



HAL
open science

Rational design of bioactive molecules for antibiotic purposes

Julia Reville Imbernon

► **To cite this version:**

Julia Reville Imbernon. Rational design of bioactive molecules for antibiotic purposes. *Cheminformatics*. Université de Strasbourg, 2023. English. NNT : 2023STRAF037 . tel-04700126

HAL Id: tel-04700126

<https://theses.hal.science/tel-04700126v1>

Submitted on 17 Sep 2024

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.

ÉCOLE DOCTORALE DES SCIENCES CHIMIQUES

Laboratoire d'Innovation Thérapeutique – UMR 7200

THÈSE

présentée par :

REVILLO IMBERNON Julia

soutenue le : 15 septembre 2023

pour obtenir le grade de : **Docteur de l'Université de Strasbourg**

Discipline/ Spécialité : Chimie informatique

Rational design of bioactive molecules for antibiotic purposes

THÈSE dirigée par :

Mme KELLENBERGER Esther

Professeur, Université de Strasbourg

RAPPORTEURS :

M MORELLI Xavier

Directeur de recherches, CNRS, Marseille

M CHALOIN Laurent

Directeur de recherches, CNRS, Montpellier

AUTRES MEMBRES DU JURY :

MITEVA Maria A.

Directeur de recherches, INSERM, Paris

MEMBRES INVITES :

DEJAEGERE Annick

Professeur, Université de Strasbourg

Acknowledgments

Une thèse doctorale nécessite de la passion, de l'implication, de l'analyse de soi, du contrôle de soi et de ses émotions, de la force pour persister dans les moments où rien ne marche... Bref, de l'endurance. Moi, cette endurance, je l'ai eu grâce aux personnes qui m'ont accompagné sur ce chemin et que j'aimerais remercier aujourd'hui ici.

Je me rappelle qu'avant de commencer j'étais convaincue que je n'en ferais jamais une. Cela me semblait un travail titanesque pour lequel je n'étais pas prête et ne le serai pas. Esther a su apercevoir cette insécurité propre aux novices et a commencé à me guider dans ce travail qui allait finalement aboutir dans ce manuscrit. Esther, je veux te remercier de m'avoir écouté quand les doutes me submergeaient, de m'avoir laissé la liberté de choisir sur quoi je voulais travailler et comment le faire, tout en veillant à ce que je ne me perde sur le chemin.

Guillaume, merci d'avoir été mon point d'appui. Merci de comprendre les non-dits, de regarder et de voir en regardant, de prendre soin. Je vais garder dans un petit coin de mon cœur toutes nos pauses café qui m'ont tellement fait rire et qui ont permis que je te découvre et qu'on devienne amis. Merci pour ton ouverture d'esprit, d'écouter mes délires et d'en rigoler. Merci de m'avoir écouté, aidé et appuyé. Merci d'être aussi têtu que moi et d'avoir poussé loin nos discussions.

Lauri, Luca, François, Teodora, and everyone in the lab. It has been a real pleasure to grow with all of you. I have loved our coffee breaks, lunch times, pizza nights, and cinema sessions. Thank you for such nice moments, interesting conversations and for being a team in which we all take care of each other with an open heart. I couldn't have wished for a better team/friends. Je ne veux pas oublier de mentionner toutes les personnes qui font partie de l'UMR ici, car, même si on était enfermés dans notre petit placard de chimoinformaticiens, ça a été un plaisir de partager des histoires, des rires et des mots dans les couloirs. Je me suis toujours sentie à l'aise et heureuse de travailler dans un labo où il y a tellement de bienveillance. Petite mention spéciale pour l'équipe escalade qui a grandi et diminué au cours des années (Masha, Carla, Antoine, Paula, Lauri, Romain, Gaëlle, Manu) mais qui a tant libéré mon esprit.

Alex, Paula, Blanca, Léa, Marie, Marie, Tagir, Pablo, Louis, Antoine, Roman, Javi, Lauri, Anniina, Franck, Hugo, Deniz, Louison, Elisa, Gary, Masha et teeeeeellement d'autres personnes qui m'ont accompagné dans cette aventure à Strasbourg, merci. Pour tous les moments qu'on a passés ensemble, tous les mojitos du Chariot, toutes les manifestations

auxquelles on est allés pour prendre un peu le soleil, toutes les discussions, moments joyeux, moments moins joyeux, partages, soirées, marchés de Noël, raclettes, bières, vins, randonnées, voyages, blagues, conversations, danses, amour. Je vous aime beaucoup.

Sajou, Camille, Theepik, Susu, Sutha, Gnanam, vous avez été ma famille à Strasbourg. Merci pour tant d'amour.

Bueno, me parece que ha llegado el momento de agradecer a los que siempre habéis estado a mi lado, me habéis apoyado, entendido (o no, pero habéis confiado), querido y mucho más. Mamá, Matu, gracias. Os quiero. Siempre habéis sido los pilares que han hecho que todos los momentos complicados sean más llevaderos. Gracias por querernos.

Gracias también a vosotros Raul, Sarza, siempre.

Sajan, merci pour tout.

Contents page

Acknowledgments.....	1
Contents page.....	3
List of tables.....	5
List of figures.....	6
List of acronyms	8
Résumé en français	11
I Introduction.....	25
I.1 General introduction to bacteria.....	25
I.2 General introduction to antibiotics and antibacterial resistance	25
I.3 Epidemiology	28
I.4 Politics to address AMR crisis	29
I.5 Organization of this thesis manuscript.....	30
II Rational design of aminoglycoside derivatives	31
II.1 The chemical structure of aminoglycosides.....	31
II.2 Structural analysis of neomycin B and kanamycin A binding AMEs and bacterial ribosomal RNA	33
II.2.1 Introduction.....	33
II.2.2 Materials and methods	36
II.2.3 Results.....	38
II.2.4 Example on amikacin.....	49
II.3 Modelling the binding of neomycin B and its synthetic derivative HL_171 into five AMEs	52
II.3.1 Introduction.....	52
II.3.2 Materials and methods	54
II.3.3 Results.....	60
II.3.4 Discussion	103
II.3.5 Conclusion	106
III β -clamp as an antibacterial target	107
III.1 Introduction.....	107
III.1.1 β -clamp structure	107
III.1.2 β -clamp as an antibacterial target	111
III.1.3 Starting point.....	113

III.1.4	Aim of the work	113
III.2	Fragment libraries and binding mode analysis	114
III.2.1	Introduction.....	114
III.2.2	Comprehensive analysis of commercial fragment libraries.....	115
III.2.3	Analysis of the Diamond Light Source fragment database	125
III.2.4	Selection of DLS fragment references	127
III.2.5	Conclusion	129
III.3	Virtual screening for the identification of small molecules inhibitors of <i>E. coli</i> β -clamp	130
III.3.1	Introduction.....	130
III.3.2	Materials and Methods.....	130
III.3.3	Results.....	133
III.3.4	Discussion	141
III.3.5	Conclusion and perspectives.....	142
IV	Conclusion	145
V	Bibliography	147
VI	Annexes.....	161
VI.1	Interactions between AME and AGA during the MD simulations.....	161
VI.2	Results of ROCS test	176
VI.3	SI Comprehensive analysis of commercial fragment libraries	182
VI.4	Mining the Protein Data Bank to inspire fragment library design.....	189
VI.5	SI Mining the Protein Data Bank to inspire fragment library design	200
	Résumé.....	204
	Abstract.....	204

List of tables

Table II.2.3-1 Studied PDB entries containing neomycin B or kanamycin A in complex with an AME or RNA.	40
Table II.3.3-1 Presence of AMEs in and sensitivity to neomycin B of clinical <i>E. coli</i> strains.	62
Table II.3.3-2 Presence of AMEs in and sensitivity to neomycin B of clinical <i>S. aureus</i> strains.	62
Table II.3.3-3 Summary of PDB entries of the selected AMEs or homologs.	63
Table II.3.5-1 MD results for neomycin B binding modes (consistent or non-consistent during the simulations) and HL_171 binding modes (similar or different from that of neomycin B).	106
Table III.2.2-1 Summary of libraries classification	116
Table III.2.2-2 Comparison of scaffolds properties and frequency.	120

List of figures

Figure II.2.1-1 Timeline of the discovery of new antibiotic classes and other events including the appearance of resistances.	26
Figure II.2.1-2 Sites of action and potential mechanisms of bacterial resistance to antimicrobial agents.	27
Figure II.2.1-3 Infographic about the bacterial interconnection between human, animals and the environment and about the spread of bacterial resistance.	28
Figure II.2.1-1 Number of deaths attributable to AMR in 2014 and prediction for 2050.	29
Figure II.2.1-1 Structures of the aminocyclitol cores and AGAs numbering,	32
Figure II.2.1-1 Action of AMEs on aminoglycosides.	34
Figure II.2.3-1 2D and 3D representation of neomycin B	43
Figure II.2.3-2 Comparison of neomycin B conformers dihedral angles.	45
Figure II.2.3-3 2D and 3D representation of kanamycin A RX-structures	47
Figure II.2.3-4 Comparison of kanamycin A conformers dihedral angles.	48
Figure II.2.4-1 2D and 3D representation of amikacin in RX-structures in complex with rRNA.	50
Figure II.3.1-1 Strategies developed on neomycin B to overcome AMEs resistance. Image from article ²⁴	52
Figure II.3.1-2 Structures of (A) neomycin B and (B) HL_171. In pink the synthetic modifications.	54
Figure II.3.2-1 Structure of the scaffold constraint.	56
Figure II.3.3-1 Protocol to model and study the AME/AGA complexes.	61
Figure II.3.3-2 Different representations of the active site of APH(3')-IIIa.	65
Figure II.3.3-3 B-factor values (\AA^2) of neomycin B atoms and its binding site in the PDB entry 2B0Q.	67
Figure II.3.3-4 Three-dimensional view of APH(3')-IIIa and key distances measurement.	68
Figure II.3.3-5 RMSD plots of neomycin B and HL_171 in complex with APH(3')-IIIa.	69
Figure II.3.3-6 Interaction profiles between neomycin B and APH(3')-IIIa during the first production.	70
Figure II.3.3-7 Stable interactions formed between APH(3')-IIIa and neomycin B or HL_171 during the MD simulations and present in more than 70% of the frames of all three productions.	71
Figure II.3.3-8 Neomycin B and HL_171 dihedral angles during the MD simulations.	72
Figure II.3.3-9 3D structures of the last MD simulations frame and key distances of APH(3')-IIIa.	73
Figure II.3.3-10 3D view of APH(3')-IIa active site (PDB ID: 1ND4).	74
Figure II.3.3-11 B-factor values (\AA^2) of kanamycin A atoms and its binding site in the PDB entry 1ND4.	75
Figure II.3.3-12 Three-dimensional view of APH(3')-IIa and key distances measurement.	76
Figure II.3.3-13 RMSD plots of neomycin B and HL_171 in complex with APH(3')-IIa.	77
Figure II.3.3-14 Stable interactions formed between APH(3')-IIa and neomycin B or HL_171 during the MD simulations and present in more than 70% of the frames of all three productions.	78
Figure II.3.3-15 Neomycin B and HL_171 dihedral angles during the MD simulations.	79
Figure II.3.3-16 3D structures of the last MD simulations frame and key distances of APH(3')-IIa.	81
Figure II.3.3-17 3D view of APH(3')-Ia active site (PDB ID: 4EJ7).	83
Figure II.3.3-18 B-factor values (\AA^2) of kanamycin A atoms and its binding site in the PDB entry 4GKH.	84
Figure II.3.3-19 Three-dimensional view of APH(3')-Ia and key distances measurement.	86
Figure II.3.3-20 RMSD plots of neomycin B and HL_171 in complex with APH(3')-Ia.	86

Figure II.3.3-21 Stable interactions formed between APH(3')-Ia and neomycin B or HL_171 during the MD simulations and present in more than 70% of the frames of all three productions.	87
Figure II.3.3-22 Neomycin B and HL_171 dihedral angles during the MD simulations.	88
Figure II.3.3-23 3D structures of the last MD simulations frame and key distances of APH(3')-Ia	89
Figure II.3.3-24 Different representations of the active site of ANT(4')-Ia.	90
Figure II.3.3-25 B-factor values (\AA^2) of neomycin B atoms and its binding site to chain B of PDB 6NML.	92
Figure II.3.3-26 Three-dimensional view of ANT(4')-Ia and key distances measurement	93
Figure II.3.3-27 RMSD plots of neomycin B and HL_171 in complex with ANT(4')-Ia.	93
Figure II.3.3-28 Stable interactions formed between ANT(4')-Ia and neomycin B or HL_171 during the MD simulations and present in more than 70% of the frames of all three productions.	94
Figure II.3.3-29 Neomycin B and HL_171 dihedral angles during the MD simulations.	95
Figure II.3.3-30 3D structures of the last MD simulations frame and key distances of ANT(4')-Ia	96
Figure II.3.3-31 Different representations of the active site of AAC(3)-IIa.	97
Figure II.3.3-32 B-factor values (\AA^2) of kanamycin A atoms and its binding site in the PDB entry 6O5U.	99
Figure II.3.3-33 Three-dimensional view of AAC(3)-Ia and key distances measurement	100
Figure II.3.3-34 RMSD plots of neomycin B and HL_171 in complex with AAC(3)-IIa.	100
Figure II.3.3-35 Stable interactions formed between AAC(3)-IIa and neomycin B or HL_171 during the MD simulations and present in more than 70% of the frames of all three productions.	101
Figure II.3.3-36 Neomycin B and HL_171 dihedral angles during the MD simulations.	101
Figure II.3.3-37 3D structures of the last MD simulations frame and key distances of AAC(3)-IIa	102
Figure III.1.1-1 Crystal structure of <i>E. coli</i> β -clamp.	108
Figure III.1.1-2 DNA sliding clamps of <i>E. coli</i> β -clamp in complex with a peptide of DNA polymerase IV	109
Figure III.1.1-3 3D view of <i>E. coli</i> CBM.	110
Figure III.1.2-1 Structure of β -clamp inhibitors	112
Figure III.2.2-1 Physico-chemical properties of 512 284 commercial fragments.	117
Figure III.2.2-2 Diversity of fragment scaffolds.	119
Figure III.2.2-3 Spatial distribution of the fragments in the GTM.	121
Figure III.2.2-4 GTM landscape of small-sized fragments subsets.	122
Figure III.2.2-5 Three-dimensional properties.	123
Figure III.2.3-1 Spatial distribution of the fragments in GTM Landscapes	126
Figure III.2.4-1 3D view of the crystallographic fragments surfaces bound to the <i>E.coli</i> B-clamp dimer	128
Figure III.2.4-2 Interaction fingerprint of each crystallographic fragment with the amino acids of <i>E.coli</i> B-clamp.	128
Figure III.2.4-3 Workflow employed for the selection of the query fragments.	129
Figure III.3.3-1 VS protocol.	133
Figure III.3.3-2 Results of ROCS test.	135
Figure III.3.3-3 His175 protonation states	137
Figure III.3.3-4 LID process.	138
Figure III.3.3-5 Properties of the 120 selected molecules.	139
Figure III.3.3-6 Molecular surfaces of the selected molecules bound to the CBM.	140

List of acronyms

2-DOS: 2-deoxystreptamine	CYP: Cytochrome P50
2D: Two-dimensional	DLS: Diamond Light Source
3D: Three-dimensional	DNA: Deoxyribonucleic acid
<i>A. baumannii</i> : Acinetobacter baumannii	<i>E. coli</i> : Escherichia coli
ADP: Adenine Diphosphate	ECFP: Extended Connectivity Fingerprints
ATCC: Americal Type Culture Collection	FAO: Food and Agriculture Organization
ATP: Adenine Triphosphate	FBDD: Fragment-Based Drug Design
AGA: Aminoglycoside	GAFF2: General Amber Force Field 2
AAC: Aminoglycoside N-Acetyltransferase	GB/VI: Generalized-Born/Volume Integral
AcCoA: Acetyl Coenzyme A	GCN5: General Control Non-repressed 5
ADMET: Absorption, Distribution, Metabolism, Excretion and Toxicity	GOLD: Genetic Optimization for Ligand Docking
ADP: Adenosine diphosphate	GPU: Graphics Processing Unit
AM1-BCC: Austin Model 1 with bond charge correction	Gram+: Gram-positive
AMBER: Assisted Model Building with Energy Refinement	Gram-: Gram-negative
AME: Aminoglycoside-Modifying Enzyme	GTM: Generative Topographic Map
AMR: Antimicrobial resistance	HABA: Hydroxy-aminobutyric acid
ANT: Aminoglycoside O-Nucleotidyltransferase	HB: Hydrogen bond
APH: Aminoglycoside O-Phosphotransferase	HBA: Hydrogen bond acceptor
ATP: Adenosine triphosphate	HBD: Hydrogen bond donor
<i>B. subtilis</i> : Bacillus subtilis	HID: Histidine with a protonated δ -N
BLAST: Basic Local Alignment Search Tool	HIE: Histidine with a protonated ϵ -N
BM: Binding mode	HIP: Histidine with both δ -N and ϵ -N protonated
BS: Binding site	HIV DIS: Human Immunodeficiency Virus Dimerization Initiation Site
CBM: β -clamp binding motif	IC ₅₀ : Half maximal Inhibitory Concentration
CN: Chimiothèque nationale	ID: Identity
CNRS: Centre Nationale de Recherches Scientifiques	IFP: Interaction Fingerprint
CoA: Coenzyme A	IN2P3: Institut National de Physique Nucléaire et de Physique des Particules
	KAN: Kanamycin A
	K _i : Inhibitor constant
	LID: Local Interaction Density

LLL: Gentamicin C1
logP: Partition logarithm
MIC: Minimum Inhibitory Concentration
M: Molar
M. tuberculosis: Mycobacterium tuberculosis
MD: Molecular dynamics
MDR: Multidrug resistance
MOE: Molecular Operating Environment
MW: Molecular Weight
NMY: Neomycin B
NRings: Number of rings
NRot: Number of rotatable bonds
ns: nanoseconds
P. aeruginosa: Pseudomonas aeruginosa
P16: 16-residue C-terminal peptide from PolIV
PCNA: Proliferating Cell Nuclear Antigen
PCR: Polymerase Chain Reaction
PDB: Protein Data Bank
Pmemd: Particle Mesh Ewald MD
ps: picoseconds
PTC: Premature Termination Codon
PSA: Polar Surface Area
PTC: Premature Termination Codon
R&D: Research and development
RDKit: Rational Discovery Kit
RCSB: Research Collaboratory for Structural Bioinformatics
RMSD: Root Mean Square Deviation
RNA: Ribonucleic Acid
rRNA: ribosomal ribonucleic acid
Ro3: Rule of Three
ROCS: Rapid Overlay of Chemical Structures
RSCC: Real-Space Correlation Coefficient
S. aureus: Staphylococcus aureus
Tc: Tanimoto coefficient
TIP3P: Transferable Intermolecular Potential with 3 Points
UN: United Nations
VIH: Human Immunodeficiency Virus
VS: Virtual Screening
WOAH: World Organization for Animal Health
WHO: World Health Organization

Résumé en français

Introduction

De nos jours, la résistance des bactéries aux médicaments est un enjeu majeur pour l'humanité en termes de santé publique, d'économie et d'écologie. Cette résistance résulte de la combinaison de différents facteurs. Pour commencer, la surconsommation d'antibiotiques a conduit à l'apparition de souches résistantes. Ensuite, la mondialisation des échanges a favorisé leur diffusion. Enfin, l'industrie pharmaceutique s'est détournée du développement de nouveaux antibiotiques depuis la fin des années 1980, et par conséquent l'arsenal thérapeutique actuel n'est pas suffisant pour faire face au développement des résistances bactériennes.¹

En 2015, l'organisation mondiale de la santé (OMS), l'organisation pour l'alimentation et l'agriculture (FAO) et l'organisation mondiale de la santé animale (OIE) ont mis en place un plan d'action mondial contre la résistance microbienne.² Un an plus tard, les Nations unies (ONU) a reconnu cette lutte comme une priorité mondiale puis, en 2017, l'OMS a publié une liste des bactéries à combattre en priorité pour orienter la recherche et le développement (R&D) de nouveaux antibiotiques.³ Deux stratégies sont envisageables : Améliorer les propriétés des molécules déjà sur le marché et relancer la production de nouveaux antibiotiques en identifiant de nouvelles cibles moléculaires bactériennes.

Ma thèse s'inscrit dans deux axes de recherche et développement de nouveaux antibiotiques et comporte deux parties : la première a pour thème l'optimisation de la néomycine, un antibiotique de la classe des aminoglycosides (AGA); et la seconde la recherche de molécules antibiotiques ayant un mode d'action original, capables d'inhiber la croissance bactérienne et l'émergence de résistance en ciblant l'anneau de réplication. Ma contribution porte sur la conception rationnelle de molécules à visée antibiotique à l'aide de méthodes chémoinformatiques et bioinformatiques, et plus particulièrement la prédiction de mode d'interaction moléculaire entre protéine et ligand, et le criblage virtuel.

La néomycine est un antibiotique qui a d'excellentes propriétés antibactériennes mais qui subit les résistances bactériennes et qui peut être ototoxique et hépatotoxique. En tant qu'AGA, il est composé d'un noyau 2-déoxystreptamine, relié à plusieurs oses. Plus précisément, il est membre de la classe possédant un groupe déoxystreptamine 4,5 disubstituée, constitué par 4

cycles. La néomycine, et plus généralement les AGA, ont une fonction bactéricide en ciblant l'acide ribonucléique ribosomique (ARNr) pour inhiber la traduction bactérienne.⁴ Bien que de nombreux mécanismes de résistance aux AGA existent, le plus commun est celui qui est dû aux enzymes modificateurs d'aminoglycosides (AME) (Figure 1A). Comme leur nom l'indique, ces enzymes sont capables de modifier chimiquement les aminoglycosides. Trois types ont été identifiés au cours des années : aminoglycoside N-Acetyltransferase (AAC), aminoglycoside O-phosphotransferase (APH) et aminoglycoside O-Nucleotidyltransferase (ANT).⁵

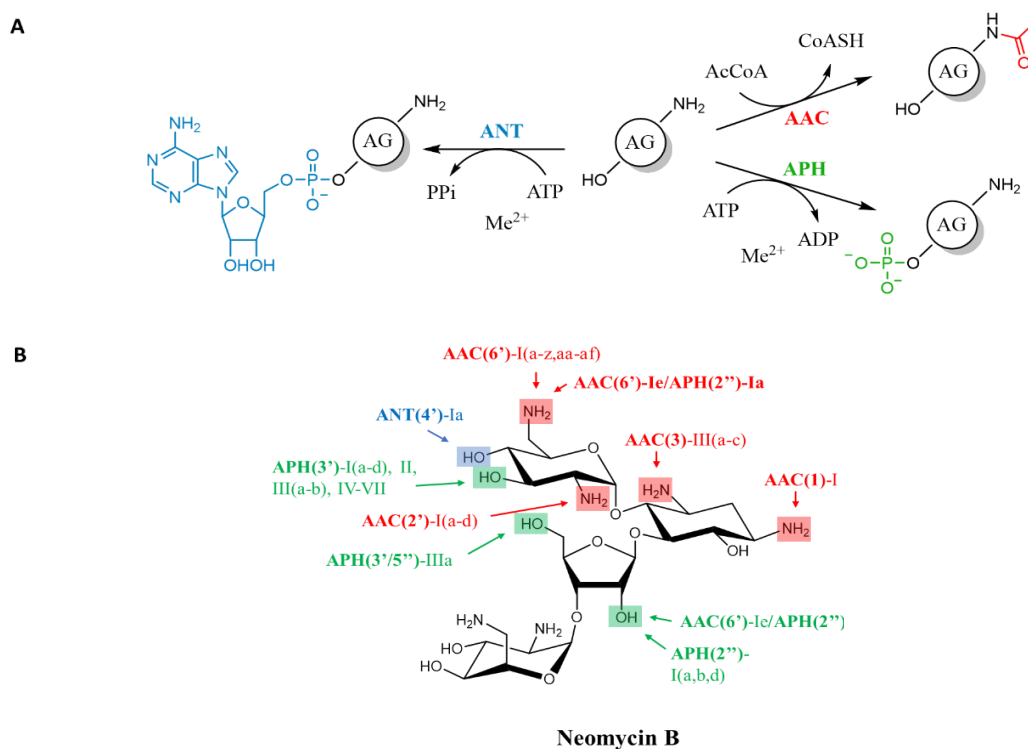


Figure 1 Action des AMEs sur les aminoglycosides. A) Reactions catalysées par les trois types d'AMEs. B) Profil de modification enzymatique de la néomycine B.

Les anneaux de réplication bactériens, ou β -clamp, confèrent une grande processivité aux ADN polymérase répliquatives et sont des plateformes de liaison pour diverses enzymes impliquées dans le métabolisme de l'ADN. L'équipe du Dr. Burnouf a identifié et validé cette cible pour le développement de nouveaux antibiotiques à l'aide de petits peptides qui se lient à son site hydrophobe (Figure 2).^{6,7} Un criblage par fragments a été réalisé et a confirmé la droguabilité du site, tout en fournissant des indications pour la conception de pharmacophores. Notre équipe a eu pour tâche d'exploiter les données structurales pour aller vers un potentiel candidat médicament.

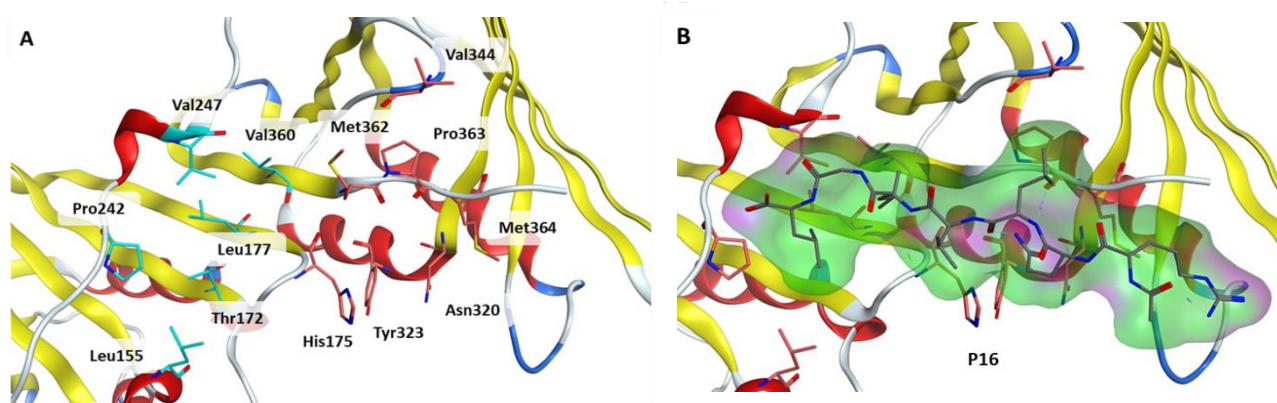


Figure 2 Vue 3D du (A) Site de liaison de l'anneau de réplication bactérien d'*E. coli* (B) peptide P16 lié au site de liaison de l'anneau de réplication bactérien d'*E. coli*. (PDB: 1OK7)

Résultats et discussion

Partie 1 : optimisation de la néomycine

Pour ce projet, nous sommes en collaboration avec trois laboratoires de disciplines différentes. L'équipe de chimistes du Pr. Weibel (LaSyROC, UMR 7177, CNRS-University of Strasbourg), qui porte le projet et explore les relations structure-activité des dérivés. A ce jour, 33 dérivés ont été synthétisés, avec des modifications sur les groupements rattachés aux atomes C-5'' et/ou N-6'. L'équipe de bactériologie du Dr. Prévost (Virulence bactérienne précoce, UR 7290, CNRS-University of Strasbourg) a recherché les AME dans les souches d'*E. coli* et *S. aureus* résistantes et a testé l'activité des dérivés sur la croissance et la survie des bactéries, ainsi que leur toxicité cellulaire. L'équipe de biophysique du Dr. Ennifar (Architecture et réactivité de l'ARN, UPR 9002, CNRS-University of Strasbourg) a résolu par cristallographie aux rayons X la structure de dérivés en complexe avec l'ARNr. Au sein de l'équipe de modélisation du Pr. Kellenberger, je travaille à rationaliser l'optimisation en modélisant l'interaction des dérivés de néomycine avec les AME et cherchant à comprendre l'impact des modifications sur le métabolisme des AGAs.

Dans un premier temps, nous avons comparé les structures d'AGAs qui sont disponibles dans la Protein Data Bank (PDB) et présents à la fois dans plusieurs AMEs et dans le site A de l'ARNr. Notre étude a porté sur la néomycine B et la kanamycine A, respectivement représentatifs des AGAs à trois et quatre cycles. Notre objectif est de décrire de manière exhaustive les modes de liaison d'un même AGA, afin de repérer dans ces AGAs les positions clés pour la reconnaissance et l'action des AMEs et pour l'activité antibactérienne. Au total,

nous avons détecté les interactions formées dans 92 complexes AGAs-AMEs et dans 12 AGAs-ARN (Figure 3). Nous avons observé que les complexes sont principalement dominés par des liaisons polaires, dont 60% avec des acides glutamiques et des acides aspartiques. Les molécules d'eau participent aux interactions à hauteur de 14% dans les complexes néomycine B-AME, et de 19% dans les complexes kanamycine A-AME. Certaines des molécules d'eau sont conservées dans les différents complexes. Les modes d'interactions adoptés par les deux AGAs dans les AMEs sont similaires à ceux identifiés dans les complexes d'ARN. Par ailleurs, nous avons calculé les angles dièdres de chacun des AGAs pour comparer leurs conformères dans les différents complexes. Nous avons ainsi mis en évidence la conservation de la conformation des cycles I et II parmi tous les complexes, AGA-AME comme AGA-ARN. Dans le cas de la néomycine B, la flexibilité observée parmi les conformères vient principalement des liaisons entre les cycles II et III, et les cycles III et IV.

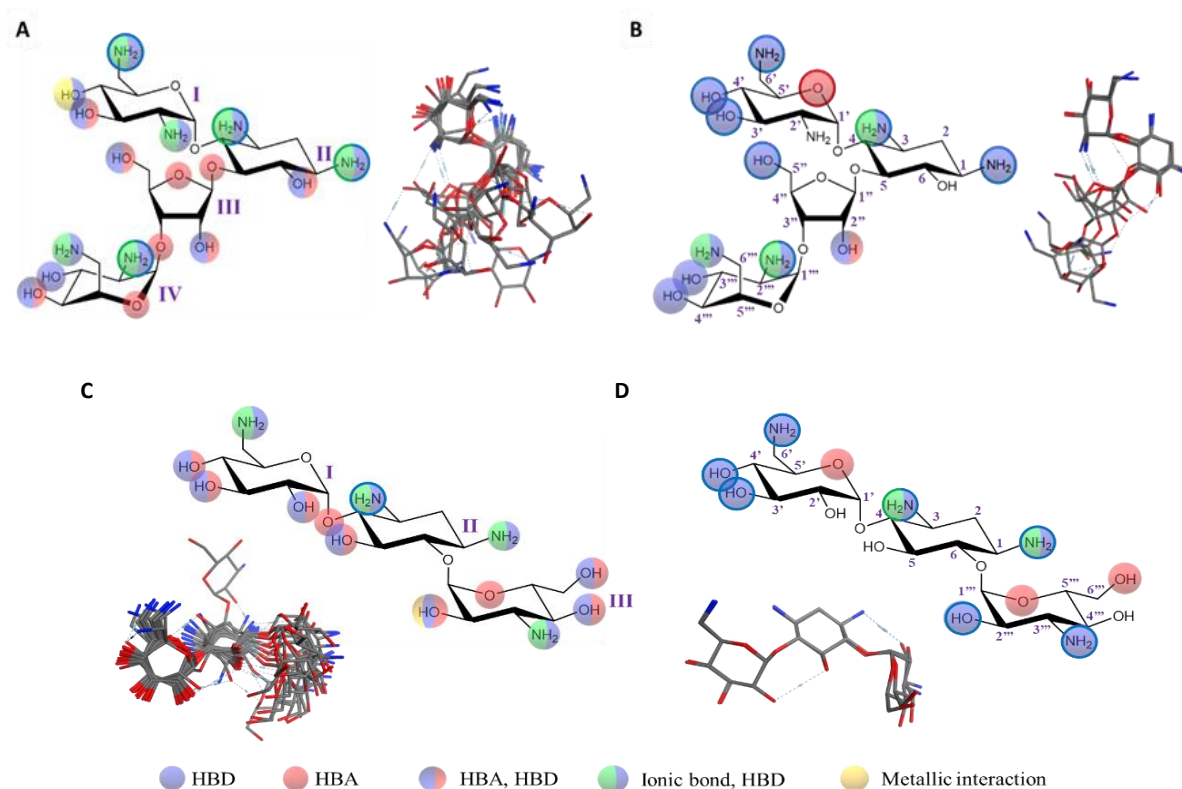


Figure 3 Représentations 2D et 3D de la néomycine B et kanamycine A (A) ensemble des interactions néomycine B-AME et conformères de la néomycine B superposés sur les cycles I et II. (B) ensemble des interactions néomycine B-ARN et conformères de la néomycine B superposés sur les cycles I et II. (C) ensemble des interactions kanamycine A-AME et conformères de la kanamycine A superposés sur les cycles I et II. (D) ensemble des interactions kanamycine A-ARN et conformères de la kanamycine A superposés sur les cycles I et II. Les interactions entourées ont été détectées dans toutes les structures cristallographiques.

Par la suite, nous avons mis en place une méthodologie pour prédire l'action des AME sur un dérivé synthétique AGA. Nous avons travaillé à la preuve de concept en choisissant la néomycine B et son dérivé HL_171 (Figure 4). HL_171 est un des nombreux dérivés synthétisés par l'équipe du Pr. Weibel. Il présente les meilleurs résultats de concentration minimale inhibitrice sur souches cliniques d'*E. coli* et *S. aureus* identifiées par le laboratoire du Dr. Gilles Prévost.

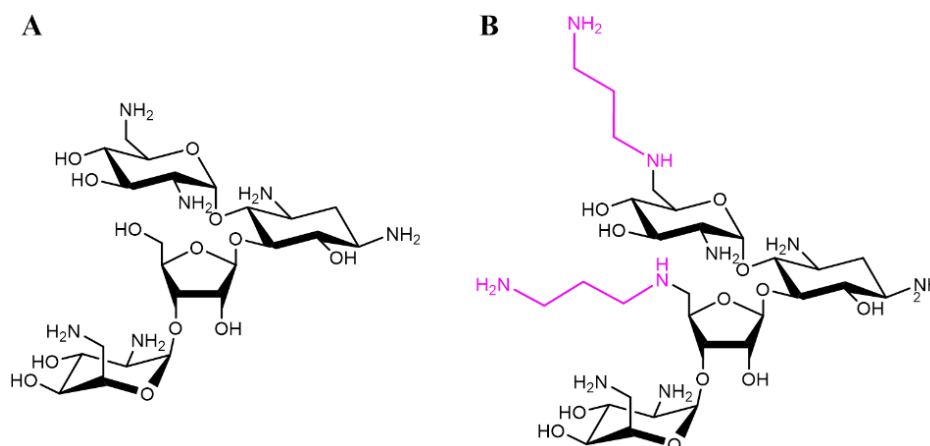


Figure 4 Structures de (A) néomycine B et (B) HL_171

Notre mode opératoire est résumé sur la figure 5 ci-contre. Nous avons modélisé par homologie cinq AMEs. La liaison de l'AGA dans le site actif de l'AME a été réalisée par docking sous contraintes. Les complexes ont été simulés par dynamique moléculaire. Pour chaque complexe,

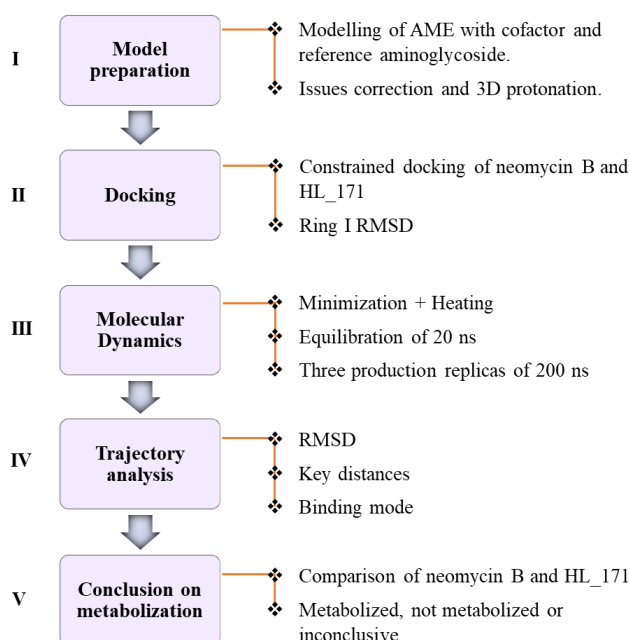


Figure 5 Protocole pour l'étude de l'influence des fonctionnalisations de la néomycine sur la métabolisation par AME.

trois trajectoires de 200ns ont été calculées. La stabilité des complexes a été étudiée à travers la déviation des coordonnées atomiques du site actif et de l'AGA. La néomycine a été prise comme référence pour déterminer un possible défaut dans la métabolisation de HL_171 par l'AME en utilisant les modes de liaison comme principaux indicateurs.

Les données obtenues visent à déterminer si les modifications appliquées à la néomycine sont favorables ou défavorables à la métabolisation. Dans l'ensemble, les modifications synthétiques apportées à la néomycine B ont eu un impact significatif sur deux AMEs (Tableau 1), ce qui est en accord avec les résultats biologiques.

Plus précisément, des interactions stables ont été observées entre APH(3')-IIIa et les chaînes synthétiques des positions C(6') et C(5''), ainsi qu'entre ANT(4')-Ia et la chaîne synthétique ajoutée à la position C(5''). Ces interactions sont responsables de la variation observée dans les modes de liaison et de leur impact subséquent sur la métabolisation de l'aminoglycoside.

En revanche, aucune interaction stable n'a été trouvée entre les chaînes synthétiques et APH(3')-Ia ou APH(3')-IIa, ce qui a mené à des modes de liaison similaires entre HL_171 et la néomycine B pour ces AMEs. Néanmoins, les chaînes synthétiques influent la configuration des conformères : l'angle ϕ de la liaison entre les cycles II et III présente des valeurs distinctes comparé à celles observées dans les structures cristallographiques.

Finalement, les résultats obtenus pour AAC(3)-IIa sont en accord avec la nature étroite du site de liaison de l'AME, empêchant ainsi l'activité catalytique sur les AGAs de la famille 2-DOS 4,5-disubstituée. De plus, ce modèle fournit des informations utiles sur les résultats potentiels liés aux AGAs qui sont anti-complémentaires au site de liaison.

	APH(3')-IIIa	APH(3')-IIa	APH(3')-Ia	ANT(4')-Ia	AAC(3)-IIa
Neomycin B	Consistent	Consistent	Consistent	Consistent	Inconstant
HL_171	Different BM	Similar BM	Inconclusive	Different BM	Different BM

Tableau 1 Résultats de simulation de dynamiques moléculaires pour les modes de liaison

En conclusion, grâce à la construction de modèles tridimensionnels des complexes et à l'étude de la stabilité et des modes de liaison par le biais de simulations de dynamique moléculaire, nous avons réussi à prédire avec succès l'impact des modifications de la néomycine B sur la reconnaissance par les AME et à mettre en évidence le comportement unique de chaque

complexe. De plus, nos simulations ont fourni des informations précieuses pour les futures modifications, mettant particulièrement en avant l'importance des positions C(1) et C(2'). Ces découvertes soulignent également l'importance d'employer simultanément plusieurs stratégies de modification pour augmenter les chances de concevoir un AGA qui échappe à la reconnaissance par les AME.

Partie 2 : recherche de ligands de l'anneau de réplication bactérien par l'approche fragment

Dans un premier temps, nous avons voulu analyser la chimiothèque de fragments Diamond, qui a été utilisée pour le screening. Pour cela, et étant donné que l'approche par fragment n'est pas encore utilisée en routine dans notre laboratoire, nous avons fait deux études permettant de comprendre comment les bases des données de fragments sont conçues, et comment évaluer leur composition, diversité, tri-dimensionalité et versatilité.

Ainsi, nous avons comparé les fragments commercialisés par les vendeurs de chimiothèques. Nous avons donc collecté et analysé les molécules issues de 86 chimiothèques de 14 fournisseurs. A l'aide de descripteurs simples, nous avons caractérisé la complexité et la tridimensionnalité des fragments pris dans leur ensemble, ou considérés par type de chimiothèques, tel qu'il est défini par les fournisseurs (ex : inspirée par les produits naturels, ou encore optimisée pour un maximum de diversité). Nous avons également évalué la diversité de chaque chimiothèque en considérant les châssis chimiques. Enfin nous avons modélisé l'espace chimique défini par les fragments en utilisant la méthode «Generative Topographic Map» (GTM), de manière à poser un référentiel pour la comparaison de deux chimiothèques ou le positionnement de toute chimiothèque par rapport à l'ensemble des fragments. Nos conclusions, publiées en 2022 dans *RSC Med. Chem*, fournissent des lignes directrices pour le choix ou conception d'une chimiothèque dans le cadre d'un projet appliqué. La chimiothèque de fragments Diamond, projetée à travers la méthode GTM, se positionne en tant que chimiothèque diverse. Diamond couvre toutes les régions de la carte GTM de manière homogène, assurant une couverture représentative de l'ensemble de l'espace chimique de fragments.⁸

Plus récemment, nous avons complété notre étude des fragments par une recherche de fragments versatiles, c'est-à-dire qui soient capables de cibler plusieurs cibles, et qui par conséquent seraient susceptibles d'être des touches plus fréquentes lors de criblage. La sélectivité pour la cible serait travaillée ultérieurement, lors de l'élaboration du fragment en une molécule de type candidat médicament. Nous avons donc exploité la PDB pour identifier les fragments qui adaptent leur mode de liaison pour lier des cavités différentes (Figure 6). Notre méthode repose sur la comparaison des ensembles d'interactions moléculaires représentées par des graphes et sur la comparaison locale de structures de protéines représentées par des points de cavité et alignées avec l'outil ProCare développé au laboratoire. Ainsi, nous avons identifié 203 fragments versatiles dans la PDB. Leur analyse, réalisée en suivant la méthodologie développée sur les chimiothèques commerciales de fragments et publiées en 2023 dans *Front. Chem.*, a révélé des motifs rappelant les produits d'origine naturelle et a mis en évidence des châssis peu ou pas représentés dans les chimiothèques de fragments commerciales, mais aussi de châssis présents dans des principes actifs de médicaments approuvés par les autorités de santé.⁹

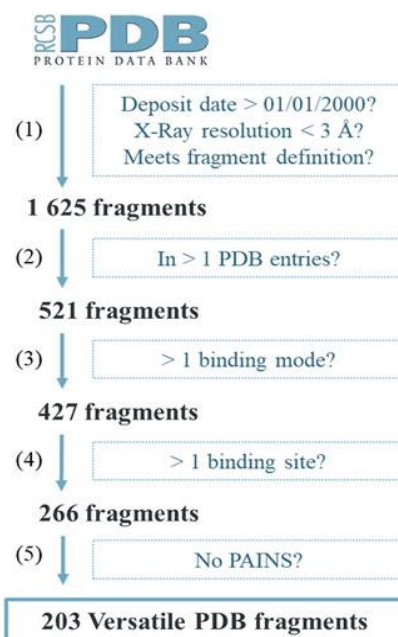


Figure 6 Protocole pour l'étude de la promiscuité des fragments

Le criblage par cristallographie aux rayons X de chimiothèque Diamond Light Source de fragments réalisé par nos collaborateurs a permis de fournir 73 structures contenant des fragments liés au sous-site I (Figure 7). Parmi ces fragments, 38 ont été sélectionnés comme références sur des critères d'interactions et d'aire de surface enfouie (Figure 8).

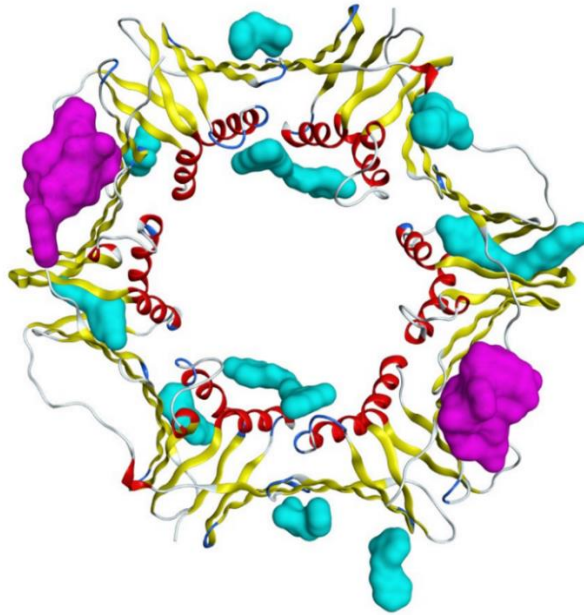


Figure 7 Vue 3D des fragments liés à l'anneau de réplication bactérien à l'issue du criblage. En rose les fragments liés au site actif.

A l'aide des données sur les fragments et les peptides liés à l'anneau bactérien, nous avons conçu une stratégie de criblage virtuel de chimiothèques de molécules « drug-like » (Figure 9). Nous avons tout d'abord filtré des chimiothèques Enamine et Chimiothèque Nationale (30.4 millions de molécules au total) pour ne conserver que les molécules dont une partie s'aligne sur la forme et les pharmacophores des fragments de référence. Les trois millions de molécules retenues après l'issue de cette étape ont été dockées dans le site ciblé. Les poses de docking ont été priorisées en fonction des modes de liaison en privilégiant ceux similaires aux peptides inhibiteurs de référence.

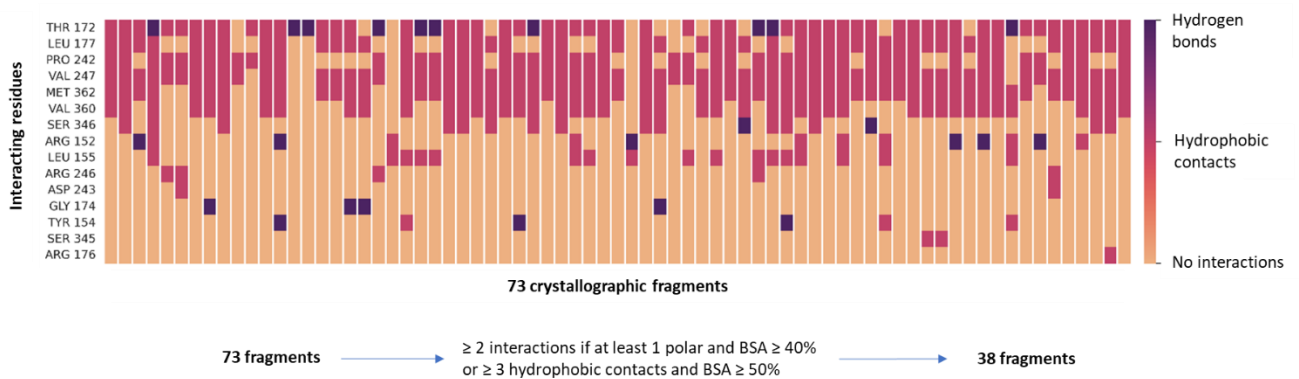


Figure 8 Sélection des fragments de référence sur des critères d'interactions et aire de surface enfouie (BSA)

L'examen visuel a permis d'écartier les molécules présentant des angles de torsion inhabituels, tout en favorisant celles montrant une bonne complémentarité avec la protéine, et en privilégiant également l'originalité et la diversité dans la sélection. Le filtrage a conclu avec la sélection de 120 molécules, dont 108 qui ont été synthétisées par Enamine et sont actuellement évaluées pour leur activité biologique par nos collaborateurs (liaison à l'anneau bactérien de *E. coli*, inhibition de la croissance bactérienne).

Les molécules sélectionnées dans leur ensemble couvrent une zone comprenant à la fois les sous-sites 1 et 2 du CBM, et s'étendaient au-delà (Figure 10). Prises individuellement, les molécules occupaient le CBM de différentes manières. Par exemple, la molécule individuelle mise en évidence sur la Figure 10A couvrait les deux sous-sites, tandis que la molécule représentée sur la Figure 10B n'occupait que le sous-site 1.

Les 120 molécules sélectionnées présentent des poids moléculaires allant de 295 à 500 et des logP allant de -1.6 à 6.0, indiquant un ensemble diversifié de molécules avec des caractéristiques hydrophiles et lipophiles. Sa comparaison à une liste de 287 antibiotiques présents dans la base de données DrugBank montre un ensemble divers aux molécules existantes dans le marché.

La suite du projet sera déterminée par les résultats biologiques obtenus lors de l'analyse des 108 composés. En cas de validation de touche, le projet s'orientera vers leur optimisation en chimie médicinale assistée par les modèles tridimensionnels des complexes ligand/protéine. Dans le cas contraire, nous reconsidérons les hypothèses faites sur la reconnaissance

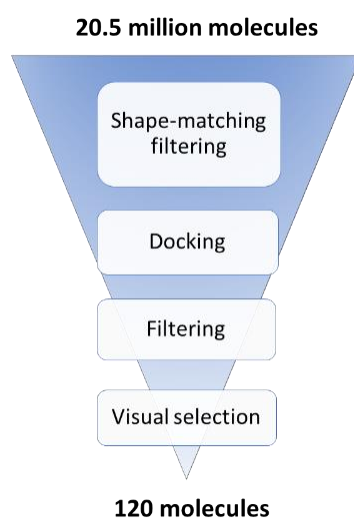


Figure 9 Protocole pour la sélection de potentiels candidats médicaments ciblant l'anneau de réplication bactérien

ligand/protéine à la lumière de la dynamique de la protéine et du rôle du solvant pour proposer une stratégie alternative basée sur la croissance de quelques fragments, préalablement sélectionnés selon des critères d'accessibilité synthétique, de propriétés pharmacophoriques et de liaison aux « hot-spots ».

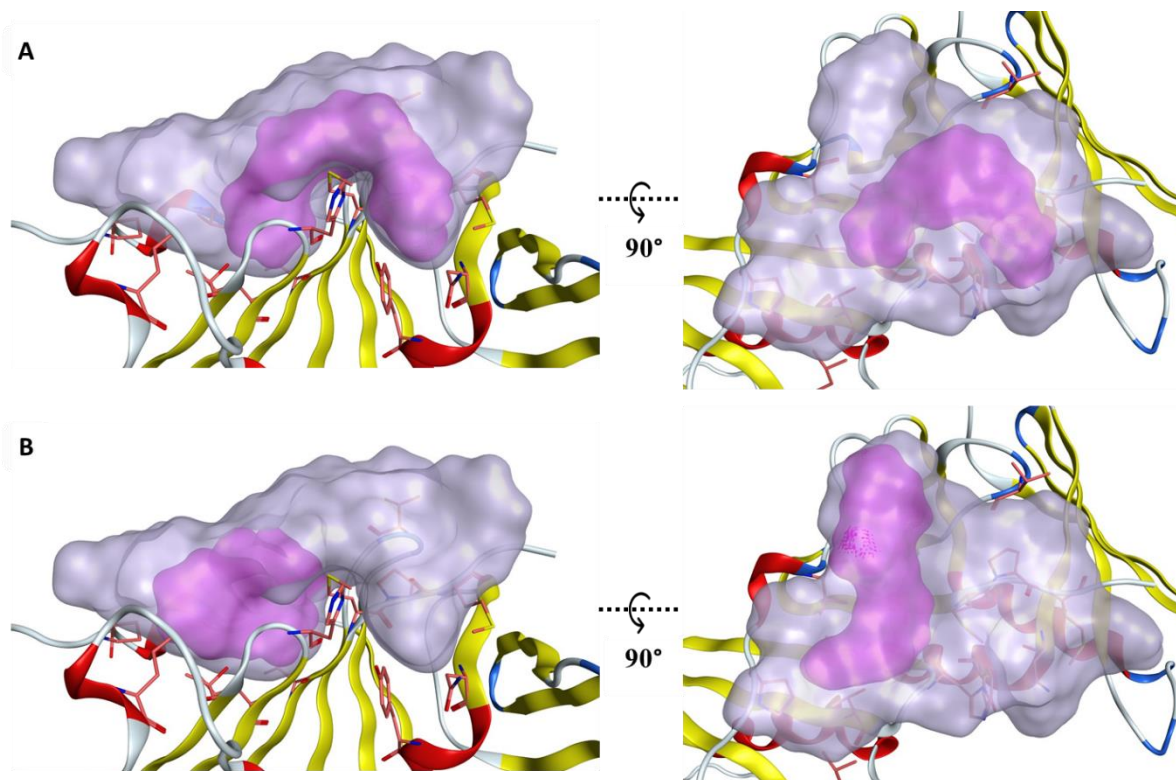


Figure 10 Surfaces moléculaires des molécules sélectionnées liées au CBM. La surface violette représente la zone couverte par l'ensemble des molécules. La surface rose représente la zone couverte par une molécule. (A) Molécule liée aux deux sous-sites. (B) Molécule liée au sous-site 1.

Conclusion générale

Pour conclure, la résistance des bactéries aux antibiotiques est un enjeu majeur pour l'humanité. Afin de lutter contre ce fléau, deux projets de recherche et de développement de nouveaux antibiotiques ont été mis en place dans le cadre de cette thèse : l'optimisation de la néomycine B et la recherche de ligands ciblant l'anneau de réplication bactérien. Ces projets ont été réalisés en collaboration avec plusieurs équipes de différentes disciplines.

Pour l'optimisation de la néomycine, une étude a été menée pour décrire de manière exhaustive les modes de liaison d'un même AGA, afin de repérer les positions clés pour la reconnaissance et l'action des AMEs et pour l'activité antibactérienne. Ensuite, une méthodologie a été mise en

place pour prédire l'action de cinq AMEs sur des dérivés synthétiques de la néomycine B, permettant ainsi de comprendre l'impact des modifications apportées. Les résultats obtenus mettent en évidence l'impact de la substitution en position C(6')-NH₂ et incitent à examiner cette position ainsi que C(1) et C(2') en tant que cibles potentielles pour inhiber l'action de certaines AMEs.

Dans le cadre de la recherche de ligands de l'anneau de réplication bactérien, une stratégie de criblage virtuel de chimiothèques de molécules "drug-like" a été mise en place. Un total de 120 molécules diverses a ainsi été sélectionné, et nous avons pu nous procurer 108. Des mesures de liaison et de croissance bactérienne sont en cours. Cette recherche s'est appuyée sur le mode de liaison de 73 fragments de la chimiothèque Diamond et de 5 peptides inhibiteurs. Afin d'explorer l'approche par fragments et d'évaluer la chimiothèque Diamond, deux analyses globales des données ont été menées sur des chimiothèques de fragments commerciales ainsi que sur les fragments versatiles présents dans la PDB.

Bibliographie :

- (1) Cook, M. A.; Wright, G. D. The Past, Present, and Future of Antibiotics. *Sci. Transl. Med.* **2022**, *14* (657), eabo7793.
- (2) World Health Organization. Global Action Plan on Antimicrobial Resistance. World Health Organization 2015.
- (3) L'OMS publie une liste de bactéries contre lesquelles il est urgent d'avoir de nouveaux antibiotiques <https://www.who.int/fr/news/item/27-02-2017-who-publishes-list-of-bacteria-for-which-new-antibiotics-are-urgently-needed> (accessed 2023 -05 -11).
- (4) Magnet, S.; Blanchard, J. S. Molecular Insights into Aminoglycoside Action and Resistance. *Chem. Rev.* **2005**, *105* (2), 477–498.
- (5) Ramirez, M. S.; Tolmasky, M. E. Aminoglycoside Modifying Enzymes. *Drug Resist. Updat. Rev. Comment. Antimicrob. Anticancer Chemother.* **2010**, *13* (6), 151–171.
- (6) Wolff, P.; Oliéric, V.; Briand, J. P.; Chaloin, O.; Dejaegere, A.; Dumas, P.; Ennifar, E.; Guichard, G.; Wagner, J.; Burnouf, D. Y. Structure-Based Design of Short Peptide Ligands Binding onto the E. Coli Processivity Ring. *J. Med. Chem.* **2011**, *54* (13), 4627–4637.
- (7) Wolff, P.; Amal, I.; Oliéric, V.; Chaloin, O.; Gygli, G.; Ennifar, E.; Lorber, B.; Guichard, G.; Wagner, J.; Dejaegere, A.; Burnouf, D. Y. Differential Modes of Peptide Binding onto Replicative Sliding Clamps from Various Bacterial Origins. *J. Med. Chem.* **2014**, *57* (18), 7565–7576.
- (8) Revillo Imbernon, J.; Jacquemard, C.; Bret, G.; Marcou, G.; Kellenberger, E. Comprehensive Analysis of Commercial Fragment Libraries. *RSC Med. Chem.* **2022**, *13*, 300-310
- (9) Revillo Imbernon, J.; Chiesa, L.; Kellenberger, E. Mining the Protein Data Bank to Inspire Fragment Library Design. *Front. Chem.* **2023**, *11*.

LISTE DES PRESENTATIONS

Communications nationales et internationales:

Orales:

06/09/2022 – Julia Reville Imbernon, Esther Kellenberger – Rationally modifying Neomycin, an antibiotic from the aminoglycosides family – Ecole d'été ITI InnoVec VIPR: Vectorization & Innovation against Pathogens Resistance; Strasbourg

23/05/2023 – Julia Reville Imbernon, Esther Kellenberger – Modelling of the binding of neomycin B and its derivative into Aminoglycoside-Modifying Enzymes – Journées du campus d'Illkirch (JCI); Illkirch-Graffenstaden

Par affiche:

29/09/2021-01/10/2021 – Julia Reville Imbernon, Célien Jacquemard, Guillaume Bret, Gilles Marcou, Esther Kellenberger – Comprehensive analysis of fragment libraries – Groupe de Graphisme et Modélisation Moléculaire – Société Française de ChémoInformatique (GGMM-SFCI); Lille (Prix Poster)

17/05/2022-20/05/2022 – Julia Reville Imbernon, Célien Jacquemard, Guillaume Bret, Gilles Marcou, Esther Kellenberger – Comprehensive analysis of commercial fragment libraries – Chemical Computing Group – User Group Meeting (CCG UGM) & Conference; Amsterdam

23/05/2022-24/05/2022 – Julia Reville Imbernon, Célien Jacquemard, Guillaume Bret, Gilles Marcou, Esther Kellenberger – Comprehensive analysis of commercial fragment libraries – Journées du Campus d'Illkirch

27/06/2022-01/07/2022 – Julia Reville Imbernon, Célien Jacquemard, Guillaume Bret, Gilles Marcou, Esther Kellenberger – Comprehensive analysis of commercial fragment libraries – Summer School of Chemoinformatics; Strasbourg

24/04/2023-25/04/2023 – Julia Reville Imbernon, Luca Chiesa, Esther Kellenberger – Mining the Protein Data Bank to inspire fragment library design - 9th French-Japanese Workshop on Computational Methods in Chemistry; Strasbourg

05/07/2023-07/07/2023 – Julia Reville Imbernon, Luca Chiesa, Esther Kellenberger – Mining the Protein Data Bank to inspire fragment library design – Rencontres Internationales de Chimie Thérapeutiques (RICT); Lille

05/07/2023-07/07/2023 – Julia Reville Imbernon, Esther Kellenberger – Modelling of the binding of neomycin B and its derivative into Aminoglycoside-Modifying Enzymes – Rencontres Internationales de Chimie Thérapeutiques (RICT); Lille

Autres présentations :

Orales:

22/03/2022 – Julia Reville Imbernon, Esther Kellenberger – Fragment-based approach applied to the search for new antibiotics – Séminaire d'UMR 7200; Strasbourg

08/04/2022 – Julia Reville Imbernon, Esther Kellenberger – Modelling the Recognition of Aminoglycoside-Modifying Enzymes by Neomycin Synthetic Derivatives – Séminaire externalisé d'UMR 7200; Obernai

24/01/2023 – Julia Reville Imbernon, Luca Chiesa, Esther Kellenberger – Versatile PDB fragments – Séminaire d'UMR 7200; Strasbourg

LISTE DES PUBLICATIONS

1. Reville Imbernon, J.; Jacquemard, C.; Bret, G.; Marcou, G.; Kellenberger, E. Comprehensive Analysis of Commercial Fragment Libraries. *RSC Med. Chem.* **2022**, *13*, 300-310, 10.1039.D1MD00363A
2. Reville Imbernon, J.; Chiesa, L.; Kellenberger, E. Mining the Protein Data Bank to Inspire Fragment Library Design. *Front. Chem.* **2023**, *11*, 10.3389/fchem.2023.1089714

I Introduction

I.1 General introduction to bacteria

Bacteria are a diverse group of microscopic organisms that have a significant impact on the planet and play crucial roles in ecological processes. They have adapted to various environments over billions of years, exhibiting remarkable diversity in terms of morphology, physiology, and ecological interactions. They are single-celled organisms that are part of the prokaryote organisms, lacking a nucleus and membrane-bound organelles.

Bacteria can be classified into two categories: Gram-positive (Gram+) and Gram-negative (Gram-). This classification is based on their response to the Gram staining technique, which provides valuable information about their cell wall composition.¹ Gram+ bacteria retain the crystal violet stain, having thick cell walls primarily composed of peptidoglycan. Gram- bacteria do not retain the stain, and instead take up a counterstain, appearing pink or red. They have thinner cell walls with an outer membrane containing lipopolysaccharides, negatively charged and forming a hydrophilic network. The presence of an outer membrane in Gram- bacteria prevents the penetration of hydrophobic molecules, and thus provides additional resistance to certain antibiotics and immune responses.²

I.2 General introduction to antibiotics and antibacterial resistance

While bacteria play an essential role in many aspects of life, they can also be responsible for a wide range of infections and diseases in humans and other hosts. When bacteria invade the body, they can cause illnesses such as pneumonia, urinary tract infections, or skin infections. To combat these bacterial infections, antibiotics are employed.

Antibiotics are compounds that can either kill bacteria or inhibit their growth. They work by blocking one or more stages of the bacterial cell cycle, which are crucial for their survival or reproduction. Most of these molecules are naturally produced and synthesized by various organisms, including plants, fungi, and prokaryotes. However, there is now a significant number of synthetic molecules available in the market as well.³

In the pre-antibiotic era, before the discovery of antibiotics, bacterial infections often resulted in severe illness and death. Ancient civilizations used various natural substances with

antimicrobial properties, such as moldy bread, to treat infections.^{3,4} However, the mechanisms of action and scientific understanding were limited during that era.

In the late 19th century, Louis Pasteur and Robert Koch established the germ theory of disease, demonstrating the role of bacteria in causing infections. In the early 20th century, Paul Ehrlich introduced the concept of a “magic bullet”, which was a substance capable of selectively targeting pathogens without harming the host. He developed the first synthetic antimicrobial agent, Salvarsan, for treating syphilis.^{4,5} Later, the discovery of penicillin, attributed to Alexander Fleming in 1929, marked a significant breakthrough. Fleming observed that a mold called *Penicillium* inhibited the growth of bacteria. In 1939, Howard Florey and Ernst Chain furthered the development of penicillin, leading to large-scale production and use as a potent antibiotic during World War II. This discovery opened the doors to the golden age of antibiotics, with the subsequent discovery and development of streptomycin, tetracycline, erythromycin, and other important antibiotics (see Figure II.2.1-1).

However, since 1970s, the discovery of new antibacterial classes has become increasingly challenging getting closer to a post-antibiotic era.^{5,6}

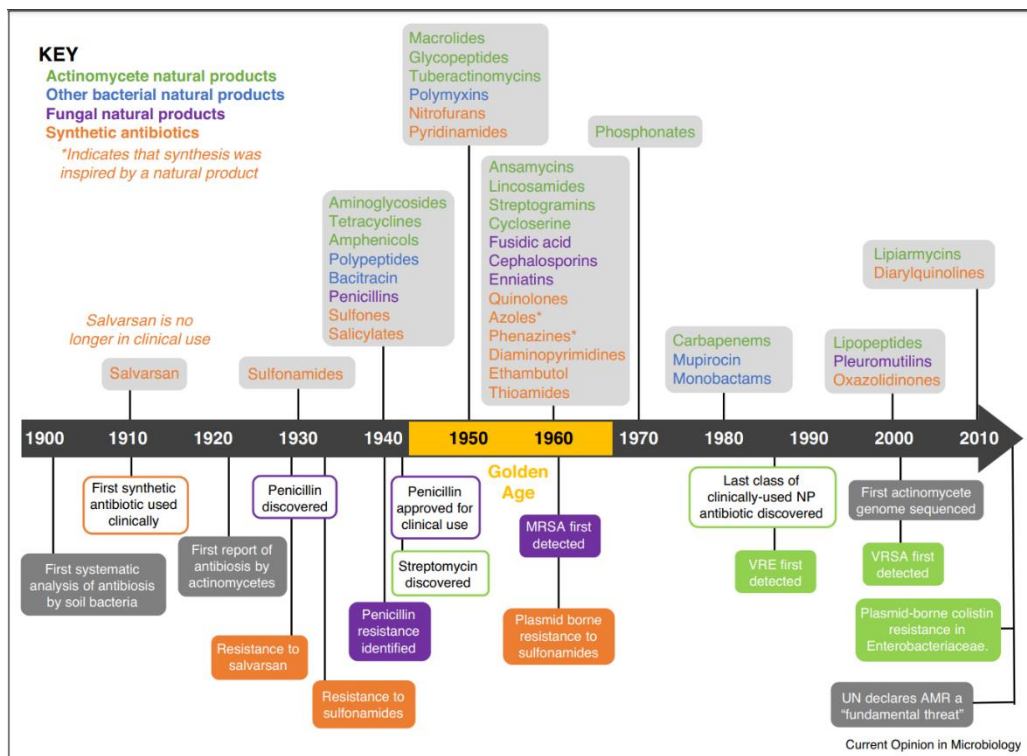


Figure II.2.1-1 Timeline of the discovery of new antibiotic classes and other events including the appearance of resistances. Image from the article³

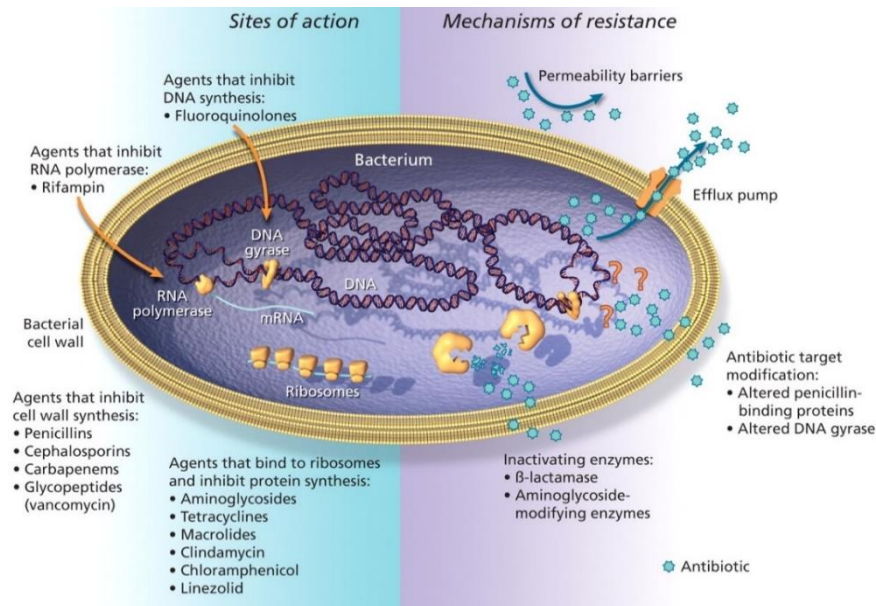


Figure II.2.1-2 Sites of action and potential mechanisms of bacterial resistance to antimicrobial agents. Image from the article⁷

Antibiotics employ various mechanisms of action, going from inhibiting the cell wall synthesis, disrupting the bacterial cell membrane, inhibiting DNA synthesis, inhibiting RNA polymerase or inhibiting protein synthesis (see Figure II.2.1-2).^{6,7}

Today, antibiotics continue to be a critical tool in healthcare, but their misuse and overuse have led to the rise of multidrug-resistant bacteria (MDR). The inappropriate dosage or duration of antibiotic treatment plays a significant role in the development of antimicrobial resistance (AMR). When microorganisms become resistant to a particular antibiotic, they possess the capacity to infect different hosts throughout the food chain or the environment. This dissemination is further facilitated by globalization, thereby intensifying the threat posed to public health⁸ (see Figure II.2.1-3).

Indeed, bacteria possess the remarkable ability to evolve and develop mechanisms to resist the effects of antibiotics. The genes responsible for antibiotic resistance can be located in plasmids or genome within bacterial cells.⁹ Thus, these resistance genes can be intrinsic and therefore transmitted vertically during bacterial reproduction or acquired through horizontal gene transfer. This transfer can occur through mechanisms such as conjugation, transformation, and transduction, allowing resistance genes to spread rapidly within bacterial populations.^{7,10}

Antibiotic resistance involves mechanisms that prevent the antibiotic from reaching its target. Three main mechanisms exist: limiting the access to the target through a permeabilization of



Figure II.2.1-3 Infographic about the bacterial interconnection between human, animals and the environment and about the spread of bacterial resistance. Image from FAO¹¹

the membranes or actively pumping out the antibiotic; modifying the antibiotic target site; or inactivating the antibiotic with enzymes (see Figure II.2.1-2).¹¹

I.3 Epidemiology

Antibiotic resistance poses a significant and growing threat to public health. In 2019, it was estimated that resistant bacteria caused approximately 1.27 million deaths and were associated with 4.95 million deaths.¹² If the current trend continues, by 2050, AMR could result in 10 million deaths annually, surpassing the number of deaths caused by cancer (see Figure II.2.1-1).¹³ Disturbingly, there have been reported cases of bacteria, such as *Acinetobacter baumannii*, demonstrating resistance to all available antibiotics.¹⁴ Another worrying example is *Klebsiella pneumoniae*, where infection with carbapenem-resistant strains has been linked to high mortality rates in both the United States of America and Europe of 30%–40% and 40%–50%, respectively.^{15,16}

Additionally, research¹⁷ has shown that antimicrobial resistance may be related to factors such as gross domestic product per capita, education, infrastructure, public health-care spending, antibiotic consumption, high temperatures, weaker governance, and a higher ratio of private to public health expenditure. Indeed, low to middle income countries, refugee camps, and conflict zones that lack stringent prescription regulations for antibiotics, proper infection prevention and control measures, and sufficient access to uncontaminated water supplies are at a higher risk of experiencing a greater number of cases of AMR.¹⁷

It is crucial to acknowledge that the AMR crisis not only poses a threat to public health but also threatens modern medicine as a whole. Antibiotics not only save lives, but they also enable essential medical procedures. If these trends continue, several predictions include increasing difficulties in operations and surgeries, higher mortality rates among immunocompromised patients, and potential alterations in human and other host microbiomes leading to the acquisition of various diseases.^{12,13}

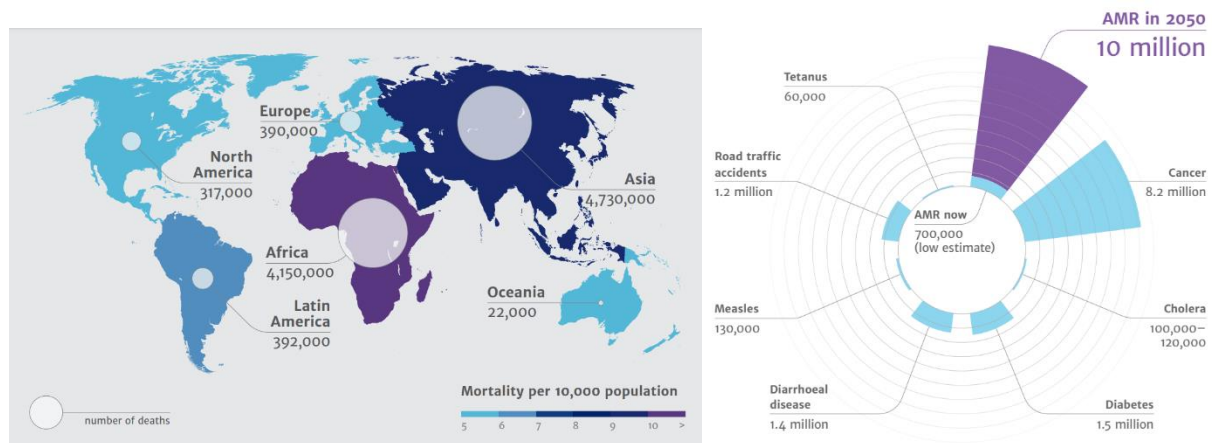


Figure II.2.1-1 Number of deaths attributable to AMR in 2014 and prediction for 2050. Image from the AMR Review Paper from 2014 and 2016.^{13,21}

I.4 Politics to address AMR crisis

The globalization of our modern world has accelerated the spread of resistant microorganisms across borders, transcending geographical boundaries, and thus requires a global and coordinated approach.

Recognizing this, key international organizations such as the World Health Organization (WHO), the Food and Agriculture Organization (FAO), and the World Organization for Animal Health (WOAH) have joined forces to establish a global action plan against microbial resistance in 2015.¹⁸ The United Nations (UN) also acknowledged the importance of addressing AMR on a global scale, recognizing it as a global priority in 2016. This recognition further amplified the significance of a coordinated and united effort to combat AMR. Finally, in 2017, the WHO took a crucial step by publishing a list of priority bacteria, highlighting the urgent need for research and development (R&D) of new antibiotics. This list serves as a guide for scientists, pharmaceutical companies, and policymakers, directing their efforts towards developing effective treatments against the most threatening pathogens (*E. faecium*, *S. aureus*,

K. pneumoniae, *A. baumannii*, *P. aeruginosa*, and *Enterobacter spp.*, also called “ESKAPE”).¹⁹

Before 2010, the strategies employed to combat antimicrobial resistance primarily revolved around human health concerns. However, with a growing realization of the connection between human health, animal welfare, and environmental stability, the “One Health” initiative gained importance. By encouraging collaboration, this policy seeks to address antibiotic misuse across all three domains, thereby promoting improved health outcomes on a global scale. Consequently, its main objectives consist of preventing zoonotic disease outbreaks, enhancing food safety and security, safeguarding global health security, and preserving biodiversity and conservation efforts.²⁰

To conclude on a positive tone, as stated in the 2014 AMR Review Paper: “This might be one of the world’s biggest problems, but it does not need to be its hardest.”²¹ In fact, we still have the power to alter the situation and prevent a regression to the pre-antibiotic era. This thesis is my contribution to it.

I.5 Organization of this thesis manuscript

This thesis explores the two fundamental strategies for the research and development of new antibiotics:

- Chapter II focuses on optimizing an existing antibiotic, neomycin B, to overcome resistance mechanisms while preserving their antimicrobial activity.
- Chapter III focuses on the discovery of novel compounds able to target a new biological target, the bacterial sliding clamp.

II Rational design of aminoglycoside derivatives

This work is part of a collaborative project aiming at the development of aminoglycoside derivatives that are active on a selection of resistant bacterial strains.

This chapter is divided into three independent, yet complementary, parts:

- Part II.1: A description of the chemical structure of aminoglycosides gives the keys to understand the nomenclature to the readers.
- Part II.2: The analysis of the three-dimensional structures of two exemplary aminoglycosides which are well described in the Protein Data Bank (PDB) provides structural insights into their functional properties.
- Part II.3: The design of a computing method to predict the antimicrobial action of aminoglycosides in a selection of resistant bacterial strains. We worked on the proof of concept using neomycin B and a synthetic derivative, the compound HL_171.

II.1 The chemical structure of aminoglycosides

Aminoglycosides (AGAs) constitute a wide class of therapeutic compounds that have historically been used for combating bacterial infections. They consist of a central structure known as an aminocyclitol, which is connected to amino sugars through glycosidic bonds.

The classification of AGAs into sub-families is primarily based on the present type of aminocyclitol. The majority of AGAs contain a motif called 2-deoxystreptamine (2-DOS), while streptomycin contains a streptidin motif, and fortimicins contain a fortamine moiety (see Figure II.2.1-1).

The second level of classification concerns the substitutions on the aminocyclitol ring and involves four categories:²²

- Mono-substituted compounds at position 4, such as neamine, paromamine, nebramine, and apramycin.
- Mono-substituted compounds at position 5, such as hygromycin.
- Di-substituted compounds at positions 4 and 5, such as neomycin, paromomycin, lividomycin B, and ribostamycin.

- Di-substituted compounds at positions 4 and 6, such as arbekacin, amikacin, dibekacin, gentamicin, geneticin, isepamicin, kanamycins A and B, nebramine, netilmycin, plazomicin, sisomicin, and tobramycin.

Moreover, to ensure an accurate description of the different aminoglycosides and eliminate any potential confusion, a precise numbering of the rings and substituents positions has been established (see Figure II.2.1-1):^{23,24}

- “Ring II” is the central aminocyclitol.
- “Ring I” is the sugar attached in position 4 of the central aminocyclitol; in the numbering of atoms, ring I is flagged by one prime sign (').
- “Ring III” is the sugar attached in position 5 or 6 of central aminocyclitol; in the numbering of atoms, ring III is flagged by two prime signs ('').
- If applicable, “Ring IV” is the sugar attached to ring III; in the numbering of atoms, ring IV is flagged with three prime signs (''').
- If applicable, “Ring V” is attached to ring IV; in the numbering of atoms, ring IV is flagged with four prime signs ('''').

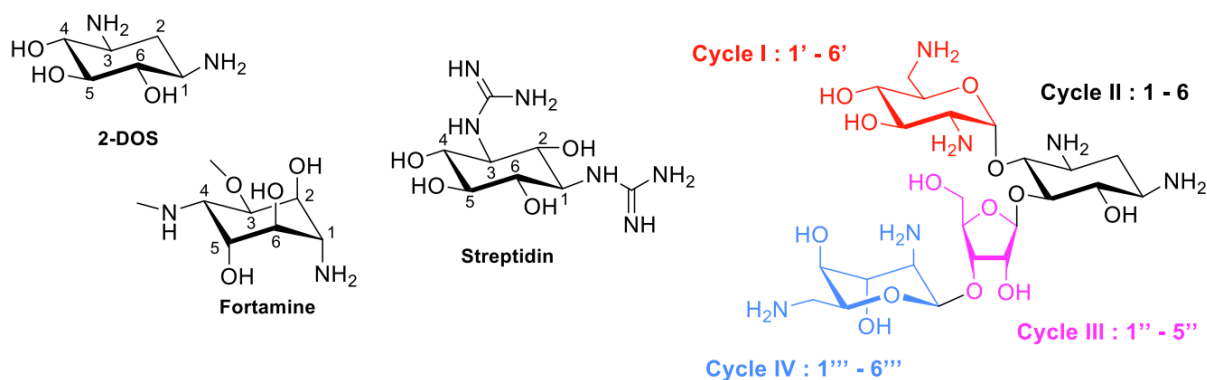


Figure II.2.1-1 Structures of the aminocyclitol cores and AGAs numbering, with neomycin B as typical aminoglycoside. Image from article ²⁴

II.2 Structural analysis of neomycin B and kanamycin A binding AMEs and bacterial ribosomal RNA

II.2.1 Introduction

Since the discovery of streptomycin in 1943, aminoglycoside antibiotics (AGAs) have been crucial in the treatment of various Gram-positive and Gram-negative bacterial infections. AGAs derive from substances produced by the species *Streptomyces* or *Micromonospora* and are effective against a wide range of bacterial pathogens, including the members of the Enterobacteriaceae family, many *Mycobacterium* spp., *S. aureus*, *Listeria*, non mucoid *P. aeruginosa*, and *A. baumannii* in a lesser extent.^{25,26} This therapeutical class enters bacteria through the membranous respiratory apparatus and binds the bacterial small 30S subunit on 16S ribosomal RNA subunit at the tRNA acceptor site A. The binding leads to the misread of mRNA, and thus to the synthesis of non-functional proteins which can ultimately cause the death of the bacterial cell.^{27,28} Aminoglycosides have shown their interest in the treatment of genetic diseases, thanks to their ability to suppress premature termination codons (PTCs). They were also studied for the treatment of HIV as they can target many *in vitro* steps of the virus life cycle.²⁹

Gentamicin, amikacin, and tobramycin are the most clinically used aminoglycosides because of their reliable activity against Gram-negative aerobic *bacilli*.^{25,30} Generally, AGAs cannot be absorbed by the intestine tract and are hence administered via topical, inhalation, or parenteral routes. Neomycin and paromomycin are the exceptions to this rule as their low absorption make them useful to suppress the intestinal bacterial flora.²⁵ Nevertheless, the prolonged use of these antibiotics can cause nephrotoxicity or ototoxicity. Indeed, the excretion of aminoglycosides is mainly operated by the kidneys. The partial reabsorption of AGAs during the nephron reabsorption step causes the death of kidney epithelial cells. However, this renal toxicity is generally reversible thanks to the regeneration capacity of proximal tubular cells.³¹ On another side, the ototoxicity occurs in patients carrying a punctual mutation of the mitochondrial 12S rRNA on the internal ear cells. The mutation A1555G or C1494U renders the sequence of the mitochondrial 12S rRNA similar to the bacterial rRNA aminoglycosides target on bacteria and thus, binds the AGAs irreversibly.³² Despite toxicity, aminoglycosides are listed as essential medicines by the WHO and are still crucial against the Gram-negative bacterial infections.^{33,34}

After 80 years on the market, another limitation tackles aminoglycosides. Different AGAs resistance mechanisms were selected and have emerged such as the modification of the active site due to ribosome mutations,³⁵ methylations by 16S rRNA methyltransferases,³⁶ drug efflux³⁷ or the decrease of bacteria membrane permeability.³⁸ However, the most common mechanism of resistance is assigned to Aminoglycoside-Modifying Enzymes.²³

Aminoglycoside-Modifying Enzymes (AMEs) are an effective resistance mechanism able to metabolize all existing aminoglycosides. In 2010, a staggering number of 120 different AMEs were reported, highlighting the diversity and complexity of these enzymes.²³ More simply, AMEs can be classified by catalytical activity resulting in only 3 types: Aminoglycoside N-Acetyltransferase (AAC), Aminoglycoside O-Phosphotransferase (APH) and Aminoglycoside O-Nucleotidyltransferase (ANT) (Figure II.2.1-1A). To accomplish their biological activity, each AME requires a cofactor, and a divalent ion in the case of APHs and ANTs. Each AME has a specific resistance profile and metabolizes one position of the targeted aminoglycosides. Thus, aminoglycosides are targeted by a large but specific enzymatic modification profile. Specific examples for neomycin B and kanamycin A are shown in Figure II.2.1-1B.²³

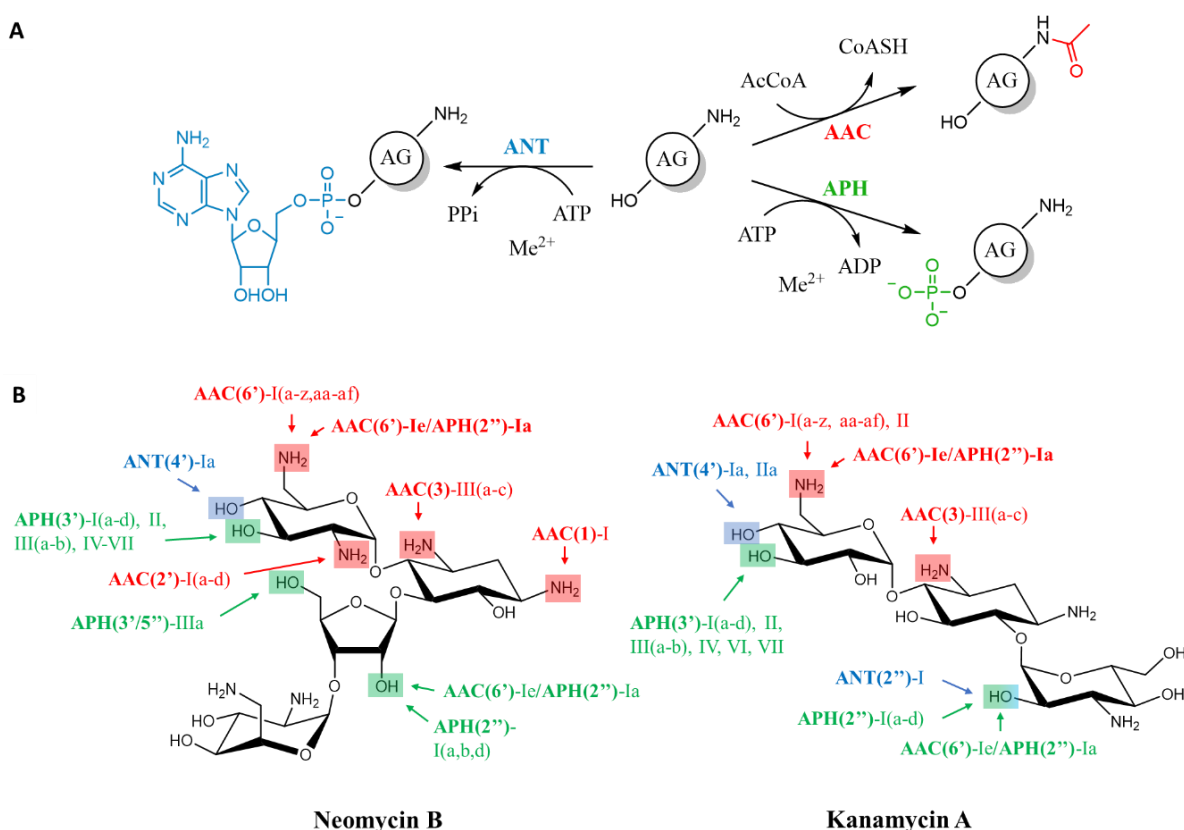


Figure II.2.1-1 Action of AMEs on aminoglycosides. A) Reactions catalyzed by the three AMEs. Image adapted from articles 24 and 26 B) Enzymatic modification profile of neomycin B and kanamycin A.

Many efforts have been made since the early 1970s to develop aminoglycoside derivatives that are highly effective, minimally toxic, and capable of overcoming bacterial resistance.^{22,24,39} Significant advancements in chemistry have led to the production of a broad range of amino-modified glycosides, culminating in the introduction of a novel drug in 2019, called plazomicin.⁴⁰ Regrettably, bacteria have rapidly developed resistance to all semi-synthetic aminoglycosides, including plazomicin, highlighting the critical need for ongoing research and development of alternative treatments.^{23,41}

In addition to investigating the relationship between structure and activity, the analysis of the three-dimensional structures of aminoglycosides has served as the foundation for comprehending their antibiotic activity. The mechanism of action of aminoglycosides was uncovered in 1965,²⁷ but it took twenty years to identify the binding site of these compounds.^{28,42} The first image of the 30S ribosomal particle bound to an aminoglycoside was obtained through X-ray crystallography in 2000.⁴³ With the advent of the genomic era and advancements in technology, the discovery and design of antibiotics have shifted towards target-based strategies.⁴⁴ Consequently, since the 2000s, there has been a growing number of crystallographic structures published. Initially, research focused on investigating the mechanism of action of aminoglycosides by crystallizing the 30S ribosomal subunits in complex with tRNA, mRNA, and AGAs.⁴⁵ Subsequently, studies of AGAs-RNA complexes that did not include the ribosomal subunit gained popularity between 2015 and 2019. Finally, the need to overcome resistance has led to an increased focus on AGAs-AMEs complexes, with 35 of the 82 existing structures published between 2015 and 2019, and 22 since the beginning of 2020.⁴⁵ In brief, research on the AGAs primary target has shed light on the molecular mechanisms underlying antibiotic action, while studies on AMEs have revealed how these enzymes can metabolize one or more substrates in the aminoglycoside family. However, only a limited number of studies have investigated the interplay between different types of target biomolecules and AGAs. In a comprehensive structural study carried out in 2003, Vicens and Westhof showed that neamine, which is composed of rings I and II and is conserved in almost all AGAs, plays a critical role in binding to the ribosomal site A and AMEs.⁴² Both binding sites have negatively charged residues on their surface, which are complementary to the positively charged amine groups in AGAs, and thus, the binding modes are dominated by electrostatic interactions and hydrogen bonds. This study was however based on a limited number of structures: 14 PDB entries, including five complexes between an AGA and the 30S

subunit, three complexes between an AGA and synthetic RNA fragments containing two A-sites, and six complexes with three different AMEs.

In this review, we provide an inventory of the structures of AMEs available in the Protein Data Bank (PDB), with a focus on the recognition mode of two leading AGAs, namely neomycin B and kanamycin A. Neomycin B and kanamycin A belong to 4,5-2-DOS disubstituted and 4,6-2-DOS disubstituted AGAs, respectively, and are among the most documented aminoglycosides in their class in the PDB. The structures of these two AGAs indeed cover the three types of AMEs, but also the RNA A-site target. We have conducted a systematic comparison of the binding modes of these two AGAs to their respective binding sites and we will here present the conclusions derived from this analysis. Firstly, we will outline the chemical groups that form non-covalent bonds with different types of AMEs with the aim of identifying common or specific recognition positions for these enzymes. Secondly, we will describe the conformations of neomycin B and kanamycin A in their bound forms to evaluate the adaptability of these two AGAs to the protein environments of AMEs. Finally, we will revisit our findings in light of the requirements needed for RNA binding and the functionalizations prone to prevent AME binding.

II.2.2 Materials and methods

II.2.2.1 Collection of AME-AGA complexes from PDB

Three dimensional structures of AMEs containing neomycin B or kanamycin A (HET codes: NMY or KAN, respectively) were selected from the PDB website and saved under mmCIF format. The published sc-PDB protocol^{46,47} was followed for the preparation, standardization, and binding site detection of the complexes. In this protocol, the all-atom description of molecules, including hydrogen, was prepared using Protoss (v4.0, ZBH, University of Hamburg, Germany).⁴⁸ Entries failing at any step of the protocol were discarded. Covalently linked aminoglycosides were ignored. The aminoglycoside binding site in the AME was defined as all the residues at a maximum distance of 6.5 Å from the aminoglycoside. For each of the AME-AGA structures, the aminoglycoside, the protein, and its binding site were recorded in three separate files in MOL2 format. Water molecules were included in the protein and the site provided forming at least two hydrogen bonds with the protein and one with the aminoglycoside. Metal cations were also included in the protein and the site. This is not the case for the other cofactors, discarded because of their inconsistent description.

II.2.2.2 Collection of RNA-AGA complexes from PDB

Three dimensional structures of the bacterial A site containing neomycin B and kanamycin A were downloaded from the PDB website. The residues constituting the bacterial ribosomal site A and having at least one atom at a distance of 6.5 Å from the aminoglycoside were extracted from the PDB structure and saved under PDB format using MOE (2022.02 Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7). The protonation was performed using Protoss (v4.0, ZBH, University of Hamburg, Germany).⁴⁸ The correct valences of all atoms were verified by using RDKit tools (RDKit: Open-source cheminformatics, <https://www.rdkit.org>, <https://doi.org/10.5281/zenodo.7415128>). Water molecules forming at least two hydrogen bonds with the RNA and one with the aminoglycoside were included in the complex.

II.2.2.3 Assessment of the quality of structures

Entries with resolution higher than 3.5 Å were discarded. The quality of structures was evaluated using Real-Space Correlation Coefficient (RSCC) following the Twilight classification: $RSCC > 0.9$ states that the model fits the electron density, $0.9 > RSCC > 0.8$ is given to a model partially fitting the electron density, $RSCC < 0.8$ advertises about models with questionable quality to be used with caution. The RSCC of the ligands was extracted from the “validation XML” file present on the PDB website for each entry. The mean RSCC of the binding site residues was calculated by using the score given for each amino acid on the “validation XML” file.

II.2.2.4 Interactions analysis

The interactions formed between the aminoglycoside and the AME binding site were detected using IChem (v5.2.8, University of Strasbourg, France).^{49,50} IChem relies on geometric rules to identify ionic bonds, hydrogen bonds, metallic interactions, aromatic stacking, and hydrophobic contacts. IChem generates a MOL2 file containing three pseudoatoms per interaction: one representing the ligand atom(s) which is (are) involved in the interaction; one representing the protein atom(s) which is (are) involved in the interaction; and one at the center of the segment defined by these two points.

The interactions formed between the aminoglycoside and the RNA were detected by calculating the distances (3.5 Å for hydrogen bonds and 4 Å for ionic bonds) and angles ($>120^\circ$)

between the atoms with RDKit (RDKit: Open-source cheminformatics, <https://www.rdkit.org>, <https://doi.org/10.5281/zenodo.7415128>).

II.2.2.5 Aminoglycoside conformational analysis

The root mean square deviation (RMSD) of the coordinates of neomycin B or kanamycin A was calculated using the pair-fit tool offered by PyMOL (v2.3.5, Schrodinger, LLC) through an all-against-all fit of all non-hydrogen atoms. The number of conformations was determined via a density-based clustering using Sklelarn ($\text{eps} = 0.5$).⁵¹

For display purposes, the 3D structures of an aminoglycoside were aligned based on rings I and II only (all non-hydrogen atoms except C(6')-NH₂).

The dihedral angles of the linkage between the rings of the aminoglycoside were determined using RDKit tools (RDKit: Open-source cheminformatics, <https://www.rdkit.org>, <https://doi.org/10.5281/zenodo.7415128>) as described by ⁵².

II.2.2.6 AME sequence comparison

Only peptide chain involved in the binding of aminoglycosides was considered for the comparison of AMEs sequences. Sequences were aligned and compared by using the water EMBOSS package (v.6.6.0.0, EMBL-EBI, Cambridgeshire, United Kingdom). The default settings were used: EMBLOSUM62 as the comparison matrix, a gap opening of 10, and a gap extension set as 0.5. Alignments with a length of less than 100 amino acids were not taken into account.

II.2.2.7 RNA sequence comparison

The identity of the sequences was calculated by the local comparison of a consensus AGA binding site, which comprises G1488 to G1497 on the direct strand and C1404 to C1412 on the reverse strand.

II.2.3 Results

In order to study how aminoglycosides bind to AMEs and bacterial RNA, we retrieved all PDB entries that featured neomycin B or kanamycin A bound to any AME or to the bacterial ribosomal A site. We identified the specific interactions present in each complex and mapped

them onto 2D structures of the aminoglycoside as pharmacophoric points. The study begins by examining the various binding modes and conformations of neomycin B, which enable it to bind to different AMEs and to the bacterial ribosomal A site. Then, we repeat these analyses for kanamycin A.

A total of 29 PDB files of aminoglycosides in complex with AMEs were explored, 10 containing neomycin B and 19 containing kanamycin A (Table II.2.3-1). Several entries in the dataset contained multiple copies of an aminoglycoside-AME complex, resulting in 92 complexes being analyzed in this study. Of these complexes, 22 involved neomycin B and 70 involved kanamycin A. Four PDB entries containing neomycin B were discarded due to covalently linked structures or because of not passing one step of the protocol (PDB ID: 6NMM, 6NMN, 6P08, and 7Q1X). Overall, the structures are of good resolution and exhibit good local precision. According to the twilight classification, in 79 of structures, the model correctly fits the electron density (RSCC >0.8). Out of the analyzed structures, only 13 had an RSCC coefficient lower than 0.8 or were missing this value altogether. As a result, these structures need to be approached with caution. However, in order to avoid losing valuable information regarding the binding of aminoglycosides to certain AMEs, these structures were still included in the study. Any new insights that may be gained from these structures will be noted and carefully interpreted.

The studied dataset covers the three AME types: AAC, APH and ANT. The AACs and APHs activities have been found in other enzymes. AACs belong to the GCN5-related N-acetyltransferases superfamily whereas APHs function as protein kinases. No enzyme superfamily has been associated with ANTs.^{23,53,54} In spite of a common function, enzymes in the same class can bear little resemblance at the level of their sequence. In the studied dataset, the highest sequence identity between AMEs binding neomycin B is 23.3% (between APH(3')-IIIa and the phosphotransferase of AAC(6')-Ie/APH(2'')-Ia). Of note crystallographic complexes containing engineered mutated versions of ANT(4')-Ia (T130K and T130K/E52D) were included in the study as the arrangement of their active site and catalytic function remained unchanged. However, the doubly mutated enzyme (T130K/E52D) modifies the binding to metal ions, resulting in a weaker binding to neomycin.⁵⁵ The dataset contains more AMEs targeting kanamycin, and not unexpectedly their comparison revealed close homolog pairs. For example, the sequence identity is 60% between the acetyltransferase protomer of AAC(6')-Ie/APH(2'')-Ia and APH(6')-Im. The three APH(3') enzymes share between 30 and

Table II.2.3-1 Studied PDB entries containing neomycin B or kanamycin A in complex with an AME or RNA. The superscript indicates the number of aminoglycoside copies. * refers to the bifunctional enzyme AAC(6')-Ie/APH(2'')-Ia.

aminoglycoside-modifying enzyme				PDB ID	
type	Targeted position	Resistance profile	Identifier	neomycin B	kanamycin A
AAC	3	III	b	6MB5 ¹ , 6MB9 ⁴	
		VI	a		6O5U ¹
	2'	I	c	7CS1 ²	1M4I ²
			d		
	6'	I	e*		4QC6 ²
m				6BFH ¹	
APH	3'	I	a	2B0Q ¹	4FEU ⁶ , 4FEV ⁶ , 4FEW ⁶ , 4FEX ⁵ , 4GKH ¹² , 4GKI ¹²
					II
		III	a		1L8T ¹
	2''	I/IV	d/a	5IQE ⁴	5IQB ⁴
					III
ANT	4'	I	a	6UN8 ²	1KNY ²
			a/T130K mutant	6NMK ² , 6NML ²	
			a-T130K/E52D mutant	6P04 ² , 6P06 ²	
	2''	I	a		4WQL ¹
RNA	A site			2A04 ² , 2ET4 ²	2ESI ²
	30S ribosomal subunit from <i>Thermus thermophilus</i>			4LF6 ¹ , 4LFB ¹	
	70S ribosome from <i>E. coli</i>			4V52 ² , 4V57 ² , 4V9C ²	

35% identical sequence. In brief, the dataset studied illustrates the diversity of AMEs binding neomycin B or kanamycin A, but also presents versatile enzymes acting on the two aminoglycosides.

Regarding RNA complexes, 8 PDB entries were studied, 7 containing neomycin B and 1 kanamycin A (Table II.2.3-1). These complexes contained a total of 12 neomycin B and 2 kanamycin A bound to site A. All structures had an RSCC > 0.8, fitting the electron density. The RNA sequences constituting the studied binding site are identical between five of the PDB entries (PDB ID: 2A04, 2ET4, 4V9C, 4V52, and 4V57) and share 90% identity with the two other sequences (PDB ID: 4LF6 and 4LFB).

Each of the analyzed structures have been prepared using a standardized protocol to obtain an all-atoms description of the AGA-AME/RNA complex. Missing hydrogen atoms have been added and placed at the most probable position based on the construction of an optimal hydrogen bonding network.⁴⁸ Based on the hydrogen configuration, hydrogen bonds, ionic bonds, and metallic interactions between AGAs and AMEs/RNA have been detected based on

geometrical rules, such as specific distances and angles. Since water molecules can play an important role in stabilizing the complex and mediating interactions between the AME/RNA and the AGA, we considered that water molecules forming two or more hydrogen bonds with the protein were tightly bound to the protein and, therefore, were part of the binding site.

II.2.3.1 Neomycin B in complex with AMEs and RNA.

The 22 structures of neomycin B and AME complexes are distributed across five enzyme classes, including two AACs, two APHs and one ANT (Table II.2.3-1). When considering all classes collectively, it has been observed that each of the four aminoglycoside rings is accommodated by at least two enzyme classes, and nearly all of the nitrogen and oxygen atoms in the aminoglycoside are utilized by at least one of the enzymes (Figure II.2.3-1A). Additionally, for each enzyme considered individually, a significant portion of the aminoglycoside interacts with the protein site. The total number of polar interactions per complex ranges from 11 with AAC(3)-Iib (PDB ID: 6MB9) to 27 with APH(2'')-Ia (PDB ID: 5IQE) (Figure II.2.3-1C), including hydrogen bond (HB), ionic bonds and an interaction with a metallic cation. When examining neomycin B, attention is directed towards its seven hydroxyl groups and six amine groups. Among hydroxyl groups, all are capable of acting as hydrogen bonding donors (HBD), with one group also interacting with a metallic cation. Additionally, five of these hydroxyl groups can also act as hydrogen bonding acceptors (HBA). When considering the amino groups, it is predicted that they will be positively charged in all of the studied complexes except for one of the complexes of ANT(4')-Ia (PDB ID: 6MNK). Furthermore, all six of these groups are involved in salt bridges that are assisted by hydrogen bonding. The other polar atoms of neomycin are oxygen atoms in the ether groups which are present in rings I, III and IV as well as in the three linkages connecting the four rings. Compared to the hydroxyl and amine groups, the ether groups of neomycin B play a minor role in the recognition of the AMEs, as only those located in or near the ring IV are recognized by the ANT(4')-Ia, and those located in or near the ring II are recognized by APH(2'')-Ia. It should be noted that the accessibility of the oxygen atom in ether groups is limited due to steric hindrance.

Water-mediated hydrogen bonds were detected in all investigated PDB files, with a maximal number of 9 and an average number of 3 ± 2 , suggesting that water plays a role in the recognition of neomycin B by all types of AMEs. The quantity and arrangement of water

molecules are contingent upon the AME, yet are not always consistent. In some PDB files that include multiple copies of the same complex, there may be both structures with and without water-mediated hydrogen bonds (e.g., PDB IDs 6MB9, 5IQE and 6P04). Several water-mediated hydrogen bonds are by contrast well conserved: with C(1)-NH₂ in four out of the five copies of AAC(3)-IIIb; with C(3), C(4'), and C(6') in the two copies of AAC(2')-Id ; with C(3)-NH₂, C(6')-NH₂, and C(6''')-NH₂ in seven out of ten copies of ANT(4')-Ia; and with C(5''')-OH in 5 of the copies of ANT(4')-Ia.

Noteworthy, the catalytic activities of APH and ANT require a metallic ion, yet the structures show a direct interaction between neomycin B and the metal only in ANT(4')-Ia.

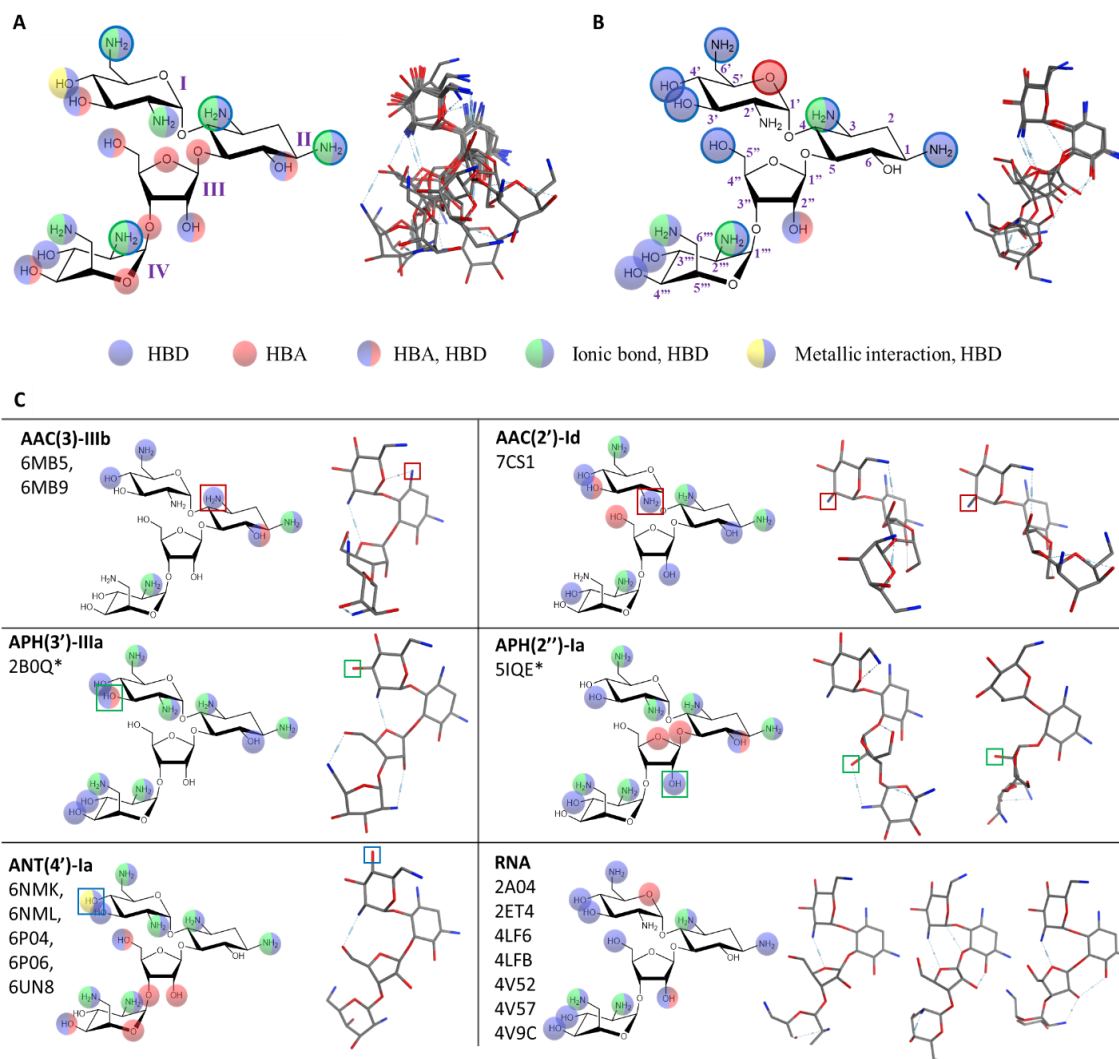


Figure II.2.3-1 2D and 3D representation of neomycin B (A) ensemble of all neomycin B-AME interactions and neomycin B conformers superposed on rings I and II. Outlined interactions are present in all AMEs. (B) ensemble of all neomycin B-RNA interactions and neomycin B conformers superimposed on rings I and II. Outlined interactions are present in all PDB entries. (C) 2D and 3D structures of neomycin B per AME and RNA. A colored square marks the targeted position. On the right, all neomycin B conformations binding the enzyme are depicted. The symbol * identifies low scored RSCC structures.

Regarding RNA complexes, neomycin B binds to several binding sites other than the ribosomal A site on the structures showing the 30S ribosomal subunit. Here, we focus on the bacterial ribosomal A site binding. All of the studied complexes except for one of the structures (PDB ID: 4V9C) have six charged amino groups. Neomycin B uses three hydroxyl groups, four amino groups, and one ether to bind the ribosomal A site in all 12 structures (Figure II.2.3-1B, see outlined interactions). Rings I and II form six out of the eight interactions which are conserved across the studied complex, while rings III and IV show a binding mode dependent

on the 3D structure. Interestingly, C(6')-NH₂ and C(1)-NH₂ do not form ionic bonds contrarily to the interactions with the AMEs. Also, C(2')-NH₂ and C(6)-OH do not interact with the RNA while they appear to be part of the binding mode for five out of four AME complexes. Among all interactions, C(3)-NH₂ forms the higher number of interactions by forming hydrogen and ionic bonds with two nucleotides (A1493 and G1494).

Only two water mediated interactions have been detected among the structures: one interacting with C(6')-NH₂ and a second one interacting with C(5''')-OH.

The comparison of the binding modes of neomycin B to the different protein classes points out the singularity of AACs (Figure II.2.3-1C). All studied enzymes interact with most of the aminoglycoside polar groups, while AAC(3)-IIIb only forms six polar interactions, including five with rings I and II. Additionally, the other AAC of the dataset, namely AAC(2')-Id, also form fewer interactions with ring IV than the APHs and ANTs. Given that rings I and II participate in numerous interactions between neomycin and its protein binding site, regardless of the enzyme type, we assessed whether this substructure of the aminoglycoside exhibits conformational diversity or instead adopts a consistent conformation across various AMEs.

Figure II.2.3-1A shows the superimposition of the 22 structures of neomycin B, after their 3D alignment for the best overlay of rings I and II. The conformation of rings I and II, i.e. neamine, is pretty conserved across the five AMEs. However, Figure II.2.3-1C shows that neomycin B adopts two different conformations when binding AAC(2')-Id and APH(2'')-Ia. The distribution of the torsion angles of the linkages between the rings presented in Figure II.2.3-2 more finely characterizes the conformational variations of the AME-bound forms of neomycin B. The linkage between rings I and II show conserved phi and psi angles, although the clustering of the values segregates the AMEs by type. While dihedral angles forming the bonds linking rings I and II move on a range of approximately 60°, bonds linking rings II-III and III-IV can differ by 180° providing a total new conformer. Despite a few exceptions, it is also noticeable that the dihedral angles are grouped by type of AME and so, that neomycin B's conformations are adapted to the environment. The biggest difference is shown by the structure represented in AAC(2')-Id (PDB code: 7CS1), where the ϕ angle of the bond linking ring II to III has rotated 180° compared to the other structures. The same thing happens for the ϕ angle of the linkage between rings III and IV of one of the two conformers of the enzyme. This mobility of rings III and IV can be appreciated in Figure II.2.3-1A. Actually, the left conformation of AAC(2')-Id forms all the interactions shown on the 2D structure except the

hydrogen bond with C(4')-OH on ring I, while the right conformation loses almost all the interactions formed by rings III and IV, only keeping the bridge salt formed by C(2''')-NH₂ of ring IV. APH(2'')-Ia (PDB code: 5IQE) also presents two conformers. The right structure, being the structure having a low RSCC, differentiates itself by adding up the hydrogen bond interaction formed by the C(3''')-OH of ring IV and the acceptor feature of the hydrogen bond formed by C(6)-OH of ring II. Neomycin B conformers binding APH(2'')-Ia are the ones showing the biggest difference of ψ angle for linkage II-III and of ϕ angle of rings III and IV on Figure II.2.3-2.

Neomycin B bound to the bacterial ribosomal A site also has a conserved neamine conformation in all complexes, while rings III and IV have a higher mobility. Interestingly, the conformations it adopts are similar to those adopted when in complex with AMEs (Figure II.2.3-2). In one of the neomycin B - RNA structures (PDB code: 4V9C), ring IV is rotated of 180° (Figure II.2.3-1A, Figure II.2.3-1B, and Figure II.2.3-2) similarly to the unique ϕ angle of the linkage between rings III and IV present in AAC(2')-Id.

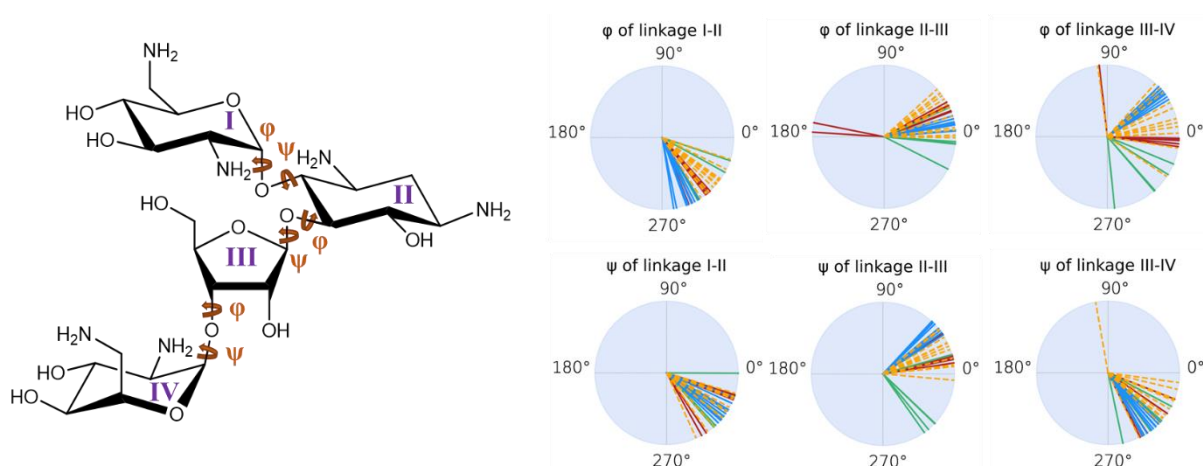


Figure II.2.3-2 Comparison of neomycin B conformers dihedral angles. Red color is attributed to conformers bound to AAC, green to APH, blue to ANT, and dashed orange to RNA. Atoms forming dihedral angles were defined following reference 52: ϕ (H1-C1-O1-Cx) and ψ (C1-O1-Cx-Hx).

In conclusion of the structural analysis, the binding of neomycin B to AME complexes is primarily governed by polar interactions, which is consistent with the highly polar and positively charged characteristics of the aminoglycoside and the complementary properties of the protein binding sites. This conclusion is supported by 448 polar interactions and hydrophobic contacts presented in Figure II.2.3-1A, where 69% of data involve glutamic and aspartic acids, i.e. negatively charged residues in AMEs. Additionally, water molecules significantly contribute to the binding, accounting for 14% of the interactions. Finally,

neomycin B acquires its flexibility from the bonds linking rings II to III and III to IV, keeping a conserved neamine conformation in all complexes. The binding mode and conformations of neomycin B when bound to RNA and AME are similar. However, ionic bonds are less common in RNA complexes and HBA are practically excluded from the binding mode.

II.2.3.2 Kanamycin A in complex with AMEs and RNA.

Figure II.2.3-3 collects all interactions with AMEs and conformations that kanamycin A presents in 70 structures (Table II.2.3-1). As for neomycin B, all rings and polar groups of kanamycin A are involved in the binding to AMEs, and the bonding is dominated by polar interactions (Figure II.2.3-1). The number of interactions formed by kanamycin A ranges from 8 to 26, with 20 being the average.

All seven kanamycin hydroxylic groups are able to adjust their role of HBD or HBA to the enzyme. One of them also involves a metallic interaction. Amino groups are essential to form ionic bonds with the AMEs. In fact, of a total of 1489 interactions when combining all structures, 885 of them occur with aspartic acid and glutamic acid. Apart from ionic bonds, all four amino functions also act as HBD. In the case of kanamycin A, only two over four ether functions act as HBA.

As mentioned, all three rings interact in all AME-kanamycin A structures. However, ANT(2'')-Ia is the only AME forming just three ionic bonds with three amino groups and a metallic interaction. All the other complexes involve a minimum of two hydrogen bonds formed by the hydroxylic groups and do not have any metallic interaction. AAC(2')-Ic is included in the analysis even though kanamycin A cannot be metabolized by the AME as the targeted position is a hydroxyl group. Though, the shown binding mode can be extrapolated to kanamycin B that has an amino group on that position.

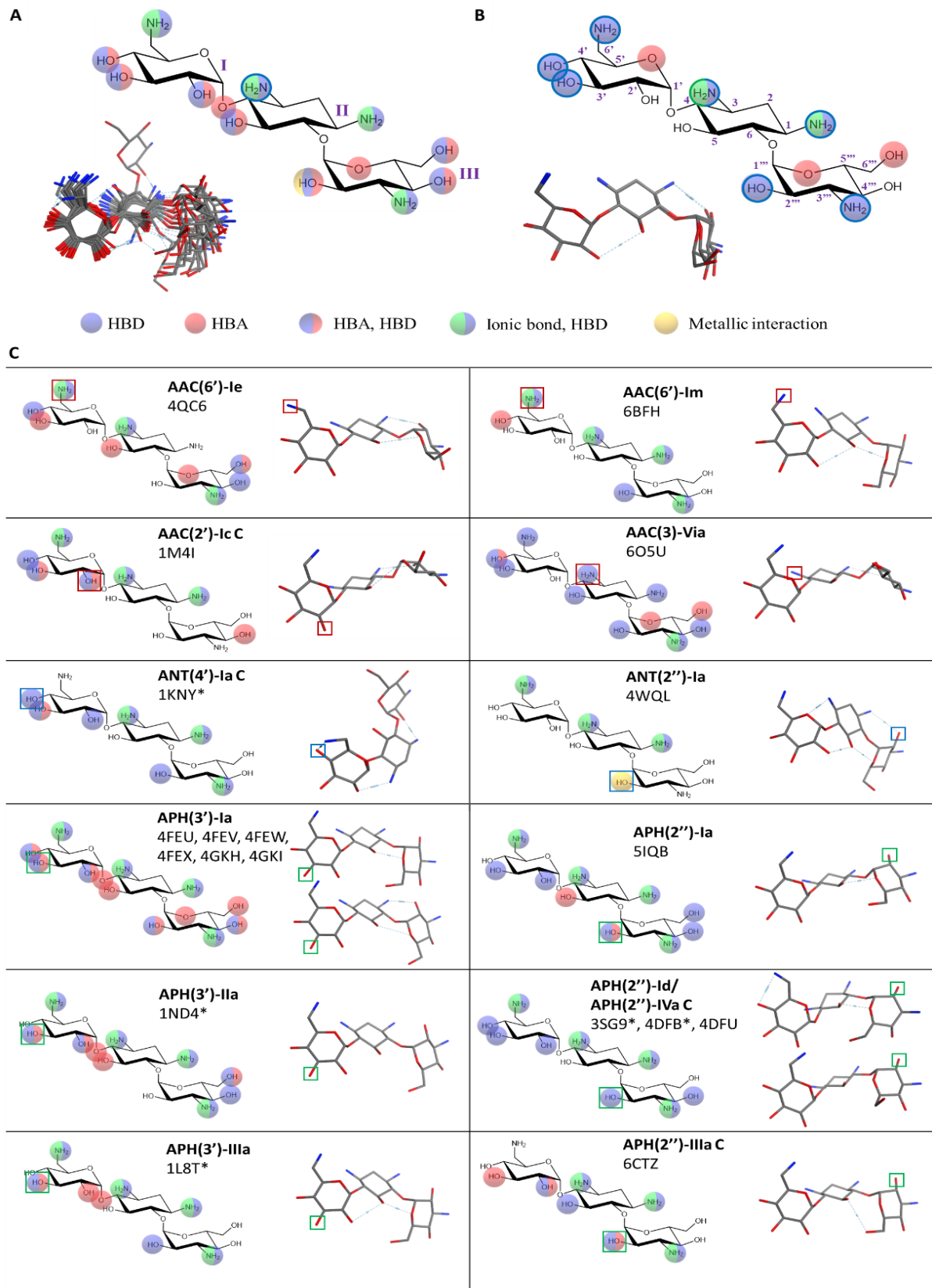


Figure II.2.3-3 2D and 3D representation of kanamycin A RX-structures (A) ensemble of all kanamycin A-AME interactions and kanamycin A conformers superimposed on rings I and II. Outlined interactions are present in all AMEs. (B) ensemble of all kanamycin A-RNA interactions and kanamycin A conformers superimposed on rings I and II. Outlined interactions are present in both RNA complexes. (C) 2D and 3D structures of kanamycin A per AME. A colored square marks the targeted position. On the right, all kanamycin A conformers binding the enzyme are depicted. The symbol * identifies low scored RSCC structures.

The water network is also important for the binding of kanamycin A as it is responsible for at least one interaction in 61 of the 70 structures. In average, it is responsible for 4 ± 2.5 interactions per complex, and for some complexes water forms up to 7 hydrogen bonds with the aminoglycoside. ANT(4')-Ia and ANT(2'')-Ia are the only two AMEs not implicating any water molecules on the binding mode. Several water-mediated hydrogen bonds are well conserved among the other complexes: with C(3)-NH₂ for all copies of AAC(2')-Ic, AAC(3)-VIa, and APH(3')-IIa, 38 copies out of 47 of APH(3')-Ia, and one copy out of two of AAC(6')-Ie; with C(6')-NH₂ for all copies of AAC(6')-Im, AAC(2')-Ic, AAC(3)-VIa, and APH(2'')-Ia, and 23 copies out of 47 of APH(3')-Ia one copy out of two of AAC(6')-Ie and APH(3')-IIa; with C(2')-OH for all copies of APH(3')-IIIa, APH(2'')-IIIa, 32 copies out of 47 of APH(3')-Ia, and one copy of AAC(2')-Ic and APH(3')-IIa; and finally with the O linking rings I and II in all copies of APH(3')-IIa, APH(3')-IIIa and 30 out of 47 of APH(3')-Ia.

Kanamycin A binds to RNA by using all four amino groups and four out of seven hydroxyl groups (Figure II.2.3-3B). Kanamycin A binding is mainly led by seven HBD interactions and two ionic bonds formed by C(3)-NH₂ as shown by the outlined interactions in Figure II.2.3-3B. In the case of kanamycin A, the three rings are implicated in the binding mode.

The superimposition of kanamycin conformers (Figure II.2.3-3A) shows a smaller flexibility of this aminoglycoside compared to neomycin B. All conformers except one are spatially very close to each other. The RMSD study, assigning new clusters if the RMSD difference is higher

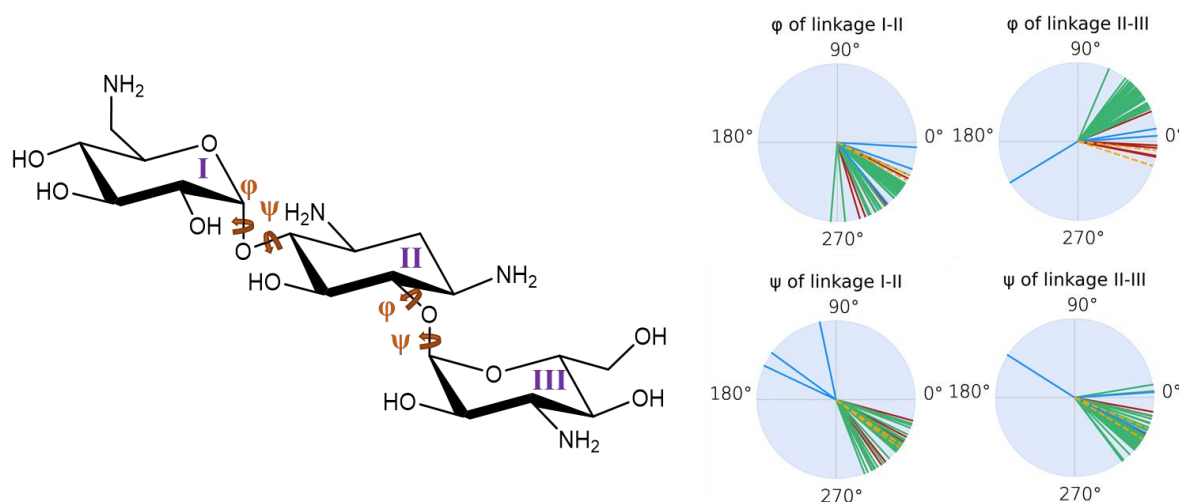


Figure II.2.3-4 Comparison of kanamycin A conformers dihedral angles. Red color is attributed to conformers bound to AAC, green to APH, blue to ANT, and dashed orange to RNA. Atoms forming dihedral angles were defined following reference 52: $\phi(\text{H1-C1-O1-Cx})$ and $\psi(\text{C1-O1-Cx-Hx})$.

than 0.5 Å, revealed only 2 clusters of conformers. Only the aminoglycoside binding ANT(4')-Ia totally differs from all other structures. All three ψ of linkage I-II, ϕ and ψ of linkage II-III dihedral angles are inversed when compared to the other conformers. On another side, both linkages seem equally responsible for flexibility, but the ϕ angle of linkage II-III seems to differentiate the kanamycin A binding APH, with angles between 24° and 67°, from those binding AAC, with angles ranging from 348° to 357° if we ignore AAC(6')-Im with an angle of 22°. This angle difference can be observed in Figure II.2.3-3A and Figure II.2.3-3C.

If we exclude the previously discussed exceptions, kanamycin A conformers bound to AME have similar dihedral angles to those bound to site A.

As with neomycin B, kanamycin A – AME complexes are governed by polar interactions. More than half of the interactions (59%) involve either glutamic acid or aspartic acid. The water network is again very important, acting as a bridge in 19% of interactions. Finally, the conformers of bound kanamycin A defined only two clusters even though the torsion angle ranges showed variations.

II.2.4 Example on amikacin

Amikacin is a semisynthetic derivative of kanamycin A introduced in the seventies. It is substituted at the position C(1)-NH₂ by a (S)-4-amino-2-hydroxybutyrate (HABA). Like the other aminoglycosides, it is active against Gram+ and Gram- bacteria, and, together with plazomicin, amikacin demonstrates higher resistance to the action of AMEs. Nevertheless, there has been an emergence of resistant strains, particularly associated with the rise of AAC(6')-Ib, which has limited its effectiveness.⁵⁶

This analysis of amikacin intends to provide insights into the potential of designing new AGAs derivatives based on the crystallographic structures.

Amikacin has been described in four crystallographic structures: two in complex with rRNA (PDB ID: 4P20 and 6YPU), one in complex with APH(2'')-Ia, GMPPNP, and Mg²⁺ (PDB ID: 6CGD), and one in complex with AAC(2')-Ia and AcCoA (PDB ID: 6VTA).

The analysis of the interactions between amikacin and rRNA revealed that the three rings of amikacin form polar interactions with the target. Indeed, all five amino groups were implicated

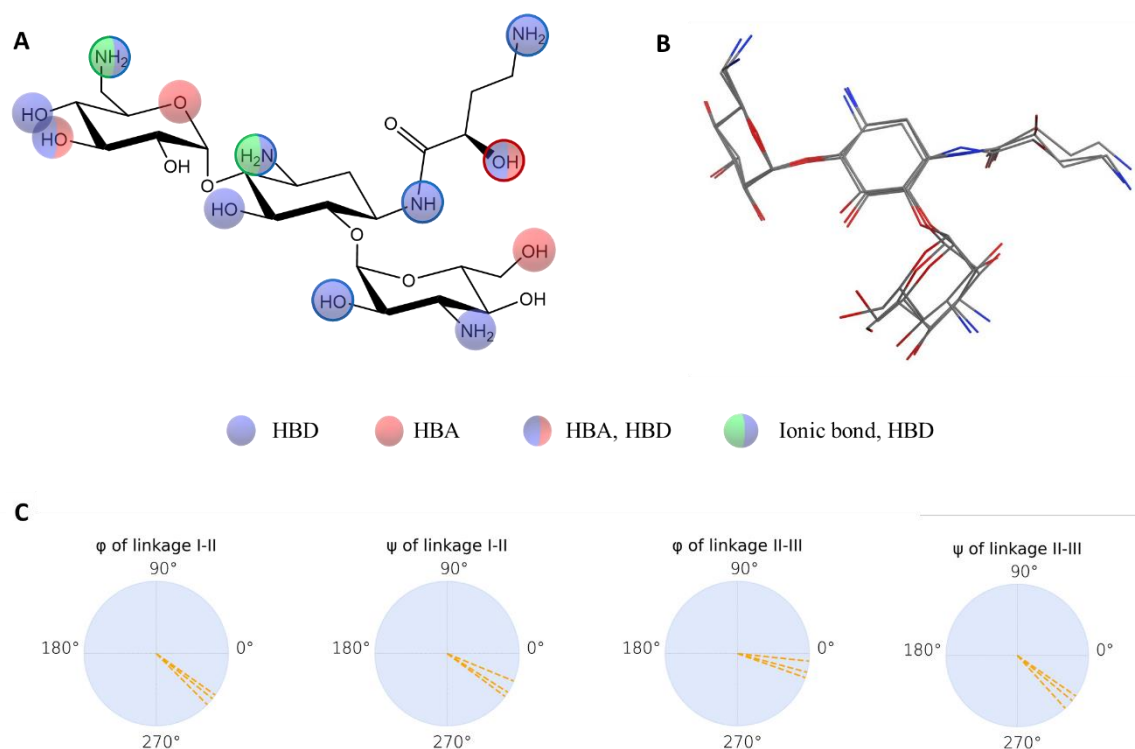


Figure II.2.4-1 2D and 3D representation of amikacin in RX-structures in complex with rRNA. **(A)** Ensemble of all amikacin-rRNA interactions. Outlined interactions are present in all RNA complexes. **(B)** amikacin 3D structures superimposed on rings I and II. **(C)** Comparison of amikacin conformers dihedral angles. Atoms forming dihedral angles were defined following reference 52: $\phi(\text{H1-C1-O1-Cx})$ and $\psi(\text{C1-O1-Cx-Hx})$.

in the binding through hydrogen bonds, and ionic bonds for two of them. Additionally, five out of seven hydroxyl groups formed hydrogen bonds with the rRNA (see Figure II.2.4-1A).

Furthermore, the conformation adopted by amikacin in all complexes had a maximum RMSD of 0.86 Å (see Figure II.2.4-1B). The comparison of the ring linkages dihedral angles revealed a high similarity between amikacin conformers and kanamycin A conformers when bound to rRNA (see Figure II.2.4-1C and Figure II.2.3-4).

The modification made to kanamycin A in order to produce amikacin could have suggested that the affinity towards rRNA could be disrupted due to the substitution of a key interacting group. Nevertheless, the presence of the C(1)-NH₂, which formed a stable hydrogen bond, was retained in amikacin. Also, the rest of the HABA chain maintained polar interactions observed in all three complexes (see Figure II.2.4-1A).

Hence, it can be hypothesized that the addition of the HABA chain introduced a steric hindrance to the binding with AMEs. Indeed, the C(1)-NH₂ was involved in the binding mode of all complexes AME/kanamycin A, except for AAC(6')-Ie (see Figure II.2.3-3).

Briefly, the substitution made in kanamycin A to form amikacin did not interfere with its binding to rRNA as the amino group at position C(1) and the HABA chain continued to establish stable interactions. However, this modification did disrupt the binding with AMEs. Although this conclusion could not have been obtained uniquely from our analysis, we can hypothesize the underlying reasons. Consequently, the next step would be to conduct a Buried Surface Area (BSA) analysis per atom to identify potential positions for further modifications that selectively hinder AME binding while preserving binding to rRNA.

II.3 Modelling the binding of neomycin B and its synthetic derivative HL₁₇₁ into five AMEs

II.3.1 Introduction

Numerous strategies have been employed to fight resistance to aminoglycosides. Considering that AMEs involve 60-70% of acute resistances, a key approach consists in chemically modifying AGAs to reduce their affinity for these enzymes.³⁹ Examples have been published showing the deletion or modification of one or more chemical groups, the synthesis of dimers, or the conjugation with a peptide, nucleotide, sugar, or another antibiotic.²⁴ However, these modifications may also lead to a decrease in affinity for the target site A.

We can cite work on amikacin as a successful example in the 2-DOS 4,6-disubstituted family of AGAs with its two derivatives arbekacin, and plazomicin, that are drugs used for the treatment of infections caused by multiresistant bacteria.^{40,57,58} By contrast, despite extensive attempts and investigating the modification of various positions, no new drugs have yet been derived from an AGA of the 2-DOS 4,5-disubstituted family, such as neomycin (see Figure II.3.1-1).^{24,39,59} Nonetheless, many of the tested modifications have provided valuable insights for future research in this field. For instance, acylation at position C(1), alkylation at positions C(6') and C(4'), and deoxygenation of ring I have shown beneficial effects. It is also necessary

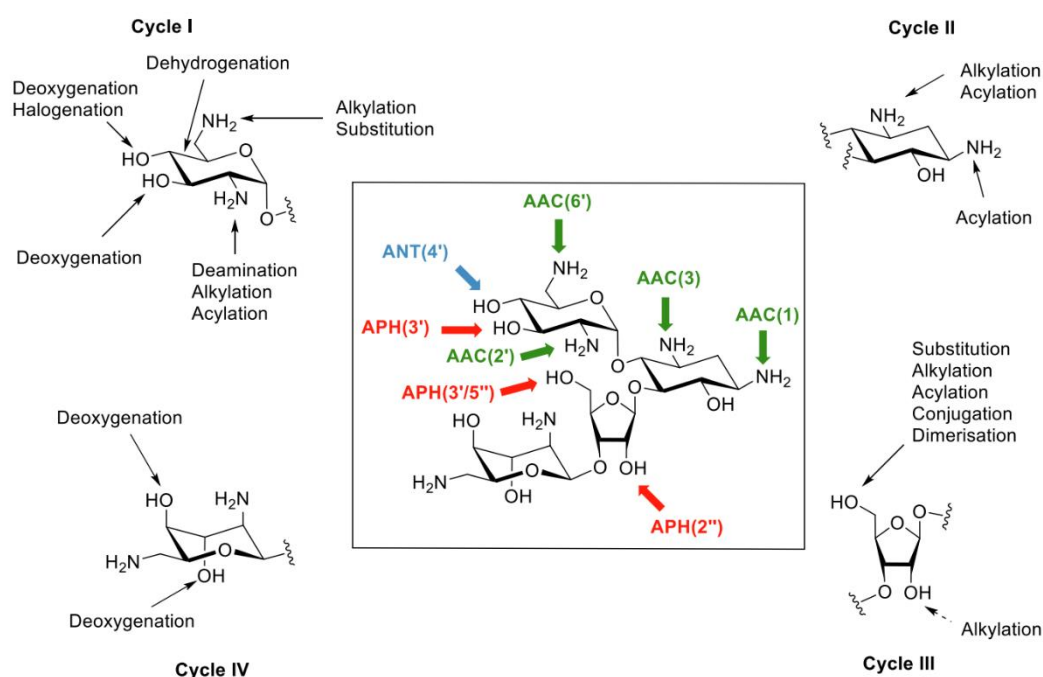


Figure II.3.1-1 Strategies developed on neomycin B to overcome AMEs resistance. Image from article ²⁴.

to maintain an amino group at position C(6') for antibacterial activity. Combining different modifications can be advantageous.²⁴

In continuation of the findings from previous modifications of aminoglycosides of the 2-DOS 4,5-disubstituted family, the primary objective of this work is to continue with the exploration and develop original derivatives of neomycin B (see Figure II.3.1-2A) that can overcome the effect of AMEs while preserving the bactericidal effect.

To do so, a transdisciplinary consortium between four laboratories was settled:

Pr. Weibel (LaSyROC, UMR 7177, CNRS-University of Strasbourg), responsible for the AGA derivative synthesis, leads the project. His experience with aminoglycosides dates back to 2003. He initially studied AGAs as molecules targeting VIH-1 RNA and transitioned to AGAs antibacterial activity a decade ago. Since 2018, Professor Weibel's team has focused on functionalizing neomycin B positions C(5'') and C(6'). They based their choice on the limited influence of this positions on the bacterial rRNA affinity, on the existence of AMEs targeting those positions (APH(3'/5''), AAC(6'), and AAC(6')/APH(2'')), and on the chemical reactivity of paromomycin. Hence, they synthesized 37 derivatives of neomycin B: 27 of them mono-substituted at position C(5'') and 10 di-substituted at positions C(5'') and C(6').

Dr. Prévost (Virulence bactérienne précoce, UR 7290, CNRS-University of Strasbourg) is responsible for conducting the *in vitro* testing of the AGAs and their derivatives. His group identified a total of 9 AMEs present in 10 clinical sets of *S. aureus* strains and 8 clinical sets of *E. coli* strains and evaluated the Minimum Inhibitory Concentration (MIC) of the synthesized molecules.

Dr. Ennifar (Architecture et réactivité de l'ARN, UPR 9002, CNRS-University of Strasbourg) focusses on assessing ribosomal binding using biophysical approaches (X-Ray crystallography and isothermal titration calorimetry). His team also worked with AGAs in the past, publishing two crystallographic structures of neomycin B bound to HIV-1 DIS RNA (PDB ID: 2FCY, 3C7R) in 2006 and 2008.

Finally, our laboratory (Laboratoire d'Innovation Thérapeutique, UMR 7200 CNRS-University of Strasbourg) focuses on molecular modelling. Our goal is to give computational support to understand and predict the impact of neomycin B modifications on its recognition by AMEs.

Among all the derivatives of neomycin B which have been synthesized and tested by the members of our consortium, the synthetic disubstituted derivative HL_171 (see Figure II.3.1-2B) showed a lower MIC than neomycin B in nine out of the eleven strains exhibiting resistance against neomycin B. Therefore, we modelled and compared the 3D-structures of the AMEs in complex with neomycin B and HL_171. Then, we examined their recognition by five AMEs: APH(3')-Ia, APH(3')-IIa, APH(3')-IIIa, ANT(4')-Ia, and AAC(3)-IIa. We predicted the binding mode of AGA to AME using docking and molecular dynamics simulations. Finally, we analyzed the productions focusing on AGA-AME binding modes, AGA distances to cofactors, and AGA distances to key catalytic residues in order to conclude on the effect of neomycin B functionalization on its recognition by AMEs.

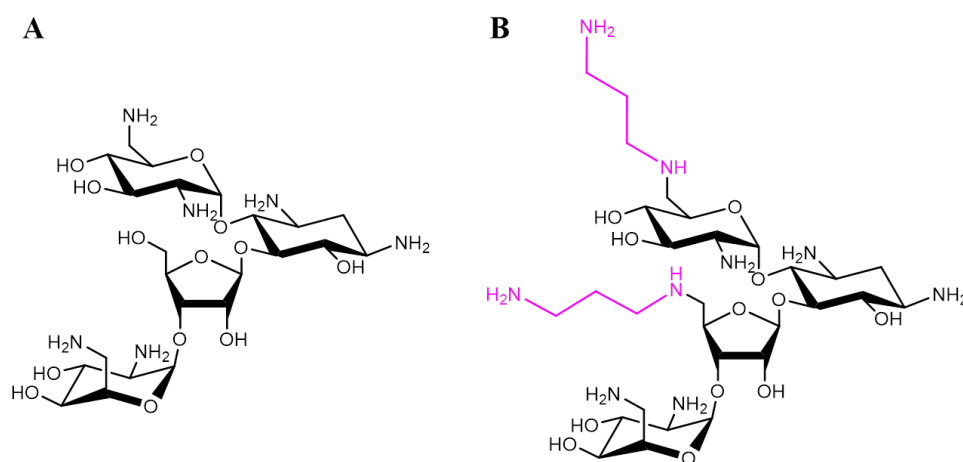


Figure II.3.1-2 Structures of (A) neomycin B and (B) HL_171. In pink the synthetic modifications.

II.3.2 Materials and methods

II.3.2.1 AME modelling

AME sequences were provided by Dr. Prévost. Their Genbank ID are:

- X13543 for AAC(3)-IIa from Plasmid pWP113a
- X04555 for ANT(2'')-Ia from Plasmid pDGO100
- V00359 for APH(3')-Ia from *E. coli*
- M20305 for APH(3')-Ib from Plasmid RP4
- V00618 for APH(3')-IIa from *E. coli*

- M28829 for APH(3'')-Ib from *E. coli*
- AB300568 for APH(3')-IIIa from *S. aureus*
- M19465 for ANT(4')-Ia from *S. aureus*
- CP065515 for AAC(6')-Ie/APH(2'')-Ia from *S. aureus*

The sequences were used as queries to search for the templates in the Protein Data Bank (PDB)⁴⁵.

If the searched AME was found in the PDB (100% sequence identity covering the full length of the protein), the structural issues such as partial occupancies, uncapped breaks, missing disulfide bridges, topology disagreements, incorrect charges or incorrect number of hydrogens were corrected using MOE (versions 2019.01, 2020.09, and 2022.02, Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7) using the forcefield Amber10:EHT. The “Protonate3D” tool was used to add missing hydrogen atoms to the complex. Water molecules forming at least one interaction with the aminoglycoside and two with the protein were extracted and saved in separated MOL2 files for the binding mode analysis.

If the searched AME was not found in the PDB, a homology model was built using MOE (versions 2019.01, 2020.09, and 2022.02, Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7). The homolog search was performed by Basic Local Alignment Search Tool (BLAST).⁶⁰ Preference was given to PDB structures including an AGA or a cofactor. The sequence of the protein to be modelled was aligned to its crystal homolog using the “Align/superpose” tool of the protein module in MOE (default settings). The three-dimensional model was built using the “Homology Model” tool of the protein module (default settings), generating ten models ranked based on the GB/VI scoring method. The model with the lowest score was saved in MOL2 format. The crystallographic proteins with higher similarity and containing the missing cofactor or AGA were identified using BLAST.⁶⁰ The pocket residues of both enzymes, defined as residues with at least one atom within a maximum distance of 4.5 Å to the aminoglycoside, were superposed using the “Align/superpose” tool in MOE. The missing cofactor and/or AGA were copied from the reference to the modelled structure. Water molecules forming at least one interaction with the AGA and two with the AME were identified and saved in MOL2 files. The structure was prepared and protonated using MOE.

II.3.2.2 Docking of the AGAs

Neomycin B derivative HL_171 was built from neomycin B by manual 3D edition from PDB file 2B0Q. The structure was minimized using MOE “Builder” and “Energy minimize” using the forcefield Amber10:EHT. Ligands atom types defined as “difficult groups” were manually corrected to match those expected in the standard MOL2 format. Namely, sp nitrogen atom has to be represented by N.1 atom type, sp² by N.2, tetrahedral neutral sp³ by N.3, positively charged sp³ by N.4, trigonal planar sp² with delocalized lone pair by N.pl3, and planar amide nitrogen by N.am.

Neomycin B and HL_171 were docked into the AMEs using GOLD Suite (v5.3, CCDC, Cambridge, United Kingdom),⁶¹ a program basing the docking poses search and optimization on a genetic algorithm. A “scaffold matching” constraint was used to guide the placement of a fragment of the molecule onto the ring II position of the aminoglycoside. It included all carbon and nitrogen atoms, and the hydrogen atoms of C(1) and C(3) (see Figure II.3.2-1). The reference conformation of ring II was taken from the crystallographic structure of the complex between the aminoglycoside and the AME:

- PDB ID: 2B0Q for APH(3’)-IIIa
- PDB ID: 1ND4 for APH(3’)-IIa
- PDB ID: 4EJ7 for APH(3’)-Ia
- PDB ID: 6NML for ANT(4’)-Ia
- PDB ID: 6O5U for AAC(3)-IIa

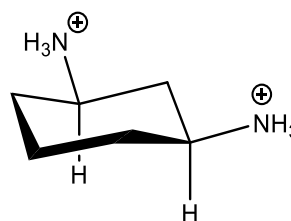


Figure II.3.2-1 Structure of the scaffold constraint

This constraint is enforced at the mapping stage in GOLD and thus, does not contribute to the fitness score to bias docking. The ligand placements are generated using a best least-squares fit with the scaffold heavy atom positions.

For the docking, a configuration file was written for each ligand. The “scaffold matching” constraint was set to 5.0, the scoring function to CHEMPLP, and an early termination of docking was allowed if the root-mean-square-deviation (RMSD) of the coordinates of the 3 top solutions was lower than 1.5 Å. The “toggle” option was activated allowing the algorithm to decide on the consideration of the water molecules in the binding site: included if the docking fitness score increased, excluded if it decreased. Default settings were maintained for the other parameters. The software outputted a maximum of ten poses per docked molecule.

a Scoring, ranking and evaluation of docking poses

For each AME, a redocking experiment was performed to assess the docking performance.

The similarity of the placement of ring I between the crystal structure and the docking poses was evaluated with the RMSD computed from non-hydrogen atom coordinates using Surflex-Dock (v3.066, BioPharmics LLC, California, United States).⁶²

The docking poses were further evaluated based on characteristic distances between the targeted atom of the docked aminoglycoside and key elements for the metabolization in the AME. Distances were calculated using the RDKit library (RDKit: Open-source cheminformatics, <https://www.rdkit.org>, <https://doi.org/10.5281/zenodo.7415128>). The key elements were selected depending on their importance for the catalysis, namely:

- OD1 or OD2 of Asp190, β -P of ADP, and Mg^{2+} for APH(3')-IIIa.
- OD1 or OD2 of Asp190, γ -P of ATP, and Ca^{2+} for APH(3')-IIa.
- OD1 or OD2 of Asp198, γ -P of ATP, and Ca^{2+} for APH(3')-Ia.
- OE1 or OE2 of Glu145, OE1 or OE2 of Glu76, α -P of AMPCPP, and Mg^{2+} for ANT(4')-Ia.
- NE2 of His189 and S of CoA for AAC(3)-IIa.

II.3.2.3 Molecular Dynamics (MD) simulations

The crystal structure and model of the AME/AGA complexes were submitted to molecular dynamics (MD) simulations to assess the stability of the AGA binding mode. In addition, AME/neomycin B data was taken as control to analyze AME/HL_171 and conclude about the potential effect of the AGA modification on its recognition by the AME.

a Data preparation

Preparation of ligands

The cofactors, and neomycin and HL_171 best docking pose were extracted from the crystal structure or modelled complexes and prepared individually. The antechamber program was used with the general AMBER force field GAFF2 to generate the topology files. The atomic point charges of the small organic molecules were calculated with the charge method AM1-

BCC. The method parmchk was used to include the missing force field parameters to the output file. Finally, the LEaP program generated the lib, topology, and coordinate files for each ligand.

Preparation of protein

The crystal structure or modelled protein were used to prepare the protein structure for the MD simulations. The pdb4amber tool was run to analyze and fix the PDB file. The "--reduce" option was activated to add hydrogen atoms. Regarding histidine residues, the protonation was manually set to maximize the number of intramolecular hydrogen bonds. The residue name was set to HID for a histidine with a protonated δ -N, to HIE for a protonated ϵ -N and to HIP for when both nitrogen atoms were protonated.

Preparation of complex

AMBER LEaP program was used to prepare the MD input files from the structure files. The general libraries for standard protein residues (ff14SB), organic molecules (GAFF2), water residues (TIP3P) were loaded. The files "frcmod.ions234lm_126_tip3p" and "atomic_ions.lib" were used to interpret metals such as Mg^{2+} . Based on the total charge of the protein, cofactor and ligand, the system for MD simulation was neutralized by adding counterions (Na^+ or Cl^-) with the option "addIons". An octahedral water box with a radius of 10.00 Å was added using the command "TIP3PBOX". Finally, the tool "addIonsRand" was used in order to set the ion concentration, based on the volume of the octahedron box, so the salt molarity in the system is similar to the one of cells: 0.15 M.

The final parameters were saved in a topology and coordinates file: prmtop and inpcrd formats respectively.

b Molecular Dynamics simulations

The program sander was used for the minimization step and pmemd.cuda for the heating, equilibration, and production steps using AMBER18 (2018, University of California, San Francisco)⁶³ and CUDA 10.1⁶⁴ toolkit. Simulations were performed at the Institut national de physique nucléaire et de physique des particules (IN2P3) computing center using GPU Nvidia® Tesla® V100 processors.

The systems were first minimized following a two-stages procedure. The first energy minimization was performed with a restraint of 1 kcal/(mol.Å²) on all atoms except hydrogen atoms and that of water molecules. The second one had no restraints applied. Both stages underwent 500 steps of steepest descent minimization followed by 500 steps of conjugate gradient minimization. Then, the systems were heated from 0 K to 300 K during 200 ps with a weak constraint of 0.5 kcal/(mol.Å²) applied to all atoms but that of aminoglycosides. The heating was controlled by the Langevin thermostat and led under constant volume. Followingly, an equilibration step of 20 ns at constant pressure allowed to reach a stationary state. Three replicas of 200 ns were then performed for each system under constant pressure and Langevin thermostat. The SHAKE algorithm was used to constrain all bonds involving hydrogen atoms.

Trajectories analysis

The deviation of the atomic coordinates during the equilibration and productions stages was evaluated by calculating the RMSD using CPPTRAJ (v.4.14.0, University of Utah, United States).⁶⁵ All RMSD calculations used the initial minimized state as reference structure. A total of four RMSD calculations were conducted for each complex in this order:

- RMSD of α -C of the fitted binding site residues.
- RMSD of non-hydrogen atoms of the AGA rings I and II (“nofit” option activated).
- RMSD of non-hydrogen atoms of the fitted AGA.
- RMSD of α -C of the fitted AME.

The non-covalent interactions between the AGA and AME were determined using an in-house python code. Hydrogen bonds were considered for a distance cutoff of 3.5 Å and an angle between -60° and 60°, ionic bonds for a distance cutoff of 4 Å, and hydrophobic contacts for a distance cutoff of 5 Å.

The dihedral angles of the linkage between the rings of the aminoglycoside were determined using CPPTRAJ as described by Corzana et al.: $\varphi(\text{H1-C1-O1-Cx})$ and $\psi(\text{C1-O1-Cx-Hx})$.⁵²

II.3.3 Results

II.3.3.1 Protocol

The objective of this study was to develop a tool able to predict the impact of neomycin B modifications on its recognition by AMEs. Therefore, we worked on the proof of concept by focusing on neomycin B and its derivative HL_171. This derivative, synthesized by Dr. Weibel laboratory, showed the best MIC results on the clinical strains of *E. coli* and *S. aureus* identified by Dr. Gilles Prévost's laboratory.

Structural insights into the complexes formed by neomycin B or HL_171 binding to five AMEs were obtained from MD simulations. The binding mode of neomycin B was used as reference to determine whether the substitutions introduced in HL_171 could potentially impact its binding to the AMEs.

Therefore, a five-step protocol was implemented (see Figure II.3.3-1):

- I) The three-dimensional structures of the AMEs in complex with the AGA and the cofactor were generated using homology modelling and corrected and protonated using MOE.
- II) A constrained docking approach was employed on ring II of neomycin B and HL_171. The poses were ranked based on the RMSD of ring I compared to the reference crystallographic AGA.
- III) MD simulations of the best docking pose were performed with three production replicas of 200 ns.
- IV) The evaluation of the stability and binding mode was done by measuring the RMSD, the key distances, and the frequency of non-covalent interactions.
- V) A conclusion determining whether the modifications made in neomycin B to obtain the derivative HL_171 had an influence on the metabolization was made.

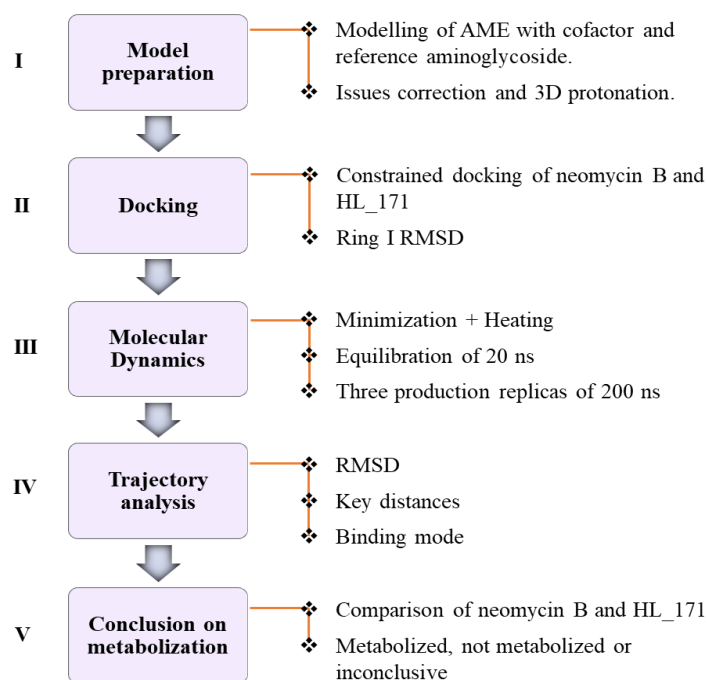


Figure II.3.3-1 Protocol to model and study the AME/AGA complexes.

II.3.3.2 Selection of AMEs and derivative to model

To obtain the three-dimensional structures of the complexes, the initial stage involved selecting the appropriate AMEs for modelling. This was done by combining the data acquired from the collaborators with the existing crystallographic structures available in the PDB.

A total of 18 clinical strains (8 *E. coli* and 10 *S. aureus*) were evaluated by polymerase chain reaction (PCR) to detect genes encoding 9 different AMEs. The 8 and 10 strains of *E. coli* and *S. aureus* species were compared to reference strains of the American type culture collection (ATCC), which are recommended for antibiotic susceptibility testing, namely ATCC 25922 and ATCC 25923, respectively.^{66,67}

All the strains were used to test the antibacterial activity of neomycin B and HL_171 on multidrug-resistant (MDR) bacteria, evaluated with minimal inhibitory concentration (MIC) assays based on the monitoring of bacterial growth. Neomycin B MIC results were used as a control (see Table II.3.3-1 and Table II.3.3-2).

Results reported in Table II.3.3-1 and Table II.3.3-2 showed that all but one clinical strains contained at least one AME. Up to four AMEs were identified in *E. coli* strains whereas a single AME was detected in *S. aureus* strains. Despite the presence of AMEs, four *E. coli* strains and two *S. aureus* strains were as sensitive to neomycin B as the reference ATCC strain.

Table II.3.3-1 Presence of AMEs in and sensitivity to neomycin B of clinical *E. coli* strains. Red shading highlights MIC values higher than the control. Bold indicates HL_171 MIC lower than neomycin B MIC.

	EC ATCC 25922	EC1	EC2	EC3	EC4	EC5	EC6	EC7	EC8
AAC(3)-IIa		X		X		X	X	X	X
ANT(2'')-Ia			X						
APH(3')-Ia			X					X	
APH(3')-Ib			X	X		X	X	X	X
APH(3')-IIa					X	X			X
APH(3'')-Ib			X	X		X	X	X	X
MIC (mg/L) Neomycin B	3	1	64	2	2	32	2	> 256	4
MIC (mg/L) HL_171	4	4	16	4	4	64	4	> 128	4

Table II.3.3-2 Presence of AMEs in and sensitivity to neomycin B of clinical *S. aureus* strains. Red shading highlights MIC values higher than the control. Bold indicates HL_171 MIC lower than neomycin B MIC.

	SA ATCC 25923	SA1	SA2	SA3	SA4	SA5	SA6	SA7	SA8	SA9	SA10
ANT(4')-Ia			X	X			X				X
APH(3')-IIIa								X	X	X	
AAC(6')-Ie / APH(2')-Ia		X				X					
MIC (mg/L) Neomycin B	0.8	1	64	64	1	1	64	256	> 256	32	32
MIC (mg/L) HL_171	1	2	8	8	1.5	1.5	8	16	16	4	4

All the clinical strains containing ANT(2'')-Ia, APH(3')-Ia, ANT(4')-Ia and APH(3')-IIIa could metabolise neomycin B. The AMEs AAC(3)-IIa, APH(3')-Ib APH(3')-IIa, and APH(3'')-Ib were detected in strains resistant to neomycin B but also in others sensitive to it, suggesting that the metabolization activity was due to other AMEs. In contrast, none of the *S. aureus* strains containing AAC(6')-Ie/APH(2')-Ia were able to metabolise neomycin B.

Interestingly, HL_171 showed reduced MIC as compared to neomycin B on 2 *E. coli* strains and 7 *S. aureus* strains suggesting that the action of ANT(2'')-Ia, APH(3')-Ia, ANT(4')-Ia and APH(3')-IIIa on neomycin B may be prevented by the modification of two positions in HL_171.

Briefly, all but two AMEs were considered for the modelling study. APH(3'')-Ib was indeed directly excluded from the analysis because neither neomycin B nor HL_171 have a hydroxyl group on the position C(3'') to be metabolized. AAC(6')-Ie/APH(2'')-Ia was discarded based on neomycin B MIC (see Table II.3.3-2).

The next conditions for modelling were based on template criteria as it follows:

- Resolution ≤ 3 Å.
- Sequence identity ≥ 25 %.
- Presence of aminoglycoside and/or cofactor in an AME template.

Table II.3.3-3 presents a summary of the PDB structures that satisfied at least one of the mentioned criteria and provided new structural data, such as an aminoglycoside or cofactor. All AMEs met the similarity and quality criteria, with all PDB references having a sequence identity ≥ 33 % and a resolution ≤ 2.70 Å. The local accuracy was systematically evaluated and will be discussed in the AMEs specific subsections.

Table II.3.3-3 Summary of PDB entries of the selected AMEs or homologs. In green, the AMEs that pass the template criteria. In blue the elements used for the modelling. KAN corresponds to kanamycin A, NMY to neomycin B, and LLL to gentamycin C1.

AME TO MODEL			TEMPLATE AME					
AME	Genbank code	Source organism	PDB references	Resolution (Å)	Sequence identity (%)		AGA	Cofactor
					All	Binding site		
AAC(3)-IIa	X13543	Plasmid pWP113a	6O5U	1.40	46.5	71.4	KAN	/
			6MN1	2.25	34.3	21.4	LLL	CoA
ANT(4')-Ia	M19465	<i>S. aureus</i>	6NML	2.00	99.4	100.0	NMY	AMPCPP 2 Mg ²⁺
			6UN8	1.65	100.0	100.0	NMY	/
			1KNY	2.50	99.4	100.0	KAN	AMPCPP 2 Mg ²⁺
ANT(2'')-Ia	X04555	plasmid pDGO100	4WQL	1.73	71.1	100.0	KAN	2 Mg ²⁺
APH(3')-Ia	V00359	<i>E. coli</i>	4GKH	1.86	99.6	100.0	KAN	/
APH(3')-Ib	M28829	<i>E. coli</i>	4EJ7	2.29	99.6	100.0	/	ATP, 2 Ca ²⁺
					57.3	100.0		
APH(3')-IIa	V00618	<i>E. coli</i>	1ND4	2.10	100.0	100.0	KAN	/
			4EJ7	2.29	33.3	58.3	/	ATP, 2 Ca ²⁺
APH(3')-IIIa	AB300568	<i>S. aureus</i>	2B0Q	2.70	100.0	100.0	NMY	ADP, 2 Mg ²⁺

In contrast, ANT(2'')-Ia did not pass the last condition as none of the similar crystallographic structures had a cofactor. APH(3')-Ib was also excluded as its PDB reference was the same as the one used for APH(3')-Ia, and the sequence identity was closer to the latter.

Therefore, five AMEs, namely AAC(3)-IIa, ANT(4')-IIa, APH(3')-Ia, APH(3')-IIa, and APH(3')-IIIa, were selected for modelling as they passed the template conditions. Regarding ANT(4')-Ia, the PDB 6NML was chosen as reference because it contained both neomycin B and the cofactor.

In brief, we modelled, examined, and compared the 3D structures of the complexes formed by neomycin B and HL_171 with five selected AMEs. The results obtained for each AME are individually detailed in the next sections of the chapter.

We firstly present the work done on APH(3')-IIIa, which was crystallized in complex with neomycin B and ADP, thus providing a reliable structural basis for modeling. We then continue with the other two phosphotransferases APH(3')-IIa and APH(3')-Ia, as their binding site is similar to APH(3')-IIIa. The next AME of interest is ANT(4')-Ia, which also has a complete crystallographic template. Finally, we detail the work on AAC(3)-IIa. Although this enzyme was present in two strains showing bacterial resistance against neomycin B (see EC5 and EC7 in Table II.3.3-1), the literature suggests that it does not metabolize the 4,5-aminoglycoside types.²³ Therefore, we considered this enzyme as a negative control.

II.3.3.3 APH(3')-IIIa

a About APH(3')-IIIa: catalytic mechanism and AGA binding mode

The enzyme APH(3')-IIIa, also called aminoglycoside 3'-phosphotransferase type IIIa, catalyzes the transfer of the γ -phosphate group from ATP to a free hydroxyl group. This phosphorylation specifically targets the hydroxyl groups located at position C(3') of kanamycin, neomycin, paromomycin, livostamycin, butirosin, amikacin, ribostamicin, gentamicin, and isepamicin; at position C(5'') of lividomycin; and at positions C(3') and C(5'') of butirosin and neomycin B leading to a di-phosphorylation.²³

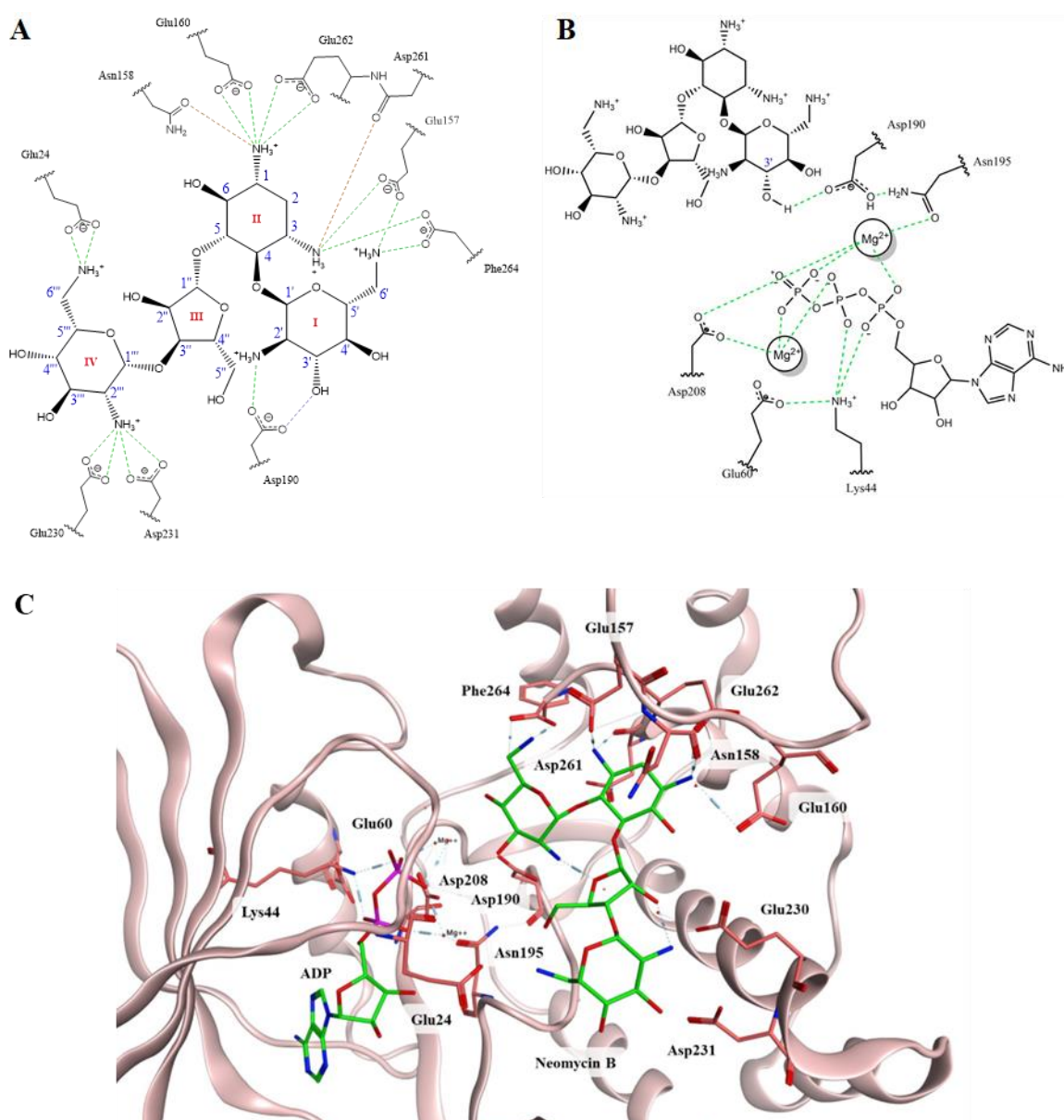


Figure II.3.3-2 Different representations of the active site of APH(3')-IIIa. (A) Hydrogen bonds (green dashed lines) and ionic bonds (orange dashed lines) between neomycin B and APH(3')-IIIa. (B) Key interactions (green dashed lines) involved in the catalysis.⁶⁸ (C) 3D view of the active site. (PDB ID: 2B0Q)

As shown in Figure II.3.3-2A, the targeted hydroxyl group C(3')-OH only binds to the residue Asp190, which is essential for both metal ion and aminoglycoside positioning. Moreover, Asp208 coordinates the two Mg²⁺ cofactors and is very important for the formation and stabilization of the transition state.⁶⁸ Lastly, the interactions between Lys44 and the α and β -phosphate groups allowed the adenosine triphosphate (ATP) to have an optimal position enabling the transfer of its γ -phosphate to the C(3')-OH of neomycin B.

Furthermore, the APH(3')-IIIa – ADP – neomycin complex (PDB ID: 2B0Q)⁶⁹, in which the ATP cofactor was replaced by ADP to avoid an instantaneous phosphorylation of the AGA, showed a highly negative charged binding pocket formed by residues that are acidic in nature (3 aspartic acids, 5 glutamic acids and 1 C-terminal carboxylic acid). The fact that the binding site mainly provides acceptor groups for hydrogen bond interactions is coherent with the nature of aminoglycosides, which are positively charged molecules (see Figure II.3.3-2A). The complex also showed that most of the binding interactions between the AME and the neomycin occur with rings I and II.

Based on this analysis, we selected the following distances to characterize the stability of the AGA in the catalytic site:

- **d1:** AGA C(3')-oxygen atom – Asp190 OD1/OD2 atom.
- **d2:** AGA C(3')-oxygen atom – ADP β -phosphorus atom.
- **d3:** AGA C(3')-oxygen atom – Mg²⁺.

b Experimental data used for modelling and AME/AGA model

APH(3')-IIIa has been characterized in a total of seven crystallographic structures: in its apoenzyme form, and in complex with: ADP; AMPPNP; ADP and kanamycin A; ADP and neomycin B; AMPPNP and butirosin A; and the inhibitor AR 3A (PDB IDs: 1J7I, 1J7L, 1J7U, 1L8T, 2B0Q, 3TM0, and 2BKK, respectively).⁴⁵

The APH(3')-IIIa of the PDB file 2B0Q shared 100% identity with the template sequence. Consequently this structure was directly used for the structural study. The structure was deposited in 2002, had a resolution of 2.7 Å and contained one chain with the crystallographic structure of the AME sourced from *Enterococcus faecalis* bacteria, a molecule of neomycin B, adenosine diphosphate (ADP) and two Mg²⁺ ions.

The geometry analysis of the structure revealed three Ramachandran outliers. Two of them (Arg5 and Thr163) were located on two different loops on the surface of the enzyme, more than 6.5 Å away from the binding pocket of the neomycin. The third outlier was the catalytic Asp190, which formed hydrogen and ionic bonds with neomycin B. A total of three clashes were also detected in neomycin B binding site, namely Glu160, Ser227, and Glu262. No bond length or bond angle outliers were detected.

Interestingly, the B-factor indicated a better definition of rings I and II compared to rings III and IV, which in addition formed fewer interactions with the AME (see Figure II.3.3-3). This feature was also observed in the other AME/AGA structures (see Structural analysis of neomycin B and kanamycin A binding AMEs and bacterial ribosomal RNA).

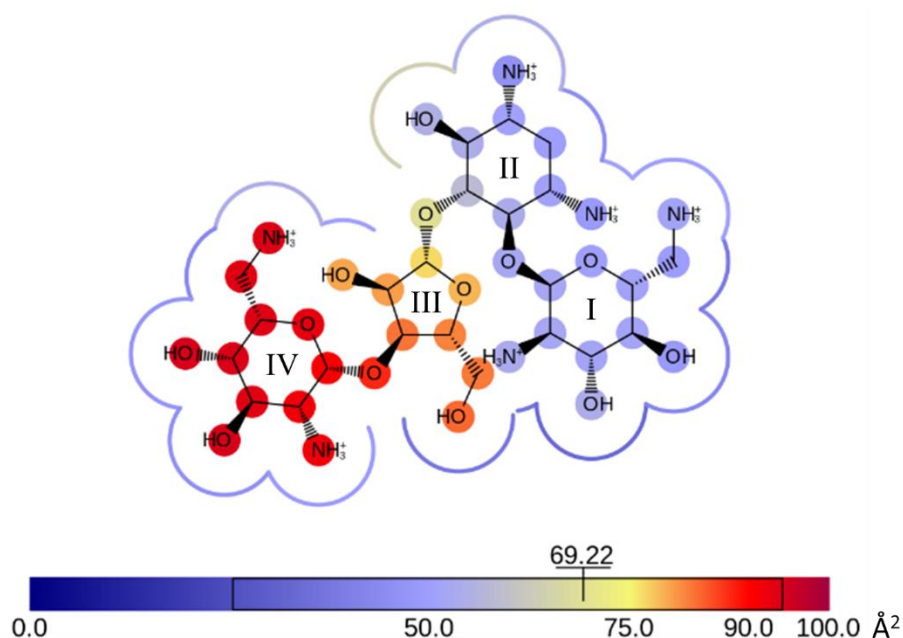


Figure II.3.3-3 B-factor values (\AA^2) of neomycin B atoms and its binding site in the PDB entry 2B0Q. The blue color indicates a smaller B-factor and thus a higher precision. The line surrounding the AGA represents the B-factor of the residues at less than 4.5 Å of neomycin B.

Globally, neomycin-binding pocket was well defined. The low resolution together with the detail given by the B-factor and the absence of length and angle bond outliers confirmed the quality of the structure. The following minimization step and MD simulations was expected to solve the observed clashes and dihedral angles exceptions. Briefly, the template 2B0Q was adequate to study the AGA binding mode and to model the binding mode of HL_171.

The redocking neomycin B under different conditions confirmed the ability of the docking program to predict an accurate pose and established the optimal docking conditions. Docking the neomycin B without constraint back into its co-crystal protein output one pose with a low RMSD for ring I (0.9 Å). However, multiple incorrect solutions were also generated, with maximum and average RMSD values of 3.9 Å and 2.3 Å, respectively. In contrast, when the docking was constrained at ring II, it issued a consistent and accurate ensemble of docking poses (RMSD maximum: 1.7 Å, average: 1.3 Å). Hence, the docking successfully predicted the correct placement of neomycin B on AME, making it suitable for docking HL_171.

Comparing the HL_171 poses to neomycin B crystal structure, the lowest RMSD were 1.6 Å and 1.1 Å if all atoms and only ring I were considered, respectively. The placement of ring I showed a slight shift compared to the crystallographic pose, resulting in a longer distance between the targeted C(3')-OH and the Asp190 (**d1**) (3.6 Å vs 3.1 Å). As expected, rings III and IV showed a bigger displacement. Overall, the key distances **d1**, **d2**, and **d3** were conserved in the APH(3')-IIIa/neomycin B crystal structure and in the docking poses of neomycin B and HL_171.

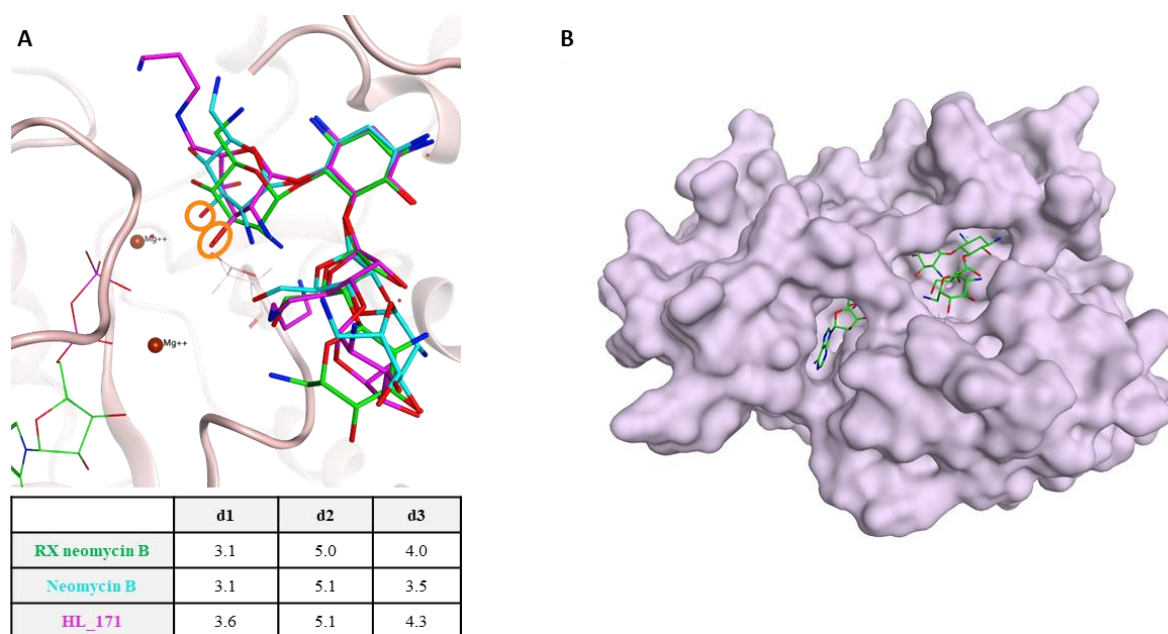


Figure II.3.3-4 Three-dimensional view of APH(3')-IIIa and key distances measurement of (A) the bound RX neomycin B (green), and the docking poses of neomycin B (blue) and HL_171 (fuchsia). (B) Global view of the crystallographic AME/AGA complex.

c AME/AGA MD simulations

After the minimization, heating, and 20 ns equilibration period, we conducted three production replicas (3 x 200 ns). The stability of the complexes was assessed by monitoring the RMSD of the α C of the protein residues, as well as the RMSD of the binding site backbone atoms. The conformational changes in the AGA were evaluated through the fitted AGA RMSD while the displacements within the cavity were observed through the AGA rings I and II RMSD fitted to the binding site (see Figure II.3.3-5).

APH(3')-IIIa/neomycin B equilibration was reached before production. The AGA showed almost no variations during the equilibration with a RMSD <1 Å, showing that the input conformation was close to the energy minimum (see Fitted aminoglycoside in Figure II.3.3-5). The binding site conformation also remained close to that of initial state with a RMSD ≤ 2 Å (see Fitted binding site in Figure II.3.3-5).

HL_171 also formed stable complexes when bound to APH(3')-IIIa. As for neomycin B, HL_171 binding site kept an RMSD ≤ 2 Å during the three productions. Nevertheless, the HL_171 conformation showed a higher variation than neomycin B with a fitted AGA RMSD close to 2 Å (see Figure II.3.3-5). Moreover, one of HL_171 replicas revealed a repositioning of rings I and II within the cavity at 150 ns (see HL_171 rings I and II RMSD in Figure II.3.3-5).

Both neomycin B and HL_171 formed stable complexes, enabling the comparison between their binding modes. To do so, we identified the hydrogen bonds, ionic bonds, and hydrophobic contacts between the AGAs and specific AME residues (see Figure II.3.3-6).

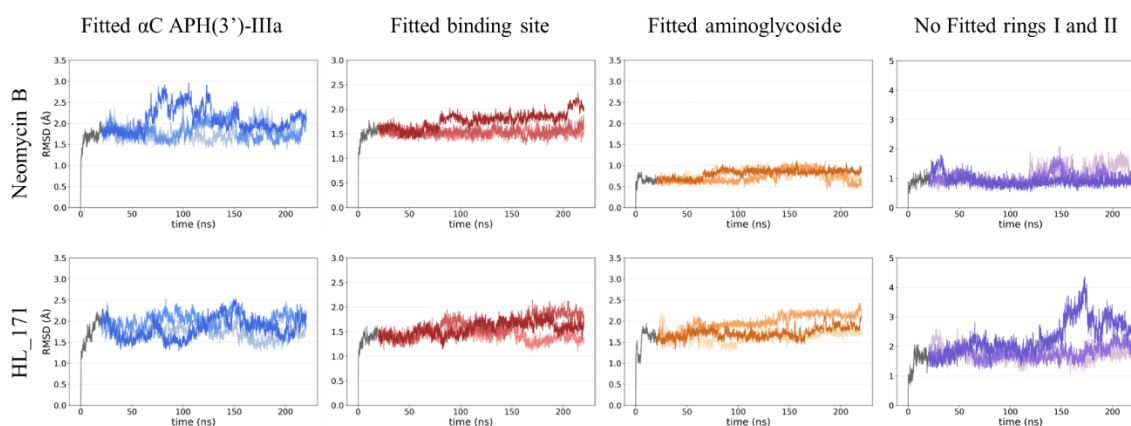


Figure II.3.3-5 RMSD plots of neomycin B and HL_171 in complex with APH(3')-IIIa. The reference is the minimized structure. The grey part corresponds to the equilibration phase. The colored parts represent the three productions. The lighter color corresponds to the first replica and the darker one to the third.

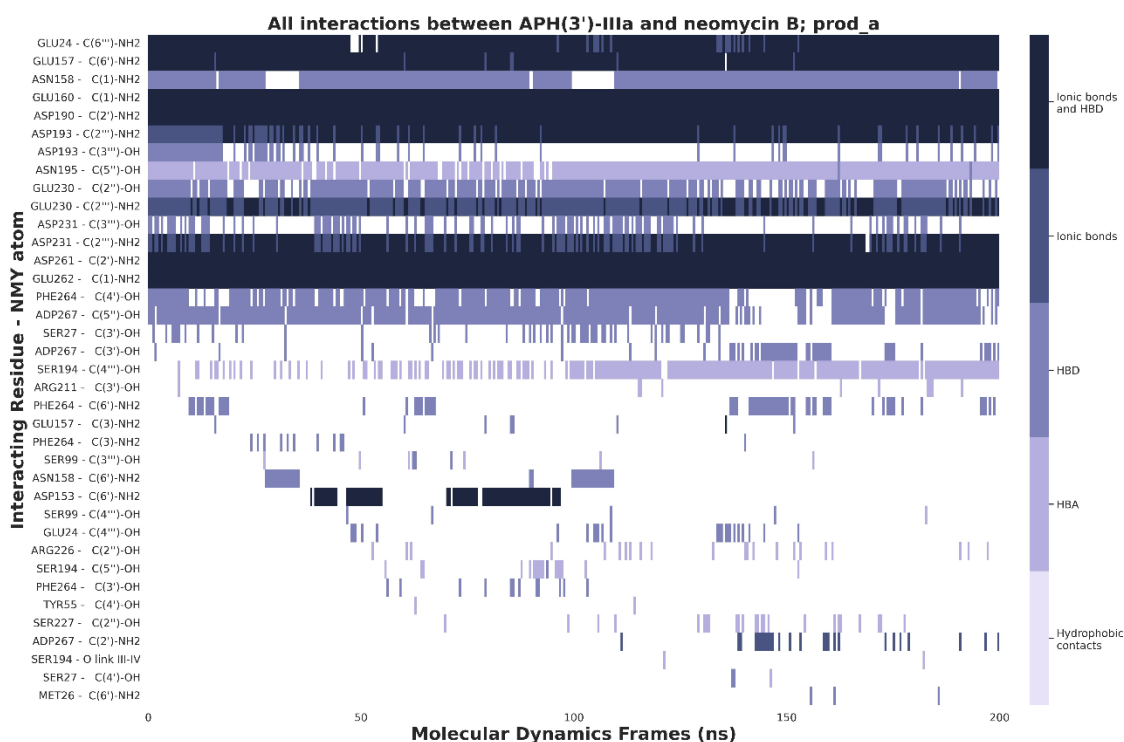


Figure II.3.3-6 Interaction profiles between neomycin B and APH(3')-IIIa during the first production.

The interaction profiles of the other MD simulations are provided in the annexes VI.1. Neomycin B binding mode was defined by thirteen interactions, mainly ionic, maintained over the entire simulation. Although the figure provided a clear visualization of the binding variations occurring during the simulations, the analysis and comparison of the binding modes was challenging. Therefore, we conducted an analysis of various complexes and determined a frequency threshold of 70% to distinguish between sporadic interactions and persistent ones (see Figure II.3.3-7).

From the previously identified thirteen interactions, eleven were responsible for the recognition in all three replicas (see Figure II.3.3-7A), including ionic bonds with the amino groups at positions C(1), C(2'), C(6'), and C(2''') and hydrogen bonds with C(5'')-OH.

Next, we searched for these specific interactions in the APH(3')-IIIa/HL_171 complex (see Figure II.3.3-7). The binding mode adopted by neomycin B bound to APH(3')-IIIa was different from the one adopted by HL_171 (see Figure II.3.3-7C). Only the interactions implying C(1)-NH₂ were present for both residues in all production replicas. HL_171 binding mode was mainly stabilized by four of its groups, including the two amino groups synthetically introduced by C(6')-NH₂ and C(5'')-OH substitutions (see outlined interactions in Figure II.3.3-7B and Figure II.3.3-7C).

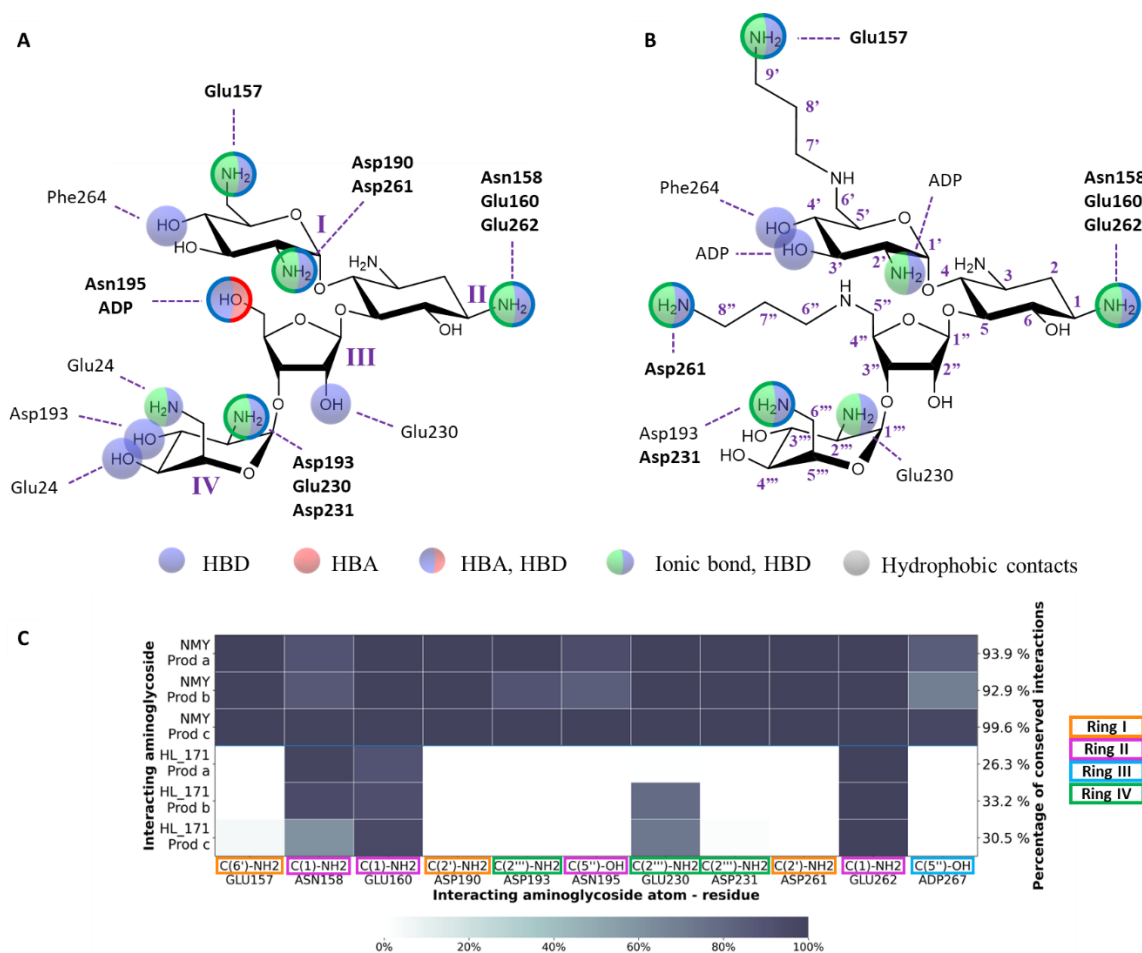


Figure II.3.3-7 Stable interactions formed between APH(3')-IIIa and neomycin B or HL_171 during the MD simulations and present in more than 70% of the frames. Outlined interactions are present in all three productions. (A) Annotations on neomycin B. (B) Annotations on HL_171. (C) Frequency of interactions between the neomycin B (NMY) or HL_171 and APH(3')-IIIa in each replica (prod_a, prod_b, and prod_c).

The conformational analysis followed the same approach as described in section II.2 Structural analysis of neomycin B and kanamycin A binding AMEs and bacterial ribosomal RNA. We calculated both phi and psi angles for both neomycin B and HL_171 ring linkages during the MD frames (see Figure II.3.3-8).

Neomycin B torsion angles shown in Figure II.3.3-8 were steady for all three simulations and had the same range of values as the crystallographic structures from previous chapter (see section II.2.3.1 Neomycin B in complex with AMEs and RNA.). Nevertheless, HL_171 presented a higher conformational variety. The three replicas showed different values for both ϕ and ψ angles of the linkage of rings III and IV, resulting in different conformers occupying the same subsite.

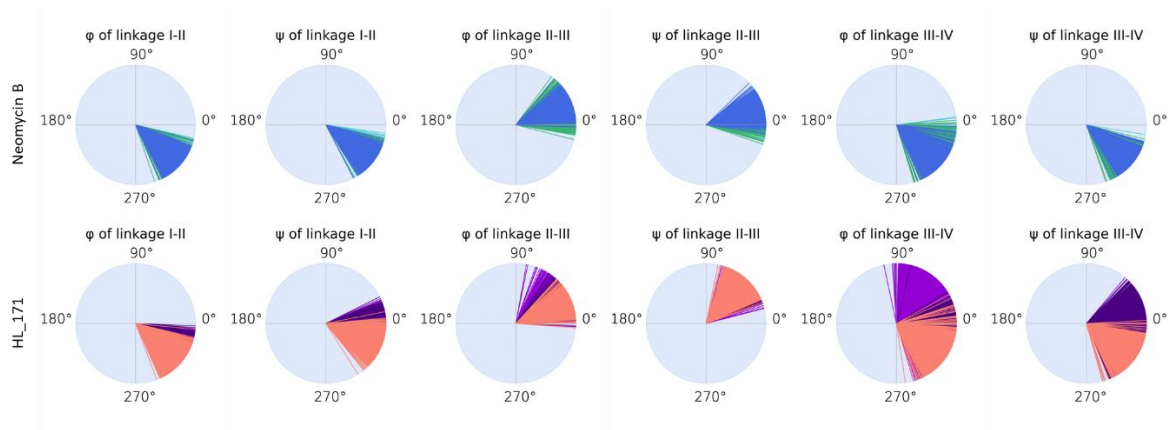


Figure II.3.3-8 Neomycin B and HL_171 dihedral angles during the MD simulations. Each color represents one of the production replicas.

Figure II.3.3-9 illustrates previous observations by representing the complexes formed by neomycin B and HL_171 complexes at the last frame of the simulations. Neomycin B showed a consistently low displacement from the crystallographic pose throughout all simulations. The distances to all key elements were consistent among the three productions and similar to the crystallographic ones (see Figure II.3.3-9B). Only the second replica showed a higher variation towards the distance to ADP (**d2**) which was due to the movement of the cofactor.

In contrast, HL_171 displayed a higher variation from the initial docking pose, but eventually reached a stable conformation and pose. The repositioning observed in the RMSD analysis was also observable when assessing the distances between the targeted C(3')-OH of HL_171 and the key elements involved in its metabolization. Nevertheless, all HL_171 replicas demonstrated distinct distances compared to those observed in the crystallographic neomycin B complex (see Figure II.3.3-9C).

In conclusion, the modifications introduced in HL_171 caused a relocation of the AGA in the protein, resulting in a different binding mode from that of neomycin B and not competent for catalysis. This finding agrees with the biological data, which showed that HL_171 inhibits the resistant strains SA7, SA8, and SA9 at lower dose compared to neomycin B.

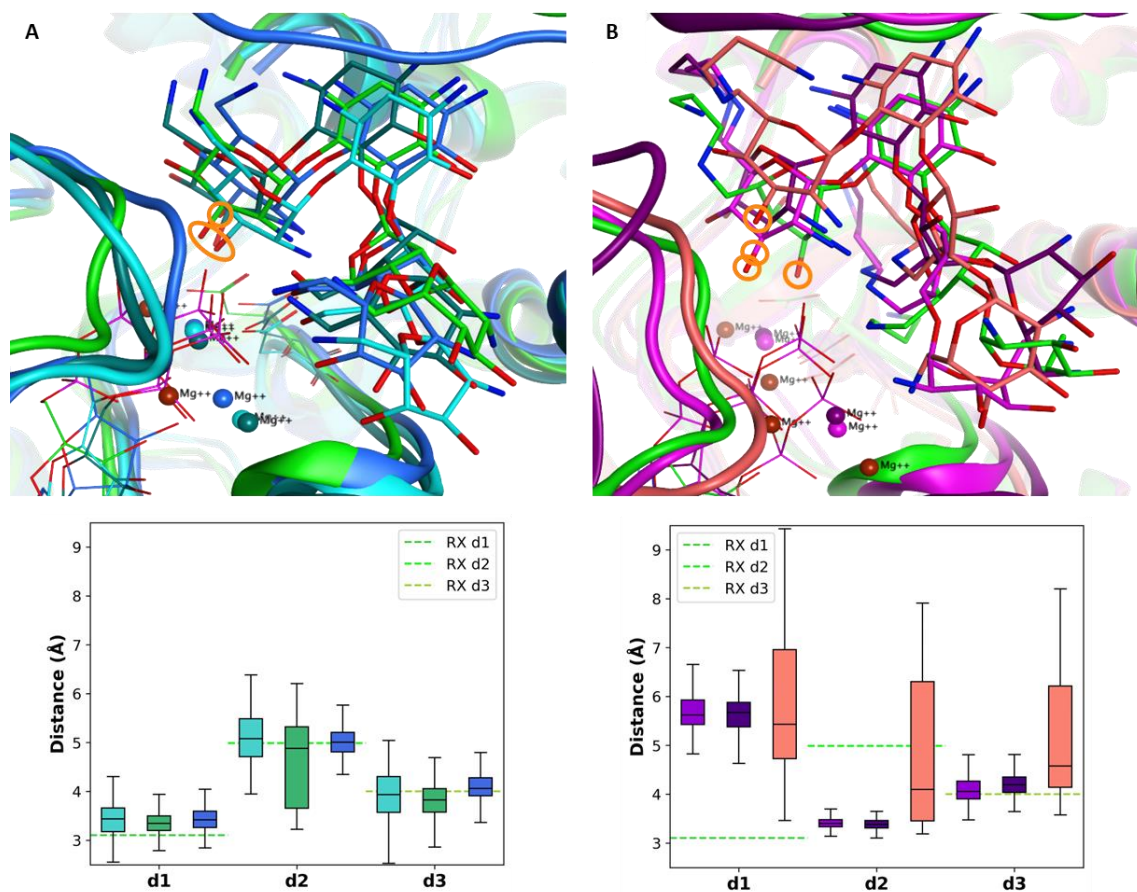


Figure II.3.3-9 3D structures of the last MD simulations frame and key distances of APH(3')-IIIa in complex with (A) neomycin B (blues) and (B) HL_171 (purples). For the sake of comparison, the initial structure is shown in green. The targeted position of the aminoglycosides is marked in orange. With **d1**: AGA C(3')-oxygen atom – Asp190 OD1/OD2 atom; **d2**: AGA C(3')-oxygen atom – ADP β -phosphorus atom; **d3**: AGA C(3')-oxygen atom – Mg²⁺.

II.3.3.4 APH(3')-IIa

a About APH(3')-IIa: catalytic mechanism and AGA binding mode

The APH(3')-IIa, or aminoglycoside 3'-phosphotransferase type IIa, catalyzes the transfer of the ATP γ -phosphate group to a free hydroxyl group at position C(3'). The enzyme shows a resistance profile to kanamycin, neomycin, butirosin, paromomycin, and ribostamycin.²³

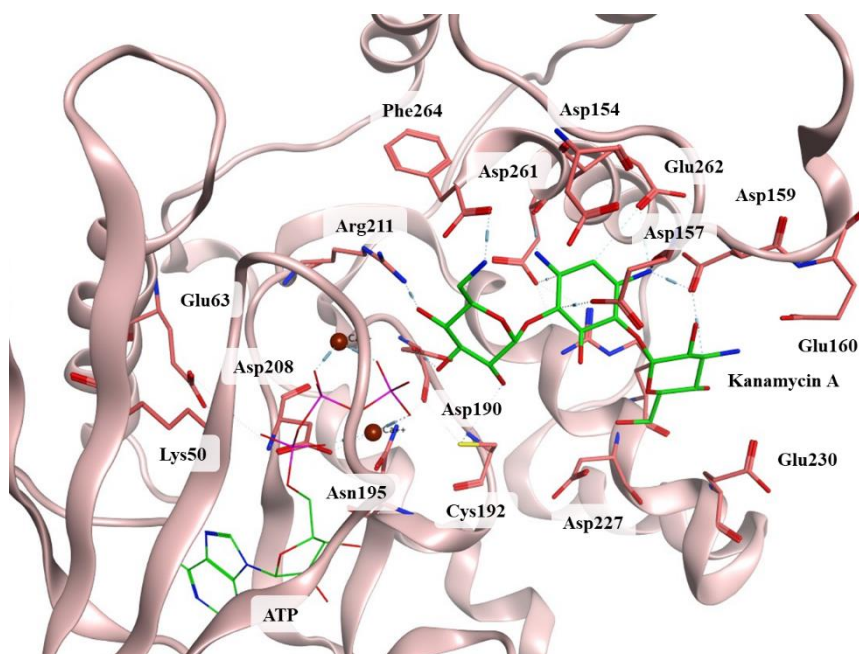


Figure II.3.3-10 3D view of APH(3')-IIa active site (PDB ID: 1ND4).

APH(3')-IIa fold has a high degree of similarity to that of APH(3')-IIIa, with an RMSD value of 2.0 Å for 30% of sequence identity.⁷⁰ The catalytic triad required for APH(3')-IIa catalytic activity was identified as identical to that of APH(3')-IIIa (see Figure II.3.3-2B), with Asp190 being responsible for C(3')-OH activation and Asn195 coordinating Asp190 and Mg^{2+} . Following the same principle, Asp208 was in charge of coordinating the two metallic cofactors, here Ca^{2+} ; Lys50 of ATP binding; and Glu63 of Lys50 orientation (see Figure II.3.3-10).⁷⁰

Therefore, following the same reasoning explained in the previous section, we selected the following distances to study the active site:

- **d1:** AGA C(3')-oxygen atom – Asp190 OD1/OD2 atom.
- **d2:** AGA C(3')-oxygen atom – ATP γ -phosphorus atom.
- **d3:** AGA C(3')-oxygen atom – Ca^{2+} .

b Experimental data used for modelling and AME/AGA model

To date, the PDB file 1ND4 is the only crystallographic structure available that characterizes APH(3')-IIa. The structure shared 100% identity with the template sequence used for the PCR identification of AMEs. Therefore, it was directly used for the study.

The PDB entry 1ND4 was deposited in 2002, had a 2.10 Å resolution and was sourced from *K. pneumoniae*. It contained two chains, a molecule of kanamycin A, and two Mg²⁺ ions. Chain A had less geometrical outliers and was thus used as the AME template. Nevertheless, the template lacked the necessary cofactor for metabolization. Therefore, we aligned the ATP binding site of the crystal structure with the one of the PDB entry 4EJ7, which had 59% of sequence identity, and incorporated the ATP and Ca²⁺ cofactors into the model. The Mg²⁺ ions from the template 1ND4 were removed from the final 3D structure.

Chain A did not present any geometrical outliers except for the backbone bond angles of the residues Gly95 and Gly99, which were located at the surface of the AME and at 3.1 Å and 6.6 Å of the ATP binding site, respectively. Furthermore, the structure presented one internal clash of the residue Glu160, part of kanamycin A binding site, and two clashes between the ATP and Ca²⁺.

The B-factor analysis indicated a good definition of all three kanamycin A cycles. The AGA average B-factor value was 44.9 Å², and the binding site one was 37.4 Å².

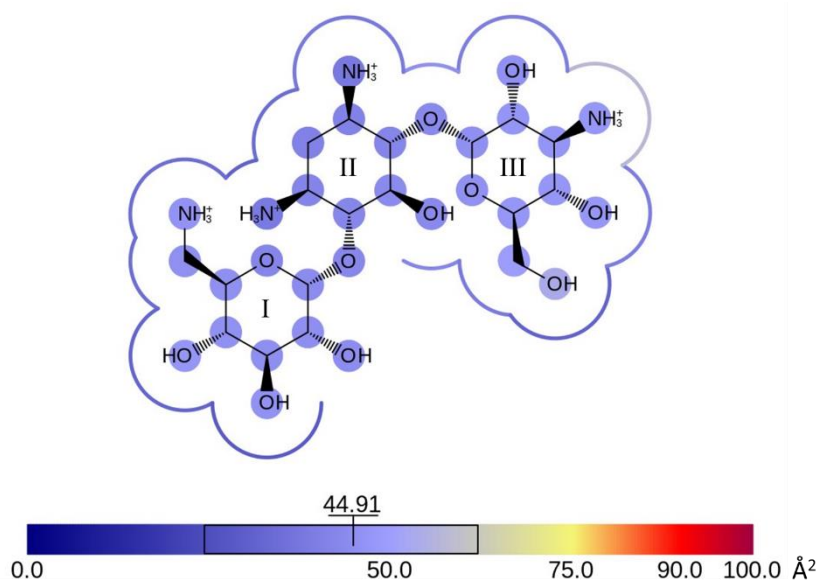


Figure II.3.3-11 B-factor values (Å²) of kanamycin A atoms and its binding site in the PDB entry 1ND4. The blue color indicates a smaller B-factor and thus a higher precision. The line surrounding the AGA represents the B-factor of the residues at less than 4.5 Å of kanamycin A.

Briefly, the active site proposed by the PDB entry 1ND4 is well defined and can be used for modelling neomycin B and HL_171 and analyzing their respective binding modes.

During docking, kanamycin A was used as reference to settle the constraint and evaluate the docking poses. Neomycin B docking positioned the targeted atom C(3')-O at the exact same distance of Asp190 (**d1**) compared to the crystallographic structure of kanamycin A. The distances **d2** and **d3** showed a slight increase of 0.2 and 0.5 Å respectively. Actually, the positioning of neomycin B ring I was similar to the template with an RMSD of 1.1 Å.

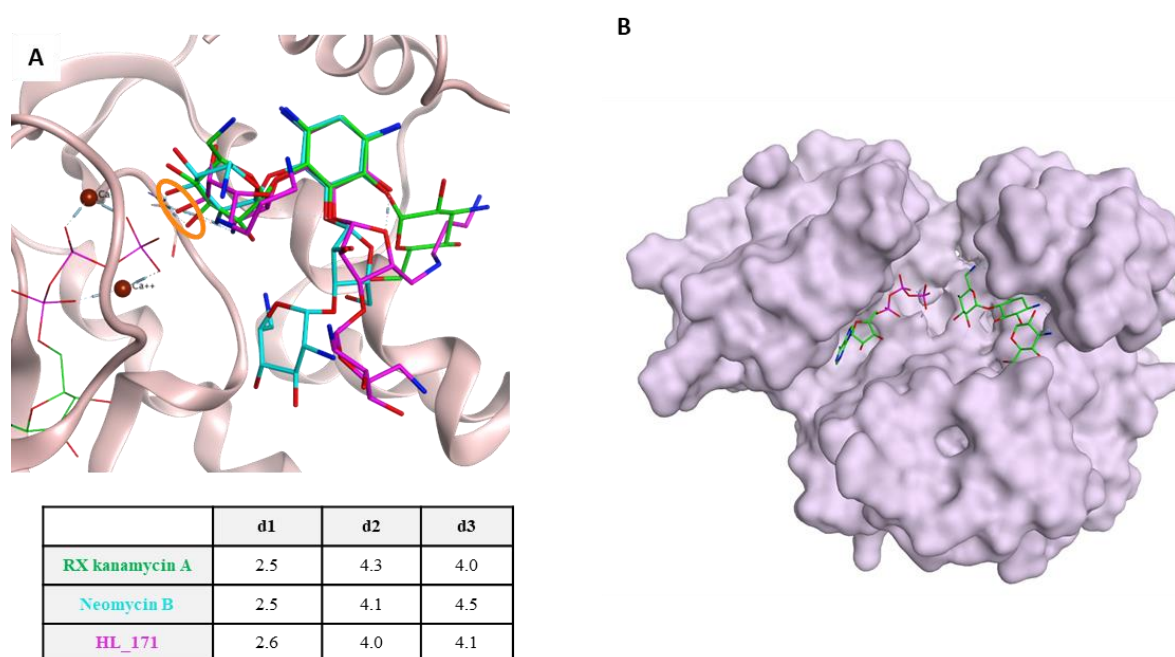


Figure II.3.3-12 Three-dimensional view of APH(3')-IIa and key distances measurement of (A) the bound RX of kanamycin A (green), and the docking poses of neomycin B (blue) and HL_171 (fuchsia). (B) Global view of the crystallographic AME/AGA complex.

The docking positioned HL_171 ring I in a similar location to that of the template, with an RMSD of 1.5 Å. Additionally, the distance **d1** of HL_171 compared to kanamycin A was only 0.1 Å greater. The distances **d2** and **d3** were also very similar, with differences of 0.2 Å and 0.1 Å, respectively, compared to kanamycin A (see Figure II.3.3-12).

Because of the structural differences between kanamycin A and neomycin B, the location of rings III and IV shifted for both AGAs during the docking. The shift illustrated the width of the subsite, which allowed for the mobility of these rings, as discussed on section II.2 Structural analysis of neomycin B and kanamycin A binding AMEs and bacterial ribosomal RNA.

Briefly, the docking of both neomycin B and HL_171 output adequate three-dimensional structures allowing to continue with the MD evaluation.

c AME/AGA MD simulations

The complex APH(3')-IIa/neomycin B reached equilibration before production, with the AGA conformation showing almost no variations compared to the initial state (RMSD ~ 1 Å), and the binding site also staying close to the initial configuration (RMSD < 1.5 Å) (see Figure II.3.3-13). Nevertheless, the RMSD evaluation of the productions exposed a conformational change of the binding site and the neomycin B. This variation was induced by the approach of the loop formed by the residues Gln27 to Ala34 to the ligand and caused a repositioning of rings I and IV.

The complex formed by HL_171 and APH(3')-IIa also presented an RMSD similar to the initial minimized structure during the equilibration phase. The binding site conformational change previously observed for neomycin B, also occurred for HL_171 complex with smaller alterations. Finally, the AGA RMSD changes depicted on the “Fitted aminoglycoside” plot (see Figure II.3.3-13) were mainly due to the synthetically substituted chains and did not exceed an RMSD of 1.75 Å.

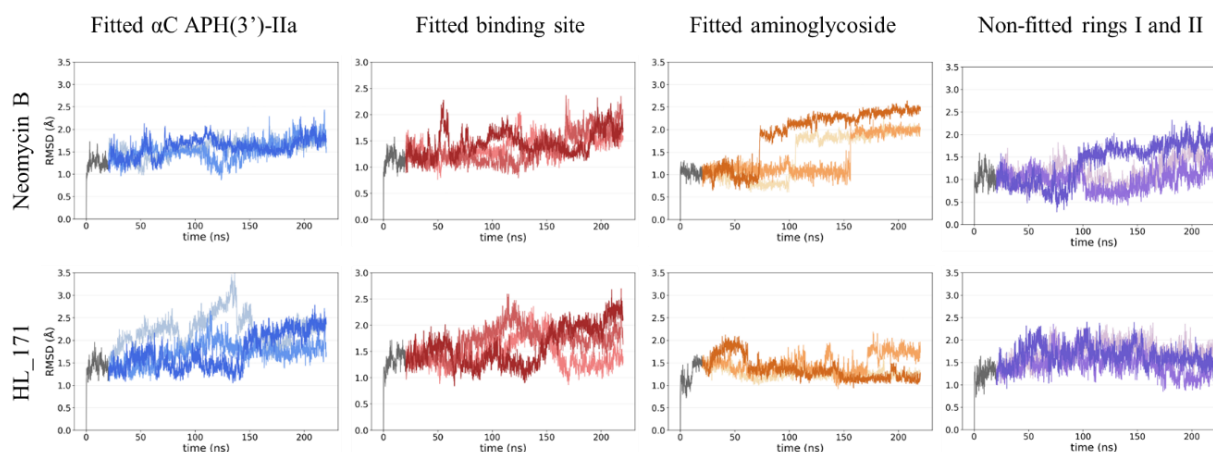


Figure II.3.3-13 RMSD plots of neomycin B and HL_171 in complex with APH(3')-IIa. The reference is the minimized structure. The grey part corresponds to the equilibration phase. The colored parts represent the three productions. The lighter color corresponds to the first replica and the darker one to the third.

Rings I and II of both neomycin B and HL_171 were relocated during the replicas, but their RMSD did not exceed 2.5 Å, suggesting a recognition of both AGAs by AME despite the variations of the binding site and the AGA conformation.

As shown in Figure II.3.3-14, neomycin B formed four interactions that remained stable during the three production replicas, involving three amino groups binding two aspartic acids, one glutamic acid and ATP. Interestingly, the previously noted conformational changes could also be observed through the interactions occurring during the simulations (Annexes VI.1.7 - VI.1.9). As the simulations progressed, the number of interactions formed by neomycin B

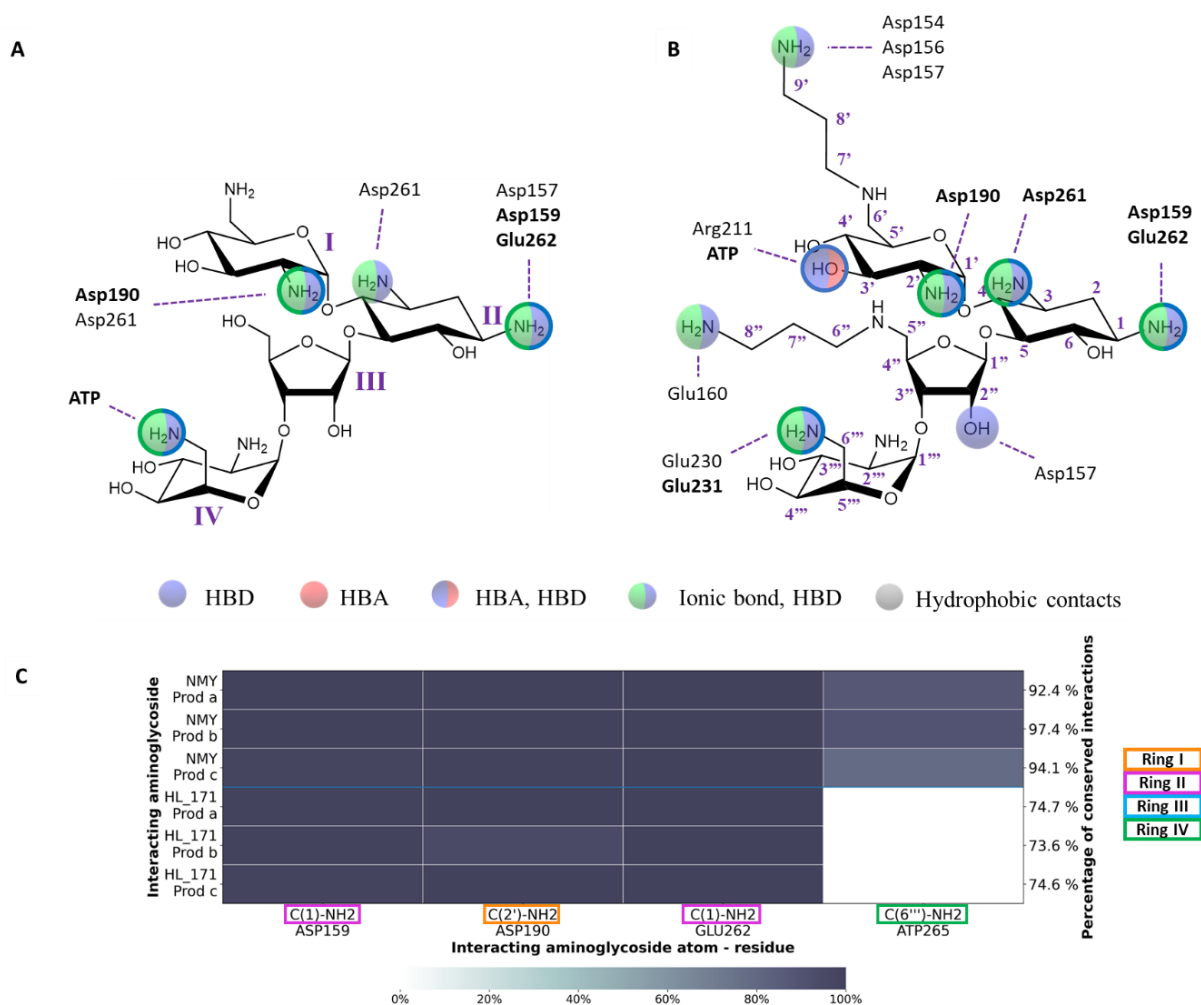


Figure II.3.3-14 Stable interactions formed between APH(3')-IIa and neomycin B or HL_171 during the MD simulations and present in more than 70% of the frames. Outlined interactions are present in all three productions. (A) Annotations on neomycin B. (B) Annotations on HL_171. (C) Frequency of interactions between the neomycin B (NMY) or HL_171 and APH(3')-IIa in each replica (prod_a, prod_b, and prod_c).

increased. Actually, for all simulations the interaction formed by C(2'')-OH and Asp157 were lost while a minimum of six interactions became stable, principally involving rings III and IV.

HL_171 and neomycin B binding modes were similar with the exception of the interaction with ATP, which involved C(3')-OH and C(6''')-NH₂, respectively (see Figure II.3.3-14). Of note, C(3')-OH is the position targeted by the AME. Finally, in addition to the interactions in common with neomycin B, HL_171 included the groups C(3')-OH and C(3)-NH₂ to its binding profile.

Briefly, neomycin B RMSD variations could be attributed to the repositioning of rings III and IV until the formation of persistent interactions. HL_171 shared three out of the four interactions defining neomycin B binding mode and involved two other groups to its binding mode.

Neomycin B conformation analysis exposed a wide range of conformations across the three productions (see Figure II.3.3-13 and Figure II.3.3-15). Specifically, the linkage between rings II and III displayed a significant variability, with both ϕ and ψ differing by 180° and 135°, respectively. Moreover, ψ angle coverage of the linkage between rings I and II was greater than 90°, indicating the presence of a binding site allowing the neamine part to adopt diverse conformations.

In contrast, HL_171 displayed a narrower range of dihedral angles. However, the synthetic modifications added to the C(5'') position resulted in an inverted ring III that differed from all

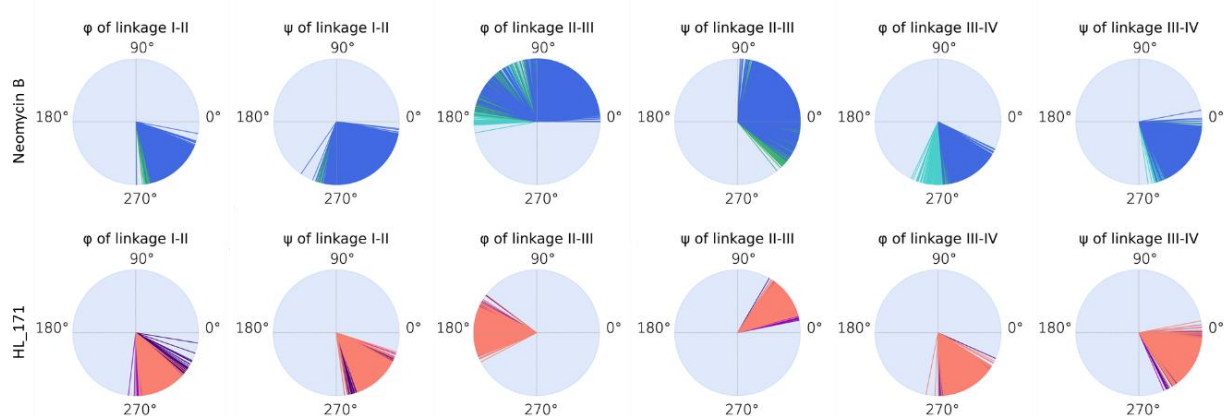


Figure II.3.3-15 Neomycin B and HL_171 dihedral angles during the MD simulations. Each color represents one of the production replicas.

the crystallographic structures presented in section II.2.3.1 showing neomycin B in complex with AMEs and RNA.

Finally, Figure II.3.3-16 illustrated the results of the docking and MD simulations. Interestingly, even though neomycin B adopted different conformations during the productions, the final frames resulted in similar complexes that kept a regular distance to the key elements (**d1**, **d2**, and **d3**) during the three runs. The distances **d1** and **d2** were similar to those of the crystallographic complex (see Figure II.3.3-16B). The distances **d3** increased for all three replicas due to a shift of Ca^{2+} in line with ATP motion.

In contrast, HL_171, the distances **d1**, **d2**, and **d3** showed a higher variation towards that of the crystal structure. Yet, the distance to the catalytic residue Asp190 was still favorable for forming hydrogen bonds in two of the replicas (**d1** < 3.5 Å).

In summary, HL_171 formed a stable complex with APH(3')-IIa with a binding mode similar to that of neomycin B. The two synthetically added chains contributed to the binding mode by

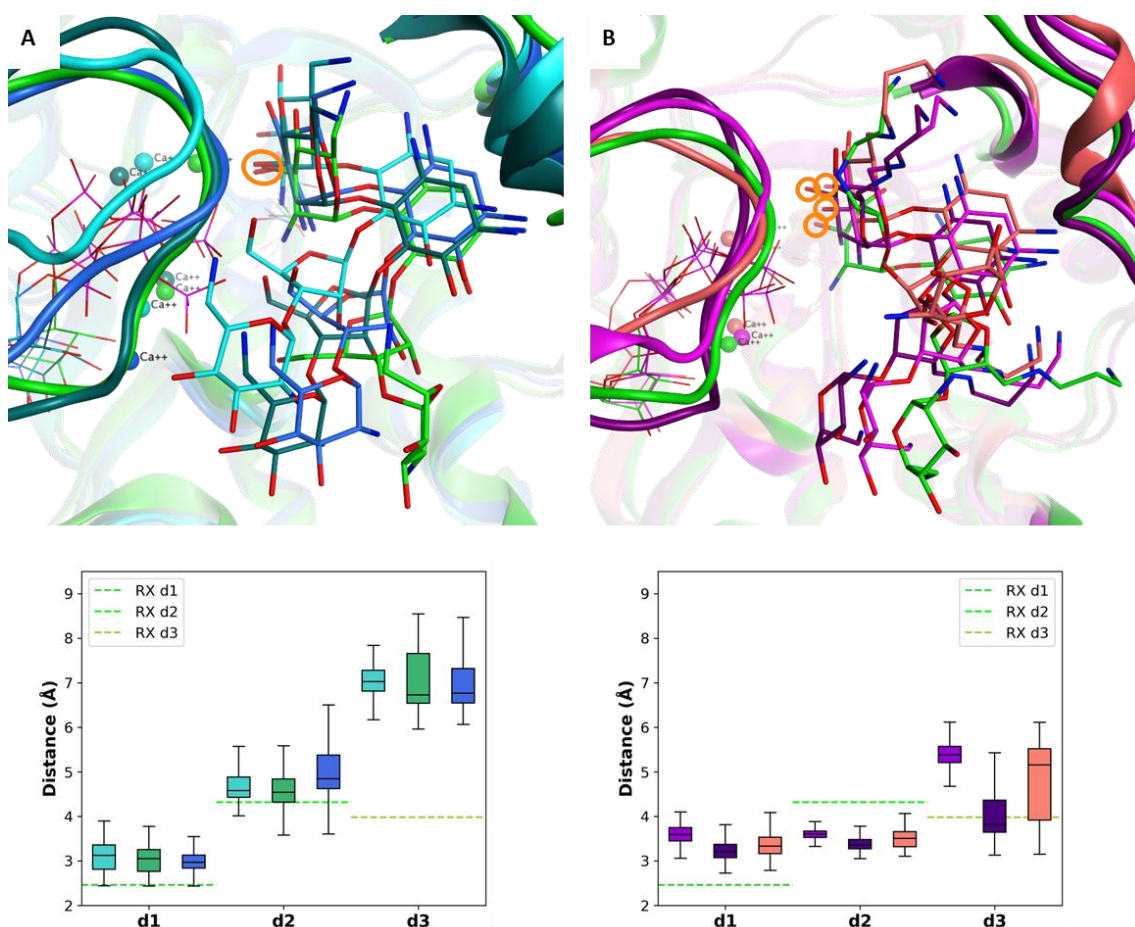


Figure II.3.3-16 3D structures of the last MD simulations frame and key distances of APH(3')-IIa in complex with (A) neomycin B (blues) and (B) HL_171 (purples). For the sake of comparison, the initial structure is shown in green. The targeted position of the aminoglycosides is marked in orange. With **d1**: AGA C(3')-oxygen atom – Asp190 OD1/OD2 atom; **d2**: AGA C(3')-oxygen atom – ATP γ -phosphorus atom; **d3**: AGA C(3')-oxygen atom – Ca^{2+} .

interacting with different aspartic and glutamic acids but did not introduce significant changes able to disrupt neomycin B binding mode. According to the results, HL_171 would still be metabolized by the AME APH(3')-IIa.

APH(3')-IIa was identified in three clinical *E.coli* strains, with two of them displaying resistance to neomycin B (see Table II.3.3-1). Interestingly, the MIC of HL_171 showed a 2-fold increase for two of the enzymes (EC4 and EC5), indicating reduced efficacy against the resistant strains. However, the MIC remained unchanged for the third enzyme (EC8). These findings align with the conclusions drawn from the molecular modeling analysis, suggesting

that the synthetic modifications made to neomycin B did not prevent recognition of the AGA by the AME.

II.3.3.5 APH(3')-Ia

a About APH(3')-Ia: catalytic mechanism and AGA binding mode

APH(3')-Ia, or aminoglycoside 3'-phosphotransferase type Ia, catalyzes the transfer of ATP or GTP γ -phosphate group to the hydroxyl group at position C(3') of kanamycin, neomycin, paromomycin, ribostamycin, and lividomycin.²³

APH(3')-Ia adopts the same fold as other phosphotransferases, including the previously studied AMEs. Its binding site is located at the convergence of the two structural lobes, which is analogous to the configuration depicted in Figure II.3.3-2.⁷¹ Indeed, the six key catalytic residues of APH(3')-IIIa are strictly conserved in APH(3')-Ia.⁶⁸ In APH(3')-Ia, Asp198 activates the targeted C(3')-OH while coordinated by Asn203, which is also in charge of coordinating one of the Mg²⁺ cofactors. Likewise, Asp216 coordinates the two Mg²⁺ ions, and Lys55 interacts with ATP, while its positioning is maintained by Glu68. Again, Ser36 facilitates the cleavage of the bond for the transfer of the γ -phosphate (see Figure II.3.3-17).

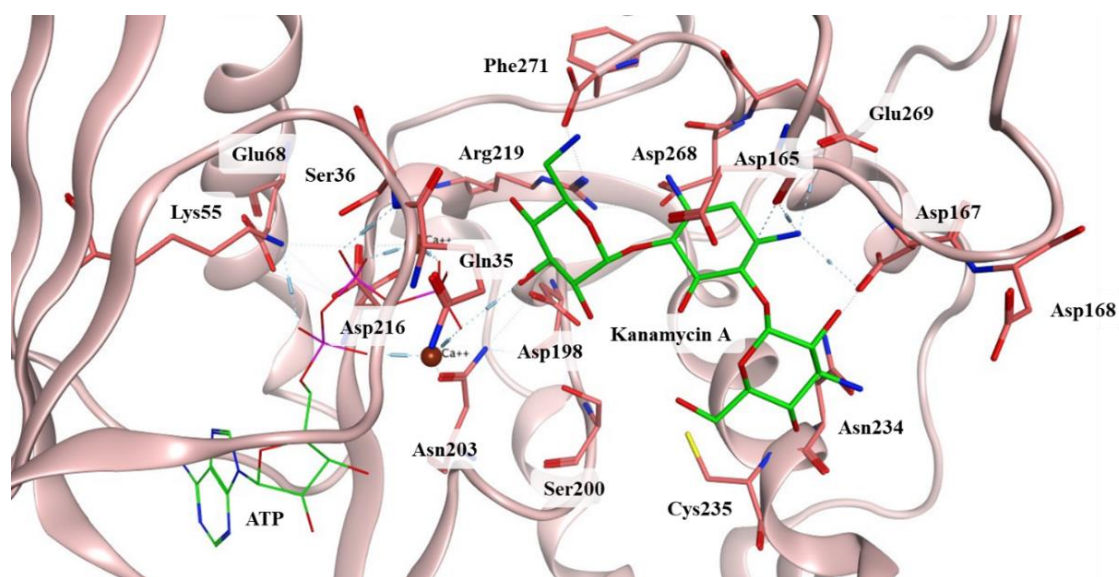


Figure II.3.3-17 3D view of APH(3')-Ia active site (PDB ID: 4EJ7).

Therefore, the following distances have been chosen to study the AME/AGA complexes:

- **d1:** AGA C(3')-oxygen atom – Asp198 OD1/OD2 atom.
- **d2:** AGA C(3')-oxygen atom – ATP γ -phosphorus atom.
- **d3:** AGA C(3')-oxygen atom – Ca²⁺.

b Experimental data used for modelling and AME/AGA model

APH(3')-Ia has been characterized in seven PDB entries, all published by the same authors: one in complex with ATP, and the other six in complex with kanamycin A and different small inhibitors (PDB IDs: 4EJ7, 4FEU, 4FEV, 4FEW, 4FEX, 4GKH, and 4GKI).

APH(3')-Ia has been modeled from the PDB entries 4EJ7 and 4GKH. These structures, deposited in 2012, having a resolution of 2.29 Å and 1.86 Å, respectively, were sourced from *E. coli* and shared 100 % sequence identity with each other and 99.6% sequence identity with the template sequence used by Dr. Prévost's team. The first structure provided the cofactors ATP and two Ca²⁺ ions, while the second structure provided the template kanamycin A. The second structure contained a small molecule inhibitor (1-NA-PP1) bound to the ATP-binding site that was not included.

The template structure (chain A of the PDB 4EJ7) and the query sequence differed in four residues, namely L19M, R27K, N48D, and A77E. All four residues were located at the surface of the enzyme at more than 12 Å from the active site.

The APH(3')-Ia model did not present any Ramachandran or bond length outlier. Yet, Asn16 and Arg170 displayed bond angle anomalies, with the second residue being part of ATP binding site. Also, two clashes were detected between ATP and both Ca²⁺ ions.

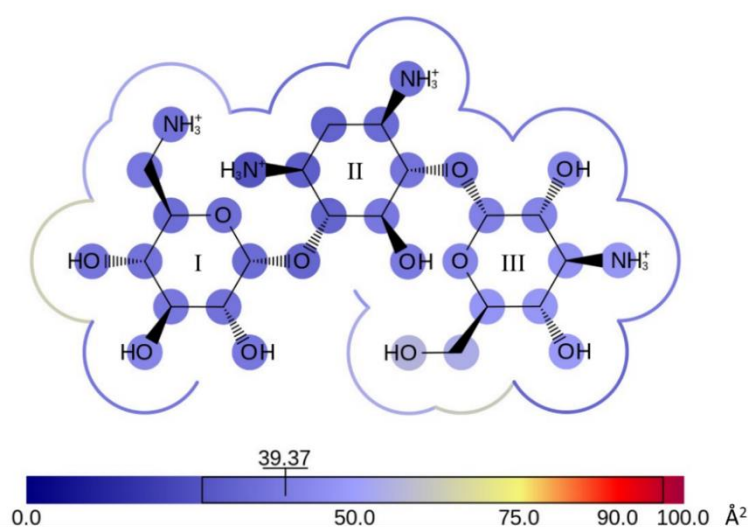


Figure II.3.3-18 B-factor values (Å²) of kanamycin A atoms and its binding site in the PDB entry 4GKH. The blue color indicates a smaller B-factor and thus a higher precision. The line surrounding the AGA represents the B-factor of the residues at less than 4.5 Å of kanamycin A.

The B-factor analysis of the atoms revealed a well-defined kanamycin A, yet with a slightly higher mobility on ring III compared to the other two rings. The residues close to C(4')-OH, namely Arg219 and Phe271, presented a high average temperature value. The same thing happened with Cys235, at less than 4.5 Å of C(6''). Additionally, residues in close proximity to the C(4')-OH group (Ser36, Arg219 and Phe271) demonstrated high average temperature values. Similarly, Cys235, located within a distance of less than 4.5 Å from C(6''), presented a high B-factor.

In summary, APH(3')-Ia active site is adequate to continue the modelling of neomycin B and HL_171.

Through the docking of neomycin B and HL_171 into APH(3')-Ia, we obtained a preliminary three-dimensional structure of both complexes. For both AGAs, the positioning of rings I and II was similar to that of kanamycin A, with a ring I RMSD of 0.9 Å and 1.0 Å, respectively. As expected, rings III and IV adopted different poses compared to kanamycin A. Interestingly, HL_171 ring III pose was inverted compared to the docked neomycin B, further supporting the width of subsite C.

c AME/AGA MD simulations

Similarly to previous APH(3')s, the complex stability was evaluated through the RMSD. Nevertheless, APH(3')-Ia contained a section of 17 residues (Ser2-Asn18) that were excluded from the α C RMSD calculation. This region lacked a defined secondary structure and was in constant motion, increasing the RMSD and impeding its interpretation.

Before production, the complex APH(3')-Ia/neomycin B reached a steady state close to the initial structure. Indeed, the binding site RMSD was lower than 2 Å, and the AGA RMSD was close to 1 Å. During production, the three replicas showed conformational changes as previously observed for APH(3')-IIa (see Fitted binding site RMSD in Figure II.3.3-20), involving the flexible loop Asp31-Ala38 and impacting the AGA binding mode (Annexes VI.1.13 - VI.1.15). Noteworthy, C(6''')-NH₂ transitioned from interacting with Gln35 to interacting with ATP, and C(5'')-OH initially interacted with Asn203 and then formed a hydrogen bond with Gln35.

HL_171 in complex with APH(3')-Ia also reached a steady configuration close to the energy minimum before production (binding site RMSD ~1.5 Å and AGA RMSD < 1.5 Å). The same

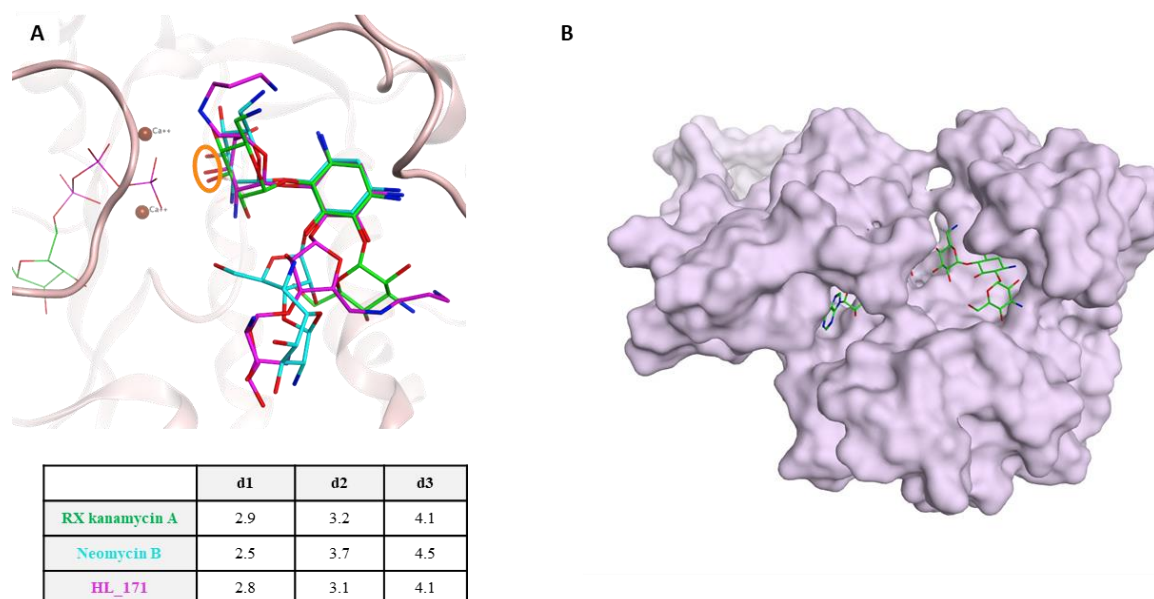


Figure II.3.3-19 Three-dimensional view of APH(3')-Ia and key distances measurement of (A) the bound RX of kanamycin A (green), and the docking poses of neomycin B (blue) and HL_171 (fuchsia). (B) Global view of the crystallographic AME/AGA complex.

conformational change observed in neomycin B complexes involving the loop Asp31-Ala38, also occurred for this complex. In the third replica, C(8'')-NH₂ interactions with Asp167 and Glu238, and ATP interactions with C(2')-NH₂ and C(3')-OH were lost after 90 ns and replaced by an interaction between Asp167 and C(1)-NH₂. These conformational changes were also responsible for rings I and II repositioning observed in their RMSD plot in Figure II.3.3-20.

Briefly, neomycin B and HL_171 initially formed transient complexes that relocated during the productions due to the movement of one of APH(3')-Ia loops. The analysis of the

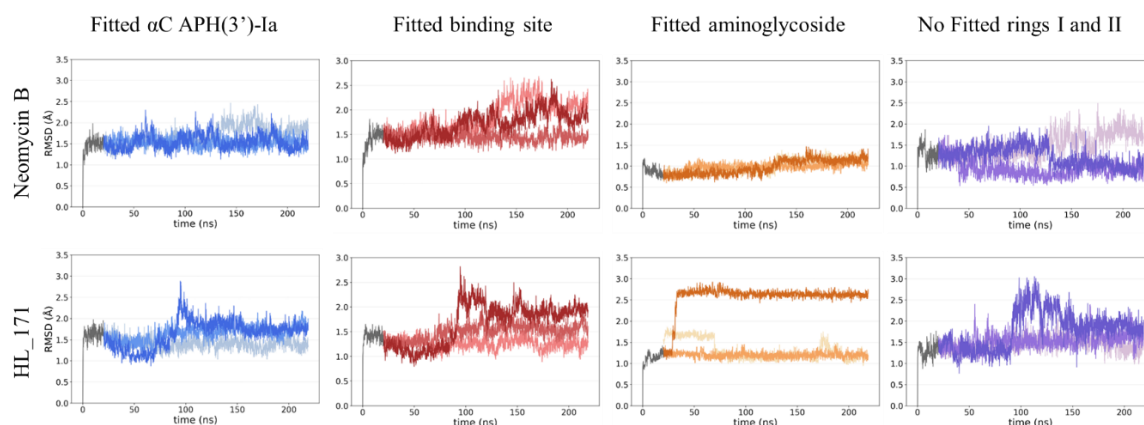


Figure II.3.3-20 RMSD plots of neomycin B and HL_171 in complex with APH(3')-Ia. The reference is the minimized structure. The grey part corresponds to the equilibration phase. The colored parts represent the three productions. The lighter color corresponds to the first replica and the darker one to the third.

interactions formed between the AGAs and the AME will help understand the impact of the conformational change.

The binding mode of neomycin B mainly relied on the four amino groups located on rings I and II (see Figure II.3.3-21), which persistently interacted with six amino acids. A steady hydrogen bond was also observed between C(5'')-OH and ATP. The implication of rings I and II to the binding mode aligned with the conformational changes, which predominantly affected rings III and IV.

When comparing the binding modes of neomycin B and HL_171, it was found that three stable interactions defining neomycin B binding mode, and involving three amino groups, also participated in HL_171 binding mode. Nevertheless, due to the synthetic substitutions

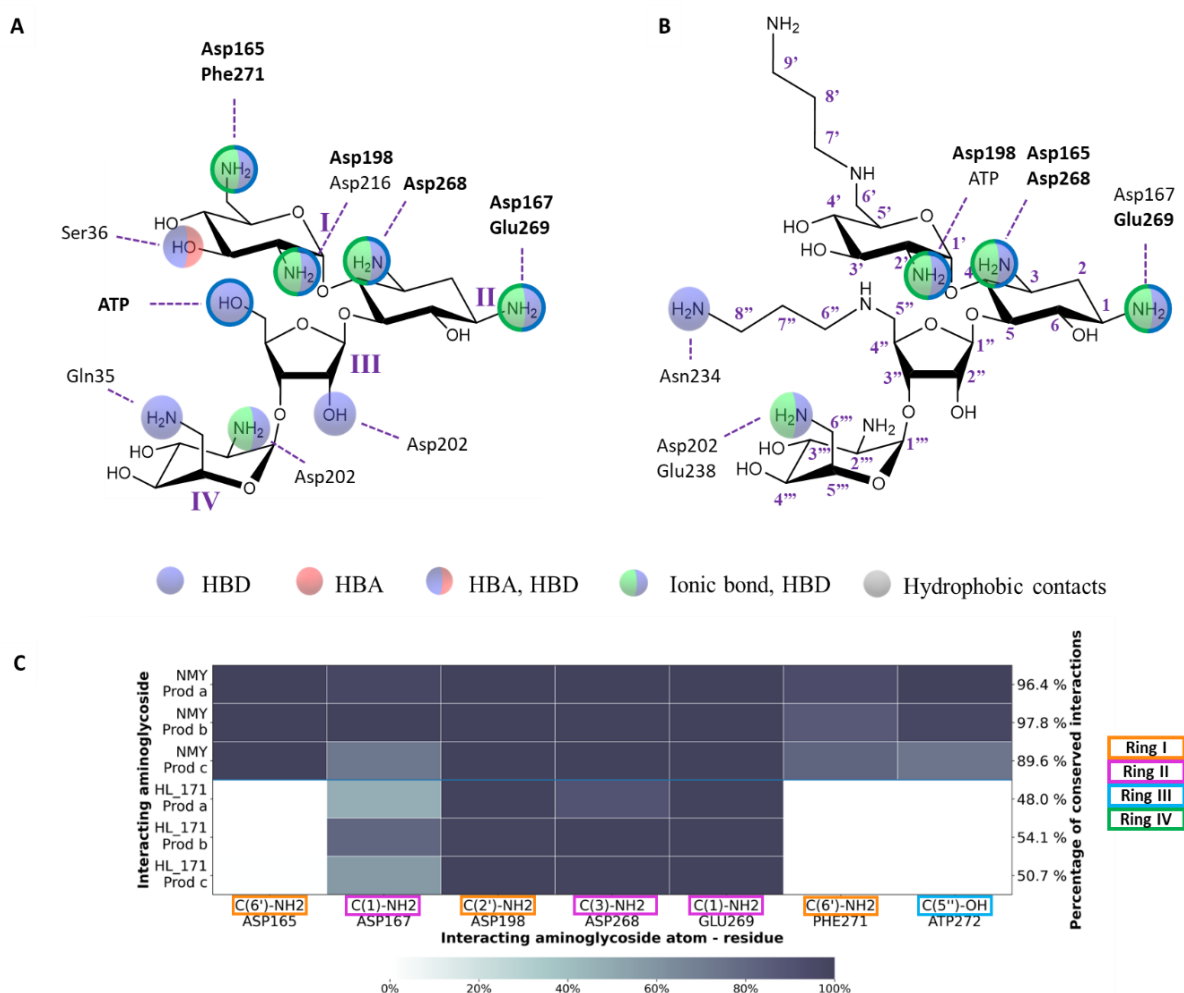


Figure II.3.3-21 Stable interactions formed between APH(3')-Ia and neomycin B or HL_171 during the MD simulations and present in more than 70% of the frames. Outlined interactions are present in all three productions. (A) Annotations on neomycin B. (B) Annotations on HL_171. (C) Frequency of interactions between the neomycin B (NMY) or HL_171 and APH(3')-Ia in each replica (prod_a, prod_b, and prod_c).

introduced in HL_171, the interactions involving C(6')-NH₂ were not part of its binding mode. Furthermore, the interaction between Asp167 and C(1)-NH₂ was not consistent anymore, resulting in a binding mode present in 54 % of the frames for the most similar HL_171 complex.

Regarding AGAs conformations, neomycin B torsion angles were consistent throughout the dynamics (see Figure II.3.3-22). In contrast, HL_171 displayed a conformational change during the first and third productions (see Fitted aminoglycoside in Figure II.3.3-20). Specifically, the ϕ angle of the linkage between rings III and IV varied up to 180° across the three replicas, showing two clearly distinct conformations during the third one (salmon-colored). Furthermore, similarly to APH(3')-IIa, HL_171 also exhibited an inverted ring III configuration.

In Figure II.3.3-23, the depiction of the last simulation frames illustrated the previous observations. Throughout all MD simulations of both neomycin B and HL_171 complexes, the distance between the targeted atom C(3')-OH and the catalytic residue Asp198 (**d1**) remained stable and similar to the referential crystallographic distance (see Figure II.3.3-23B and C). Notably, neomycin B second replica, and HL_171 first and second replica showed very stable distances towards all key elements indicating a consistent position for C(3')-OH, while the remaining three simulations demonstrated higher distance variations.

In summary, both AGAs shared common interactions, mainly implicating rings I and II. Furthermore, all complexes kept the targeted C(3')-OH at an adequate distance from the catalytic residue Asp198 (**d1**), as referred to the crystallographic reference. However, because of the conformational changes, rings I and II repositioning during APH(3')-Ia/HL_171 third

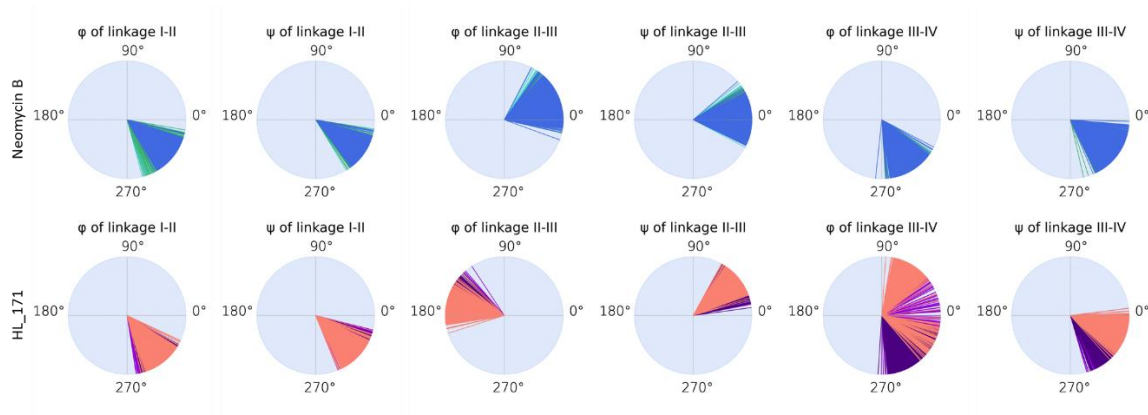


Figure II.3.3-22 Neomycin B and HL_171 dihedral angles during the MD simulations. Each color represents one of the production replicas.

replica, the inconsistency of Asp167 - C(1)-NH₂ interaction, combined with the logically missing C(6')-NH₂ and C(5'')-OH interactions, it was not possible to draw a clear conclusion about the possible action of the AME on HL_171.

The biological team detected the presence of APH(3')-Ia in two *E.coli* clinical strains exhibiting resistance towards neomycin B (see Table II.3.3-1). Interestingly, the MIC of HL_171 decreased compared to neomycin B. The here-employed methodology did not provide strong evidences to support the reduced activity of APH(3')-Ia on HL_171.

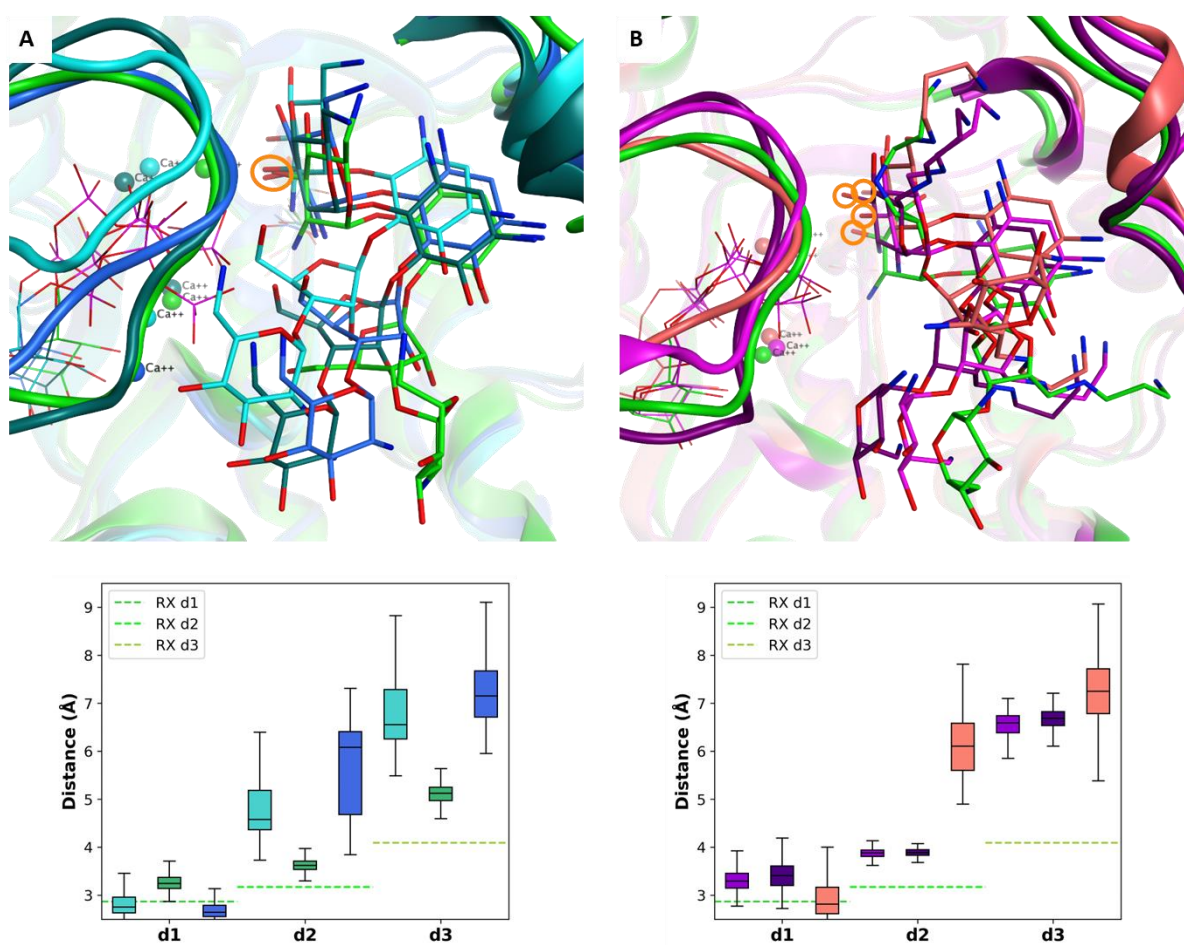


Figure II.3.3-23 3D structures of the last MD simulations frame and key distances of APH(3')-Ia in complex with (A) neomycin B (blues) and (B) HL_171 (purples). For the sake of comparison, the initial structure is shown in green. The targeted position of the aminoglycosides is marked in orange. With **d1**: AGA C(3')-oxygen atom – Asp198 OD1/OD2 atom; **d2**: AGA C(3')-oxygen atom – ATP γ -phosphorus atom; **d3**: AGA C(3')-oxygen atom – Ca²⁺.

II.3.3.6 ANT(4')-Ia

a About ANT(4')-Ia: catalytic mechanism and AGA binding mode

ANT(4')-Ia, also known as kanamycin nucleotidyltransferase, activates the hydroxylic group at position C(4') of tobramycin, amikacin, dibekacin, isepamicin, kanamycin, neomycin, paromomycin, and butirosin and at position C(4'') of dikekacin.²³

Its catalytic activity relies on the presence of two Mg^{2+} ions and a nucleotide triphosphate, namely ATP, GTP, CTP, TTP, or UTP.⁷² According to the catalytic mechanism proposed by Chen-Goodspeed et al., the AGA C(4')-OH is activated by Glu145 and attacks the ATP α -P forming a phosphodiester bond with AMP, and releasing of a pyrophosphate group. Lys149 would be in charge of increasing the α -P electrophilic character by interacting with the α -

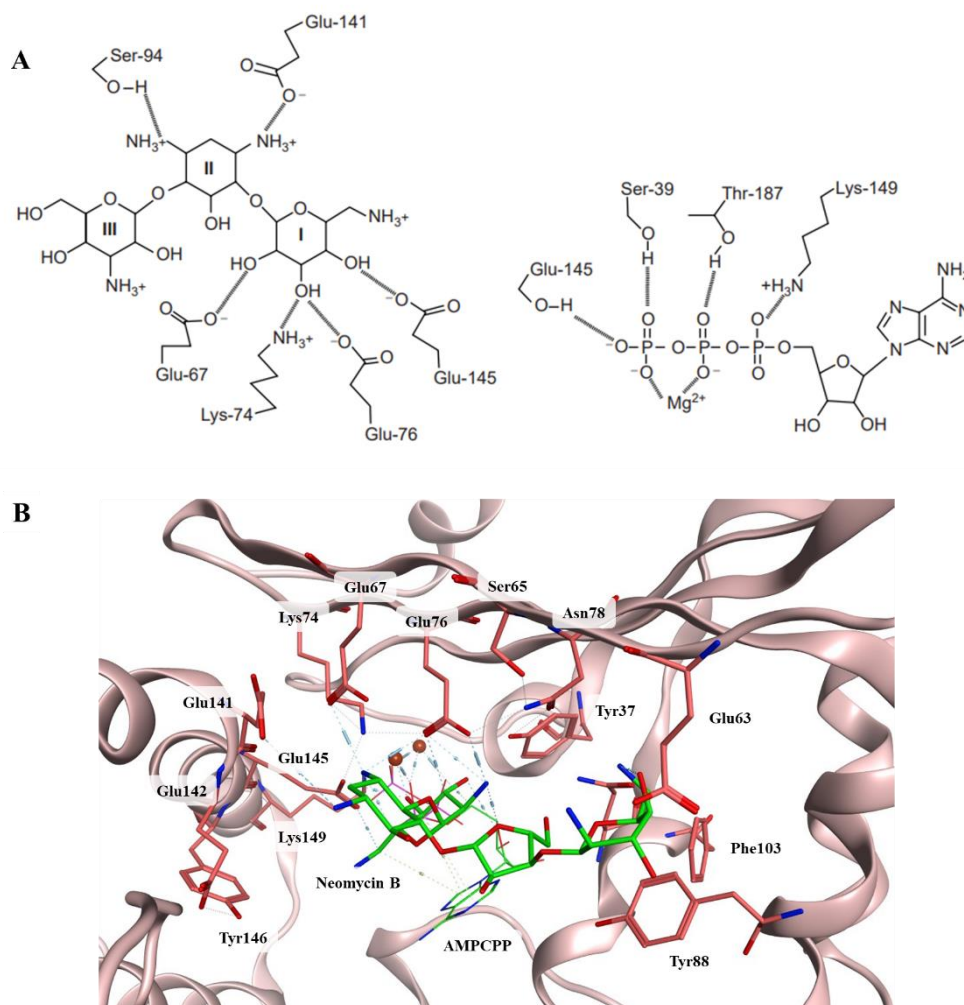


Figure II.3.3-24 Different representations of the active site of ANT(4')-Ia. (A) 2D view of the interactions involved in the catalysis. Image from the article 74. (B) 3D view of the active site.

phosphoryl oxygen atoms.^{73,74} Moreover, Glu76 was considered to be an alternate basic catalyst assisting the reaction.⁷⁵

Following the same reasoning as for previous enzymes, we selected the following distances to study the complexes formed by neomycin B and HL_171:

- **d1:** AGA C(4')-oxygen atom – Glu145 OE1/OE2 atom.
- **d2:** AGA C(4')-oxygen atom – Glu76 OE1/OE2 atom.
- **d3:** AGA C(4')-oxygen atom – AMPCPP α -phosphorus atom.
- **d4:** AGA C(4')-oxygen atom – Mg²⁺.

b Experimental data used for modelling and AME/AGA model

To date, nine crystal structures have been deposited in the PDB characterizing ANT(4')-Ia. Three with the wild AME sequence: in its apoenzyme form; in complex with AMPCPP and kanamycin A; and in complex with neomycin B (PDB IDs: 1KAN, 1KNY, and 6UN8 respectively). Four other structures had a mutation of T130K and were in complex with neomycin B; neomycin B and AMPCPP; neomycin B, AMP, and pyrophosphate; and neomycin B and AMP (PDB IDs: 6NMK, 6NML, 6NMM, and 6NMN, respectively). Finally the last two structures had two mutations (T130K and E52D) and were in complex with neomycin B; and neomycin B and AMPCPP (PDB IDs: 6P04 and 6P06, respectively).

The structure used for modelling this enzyme was obtained from the PDB entry 6NML, which was deposited in 2019, had a resolution of 2.0 Å and sourced from *E. coli*. This particular PDB file contained a ternary complex formed between neomycin B and AMPCPP, chosen to prevent the spontaneous formation of adenylate neomycin. The structure also included the two essential Mg²⁺ cofactors. It is worth reminding that the PDB structure corresponded to a thermostable mutant T130K. Nevertheless, the mutation was located on the enzyme surface and more than 21 Å away from the binding site. It did not introduce any modifications to the enzyme folding when compared to the wild-type ANT(4')-Ia.⁵⁵ Both chains of the homodimer were kept as they were part of the active site and interacted with the AGA.

During the evaluation of the geometry, Val186 was found to be an outlier according to the Ramachandran plot and to the standard bond angle values. However, the residue was located at more than 17.5 Å away from the active site, not affecting the analysis of the complexes

binding mode. Five other angle anomalies were identified in both chains of the structure: Thr39 and Tyr105 in chain A, and Ile29, Thr39, and Asp45 in chain B. Among these residues, only Asp45 was located within the AMPCPP binding site. No bond length outliers or clashes were observed.

The average B-factor for the neomycin B molecule bound to chain A was calculated as 42.4 Å², whereas for chain B, it was determined as 30.8 Å². Furthermore, the binding site B-factor in chain B was also lower than that of chain A (36.2 Å² vs 26.4 Å²). Hence, neomycin B bound to chain B was chosen to be the AGA template (see Figure II.3.3-25).

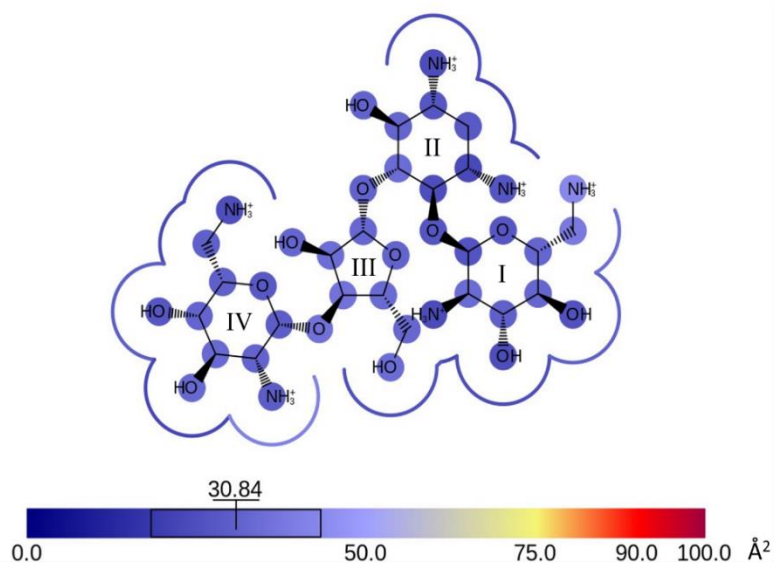


Figure II.3.3-25 B-factor values (Å²) of neomycin B atoms and its binding site to chain B of PDB 6NML. The blue color indicates a smaller B-factor and thus a higher precision. The line surrounding the AGA represents the B-factor of the residues at less than 4.5 Å of kanamycin A.

Briefly, the active sites of both chains were correctly defined to study neomycin B binding mode and model a complex with HL_171. Based on the B-factor analysis, we decided to work with ANT(4')-Ia chain B, as it showed a smaller mobility.

Neomycin B redocking under constraints predicted a pose of ring I almost identical to the crystallographic structure, with an RMSD equal to 0.24 Å. The global RMSD, equal to 2.0 Å, was also low, with all the rings mimicking the location of the template. Furthermore, the distances **d1**, **d2**, **d3** and **d4** only differed by a maximum of 0.2 Å compared to the template (see Figure II.3.3-26A).

HL_171 docking into ANT(4')-Ia presented a ring I RMSD of 0.9 Å, also adopting a close conformation to that of the neomycin B template. However, HL_171 targeted C(4')-OH shifted, resulting in a distance **d3** 2 Å higher from that of the template (see Figure II.3.3-26A).

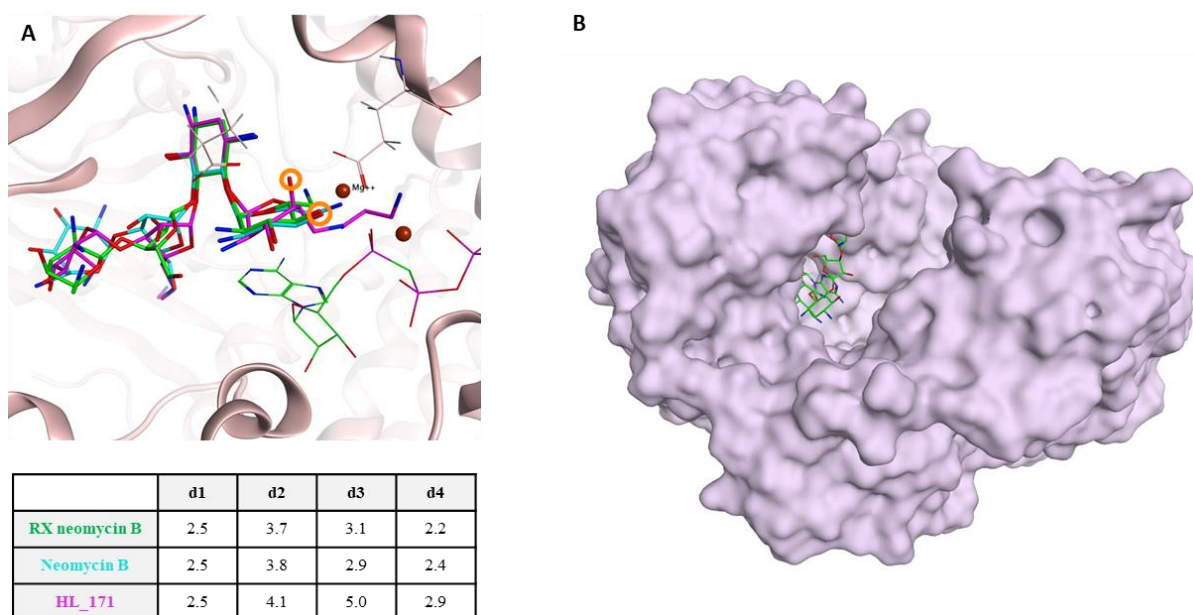


Figure II.3.3-26 Three-dimensional view of ANT(4')-Ia and key distances measurement of (A) the bound RX of kanamycin A (green), and the docking poses of neomycin B (blue) and HL_171 (fuchsia). (B) Global view of the crystallographic AME/AGA complex.

c AME/AGA MD simulations

All RMSD plots showed very stable complexes for both neomycin B and HL_171 (see Figure II.3.3-27). The configuration adopted by neomycin B complexes was similar to the minimized state as the RMSD of the binding site and of the AGA did not exceed 1.5 Å. No conformational changes were observed during the productions of ANT(4')-Ia/neomycin B.

HL_171 also showed a stable complex during the MD simulations with the binding site RMSD being also lower than 1.5 Å. However, the RMSD of rings I and II was higher in the complex

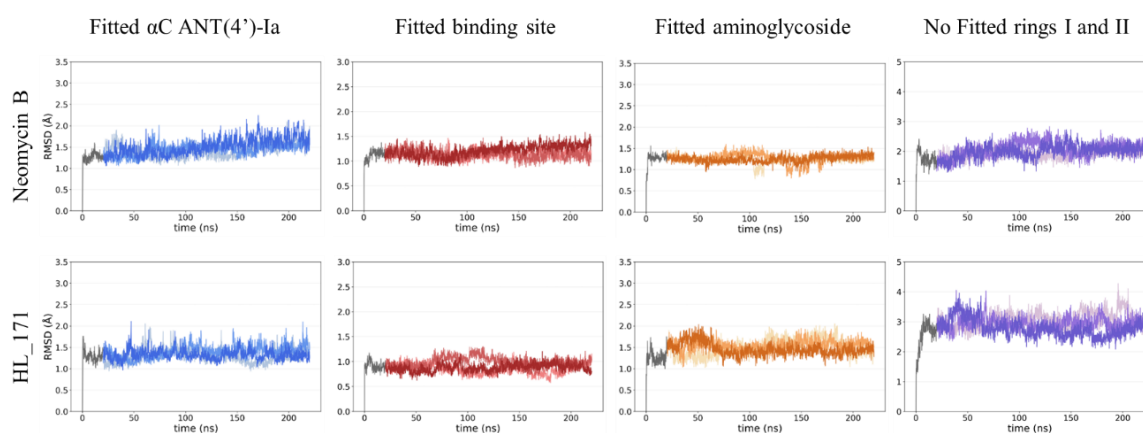


Figure II.3.3-27 RMSD plots of neomycin B and HL_171 in complex with ANT(4')-Ia. The reference is the minimized structure. The grey part corresponds to the equilibration phase. The colored parts represent the three productions.

formed by the derivative compared to that of neomycin B, suggesting a shift of the rings inside of the binding site during the equilibration phase.

Briefly, both AGAs formed stable complexes with ANT(4')-Ia allowing the comparison of the binding modes.

Neomycin B formed interactions with the enzyme using four of its six amino groups and the targeted C(4') hydroxyl group. Most of the interacting atoms were located in rings I and II, with only one interaction occurring in ring IV. The amino group at position C(6') appeared to be important for recognition as it formed hydrogen bonds and ionic bonds with three residues (see Figure II.3.3-28A).

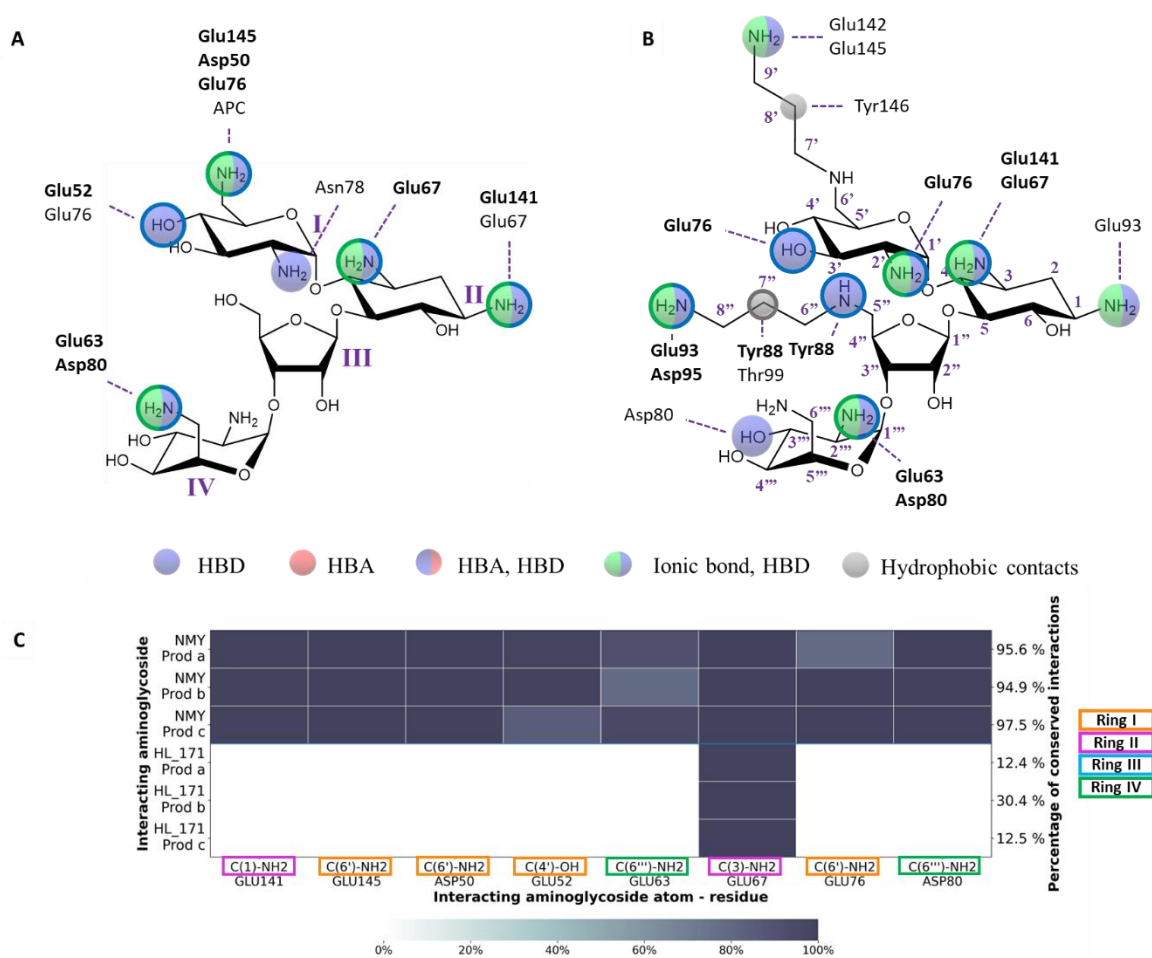


Figure II.3.3-28 Stable interactions formed between ANT(4')-Ia and neomycin B or HL_171 during the MD simulations and present in more than 70% of the frames. Outlined interactions are present in all three productions. (A) Annotations on neomycin B. (B) Annotations on HL_171. (C) Frequency of interactions between the neomycin B (NMY) or HL_171 and ANT(4')-Ia in each replica (prod_a, prod_b, and prod_c).

In contrast, HL_171 displayed a completely different binding mode, except for the shared interaction between Glu67 and C(3)-NH₂ (see Figure II.3.3-28C). The recognition of HL_171 was based on five amino groups, with three of them located in rings III and IV, along with the C(3')-hydroxylic group. Notably, the substituted chain at position C(5'') in HL_171 played a significant role by forming stable interactions with three specific residues that persisted throughout all three productions. This chain introduced hydrophobic contacts to the binding mode profile, contributing to it in a novel manner (see Figure II.3.3-28B).

In summary, both AGAs formed stable complexes with ANT(4')-Ia but with different binding modes.

The torsion angles of both AGAs followed the pattern of the crystallographic structures analyzed in section II.2.3.1 Neomycin B in complex with AMEs and RNA. Only neomycin B φ angle of the linkage between rings III and IV demonstrated higher mobility with a range of angles of 150°. This increased mobility could be attributed to the absence of stable interactions between ring III and the AME (see Figure II.3.3-29).

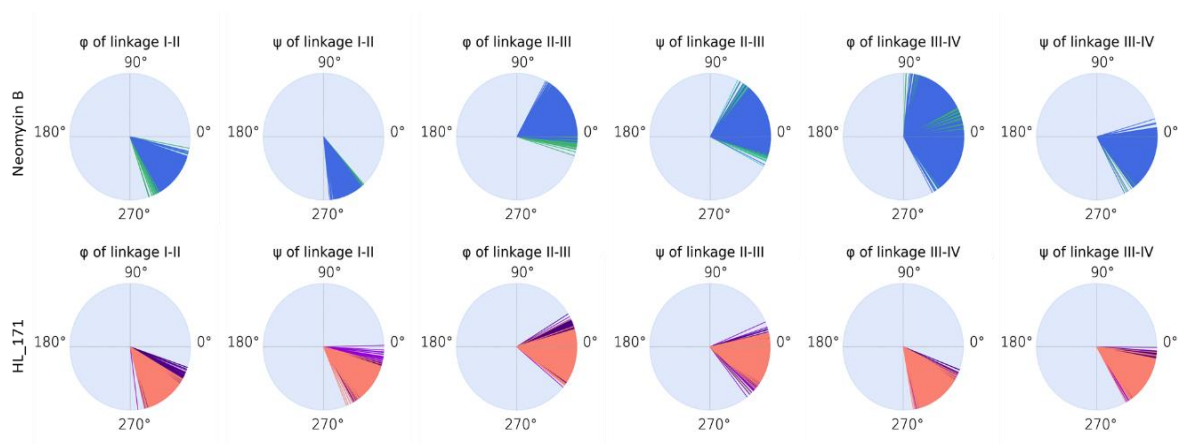


Figure II.3.3-29 Neomycin B and HL_171 dihedral angles during the MD simulations. Each color represents one of the production replicas.

Figure II.3.3-30 illustrated the end points of the MD productions. The distances of the complexes formed by both AGAs appeared to be different from that of the crystallographic structure. Neomycin B slightly shifted on the cavity elongating the distance all key elements (**d1**, **d2**, and **d3**). Nevertheless, the alternate basic catalyst Glu76 remained at distance **d2** < 3.5 Å, allowing for the formation of a hydrogen bond. Finally, another glutamic acid steadily interacted with both the targeted C(4')-OH and the Mg²⁺: Glu52.

HL_171 also showed very different distances **d1**, **d2**, and **d3** compared to the template. The relocation of the derivative inside of the cavity can be observed in Figure II.3.3-30B. All HL_171 complexes adopted a similar and steady pose, but distant from that of neomycin B and from the key elements responsible for the AMP transfer.

Briefly, both neomycin B and HL_171 formed stable complexes with ANT(4')-Ia. However, they exhibited distinct binding profiles. HL_171, with its additional synthetic chains, and particularly the one at position C(5''), experienced a shift in the binding site. This suggested that the modification interfered with the metabolism activity of the AME. These findings aligned with the biological results, which showed an 8-fold decrease of the MIC for HL_171 tested in strains containing ANT(4')-Ia as compared to neomycin B (see Table II.3.3-2).

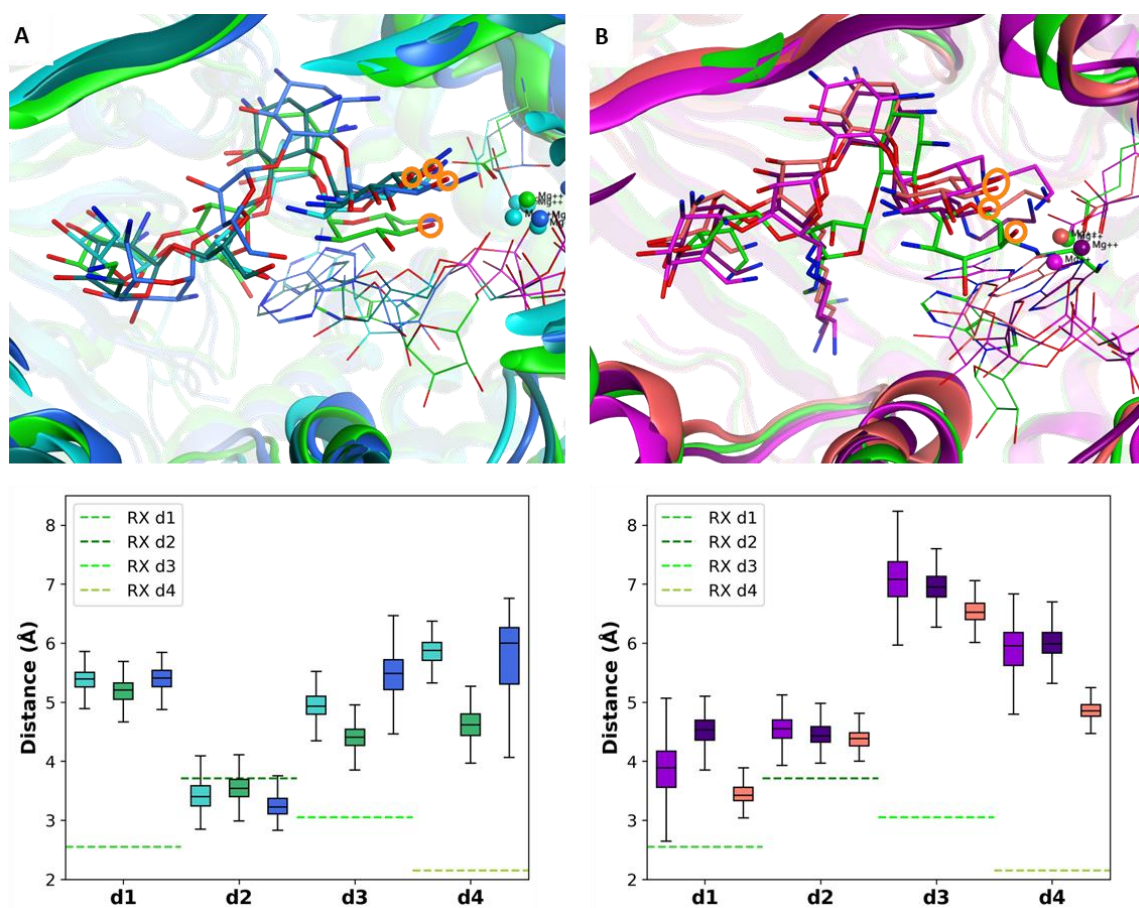


Figure II.3.3-30 3D structures of the last MD simulations frame and key distances of ANT(4')-Ia in complex with (A) neomycin B (blues) and (B) HL_171 (purples). For the sake of comparison, the initial structure is shown in green. The targeted position of the aminoglycosides is marked in orange. With **d1**: AGA C(4')-oxygen atom – Glu145 OD1/OD2 atom; **d2**: AGA C(4')-oxygen atom – Glu76 OD1/OD2 atom; **d3**: AGA C(4')-oxygen atom – ATP α -phosphorus atom; and **d4**: AGA C(4')-oxygen atom – Mg^{2+} .

II.3.3.7 AAC(3)-IIa

a About AAC(3)-IIa: catalytic mechanism and AGA binding mode

AAC(3)-IIa, also called aminoglycoside 3-N-acetyltransferase type IIa, belongs to the family of GCN5-related N-acetyltransferase (GNAT). Acetyltransferases require acetyl coenzyme A as a cofactor to acetylate AGA amino groups, and, unlike the previously studied AMEs, they do not depend on metal cofactors for their activity. AAC(3)-IIa targets the amino group at position C(3) of tobramycin, gentamicin, netilmicin, dibekacin, and sisomicin.

The catalytic mechanism of acetyltransferases has been described for AAC(3)-VIa (see Figure II.3.3-31). The AME contains a non-canonical catalytic triad composed of Glu192, His189, and the AGA with an amine at position C(3). Upon activation, the amine reacts with the acetyl

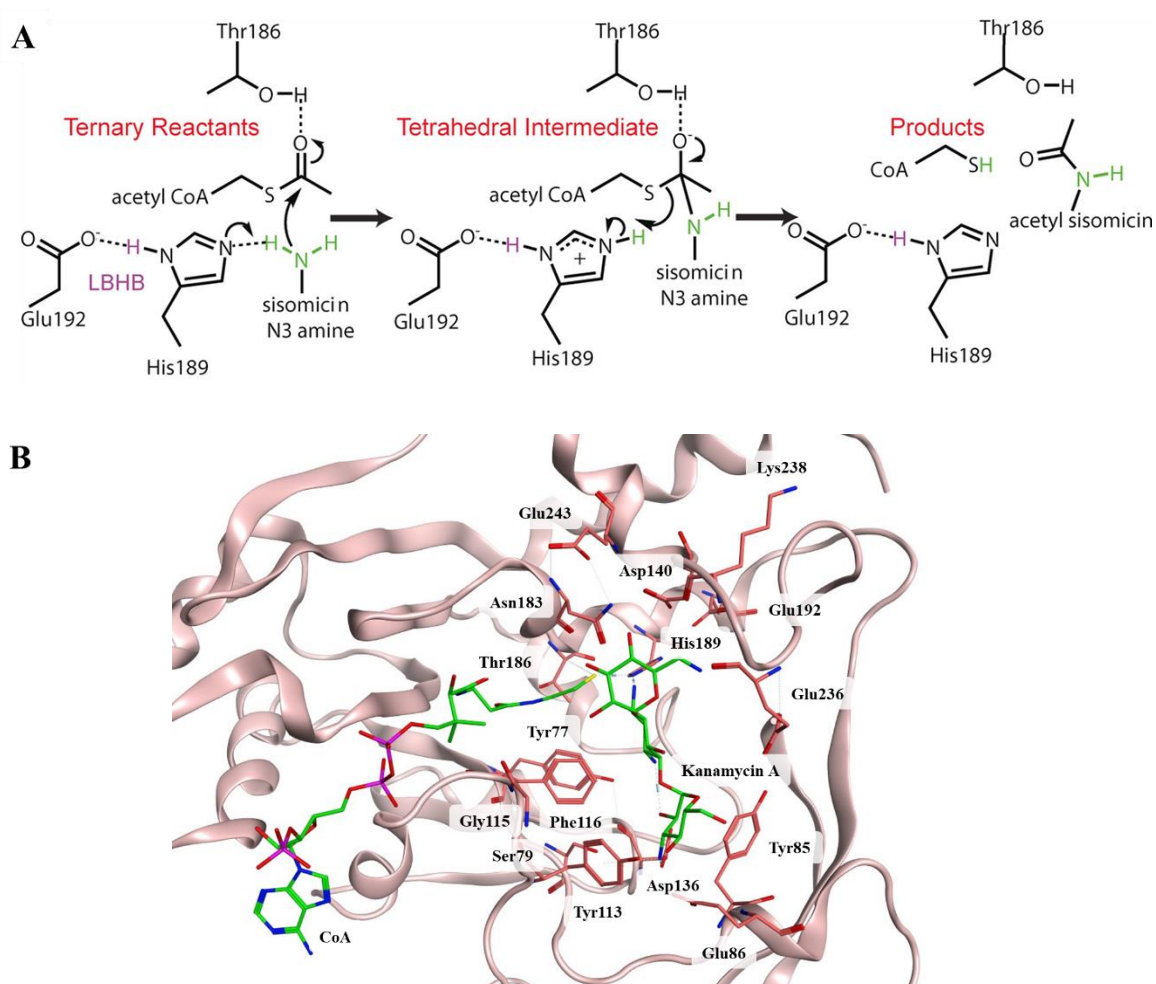


Figure II.3.3-31 Different representations of the active site of AAC(3)-IIa. (A) 2D view of the interactions involved in the catalysis of AAC(3)-VIa. Image from the article 76 (B) 3D view of AAC(3)-IIa active site.

of acetyl CoA, resulting in the production of coenzyme A and an acetylated AGA that loses its antibacterial activity.⁷⁶

Therefore, the chosen key elements to evaluate the distances to the atoms required for metabolization are:

- **d1:** AGA C(3)-nitrogen atom – His189 NE2 atom.
- **d2:** AGA C(3)-nitrogen atom – CoA sulphur atom.

Compared to other AMEs, AAC(3)-IIa and AAC(3)-VIa have a narrow substrate specificity profile, only targeting structurally similar AGAs. Therefore, neomycin B and other aminoglycosides from the 4,6-AGA family are not affected by AAC(3) metabolic activity.

Our decision to model and study the complexes of AAC(3)-IIa with neomycin B and HL_171 came from the presence of this enzyme in three of the strains demonstrating resistance (see Table II.3.3-1) and its potential to serve as a negative control to validate the ability of our method to predict the absence of metabolism on neomycin B.

b Experimental data used for modelling and AME/AGA model

The structure of AAC(3)-IIa has not been characterized in a crystal structure yet. Therefore, the AME was modeled using AAC(3)-VIa as template, which shared 46.5% sequence identity with our target sequence (PDB ID: 6O5U). The PDB entry was deposited in 2019, had a resolution of 1.40 Å, and was sourced from *E. coli*. The reference AGA kanamycin A was also obtained from this entry. The coenzyme A, lacking the acetyl group, was extracted from the PDB file 6MN1, which describes the structure of the mutant H168A AAC(3) in complex with gentamicin and CoA.

The modelled structure identified Ala92 as a Ramachandran outlier located on a loop at the surface of the AME at 15 Å from kanamycin A binding site. Three bond angle outliers were also identified: Asp81, Ile197, and Glu236. Among these residues, only Glu236 was located within the AGA binding site.

Kanamycin A and its binding site showed low B-factor values (23.1 Å² and 23.0 Å², respectively) that could be connected with the narrowness of the binding site (see Figure II.3.3-32).

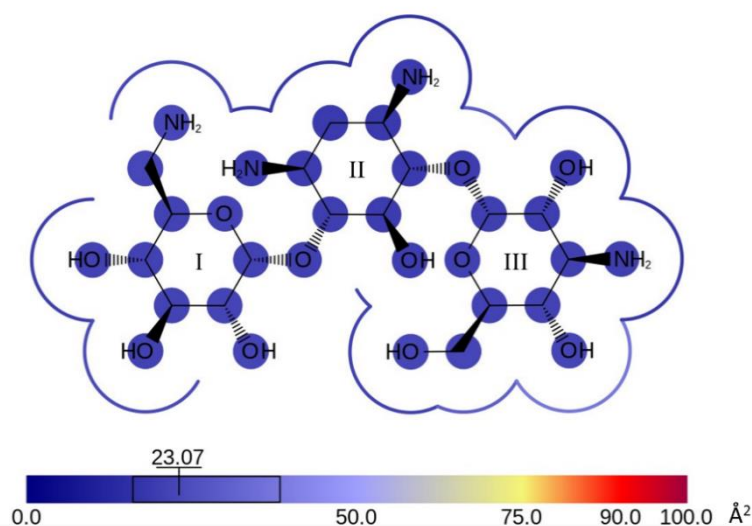


Figure II.3.3-32 B-factor values (\AA^2) of kanamycin A atoms and its binding site in the PDB entry 6O5U. The blue color indicates a smaller B-factor and thus a higher precision. The line surrounding the AGA represents the B-factor of the residues at less than 4.5 \AA of neomycin B.

Kanamycin A binding site is adequate for the study of AAC(3)-IIa in complex with neomycin B and HL_171.

Neomycin B and HL_171 docking poses precisely matched the reference pose of the RX kanamycin A, with the targeted C(3)- NH_2 positioned at the same location of the template (see Figure II.3.3-33). Neomycin B pose presented an RMSD for ring I equal to 1.4 \AA . However, HL_171 ring I RMSD was equal to 2.8 \AA , suggesting an impact of the synthetic substitution at position C(6') on the location of ring I.

c AME/AGA MD simulations

The RMSD profile obtained from the production simulations clearly indicated that neomycin B and AAC(3)-IIa did not form a stable complex. Throughout the simulations, the binding site underwent constant changes, and the AGA RMSD analysis revealed multiple conformations. Additionally, rings I and II of neomycin B continuously relocated during the dynamics (see neomycin B No Fitted rings I and II Figure II.3.3-34).

Among the replicas, only the first one presented some stability regarding rings I and II location. Surprisingly, HL_171 presented a more stable binding to AAC(3)-IIa, with a binding site RMSD < 2.5 \AA . Nevertheless, some motion was still visible in the placement of rings I and II with the RMSD going up to 5 \AA (see and Figure II.3.3-34). The observed instability of the AAC(3)-IIa/neomycin B complex was supported by the analysis of the interactions during the

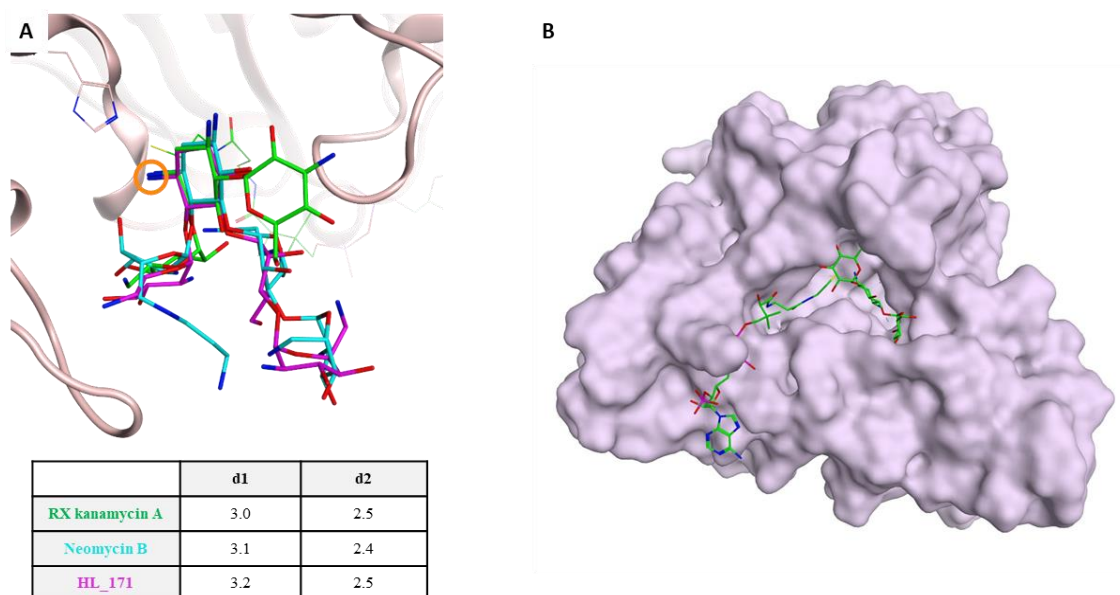


Figure II.3.3-33 Three-dimensional view of AAC(3)-Ia and key distances measurement of (A) the bound RX of kanamycin A (green), and the docking poses of neomycin B (blue) and HL_171 (fuchsia). (B) Global view of the crystallographic AME/AGA complex.

MD simulations. Each replica displayed a maximum of four persistent interactions, primarily involving the C(6')-NH₂ group. However, none of them was conserved in all three runs (see Figure II.3.3-35A).

In contrast, HL_171 demonstrated a higher stability due to the presence of two common interactions across all replicas involving hydrogen and ionic bonds with C(1)-NH₂ and C(6''')-NH₂ (see Figure II.3.3-35B). When analyzing the impact of the synthetically added chains on

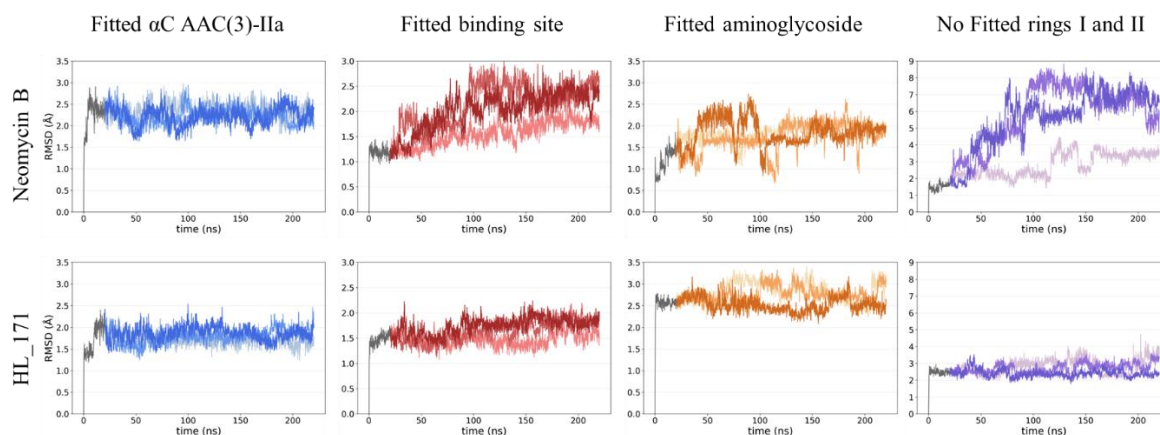


Figure II.3.3-34 RMSD plots of neomycin B and HL_171 in complex with AAC(3)-IIa. The reference is the minimized structure. The grey part corresponds to the equilibration phase. The colored parts represent the three productions. The lighter color corresponds to the first replica and the darker one to the third.

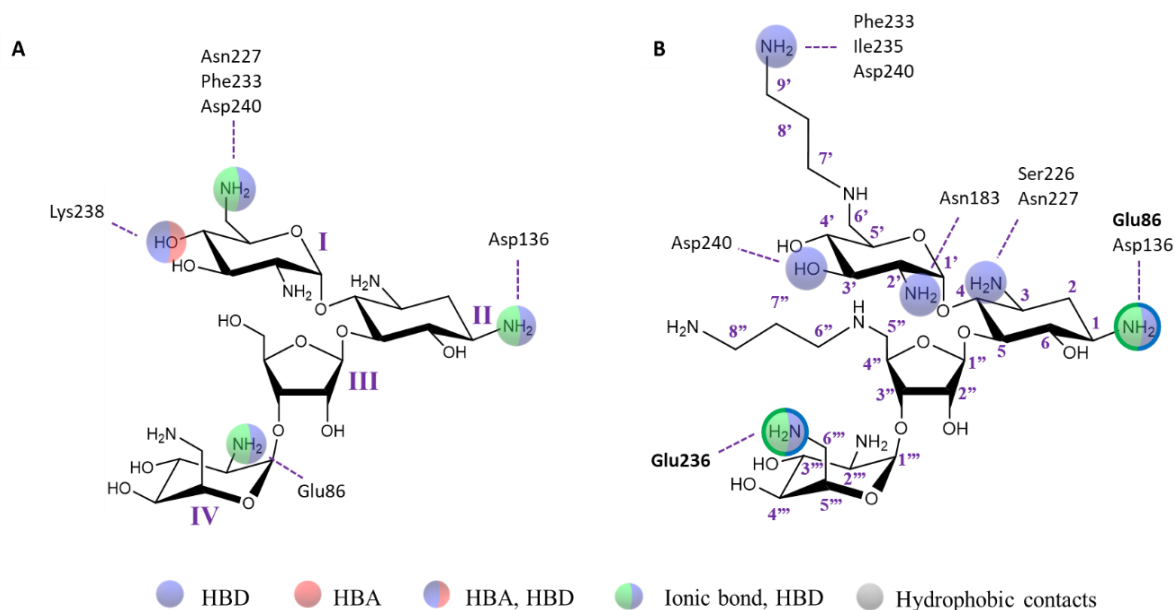


Figure II.3.3-35 Stable interactions formed between AAC(3)-IIa and neomycin B or HL_171 during the MD simulations and present in more than 70% of the frames. Outlined interactions are present in all three productions. (A) Annotations on neomycin B. (B) Annotations on HL_171.

the binding mode, it was notable that the C(9') position of the aminoglycoside could form persistent interactions with three residues during the MD simulations (Ile235, Phe233, and Asp240).

As a consequence of the lack of a stable binding mode, neomycin B and HL_171 also displayed a large range of conformations, especially boosted by the φ angle of linkage between rings III and IV (see Figure II.3.3-36).

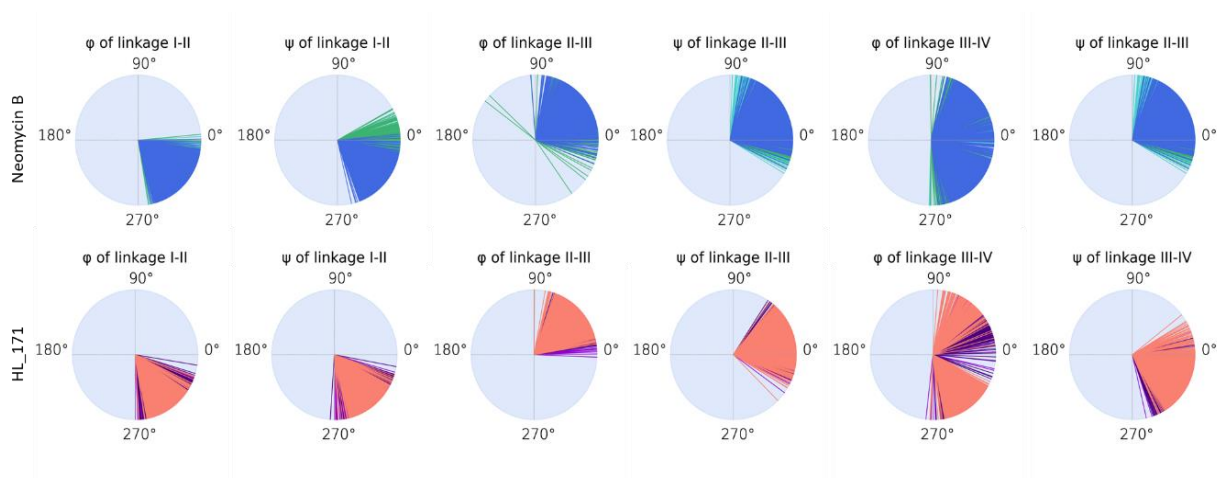


Figure II.3.3-36 Neomycin B and HL_171 dihedral angles during the MD simulations. Each color represents one of the production replicas.

Finally, the illustration of the last simulation frame in Figure II.3.3-37A showed neomycin B shift on the cavity. Indeed, the targeted position C(3)-N was found to be as far as 14 Å away from the catalytic residue His189. The first simulation, showing a higher stability in Figure II.3.3-34, also presented a consistent distance **d1**. However, the complex cofactor CoA exited the binding cavity, reaching distances greater than 40 Å in some frames.

HL_171 underwent a smaller relocation compared to neomycin B, but also showed different distances (**d1** and **d2**) from that of the crystallographic kanamycin A (see Figure II.3.3-37B).

Briefly, the employed method accurately predicted that neomycin B was not a substrate of AAC(3)-IIa. The AGA did not exhibit a specific binding mode, as none of the interactions it formed were consistently observed in all three replicas. Moreover, the repositioning of neomycin B led to a further separation from the key elements involved in the catalytic process. In contrast, HL_171 was able to bind AAC(3)-IIa with common features among the different replicas. Nevertheless, the distances between the AGA and the key catalytic residues were still too far apart, indicating that this derivative would not be metabolized either

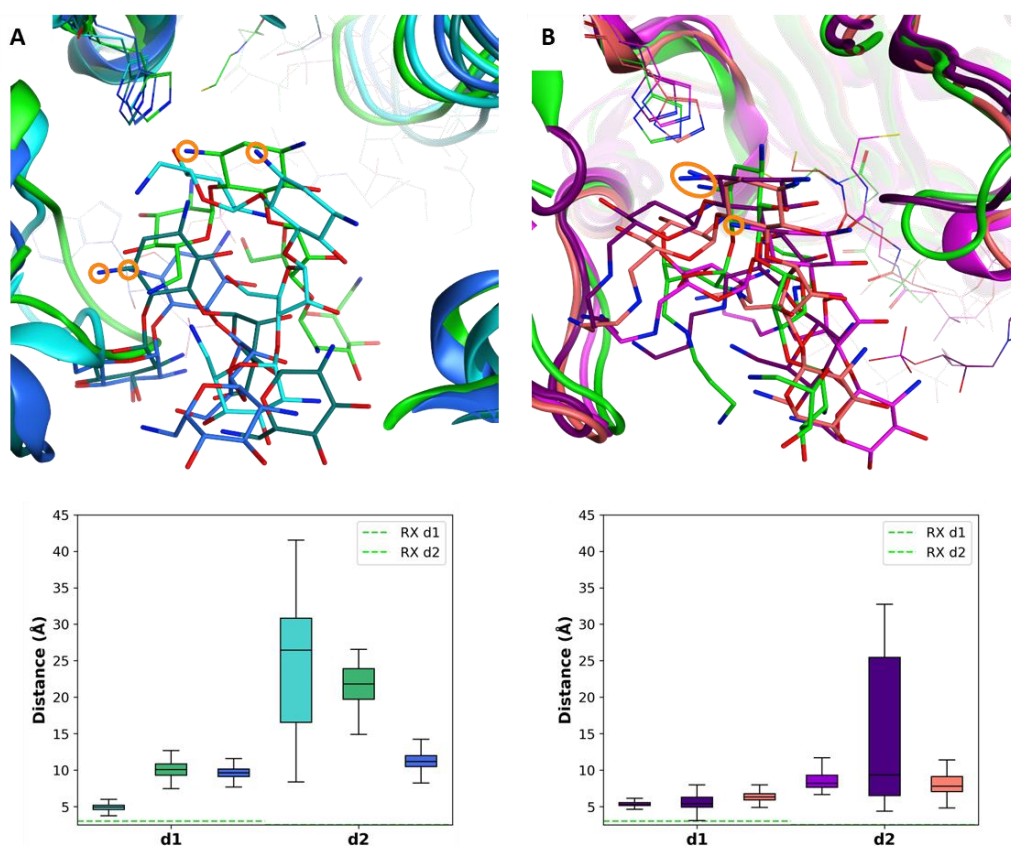


Figure II.3.3-37 3D structures of the last MD simulations frame and key distances of AAC(3)-IIa in complex with (A) neomycin B (blues) and (B) HL_171 (purples). For the sake of comparison, the initial structure is shown in green. The targeted position of the aminoglycosides is marked in orange. With **d1**: AGA C(3)-oxygen atom – His189 NE2 atom; **d2**: AGA C(3)-oxygen atom – CoA sulfur atom.

II.3.4 Discussion

II.3.4.1 About the methodology

In this study, we developed a five-step methodology to compare the recognition of neomycin B and its derivative, HL_171, by five different AMEs. The methodology involved the construction of a three-dimensional model of the AME/AGA complex using homology modelling and docking, followed by stability evaluations through MD simulations.

The assessment of HL_171 recognition by an AME was based on multiple factors, including the consistency of its binding mode throughout the MD simulations, the consistency with the neomycin B binding mode, and the distances to key elements involved in metabolization. Based on our findings, HL_171 was favorable for metabolization when bound to APH(3')-IIa. In contrast, in presence of APH(3')-IIIa, ANT(4')-Ia, and AAC(3)-IIa, HL_171 demonstrated unfavorable metabolization. Finally, the recognition of HL_171 by APH(3')-Ia was inconclusive due to the partial conservation of neomycin B binding mode and the appropriate distance to key metabolization elements observed in only two of the simulation replicas.

Briefly, these results highlighted the potential variations in HL_171 recognition among different AMEs and emphasized the importance of considering multiple factors in assessing the effectiveness of neomycin B derivatives. Furthermore, these results agreed with previous research in the field aiming at understanding the origin or AME promiscuity.⁷⁷

The methodology developed in this study enabled us to examine the stability of AME/AGA complexes and draw conclusions regarding AGA metabolization. MD simulations played a crucial role in confirming the stability of the complexes and the consistency of their binding modes. Moreover, they provided insights into conformational variations within the binding sites, offering a deeper understating of the binding site and how modifications to neomycin B could impact binding. However, a significant drawback of this procedure was the need for individual calculations and analysis for each complex, making impossible the automation of the process and requiring a considerable amount of time for each complex.

Indeed, during the procedure, the preparation of initial three-dimensional models and the subsequent analysis of the MD simulation results were particularly time-consuming. Once the initial AME model was constructed, each MD simulation replica took approximately one day

to complete. Therefore, although the methodology provided valuable insights, its applicability may be limited to the specific AMEs studied in this work. However, it opens up the possibility of studying the recognition of new derivatives for these enzymes.

Furthermore, our findings suggested that three 200ns MD replicas are sufficient to assess the stability of the binding modes predicted by docking. This was obvious in the simulation of neomycin B with AAC(3)-IIa, where none of the replicas exhibited the same binding mode. Similarly, the significance of the 200ns duration was demonstrated in the simulation of recognition by APH(3')-IIa, where the binding mode changed over the course of the simulation.

II.3.4.2 About the prediction on HL₁₇₁ recognition

The results obtained from the modeling of HL₁₇₁ were consistent with the biological findings.

For APH(3')-IIIa and ANT(4')-Ia, which were the only AMEs detected in three *S. aureus* strains, our calculations predicted an unfavorable metabolism of HL₁₇₁. This prediction was attributed to the different binding mode compared to neomycin B and the considerable distance from the essential elements required for AME catalytic activity. The biological results confirmed this prediction by indicating an 8-fold to 16-fold decrease in MIC activity for these two AMEs.

APH(3')-IIa, which was detected in three *E. coli* strains, was the only AME detected in one of the strains, while the other two strains included additional AMEs. For two of the strains, including the one where APH(3')-IIa was the only AME present, the MIC of HL₁₇₁ was found to be 2-fold higher than that of neomycin B. This observation indicated an increase in the AME activity. The results agreed with our findings, suggesting that HL₁₇₁ was a substrate of APH(3')-IIa.

APH(3')-Ia was found in two *E. coli* strains but always co-occurring with other AMEs, rendering the interpretation of the results more complicated. Among the other AMEs present in these strains, only ANT(2'')-Ia and APH(3')-Ib had the potential to metabolize neomycin B or HL₁₇₁. However, since we did not model the complexes involving these AMEs, a direct comparison could not be made. Nevertheless, both strains showed a MIC decrease when exposed to HL₁₇₁ compared to neomycin B. The variation in activity could potentially be attributed to the differences observed in the binding mode of the APH(3')-Ia/HL₁₇₁ complex

during the MD simulations. However, it is important to note that our methodology relied on various factors for drawing conclusions, and thus could not provide a clear conclusion in this regard.

In the case of AAC(3)-IIa, the results agreed with the narrow binding site nature of the AME preventing the catalytic activity on AGAs from the 2-DOS 4,5 disubstituted family. Additionally, this model provided insights into the potential outcomes for AGAs that are anti-complementary to the binding site.

Overall, the synthetic modifications made to neomycin B had a significant impact on two specific AMEs. Specifically, stable interactions were observed between APH(3')-IIIa and the synthetic chains at positions C(6') and C(5''), as well as between ANT(4')-Ia and the synthetic chain at position C(5''). These interactions were responsible for the observed variation in binding modes and subsequent impact on metabolization activity. In contrast, no stable interactions were found between the synthetic chains and APH(3')-Ia or APH(3')-IIa, resulting in similar binding modes between HL_171 and neomycin B for these AMEs. Nonetheless, the synthetic chains did have an influence when considering the conformers, as the ϕ angle of the linkage between rings II and III presented distinct values from those observed in crystallographic structures.

II.3.4.3 Suggestions for future modifications

In terms of future modifications, it is suggested that combining multiple modifications could enhance the potential to reduce affinity towards AMEs. However, it is important to acknowledge the challenge of finding modifications that can impact all AMEs due to the significant differences in their binding sites. Nevertheless, the analysis of persistent interactions between neomycin B and the AMEs highlighted the importance of the amino group at position C(1), which was consistently involved in the binding across all models. Furthermore, amikacin confirmed the possibility of growing the AGA from that position without altering the binding to RNA. Similarly, C(2')-NH₂ formed stable interactions with three AMEs and thus is also an interesting modification target. Finally, the most effective modifications are those directly preventing the AME catalytic activity, achieved by removing or modifying the amino or hydroxyl groups from the targeted positions.

II.3.5 Conclusion

Bacterial resistance is a significant threat to human health, with resistance being observed across all classes of antibiotics. In our study, we addressed this pressing issue by developing a comprehensive five-step protocol to evaluate the impact of modifications made to neomycin B on the recognition by five AMEs.

Through the construction of three-dimensional models of the complexes and the study of the stability and binding modes through MD simulations, we successfully predicted the impact of neomycin B modifications on AME recognition (see Table II.3.5-1) and highlighted the unique behavior of each complex.

Table II.3.5-1 MD results for neomycin B binding modes (consistent or non-consistent during the simulations) and HL_171 binding modes (similar or different from that of neomycin B).

	APH(3')-IIIa	APH(3')-IIa	APH(3')-Ia	ANT(4')-Ia	AAC(3)-IIa
Neomycin B	Consistent	Consistent	Consistent	Consistent	Non-consistent
HL_171	Different BM	Similar BM	Similar BM	Different BM	Different BM

Furthermore, our simulations provided valuable insights for future modifications, particularly highlighting the significance of positions C(1) and C(2'). These findings also emphasize the importance of simultaneously employing multiple modification strategies to enhance the chances of developing an AGA which avoids AMEs recognition.

A follow-up study on new derivatives would involve docking the derivatives into the modelled AMEs (APH(3')-IIIa, APH(3')-IIa, APH(3')-Ia, ANT(4')-Ia, and AAC(3)-IIa), conducting MD simulations, and comparing their binding modes to neomycin B. Each MD replica would require approximately one day of calculations, in addition to the necessary analysis time.

Moreover, it would be of great interest to calculate the buried surface area of each AGA atom within the AMEs. This analysis would help identify the specific modifications that could have the greatest impact on the binding mode and subsequently guide the design of more effective derivatives. Additionally, exploring the buried surface area of each AGA atom within the RNA molecule could provide valuable information about modifications that may lead to a reduced affinity towards the bacterial target and thus should be avoided.

III β -clamp as an antibacterial target

III.1 Introduction

The β -clamp protein, also known as the sliding clamp, processivity factor or β subunit of the polymerase III holoenzyme, plays a crucial role in DNA replication in prokaryotes where it ensures an accurate and efficient DNA synthesis, repair, recombination and mutagenesis.^{78–80}

During DNA replication, the double-stranded DNA molecule unwinds to expose the individual strands. The β -clamp is a ring-shaped protein that surrounds one of the DNA strands, providing a stable platform for DNA polymerase to synthesize new DNA strands. It acts as a molecular "sliding clamp" by sliding along the DNA strand, keeping the DNA polymerase firmly attached and preventing its dissociation from the template strand. This processivity, which is the ability of the enzyme to remain attached to the DNA template and synthesize long stretches of DNA, is essential for the rapid and accurate replication of the genome.^{81,82}

III.1.1 β -clamp structure

III.1.1.1 Overall structure

Sliding clamps are homodimer proteins with a head-to-tail arrangement, resulting in a ring with a diameter of approximately 80 Å, and a hole of approximately 35 Å. Each monomer of the β -clamp is composed of three globular domains: an N-terminal domain, a central domain, and a C-terminal domain. Each domain consists of two adjacent and anti-parallel α -helices, with two antiparallel β -sheets supporting them. The overall structure of the β -clamp exhibits a repetitive motif throughout, resulting in a rigid protein composed of 12 α -helices and 6 β -sheets forming the outer surface (see Figure III.1.1-1).⁸¹

Despite the overall negative charge of the protein, the electrostatic field has a concentration of positive charge in the center of the ring, which matched with the expected location of the negatively charged phosphate backbone of DNA.⁸¹

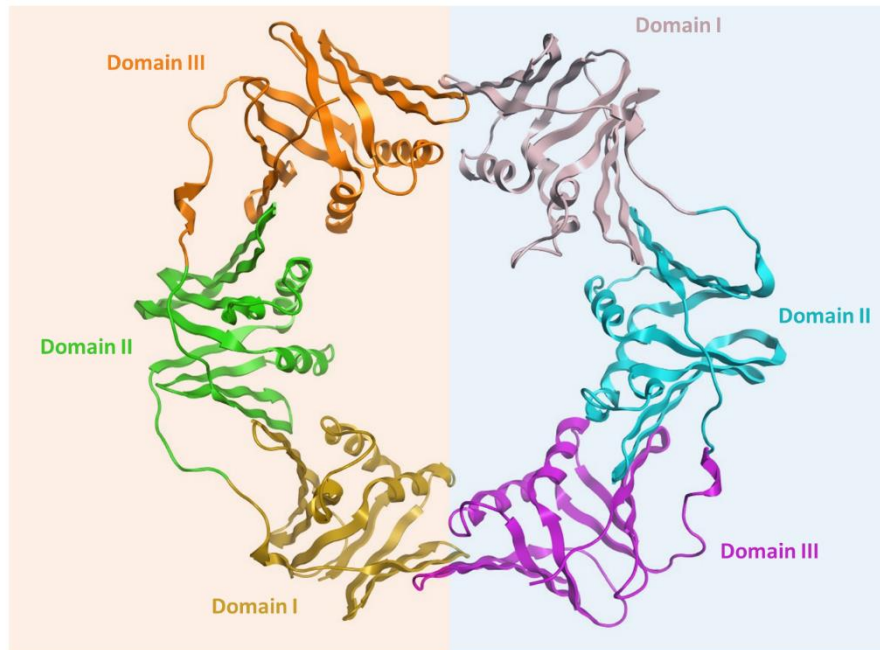


Figure III.1.1-1 Crystal structure of *E. coli* β -clamp. (PDB ID: 1OK7). The orange and blue backgrounds differentiate the two protein monomers.

III.1.1.2 Comparison with PCNA

The proliferating cell nuclear antigen (PCNA) is the eukaryotic counterpart of the β -clamp. Both proteins are ring-shaped and function as processivity factors in DNA replication. Despite originating from different organisms, they present structural and functional similarities.

PCNA is a homotrimer, where each subunit consists of two similar domains, termed the N-terminal domain and the C-terminal domain, resulting also in a six-domain ring-shaped structure. Both the β -clamp and PCNA have conserved a positively charged inner surface of the ring structure to interact with the negatively charged DNA. β -clamp and PCNA have a very similar fold even though they share less than 15% of sequence identity (see Figure III.1.1-2).⁸⁰

Regarding their functional roles, both the β -clamp and PCNA act as processivity factors for DNA polymerase during replication. They increase the processivity and efficiency of DNA synthesis by tethering DNA polymerase to the DNA template strand.⁸⁰

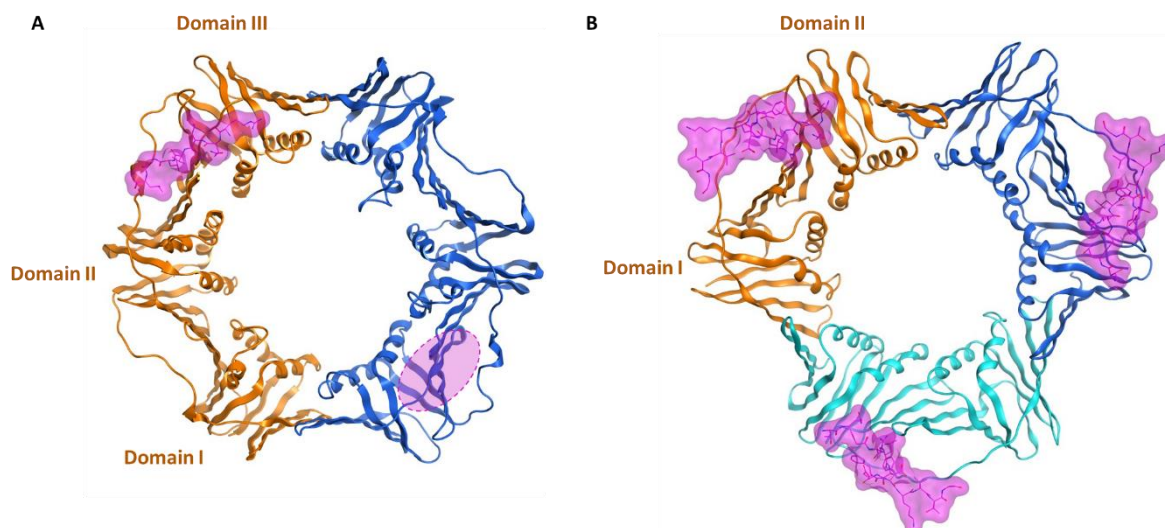


Figure III.1.1-2 DNA sliding clamps of (A) *E. coli* β -clamp in complex with a peptide of DNA polymerase IV (PDB ID: 1OK7). The second binding site is signaled by a fuchsia oval. (B) PCNA in complex with the peptides of flap endonuclease-1 (PDB ID: 1U7B). Each color corresponds to a monomer. The peptides, with its surface colored in fuchsia, show the location of the binding sites. Image adapted from the article 80.

III.1.1.3 Relationship between structure and function

The sliding clamp is considered a binding “hub” for interacting with all DNA polymerases and, thus, presents a high binding site versatility. Its binding site, called β -clamp binding motif (CBM), is located near the inter-domain connector loop between the domains II and III and is formed by two sub binding sites.⁸⁰ The sub-site 1 is a hydrophobic cleft located between the domains II and III and formed by the residues Leu155, Leu177, Pro242, Val247, and Val360. This sub-site is connected to sub-site 2 via a groove. Sub-site 2 is located in domain III and composed by the residues His175, Asn320, Tyr323, Val344, Pro363, Met362, and Met364 (see Figure III.1.1-3).⁷⁸ Furthermore, the alignment of 128 *E. coli* CBM sequences from various phylae (Proteobacteria, Actinobacteria, Firmicutes, Chlamidiae...) revealed a highly conserved binding site.⁷⁸

Gram-positive (Gram+) and Gram-negative (Gram-) bacteria share the same ring-shaped β -clamp structure, but have low sequence identity.^{83,84} When comparing the peptide binding pocket, the sequence identity between *E. coli* and Gram+ bacteria like *S. aureus*, *M. tuberculosis*, or *B. subtilis* is no higher than 56%. However, there is a good superimposition of the pockets (RMSD = 0.37 Å), except for residues in the loops.⁸⁵ Actually, certain mutations in Gram+ bacteria sub-site 1 mutations are conservative when compared to *E. coli*, maintaining the hydrophobic character of the pocket.

In fact, native peptide binders like DnaE, PolC, PolB, UmuC, MutS, or DinB often share common characteristics, including an N-terminal glutamine and two C-terminal hydrophobic residues, which are frequently aromatic. As a result, a universal peptide-CBM interaction motif was proposed based on a five-residue peptide consensus: QL(S/D)LF.⁸² Nevertheless, the peptides containing this sequence demonstrate different affinities towards Gram+ and Gram- CBM, despite their structural similarities. However, 3D-structure provides insights into the selectivity since specific residues in Gram+ bacteria prevent the peptide from lying into the groove between sub-site 1 and sub-site 2.⁸⁵

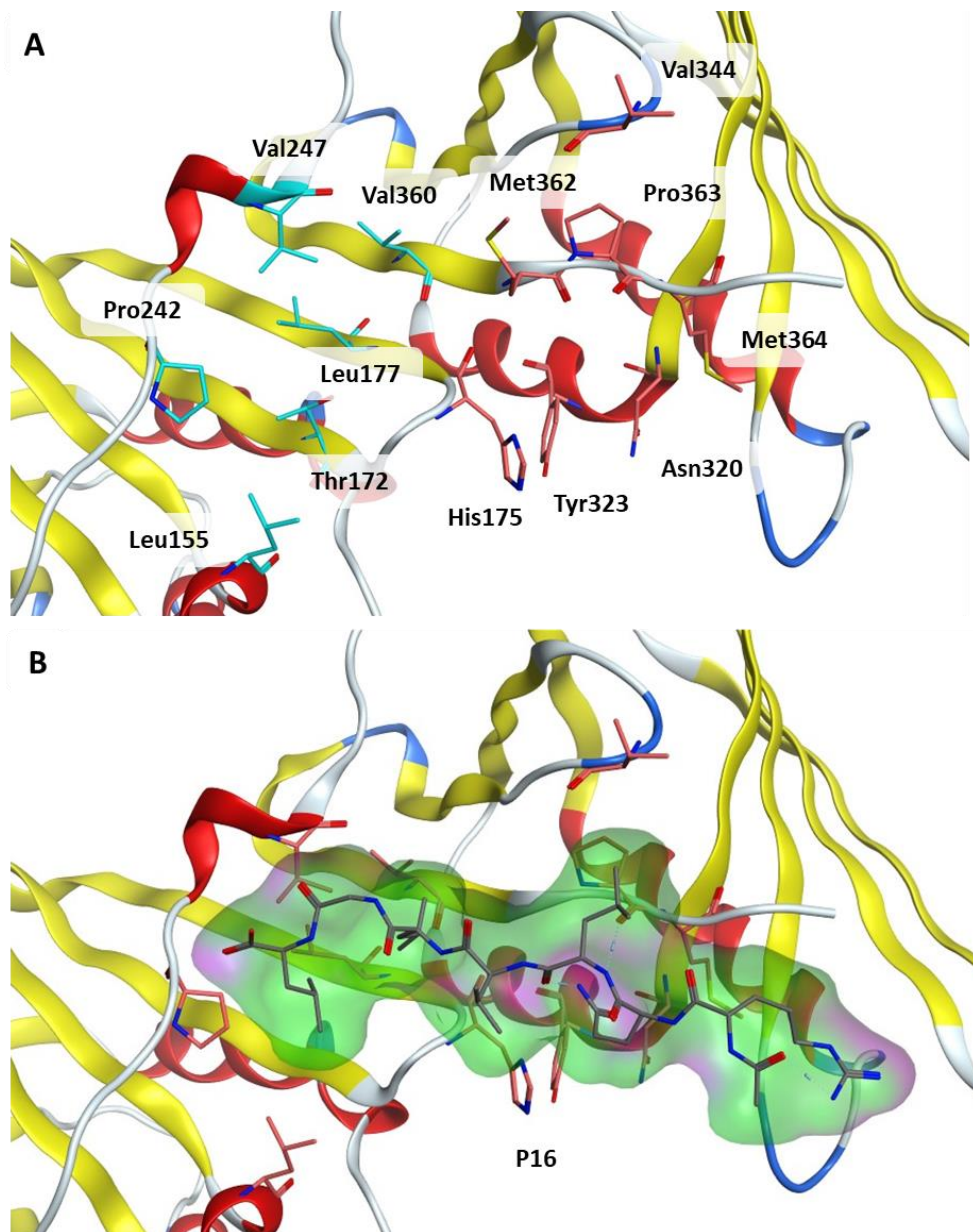


Figure III.1.1-3 3D view of *E. coli* CBM. (A) Sub-site 1 residues shown in blue, sub-site 2 residues shown in red. (B) P16 peptide bound to CBM. (PDB: 1OK7)

III.1.2 β -clamp as an antibacterial target

The CBM is the only interacting site for the β -clamp partners and is highly conserved among bacteria. Therefore, β -clamp inhibition causes bacterial cell death, stops its replication, and has also the potential to limit the emergence of resistance by blocking mutational adaptation of bacteria.^{86,87}

The point mutation of the *E. coli* β -clamp in the CBM resulted in either the inactivation of bacterial replication activity (P363A, L366A) or a decrease in activity (M364A).⁸⁸

Furthermore, an *in vitro* competition assay showed the impact of P16, a 16-residue C-terminal peptide extracted from Pol IV (see Figure III.1.1-3B), when associated to the β -clamp. The assay revealed that when used at a high concentration (25 mM), P16 effectively inhibited the β -stimulated polymerase activities of both Pol IV and the α -subunit of Pol III. Importantly, P16 did not disrupt the intrinsic activity of the polymerases, indicating that its inhibitory effect was specific to the interaction with the β -clamp. Noteworthy, the inhibition of the polymerase activity by P16 likely requires simultaneous occupancy of both binding sites on the β -clamp ring.⁷⁸

Therefore, the design of new peptides destined to be β -clamp inhibitors is normally inspired by the native CBM peptides by keeping their linear backbone and modifying the side chains to increase the affinity. The reported modifications improving the consensus peptide Q₁L₂D₃L₄F₅ (IC₅₀ = 63.2 μ M and K_d = 29.5 μ M measured by fluorescence polarization), and the native P16 (IC₅₀ = 8.85 μ M and K_d = 0.15 μ M measured by surface plasmon resonance) binding are:

- The acetylation of the glutamine Q₁, which forms a new hydrogen bond with an arginine in sub-site 2 and improves the binding by about 30-fold when compared to the consensus.^{89,90}
- The addition of a 3,4-dichlorination of the F₅ to the previous modification, which improves the hydrophobic contacts with sub-site 1 and the binding by about 3-fold.⁹¹
- The combination of the two previous modifications together with the introduction of a cyclohexyl-L-alanyl at the L₂ position, which makes additional interactions with the hydrophobic pocket and improves the binding by about 100-fold when compared to the native P16.^{89,90}

It is worth noting that peptides binding the β -clamp consist of five residues and do not bind PCNA, while PCNA peptides consist of eight residues and do not bind to the β -clamp.^{80,92}

Nevertheless, peptides often encounter challenges related to delivery and stability within the human body, making small molecule inhibitors preferable in therapeutic applications.

Among natural products, griselimycin is a partially cyclic peptide isolated from *Streptomyces* that has bacterial activity against *Mycobacteria*. Griselimycin binds with sub-site 1 and mainly forms hydrophobic interactions, with the exception of two hydrogen bonds with two arginine.⁹³

Three non-steroidal drugs, namely carprofen, vedaprofen, and bromfenac, alter *E. coli* DNA replication by interacting with CBM hydrophobic sub-site 1. However, their MIC, higher than 2 500 μ M, remains weak compared to other antibiotics (see Figure III.1.2-1).^{80,94}

Finally, some examples of small molecules inhibiting the β -clamp have also been reported: RU7, obtained by screening the Rockefeller University chemical library and has a K_i of \sim 10 μ M;⁹⁵ “compound 4”, obtained by peptide mimetics and with a IC_{50} of \sim 40 μ M,⁹¹ and “compound 8” obtained by a the screening of the Zenobia Fragment Libraries and having a K_i of 64 μ M⁹⁶ (see Figure III.1.2-1).

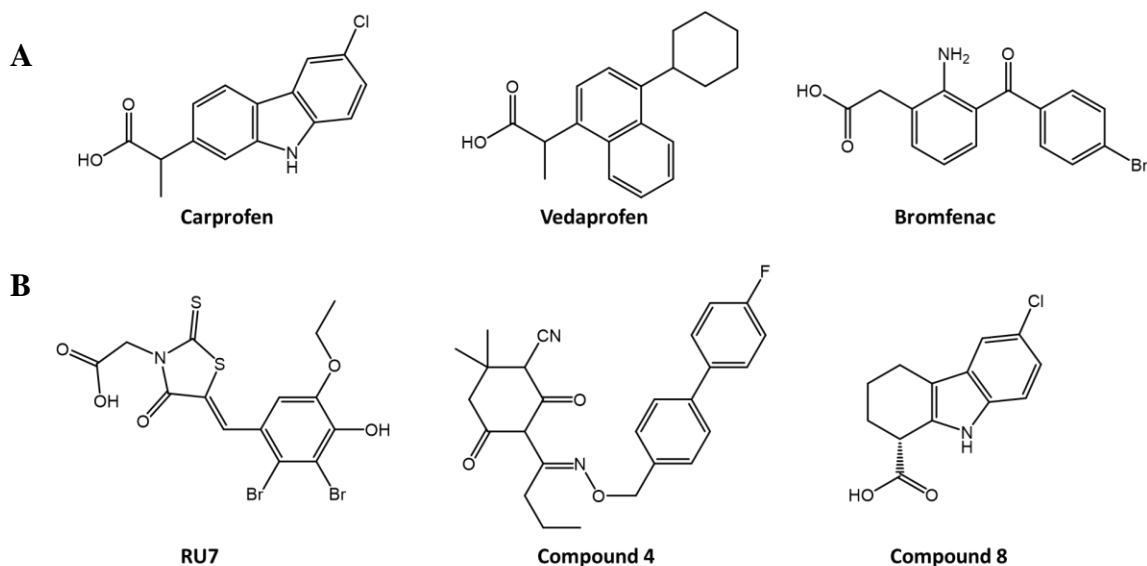


Figure III.1.2-1 Structure of β -clamp inhibitors (A) Non-steroidal anti-inflammatory drugs (B) Small molecules.^{91, 94-96}

Briefly, the β -clamp has been validated as a bacteriostatic and bactericidal target that can also limit the emergence of resistance. Various peptides have demonstrated their efficacy in inhibiting the binding site of the β -clamp by engaging with its two sub-sites. However, the

discovery of highly efficient and potent small molecules targeting the β -clamp remains yet to be accomplished.

III.1.3 Starting point

The objective of validating the CBM as a therapeutical target,⁷⁸ studying and developing peptides able to prevent the replication of the genome *in vitro*,⁸⁹ and comparing the binding of those peptides to different bacterial strains⁸⁵ was performed during the last 20 years by Dr. Burnouf's laboratory (Architecture et Réactivité de l'ARN, UPR 9002, CNRS-University of Strasbourg).

Since 2004, Burnouf team has published the structure of five peptides binding the *E. coli* β -clamp (PDB IDs: 1KO7, 3Q4K, 3Q4J, 3Q4L, and 6FVL). Furthermore, they performed a screening of a fragment library proposed by the company Diamond Light Source (DLS) on the sliding clamp. After the refinement, the crystals with a potential binding event at less than 20 Å from the binding site were kept. Finally, the filtering of the crystals without a sufficiently strong signal to build the ligand led to 94 crystallographic poses for 57 unique fragments.

III.1.4 Aim of the work

The aim of this work was to perform a virtual screening based on the crystallographic structures obtained by our collaborators in order to discover small molecules capable of targeting the sliding clamp binding site.

To do so, we first wanted to analyze the DLS fragment library that was used for the screening. Since the fragment approach is not yet used in our laboratory, we conducted one study to evaluate fragment libraries composition, diversity, and three-dimensionality. Then, the crystal DLS fragments and the small peptides binding the CBM were taken as references for guiding the virtual screening of the libraries Enamine and CN.

Hence, this chapter is organized as it follows:

- Part III.2 Fragment libraries and binding mode analysis:
 - Study of the diversity and three-dimensionality in commercial fragment libraries, offering guidelines for selecting or conceiving a fragment library. “Comprehensive

analysis of commercial fragment libraries” was published in *RSC Chem Med* in 2022.

- Analysis of the Diamond Light Source fragment database.
- Selection of DLS fragment references to be used during the VS.
- Part III.3 Virtual screening for the identification of small molecules inhibitors of *E. coli* β -clamp.

III.2 Fragment libraries and binding mode analysis

III.2.1 Introduction

The aim of this section was to step into the fragment-based drug discovery (FBDD) realm, and, through it, analyze the DSL fragment library used for screening the β -clamp. To do so, we focused on different questions:

- How can we evaluate the content of a fragment library? With a focus on diversity or three-dimensionality.
- How can we compare different fragment libraries?
- What scaffolds are rarely present in commercial fragment libraries and, thus, could bring novelty to a fragment-based project?

And finally:

- How is the DLS fragment library composed? What kind of library is it? Is it suitable for this fragment-based project? What fragments should we use to guide the virtual screening of β -clamp?

III.2.2 Comprehensive analysis of commercial fragment libraries

Julia Reville Imbernon,^a Celien Jacquemard,^a Guillaume Bret,^a Gilles Marcou^{*b} and Esther Kellenberger^{*a}

Received 11th November 2021,
Accepted 22nd December 2021

DOI: 10.1039/d1md00363a

rsc.li/medchem

Screening of fragment libraries is a valuable approach to the drug discovery process. The quality of the library is one of the keys to success, and more particularly the design or choice of a library have to meet the specificities of the research program. In this study, we made an inventory of the commercial fragment libraries and we established a methodology which allows to position any library in relation to the complete offer currently on the market, by addressing the following questions: Does this chemical library look like another chemical library? What is the coverage of the current chemical space by this chemical library? What are the characteristic structural features of the fragments of this chemical library? We based our analysis on 2D and 3D chemical descriptors, framework class generation and Generative Topographic Map. We identified 59 270 scaffolds, which can be searched in a dedicated web site (<https://gtmfrag.drugdesign.unistra.fr>) and developed a model which account to fragment diversity while being easy to interpret (download at 10.5281/zenodo.5534434).

III.2.2.1 Introduction

For the last 25 years, Fragment-Based Drug Discovery (FBDD) has widely increased in popularity as an alternative or complement to high throughput screening (HTS).^{97,98} The approach has allowed to the success of many medicinal chemistry programs, with five fragment drugs approved by the FDA and many molecules having entered clinical trials.⁹⁹ FBDD is an interdisciplinary field organized around library design, fragment screening, computational methods and optimization.¹⁰⁰ These topics are strongly interconnected. For example, the biophysical techniques suitable for the study of low affinity binders, such as X-ray crystallography or isothermal calorimetry, have a low throughput, therefore are limited to testing small-sized libraries. The screening method can also impose a bias in the chemical structure. For example, Nuclear Magnetic Resonance (NMR), a pioneer method in FBDD,¹⁰¹ can exploit fluorine magnetic properties for the detection of fluorinated ligand binding.¹⁰² More generally, the physico-chemical properties are determinant for the selection of fragments depending on the screening objective. An illustrative example is provided by the X-ray crystallography which allows the mapping of hot spots in protein using molecules with ultra-low molecular weight.¹⁰³

Fragment definition. A widely accepted definition is that fragments are compounds following the rule of three (RO3),¹⁰⁴ analogous to Lipinski's rule of five (RO5):¹⁰⁵ MW < 300 Da, AlogP ≤ 3, number of hydrogen bond acceptors ≤ 3 and number of hydrogen bond donors ≤ 3. Additional criteria have also been proposed such as number of rotatable bonds smaller than 3 and

polar surface area smaller than 60 Å². The MW limit has also fluctuated tending towards smaller molecules, which generally present a lower hydrophobicity.^{106,107} More complex rules have been developed to quantify the desirability of each compound, such as the quantitative metric of drug-likeness (QED), which ranges from 0 to 1.¹⁰⁸ However, QED is designed for full-sized drug-like molecules: as a consequence, QED values are biased when applied to fragment-like compounds. In this work, we expect this bias to be equal on all investigated situations. Recently, Konteatis has proposed eight guidelines for the selection of fragments, stressing that priority should be set on the enthalpy-driven binding interactions, followed by the diversity molecular shape of the target-complementary fragments.¹⁰⁹ Indeed, Keserü *et al.* have theoretically and experimentally confirmed the importance of enthalpic contribution to prioritize fragments forming key interactions.¹¹⁰ The success of fragment-to-lead optimization depends on the target-anchoring capacity of fragments, which in particular determines if the binding mode is conserved during fragment growing. Analysis of fragments binding modes in the Protein Data Bank suggested to prioritize fragments that are able to form polar interactions and have a molecular weight larger than 150 Da.¹¹¹

Fragment libraries. Suppliers of chemical libraries offer a large panel of screening collections of fragments. The content of fragment libraries varies depending on their purpose. A fragment library can be tailored to a precise screening method, such as, for example, NMR.¹⁰² Fragment screening can not only deliver hits, yet also suits for protein binding site detection, druggability assessment or study of versatile binding modes.¹¹² Library design in general focus on a specific need, such as, for example, the ability of fragment to form covalent bond, which

^aLaboratoire d'Innovation Thérapeutique, UMR7200 CNRS Université de Strasbourg, Institut du Médicament de Strasbourg, Université de Strasbourg, faculté de Pharmacie, Illkirch-Graffenstaden France.

^bLaboratory of Chemoinformatics, UMR7177 CNRS Université de Strasbourg, Strasbourg, France.

imposes particular functional group, or else the ability of the library to offer a wide diversity of chemical structures.

The size of the chemical libraries is variable, with a number of fragments ranging from a few hundred to over a hundred thousand. From the early days of FBDD, Hann *et al.* have noticed the relationship between molecular complexity and the likelihood of success in hit discovery,¹¹³ suggesting that screening a thousand fragments cover the chemical space in a more effective way than a million HTS compounds. More recently, Shi and von Itzstein engaged in a theoretical exercise aiming to determine how many fragments are sufficient to represent the 227 787 commercial fragments of the ZINC database.¹¹⁴ Their comparison of different sized subsets, using a measure of diversity taking into account both the number of unique structural fingerprints and their relative abundance in the subset, revealed that a subset of approximately 2 000 fragments can have the same level of diversity as the full set of fragments.

Three dimensionality. For the past ten years, three-dimensionality has been a major concern in medicinal chemistry to produce molecules that improve both the originality of drug candidates and their selectivity. In the seminal article by Lovering *et al.*, the analysis of a huge number of compounds sampled at different stages of the drug development process suggested that complex molecules are more likely to become drug candidates than simpler molecules. This parameter was evaluated by the fraction of sp³ hybridized carbons and the number of chiral centers.¹¹⁵ The interest in non-planar scaffolds has given rise to numerous developments in synthetic methods.^{116–120} In FBDD programs, planar fragments are often avoided, since they tend to bind multiple targets, and therefore are potentially more difficult to develop into a selective drug-like compound.¹¹² Nonetheless, the successful ligand design from flat fragments have fueled the debate, demonstrating that the flatness of the fragments does not necessarily transfer to the lead.¹²¹

In this study, we propose a comprehensive analysis of commercial fragment libraries available in 2021, with emphasis set in diversity and three-dimensionality. This analysis is designed so that it can be updated and made available on a yearly basis, to follow the evolutions of the content and availability of commercial fragment libraries.

III.2.2.2 Materials and methods

Collection of fragment libraries and classification. A total of 86 different fragment libraries from fourteen vendors were collected between February 22nd-24th, 2021. These libraries were categorized by composition, as annotated by the supplier, into ten main categories (Table III.2.2-1): general (GEN), diverse (DIV), nature-product like (NP), 3D-shaped (3D), sp³ hybridization (SP3), metal chelating (CHE), protein-protein interactions (PPI) fluorine-rich (FLU), halogen-rich (HAL), and covalent binders (COV). The chemical libraries that did not fall into any of these categories were called miscellaneous (MIS).

Table III.2.2-1 Summary of libraries classification

Type	Description	Number of libraries		
		Small-sized	Middle-sized	Large-sized
GEN	General	2	3	9
DIV	Diverse	7	1	0
NP	Natural product-like	2	2	0
3D	3D-shaped	3	1	0
SP3	Sp ³ hybridized	1	2	2
CHE	Metal chelators	3	0	0
PPI	Protein-protein interacting	1	3	0
FLU	Fluorine enriched	7	3	1
HAL	Halogen enriched	3	4	0
COV	Reactive covalent linking	4	5	0
MIS	Miscellaneous	8	7	2

Standardization of fragments and calculation of molecular properties. Pipeline Pilot (v.19.1.0.1964; BIOVIA) was used to generate the canonical SMILES of each molecule from the structure files provided by the suppliers, then to remove counter ions of salts. Filter (v.2.5.1.4; Openeye Scientific Software, Santa Fe) was used to filter out simple ions and invalid chemical structures (for example containing a pentavalent carbon atom) and to ionize the molecules at pH=7.4. The pyridone, pyridine and sulfonamide description were corrected, and fragments were aromatized, using basic aromatic scheme, with the Standardizer (v16.10.17.0; ChemAxon Ltd.) software. Pipeline Tools was used to calculate 2D descriptors: number of hydrogen bond acceptors (HBA) and donors (HBD), as defined by Lipinski's Rule; AlogP, computed with the Ghose/Crippen group-contribution estimate for LogP¹²²; polar surface area (PSA); number of rotatable bonds (Nrot); quantitative estimate of drug-likeness (QED)¹⁰⁸. 3D descriptors were calculated by using Openeye libraries (v.2019.10.2) from a low energy conformer generated for each molecule by the Corina (v.3.40; Molecular Networks GmbH, Nürnberg, Germany) software with its default settings: solvent-accessible, surface area (SASA), polar surface area (3D-PSA), plane of best fit (PBF)¹²³.

Generation of fragments classes. The fragments from all libraries except those built with a chemical bias (i.e. COV, PPI, HAL, FLU and the MIS constructed around a specific functional group) were merged into a single file. Duplicates, as identified from identical canonical SMILES were removed using Pipeline Pilot (BIOVIA, Dassault Systèmes, Pipeline Pilot, Release 2019, San Diego: Dassault Systèmes).

Fragments were imported into ADMET Predictor 2019 (ADMET Predictor™, Simulations Plus, Inc., Lancaster, California, USA), neutralized, standardized, and classified using the option "Frameworks". The method is based on ring systems and connecting chains, similar to the so-called "Murcko assemblies" described by Bemis and Murcko¹²⁴. It yields only one scaffold per molecule. Each fragment belongs to only one class,

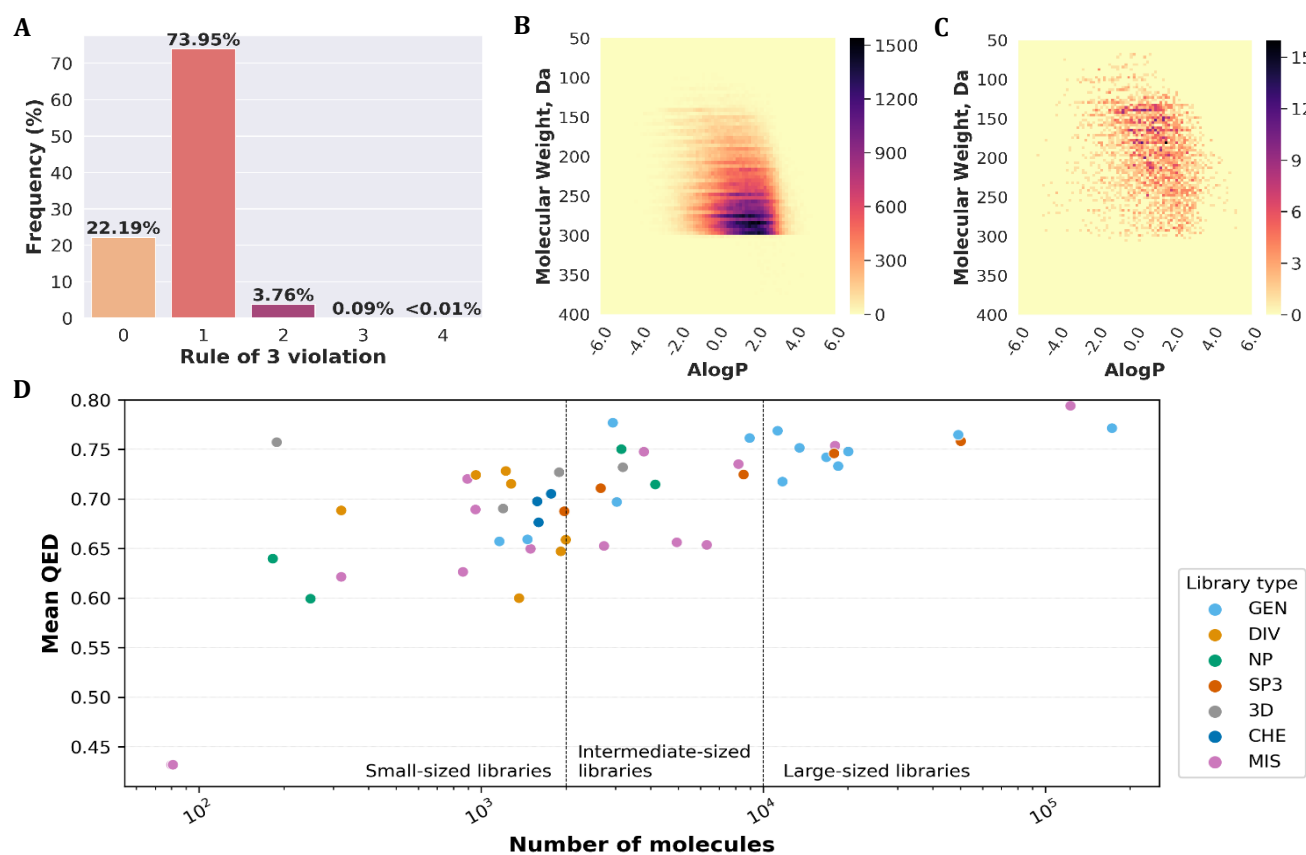


Figure III.2.2-1 Physico-chemical properties of 512 284 commercial fragments. (A) Percentage of fragments which comply with the RO3 and which violate one to four of the rules (B) Heatmap showing the distribution of MW and AlogP for the entire fragment set (C) Heatmap showing the distribution of MW and AlogP for the subset of fragments in the small-sized libraries (D) Logarithmic scatter plot representing the distribution of mean QED depending on the number of molecules and the type of library. The black dashed lines indicate the limit between small-sized libraries, intermediate-sized libraries, and large sized-libraries.

corresponding to its major ring system, and each class is defined by a specific scaffold.

Generative Topographic Maps (GTM). The training set was constituted of 6 017 fragments randomly picked in the small databases. Molecular descriptors were generated using ISIDA/Fragmentor (v.2019, Faculté de Chimie, Université de Strasbourg, France),¹²⁵ with atom centered fragments based on sequences of atoms and bonds, a length ranging from 2 to 4 and tagging carbon atom if part of a ring and/or of an aromatic system, IAB(2-4)_cycle. Those descriptors present in less than 5% of the dataset have been discarded.

The manifold generation and projections were carried out with the GTMTool (v.2018, Faculté de Chimie, Université de Strasbourg, France). A pre-processing procedure, to center the dataset, was used and the total number of traits was set to 301. The best model was chosen based on the likelihood. The GTMClass tool was used to generate property landscapes specific of supplier categories (Table III.2.2-1). Landscape depiction was performed using the GTMLandscape software.

Dataset availability. Datasets are available on Zenodo (doi: 10.5281/zenodo.5534434). The deposit contains the chemical structures, annotations, the computed molecular descriptors, the scaffolds, and models.

A web interface was designed to search the scaffolds. It is freely available at: <https://gtmfrag.drugdesign.unistra.fr>. Molconvert

was used to standardize structures and convert between file types, JChem base was used to store structures in a postgresql instance and is used to query the database and Marvin.js was used for drawing structures, LTS-helium (21.4.0), ChemAxon (<https://www.chemaxon.com>)

III.2.2.3 Results and discussion

Overview of commercially available fragments. Our survey identified 86 fragment libraries, provided by fourteen 14 chemical suppliers, for a total of 754 646 molecules with 512 284 unique chemical 2D-structures.

These libraries have been designed with various specifications, for purposes which are generally well described, thereby allowing us to classify most of the libraries by type (Table III.2.2-1). Firstly, twenty-seven libraries show chemical bias: enrichment in fluorine for screening by ¹⁹F-NMR (FLU type, eleven libraries), enrichment in other halogen elements for harnessing halogen bonding (HAL type, seven libraries)¹²⁶, and presence of a reactive group for covalent linking to the target (COV type, nine libraries). For eleven other chemical libraries, the construction rules consider the whole chemical structure of fragment to focus on particular chemotypes: natural product or natural product-like fragments (NP type, four libraries), compounds targeting protein-protein interactions (PPI type,

four libraries), or potential metal-chelating compounds (CHE type, three libraries). There are also eight libraries which aim at maximizing the coverage of the chemical space by optimizing molecular diversity within the set of fragments (DIV type). Avoiding an overrepresentation of flat fragments was the determining factor in the design of nine libraries, by favoring fragments with a non-flat 3D-shaped (3D type, four libraries) or fragments with a high content in sp³-hybridized atoms (SP3 type, five libraries). The other libraries have each their own special properties and thus were qualified as miscellaneous (MIS type, seventeen libraries, e.g. library containing fragments which are well soluble in both dimethylsulfoxide and phosphate-buffered saline buffer), or do not present specific characteristics and thus were qualified as general (GEN type, fourteen libraries).

Libraries were also classified according to their number of fragments into the three following categories: small ($\leq 2\,000$ fragments, 41 libraries), intermediate (2 000 – 10 000 fragments, 31 libraries) and large ($> 10\,000$ fragments, 14 libraries). Only small-sized libraries are suitable for testing by biophysical approaches, such as NMR. Intermediate ones are accessible for medium throughput screening techniques. The large libraries offer a wide choice for the custom design of small chemical library. The small-sized libraries taken as a whole contain 33 023 different fragments. This subset covers 6.4 % of all commercial fragments. By comparison, the intermediate libraries taken as a whole contain 94 471 different fragments, covering 18.4 % of all commercial fragments.

Commercially available fragments conform with RO3. In 2003, Congreve et al. defined the “Rule of 3” (RO3) as follows: $MW \leq 300$ Da, $AlogP \leq 3$, $HBA \leq 3$, $HBD \leq 3$.¹⁰⁴ As shown in Figure III.2.2-1A, less than 23 % of the fragments strictly fulfill the four conditions and 74 % break one rule, mainly that about HBA (60 % of the cases). Most of the fragments have a MW in the 200-300 Da range (Figure III.2.2-1B). Only 13.43 % of the fragments have a MW lower than 200 Da. These small fragments, which generally have a smaller number of atoms and a lower molecular complexity, have an increased chance of binding to a protein.¹¹³ Moreover, as low affinity binders with low specificity, these small fragments constitute chemical probes useful for the detection of binding site and hotspots as well as druggability studies.¹¹² Several suppliers have developed libraries of low MW fragments, sometimes called the Mini-Frag libraries. Noteworthy, the proportion of small fragments has surprisingly decreased over the two last years (24.29 % in 2019). Together with MW, lipophilicity is an important parameter to prioritize compounds in a medicinal chemistry program.^{127,128} Considering logP, again most of the fragments fall in a narrow range (0-2), with a high proportion of weakly hydrophobic molecules (46 % with $AlogP < 1$, Figure III.2.2-1B). Considering only the small-sized libraries (Figure III.2.2-1C), MW and logP distributions are more spread out, with a shift of the maximum of the MW distribution to a lower value, as compared to the full set of fragments.

The drug-likeness was further evaluated using QED. In their publication presenting the QED, Bickerton et al. set at 0.67 ± 0.16 the threshold value that defines whether a compound is an attractive starting point for hit optimization.¹⁰⁸ They observed that “too simple” unattractive compounds have mean QED equal to 0.49 ± 0.23 , “too complex” ones have mean QED equal to 0.34 ± 0.24 . The QED mean values computed here for each fragment library are represented in Figure III.2.2-1D. All libraries but Mini-Frag show mean QED exceeding 0.6, i.e. close or passing the threshold proposed for drug-like compounds, although MW, which is taken into account in the QED, is lower in fragments. QED mean values are not related to the library type.

Diversity of fragments. In order to assess the chemical diversity of the entire fragments collection, we used two different and complementary approaches: the classification of the fragments by common chemical scaffolds and the mapping of the chemical space covered by the fragments using GTM. These analyses have taken into account the libraries of the following types: GEN, DIV, NP, CHE, SP3, 3D and MIS (list in Table S1). PPI libraries were removed because their MW is significantly higher than the MW given by the fragment definition used for this study. FLU and HAL libraries were discarded because they would have introduced a bias in the representation of the chemical space. COV libraries were removed because the recurrent substructures present on reactive covalent linking groups fragments would also have caused a bias in the analysis of the chemical space. The diversity analyses were hence performed on an ensemble of 433 433 unique fragments from fifty libraries provided by fourteen suppliers (Table S2). The classification of all fragments by common substructure yielded to the inventory of 59 270 scaffolds. Not surprisingly, the number of scaffolds represented in a chemical library depends on the total number of fragments in the library (Figure III.2.2-2A and Figure III.2.2-2B). If the relationship is generally linear with a ratio between the number of molecules and the number of scaffolds approximatively equal to five, variations are notable between the small-sized chemical libraries, with a ratio ranging from 1.30 to 5.01. Interestingly, small-sized DIV libraries include both some of the most and the least diverse scaffolds as illustrated by the range of ratio values they cover (Figure III.2.2-2B). This suggests that some commercial collections of DIV fragments are not obtained by maximizing the number of different scaffolds. More generally, the type of a library does not allow to presume of the number of scaffolds it contains.

If we consider the 59 270 scaffolds as a whole, their representation among the complete set of 433 433 fragments is highly uneven. About 0.1 % of the scaffolds are present in more than a thousand of fragments, hence covering each 25 % of the entire set (Figure III.2.2-2C). These scaffolds are mainly simple and aromatic, such as benzene and nitrogen heterocycles. The structures of the fifty more popular scaffolds are shown in Table S3. At the other end of the distribution of scaffolds by frequency, there are 36 555 scaffolds defined by a single fragment and here called singletons.

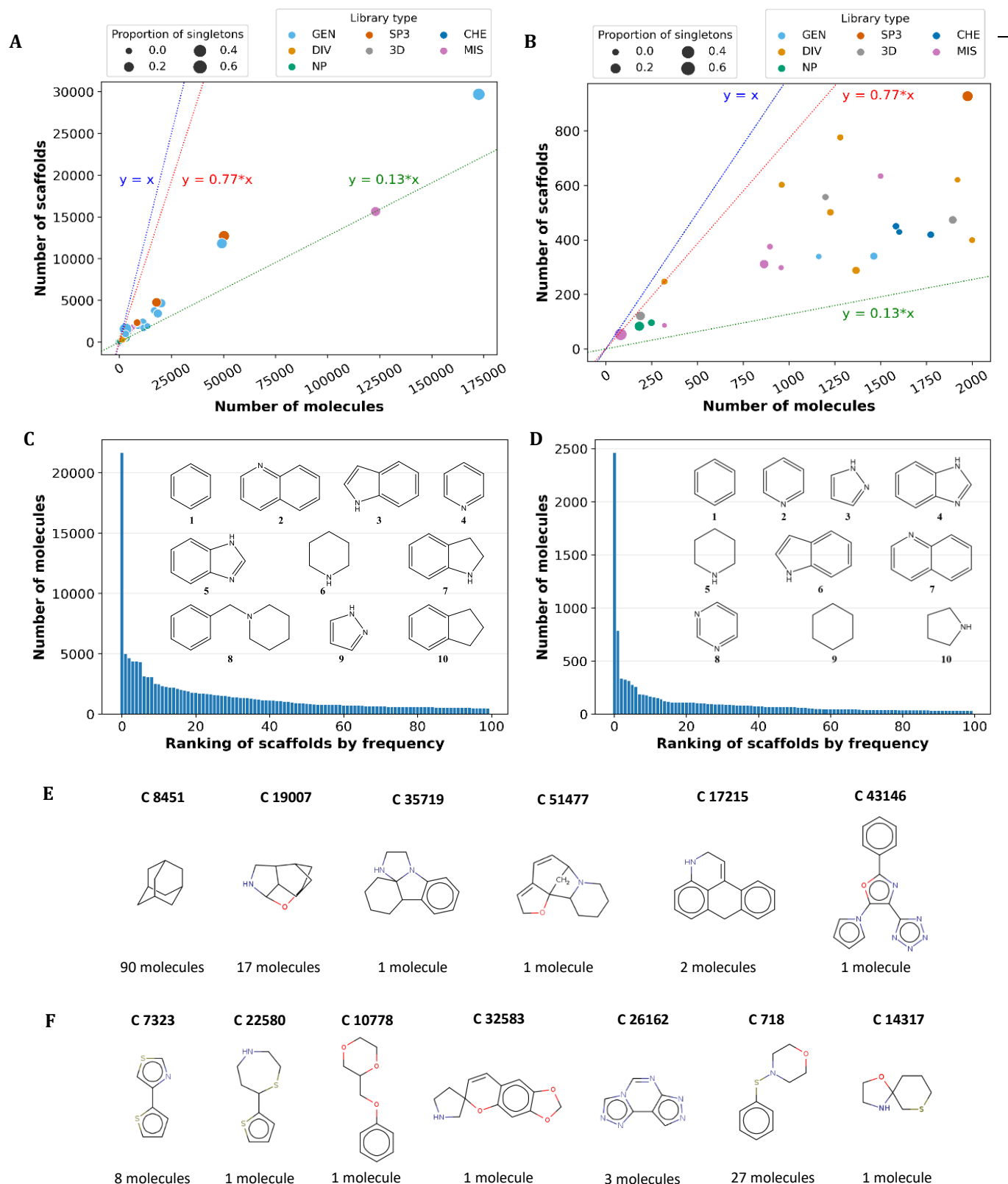


Figure III.2.2-2 Diversity of fragment scaffolds. (A) Scatter plot showing the number of scaffolds in a library as a function of the number of fragments in the library. The dot size indicates the proportion of singletons, *i.e.* scaffolds present in a single copy in the full set of fragments. The blue dotted line depicts $x=y$, the red and green dotted lines respectively depict the maximum and minimum ratio between the number of classes and the number of molecules. (B) Zoom on the portion of the scatter plot (A) corresponding to small-sized libraries. (C) Bar plot showing the number of molecules that contain a scaffold as a function of the resulting rank of the scaffold, for the hundred most frequent scaffolds in the full set of fragments. The structures of the ten most common scaffolds are shown in the insert. (D) Bar plot showing the number of molecules that contain a scaffold as a function of the resulting rank of the scaffold, for the hundred most frequent scaffolds in small-sized libraries. The structures of the ten most common scaffolds are shown in the insert. (E) Selection of complex scaffolds. The upper number corresponds to the identification number of the scaffold. The number of molecules state the number of compounds containing the scaffold. (F) Selection of scaffolds rich in heteroatoms. The upper number correspond to the identification number of the scaffold. The number of molecules state the number of compounds containing the scaffold.

Table III.2.2-2 Comparison of scaffolds properties and frequency.

	Frequency	Number of scaffolds	Number of non-hydrogen atoms (Median)	Number of rings (Mean \pm SD)	PBF (Mean \pm SD)
Scaffolds present in all libraries	Singletons	36555	17	2.79 \pm 0.65	0.62 \pm 0.17
	Intermediate	22669	15	2.42 \pm 0.56	0.56 \pm 0.17
	Popular ^a	46	9	1.76 \pm 0.43	0.46 \pm 0.17
Scaffolds present in small-sized libraries	Singletons	2 356	13	2.30 \pm 0.58	0.49 \pm 0.20
	Intermediate	1 426	11	2.05 \pm 0.42	0.43 \pm 0.20
	Popular ^a	46	9	1.70 \pm 0.58	0.30 \pm 0.17

^aScaffolds appearing in at least 1 000 times in all libraries or 66 times in small-sized libraries. These thresholds have been set as the top 1/5 of the highest frequency value (most common scaffolds have been previously discarded: benzene for all libraries set and benzene and pyridine for small-sized libraries set).

Interestingly, the proportion of singleton may be related to the library specificity (Figure III.2.2-2B and Table S2). For example, small and diverse-typed libraries, which are designed to maximally cover the chemical space, have a low proportion of singleton-scaffolds. By contrast, sp^3 -typed libraries, which are focused on a more specific chemical space, present a higher proportion of singletons.

Focusing on small-sized library, about 6.4 percent of the 59 270 scaffolds are represented among the set of 20 708 unique fragments. Even though the ratio of scaffolds per number of molecules indicates that small-sized libraries are more diverse than the entire dataset (0.18 against 0.14 respectively), the distribution is even more unbalanced. In fact, the most frequent scaffold, also corresponding to the benzene, represents 12 % of the molecules instead of 5 %. The structures and frequencies of the fifty more common scaffolds can be seen in Table S3. The diversity has been obtained through the singletons, which represent a 62 % of the scaffolds.

Less frequent scaffolds generally have a higher complexity. Their number of non-hydrogen atoms, number of rings, and PBF are on average higher than for the most common scaffolds (Table III.2.2-2). By focusing on the scaffolds present in small-sized libraries, this difference in complexity between common and rare scaffolds is maintained. If the comparison of common and less frequent scaffolds again shows that less frequent scaffolds have on average a higher number of non-hydrogen atoms, a higher number of rings and a higher PBF, it also reveals that the average complexity of small-sized libraries scaffolds is reduced (Table III.2.2-2 and examples in Figure III.2.2-2D).

For a more qualitative description of the scaffolds in the small-sized libraries, we analysed the number of rings and the number of nitrogen, oxygen and sulfur atoms. The majority of scaffolds contain two (76.8%) or three rings (17.0%). A hundred complex scaffolds contain four or five, which are either fused or bridged rings, forming a compact geometry (adamantane-like, as in C 8451 and C 19007), 3D (as in C 35719 and C 51477) or linear (as in C 17215), or which are separated by a central aliphatic linker of length 1 to 4. A unique scaffold shows a central ring acting as a platform on which are attached, via linkers, three other rings exploring three directions in space (C 43146). This same scaffold (C 43146) is also remarkable because it contains six nitrogen atoms and one oxygen atom. The examples are shown in Figure III.2.2-2E. While almost all scaffolds contain at

least one atom of nitrogen, oxygen or sulfur, nitrogen is by far the most common heteroatom (1-6 N in 93.4% of scaffolds), ahead of oxygen (1-3 O in 30.7% of scaffolds) and sulfur (1-2 S in 21.9% of scaffolds). Structures illustrating the polysulfurized (C 7323; C 22580), polyoxygenated (C 10778; C 32583), polynitrogenous (C 26162), and containing both a nitrogen atom, oxygen and sulfur (C 718; C 14317) are shown in Figure III.2.2-2F.

Fragment space mapping. To further study the chemical diversity covered by commercial fragments, we developed a Generative Topographic Map (GTM) model able to represent the chemical space in a landscape.¹²⁵ This method has the advantage of allowing a simple visual analysis. At a glance, the overall comparison of two maps gives the degree of resemblance between two sets of fragments. Local inspection of a map reveals information at a finer scale, such as chemical substructures which are well represented in a library. The GTM model was used to evaluate each library against all the fragments and to compare the libraries with each other.

Figure III.2.2-3 shows how the constructed GTM model has distributed the fragments on the chemical landscape. Detailed visual analysis of the GTM map validated the consistency between the spatial proximity of the projection points of the fragments and the similarity of chemical structure of these fragments. Thus, it is possible to associate particular chemical characteristics with the different areas of the map, as illustrated in Figure III.2.2-3A with an area rich on unsaturated molecule (**R1**), an area populated by derivatives of benzoimidazole (**R2**) and an area where all fragments contain fluorine (**R3**). Within an area, the fragments are distributed in space depending on complexity (number of rings, presence of heteroatoms, number of substituents) and the type of functional groups. For example, as shown in Figure III.2.2-3B, in the lower part of **R3**, a path going from right to left will successively meet fragments containing a trifluoromethyl group and a phenyl group, then those containing a trifluoromethyl group and a nitrogenous heterocycle, and finally those containing only a single fluorine atom and a nitrogenous heterocycle.

In the GTM landscapes, likelihood is represented through a gradient color and depicts the probability distribution of the data. Figure III.2.2-3C shows that the density of the entire dataset covers the entire map, yet the distribution is clearly heterogeneous. Purple regions correspond to the most

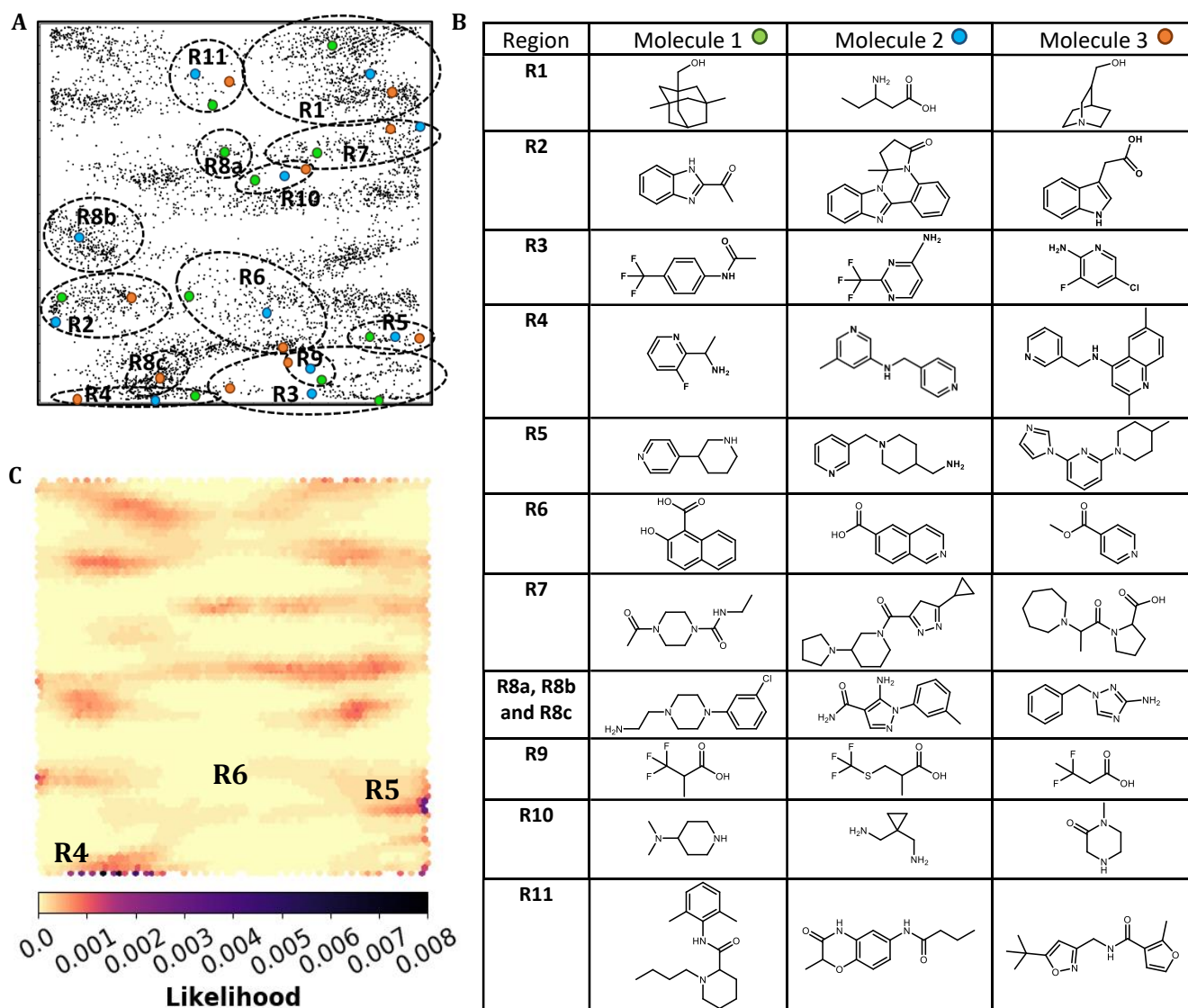


Figure III.2.2-3 Spatial distribution of the fragments in the GTM. (A) Visualization of the composition of the GTM model. Each point in the map corresponds to a single compound. The regions explained on the text are delimited by a black dashed line. The green, blue and orange dots correspond to the position of the molecules depicted in (B) Table depicting three molecules per region. The colored dot link the molecule to its location in the GTM model (A). R8a, R8b and R8c correspond to different chemical regions but have been grouped for the discussion. (C) GTM Landscape of the full dataset – 433 433 fragments. The likelihood represents the probability distribution of data in the chemical space. The annotations represent specific regions to help the reader locating them during the discussion.

populated ones. The widely dense south-west area (R4) collects a broad set of pyridine derivatives connected to nitrogenated substitutes. As the populated south-east region (R5) is rich in fragments containing a piperazine connected to a pyridine, both regions could be related. The center-south region (R6), enclosing the benzoic acid derivatives, appears to be the less occupied one.

Coverage of the fragment landscape by libraries. Figure III.2.2-4A–D show the GTM landscape of the different fragments subsets as defined by the well-characterized library types DIV, NP, CHE, SP3 and 3D. Fragments in SP3 and 3D libraries, which both focus on three-dimensionality of molecules, were merged into a single set. The projections of the subsets on the GTM show different trends, most of which are

expected from the purpose of the libraries that compose them. The DIV subset is representative of the whole chemical space as it contains almost all the chemical clusters of the GTM model applied on the full fragments' dataset (Figure III.2.2-4A). Nevertheless, the area containing the unsaturated cyclic and linear fragments (R1 in Figure III.2.2-3A and R1 in Figure III.2.2-4A) is less populated on its left part, where the more complex molecules are located. Also, right south to this region (R7), an elongated area grouping molecules rich in piperidine amides is almost empty. The SP3-3D landscape also well represents the whole data distribution (Figure III.2.2-4B), yet as expected, chart density is concentrated on the upper area, where the unsaturated compounds are located (R1). The NP landscape (Figure III.2.2-4C) also has a great density among the

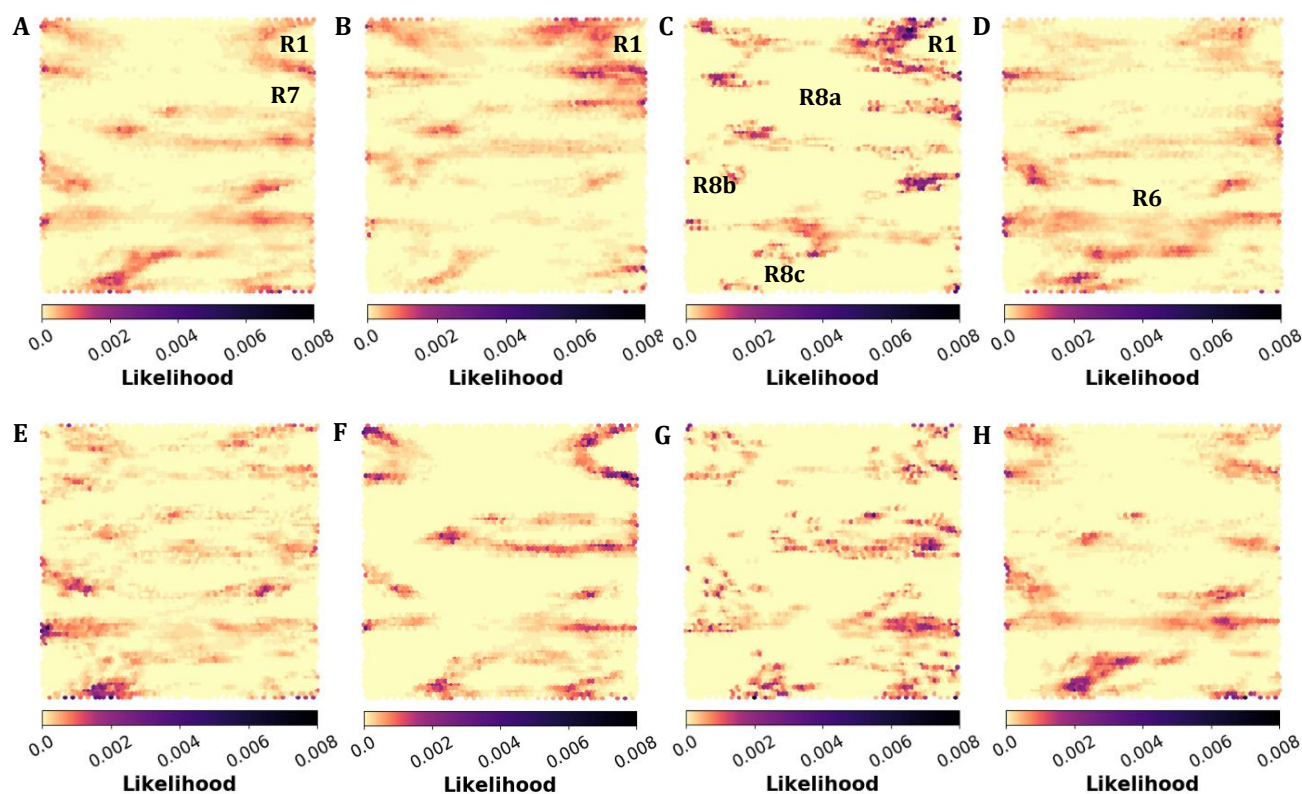


Figure III.2.2-4 GTM landscape of small-sized fragments subsets. (A) DIV libraries – 7 454 fragments (B) SP3 and 3D – 5 236 fragments (C) NP libraries – 431 fragments (D) CHE libraries diverse – 3 758 (E) DIV library 3 – 1 226 fragments (F) DIV library 6 – 1 920 fragments (G) DIV library 1 – 320 fragments (H) DIV library 7 – 2 000 fragments. The likelihood represents the probability distribution of data in the chemical space. The annotations represent specific regions to help the reader locating them during the discussion.

unsaturated compounds (**R1** in Figure III.2.2-3A and **R1** in Figure III.2.2-4C), including the area where the most complex ones are found, yet misses other regions of whole fragment landscape, in particular the regions with a benzene linked to multiple-nitrogenated cycles are almost empty (**R8a**, **R8b** and **R8c**). For example, the piperazine area is not populated. The CHE subset better covers the whole fragments map than the NP subset (Figure III.2.2-4D). In addition, its projection on the GTM well fills the center-south area (**R6**), which is sparsely populated in (**R6**) of Figure III.2.2-3C. Hence, metal chelators libraries contain specific fragments that are not common in the general dataset. If the DIV libraries taken together well represent the whole fragment space, they are not all equivalent (Figure III.2.2-4E–H). Comparison of the DIV libraries 3 and 6 gives a good illustration (Figure III.2.2-4E and Figure III.2.2-4F respectively). The two sets, respectively made of 1 226 and 1 920 fragments, occupy the space in a very different way, although both contain fragments representing every populated cluster present in the full dataset (Figure III.2.2-3C). The DIV library 3 has a homogenous coverage of the chemical space, while the DIV library 6 is more representative of the density. The comparison between DIV library 1 and DIV library 7 (Figure III.2.2-4G – 320 fragments and Figure III.2.2-4H – 2 000 fragments respectively) illustrates that small-sized libraries can correctly cover the ensemble of the chemical space. Most clusters are covered by both libraries. However, DIV library 1 has a bigger presence on the north part than DIV library 7, despite being 6 times smaller.

Fragment three-dimensionality. The 3D character of fragments was evaluated using the Plane of Best Fit (PBF) method, which fits a plane on the 3D structure of the molecule to issue the average distance of atomic coordinates from the plane, the PBF score.¹²³ If molecule lies on the plane, i.e. is flat, PBF equals 0. Otherwise, PBF increases as a function of the scaffold geometry and the position and size of substituents. The distribution of PBF among the complete dataset of fragments is bimodal with a major peak at 0.62 ± 0.20 and a minor peak at 0.01 ± 0.03 , allowing the distinction of flat fragments (7.8 % of fragments with PBF < 0.1 Å) and fragments with non-planar geometry (92.2 % of fragments with PBF \geq 0.1 Å). A large part of flat fragments has no rotatable bonds (Figure III.2.2-5A). The proportion of this type of fragments is greater in the subset of small-sized libraries than in the whole dataset (8.6 % against 21.9 %), suggesting that currently, planarity is not a strong filter in the design of many fragment libraries (Figure III.2.2-5A and Figure III.2.2-5B). Accordingly, the percentage of flat fragments in small-sized libraries has significantly raised since 2019 (16 %). Considering each library individually, the annotations by type and the three-dimensionality of fragments are consistent (Table S2). As expected, the 3D and SP3 libraries show the highest average PBF (\geq 0.6 Å). The NP libraries also show high average PBF, in agreement with the known complexity of many natural products.¹²⁹ On the other hand, the MiniFrag libraries, which are overall poorly diverse (MIS libraries 1 and 2 in Table S2), are mostly composed flat fragments (mean PBF \leq 0.17 Å).

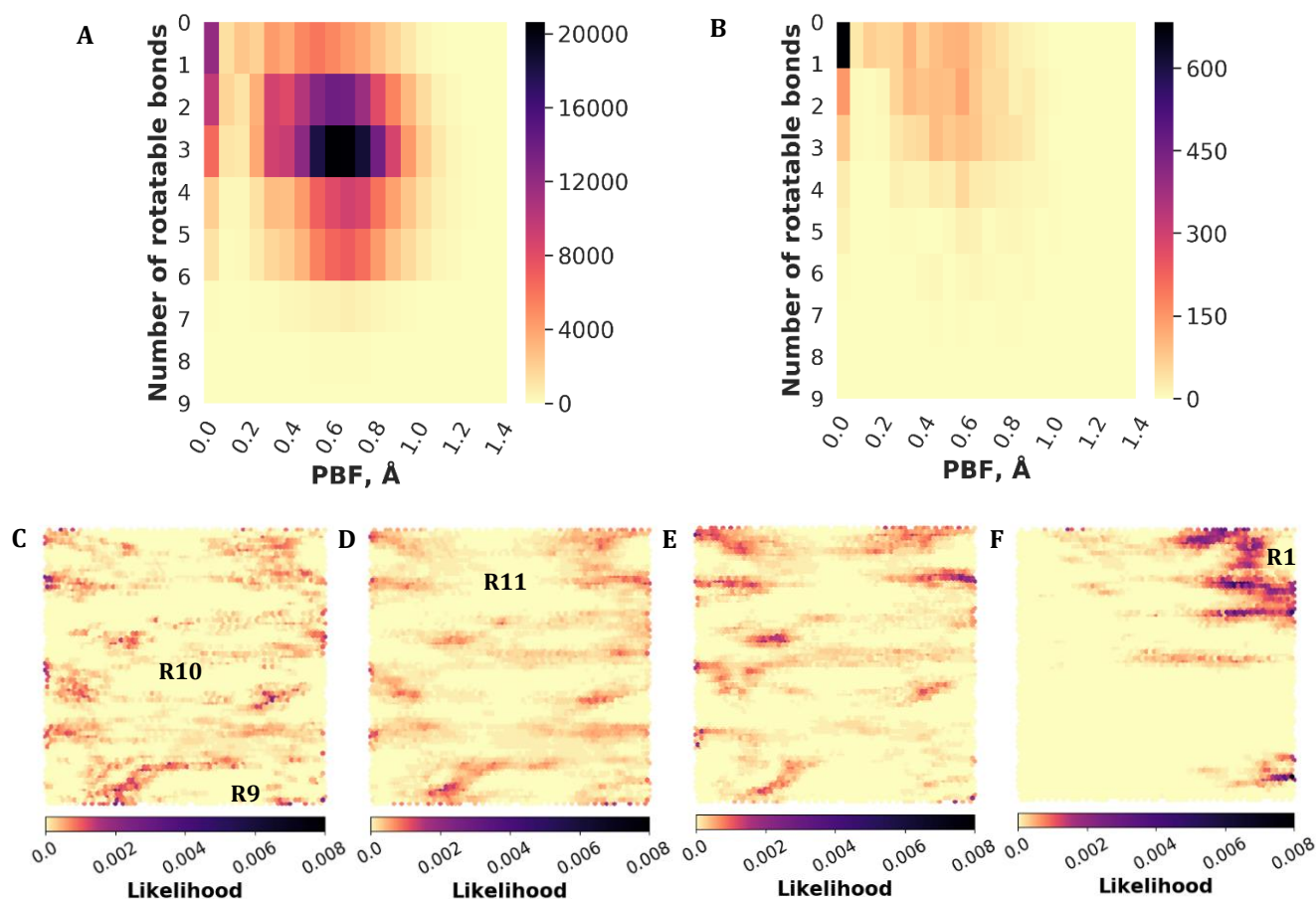


Figure III.2.2-5 Three-dimensional properties. (A) Heat map showing the distribution of the number of rotatable bonds and PBF for the entire fragment set. (B) Heat map of NRot vs PBF for the subset of fragments in the small-sized libraries. (C) GTM Landscape of fragments with PBF < 0.1. (D) GTM Landscape of fragments with PBF \geq 0.1. (E) GTM Landscape of 3D libraries. (F) GTM Landscape of SP3 libraries. The likelihood represents the probability distribution of data in the chemical space. The annotations represent specific regions to help the reader locating them during the discussion.

The mean PBF values of DIV and CHE libraries are smaller than the global average. Actually, the percentage of flat fragments in both libraries is 40.8 % and 31.7 % respectively.

To further study of three-dimensionality of fragments, we observed the GTM landscape of flat fragments (Figure III.2.2-5C) and of non-planar fragments (Figure III.2.2-5D). Interestingly, flat fragments are evenly distributed in nearly the entire fragment space. The same observation is made for non-planar fragments, suggesting that flat fragment can have non-planar counterpart. Nevertheless, flat fragments show a smaller density in the south-east part (R9), corresponding to the trifluoromethyl derivatives, and in the main center of the map, where the simplest non-aromatic compounds are located (R10). The non-planar molecules also appear to be absent from regions rich in amide groups, such as the center-north area (R11).

The 3D and SP3 libraries, which both are designed accounting for three-dimensionality, occupy nevertheless different spaces. The 3D libraries tend to reproduce the picture of the entire dataset (Figure III.2.2-5E) while the fragment of the SP3 libraries mainly populate in the east region of the landscape (R1), notably the north where the unsaturated molecules are

clustered (Figure III.2.2-5F). Whether this distribution may be biased by the molecular descriptors, it is consistent with the recent study lead by Downes which states that the three dimensionality of a molecule cannot be well described by sp3 descriptors.¹²³

III.2.2.4 Conclusions

We have proposed here a methodology for analyzing fragment libraries based on the classification of scaffolds and the modeling of landscapes by GTM. Combined, these two approaches have allowed the comparison of chemical libraries with one another as well as the positioning of each of them with respect to the set of fragments currently available commercially. We have validated that a small-sized chemical library (< 2 000 fragments) can adequately represent the chemical diversity of small-sized molecules (MW < 300 Da). We also observed unexpected characteristics of specialized libraries, such as the low proportion of non-flat fragments in diverse libraries. Lastly, our results show that libraries are not interchangeable, even if they were built for the same purpose (e.g., three-dimensionality).

Author Contributions

Conceptualization: G.M. and E.K. Methodology: J.R.I., C.J., G.B., G.M. and E.K. Implementation of the protocol: J.R.I. Data preparation: J.R.I. Formal analysis: J.R.I., G.M. and E.K. Writing—original draft preparation: J.R.I. and E.K. Writing—review and editing: J.R.I., C.J., G.M. and E.K.; project coordination: E.K.

Conflicts of interest

There are no conflicts to declare.

Funding Sources

Programme d'Investissements d'Avenir de l'Agence Nationale de Recherche (grant IDEX W19RHUS3, Université de Strasbourg).

Acknowledgements

The authors thank Fanny Bonachera and Dragos Horvart for technical assistance and helpful discussion, and Xuechen Tang for preliminary studies. The authors thank the Institut du Medicament (IMS) for financial support.

Abbreviations

3D, 3D-shaped; 3D-PSA, Polar surface area; CHE, Metal chelating; COV, Covalent; DIV, Diverse; F-NMR, Fluorine nuclear magnetic resonance; FBDD, Fragment-Based Drug Discovery; FDA, Food and drug administration; HBA, Hydrogen bond acceptor; HBD, Hydrogen bond donor; HTS, high throughput screening; FLU, Fluorine-rich; GEN, General; GTM, Generative topographic map; HAL, Halogen-rich; logP, Partition logarithm; MIS, Miscellaneous; MW, Molecular weight; NMR, Nuclear magnetic resonance; NP, Nature-product like; Nrot, Number of rotatable bonds; PBF, Plane of best fit; PIM, Principal moments of inertia; PPI, protein-protein interactions; PSA, Polar surface area; RO3, Rule of three; RO5, Rule of five; SASA, Solvent-accessible surface area; SD, Standard deviation; SP3, sp³ hybridization; QED, Quantitative estimate drug-likeness.

III.2.3 Analysis of the Diamond Light Source fragment database

III.2.3.1 Introduction

The Diamond Light Source (DLS) fragment library has been designed for hit identification as well as optimization, with a rapid follow-up synthesis. It is composed of poised fragments; i.e. small molecules with a molecular weight (MW) below 300 that contain at least one chemical group allowing fragment growing with simple synthetic reaction, such as a reaction between an amine and an acid chloride to form an amide bond.¹³⁰

Hence, the aim of this section is to evaluate the DLS fragment library composition and diversity in order to determine its suitability for a fragment-based strategy.

III.2.3.2 Material and methods

The DLS fragments were provided by our collaborators in SMILES format. Pipeline Pilot (v.22.1.0.2935, BIOVIA, Dassault Systèmes, Pipeline Pilot) was used to convert the SMILES in SDF format. Standardization of fragments, calculation of molecular properties and GTM generation protocols were described in the previous section (on pages 116 – 117).

III.2.3.3 Results

The DLS fragment library used to screen the β -clamp contained 777 fragments. The maximal MW was 260, in agreement with the Rule of Three (Ro3) guidelines,¹⁰⁴ while the minimal MW of 175, recommended for fragment growing,¹¹¹ was respected in 84% of the set. In terms of lipophilicity, the DLS fragment library followed a tendency similar to that of commercial fragment libraries. Approximately, 41% of the fragments had an AlogP ≤ 1 , demonstrating to be weakly hydrophobic, and only 25% of them presented an AlogP ≥ 2 . Furthermore, 7% of these fragments violated one of the Ro3 thresholds, with an AlogP > 3 . Globally, 83% of the fragments did not break any of the Ro3 (MW ≤ 300 Da, AlogP ≤ 3 , hydrogen bond acceptor (HBA) ≤ 3 , hydrogen bond donor (HBD) ≤ 3).

The projection of the DLS fragment set in Figure III.2.3-1A, presented a homogeneously distributed chemical space. All the clusters present in the full set of commercially available fragments (see Figure III.2.3-1B), were populated by the DLS. Nevertheless, some differences could still be found. For instance, the area between regions **A2** and **A3**, representing the

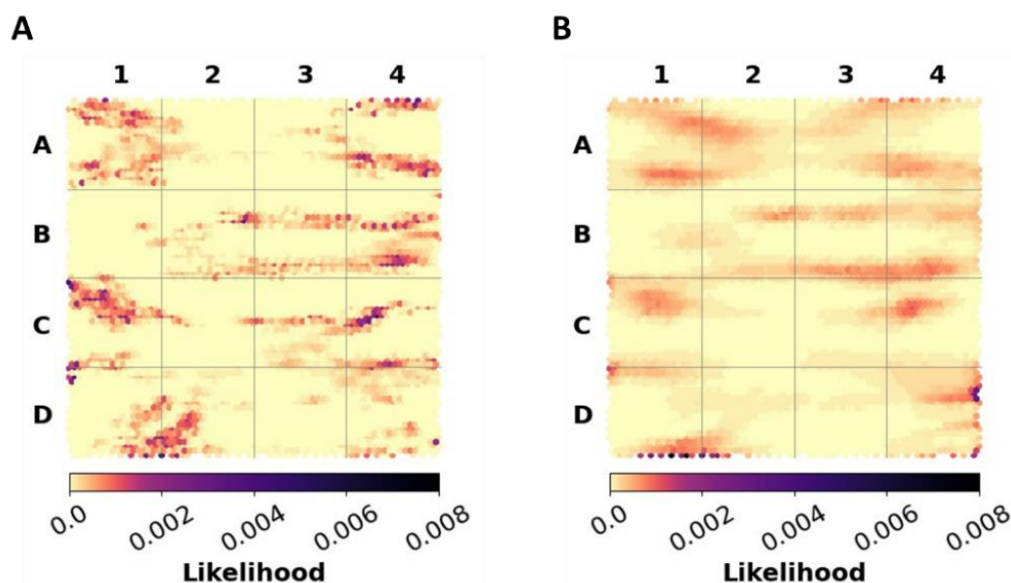


Figure III.2.3-1 Spatial distribution of the fragments in GTM Landscapes (A) DLS fragment library (B) all commercially available fragments.

molecules with an aromatic ring bound to a non-planar feature (see **R11** in Comprehensive analysis of commercial fragment libraries), only contained one fragment in the DLS library. In contrast, the region **C3**, grouping the molecules containing two hetero-substituted aromatic rings bound by a chain with at least one heteroatom, showed a higher representation among the DLS fragments. In fact, among the DLS fragments, the predominant group present in the connecting chain were amino groups, aligning with the nature of poised fragments.

III.2.3.4 Conclusion

Briefly, most DLS fragments respected fragment guidelines provided by the Ro3.¹⁰⁴ Moreover, a high percentage of the fragments have a MW >175, procuring a solid base for fragment-based approaches.¹¹¹ The fragment set consistently covered the entire chemical space, positioning itself as a diverse library, and did not introduce any chemical bias.

Therefore, the DLS library was considered a well-suited fragment library for following a fragment-based strategy to discover small molecules inhibitors of the β -clamp. In fact, employing a diverse library like DLS in the screening process is favorable for accurately mapping the protein, validating the active site, identifying the hot spots, and detecting the interactions between the fragments and the binding site.

III.2.4 Selection of DLS fragment references

III.2.4.1 Introduction

The selection of a suitable fragment set for a drug discovery campaign is crucial for its success. Therefore the aim of this section is to study the binding mode of the crystallographic fragments to select those that will be used as references during the VS.

The selection of the referential fragments was based on three conditions:

- The fragment needed to be bound to the β -clamp active site (CBM).
- Fragments forming polar interactions with the active site were privileged.
- Fragments with a high buried surface area were privileged.

III.2.4.2 Materials and methods

The crystallographic structures of β -clamp /DLS fragments were provided by Burnouf's team in PDB format. A visual analysis of the structures was done using MOE (2020.09, Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7).¹³¹ Only fragments bound to the active site were kept.

Fragments protonation was verified and corrected by Protoss (v4.0, ZBH, University of Hamburg, Germany),^{48,132} which analyzes the hydrogen bonds, the metal interactions, and the repulsive atom contacts of all states to calculate the optimal hydrogen bonding network.

The interaction fingerprints (IFP) and the buried surface area (BSA) of the fragments were calculated by IChem (v.5.9.1, University of Strasbourg, France) with the default settings.^{50,133}

III.2.4.3 Results

The screening of the Diamond Light Source fragment database against the *E.coli* β -clamp successfully identified 94 crystallographic fragments bound to the protein (see Figure III.2.4-1A). A total of 75 fragments bound the sub-site 1, with 57 of them being unique. In contrast, none of them bound the polar sub-site 2. The other 19 fragments binding different regions of the protein were excluded from the analysis. A specific fragment located at the edge of the hydrophobic sub-site 1, at the opposite side of sub-site 2, was present in both chains of the homodimer. This fragment was also excluded from the analysis (see Figure III.2.4-1B).

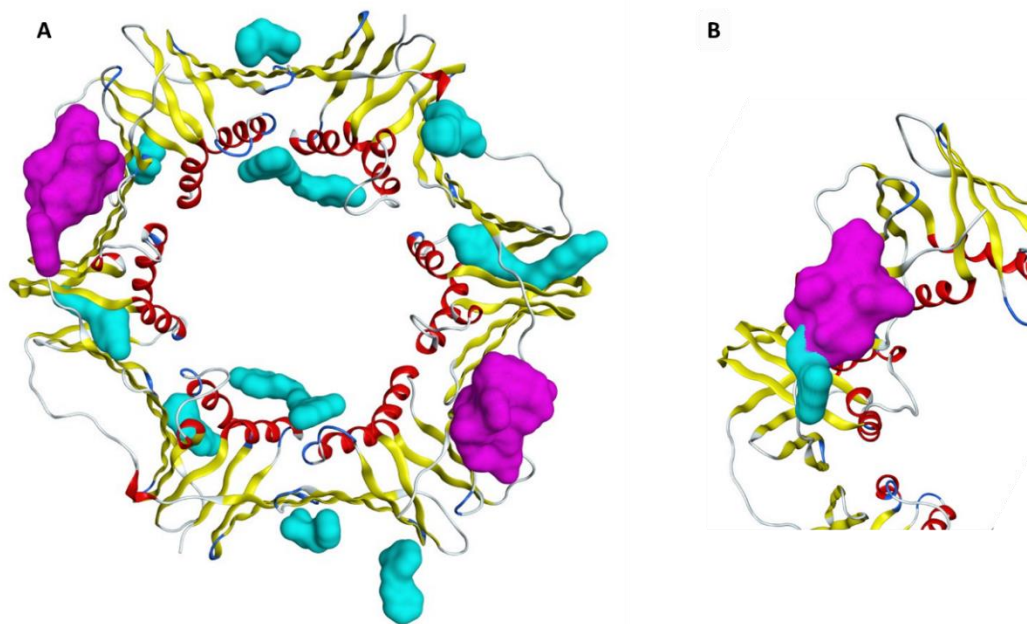


Figure III.2.4-1 3D view of the crystallographic fragments surfaces bound to the *E.coli* B-clamp dimer

(A) Global view. In fuchsia the surface of the fragments bound to the CBM. (B) Zoom to the sub-site 1 of chain B. The surface of the outlier fragment is depicted in blue

The screening revealed the presence of 12 distinct binding spots on the protein. Nevertheless, 80% of the fragments were found to bind to the CBM sub-site 1, indicating its significance as the target site. In contrast, the remaining sites accommodated a maximum of two fragments each.

As shown in Figure III.2.4-2, hydrogen bonds were relatively infrequent, with only 33% of the fragments forming one or two bonds with the sub-site 1. No ionic bonds or aromatic stacking interactions were detected among the complexes. Nevertheless, the main interaction fingerprints were formed by hydrophobic contacts, considered for a distance cutoff of 5 Å. The number of hydrophobic contacts varied among the complexes, ranging from 0 contacts for two complexes to 8 contacts for two others. The most interacting residues (Leu155, Thr172, Leu177, Pro242, Val247, and Val360) constituted the sub-site 1 described in Figure III.1.1-3.

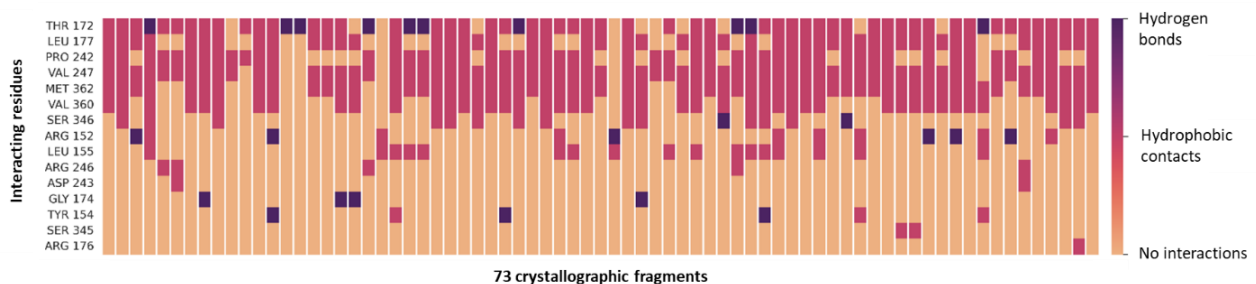


Figure III.2.4-2 Interaction fingerprint of each crystallographic fragment with the amino acids of *E. coli* β -clamp.

Furthermore, the residue Met362, which is part of sub-site 2, is located at the interface between the two the sub-sites.

The Buried Surface Area (BSA) of the 73 crystallographic fragments bound to sub-site 1 ranged from 34% to 66%. Among these fragments, 11 had a BSA of $\leq 40\%$, indicating they were partially buried. Of the remaining 62 well-buried fragments, 41 did not form any polar contacts, and 22 of them had a BSA smaller than 50%.

The 73 crystallographic fragments were filtered based on their interaction fingerprint (IFP) and buried surface area (BSA) following the workflow depicted in Figure III.2.4-3:

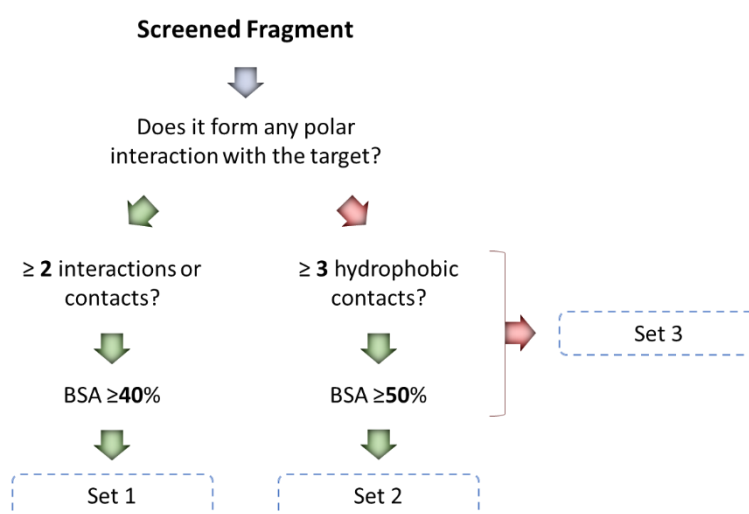


Figure III.2.4-3 Workflow employed for the selection of the query fragments. The green arrows correspond to an affirmative answer, the red arrows to a negative answer.

Overall, the filtering led to 38 crystallographic fragments (27 unique), being selected as references. Among the referential fragments 19 belonged to set 1, involving the fragment forming at least one polar interaction, and 19 to set 2, including the fragments forming only hydrophobic contacts. A total of 35 fragments were discarded from the reference sets and kept in the set 3 (see Figure III.2.4-3).

III.2.5 Conclusion

Briefly, 38 crystallographic fragments, bound to β -clamp sub-site 1, were selected as references for the VS. The selection process aimed to prioritize complexes forming polar interactions and fragments that were well buried into the CBM. Overall, the binding modes predominantly involved hydrophobic contacts, aligning with the hydrophobic nature of the sub-site 1.

III.3 Virtual screening for the identification of small molecules inhibitors of *E. coli* β -clamp

III.3.1 Introduction

Numerous crystallographic CBM binders have been characterized over the years, mainly involving peptides. Dr. Burnouf's team determined nine β -clamp structures in complex with small peptides coming from different bacterial origins: *E.coli* (PDB ID: 1OK7, 3Q4K, 3Q4J, 3Q4L, and 6FVL); *B.subtilis* (PDB ID: 4TR6); *M.tuberculosis* (PDB ID: 4TR7); and *P.aeruginosa* (PDB ID: 4TR8 and 4TSZ) proving that peptides with high affinity for one bacterial strain did not necessarily served as ligands for another bacterial strain.^{85,89}

The aim of this work was to identify small molecules able to inhibit the *E. coli* β -clamp. Therefore, the five peptides and 73 fragments binding the *E. coli* CBM were chosen as the starting point. Noteworthy, the PDB entry 1OK7 was used as reference for this VS because it was in complex with a native peptide, serving as guide for mimicking its binding mode. Furthermore, this entry has been the reference used for the design of numerous peptides.⁸⁹

Hence, two approaches were possible:

- Identification of small molecules mimicking the peptides and fragments binding mode through a virtual screening of drug-like molecules.
- *De novo* design of small molecules by growing the crystallographic fragments.

This work presents the first approach.

III.3.2 Materials and Methods

III.3.2.1 Preparation of the database molecules for the virtual screening

The databases Diverse REAL drug-like from Enamine¹³⁴ (accessed the 23/11/21, 30.7 million molecules), Screening collection from Enamine¹³⁵ (accessed the 11/10/21, 2.9 million molecules) and Chimiothèque nationale (accessed the 11/10/21, 81 130 molecules) were downloaded and saved in SMILES and SDF formats.¹³⁶

Pipeline Pilot (v.22.1.0.2935, BIOVIA, Dassault Systèmes, Pipeline Pilot) was used to run the “Bioinfo protocol”, created in 2005 to clean up, prepare and select drug-like candidates.¹³⁷ The protocol creates a unique identifier for all molecules, standardizes the charges and the stereo of the molecules, removes the counterions, discards the compounds containing more than 2 undefined stereo atoms, and merges the duplicates conserving the original data, such as the suppliers. Later, Filter (v.2.5.1.4; Openeye Scientific Software, Santa Fe) was used to obtain an adequate ionization state at physiological pH and remove the non-druggable molecules by applying a set of 220 rules (minimum and maximum values of several descriptors). Finally, the pyridone, pyridine or sulfonamide were standardized, and the 2D descriptors of the compounds were calculated (MW, AlogP, PSA, HBA, HBD, logS, NRot, NRings, Molecular Formula) with Pipeline Pilot (v.22.1.0.2935, BIOVIA, Dassault Systèmes, Pipeline Pilot).

Corina¹³⁸ (v.3.40; Molecular Networks GmbH, Nürnberg, Germany) was used to generate a low energy 3D conformer for each molecule. OMEGA¹³⁹ Rocs (v.4.2.1.1: OpenEye Scientific Software, Santa Fe, NM, USA) was used to generate a maximum of 50 conformers per molecule (default settings). Each conformer was saved in SDF format.

III.3.2.2 Preparation of the β -ring protein structure and peptides

The PDB entry 1OK7 was downloaded from the Protein Data Bank⁴⁵ and prepared using MOE (2020.09, Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7).¹³¹ The chain A, its ligand, and water molecules were deleted from the structure. The option “Structure preparation” was used to correct the errors of the structure and “Protonate3D” (default settings), such as setting adequate ionization states, position of the hydrogen atoms, and atom types. The option “Geometry” was used to detect bond lengths, bond angles, and dihedral angles outliers and atom clashes. Either delta (HID) or epsilon (HIE) positions of the His175 were protonated, thus two different protein files were saved.

Peptides in PDB entries were extracted for the PDB files, aligned, and superposed by MOE “Protein Align/Superpose” tool (2020.09, Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7)¹³¹ using the α -carbons of PDB 1OK7 chain B as reference. Protoss (v4.0, ZBH, University of Hamburg, Germany)⁴⁸ was used to set an adequate protonation of the proteins and peptides (default settings). Finally, the peptides were extracted and saved separately in PDB files.

III.3.2.3 Virtual screening

a Shaped-matching filtering

The shaped-matching filtering was performed using the software ROCS (v.4.2.1.1: OpenEye Scientific Software, Santa Fe, NM, USA)¹⁴⁰ with its default settings and the option “stats” set to “best”. The results were ranked by the RefTverskyCombo score. A maximum of 50 conformers were generated per fragment using OMEGA Rocs¹³⁹ (v.4.2.1.1: OpenEye Scientific Software, Santa Fe, NM, USA) (default settings).

b Docking

The docking calculations were performed with the ant colony optimization (ACO) algorithm PLANTS (v1.2).¹⁴¹ “ChemPLP” was used as the scoring function and “speed1” as the speed option for the screen mode. A maximum of 10 docking poses were issued, with the RMSD cluster set at 2 Å. A second docking was performed using GOLD Suite (v5.3, CCDC, Cambridge, United Kingdom), with “ChemPLP” used as scoring function and a maximum of 10 docking poses.⁶¹

The center of mass of the peptide from the PDB 1OK7 was calculated with PYMOL (v2.3.5, Schrodinger, LLC) and defined the binding site center.

c Rescoring

The referential crystallographic fragments and peptides were aligned on the chain B of the PDB entry 1OK7 using MOE “Align/Superpose” tool (2020.09, Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7).¹³¹

IChem “IFP” (v5.2.8, University of Strasbourg, France)^{49,50} was used to generate the interaction fingerprints (IFP) and calculate the Tanimoto coefficient (Tc) between the docking pose and the reference. The option “--polar” could be used to generate the IFP of hydrogen and ionic bonds.

The docking poses were rescored using LID.¹⁴² The option “lid” was used to generate an ensemble of pseudo-atoms representing the interactions and a grid where all the interactions were merged. The option “rescoring” calculated a similarity score based on the interactions of the docking poses.

d Filtering

The molecules were neutralized, standardized, and their ADMET Risk was calculated by the software ADMET Predictor[®] (v.10.0; Simulation Plus Inc., Lancaster, CA, USA).

The molecules were individually visualized by using MOE (2020.09, Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7).¹³¹

e Analysis of the selected molecules

The Tanimoto similarity score between the molecules was calculated using Pipeline Pilot (v.22.1.0.2935, BIOVIA, Dassault Systèmes, Pipeline Pilot) based on Extended Connectivity Fingerprints (ECFP_4). The same comparison was performed between the molecules and a list of existing antibiotics downloaded from the DrugBank website.¹⁴³

III.3.3 Results

III.3.3.1 Protocol

The virtual screening of drug-like molecules against the *E. coli* β -clamp CBM followed the forthcoming protocol (see Figure III.3.3-1):

- I) Download of the libraries “diverse REAL drug-like” from Enamine; “screening collection” from Enamine, and Chimiothèque Nationale. Clean up and selection of drug-like compounds. The step I resulted in a starting point with 20.5 million molecules.
- II) Filtering of the drug-like compounds based on a shape and color similarity with the chosen crystallographic fragments from sets 1 and 2. Polar interactions were privileged.

As a reminder, the crystallographic fragments are distributed into three sets (see Selection of DLS fragment references and Figure III.2.4-3):

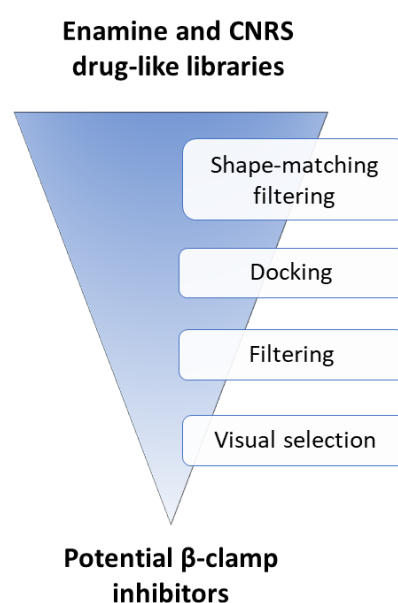


Figure III.3.3-1 VS protocol.

- Set 1: 19 fragments forming at least 1 polar interaction with the target.
 - Set 2: 19 fragments forming only hydrophobic contacts with the target.
 - Set 3: 35 fragments that did not fulfill the interactions and BSA conditions.
- III)** Docking of the drug-like compounds into the β -clamp under four different conditions:
- a. His175 HD1 protonation and active site radius equal to 20 Å
 - b. His175 HE2 protonation and active site radius equal to 20 Å
 - c. His175 HD1 protonation and active site radius equal to 8 Å
 - d. His175 HE2 protonation and active site radius equal to 8 Å.
- IV)** Rescoring of docking poses based on the similarity of the interactions fingerprint with that of the five crystallographic peptides and fragments from the three sets (see Selection of DLS fragment references).
- V)** Filtering of the drug-like compounds with a high ADMET risk score, which was calculated to give an overall score based on the predicted low absorption risk, the high CYP metabolism risk, toxicity risk, high plasma protein binding risk, and high steady-state volume of distribution risk.
- VI)** Final drug-like molecule visual selection of the potential candidates.

III.3.3.2 Shape-similarity filtering

The molecules issued from the databases Enamine REAL drug-like, Enamine screening collection and the chimiothèque nationale, accounting for more than 34.4 million molecules, underwent a first filter to select the drug-like candidates, and a second one to obtain a subset of molecules similar to the shapes and pharmacophoric features of the queries from sets 1 and 2. To do so, the method used was based on ROCS procedure, or shape-similarity filtering.

ROCS¹⁴⁰ is a rigid ligand-centered approach conceived for comparing the shape of superimposed ligands. The ligands are aligned by a Gaussian-based volume overlap maximization, scored, and ranked by similarity towards a query molecule. The score is based

on the shape and the pharmacophoric features (hydrogen donor, hydrogen acceptor, anion, cation, hydrophobic contacts, rings) similarity between the query and the molecule of the database. The similarity was measured by a Tversky coefficient (*RefTverskyCombo*), since the query is composed of fragments and the database of drug-like molecules.¹⁴⁴

The method was tested by comparing each fragment of the reference set (set 1 +2) with all the fragments of the reference set, including itself. The query fragment was represented by its crystallographic structure as found in complex with the β -clamp, while the compared fragments were represented by 50 conformers. The goal of this retrospective exercise was to establish a threshold for the library filtering of step II.

Overall, ROCS successfully assigned the higher score to fragment conformers, with scores ranging from 1.37 to 1.99 (Annexes VI.2). Some fragments exhibited common pharmacophores, leading to high scores for other fragments (see Figure III.3.3-2A), while others likely possessed unique features as suggested by the large score gap between the self-comparison and the other fragments (see pink and blue dots on Figure III.3.3-2B, respectively). However, there was one exception observed for query 24, where ROCS assigned the highest score to a cross-comparison and ranked second the self-comparison (see Figure III.3.3-2C).

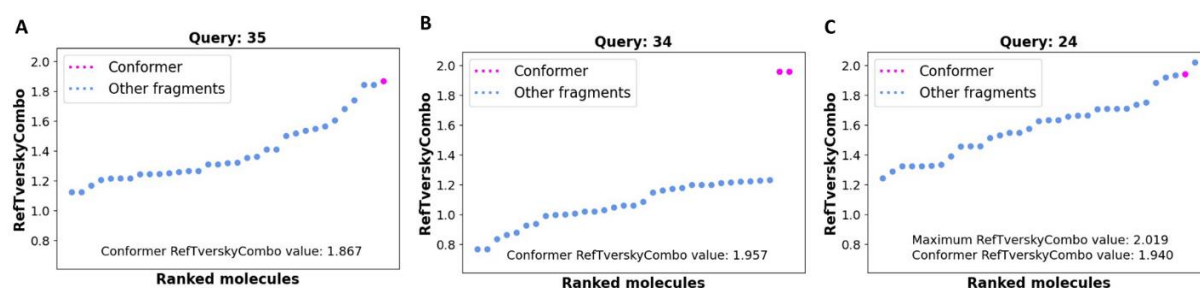


Figure III.3.3-2 Results of ROCS test

Briefly, ROCS score attribution was effective in recognizing fragments with similar shapes and pharmacophores. As a consequence of the score variation between the queries, the selection threshold for selecting the molecules similar to the fragments was defined as a percentage rather than a specific value. Furthermore, following the previously mentioned strategy of prioritizing the formation of polar interactions, we favored the molecules selected by query fragments belonging to set 1:

- Top 1% if compared to a query belonging to set 1.

- Top 0.1% if compared to a query belonging to set 2.

Upon duplicate filtration, the filtering step resulted in the reduction of the initial collection of 20.5 million molecules to 2.9 million molecules.

III.3.3.3 Analysis of the PDB structure

The *E. coli* β -clamp model used for the virtual screening was the chain B of the PDB entry 1OK7. The file was deposited in 2003 and had a resolution of 1.65 Å. Within the file, the peptide P16, a 16-amino acid segment of DNA polymerase IV, was bound to chain B.

Chain B presented four bond angle anomalies involving glycine residues; but, none of these anomalies were in close proximity to the binding site. Additionally, five clashes were observed, with Asp243 being part of the binding site on the same side as P16 Leu16.

The referential chain B of 1OK7 had a sequence identity of 100% and an α -carbon RMSD of 0.54 Å when compared to the reference structure used by Dr. Burnouf's team to construct the models with bound fragments. The CBM displayed an RMSD of 0.46 Å and was analyzed in the absence of a bound ligand. The loop spanning Ala147 to Tyr153, located near the binding site, showed the greatest difference with an RMSD of 1.05 Å, but did not interact with the peptide.

III.3.3.4 Virtual Screening

As mentioned, the aim of this virtual screening was to identify molecules able to bind the CBM. Currently, peptides have shown the highest affinity, with the ability to bind the two sub-binding sites being the biggest difference with the discovered small molecules.

With this in mind, the initial virtual screening aimed to identify longer molecules capable of mimicking the binding mode of peptides. To do so, we established the coordinates of the active site by calculating the center of mass of the peptide bound to chain B in the PDB entry 1OK7. Then, we determined the distance between the center of mass and the furthest atom of the peptide establishing a binding site radius of 20 Å involving the two CBM sub-sites.

His175 was one of the residues constituting sub-site 2 and could be protonated at positions HID or HIE to form different hydrogen bonds (see Figure III.3.3-3A and Figure III.3.3-3B,

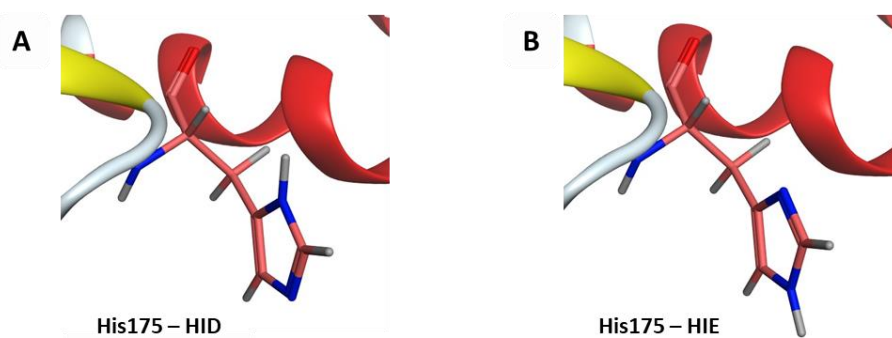


Figure III.3.3-3 His175 protonation states (A) HID protonation of His175 (B) HIE protonation of His175. (PDB ID: 1OK7)

respectively). Thus, the first virtual screening was run twice (once per Hist175 protonation state) with an active site radius of 20 Å.

a First VS

A total of 2.9 million molecules were subjected to docking using the PLANTS software and classified using ChemPLP. Then, a rescoring technique called LID (Local Interaction Density) was employed to rescore the docking poses by comparing the interactions of all known binders, including five peptides binding to *E. coli* β -clamp (PDB ID: 1OK7, 3Q4K, 3Q4J, 3Q4L, and 6FVL) and all 73 crystallographic fragments (sets 1, 2, and 3), with the interactions of the docking pose (see Figure III.3.3-4). LID threshold of 0.025 was established by calculating the mean LID value of the top 1000 molecules ranked by ChemPLP score.

The top 1000 molecules ranked by ChemPLP score for each VS and having a LID score higher than 0.025 (His175 HID and His175 HIE) were retained.

To validate the selection, the molecules were redocked using the GOLD software. The redocking process was performed with a binding site radius of 8 Å, ensuring that the docking was focused on the CBM. Among these, only the docking poses with at least one common polar interaction with the five peptides (polar IFP $T_c > 0$) and an ADMET Risk below 5 were kept. This filtering process resulted in 1990 conformers for His175 HID protonation and 1977 for His175 HIE. Finally, for each molecule, only the docking pose displaying the highest non-polar IFP Tanimoto coefficient when compared to the five peptides was retained.

The filtering process yielded 240 unique molecules for the His175 HID protonation state and 246 for the HIE protonation state. When combining the results, only two molecules were found to be in common, resulting in a total of 484 unique molecules.

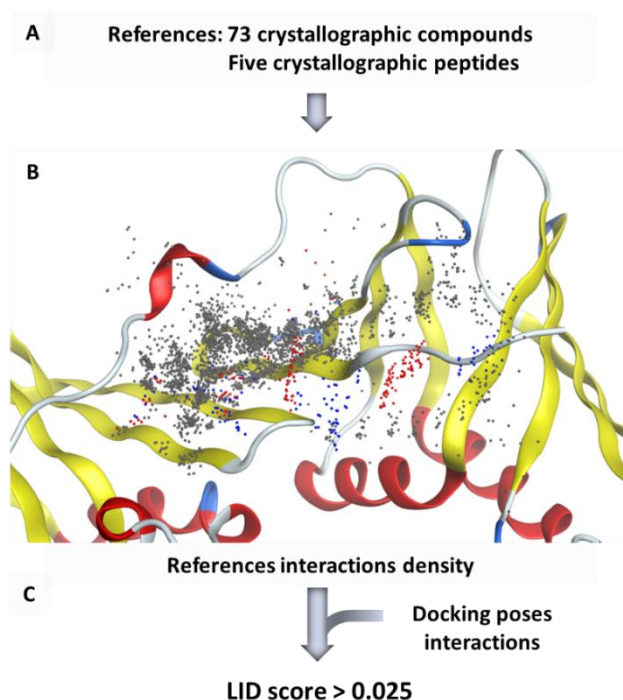


Figure III.3.3-4 LID process. (A) Feed of the references, (B) generation of the interactions, (C) and LID comparison of the docking poses interactions.

b Second VS

A second VS was carried out to extend the set of selection. A radius of 8\AA was chosen, since we noted during the first VS that the 20\AA radius allowed docking pose out of the CBM.

In this second VS, all molecules issued by step **II** were again docked using the PLANTS. However, in this iteration, the thresholds established during the first virtual screening were retained, allowing the evaluation and retention of the optimal docking poses immediately after the docking process. The conditions used to determine which docking poses to keep were the following:

- ChemPLP score < -75 , which was the lowest score among the 484 molecules filtered during the first virtual screening.
- LID score > 0.025 . The threshold value was established by calculating the mean LID value of the top 1000 molecules ranked by ChemPLP score in the first VS.
- Polar IFP $T_c > 0$.

As a result of the conditions set for the VS performed on His175 HID and His175 HIE, a total of 422 and 465 non-unique molecules were retained, respectively. For each molecule, the

docking pose with the highest IFP Tc when compared to the five peptides was selected. The maximal ADMET Risk threshold was maintained at 5.

It is worth noting that none of the molecules retained in this second virtual screening had been obtained from the first virtual screening. After applying these criteria, the workflow yielded to a total of 98 unique molecules.

III.3.3.5 Visual selection and evaluation of potential hits

The visual analysis discarded molecules with unusual torsion angles, while favoring molecules showing good complementarity with the protein, and also prioritizing originality and diversity in the selection.

From the first virtual screening, a total of 68 molecules were selected and ordered, with 43 molecules selected from the Enamine screening collection and 25 molecules selected from Enamine's Diverse REAL drug-like database. In the second virtual screening, 52 molecules were selected, consisting of 5 molecules from the Enamine screening collection and 47 molecules from Enamine's Diverse REAL drug-like database.

The 120 molecules that were selected had different molecular weights (MW), ranging from 295 to 500. Approximately 52% of these molecules had similar molecular weights, falling within the range of 375 to 400. The logP values showed a wide range from -1.6 to 6.0 indicating

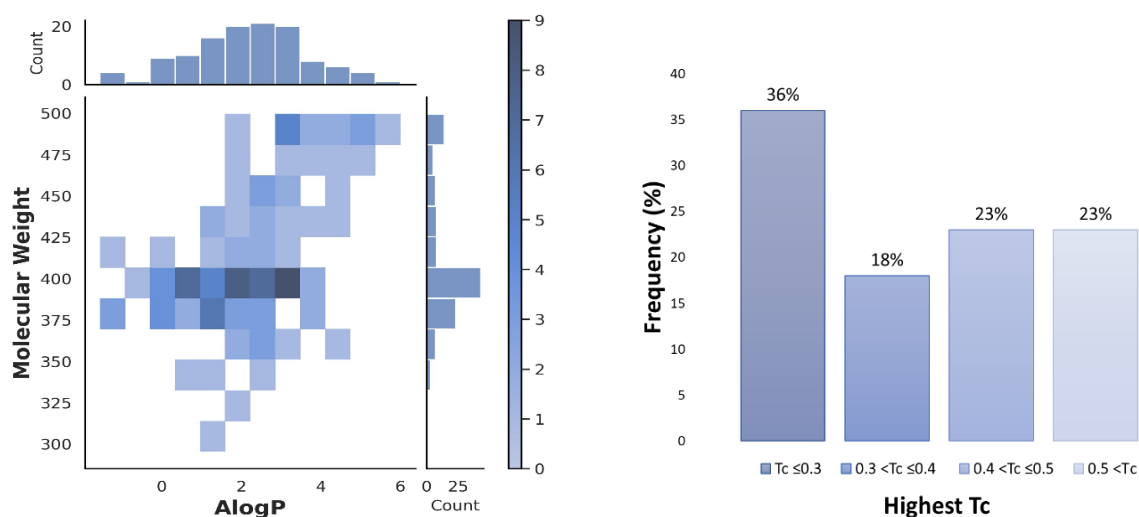


Figure III.3.3-5 Properties of the 120 selected molecules. (A) MW vs. AlogP (B) Frequency of the highest Tc between the selected molecules

a diverse set of molecules with both hydrophilic and lipophilic characteristics. Out of the 120 molecules, 9 compounds were hydrophilic (AlogP below 0), and 82 fragments were classified as weakly lipophilic (AlogP values between 0 and 3). Only 2 fragments exceeded the Lipinski limit, with logP values higher than 5 (see Figure III.3.3-5A).

When comparing the 120 molecules to each other, the highest Tanimoto Coefficient (Tc), calculated using ECFP₄ molecular fingerprints, ranged from 0.75 to 0.14. A total of 91 molecules had Tc < 0.5, indicating that most of the molecules were structurally different from each other (see Figure III.3.3-5B). Two-thirds of the molecules with a Tc > 0.5 showed a Tc lower than 0.5 when evaluating its second highest similarity.

Furthermore, when comparing the selected molecules to a list of 287 antibiotics in the DrugBank database, the Tc similarity comparison revealed that the molecules were different from all the antibiotics listed, with the highest Tc value obtained in this comparison equal to 0.26.

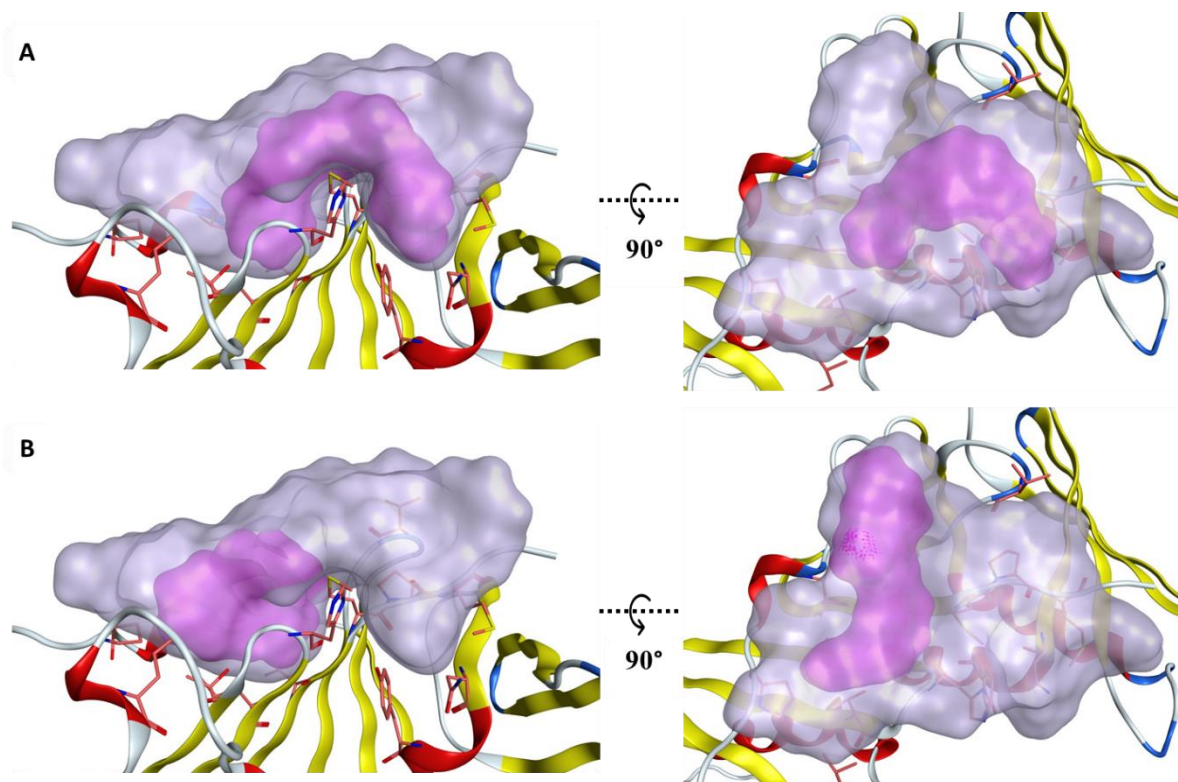


Figure III.3.3-6 Molecular surfaces of the selected molecules bound to the CBM. The purple surface depicted the area covered by the entire set of molecules. The pink surface depicted the area covered by one molecule. (A) Molecule binding the two sub-sites. (B) Molecule binding the sub-site 1.

The selected molecules taken as a whole covered an area which included both sub-sites 1 and 2 of the CBM, and extended beyond (Figure III.3.3-6). Considered individually, the molecules occupied the CBM in different ways. For example, the individual molecule highlighted on Figure III.3.3-6A spanned the two sub-sites, while the molecule shown on Figure III.3.3-6B occupied only the sub-site 1.

Out of the 120 molecules selected, 108 of them were synthesized by the external company Enamine. Currently, these 108 molecules are undergoing testing using the minimum inhibitory concentration (MIC) assay on *E. coli* bacteria. The purpose of this testing is to identify any initial signs of activity or potential antibacterial effects exhibited by the molecules. To date, we still do not have the biological results informing about the biological interest of the selected molecules.

III.3.4 Discussion

The β -clamp is a protein that plays a crucial role in bacterial DNA replication, repair, recombination, and mutagenesis, and that structurally differs from its eukaryotic equivalent, PCNA.⁸⁰ Several peptide studies have proven the potential of the target for the development of new antibacterial treatments. Nevertheless, despite the structural similarities existing between the CBM of Gram+ and Gram- bacteria, the peptides demonstrated a reduced affinity towards Gram+ bacteria due to a groove located between sub-site 1 and sub-site 2.⁸⁵

A VS approach was employed to target the CBM, with the aim of mimicking the binding mode of peptides while favoring the binding to sub-site 1.

To achieve this objective, several decisions were made during the screening process:

- An initial filtering step to favor molecules having pharmacophoric similarities with the crystallographic fragments forming polar interactions. This filtering aimed to select molecules capable of binding sub-site 1 and to prioritize the formation of polar bonds. Although sub-site 1 is primarily hydrophobic, the inclusion of polar interactions would increase the molecules affinity for the target.
- The screening involved two docking steps. Initially, a docking with a larger radius was performed to allow the molecules to potentially bind to both sub-sites. Then, the best-

ranked molecules were docked using another software with a smaller radius to validate their binding.

- In the second screening, a radius of 8 Å was used from the beginning to enable the selection of smaller molecules that do not need to bridge between the two sub-sites.
- A rescoring process was conducted based on the density of interactions observed in all known crystallographic peptides and fragments. Crystallographic fragments from all three sets were employed to increase the density of sub-site 1 interactions and further promote the binding to this region.
- A final filtering step was implemented based on the similarity of interactions with peptides, to continue the search for peptidomimetics.

Due to the structural configuration of the CBM, the final selected molecules were found to be relatively long. However, they still comply with the rule of five guideline for molecular weight, which is an important criterion for drug-likeness and oral bioavailability. Additionally, the final sets of molecules underwent a filtering based on ADMET risk assessment to mitigate potential toxicity if any of the molecules were found to be active against the target.

Finally, while our work on the β -clamp provided a list of 120 potential hits targeting *E. coli* CBM, it is important to acknowledge some limitations and potential drawbacks in our approach. One notable limitation is that we did not consider the presence of water molecules in our simulations, which could significantly influence the binding interactions and overall behavior of the β -clamp. Furthermore, our study did not consider the dynamic nature of the protein. Despite the rigidity conferred to the protein and binding site through the β -sheets, neglecting the mobility of the side chains, such as Met362, could importantly influence the binding of the ligands. Additionally, while we focused on selecting molecules mimicking the peptides, we did not consider the entire spectrum of reference peptide structures. By limiting our analysis to a subset of reference peptides, we might have missed out on potential ligands.

III.3.5 Conclusion and perspectives

This work aimed at selecting drug-like molecules able to target the CBM. To do so, we implemented several strategies to guide the search for compounds that can replicate the binding of peptides to the β -clamp, with a specific focus on promoting binding to sub-site 1.

First, we filtered the Enamine and CN compound libraries, which consisted of a total of 34.4 million molecules. The filtering process aimed to retain only the molecules that exhibited partial alignment with the shape and pharmacophores of well-buried and anchored crystallographic fragments. After this step, 2.9 million molecules were selected and subjected to docking within the targeted site. The docking poses were prioritized based on their binding modes, with a preference for those resembling the binding modes of reference inhibitory peptides and crystallographic fragments.

A visual examination of the virtual hits led to the selection of 120 molecules, out of which 108 were synthesized by Enamine. These synthesized molecules are currently being evaluated by our collaborators to assess their biological activity, specifically their binding to the bacterial clamp of *E. coli* and their inhibition of bacterial growth.

If the hits are validated, the project will proceed towards their optimization using medicinal chemistry and aided by three-dimensional models of the protein/ligand complexes. In case of unsuccessful validation, we will reevaluate the assumptions made regarding protein/ligand recognition, considering protein dynamics and the role of solvents. This reevaluation will allow us to propose an alternative strategy based on the growth of selected fragments, chosen based on criteria such as synthetic accessibility, pharmacophoric properties, and binding to hot-spots.

IV Conclusion

The resistance of bacteria to antibiotics is a major issue for humanity. In order to overcome this problem, two research and development projects for new antibiotics have been implemented as part of this thesis: the optimization of neomycin B and the search for ligands targeting the bacterial β -clamp. These projects were carried out in collaboration with several teams from different disciplines.

For the optimization of neomycin, a study was conducted to comprehensively describe the binding modes of a specific AGA in order to identify key positions for the recognition and action of AMEs and for antibacterial activity. Then, a methodology was established to predict the action of five AMEs on synthetic derivatives of neomycin B, allowing for an understanding of the impact of the modifications made. The obtained results highlighted the impact of substitution at the C(6')-NH₂ position and suggested examining two other positions, C(1) and C(2'), as a potential targets for inhibiting the action of certain AMEs.

Now, a necessary question arises: Should we continue the search for new AGA derivatives able to overcome AMEs resistance? My opinion on this matter is divided:

Aminoglycosides, as a class of drugs, pose significant challenges due to their polarity, intrinsic toxicity, the large number of reported AMEs (over 120), and their promiscuity. Furthermore, studies have suggested that the AMEs binding site mimics the rRNA binding site, which is essential for the AGAs mechanism of action. Therefore, modifying a position to reduce affinity towards AMEs may potentially result in a loss of antibacterial activity. Moreover, researchers have already invested considerable effort in this pursuit, yet no neomycin B derivatives have been successfully developed this far. The rapid emergence of resistance to new derivatives, as demonstrated by plazomicin, is also a concern.

Nevertheless, aminoglycosides are highly effective antibacterials, and particularly interesting when referring to Gram- infections. Furthermore, given the urgent need to combat bacterial resistance, every hope to find new and effective antibiotics resistant-free remains important. In this regard, I want to think that our collaborative project is the beginning of a novel and faster approach for testing derivatives, which could provide a renewed hope to researchers in this

field. Nevertheless, many efforts, such as modelling more AMEs, predicting the binding to rRNA, or making the tool easily usable to anyone, still need to be made.

Regarding the search for ligands targeting the bacterial replication ring, a strategy for virtual screening of drug-like molecule libraries was implemented. This research relied on the binding mode of 73 fragments from the Diamond library and 5 inhibitory peptides. To explore the fragment-based approach and evaluate the Diamond library, a comprehensive analyses of the commercial fragment libraries was conducted. We selected 120 molecules and 108 of them were synthesized.

At the beginning of this thesis, there was a natural outcome haunting me: no matter what antibiotic we optimized or discovered, bacteria were going to find a way to develop resistances against them. Actually, bacteria were able to develop resistance mechanisms faster than researchers were able to discover new antibiotics. I increasingly realized the magnitude of the problem and started to look for the established political measures adopted to counter this problem. The results of this search did not relieve my concern: we are facing a silent pandemic for which only academia, NGOs such as CARB-X or GARDP, and some governments (special mention to UK's) are acting. Nowadays, drugs discovery projects in pharmaceutical companies are not only directed by a scientific committee but also by a budgetary one. Decision-makers decide on the direction of Med Chem programs mostly on economic profitability; and, sadly, antibiotics are prone to resistance and do not provide prolonged treatment. Worryingly, the last example of a pharmaceutical company bringing antibacterial to the market concluded with the bankruptcy of the company, (i.e. Achaogen). As mentioned at the beginning of the manuscript, "This might be one of the world's biggest problems, but it does not need to be its hardest."²¹. New policies need to be urgently made to bring pharmaceutical companies back to the race against bacterial resistance. Meanwhile, let's continue raising awareness about the proper use of antibiotics and this endless race against resistance.

V Bibliography

- (1) Beveridge, T. Use of the Gram Stain in Microbiology. *Biotech. Histochem.* **2001**, *76* (3), 111–118.
- (2) Silhavy, T. J.; Kahne, D.; Walker, S. The Bacterial Cell Envelope. *Cold Spring Harb. Perspect. Biol.* **2010**, *2* (5), a000414.
- (3) Hutchings, M. I.; Truman, A. W.; Wilkinson, B. Antibiotics: Past, Present and Future. *Curr. Opin. Microbiol.* **2019**, *51*, 72–80.
- (4) Gould, K. Antibiotics: From Prehistory to the Present Day. *J. Antimicrob. Chemother.* **2016**, *71* (3), 572–575.
- (5) Mohr, K. I. History of Antibiotics Research. In *How to Overcome the Antibiotic Crisis : Facts, Challenges, Technologies and Future Perspectives*; Stadler, M., Dersch, P., Eds.; Current Topics in Microbiology and Immunology; Springer International Publishing: Cham, 2016; pp 237–272.
- (6) Lewis, K. Platforms for Antibiotic Discovery. *Nat. Rev. Drug Discov.* **2013**, *12* (5), 371–387.
- (7) Mulvey, M. R.; Simor, A. E. Antimicrobial Resistance in Hospitals: How Concerned Should We Be? *CMAJ Can. Med. Assoc. J.* **2009**, *180* (4), 408–415.
- (8) Antimicrobial Resistance | Food and Agriculture Organization of the United Nations <https://www.fao.org/antimicrobial-resistance/en/> (accessed 2022 -04 -12).
- (9) Tenover, F. C. Mechanisms of Antimicrobial Resistance in Bacteria. *Am. J. Infect. Control* **2006**, *34* (5 Suppl 1), S3-10; discussion S64-73.
- (10) von Wintersdorff, C. J. H.; Penders, J.; van Niekerk, J. M.; Mills, N. D.; Majumder, S.; van Alphen, L. B.; Savelkoul, P. H. M.; Wolffs, P. F. G. Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. *Front. Microbiol.* **2016**, *7*, 173.
- (11) Blair, J. M. A.; Webber, M. A.; Baylay, A. J.; Ogbolu, D. O.; Piddock, L. J. V. Molecular Mechanisms of Antibiotic Resistance. *Nat. Rev. Microbiol.* **2015**, *13* (1), 42–51.
- (12) Murray, C. J. L.; Ikuta, K. S.; Sharara, F.; Swetschinski, L.; Aguilar, G. R.; Gray, A.; Han, C.; Bisignano, C.; Rao, P.; Wool, E.; Johnson, S. C.; Browne, A. J.; Chipeta, M. G.; Fell, F.; Hackett, S.; Haines-Woodhouse, G.; Hamadani, B. H. K.; Kumaran, E. A. P.; McManigal, B.; Achalapong, S.; Agarwal, R.; Akech, S.; Albertson, S.; Amuasi, J.; Andrews, J.; Aravkin, A.; Ashley, E.; Babin, F.-X.; Bailey, F.; Baker, S.; Basnyat, B.; Bekker, A.; Bender, R.; Berkley, J. A.; Bethou, A.; Bielicki, J.; Boonkasidecha, S.; Bukosia, J.; Carvalheiro, C.; Castañeda-Orjuela, C.; Chansamouth, V.; Chaurasia, S.; Chiurchiù, S.; Chowdhury, F.; Donatien, R. C.; Cook, A. J.; Cooper, B.; Cressey, T. R.; Criollo-Mora, E.; Cunningham, M.; Darboe, S.; Day, N. P. J.; Luca, M. D.; Dokova, K.; Dramowski, A.; Dunachie, S. J.; Bich, T. D.; Eckmanns, T.; Eibach, D.; Emami, A.; Feasey, N.; Fisher-Pearson, N.; Forrest, K.; Garcia, C.; Garrett, D.; Gastmeier, P.; Giref, A. Z.; Greer, R. C.; Gupta, V.; Haller, S.; Haselbeck, A.; Hay, S. I.; Holm, M.; Hopkins, S.; Hsia, Y.; Iregbu, K. C.; Jacobs, J.; Jarovsky, D.; Javanmardi, F.; Jenney, A. W. J.; Khorana, M.; Khusuwan, S.; Kissoon, N.; Kobeissi, E.; Kostyanov, T.; Krapp, F.; Krumkamp, R.; Kumar, A.; Kyu, H. H.; Lim, C.; Lim, K.; Limmathurotsakul, D.; Loftus, M. J.; Lunn, M.; Ma, J.; Manoharan, A.; Marks, F.; May, J.; Mayxay, M.; Mturi, N.; Munera-Huertas, T.; Musicha, P.; Musila, L. A.; Mussi-Pinhata, M. M.; Naidu, R. N.; Nakamura, T.; Nanavati, R.; Nangia, S.; Newton, P.; Ngoun, C.; Novotney, A.; Nwakanma, D.; Obiero, C. W.; Ochoa, T. J.; Olivas-

- Martinez, A.; Olliaro, P.; Ooko, E.; Ortiz-Brizuela, E.; Ounchanum, P.; Pak, G. D.; Paredes, J. L.; Peleg, A. Y.; Perrone, C.; Phe, T.; Phommasone, K.; Plakkal, N.; Ponce-de-Leon, A.; Raad, M.; Ramdin, T.; Rattanavong, S.; Riddell, A.; Roberts, T.; Robotham, J. V.; Roca, A.; Rosenthal, V. D.; Rudd, K. E.; Russell, N.; Sader, H. S.; Saengchan, W.; Schnall, J.; Scott, J. A. G.; Seekaew, S.; Sharland, M.; Shivamallappa, M.; Sifuentes-Osornio, J.; Simpson, A. J.; Steenkeste, N.; Stewardson, A. J.; Stoeva, T.; Tasak, N.; Thaiprakong, A.; Thwaites, G.; Tigoi, C.; Turner, C.; Turner, P.; Doorn, H. R. van; Velaphi, S.; Vongpradith, A.; Vongsouvath, M.; Vu, H.; Walsh, T.; Walson, J. L.; Waner, S.; Wangrangsimakul, T.; Wannapinij, P.; Wozniak, T.; Sharma, T. E. M. W. Y.; Yu, K. C.; Zheng, P.; Sartorius, B.; Lopez, A. D.; Stergachis, A.; Moore, C.; Dolecek, C.; Naghavi, M. Global Burden of Bacterial Antimicrobial Resistance in 2019: A Systematic Analysis. *The Lancet* **2022**, *399* (10325), 629–655.
- (13) O'Neill, J. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations*; Report; Government of the United Kingdom, 2016.
- (14) Göttig, S.; Gruber, T. M.; Higgins, P. G.; Wachsmuth, M.; Seifert, H.; Kempf, V. A. J. Detection of Pan Drug-Resistant *Acinetobacter Baumannii* in Germany. *J. Antimicrob. Chemother.* **2014**, *69* (9), 2578–2579.
- (15) Xu, L.; Sun, X.; Ma, X. Systematic Review and Meta-Analysis of Mortality of Patients Infected with Carbapenem-Resistant *Klebsiella Pneumoniae*. *Ann. Clin. Microbiol. Antimicrob.* **2017**, *16* (1), 18.
- (16) Lewis, K. The Science of Antibiotic Discovery. *Cell* **2020**, *181* (1), 29–45.
- (17) Nadimpalli, M. L.; Marks, S. J.; Montealegre, M. C.; Gilman, R. H.; Pajuelo, M. J.; Saito, M.; Tsukayama, P.; Njenga, S. M.; Kiiru, J.; Swarthout, J.; Islam, M. A.; Julian, T. R.; Pickering, A. J. Urban Informal Settlements as Hotspots of Antimicrobial Resistance and the Need to Curb Environmental Transmission. *Nat. Microbiol.* **2020**, *5* (6), 787–795.
- (18) World Health Organization. Global Action Plan on Antimicrobial Resistance. World Health Organization 2015.
- (19) Tacconelli, E.; Carrara, E.; Savoldi, A.; Harbarth, S.; Mendelson, M.; Monnet, D. L.; Pulcini, C.; Kahlmeter, G.; Kluytmans, J.; Carmeli, Y.; Ouellette, M.; Outterson, K.; Patel, J.; Cavaleri, M.; Cox, E. M.; Houchens, C. R.; Grayson, M. L.; Hansen, P.; Singh, N.; Theuretzbacher, U.; Magrini, N.; Aboderin, A. O.; Al-Abri, S. S.; Jalil, N. A.; Benzonana, N.; Bhattacharya, S.; Brink, A. J.; Burkert, F. R.; Cars, O.; Cornaglia, G.; Dyar, O. J.; Friedrich, A. W.; Gales, A. C.; Gandra, S.; Giske, C. G.; Goff, D. A.; Goossens, H.; Gottlieb, T.; Blanco, M. G.; Hryniewicz, W.; Kattula, D.; Jinks, T.; Kanj, S. S.; Kerr, L.; Kieny, M.-P.; Kim, Y. S.; Kozlov, R. S.; Labarca, J.; Laxminarayan, R.; Leder, K.; Leibovici, L.; Levy-Hara, G.; Littman, J.; Malhotra-Kumar, S.; Manchanda, V.; Moja, L.; Ndoye, B.; Pan, A.; Paterson, D. L.; Paul, M.; Qiu, H.; Ramon-Pardo, P.; Rodríguez-Baño, J.; Sanguinetti, M.; Sengupta, S.; Sharland, M.; Si-Mehand, M.; Silver, L. L.; Song, W.; Steinbakk, M.; Thomsen, J.; Thwaites, G. E.; Meer, J. W. van der; Kinh, N. V.; Vega, S.; Villegas, M. V.; Wechsler-Fördös, A.; Wertheim, H. F. L.; Wesangula, E.; Woodford, N.; Yilmaz, F. O.; Zorzet, A. Discovery, Research, and Development of New Antibiotics: The WHO Priority List of Antibiotic-Resistant Bacteria and Tuberculosis. *Lancet Infect. Dis.* **2018**, *18* (3), 318–327.
- (20) One Health | CDC <https://www.cdc.gov/onehealth/index.html> (accessed 2023 -06 -27).
- (21) O'Neill, J. *Tackling a Crisis for the Health and Wealth of Nations*; Report; Government of the United Kingdom, 2014.
-

-
- (22) Tevyashova, A. N.; Shapovalova, K. S. Potential for the Development of a New Generation of Aminoglycoside Antibiotics. *Pharm. Chem. J.* **2021**, *55* (9), 860–875.
- (23) Ramirez, M. S.; Tolmasky, M. E. Aminoglycoside Modifying Enzymes. *Drug Resist. Updat. Rev. Comment. Antimicrob. Anticancer Chemother.* **2010**, *13* (6), 151–171.
- (24) Obszynski, J.; Loidon, H.; Blanc, A.; Weibel, J.-M.; Pale, P. Targeted Modifications of Neomycin and Paromomycin: Towards Resistance-Free Antibiotics? *Bioorganic Chem.* **2022**, *126*, 105824.
- (25) Avent, M. L.; Rogers, B. A.; Cheng, A. C.; Paterson, D. L. Current Use of Aminoglycosides: Indications, Pharmacokinetics and Monitoring for Toxicity. *Intern. Med. J.* **2011**, *41* (6), 441–449.
- (26) Krause, K. M.; Serio, A. W.; Kane, T. R.; Connolly, L. E. Aminoglycosides: An Overview. *Cold Spring Harb. Perspect. Med.* **2016**, *6* (6), a027029.
- (27) Davies, J.; Gorini, L.; Davis, B. D. Misreading of RNA Codewords Induced by Aminoglycoside Antibiotics. *Mol. Pharmacol.* **1965**, *1* (1), 93–106.
- (28) Moazed, D.; Noller, H. F. Interaction of Antibiotics with Functional Sites in 16S Ribosomal RNA. *Nature* **1987**, *327* (6121), 389–394.
- (29) Houghton, J. L.; Green, K. D.; Chen, W.; Garneau-Tsodikova, S. The Future of Aminoglycosides: The End or Renaissance? *ChemBioChem* **2010**, *11* (7), 880–902.
- (30) Uis S. Gonzalez, I.; Spencer, J. P. Aminoglycosides: A Practical Review. *Am. Fam. Physician* **1998**, *58* (8), 1811–1820.
- (31) Mingeot-Leclercq, M.-P.; Tulkens, P. M. Aminoglycosides: Nephrotoxicity. *Antimicrob. Agents Chemother.* **1999**, *43* (5), 1003–1012.
- (32) Hutchin, T.; Haworth, I.; Higashi, K.; Fischel-Ghodsian, N.; Stoneking, M.; Saha, N.; Arnos, C.; Cortopassi, G. A Molecular Basis for Human Hypersensitivity to Aminoglycoside Antibiotics. *Nucleic Acids Res.* **1993**, *21* (18), 4174–4179.
- (33) Böttger, E. C.; Crich, D. Aminoglycosides: Time for the Resurrection of a Neglected Class of Antibacterials? *ACS Infect. Dis.* **2020**, *6* (2), 168–172.
- (34) WHO Model Lists of Essential Medicines <https://www.who.int/groups/expert-committee-on-selection-and-use-of-essential-medicines/essential-medicines-lists> (accessed 2023 -03 -21).
- (35) Galimand, M.; Sabtcheva, S.; Courvalin, P.; Lambert, T. Worldwide Disseminated ArmA Aminoglycoside Resistance Methylase Gene Is Borne by Composite Transposon Tn1548. *Antimicrob. Agents Chemother.* **2005**, *49* (7), 2949–2953.
- (36) Doi, Y.; Wachino, J.-I.; Arakawa, Y. Aminoglycoside Resistance: The Emergence of Acquired 16S Ribosomal RNA Methyltransferases. *Infect. Dis. Clin. North Am.* **2016**, *30* (2), 523–537.
- (37) Poole, K. Efflux Pumps as Antimicrobial Resistance Mechanisms. *Ann. Med.* **2007**, *39* (3), 162–176.
- (38) Hancock, R. E. W. Aminoglycoside Uptake and Mode of Action—with Special Reference to Streptomycin and Gentamicin: I. Antagonists and Mutants. *J. Antimicrob. Chemother.* **1981**, *8* (4), 249–276.
-

- (39) Chandrika, N. T.; Garneau-Tsodikova, S. Comprehensive Review of Chemical Strategies for the Preparation of New Aminoglycosides and Their Biological Activities. *Chem. Soc. Rev.* **2018**, *47* (4), 1189–1249.
- (40) Shaer, K. M.; Zmarlicka, M. T.; Chahine, E. B.; Piccicacco, N.; Cho, J. C. Plazomicin: A Next-Generation Aminoglycoside. *Pharmacother. J. Hum. Pharmacol. Drug Ther.* **2019**, *39* (1), 77–93.
- (41) Golkar, T.; Bassenden, A. V.; Maiti, K.; Arya, D. P.; Schmeing, T. M.; Berghuis, A. M. Structural Basis for Plazomicin Antibiotic Action and Resistance. *Commun. Biol.* **2021**, *4* (1), 1–8.
- (42) Vicens, Q.; Westhof, E. Molecular Recognition of Aminoglycoside Antibiotics by Ribosomal RNA and Resistance Enzymes: An Analysis of x-Ray Crystal Structures. *Biopolymers* **2003**, *70* (1), 42–57.
- (43) Carter, A. P.; Clemons, W. M.; Brodersen, D. E.; Morgan-Warren, R. J.; Wimberly, B. T.; Ramakrishnan, V. Functional Insights from the Structure of the 30S Ribosomal Subunit and Its Interactions with Antibiotics. *Nature* **2000**, *407* (6802), 340–348.
- (44) Cook, M. A.; Wright, G. D. The Past, Present, and Future of Antibiotics. *Sci. Transl. Med.* **2022**, *14* (657), eabo7793.
- (45) RCSB PDB: Homepage <https://www.rcsb.org/> (accessed 2022 -04 -01).
- (46) Kellenberger, E.; Muller, P.; Schalon, C.; Bret, G.; Foata, N.; Rognan, D. Sc-PDB: An Annotated Database of Druggable Binding Sites from the Protein Data Bank. *J. Chem. Inf. Model.* **2006**, *46* (2), 717–727.
- (47) Desaphy, J.; Bret, G.; Rognan, D.; Kellenberger, E. Sc-PDB: A 3D-Database of Ligandable Binding Sites—10 Years On. *Nucleic Acids Res.* **2015**, *43* (D1), D399–D404.
- (48) Bietz, S.; Urbaczek, S.; Schulz, B.; Rarey, M. Protoss: A Holistic Approach to Predict Tautomers and Protonation States in Protein-Ligand Complexes. *J. Cheminformatics* **2014**, *6* (1), 12.
- (49) Desaphy, J.; Raimbaud, E.; Ducrot, P.; Rognan, D. Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. *J. Chem. Inf. Model.* **2013**, *53* (3), 623–637.
- (50) Da Silva, F.; Desaphy, J.; Rognan, D. IChem: A Versatile Toolkit for Detecting, Comparing, and Predicting Protein–Ligand Interactions. *ChemMedChem* **2018**, *13* (6), 507–510.
- (51) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, É. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12* (85), 2825–2830.
- (52) Corzana, F.; Cuesta, I.; Freire, F.; Revuelta, J.; Torrado, M.; Bastida, A.; Jiménez-Barbero, J.; Asensio, J. L. The Pattern of Distribution of Amino Groups Modulates the Structure and Dynamics of Natural Aminoglycosides: Implications for RNA Recognition. *J. Am. Chem. Soc.* **2007**, *129* (10), 2849–2865.
- (53) Daigle, D. M.; McKay, G. A.; Thompson, P. R.; Wright, G. D. Aminoglycoside Antibiotic Phosphotransferases Are Also Serine Protein Kinases. *Chem. Biol.* **1999**, *6* (1), 11–18.

-
- (54) Vetting, M. W.; S. de Carvalho, L. P.; Yu, M.; Hegde, S. S.; Magnet, S.; Roderick, S. L.; Blanchard, J. S. Structure and Functions of the GNAT Superfamily of Acetyltransferases. *Arch. Biochem. Biophys.* **2005**, *433* (1), 212–226.
- (55) Selvaraj, B.; Kocaman, S.; Trifas, M.; Serpersu, E. H.; Cuneo, M. J. “Catch and Release”: A Variation of the Archetypal Nucleotidyl Transfer Reaction. *ACS Catal.* **2020**, *10* (6), 3548–3555.
- (56) Ramirez, M. S.; Tolmasky, M. E. Amikacin: Uses, Resistance, and Prospects for Inhibition. *Molecules* **2017**, *22* (12), 2267.
- (57) Price, K. E. The Potential for Discovery and Development of Improved Aminoglycosides. *Am. J. Med.* **1986**, *80* (6, Supplement 2), 182–189.
- (58) Matsumoto, T. Arbekacin: Another Novel Agent for Treating Infections Due to Methicillin-Resistant Staphylococcus Aureus and Multidrug-Resistant Gram-Negative Pathogens. *Clin. Pharmacol. Adv. Appl.* **2014**, *6*, 139–148.
- (59) Bera, S.; Mondal, D.; Palit, S.; Schweizer, F. Structural Modifications of the Neomycin Class of Aminoglycosides. *MedChemComm* **2016**, *7* (8), 1499–1534.
- (60) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.
- (61) Jones, G.; Willett, P.; Glen, R. C.; Leach, A. R.; Taylor, R. Development and Validation of a Genetic Algorithm for Flexible Docking. *J. Mol. Biol.* **1997**, *267* (3), 727–748.
- (62) Spitzer, R.; Jain, A. N. Surflex-Dock: Docking Benchmarks and Real-World Application. *J. Comput. Aided Mol. Des.* **2012**, *26* (6), 687–699.
- (63) D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R. Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J.; Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R.; Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. SalomonFerrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman. *AMBER 2018*; University of California: San Francisco, 2018.
- (64) NVIDIA; Vingelmann, P.; Fitzek, F. H. P. *CUDA, Release: 10.1*; 2020.
- (65) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theory Comput.* **2013**, *9* (7), 3084–3095.
- (66) Minogue, T. D.; Daligault, H. A.; Davenport, K. W.; Bishop-Lilly, K. A.; Broomall, S. M.; Bruce, D. C.; Chain, P. S.; Chertkov, O.; Coyne, S. R.; Freitas, T.; Frey, K. G.; Gibbons, H. S.; Jaissle, J.; Redden, C. L.; Rosenzweig, C. N.; Xu, Y.; Johnson, S. L. Complete Genome Assembly of Escherichia Coli ATCC 25922, a Serotype O6 Reference Strain. *Genome Announc.* **2014**, *2* (5), 10.1128/genomea.00969-14.
- (67) Treangen, T. J.; Maybank, R. A.; Enke, S.; Friss, M. B.; Diviak, L. F.; Karaolis, D. K. R.; Koren, S.; Ondov, B.; Phillippy, A. M.; Bergman, N. H.; Rosovitz, M. J. Complete Genome Sequence of the Quality Control Strain Staphylococcus Aureus Subsp. Aureus ATCC 25923. *Genome Announc.* **2014**, *2* (6), 10.1128/genomea.01110-14.
-

- (68) Kim, C.; Mobashery, S. Phosphoryl Transfer by Aminoglycoside 3'-Phosphotransferases and Manifestation of Antibiotic Resistance. *Bioorganic Chem.* **2005**, *33* (3), 149–158.
- (69) Fong, D. H.; Berghuis, A. M. Substrate Promiscuity of an Aminoglycoside Antibiotic Resistance Enzyme via Target Mimicry. *EMBO J.* **2002**, *21* (10), 2323–2331.
- (70) Nurizzo, D.; Shewry, S. C.; Perlin, M. H.; Brown, S. A.; Dholakia, J. N.; Fuchs, R. L.; Deva, T.; Baker, E. N.; Smith, C. A. The Crystal Structure of Aminoglycoside-3'-Phosphotransferase-IIa, an Enzyme Responsible for Antibiotic Resistance. *J. Mol. Biol.* **2003**, *327* (2), 491–506.
- (71) Stogios, P. J.; Spanogiannopoulos, P.; Evdokimova, E.; Egorova, O.; Shakya, T.; Todorovic, N.; Capretta, A.; Wright, G. D.; Savchenko, A. Structure-Guided Optimization of Protein Kinase Inhibitors Reverses Aminoglycoside Antibiotic Resistance. *Biochem. J.* **2013**, *454* (2), 191–200.
- (72) Chen-Goodspeed, M.; Vanhooke, J. L.; Holden, H. M.; Raushel, F. M. Kinetic Mechanism of Kanamycin Nucleotidyltransferase from *Staphylococcus Aureus*. *Bioorganic Chem.* **1999**, *27* (5), 395–408.
- (73) Pedersen, L. C.; Benning, M. M.; Holden, H. M. Structural Investigation of the Antibiotic and ATP-Binding Sites in Kanamycin Nucleotidyltransferase. *Biochemistry* **1995**, *34* (41), 13305–13311.
- (74) Azucena, E.; Mobashery, S. Aminoglycoside-Modifying Enzymes: Mechanisms of Catalytic Processes and Inhibition. *Drug Resist. Updat.* **2001**, *4* (2), 106–117.
- (75) Matesanz, R.; Diaz, J. F.; Corzana, F.; Santana, A. G.; Bastida, A.; Asensio, J. L. Multiple Keys for a Single Lock: The Unusual Structural Plasticity of the Nucleotidyltransferase (4')/Kanamycin Complex. *Chem. – Eur. J.* **2012**, *18* (10), 2875–2889.
- (76) Kumar, P.; Agarwal, P. K.; Waddell, M. B.; Mittag, T.; Serpersu, E. H.; Cuneo, M. J. Low-Barrier and Canonical Hydrogen Bonds Modulate Activity and Specificity of a Catalytic Triad. *Angew. Chem. Int. Ed.* **2019**, *58* (45), 16260–16266.
- (77) Romanowska, J.; Reuter, N.; Trylska, J. Comparing Aminoglycoside Binding Sites in Bacterial Ribosomal RNA and Aminoglycoside Modifying Enzymes. *Proteins Struct. Funct. Bioinforma.* **2013**, *81* (1), 63–80.
- (78) Burnouf, D. Y.; Olieric, V.; Wagner, J.; Fujii, S.; Reinbolt, J.; Fuchs, R. P. P.; Dumas, P. Structural and Biochemical Analysis of Sliding Clamp/Ligand Interactions Suggest a Competition Between Replicative and Translesion DNA Polymerases. *J. Mol. Biol.* **2004**, *335* (5), 1187–1197.
- (79) Becherel, O. J.; Fuchs, R. P. P.; Wagner, J. Pivotal Role of the β -Clamp in Translesion DNA Synthesis and Mutagenesis in *E. Coli* Cells. *DNA Repair* **2002**, *1* (9), 703–708.
- (80) Altieri, A. S.; Kelman, Z. DNA Sliding Clamps as Therapeutic Targets. *Front. Mol. Biosci.* **2018**, *5*.
- (81) Kong, X. P.; Onrust, R.; O'Donnell, M.; Kuriyan, J. Three-Dimensional Structure of the Beta Subunit of *E. Coli* DNA Polymerase III Holoenzyme: A Sliding DNA Clamp. *Cell* **1992**, *69* (3), 425–437.
- (82) Dalrymple, B. P.; Kongsuwan, K.; Wijffels, G.; Dixon, N. E.; Jennings, P. A. A Universal Protein–Protein Interaction Motif in the Eubacterial DNA Replication and Repair Systems. *Proc. Natl. Acad. Sci.* **2001**, *98* (20), 11627–11632.
-

-
- (83) Argiriadi, M. A.; Goedken, E. R.; Bruck, I.; O'Donnell, M.; Kuriyan, J. Crystal Structure of a DNA Polymerase Sliding Clamp from a Gram-Positive Bacterium. *BMC Struct. Biol.* **2006**, *6* (1), 2.
- (84) Gui, W.-J.; Lin, S.-Q.; Chen, Y.-Y.; Zhang, X.-E.; Bi, L.-J.; Jiang, T. Crystal Structure of DNA Polymerase III β Sliding Clamp from Mycobacterium Tuberculosis. *Biochem. Biophys. Res. Commun.* **2011**, *405* (2), 272–277.
- (85) Wolff, P.; Amal, I.; Oliéric, V.; Chaloin, O.; Gygli, G.; Ennifar, E.; Lorber, B.; Guichard, G.; Wagner, J.; Dejaegere, A.; Burnouf, D. Y. Differential Modes of Peptide Binding onto Replicative Sliding Clamps from Various Bacterial Origins. *J. Med. Chem.* **2014**, *57* (18), 7565–7576.
- (86) Cirz, R. T.; O'Neill, B. M.; Hammond, J. A.; Head, S. R.; Romesberg, F. E. Defining the Pseudomonas Aeruginosa SOS Response and Its Role in the Global Response to the Antibiotic Ciprofloxacin. *J. Bacteriol.* **2006**, *188* (20), 7101–7110.
- (87) Smith, P. A.; Romesberg, F. E. Combating Bacteria and Drug Resistance by Inhibiting Mechanisms of Persistence and Adaptation. *Nat. Chem. Biol.* **2007**, *3* (9), 549–556.
- (88) Naktinis, V.; Turner, J.; O'Donnell, M. A Molecular Switch in a Replication Machine Defined by an Internal Competition for Protein Rings. *Cell* **1996**, *84* (1), 137–145.
- (89) Wolff, P.; Oliéric, V.; Briand, J. P.; Chaloin, O.; Dejaegere, A.; Dumas, P.; Ennifar, E.; Guichard, G.; Wagner, J.; Burnouf, D. Y. Structure-Based Design of Short Peptide Ligands Binding onto the E. Coli Processivity Ring. *J. Med. Chem.* **2011**, *54* (13), 4627–4637.
- (90) Yin, Z.; Kelso, M. J.; Beck, J. L.; Oakley, A. J. Structural and Thermodynamic Dissection of Linear Motif Recognition by the E. Coli Sliding Clamp. *J. Med. Chem.* **2013**, *56* (21), 8665–8673.
- (91) Wijffels, G.; Johnson, W. M.; Oakley, A. J.; Turner, K.; Epa, V. C.; Briscoe, S. J.; Polley, M.; Liepa, A. J.; Hofmann, A.; Buchardt, J.; Christensen, C.; Prosselkov, P.; Dalrymple, B. P.; Alewood, P. F.; Jennings, P. A.; Dixon, N. E.; Winkler, D. A. Binding Inhibitors of the Bacterial Sliding Clamp by Design. *J. Med. Chem.* **2011**, *54* (13), 4831–4838.
- (92) Flores-Rozas, H.; Kelman, Z.; Dean, F. B.; Pan, Z. Q.; Harper, J. W.; Elledge, S. J.; O'Donnell, M.; Hurwitz, J. Cdk-Interacting Protein 1 Directly Binds with Proliferating Cell Nuclear Antigen and Inhibits DNA Replication Catalyzed by the DNA Polymerase Delta Holoenzyme. *Proc. Natl. Acad. Sci.* **1994**, *91* (18), 8655–8659.
- (93) Kling, A.; Lukat, P.; Almeida, D. V.; Bauer, A.; Fontaine, E.; Sordello, S.; Ziburanyi, N.; Herrmann, J.; Wenzel, S. C.; König, C.; Ammerman, N. C.; Barrio, M. B.; Borchers, K.; Bordon-Pallier, F.; Brönstrup, M.; Courtemanche, G.; Gerlitz, M.; Geslin, M.; Hammann, P.; Heinz, D. W.; Hoffmann, H.; Klieber, S.; Kohlmann, M.; Kurz, M.; Lair, C.; Matter, H.; Nuermberger, E.; Tyagi, S.; Fraisse, L.; Grosset, J. H.; Lagrange, S.; Müller, R. Targeting DnaN for Tuberculosis Therapy Using Novel Griselimycins. *Science* **2015**, *348* (6239), 1106–1112.
- (94) Yin, Z.; Wang, Y.; Whittell, L. R.; Jergic, S.; Liu, M.; Harry, E.; Dixon, N. E.; Kelso, M. J.; Beck, J. L.; Oakley, A. J. DNA Replication Is the Target for the Antibacterial Effects of Nonsteroidal Anti-Inflammatory Drugs. *Chem. Biol.* **2014**, *21* (4), 481–487.
- (95) Georgescu, R. E.; Yurieva, O.; Kim, S.-S.; Kuriyan, J.; Kong, X.-P.; O'Donnell, M. Structure of a Small-Molecule Inhibitor of a DNA Polymerase Sliding Clamp. *Proc. Natl. Acad. Sci.* **2008**, *105* (32), 11116–11121.
-

- (96) Yin, Z.; Whittell, L. R.; Wang, Y.; Jergic, S.; Liu, M.; Harry, E. J.; Dixon, N. E.; Beck, J. L.; Kelso, M. J.; Oakley, A. J. Discovery of Lead Compounds Targeting the Bacterial Sliding Clamp Using a Fragment-Based Approach. *J. Med. Chem.* **2014**, *57* (6), 2799–2806.
- (97) Carr, R. A. E.; Congreve, M.; Murray, C. W.; Rees, D. C. Fragment-Based Lead Discovery: Leads by Design. *Drug Discov. Today* **2005**, *10* (14), 987–992.
- (98) Murray, C. W.; Rees, D. C. The Rise of Fragment-Based Drug Discovery. *Nat. Chem.* **2009**, *1* (3), 187–192.
- (99) Jahnke, W.; Erlanson, D. A.; de Esch, I. J. P.; Johnson, C. N.; Mortenson, P. N.; Ochi, Y.; Urushima, T. Fragment-to-Lead Medicinal Chemistry Publications in 2019. *J. Med. Chem.* **2020**, *63* (24), 15494–15507.
- (100) Romasanta, A. K. S.; van der Sijde, P.; Hellsten, I.; Hubbard, R. E.; Keseru, G. M.; van Muijlwijk-Koezen, J.; de Esch, I. J. P. When Fragments Link: A Bibliometric Perspective on the Development of Fragment-Based Drug Discovery. *Drug Discov. Today* **2018**, *23* (9), 1596–1609.
- (101) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274* (5292), 1531–1534.
- (102) Singh, M.; Tam, B.; Akabayov, B. NMR-Fragment Based Virtual Screening: A Brief Overview. *Molecules* **2018**, *23* (2), 233.
- (103) O'Reilly, M.; Cleasby, A.; Davies, T. G.; Hall, R. J.; Ludlow, R. F.; Murray, C. W.; Tisi, D.; Jhoti, H. Crystallographic Screening Using Ultra-Low-Molecular-Weight Ligands to Guide Drug Design. *Drug Discov. Today* **2019**, *24* (5), 1081–1086.
- (104) Congreve, M.; Carr, R. A. E.; Murray, C. W.; Jhoti, H. A 'Rule of Three' for Fragment-Based Lead Discovery? *Drug Discov. Today* **2003**, *8* (19), 876–877.
- (105) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **2001**, *24*.
- (106) Jhoti, H.; Williams, G.; Rees, D. C.; Murray, C. W. The "rule of Three" for Fragment-Based Drug Discovery: Where Are We Now? *Nat. Rev. Drug Discov.* **2013**, *12* (8), 644–644.
- (107) Erlanson, D. A.; Fesik, S. W.; Hubbard, R. E.; Jahnke, W.; Jhoti, H. Twenty Years on: The Impact of Fragments on Drug Discovery. *Nat. Rev. Drug Discov.* **2016**, *15* (9), 605–619.
- (108) Bickerton, G. R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A. L. Quantifying the Chemical Beauty of Drugs. *Nat. Chem.* **2012**, *4* (2), 90–98.
- (109) Konteatis, Z. What Makes a Good Fragment in Fragment-Based Drug Discovery? *Expert Opin. Drug Discov.* **2021**, *16* (7), 723–726.
- (110) Ferenczy, G. G.; Keserü, G. M. On the Enthalpic Preference of Fragment Binding. *MedChemComm* **2016**, *7* (2), 332–337.
- (111) Jacquemard, C.; Kellenberger, E. A Bright Future for Fragment-Based Drug Discovery: What Does It Hold? *Expert Opin. Drug Discov.* **2019**, *14* (5), 413–416.
- (112) Keserü, G. M.; Erlanson, D. A.; Ferenczy, G. G.; Hann, M. M.; Murray, C. W.; Pickett, S. D. Design Principles for Fragment Libraries: Maximizing the Value of Learnings from Pharma
-

- Fragment-Based Drug Discovery (FBDD) Programs for Use in Academia. *J. Med. Chem.* **2016**, *59* (18), 8189–8206.
- (113) Hann, M. M.; Leach, A. R.; Harper, G. Molecular Complexity and Its Impact on the Probability of Finding Leads for Drug Discovery. *J. Chem. Inf. Comput. Sci.* **2001**, *41* (3), 856–864.
- (114) Shi, Y.; von Itzstein, M. How Size Matters: Diversity for Fragment Library Design. *Molecules* **2019**, *24* (15), 2838.
- (115) Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem.* **2009**, *52* (21), 6752–6756.
- (116) Péron, F.; Riché, S.; Lesur, B.; Hibert, M.; Breton, P.; Fourquez, J.-M.; Girard, N.; Bonnet, D. Versatile Synthetic Approach for Selective Diversification of Bicyclic Aza-Diketopiperazines. *ACS Omega* **2018**, *3* (11), 15182–15192.
- (117) Downes, T. D.; Jones, S. P.; Klein, H. F.; Wheldon, M. C.; Atobe, M.; Bond, P. S.; Firth, J. D.; Chan, N. S.; Waddelove, L.; Hubbard, R. E.; Blakemore, D. C.; De Fusco, C.; Roughley, S. D.; Vidler, L. R.; Whatton, M. A.; Woolford, A. J. -A.; Wrigley, G. L.; O'Brien, P. Design and Synthesis of 56 Shape-Diverse 3D Fragments. *Chem. – Eur. J.* **2020**, *26* (41), 8969–8975.
- (118) Tang, S.-Q.; Leloire, M.; Schneider, S.; Mohr, J.; Bricard, J.; Gizzi, P.; Garnier, D.; Schmitt, M.; Bihel, F. Diastereoselective Synthesis of Nonplanar 3-Amino-1,2,4-Oxadiazine Scaffold: Structure Revision of Alchornedine. *J. Org. Chem.* **2020**, *85* (23), 15347–15359.
- (119) Li Petri, G.; Raimondi, M. V.; Spanò, V.; Holl, R.; Barraja, P.; Montalbano, A. Pyrrolidine in Drug Discovery: A Versatile Scaffold for Novel Biologically Active Compounds. *Top. Curr. Chem.* **2021**, *379* (5), 34.
- (120) Hiesinger, K.; Dar'in, D.; Proschak, E.; Krasavin, M. Spirocyclic Scaffolds in Medicinal Chemistry. *J. Med. Chem.* **2021**, *64* (1), 150–183.
- (121) Hall, R. J.; Mortenson, P. N.; Murray, C. W. Efficient Exploration of Chemical Space by Fragment-Based Screening. *Prog. Biophys. Mol. Biol.* **2014**, *116* (2–3), 82–91.
- (122) Ghose, A. K.; Viswanadhan, V. N.; Wendoloski, J. J. Prediction of Hydrophobic (Lipophilic) Properties of Small Organic Molecules Using Fragmental Methods: An Analysis of ALOGP and CLOGP Methods. *J. Phys. Chem. A* **1998**, *102* (21), 3762–3772.
- (123) Firth, N. C.; Brown, N.; Blagg, J. Plane of Best Fit: A Novel Method to Characterize the Three-Dimensionality of Molecules. *J. Chem. Inf. Model.* **2012**, *52* (10), 2516–2525.
- (124) Bemis, G. W.; Murcko, M. A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39* (15), 2887–2893.
- (125) Kireeva, N.; Baskin, I. I.; Gaspar, H. A.; Horvath, D.; Marcou, G.; Varnek, A. Generative Topographic Mapping (GTM): Universal Tool for Data Visualization, Structure-Activity Modeling and Dataset Comparison. *Mol. Inform.* **2012**, *31* (3–4), 301–312.
- (126) Zimmermann, M. O.; Lange, A.; Wilcken, R.; Cieslik, M. B.; Exner, T. E.; Joerger, A. C.; Koch, P.; Boeckler, F. M. Halogen-Enriched Fragment Libraries as Chemical Probes for Harnessing Halogen Bonding in Fragment-Based Lead Discovery. *Future Med. Chem.* **2014**, *6* (6), 617–639.

- (127) Hann, M. M.; Keserü, G. M. Finding the Sweet Spot: The Role of Nature and Nurture in Medicinal Chemistry. *Nat. Rev. Drug Discov.* **2012**, *11* (5), 355–365.
- (128) Hann, M. M. Molecular Obesity, Potency and Other Addictions in Drug Discovery. *MedChemComm* **2011**, *2* (5), 349–355.
- (129) Over, B.; Wetzel, S.; Grütter, C.; Nakai, Y.; Renner, S.; Rauh, D.; Waldmann, H. Natural-Product-Derived Fragments for Fragment-Based Ligand Discovery. *Nat. Chem.* **2013**, *5* (1), 21–28.
- (130) Cox, O. B.; Krojer, T.; Collins, P.; Monteiro, O.; Talon, R.; Bradley, A.; Fedorov, O.; Amin, J.; Marsden, B. D.; Spencer, J.; Delft, F. von; Brennan, P. E. A Poised Fragment Library Enables Rapid Synthetic Expansion Yielding the First Reported Inhibitors of PHIP(2), an Atypical Bromodomain. *Chem. Sci.* **2016**, *7* (3), 2322–2330.
- (131) *Molecular Operating Environment (MOE)*; Chemical Computing Group ULC: 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2022.
- (132) Lippert, T.; Rarey, M. Fast Automated Placement of Polar Hydrogen Atoms in Protein-Ligand Complexes. *J. Cheminformatics* **2009**, *1* (1), 13.
- (133) Tran-Nguyen, V.-K.; Da Silva, F.; Bret, G.; Rognan, D. All in One: Cavity Detection, Druggability Estimate, Cavity-Based Pharmacophore Perception, and Virtual Screening. *J. Chem. Inf. Model.* **2019**, *59* (1), 573–585.
- (134) <https://enamine.net/compound-collections/real-compounds/real-compound-libraries>.
- (135) <https://enamine.net/compound-collections/screening-collection>.
- (136) ChemBioFrance - Infrastructure de recherche <https://chembiofrance.cn.cnrs.fr/fr/composante/chimiotheque> (accessed 2022 -02 -21).
- (137) Rognan, D. BioinfoDB: Un Inventaire de Molécules Commercialement Disponibles à Des Fins de Criblage Biologique. *Gaz. CINES* **2005**, 1–4.
- (138) *Corina*.
- (139) *OMEGA 2.5.1.4*; OpenEye Scientific Software, Santa Fe, NM. [Http://www.eyesopen.com](http://www.eyesopen.com). Hawkins, P.C.D; Skillman, A.G.; Warren, G.L.; Ellingson, B.A.; Stahl, M.T.
- (140) *ROCS 3.2.0.4*; OpenEye Scientific Software, Santa Fe, NM. [Http://www.eyesopen.com](http://www.eyesopen.com).
- (141) Korb, O.; Stützel, T.; Exner, T. E. Empirical Scoring Functions for Advanced Protein–Ligand Docking with PLANTS. *J. Chem. Inf. Model.* **2009**, *49* (1), 84–96.
- (142) Jacquemard, C.; Tran-Nguyen, V.-K.; Drwal, M. N.; Rognan, D.; Kellenberger, E. Local Interaction Density (LID), a Fast and Efficient Tool to Prioritize Docking Poses. *Molecules* **2019**, *24* (14), 2610.
- (143) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: A Knowledgebase for Drugs, Drug Actions and Drug Targets. *Nucleic Acids Res.* **2008**, *36* (Database issue), D901–D906.
- (144) Hawkins, P. C. D.; Skillman, A. G.; Nicholls, A. Comparison of Shape-Matching and Docking as Virtual Screening Tools. *J. Med. Chem.* **2007**, *50* (1), 74–82.
-

-
- (145) de Esch, I. J. P.; Erlanson, D. A.; Jahnke, W.; Johnson, C. N.; Walsh, L. Fragment-to-Lead Medicinal Chemistry Publications in 2020. *J. Med. Chem.* **2022**, *65* (1), 84–99.
- (146) Erlanson, D. Practical Fragments: Fragments in the Clinic: 2021 Edition. *Practical Fragments*, 2021.
- (147) Li, Q. Application of Fragment-Based Drug Discovery to Versatile Targets. *Front. Mol. Biosci.* **2020**, *7*.
- (148) Revillo Imbernon, J.; Jacquemard, C.; Bret, G.; Marcou, G.; Kellenberger, E. Comprehensive Analysis of Commercial Fragment Libraries. *RSC Med. Chem.* **2022**, *13* (3), 300–310.
- (149) Troelsen, N. S.; Clausen, M. H. Library Design Strategies To Accelerate Fragment-Based Drug Discovery. *Chem. – Eur. J.* **2020**, *26* (50), 11391–11403.
- (150) Shinya, S.; Katahira, R.; Furuita, K.; Sugiki, T.; Lee, Y.-H.; Hattori, Y.; Takeshita, K.; Nakagawa, A.; Kokago, A.; Akagi, K.; Oouchi, M.; Hayashi, F.; Kigawa, T.; Takimoto-Kamimura, M.; Fujiwara, T.; Kojima, C. ¹⁹F Chemical Library and ¹⁹F-NMR for a Weakly Bound Complex Structure. *RSC Med. Chem.* **2022**, *13* (9), 1100–1111.
- (151) Kidd, S. L.; Osberger, T. J.; Mateu, N.; Sore, H. F.; Spring, D. R. Recent Applications of Diversity-Oriented Synthesis Toward Novel, 3-Dimensional Fragment Collections. *Front. Chem.* **2018**, *6*, 460.
- (152) Heidrich, J.; Sperl, L. E.; Boeckler, F. M. Embracing the Diversity of Halogen Bonding Motifs in Fragment-Based Drug Discovery—Construction of a Diversity-Optimized Halogen-Enriched Fragment Library. *Front. Chem.* **2019**, *7*.
- (153) Ferri, M.; Alunno, M.; Greco, F. A.; Mammoli, A.; Saluti, G.; Carotti, A.; Sardella, R.; Macchiarulo, A.; Camaioni, E.; Liscio, P. Fragment Based Drug Design and Diversity-Oriented Synthesis of Carboxylic Acid Isosteres. *Bioorg. Med. Chem.* **2020**, *28* (22), 115731.
- (154) Wollenhaupt, J.; Metz, A.; Barthel, T.; Lima, G. M. A.; Heine, A.; Mueller, U.; Klebe, G.; Weiss, M. S. F2X-Universal and F2X-Entry: Structurally Diverse Compound Libraries for Crystallographic Fragment Screening. *Structure* **2020**, *28* (6), 694-706.e5.
- (155) Carbery, A.; Skyner, R.; von Delft, F.; Deane, C. M. Fragment Libraries Designed to Be Functionally Diverse Recover Protein Binding Information More Efficiently Than Standard Structurally Diverse Libraries. *J. Med. Chem.* **2022**, *65* (16), 11404–11413.
- (156) Bajusz, D.; Wade, W. S.; Satała, G.; Bojarski, A. J.; Ilaš, J.; Ebner, J.; Grebien, F.; Papp, H.; Jakab, F.; Douangamath, A.; Fearon, D.; von Delft, F.; Schuller, M.; Ahel, I.; Wakefield, A.; Vajda, S.; Gerencsér, J.; Pallai, P.; Keserű, G. M. Exploring Protein Hotspots by Optimized Fragment Pharmacophores. *Nat. Commun.* **2021**, *12* (1), 3201.
- (157) Desaphy, J.; Bret, G.; Rognan, D.; Kellenberger, E. Sc-PDB: A 3D-Database of Ligandable Binding Sites--10 Years On. *Nucleic Acids Res.* **2015**, *43* (Database issue), D399-404.
- (158) Drwal, M. N.; Bret, G.; Perez, C.; Jacquemard, C.; Desaphy, J.; Kellenberger, E. Structural Insights on Fragment Binding Mode Conservation. *J. Med. Chem.* **2018**, *61* (14), 5963–5973.
- (159) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28* (1), 235–242.
-

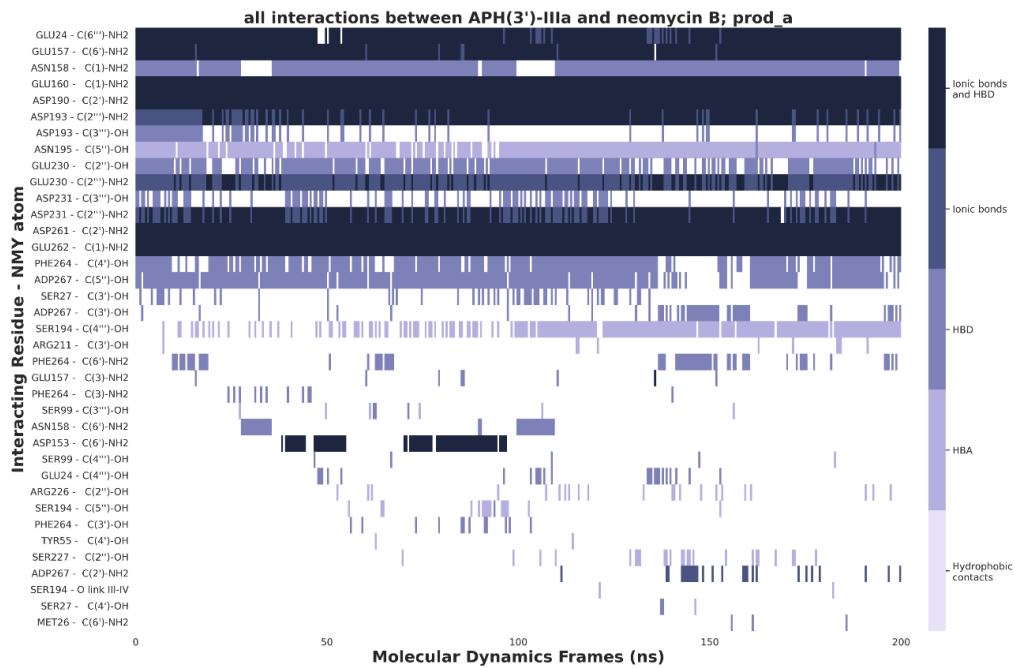
- (160) Desaphy, J.; Raimbaud, E.; Ducrot, P.; Rognan, D. Encoding Protein–Ligand Interaction Patterns in Fingerprints and Graphs. *J. Chem. Inf. Model.* **2013**, *53* (3), 623–637.
- (161) Siglidis, G.; Nikolentzos, G.; Limnios, S.; Giatsidis, C.; Skianis, K.; Vazirgiannis, M. GraKeL: A Graph Kernel Library in Python. *J. Mach. Learn. Res.* **2020**, *21*, 1–5.
- (162) Desaphy, J.; Azdimousa, K.; Kellenberger, E.; Rognan, D. Comparison and Druggability Prediction of Protein–Ligand Binding Sites from Pharmacophore-Annotated Cavity Shapes. *J. Chem. Inf. Model.* **2012**, *52* (8), 2287–2299.
- (163) Eguida, M.; Rognan, D. A Computer Vision Approach to Align and Compare Protein Cavities: Application to Fragment-Based Drug Design. *J. Med. Chem.* **2020**, *63* (13), 7127–7142.
- (164) Baell, J. B.; Holloway, G. A. New Substructure Filters for Removal of Pan Assay Interference Compounds (PAINS) from Screening Libraries and for Their Exclusion in Bioassays. *J. Med. Chem.* **2010**, *53* (7), 2719–2740.
- (165) Weichenberger, C. X.; Pozharski, E.; Rupp, B. Visualizing Ligand Molecules in Twilight Electron Density. *Acta Crystallograph. Sect. F Struct. Biol. Cryst. Commun.* **2013**, *69* (Pt 2), 195–200.
- (166) Deller, M. C.; Rupp, B. Models of Protein–Ligand Crystal Structures: Trust, but Verify. *J. Comput. Aided Mol. Des.* **2015**, *29* (9), 817–836.
- (167) Perebyinis, M.; Rognan, D. Overlap of On-Demand Ultra-Large Combinatorial Spaces with On-the-Shelf Drug-like Libraries. *Mol. Inform.* **2022**, *41* (2200163).
- (168) Wishart, D. S.; Feunang, Y. D.; Guo, A. C.; Lo, E. J.; Marcu, A.; Grant, J. R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; Assempour, N.; Iynkkaran, I.; Liu, Y.; Maciejewski, A.; Gale, N.; Wilson, A.; Chin, L.; Cummings, R.; Le, D.; Pon, A.; Knox, C.; Wilson, M. DrugBank 5.0: A Major Update to the DrugBank Database for 2018. *Nucleic Acids Res.* **2018**, *46* (Database issue), D1074–D1082.
- (169) Wilson, K. A.; Kung, R. W.; D’souza, S.; Wetmore, S. D. Anatomy of Noncovalent Interactions between the Nucleobases or Ribose and π -Containing Amino Acids in RNA–Protein Complexes. *Nucleic Acids Res.* **2021**, *49* (4), 2213–2225.
- (170) Westbrook, J.; Burley, S. K. How Structural Biologists and the Protein Data Bank Contributed to Recent US FDA New Drug Approvals. *Struct. Lond. Engl. 1993* **2019**, *27* (2), 211–217.
- (171) Young, R. J.; Flitsch, S. L.; Grigalunas, M.; Leeson, P. D.; Quinn, R. J.; Turner, N. J.; Waldmann, H. The Time and Place for Nature in Drug Discovery. *JACS Au* **2022**, jacsau.2c00415.
- (172) Ertl, P.; Schuhmann, T. A Systematic Cheminformatics Analysis of Functional Groups Occurring in Natural Products. *J. Nat. Prod.* **2019**, *82* (5), 1258–1263.
- (173) Seley-Radtke, K. L.; Yates, M. K. The Evolution of Nucleoside Analogue Antivirals: A Review for Chemists and Non-Chemists. Part 1: Early Structural Modifications to the Nucleoside Scaffold. *Antiviral Res.* **2018**, *154*, 66–86.
- (174) Yates, M. K.; Seley-Radtke, K. L. The Evolution of Antiviral Nucleoside Analogues: A Review for Chemists and Non-Chemists. Part II: Complex Modifications to the Nucleoside Scaffold. *Antiviral Res.* **2019**, *162*, 5–21.

- (175) Chen, X.; Liu, M.; Gilson, M. BindingDB: A Web-Accessible Molecular Recognition Database. *Comb. Chem. High Throughput Screen.* **2001**, *4* (8), 719–725.
- (176) Sturm, N.; Desaphy, J.; Quinn, R. J.; Rognan, D.; Kellenberger, E. Structural Insights into the Molecular Basis of the Ligand Promiscuity. *J. Chem. Inf. Model.* **2012**, *52* (9), 2410–2421.
- (177) Barelier, S.; Sterling, T.; O’Meara, M. J.; Shoichet, B. K. The Recognition of Identical Ligands by Unrelated Proteins. *ACS Chem. Biol.* **2015**, *10* (12), 2772–2784.
- (178) Pottel, J.; Levit, A.; Korczynska, M.; Fischer, M.; Shoichet, B. K. The Recognition of Unrelated Ligands by Identical Proteins. *ACS Chem. Biol.* **2018**, *13* (9), 2522–2533.
- (179) Feldmann, C.; Bajorath, J. Biological Activity Profiles of Multitarget Ligands from X-Ray Structures. *Molecules* **2020**, *25* (4), 794.
- (180) Pinzi, L.; Rastelli, G. Identification of Target Associations for Polypharmacology from Analysis of Crystallographic Ligands of the Protein Data Bank. *J. Chem. Inf. Model.* **2020**, *60* (1), 372–390.
- (181) Chaudhari, R.; Tan, Z.; Huang, B.; Zhang, S. Computational Polypharmacology: A New Paradigm for Drug Discovery. *Expert Opin. Drug Discov.* **2017**, *12* (3), 279–291.
- (182) Proschak, E.; Stark, H.; Merk, D. Polypharmacology by Design: A Medicinal Chemist’s Perspective on Multitargeting Compounds. *J. Med. Chem.* **2019**, *62* (2), 420–444.

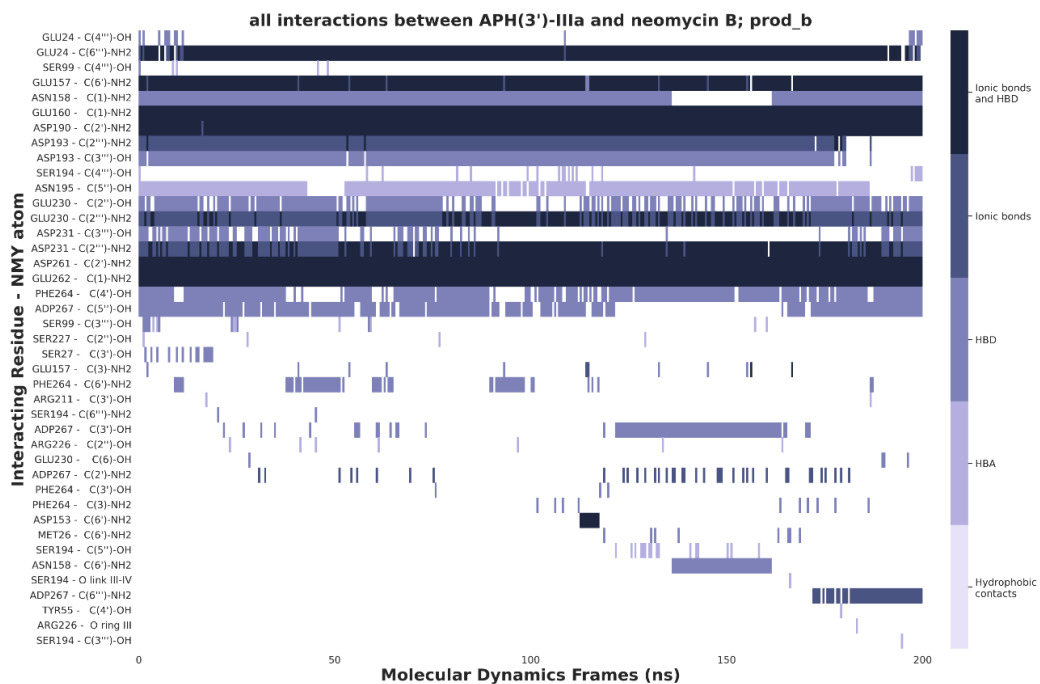
VI Annexes

VI.1 Interactions between AME and AGA during the MD simulations

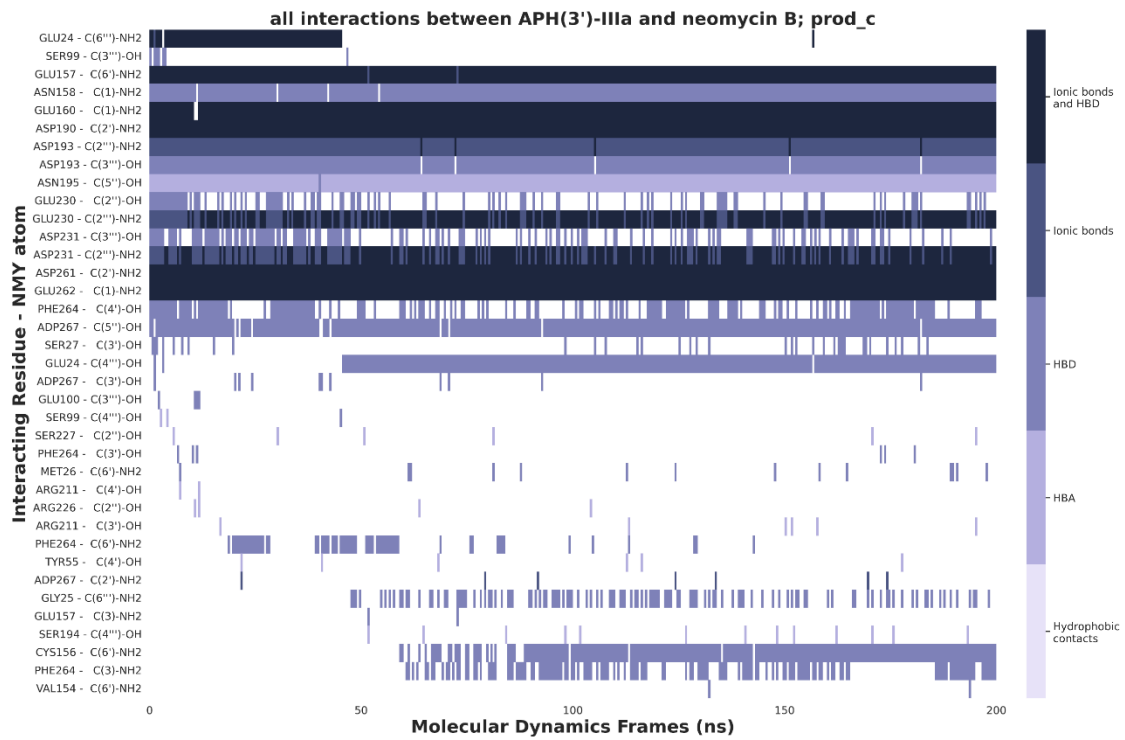
VI.1.1



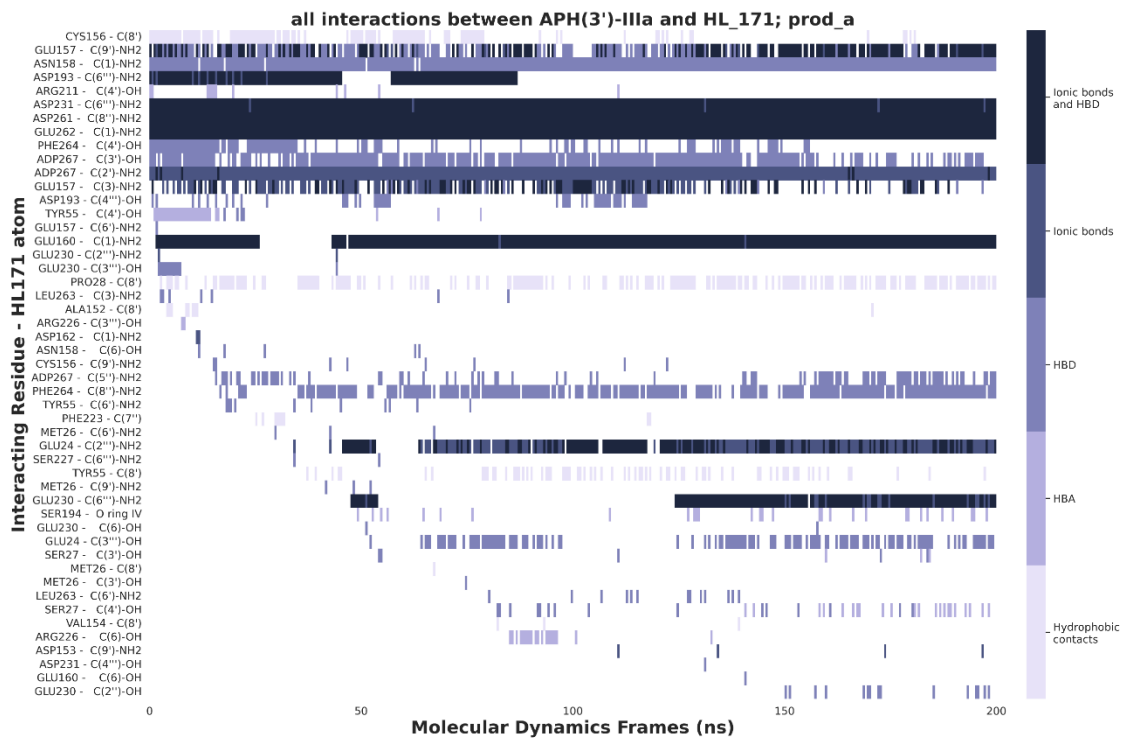
VI.1.2



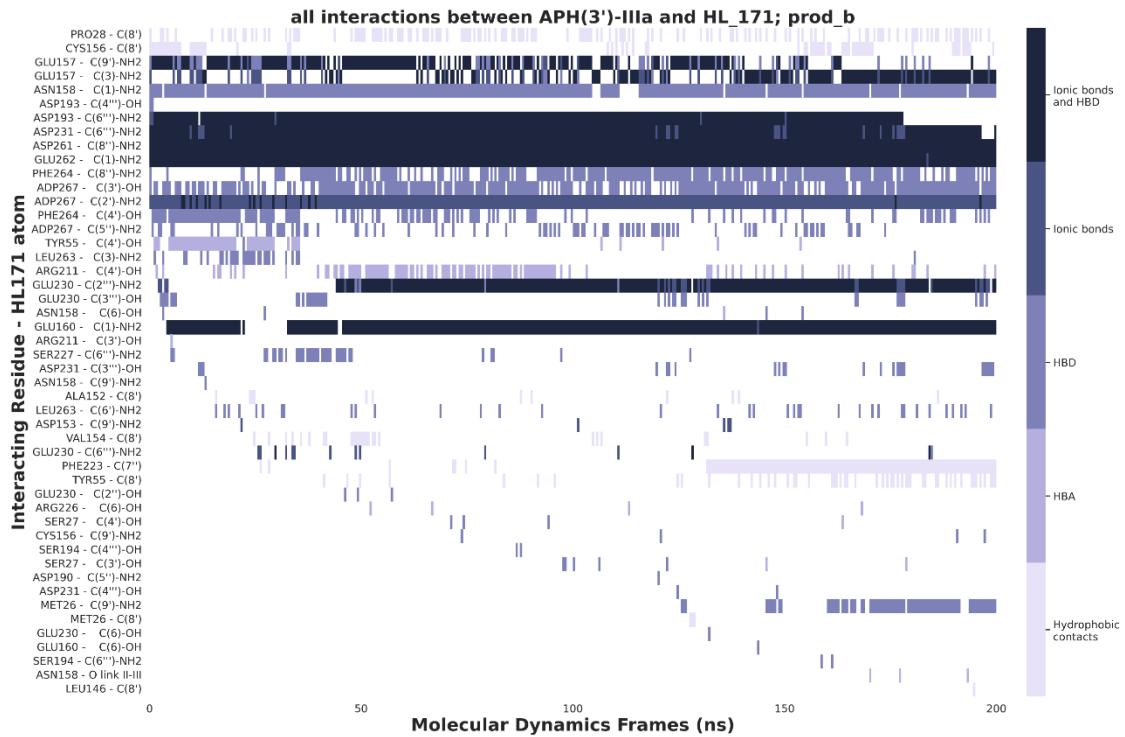
VI.1.3



VI.1.4



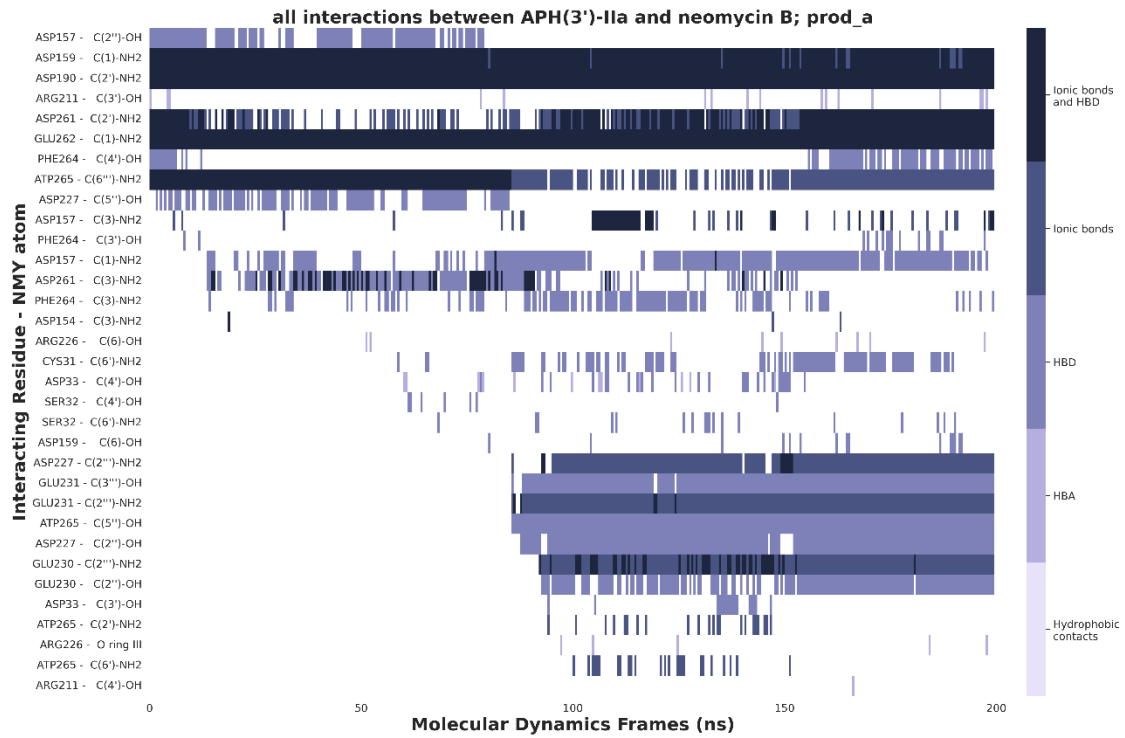
VI.1.5



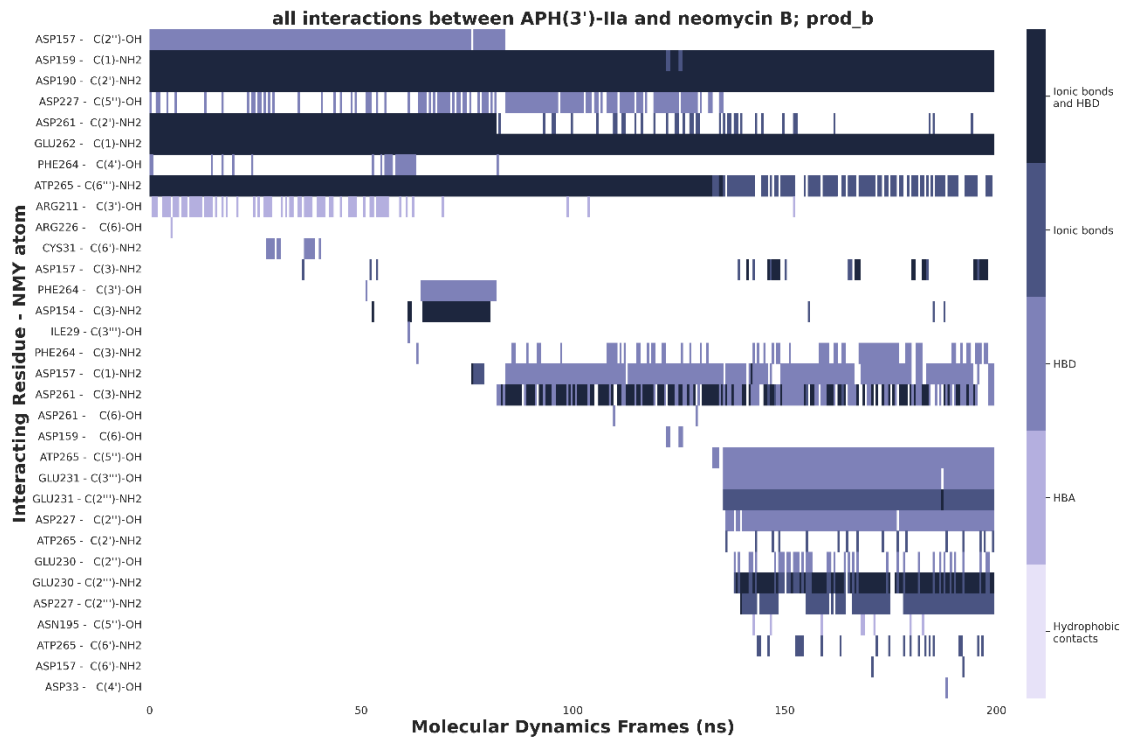
VI.1.6



VI.1.7



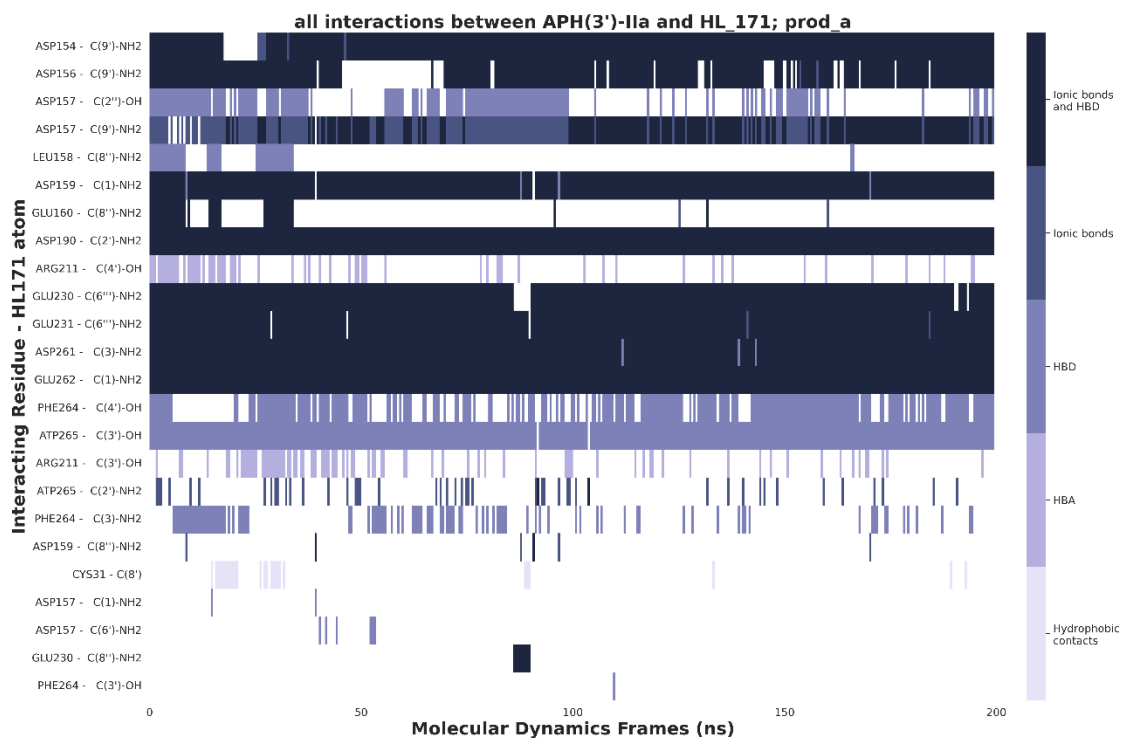
VI.1.8



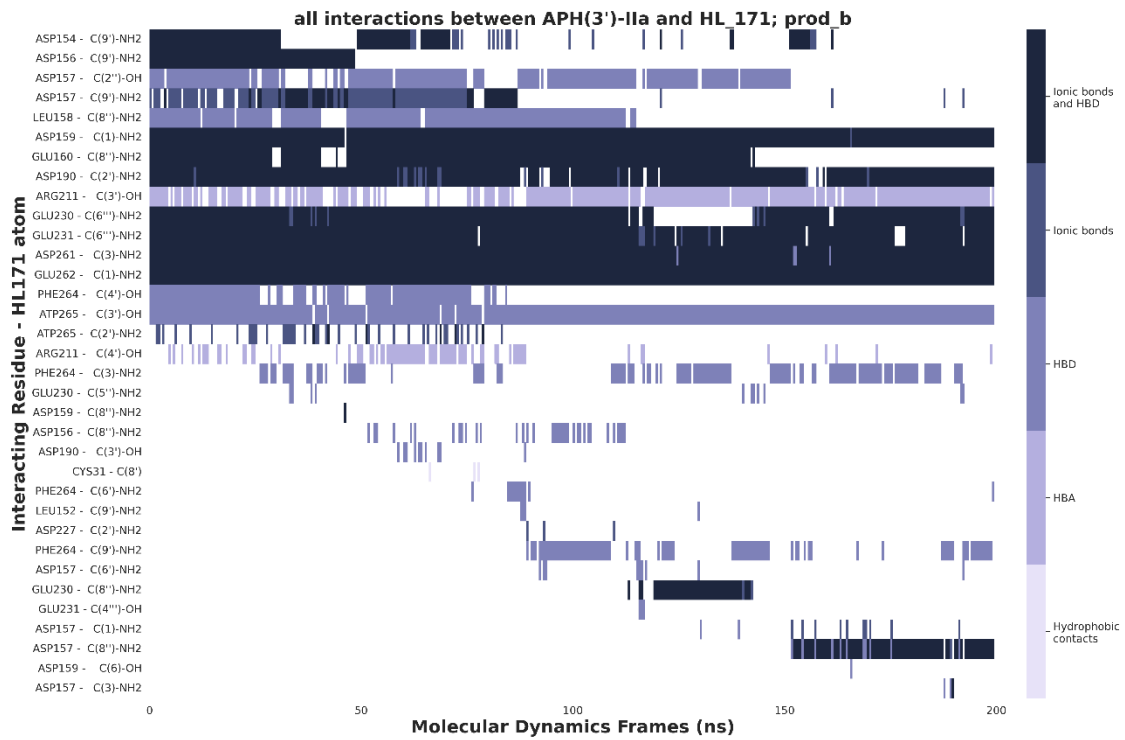
VI.1.9



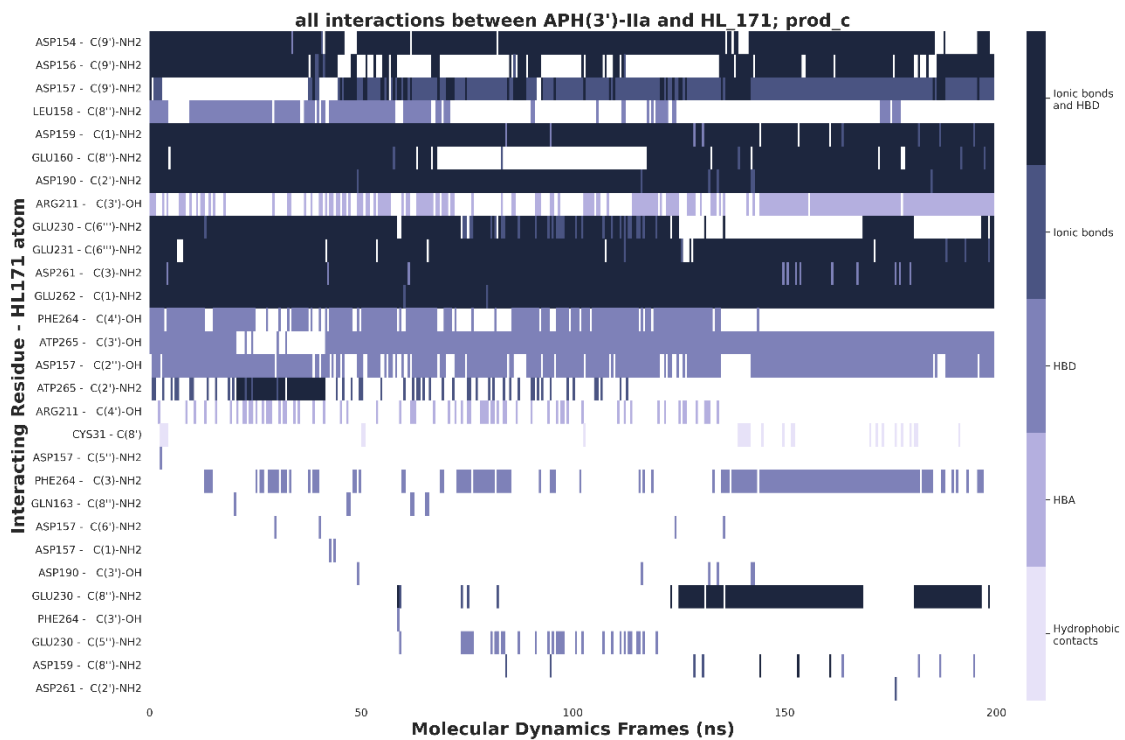
VI.1.10



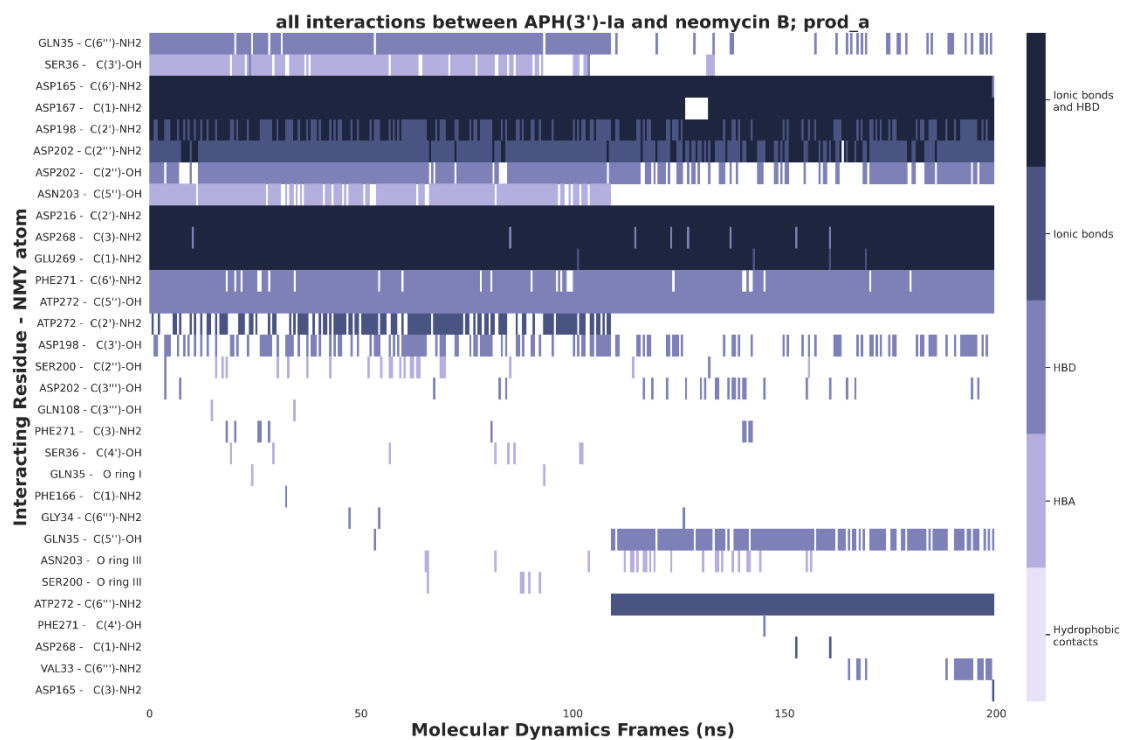
VI.1.11



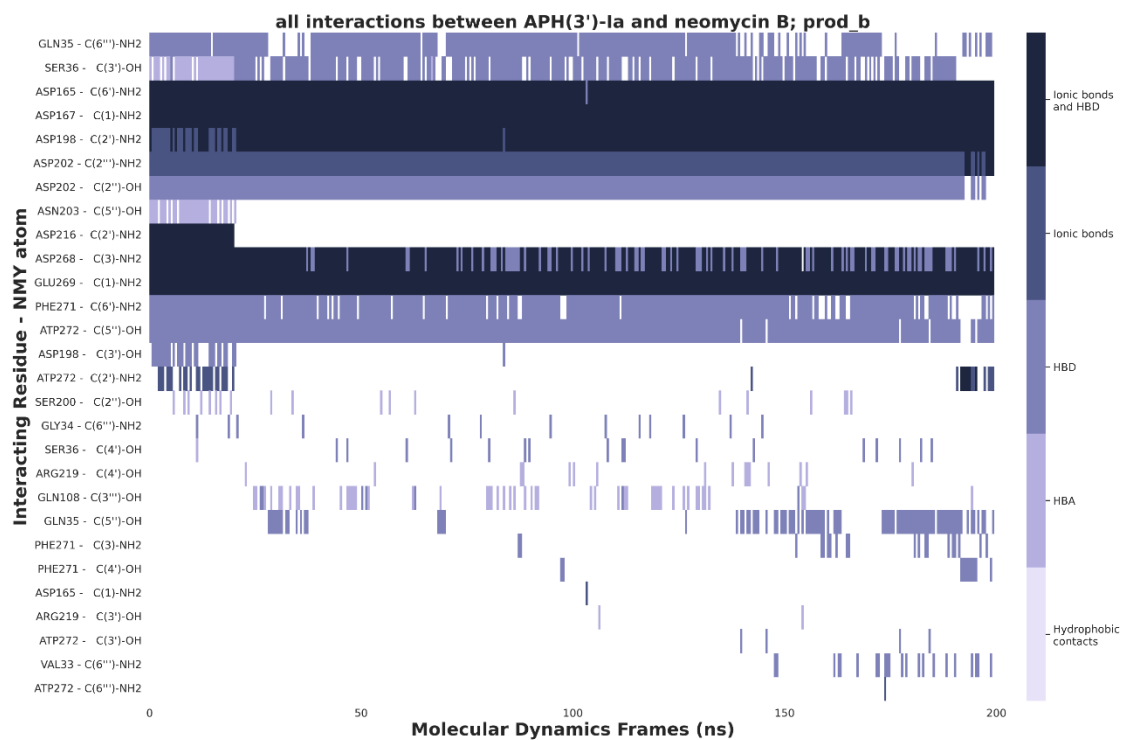
VI.1.12



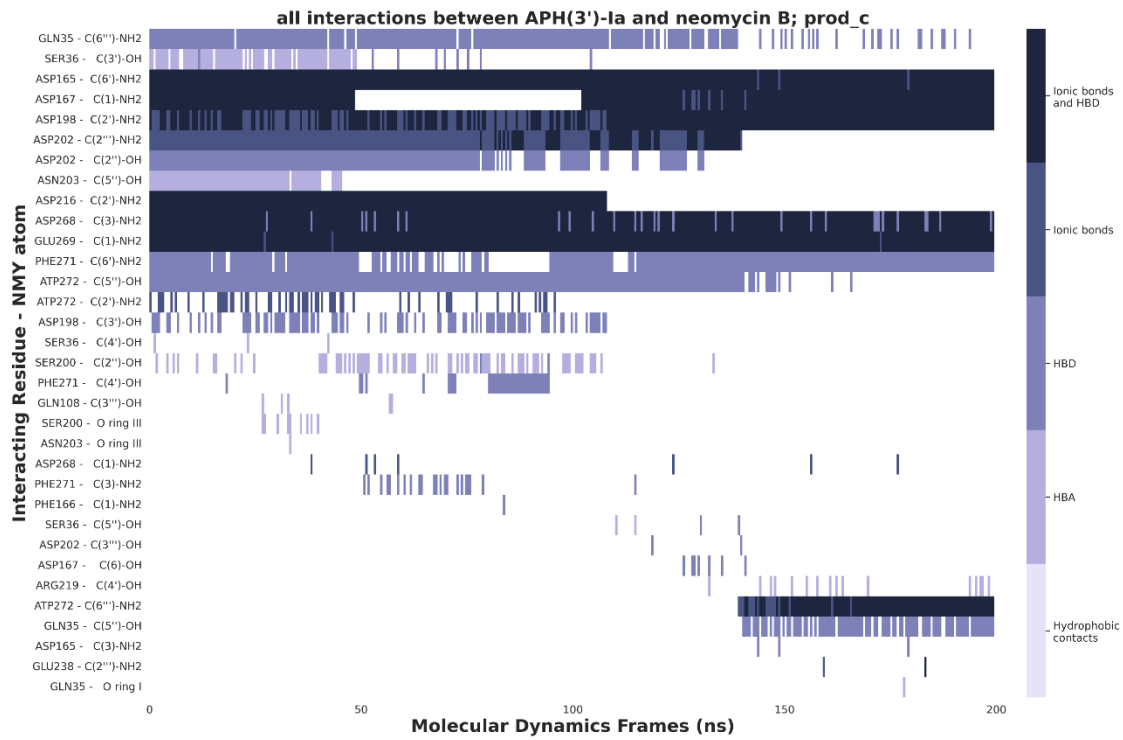
VI.1.13



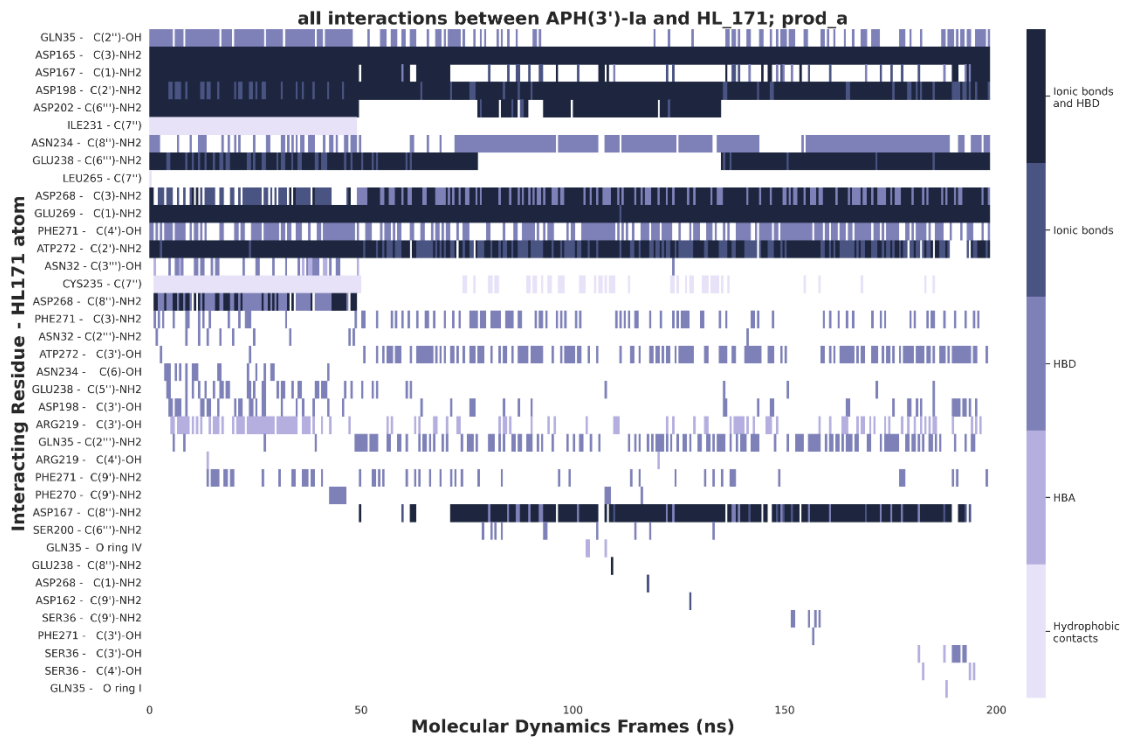
VI.1.14



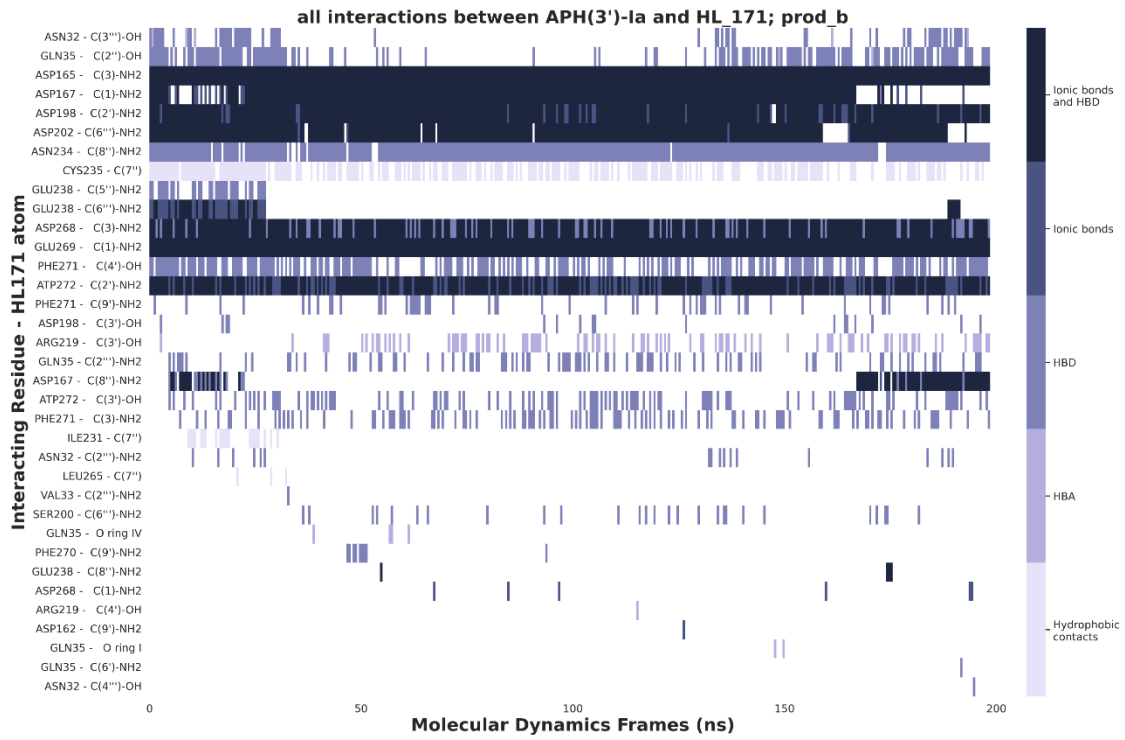
VI.1.15



VI.1.16



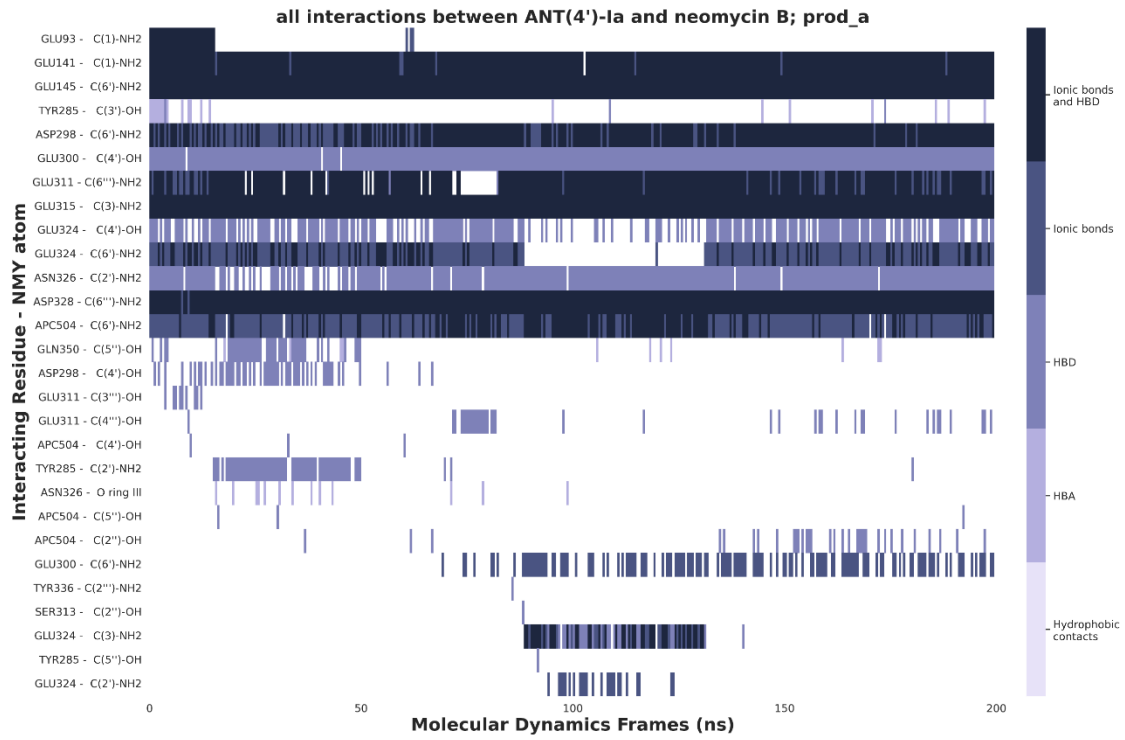
VI.1.17



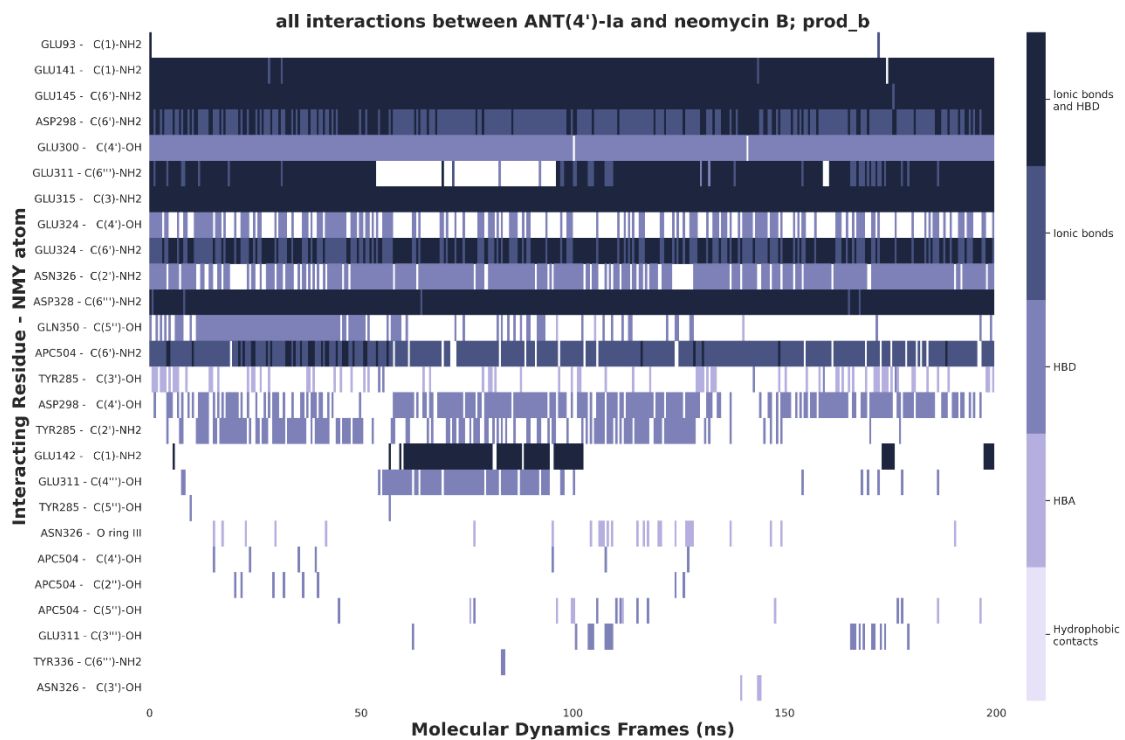
VI.1.18



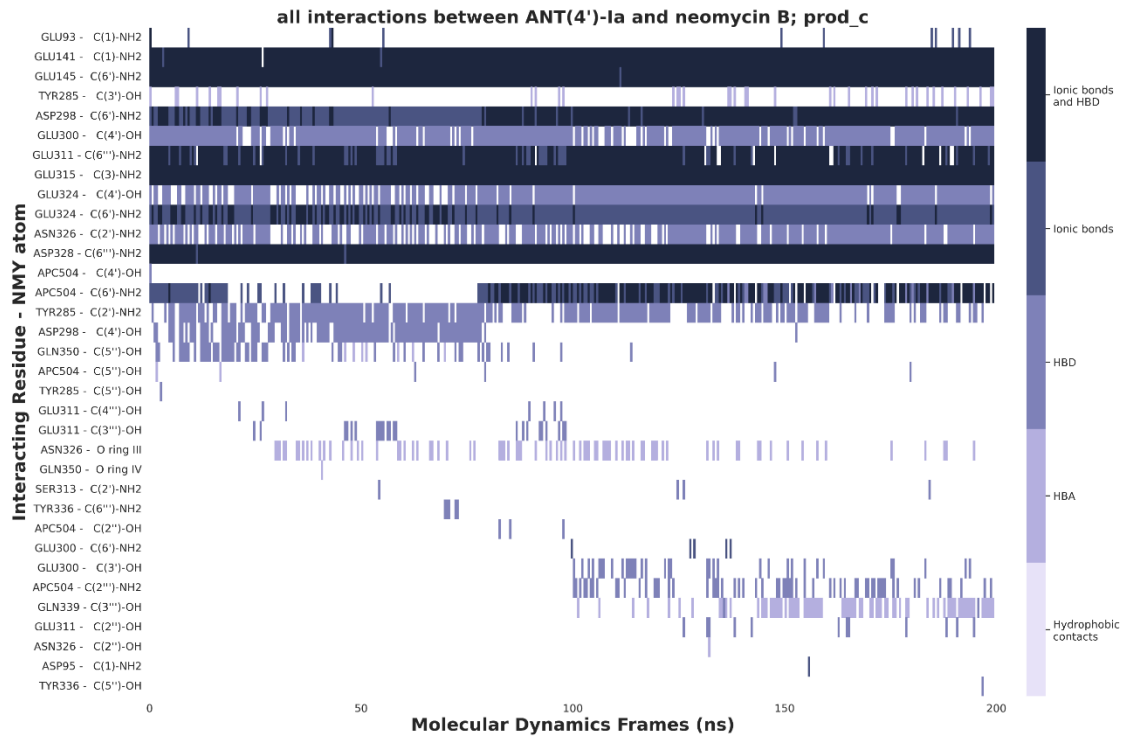
VI.1.19



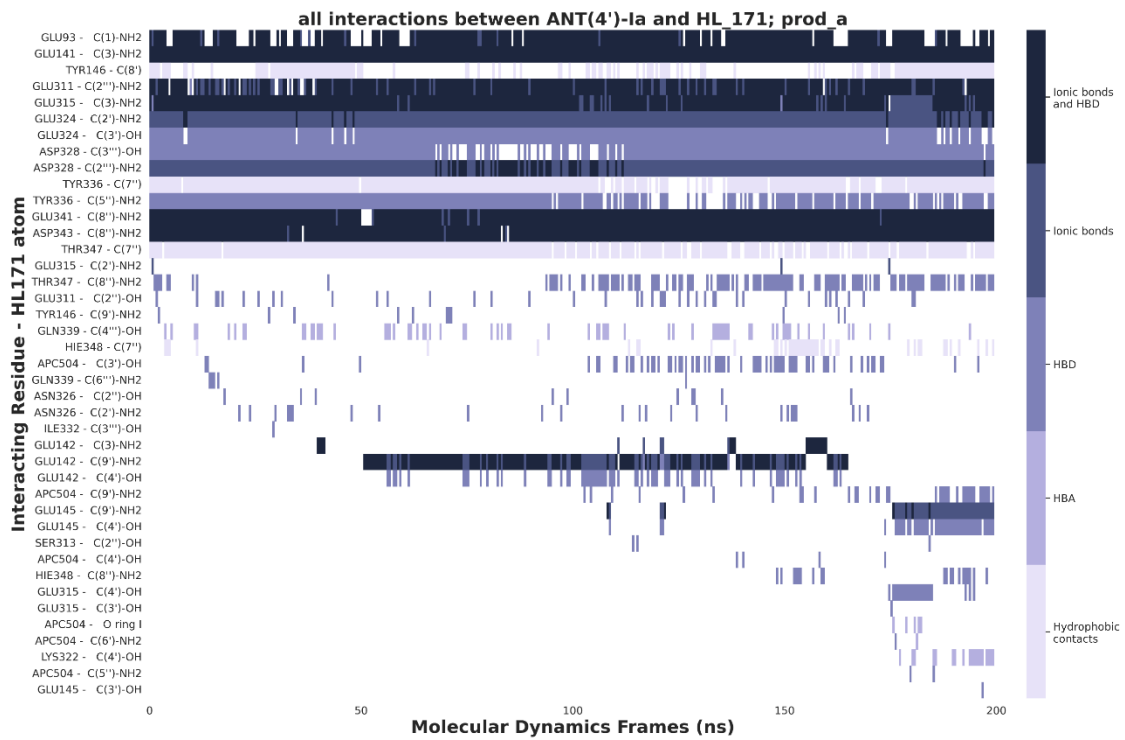
VI.1.20



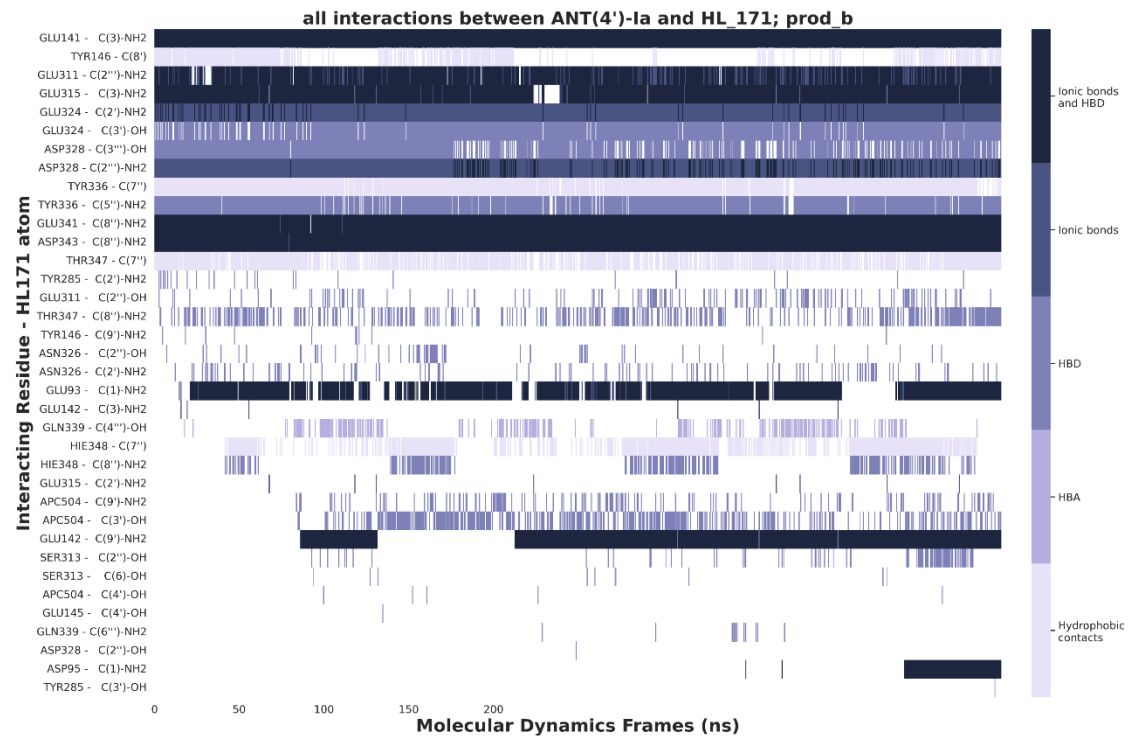
VI.1.21



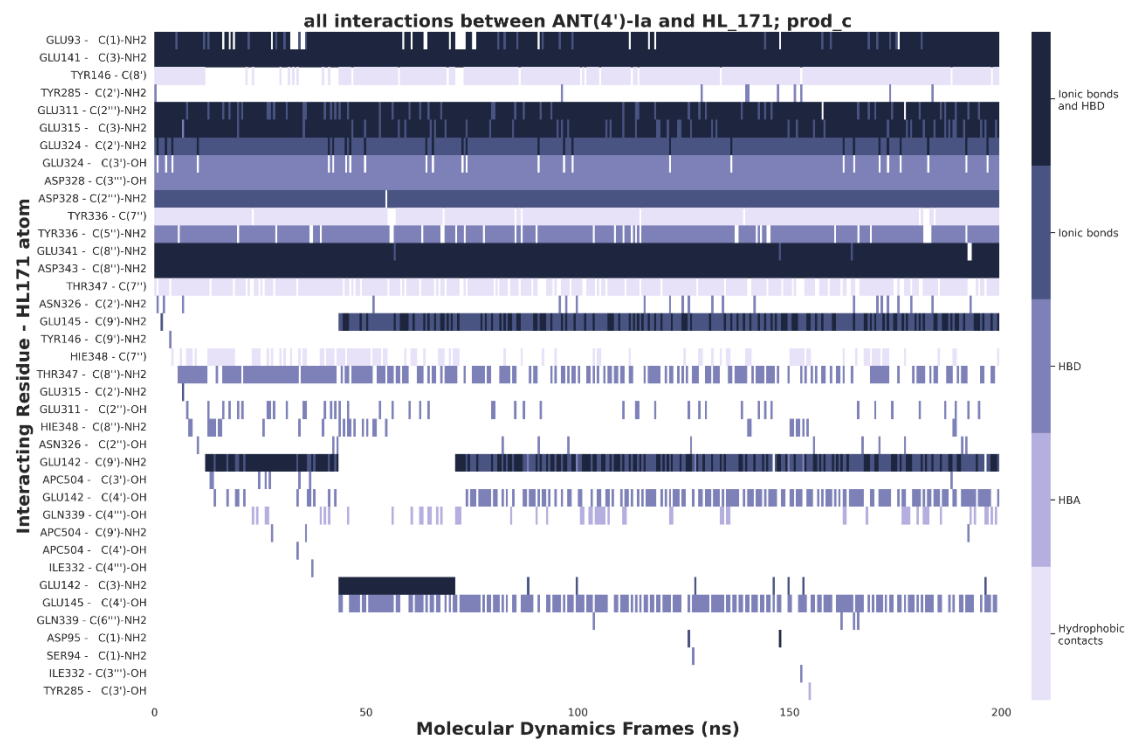
VI.1.22



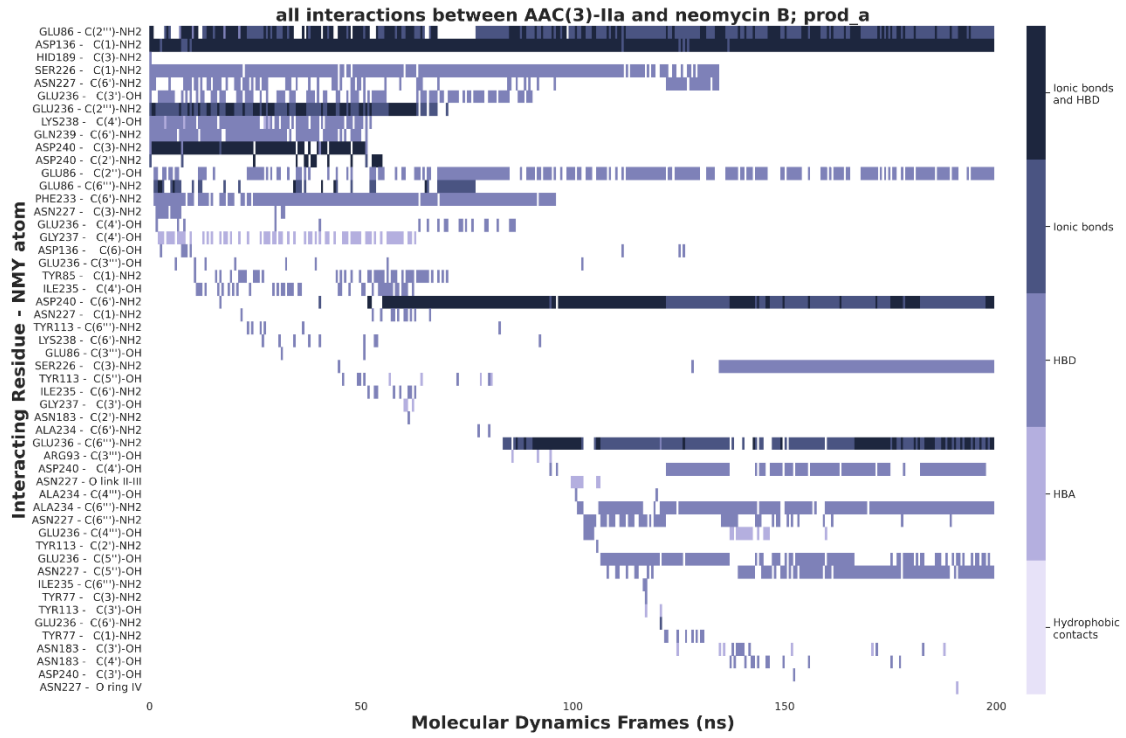
VI.1.23



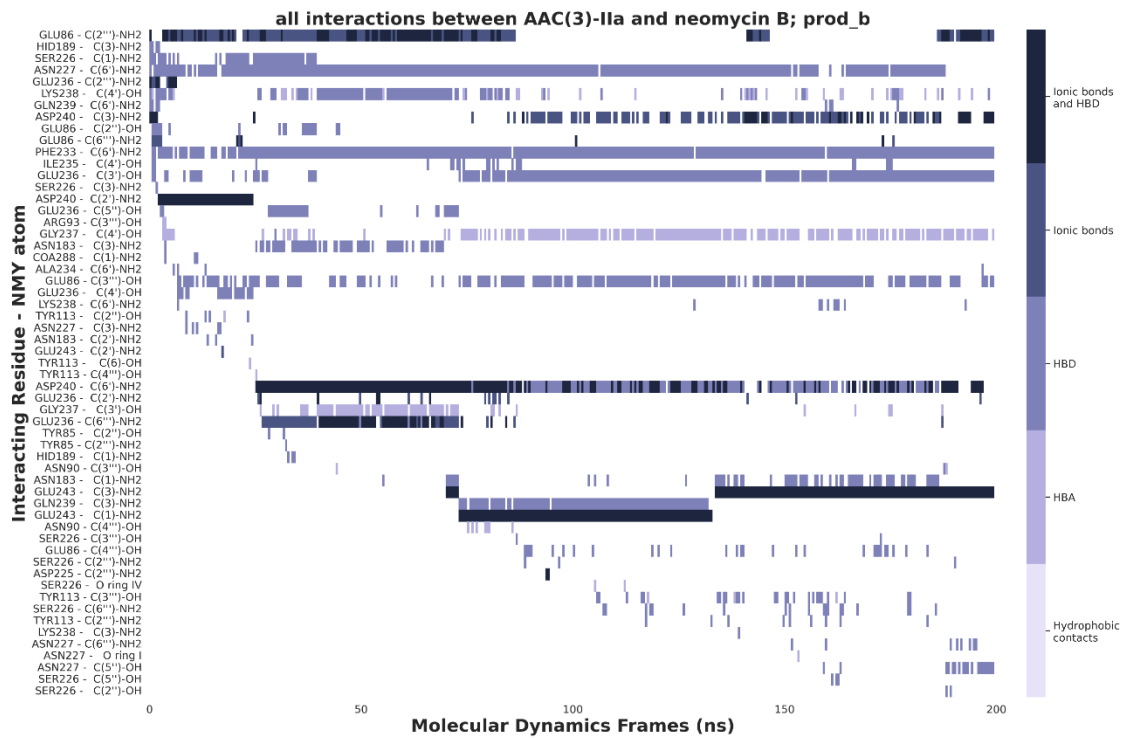
VI.1.24



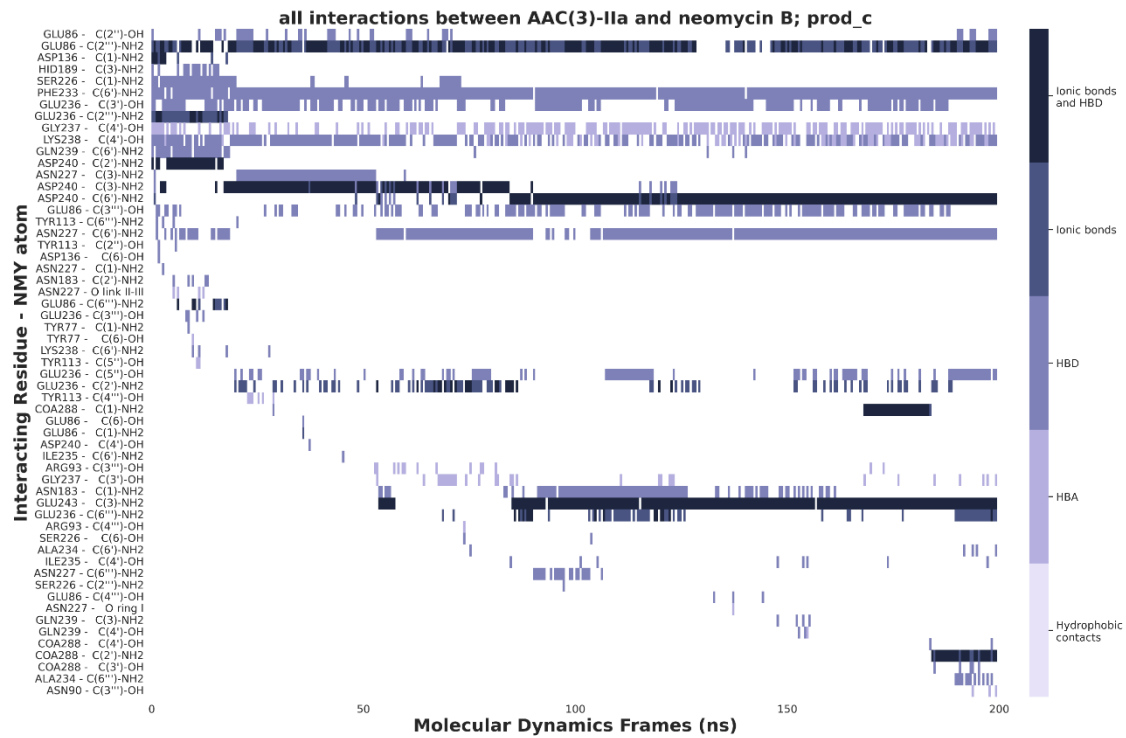
VI.1.25



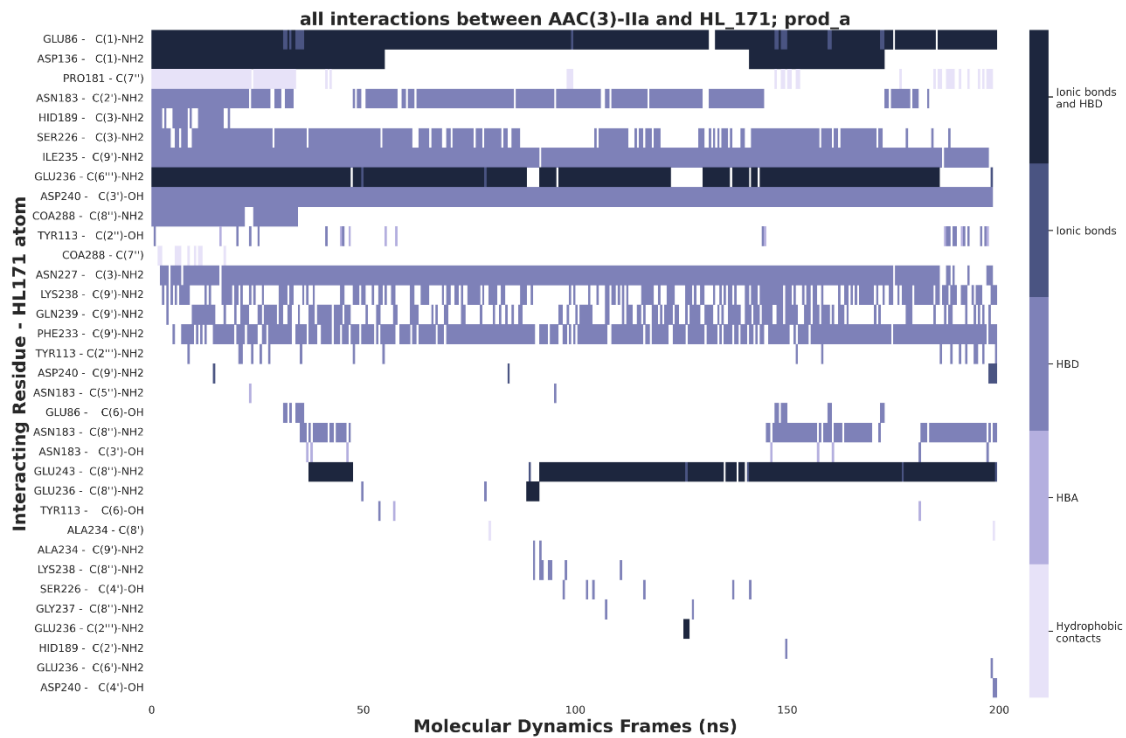
VI.1.26



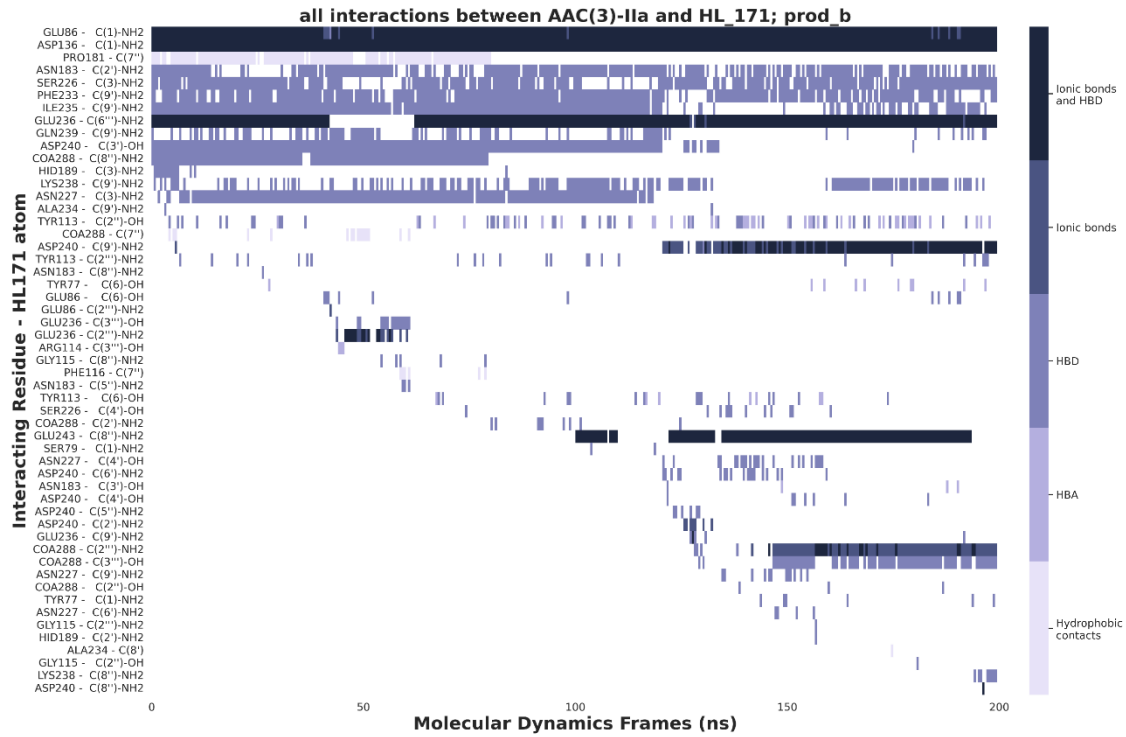
VI.1.27



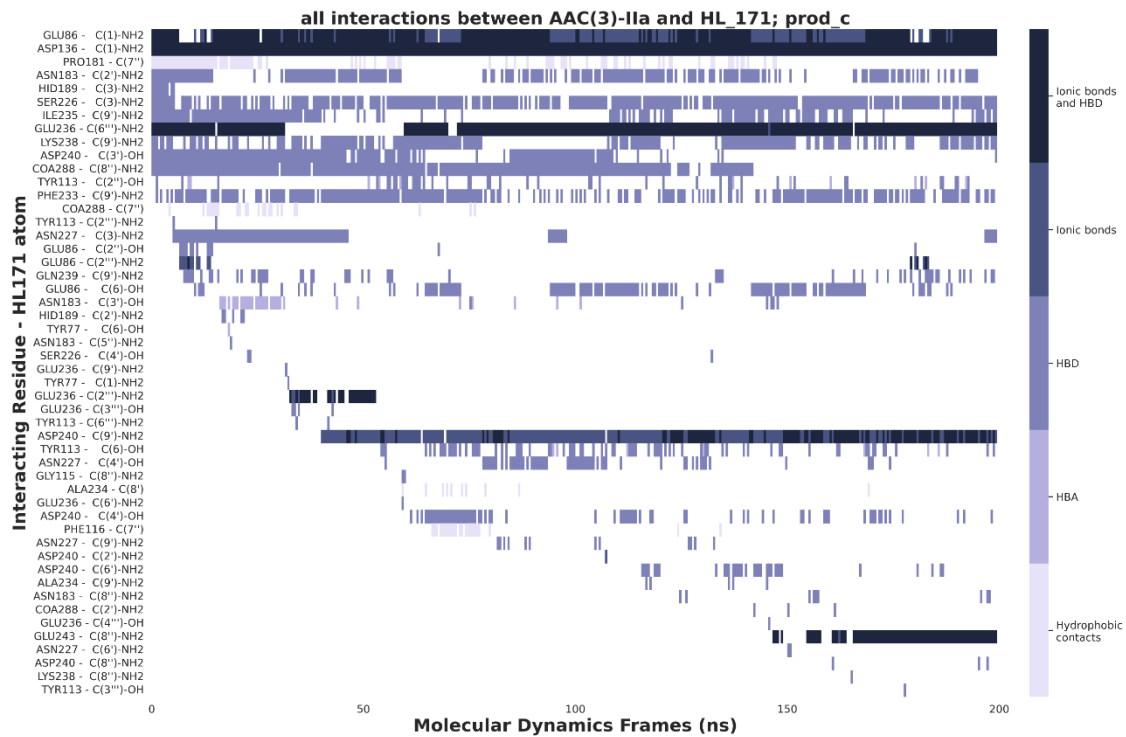
VI.1.28



VI.1.29

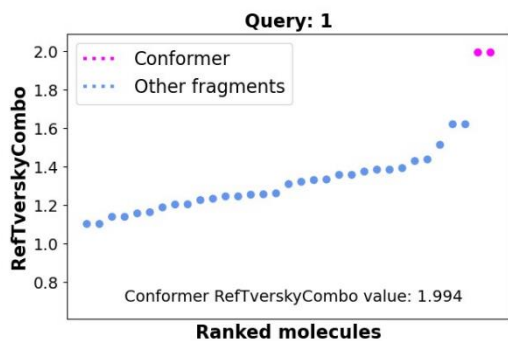


VI.1.30

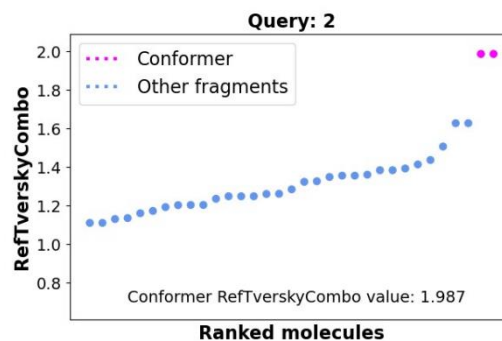


VI.2 Results of ROCS test

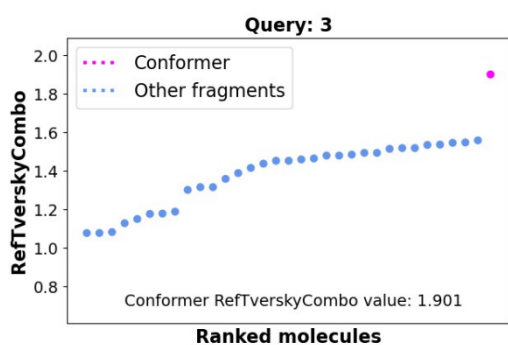
VI.2.1



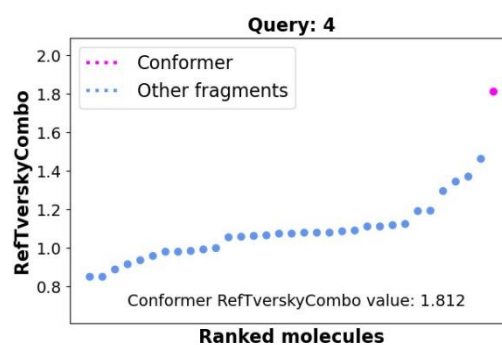
VI.2.2



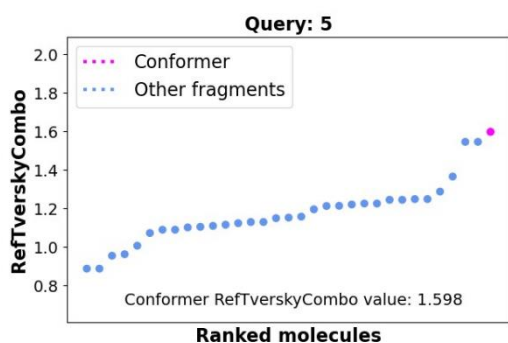
VI.2.3



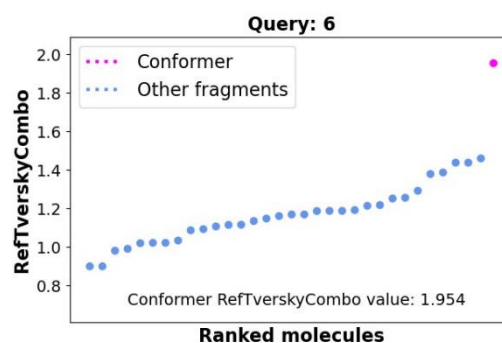
VI.2.4



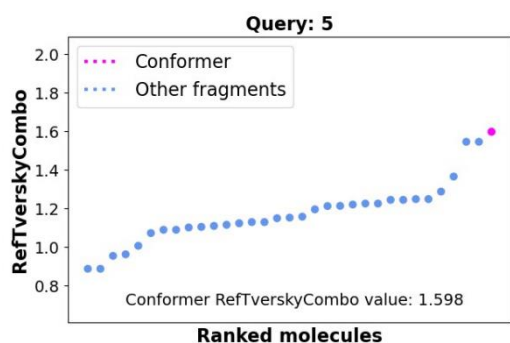
VI.2.5



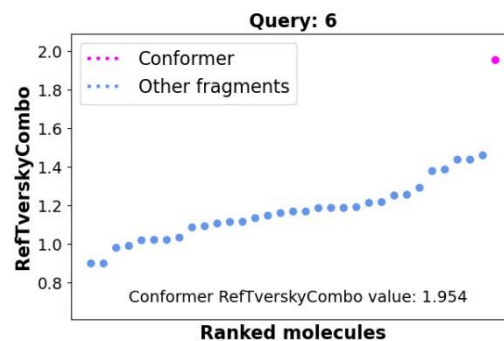
VI.2.6



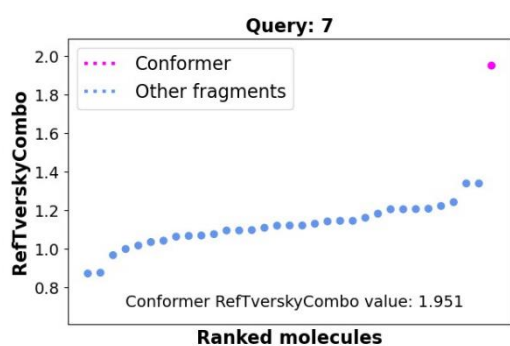
VI.2.7



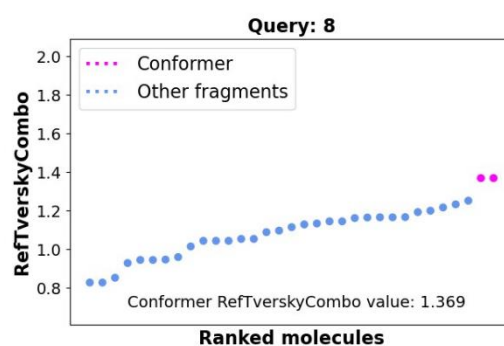
VI.2.8



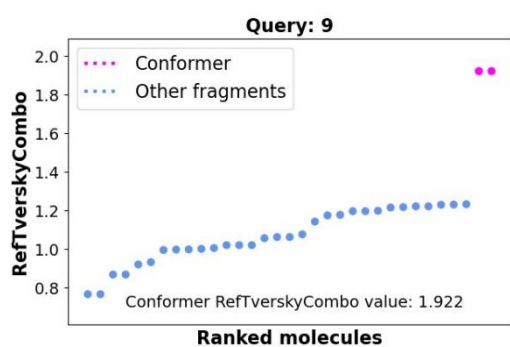
VI.2.9



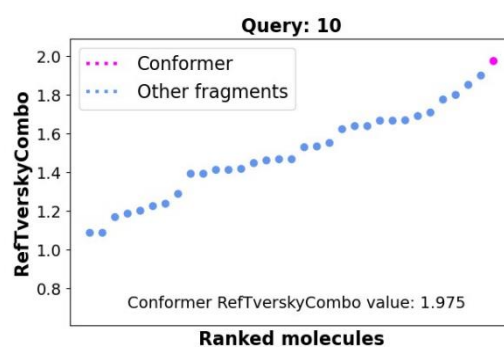
VI.2.10



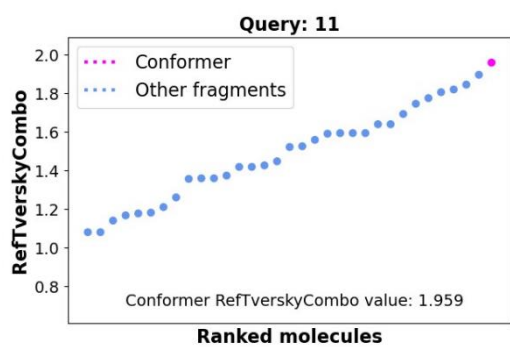
VI.2.11



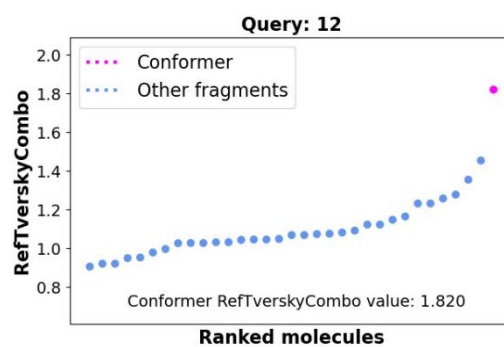
VI.2.12



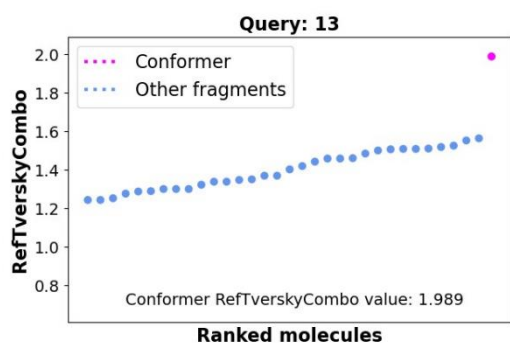
VI.2.13



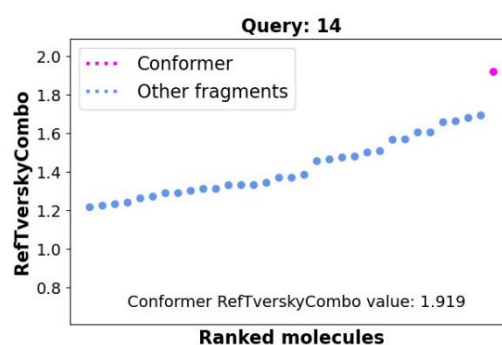
VI.2.14



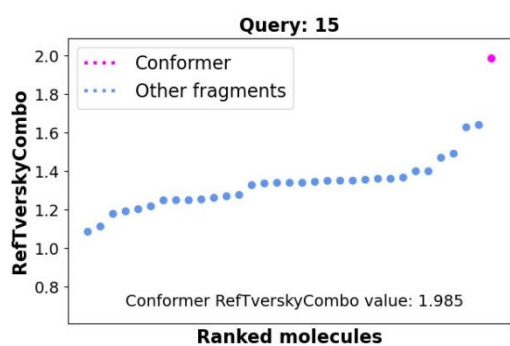
VI.2.15



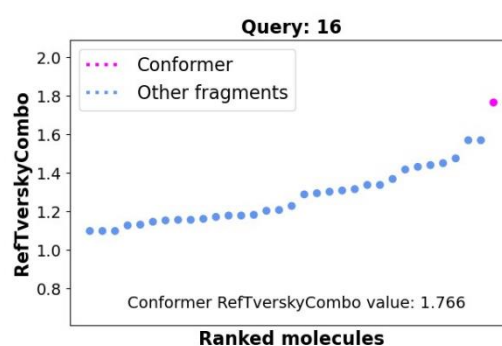
VI.2.16



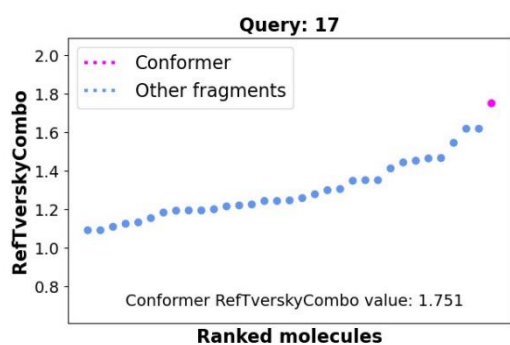
VI.2.17



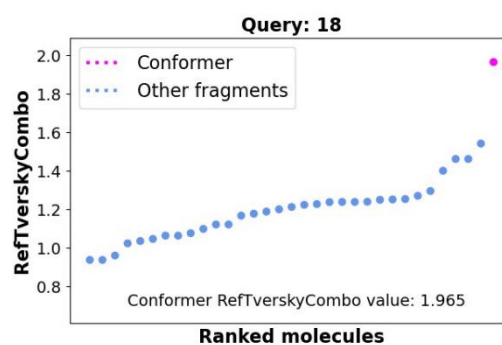
VI.2.18



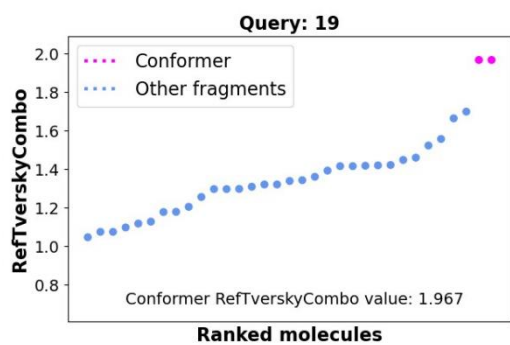
VI.2.19



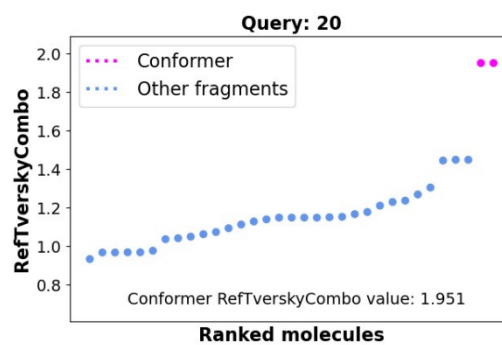
VI.2.20



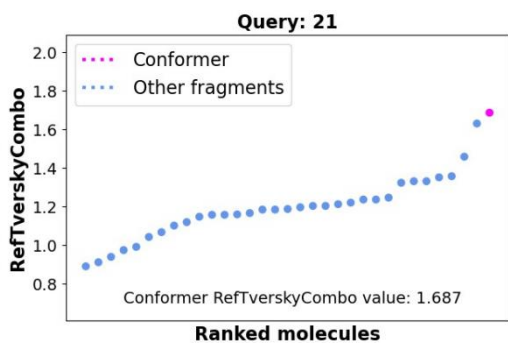
VI.2.21



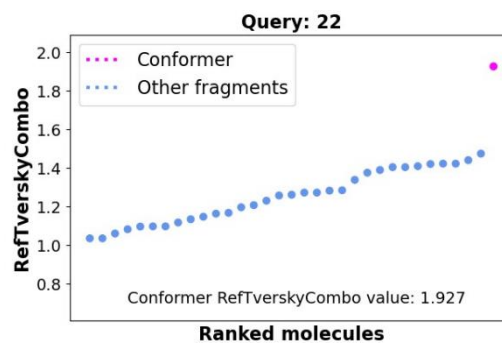
VI.2.22



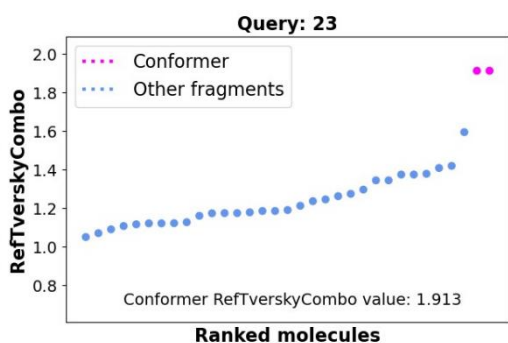
VI.2.23



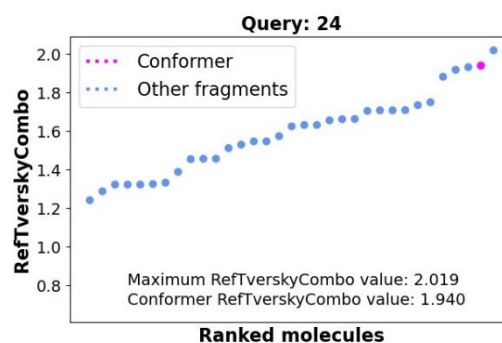
VI.2.24



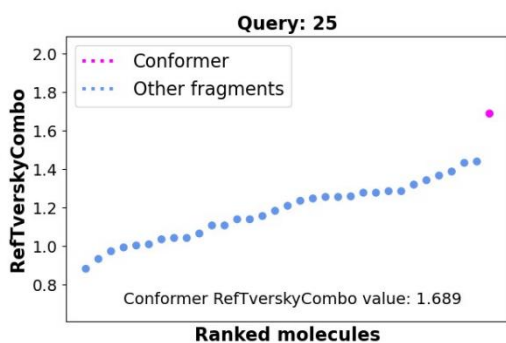
VI.2.25



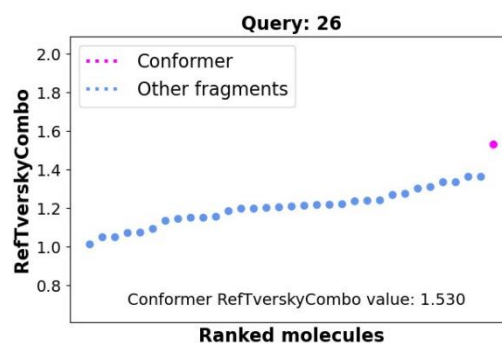
VI.2.26



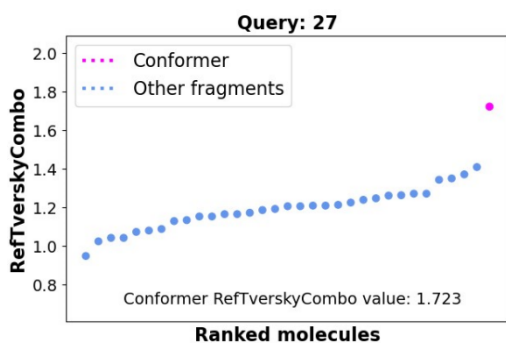
VI.2.27



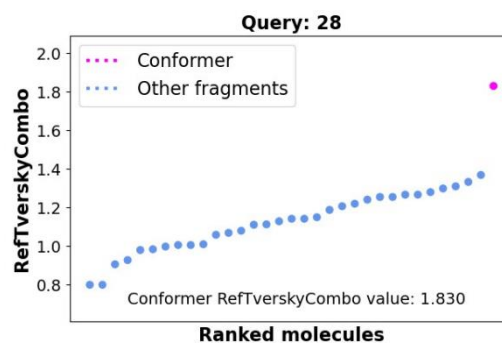
VI.2.28



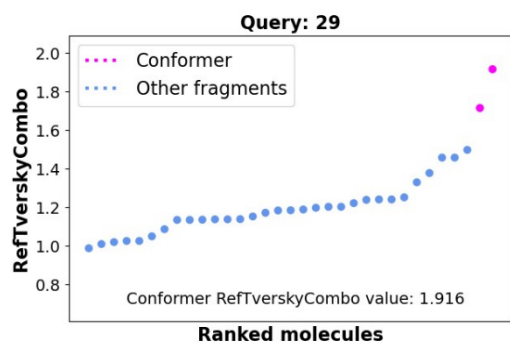
VI.2.29



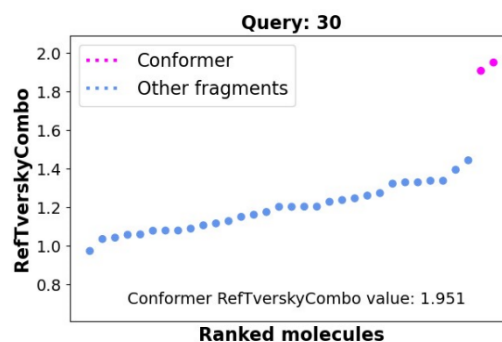
VI.2.30



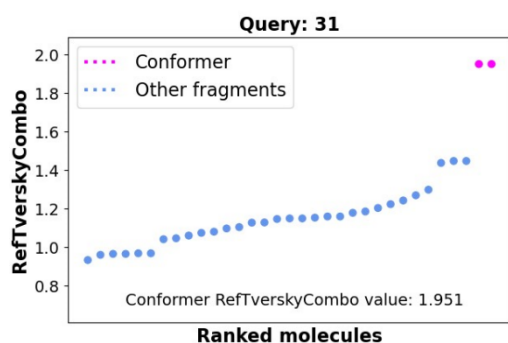
VI.2.31



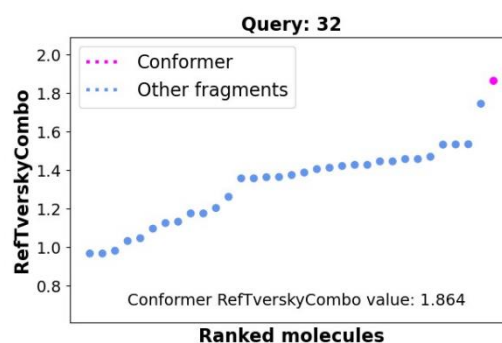
VI.2.32



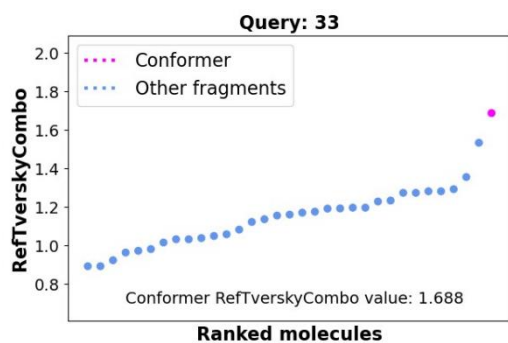
VI.2.33



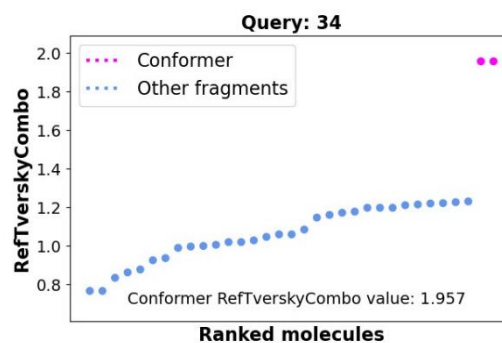
VI.2.34



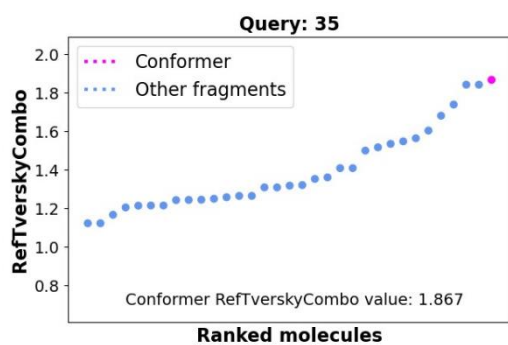
VI.2.35



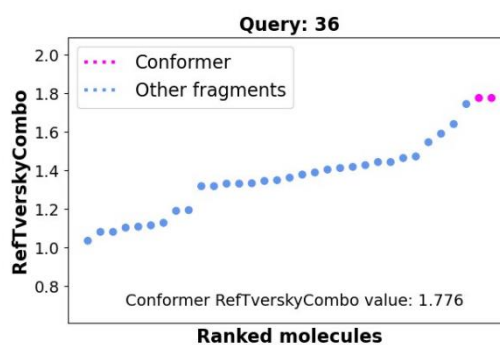
VI.2.36



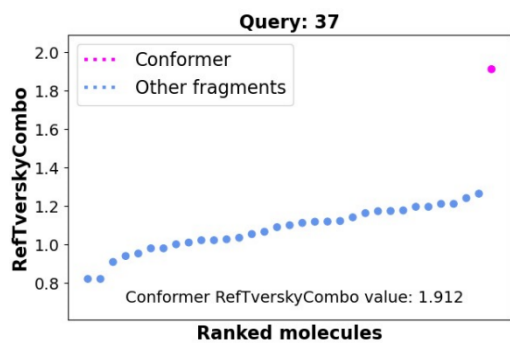
VI.2.37



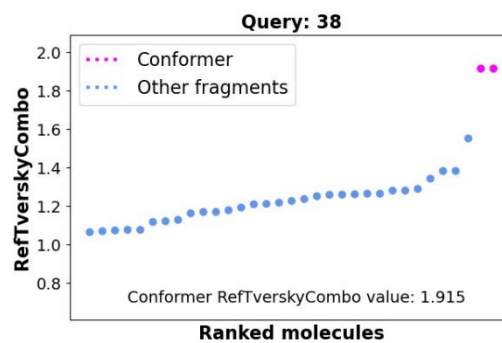
VI.2.38



VI.2.39



VI.2.40



VI.3 SI Comprehensive analysis of commercial fragment libraries

Table S1: List and annotation of analyzed commercial libraries

Supplier	Chemical Libraries		
	Small ($\leq 2\ 000$)	Intermediate	Large ($>10\ 000$)
General	<p>BIONET 2nd Generation Premium Fragment Library – KeyOrganics * Purity > 90% 1 166 fragments on 22/02/2021</p> <p>Drug-Fragment Library – Prestwick * Purity $\geq 95\%$ 1 463 fragments on 22/02/2021</p>	<p>Chembridge Fragment Library Part 2 – Chembridge * Purity $\geq 90\%$ 2 942 fragments on 22/02/2021</p> <p>Fragment-Based Library – Timtec * 3 032 fragments on 22/02/2021</p> <p>All Purpose Fragment Library – FCH Group * Purity $\geq 90\%$ 8 969 fragments on 24/02/2021</p>	<p>Fragments Library – ChemDiv * Purity $\geq 92\%$ 11 269 fragments on 24/02/2021</p> <p>General Fragment Library – OTAVAChemicals * Purity $\geq 95\%$ 11 726 fragments on 22/02/2021</p> <p>Chembridge Fragment Library Part 1 – Chembridge * Purity $\geq 90\%$ 13 475 fragments on 22/02/2021</p> <p>Advanced Subset General Fragment Library – LifeChemicals * Purity > 90% 16 774 fragments on 22/02/2021</p> <p>Allium – VitasmLab * 18 505 fragments on 22/02/2021</p> <p>Asinex Fragments – Asinex * Purity > 90% 20 061 fragments on 22/02/2021</p> <p>Fragment Collection – Maybridge Purity $\geq 90\%$ 30 099 fragments on 29/01/2019</p> <p>General Fragment Library – LifeChemicals * Purity > 90% 49 236 fragments on 22/02/2021</p> <p>Ro3 Library – Enamine * Purity $\geq 90\%$ 172 723 fragments on 22/02/2021</p>
Diverse	<p>Fragment Diversity Set 3 – LifeChemicals * Purity > 90% 320 fragments on 22/02/2021</p> <p>Fragment Diversity Set 2 – LifeChemicals * Purity > 90% 960 fragments on 22/02/2021</p> <p>Broad Spectrum Fragment Library – Reaxense * Purity > 90% 1 226 fragments on 23/02/2021</p> <p>Fragment Diversity Set 1 – LifeChemicals * Purity > 90% 1 280 fragments on 22/02/2021</p>	<p>Ultimate Fragment Library – LifeChemicals * Purity > 90% 8 187 fragments on 22/02/2021</p>	

	<p>Selcia Fragment Library – Bioascent * Purity > 95% 1 366 fragments on 22/02/2021</p> <p>High Fidelity Fragments – Enamine * Purity ≥ 90% 1 920 fragments on 22/02/2021</p> <p>Ro3 diversity Sets – Maybridge * Purity ≥ 90% 2 000 fragments on 22/02/2021</p>		
Natural Product-Like	<p>BIONET Fragments from Nature – KeyOrganics * Purity ≥ 95% 183 fragments on 22/02/2021</p> <p>Natural Fragment Library – Reaxense * Purity > 90% 250 fragments on 23/02/2019</p>	<p>Natural Product Like Fragment Library – LifeChemicals * Purity > 90% 3 148 fragments on 22/02/2021</p> <p>Natural Product-like Fragments – Enamine * Purity ≥ 90% 4 160 fragments on 22/02/2021</p>	
3D shape	<p>Spiro Fragments – FCH Group * Purity ≥ 90% 189 fragments on 24/02/2021</p> <p>3D Shape Diverse Fragments – Enamine * Purity ≥ 90% 1 200 fragments on 22/02/2021</p> <p>Stereogenic Centers Fragment Library – OTAVACHemicals * Purity ≥ 95% 1 902 fragments on 22/02/2021</p>	<p>3D-shaped Fragment Library – LifeChemicals * Purity > 90% 3 186 fragments on 22/02/2021</p>	
sp3	<p>High Fsp3 Fragment Library – FCH Group * Purity ≥ 90% 1 975 fragments on 24/02/2021</p>	<p>High Fsp3 Fragment Library – OTAVACHemicals * Purity ≥ 95% 2 662 fragments on 22/02/2021</p> <p>Advanced Subset of Fsp3-Enriched Fragment Library – LifeChemicals * Purity > 90% 8 536 fragments on 22/02/2021</p>	<p>Fsp3 Enriched Fragment Library – LifeChemicals * Purity > 90% 17 862 fragments on 22/02/2021</p> <p>Special sp3-rich Fragment Set – Enamine * Purity ≥ 90% 50 272 fragments on 22/02/2021</p>
Metal	<p>Chelator Fragment Library – LifeChemicals * Purity > 90% 1 584 fragments on 22/02/2021</p> <p>Fragment Library Chelators 2019 subset – Reaxense * Purity > 90% 1 602 fragments on 23/02/2021</p> <p>Chelator Fragment Library – OTAVACHemicals * Purity ≥ 95% 1 774 fragments on 22/02/2021</p>		
PPI	<p>Fragment-Like Aminoacids – FCH Group Purity ≥ 90% 48 fragments on 24/02/2021</p>	<p>PPI Fragment Library – Enamine Purity ≥ 90% 3 600 fragments on 22/02/2021</p> <p>PPI Fragment Library – LifeChemicals Purity > 90%</p>	

		3 882 fragments on 22/02/2021 PPI Fragment Library – Reaxense Purity > 90% 6 298 fragments on 23/02/2021	
Fluor	<p>F19 Library – Maybridge Purity ≥ 90% 387 fragments on 22/02/2021</p> <p>BIONET Fluorine Fragment Set – KeyOrganics Purity ≥ 95% 462 fragments on 22/02/2021</p> <p>Fluorinated Fragments Diversity Set – Enamine Purity ≥ 90% 1 000 fragments on 22/02/2021</p> <p>19F NMR Fluorine containing Fragment Library – OTAVACchemicals Purity ≥ 95% 1 151 fragments on 22/02/2021</p> <p>Fluorine Fragment Cocktail – LifeChemicals Purity > 90% 1 340 fragments on 22/02/2021</p> <p>Advanced Subset of Fluorine Fragment Library – LifeChemicals Purity > 90% 1 762 fragments on 22/02/2021</p> <p>Fluorine Fragment Library – FCH Group Purity ≥ 90% 1 821 fragments on 24/02/2021</p>	<p>Chembridge Fragment Fluor Library – Chembridge Purity ≥ 90% 2 321 fragments on 22/02/2021</p> <p>Fluorine Fragment Library – LifeChemicals Purity > 90% 5 808 fragments on 22/02/2021</p> <p>Fluorinated Fragments – Enamine Purity ≥ 90% 6 377 fragments on 22/02/2021</p>	Fluoro Fragments – Maybridge Purity ≥ 90% 27 045 fragments on 22/02/2021
Halogen	<p>BIONET Bromine Fragments – KeyOrganics Purity > 90% 275 fragments on 22/02/2021</p> <p>Halogen-Enriched Fragment Library – OTAVACchemicals Purity ≥ 95% 596 fragments on 22/02/2021</p> <p>Advanced Subset of Bromine – LifeChemicals Purity > 90% 1 314 fragments on 22/02/2021</p>	<p>Bromine Fragment Library – Life Chemicals Purity > 90% 2 031 fragments on 22/02/2021</p> <p>Chembridge Fragment Bromo Library – Chembridge Purity ≥ 90% 2 378 fragments on 22/02/2021</p> <p>Halogen-Enriched Fragment Library – Reaxense Purity > 90% 2 659 fragments on 23/02/2021</p> <p>Bromo Fragments – Maybridge Purity ≥ 90% 7 270 fragments on 22/02/2021</p>	
Covalent	<p>Specific Covalent Fragments – LifeChemicals Purity > 90% 598 fragments on 22/02/2021</p> <p>BIONET Covalent Fragments – KeyOrganics Purity > 90% 687 fragments on 22/02/2021</p> <p>Serine Focused Covalent Fragments – Enamine</p>	<p>Covalent Fragment Library – Reaxense Purity > 90% 2 506 fragments on 23/02/2021</p> <p>Advance Subset of Covalent Fragment Library – LifeChemicals Purity > 90% 2 851 fragments on 22/02/2021</p> <p>Cystein Focused Covalent Fragments – Enamine Purity ≥ 90%</p>	

	<p>Purity \geq 90% 1 280 fragments on 22/02/2021</p> <p>Lysine Focused Covalent Fragments – Enamine Purity \geq 90% 1 600 fragments on 22/02/2021</p>	<p>3 200 fragments on 22/02/2021</p> <p>Covalent Fragment Library – LifeChemicals Purity > 90% 4 077 fragments on 22/02/2021</p> <p>Covalent Fragments Library Plated – Enamine Purity \geq 90% 6 120 fragments on 22/02/2021</p>	
Miscellaneous	<p>Mini-Fragment Library – Enamine * Purity \geq 90% 80 fragments on 22/02/2021</p> <p>BIONET MiniFrag – KeyOrganics * Purity > 90% 81 fragments on 22/02/2021</p> <p>Essential Fragments – Enamine * Purity \geq 90% 320 fragments on 22/02/2021</p> <p>Boronic Acids Fragments – KeyOrganics Purity > 90% 514 fragments on 22/02/2021</p> <p>Special Selection Fragment Library – FCH Group * Purity \geq 90% 864 fragments on 24/02/21</p> <p>DSI-POISED Library Plated – Enamine * Purity \geq 90% 896 fragments on 22/02/2021</p> <p>Solubility Fragment Library – OTAVACHemicals * Purity \geq 95% 957 fragments on 22/02/2021</p> <p>Single Pharmacophore Fragments – Enamine * Purity \geq 90% 1 500 fragments on 22/02/2021</p>	<p>Fragment-Like Amines – FCH Group Purity \geq 90% 2 155 fragments on 24/02/2021</p> <p>Ro3 Compliant for CherryPicking – Maybridge * Purity \geq 95% 2 733 fragments on 22/02/2021</p> <p>Fragment-Like Acids – FCH Group Purity \geq 90% 2 898 fragments on 24/02/2021</p> <p>High Solubility Fragment Subset – LifeChemicals * Purity > 90% 3 783 fragments on 22/02/2021</p> <p>Carboxylic Acids Fragments – Enamine Purity \geq 90% 4 000 fragments on 22/02/2021</p> <p>Advanced Subset of Low MW Fragment Library – LifeChemicals * Purity > 90% 4 949 fragments on 22/02/2021</p> <p>Low MW Fragment Library – LifeChemicals * Purity > 90% 6 323 fragments on 22/02/2021</p>	<p>Fragment Library with Experimental Solubility Data – LifeChemicals * Purity > 90% 18 027 fragments on 22/02/2021</p> <p>Tangible Fragments – LifeChemicals * Purity > 90% 123 026 fragments on 22/02/2021</p>

* Non-chemically biased libraries.

Purity has been determined by HPLC and/or ¹H-NMR

Table S2: Summary of non-chemically biased commercial libraries properties. ^aThe number of scaffolds which are present in a single copy in the full set of unique fragments is indicated in brackets

Type	Index of library	Number of fragments	Number of scaffolds ^a	MW (Mean ± SD)	logP (Mean ± SD)	PBF (Mean ± SD)	QED (Mean ± SD)
GEN	1	1 163	339 (15)	176.57 ± 34.57	1.06 ± 1.05	0.29 ± 0.27	0.66 ± 0.10
	2	1 463	340 (50)	195.17 ± 48.02	1.00 ± 1.46	0.42 ± 0.29	0.66 ± 0.13
	3	2 936	1 565 (867)	256.25 ± 24.66	0.59 ± 1.25	0.65 ± 0.19	0.78 ± 0.09
	4	3 032	974 (211)	217.05 ± 40.88	0.88 ± 1.30	0.38 ± 0.28	0.70 ± 0.12
	5	8 968	1 986 (307)	245.78 ± 42.68	0.35 ± 1.12	0.62 ± 0.26	0.76 ± 0.11
	6	11 268	2 387 (631)	250.27 ± 38.54	1.56 ± 1.23	0.45 ± 0.29	0.77 ± 0.12
	7	11 724	1 349 (364)	220.60 ± 45.35	1.38 ± 1.18	0.40 ± 0.30	0.72 ± 0.12
	8	13 475	1 910 (278)	244.33 ± 48.98	1.48 ± 1.29	0.50 ± 0.26	0.75 ± 0.12
	9	16 774	3 779 (837)	226.46 ± 45.58	0.88 ± 1.20	0.51 ± 0.25	0.74 ± 0.11
	10	18 473	3 431 (1 191)	236.18 ± 46.49	1.83 ± 1.20	0.43 ± 0.28	0.73 ± 0.13
	11	20 061	4 656 (1 722)	252.27 ± 36.25	0.85 ± 1.43	0.54 ± 0.27	0.75 ± 0.12
	12	49 234	11 804 (5 454)	253.56 ± 42.32	1.12 ± 1.22	0.54 ± 0.26	0.76 ± 0.12
	13	172 631	29 672 (16 553)	246.45 ± 41.48	0.83 ± 1.28	0.60 ± 0.25	0.77 ± 0.10
DIV	1	320	247 (20)	196.12 ± 46.68	0.84 ± 1.31	0.40 ± 0.28	0.69 ± 0.12
	2	960	602 (48)	211.91 ± 45.21	0.98 ± 1.16	0.46 ± 0.26	0.72 ± 0.11
	3	1 226	501 (59)	224.58 ± 45.12	1.31 ± 1.22	0.42 ± 0.32	0.73 ± 0.11
	4	1 280	776 (68)	207.96 ± 46.10	0.94 ± 1.20	0.44 ± 0.27	0.72 ± 0.12
	5	1 366	288 (46)	169.86 ± 32.58	0.70 ± 1.23	0.26 ± 0.27	0.60 ± 0.12
	6	1 920	620 (28)	176.40 ± 26.89	0.25 ± 1.06	0.45 ± 0.23	0.65 ± 0.09
	7	2 000	399 (28)	180.00 ± 38.92	1.25 ± 1.13	0.32 ± 0.27	0.66 ± 0.11
	8	8 187	2 153 (299)	219.22 ± 38.38	0.80 ± 1.16	0.52 ± 0.23	0.74 ± 0.10
NP	1	183	83 (23)	249.04 ± 35.80	-1.14 ± 1.63	0.66 ± 0.17	0.64 ± 0.14
	2	249	96 (12)	195.97 ± 29.63	0.43 ± 1.28	0.46 ± 0.27	0.60 ± 0.12
	3	3 148	512 (149)	253.07 ± 41.92	1.09 ± 1.21	0.51 ± 0.28	0.74 ± 0.11
	4	4 152	1 307 (382)	229.24 ± 52.72	0.77 ± 1.34	0.60 ± 0.26	0.72 ± 0.11
3D	1	189	121 (30)	264.15 ± 29.44	0.35 ± 1.17	0.73 ± 0.16	0.76 ± 0.10
	2	1 199	557 (53)	209.25 ± 44.52	0.41 ± 1.29	0.70 ± 0.19	0.69 ± 0.14
	3	1 894	473 (86)	232.46 ± 43.17	0.97 ± 1.27	0.63 ± 0.21	0.73 ± 0.12
	4	3 186	1 182 (272)	237.02 ± 45.39	0.65 ± 1.32	0.64 ± 0.16	0.73 ± 0.12
SP3	1	1 975	927 (298)	231.67 ± 48.88	-0.22 ± 1.30	0.67 ± 0.18	0.69 ± 0.12
	2	2 660	703 (128)	223.23 ± 47.26	0.87 ± 1.37	0.58 ± 0.24	0.71 ± 0.13
	3	8 536	2 333 (503)	224.20 ± 46.17	0.65 ± 1.32	0.60 ± 0.17	0.72 ± 0.12
	4	17 862	4 759 (1 850)	247.24 ± 44.58	0.71 ± 1.28	0.61 ± 0.19	0.75 ± 0.12
	5	50 268	612 721 (6 090)	257.61 ± 25.74	0.34 ± 1.37	0.69 ± 0.18	0.76 ± 0.10
CHE	1	1 584	449 (48)	207.36 ± 46.58	0.64 ± 1.31	0.38 ± 0.30	0.70 ± 0.12
	2	1 602	429 (28)	200.97 ± 44.30	0.36 ± 1.33	0.38 ± 0.29	0.68 ± 0.12
	3	1 774	419 (51)	218.78 ± 46.34	0.96 ± 1.24	0.32 ± 0.28	0.71 ± 0.13
MIS	1	80	52 (21)	94.05 ± 15.66	-0.64 ± 1.10	0.15 ± 0.19	0.43 ± 0.06
	2	81	53 (21)	94.33 ± 16.63	-0.60 ± 1.15	0.17 ± 0.19	0.43 ± 0.06
	3	320	86 (0)	166.46 ± 29.55	0.63 ± 1.11	0.22 ± 0.26	0.62 ± 0.10
	4	864	311 (72)	178.83 ± 28.58	-0.31 ± 1.06	0.40 ± 0.27	0.63 ± 0.09
	5	896	375 (22)	205.43 ± 31.84	0.70 ± 1.35	0.48 ± 0.23	0.72 ± 0.11
	6	957	298 (9)	195.19 ± 39.61	1.00 ± 1.01	0.33 ± 0.28	0.69 ± 0.10
	7	1 500	634 (32)	171.40 ± 22.55	0.52 ± 0.93	0.39 ± 0.24	0.65 ± 0.08
	8	2 732	497 (51)	179.77 ± 39.10	1.24 ± 1.19	0.34 ± 0.27	0.65 ± 0.11
	9	3 783	1 311 (215)	224.33 ± 43.13	1.05 ± 1.04	0.54 ± 0.23	0.75 ± 0.10
	10	4 949	1 462 (103)	184.76 ± 25.66	0.47 ± 1.22	0.40 ± 0.24	0.66 ± 0.09
	11	6 321	1 782 (141)	185.70 ± 25.96	0.55 ± 1.23	0.39 ± 0.25	0.65 ± 0.09
	12	18 004	4 761 (1 359)	236.63 ± 45.22	1.08 ± 1.20	0.53 ± 0.25	0.75 ± 0.11
	13	123 026	15 642 (6 394)	252.58 ± 33.99	1.20 ± 1.23	0.58 ± 0.24	0.79 ± 0.09


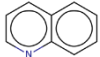









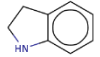









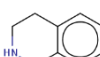









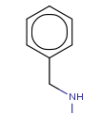










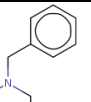













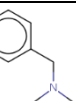
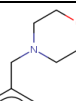



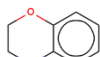









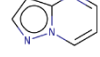









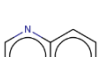









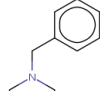








Most popular scaffolds in all libraries					Most popular scaffolds in small-sized libraries				
 21660	 4987	 4625	 4374	 4371	 2461	 782	 331	 322	 310
 4308	 3104	 3064	 3030	 2489	 274	 255	 183	 182	 174
 2461	 2294	 2267	 2195	 2162	 165	 155	 151	 138	 118
 2092	 2010	 1960	 1854	 1781	 112	 110	 109	 108	 108
 1752	 1703	 1695	 1665	 1632	 107	 106	 105	 103	 102
 1558	 1535	 1518	 1486	 1470	 102	 95	 91	 90	 90
 1381	 1370	 1335	 1334	 1307	 89	 88	 87	 84	 81
 1277	 1227	 1177	 1133	 1129	 81	 79	 79	 77	 72
 1107	 1105	 1094	 1070	 1022	 72	 72	 68	 68	 67
 1004	 944	 917	 867	 866	 67	 66	 64	 62	 62

Table S3: Fifty most common scaffolds in all libraries and in small-sized libraries. The numbers below the structures correspond to the number of molecules containing the scaffold

VI.4 Mining the Protein Data Bank to inspire fragment library design

EDITED BY

Rachelle J. Bienstock,
RJB Computational Modeling LLC,
United States

REVIEWED BY

Emanuele Carosati,
University of Trieste, Italy
Marcus Fischer,
St. Jude Children's Research Hospital,
United States

*CORRESPONDENCE

Esther Kellenberger,
ekellen@unistra.fr

SPECIALTY SECTION

This article was submitted to Theoretical
and Computational Chemistry,
a section of the journal

Frontiers in Chemistry

RECEIVED 04 November 2022

ACCEPTED 27 January 2023

PUBLISHED 10 February 2023

CITATION

Revilla Imbernon J, Chiesa L and
Kellenberger E (2023), Mining the Protein
Data Bank to inspire fragment
library design.

Front. Chem. 11:1089714.

doi: 10.3389/fchem.2023.1089714

COPYRIGHT

© 2023 Revilla Imbernon, Chiesa and
Kellenberger. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License
\(CC BY\)](https://creativecommons.org/licenses/by/4.0/). 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.

Julia Revilla Imbernon, Luca Chiesa and Esther Kellenberger*

Laboratoire d'Innovation Thérapeutique, Faculté de Pharmacie, UMR7200 CNRS Université de Strasbourg,
Illkirch-Graffenstaden, France

The fragment approach has emerged as a method of choice for drug design, as it allows difficult therapeutic targets to be addressed. Success lies in the choice of the screened chemical library and the biophysical screening method, and also in the quality of the selected fragment and structural information used to develop a druglike ligand. It has recently been proposed that promiscuous compounds, i.e., those that bind to several proteins, present an advantage for the fragment approach because they are likely to give frequent hits in screening. In this study, we searched the Protein Data Bank for fragments with multiple binding modes and targeting different sites. We identified 203 fragments represented by 90 scaffolds, some of which are not or hardly present in commercial fragment libraries. By contrast to other available fragment libraries, the studied set is enriched in fragments with a marked three-dimensional character (download at 10.5281/zenodo.7554649).

KEYWORDS

scaffold, binding mode, site comparison, interaction graph, similarity, FBDD, PDB

VI.4.1 Introduction

Over the past twenty-five years, the fragment-based approach has demonstrated its value to drug discovery by producing fifty-eight successful fragment-to-lead developments, including six approved drugs^{145,146}. A fragment-based drug discovery (FBDD) project begins with the discovery of a molecule characterized by low molecular weight, low chemical structure complexity, and low binding affinity for the assayed protein target^{112,147}. It generally relies on biophysical methods, which are sensitive, and three-dimensional binding mode information, which facilitates the elaboration of the fragment hit into a drug-like lead. Compared to high-throughput screening campaigns, fragment screening has the advantage of a reduced scale for screening while efficiently exploring chemical space, the possible combinations between fragments giving access to multiple complex compounds¹¹⁴.

The fragment quality is critical to the success of drug design. Fragment must be stable in buffers and biological media, have high solubility in water, and have a well-defined synthetic pathway to allow for follow-up discovery to develop a drug candidate. Accordingly, fragment libraries present a preferred range of physico-chemical descriptors, e.g., following the rule of three (molecular weight ≤ 300 ; number of hydrogen bond donors ≤ 3 ; number of hydrogen bond acceptors ≤ 3 ; $\log P \leq 3$; number of rotatable bonds ≤ 3), and have generally been filtered to exclude unwanted compounds, e.g., reactive or poorly soluble compounds¹⁰⁴. Recommended cut-off values for molecular weight have fluctuated in response to questions about ideal size and complexity.

In 2021, approximately half a million fragments were marketed in 81 libraries by 14 chemical suppliers¹⁴⁸. Collections vary both in size and content. Large catalogs, which

contain up to 172 000 fragments, allow the assembly of custom libraries based on specific criteria. Numerous medium to small size libraries (80 – 8 968 fragments) are adapted to particular purposes, e.g. collections of fluorinated fragments are suitable for screening by NMR^{149,150}. Several commercial libraries focus on diversity, aiming for good fragment space coverage and the ability to deliver hits to most targets. Academic initiatives have also proposed libraries of fragments built by maximizing structural or shape diversity, such as libraries developed based on diversity-oriented synthetic (DOS) strategy¹⁵¹. The diverse fragment libraries can serve different purposes, ranging from being a tool to study ligand-protein interactions or to facilitate bioisosteric replacement^{152,153} to the screening collection applicable to a wide variety of target proteins, such as the Diamond iNext Poised library of the Structural Genomics Consortium (DSiP) and the F2X chemical library, which both have been designed taking into account the diversity of chemical structures^{130,154}. The diversity of chemical structures, however, does not presage the ability to provide hits for different proteins, as structurally dissimilar fragments may exhibit the same biological activity. XChem's analysis of DSiP screening campaigns thus demonstrated that structurally dissimilar fragments can form the same interactions, consistent with the pharmacophore concept¹⁵⁵. The investigation of 309 protein-fragment structures obtained for 10 unrelated targets and 225 fragments led the authors to propose a strategy for selecting functionally diverse fragments based on the interactions formed with multiple proteins, as encoded in interaction fingerprints. Applied to DSiP hits, it yielded an ensemble of hundred functionally diverse fragments. The pharmacophore concept is also the basis for the development of the SpotXplorer library, which is composed of 96 fragments selected from commercial collections in order to maximize pharmacophore diversity while covering as much as possible of the 425 2-point and 3-point binding pharmacophores present in the Protein Data Bank (PDB) hotspots¹⁵⁶. SpotXplorer library has delivered hits for both well-studied proteins (G-protein coupled receptors and kinases) and novel targets (e.g., histone methyl transferase).

In this study, we further mine the PDB for relevant fragments, looking for small molecules able to adapt their binding mode to the protein context. We thus systematically compared the binding modes as well as the binding sites of a fragment if present in more than one PDB entry. We identified 203 fragments exhibiting two or more binding modes and binding to two or more different protein sites. This set of versatile PDB fragments overlaps with commercial libraries, contains approved drugs and approved drugs substructures, and reveals scaffolds that are still little exploited.

VI.4.2 Material and methods

a Collection of versatile PDB fragments

Fragment is here defined as a small molecular compound (molecular weight below 300 Da, non-hydrogen atoms count ranging from 2 to 18) that is not a monosaccharide, an organometallic compound, a prosthetic group, a crystallization additive (solvent, detergent, buffer, polyalcohol, etc...) ^{157,158}, a simple polyatomic ion (phosphate, sulfate or carboxylic acid), a small polyhalogenated compound, or a linear aliphatic compound.

Three-dimensional structures of protein in complex with a fragment were selected in the PDB if the deposition date is after January 1, 2000 and the resolution is below 3 Å. Their mmCIF files were downloaded from the RCSB website ^{45,159}. The preparation and standardization of the data followed the previously published protocol for the creation and maintenance of the sc-PDB, which provides, in an all-atom and annotated description, the druggable sites of the PDB and their ligands ^{46,157}. Briefly, the entries were filtered on the basis of molecular completeness and hydrogen atoms were added using protoss (v4.0, ZBH, University of Hamburg, Germany) ⁴⁸. For each PDB entry, the fragments, and their protein target (here including only amino acid residues) were saved in separate MOL2 files.

The frequency of each fragment in the data sample was evaluated by considering the occurrence of its HET code. Only fragments present in more than one PDB file were further considered.

The binding mode of a fragment to a protein site was described as an interaction graph. Hydrogen bond, ionic bond, aromatic stacking, and hydrophobic contacts were detected based on geometric rules using IChem (v5.2.8, University of Strasbourg, France) ^{50,160}. For each interaction, a pseudo-atom is placed on the ligand interacting atom and labeled according to the nature of the interaction. In the case of hydrogen bond, a second pseudo-atom is defined halfway to the protein interacting atom to account for the directionality of the bond. The pseudo-atoms constitute the nodes of the interaction graph. Edges are defined between all pairs of nodes and labeled with the corresponding Euclidian distance. All graphs generated for the same fragment were compared to each other using a Subgraph Matching Kernel (max nodes = 3). Edge similarity was obtained via a normalized distance function (max distance = 1) ¹⁶¹. The kernel yields a similarity score value ranging between 0 and 1. The number of binding modes of a fragment was inferred from density-based clustering using Sklearn (eps=0.23, N=2) ⁵¹. A correction of this number was applied based on shape similarity using ROCS (v3.4.1, OpenEye, Cadence Design Systems, Inc). Two clusters were merged if the graphs of one cluster are subgraphs of the other cluster (TverskyCombo score ≥ 2). Only fragments showing two or more binding modes were further considered.

The binding cavity of a fragment was described as a cloud of points colored with pharmacophoric properties using Volsite tool in IChem and considering a maximal distance of 4 Å between the points and the non-hydrogen atoms of the fragment ^{50,162}. All clouds of points generated for the same fragment were compared to each other using ProCare ¹⁶³. Procare

benchmarking, conducted by its authors, indicated that the comparison of dissimilar cavities yields a score below 0.47. Comparisons were made twice (cavity 1 vs cavity 2 and cavity 2 vs cavity 1), and the best score was retained. The number of cavities of a fragment was inferred from density-based clustering using Sklearn ($\text{eps}=0.47$, $N=2$)⁵¹. Only fragments showing two or more binding cavities were further considered.

Pan-assay interference compounds (PAINS)¹⁶⁴ were identified based on 652 rules using Filter (v.2.5.1.4; Openeye Scientific Software, Santa Fe) and removed from the dataset.

b Analysis of versatile PDB fragments

Molecular descriptors were computed for the PDB fragments using Pipeline Pilot (v.22.1.0.2935, BIOVIA, Dassault Systèmes, Pipeline Pilot): molecular weight, AlogP, molecular polar surface area, number of hydrogen bond acceptors and donors, aqueous solubility, number of positively and negatively charged atoms, number of rotatable bonds, number of rings, number of stereocenters, number of heavy atoms and solvent accessible surface area. The Plane of Best Fit (PBF) was calculated using Openeye libraries (v2019.10.2, OpenEye, Cadence Design Systems, Inc) from a low energy conformer generated by corina (v3.40, Molecular Networks GmbH, Nürnberg, Germany).

The modeling confidence of the binding sites and fragments was assessed by the Real Space Correlation Coefficient (RSCC). The validation XML files of each PDB entry was downloaded from the PDB website⁴⁵, and the mean RSCC of the residues in the fragment binding site was calculated. The coefficient of the fragment was directly extracted from the validation XML file. Model's quality was assigned based on Twilight classification: $\text{RSCC} > 0.9$ for a model that fits the density, $0.9 \geq \text{RSCC} \geq 0.8$ for a model that partially fits the electronic density and $\text{RSCC} < 0.8$ for a model with significant parts are missing from the density^{165,166}.

The sequence conservation between the proteins binding the same fragment was evaluated by local alignment using the EMBOSS water package (v.6.6.0.0, EMBL-EBI, Cambridgeshire, United Kingdom). EBLOSUM62 was used as score matrix. Penalties for gap opening and gap extension were set to 10 and 0.5, respectively. For each PDB complex of a fragment, the protein chain containing more than half of the amino-acids forming the fragment binding site was saved in FASTA format. Proteins were clustered based on sequence identity through density-based clustering using Sklearn ($\text{eps} = 0.10$ or $\text{eps} = 0.75$, $N = 2$)⁵¹. Only alignments which contains at least hundred amino-acids were considered.

The structural conservation between the proteins binding the same fragment was evaluated by local 3D-alignment using CE with default settings (v1.02.2, San Diego Supercomputer Center, United States). The root mean square deviation (RMSD) between backbone atom coordinates was computed between identical and homologous proteins only. The RMSD analysis

focused on the maximal values obtained from the all-against-all comparison of proteins binding the same fragment.

The conservation of the conformation of the same ligand in all its bound forms was evaluated by rigid body fit using PyMOL (v2.3.5, Schrodinger, LLC).

Structural similarity to XChem's functionally diverse fragments and SpotXplorer library was evaluated using ECFP4 molecular fingerprints and Tanimoto's coefficient (Tc) using Pipeline Pilot. The maximal similarity was reported for each versatile PDB fragment for the two compared sets.

Chemical scaffolds in the versatile PDB fragments were identified following the Bemis and Murcko approach¹²⁴ implemented in ADMET predictor (ADMET Predictor, Simulations Plus, Inc., Lancaster, California, United States). Fragments were standardized, neutralized, then classified using the option "Frameworks".

How the versatile PDB fragments populate the chemical space defined by the commercial fragments libraries was investigated by projecting the versatile PDB fragments in the published GTM map¹⁴⁸. Fragments were standardized using Standardizer (v16.10.17.0; ChemAxon Ltd.) and transformed into IIAB(2-4)_cycle ISIDA descriptors using ISIDA/Fragmentor (v.2019, Faculté de Chimie, Université de Strasbourg, France). The map was explored using GTM tools¹²⁵.

The versatile PDB fragments were searched in the drug-like collection of on-the-shelf commercially available compounds¹⁶⁷ and in Drugbank¹⁶⁸. Compound standardization and substructure search were performed using RDKit (RDKit: Open-source cheminformatics, <https://www.rdkit.org>, <https://doi.org/10.5281/zenodo.7415128>).

The molecules were neutralized, and their canonical tautomeric state was determined. The substructure search was performed using the fragment as the substructure query, and the commercial library or DrugBank compounds as the target. The chirality of both the query and the target were considered during the search. The contribution of each fragment to its corresponding superstructure was determined as the fraction of bonds in the query, over the number of bonds in the target. The result of the search was considered an exact match if the ratio between the number of bonds in the query and the target was equal to 1, otherwise it was considered as a substructure.

Datasets are available on Zenodo (doi: 10.5281/zenodo.7554649). The repository contains the MOL2 files of the crystallographic structures of the PDB versatile fragments, their binding proteins, their binding cavity, and interaction pseudoatoms. The repository also contains five CSV files: one including the versatile PDB fragments, their SMILES string, their number of binding modes and their number of binding cavities; another one containing the HET code and SMILES of the 521 fragments maintained after the PDB and ligand type filtering; the third CSV file groups the starting PDB information downloaded from the website⁴⁵; another one containing the RSCC scores of fragments and binding sites; the

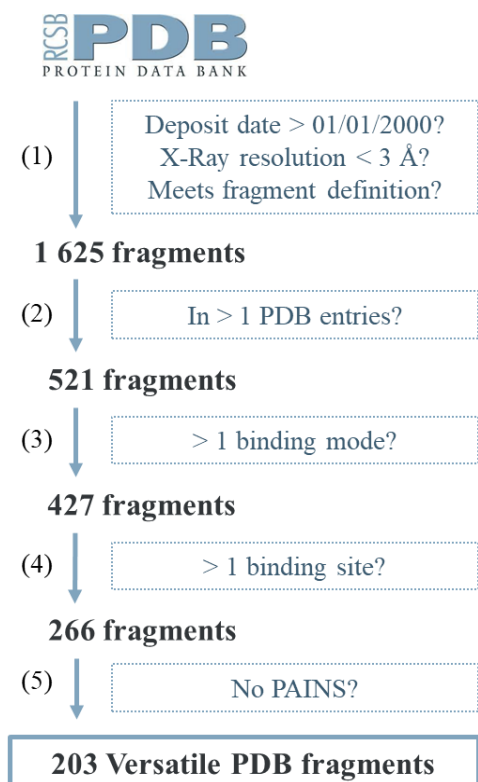


FIGURE 1

last one summarizes the results of the search for the versatile PDB fragments in Drugbank and commercial drug-like libraries.

VI.4.3 Results

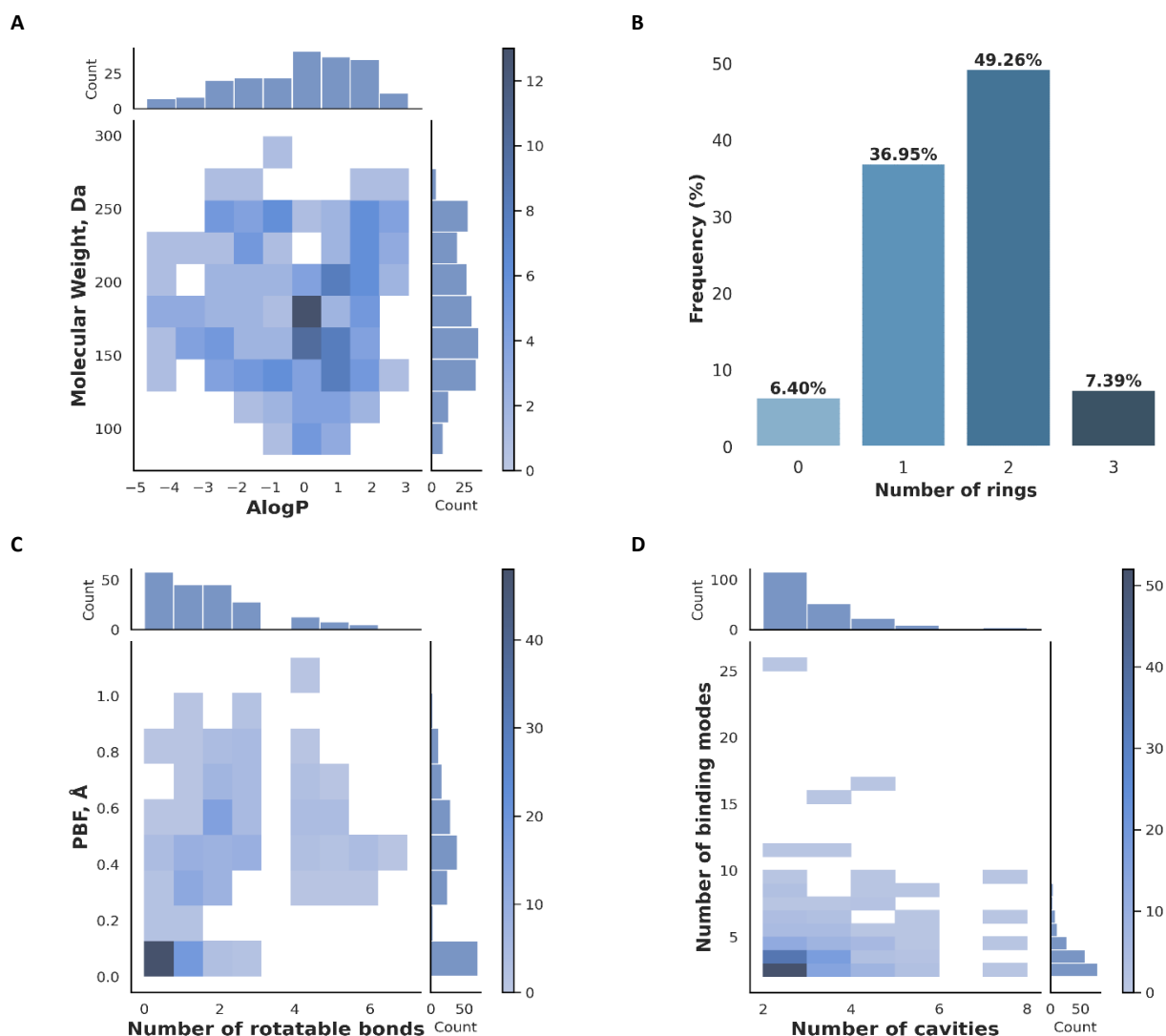
The versatile fragments were selected from the PDB following a protocol comprising five filtering steps: 1) Unwanted molecules that do not meet the definition of a fragment, 2) molecules present in a single entry, 3) molecules that show only a single binding mode, 4) molecules that are described only in binding sites that are similar, and 5) PAINS compounds which are likely to interfere with bioassays (Figure 1A, Figure III.2.2-1). At the end of step 2, a total of 521 fragments were retained. For each of them, binding modes were grouped following a density-based clustering using pairwise comparisons by subgraph matching. The distance threshold for clustering was defined on the inflection point in a polynomial projection of the distribution of the comparison data of all the 521 fragments. The visual examination of the clusters revealed cases of separation of binding modes when one is included in the other. Therefore, we applied a correction aiming to group them together. A total of 427 fragments showed two or more binding modes. The ability of these fragments to adapt to different protein environments was then assessed using a new protein comparison method which focuses on local 3D alignment. Protein cavities were delineated based on the shape and size of the bound fragment.

The distribution of similarity scores is bimodal, with a first maximum at 0.86 corresponding to the median value obtained for the comparison of two copies of the same site in the same protein, and a second maximum at 0.34 corresponding to the median value for the random comparison of two unrelated sites. Density-based clustering using the recommended threshold for detecting similar sites (0.47) tended to cluster overlapping sub-sites of a pocket while well separating cavities with no visible commonalities. A total of 266 fragments were found in complex with two or more dissimilar binding sites. Of this set, 63 fragments were discarded because they contained quaternary pyridines, phosphorus, catechol among other structural motifs prevalent in problematic screening compounds.

The structure quality was assessed based on the correlated quality indicator RSCC, available for 4,997 of the 5,220 structures of the 203 versatile fragments. The protein binding site structures are mostly very well resolved (97% of the binding site structures fit to the density). Overall, the fragment structures are also well defined (72% and 18% of the fragment structures fit and partly fit the electron density, respectively). The 10% of structures where part of the fragment is not in the density concern 115 fragments. All data was used in the analysis that follows. If the lower quality structures had been discarded, one fragment would have been excluded for lack of binding mode diversity (HET code: NBO) and eleven fragments would have been excluded for lack of binding site diversity (HET code: IBP, M2S, PEA, 2EC, LTN, 4BX, 2UP, 1FL, MOK, FMK, and AJD).

The 203 versatile fragments identified in the PDB mostly comply with the rule of three, with 130 presenting no rules' violation, 46 one rules' violation, 26 two rules' violations, and only one three rules' violations. Rules violations are mostly linked to the number of hydrogen bond donor (24) and acceptor groups (49), the number of rotatable bonds (27), and only in one instance the logP. Figure 2 shows the absence of molecular obesity in the set (panel A), and provides an overview of the molecular complexity (panel B) and three-dimensionality (panel C) of the selected fragments. For more details, the molecular weight of 91 fragments falls in the 175-250 range. All fragments contain 6 to 18 non-hydrogen atoms (Figure 2A). Fragments are weakly lipophilic (107 fragments with $0 \leq \log P < 3$) or hydrophilic (95 fragments with $\log P < 0$). A total of 175 fragments contains one or two rings (Figure 2B). Consistent with ring aromaticity, there are 66 flat fragments, as indicated by the zero PBF value (Figure 2C), yet they only represent one-third of the entire set. The number of rotatable bonds does not exceed three for almost 87% of the fragments.

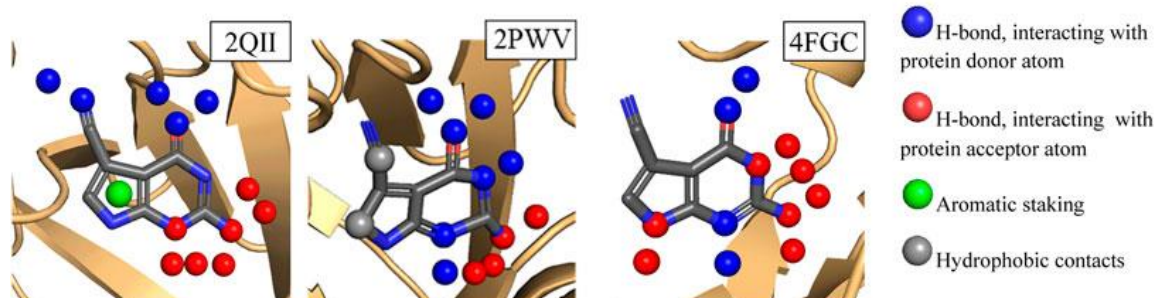
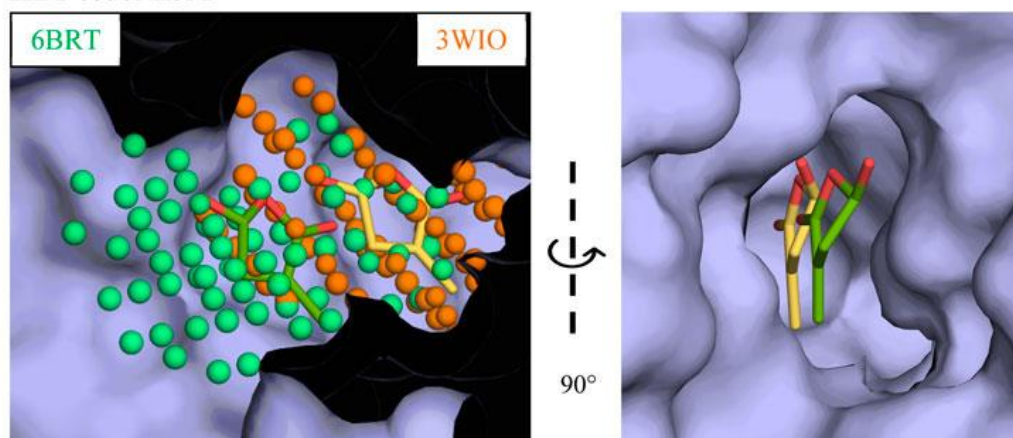
The 203 versatile fragments are present in 2,303 crystallographic PDB structures, representing 5,220 structures of fragment-protein complexes. About 82% of versatile PDB fragments bind to two or three different cavities, and accordingly using a small number of binding modes (Figure 2D). Exceptions are nucleosides such as thymidine, which shows 26 binding modes in two cavities, and flexible linear fragments such

**FIGURE 2**

Properties of the versatile PDB fragments. (A) Molecular weight vs AlogP (B) Number of rings (C) PBF vs Number of rotatable bonds (D) Number of binding modes vs Number of binding cavities. The distributions of individual properties are given on the top and side of the heat maps in (A), (C) and (D). The color scale on the side of the heat maps in (A), (C) and (D) refers to the number of fragments.

as the amino acid arginine, which shows six binding modes in two cavities. A total of twelve fragments are present in five or more cavities: six amino acids (lysine, canavanin, methionine, phenylalanine, leucine and pidolic acid), one nucleobase (guanine), two molecules naturally present in the organism and three exogenous molecules. The highly versatile natural molecules are both small ring system substituted with oxygenated group: the gamma-butyrolactone, which is a precursor of gamma-hydroxybutyrate (HET code: GBL) and the vitamin C or ascorbic acid (HET code: ASC). The three exogenous molecules target five to seven cavities: Afegostat, an iminosugar which failed a phase II clinical trial for Gaucher's Disease (PDB code: IFM), and 4-bromo-1H-pyrazole (HET code: BYZ), thiophene-2-carboxylic acid (HET code: C21).

Sequence comparison between the proteins binding the same fragment revealed that 140 fragments bind to non-related proteins (identity <25%), specifically with two to three proteins for 93 fragments and more than 40 proteins for four fragments (resorcinol RCO: 82, styrene oxide RSO: 58, uridine URI: 41 and phenylalanine PHE: 64). Only eight of the fragments are in complex with the same protein or close homologs (identity $\geq 90\%$), whose overall structure is well conserved (maximum pairwise RMSD ≤ 1.42 Å). The remaining 55 fragments bind to more distant homologs ($25\% \leq \text{identity} \leq 90\%$), whose overall structure is either well conserved (maximal pairwise RMSD ~ 1 Å for 14 fragments) or else very different (maximal pairwise RMSD ~ 6 Å for 41 fragments). Thus, a total of 22 of the fragments explore several local sites in proteins that are similar in sequence and overall structure.

A HET code: PQ0**B HET code: H3M****FIGURE 3**

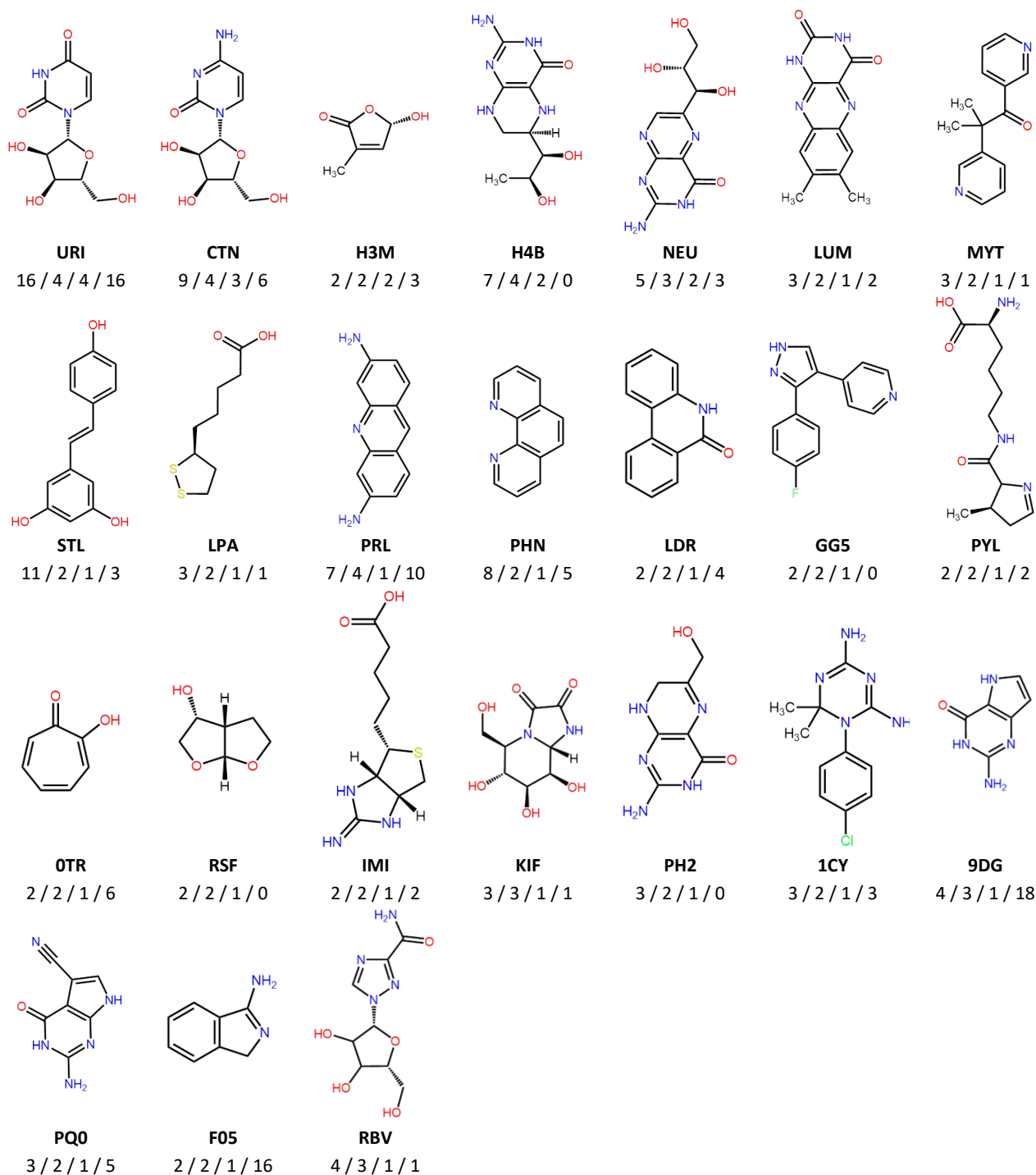
Examples of three-dimensional structures of PDB versatile fragments **(A)** Interactions formed by pyrrolo-pyrimidine in tRNA-guanine transglycolase. **(B)** Substituted furanone binding to two adjacent subpockets in strigolactone esterase D14. The cavity points representing the two subpockets and used for the comparison of the two sites are shown in the left image.

The conformations of the fragments were compared to assess whether structural plasticity can account for the versatility of binding. About 43% of the 203 versatile fragments show a conformation which is conserved in all their binding sites (maximum pairwise RMSD ≤ 0.5 Å) and only 11 fragments use very different conformations to adapt to their different protein environments (maximum pairwise RMSD > 2 Å). On average, fragment conformation is more variable for fragments that bind to proteins (Figure 3A), and to different sites in the same protein (Figure 3B). The first example involves a pyrrolo-pyrimidine fragment shown in Figure 3A (HET code: PQ0). This fragment shows three binding modes in two protein cavities found in two types of proteins, tRNA-guanine transglycolase (PDB codes: 1IT8, 1POB, 2PWV, 2QII) and nitril reductase QueF (PDB code: 4FGC). The three binding modes are dominated by hydrogen bonds. Changes involve the pyrrole moiety, which does not interact with the protein or interact with a tyrosine side chain depending on local structural variation of tRNA-guanine transglycolases binding site. More importantly, different protonation states of

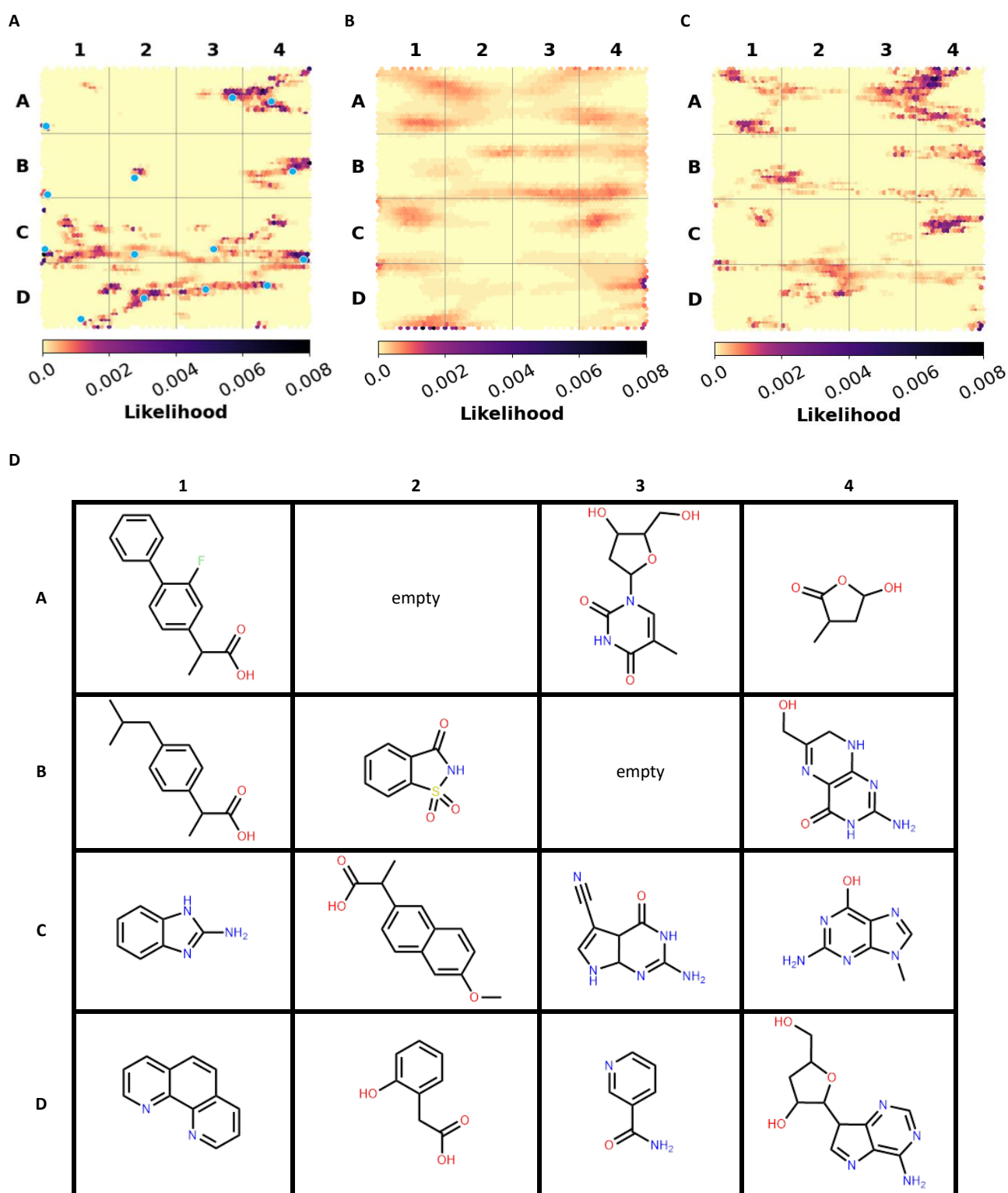
to homologs than for fragments that bind to proteins that are unrelated in sequence: the mean value of the maximum pairwise RMSD is equal to $0.78 \text{ Å} \pm 0.69$ for the 140 fragments binding to non-related proteins, to $0.88 \text{ Å} \pm 0.72$ for 55 fragment binding to distant homologs, and $0.97 \text{ Å} \pm 0.58$ for the eight fragments binding to the same protein or close homologs.

Two examples of versatile fragments are discussed below and illustrate how a fragment adapts its binding mode to unrelated the fragment allow for different hydrogen bonding patterns in the two proteins. Last, although it is not the focus of the present study, this fragment also binds to RNA (PDB code: 3GCA)¹⁵⁹.

The second example consists of a furanone (HET code: H3M). The furanone is in complex twice with the strigolactone esterase D14 (PBD code: 6BRT and 3WIO), yet shows a 4 Å shift in the main cavity, thereby defining two non-overlapping subpockets (Figure 3B). The two corresponding binding modes are dominated by π interactions which are frequently observed in RNA-protein complexes¹⁶⁹, but they differ in the hydrogen bonds formed with the protein, which either involve a


FIGURE 4

Representative chemical structures of the versatile PDB fragments which are absent or poorly represented in commercial fragments libraries. Below each fragment structure are indicated the HET code, then the number of binding modes / the number of different cavities / the number of versatile PDB fragments containing the same scaffold / the number of commercial fragments containing the same scaffold.

**FIGURE 5**

Spatial distribution of the fragments in GTM Landscapes **(A)** Versatile PDB fragments **(B)** All commercially-available fragments **(C)** Ensemble of nature-product like commercial libraries. **(D)** Examples of versatile PDB fragments. Their projection points are marked with blue dots on the map **(A)**. Figures 5B and 5C are adapted from ¹⁴⁸

tryptophan indole ring, or a cysteine sulfhydryl and a protein-bound water molecule. Of note, the binding mode comparison method used here ignored weak hydrogen bonds and water mediated interactions yet was nevertheless able to differentiate the two binding modes. Interestingly, the three-dimensional structures of furanone ring are not identical in the two complexes, suggesting that despite the absence of rotatable bonds this scaffold is able to adapt its shape to the local environment.

The versatile PDB fragments were compared to two academic collections, the XChem's functionally diverse fragments and SpotXplorer library. The distribution of maximal similarity between versatile PDB fragments and XChem's functionally diverse fragments had a median maximal value equal to 0.21 (Supplementary Figure 1). Thirteen versatile PDB fragments share common chemical structures with XChem's functionally diverse fragments ($T_c > 0.45$). One single fragment is common to both sets. This fragment is a sulfapyridine (HET code: SFY), which shows two binding modes in three cavities. It contains multiple pharmacophoric features (two hydrophobic rings, three hydrogen bond donors, two hydrogen bond acceptors), allowing adaptation to different protein environments. The chemical structures of versatile PDB fragments and SpotXplorer library do not overlap, with a median maximal similarity value equal to 0.23. A total of six versatile PDB fragments share common chemical structures with SpotXplorer library ($T_c > 0.45$).

The versatile PDB fragments were compared to the commercially available fragments using two approaches: shared Bemis and Murcko scaffolds and generative topographic map (GTM), which represents the chemical space in a 2D map, here based on topological descriptors¹²⁵. The 203 fragments are covered by 90 scaffolds. The commercial chemical space of fragments was described with about 60,000 scaffolds (<https://gtmfrag.drugdesign.unistra.fr/>) ranging from very simple (benzene) to very complex (adamantane)¹⁴⁸. We identified four absent scaffolds and 20 rare scaffolds, which were observed in twenty or less commercially available fragments (Figure 4). These 24 scaffolds represent a total of 32 fragments (supplementary Figure 2). They show on average more oxygen atoms than commercial fragments (1–6 O in 79% of PDB scaffolds; 1–3 O in 31% of commercial scaffolds), and conversely fewer nitrogen atoms (1–5 N in 79% of PDB scaffolds; 1–6 N in 93% of commercial scaffolds). The coverage of the fragment space by the versatile PDB fragments was further analyzed through a GTM model recently developed to compare commercial libraries, based on the likelihood of the compounds projected onto the common frame (Figure 5). The versatile PDB fragments populate three regions which show a low density among the general commercial fragments set (Figure 5B): **A4**, **C2**, and **C4**. The region **A4** groups together three-dimensional molecules, mostly non-aromatic heterocycles which contain oxygen atoms. The region **C2** includes aromatic rings fused and substituted with alcohol,

ketone or carboxylic acid groups. The region **C4** shows many purine derivatives with amino substituents on the pyrimidine ring. Noteworthy, the patterns observed in regions **A4** and **C2** have a low likelihood among the general commercial fragments set (Figure 5B) but are characteristic of nature-product like fragments of commercial libraries (Figure 5C).

The 203 versatile PDB fragments were further characterized by querying a repository of approved drugs, Drugbank¹⁶⁸, and screening collections of commercial catalogs. A total of 45 fragments are registered as approved small molecule drugs in DrugBank. Among these, seven are also nutraceuticals and sixteen in clinical trials. Around half of the fragments (98) were identified as substructures of 1,436 approved drugs. The fragments generally represent a small portion of the corresponding superstructures, with the bonds in the fragment representing less than 50% of the bonds in the superstructure for 88% of the fragment-superstructure pairs. While 22 fragments are substructures of only a single drug, 48 are found in at least five drugs. Larger fragments ($MW > 175$) account for 28 of the 45 identified drugs, while 37 represent a substructure of 134 drugs. The exploration of commercial catalogs identified 95 purchasable fragments, and 163 fragments substructures of a huge number of commercially available screening compounds (4,769,862). The fragments represent, however, less than 50% of the superstructures' bonds in 97% of the fragment-compound pairs.

VI.4.4 Discussion

The PDB is an essential resource for structure-based drug discovery. Most drugs that have been approved by the FDA over the past decade were developed when structural information was freely available for the protein target¹⁷⁰. Additionally, with 38,790 entries in the Chemical Component Dictionary in 2022, the PDB provides important material for characterizing recognition between small molecular entities and proteins. The number of fragment-sized chemical entities ($MW \leq 300$ Da) amounts to 13,965. Our analysis of the PDB identified only 203 versatile fragments, able to adopt at least two binding modes and to bind to at least two distinct sites. Most of these fragments target multiple proteins that differ in sequence, but we also highlighted eight fragments that bind differently to the same protein or close homologs.

Do the properties of the 203 versatile PDB fragments meet FBDD expectations? Lipophilicity and molecular weight are two widely used criteria, as they are related to solubility, metabolism and selectivity, and a *sweet spot* has been suggested around a molecular weight 400 Da and a logP around 3^{127,128}. These values are shifted down in commercial fragment libraries with molecular weight in the range 200–300 Da and logP in the range 0–2¹⁴⁸. A minimal molecular weight around 175 Da was suggested by the study of the conservation of the binding mode of the fragments in their drug-like superstructures¹¹¹. The same threshold was proposed

following the analysis of XChem's screening campaigns which revealed that fragments that have never been shown to bind to a target tend to have a low molecular weight compared to fragments that bind to one or more targets¹⁵⁵. According to the size criterion, about half of the versatile PDB fragments may provide a valid base for fragment growing (105 fragments with MW > 175 Da). A total of 38 fragments shows a logP in the 0–2 range, yet there are also 50 hydrophilic fragments (logP < 0). The combinations of lipophilicity and molecular weight which are little explored by recently developed synthetic oral drugs should not be overlooked since they may correspond to molecules of therapeutic interest such as natural products (e.g., antibiotics)¹⁷¹. Thus, we did not exclude hydrophilic fragments from the analysis. In addition, these fragments have the advantage of good aqueous solubility.

Are the 203 versatile PDB fragments original? To answer this question, we compared this set to XChem's functionally diverse fragments and SpotXplorer library, to the chemical space covered by commercial fragment libraries, and to screening collections of drug-like molecules. The XChem's functionally diverse fragments and SpotXplorer library are two collections comprising a wide diversity of pharmacophores evidenced by the binding modes observed in the structural data^{155,156}. The comparison of chemical structures revealed that these two collections and the 203 versatile PDB fragments hardly overlap. The XChem's functionally diverse fragments and SpotXplorer library being assembled from commercial catalogs, we therefore sought to assess whether the scaffolds of the 203 versatile PDB fragments were indeed present in the fragment's libraries provided by chemical suppliers. According to our 2021 inventory¹⁴⁸, 198 versatile PDB fragments or structural analogs are commercially available, but 27 of them are rare in commercial fragment libraries. The five fragments whose scaffold is missing from commercial fragment libraries also cannot be purchased from commercial drug-like collections. These fragments are all the more original as only two of them are substructures of an available compound (HET code GG5 and RSF, representing 71% and 47% of the superstructure bonds in their respective best fragment/compound pair). In summary, if a large part of the versatile PDB fragments can be purchased from chemical suppliers, 24 of their scaffolds are little or not represented in commercial fragment libraries (Figure 4).

Many of the versatile PDB fragments that are not or rarely present in commercial fragment libraries are enriched in oxygen atom. Oxygen containing functional groups, such as ether, are more common in natural products than in synthetic molecules¹⁷². Among oxygen-containing versatile PDB fragments are eight nucleosides and close analogs. The phosphorylated nucleoside, i.e. nucleotides, are common cofactors, therefore recognized by numerous protein sites. They have already widely inspired medicinal chemists, allowing for example the development of many anti-viral drugs^{173,174}.

Have all the 203 versatile PDB fragments already been exploited in drug discovery? To answer this question, we

explored Drugbank thus identifying 45 approved drugs. According to Drugbank annotations, 29 of these drugs have more than one known target. For example, acetazolamide used to treat edema, certain types of epilepsy and glaucoma, has been reported to target nine types of carbonic anhydrase and aquaporin-1. The PDB structures provide eight binding modes of acetazolamide in multiple enzyme types: carbonic anhydrase, chitinase, deacetylase and synthase (HET code: AZM). The 16 other drugs have a single known target or no targets. For example, sulfapyridine is a sulfonamide antibiotic targeting dihydropteroate synthase type-1. The PDB structures provide four binding modes of sulfapyridine in multiple protein types: reductase, kinase, proteinase, erythrocyte membrane protein and synthase (HET code SFY). As an example of drug with no known target, piracetam was crystallized in glutamate receptor 2 and glutamate receptor 3 (HET code PZI). Interestingly, 123 of the versatile PDB fragments are substructures of approved drugs, including drugs with a single known target. The data revealed examples of promiscuous fragment embedded into a selective drug, such as indole-3-carboxylic acid (HET code ICO) which was crystallized in complex with four types of protein (androgen receptor, integrase, malate synthase G, and a transcriptional regulator), and which also represents about half of tropisetron, a 5HT-3 receptor antagonist used as an antiemetic in the treatment of chemotherapy-induced nausea and vomiting (K_i = 2.70 nM, BindingDB entry BDBM50108392¹⁷⁵). In summary, we have identified 80 versatile PDB fragments that have not been yet exploited in drugs. Moreover, structural data on the 203 versatile PDB fragments reveal new off-targets for existing drugs and show that a versatile fragment can evolve into a drug with high affinity for its target. This data, freely available on zenodo, provides fragment-focused insights complementing previous studies of multi-target PDB ligands^{176–180} that can be used to better understand polypharmacology^{181,182}.

VI.4.5 Conclusion

The clustering of the binding modes and binding sites of PDB fragments issued a set of 203 low molecular weight compounds able to adapt to different protein environments and to exploit several combinations of interacting atoms. The versatile PDB fragments constitute an ensemble which contains many structures that can be purchased directly yet also a few scaffolds that are poorly represented in fragment screening collections. The presence of oxygenated compounds in rare and original scaffolds calls for renewed interest in natural products. Since the PDB provides structural insights into well-studied biological functional systems, a significant amount of versatile PDB fragments (45/203) are approved drugs. Moreover, many promiscuous fragments are a substructure of selective drug, suggesting their utility for FBDD.

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.

Author Contributions

Conceptualization: E. K. Methodology: J. R. I. and E. K. Implementation of the protocol: J. R. I. and L. C. Data preparation: J. R. I. Formal analysis: J. R. I. and E. K. Writing—

original draft preparation: J. R. I. and E. K. Writing—review and editing: J. R. I., L. C., and E. K.; project coordination: E. K.

Acknowledgments

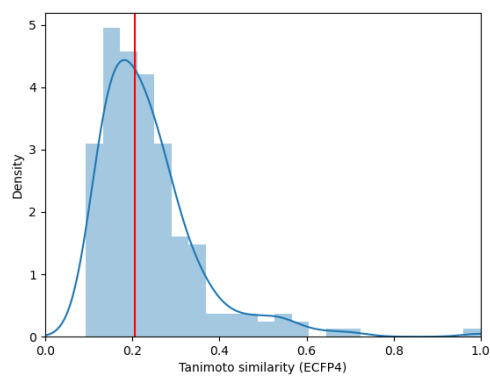
The authors thank Guillaume Bret for technical assistance and helpful discussion. The authors thank the Institut du Médicament (IMS) for financial support.

VI.5 SI Mining the Protein Data Bank to inspire fragment library design

Table 1: Visually discarded fragments by category.

Discarded group	Discarded HET
Aliphatic	DCX
Ammonium	144, BTM, TBA, 211, DCD, N6C
Monosaccharides	IPT, G4D, ISD, IPD, KDG, X1P, ASO, LGC, M7P, GAR, GDL, GLR, GPM, SOE, 1KM, NHF, R2B, M7B, LIP, 149
Phosphate	PPF, 3PP, PEP, 13P, 2PG, DPF, DPJ, P3S, G88, SUF, AS9, FOM, PGH, OPE, CP, PPR, EIP, IPR, IPE, 210, PC, KPC, PCT, S0H, UVW, P23, P25, DST, MPJ, FPE, DED, CDI, PLU, PAE, MDN, TCE, ISY, 2HE, M44, ODV, TB6
Polyhalogenated	PCI, HLT, ICF, PFB, CFH, T6C
Sulfate	ESA, TAU, COM, SPV, PS9
Too small	TBU, DMF, PTL, BTL, AKR, MOE, PTD, TBF, BMD, EEE, NMU, OXM
Two in site	ID3, 2MY

A



B

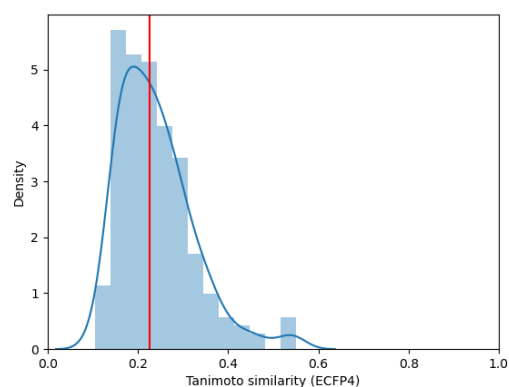


Figure SI 1: Distribution of Tanimoto Similarity between the versatile PDB fragments and (A) XChem top 100 fragment selection (B) SpotXplorer. The red line represents the median maximal value.

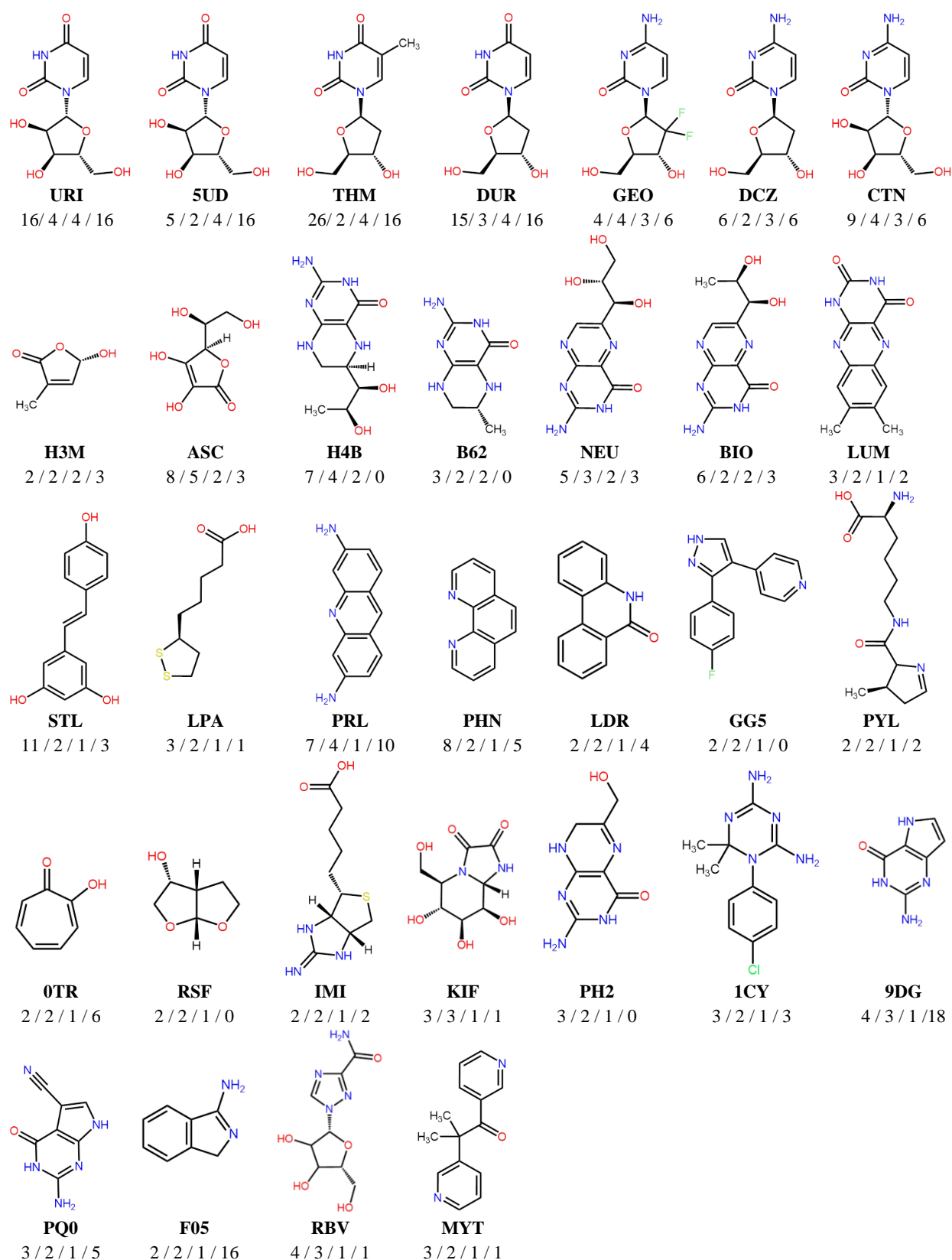


Figure SI 2: Representative chemical structure of versatile and original PDB fragments. Below each fragment structure are indicated the HET code, then the number of binding modes / the number of different cavities / the number of versatile PDB fragments containing the same scaffold / the number of commercial fragments containing the same scaffold.

Julia REVILLO IMBERNON
**Rational design of bioactive
molecules for antibiotic purposes**

Résumé

La montée de la résistance aux antibiotiques représente une menace significative pour la santé humaine. Ce manuscrit aborde cette problématique à travers deux projets de recherche collaboratifs et interdisciplinaires : (1) l'optimisation de la néomycine B, un antibiotique de la classe des aminoglycosides, et (2) la recherche de petites molécules pour une nouvelle cible, l'anneau β bactérien.

(1) Nous avons développé une méthodologie basée sur la structure tridimensionnelle pour prédire l'effet de modification de la néomycine B sur l'action des enzymes bactériennes de modification des aminoglycosides (AME). Cette méthodologie utilise la simulation de dynamique moléculaire pour évaluer la stabilité des modes de liaison, et a été mise au point en considérant cinq AMEs présentes dans des souches cliniques et le dérivé HL_171.

(2) Nous avons mis en place une stratégie de criblage virtuel inspirée des données structurales obtenues pour 73 fragments de la chimiothèque Diamond DLS et cinq peptides inhibiteurs issus de protéines liant l'anneau β bactérien et impliquées dans la réplication de l'ADN. Nous avons priorisé une centaine de molécules disponibles commercialement, et qui sont actuellement évaluées par nos collaborateurs biologistes.

Abstract

The emergence of antibiotic resistance represents a significant threat to human health. This manuscript addresses this issue through two collaborative and interdisciplinary research projects: (1) the optimization of neomycin B, an antibiotic belonging to the aminoglycoside family, and (2) the search for small molecules targeting a novel bacterial target: the β -clamp.

(1) We have developed a methodology based on three-dimensional structures to predict the effect of modifying neomycin B on the activity of Aminoglycoside-Modifying Enzymes (AME). This methodology uses molecular dynamics simulations to evaluate the stability of the binding modes. It has been constructed by considering five AMEs present in clinical strains and the HL_171 derivative.

(2) We have implemented a virtual screening strategy inspired by the structural data obtained from 73 fragments from the Diamond Light Source chemical library and five inhibitory peptides derived from proteins that bind the β -clamp and play a role in DNA replication. We have prioritized a hundred commercially available molecules, which are currently being evaluated by our biologists' collaborators.