avons conçus dans ce groupe ont un groupement NH₂ en para en présence d'un halogène ou deux. Nous trouvons le fluor en ortho pour A4, le fluor en méta pour B4 et deux atomes de fluor en ortho et méta pour C4.

Les choix de NHCH₃ et NH₂ comme groupements donneurs d'électrons et du fluor, brome et CF₃ comme groupements attracteurs d'électrons ont été inspirés d'une étude antérieurement réalisée au sein de l'équipe en 2014 par El Hage et al [3] sur des substitutions au niveau du noyau d'Elvitégravir (EVG) et leur effet sur son interaction avec les bases virales de l'ADN processé où CF₃ et NHCH₃ se sont avérés être respectivement, les meilleurs groupements attracteur et donneur d'électrons. Cette étude s'est basée sur les propriétés et les effets du trou sigma sur l'interaction de l'halogène avec les bases de l'ADN viral.

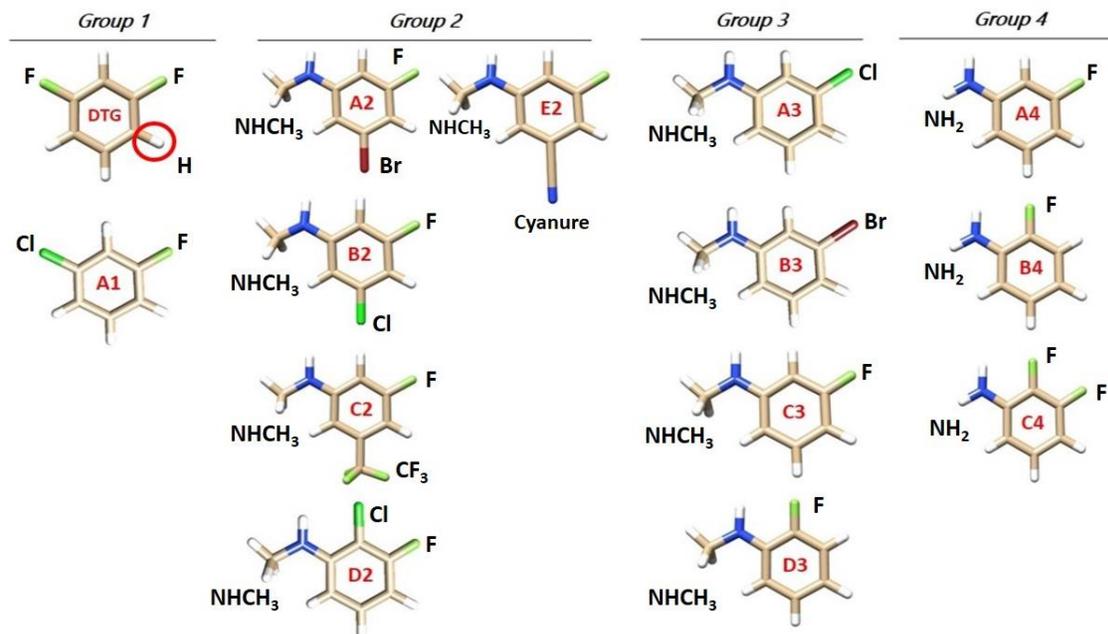


Figure 40 : Structures moléculaires du noyau halogéné du Dolutégravir (DTG) et des nouveaux noyaux que nous avons conçus avec le logiciel Chimera 1.11.2, réparties en quatre groupes selon les substitutions effectuées. Le code couleur est le même que celui indiqué à la Figure 39, avec en plus le chlore en vert foncé. Les substitutions ajoutées sont en noir et gras, indiquées par leurs symboles mentionnés dans le texte.

A.III. Evaluation de l'énergie d'interaction des nouveaux noyaux halogénés avec les bases terminales de l'ADN viral du VIH-1

Une fois les dérivés halogénés du DTG obtenus, l'étape suivante dans nos travaux de thèse, fut d'évaluer leurs énergies d'interaction avec les noyaux des bases C₁₆ et G₄, dans le but d'explorer si les substitutions effectuées contribueraient à une meilleure interaction entre l'inhibiteur et les bases virales ?

A cette fin, nous avons calculé l'énergie d'interaction du mini-complexe ternaire formé par le noyau de C₁₆, le noyau de G₄ et le noyau halogéné de nos composés (y inclus le DTG), par des méthodes quantiques et des méthodes de mécanique moléculaire, espérant en premier lieu trouver un accord entre les deux types de méthodes, confirmant alors la validité de nos résultats, et prouver ensuite l'efficacité de nos dérivés. La Figure 41 présente un nombre des mini-complexes étudiés. Nous y trouvons le noyau de C₁₆, le noyau de G₄ en présence d'un noyau halogéné représentatif de chaque groupe ; le DTG du groupe 1, les composés C2 et E2 du groupe 2, le composé B3 du groupe 3 et le composé C4 du groupe 4. Similairement au DTG, nos nouveaux noyaux présentent la même position vis-à-vis de C₁₆ et G₄, effectuant des empilements π - π avec la première base et des interactions électrostatiques via leurs groupements NHCH₃ et NH₂ avec la deuxième base.

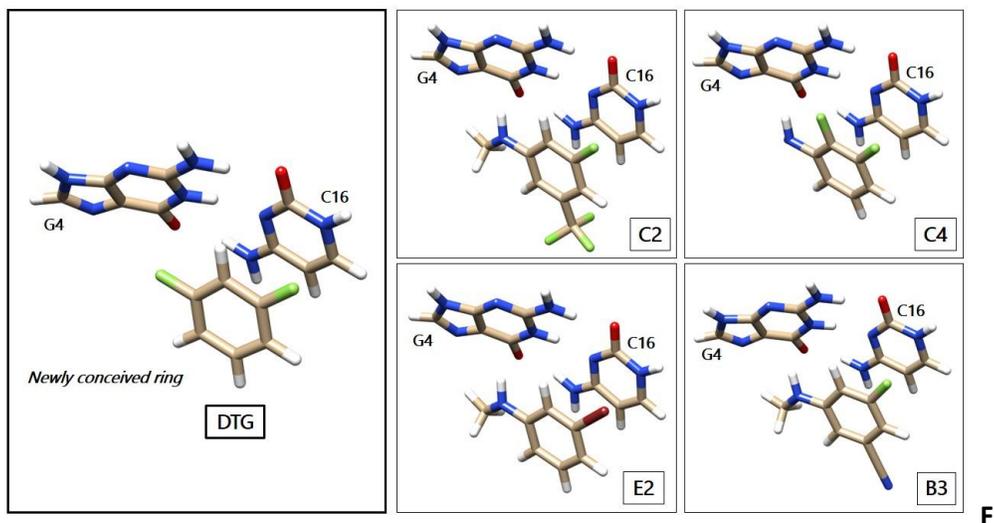


Figure 41 : Représentation de certains complexes étudiés, formés par le noyau C₁₆ – noyau G₄ – noyau halogéné, visualisés par le logiciel Chimera1.11.2. Les coordonnées des noyaux de C₁₆, G₄ et du Dolutégravir (DTG) sont issues de la structure PDB 3S3M, correspondant à l’intasome du Prototype Foamy Virus (PFV) en présence du Dolutégravir et des ions Mg²⁺. Le code couleur pour les atomes reste le même que celui mentionné pour les figures 39 et 40.

A.III.1. Calculs mécanique et quantiques corrélés de l’énergie d’interaction totale du complexe C₁₆-G₄-nouveau noyau

Comme déjà mentionné, les coordonnées des atomes constituant les noyaux de C₁₆, G₄ et du DTG sont issues de la structure PDB 3S3M. Les coordonnées de nos nouveaux dérivés sont identiques à celles du DTG. Pour les atomes nouvellement ajoutés, leurs coordonnées sont directement données par Chimera suite à leur génération. Avant de commencer tout calcul énergétique, nous avons effectué une optimisation de la géométrie de nos structures à l’aide du logiciel GAUSSIAN09 en utilisant la base cc-PVTZ (-f). Durant l’étape d’optimisation, les bases virales C₁₆ et G₄ ont été figées ainsi que l’hydrogène connecté au fragment dicétonique de la drogue dans le but de mimer et de conserver leurs positions à l’image de celles en présence de l’intégrase virale dans la structure PDB. Les atomes restants au niveau des noyaux modifiés, ont été relaxés afin de leur permettre de trouver la position optimale au sein du complexe C₁₆-G₄. L’étape de calcul permettrait ensuite de

déterminer l'énergie d'interaction du DTG et de ses dérivés avec les deux bases terminales de l'ADN viral traité et évaluer ainsi l'efficacité des noyaux substitués par rapport au DTG.

Premièrement, nous avons calculé les énergies d'interaction (ΔE) de chaque noyau substitué en présence de C_{16} et G_4 , au niveau corrélé, en utilisant la méthode DFT en présence de deux types de fonctionnelles ; B97-D3 et B3LYP-D3. Nous avons gardé la même fonction de base cc-PVTZ (-f), implémenté dans le logiciel GAUSSIAN09. Nous avons obtenu deux valeurs ; une valeur énergétique brute et une autre après correction pour l'erreur de superposition de bases (BSSE). ΔE a été de même générée par la méthode ALMOEDA avec deux types de fonctionnelles ; ω -B97D et B97-D3 (logiciel Q-Chem ^[4]). Les colonnes 2 et 3 du Tableau 1 correspondent aux valeurs de ΔE générées par les fonctionnelles B97-D3 et B3LYP-D3 de la méthode DFT, respectivement. Les colonnes 4 et 5 correspondent aux valeurs de ΔE obtenues par ALMOEDA avec les fonctionnelles B97D3 et ω -B97D respectivement. A noter que nous avons calculé de même les énergies de chaque partenaire pris seul afin d'obtenir les valeurs brutes données dans les colonnes de la manière suivante :

$$\text{Energie d'interaction totale } (\Delta E) = \text{Energie du complexe (noyau substitué + } C_{16} + G_4) - (\text{Energie inhibiteur + Energie } C_{16} + \text{Energie } G_4)$$

De plus, nous avons calculé l'énergie de solvation (ΔG_{solv}) de chaque noyau halogéné seul (DTG + les treize dérivés) suivant la procédure du modèle de continuum polarisable. ΔG_{solv} représente l'énergie de solvation réelle de chaque noyau avant sa complexation. Sa contribution sera prise en considération dans la valeur finale de l'énergie totale d'interaction, pour des résultats plus précis

et une meilleure reproduction de la situation *in vivo*. Ses valeurs sont indiquées dans la colonne 7 du Tableau 1. Il faut noter que le calcul des énergies libres de solvation réelles du DTG entier et des dérivés substitués entiers dans une boîte de molécules d'eau simulée dynamiquement sera considéré dans la prochaine étape de nos études (Chapitre IV partie C). Outre les méthodes de mécanique quantique, la mécanique moléculaire polarisable, via le champ de force SIBFA, nous a permis aussi de déterminer les énergies d'interaction totales au niveau des complexes ternaires (noyau halogéné – C₁₆ – G₄), indiquées dans la colonne 6 du Tableau 1. La valeur totale est directement donnée par la méthode SIBFA.

Tableau 1 : Valeurs des énergies d'interactions intermoléculaires en kcal/mol des complexes ; nouveaux noyaux halogénés – C₁₆-G₄, calculées par des méthodes de chimie quantique (DFT et ALMOEDA) et de mécanique moléculaire polarisable (SIBFA). L'énergie de solvation ($\Delta G_{\text{sol}}^{\text{v}}$) est donnée de même. Les valeurs énergétiques du DTG et des meilleurs noyaux halogénés sont indiquées en rouge.

Compounds	ΔE B97-D3 (raw value/BSSE corrected)	ΔE B3LYP-D3 (raw value/BSSE corrected)	ΔE ALMOEDA (B97D3)	ΔE ALMOEDA (ω -B97D)	SIBFA	$\Delta G_{\text{sol}}^{\text{v}}$	Total interaction energy (ΔE B97-D3 - $\Delta G_{\text{sol}}^{\text{v}}$) (raw value/BSSE corrected)
DTG	-37.4/-33.2	-40.6/-36.5	-33.2	-35.9	-39.6	-6.1	-31.3/-27.1
A1	-37.5/-33.8	-40.6/-37.0	-33.8	-35.8	-38.4	-5.9	-31.6/-27.9
A2	-45.1/-40.7	-48.1/-43.7	-40.6	-43.3	-45.9	-9.7	-35.4/-31.0
B2	-44.9/-40.5	-47.9/-43.5	-40.4	-43.1	-46.6	-9.5	-35.4/-31.0
C2	-46.0/-41.3	-49.1/-44.5	-41.2	-44.1	-47.3	-10.1	-35.9/-31.2
D2	-43.0/-38.6	-46.3/-41.9	-38.5	-41.0	-45.2	-7.8	-35.2/-30.8
E2	-45.3/-40.9	-48.2/-43.9	-40.6	-43.4	-47.7	-7.8	-37.5/-33.1
A3	-44.7/-40.4	-47.6/-43.4	-40.3	-42.7	-44.6	-9.4	-35.3/-31.0
B3	-44.9/-40.7	-47.9/-43.7	-40.6	-43.1	-46.9	-9.5	-35.4/-31.2
C3	-44.1/-39.6	-47.0/-42.6	-38.8	-41.5	-45.2	-9.7	-34.3/-29.9
D3	-42.6/-37.8	-45.7/-41.1	-38.4	-41.2	-44.0	-7.8	-34.8/-30.0
A4	-42.5/-38.2	-45.2/-40.9	-38.0	-40.6	-43.7	-11.2	-31.3/-27.0
B4	-41.5/-37.0	-44.2/-39.7	-37.4	-39.7	-41.1	-9.7	-31.8/-27.3
C4	-42.1/-37.5	-45.0/-40.4	-41.2	-41.0	-42.6	-10.3	-31.9/-27.2

DTG as well as the compounds displaying the best energetic values are highlighted in red.

Il faut noter que plus la valeur de l'énergie d'interaction est basse, plus le complexe est stable, et par conséquent présente de meilleures interactions entre les différents partenaires. Ainsi, et dans le cas de la méthode DFT avec la fonctionnelle B97-D3 et après correction par BSSE, nous remarquons que tous les nouveaux composés affichent une meilleure énergie d'interaction avec C₁₆ et G₄ comparés au DTG ; les valeurs de ΔE corrigées sont proches pour les dérivés halogénés, dans la marge de -33,8 kcal/mol à -41,3 kcal/mol alors que ΔE (DTG) a une valeur de -33,2 kcal/mol. De même avec la fonctionnelle B3LYP-D3 et après correction par BSSE, les dérivés halogénés montrent des valeurs plus stables (entre -37,0 kcal/mol et -44,5 kcal/mol) que le DTG (-36,5 kcal/mol). Ceci est observable également avec ALMOEDA où pour la fonctionnelle B97-D3, l'énergie d'interaction de DTG est de -33,2 kcal/mol et celles des nouveaux dérivés varient entre -33,8 kcal/mol et -41,2 kcal/mol et pour la fonctionnelle ω -B97D où DTG affiche une valeur de -35,9 kcal/mol comparée à d'autres qui varient entre -35,8 kcal/mol et -44,1 kcal/mol pour les noyaux dérivés.

D'autre part, les valeurs obtenues avec SIBFA indiquent une plus faible énergie d'interaction du dérivé A1 avec les deux bases virales C₁₆ et G₄, d'une valeur de -38,4 kcal/mol, suivie du DTG qui possède une valeur de -39,6 kcal/mol. Les valeurs énergétiques des autres dérivés halogénés se situent entre -41,1 kcal/mol et -47,7 kcal/mol. Pour l'énergie de solvation (ΔG_{solv}), le groupe 1 comprenant DTG et le composé A1 ont les valeurs les plus faibles, de -6,1 kcal/mol et -5,9 kcal/mol respectivement. Les autres groupes montrent des valeurs qui varient entre -7,8 kcal/mol et -11,2 kcal/mol.

Enfin et afin d'obtenir les valeurs de l'énergie d'interaction totale à la colonne 8 du Tableau 1 (brute et corrigée par BSSE), nous avons soustrait la valeur de ΔG solv de la valeur de ΔE B97-D3 obtenue par DFT. Nous remarquons alors que l'ensemble des nouveaux dérivés présente des interactions plus fortes avec les deux bases virales, comparés au DTG, les valeurs de l'énergie d'interaction totale étant comprises entre -27,9 kcal/mol et -33,1 kcal/mol versus -27,1 kcal/mol pour le DTG.

De plus, nous concluons que les meilleurs dérivés issus de DTG présentant les énergies d'interaction totale les plus favorables sont les composés A2, B2, C2 et E2 du groupe 2 ainsi que le composé B3 du groupe 3. Pour rappel, les composés du groupe 2 et 3 présentent un NHCH_3 en para comme donneur d'électrons avec un fluor en ortho et un groupement attracteur d'électrons au niveau de la deuxième position ortho pour les dérivés du groupe 2, et un halogène en ortho ou en méta pour les dérivés du groupe 3. E2 est le dérivé le plus prometteur, doté d'une valeur de ΔE totale après correction égale à -33,1 kcal/mol. Il est suivi de C2 et de B3 avec une valeur de ΔE totale similaire égale à -31,2 kcal/mol. Pour les composés A2 et B2, nous avons trouvé une valeur de -31,0 kcal/mol.

Les composés du groupe 4, avec un NH_2 en position para par rapport au DTG, présentent quant à eux des valeurs énergétiques plus faibles que celles des composés des groupes 2 et 3, mais restent quand même plus efficaces que les dérivés du groupe 1, avec des énergies de 4 à 5 kcal/mol plus favorables que celle du DTG. Cependant, leurs valeurs absolues de ΔG solv sont élevées, ce qui a compensé l'écart entre leurs ΔE totales et celle du DTG, avec une différence finale de 0,1-0,2 kcal/mol. Concernant les valeurs de BSSE, leur moyenne est comprise entre 1,5 et 2 kcal/mol dans le cas de la DFT et elle est de 4,14 kcal/mol pour la

fonctionnelle B97-D3 ainsi qu'une valeur de 3,41 kcal/mol pour la fonctionnelle ω -B97D dans le cas d'ALMOEDA. Comme récapitulatif, le classement des groupes par ordre décroissant d'efficacité sera le suivant : groupe 2 > groupe 3 > groupe 4 > groupe 1, et le classement des cinq meilleurs dérivés sera ainsi : composé E2 > composé C2 > composé B3 > composé A2 > composé B2.

Nous proposons par la suite qu'un donneur d'électrons en para (NHCH_3) et deux attracteurs d'électrons en ortho contribueraient à augmenter la stabilité du complexe inhibiteur – bases terminales de l'ADN viral du VIH-1.

La Figure 42 représente les courbes des énergies d'interaction totale (valeurs corrigées) calculées pour chaque dérivé via l'ensemble des méthodes indiquées dans le Tableau 1. Elles soulignent l'accord entre les méthodes quantiques et les méthodes de mécanique moléculaire.

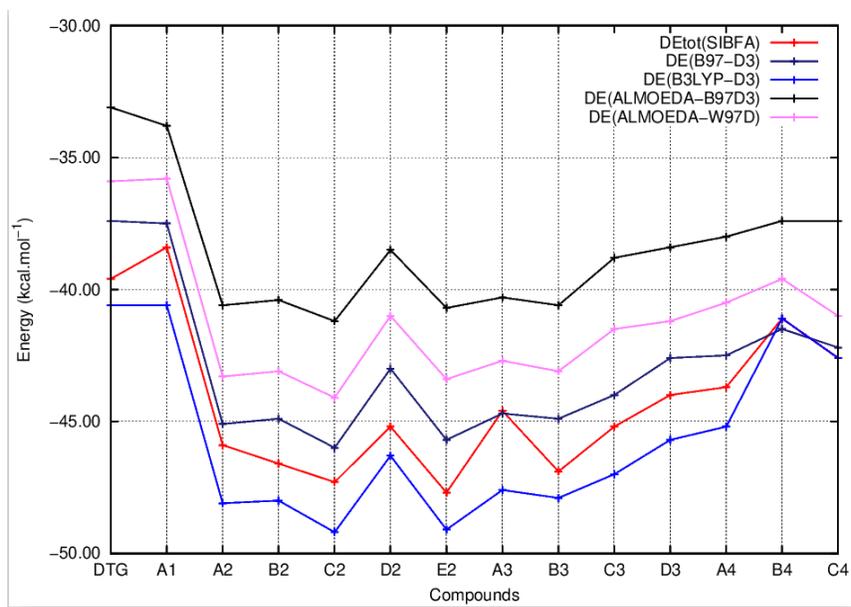


Figure 42 : Evolution des énergies d'interaction intermoléculaire du complexe ternaire C₁₆-G₄-noyau substitué en fonction du cycle halogéné du Dolutégravir (DTG) et des cycles des dérivés nouvellement conçus (Composés A1 à C4). Les valeurs de ΔE (DE) présentées sont calculées via des calculs corrélés de chimie quantique, se basant sur la méthode de DFT avec les fonctionnelles B97-D3 (DE B97-D3, courbe bleu marine) et B3LYP-D3 (DE B3LYP-D3, courbe bleu clair). ΔE obtenue via la méthode des orbitales moléculaires absolument localisées (ALMOEDA) est également donnée, en utilisant deux fonctionnelles ; B97-D3 (DE ALEMODA B97D3, courbe noire) et ω -B97D (DE ALMOEDA W97D, courbe rose). Les valeurs de ΔE calculées via des calculs de mécanique moléculaire avec le champ de force polarisable SIBFA (DE tot SIBFA, courbe rouge) sont indiquées aussi. Tous les composés présentent une énergie d'interaction totale avec C₁₆ et G₄ plus stable que celle de DTG, avec comme meilleurs dérivés dans l'ordre décroissant les composés E2, C2, B3, A2 et B2.

Discussion des valeurs énergétiques obtenues

- Nous pouvons élaborer, d'après les valeurs totales obtenues par les calculs quantiques DFT et ALMOEDA et celles obtenues par SIBFA, et d'après la formule de calcul de ΔE totale (indiquée à la page 185) que pour tous les composés, les énergies d'interaction des complexes ternaires sont non-additives : ΔE totale est supérieure en valeurs absolues à la somme de ses valeurs dans les trois complexes « binaires » pris deux à deux ; noyau halogéné – Guanine 4 , noyau halogéné – Cytosine 16 et Guanine 4 – Cytosine 16. Ceci est une caractéristique des interactions coopératives ^[5]. Cette

coopérativité de ΔE peut impacter favorablement les affinités des nouveaux noyaux conçus pour leur cible. Nous pouvons donc affirmer que la présence des noyaux halogénés au sein du complexe formé par C_{16} et G_4 stabilise leur position et que leur interaction simultanée avec les deux bases est plus stable que celle établie avec chaque base séparément.

- La totalité de nos 13 composés présente une énergie d'interaction totale plus stable que celle du DTG. Ceci souligne l'importance de nos substitutions au niveau de l'interaction de la drogue avec les bases terminales virales. En fait, la présence d'un groupement électro donneur, riche en atomes d'hydrogènes ($NHCH_3$ dans le cas du groupe 3 et NH_2 dans le cas du groupe 4) a renforcé l'interaction électrostatique entre la drogue et le noyau riche en électrons de G_4 , par rapport à l'interaction du trou sigma de la liaison C-F du DTG avec G_4 .
- A1 qui fait partie du groupe 1 avec le DTG, est le dérivé le moins efficace, avec une valeur de ΔE totale proche de celle du DTG de -27,9 kcal/mol. Ceci est expliqué par le fait que leurs structures sont très similaires ; A1 présente un chlore en para à la place du fluor. Dépourvu d'un groupement donneur d'électrons en para, il reste moins efficace que les dérivés des groupes 2, 3 et 4. Ceci souligne davantage l'importance d'avoir un groupement donneur d'électrons en para, faisant face à G_4 .
- Les interactions des dérivés des groupes 2 et 3 sont plus stables par rapport à celles des dérivés du groupe 4. En moyenne, les groupes 2 et 3 ont des valeurs de ΔE totales variant entre -34 kcal/mol et -35 kcal/mol alors que le groupe 4 a une valeur moyenne de ΔE totale de -31 kcal/mol. Ces résultats favorisent la présence du groupement $NHCH_3$ par rapport à NH_2 en position

para. En effet, le premier est plus riche en protons que le second, il constitue alors une région plus électropositive que celle formée par le groupement NH_2 , ce qui renforce l'interaction avec G_4 .

- Le groupe 2 quant à lui renferme les dérivés les plus efficaces ; C2, E2, A2 et B2. Outre le groupement NHCH_3 , il présente deux groupements attracteurs d'électrons par rapport aux dérivés du groupe 3 qui ne présentent qu'un seul. Ces deux groupements riches en électrons font face à la cytosine, favorisant des interactions par empilements π - π avec elle. Ce qui explique leur meilleure efficacité, notamment dans le cas du meilleur noyau dérivé, E2 qui avec son groupement cyanure, possède une région très enrichie en électrons. D2 dans ce groupe, reste cependant moins efficace que les autres. Ceci peut-être expliqué par le fait que ses deux groupements attracteurs d'électrons se trouvent en ortho et méta par rapport à l'hydrogène le reliant au reste de la drogue. Une forte répulsion peut exister ainsi entre les deux atomes et contribuer à déstabiliser légèrement ΔE totale.
- De plus et comme déjà mentionné, El Hage et collaborateurs au sein de notre équipe ont travaillé précédemment sur l'importance des substitutions de NHCH_3 et de CF_3 au niveau du noyau d'EVG afin d'améliorer son efficacité d'interaction avec l'ADN viral ^[3]. Nos résultats restent en accord avec cette étude puisque nos meilleurs dérivés appartenant au groupe 2 présentent ces groupements (cas du composé C2).
- D'autre part, les valeurs quantiques obtenues avec DFT et ALMOEDA sont très proches même presque identiques, notamment dans le cas où la même fonctionnelle B97-D3 est utilisée. Cet accord entre nos méthodes quantiques confirme l'exactitude de nos résultats.

- Enfin, il existe un accord entre les méthodes de calcul quantique (DFT et ALMOEDA) et SIBFA adoptée en mécanique moléculaire. Ce qui confirme des résultats prouvés antérieurement par notre équipe, sur la capacité de notre champ de force polarisable SIBFA à reproduire les résultats obtenus par chimie quantique, plus rapidement et plus efficacement ^{[6][7]}. De tels accords sont encourageants étant donné qu'aucun effort de calibrage supplémentaire n'a été fait sur les dérivés du DTG.

A.III.2. Décomposition de l'énergie d'interaction totale des composés

Comme montré par les calculs quantiques effectués via les approches DFT, ALMOEDA et SIBFA, nos nouveaux noyaux substitués présentent une interaction avec les bases virales plus stable que celle de DTG. Ils promettent donc d'être plus efficaces que ce dernier. Dans le but de décrypter les origines de ces améliorations énergétiques, nous avons procédé à une décomposition de l'énergie d'interaction totale en différentes composantes par la méthode RVS qui décompose ΔE totale obtenue par DFT, par la méthode ALMOEDA qui décompose sa propre énergie fournie par le potentiel SIBFA qui assure de même une décomposition de l'énergie d'interaction totale calculée.

La première contribution donnée par ces méthodes est l'énergie du premier ordre ; E1. Pour RVS, E1 correspond à la somme de l'énergie de Coulomb (E_c) et l'énergie d'échange-répulsion (E_x). Pour ALMOEDA, cette énergie est donnée directement sous forme d'une contribution énergétique gelée (FRZ). Quant à SIBFA, E1 est égale à la somme de l'énergie électrostatique augmentée par un terme de pénétration (E_{MTP}) avec l'énergie de répulsion à courte portée (E_{rep}). Les valeurs d'E1 obtenues

sont indiquées dans le Tableau 2 et les courbes correspondantes sont représentées dans la Figure 43.

Tableau 2 : Valeurs de la contribution du premier ordre E1 calculées par les méthodes suivantes : RVS et ALMOEDA (B97-D3 et ω -B97D) et SIBFA pour le complexe G₄-C₁₆-noyau halogéné. Les valeurs du Dolutégravir (DTG) et des cinq meilleurs dérivés sont mis en relief en gras et rouge.

<u>Compounds</u>	<u>E1 (RVS)</u> (kcal/mol)	<u>E1 (SIBFA)</u> (kcal/mol)	<u>E1 (ALMOEDA – B97-D3)</u> (kcal/mol)	<u>E1 (ALMOEDA – W97D)</u> (kcal/mol)
DTG	- 1,5	-2,8	-12,8	-17,4
A1	- 0,9	-3,3	-13,0	-17,1
A2	-3,8	-7,0	-17,1	-21,9
B2	-3,7	-7,5	-16,9	-21,7
C2	-4,2	-8,0	-17,6	-22,5
D2	-0,1	-4,8	-15,8	-20,3
E2	-2,6	-8,7	-17,2	-21,8
A3	-0,8	-6,0	-16,9	-21,4
B3	-2,5	-8,5	-17,3	-21,8
C3	-1,3	-5,5	-15,6	-20,2
D3	-0,8	-3,4	-15,6	-20,2
A4	-4,4	-5,2	-15,8	-20,1
B4	-3,1	-3,3	-14,8	-19,1
C4	-3,4	-2,6	-20,2	-19,1

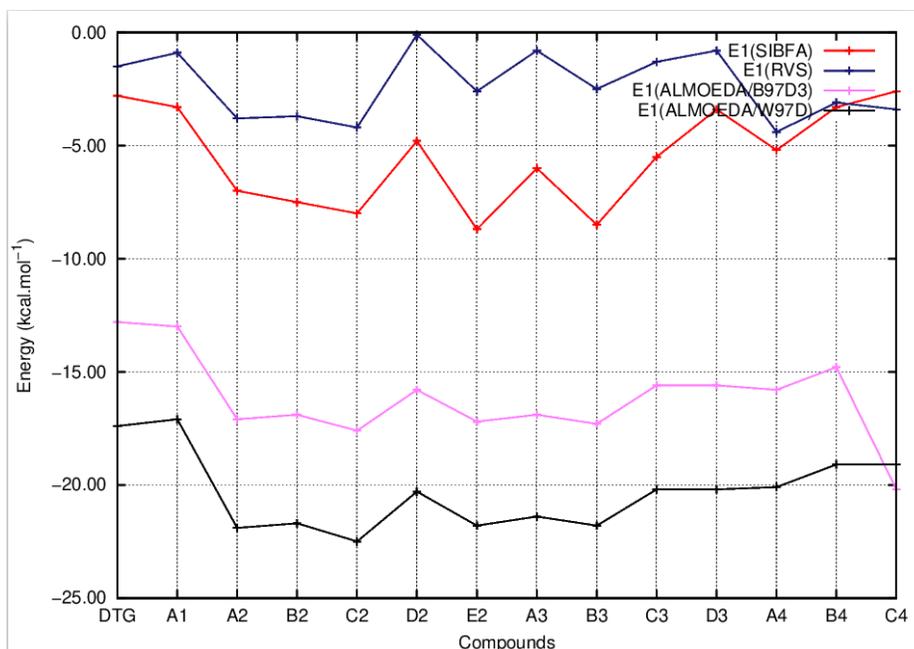


Figure 43: Courbes de l'évolution de l'énergie du premier ordre (E1) du complexe ternaire G_4-C_{16} -noyau substitué en fonction des composés ; Dolutégravir (DTG) et nos noyaux conçus (Composés A1 à C4). Les valeurs de E1 sont calculées via la chimie quantique par les méthodes de décomposition d'énergie ; RVS (E1 RVS ; Reduced Variational Space, courbe bleu marine) et ALMOEDA (Absolutely Localized Molecular Orbitals). Cette dernière utilise deux fonctionnelles B97-D3 (E1 ALMOEDA B97D3, courbe rose) et ω -B97D (E1 ALMOEDA W97D, courbe noire). Les valeurs de E1 sont obtenues aussi via des calculs de mécanique moléculaire avec le champ de force polarisable SIBFA (E1 SIBFA, courbe rouge).

Nous pouvons constater que les profils des courbes d'E1 de RVS et SIBFA sont similaires, avec des valeurs variant entre -0,1 kcal/mol et -4 kcal/mol pour RVS et entre -2,6 kcal/mol et -8,5 kcal/mol pour SIBFA. Les valeurs les plus stables correspondent aux dérivés A2, B2, C2 et E2 du groupe 2 et au dérivé B3 du groupe 3. Elles sont entre 2,5 kcal/mol et 4,2 kcal/mol dans le cas de la méthode RVS et entre 7,0 kcal/mol et 8,7 kcal/mol dans le cas de SIBFA. Les valeurs les moins stables correspondent notamment aux éléments du groupe 1 ; le DTG et A1. Quant aux valeurs d'E1 données par la méthode ALMOEDA, les deux courbes présentent des profils proches quel que soit la fonctionnelle, avec les valeurs les plus stables attribuées aussi aux dérivés A2, B2, C2 et E2 ainsi que B3. De même, on trouve les

valeurs les moins stables dans le cas du DTG et d'A1, de moyenne -13 kcal/mol pour la fonctionnelle B97-D3 et -17 kcal/mol pour la fonctionnelle ω -97D.

La deuxième contribution fournie par les méthodes de décomposition d'énergie est celle du deuxième ordre ; E2 qui correspond dans toutes nos méthodes utilisées à la somme de l'énergie de polarisation (E_{pol}) et l'énergie de transfert de charge (E_{ct}). Les valeurs et courbes correspondantes sont représentées dans le Tableau 3 ainsi que dans la Figure 44, respectivement.

Tableau 3 : Valeurs de la contribution du second ordre (E2) calculées par les méthodes suivantes : RVS et ALMOEDA (B97-D3 et ω -B97D) et SIBFA pour le complexe G_4 -C₁₆-noyau halogéné. Les valeurs du Dolutégravir (DTG) et des cinq meilleurs dérivés sont mises en relief en gras et rouge.

<u>Compounds</u>	<u>E2 (RVS)</u> (kcal/mol)	<u>E2 (SIBFA)</u> (kcal/mol)	<u>E2 (ALMOEDA – B97-D3)</u> (kcal/mol)	<u>E2 (ALMOEDA – W97D)</u> (kcal/mol)
DTG	-17,3	-13,5	-20,3	-18,5
A1	-17,1	- 13,6	-20,7	-18,8
A2	-19,4	-16,5	-23,4	-21,4
B2	-19,3	-16,6	-23,4	-21,4
C2	-19,5	-16,7	-23,6	-21,5
D2	-19,1	-16,0	-22,7	-20,6
E2	-19,2	-16,4	-23,6	-21,6
A3	-19,7	-16,9	-23,4	-21,3
B3	-19,2	-16,9	-23,3	-21,3
C3	-19,6	-15,4	-23,2	-21,2
D3	-19,2	-16,3	-22,8	-20,9
A4	-18,4	-15,7	-22,2	-20,4
B4	-18,3	-15,4	-22,6	-20,4
C4	-18,5	-15,4	-21,0	-21,9

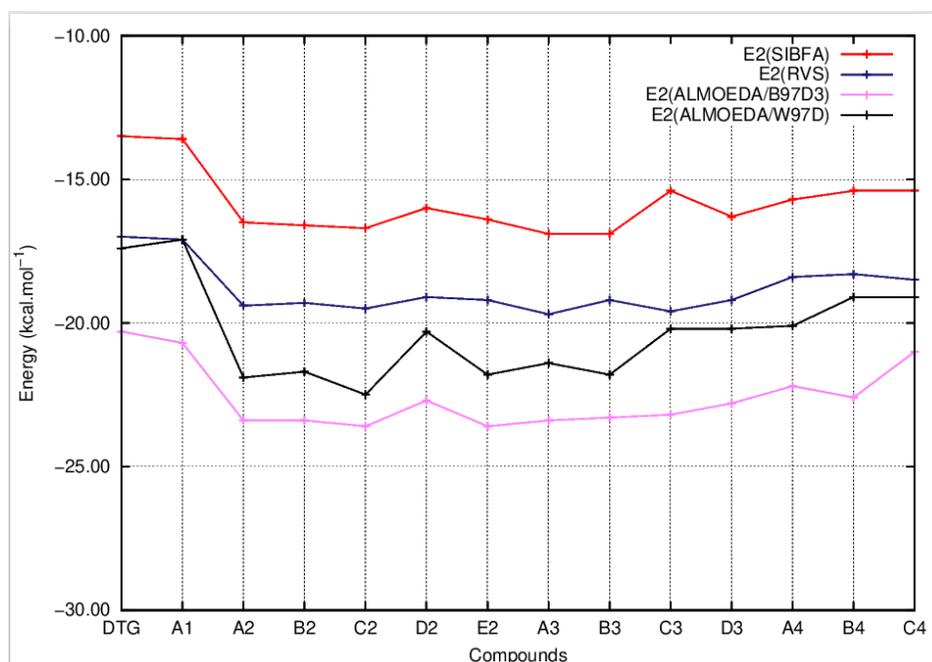


Figure 44: Courbes de l'évolution de l'énergie du second ordre (E2) du complexe ternaire G_4-C_{16} -noyau substitué en fonction des composés ; Dolutégravir (DTG) et nos noyaux conçus (Composés A1 à C4). Les valeurs d'E2 sont calculées via la chimie quantique par les méthodes de décomposition d'énergie ; RVS (E2 RVS ; Reduced Variational Space, courbe bleu marine) et ALMOEDA (Absolutely Localized Molecular Orbitals). Cette dernière utilise deux fonctionnelles B97-D3 (E2 ALMOEDA B97D3, courbe rose) et ω -B97D (E2 ALMOEDA W97D, courbe noire). Les valeurs d'E2 sont obtenues aussi via des calculs de mécanique moléculaire avec le champ de force polarisable SIBFA (E2 SIBFA, courbe rouge).

Le profil des quatre courbes de la Figure 44 est à peu près identique, avec DTG et A1 ayant les valeurs énergétiques les moins stables de moyenne de 17 kcal/mol en RVS, -13 kcal/mol en SIBFA, -20 kcal/mol en ALMOEDA B97-D3 et -18 kcal/mol en ALMOEDA ω -97D. D'après l'aspect des courbes ainsi que les valeurs du Tableau 3, nous pouvons affirmer que tous les dérivés appartenant aux groupes 2, 3 et 4 montrent des valeurs d'E2 plus stables que celles des composés du groupe 1, avec des moyennes aux alentours de -19 kcal/mol dans le cas de RVS, -16 kcal/mol dans le cas de SIBFA, -23 kcal/mol et -21 kcal/mol dans le cas d'ALMOEDA, avec les fonctionnelles B97-D3 et ω -97D respectivement.

La troisième et dernière contribution donnée par les méthodes de décomposition d'énergie est l'énergie de dispersion (E_{disp}). Etant donné que ce terme de dispersion n'est pas calculable au niveau non corrélé (niveau Hartree-Fock), nous avons obtenu une version approchée de cette énergie en soustrayant l'énergie d'interaction calculée au niveau HF (valeurs non présentées) à l'énergie d'interaction obtenue par DFT avec les fonctionnelles B97-D3 et B3LYP-D3. Nous avons obtenu alors E_{disp} (B97-D3) et E_{disp} (B3LYP-D3). Quant à SIBFA, elle fournit directement la valeur de cette contribution. Les valeurs d' E_{disp} figurent dans le Tableau 4 pour chaque complexe C₁₆-G₄-noyau halogéné. La Figure 45 représente les courbes respectives aux valeurs obtenues.

Tableau 4 : Valeurs de l'énergie de dispersion E_{disp} calculée par RVS et SIBFA pour le complexe G₄-C₁₆-noyau halogéné. Deux fonctionnelles sont utilisées dans le cas de RVS ; la fonctionnelle B97-D3 et la fonctionnelle B3LYP-D3. Les valeurs du Dolutégravir (DTG) et des cinq meilleurs dérivés sont mis en relief en gras et rouge.

<u>Compounds</u>	<u>E_{disp} (RVS B97-D3)</u> (kcal/mol)	<u>E_{disp} (RVS – B3LYP-D3)</u> (kcal/mol)	<u>E_{disp} (SIBFA)</u> (kcal/mol)
DTG	-17,4	-20,6	-23,2
A1	-18,2	- 21,3	-21,5
A2	-20,3	-23,3	-22,4
B2	-20,3	-23,4	-22,5
C2	-20,5	-23,7	-22,6
D2	-22,2	-25,5	-24,4
E2	-21,8	-25,2	-22,6
A3	-22,7	-25,6	-21,6
B3	-21,6	-24,6	-21,4
C3	-21,5	-24,4	-24,4
D3	-20,9	-24,0	-24,4
A4	-18,2	-20,9	-22,9
B4	-18,4	-18,0	-22,5
C4	-18,5	-19,2	-24,8

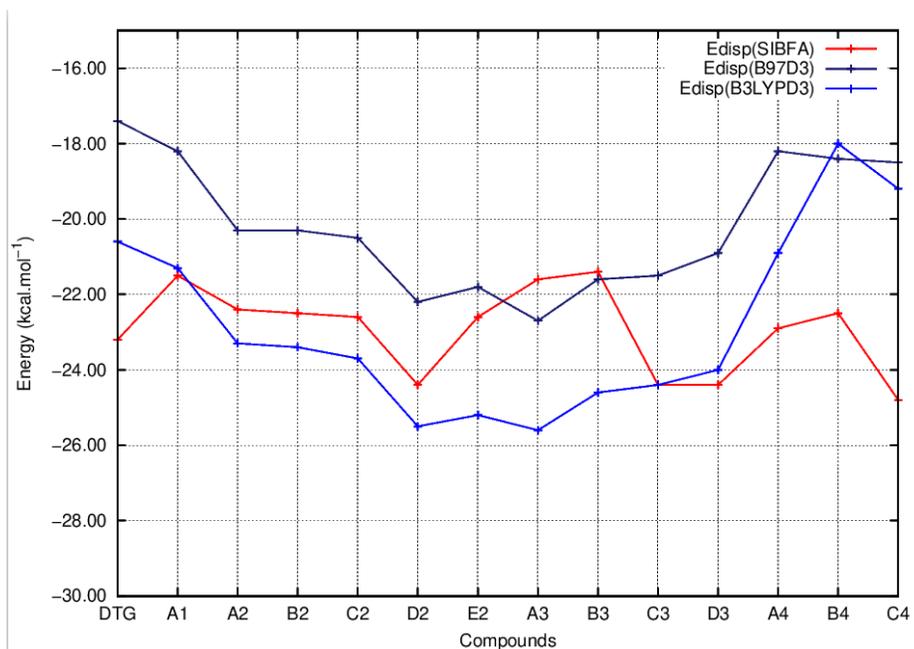


Figure 45: Courbes de l'évolution de l'énergie de dispersion (E_{disp}) du complexe ternaire $G_4\text{-C}_{16}$ -noyau substitué en fonction des composés ; Dolutégravir (DTG) et nos noyaux conçus (Composés A1 à C4). Les valeurs d' E_{disp} sont calculées via la chimie quantique par la méthode de décomposition d'énergie ; RVS utilisant deux fonctionnelles ; B97D3 (courbe violette) et B3LYPD3 (courbe bleue). Les valeurs d' E_{disp} sont obtenues aussi via des calculs de mécanique moléculaire avec le champ de force polarisable SIBFA (courbe rouge).

Le profil des courbes de l'énergie de dispersion est identique dans le cas de la démarche quantique pour les deux fonctionnelles adoptées. Cependant, il existe un faible accord entre les valeurs obtenues par SIBFA et celles obtenues par RVS, notamment pour les composés A1 jusqu'à E2 qui possèdent des valeurs de E_{disp} (SIBFA) intermédiaires à celles obtenues par RVS ; pour les composés A3 et B3, ces valeurs sont proches de celles de E_{disp} (B97D3) avec des différences de 1,2 kcal/mol et 0,2 kcal/mol respectivement. Pour C3 et D3, E_{disp} (SIBFA) est plus proche des valeurs d' E_{disp} (B3LYP-D3) avec des différences aux alentours de 0,4 kcal/mol. Quant aux composés du groupe 4, les valeurs de l'énergie de dispersion données par SIBFA sont plus faibles que celles fournies par les deux fonctionnelles de RVS. Nous remarquons aussi que les valeurs de cette énergie sont légèrement plus stables dans le cas des composés D2, A3 et C3 qui présentent pour chaque méthode des

valeurs d' E_{disp} faiblement moins élevées que les meilleures dérivés identifiés. Par exemple, pour une valeur d' E_{disp} (B97-D3) moyenne de -20 kcal/mol pour A2, B2 et C2, D3 et A3 présentent une valeur moyenne de -22 kcal/mol.

Finalement, nous avons calculé la valeur de l'énergie d'interaction totale au niveau non corrélé avec la méthode RVS (ΔE RVS) et avec SIBFA sans la contribution de dispersion (ΔE SIBFA) afin d'analyser l'effet de cette troisième contribution. ΔE RVS est obtenue par soustraction de la valeur de BSSE de l'énergie d'interaction HF (valeurs de ΔE HF non présentées). ΔE (SIBFA) est donnée directement. Nous présenterons dans le Tableau 5 les valeurs que nous avons obtenues et leurs courbes correspondantes à la Figure 46.

Tableau 5 : Valeurs de l'énergie (ΔE) calculée par RVS (au niveau non corrélé) et SIBFA (sans la contribution de dispersion) pour le complexe G_4-C_{16} -noyau halogéné. Les valeurs du Dolutégravir (DTG) et des cinq meilleurs dérivés sont mis en relief en gras et rouge.

<u>Compounds</u>	<u>ΔE (RVS)</u> (kcal/mol)	<u>ΔE (SIBFA)</u> (kcal/mol)
DTG	-18,4	-16,3
A1	-18,1	-16,9
A2	-23,2	-23,5
B2	-23,0	-24,1
C2	-23,4	-24,7
D2	-19,3	-20,8
E2	-21,8	-25,1
A3	-20,4	-22,9
B3	-21,7	-25,5
C3	-21,0	-21,0
D3	-20,0	-19,7
A4	-22,8	-21,0
B4	-21,5	-18,8
C4	-21,8	-18,0

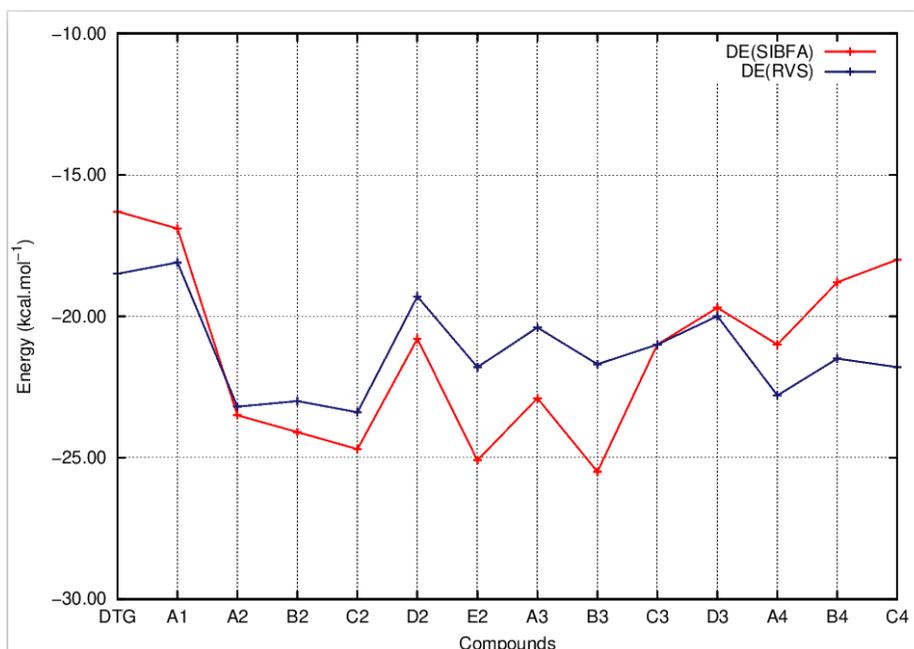


Figure 46: Courbes de l'évolution de l'énergie d'interaction (ΔE) du complexe ternaire G_4 - C_{16} -noyau substitué en fonction des composés ; Dolutégravir (DTG) et nos noyaux conçus (Composés A1 à C4). Ces valeurs sont calculées via la méthode quantique de décomposition d'énergie RVS (Reduced Variational Space ; DE RVS, courbe bleu marine). ΔE est aussi donnée via des calculs de mécanique moléculaire avec le champ de force polarisable SIBFA (DE SIBFA, courbe rouge).

Les valeurs obtenues suite au calcul de ΔE RVS et ΔE SIBFA sont semblables aux valeurs de l'énergie d'interaction totale. Les deux courbes de la Figure 46 présentent en fait des minima avec les dérivés A2, B2, C2, E3 et B3 (valeurs soulignées en rouge dans le Tableau 5). ΔE RVS a un minimum supplémentaire avec le composé A4 de la série 4. Les valeurs de SIBFA restent majoritairement en accord avec les valeurs de RVS, mettant en évidence de même les cinq meilleurs dérivés déjà identifiés.

Discussion des décompositions énergétiques réalisées

- Les contributions énergétiques obtenues par SIBFA et représentées dans les Figures 38 à 41 reproduisent assez étroitement celles fournies par les calculs quantiques (RVS et ALMOEDA). Ceci confirme la capacité de SIBFA à refléter

les résultats de l'énergie d'interaction totale calculée par les méthodes quantique en reproduisant ses différentes contributions énergétiques efficacement.

- Les analyses de décomposition d'énergie par les méthodes RVS, ALMOEDA et SIBFA nous permettent de constater que les principaux déterminants de l'affinité des noyaux substitués pour les bases de l'ADN viral sont les contributions du premier et du second ordre. En fait, les valeurs d'E1 et E2 conservent l'ordre observé avec les valeurs de ΔE totale, avec E2, C2, B3, A2 et B2 comme les dérivés les plus potentiels, contrairement à E_{disp} qui a présenté des valeurs plus stables pour d'autres dérivés. De plus, même en l'absence d' E_{disp} , ΔE (RVS) et ΔE (SIBFA) restent similaires à ΔE totale SIBFA et ΔE totale RVS ; ce qui souligne la contribution majeure des énergies du premier ordre et du second ordre.
- Même si les résultats de SIBFA apparaissent satisfaisants, l'accord entre E1 (SIBFA) et E1 (RVS et ALMOEDA) ainsi que l'accord entre E2 (SIBFA) et E2 (RVS et ALMOEDA) peuvent être encore améliorés. En fait, ces améliorations sont en cours, se basant sur la reconstruction d'une nouvelle bibliothèque de fragments protéiques et d'ADN ayant recours à des multipôles et à des polarisabilités issus de calculs corrélés. Une remise à l'échelle des contributions énergétiques individuelles doit être de même effectuée, et ceci par la méthode SAPT-DFT (Symmetry Adapted Perturbation Theory – Density Functional Theory) qui se base sur des analyses corrélées de décomposition d'énergie ^{[8][9]}.

A.IV. Evaluation qualitative de l'interaction des nouveaux noyaux halogénés avec les bases terminales de l'ADN viral du VIH-1

A.IV.1. Détermination des contours de potentiel électrostatique

Après avoir évalué nos nouveaux dérivés par une démarche quantitative en calculant leur énergie d'interaction avec la paire de bases virales terminales C₁₆/G₄, nous sommes passés à leur évaluation qualitative par détermination de leurs contours de potentiel électrostatique (MEP), à l'aide du logiciel de chimie quantique GAUSSIAN09. Ceci nous permettra d'explorer les effets électroniques résultants des halogénations et des substitutions ponctuelles effectuées et d'expliquer ainsi les valeurs énergétiques obtenues. La visualisation des contours de potentiel électrostatique après leur détermination se fait par le logiciel GAUSSVIEW ^[10]. La Figure 47 illustre les MEP obtenus pour nos dérivés. Nous avons choisi de représenter un ou deux composés de chaque groupe étant donné que les contours obtenus sont très similaires pour les membres d'un même groupe. Du coup, nous montrons les résultats pour le DTG du groupe 1, les composés C2 et E2 du groupe 2, le composé B3 du groupe 3 et le composé C4 du groupe 4. Pour chaque composé deux structures sont affichées ; l'une avec les contours électrostatiques uniquement (structure en haut pour le DTG et à gauche pour les autres composés) et l'autre avec les contours en présence des atomes de la molécule (en bas pour le DTG et à droite pour les autres composés). Les structures des noyaux obtenues par Chimera sont également présentées comme repère pour les positions des atomes.

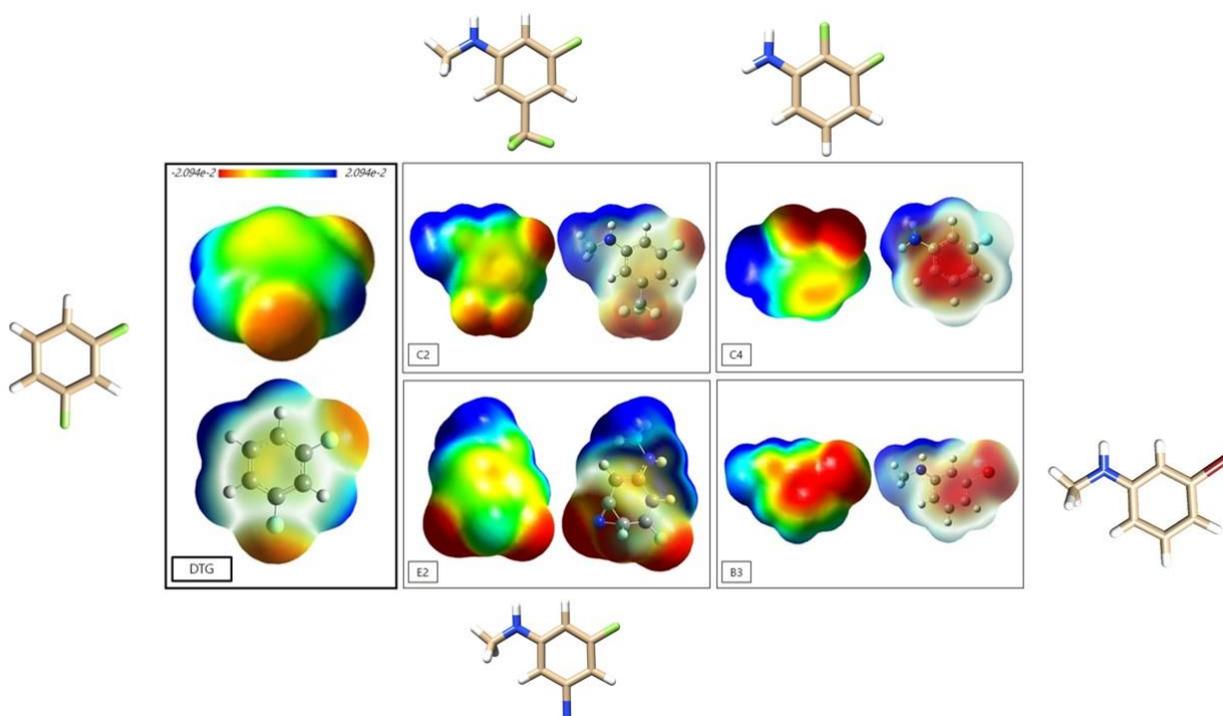


Figure 47 : Contours des potentiels électrostatiques (MEP) déterminés pour le Dolutégravir et des dérivés représentatifs de chaque groupe ; C2 et E2 du groupe 2, B3 du groupe 3 et C4 du groupe 4. La surface d'isodensité est de $0,004 \text{ électrons Bohr}^{-3}$. La barre colorée en dessus de la Figure indique la variation de la densité électronique entre $-2,094 \times 10^{-2}$ (correspondant à la couleur rouge) et $2,094 \times 10^{-2}$ (correspondant à la couleur bleue). Les structures des noyaux sans contours sont représentées de même comme référence pour les positions des atomes.

Un examen des contours de potentiel électrostatique générés montre :

- Une région de couleur bleue (foncé et clair) qui correspond à une zone de faible densité électronique (δ^+). Elle est située principalement aux alentours des hydrogènes du DTG et au niveau des groupements NHCH_3 (pour les composés des groupes 2 et 3) et NH_2 pour les composés du groupe 4.
- Une région de couleur rouge orange qui correspond à une zone de densité électronique élevée ($-\delta$), localisée notamment aux alentours des groupements attracteurs d'électrons ; le fluor dans le cas du DTG et du composé C4, le fluor et le CF_3 dans le cas du composé C2, le fluor et le cyanure dans le cas du composé E2 et finalement le brome dans le cas de B3.

Discussion des résultats des contours de potentiel électrostatique

- Premièrement, nous observons une extension beaucoup plus large de la zone bleue de potentiel positif autour des substituants en position para (NHCH_3 ou NH_2) qu'autour du fluor en para du DTG. La zone autour de ce dernier est de couleur orange, par contre elle est de couleur bleue très prononcée au niveau des groupements NHCH_3 et NH_2 de nos nouveaux dérivés. Cette zone positive ($+\delta$) qui paraît chez les dérivés favorise alors les interactions électrostatiques avec la structure polycyclique riche en électrons de G_4 . Ce qui confirme les valeurs des énergies d'interaction de nos noyaux avec l'ADN viral par rapport au DTG (ΔE totale QC et ΔE totale SIBFA) et souligne l'effet important qu'apporte nos substitutions de groupements donneurs d'électrons en para à l'amélioration de l'interaction des noyaux des drogues avec les bases de l'ADN viral.
- Pour les composés du groupe 2, le groupement CF_3 du composé C2 génère une zone rouge dotée d'un potentiel négatif beaucoup plus large que le cyanure du composé E2. De plus, le potentiel positif autour du groupement NHCH_3 de C2 est plus important que celui d'E2. Ces constatations peuvent expliquer l'énergie de solvabilité (ΔE solv) supérieure de 2,1 kcal/mol de C2 par rapport à celle d'E2 (-10,1 kcal/mol vs -7,8 kcal/mol), impactant le bilan énergétique en faveur de ce dernier malgré que les deux dérivés présentent des valeurs proches de ΔE totale.
- Les zones rouges de potentiel négatif ($-\delta$) sont plus prononcées au niveau des substitutions de C2, E2 et B3 par rapport à C4. En fait, le groupement cyanure avec sa triple liaison, le CF_3 avec ses trois fluors et le brome avec son nombre d'électrons plus élevé que le fluor sont plus riches en électrons que

le fluor de C4. Ceci renforce l'interaction du cycle de ces dérivés, riche en électrons, avec le noyau de la cytosine C₁₆ par empilement π - π . Quant au composé C4, nous constatons qu'il présente une faible densité électronique au milieu de son noyau du fait de la présence de deux atomes de fluor moins attracteurs d'électrons, lui conférant l'interaction la moins favorable avec les paires virales de l'ADN. De plus, la position proche des deux fluors engendre une répulsion qui contribue au faible rendement énergétique de C4.

- Ajoutons que malgré la présence d'une région de potentiel électronégatif ($-\delta$) plus prononcé au niveau du centre du composé B3 par rapport à C2 et E2, il reste moins efficace que ces deux derniers. Ceci peut être expliqué par le fait que le groupement cyanure d'E2 et CF3 de C2 sont plus riches en électrons que le brome, ce qui compense un faible centre électronégatif de leur noyau par une région en ortho plus riche en électrons.

A.V. Conclusions du chapitre III partie A

Le travail accompli dans cette partie a pour objectif la conception de nouveaux noyaux halogénés inhibiteurs de l'étape de transfert de brin, ayant des affinités augmentées aux bases réactives C₁₆ et G₄ de l'ADN viral processé. Le but ultime serait de contourner les mutations de résistance de l'intégrase suite à l'administration des médicaments. Notre stratégie de conception s'est basée sur l'optimisation du Dolutégravir, l'INSTI le plus efficace mis sur le marché, par des halogénations sélectives (chlore et brome) et par des substitutions appropriées (groupements électropositifs comme le NH₂ et le NHCH₃ et groupements électronégatifs tels que le CF₃ et le cyanure). C'est une stratégie de travail simple mais rigoureuse, dans laquelle nous avons adopté comme modèle d'étude un mini-système formé par nos nouveaux noyaux + C₁₆ + G₄. Nous avons ensuite calculé

l'énergie d'interaction totale (ΔE) des trois partenaires par des calculs de chimie quantique post-HF basés sur la Théorie de la Fonctionnelle de la Densité (DFT), utilisant les fonctionnelles B97-D3 et B3LYP-D3. La méthode quantique ALMOEDA basée sur les orbitales moléculaires a été de même adoptée pour générer les valeurs de ΔE . Nos calculs ont donné des énergies d'interaction plus stables pour nos nouveaux noyaux avec les bases virales par rapport au DTG. Une analyse qualitative de ces interactions par détermination des contours de potentiel électrostatique de chaque noyau substitué a validé les énergies d'interaction calculées. Nos résultats ont prouvé en effet l'efficacité de la présence de forts groupements donneurs d'électrons NHCH_3 et NH_2 à assurer une meilleure interaction électrostatique avec les noyaux aromatiques de G_4 de densité électronique élevée, notamment le groupement NHCH_3 qui a un pouvoir électro donneur plus important que NH_2 . Ces groupements remplacent le fluor en para du DTG. Ajoutons que la présence de groupements attracteurs d'électrons en position ortho notamment à la place du fluor du DTG contribue à renforcer l'interaction de nos nouvelles molécules avec C_{16} par empilement π - π , en créant des zones riches en électrons. C'est le cas de nos meilleurs composés qui ont présenté des groupements électroattracteurs importants en ortho comme le cyanure (composé E2), le CF_3 (composé C2), le brome (composé B3 et A2) ou le chlore (composé B2). Finalement, une décomposition des énergies d'interaction totale par l'analyse RVS et ALMOEDA a déterminé les contributions responsables des meilleurs rendements énergétiques observés notamment celles du premier ordre (énergie électrostatique + énergie de répulsion) et du second ordre (énergie de polarisation + énergie de transfert de charge). Ces résultats ont été reproduits correctement par le champ de force polarisable SIBFA, qui a correctement pris en compte l'évolution des

contributions quantiques individuelles ainsi que celle de l'énergie d'interaction totale. Quelques calibrations supplémentaires restent cependant nécessaires afin de raffiner davantage ce potentiel.

De tels résultats prouvent donc que des halogénations et substitutions sélectives au niveau du noyau aromatique de DTG peuvent accroître et d'une manière significative les affinités de liaison de ce ligand inhibiteur pour un domaine d'acides nucléiques bien défini et hautement conservé. Elles pourraient ainsi donner lieu à de nouvelles drogues inhibitrices plus potentielles que le DTG, interagissant de préférence avec l'ADN viral au détriment de l'enzyme et pouvant surmonter l'obstacle de la résistance virale et être donc plus efficaces face à l'intégrase du VIH-1.

Chapitre IV

Résultats et Discussion

Partie B : Simulations par dynamique moléculaire des complexes biologiques cibles du VIH-1

B.I. Introduction générale de la partie B

Dans le cadre de l'amélioration du mode d'action des inhibiteurs de transfert de brin ciblant l'intégrase, nous avons commencé en premier lieu par l'optimisation de la structure du DTG de façon à renforcer ses interactions avec l'ADN viral ce qui lui permettra de contourner les mutations de résistance développées par l'intégrase. L'étude de l'interaction des nouveaux noyaux halogénés dérivés du noyau de DTG avec les noyaux des deux bases virales terminales situées à l'extrémité 3' de l'ADN viral processé ; C₁₆ et G₄ s'est faite en deuxième lieu. Les calculs énergétiques de mécanique quantique et de mécanique moléculaire ont mis en évidence les meilleures interactions de nos dérivés avec les deux bases, par rapport au DTG. Il est important par la suite d'extrapoler notre étude à une échelle plus élevée où nous analyserons l'interaction de la structure entière de nos nouvelles molécules (et non seulement leurs noyaux halogénés) avec l'ADN viral entier en présence aussi de la structure complète de l'intégrase. En fait, étant donné que nos inhibiteurs agissent à l'interface ADN-IN, il est primordial de considérer l'intasome du VIH en entier. Pour ce faire, nous avons procédé à des simulations de dynamique moléculaire conventionnelle des complexes comportant les trois partenaires ; nouvelles drogues – ADN viral – Intégrase, afin d'élucider leur mécanisme d'action *in silico*, ce qui nous permettra en premier lieu de parfaire la compréhension du mécanisme de l'intégration ensuite de mieux explorer le mode d'inhibition du VIH-1 par les INSTIs. En effet, ces dynamiques contribueraient à expliquer la meilleure efficacité du DTG par rapport aux INSTIs précédents puis de comparer l'effet thérapeutique de nos nouveaux dérivés à celui du DTG. A noter que nous passons d'une échelle de cinquantaine d'atomes correspondant aux petits complexes étudiés antérieurement à de larges systèmes de 200 000 atomes ;

l'étude quantique sera alors très coûteuse en temps de calcul, d'où l'intérêt d'avoir recours à la dynamique moléculaire.

B.II. Présentation des systèmes à simuler

Outre le DTG, nous avons choisi pour la suite, parmi les treize nouveaux noyaux conçus, les trois molécules B3, A2 et B2 présentant le meilleur rendement énergétique au niveau de leur interaction avec C₁₆ et G₄. Pour rappel, le protocole utilisé pour les simulations de dynamique moléculaire conventionnelle a été largement détaillé dans le chapitre III Partie B.III. Le choix de B3, A2 et B2 a été appuyé par le fait qu'ils présentent des substitutions halogénés simples de chlore et de brome comparées aux substitutions complexes de cyanure et de CF₃ trouvées chez E2 et C2 respectivement, qui sont également pertinents au niveau énergétique. En effet, le but ultime de nos travaux de thèse étant de concevoir et de tester expérimentalement nos nouvelles molécules (tests *in vitro* et éventuellement *in vivo*), nous avons choisi les meilleurs dérivés au niveau énergétique ayant une structure adéquate permettant leur synthèse et leur administration aux patients. Par conséquent, nous avons substitué le noyau du DTG par le noyau de B3, A2 et B2 en gardant les mêmes coordonnées pour le reste de la drogue au niveau du complexe 3S3M. Un cinquième dérivé a été ajouté, identique à la molécule de DTG mais dépourvu du noyau halogéné (Figure 48) ; c'est une molécule « contrôle » qui permettrait de mettre en relief l'implication du noyau halogéné dans les interactions de la drogue au sein de l'intasome.

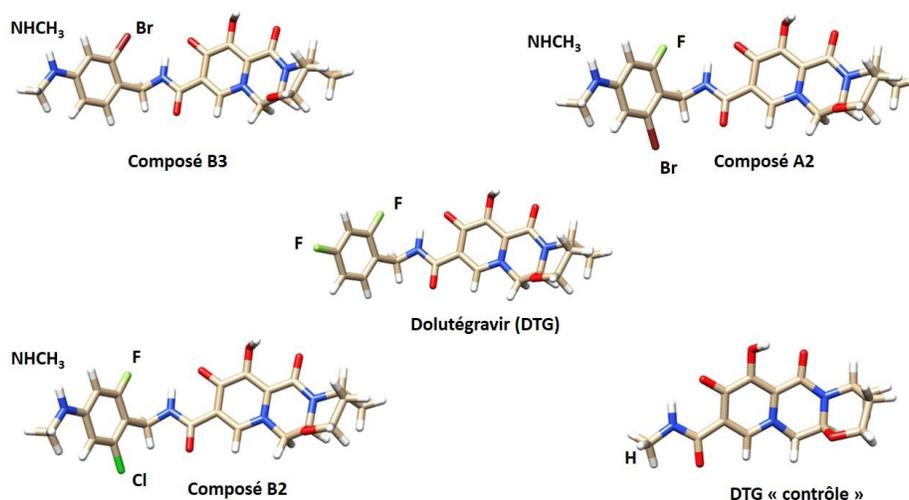


Figure 48 : Structures des cinq inhibiteurs étudiés par dynamique moléculaire. Le code couleur reste le même que celui des Figures 39 et 40. Les substitutions au niveau des noyaux halogénés sont indiquées en gras et en noir ; Br correspond au brome, F au fluor, Cl au chlore et H à l'hydrogène.

Nos cinq dérivés ont donc le même échafaud avec des noyaux halogénés différents. (Figure 49).

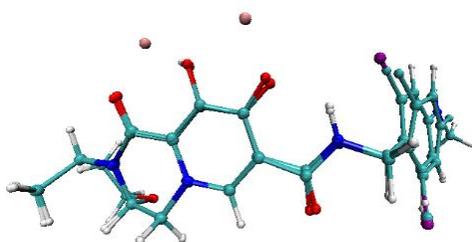


Figure 49 : Superposition des structures de nos cinq ligands avec le logiciel de visualisation VMD (Visual Molecular Dynamics). Ils présentent le même échafaud (même coordonnées) mais des substitutions différentes au niveau des noyaux halogénés (coordonnées différents). Les deux ions de Mg^{2+} chélatés par les drogues sont représentés par des sphères roses. Le code couleur des atomes est le même que celui adopté précédemment.

Nous avons alors cinq complexes initiaux de départ ; ADNv-Intégrase-DTG, ADNv-Intégrase-B3, ADNv-Intégrase-A2, ADNv-Intégrase-B2 et ADNv-Intégrase-DTG « contrôle ». Comme mentionné précédemment et afin de compléter la compréhension de l'intégration virale ainsi que le mécanisme d'inhibition par DTG, nous avons ajouté trois complexes supplémentaires à simuler; ADNv-DTG Intégrase-DTG et Intégrase-ADNv. Les deux premiers ont été obtenus à partir de la

structure 3S3M en éliminant respectivement l'intégrase et l'ADNv. Pour le complexe Intégrase-ADNv, nous avons adopté la structure PDB de code 3L2R du cristal de l'intasome du PFV en présence des ions Mg^{2+} . Cette structure apo (sans ligands) nous permettra de comparer l'interaction entre l'IN et l'ADNv en absence de tout ligand ce qui soulignera l'effet de ce dernier sur le complexe bipartite. Notons que chacun de nos systèmes fait à peu près 200 000 atomes.

Les Figures 50, 51 et 52 illustrent les positions initiales des ligands construits et situés au niveau de leur site d'action, à l'interface ADNv-IN. Nous retrouvons à la Figure 50 le complexe ternaire ; ADN-Intégrase-DTG et le complexe ADNv-IN en l'absence du DTG ; en comparant les deux sites, nous remarquons comment en l'absence d'inhibiteur, A_{17} présente des empilements avec C_{16} au niveau des noyaux de leurs squelettes, alors que dans le cas du complexe ternaire, l'inhibiteur se situe au niveau de la poche occupée précédemment par A_{17} . Par conséquent, il effectue des empilements avec les noyaux de C_{16} et A_{17} via son noyau halogéné et les noyaux de son squelette respectivement, avec son fluor en para pointé vers G_4 .

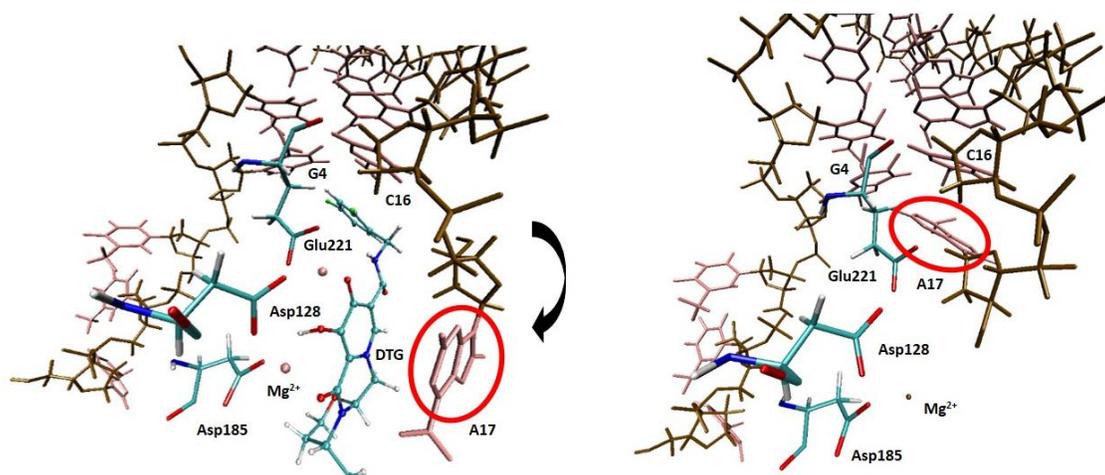


Figure 50 : Positions de départ des structures des complexes ADN-Intégrase-DTG (à gauche) et ADN-Intégrase (à droite) pour le Prototype Foamy Virus (PFV). L'ADN double brin est représenté en bâtons de couleur rose et marron, le DTG est représenté en boules pour les atomes et en bâtons pour les liaisons entre eux, avec les cations divalents de magnésium (Mg^{2+}) représentés en sphères roses. La triade catalytique de l'intégrase est schématisée en bâtons, avec Asp128, Asp185 et Glu221 correspondant à Asp64, Asp185 et Glu152 du VIH-1 respectivement. Le code couleur pour le DTG et pour la triade catalytique reste le même que celui utilisé précédemment pour les Figures 39 et 40.

La Figure 51 schématise la position initiale du DTG en présence de l'ADN seul et de la protéine seule respectivement, et finalement la Figure 52 présente les positions de départ des ligands DTG « contrôle et B3. Celles correspondant aux dérivés A2 et B2 au sein des complexes ternaires sont présentées en annexe vu qu'elles sont très similaires à la position de départ du composé B3 (Figure 120 en annexe).

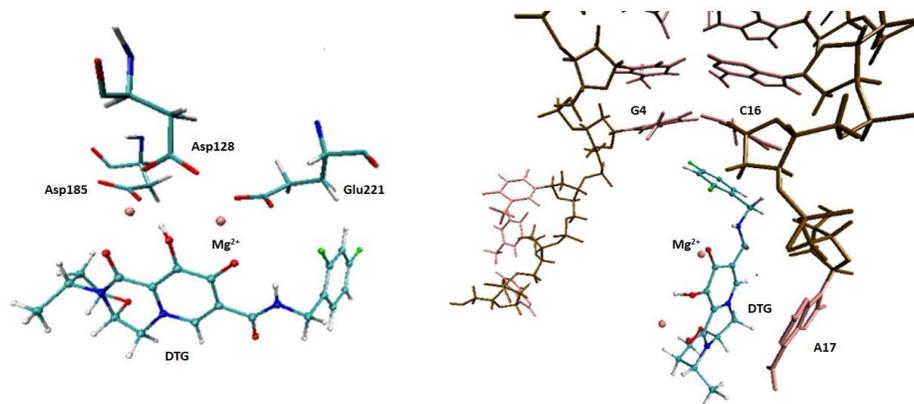


Figure 51 : Positions de départ des structures des complexes Intégrase-DTG (à gauche) et ADN-DTG (à droite) pour le Prototype Foamy Virus (PFV). L'ADN double brin est représenté en bâtons de couleur rose et marron, le DTG est représenté en boules pour les atomes et en bâtons pour les liaisons entre eux, avec les cations divalents de magnésium (Mg^{2+}) représentés en sphères roses. La triade catalytique de l'intégrase est schématisée en bâtons, avec Asp128, Asp185 et Glu221 correspondant à Asp64, Asp185 et Glu152 du VIH-1 respectivement. Le code couleur pour le DTG et pour la triade catalytique reste le même que celui utilisé précédemment pour les Figures 39 et 40.

Les nouveaux dérivés s'empilent aussi avec C_{16} et présentent leur groupement $NHCH_3$ pointé vers G_4 .

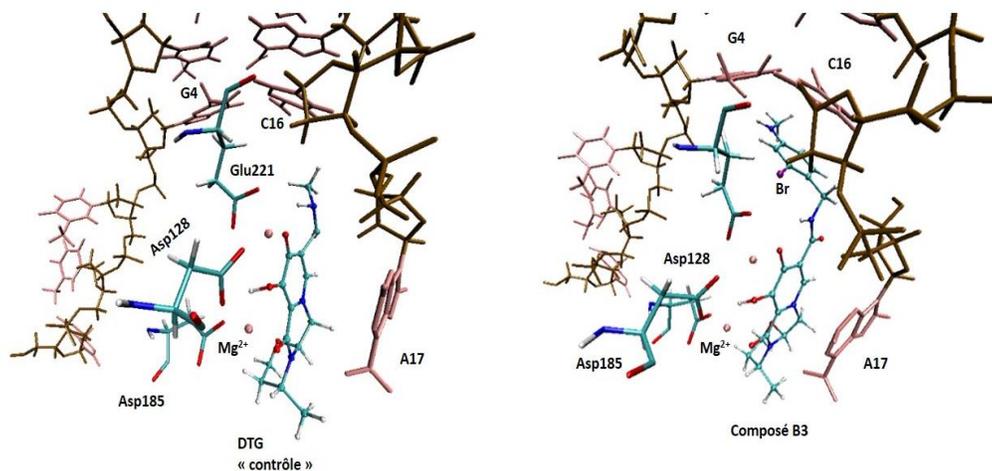


Figure 52 : Positions de départ des structures des complexes ADN-Intégrase-DTG « contrôle » (à gauche) et ADN-Intégrase-B3 (à droite) pour le Prototype Foamy Virus (PFV). L'ADN double brin est représenté en bâtons de couleur rose et marron, le DTG contrôle et B3 sont représentés en boules pour les atomes et en bâtons pour les liaisons entre eux, avec les cations divalents de magnésium (Mg^{2+}) représentés en sphères roses. La triade catalytique de l'intégrase est schématisée en bâtons, avec Asp128, Asp185 et Glu221 correspondant à Asp64, Asp185 et Glu152 du VIH-1 respectivement. Le code couleur pour le DTG et pour la triade catalytique reste le même que celui utilisé précédemment pour les Figures 39 et 40, sauf pour le brome coloré en violet dans ce cas.

B.III. Analyse des simulations de dynamique moléculaire conventionnelle

Nos avons effectué de longues simulations de dynamique moléculaire pour chaque système, d'une durée totale de 400 ns, à une température de 300K, en NVT, avec un pas de temps de 10 fs. A noter que nos simulations sont une première à plusieurs niveaux ; la simulation du complexe entier ADNv-IN-DTG du PFV ou de VIH-1 par dynamique moléculaire n'a jamais été faite auparavant ainsi que les simulations des drogues inhibitrices issues du DTG, en champ de force polarisable (AMOEBA) et en utilisant le logiciel Tinker-HP. Nous avons procédé ensuite à analyser nos simulations en se focalisant sur des interactions clés qui s'effectuent à plusieurs niveaux ; entre des résidus importants de l'ADN et les inhibiteurs, entre des résidus actifs de l'intégrase et les inhibiteurs et entre l'ADN viral et l'intégrase en absence et en présence de molécules inhibitrices. Ces analyses nous permettront de mieux comprendre les points suivants :

- Le mécanisme d'action du DTG à l'interface ADNv-IN, en comparant notamment les dynamiques des systèmes DTG-ADNv-IN (système 3S3M) et ADNv-IN (système 3L2R).
- Le mécanisme d'action des nouveaux inhibiteurs B3, A2 et B2 à l'interface ADNv-IN en évaluant leur rendement énergétique déjà observé au niveau des mini-systèmes par calcul de ΔE au niveau du complexe entier.
- Le rôle que joue le noyau halogéné des inhibiteurs et ceci par comparaison des simulations des complexes ADNv-IN-inhibiteurs avec le complexe ADNv-IN-DTG « contrôle ».

B.III.1. Etude des interactions ADN viral – Intégrase

La première partie, dans le cadre des analyses de nos simulations de dynamique conventionnelle, s’articule autour de l’étude des interactions entre des résidus de l’intégrase et des bases de l’ADN viral dans les complexes suivants : ADN-Intégrase (Code PDB : 3L2R), ADN-Intégrase-DTG (Code PDB : 3S3M), ADN-Intégrase-B3, ADN-Intégrase-A2 et ADN-Intégrase-B2, afin de : (i) compléter notre compréhension du processus de l’intégration et plus particulièrement des sites d’interaction au niveau des deux partenaires ; (ii) évaluer l’effet des inhibiteurs à l’interface ADN-Intégrase sur le comportement des deux partenaires et la formation de leur complexe au niveau de ces sites d’interaction. Pour rappel, les résidus de l’intégrase impliqués directement dans les interactions avec l’ADN viral se trouvent majoritairement au niveau de l’hélice $\alpha 4$ de l’intégrase (Résidus Ser147 \rightarrow Arg 166) ainsi qu’au niveau de la boucle en amont de cette hélice (Résidus Phe139 \rightarrow Gln146) (Figure 53) [11].

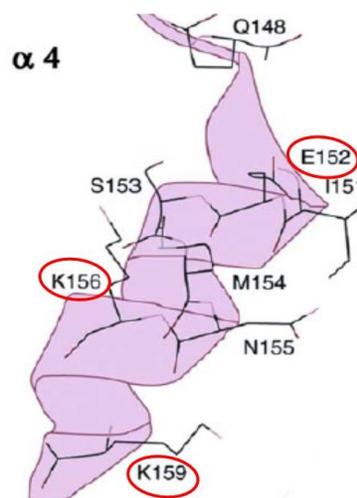


Figure 53 : Schéma illustratif de l’hélice $\alpha 4$ du VIH-1, où sont indiqués les résidus hydrophiles situés à sa surface. Nous avons encadré en rouge les résidus hydrophiles dont les interactions avec l’ADN viral ont été étudiées. La Glycine en position 149 (G149) et la Lysine 160 (K160) étudiées de même, manquent dans le schéma.

Pour les bases de l'ADN, nous rappelons que celles impliquées dans les liaisons avec l'intégrase sont notamment les six dernières paires de base au niveau 3' du brin clivé ainsi que les résidus à l'extrémité 5' du brin viral non clivé, complémentaires à ceux en 3' du brin clivé (Figure 16). Les liaisons étudiées sont décrites ci-dessous et rassemblées dans le Tableau 7.

- Distance entre l'hydrogène saillant d'A₁₇ et l'oxygène de Glu152 (Figure 54). Il faut noter qu'un empilement existe entre C₁₆ et A₁₇ au niveau de l'extrémité 3', en absence d'inhibiteur, avec le groupement réactif OH de l'adénine qui pointe vers le site catalytique de l'intégrase pour déclencher l'étape de transfert de brin [12]. Cependant, en présence du DTG, la poche occupée précédemment par l'adénine sera libérée permettant à la drogue de se situer à l'interface ADN-IN et d'effectuer un empilement avec C₁₆. 3'OH de l'Adénine est ainsi éloigné du site catalytique et la réaction de transfert de brin est alors inhibée [16].

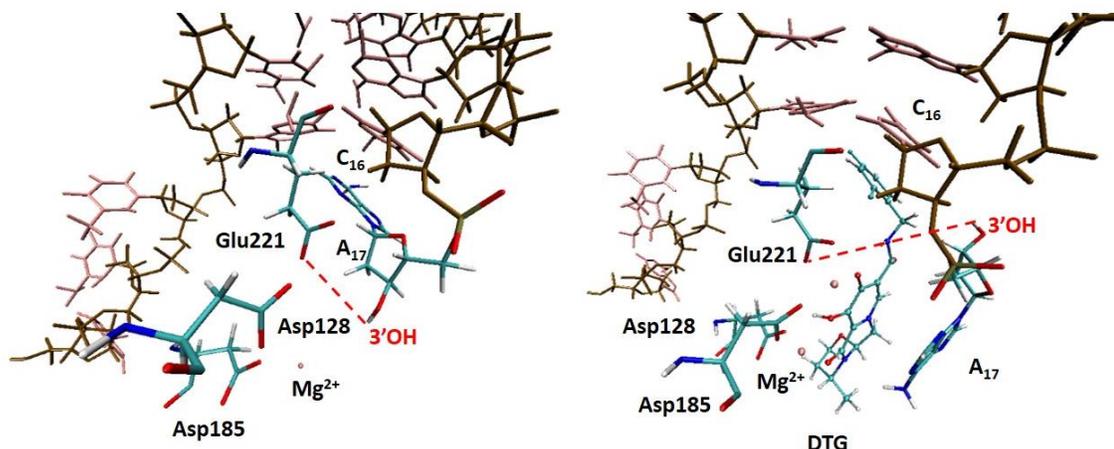


Figure 54 : Positions de départ de nos systèmes simulés ; le complexe ADN-Intégrase du PFV (à gauche) et le complexe ADN-Intégrase-DTG (à droite). Les atomes étudiés sont l'hydrogène (en blanc) d'A₁₇ et l'oxygène (en rouge) du résidu catalytique Glu221 correspondant au Glu152 du VIH-1. La distance entre eux est représentée par une ligne pointillée rouge. L'empilement entre les bases C₁₆ et A₁₇ peut être observé dans la figure à gauche, alors que dans la figure à droite, nous observons le DTG enfoncé dans la poche occupée précédemment par A₁₇, éloignant son extrémité 3'OH du site actif de l'intégrase et effectuant à sa place des empilements π-π avec la cytosine. Les cations divalents de Mg²⁺ sont représentés par des sphères roses (deux sont chélatés par le DTG à droite alors que dans le complexe ADN-Intégrase un cation a été identifié uniquement). Les deux autres résidus de la triade catalytique sont aussi représentés en bâtons ; Asp128 (Asp64 du VIH-1) et Asp185 (Asp116 du VIH-1). Le code couleur est le même que celui des Figures 50, 51 et 52 avec l'ADN viral en bâtons de couleur rose et marron et le DTG en bâtons et boules avec le même code couleur que celui des figures 34 et 35. Nous présentons uniquement les positions initiales dans le cas de DTG puisque celles de B3, A2 et B2 lui sont identiques.

- Liaison hydrogène entre la Glycine 149 (conservée chez le VIH-1 et le PFV sous forme de Gly218) et la base G₄ en 5' du brin d'ADN viral non clivé, comme démontré dans le cristal 3L2R de Hare et al ^[12] (Figure 55 et Figure 121 en annexe).

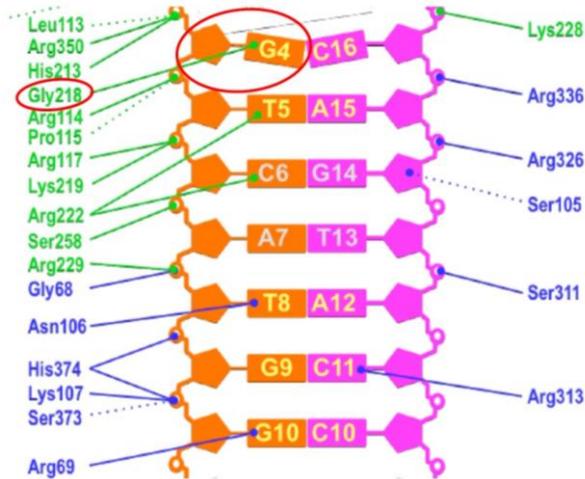


Figure 55 : Schéma représentant les contacts ADN-protéine identifiés dans le complexe de code PDB : 3L2R par Hare et ses collaborateurs. Le double brin d'ADN viral est représenté en orange pour le brin non traité et en magenta pour le brin traité. Les traits pleins indiquent les liaisons hydrogènes entre le résidu de l'intégrase et la base virale alors que les traits pointillés correspondent aux interactions de vdW. Nous avons encerclé la G₄ dont les interactions avec la Glycine 218 (Gly 218) correspondant à Gly149 chez le VIH-1, ont été suivies.

- Distance entre la Lysine 156 (K156) et C₁₆ en face d'elle (Figure 122 en annexe). La Sérine 225 (S225) est l'équivalent de K156 chez le PFV.
- Liaison hydrogène entre la Lysine 159 (K159) et le groupement phosphate situé entre les deux bases actives C₁₆ et A₁₇ du brin traité (Figure 56 et Figure 122 en annexe).

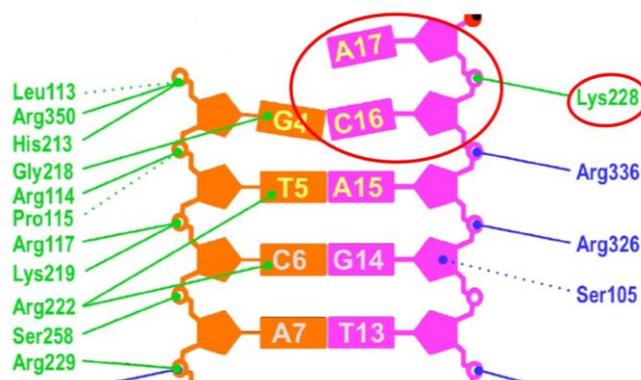


Figure 56 : Schéma représentant les contacts ADN-protéine identifiés dans le complexe de code PDB : 3L2R. Le double brin d'ADN viral est représenté en orange pour le brin non traité et en magenta pour le brin traité. Les traits pleins indiquent les liaisons hydrogènes entre le résidu de l'intégrase et la base virale alors que les traits pointillés correspondent aux interactions de vdW. Nous avons encerclé les bases A₁₇ et C₁₆ dont les interactions avec la Lysine 228 (Lys 228) correspondant à Lys159 chez le VIH-1, ont été suivies.

- Liaison hydrogène entre la Lysine 160 (K160) dont l'équivalent chez le PFV est l'Arginine 229 (R229) et le groupement phosphate situé entre les bases A₇ et T₈ du brin d'ADN non traité (Figures 122 et 123 en annexe). Ajoutons que cette triade de Lysine a été identifiée comme très importante au niveau des interactions ADN-Intégrase, par des études antérieures effectuées entre autres par notre équipe ^{[13] [14] [15]}.
- La dernière interaction analysée associe la Glutamine 146 (Gln146), conservée chez l'intégrase du VIH-1 et celle du PFV où elle correspond à la Glutamine 215 (Gln215) et A₁₇. Gln146 située dans la boucle en amont de l'hélice α ₄, déplace la Thymine T₃ du brin d'ADN viral non clivé afin d'effectuer des liaisons hydrogènes avec l'Adénine A₁₇ complémentaire à T₃, et faciliter ainsi l'étape du 3'processing (Figure 124 en annexe).

L'étude de ces interactions ainsi que celles présentées par la suite ont été réalisées en utilisant un algorithme développé au sein de notre laboratoire, qui génère les résultats sous forme de deux graphiques ; le premier présente l'évolution de la distance en Å en fonction des trajectoires (frames) de nos simulations de dynamique moléculaire. A noter que toutes nos simulations de 400 ns comportent 2000 frames et qu'au niveau de l'axe des abscisses, chaque 100 frames correspondent à 20 ns. Le deuxième graphique obtenu indique la variation de la densité des points correspondant à nos trajectoires en fonction des distances. Il permet de déterminer la distance la plus récurrente observée pour la majorité des positions. Ce graphique donne de même une idée sur la stabilité des systèmes simulés selon l'aspect des courbes obtenues. Les figures ci-dessous illustrent nos résultats. Ceux concernant les composés ADN-Intégrase-A2 et ADN-Intégrase-B2 sont en annexe, étant similaires aux résultats de B3 (Figures 125 à 129 en annexe).

Ajoutons que les résultats pour le suivi de la liaison entre K160 et les bases virales A7-T8 sont mises en annexe aussi (Figures 129 à 131 en annexe).

Tableau 7 : Tableau résumant les interactions étudiées entre les résidus de l'intégrase du VIH-1 et les bases virales de l'ADN du VIH-1. Nos modèles d'études sont les systèmes ADN-Intégrase et ADN-Intégrase-Inhibiteurs.

Résidus au niveau de l'intégrase du VIH-1	Bases au niveau l'ADN viral du VIH-1
Glu152 (hélice α 4)	A ₁₇ (3' brin viral traité)
Gly149 (hélice α 4)	G ₄ (5' brin viral non traité)
Lys159 (hélice α 4)	Groupement phosphate entre C ₁₆ -A ₁₇ en 3' du brin viral non traité
Lys160 (hélice α 4)	Groupement phosphate entre A ₇ -T ₈ en 5' du brin viral non traité
Lys156 (hélice α 4)	C ₁₆ (3' brin viral traité)
Gln146 (boucle en amont de l'hélice α 4)	A ₁₇ (3' brin viral traité)

En observant les graphiques de la Figure 57 nous déduisons la présence de contacts proches (3,6 Å et 4,2 Å) au cours des 400 ns de simulation au sein du complexe ADN-Intégrase, entre l'hydrogène réactif terminal d'A₁₇ et Glu152.

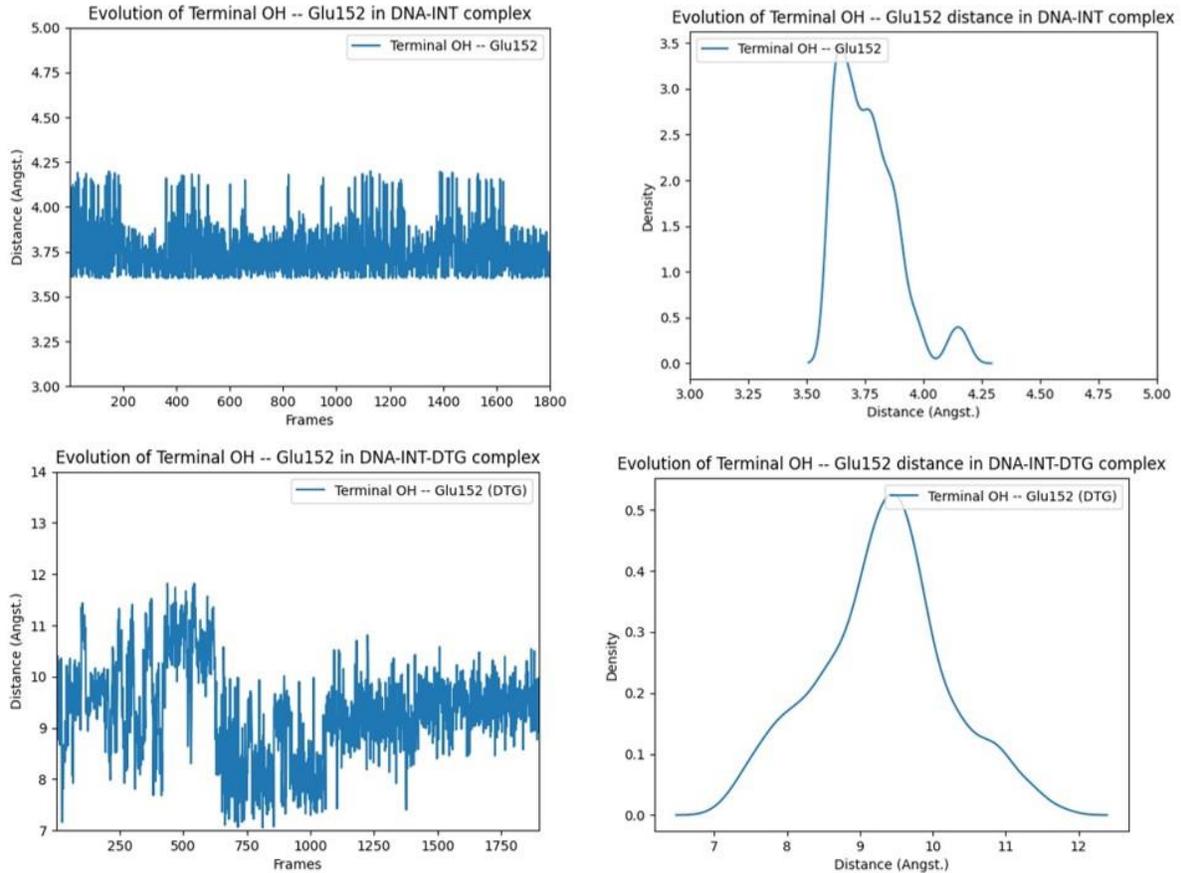


Figure 57 : Graphiques représentant l'évolution de la distance entre 3'OH et Glu152 du VIH-1 (Glu221 dans le cas du PFV) dans le complexe ADN-Intégrase (à gauche en haut) et dans le complexe ADN-Intégrase - DTG (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas du complexe ADN-Intégrase et en bas dans le cas du complexe ADN-Intégrase-DTG).

Alors qu'en présence de DTG et de B3, nous confirmons que l'éloignement de l'hydrogène terminal du Glutamate 152 est maintenu tout au long de nos simulations, avec des variations de distance entre 9 Å et 11 Å jusqu'à 50 ns de simulation, suivies d'une diminution des distances à 8-9 Å avant d'atteindre des distances stables aux alentours de 9,5 Å à 220 ns (Figures 57 et 58).

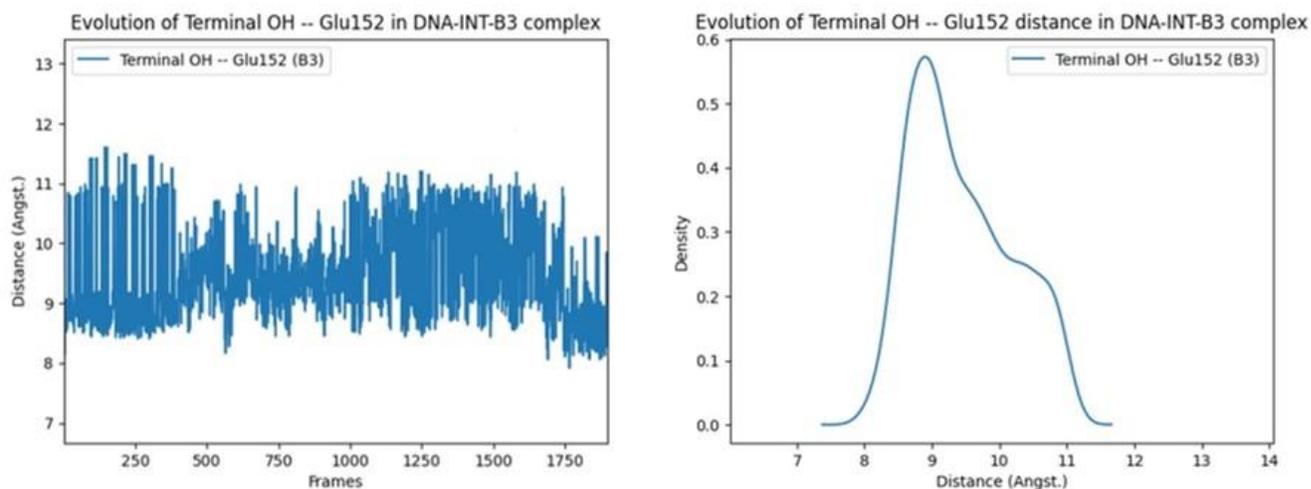


Figure 58 : Graphique à gauche représentant l'évolution de la distance entre 3'OH et Glu152 du VIH-1 (Glu221 dans le cas du PFV) dans le complexe ADN-Intégrase-B3 en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons le graphique de la densité des pauses en fonction des distances obtenues.

Pour les interactions Gly149-G₄, dans le complexe ADN-Intégrase, la liaison hydrogène est conservée tout au long des simulations, variant entre 2,8 Å et 3,8 Å, avec la plupart des distances aux alentours de 3,2 Å (Figure 59, graphiques en haut). Alors qu'en présence de DTG et de B3, la liaison hydrogène n'est préservée que pendant 50 ns de simulation. Elle disparaît par la suite avec une variation des distances entre 4,5 Å et 5,5 Å dans le cas de DTG (Figure 59, graphiques en bas) et entre 5-6 Å dans le cas de B3 (Figure 60).

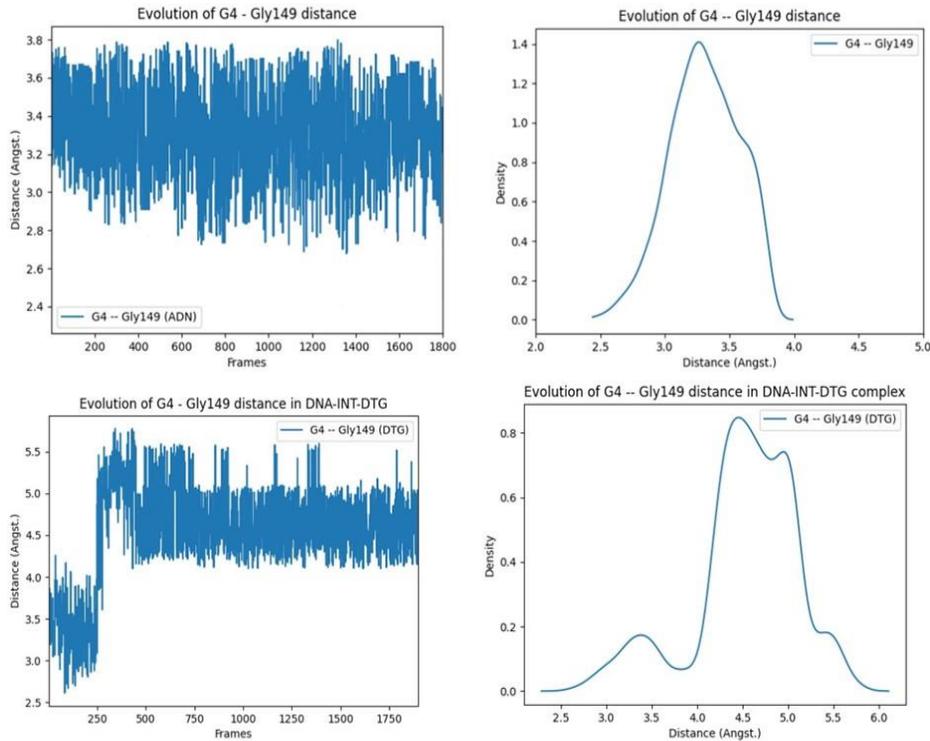


Figure 59 : Graphiques représentant l'évolution de la distance entre G_4 et Gly149 du VIH-1 (Gly218 dans le cas du PFV) dans le complexe ADN-Intégrase (graphique en haut à gauche) ainsi que dans le complexe ADN-Intégrase-DTG (graphique en bas à gauche) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues dans le cas des deux complexes ; en haut pour le complexe binaire et en bas pour le complexe ternaire.

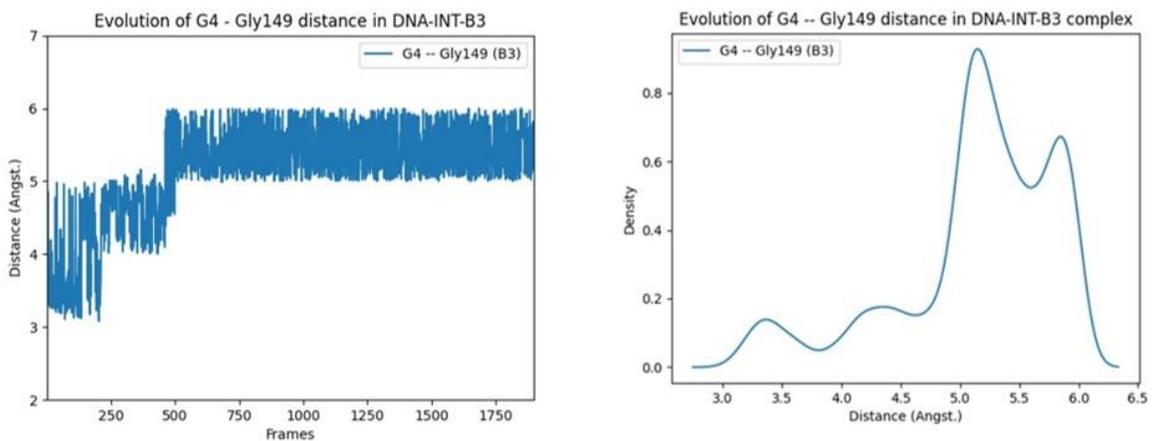


Figure 60 : Graphique à gauche représentant l'évolution de la distance entre G_4 et Gly149 du VIH-1 (Gly218 dans le cas du PFV) dans le complexe ADN-Intégrase-B3 en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons le graphique de la densité des pauses en fonction des distances obtenues.

Pour les interactions avec la Lysine, La liaison hydrogène entre Lys156 et C₁₆ est maintenue au cours de nos simulations dans le cas du complexe ADN-Intégrase avec une distance majoritairement de 2,9 Å (Figure 61, graphiques en haut).

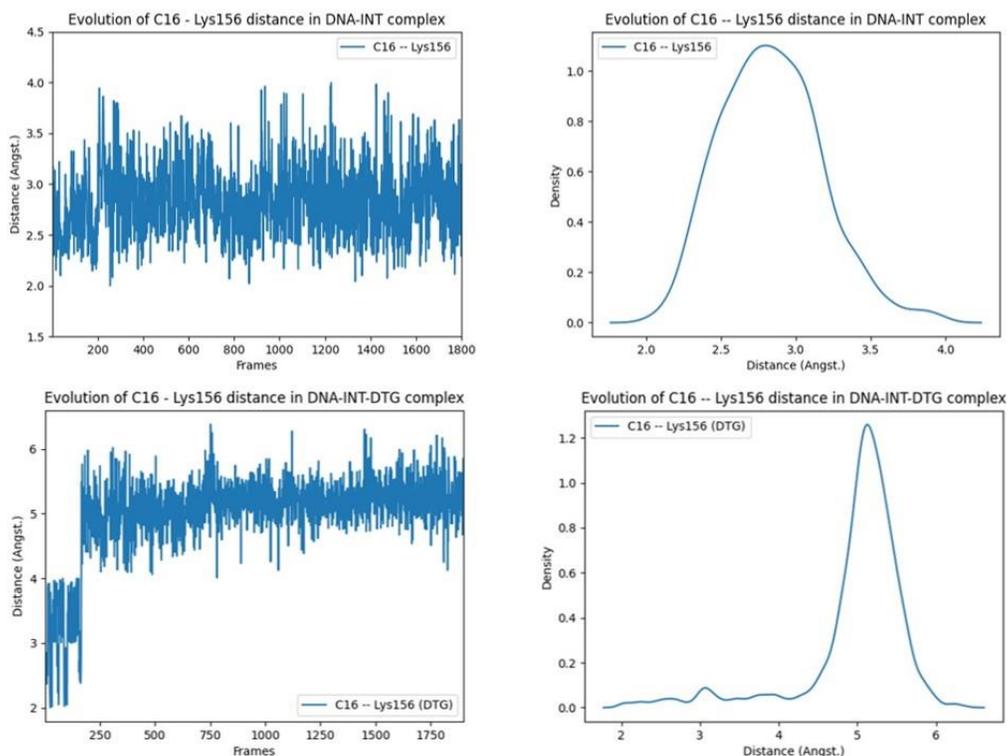


Figure 61 : Graphiques représentant l'évolution de la distance entre C₁₆ et Lys156 du VIH-1 (Ser225 dans le cas du PFV) dans le complexe ADN-Intégrase (à gauche en haut) ainsi que dans le complexe ADN-Intégrase-DTG (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues dans le cas des deux complexes ; en haut pour le complexe ADN-Intégrase et en bas pour le complexe en présence du DTG.

En ajoutant nos inhibiteurs ; DTG et B3 au complexe, cette liaison hydrogène est maintenue pendant 50 ns de simulation (valeurs entre 2,5 Å et 3,5 Å) avant que les deux résidus ne s'éloignent pour atteindre une distance approximative de 5 Å jusqu'à la fin de la simulation (Figure 61, graphiques en bas et Figure 62).

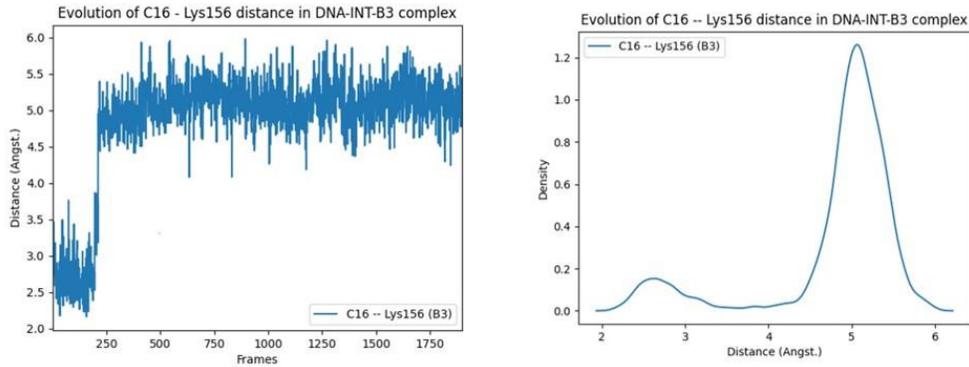


Figure 62 : Graphique représentant l'évolution de la distance entre C₁₆ et Lys156 du VIH-1 (Ser225 dans le cas du PFV) dans le complexe ADN-Intégrase-B3 en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons le graphique de la densité des pauses en fonction des distances obtenues.

Quant à la liaison entre la Lys159 et le squelette de C₁₆-A₁₇, elle reste stable pendant les 400 ns de simulation avec des valeurs variant entre 3,2 Å et 4 Å dont la majorité se situe aux alentours de 3,6 – 4 Å (Figure 63, graphiques en haut).

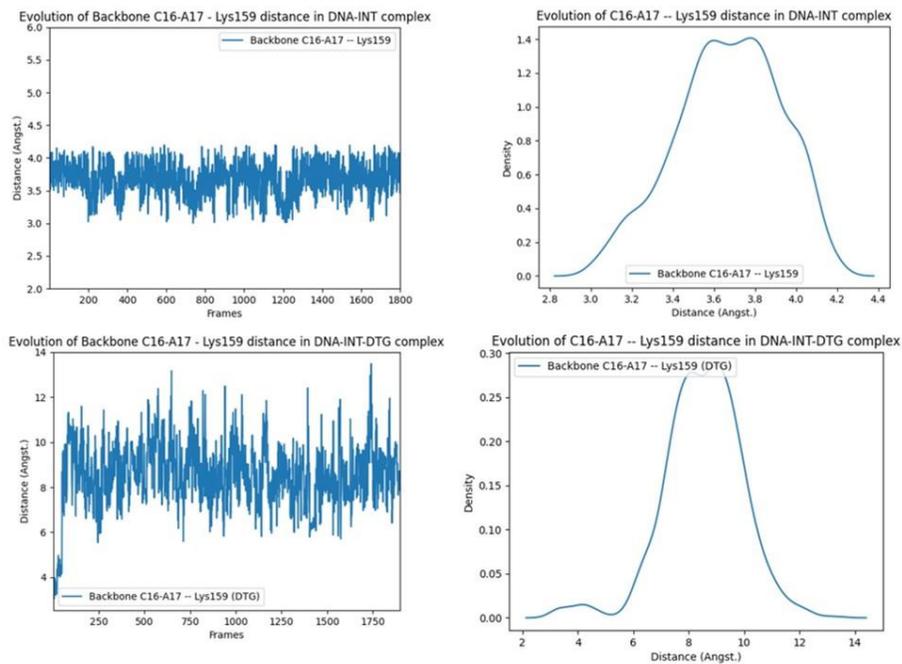


Figure 63 : Graphiques représentant l'évolution de la distance entre C₁₆-A₁₇ et Lys159 du VIH-1 (Lys228 dans le cas du PFV) dans le complexe ADN-Intégrase (à gauche en haut) ainsi que dans le complexe ADN-Intégrase-DTG (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues dans le cas des deux complexes ; en haut pour le complexe ADN-Intégrase et en bas pour le complexe en présence du DTG.

Alors qu'une fois le DTG et B3 présents, nous observons une perte de ces liaisons hydrogènes ; les distances augmentent pour atteindre des valeurs variant entre 6 Å et 11 Å pour DTG et entre 6-8 Å pour B3 avec des distances majoritaires de 9-10 Å (Figure 63, graphiques en bas et Figure 64).

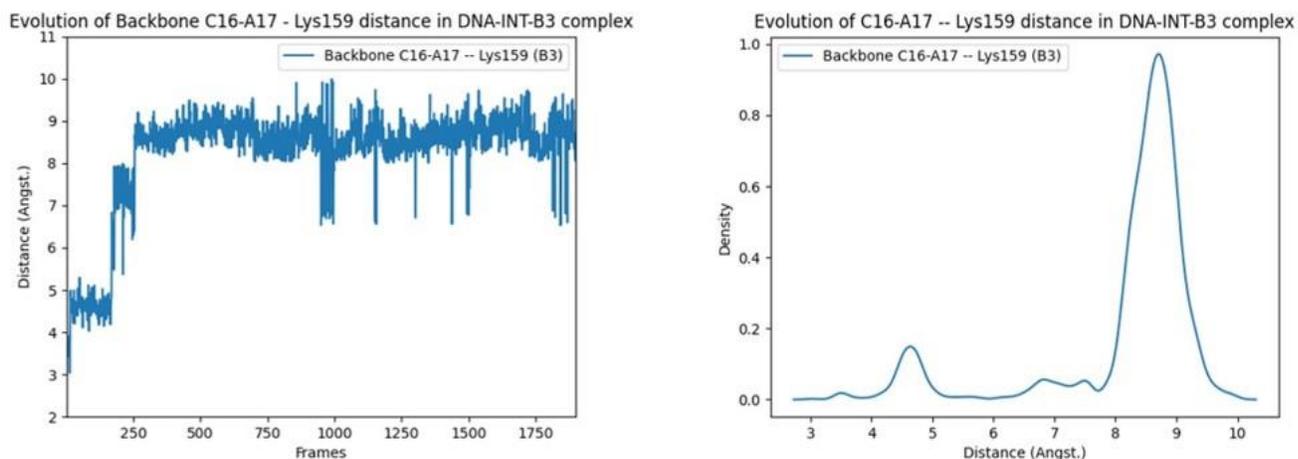


Figure 64 : Graphique représentant l'évolution de la distance entre C₁₆-A₁₇ et Lys159 du VIH-1 (Lys228 dans le cas du PFV) dans le complexe ADN-Intégrase-B3 en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons le graphique de la densité des pauses en fonction des distances obtenues.

Finalement, pour les interactions Gln146-A₁₇, la Figure 65 indique qu'au cours des 400 ns de simulation du complexe ADN-Intégrase, la liaison hydrogène est conservée entre Gln146 et A₁₇ avec la plupart des distances aux alentours de 2,8-2,9 Å.

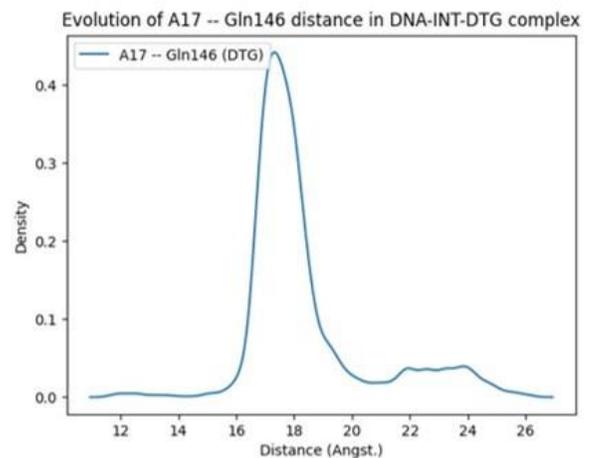
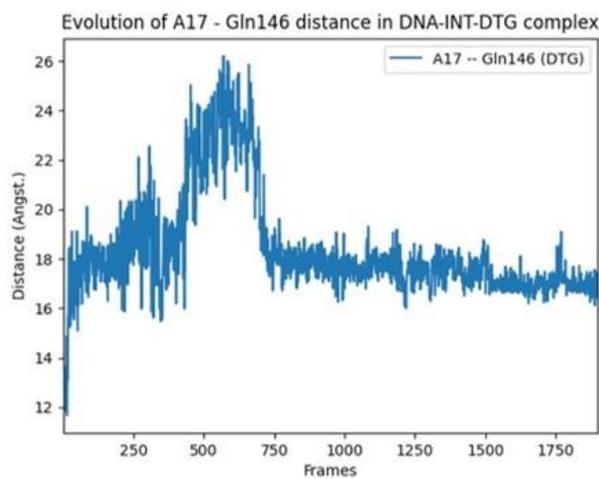
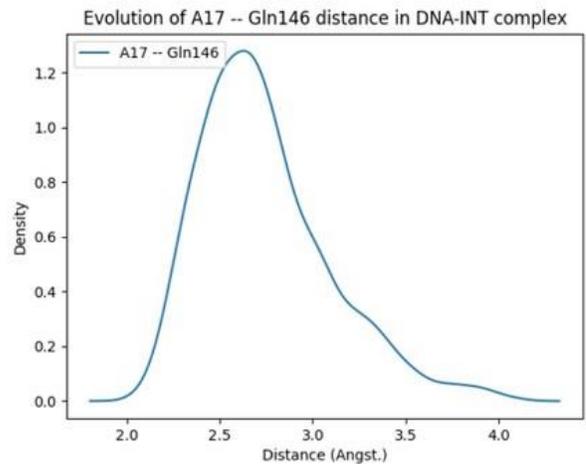
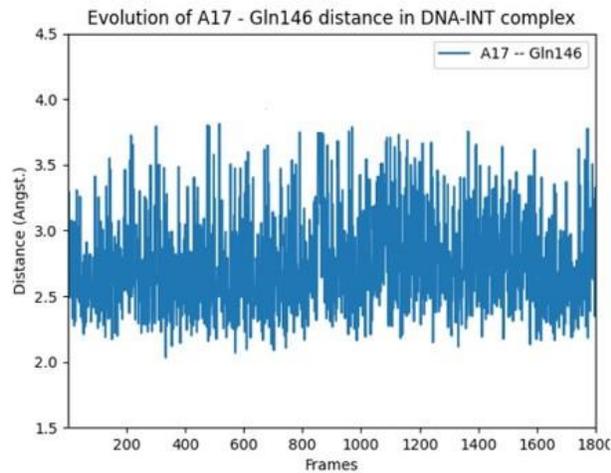


Figure 65 : Graphiques à gauche représentent l'évolution de la distance entre A₁₇ et Gln146 du VIH-1 (Gln215 dans le cas du PFV) dans le complexe ADN-Intégrase (en haut) et dans le complexe ADN-Intégrase-DTG (en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas du complexe ADN-Intégrase et en bas dans le cas du complexe ADN-Intégrase-DTG).

Alors que dans le complexe ADN-Intégrase-DTG et ADN-Intégrase-B3, (Figure 65, graphiques en bas et Figure 66), ces liaisons sont immédiatement perdues, où les distances entre l'acide aminé et la base virale deviennent majoritairement aux alentours de 18-19 Å après 250 ns de simulation.

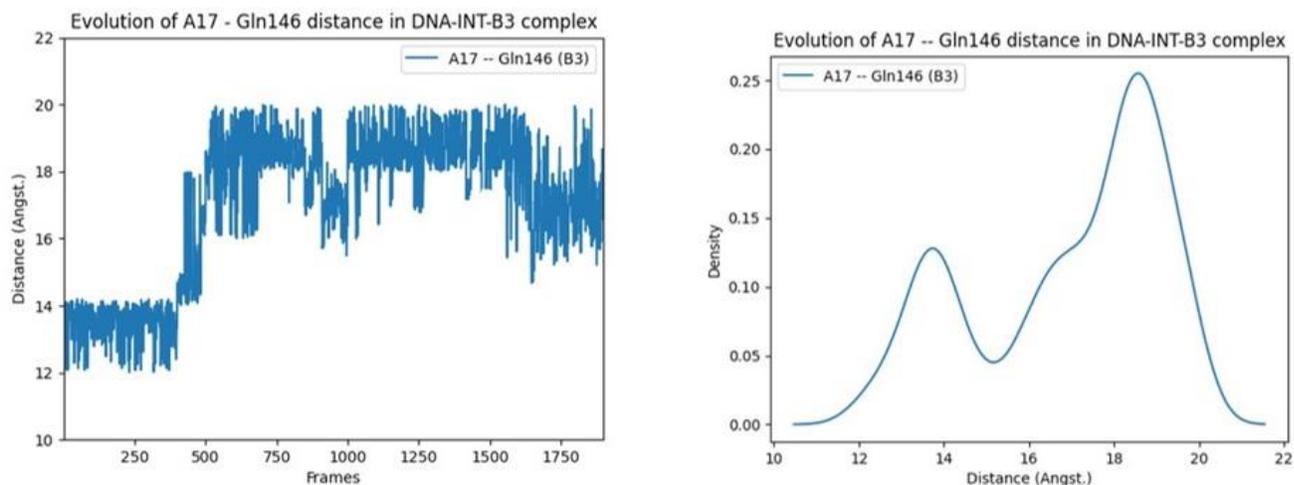


Figure 66 : Graphique à gauche représentant l'évolution de la distance entre A₁₇ et Gln146 du VIH-1 (Gln215 dans le cas du PFV) dans le complexe ADN-Intégrase-B3 en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons le graphique de la densité des pauses en fonction des distances obtenues.

Discussion des résultats obtenus pour les interactions ADN-intégrase en absence et en présence d'inhibiteurs

- Premièrement, les courbes et les graphiques obtenus présentent peu de fluctuations, indiquant que nos complexes sont caractérisés par des distances moyennes fiables et maintenues le long des simulations.
- En absence de tout inhibiteur et au sein du complexe ADN-Intégrase de code 3L2R, des contacts proches entre l'extrémité 3'OH d'A₁₇ et le résidu catalytique Glu152 sont conservées, alors qu'il y a perte de ces contacts en présence des quatre inhibiteurs, respectivement. Ces résultats sont très importants, puisqu'ils mettent en valeur le mécanisme d'inhibition de l'étape de transfert de brin par DTG et qui est maintenue par B3, A2 et B2. En fait, au cours de toutes les simulations, le DTG et les nouveaux inhibiteurs gardent leur position au niveau de la poche occupée antérieurement par A₁₇, remplaçant l'interaction du Glu152 avec l'extrémité 3'OH de l'Adénine. Ainsi, en éloignant cette dernière et diminuant ses interactions avec la triade

catalytique de l'IN, ils assurent une inhibition efficace de l'étape de transfert de brin.

- Les liaisons hydrogènes entre G₄ et Gly149 observées généralement au sein du complexe ADN-Intégrase, sont perdues en présence des inhibiteurs quelques ns après le début des simulations. Ceci serait dû au positionnement de DTG à proximité de Gly149 et G₄ et aux interactions établies entre nos inhibiteurs et G₄ (démontrées dans la partie A du chapitre III par les valeurs de ΔE) contribuant alors à la diminution des contacts entre la base virale et Gly149.
- DTG ainsi que les nouveaux inhibiteurs perturbent le réseau d'interactions établies entre la Lysine156 et C₁₆ ainsi qu'entre la Lysine 160 et le squelette A₇-T₈ puisqu'en absence d'inhibiteurs, dans le complexe ADN-Intégrase nous observons des liaisons hydrogènes stables et constantes au cours des simulations de dynamique moléculaire alors qu'une fois les inhibiteurs présents, ces contacts proches s'estompent entre la Lysine156 et C₁₆ et entre la Lysine160 et le squelette de l'ADN viral au niveau des bases A₇ et T₈. En fait, nos inhibiteurs interagissant par empilement π - π avec la Cytosine, affaibliraient ses interactions avec la Lysine 156.
- DTG et les nouveaux inhibiteurs influencent aussi les interactions entre la Lysine 159 et le squelette C₁₆-A₁₇ puisqu'ils affaiblissent les liaisons hydrogènes entre les deux. En effet et vu que, DTG, B3, A2 et B2 éloignent l'Adénine terminale (A₁₇) du site actif de l'intégrase et perturbent sa position initiale en s'enfonçant dans la poche qu'elle occupait préalablement, A₁₇ devient alors peu stable et présente une flexibilité importante ce qui perturbe ses interactions avec la Lysine159 au niveau du groupement

phosphate. De plus, par l'interaction des inhibiteurs avec C₁₆ (démontrées de même par les calculs de ΔE), les contacts entre l'adénine et Lys159 sont perturbées.

- En envahissant la poche préalablement occupée par A₁₇, DTG et les molécules qui en dérivent contribuent à éloigner aussi A₁₇ de la Gln146. Ajoutons les empilements des noyaux halogénés avec la base virale qui remplacent les interactions de cette dernière avec la Glutamine 146.

B.III.2. Etude des interactions ADN viral – Inhibiteurs

Nous avons continué par une étude approfondie des interactions établies entre nos inhibiteurs et des nucléotides actifs de l'ADN viral. Ces derniers correspondent aux bases C₁₆ et A₁₇, situées à l'extrémité 3' du brin d'ADN viral clivé où s'effectue la réaction de 3'processing catalysée par l'intégrase. On trouve aussi la Guanine G₄, complémentaire à C₁₆ au niveau de l'extrémité 5' du brin viral non clivé. Nous avons étudié, dans les systèmes ADN-Intégrase-inhibiteurs (Inhibiteurs correspondant au DTG, B3, A2 et B2), ADN-Intégrase-DTG « contrôle » et ADN-DTG, l'évolution des interactions suivantes :

- L'empilement π - π entre nos inhibiteurs et C₁₆ (Figure 67)
- L'empilement entre nos inhibiteurs et A₁₇ (Figure 50 à gauche).
- La distance entre G₄ et la substitution en para des inhibiteurs ; le fluor dans le cas du DTG, le NHCH₃ dans le cas de B3, A2 et B2 et l'hydrogène substituant le noyau halogéné dans le cas de la molécule « contrôle » (Figure 132 en annexe).

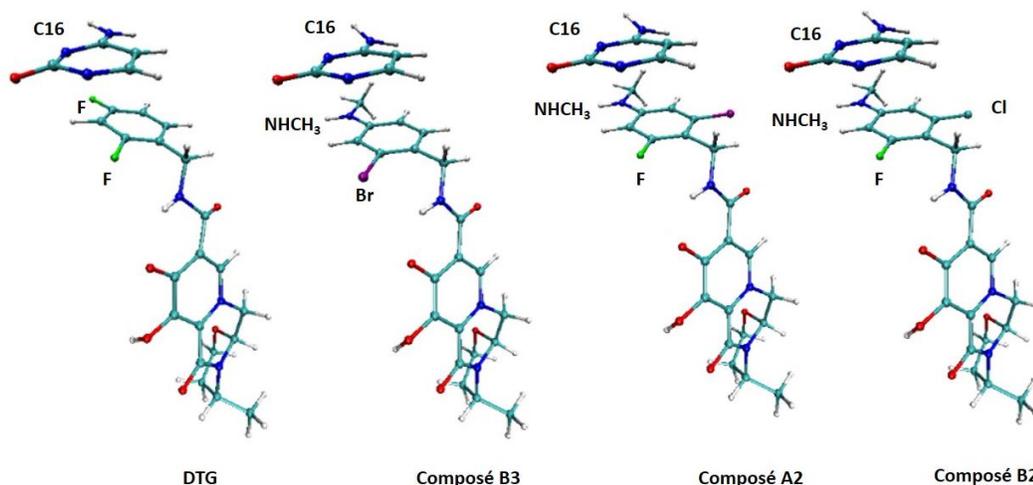


Figure 67 : Représentation de l'empilement entre le noyau de C₁₆ et celui de nos inhibiteurs ; Dolutégravir (DTG), B3, A2 et B2. Le code couleur est le même que celui adopté précédemment, avec les substitutions des noyaux des dérivés indiquées en gras et noir. Le chlore (Cl) dans le logiciel VMD est coloré de la même couleur que le carbone en cyan.

L'étude de la molécule « contrôle » mettrait en relief le rôle joué par le noyau halogéné dans les interactions avec l'ADNv alors que l'étude du complexe binaire ADN-DTG en l'absence d'intégrase nous permettrait de voir si l'absence de la protéine virale a une influence sur l'interaction de la drogue avec l'ADN viral.

Le suivi des empilements est réalisé grâce à un second algorithme développé au sein de notre laboratoire, qui effectue un suivi de l'indice d'empilement π (π -stacking index) entre la base virale et chaque noyau étudié, et attribue une valeur à cet indice, comprise entre 0 (absence d'empilement) et 0,6 (empilement parfait). Les résultats sont donnés sous forme d'un diagramme (Figures 68 et 69) qui représente la densité des points dont chacun correspond à une trajectoire (Density dans le graphique) en fonction de l'indice d'empilement π pour chaque noyau, pour une durée totale de simulation de 400 ns.

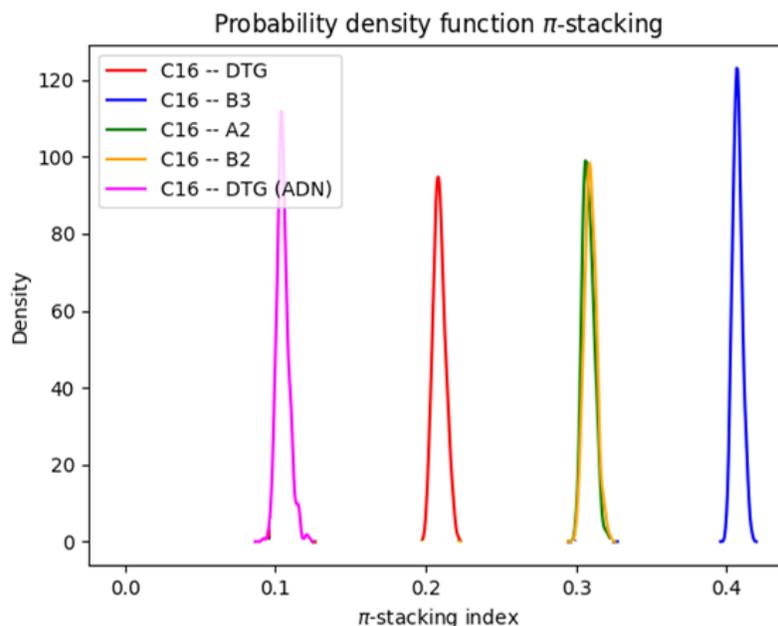


Figure 68 : Diagrammes représentant la variation de la densité des positions simulées en fonction de l'indice d'empilement π entre le noyau de C_{16} et celui des dérivés inhibiteurs pour une durée totale de simulation de 400 ns. Un indice de 0,1 est obtenu dans le cas du complexe DTG-ADN viral (courbe rose) et un indice de 0,2 est obtenu dans le cas de C_{16} avec DTG dans le complexe ternaire (courbe rouge). Pour les nouveaux dérivés, l'indice d'empilement dans le cas de A2 et B2 est le même, égal à 0,3 (courbes verte et jaune respectivement) et de 0,4 dans le cas du composé B3 avec C_{16} (courbe bleue).

Nous constatons suite à l'analyse de la Figure 68 que le composé B3 présente la valeur d'indice d'empilement la plus élevée, égale à 0,4 (courbe bleue). Il est suivi par les composés A2 et B2 ayant le même indice de 0,3 ; leurs courbes respectives verte et jaune sont superposées à la Figure 54. Le DTG a présenté les indices d'empilement π les plus faibles de 0,2 dans le complexe ternaire (courbe rouge) et de 0,1 quand il est présent avec l'ADN viral uniquement (courbe rose). Nous pouvons énoncer donc que le noyau halogéné de B3 effectue le meilleur empilement avec C_{16} , suivi d'A2 et B2 puis du DTG dans le complexe ternaire. Le DTG en présence de l'ADN viral seul effectue l'empilement le plus faible avec C_{16} appuyant l'efficacité de DTG à l'interface ADN-IN.

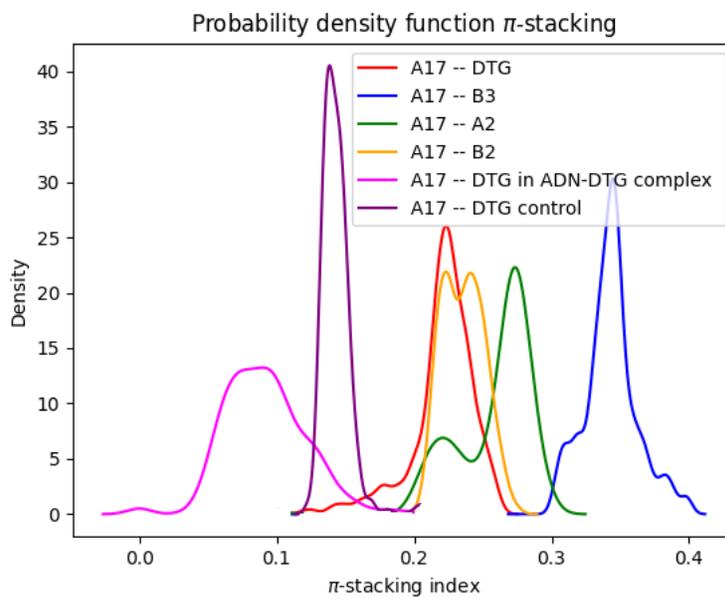


Figure 69 : Diagrammes illustrant la densité des points représentant nos trajectoires pour une simulation de 400 ns en fonction de l'indice d'empilement π entre le noyau d'A₁₇ et celui des dérivés inhibiteurs et du dérivé « contrôle ». Un indice de 0,1 est obtenu dans le cas du complexe DTG-ADN viral (courbe rose) et un indice de 0,21 est obtenu dans le cas d'A₁₇ avec DTG dans le complexe ternaire. Pour les nouveaux dérivés, l'indice d'empilement dans le cas d'A2 et B2 est de 0,28 et 0,25, respectivement (courbes verte et jaune respectivement) et il est égal à 0,35 dans le cas du composé B3 avec A₁₇. Le DTG « contrôle » quant à lui présente une valeur de 0,15.

Les diagrammes de la Figure 69 montrent que le composé B3 présente l'indice d'empilement avec A₁₇ le plus élevé, égal à 0,35 (courbe bleue), suivi par A2 puis B2 ayant des indices respectifs de 0,28 (courbe verte) et 0,25 (courbe jaune). Le DTG quant à lui présente les valeurs d'indice d'empilement les plus faibles ; de 0,21 dans le complexe ternaire (courbe rouge), de 0,1 avec l'ADN seul (courbe rose) et de 0,15 quand il est dépourvu du noyau d'halobenzène (courbe en violet).

Concernant les interactions inhibiteurs-G₄, Hare et ses collaborateurs ont en fait observé des contacts proches, aux alentours de 3 Å, entre le fluor en para du DTG et le carbone en position 6 (C6) de G₄ [16] au niveau du cristal 3S3M de l'intasome du PFV en présence du DTG et Mg²⁺. Nous avons donc choisi d'analyser l'évolution de la distance entre C6 de la guanine et les substitutions en para de nos dérivés ; le fluor du DTG, l'azote du groupement NHCH₃ de B3, A2 et B2 et l'hydrogène

substituant le noyau halogéné dans le cas du DTG «contrôle» (Figure 132 en annexe et Figures 70, 71 et 72).

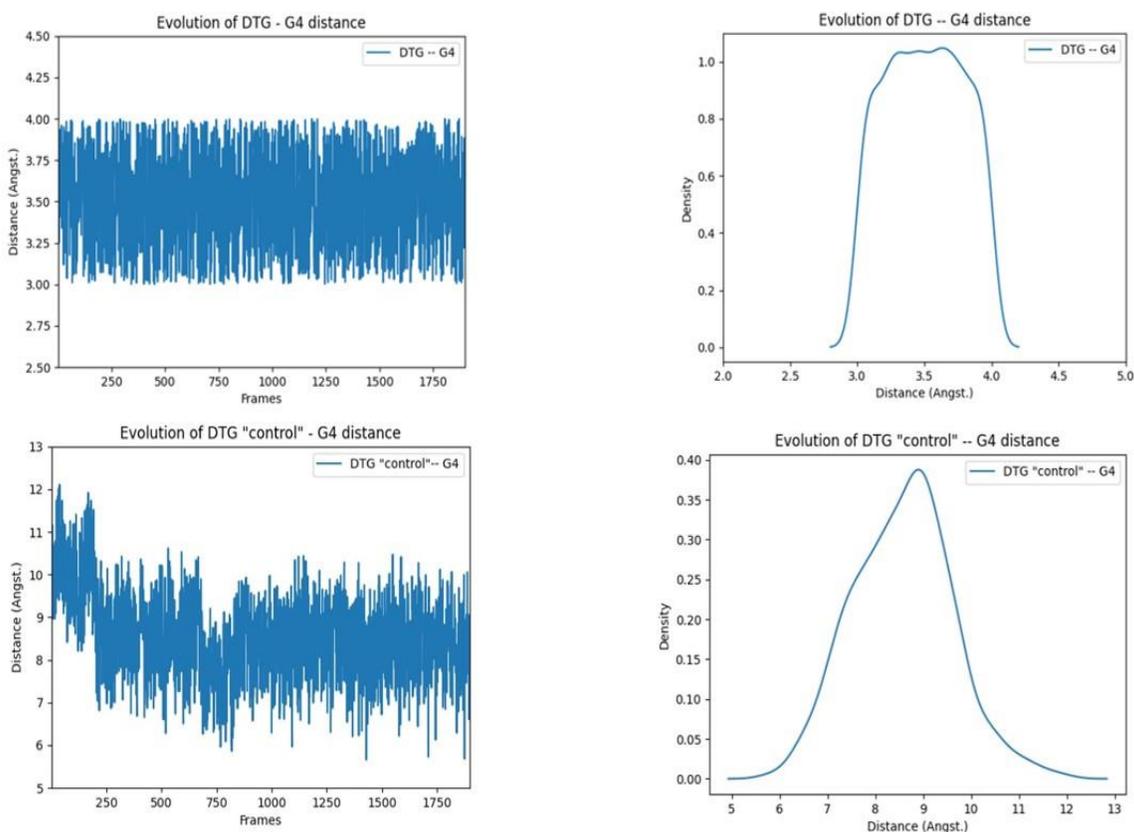


Figure 70 : Graphiques représentant l'évolution de la distance entre DTG et G₄ (à gauche en haut) et DTG «contrôle» et G₄ (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas du DTG avec G₄ et en bas dans le cas du DTG « contrôle » avec G₄).

L'analyse des graphiques de la Figure 70 nous fournit les informations suivantes : la variation de la distance DTG-G₄ est entre 3 Å et 4 Å (graphique à gauche en haut), avec la majorité des distances comprise entre 3,2 Å et 3,8 Å (graphique à droite en haut). Pour le DTG sans noyau halogéné, les distances commencent entre 9 Å et 12 Å avant de diminuer et atteindre des valeurs comprises entre 7 Å et 10 Å (graphique à gauche en bas) avec une majorité aux alentours de 9 Å (graphique à droite en bas).

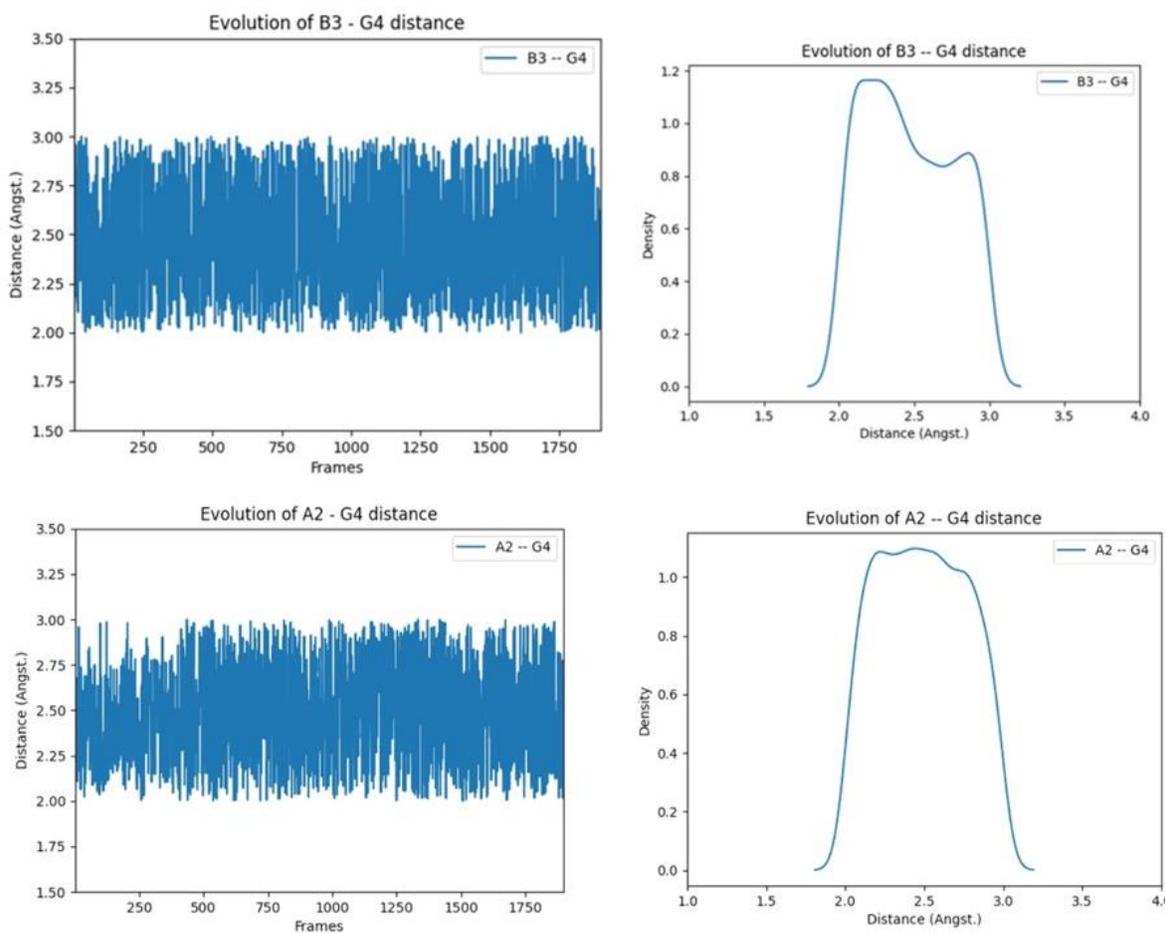


Figure 71 : Graphiques représentant l'évolution de la distance entre B3 et G₄ (à gauche en haut) et A2 et G₄ (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B3 avec G₄ et en bas dans le cas d'A2 avec G₄).

Concernant les distances entre N de NHCH₃ pour B3, A2 et B2 et C6 de la guanine (Figures 71 et 72), elles varient entre 2 Å et 3 Å avec la majorité des distances égale à 2,2 Å pour B3 (Figure 71, graphique à droite en haut), comprise entre 2,5 Å et 2,7 Å pour B2 (Figure 72, graphique à droite en haut) et située entre 2,2 Å et 2,6 Å pour A2 (Figure 71, graphique à droite en bas).

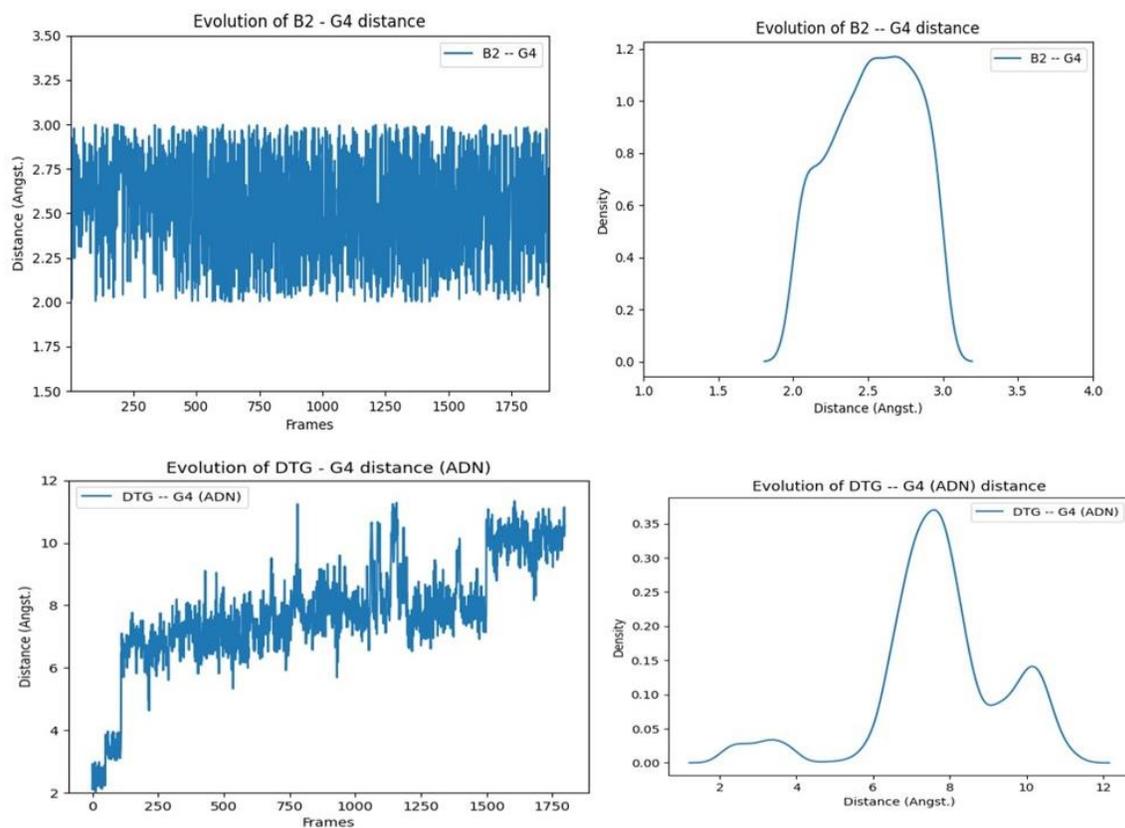


Figure 72 : Graphiques représentant l'évolution de la distance entre B2 et G₄ (à gauche en haut) et DTG et G₄ dans le complexe ADNv-DTG (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B2 avec G₄ et en bas dans le cas du DTG avec G₄ dans le complexe ADNv-DTG).

En l'absence d'ADNv (Figure 72, graphiques en bas), nous remarquons que les contacts DTG-G₄ s'établissent à une distance proche aux alentours de 3 Å puis s'éloignent après 50 ns à une distance entre 6 Å et 8 Å avant d'atteindre une valeur de 10 Å. Le graphique de la Figure 74 à droite en bas illustre ces fluctuations de distance.

Discussion des résultats obtenus concernant les interactions ADN-inhibiteurs

- En premier lieu, l'aspect droit des courbes en étudiant l'empilement entre C₁₆ et les inhibiteurs ainsi que les interactions électrostatiques entre G₄ et les inhibiteurs indique que la majorité des trajectoires de nos composés

présente principalement le même indice d'empilement et la même distance, confirmant que nos structures restent stables tout au long des simulations affichant des valeurs précises et fiables.

- B3, A2 et B2 présentent de meilleurs indices d'empilement π avec C_{16} ainsi que des indices d'empilement avec A_{17} plus élevés par rapport au DTG au sein du complexe ternaire. Ils interagissent donc d'une manière plus efficace et spécifique que le DTG avec ces deux bases terminales virales. Ceci expliquerait les meilleures énergies d'interactions de nos dérivés (ΔE) obtenues précédemment par calcul de chimie quantique et de SIBFA ; en effet des interactions plus spécifiques de B3, A2 et B2 avec C_{16} particulièrement ont contribué à un rendement énergétique plus important de ces dérivés par rapport au DTG.
- Les contours de potentiel électrostatique générés pour DTG, B3, A2 et B2 appuient notre point précédent. Il existe en effet un accord entre les indices d'empilement des noyaux du DTG, B3, A2 et B2 avec la cytosine dans les complexes ternaires et leurs MEP. La Figure 73 rappelant ces contours, montre une région de potentiel électrostatique plus élevée au centre du noyau de B3 (région rouge) par rapport au centre des noyaux d'A2 et B2 (région jaune orange) et DTG (région jaune vert). Ceci explique également le meilleur empilement π - π effectué entre B3 et C_{16} par rapport aux autres dérivés. C'est en accord avec la valeur de ΔE trouvée pour B3 (-31,2 kcal/mol) par rapport aux valeurs de DTG, A2 et B2 de -27,1 kcal/mol et -31,0 kcal/mol respectivement.

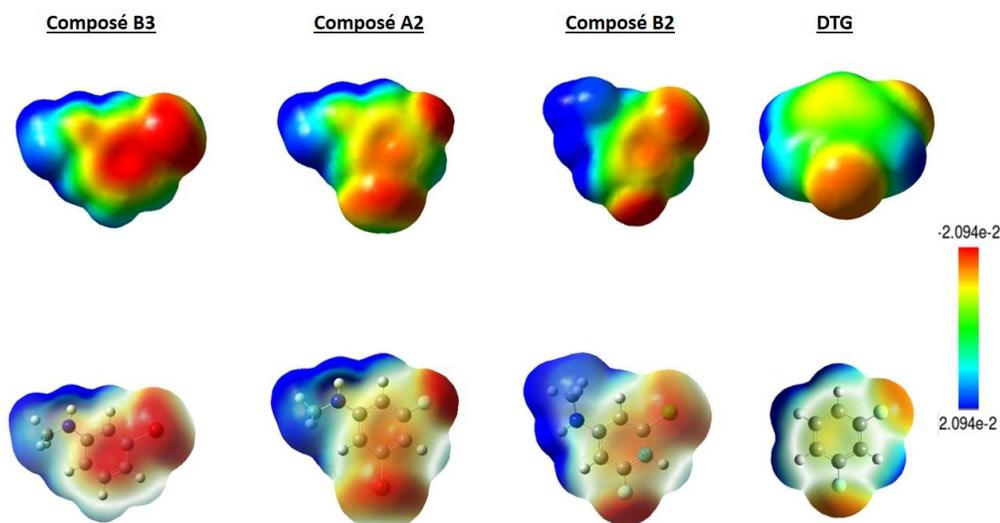


Figure 73 : Contours des potentiels électrostatiques des composés B3, A2, B2 et du Dolutégravir (DTG). La barre colorée à droite indique la variation de la densité électronique entre $-2,094 \times 10^{-2}$ (correspondant à la couleur rouge) et $2,094 \times 10^{-2}$ (correspondant à la couleur bleue).

- Nous pouvons déduire que les interactions des dérivés inhibiteurs avec C₁₆ aboutissent à une diminution des contacts proches entre cette cytosine et la Lysine 156 ainsi que la Lysine 159, comme démontré dans la partie précédente (B.I.1). De plus, leurs contacts proches avec G4 via leur groupement en para contribuent à la diminution des interactions de la Guanine avec Gly149.
- En plus de l'éloignement de l'Adénine A₁₇ du site actif de l'intégrase par DTG, B3, A2 et B2, ces derniers interagissent avec elle par empilement. Ceci peut expliquer de même pourquoi les interactions d'A17 avec Glu152, avec Lys159 ainsi qu'avec Gln146 diminuent une fois les inhibiteurs présents à l'interface ADN-Intégrase.
- Les structures des noyaux d'A2 et B2 sont très proches avec un fluor en ortho, un halogène en ortho et NHCH₃ en para. Cette analogie au niveau de la structure explique la similitude de l'indice d'empilement de leur noyau halogéné avec C₁₆ et par conséquent la superposition de leurs courbes (de

couleur verte pour A2 et jaune pour B2) dans la Figure 56 et les mêmes valeurs de ΔE de -31,0 kcal/mol calculées dans la partie précédente.

- Des contacts plus proches s'effectuent entre l'azote de NHCH_3 que nous trouvons chez B3, A2 et B2 et le carbone 6 de la guanine ; les distances varient entre 2 Å et 3 Å par rapport aux distances entre 3 Å et 4 Å du fluor du DTG avec C6 de la guanine. Ceci souligne l'importance du groupement NHCH_3 que nous avons ajouté, qui rapproche le noyau halogéné de la base virale de l'ADN, assurant des contacts plus spécifiques entre les deux.
- B3 est le dérivé présentant la meilleure interaction avec C_{16} ainsi qu'avec A_{17} puisqu'il a les indices d'empilement les plus élevés, 0,4 et 0,35 respectivement. De plus, comparé à A2 et B2, B3 est le plus proche de G_4 via son groupement NHCH_3 avec des distances approximatives de 2,2 Å. Ainsi son meilleur rendement énergétique ($\Delta E = -31,2$ kcal/mol) par rapport aux DTG, A2 et B2 est expliqué par des interactions plus spécifiques avec les trois bases virales étudiées. B3 s'avère être donc le dérivé le plus efficace parmi les nouveaux inhibiteurs. Ceci met en relief la présence du brome comme substitution en ortho que nous explorerons par la suite.
- Notons que les aspects des courbes obtenues à la Figure 68 ne sont pas identiques à celles obtenues à la Figure 69 ; les valeurs de l'indice d'empilement fluctuent plus dans le cas d' A_{17} que dans le cas de C_{16} . Ceci est dû au fait que la structure d' A_{17} présente des degrés variables de désordre observés par Hare et al dans le cristal de l'intasome du PFV avec le DTG et Mg^{2+} [14], suite à son éloignement de sa position initiale. Cette flexibilité de la structure d' A_{17} explique alors les variations des valeurs d'indice d'empilement entre l'Adénine et les noyaux du squelette de nos inhibiteurs.

- Le DTG au sein du complexe ternaire s'empile d'une manière plus efficace avec le noyau de C₁₆ ainsi qu'avec les noyaux d'A₁₇ qu'en présence de l'ADN viral seul. En effet, en présence de l'ADN uniquement, l'indice d'empilement trouvé pour le DTG est de 0,1, valeur plus faible que trouvée en présence de l'ADNv et l'IN (valeur de 0,2). De plus, les contacts spécifiques entre le DTG et G₄ sont affaiblis peu à peu en l'absence de l'intégrase, éloignant le noyau de DTG de la Guanine. Ceci est en accord avec des études précédentes effectuées sur les inhibiteurs de transfert de brin, révélant que ces derniers se fixent au niveau de l'intégrase du VIH-1 quand l'enzyme est engagée avec l'ADN viral, après l'étape du 3' processing et avant l'étape de transfert de brin ^[15]. Du coup, un mode d'action efficace du DTG nécessite la présence simultanée de ses deux partenaires. Ainsi, en absence de l'IN, l'interaction respective de DTG avec C₁₆, A₁₇ et G₄ s'est affaiblie.
- En absence de noyau halogéné, le DTG «contrôle » dans le complexe ternaire présente un indice d'empilement avec A₁₇ plus faible (0,15) que celui de DTG (0,21) dans le complexe ternaire. Nous remarquons aussi chez la molécule « contrôle » que son hydrogène remplaçant le noyau halogéné est éloigné de G₄ à des distances allant jusqu'à 12 Å. Cet éloignement marque l'implication du noyau halogéné dans les interactions spécifiques des drogues inhibitrices avec C₁₆, G₄ et A₁₇, l'inhibiteur se positionnant au niveau de la poche libérée par l'Adénine.

B.III.3. Etude des interactions Intégrase – Inhibiteurs

Les inhibiteurs de transfert de brin agissant à l'interface ADN-IN, et après l'analyse des interactions entre l'ADN viral et les inhibiteurs, nous nous intéressons à l'étude

des interactions entre nos inhibiteurs et l'intégrase du PFV, particulièrement au niveau de résidus clés (conservés chez le VIH-1).

B.III.3.1. Evaluation de la distance entre des résidus clés de l'intégrase virale et les inhibiteurs

Dans le cristal de l'intasome du PFV et en présence du DTG et des cations Mg^{2+} , Hare et ses collaborateurs ont identifié des interactions entre des résidus clés de l'intégrase du PFV et le DTG [16]. En premier lieu, ils ont identifié des contacts de van der Waals entre le fluor en position ortho du DTG et les carbones $C\gamma$ et $C\delta$ du Glutamate en position 221 de la triade catalytique du PFV, correspondant au Glutamate en position 152 de l'intégrase du VIH-1 (Figure 47 et Figure 132 en annexe). Deuxièmement, bien que le DTG résiste à certaines mutations au niveau de l'IN, il affiche une faible résistance face à la substitution du résidu Glycine 118 en Arginine (Gly118R) [14]. En effet, Hare et al ont observé que le DTG s'étend vers Gly118, établissant avec elle une liaison hydrogène (Figure 74). Cette Gly118 est conservée au niveau de l'intégrase du VIH-1 et celle du PFV (Gly187 dans la Figure 47). Elle est située dans la boucle $\beta 4-\alpha 2$ de l'intégrase ; cette boucle contient le résidu catalytique Asp116 (correspondant à Asp185 du PFV).

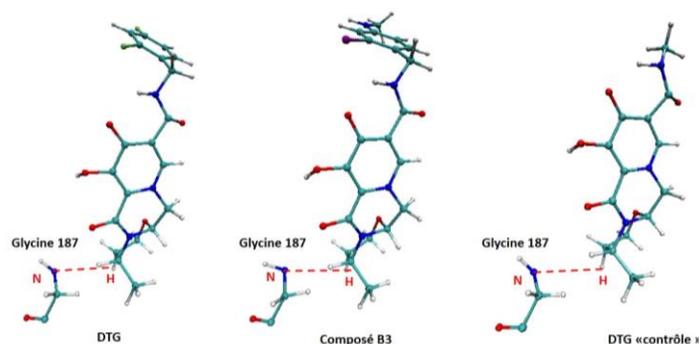


Figure 74 : Représentation des positions de départ dans nos simulations des dérivés DTG, B3 et DTG « contrôle » par rapport à la Glycine 118, avec en gras et rouge les atomes dont les distances ont été suivies au cours des simulations. La ligne rouge pointillée indique la distance entre eux. Les composés A2 et B2 ayant des positions de départ similaires au composé B3 ne sont pas représentés dans la figure.

Une autre interaction clé identifiée implique le carbone du « linker » du DTG et le C γ de la Proline 145 du VIH-1, résidu situé au niveau de la boucle en amont de l'hélice α 4 et conservé chez l'intégrase du PFV (Proline 214) (Figure 47 et Figure 133 en annexe). On a constaté que le DTG ajuste sa position en fonction de celle de Pro145 ; en effet, on a observé dans la structure du DTG en présence de l'intégrase mutée S217H (Code PDB : 3S3N) un déplacement de la Proline 145 de 0,49 Å par rapport à sa position dans la structure de l'intégrase sauvage en présence du DTG (code 3S3M) suite à la mutation S217H. Par conséquent, les atomes du « linker » du DTG se sont déplacés 0,4 Å afin de préserver leurs contacts de vdW avec la proline ^[16]. Ainsi, nous avons suivi l'évolution de la distance entre les carbones C γ et C δ de Glu152 et l'atome en position ortho de nos inhibiteurs, séparément. Ce dernier correspond au fluor dans le cas du DTG, A2 et B2, au brome dans le cas de B3 et à l'hydrogène dans le cas du DTG « contrôle » (Figure 133 en annexe). Ensuite, nous avons étudié la variation de la liaison hydrogène établie entre la Gly118 et le DTG vérifiant si la susceptibilité de ce dernier à la Glycine 118 est due à ses contacts proches avec elle. La distance entre la Gly118 et B3, A2 et B2 a été de même évaluée séparément afin de mieux décrire le comportement de chaque dérivé vis-à-vis de ce résidu. Finalement, nous avons relevé les distances entre le C γ de la Pro145 et le carbone du « linker » de nos inhibiteurs (Figure 134 en annexe) afin de voir non seulement si les liaisons de vdW entre le DTG et ce résidu sont maintenues au cours de nos simulations, mais aussi si nos nouveaux inhibiteurs présentent des contacts proches avec Pro145. Le DTG « contrôle » sera de même pris en considération toujours dans le but de mettre en évidence le rôle joué par le noyau halogéné dans les contacts inhibiteur-IN. Enfin, la formation du complexe ADNv-IN étant nécessaire pour une activité optimale des INSTIs, l'évolution des trois

distances mentionnées ci-dessus a été suivie également au niveau du complexe Intégrase-DTG. L'étude de ce complexe binaire montrerait la contribution de l'ADN viral à ces interactions. L'ensemble des résultats est présenté aux figures ci-dessous. Concernant les interactions Glu152-inhibiteurs, les résultats étant similaires pour le C γ et le C δ , nos graphiques afficheront les distances inhibiteurs-C γ uniquement ; celles avec le C δ ainsi que la distance C γ -B2 sont présentées en annexe (Figures 135 à 138). Pour les distances inhibiteurs-Gly118 et inhibiteurs-Pro145, nous présentons les résultats pour B3 uniquement étant donné que les trois dérivés B3, A2 et B2 ont généré des profils de graphiques similaires (Figures 139 et 140 en annexe).

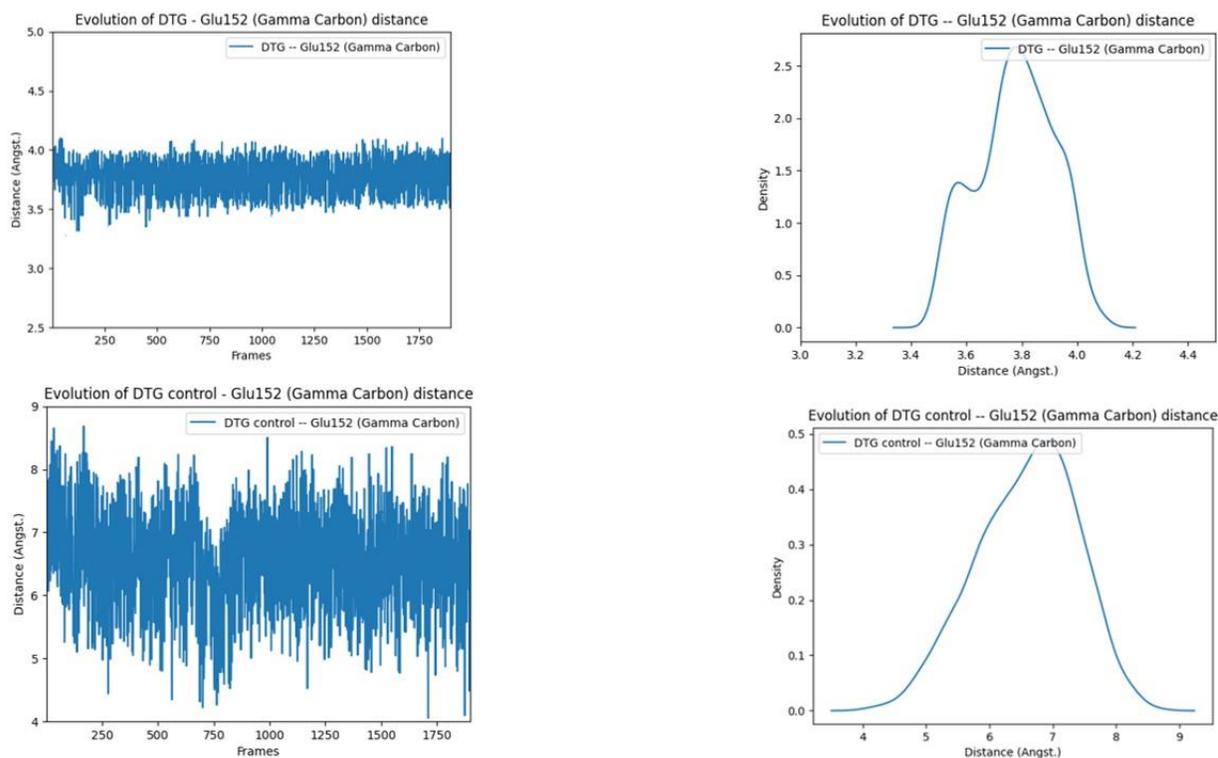


Figure 75 : Graphiques représentant l'évolution de la distance entre DTG et Glu152 (correspondant au Glu221 du PFV) (à gauche en haut) et entre DTG « contrôle » et Glu152 (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas du DTG avec Glu152 et en bas dans le cas du DTG «contrôle » avec Glu152.

Nous observons une variation de la distance entre le fluor en ortho du DTG et le Cy de Glu152 comprise entre 3,5 et 4 Å tout au long des simulations (Figure 75, graphique en haut à gauche) avec des contacts majoritairement aux alentours de 3,8-3,9 Å (Figure 75, graphique en haut à droite). Pour le DTG «contrôle», la distance qui sépare son hydrogène du Glu152 varie entre 5 Å et 8 Å avec des valeurs majoritairement égales à 6,8 Å (Figure 75, graphiques en bas respectivement).

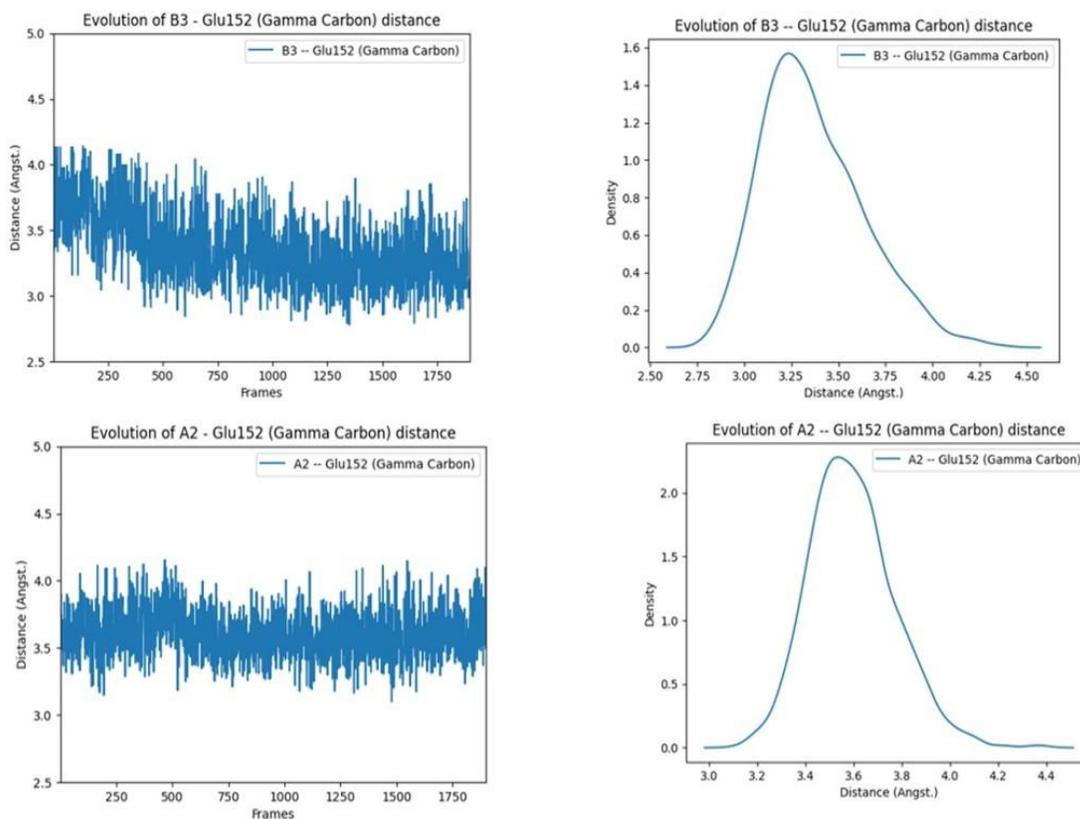


Figure 76 : Graphiques représentant l'évolution de la distance entre B3 et Glu152 (correspondant à Glu221 pour le PFV) (à gauche en haut) et entre A2 et Glu152 (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B3 avec Glu152 et en bas dans le cas d'A2 avec Glu152).

Quant à nos nouveaux dérivés, dans le cas de B3 et A2, la distance respective entre le Cy du résidu catalytique et le brome en ortho de B3 ainsi que le fluor en ortho d'A2 est aux alentours de 3,5-4 Å (Figure 76, graphique à gauche). Elle diminue à une valeur de 3 Å après 80 ns de simulation dans le cas de B3 où elle reste

majoritairement aux alentours de 3,2-3,3 Å (Figure 76, graphiques en haut à droite). Tandis que la majorité des trajectoires d'A2 affichent des distances de 3,5-3,6 Å (Figure 76, graphique en bas à droite).

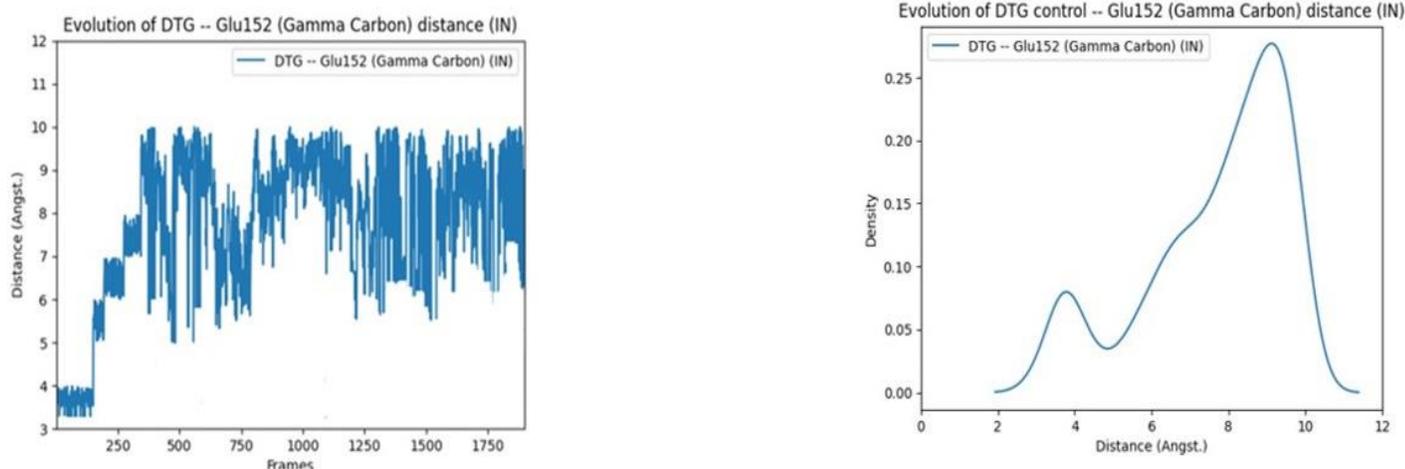


Figure 77 : Graphique à gauche représentant l'évolution de la distance entre DTG et Glu152 (correspondant au Glu221 du PFV) au sein du complexe Intégrase-DTG en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons le graphique de la densité des pauses en fonction des distances obtenues

Les deux graphiques de la Figure 77 correspondent à l'évolution de la distance fluor – C γ dans le cas du DTG en présence de l'intégrase uniquement. Ils indiquent des contacts proches entre les deux (entre 3,5 et 4 Å) au début de la simulation, qui augmentent par la suite après 50 ns et fluctuent entre 6,5 Å et 9 Å, avec la plupart des distances égales à 8,5 Å, comme l'indique la courbe du graphique à droite.

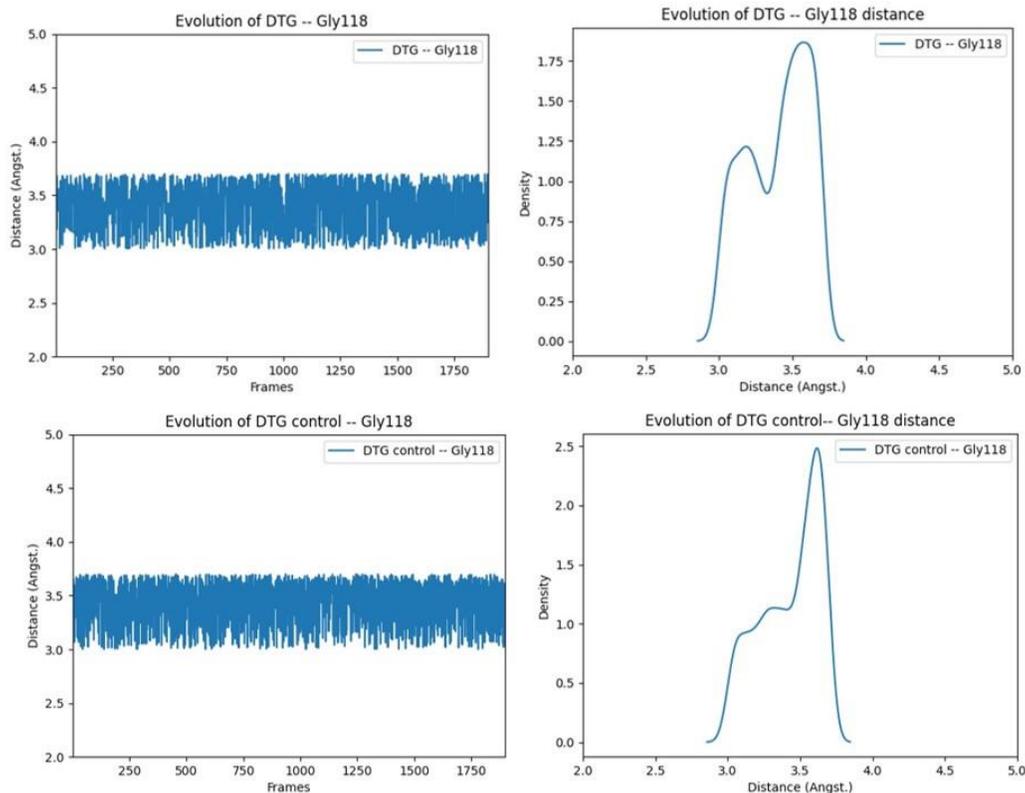


Figure 78 : Graphiques représentant l'évolution de la distance entre DTG et la Glycine 118 (correspondant à la Glycine 187 du PFV) (à gauche en haut) et entre le DTG « contrôle » et la Glycine 118 (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas du DTG avec Gly118 et en bas dans le cas du DTG « contrôle » en présence de Gly118).

Pour les interactions de DTG et DTG « contrôle » avec Gly118 et Pro145, nous remarquons que les liaisons hydrogènes établies entre le DTG et la Gly118 ainsi que celles avec la Pro145 sont conservées au cours des 400 ns de simulation, avec la majorité des distances variant entre 3,2 Å et 3,6 Å dans le cas de la Glycine (Figure 78, graphiques en haut) et entre 3,5 Å et 4 Å dans le cas de la Proline avec une majorité de distances égale à 3,9 Å (Figure 79, graphiques en haut). Ceci est de même observé dans le cas du DTG « contrôle » qui maintient des contacts proches entre 3,4 Å et 3,6 Å avec la Gly118 (Figure 78, graphiques en bas) et entre son « linker » et la Pro145 où la majorité des distances est égale à 3,8 Å (Figure 79, graphiques en bas).

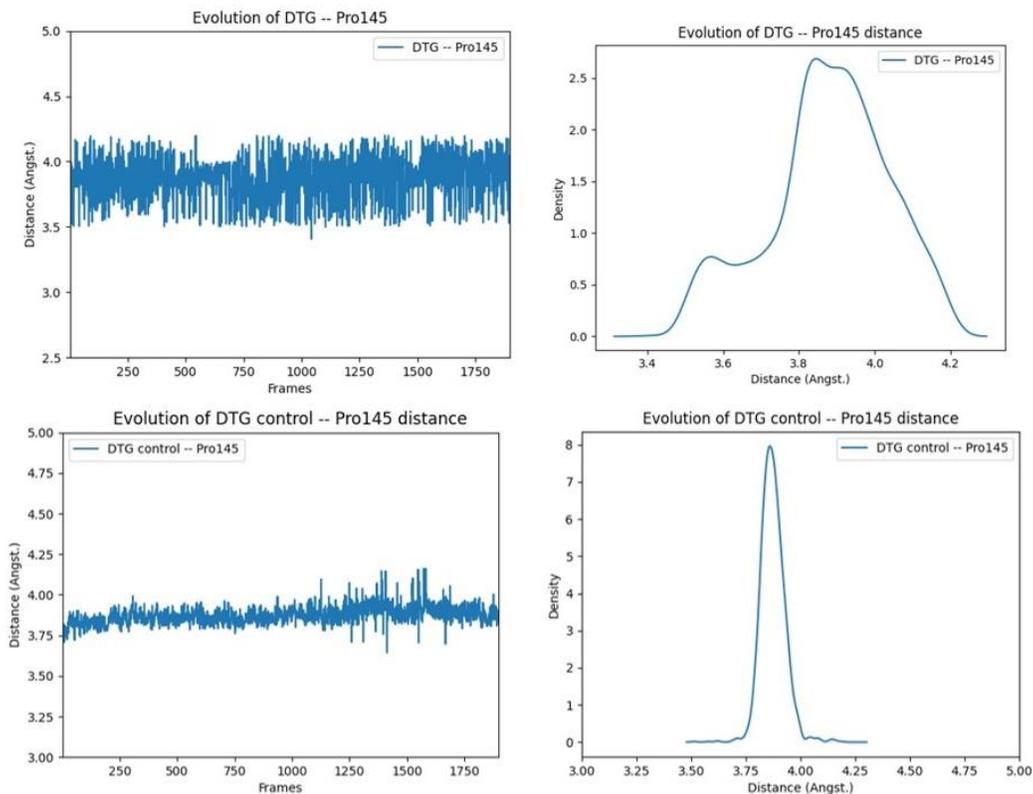


Figure 79 : Graphiques représentant l'évolution de la distance entre DTG et la Proline 145 (à gauche en haut) et DTG « contrôle » et la Proline 145 (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas du DTG avec Pro145 et en bas dans le cas de DTG « contrôle » en présence de Pro145).

Concernant le complexe ADN-Intégrase-B3, les liaisons hydrogènes avec Gly118 et les contacts de vdW entre son « linker » et Pro145 sont maintenues au cours des simulations, à des distances majoritairement égales à 3,5 Å dans le cas de la Glycine (Figure 80, graphiques en haut). Elles varient entre 3,7 Å et 3,9 Å dans le cas de la Proline (Figure 81, graphiques en haut).

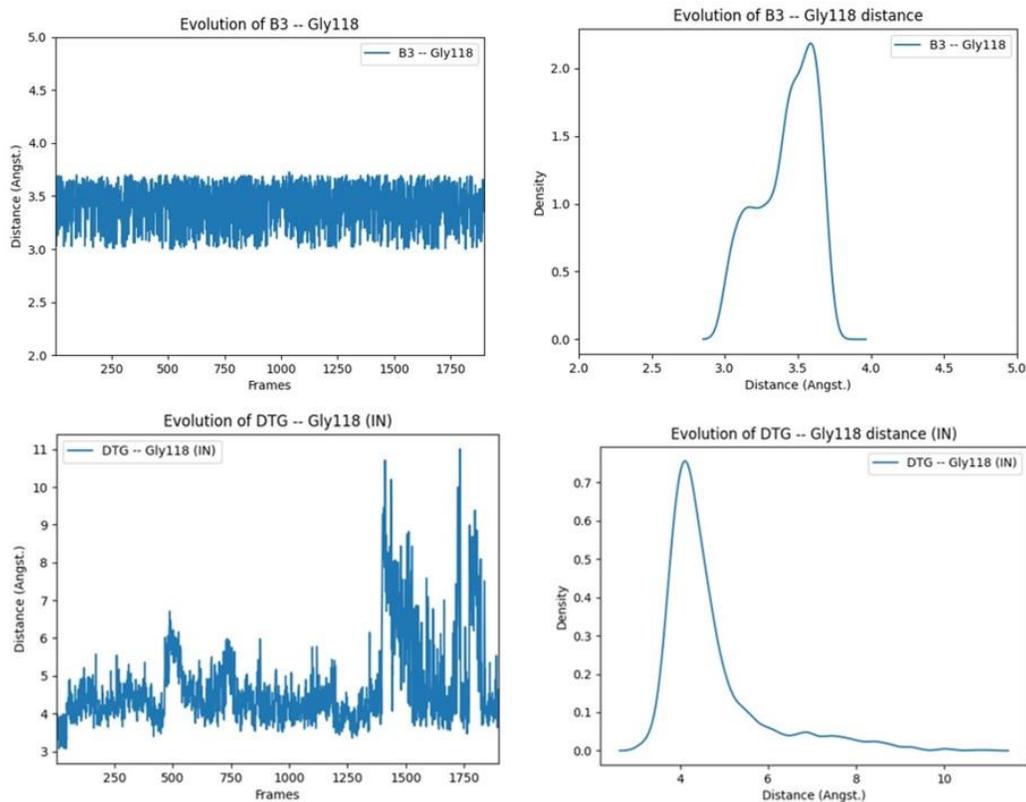


Figure 80 : Graphiques représentant l'évolution de la distance entre B3 et la Glycine 118 (correspondant à la Glycine 187 du PFV) (à gauche en haut) et DTG et la Glycine 118 dans le complexe Intégrase-DTG (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B3 avec Gly118 et en bas dans le cas du DTG en présence de Gly118 dans le complexe DTG-Intégrase).

Cependant en l'absence de l'ADNv, nous remarquons une perte des contacts proches effectués entre le DTG et les deux résidus ; la liaison hydrogène est perdue entre l'inhibiteur et Gly118 où la distance entre les deux est de 5 Å à peu près jusqu'à 280 ns de simulation avant d'augmenter à des valeurs entre 8 Å et 10 Å vers la fin des simulations (Figure 80, graphiques en bas). La liaison de vdW est maintenue entre le linker du DTG et Pro145 uniquement au cours des premières 50 ns de simulation. Après cette durée, la distance entre les deux est située majoritairement à 15 Å (Figure 80, graphiques en bas).

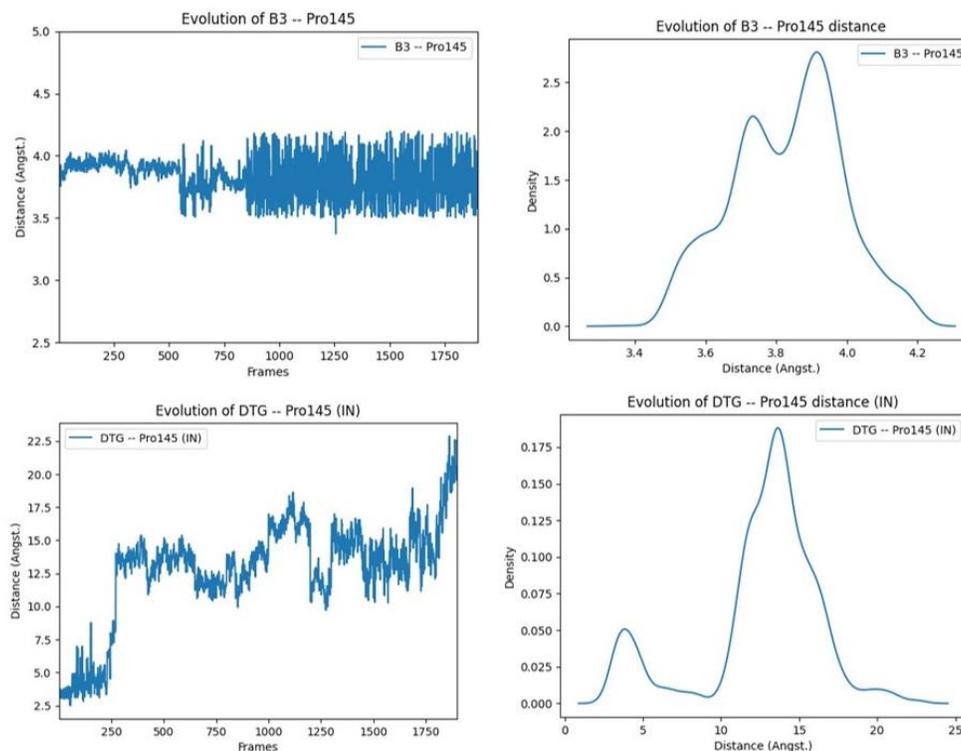


Figure 81: Graphiques représentant l'évolution de la distance entre B3 et la Proline 145 (à gauche en haut) et entre le DTG et la Proline 145 dans le complexe Intégrase-DTG (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B3 avec Pro145 et en bas dans le cas du DTG en présence de Pro145 dans le complexe Intégrase-DTG).

Discussion des résultats obtenus concernant les interactions des inhibiteurs avec les résidus clés de l'intégrase

- Les graphiques obtenus présentent peu de fluctuations au niveau des valeurs des distances attribuées aux trajectoires des systèmes simulés, ce qui indique que nos structures restent stables tout au long des simulations et que les valeurs affichées sont correctes et fiables.
- DTG interagit d'une manière stable avec la Gly118 et la Pro145, effectuant avec elles des contacts proches conservés le long des simulations. Ceci explique la position de départ du DTG, observée par Hare et ses collaborateurs dans le cristal 3S3M où la drogue tend vers Gly118 et se trouve à proximité d'elle. Sa susceptibilité à la mutation G118R est donc due

aux contacts spécifiques avec ce résidu. De plus, les contacts proches et conservés entre la Pro145 et DTG expliquent pourquoi ce dernier ajuste sa position suite au mouvement de la Proline dans le cas du complexe muté S217/H^[16].

- Nos nouveaux dérivés B3, A2 et B2 ainsi que le DTG présentent des liaisons stables entre leur atome en ortho et le résidu Glu152 de la triade catalytique. Ces liaisons sont aux alentours de 3,5-4 Å, distances caractéristiques des liaisons de vdW. Ce qui explique la diminution des interactions entre Glu152 et A₁₇ démontrée dans la partie B.I. Ajoutons que B3, A2 et B2 maintiennent des contacts proches avec la Gly118 et la Pro145 au cours des simulations. Ainsi, les interactions des trois dérivés avec ces résidus clés sont conservées tout au long de nos dynamiques.
- B3 effectue des contacts légèrement plus proches via son brome avec le C γ du Glu152 par rapport au fluor de DTG, A2 et B2. En effet, nous observons les distances suivantes ; Brome – C γ = 3,3 Å / Fluor (A2) – C γ = 3,5 Å / Fluor (B2) – C γ = 3,7 Å (résultats en annexe) / Fluor (DTG) – C γ = 3,9 Å. B3 présente donc l'interaction la plus forte avec ce résidu catalytique de l'intégrase, confirmant notre choix du brome comme substituant. B3 possède donc le meilleur profil inhibiteur, via ses contacts plus proches avec les bases de l'ADN (démontrés précédemment) ainsi qu'avec Glu152.
- Le DTG « contrôle », dépourvu de son noyau d'halobenzène présente des distances très élevées (aux alentours de 7 Å) entre son hydrogène et C γ de Glu152. Ces dernières ne caractérisent plus des liaisons de vdW. Il y a donc eu une perte des contacts entre la drogue et Glu152 ce qui souligne l'importance du noyau d'halobenzène non seulement pour interagir avec

l'ADN viral mais aussi avec la triade catalytique de l'intégrase. Cependant, les contacts sont conservés entre la molécule « contrôle » et Gly118 et Pro145 respectivement. Nous pouvons conclure par la suite que la structure halogénée ne joue pas de rôle direct et majeur dans l'interaction de l'inhibiteur avec ces résidus de l'intégrase

- Le complexe binaire Intégrase-DTG a manifesté des pertes des liaisons de vdW entre le fluor du DTG et le résidu catalytique Glu152, après quelques ns de simulations. Il y a eu aussi une perte des contacts proches entre le DTG et les Gly118 et Pro145. Ceci souligne l'instabilité et la non spécificité de l'interaction entre le DTG et l'intégrase en absence d'ADNv. En fait, la fixation des inhibiteurs de transfert de brin à l'intégrase virale nécessite une enzyme fonctionnelle liée à l'ADNv ^{[16][17]}.

B.III.3.2. Evaluation de la distance entre les résidus causant des mutations de résistance primaires et secondaires chez l'intégrase du VIH-1 et nos inhibiteurs

L'intégrase, suite au contact des inhibiteurs de transfert de brin, développe des mutations de résistance primaires et secondaires. Or, le but ultime de notre stratégie de conception de nouvelles molécules inhibitrices étant de renforcer leurs interactions respectives avec l'ADN viral afin de compenser une perte éventuelle des contacts de l'IN en cas de résistance. Par les calculs de ΔE et nos analyses précédentes des interactions ADN-inhibiteurs, nous avons validé les interactions spécifiques de nos drogues avec l'ADN viral. DTG est déjà connu pour surmonter la majorité des mutations de résistance ciblant RAL et EVG ^{[18][19]} et notre équipe a démontré dans des études précédentes *in vitro* que cette caractéristique du DTG est due à une interaction plus spécifique avec l'ADN viral clivé par rapport à RAL et EVG ^{[20][21]}. Dans cette partie, et par un suivi de l'évolution de la distance entre DTG

et les résidus de l'intégrase qui subissent une mutation de résistance primaire ou secondaire, nous cherchons à démontrer que des contacts de DTG moins importants avec IN seraient à la base de sa moindre sensibilité aux mutations de résistance. Nous évaluerons de même cette caractéristique chez B3, A2 et B2, afin de vérifier l'effet des substitutions réalisées sur les interactions des dérivés avec l'enzyme. Les résidus de IN impliqués dans des mutations de résistance primaires au contact de RAL et EVG sont les suivants ^[22] :

- la Tyrosine 143 mutée en Histidine ou Arginine ou Cystéine (Y143 H/R/C). Elle est située dans la boucle catalytique en amont de l'hélice α 4 et est conservée chez les intégrase du VIH-1 et du PFV ; son équivalent correspond à la Tyrosine 212 du PFV (Figure 47). Dans le cristal de code 3S3M, Hare et collaborateurs ont identifié des contacts de vdW entre l'hydrogène du groupement hydroxyle (OH) du noyau aromatique de Tyr143 et le carbone du squelette des drogues (Figure 141 en annexe).
- l'Asparagine 155 mutée en Histidine (N155H), est située au niveau de l'hélice α 4 et est conservée chez les intégrases du VIH-1 et du PFV, avec l'Asn 224 comme son équivalent chez ce dernier (Figure 47).
- La Glutamine 148 mutée en Histidine ou Arginine ou Lysine (Q148 H/R/K), non conservée chez le PFV ; la Sérine 217 lui correspond dans l'intégrase de ce dernier (Figure 47). Elle est située au niveau de l'hélice α 4. Elle est souvent accompagnée d'une mutation de résistance secondaire au niveau de la Glycine 140 (G140 S/A). Notons que dans le cas du PFV, la mutation S217H est l'équivalent de la double mutation Q148H/G140S chez le VIH-1 puisque le PFV présente déjà une sérine en position 140.

Nous suivrons l'évolution de la distance entre les résidus de mutations primaires (Tyr143, Asn155 et Gln148) et les inhibiteurs anti-intégrase DTG, B3, A2, B2 et le DTG «contrôle ». Ce dernier nous donnera un aperçu sur l'implication du noyau halogéné dans ces interactions. Les distances suivies sont : l'hydrogène du groupement hydroxyle (OH) du noyau aromatique de Tyr143 avec le carbone du squelette des drogues, le Carbone α d'Asn155 avec le carbone du « linker » de nos inhibiteurs, situé en face de ce résidu et enfin le carbone α de S217 avec le carbone du « linker » de nos inhibiteurs, en face de lui (Figures 141 à 143 en annexes). Dans le cas de la mutation de résistance secondaire, les résultats du suivi de l'évolution de la distance entre la Gly140 et les inhibiteurs seront indiqués en annexe (Figures 144 à 147 en annexe).

Premièrement, l'analyse de la Figure 82 montre des contacts proches de 4 Å s'effectuant entre l'hydrogène de Tyr143 et le carbone du DTG, ainsi qu'entre la Tyrosine et le DTG « contrôle ». Ces contacts sont perdus respectivement après 40 ns de simulation et après 20 ns. Dans le cas du DTG, les distances augmentent et varient entre 5 Å et 6,5 Å. Ensuite, à partir de 150 ns de simulation, le DTG s'éloigne davantage de la Tyrosine et se situe à 8,5 Å (Figure 82, graphiques en haut). Alors que dans le cas du DTG « contrôle », la distance devient proche de 8 Å, avant de se stabiliser à une valeur entre 10 Å et 11 Å après 50 ns de simulation.

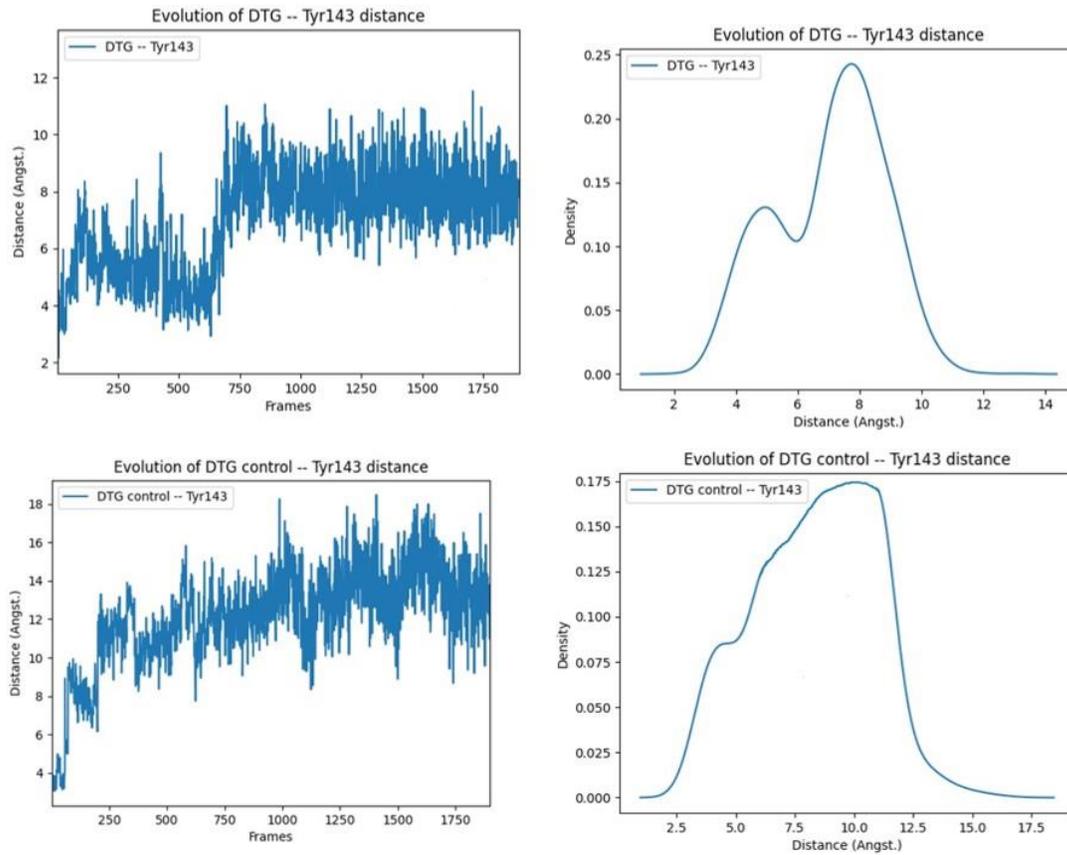


Figure 82 : Graphiques représentant l'évolution de la distance entre DTG et la Tyrosine 143 (à gauche en haut) et le DTG « contrôle » et la Tyrosine 143 (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas du DTG avec la Tyrosine 143 et en bas dans le cas du DTG «contrôle » avec la Tyrosine 143).

Concernant les résidus Asn155 et Ser217, la distance entre eux et le DTG est stable tout au long des simulations, variant respectivement aux alentours de 9-10 Å (majorité des distances entre 9,3-9,5 Å dans le cas d'Asn et entre 9 Å et 9,5 Å dans le cas de Ser) (Figure 83 et 84, graphiques en haut). Quant au DTG «contrôle », sa distance avec ces deux résidus varie respectivement autour de 10-10,5 Å et 9-11,5 Å (Figure 83 et 84, graphiques en bas).

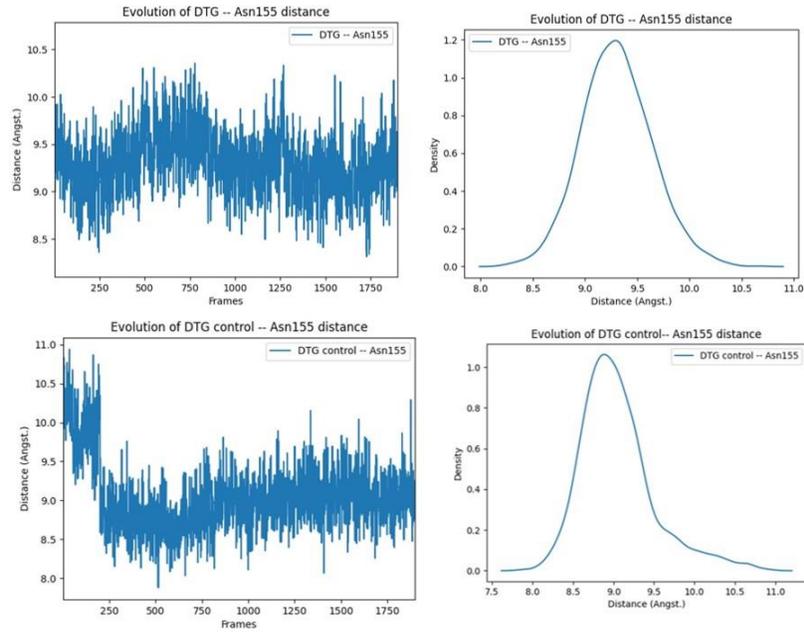


Figure 83 : Graphiques représentant l'évolution de la distance entre le DTG et l'Asparagine 155 (Asn155) (correspondant à Asn224 chez le PFV) (à gauche en haut) et entre le DTG «contrôle » et l'Asparagine 155 (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas du DTG avec l'Asparagine 155 et en bas dans le cas du DTG « contrôle » avec l'Asparagine 155).

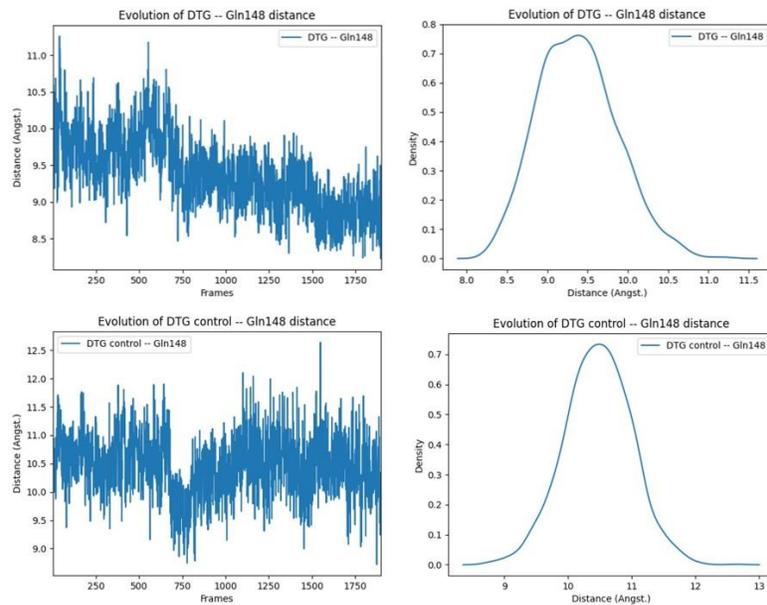


Figure 84 : Graphiques représentant l'évolution de la distance entre le DTG et la Sérine 217 du PFV correspondant à la Glutamine 148 (Gln148) du VIH-1 (à gauche en haut) et entre le DTG «contrôle » et la Sérine 217 du PFV (Gln148 du VIH-1) (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas du DTG avec Gln148 et en bas dans le cas du DTG «contrôle » en présence de Gln148).

Au sein des complexes ADN-Intégrase-nouveaux dérivé (Figures 85 et 86), le comportement des dérivés vis-à-vis la Tyr143 sont presque similaires ; les liaisons s'établissant sont proches, à une distance de 4 Å jusqu'à 20 ns de simulation puis, dans le cas de B3 et B2, la distance augmente et devient de 6 Å avant de varier entre 7 Å et 10 Å jusqu'à 200 ns de simulation dans le cas de B3 (Figure 85, graphique en haut à gauche) et de 10 Å à 320 ns de simulation dans le cas de B2 (Figure 86, graphique en bas à gauche). B3 s'éloigne alors davantage et se trouve à 12,5 Å majoritairement (Figure 85, graphique en haut à droite) alors que la plupart des distances entre B2 et Tyr143 sont de 9 Å (Figure 86, graphique en bas à droite).

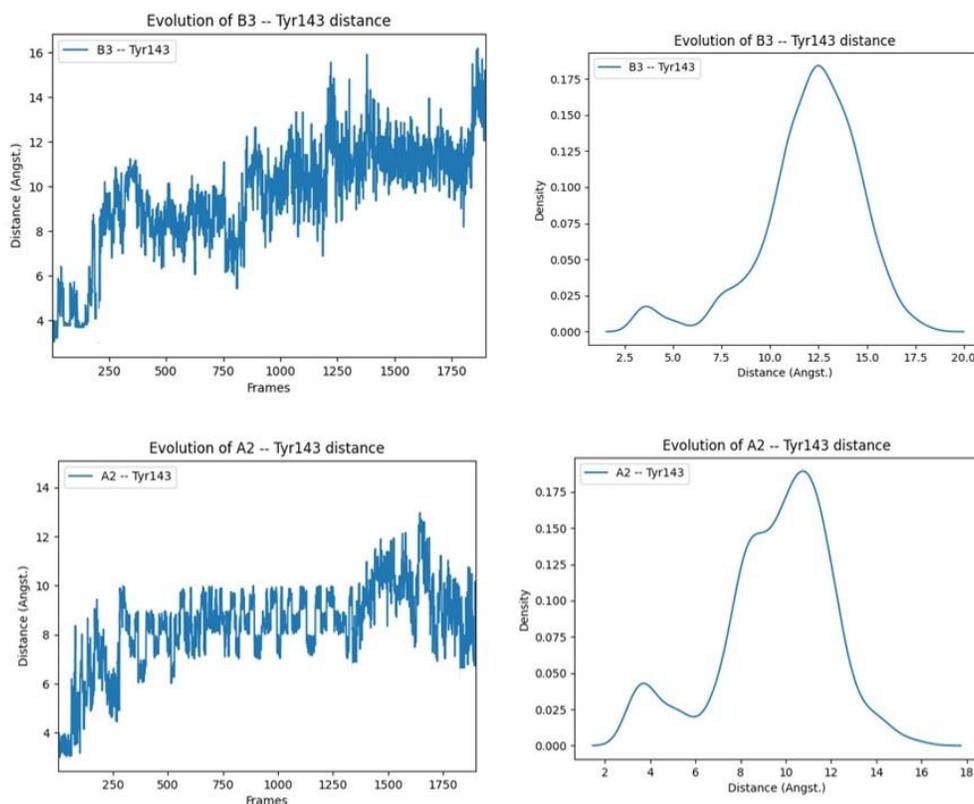


Figure 85 : Graphiques représentant l'évolution de la distance entre B3 et la Tyrosine 143 (correspondant à la Tyrosine 212 du PFV) (à gauche en haut) et entre A2 et la Tyrosine 143 (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B3 avec la Tyrosine 143 et en bas dans le cas d'A2 avec la Tyrosine 143).

Quant à A2, la distance augmente et varie entre 6 Å et 8 Å jusqu'à 60 ns, puis se stabilise aux alentours de 9 Å avant d'augmenter à 11,5 – 12 Å après 260 ns de simulation. La majorité des distances Tyr 143 – A2 sont égales à 11 Å (Figure 85, graphiques en bas).

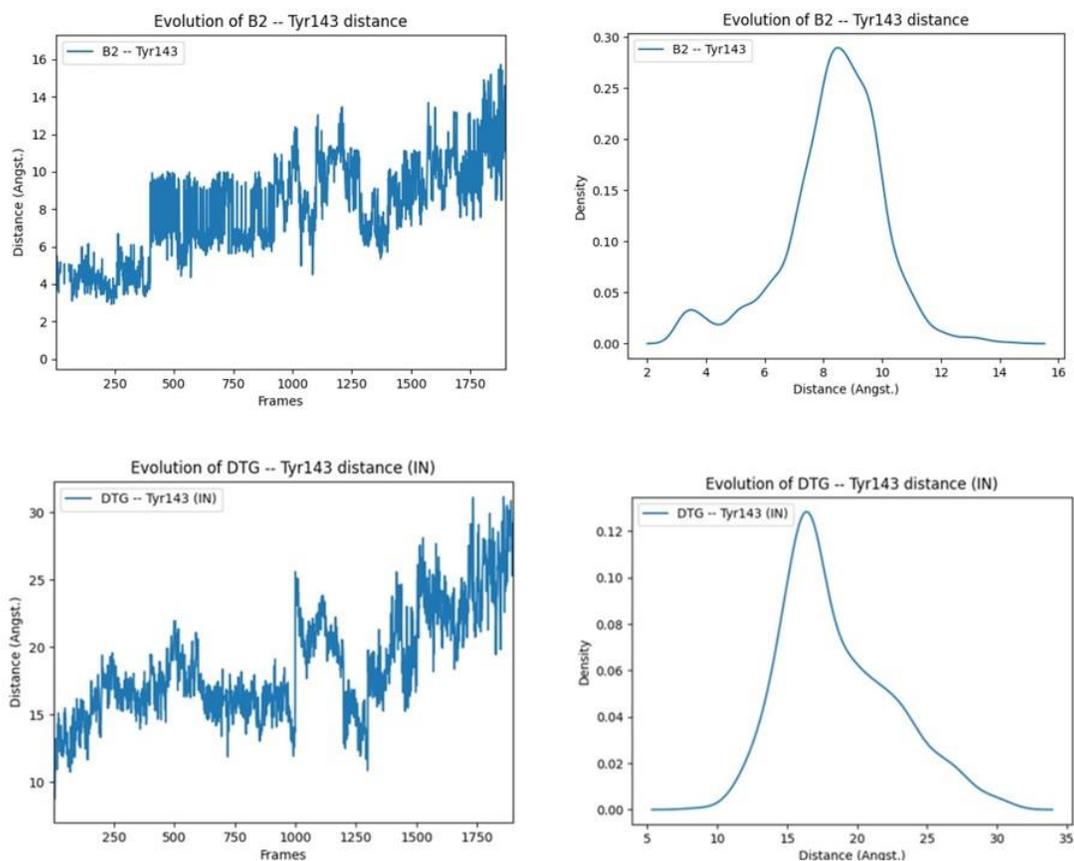


Figure 86 : Graphiques représentant l'évolution de la distance entre B2 et la Tyrosine 143 (correspondant à la Tyrosine 212 du PFV) (à gauche en haut) et DTG et la Tyrosine 143 au sein du complexe Intégrase – DTG (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B2 avec la Tyrosine 143 et en bas dans le cas du DTG avec la Tyrosine 143 dans le complexe Intégrase -DTG).

Finalement, en considérant les complexes des inhibiteurs-IN en absence de l'ADNv, les graphiques de la Figure 86 montrent une instabilité de la liaison DTG-Tyr143 avec une perte rapide des contacts de vdW et des distances de liaisons éloignées de 16 Å dans leur majorité.

Pour les interactions avec les résidus Asn et Ser, la distance varie dans le cas de B3 entre 10,5 Å et 13,5 Å et entre 11 Å et 12 Å respectivement avec la majorité des valeurs égales à 12 Å (Figure 87 et 89, graphiques en haut). La distance A2-Asn155 est quant à elle entre 9,5 Å et 10,5 Å pour l'Asparagine et entre 10,5 Å et 11,5 Å pour la Sérine (Figure 87 et 89, graphiques en bas), avec des distances majoritaires de 11 Å.

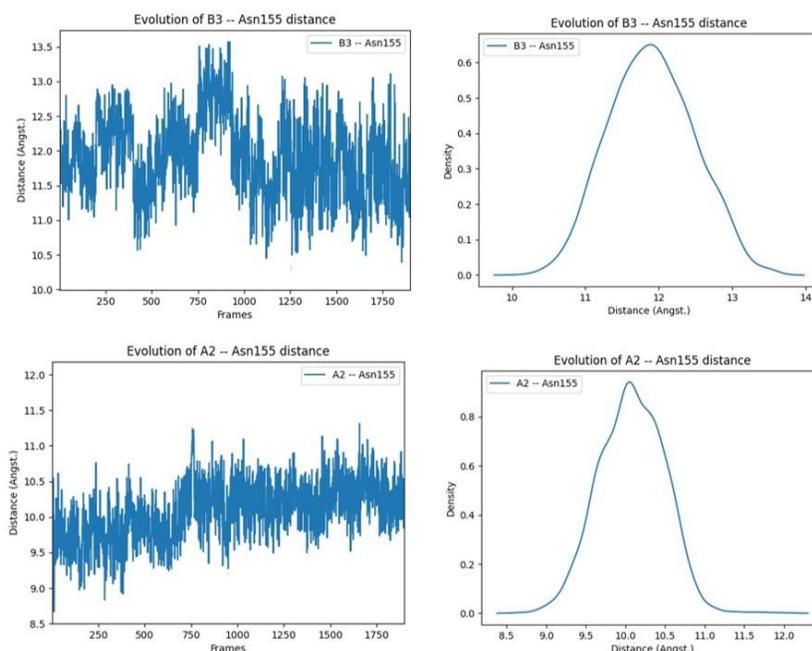


Figure 87 : Graphiques représentant l'évolution de la distance entre B3 et la l'Asparagine 155 (correspondant à Asn224 chez le PFV) (à gauche en haut) et entre A2 et l'Asparagine 155 (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B3 avec l'Asparagine 155 et en bas dans le cas d'A2 en présence de l'Asparagine 155).

Concernant la distance B2-Asn155, nous trouvons qu'elle varie entre 9,5-10,5 Å avant de se stabiliser à 8 Å après 50 ns de simulation avec une distance majoritairement aux alentours de 9,5 Å (Figure 88, graphiques en haut). De plus, B2 est à 10 Å à peu près de la Ser217 jusqu'à 100 ns de simulation. Il passe à 12 Å à 200 ns de simulation avant de revenir à une distance de 11 Å. Il est situé majoritairement à 10,5 Å (Figure 90, graphiques en haut).

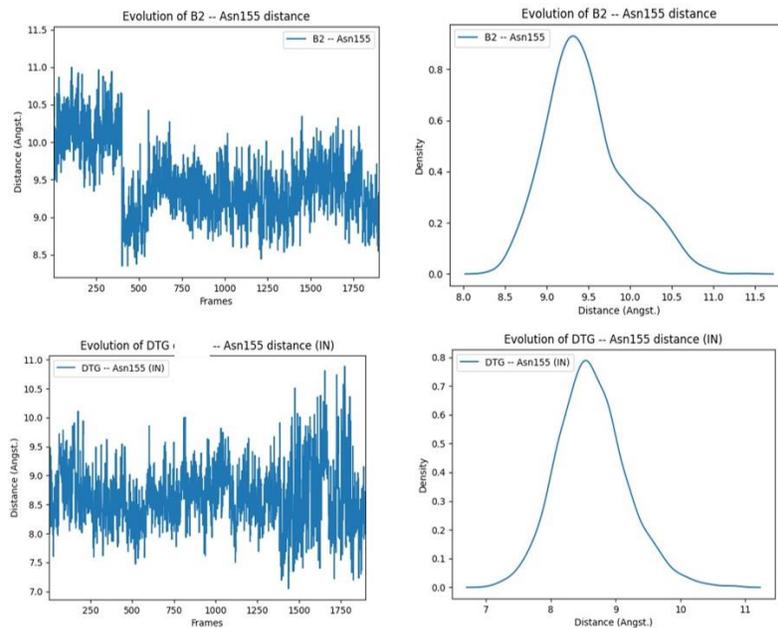


Figure 88 : Graphiques représentant l'évolution de la distance entre B2 et l'Asparagine 155 (correspondant à Asn224 chez le PFV) (à gauche en haut) et entre DTG et l'Asparagine 155 dans le complexe Intégrase-DTG (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B2 avec l'Asparagine 155 et en bas dans le cas du DTG en présence de l'Asparagine 155 dans le complexe Intégrase - DTG).

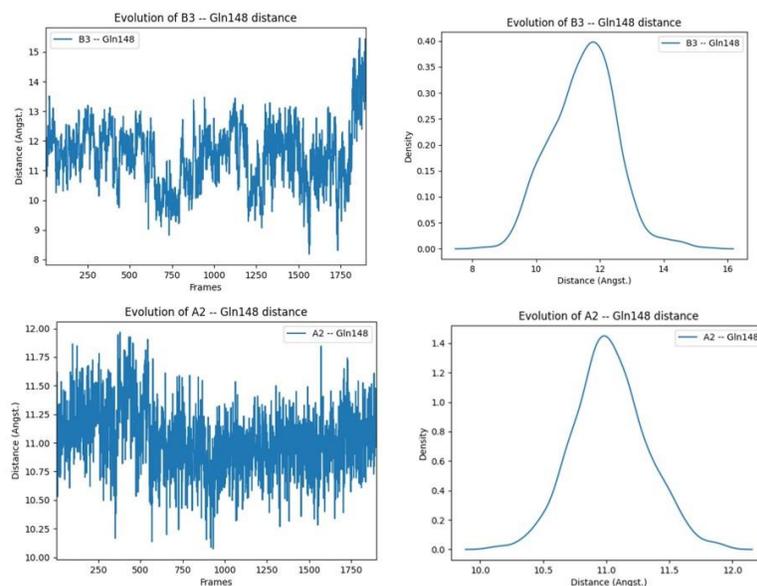


Figure 89 : Graphiques représentant l'évolution de la distance entre B3 et la Sérine 217 du PFV correspondant à la Glutamine 148 (Gln148) (à gauche en haut) et entre A2 et la Sérine 217 du PFV (Gln 148 du VIH-1) (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B3 avec Gln148 et en bas dans le cas d'A2 en présence de Gln148).

Pour le DTG dans le complexe Intégrase-DTG, la distance entre la drogue Asn155 varie entre 8 Å et 10,5 Å avec la plupart des distances égales à 8,5 Å dans nos simulations (Figure 88, graphiques en bas).

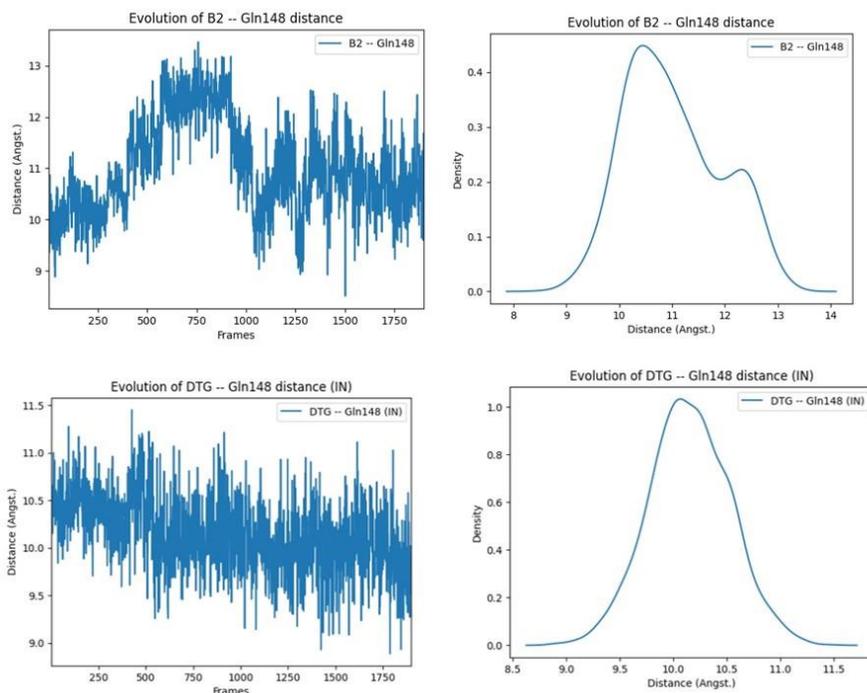


Figure 90 : Graphiques représentant l'évolution de la distance entre B2 et la Sérine 217 du PFV correspondant à la Glutamine 148 (Gln148) du VIH-1 (à gauche en haut) et entre DTG et la Sérine 217 (Gln148 du VIH-1) dans le complexe Intégrase-DTG (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B2 avec Gln148 et en bas dans le cas du DTG en présence de Gln148 dans le complexe Intégrase - DTG).

Alors que pour le DTG dans le complexe Intégrase – DTG, la distance Ser217-DTG fluctue entre 9,3-11,3 Å. La plupart des trajectoires situent le DTG à une distance approximative de 10 Å de la Sérine 217 du PFV (Figure 90, graphiques en bas).

Discussion des valeurs obtenues concernant l'interaction des inhibiteurs avec les résidus subissant des mutations de résistance primaire

- DTG, B3, A2 et B2 perdent leurs liaisons de vdW avec la Tyrosine 143 au cours des simulations, vu que leur distance par rapport à ce résidu augmente au

fur et à mesure, par rapport à la valeur de départ située entre 3,5 Å et 4 Å. En absence d'interaction directe avec ce résidu, ils peuvent échapper ainsi à sa mutation de résistance. Ce qui peut notamment contribuer à la moindre sensibilité de DTG par rapport à RAL et EVG, démontrée antérieurement et conservée chez les nouveaux dérivés.

- DTG, B3, A2 et B2 n'entrent pas en contacts directs avec les résidus Asn155 et Gln148 (dans le cas du PFV, il s'agit de Ser217). Ils échappent ainsi aux mutations de résistance primaires au niveau de ces résidus, ce qui explique aussi l'efficacité de DTG par rapport à RAL et EVG et prouve en plus que nous avons conservé par nos substitutions la capacité de nos dérivés à échapper aux mutations de résistance de l'IN. Le premier et le second point indique donc un accord entre nos résultats *in silico* et les résultats *in vitro* de notre équipe, mentionnés précédemment.
- Notons que les graphiques et les courbes obtenues dans le cas de nos quatre ligands présentent peu de fluctuations. Ceci indique que nos systèmes simulés par dynamique moléculaire sont stables au cours des 400 ns et que les distances obtenues sont correctes reflétant bien la réalité au niveau des complexes.
- En récapitulant dans le Tableau 6 l'ensemble des distances majoritaires de nos dérivés par rapport à Tyr143, Asn155 et Gln148, nous remarquons que B3 est le composé le plus éloigné des résidus sujets aux mutations de résistance primaire, avec une moyenne des distances de 12 Å, suivi du composé A2 puis de B2 et finalement du DTG. Le fait que nos dérivés sont encore plus éloignés des résidus de l'intégrase par comparaison au DTG, contribue à leur meilleure efficacité par rapport à ce dernier. Ce nouveau

paramètre renforce encore le profil plus le profil inhibiteur de B3. Des calculs de l'énergie libre d'interaction des dérivés au sein de l'intasome entier seraient effectués dans la partie C du chapitre IV pour compléter les observations faites par dynamique moléculaire.

Tableau 6 : Les valeurs majoritaires des distances entre nos inhibiteurs et les résidus de l'intégrase du VIH-1 subissant des mutations de résistance primaires ; Tyrosine 143 (Tyr143), Asparagine 155 (Asn155) et Glutamine 148 (Gln148) correspondant à la Sérine 217 du Prototype Foamy Virus (PFV). Ces distances sont obtenues à partir des simulations de dynamique moléculaire des complexes ADN-Intégrase-Inhibiteurs, d'une durée totale de 400 ns.

Inhibiteurs	Distance Inhibiteur-Tyr143	Distance Inhibiteur-Asn155	Distance Inhibiteur-Gln148
DTG	8,5 Å	9,3 Å	9,5 Å
B3	12,5 Å	12 Å	12 Å
A2	11 Å	10 Å	11 Å
B2	9 Å	9,5 Å	10,5 Å

- Le DTG « contrôle » affichant envers les résidus protéiques étudiés un comportement proche de celui des inhibiteurs halogénés ; nous observons une perte des interactions de vdW avec la Tyr143 et des distances éloignées des résidus Asn155 et Ser217 (Gln148 du VIH-1). Nous en concluons que le noyau halogéné ne joue pas de rôle direct dans les interactions des inhibiteurs avec ces résidus et que sa contribution n'est pas primordiale.
- Dans le cas du DTG en présence de l'intégrase virale et en absence d'ADN, les contacts de vdW sont perdus immédiatement dès le début des simulations, ce qui souligne davantage l'instabilité du DTG en présence d'un seul partenaire seulement (la protéine virale dans ce cas) et la

nécessité de formation du complexe binaire IN-ADNv pour une activité optimale des INSTIs.

B.IV. Conclusions du chapitre III partie B

Nous avons effectué des simulations de dynamique moléculaire d'une durée de 400 ns des systèmes suivants : Intégrase-ADN, Intégrase-ADN-DTG, Intégrase-ADN-B3, Intégrase-ADN-A2, Intégrase-ADN-B2, Intégrase-ADN-DTG « contrôle », ADN-DTG et Intégrase-DTG. Nous avons ensuite analysé les variations de distances et d'empilement entre des résidus de l'ADN et de l'intégrase en absence et en présence de molécules inhibitrices, entre des résidus de l'ADN et nos inhibiteurs et des résidus de l'intégrase et nos inhibiteurs. Premièrement, l'analyse des interactions identifiées entre des résidus actifs de l'intégrase et des bases primordiales de l'ADN viral, en absence et en présence des inhibiteurs a souligné le rôle important que jouent ces derniers dans la perturbation de dites interactions, notamment au niveau de l'Adénine A₁₇ ; base virale terminale qui est le siège de la réaction de transfert de brin. En fait, DTG, B3, A2 et B2 éloignent cette base du résidu de la triade catalytique Glu152 et perturbent ses interactions avec la Lysine 159 et la Gln146. A₁₇ devient alors peu stable avec une flexibilité élevée et n'a plus de contacts proches avec les résidus de l'intégrase. Ayant une influence importante sur les interactions réalisées par l'Adénine, DTG et nos dérivés méritent alors l'appellation d'inhibiteurs de transfert de brin. De plus, nos inhibiteurs affectent la liaison entre la Lysine 156 et C₁₆, entre la Lysine 160 et le squelette A₇-T₈ et entre la Glycine 149 et G₄, affectant davantage le réseau d'interactions établies entre la protéine et l'ADN notamment au niveau des bases C₁₆ et G₄. Nous avons donc pu en renforçant les interactions des drogues avec les bases d'ADN viral diminuer les

contacts de ces dernières avec l'intégrase, favorisant ainsi un mode d'inhibition spécifique et efficace. Nous nous intéressons également aux interactions ADN-inhibiteurs et l'étude menée à ce niveau nous permet de conclure que nos nouveaux dérivés B3, A2 et B2 effectuent un meilleur empilement avec les noyaux de C₁₆ et A₁₇ comparés au DTG ainsi que des contacts plus proches avec G₄ via leur groupement en para (NHCH₃) par rapport au contact du fluor du DTG avec G₄. Ils sont donc caractérisés par une interaction plus spécifique et plus efficace avec les bases terminales primordiales de l'ADN viral ; C₁₆, A₁₇ et G₄. Ces résultats confirment les meilleures valeurs de l'énergie d'interaction des noyaux inhibiteurs de nos dérivés avec C₁₆ et G₄, calculées dans le chapitre IV partie A. Ils expliquent de même la diminution d'interactions clés entre l'intégrase et l'ADN viral. En effet, les interactions spécifiques de nos inhibiteurs avec C₁₆ par empilement π - π perturbent les contacts proches de cette dernière avec la Lysine156 et la Lysine 159. De plus, les interactions de vdW de nos inhibiteurs avec G₄ affectent les contacts de la Guanine avec Gly149. Finalement, en interagissant par empilement avec A₁₇, nos dérivés inhibiteurs et le DTG diminuent les contacts de cette base terminale avec la triade catalytique (Glu152), avec la Lysine159 et la Glutamine146, soulignant leur rôle d'inhibiteurs de transfert de brin.

Outre leurs interactions spécifiques avec les bases virales de l'ADN, nos inhibiteurs effectuent des interactions avec certains résidus clés de l'intégrase, situés au niveau de l'hélice α 4 et de la boucle qui la précède. Ainsi ils présentent toujours une sensibilité vis-à-vis de certaines mutations au dépend d'autres. En effet, nous avons identifié la conservation des liaisons de vdW entre le groupement en ortho du DTG, B3, A2 et B2 et le Glutamate 152, ainsi qu'entre nos dérivés et la Proline 145. De plus, des liaisons hydrogènes sont observées entre ces dérivés et la Glycine

118. Cependant, ils n'interagissent pas étroitement avec la Tyrosine 143, l'Asparagine 155 et la Glutamine 148 ce qui leur permet d'échapper aux mutations de résistance de ces résidus et favorise alors leur efficacité par rapport aux INSTIs mis antérieurement sur le marché (RAL et EVG). Nous pouvons alors conclure que les substitutions réalisées sur les nouveaux dérivés ne renforcent pas seulement leur interaction avec l'ADN viral mais diminuent également leur sensibilité vis-à-vis de certaines mutations de résistance développées au niveau de l'intégrase.

Nos résultats prouvent de même l'importance de la présence simultanée de la protéine et de l'ADN pour un mode d'inhibition optimal du DTG, B3, A2 et B2. A titre d'exemple, au niveau du complexe binaire étudié, les interactions entre le DTG et l'IN sont perdues, notamment au niveau de la Gly118, Pro145 et Glu152, en absence d'ADNv, confirmant l'hypothèse que nos inhibiteurs interfaciaux ont besoin de l'IN et de l'ADN en conformation active pour un mode d'action efficace. En effet, même si nos inhibiteurs affichent des interactions spécifiques avec les bases virales, ils restent positionnés à l'interface ADN-Intégrase, effectuant des liaisons avec des résidus clés de cette dernière, pour une meilleure efficacité.

De plus, la structure halogénée au niveau de nos dérivés, quant à elle, joue un rôle indirect et peu spécifique dans l'interaction des dérivés avec Gly118 et Pro145. Cependant son rôle est direct et important dans l'interaction de DTG, B3, A2 et B2 avec Glu152 puisqu'elle contient la substitution en ortho. Ceci souligne le rôle primordial que joue le noyau halogéné de par son interaction avec les bases réactives de l'ADN et d'autre part son implication directe dans les interactions des drogues avec le site catalytique de l'IN. Enfin, les trois dérivés retenus prouvent une fois de plus leur efficacité accrue comparés au DTG, avec particulièrement B3

qui présente des interactions renforcées avec C₁₆, A₁₇ et G₄ d'une part et Glu152 d'autre part. Ces affinités accrues spécifiquement pour les bases virales se sont manifestées par son rendement énergétique évalué auparavant et sont notamment dues à la présence de brome en ortho qui contribue à la création d'une zone très riche en électrons au niveau de son noyau, favorisant surtout les interactions avec la Cytosine. De plus, B3 a présenté les distances les plus éloignées des résidus de l'IN concernées par les mutations ce qui promet également de réduire sa sensibilité aux mutations de résistance. Cette efficacité remarquable de B3 est à exploiter davantage par des calculs d'énergie libre de ses interactions au sein du complexe ADN-Intégrase, que nous comparerons aux énergies libres de liaison de nos autres complexes, pour compléter les résultats décrits par des affinités quantitatives.

Chapitre IV

Résultats et Discussion

Partie C : Calculs d'énergie libre alchimique absolue des interactions au sein des complexes simulés du VIH-1

C.I. Introduction générale de la partie C

La dernière partie de notre projet de travail porte sur les calculs d'énergie libre alchimique absolue (ΔG) des interactions de nos inhibiteurs (DTG, B3, A2, B2 et DTG «contrôle») avec l'intégrase et l'ADN viral au sein des intasomes simulés par dynamique moléculaire. Cette approche est aussi une première au niveau des calculs de ΔG effectués en champ de force polarisable (AMOEBA) pour de grands systèmes complexes de près de 200 000 atomes. En effet, après avoir calculé les énergies d'interaction (ΔE) du noyau halogéné de nos inhibiteurs avec les noyaux de C₁₆ et G₄ et après avoir simulé par dynamique moléculaire les complexes entiers des intasomes du PFV et analysé les interactions à leur niveau, nous désirons voir si la meilleure efficacité de nos noyaux par rapport au DTG, observée par les valeurs de ΔE , peut être reproduite à un niveau plus complexe par calcul des énergies libres absolues des dérivés entiers (et non seulement leurs noyaux) avec l'ADN viral entière (et non seulement les noyaux de C₁₆ et G₄) et l'intégrase complète dans un milieu aqueux mimant le milieu physiologique cellulaire. Ceci nous permettra de s'approcher plus de la réalité en générant des valeurs de ΔG fiables et correctes. Nous voulons aussi voir si les interactions et les liaisons plus spécifiques qu'effectuent B3, A2 et B2 par rapport au DTG au niveau des complexes entiers simulés peuvent être traduites par les valeurs énergétiques de ΔG , notamment la meilleure efficacité de B3 observée dans nos dynamiques. Notre approche alchimique devrait donc compléter et soutenir les résultats trouvés dans les parties

précédentes. Ce travail sera présenté dans un article qui est en cours de rédaction. L'intitulé ainsi que son résumé se trouvent en annexe à la page 295.

C.II. Notre protocole pour le calcul des énergies libres alchimiques absolues

Le calcul de l'énergie libre absolue se base sur des simulations de dynamiques moléculaires non conventionnelles des systèmes étudiés. La démarche comprend plusieurs étapes que nous présenteront ci-dessous.

- *Etape 1 : Choix des systèmes étudiés*

Nous commençons en premier lieu par définir les systèmes pour lesquelles ΔG sera calculée. Nous avons utilisé les systèmes du PFV, déjà adoptés pour nos analyses de dynamique moléculaire conventionnelle. Ils sont les suivants : ADN-Intégrase-DTG, ADN-Intégrase-B3, ADN-Intégrase-A2, ADN-Intégrase-B2 et ADN-Intégrase-DTG «contrôle ». Ils nous permettront de calculer les énergies d'interaction des inhibiteurs avec leurs deux partenaires ; l'ADN viral et l'intégrase. Pour rappel, l'absence de cristal de l'intégrase entière du VIH-1 et la grande homologie entre l'intégrase du VIH-1 et celle du PFV nous permettent d'adopter le cristal de ce dernier dans nos calculs de ΔG et d'avoir des résultats fiables et précis reproductibles chez le VIH-1. De plus, nous avons calculé les énergies libres d'interaction au sein des systèmes ADN-B3, ADN-A2, ADN-B2 et ADN-DTG dans le but de voir si la meilleure interaction des trois dérivés avec les bases de l'ADN que nous avons observée par les calculs de ΔE sera aussi reproduite au niveau des drogues entières avec l'ADN viral entière. Nous avons aussi eu recours aux

systemes ADN-DTG et Intégrase-DTG pour calculer l'énergie libre de la drogue en présence de chacun d'eux, afin de voir premièrement si le DTG interagit mieux au sein du complexe ternaire qu'au sein des complexes binaires (comme identifié par nos analyses des dynamiques moléculaires conventionnelles) et deuxièmement, s'il présente une meilleure interaction avec l'ADN viral qu'avec l'intégrase. Ceci confirmera l'hypothèse énoncée précédemment par El Khoury et al au sein de notre équipe, que la meilleure efficacité du DTG par rapport aux inhibiteurs de transfert de brin précédents est due à sa meilleure interaction spécifique avec l'ADN viral, l'aidant à surmonter les mutations de résistance de la protéine [20][21].

- Etape 2 : Préparation des systèmes

Les étapes de préparation des systèmes à simuler pour les calculs de ΔG sont identiques à celles que nous avons suivies pour les simulations de dynamique moléculaire conventionnelle. Les structures de départ pour nos calculs sont donc les onze systèmes solvatés dans des boîtes d'eau cubiques en présence de 150 mM de KCl, minimisés puis équilibrés à 300K et au pH physiologique (égal à 7) avec le champ de force polarisable AMOEBA, implémenté dans Tinker-HP.

- Etape 3 : Détermination des états intermédiaires

Avant de commencer les simulations de dynamique moléculaire, le cycle thermodynamique alchimique doit être déterminé ; nous devons donc définir les états initial et final de nos complexes ainsi que les états intermédiaires entre les deux, permettant le passage de la position de départ à la position finale pour une estimation correcte de l'énergie libre de liaison. Pour tous nos complexes mentionnés ci-dessus, l'état initial correspond aux ligands (DTG, B3, A2, B2 et DTG « contrôle ») seuls dans la boîte de simulation et l'intégrase et l'ADN dans une aute

boîte identique, et l'état final correspond aux ligands situés dans leur site de liaison à l'interface ADN-Intégrase. Nous aurons donc deux ensembles à simuler ; les ligands seuls dans la boîte d'eau ce qui va générer leur énergie libre de déshydratation ($\Delta G_{\text{déshyd}}$) et les ligands en présence de leurs partenaires (ADN ou Intégrase ou les deux) pour obtenir leur énergie libre de liaison ($\Delta G_{\text{liaison}}$).

Pour les énergies libres de déshydratation, les ligands commencent en interaction totale avec les molécules d'eau ; les systèmes sont caractérisés par $\lambda_{\text{électrostatique}}$ et λ_{vdW} égales à 1. Ensuite, nous effectuerons leur découplage théorique afin d'éliminer tout d'abord leurs interactions électrostatiques en variant $\lambda_{\text{électrostatique}}$ de 1 à 0 tout en gardant $\lambda_{\text{vdW}} = 1$ puis nous procéderons à l'élimination des interactions de VdW du ligand avec les molécules d'eau en variant λ_{vdW} de 1 à 0. Ces variations de λ de 1 à 0 engendrent la « disparition » des ligands de leur boîte de simulation et donc leur passage de la phase liquide à la phase gazeuse. Vient après le calcul de $\Delta G_{\text{liaison}}$ où le ligand en phase gazeuse « réapparaît » au niveau de son site de liaison dans la boîte contenant l'ADN et l'Intégrase, suite à la variation de λ de 0 à 1, en commençant dans ce cas par le passage de λ_{vdW} de 0 à 1 en gardant $\lambda_{\text{électrostatique}}$ nulle qui sera ensuite à son tour mise à l'échelle par un passage de 0 à 1. Avec les ensembles de λ égales à 1, nos ligands sont en interaction totale au sein de l'intasome. Notons que lors de la variation du paramètre thermodynamique λ , chaque valeur qu'il adopte correspond à un état intermédiaire alchimique (n'existant pas en réalité) qui doit être simulé dynamiquement. Plus on ajoute des simulations intermédiaires, plus on augmente le chevauchement de l'espace de phase entre les états intermédiaires. Ceci permet alors une estimation plus correcte de l'énergie libre de liaison. Le nombre et

l'espacement des états intermédiaires sont adaptés spécifiquement au système d'intérêt.

- Etape 4 : Détermination des contraintes

Les contraintes que nous avons utilisées pour fixer le ligand au sein du site actif au cours des simulations des systèmes ADN-Intégrase-Inhibiteurs sont les contraintes de Boresch, étant donné que la position optimale du DTG est connue et est décrite dans la structure PDB de code 3S3M. Ce type de contraintes ajoutées, comme déjà mentionné, nous permet de garder les ligands en leurs positions et orientations initiales bien définies dans la structure PDB. Nous avons déterminé trois atomes de nos ligands et trois atomes de l'intégrase et de l'ADN viral pour y appliquer les contraintes au niveau d'une distance, de deux angles et de trois dièdres formés par les atomes choisis. Pour le choix des atomes, il convient à choisir ceux qui se trouvent dans des parties rigides de la protéine et de l'ADN viral comme le squelette des acides aminés et le squelette phosphodiester respectivement. Quant aux ligands il convient de choisir des atomes de référence au niveau de leurs cycles aromatique. Ceci leur permettra de rester fixés dans leurs positions et orientations de départ trouvées dans le fichier PDB assurant une estimation correcte et exacte des énergies libres. La Figure 117 illustre les six atomes de référence pour l'application des contraintes de Boresch, dans le cas du complexe ADN-Intégrase-DTG (les atomes de référence sont les mêmes pour tous les complexes). Nous avons gardé le noyau d'halobenzène libre afin de lui permettre de bien prendre sa position optimale au cours des simulations.

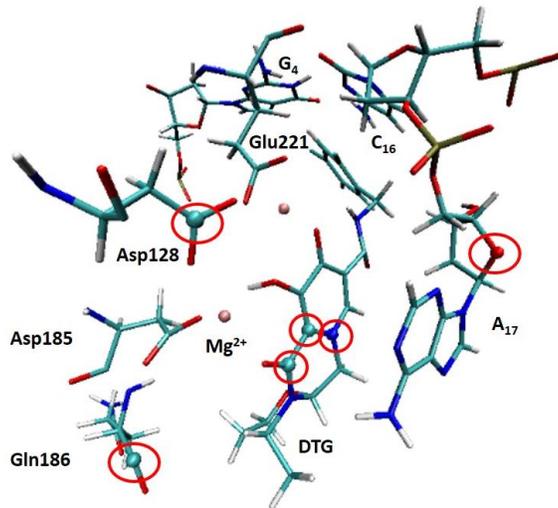


Figure 117 : Position de départ de l'intasome du Prototype Foamy Virus (PFV) en présence du Dolutégravir (DTG) et des ions Mg^{2+} . Le DTG est présent à l'interface ADN-Intégrase chélatant deux ions Mg^{2+} représentés par des sphères roses. La triade catalytique Asp128, Asp185, Glu221 et le résidu Gln186 sont schématisés en bâtons (ils correspondent à Asp64, Asp116, Glu152 et Asn117 chez le VIH-1 respectivement). Les bases réactives de l'ADN viral au niveau de l'extrémité 3' terminal sont représentées de même en bâtons ; G₄, C₁₆ et A₁₇. Les atomes fixés pour les contraintes de Borech sont sous forme de boules et encerclés.

Pour les valeurs de référence des distances, des angles et des dièdres, elles correspondent aux valeurs trouvées dans les complexes après leur équilibration ; en effet la position équilibrée des systèmes est le point de départ pour nos simulations puisque notre approche adoptée pour le calcul des ΔG nécessite des systèmes déjà équilibrés. Pour la constante de force (κ) indiquée dans la formule de la page 143, nous avons utilisé une valeur de $10 \text{ kcal/mol/\AA}^2$. En effet, il a été démontré que cette valeur est suffisante pour bien fixer le ligand et fournir des valeurs exactes de ΔG_{rest} ; Borech et ses collaborateurs ont testé des constantes de force allant de 5 à $50 \text{ kcal/mol/\AA}^2$, prouvant que la valeur de l'énergie libre des contraintes est indépendante de la rigidité de ces dernières. Cependant, une valeur trop faible (5 kcal/mol/\AA^2) ou trop élevée ($50 \text{ kcal/mol/\AA}^2$) peut affecter la position

du ligand et la précision des résultats. Pour cela, 10 kcal/mol/Å² est en pratique la valeur idéale à utiliser [23] [24].

- Etape 5 : Déroulement des simulations

Comme mentionné précédemment, deux ensembles de simulations de dynamique moléculaire seraient effectuées pour chaque système afin de générer l'énergie de déshydratation du ligand et son énergie de complexation dans son site de liaison. A chaque simulation est attribuée une valeur de $\lambda_{\text{électrostatique}}$ et λ_{vdW} pour permettre une transition correcte de l'état initial à l'état final. Les simulations sont effectuées parallèlement et indépendamment les unes des autres. La Figure 118 illustre un exemple du chemin thermodynamique adopté (cas du complexe ADN-Intégrase). Tout d'abord, nous commençons par le DTG seul et libre dans la boîte d'eau, effectuant des interactions intermoléculaires avec les molécules du solvant ($\lambda_{\text{électrostatique}}$ et $\lambda_{\text{vdW}} = 1$). Nous le découplons ensuite progressivement en commençant par l'élimination de ses interactions électrostatiques puis de ses interactions de vdW. Nous avons choisi d'espacer les valeurs de λ de 0,1. Ainsi nous aurons les états intermédiaires suivants : $\lambda_{\text{électrostatique}} = 1 \lambda_{\text{vdW}} = 1 / \lambda_{\text{électrostatique}} = 0,9$ $\lambda_{\text{vdW}} = 1 / \lambda_{\text{électrostatique}} = 0,8$ $\lambda_{\text{vdW}} = 1 / \lambda_{\text{électrostatique}} = 0,7$ $\lambda_{\text{vdW}} = 1$ jusqu'à arriver à $\lambda_{\text{électrostatique}} = 0 \lambda_{\text{vdW}} = 1$. Ensuite, l'élimination des interactions de VdW peut débuter avec les étapes intermédiaires suivantes : $\lambda_{\text{électrostatique}} = 0 \lambda_{\text{vdW}} = 1 / \lambda_{\text{électrostatique}} = 0$ $\lambda_{\text{vdW}} = 0,9 / \lambda_{\text{électrostatique}} = 0$ $\lambda_{\text{électrostatique}} = 0,8 / \lambda_{\text{électrostatique}} = 0$ $\lambda_{\text{vdW}} = 0,7$ jusqu'à arriver à $\lambda_{\text{électrostatique}} = 0 \lambda_{\text{vdW}} = 0$. L'état final du DTG correspond à un ligand sans interactions intermoléculaires, retenu dans le vide. Nous passons ensuite à la simulation du ligand au sein du complexe ternaire, en commençant par des valeurs de $\lambda_{\text{électrostatique}}$ et λ_{vdW} égales à zéro. Nous effectuons des simulations en variant tout

d'abord λ_{vdW} de 0 à 1, puis $\lambda_{\text{électrostatique}}$ de 0 à 1. Les étapes intermédiaires ont les mêmes valeurs de λ que celles utilisées pour générer l'énergie libre de déshydratation, espacées de 0,1. Les contraintes de Boresch sont appliquées au ligand pendant ce deuxième ensemble de simulations. L'énergie libre de la contribution de ces contraintes est calculée analytiquement et est prise en considération dans le calcul de l'énergie libre finale. Pour les paramètres des calculs au cours des simulations de dynamique moléculaire, nous avons utilisé les composantes suivantes : l'intégrateur BAOABRESPA1, le thermostat Bussi pour le maintien de la température à 300K et le barostat Berendsen pour le maintien de la pression à 1 atm. La durée de chaque état simulé est de 10 ns dans un ensemble isotherme-isobarique (NPT), avec un pas de temps de 10 fs et une impression des données chaque 5 ps.

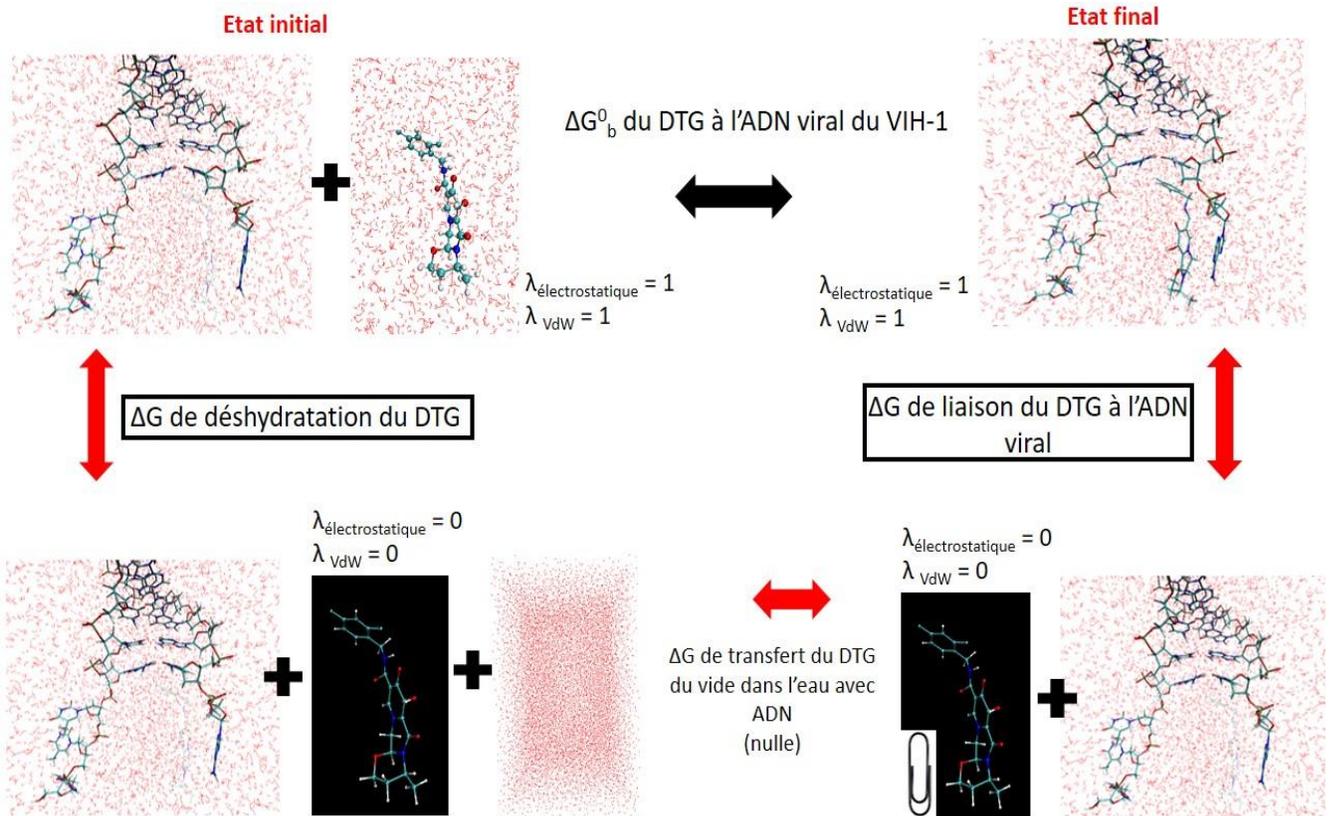


Figure 118 : Exemple d'un cycle thermodynamique que nous avons utilisé pour calculer l'énergie libre de liaison du Dolutégravir (DTG) au niveau de l'ADN viral. Les étapes sont les mêmes pour les autres systèmes. Le cycle comprend majoritairement deux parties ; la partie à gauche représentant le calcul de l'énergie libre de déshydratation du DTG et la partie à droite comprenant le calcul de l'énergie de liaison du DTG au niveau de l'ADN viral. Au cours du deuxième calcul, le DTG est restreint (représentation des contraintes par un trombone) afin de maintenir sa position à l'équilibre fixe au cours des simulations. Le paramètre lambda permet de contrôler la variation des interactions intermoléculaires ; les interactions électrostatiques et de vdW. Pour une valeur égale à 1, les interactions sont totales alors que pour une valeur égale à 0, il y a absence d'interactions intermoléculaires.

- Etape 6 : Analyse des données

Une fois les simulations de dynamique moléculaire de nos complexes sont lancées et achevées, nous passons à l'analyse des données collectées. Nous devons estimer l'énergie libre totale de déshydratation de nos ligands ainsi que l'énergie libre totale de leur liaison à leur cible. La première correspond à la somme des différences d'énergie libre de chaque état intermédiaire lors de déshydratation du

ligand et la seconde est égale à la somme des différences d'énergie libre de chaque état intermédiaire lors de son « apparition » au niveau du site de liaison. Tinker-HP, notre progiciel de simulation fournit le programme BAR incluant les méthodes BAR et FEP afin d'analyser nos résultats. Pour rappel, cet estimateur requiert les différences d'énergie potentielle entre deux états intermédiaires voisins afin de générer les valeurs de ΔG . Nous ajouterons à ces contributions l'énergie libre des contraintes de Boresch calculées analytiquement ; ΔG_{rest} (comprenant la correction de l'état standard). L'énergie libre absolue totale à l'état standard (ΔG^0_b) sera calculée par la formule suivante :

$$\Delta G^0_b = \Delta G_{\text{liaison totale}} + \Delta G_{\text{déshydratation}} + \Delta G_{rest}$$

Notons qu'une incertitude est attribuée à chaque estimation de l'énergie libre, qui devra être prise en compte dans le résultat final ^[25].

C.III. Résultats des calculs des énergies libres absolues alchimiques

C.III.1. Calculs des énergies libres absolues alchimiques des complexes ADN-inhibiteurs

Nous avons effectué en premier lieu les simulations des ligands B3, A2, B2 et DTG en présence de l'ADN viral uniquement (nous avons éliminé l'intégrase de la structure initiale). Les résultats des valeurs de ΔG calculées sont présentés dans le Tableau 8 ci-dessous. $\Delta G_{\text{liaison}}$ (ADN-ligand) correspond à l'énergie libre absolue de l'interaction de l'ADN avec le ligand en présence des molécules d'eau et des contraintes. $\Delta G_{\text{déshydratation}}$ correspond quant à elle à l'énergie libre de l'élimination des interactions intermoléculaires entre le ligand et l'eau d'où sa valeur positive. ΔG_{rest} représente l'énergie libre d'ajout de contraintes aux ligands pour les fixer au niveau de leur site. Elle comprend la correction à l'état standard. Finalement,

l'énergie libre absolue totale à l'état standard (ΔG^0_b) est obtenue en soustrayant de $\Delta G_{\text{liaison}}$ les contributions de $\Delta G_{\text{déshydratation}}$ et ΔG_{rest} . Nous avons ajouté dans la première ligne du Tableau 8 les valeurs de l'énergie d'interaction intermoléculaire (ΔE) que nous avons calculées dans la partie A de ce chapitre pour les comparer aux valeurs de ΔG^0_b .

Tableau 8 : Tableau représentant les valeurs de l'énergie libre absolue calculée pour les complexes ADN-DTG, ADN-B3, ADN-A2 et ADN-B2. Les valeurs de l'énergie d'interaction intermoléculaire (ΔE) que nous avons calculées précédemment sont indiquées aussi. ΔG totale d'interaction est obtenue par la formule suivante : $\Delta G^0_b = \Delta G_{\text{liaison}} + \Delta G_{\text{déshydratation}} + \Delta G_{\text{rest}}$

	ADN-Composé B3	ADN-Composé A2	ADN-Composé B2	ADN-DTG
ΔE totale d'interaction (C ₁₆ -G ₄ -ligand)	-31,2 kcal/mol	-31 kcal/mol	-31 kcal/mol	-27,1 kcal/mol
$\Delta G_{\text{liaison}}$ (ligand – ADN)	-108,49 kcal/mol	-107,85 kcal/mol	-107,53 kcal/mol	-106,94 kcal/mol
$\Delta G_{\text{déshydratation}}$	94,12 kcal/mol	95,66 kcal/mol	95,83 kcal/mol	96,69 kcal/mol
ΔG_{rest}	7,58 kcal/mol	7,46 kcal/mol	7,22 kcal/mol	7,42 kcal/mol
ΔG totale d'interaction ΔG^0_b	-6,79 kcal/mol +/- 0.1	-4,73 kcal/mol +/- 0.1	-4,48 kcal/mol +/- 0.1	-2,83 kcal/mol +/- 0.1

Les énergies libres absolues de la liaison des ligands à l'ADN viral sont relativement proches entre elles et varient entre -106,94 kcal/mol et 108,49 kcal/mol. La contribution énergétique des molécules d'eau (énergie libre de déshydratation) présente des valeurs entre 94,12 kcal/mol et 96,96 kcal/mol et la contribution des contraintes est à peu près la même pour tous les composés avec une moyenne de 7 kcal/mol.

Discussion des valeurs des énergies libres absolues obtenues pour les complexes ADN-inhibiteurs

- Les valeurs des énergies libres absolues de liaison des ligands à la totalité de l'ADN dans l'eau sont des valeurs négatives (deuxième ligne du Tableau 8),

ce qui signifie que l'interaction entre les deux est stable et favorable. C'est une interaction coopérative.

- L'énergie libre absolue totale à l'état standard (ΔG^0_b) présente de même des valeurs négatives pour tous les composés (dernière ligne du Tableau 8), indiquant que l'interaction spécifique entre les ligands et l'ADN reste stable et coopérative après élimination des contributions des molécules d'eau et des contraintes.
- Il existe des critères d'évaluation de la fiabilité des valeurs de ΔG totale obtenues ; tout d'abord, toute valeur inférieure ou égale à -15 kcal/mol n'est pas correcte. De plus, un taux d'erreur supérieur à 1 kcal/mol indique une valeur non fiable [27]. Ceci dit, nos valeurs énergétiques obtenues étant supérieures à -15 kcal/mol (entre -2,83 kcal/mol et -6,79 kcal/mol) avec un taux d'erreur de +/- 0,1 kcal/mol sont correctes.
- Les résultats obtenus pour les calculs de ΔG au niveau des complexes entiers inhibiteurs-ADN sont en parfait accord avec ceux obtenus pour les calculs de ΔE au niveau des mini-complexes (noyau C₁₆-noyau G₄- noyau inhibiteur). Tous nos nouveaux dérivés présentent des valeurs de $\Delta G_{liaison}$ et ΔG^0_b plus grandes en valeurs absolues que les valeurs du DTG. L'interaction plus stable et plus spécifique de leurs noyaux avec C₁₆ et G₄ par rapport au DTG s'est reproduite au niveau du complexe entier en milieu aqueux. Ce qui met en relief davantage l'efficacité de nos nouveaux inhibiteurs par rapport au DTG.
- B3 présente l'interaction la plus spécifique avec l'ADN virale, ayant un ΔG^0_b de -6,79 kcal/mol, le plus élevé en valeur absolue par rapport aux valeurs du DTG, A2 et B2. Sa meilleure efficacité que nous avons observée avec C₁₆ et G₄ (valeur de ΔE la plus stable de -31,2 kcal/mol) ainsi qu'au niveau des

simulations de dynamique conventionnelle est observée aussi au niveau de sa valeur d'énergie libre d'interaction avec l'ADN virale entière, confirmant ainsi les observations faites dans la partie B et les calculs effectués dans la partie A de ce chapitre et soulignant l'efficacité de la présence du brome en para comme substitution.

C.III.2. Calculs des énergies libres absolues alchimiques des complexes ADN-Intégrase-Inhibiteurs

Nous avons simulé ensuite, selon le chemin thermodynamique expliqué dans la partie C.II, les complexes ternaires suivants : ADN-Intégrase-DTG, ADN-Intégrase-B3, ADN-Intégrase-A2, ADN-Intégrase-B2 et ADN-Intégrase-DTG « contrôle », afin de calculer la valeur de l'énergie libre absolue d'interaction totale de chaque ligand avec l'ADN et l'intégrase. Les résultats obtenus sont affichés dans le Tableau 9 ci-après. Ils présentent les mêmes lignes que celles du Tableau 8 mais sans les valeurs de ΔE .

Tableau 9 : Tableau représentant les valeurs de l'énergie libre absolue calculée pour les complexes ADN-Intégrase-DTG, ADN-Intégrase-B3, ADN-Intégrase-A2, ADN-Intégrase-B2 et ADN-Intégrase-DTG « contrôle ». ΔG totale d'interaction est obtenue par la formule suivante : $\Delta G_b^0 = \Delta G_{\text{liaison}} + \Delta G_{\text{déshydratation}} + \Delta G_{\text{rest}}$.

	ADN-Intégrase-Composé B3	ADN-Intégrase-Composé A2	ADN-Intégrase-Composé B2	ADN-Intégrase-DTG	ADN-Intégrase-DTG « contrôle »
$\Delta G_{\text{liaison}}$ ADN-Intégrase-ligand	-110,28 kcal/mol	-109,93 kcal/mol	-109,47 kcal/mol	-109,02 kcal/mol	-100,55 kcal/mol
$\Delta G_{\text{déshydratation}}$	93,92 kcal/mol	95,66 kcal/mol	95,83 kcal/mol	96,70 kcal/mol	88,71 kcal/mol
ΔG_{rest}	7,40 kcal/mol	7,27 kcal/mol	7,22 kcal/mol	7,27 kcal/mol	7,46 kcal/mol
ΔG totale d'interaction ΔG_b^0	-8,96 kcal/mol +/- 0.1	-7,00 kcal/mol +/- 0.1	-6,42 kcal/mol +/- 0.1	-5,05 kcal/mol +/- 0.1	-4,38 kcal/mol +/- 0.1

Les valeurs de $\Delta G_{\text{liaison}}$ (ADN-Intégrase-ligand) varient entre -100,55 kcal/mol (cas du DTG « contrôle ») et -110,28 kcal/mol (cas du complexe ternaire en présence de B3). L'énergie libre de déshydratation varie entre 88,71 kcal/mol et 96,70 kcal/mol et celle de la contribution des contraintes reste la même approximativement pour tous les ligands avec une moyenne de 7 kcal/mol.

Discussion des valeurs des énergies libres absolues obtenues pour les complexes ADN-Intégrase-inhibiteurs

- Les valeurs des énergies libres absolues de liaison des ligands à l'interface ADN-Intégrase en milieu aqueux sont des valeurs négatives (première ligne du Tableau 9) indiquant une interaction coopérative et favorable entre chaque ligand et ses partenaires. Cette interaction stable est maintenue après élimination des énergies libres de déshydratation et des contraintes puisque les valeurs de l'énergie libre absolue totale à l'état standard (ΔG^0_b) sont de même négatives pour tous les composés (deuxième ligne du Tableau 9). Ceci signifie que l'interaction spécifique entre les ligands et l'intasome viral reste stable et coopérative même après élimination des contributions des molécules d'eau et des contraintes.
- Les valeurs de ΔG^0_b obtenues pour les complexes ternaires (dernière ligne du Tableau 9) remplissent les conditions exigées pour des valeurs fiables et correctes de l'énergie libre absolue totale, puisqu'elles sont supérieures à -15 kcal/mol (entre -4,38 kcal/mol et -8,96 kcal/mol) avec un taux d'erreur de +/- 0,1 kcal/mol.
- Les nouveaux dérivés B3, A2 et B2 présentent des valeurs de ΔG^0_b plus grandes en valeurs absolues que celle du DTG, ce qui indique qu'ils interagissent d'une manière plus spécifique et efficace avec l'intégrase et

l'ADN que le DTG, confirmant leur meilleure efficacité observée dans nos démarches antérieures. En effet, leur meilleure interaction avec l'ADN viral déduite à partir des valeurs de ΔE , ΔG^0_b (ADN-ligand) et leurs comportements pendant les dynamiques conventionnelles (contacts proches avec des bases virales de l'ADN ainsi qu'avec des résidus clés de l'intégrase) ont contribué à leur meilleure interaction représentée par les valeurs des énergies libres absolues totales obtenues pour les complexes ternaires.

- DTG « contrôle » a présenté la plus faible interaction avec l'ADN et l'intégrase par rapport aux autres ligands puisqu'il a la valeur de ΔG^0_b la plus faible en valeur absolue, de valeur -4,38 kcal/mol. En effet, dans les simulations de dynamique moléculaire conventionnelle, le DTG « contrôle » a présenté le moins de contacts proches avec Glu152 de la triade catalytique et un faible empilement avec A₁₇. Ces interactions peu spécifiques avec des résidus importants au niveau de l'intégrase et de l'ADN respectivement ont aboutit à sa faible valeur d'énergie libre absolue. De plus, en absence de noyau halogéné au sein de sa structure, notre molécule « contrôle » n'effectue pas d'empilement avec C₁₆ ni des interactions électrostatiques avec G₄. Ces absences d'interactions spécifiques avec l'ADN viral ont contribué aussi à ses interactions les moins stables avec l'intasome virale, qui ont traduits par sa valeur de ΔG^0_b la plus faible en valeur absolue.
- Comme dans le cas des interactions ADN-inhibiteurs, l'interaction la plus spécifique et efficace avec l'intasome du PFV correspond au composé B3 qui possède une valeur de ΔG^0_b la plus grande en valeur absolue de -8,96 kcal/mol. Ceci est en accord avec tous les résultats précédents que nous avons trouvés. En fait, cette meilleure interaction de B3 observée en

présence de l'ADN viral et de l'intégrase est due à notamment à ses interactions spécifiques avec C₁₆ et G₄, son meilleur empilement avec A₁₇ et ses contacts les plus proches avec Glu152, comme nous les avons démontrés dans les parties précédentes. Sa valeur d'énergie libre absolue confirme donc les observations faites par dynamique.

- Notons que pour chacun des ligands, DTG, B3, A2 et B2, leurs valeurs de ΔG^0_b en présence de l'intégrase et de l'ADN sont plus stables qu'en présence de l'ADN uniquement. Par exemple, nous trouvons pour B3 au sein du complexe ternaire une valeur énergétique totale de -8,96 kcal/mol alors qu'en présence de l'ADN uniquement, sa valeur de ΔG^0_b est égale à -6,79 kcal/mol et pour A2, son énergie libre absolue totale vaut -7,0 kcal/mol en présence de l'ADN et de l'intégrase et est égale à -4,73 kcal/mol pour son interaction avec l'ADN. Ceci confirme l'hypothèse de Hare et al que nous avons observé dans nos simulations de dynamiques, que les inhibiteurs de transfert de brin ont besoin de la présence simultanée de l'intégrase et de l'ADN viral pour un mode d'action optimale. Nos ligands présentent donc des interactions absolues plus stables quand ils se trouvent à l'interface ADN-Intégrase.

C.III.3. Calculs des énergies libres absolues alchimiques des complexes ADN-DTG et Intégrase-DTG

Finalement, nous avons calculé via le chemin thermodynamique expliqué précédemment, l'énergie libre absolue au niveau du complexe DTG-Intégrase et nous l'avons comparée à la valeur correspondante au système DTG-ADN viral afin de voir si cet inhibiteur interagit plus efficacement avec l'ADN qu'avec la protéine.

Les valeurs obtenues sont présentées dans le Tableau 10 ci-dessous.

Tableau 10 : Tableau représentant les valeurs de l'énergie libre absolue calculée pour les complexes Intégrase-DT et ADN-DTG. ΔG totale d'interaction est obtenue par la formule suivante : $\Delta G^0_b = \Delta G_{\text{liaison}} + \Delta G_{\text{déshydratation}} + \Delta G_{\text{rest}}$.

	Intégrase-DTG	ADN-DTG
$\Delta G_{\text{liaison}}$ (ligand –Intégrase)	-105,30 kcal/mol	-106,94 kcal/mol
$\Delta G_{\text{déshydratation}}$	96,70 kcal/mol	96,69 kcal/mol
ΔG_{rest}	7,20 kcal/mol	7,42 kcal/mol
ΔG totale d'interaction ΔG^0_b	-1,40 kcal/mol +/- 0.1	-2,83 kcal/mol +/- 0.1

Discussion des valeurs des énergies libres absolues obtenues pour les complexes Intégrase-DTG et ADN-DTG

- Par comparaison tout d'abord des valeurs de la première ligne du Tableau 10 correspondant aux énergies absolues de liaison du DTG à l'intégrase et à l'ADN séparément, nous constatons que $\Delta G_{\text{liaison}}$ du DTG avec l'intégrase (-105,3 kcal/mol) est moins stable que celle du DTG avec l'ADN (-106,94 kcal/mol). Ceci est le cas aussi pour les valeurs de ΔG^0_b où nous avons obtenu dans le cas du complexe Intégrase-DTG une valeur de -1,4 kcal/mol moins stable que l'énergie libre du complexe ADN-DTG, de -2,83 kcal/mol. Le DTG interagit donc d'une manière plus efficace avec l'ADN viral qu'avec l'intégrase ; hypothèse énoncée auparavant par notre équipe (El Khoury et al). En effet, par ses interactions proches et spécifiques avec les bases terminales virales C₁₆, A₁₇ et G₄ et par ses distances éloignées des résidus de l'intégrase susceptibles à se muter (Tyr143, Asn155, Gln148 et Gly140), le DTG effectue des interactions plus potentielles avec l'ADN viral. Ceci met en

valeur le rôle joué par le noyau halogéné qui est le plus impliqué dans les interactions spécifiques de la drogue avec l'ADN viral.

- Les valeurs de ΔG^0_b obtenues pour les complexes binaires (dernière ligne du Tableau 10) sont des valeurs fiables et correctes puisqu'elles sont supérieures à -15 kcal/mol (valeurs de -1,4 kcal/mol et -2,83 kcal/mol) avec un taux d'erreur de +/- 0,1 kcal/mol.
- Nous avons remarqué au cours des simulations des états intermédiaires du complexe Intégrase-DTG, un changement de la position du noyau halogéné qui au début des simulations était en bonne position avec son fluor en para en face du Glu152 de la triade catalytique de l'intégrase. Cependant, il tourne par la suite ce qui éloigne le fluor en para du résidu actif, diminuant les contacts proches entre les deux (Figure 119). Ceci explique en premier lieu, les distances élevées que nous avons observées après 50 ns de simulation, lors du suivi de la distance entre Glu152 et F en para du DTG dans le complexe Intégrase-DTG (Figure 66). Et en deuxième lieu, ceci éclaire le meilleur rendement énergétique du DTG en présence de l'ADN par rapport à celui en présence de l'intégrase.

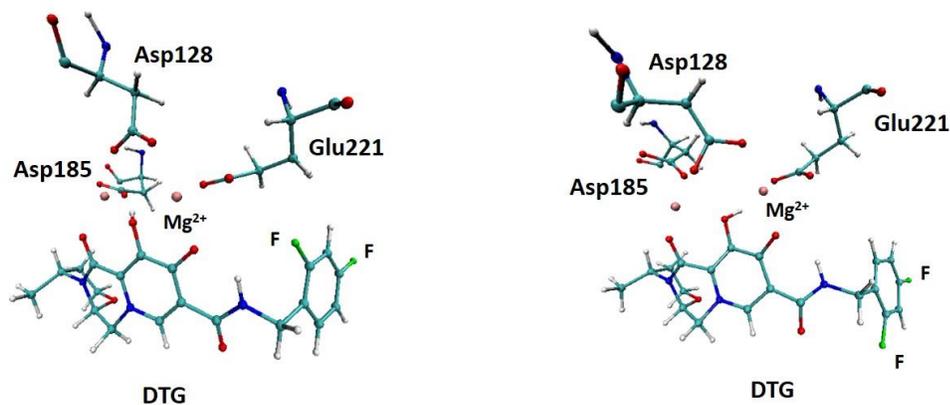


Figure 119 : Positions du Dolutégravir au niveau de son site de fixation dans l'intégrase du PFV en présence des ions Mg^{2+} . A gauche, nous trouvons la position de départ du DTG et à droite nous trouvons sa position après 50 ns de simulation. Le DTG, représenté en boules et en bâtons, est présent au niveau de la triade catalytique de l'intégrase, schématisée en boules et bâtons et représentée par Asp128, Asp185 et Glu221 correspondant respectivement à Asp64, Asp116 et Glu152 chez l'intégrase du VIH-1. Les deux ions Mg^{2+} chélatés par le DTG sont représentés par des sphères roses. Nous pouvons voir comment le noyau halogéné du DTG subit une rotation après 50 ns de simulation, de manière à éloigner ses substitutions de fluor du résidu catalytique Glu221.

- L'interaction du DTG à l'interface ADN-IN est plus stable que son interaction avec l'ADN ou l'IN seuls ; ΔG^0_b (ADN-Intégrase-DTG) = -109,02 kcal/mol) alors que ΔG^0_b (Intégrase-DTG) = -105,3 kcal/mol et ΔG^0_b (ADN-DTG) = -106,94 kcal/mol. Le DTG a donc un meilleur mode d'action et un meilleur profil inhibiteur en présence de ses deux partenaires qu'en présence de chacun d'eux seul. Nous pouvons donc confirmer les résultats des suivis des liaisons effectués pour nos simulations de dynamique ; le DTG nécessite la présence de l'intégrase en conformation active liée à l'ADN traité pour bien se situer à leur interface ce qui lui permet d'interagir efficacement avec leurs résidus clés et d'inhiber ainsi l'étape de transfert de brin.

C.IV. Conclusions du chapitre III partie C

Nous avons effectué des simulations de dynamique moléculaire non conventionnelles dans le but de calculer les énergies libres absolues alchimiques (ΔG^0_b) de l'interaction de nos ligands ; DTG, B3, A2, B2 avec l'ADN viral ainsi qu'avec l'ADN viral lié à l'intégrase. Nous avons de même simulé le complexe DTG-Intégrase afin de calculer l'énergie libre absolue de la liaison du DTG à l'IN pour pouvoir comparer son comportement en présence de la protéine avec celui en présence de l'ADN viral. Le dernier complexe simulé comprenait le DTG sans noyau d'halobenzène dans le but d'étudier le rôle de cette structure dans les interactions de la drogue avec ses partenaires. Nos résultats énergétiques obtenus sont conformes aux valeurs des énergies d'interaction intermoléculaire des noyaux halogénés avec les bases C₁₆ et G₄. En effet, d'après leurs valeurs de ΔG^0_b totales, B3, A2 et B2 ont présenté des interactions plus stables avec l'ADN viral par rapport au DTG ; il y a un accord entre les valeurs de ΔG^0_b totales et ΔE ce qui signifie que nos observations au niveau des mini-systèmes supportent celles au niveau des systèmes entiers. La meilleure efficacité de nos dérivés a été aussi observée au niveau du complexe ternaire où ils ont présenté une meilleure interaction avec le complexe ADN-Intégrase par rapport au DTG, ce qui confirme que nos nouvelles substitutions assurent leur efficacité plus importante. En absence de noyau halogéné, le DTG « contrôle » a affiché la valeur de ΔG^0_b la moins stable, soulignant le rôle que joue la structure halogénée à effectuer des interactions spécifiques avec l'ADN et l'IN. Finalement, en absence d'ADN, le DTG interagit moins efficacement avec l'Intégrase par comparaison avec son interaction avec l'ADN. Cependant, il nécessite la présence des deux pour entamer des contacts stables aboutissant à un mode d'inhibition plus efficace.

Références bibliographiques

- [1] J. S. Murray, P. Lane, and P. Politzer, "Expansion of the sigma-hole concept," *J. Mol. Model.*, vol. 15, no. 6, pp. 723–729, Jun. 2009, doi: 10.1007/s00894-008-0386-9.
- [2] E. F. Pettersen *et al.*, "UCSF Chimera—A visualization system for exploratory research and analysis," *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–1612, 2004, doi: 10.1002/jcc.20084.
- [3] K. El Hage, J.-P. Piquemal, Z. Hobaika, R. G. Maroun, and N. Gresh, "Could the 'Janus-like' properties of the halobenzene CX bond (X=Cl, Br) be leveraged to enhance molecular recognition?," *J. Comput. Chem.*, vol. 36, no. 4, pp. 210–221, Feb. 2015, doi: 10.1002/jcc.23786.
- [4] Y. Shao *et al.*, "Advances in molecular quantum chemistry contained in the Q-Chem 4 program package," *Mol. Phys.*, vol. 113, no. 2, pp. 184–215, Jan. 2015, doi: 10.1080/00268976.2014.952696.
- [5] K. El Hage, J.-P. Piquemal, Z. Hobaika, R. Maroun, and N. Gresh, "Approaching the double-faceted nature of the CX bond in halobenzenes with a bifunctional probe," *Chem. Phys. Lett.*, vol. 637, pp. 51–57, Aug. 2015, doi: 10.1016/j.cplett.2015.07.047.
- [6] K. El Hage, J.-P. Piquemal, Z. Hobaika, R. G. Maroun, and N. Gresh, "Substituent-Modulated Affinities of Halobenzene Derivatives to the HIV-1 Integrase Recognition Site. Analyses of the Interaction Energies by Parallel Quantum Chemical and Polarizable Molecular Mechanics," *J. Phys. Chem. A*, vol. 118, no. 41, pp. 9772–9782.
- [7] K. El Hage, J.-P. Piquemal, Z. Hobaika, R. Maroun, and N. Gresh, "Could an anisotropic molecular mechanics/dynamics potential account for sigma hole effects in the complexes of halogenated compounds?," *J. Comput. Chem.*, vol. 34, p. 1125, May 2013, doi: 10.1002/jcc.23242.
- [8] A. J. Misquitta and K. Szalewicz, "Intermolecular forces from asymptotically corrected density functional description of monomers," *Chem. Phys. Lett.*, vol. 357, no. 3, pp. 301–306, May 2002, doi: 10.1016/S0009-2614(02)00533-X.
- [9] R. Podszwa, R. Bukowski, and K. Szalewicz, "Density-Fitting Method in Symmetry-Adapted Perturbation Theory Based on Kohn–Sham Description of Monomers," *J. Chem. Theory Comput.*, vol. 2, no. 2, pp. 400–412, Mar. 2006, doi: 10.1021/ct050304h.
- [10] GaussView, Version 6, Dennington, Roy; Keith, Todd A.; Millam, John M. Semichem Inc., Shawnee Mission, KS, 2016.
- [11] Zargarian, L., Benleumi, M. S., Renisio, J.-G., Merad, H., Maroun, R. G., Wieber, F., et al. (2003). Strategy to discriminate between high and low affinity bindings of human immunodeficiency virus, type 1 integrase to viral DNA. *J. Biol. Chem.* 278, 19966–19973. doi:10.1074/jbc.M211711200.
- [12] S. Hare, S. S. Gupta, E. Valkov, A. Engelman, and P. Cherepanov, "Retroviral intasome assembly and inhibition of DNA strand transfer," *Nature*, vol. 464, no. 7286, pp. 232–236, Mar. 2010, doi: 10.1038/nature08784.
- [13] Z. Hobaika, L. Zargarian, Y. Boulard, R. G. Maroun, O. Mauffret, and S. Fermandjian, "Specificity of LTR DNA recognition by a peptide mimicking the HIV-1 integrase $\alpha 4$ helix," *Nucleic Acids Res.*, vol. 37, no. 22, pp. 7691–7700, Dec. 2009, doi: 10.1093/nar/gkp824.
- [14] R. G. Maroun *et al.*, "Conformational aspects of HIV-1 integrase inhibition by a peptide derived from the enzyme central domain and by antibodies raised against this peptide," *Eur. J. Biochem.*, vol. 260, no. 1, pp. 145–155, 1999, doi: 10.1046/j.1432-1327.1999.00130.x.
- [15] T. M. Jenkins, D. Esposito, A. Engelman, and R. Craigie, "Critical contacts between HIV-1 integrase and viral DNA identified by structure-based analysis and photo-crosslinking.," *EMBO J.*, vol. 16, no. 22, pp. 6849–6859, Nov. 1997, doi: 10.1093/emboj/16.22.6849.

- [16] S. Hare *et al.*, “Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572),” *Mol. Pharmacol.*, vol. 80, no. 4, pp. 565–572, Oct. 2011, doi: 10.1124/mol.111.073189.
- [17] A. S. Espeseth *et al.*, “HIV-1 integrase inhibitors that compete with the target DNA substrate define a unique strand transfer conformation for integrase,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 97, no. 21, pp. 11244–11249, Oct. 2000.
- [18] C. Marchand *et al.*, “Metal-dependent inhibition of HIV-1 integrase by beta-diketo acids and resistance of the soluble double-mutant (F185K/C280S),” *Mol. Pharmacol.*, vol. 64, no. 3, pp. 600–609, Sep. 2003, doi: 10.1124/mol.64.3.600.
- [19] J. M. Llibre and B. Clotet, “[Resistance profile and genetic barrier of dolutegravir],” *Enferm. Infecc. Microbiol. Clin.*, vol. 33 Suppl 1, pp. 20–25, Mar. 2015, doi: 10.1016/S0213-005X(15)30005-
- [20] L. E. Khoury *et al.*, “Spectrometric and computational studies of the binding of HIV-1 integrase inhibitors to viral DNA extremities,” *PeerJ Phys. Chem.*, vol. 1, p. e6, Dec. 2019, doi: 10.7717/peerj-pchem.6.
- [21] L. El Khoury, J.-P. Piquemal, S. Fermandjian, R. G. Maroun, N. Gresh, and Z. Hobaika, “The inhibition process of HIV-1 integrase by diketoacids molecules: Understanding the factors governing the better efficiency of dolutegravir,” *Biochem. Biophys. Res. Commun.*, vol. 488, no. 3, pp. 433–438, Jul. 2017, doi: 10.1016/j.bbrc.2017.05.001.
- [22] S. Hare, A. M. Vos, R. F. Clayton, J. W. Thuring, M. D. Cummings, and P. Cherepanov, “Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, no. 46, pp. 20057–20062, Nov. 2010, doi: 10.1073/pnas.1010246107.
- [23] S. Boresch, and F. Tettinger, M. Leitgeb, and M. Karplus, “Absolute Binding Free Energies: A Quantitative Approach for Their Calculation,” 2003, doi: 10.1021/JP0217839.
- [24] A. S. J. S. Mey *et al.*, “Best Practices for Alchemical Free Energy Calculations,” *Living J. Comput. Mol. Sci.*, vol. 2, no. 1, 2020, doi: 10.33011/livecoms.2.1.18378.
- [25] S. Boresch, and F. Tettinger, M. Leitgeb, and M. Karplus, “Absolute Binding Free Energies: A Quantitative Approach for Their Calculation,” 2003, doi: 10.1021/JP0217839.

Chapitre V

Conclusions et Perspectives

V.I. Conclusions des travaux de thèse

L'intégrase du virus VIH-1 qui est l'agent causal de la maladie du SIDA, est une enzyme, responsable de l'intégration de l'ADN viral au niveau de l'ADN cellulaire en deux étapes ; le 3' processing suivi du transfert de brin, afin d'initier la réplication du virus. Etant primordiale au cycle du VIH-1 et n'ayant pas de contrepartie cellulaire, elle constitue une excellente cible pour le développement de médicaments antirétroviraux. Trois molécules principales anti-intégrase ont été mises sur le marché depuis 2007, ciblant l'étape de transfert de brin orchestrée par l'IN. Ces molécules sont le RAL, l'EVG et le DTG qui agissent à l'interface ADN-IN afin d'inhiber le transfert de l'ADN viral après son traitement au niveau 3'. Le DTG est l'inhibiteur le plus potentiel parmi ses antécédents. Cependant, malgré leur succès initial, l'intégrase a développé des mutations de résistance primaires et secondaires contre ces drogues, diminuant ainsi leur efficacité chez les patients traités. Pour cela, nous avons visé durant ces trois années de thèse, à concevoir de nouveaux inhibiteurs de transfert de brin ayant des affinités augmentées pour l'intasome du VIH-1, en se focalisant notamment à améliorer leur affinité envers l'ADN viral ce qui leur permettrait de contourner l'obstacle des mutations de résistance de l'IN. Notre stratégie d'optimisation s'est basée sur la structure du DTG comme référence, où nous avons effectué des halogénations sélectives et des substitutions appropriées de groupements électrodonneurs et électroattracteurs, au niveau de son noyau halogéné (noyau de benzène avec deux fluors) puisque ce dernier est la structure du DTG qui entre en contact direct avec les bases virales réactives situées à l'extrémité 3' de l'ADN viral. Ces bases sont la Cytosine C₁₆ qui effectue des empilements π - π avec le noyau halogéné du DTG et la Guanine G₄ qui par le biais de ses noyaux riches en électrons, fait des interactions électrostatiques

avec les substitutions du noyau d'halobenzène de la drogue. Nous avons ainsi conçu treize nouveaux noyaux halogénés substitués que nous avons évalués par la suite afin de voir si nos substitutions ont contribué à un meilleur rendement thérapeutique de nos dérivés. Tout d'abord, nous avons adopté un modèle d'étude simplifié formé par le noyau de C₁₆, le noyau de G₄ et nos noyaux halogénés pour lequel nous avons calculé les énergies d'interaction intermoléculaire (ΔE) via des approches de mécanique quantique (méthodes DFT et ALMOEDA) ainsi que des approches de mécanique moléculaire (méthode basée sur le champ de force polarisable SIBFA). Les résultats énergétiques obtenus ont montré une meilleure interaction de la totalité de nos nouveaux dérivés avec les deux bases virales par rapport au DTG, avec un accord entre les méthodes quantiques et moléculaires. Une décomposition des ces énergies intermoléculaires via les méthodes RVS et SIBFA ont mis en relief la contribution des énergies du premier et du second ordre dans ces interactions, avec un accord entre nos deux approches. En effet, la méthode SIBFA a pu reproduire avec précision les valeurs et les évolutions des contributions quantiques individuelles ce qui confirme la fiabilité de nos résultats. Une exploration des contours de potentiel électrostatique de nos dérivés a permis l'exploration des effets électroniques de nos substitutions ; nos groupements électrodonneurs en position para des noyaux, pointant vers G₄, ont favorisé leurs interactions électrostatiques avec cette base alors que les halogènes électroattracteurs substitués ont amélioré l'empilement π - π avec C₁₆. L'étape suivante a été de construire l'intasome entier du PFV (homologue à l'intasome du VIH-1) en présence du DTG et des cations de Magnésium ainsi que d'intégrer dans cette structure complète, à la place du DTG, nos meilleurs noyaux inhibiteurs présélectionnés à partir des calculs de ΔE . Nous avons choisi donc, autre que le

DTG, de construire l'intasome viral en présence de B3, A2 et B2 respectivement (nos trois inhibiteurs les plus efficaces choisis) ainsi qu'en présence du DTG sans noyau d'halobenzène. Trois systèmes supplémentaires ont été construits ; les systèmes ADN-Intégrase, ADN-DTG et Intégrase-DTG. Cette démarche est une première où nos complexes construits et solvatés dans des boîtes d'eau ont été simulés ensuite par dynamique moléculaire conventionnelle. Le suivi des liaisons et des interactions au niveau des trajectoires obtenues a montré de meilleures interactions de nos inhibiteurs avec les bases virales primordiales C₁₆, G₄ et A₁₇ par rapport au DTG, confirmant les résultats de ΔE et soulignant le rôle du noyau halogéné à interagir spécifiquement avec l'ADN viral. Par leurs interactions spécifiques avec ces bases virales, nos inhibiteurs ainsi que le DTG ont perturbé un réseau d'interactions établies entre l'IN et l'ADN, notamment au niveau des résidus importants tels que Glu152, Gly149, Lys156, Lys159, Lys160 et Gln146. De plus, nos composés ont effectué des contacts proches avec des résidus clés de l'intégrase notamment Glu152 de la triade catalytique, Gly118 et Pro145, prouvant leur mode d'action à l'interface ADN-IN et leur inhibition de l'étape de transfert de brin, puisqu'en absence de l'ADN ou de la protéine, les complexes simulés n'étaient plus stables. De plus, la meilleure efficacité de nos dérivés et notamment celle de DTG par rapport à RAL et EVG a été confirmée par l'absence de contacts proches entre eux avec les résidus subissant des mutations de résistance primaire et secondaire. Finalement, nous avons effectué sur ces complexes simulés, des calculs des énergies libres absolues alchimiques des interactions de nos ligands avec l'ADN entier et avec l'intasome viral entier. Nos calculs ont présenté un parfait accord avec nos résultats quantitatifs et qualitatifs précédents, indiquant des interactions plus spécifiques de B3, A2, et B2 avec l'ADN viral et avec l'ADN lié à l'intégrase par

rapport au DTG. Ce dernier a été aussi caractérisé par une meilleure spécificité envers l'ADN viral qu'envers l'intégrase entière, confirmant davantage que les contacts spécifiques de l'inhibiteur avec l'ADN viral via son noyau halogéné contribuent à sa meilleure efficacité et à un mode d'action plus efficace pouvant surmonter les mutations de résistance. L'ensemble de nos résultats prouvent donc que des halogénations et substitutions sélectives du noyau halogéné du DTG peuvent renforcer d'une manière significative les affinités des ligands pour leurs cibles situées notamment à l'extrémité de l'ADN viral traité. Nos démarches ont donc donné lieu à la mise au point de nouvelles drogues inhibitrices potentiellement plus efficaces que celles mises sur le marché, ayant la capacité de surmonter l'obstacle de la résistance virale et assurer un mode d'action plus efficace.

V.II. Perspectives des travaux de thèse

La suite de ces travaux se basera sur les points suivants :

- Synthétiser nos meilleures dérivés conçus, les composés B3, A2 et B2, ce qui nous permettra de les tester par la suite *in vitro* et *in vivo*. Nous avons déjà commencé à collaborer avec l'Institut de Biologie-Paris-Seine à Sorbonne Université pour la synthèse de nos trois dérivés.
- Passer à l'application expérimentale afin de tester l'efficacité des trois meilleures composées, déjà démontrées *in silico*. Ceci peut se faire par des expériences d'anisotropie de fluorescence mettant en jeu la structure entière de chaque inhibiteur avec une série de nucléotides mimant l'extrémité 3' du brin d'ADN viral traité du VIH-1. Ceci nous permettra de voir si nos composés présentent aussi *in vitro* de meilleures interactions avec l'ADN par rapport au DTG dont les interactions par anisotropie de

fluorescences avec l'ADN viral ont été déjà étudiées par El Khoury et al au sein de notre équipe ^[19].

- Simuler par dynamique moléculaire, les complexes B3-Intégrase, A2-Intégrase et B2-Intégrase et calculer leurs énergies libres absolues d'interaction afin de voir en premier lieu s'ils interagissent plus spécifiquement avec la protéine virale par rapport au DTG en présence de l'intégrase puis, en deuxième lieu, comparer les valeurs énergétiques obtenues aux énergies libres d'interaction des inhibiteurs avec l'ADN viral dans le but de voir si, comme le DTG, nos nouvelles substitutions ont gardé une meilleure interaction avec l'ADNv qu'avec l'intégrase.
- Passer à l'évaluation du rôle des cations de Mg^{2+} en étudiant via des approches de dynamique moléculaire (simulations conventionnelles, calculs de ΔG ...) l'interaction de ces cations avec la fonction dicétone responsable de leur chélation par les inhibiteurs. En effet, les inhibiteurs de transfert de brin entrent en compétition avec les résidus de la triade catalytique afin de chélater les ions Mg^{2+} nécessaires à l'étape d'intégration. Nous pourrions alors comparer le comportement de ces cations en absence et en présence des inhibiteurs afin de voir l'influence que possèdent ces derniers sur la fonction des ions de Magnésium.

Chapitre VI

ANNEXES



Quantum-Chemistry Based Design of Halobenzene Derivatives With Augmented Affinities for the HIV-1 Viral G₄/C₁₆ Base-Pair

Perla El Darazi^{1,2}, Léa El Khoury^{1,2†}, Krystal El Hage³, Richard G. Maroun², Zeina Hobaika², Jean-Philip Piquemal^{1,4,5*} and Nohad Gresh^{1*}

OPEN ACCESS

Edited by:

Nino Russo,
University of Calabria, Italy

Reviewed by:

Tiziana Marino,
University of Calabria, Italy
Pedro Alexandrino Fernandes,
University of Porto, Portugal

*Correspondence:

Jean-Philip Piquemal
jean-philip.piquemal@
sorbonne-universite.fr
Nohad Gresh
nohad.gresh@ict.jussieu.fr

† Present address:

Léa El Khoury,
Department of Pharmaceutical
Sciences, University of California,
Irvine, Irvine, CA, United States

Specialty section:

This article was submitted to
Theoretical and Computational
Chemistry,
a section of the journal
Frontiers in Chemistry

Received: 25 November 2019

Accepted: 27 April 2020

Published: 19 June 2020

Citation:

El Darazi P, El Khoury L, El Hage K,
Maroun RG, Hobaika Z, Piquemal J-P
and Gresh N (2020)
Quantum-Chemistry Based Design of
Halobenzene Derivatives With
Augmented Affinities for the HIV-1
Viral G₄/C₁₆ Base-Pair.
Front. Chem. 8:440.
doi: 10.3389/fchem.2020.00440

¹ Sorbonne Université, Laboratoire de Chimie Théorique, UMR7616 CNRS, Paris, France, ² UR EGP, Centre d'Analyses et de Recherche, Faculté des Sciences, Université Saint-Joseph de Beyrouth, Beirut, Lebanon, ³ SABNP, Univ. Evry, INSERM U1204, Université Paris-Saclay, Evry, France, ⁴ Institut Universitaire de France, Paris, France, ⁵ Department of Biomedical Engineering, The University of Texas at Austin, Austin, TX, United States

The HIV-1 integrase (IN) is a major target for the design of novel anti-HIV inhibitors. Among these, three inhibitors which embody a halobenzene ring derivative (HR) in their structures are presently used in clinics. High-resolution X-ray crystallography of the complexes of the IN-viral DNA transient complex bound to each of the three inhibitors showed in all cases the HR ring to interact within a confined zone of the viral DNA, limited to the highly conserved 5' CpA 3'/5' TpG 3' step. The extension of its extracyclic CX bond is electron-depleted, owing to the existence of the "sigma-hole." It interacts favorably with the electron-rich rings of base G₄. We have sought to increase the affinity of HR derivatives for the G₄/C₁₆ base pair. We thus designed thirteen novel derivatives and computed their Quantum Chemistry (QC) intermolecular interaction energies (ΔE) with this base-pair. Most compounds had ΔE values significantly more favorable than those of the HR of the most potent halobenzene drug presently used in clinics, Dolutegravir. This should enable the improvement in a modular piece-wise fashion, the affinities of halogenated inhibitors for viral DNA (vDNA). In view of large scale polarizable molecular dynamics simulations on the entirety of the IN-vDNA-inhibitor complexes, validations of the SIBFA polarizable method are also reported, in which the evolution of each ΔE (SIBFA) contribution is compared to its QC counterpart along this series of derivatives.

Keywords: HIV-1 integrase, quantum chemistry, polarizable force fields, viral DNA, Dolutegravir, INSTIs

INTRODUCTION

The HIV-1 integrase (IN) catalyzes the transfer of a viral DNA (vDNA) strand into the genome of the host cell (Lesbats et al., 2016). It is also involved in reverse transcription (Hironori et al., 2009), nuclear import (Mouscadet et al., 2007), and HIV-1 particle maturation (Kessl et al., 2016). It has no counterpart in human cells and thus constitutes an emerging target for the design of novel anti-retroviral inhibitors (Liao et al., 2010).

Three integrase inhibitors have been approved by the FDA in anti-HIV therapies, Raltegravir (RAL) (Summa et al., 2008), Elvitegravir (EVG) (Shimura et al., 2008), and Dolutegravir (DTG) (Underwood et al., 2012). All three act as integrase strand transfer inhibitors (INSTIs)

(Ammar et al., 2016). They embody two distinct structural motives, namely a large diketo acid pharmacophore, and a halobenzene derivative. A major advance toward the design of novel derivatives is enabled by high-resolution X-ray crystal structures of the ternary complexes of IN, vDNA, and each of the three INSTI's (Hare et al., 2010a, 2011). These show all three drug complexes to be two-pronged. The keto oxygen and a coplanar neighboring oxygen both coordinate two IN catalytic Mg (II) cations, structural water molecules, and, either directly or through water, IN residues. The halobenzyl moiety is confined in a narrow cleft, binding to the G₄ and C₁₆ bases of the highly conserved 5' CpA 3'/5' TpG 3' step on the viral DNA ends (Hare et al., 2010b). The emergence of IN mutations weakening IN-drug interactions is a major limitation for INSTI-based therapies (Wainberg et al., 2011). Such mutations were reported to occur solely on IN and not on vDNA: this thus leaves open the possibility that additional enhancements of INSTI-vDNA binding should not be adversely impacted by IN mutations.

The present study was motivated by two findings. First, reports from one of our Laboratories showed the DTG > EVG > RAL ranking of affinities for the IN-vDNA complex (denoted as the intasome, INT) to be paralleled by their corresponding affinities for the sole vDNA (El Khoury et al., 2017). Could, thus, increases of the INSTI-INT binding affinities be attempted upon focusing on the sole "ternary" complexes of G₄, C₁₆, and a halobenzene ring? Second, recent spectrometric and computational studies of the binding of INSTIs to viral DNA extremities showed the ranking of affinities to be governed by the enthalpy (ΔH) component of the binding free energies (ΔG), the entropy component (TAS) being similar for all three complexes (El Khoury et al., 2019), a "signature" for intercalative binding (Chaires, 2006). Furthermore, the ΔH ranking of the three inhibitor affinities was itself paralleled by the corresponding ranking of the *ab initio* quantum chemistry (QC) intermolecular interaction energies, $\Delta E(QC)$, of their halobenzene rings with the *sole* G₄/C₁₆ base pair. Upon focusing on the halobenzene ring of the best bound compound, namely DTG, could, then, $\Delta E(QC)$ Energy Decomposition Analysis (EDA) along with electronic structure considerations offer insight for affinity-enhancing chemical substitutions?

The full structures of RAL, EVG and DTG are presented in **Figure 1**. There is a zone of electron depletion along the extension of the C-F bond of DTG, denoted as the "sigma-hole" (Murray et al., 2008). This bond is *para* to the C-C bond connecting the ring to the central diketo acid group (**Figure 1**). In the crystal structure of the DTG-INT complex, it points toward the electron-rich bicyclic ring of G₄. This could constitute a key stabilizing feature of the DTG-G₄ complex.

We have considered and analyzed several analogs of the DTG halobenzene ring (HR), 13 of which will be reported in this paper. It is noted that with the exception of compound A1, all compounds have an extracyclic -NHCH₃ or -NH₂ proton donor replacing in *para* the extracyclic CF bond. All additional substitutions were done *ortho* or *para* to the C-C connector by electron-withdrawing groups. Each substituent can impact $\Delta E(QC)$ by a combination of different factors:

- on account of its electron-withdrawing character, a favorable increase of the electrostatic contribution, E_C , of $\Delta E(QC)$, of the *para* substituent with G₄;
- a favorable increase of the polarization contribution, E_{pol} , of $\Delta E(QC)$, due to the contribution of its polarizability to the polarization energy of the ligand;
- when in *meta*, it could contribute to additional electrostatic interactions with sites belonging to G₄ or C₁₆, as reflected by E_C and E_{pol} . There is a cone of electron-rich density around the CX bonds of halogens. Halogen substituents could enhance E_C if such a cone was in the vicinity of electron-deficient sites of C₁₆.
- these three factors could be counteracted in some cases by increases of the short-range repulsion, E_X , and, in the energy balances, by a larger solvation penalty than DTG, since most designed derivatives bear a more polar character than DTG.

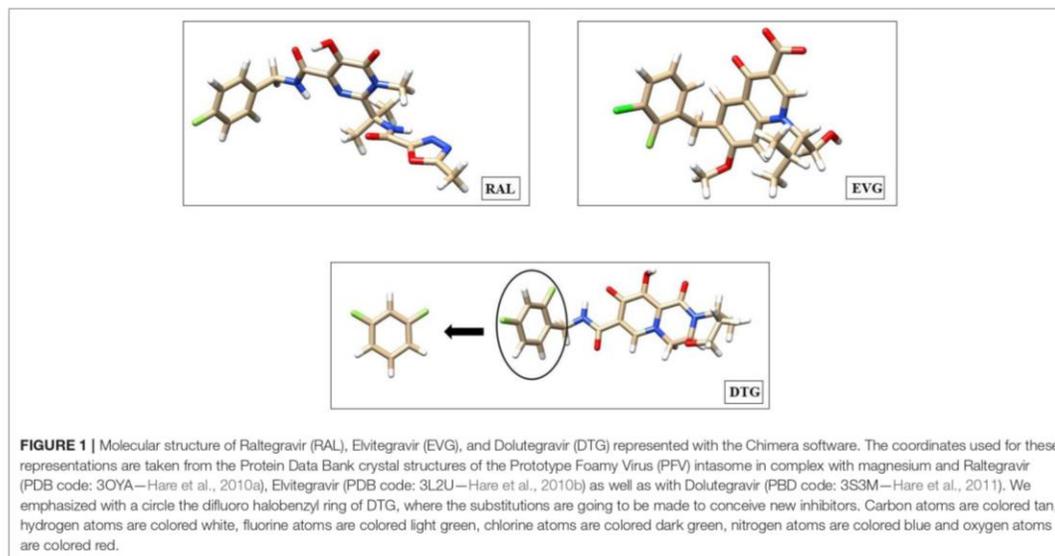
EDA unravels the relative magnitudes of the individual $\Delta E(QC)$ contributions along the series investigated, and how each contribution can be impacted by each substituent. Optimizing the localized HR-G₄-C₁₆ interactions could contribute to a modular design of INSTIs as a preliminary to long-duration molecular dynamics (MD) of the entirety of the INSTI with the entirety of the INT. The binding site includes highly polar protein residues, two Mg (II) cations, and structural, highly polarizable water molecules. Polarizable, multipolar Molecular Mechanics/Dynamics approaches, such as SIBFA (Gresh et al., 2007) or AMOEBA (Ponder et al., 2010) should be adapted to these simulations. Such procedures were demonstrated to reliably account for the impact of the sigma hole and the dual character of the CX bond of HR on electrostatics (El Hage et al., 2013). Very large macromolecular complexes, such as the INT-INSTI ones, are now amenable to long-duration MD upon resorting to the massively parallel code Tinker-HP (Lagardère et al., 2018). Prior to these, and in the context of the present study, it was thus essential to evaluate the accuracy of one of these procedures, SIBFA. Specifically, how well will the evolutions of $\Delta E(QC)$ and of each of its contributions, be paralleled by the SIBFA contributions along the series of the thirteen HR ligands?

As a complement to EDAs, we will also report the contours of their electrostatic potential maps.

METHODS

Quantum Chemistry Calculations Energy Decomposition Analysis

The decomposition of the *ab initio* SCF interaction energy is done using the Reduced Variational Space (RVS) analysis (Stevens and Fink, 1987), where the intermolecular interaction energy is separated into four contributions: Coulomb (E_C) and short-range exchange-repulsion (E_X) in first order (E_1) and polarization (E_{pol}) and charge-transfer (E_{ct}) in second order (E_2). Finally, the dispersion contribution is assessed as the difference between the BSSE-corrected B97-D3 intermolecular interaction energies and the Hartree-Fock (HF) ones. The basis set superposition error (BSSE) is taken into consideration in the final energy values. The GAMESS software with the cc-pVTZ (-f) basis set (Schmidt



et al., 1993) were used in this analysis. We denote below as E_{pol} (VR), a “variational” value of E_{pol} , obtained as the difference between $\Delta E(\text{RVS})$ and the sum of E_1 and E_{ct} . This is to be contrasted to $E_{\text{pol}}(\text{RVS})$ at the outcome of the RVS procedure, the sum of all individual ligand polarization energies computed in a process when the occupied molecular orbitals (MO) of this ligand is relaxed toward its own virtual MO’s, the other ligands being frozen.

Energy decomposition analyses were also performed at the correlated levels and the ω -B97-D3 functional (Grimme, 2006) resorting to the Absolutely Localized Molecular Orbitals method (ALMOEDA) (Azar and Head-Gordon, 2011) using the Q-Chem software (Shao et al., 2015). ΔE is decomposed into a “frozen density” component (FRZ), namely the sum of the Coulomb and short-range contribution, and a polarization (E_{pol}) and a charge transfer contribution (E_{ct}) in second order.

Correlated Calculations

The intermolecular interaction energies (ΔE) of the complexes formed by the substituted rings with G₄ and C₁₆ were computed at the correlated level using the dispersion-corrected functionals B97-D3 and B3LYP-D3 (Goerigk and Grimme, 2011), with the cc-pVTZ basis set (-f) (Feller, 1996) and the Gaussian software G09 (Frisch et al., 2009) software. The obtained values were corrected for BSSE (Simon and Duran, 1996). The values of the corresponding correlation and dispersion contributions were computed as the differences between the dispersion-corrected correlated ΔE and uncorrelated Hartree-Fock (HF) ΔE values.

G09 was used as well for energy-minimization of the complexes, using a starting structure determined by X-ray crystallography and taken from Protein Data Base site (PDB code: 3S3M, Hare et al., 2011). The C-C bond connecting the HR to

the diketoacid (DKA) ring was replaced by a CH bond, since the DKA ring was removed.

The halobenzene rings were relaxed except for the hydrogen atom of this CH bond. This choice was made to avoid wanderings over the G₄/C₁₆ bases prevented in the complete complexes by HR anchoring to the rest of the drug. We also chose to hold both guanine and cytosine frozen in their X-ray geometry to account for their anchoring in the DNA backbone. We did not allow for conformational relaxation around the glycosidic bonds of G₄ and C₁₆ since we considered that this viral DNA base pair is held in an experimental position “tailored” for DTG: any additional positional relaxation in complexes with derivatives with bulkier ligands could *a priori* be expected to further optimize, rather than penalize, the binding of such derivatives, with possibly an even more favorable outcome than the one from the present study. This could clearly, only be clarified at the outcome of long-duration MD simulations on the ternary DNA-IN-ligand complexes, enabling to explore the widths and flexibility of the energy basins in the G₄/C₁₆/HR zone. Such studies will be reported in due course.

A Continuum desolvation energy ΔG_{solv} , computed following the Polarizable Continuum Model (PCM) procedure (Mennucci et al., 2002) was also calculated for each HR ring. It was considered as an upper bound to its actual PCM desolvation energy prior to its complexation.

SIBFA Computations

In the context of the SIBFA procedure, the intermolecular interaction energy (ΔE_{tot}) is computed as the sum of five contributions: electrostatic multipolar (E_{MTP}), short-range repulsion (E_{rep}), polarization (E_{pol}), charge transfer (E_{CT}), and dispersion (E_{disp}). E_{MTP} is computed with distributed multipoles

(up to quadrupoles) derived from the QC molecular orbitals precomputed for each individual molecule, derived from the Stone analysis (Stone and Alderton, 1985) and distributed on the atoms using a procedure developed by Vigné-Maeder and Claverie (1988). E_{MTP} is augmented with a penetration term (Piquemal et al., 2007). The anisotropic polarizabilities are distributed on the centroids of the localized orbitals (heteroatom lone pairs and bond barycenters) using a procedure due to Garmer and Stevens (1989). E_{rep} and E_{CT} , the two short-range contributions, are computed using representations of the molecular orbitals localized on the chemical bonds and on localized lone-pairs. E_{disp} is computed as an expansion into $1/R^6$, $1/R^8$, and $1/R^{10}$ and embodies an explicit exchange–dispersion term (Creuzet et al., 1991).

Contours of Electrostatic Potentials

The contours of molecular electrostatic potentials (MEPs) of the HRs derived from their electronic densities were displayed by the Gaussview software, implemented in the Gaussian software (Frisch et al., 2009). The colored zones are based on their electronic densities.

RESULTS AND DISCUSSION

All DTG derivatives have at least one extracyclic halogen atom; F, Cl, or Br. The DTG coordinates are derived from the crystal structure of the viral intasome with Dolutegravir and two magnesium ions (PDB code: 3S3M, Hare et al., 2011). The first derivative considered, A1, has a chlorine substituent replacing the second DTG fluorine, which is *para* to the CH bond connecting it to the diketo moiety. This did not bring a significant ΔE_{tot} increase with respect to DTG (see Table 1 below). Several alternative substituents could be considered. However, for the present study we deemed it more instructive to replace upfront the *para*-F group, which binds G_4 by its sigma-hole prolonging the CF bond, by another electron-deficient group, namely a proton donor. We could thus leverage these C-X bonds (X halogen or electron-deficient group) interactions and increase their magnitudes by depleting this donor electron-density with selected electron-attracting substituents. A natural choice bore on substituents such as $-NHCH_3$ or $-NH_2$. Both groups have the additional advantage of acting as electron-donating substituents: this is in contrast to fluorine, which is electron-withdrawing. This could favor the interaction of the electron-deficient sites of C_{16} with the halobenzene ring and/or with the electron-rich cone around the halogen substituent. It is another means of leveraging the “Janus-like” character of the halobenzene derivatives (El Hage et al., 2014, 2015). The $-NHCH_3$ substituted derivatives were split into two groups, having two or one electron-withdrawing substituents, respectively. The search for $-NH_2$ substituted derivatives was limited to three derivatives, with one or two fluorine substituents. The four groups are indexed as Group 1, 2, 3, and 4, and the derivatives in each group are indexed with capital letters from A to E. All derivatives are represented in Figure 2.

First Group

Para-Electron Attracting Substituted Derivatives

This group contains the compound A1 which has a similar structure to DTG, but with a chlorine atom in *para* instead of a fluorine one.

Second Group

Para- $NHCH_3$ Substituted Derivatives

Five derivatives are considered. They all have, similar to DTG, fluorine in *ortho*, and a third, electron-withdrawing substituent:

A2, B2, and C2 have in the second *ortho* position: Br, Cl, and $-CF_3$, respectively. D2 has chlorine in *meta*, and E2 has a cyano substituent in the second *ortho* position.

Third Group

Para- $NHCH_3$ Substituted Derivatives

A3, B3, and C3 have as *ortho* substituent chlorine, bromine and fluorine, respectively, while D3 has fluorine in *meta* instead of *ortho*.

Fourth Group

Para- NH_2 Substituted Derivatives

A4 and B4 have a fluorine in *ortho* and in *meta*, respectively, while C4 has the two fluorines in *ortho* and *meta*.

Figure 3 gives a representation of the complexes of the G_4/C_{16} vDNA base pair with representative compounds of series 1 to 4. We have displayed the parent compound, DTG from the first group, C2 and E2 from the second, B3 from the third and C4 from the fourth group. As seen below, the inhibitory rings face the ring of C_{16} , interacting with the latter via π - π stacking, while the moiety in *para* position face the electron-rich rings of G_4 . Therefore, the electron donor groups in C2, E2, B3, and C4 as well as the electron-depleted region in the prolongation of the sigma-hole in DTG interacts with G_4 via electrostatic interactions.

Table 1 lists the ΔE (QC) values at the B97-D3, B3LYP-D3, and ω -B97D levels together with the ΔE (SIBFA) values. It also lists the continuum solvation energies, ΔG_{solv} , of each isolated derivative prior to complexation, and the resulting values of partial B97-D3 energy balances after ΔG_{solv} subtraction from ΔE (B97-D3). These ΔG_{solv} values should be considered as upper bounds to the actual PCM solvation energies of the halobenzene ring of DTG and its derivatives, because in the “complete” inhibitor, the exposure of the polar group is less on account of a lesser accessibility and possible intramolecular interactions. The MD calculation of the actual solvation free energies of DTG and its derivatives in a box of water molecules will be considered in the next step of our studies in parallel with the MD simulations of their complexes with the intasome (El Darazi et al., work in progress).

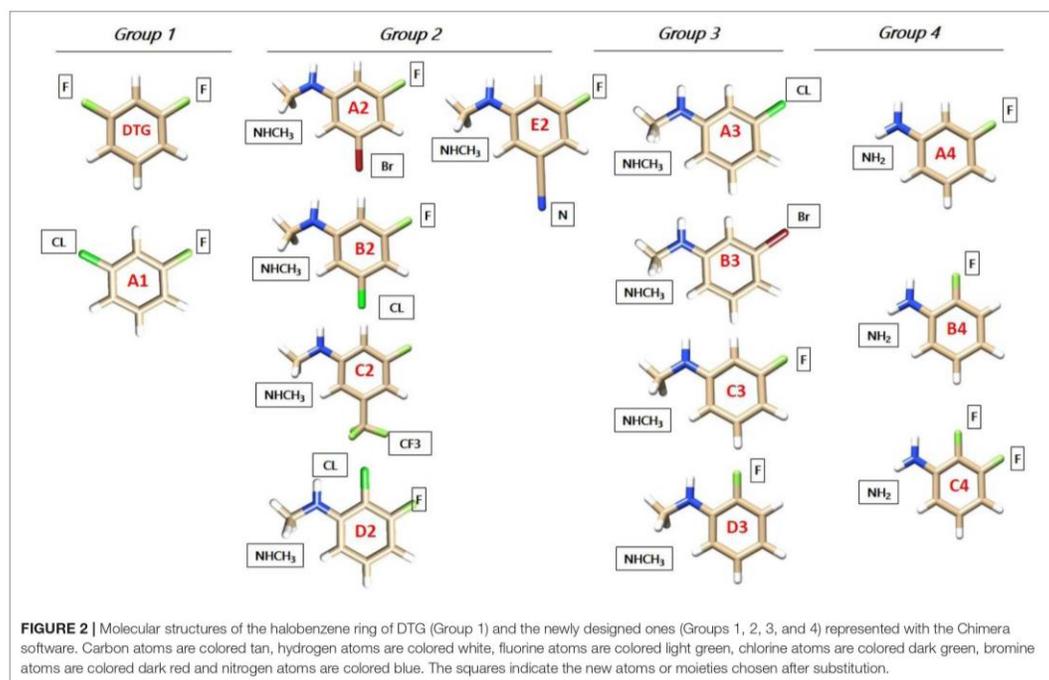
The listed QC ΔE values take into account the BSSE correction as the HF and at the correlated levels. It has values in the 1.5–2 kcal/mol range at the HF level, and 3.5–4.6 kcal/mol range at the correlated levels, depending upon the functional, but changing little for a given functional with the ligand.

The evolution of ΔE as a function of the compound number is plotted in Figure 3 for each functional.

TABLE 1 | Intermolecular interaction energies (kcal/mol) of halogenated derivatives with the G₄/C₁₆ vDNA base pair.

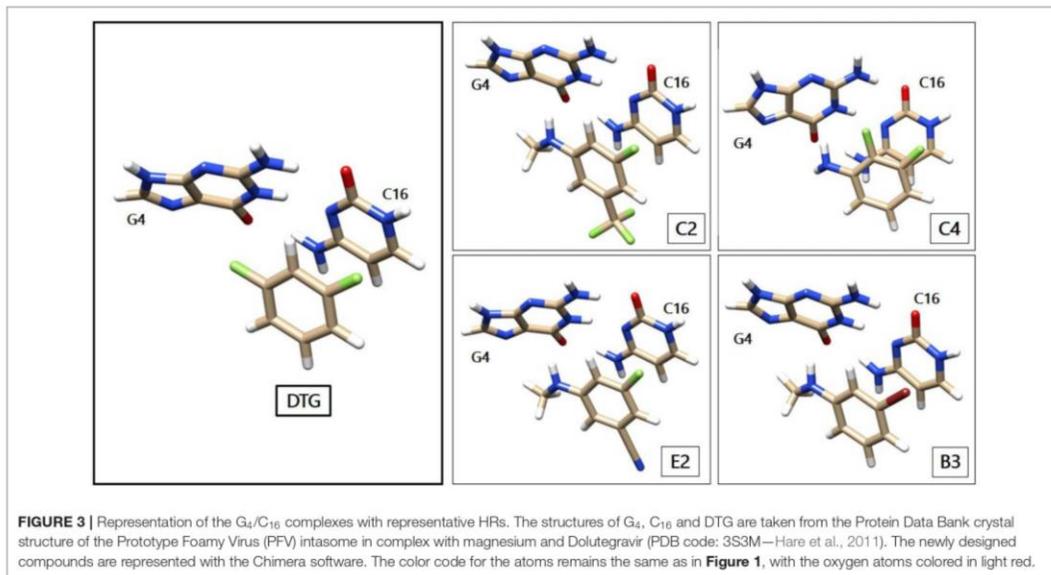
Compounds	ΔE B97-D3 (raw value/BSSE corrected)	ΔE B3LYP-D3 (raw value/BSSE corrected)	ΔE ALMOEDA (B97D3)	ΔE ALMOEDA (ω -B97D)	SIBFA	ΔG solv	Total interaction energy (ΔE B97-D3 - ΔG solv) (raw value/BSSE corrected)
DTG	-37.4/-33.2	-40.6/-36.5	-33.2	-35.9	-39.6	-6.1	-31.3/-27.1
A1	-37.5/-33.8	-40.6/-37.0	-33.8	-35.8	-38.4	-5.9	-31.6/-27.9
A2	-45.1/-40.7	-48.1/-43.7	-40.6	-43.3	-45.9	-9.7	-35.4/-31.0
B2	-44.9/-40.5	-47.9/-43.5	-40.4	-43.1	-46.6	-9.5	-35.4/-31.0
C2	-46.0/-41.3	-49.1/-44.5	-41.2	-44.1	-47.3	-10.1	-35.9/-31.2
D2	-43.0/-38.6	-46.3/-41.9	-38.5	-41.0	-45.2	-7.8	-35.2/-30.8
E2	-45.3/-40.9	-48.2/-43.9	-40.6	-43.4	-47.7	-7.8	-37.5/-33.1
A3	-44.7/-40.4	-47.6/-43.4	-40.3	-42.7	-44.6	-9.4	-35.3/-31.0
B3	-44.9/-40.7	-47.9/-43.7	-40.6	-43.1	-46.9	-9.5	-35.4/-31.2
C3	-44.1/-39.6	-47.0/-42.6	-38.8	-41.5	-45.2	-9.7	-34.3/-29.9
D3	-42.6/-37.8	-45.7/-41.1	-38.4	-41.2	-44.0	-7.8	-34.8/-30.0
A4	-42.5/-38.2	-45.2/-40.9	-38.0	-40.6	-43.7	-11.2	-31.3/-27.0
B4	-41.5/-37.0	-44.2/-39.7	-37.4	-39.7	-41.1	-9.7	-31.8/-27.3
C4	-42.1/-37.5	-45.0/-40.4	-41.2/	-41.0	-42.6	-10.3	-31.9/-27.2

DTG as well as the compounds displaying the best energetic values are highlighted in red.



All three ΔE (QC) curves run parallel. **Table 1** and **Figure 4** show that the ΔE (QC) values of all compounds in series 2–4 have significantly larger magnitudes than those of either DTG or its chlorinated derivative A1. Taking the B97-D3 results as an example, the differences are in the 4–8.3 kcal/mol range. All

three QC procedures concur into having five derivatives which stand out from the rest. Four belong to series 2, namely A2, B2, C2, and E2, and one belongs to series 3, namely B3. In series 2, all four derivatives have their two substituents on both *ortho* sides of the connecting -CH bond, or, equivalently, on



both sides *meta* to the $-\text{NHCH}_3$ para substituent. The two best-bound compounds are C2 and E2, with either a $-\text{CF}_3$ or a cyano substituent $-\text{CN}$ in *meta* to the $-\text{NHCH}_3$ group, respectively. C2 has a larger ΔG_{solv} value than E2, namely -10.1 compared to -7.8 kcal/mol. As mentioned above, this 2.3 kcal/mol difference should represent an upper bound to the actual desolvation energy difference between the two derivatives. It could reflect a greater “hydrophilicity” of the C2 than the E2 ring. Including such a difference in partial relative energy balances would give rise to a 2 kcal/mol preference for the cyano derivative compared to the trifluoromethyl one.

In series 3, the best-bound derivative is B3, with a bromine substituent. Owing to the large hydrophilic character of this atom compared to fluorine in DTG, the values of ΔG_{solv} are correspondingly larger in series 2 and 3 than that of DTG, in the range 2.8–4.0 kcal/mol. This leaves out relative energy preferences for A2–C2, E2, and B3 over DTG in the range 4–6.7 kcal/mol.

The ΔE values in series 4 with an $-\text{NH}_2$ para substituent have smaller magnitudes than in series 2 and 3. These are 4–5 kcal/mol more favorable than DTG, but these are virtually fully compensated for by correspondingly larger ΔG_{solv} values, leaving partial energy balances <0.5 kcal/mol more favorable than DTG, which is inconclusive in terms of augmented affinities.

As for the BSSE values, we wish to note that the average values of the BSSE amounts to 4.14 kcal/mol for the B97D-3 functional and to 3.41 kcal/mol for the W97D functional, both used for the ALMOEDA energetic calculations. With each, the relative variations among the compounds are around 0.5

kcal/mol. Whereas, at the HF level, the average value of BSSE is in the range of 1.5–2 kcal/mol.

In **Figures 5–8** we report the evolution of the individual QC contributions along the series. We denote by E_1 the “frozen” energy contribution from ALMOEDA and the sum of the Coulomb, E_C , and exchange, E_X , contributions from the RVS analyses. E_1 (SIBFA) is correspondingly the sum of its penetration-augmented electrostatic and short-range repulsion contributions. We denote by E_2 the sum of the polarization and charge-transfer contributions in all approaches.

In the perspective of long-duration polarizable MD on the complexes of vDNA and of the Integrase-vDNA assembly, it is essential to evaluate how well could a procedure such as SIBFA account for the ΔE (QC) trends. This alone could lend credence to comparative energy balances and prospective free energy calculations bearing on complexes out of reach of QC calculations, but amenable to this, and related procedures such as those between “improved” DTG derivatives with v-DNA, let alone with INT, which total several thousands of atoms. **Table 1** and **Figure 4** show that ΔE_{tot} (SIBFA) is fully able to recover the trends from $\Delta E(\text{QC})$. C2 and E2 are found to be the two best-bound compounds, while B2, B3, and A2 come next with a small margin. Such agreements are encouraging, considering that no extra calibration effort was done on the DTG derivatives.

Figure 5 reports the evolutions of E_1 at the uncorrelated RVS and correlated B97D3 and ω -B97D levels and those of E_1 (SIBFA). **Figure 6** reports the corresponding evolutions of E_2 . **Figure 7** compares the evolution of uncorrelated ΔE (RVS) and ΔE (SIBFA) without the dispersion contribution E_{disp} . **Figure 8**

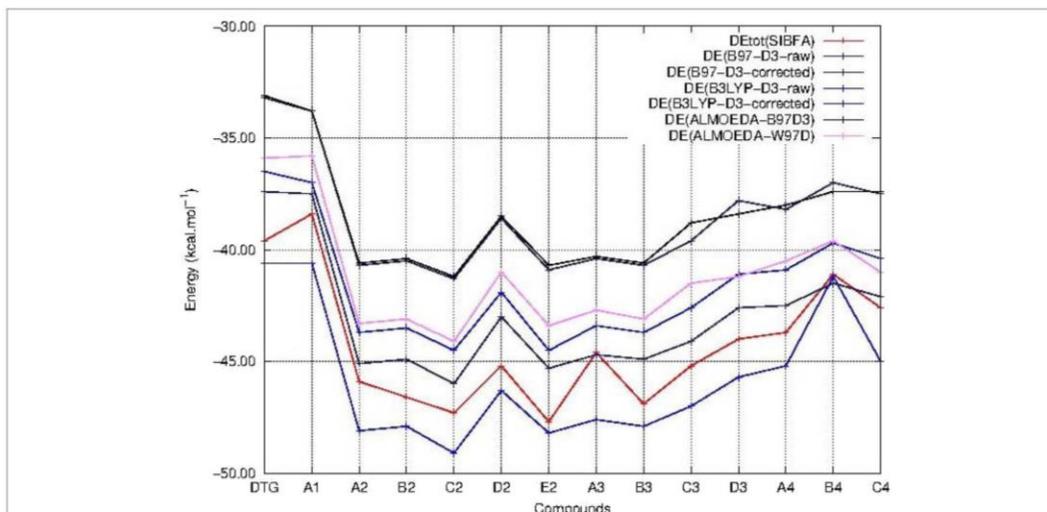


FIGURE 4 | Evolution of the intermolecular interaction energies (ΔE) of the ternary complex; G_4/C_{16} /substituted ring as a function of the halobenzene ring of DTG and the newly conceived rings (Compounds A1 to C4). ΔE values are calculated via quantum chemistry correlated calculations with the B3LYP-D3 functional (ΔE B3LYP-D3-raw before BSSE correction and ΔE B3LYP-D3-corrected after BSSE correction) and the B97-D3 functional (ΔE B97-D3-raw before BSSE correction and ΔE B97-D3-corrected after BSSE correction). ΔE obtained via quantum chemistry energy decomposition analyses methods (EDA) are also presented, precisely the Absolutely Localized Molecular Orbitals method using two functionals; B97D3 (ΔE ALMOEDA-B97D3) and ω -B97D (ΔE ALMOEDA- ω -B97D). ΔE values are obtained as well via molecular mechanics calculations with the polarizable force field SIBFA (ΔE_{tot} SIBFA).

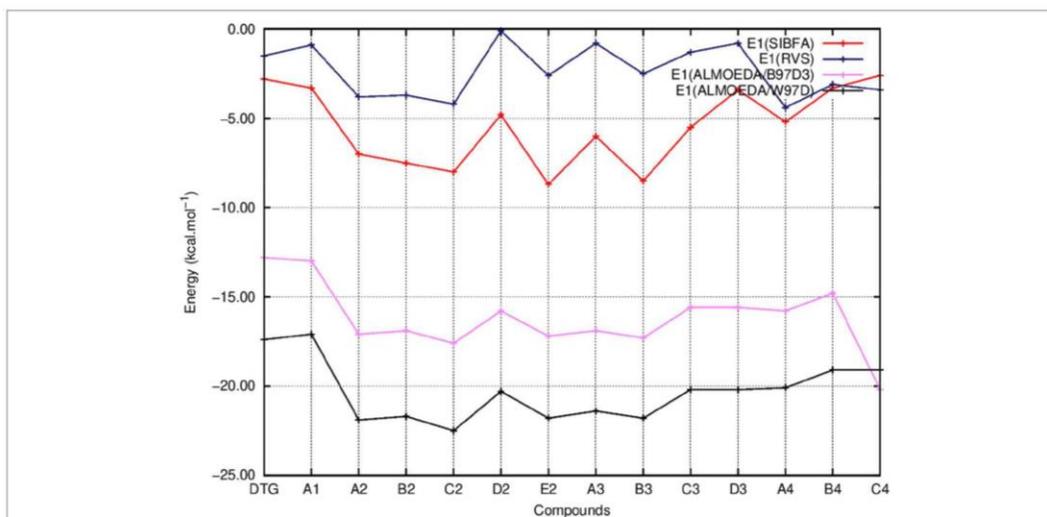


FIGURE 5 | Evolution of the first order energy (E_1) of the ternary complex; G_4/C_{16} /substituted ring as a function of the halobenzene ring of DTG and the newly conceived rings (Compounds A1 to C4). E_1 values are calculated via quantum chemistry Energy Decomposition Analysis methods (EDA) precisely the Reduced Variational Space method (E_1 RVS) and Absolutely Localized Molecular Orbitals method using two functionals; B97D3 (E_1 ALMOEDA/B97D3) and ω 97D (E_1 ALMOEDA- ω -B97D). E_1 values are obtained as well via molecular mechanics calculations with the polarizable force field SIBFA (E_1 SIBFA).

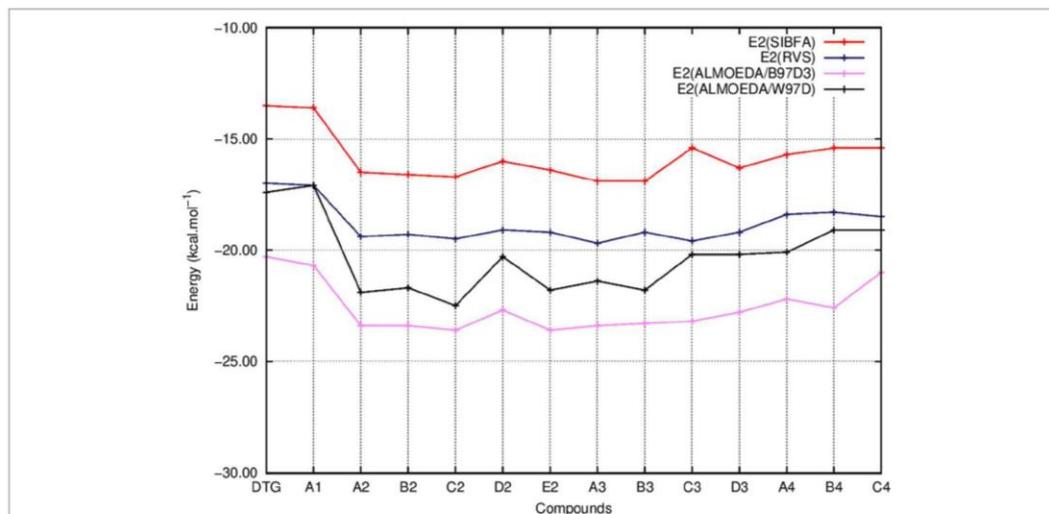


FIGURE 6 | Evolution of the second order energy (E_2) of the ternary complex; G_4/C_{16} /substituted ring as a function of the halobenzene ring of DTG and the newly conceived rings (Compounds A1 to C4). E_2 values are calculated via quantum chemistry Energy Decomposition Analysis methods (EDA), precisely the Reduced Variational Space method (E_2 RVS) and Absolutely Localized Molecular Orbitals method using two functionals; B97D3 (E_2 ALMOEDA/B97D3) and ω -B97D (E_2 ALMOEDA/ ω -B97D). E_2 values were obtained as well via molecular mechanics calculations with the polarizable force field SIBFA (E_2 SIBFA).

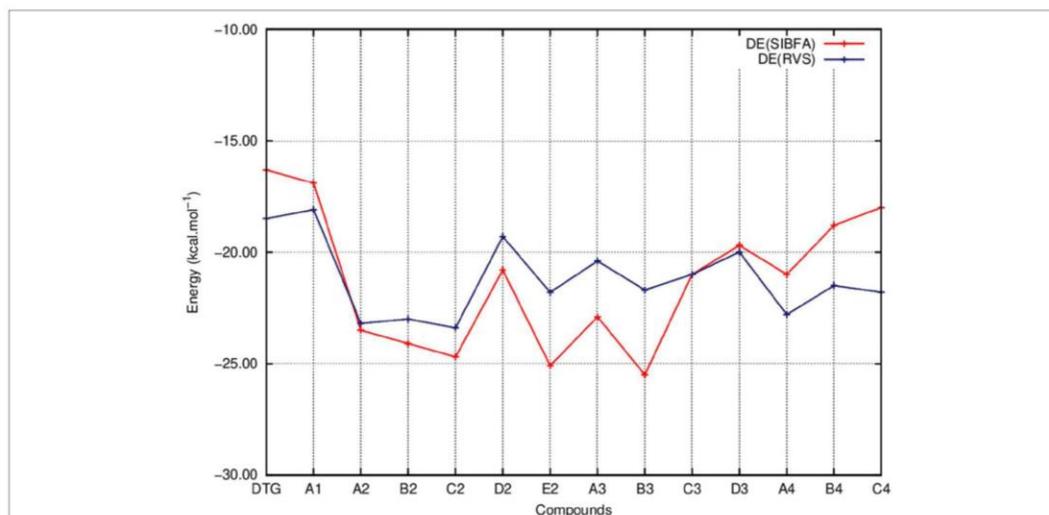


FIGURE 7 | Evolution of the interaction energy of the ternary complex; G_4/C_{16} /substituted ring as a function of the halobenzene ring of DTG and the newly conceived rings (Compounds A1 to C4). ΔE values are calculated via quantum chemistry Energy Decomposition Analysis method (EDA), precisely the Reduced Variational Space method (ΔE RVS) and via molecular mechanics calculations with the polarizable force field SIBFA (ΔE SIBFA).

reports the evolution of a contribution denoted as “ $E_{\text{disp}}/E_{\text{corr}}$ ” the gain in ΔE upon passing from the uncorrelated RVS ΔE to the correlated B3LYP-D3, B97-D3, and ω -B97D levels, along

with E_{disp} (SIBFA). There is no explicit ALMOEDA “dispersion” contribution, as it is included in the van der Waals kernel for both E_1 and E_2 .

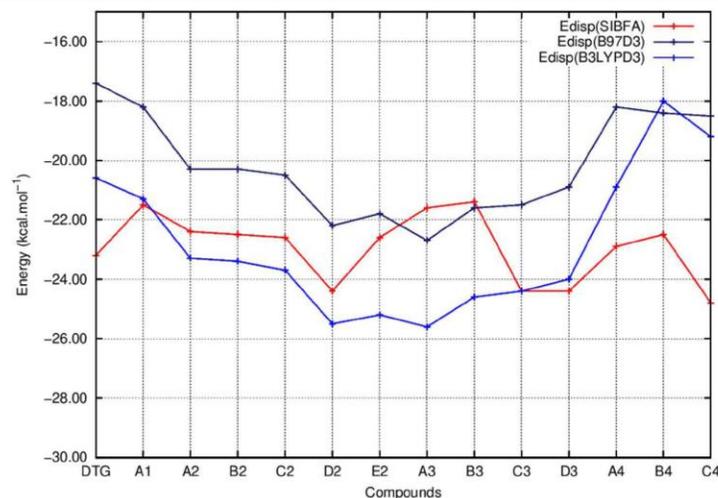


FIGURE 8 | Evolution of the dispersion energy (E_{disp}) of the ternary complex; G_4/C_{16} /substituted ring as a function of the halobenzene ring of DTG and the newly conceived rings (Compounds A1 to C4). E_{disp} values correspond to the ones calculated via quantum chemistry correlated calculations using two functionals; B97D3 (E_{disp} B97D3) and B3LYPD3 (E_{disp} B3LYPD3) and via molecular mechanics calculations with the polarizable force field SIBFA (E_{disp} SIBFA).

Throughout we will denote by ΔE (SIBFA) and ΔE_{tot} (SIBFA) the SIBFA intermolecular interaction energies without and with the dispersion contribution. ΔE (RVS) denotes the BSSE-corrected RVS intermolecular interaction energy. ΔE (QC) denotes the correlated QC intermolecular interaction energy with the B3LYP-D3, B97-D3, or ω -B97D functionals. The QC-derived “dispersion” energy is the difference between ΔE (QC) and ΔE (HF), the latter being derived from a Hartree-Fock computation without removing the BSSE.

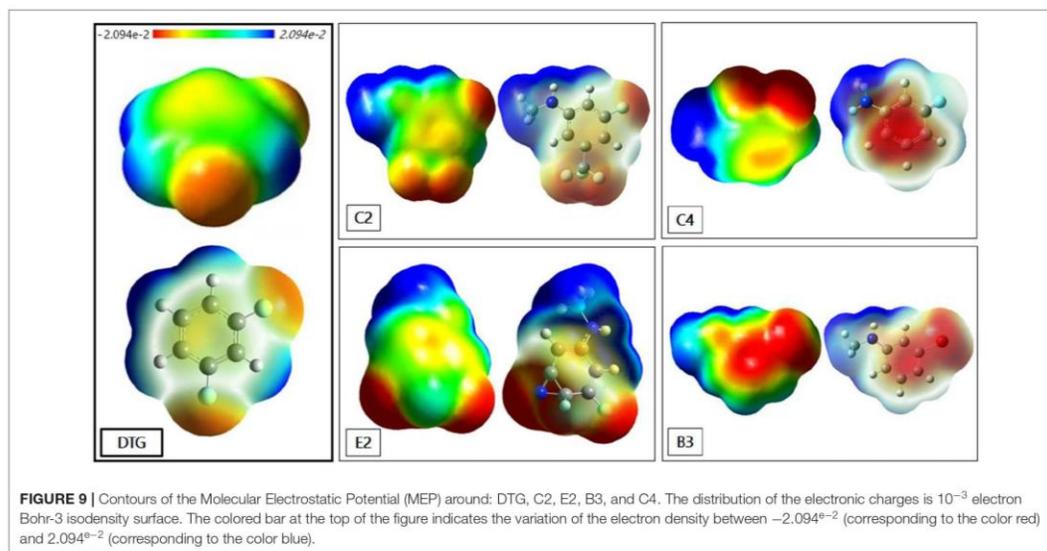
Figure 5 shows that the trends in ΔE (QC) and ΔE_{tot} (SIBFA) giving distinct preferences for derivatives A2, B2, C2, E2, and B3 are retrieved by E_1 , while the E_2 curves (Figure 6) are much shallower. Figure 7 shows that similar to ΔE (QC) and ΔE_{tot} (SIBFA), both ΔE (RVS) and ΔE (SIBFA) curves have minima with derivatives A2–C2, E2, and B3. ΔE (RVS) has an additional minimum with derivative A4 from series 4. The corresponding ΔE (SIBFA) minimum is higher in energy. There is a lesser correspondence between E_{disp} (SIBFA) and those found from the B97D3 and B3LYP-D3 calculations (Figure 8). For compounds A1–E2, it has values intermediate between those from both functionals. For A3 and B3, it is close to the B97D3 values, with differences of 1.2 and 0.2 kcal/mol, respectively. For C3 and D3, it is close to the B3LYP-D3 values, with differences of 0.0 and 0.4 kcal/mol, respectively. For all three compounds in series 4 having the *para* $-\text{NH}_2$ substituent, E_{disp} (SIBFA) has smaller values than either functional. In fact, the trends are not consistent between the two functionals. Thus, while E_{disp} (B3LYP-D3) is larger by 2.7 kcal/mol than E_{disp} (B97-D3), both functionals give very close E_{disp} values in both B3 and C3 complexes.

Even though it appears satisfactory, the agreement between ΔE_{tot} (SIBFA) and ΔE (QC) could be further improved in the near future. There are ongoing SIBFA refinements in the context of reconstruction of a new library of protein, DNA, and ligand constitutive fragments. They resort to multipoles and polarizabilities from correlated calculations, and a rescaling of the individual energy contributions on the basis of correlated Symmetry Adapted Perturbation Theory SAPT-DFT (Misquitta and Szalewicz, 2002; Podeszwa et al., 2006) energy decomposition analyses. They will be reported elsewhere.

Molecular Electrostatic Potentials (MEP)

It is instructive to represent the impact of some of the reported substitutions on the MEP contours. Figure 9 presents such contours around five representative compounds: DTG; C2 and E2; the *meta*-substituted trifluoro- and cyano derivatives of group 2, respectively; B3, the *meta*-substituted bromine derivative of series 3; and C4, the difluorine-substituted derivative of group 4.

There is a much wider extension of the “blue” zone of positive potential around the *para* $-\text{NHCH}_3$ or $-\text{NH}_2$ substituents than around the *para*-fluorine substituent of DTG. This explains the better efficiency observed in the case of the newly conceived compounds in comparison with DTG. In fact, the lower electron density of NHCH_3 and NH_2 confirms the better electrostatic interaction of these *para* rings with the electron rich polycyclic structure of G_4 . For the compounds of group 2, the $-\text{CF}_3$ substituted derivative C2 generates a much wider and delocalized “red” zone of negative potential than the cyano-substituted derivative E2. The extension of positive potential around the $-\text{NHCH}_3$ group is also larger for C2 than for E2. To a



large extent, the first feature should explain the 2.1 kcal/mol larger ΔG_{solv} energy of C2 than E2, impacting the energy balance in favor of E2 while both derivatives had close ΔE (QC) and ΔE_{tot} (SIBFA) values. Derivative B3 also displays a wide zone of negative electrostatic potential around the -Br substituent, and this zone is significantly more extended than around the two -F substituents of derivative C4. The latter presents a low electron density in the middle of its ring due to the presence of its two-electron attracting fluorines, giving it the least favorable attraction with the DNA viral base pairs.

CONCLUSIONS AND PERSPECTIVES

We have designed a series of derivatives of the halobenzene ring of dolutegravir (DTG), the most potent inhibitor of the HIV-1 integrase to date. These derivatives target the highly conserved G_4/C_{16} base pair of viral DNA. The *-para* fluorine ring of DTG is replaced by a $-NHCH_3$ or a $-NH_2$ substituent. All derivatives reported in this study had more favorable intermolecular interaction energies with this base pair than DTG, as computed by quantum chemistry, and a polarizable molecular mechanics procedure, SIBFA. The two compounds with the highest G_4/C_{16} affinities had a *para* $-NHCH_3$ substituent and two substituents *meta* to it: fluorine and trifluoromethyl, denoted as C2, and fluorine and cyano, denoted as E2. ΔE_{tot} (SIBFA) displayed trends consistent with ΔE (QC). This study, along with our previous ones (El Hage et al., 2014, 2015; El Khoury et al., 2019) shows that it is possible to design, in a piece-wise fashion, novel derivatives targeting a well-defined, highly conserved, subset of the recognition site of a large macromolecular

target. It also constitutes an essential validation step prior to large-scale polarizable molecular dynamics simulations of the complex of the entirety of the drug with the entirety of the target.

DATA AVAILABILITY STATEMENT

The datasets generated for this study are available on request to the corresponding author.

AUTHOR CONTRIBUTIONS

PD, LK, and KH performed computations. PD, RM, ZH, J-PP, and NG analyzed results. PD, J-PP, and NG wrote the paper.

FUNDING

This work has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 Research and Innovation Program (grant agreement no. 810367), project EMC2.

ACKNOWLEDGMENTS

We wish to thank the Grand Equipement National de Calcul Intensif (GENCI): Institut du Développement et des Ressources en Informatique Scientifique (IDRIS), Centre Informatique de l'Enseignement Supérieur (CINES), France (Project No. x2009-075009), and the Centre Régional Informatique et d'Applications Numériques de Normandie, Rouen, France (Project 1998053).

We thank also for funding this work the Research Council of Saint-Joseph University of Beirut, Lebanon (Project FS71), the Lebanese National Council for Scientific Research, CNRS-L

(Projects CNRS-FS80 and CNRS-FS 116), and the French Institute-Lebanon and the French-Lebanese Program CEDRE (Project 35327U)).

REFERENCES

- Ammar, F. F., Hobaika, Z., Abdel-Azeim, S., Zargarian, L., Maroun, R. G., and Fermandjian, S. (2016). A targeted DNA substrate mechanism for the inhibition of HIV-1 integrase by inhibitors with antiretroviral activity. *FEBS Press* 6, 234–250. doi: 10.1002/2211-5463.12025
- Azar, R. J., and Head-Gordon, M. (2011). An energy decomposition analysis for intermolecular interactions from an absolutely localized molecular orbital reference at the coupled-cluster singles and doubles level. *J. Chem. Phys.* 132:024103. doi: 10.1063/1.3674992
- Chaires, J. B. (2006). A thermodynamic signature for drug-DNA binding mode. *Arch. Biochem. Biophys.* 453, 26–31. doi: 10.1016/j.abb.2006.03.027
- Creuzet, S., Langlet, J., and Gresh, N. (1991). Adjustment of the SIBFA method for potential maps to study hydrogen bonding vibrational frequencies. *J. Chem. Phys.* 88, 2399–2409. doi: 10.1051/jcp/1991882399
- El Hage, K., Piquemal, J. P., Hobaika, Z., Maroun, R. G., and Gresh, N. (2013). Could an anisotropic molecular mechanics/dynamics potential account for sigma hole effects in the complexes of halogenated compounds? 34, 1125–1135. doi: 10.1002/jcc.23242
- El Hage, K., Piquemal, J. P., Hobaika, Z., Maroun, R. G., and Gresh, N. (2014). Substituent-modulated affinities of halobenzene derivatives to the HIV-1 integrase recognition site. Analyses of the interaction energies by parallel quantum chemical and polarizable molecular mechanics. *J. Phys. Chem. A* 118, 9772–9782. doi: 10.1021/jp5079899
- El Hage, K., Piquemal, J. P., Hobaika, Z., Maroun, R. G., and Gresh, N. (2015). Could the “Janus-like” properties of the halobenzene CX bond (X = Cl, Br) be leveraged to enhance molecular recognition? 36, 210–221. doi: 10.1002/jcc.23786
- El Khoury, L., El Hage, K., Piquemal, J. P., Fermandjian, S., Maroun, R. G., Gresh, N., et al. (2019). Spectrometric and computational studies of the binding of HIV-1 integrase inhibitors to viral DNA extremities. *Peer J Phys. Chem.* 1:e6. doi: 10.7717/peerj-pchem.6
- El Khoury, L., Piquemal, J. P., Fermandjian, S., Maroun, R. G., Gresh, N., and Hobaika, Z. (2017). The inhibition process of HIV-1 integrase by diketocoids molecules: understanding the factors governing the better efficiency of dolutegravir. *Biochem. Biophys. Res. Commun.* 488, 433–438. doi: 10.1016/j.bbrc.2017.05.001
- Feller, D. (1996). The role of databases in support of computational chemistry calculations. *J. Comput. Chem.* 17, 1571–1586.
- Frisch, M. J., Trucks, G. W., Schlegel, H. B., Scuseria, G. E., Robb, M. A., Cheeseman, J. R., et al. (2009). *Gaussian 09*. Wallingford, CT: Gaussian Inc.
- Garmer, D. R., and Stevens, W. J. (1989). Transferability of molecular distributed polarizabilities from a simple localized orbital based method. *J. Phys. Chem.* 93, 8263–8270. doi: 10.1021/j100362a023
- Goerigk, L., and Grimme, S. (2011). A thorough benchmark of density functional methods for general main group thermochemistry, kinetics, and noncovalent interactions[†]. *J. Phys. Chem. Chem. Phys.* 13, 6670–6688. doi: 10.1039/c0cp02984j
- Gresh, N., Cisneros, G. A., Darden, T. A., and Piquemal, J. P. (2007). Anisotropic, polarizable molecular mechanics studies of inter- and intramolecular interactions and ligand-macromolecule complexes. a bottom-up strategy. *J. Chem. Theory Comput.* 3, 1960–1986. doi: 10.1021/ct700134r
- Grimme, S. (2006). Semiempirical GGA-type density functional constructed with a long-range dispersion correction. *J. Comput. Chem.* 27, 1787–1799. doi: 10.1002/jcc.20495
- Hare, S., Gupta, S. S., Valkov, E., Engelman, A., and Cherepanov, P. (2010b). Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature* 464, 232–236. doi: 10.1038/nature08784
- Hare, S., Smith, S. J., Métifiot, M., Jaxa-Chamiec, A., Pommier, Y., Hughes, S. H., et al. (2011). Structural and functional analyses of the second-generation integrase strand transfer inhibitor dolutegravir (S/GSK1349572). *Mol. Pharmacol.* 80, 565–572. doi: 10.1124/mol.111.073189
- Hare, S., Vos, A. M., Clayton, R. F., Thuring, J. W., Cummings, M. D., and Cherepanov, P. (2010a). Molecular mechanisms of retroviral integrase inhibition and the evolution of viral resistance. *PNAS* 107, 20057–20062. doi: 10.1073/pnas.1010246107
- Hironori, N., Hayashi, T., Takahashi, T., Miyano, M., Kannagi, M., and Masuda, T. (2009). Augmentation of reverse transcription by integrase through an interaction with host factor, SIP1/Gemin2 is critical for HIV-1 infection. *PLoS ONE* 4:e7825. doi: 10.1371/journal.pone.0007825
- Kessl, J. J., Kutluay, S. B., Townsend, D., Rebersburg, S., Slaughter, A., Larue, R. C., et al. (2016). HIV-1 integrase binds the viral RNA genome and is essential during virion morphogenesis. *Cell* 166, 1257–1268. doi: 10.1016/j.cell.2016.07.044
- Magardère, L., Jolly, L. H., Lipparini, F., Aviat, F., Stamm, B., Jing, Z. F., et al. (2018). Tinker-HP: a massively parallel molecular dynamics package for multiscale simulations of large complex systems with advanced point dipole polarizable force fields. *Chem. Sci.* 9, 956–972. doi: 10.1039/C7SC04531J
- Lesbats, P., Engelman, A. N., and Cherepanov, P. (2016). Retroviral DNA integration. *Chem. Rev.* 116, 12730–12757. doi: 10.1021/acs.chemrev.6b00125
- Liao, C., Marchand, C., Burke, T. R. Jr., Pommier, Y., and Nicklaus, M. C. (2010). Authentic HIV-1 integrase inhibitors. *Future Med. Chem.* 2, 1107–1122. doi: 10.4155/fmc.10.199
- Mennucci, B., Tomasi, J., Cammi, R., Cheeseman, J. R., Frisch, M. J., Devlin, F. J., et al. (2002). Polarizable Continuum Model (PCM) calculations of solvent effects on optical rotations of chiral molecules. *J. Phys. Chem. A* 106, 6102–6113. doi: 10.1021/jp020124t
- Misquitta, A., and Szalewicz, K. (2002). Intermolecular forces from asymptotically corrected density functional description of monomers. *Chem. Phys. Lett.* 357, 301–306. doi: 10.1016/S0009-2614(02)00533-X
- Mouscadet, J. F., Tchertanov, L., and Deprez, E. (2007). Structures et rôles de l'intégrase du VIH dans le cycle viral. *Virologie* 11, 143–152. doi: 10.1684/vir.2007.0127
- Murray, J. S., Lane, P., and Politzer, P. (2008). Expansion of the sigma-hole concept. *J. Mol. Model* 15, 723–729. doi: 10.1007/s00894-008-0386-9
- Piquemal, J. P., Chevreau, H., and Gresh, N. (2007). Toward a separate reproduction of the contributions to the hartree-fock and DFT intermolecular interaction energies by polarizable molecular mechanics with the SIBFA potential. *J. Chem. Theory Comput.* 3, 824–837. doi: 10.1021/ct7000182
- Podeszwa, R., Bukowski, R., and Szalewicz, K. (2006). Density-fitting method in symmetry-adapted perturbation theory based on kohn–sham description of monomers. *J. Chem. Theory Comput.* 2, 400–412. doi: 10.1021/ct050304h
- Ponder, J., Wu, C., Ren, P., Pande, V. S., Chodera, J. D., Schieders, M. J., et al. (2010). Current status of the AMOEBA polarizable force field. *J. Phys. Chem.* 114, 2549–2564. doi: 10.1021/jp910674d
- Schmidt, M. W., Baldridge, K. K., Boatz, J. A., Elbert, S. T., Gordon, M. S., Jensen, J. H., et al. (1993). General atomic and molecular electronic structure system. *J. Comput. Chem.* 14, 1347–1363. doi: 10.1002/jcc.540141112
- Shao, Y., Gan, Z., Epifanovsky, E., Gilbert, A. T. B., Wormit, M., Kusmann, J., et al. (2015). Advances in molecular quantum chemistry contained in the Q-Chem 4 program package. *Mol. Phys.* 113, 184–215. doi: 10.1080/00268976.2014.952696
- Shimura, K., Kodama, E., Sakagami, Y., Matsuzaki, Y., Watanabe, W., and Yamataka, K., et al. (2008). Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137). *J. Virol.* 82, 764–774. doi: 10.1128/JVI.01534-07
- Simon, S., and Duran, M. (1996). How does basis set superposition error change the potential surfaces for hydrogen-bonded dimers? *J. Chem. Phys.* 105:11024. doi: 10.1063/1.472902
- Stevens, W. J., and Fink, W. H. (1987). Frozen fragment reduced variational space analysis of hydrogen bonding interactions. Application to the water dimer. *Chem Phys Lett.* 139, 15–22. doi: 10.1016/0009-2614(87)80143-4

- Stone, A. J., and Alderton, M. (1985). Distributed multipole analysis methods and applications. *Mol. Phys.* 100, 221–233. doi: 10.1080/00268970110089432
- Summa, V., Petrocchi, A., Bonelli, F., Crescenzi, B., Donghi, M., and Ferrara, M., et al. (2008). Discovery of raltegravir, a potent, selective orally bioavailable HIV-integrase inhibitor for the treatment of HIV-AIDS infection. *J. Med. Chem.* 51, 5843–5855. doi: 10.1021/jm800245z
- Underwood, M. R., Johns, B. A., Sato, A., Martin, J. N., Deeks, S. G., and Fujiwara, T. (2012). The activity of the integrase inhibitor dolutegravir against HIV-1 variants isolated from raltegravir-treated adults. *J. Acquired Immunodefic. Synd.* 61, 297–301. doi: 10.1097/QAI.0b013e31826bfd02
- Vigné-Maeder, F., and Claverie, P. (1988). The exact multicenter multipolar part of a molecular charge distribution and its simplified representations. *J. Chem. Phys.* 88, 4934–4948. doi: 10.1063/1.454705
- Wainberg, M. A., Zaharatos, G. J., and Brenner, B. G. (2011). Development of antiretroviral drug resistance. *N. Engl. J. Med.* 365, 637–646. doi: 10.1056/NEJMra1004180

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2020 El Darazi, El Khoury, El Hage, Maroun, Hobaika, Piquemal and Gresh. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Publication 2 (En cours de rédaction)

Evaluating the interaction between HIV-1 viral intasome complex and novel suggested integrase inhibitors: An absolute alchemical free energy approach

Perla El Darazi^{1,2, 1, 1, 1, 1}, Zeina Hobaika², Nohad Gresh¹, Richard G. Maroun², and Jean-Philip Piquemal^{*1,4,5}

¹Sorbonne Université, Laboratoire de Chimie Théorique, UMR 7616 CNRS, 75005, Paris, France

²Equipe Structure et Interactions des macromolécules, UR EGP, Centre d'Analyses et de Recherche, Faculté des Sciences, Université Saint-Joseph de Beyrouth, Beirut, Lebanon

³Sorbonne Université, Institut Parisien de Chimie Physique et Théorique, FR 2622 CNRS, 75005, Paris, France

⁴UT Austin, Department of Biomedical Engineering, TX 78712, USA.

⁵Institut Universitaire de France, 75005, Paris, France.

ABSTRACT

We studied, using computational chemistry approaches, novel inhibitors targeting the integrase (IN) of the Human Immunodeficiency Virus-1 (HIV-1); causal agent of the Acquired Immunodeficiency Syndrome (AIDS). IN is in fact an essential viral enzyme that allows for the smooth integration of the viral DNA into target cells via two steps; 3'-processing followed by strand transfer (ST), making it a major HIV-1 therapeutic target. Recent anti-AIDS therapies resort to the usage of IN inhibitors belonging to the Diketoacid family (DKAs), the most recent of which is the Dolutegravir (DTG). These DKAs act on the IN-DNA interface and are known as Integrase Strand Transfer Inhibitors (INSTIs), inhibiting specifically the strand transfer step. Unfortunately, resistance mutations have been observed at the IN level after its contact with these inhibitory molecules, including DTG, compromising the long-term efficiency of anti-HIV therapies. One solution to overcome this problem is to reinforce the interactions of DTG with viral DNA, at the interface IN-viral DNA; this could circumvent the need for specific interactions with the protein and hence, bypass resistance mutations. We had previously designed thirteen novel molecules, via substitutions made on the halobenzene ring of DTG and evaluated their intermolecular interaction energy (E) with terminal base pairs of HIV-1 DNA (C16 and G4) via quantum mechanics (QM) and molecular mechanics (MM) calculations. Based on their energetic yield, all newly conceived molecules showcased a better interaction with these terminal base pairs than DTG.

For this work, we have selected three of our best derivatives (compounds named B3, A2 and B2) to further evaluate their interaction with viral DNA in comparison with DTG. This was done using absolute alchemical free energy calculations (G) based on all-atom molecular dynamics (MD) simulations. With this rigorous approach, we calculated the free energy of interaction of our four ligands (DTG, B3, A2 and B2) with the viral intasome complex (IN-viral DNA). We have evaluated as well the free energy of interaction of DTG with the IN alone and viral DNA alone, respectively. These calculations were based on the crystallographic structure of the HIV-1 intasome in complex with DTG (PDB code: 3S3M). Our resulting free energy values indicated firstly a more stable interaction of DTG with the viral DNA, compared to its free energy of interaction with HIV-1 integrase, clarifying the better efficiency of DTG in regards to its INSTIS predecessors. Furthermore, B3, A2 and B2 all presented better free energy values than DTG in regards to their interaction within the intasome. The obtained G values were in agreement with our previous E calculations, exhibiting newly designed INSTIs more potent than DTG, with an ability to better bypass IN resistance mutations.

Figures du chapitre IV Partie B

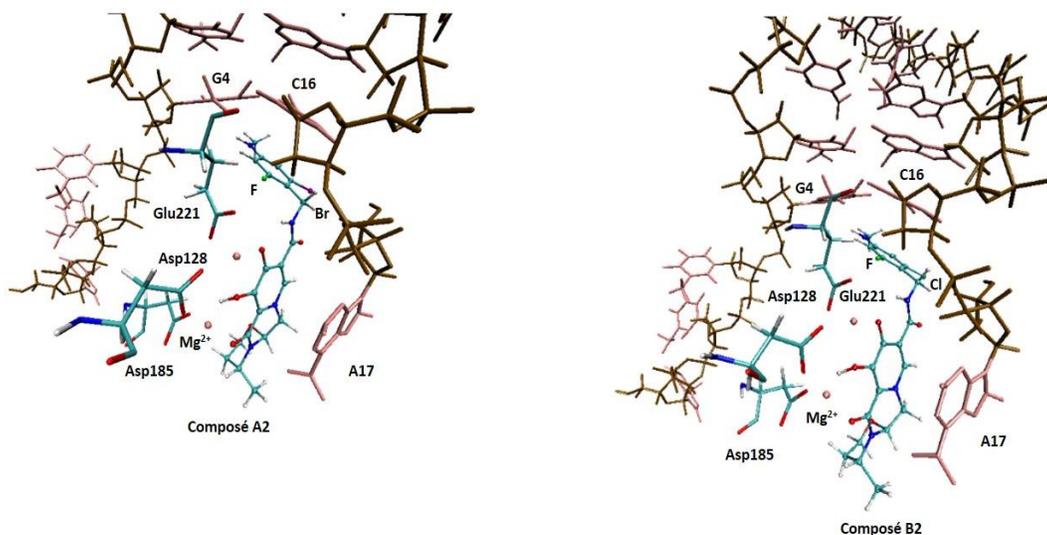


Figure 120 : Positions de départ des structures des complexes ADN-Intégrase-A2 (à gauche) et ADN-Intégrase-B2 (à droite) pour le Prototype Foamy Virus (PFV). L'ADN double brin est représenté en bâtons de couleur rose et marron, A2 et B2 sont représentés en boules pour les atomes et en bâtons pour les liaisons entre eux, avec les cations divalents de magnésium (Mg^{2+}) représentés en sphères roses. La triade catalytique de l'intégrase est schématisée en bâtons, avec Asp128, Asp185 et Glu221 qui correspondent à Asp64, Asp185 et Glu152 du VIH-1 respectivement. Le code couleur pour le DTG et pour la triade catalytique reste le même que celui utilisé précédemment pour les Figures 39 et 40, sauf pour le brome coloré en violet dans ce cas.

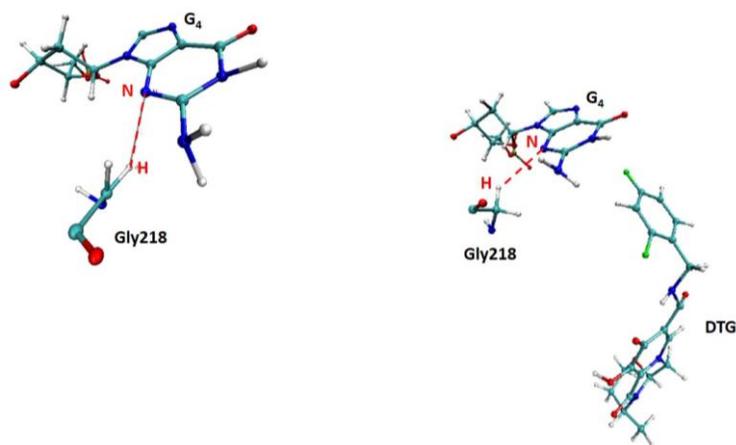


Figure 121 : Représentation des positions de départ de Gly149 du VIH-1 (correspondant à Gly218 du PFV) et de G_4 au niveau des complexes ADN-Intégrase (à gauche) et ADN-Intégrase-DTG (à droite), avec en gras et rouge les atomes dont les distances ont été suivies au cours des simulations. La ligne rouge pointillée indique la distance entre eux. Les composés B3, A2 et B2 ayant des positions de départ similaires au DTG ne sont pas représentés dans la figure.

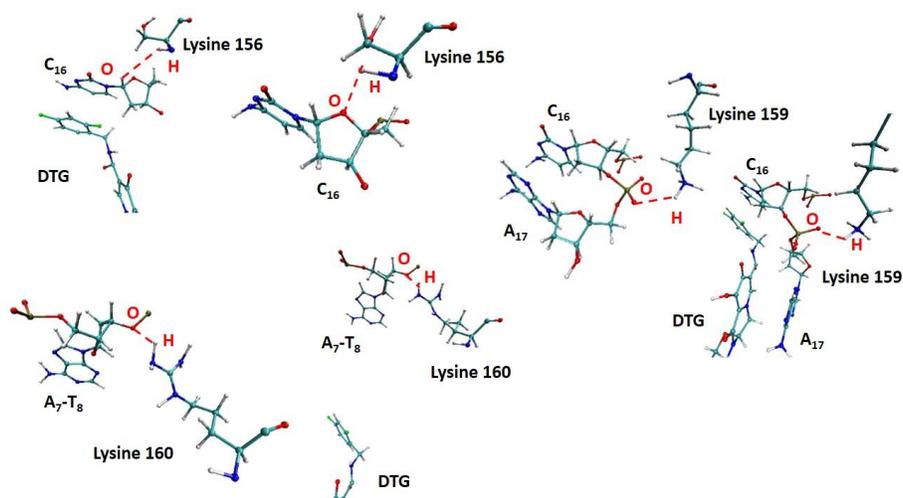


Figure 122 : Représentation des positions de départ de la Lysine 156 (correspondant à la Sérine 225 du PFV) avec C₁₆ au niveau des complexes ADN-Intégrase et ADN-Intégrase-DTG (complexes en haut à gauche), ainsi que les positions de départ de la Lysine 159 avec C₁₆ et A₁₇ au niveau des complexes ADN-Intégrase et ADN-Intégrase-DTG (complexes à droite). Finalement nous trouvons la Lysine 160 avec les bases A₇ et T₈ au niveau des complexes ADN-Intégrase et ADN-Intégrase-DTG (complexes en bas à gauche). Les atomes dont les distances ont été suivies au cours des simulations sont marqués en gras et rouge avec une ligne rouge pointillée indiquant la distance entre eux. Les composés B3, A2 et B2 ayant des positions de départ similaires au DTG ne sont pas représentés dans la figure.

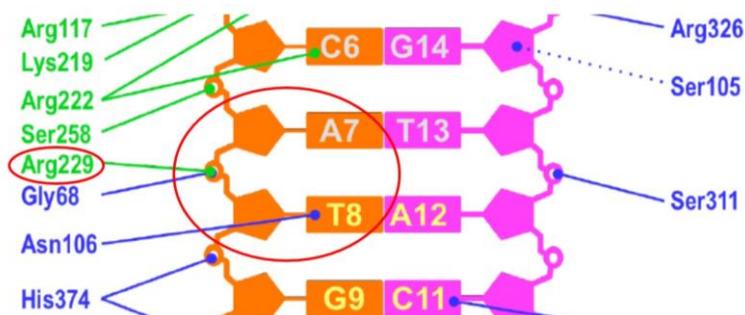


Figure 123 : Schéma représentant les contacts ADN-protéine identifiés dans le complexe de code PDB : 3L2R. Le double brin d'ADN viral est représenté en orange pour le brin non traité et en magenta pour le brin traité. Les traits pleins indiquent les liaisons hydrogènes entre le résidu de l'intégrase et la base virale alors que les traits pointillés correspondent aux interactions de vdW. Nous avons encadré les bases A₇ et T₈ dont les interactions avec la Lysine 228 (Lys 228) correspondant à Lys159 chez le VIH-1, ont été suivies.

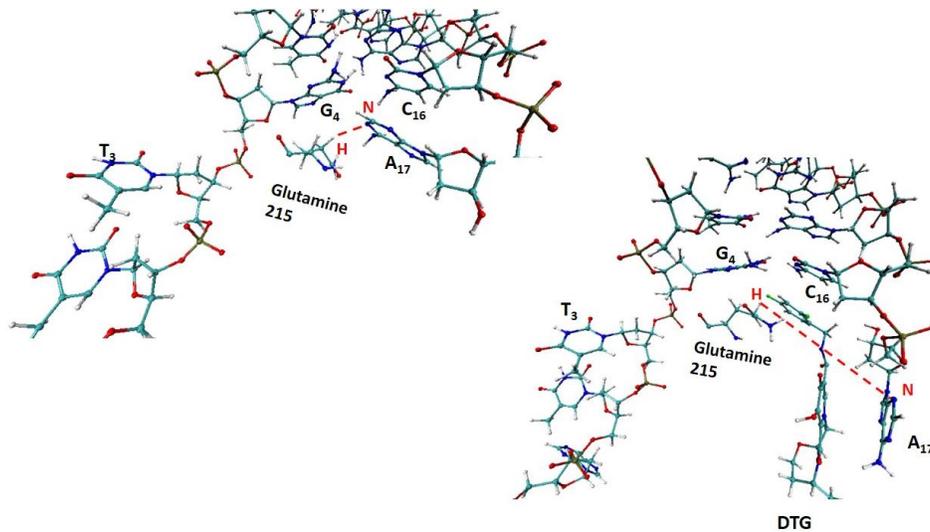


Figure 124 : Représentation des positions de départ de la Glutamine 146 du VIH-1 (correspondant à la Glutamine 215 du PFV) et d'A₁₇ au niveau des complexes ADN-Intégrase (à gauche) et ADN-Intégrase-DTG (à droite), avec en gras et rouge les atomes dont les distances ont été suivies au cours des simulations. La ligne rouge pointillée indique la distance entre eux. Les bases C₁₆, G₄ et T₃ sont de même représentées. Les composés B3, A2 et B2 ayant des positions de départ similaires au DTG ne sont pas représentés dans la figure.

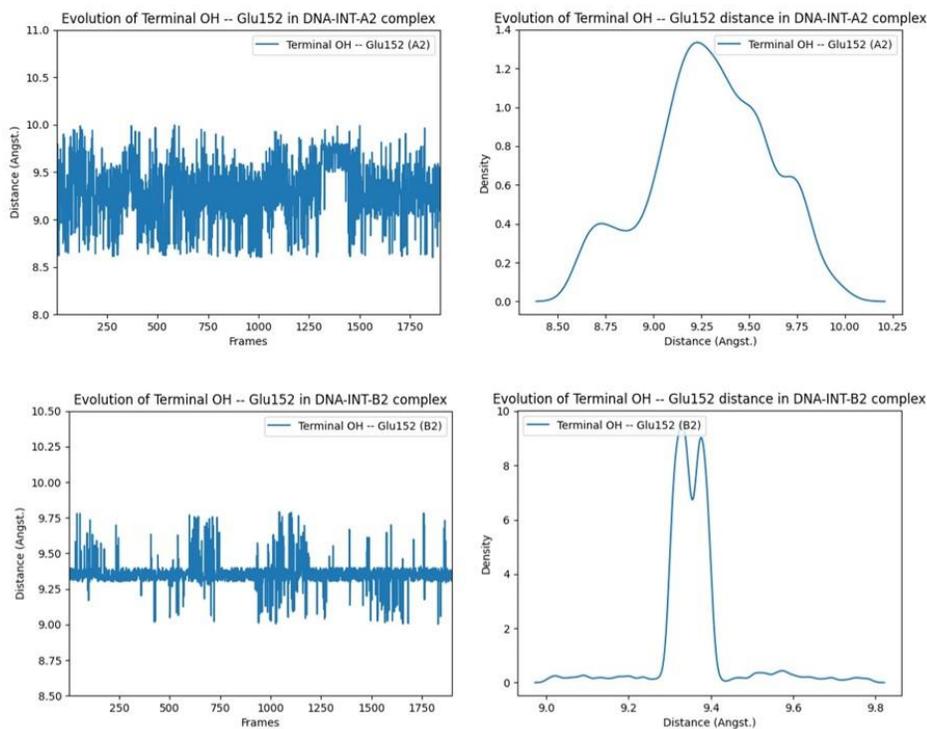


Figure 125 : Graphiques représentant l'évolution de la distance entre l'extrémité terminale 3'OH d'A₁₇ et Glu152 (correspondant à Glu221 pour le PFV) dans les complexes ADN-Intégrase-A2 et ADN-Intégrase-B2 (à gauche). Nous trouvons aussi à droite les graphiques de la densité des pauses en fonction des distances obtenues au niveau des mêmes complexes.

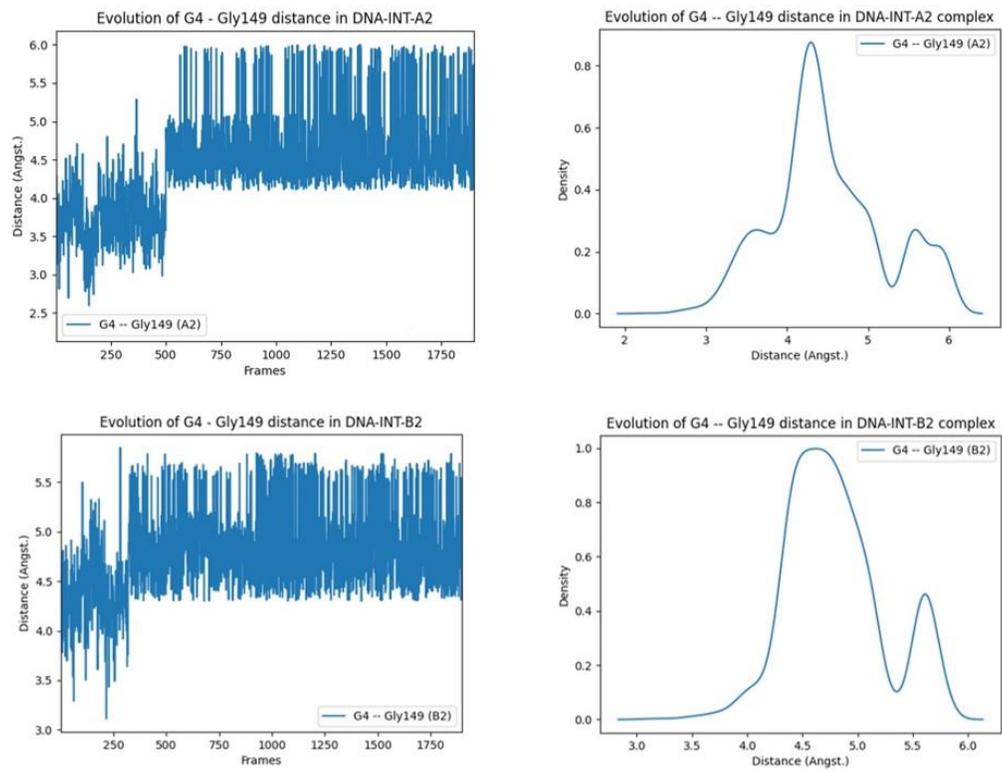


Figure 126 : Graphiques représentant l'évolution de la distance entre la base virale G_4 et Gly149 (correspondant à Gly218 pour le PFV) dans les complexes ADN-Intégrase-A2 et ADN-Intégrase-B2 (à gauche). Nous trouvons aussi à droite les graphiques de la densité des pauses en fonction des distances obtenues au niveau des mêmes complexes.

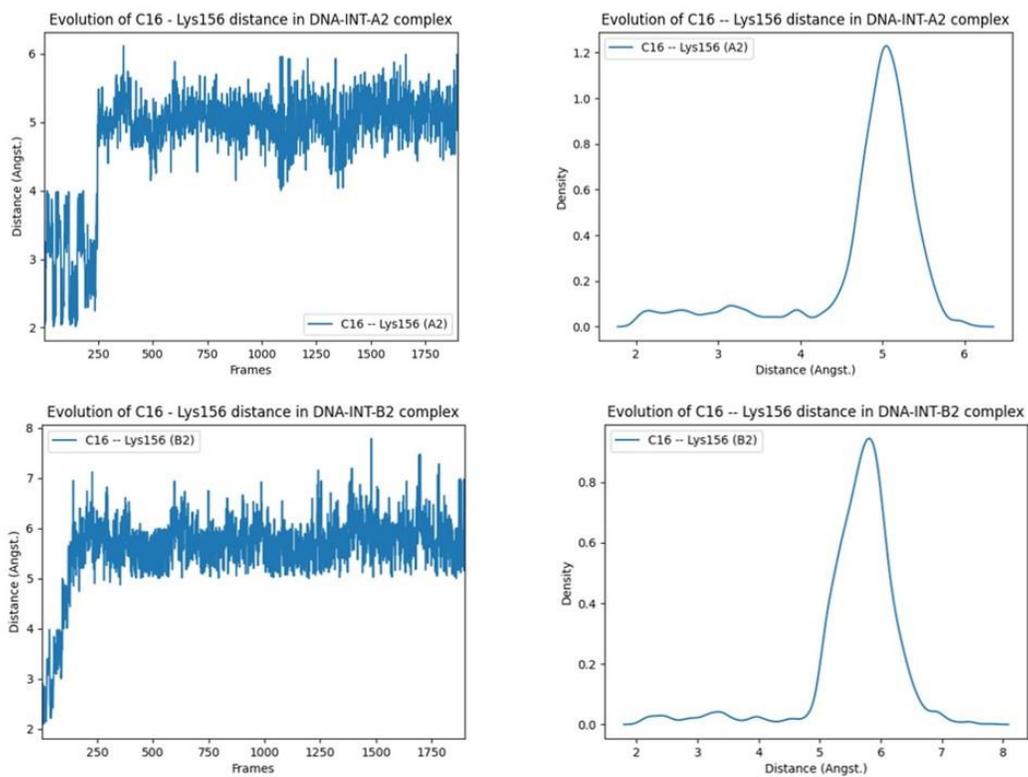
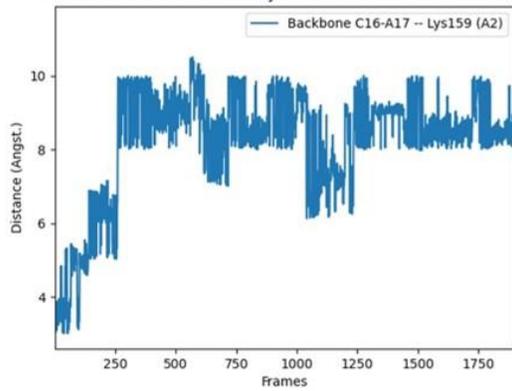
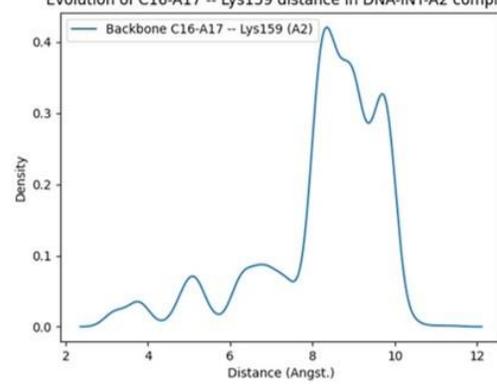


Figure 127 : Graphiques représentant l'évolution de la distance entre la base virale C₁₆ et Lys156 (correspondant à Ser225 pour le PFV) dans les complexes ADN-Intégrase-A2 et ADN-Intégrase-B2 (à gauche). Nous trouvons aussi à droite les graphiques de la densité des pauses en fonction des distances obtenues au niveau des mêmes complexes.

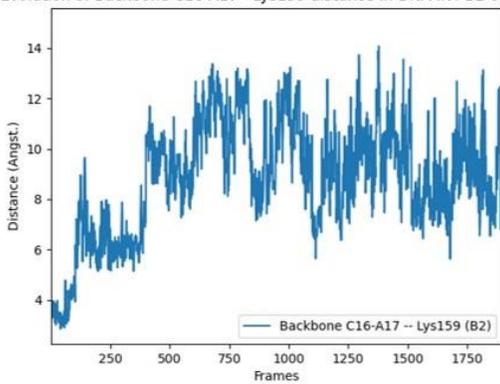
Evolution of Backbone C16-A17 - Lys159 distance in DNA-INT-A2 complex



Evolution of C16-A17 -- Lys159 distance in DNA-INT-A2 complex



Evolution of Backbone C16-A17 - Lys159 distance in DNA-INT-B2 complex



Evolution of C16-A17 -- Lys159 distance in DNA-INT-B2 complex

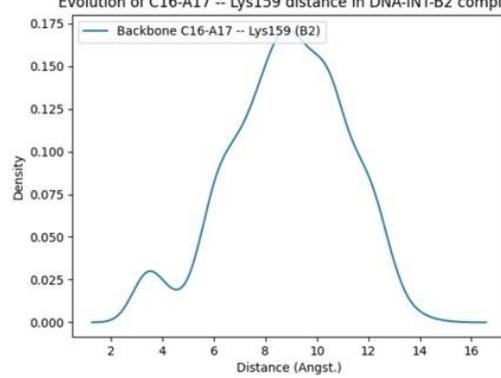


Figure 128 : Graphiques représentant l'évolution de la distance entre les bases virales C₁₆-A₁₇ et Lys159 (correspondant à Lys228 pour le PFV) dans les complexes ADN-Intégrase-A2 et ADN-Intégrase-B2 (à gauche). Nous trouvons aussi à droite les graphiques de la densité des pauses en fonction des distances obtenues au niveau des mêmes complexes.

Suivi de la liaison entre la Lysine160 (correspondant à l'Arginine229 chez le PFV) et le squelette phosphodiester situé entre les bases virales A7-T8 pour les complexes ADN-Intégrase et ADN-Intégrase-Inhibiteurs (Inhibiteurs correspondent au DTG, B3, A2 et B2).

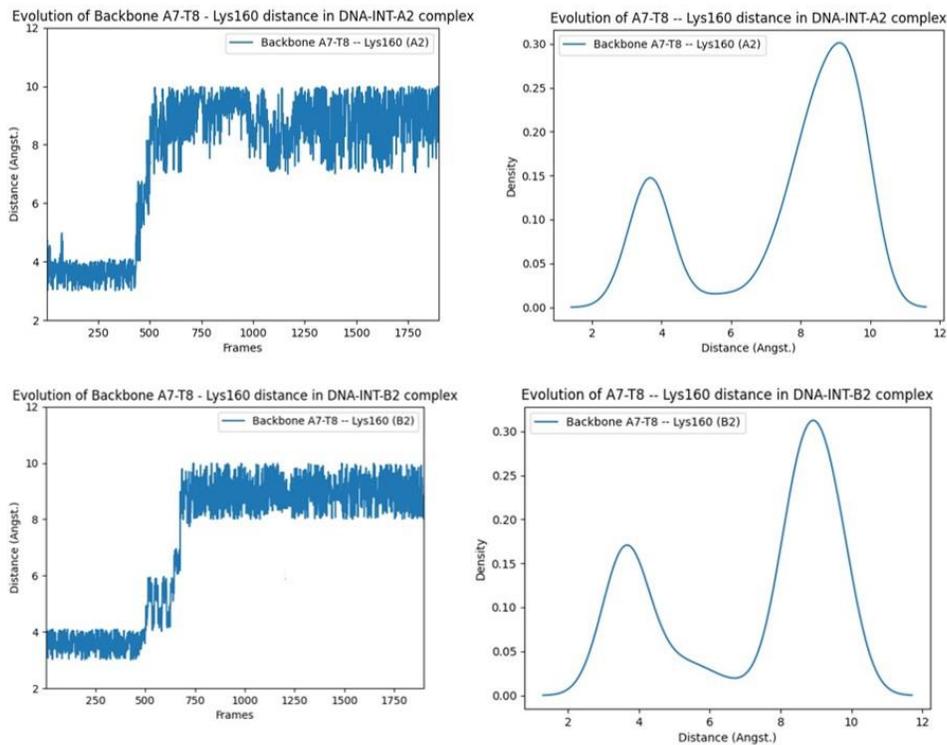


Figure 129 : Graphiques représentant l'évolution de la distance entre les bases virales A₇-T₈ et Lys160 (correspondant à Arg229 pour le PFV) dans les complexes ADN-Intégrase-A2 et ADN-Intégrase-B2 (à gauche). Nous trouvons aussi à droite les graphiques de la densité des pauses en fonction des distances obtenues au niveau des mêmes complexes.

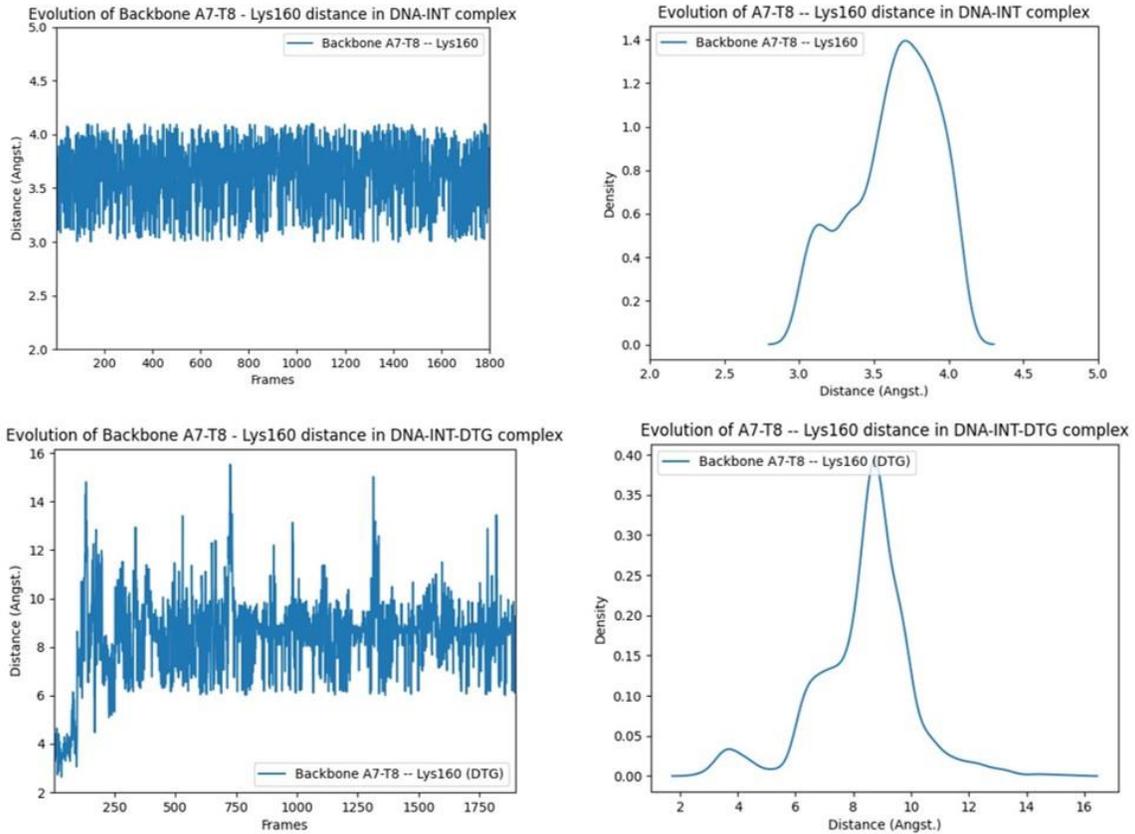


Figure 130 : Graphiques représentant l'évolution de la distance entre les bases virales A₇-T₈ et Lys160 (correspondant à Arg229 pour le PFV) dans les complexes ADN-Intégrase et ADN-Intégrase-DTG (à gauche). Nous trouvons aussi à droite les graphiques de la densité des pauses en fonction des distances obtenues au niveau des mêmes complexes.

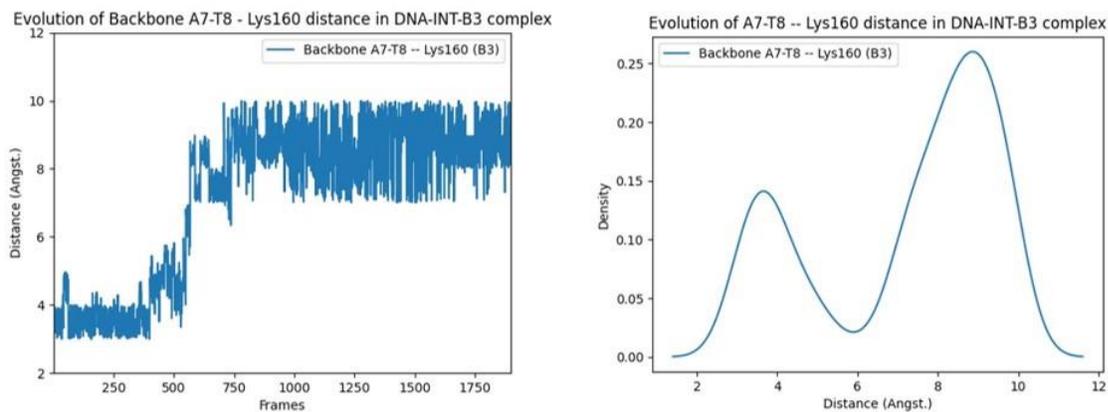


Figure 131 : Graphique représentant l'évolution de la distance entre les bases virales A₇-T₈ et Lys160 (correspondant à Arg229 pour le PFV) dans le complexe ADN-Intégrase-B3 (à gauche). Nous trouvons aussi à droite le graphique de la densité des pauses en fonction des distances obtenues au niveau du même complexe.

Les graphiques de la figure 130 indiquent que la liaison hydrogène est maintenue entre la Lysine 160 et le squelette A₇-T₈ au cours de nos simulations (valeurs entre 3 et 4 Å) avec la majorité des distances de 3,6 Å. Cependant, en présence du DTG, cette liaison hydrogène est perdue après 50 ns de simulation où la distance séparant les résidus concernés devient majoritairement de 9 Å. Dans le cas des trois nouveaux inhibiteurs ; B3, A2 et B2, les liaisons hydrogènes entre K156 et le squelette A₇-T₈ sont maintenues pendant 120 ns de simulation (valeurs variant entre 3 Å et 4 Å) puis sont perdues (des distances de 9 Å apparaissent). Les graphiques et les courbes des figures 129 et 131 illustrent bien ces variations, notamment les courbes à droite qui présentent des pics aux alentours de 3 Å (liaisons hydrogènes) suivis de pics de 9 Å (perte des contacts proches).

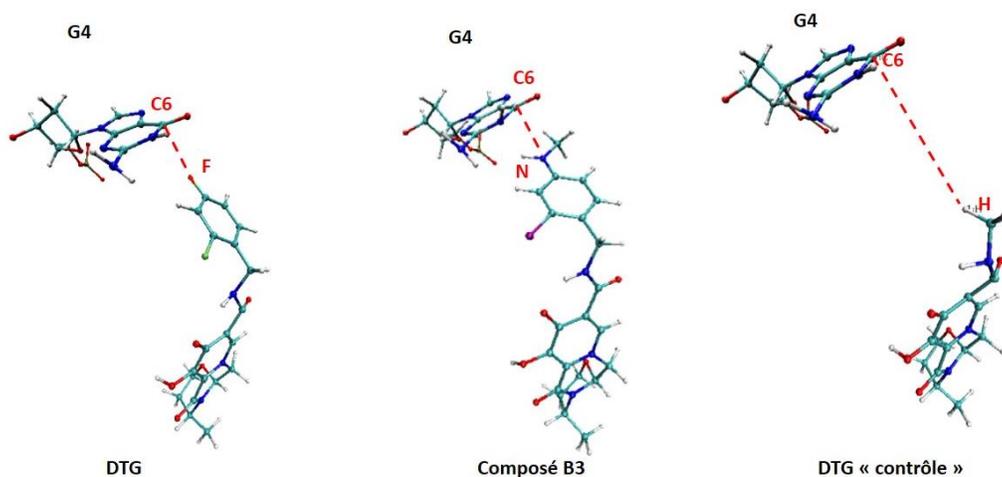


Figure 132 : Représentation des positions de départ dans nos simulations des dérivés DTG, B3 et DTG « contrôle » par rapport à G4, avec en gras et rouge les atomes effectuant des contacts proches entre les dérivés et la base virale. La ligne rouge pointillée indique la distance entre eux. Les composés A2 et B2 ne sont pas représentés puisque leurs positions de départ sont similaires à celle de B3, avec un NHCH₃ en para au niveau de leur noyau halogéné.

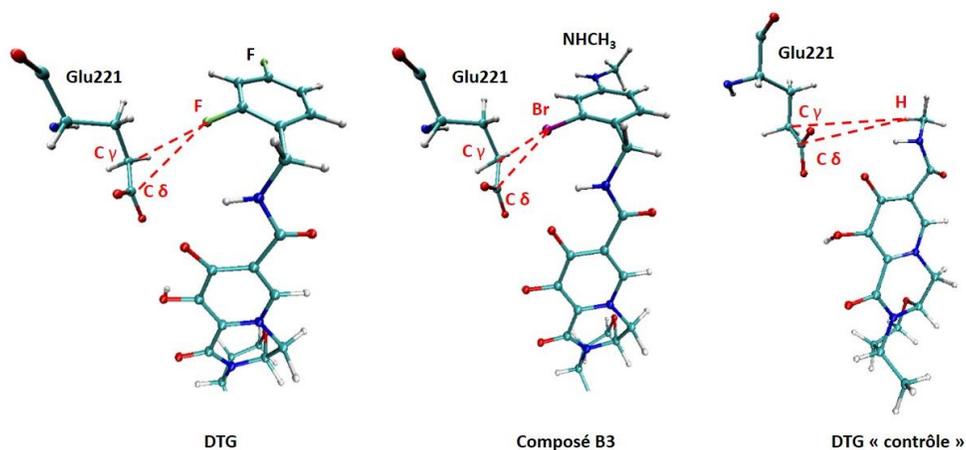


Figure 133 : Représentation des positions de départ dans nos simulations des dérivés DTG, B3 et DTG « contrôle » par rapport à Glu221 chez le Prototype Foamy Virus (PFV), correspondant à Glu152 chez le VIH-1. Les atomes effectuant des contacts proches chez les dérivés et Glu221 sont indiqués en rouge et gras. La ligne rouge pointillée représente la distance entre eux. Etant donné qu'A2 et B2 ont un fluor en ortho face à Glu221 comme le DTG, nous avons uniquement représenté le noyau du DTG.

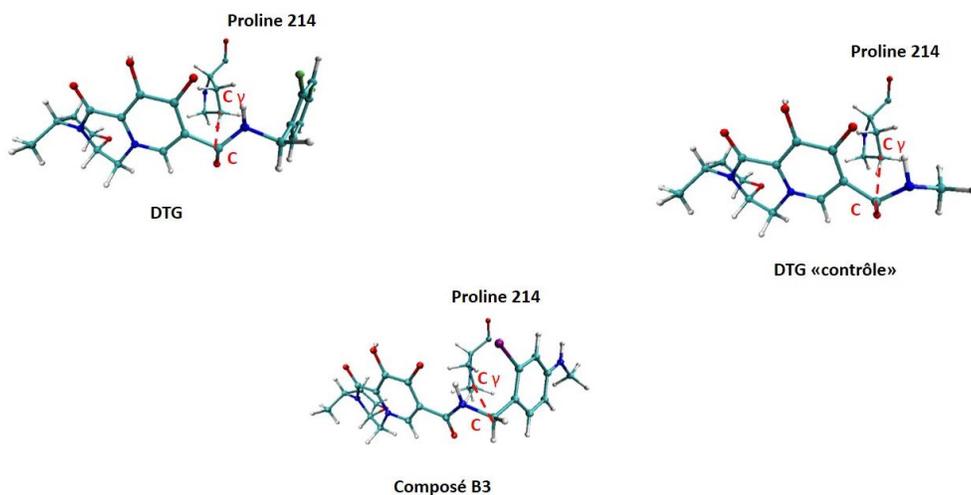


Figure 134 : Représentation des positions de départ dans nos simulations des dérivés DTG, B3 et DTG « contrôle » par rapport à la Proline 145, avec en gras et rouge les atomes dont les distances ont été suivies au cours des simulations. La ligne rouge pointillée indique la distance entre eux. Les composés A2 et B2 ayant des positions de départ similaires au composé B3 ne sont pas représentés dans la figure.

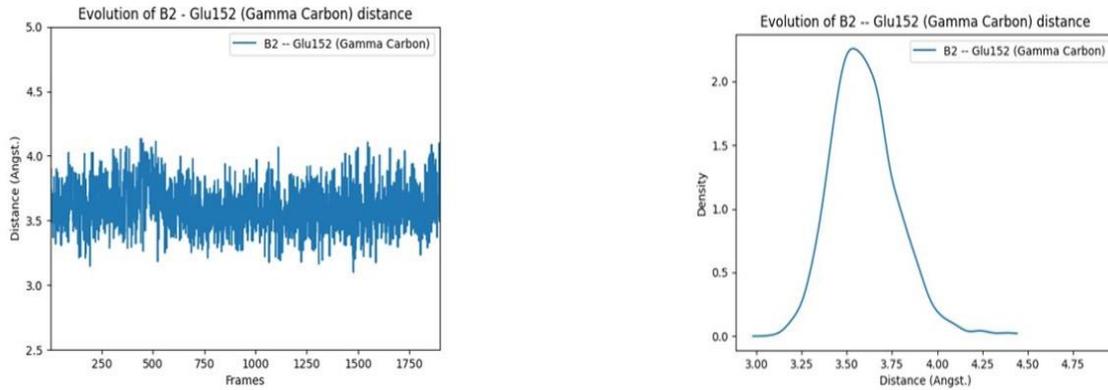


Figure 135 : Graphique représentant l'évolution de la distance entre B2 et Glu152 (correspondant au Glu221 du PFV) (à gauche) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons le graphique de la densité des pauses en fonction des distances obtenues.

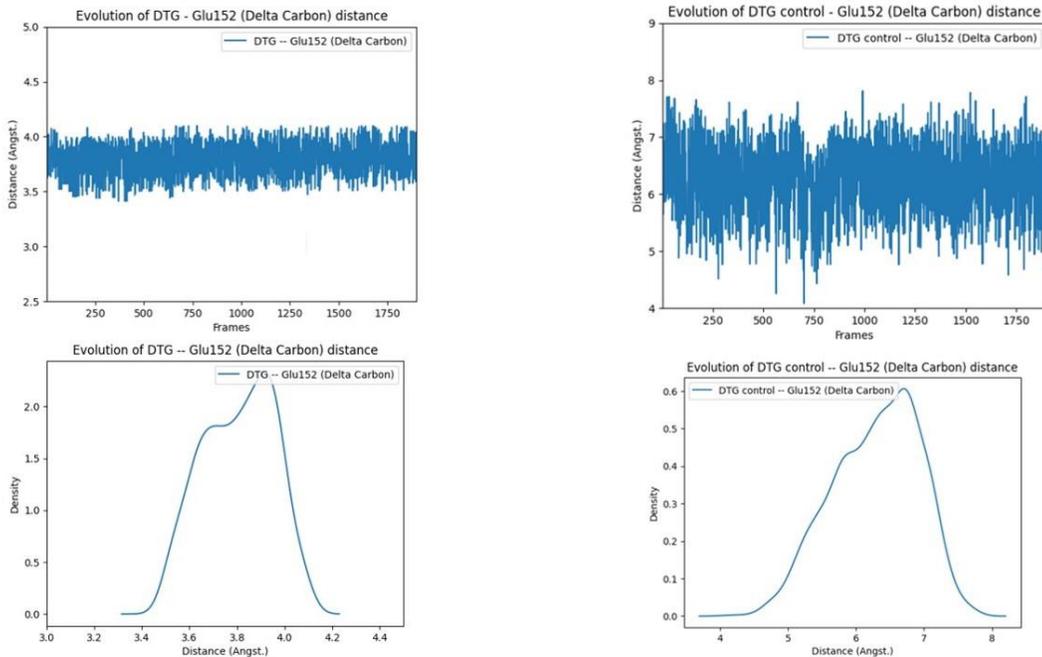


Figure 136: Graphiques représentant l'évolution de la distance entre DTG et Glu152 (correspondant à Glu221 pour le PFV) (à gauche) et entre le DTG « contrôle » et Glu152 (à droite) en fonction des pauses de dynamique moléculaire (Frames). Nous trouvons aussi les graphiques de la densité des pauses en fonction des distances obtenues (à gauche en bas dans le cas du DTG avec Glu152 et à droite en bas dans le cas du DTG «contrôle » avec Glu152.

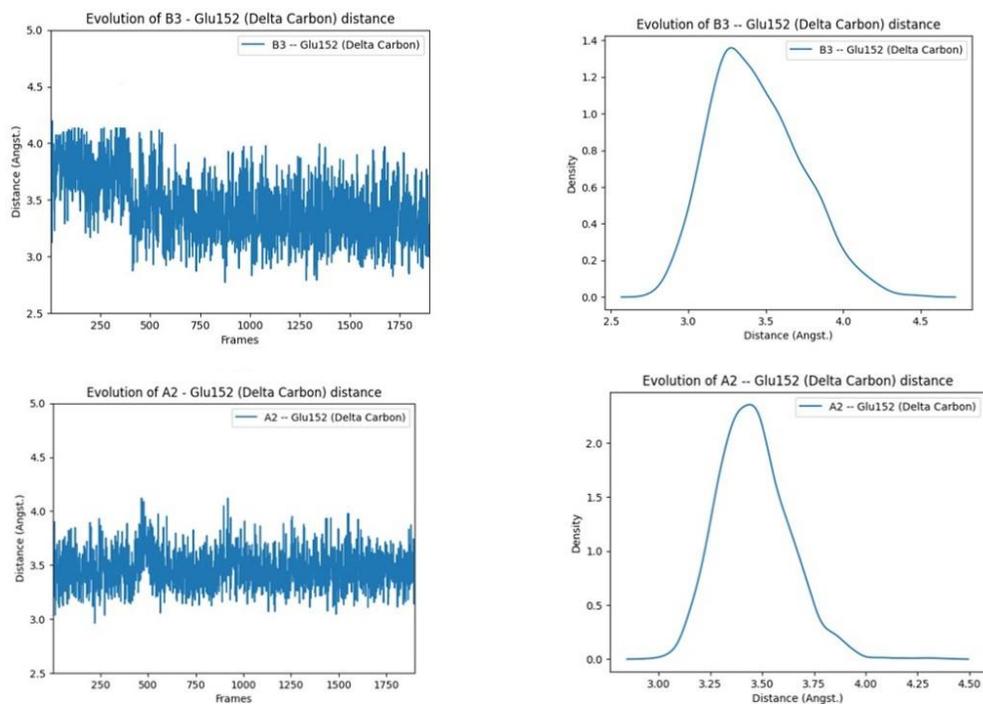


Figure 137 : Graphiques représentant l'évolution de la distance entre B3 et Glu152 (correspondant à Glu221 pour le PFV) (à gauche en haut) et entre A2 et Glu152 (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). Nous trouvons aussi les graphiques de la densité des pauses en fonction des distances obtenues (à droite en haut dans le cas de B3 avec Glu152 et à droite en bas dans le cas d'A2 avec Glu152).

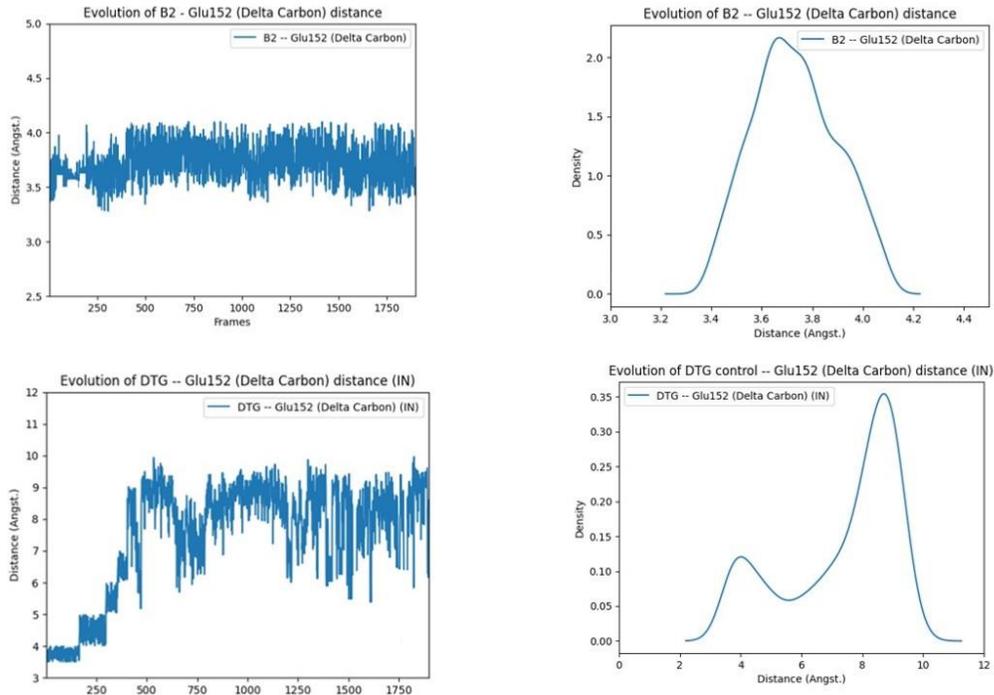


Figure 138 : Graphiques représentant l'évolution de la distance entre B2 et Glu152 (correspondant à Glu221 pour le PFV) (en haut à gauche) et entre le DTG et Glu152 dans le complexe Intégrase-DTG (en bas à gauche) en fonction des pauses de dynamique moléculaire (Frames). Nous trouvons aussi les graphiques de la densité des pauses en fonction des distances obtenues (à droite en haut dans le cas de B2 avec Glu152 et à droite en bas dans le cas du DTG avec Glu152 au sein du complexe Intégrase-DTG).

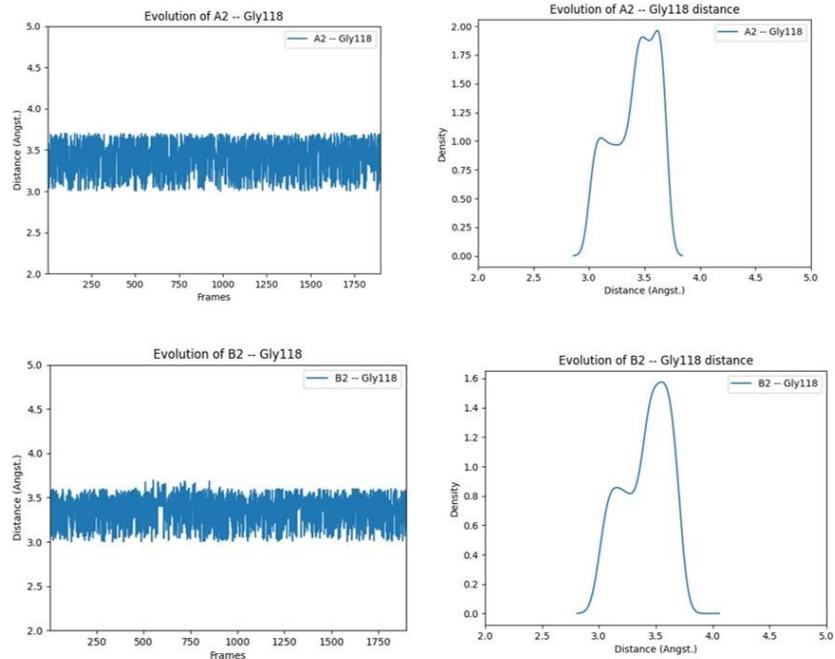


Figure 139 : Graphiques représentant l'évolution de la distance entre A2 et Gly118 (correspondant à Gly187 pour le PFV) (en haut à gauche) et entre B2 et Gly118 (en bas à gauche) en fonction des pauses de dynamique moléculaire (Frames). Nous trouvons aussi à droite les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas d'A2 avec Gly118 et en bas dans le cas de B2 avec Gly118).

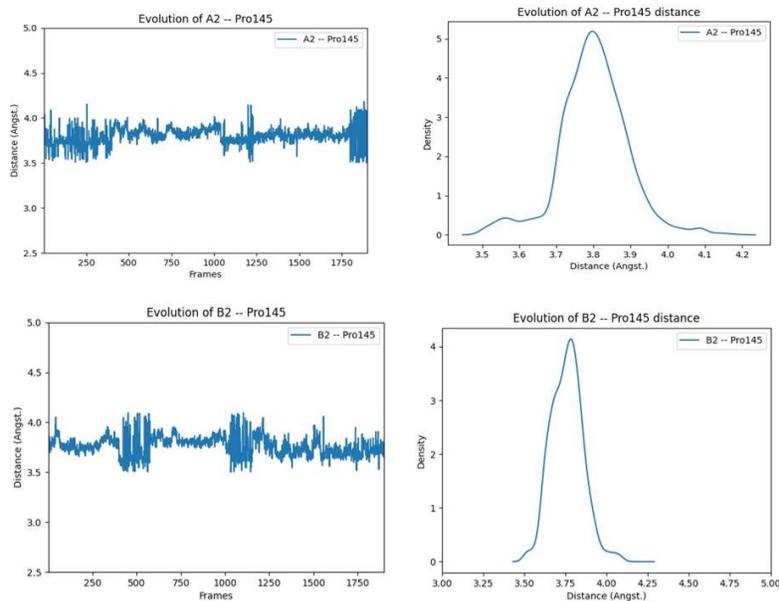


Figure 140 : Graphiques représentant l'évolution de la distance entre A2 et Pro145 (correspondant à Pro214 pour le PFV) (en haut à gauche) et entre B2 et Pro145 (en bas à gauche) en fonction des pauses de dynamique moléculaire (Frames). Nous trouvons aussi à droite les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas d'A2 avec Pro145 et en bas dans le cas de B2 avec Pro145).

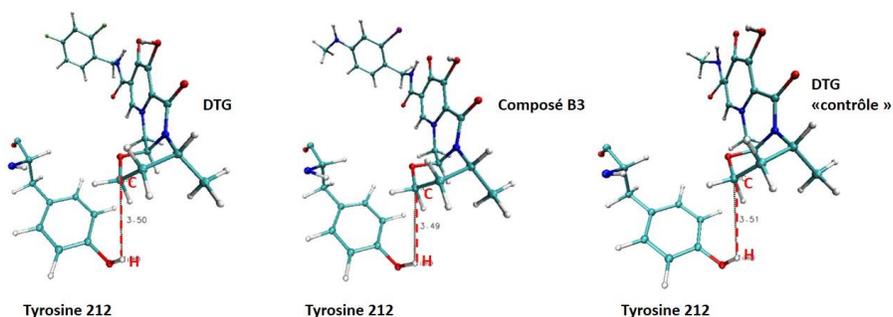


Figure 141 : Représentation des positions de départ dans nos simulations des dérivés DTG, B3 et DTG « contrôle » par rapport à la Tyrosine 212 du Prototype Foamy Virus (PFV) (correspondant à la Tyrosine 143 du VIH-1), avec en gras et rouge les atomes effectuant des contacts de vdW entre eux. La ligne rouge pointillée indique la distance entre eux. Les composés A2 et B2 ayant des positions de départ similaires au composé B3 ne sont pas représentés dans la figure.

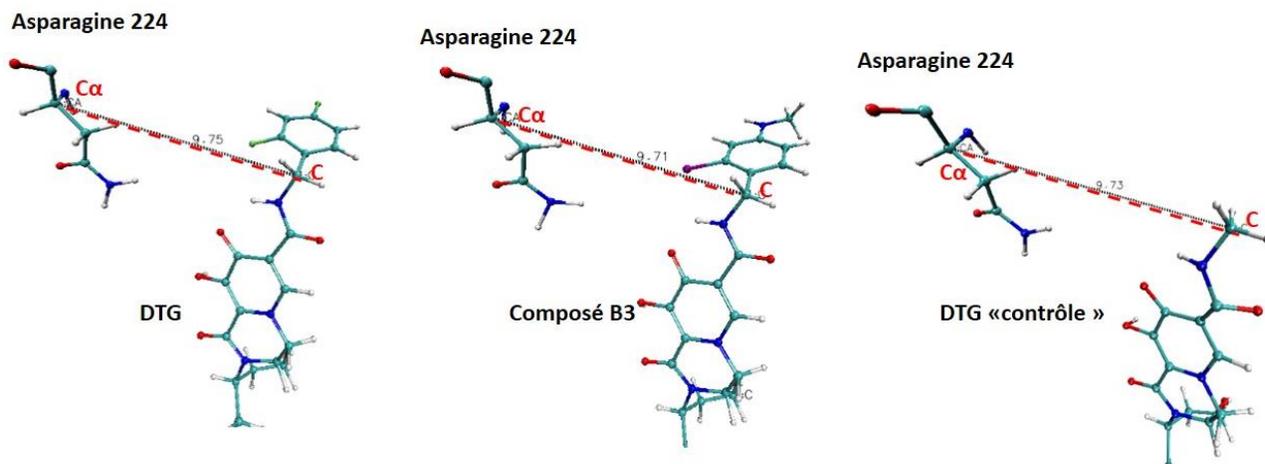


Figure 142 : Représentation des positions de départ dans nos simulations des dérivés DTG, B3 et DTG « contrôle » par rapport à l'Asparagine 155, avec en gras et rouge les atomes dont les distances ont été suivies au cours des simulations. La ligne rouge pointillée indique la distance entre eux. Les composés A2 et B2 ayant des positions de départ similaires au composé B3 ne sont pas représentés dans la figure.

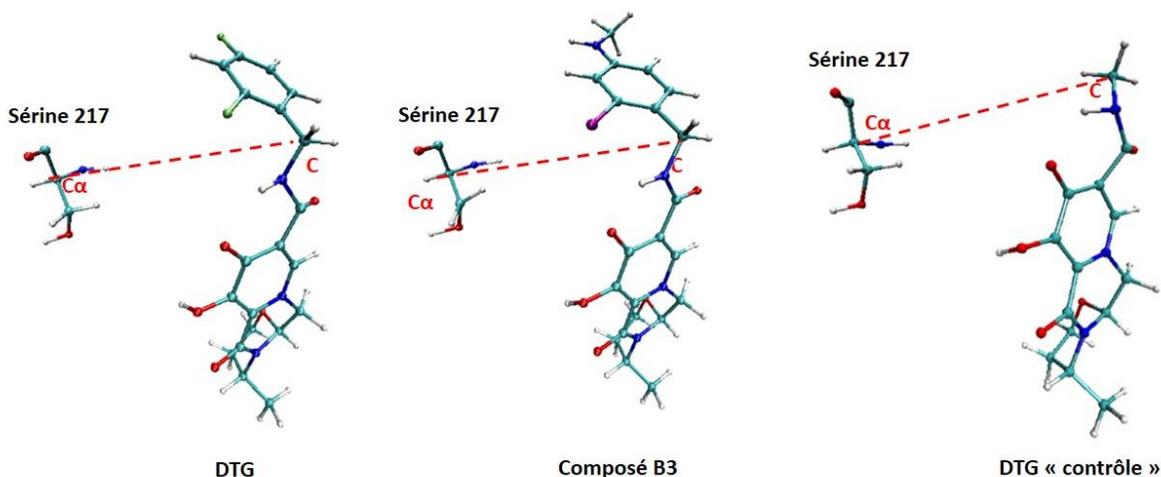


Figure 143 : Représentation des positions de départ dans nos simulations des dérivés DTG, B3 et DTG « contrôle » par rapport à la Sérine 217 (correspondant à la Glutamine 148), avec en gras et rouge les atomes dont les distances ont été suivies au cours des simulations. La ligne rouge pointillée indique la distance entre eux. Les composés A2 et B2 ayant des positions de départ similaires au composé B3 ne sont pas représentés dans la figure.

Suivi de la liaison entre la Glycine140 (correspondant à la Sérine209 chez le PFV) et nos inhibiteurs ; DTG, B3, A2, B2 et DTG « contrôle ».

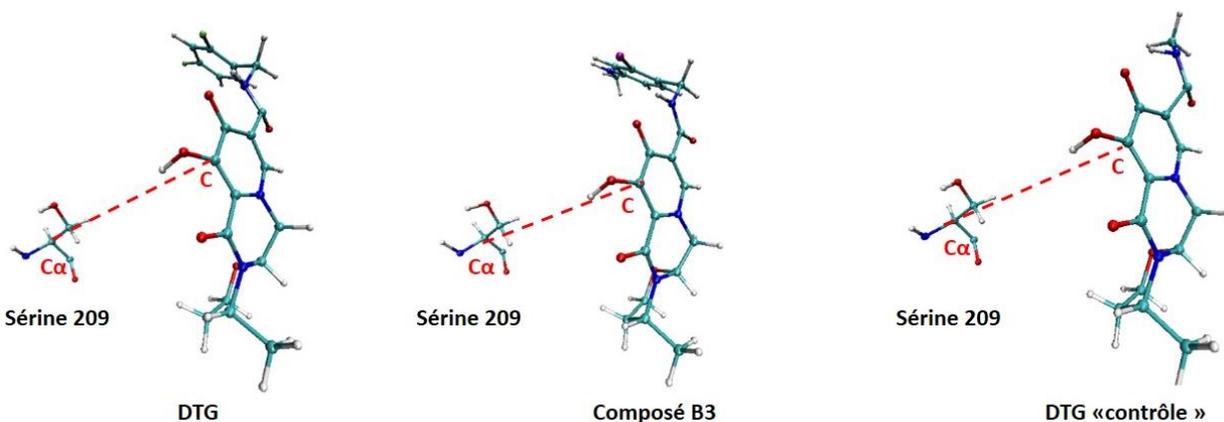


Figure 144 : Représentation des positions de départ dans nos simulations des dérivés DTG, B3 et DTG « contrôle » par rapport à la Sérine 209 chez le PFV (correspondant à Gly140 chez le VIH-1) avec en gras et rouge les atomes dont les distances ont été suivies au cours des simulations. La ligne rouge pointillée indique la distance entre eux. Les composés A2 et B2 ayant des positions de départ similaires au composé B3 ne sont pas représentés dans la figure.

La Glycine 140 (Gly140) du VIH-1 est non conservée chez le PFV ; son équivalent est la Sérine 209 (S209) dans l'intégrase de ce dernier. Elle se trouve au niveau de la boucle en amont de l'hélice α_4 et subit une mutation de résistance secondaire (G140 S/A) accompagnant la mutation de résistance primaire Q148 H/R/K qui fait face au RAL et à l'EVG. Nous avons suivi la variation de la distance entre le C α de S209 et le carbone en face de lui au niveau des inhibiteurs (Figure 144). Les résultats obtenus sont illustrés dans les graphiques ci-dessous.

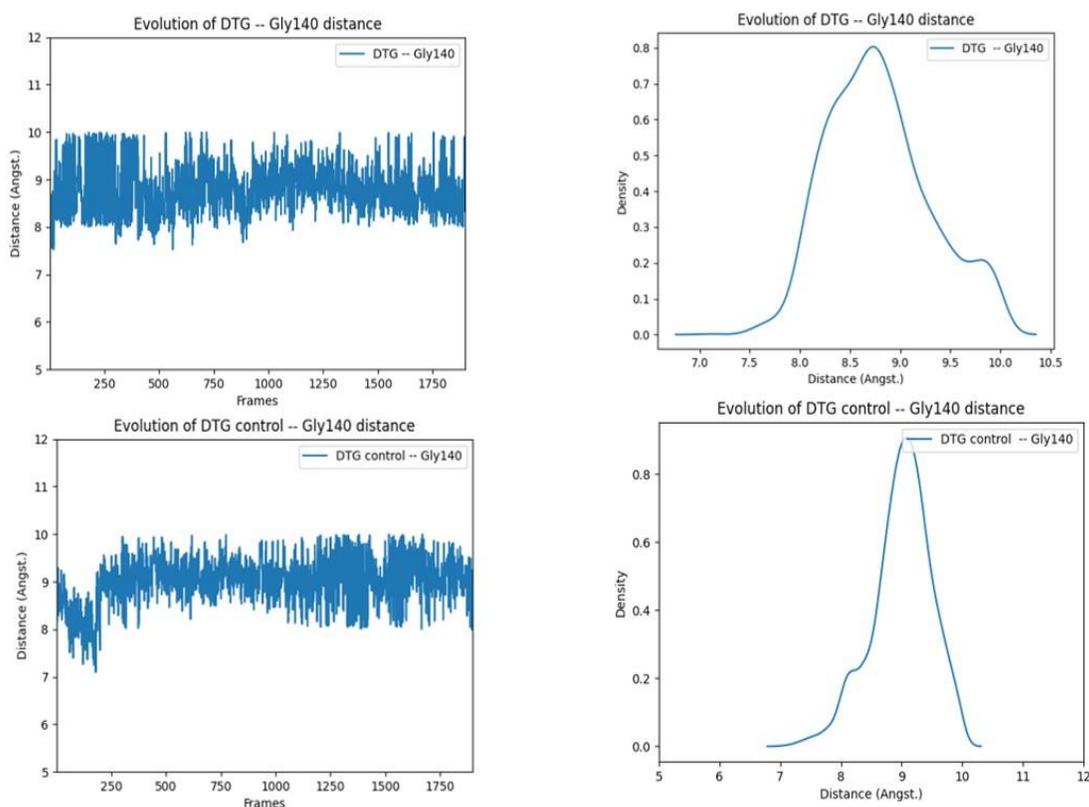


Figure 145 : Graphiques représentant l'évolution de la distance entre le DTG et la Sérine 209 du PFV correspondant à la Glycine 140 (Gly140) du VIH-1 (à gauche en haut) et entre le DTG « contrôle » et la Sérine 209 (Gly140 du VIH-1) (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas du DTG avec Gly140 et en bas dans le cas du DTG « contrôle » en présence de Gly140).

La distance entre le DTG et Ser209 est stable au cours des simulations, variant entre 8 Å et 10 Å avec des valeurs majoritaires aux alentours de 8,7 Å (Figure 145) alors

que la distance DTG «contrôle» - Ser209 est aux alentours de 8 Å au début des simulations puis augmente pour atteindre des valeurs approximatives de 9,5 Å après 50 ns de simulation. La majorité des trajectoires présentent dans ce cas des distances de 9 Å (Figure 145).

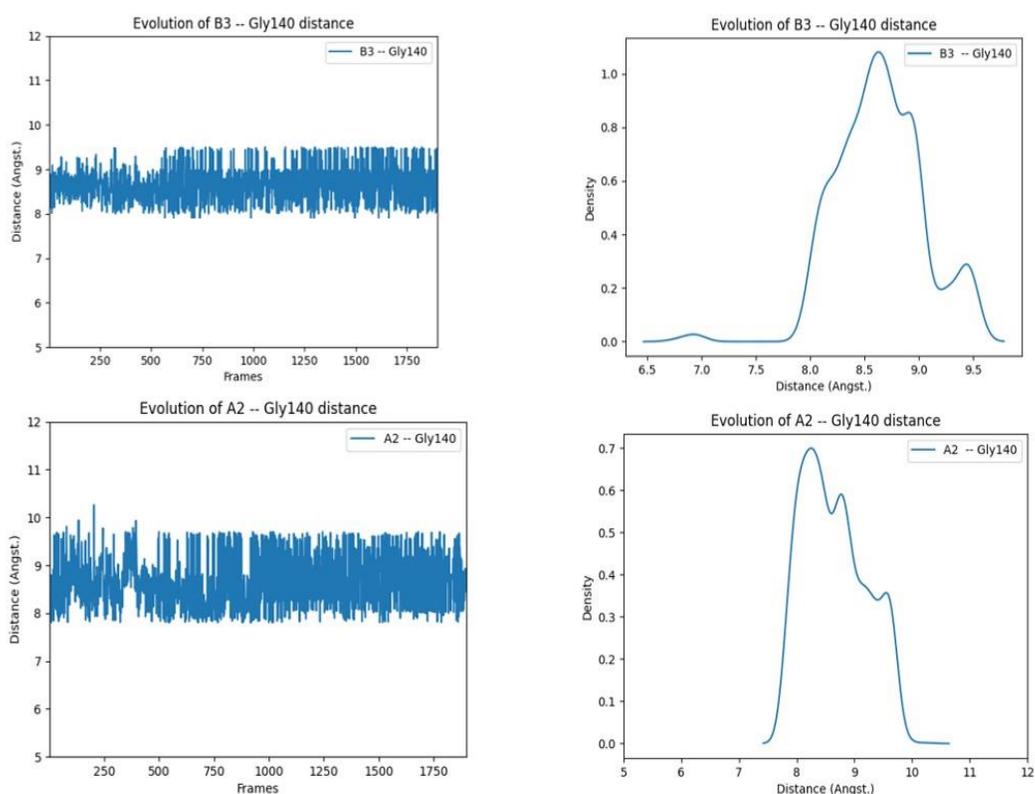


Figure 146 : Graphiques représentant l'évolution de la distance entre B3 et la Sérine 209 du PFV correspondant à la Glycine 140 (Gly140) du VIH-1 (à gauche en haut) et A2 et la Sérine 209 (Gly140 du VIH-1) (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B3 avec Gly140 et en bas dans le cas d'A2 en présence de Gly140).

L'analyse des graphiques de la figure 146 indique des variations similaires de la distance entre B3 et Ser209 et A2 et le même résidu ; ces distances sont stables variant entre 8 Å et 9,5 Å pour B3 et 8 Å et 10 Å pour A2. Le graphique en haut de la figure 147 indique de même des distances stables entre B2 et Ser209 variant entre 8 Å et 10 Å avec des valeurs majoritairement égales à 8,2-8,3 Å. Dans le cas

du DTG en absence d'ADN viral, sa distance avec Ser209 commence entre 8 Å et 10 Å puis après 50 ns de simulation puis augmente pour atteindre des valeurs situées entre 12 Å et 14 Å (Figure 147, graphique en bas à gauche). La majorité des distances dans ce cas sont égales à 12 Å (Figure 147, graphique en bas à droite).

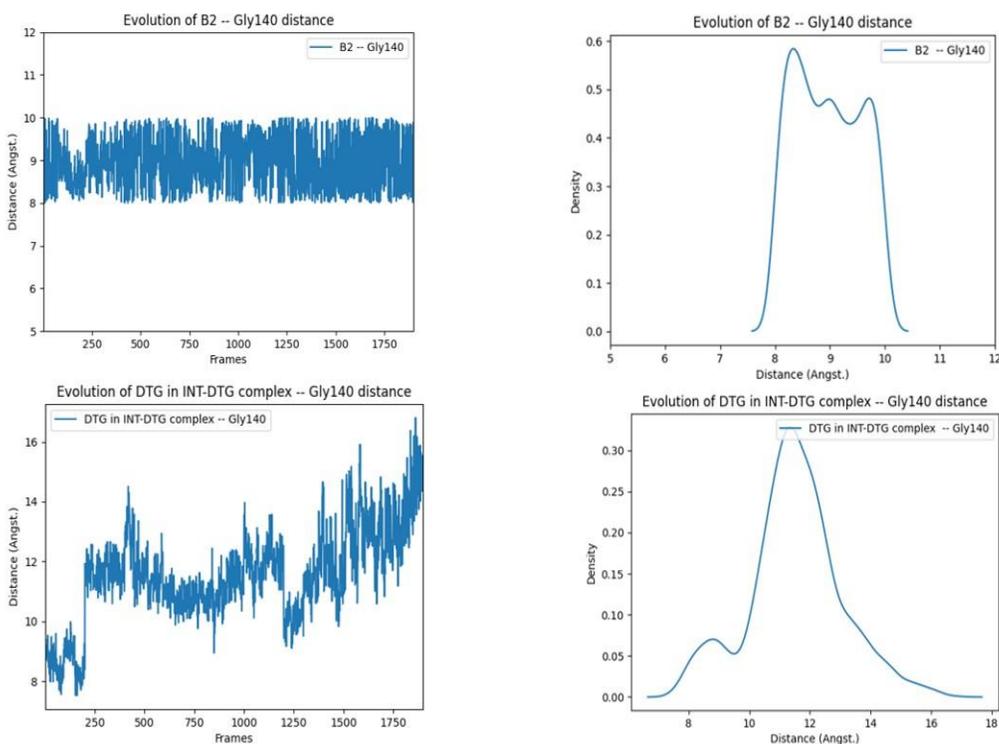


Figure 147 : Graphiques représentant l'évolution de la distance entre B2 et la Sérine 209 du PFV correspondant à la Glycine 140 (Gly140) du VIH-1 (à gauche en haut) et DTG et la Sérine 209 (Gly140 du VIH-1) dans le complexe Intégrase - DTG (à gauche en bas) en fonction des pauses de dynamique moléculaire (Frames). A droite, nous trouvons les graphiques de la densité des pauses en fonction des distances obtenues (en haut dans le cas de B2 avec Gly140 et en bas dans le cas du DTG en présence de Gly140 dans le complexe Intégrase – DTG).