UNIVERSITAT DE BARCELONA

Computational exploration of protein-ligand interactions: substrate selectivity and small molecule inhibition

Aitor Valdivia

Aquesta tesi doctoral està subjecta a la llicència ***Reconeixement 4.0. Espanya de Creative Commons***.

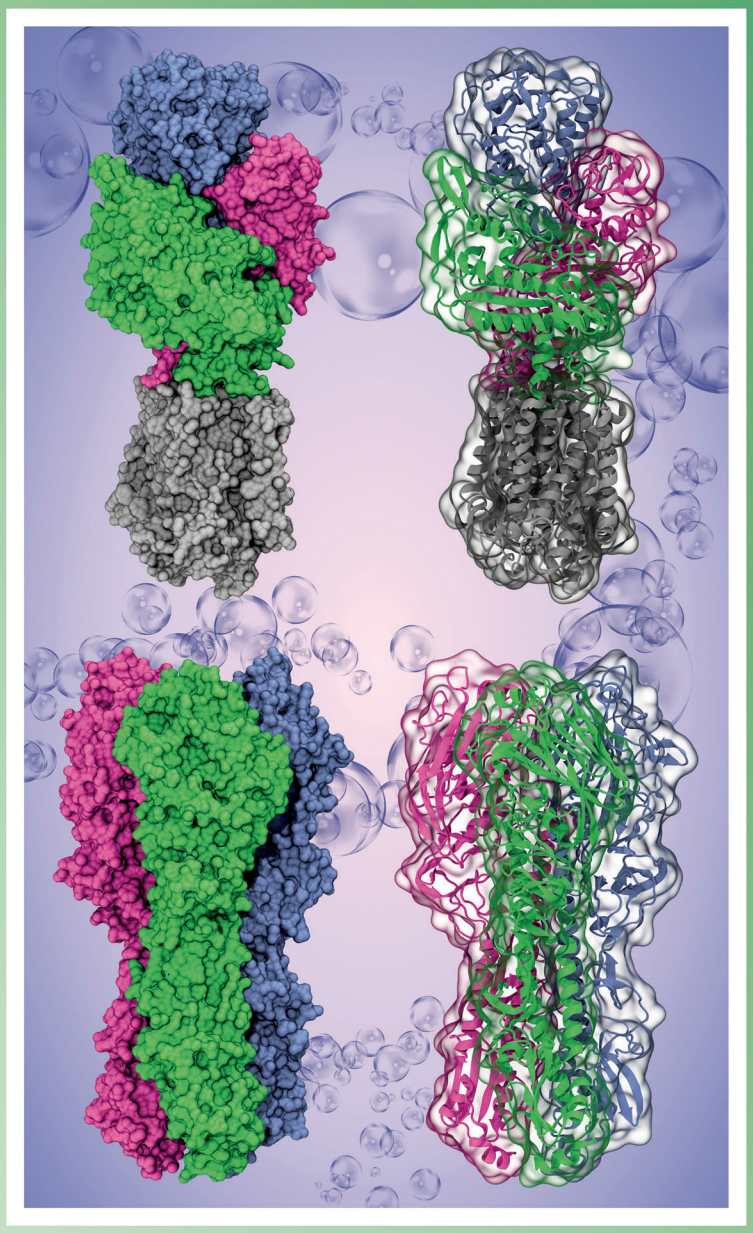
Esta tesis doctoral está sujeta a la licencia ***Reconocimiento 4.0. España de Creative Commons***.

This doctoral thesis is licensed under the ***Creative Commons Attribution 4.0. Spain License***.

UNIVERSITAT DE BARCELONA

Computational exploration of protein-ligand interactions:

substrate selectivity and small molecule inhibition



Aitor Valdivia

2025

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

Computational exploration of protein-ligand interactions: substrate selectivity and small molecule inhibition

Exploració computacional d'interaccions proteïna-lligand: selectivitat de substrat i inhibició amb molècules petites

Aitor Valdivia

Barcelona, April 2025

UNIVERSITAT DE BARCELONA

FACULTAT DE FARMÀCIA I CIÈNCIES DE L'ALIMENTACIÓ

Departament de Ciències de l'Alimentació i Gastronomia

Programa de Doctorat en Biotecnologia

Computational exploration of protein-ligand interactions: substrate selectivity and small molecule inhibition

Exploració computacional d'interaccions proteïna-lligand: selectivitat de substrat i inhibició amb molècules petites

Memòria presentada per Aitor Valdivia per optar al títol de doctor per la Universitat de Barcelona

Dra. Salomé Llabrés (Directora)

Prof. F. Javier Luque (Co-Director)

Prof.ª Laura Herrero (Tutora)

Aitor Valdivia

Barcelona, April 2025

A mi madre.



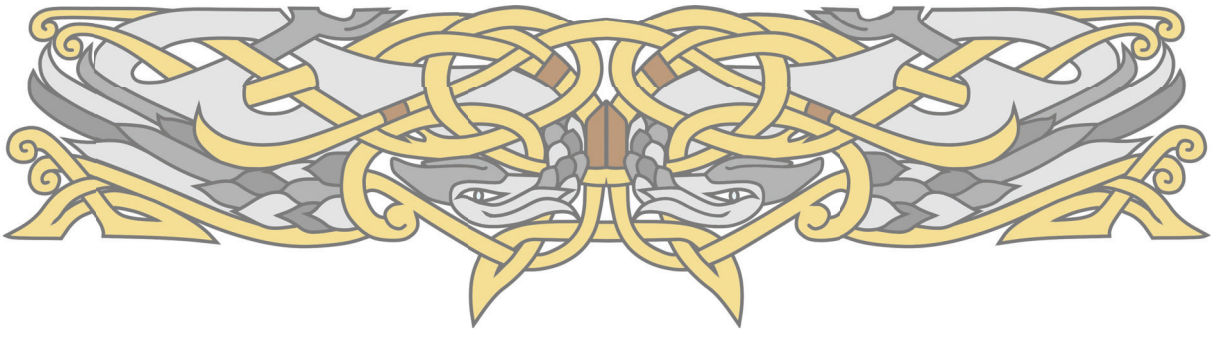
I know that I hung on a windy tree
for all of nine nights,
wounded by a spear and given to Óðinn,
myself to myself,
on that tree of which no one knows
the origin of its roots.

They blessed me with neither bread nor horn,
I peered down, I took up the runes,
screaming I took them.
And thence I fell.

Nine mighty songs I learnt from the famous son
of Bǫlþórr, father of Bestla,
and I got a draught of the precious mead,
poured from Óðrerir.

Then I began to germinate and to be wise,
and to grow and feel well.
One word sought another in my words,
one action sought another in my actions.

Hár (Hávamál, stanzas 138-141)



ABSTRACT

The continuous development of molecular modelling and computational methods is pivotal in addressing the challenges encountered in the fields of biomedicine and drug design. Among them, the lack of structural and mechanistic information underlying protein dysfunction and drug resistance stands out. In this context, the present doctoral thesis focuses on elucidating the molecular interactions behind the mechanism of action (MOA) of two druggable targets that are directly involved in major public health concerns highlighted by the World Health Organization (WHO), namely cardiovascular disease (CVD) and infectious diseases.

In the first project, the mechanism and selectivity of the main cholesterol transporter in mammals, the Niemann-Pick C1-Like 1 transporter (NPC1L1) has been investigated. The imbalance of the correct blood lipid profile, caused by a dysregulation on cholesterol homeostasis, plays a key role in the development of CVD. Cholesterol is a fat-like substance endowed with a high physiological relevance in humans, and its homeostasis is tightly regulated by various cellular processes including the import in the small intestine and the reabsorption in the biliary ducts by the NPC1L1 importer. NPC1L1 can mediate the absorption of a variety of sterols but strikingly exhibits a large sensitivity to cholesterol epimerization. The existence of two distinct binding modes for cholesterol and epi-cholesterol to the N-terminal domain (NTD) of NPC1L1 provides the structural basis to rationalize the affinity difference between the two substrates. The configuration of the hydroxyl group of cholesterol and the presence of a structural water molecule are key features for the effective binding to NTD. In addition, the existence of a gating mechanism involved in the sterol binding to this protein domain was confirmed by the conformational rearrangements of the α8/β7 loop. Furthermore, distinct single-point mutations that impair cholesterol transport show a mild effect on the binding of cholesterol to the NTD, but a drastic influence on the conformational landscape of the α8/β7 loop in the apo species compared to the wild-type protein. Overall, these results support the functional role played by the α8/β7 loop in regulating the access of ligands to NPC1L1, and hence in assessing the impact of these mutations on diseases related to disruption of sterol absorption.

In the second project, the MOA of an anti-viral inhibitor against the Hemagglutinin (HA) of Influenza A virus (IAV) has been investigated. We explored the putative binding site and the binding mode of the pinanamine-based inhibitor M090, able to block the HA-mediated membrane fusion mechanism in early stages of the infection and active against a wide range of

Keywords: cholesterol transport, binding affinity, enhanced sampling techniques, fusion inhibitors, hemagglutinin, influenza A, selectivity, molecular dynamics, relative free energy calculations, structure-based drug design

HA variants. The predicted binding site is located at the interface defined by the α-helix $L80_2$-$A96_2$ and by the loops $P301_1$-$K318_1$ and $M59_2$-$L73_2$, i.e. in close proximity to the main stem region of the HA trimer. The results obtained from relative binding free energy (RBFE) calculations show close agreement with the differences in inhibitory potency determined in antiviral assays for M090 and its derivatives, providing *in silico* validation of the proposed binding mode. In addition, the $E74_2 \rightarrow D$ drug-resistant mutation weakens the optimal hydrogen bonding network established between the loops forming the M090 binding site. Altogether, a change in the overall packing against the stem region of HA leads to a disruption of the binding mode thus explaining the loss of activity.

In the third project, we examined the binding site and the binding mode of the sulphonamide derivative VF-57a, synthesised by the group of Prof. Santiago Vázquez (Departament de Farmacologia, Toxicologia i Química Terapèutica, UB) and tested against IAV-infected MDKC cells by the group of Prof. Lieve Naesens (Department of Microbiology, Immunology and Transplantation, KUL). The compound shows sub-micromolar activities against A/H1N1 strains (EC_{50} = 0.31 μM and EC_{50} = 0.92 μM against A/Virginia/ATCC3/2009 and A/Puerto Rico/8/1934, respectively) and H5N1 strains (EC_{50} = 0.81 μM against A/bald eagle/Florida/W22-134-OP/2022). We predicted the binding site for VF-57a to be in close proximity to the binding site for the broad-spectrum inhibitor Arbidol (Umifenovir). In this pocket, the thiophene ring of VF-57a is stabilised by van der Waals contacts with $V36_1$, $L49_1$, $P308_1$, $F309_1$, $V55_2$ and $L99_2$ from one protomer and with $L98_2$ and $L102_2$ from a second protomer. The ring containing the trifluoromethyl moiety is stabilised by π-stacking interactions with $Y94_2$ and $F309_1$. In addition, RBFE predictions for VF-57a and some derivatives support the predicted MOA by showing a good correlation with the experimental activities.

By combining state-of-the-art computational chemistry methods with experimental data, this doctoral thesis facilitates the understanding of certain cholesterol-related physiological disorders and contributes to the development of new small molecules as antivirals against IAV.

Keywords: cholesterol transport, binding affinity, enhanced sampling techniques, fusion inhibitors, hemagglutinin, influenza A, selectivity, molecular dynamics, relative free energy calculations, structure-based drug design

RESUM

El desenvolupament continu de mètodes de modelització molecular és fonamental per afrontar els reptes que es plantegen en els camps de la biomedicina i el disseny de fàrmacs. Entre ells, destaquen la falta d'informació estructural i mecanística darrera de la disfunció de les proteïnes i la resistència als fàrmacs. En aquest context, la present tesi doctoral es centra en elucidar les interaccions moleculars responsables del mecanisme d'acció de dues dianes farmacològiques directament implicades en dos dels principals problemes de salut pública identificats per l'Organització Mundial de la Salut: les malalties cardiovasculars i les malalties infeccioses.

En el primer projecte, s'ha investigat el mecanisme i la selectivitat del principal transportador de colesterol en mamífers, el transportador el transportador Niemann-Pick C1-Like 1 (NPC1L1). El desequilibri del perfil lipídic sanguini, causat per una disrupció de l'homeòstasi del colesterol, té un paper clau en el desenvolupament de les malalties cardiovasculars. El colesterol és una substància similar al greix amb una importància fisiològica fonamental en els éssers humans, i, per tant, i la seva homeòstasi està estretament regulada per diversos processos cel·lulars, que inclouen la importació en l'intestí prim i la reabsorció en els conductes biliars per l'importador NPC1L1. NPC1L1 pot mediar en l'absorció d'una varietat de esterols, però sorprenentment mostra una gran sensibilitat a l'epimerització del colesterol. Hem descrit l'existència de dos modes d'unió diferents per al colesterol i l'epicolesterol al domini N-terminal (NTD) de NPC1L1, que expliquen les característiques moleculars responsables de la diferència d'afinitat entre tots dos substrats. La configuració del grup hidroxil del colesterol i la presència d'una molècula d'aigua estructural al lloc d'unió, són imprescindibles per a l'absorció efectiva. A més, confirmem l'existència d'un mecanisme d'obertura implicat en la unió dels esterols a aquest domini de la proteïna mitjançant el reordenament conformacional del loop α8/β7. D'altra banda, s'ha determinat que una sèrie de mutacions puntuals que disminueixen l'absorció de colesterol exerceixen un efecte lleu en la unió del colesterol al NTD, però al seu torn tenen una influència dràstica en el rang conformacional del loop α8/β7 en l'estat apo en comparació amb la proteïna no mutada. En conjunt, els resultats recolzen el paper funcional del loop α8/β7 en la regulació de l'accés de lligands a NPC1L1 i, per tant, en l'avaluació de l'impacte d'aquestes mutacions en malalties relacionades amb l'alteració de l'absorció d'esterols.

Paraules clau: transport de colesterol, afinitat d'unió, tècniques de mostreix avançat, inhibidors de fusió, hemaglutinina, grip A, selectivitat, dinàmica molecular, càlculs d'energia lliure relativa, diseny de fàrmacs basat en l'estructura

En el segon projecte, s'ha investigat el mode d'unió d'un d'inhibidor antiviral de la hemaglutinina (HA) del virus de la grip A (IAV). En aquest projecte, identifiquem el lloc d'unió putatiu i el mode d'unió de l'inhibidor M090 basat en la pinanamina, capaç de bloquejar el mecanisme de fusió de membrana mediat per HA en les primeres fases de la infecció i actiu contra una àmplia gamma de variants de HA. El lloc d'unió previst es localitza en la interfície definida per la α-hèlix $L80_2$-$A96_2$ i pels loops $P301_1$-$K318_1$ i $M59_2$-$L73_2$, és a dir, molt a prop de la regió principal del trímer de HA. Les prediccions d'energia lliure d'unió relativa (RBFE) mostren una estreta concordança amb les diferències en la potència inhibidora determinada en assajos antivirals per a M090 i els seus derivats, proporcionant així una validació *in silico* del mode d'unió proposat. A més, observem que la mutació $E74_2 \rightarrow D$ resistent al fàrmac, debilita la xarxa òptima d'enllaços d'hidrogen establerta entre els loops que formen el lloc d'unió de M090. En conjunt, demostrem que un canvi en l'empaquetament global enfront de la regió de la tija de l'HA condueix a una alteració del mode d'unió que explica la pèrdua d'activitat.

En el tercer projecte, identifiquem el lloc d'unió putatiu i el mode d'unió del derivat sulfonamida VF-57a, sintetitzat pel grup del Prof. Santiago Vázquez (Departament de Farmacologia, Toxicologia i Química Terapèutica, UB) i assajat enfront de cèl·lules MDKC infectades amb IAV pel grup de la Prof. Lieve Naesens (Department of Microbiology, Immunology and Transplantation, KUL). El compost mostra activitats submicromolars contra soques A/H1N1 (EC_{50} = 0.31 μM i EC_{50} = 0.92 μM contra A/Virginia/ATCC3/2009 i A/Puerto Rico/8/1934, respectivament) i soques H5N1 (EC_{50} = 0.81 μM contra A/bald eagle/Florida/W22-134-OP/2022). Els resultats indiquen que el lloc d'unió de VF-57a es troba molt prop del lloc d'unió de l'inhibidor d'ampli espectre Arbidol (Umifenovir). En aquest lloc d'unió, l'anell de tiofè de VF-57a s'estabilitza mitjançant contactes de van der Waals amb $V36_1$, $L49_1$, $P308_1$, $F309_1$, $V55_2$ i $L99_2$ d'un protòmer i amb $L98_2$ i $L102_2$ d'un segon protòmer. L'anell que conté el fragment trifluorometil s'estabilitza mitjançant interaccions π amb $Y94_2$ i $F309_1$. A més, les RBFE per a VF-57a i els seus derivats de l'estudi de relacions estructura-activitat (SAR) donen suport al mode d'unió predit en mostrar una bona correlació amb les activitats experimentals.

Mitjançant la combinació dels últims avenços en mètodes de química computacional amb dades experimentals, aquesta tesi doctoral facilita la comprensió d'uns certs trastorns fisiològics relacionats amb el colesterol i contribueix al desenvolupament de noves molècules com a antivirals front el IAV.

Paraules clau: transport de colesterol, afinitat d'unió, tècniques de mostreix avançat, inhibidors de fusió, hemaglutinina, grip A, selectivitat, dinàmica molecular, càlculs d'energia lliure relativa, diseny de fàrmacs basat en l'estructura

RESUMEN

El continuo desarrollo de métodos de modelización molecular es fundamental para afrontar los retos que se plantean en los campos de la biomedicina y el diseño de fármacos. Entre ellos, destacan la falta de información estructural y mecanística que provoca la disfunción de las proteínas y la resistencia a los fármacos. En este contexto, la presente tesis doctoral se centra en elucidar las interacciones moleculares que subyacen al mecanismo de acción de dos dianas farmacológicas directamente implicadas en dos de los principales problemas de salud pública identificados por la Organización Mundial de la Salud: las enfermedades cardiovasculares y las enfermedades infecciosas.

En el primer proyecto, se ha investigado el mecanismo y la selectividad del principal transportador de colesterol en mamíferos, el transportador Niemann-Pick C1-Like 1 (NPC1L1). El desequilibrio del perfil lipídico sanguíneo óptimo, causado por una desregulación de la homeostasis del colesterol, desempeña un papel clave en el desarrollo de las enfermedades cardiovasculares. El colesterol es una sustancia similar a la grasa con una importancia fisiológica fundamental en los seres humanos, y su homeostasis está estrechamente regulada por diversos procesos celulares, incluida la importación en el intestino delgado y la reabsorción en los conductos biliares por el importador NPC1L1. NPC1L1 puede mediar en la absorción de una variedad de esteroles, pero sorprendentemente muestra una gran sensibilidad a la epimerización del colesterol. Hemos revelado la existencia de dos modos de unión distintos para el colesterol y el epicolesterol al dominio N-terminal (NTD) de NPC1L1, proporcionando las características moleculares responsables de la diferencia de afinidad entre ambos sustratos. La configuración del grupo hidroxilo del colesterol y la presencia de una molécula de agua estructural, son características clave para la absorción efectiva. Además, confirmamos la existencia de un mecanismo de apertura implicado en la unión del esterol a este dominio de la proteína mediante los reordenamientos conformacionales del loop α8/β7. Por otra parte, se ha determinado que una serie de mutaciones puntuales que alteran el transporte de colesterol ejercen un efecto leve en la unión del colesterol al NTD, pero a su vez ejercen una influencia drástica en el rango conformacional del loop α8/β7 en la especie apo en comparación con la proteína no mutada. En conjunto, los resultados respaldan el papel funcional desempeñado por el loop α8/β7 en la regulación del acceso de ligandos a NPC1L1 y, por tanto, en la evaluación del impacto de estas mutaciones en enfermedades relacionadas con la alteración de la absorción de esteroles.

Palabras clave: transporte de colesterol, afinidad de unión, técnicas de muestreo avanzado, inhibidores de fusión, hemaglutinina, gripe A, selectividad, dinámica molecular, cálculos de energía libre relativa, diseño de fármacos basado en estructura

En el segundo proyecto, se ha investigado el modo de unión de un inhibidor antiviral frente la hemaglutinina (HA) del virus de la gripe A (IAV). En este proyecto, desvelamos el sitio de unión putativo y el modo de unión del inhibidor M090 basado en la pinanamina, capaz de bloquear el mecanismo de fusión de membrana mediado por HA en las primeras fases de la infección y activo frente a una amplia gama de variantes de HA. El sitio de unión previsto se localiza en la interfaz definida por la α-hélice $L80_2$-$A96_2$ y por los loops $P301_1$-$K318_1$ y $M59_2$-$L73_2$, es decir, muy cerca de la región principal del trímero de HA. Las predicciones de energía libre de unión relativa (RBFE) muestran una estrecha concordancia con las diferencias en la potencia inhibidora determinada en ensayos antivirales para M090 y sus derivados, proporcionando una validación *in silico* del modo de unión propuesto. Además, descubrimos que la mutación $E74_2 \rightarrow D$, resistente al fármaco, debilita la red óptima de enlaces de hidrógeno establecida entre los loops que forman el sitio de unión de M090. En conjunto, un cambio en el empaquetamiento global frente a la región del tallo de la HA conduce a una alteración del modo de unión que explica la pérdida de actividad.

En el tercer proyecto, revelamos el sitio de unión putativo y el modo de unión del derivado sulfonamida VF-57a, sintetizado por el grupo del Prof. Santiago Vázquez (Departament de Farmacologia, Toxicologia i Química Terapèutica, UB) y ensayado frente a células MDKC infectadas con IAV por el grupo de la Prof. Lieve Naesens (Department of Microbiology, Immunology and Transplantation, KUL). El compuesto muestra actividades submicromolares frente cepas A/H1N1 (EC_{50} = 0.31 μM y EC_{50} = 0.92 μM frente A/Virginia/ATCC3/2009 y A/Puerto Rico/8/1934, respectivamente) y cepas H5N1 (EC_{50} = 0.81 μM frente A/bald eagle/Florida/W22-134-OP/2022). Los resultados revelan que el sitio de unión del VF-57a se encuentra muy cerca del sitio de unión del inhibidor de amplio espectro Arbidol (Umifenovir). En este bolsillo, el anillo de tiofeno del VF-57a se estabiliza mediante contactos de van der Waals con $V36_1$, $L49_1$, $P308_1$, $F309_1$, $V55_2$ y $L99_2$ de un protómero y con $L98_2$ y $L102_2$ de un segundo protómero. El anillo que contiene el fragmento trifluorometilo se estabiliza mediante interacciones π con $Y94_2$ y $F309_1$. Además, las RBFE para VF-57a y algunos derivados del estudio SAR apoyan el modo de unión predicho al mostrar una buena correlación con las actividades experimentales.

Mediante la combinación de los últimos avances en métodos de química computacional con datos experimentales, esta tesis doctoral facilita la comprensión de ciertos trastornos fisiológicos relacionados con el colesterol y contribuye al desarrollo de nuevas moléculas como antivirales frente el IAV.

Palabras clave: transporte de colesterol, afinidad de unión, técnicas de muestreo avanzado, inhibidores de fusión, hemaglutinina, gripe A, selectividad, dinámica molecular, cálculos de energía libre relativa, diseño de fármacos basado en estructura

Table of contents

"Hearing I ask for from all the tribes,
both high and low, offspring of Heimdallar.
At Valfǫðr bidding, I will now speak the tales,
the oldest of all that I remember."

In ancient times, when Ymir lived,
there was neither sand nor sea nor cool waves.
No earth existed at all nor the sky above.
The gap was Ginnunga, and grass was there nowhere."

Vǫlva (Vǫluspá, stanzas 1, 3)

ABSTRACT ix
ABSTRACT (CATALAN) xi
ABSTRACT (SPANISH) xiii
TABLE OF CONTENTS xvii
LIST OF FIGURES AND TABLES xix
ABBREVIATIONS xxiii
1. Introduction 3
1.1. Proteins: The gears of life and beyond 3
1.2. Computational chemistry: where modern biomedicine and biotechnology meet 6
1.2.1. Molecular Dynamics (MD) 8
1.2.2. Steered Molecular Dynamics (SMD) 10
1.2.3. Well-Tempered Metadynamics (WT-MetaD) 12
1.2.4. Thermodynamic Integration (TI) 13
1.3. Two drug targets of public health concern 16
1.3.1. Niemann-Pick C1-Like 1 (NPC1L1) transporter 16
1.3.1.1. NPC1L1 metabolic function and mechanism of action 19
1.3.1.2. NPC1L1 structure 21
1.3.1.3. NPC1L1 inhibition 24
1.3.2. Influenza A virus (IAV) 26
1.3.2.1. IAV structure and genotypic classification 28
1.3.2.2. IAV viral cycle 31
1.3.2.3. IAV-HA inhibition overview 33
2. Objectives 39
3. Results 43
Informe dels directors 43
Project 1 49
Project 2 103
Project 3: in preparation 141
4. Discussion 233
4.1. *"Binding of Cholesterol to the N-Terminal Domain of the NPC1L1 Transporter: Analysis of the Epimerization-Related Binding Selectivity and Loop Mutations"* 233
4.1.1. NPC1L1-NTD protein dynamics upon cholesterol binding 234
4.1.2. Molecular features driving the sterol selectivity of the NPC1L1-NTD 235
4.1.3. Mutations impairing cholesterol transport alter the NTD dynamics 237

4.2. *"Mining druggable sites in influenza A hemagglutinin: Binding of the pinanamine-based inhibitor M090"* 239
4.2.1. M090 binding site and binding mode 239
4.2.2. In silico validation of the binding mode using free energy predictions 242
4.2.3. Effect of the $E74_2 \rightarrow D$ drug-resistant mutation 244
4.3. *"N-[(Thiophen-3-yl)methylbenzamides as Influenza Virus Fusion Inhibitors Acting on H1 and H5 Hemagglutinins"* 247
4.3.1. Benzamides as HA-mediated fusion inhibitors: SAR study 247
4.3.2. Pseudovirus entry and cell-cell fusion assays: Validation of VF-57a 249
4.3.3. VF-57a-resistant markers in H1 HA 250
4.3.4. VF-57a binding site and binding mode 252
5. Conclusions 257
Bibliography 263
APPENDIX I - Other publications 275
"Influenza A virus hemagglutinin: from classical fusion inhibitors to proteolysis targeting chimera-based strategies in antiviral drug discovery" 275

LIST OF FIGURES AND TABLES

Figure 1.1 Representation of the four main classes of macromolecules 3

Figure 1.2 Representation of the four levels of organisation used to describe the structure of proteins: *Primary structure*: the amino acid sequence; *Secondary structure*: the local arrangements (black dotted lines represent hydrogen bonds within the protein backbone); *Tertiary structure*: the global fold; *Quaternary structure:* the domain assembly 4

Figure 1.3 Scheme representing the main steps involved in CADD approaches 7

Figure 1.4 General expression of the energetic contributions in the AMBER force field 9

Figure 1.5 Scheme illustrating the binding and unbinding process of a protein-ligand complex using SMD in combination with US 11

Figure 1.6 Scheme illustrating the fundamental aspects behind Metadynamics Simulations 12

Figure 1.7 Scheme of the thermodynamic cycle used in Relative Binding Free Energy (RBFE) calculations and the application of Thermodynamic Integration (TI)..... 14

Figure 1.8 Scheme of the cholesterol journey through the body and the main roles of the small intestine and the liver. FC: free cholesterol; PL: phospholipids; BA: bile acids; CE: cholesterol ester 18

Figure 1.9 Representation of the organs and cell types where NPC1L1 is expressed, i.e. enterocytes in lumen of the small intestine and hepatocytes in the liver. The surface structure of NPC1L1 is shown in the centre..... 19

Figure 1.10 Representation of the NPC1L1-mediated enterocytic cholesterol transport..... 21

Figure 1.11 3D model of the full-length NPC1L1 (left) and schematic representation of its constituent domains (right) 22

Figure 1.12 Chemical structure of some of the known inhibitors of NPC1L1-mediated cholesterol absorption discussed in section 1.3.1.1..... 25

Figure 1.13 Summary of pandemic outbreaks of the last century from 1918 to 2009 great pandemic events. In 1977, the human H1N1 virus re-emerged and enough immunity in the population prevented an uncontrolled pandemic event. The 2009 pandemic was a recombinant virus including RNA from swine, avian and human viruses. All pandemic viruses have included gene segments from the 1918 virus represented in red 27

Figure 1.14 Representation of Influenza A virus and its major components. In the lipid envelope: the two viral glycoproteins HA and NA and the ion channel M2. The structural core of the virion consists of M1 proteins that interact with surface proteins and vRNPs. The eight gene segments form the vRNPs, which are composed of NP and the polymerase complex (PA, PB1, PB2). The nuclear export protein (NEP) is also present in the virion 28

Figure 1.15 Phylogenetic classification of HA and NA 31

Figure 1.16 Schematic representation of the IAV replication cycle 32

Figure 1.17 Representation of the monomeric (surface) and trimeric (cartoon) structures of HA. The two binding sites experimentally observed by X-ray crystallography are shown as the orange surface for Site A and as the blue surface for Site B, together with the molecular structures of the main co-crystallised inhibitors mentioned in this section 35

Figure 4.1 A) X-ray structure of the NTD of NPC1L1 (PDBid: 3QNT). Relevant residues are highlighted in gray sticks, L213 is highlighted in yellow, α8/β7 loop in blue and the α3 helix in red; B) Superposition of the X-ray structures of the apo NTD of NPC1L1 and the NTD-CHL complex of NPC1 (PDBid: 3GKI). CHL is highlighted in orange sticks; C) S-opening view for the superposition of X-ray structures. Q200 and N86 of NPC1 are highlighted in magenta sticks and the crystallographic water molecule on top of CHL's hydroxyl has been added for clarity; D) Surface representation of the NTD of NPC1L1 with CHL in the binding pocket 234

Figure 4.2 A) Representative snapshot of the *Q95 binding mode* for ECL. Q95 is highlighted in magenta. B) Representative snapshot of the *T128 binding mode* for ECL. T128 is highlighted in magenta; In both representations the α8/β7 loop is highlighted in blue, L213 is highlighted in yellow and the h-bond interaction between ECL and Q95/T128 is shown as black dots for clarity. Below each panel, the thermodynamic cycle considered for the RBFE calculations. The final averaged relative free energy values for each transformation are shown in kcal·mol-1 and the cycle closure error is shown in green 236

Figure 4.3 Initial conformation of M090 in the three different binding modes studied in this work. The front view is shown at the top of each panel, while the side view is shown at the bottom. The black dotted lines indicate the hydrogen bond between M090 and $T309_1$. The three explored binding modes are shown: BM-I (A), corresponds to the binding mode predicted by Zhao et al., BM-R (B) and BM-II (C) 240

Figure 4.4 A) Representative snapshot of the final binding mode achieved for M090 in the BM-II systems. The front view of the bag is shown above, while the side view is shown below.

B) Superposition of snapshots along the simulation for the four independent replicates of the BM-II systems 242

Figure 4.5 Final thermodynamic cycle used in the RBFE simulations. The numbering of the derivatives follows the notation used by Zhao et al. Green values denote the closure error of each inner cycle. The plot shows the comparison between the differences in binding affinity estimated from the experimental EC_{50} data and the relative free energy differences predicted from TI calculations. Bars denote the errors of calculated and experimental relative free energy changes. Dotted and dashed lines denote the 95% confidence and prediction intervals. All values (kcal·mol^{-1}) are relative to M090. MUE and RMSE stand for mean unsigned error and root-mean-square error, respectively 243

Figure 4.6 A) Representative snapshot of the predicted M090 binding mode (BM-II). H-BS, and L2-BS regions are highlighted in green and the main amino acids forming the hydrogen bond network are shown in sticks. Green regions account for one protomer while grey regions account for the adjacent protomer. B) Representation of the two main essential motions of the backbone atoms that shape the binding pocket for both the wild type and the E742 → D mutant 245

Figure 4.7 A) Dose-response curves for the pseudovirus entry and cell-cell fusion assays for measuring the inhibitory potency of VF-57a, RL-007 and Arbidol. B) Inhibitor binding response in the presence of HA-binding antibodies indicating whether HA is in the prefusion or in the postfusion conformation at physiological or acidic pH 249

Figure 4.8 Impact of HA mutations on viral sensitivity to VF-57a and RL-007 and on acid lability of HA. The hemolysis pH was defined as the pH where 50% of virus-induced hemolysis occurred relative to pH 5.0. Amino acid mutations are highlighted in yellow spheres in a model of A/Virginia/ATCC3/2009 HA. The mutations, emerging from the presence of VF-57a are highlighted in red, while the mutations emerging from the presence of RL-007 are highlighted in blue. The suffix refers to the HA1 (green) or HA2 (cyan) domain of HA 251

Figure 4.9 A) Initial conformation of (*S*)-F0045 and several snapshots along the MD simulation in one of the three pockets; B) Initial conformation of VF-57a in the BMA and several snapshots along the MD simulation in one of the three pockets; C) Initial conformation of VF-57a in the BMB and several snapshots along the MD simulation in one of the three pockets; D) H1 HA in complex with the anti-HA stem antibody C179 bound to Site B. X-ray structures of HA in complex with HA-mediated fusion inhibitors CBS1117, JNJ4776 and (*S*)-F0045 are included for clarity 252

Figure 4.10 A) Final conformation of VF-57a in Site B and several snapshots along the MD trajectories for each of the three pockets of the HA trimer; B) results of the predicted change in free energy in kcal· mol-1 for both directions of each transformation. 253

Table 4.1 SAR for the antiviral (A/H1N1) activity of the explored benzamides in MDCK cells..... 248

ABBREVIATIONS

(-)vRNA	Negative-polarity viral RNA
(+)cRNA	Positive-polarity complementary RNA
(+)mRNA	Positive-polarity messenger RNA
ABC	ATP-binding cassette
ACAT2	Acyl-CoA:cholesterol acetyltransferase
Acetyl-CoA	Acetyl coenzyme A
AI	Artificial Intelligence
BM	Binding Mode
CADD	Computer-aided Drug Design
CHL	Cholesterol
Cryo-EM	Cryogenic Electron Microscopy
CTD	C-terminal domain
CV	Collective variable
CVD	Cardiovascular Disease
DNA	Deoxyribonucleic acid
ECL	Epi-cholesterol
EGCG	Quercetin and epigallocatechin gallate
ER	Endoplasmic reticulum
ERC	Endocytic recycling compartment
FEP	Free Energy Perturbation
FES	Free Energy Surface
FP	Fusion peptide
HA	Hemagglutinin
h-bond	Hydrogen bond
HDL	High-density lipoprotein
HHL	25-hydroxycholesterol
HMG-CoA	3-hydroxy-3-methyl-glutaryl coenzyme A
IAV	Influenza A Virus
IV	Influenzavirus

LAIV	Live-attenuated influenza virus
LBDD	Ligand-Based Drug Design
LDL	Low-density lipoprotein
LDLR	LDL receptors
LIMA1	LIM domain and actin binding 1
LNL	Lanosterol
LRP	LDLR-related protein
M1	Matrix protein 1
M2	Matrix protein 2
MD	Molecular Dynamics
MDCK	Madin-Darby canine kidney
MetaD	Metadynamics
MLD	Middle luminal domain
MM	Molecular Mechanics
MMPBSA	Molecular Mechanics Poisson-Boltzmann Surface Area
MOA	Mechanism of action/Mode of Action
MSA	Multiple sequence alignment
MTP	Microsomal triglyceride transfer protein
MUE	Mean unsigned error
NA	Neuraminidase
NEP	Nuclear export protein
NMR	Nuclear magnetic resonance
NP	Nucleoprotein
NPC1	Niemann-Pick type C1
NPC1L1	Niemann-Pick C1-Like 1
NPDC	Niemann-Pick type C1 disease
NS2	Non-structural protein 2
NTD	N-terminal Domain
PA	Polymerase acidic protein
PB1	Polymerase basic 1
PB2	Polymerase basic 2

PC	Principal component
PCA	Principal component analysis
PCSK9	Anti-proprotein convertase subtilisin/kexin type 9
PDB	Protein Data Bank
PMF	Potential of Mean Force
QM	Quantum Mechanics
RBFE	Relative Binding Free Energy
RMSD	Root-mean-square deviation
RMSE	Root-mean-square error
RNA	Ribonucleic acid
SAR	Structure-activity relationship
SBDD	Structure-Based Drug Design
SCAP	Sterol regulatory element–binding protein cleavage-activating protein
SMD	Steered Molecular Dynamics
SPR	Surface Plasmon Ressonance
SR-BI	Receptor scavenger class B type I
SSD	Sterol-sensing domain
TI	Thermodynamic Integration
TIV	Trivalent inactivated virus
TMD	Transmembrane domain
TMs	Transmembrane helices
US FDA	United States Food and Drug Administration
US	Umbrella Sampling
vdW	van der Waals
VLDL	Very low-density lipoprotein
vRNA	Viral RNA
vRNP	Viral ribonucleoprotein
WHAM	Weighted Histogram Analysis Method
WHO	World Health Organization
WT-MetaD	Well-Tempered Metadynamics

Chapter 1

Introduction

"Then sought the gods, the mighty gods,
their assembly seats, and council held:
they gave names to night and its kindred,
called them morning and midday,
afternoon and evening, to count the years.

They played at tables in the meadow, were merry,
there was for them no whit of a want of gold.
Until three maidens of giants came,
immensely mighty, from Jǫtunheimr."

Vǫlva (Vǫluspá, stanzas 6, 8)

"Other biological sciences take up the study at other levels of organisation: dissecting the individual into organs and tissues and seeing how these work together, as in physiology; reaching down still further to the level of cells, as in cytology; and reaching the final biological level with the study of living molecules and their interactions, as in biochemistry. No one of these levels can be considered as more important than any other."

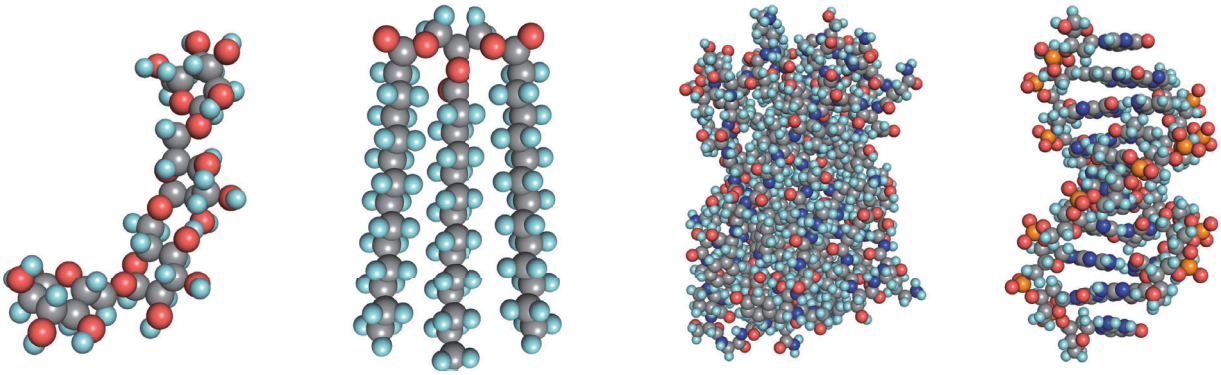
(Marston Bates, 1990)

1.1 Proteins: The gears of life and beyond

The secret of life lies in the interplay between biological macromolecules. Through decades of research, mankind has discovered and classified thousands of biological macromolecules into four main classes: carbohydrates, lipids, nucleic acids and proteins (Figure 1.1).[1]

Carbohydrates are an essential source of energy for the human body, playing a role in maintaining blood glucose and insulin balance, facilitating cholesterol and triglyceride metabolism and acting as substrates in fermentation processes. The microbiota present in our digestive tract has evolved to break down carbohydrates into glucose, which can be stored in the liver and muscle tissue until energy supply is needed in the body tissues.[2,3]

Lipids also provide metabolic energy and are the major structural components of cellular membranes. They are also involved in signalling and metabolic regulation, as shown by the effect triggered by the changes in the concentration of certain lipids and their derivatives. For example, prostaglandins are involved in arterial pressure regulation, coagulation and inflammatory response.[4]



Carbohydrate (isomaltotetraose) **Lipid (triacylglycerol)** **Protein (enzyme)** **Nucleic acid (DNA)**

Figure 1.1 Representation of the four main classes of macromolecules.

Nucleic acids, known as deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), encode the blueprints for life and other biological entities not classified as living such as viruses and prions. In the case of humans, our genome is packed into 23 pairs of chromosomes in the cell nucleus, plus an extra small amount in the mitochondria. DNA has the famous shape of a double-helix polymer made up of monomeric units called nucleotides, which in turn are made up of a 5-carbon sugar deoxyribose, a nitrogenous base and one or more phosphate groups. RNA usually is comprised a single chain of nucleotides containing the 5-carbon sugar ribose, a phosphate and a nitrogenous base.[5] Nucleic acids encode genetic information needed to produce proteins.

Proteins are the cellular machinery responsible for the wide diversity of molecular functions found in all known living organisms, as well as non-living forms such as viruses.[6] All these functions are made possible by the precise three-dimensional arrangements that proteins adopt, which are held together by hydrogen and disulfide bonds, salt bridges between charged amino acid side chains, and van der Waals (vdW) forces.[7] Three levels of organisation are used to describe the structure of a single protein chain, also named *monomer*. The first level is known as the *primary structure*, in which a continuous chain of building blocks called *amino acids* are linked by a covalent bond between the last carbon atom and the nitrogen atom of the next amino acid, known as peptide bond (Figure 1.2).[8] There are twenty natural amino acids that encode for proteins, with different side chains linked to the central carbon atom, called Cα. The repeated motif N-Cα-C is known as the backbone of the protein and each amino acid is named according to the side chain they have.[9]

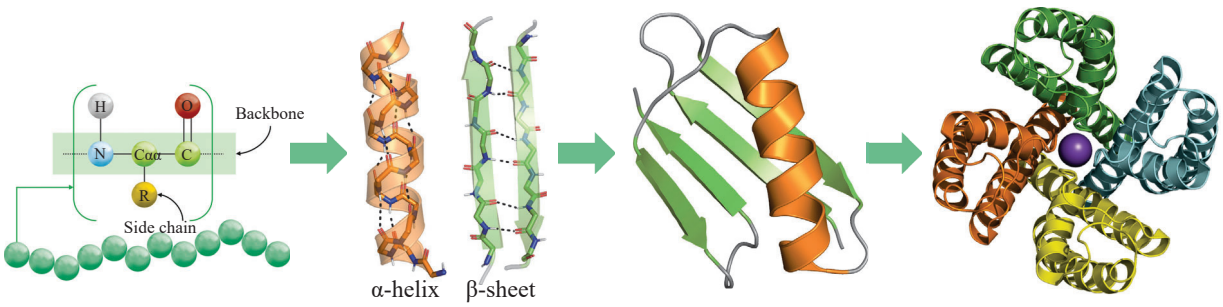


Figure 1.2 Representation of the four levels of organisation used to describe the structure of proteins: *Primary structure*: the amino acid sequence; *Secondary structure*: the local arrangements (black dotted lines represent hydrogen bonds within the protein backbone); *Tertiary structure*: the global fold; *Quaternary structure:* the domain assembly.

The second level emerges when the primary structure folds into local 3D structures, giving rise to the *secondary structure*. Two main secondary structures exist: *α-helices* and *β-sheets*. The latter are formed by the parallel arrangement of multiple *β-strands*.[10] These structures, connected between them by flexible unstructured portions of sequence called *loops*, are further stabilised in nature by hydrogen bonds between vicinal amide moieties and disulphide bonds. Finally, all these secondary structures fold back on themselves to form the final three-dimensional level of structural order, known as the *tertiary structure*, which is determined by the spatial position of the secondary elements and the chemical interactions between the individual amino acid side chains and the chemical environment.[11]

There is a fourth level of structural organisation known as the *quaternary structure*, that arises when there is a spatial association of multiple tertiary structures held together by molecular interactions. The strength of these interactions determines the overall structural flexibility of the tertiary and quaternary structures leading to several possible spatial arrangements, or conformations, which are usually fine-tuned for their molecular function. The dynamical ensemble of conformations enables proteins to participate in specific molecular processes.

Protein function is usually the result of the interaction with other proteins and/or small molecules.[12] For example, enzymes are complex proteins that act as highly specific biological catalysts accelerating the biochemical reactions.[13] The importance of proteins in biological functions, reflected in health and disease, has led to the development of a wide variety of disciplines such as biotechnology, biomedicine or biochemistry, which aim to better understand the mechanisms behind protein function. Many diseases arise from changes in protein-substrate interactions, often caused by mutations that disrupt the recognition and binding affinity between protein and other biomolecules.[14] In this context, molecular recognition becomes a fundamental step in any biological process, such as cellular signalling, enzymatic catalysis or protein-protein association, to name a few. In all these processes, molecular recognition is key in determining the final outcome of a protein or set of protein functions, leading to their association or rejection, depending on the nature of the process. Often, molecular association events between proteins or protein-ligand systems lead to a specific biological function. On the other hand, molecular rejection events often reflect biomolecular selectivity and specificity in biological systems.

Due to the complexity of biological systems, multidisciplinary approaches are needed to fully understand the intricacies of protein function. Computational approaches, such as Quantum Mechanics (QM), Molecular Dynamics (MD) and Molecular Docking, provide atomistic information that can complement experimental insights. For example, they can predict the behaviour of molecules, elucidate reaction mechanisms and identify potential drug candidates, thereby guiding and improving drug discovery projects. However, computational approaches need to be experimentally validated using a wide range of experimental techniques, such as X-ray crystallography, binding assays or NMR spectroscopy. Therefore, the combination of experimental and computational approaches has become essential in modern drug discovery pipelines.[15] This thesis showcases the potential of using computational tools to elucidate the mechanism and target biologically relevant proteins with small molecules.

1.2 Computational chemistry: where modern biomedicine and biotechnology meet

With the speed up of computer technology in recent decades, the predictive understanding between amino acid sequence and protein structure has dramatically improved leading to great advances in biotechnology and biomedicine. For example, we are now able to design *de novo* protein structures and complexes,[16,17] some of which with promising therapeutic applications.[18] Advances in software development and the emergence of deep learning algorithms have not only accelerated the growth of experimentally determined protein structures in the Protein Data Bank (PDB),[19] but have also enabled the prediction and classification of druggable proteins, boosting the *in silico* drug discovery and the treatment of genetic diseases.[20]

A recent example of the most relevant computational advances related to protein structure prediction is AlphaFold (being AlphaFold 3 the latest version).[21–23] This artificial intelligence (AI) program, developed by DeepMind, is able to predict the 3D structure of proteins from their amino acid sequence and has solved thousands of previously unknown protein structures that were extremely difficult or impossible to solve experimentally with our current technology.[24] However, it requires high-quality multiple sequence alignments (MSAs) as input and computational power to retrieve accurate structures close to the experiment from X-ray or cryo-EM analyses. While AlphaFold2 was developed, other open-source protein structure prediction tools were introduced, being RoseTTAFold developed by the Baker Lab at the University of Washington the most relevant.[25] Designed to run on standard laboratory hardware with GPU support, RoseTTAFold provides a more accessible and flexible alternative, although it is not as accurate in single protein prediction. On the other hand, it has been shown to outperform

AlphaFold when the input MSAs are sparse and is able to predict protein-protein interactions and even protein complexes, making it particularly useful for broader applications in the fields of protein design and synthetic biology.[26,27] Recognising the outstanding value of such tools, Dr. Demis Hassabis and Dr. John Jumper from DeepMind as well as Prof. David Baker of the University of Washington were awarded the Nobel Prize in Chemistry in 2024 for their contributions to protein structure prediction and computational protein design, respectively.

In the context of drug discovery, understanding the interactions between drug-like compounds and their target receptors is fundamental. Computational methods have revolutionised the field giving rise to Computer-aided Drug Design (CADD) approaches for the discovery, development and analysis of potential drug candidates (Figure 1.3).[28,29] CADD can be divided in two areas: i) Ligand-Based Drug Design (LBDD), which involves the analysis and screening of library of compounds that are known to interact or potentially interact with a target of interest of which we have no structural information, and ii) Structure-Based Drug Design (SBDD), which involves the design and optimisation of potential drug candidates, using the structure of protein of interest to determine or identify complexes between the screened analogues.

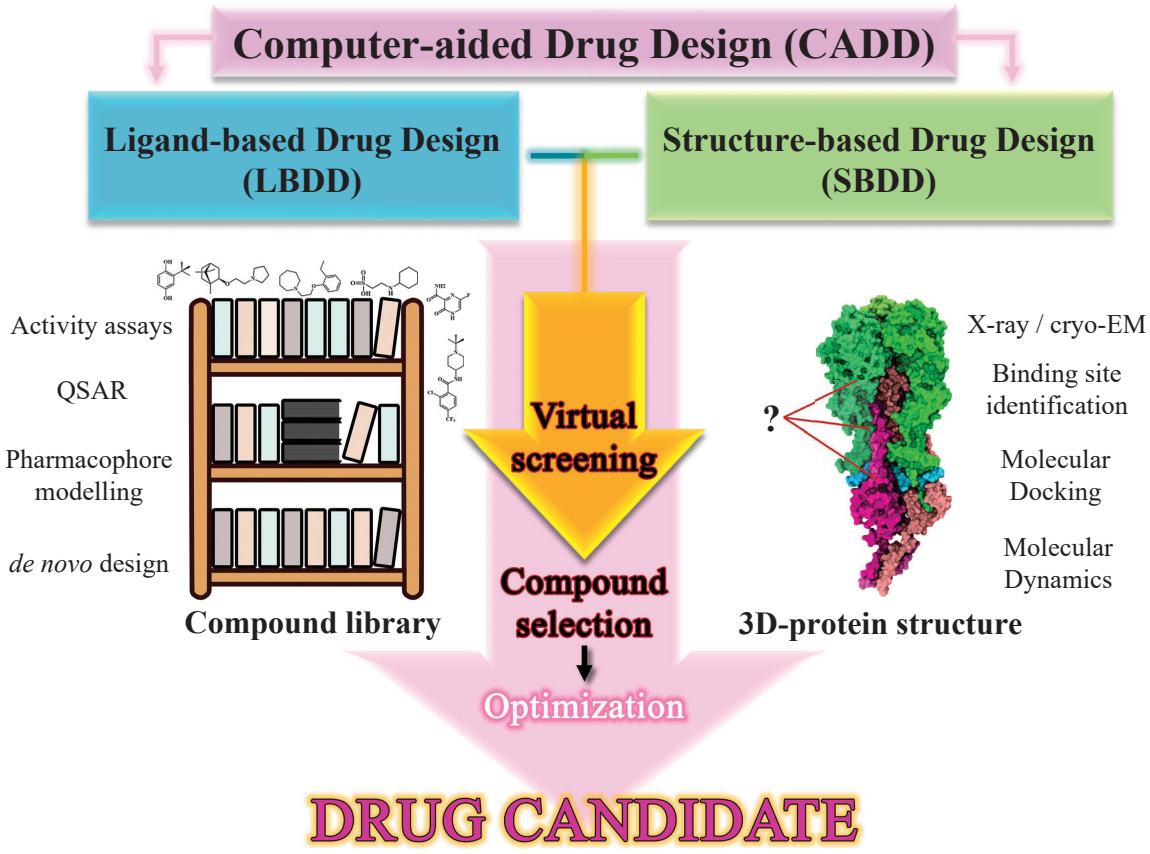


Figure 1.3 Scheme representing the main steps involved in CADD approaches.

Most CADD projects start at the lead generation stage, when a library of potential hits is identified and handed to chemists for synthesis, optimisation and transfer to biochemists and pharmacologists for activity assays. During this stage, chemical libraries are compiled to obtain Quantitative Structure-Activity Relationship models (QSAR) using computational virtual screening approaches.[30] In these models, different types of scoring functions are used in combination with known molecular properties, such as measured activities or structural information of the target. Two scenarios are then possible. When the structure of the target is unknown (LBDD), the experimental activity data can be used to optimise the scoring functions and further filter the library, modelling a pharmacophore containing key molecular features such as hydrophobic groups, hydrogen bond donors or acceptors, etc. and using it as a molecular template to screen for compounds with similar physicochemical properties. The other possibility is to have structural information of the target in advance (SBDD), by experimental techniques (X-ray crystallography or cryo-EM experiments) or computational tools such as homology modelling and the protein structure prediction tools mentioned earlier.

The structural information enables the application of a wide variety of methods such as Molecular Docking, *de novo* compound generation, or fragment-growing methods. Molecular Docking is based on virtually positioning the library of potential hits in the 3D binding site of the target and using a scoring function to rank them according to a predictive activity scale.[31] The *de novo* compound design methods are able to generate new molecules from scratch based on the chemical environment of the target protein.[32] Also, fragment growing strategies are able to construct the ligand in the binding site by joining experimentally identified fragments or building blocks with the desired molecular properties.[33] All these strategies share the same counterparts, namely the scoring functions and their limited transferability between different protein targets. The identified putative binders must be purchased or synthesised for further experimental assays. In this thesis, I focused on the use of SBDD methods for two protein targets of biomedical interest. To illustrate the variety and potential to these methods, I will briefly describe the main computational techniques I used in my work.

1.2.1 Molecular Dynamics (MD)

For large systems such as proteins, the methods based on Molecular Mechanics (MM) have become the standard approach in the field. In this context, molecules are described by a *force field*, i.e. theoretical constructs used to calculate their potential energy, describing them as a collection of classical particles bound together by virtual harmonic springs and neglecting

the explicit treatment of electrons.[34] These force fields are explicit energy functions that describe bonded (bond stretching, angle bending and dihedral torsions) and non-bonded (vdW and electrostatic interactions) contributions for each classical particle in the system (Figure 1.4 and Equations 1.1, 1.2, 1.3 and 1.4). It allows to model atom motion by calculating the forces between the particles and their potential energies using the aforementioned force field equations. This approach is known as Molecular Dynamics (MD),[35] where the conformational sampling of structures is used to calculate the potential energy of the system and its evolution over time (i.e., a trajectory). The sum of all the potential energies involved in each time step of a MD simulation gives rise to a Hamiltonian, in which the bonded and non-bonded interactions between particles are described by analytical expressions. There is a whole field within computational chemistry concerned with improving force fields to describe biological systems and small molecules.[36]

In the context of this thesis, I have used the AMBER force field,[37] which excels in obtaining the conformational energies of complex biological systems and consists of four general terms, which are illustrated in Figure 1.4.[38] In this model, the interactions between bonded atoms are described by simple harmonic expressions for stretching and bending (Equation 1.1 and 1.2, respectively), and a combination of cosine functions to describe the torsional potential energy (Equation 1.3). On the other hand, the non-bonded interactions are described by a 6-12 Lennard-Jones potential expression[39,40] for the vdW interactions and a Coulombic expression for the electrostatic interactions between particle-centred (atom model) point charges (Equation 1.4).[41] Ultimately, the choice of a force field depends on the nature and characteristics behind the simulated system. This is the fact behind the catalogue of specific force fields designed to treat different kinds of systems such as proteins, DNA and RNA,[42] carbohydrates,[43] lipids,[44] organic molecules,[45] ions[46] and metal-containing complexes.[47] This method provides close and validated approximations at an affordable cost and with a high degree of reliability, making it possible to study complex biological systems at the atomic level.

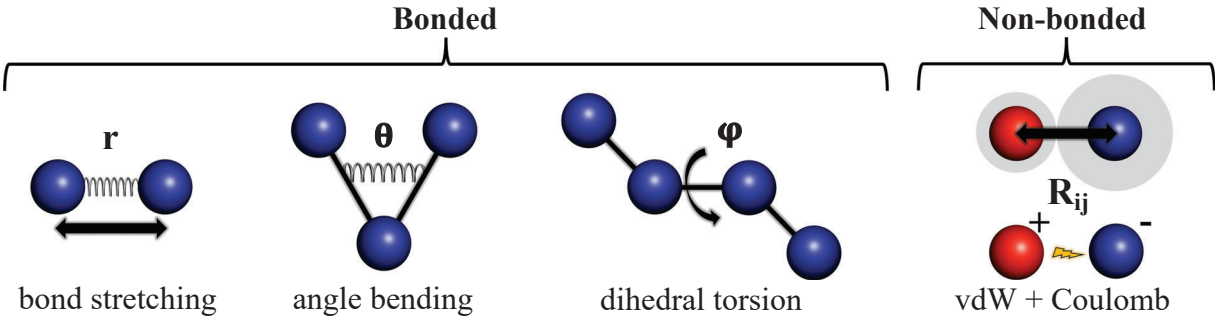


Figure 1.4 General expression of the energetic contributions in the AMBER force field.

$$E_{bonds} = \sum_{bond} K_r \left(r - r_{eq}\right)^2 \tag{1.1}$$

$$E_{angles} = \sum_{angle} K_\theta \left(\theta - \theta_{eq}\right)^2 \tag{1.2}$$

$$E_{dihedrals} = \sum_{dihedral} \frac{V_n}{2}[1 + cos(n\varphi - \gamma)] \tag{1.3}$$

$$E_{vdW} + E_{coulomb} = \sum_{i<j} \left[\frac{A_{ij}}{R_{ij}^{12}} - \frac{B_{ij}}{R_{ij}^{6}} + \frac{q_i q_j}{\in R_{ij}}\right] \tag{1.4}$$

MD simulations are an excellent tool for obtaining conformational samples of protein-ligand complexes as well as their energetic components. However, some processes relevant to understanding these interactions may be slower than the time scale provided by the technique, typically on the order of nanoseconds. For example, there may be conformational rearrangements in the protein that allow or restrict access to the binding site, but the energetic barrier may be high enough to require very long simulation times, such as microseconds or even longer. In these cases, we need to resort to enhanced sampling techniques, which are methods that use approximations to explore processes with timescales beyond those provided by MD. There are two types of enhanced sampling techniques: those that use reaction coordinates or collective variables and those that consider the endpoint states of the process. Reaction coordinates are variables that describe the collective motions by which the event proceeds. I will briefly describe the methods used in this thesis: Steered MD, Metadynamics and Relative Binding Free Energy (RBFE) calculations.

1.2.2 Steered Molecular Dynamics (SMD)

In the context of a protein-ligand complex, reaction coordinates are usually geometric descriptors that account for relevant processes of the system, such as the distance between the ligand and the binding pocket. These set of reaction coordinates must account for all the slow degrees of freedom of the process. Then we can calculate the free energy profile along the reaction coordinate(s) through a Boltzmann-weighted averaging over the system degrees of freedom other than the ones defining the reaction coordinate(s). This process generates the Potential of Mean Force (PMF).[48] The motion along the reaction coordinate is then approximated as a diffusive motion on the effective potential determined by the PMF.

When a high energy barrier limits the sampling along the reaction coordinate, an external force or bias can be applied to accelerate the process and combined with conventional MD simulations. In Steered Molecular Dynamics (SMD),[49] the applied bias has the form of a force vector acting as a virtual spring. The applied force is measured over the simulation time and recorded as the force distribution at different conformational states of the system (Figure 1.5). This description follows a non-equilibrium process whereas the PMF is an equilibrium property. Therefore, some numerical treatment must be applied in order to obtain the free energy profile of the process of interest. This treatment receives the name of Jarzynski's equality (Equation 1.5),[50] where an exact relation between free energy differences and the work done through equilibrium processes is correlated:

$$\Delta F = -\frac{1}{\beta} ln\langle e^{-\beta W}\rangle \tag{1.5}$$

where $\beta = 1/k_B T$, with $k_B T$and T being the Boltzmann constant and temperature. In addition to obtain the free energy profiles of the process, SMD can be combined with other techniques such Umbrella Sampling (US). This method splits the reaction coordinate in several windows, at each of which a biased MD simulation is usually performed with harmonic potentials.[51] The windows are then combined by methods like the Weighted Histogram Analysis Method (WHAM)[52] or umbrella integration.[53]

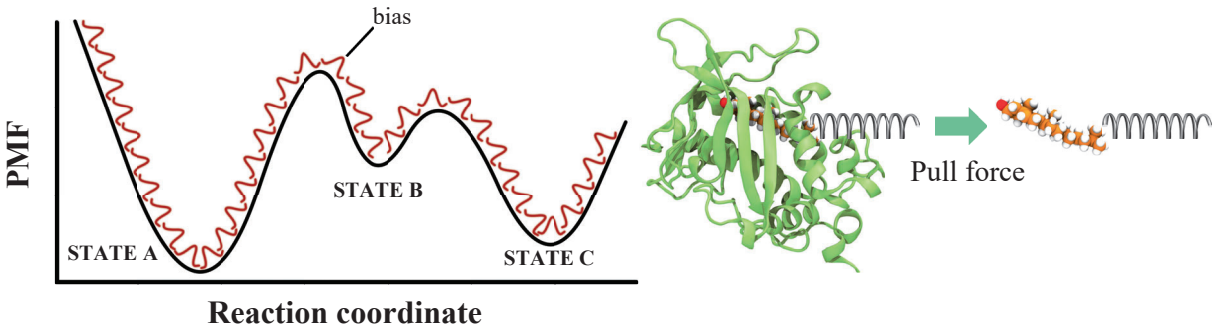


Figure 1.5 Scheme illustrating the binding and unbinding process of a protein-ligand complex using SMD in combination with US.

1.2.3 Well-Tempered Metadynamics (WT-MetaD)

There are cases where the desired process cannot be described as one-dimensional reaction coordinate. For example, studying the effect of conformational rearrangements that an enzyme undergoes over time on ligand binding. However, it is not always possible to determine the timescales of these processes, as several energetic barriers may have to be overcome to sample all the states that make up the mechanism of this protein-ligand complex. In these cases, we can facilitate the conformational sampling by introducing an additional bias potential (or force) acting on selected CVs over time, therefore promoting the spontaneous evolution of the process in the context of an MD simulation. This method is called Metadynamics (MetaD).[54]

MetaD introduces an external history-dependent bias potential to the Hamiltonian of the system as function of the CVs (Figure 1.6). This potential is described by the sum of Gaussian functions deposited at certain time intervals in an additive manner at the location of the CVs (Equation 1.6), avoiding the system to revisit sampled regions:

$$V(\vec{s},t) = \sum_{k\tau<t} W(k\tau)\, exp\left(-\sum_{i=1}^{d} \frac{\left(S_i - S_i(q(k\tau))\right)^2}{2\sigma_i^2}\right) \tag{1.6}$$

where τ is the Gaussian deposition stride, σ_i the width of the Gaussian for the ith CV, and $W(k\tau)$ the height of the Gaussian. As time progresses, minima are slowly filled, forcing the system to explore other configurations of the Free Energy Surface (FES) while maintaining spontaneous evolution.[55] This method allows for the exploration of the lowest free-energy

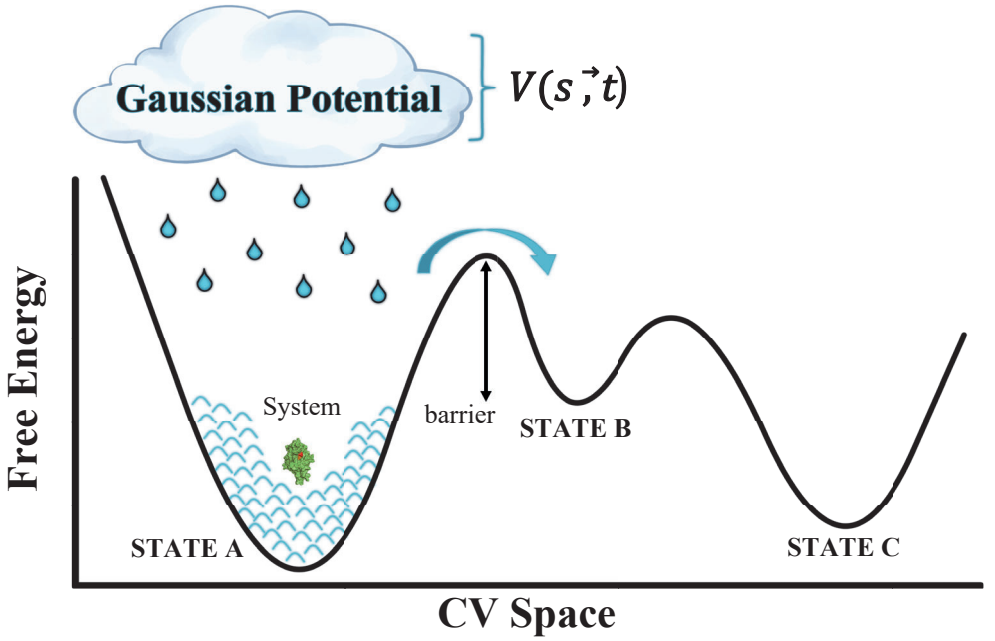


Figure 1.6 Scheme illustrating the fundamental aspects behind Metadynamics simulations.

minimum path of the studied process. When the exploration is complete and all the energetic barriers are overcome, the sampling of the reaction coordinate becomes diffuse, i.e. convergence is achieved.

The bias potential constructed by the sum of the Gaussian functions provides an accurate unbiased estimate of the free energy of the process. However, in the initial implementation of MetaD, the fixed height of the Gaussian functions can overfill some of the local minima and lead to an inaccurate representation of the FES. This overfilling can prevent the system from properly exploring these local minima, resulting in simulation artifacts. In addition, the continuous addition of a bias can hinder convergence by moving the system too vigorously through conformational space. To address these drawbacks, several variations of the original MetaD approach have been developed to improve the control over convergence and avoid such artifacts. In this thesis, we used the Well-Tempered Metadynamics approach (WT-MetaD)[56,57] where the height of the Gaussian potentials varies and becomes smaller over time as one approaches the edge of the wells in the FES (tempering), thus avoiding the overfilling problem. It also reduces the rate of Gaussian addition as the system approaches equilibrium, overcoming the drastic conformational changes near well transitions. In addition, a more aggressive exploration of the FES is performed at the beginning of the simulation, preventing some regions from being unexplored and refining the free energy estimate of regions already visited. As the tempering also operates at this initial stage, it prevents the introduction of unrealistic transitions between FES wells. Altogether, it improves the computational efficiency of the MetaD method.

1.2.4 Thermodynamic Integration (TI)

The power of predictive tools such as Molecular Docking and MD simulations has been discussed in the context of predicting the binding modes of potential drug candidates. Despite the utility of Molecular Docking in this area, its high efficiency is offset by its accuracy, which depends on the robustness of the scoring functions that define the algorithms. The score provided by Molecular Docking approaches is known to be quite inaccurate, particularly when comparing molecules with subtle differences in binding affinity. This scenario is common when comparing molecules with the same molecular core but with different substituents or functional groups, resulting in affinity differences below the tenfold range.[12] MD simulations offer a significant improvement over Molecular Docking techniques in that they can sample the conformational space of a protein-ligand complex. However, due to time scale limitations, it is not possible to obtain accurate binding affinities. This obeys to the fact that binding affinity

calculations involve integration over all possible states of the system to correctly sample the entire free energy landscape and obtain a sufficiently accurate prediction comparable to experimental data.

To overcome these limitations, computational methods based on alchemical free energy transformations, such as Free Energy Perturbation (FEP) and Thermodynamic Integration (TI), have emerged as standards of accuracy in the context of predicting accurate binding affinities.[58] In terms of protein-ligand binding affinities, these methods transform a ligand A into a ligand B by applying a series of perturbations to the force field parameters describing each molecule, following a thermodynamic cycle as shown in Figure 1.7. The relative free energy change associated with the binding of the ligands ($\Delta\Delta G$) is then obtained by performing the transformation in the ligand-bound complex and in the unbound state as shown in Equation 1.7 (i.e. with ligands in the solvent).

$$\Delta\Delta G = \Delta G^{B}_{binding} - \Delta G^{A}_{binding} = \Delta G^{A\to B}_{complex} - \Delta G^{A\to B}_{solvent} \tag{1.7}$$

The perturbations are applied gradually by introducing a dimensionless coupling parameter, λ, which serves as a weight constant for the interconversion of the force field parameters of ligand A into ligand B. Therefore, at $\lambda = 0$, the system corresponds to ligand A and at $\lambda = 1$ the system corresponds to ligand B. This process is known as Relative Binding Free Energy (RBFE) calculation. Using this methodology, the free energy changes along the thermodynamic pathway can be accurately calculated using Statistical Mechanics principles and including the force field components as well as entropic contributions arising from

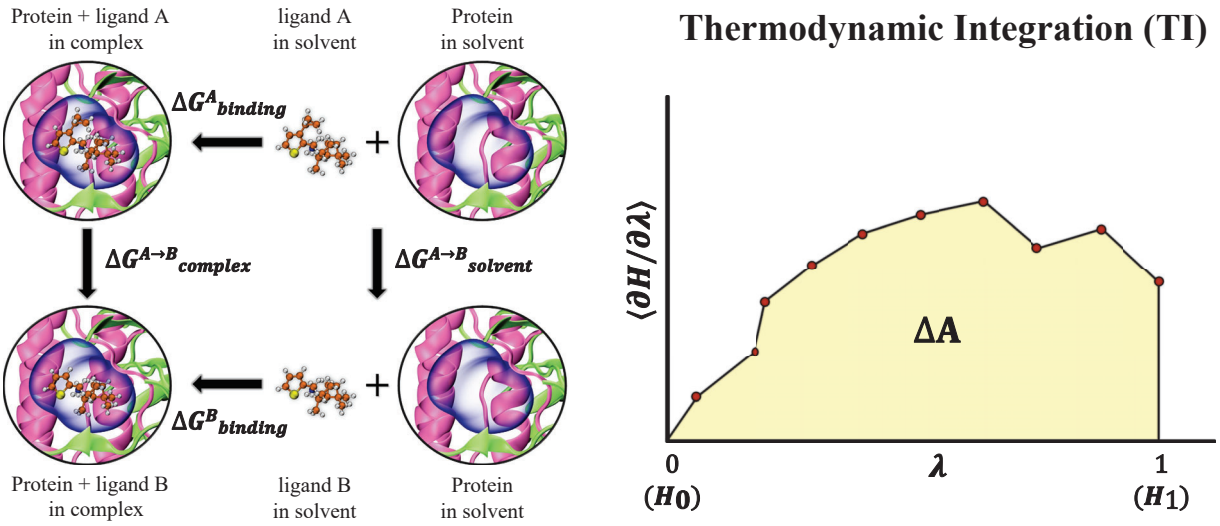


Figure 1.7 Scheme of the thermodynamic cycle used in Relative Binding Free Energy (RBFE) calculations and the application of Thermodynamic Integration (TI).

conformational sampling. In addition, the techniques can be used to study protein mutations following the same principle.

In this thesis, I used the Thermodynamic Integration (TI) approach, based on the evaluation of the free energy difference between the initial and final states of the transformation defined by a Hamiltonian (H_0 and H_1 in Figure 1.7). Since the change is done in the MD simulation context, total potential energies are calculated (U). The relative free energy change of the alchemical transformation is obtained by evaluating the integration area defined by all the intermediate Hamiltonians with respect to λ, $\langle \partial H/\partial \lambda \rangle$ (Figure 1.7 and Equation 1.8):

$$H_\lambda = \lambda H_1 + (1-\lambda)H_0 \tag{1.8}$$

For clarification, Helmholtz free energies (A or F) are predicted by TI, since such simulations are commonly performed under canonical (NVT) ensemble conditions, where the Hamiltonian corresponds to the Helmholtz free energy expression.[59] The Helmholtz free energy is expressed as:

$$A = U - TS \tag{1.9}$$

where U is the internal potential energy, T is the system temperature and S is the entropy. A is directly related to the partition function Z of the canonical ensemble calculated from the Hamiltonian since:

$$A = -k_B T \ln Z \tag{1.10}$$

where k_B is the Boltzmann's constant. The interconversion to Gibbs free energy is then trivial considering the canonical ensemble conditions since:

$$G = A + PV \tag{1.11}$$

where G is the Gibbs free energy, P is the pressure and V is the volume. Although PV contributions are present in the Gibbs expression, they are considered small for liquids and solids at constant temperature and pressure (incompressible). Therefore, the equality between the Helmholtz and Gibbs free energies can be taken as a good approximation in the context of MD simulations performed on the canonical NVT ensemble in condensed phases (Equation 1.12).

$$\Delta A_{A\to B} = \int_0^1 \left(\frac{\partial A}{\partial \lambda}\right) \partial\lambda = \int_0^1 \left\langle \frac{\partial U}{\partial \lambda} \right\rangle_\lambda \partial\lambda = \sum_{\lambda=0}^{1} \left\langle \frac{\partial H}{\partial \lambda} \right\rangle_\lambda \Delta\lambda \tag{1.12}$$

"I think that we shall be able to get a more thorough understanding of the nature of disease in general by investigating the molecules that make up the human body, including the abnormal molecules, and that this understanding will permit...the problem of disease to be attacked in a more straightforward manner such that new methods of therapy will be developed."

(Linus Pauling, 1960)

1.3 Two drug targets of public health concern

The constant development of molecular modelling methods contributes to addressing the challenges in the fields of biomedicine and drug design, such as the lack of structural and mechanistic information behind protein dysfunction and drug resistance.[60] In this thesis, I have employed state-of-the-art computational techniques to understand two known drug targets for two major health problems discussed in the following sections: cardiovascular diseases (CVD) and infectious diseases. First, I have studied the mechanism and selectivity of the main cholesterol transporter in mammals, the Niemann-Pick C1-Like 1 transporter (NPC1L1). Second, I have explored new inhibition strategies against the Influenza A virus (IAV) specifically targeting Hemagglutinin (HA).

1.3.1 Niemann-Pick C1-Like 1 (NPC1L1) transporter

A recent study based on the reported mortality data from the World Health Organisation (WHO) (Global Burden of Disease 2019 including statistics from 1990 to 2019)[61] forecasts that between 2025 and 2050 the cardiovascular prevalence will increase a 90.0%, resulting in 73.4% increase in crude mortality, with 35.6 million cardiovascular-related deaths expected in 2050 (from 20.5 million in 2025).[62] In 2019, over 65% of the global burden was caused by non-communicable diseases, with Cardiovascular Diseases (CVD) (ischaemic heart disease and stroke) leading the ranking. Among the risk factors attributed to CVD, genetic predisposition, age and gender are considered to be non-modifiable factors that cannot be influenced by either therapy or lifestyle changes. On the other hand, other common factors directly associated with the prevalence of CVD include smoking, pollution, excessive alcohol consumption, chronic stress, physical inactivity, sleep disturbances, poor diet, obesity, diabetes, high blood pressure and high plasma cholesterol levels.[63] All of these factors are known as modifiable risk factors and can be addressed by adopting a healthy lifestyle that systematically reduces the overall risk. CVD is actually a combination of related conditions affecting the heart and blood vessels,

caused by a reduction on the blood flow into the heart and brain, resulting in a lack of oxygen to vital organs.[64] This situation is usually the result of a blockage caused by the build-up of fatty deposits on the inner walls of the blood vessels, a syndrome known as atherosclerosis.[65] The imbalance of the correct blood lipid profile plays a key role in the development of CVD. Elevated levels of triglycerides and low-density lipoprotein (LDL)-cholesterol have been shown to contribute to the progression of atherosclerosis by infiltrating and accumulating in the layers that make up the arteries, leading to an inflammatory response that worsen the vessel blockage.[66]

The cause for altered cholesterol blood levels is the dysregulation of cholesterol homeostasis. The efficient biosynthesis, transport and metabolism of cholesterol depends on complex regulation system. Malfunctions and defects in the proteins that regulate these systems, such as the 3-hydroxy-3-methyl-glutaryl coenzyme A (HMG-CoA) reductase, are involved in the etiology of the disease.[67] In mammals, cholesterol plays a key role as structural component of cell membranes, maintaining their integrity, fluidity, and permeability. It is essential for membrane trafficking, signalling and cell proliferation; it is also the precursor for the biosynthesis of steroid hormones, bile salts and vitamin D. Cholesterol can be synthesised in the liver, involving a complex combination of nearly 30 enzymatic reactions with acetyl coenzyme A (acetyl-CoA) as the starting point.[68] The liver also plays a crucial role in sorting LDL-cholesterol from blood and converting it into bile salts, which increase its solubility and allow it to be excreted efficiently when there is an excess, thereby facilitating the digestion of dietary fats.

Figure 1.8 shows a schematic representation of the biogenesis, absorption, and recycling processes of cholesterol through the body. Once cholesterol has entered enterocytes, it must be sorted at the endoplasmic reticulum, esterificated by the acyl-CoA:cholesterol acetyltransferase 2 (ACAT2), and then packaged in a microsomal triglyceride transfer protein (MTP)-dependent manner into very low density lipoproteins (VLDL). VLDL will be further grouped in a class of lipoproteins known as chylomicrons, which will be secreted into the circulation via the lymphatic system.[69] In the enterocyte, cholesterol can also be pumped back into the intestinal lumen by the efflux transporter ATP-binding cassette (ABC) transporters G5 and G8 (ABCG5/ABCG8) or further transported to the basolateral membrane for conversion by ABCA1 to high-density lipoprotein (HDL), also known as 'good cholesterol' because it transports excess cholesterol to the liver for further metabolism and elimination.[70] However, due to high energetic cost of cholesterol biosynthesis, the intake of dietary cholesterol, i.e. a

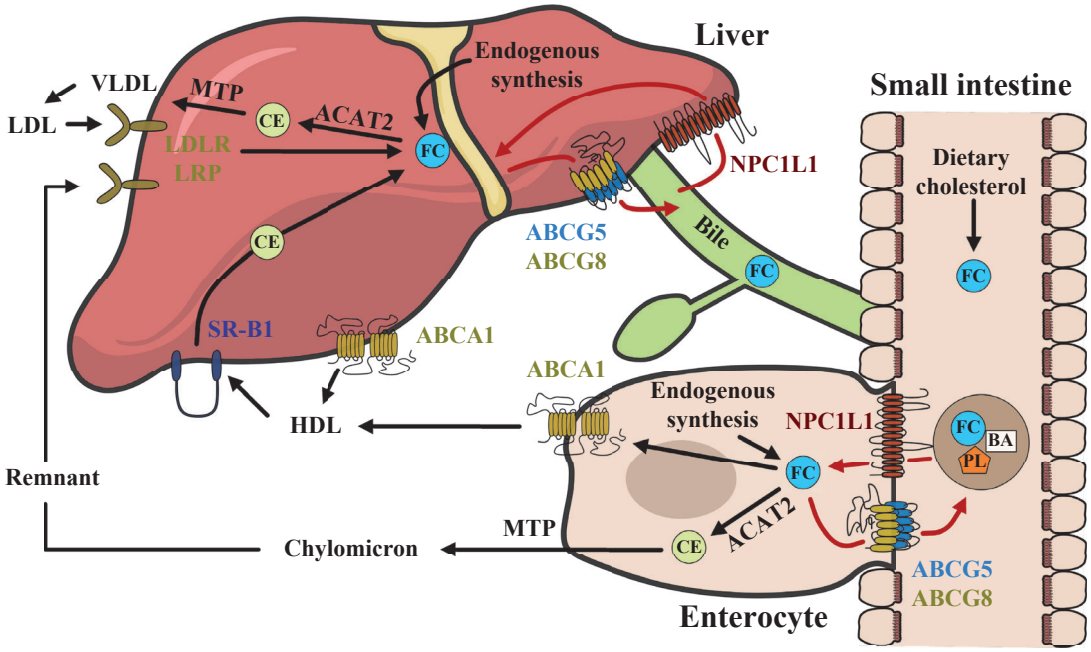


Figure 1.8 Scheme of the cholesterol journey through the body and the main roles of the small intestine and the liver. FC: free cholesterol; PL: phospholipids; BA: bile acids; CE: cholesterol ester.

process of cholesterol absorption in the intestinal lumen, contributes to save energy that can be used in other metabolic pathways. The biosynthetic production and intestinal absorption of cholesterol work together in such a way that high cholesterol levels induce a down-regulation of its biosynthesis, preventing over-accumulation and ensuring homeostasis. Furthermore, the conversion to bile acids by the liver for its secretion into the intestine can be reverted by reabsorption, facilitating the reuptake of cholesterol and reducing the need for constant biosynthesis.[71]

The need for new drugs capable of lowering LDL-cholesterol led to the discovery of ezetimibe (Zetia; SCH 58235; 1-(4-fluorophenyl)-(3R)-[3-(4-fluorophenyl)-(3S)-hydroxypropyl]-(4S)-(4-hydroxyphenyl)-2-azetidinone),[72] a drug approved in 2002 by the US FDA for the treatment of hypercholesterolaemia (alone and in combination with statins), that acts as a potent inhibitor of the intestinal cholesterol absorption.[73,74] The empirical evidence led to the conclusion that the physiological pathway controlling this process was a protein-mediated mechanism. The next steps were to find the key dedicated protein(s) governing this absorption process with the only hints of being the target of ezetimibe and probably highly expressed in the cells of the intestinal lumen. This search disclosed the Niemann-Pick C1-Like 1 (NPC1L1) transporter, highly expressed in the small intestine and in the liver (Figure 1.9), as the principal importer of free cholesterol in mammals.[75,76]

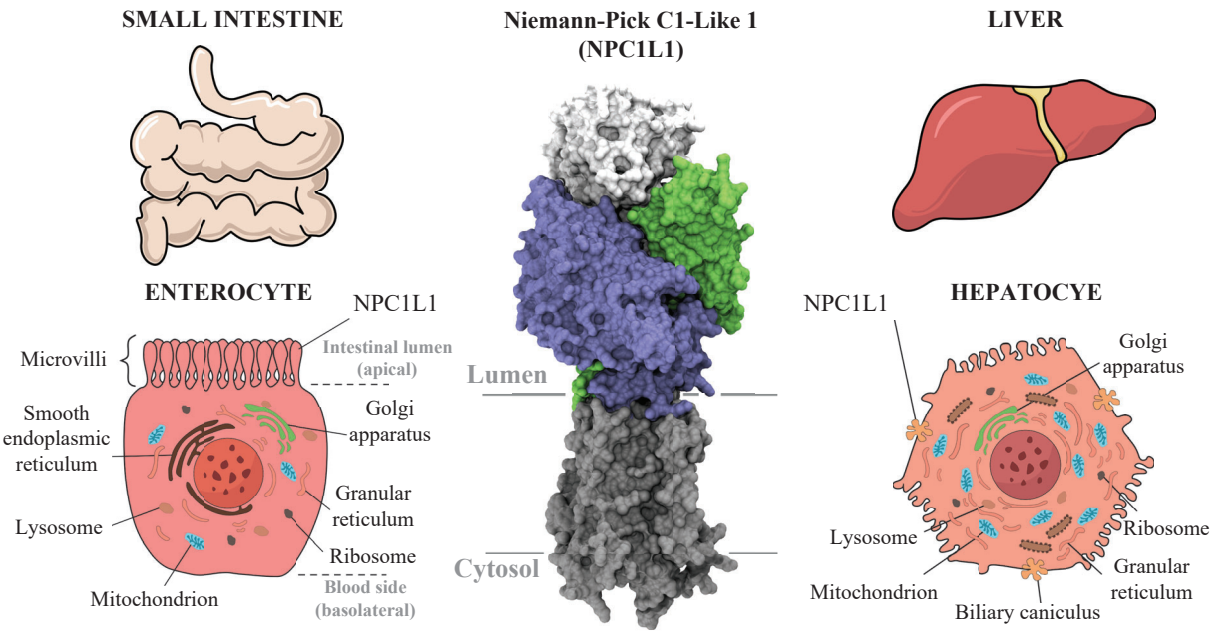


Figure 1.9 Representation of the organs and cell types where NPC1L1 is expressed, i.e. enterocytes in lumen of the small intestine and hepatocytes in the liver. The surface structure of NPC1L1 is shown in the centre.

1.3.1.1 NPC1L1 metabolic function and mechanism of action

Cholesterol absorption takes place in the small intestine, with 'absorption' covering all the steps from the ingestion and metabolism to transfer into the bloodstream. The overall process can then be divided into two main steps: first the uptake, which refers to the internalisation of cholesterol into the epithelial cells of the small intestine, and second the assimilation, in which cholesterol is transferred from the cytosol to the lymph through the basolateral membrane. For many years, the intestinal absorption of cholesterol was attributed to a passive process, but the second-order kinetics, the sterol specificity of the process and the *in vivo* effect of ezetimibe have shown that NPC1L1 mediates the cholesterol transport into enterocytes.[77] Genetic inactivation of this protein in mice led to a reduction in intestinal cholesterol absorption similar to that observed in mice treated with ezetimibe.[78] Although NPC1L1 is abundantly expressed in the small intestine of mammals, in the case of humans and primates it is expressed at significant levels in the liver, whereas in rodents NPC1L1 is expressed mainly in the small intestine.[79] In the liver, hepatic NPC1L1 facilitates the recapture of biliary secreted cholesterol by hepatocytes, leading to a recirculating enterohepatic process.[71] Liver hepatocytes are also responsible for removing excess of LDL-cholesterol from the circulation via LDL receptors (LDLR) and are also able to recycle HDL and chylomicron remnants via HDL receptor scavenger class B type I (SR-BI) and LDLR-related protein (LRP), respectively. However, they

are also responsible for atherosclerotic plaques when there is an ineffective clearance of plasma LDL.[80] The endocytic mechanisms behind cholesterol regulation have already been elucidated in hepatocytes, and it was thought that similar mechanisms might be at work in the small intestine.[81] NPC1L1 has been identified as the major importer of cholesterol and much progress has been made in understanding its mechanism over the last two decades.[82] However, the exact molecular mechanism of unesterified cholesterol uptake and transport into the interior of enterocytes remains unknown. Previous studies have elucidated several aspects of its complex mechanism of action, including the discovery that NPC1L1 internalises dietary cholesterol via vesicular endocytosis.[83,84]

Once cholesterol binds to the N-terminal domain (NTD) of NPC1L1, the carboxyl terminus on the cytoplasmic side is released from the membrane. The C-terminus contains a YVNxxF sequence motif, where 'x' can be any amino acid. This motif is recognised by the endocytic adaptor protein NUMB, which in turn recruits the clatrhin-AP2 complex for the endocytosis-mediated internalisation.[85] One of the unresolved steps is how the binding of cholesterol to the N-terminal domain induces the conformational change in NPC1L1 that releases the YVNxxF motif from the membrane to promote endocytosis. NPC1L1 has an internal tunnel connecting the luminal domains to the sterol-sensing domain (SSD), which is located in the transmembrane region and is also present in the structure of the Niemann-Pick C-related protein 1 (NCR1),[86] Niemann-Pick C1 (NPC1),[87] Hedgehog proteins Patched and Dispached,[88,89] sterol regulatory element–binding protein cleavage-activating protein (SCAP),[90] and HMG-CoA reductase (HMG-CoA).[91]

Sterol-like molecules have been resolved in the crystallographic structures of these proteins at different positions in the tunnel suggesting an internal structural transport pathway from the NTD to the membrane. However, the relationship between the transport process and the interaction of cholesterol with the SSD remains to be elucidated. It is known that NPC1L1 interacts with Flotillins and diffuses to cholesterol-rich domains in the membrane.[92] This coordination may provide an environment of high cholesterol concentration, sensed by the SSD, which promotes the recruitment of the clathrin-AP2 complex (Figure 1.10). In this scenario, the role of the NTD upon cholesterol binding is again questioned. Once the endocytosis is activated, NPC1L1 is internalised along with large amounts of cholesterol, which is transported to the endocytic recycling compartment (ERC), extracted by bile acids and diffused to the endoplasmic reticulum (ER). Lastly, cholesterol depletion in the ERC results in the recycling

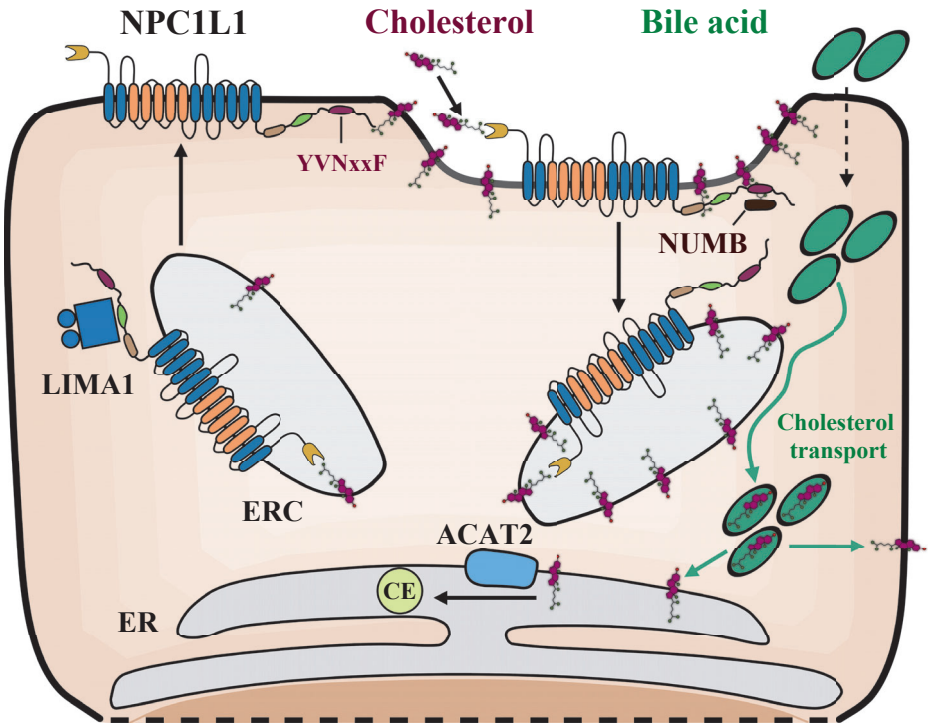


Figure 1.10 Representation of the NPC1L1-mediated enterocytic cholesterol transport.

or translocation of NPC1L1 back to the plasma membrane mediated by LIM domain and actin binding 1 (LIMA1) and Myosin-Vb.[83,93,94]

Further studies have shown that cholesterol is not the only molecule transported by NPC1L1. NPC1L1 has also been reported to be involved in the intestinal absorption of several fat-soluble nutrients, such as vitamin E and vitamin K_1, and other animal and plant sterols.[95] Interestingly, in addition to the fact that the translocation of NPC1L1 to the ERC is inhibited by ezetimibe, it has also been reported to be specific for cholesterol, and other sterols from animal sources, on phytosterols.[83] In addition, it has been reported that the absorption of phytosterols is dependent on NPC1L1, but the rate of absorption is lower than that of cholesterol and inversely proportional to the size of the sterol.[96]

1.3.1.2 NPC1L1 structure

The Niemann-Pick C1-Like 1 (NPC1L1) protein was first identified about two decades ago using a combination of genomics and bioinformatics approaches.[75,97] The discovery was driven by the search of proteins related to the Niemann-Pick type C1 disease (NPDC),[98] an autosomal recessive human lipidosis caused by defects in the Niemann-Pick type C1 (NPC1) protein, which are associated with an accumulation of unesterified cholesterol and other lipids in lysosomes. The complete NPC1L1 cDNA encodes a protein made of 1359 amino acids, but

it has been shown that the predominant mRNA transcript often skips exon 15 in the NPC1L1 gene, producing a protein made of 1332 amino acids, the most experimentally studied form of the protein.[97,99,100] NPC1L1 shares 42% of identity and 51% of sequence similarity with NPC1. Like its homologue, it has three luminal domains called the N-terminal domain (NTD), the middle luminal domain (MLD), and the C-terminal domain (CTD), and a transmembrane domain (TMD) composed of 13 transmembrane helices (TMs) (Figure 1.11). In addition, 5 of the 13 helices forming the TMD (TM3 to TM7) form the sterol-sensing domain (SSD).

To date, there is no complete crystallographic structure for human NPC1L1 and the most studied models are based on mice models. Two cryo-electron microscopy (cryo-EM) structures are available for the complete recombinant NPC1L1 protein in the PDB: PDBid 6V3H (resolution 3.50 Å) and PDBid 6V3F (resolution 3.70 Å).[101] Mouse knockout studies have established the role of NPC1L1 as a key protein involved in intestinal cholesterol absorption and the efficacy of ezetimibe as its main inhibitor.[75,96] In addition, there is an X-ray structure available in the PDB for the isolated human NTD of NPC1L1, PDBid 3QNT (resolution 2.83 Å),[102] which has been of high relevance for understanding some details behind the mechanism of cholesterol absorption, transport, as well as the inhibition and the sterol selectivity of NPC1L1. Other human NPC1L1 structures have been reported in the PDB, but none of them contains the NTD, suggesting that this domain has a high degree of mobility as noted upon the

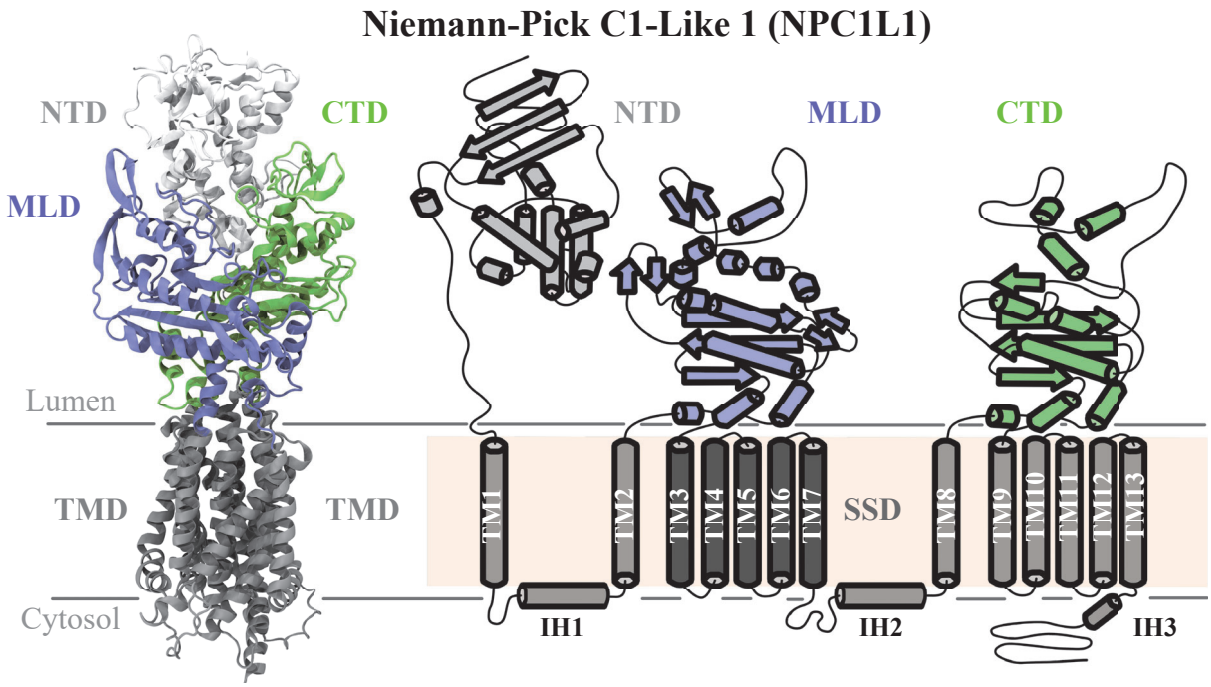


Figure 1.11 3D model of the full-length NPC1L1 (left) and schematic representation of its constituent domains (right).

inspection of PDBids 7N4U, 7N4V and 7N4X (resolution 3.34, 3.58 and 3.33 Å, respectively),[103] and 7DF8, 7DFW and 7DFZ (resolution 3.03, 2.69 and 3.58, respectively).[104] Another important aspect is the structural similarity of NPC1L1 to NPC1, for which there are 9 full-length human structures in the PDB: PDBids 3JD8 (resolution 4.43 Å, unpublished), 5U73 and 5U74 (resolution 3.35 and 3.33 Å, respectively),[105] 6UOX (resolution 4.02 Å),[106] and 6W5R, 6W5S, 6W5T, 6W5U, 6W5V (resolution 3.60, 3.00, 3.70, 3.90 and 4.00 Å, respectively).[87] In addition, there is a structure of the isolated NTD in complex with cholesterol, PDBid 3GKI (resolution 1.80 Å).[107] The high degree of similarity between the two proteins, which belong to the same NPC1-like family and are both involved in cholesterol homeostasis and regulation are among the reasons why NPC1 is considered the most appropriate structural and functional protein for comparison with NPC1L1.

Of the luminal domains, the NTD has been shown to directly bind free cholesterol in *in vitro* studies. It is also capable of binding cholesterol in the absence of other domains of the protein, and mutations near the binding cavity or the removal of the NTD drastically reduces the cholesterol uptake by NPC1L1, suggesting that its essential for the absorption mechanism.[108] The difficulty in solving the full-length structure with the NTD and the experimental evidence are behind the ascribed structural flexibility of this domain. Furthermore, the cryo-EM structure shows that the NTD fits into the space between MLD and CTD and that its binding cavity is perfectly articulated with the internal tunnel connecting the luminal domains to the SSD. This tunnel is also present in all the previously mentioned proteins that share the conserved SSD and from which similar sterol-like molecules have been solved at different positions.[101] All these findings support the theory that the tunnel is part of a sterol transport mechanism from the NTD to the membrane. To support this theory, the cryo-EM structure was solved with a slightly more potent ezetimibe derivative, EZE-phosphatidylserine (PS),[101] attached to the cavity formed between MLD and CTD, where the NTD fits, and at the point where the NTD-binding cavity and the internal tunnel of the protein meet. Given that ezetimibe has been shown to significantly inhibit NPC1L1-mediated cholesterol uptake, the proposed transport mechanism from NTD to SSD may be important as a signalling step to promote the already demonstrated endocytosis.[83]

1.3.1.3 NPC1L1 inhibition

There are several treatments for hypercholesterolemia, which is one of the risk factors for developing CVD. One of the most effective drugs for lowering plasma LDL cholesterol levels is ezetimibe, alone or in combination with statins. In brief, statins inhibit cholesterol synthesis in the liver, while ezetimibe inhibits cholesterol transport in the small intestine. Ezetimibe acts as a selective inhibitor of NPC1L1 by blocking the NPC1L1-mediated cholesterol absorption mechanism, although some of the mechanistic details behind this process remain unknown.[72–74] The trio of approved cholesterol-lowering drugs is completed by evolocumab (AMG 145), a monoclonal anti-proprotein convertase subtilisin/kexin type 9 (PCSK9) antibody that prevents PCSK9 from binding to the hepatic LDL receptor and reduces LDL receptor degradation, thereby increasing the clearance of serum circulating LDL-cholesterol.[109] The search for new potential LDL cholesterol-lowering agents is ongoing, as statin therapy is known to have undesirable side-effects in patients who require high doses,[110] as well as in patients prescribed ezetimibe who suffer from myalgia, high levels of creatine kinase in the blood, elevated transaminases and thrombocytopenia.[111] Over the past few years, other molecular candidates have been identified that show a lowering effect on plasma cholesterol levels and are related to NPC1L1 inhibition.

Polyphenols have been shown to attenuate hypercholesterolaemia through inhibition of NPC1L1 both *in vivo* and *in vitro*. In particular, seven polyphenols have showed specific inhibition of NPC1L1: luteolin, curcumin, cyanidin-3-glucoside, chlorogenic acid and catechin reduce NPC1L1 mRNA levels, while luteolin, quercetin and epigallocatechin gallate (EGCG) inhibit NPC1L1 transport by direct binding (Figure 1.12).[112–120] However, as with ezetimibe, their exact mechanism of inhibition remains unclear.

Some efforts have also been made to develop cholesterol-lowering drugs from natural products. Parthenolide (PTL in Figure 1.12), a sesquiterpene lactone extracted from the leaves of *Tanacetum parthenium*,[121] has recently been reported to have cholesterol-lowering activity associated with NPC1L1,[122] although the exact mode of action and its molecular relationship with the NPC1L1 mechanism remain to be elucidated. Another interesting example of natural products targeting NPC1L1 is fomiroid A (Figure 1.12), a sterol-like compound isolated from the mushroom *Fomitopsis nigra*,[123] which is structurally similar to lanosterol, a precursor of cholesterol biosynthesis that binds to the NTD in the same way as cholesterol.[102] Fomiroid A is as effective as ezetimibe in inhibiting the NPC1L1-mediated cholesterol absorption.

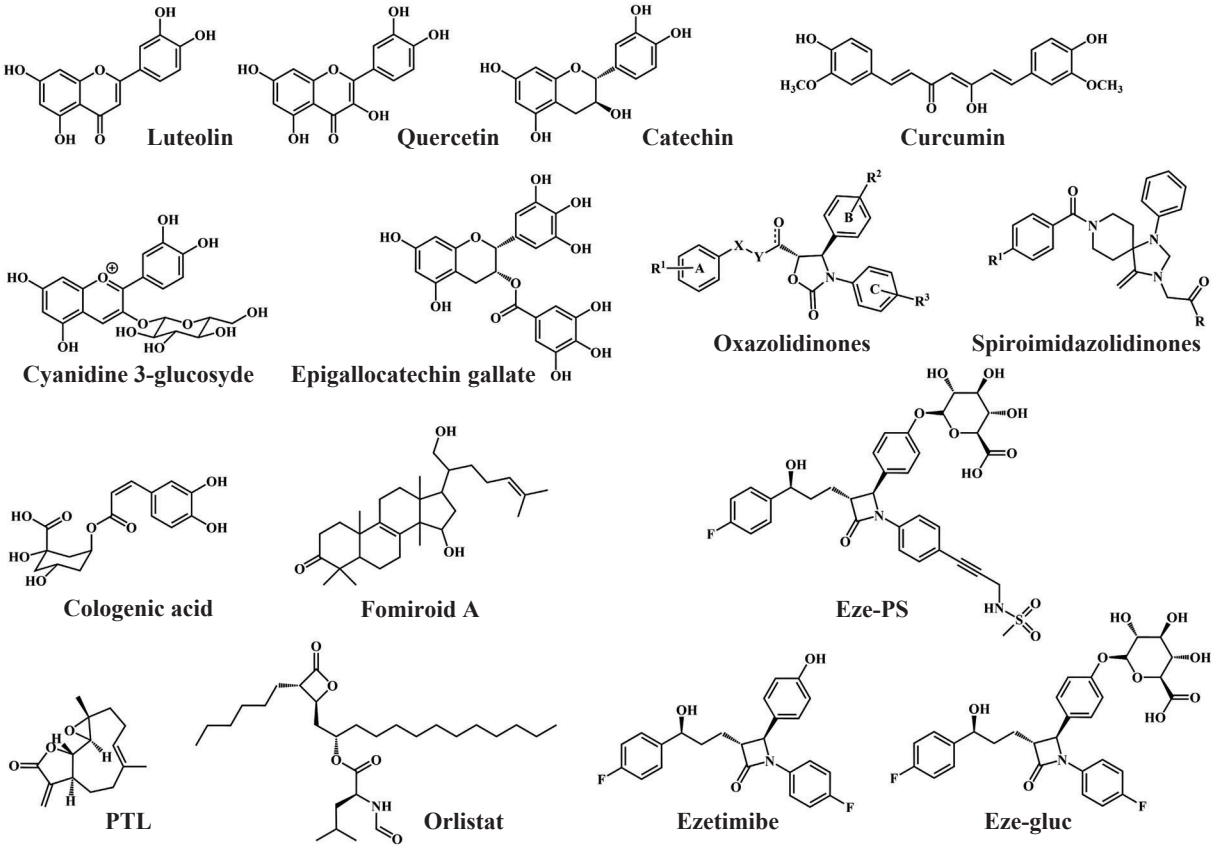


Figure 1.12 Chemical structure of some of the known inhibitors of NPC1L1-mediated cholesterol absorption discussed in section 1.3.1.1.

Based on the structure of ezetimibe, several derivatives and molecules with similar scaffolds have also been reported to act against NPC1L1-mediated cholesterol absorption. Ezetimibe itself has been reported to be rapidly glucuronidated (Eze-gluc in Figure 1.12) after administration, prior to reaching its target in the small intestine, and to be at least as potent as ezetimibe.[76] Another example already discussed is the ezetimibe derivative solved along the full-length cryo-EM structure of rat recombinant NPC1L1 by Huang et al,[101] which contains the glucuronide moiety as well as a sulphonamide moiety not present in the parent compound (Eze-PS in Figure 1.12). Other scaffolds with similar properties include β-lactone derivatives such as orlistat,[124] which has been reported to inhibit cholesterol uptake and acts as a non-competitive inhibitor of NPC1L1; spiroimidazolidinone derivatives discovered by virtual screening,[125,126] which inhibit cholesterol uptake by NPC1L1 and reduce plasma cholesterol levels by 67% in mice; and oxazolidinones,[127] which have NPC1L1 affinities in the micromolar range and acute *in vivo* efficacy in reducing plasma cholesterol levels by more than 50%.

1.3.2 Influenza A virus (IAV)

Back in ancient Greece, in 412 BC, Hippocrates described a flu-like syndrome called the "Cough/Fever of Perinthus" in the sixth volume of a book attributed to him called (translated as) Epidemics.[128] He described the illness as a winter and spring respiratory infection occurring every year at Perinthus, Greece (nowadays Turkey).

The first description of an influenza epidemic likely dates back to 1173 in both France and Italy.[129] The origin of the name "influenza" is often traced back to 15th century in Italy, where an epidemic fitting the influenza description was attributed to the "Influenza of the Stars" that would have spread across Europe, Asia and Africa.[130] Other epidemics described with similar symptoms may have occurred in 1414, 1557 and 1675-1676, and confirmed and documented epidemics, probably related to the influenza virus, occurred in France in 1788-1790 and 1830-1832.[131] In modern times, the most devastating and documented influenza pandemic occurred in 1918-1919 and has been known since then as the "Spanish Flu", suspected to have been originated in China. Recent estimates suggest that 17.4 million people died, or approximately 1% of the world's population at the time.[132] Modern phylogenetic analyses have attempted to explain the virulence of the strain from conserved samples of the virus from this pandemic with unclear results.[133,134] RNA sequencing of samples taken from birds preserved in alcohol since 1917 concluded that the strain was not entirely derived from avian influenza.[135]

The high degree of variability that displays the influenza virus (IV) allows it to escape immune response leading to annual epidemic events known as “seasonal influenza”. Eventually, a new highly pathogenic variant arises causing pandemic events with a high incidence in mortality. IV can be classified into types A, B, C, and D, but types A and B are responsible for most mortal cases. Type A (IAV) is the only one displaying pandemic potential for humans due to its animal reservoir.[136] In addition, IV subtypes are distinguished based on the antigenic properties of two surface glycoproteins: Hemagglutinin (HA) and Neuraminidase (NA), which promote and coordinate host cell entry and exit, respectively. Today, influenza is classified as a respiratory disease that contributes significantly to the global burden of disease. The WHO estimates that there are between 3 to 5 million severe influenza cases and between 290,000 to 650,000 influenza-related deaths every year.[137] There have been four highly relevant pandemic episodes involving new highly pathogenic IAV variants for which humans have no

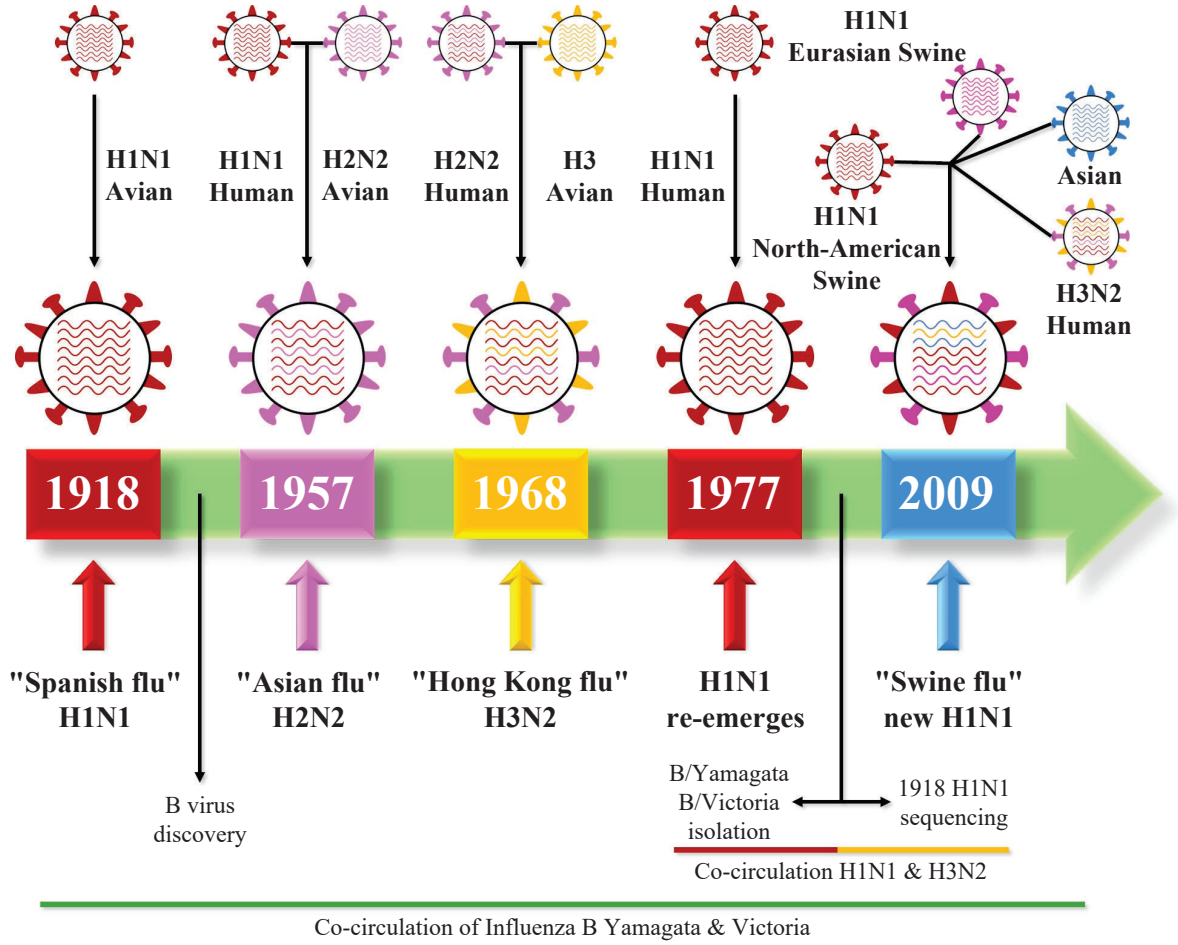


Figure 1.13 Summary of pandemic outbreaks of the last century from 1918 to 2009 great pandemic events. In 1977, the human H1N1 virus re-emerged and enough immunity in the population prevented an uncontrolled pandemic event. The 2009 pandemic was a recombinant virus including RNA from swine, avian and human viruses. All pandemic viruses have included gene segments from the 1918 virus represented in red.

natural immunity and shown in Figure 1.13: the aforementioned H1N1 “Spanish flu” (1918), the H2N2 “Asian flu” (1957), the H3N2 “Hong Kong” flu (1968) and the H1N1 “Swine” flu (2009).[138] Since the 2009 pandemic, WHO has developed global epidemiological standards for influenza spread control to standardise public health measures for prevention.[139] This threatening scenario calls for continued efforts to promote influenza prevention, to understand the molecular mechanisms involved in viral infection, and to develop new effective treatments for influenza especially in view of the emergence of strains resistant to current treatments.[140]

1.3.2.1 IAV structure and genotypic classification

Influenza A virus (*Alphainfluenza virus*), the primary etiological agent of influenza, is a member of the Orthomyxoviridae family. This family includes seven genera of segmented, negative-sense, single-stranded RNA viruses: Influenza A, Influenza B (IBV), Influenza C (ICV), Influenza D (IDV), Isavirus, Thogotovirus and Quaranjavirus.[141]

Structurally, both IAV and IBV virions are similar and pleomorphic, whereas ICV and IDV are more divergent. IAV virions tend to form spherical (approximately 100 nm in diameter) or filamentous (approximately 100 nm wide and up to 20 µm long) morphologies, both of which have been observed experimentally.[142,143] Three major subviral components make up an IAV virion as shown in Figure 1.14: the viral envelope, the viral core and the ribonucleoproteins (vRNPs). The viral lipidic envelope, derived from the host plasma membrane, contains three transmembrane proteins: the aforementioned surface glycoproteins, HA and NA, and matrix protein 2 (M2), commonly known as the M2 proton channel. The viral core is a monolayer of matrix protein 1 (M1) located beneath the viral envelope that provides structural support to the virion, interacts with cytoplasmic regions near HA and NA on the apical surface of the host cell, and encloses the approximately 13,500 nucleotide viral genome.

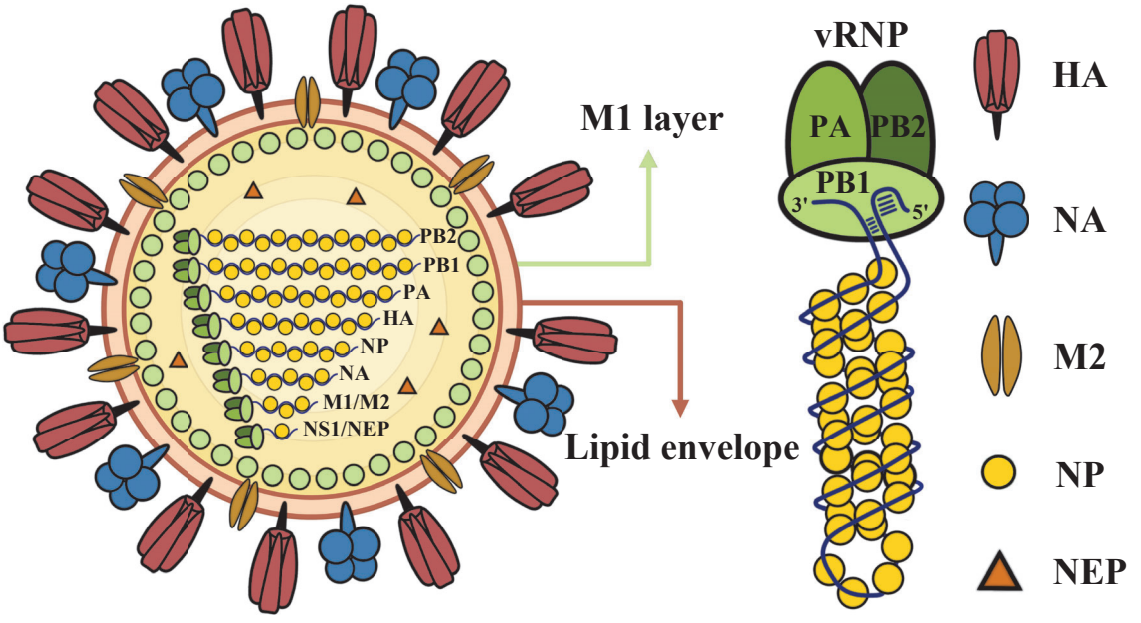


Figure 1.14 Representation of Influenza A virus and its major components. In the lipid envelope: the two viral glycoproteins HA and NA and the ion channel M2. The structural core of the virion consists of M1 proteins that interact with surface proteins and vRNPs. The eight gene segments form the vRNPs, which are composed of NP and the polymerase complex (PA, PB1, PB2). The nuclear export protein (NEP) is also present in the virion.

The viral genome is divided into eight individual gene segments called viral ribonucleoproteins (vRNPs).[144] Each of these vRNPs is in turn a complex formed by viral RNA (vRNA) wrapped around a double helical scaffold of nucleoprotein (NP), the heterotrimeric polymerase complex at the 3′ and 5′ ends (polymerase acidic protein, PA; polymerase basic 1, PB1; and polymerase basic 2, PB2), and a loop at the other end. In addition, small amounts of the nuclear export protein NEP (formerly non-structural protein 2, NS2)[145] are found in the closed environment of the viral core.[146] Each of the eight vRNPs encodes at least one protein. Of these, 10 have been identified as essential for viral replication and will be briefly discussed next: HA, NA, M1, M2, NP, PA, PB1, PB2, non-structural protein 1 (NS1) and NEP. However, the IAV genome is known to encode a further 11 accessory proteins: PB2-S1; PB1-N40; PB1-F2; PA-X; PA-N155; PA-N182; eNP; NA43; M42; NS3; and tNS1.[147]

HA is a homotrimeric glycosylated type I integral membrane protein with the ability to bind to sialic acids (N-Acetylneuraminic acid) found in host cell glycoproteins. The orientation of the glycosidic linkage between the sialic acid and the sugar is critical for the recognition mechanism and can vary depending on the nature of the IAV strain. For example, HA from IAVs from human samples have a higher specificity for the α2-6 or "bent" orientation, whereas IAVs adapted to avian species have HA proteins specific for the α2-3 or "flat" orientation.[148,149] Once HA is bound to the host cell, the virion is internalised via receptor-mediated endocytosis. The low pH inside the endosome promotes a large-scale and irreversible conformational rearrangement of the HA trimeric structure involving two crucial steps: i) the cleavage of each HA inactive monomer (HA_0) by host cell proteases into two subunits HA_1 (containing the receptor binding site for sialic acids) and HA_2, and ii) the subsequent release of the fusion peptide (FP) from the HA_2 subunit. Typically, a host-derived serine protease cleaves the FP by recognising the Q/E-x-R motif at the interface of HA_1 and HA_2.[150] The release of the FP ultimately mediates fusion of the viral envelope with the endosomal membrane releasing the vRNPs into the cytosol.[151]

NA is a tetrameric type II integral membrane protein that acts as a sialidase, hydrolysing the glycosidic linkage between sialic acids and the terminal saccharides in the host cell. While the recognition function of HA initiates the infection process, NA inhibits the release of newly formed virions bound to sialic acids of the host cell. NA selectively cleaves unwanted interactions with sialic acid both on the host cell membrane and between nascent virions, preventing aggregation.[152]

The matrix proteins M1 and M2 are both produced by the 7th gene segment. As mentioned above, M1 is located beneath the viral envelope as a matrix monolayer providing structural support, interacting with both vRNPs and surface glycoproteins, and promoting assembly and budding of new virions. On the other hand, M2 is a small tetrameric type III integral membrane protein that acts as a proton channel when exposed to the low pH of the endosome. The influx of protons through the M2 channels into the virion is crucial for the early steps of the infection cycle, allowing dissociation between M1 and vRNPs during viral uncoating.[153]

NP serves as the structural backbone of vRNPs encircling all RNA gene segments. It is involved in the intracellular trafficking of vRNPs and in the transcription process, stabilising template-polymerase interactions. NP interactions with host cell proteins are crucial for proper subcellular localisation of vRNA and the most important are the importin α nuclear localisation signalling receptors, filamentous actin, the nuclear export receptor CRMI and a DEAD-box helicase.[154] The NPs with the RNA wrapped around them are packaged together with the largest gene products of the virus, polymerase proteins, to form the vRNPs: PB1, PB2 and PA. PB1 is the RNA-dependent RNA polymerase for the virus and PB2 is needed for the mRNA synthesis.[155] The exact biological role of PA is not fully understood, but it has recently been suggested to modulate polymerase activity.[156]

The final essential components for viral replication are the non-structural proteins NS1 and NEP, which are alternative splice products of the 8th gene segment. The roles of NS1 are multiple and varied due to the interactions of its multiple domains with various cellular proteins and RNA, but mainly it suppresses cellular gene expression by blocking nuclear mRNA processing.[157] Lastly, NEP facilitates the nuclear export and transport of vRNP complexes to the assembly site at the apical cell membrane.

Without considering any of the accessory proteins encoded by the IAV genome and is high genotypic variability, a proper nomenclature must be used to classify the different strains by its phylogenetic origin. Correct phylogenetic classification is essential to track the emergence and spread of new variants, and to know the circulating strains in humans and other species, as well as the zoonotic transmission potential.[158] In the case of IAV, the different strains are classified depending on sequence and the antigenic properties of HA and NA (Figure 1.15). IAV have been classified into two phylogenetic groups, group 1 and group 2, on the basis of 18 HA subtypes and 9 NA subtypes, giving the known nomenclature HXNY (being X and Y the numerated HA and NA subtypes, respectively). However, very recently it was reported a 19th

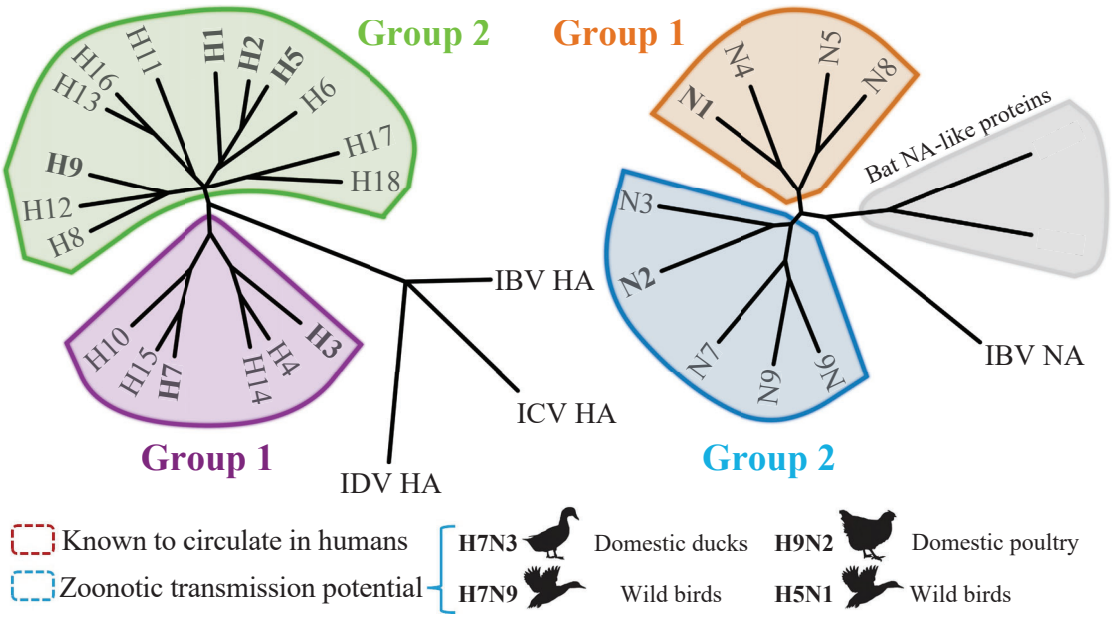


Figure 1.15 Phylogenetic classification of HA and NA.

variant of HA with a large genetic distance from the previously described subtypes.[159] Thus, the updated classification stablishes that group 1 contains the subtypes H1, H2, H5, H6, H8, H9, H11, H12, H13, H16, H17, H18 and H19, and group 2 encompasses the subtypes H3, H4, H7, H10, H14, and H15.

1.3.2.2 IAV viral cycle

The IAV replication cycle involves several steps and begins with the attachment of an IAV virion to the sialylated surface receptors of the host cell (step 1, Figure 1.16). Following attachment, the virion is internalised via receptor-mediated endocytosis and moves through the endosomal pathway in a clathrin-dependent manner (step 2, Figure 1.16). Upon acidification to a pH close to 5.0 in the late endosome, two structural changes occur in the virion: 1) the M2 ion channel facilitates the entry of protons into the viral particle, disrupting the M1-vRNP interactions, and 2) a conformational stepwise rearrangement in HA exposes the fusion peptide FP, which inserts into the endosomal membrane and ultimately promotes fusion with the viral envelope (step 3, Figure 1.16). Following fusion, vRNPs are released into the host cytosol and imported into the nucleus via the NP signalling and nuclear import pathway.

In the nucleus, the virus synthesises three types of RNA from vRNA: positive-polarity messenger RNA ((+)mRNA), negative-polarity viral RNA ((-)vRNA), which is part of new virions, and positive-polarity complementary RNA ((+)cRNA), which serves as a full-length

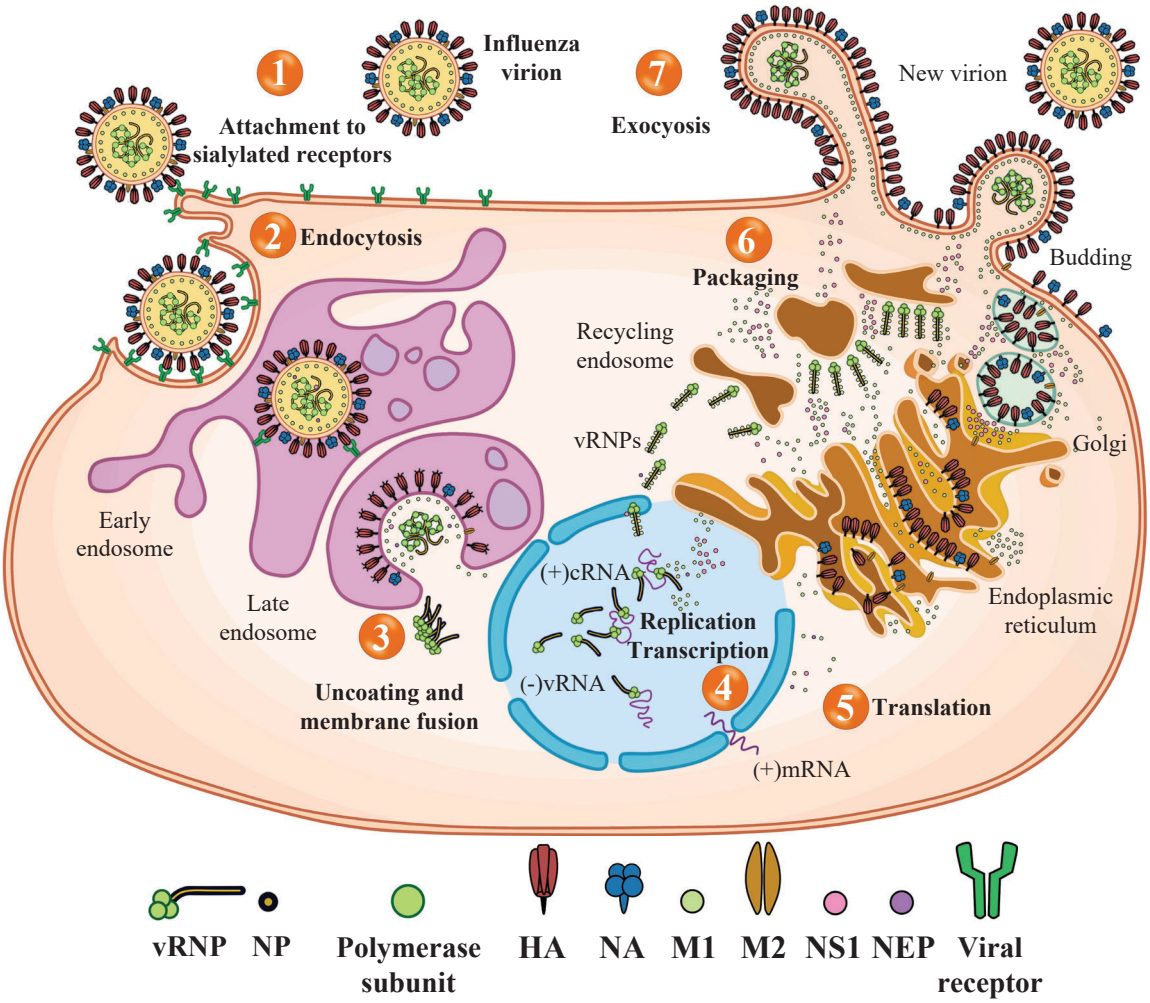


Figure 1.16 Schematic representation of the IAV replication cycle.

template for the production of new vRNAs. The synthesis of (-)vRNA and (+)cRNA is promoted by NP, which also binds to the elongating RNA strands and inhibits the formation of (+)mRNA by stabilizing the polymerase-template interaction. The new (-)vRNA is then used as template for the production of the positive polarity RNAs (step 4, Figure 1.13), (+)cRNA (replication) and (+)mRNA (transcription). Transcription of (+)mRNA from the (-)vRNA template uses host-derived 5′-methyl caps to efficiently bind host cellular proteins involved in RNA processing, nuclear export and protein production, while the 3′-terminus of each segment is polyadenylated. This process is known as “cap-snatching”, where PB2 binds to host cell mRNA and PB1 cleaves that RNA to serve as prime to new (-)vRNA. While still in the nucleus, the new viral (+)mRNA is spliced to produce the coding regions for M2 channels and NEP, and then exported to the cytosol for viral protein production.

When (+)mRNA is exported to the cytosol, viral proteins are translated by the host protein production machinery (step 5, Figure 1.16). Among the major essential proteins for the virus,

PB1, PB2, PA, NP, NS1, NEP and M1 are produced in the cytosol by host ribosomes and transported back to the nucleus, where they will be involved in further transcription, gene replication and formation of vRNPs that will be packed into nascent virions (step 6, Figure 1.16). In parallel, HA, NA and M2 are produced by ribosomes bound to the membrane of the endoplasmic reticulum (ER), inserted into the ER and glycosylated before being transported to the Golgi apparatus. The viral surface proteins are then transported to the cell membrane via the secretory pathway and assembled with the new vRNP complexes coming from the nucleus. The formation of vRNP takes place in the nucleus and involves the binding of the viral polymerase complex (PB1, PB2 and PA), NP and NEP to the new (-)vRNA. M1 then interacts with vRNPs and NEPs to export the M1-vRNP complexes to the cytosol (M1 prevents the import of vRNPs back into the nucleus). M1 directs the localisation pathway to the host membrane, where the M1-vRNP complexes will interact with the cytoplasmic regions of new HA and NA (coming from the Golgi apparatus) leading to the formation of viral buds with no more than 8 gene segments in each nascent particle (step 6, Figure 1.16). Finally, the exocytosis of a new virion occurs (step 7, Figure 1.16).[160] Host membrane proteins are largely excluded from the nascent viral envelopes due to NA's ability to cleave sialic acids, and the exact mechanism behind the specific packaging of the new viral material is thought to be the result of RNA signalling from each of the gene segments, ensuring that most newly formed virions are infectious.[161,162]

1.3.2.3 IAV-HA inhibition overview

To address the burden of IAV, several strategies focusing on reducing the severity of infection, transmission and limiting viral replication rates have been adopted through years of research. Two main strategies are used to combat IAV infection. Preventive measures such as annual vaccination are essential to reduce the spread of the virus in the population and to increase immunity. On the other hand, antiviral drug treatments are administered with the aim of interfering directly with the replication cycle of the virus, even disrupting its ability to make copies of itself and spread.

From the inception of anti-influenza strategies to the present day, the primary shield against influenza infection has been neutralising and receptor blocking antibodies to HA or, at a secondary level, to NA. Although immunity to multiple proteins can attenuate disease, only antibodies specific to HA can prevent infection. Inhibition of NA blocks the enzymatic activity required for particle release in the final stages of the replication cycle, reducing viral spread

only after initial infection.[163] Influenza vaccines have been used for more than 70 years,[164] of which two types are licensed for use in humans depending on their contents: Trivalent inactivated virus (TIV) and live-attenuated influenza virus (LAIV). TIV contains components of the three major strains circulating in the human population: H1N1 IAV, H3N2 IAV and the Victoria lineage of IBV. For IBV, only two strains are known, the Victoria and Yamagata lineages. However, it has been reported that the Yamagata lineage may have become extinct during the COVID-19 pandemic, as it has not been detected in global influenza surveillance since March 2020.[165] Therefore, its components are not administered in anti-influenza immunisations. LAIV, on the other hand, is produced from cold-adapted influenza viruses. Both types use polyvalent formulations, i.e. containing mixtures of the same antigen from different strains. However, the increasing limitations in developing new vaccines and the emergence of new resistant strains has stimulated the development of new antiviral strategies to complement immunisation, such as the emergence of small molecule antivirals.[166]

The first small molecules to enter the anti-influenza drug catalogue were the adamantane-related amantadine and rimantadine.[167] These two molecules can inhibit viral replication by blocking the M2 proton channels, which are crucial for disrupting the M1-vRNP complex interaction. Today, both inhibitors are no longer recommended for human therapy due to their low activity against current circulating adamantane-resistant strains and undesirable side effects in both the gastrointestinal tract and the central nervous system.

Since then, new inhibitors were designed to target the M2 proton channel, but other drug targets have become more relevant (NA, RNA polymerase and HA) as new resistant strains emerged.[168,169] The most effective NA inhibitor is Laninamivir,[170] a prodrug that has a high affinity for NA and acts as a competitive inhibitor. One of the most important features of Laninamivir is its long-lasting action, as it was designed to be administered by inhalation to facilitate its delivery directly to the respiratory tract. However, in recent years, NA has received less attention as an anti-influenza target due to the rapid adaptation of NA by resistant mutations.

RNA polymerase inhibitors are designed to target the RNA-dependent RNA polymerase due to its high degree of conservation among all four IVs (A, B, C and D). Several inhibitors have been developed targeting one of the three subunits of the complex.[171] For example, faripiravir is able to inhibit RNA elongation by incorporating into nascent viral RNA strands produced by PB1; pimodivir targets the PB2 domain by occupying the cap binding domain,

inhibiting RNA binding; and baloxavir marboxil binds to the PA domain, inhibiting RNA cleavage.

Due to its structural flexibility and phylogenetic variation, HA is a challenging drug target to modulate with small molecules. A feasible inhibition strategy would be to stabilise the pre-fusion conformation of HA, preventing the release of the FP and blocking the fusion of the viral and endosomal membranes. Since the pH-induced structural rearrangement in the late endosome involves the entire trimeric structure, fusion inhibitors target different binding pockets in the HA structure depending on their physico-chemical properties. The X-ray structures available in the PDB show two major binding sites for small molecules inhibiting HA, both in the HA1-HA2 stem region, shown as Site A and Site B in Figure 1.17.

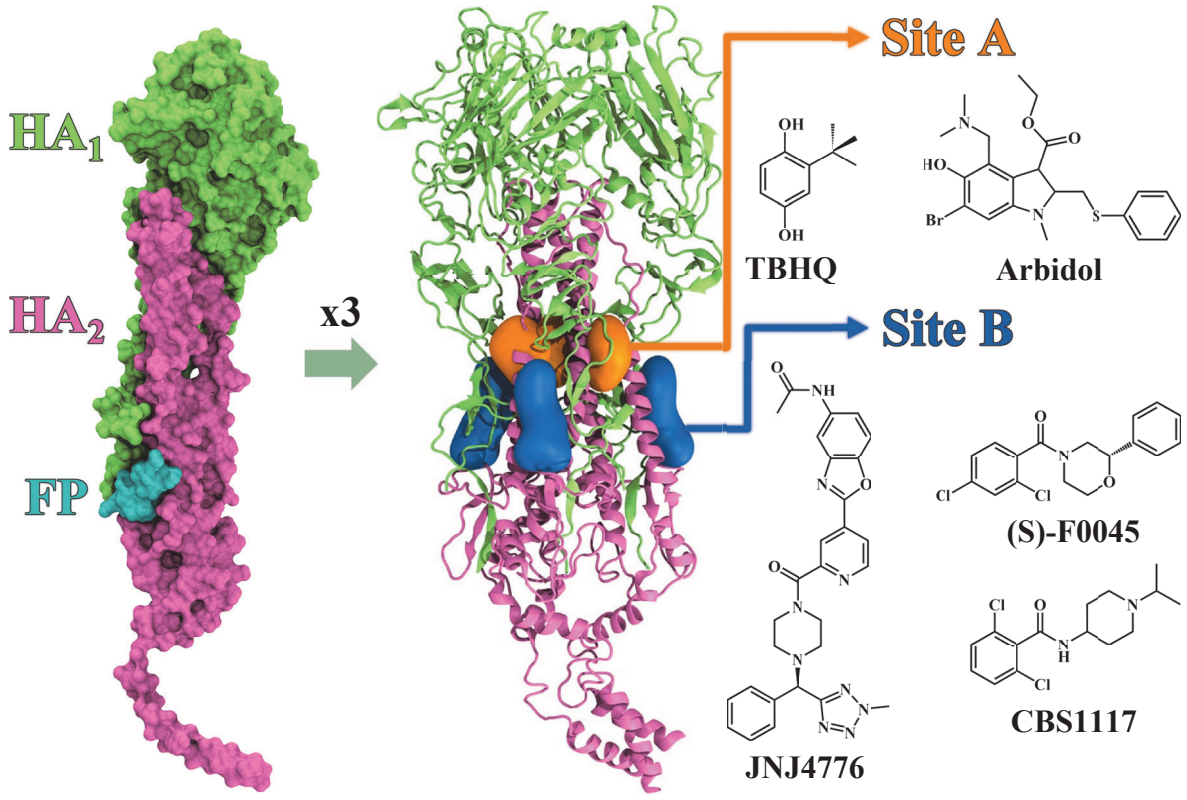


Figure 1.17 Representation of the monomeric (surface) and trimeric (cartoon) structures of HA. The two binding sites experimentally observed by X-ray crystallography are shown as the orange surface for Site A and as the blue surface for Site B, together with the molecular structures of the main co-crystallised inhibitors mentioned in this section.

Site A is reported to be the binding site for tert-butylhydroquinone (TBHQ; PDBids: 3EYK and 3EYM, resolution 2.50 and 2.80 Å, respectively)[172] and Arbidol (PDBids: 5T6N and 5T6S, resolution 2.54 and 2.36 Å, respectively).[173] Although TBHQ is reported to be a weak antiviral compound, its mode of action inhibits the conformational rearrangement that HA performs to promote membrane fusion by stabilising the neutral pH structure prior to acidification. Its binding position at 16 Å from the FP allows the formation of a TBHQ-HA complex with a stoichiometric ratio of 3:1. On the other hand, Arbidol is considered the most successful small-molecule inhibitor of IAV and is approved for human use in Russia and China, although its exact mechanism of action is unclear. Arbidol is able to target both phylogenetic groups of HA, with a higher affinity for group 2 HAs, as well as other viruses such as hepatitis C virus (HCV).[174] Experimental assays have determined that Arbidol also stabilizes the pre-fusion state of HA and inhibits the conformational rearrangement that releases the FP. However, the need for higher dosages to exhibit a substantial inhibition rate stimulated the search for novel spirothiazolidinone derivatives based on its structure, exhibiting remarkable activities against H3 HA subtype.[175]

Site B is known to be the binding site of several HA neutralising antibodies, which are able to recognise its highly conserved sequence between different HA subtypes and block the fusion mechanism.[176] For small molecules, Site B has been reported to accommodate the fusion inhibitors JNJ4796 (PDBids: 6CFG and 6CF7, resolution 2.32 and 2.72 Å, respectively),[177] (*S*)-F0045 (PDBid: 6WCR, resolution 2.68 Å),[178] and CBS1117 (PDBid: 6VMZ, resolution 2.20 Å).[179] X-ray structures and further SAR studies of several derivatives of JNJ4796 have yielded compounds with potent *in vivo* antiviral activity against H1 HA subtype (group 1) in the sub-micromolar range.[180,181] Similar trends can be found in the reported activities in the micromolar range for (*S*)-F0045, CBS1117 and further analogues, which display a high degree of specificity for group 1 HA.[182]

Chapter 2

Objectives

"I know a standing ash, Yggdrasill its name is,
a tall tree doused with white bright mud.
From there come dews, those that fall in dales.
Does it ever grow, green, above Urðr's spring.

From there come the maidens, mighty in wisdom,
three from the dwelling near the lake beneath the tree.
Urðr is one named, Verðandi the next,
- on the wood they scored - and Skuld the third.
Laws they made there, and life allotted
to the sons of men, and set their fates"

Vǫlva (Vǫluspá, stanzas 19, 20)

Despite the development of novel therapeutic strategies against disease and the evolution of technology, humanity still suffers from treatment failure, drug resistance and other causes of widespread disease. Understanding the molecular mechanisms behind disease is the first step in addressing these challenges. In this thesis, I have used state-of-the-art computational techniques to investigate two druggable targets that are directly involved in major public health problems identified by the WHO. First, I have investigated the unknown mechanism and selectivity of the major mammalian free cholesterol transporter, Niemann-Pick C1-like 1 (NPC1L1), which is implicated in the increasing burden of cardiovascular disease (CVD). Second, I have explored novel inhibition strategies against the influenza A virus (IAV), specifically targeting one of its antigens, haemagglutinin (HA). In this context, the main objectives of this thesis are outlined below.

Project 1: The NPC1L1 project focuses on investigating the molecular mechanism and the substrate selectivity of the N-terminal domain (NTD) of the NPC1L1 transporter. The findings will pave the way for the understanding of certain cholesterol-related physiological disorders. In particular, the aims of this project are:

- To identify key amino acid residues and structural motifs responsible for selective sterol binding.
- To explain the effect of single point mutations that impair cholesterol transport by disrupting protein dynamics of the NTD.

Project 2 and Project 3: The IAV HA inhibition project aims to explore novel inhibition strategies targeting Hemagglutinin (HA) contributing to the development of small molecules as antivirals. The focus is on predicting a putative binding site and evaluating the structural stability of HA-mediated fusion inhibitors by correlating computational predictions with the available structure-activity relationship (SAR) data of two potential inhibitors:

- The pinanamine-based inhibitor M090 discovered by Zhao et al.[183] with sub-micromolar activities against group 1 and group 2 HA subtypes.
- To provide SAR of a series of benzamide derivatives synthesised by the group of Prof. Santiago Vázquez (Departament de Farmacologia, Toxicologia i Química Terapèutica, UB) and tested against IAV-infected MDKC cells by the group of Prof. Lieve Naesens (Department of Microbiology, Immunology and Transplantation, KUL).

Chapter 3

Results

"A hall I see, far from the sun,
on Nástrǫnd it stands, the doors face north.
Venom drops, through the smoke-vent fall,
for around the walls, winding snakes.

There I see wading, through swift currents
perjured people and murder wolves,
and workers of ill, with the wives of men.
There Níðhǫggr sucked, the blood of the deceased,
and the wolf tore men. Would you know yet more?"

Vǫlva (Vǫluspá, stanzas 37, 38)

F. Javier Luque Garriga
Departament de Nutrició, Ciències de l'Alimentació i Gastronomia.

UNIVERSITAT DE BARCELONA

Campus de l'Alimentació de Torribera
Facultat de Farmàcia

Avd. Prat de la Riba 171
08921 Santa Coloma de Gramenet

Tel: +34 93 403 37 88
e-mail: fjluque@ub.edu
http://www.ub.edu/cbdd

Santa Coloma de Gramenet, 25 de Març de 2025

Informe dels directors de tesi

Com a supervisors de la Tesi Doctoral, volem remarcar que el Sr. Aitor Valdivia Escribá ha fet un contribució plenament original i significativa, avalada per les següents evidències:

- El seu treball ha dut a quatre publicacions, tres en forma d'article de recerca i un altre en forma d'article de revisió. Dels articles de recerca publicats, la seva contribució ha estat fonamental, portant a terme el pes del seu desenvolupament, motiu pel qual figura com a primer autor o co-autor.

  Aitor Valdivia, F. Javier Luque, and Salomé Llabrés. Binding of Cholesterol to the N-Terminal Domain of the NPC1L1 Transporter: Analysis of the Epimerization-Related Binding Selectivity and Loop Mutations. ***Journal of Chemical Information and Modeling***, 2024, 64, 180-204
  (Revista Q1 en Enginyeria Química (miscel·lània), Química (miscel·lània), Aplicacions en Ciències Computacionals i Bibliotèques i Ciències de la Informació).

  Aitor Valdivia, Maria Rocha and F. Javier Luque. Mining Druggable Sites in Influenza A Hemagglutinin: Binding of the Pinanamine Based Inhibitor M090. ***ACS Medicinal Chemistry Letters***, 2025, 16, 1, 126-135
  (Revista Q1 en Descobriment de Fàrmacs i Química Orgànica, Rewvista Q2 en Química Mèdica; aquest treball va ser publicat en el número especial *"Structural Biology in Drug Discovery and Development"*, i seleccionat per il·lustrar la coberta del número on va ser publicat).

Els factors d'impacte i quartils (actualitzats a data de juny de 2024) s'han consultat a través del cercador Journal Searches, on es mostren les dades recollides a Journal Citation Reports (JCR) (Clarivate).

- A més, és el co-autor d'un manuscrit més presentat en aquesta tesi, fet en col·laboració amb el grup de recerca del Prof. S. Vázquez de la Facultat de Farmàcia i Ciències de l'Alimentació, i que es troba en la fase final de redacció. En aquest treball multidisciplinari, que combina modelització, síntesi química i avaluació farmacològica, Aitor Valdivia ha dut a terme tots els estudis de modelització i simulació molecular.

  N-[(Thiophen-3-yl)methylbenzamides as Influenza Virus Fusion Inhibitors Acting on H1 and H5 Hemagglutinins
  Juan Martín-López, Silke Rimaux, **Aitor Valdivia**, Valeria Francesconi, Celia Escriche, Cato Mestdagh, Ria Van Berwaer, Kaat Verleye, Samuel Noppen, Manon Laporte, Óscar Lozano, Annelies Stevaert, F. Javier Luque, Lieve Naesens, Santiago Vázquez. ***In preparation***.

- També va contribuir a d'un article de revisió en el que n'és el segon autor, que no està inclòs dins de la seva tesi:

  Francisco Javier Hermoso-Pinilla, **Aitor Valdivia**, Maria-José Camarasa, Tiziana Ginex, and Francisco Javier Luque. Influenza A virus hemagglutinin: from classical fusion inhibitors to proteolysis targeting chimera-based strategies in antiviral drug Discovery. ***Exploration of Drug Science***, 2024, 2, 85-116.

- Cal dir que el Sr. Valdivia ha supervisat el treball de dos estudiants estrangeres que han fet estades al grup:

  i) Maria Rocha (Univ. Coimbra), que va participar en el treball publicat a *ACS Medicinal Chemistry Letters*, i

  ii) Ilenia Lupinu (Univ. Sassari), qui està duent a terme estudis sobre el mecanisme d'inhibició de compostos antivirals, del qual esperem que en sortirà una publicació addicional en 2025.

- Finalment, cal dir que la formació acadèmica del Sr. Valdivia ha estat complementada amb la participació en un projecte de col·laboració amb l'empresa Pharmacelera, a través de la Fundació Bosch Gimpera, realitzant estudis per optimitzar la interacció de lligands amb les dianes d'interès, amb una durada de dos anys. Tot i que aquest treball no està recollit en cap publicació, aquest projecte li ha permès guanyar formació en l'abordatge de la recerca aplicada des d'una perspectiva industrial.

Atentament:

F. Javier Luque

FRANCISCO JAVIER LUQUE GARRIGA - DNI 35047889Y

Digitally signed by FRANCISCO JAVIER LUQUE GARRIGA - DNI 35047889Y
Date: 2025.04.15 11:06:08 +02'00'

Salomé Llabrés

The results of the present thesis are presented in this chapter as a compendium of publications and divided in three separated Projects that involve two different topics.

Project 1:

Valdivia, A., Luque, F. J., & Llabrés, S. (2024). Binding of Cholesterol to the N-Terminal Domain of the NPC1L1 Transporter: Analysis of the Epimerization-Related Binding Selectivity and Loop Mutations. *Journal of Chemical Information and Modeling*, *64*(1), 189–204.

In **Project 1**, I used a variety of computational chemistry techniques, including Molecular Dynamics (MD), Relative Binding Free Energy (RBFE), Well-Tempered Metadynamics (WT-MetaD) simulations, among others, to determine the molecular features behind the sterol affinity and selectivity of the major mammalian free cholesterol transporter, the Niemann-Pick C1-Like 1 (NPC1L1) transporter.

Project 2 and Project 3:

Valdivia, A., Rocha, M., & Luque, F. J. (2025). Mining Druggable Sites in Influenza A Hemagglutinin: Binding of the Pinanamine-Based Inhibitor M090. *ACS Medicinal Chemistry Letters*, *16*(1), 126–135.

Draft in preparation: Martín-López, J., Rimaux, S., Valdivia, A., Francesconi, V., Escriche, C., Mestdagh, C., Van Berwaer, R., Verleye, K., Noppen, S., Laporte, M., Lozano, O., Stevaert, A., Luque, F. J., Naesens, L., Vázquez, S. (2025). N-[(Thiophen-3-yl)methylbenzamides as Influenza Virus Fusion Inhibitors Acting on H1 and H5 Hemagglutinins.

In **Project 2** and **Project 3**, I used Molecular Dynamics (MD) and Relative Binding Free Energy (RBFE) simulations to explore and suggest a putative binding site for two series of compounds with HA-mediated fusion inhibition potential against the Influenza A virus. In Project 2, I explored pinanamine-based inhibitors exhibiting potent activities against both HA phylogenetic groups. In Project 3, (draft ready for submission and included in this thesis), I conducted similar simulations for a series of benzamide derivatives synthesised by the group of Prof. Santiago Vázquez (Departament de Farmacologia, Toxicologia i Química Terapèutica, UB) and tested against IAV-infected MDKC cells by the group of Prof. Lieve Naesens (Department of Microbiology, Immunology and Transplantation, KUL).

Project 1

"Binding of Cholesterol to the N-Terminal Domain of the NPC1L1 Transporter: Analysis of the Epimerization-Related Binding Selectivity and Loop Mutations"

"Binding of Cholesterol to the N-Terminal Domain of the NPC1L1 Transporter: Analysis of the Epimerization-Related Binding Selectivity and Loop Mutations"

Aitor Valdivia,[†,‡] F. Javier Luque,[‡,§,||] and Salomé LLabrés[‡,||]

† Doctorate in Biotechnology, Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació - Campus Torribera, Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain

‡ Institut de Biomedicina (IBUB), Universitat de Barcelona, 08028 Barcelona, Spain

§ Departament de Nutrició, Ciències de l'Alimentació I Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació - Campus Torribera, Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain

|| Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, 08921 Barcelona, Spain

Published in: Journal of Chemical Information and Modeling, **2024**, *64,* 180-204

Summary

Cholesterol is a fat-like substance with a pivotal physiological relevance in humans, and its homeostasis is tightly regulated by various cellular processes, including the import in the small intestine and the reabsorption in the biliary ducts by the Niemann−Pick C1 Like 1 (NPC1L1) importer. NPC1L1 can mediate the absorption of a variety of sterols but strikingly exhibits a large sensitivity to cholesterol epimerization. This study examines the molecular basis of the epimerization-related selective binding of cholesterol by combining extended unbiased molecular dynamics simulations of the apo and holo species of the N-terminal domain of wild-type NPC1L1, in conjunction with relative binding free energy, umbrella sampling, and well-tempered Metadynamics calculations. The analysis of the results discloses the existence of two distinct binding modes for cholesterol and epi-cholesterol. The former binds deeper in the cavity, forming key hydrogen-bond interactions with Q95, S56, and a water molecule. In contrast, epi-cholesterol is shifted ca. 3 Å to the mouth of the cavity and the transition to the Q95 site is prevented by an energetic barrier of 4.1 kcal·mol−1. Thus, the configuration of the hydroxyl group of cholesterol, together with the presence of a structural water molecule, is a key feature for effective absorption. Finally, whereas these findings may seemingly be challenged by single-point mutations that impair cholesterol transport but have a mild impact on the binding of cholesterol to the Q95 binding site, our results reveal that they have a drastic influence on the conformational landscape of the $\alpha 8/\beta 7$ loop in the apo species compared to the wild-type protein. Overall, the results give support to the functional role played by the $\alpha 8/\beta 7$ loop in regulating the access of ligands to NPC1L1, and hence to interpreting the impact of these mutations on diseases related to disruption of sterol absorption, paving the way to understanding certain physiological dysfunctions.

JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Open Access
This article is licensed under CC-BY 4.0

pubs.acs.org/jcim Article

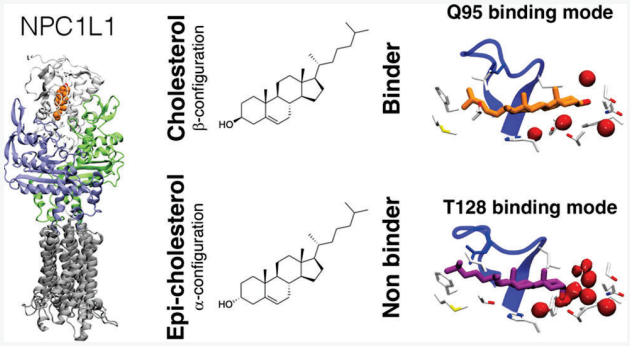
Binding of Cholesterol to the N-Terminal Domain of the NPC1L1 Transporter: Analysis of the Epimerization-Related Binding Selectivity and Loop Mutations

Aitor Valdivia, F. Javier Luque, and Salomé Llabrés*

Cite This: J. Chem. Inf. Model. 2024, 64, 189−204 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Cholesterol is a fat-like substance with a pivotal physiological relevance in humans, and its homeostasis is tightly regulated by various cellular processes, including the import in the small intestine and the reabsorption in the biliary ducts by the Niemann−Pick C1 Like 1 (NPC1L1) importer. NPC1L1 can mediate the absorption of a variety of sterols but strikingly exhibits a large sensitivity to cholesterol epimerization. This study examines the molecular basis of the epimerization-related selective binding of cholesterol by combining extended unbiased molecular dynamics simulations of the apo and holo species of the N-terminal domain of wild-type NPC1L1, in conjunction with relative binding free energy, umbrella sampling, and well-tempered metadynamics calculations. The analysis of the results discloses the existence of two distinct binding modes for cholesterol and epi-cholesterol. The former binds deeper in the cavity, forming key hydrogen-bond interactions with Q95, S56, and a water molecule. In contrast, epi-cholesterol is shifted ca. 3 Å to the mouth of the cavity and the transition to the Q95 site is prevented by an energetic barrier of 4.1 kcal·mol^{-1}. Thus, the configuration of the hydroxyl group of cholesterol, together with the presence of a structural water molecule, is a key feature for effective absorption. Finally, whereas these findings may seemingly be challenged by single-point mutations that impair cholesterol transport but have a mild impact on the binding of cholesterol to the Q95 binding site, our results reveal that they have a drastic influence on the conformational landscape of the $\alpha 8/\beta 7$ loop in the apo species compared to the wild-type protein. Overall, the results give support to the functional role played by the $\alpha 8/\beta 7$ loop in regulating the access of ligands to NPC1L1, and hence to interpreting the impact of these mutations on diseases related to disruption of sterol absorption, paving the way to understanding certain physiological dysfunctions.



Downloaded via 37.223.48.164 on March 9, 2024 at 11:40:09 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

INTRODUCTION

Cholesterol is an essential component of mammalian cell membranes and a key precursor of several metabolites, such as steroid hormones, vitamin D, and bile salts. It plays an important role in the rigidity and fluidity of the cell membrane and also in regulating the functional conformations of membrane proteins. The multiple physiological roles modulated by cholesterol demand a tight regulation of cholesterol homeostasis, which involves three main processes: endogenous biosynthesis, dietary absorption in the intestine, and net biliary excretion and reabsorption.[1,2] Alterations in the metabolic pathways that control cholesterol homeostasis are implicated in the etiology of various diseases and particularly have been associated with a higher risk of cardiovascular disease, neurodegeneration, and cancer.[1,3,4]

The Niemann−Pick C1 Like 1 (NPC1L1) protein is the primary importer of cholesterol in the small intestine and the reabsorption of sterols in the biliary duct.[5] NPC1L1 is also the molecular target of ezetimibe, which is one of the most prescribed drugs to lower high blood cholesterol levels.[6] From a structural point of view, NPC1L1 consists of three large extracellular (N-terminal, middle-luminal, and C-terminal) domains and a large transmembrane region that comprises 13 helices (Figure 1A).[7] The N-terminal domain (NTD) is free to unbind from the NPC1L1 core to capture cholesterol in extracellular media. Upon binding, the NTD folds over the other luminal domains, creating a continuous hydrophobic tunnel between the NTD and the membrane, thus enabling cholesterol to reach the cellular membrane.[7,8] Mutations in the NPC1L1 gene are associated with lower cholesterol absorption, which results in lower coronary heart disease and infarction risk.[9] Therefore, NPC1L1 inhibitors, such as ezetimibe,[10] open avenues for the effective regulation of

Received: August 17, 2023
Revised: December 14, 2023
Accepted: December 15, 2023
Published: December 28, 2023



ACS Publications
© 2023 The Authors. Published by American Chemical Society

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

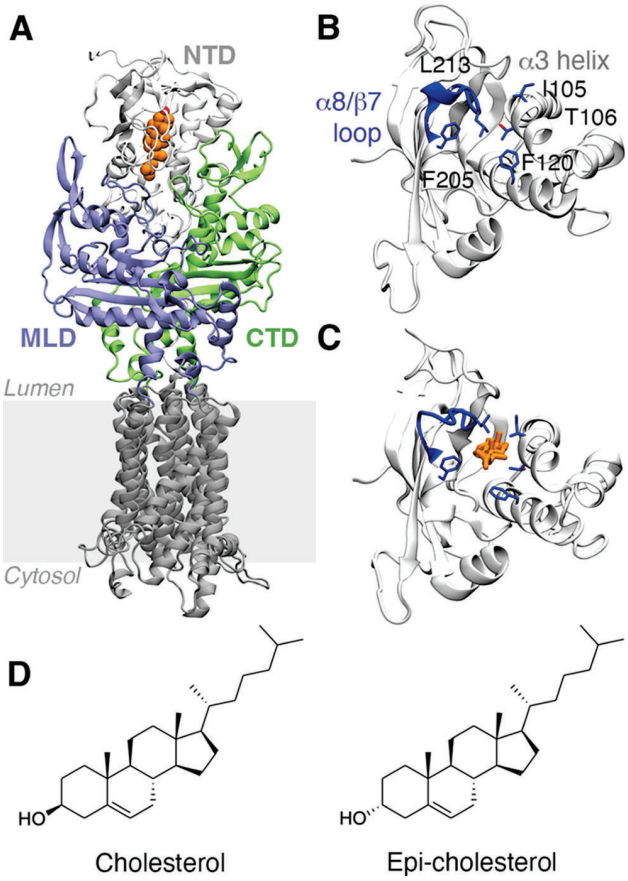


Figure 1. (A) Representation of the N-terminal (NTD), middle (MLD), and C-terminal (CTD) domains (shown as white, blue, and green cartoons, respectively) of NPC1L1 present in the luminal space and the transmembrane region. Detail of the entrance to the binding site of the NPC1L1-NTD in (B) the apo (X-ray structure; PDB 3QNT)[15] and (C) cholesterol-bound state. These structures illustrate the open-closed conformational change of the $\alpha 8/\beta 7$ loop, which regulates the access of sterols to the binding site in conjunction with residues I105, T106, F120, L213, and F205 (shown as blue sticks). The crystallographic position of cholesterol in NPC1 is shown as sticks (C atoms in orange). (D) Structure of cholesterol and epi-cholesterol.

cholesterol levels in plasma in patients with hypercholesterolemia[11] and sitosterolaemia[12] and can be used as coadjuvants in combination chemotherapies.[13,14]

NPC1L1 imports selectively cholesterol as well as sterol precursors and vitamins.[5] The molecular basis of its ligand specificity is encoded into the structural features of the NTD binding site.[16,17] The X-ray structure of the apo NTD has been solved in its closed conformation, which prevents access to the binding site via the steric blockage by the side chain of L213 of the $\alpha 8/\beta 7$ loop (Figure 1B). Although there is no X-ray structure for the cholesterol-bound NTD, details of the binding mode have been gained from the holo species of the homologous Niemann−Pick C1 (NPC1) protein (40% identity and 57% similarity; 41% identity and 64% similarity considering only the residues involved in the cholesterol-binding site), which has been solved in complex with cholesterol and 25-hydroxycholesterol.[17] Whereas the NPC1L1 structure shows an empty binding site, the structures of cholesterol-bound NPC1 reveal the identification of the binding mode of this compound as well as the hydration pattern within the binding site (Figures S1 and S2). Inspection of these NPC1 structures shows that residue Q200, equivalent to L213 in NPC1L1, is exposed to the solvent and interacts with helix $\alpha 3$, stabilizing the open conformation that enables access to the binding pocket (Figure 1C). Therefore, it has been hypothesized that the loop $\alpha 8/\beta 7$ plays a key role in the gating mechanism that regulates the sterol binding to NPC1L1.[16,17]

The structural similarity between the homologous domains of NPC1 and NPC1L1 paves the way to identifying the molecular basis of the selective binding of cholesterol and related sterols to the NPC1L1 transporter. In this regard, the drastic reduction in transport efficiency triggered upon epimerization of the hydroxyl group bound at carbon 3 in cholesterol (Figure 1D) is in contrast with the effective competition observed for other natural sterols that retain the β configuration of the hydroxyl group in cholesterol.[15,18] To gain insight into the molecular basis of the epimerization-related selective binding of cholesterol to NPC1L1, this study reports a series of extensive molecular dynamics (MD) simulations that have been used to explore the binding mode of cholesterol and epi-cholesterol. The analysis of the trajectories reveals the existence of distinct binding modes to the NTD domain of the NPC1L1. Free energy calculations have been used to examine the structural stability of the binding mode of these compounds and to estimate the differences in binding affinity between cholesterol and epi-cholesterol. The binding mode of cholesterol has been confirmed for different sterols that retain the β configuration of the hydroxyl group. Finally, the effect of distinct mutations on the binding of cholesterol is discussed by considering the impact of the mutated variants on the structural features implicated in ligand binding. The results highlight the subtle interplay between different factors in modulating the binding mode and affinity of sterols for NPC1L1, including the epimerization at the C3 position, the role of water molecules in assisting the stability of the binding mode, and the impact of mutations on the accessibility to the binding pocket.

METHODS

Setup of the Systems. The NPC1L1-NTD domain was modeled from the X-ray structure solved by Kwon et al. (PDB entry: 3QNT, resolution 2.83 Å),[15] which corresponds to the closed conformation that occludes the sterol binding pocket from solvent. The crystallographic structure of cholesterol in the NTD of the homologous protein human NPC1 (PDB entry: 3GKI, resolution 1.80 Å)[18] was used as a template for the binding mode of sterols in the ligand-bound complexes of NPC1L1-NTD. Given the almost perfect overlay of the residues that shape the binding pocket in the NPC1−cholesterol complex and those present in the NPC1L1-NTD, cholesterol was placed within the binding site of NPC1L1 by translating the coordinates of the ligand bound to NPC1. Indeed, only a slight steric overlap was detected between the side chain of L213 and the aliphatic chain of cholesterol. Nevertheless, it was alleviated through energy minimization of the system, which was the first step of the heating and equilibration protocol. The coordinates of the refined cholesterol−NPC1L1 system were subsequently used to model the binding of the sterol compounds.

The systems were solvated using a dodecahedral box of TIP3P water molecules accounting for a 12 Å spacing between the protein and the box edge. Counterions (Na^+ and Cl^-) were

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

added to maintain the neutrality of the simulated system at an ionic concentration of 0.15 M following the SPLIT method.[19] The simulation systems account for a total of ∼53,500 atoms in a simulation box of 88 × 83 × 72 Å^3. No crystallographic waters were included in the initial setup, as none are reported in any of the experimentally solved structures of the NPC1L1-NTD (Tables S1 and S2). Nevertheless, a 15 ns long equilibration protocol was used to ensure proper hydration of the binding site (see the next section for additional details).

The amberff14sb force field[20] was used for the protein, and the sterol molecules were parametrized using the Lipid14 force field.[21] Joung and Cheatham III parameters were used for the counterions,[22] and the TIP3P model[23] was used for the solvent. Partial atomic charges for the ligands were derived using the RESP protocol[24] at the HF/6-31G(d) level of theory, as calculated with Gaussian09.[25] Finally, the parameters were converted to Gromacs file-format using the AnteChamber PYthon Parser interfacE (ACPYPE) tool.[26]

Molecular Dynamics Simulations. Each system was minimized using 15,000 steps of steepest descent, followed by 5000 steps of conjugate gradient algorithm. Then, each system was gradually equilibrated in 9 steps. The systems were heated in the NVT ensemble from 100 to 300 K in three stages of 1 ns (100−150, 150−250, 250−300 K) with 5 kcal·mol^{-1} positional restraints on the backbone atoms of the protein and all of the atoms of the ligand. Subsequently, the density of the system was equilibrated for 12 ns in the NPT ensemble (pressure: 1 bar, T: 300 K). During the last 5 ns of the equilibration run, the positional restraints were slowly released, decreasing 1.0 kcal·mol^{-1} per ns.

Production runs encompassed 4 replicas of 0.5 μs per system, accounting for a total simulation time of 2 μs per system (for a total of 28 μs of simulation time). Temperature control was achieved using a V-rescale thermostat.[27] Pressure control was maintained by using the Parrinello−Rahman method. All bonds involving hydrogen atoms were constrained by the LINCS algorithm[28] to allow for a time step of 2 fs. All simulations were performed with the Gromacs 2022.4 package.[29]

Molecular Mechanics Poisson−Boltzmann Surface Area (MM-PBSA) Calculations. The binding free energy of sterol molecules to NPC1L1-NTD was estimated using MM-PBSA calculations with the MMPBSA.py script available in AmberTools22.[30] The theoretical details of the method are described in detail elsewhere.[31] For each sterol-bound complex, a total of 400 configurations (100 taken from each independent replicate) were extracted and preprocessed to include key structural water present in most of the simulations (see below) as part of the receptor. Dielectric constants of 2 and 80 were used for the interior of the protein and the bulk solvent, respectively, in conjunction with an ionic strength of 0.15 M.[32] No entropic contributions were calculated since the final aim was to estimate the relative binding free energies (RBFE) of structurally related, largely rigid ligands that adopt a similar arrangement in the binding pocket.

The *cpptraj* package was used for the analysis of the simulated systems. The volume of the binding pocket was calculated with MDPocket.[33]

Binding Free Energy Calculations. The relative binding free energy (RBFE)[34] between sterols was determined through alchemical transformations where a ligand (L1) is converted into a structurally related analogue (L2) in both the protein-bound complex and in the unbound state in aqueous solution (eq 1). Accordingly, independent sets of simulations were carried out for the transformation between L1 and L2 in water and in the protein.

$$\Delta\Delta G_{\mathrm{bind}}(\mathrm{L1} \rightarrow \mathrm{L2}) = \Delta G_{\mathrm{bound}}(\mathrm{L1} \rightarrow \mathrm{L2}) - \Delta G_{\mathrm{water}}(\mathrm{L1} \rightarrow \mathrm{L2}) \quad (1)$$

The transformation from L1 to L2 was split in a set of windows, where $\lambda = 0$ and $\lambda = 1$ correspond to L1 and L2, respectively, and the intermediate λ values correspond to a linear interpolation between the parameters of the initial and final systems.[35] A dual topology alchemical perturbation protocol was employed by using 21 evenly spaced λ-windows. To smoothly transform L1 to L2, we applied soft-core potentials using soft-core electrostatics[36] as described in AMBER22[35] to the ring A of the sterol core. All of the λ simulations were run for 6 ns, accounting for a total of 126 ns per transformation (12 transformations, over 1.5 μs of RBFE simulations). The Thermodynamic Integration (TI) estimator, as implemented in *alchemlyb*,[37] was used to estimate the free energy change for each λ simulation. All simulations were carried out with the GPU-accelerated TI implementation[34,38] of Amber22.

The initial coordinates for each binding mode were extracted from the unbiased MD runs. All of the systems were carefully equilibrated with the dual topology using the TI code in each transformation with $\lambda = 0.5$. Each system was first heated in the NVT ensemble from 5 to 300 K (200 ps) and subsequently equilibrated at 1 bar in the NPT ensemble (300 ps). During the equilibration protocol, the heavy atoms of the protein and ligand were restrained with positional restraints (5 kcal·mol^{-1}). Each λ state was then simulated in the NVT ensemble at 300 K by using the Langevin thermostat and the SHAKE algorithm for bonds involving hydrogen atoms.

Each simulation was repeated three times to assess reproducibility of the free energy changes. The resulting binding free energies were averaged, and their standard deviations were used for the error analysis. Additionally, we calculated the cycle closure error to assess the convergence of the simulations. Convergence was reached when the cycle closure errors were below 0.3 kcal·mol^{-1}.

Umbrella Sampling (US) Calculations. To estimate the potential of mean force (PMF) for the transition of epi-cholesterol between the two binding modes in the ligand cavity, a series of five steered MD simulations were performed to sample the shift from the T128 anchoring site to the Q95 binding one.[39] To this end, the distance between the N atom of the Q95 side chain amide group and the O atom of epi-cholesterol was chosen as a transition coordinate. Steered MD simulations were started by using snapshots taken from the unbiased MD runs, where the T128 binding mode was stable. The transition of epi-cholesterol between the two anchoring sites was driven by using a speed of 0.5 Å·ns^{-1} in conjunction with a force constant of 500 kcal·mol^{-1} Å^{-2} (hence using the stiff-spring approximation).

The trajectory that showed lower work profiles was selected for the umbrella sampling (US) calculations, which were performed to estimate the potential of mean force (PMF) for the transition between the Q95 and T128 binding modes of epi-cholesterol. The distance between the N atom of the Q95 side chain and the O atom of the epi-cholesterol molecule was varied within a range of 2.5−8 Å and an interval of 0.5 Å, accounting for a total of 12 bins. The initial coordinates of

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

each bin were selected by choosing the snapshot of the steered MD trajectories with the closest value to the ideal transition coordinate. Harmonic potential restraints of 1 kcal·mol^{-1} Å^{-2} were applied to avoid deviations from the target values along the transition coordinate that connects the two binding states. Each bin was run for 6 ns using the same conditions described in the MD section. Values of the reaction coordinates were stored every 10 fs for postprocessing. The WHAM method[40] was used to obtain the PMF profile as implemented in the GROMACS package (*gmx wham*), and the error bars were computed by bootstrapping 100 times.

Well-Tempered Metadynamics Simulations. To further check the results obtained from the US calculations, multiple-walkers well-tempered metadynamics simulations (MW-WTM)[41−43] were used to examine the transition of epi-cholesterol between the two binding modes defined by interactions with Q95 and T128 in the ligand cavity. Two collective variables were defined to guide conformational sampling. The first collective variable (CV1) accounts for the difference between the distances of the H-bonds formed by the oxygen atom of epi-cholesterol with the side chains of Q95 and T128 (the Q95 and T128 binding modes are mapped onto the negative and positive values of CV1, respectively). The second collective variable (CV2) was chosen to follow the conformational fluctuations of the S-entrance by measuring the distance between the center-of-mass (COM) of residues 212−214 of the $\alpha 8\beta 7$ loop and the COM of residues 104−106 of α-helix 3. Lower and upper walls were placed to keep the distances between the O atom of epi-cholesterol and the N/O atom of the Q95/T128 side chain within the range of 2−10 Å. Likewise, CV2 was maintained between 7 and 17 Å to sample the conformational space of the S-entrance in the unbiased MD simulations and to avoid drastic changes (i.e., unfolding) in the $\alpha 8\beta 7$ loop. Six independent walkers[44] were used to contribute simultaneously to the same history-dependent bias potential and to explore the free energy surface (FES). The deposition rate of the Gaussian biasing terms was set to 1 ps, their initial height to 0.4 kcal mol^{-1}, and a bias factor of 15. Initial positions for each walker were set along the path that connects the Q95 and T128 binding modes obtained from US calculations. The system was simulated for a total of 1.2 μs, and the bias potential was updated every 0.2 ps. The minimum energy path between the Q95 and T129 binding modes was obtained using the string method[45,46] over the FES.

Principal Component Analysis (PCA). To explore the conformational landscape of the S-opening in the NPC1L1-NTD upon cholesterol binding and absorption impairing mutations, we performed principal component analysis (PCA) onto the apo and cholesterol-bound trajectories of wild-type and mutant domains. We only included the main-chain atoms (C, Cα, N, and O atoms) of the $\alpha 8/\beta 7$ loop (from Q206 to L216) and the $\alpha 3$ helix (from A93 to L110). The trajectories were properly aligned to the equilibrated structure of NPC1L1-NTD in the apo state. Then, the eigenvalues and eigenvectors were calculated by diagonalizing the mass-weighted covariance matrix of the system using the *gmx covar* and *gmx anaeig* modules of the GROMACS 2022.4 package.

The relative population of each conformational state sampled for the apo- and cholesterol-bound systems was estimated after projection in the space defined by the first two principal components (PCs), which account for 66% of the total structural variance (see below). To this end, we used the k-means clustering with 100 clusters to discretize the PC space, and then these clusters were further classified into eight conformational states (five for the apo and three for the cholesterol-bound systems) using the Robust Perron Cluster Analysis method (PCCA+)[47] as implemented in the PyEMMA 2.0 module.[48] The error bars of these populations were obtained by bootstrapping 5 times. Fifty structures of each of these conformational states were extracted for visualization and structural analysis purposes.

RESULTS AND DISCUSSION

Conformational Changes of the $\alpha 8/\beta 7$ Loop of NPC1L1 upon Ligand. The NTD domains of both NPC1L1 and NPC1 have a common protein fold, as noted in a positional root-mean-square deviation (RMSD) of 1.6 Å determined for the Cα atoms (Figures S1 and S2); however, comparison of their structures highlights the subtle but relevant rearrangement of the $\alpha 8/\beta 7$ loop, which affects the accessibility to the binding pocket. In the NPC1L1-NTD, the side chain of L213 is buried into the entrance of the binding site, thus preventing ligand binding. In contrast, residue Q200 of the $\alpha 8/\beta 7$ loop of NPC1-NTD, equivalent to L213 of NPC1L1-NTD, interacts with N86 of the $\alpha 3$ helix and stabilizes the open conformation.[17] This structural rearrangement enlarges the entrance to the binding pocket in NPC1 compared to NPC1L1, as noted in the width of the cavity mouth, which increases from 7.6 Å in NPC1L1-NTD (X-ray structure 3QNT) to 9.3 Å in NPC1-NTD (X-ray structure 3GKI) measured as the distance between the C$_\beta$ atoms of L213 and T106 of NPC1L1-NTD, and the C$_\beta$ atoms of Q200 and P90 of NPC1-NTD.

The structural stability of the apo- and cholesterol-bound states of NPC1L1-NTD was examined by means of extensive unbiased MD simulations. Four independent replicates, each covering 0.5 μs, were run for each system, thus leading to a cumulative simulation time of 2 μs for the apo and cholesterol-bound forms of NPC1L1-NTD. Stable RMSD profiles were obtained in the four replicates run for both apo and holo systems (Figures S3 and S4)

Our simulations confirm that, in the apo state, the side chain of L213 interacts with L103, T106, F120, and F205, thus forming a stable hydrophobic cluster that keeps the closed conformation, which hinders the entrance to the binding site (Figure 2; see also Figure S5), also known as S-opening. These residues bury the side chain of L213 in the binding site close to the phenyl ring of F120 for most of the simulation time (average distance of 5.4 ± 0.6 Å; present in 58% of the simulation time). This conformation is further stabilized by a H-bond interaction between the backbone of L213 and S102 on the $\alpha 3$ helix (average distance of 2.7 ± 0.1 Å; 45% of the simulation time), which brings the $\alpha 8/\beta 7$ loop upon the $\alpha 3$ helix and tightens this hydrophobic cluster. This causes a slight distortion from the X-ray structure, as the RMSD of the Cα atoms of the $\alpha 8/\beta 7$ loop is on average 2.0 ± 0.9 Å, accompanied by a decrease in the width of the cavity mouth, which amounts on average to 6.5 ± 0.6 Å (average distance from the Cβ atoms of the L213 and T106), compared to the X-ray structure (distance of 7.6 Å). The side chain of L213 can also adopt a distinct conformation that enables the interaction with I105 (average distance of 5.5 ± 0.3 Å; 37% of the simulation time) but without altering the entrance of the binding site (6.3 ± 0.8 Å). Altogether, the two conformations

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

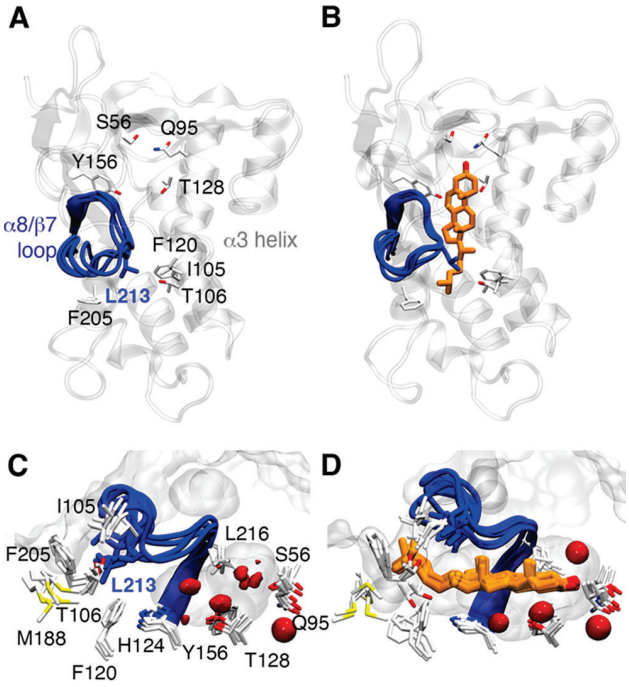


Figure 2. Superposition of the final structures of the MD trajectories of the apo (A, C) and cholesterol-bound states (B, D). (A, B) The conformations sampled by the $\alpha 8/\beta 7$ loop are shown in a blue cartoon. (C, D) Shape and solvation pattern of the binding site of NPC1L1-NTD for the apo and cholesterol-bound states. The protein surface is shown as a gray surface. The red isosurface accounts for 50% of water occupancy in the binding site during the last 100 ns of each replicate (aggregated 400 ns of simulation time).

adopted by L213 restrict the volume of the binding site, which amounts to 374 ± 4 Å^3.

Despite the hydrophobic nature of the cavity, between 6 and 16 water molecules were found in the binding site, with an average number of 11 ± 2 waters for the four replicates (Figure S6). The access of water molecules to the binding site occurs through the W opening, which is shaped by the $\alpha 1$ and $\alpha 2$ helices that lie on top of the cavity. Also, the binding pocket has a subset of polar residues composed of Q95, S56, Y156, T128, and H124 (Figure 2) that interact with a network of water molecules that fill the binding pocket. This network overlaps nicely with the waters detected in the crystallographic structure of the apo state of the NPC1 (Figure S7).[17]

The binding of cholesterol requires the conformational rearrangement of the side chain of L213, which interacts with the ligand's side chain (average distance of 4.9 ± 0.1 and 5.2 ± 0.1 Å with the carbon atom of the methyl group in position C18 of the sterol moiety and the first carbon atom of the aliphatic side chain, respectively). This interaction is facilitated by the fact that L213 adopts a conformation that mimics the arrangement of Q200 in NPC1-NTD, where its side chain points to I105 in the $\alpha 3$ helix (Figure S7). The presence of cholesterol in the binding site forces the open conformation of the pocket, which is stable and exhibits small fluctuations in the four replicates (Figure 2; see Figure S3) in agreement with previous studies.[16] Thus, the average RMSD determined for the Cα atoms of the $\alpha 8/\beta 7$ loop relative to the X-ray structure of NPC1-NTD is 2.3 ± 0.1 Å. Compared to the apo species, the expansion of the cavity is reflected in the width of the mouth of the binding site, which on average amounts to 8.3 ± 0.5 Å, and in the enlargement of the cavity volume (average value of 1031 ± 18 Å^3), it being around 3-fold larger than the volume determined for the apo species.

The crystallographic binding mode of cholesterol was fully maintained in the simulations of the complex with NPC1L1-NTD (Figure 2). The hydroxyl moiety of cholesterol can establish three H-bond interactions with neighboring residues: the amide NH of Q95, the hydroxyl group of S56, and a water-mediated interaction with the backbone of Q54 [let us note that this water molecule is also present in the experimental structure of NPC1 (Figure S7)].[17] While the H-bond interactions with Q95 are stable along all of the replicates (average distance of 2.9 ± 0.1 Å; 66% of the simulation time), the S56 side chain fluctuates between its interaction with the ligand (average distance of 2.8 ± 0.1 Å; detected in 17% of the simulation time) and the backbone carbonyl oxygen of D37. The hydrophobic tetracyclic system of cholesterol interacts with residues L99, L103, and H124 of the α-helices of the NTD core and residues A214 and P215 of the $\alpha 8/\beta 7$ loop. Finally, the aliphatic chain is partially exposed to the bulk solvent at the S-opening, where it interacts with L213 and F205 (Figure S8). Hereafter, given the increased stability of the interaction with Q95, this binding mode will be referred to as the Q95 binding mode.

The binding of cholesterol displaces most of the water molecules present in the binding site, which, however, still contains an average number of 5 ± 1 water molecules (Figures S7 and S9). The 50% occupancy isocontour of the water density accounts for the water wire that extends from H124 to Q95, highlighting the position of three structural waters. The first two mediate interactions between residues H124 and Y156, and the third establishes an H-bond interaction with the ß-hydroxyl moiety of cholesterol. Let us note that the presence of water molecules is not uncommon in cholesterol-binding sites,[49] and the homologous NPC1-NTD contains two crystallographic water molecules that overlap with the water wire positions (Figure S7).

Epi-Cholesterol Binds Using Two Binding Modes in the NPC1L1-NTD Domain. To understand the molecular basis of the inability of epi-cholesterol to effectively compete with cholesterol for binding to NPC1L1-NTD, four independent unbiased MD simulations of the epi-cholesterol-bound complex were performed. We used the crystallographic binding mode of cholesterol to NPC1-NTD as a template to model the starting binding mode of epi-cholesterol in the NPC1L1-NTD complexes.

The analysis of the trajectories showed that epi-cholesterol displays two distinct binding modes (Figure 3; see also Figure S10). In the first one, epi-cholesterol tries to adopt the Q95 binding mode of cholesterol. However, to keep the H-bond with the side chain of Q95, the orientation of the hydroxyl group forces the sterol core to move closer to the loops with a concomitant displacement of the crystallographic water (Figure 3A,C). In the second, epi-cholesterol is shifted toward the S-opening to establish a H-bond interaction with the side chain of T128 (Figure 3B,D; average distance of 2.8 ± 0.1 Å; present in 72% of the simulation time). We have named this alternative pose as the T128 binding mode. Binding to T128 enlarges the exposure of the aliphatic chain of epi-cholesterol to the bulk solvent, although this effect is counterbalanced by the interaction of the aliphatic side chain with M188 (6.4 ± 1.9% SASA of epi-cholesterol compared to the 3.8 ± 0.6% SASA of epi-cholesterol in the Q95 binding mode).

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

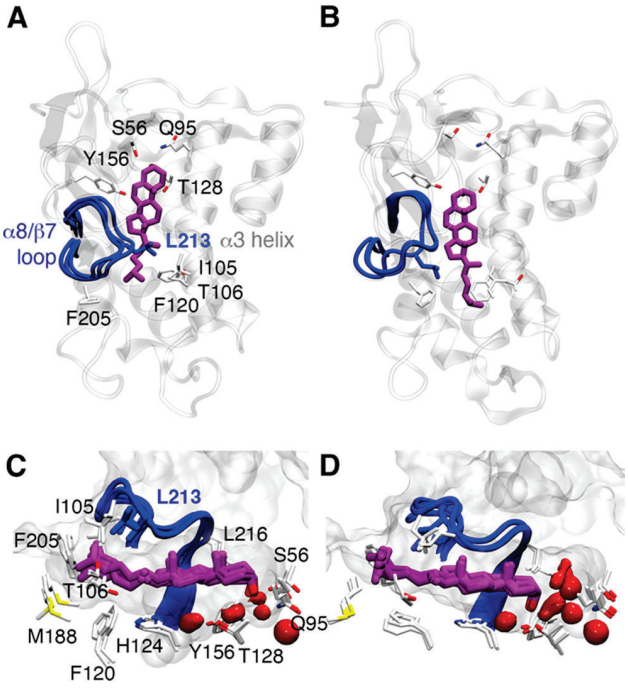
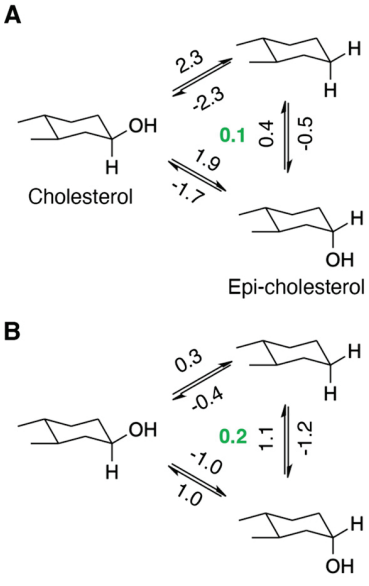


Figure 3. Superposition of the final structures of the MD trajectories of the Q95 (A, C) and T128 binding modes (B, D) of the epi-cholesterol-bound states. (A, B) The conformations sampled by the $\alpha 8/\beta 7$ loop are shown in a blue cartoon. Epi-cholesterol is shown in purple sticks. (C, D) Shape and solvation pattern of the binding site of NPC1L1-NTD for the epi-cholesterol-bound states. The protein surface is shown as a gray surface. The red isosurface accounts for 50% of water occupancy in the binding site during the last 100 ns of each replicate (accumulated 200 ns of simulation time).

Furthermore, the shift of epi-cholesterol to the T128 binding mode causes the side chain of L213 to move away from I105 while it becomes closer to F205. Therefore, the $\alpha 8/\beta 7$ loop opens slightly when epi-cholesterol binds to T128 compared to the Q95 binding mode, leading to a slight increase of the cavity volume (1153.2 ± 30.9 and 1095.0 ± 7.7 Å^3, respectively) and filling the cavity with a higher number of water molecules (9 ± 1 and 7 ± 1 water molecules; Figure S11).

It is worth noting that the binding of epi-cholesterol to the Q95 and T128 binding sites exhibits different traits between the four independent replicas. Thus, epi-cholesterol remains primarily bound to T128 in replicas 1 and 3, whereas the occupancy of the Q95 binding site is larger in replicas 2 and 4. Furthermore, differences are also observed in the number of exchanges between the two binding sites, as measured from the difference between the distances of the H-bond interactions formed by epi-cholesterol with Q95 and T128 in conjunction with a threshold of 1.5 Å to split the two binding modes. This analysis revealed that the transitions of epi-cholesterol between Q95 and T128 sites varied from ∼30 exchanges in replica 2 to almost 560 exchanges in replica 3. These differences seem to arise from the behavior of the $\alpha 8/\beta 7$ loop, which must open slightly to facilitate the transitions of epi-cholesterol in the interior of the cavity (see above and Figure 3). Therefore, the conformational plasticity of this loop appears to be a relevant factor in assisting the binding of the ligands to NPC1L1-NTD.

On the other hand, these results point out that the sampling obtained from the four independent MD simulations is not complete, and a proper quantitative analysis of the preferences for binding to the Q95 and T128 sites should be gained by using enhanced sampling techniques. In this context, RBFE calculations[50,51] were first carried out to obtain a quantitative estimate of the change in binding affinity due to epimerization of the hydroxyl group in cholesterol. To this end, alchemical transformations were performed for the conversion of cholesterol and epi-cholesterol to cholestane, which lacks the hydroxyl group and can therefore be used as a reference to estimate the contribution of the H-bond to the binding of



C

Perturbation	ΔΔG / kcal mol^{-1}				
Q95 binding site	R1	R2	R3	Mean	StDev
CHL → C5N	2.4	2.1	2.5	2.3	0.2
C5N → CHL	-2.4	-2.0	-2.5	-2.3	0.2
ECL → C5N	0.7	0.2	0.3	0.4	0.3
C5N → ECL	-0.4	-0.6	-0.5	-0.5	0.1
CHL → ECL	1.8	1.5	2.3	1.9	0.4
ECL → CHL	-1.8	-1.4	-1.7	-1.7	0.2
Cycle closure	0.2	-0.4	0.3	0.1	0.4
T128 binding site	R1	R2	R3	Mean	StDev
CHL → C5N	0.1	0.5	0.4	0.3	0.2
C5N → CHL	-0.3	-0.8	-0.1	-0.4	0.4
ECL → C5N	1.2	1.2	1.1	1.1	0.1
C5N → ECL	-1.2	-1.1	-1.3	-1.2	0.1
CHL → ECL	-1.5	-0.8	-0.8	-1.0	0.4
ECL → CHL	1.1	0.7	1.3	1.0	0.3
Cycle closure	-0.6	-0.5	0.5	0.2	0.6

Figure 4. (A, B) Perturbation network of the chemical changes on position 3 of the cholesterol ring for the (A) Q95 binding mode and (B) T128 binding mode. Arrows indicate alchemical simulations carried out with computed relative free energies in kcal mol^{-1}. Green numbers indicate average cycle closure errors. (C) Relative free energy changes for the three replicates and the mean and standard deviation of the $\Delta\Delta G$ for each transformation. CHL: cholesterol, ECL: epi-cholesterol, and C5N: cholestane.

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

cholesterol and epi-cholesterol to the Q95 and T128 binding sites. The simulation systems included the key water molecule that assists the binding of sterols in the Q95 binding mode.[52,53] The relative free energy values are obtained from the average values of three independent runs (Figure 4).

The results point out that the H-bond formed by cholesterol with Q95 has a net stabilizing contribution of 2.3 kcal·mol^{-1} relative to cholestane, whereas the contribution of the hydroxyl group in epi-cholesterol is around 5-fold less favorable. Indeed, cholesterol binds more favorably to the Q95 binding site than epi-cholesterol by 1.8 kcal·mol^{-1}, thus reflecting the inability of the α-configuration of the hydroxyl group to form an effective H-bond in the Q95 binding mode. On the contrary, epi-cholesterol binds more favorably than cholesterol by 1.0 kcal·mol^{-1} in the T128 binding mode, reflecting the formation of the H-bond with T128, which is unfeasible for the β-configuration of the hydroxyl group. In fact, the difference in binding free energy between cholesterol and cholestane is ca. 0.3 kcal·mol^{-1}.

These results provide a rationale for the stability of the binding mode observed for cholesterol in the unbiased MD simulations, which showed that the ligand remains tightly bound to Q95, in contrast to the transition observed for epi-cholesterol to the T128 binding mode in two trajectories. Nevertheless, keeping in mind the limited sampling attained from the four independent unbiased simulations, we employed two enhanced sampling techniques to estimate the energetic cost of the transition: one-dimensional umbrella sampling calculations (US) using only the formation of the H-bond with Q95 side, and bidimensional well-tempered metadynamics simulations that also consider the opening of the $\alpha 8\beta 7$ loop. First, the potential of mean force (PMF) determined from US calculations reveals the presence of two minima at 2.9 and 6.5 Å, which correspond to the Q95 and T128 binding modes, respectively. The Q95 binding mode is destabilized by approximately 1.1 kcal mol^{-1} with respect to the global minima at the T128 binding mode. The energetic barrier for the transition from the T128 binding site to the Q95 one is 3.6 kcal mol^{-1}, whereas the barrier for the Q95 →T128 transition is roughly 2.5 kcal mol^{-1}. These energetic barriers were refined using well-tempered metadynamics simulations, which exploit the difference of distances between the O atom of epi-cholesterol and the Q95 and T128 side chains (CV1) and the opening of the $\alpha 8\beta 7$ loop (CV2), defined as the distance between the center-of-mass (COM) of residues 212−214 within the $\alpha 8\beta 7$ loop and residues 104−106 located at the $\alpha 3$ helix.[54] The resulting FES (Figure 5) shows two minima that correspond to the Q95 and T128 binding modes. The global minimum located at the T128 binding mode is more stable than the Q95 binding mode by ca. 2.0 kcal mol^{-1}. While the Q95 binding mode shows a preference for closed conformations of the loop, the T128 binding site can accommodate a larger range of loop conformations. The minimum energy path between the two minima obtained by the String method[45] is shown in a black line on top of the FES (Figure 5). The barrier for the passage from Q95 to T128 amounts to ca. 2.1 kcal mol^{-1}, whereas the barrier for the T128 →Q95 transition is close to 4.1 kcal mol^{-1}. The energetic profiles obtained by the well-tempered metadynamics simulations and US calculations are qualitatively similar. However, it is also clear that the motions of the S-entrance may have an impact on the quantitative features of the FES, thus highlighting the coupling between the fluctuations between the two H-bond states of

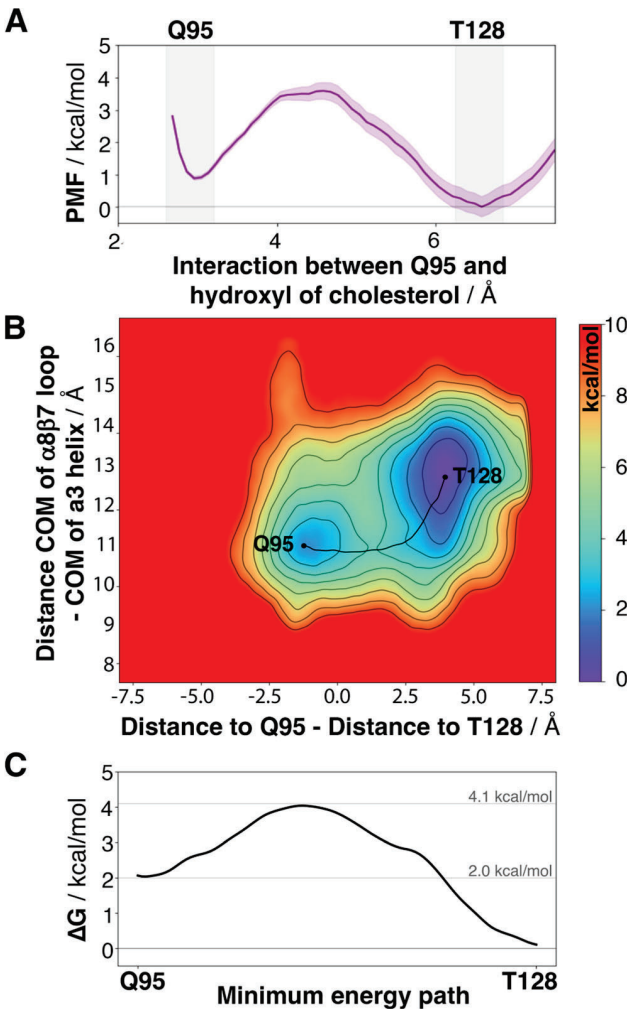


Figure 5. (A) PMF of the interconversion between the Q95 and T128 binding modes obtained by US is in kcal mol^{-1}. (B) Free energy surface of the transition between the Q95 and T128 binding modes for epi-cholesterol obtained via two-dimensional (2D)-well-tempered metadynamics. The distances that account for the formation of respective H-bonds between Q95 and T128 with the hydroxyl of epi-cholesterol. The minimum energy path between Q95 and T128 is shown as a black line. The color bar is in kcal mol^{-1}. (C) The energetic profile of the minimum energy path between binding modes sampled from the bidimensional free energy surface in kcal mol^{-1}.

epi-cholesterol and the conformational flexibility of the $\alpha 8\beta 7$ loop and the $\alpha 3$ helix. Overall, these results support the existence of distinct binding modes for cholesterol and epi-cholesterol, revealing that epimerization of cholesterol weakens the binding affinity by promoting the binding to a less buried site in the cavity (Figures 4 and 5).

Binding Affinity of 25-Hydroxycholesterol and Lanosterol to NP1CL1-NTD. To further explore these findings, an additional set of four independent MD simulations was performed for the complexes of NPC1L1-NTD with 25-hydroxycholesterol and lanosterol (Figure 5) since ^{3}H-cholesterol competitive binding assays have shown that these compounds can displace cholesterol from the NPC1L1-NTD binding site.[15,18] Following the previous discussion, the analysis was focused on the effect of the chemical differences present in these compounds relative to cholesterol (i.e., the location or the double bond and the presence of methyl groups

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

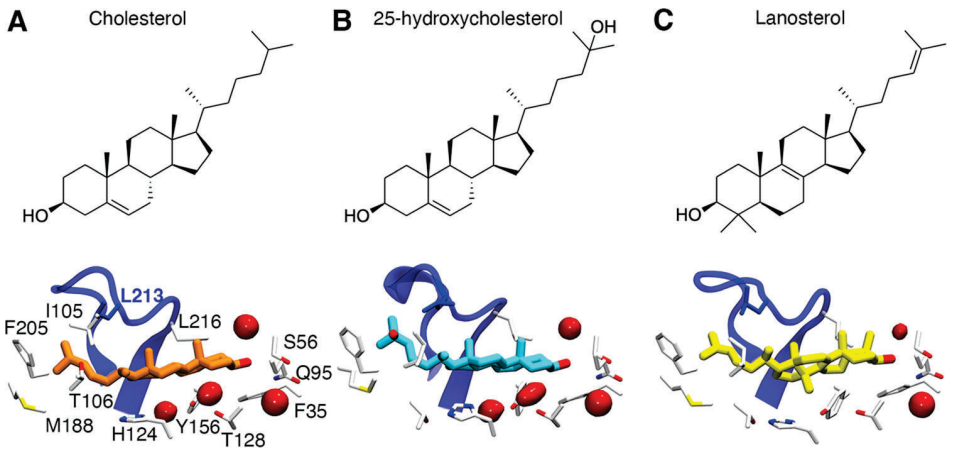


Figure 6. Chemical structures and representative structures of the binding modes of 25-hydroxy-cholesterol and lanosterol bound to the NPC1L1-NTD. Sterols are shown as colored sticks: cholesterol (A, orange), 25-hydroxycholesterol (B, blue), and lanosterol (C, yellow). NPC1L1-NTD is shown as a white cartoon. The position of the $\alpha 8/\beta 7$ loop is shown as a blue cartoon. The occupancy of water molecules in the binding site over 50% of the simulation time is shown as a red isosurface.

Table 1. Binding Free Energies [Average and Standard Deviation (SD); kcal·mol^{-1}] Obtained from MM-PBSA Calculations for the Studied Sterols in Complex with NPC1L1-NTD

	MM-PBSA without water molecule			MM-PBSA with water molecule		
	mean	SD	standard error of the mean	mean	SD	standard error of the mean
cholesterol	−16.0	0.6	0.3	−20.8	0.9	0.5
epi-cholesterol Q95 binding mode	−15.6	2.2	1.5			
epi-cholesterol T128 binding mode	−19.2	0.2	0.2			
25-hydroxy-cholesterol	−15.0	1.5	0.7	−19.3	2.6	1.3
lanosterol	−15.0	2.4	1.2	−19.2	2.3	1.1

in the tetracyclic core and the hydroxyl or double bond introduced in the aliphatic chain) on the binding to NPC1L1-NTD.

The RMSD of the protein backbone was stable in all cases, with values around 2 Å for the two sterol-bound complexes (see Figures S13 and S14). Compared to cholesterol, which is stably bound (RMSD close to 1 Å), slightly larger RMSD values, generally lower than 3 Å, were observed for the ligand in the complexes with NPC1L1-NTD, which in most cases reflects fluctuations of the aliphatic side chain of the sterols outside the binding pocket.

The chemical modifications introduced in these compounds do not alter the Q95 binding mode of the sterols (Figure 6). The binding mode of 25-hydroxycholesterol retains all of the described interactions for cholesterol, especially regarding the H-bond with Q95. On the other hand, the presence of the additional hydroxyl group bound at the C25 position can establish additional interactions with the solvent, the side chain of T106 (average distance of 2.8 ± 0.1 Å; present in 25% of the simulation time), the backbone of S102 (average distance of 2.8 ± 0.1 Å; 17% of the simulation time), and the side chain of Q206 (average distance of 2.9 ± 0.1 Å; 8% of the simulation time) (Figure S13 and Table S3).

Despite the chemical changes introduced in the core of the sterol moiety, lanosterol establishes the same interactions described for cholesterol in the Q95 binding mode (Figure 6). The additional methyl groups in the C4 position fit well into the pocket shaped by residues L216, Y156, and F35 and do not impair the formation of H-bonds with Q95 and S56 (Figure S14 and Table S3). Thus, an average distance of 2.9 ± 0.1 Å was determined for the H-bond interaction with Q95 (present in 70% of the simulation time), whereas the interaction with the S56 side chain (average distance of 2.8 ± 0.1 Å) is found in 13% of the simulation time.

The number of water molecules found in the binding site is similar for all of the sterols in the Q95 binding mode, as it amounts on average to 5 ± 1, 6 ± 1, and 4 ± 1 for cholesterol, 25-hydroxycholesterol, and lanosterol, respectively (Figures 6 and S15). In the latter case, the lower content of water molecules can be attributed to the disruption of the water wire caused by the two methyl groups bound at the C4 position. In all sterol complexes, the hydroxyl moiety of the sterol core establishes an H-bond interaction with a water molecule, which, as mentioned before, occupies the same position found in the crystallographic structure of the cholesterol−NPC1 complex (Figure S7).

Finally, MM-PBSA calculations were performed to examine the contribution of the residues in the binding pocket as well as of the water molecule to the binding affinity and selectivity of cholesterol, epi-cholesterol, 25-hydroxycholesterol, and lanosterol (Table 1). Including the water molecule, the binding affinity of cholesterol is estimated to be −20.8 ± 0.9 kcal·mol^{-1}, which is close to the values obtained for 25-hydroxycholesterol and lanosterol (−19.3 ± 2.6 and −19.2 ± 2.3 and, respectively). This agrees with the experimental evidence,[7,15,18] which shows that both 25-hydroxycholesterol and lanosterol exhibit a direct competition with cholesterol for the binding to NPC1L1-NTD. Notably, the absence of the water molecule reduces the binding affinity to 15−16 kcal·mol^{-1}. In contrast, the binding affinity estimated for the two binding modes of epi-cholesterol is −15.6 ± 2.2 (Q95 binding site) and −19.2 ± 0.2 kcal·mol^{-1} (T128 binding site).

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

Table 2. Per-Residue Decomposition to the Binding Affinity (kcal·mol^{-1}) Obtained from MM-PBSA for All of the Studied Sterols in Complex with NPC1L1-NTD[a]

residue	cholesterol	epi-cholesterol Q95 binding mode	epi-cholesterol T128 binding mode	25-hydroxy-cholesterol	lanosterol
E38	−0.6 ± 0.4	0.5 ± 0.6	−0.4 ± 0.2	−0.6 ± 0.3	−0.5 ± 0.5
N54	−0.5 ± 0.3	0.1 ± 0.4	−0.1 ± 0.1	−0.4 ± 0.3	−0.2 ± 0.3
S56	−1.6 ± 0.9	−0.7 ± 0.3	−0.0 ± 0.0	−1.4 ± 0.8	−0.7 ± 0.2
T61	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**
K94	−0.2 ± 0.1	−0.2 ± 0.1	0.1 ± 0.1	−0.2 ± 0.1	−0.1 ± 0.1
Q95	−3.4 ± 0.5	−3.2 ± 0.5	0.0 ± 0.2	−3.5 ± 0.5	−3.6 ± 0.5
S98	−0.5 ± 0.3	−0.7 ± 0.4	−0.1 ± 0.1	−0.6 ± 0.3	−0.6 ± 0.3
L99	−1.9 ± 0.3	−2.0 ± 0.3	−1.4 ± 0.3	−1.9 ± 0.3	−2.2 ± 0.3
L103	−1.2 ± 0.2	−1.2 ± 0.2	−1.6 ± 0.3	−1.0 ± 0.2	−1.1 ± 0.3
I105	−0.2 ± 0.1	-0.2 ± 0.1	−1.1 ± 0.3	−0.4 ± 0.2	−0.3 ± 0.2
L110	**0.0 ± 0.0**	**0.0 ± 0.0**	−0.2 ± 0.1	**0.0 ± 0.0**	**0.0 ± 0.0**
F120	−0.6 ± 0.3	−0.6 ± 0.2	−1.0 ± 0.2	−0.6 ± 0.2	−0.6 ± 0.2
H124	−1.3 ± 0.4	−1.0 ± 0.4	−1.2 ± 0.5	−1.0 ± 0.3	−0.6 ± 0.5
T128	−0.3 ± 0.3	−0.5 ± 0.4	−2.1 ± 0.1	−0.3 ± 0.2	−0.3 ± 0.2
N132	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**	**0.0 ± 0.0**
F205	−0.2 ± 0.3	−0.2 ± 0.3	−0.6 ± 0.2	−0.3 ± 0.3	−0.3 ± 0.2
L213	−1.6 ± 0.4	−1.0 ± 0.3	−1.5 ± 0.3	−1.6 ± 0.4	−1.5 ± 0.4
A214	−1.0 ± 0.2	−0.8 ± 0.2	−1.0 ± 0.2	−0.9 ± 0.3	−0.9 ± 0.2
P215	−1.6 ± 0.3	−1.5 ± 0.3	−3.2 ± 0.1	−1.6 ± 0.3	−1.2 ± 0.3
L216	−0.8 ± 0.2	−0.9 ± 0.2	−1.6 ± 0.3	−0.7 ± 0.2	−1.1 ± 0.2
water	−1.6 ± 0.6			−1.7 ± 0.5	−1.3 ± 0.6

[a]Residues highlighted in gray are key positions for cholesterol absorption identified in the literature.

Although caution is required to not overemphasize the significance of the differences in the calculated binding affinities, which nevertheless benefit from the cancellation of uncertainties due to the comparison between structurally related sterols, these results point out that the suboptimal geometry of the C3 position to establish H-bonds and the loss of the interaction with the water molecule destabilizes the binding of epi-cholesterol to the Q95 binding site.

A per-residue decomposition of the binding affinity provided a consistent view of the most relevant residues for the binding of sterols to NPC1L1-NTD (Table 2). The key residues involved in the H-bond, namely, Q95 and S56, contribute around −3.4 ± 0.5 and −1.6 ± 0.9 kcal·mol^{-1}, respectively. The interaction with T128 is only relevant for epi-cholesterol in the T128 binding mode (−2.1 ± 0.1 kcal·mol^{-1}), whereas this residue has a negligible effect for the other sterols (≥−0.5 kcal·mol^{-1}). Other polar residues are common to both binding modes, such as H124, with an average value of −1.1 ± 0.9 kcal·mol^{-1}. The hydrophobic contacts between the tetracyclic core of the sterols and the binding site include L99 (−1.9 ± 0.3 kcal·mol^{-1}), L103 (−1.2 ± 0.2 kcal·mol^{-1}), and F120 (−0.6 ± 0.3 kcal·mol^{-1}), and the $\alpha 8/\beta 7$ loop residues L213 (−1.6 ± 0.4 kcal·mol^{-1}), A214 (−1.0 ± 0.2 kcal·mol^{-1}), P215 (−1.6 ± 0.3 kcal·mol^{-1}), and L216 (−0.8 ± 0.2 kcal·mol^{-1}) (see Table 2).

Mutations that Impair Cholesterol Transport Distort the $\alpha 8/\beta 7$ Loop. Previous studies have disclosed a variety of single-point mutations in the NPC1L1-NTD that have a detrimental effect on cholesterol absorption.[18,55,56] In light of the previous results regarding the binding mode of cholesterol, a discussion of the potential impact of mutations that affect specific residues in the binding pocket of NPC1L1-NTD deserves interest. For our purposes, the T128A, I105A, F205A, and P215A variants, which trigger a reduction in cholesterol absorption varying from 35 to 85% (Table 3),[18] were selected to explore their effect on cholesterol binding. Noteworthily,

Table 3. Binding Free Energies (Average and Standard Deviation (SD); kcal·mol^{-1}) Obtained Using MM-PBSA Calculations for Cholesterol Bound to the Mutated Variants of NPC1L1-NTD

		MM-PBSA with water molecule kcal mol^{-1}		
	reduction of cholesterol absorption %	mean	SD	standard error of the mean
wild-type	–	−20.8	0.9	0.5
I105A	35	−22.1	1.9	1.0
T128A	85	−18.1	2.7	1.4
F205A	85	−19.8	1.5	0.7
P215A	52	−15.7	2.2	1.1

although the detrimental effect of P215A may be justified by the loss of the contribution of P215 to the binding affinity, which varies from −1.6 to −1.2 kcal·mol^{-1} for cholesterol, 25-hydroxycholesterol, and lanosterol (see Table 2), the effect of the other three variants is not apparent, as their contribution to the binding affinity is around −0.3 kcal·mol^{-1} for these compounds. In turn, this suggests that other factors beyond the direct interaction between the ligand and the residues in the binding pocket may influence the efficiency of binding of cholesterol to NPC1L1-NTD.

To gain insight into this question, four independent MD simulations were performed for both the apo and cholesterol-bound states of the mutated (T128A, I105A, F205A, and P215A) NPC1L1-NTD and a comparison was made with the structural and energetic features determined for the wild-type protein (see above). The RMSD profile of the protein backbone of the mutated variants showed no significant alterations for both the apo and cholesterol-bound simulations (RMSD values around 2 Å; see Figures S16−S19 and S20−S23 for apo and ligand-bound species, respectively). The RMSD values of cholesterol for all of the mutants were close to 1 Å and, therefore, comparable to the cholesterol-bound

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

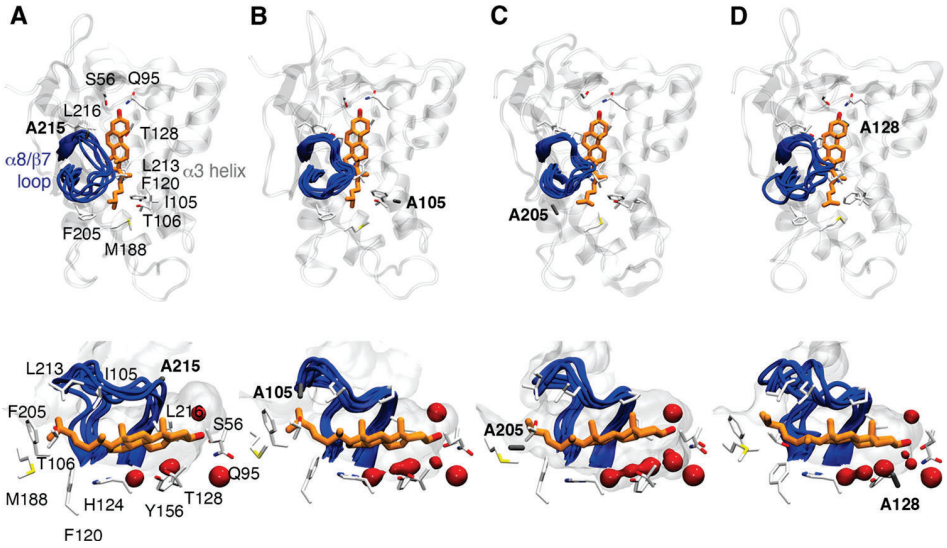


Figure 7. (Top) Q95 binding mode of cholesterol is maintained in the (A) P215A, (B) I105A, (C) F205A, and (D) T128A mutated variants of NPC1L1-NTD. (Bottom) Shape and solvation pattern of the binding site of NPC1L1-NTD for cholesterol-bound states. The conformations sampled by the $\alpha 8/\beta 7$ loop are shown in a blue cartoon. Cholesterol molecules are shown as orange sticks. Mutated residues are highlighted using gray sticks. The protein surface is shown as a gray surface. The red isosurface accounts for 50% of water occupancy in the binding site during the last 100 ns of each replicate (accumulated 400 ns of simulation time).

complex formed with the wild-type NPC1L1-NTD (Figure S4), with the only exception of a few transient peaks, always below 3 Å, which account for rearrangements of the aliphatic side chain.

Compared to the wild-type complex, there is a slight reduction in the volume of the cavity, which varies from 940 to 990 Å^3 for the mutated variants (volume of 1024 ± 19 Å^3 for the wild-type protein). Nevertheless, the Q95 binding mode is stable along the simulation time in all of the cholesterol-bound simulations (Figures 7 and S20−S23). Indeed, the three key H-bond interactions with Q95, S56, and water molecules are maintained in all replicates. We only observe two short-lived destabilizations of the Q95 binding mode in the T128A mutant, where cholesterol moves toward the S-opening with a water molecule bridging the interaction with Q95.

The effect of each mutation on the cholesterol binding to NPC1L1-NTD was examined using MM-PBSA calculations (Table 3). In agreement with the trends discussed from the data in Table 2, the P215A mutation gives rise to a significant decrease in the binding affinity of cholesterol (−15.7 ± 2.2 kcal·mol^{-1}) compared to the wild-type protein (−20.8 ± 0.9 kcal·mol^{-1}), which partially accounts for the loss of the proline side chain contribution of 1.6 ± 0.3 kcal·mol^{-1}. Nevertheless, the values obtained for I105A, T128A, and F205A (−19.2 ± 2.3, −19.3 ± 2.6, and −21.0 ± 1.9 kcal·mol^{-1}, respectively) are close to the binding affinity estimated for the wild-type complex, suggesting that other factors may affect the binding process.

Considering these results, principal components analysis (PCA) was used to examine the impact of the mutations on the conformational dynamics of the $\alpha 8/\beta 7$ loop and the $\alpha 3$ helix, which contribute to shaping the binding pocket in the apo and cholesterol-bound species (Figure 8). The two first PCs explain almost 66% of the total structural variance. The first PC (PC1) accounts for 39% of the structural variance and describes a concerted lateral motion of the $\alpha 8$-$\beta 7$ loop and the $\alpha 3$ helix, which opens the entrance of the binding site. The second PC (PC2) explains 27% of the variance and describes the concerted motion that brings the $\alpha 8$-$\beta 7$ loop into the binding pocket and closer to the $\alpha 3$ helix, which is the main driver of the narrowing of the binding pocket entrance. The third and the fourth components only explained 8 and 7% of the total variance.

The apo- and cholesterol-bound trajectories sample two distinct populations in the PC space. PC2 discriminates between apo and holo states of the NPC1L1-NTD, as apo and cholesterol-bound trajectories populate positive and negative values along the PC2 projection, respectively (Figure 8). On the other hand, the different conformations of the $\alpha 8/\beta 7$ loop sampled in the simulations spread along positive and negative values of PC1, leading to two clusters for the wild-type apo species (Figure 9; see also Table S3): in one case the $\alpha 8/\beta 7$ loop is buried in the binding site placing the L213 side chain close to F120 (A1: 24 ± 2%, pocket volume of 483 ± 116 Å^3) and in the other, the L213 backbone forms a H-bond with S102 (3.0 ± 0.7 Å) and a hydrophobic contact with I105A (A2: 75 ± 2%, pocket volume of 491 ± 145 Å^3). In the holo species, NPC1L1-NTD populates a single conformation (B1: 97 ± 2%, pocket volume of 1025 ± 61 Å^3).

The P215A variant distorts the dynamics of the $\alpha 8/\beta 7$ loop, leading to four conformations of the entrance of the binding site in the apo species (Figures 8 and 9; see also Table S4). Compared to the wild-type protein, whereas the A1 conformation is not significantly altered (A1: 24 ± 1%, pocket volume of 361 ± 76 Å^3), the A2 conformation is less populated (A2: 42 ± 1%, pocket volume of 390 ± 55 Å^3) in favor of two new conformational states. In one of them (A3: 25 ± 2%, pocket volume of 444 ± 65 Å^3), the $\alpha 8/\beta 7$ loop is stabilized by a H-bond between A214 and S102 (3.6 ± 0.9 Å), which shifts the L213 side chain from the entrance and interacts with F205. In the other (A5: 9 ± 1%, pocket volume of 288 ± 22 Å^3), the $\alpha 8/\beta 7$ loop binds deeper into the binding site, with L213 being located closer to H124. In the holo state, however, the impact of the P215A mutation is weaker, as the B1 conformation (79 ± 1%, pocket volume of 902 ± 51 Å^3) is largely populated, and a minor conformation (B2: 20 ± 1%,

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

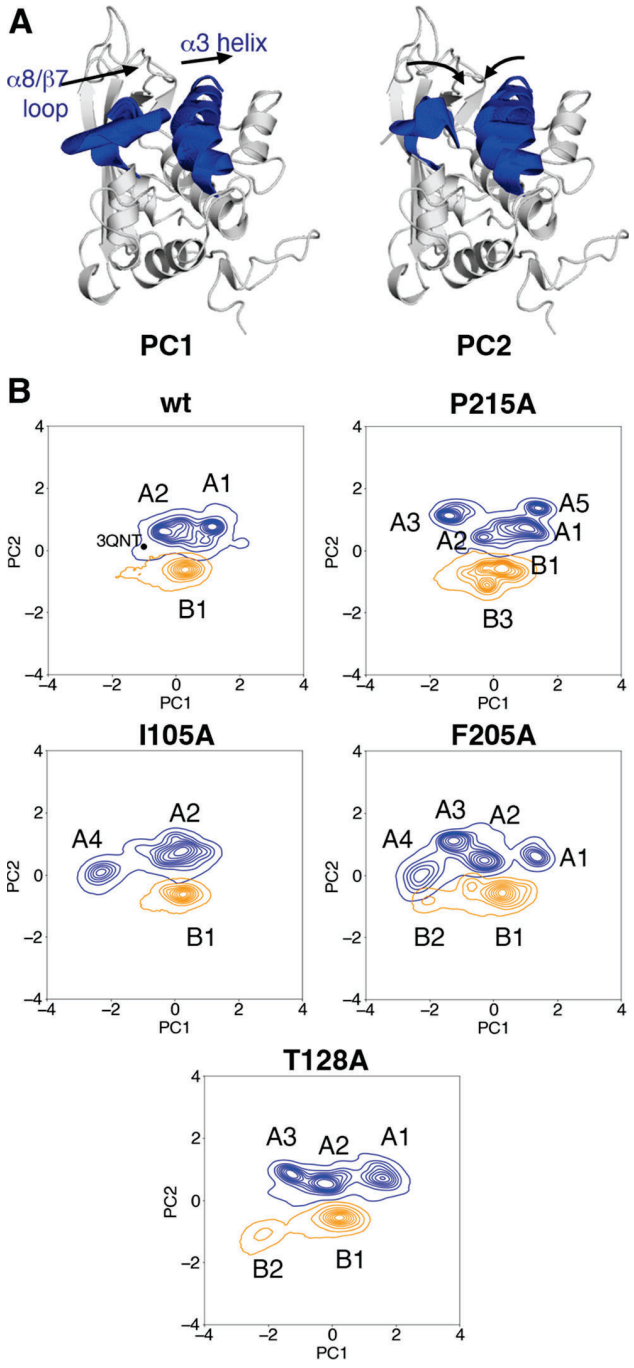


Figure 8. Principal component analysis of the wild-type NPC1L1-NTD domain and its mutated variants. (A) Structural representation of the two main principal components of the PCA. PC1 accounts for the sideways motion between the $\alpha 8/\beta 7$ loop and helix $\alpha 3$, and PC2 accounts for the motion that reduces the distance between the $\alpha 8/\beta 7$ loop and helix $\alpha 3$. (B) Projections of the wild-type protein and the mutated variants on the space defined by the first two Principal Components. The density of visited conformations of the trajectories of the apo and cholesterol-bound are shown as blue and orange-shaded contours, respectively. For the wild-type protein, the location of the X-ray structure of the apo species is also indicated.

pocket volume of 940 ± 72 Å^3), where the $\alpha 8/\beta 7$ loop adopts a slightly tighter conformation around the sterol molecule, retains the Q95 binding mode (Figure 7D). Altogether, the P215A mutation alters primarily the conformational space of the apo species, stabilizing alternative conformations with significantly reduced volume of the binding site and reducing the binding affinity of cholesterol in the holo species (see Table 3). Altogether, these results provide a structural basis to explain the experimental reduction (48%) of cholesterol absorption.[18,56]

Mutations on residues I105 and F205, which stabilize the hydrophobic cluster found in the closed state of the apo state, are likely to alter the conformational preferences of the $\alpha 8/\beta 7$ loop. With regard to the I105A variant, although the conformation A2 is the most populated (76 ± 5%) and resembles the cluster found in the wild-type apo species, there is a marked reduction in the pocket volume (264 ± 43 Å^3). Furthermore, the A1 conformation disappears, and a minor conformational state (A4: 18 ± 3%, pocket volume of 525 ± 44 Å^3) emerges (Figures 8 and 9; see also Table S4). This reflects the weaker stability of the hydrophobic cluster resulting from the Ile → Ala mutation, which favors the opening of the binding site and the enlargement of the cavity. In contrast, the I105A mutation does not significantly alter the conformation dynamics of the cholesterol-bound state (B1: 93 ± 1%, pocket volume of 985 ± 52 Å^3). These results are in qualitative agreement with the computed binding affinity of cholesterol for this mutated variant, which compares with the value obtained for the wild type (see Table 3), the mild effect on the cholesterol absorption, as the I105A mutation leads to the lowest reduction (35%) among the four selected variants.[18,56]

Compared to I105A, the F205A variant gives rise to a larger reduction of cholesterol absorption (~85%).[18,56] This trait agrees with the larger remodeling of the conformational preferences of the $\alpha 8/\beta 7$ loop observed for this mutation (Figures 8 and 9; see also Table S4). Thus, conformations A2 (35 ± 1.0%, pocket volume of 475 ± 126 Å^3) and A1 (12 ± 1%, pocket volume of 488 ± 123 Å^3) are less populated relative to the wild-type apoprotein. Furthermore, the $\alpha 8/\beta 7$ loop populates two novel states: A3 (31 ± 1%, pocket volume of 366 ± 43 Å^3) and A4 (23 ± 1%, pocket volume of 745 ± 95 Å^3). Since the side chain of F205 is a key component of the hydrophobic cluster that stabilizes the closed state, the mutation to Ala weakens the stability of the cluster, favoring the adoption of other conformations. In contrast, the F205A effect is much less apparent in the cholesterol-bound state, as the B1 conformation is the main cluster (90 ± 1%, pocket volume of 994 ± 45 Å^3). Thus, although F205 has a residual contribution to the cholesterol-binding affinity, it is key to stabilizing the conformation of the $\alpha 8/\beta 7$ loop in the apo state.

The T128A mutant, which also has a drastic impact on cholesterol absorption (reduction of 85%), is more challenging since T128 does not directly interact with the $\alpha 8/\beta 7$ loop. Nevertheless, our results point out that this mutation changes the dynamics of the S-opening, especially in the species. Thus, the main conformation A2 (52 ± 4%, pocket volume of 434 ± 37 Å^3) is partially shifted to a novel state (A3: 17 ± 3%, pocket volume of 347 ± 86 Å^3) due to a remodeling of the hydration pattern in the binding site (Figure 10). Compared to the wild-type species, T128A promotes the loss of water located between the backbone of L213 and the side chain of S102, which bridges the interaction between the L213 backbone and the H124 side chain. Of note, in the A1 conformation, the H124 side chain adopts a distinct arrangement that enables this interaction without the water molecule. Upon ligand binding, the S-entrance mainly populates the B1 conformation (86 ± 1%, pocket volume of 935 ± 54 Å^3), which supports the

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

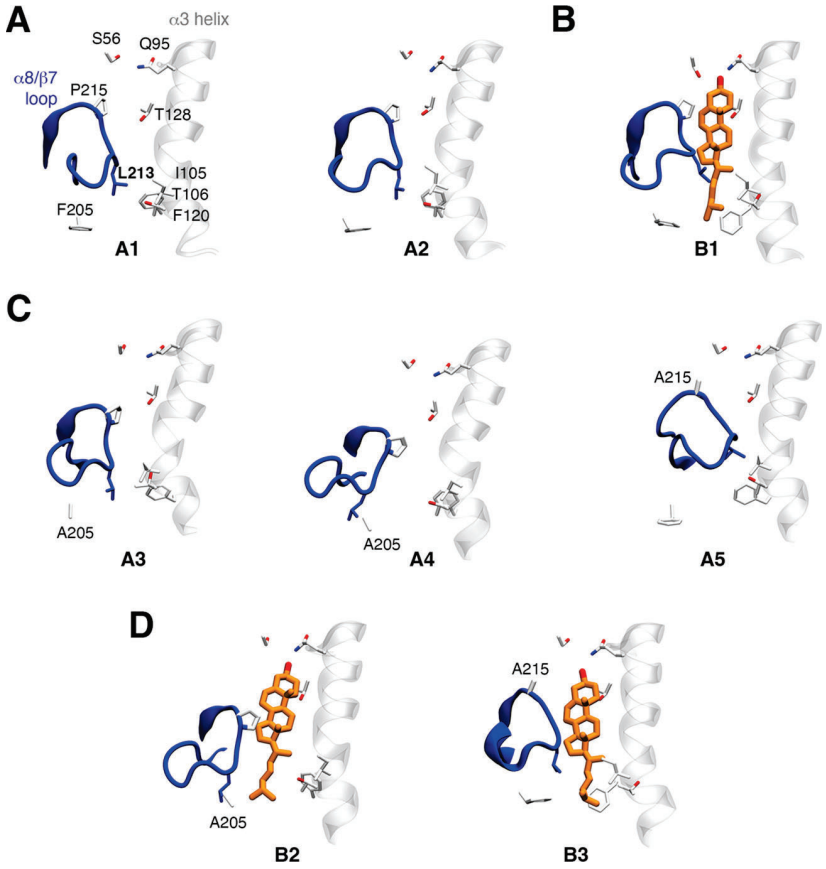


Figure 9. Representative structures of the main conformations of the $\alpha 8/\beta 7$ loop and $\alpha 3$ helix in (A, B) the wild-type and (C, D) the mutated variants of NPC1L1-NTD. A representative snapshot of the $\alpha 8/\beta 7$ loop and $\alpha 3$ helix sampled in (A, C) the apo and (B, D) cholesterol-bound trajectories is shown as blue and white cartoons, respectively. Cholesterol is shown as orange sticks. Representative structures of the alternative conformations A3, A4, and B2 are extracted from the F205A mutant. Representative structures of the conformations A5 and B3 are only present in the P215A variant.

Q95 binding mode, although a minor conformation (B3: 8 ± 1%, pocket volume of 968 ± 54 Å^3), which alters the Q95 binding mode by shifting cholesterol toward the S-opening due to a water molecule that bridges the interaction with Q95, is also observed (Figure S24).

Altogether, the analysis of the trajectories sampled for the wild-type NPC1L1-NTD and its mutated variants suggest that the mutations have a minor impact on the conformational preferences of the $\alpha 8/\beta 7$ loop in the cholesterol-bound species, as the main conformation (B1) observed in the wild-type protein is preserved in the mutated proteins (Figure 11). Indeed, the mutations do not substantially alter the binding mode of cholesterol in the Q95 binding site, and in most of the cases, they do not even affect its binding affinity, with the only exception of P215A (Table 3). However, the present results point out that the single-point mutations in key residues with a minor contribution to the binding affinity of cholesterol may have a large impact on the efficiency of cholesterol absorption by altering the conformational landscape of the S-opening in the apo species of NPC1L1-NTD (Figure 11). These findings agree with previous studies that have highlighted the relevant role played by mutations introduced in enzymes to remodel the dynamic equilibrium between the ensemble of differently populated conformational substates, thus strongly modifying not only the catalytic efficiency but also the ability to accommodate alternative substrates.[57−59] In this context, our results strengthen the role played by the $\alpha 8/\beta 7$ loop, particularly L213 and F205, in modulating the access of cholesterol to the binding pocket of NPC1L1-NTD.[15,17]

CONCLUSIONS

The molecular basis of the epimerization-related selectivity in the binding and transport of cholesterol by the N-terminal domain of the NPC1L1 transporter has been explored by combining molecular dynamics simulations of the apo and holo species, in conjunction with free energy calculations for the wild-type protein and four mutated variants known to cause a detrimental effect on the cholesterol import.

The results have shown the drastic influence of the epimerization of the hydroxyl group on the binding mode of cholesterol and epi-cholesterol. The analysis of the independent MD simulations run for cholesterol consistently supports the stability of the Q95 binding mode, where the hydroxyl moiety establishes H-bond interactions with the side chain of Q95, S56, and a neighboring water molecule. The main features of this binding mode are also stably reproduced for 25-hydroxycholesterol and lanosterol, although additional interactions can be found due to the chemical modifications on the aliphatic tail of 25-hydroxycholesterol and the methyl groups of lanosterol. In contrast, the α-configuration of the C3 atom in epi-cholesterol leads to the T128 binding mode, which is shifted ca. 3 Å to the mouth of the cavity. Indeed, US calculations and well-tempered metadynamics simulations

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

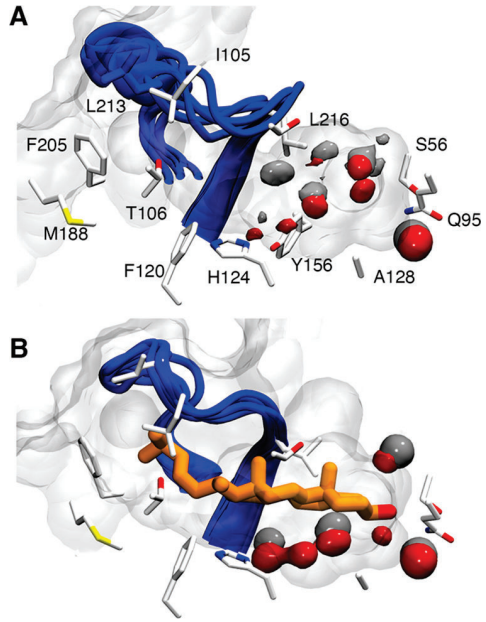


Figure 10. Superposition of the major conformations of (A) the apo and (B) cholesterol-bound species of the T128A variant. The conformations sampled by the $\alpha 8/\beta 7$ loop are shown as a blue cartoon. The protein surface is shown as a gray surface. The isosurface accounts for 50% of water occupancy in the binding site during the last 100 ns of each replicate (accumulated 400 ns of simulation time). The water occupancy for the wild-type protein and T128A mutant are shown as gray and red isocontours, respectively.

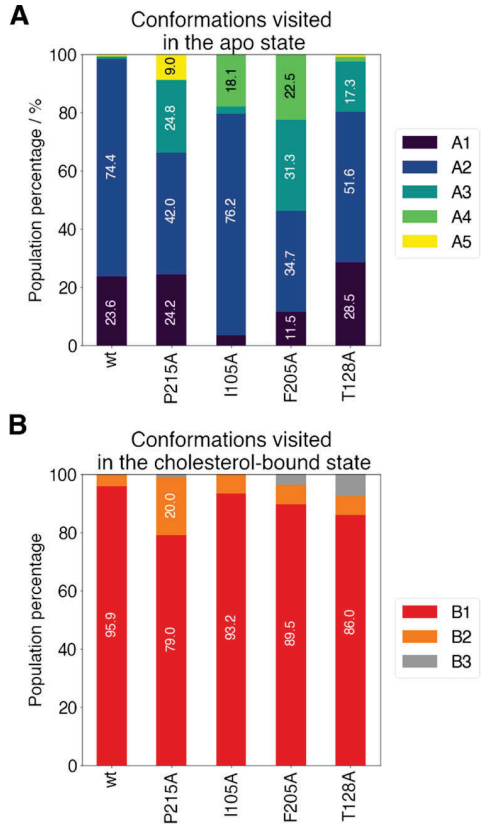


Figure 11. Relative population of each conformation is shown as a stacked bar plot for (A) apo and (B) cholesterol-bound simulations.

reveal the existence of an energetic barrier (i.e., 4.1 kcal·mol^{-1}) for the transition from this binding mode to the Q95 one, thus preventing epi-cholesterol from penetrating deeper in the binding pocket. Altogether, this reflects that the α-configuration of the C3 position is unable to properly bind to the Q95 binding mode and can move to a metastable state of the entrance pathway of sterols.

The binding site of cholesterol is seemingly challenged by the detrimental effect of several single-point mutations in the binding site (T128A, I105A, F205A, and P215A) that impair cholesterol absorption.[18] Our results showed that these mutations have a mild effect on the conformational space of the $\alpha 8/\beta 7$ loop in the cholesterol-bound state and that only P215A contributes to weaken the binding affinity of cholesterol. In contrast, the results also revealed the impact of these mutations in distorting the conformational landscape of the $\alpha 8/\beta 7$ loop in the apo state, which is key for regulating the access of cholesterol to the binding pocket through the formation of a hydrophobic cluster formed by L213, L103, T106, F120, and F205. The reshaping of the conformational preferences of the $\alpha 8/\beta 7$ loop should destabilize its role in mediating the access of cholesterol to the binding pocket, thus strengthening the hypothesis of the gating mechanism played by the $\alpha 8/\beta 7$ loop.[15,17]

■ ASSOCIATED CONTENT

Data Availability Statement

Gaussian09 (Revision B.01) is a licensed software for electronic structure calculations used in the parametrization of sterol molecules. The Gromacs 2022.4 package is a free, open-source software suite for high-performance molecular dynamics and output analysis. RBFE (Thermodynamic Integration) calculations were performed with the GPU-accelerated implementation of Amber22, and MM-PBSA computations were determined with the MMPBSA.py script available in AmberTools22. Clustering of the ensemble of conformational states was performed using the Robust Perron Cluster Analysis method (PCCA+) as implemented in the PyEMMA 2.0 module.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jcim.3c01319.

Detailed analysis of the independent trajectories run for apo and ligand-bound species of wild-type NPC1L1-NTD and its mutated variants; files with topologies and parameters of cholesterol, epi-cholesterol, 25-hydroxycholesterol, and lanosterol, initial configurations of simulated systems (GRO files) and trajectories are available at Zenodo (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Salomé Llabrés − *Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació—Campus Torribera, Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain; Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, 08921 Barcelona, Spain;* orcid.org/0000-0002-2039-7821; Email: salome.llabres@ub.edu

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

Authors

Aitor Valdivia − *Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació—Campus Torribera, Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain; Institut de Biomedicina (IBUB), Universitat de Barcelona, 08028 Barcelona, Spain;* orcid.org/0000-0003-4862-0850

F. Javier Luque − *Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació—Campus Torribera, Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain; Institut de Biomedicina (IBUB), Universitat de Barcelona, 08028 Barcelona, Spain; Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, 08921 Barcelona, Spain;* orcid.org/0000-0002-8049-3567

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jcim.3c01319

Author Contributions

A.V.: Investigation and formal analysis. F.J.L.: Conceptualization, resources, funding acquisition, supervision, and writing—review and editing. S.L.: Investigation, conceptualization, investigation, formal analysis, funding acquisition, supervision, and writing—Original Draft.

Funding

PID2020-117646RB-I00 MCIN/AEI/10.13039/501100011033, Maria de Maetzu CEX2021-001202-M, 2021SGR00671, IJC2018-037291-I

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The authors thank the Spanish Ministerio de Ciencia e Innovación (PID2020-117646RB-I00 MCIN/AEI/10.13039/501100011033; Maria de Maetzu CEX2021-001202-M) and Generalitat de Catalunya (2021SGR00671) for financial support. S.L. is supported by a Juan de la Cierva-Incorporación 2018 fellowship (IJC2018-037291-I) financed by the Spanish Ministerio de Ciencia e Innovación. The authors acknowledge access to the Memento and Lusitania HPC facilities granted through the Red Española de Supercomputación (BCV-2021-3-0005 and BCV-2022-3-0011) and thank M. Papadourakis for helpful discussions on RFBE methodologies.

■ ABBREVIATIONS

NPC1L1, Niemann−Pick type C1 like 1 transporter; NPC1, Niemann−Pick type C1 transporter; NTD, N-terminal domain; RBFE, relative binding free energies

■ REFERENCES

(1) Luo, J.; Yang, H.; Song, B.-L. Mechanisms and Regulation of Cholesterol Homeostasis. *Nat. Rev. Mol. Cell Biol.* **2020**, *21* (4), 225−245.

(2) Silva Afonso, M.; Marcondes Machado, R.; Ferrari Lavrador, M. S.; Rocha Quintao, E. C.; Moore, K. J.; Lottenberg, A. M. Molecular Pathways Underlying Cholesterol Homeostasis. *Nutrients* **2018**, *10*, No. 760, DOI: 10.3390/nu10060760.

(3) Kuzu, O. F.; Noory, M. A.; Robertson, G. P. The Role of Cholesterol in Cancer. *Cancer Res.* **2016**, *76* (8), 2063−2070.

(4) Virani, S. S.; Alonso, A.; Aparicio, H. J.; Benjamin, E. J.; Bittencourt, M. S.; Callaway, C. W.; Carson, A. P.; Chamberlain, A. M.; Cheng, S.; Delling, F. N.; Elkind, M. S. V.; Evenson, K. R.; Ferguson, J. F.; Gupta, D. K.; Khan, S. S.; Kissela, B. M.; Knutson, K. L.; Lee, C. D.; Lewis, T. T.; Liu, J.; Loop, M. S.; Lutsey, P. L.; Ma, J.; Mackey, J.; Martin, S. S.; Matchar, D. B.; Mussolino, M. E.; Navaneethan, S. D.; Perak, A. M.; Roth, G. A.; Samad, Z.; Satou, G. M.; Schroeder, E. B.; Shah, S. H.; Shay, C. M.; Stokes, A.; VanWagner, L. B.; Wang, N.-Y.; Tsao, C. W. Heart Disease and Stroke Statistics 2021 Update. *Circulation* **2021**, *143* (8), e254−e743.

(5) Betters, J. L.; Yu, L. NPC1L1 and Cholesterol Transport. *FEBS Lett.* **2010**, *584* (13), 2740−2747.

(6) Altmann, S. W.; Davis, H. R.; Zhu, L.; Yao, X.; Hoos, L. M.; Tetzloff, G.; Iyer, S. P. N.; Maguire, M.; Golovko, A.; Zeng, M.; Wang, L.; Murgolo, N.; Graziano, M. P. Niemann-Pick C1 Like 1 Protein Is Critical for Intestinal Cholesterol Absorption. *Science* **2004**, *303* (5661), 1201−1204.

(7) Huang, C.-S.; Yu, X.; Fordstrom, P.; Choi, K.; Chung, B. C.; Roh, S.-H.; Chiu, W.; Zhou, M.; Min, X.; Wang, Z. Cryo-EM Structures of NPC1L1 Reveal Mechanisms of Cholesterol Transport and Ezetimibe Inhibition. *Sci. Adv.* **2020**, *6* (25), No. eabb1989.

(8) Saha, P.; Shumate, J. L.; Caldwell, J. G.; Elghobashi-Meinhardt, N.; Lu, A.; Zhang, L.; Olsson, N. E.; Elias, J. E.; Pfeffer, S. R. Inter-Domain Dynamics Drive Cholesterol Transport by Npc1 and Npc1l1 Proteins. *eLife* **2020**, *9*, 1−28.

(9) Muendlein, A.; Leiherer, A.; Saely, C. H.; Rein, P.; Zanolin, D.; Kinz, E.; Brandtner, E.-M.; Fraunberger, P.; Drexel, H. Common Single Nucleotide Polymorphisms at the NPC1L1 Gene Locus Significantly Predict Cardiovascular Risk in Coronary Patients. *Atherosclerosis* **2015**, *242* (1), 340−345.

(10) Qu, Y.-X.; Ge, L.; Wang, J.; Qi, W.; Song, B.-L.; Cao, J.; Miao, H.-H.; Li, B.-L. The Cholesterol Absorption Inhibitor Ezetimibe Acts by Blocking the Sterol-Induced Internalization of NPC1L1. *Cell Metab.* **2008**, *7* (6), 508−519.

(11) Davidson, M. H.; McGarry, T.; Bettis, R.; Melani, L.; Lipka, L. J.; LeBeaut, A. P.; Suresh, R.; Sun, S.; Veltri, E. P. Ezetimibe Coadministered with Simvastatin in Patients with Primary Hypercholesterolemia. *J. Am. Coll. Cardiol.* **2002**, *40* (12), 2125−2134.

(12) Tsubakio-Yamamoto, K.; Nishida, M.; Nakagawa-Toyama, Y.; Masuda, D.; Ohama, T.; Yamashita, S. Current Therapy for Patients with Sitosterolemia-Effect of Ezetimibe on Plant Sterol Metabolism. *J. Atheroscler. Thromb.* **2010**, *17* (9), 891−900.

(13) Zhang, Z.; Qin, S.; Chen, Y.; Zhou, L.; Yang, M.; Tang, Y.; Zuo, J.; Zhang, J.; Mizokami, A.; Nice, E. C.; Chen, H.-N.; Huang, C.; Wei, X. Inhibition of NPC1L1 Disrupts Adaptive Responses of Drug-Tolerant Persister Cells to Chemotherapy. *EMBO Mol. Med.* **2022**, *14* (2), No. e14903.

(14) Zhang, R.; Zeng, J.; Liu, W.; Meng, J.; Wang, C.; Shi, L.; Yang, S.; Chang, J.; Xing, D. The Role of NPC1L1 in Cancer. *Front. Pharmacol.* **2022**, *13*, No. 956619, DOI: 10.3389/fphar.2022.956619.

(15) Kwon, H. J.; Palnitkar, M.; Deisenhofer, J. The Structure of the NPC1L1 N-Terminal Domain in a Closed Conformation. *PLoS One* **2011**, *6* (4), No. e18722, DOI: 10.1371/journal.pone.0018722.

(16) Poongavanam, V.; Kongsted, J.; Wüstner, D. Computational Modeling Explains the Multi Sterol Ligand Specificity of the N-Terminal Domain of Niemann−Pick C1-Like 1 Protein. *ACS Omega* **2019**, *4* (25), 20894−20904.

(17) Kwon, H. J.; Abi-Mosleh, L.; Wang, M. L.; Deisenhofer, J.; Goldstein, J. L.; Brown, M. S.; Infante, R. E. Structure of N-Terminal Domain of NPC1 Reveals Distinct Subdomains for Binding and Transfer of Cholesterol. *Cell* **2009**, *137* (7), 1213−1224.

(18) Zhang, J.-H.; Ge, L.; Qi, W.; Zhang, L.; Miao, H.-H.; Li, B.-L.; Yang, M.; Song, B.-L. The N-Terminal Domain of NPC1L1 Protein Binds Cholesterol and Plays Essential Roles in Cholesterol Uptake. *J. Biol. Chem.* **2011**, *286* (28), 25088−25097.

(19) Machado, M. R.; Pantano, S. Split the Charge Difference in Two! A Rule of Thumb for Adding Proper Amounts of Ions in MD Simulations. *J. Chem. Theory Comput.* **2020**, *16* (3), 1367−1372.

(20) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from Ff99SB. *J. Chem. Theory Comput.* **2015**, *11* (8), 3696−3713.

(21) Dickson, C. J.; Madej, B. D.; Skjevik, ÅA.; Betz, R. M.; Teigen, K.; Gould, I. R.; Walker, R. C. Lipid14: The Amber Lipid Force Field. *J. Chem. Theory Comput.* **2014**, *10*, 865−879.
(22) Joung, I. S.; Cheatham, T. E. Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations. *J. Phys. Chem. B* **2008**, *112* (30), 9020−9041.
(23) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954−9960.
(24) Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21* (12), 1049−1074.
(25) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi, R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J.*Gaussian 09*, revision B.01; Gaussian, Inc., Wallingford, CT, 2009.
(26) Sousa da Silva, A. W.; Vranken, W. F. ACPYPE—AnteChamber PYthon Parser InterfacE. *BMC Res. Notes* **2012**, *5* (1), No. 367.
(27) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126* (1), No. 014101, DOI: 10.1063/1.2408420.
(28) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: A Linear Constraint Solver for Molecular Simulations. *J. Comput. Chem.* **1997**, *18* (12), 1463−1472.
(29) Abraham, M. J.; Murtola, T.; Schulz, R.; Páall, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1−2*, 19−25.
(30) Miller, B. R.; McGee, T. D., Jr.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.Py: An Efficient Program for End-State Free Energy Calculations. *J. Chem. Theory Comput.* **2012**, *8* (9), 3314−3321.
(31) Kuhn, B.; Gerber, P.; Schulz-Gasch, T.; Stahl, M. Validation and Use of the MM-PBSA Approach for Drug Discovery. *J. Med. Chem.* **2005**, *48* (12), 4040−4048.
(32) Wang, E.; Weng, G.; Sun, H.; Du, H.; Zhu, F.; Chen, F.; Wang, Z.; Hou, T. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 10. Impacts of Enhanced Sampling and Variable Dielectric Model on Protein−Protein Interactions. *Phys. Chem. Chem. Phys.* **2019**, *21* (35), 18958−18969.
(33) Schmidtke, P.; Bidon-Chanal, A.; Luque, F. J.; Barril, X. MDpocket: Open Source Cavity Detection and Characterization on Molecular Dynamics Trajectories. *Bioinformatics* **2011**, *27*, No. btr550.
(34) Song, L. F.; Lee, T.-S.; Zhu, C.; York, D. M.; Merz, K. M., Jr. Using AMBER18 for Relative Free Energy Calculations. *J. Chem. Inf. Model.* **2019**, *59* (7), 3128−3135.
(35) Mey, A. S. J. S.; Allen, B. K.; Bruce McDonald, H. E.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best Practices for Alchemical Free Energy Calculations [Article v1.0. *Living J. Comput. Mol. Sci.* **2020**, *2* (1), 18378.
(36) Tsai, H.-C.; Tao, Y.; Lee, T.-S.; Merz, K. M., Jr.; York, D. M. Validation of Free Energy Methods in AMBER. *J. Chem. Inf. Model.* **2020**, *60* (11), 5296−5300.
(37) Chodera, J. D. A Simple Method for Automated Equilibration Detection in Molecular Simulations. *J. Chem. Theory Comput.* **2016**, *12* (4), 1799−1805.
(38) Hu, Y.; Muegge, I. In Silico Positional Analogue Scanning with Amber GPU-TI. *J. Chem. Inf. Model.* **2022**, *62* (18), 4448−4459.
(39) Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular Dynamics Study of Unbinding of the Avidin-Biotin Complex. *Biophys. J.* **1997**, *72* (4), 1568−1581.
(40) Grossfield, A. WHAM: The Weighted Histogram Analysis Method, 2004. http://membrane.urmc.rochester.edu/content/wham.
(41) Dama, J. F.; Parrinello, M.; Voth, G. A. Well-Tempered Metadynamics Converges Asymptotically. *Phys. Rev. Lett.* **2014**, *112* (24), No. 240602.
(42) Barducci, A.; Bussi, G.; Parrinello, M. Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Phys. Rev. Lett.* **2008**, *100* (2), No. 020603, DOI: 10.1103/PhysRevLett.100.020603.
(43) Bussi, G.; Laio, A. Using Metadynamics to Explore Complex Free-Energy Landscapes. *Nat. Rev. Phys.* **2020**, *2* (4), 200−212.
(44) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. Efficient Reconstruction of Complex Free Energy Landscapes by Multiple Walkers Metadynamics. *J. Phys. Chem. B* **2006**, *110* (8), 3533−3539.
(45) Ren, W.; Vanden-Eijnden, E. Simplified and Improved String Method for Computing the Minimum Energy Paths in Barrier-Crossing Events. *J. Chem. Phys.* **2007**, *126* (16), No. 164103.
(46) Prajapati, J. D.; Fernández Solano, C. J.; Winterhalter, M.; Kleinekathöfer, U. Characterization of Ciprofloxacin Permeation Pathways across the Porin OmpC Using Metadynamics and a String Method. *J. Chem. Theory Comput.* **2017**, *13* (9), 4553−4566.
(47) Röblitz, S.; Weber, M. Fuzzy Spectral Clustering by PCCA+: Application to Markov State Models and Data Classification. *Adv. Data Anal. Classif.* **2013**, *7* (2), 147−179.
(48) Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* **2015**, *11* (11), 5525−5542.
(49) Bukiya, A. N.; Dopico, A. M. Common Structural Features of Cholesterol Binding Sites in Crystallized Soluble Proteins. *J. Lipid Res.* **2017**, *58* (6), 1044−1054.
(50) Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57* (12), 2911−2937.
(51) Kuhn, M.; Firth-Clark, S.; Tosco, P.; Mey, A. S. J. S.; Mackey, M.; Michel, J. Assessment of Binding Affinity via Alchemical Free-Energy Calculations. *J. Chem. Inf. Model.* **2020**, *60* (6), 3120−3130.
(52) Wahl, J.; Smieško, M. Assessing the Predictive Power of Relative Binding Free Energy Calculations for Test Cases Involving Displacement of Binding Site Water Molecules. *J. Chem. Inf. Model.* **2019**, *59* (2), 754−765.
(53) Mey, A. S. J. S.; Juárez-Jiménez, J.; Hennessy, A.; Michel, J. Blinded Predictions of Binding Modes and Energies of HSP90-α Ligands for the 2015 D3R Grand Challenge. *Bioorg. Med. Chem.* **2016**, *24* (20), 4890−4899.
(54) Estarellas, C.; Scaffidi, S.; Saladino, G.; Spyrakis, F.; Franzoni, L.; Galdeano, C.; Bidon-Chanal, A.; Gervasio, F. L.; Luque, F. J. Modulating Ligand Dissociation through Methyl Isomerism in Accessory Sites: Binding of Retinol to Cellular Carriers. *J. Phys. Chem. Lett.* **2019**, *10* (23), 7333−7339.
(55) Cohen, J. C.; Pertsemlidis, A.; Fahmi, S.; Esmail, S.; Vega, G. L.; Grundy, S. M.; Hobbs, H. H. Multiple Rare Variants in NPC1L1 Associated with Reduced Sterol Absorption and Plasma Low-Density Lipoprotein Levels. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103* (6), 1810−1815.
(56) Wang, L.-J.; Wang, J.; Li, N.; Ge, L.; Li, B.-L.; Song, B.-L.; et al. Molecular Characterization of the NPC1L1 Variants Identified from Cholesterol Low Absorbers. *J. Biol. Chem.* **2011**, *286* (9), 7397−7408.

(57) Corbella, M.; Pinto, G. P.; Kamerlin, S. C. L. Loop Dynamics and the Evolution of Enzyme Activity. *Nat. Rev. Chem.* **2023**, *7* (8), 536−547.
(58) Juárez-Jiménez, J.; Gupta, A. A.; Karunanithy, G.; Mey, A. S. J. S.; Georgiou, C.; Ioannidis, H.; De Simone, A.; Barlow, P. N.; Hulme, A. N.; Walkinshaw, M. D.; Baldwin, A. J.; Michel, J. Dynamic Design: Manipulation of Millisecond Timescale Motions on the Energy Landscape of Cyclophilin A. *Chem. Sci.* **2020**, *11* (10), 2670−2680.
(59) Maria-Solano, M. A.; Serrano-Hervás, E.; Romero-Rivera, A.; Iglesias-Fernández, J.; Osuna, S. Role of Conformational Dynamics in the Evolution of Novel Enzyme Function. *Chem. Commun.* **2018**, *54* (50), 6622−6634.

NOTE ADDED AFTER ASAP PUBLICATION

Corrections were made to the second column of Table 3 on January 8, 2024.

Recommended by ACS

NPC1L1 Plays a Novel Role in Nonalcoholic Fatty Liver Disease
ChongLi Xu, ChongBo Xu, *et al.*
DECEMBER 13, 2023
ACS OMEGA
READ

Estimating the Cholesterol Affinity of Integral Membrane Proteins from Experimental Data
Theodore L. Steck, Yvonne Lange, *et al.*
DECEMBER 15, 2023
BIOCHEMISTRY
READ

Cation Chloride Cotransporter NKCC1 Operates through a Rocking-Bundle Mechanism
Manuel José Ruiz Munevar, Marco De Vivo, *et al.*
DECEMBER 25, 2023
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
READ

Interplay of Cholesterol and Actin in Neurotransmitter GPCR Signaling: Insights from Chronic Cholesterol Depletion Using Statin
Parijat Sarkar and Amitabha Chattopadhyay
OCTOBER 07, 2023
ACS CHEMICAL NEUROSCIENCE
READ

Get More Suggestions >

https://doi.org/10.1021/acs.jcim.3c01319
J. Chem. Inf. Model. 2024, 64, 189−204

SUPPORTING INFORMATION

Binding of Cholesterol to the N-terminal Domain of the NPC1L1 Transporter: Analysis of the Epimerisation-Related Binding Selectivity and Loop Mutations

Aitor Valdivia [1,3], F Javier Luque [1,2,3], Salomé Llabrés [1,3]*

1 Department of Nutrition, Food Sciences and Gastronomy, Campus Torribera, University of Barcelona, Prat de la Riba 171, 08921 Santa Coloma de Gramenet, Spain.

2 Institute of Biomedicine (IBUB), University of Barcelona, 08028 Barcelona, Spain.

3 Institute of Theoretical and Computational Chemistry (IQTC-UB), University of Barcelona, 08028 Barcelona, Spain.

*** e-mail: salome.llabres@ub.edu**

TABLE OF CONTENTS

Figure S1. Structural comparison of the NTD domains of NPC1L1 and NPC1. S4

Figure S2. Superposition of the X-ray structures of the NTD domains of NPC1L1 and NPC1.S5

Figure S3. RMSD profiles of apo NPC1L1-NTD. S6

Figure S4. RMSD profiles of cholesterol-bound NPC1L1-NTD and H-bond interactions of cholesterol over simulation time. S7

Figure S5. Selected interactions of L213 in the apo NPC1L1-NTD over simulation time. S8

Figure S6. Number of water molecules in the binding site of the apo state of NPC1L1-NTD over simulation time. S9

Figure S7. Overlap between the water network of NPC1L1-NTD with the X-ray structures of the NPC1-NTD. S10

Figure S8. Selected interactions of L213 in the cholesterol-bound NPC1L1-NTD over simulation time. S11

Figure S9. Number of water molecules in the binding site of the cholesterol-bound species of NPC1L1-NTD over simulation time. S12

Figure S10. RMSD profiles of epi-cholesterol bound NPC1L1-NTD and H-bond interactions of epi-cholesterol over simulation time. S13

Figure S11. Number of water molecules in the binding site of the epi-cholesterol-bound species of NPC1L1-NTD over simulation time. S14

Figure S12. Representation of the two collective variables along the well-tempered metadynamics over the Collective Variable space. S15

Figure S13. RMSD profiles of 25-hydroxy-cholesterol-bound NPC1L1-NTD and H-bond interactions of 25-hydroxy-cholesterol over simulation time. S15

Figure S14. RMSD profiles of lanosterol-bound NPC1L1-NTD and H-bond interactions of lanosterol over simulation time. S16

Figure S15. Number of water molecules in the binding site of (*left*) 25-hydroxy-cholesterol- and (*right*) lanosterol-bound species of NPC1L1-NTD over simulation time S17

Figure S16. RMSD profiles of apo P215A NPC1L1-NTD over simulation time. S18

Figure S17. RMSD profiles of apo I105A NPC1L1-NTD over simulation time. S19

Figure S18. RMSD profiles of apo F205A NPC1L1-NTD over simulation time. S20

Figure S19. RMSD profiles of apo T128A NPC1L1-NTD over simulation time. S21

Figure S20. RMSD profiles of cholesterol-bound P215A NPC1L1-NTD and H-bond interactions of cholesterol over simulation time. S22

Figure S21. RMSD profiles of cholesterol-bound I105A NPC1L1-NTD and H-bond interactions of cholesterol over simulation time. S23

Figure S22. RMSD profiles of cholesterol-bound F205A NPC1L1-NTD and H-bond interactions of cholesterol over simulation time. S24

Figure S23. RMSD profiles of cholesterol-bound T128A NPC1L1-NTD and H-bond interactions of cholesterol over simulation time. S25

Figure S24. Superposition of the major conformations of (**A,C**) the apo and (**B,D**) cholesterol-bound states of the T128A mutant. S26

Table S1. Available structural information of the NPC1L1 protein on PDB database. S27

Table S2. Available structural information of the NPC1 protein on PDB database. S28

Table S3. Average distance (A) and preservation (%) of the interactions between wild-type NPC1L1-NTD and the studied sterol molecules. S29

Table S4. Description of the main conformational states of the apo and cholesterol-bound systems found for the wildtype NPC1L1-NTD and its mutated variants. S30

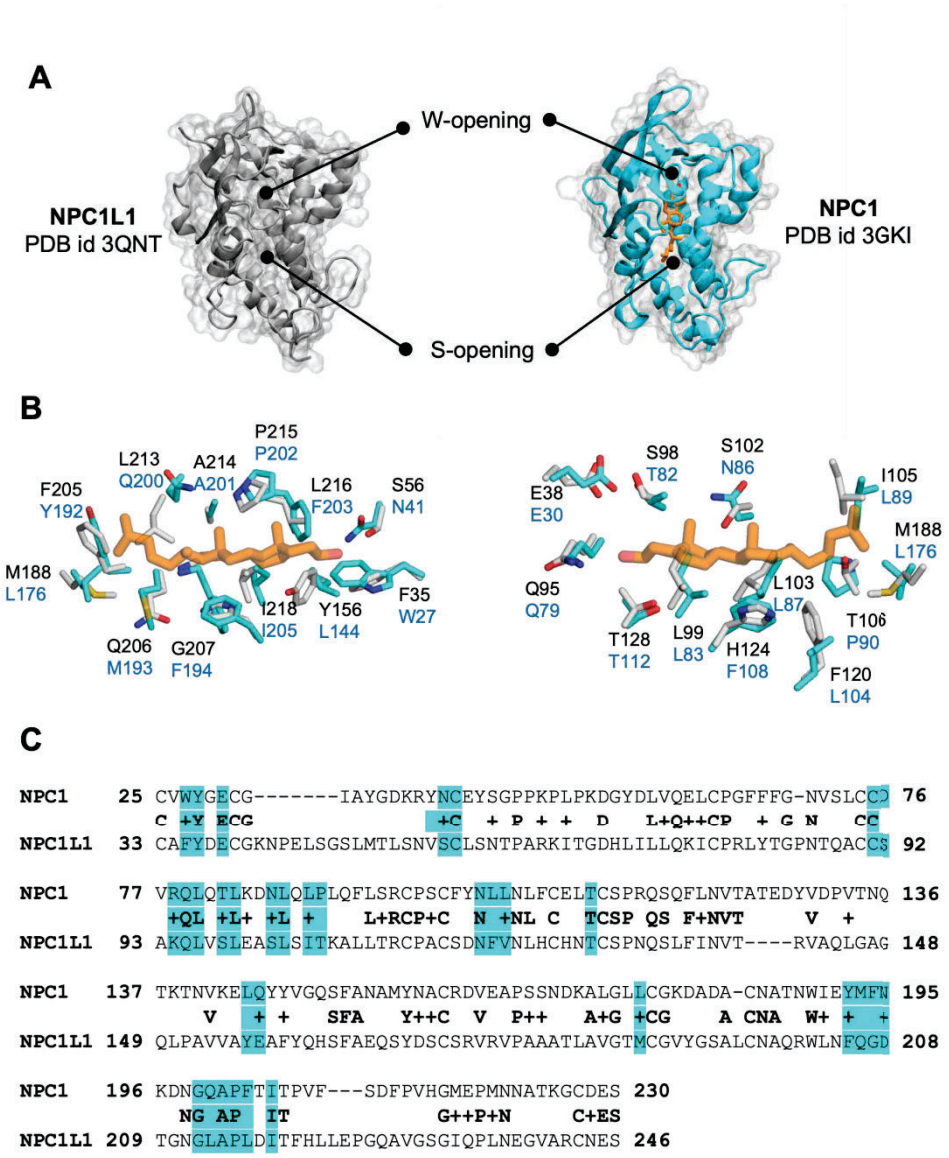


Figure S1. Structural comparison of the NTD domains of NPC1L1 and NPC1. (**A**) The global fold of both NTD domains of NPC1L1 and NPC1 are shown in grey and cyan cartoon respectively. (**B**) Comparison of the residues involved in the cholesterol binding site for both proteins. The crystallographic position of cholesterol in NPC1 is shown in orange sticks. (**C**) Sequence alignment between the NTDs of NPC1L1 and NPC1. Residues in the binding site of cholesterol are highlighted in cyan.

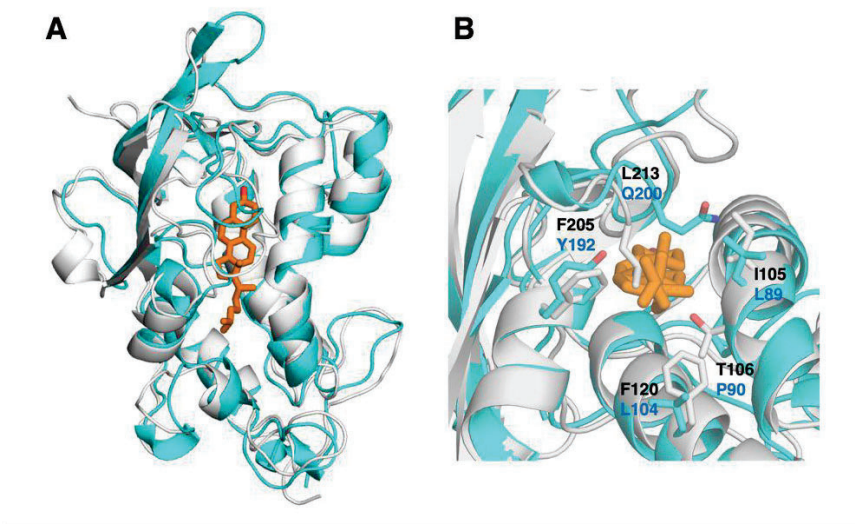


Figure S2. Superposition of the X-ray structures of the NTD domains of NPC1L1 and NPC1. (**A**) The global fold of both NTD domains of NPC1L1 and NPC1 are shown in grey and cyan cartoon respectively. (**B**) Comparison of the residues involved in the entrance of the binding site for both proteins. The crystallographic position of cholesterol in NPC1 is shown in orange sticks.

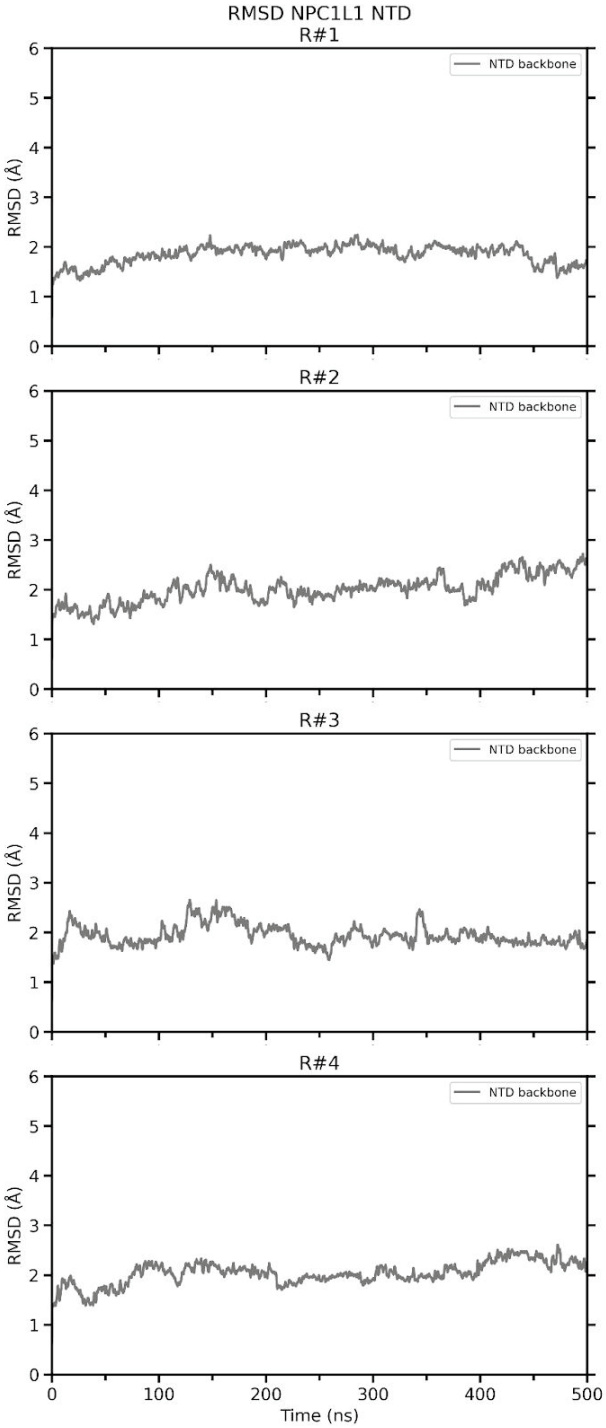


Figure S3. RMSD profiles of the backbone atoms of apo NPC1L1-NTD.

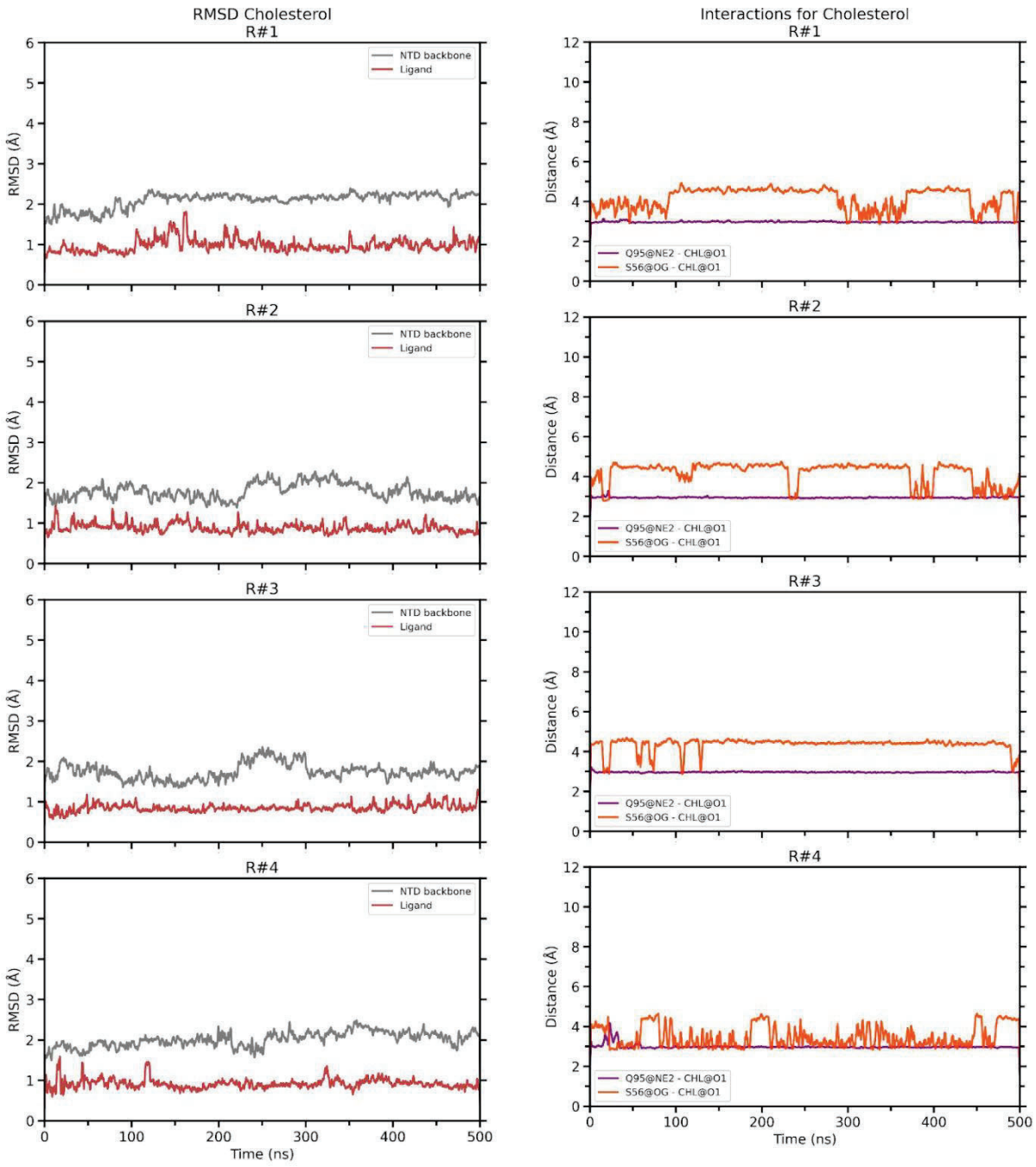


Figure S4. RMSD profiles of the cholesterol-bound NPC1L1-NTD and H-bond interactions of cholesterol over simulation time. (*Left*) Grey and red lines correspond to the RMSD of the backbone of the protein and heavy atoms of the ligand. (*Right*) Purple and orange lines correspond to the H-bonds between the hydroxyl moiety of the ligand and the sidechain of Q95 and S35.

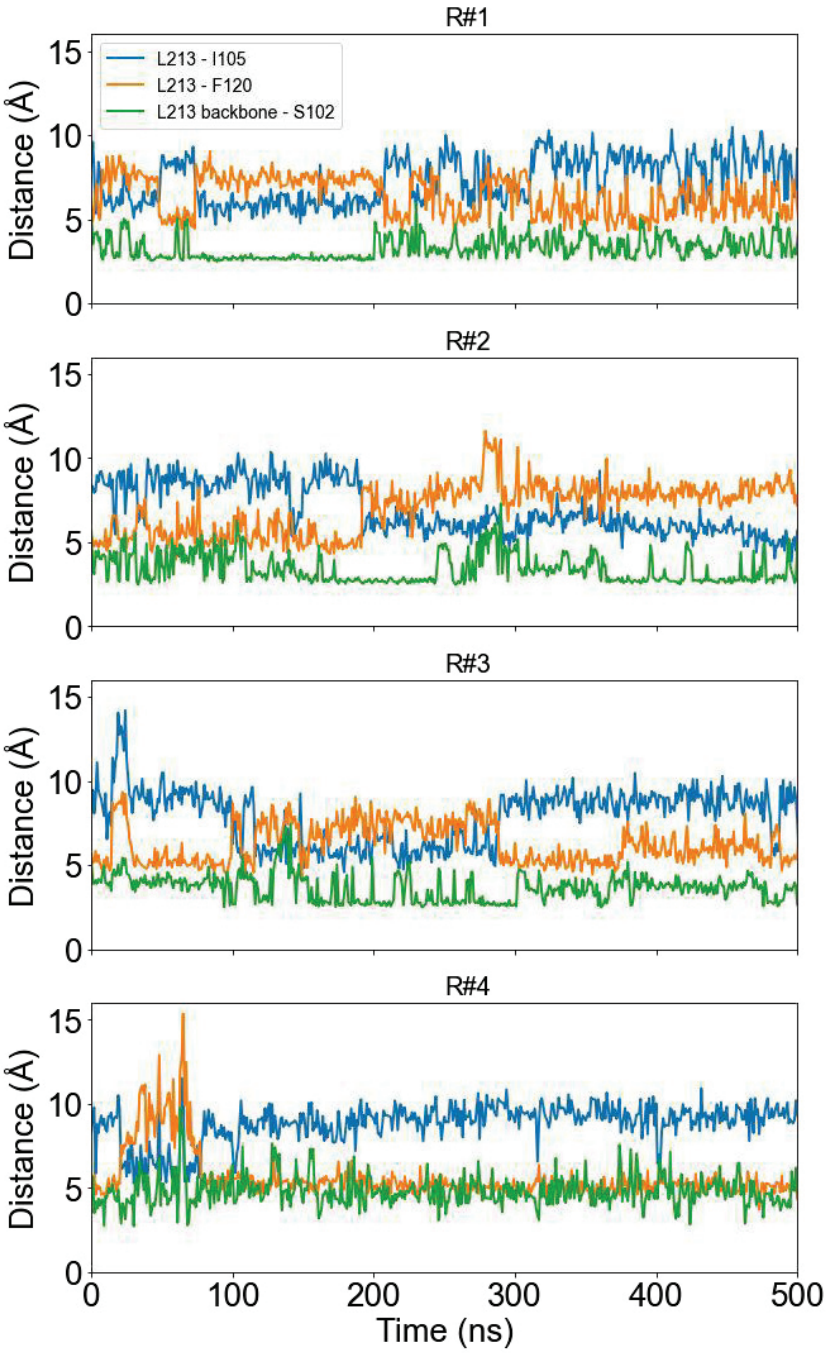


Figure S5. Selected interactions of L213 in the apo NPC1L1-NTD over simulation time. Blue and orange lines correspond to the distances of the CG atom of L213 to the CG atom of I105 and the center of the aromatic ring of F120, respectively. Green lines account for the H-bond interaction between the O atom of the backbone of L213 to the sidechain of S102.

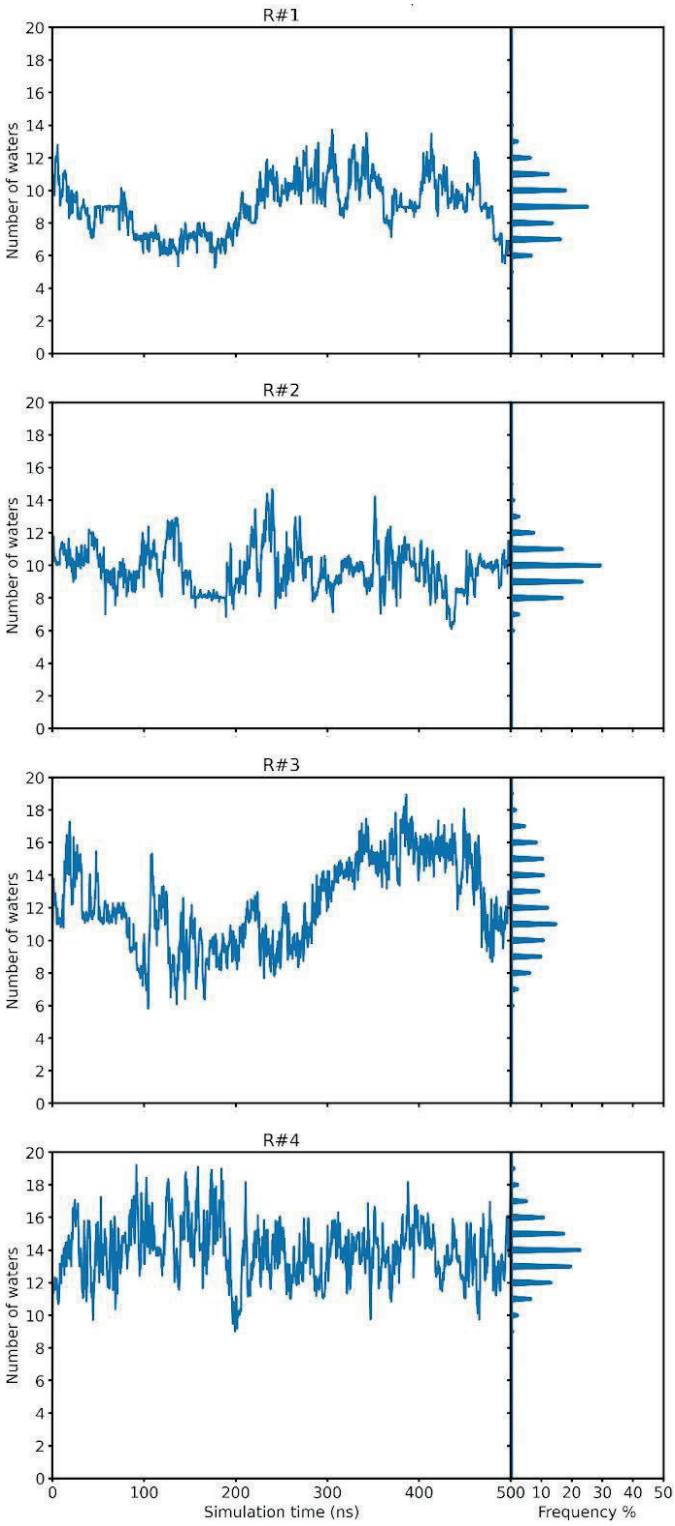


Figure S6. Number of water molecules in the binding site of the apo state of NPC1L1-NTD over simulation time.

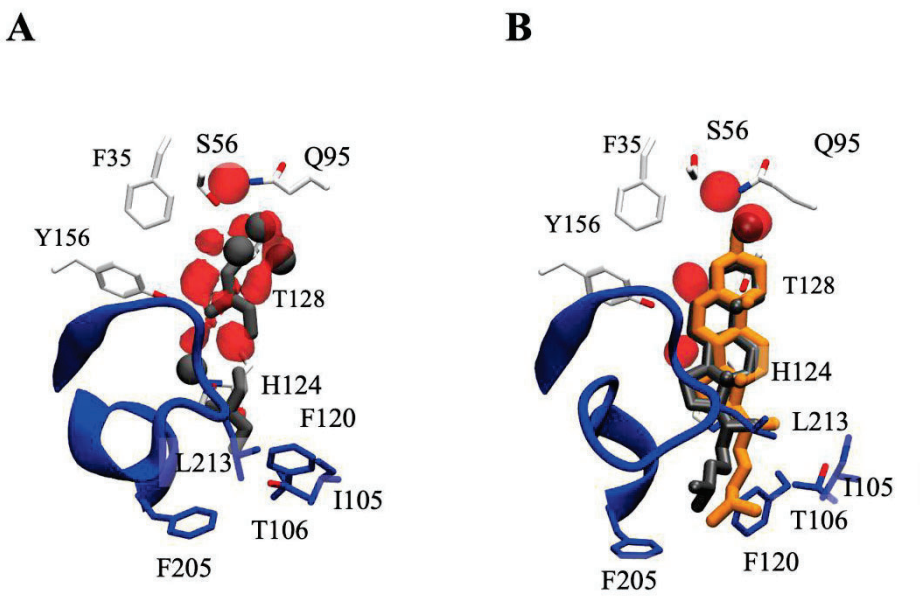


Figure S7. Overlap between the water network of NPC1L1-NTD with the X-ray structures of the NPC1-NTD. Close up views of the binding site in the (**A**) apo and (**B**) cholesterol-bound states. The occupancy of water molecules in the binding site over 50% of the simulation time is shown in the transparent red isosurface. The crystallographic molecules of water, cholesterol and glycerol found in the NPC1-NTD are shown as gray sticks and spheres. The NPC1L1-NTD is shown as white cartoon and sticks. The α8/β7 loop is shown as blue cartoon. The residues of the binding site of NPC1L1-NTD are shown as white and blue sticks.

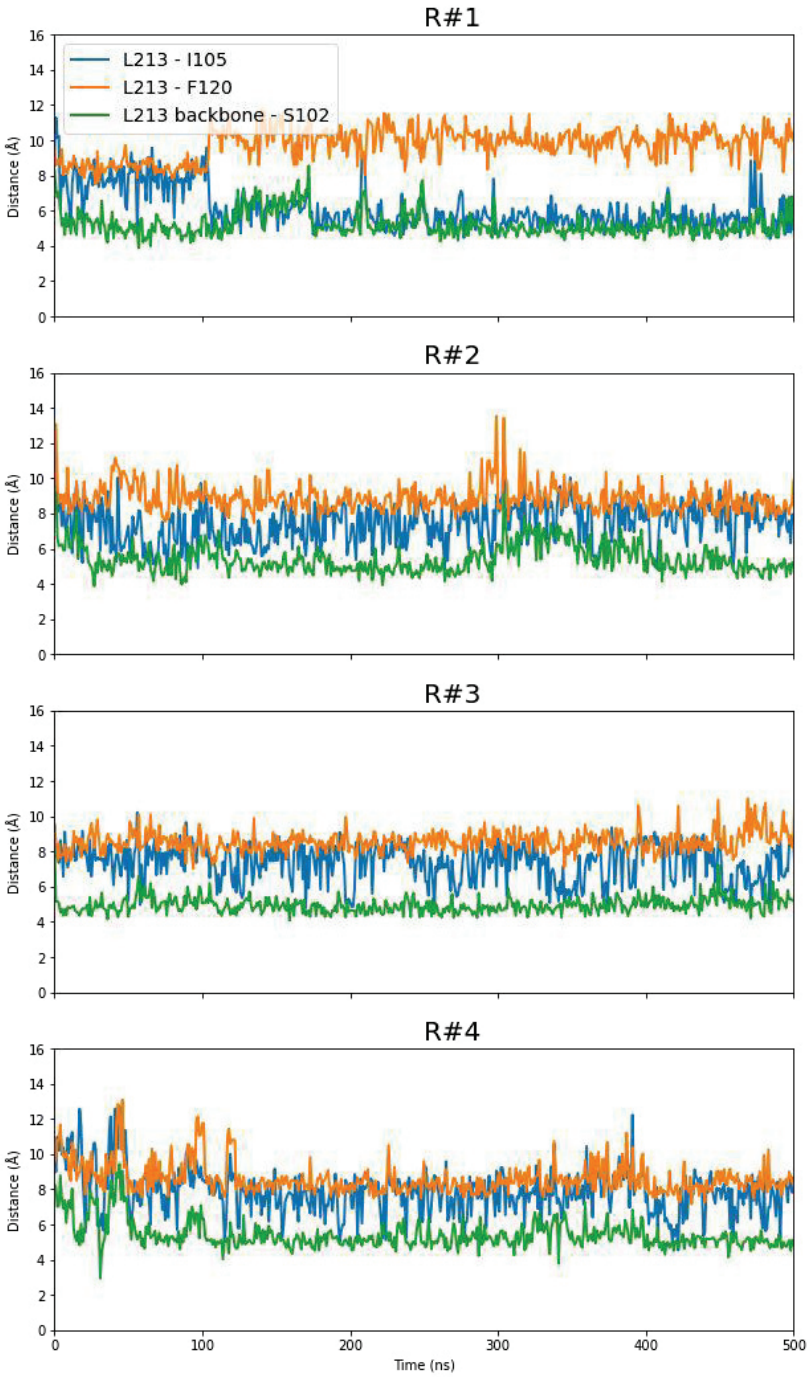


Figure S8. Selected interactions of L213 in the cholesterol-bound NPC1L1-NTD over simulation time. Blue and orange lines correspond to the distances of the CG atom of L213 to the CG atom of I105 and the center of the aromatic ring of F120 respectively. Green lines account for the H-bond interaction between the O atom of the backbone of L213 to the sidechain of S102.

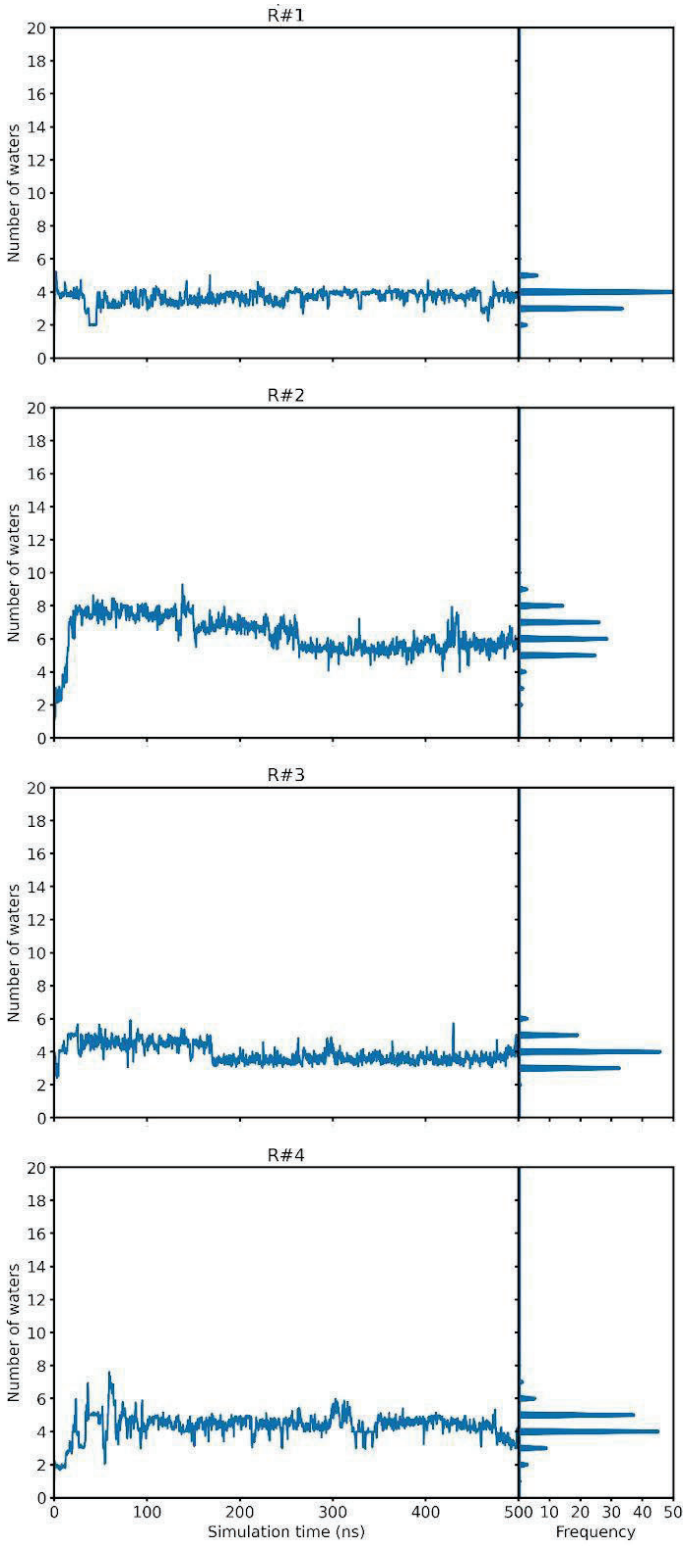


Figure S9. Number of water molecules in the binding site of the cholesterol-bound species of NPC1L1-NTD over simulation time.

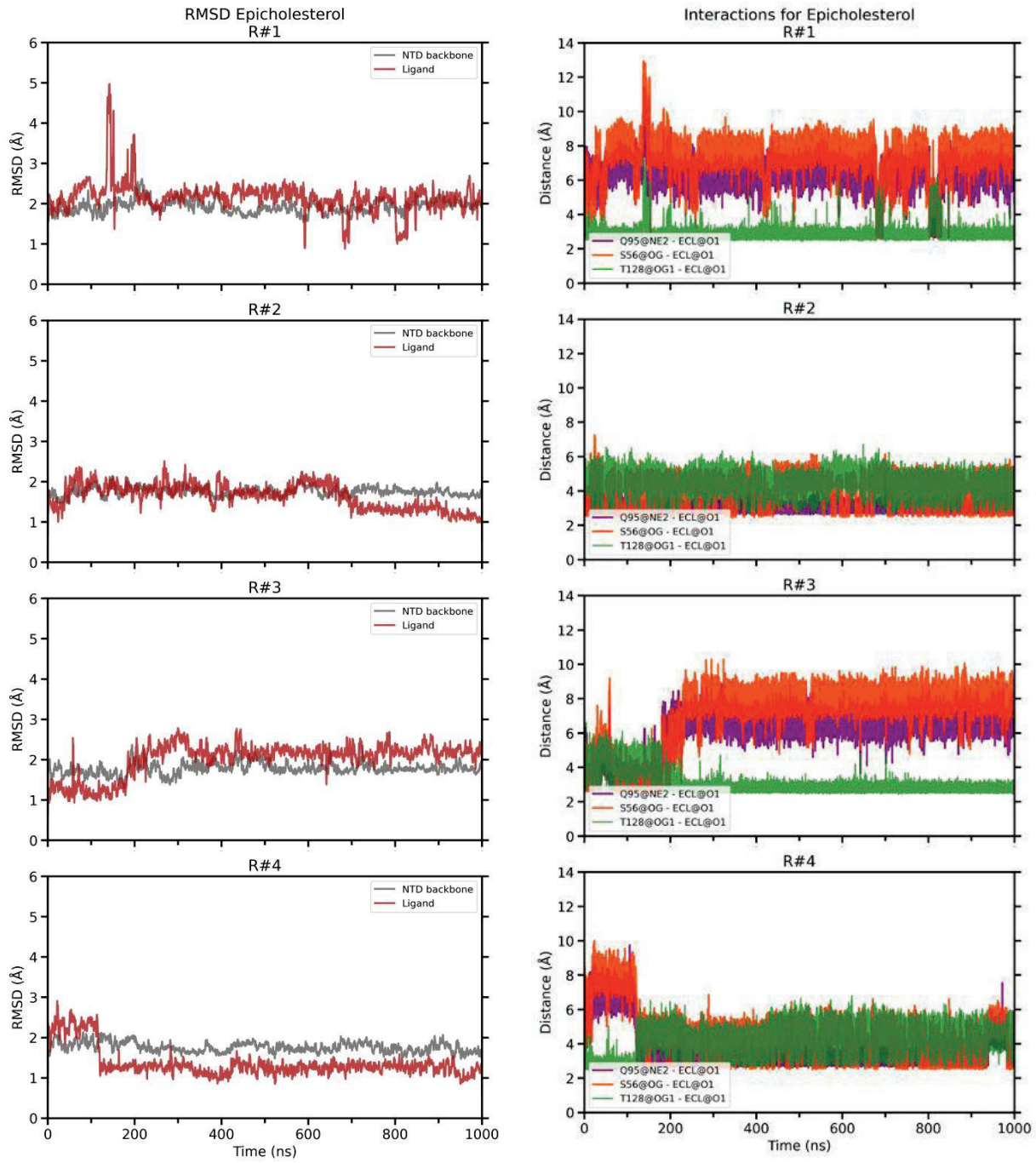


Figure S10. RMSD profiles of the epi-cholesterol-bound NPC1L1-NTD and H-bond interactions of the sterol over simulation time. (*Left*) Grey and red lines correspond to the RMSD of the backbone of the protein and heavy atoms of the ligand. (*Right*) Green, purple and orange lines correspond to the H-bonds between the hydroxyl moiety of the ligand and to the T128, Q95 and S35 sidechain.

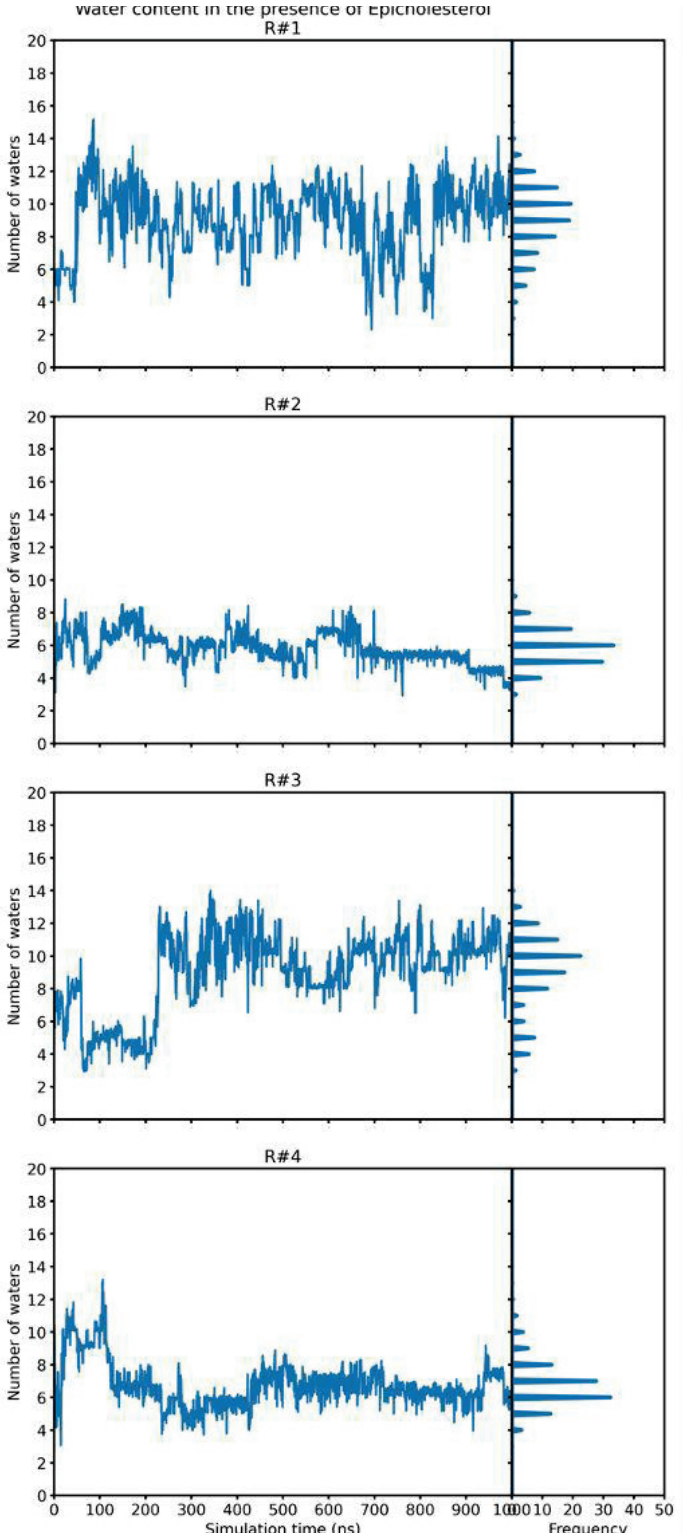


Figure S11. Number of water molecules in the binding site of the epi-cholesterol-bound species of NPC1L1-NTD over simulation time.

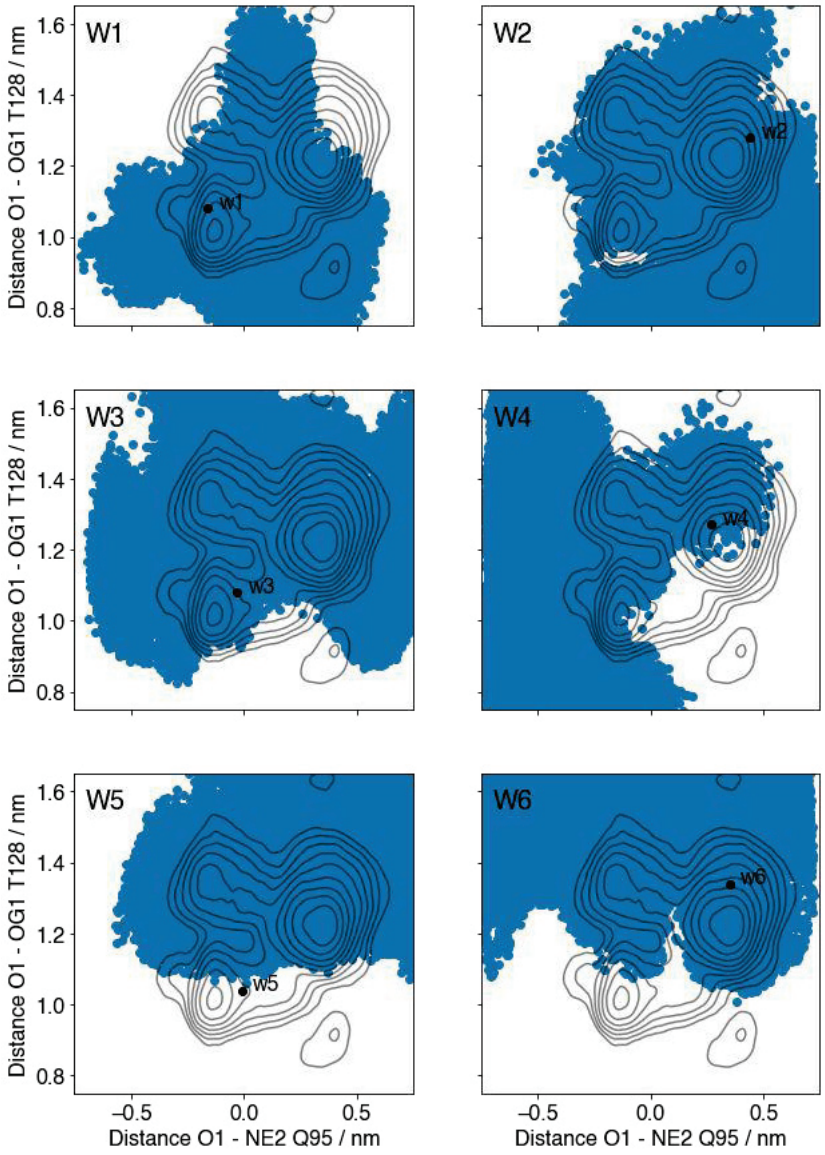


Figure S12. Representation of the two collective variables along the well-tempered metadynamics over the Collective Variable space. CV1 accounts for the difference of distances between the O atom of epi-cholesterol and the sidechains of residues Q95 and T128. CV2 accounts for the distance between the COM of the α8β7 loop and the α3 helix.

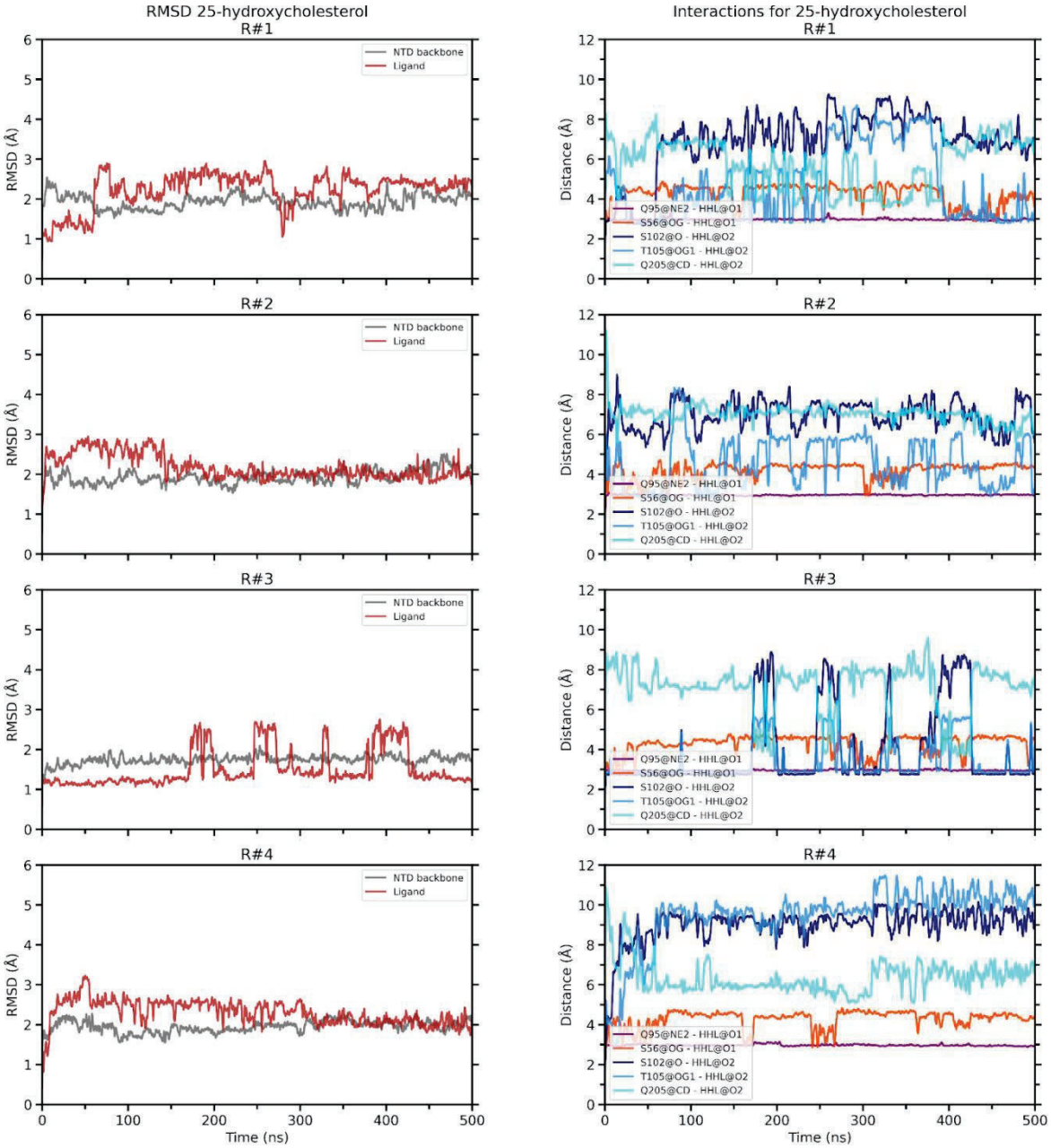


Figure S13. RMSD profiles of the 25-hydroxy-cholesterol-bound NPC1L1-NTD and H-bond interactions of the sterol over simulation time. (*Left*) Grey and red lines correspond to the RMSD of the backbone of the protein and heavy atoms of the ligand. (*Right*) Purple and orange lines correspond to the H-bonds between the hydroxyl moiety of the ligand and the sidechain of Q95 and S35. Blue lines show H-bond interaction of the 25-hydroxyl moiety.

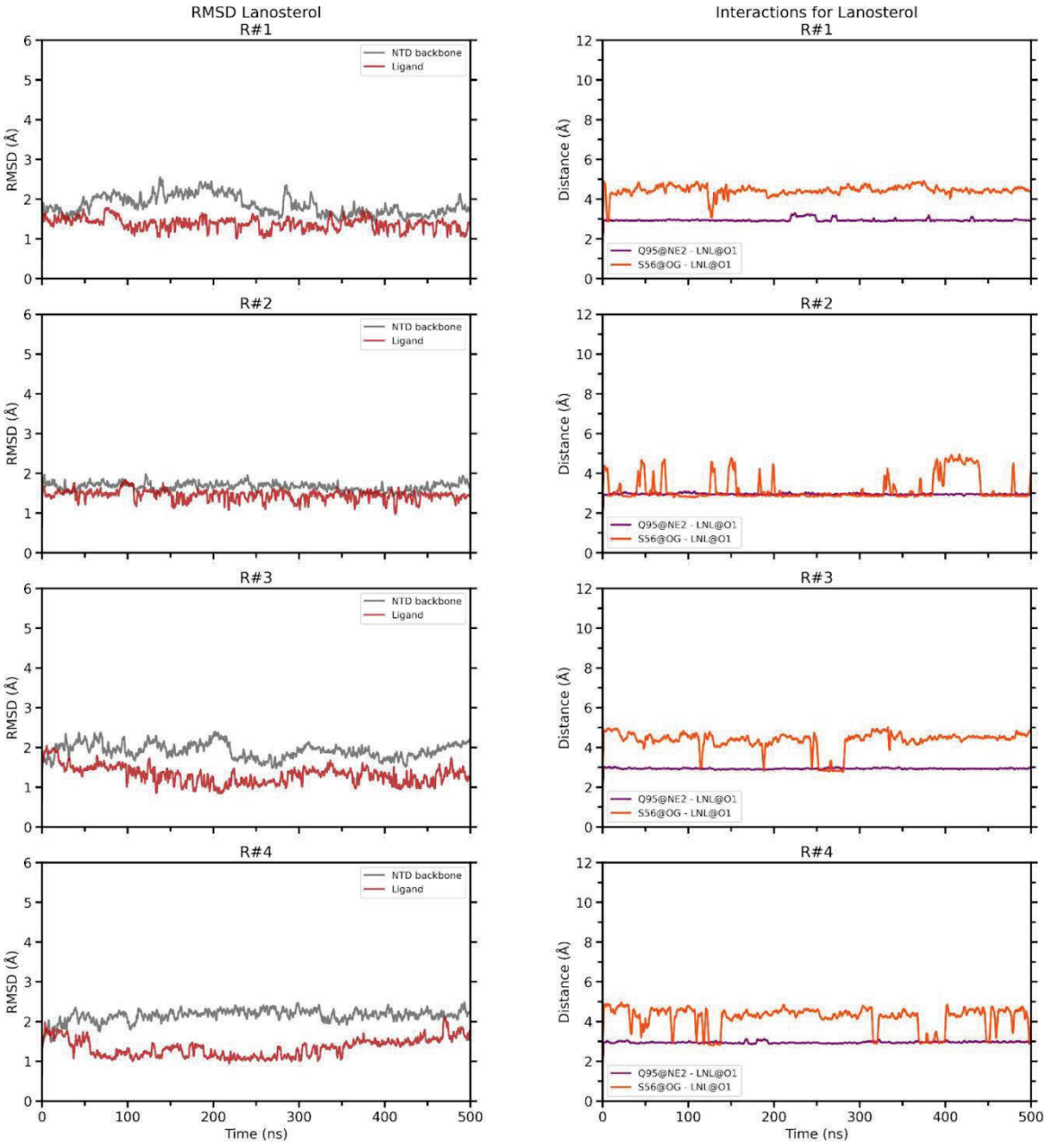


Figure S14. RMSD profiles of the lanosterol-bound NPC1L1-NTD and H-bond interactions of the sterol over simulation time. (*Left*) Grey and red lines correspond to the RMSD of the backbone of the protein and heavy atoms of the ligand. (*Right*) Purple and orange lines correspond to the H-bonds between the hydroxyl moiety of the ligand and the sidechain of Q95 and S35.

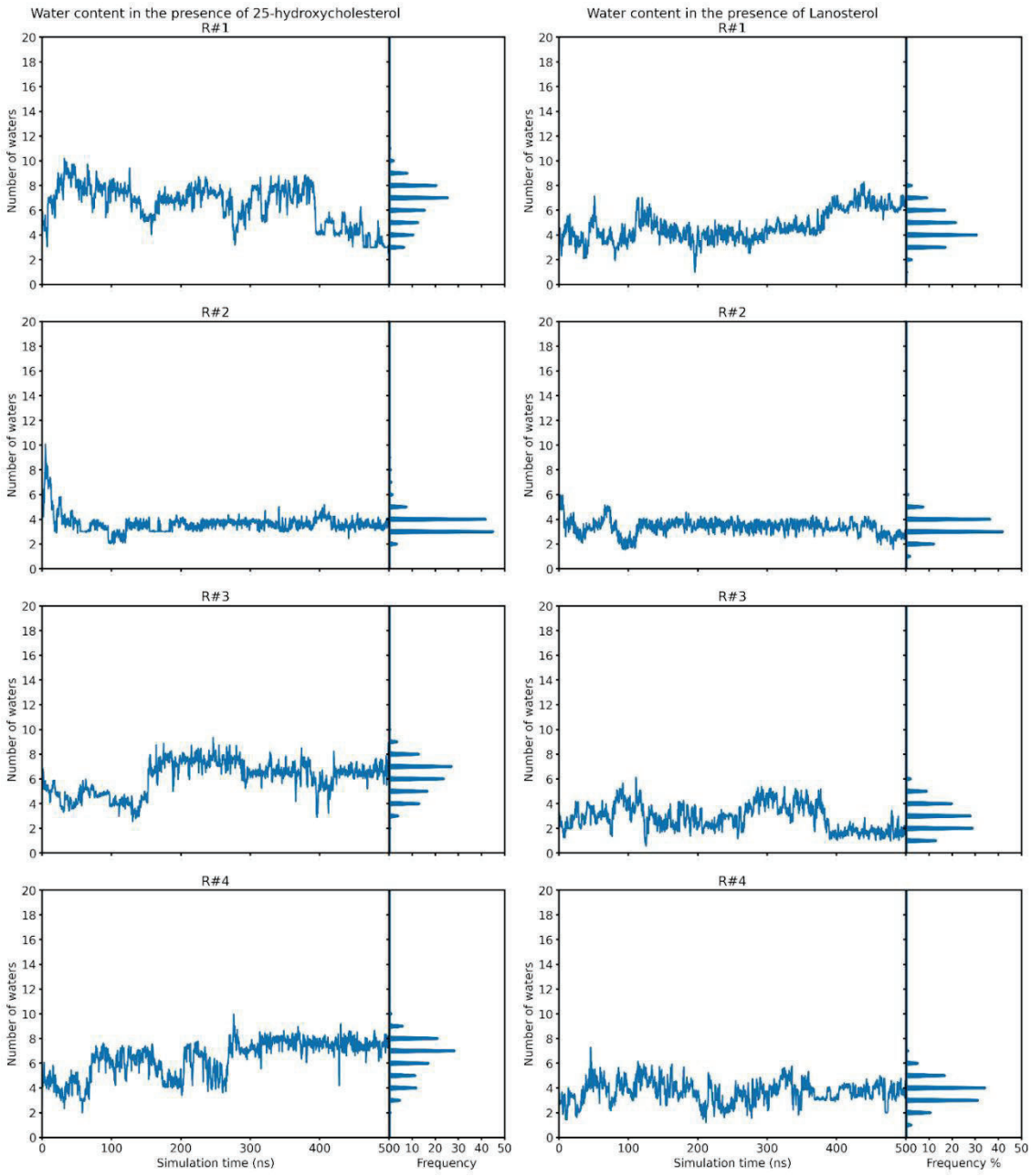


Figure S15. Number of water molecules in the binding site of (*left*) 25-hydroxy-cholesterol- and (*right*) lanosterol-bound species of NPC1L1-NTD over simulation time.

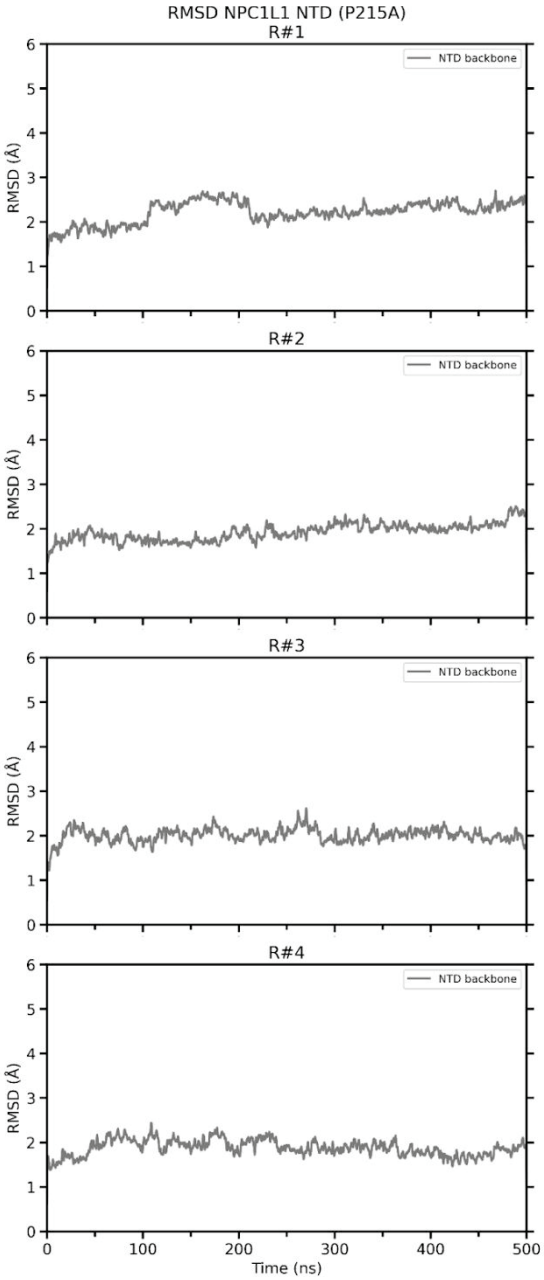


Figure S16. RMSD profiles of apo P215A NPC1L1-NTD over simulation time. Grey and red lines correspond to the RMSD of the backbone of the protein and heavy atoms of the ligand.

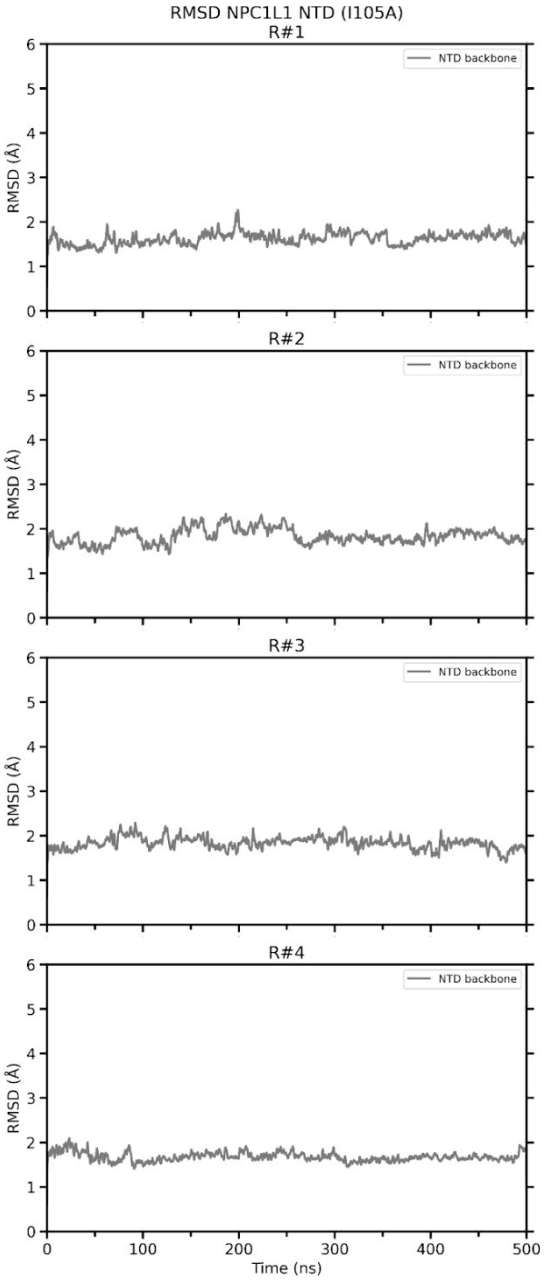


Figure S17. RMSD profiles of apo I105A NPC1L1-NTD over simulation time. Grey and red lines correspond to the RMSD of the backbone of the protein and heavy atoms of the ligand.

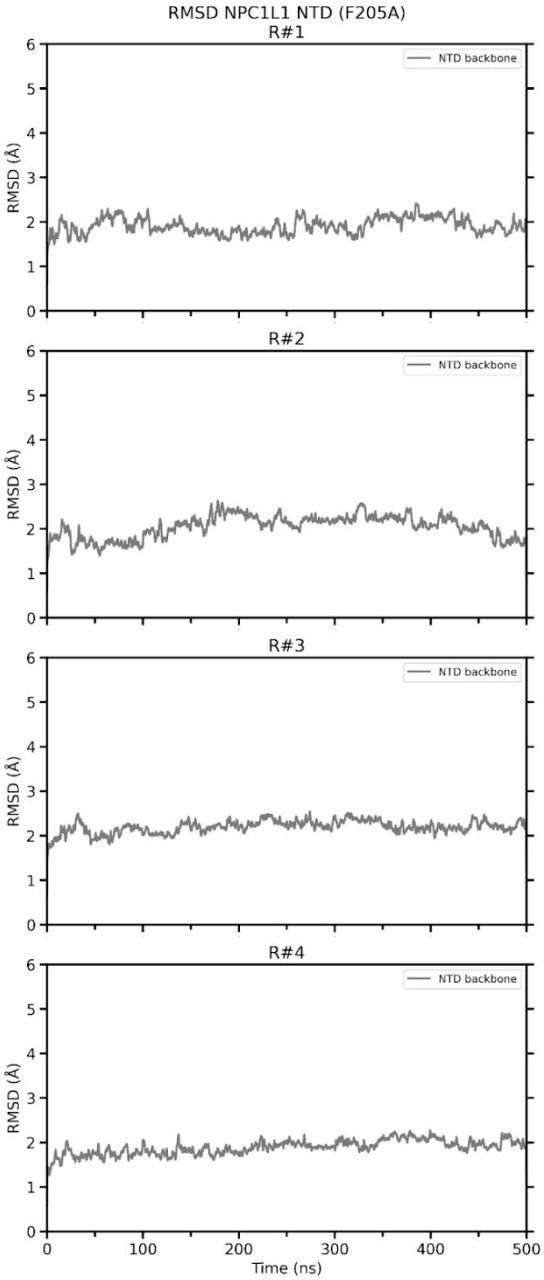


Figure S18. RMSD profiles of apo F205A NPC1L1-NTD over simulation time. Grey and red lines correspond to the RMSD of the backbone of the protein and heavy atoms of the ligand.

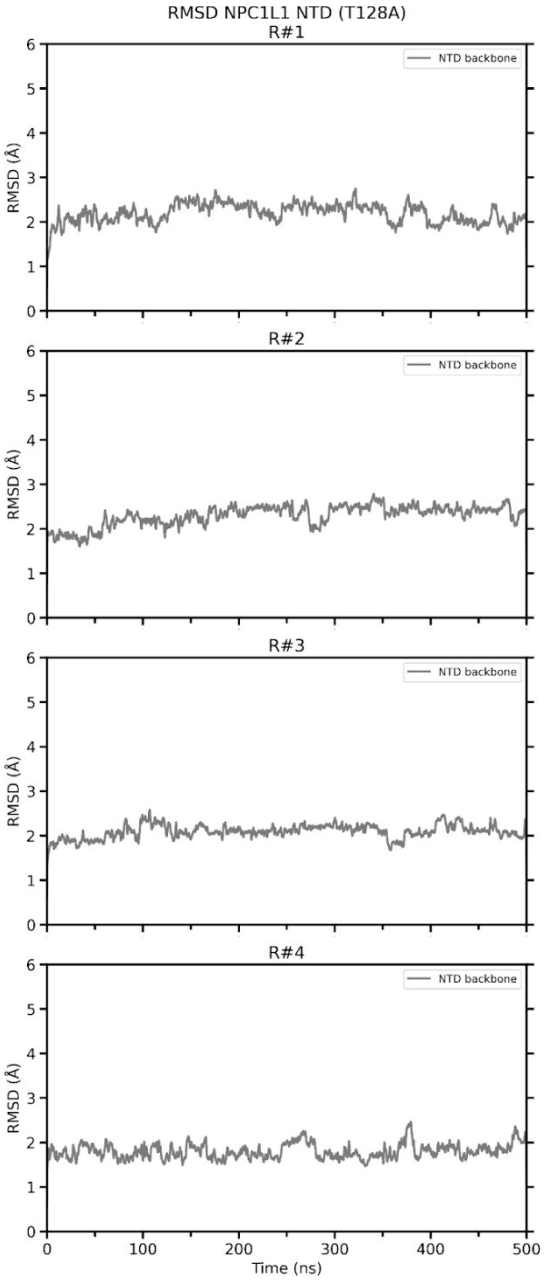


Figure S19. RMSD profiles of apo T128A NPC1L1-NTD over simulation time. Grey and red lines correspond to the RMSD of the backbone of the protein and heavy atoms of the ligand.

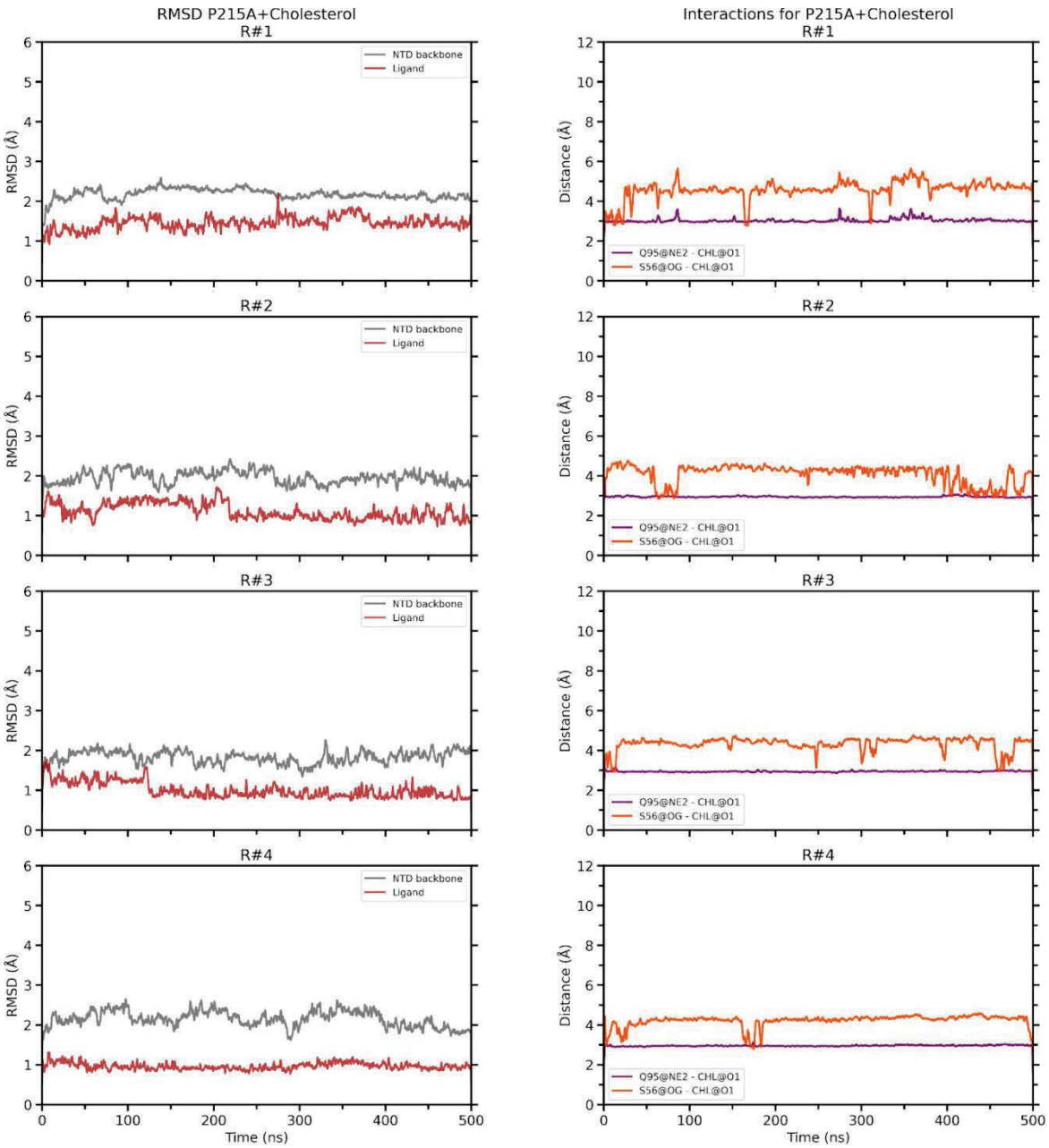


Figure S20. RMSD profiles of cholesterol-bound P215A NPC1L1-NTD and H-bond interactions of cholesterol over simulation time. Grey and red lines correspond to the RMSD of the backbone of the protein and heavy atoms of the ligand.

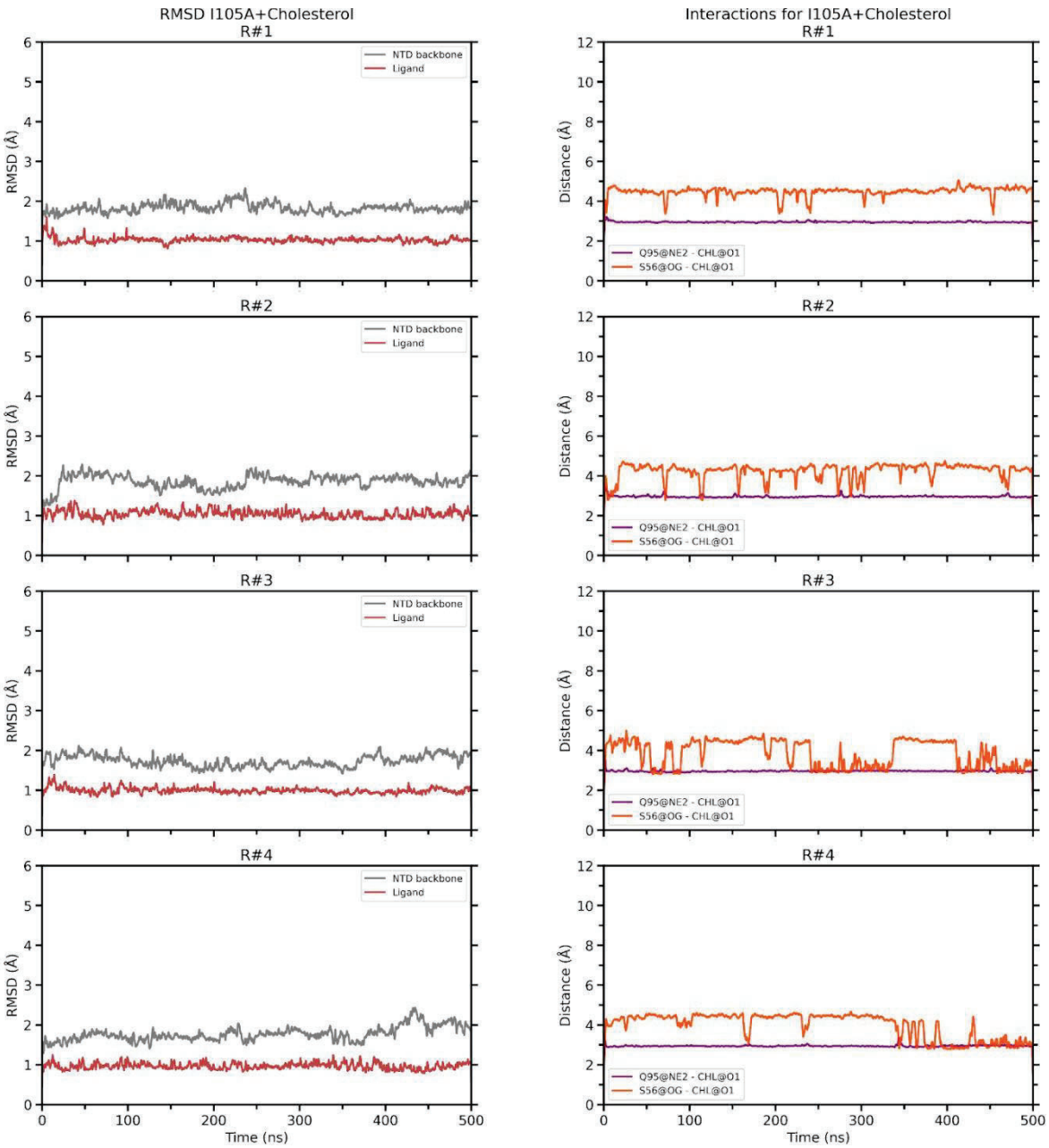


Figure S21. RMSD profiles of cholesterol-bound I105A NPC1L1-NTD and H-bond interactions of cholesterol over simulation time. Grey and red lines correspond to the RMSD of the backbone of the protein and heavy atoms of the ligand.

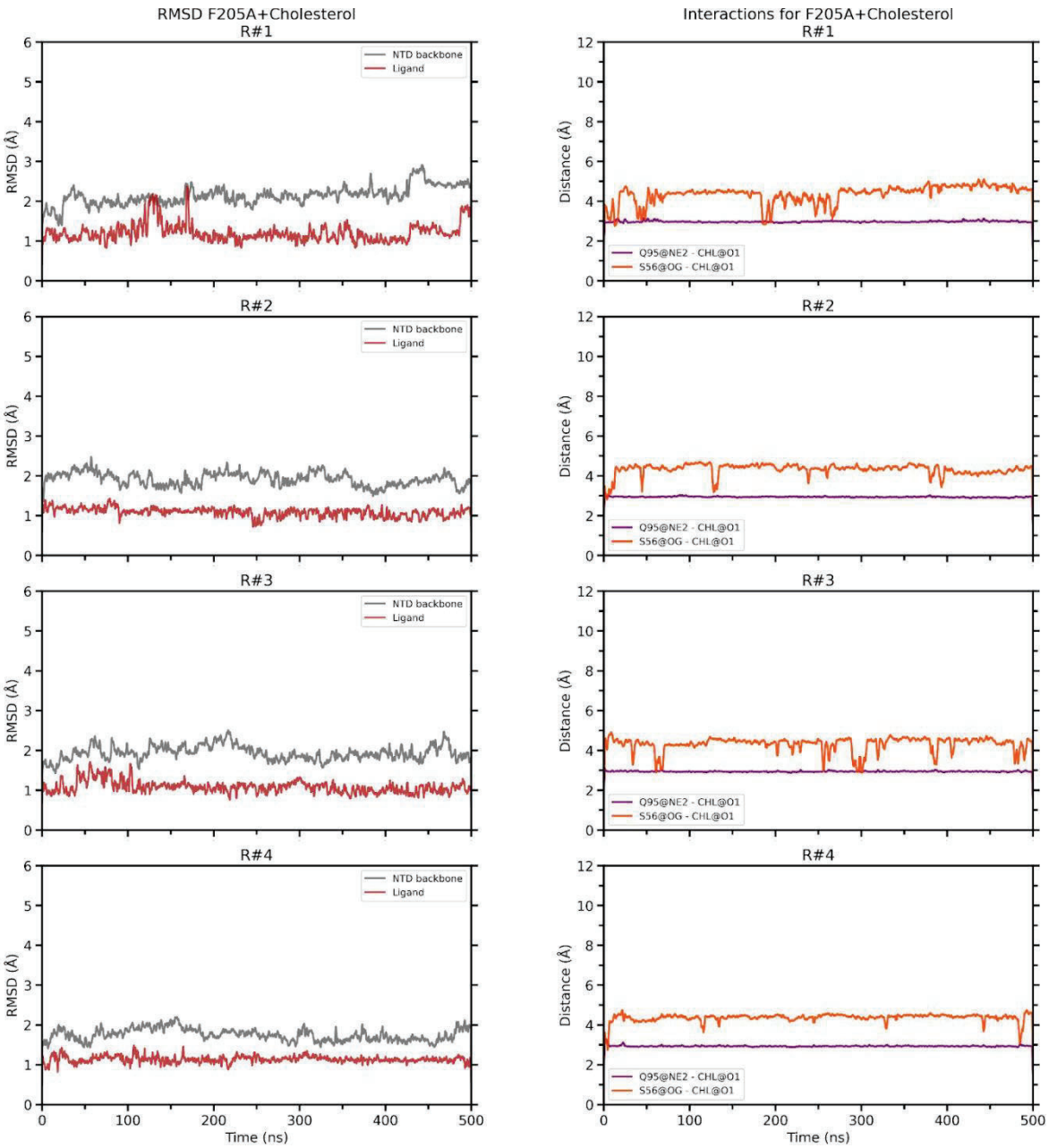


Figure S22. RMSD profiles of cholesterol-bound F205A NPC1L1-NTD and H-bond interactions of cholesterol over simulation time. Grey and red lines correspond to the RMSD of the backbone of the protein and heavy atoms of the ligand.

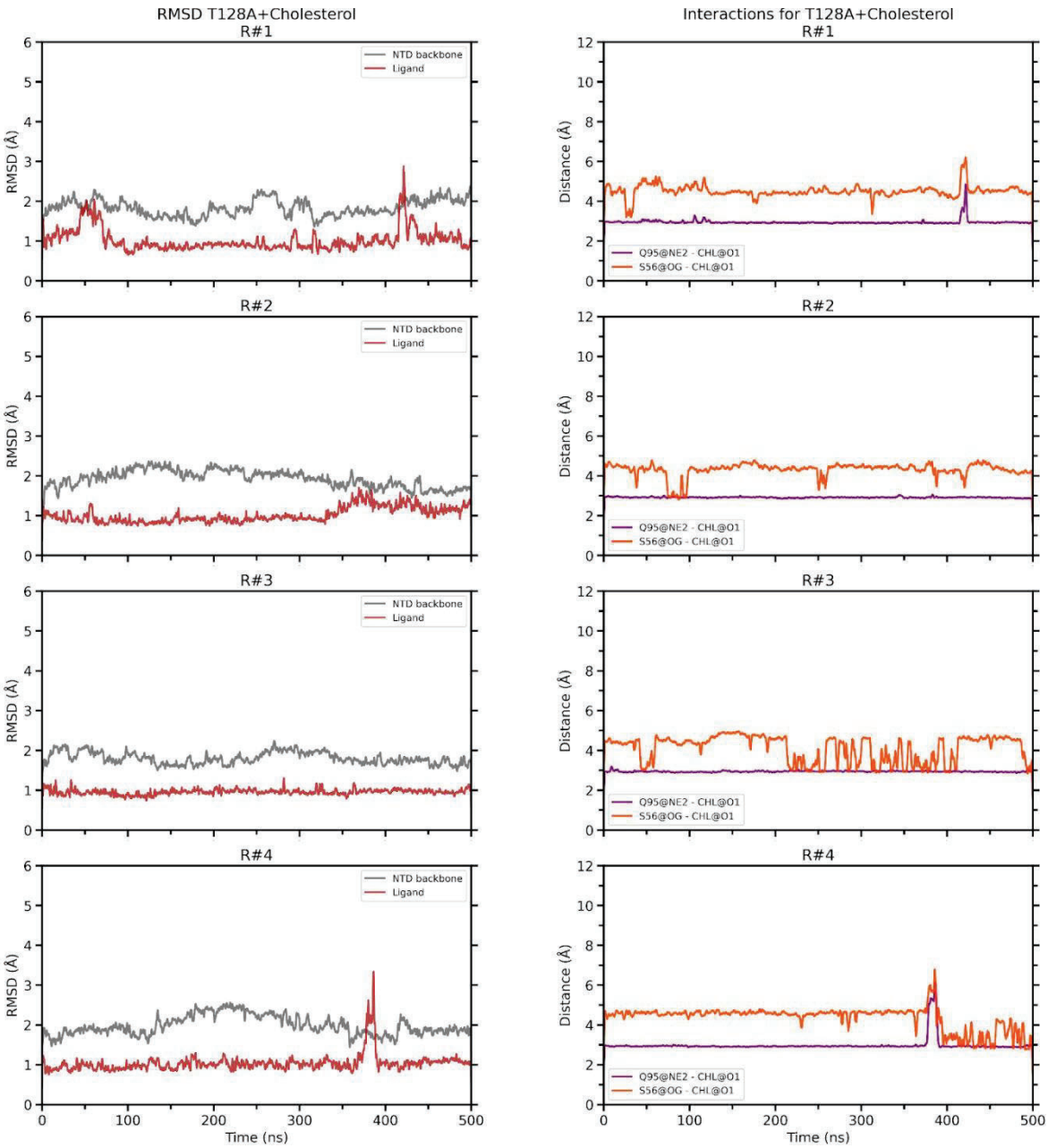


Figure S23. RMSD profiles of cholesterol-bound T128A NPC1L1-NTD and H-bond interactions of cholesterol over simulation time. Grey and red lines correspond to the RMSD of the backbone of the protein and heavy atoms of the ligand.

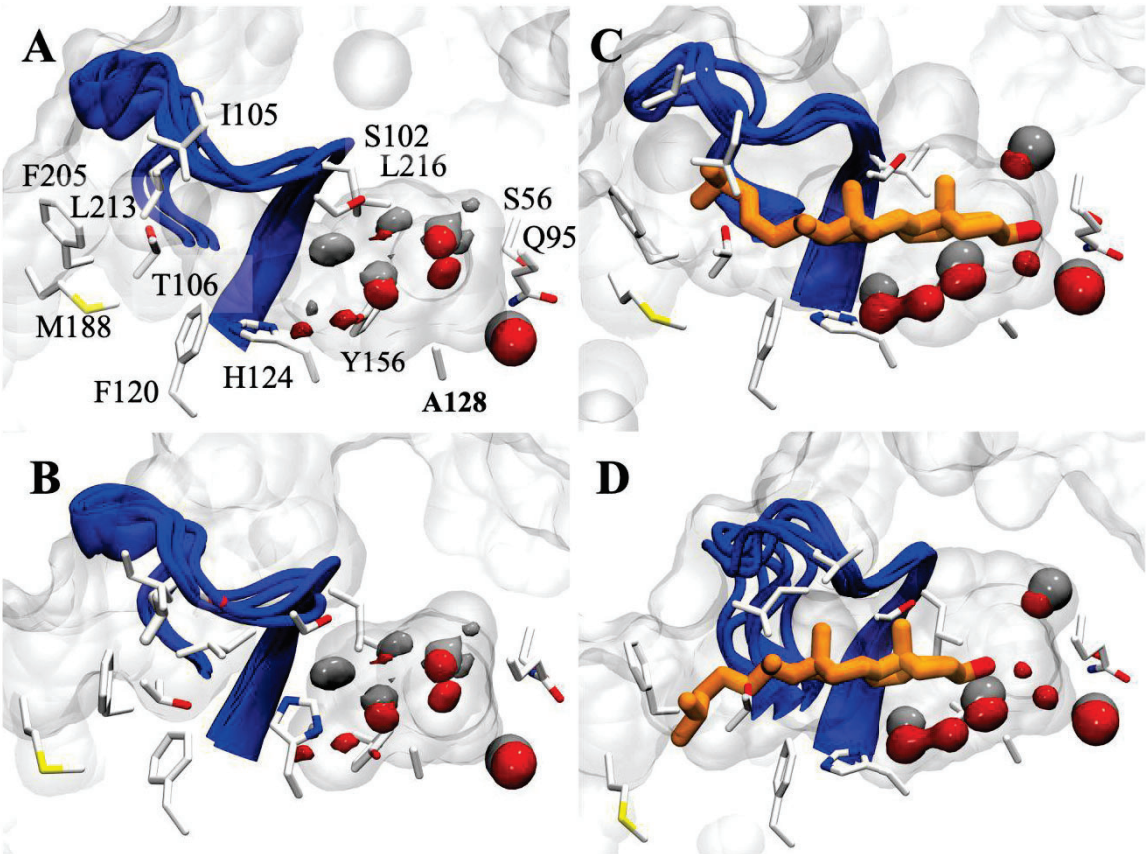


Figure S24. Superposition of the major conformations of the (**A,C**) apo and (**B,D**) cholesterol-bound states of the T128A mutant. (**A**) The major A2 and minor A1 (**C**) conformational states of the apo T128A mutant. **(B,D)** The two binding modes of cholesterol in the T128A mutants. The conformations sampled by the α8/β7 loop are shown as blue cartoon. The protein surface is shown as grey surface. The isosurface accounts for 50% of water occupancy in the binding site during the last 100 ns of each replicate (accumulated 400 ns of simulation time). The water occupancy for the wildtype protein and T128A mutant are shown as grey and red isocontours, respectively.

Table S1. Available structural information of the NPC1L1 protein on PDB database.

NPC1L1 (Uniprot Q9UHC9 & Uniprot Q6T3U3)						
PDB id	**Method**	**Resolution Å**	**Spp**	**NTD-included**	**Ligand bound**	**Reference**
7DFW	EM	2.69	*Homo sapiens*	NO	Cholesterol	Long T. *et al* *Sci. Adv.* (**2021**) 7, eabh3997
7DFZ	EM	3.58	*Homo sapiens*	NO	Ezetimibe	
7DF8	EM	3.03	*Homo sapiens*	NO	---	
7N4V	EM	3.58	*Homo sapiens*	NO	Cholesterol	
7N4X	EM	3.33	*Homo sapiens*	NO	Cholesterol	
7N4U	EM	3.34	*Homo sapiens*	NO	α-tocopherol	
6V3F	EM	3.70	*Rattus norvegicus*	YES	Cholesterol	Huang CS. *et al* *Sci. Adv.* (**2020**) 6, eabb1989
6V3H	EM	3.50	*Rattus norvegicus*	YES	Ezetimibe analog	
3QNT	X-Ray	2.83	*Homo sapiens*	YES	---	Kwon HJ, *et al* *PLoS ONE* (**2011**) 6: e18722

Table S2. Available structural information of the NPC1 protein on PDB database.

NPC1 (Uniprot O15118)						
PDB id	**Method**	**Resolution Å**	**Spp**	**NTD-included**	**Ligand bound**	**Reference**
8EUS	X-Ray	2.30	*Homo sapiens*	NO	---	Odongo L. *et al* *Acta Cryst. F* (**2023**) 79, 45–50.
6UOX	EM	4.02	*Homo sapiens*	YES	Itraconazole	Long T. *et al* *Nat Commun* (**2020**) **11**, 152.
6W5R	EM	3.60	*Homo sapiens*	YES	Cholesterol	Qian H. *et al* *Cell* (**2020**) *182*, 98-111
6W5S	EM	3.00	*Homo sapiens*	YES	Cholesterol	
6W5T	EM	3.70	*Homo sapiens*	YES	Cholesterol	
6W5U	EM	3.90	*Homo sapiens*	YES	Cholesterol	
6W5V	EM	4.00	*Homo sapiens*	YES	Cholesterol	
5U73	X-Ray	3.35	*Homo sapiens*	YES	---	Li, X. *et al* *Proc. Natl. Acad. Sci.* (**2017**), *114*, 9116-9121.
5U74	X-Ray	3.34	*Homo sapiens*	YES	---	
5KWY	X-Ray	2.41	*Homo sapiens*	NO	Cholesterol-sulfate	Li X. *et al* *Proc. Natl. Acad. Sci.* (**2016**) *113*, 10079–10084.
5JNX	EM	6.56	*Homo sapiens*	YES	---	Gong, X. *et al* *Cell* (**2016**) *165*, 1467–1478.
5HNS	X-Ray	2.45	*Homo sapiens*	NO	---	Zhao, Y. *et al* *FEBS Lett* (**2016**) 590: 605-612
5F18	X-Ray	2.00	*Homo sapiens*	NO	---	Wang, H. *et al* *Cell* (**2016**), *164*, 258–268.
5F1B	X-Ray	2.30	*Homo sapiens*	NO	---	
3GKH	X-Ray	1.81	*Homo sapiens*	YES	---	Kwon H. J. *et al* *Cell* (**2009**) *137*, 1213–1224.
3GKI	X-Ray	1.80	*Homo sapiens*	YES	Cholesterol	
3GKJ	X-Ray	1.60	*Homo sapiens*	YES	25-hydroxy-cholesterol	
3JD8	EM	4.43	*Homo sapiens*	YES	Cholesterol	To be published.

Table S3. Average distance (A) and preservation (% of simulation time) of the interactions between wild-type NPC1L1-NTD and the studied sterol molecules.

			Replicas			
Complex	**H-bond interaction**		**R1**	**R2**	**R3**	**R4**
Cholesterol	**Q95 and OH of sterol**	**Average distance**	2.9	2.9	2.9	2.9
		%	64	71	66	64
	S56 and OH of sterol	**Average distance**	2.8	2.8	2.8	2.8
		%	15	10	4	40
25-hydroxycholesterol	**Q95 and OH of sterol**	**Average distance**	2.9	2.9	2.9	2.9
		%	64	67	70	65
	S56 and OH of sterol	**Average distance**	2.8	2.8	2.8	2.8
		%	11	10	7	8
	T106 and OH of aliphatic chain	**Average distance**	2.8	2.8	2.8	2.8
		%	24	16	57	1
	S102 and OH of aliphatic chain	**Average distance**	2.8	2.8	2.7	2.8
		%	8	0	64	0
	Q206 and OH of aliphatic chain	**Average distance**	2.8	2.9	2.8	2.9
		%	23	0	9	0
Lanosterol	**Q95 and OH of sterol**	**Average distance**	2.9	2.9	2.9	2.9
		%	70	67	74	67
	S56 and OH of sterol	**Average distance**	2.8	2.8	2.8	2.8
		%	1	41	4	6

Table S4. Description of the main conformational states of the apo and cholesterol-bound systems found for the wildtype NPC1L1-NTD and its mutated variants.

	Conformation	**Population %**	**Volume** Å
wildtype	A1	24 ± 2	483 ± 116
	A2	75 ± 2	491 ± 145
	B1	97 ± 2	1025 ± 61
I105A	A2	76 ± 5	264 ± 43
	A4	18 ± 3	525 ± 44
	B1	93 ± 1	985 ± 52
T128A	A1	29 ± 1	459 ± 47
	A2	52 ± 4	434 ± 37
	A3	17 ± 3	347 ± 86
	B1	86 ± 1	935 ± 54
	B3	8 ± 1	968 ± 54
F205A	A1	12 ± 1	488 ± 123
	A2	35 ± 1	475 ± 126
	A3	31 ± 1	366 ± 43
	A4	23 ± 1	745 ± 95
	B1	90 ± 1	994 ± 45
	B3	4 ± 1	973 ± 72
P215A	A1	24 ± 1	361 ± 76
	A2	42 ± 1	390 ± 55
	A3	25 ± 2	444 ± 65
	A5	9 ± 1	288 ± 22
	B1	79 ± 1	902 ± 51
	B2	20 ± 1	941 ± 73

Project 2

"Mining Druggable Sites in Influenza A Hemagglutinin:
Binding of the Pinanamine Based Inhibitor M090"

"Mining Druggable Sites in Influenza A Hemagglutinin: Binding of the Pinanamine Based Inhibitor M090"

Aitor Valdivia,[†,‡] Maria Rocha,[§,||] and F. Javier Luque[‡,§,#]

† Doctorate in Biotechnology, Departament de Nutrició, Ciències de l'Alimentació I Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació - Campus Torribera, Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain

‡ Institut de Biomedicina (IBUB), Universitat de Barcelona, 08028 Barcelona, Spain

§ Departament de Nutrició, Ciències de l'Alimentació I Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació - Campus Torribera, Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain

|| Department of Life Sciences, University of Coimbra, Calçada Martim de Freitas, 3000-456 Coimbra, Portugal

Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, 08921 Barcelona, Spain

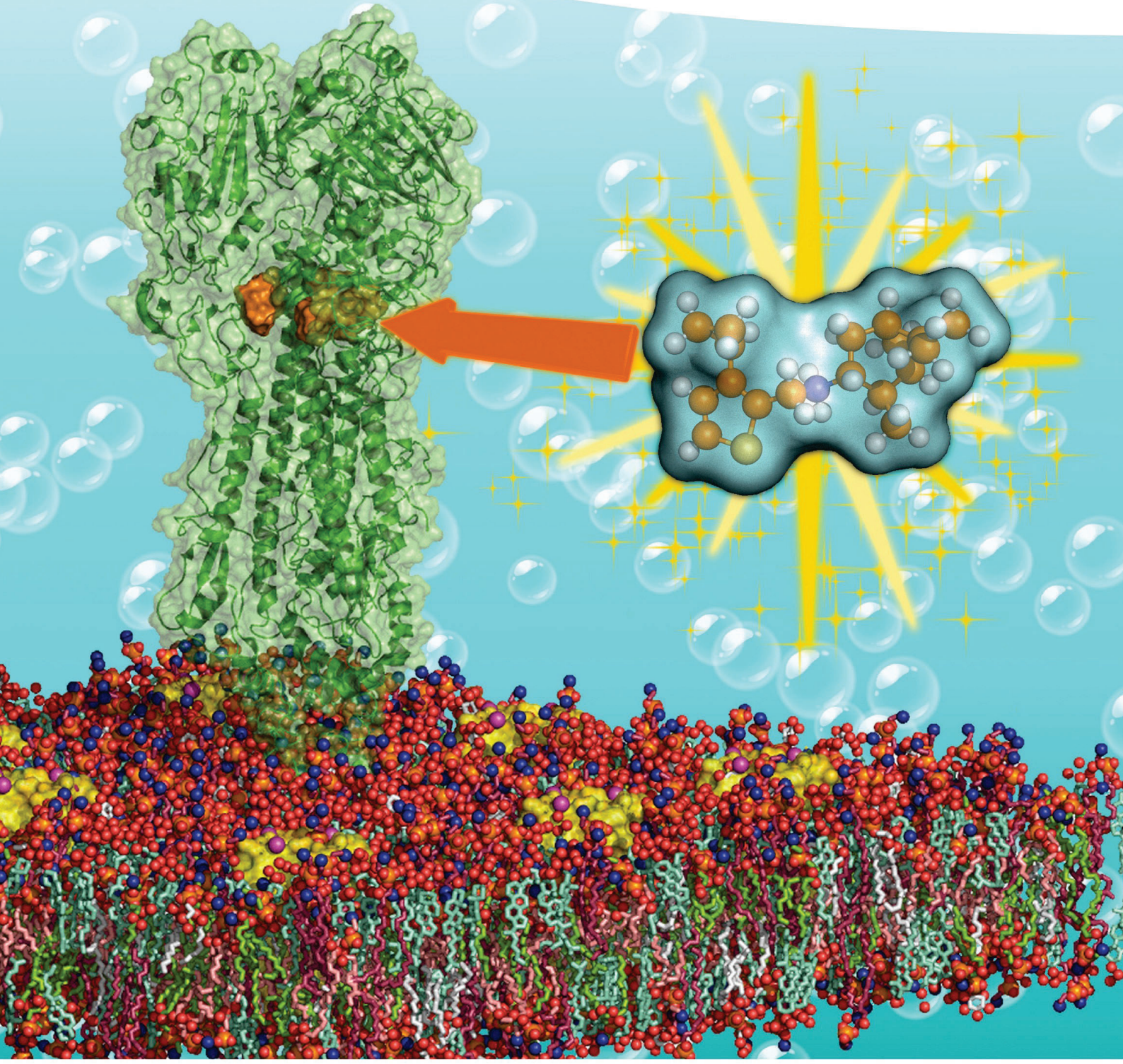
Published in: ACS Medicinal Chemistry Letters, **2024**, 16, 1, 126-135

Note: A Front Cover Art was prepared and selected by the journal editors to be the front cover of the issue. The cover is included in the next page of this manuscript.

ACS Medicinal Chemistry Letters

January 2025 Volume 16, Issue 1

pubs.acs.org/acsmedchemlett



ACS Publications
Most Trusted. Most Cited. Most Read.

www.acs.org

Summary

Assessing the binding mode of drug-like compounds is key in structure-based drug design. However, this may be challenged by factors such as the structural flexibility of the target protein. In this case, state-of-the- art computational methods can be valuable to explore the linkages between structural and pharmacological data. Following this strategy, extended molecular dynamics simulations and thermodynamic integration calculations are used to examine the binding of the potent antiviral inhibitor M090 and related pinanamine-based analogues, covering a 250-fold difference in inhibitory potency to the influenza A hemagglutinin, which is essential for virus entry and membrane fusion. This analysis has disclosed the hydrophobic shielding effect played by the 3-cyclopropylthiophene moiety in M090. Furthermore, the results support the negative effect of the resistance-induced E742 → D mutation, which should weaken the binding by increasing the structural flexibility of the L2-BS loop. The results pave the way to exploration of the antiviral activity of novel compounds.

ACS Medicinal Chemistry Letters

Open Access
This article is licensed under CC-BY 4.0

pubs.acs.org/acsmedchemlett
Letter

Mining Druggable Sites in Influenza A Hemagglutinin: Binding of the Pinanamine-Based Inhibitor M090

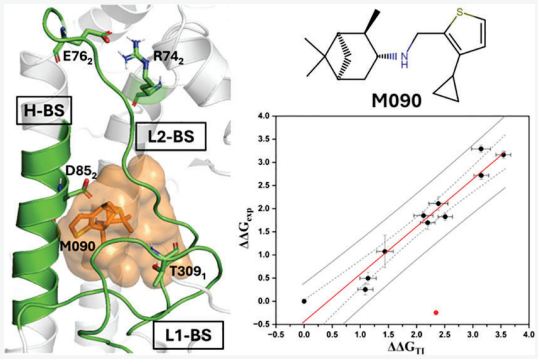
Published as part of ACS Medicinal Chemistry Letters special issue "Structural Biology in Drug Discovery and Development".

Aitor Valdivia, Maria Rocha, and F. Javier Luque*

Cite This: https://doi.org/10.1021/acsmedchemlett.4c00502
Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Assessing the binding mode of drug-like compounds is key in structure-based drug design. However, this may be challenged by factors such as the structural flexibility of the target protein. In this case, state-of-the-art computational methods can be valuable to explore the linkages between structural and pharmacological data. Following this strategy, extended molecular dynamics simulations and thermodynamic integration calculations are used to examine the binding of the potent antiviral inhibitor M090 and related pinanamine-based analogues, covering a 250-fold difference in inhibitory potency to the influenza A hemagglutinin, which is essential for virus entry and membrane fusion. This analysis has disclosed the hydrophobic shielding effect played by the 3-cyclopropylthiophene moiety in M090. Furthermore, the results support the negative effect of the resistance-induced $E74_2 \rightarrow D$ mutation, which should weaken the binding by increasing the structural flexibility of the L2-BS loop. The results pave the way to exploration of the antiviral activity of novel compounds.



KEYWORDS: *Influenza A, Hemagglutinin, Ligand binding, Inhibition, Molecular dynamics, Free energy calculations*

Downloaded via 37.222.105.83 on November 28, 2024 at 11:12:33 (UTC).
See https://pubs.acs.org/sharingguidelines for options on how to legitimately share published articles.

Mapping druggable pockets in therapeutically relevant proteins is a fundamental step in structure-based drug design. They should encode geometrical and physicochemical features suitable to accommodate drug-like compounds,[1,2] including hot spots that can make major contributions to the binding affinity. Moreover, a proper balance between regions of high stability and high flexibility is also required.[3−7] However, the structural plasticity of targets such as the influenza A hemagglutinin (HA), which is crucial for viral infectivity,[8−13] may severely limit the success of drug discovery. However, the notion of HA as a druggable target has gained increasing support due to the discovery of small molecules that interfere with the role played in virus entry and membrane fusion.[14]

HA promotes entry of the virus into the cell by recognizing the sialic acid (SA) residues attached to cell membrane receptors. On the other hand, in the interior of the endosome, a large-scale conformational rearrangement of HA is triggered upon acidification, leading to the release of the fusion peptide (FP), which is hidden in the stem. The pH-induced structural remodeling promotes the fusion of viral and host membranes, enabling the release of the viral genetic material.[15,16]

From a structural point of view, HA is a homotrimer located on the viral envelope. Each protomer is expressed as a single, inactive precursor known as HA_0, and maturation to the active form requires proteolytic cleavage by host cell proteases, leading to chains HA_1 and HA_2 (Figure 1).[17] Two structural domains can be distinguished in HA: the globular head, which contains the receptor binding domain that mediates binding to the host cell membrane, and the stem region, which maintains the whole structure bound to the viral membrane. Three main strategies have been explored in the search of small molecules targeting HA:[18,19] (1) altering the HA maturation via inhibition of host proteases required for proteolytic cleavage or alternatively through blockage of HA glycosylation;[20] (2) impairing the recognition of the host receptor, resorting, for instance, to SA-mimic peptides;[21] and (3) preventing the fusion between viral and host cell membranes by impeding the dynamical changes that trigger the release of the FP.[22−24]

Inspection of the available X-ray structures reveals that fusion inhibitors bind to two sites in the stem region. One

Received: October 17, 2024
Revised: November 11, 2024
Accepted: November 21, 2024

ACS Publications
© XXXX The Authors. Published by American Chemical Society

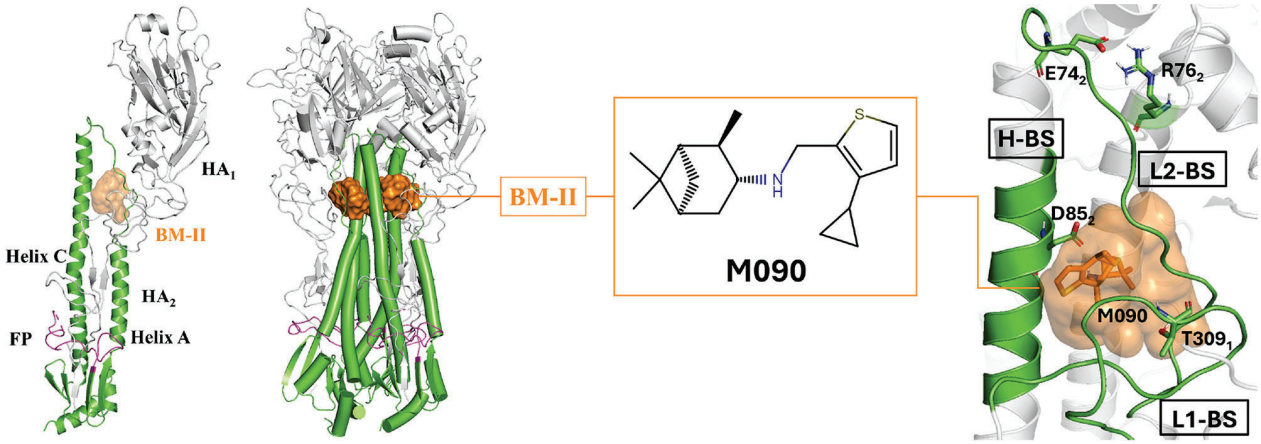


Figure 1. Structural representation of HA. The monomeric chains HA_1 (gray) and HA_2 (green) formed after proteolytic cleavage assemble, forming a trimer where the fusion peptide (FP) (magenta) remains buried until acidification in the endosome. The putative binding site of M090 (BM-II), which is shaped by an α-helix (H-BS) and two loops (L1-BS and L2-BS), is displayed as an orange surface that encloses M090 (shown as sticks).

corresponds to the pocket occupied by *tert*-butylhydroquinone (TBHQ) (PDB IDs 3EYK and 3EYM)[25] and arbidol (PDB IDs 5T6N and 5T6S),[26] consisting of residues located at helices A and C from protomer 1 and C′ from protomer 2. The other is the groove filled by JNJ4796 (PDB IDs 6CFG and 6CF7),[27] (*S*)-F0045 (PDB ID 6WCR),[28] and CBS1117 (PDB ID 6VMZ).[29] This information is valuable not only to explore the structure−activity relationships but also to assess the putative binding mode of novel compounds.

Given the structural complexity of HA and its conformational plasticity, the occurrence of novel druggable sites cannot be ruled out. Indeed, other sites have been suggested for a diverse set of compounds, such as *N*-benzyl-4,4-disubstituted piperidines[30] and pinanamine-based antivirals.[31,32] The most potent inhibitor of this latter series of compounds, M090 (Figure 1), showed an EC_{50} of 0.3 μM against H1N1 A/Guangzhou/GIRD/07/2009 virus-infected MDCK cells, EC_{50} values of 0.34 and 0.1 μM in a plaque reduction assay against H1N1 A/California/07/2009 and A/Texas/04/2009, respectively, and EC_{50} values of 1.5−6.9 μM in cytopathic assays against several strains (H1N1, H3N2, H7N3, and H9N2). Moreover, M090 exerts its inhibitory activity in the early steps of the virus life cycle by targeting HA, and resistance selection experiments revealed the occurrence of the $E74_2 \rightarrow D$ mutation in HA_2. Finally, molecular modeling studies reported by Zhao et al.[32] suggested that M090 binds at the interface defined by the α-helix $L80_2$−$A96_2$ (H-BS) and by loops $P301_1$−$K318_1$ and $M59_2$−$L73_2$ (L1-BS and L2-BS, respectively) (Figure 1). In this putative binding mode, the thiophene ring forms contacts with $I308_1$, $Y316_1$, and $W92_2$, the pinane moiety fills the region shaped by $A65_2$ and $V66_2$, and the cyclopropyl ring interacts with $F88_2$ and $P307_1$. In addition, the N−H group forms a hydrogen bond (HB) to $T309_1$.

In the absence of experimental structural evidence, this work aims to assess the proposed binding mode of M090 by synergistically combining the structural information on HA with computational studies—atomistic MD simulations (Table 1) and free energy calculations—and the pharmacological information gained from antiviral assays for M090-related compounds. The results are examined to gain insight into the effect of chemical changes on the inhibition potency and the impact of resistant-associated mutations.

Table 1. Summary of the Simulated Systems Used in This Work[a]

system	binding mode	code	starting structure	simulated time	ligand release
apo, wild type	–	Awt	homology model	2 μs	–
apo, E74→D mutant	–	Amut	snapshot at 1 μs of Awt	2 μs	–
holo, wild type	BM-I	H1wt-I	homology model	1 μs	no
		H2wt-I	snapshot at 1 μs of Awt	1 μs	1 out of 3
	BM-II	H1wt-II	homology model	2 μs	no
		H2wt-II	snapshot at 1 μs of Awt	1 μs	no
		H3wt-II	homology model	1 μs	no
		H4wt-II	snapshot at 2 μs of Awt	1 μs	no
holo, E74→D mutant	BM-II	H1mut-II	homology model	1 μs	no
		H2mut-II	snapshot at 2 μs of Awt	1 μs	no
holo, wild type	BM-R	H1wt-R	homology model	1 μs	1 out of 3
		H2wt-R	snapshot at 1 μs of Awt	0.5 μs	1 out of 3

[a]The representations of the three binding modes (BM-I, BM-II, and BM-R) are shown in Figure 2A1, Figure 2B1, and Figure 2C1, respectively.

To assess the structural stability of the binding mode proposed by Zhao et al. (denoted BM-I, Figure 2A1),[32] a complex formed by HA (H1N1 A/Virginia/ATCC3/2009) and three M090 molecules (stoichiometric ratio 1:3), each bound to one of the three sites in the trimer, was built up. This strategy was also adopted in previous studies,[33] thus enabling a check of the consistency of the binding mode of the ligand at the three pockets of HA in a single MD simulation. The initial pose of M090 in each pocket was manually adjusted to retain the key contacts described by Zhao et al.,[32] especially regarding the HB formed with $T309_1$. At this point, let us note that the

https://doi.org/10.1021/acsmedchemlett.4c00502
ACS Med. Chem. Lett. XXXX, XXX, XXX−XXX

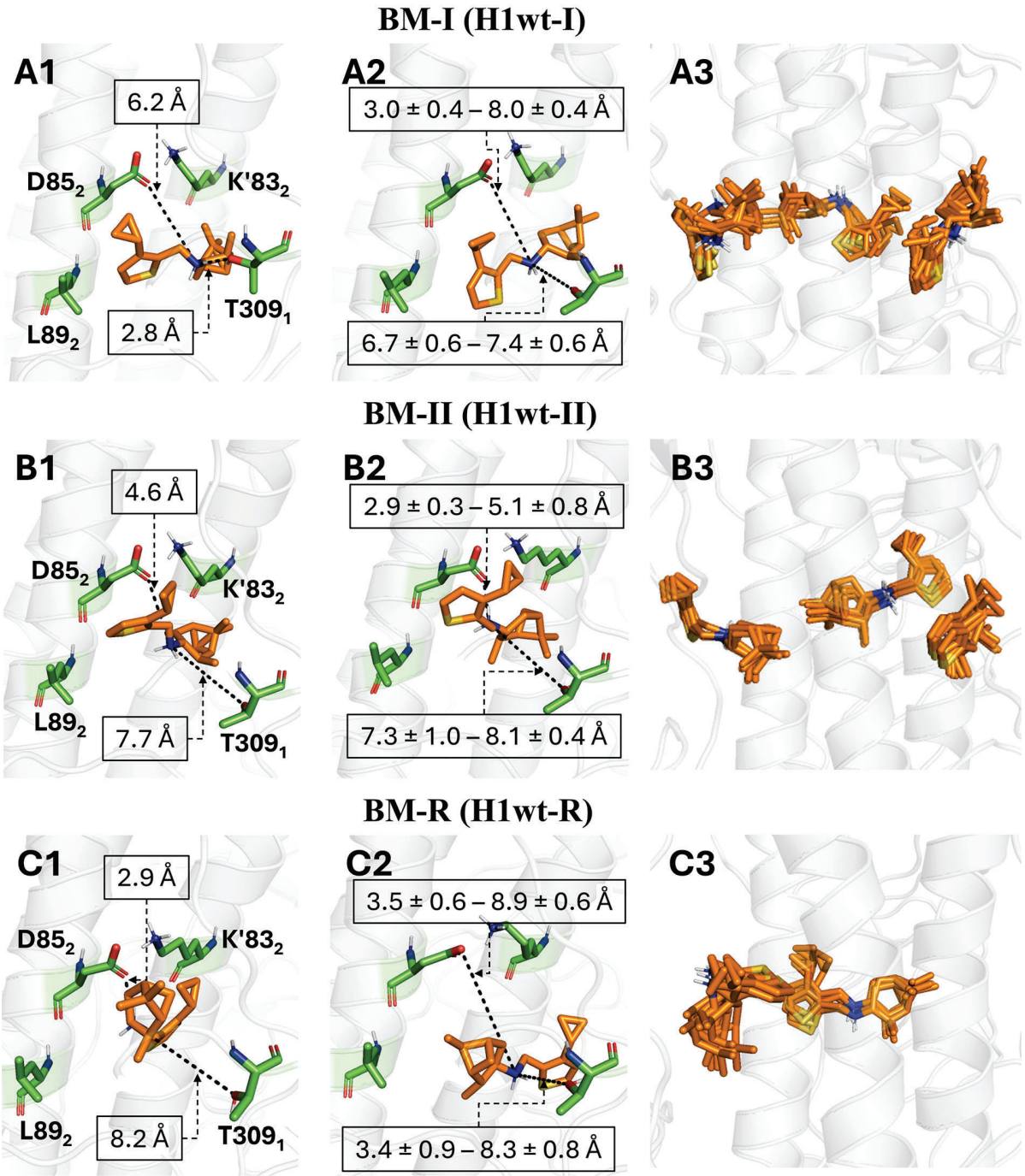


Figure 2. Representation of the three initial binding modes of M090 investigated in this study: (A1) the pose with the restrained HB between the amine nitrogen and T309$_1$; (B1) the unrestrained arrangement in the interior of the pocket; (C1) the reversed pose where the thiophene and pinanamine units have been exchanged relative to A1 and B1. Representative snapshots of the poses obtained at the end of the MD simulations with distances to D85$_2$ and T309$_1$ are shown in the central column (A2, B2, and C2). Finally, superpositions of the ligands taken every 100 ns along the last 500 ns of the MD simulation are displayed in the right column (A3, B3, C3). Note that only two pockets containing overlapped ligands are shown in C3 due to the release of M090 from the third binding pocket. Values denote the distance averaged for the three distinct M090 molecules bound to HA.

residues that shape the binding pocket in this strain are identical to those found in other H1N1 strains (see Homology Modeling, Figure S1, and Table S1 in the Supporting Information (SI)), such as A/Guangzhou/GIRD/07/2009 and A/PR/8/34.

Using this model system, two independent MD simulations (H1wt-I and H2wt-I, 1 μs) were run to examine the stability of the BM-I binding mode (see Molecular Dynamics Simulations in the SI). The N–H···O(T309$_1$) HB was restrained during equilibration and the first 200 ns of the MD production run to avoid artifactual changes due to structural relaxation during thermalization. The root-mean-square deviation (RMSD) profile determined for the two trajectories supports the structural stability of the binding site (RMSD values ranging

https://doi.org/10.1021/acsmedchemlett.4c00502
ACS Med. Chem. Lett. XXXX, XXX, XXX–XXX

from 1.6 to 2.2 Å for the six binding pockets in the two independent MD simulations; Figure S2). However, after release of the restraints, the HB between M090 and $T309_1$ was lost in all cases, and the ligand was slightly shifted in the interior of the pocket (RMSD of 2.3−4.9 Å; Figure S2). The new arrangement was stable in five out of six cases (Figure 2A2,A3), i.e., three equivalent binding modes on the HA:M090 (1:3 ratio) complex simulated in two independent MD simulations, with the amine nitrogen being roughly equidistant from residues $T309_1$ (distances of 6.7 ± 0.6 to 7.4 ± 0.6 Å) and $D85_2$, forming transient salt bridges with this latter residue (distances of 3.0 ± 0.4 to 8.0 ± 0.4 Å). In the remaining pocket, the ligand was released to the bulk solvent. This was facilitated by the displacement of the L2-BS loop, which enabled flipping of the ligand so that the pinanamine ring was facing the mouth of the pocket (shaped by H-BS and L1-BS) at the beginning of the trajectory, and the ligand was finally released at the end of the trajectory (ca. 840 ns).

Because of the previous results, an additional MD simulation (H1wt-II, 2 μs) was run for the HA:M090 (1:3 ratio) complex, where M090 was arranged to mimic the original binding mode (i.e., with the pinane moiety filling the innermost region). However, rather than imposing the HB with $T309_1$, the protonated amine was oriented to the carboxylate group of $D85_2$, thus defining the binding mode denoted as BM-II (Figure 2B1). The RMSD profiles confirmed the stability of the binding pocket (RMSD of 1.2−1.5 Å) and ligand (RMSD of 1.7−2.7 Å) (Figure S3). M090 adopted a similar pose in the three pockets (Figure 2B2,B3), forming a salt bridge between the amine nitrogen and $D85_2$ (distances of 3.1 ± 0.5 to 5.0 ± 0.7 Å for the three cases), which in turn forms an electrostatic interaction with $K'83_2$ (distances of 2.9 ± 0.3 to 3.0 ± 0.3 Å) in the H-BS of the neighboring protomer. Let us note that these two residues are preserved in all of the H1N1 strains (see Table S1). The thiophene ring forms van der Waals contacts with the side chains of $P307_1$, $I308_1$, and $L89_2$, and the pinane moiety fills the interior of the pocket through interactions with $F'88_2$, $A65_2$, and the methylene chain of $K'83_2$.

Further support for the BM-II binding mode came from three independent MD simulations (H2wt_II, H3wt-II, and H4wt-II, 1 μs), which were prepared using distinct starting structures (i.e., the homology model and the frames taken at 1 and 2 μs from a simulation of the apo species of HA; see Table 1). M090 was introduced in the center of the three binding sites, with the pinane moiety filling the innermost region and the thiophene ring close to the pocket mouth without imposing restraints with either $T309_1$ or $D85_2$ (Figure 2B1). The structural integrity of the binding pocket was maintained in all cases (RMSD of 0.8−2.0 Å; Figure S3). Furthermore, the binding mode of M090 was preserved in all of the pockets (RMSD of 1.0−3.6 Å; Figure S3). At this point, let us note that the changes observed in the RMSD profile of the ligand in some trajectories merely reflected reorientations of the pinanamine moiety through rotations around the N−C bond. Finally, the distance from the amine nitrogen to the $D85_2$ carboxylate oxygens ranged from 2.9 ± 0.3 to 5.1 ± 0.8 Å for the nine binding sites (three independent MD simulations, each with three molecules of M090 per HA trimer), which supports the electrostatic stabilization of the binding mode exerted by the interaction with $D85_2$.

As a final test, we explored the stability of the complex formed with M090 but adopted a reverse binding mode (denoted BM-R) that exchanges the pinane and thiophene moieties in the binding pocket (Figure 2C1). To this end, two additional MD simulations (H1wt-R, 1 μs; H2wt-R, 0.5 μs; Table 1) were run for the ligand in this reverse binding mode without imposing positional restraints. As in the preceding cases, the RMSD profiles supported the structural stability of the binding pocket (RMSD of 1.2−1.6 Å in the two independent MD simulations; Figure S4). After a slight rearrangement, the ligand remained stable in four out of the six cases, i.e., three equivalent binding modes on the HA:M090 (1:3 ratio) complex simulated in two independent MD simulations. However, a precise binding mode was not observed. Thus, in two cases, the protonated amine formed direct interactions with the amide carbonyl oxygen of $Q62_2$ (distance of 3.3 ± 1.2 Å; Figure S5), assisted by electrostatic interactions with $D'90_2$ (distance of 6.7 ± 1.8 Å). Alternatively, the protonated amine formed a salt bridge with $D85_2$ (distance of 3.7 ± 0.8 Å). Finally, in two cases, M090 was released from the binding site, arguing against the feasibility of the reversed binding mode.

The preceding results suggest that M090 adopts a stable binding mode when the pinanamine moiety fills the innermost region of the pocket, as suggested by Zhao et al.,[32] but with the amine nitrogen forming an electrostatic interaction with $D85_2$. To further validate the BM-II binding mode predicted for M090, we planned to perform a series of alchemical transformations taking advantage of the inhibitory potencies reported for a series of M090-related compounds, which affect both the nature of the thiophene ring and the attached cyclopropyl chain (Figure 3A).[32]

Thermodynamic integration (TI) calculations (see Alchemical Calculations in the SI)[34−37] were carried out to estimate the differences in binding free energy between M090 and its structural analogues. To this end, the free energy changes were determined for the alchemical transformations between ligands bound to HA relative to the same process in an aqueous solution. The changes in the binding free energy were compared with the experimental values, which were approximated from the EC_{50} values reported by Zhao et al.[32] using the Cheng−Prusoff equation.[38] With this aim, the inhibitory potency measured from a consistent set of virus (A/Guangzhou/GIRD/07/2009)-infected MDCK cell assays, leading to EC_{50} values ranging from 0.30 μM (M090) to 75 μM (M10; see below), were used.[32] Let us note that this choice may introduce some discrepancies between the results derived from *in silico* calculations and cell-based assays due to the potential effect of factors unrelated to the binding of the compounds to the putative site in HA.

The free energy changes predicted for the alchemical transformations are shown in Figure 3A. It is worth noting the consistency between the free energy changes determined for the "forward" and "backward" processes for each transformation. Furthermore, the addition of the free energy changes for the combination of processes that define the closure of the thermodynamic cycles is close to zero. Finally, one may also notice the consistency between the free energy differences obtained following distinct pathways that connect M090, which was taken as a reference, to other derivatives relative to M090 (see Table 2). The only exception to this behavior is compound M24, which will be discussed below.

Figure 3B shows the comparison between the differences in the binding free energy estimated from experimental data and the values predicted from TI calculations. The results reveal that there is a close correspondence between predicted and

https://doi.org/10.1021/acsmedchemlett.4c00502
ACS Med. Chem. Lett. XXXX, XXX, XXX−XXX

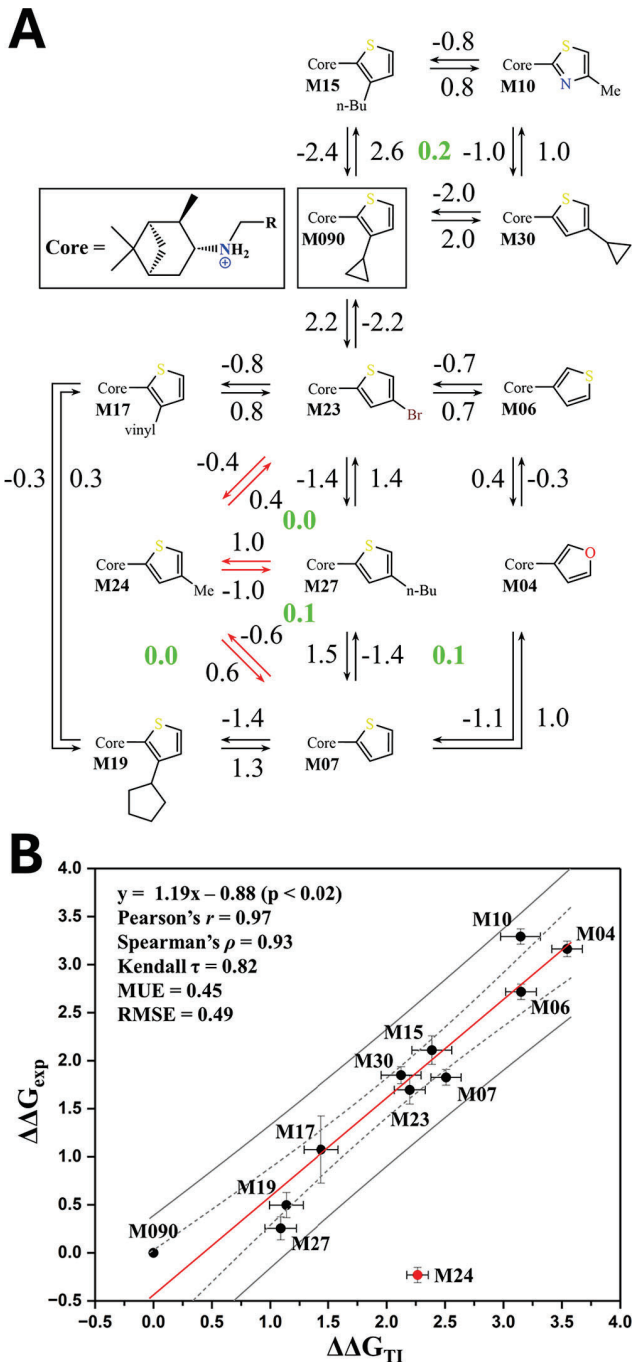


Figure 3. (A) Representation of the free energy changes determined from TI calculations for the alchemical (forward and backward) transformations between M090-related derivatives. The numbering of the derivatives follows the notation used by Zhao et al. in ref 32. Green values denote the sum of the free energy changes for the processes implicated in the closure of the thermodynamic cycle. (B) Comparison between the differences in binding affinity estimated from the experimental EC_{50} data and the relative free energy differences predicted from TI calculations. Bars denote the errors of calculated and experimental relative free energy changes. Dotted and dashed lines denote the 95% confidence and prediction intervals. All values (kcal mol^{-1}) are relative to M090. MUE and RMSE stand for mean unsigned error and root-mean-square error, respectively.

experimental values, which is remarkable keeping in mind the potential uncertainties arising from the virus-infected MDCK cell assays. Likewise, the error between predicted and experimentally derived values amounts to 0.5 kcal/mol.

The TI results point out the relevance of the 3-cyclopropylthiophene unit to the inhibitory activity. This may be attributed to a hydrophobic shielding effect,[39−42] which prevents the access of water molecules to the interior of the binding pocket, where they could compete with the salt bridge formed between the amine nitrogen and the carboxylate group of $D85_2$. In this context, replacement of the cyclopropyl unit by a linear *n*-butyl chain (compound M15; EC_{50} of 10.33 μM) leads to a 33-fold reduction in the inhibitory potency, which reflects the enhanced steric clashes with $G67_2$ at the L2-BS loop (Figure 4A), together with a larger conformational penalty of the flexible butyl chain compared to the more rigid cyclopropyl unit, which only has one rotatable bond. In contrast, the inhibitory potency of M090 remains almost unaltered upon translation of the *n*-butyl chain from position 3 (M15) to 4 (M27; EC_{50} of 0.46 μM), as this reorients the aliphatic chain toward the mouth of the pocket, adopting a more extended conformation. This permits the avoidance of the steric clashes with L2-BS and favors the formation of van der Waals contacts with $P307_1$ and $I308_1$ in L1-BS and the methylene chain of $K82_2$ in H-BS (Figure 4B). Similarly, elimination of the alkyl chain (M07; EC_{50} of 6.42 μM) reduces the hydrophobic shielding, which would justify the 21-fold decrease in potency.

The results also point out the feasibility of accommodating the cyclopentyl unit (M19; EC_{50} of 0.69 μM) instead of the cyclopropyl one, which agrees with the ∼2-fold reduction in inhibitory potency (Figure 4C). Finally, the relevance of the cyclopropyl chain is also reflected in the loss of potency generated by changing the linkage between the pinanamine and thiophene moieties from position 2 in M090 to position 3 in M30 (EC_{50} of 6.68 μM) and even upon deletion of the cyclopropyl unit in M06 and M07 (EC_{50} of 28.60 and 6.42 μM, respectively), thus suppressing the hydrophobic shielding due to the larger accessibility of waters to the interior of the pocket (Figure 4D). This also justifies the drastic reduction in potency observed upon replacement of the thiophene ring by more polar moieties, such as furan and thiazole, leading to 200-fold (M04; EC_{50} of 60.46 μM) and 250-fold (M10; EC_{50} of 75 μM) reductions in inhibitory potency, respectively.

The only exception to this behavior is compound M24 (red dot in Figure 3B), which is predicted to be much less active compared to the experimental EC_{50} (0.20 μM). As noted in Figure 3A, consistent trends were obtained in alchemical transformations leading from M23, M27, and M07 to M24. For instance, shortening of the *n*-butyl chain (M27) to methyl (M24) should destabilize the hydrophobic shielding effect, whereas insertion of methyl in M07 to generate M24 should increase the shielding effect. We speculate that the activity of M24, which is slightly more potent than M090, may reflect the ability to interact with another viral target, particularly the M2 proton channel.[43] In this regard, one may notice the close hydrophobic/philic-guided overlap determined with Pharmscreen[44] between M24 and *N*-[(5-bromothiophen-2-yl)methyl]adamantan-1-amine,[45] a compound that showed ∼77% inhibition of the wild-type M2 channel and its S31N mutant (EC_{50} of 1.8 mM against the A/WSN/33 (S31N) strain).[46] Indeed, superposition of the overlapped molecules in the X-ray crystallographic structure of the complex formed by this compound with the S31N mutant (PDB ID 2MUV) reflects a fine accommodation in the pore of the channel

https://doi.org/10.1021/acsmedchemlett.4c00502
ACS Med. Chem. Lett. XXXX, XXX, XXX−XXX

Table 2. Free Energy Differences ($\Delta\Delta G$, in kcal·mol^{-1}) in of the Set of Pinanamine-Based Compounds Relative to M090 (Reference Compound[a])

compound (EC_{50})[b]	path	$\Delta\Delta G_{TI}$[c]
M04 (60.46 ± 1.75)	M090 → M23 → M06 → M04	3.3
	M090 → M23 → M17 → M19 → M07 → M04	3.5
	M090 → M23 → M27 → M07 → M04	3.2
		3.3
M06 (28.60 ± 0.24)	M090 → M23 → M06	2.9
	M090 → M23 → M17 → M19 → M07 → M04 → M06	3.1
	M090 → M23 → M27 → M04 → M06	**2.9**
		2.9
M07 (6.42 ± 0.18)	M090 → M23 → M06 → M04 → M07	2.3
	M090 → M23 → M17 → M19 → M07	2.5
	M090 → M23 → M27 → M07	2.2
		2.3
M10 (75.00 ± 0.08)	M090 → M30 → M10	3.0
	M090 → M15 → M10	3.3
		3.1
M15 (10.33 ± 2.17)	M090 → M15	2.5
	M090 → M30 → M10 → M15	2.3
		2.4
M17 (1.82 ± 1.04)	M090 → M23 → M17	1.4
	M090 → M23 → M06 → M04 → M07 → M19 → M17	1.2
	M090 → M23 → M27 → M07 → M19 → M17	1.1
		1.2
M19 (0.69 ± 0.12)	M090 → M23 → M17 → M19	1.1
	M090 → M23 → M06 → M04 → M07 → M19	0.9
	M090 → M23 → M27 → M07 → M19	0.8
		0.9
M23 (5.16 ± 1.08)	M090 → M23	2.2
M24 (0.20)	M090 → M23 → M27 → M24	1.8
	M090 → M23 → M06 → M04 → M07 → M24	2.9
	M090 → M23 → M06 → M04 → M07 → M27 → M24	3.3
	M090 → M23 → M17 → M19 → M07 → M24	1.8
	M090 → M23 → M17 → M19 → M07 → M27 → M24	1.9
		1.8
M27 (0.46 ± 0.07)	M090 → M23 → M06 → M04 → M07 → M27	0.9
	M090 → M23 → M17 → M19 → M07 → M27	1.0
	M090 → M23 → M27	0.8
		0.9
M30 (6.68 ± 0.42)	M090 → M30	2.0
	M090 → M15 → M10 → M30	2.3
		2.1

[a]EC_{50} = 0.30 ± 0.04 μM. [b]Experimental EC_{50} values (μM) taken from ref 32. [c]Values averaged for distinct pathways are highlighted in bold.

(Figure 5). However, severe steric clashes can be anticipated upon attachment of the cyclopropyl unit at position 3 of the thiophene ring in M090 (Figure 5), which agrees with the lack of inhibition against the M2(S31N) channel reported by Zhao et al.[32]

Finally, to further assess the feasibility of the binding site proposed for M090, we examined the effect of the drug-resistant mutation $E74_2$ → D disclosed from resistance-induction assays using the A/Guangzhou/GIRD/07/2009 (H1N1) virus in MDCK cells.[32] It is noteworthy that this is the sole mutation found in the mature M090-resistant virus strain. $E74_2$ is placed at the top of the α-helix (H-BS) of the HA_2 domain, where it forms a salt bridge with $R76_2$ (Figure 1). Although $E74_2$ is distant from the binding pocket, as noted in the distances from the Cα atom of $E74_2$ to the Cα atoms of $D85_2$ and T3091 (~17 and 24 Å, respectively; see Table S2), the salt bridge is located at the hinge between H-BS and L2-BS, and hence, its stability may affect the structural flexibility of the binding pocket.

To check the impact of the $E74_2$ → D mutation, MD simulations were run for the wild type and mutated variants of apo HA (Awt and Amut, 2 μs; Table 1). The RMSD profiles support the structural stability of the binding pocket (RMSD values ranging from 1.1 to 2.4 Å; Figure S8). As expected, the salt bridge between $E74_2$ and $R76_2$ is almost fully maintained in the wild type species (average population of 96%), reflecting primarily a bidentate interaction between the carboxylate oxygens and guanidinium NH groups. However, its population is reduced to ~50% in the mutated species. Similar trends were observed in the comparison of MD simulations run for the wild type holo species (H1wt-II, H2wt-II, H3wt-II, and H4wt-II) and two MD simulations run by placing M090 in the three pockets of the mutated HA preserving the binding pose BM-II (H1mut-II and H2mut-II, 1 μs; Table 1; see Figure S9 for the

https://doi.org/10.1021/acsmedchemlett.4c00502
ACS Med. Chem. Lett. XXXX, XXX, XXX−XXX

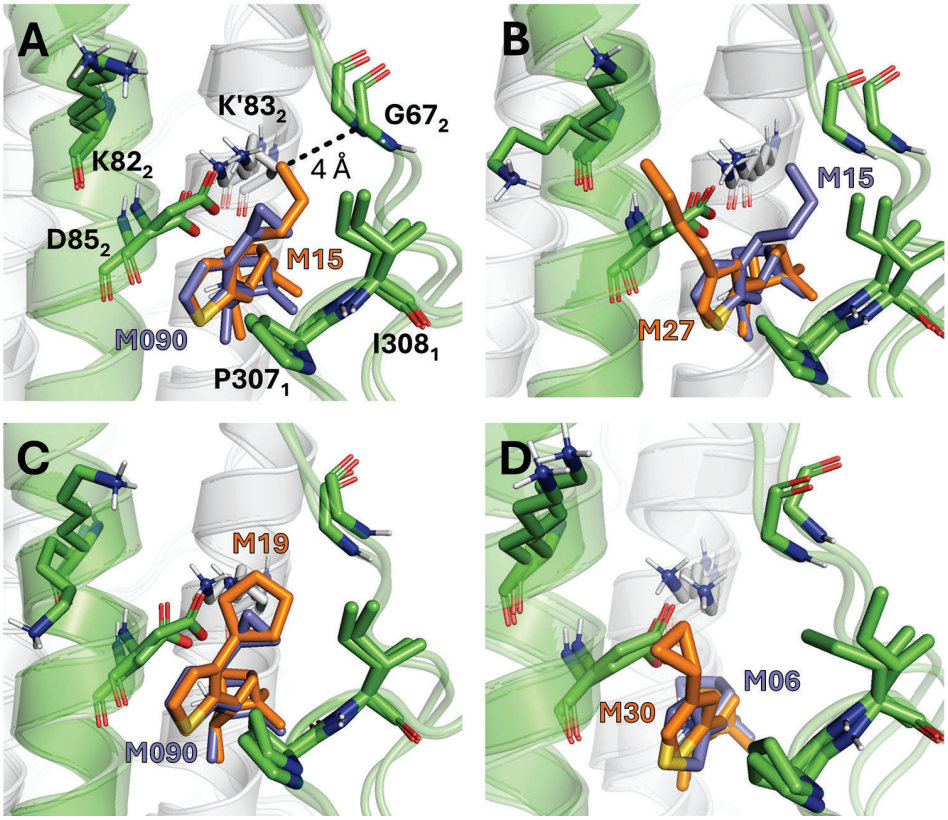


Figure 4. Representative snapshots of the binding pose obtained for compounds (A) M15, (B) M27, (C) M19, and (D) M30 (shown as sticks) at the end of the alchemical transformations. (A) The *n*-butyl chain attached to position 3 of the thiophene unit adopts a folded conformation due to the steric clash with loop L2-BS. (B) Superposition of the complexes with *n*-butyl derivatives M15 (bound to position 3) and M27 (attached to position 4). The steric clash observed for M15 with L2-BS is alleviated in M27, where the *n*-butyl chain adopts an extended conformation and forms van der Waals contacts with P307$_1$, I308$_1$, and the side chain of K82$_2$. (C, D) The cyclopentyl unit in M19 occludes the access of water molecules, thus creating hydrophobic shielding of the electrostatic interaction between D85$_2$ and K83$_2$ in the neighboring protomer, mimicking the protecting effect of the cyclopropyl unit in M090. This effect is lost in compounds M04, M06, and M07, which have unsubstituted furan and thiophene rings.

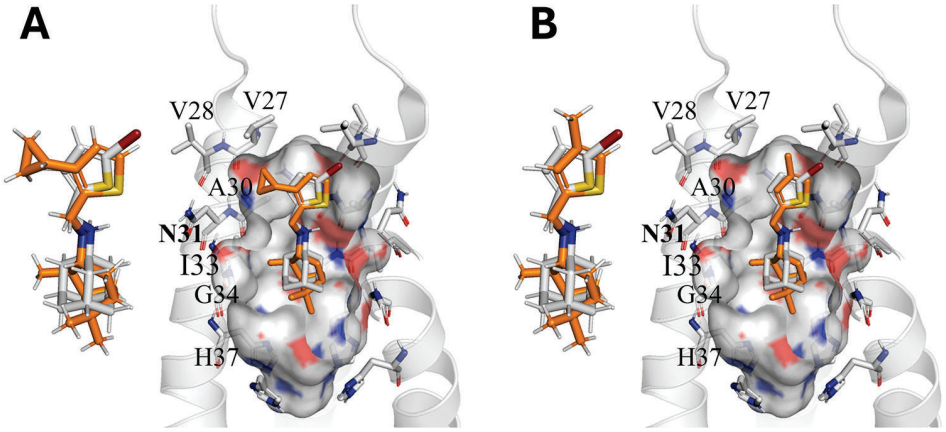


Figure 5. Overlaid structures of (A) M090 and (B) M24 (orange sticks) with *N*-[(5-bromothiophen-2-yl)methyl]adamantan-1-amine (gray sticks) and superposition of the overlaid compounds onto the X-ray crystallographic structure of the S31N mutated variant of the M2 proton channel (PDB ID 2MUV).[46]

RMSD profiles of the holo complexes run for the mutated HA). The salt bridge between E74$_2$ and R76$_2$ is observed in 87−95% of the four trajectories run for the wild type M090-bound HA. However, this interaction is reduced to 46% and 54% in the two independent MD simulations of the mutated holo species. It is noteworthy that the electrostatic destabilization caused by the E74$_2$ → D mutation also affects secondary interactions, such as the contact between R76$_2$ and E69$_2$, which is located in L2-BS. Thus, this contact is found, on average, in 66% of the snapshots analyzed for the apo and holo wild type simulations, but it is reduced to 41% in the mutated strain.

https://doi.org/10.1021/acsmedchemlett.4c00502
ACS Med. Chem. Lett. XXXX, XXX, XXX−XXX

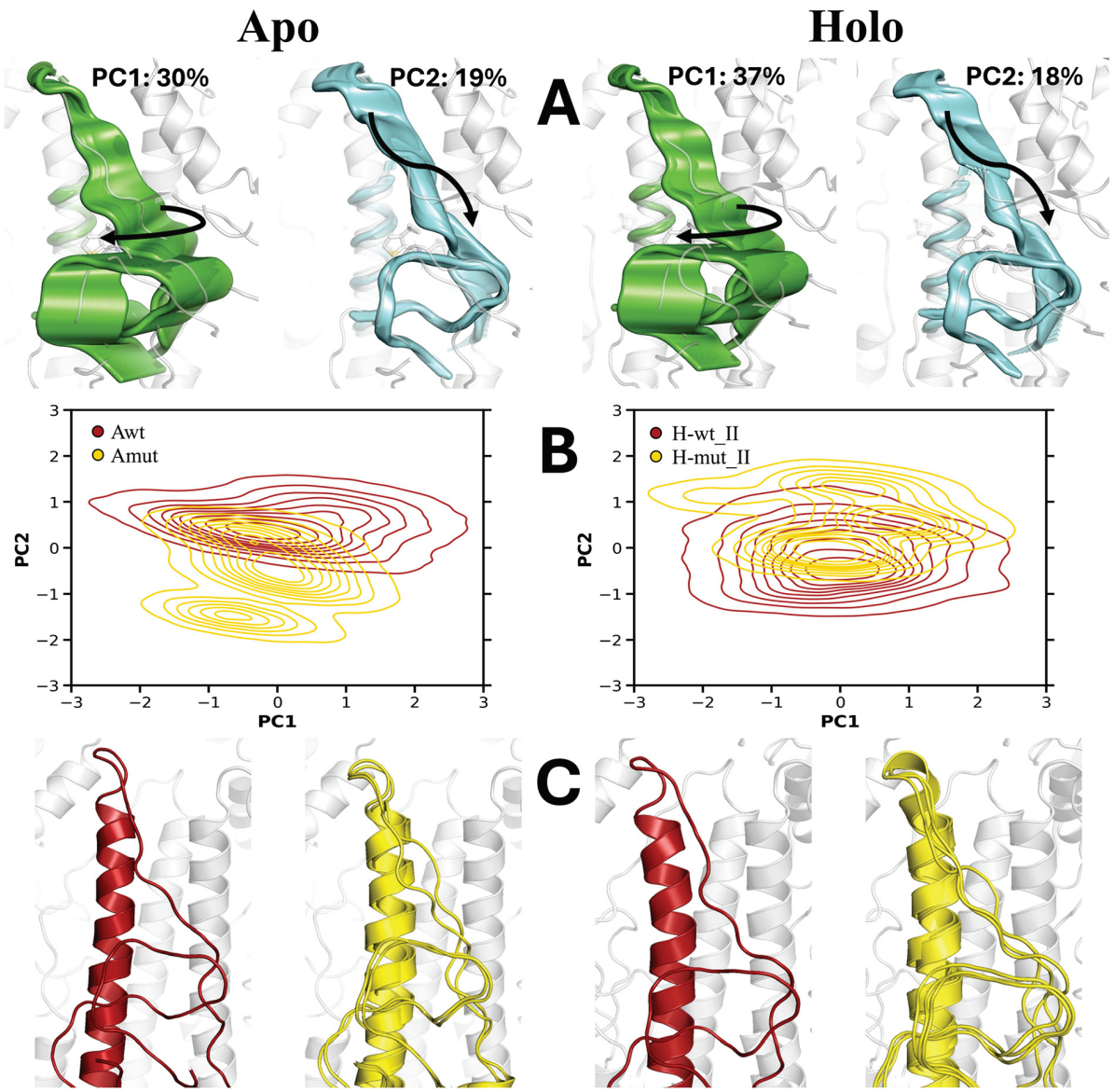


Figure 6. (A) Representation of the two main essential motions of the backbone atoms that shape the binding pocket for the (left) apo and (right) holo species of wild type and $E74_2 \rightarrow D$ mutated HA. (B) Projections of the snapshots sampled for the apo wild type HA (red) and its mutated variant (yellow). (C) Centroids of the clusters observed for the apo wild type and mutated HA (backbone of the binding pocket in red and yellow, respectively).

Principal component (PC) analysis was used to examine the impact of the $E74_2 \rightarrow D$ mutation on the motions of the backbone atoms of H-BS, L1-BS, and L2-BS in the apo wild type and mutated HA. The first two PCs explain ∼50% of the total variance (the contributions of higher-order components are lower than 7%). The first PC accounts for 30% of the structural variance and describes a rotational motion of H-BS, L1-BS, and L2-BS, following the C_3 symmetry axis of HA (Figure 6A). In contrast, the second PC (19%) describes a conformational remodeling of L2-BS. In the case of the holo species, the two main essential motions (37% and 18% of the structural variance) exhibit a large resemblance to corresponding motions found for the apo species (Figure 6A).

The projections of the snapshots collected from the simulations of wild type HA and the mutated variant reveal that the main difference involves the second PC (Figure 6B). Indeed, whereas a single cluster is observed for the wild type HA in both the apo and holo forms, two (apo) and three (holo) clusters are found in the projection of the snapshots sampled for the mutated HA (Figure 6C). This reflects that the $E74_2 \rightarrow D$ mutation induces a larger conformational flexibility in the L2-BS loop, which should alter the packing against H-BS and L1-BS and weaken the proper accommodation of M090 in the binding pocket. Furthermore, we cannot rule out larger spatial rearrangements on longer time scales, especially upon acidification.

Overall, this study provides support for the involvement of the putative binding site proposed by Zhao et al.[32] for the binding of M090. MD simulations reveal a stable binding pose characterized by the insertion of the pinanamine moiety in the innermost region of the pocket, whereas the 3-cylopropylthiophene ring occupies the mouth of the pocket, occluding the accessibility of water molecules. Instead of the HB formed between the amine nitrogen and $T309_1$, our results support the

https://doi.org/10.1021/acsmedchemlett.4c00502
ACS Med. Chem. Lett. XXXX, XXX, XXX−XXX

electrostatic stabilization of the protonated nitrogen with $D85_2$, a residue conserved in all the H1N1 strains considered in the experimental assays. Furthermore, this binding mode provides a structural basis to justify the effect of the chemical changes introduced in both the cyclopropyl and thiophene units, since the alchemical transformations discussed here generally exhibit a close correspondence with the differences in inhibitory potency determined in antiviral assays. Finally, as hypothesized by Zhao et al.,[32] the $E74_2 \rightarrow D$ mutation has a negative impact on the stability of the salt bridge formed with $R76_2$, which in turn also affects other secondary interactions, leading ultimately to an increase in the structural flexibility of the L2-BS loop. Taken together, these results support the occurrence of a binding pocket alternative to sites A and B in HA (Figure 1), which may be valuable to explore the fusion inhibition by other compounds with an unknown mechanism of action.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsmedchemlett.4c00502.

Details of MD simulations, thermodynamic integration calculations, and supplementary figures (PDF)

■ AUTHOR INFORMATION

Corresponding Author

F. Javier Luque − *Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació - Campus Torribera, Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain; Institut de Biomedicina (IBUB) and Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, 08028 Barcelona, Spain;* orcid.org/0000-0002-8049-3567; Email: fjluque@ub.edu

Authors

Aitor Valdivia − *Doctorate in Biotechnology, Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació - Campus Torribera, Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain; Institut de Biomedicina (IBUB), Universitat de Barcelona, 08028 Barcelona, Spain;* orcid.org/0000-0003-4862-0850

Maria Rocha − *Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació - Campus Torribera, Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain; Department of Life Sciences, University of Coímbra, 3000-456 Coímbra, Portugal*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsmedchemlett.4c00502

Author Contributions

The manuscript was written through contributions of all authors. All of the authors approved the final version of the manuscript.

Notes

Safety statement: No safety hazards were encountered.
The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to Dr. S. Llabrés and Dr. J. Vázquez for technical assistance and valuable suggestions. The authors acknowledge the financial support from the Spanish Ministerio de Ciencia e Innovación (PID2020-117646RB-I00 MCIN/AEI/10.13039/501100011033; Maria de Maetzu CEX2021-001202-M) and Generalitat de Catalunya (2021SGR00671) and the access to supercomputation resources provided by the Consorci de Serveis Universitaris de Catalunya (CSUC) (Molecular Recognition Project).

■ ABBREVIATIONS

FP, fusion peptide; HA, hemagglutinin; HB, hydrogen bond; MD, molecular dynamics; PC, principal component; RMSD, root-mean-square deviation; SA, syalic acid; TI, thermodynamic integration

■ REFERENCES

(1) Wakefield, A. E.; Kozakov, D.; Vajda, S. Mapping the binding sites of challenging drug targets. *Curr. Opin. Struct. Biol.* **2022**, *75*, 102396.

(2) Feng, T.; Barakat, K. Molecular dynamics simulations and predictions of druggable binding sites. *Methods Mol. Biol.* **2018**, *1762*, 87−103.

(3) Luque, I.; Freire, E. Structural stability of binding sites: Consequences for binding affinity. *Proteins* **2000**, *41* (S4), 63−71.

(4) Cozzini, P.; Kellogg, G. E.; Spyrakis, F.; Abraham, D. J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L. A.; Morris, G. M.; Orozco, M.; Pertinhez, T. A.; Rizzi, M.; Sotriffer, C. A. Target flexibility: An emerging consideration in drug discovery and design. *J. Med. Chem.* **2008**, *51*, 6237−6255.

(5) Stank, A.; Kokh, D. B.; Fuller, J. C.; Wade, R. C. Protein binding pocket dynamics. *Acc. Chem. Res.* **2016**, *49*, 809−815.

(6) Feixas, F.; Lindert, S.; Sinko, W.; McCammon, J. A. Exploring the role of receptor flexibility in structure-based drug discovery. *Biophys. Chem.* **2014**, *186*, 31−45.

(7) Spyrakis, F.; Cavasotto, C. N. Open challenges in structure-based virtual screening: Receptor modeling, target flexibility consideration and active site water molecules description. *Arch. Biochem. Biophys.* **2015**, *583*, 105−119.

(8) Di Lella, S.; Herrmann, A.; Mair, C. M. Modulation of the pH stability of influenza virus hemagglutinin: A host cell adaptation strategy. *Biophys. J.* **2016**, *110*, 2293−2301.

(9) Benton, D. J.; Gamblin, S. J.; Rosenthal, P. B.; Skehel, J. J. Structural transitions in influenza haemagglutinin at membrane fusion pH. *Nature* **2020**, *583*, 150−153.

(10) Garcia-Moro, E.; Zhang, J.; Calder, L. J.; Brown, N. R.; Gamblin, S. J.; Skehel, J. J.; Rosenthal, P. B. Reversible structural changes in the influenza hemagglutinin precursor at membrane fusion pH. *Proc. Natl. Acad. Sci. U.S.A.* **2022**, *119*, No. e2208011119.

(11) Lu, W.; Du, W.; Somovilla, V. J.; Yu, G.; Haksar, D.; de Vries, E.; Boons, G.-J.; de Vries, R. P.; de Haan, C. A. M.; Pieters, R. J. Enhanced inhibition of influenza A virus adhesion by di- and trivalent hemagglutinin inhibitors. *J. Med. Chem.* **2019**, *62*, 6398−6404.

(12) Guan, S.; Wang, T.; Kuai, Z.; Qian, M.; Tian, X.; Zhang, X.; Yu, Y.; Wang, S.; Zhang, H.; Li, H.; Kong, W.; Shan, Y. Exploration and binding inhibition mechanism of a small molecule inhibitor of influenza virus H1N1 hemagglutinin by molecular dynamics simulation. *Sci. Rep.* **2017**, *7*, 3786.

(13) Yanagita, H.; Yamamoto, N.; Fuji, H.; Liu, X.; Ogata, M.; Yokota, M.; Takaku, H.; Hasegawa, H.; Odagiri, T.; Tashiro, M.; Hoshino, T. Mechanism of drug resitance of hemagglutinin of influenza virus and potent scaffolds inhibiting its function. *ACS Chem. Biol.* **2012**, *7*, 552−562.

(14) Das, D. K.; Govindan, R.; Nikic-Spiegel, I.; Krammer, F.; Lemke, E. A.; Munro, J. B. Direct visualization of the conformational

dynamics of single influenze hemagglutinin trimers. *Cell* **2018**, *174*, 926−937.
(15) Rossman, J. S.; Lamb, R. A. Influenza virus assembly and budding. *Virology* **2011**, *411*, 229−236.
(16) York, A.; Fodor, E. Biogenesis, assembly, and export of viral messenger ribonucleoproteins in the influenza A virus infected cell. *RNA Biol.* **2013**, *10*, 1274−1282.
(17) Laporte, M.; Naesens, L. Airway proteases: An emerging drug target for influenza and other respiratory virus infections. *Curr. Opin. Virol.* **2017**, *24*, 16−24.
(18) Hermoso-Pinilla, F. J.; Valdivia, A.; Camarasa, M. J.; Ginex, T.; Luque, F. J. Influenza A virus hemagglutinin: From classical fusion inhibitors to proteolysis targeting chimera-based strategies in antiviral drug discovery. *Explor. Drug Sci.* **2024**, *2*, 85−116.
(19) Kumari, R.; Sharma, S. D.; Kumar, A.; Ende, Z.; Mishina, M.; Wang, Y.; Falls, Z.; Samudrala, R.; Pohl, J.; Knight, P. R.; Sambhara, S. Antiviral approaches against influenza virus. *Clin. Microbiol. Rev.* **2023**, *36*, No. e00040-22.
(20) Rossignol, J. F.; La Frazia, S.; Chiappa, L.; Ciucci, A.; Santoro, M. G. Thiazolides, a new class of anti-influenza molecules targeting viral hemagglutinin at the post-translational level. *J. Biol. Chem.* **2009**, *284*, 29798−29808.
(21) Zhang, Q.; Liang, T.; Nandakumar, K. S.; Liu, S. Emerging and state of the art hemagglutinin-targeted influenza virus inhibitors. *Expert Opin. Pharmacother.* **2021**, *22*, 715−728.
(22) Chen, Z.; Cui, Q.; Caffrey, M.; Rong, L.; Du, R. Small molecule inhibitors of influenza virus entry. *Pharmaceuticals (Basel).* **2021**, *14*, 587.
(23) Ginex, T.; Luque, F. J. Searching for effective antiviral small molecules against influenza A virus: a patent review. *Expert Opin Ther Pat.* **2021**, *31*, 53−66.
(24) Liu, H. Y.; Yang, P. L. Small-molecule inhibition of viral fusion glycoproteins. *Annu. Rev. Virol.* **2021**, *8*, 459−489.
(25) Russell, R. J.; Kerry, P. S.; Stevens, D. J.; Steinhauer, D. A.; Martin, S. R.; Gamblin, S. J.; Skehel, J. J. Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 17736−17741.
(26) Kadam, R. U.; Wilson, I. A. Structural basis of influenza virus fusion inhibition by the antiviral drug Arbidol. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 206−214.
(27) van Dongen, M. J. P.; Kadam, R. U.; Juraszek, J.; Lawson, E.; Brandenburg, B.; Schmitz, F.; Schepens, W. B. G.; Stoops, B.; van Diepen, H. A.; Jongeneelen, M.; Tang, C.; Vermond, J.; van Eijgen-Obregoso Real, A.; Blokland, S.; Garg, D.; Yu, W.; Goutier, W.; Lanckacker, E.; Klap, J. M.; Peeters, D. C. G.; Wu, J.; Buyck, C.; Jonckers, T. H. M.; Roymans, D.; Roevens, P.; Vogels, R.; Koudstaal, W.; Friesen, R. H. E.; Raboisson, P.; Dhanak, D.; Goudsmit, J.; Wilson, I. A. A small-molecule fusion inhibitor of influenza virus is orally active in mice. *Science* **2019**, *363*, eaar6221.
(28) Yao, Y.; Kadam, R. U.; Lee, C. D.; Woehl, J. L.; Wu, N. C.; Zhu, X.; Kitamura, S.; Wilson, I. A.; Wolan, D. W. An influenza A hemagglutinin small-molecule fusion inhibitor identified by a new high-throughput fluorescence polarization screen. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117*, 18431−18438.
(29) Antanasijevic, A.; Durst, M. A.; Cheng, H.; Gaisina, I. N.; Perez, J. T.; Manicassamy, B.; Rong, L.; Lavie, A.; Caffrey, M. Structure of avian influenza hemagglutinin in complex with a small molecule entry inhibitor. *Life Sci. Alliance* **2020**, *3*, No. e202000724.
(30) De Castro, S.; Ginex, T.; Vanderlinden, E.; Laporte, M.; Stevaert, A.; Cumella, J.; Gago, F.; Camarasa, M. J.; Luque, F. J.; naesens, L.; Velázquez, S. *N*-benzyl 4,4-disubstituted piperidines as a potent class of influenza H1N1 virus inhibitors showing a novel mechanism of hemagglutinin fusion peptide interaction. *Eur. J. Med. Chem.* **2020**, *194*, 112223.
(31) Hu, W.; Zeng, S.; Li, C.; Jie, Y.; Li, Z.; Chen, L. Identification of hits as matrix-2 protein inhibitors through the focused screening of a small primary amine library. *J. Med. Chem.* **2010**, *53*, 3831−3834.
(32) Zhao, X.; Li, R.; Zhou, Y.; Xiao, M.; Ma, C.; Yang, Z.; Zeng, S.; Du, Q.; Yang, C.; Jiang, H.; Hu, Y.; Wang, K.; Mok, C. K. P.; Sun, P.; Dong, J.; Cui, W.; Wang, J.; Tu, Y.; Yang, Z.; Hu, W. Discovery of highly potent pinanamine-based inhibitors against amantadine- and oseltamivir-resistant influenza a viruses. *J. Med. Chem.* **2018**, *61*, 5187−5198.
(33) Leiva, R.; Barniol-Xicota, M.; Codony, S.; Ginex, T.; Vanderlinden, E.; Montes, M.; Caffrey, M.; Luque, F. J.; Naesens, L.; Vázquez, S. Aniline-based inhibitors of influenza H1N1 virus acting on hemagglutinin-mediated fusion. *J. Med. Chem.* **2018**, *61*, 98−118.
(34) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813−1818.
(35) Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous free energy simulations in virtual screening. *J. Chem. Inf. Model.* **2020**, *60*, 4153−4169.
(36) Tsai, H.-C.; Tao, Y.; Lee, T.-S.; Merz, K. M., Jr; York, D. M. Validation of free energy methods in AMBER. *J. Chem. Inf. Model.* **2020**, *60*, 5296−5300.
(37) Mey, A. S. J. S.; Allen, B. K.; Bruce Macdonald, H. E.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best practices for alchemical free energy calculations. *Living J. Comput. Mol. Sci.* **2020**, *2*, 18378.
(38) Cheng, Y.-C.; Prusoff, W. H. Relationship between the inhibition constant (K_1) and the concentration of inhibitor which causes 50% inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.* **1973**, *22*, 3099−3108.
(39) Schmidtke, P.; Luque, F. J.; Murray, J. B.; Barril, X. Shielded hydrogen bonds as structural determinants of binding kinetics: Application in drug design. *J. Am. Chem. Soc.* **2011**, *133*, 18903−18910.
(40) Ruiz-Carmona, S.; Schmidtke, P.; Luque, F. J.; Baker, L.; Matassova, N.; Davis, B.; Roughley, S.; Murray, J.; Hubbard, R.; Barril, X. Dynamic undocking and the quasi-bound state as tools for drug discovery. *Nat. Chem.* **2017**, *9*, 201−206.
(41) Ferenczy, G. G.; Kellermayer, M. Contribution of hydrophobic interactions to protein mechanical stability. *Comput. Struct. Biotechnol. J.* **2022**, *20*, 1946−1956.
(42) Miñarro-Lleonar, M.; Bertran-Mostazo, A.; Duro, J.; Barril, X.; Juárez-Jiménez, J. Lenalidomide stabilizes protein-protein complexes by turning labile intermolecular H-bonds into robust interactions. *J. Med. Chem.* **2023**, *66*, 6037−6046.
(43) Aledavood, E.; Selmi, B.; Estarellas, C.; Masetti, M.; Luque, F. J. From acid activation mechanisms of proton conduction to design of inhibitors of the M2 proton channel of influenza A virus.. Contribution of hydrophobic interactions to protein mechanical stability. *Front. Mol. Biol.* **2022**, *8*, 796229.
(44) Vázquez, J.; Deplano, A.; Herrero, A.; Ginex, T.; Gibert, E.; Rabal, O.; Oyarzabal, J.; Herrero, E.; Luque, F. J. Development and validation of molecular overlays derived from three-dimensional hydrophobic similarity with PharmScreen. *J. Chem. Inf. Model.* **2018**, *58*, 1596−1609.
(45) Vázquez, J.; Deplano, A.; Herrero, A.; Ginex, T.; Gibert, E.; Rabal, O.; Oyarzabal, J.; Herrero, E.; Luque, F. J. Development and validation of molecular overlays derived from three-dimensional hydrophobic similarity with PharmScreen. *J. Chem. Inf. Model.* **2018**, *58*, 1596−1609.
(46) Wu, Y.; Canturk, B.; Jo, H.; Ma, C.; Gianti, E.; Klein, M. L.; Pinto, L. H.; Lamb, R. A.; Fiorin, G.; Wang, J.; DeGrado, W. F. Flipping in the pore: Discovery of dual inhibitors that bind in different orientations to the wild-type versus the amantadine-resistant S31N mutant of the influenza A virus M2 proton channel. *J. Am. Chem. Soc.* **2014**, *136*, 17987−17995.

https://doi.org/10.1021/acsmedchemlett.4c00502
ACS Med. Chem. Lett. XXXX, XXX, XXX−XXX

Supporting Information to

Mining druggable sites in influenza A hemagglutinin: Binding of the pinanamine-based inhibitor M090

Aitor Valdivia,[†,‡] Maria Rocha,[†,¶] F. Javier Luque[†,‡,§,*]

[†] Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació - Campus Torribera, Universitat de Barcelona, Prat de la Riba 171, 08921 Santa Coloma de Gramenet, Spain

[‡] Institut de Biomedicina (IBUB), Universitat de Barcelona

[¶] Department of Life Sciences, University of Coímbra, Calçada Martim de Freitas, 3000-456 Coímbra, Portugal

[§] Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona

[*] Email: fjluque@ub.edu

Index

Homology modeling S2
Figure S1 S2
Table S1 S3
Molecular Dynamics simulations S4
Figure S2 S6
Figure S3 S7
Figure S4 S8
Figure S5 S9
Alchemical transformations S10
Figure S6 S13
Figure S7 S14
Principal component analysis S15
Table S2 S15
Figure S8 S16
Figure S9 S17
References S18

Homology modelling

The 3D structure of HA (H1N1 A/Virginia/ATCC3/2009) was modelled using SWISS-MODEL[S1] and the crystallographic structure of the A/PR/8/34 H1 HA as structural template (PDB ID 6WCR,[S2] resolution 2.68 Å). This was chosen to take advantage of previous studies focused on the interaction of aniline-based antiviral compounds with HA.[S3] Furthermore, the residues that shape the pocket proposed for the binding of M090 are preserved in distinct H1N1 HA subtypes (see Table S1). Thus, comparison of the sequence content of the α-helix formed by $L80_2$–$A96_2$ (H-BS) and the loops $P301_1$–$K318_1$ and $M59_2$–$L73_2$ (L1-BS and L2-BS, respectively) for A/Virginia/ATCC3/2009, A/California/04/2009, A/Texas/04/2009, A/Guangzhou/GIRD/07/2009, A/California/07/2009, A/WSN/1933 and A/PR/8/1934 shows very few changes, such as the replacement of $K312_1$ by E in A/WSN/1933 and A/PR/8/1934, $T61_2$ and $H72_2$ by I and K in A/PR/8/193, and $D86_2$ by G in A/Texas/04/2009. Most of these changes occur in solvent-exposed areas of the binding pocket (Figure S1). In contrast, larger changes are observed in H3N2, H7N3 and H9N2 strains, which may contribute to the weakening in the inhibitory potency.

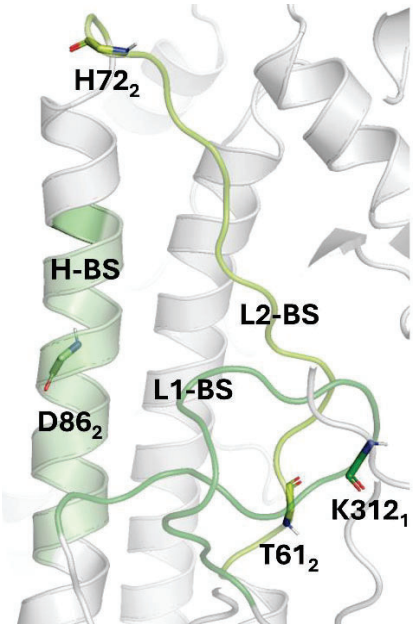


Figure S1. Representation of the structural elements that shape the putative binding pocket of M090 in HA: the α-helix formed by $L80_2$–$A96_2$ (H-BS) and the loops $P301_1$–$K318_1$ and $M59_2$–$L73_2$ (L1-BS and L2-BS, respectively). The location of the residues affected by the few non-conservative mutations between the viral H1N1 strains are highlighted with the backbone represented as sticks.

Table S1. Sequence alignment of the structural elements that shape the proposed binding pocket of M090 in hemagglutinin: the α-helix formed by $L80_2$–$A96_2$ (H-BS) and the loops $P301_1$–$K318_1$ and $M59_2$–$L73_2$ (L1-BS and L2-BS, respectively). Conservative changes in the nature of the residues are highlighted in green, whereas non-conservative changes are shown in red.

HA strain	EC_{50} (μM)	Sequence (**L1-BS**)	Sequence (**L2-BS**)	Sequence (H-**BS**)
A/Virginia/ATCC3/2009 (H1N1)	--- [a]	PFQNIHPITIGKCPKYVK	MNTQFTAVGKEFNHL	LNKKVDDGFLDIWTYNA
A/California/04/2009 (H1N1)	--- [b]	PFQNIHPITIGKCPKYVK	MNTQFTAVGKEFNHL	LNKKVDDGFLDIWTYNA
A/Texas/04/2009 (H1N1)	0.10 [c]	PFQNIHPITIGKCPKYVK	MNTQFTAVGKEFNHL	LNKKVDGGFLDIWTYNA
A/Guangzhou/GIRD/07/2009 (H1N1)	0.30 [d]	PFQNIHPITIGKCPKYVK	MNTQFTAVGKEFNHL	LNKKVDDGFLDIWTYNA
A/California/07/2009 (H1N1)	0.34 [c]	PFQNIHPITIGKCPKYVK	MNTQFTAVGKEFNHL	LNKKVDDGFLDIWTYNA
A/WSN/1933 (H1N1)	1.48 [e]	PFQNIHPVTIGECPKYVR	MNTQFTAVGKEFNNL	LNKKVDDGFLDIWTYNA
A/PR/8/1934 (H1N1)	2.23 [e]	PYQNIHPVTIGECPKYVR	MNIQFTAVGKEFNKL	LNKKVDDGFLDIWTYNA
A/Aichi/2/1968 (H3N2)	4.55 [e]	PFQNVNKITYGACPKYVK	TNEKFHQIEKEFSEV	LEKYVEDTKIDLWSYNA
A/duck/Guangdong/1/1996 (H7N3)	4.58 [e]	PFQNINPRTVGRCPRYVK	TNQQFELIDNEFSEI	VINWTRDSMTEVWSYNA
A/Hong Kong/1968 (H3N2)	5.38 [e]	PFQNVNKITYGACPKYVK	TNEKFHQIEKEFSEV	LEKYVEDTKIDLWSYNA
A/duck/HK/Y280/1997 (H9N2)	6.79 [e]	PFHNVSKYAFGNCPKYVG	MNKQYEIIDHEFSEV	INNKIDDQIQDIWAYNA

[a] Structural model used in this work. [b] Structural used in ref. S3. [c] Activity measured in plaque reduction assay. [d] Activity determined in virus-Infected MDCK cells. [e] Activity measured in cytophatic effect assay.

Molecular Dynamics simulations

The amberff14sb force field[S4] was used for the protein, and M090 and its derivatives were parameterised using the gaff2 force field.[S5] Partial charges for the ligands were derived using the RESP protocol[S6] at the B3LYP/6-31G(d) level of theory, as calculated using Gaussian16.[S7] The systems were solvated with a truncated octahedral box of TIP3P[S8] water molecules, allowing a distance of 12 Å from the protein to the box edge. Counterions (K^+, Cl^-)[S9] were added to maintain the neutrality of the simulated system and to keep the ion concentration at 0.15 M using the SPLIT method.[S10] The simulated systems contain ~246,000 atoms.

Each system was minimised using 7500 steps of steepest descent, followed by 7500 steps of conjugate gradient algorithm. Equilibration was performed in 3 steps. The systems were slowly heated in the NVT ensemble from 5 K to 300 K in a temperature ramp for 250 ps and held at 300K for 50 ps imposing positional restraints (5 kcal·mol^{-1}) on the ligands. The density of the system was then equilibrated for 4.7 ns in the NPT ensemble (pressure: 1 bar, T: 300 K). Finally, after density equilibration, the system was simulated using the NVT ensemble for 1 ns and prepared for the MD production.

During the production runs, temperature was controlled using Langevin dynamics with a collision frequency of 2 ps^{-1}, and pressure was maintained using the Berendsen barostat during the density equilibration step. All bonds involving hydrogen atoms were constrained by the SHAKE algorithm[S11] to allow a timestep of 2 fs. Since the X-ray structural information is limited to the ectodomain, a series of distance NMR restraints were applied to avoid an artefactual increase of flexibility at the C-termini of the HA trimer. In particular, restraints (5 kcal·mol^{-1}) were applied to the C_α of two residues ($N128_2$ and $R170_2$) of each HA monomer when a displacement greater than 2Å from the crystallographic distance occurred (the restraint is applied progressively until the restraint is fixed at 1.5 Å beyond 2Å from the crystallographic distance, i.e., at 3.5 Å from the crystallographic distance). All simulations were performed using the AMBER20 package.[S12]

The structural flexibility of the binding pocket and the stability of M090 was assessed through the analysis of a variety of MD simulations for the apo and holo systems (Table 1). The initial model was simulated in the apo state for 2 μs of unbiased MD.

For the ligand-bound complexes, three binding modes were considered (Figure S2). The first one corresponds to the binding pose reported by Zhao *et al.,*[S13] which will be denoted

BM-I and is characterized by the formation of the hydrogen bond between the amine nitrogen of M090 and $T309_1$. The second binding mode (BM-II) is characterized by the formation of a salt bridge between the charged nitrogen of M090 and $D85_2$. Finally, the third binding pose (BM-R) is characterized by the exchange of the pinanamine and thiophene units in the binding pocket. The holo systems were simulated using at least two independent MD simulations, each simulated for 1 μs (the only exception is one of the simulations performed for the BM-R binding mode, which was run for 0.5 μs due to the release of one of the ligands from the binding pocket). In one case the starting structure was the equilibrated HA model obtained from the SWISS MODEL web server. Alternatively, the starting structure was the frame obtained after 1 or 2 μs of the MD simulation run for the apo system.

Finally, the apo species of the E74→$D74_2$ mutant was simulated for a total of 2 μs, and the ligand-bound complex formed with the E74→$D74_2$ mutant was examined considering two MD simulations extended up to 1 μs each.

Figure S2. RMSD profiles determined for (left) the backbone of the residues that define the three binding pockets and (right) the bound ligand in the two MD simulations (H1wt-I and H2wt-I) run for the complex between M090 (binding mode BM-I) and HA.

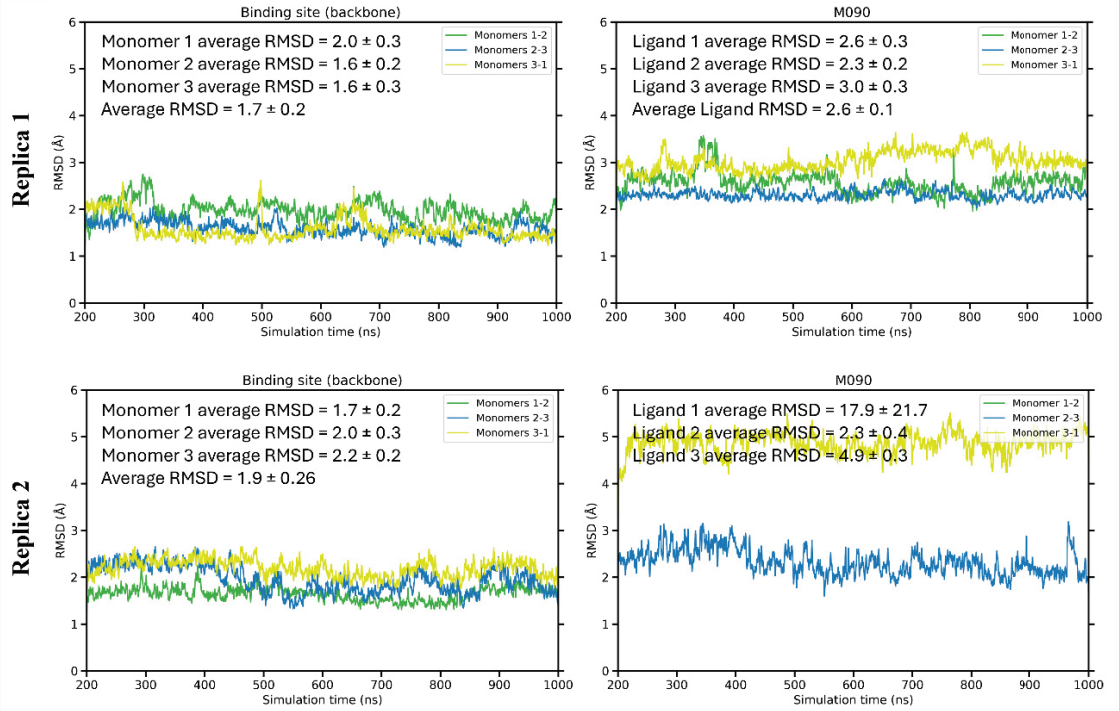


Figure S3. RMSD profiles determined for (left) the backbone of the residues that define the three binding pockets and (right) the bound ligand in the four MD simulations (H1wt-II, H2wt-II, H3wt-II and H4wt-II) run for the complex between M090 (binding mode BM-II) and HA.

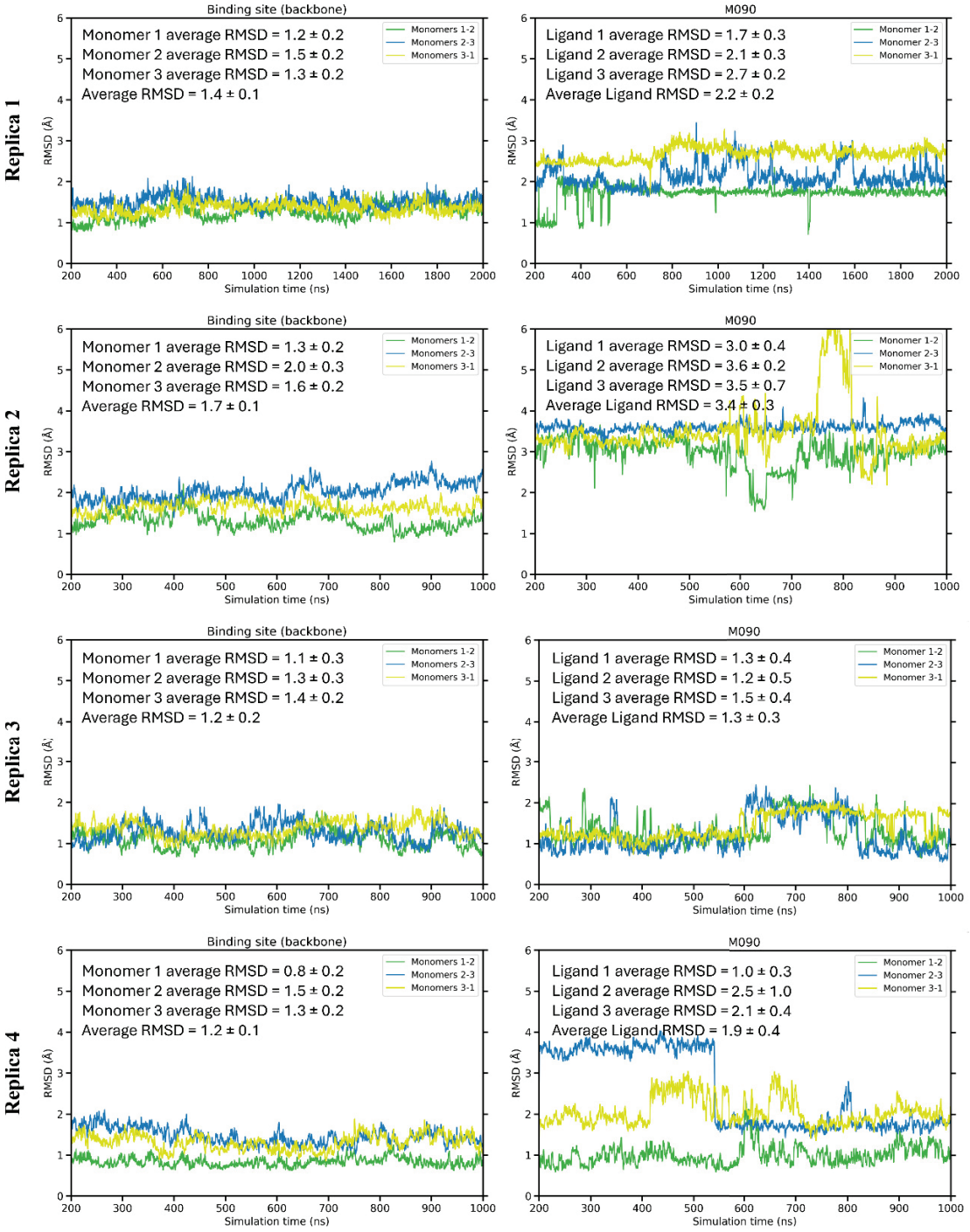


Figure S4. RMSD profiles determined for (left) the backbone of the residues that define the three binding pockets and (right) the bound ligand in the two MD simulations (H1wt-R and H2wt-R) run for the complex between M090 (binding mode BM-R) and HA.

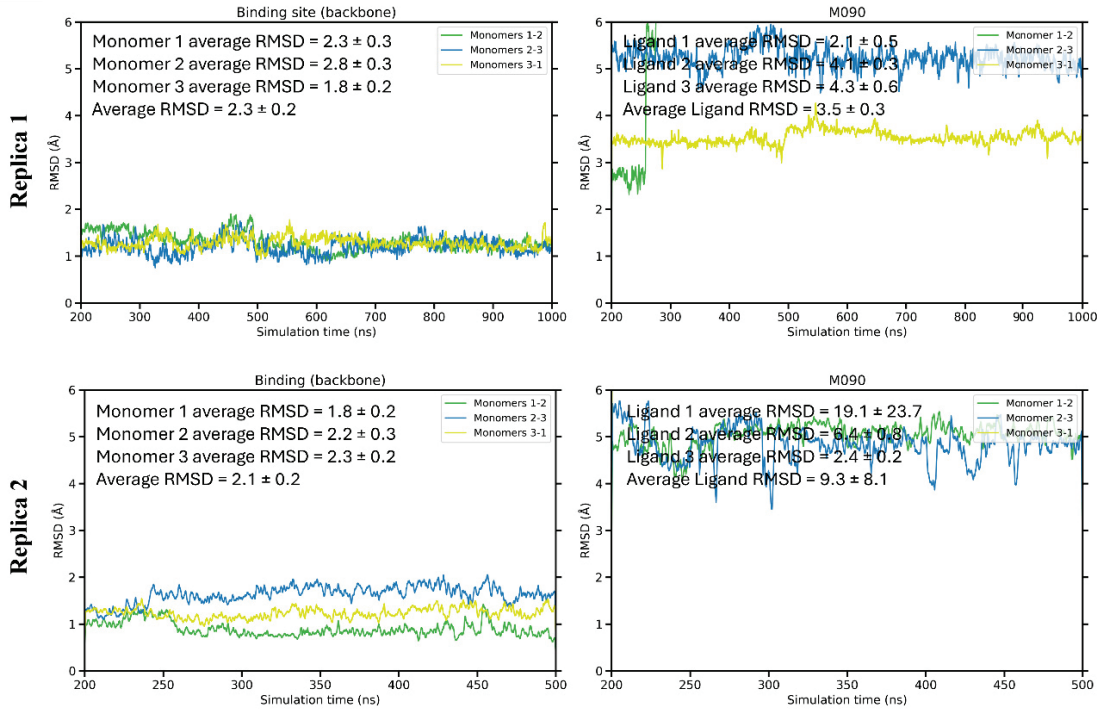
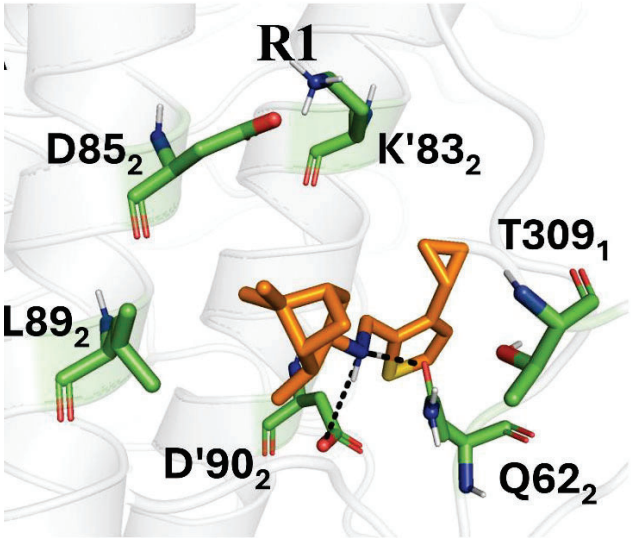


Figure S5. Representation of the alternative binding mode found for M090 in the simulations run for the reversed binding mode (H1wt-R and H2wt-R), where the protonated amine formed direct interactions with the amide carbonyl oxygen of $Q62_2$, assisted by electrostatic interactions with $D'90_2$.



Alchemical calculations

The relative binding free energy (RBFE)[S14] between derivatives of M090 was determined by alchemical transformations, where a ligand (L_1) is converted to a structurally related analogue (L_2) both in the protein-bound complex and in the unbound state in aqueous solution (Eq. 1).

$$\Delta\Delta G_{bind} = \Delta\Delta G_{complex}(L_1 \rightarrow L_2) - \Delta\Delta G_{solvent}(L_1 \rightarrow L_2) \quad (1)$$

The transformation from L_1 to L_2 was divided into a series of windows, where the states corresponding to $\lambda = 0$ and $\lambda = 1$ represent the holo species with L_1 and L_2, respectively, and the intermediate λ values denote a linear interpolation between the parameters of the initial and final systems. Alchemical transformations were performed using 29 equally spaced λ windows, including states $\lambda = 0$ and $\lambda = 1$ (i.e., λ = 0.000000, 0.035714, 0.071429, 0.107143, 0.142857, 0.178571, 0.214286, 0.250000, 0.285714, 0.321429, 0.357143, 0.392857, 0.428571, 0.464286, 0.500000, 0.535714, 0.571429, 0.607143, 0.642857, 0.678571, 0.714286, 0.750000, 0.785714, 0.821429, 0.857143, 0.892857, 0.928571, 0.964286, 1.000000).

To smoothly transform L_1 into L_2, we followed a concerted alchemical transformation protocol using a dual-topology scheme. During the transformation, two copies of the common core (CC) of the molecules are present with identical sets of atomic coordinates to facilitate phase space overlap between states during the alchemical transformation. The non-common regions for each pair of molecules (i.e., the atoms with different sets of coordinates) were evaluated with soft-core potentials.[S15] All the λ simulations were run for 12ns - 18ns, depending on the needs to reach the convergence criteria. Calculations were performed using the GPU-accelerated TI implementation of AMBER20.

The Thermodynamic Integration (TI) estimator as implemented in *alchemlyb* was used to estimate the free energy change for each λ simulation. In addition, each transformation was simulated in both directions to assess hysteresis. For each alchemical transformation, we determined the average value of $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda}$ and the standard error (and variance) at each lambda. The free energy difference between initial and final states was estimated through integration of the set of $\left\langle \frac{\partial U}{\partial \lambda} \right\rangle_{\lambda}$ values using the trapezoidal rule. Similarly, the error was estimated from the propagation of the variances determined at each lambda between the initial and final states. Since each alchemical transformation was performed in forward and backward directions, the free energy change was estimated as the average of the two

values, and the error was determined as the square root of the sum of the corresponding variances.

In the study, all the free energy differences are given relative to M090, which was used as reference. Accordingly, when comparison of a given compound relative to M090 involves several alchemical transformations (i.e, M090 → M23 → M06 → M04; see Figure 3), the global free energy difference was determined from the addition of the free energy changes for each individual transformation, and the global error from the cumulative propagation of the variances.

In few instances several pathways may be used to link a given compound to M090. For instance, the transformation M090→M04 may be completed using 3 possible paths following the perturbation map in Figure 3:

1. M090 → M23 → M06 → M04
2. M090 → M23 → M17 → M19 → M07 → M04
3. M090 → M23 → M27 → M07 → M04

In this case, the final free energy difference was taken as the average value of the 3 values corresponding to each path, and the final error was estimated from the square root of the sum of the variances for the distinct paths.

The initial coordinates of the system were extracted from the H1wt- II simulation at 1 μs (Table 1), keeping the positions of the water molecules at 8 Å of each M090 molecule to improve the reproducibility of the binding mode during the equilibration of the system after changing the topology. All the systems were then carefully equilibrated with the dual topology using the TI code for each transformation with its respective λ value. Each system was first heated in the NVT ensemble from 5 K to 300 K (150 ps), equilibrated at 1 bar in the NPT ensemble (300 ps), and finally switched back to the NVT (100 ps) ensemble to avoid artifacts at the beginning of the TI simulations.

During the equilibration protocol, the heavy atoms of the protein and ligand were restrained with positional restraints (10 kcal·mol^{-1}) to maintain the binding mode obtained during the MD. Each λ state was then simulated in the NVT ensemble at 300 K using the Langevin thermostat and the SHAKE algorithm for bonds involving hydrogen. The analysis and integration of the final free energy values was performed using the last 6ns of the total sampling for each λ state.

The experimental binding free energies, $\Delta\Delta G_{exp}$, were approximated from the experimental EC_{50} reported by Zhao *et al.*[S13] using influenza virus-infected MDCK cell assays (Eq. 2). It should be noted that this choice may introduce some discrepancies

between the results from in silico calculations and cell-based assays due to the potential effect of other factors not directly related to the binding of the compounds to the putative binding site, such as the effective concentration of the ligand and non-specific binding events, in the experimental assays.

$$\Delta\Delta G_{456} = -RTln\frac{78!"(:\#)}{78_{!"}(:\$)} \quad (2)$$

where $\Delta\Delta G_{exp}$ is in kcal·mol^{-1}, R ≈ 1.987·10^{-3} kcal·mol^{-1}·K^{-1} and T is the temperature in K (300K).

Figures S6 and S7 show representative plots used in the analysis of the free energy changes determined from TI calculations for the alchemical transformations between M04 and M06 in aqueous solution and in the protein-ligand complex.

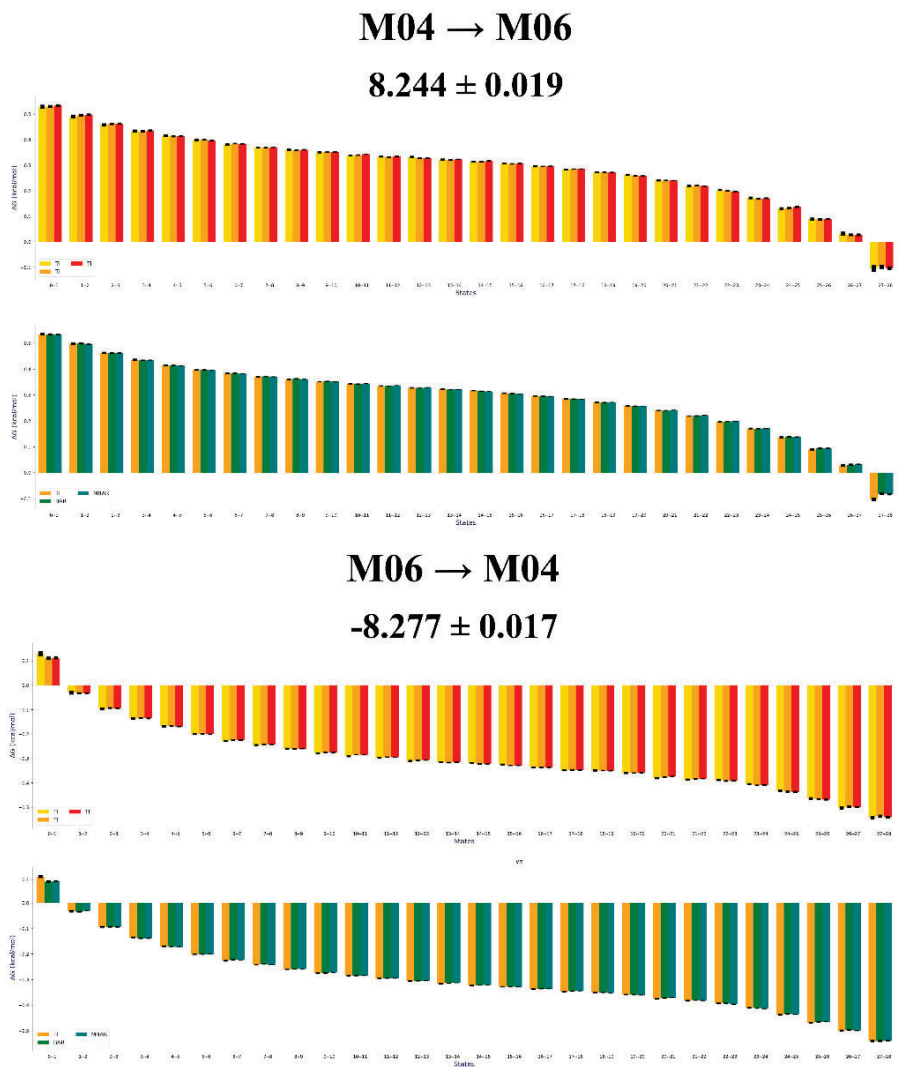


Figure S6. TI analysis for the M04↔M06 (forward and backward) transformation in aqueous solution (values in kcal·mol^{-1}). Representation of (top) the free energy change determined from the last 6ns of each λ divided in steps of 2 ns (each in yellow, orange and red), and (bottom) correlation between the TI estimator used in the calculations and the BAR and MBAR estimators.

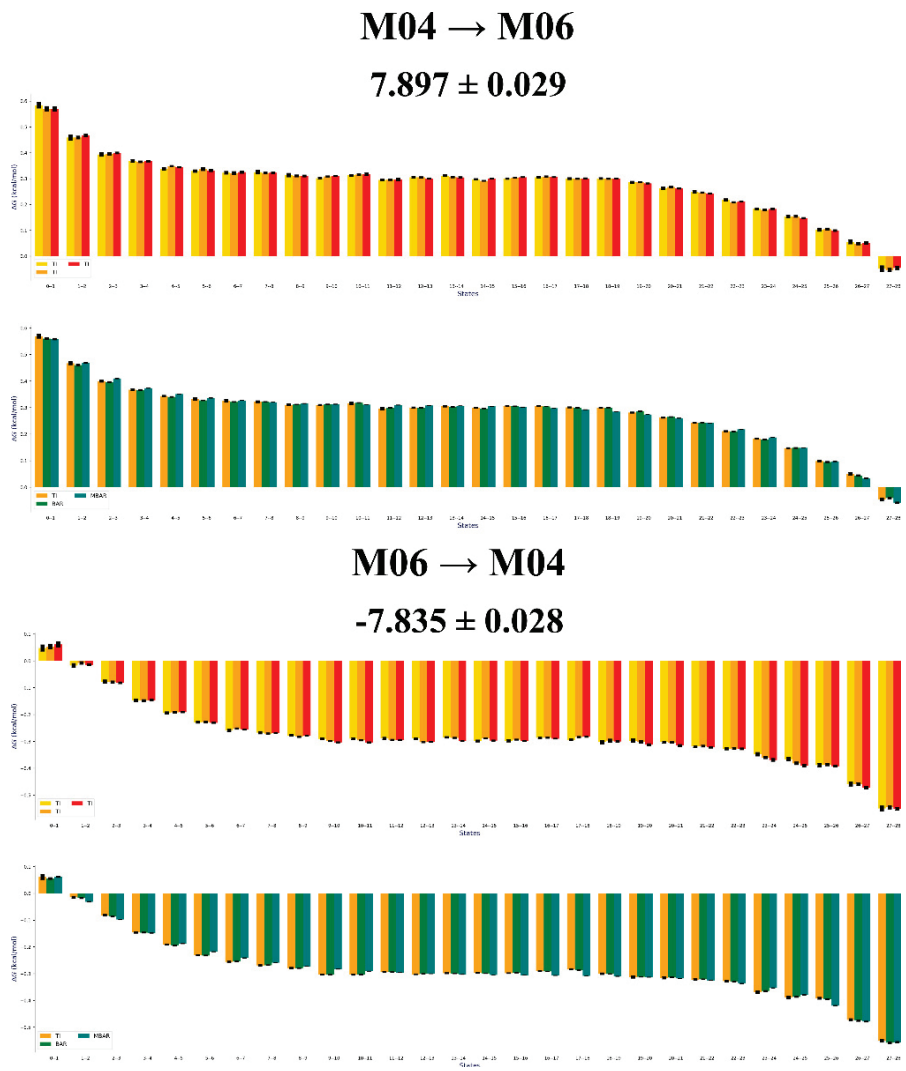


Figure S7. TI analysis for the M04↔M06 (forward and backward) transformation in the ligand-protein complex (values in kcal·mol^{-1}). Representation of (top) the free energy change determined from the last 6ns of each λ divided in steps of 2 ns (each in yellow, orange and red), and (bottom) correlation between the TI estimator used in the calculations and the BAR and MBAR estimators.

Principal component analysis

To explore the conformational flexibility of the structural elements of HA that shape the putative binding pocket for M090 in the wild type and $E74_2 \rightarrow D$ mutated variant, Principal Component Analysis (PCA) was used to examine the structural plasticity of the apo systems and the M090-bound systems in BM-II. Note that the residue $E74_2$ is located distant from the binding pocket, as noted in the distances from the Cα atom from this latter residue to the Cα atom of both D852 and T3091 (see Table S2).

Computation of the PCA was performed for the main chain atoms (C, Cα, N and O atoms) of L1-BS, L2-BS and H-BS. To this end, the trajectories of each system were properly aligned to the α-helix backbone of an average structure computed with the last 100 ns of the H1-WT_II system, which was used as a reference structure. The eigenvalues and eigenvectors were calculated diagonalizing the mass-weighted covariance matrix of the system using the *gmx covar* and *gmx anaeig* modules of the GROMACS 2023.2[S16] package.

Table S2. Distance (Å) between the Cα atom of residues $E74_2$ ($D74_2$ in the mutated variant) and $D85_2$ and $T309_1$ in the binding pocket of M090.

Simulation	$D85_2$ – E/$D74_2$ (Å)	$T309_1$ – E/$D74_2$ (Å)
Awt	17.1 ± 0.4	24.1 ± 0.6
Amut	16.7 ± 0.3	23.8 ± 0.6
H1wt-II	17.2 ± 0.4	24.2 ± 0.4
H2wt-II	16.8 ± 0.3	24.4 ± 0.5
H3wt-II	17.2 ± 0.3	24.5 ± 0.4
H4wt-II	17.2 ± 0.4	24.3 ± 0.4
H1mut-II	16.9 ± 0.3	24.1 ± 0.5
H2mut-II	16.8 ± 0.3	24.0 ± 0.5

Figure S8. RMSD profiles determined for the backbone of the residues that define the three binding pockets the MD simulations run for the (top) apo wild type HA and (bottom) its mutated variant.

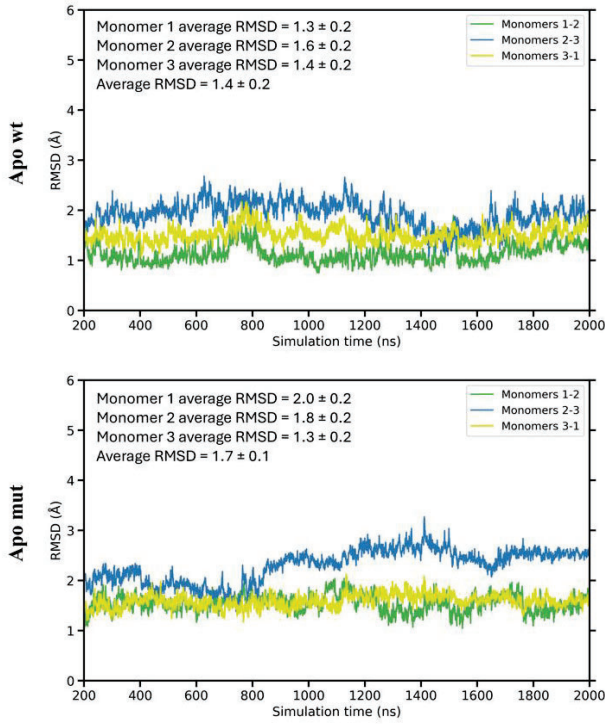
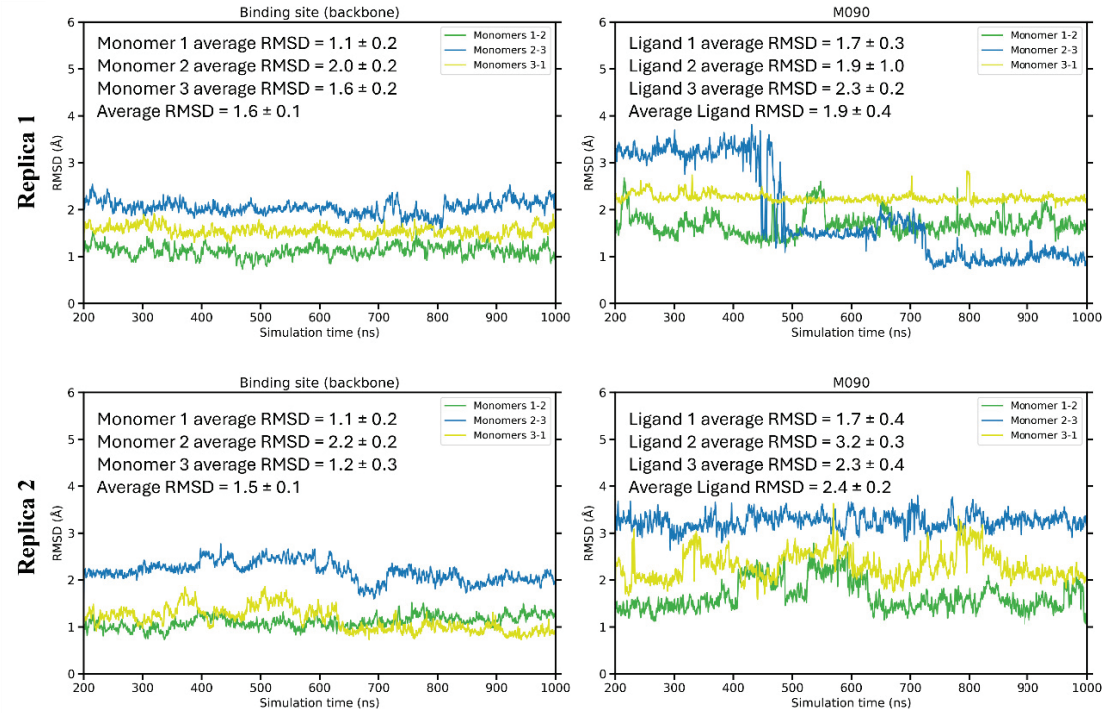


Figure S9. RMSD profiles determined for (left) the backbone of the residues that define the three binding pockets and (right) the bound ligand in the two MD simulations (H1mut-II and H2mut-II) run for the complex between M090 (binding mode BM-R) and the mutated variant of HA.



References

(S1) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: Homology modelling of protein structures and complexes. *Nuc. Acids Res*. **2018**, *46*, W296–W303.
(S2) Yao, Y.; Kadam, R. U.; Lee, C. D.; Woehl, J. L.; Wu, N. C.; Zhu, X.; Kitamura, S.; Wilson, I. A.; Wolan, D. W. An influenza A hemagglutinin small-molecule fusion inhibitor identified by a new high-throughput fluorescence polarization screen. *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 18431–18438.
(S3) Leiva, R.; Barniol-Xicota, M.; Codony, S.; Ginex, T.; Vanderlinden, E.; Montes, M.; Caffrey, M.; Luque, F. J.; Naesens, L.; Vázquez, S. Aniline-based inhibitors of influenza H1N1 virus acting on hemagglutinin-mediated fusion. *J. Med. Chem*. **2018**, *61*, 98–118.
(S4) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. FF14SB: Improving the accuracy of protein side chain and backbone parameters from FF99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
(S5) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
(S6) Wang, J.; Cieplak, P.; Kollman, P. A. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.
(S7) Gaussian 16, Revision B.01, Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V.; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V.; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery, J. A., Jr.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian, Inc., Wallingford CT, 2016.
(S8) Mark, P.; Nilsson, L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960.
(S9) Joung, I. S.; Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041.
(S10) Machado, M. R.; Pantano, S. Split the charge difference in two! A rule of thumb for adding proper amounts of ions in MD simulations. *J. Chem. Theory Comput.* **2020**, *16*, 1367–1372.
(S11) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of N-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
(S12) Case, D. A.; K. Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Roe, D. R.;

Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wilson, L.; Wolf, R. M.; Wu, X.; Xiong, Y.; Xue, Y.; York, D. M.; Kollman, P. A. (2020), AMBER 2020, University of California, San Francisco.
(S13) Zhao, X.; Li, R.; Zhou, Y.; Xiao, M.; Ma, C.; Yang, Z.; Zeng, S.; Du, Q.; Yang, C.; Jiang, H.; Hu, Y.; Wang, K.; Mok, C. K. P.; Sun, P.; Dong, J.; Cui, W.; Wang, J.; Tu, Y.; Yang, Z.; Hu, W. Discovery of highly potent pinanamine-based inhibitors against amantadine- and oseltamivir-resistant influenza A viruses. *J. Med. Chem.* **2018**, *61*, 5187–5198.
(S14) Röblitz, S.; Weber, M. Fuzzy Spectral Clustering by PCCA+: Application to Markov State Models and Data Classification. *Adv. Data Anal. Classif.* **2013**, *7*, 147–179.
(S15) Tsai, H.-C.; Tao, Y.; Lee, T.-S.; Merz, K. M. Jr.; York, D. M. Validation of free energy methods in AMBER. *J. Chem. Inf. Model.* **2020**, *60*, 5296–5300.
(S16) Abraham, M. J.; Murtola, T.; Schulz, R.; Páall, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High Performance Molecular Simulations through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1–2*, 19–25.

Project 3

"N-[(Thiophen-3-yl)methylbenzamides as Influenza Virus Fusion Inhibitors Acting on H1 and H5 Hemagglutinins"

"N-[(Thiophen-3-yl)methylbenzamides as Influenza Virus Fusion Inhibitors Acting on H1 and H5 Hemagglutinins"

Silke Rimaux[1‡], Aitor Valdivia[2,3‡], Juan Martín-López[4,5], Valeria Francesconi[5,#], Celia Escriche[4,5], Cato Mestdagh[1], Ria Van Berwaer[1], Kaat Verleye[1], Samuel Noppen[1], Óscar Lozano[4,5†], Annelies Stevaert[1], Francisco Javier Luque[2,3,4*], Lieve Naesens[1*], Santiago Vázquez[4,5*]

[1]Rega Institute, Department of Microbiology, Immunology and Transplantation, KU Leuven, B-3000 Leuven, Belgium.

[2]Doctorate in Biotecnology, Department of Nutrition, Food Science and Gastronomy, Faculty of Pharmacy and Food Sciences, Universitat de Barcelona, Av. Prat de la Riba 171, Santa Coloma de Gramanet, E-08921, Spain.

[3]Institute of Theoretical and Computational Chemistry (IQTCUB), Universitat de Barcelona, 08028 Barcelona, Spain.

[4]Institute of Biomedicine of the University of Barcelona (IBUB), Universitat de Barcelona, 08028 Barcelona, Spain.

[5]Laboratori de Química Farmacèutica, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Av. Joan XXIII, 27-31, 08028 Barcelona, Spain.

[‡]Equal contribution to this manuscript.

[#]Current address: Department of Pharmacy, University of Genoa, viale Benedetto XV n.3, Genoa 16132, Italy.

[†]Deceased.

KEYWORDS. Antivirals, benzamides, drug design, fusion inhibitor, hemagglutinin, influenza virus.

Note: the manuscript is ready for submission.

Summary

To adequately prepare against severe infections from seasonal or pandemic influenza A virus (IAV), additional drug classes are highly needed. We here report a new class of N-[(thiophen-3-yl)methylbenzamides which target the hemagglutinin (HA)-mediated fusion process. The most active compound, VF-57a, displays a 50% effective concentration (EC_{50}) of ~0.8 μM and antiviral selectivity index >130, in Madin-Darby canine kidney (MDCK) cells infected with A/H1N1 virus. No activity was observed for influenza A/H3N2 and B viruses. VF-57a proved to be a strong inhibitor of A/H1N1- and A/H5N1-pseudovirus entry, with EC_{50} values of 0.3 and 0.8 μM, respectively. The molecule prevents the conformational change of H1 or H5 HA at acidic pH, as evident from cell-cell fusion assays in HA-expressing cells, and from surface plasmon resonance-based assessment of HA protein refolding at low pH. A(H1N1)pdm09 virus, selected for resistance to VF-57a, carried three muta-tions in H1-HA, i.e. $F128_1L$, $T316_1I$ and $S54_2P$, indicating a potential role for these residues in compound binding. Molecular modeling identi-fied the putative binding cavity for VF-57a, which lies close to the binding site of arbidol, a broad inhibitor of HA-mediated fusion, and to the $S54_2P$ resistance mutation. The binding mode, which involves π-stacking interactions with $F309_1$ and $Y94_2$ in the HA2 subunit, was supported by the binding affinity predictions estimated from relative binding free energy calculations, which were able to rationalize the structure-activity relationship (SAR) of selected compounds. In combination, our findings support further development of this novel class of IAV fusion inhibitors having strong activity against A/H1N1 and A/H5N1 viruses.

N-[(Thiophen-3-yl)methyl]benzamides as Influenza Virus Fusion Inhibitors Acting on H1 and H5 Hemagglutinins

Silke Rimaux[1‡], Aitor Valdivia[2,3‡], Juan Martín-López[4,5], Valeria Francesconi[5,#], Celia Escriche[4,5], Cato Mestdagh[1], Ria Van Berwaer[1], Kaat Verleye[1], Samuel Noppen[1], Óscar Lozano[4,5†], Annelies Stevaert[1], Francisco Javier Luque[2,3,4*], Lieve Naesens[1*], Santiago Vázquez[4,5*]

[1]Rega Institute, Department of Microbiology, Immunology and Transplantation, KU Leuven, B-3000 Leuven, Belgium.

[2]Doctorate in Biotecnology, Department of Nutrition, Food Science and Gastronomy, Faculty of Pharmacy and Food Sciences, Universitat de Barcelona, Av. Prat de la Riba 171, Santa Coloma de Gramanet, E-08921, Spain.

[3]Institute of Theoretical and Computational Chemistry (IQTCUB), Universitat de Barcelona, 08028 Barcelona, Spain.

[4]Institute of Biomedicine of the University of Barcelona (IBUB), Universitat de Barcelona, 08028 Barcelona, Spain.

[5]Laboratori de Química Farmacèutica, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Av. Joan XXIII, 27-31, 08028 Barcelona, Spain.

[‡]Equal contribution to this manuscript.

[#]Current address: Department of Pharmacy, University of Genoa, viale Benedetto XV n.3, Genoa 16132, Italy.

[†]Deceased.

KEYWORDS. Antivirals, benzamides, drug design, fusion inhibitor, hemagglutinin, influenza virus.

ABSTRACT: To adequately prepare against severe infections from seasonal or pandemic influenza A virus (IAV), additional drug classes are highly needed. We here report a new class of *N*-[(thiophen-3-yl)methylbenzamides which target the hemagglutinin (HA)-mediated fusion process. The most active compound, **VF-57a**, displays a 50% effective concentration (EC_{50}) of ~0.8 μM and antiviral selectivity index >130, in Madin-Darby canine kidney (MDCK) cells infected with A/H1N1 virus. No activity was observed for influenza A/H3N2 and B viruses. **VF-57a** proved to be a strong inhibitor of A/H1N1- and A/H5N1-pseudovirus entry, with EC_{50} values of 0.3 and 0.8 μM, respectively. The molecule prevents the conformational change of H1 or H5 HA at acidic pH, as evident from cell-cell fusion assays in HA-expressing cells, and from surface plasmon resonance-based assessment of HA protein refolding at low pH. A(H1N1)pdm09 virus, selected for resistance to **VF-57a**, carried three mutations in H1-HA, i.e. $F128_1L$, $T316_1I$ and $S54_2P$, indicating a potential role for these residues in compound binding. Molecular modeling identified the putative binding cavity for **VF-57a**, which lies close to the binding site of arbidol, a broad inhibitor of HA-mediated fusion, and to the $S54_2P$ resistance mutation. The binding mode, which involves π-stacking interactions with $F309_1$ and $Y94_2$ in the HA2 subunit, was supported by the binding affinity predictions estimated from relative binding free energy calculations, which were able to rationalize the structure-activity relationship (SAR) of selected compounds. In combination, our findings support further development of this novel class of IAV fusion inhibitors having strong activity against A/H1N1 and A/H5N1 viruses.

INTRODUCTION

According to estimates of the World Health Organization, the annual epidemics, caused by influenza A and B viruses, are globally responsible for ~470,000 respiratory deaths per year.[1] In addition, antigenically distinct influenza A viruses (IAVs), emerging from zoonotic sources, can cause pandemics with potentially grave consequences,[2] as evident from the estimated toll of ~200,000 respiratory deaths and ~83,000 cardiovascular deaths, during the last IAV pandemic that occurred in 2009.[3] Nowadays, a highly pathogenic avian influenza A/H5N1 virus is causing great concern, after recent outbreaks in dairy cattle in the USA and fur farms in Europe, with regular spillover to human individuals.[4,5]

Although annual influenza vaccination represents the main prophylaxis for people at risk (such as the elderly or persons suffering from comorbidities), the overall effectiveness of the current vaccines is only between 30 and 60%.[6] Therefore, antiviral drugs are crucial to treat influenza-infected persons who are seriously ill or at high risk for severe complications.[7,8] Oseltamivir and other neuraminidase

inhibitors are the standard-of-care since many years, while the polymerase inhibitors baloxavir marboxil and favipiravir are also available in certain countries.[9] For both pharmacological classes, close monitoring of potential resistance is needed.[10] Other classes of easily accessible and cost-effective antiviral molecules remain urgently needed to prepare against the threat of influenza pandemics. This includes exploration of alternative druggable targets, such as the viral hemagglutinin (HA).[11] The small molecule arbidol (**1**) (Chart 1) acts on HA besides other targets,[12,13] and is currently approved in two countries (Russia and China).

HA is a homotrimeric protein embedded in the viral envelope. As the key player for viral entry into the host cells, HA forms the initial binding interaction with sialylated glycans on the host cell surface, which results in uptake of the virus by endocytosis.[2] As the endosomes mature to reach a pH of ~5.5, HA is triggered to undergo a drastic conformational change in its stem region,[11,14-17] which promotes the release of the hydrophobic fusion peptide and, ultimately, fusion of the endosomal membrane and viral envelope. Diverse strategies are being developed to interfere with the viral entry process,[18-21] yet many of these are challenged by the high variability of HA. The 19 currently known IAV HA subtypes fall into two phylogenetic groups. The H1, H2 and H5 HAs belong to group 1, whereas the H3 and H7 HAs fall in group 2.[19] Broadly neutralizing anti-HA antibodies with group 1-, group 2-, or pan-IAV activity have sparked major interest from the pharmaceutical industry, with a handful having progressed towards clinical evaluation.[22]

Besides, there has been significant interest in small molecule antivirals which prevent the HA-mediated membrane fusion process. Several fusion inhibitors have been reported in the literature, with extensive variety in their scaffold structures (Chart 1). In almost all cases, the activity proved HA subtype-dependent, covering either H1 or H3 HA, but not both, although several molecules showed inhibition of more than one HA subtype of the same group (Chart 1). A few fusion inhibitors were also validated in influenza mouse models.[23-25] So far, arbidol, **1**,[12,13]; M090, **2**,[26]; OA-10, **3**,[27]; and **4**,[28] are exceptional in having activity against both H1 HA and H3 HA, albeit at potencies that are quite low compared to the subtype- or group-restricted inhibitors.

Up to now, structural studies have revealed that fusion inhibitors bind to two sites in the HA stem region. One site corresponds to the pocket that is occupied by *t*-butylhydroquinone, **5**, (TBHQ; PDB IDs 3EYK and 3EYM)[29] and arbidol (PDB IDs 5T6N and 5T6S),[12] and consists of residues located at helices A and C from protomer 1 and C' from protomer 2. The other is the groove filled by JNJ4796, **9**, (PDB IDs 6CFG and 6CF7),[23] (*S*)-F0045, **10**, (PDB ID 6WCR)[30] and CBS1117, **11a** (PDB ID 6VMZ).[31] Considering the substantial chemical diversity among the reported fusion inhibitors (Chart 1) and the structural complexity of HA, the identification of the mechanism of action is challenging, although disclosing the putative binding site is crucial to rationalize the structure-activity relationships (SAR) and lead optimization.

This study elaborates on the chemotype for which the first prototype, CL385319, **19**, (Chart 1), was published in 1999.[32] The authors demonstrated antiviral activity against A/H1N1 and A/H2N2 viruses (group 1), with A/H3N2 (group 2) being much less sensitive. Several years later, another team found that CL385319 is also effective against A/H5N1 virus.[33] Subsequent studies explored the effect of replacing the piperidine ring of CL385319 by a variety of heteroaromatic rings. This led to submicromolar potencies, as in the case of the 2-(thiophenyl-2-yl)ethyl derivative **20** (Chart 1), which exhibited an EC_{50} of 0.22 μM against A/H5N1 pseudovirus.[34] These researchers also identified an oligothiophene compound having 58-fold higher potency than CL385319 in A/H1N1 virus-infected cell cultures and 360-fold higher activity in an A/H5N1 pseudovirus entry assay.[35] Finally, in 2019, Yu *et al.* performed a systematic SAR investigation on a series of heteroaromatic-based benzenesulfonamides, leading to compound **21**, which exhibited an EC_{50} of 0.47 μM against A/H5N1 virus.[36]

In light of the promising anti-IAV activities of benzamides **19** and **20** and of compound **21**, we decided to further investigate the SAR, considering a series of hybrid molecules featuring, as left-hand side, the benzamide core of CL385319 and its analog **20** and, as right-hand side, the *N*-[(thiophen-3-yl)methyl]amino unit of **21** (Scheme 1). Starting from the hybrid molecule **22**, chemical variations were designed to explore (i) the effects of substitutions attached to the phenyl ring (Charts 2-4); (ii) the optimal substitution for the thiophene ring (Charts 5-6); and (iii) the nature of the linker between the thiophene and the phenyl moieties (Chart 7).

The anti-IAV activity was evaluated in cell-based assays with different A/H1N1 and A/H5N1 (pseudo)viruses, and the fusion-inhibiting effect was determined in HA-expressing cells. To understand the HA binding mode, resistant virus was selected by serial passaging. Finally, molecular dynamics simulations and free energy calculations were conducted to identify the plausible binding site of the lead compound within the HA protein.

RESULTS AND DISCUSSION

Synthesis and SAR analysis. The benzamides were easily synthesized through the reaction of the required acyl chloride, either commercially available or freshly synthesized from its corresponding benzoic acid derivative and thionyl chloride, with the required amine, typically (2,5-dimethylthiophen-3-yl)methanamine. Some particular compounds, such as ester **63**, secondary amine **64** and inverse amide **65**, were synthesized using specific procedures as reported in the Experimental section. All the new compounds were fully characterized through their spectroscopic data and elemental analyses or HPLC/MS (see Experimental section and Supporting Information for further details)..

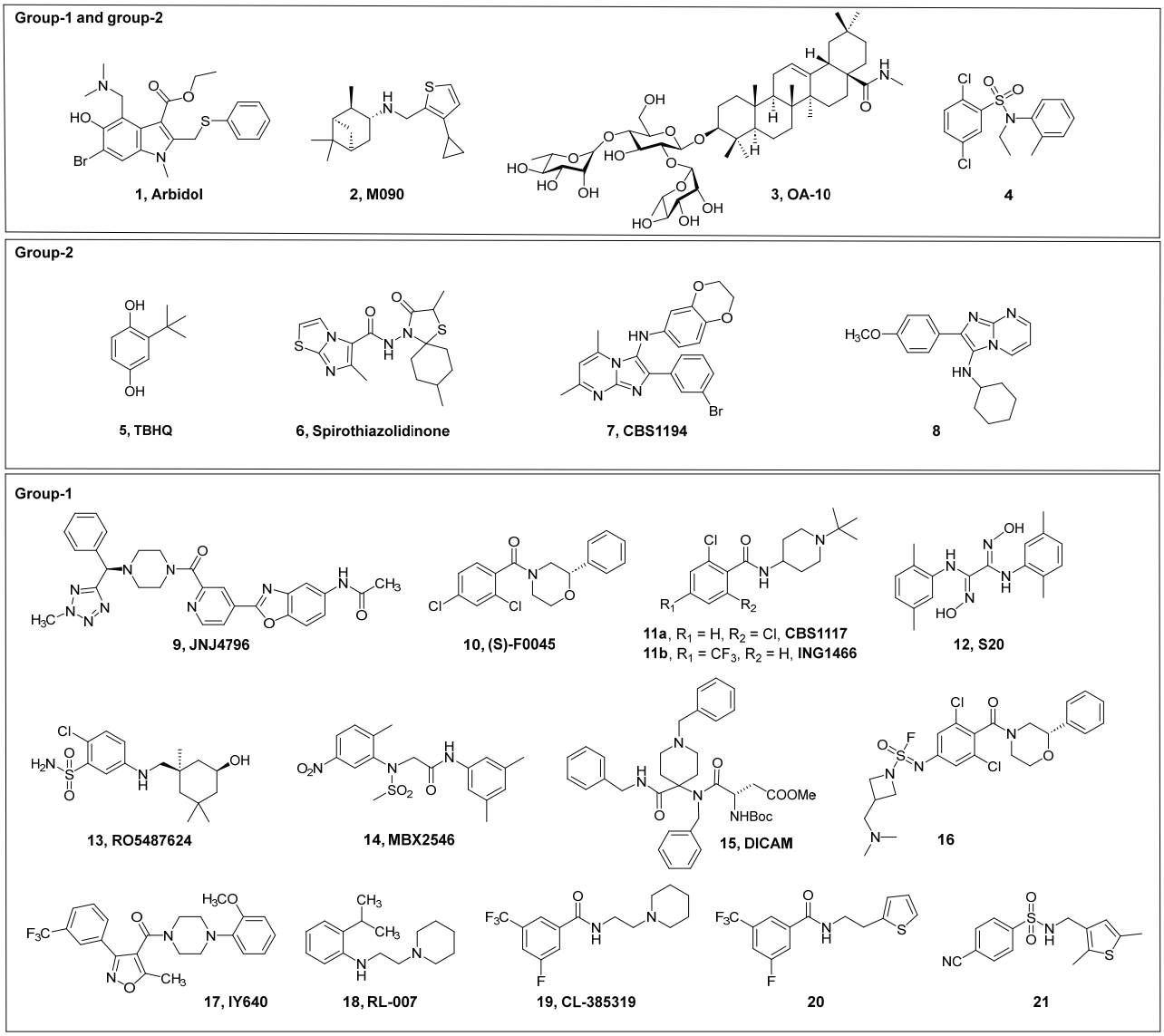
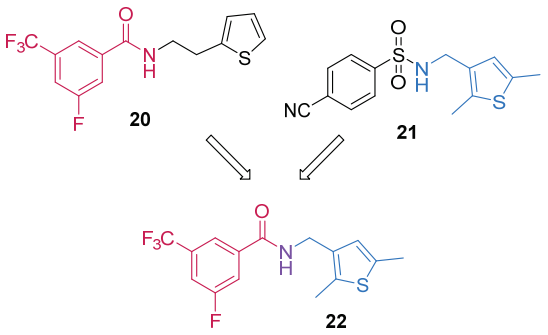


Chart 1. Structures of reported inhibitors of IAV HA-mediated fusion, specifically: arbidol[12,13]; M090[26]; OA-10[27]; **4**[28];TBHQ[29]; spirothiazolidinones[37,38]; CBS1194[39]; **8**[40]; JNJ4796[23]; (*S*)-F0045[30]; CBS1117[31]; ING1466[24]; S20[41]; RO5487624[25]; MBX-2456[42]; DICAM[43]; **16**[44]; IY640[45]; RL-007[46]; CL-385319[32]; **20**[34]; and **21**[36].

From the hybrid molecule **22** (Scheme 1), the SAR analysis of the *N*-[(thiophen-3-yl)methylbenzamides aimed to gain insight into how anti-IAV activity depends on the thiophene and phenyl rings, and the linker that joins these two moieties.

With regard to the benzene ring, starting from **22**, besides the replacement of fluorine by bromine, chlorine or trifluoromethyl in *meta* position and some related combinations shown in Chart 2, we also explored the effect of attaching additional electron-withdrawing groups in position *ortho* and *para* (Chart 3). On the other hand, since previous works on CL-385319 and **20** did not perform a full SAR study of the benzene ring, a classical Topliss approach was adopted to consider a more diverse range of substituents attached to the benzene ring (Chart 4).[47,48]



Scheme 1. Structures of anti-IAV compounds **20** and **21** and newly designed hybrid **22**.

Regarding the thiophene ring, attention was paid first to the influence of the methyl groups present in positions 2 and 5, which were removed either separately or simultaneously (Chart 5). Furthermore, the replacement of the thiophene moiety by more polar rings, such as furan, thiazole, isoxazole, pyrazole, oxazole and pyridine, was also examined (Chart 6).

Finally, we envisaged the replacement of the methylamide linker by other chemical moieties, keeping the total length of the tether between both aromatic rings (Chart 7).

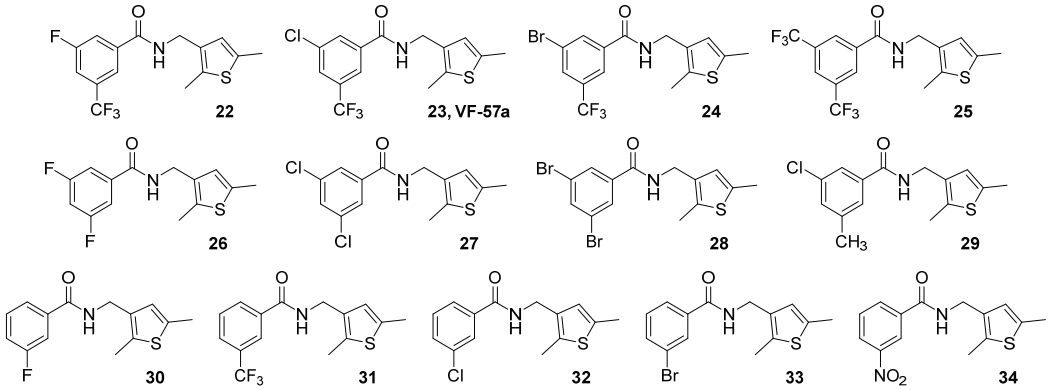


Chart 2. Structures of benzamides with a mono- or di-substituted phenyl ring.

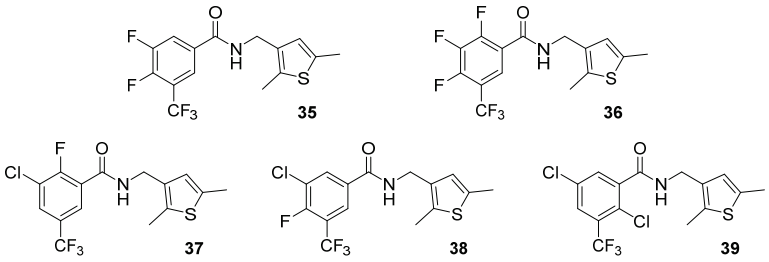


Chart 3. Structures of benzamides with tri- and tetra-substituted phenyl ring.

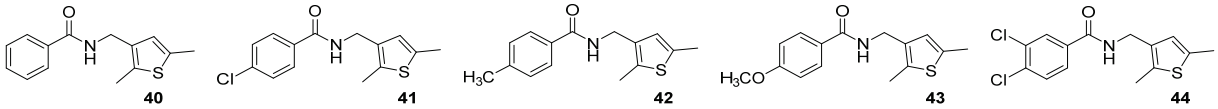


Chart 4. Structures of benzamides explored for a Topliss approach of the benzene ring.

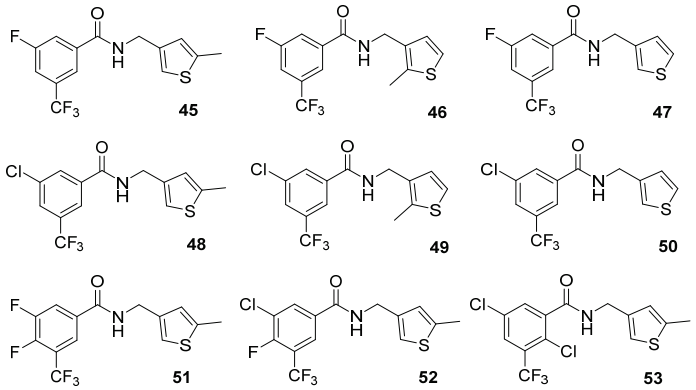


Chart 5. Structures of benzamides synthesized for exploring the influence of the thiophene methyl groups.

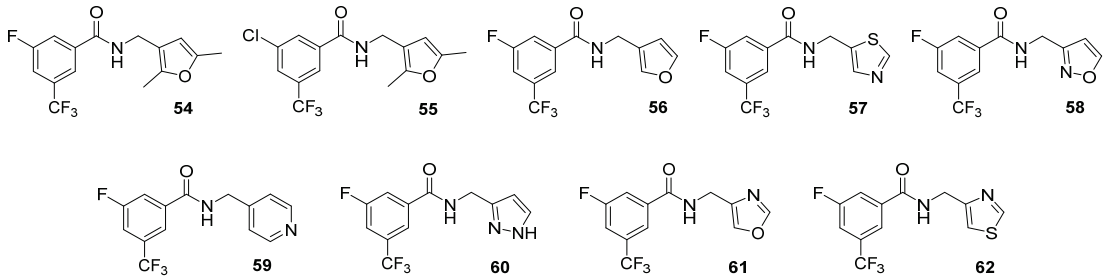


Chart 6. Structures of benzamides featuring different heterocycles on the right-hand side of the molecule.

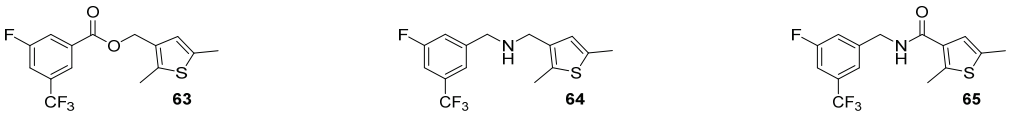


Chart 7. Structures of benzamides with alternative bridges between the phenyl and thiophene rings.

To determine the anti-influenza virus activity, we conducted cytopathic effect (CPE) reduction assays in MDCK cells, in which we used the colorimetric MTS cell viability method to measure the compounds' protective effect against viral CPE, as well as their cytotoxicity in mock-infected cells. Parallel with the MTS readout, we scored the CPE and cytotoxicity by microscopic inspection; the two readout methods gave very similar results. The virus test panel included two A/H1N1 (PR8 and Virg09), two A/H3N2 (A/Victoria/361/11 and A/Hong Kong/7/87) and one influenza B strain (B/Ned/537/05; Yamagata lineage). This evaluation showed that several members of the series exhibited strong activity against A/H1N1 virus (see Table 1 showing the results for the most active compounds). In line with the reported data on the congeners CL-385319,[32] **20**,[34] and **21**,[36] we saw no inhibitory effect on A/H3N2 and influenza B viruses (highest tested concentration: 100 μM, data not shown).

The starting benzamide, **22**, had antiviral EC_{50} values ~1 μM for the PR8 and Virg09 virus, without showing any cytotoxicity at 100 μM (the highest concentration tested; Table 1), thus yielding a selectivity index (= ratio of CC_{50} to EC_{50}, SI) of >100. Interestingly, while substituting the fluorine atom of **22** by chlorine, as in **23**, led to a slightly more potent compound (Table 1), its replacement by a bromine (**24**) or a trifluoromethyl group (**25**) fully abolished the antiviral activity. Three analogs featuring the same halogen atom, **26**, **27**, and **28**, were also inactive. No activity was also noted for compound **29**, bearing a methyl group instead of the trifluoromethyl group of **23**. Finally, five monosubstituted benzamides, **30**–**34**, were inactive, showing that the antiviral activity requires the presence of two electron-withdrawing groups in the *meta* positions (i.e. at R_1 and R_4; Chart 2).

Next, from active benzamides **22** and **23**, the incorporation of one or two additional halogen atoms on the benzene ring was explored. Of note, the introduction of a fluorine atom at the *para* position of **22**, leading to **35**, slightly increased the activity. However, the introduction of a further fluorine atom at the *ortho* position, as in **36**, fully abolished the antiviral effect. Regarding **23**, the effects of introducing additional halogen atoms paralleled those seen with the derivatives of **22**, with compound **38** showing good antiviral activity and the *ortho*-substituted derivatives **37** and **39** being less active.

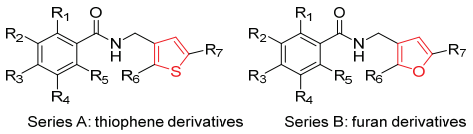
A classical medicinal chemistry approach to optimize the substitution pattern of a phenyl ring, is the Topliss batchwise scheme (TBS).[47,48] Therefore, we synthesized and tested compounds **40**–**44**, but these were inactive, thus preventing further TBS-guided design (Chart 4).

Next, we explored the right-hand side of the molecule. First, a few analogues of the more potent compounds, **22**, **23**, **35** and **38**, were synthesized to explore the effect of the methyl groups (Chart 5). All the compounds lacking one or both of the methyl groups (i.e. **45**-**53**) were active, although the 5-methyl appeared more relevant for the antiviral activity, particularly against the Virg09 virus (see Table 1). Next, several heterocycles were explored as replacements of the thiophene ring. However, except for the demethylated furan derivatives **54** and **55**, all these analogs were inactive (Chart 6 and Table 1).

Finally, the importance of the benzamide bridge of **22** was assessed. Neither the ester **63**, nor the secondary amine **64** showed antiviral activity. Finally, also the inverse amide **65** was fully inactive (Chart 7).

Based on the EC_{50} and SI values, we selected **23**, dubbed **VF-57a**, for mechanistic investigations. This lead compound had an antiviral EC_{50} value of 1 μM and 0.5 μM for the PR8 and Virg09 virus, respectively, and a selectivity index of >100 (Table 1). For comparison, RL-007 (Chart 1), an aniline-based IAV fusion inhibitor which we identified before[46] and included as a reference compound, had an EC_{50} of 0.6–1.5 μM and CC_{50} of 33 μM (Table 1), giving a selectivity index of ~31.

Table 1. Antiviral (A/H1N1) activity and cytotoxicity in MDCK cells.



Series A: thiophene derivatives Series B: furan derivatives

Cpd	R_1	R_2	R_3	R_4	R_5	R_6	R_7	Antiviral EC_{50}[a] against A/H1N1 virus (μM)				Cytotoxicity (μM)	
								PR8 strain		Virg09 strain		CC_{50}[b]	MCC[c]
								MTS	Microscopy	MTS	Microscopy	MTS	Microscopy
Series A													
22	H	F	H	CF_3	H	Me	Me	1.1 ± 0.3	1.5 ± 0.6	0.44 ± 0.19	1.7 ± 1.1	>100	>100
23	H	Cl	H	CF_3	H	Me	Me	0.92 ± 0.19	1.1 ± 0.2	0.31 ± 0.12	0.71 ± 0.45	>100	>100
35	H	F	F	CF_3	H	Me	Me	0.22 ± 0.03	0.86 ± 0.41	0.20 ± 0.05	0.89 ± 0 .13	>100	>100
37	F	Cl	H	CF_3	H	Me	Me	2.9	1.6 ± 0.8	>100	>100	>100	56 ± 22
38	H	Cl	F	CF_3	H	Me	Me	0.27 ± 0.12	1.6 ± 0.8	0.18 ± 0.04	3.1 ± 0.5	>100	40
39	H	Cl	H	CF_3	Cl	Me	Me	5.9 ± 3.2	11 ± 6	0.82 ± 0.61	8.1 ± 0.5	>100	52 ± 25
45	H	F	H	CF_3	H	H	Me	0.68 ± 0.51	0.26 ± 0.02	0.39 ± 0.18	1.6 ± 0.5	67 ± 1	78 ± 22
46	H	F	H	CF_3	H	Me	H	0.23 ± 0.08	0.17 ± 0.04	6.8 ± 6.5	8.3 ± 5.1	>100	67 ± 33
47	H	F	H	CF_3	H	H	H	0.42 ± 0.10	0.44 ± 0.16	7.2 ± 4.1	20 ± 8	>100	>100
48	H	Cl	H	CF_3	H	H	Me	0.34 ± 0.18	0.46 ± 0.14	0.15 ± 0.12	1.9 ± 1.1	>100	>100
49	H	Cl	H	CF_3	H	Me	H	0.14 ± 0.04	0.14 ± 0.05	0.37 ± 0.16	0.48 ± 0.07	>100	78 ± 22
50	H	Cl	H	CF_3	H	H	H	2.0 ± 1.0	1.0 ± 0.6	1.1 ± 0.5	2.1 ± 0.0	70 ± 2	100 ± 0
51	H	F	F	CF_3	H	H	Me	2.9 ± 1.1	3.5 ± 1.2	5.3 ± 0.8	>100	24 ± 2	33 ± 0
52	H	Cl	F	CF_3	H	H	Me	8.5 ± 5.6	13 ± 8	>100	>100	>100	>100
53	H	Cl	H	CF_3	Cl	H	Me	17 ± 8	>100	31 ± 3	>100	>100	>100
Series B													
54	H	F	H	CF_3	H	Me	Me	0.47 ± 0.19	0.54 ± 0.17	0.54 ± 0.06	3.6 ± 2.8	79 ± 5	100 ± 0
55	H	Cl	H	CF_3	H	Me	Me	0.24 ± 0.08	0.54 ± 0.37	1.1 ± 0.3	>100	29 ± 0	33 ± 0
Reference compounds													
RL-007								1.4 ± 0.0	1.5 ± 0.1	0.61 ± 0.06	0.69 ± 0.09	33 ± 5	>40
Ribavirin								38 ± 3	27 ± 1	20 ± 2	25 ± 0	>250	80 ± 20
BXA (nM)[d]								6.1 ± 1.7	6.1 ± 2.0	3.2 ± 0.4	3.5 ± 0.0	>100	>100

[a]EC_{50}: half-maximal effective concentration for protection against virus-induced CPE, as determined by the MTS cell viability assay or by microscopic scoring. A/H1N1 virus strains: A/PR/8/34 (PR8) and A/Virginia/ATCC3/2009 (Virg09). [b]CC_{50}: 50% cytotoxic concentration based on the MTS assay. [c]MCC: minimum cytotoxic concentration, i.e. compound concentration causing minimal changes in cell morphology. [d]BXA: baloxavir acid; concentrations expressed in nM. Compounds **24–34**, **36**, **40–44**, **56–65** were inactive (EC_{50} > 100 μM) against the PR8 and Virg09 strains. All the compounds were inactive against the A/Hong Kong/7/87 (A/H3N2) and B/Ned/537/05 strains.

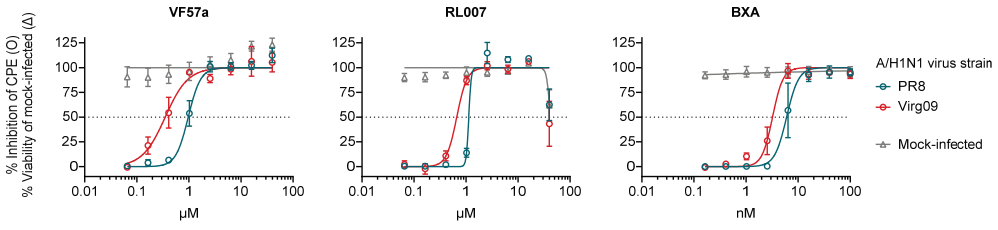


Figure 1. Dose-response curves for antiviral (A/H1N1) activity of VF-57a (23) and RL-007. Data points are the mean values ± SEM (n =3) for inhibition of CPE (O) or for viability of mock-infected cells (Δ), both based on the MTS assay. Curve fitting with GraphPad Prism 10.2.2 software.

Inhibition of H1 and H5 HA-mediated membrane fusion. Since the subtype-dependent antiviral activity aligned with HA being the target of the *N*-[(thiophene-3-yl)methyl]benzamides, we next determined the inhibitory effect of **VF-57a** in a luciferase-based pseudovirus entry assay in MDCK cells. This method enabled us to include, at BSL2 level, MLV-based pseudovirus bearing the HA and NA of highly pathogenic A/H5N1 virus, specifically strain FL22[49] which belongs to clade 2.3.4.4b alike the A/H5N1 viruses that are causing the current outbreaks in cattle in the USA.[5] As shown in Figure 2A, **VF-57a** proved a strong inhibitor of A/H1N1- and A/H5N1-pseudovirus entry, with EC_{50} values of 0.30 and 0.81 μM, respectively. RL-007, on the other hand, had almost 127-fold higher activity against A/H1N1 than A/H5N1 (EC_{50}: 0.079 and 10 μM, respectively). Arbidol, included as reference compound, had a comparable EC_{50} for A/H1N1 and A/H5N1 (Figure 2A), but ~20-fold lower potency than **VF-57a**.

More direct evidence that **VF-57a** acts on membrane fusion was obtained via cell-cell fusion assays in HA-expressing HeLa cells exposed to pH ~5. **VF-57a** and RL-007 were preincubated with the HeLa cells for 15 min, then further present during the 5-min acidic stage. Using a luciferase-based readout (see Figure 2B for an assay scheme), we determined an EC_{50} value for H1 HA of 0.95 μM for **VF-57a** and 4.1 μM for RL007. Both compounds proved also effective against H5 HA, giving complete inhibition of cell-cell fusion at 100 μM (see microscopic pictures in Figure 2C). At lower concentrations, **VF-57a** still gave almost complete (25 μM) or partial (6.25 μM) inhibition, while RL-007 was partially effective at 25 μM and inactive at 6.25 μM.

Hence, these pseudovirus entry and cell-cell fusion data validate **VF-57a** as a strong fusion inhibitor of H1 and H5 HAs. In contrast, our previously identified fusion inhibitor RL-007 showed stronger inhibition towards H1 than H5 HA.

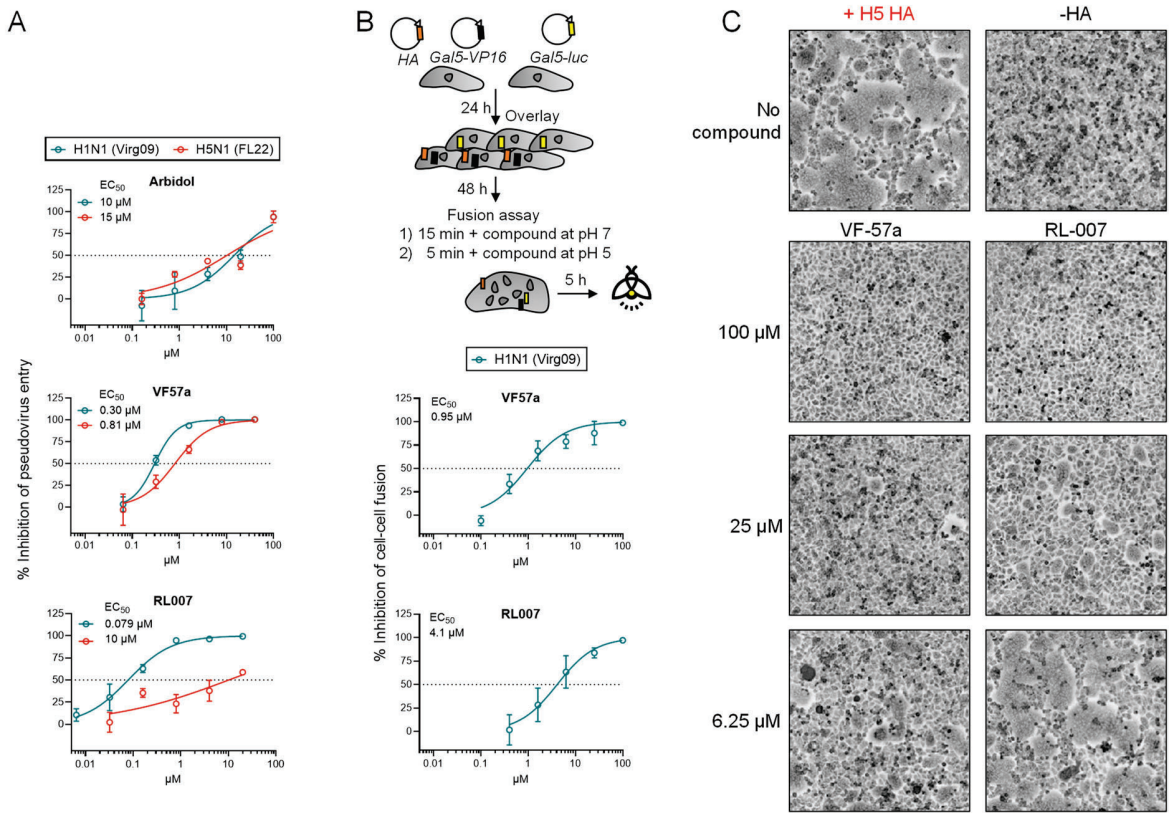


Figure 2 Dose-response curves for inhibition of HA-mediated pseudovirus entry and cell-cell fusion. (A) MDCK cells were transduced with luciferase-expressing A/H1N1- and A/H5N1-pseudoviruses, in the presence of the compounds. Three days later, luminescence was measured to assess the compounds' inhibitory effect on pseudovirus entry. (B) Cell-cell fusion assay in H1 HA-expressing HeLa cells. The scheme shows the assay setup. The compounds were present during the 15-min preincubation and 5-min acidic stage. Luminescence was measured 5 h after inducing cell-cell fusion. Data points are the mean ± SEM of three independent experiments. (C) Microscopic images of HeLa cells transfected with H5 HA and briefly exposed to pH 5.3. The compounds were present during the 15-min preincubation and 5-min acidic stage. In panels A and B, curve fitting was done with GraphPad Prism 10.2.2 software.

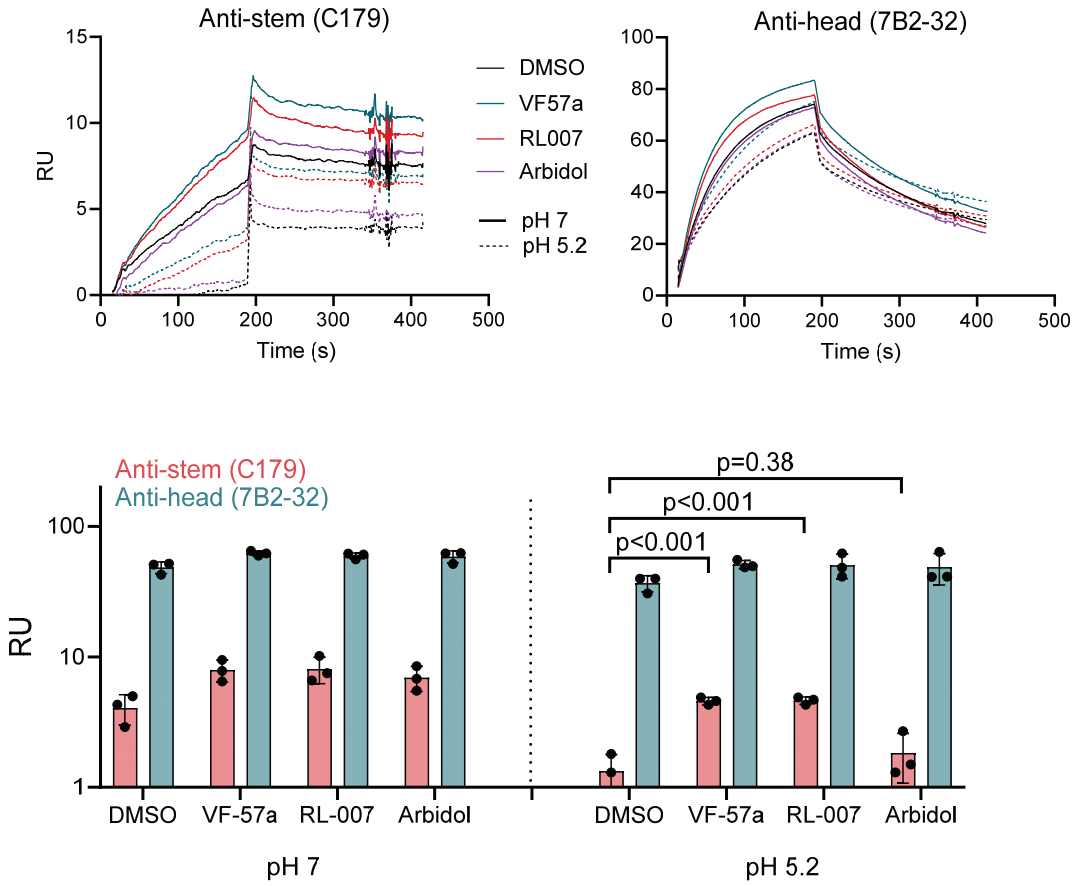


Figure 3. VF-57a and RL-007 inhibit HA refolding at pH 5.2. SPR was used to measure the binding of H1 HA to stem-directed antibody C179 and head-directed antibody 7B2-32. H1 HA was pre-incubated with DMSO or compound, then exposed to neutral or low pH, prior to SPR analysis. (A) Sensorgrams of a representative experiment. (B) Average binding response of three biological replicates. Statistical significance was analysed by a two-sided unpaired t-test. Isotype controls yielded RU values lower than 1 and are therefore not shown. RU: resonance units.

Finally, we conducted SPR analysis with two anti-HA antibodies, to evaluate the compounds' direct effect at preventing the conformational change of HA at low pH. Whereas the anti-HA stem antibody C179 only binds when the protein is in its prefusion conformation,[50,51] the anti-head antibody 7B2-32 binds to both pre- and post-fusion HA. Recombinant H1 HA protein, incubated in an acidic buffer of pH 5.2 to induce the post-fusion structure, lost the ability to bind to the C179-coated sensor chip (Figure 3). Its binding to the 7B2-32 antibody was not affected by acidic treatment. When **VF-57a** or RL-007 was present during the acidic incubation stage, the binding interaction between HA and antibody C179 remained significantly higher (P<0.001 for comparison to the DMSO control). This proves that the two molecules act by stabilizing the H1 HA protein in its prefusion structure. Arbidol, on the other hand, only slightly protected against the loss of C179 binding, and its effect was not significant. This is consistent with our pseudovirus entry data, where arbidol was much less effective than **VF-57a** and RL-007.

Phenotypic characterization of VF-57a-resistance markers in H1 HA protein. To fully understand the mechanism of **VF-57a** vis-à-vis RL007, we selected resistant viruses via serial passaging of Virg09 virus under increasing concentrations of these inhibitors. After three passages, breakthrough viruses were plaque-purified. In CPE reduction assays, all the virus clones, selected under either **VF-57a** or RL-007, proved at least 60-fold resistant to both compounds (EC_{50} >40 μM; Table 2), indicating cross-resistance between these two inhibitors. Baloxavir, included as control, was equally active against all viruses tested.

As evident from Figure 4, the Virg09 clones carried amino acid changes at different parts of HA. In a previous resistance study with an H3 HA-specific fusion inhibitor, we found that this class of agents selects two mechanistically distinct types of amino acid substitutions.[37] While some are directly located at the compound's binding pocket, other changes are lying at remote sites and associated with higher acid-lability of the HA protein. Such mutant viruses escape from the fusion inhibitor by undergoing fusion in less acidic

endosomes. Hence, we used a hemolysis assay to determine which of the mutations, selected under **VF-57a** or RL-007, rendered Virg09 HA more acid-labile. This was the case for mutants $T316_1I$ and $D90_2G$, which showed a hemolysis pH of 5.48 and 5.41, compared to 5.24 for the virus that was passaged without compound and did not acquire any HA mutations during this process (Table 2). Mutant $S54_2P$ showed a hemolysis pH of 5.30, which is a slight increase compared to the control (p=0.030). The $F128_1L$ virus had the same hemolysis pH as the control, while mutation $A90_1V$ reduced the acid-lability (pH 5.10; p<0.0001). This suggests that these sites may lie close to the binding pocket of both **VF-57a** and RL-007.

Table 2. Impact of the HA mutations on viral sensitivity to VF-57a or RL-007 and on acid-lability of HA.

Selecting compound[a] and HA mutation[b] vs parent virus	Antiviral EC_{50} by MTS assay[c]			Hemolysis pH (p *vs* control)[d]
	VF-57a (μM)	RL-007 (μM)	BXA (nM)	
VF57a: $F128_1L$	>40	>40	2.5 ± 0.5	5.25 (p=0.81)
VF57a: $T316_1I$	>40	>40	1.9 ± 0.1	5.48 (p<0.0001)
VF57a: $S54_2P$	>40	>40	1.7 ± 0.8	5.30 (p=0.030)
RL007: $A90_1V$	>40	>40	3.0 ± 0.7	5.10 (p<0.0001)
RL007: $D90_2G$	>40	>40	1.7 ± 0.2	5.41 (p=0.030)
No compound control: none	2.5 ± 1.0	0.94 ± 0.4	2.6 ± 0.7	5.24

[a]Virg09 virus was passaged three times under **VF-57a** and RL-007 (at successively 1.6, 4 and 10 μM for **VF-57a** and at 0.64, 1.6 and 4.0 μM for RL-007); a control was included that was passaged in the absence of a compound. [b]The residues are numbered as in ref. 46, i.e. per HA1 and HA2 polypeptide part (indicated by the suffix) and starting from the first methionine for HA1. In H3 numbering, the corresponding sites are: $F115_1$ ($F128_1$); $T301_1$ ($T316_1$); $S54_2$ ($S54_2$); $A81/82_1$ ($A90_1$); and $D90_2$ ($D90_2$). [c]Antiviral EC_{50} value in MDCK cells, using the MTS-based CPE assay. Mean ± SEM of three independent experiments. [d]The hemolysis pH was defined as the pH where 50% of virus-induced hemolysis occurred relative to pH 5.0, as based on the pooled data from four independent experiments. Statistical significance of differences between the hemolysis pH of different mutants *versus* the no compound control, were determined with an extra-sum-of-squares F test (GraphPad Prism 10.2.2).

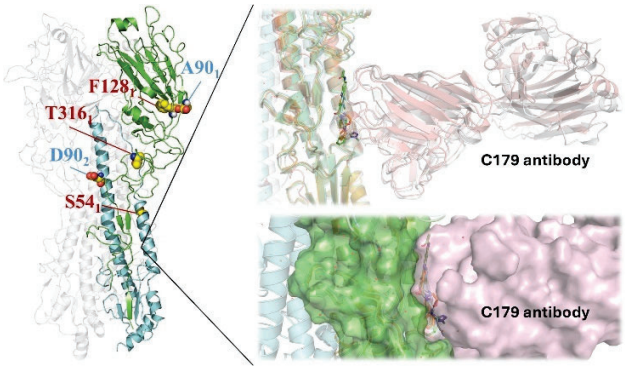


Figure 4. Location of the HA mutations that Virg09 virus acquired when passaged under **VF-57a** (red) or RL-007 (blue) (indicated using the X-ray crystallographic structure with PDB ID 3M6S).[52] The suffix refers to the HA1 or HA2 polypeptide chain. The zoom shows the binding site of anti-stem antibody C179 (revealed by X-ray crystallography; PDB ID 5C0R) and the aligned crystallographic structures of the HA-inhibitor complexes for JNJ4796, (*S*)-F0045 and CBS1117 (PDB IDs: 6CF7,[23] 6WCR[30] and 6WMZ,[31] respectively).

Exploring the binding of VF-57a to the ligand binding pockets in the HA stem. *Binding of (S)-F0045 and VF-57a to HA Site B.* As noted above, inspection of the available X-ray crystallographic structures shows the existence of two binding sites in the HA stem. One of them corresponds to the pocket that binds THBQ and arbidol, and the other mediates the recognition of JNJ4796, (*S*)-F0045 and CBS1117, which will be denoted Sites A and B, respectively, hereafter. Based on this evidence, Molecular Dynamics (MD) simulations were performed to assess the potential binding of **VF-57a** to the two sites. To this end, simulations were also performed for the HA complexes with arbidol and (*S*)-F0045, which were considered as control systems. Taking advantage of the trimeric nature of HA, MD simulations were carried out positioning the ligands in the three binding sites available in HA, enabling to calibrate the stability of the binding mode in triplicate in a single trajectory.

The X-ray structure of the complex between (*S*)-F0045 and HA (H1 A/Puerto Rico/8/1934, PDB ID 6WCR)[30] reveals that the ligand fills a groove shaped by residues $H18_1$, $H38_1$, $V40_1$, $W21_1$, $T318_1$, $D19_2$, $T41_2$, $Q42_2$, $I45_2$, $T49_2$ and $V52_2$. MD simulations performed for this complex showed that (*S*)-F0045 exhibits a stable binding mode, as noted in root-mean square deviation (RMSD) values of 1.3 ± 0.3, 1.3 ± 0.4 and 1.3 ± 0.3 Å for the three ligands (Figure 5). In particular, the hydrogen bond interaction formed between the amide carbonyl oxygen of (*S*)-F0045 and the hydroxyl group of $T318_1$ is preserved (average distance of 2.8 ± 0.1 in the three binding sites).

The potential binding of **VF-57a** to site B was examined considering two distinct binding modes, which differ in the spatial arrangement of the phenyl and thiophene rings along the groove (see Supporting Information Figure S1). Let us note that the two arrangements enable the formation of the hydrogen bond between the amide carbonyl oxygen of the ligand and the hydroxyl group of $T318_1$, while ensuring the overlap of the molecular skeleton of **VF-57a** and (*S*)-F0045.

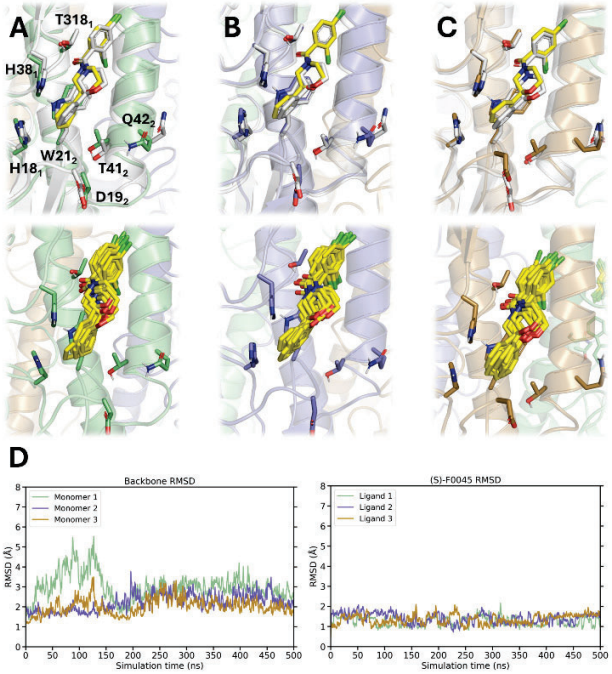


Figure 5. Binding mode of (*S*)-F0045 to Site B of HA. A), B) and C) Overlay of (*S*)-F0045 at the end of the MD simulation (orange sticks) and the crystallographic pose (grey sticks) in one of the three sites of HA (H1N1, A/Puerto Rico/8/1934). Below, the superposition of several snapshots along the whole trajectory. D) RMSD plots for each HA monomer backbone atoms and for each of the three ligands.

In contrast to the behavior found for (*S*)-F0045, none of the two poses of **VF-57a** were stable along the MD simulations performed for the complex with H1 (A/Virginia/ATCC3/2009 model) HA (Figure 6). Thus, the average RMSD values determined for the ligand range from 5.5 ± 2.6 Å to 9.0 ± 3.1 Å for the 6 ligands (3 ligands bound simultaneously to HA in the two binding modes). In addition, the structural instability of the binding mode is also reflected in the hydrogen bond formed with T333$_1$, since the average distance ranges from 5.2 ± 1.8 to 7.0 ± 3.8 Å. Compared to (*S*)-F0045, the instability of **VF-57a** could reflect the increased flexibility due to the methylene between the amide and thiophene moieties and the loss of the CH-π interactions formed by (*S*)-F0045 with H18$_1$, H38$_1$ and W21$_2$ due to the presence of the trifluoromethyl and methyl groups in the phenyl and thiophene rings, respectively. These results suggest that site B is not the binding pocket of **VF-57a** in H1 HA. This finding agrees with the distinctive behavior observed for the pH-dependent binding of antibody C179 in the absence and presence of **VF-57a**, since the binding of C179 to the stem would have been not compatible with the presence of **VF-57a** at site B.

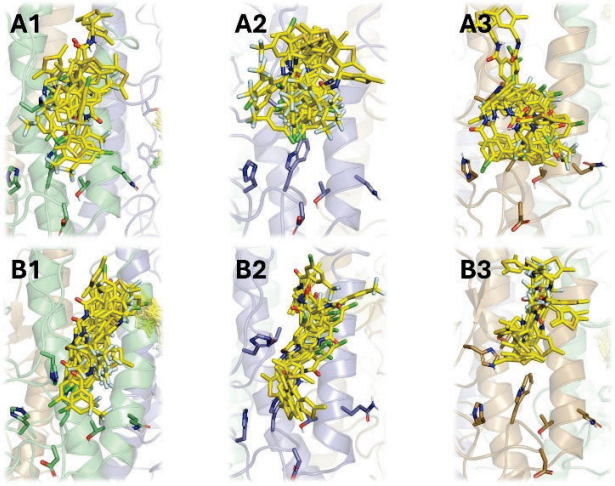


Figure 6. A1, A2 and A3) Superposition of snapshots along the trajectory for VF-57a in Site B binding mode A. B1, B2 and B3) Superposition of snapshots along the trajectory for VF-57a in Site B binding mode B. In both cases, a high displacement of the molecules is observed suggesting the lack of stabilizing interactions in Site B compared with the control MD simulation for (S)-F0045.

*Binding of arbidol and **VF-57a** to HA site A.* As described in the crystallographic structure for arbidol bound to H3 HA A/Hong Kong/1/1968, the binding pocket is shaped by residues of two HA monomers. Monomer 1 contributes with P293$_1$, F294$_1$, K307$_1$ and R54$_2$, Q57$_2$, K58$_2$, N60$_2$, W92$_2$ and Q103$_2$, whereas monomer 2 contributes with D90$_1$, A101$_1$, K310$_1$ and L29$_2$.[12]

Following the same strategy outlined above, MD simulations were performed for the complex of trimeric HA with the molecules of arbidol, each located at the binding pocket found in the X-ray crystallographic structure of H3 A/Hong Kong/1/1968. Arbidol exhibits a remarkable stable behavior in two of the three pockets during the whole MD simulation. The protein dynamics of the HA trimer are stable with average RMSD values of 2.2 ± 0.3, 2.5 ± 0.4 and 2.3 ± 0.3 Å for each monomer (Figure 7). Furthermore, the crystallographic binding mode of arbidol is maintained in two pockets, as noted in averaged RMSD values of 2.4 ± 0.6 Å and 3.0 ± 0.5 Å. We observed the formation of transient hydrogen bonds between the tertiary amine moiety of arbidol with E57$_2$ and the backbone carbonyl oxygen of K58$_2$. Nonetheless, these interactions compete with the intramolecular hydrogen bond formed with the carbonyl oxygen of arbidol. In the third case, the RMSD profile (green line in Figure 7) increased after 80-90 ns and achieved a stable value of 5.3 ± 1.8 Å after 210 ns. This reflects a rearrangement of the molecule within the binding pocket, where the indole ring flips 180° upside down and the thiophenyl group ends even more buried in the cavity.

The results observed in the simulation are aligned with the conclusions provided by the authors regarding the role of arbidol as a molecular glue for HA trimer, interacting through hydrophobic interactions and inducing the formation of salt bridges between amino acids.[12] Indeed, the salt bridge between the carboxylate group of D90$_2$ and the amine nitrogen of K310$_1$ is fully maintained during the simulation time (average N···O distance of 4.1 ± 1.4 Å), and the

guanidinium group of $R54_2$ forms transient interactions with either $E97_2$ or $E103_2$: when the salt bridge with $E97_2$ is formed (average distance of ~4.1 ± 1.0 Å), the distance from $E103_2$ is ~11.6 ± 1.0 Å, and when the interaction with $E97_2$ is broken (average distance of ~10.4 ± 1.2 Å), the sidechain of $R54_2$ rotates and interacts with $E103_2$ (average distance of ~3.4 ± 0.7 Å). However, the salt bridge between $R54_2$ and $E57_2$ observed in the crystallographic structure is lost at the beginning of the simulation. Altogether, this analysis reveals the structural plasticity of the binding site of arbidol, which enables the flip of the indole ring in one of the three simulated binding modes.

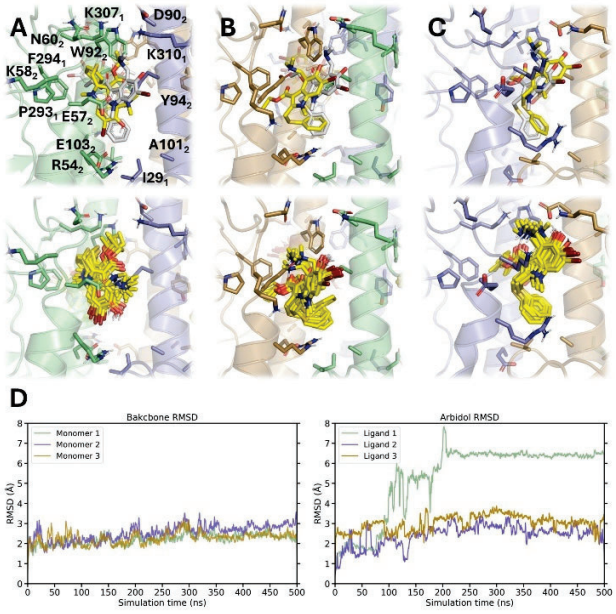


Figure 7. A), B) and C) Above, final binding mode of arbidol (yellow sticks) with respect the crystallographic position (grey sticks) in site A binding pocket of the A/Hong Kong/1/1968 Influenza A strain (H3N2). Below, the superposition of several snapshots along the whole trajectory. D) RMSD plots for each HA monomer backbone atoms and for each of the three ligands.

In light of these results, we explored the binding of **VF-57a** to the arbidol binding site in HA H1 (A/ATCC3/Virginia/2009), which was previously modelled to unfold the last helical turn of the short α-helix in HA2 to simulate the local structure observed in the complex between HA and arbidol (PDB ID 5T6N),[12] and hence enabling the binding of the ligands to the arbidol binding site in the PR8 strain (see Experimental section. Molecular modeling: Homology modelling and setup of simulated systems, and Supporting Information Figure S2). The HA trimer is stable with averaged RMSD values of 2.3 ± 0.4, 2.6 ± 0.5 and 2.7 ± 0.6 Å for each of the HA monomers, although small fluctuations were observed in the last helical turn of the short α-helix in HA2 and loops in the HA1 subunit (Figure 8). The overall structural stability of the HA trimer was confirmed by averaged RMSD values of 1.0 ± 0.2, 1.3 ± 0.2 and 1.2 ± 0.2 Å for the main stem helices for each of the protomers (see SI Figure S3).

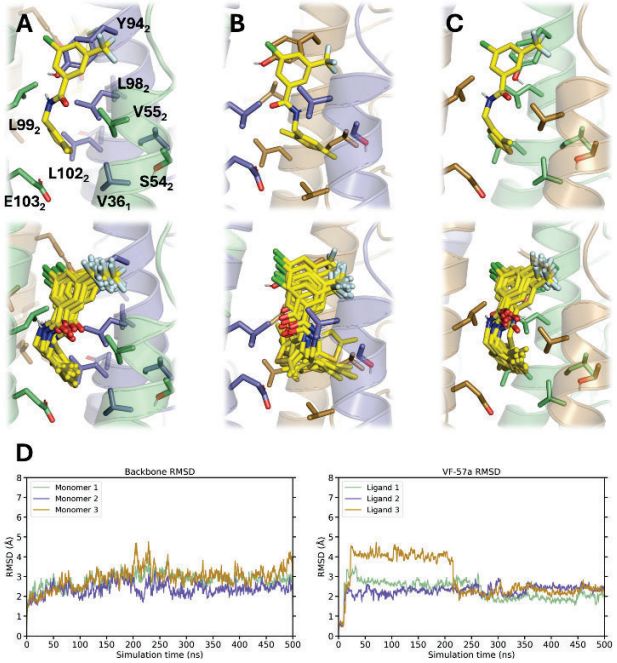


Figure 8. A), B) and C) Above, final position of **VF-57a** (yellow sticks) in site A binding pocket of the A/Virginia/ATCC3/2009 Influenza A strain (H1N1) modified homology model. Below, the superposition of several snapshots along the whole trajectory. D) RMSD plots for each HA monomer backbone atoms and for each of the three ligands.

The initial position of **VF-57a** was rearranged at the beginning of the MD simulation in the three pockets of H1, achieving a stable pose after the first 200 ns with RMSD values of 2.3 ± 0.5, 2.3 ± 0.4 and 2.9 ± 1.0 Å (Figure 8). This rearrangement places the ligand more deeply in the pocket, adopting a relatively stable pose after 250 ns in the three cases. The thiophene ring is surrounded by hydrophobic amino acid residues in the lower part of the pocket, such as $V36_1$, $L49_1$, $P308_1$, $F309_1$, $V55_2$ $L99_2$ from monomer 1, and $L98'_2$ and $L102'_2$ from monomer 2. On the other hand, the phenyl ring containing the trifluoromethyl moiety is stabilized by π-stacking interactions with $Y94_2$ and $F309_1$.

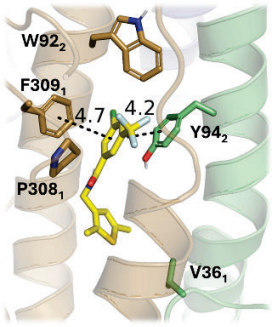
Relative Binding Free Energies (RBFE). To verify whether our predicted binding mode for **VF-57a** in H1 corresponds to its actual binding mode, and to leverage experimental activity data on cells, we conducted Relative Binding Free Energy calculations. These calculations aimed to quantitatively estimate the change in binding between **VF-57a** and **22**, **38**, and **50**, and assess the correlation with experimental data.

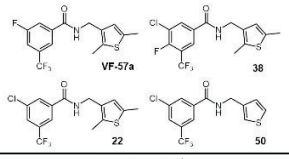
Regarding the transformation from **VF-57a** to **22**, our results indicate that the latter is destabilized by 0.6 kcal/mol (see Figure 9). This agrees with the experimental difference in inhibitory activity determined from antiviral assays for the two compounds (EC_{50} values of 0.92/1.1 (PR8) and 0.31/0.71 (Virg09) μM for **VF-57a**, and 1.1/1.5 (PR8) and 0.44/1.7 (Virg09) μM for **22**; see Table 1). Let us note that this chemical modification occurs in a region of the molecule proximal to the sidechains of $N95_2$ and $W92_2$, as reflected by an average distance of 3.9 ± 0.8 Å between the chlorine atom of **VF-57a** and the side chain of $N95_2$, and by an average distance of 4.0 ±

0.4 Å and 4.1 ± 0.4 Å from the chlorine atom of **VF-57a** and the indole ring of $W92_2$, respectively. Thus, the slight destabilization caused by the replacement of chlorine to fluorine likely reflects a balance between the steric fitting to the side chains of $N95_2$ and $W92_2$, and the modulation of the hydrophobicity of the substituted phenyl ring.

Regarding the transformation from **VF-57a** to **38**, a fluorine atom is introduced in the *para* position relative to the carbonyl group of **VF-57a** (see Figure 9). Our simulations indicate that the latter is stabilized by 0.4 kcal/mol. Within the uncertainty of the experimental data, this estimate agrees with the similar antiviral activity determined for **VF-57a** (see above) and the EC_{50} values measured for **38** (0.27/1.6 (PR8) and 0.18/3.1 (Virg09) µM for **38**; see Table 1). In this transformation, the fluorine atom is interacting with the indole ring of $W92_2$ (the shortest average distance from the H atom in para position of **VF-57a** to the indole ring of $W92_2$ is 3.5 ± 0.7 Å; this value varies to 3.3 ± 0.5 Å from the F atom in **38**).

Finally, the transformation from VF-57a to **50** involves the removal of the two methyl groups of the thiophene moiety (see Figure 9). Our simulations predict that the latter is destabilized by 1.6 kcal/mol. The predicted change in the binding affinity is consistent with the observed reduction in antiviral activity determined for **50** (EC_{50} values of 2.0/1.0 (PR8) and 1.1/2.1 (Virg09) µM; see Table 1). This trend can be justified by taking into account the packing of the thiophene moiety in a hydrophobic environment, and the concomitant decrease in stabilizing van der Waals interactions upon removal of the methyl groups.





ΔΔG (kcal·mol⁻¹)

Change	complex	solvent	ΔΔG
VF-57a → 38	-19.1 ± 0.0	-18.6 ± 0.0	-0.4 ± 0.0
38→ VF-57a	19.1 ± 0.0	18.6 ± 0.0	0.4 ± 0.0
VF-57a → 22	9.3 ± 0.0	8.7 ± 0.0	0.6 ± 0.0
22 → VF-57a	-9.4 ± 0.0	-8.7 ± 0.0	-0.6 ± 0.0
VF-57a → 50	6.2 ± 0.1	4.5 ± 0.0	1.6 ± 0.1
50→ VF-57a	-6.2 ± 0.1	-4.5 ± 0.0	-1.6 ± 0.1

Figure 9. At left, binding mode for **VF-57a** considered to perform the RBFE calculations obtained from our MD simulations. At right, the results of the complex and solvent RBFE calculations in kcal·mol^{-1} as well as the predicted change in free energy for each transformation.

CONCLUSIONS

This study reports the synthesis, antiviral evaluation and mechanism of action of an original series of *N*-[(thiophen-3-yl)methylbenzamides acting as strong inhibitors of group 1 HA-mediated fusion. The lead compound, **VF-57a** exhibited sub-micromolar activity against authentic A/H1N1 virus and a similarly high potency for inhibiting the cell entry of A/H1N1 and A/H5N1 (i.e. highly pathogenic IAV-derived) pseudoviruses.

The strong inhibitory activity of **VF-57a** is proposedly mediated through its binding to a pocket in HA2 that lies close to the binding site of the broad IAV HA inhibitor arbidol. Particularly, our MD calculations predict that the binding of VF-57a is stabilised by π-stacking interactions with $F309_1$ and $Y94_2$ and by hydrophobic interactions between its dimethylthiophene moiety and $V36_1$, $L49_1$, $P308_1$, $F309_1$, $V55_2$, $L99_2$, $L98_2$ and $L101_2$. Our MD and relative binding free energy predictions also rationalise the SAR, strong activity against the Virg09 virus and role of mutation $S54_2P$ in defining resistance to this lead molecule. In combination, these findings warrant further follow-up of **VF-57a** for anti-influenza drug development, including its validation in *in vivo* models.

EXPERIMENTAL SECTION

Chemical synthesis. Commercially available reagents and solvents were used without further purification unless stated otherwise. Preparative normal phase chromatography was performed on a CombiFlash Rf 150 (Teledyne Isco) with pre-packed RediSep Rf silica gel cartridges. Thin-layer chromatography was performed with aluminum-backed sheets with silica gel 60 F254 (Merck, ref 1.05554), and spots were visualized with UV light and 1% aqueous solution of $KMnO_4$. All compounds showed a sharp melting point and a single spot on TLC. Melting points were determined in open capillary tubes with a MFB 595010M Gallenkamp. 400 MHz ^{1}H and 100.6 MHz ^{13}C NMR spectra were recorded on a Varian Mercury 400 or on a Bruker 400 Avance III spectrometers. The chemical shifts are reported in ppm (δ scale) relative to internal tetramethylsilane, and coupling constants are reported in Hertz (Hz). Assignments given for the NMR spectra of selected new compounds have been carried out on the basis of DEPT, COSY ^{1}H/^{1}H (standard procedures), and COSY ^{1}H/^{13}C (gHSQC and gHMBC sequences) experiments. IR spectra were run on Perkin-Elmer Spectrum Two FT-IR spectrophotometer. Absorption values are expressed as wavenumbers (cm^{-1}); only significant absorption bands are given. High-resolution mass spectrometry (HRMS) analyses were performed with an LC/MSD TOF Agilent Technologies spectrometer. The elemental analyses were carried out in a Flash 1112 series Thermofinnigan elemental microanalyzator (AS) to determine C, H and N. The structure of all new compounds was confirmed by elemental analysis and/or accurate mass measurement, IR, ^{1}H NMR and ^{13}C NMR. The analytical samples of all the new compounds, which were subjected to pharmacological evaluation, possessed purity ≥95% as evidenced by their elemental analyses (Table S1) or HPLC/UV. HPLC/UV were determined with a HPLC Agilent 1260 Infinity II LC/MSD coupled to a photodiode array. 5 µL of sample 0.5 mg/mL in methanol:acetonitrile were injected, using an Agilent Poroshell 120 EC-C18, 2.7 µm, 50 mm x 4.6 mm column at 40 °C. The mobile phase was a mixture of A = water with 0.05% formic acid and B = acetonitrile with 0.05% formic acid, with the method described as follows: flow 0.6 mL/min, 5% B-95% A 3 min, 100% B 4 min, 95% B-5% A 1 min. Purity is given as % of absorbance at 220 nm.

General Procedure A. To a solution of (2,5-dimethylthiophen-3-yl)methanamine (1 equiv.) in anh. dichlorometane (3 mL), triethylamine was added (1.2 equiv.). Then, the solution was cooled to 0 °C and a solution of the required commercial acyl chloride (1.1 equiv.) in anh. dichloromethane (2 mL) was slowly added dropwise. The mixture was kept under stirring at room temperature for 5 h. Then

the reaction mixture was concentrated *in vacuo*. Column chromatography (hexane/ethyl acetate mixtures from 0 to 10%) gave the expected product.

General Procedure B. To a solution of the commercially available carboxylic acid (1 equiv.) in dry acetonitrile (5 mL), thionyl chloride (5 equiv.) was added dropwise. The mixture was heated under reflux for 2h. Then, the solvent was evaporated *in vacuo* to yield the crude product, which was used directly in the next step.

To a solution of (2,5-dimethylthiophen-3-yl)methanamine (1 equiv.) in anh. dichlorometane (3 mL), triethylamine was added (1.2 equiv.). Then, the solution was cooled to 0 °C and a solution of the required commercial acyl chloride (1.1 equiv.) in anh. dichloromethane (2 mL) was slowly added dropwise. The mixture was kept under stirring at room temperature for 5 h. Then the reaction mixture was concentrated *in vacuo*. Column chromatography in silica gel (using as eluent mixtures of EtOAc in hexane from 0 to 10%) gave the expected product.

General Procedure C. To a solution of the commercially available carboxylic acid in anh. toluene, thionyl chloride (5 equiv.) followed by some drops of DMF were added dropwise. The mixture was heated under reflux for 2h. Then, the solvent was evaporated *in vacuo* to yield the crude product, which was used directly in the next step.

To a suspension of the required amine and triethylamine in anh. DCM, a solution of the proper acyl chloride in anh. DCM was slowly added. The mixture was kept under stirring at room temperature for 5 h. Then, water (15 mL) followed by EtOAc (15 mL) were added and the mixture was extracted. The organic layer was washed with brine (15 mL), dried over anh. Na_2SO_4 and filtered. Solvents were concentrated *in vacuo* and the resulting crude was purified by column chromatography in silica gel (using as eluent mixtures of EtOAc in hexane from 0% to 10%) to obtain the desired product.

General Procedure D. To a solution of the commercially available carboxylic acid in anh. toluene, thionyl chloride (5 equiv.) followed by some drops of DMF were added dropwise. The mixture was heated under reflux for 2h. Then, the solvent was evaporated *in vacuo* to yield the crude product, which was used directly in the next step.

To a suspension of the required amine and triethylamine in anh. DCM, a solution of the proper acyl chloride in anh. DCM was slowly added. The mixture was kept under stirring at room temperature for 16 h. Then, water (15 mL) followed by EtOAc (15 mL) were added and the mixture was extracted. The organic layer was washed with brine (15 mL), dried over anh. Na_2SO_4 and filtered. Solvents were concentrated *in vacuo* and the resulting crude was purified by column chromatography in silica gel (using as eluent mixtures of EtOAc in hexane from 0% to 10%) to obtain the desired product.

***N*-[(2,5-dimethylthiophen-3-yl)methyl]-3-fluoro-5-(trifluoromethyl)benzamide, 22**. Following general procedure A, 3-fluoro-5-(trifluoromethyl)benzoyl chloride (88 mg, 0.39 mmol) was reacted with (2,5-dimethylthiophen-3-yl)methanamine (50 mg, 0.35 mmol) giving the product as a white solid (101 mg, 86% yield). m.p. 115-116 °C. ν: 3232, 3086, 2922, 1767, 1643, 1602, 1552, 1468, 1446, 1427, 1367, 1346, 1293, 1252, 1220, 1174, 1134, 1125, 1093, 1042, 998, 932, 909, 882, 837, 757, 733, 692 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.39 (s, 6H, —C$\underline{H_3}$), 4.45 (s, 2H, —NH—C$\underline{H_2}$), 6.30 (broad s, 1H, NH), 6.59 (s, 1H, C$\underline{H}$thiophene), 7.45 (d, J= 8.0 Hz, 1H, H(4)arom.), 7.68 (d, J = 8.6 Hz, 1H, H(2)arom.), 7.79 (pseudo s, 1H, H(6)arom.). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.8, 15.0, 37.5, 115.7 (dq, J_{CF} = 24.4, 3.7 Hz), 117.9 (d, J_{CF} = 22.8 Hz), 119.4 (p, J_{CF} = 3.5 Hz), 122.8 (d, J = 271.7 Hz), 126.2, 132.4, 132.9 (dd, J_{CF} = 34.1, 7.9 Hz), 134.3, 136.7, 137.7 (d, J_{CF} = 6.9 Hz), 162.5 (d, J_{CF} = 251.1 Hz), 164.4 (d, J_{CF}= 2.1 Hz).

3-chloro-*N*-[(2,5-dimethylthiophen-3-yl)methyl]-5-(trifluoromethyl)benzamide, 23 (VF-57a). Following general procedure B, 3-chloro-5-(trifluoromethyl)benzoyl chloride (106 mg, 0.44 mmol) was reacted with (2,5-dimethylthiophen-3-yl)methanamine (56 mg, 0.40 mmol) giving the product as a white solid (62 mg, 45% yield). m.p. 121-122 °C. ν: 3290, 3086, 2920, 1638, 1613, 1591, 1581, 1549, 1454, 1433, 1357, 1324, 1288, 1261, 1250, 1177, 1129, 1111, 1039, 980, 921, 889, 839, 830, 739, 715, 699, 689, 665 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.39 (s, 6H, —C$\underline{H_3}$), 4.46 (d, J = 5.2 Hz, 2H, —NH—C$\underline{H_2}$), 6.30 (broad s, 1H, NH), 6.59 (s, 1H, C$\underline{H}$ thiophene), 7.72 (s, 1H, H(4)arom.), 7.89 (s, 1H, H(6)arom.), 7.92 (s, 1H, H(2)arom.). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.8, 15.1, 37.5, 122.1 (d, J_{CF} = 3.5 Hz), 126.2, 128.3 (d, J_F = 3.5 Hz, 2C), 130.6, 132.4, 132.5, 134.3, 135.6 (d, J_{CF} = 35.3 Hz), 136.7, 136.9, 164.4.

3-bromo-*N*-[(2,5-dimethylthiophen-3-yl)methyl]-5-(trifluoromethyl)benzamide, 24. Following general procedure B, 3-bromo-5-(trifluoromethyl)benzoyl chloride (150 mg, 0.39 mmol) was reacted with (2,5-dimethylthiophen-3-yl)methanamine (50mg, 0.35mmol) giving the product as a white solid (113 mg, 82% yield). m.p. 124-125 °C. ν: 3289, 3082, 2919, 1636, 1614, 1581, 1547, 1448, 1432, 1356, 1321, 1285, 1263, 1250, 1174, 1130, 1110, 1038, 977, 924, 888, 839, 807, 723, 698, 687, 665 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.39 (s, 6H, —C$\underline{H_3}$), 4.46 (d, J = 5.2 Hz, 2H, —NH—C$\underline{H_2}$), 6.28 (broad s, 1H, NH), 6.59 (s, 1H, C$\underline{H}$thiophene), 7.87 (s, 1H, H(4)arom.), 7.94 (s, 1H, H(6)arom.), 8.07 (s, 1H, H(2)arom.). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.9, 15.1, 37.6, 122.8 (d, J_{CF} = 274.7 Hz), 122.7 (d, J_{CF} = 4.2 Hz), 123.2, 126.3, 131.2 (d, J_{CF} = 4.2 Hz), 132.4, 132.8 (d, J_{CF} = 31.8 Hz), 133.5, 134.3, 136.7, 137.1, 164.3.

***N*-[(2,5-dimethylthiophen-3-yl)methyl]-3,5-bis(trifluoromethyl)benzamide, 25.** Following general procedure B, 3,5-*bis(trifluoromethyl*)benzoyl chloride (130 mg, 0.47 mmol) was reacted with (2,5-dimethylthiophen-3-yl)methanamine (60 mg, 0.43 mmol) giving the product as a white solid (100 mg, 70% yield). m.p. 163-165 °C. ν: 3233, 3090, 2930, 1644, 1617, 1552, 1429, 1385, 1350, 1329, 1286, 1276, 1250, 1183, 1168, 1158, 1131, 1040, 987, 909, 871, 837, 756, 734, 699, 682 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.39 (s superimposed, 3H, —C$\underline{H_3}$), 2.40 (s superimposed, 3H, —C$\underline{H_3}$), 4.49 (d, J = 5.2 Hz, 2H, —NH—C$\underline{H_2}$), 6.38 (broad s, 1H, NH), 6.60 (s, 1H, C$\underline{H}$thiophene), 8.00 (s, 1H, H(4)arom.), 8.21 (s, 2H, H(2,6)arom.). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.8, 15.1, 37.6, 122.9 (d, 2C, J_{CF} = 272.6 Hz), 125.0 (t, J_{CF} = 3.6 Hz), 126.2, 127.3, 127.3 (d, 2C, J_{CF} = 2.5 Hz), 132.2 (q, J_{CF} = 34.3 Hz), 132.3, 134.4, 136.4, 136.8, 164.2.

3,5-difluoro-*N*-[(2,5-dimethylthiophen-3-yl)methyl]benzamide, 26. Following general procedure B, 3,5-difluorobenzoyl

chloride (118 mg, 0.47 mmol) was reacted with (2,5-dimethylthiophen-3-yl)methanamine (60 mg, 0.43 mmol) giving the product as a white solid (35 mg, 29% yield). m.p. 120-121 °C. ν: 3249, 3092, 3068, 2924, 2868, 1635, 1591, 1542, 1465, 1435, 1354, 1323, 1310, 1257, 1226, 1214, 1146, 1125, 1111, 1057, 1042, 984, 899, 878, 868, 853, 836, 820, 783, 761, 717, 662, 643 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.38 (s, 6H, —C$\underline{H_3}$), 4.43 (d, J = 5.2 Hz, 2H, —NH—C$\underline{H_2}$), 6.25 (broad s, 1H, NH), 6.57 (s, 1H, C$\underline{H}$thiophene), 6.93 (t, J = 8.2 Hz, 1H, H(4)arom.), 7.28 (s, 2H, H(2,6)arom.). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.8, 15.1, 37.5, 106.8 (t, J_{CF} = 27.4 Hz), 110.2 (d, 2C, J_{CF} = 26.8 Hz), 126.2, 132.6, 134.1, 136.6, 137.6, 161.7 (d, J_{CF} = 252.1 Hz), 162.9 (d, J_{CF} = 251.2 Hz), 164.7.

3,5-dichloro-*N*-[(2,5-dimethylthiophen-3-yl)methyl]benzamide, 27. By following general procedure A, 3,5-dichlorobenzoyl chloride (81 mg, 0.39 mmol) was reacted with with (2,5-dimethylthiophen-3-yl)methanamine (50 mg, 0.35 mmol) giving the product as a white solid (56 mg, 46% yield). m.p. 136-137 °C. ν: 3233, 3058, 2916, 2865, 1630, 1592, 1566, 1538, 1458, 1433, 1351, 1304, 1283, 1216, 1146, 1115, 1096, 1053, 935, 889, 864, 820, 801, 762, 745, 700, 685, 660, 638 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.39 (s, 3H, —C$\underline{H_3}$), 2.38 (s, 3H, —C$\underline{H_3}$), 4.44 (d, J = 5.2 Hz, 2H, —NH—C$\underline{H_2}$), 6.15 (broad s, 1H, NH), 6.58 (s, 1H, C$\underline{H}$thiophene), 7.47 (t, J = 1.8 Hz, 1H, H(4)arom.), 7.66-7.60 (m, 2H, H(2,6)arom.). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.8, 15.1, 37.5, 125.6 (2C), 126.2, 131.4, 132.5, 134.2, 135.5 (2C), 136.7, 137.3, 164.6.

3,5-dibromo-*N*-[(2,5-dimethylthiophen-3-yl)methyl]benzamide, 28. Following general procedure B, 3,5-dibromobenzoyl chloride (178 mg, 0.60 mmol) was reacted with (2,5-dimethylthiophen-3-yl)methanamine (70 mg, 0.50 mmol) giving the product as a white solid (101 mg, 46% yield). m.p. 134-135 °C. ν: 3233, 3058, 2916, 2865, 1630, 1592, 1566, 1538, 1458, 1433, 1351, 1304, 1283, 1216, 1146, 1115, 1096, 1053, 935, 889, 864, 820, 801, 762, 745, 700, 685, 660, 638 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.39 (s, 3H, —C$\underline{H_3}$), 2.40 (s, 3H, —C$\underline{H_3}$), 4.44 (d, J = 5.2 Hz, 2H, —NH—C$\underline{H_2}$), 6.09 (broad s, 1H, NH), 6.60 – 6.52 (m, 1H, C$\underline{H}$thiophene), 7.78 (t, J = 1.7 Hz, 1H, H(4)arom.), 7.81 (d, J = 1.7 Hz, 2H, H(2,6)arom.).

3-chloro-*N*-[(2,5-dimethylthiophen-3-yl)methyl]-5-methylbenzamide, 29. Following general procedure B, 3-chloro-5-methylbenzoyl chloride (61 mg, 0.32 mmol) was reacted with (2,5-dimethylthiophen-3-yl)methanamine (35 mg, 0.25 mmol) giving the product as a white solid (65 mg, 90% yield). m.p. 102-103 °C. ν: 3282, 2918, 2856, 1628, 1600, 1578, 1527, 1455, 1436, 1371, 1346, 1316, 1289, 1223, 1146, 1056, 1035, 994, 864, 770, 748, 677, 644 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.36 (s, 3H, C$\underline{H_3}$—C(5)arom.), 2.38 (s superimposed, 3H, —C$\underline{H_3}$), 2.39 (s superimposed, 3H, —C$\underline{H_3}$), 4.43 (d, J = 5.3 Hz, 2H, —NH—C$\underline{H_2}$), 6.19 (broad s, 1H, NH), 6.56 (s, 1H, C$\underline{H}$ thiophene), 7.28 (s, 1H, H(4)arom.), 7.45 (s, 1H, H(6)arom.), 7.51 (s, 1H, H(2)arom.). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.8, 15.2, 21.1, 37.4, 124.2, 125.9, 126.3, 132.0, 132.9, 133.9, 134.4, 135.9, 136.5, 140.3, 166.0.

***N*-[(2,5-dimethylthiophen-3-yl)methyl]-3-fluorobenzamide, 30.** Following general procedure A, 3-fluorobenzoyl chloride (63 mg, 0.39 mmol) was reacted with (2,5-dimethylthiophen-3-yl)methanamine (50 mg, 0.35 mmol) giving the product as a white solid (93 mg, 94% yield). m.p. 93-94 °C. ν: 3245, 3072, 2920, 1632, 1587, 1548, 1485, 1445, 1426, 1358, 1316, 1306, 1274, 1256, 1225, 1144, 1117, 1039, 980, 904, 891, 832, 808, 787, 726, 708, 674, 666 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.38 (s, 6H, —C$\underline{H_3}$), 4.45 (d, J = 5.2 Hz, 2H, —NH—C$\underline{H_2}$), 6.26 (broad s, 1H, NH), 6.59 (s, 1H, C$\underline{H}$ thiophene), 7.23-7.13 (m, J = 8.2, 0.8 Hz, 1H, H(4)arom.), 7.43-7.33 (m, J = 8.1, 2.1 Hz, 1H, H(5)arom.), 7.53–7.46 (m, 2H, H(2,6)arom.). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.8, 15.1, 37.4, 114.4 (d, J_{CF} = 23.4 Hz), 118.5 (d, J_{CF} = 21.1 Hz), 122.4, 126.3, 130.2 (d, J_{CF} = 8.1 Hz), 132.9, 133.9, 136.5, 136.7, 162.7 (d, J_{CF} = 248.4 Hz), 165.9.

***N*-[(2,5-dimethylthiophen-3-yl)methyl]-3-(trifluoromethyl)benzamide, 31.** Following general procedure A, 3-(trifluoromethyl) benzoyl chloride (81 mg, 0.39 mmol) was reacted with (2,5-dimethylthiophen-3-yl)methanamine (50 mg, 0.35 mmol) giving the product as a white solid (63 mg, 54% yield). m.p. 111-112 °C. ν: 3225, 3071, 2924, 1636, 1542, 1487, 1443, 1427, 1360, 1336, 1319, 1301, 1281, 1256, 1213, 1166, 1115, 1071, 1040, 975, 921, 837, 817, 785, 735, 690, 653 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.39 (s, 6H, —CH_3), 4.47 (d, J = 5.2 Hz, 2H, —NH—CH_2), 6.28 (broad s, 1H, NH), 6.60 (s, 1H, CH thiophene), 7.56 (t, J = 7.8 Hz, 1H, H(5)arom.), 7.75 (dt, J = 7.9, 0.7 Hz, 1H, H(6)arom.), 7.95 (d, J = 7.8, 1H, H(4)arom.), 8.03 (s, 1H, H(2)arom.). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.81, 15.07, 37.47, 123.68 (d, J_{CF} = 273 Hz), 123.95 (q, J_{CF} = 3.8 Hz), 126.29, 128.09 (q, J_{CF} = 3.8 Hz), 129.19, 130.23, 131.13 (d, J_{CF} = 33.0 Hz), 132.7, 134.1, 135.2, 136.6, 165.7.

3-chloro-*N*-[(2,5-dimethylthiophen-3-yl)methyl]benzamide, 32. Following general procedure A, 3-chlorobenzoyl chloride (62 mg, 0.39 mmol) was reacted with (2,5-dimethylthiophen-3-yl)methanamine (50 mg, 0.35 mmol) giving the product as a white solid (74 mg, 79% yield). m.p. 89-90 °C. ν: 3240, 3071, 2916, 1631, 1549, 1474, 1430, 1357, 1297, 1255, 1212, 1162, 1145, 1077, 1040, 974, 911, 832, 808, 728, 707, 660 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.38 (s, 6H, —C$\underline{H_3}$), 4.44 (d, J = 5.2 Hz, 2H, —NH—C$\underline{H_2}$), 6.26 (broad s, 1H, NH), 6.58 (s, 1H, C$\underline{H}$thiophene), 7.34 (t, J = 7.9 Hz, 1H, H(5)arom.), 7.45 (ddd, J = 7.9, 2.0, 1.0 Hz, 1H, H(4)arom.), 7.63 (dt, J = 7.8, 1.3 Hz, 1H, H(6)arom.), 7.75 (t, J = 1.7 Hz, 1H, H(2)arom.) ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.8, 15.1, 37.4, 125.0, 126.3, 127.3, 129.9, 131.5, 132.8, 133.9, 134.7, 136.2, 136.5, 165.8.

3-bromo-*N*-[(2,5-dimethylthiophen-3-yl)methyl]benzamide, 33. Following general procedure A, 3-fluorobenzoyl chloride (68 mg, 0.31mmol) was reacted with (2,5-dimethylthiophen-3-yl)methanamine (40 mg, 0.28 mmol) giving the product as a white solid (55 mg, 60% yield). m.p. 91-92 °C. ν: 3233, 3067, 2948, 1631, 1549, 1471, 1428, 1357, 1314, 1295, 1255, 1211, 1146, 1072, 1039, 971, 912, 833, 808, 731, 709 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.38 (s, 6H, —C$\underline{H_3}$), 4.44 (d, J = 5.2 Hz, 2H, —NH—C$\underline{H_2}$), 6.21 (broad s, 1H, NH), 6.58 (s, 1H, C$\underline{H}$thiophene), 7.29 (t, J = 7.8 Hz, 1H, H(5)arom.), 7.61 (dd, J = 8.0, 2.0 Hz, 1H, H(4)arom.), 7.68 (dd, J = 7.8, 1.6 Hz, 1H, H(6)arom.), 7.90 (t, J = 1.7 Hz, 1H, H(2)arom.). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.8, 15.1, 37.4, 122.7, 125.5, 126.3, 130.1 (2C), 132.8, 133.9, 134.4, 136.4, 136.5, 165.7.

***N*-[(2,5-dimethylthiophen-3-yl)methyl]-3-nitrobenzamide, 34.** Following general procedure A, 3-nitrobenzoyl chloride (72 mg, 0.39 mmol) was reacted with (2,5-dimethylthiophen-3-

yl)methanamine (50 mg, 0.35 mmol) giving the product as a white solid (86 mg, 84% yield). m.p. 156-157 °C. *ν*: 3233, 3086, 2911, 1635, 1618, 1580, 1552, 1523, 1474, 1426, 1349, 1322, 1307, 1275, 1256, 1214, 1145, 1086, 1042, 977, 921, 885, 842, 816, 750, 723, 712, 702 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.38 (s, 3H, —C$\underline{H_3}$), 2.40 (s, 3H, —C$\underline{H_3}$), 4.49 (d, *J* = 5.4 Hz, 2H, —NH—C$\underline{H_2}$), 6.44 (broad s, 1H, NH), 6.60 (s, 1H, C$\underline{H}$thiophene), 7.63 (t, *J* = 8.0 Hz, 1H, H(5)arom.), 8.16 (dt, *J* = 7.7, 1.2 Hz, 1H, H(6)arom.), 8.34 (ddd, *J* = 8.2, 2.0, 0.9 Hz, 1H, H(4)arom.), 8.57 (t, *J* = 1.8 Hz, 1H, H(2)arom.). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.8, 15.1, 37.6, 121.74, 126.1, 126.3, 129.8, 132.5, 133.3, 134.2, 135.9, 136.7, 148.2, 164.7.

***N*-((2,5-dimethylthiophen-3-yl)methyl)-3,4-difluoro-5-(trifluoromethyl)benzamide, 35.** Following general procedure C, 3,4-difluoro-5-(trifluoromethyl)benzoic acid (106 mg, 0.47 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (170 µL, 279 mg, 2.35 mmol). Then, (2,5-dimethylthiophen-3-yl)methanamine hydrochloride (100 mg, 0.56 mmol) and triethylamine (234 µL, 170 mg, 1.68 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride (115 mg, 0.47 mmol) in anh. DCM (1.0 mL) giving the product as a beige solid (38 mg, 23% yield). m.p. 130-131 °C. *ν*: 3239, 1647, 1627, 1552, 1504, 1436, 1371, 1349, 1295, 1276, 1192, 1139, 1042, 1001, 942, 885, 838, 731, 674 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.38 (m, 6H, 2'-C$\underline{H_3}$, 5'-C$\underline{H_3}$), 4.44 (d, *J* = 5.2 Hz, 2H, C$\underline{H_2}$), 6.34 (broad s, 1H, N$\underline{H}$), 6.57 (s, 1H, 4'-H), 7.78 (m, 1H, 6-H), 7.85 (ddd, *J* = 9.6 Hz, *J'* = 6.9 Hz, *J''* = 2.2 Hz, 1H, 2-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.9 (C2'-$\underline{C}H_3$), 15.2 (C5'-$\underline{C}H_3$), 37.8 (CH_2, $\underline{C}H_2$), 120.5 ($^2J_{CF}$ = 18.8 Hz, CH, C2), 120.7 (m, CH, C6), 120.8 (qd, $^2J_{CF}$ = 33.9 Hz, $^2J_{CF}$ = 9.6 Hz, C, C5), 121.7 (qd, $^1J_{CF}$ = 272.9 Hz, $^4J_{CF}$ = 3.5 Hz, C, $\underline{C}F_3$), 126.3 (CH, C4'), 131.4 (t, $^3J_{CF}$ = 4.8 Hz, C, C1), 132.5 (C, C3'), 134.4 (C, C2'), 136.9 (C, C5'), 150.2 (dd, $^1J_{CF}$ = 263.7 Hz, $^2J_{CF}$ = 13.7 Hz, C, C3), 150.8 (dd, $^1J_{CF}$ = 251.1 Hz, $^2J_{CF}$ = 9.6 Hz, C, C4), 163.8 (C, $\underline{C}O$). HRMS: Calcd for $[C_{15}H_{12}F_5NOS+H]^+$: 350.0633, found: 350.0636.

***N*-((2,5-dimethylthiophen-3-yl)methyl)-2,3,4-trifluoro-5-(trifluoromethyl)benzamide, 36.** Following general procedure C, 2,3,4-trifluoro-5-(trifluoromethyl)benzoic acid (150 mg, 0.61 mmol) in anh. toluene (4.0 mL) and drops of DMF was reacted with thionyl chloride (445 µL, 726 mg, 6.14 mmol). Then, (2,5-dimethylthiophen-3-yl)methanamine hydrochloride (131 mg, 0.74 mmol) and triethylamine (411 µL, 298 mg, 2.95 mmol) in anh. DCM (2.0 mL) were reacted with the crude acyl chloride (161 mg, 0.61 mmol) in anh. DCM (1.0 mL) giving the product as a white solid (110 mg, 49% yield). m.p. 109-110 °C. *ν*: 3282, 2925, 1644, 1553, 1487, 1369, 1280, 1205, 1168, 1133, 1054, 994, 905, 721, 709, 672, 632, 574 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.39 (m, 6H, 2'-C$\underline{H_3}$, 5'-C$\underline{H_3}$), 4.49 (d, *J* = 5.3 Hz, 2H, C$\underline{H_2}$), 6.58 (s, 1H, 4'-H), 6.60 (s, 1H, N$\underline{H}$), 8.23 (td, *J* = 7.5 Hz, *J'* = 1.7 Hz, 1H, 6-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 13.0 (C2'-$\underline{C}H_3$), 15.2 (C5'-$\underline{C}H_3$), 37.8 (CH_2, $\underline{C}H_2$), 116.8 (m, C, C5), 118.7 (d, $^2J_{CF}$ = 10.1 Hz, C, C1), 121.4 (q, $^1J_{CF}$ = 272.0 Hz, C, $\underline{C}F_3$), 123.9 (CH, C6), 126.2 (CH, C4'), 132.3 (C, C3'), 134.3 (C, C2'), 136.9 (C, C5'), 140.5 (ddd, $^1J_{CF}$ = 256.5 Hz, $^2J_{CF}$ = 17.5 Hz, $^3J_{CF}$ = 2.6 Hz, C, C2), 150.9 (dd, $^1J_{CF}$ = 265.7 Hz, $^2J_{CF}$ = 12.0 Hz, C, C3), 151.85 (ddd, $^1J_{CF}$ = 259.1 Hz, $^2J_{CF}$ = 11.0 Hz, $^3J_{CF}$ = 5.3 Hz, C, C4), 159.9 (C, $\underline{C}O$). HRMS: Calcd for $[C_{15}H_{11}F_6NOS+H]^+$: 368.0538, found: 368.0532.

3-chloro-*N*-((2,5-dimethylthiophen-3-yl)methyl)-2-fluoro-5-(trifluoromethyl)benzamide, 37. Following general procedure C, 3-chloro-2-fluoro-5-(trifluoromethyl)benzoic acid (68 mg, 0.28 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (101 µL, 166 mg, 1.40 mmol). Then, (2,5-dimethylthiophen-3-yl)methanamine hydrochloride (50 mg, 0.28 mmol) and triethylamine (117 µL, 85 mg, 0.84 mmol) in anh. DCM (0.5 mL), were reacted with the crude acyl chloride (73 mg, 0.28 mmol) in anh. DCM (0.5 mL) giving the product as a white solid (98 mg, 95% yield). m.p. 118-119 °C. *ν*: 3285, 2920, 1633, 1544, 1345, 1328, 1277, 1161, 1144, 1122, 1055, 896, 867, 767, 723, 661, 648, 578 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.40 (m, 6H, 2'-C$\underline{H_3}$, 5'-C$\underline{H_3}$), 4.50 (d, *J* = 5.2 Hz, 2H, C$\underline{H_2}$), 6.59 (s, 1H, 4'-H), 6.68 (broad s, 1H, N$\underline{H}$), 7.80 (dd, *J* = 6.5 Hz, *J'* = 2.4 Hz, 1H, 4-H), 8.30 (dd, *J* = 6.2 Hz, *J'* = 2.5 Hz, 1H, 6-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 13.0 (C2'-$\underline{C}H_3$), 15.2 (C5'-$\underline{C}H_3$), 37.9 (CH_2, $\underline{C}H_2$), 122.8 (q, $^1J_{CF}$ = 272.6 Hz, C, $\underline{C}F_3$), 123.1 (d, $^2J_{CF}$ = 21.4 Hz, C, C1), 123.6 (d, $^2J_{CF}$ = 13.7 Hz, C, C3), 126.2 (CH, C4'), 128.1 (m, C, C5), 128.2 (p, $^3J_{CF}$ = 3.6 Hz, CH, C6), 130.8 (CH, C4), 132.4 (C, C3'), 134.3 (C, C2'), 136.8 (C, C5'), 157.8 (d, $^1J_{CF}$ = 253.7 Hz, C, C2), 160.8 (d, $^3J_{CF}$ = 3.5 Hz, C, $\underline{C}O$). HRMS: Calcd for $[C_{15}H_{12}ClF_4NOS+H]^+$: 366.0337, found: 366.0339.

3-chloro-*N*-((2,5-dimethylthiophen-3-yl)methyl)-4-fluoro-5-(trifluoromethyl)benzamide, 38. Following general procedure C, 3-chloro-4-fluoro-5-(trifluoromethyl)benzoic acid (68 mg, 0.28 mmol) in anh. toluene (1.0 mL) and drops of DMF was reacted with thionyl chloride (101 µL, 166 mg, 1.40 mmol). Then, (2,5-dimethylthiophen-3-yl)methanamine hydrochloride (50 mg, 0.28 mmol) and triethylamine (117 µL, 85 mg, 0.84 mmol) in anh. DCM (0.5 mL), were reacted with the crude acyl chloride (73 mg, 0.28 mmol) in anh. DCM (0.5 mL) giving the product as a white solid (43 mg, 42% yield). m.p. 133-134 °C. *ν*: 3273, 1636, 1552, 1481, 1417, 1327, 1317, 1262, 1221, 1140, 1037, 919, 904, 833, 744, 716, 672, 626, 573 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.38 (m, 6H, 2'-C$\underline{H_3}$, 5'-C$\underline{H_3}$), 4.44 (dd, *J* = 5.3 Hz, *J'* = 2.0 Hz, 2H, C$\underline{H_2}$), 6.29 (broad s, 1H, N$\underline{H}$), 6.57 (s, 1H, 4'-H), 7.92 (dd, *J* = 5.9 Hz, *J'* = 2.2 Hz, 1H, 6-H), 8.04 (dd, *J* = 6.3 Hz, *J'* = 2.2 Hz, 1H, 2-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 13.0 (C2'-$\underline{C}H_3$), 15.2 (C5'-$\underline{C}H_3$), 37.8 (CH_2, $\underline{C}H_2$), 120.2 (qd, $^2J_{CF}$ = 34.0 Hz, $^2J_{CF}$ = 13.0 Hz, C, C5), 121.8 (q, $^1J_{CF}$ = 273.3 Hz, C, $\underline{C}F_3$), 123.6 (d, $^2J_{CF}$ = 17.6 Hz, C, C3), 124.5 (m, CH, C6), 126.3 (CH, C4'), 131.4 (d, $^4J_{CF}$ = 4.6 Hz, C, C1), 132.5 (C, C3'), 133.5 (CH, C2), 134.4 (C, C2'), 136.9 (C, C5'), 157.3 (d, $^1J_{CF}$ = 264.0 Hz, C, C4), 163.80 (C, $\underline{C}O$). HRMS: Calcd for $[C_{15}H_{12}ClF_4NOS+H]^+$: 366.0337, found: 366.0335.

2,5-dichloro-*N*-((2,5-dimethylthiophen-3-yl)methyl)-3-(trifluoromethyl)benzamide, 39. Following general procedure C, 2,5-dichloro-3-(trifluoromethyl)benzoic acid (122 mg, 0.47 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (170 µL, 279 mg, 2.35 mmol). Then, (2,5-dimethylthiophen-3-yl)methanamine hydrochloride (100 mg, 0.56 mmol) and triethylamine (234 µL, 170 mg, 1.68 mmol) in anh. DCM (1.0 mL) were reacted with the crude acyl chloride (130 mg, 0.47 mmol) in anh. DCM (1.0 mL) giving the product as a white solid (110 mg, 49% yield). m.p. 157-158 °C. *ν*: 3268, 1646, 1544, 1482, 1429, 1316, 1289, 1258, 1172, 1142, 1051, 886, 830, 742, 711, 689, 830, 576 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.39 (m, 6H, 2'-C$\underline{H_3}$, 5'-C$\underline{H_3}$),

4.46 (d, J = 5.2 Hz, 2H, C$\underline{H}_2$), 6.03 (s, broad, 1H, N$\underline{H}$), 6.59 (s, 1H, 4'-H), 7.67 (d, J = 2.5 Hz, 1H, 2-H), 7.72 (d, J = 2.5 Hz, 1H, 4-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 13.0 (C2'-$\underline{C}H_3$), 15.2 (C5'-$\underline{C}H_3$), 37.7 (CH_2, $\underline{C}H_2$), 121.9 (q, $^1J_{CF}$ = 274.2 Hz, C, $\underline{C}F_3$), 126.3 (CH, C4'), 127.5 (C, C6), 129.2 (q, $^3J_{CF}$ = 5.6 Hz, CH, C4), 130.9 (q, $^2J_{CF}$ = 32.0 Hz, C, C5), 132.1 (C, C3'), 132.6 (CH, C2), 133.5 (C, C3), 134.5 (C, C2'), 136.8 (C, C5'), 139.8 (C, C1), 164.3 (C, $\underline{C}O$). HRMS: Calcd for $[C_{15}H_{12}Cl_2F_3NOS+H]^+$: 382.0042, found: 382.0047.

***N*-((2,5-dimethylthiophen-3-yl)methyl)benzamide, 40.** Following general procedure D, benzoic acid (41 mg, 0.34 mmol) in anh. toluene (1.0 mL) and drops of DMF was reacted with thionyl chloride (245 µL, 402 mg, 3.38 mmol). Then, (2,5-dimethylthiophen-3-yl)methanamine hydrochloride (50 mg, 0.28 mmol) and triethylamine (118 µL, 85 mg, 0.84 mmol) in anh. DCM (1.0 mL) reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a white solid (26 mg, 39% yield). m.p. 98-99 °C. *v*: 3237, 1627, 1536, 1490, 1356, 1307, 1253, 1210, 1141, 1038, 961, 931, 830, 807, 718, 694, 568 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.38 (s, 6H, 2'-C$\underline{H}_3$, 5'-C$\underline{H}_3$), 4.46 (d, J = 5.2 Hz, 2H, C$\underline{H}_2$), 6.24 (broad s, 1H, N$\underline{H}$), 6.60 (m, 1H, 4'-H), 7.42 [m, 2H, 3(5)-H], 7.49 (m, 1H, 4-H), 7.77 [m, 2H, 2(6)-H]. ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.9 (C2'-$\underline{C}H_3$), 15.2 (C5'-$\underline{C}H_3$), 37.5 (CH_2, $\underline{C}H_2$), 126.5 (CH, C4'), 127.1 [CH, C2(6)], 128.7 [CH, C3(5)], 131.6 (CH, C4), 133.3 (C, C3'), 133.9 (C, C1), 134.5 (C, C2'), 136.5 (C, C5'), 167.3 (C, $\underline{C}O$). HRMS: Calcd for $[C_{14}H_{15}NOS+H]^+$: 246.0947, found: 246.0944.

4-chloro-*N*-((2,5-dimethylthiophen-3-yl)methyl)benzamide, 41. Following general procedure D, 4-chlorobenzoic acid (53 mg, 0.34 mmol) in anh. toluene (1.0 mL) and drops of DMF was reacted with thionyl chloride (245 µL, 402 mg, 3.38 mmol). Then, (2,5-dimethylthiophen-3-yl)methanamine hydrochloride (50 mg, 0.28 mmol) and triethylamine (118 µL, 85 mg, 0.84 mmol) in anh. DCM (1.0 mL) reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as an off-white solid (59 mg, 75% yield). m.p. 133-134 °C. *v*: 3313, 2915, 1715, 1632, 1544, 1486, 1447, 1357, 1318, 1214, 1090, 1039, 1013, 851, 514, 759, 708, 635, 570 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.37-2.39 (complex signal, 6H, 2'-C$\underline{H}_3$, 5'-C$\underline{H}_3$), 4.44 (d, J = 5.2 Hz, 2H, C$\underline{H}_2$), 6.19 (broad s, 1H, N$\underline{H}$), 6.58 (q, J = 1.1 Hz, 1H, 4'-H), 7.38 [d, J = 8.8 Hz, 2H, 3(5)-H], 7.70 [d, J = 8.8 Hz, 2H, 2(6)-H]. ^{13}C NMR (101 MHz, $CDCl_3$) δ: 13.0 (C2'-$\underline{C}H_3$), 15.2 (C5'-$\underline{C}H_3$), 37.5 (CH_2, $\underline{C}H_2$), 126.4 (CH, C4'), 128.5 [CH, C2(6)], 128.9 [CH, C3(5)], 132.9 (C, C1), 133.1 (C, C3'), 134.1 (C, C2'), 136.6 (C, C5'), 137.9 (C, C4), 166.2 (C, $\underline{C}O$). HRMS: Calcd for $[C_{14}H_{14}ClNOS+H]^+$: 280.0557, found: 280.0558.

***N*-((2,5-dimethylthiophen-3-yl)methyl)-4-methylbenzamide, 42.** Following general procedure D, 4-methylbenzoic acid (46 mg, 0.34 mmol) in anh. toluene (1.0 mL) and drops of DMF was reacted with thionyl chloride (245 µL, 402 mg, 3.38 mmol). Then, (2,5-dimethylthiophen-3-yl)methanamine hydrochloride (50 mg, 0.28 mmol) and triethylamine (118 µL, 85 mg, 0.84 mmol) in anh. DCM (1.0 mL) reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a beige solid (39 mg, 53% yield). m.p. 132-133 °C. *v*: 277, 2917, 1631, 1612, 1532, 1501, 1348, 1292, 1279, 1211, 1187, 1140, 1118, 1062, 1021, 966, 907, 835, 752, 655, 632, 607, 571 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.38-2.38 (complex signal, 9H, 4-C$\underline{H}_3$, 2'-C$\underline{H}_3$, 5'-C$\underline{H}_3$), 4.44 (d, J = 5.2 Hz, 2H, C$\underline{H}_2$), 6.25 (broad s, 1H, N$\underline{H}$), 6.59 (q, J = 1.2 Hz, 1H, 4'-H), 7.21 [dd, J = 8.6, 0.7 Hz, 2H, 3(5)-H], 7.66 [d, J = 8.2 Hz, 2H, 2(6)-H]. ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.9 (C2'-$\underline{C}H_3$), 15.2 (C5'-$\underline{C}H_3$), 21.5 (C4-$\underline{C}H_3$), 37.4 (CH_2, $\underline{C}H_2$), 126.5 (CH, C4'), 127.0 [CH, C2(6)], 129.3 [CH, C3(5)], 131.6 (C, C1), 133.5 (C, C3'), 133.8 (C, C2'), 136.4 (C, C5'), 142.0 (C, C4), 167.2 (C, $\underline{C}O$). HRMS: Calcd for $[C_{15}H_{17}NOS+H]^+$: 260.1104, found: 260.1109.

***N*-((2,5-dimethylthiophen-3-yl)methyl)-4-methoxybenzamide, 43.** Following general procedure D, 4-methoxybenzoic acid (83 mg, 0.54 mmol) in anh. toluene (1.0 mL) and drops of DMF was reacted with thionyl chloride (391 µL, 641 mg, 5.39 mmol). Then, (2,5-dimethylthiophen-3-yl)methanamine hydrochloride (80 mg, 0.45 mmol) and triethylamine (211 µL, 152 mg, 1.50 mmol) in anh. DCM (1.0 mL) reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a brownish solid (46 mg, 38% yield). m.p. 117-118 °C. *v*: 3309, 2914, 2838, 1633, 1606, 1549, 1504, 1462, 1317, 1252, 1214, 1174, 1112, 1031, 974, 844, 825, 770, 722, 661, 632, 606 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.36-2.39 (complex signal, 6H, 2'-C$\underline{H}_3$, 5'-C$\underline{H}_3$), 3.83 (s, OC$\underline{H}_3$), 4.43 (d, J = 5.2 Hz, 2H, C$\underline{H}_2$), 6.17 (broad s, 1H, N$\underline{H}$), 6.60 (m, 1H, 4'-H), 6.90 [d, J = 8.8 Hz, 2H, 3(5)-H], 7.73 [d, J = 8.9 Hz, 2H, 2(6)-H]. ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.9 (C2'-$\underline{C}H_3$), 15.2 (C5'-$\underline{C}H_3$), 37.4 (CH_2, $\underline{C}H_2$), 55.5 (CH_3, O$\underline{C}H_3$), 113.8 [CH, C3(5)], 126.6 (CH, C4'), 126.8 (C, C1), 128.9 [CH, C2(6)], 133.5 (C, C3'), 133.8 (C, C2'), 136.4 (C, C5'), 162.3 (C, C4), 166.8 (C, $\underline{C}O$). HRMS: Calcd for $[C_{15}H_{17}NO_2S+H]^+$: 279.1053, found: 279.1059.

3,4-dichloro-*N*-((2,5-dimethylthiophen-3-yl)methyl)benzamide, 44. Following general procedure D, 3,4-dichlorobenzoic acid (90 mg, 0.47 mmol) in anh. toluene (1.0 mL) and drops of DMF was reacted with thionyl chloride (339 µL, 556 mg, 4.67 mmol). Then, (2,5-dimethylthiophen-3-yl)methenamine hydrochloride (70 mg, 0.39 mmol) and triethylamine (164 µL, 118 mg, 1.17 mmol) in anh. DCM (1.0 mL) reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a brownish solid (43 mg, 35% yield). m.p. 102-103 °C. *v*: 3241, 2917, 1634, 1593, 1538, 1469, 1425, 1375, 1354, 1309, 1251, 1213, 1142, 1131, 1029, 971, 886, 858, 833, 760, 724, 683, 670, 567 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.38 (s, 3H, 2'-C$\underline{H}_3$), 2.39 (m, 3H, 5'-C$\underline{H}_3$), 4.43 (d, J = 5.2 Hz, 2H, C$\underline{H}_2$), 6.22 (broad s, 1H, N$\underline{H}$), 6.58 (d, J = 1.2 Hz, 1H, 4'-H), 7.48 (d, J = 8.3 Hz, 1H, 5-H), 7.58 (dd, J = 8.3 Hz, J' = 2.1 Hz, 1H, 6-H), 7.85 (d, J = 2.1 Hz, 1H, 2-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 13.0 (C2'-$\underline{C}H_3$), 15.2 (C5'-$\underline{C}H_3$), 37.6 (CH_2, $\underline{C}H_2$), 126.3 (CH, C6), 126.4 (CH, C4'), 129.3 (CH, C2), 130.8 (CH, C5), 132.8 (C, C1), 133.2 (C, C3'), 134.2 (C, C2'), 134.3 (C, C3), 136.1 (C, C4), 136.7 (C, C5'), 165.1 (C, $\underline{C}O$). HRMS: Calcd for $[C_{14}H_{13}Cl_2NOS+H]^+$: 314.0168, found: 314.0165.

3-fluoro-*N*-((5-methylthiophen-3-yl)methyl)-5-(trifluoromethyl)benzamide, 45. Following general procedure C, 3-fluoro-5-(trifluoromethyl)benzoic acid (100 mg, 0.48 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (174 µL, 285 mg, 2.40 mmol). Then, (5-methylthiophen-3-yl)methanamine (73 mg, 0.58 mmol) and triethylamine (162 µL, 118 mg, 1.16 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride (109 mg, 0.48 mmol) in anh. DCM (1.0 mL) giving the product as a yellow solid (30 mg, 20% yield). m.p. 86-87 °C. *v*: 3274, 1639, 1603, 1549, 1446, 1357, 1279, 1219, 1181, 1164, 1124, 1094,

1068, 1006, 952, 885, 835, 779, 710, 964, 626, 584 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.45 (s, 3H, 5'-C<u>H</u>$_3$), 4.53 (d, *J* = 5.5 Hz, 2H, C<u>H</u>$_2$), 6.57 (broad s, 1H, N<u>H</u>), 6.73 (s, 1H, 4'-H), 6.94 (s, 1H, 2'-H), 7.45 (dt, *J* = 8.3 Hz, *J'* = 1.9 Hz, 1H, 4-H), 7.70 (dt, *J* = 8.7 Hz, *J'* = 2.0 Hz, 1H, 2-H), 7.81 (s, 1H, 6-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 15.4 (C5'-<u>C</u>H_3), 39.9 (CH_2, <u>C</u>H_2), 115.9 (dq, $^2J_{CF}$ = 24.6 Hz, $^3J_{CF}$ = 3.7 Hz, CH, C4), 118.1 (d, $^2J_{CF}$ = 22.8 Hz, CH, C2), 119.6 (p, $^3J_{CF}$ = 3.7 Hz, CH, C6), 120.8 (CH, C2'), 123.0 (qd, $^1J_{CF}$ = 272.8 Hz, $^4J_{CF}$ = 2.9 Hz, C, <u>C</u>F_3), 125.7 (CH, C4'), 133.1 (qd, $^2J_{CF}$ = 33.8 Hz, $^3J_{CF}$ = 7.7 Hz, C, C5), 137.9 (d, $^3J_{CF}$ = 6.9 Hz, C1), 138.0 (C, C3'), 141.5 (C, C5'), 162.6 (d, $^1J_{CF}$ = 251.1 Hz, C, C3), 164.7 (d, $^4J_{CF}$ = 2.3 Hz, C, <u>C</u>O). HRMS: Calcd for $[C_{14}H_{11}F_4NOS+H]^+$: 318.0570, found: 318.0576.

3-fluoro-*N*-((2-methylthiophen-3-yl)methyl)-5-(trifluoromethyl)benzamide, 46. Following general procedure C, 3-fluoro-5-(trifluoromethyl)benzoic acid (100 mg, 0.48 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (349 μL, 572 mg, 4.81 mmol). Then, ((2-methylthiophen-3-yl)methanamine hydrochloride (95 mg, 0.58 mmol) and triethylamine (323 μL, 235 mg, 2.32 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride (109 mg, 0.48 mmol) in anh. DCM (1.0 mL) giving the product as a beige solid (45 mg, 30% yield). m.p. 109-110 °C. *v*: 3325, 3227, 3066, 2923, 1635, 1604, 1539, 1466, 1443, 1364, 1335, 1283, 1220, 1170, 1122, 1093, 1052, 895, 875, 775, 722, 689, 654 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.47 (s, 3H, 2'-C<u>H</u>$_3$), 4.54 (d, *J* = 5.3 Hz, 2H, C<u>H</u>$_2$), 6.44 (broad s, 1H, N<u>H</u>), 6.94 (d, *J* = 5.3 Hz, 1H, 4'-H), 7.07 (d, *J* = 5.2 Hz, 1H, 5'-H), 7.45 (d, *J* = 8.1 Hz, 1H, 4-H), 7.68 (dt, *J* = 8.6 Hz, *J'* = 2.0 Hz, 1H, 2-H), 7.79 (s, 1H, 6-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 13.1 (C2'-<u>C</u>H_3), 37.6 (CH_2, <u>C</u>H_2), 115.9 (dq, $^2J_{CF}$ = 24.6 Hz, $^3J_{CF}$ = 3.7 Hz, CH, C4), 118.1 (d, $^2J_{CF}$ = 22.8 Hz, CH, C2), 119.6 (p, $^3J_{CF}$ = 3.7 Hz, CH, C6), 122.5 (CH, C5'), 123.0 (qd, $^1J_{CF}$ = 272.7 Hz, $^4J_{CF}$ = 2.9 Hz, C, <u>C</u>F_3), 128.5 (CH, C4'), 133.0 (C, C3'), 133.1 (qd, $^2J_{CF}$ = 33.9 Hz, $^3J_{CF}$ = 7.7 Hz, C, C5), 136.8 (C, C2'), 137.8 (d, $^3J_{CF}$ = 6.9 Hz, C1), 162.6 (d, $^1J_{CF}$ = 251.1 Hz, C, C3), 164.6 (d, $^4J_{CF}$ = 2.3 Hz, C, <u>C</u>O). HRMS: Calcd for $[C_{14}H_{11}F_4NOS-H]^-$: 316.0425, found: 316.0431.

3-fluoro-*N*-(thiophen-3-ylmethyl)-5-(trifluoromethyl)benzamide, 47. Following general procedure C, 3-fluoro-5-(trifluoromethyl)benzoic acid (100 mg, 0.48 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (349 μL, 572 mg, 4.81 mmol). Then, thiophen-3-ylmethanamine (65 mg, 0.58 mmol) and triethylamine (323 μL, 235 mg, 2.32 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride (109 mg, 0.48 mmol) in anh. DCM (1.0 mL) giving the product as an off-white solid (65 mg, 45% yield). m.p. 75-76 °C. *v*: 3231, 3103, 2938, 1642, 1603, 1550, 1469, 1447, 1368, 1348, 1291, 1251, 1219, 1169, 1130, 1096, 1046, 1010, 945, 925, 886, 784, 691, 635, 563 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 4.61 (d, *J* = 5.7 Hz, 2H, C<u>H</u>$_2$), 6.78 (broad s, 1H, N<u>H</u>), 7.06 (dd, *J* = 4.9 Hz, *J'* = 1.3 Hz, 1H, 4'-H), 7.19 (m, 1H, 2'-H), 7.31 (dd, *J* = 5.0 Hz, *J'* = 3.0 Hz, 1H, 5'-H), 7.44 (d, *J* = 8.1 Hz, 1H, 4-H), 7.68 (dt, *J* = 8.7 Hz, *J'* = 2.0 Hz, 1H, 2-H), 7.81 (s, 1H, 6-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 39.6 (CH_2, <u>C</u>H_2), 115.9 (dq, $^2J_{CF}$ = 24.6 Hz, $^3J_{CF}$ = 3.7 Hz, CH, C4), 118.1 (d, $^2J_{CF}$ = 22.8 Hz, CH, C2), 119.7 (p, $^3J_{CF}$ = 3.7 Hz, CH, C6), 123.0 (qd, $^1J_{CF}$ = 272.8 Hz, $^4J_{CF}$ = 2.9 Hz, C, <u>C</u>F_3), 123.1 (CH, C2'), 126.9 (CH, C5'), 127.5 (CH, C4'), 133.1 (qd, $^2J_{CF}$ = 33.8 Hz, $^3J_{CF}$ = 7.7 Hz, C, C5), 137.8 (d, $^3J_{CF}$ = 6.9 Hz, C1), 138.2 (C, C3'), 162.6 (d, $^1J_{CF}$ = 251.1 Hz, C, C3), 164.8 (d, $^4J_{CF}$ = 2.3 Hz, C, <u>C</u>O). HRMS: Calcd for $[C_{13}H_9F_4NOS-H]^-$: 302.0268, found: 302.0270.

3-chloro-*N*-((5-methylthiophen-3-yl)methyl)-5-(trifluoromethyl)benzamide, 48. Following general procedure C, 3-chloro-5-(trifluoromethyl)benzoic acid (100 mg, 0.45 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (163 μL, 268 mg, 2.25 mmol). Then, (5-methylthiophen-3-yl)methanamine (68 mg, 0.53 mmol) and triethylamine (148 μL, 107 mg, 1.06 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride (109 mg, 0.45 mmol) in anh. DCM (1.0 mL) giving the product as a pale-yellow solid (87 mg, 58% yield). m.p. 130-131 °C. *v*: 3268, 1632, 1583, 1538, 1437, 1362, 1322, 1275, 1171, 1136, 1101, 1056, 886, 828, 756, 691, 648, 577, 560 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.44 (d, *J* = 1.1 Hz, 3H, 5'-C<u>H</u>$_3$), 4.51 (d, *J* = 5.6 Hz, 2H, C<u>H</u>$_2$), 6.71 (s, 1H, 4'-H), 6.75 (broad s, 1H, N<u>H</u>), 6.92 (s, 1H, 2'-H), 7.71 (s, 1H, 4-H), 7.91 (s, 1H, 6-H), 7.93 (s, 1H, 2-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 15.4 (C5'-<u>C</u>H_3), 39.8 (CH_2, <u>C</u>H_2), 120.7 (CH, C2'), 122.3 (q, $^3J_{CF}$ = 3.7 Hz, CH, C6), 123.0 (q, $^1J_{CF}$ = 273.2 Hz, C, <u>C</u>F_3), 125.6 (CH, C4'), 128.4 (q, $^3J_{CF}$ = 3.8 Hz, CH, C4), 130.8 (CH, C2), 132.7 (q, $^2J_{CF}$ = 33.6 Hz, C, C5), 135.7 (C, C3), 137.1 (C, C1), 138.0 (C, C3'), 141.4 (C, C5'), 164.7 (C, <u>C</u>O). HRMS: Calcd for $[C_{14}H_{11}ClF_3NOS+H]^+$: 334.0275, found: 334.0282.

3-chloro-*N*-((2-methylthiophen-3-yl)methyl)-5-(trifluoromethyl)benzamide, 49. Following general procedure C, 3-chloro-5-(trifluoromethyl)benzoic acid (100 mg, 0.45 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (323 μL, 530 mg, 4.45 mmol). Then, (2-methylthiophen-3-yl)methanamine hydrochloride (87 mg, 0.53 mmol) and triethylamine (298 μL, 217 mg, 2.14 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride (108 mg, 0.45 mmol) in anh. DCM (1.0 mL) giving the product as a beige solid (81 mg, 55% yield). m.p. 117-118 °C. *v*: 3299, 3082, 2935, 1639, 1540, 1440, 1322, 1279, 1166, 1124, 887, 827, 770, 745, 721, 690, 658, 611 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.46 (s, 3H, 2'-C<u>H</u>$_3$), 4.53 (d, *J* = 5.3 Hz, 2H, C<u>H</u>$_2$), 6.55 (broad s, 1H, N<u>H</u>), 6.93 (d, *J* = 5.3 Hz, 1H, 4'-H), 7.06 (d, *J* = 5.2 Hz, 1H, 5'-H), 7.71 (m, 1H, 4-H), 7.89 (m, 1H, 6-H), 7.91 (t, *J* = 1.8 Hz, 1H, 2-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 13.1 (C2'-<u>C</u>H_3), 37.6 (CH_2, <u>C</u>H_2), 122.3 (q, $^3J_{CF}$ = 3.7 Hz, CH, C6), 122.5 (CH, C5'), 123.0 (q, $^1J_{CF}$ = 273.1 Hz, C, <u>C</u>F_3), 128.4 (q, $^3J_{CF}$ = 3.8 Hz, CH, C4), 128.5 (CH, C4'), 130.8 (CH, C2), 132.7 (q, $^2J_{CF}$ = 33.6 Hz, C, C5), 133.0 (C, C3'), 135.7 (C, C3), 136.8 (C, C2'), 137.0 (C, C1), 164.7 (C, <u>C</u>O). HRMS: Calcd for $[C_{14}H_{11}ClF_3NOS-H]^-$: 332.0129, found: 332.0134.

3-chloro-*N*-(thiophen-3-ylmethyl)-5-(trifluoromethyl)benzamide, 50. Following general procedure C, 3-chloro-5-(trifluoromethyl)benzoic acid (100 mg, 0.45 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (323 μL, 530 mg, 4.45 mmol). Then, thiophen-3-ylmethanamine (60 mg, 0.53 mmol) and triethylamine (149 μL, 108 mg, 1.07 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride (108 mg, 0.45 mmol) in anh. DCM (1.0 mL) giving the product as a white solid (72 mg, 51% yield). m.p. 91-92 °C. *v*: 3246, 3091, 2927, 1635, 1549, 1429, 1349, 1324, 1284, 1175, 1124, 1050, 1004, 889, 826, 788, 738, 689, 637, 621 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 4.60

(d, J = 5.6 Hz, 2H, C<u>H</u>$_2$), 6.84 (broad s, 1H, N<u>H</u>), 7.05 (dd, J = 4.9 Hz, J' = 1.3 Hz, 1H, 4'-H), 7.18 (m, 1H, 2'-H), 7.30 (dd, J = 5.0 Hz, J' = 3.0 Hz, 1H, 5'-H), 7.71 (m, 1H, 4-H), 7.90 (m, 1H, 6-H), 7.93 (t, J = 1.8 Hz, 1H, 2-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 39.6 (CH_2, <u>C</u>H$_2$), 122.3 (q, $^3J_{CF}$ = 3.7 Hz, CH, C6), 123.0 (q, $^1J_{CF}$ = 273.2 Hz, C, <u>C</u>F$_3$), 123.0 (CH, C2'), 126.8 (CH, C5'), 127.4 (CH, C4'), 128.5 (q, $^3J_{CF}$ = 3.7 Hz, CH, C4), 130.8 (CH, C2), 132.7 (q, $^2J_{CF}$ = 33.6 Hz, C, C5), 135.7 (C, C3), 137.0 (C, C1), 138.2 (C, C3'), 164.8 (C, <u>C</u>O). HRMS: Calcd for $[C_{14}H_{11}ClF_3NOS\text{-}H]^-$: 317.9973, found: 317.9975.

3,4-difluoro-*N*-((5-methylthiophen-3-yl)methyl)-5-(trifluoromethyl)benzamide, 51. Following general procedure C, 3,4-difluoro-5-(trifluoromethyl)benzoic acid (85 mg, 0.38 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (276 µL, 452 mg, 3.80 mmol). Then, (5-methylthiophen-3-yl)methanamine hydrochloride (40 mg, 0.24 mmol) and triethylamine (175 µL, 127 mg, 1.26 mmol) in anh. DCM (1.0 mL) were reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a white solid (59 mg, 72% yield). m.p. 93-94 °C. *v*: 3291, 3095, 2931, 1625, 1605, 1545, 1501, 1435, 1373, 1337, 1273, 1186, 1156, 1134, 1056, 1000, 966, 904, 895, 838, 772, 746, 739, 672, 641, 573 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.45 (d, J = 1.1 Hz, 3H, 5'-C<u>H</u>$_3$), 4.51 (d, J = 5.6 Hz, 2H, C<u>H</u>$_2$), 6.61 (broad s, 1H, N<u>H</u>), 6.71 (p, J = 1.1, 1H, 4'-H), 6.93 (m, 1H, 2'-H), 7.80 (m, 1H, 6-H), 7.86 (ddd, J = 9.6 Hz, J' = 7.0 Hz, J'' = 2.2 Hz, 1H, 2-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 15.4 (C5'-<u>C</u>H$_3$), 39.9 (CH_2, <u>C</u>H$_2$), 120.6 (d, $^2J_{CF}$ = 18.6 Hz, CH, C2), 120.7 (m, CH, C6), 120.7 (m, C, C5), 120.8 (CH, C2'), 121.7 (qd, $^1J_{CF}$ = 273.3 Hz, $^3J_{CF}$ = 3.5 Hz, C, <u>C</u>F$_3$), 125.6 (CH, C4'), 131.4 (t, $^3J_{CF}$ = 4.8 Hz, C, C1), 137.8 (C, C3'), 141.6 (C, C5'), 150.2 (dd, $^1J_{CF}$ = 263.7 Hz, $^2J_{CF}$ = 13.1 Hz, C, C3), 150.8 (d, $^1J_{CF}$ = 241.5 Hz, C, C4), 163.9 (C, <u>C</u>O). HRMS: Calcd for $[C_{14}H_{10}F_5NOS\text{-}H]^-$: 334.0330, found: 334.0333.

3-chloro-4-fluoro-*N*-((5-methylthiophen-3-yl)methyl)-5-(trifluoromethyl)benzamide, 52. Following general procedure C, 3-chloro-4-fluoro-5-(trifluoromethyl)benzoic acid (92 mg, 0.38 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (276 µL, 452 mg, 3.80 mmol). Then, (5-methylthiophen-3-yl)methanamine hydrochloride (40 mg, 0.24 mmol) and triethylamine (175 µL, 127 mg, 1.26 mmol) in anh. DCM (1.0 mL) were reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a white solid (56 mg, 65% yield). m.p. 127-128 °C. *v*: 3273, 3097, 2928, 1631, 1543, 1477, 1374, 1337, 1318, 1274, 1218, 1177, 1153, 1055, 1000, 967, 894, 832, 744, 684, 668, 638, 569 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.45 (d, J = 1.1 Hz, 3H, 5'-C<u>H</u>$_3$), 4.51 (d, J = 5.5 Hz, 2H, C<u>H</u>$_2$), 6.57 (broad s, 1H, N<u>H</u>), 6.71 (p, J = 1.1 Hz, 1H, 4'-H), 6.93 (m, 1H, 2'-H), 7.93 (dd, J = 5.9 Hz, J' = 1.5 Hz, 1H, 6-H), 8.05 (dd, J = 6.4 Hz, J' = 2.2 Hz, 1H, 2-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 15.4 (C5'-<u>C</u>H$_3$), 39.9 (CH_2, <u>C</u>H$_2$), 120.1 (qd, $^2J_{CF}$ = 34.1 Hz, $^2J_{CF}$ = 13.0 Hz, C, C5), 120.9 (CH, C2'), 121.7 ($^1J_{CF}$ = 273.0 Hz, C, <u>C</u>F$_3$), 123.6 (d, $^2J_{CF}$ = 17.6 Hz, C, C3), 124.5 (m, C, C6), 125.6 (CH, C4'), 131.4 (d, $^4J_{CF}$ = 4.6 Hz, C, C1), 133.5 (C, C2), 137.8 (C, C3'), 141.6 (C, C5'), 157.4 (d, $^1J_{CF}$ = 264.3 Hz, C, C4), 163.9 (C, <u>C</u>O). HRMS: Calcd for $[C_{14}H_{10}ClF_4NOS\text{-}H]^-$: 350.0035, found: 350.0043.

2,5-dichloro-*N*-((5-methylthiophen-3-yl)methyl)-3-(trifluoromethyl)benzamide, 53. Following general procedure C, 2,5-dichloro-3-(trifluoromethyl)benzoic acid (98 mg, 0.38 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (276 µL, 452 mg, 3.80 mmol). Then, (5-methylthiophen-3-yl)methanamine hydrochloride (40 mg, 0.24 mmol) and triethylamine (175 µL, 127 mg, 1.26 mmol) in anh. DCM (1.0 mL) were reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a white solid (45 mg, 50% yield). m.p. 152-153 °C. *v*: 3326, 3091, 2924, 1651, 1537, 1426, 1331, 1294, 1259, 1220, 1172, 1138, 1043, 1008, 891, 824, 735, 697, 638, 601, 583 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.47 (d, J = 1.1 Hz, 3H, 5'-C<u>H</u>$_3$), 4.55 (d, J = 5.5 Hz, 2H, C<u>H</u>$_2$), 6.22 (broad s, 1H, N<u>H</u>), 6.75 (p, J = 1.2 Hz, 1H, 4'-H), 6.96 (m, 1H, 2'-H), 7.68 (d, J = 2.6 Hz, 1H, 2-H), 7.73 (d, J = 2.6 Hz, 1H, 4-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 15.5 (C5'-<u>C</u>H$_3$), 39.9 (CH_2, <u>C</u>H$_2$), 120.9 (CH, C2'), 121.9 (q, $^1J_{CF}$ = 273.2 Hz, C, <u>C</u>F$_3$), 125.6 (CH, C4'), 127.5 (C, C6), 129.2 (q, $^3J_{CF}$ = 5.6 Hz, CH, C4), 130.9 (q, $^2J_{CF}$ = 32.0 Hz, C, C5), 132.6 (CH, C2), 133.6 (C, C3), 137.5 (C, C3'), 139.8 (C, C1), 141.6 (C, C5'), 164.4 (C, <u>C</u>O). HRMS: Calcd for $[C_{14}H_{10}Cl_2F_3NOS+H]^+$: 367.9885, found: 367.9879.

***N*-((2,5-dimethylfuran-3-yl)methyl)-3-fluoro-5-(trifluoromethyl)benzamide, 54.** Following general procedure C, 3-fluoro-5-(trifluoromethyl)benzoic acid (100 mg, 0.48 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (349 µL, 572 mg, 4.81 mmol). Then, (2,5-dimethylfuran-3-yl)methanamine (72 mg, 0.58 mmol) and triethylamine (323 µL, 235 mg, 2.32 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride (109 mg, 0.48 mmol) in anh. DCM (1.0 mL) giving the product as a white solid (46 mg, 30% yield). m.p. 71-72 °C. *v*: 3312, 3095, 2921, 1647, 1604, 1557, 1443, 1348, 1294, 1249, 1174, 1142, 1094, 1038, 926, 887, 803, 753, 691, 621 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 2.22 (s, 3H, 5'-C<u>H</u>$_3$), 2.25 (s, 3H, 2'-C<u>H</u>$_3$), 4.33 (d, J = 5.2 Hz, 2H, C<u>H</u>$_2$), 5.89 (s, 1H, 4'-H), 6.35 (broad s, 1H, N<u>H</u>), 7.44 (m, 1H, 4-H), 7.67 (dt, J = 8.7 Hz, J' = 2.1 Hz, 1H, 2-H), 7.78 (s, 1H, 6-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 11.6 (C2'-<u>C</u>H$_3$), 13.5 (C5'-<u>C</u>H$_3$), 35.7 (CH_2, <u>C</u>H$_2$), 106.9 (CH, C4'), 115.8 (C, C3'), 115.8 (dq, $^2J_{CF}$ = 24.6 Hz, $^3J_{CF}$ = 3.8 Hz, CH, C4), 118.0 (d, $^2J_{CF}$ = 22.8 Hz, CH, C2), 119.6 (p, $^3J_{CF}$ = 3.7 Hz, CH, C6), 123.0 (qd, $^1J_{CF}$ = 272.8 Hz, $^4J_{CF}$ = 2.9 Hz, C, <u>C</u>F$_3$), 133.1 (qd, $^2J_{CF}$ = 33.7 Hz, $^3J_{CF}$ = 7.6 Hz, C, C5), 137.9 (d, $^3J_{CF}$ = 6.8 Hz, C1), 147.7 (C, C2'), 150.5 (C, C5'), 162.6 (d, $^1J_{CF}$ = 251.0 Hz, C, C3), 164.7 ($^4J_{CF}$ = 2.3 Hz, C, <u>C</u>O). HRMS: Calcd for $[C_{15}H_{13}F_4NO_2\text{-}H]^-$: 314.0810, found: 314.0819.

3-chloro-*N*-((2,5-dimethylfuran-3-yl)methyl)-5-(trifluoromethyl)benzamide, 55. Following general procedure C, 3-chloro-5-(trifluoromethyl)benzoic acid (100 mg, 0.45 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (323 µL, 530 mg, 4.45 mmol). Then, (2,5-dimethylfuran-3-yl)methanamine (67 mg, 0.53 mmol) and triethylamine (147 µL, 107 mg, 1.06 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride (109 mg, 0.45 mmol) in anh. DCM (1.0 mL) giving the product as a pale-yellow solid (86 mg, 58% yield). m.p. 88-89 °C. *v*: 3304, 3079, 2922, 2441, 1636, 1542, 1428, 1321, 1289, 1174, 1129, 1031, 924, 889, 794, 691, 624 cm^{-1}. ^{1}H NMR (400 MHz, CD_3OD) δ: 2.17 (s, 3H, 5'-C<u>H</u>$_3$), 2.24 (s, 3H, 2'-C<u>H</u>$_3$), 4.26 (s, 2H, C<u>H</u>$_2$), 5.92 (s, 1H, 4'-H), 7.85 (m, 1H, 4-H), 8.06 (m, 1H, 6-H), 8.08 (m, 1H, 2-H). ^{13}C NMR (101 MHz, CD_3OD) δ: 11.4 (C2'-<u>C</u>H$_3$), 13.3 (C5'-<u>C</u>H$_3$), 35.9 (CH_2, <u>C</u>H$_2$), 108.1 (CH, C4'), 117.8 (C, C3'), 123.7 (q,

$^{3}J_{CF}$ = 3.9 Hz, CH, C6), 124.5 (q, $^{1}J_{CF}$ = 272.0 Hz, C, CF$_3$), 129.1 (q, $^{3}J_{CF}$ = 3.8 Hz, CH, C4), 132.2 (CH, C2), 133.5 (q, $^{2}J_{CF}$ = 33.3 Hz, C, C5), 136.5 (C, C3), 138.7 (C, C1), 148.3 (C, C2'), 150.9 (C, C5'), 166.5 (C, CO). HRMS: Calcd for $[C_{15}H_{13}ClF_3NO_2\text{-}H]^-$: 330.0514, found: 330.0519.

3-fluoro-*N*-(furan-3-ylmethyl)-5-(trifluoromethyl)benzamide, 56. Following general procedure D, 3-fluoro-5-(trifluoromethyl)benzoic acid (100 mg, 0.48 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (346 µL, 567 mg, 4.81 mmol). Then, furan-3-ylmethanamine (51 mg, 0.53 mmol) and triethylamine (268 µL, 194 mg, 1.92 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a reddish syrup (67 mg, 49% yield). *ν*: 3295, 3087, 1645, 1604, 1544, 1467, 1445, 1341, 1285, 1249, 1219, 1171, 1128, 1093, 1021, 967, 925, 884, 874, 768, 692, 619, 599 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 4.50 (d, *J* = 5.5 Hz, 2H, CH$_2$), 6.38 (broad s, 1H, NH), 6.43 (m, 1H, 4'-H), 7.42 (t, *J* = 1.7 Hz, 1H, 2'-H), 7.45-7.48 (complex signal, 2H, 4-H, 5'-H), 7.69 (dt, *J* = 8.6 Hz, *J'* = 2.0 Hz, 1H, 2-H), 7.79 (s, 1H, 6-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 35.4 (CH_2, CH_2), 110.4 (CH, C4'), 116.0 (dq, $^{2}J_{CF}$ = 24.4 Hz, $^{3}J_{CF}$ = 3.7 Hz, CH, C4), 118.1 (d, $^{2}J_{CF}$ = 22.9 Hz, CH, C2), 119.6 (p, $^{3}J_{CF}$ = 3.8 Hz, CH, C6), 121.6 (C, C3'), 123.0 (q, $^{1}J_{CF}$ = 271.5 Hz, C, CF$_3$), 133.2 (qd, $^{2}J_{CF}$ = 33.9 Hz, $^{3}J_{CF}$ = 7.3 Hz, C, C5), 137.8 (d, $^{3}J_{CF}$ = 6.8 Hz, C, C1), 140.7 (CH, C5'), 143.9 (CH, C2'), 162.7 (d, $^{1}J_{CF}$ = 251.2 Hz, C, C3), 164.8 (C, CO). HRMS: Calcd for $[C_{13}H_9F_4NO_2\text{-}H]^-$: 286.0497, found: 286.0494.

3-fluoro-*N*-(thiazol-5-ylmethyl)-5-(trifluoromethyl)benzamide, 57. Following general procedure D, 3-fluoro-5-(trifluoromethyl)benzoic acid (100 mg, 0.48 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (279 µL, 457 mg, 3.84 mmol). Then, thiazol-5-ylmethanamine (60 mg, 0.53 mmol) and triethylamine (268 µL, 194 mg, 1.92 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a reddish syrup (51 mg, 35% yield). *ν*: 3269, 3080, 1646, 1604, 1543, 1520, 1467, 1446, 1408, 1348, 1280, 1220, 1169, 1126, 1093, 1039, 1003, 926, 881, 797, 752, 692, 633, 601 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 4.86 (dd, *J* = 5.9 Hz, *J'* = 0.9 Hz, 2H, CH$_2$), 6.85 (broad m, 1H, NH), 7.48 (dt, *J* = 8.0 Hz, *J'* = 1.7 Hz, 1H, 4-H), 7.72 (dt, *J* = 8.5 Hz, *J'* = 2.1 Hz, 1H, 2-H), 7.81 (d, *J* = 0.8 Hz, 1H, 4'-H), 7.82 (s, 1H, 6-H), 8.76 (s, 1H, 2'-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 36.3 (CH_2, CH_2), 116.3 (dq, $^{2}J_{CF}$ = 24.4 Hz, $^{3}J_{CF}$ = 3.7 Hz, CH, C4), 118.2 (d, $^{2}J_{CF}$ = 23.0 Hz, CH, C2), 119.7 (p, $^{3}J_{CF}$ = 3.7 Hz, CH, C6), 122.9 (q, $^{1}J_{CF}$ = 272.7 Hz, C, CF$_3$), 133.3 (qd, $^{2}J_{CF}$ = 34.0 Hz, $^{3}J_{CF}$ = 7.7 Hz, C, C5), 135.0 (C, C5'), 137.2 (d, $^{3}J_{CF}$ = 6.9 Hz, C, C1), 142.5 (CH, C4'), 154.1 (CH, C2'), 162.7 (d, $^{1}J_{CF}$ = 251.5 Hz, C, C3), 164.8 (C, CO). HRMS: Calcd for $[C_{12}H_8F_4N_2OS\text{+}H]^+$: 285.0304, found: 285.0312.

3-fluoro-*N*-(isoxazol-3-ylmethyl)-5-(trifluoromethyl)benzamide, 58. Following general procedure D, 3-fluoro-5-(trifluoromethyl)benzoic acid (78 mg, 0.37 mmol) in anh. toluene (1.5 mL) and drops of DMF was reacted with thionyl chloride (272 µL, 446 mg, 3.75 mmol). Then, isoxazol-3-ylmethylamine (37 mg, 0.37 mmol) and triethylamine (209 µL, 152 mg, 1.50 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a brownish solid (34 mg, 32% yield). m.p. 110-111 °C. *ν*: 267, 3114, 1670, 1630, 1606, 1573, 1550, 1501, 1467, 1438, 1422, 1354, 1293, 1245, 1222, 1162, 1128, 1093, 1059, 1044, 1020, 1003, 938, 888, 862, 795, 713, 692, 651, 592, 555 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 4.77 (d, *J* = 5.6 Hz, 2H, CH$_2$), 6.43 (d, *J* = 1.7 Hz, 1H, 4'-H), 6.97 (broad s, 1H, NH), 7.49 (m, 1H, 4-H), 7.74 (dt, *J* = 8.5 Hz, *J'* = 1.8 Hz, 1H, 2-H), 7.86 (s, 1H, 6-H), 8.41 (d, *J* = 1.6 Hz, 1H, 5'-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 36.3 (CH_2, CH_2), 104.1 (CH, C4'), 116.3 (dq, $^{2}J_{CF}$ = 24.6 Hz, $^{3}J_{CF}$ = 3.8 Hz, CH, C4), 118.2 (d, $^{2}J_{CF}$ = 22.9 Hz, CH, C2), 119.8 (p, $^{3}J_{CF}$ = 3.8 Hz, CH, C6), 122.9 (q, $^{1}J_{CF}$ = 272.7 Hz, C, CF$_3$), 133.3 (qd, $^{2}J_{CF}$ = 34.0 Hz, $^{3}J_{CF}$ = 7.7 Hz, C, C5), 137.2 (d, $^{3}J_{CF}$ = 6.9 Hz, C, C1), 159.4 (CH, C5'), 159.7 (C, C3'), 162.7 (d, $^{1}J_{CF}$ = 251.3 Hz, C, C3), 165.0 (C, CO). HRMS: Calcd for $[C_{12}H_8F_4N_2O_2\text{-}H]^-$: 287.0449, found: 287.0446.

3-fluoro-*N*-(pyridin-4-ylmethyl)-5-(trifluoromethyl)benzamide, 59. Following general procedure D, 3-fluoro-5-(trifluoromethyl)benzoic acid (100 mg, 0.48 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (346 µL, 567 mg, 4.81 mmol). Then, pyridin-4-ylmethanamine (54 µL, 57 mg, 0.53 mmol) and triethylamine (268 µL, 194 mg, 1.92 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a pale-yellow syrup (81 mg, 57% yield). *ν*: 3282, 3061, 1650, 1602, 1543, 1467, 1445, 1417, 1351, 1325, 1283, 1219, 1169, 1126, 1094, 1059, 1002, 928, 883, 799, 754, 691, 615 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 4.65 (d, *J* = 6.0 Hz, 2H, CH$_2$), 7.12 (broad s, 1H, NH), 7.23 [d, *J* = 6.1 Hz, 2H, 2'(6')-H], 7.49 (m, 1H, 4-H), 7.76 (dt, *J* = 8.6 Hz, *J'* = 2.1 Hz, 1H, 2-H), 7.86 (s, 1H, 6-H), 8.53 [d, *J* = 6.1 Hz, 2H, 3'(5')-H]. ^{13}C NMR (101 MHz, $CDCl_3$) δ: 43.2 (CH_2, CH_2), 116.3 (dq, $^{2}J_{CF}$ = 24.3 Hz, $^{3}J_{CF}$ = 3.7 Hz, CH, C4), 118.3 (d, $^{2}J_{CF}$ = 22.8 Hz, CH, C2), 119.7 (p, $^{3}J_{CF}$ = 4.0 Hz, CH, C6), 122.6 [CH, C2'(6')], 122.9 (qd, $^{1}J_{CF}$ = 272.9 Hz, $^{3}J_{CF}$ = 2.9 Hz, C, CF$_3$), 133.3 (qd, $^{2}J_{CF}$ = 33.7 Hz, $^{3}J_{CF}$ = 7.6 Hz, C, C5), 137.3 (d, $^{3}J_{CF}$ = 6.7 Hz, C, C1), 146.9 (C, C1'), 150.2 [CH, C3'(5')], 162.7 (d, $^{1}J_{CF}$ = 251.5 Hz, C, C3), 165.2 (d, $^{4}J_{CF}$ = 2.4 Hz, C, CO). HRMS: Calcd for $[C_{14}H_{10}F_4N_2O\text{+}H]^+$: 299.0802, found: 299.0805.

***N*-((1*H*-pyrazol-3-yl)methyl)-3-fluoro-5-(trifluoromethyl)benzamide, 60.** Following general procedure D, 3-fluoro-5-(trifluoromethyl)benzoic acid (100 mg, 0.48 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (279 µL, 457 mg, 3.84 mmol). Then, (1*H*-pyrazol-3-yl)methanamine (51 mg, 0.53 mmol) and triethylamine (201 µL, 146 mg, 1.44 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a pale-yellow solid (75 mg, 54% yield). m.p. 130-131 °C. *ν*: 3197, 3069, 2945, 1645, 1599, 1556, 1471, 1441, 1428, 1343, 1297, 1240, 1212, 1174, 1131, 1096, 1051, 1035, 1002, 921, 692, 859, 790, 765, 748, 716, 664, 615 cm^{-1}. ^{1}H NMR (400 MHz, $CDCl_3$) δ: 4.70 (d, *J* = 5.6 Hz, 2H, CH$_2$), 6.30 (d, *J* = 2.2 Hz, 1H, 4'-H), 7.31 (broad s, 1H, NH), 7.45 (m, 1H, 4-H), 7.54 (d, *J* = 2.2 Hz, 1H, 5'-H), 7.74 dt, *J* = 9.2 Hz, *J'* = 2.2 Hz, 1H, 2-H), 7.85 (td, *J* = 1.6 Hz, *J'* = 0.8 Hz, 1H, 6-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 37.1 (CH_2, CH_2), 104.5 (CH, C4'), 115.8 (dq, $^{2}J_{CF}$ = 24.5 Hz, $^{3}J_{CF}$ = 4.0 Hz, CH, C4), 118.1 (d, $^{2}J_{CF}$ = 22.9 Hz, CH, C2), 119.7 (m, CH, C6), 122.8 (qd, $^{1}J_{CF}$ = 270.5 Hz, $^{3}J_{CF}$ = 2.8 Hz, C, CF$_3$), 132.3 (CH, C5'), 132.9 (qd, $^{2}J_{CF}$ = 33.9 Hz, $^{3}J_{CF}$ = 7.5 Hz, C, C5), 137.4 (d, $^{3}J_{CF}$ = 6.9 Hz, C, C1), 146.3 (C, C3'), 162.4 (d, $^{1}J_{CF}$ = 250.9

Hz, C, C3), 165.0 (d, J = 2.2 Hz, C, CO). HRMS: Calcd for $[C_{12}H_9F_4N_3O\text{-}H]^-$: 286.0609, found: 286.0605.

3-fluoro-*N*-(oxazol-4-ylmethyl)-5-(trifluoromethyl)benzamide, 61. Following general procedure D, 3-fluoro-5-(trifluoromethyl)benzoic acid (100 mg, 0.48 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (279 μL, 457 mg, 3.84 mmol). Then, oxazol-4-ylmethanamine hydrochloride (71 mg, 0.53 mmol) and triethylamine (268 μL, 194 mg, 1.92 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a yellowish solid (107 mg, 77% yield). m.p. 99-100 °C. *ν*: 3272, 3117, 1659, 1608, 1547, 1506, 1470, 1442, 1427, 1348, 1328, 1295, 1280, 1247, 1232, 1210, 1171, 1122, 1095, 1068, 1045, 1004, 968, 923, 916, 881, 859, 796, 774, 748, 690, 664, 621 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 4.56 (d, J= 5.1 Hz, 2H, $C\underline{H}_2$), 7.28 (broad s, 1H, N<u>H</u>), 7.43 (m, 1H, 4-H), 7.68-7.72 (complex signal, 2H, 2-H, 2'-H), 7.82 (s, 1H, 6-H), 7.86 (s, 1H, 5'-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 35.8 (CH_2, $\underline{C}H_2$), 116.0 (dq, $^2J_{CF}$ = 24.5 Hz, $^3J_{CF}$ = 3.9 Hz, CH, C4), 118.2 (d, $^2J_{CF}$ = 22.7 Hz, CH, C2), 119.8 (p, $^3J_{CF}$ = 3.7 Hz, CH, C6), 122.9 (qd, $^1J_{CF}$ = 273.0 Hz, $^3J_{CF}$ = 3.1 Hz, C, $\underline{C}F_3$), 133.1 (qd, $^2J_{CF}$ = 33.8 Hz, $^3J_{CF}$ = 7.7 Hz, C, C5), 136.4 (CH, C2'), 136.4 (CH, C3'), 137.5 (d, $^3J_{CF}$ = 6.9 Hz, C, C1), 151.6 (CH, C5'), 162.6 (d, $^1J_{CF}$ = 251.0 Hz, C, C3), 164.9 (d, $^4J_{CF}$ = 2.4 Hz, C, <u>C</u>O). HRMS: Calcd for $[C_{12}H_8F_4N_2O_2\text{-}H]^-$: 287.0449, found: 287.0445.

3-fluoro-*N*-(thiazol-4-ylmethyl)-5-(trifluoromethyl)benzamide, 62. Following general procedure D, 3-fluoro-5-(trifluoromethyl)benzoic acid (100 mg, 0.48 mmol) in anh. toluene (2.0 mL) and drops of DMF was reacted with thionyl chloride (279 μL, 457 mg, 3.84 mmol). Then, thiazol-4-ylmethanamine dihydrochloride (99 mg, 0.53 mmol) and triethylamine (268 μL, 194 mg, 1.92 mmol) in anh. DCM (1.0 mL), were reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a white solid (61 mg, 42% yield). m.p, 98-99 °C. *ν*: 3268, 3138, 3069, 1664, 1611, 1543, 1520, 1471, 1428, 1407, 1348, 1289, 1247, 1209, 1170, 1123, 1093, 1041, 1002, 943, 921, 882, 859, 825, 765, 750, 729, 689, 643, 584 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 4.77 (dd, J= 5.5 Hz, J'= 0.7 Hz, 2H, $C\underline{H}_2$), 7.31 (dt, J= 2.0 Hz, J'= 0.8 Hz, 1H, 2'-H), 7.40 (broad m, 1H, N<u>H</u>), 7.43 (m, 1H, 4-H), 7.72 (dt, J= 8.7 Hz, J'= 2.1 Hz, 1H, 2-H), 7.84 (td, J= 1.6 Hz, J'= 0.8 Hz, 1H, 6-H), 8.77 (d, J= 2.0 Hz, 1H, 5'-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 40.1 (CH_2, $\underline{C}H_2$), 115.9 (dq, $^2J_{CF}$ = 24.8 Hz, $^3J_{CF}$ = 3.7 Hz, CH, C4), 116.1 (CH, C2'), 118.2 (d, $^2J_{CF}$ = 22.7 Hz, CH, C2), 119.8 (p, $^3J_{CF}$ = 3.6 Hz, CH, C6), 123.0 (qd, $^1J_{CF}$ = 272.7 Hz, $^3J_{CF}$ = 2.8 Hz, C, $\underline{C}F_3$), 133.1 (qd, $^2J_{CF}$ = 33.9 Hz, $^3J_{CF}$ = 7.7 Hz, C, C5), 137.6 (d, $^3J_{CF}$ = 6.8 Hz, C, C1), 153.0 (C, C3'), 153.7 (CH, C5'), 162.6 (d, $^1J_{CF}$ = 251.0 Hz, C, C3), 164.8 (d, $^4J_{CF}$ = 2.4 Hz, C, <u>C</u>O). Anal. calcd for $C_{12}H_8F_4N_2OS$: C 47.37, H 2.65, N 9.21. HRMS: Calcd for $[C_{12}H_8F_4N_2OS\text{-}H]^-$: 303.0221, found: 303.0219.

(2,5-dimethylthiophen-3-yl)methyl 3-fluoro-5-(trifluoromethyl)benzoate, 63. To a solution of (2,5-dimethylthiophen-3-yl)methanol (88 mg, 0.62 mmol), EDC·HCl (178 mg, 0.93 mmol) and DMAP (30 mg, 0.25 mmol) in anh. DCM (2.0 mL), 3-fluoro-5-(trifluoromethyl)benzoic acid (155 mg, 0.74 mmol) was added under Ar atmosphere and the mixture was stirred at RT for 16 h. Then, DCM (10 mL) was added and the mixture was washed with brine (2 x 20 mL). The organic layer was dried over anh. Na_2SO_4, filtered and concentrated *in vacuo*. The resulting crude was purified by column chromatography in silica gel (using as eluent mixtures of EtOAc in hexane from 0% to 10%) to obtain the product as a colorless oil (96 mg, 47% yield). *ν*: 2923, 1727, 1606, 1453, 1363, 1347, 1249, 1238, 1202, 1171, 1131, 1103, 1091, 959, 906, 887, 828, 768, 692, 582 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.42 (m, 3H, 5'-$C\underline{H}_3$), 2.46 (s, 3H, 2'-$C\underline{H}_3$), 5.24 (s, 2H, $C\underline{H}_2$), 6.69 (d, J= 1.2 Hz, 1H, 4'-H), 7.51 (dddd, J= 8.1 Hz, J'= 2.5 Hz, J''= 1.6 Hz, J'''= 0.7 Hz, 1H, 4-H), 7.91 (ddd, J= 8.7 Hz, J'= 2.6 Hz, J''= 1.4 Hz, 1H, 2-H), 8.11 (s, 1H, 6-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 13.0 (C2'-$\underline{C}H_3$), 15.2 (C5'-$\underline{C}H_3$), 61.0 (CH_2, $\underline{C}H_2$), 117.3 (dq, $^2J_{CF}$ = 24.6 Hz, $^3J_{CF}$ = 3.7 Hz, CH, C4), 120.2 (d, $^2J_{CF}$ = 23.0 Hz, CH, C2), 122.5 (p, $^3J_{CF}$ = 3.9 Hz, CH, C6), 123.0 (qd, $^1J_{CF}$ = 272.8 Hz, $^4J_{CF}$ = 2.9 Hz, C, $\underline{C}F_3$), 126.9 (CH, C4'), 131.0 (C, C3'), 133.0 (qd, $^2J_{CF}$ = 33.9 Hz, $^3J_{CF}$ = 7.5 Hz, C, C5), 133.7 (d, $^3J_{CF}$ = 7.6 Hz, C1), 136.5 (C, C2'), 136.8 (C, C5'), 162.4 (d, $^1J_{CF}$ = 250.4 Hz, C, C3), 164.2 (d, $^4J_{CF}$ = 2.3 Hz, C, <u>C</u>O). HRMS: Calcd for $[C_{15}H_{12}F_4O_2S+Na]^+$: 355.0386, found: 355.0389.

1-(2,5-dimethylthiophen-3-yl)-*N*-(3-fluoro-5-(trifluoromethyl)benzyl)methanamine, 64. To a solution 1-(bromomethyl)-3-fluoro-5-(trifluoromethyl)benzene (50 mg, 0.19 mmol) and triethylamine (81 μL, 59 mg, 0.58 mmol) in DMF (1.0 mL) was added (2,5-dimethylthiophen-3-yl)methanamine hydrochloride (35 mg, 0.19 mmol) and the mixture was kept under stirring at RT for 16 h. Water (15 mL) followed by EtOAc (10 mL) were added and the mixture was extracted. The organic layer was washed again with brine (15 mL) and then it was dried over anh. Na_2SO_4 and filtered. Solvents were concentrated *in vacuo* and the resulting crude was purified by column chromatography in silica gel (using as eluent mixtures of EtOAc in hexane from 0% to 15%) to afford the product as a colorless oil (27 mg, 45% yield). *ν*: 2921, 2860, 1605, 1452, 1342, 1227, 1166, 1125, 1091, 975, 869, 830, 761, 721, 698, 712 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.30 (s, 3H, 2'-$C\underline{H}_3$), 2.40 (m, 3H, 5'-$C\underline{H}_3$), 3.62 (s, 2H, thio-$C\underline{H}_2$), 3.84 (s, 2H, aryl-$C\underline{H}_2$), 6.59 (q, J= 1.2 Hz, 1H, 4'-H), 7.20 (m, 1H, 4-H), 7.29 (dddd, J= 9.4 Hz, J'= 2.6 Hz, J''= 1.4 Hz, J'''= 0.7 Hz, 1H, 2-H), 7.41 (tq, J= 1.5 Hz, J'= 0.7 Hz, 1H, 6-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 12.9 (C2'-$\underline{C}H_3$), 15.2 (C5'-$\underline{C}H_3$), 46.1 (CH_2, thio-$\underline{C}H_2$), 52.3 (CH_2, aryl-$\underline{C}H_2$), 111.4 (dq, $^2J_{CF}$ = 24.6 Hz, $^3J_{CF}$ = 3.8 Hz, CH, C4), 118.4 (d, $^2J_{CF}$ = 21.3 Hz, CH, C2), 120.6 (p, $^3J_{CF}$ = 3.6 Hz, CH, C6), 123.5 (qd, $^1J_{CF}$ = 272.5 Hz, $^3J_{CF}$ = 3.1 Hz, C, $\underline{C}F_3$), 126.7 (CH, C4'), 132.4 (qd, $^2J_{CF}$ = 33.0 Hz, $^3J_{CF}$ = 8.1 Hz, C, C5), 132.9 (C, C3'), 135.4 (C, C2'),, 136.0 (C, C5'), 144.8 (d, $^3J_{CF}$ = 7.1 Hz, C, C1), 162.7 (d, $^1J_{CF}$ = 248.2 Hz, C, C3). HRMS: Calcd for $[C_{15}H_{15}F_4NS+H]^+$: 318.0934, found: 318.0934.

***N*-(3-fluoro-5-(trifluoromethyl)benzyl)-2,5-dimethylthiophene-3-carboxamide, 65.** Following general procedure D, 2,5-dimethylthiophene-3-carboxylic acid (100 mg, 0.64 mmol) in anh. toluene (3.0 mL) and drops of DMF was reacted with thionyl chloride (279 μL, 457 mg, 3.84 mmol). Then, (3-fluoro-5-(trifluoromethyl)phenyl)methanamine (148 mg, 0.77 mmol) and triethylamine (268 μL, 194 mg, 1.92 mmol) in anh. DCM (2.0 mL) reacted with the crude acyl chloride in anh. DCM (1.0 mL) giving the product as a beige solid (85 mg, 40% yield). m.p. 105-106 °C. *ν*: 3293, 2925, 1634, 1609, 1563, 1515, 1454, 1424, 1351, 1341, 1315, 1280, 1229, 1169, 1120, 1084, 1041, 994, 973, 875, 852, 834, 785, 768, 716, 701, 690, 650, 601 cm^{-1}. ^1H NMR (400 MHz, $CDCl_3$) δ: 2.37 (s, 3H, 2'-$C\underline{H}_3$), 2.63 (s, 3H, 5'-$C\underline{H}_3$), 4.59 (d, J = 6.1 Hz, 2H, $C\underline{H}_2$), 6.35

(broad s, 1H, NH), 6.76 (d, J = 1.2 Hz, 1H, 4'-H), 7.22 (m, 2H, 2-H, 4-H), 7.35 (s, 1H, 6-H). ^{13}C NMR (101 MHz, $CDCl_3$) δ: 15.0 (C2'-CH_3), 15.1 (C5'-CH_3), 42.8 (d, $^4J_{CF}$ = 1.7 Hz, CH_2, CH_2), 111.9 (dq, $^2J_{CF}$ = 24.6 Hz, $^3J_{CF}$ = 3.9 Hz, CH, C4), 118.1 (d, $^2J_{CF}$ = 21.9 Hz, CH, C2), 120.1 (m, CH, C6), 123.3 (qd, $^1J_{CF}$ = 272.5 Hz, $^3J_{CF}$ = 3.3 Hz, C, CF_3), 123.9 (CH, C4'), 130.6 (C, C3'), 132.9 (qd, $^2J_{CF}$ = 33.3 Hz, $^3J_{CF}$ = 8.2 Hz, C, C5), 136.5 (C, C5'), 142.9 (d, $^3J_{CF}$ = 7.2 Hz, C, C1), 143.8 (C, C2'), 162.8 (d, $^1J_{CF}$ = 249.4 Hz, C, C3), 164.7 (C, CO). HRMS: Calcd for $[C_{15}H_{13}F_4NOS\text{-}H]^-$: 330.0581, found: 330.0573.

Virological experiments. *Cells and viruses*. Madin-Darby canine kidney (MDCK) cells were a kind gift from M. Matrosovich (Philipps University, Marburg). Human embryonic kidney (HEK293T) cells were purchased from Thermo Scientific and HeLa cells were from ATCC (CCL-2). The virus panel consisted of: A/PR/8/34 [A/H1N1; reverse-engineered from the plasmids which were kindly provided by M. Kim (Korea Research Institute of Chemical Technology); hereunder abbreviated as PR8]; A/Virginia/ATCC3/2009 [A/H1N1; ATCC VR-1738; hereunder abbreviated as Virg09]; A/Victoria/361/11 [A/H3N2; a kind gift from G. Rimmelzwaan (Erasmus Medical Center, Rotterdam)]; and B/Ned/537/05 (Yamagata lineage; a kind gift from R. Fouchier (Erasmus Medical Center, Rotterdam)]. The virus stocks were prepared in MDCK cells or embryonated hen eggs, and titrated by the 50% cell culture infective dose ($CCID_{50}$) method.

Assessment of antiviral activity based on reduction of viral cytopathic effect (CPE). The infection medium consisted of UltraMDCK medium (Lonza), supplemented with 225 mg/L sodium bicarbonate, 2 mM L-glutamine and 2 μg/mL TPCK (tosylphenylalanylchloromethyl-keton)-treated trypsin (Sigma-Aldrich). Our detailed method can be found elsewhere.[53] In short, MDCK cells were seeded at 7,500 cells per well in 96-well plates. On the next day, the virus was added at an MOI of 50 $CCID_{50}$ per well, immediately followed by serial dilutions of the test compounds. After three days incubation at 35 °C, microscopy was performed to score virus-induced CPE and compound cytotoxicity. Next, the CellTiter 96® AQ_{ueous} MTS Reagent (Promega) was added to the cells and, four hours later, the absorbance at 490 nm was measured in a plate reader. The compounds' antiviral activity was expressed as the half-maximal effective concentration (EC_{50}) in the MTS or microscopic scoring assay (see ref. 46 for calculation details]. Cytotoxicity was expressed as the CC_{50}, i.e. 50% cytotoxic concentration by MTS assay and the MCC (minimum cytotoxic concentration), i.e. concentration producing minimal changes in cell morphology.

Selection of resistant influenza viruses by serial passaging. MDCK cells were infected with Virg09 virus as above, and exposed to different concentrations of VF-57a or RL-007. Three days later, all wells were inspected to select the highest compound concentrations at which virus-induced CPE was visible, and freeze these supernatants at -80 °C. The harvests were further passaged under gradually higher compound concentrations, until resistance was reached (i.e. virus breakthrough at 40 μM of compound). A no compound control was passaged in parallel. Virus clones were obtained by plaque purification under 0.6 % agarose and 4 μM VF-57a or 10 μM RL-007. After expansion in MDCK cells, the clones underwent RNA extraction, reverse transcription and high-fidelity PCR, followed by Sanger sequencing (by Macrogen) of the HA gene. To determine the hemolysis pH of the wild-type (WT) and mutant Virg09 viruses, we first prepared allantoic stocks. The method from Leiva *et al.*,[42] was adapted to round-bottom 96-well format. Briefly, virus was added to the wells together with an equal volume of 2% chicken red blood cell (RBC) suspension in PBS. After 10 min incubation at 37 °C, unbound virus was removed by centrifugation. The cell pellets were resuspended in acidic buffer, i.e. PBS that was acidified with acetic acid to a pH ranging from 5.0 to 6.0, with 0.1 increments. After 25 min incubation at 37 °C, the samples were neutralized with NaOH and the plates were centrifugated to pellet intact RBC. The supernatants were transferred to a fresh plate and the absorbance at 540 nm was measured in a plate reader. The hemolysis pH was defined as the pH at which 50% hemolysis occurred, relative to the value at pH 5.0.

Pseudovirus entry assay. The pCAGEN plasmids encoding the H1 HA and N1 NA of Virg09 virus were previously described.[54] The codon-optimized DNAs encoding H5 HA and N1 NA of the highly pathogenic avian A/H5N1 virus (A/bald eagle/FL/W22-114/2022, hereunder abbreviated FL22) were purchased from Life Technologies and cloned into pCAGEN [provided by C. Cepko (Boston, MA) via Addgene (plasmid 11160)], using NEBuilder HiFi DNA Assembly mix. To prepare murine leukemia virus (MLV) pseudoviruses bearing these HA and NA proteins, HEK293T cells were transfected with a mixture of the pCAGEN and MLV backbone plasmids (kind gift from S. Pöhlmann, German Primate Center-Leibniz Institute for Primate Research, Göttingen). The details were published before.[54] Three days after transfection, the Virg09 pseudoviruses were incubated for 15 min with 80 μg/mL trypsin, followed by 80 μg/mL soybean inhibitor. Finally, the Virg09 and FL22 pseudoviruses were stored in aliquots at -80 °C.

To determine the compounds' inhibitory effect on pseudovirus entry, MDCK cells were seeded in white 96-well plates and, on the next day, preincubated with serial compound dilutions (in medium with 2% FCS) for 20 min. The pseudovirus was added and the plates were spinoculated at 37 °C for 45 min at 450 x *g*, followed by 60 min incubation at 37 °C. After replacing the supernatants by fresh medium without compound, the plates were incubated for three days at 37 °C. To quantify the expression of the firefly luciferase reporter, we used a luciferase assay system kit and Glomax Navigator instrument, both from Promega.

Cell-cell fusion assay. To quantify the inhibitory effect on H1 HA-mediated membrane fusion, we used the pGal5-luc and pGal4-VP16 plasmids (kind gift from S. Pöhlmann) and transactivation setup reported by his team,[55] with several modifications. Briefly, suspensions of HeLa cells were prepared in two tubes in which either the pGAL5-luc or pGAL5-VP16 plus HA-pCAGEN plasmid were added (or pCAGEN-empty plasmid for the mock control), together with Fugene® transfection reagent and growth medium (= MEM supplemented with non-essential amino acids, HEPES, L-glutamine and 10% FCS). The cells were seeded in 6- and 96-well plates, respectively, and incubated for one day. From the 6-well plate, the cells were detached with non-enzymatic cell dissociation solution (Sigma), then overlaid on the HA-expressing cells in the 96-well plate. On the next day, the cells were sequentially incubated at 37 °C with the following reagents (with removal after each step): (i) for 15 min: 5 μg/mL of TPCK-trypsin to activate HA; (ii) for 15 min: serial compound dilutions; (iii) for exactly 5 min: fresh compound diluted

in PBS-CM (= PBS supplemented with Ca^{2+} and Mg^{2+}) adjusted to pH 5.3; (iv) washing with growth medium; (v) for 5 h: growth medium to allow cell-cell fusion; and (vi) cell culture lysis reagent from Promega. After reading the luminescence with the luciferase assay system kit and Glomax Navigator instrument, the relative luminescence units (RLU) values were subtracted for the value seen in mock-transfected cells, and the % luminescence at each compound concentration was calculated, relative to the condition which received medium instead of compound.

For H5 HA, we used a microscopic readout as established in a previous study.[54] HeLa cells were seeded in 96-well plates and transfected with H5 HA (strain A/duck/Hunan/795/2002). On the next day, the cells bearing surface-exposed HA (already activated by endogenous furin) were preincubated for 15 min with compound; exposed for 5 min to PBS-CM at pH 5.2 under continued presence of compound; and then washed with PBS-CM. Cell culture medium was added and after 3 h incubation, the cells were fixated and stained with Giemsa solution to allow microscopic assessment of polykaryon formation.

Surface plasmon resonance (SPR)-based assessment of HA refolding. H1 HA protein (strain A/California/06/09; ectodomain 99.4% identical to that of Virg09-HA), produced in HEK293 cells, was purchased from eEnzyme (IA-H1-11SWt). The mouse monoclonal antibodies recognizing the HA head (7B2-32) or stem region (C179) were from Kerafast and Takara Bio, respectively. The HA protein was diluted in PBS to 50 μg/mL and pre-incubated with 100 μM VF-57a or RL007 for 10 min at 37 °C. Next, a pre-determined amount of 1 M citric acid buffer of pH 4.7 was added to install a pH of 5.2. The protein was incubated for 1 h at 37 °C to induce the conformational change in HA, then re-neutralized using citric acid buffer pH 7.

To conduct SPR analysis, the protein sample was diluted to 5 μg/mL, in HEPES-buffered saline supplemented with 1 mg/mL BSA and 0.05% v/v Tween20. First, the antibodies were coupled onto a rabbit anti-mouse C1 sensor chip (Cytiva) at a density of around 100 resonance units (RU) with a flow rate of 10 μL/min for 120 s. Next, the sample was injected for 180 s, at a flow rate of 10 μL/min and with a dissociation time of 150 s. Isotype controls were used to assess non-specific binding and buffer injections for refractive index changes. After each sample injection, a regeneration was performed using three 10 mM Glycine.HCl pH 1.7 injections at a flow rate of 30 μL/min for 20 s and a 0.5% Triton X-100 injection at a flow rate of 30 μL/min for 15 s. All binding experiments were performed on a Biacore T200 instrument (Cytiva) and at 25 °C, using a binding buffer composed of 10 mM HEPES, 150 mM NaCl and 0.05% v/v Tween20. Data were analysed using Biacore T200 Evaluation Software 3.1 and a report point at 100 s after the injection stop was used as the binding response.

Molecular modeling. *Ligand parametrization with QM calculations*. The 3D structure of VF-57a was prepared using GaussView 6.0 and optimized at the B3LYP/6-31G(d) level employing the Gaussian 16.0 software package.[56] The minimum energy nature of the optimized compound was verified upon inspection of the vibrational frequencies, which were positive. The ligand was parameterized using the gaff2 force field.[57] Partial atomic charges were derived following the RESP-charges protocol[58,59] at the B3LYP/6-31G(d) level. Arbidol and (*S*)-F0045 were parametrized using the same protocol. In this case, the geometry used in geometry optimization was taken from the crystallographic structures available in the Protein Data Bank (PDB ID 5T6N[12] and 6WCR,[30] respectively).

Molecular modeling: Homology modelling and system setup. The 3D models of H1 and H3 HA were built by homology-modelling using the SWISSMODEL.[60] Two models were built up for the H1 HA. The first relied on the complex formed by (*S*)-F0045 with PR8 (PDB ID 6WCR),[30] where modeling was used to unfold the last helical turn of the short α-helix in HA_2 in order to simulate the local structure observed in this region in the complex between HA and arbidol (PDB ID 5T6N),[12] and hence enabling the binding of the ligands to the arbidol binding site in the PR8 strain. The second was the homology model obtained for the sequence of Virg09 strain using 6WCR[30] as structural template. For the H3 model, the sequence of the H3N2 A/Hong Kong/7/1987 strain was employed using the X-ray structure of Arbidol bound to H3 HA A/Hong Kong/1/1968 as structural template (PDB ID 5T6N).[12]

For the H1 model, two distinct binding sites were explored for **VF-57a** (namely Site A and Site B). In Site A, the X-ray structure of the Arbidol-bound HA and the modified H1 model (with the aforementioned last turn of the α-helix in HA_2 subunit unfolded) were aligned. Then, **VF-57a** was placed in the binding pocket (see SI, Figure S2) through overlay of the thiophene ring of **VF-57a** onto the thiophenyl ring of arbidol, taking advantage of the hydrophobic nature of the residues surrounding this region of the binding site. In Site B, **VF-57a** was overlaid on the chemical skeleton of (*S*)-F0045 (see SI, Figure S1) retaining the hydrogen bond between the carbonyl oxygen and the hydroxyl group of $T318_2$. However, two distinct binding modes were explored depending on the overlay of the halogenated phenyl (binding mode A) or the thiophene unit (binding mode B) onto the 2,5-dichlorophenyl moiety of (*S*)-F0045. For the H3 model, **VF-57a** was simply docked into the arbidol binding pocket.

To calibrate the structural stability of the complexes, the crystallographic complexes of (*S*)-F0045 in PR8 HA and arbidol in HK HA were used as control systems.

As a final remark, let us note that due to the trimeric nature of HA, the complexes formed by HA with either arbidol, (*S*)-F0045 and **VF-57a** were built up using a stoichiometric ratio 1:1, that is, with 3 ligands bound to the trimeric HA.

Sequence alignment. Sequences of the studied Influenza A strains (see SI, Figure S4) were aligned using the multiple sequence alignment tool of CLUSTAL OMEGA (v. 1.2.4) in the EMBL's European Bioinformatics Institute (EMBL-EBI) web server.[61]

Molecular Dynamics Simulations. Each system was minimized using 10000 steps of steepest descent in combination with 5000 steps of conjugate gradient algorithm. Then, each system was equilibrated in 2 steps for a total simulation time of 5 ns. The systems were heated in the NVT ensemble from 5 K to 300 K in a 250 ps temperature ramp and kept at 300K fot 50 ps more. Subsequently, the density of the system was equilibrated for 4.7 ns in the NPT ensemble (pressure: 1 bar, T: 300 K).

Production runs were done for 500 ns. Temperature control was achieved using Langevin dynamics and pressure control was maintained using the Berendsen barostat. All bonds involving hydrogen atoms were constrained by the SHAKE algorithm[62] in order to employ a timestep of 2 fs. The amberff14sb force field[63] was used for the protein, Joung and Cheatham III parameters were used for the counterions[64] and the TIP3P model for water.[65] Counterions (K^+ and Cl^-) were added to maintain the neutrality of the simulated system and to maintain the ionic concentration at 0.15 M following the SPLIT method.[66] All simulations were performed with *AMBER20* package.[67] During all the simulations a set of distance NMR restraints were applied to maintain the structural stability of the HA trimer, since the models (constructed from the mentioned X-ray structures above) miss the transmembrane part of the protein. The restraints involved one residue of each HA monomer located at the end-chains. A force constant of 5 kcal/mol was applied gradually when a displacement higher than 2Å from the crystallographic distance occurs (the restraint becomes fixed at 5kcal/mol 1.5 Å after the first cutoff).

Relative Binding Free Energy (RBFE) Simulations. The relative binding free energy (RBFE)[68] between derivatives was determined through alchemical transformations, wherein a ligand (L1) is converted into a structurally related analogue (L2) both in the protein-bound complex and in the unbound state in aqueous solution (see Eq. 1).

$$\Delta\Delta G_{bind}(L1 \rightarrow L2) = \Delta G_{bound}(L1 \rightarrow L2) - \Delta G_{water}(L1 \rightarrow L2) \quad \text{(Eq. 1)}$$

The transformation from L1 to L2 was divided into a series of windows, where λ = 0 and λ = 1 represent L1 and L2, respectively, and intermediate λ values denote a linear interpolation between the parameters of the initial and final systems.[69] A dual-topology alchemical perturbation protocol was employed, utilizing 23 λ-windows evently spaced following the trapezoidal rule. To smoothly transition from L1 to L2, softcore potentials were applied to the atoms involved in a chemical change. Each λ simulation initially ran for 6 ns, adding extensions in the windows that were not converged when necessary. The Thermodynamic Integration (TI) estimator, implemented in alchemlyb,[70] was utilized to estimate the free energy change for each λ simulation. All simulations were conducted using the GPU-accelerated TI implementation[68,71] of Amber20.[67]

Initial coordinates for each binding mode were extracted from unbiased MD runs. All systems were meticulously equilibrated with the dual topology using the TI code in each λ before production runs. Each system underwent heating in the NVT ensemble from 5 K to 300 K (150 ps), followed by equilibration at 1 bar in the NPT ensemble (300 ps). Finally, the density equilibrated system was simulated for 100 ps more switching back to the NVT ensemble preparing the system to conduct the actual RBFE production. During the equilibration process, the heavy atoms of the protein and ligand were restrained with positional restraints (5 kcal·mol^{-1}) additionally to the NMR restraints applied to the bottom of the HA model as in the unbiased MD simulations. Subsequently, each λ state was simulated in the NVT ensemble at 300 K using the Langevin thermostat and the SHAKE algorithm for bonds involving hydrogen atoms.

ASSOCIATED CONTENT

Supporting Information. Synthesis of some starting materials, ^{1}H and ^{13}C NMR spectra and elemental analysis data of the new compounds, Table S1 (elemental analysis data), Figures S1-S4 (PDF). Molecular formula string and data (CSV). This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Authors

E-mail: lieve.naesens@kuleuven.be (biology) Phone: +32 16322098.
ORCID 0000-0001-9742-9302.
E-mail: fjluque@ub.edu (modeling) Phone: +34 934033788.
ORCID 0000-0002-8049-3567.
E-mail: svazquez@ub.edu (chemistry) Phone: +34 934024533.
ORCID 0000-0002-9296-6026.

Author Contributions

‡These authors (S.R. and A.V.) contributed equally.

†Deceased. Óscar Lozano passed away in 2021 while this work was ongoing.

L.N. and S.V. conceived the idea. V.F., J.M., O.L., and C.E.M. synthesized and chemically characterized the compounds. A.V. and F.J.L. performed MD calculations. S.R., R.V.B., C.M., K.V., S.N., A.S. and L.N. performed the antiviral assays and the cytotoxicity studies. L.N., F.J.L., and S.V. analyzed the data. L.N., F.J.L., and S.V. wrote, edited and reviewed the manuscript with feedback from all the authors. All authors have given approval to the final version of the manuscript.

Notes

None of the authors has any disclosures to declare.

ACKNOWLEDGMENT

This work was funded by the Spanish Ministerio de Ciencia, Innovación y Universidades, MICIU/AEI/10.13039/501100011033: Grants PID2023-147004OB-I00 (to S.V.), PID2020-117646RB-I00, PID2023-147942OB-I00 and Maria de Maetzu CEX2021-001202-M) (to F.J.L.). This study was supported by funding from Fundació La Marató de TV3 (Nos. 201832 and 202135 to L.N. and S.V.), and Agència de Gestió d'Ajuts Universitaris i de Recerca (grant 2021SGR00671). C.E.M. was supported by the Research Personnel Training Program (FPI grant PRE2022-104952) of the Spanish *Ministerio de Ciencia, Innovación y Universidades*.

ABBREVIATIONS

ATR, Attenuated Total Reflectance; BXA: baloxavir acid; CPE, Cytopathic Effect, CPE; HA, Hemagglutinin; MCC, Minimum Cytotoxic Concentration; MDCK, Madin-Darby canine kidney cells; ND, Not Determined; SAR, structure-activity relationships; SI, Selectivity Index; TBS, Topliss Batchwise Scheme.

REFERENCES

(1) Iuliano, A. D.; Roguski, K. M.; Chang, H. H.; Muscatello, D. J.; Palekar, R.; Tempia, S.; Cohen, C.; Gran, J. M.; Schanzer, D.; Cowling, B. J.; Wu, P.; Kyncl, J.; Ang, L. W.; Park, M.; Redlberger-Fritz, M.; Yu, H.; Espenhain, L.; Krishnan, A.; Emukule, G.; van Asten, L.; Pereira da Silva, S.;

Aungkulanon, S.; Buchholz, U.; Widdowson, M. A.; Bresee, J. S.; . Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *Lancet* **2018**, *391*, 1285–1300. DOI: 10.1016/s0140-6736(17)33293-2.

(2) Krammer, F.; Smith, G. J. D.; Fouchier, R. A. M.; Peiris, M.; Kedzierska, K.; Doherty, P. C.; Palese, P.; Shaw, M. L.; Treanor, J.; Webster, R. G.; García-Sastre, A. Influenza. *Nat. Rev. Dis. Primers* **2018**, *4*, 3. DOI: 10.1038/s41572-018-0002-y.

(3) Dawood, F. S.; Iuliano, A. D.; Reed, C.; Meltzer, M. I.; Shay D.K.; Cheng, P.-Y.; Bandaranayake, D.; Breiman, R. F.; Brooks, W. A.; Buchy, P.; Feikin, D. R.; Fowler, K. B.; Gordon, A.; Hien N.T.; Horby, P.; Huang, Q. S.; Katz, M. A.; Krishnan, A.; Lal, R.; Montgomery, J. M.; Mølbak, K.; Pebody, R.; Presanis, A. M.; Razuri, H.; Steens, A.; Tinoco, Y. O.; Wallinga, J.; Yu, H.; Vong, S.; Bresee, J.; Widdowson, M.-A. Estimated global mortality associated with the first 12 months of 2009 pandemic influenza A H1N1 virus circulation: a modelling study. *Lancet Infect. Dis.* **2012**, *12*, 687–695. DOI: 10.1016/s1473-3099(12)70121-4.

(4) Neumann, G.; Kawaoka, Y. Highly pathogenic H5N1 avian influenza virus outbreak in cattle: the knowns and unknowns. *Nat. Rev. Microbiol.* **2024**, *22*, 525–526. DOI: 10.1038/s41579-024-01087-1.

(5) Peacock, T. P.; Moncla, L.; Dudas, G.; VanInsberghe, D.; Sukhova, K.; Lloyd-Smith, J. O.; Worobey, M.; Lowen, A. C.; Nelson, M. I. The global H5N1 influenza panzootic in mammals. *Nature* **2025**, *637*, 304–313. DOI: 10.1038/s41586-024-08054-z.

(6) Centers for Disease Control and Prevention. *CDC seasonal flu vaccine effectiveness studies.* https://www.cdc.gov/flu-vaccines-work/php/effectiveness-studies/index.html (accessed 22/03/2025).

(7) Nguyen-Van-Tam, J. S.; Venkatesan, S.; Muthuri, S. G.; Myles, P. R. Neuraminidase inhibitors: who, when, where? *Clin. Microbiol. Infect.*, **2015**, *21*, 222–225. DOI: 10.1016/j.cmi.2014.11.020.

(8) Uyeki, T. M.; Hui, D. S.; Zambon, M.; Wentworth, D. E.; Monto, A. S. Influenza. *Lancet*, **2022**, *400*, 693–706. DOI: 10.1016/s0140-6736(22)00982-5.

(9) Beigel, J. H.; Hayden, F.G. Influenza therapeutics in clinical practice-challenges and recent advances. *Cold Spring Harb. Perspect. Med.*, **2021**, *11*, a038463. DOI: 10.1101/cshperspect.a038463.

(10) Govorkova, E. A.; Takashita, E.; Daniels, R. S.; Fujisaki, S.; Presser, L. D.; Patel, M. C.; Huang, W.; Lackenby, A.; Nguyen, H. T.; Pereyaslov, D.; Rattigan, A.; Brown,, S. K.; Samaan, M.; Subbarao, K.; Wong, S.; Wang, D.; Webby, R. J.; Yen H. L.; Zhang, W.; Meijer, A.; Gubareva, L.V. Global update on the susceptibilities of human influenza viruses to neuraminidase inhibitors and the cap-dependent endonuclease inhibitor baloxavir, 2018-2020. *Antiviral Res.*, **2022**, *200*, 105281. DOI: 10.1016/j.antiviral.2022.105281.

(11) Gamblin, S.J.; Vachieri, S.G.; Xiong, X.; Zhang, J.; Martin, S.R.; Skehel, J.J. Hemagglutinin structure and activities. *Cold Spring Harb. Perspect. Med.*, **2020**, *11*, a038638. DOI: 10.1101/cshperspect.a038638.

(12) Kadam, R. U.; Wilson, I. A. Structural basis of influenza virus fusion inhibition by the antiviral drug Arbidol. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 206–214. DOI: 10.1073/pnas.1617020114.

(13) Blaising, J.; Polyak, S. J.; Pecheur, E. I. Arbidol as a broad-spectrum antiviral: an update. *Antiviral Res.* **2014**, *107*, 84–94. DOI: 10.1016/j.antiviral.2014.04.006.

(14) Di Lella, S.; Herrmann, A.; Mair, C. M. Modulation of the pH stability of influenza virus hemagglutinin: A host cell adaptation strategy. *Biophys. J.* **2016**, *110*, 2293–2301. DOI: 10.1016/j.bpj.2016.04.035.

(15) Benton, D. J.; Gamblin, S. J.; Rosenthal, P. B.; Skehel, J. J. Structural transitions in influenza haemagglutinin at membrane fusion pH. *Nature* **2020**, *583*, 150–153. DOI: 10.1038/s41586-020-2333-6.

(16) Garcia-Moro, E.; Zhang, J.; Calde,r L. J.; Brown N. R.; Gamblin, S. J.; Skehel, J. J.; Rosenthal, P. B. Reversible structural changes in the influenza hemagglutinin precursor at membrane fusion pH. *Proc. Nat. Acad. Sci. U. S. A.* **2022**, *119*, e2208011119. DOI: 10.1073/pnas.2208011119.

(17) Stevaert, A.; Naesens, L. The influenza virus polymerase complex: an update on its structure, functions, and significance for antiviral drug design. *Med. Res. Rev.* **2016**, *36*, 1127–1173. DOI: 10.1002/med.21401.

(18) Vanderlinden, E.; Naesens, L. Emerging antiviral strategies to interfere with influenza virus entry. *Med. Res. Rev.* **2014**, *34*, 301–339. DOI: 10.1002/med.21289.

(19) Chen, Z.; Cui, Q.; Caffrey, M.; Rong, L.; Du, R. Small molecule inhibitors of influenza virus entry. *Pharmaceuticals (Basel)* **2021**, *14*, 587. DOI: 10.3390/ph14060587.

(20) Liu, H. Y.; Yang, P. L. Small-molecule inhibition of viral fusion glycoproteins. Small-Molecule Inhibition of Viral Fusion Glycoproteins. *Ann. Rev. Virol.,* **2021**, *8*, 459–489. DOI: 10.1146/annurev-virology-022221-063725.

(21) Hermoso-Pinilla, F. J.; Valdivia, A.; Camarasa, M. J.; Ginex, T.; Luque, F. J. Influenza A virus hemagglutinin: From classical fusion inhibitors to proteolysis targeting chimera-based strategies in antiviral drug discovery. *Explor. Drug Sci.* **2024**, *2*, 85–116. DOI: 10.37349/eds.2024.00037.

(22) Sun X.; Ma, H.; Wang, X.; Bao, Z.; Tang, S.; Yi, C.; Sun, B. Broadly neutralizing antibodies to combat influenza virus infection. *Antiviral Res.* **2024**, *221*, 105785. DOI: 10.1016/j.antiviral.2023.105785.

(23) van Dongen, M. J. P.; Kadam, R. U.; Juraszek, J.; Lawson, E.; Brandenburg, B.; Schmitz, F.; Schepens, W. B. G.; Stoops, B.; van Diepen, H. A.; Jongeneelen, M.; Tang C.; Vermond, J.; van Eijge-Obregoso, Real A.; Blokland, S.; Garg, D.; Yu W.; Goutier,, W.; Lanckacker, E.; Klap, J. M.; Peeters, D. C. G.; Wu, J.; Buyck,, C.; Jonckers T. H. M.; Roy-mas, D.; Roevens,

P.; Vogels,, R.; Koudstaal W.; Friesen, R. H. E.; Raboisson, P.; Dhanak, D.; Goudsmit, J.; Wilson, I. A. A small-molecule fusion inhibitor of influenza virus is orally active in mice. *Science* **2019**, *363*, eaar6221. DOI: 10.1126/science.aar6221.

(24) Gaisina, I.; Li, P.; Du, R.; Cui, Q.; Dong, M.; Zhang, C.; Manicassamy, B.; Caffrey, M.; Moore, T.; Cooper, L.; Rong, L. An orally active entry inhibitor of influenza A viruses protects mice and synergizes with oseltamivir and baloxavir marboxil. *Sci. Adv.* **2024**, *10*, eadk9004. DOI: 10.1126/sciadv.adk9004.

(25) Zhu, L.; Li, Y.; Li, S.; Li, H.; Qiu, Z.; Lee, C.; Lu, H.; Lin, X.; Zhao, R.; Chen, L.; Wu, J. Z.; Tang, G.; Yang, W. Inhibition of influenza A virus (H1N1) fusion by benzenesulfonamide derivatives targeting viral hemagglutinin. *PLoS One* **2011**, *6*, e29120. DOI: 10.1371/journal.pone.0029120.

(26) Zhao X.; Li,, R.; Zhou, Y.; Xiao, M.; Ma, C.; Yang, Z.; Zeng, S.; Du Q.; Yang,, C.; Jiang, H.; Hu, Y.; Wang, K.; Mok, C. K. P.; Sun, P.; Dong, J.; Cui, W.; Wang, J.; Tu, Y.; Yang, Z.; Hu, W. Discovery of highly potent pinanamine-based inhibitors against amantadine- and oseltamivir-resistant influenza a viruses. *J. Med. Chem.* **2018**, *61*, 5187–5198. DOI: 10.1021/acs.jmedchem.8b00042.

(27) Ye, M.; Liao, Y.; Wu, L.; Qi, W.; Choudhry, N.; Liu, Y.; Chen, W.; Song, G.; Chen, J. An oleanolic acid derivative inhibits hemagglutinin-mediated entry of influenza A virus. *Viruses* **2020**, *12*, 225. DOI: 10.3390/v12020225.

(28) White, K.; Esparza, M.; Liang, J.; Bhat, P.; Naidoo, J.; McGovern, N. L.; Williams, M. A. P.; Alabi, B. R.; Shay, J.; Niederstrasser, H.; Posner, B.; García-Sastre, A.; Ready, J.; Fontoura, B. M. A. Aryl sulfonamide inhibits entry and replication of diverse influenza viruses via the hemagglutinin protein. *J. Med. Chem.* **2021**, *64*, 10951–10966. DOI: 10.1021/acs.jmedchem.1c00304

(29) Russell, R. J.; Kerry, P. S.; Stevens, D. J.; Steinhauer, D. A.; Martin S. R.; Gamblin, S. J.; Skehel, J. J. Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 17736–17741. DOI: 10.1073/pnas.0807142105.

(30) Yao, Y.; Kadam, R. U.; Lee, C. D.; Woehl J. L.; Wu, N. C.; Zhu, X.; Kitamura, S.; Wilson, I. A.; Wolan, D. W. An influenza A hemagglutinin small-molecule fusion inhibitor identified by a new high-throughput fluorescence polarization screen. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117*, 18431–18438. DOI: 10.1073/pnas.2006893117.

(31) Antanasijevic A.; Durst M. A.; Cheng H.; Gaisina I. N.; Perez J. T.; Manicassamy B.; Rong L.; Lavie A.; Caffrey M. Structure of avian influenza hemagglutinin in complex with a small molecule entry inhibitor. *Life Sci. Alliance* **2020**, *3*, e202000724. DOI: 10.26508/lsa.202000724.

(32) Plotch, S. J.; O'Hara, B.; Morin, J.; Palant, O.; LaRocque, J.; Bloom, J. D.; Lang, S. A Jr.; DiGrandi, M. J.; Bradley, M.; Nilakanta, R.; Gluzman, Y. Inhibition of influenza A virus replication by compounds interfering with the fusogenic function of the viral hemagglutinin. *J. Virol.* **1999**, *73*, 140–151. DOI: 10.1128/JVI.73.1.140-151.1999.

(33) Liu S.; Li, R.; Zhang, R.; Chan, C. C.; Xi, B.; Zhu, Z.; Yang, J.; Poon,, V. K.; Zhou, J.; Chen,, M.; Munch, J.; Kirchhoff, F.; Pleschka, S.; Haarmann T.; Dietrich, U.; Pan, C.; Du, L.; Jiang, S.; Zheng, B. CL-385319 inhibits H5N1 avian influenza A virus infection by blocking viral entry. *Eur. J. Pharmacol.* **2011**, *660*, 460–467. DOI: 10.1016/j.ejphar.2011.04.013.

(34) Zhu, Z.; Li, R.; Xiao, G.; Chen, Z.; Yang, J.; Zhu, Q.; Liu, S. Design, synthesis and structure-activity relationship of novel inhibitors against H5N1 hemagglutinin-mediated membrane fusion. *Eur. J. Med. Chem.* **2012**, *57*, 211–216. DOI: 10.1016/j.ejmech.2012.08.041.

(35) Zhu, Z.; Yao, Z.; Shen, X.; Chen, Z.; Liu, X.; Parquette, J. R.; Liu, S. Oligothiophene compounds inhibit the membrane fusion between H5N1 avian influenza virus and the endosome of host cell. *Eur. J. Med. Chem.* **2017**, *130*, 185–194. DOI: 10.1016/j.ejmech.2017.02.040.

(36) Yu, Y.; Tazeem, Xu, Z.; Du, L.; Jin, M.; Dong, C.; Zhou, H.-B.; Wu, S. Design and synthesis of heteroaromatic-based benzenesulfonamide derivatives as potent inhibitors of H5N1 influenza A virus. *MedChemComm* **2019**, *10*, 89–100. DOI: 10.1039/c8md00474a.

(37) Vanderlinden, E.; Göktas, F.; Cesur, Z.; Froeyen, M.; Reed, M. L.; Russell, C. J.; Cesur, N.; Naesens, L. Novel inhibitors of influenza virus fusion: structure-activity relationship and interaction with the viral hemagglutinin. *J. Virol.* **2010**, *84*, 4277–4288. DOI: 10.1128/JVI.02325-09.

(38) Cihan-Ustundag, G.; Zopun, M.; Vanderlinden, E.; Ozkirimli, E.; Persoons, L.; Capan, G.; Naesens, L. Superior inhibition of influenza virus hemagglutinin-mediated fusion by indole-substituted spirothiazolidinones. *Bioorg. Med. Chem.* **2020**, *28*, 115130. DOI: 10.1016/j.bmc.2019.115130.

(39) Du, R.; Cheng, H.; Cui, Q.; Peet, N. P.; Gaisina, I. N.; Rong, L. Identification of a novel inhibitor targeting influenza A virus group 2 hemagglutinins. *Antiviral Res.* **2021**, *186*, 105013. DOI: 10.1016/j.antiviral.2021.105013.

(40) Alqarni, S.; Cooper, L.; Achi, J. G.; Bott, R.; Sali, V. K.; Brown, A.; Santarsiero, B. D.; Krunic, A.; Manicassamy, B.; Peet, N. P.; Zhang, P.; Thatcher, G. R. J.; Gaisina, I. N.; Rong, L.; Moore, T. W. Synthesis, optimization, and structure-activity relationships of imidazo[1,2-*a*]pyrimidines as inhibitors of group 2 influenza A viruses. *J. Med. Chem.* **2022**, *65*, 14104–14120. DOI: 10.1021/acs.jmedchem.2c01329

(41) White, K. M.; De Jesus, P.; Chen, Z.; Abreu, P., Jr.; Barile, E.; Mak, P. A.; Anderson, P.; Nguyen, Q. T.; Inoue, A.; Stertz, S.; Koening, R.; Pellecchia, M.; Palese, P.; Kuhen, K.; García-Sastre, A.; Chanda, S. K.; Shaw, M. L. A potent anti-influenza compound blocks fusion through stabilization of the prefusion conformation of the hemagglutinin protein. *ACS Infect. Dis.* **2015**, *1*, 98–109. DOI: 10.1021/id500022h.

(42) Basu, A.; Antanasijevic A.; Wang, M.; Li, B.; Mills, D. M.; Ames, J. A.; Nash, P. J.; Williams, J. D.; peet, N. P.; Moir, D. T.; Prichard, M. N.; Keith, K. A.; Barnard, D. L.; Caffrey, M.; Rong, L.; Bowlin, T. L. New small molecule entry inhibitors targeting hemagglutinin-mediated influenza A virus fusion. *J. Virol.* **2014**, *88*, 1447–1460. DOI: https://doi.org/10.1128/jvi.01225-13.

(43) De Castro, S.; Ginex, T.; Vanderlinden, E.; Laporte, M.; Stevaert, A.; Cumella, J.; Gago, F.; Camarasa, M. J.; Luque, F. J.; Naesens, L.; Velázquez, S. N-benzyl 4,4-disubstituted piperidines as a potent class of influenza H1N1 virus inhibitors showing a novel mechanism of hemagglutinin fusion peptide interaction. *Eur. J. Med. Chem.* **2020**, *194*, 112223. DOI: 10.1016/j.ejmech.2020.112223.

(44) Kitamura, S.; Lin, T.-H.; Lee, C.-C. D.; Takamura, A.; Kadam, R. U.; Zhang, D.; Zhu, X.; Dada, L.; Nagai, E.; Yu, W.; Yao, Y.; Sharpless, K. B.; Wilson, I. A.; Wolan, D. W. Ultrapotent influenza hemagglutinin fusion inhibitors developed through SuFEx-enabled high-throughput medicinal chemistry. *Proc. Nat. Acad. Sci. U. S. A.* **2024**, *121*, e2310677121. DOI: 10.1073/pnas.2310677121.

(45) Kim, J. I.; Lee, S.; Lee, G. Y.; Park, S.; Bae, J.-Y.; Heo, J.; Kim, H.-Y.; Woo, S.-H.; Lee, H. U.; Ahn, C. A.; Bang, H. J.; Ju, H. S.; Ok, K.; Byun, Y.; Cho, D.-J.; Shin, J. S.; Kim, D.-Y.; Park, M. S.; Park, M.-S. Novel small molecule targeting the hemagglutinin stalk of influenza viruses. *J. Virol.* **2019**, *93*, e99878-19. DOI: 10.1128/JVI.00878-19

(46) Leiva, R.; Barniol-Xicota, M.; Codony, S.; Ginex, T.; Vanderlinden, E.; Montes, M.; Caffrey, M.; Luque, F. J.; Naesens, L.; Vazquez, S. Aniline-based inhibitors of influenza H1N1 virus acting on hemagglutinin-mediated fusion. *J. Med. Chem.*, **2018**, *61*, 98–118. 10.1021/acs.jmedchem.7b00908.

(47) Topliss, J. G. A Manual method for applying the Hansch approach to drug design. *J. Med. Chem.* **1977**, *20*, 463–469. DOI: https://doi.org/10.1021/jm00214a001.

(48) Richter, L. Topliss batchwise schemes reviewed in the era of open data reveal significant differences between enzymes and membrane receptors. *J. Chem. Inf. Model.* **2017**, *57*, 2575–2583. DOI: 10.1021/acs.jcim.7b00195.

(49) Kandeil, A.; Patton, C.; Jones, J. C.; Jeevan, T.; Harrington, W. N.; Trifkovic, S.; Seiler, J. P.; Fabrizio, T.; Woodard, K.; Turner, J. C.; Crumpton, J.C-.; Miller, L.; Rubrum, A.; DeBeauchamp, J.; Russell, C. J.; Govorkova, E. A.; Vogel, P.; Kim-Torchetti, M.; Berhane, Y.; Stallknecht, D.; Poulson, R.; Kercher, L.; Webby, R. J. Rapid evolution of A(H5N1) influenza viruses after intercontinental spread to North America. *Nat. Commun.* **2023**, *14*, 3082. DOI: 10.1038/s41467-023-38415-7.

(50) . Okuno, Y.; Isegawa, Y.; Sasao, F.; Ueda, S. A common neutralizing epitope conserved between the hemagglutinins of influenza A virus H1 and H2 strains. *J. Virol.* **1993**, *67*, 2552-2558. DOI: 10.1128/JVI.67.5.2552-2558.1993.

(51) Dreyfus, C.; Ekiert, D. C.; Wilson, I. A. Structure of a classical broadly neutralizing stem antibody in complex with a pandemic H2 influenza virus hemagglutinin. *J. Virol.* **2013**, *87*, 7149-7154. DOI: 10.1128/JVI.02975-12.

(52) Yang, H.; Carney, P. J.; Chang, J. C.; Guo, Z.; Villanueva, J. M.; Stevens, J. Structure and receptor binding preferences of recombinant human A(H3N2) virus hemagglutinins. *Virology* **2015**, *477*, 18-31. DOI: 10.1016/j.virol.2014.12.024.

(53) Vrijens, P.; Noppen, S.; Boogaerts, T.; Vanstreels, E.; Ronca, R.; Chiodelli, P.; Laporte, M.; Vanderlinden, E.; Liekens, S.; Stevaert, A.; Naesens, L. Influenza virus entry via the GM3 ganglioside-mediated platelet-derived growth factor receptor beta signalling pathway. *J. Gen. Virol.* **2019**, *100*, 583-601. DOI: 10.1099/jgv.0.001235.

(54) Laporte, M.; Stevaert, A.; Raeymaekers, V.; Boogaerts, T.; Nehlmeier, I.; Chiu, W.; Benkheil, M.; Vanaudenaerde, B.; Pohlmann, S.; Naesens, L. Hemagglutinin cleavability, acid stability, and temperature dependence optimize influenza B virus for replication in human airways. *J. Virol.* **2019**, *94*, e01430-01419. DOI: 10.1128/JVI.01430-19.

(55) Glowacka, I.; Bertram, S.; Muller, M. A.; Allen, P.; Soilleux, E.; Pfefferle, S.; Steffen, I.; Tsegaye, T. S.; He, Y.; Gnirss, K.; Niemeyer, D.; Schneider, H.; Drosten, C.; Pöhlmann, S. Evidence that TMPRSS2 activates the severe acute respiratory syndrome coronavirus spike protein for membrane fusion and reduces viral control by the humoral immune response. *J. Virol.* **2011**, *85* (9), 4122-4134. DOI: 10.1128/jvi.02232-10.

(56) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H.; Li, X.; Caricato, M.; Marenich, A. V; Bloino, J.; Janesko, B. G.; Gomperts, R.; Mennucci, B.; Hratchian, H. P.; Ortiz, J. V; Izmaylov, A. F.; Sonnenberg, J. L.; Williams-Young, D.; Ding, F.; Lipparini, F.; Egidi, F.; Goings, J.; Peng, B.; Petrone, A.; Henderson, T.; Ranasinghe, D.; Zakrzewski, V. G.; Gao, J.; Rega, N.; Zheng, G.; Liang, W.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Throssell, K.; Montgomery Jr., J. A.; Peralta, J. E.; Ogliaro, F.; Bearpark, M. J.; Heyd, J. J.; Brothers, E. N.; Kudin, K. N.; Staroverov, V. N.; Keith, T. A.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A. P.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Millam, J. M.; Klene, M.; Adamo, C.; Cammi, R.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Farkas, O.; Foresman, J. B.; Fox, D. J. Gaussian16 Revision C.01. 2016.

(57) He, X.; Man, V. H.; Yang, W.; Lee, T.−S.; Wang, J. A fast and high-quality charge model for the next generation general AMBER force field. *J. Chem. Phys.* **2020**, *153*, 114502. DOI: 10.1063/5.0019056.

(58) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Kollman, P. A. A Second Generation Force Field for the Simulation Of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1993**, *115*, 9620–9631. DOI: 10.1021/ja00124a002.

(59) Bayly, C. I.; Cieplak, P.; Cornell, W. D.; Kollman, P. A. *J. Phys. Chem.* **1993**, *97*, 10269–10280. Bayly, C. I.; Cieplak,

P.; Cornell, W. D.; Kollman, P. A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **1993**, *97*, 10269–10280.

(60) Waterhouse, A.; Bertoni, M.; Bienert, S.; Studer, G.; Tauriello, G.; Gumienny, R.; Heer, F. T.; de Beer, T. A. P.; Rempfer, C.; Bordoli, L.; Lepore, R.; Schwede, T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **2018**, *46*, W296–W303. DOI: 10.1093/nar/gky427.

(61) Madeira, F.; Pearce, M.; Tivey, A. R. N.; Basutkar, P.; Lee, J.; Edbali, O.; Madhusoodanan, N.; Kolesnikov, A.; Lopez, R. Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.* **2022**, *50*, W276–W279. DOI: 10.1093/nar/gkac240.

(62) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977,** 23, 327. DOI: 10.1016/0021-9991(77)90098-5.

(63) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. Evaluation of all-atom force fields in viral capsid simulations and properties. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713. DOI: 10.1039/d1ra08431c.

(64) Joung, I. S.; Cheatham, T. E. Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *J. Phys. Chem. B* **2008**, *112*, 9020–9041. DOI: 10.1021/jp8001614.

(65) Mark, P.; Nilsson, L. Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K. *J. Phys. Chem. A* **2001**, *105*, 9954–9960. DOI: 10.1021/jp003020w.

(66) Machado, M. R.; Pantano, S. Split the Charge Difference in Two! A Rule of Thumb for Adding Proper Amounts of Ions in MD Simulations. *J. Chem. Theory Comput.* **2020**, *16*, 1367–1372. DOI: 10.1021/acs.jctc.9b00953.

(67) Case, D. A.; K. Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Kasavajhala, K.; Kovalenko, A.; Krasny, R.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; Luchko, T.; Luo, R.; Man, V.; Merz, K. M.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K. A.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shajan, A.; Shen, J.; Simmerling, C. L.; Skrynnikov, N. R.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wang, J.; Wei, H.; Wu, X.; Wu, Y.; Xiong, Y.; Xue, Y.; York, D. M.; Zhao, S.; Zhu, Q.; Kollman, P. A. AMBER 2020. University of California: San Francisco. 2020.

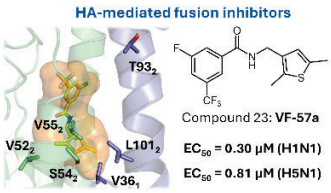
(68) Song, L. F.; Lee, T.-S.; Zhu, C.; York, D. M.; Merz, K. M. Jr. Using AMBER18 for Relative Free Energy Calculations. *J Chem Inf Model* **2019**, *59*, 3128–3135. DOI: 10.1021/acs.jcim.9b00105.

(69) Mey, A. S. J. S.; Allen, B. K.; Bruce McDonald, H. E.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. *Living J.* Best Practices for Alchemical Free Energy Calculations. *Comput. Mol. Sci.* **2020**, *2*, 18378. DOI: 10.33011/livecoms.2.1.18378.

(70) Chodera, J. D. A Simple Method for Automated Equilibration Detection in Molecular Simulations. *J. Chem. Theory Comput.* **2016**, *12*, 1799–1805. DOI: 10.1021/acs.jctc.5b00784.

(71) Hu, Y.; Muegge, I. In Silico Positional Analogue Scanning with Amber GPU-TI. *J. Chem. Inf. Model.* **2022**, *62*, 4448–4459. DOI: 10.1021/acs.jcim.2c00860.

Insert Table of Contents artwork here



Supporting Information

N-[(Thiophen-3-yl)methyl]benzamides as Influenza Virus Fusion Inhibitors Acting on H1 and H5 Hemagglutinins

Silke Rimaux[1‡], Aitor Valdivia[2,3‡], Juan Martín-López[4,5], Valeria Francesconi[5,#], Celia Escriche[4,5], Cato Mestdagh[1], Ria Van Berwaer[1], Kaat Verleye[1], Samuel Noppen[1], Óscar Lozano[4,5†], Annelies Stevaert[1], Francisco Javier Luque[2,3,4*], Lieve Naesens[1*], Santiago Vázquez[4,5*]

[1]Rega Institute, Department of Microbiology, Immunology and Transplantation, KU Leuven, B-3000 Leuven, Belgium.

[2]Doctorate in Biotecnology, Department of Nutrition, Food Science and Gastronomy, Faculty of Pharmacy and Food Sciences, Universitat de Barcelona, Av. Prat de la Riba 171, Santa Coloma de Gramanet, E-08921, Spain.

[3]Institute of Theoretical and Computational Chemistry (IQTCUB), Universitat de Barcelona, 08028 Barcelona, Spain.

[4]Institute of Biomedicine of the University of Barcelona (IBUB), Universitat de Barcelona, 08028 Barcelona, Spain.

[5]Laboratori de Química Farmacèutica, Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Av. Joan XXIII, 27-31, 08028 Barcelona, Spain.

[‡]Equal contribution to this manuscript.

[#]Current address: Department of Pharmacy, University of Genoa, viale Benedetto XV n.3, Genoa 16132, Italy.

[†]Deceased.

Table of contents

Synthesis of starting materials	Page S4
NMR spectra of key intermediates	Page S8
^{1}H and ^{13}C NMR spectra of compound **22**	Page S10
^{1}H and ^{13}C NMR spectra of compound **23**	Page S11
^{1}H and ^{13}C NMR spectra of compound **24**	Page S12
^{1}H and ^{13}C NMR spectra of compound **25**	Page S13
^{1}H and ^{13}C NMR spectra of compound **26**	Page S14
^{1}H and ^{13}C NMR spectra of compound **27**	Page S15
^{1}H and ^{13}C NMR spectra of compound **28**	Page S16
^{1}H and ^{13}C NMR spectra of compound **29**	Page S17
^{1}H and ^{13}C NMR spectra of compound **30**	Page S18
^{1}H and ^{13}C NMR spectra of compound **31**	Page S19
^{1}H and ^{13}C NMR spectra of compound **32**	Page S20
^{1}H and ^{13}C NMR spectra of compound **33**	Page S21
^{1}H and ^{13}C NMR spectra of compound **34**	Page S22
^{1}H and ^{13}C NMR spectra of compound **35**	Page S23
^{1}H and ^{13}C NMR spectra of compound **36**	Page S24
^{1}H and ^{13}C NMR spectra of compound **37**	Page S25
^{1}H and ^{13}C NMR spectra of compound **38**	Page S26
^{1}H and ^{13}C NMR spectra of compound **39**	Page S27
^{1}H and ^{13}C NMR spectra of compound **40**	Page S28
^{1}H and ^{13}C NMR spectra of compound **41**	Page S29
^{1}H and ^{13}C NMR spectra of compound **42**	Page S30
^{1}H and ^{13}C NMR spectra of compound **43**	Page S31
^{1}H and ^{13}C NMR spectra of compound **44**	Page S32
^{1}H and ^{13}C NMR spectra of compound **45**	Page S33
^{1}H and ^{13}C NMR spectra of compound **46**	Page S34
^{1}H and ^{13}C NMR spectra of compound **47**	Page S35
^{1}H and ^{13}C NMR spectra of compound **48**	Page S36
^{1}H and ^{13}C NMR spectra of compound **49**	Page S37
^{1}H and ^{13}C NMR spectra of compound **50**	Page S38
^{1}H and ^{13}C NMR spectra of compound **51**	Page S39
^{1}H and ^{13}C NMR spectra of compound **52**	Page S40
^{1}H and ^{13}C NMR spectra of compound **53**	Page S41
^{1}H and ^{13}C NMR spectra of compound **54**	Page S42
^{1}H and ^{13}C NMR spectra of compound **55**	Page S43
^{1}H and ^{13}C NMR spectra of compound **56**	Page S44
^{1}H and ^{13}C NMR spectra of compound **57**	Page S45
^{1}H and ^{13}C NMR spectra of compound **58**	Page S46
^{1}H and ^{13}C NMR spectra of compound **59**	Page S47
^{1}H and ^{13}C NMR spectra of compound **60**	Page S48

^{1}H and ^{13}C NMR spectra of compound **61** Page S49
^{1}H and ^{13}C NMR spectra of compound **62** Page S50
^{1}H and ^{13}C NMR spectra of compound **63** Page S51
^{1}H and ^{13}C NMR spectra of compound **64** Page S52
^{1}H and ^{13}C NMR spectra of compound **65** Page S53
Table S1. Elemental analysis data. Page S54
Figure S1. Representation of the two binding modes examined for VF-57a upon binding to site B in the H1 HA (A/Virginia/ATCC3/2009). Page S55
Figure S2. Representation of the binding mode examined for VF-57a upon binding to site A. Page S55
Figure S3. RMSD plot for each HA-stem backbone atoms for the simulation of VF-57a in site A binding pocket of the A/Virginia/ATCC3/2009 Influenza A strain (H1N1). Page S56
Figure S4. Sequence alignment for the studied Influenza A strains. Page S57

SYNTHESIS OF STARTING MATERIALS

(2,5-dimethylthiophen-3-yl)methanamine. Thionyl chloride (2.32 mL, 3.81 g, 32.00 mmol) was added to a solution of 2,5-dimethylthiophen-3-carboxylic acid (1.0 g, 6.40 mmol) in anh. toluene (15.0 mL) and the complex mixture was kept under reflux for 2 h. After reaction completed, the solution was concentrated. Without further purification, the crude product 2,5-dimethylthiophen-3-carbonyl chloride was immediately used for the next step.

Ammonium hydroxide 25% aqueous solution (10.0 mL, 64.00 mmol) was added dropwise to a solution of 2,5-dimethylthiophen-3-carbonyl chloride in DCM (5.0 mL) at 0 °C and the mixture was allowed to warm up to room temperature and stirred for 1 h. Brine (30 mL) followed by DCM (20 mL) were added and the mixture was extracted. The aqueous layer was extracted again with DCM (2 x 20 mL). All the organic layers were joined, dried over anh. Na_2SO_4 and filtered to afford the crude product 2,5-dimethylthiophene-3-carboxamide that was used as such without further purification.

To a solution of 2,5-dimethylthiophene-3-carboxamide in anh. THF (10.0 mL) at 0 °C, $LiAlH_4$ 1 M in THF (32.0 mL, 32.00 mmol) was added dropwise. Then, the suspension was heated to reflux and stirred for 4 h. The reaction suspension was quenched adding water dropwise under ice-bath until no more bubbling was observed. Then, anh. Na_2SO_4 was added and the mixture was filtered through a pad with Celite® using MeOH (3 x 30 mL) as eluting agent. The solvent was concentrated *in vacuo* and the resulting crude was purified by column chromatography in silica gel (using as eluent mixtures of MeOH in DCM from 0% to 7%). Fractions containing the desired product were collected and concentrated *in vacuo*.

An excess of HCl 2 M in Et_2O was added to the suspension of the amine in Et_2O (5.0 mL) to form its hydrochloride salt, followed by filtration of the solid to afford the title

compound as a brownish solid (620 mg, 55% yield). ^{1}H NMR (400 MHz, CD_3OD) δ: 2.40 (m, 3H), 2.41 (s, 3H), 3.97 (s, 2H), 6.71 (s, 1H).

(2-methylthiophen-3-yl)methanamine. Thionyl chloride (6.96 mL, 11.42 g, 96.00 mmol) was added to a solution of 2-methylthiophen-3-carboxylic acid (1.36 g, 9.60 mmol) in anh. toluene (25.0 mL) followed by 20 drops of DMF and the mixture was kept under reflux for 2 h. After reaction completed, the solution was concentrated. Without further purification, the crude product 2-methylthiophen-3-carbonyl chloride was immediately used for the next step.

Ammonium hydroxide 25% aqueous solution (14.4 mL, 192.00 mmol) was added dropwise to a solution of 2-methylthiophen-3-carbonyl chloride in DCM (15 mL) at 0 °C and the mixture was allowed to warm up to room temperature and stirred for 1 h. Brine (40 mL) followed by DCM (30 mL) were added and the mixture was extracted. The aqueous layer was extracted again with DCM (2 x 30 mL). All the organic layers were joined, dried over anh. Na_2SO_4 and filtered to afford the crude product 2-methylthiophene-3-carboxamide that was used as such without further purification.

To a solution of 2-methylthiophene-3-carboxamide in anh. THF (40.0 mL) at 0 °C, $LiAlH_4$ (1.82 g, 48.00 mmol) was added portionwise. Then, the suspension was heated to reflux and stirred for 4 h. The reaction suspension was quenched adding water dropwise under ice-bath until no more bubbling was observed. Then, anh. Na_2SO_4 was added and the mixture was filtered through a pad with Celite® using MeOH (3 x 50 mL) as eluting agent. The solvent was concentrated *in vacuo* and the resulting crude was purified by column chromatography in silica gel (using as eluent mixtures of MeOH in DCM from 1% to 4%). Fractions containing the desired product were collected and concentrated *in vacuo*.

An excess of HCl 2 M in Et_2O was added to the suspension of the amine in Et_2O (10.0 mL) to form its hydrochloride, followed by filtration of the solid to afford the title compound as a beige solid (606 mg, 53% yield). ^{1}H NMR (400 MHz, CD_3OD) δ: 2.50 (s, 3H), 4.07 (s, 2H), 7.05 (d, *J* = 5.3 Hz, 1H), 7.25 (d, *J* = 5.3 Hz, 1H).

(5-methylthiophen-3-yl)methanamine. Thionyl chloride (3.48 mL, 5.71 g, 48.00 mmol) was added to a solution of 5-methylthiophen-3-carboxylic acid (680 mg, 4.80 mmol) in anh. toluene (12.0 mL) followed by 10 drops of DMF and the mixture was kept under reflux for 2 h. After reaction completed, the solution was concentrated. Without further purification, the crude product 5-methylthiophen-3-carbonyl chloride was immediately used for the next step.

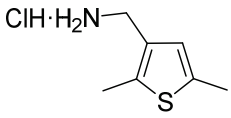
Ammonium hydroxide 25% aqueous solution (7.2 mL, 96.00 mmol) was added dropwise to a solution of 5-methylthiophen-3-carbonyl chloride in DCM (7 mL) at 0 °C and the mixture was allowed to warm up to room temperature and stirred for 1 h. Brine (20 mL) followed by DCM (15 mL) were added and the mixture was extracted. The aqueous layer was extracted again with DCM (2 x 15 mL). All the organic layers were joined, dried over anh. Na_2SO_4 and filtered to afford the crude product 5-methylthiophene-3-carboxamide that was used as such without further purification.

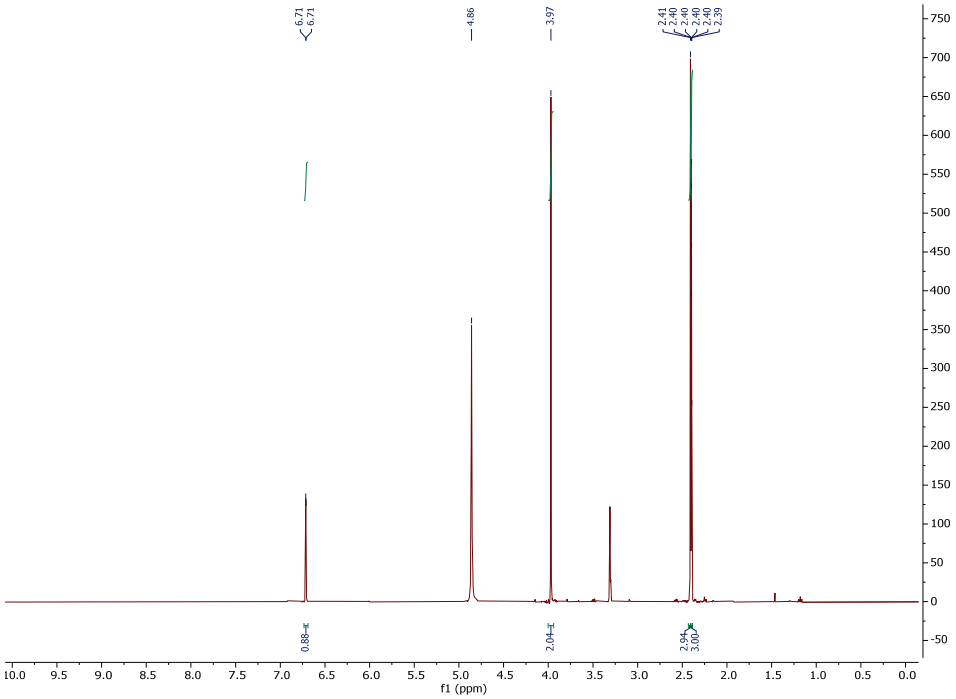
To a solution of 5-methylthiophene-3-carboxamide in anh. THF (20.0 mL) at 0 °C, $LiAlH_4$ (910 mg, 24.00 mmol) was added portionwise. Then, the suspension was heated to reflux and stirred for 4 h. The reaction suspension was quenched adding water dropwise under ice-bath until no more bubbling was observed. Then, anh. Na_2SO_4 was added and the mixture was filtered through a pad with Celite® using MeOH (3 x 25 mL) as eluting agent. Solvent was concentrated *in vacuo* and the resulting crude was purified by column chromatography in silica gel (using as eluent mixtures of MeOH in DCM from 1% to 5%). Fractions containing the desired product were collected and concentrated *in vacuo*.

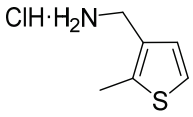
An excess of HCl 2 M in Et_2O was added to the suspension of the amine in Et_2O (5.0 mL) to form its hydrochloride, followed by filtration of the solid to afford the title compound as a beige solid (127 mg, 16% yield). ^1H NMR (400 MHz, CD_3OD) δ: 2.48 (d, J = 1.0 Hz, 3H), 4.05 (s, 2H), 6.87 (m, 1H), 7.26 (m, 1H).

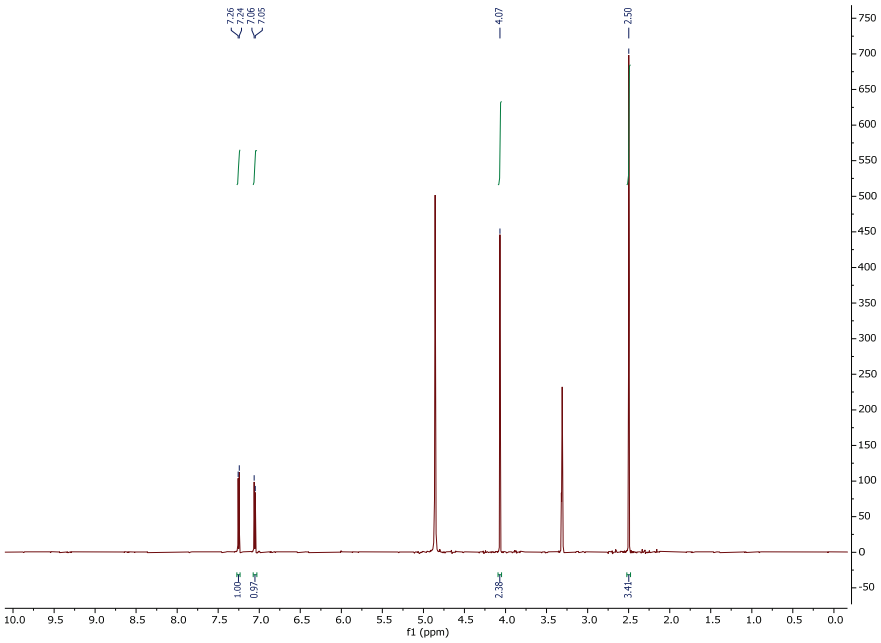
(2,5-dimethylthiophen-3-yl)methanol. To a solution of ethyl 2,5-dimethylthiophene-3-carboxylate (100 mg, 0.64 mmol) in anh. THF (3.0 mL), $LiAlH_4$ (73 mg, 1.92 mmol) was added portinonwise under ice-bath. The mixture was stirred at RT for 16 h. Under an ice-bath, HCl 2 M (2.0 mL) was added and the mixture was stirred for 10 more min. Then, water (10 mL) followed by EtOAc (15 mL) were added and the mixture was extracted. The organic layer was washed again with brine (15 mL), then it was dried over anh. Na_2SO_4 and filtered. Solvents were concentrated *in vacuo* and the resulting crude was purified by column chromatography in silica gel (using as eluent mixtures of EtOAc in hexane from 0% to 15%). Fractions containing the desired product were collected and concentrated *in vacuo* to afford a colorless oil (88 mg, 97% yield). Spectroscopic data matched with those reported in SciFinder® (taken from Enamine Ltd., spectrum ID: EN300-55412).

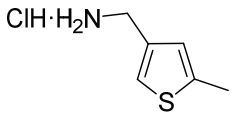
NMR spectra of key intermediates

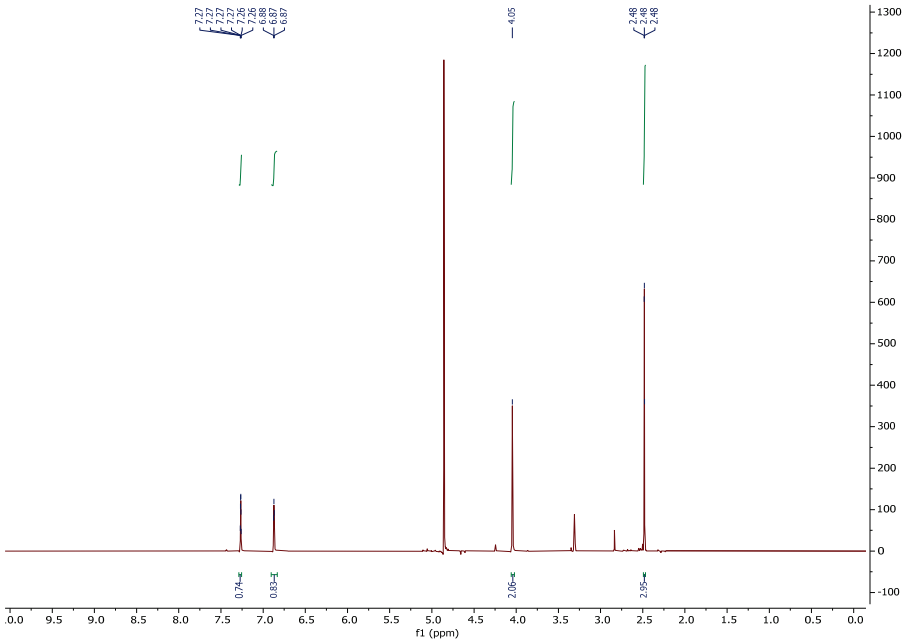


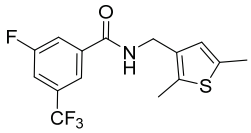




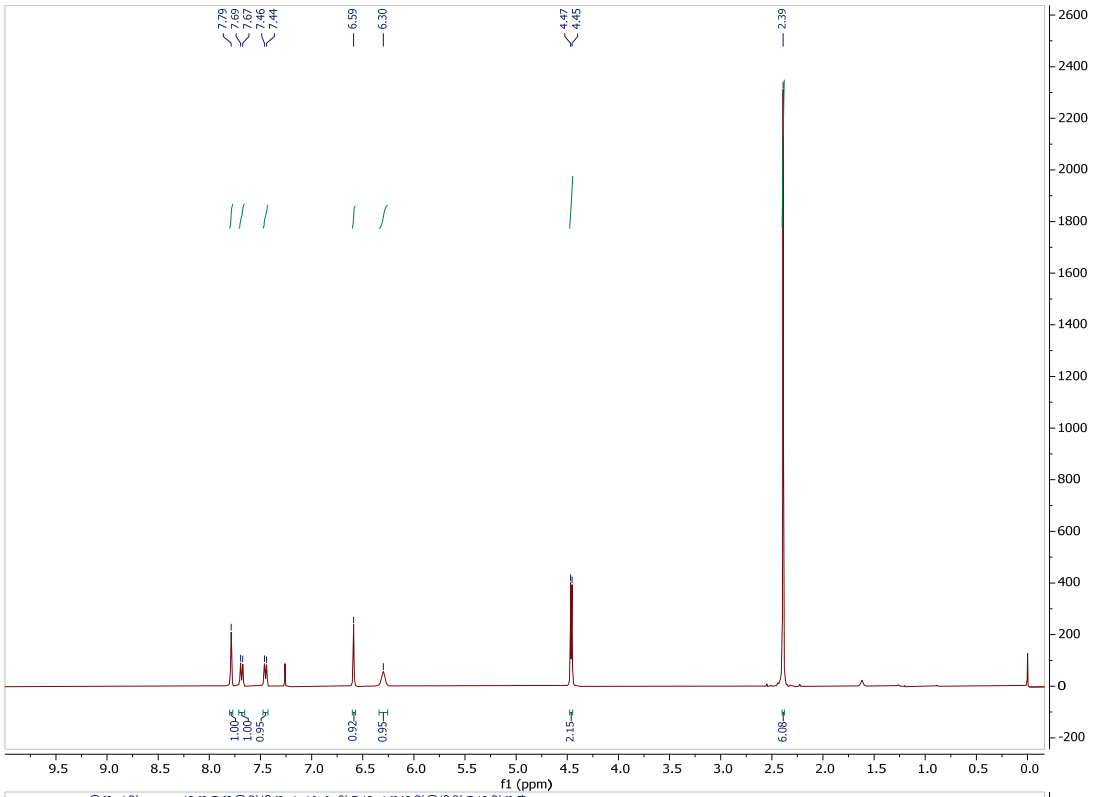


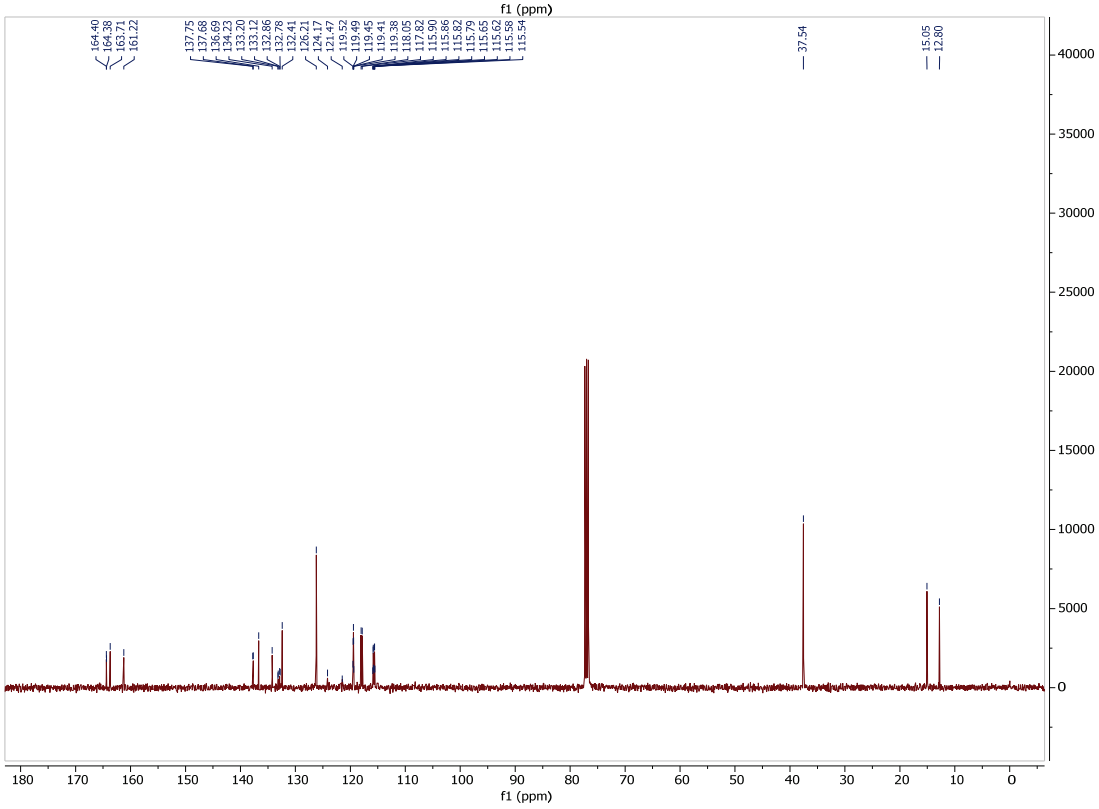

ClH·H2N
S

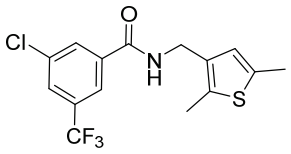

7.27
7.27
7.27
7.27
7.26
7.26
6.88
6.87
6.87
4.05
2.48
2.48
2.48
0.74
0.83
2.06
2.95
f1 (ppm)


O
F
N
H
S
CF3

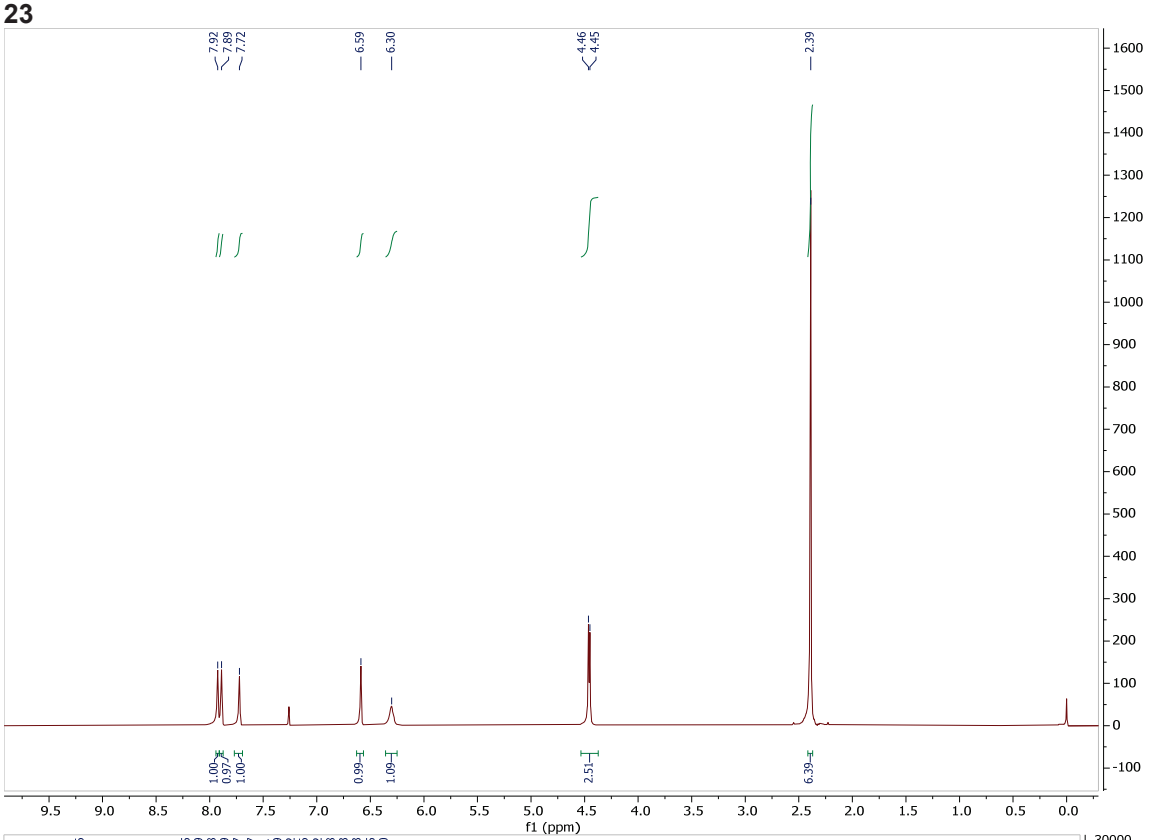
22

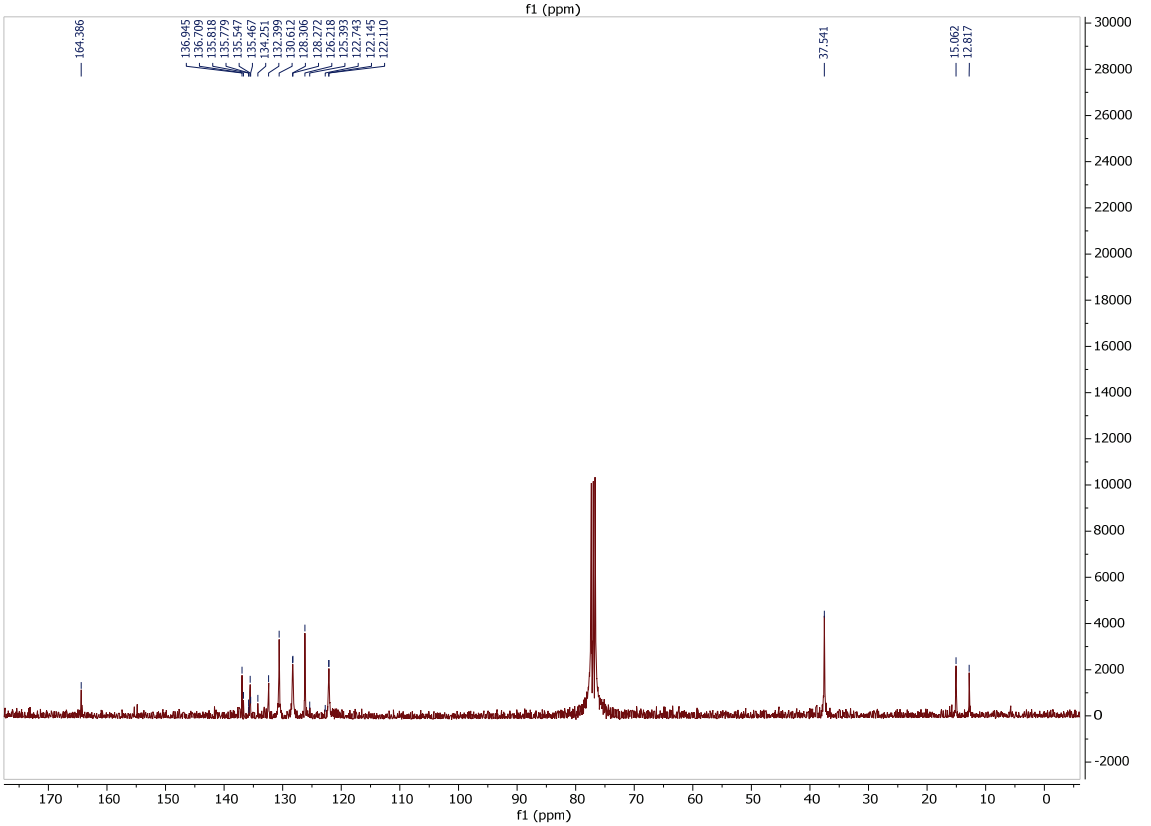

7.79
7.69
7.67
7.46
7.44
6.59
6.30
4.47
4.45
2.39
1.00
1.00
0.95
0.92
0.95
2.15
6.08
f1 (ppm)

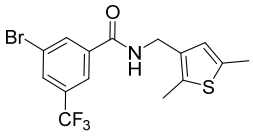

164.40
164.38
163.71
161.22
137.75
137.68
136.69
134.23
133.20
133.12
132.86
132.78
132.41
126.21
124.17
121.47
119.52
119.49
119.45
119.41
118.08
118.05
117.82
115.90
115.86
115.82
115.79
115.65
115.62
115.58
115.54
37.54
15.05
12.80
f1 (ppm)


O
Cl
N
H
S
CF3

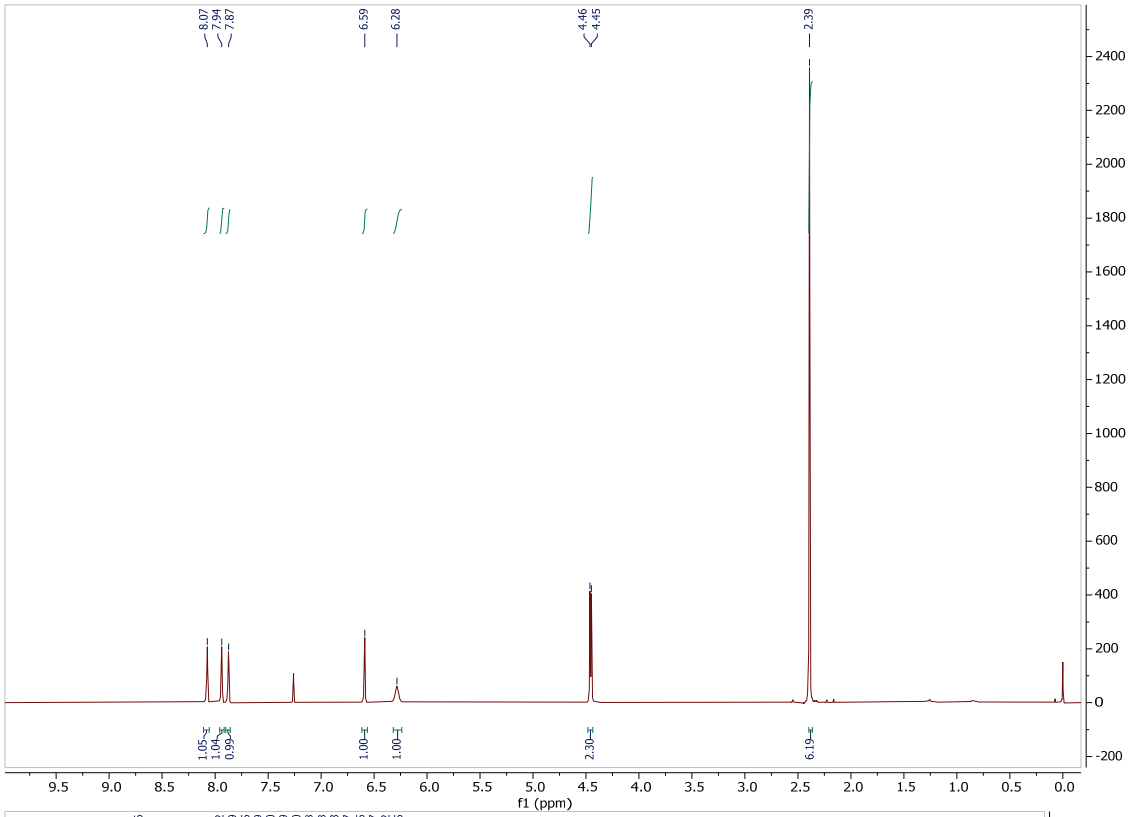
23

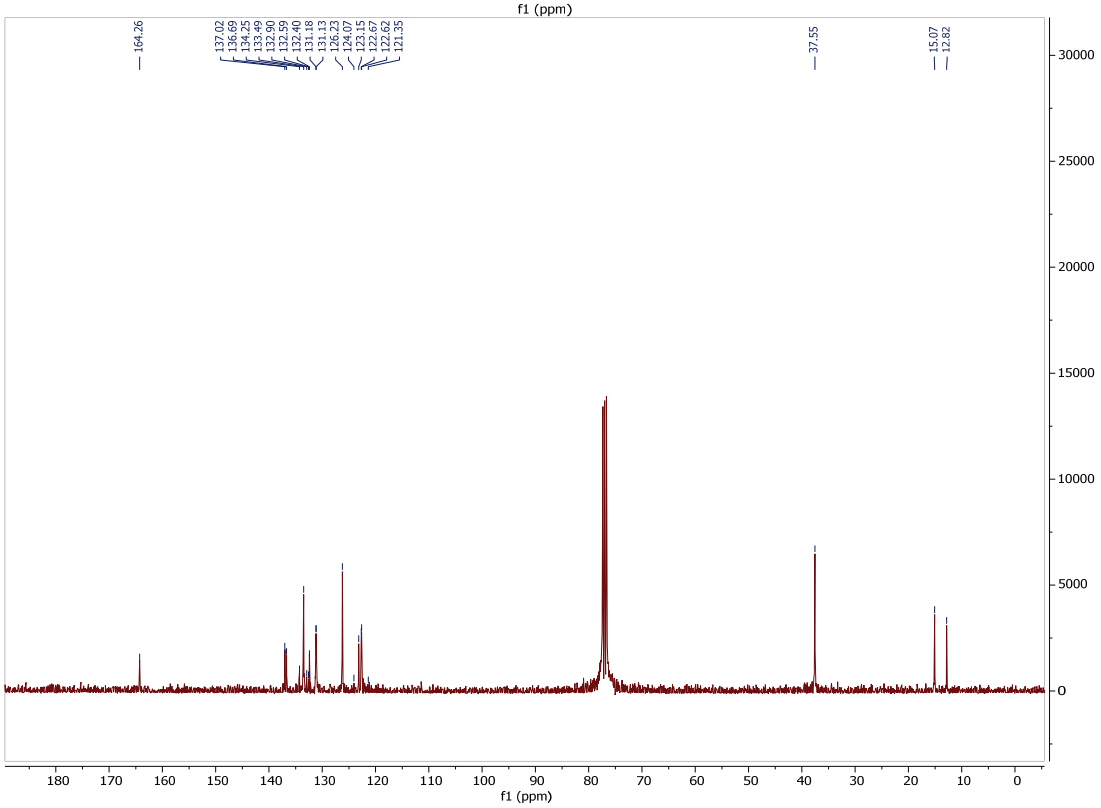

7.92
7.89
7.72
6.59
6.30
4.46
4.45
2.39
1.00
0.97
1.00
0.99
1.09
2.51
6.39
f1 (ppm)

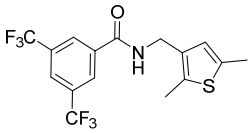

164.386
136.945
136.709
135.618
135.779
135.547
135.467
134.251
132.399
130.612
128.306
128.272
126.218
125.393
122.743
122.145
122.110
37.541
15.062
12.817
f1 (ppm)


Br
O
N
H
S
CF3

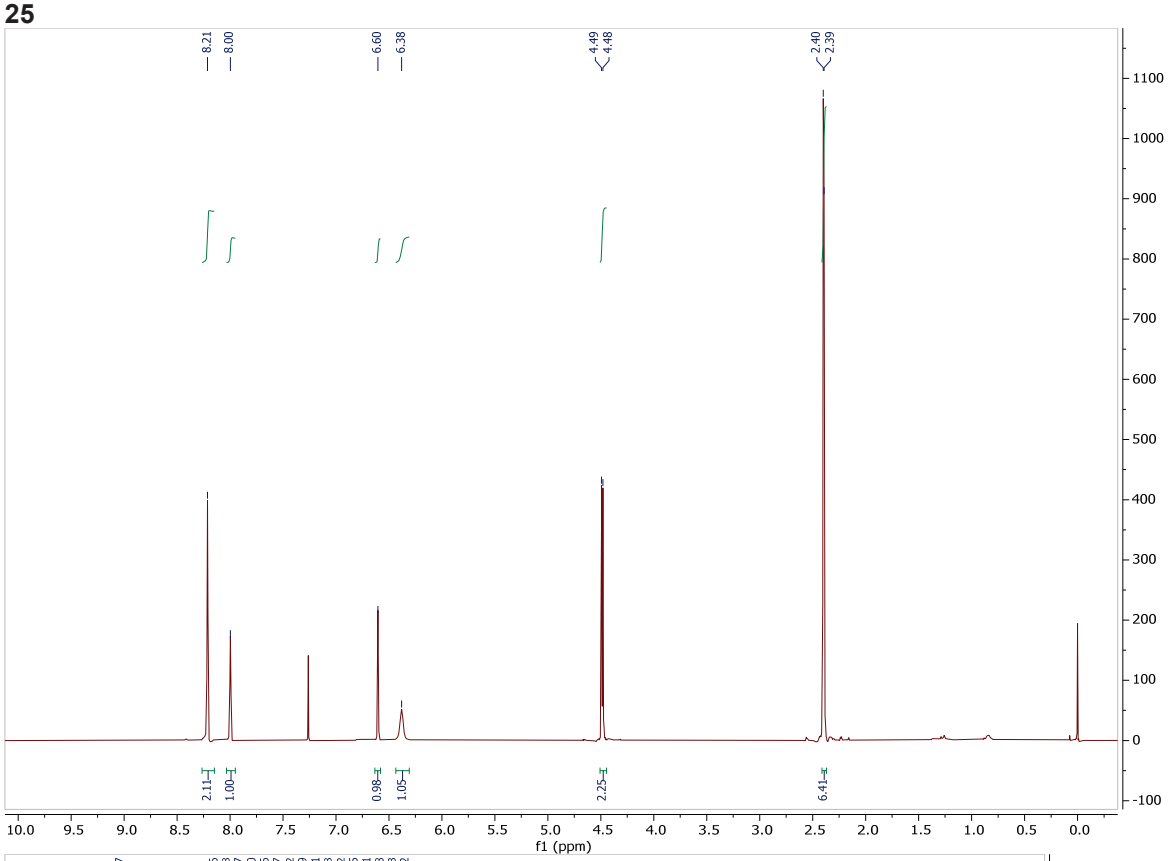
24

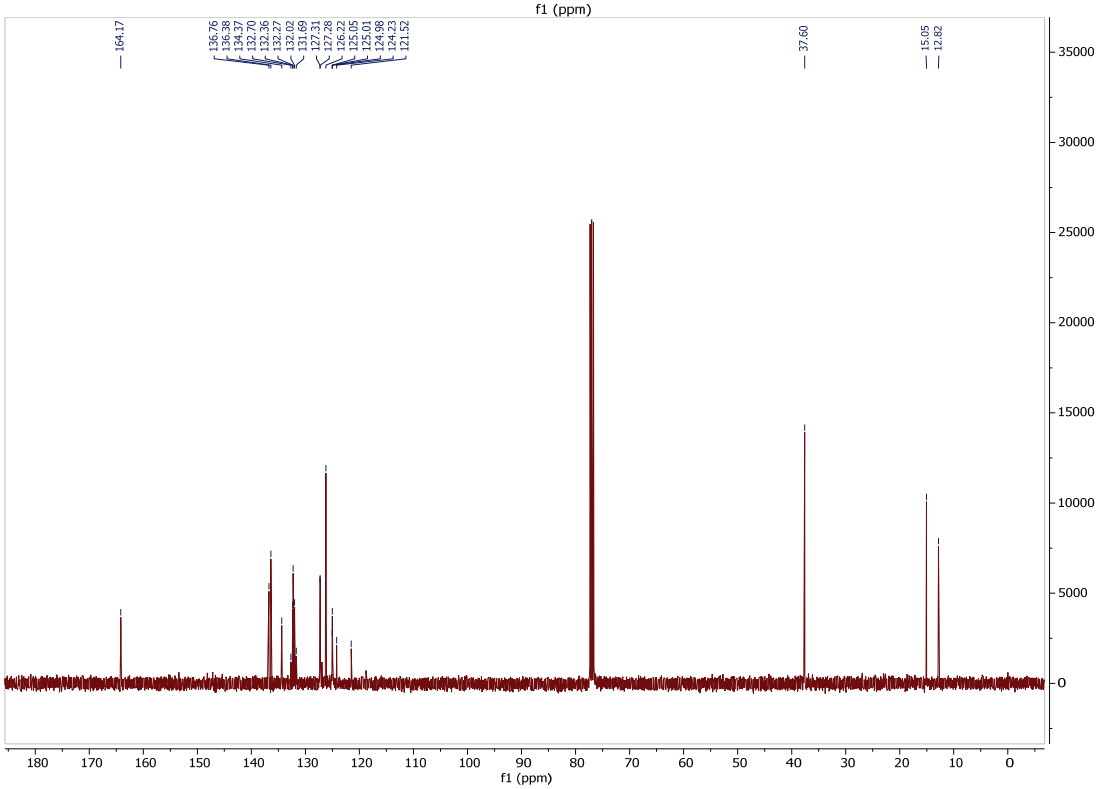

8.07
7.94
7.87
6.59
6.28
4.46
4.45
2.39
1.05
1.04
0.99
1.00
1.00
2.30
6.19
f1 (ppm)

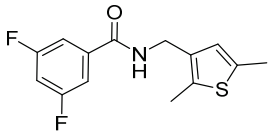

164.26
137.02
136.69
134.25
133.49
132.90
132.59
132.40
131.18
131.13
126.23
124.07
123.15
122.67
122.62
121.35
37.55
15.07
12.82
f1 (ppm)


O
F3C
N
H
S
CF3

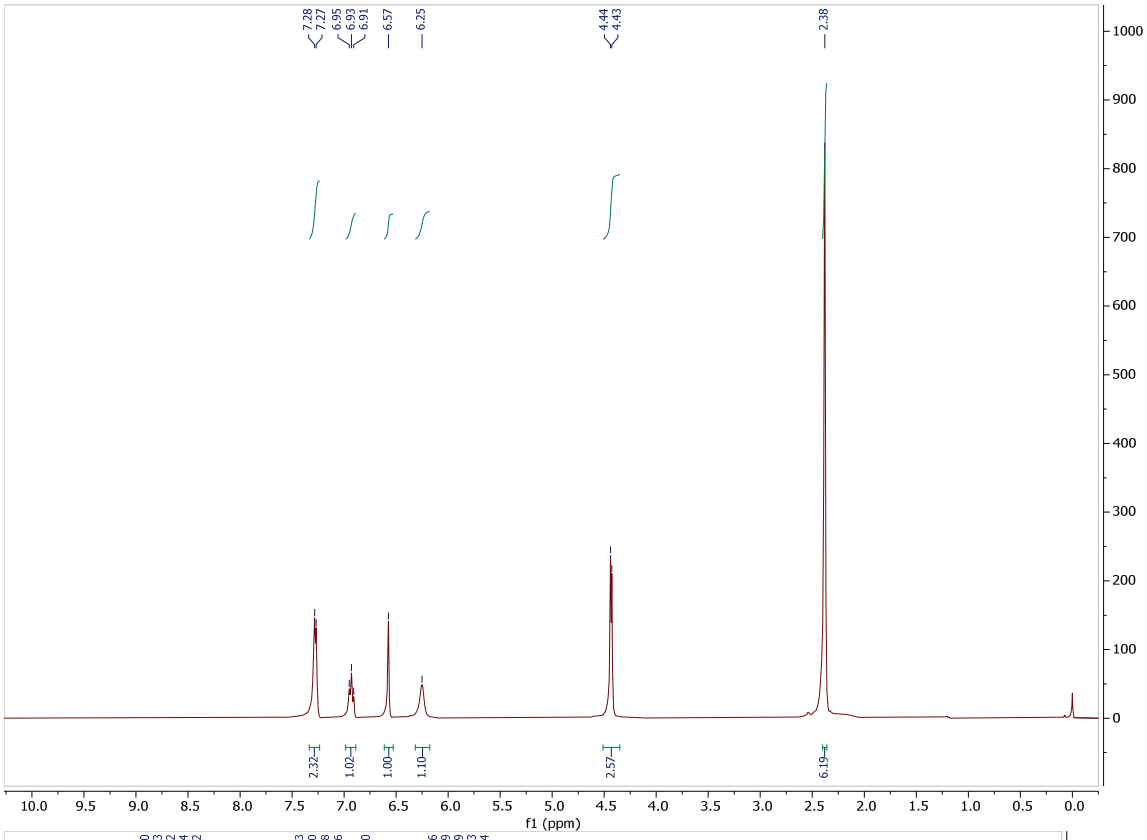
25

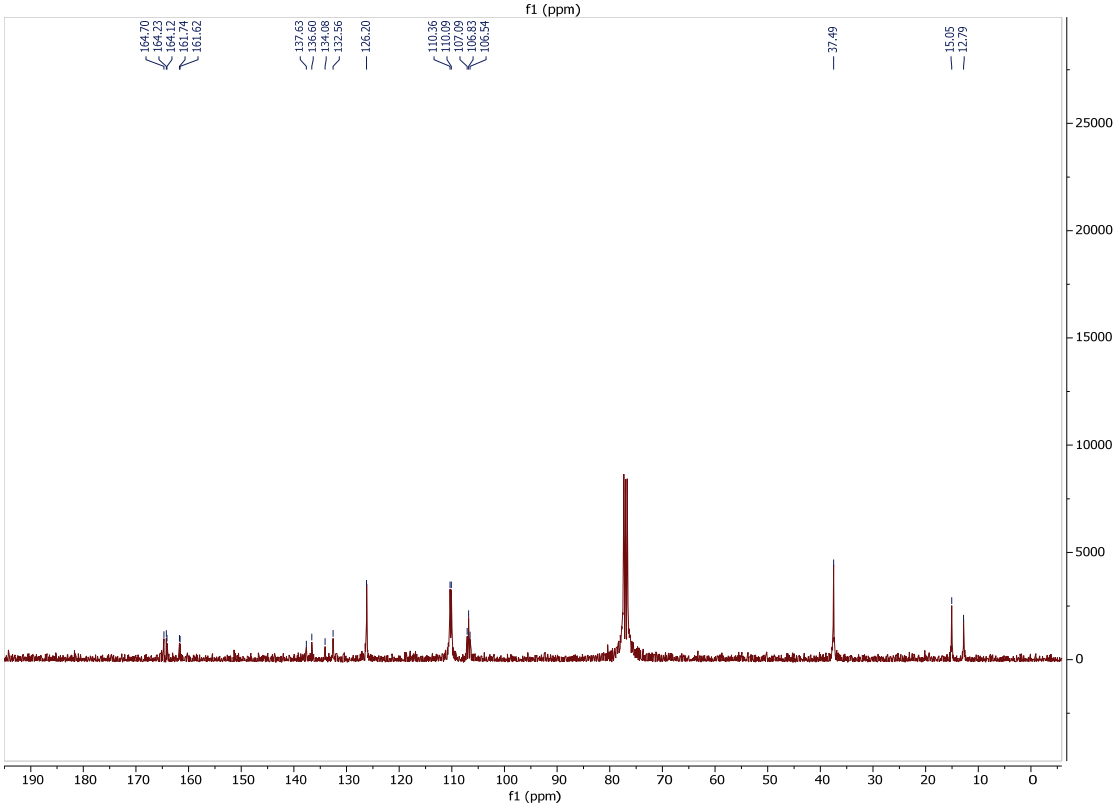

8.21
8.00
6.60
6.38
4.49
4.48
2.40
2.39
2.11
1.00
0.98
1.05
2.25
6.41
f1 (ppm)

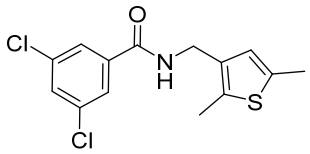

164.17
136.76
136.38
134.37
132.70
132.36
132.27
132.02
131.69
127.31
127.28
126.22
125.05
125.01
124.98
124.23
121.52
37.60
15.05
12.82
f1 (ppm)


O
F
N
H
S
F

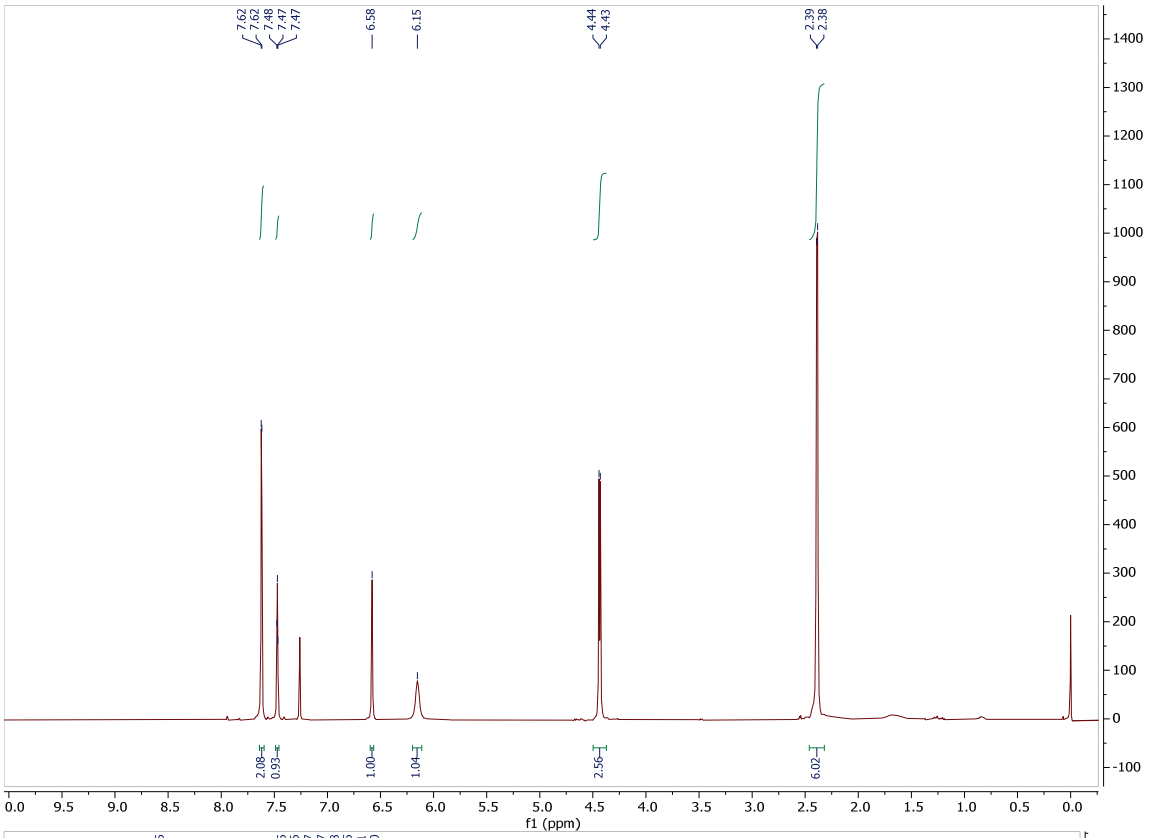
26

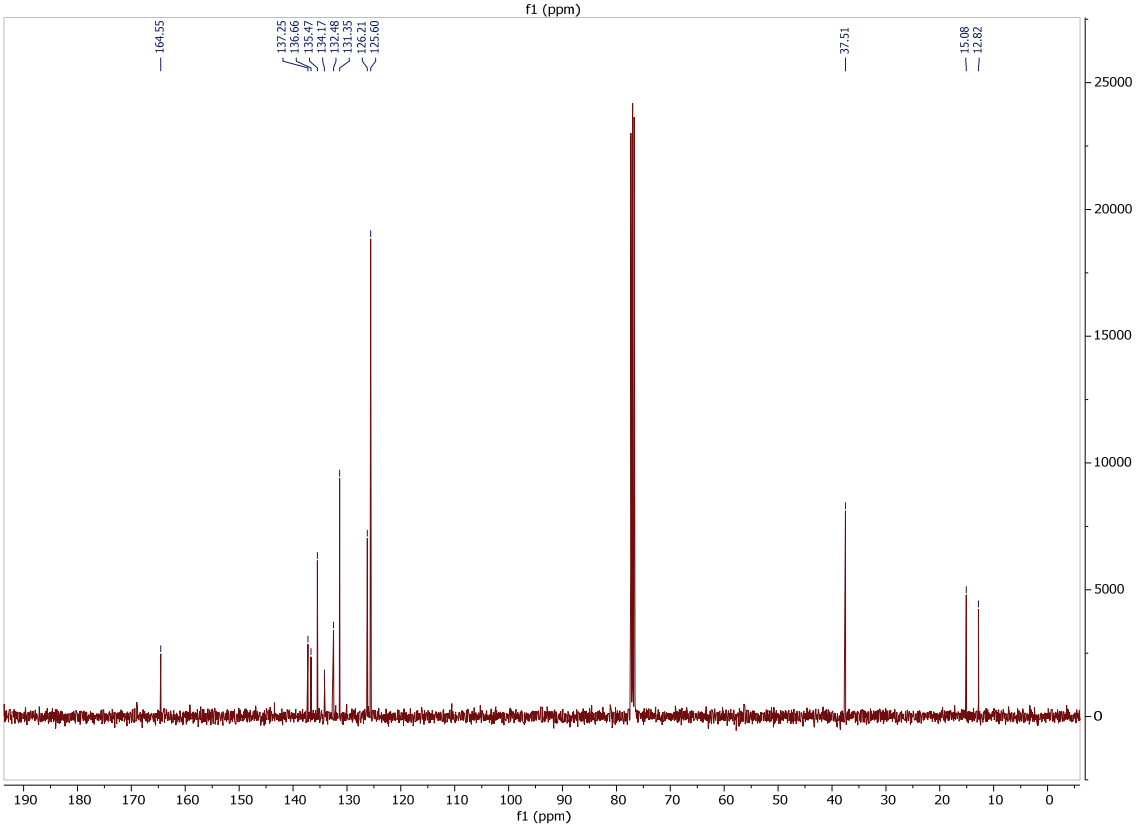

7.28
7.27
6.95
6.93
6.91
6.57
6.25
4.44
4.43
2.38
2.32
1.02
1.00
1.10
2.57
6.19
f1 (ppm)

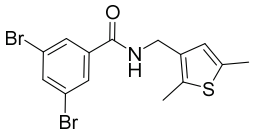

164.70
164.23
164.12
161.74
161.62
137.63
136.60
134.08
132.56
126.20
110.36
110.09
107.09
106.83
106.54
37.49
15.05
12.79
f1 (ppm)


O
Cl
N
H
S
Cl

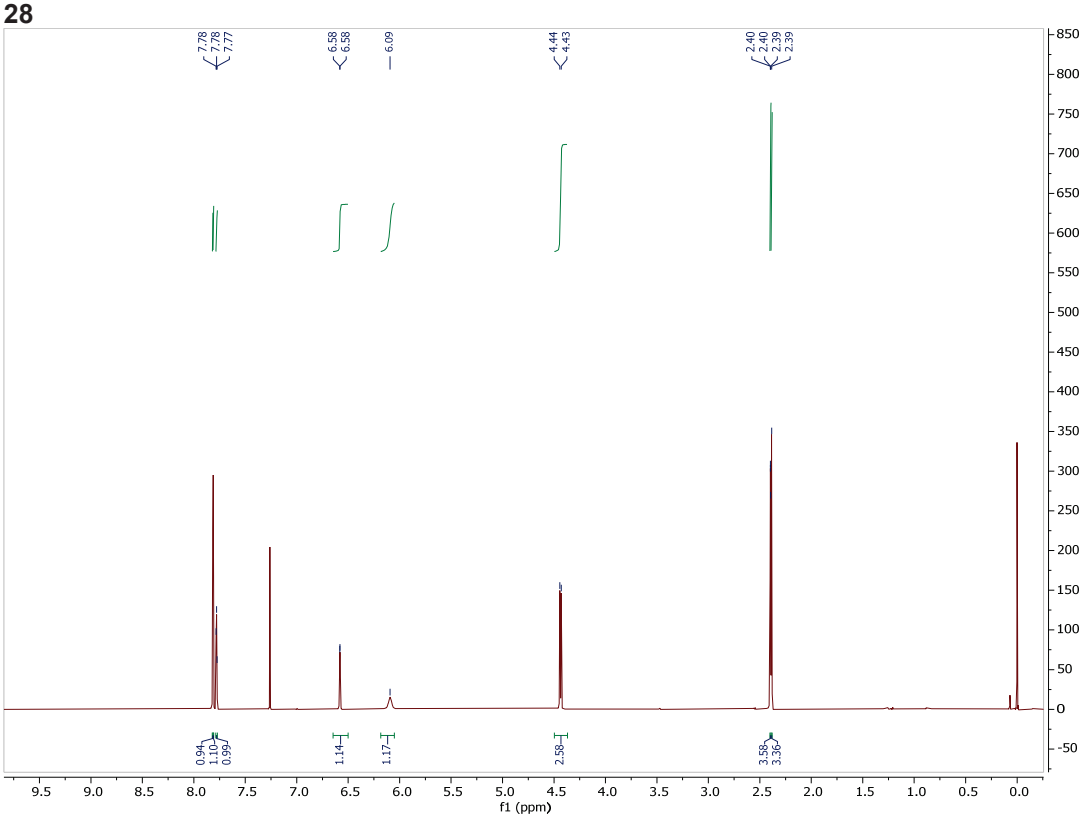
27

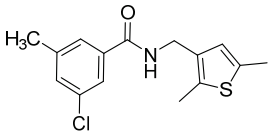

7.62
7.62
7.48
7.47
7.47
6.58
6.15
4.44
4.43
2.39
2.38
2.08
0.93
1.00
1.04
2.56
6.02
f1 (ppm)


164.55
137.25
136.66
135.47
134.17
132.48
131.35
126.21
125.60
37.51
15.08
12.82
f1 (ppm)

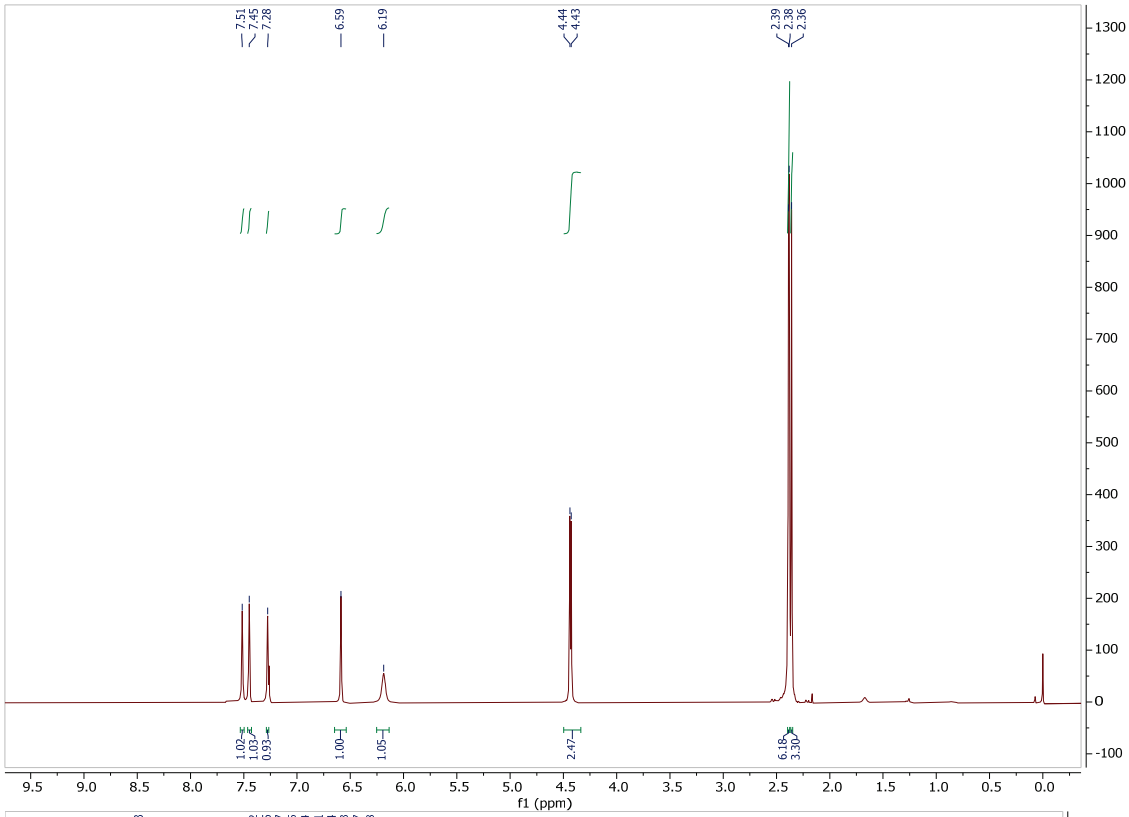

Br
O
N
H
S
Br

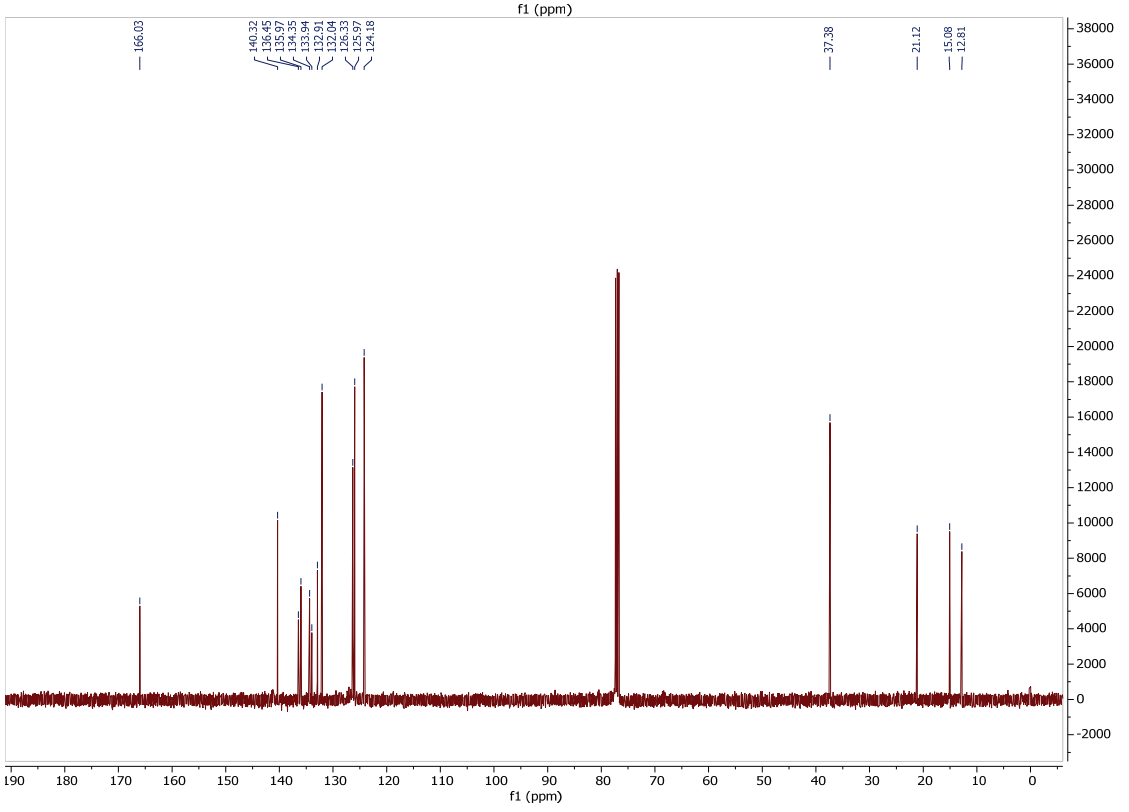
28

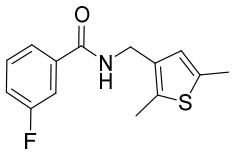

7.78
7.78
7.77
6.58
6.58
6.09
4.44
4.43
2.40
2.40
2.39
2.39
0.94
1.10
0.99
1.14
1.17
2.58
3.58
3.36
f1 (ppm)


O
H3C
N
H
S
Cl

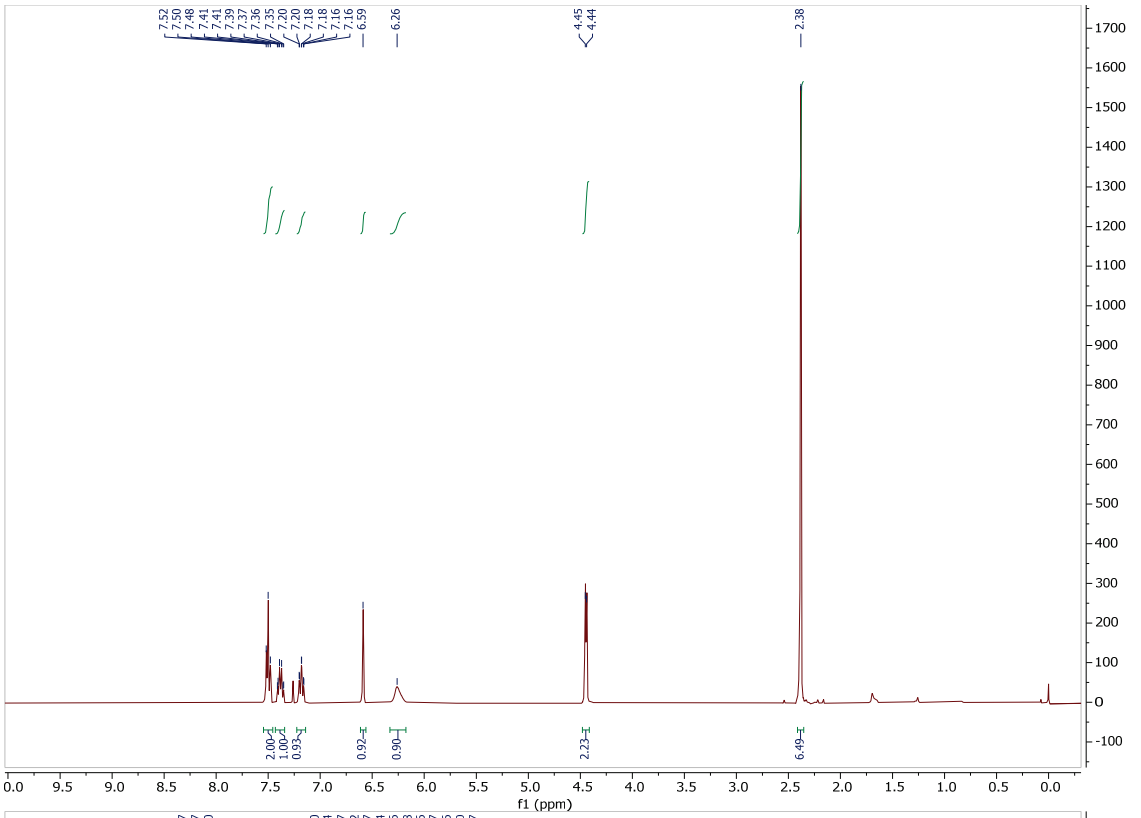
29

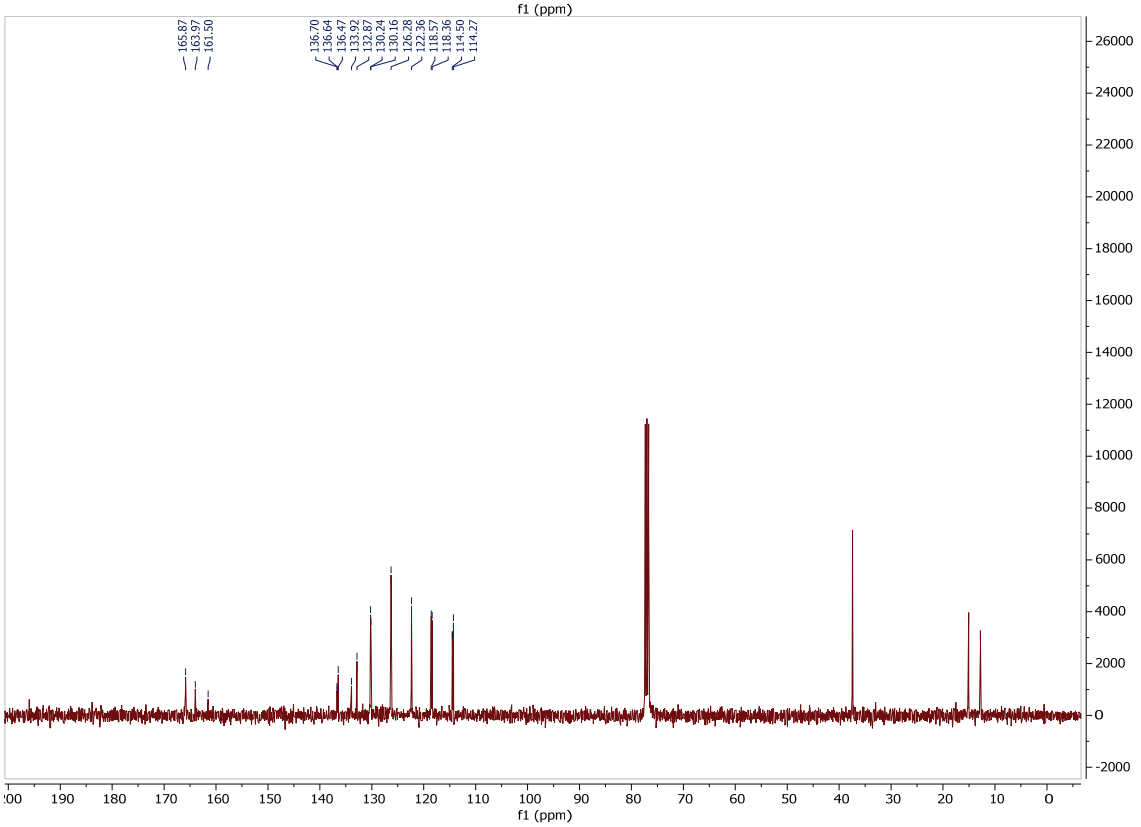

7.51
7.45
7.28
6.59
6.19
4.44
4.43
2.39
2.38
2.36
1.02
1.02
0.93
1.00
1.05
2.47
6.18
3.30
f1 (ppm)

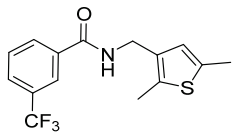

166.03
140.32
136.45
135.97
134.35
133.94
132.91
132.04
126.33
125.97
124.18
37.38
21.12
15.08
12.81
f1 (ppm)


O
N
H
S
F

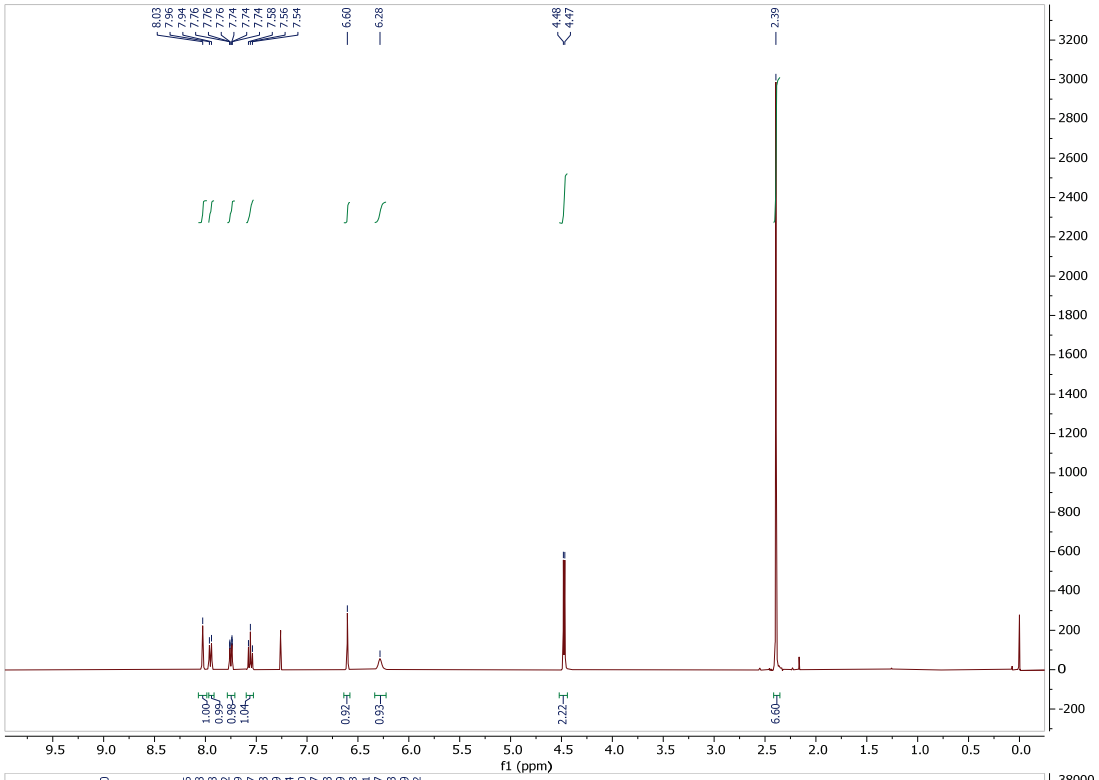
30

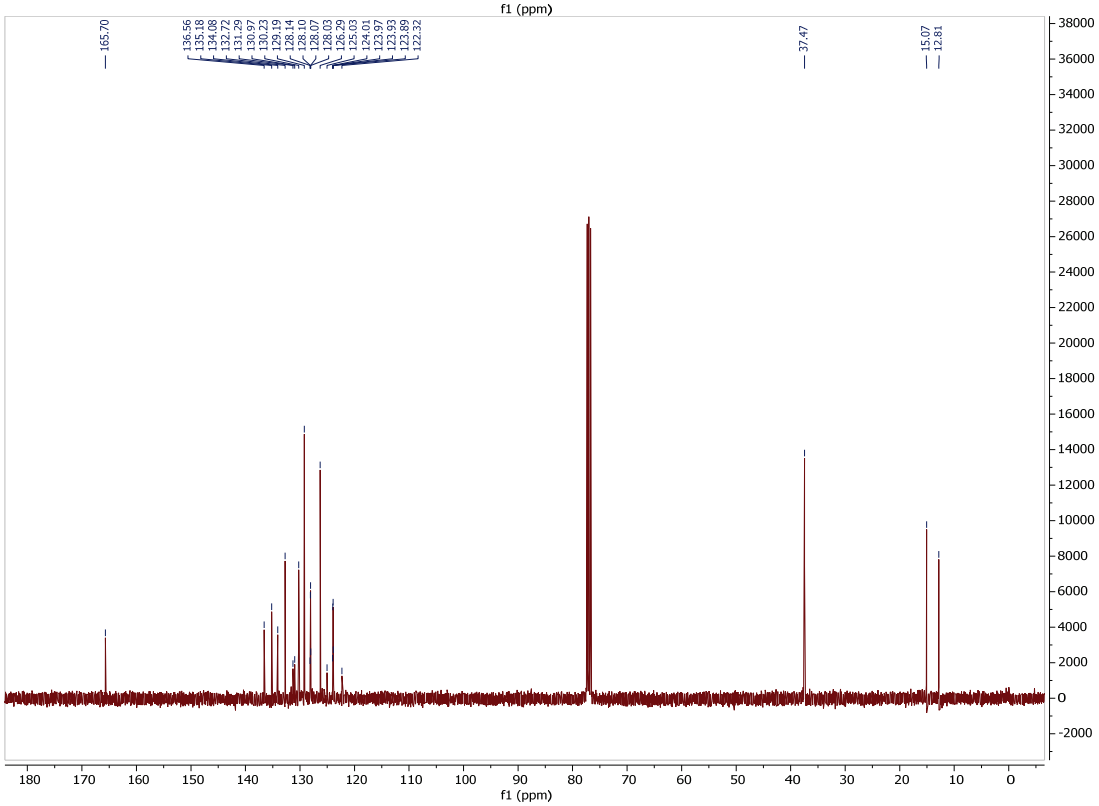

7.52
7.50
7.48
7.41
7.41
7.39
7.37
7.36
7.35
7.20
7.20
7.18
7.16
7.16
6.59
6.26
4.45
4.44
2.38
2.00
1.00
0.93
0.92
0.90
2.23
6.49
f1 (ppm)

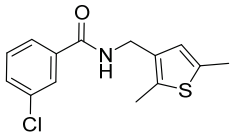

165.87
163.97
161.50
136.70
136.64
136.47
133.92
132.87
130.24
130.16
126.28
122.36
118.57
118.36
114.50
114.27
f1 (ppm)


O
N
H
S
CF3

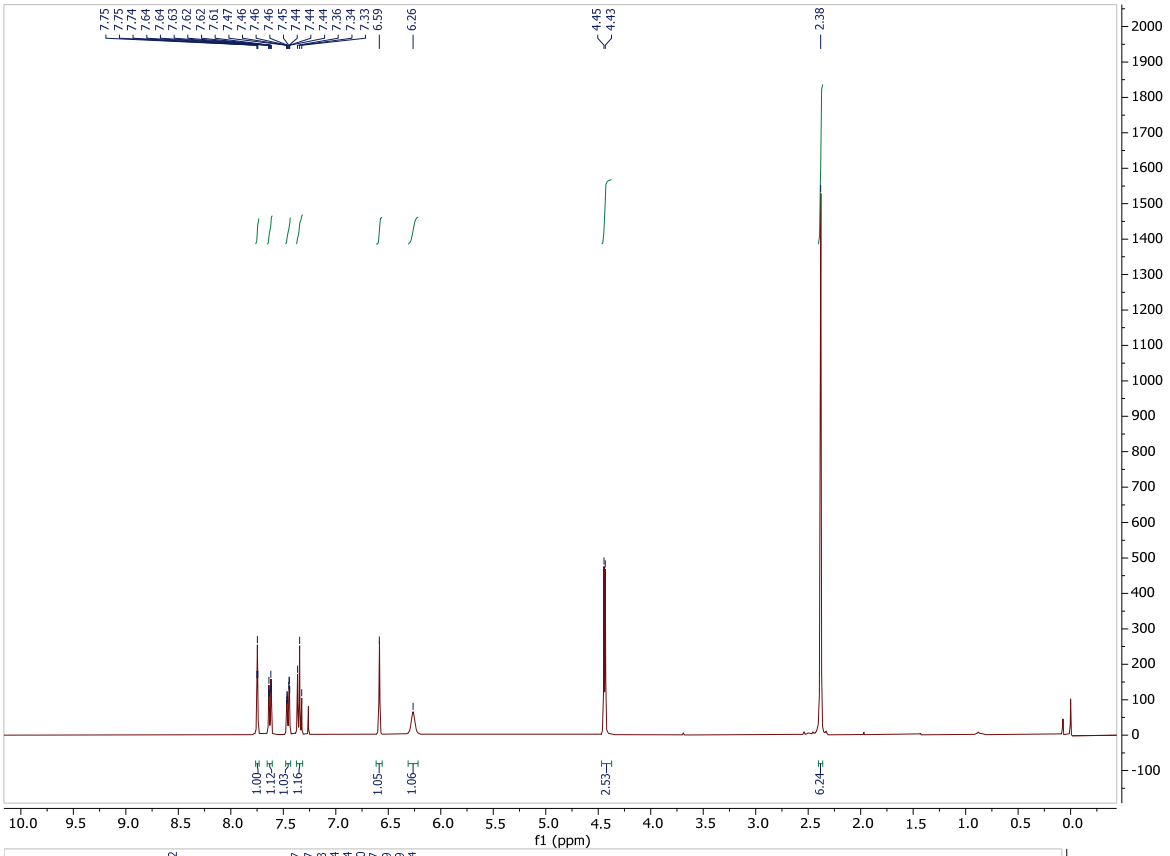
31

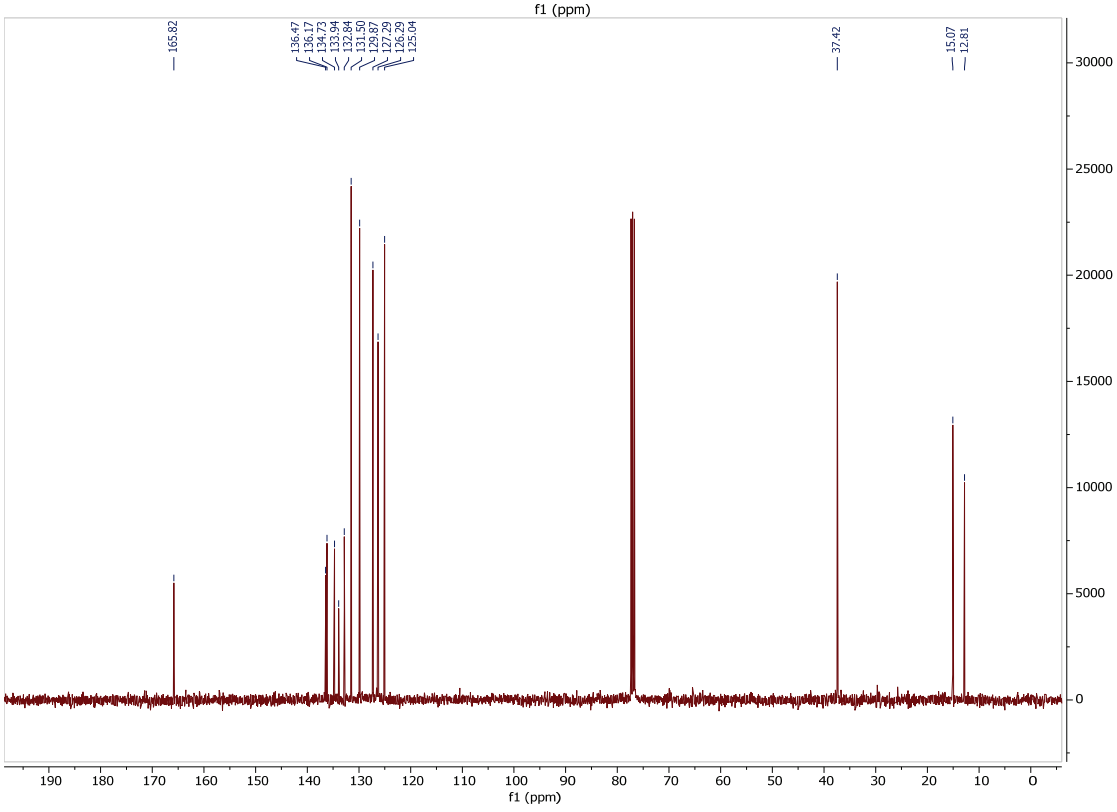

8.03
7.95
7.94
7.76
7.76
7.76
7.74
7.74
7.74
7.58
7.56
7.54
6.60
6.28
4.48
4.47
2.39
1.00
0.99
0.98
1.04
0.92
0.93
2.22
6.60
f1 (ppm)

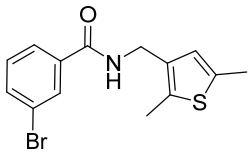

165.70
136.56
135.18
134.08
132.72
131.26
130.97
130.23
129.19
128.14
128.10
128.07
128.03
126.29
125.03
124.01
123.97
123.93
123.89
122.32
37.47
15.07
12.81
f1 (ppm)


O
N
H
S
Cl

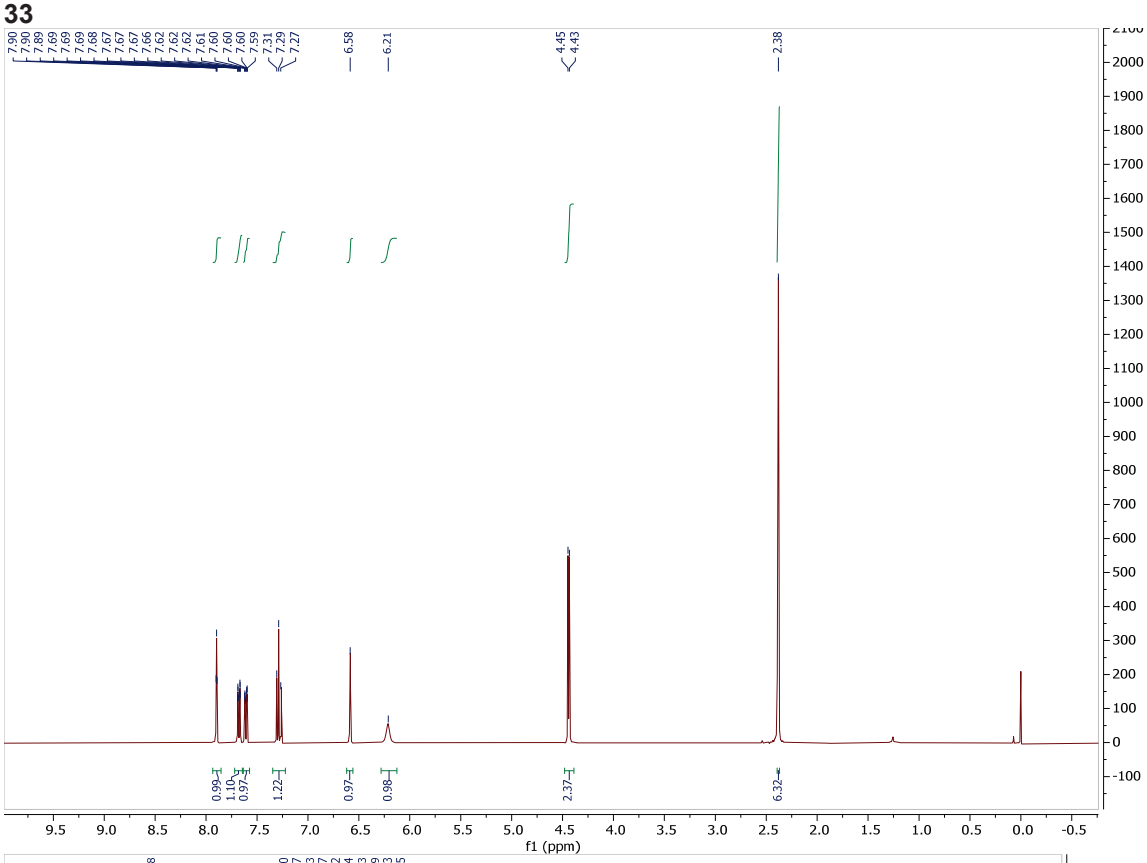
32

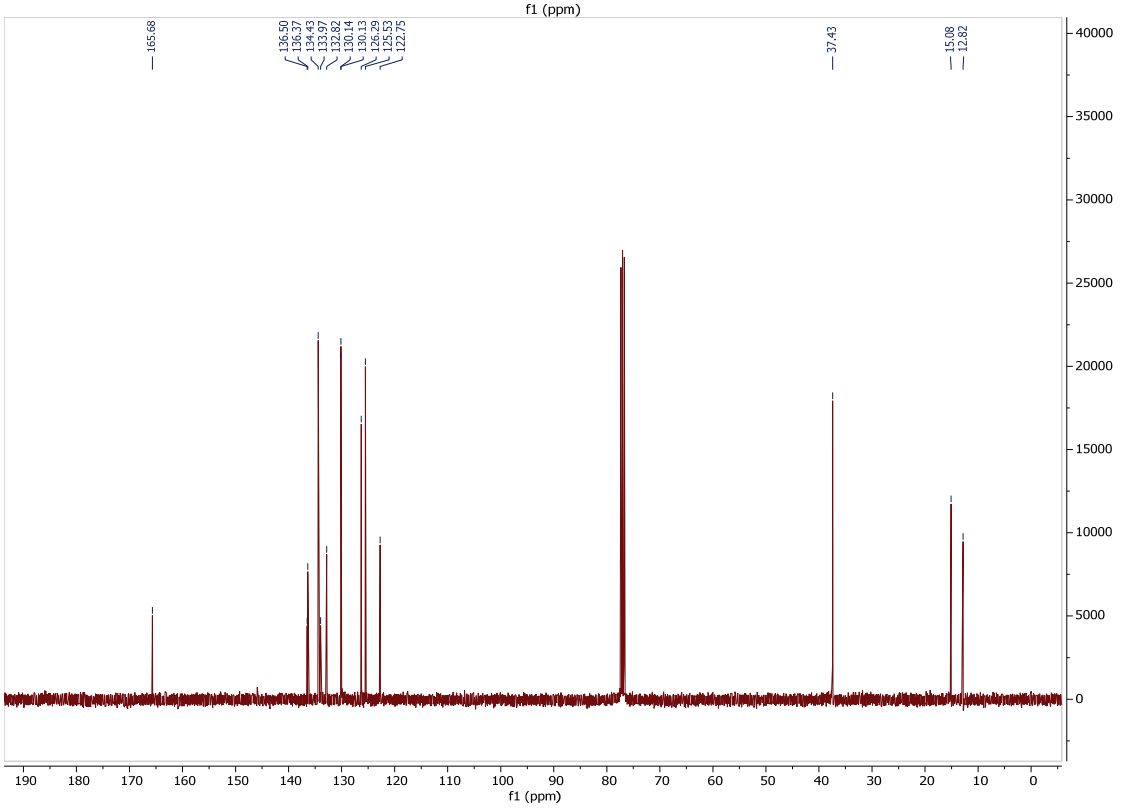

7.75
7.75
7.74
7.64
7.64
7.63
7.62
7.62
7.61
7.47
7.46
7.46
7.46
7.45
7.44
7.44
7.44
7.36
7.34
7.33
6.59
6.26
4.45
4.43
2.38
1.00
1.12
1.03
1.16
1.05
1.06
2.53
6.24
f1 (ppm)

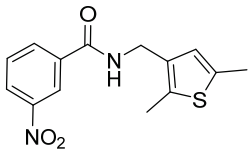

165.82
136.47
136.17
134.73
133.94
132.84
131.50
129.87
127.29
126.29
125.04
37.42
15.07
12.81
f1 (ppm)


O
N
H
S
Br

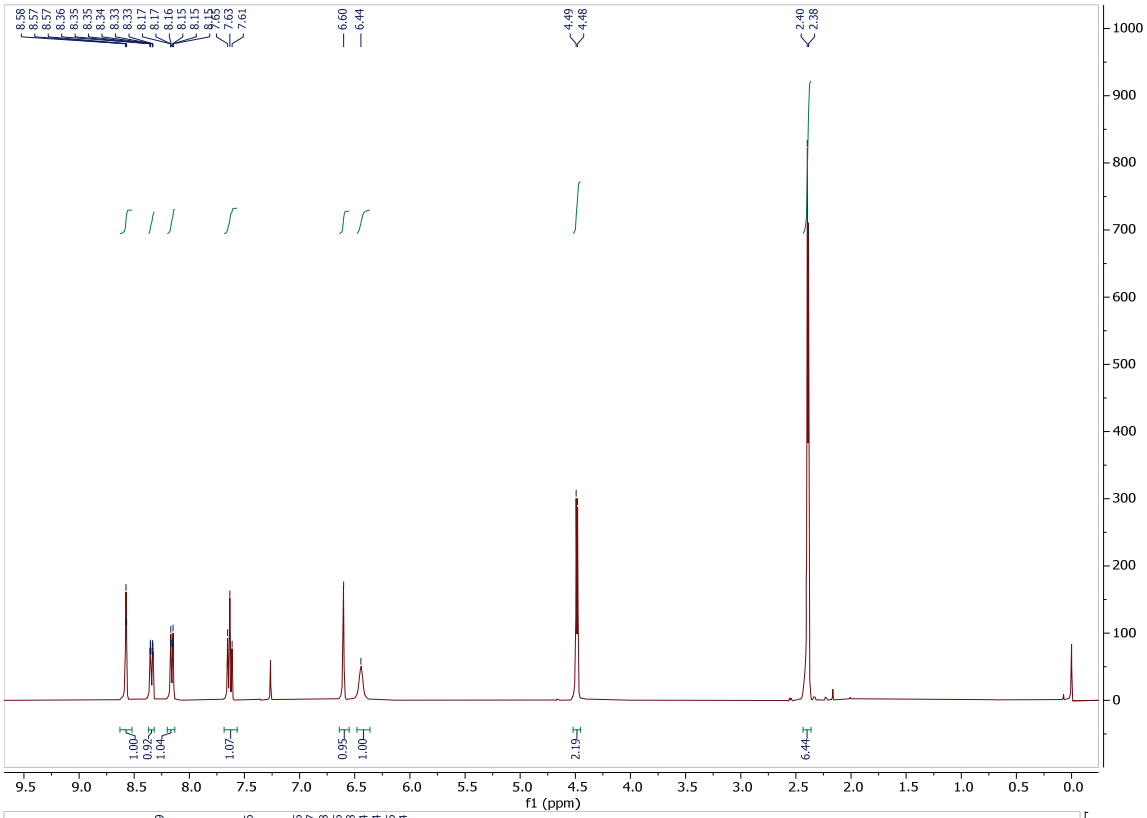
33

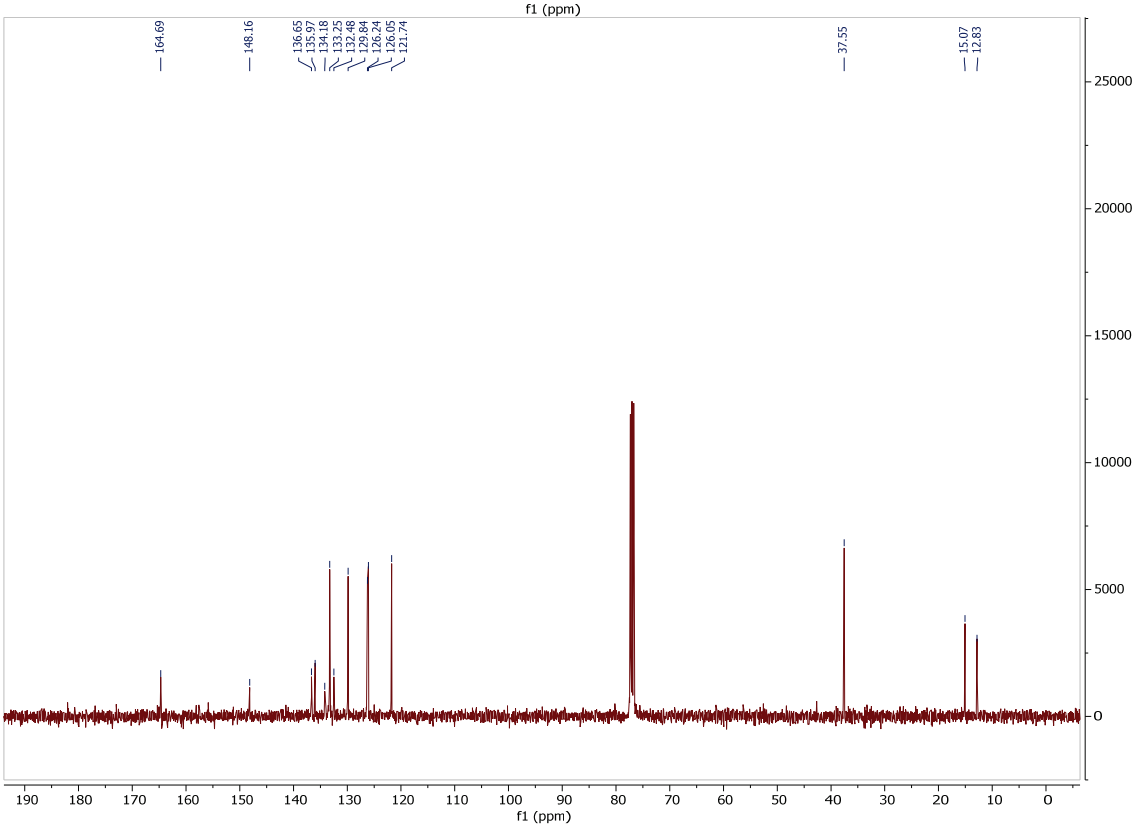

7.90
7.90
7.89
7.69
7.69
7.69
7.68
7.67
7.67
7.67
7.66
7.62
7.62
7.62
7.61
7.60
7.60
7.60
7.59
7.31
7.29
7.27
6.58
6.21
4.45
4.43
2.38
0.99
1.10
0.97
1.22
0.97
0.98
2.37
6.32
f1 (ppm)

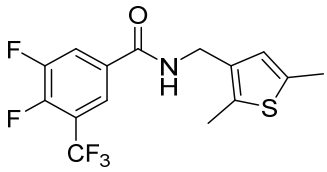

165.68
136.50
136.37
134.43
133.07
132.82
130.14
130.13
126.29
125.53
122.75
37.43
15.08
12.82
f1 (ppm)


O
N
H
S
NO2

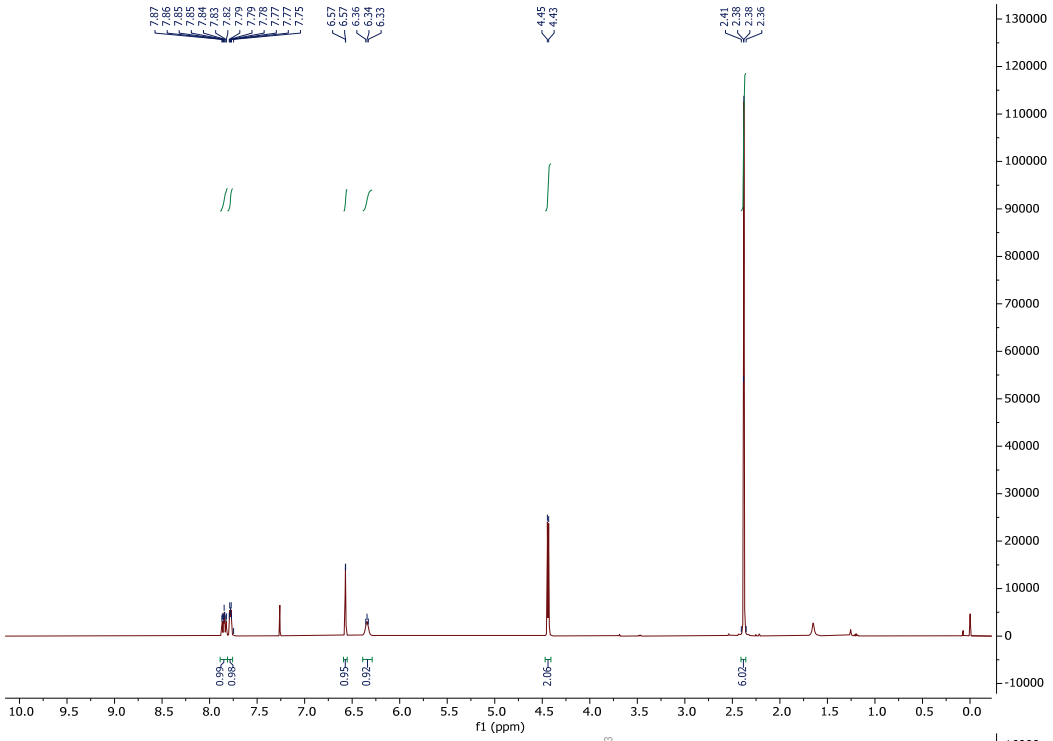
34

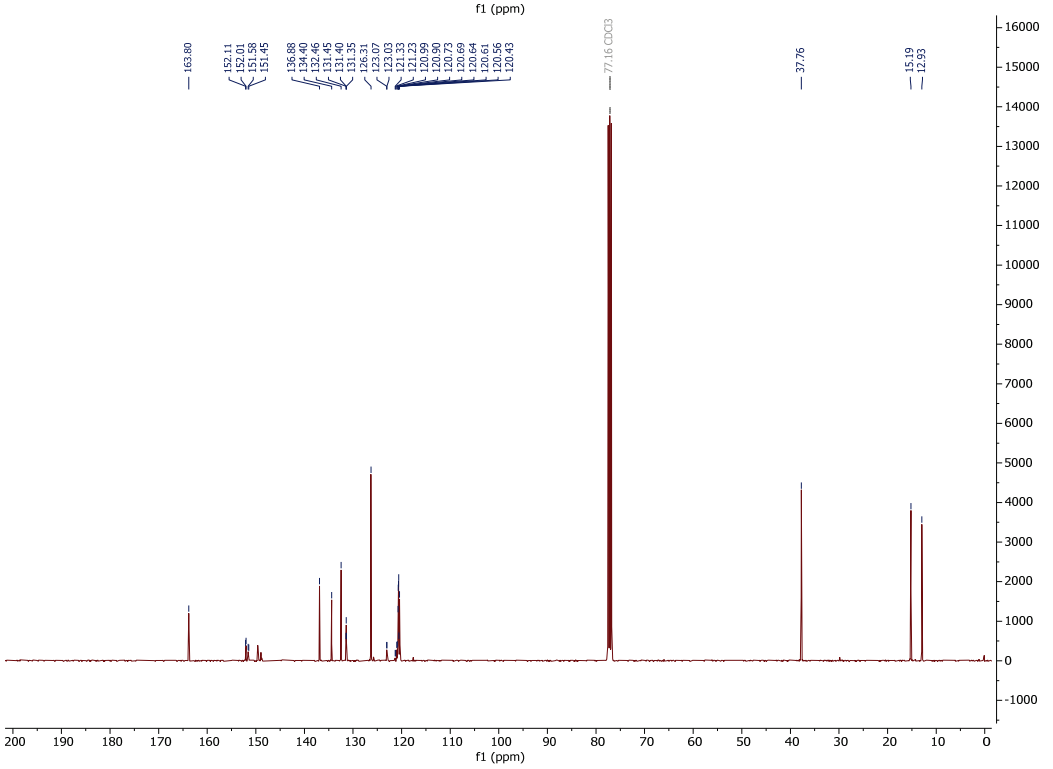

8.58
8.57
8.57
8.36
8.35
8.35
8.34
8.33
8.33
8.17
8.17
8.16
8.15
8.15
8.15
7.65
7.63
7.61
6.60
6.44
4.49
4.48
2.40
2.38
1.00
0.92
1.04
1.07
0.95
1.00
2.19
6.44
f1 (ppm)

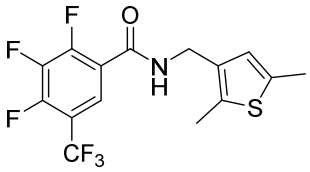

164.69
148.16
136.65
135.97
134.18
133.25
128.48
126.84
126.24
126.05
121.74
37.55
15.07
12.83
f1 (ppm)


O
F
N
H
F
S
CF3

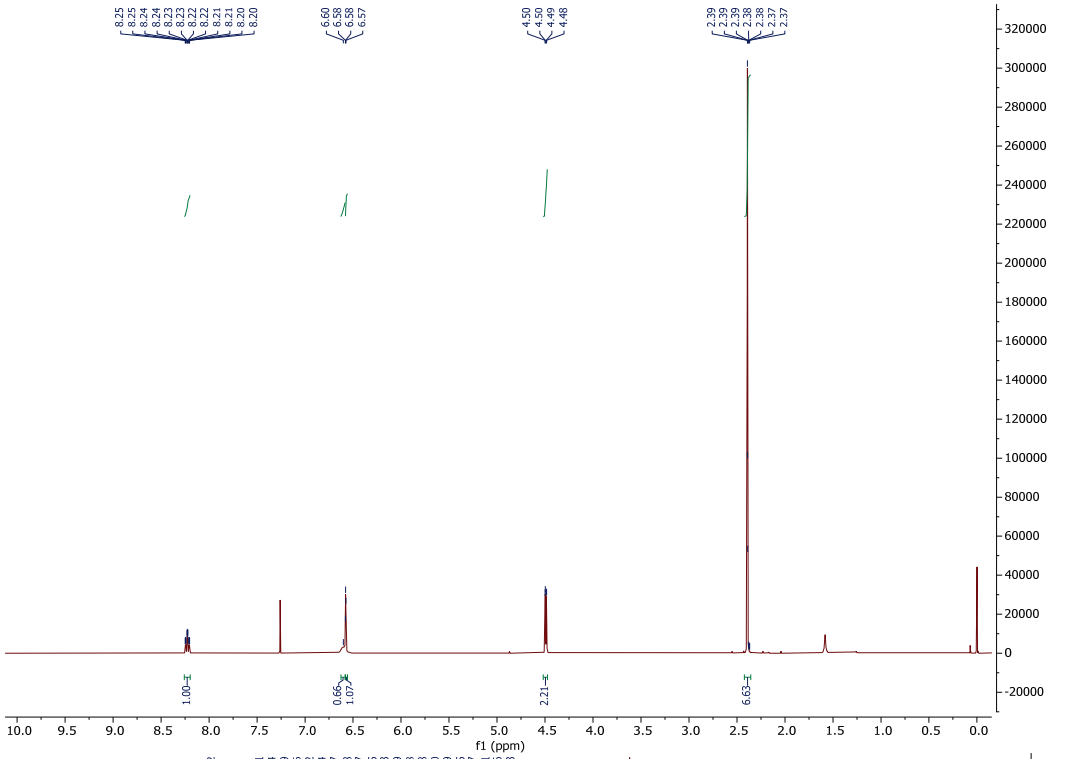
35

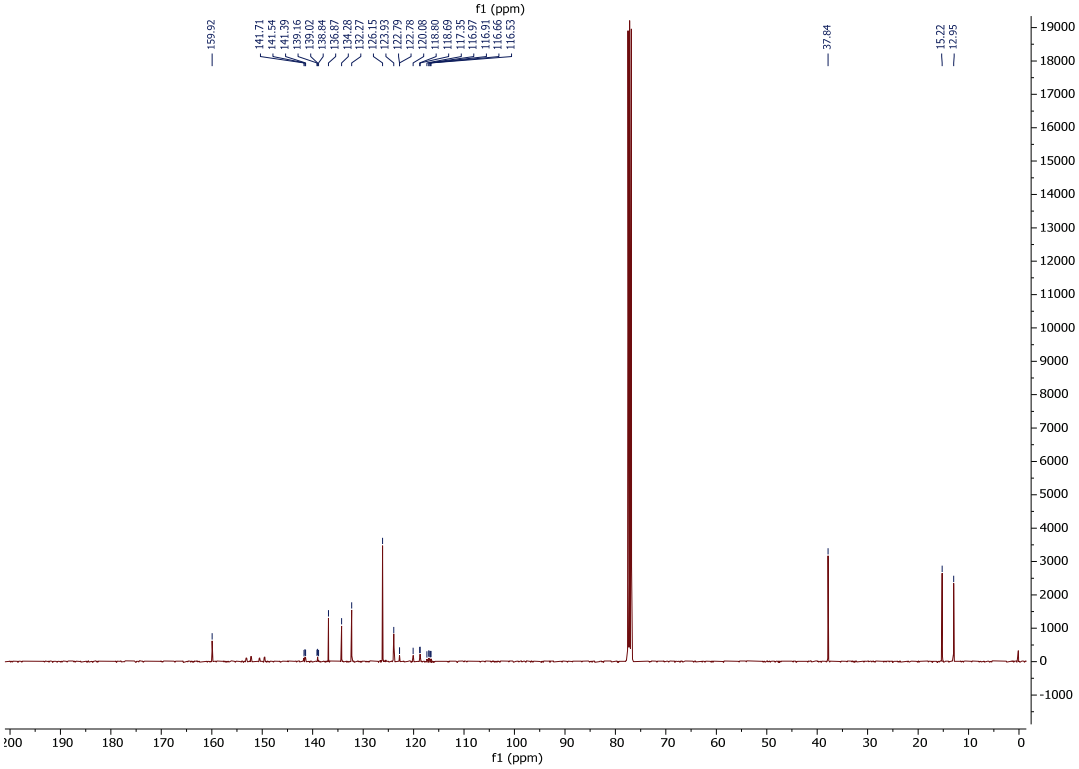

7.87
7.86
7.85
7.85
7.84
7.83
7.82
7.79
7.79
7.78
7.77
7.77
7.75
6.57
6.57
6.36
6.34
6.33
4.45
4.43
2.41
2.38
2.38
2.36
0.99
0.98
0.95
0.92
2.06
6.02
f1 (ppm)

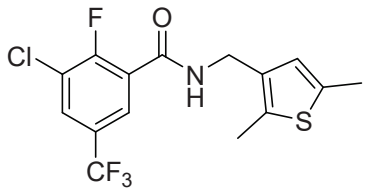

163.80
152.11
152.01
151.88
151.45
136.88
134.40
132.46
131.45
131.40
131.35
126.31
123.07
123.03
121.33
121.23
120.99
120.90
120.73
120.69
120.64
120.61
120.56
120.43
77.16 CDCl3
37.76
15.19
12.93
f1 (ppm)


F
O
F
N
H
F
S
CF3

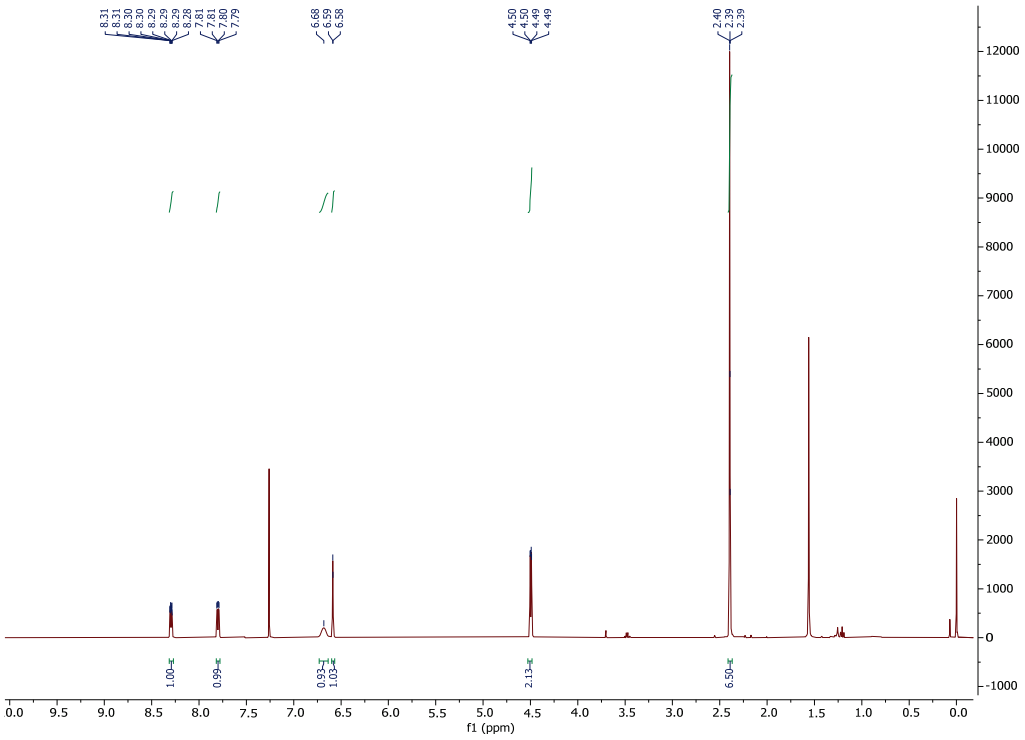
36

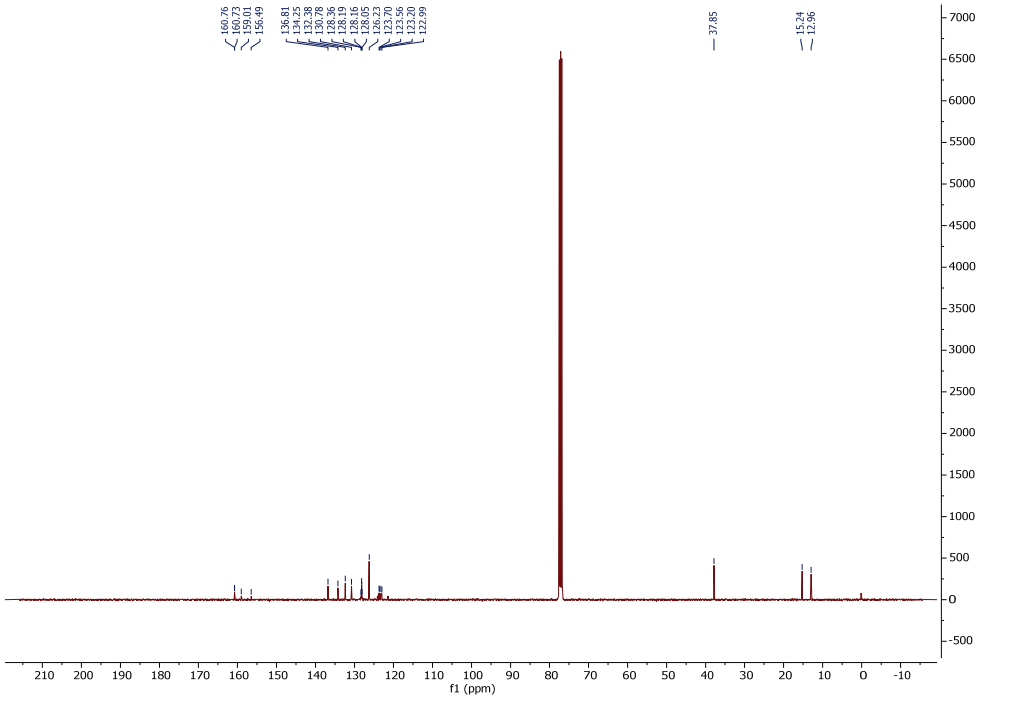

8.25
8.24
8.23
8.22
8.21
8.20
6.60
6.58
6.57
4.50
4.49
4.48
2.39
2.38
2.37
1.00
0.66
1.07
2.21
6.63
f1 (ppm)

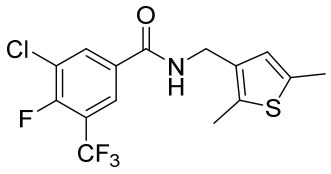

159.92
141.71
141.54
141.39
139.16
139.02
138.84
136.87
134.28
132.27
126.15
123.93
122.79
122.78
120.88
118.80
118.69
117.35
116.97
116.91
116.66
116.53
37.84
15.22
12.95
f1 (ppm)


F
O
Cl
N
H
S
CF3

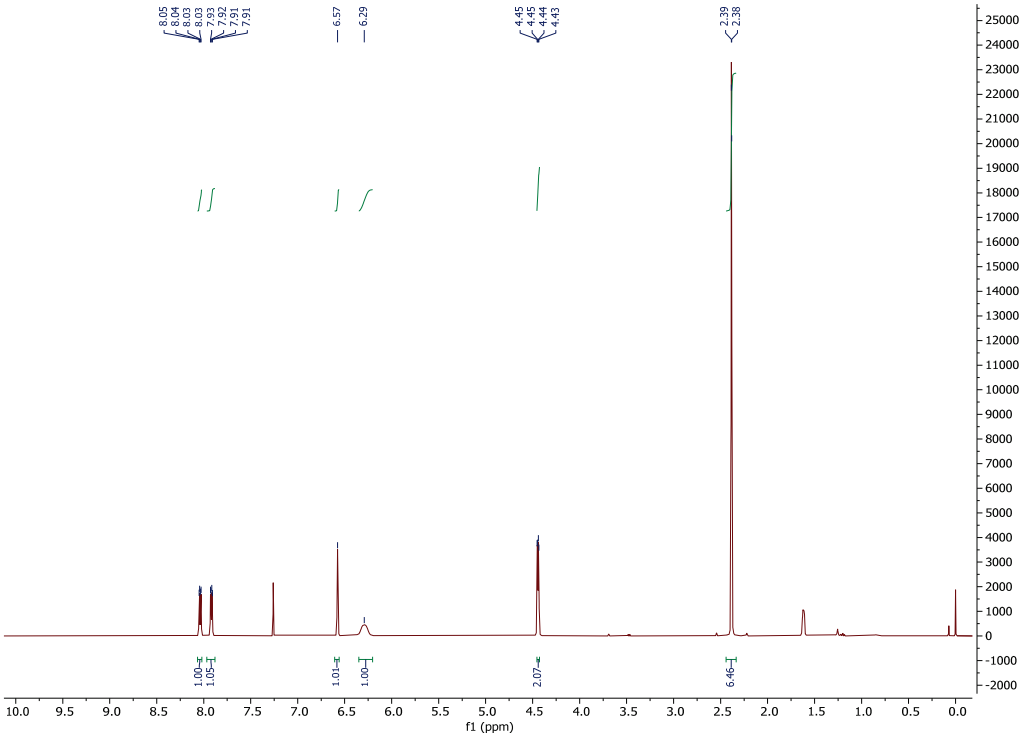
37

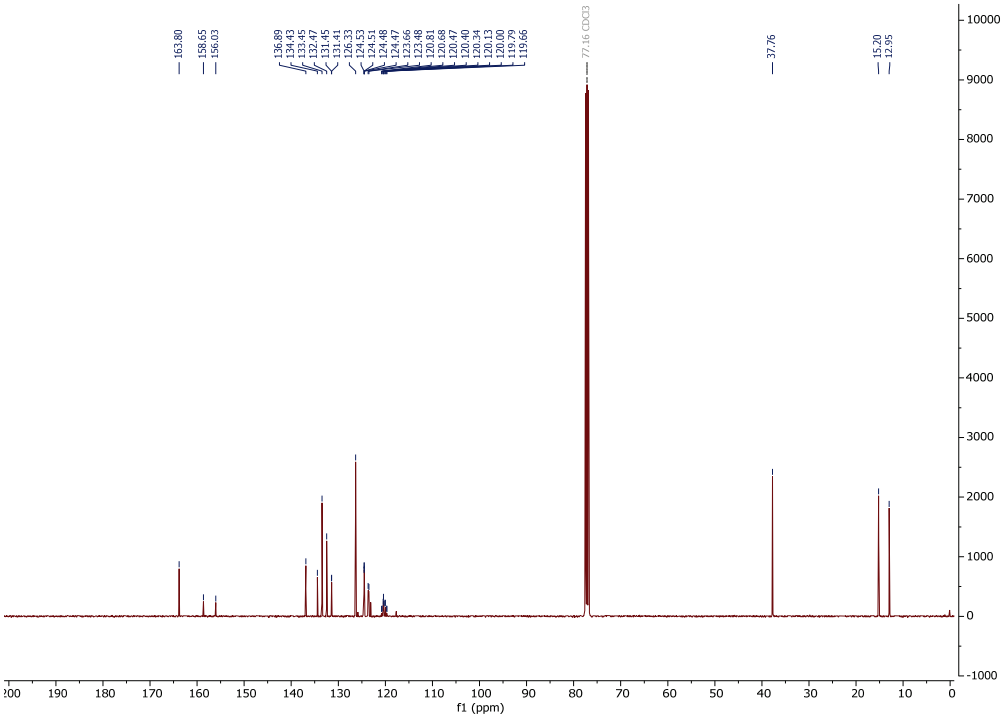

8.31
8.31
8.30
8.30
8.29
8.29
8.29
8.28
7.81
7.81
7.80
7.79
6.68
6.59
6.58
4.50
4.50
4.49
4.49
2.40
2.39
2.39
1.00
0.99
0.93
1.03
2.13
6.50
f1 (ppm)

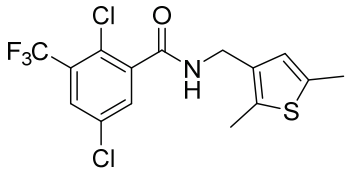

160.76
160.73
159.01
156.49
136.81
134.25
132.38
130.78
128.36
128.19
128.16
128.05
126.23
123.70
123.56
123.20
122.99
37.85
15.24
12.96
f1 (ppm)


O
Cl
N
H
F
S
CF3

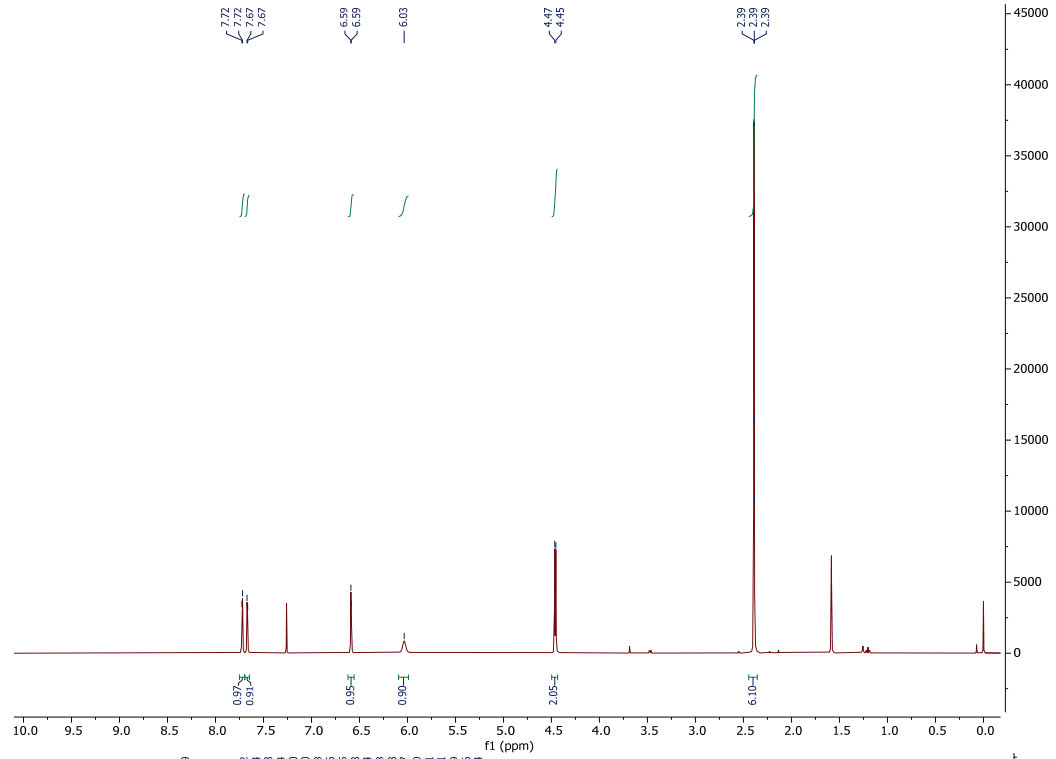
38

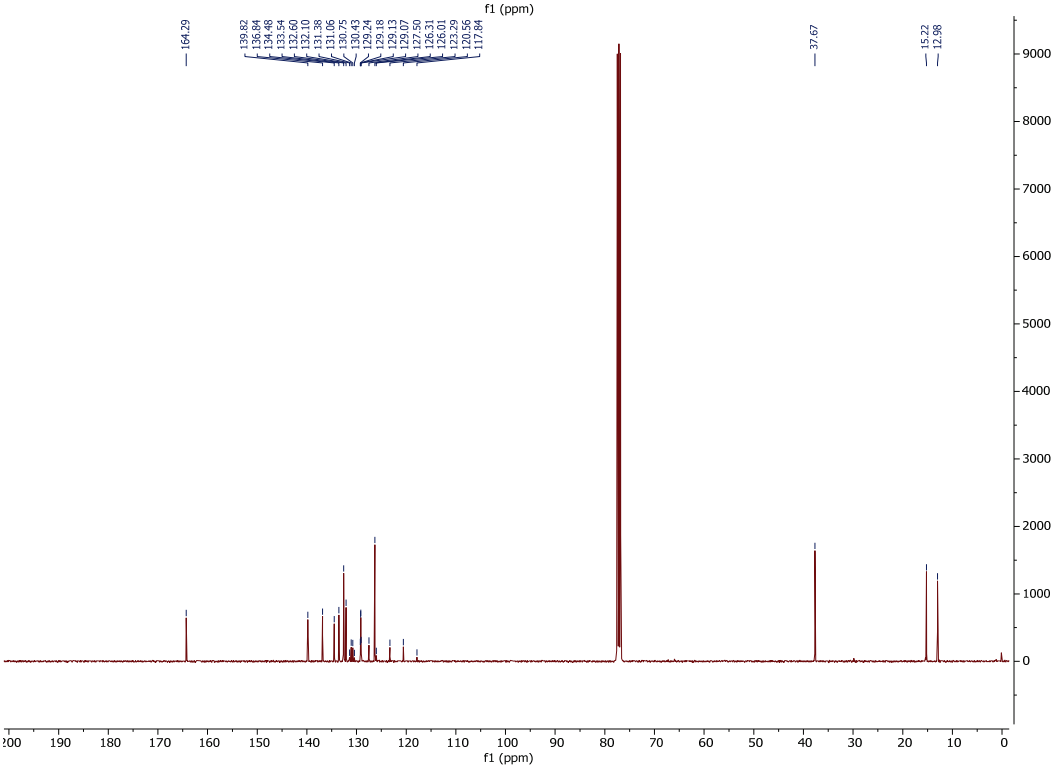

8.05
8.04
8.03
8.03
7.92
7.91
7.91
6.57
6.29
4.45
4.45
4.44
4.43
2.39
2.38
1.00
1.05
1.01
1.00
2.07
6.46
f1 (ppm)

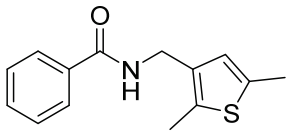

163.80
158.65
156.03
136.89
134.43
133.45
132.47
131.45
131.41
126.33
124.53
124.51
124.48
124.47
124.47
123.46
123.48
120.81
120.68
120.47
120.40
120.34
120.13
120.00
119.79
119.66
77.16 CDCl3
37.76
15.20
12.95
f1 (ppm)


Cl
O
F3C
N
H
S
Cl

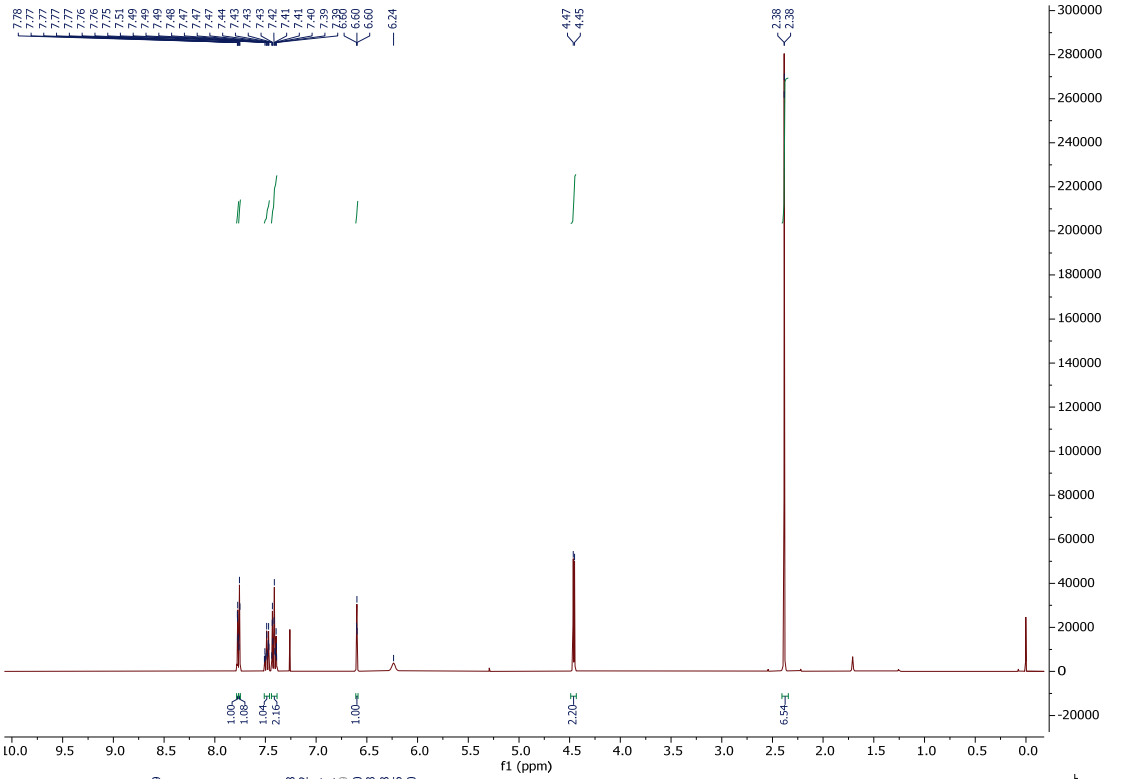
39

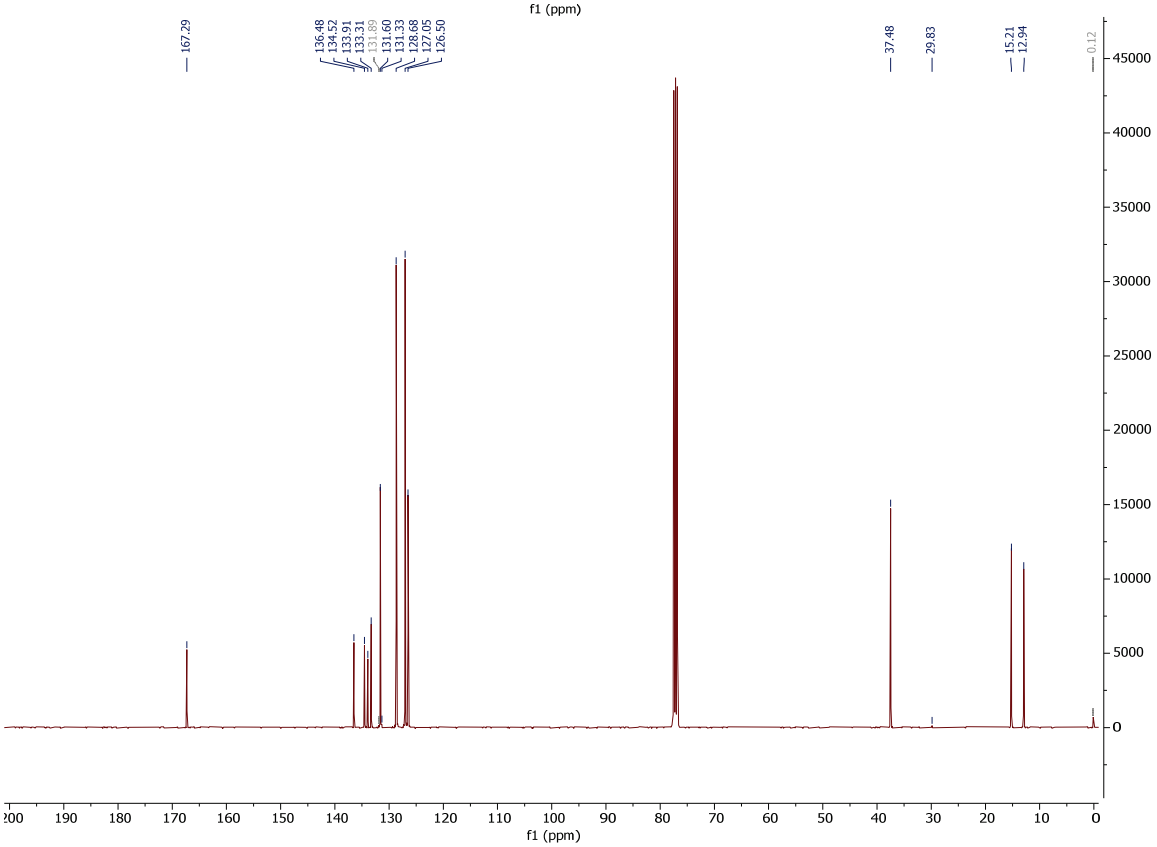

7.72
7.72
7.67
7.67
6.59
6.59
6.03
4.47
4.45
2.39
2.39
2.39
0.97
0.91
0.95
0.90
2.05
6.10
f1 (ppm)

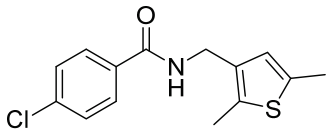

164.29
139.82
136.84
134.48
133.54
132.60
132.10
131.38
131.06
130.75
130.43
129.24
129.18
129.13
129.07
127.50
126.31
126.01
123.29
120.56
117.84
37.67
15.22
12.98
f1 (ppm)



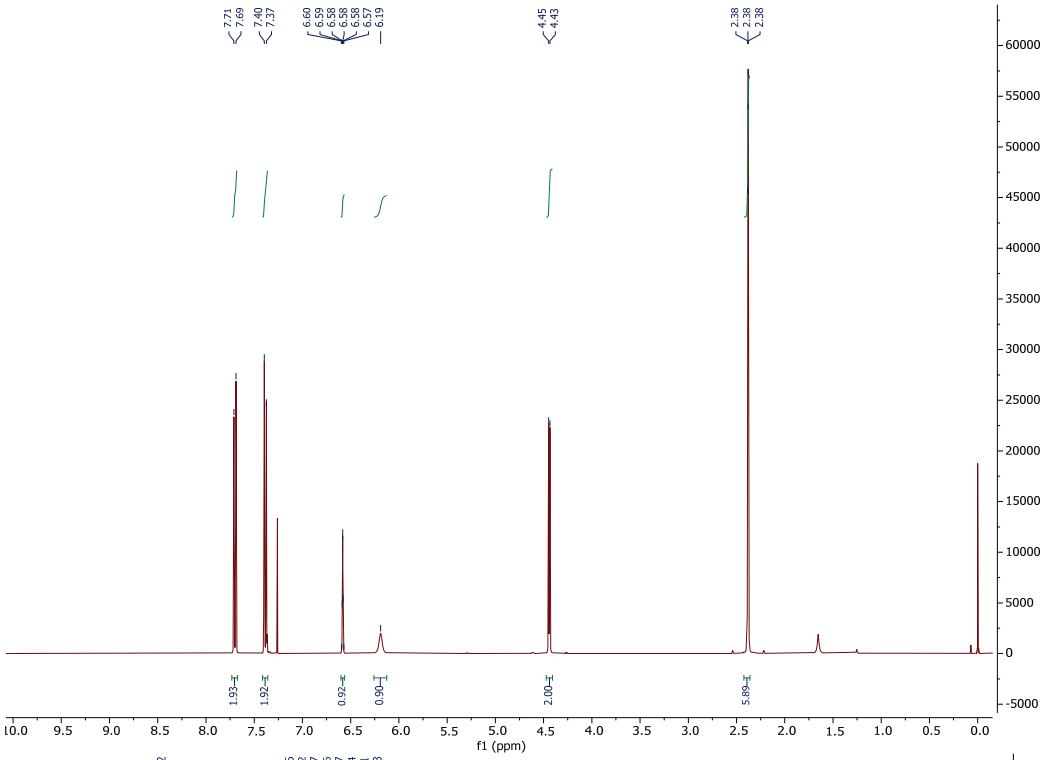
40

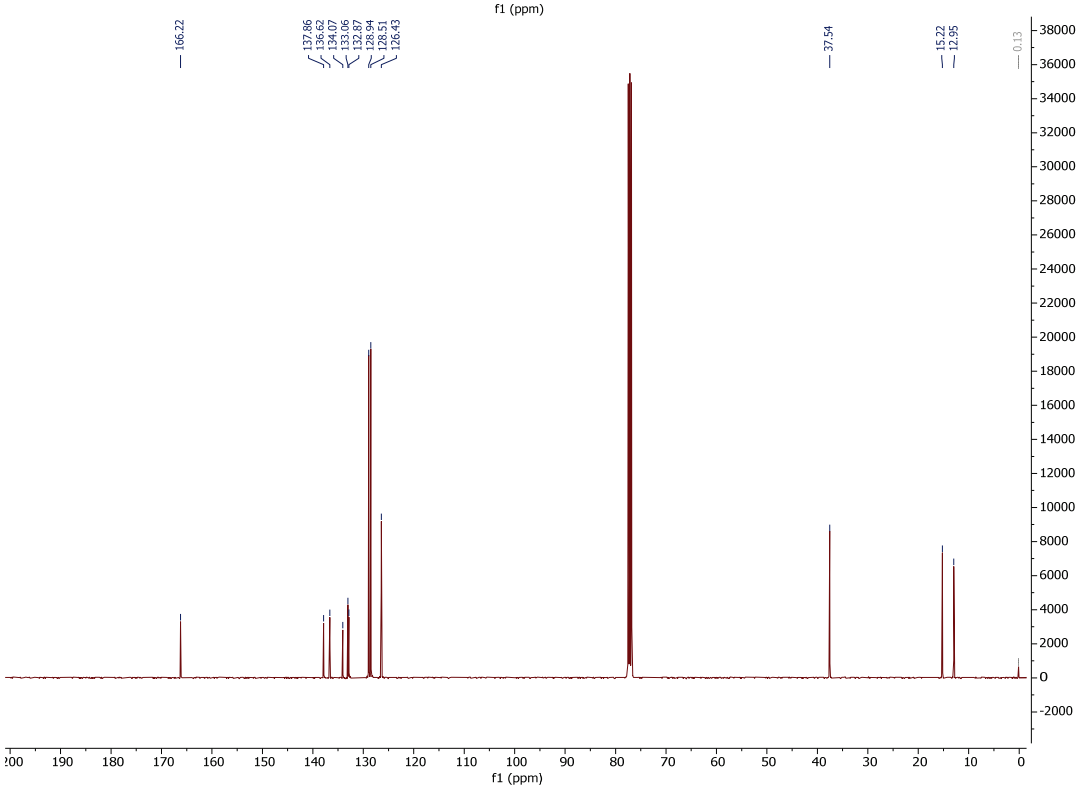

7.78
7.77
7.77
7.77
7.77
7.76
7.76
7.75
7.51
7.49
7.49
7.48
7.47
7.47
7.47
7.44
7.43
7.43
7.42
7.41
7.41
7.40
7.39
6.60
6.60
6.60
6.24
4.47
4.45
2.38
2.38
1.00
1.08
1.04
2.16
1.00
2.20
6.54
f1 (ppm)

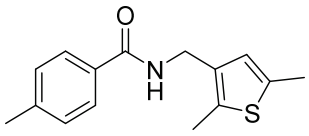

167.29
136.48
134.52
133.91
133.31
131.89
131.60
131.33
128.68
127.05
126.50
37.48
29.83
15.21
12.94
0.12
f1 (ppm)


O
N
H
Cl
S

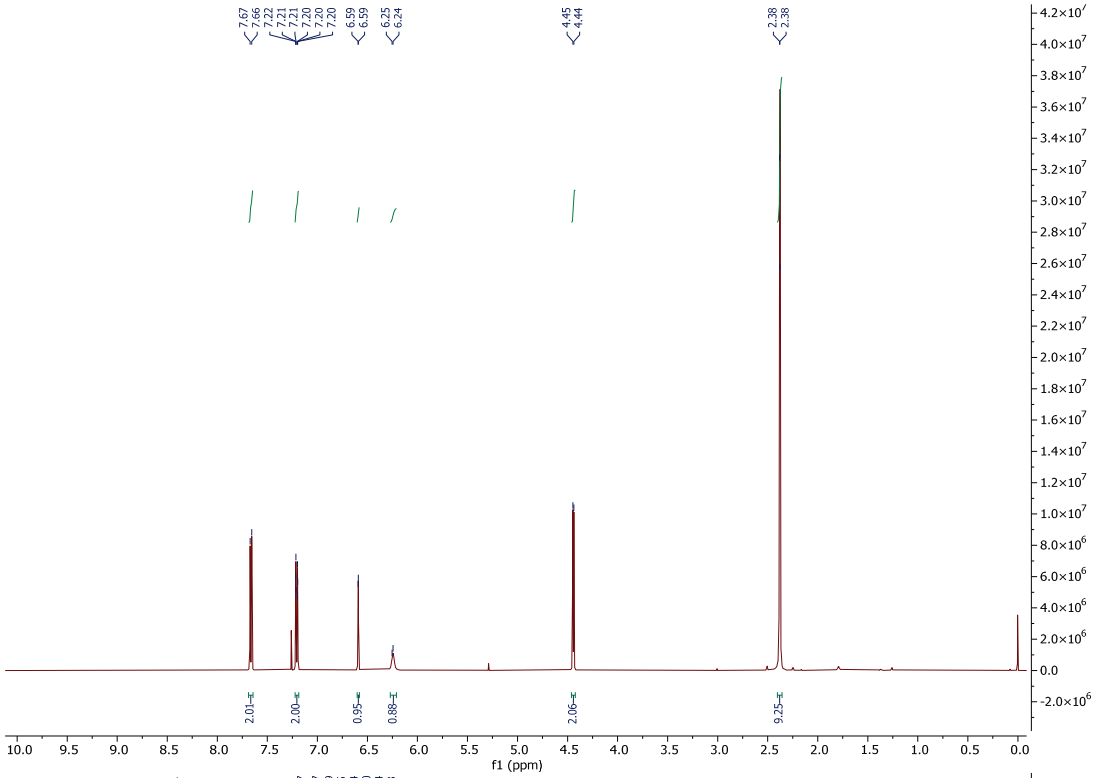
41

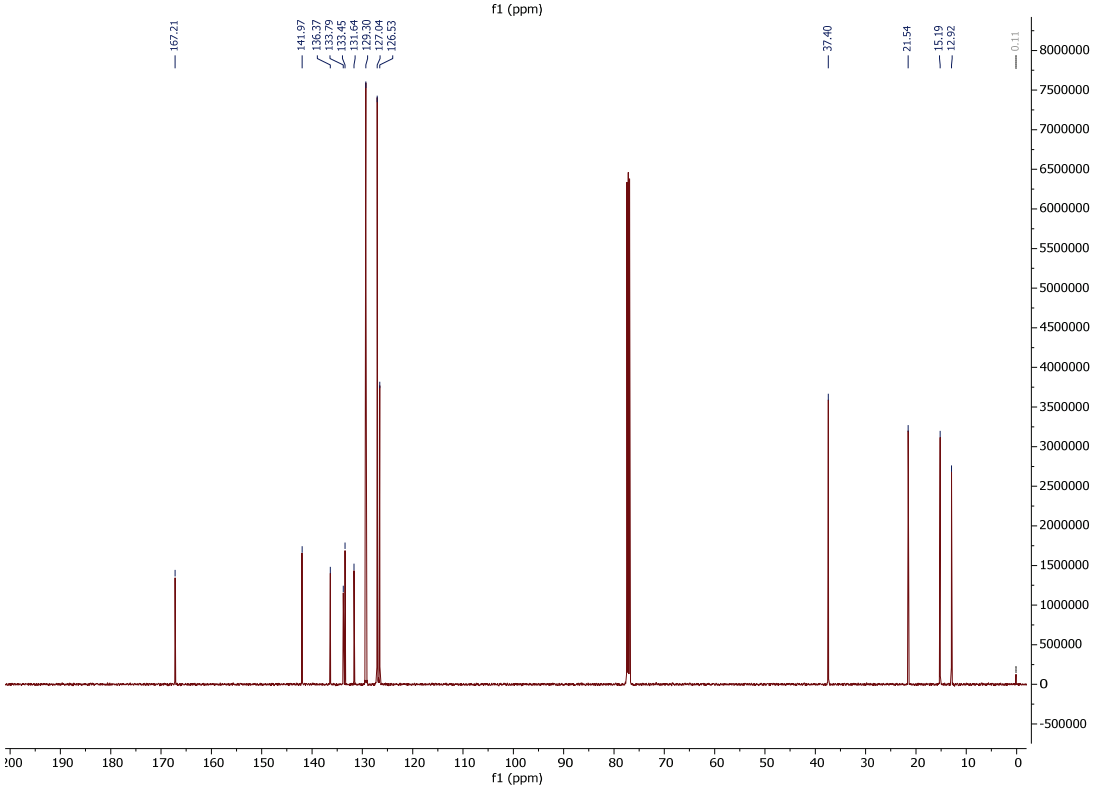

7.71
7.69
7.40
7.37
6.60
6.59
6.58
6.58
6.58
6.57
6.19
4.45
4.43
2.38
2.38
2.38
1.93
1.92
0.92
0.90
2.00
5.89
f1 (ppm)

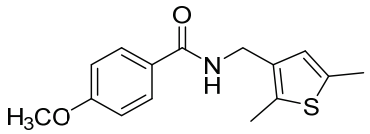

166.22
137.86
136.62
134.07
133.06
132.87
128.84
128.51
126.43
37.54
15.22
12.95
0.13
f1 (ppm)


O
N
H
S

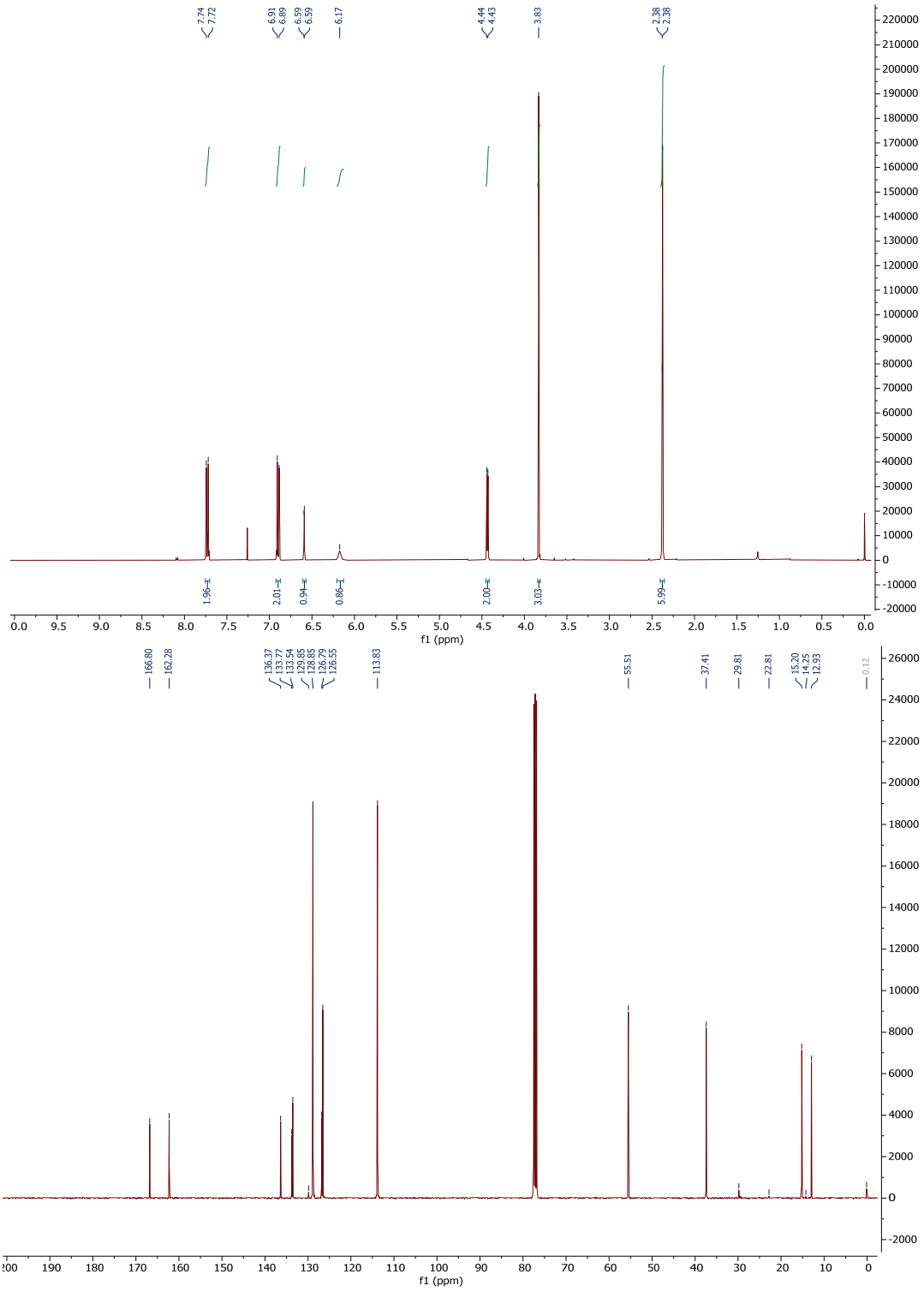
42

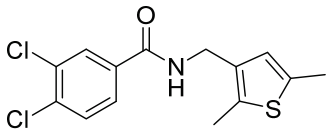

7.67
7.66
7.22
7.21
7.21
7.20
7.20
7.20
6.59
6.59
6.25
6.24
4.45
4.44
2.38
2.38
2.01
2.00
0.95
0.88
2.06
9.25
f1 (ppm)


167.21
141.97
136.37
133.45
131.64
129.30
127.04
126.53
37.40
21.54
15.19
12.92
0.11
f1 (ppm)

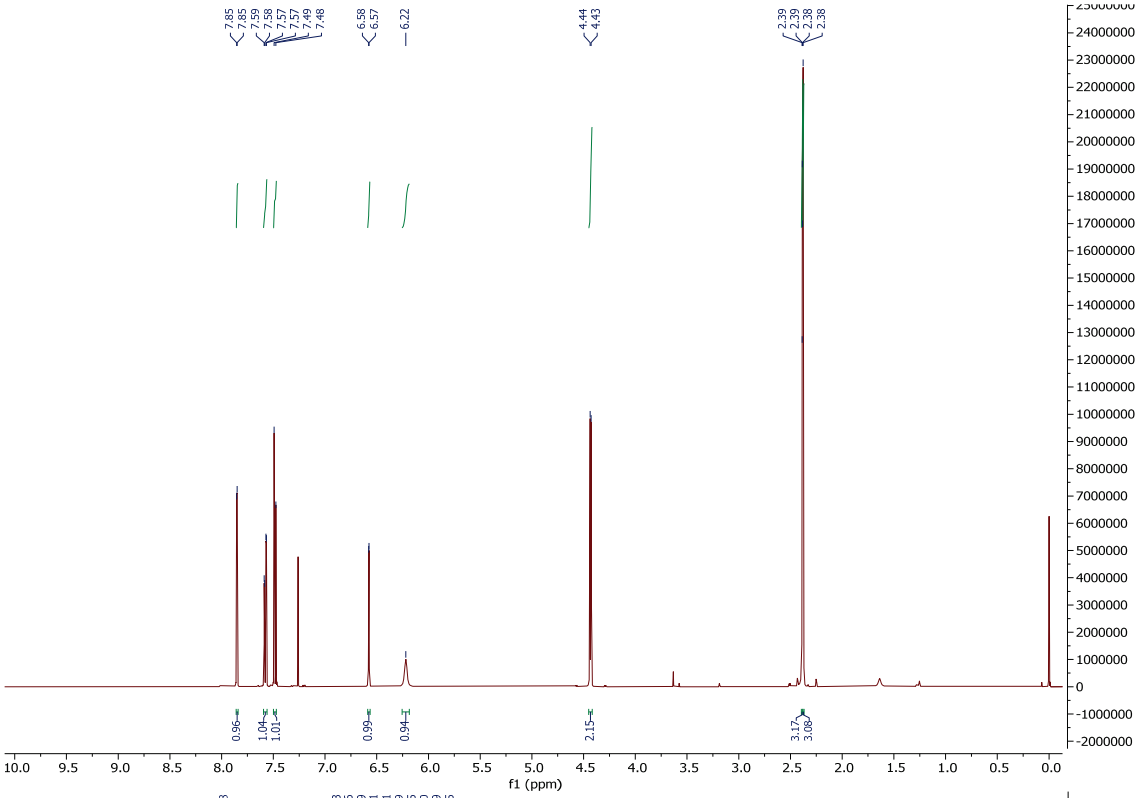

O
N
H
H3CO
S

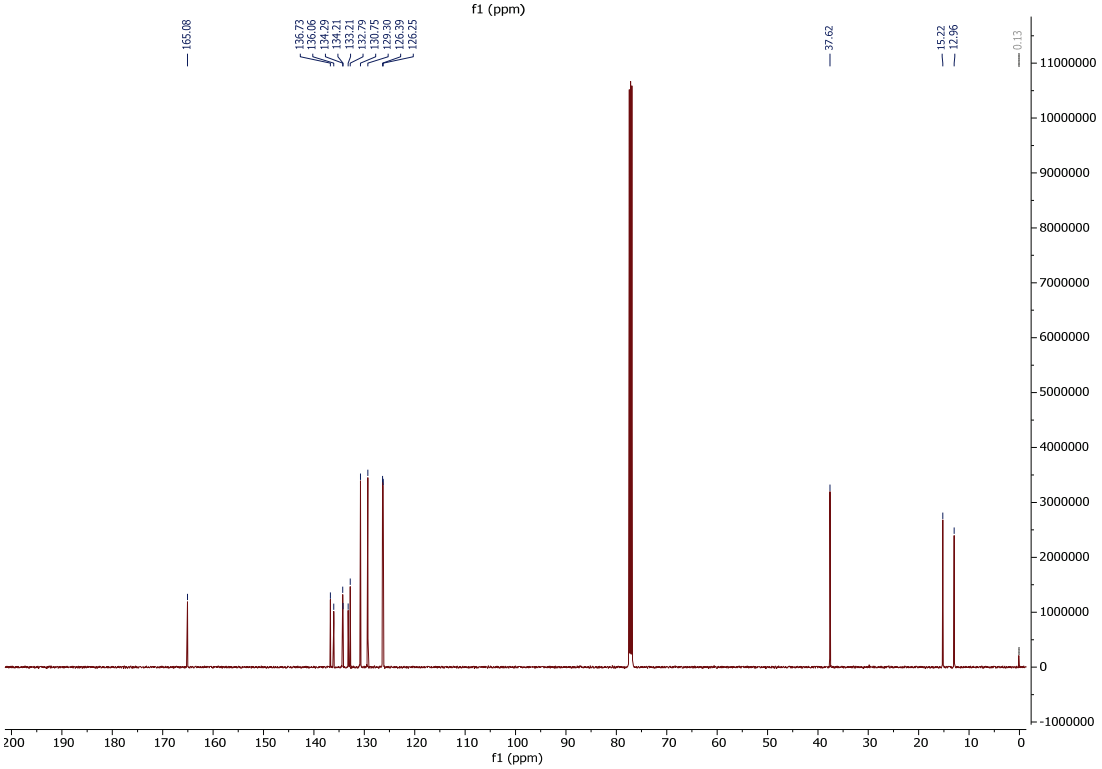
43

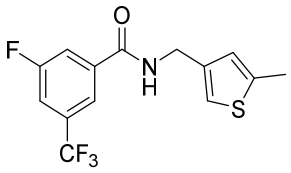

7.74
7.72
6.91
6.89
6.59
6.59
6.17
4.44
4.43
3.83
2.38
2.38
1.96
2.01
0.94
0.86
2.00
3.03
5.99
f1 (ppm)
166.80
162.28
136.37
133.77
133.54
129.85
128.85
126.79
126.55
113.83
55.51
37.41
29.81
22.81
15.20
14.25
12.93
0.12
f1 (ppm)


O
Cl
N
H
Cl
S

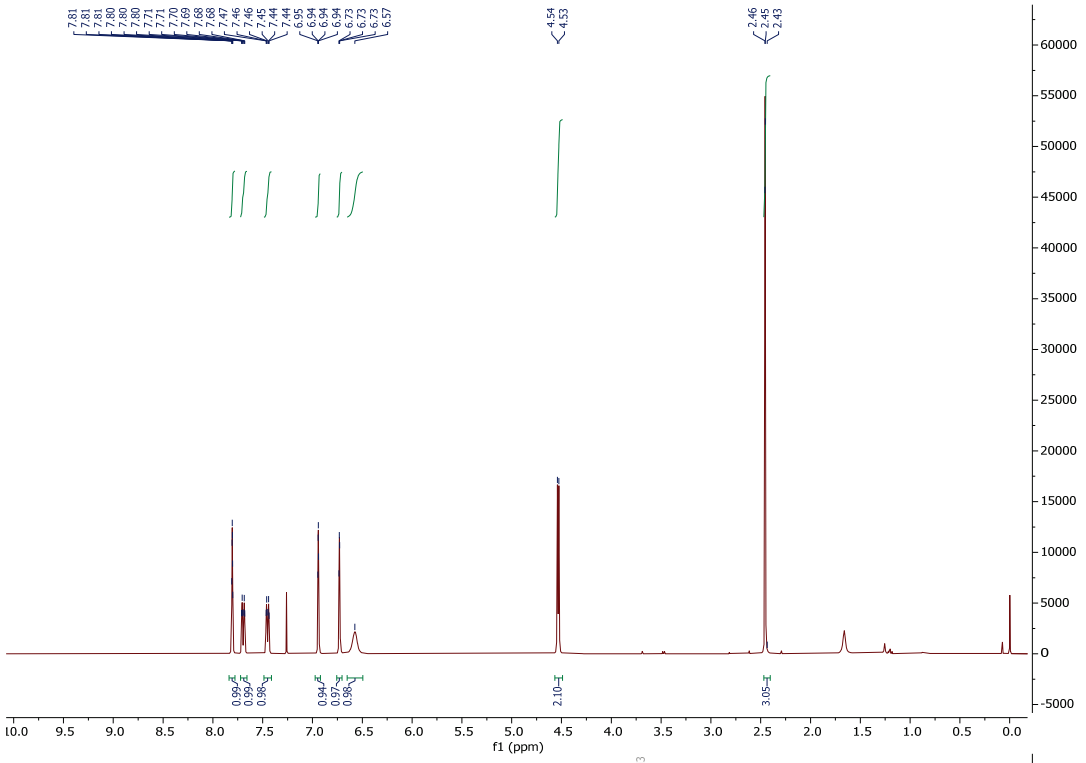
44

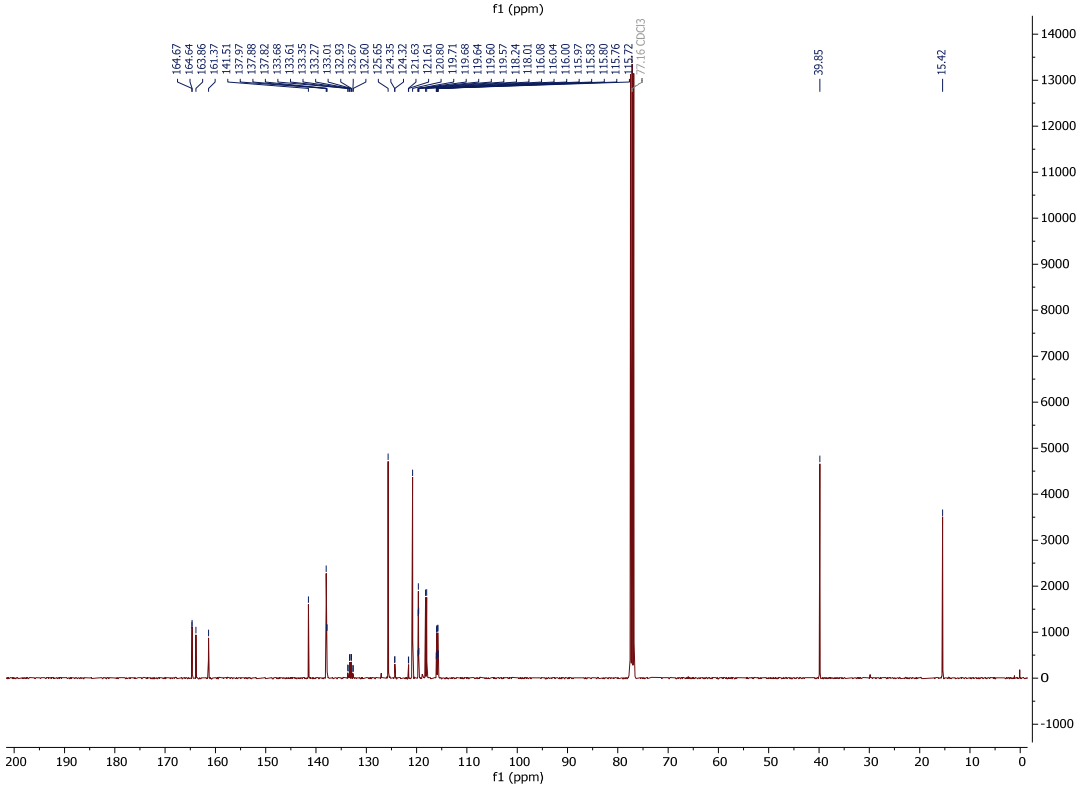

7.85
7.85
7.59
7.58
7.57
7.57
7.49
7.48
6.58
6.57
6.22
4.44
4.43
2.39
2.39
2.38
2.38
0.96
1.04
1.01
0.99
0.94
2.15
3.17
3.08
f1 (ppm)

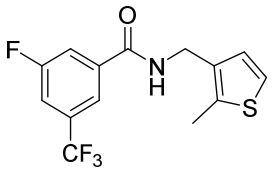

165.08
136.73
136.06
134.29
134.21
133.21
132.79
130.75
129.30
126.39
126.25
37.62
15.22
12.96
0.13
f1 (ppm)


O
F
N
H
S
CF3

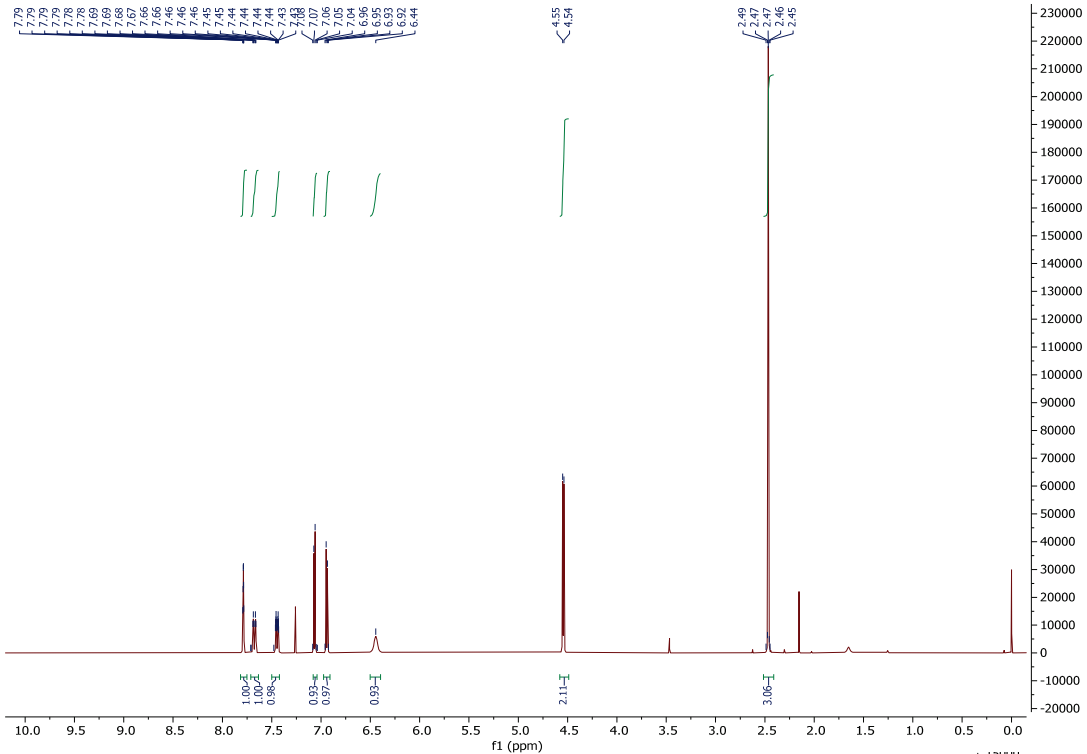
45

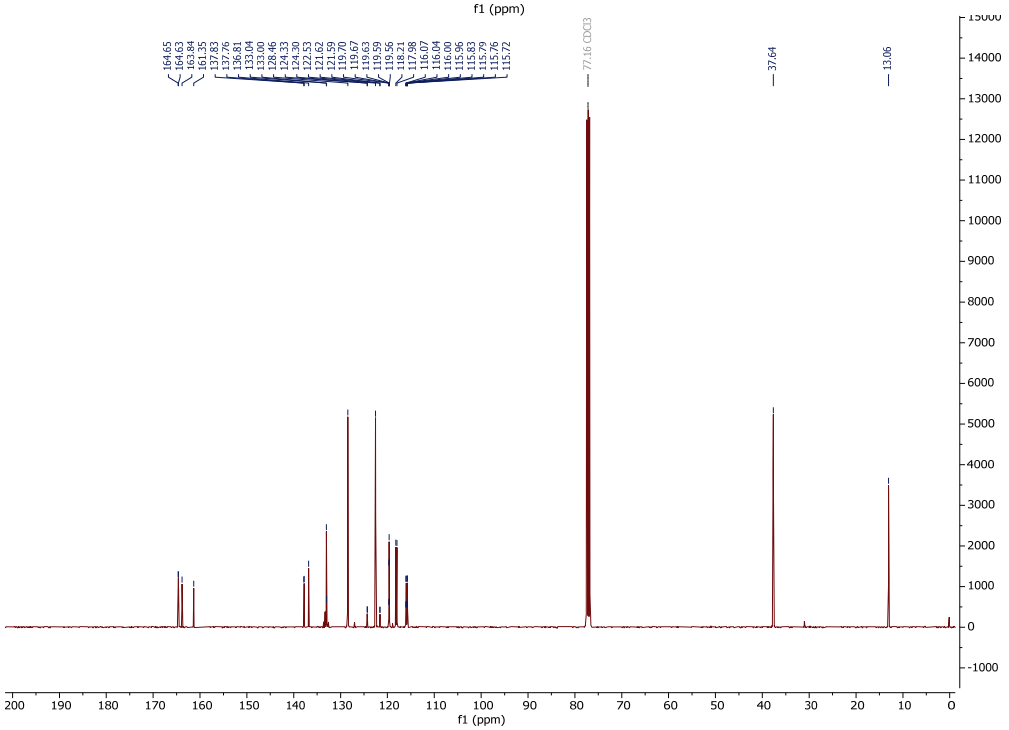

7.81
7.81
7.81
7.80
7.80
7.80
7.71
7.71
7.70
7.69
7.68
7.68
7.47
7.46
7.46
7.45
7.44
7.44
6.95
6.94
6.94
6.94
6.73
6.73
6.57
4.54
4.53
2.46
2.45
2.43
0.99
0.99
0.98
0.94
0.97
0.98
2.10
3.05
f1 (ppm)

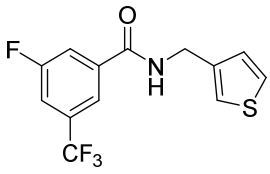

164.67
164.64
163.86
161.37
141.51
137.97
137.88
137.82
133.68
133.61
133.35
133.27
133.01
132.93
132.67
132.60
125.65
124.35
124.32
121.63
121.61
120.89
119.71
119.68
119.64
119.60
119.57
118.24
118.01
116.08
116.04
116.00
115.97
115.83
115.80
115.76
115.72
77.16 CDCl3
39.85
15.42
f1 (ppm)


O
F
N
H
S
CF3

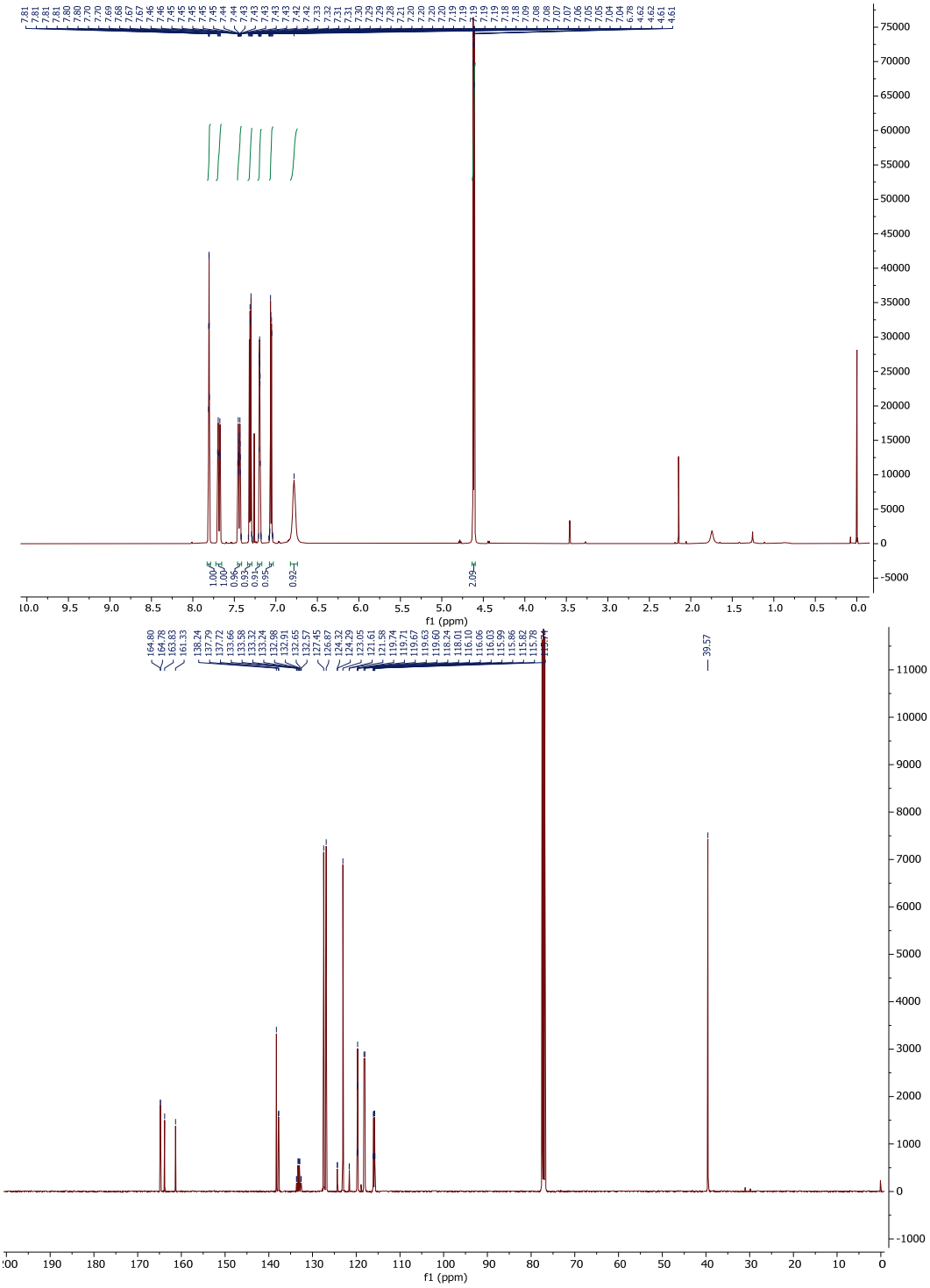
46

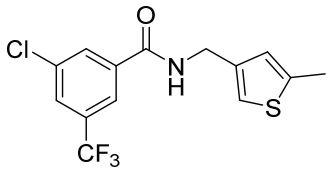

7.79
7.79
7.79
7.79
7.78
7.78
7.69
7.69
7.68
7.67
7.66
7.66
7.46
7.46
7.46
7.45
7.45
7.44
7.44
7.44
7.44
7.43
7.43
7.08
7.07
7.06
7.05
7.04
6.96
6.95
6.93
6.92
6.44
4.55
4.54
2.49
2.47
2.47
2.46
2.45
1.00
1.00
0.98
0.93
0.97
0.93
2.11
3.06
f1 (ppm)


164.65
164.63
163.84
161.35
137.83
137.76
136.81
133.04
133.00
128.46
124.33
124.30
122.53
121.62
121.59
119.70
119.67
119.63
119.59
119.56
118.21
117.98
116.07
116.04
116.00
115.96
115.83
115.79
115.76
115.72
77.16 CDCl3
37.64
13.06
f1 (ppm)

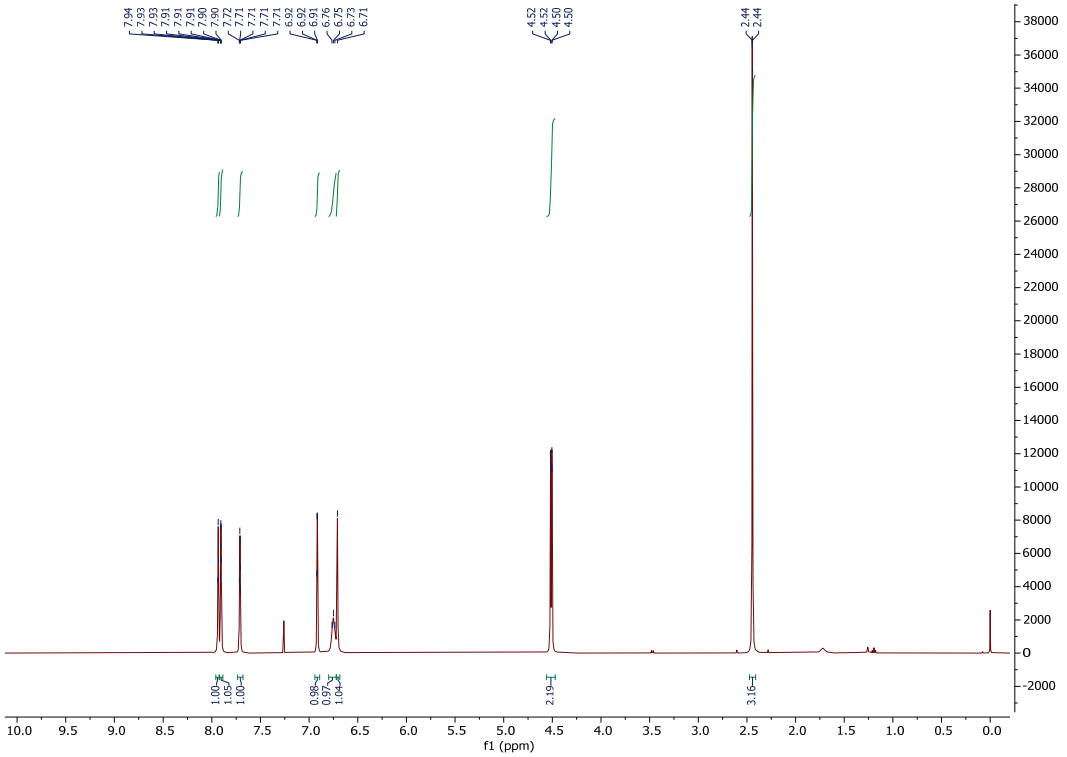


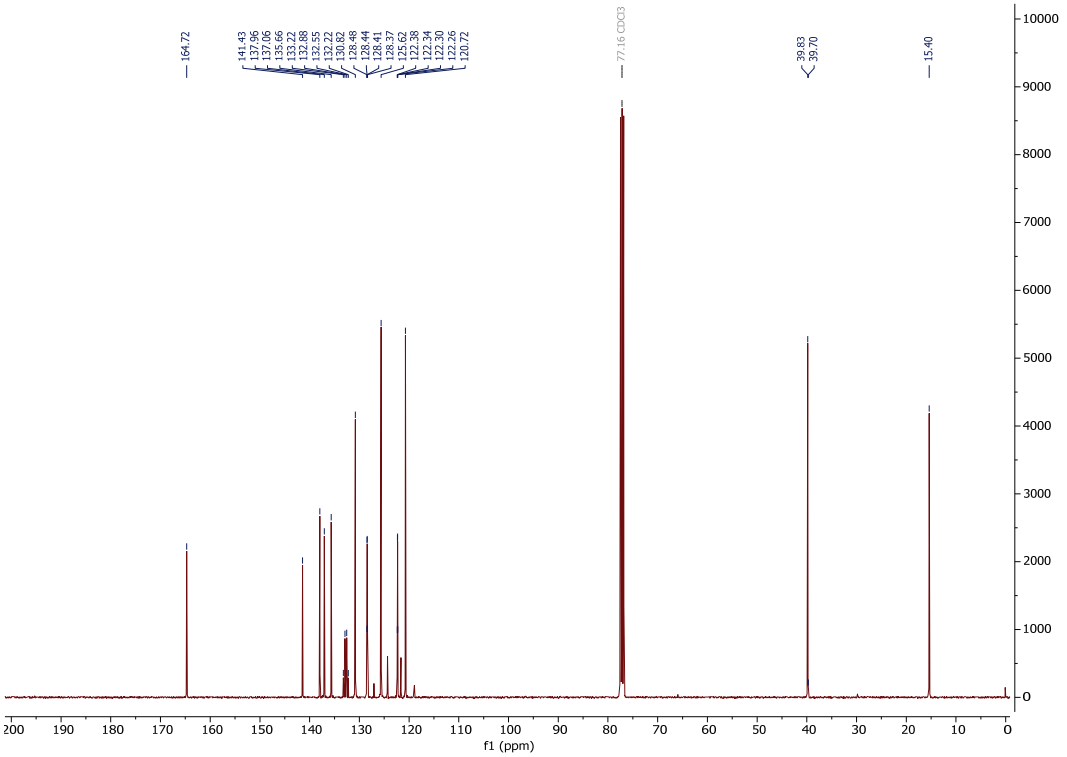
47

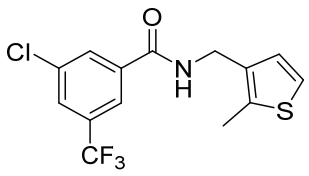



O
Cl
N
H
S
CF3

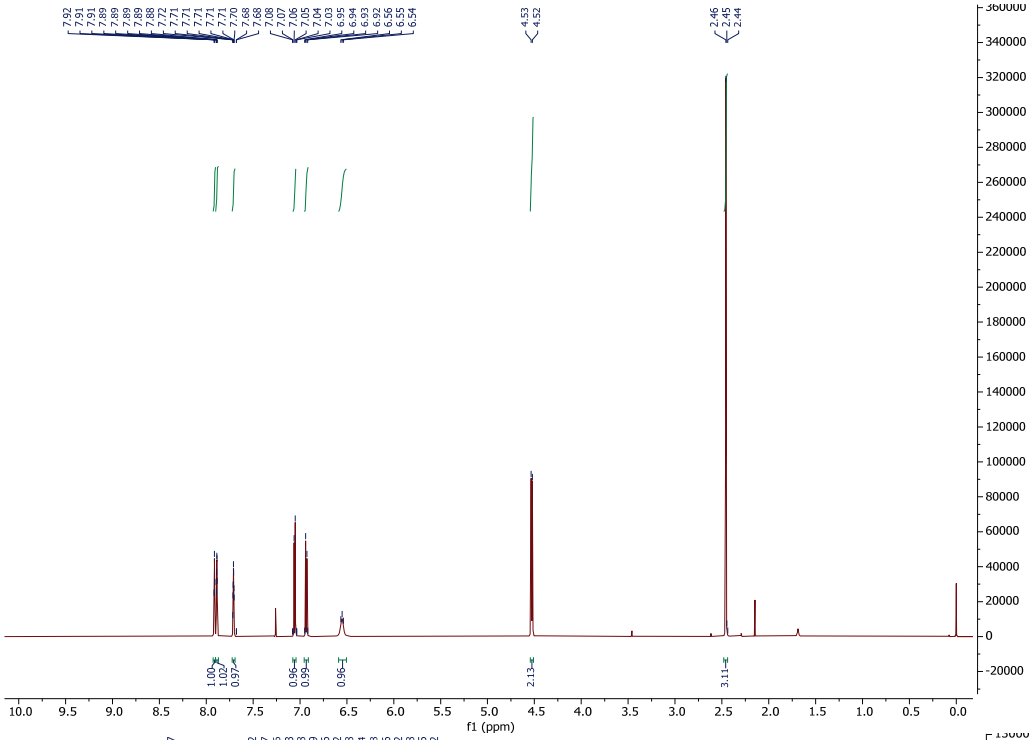
48

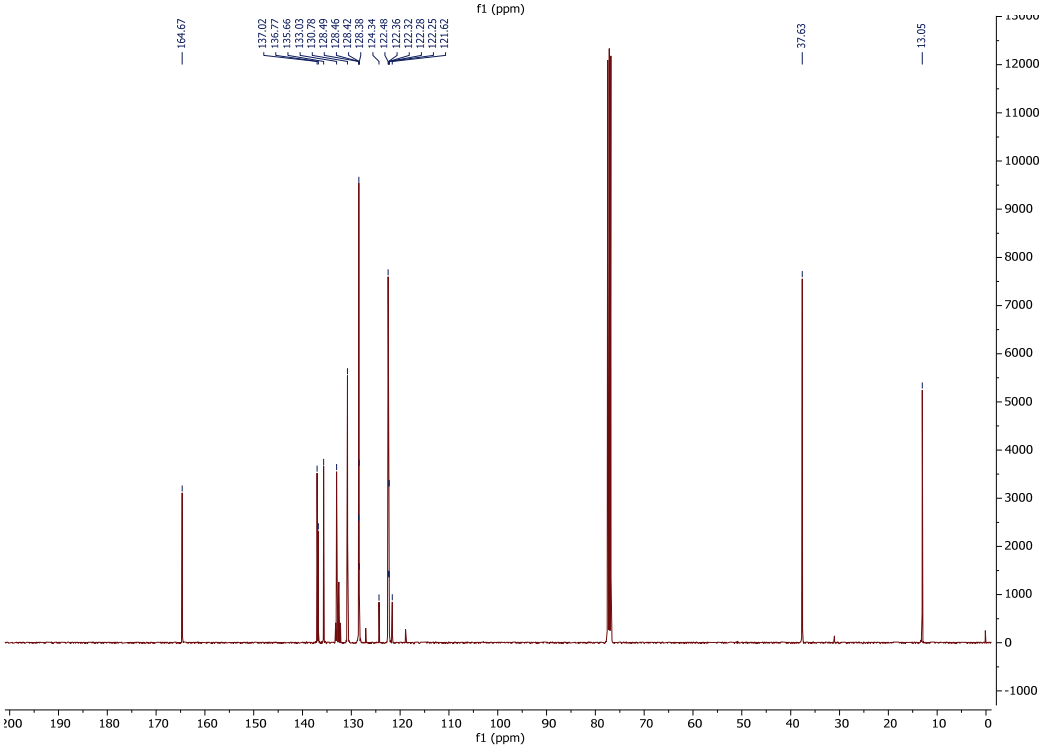

7.94
7.93
7.93
7.91
7.91
7.91
7.90
7.90
7.72
7.71
7.71
7.71
7.71
6.92
6.92
6.91
6.76
6.75
6.73
6.71
4.52
4.52
4.50
4.50
2.44
2.44
1.00
1.05
1.00
0.98
0.97
1.04
2.19
3.16
f1 (ppm)

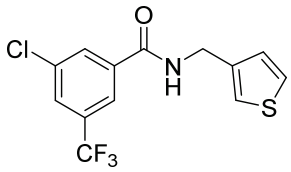

164.72
141.43
137.96
137.06
135.66
133.22
132.88
132.55
132.22
130.82
128.48
128.44
128.41
128.37
125.62
122.38
122.34
122.30
122.26
120.72
77.16 CDCl3
39.83
39.70
15.40
f1 (ppm)


O
Cl
N
H
S
CF3

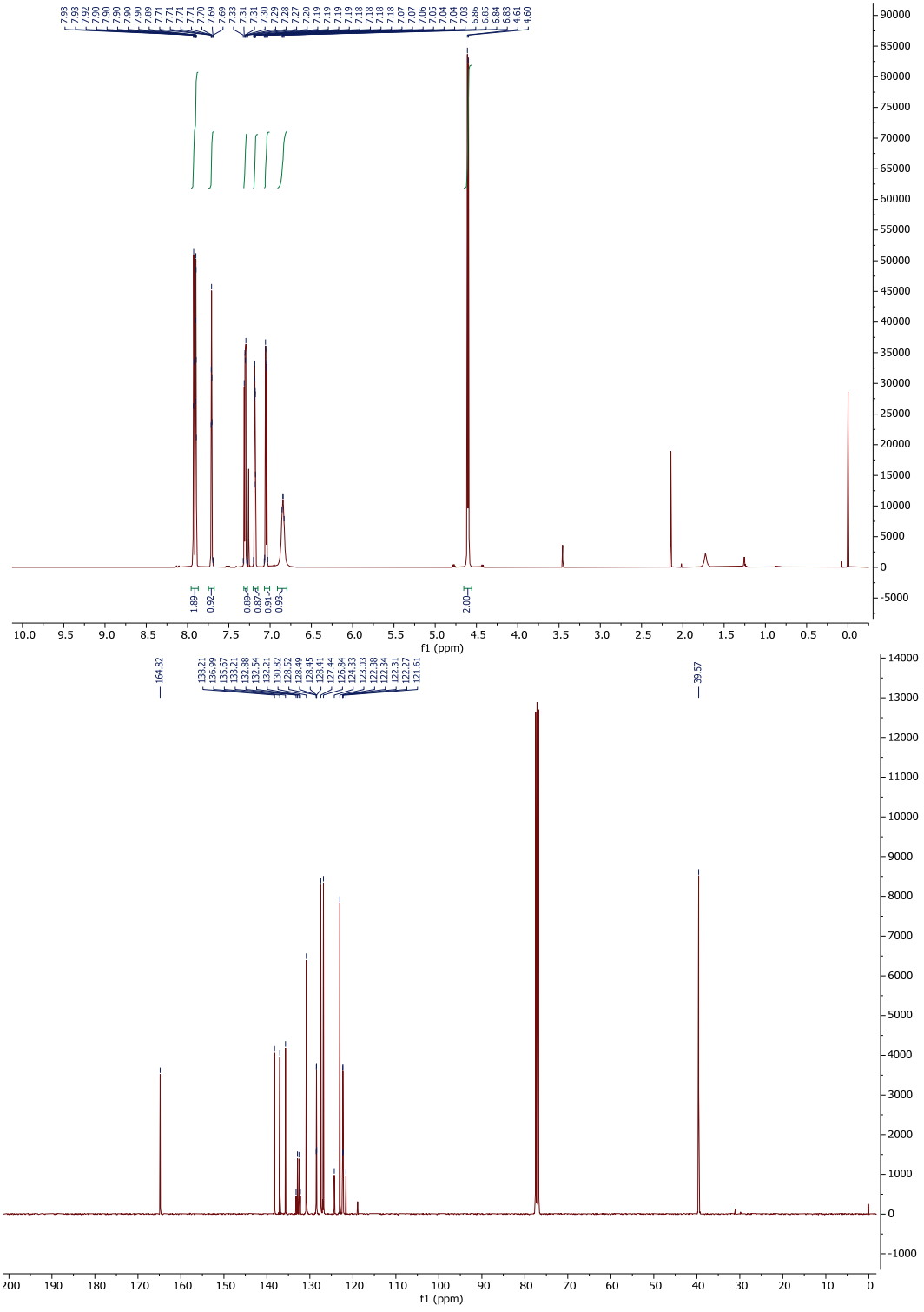
49

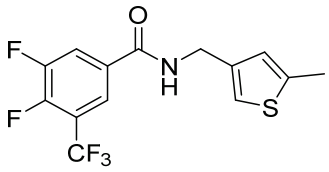

7.92
7.91
7.91
7.89
7.89
7.89
7.88
7.72
7.71
7.71
7.71
7.71
7.71
7.70
7.68
7.68
7.68
7.07
7.06
7.05
7.04
7.03
6.95
6.94
6.93
6.92
6.56
6.55
6.54
4.53
4.52
2.46
2.45
2.44
1.00
1.02
0.97
0.96
0.99
0.96
2.13
3.11
f1 (ppm)


164.67
137.02
136.77
135.66
133.03
130.78
129.49
128.46
128.42
128.38
124.34
122.48
122.36
122.32
122.28
122.25
121.62
37.63
13.05
f1 (ppm)

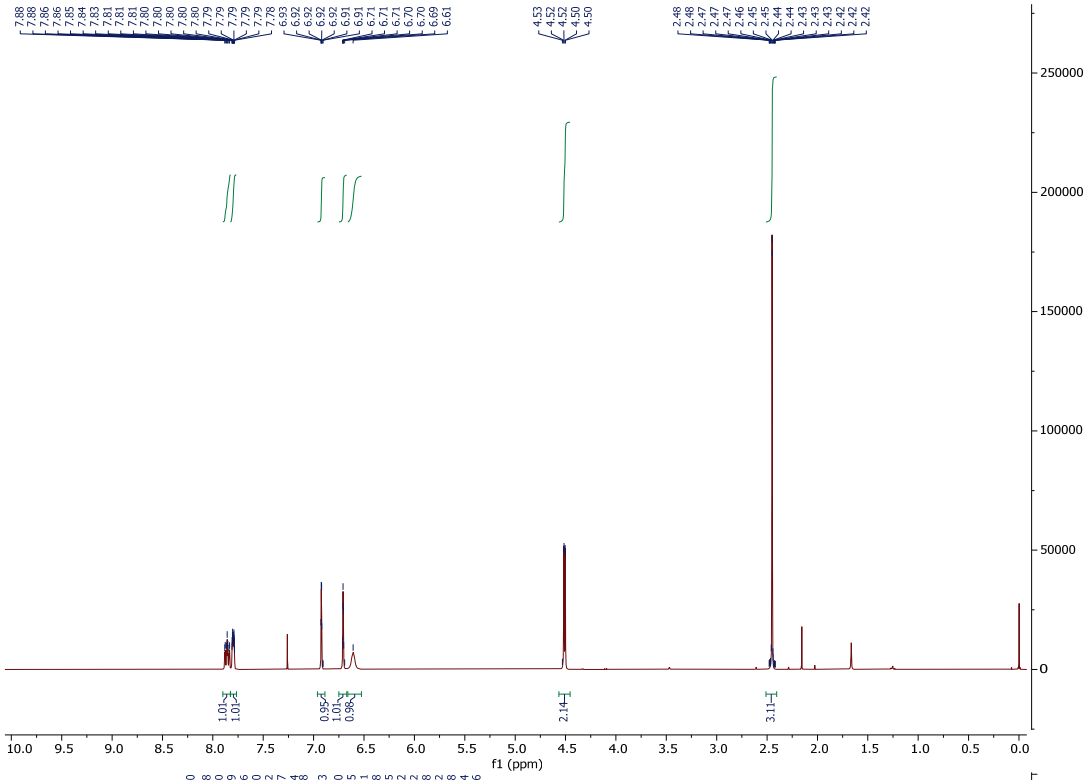


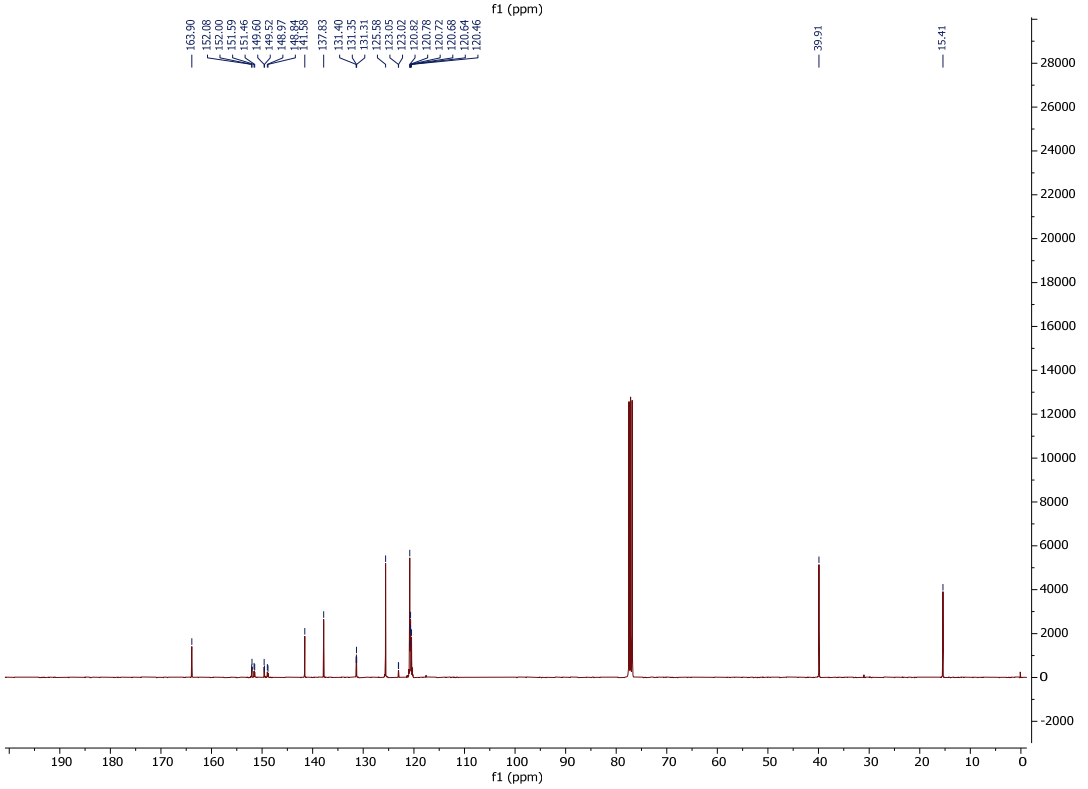
50

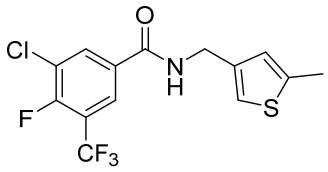



O
F
N
H
F
S
CF3

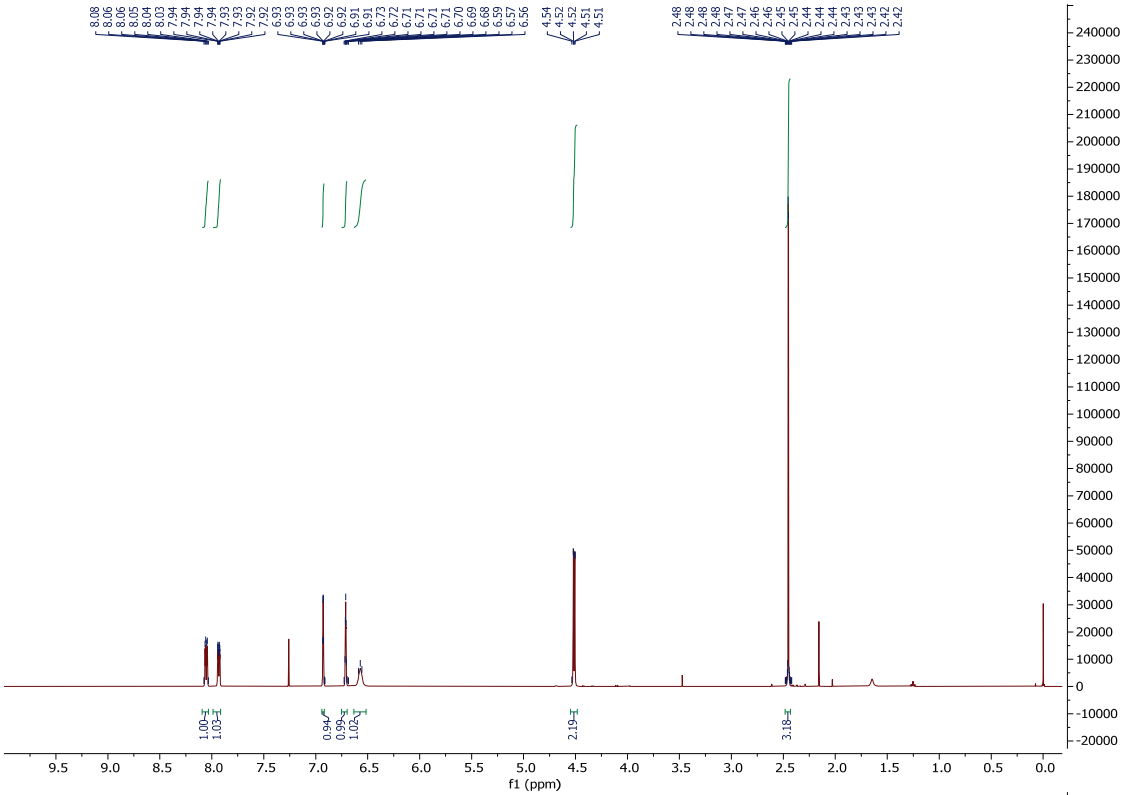
51

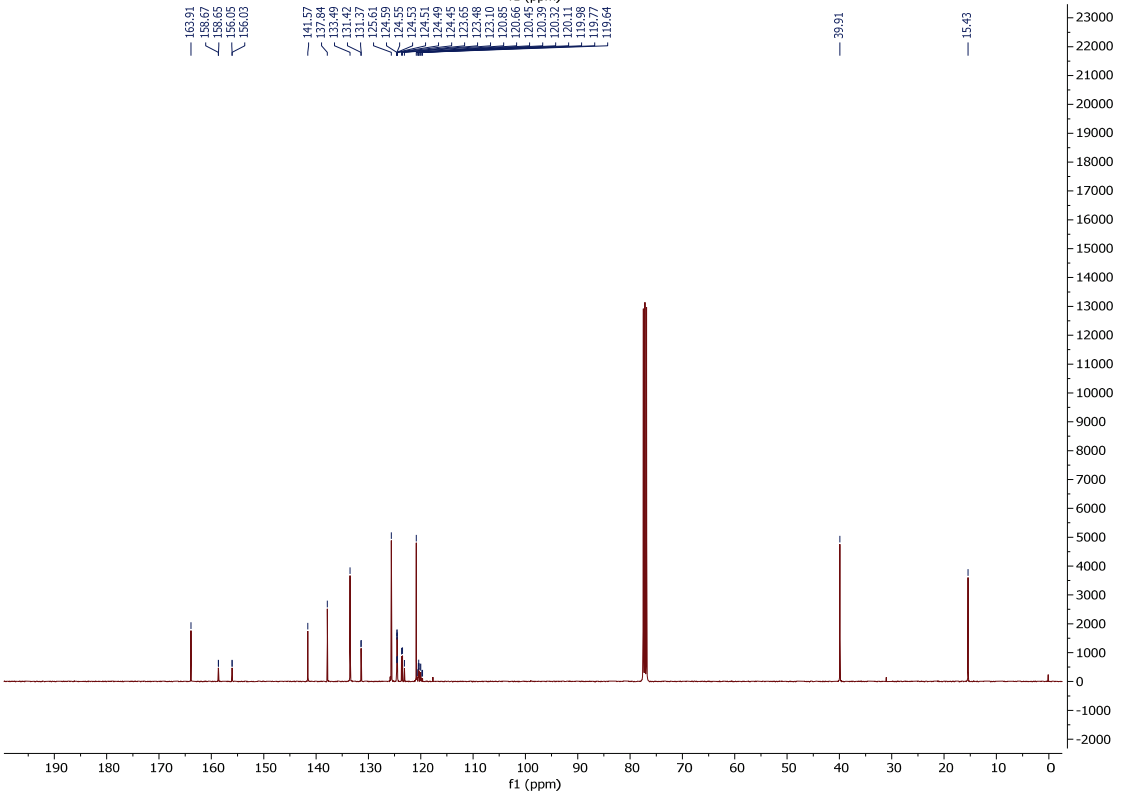

7.88
7.88
7.86
7.86
7.85
7.84
7.83
7.81
7.81
7.81
7.80
7.80
7.80
7.80
7.80
7.79
7.79
7.79
7.79
7.78
6.93
6.92
6.92
6.92
6.92
6.91
6.91
6.71
6.71
6.71
6.70
6.70
6.69
6.61
4.53
4.52
4.52
4.51
4.50
4.50
2.48
2.48
2.47
2.47
2.47
2.46
2.45
2.45
2.44
2.44
2.43
2.43
2.43
2.42
2.42
2.42
250000
200000
150000
100000
50000
0
1.01
1.01
0.95
1.01
0.98
2.14
3.11
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

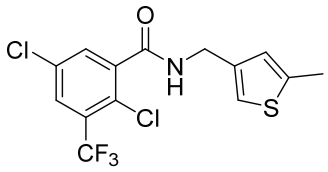

163.90
152.08
152.00
151.59
151.46
149.60
149.52
149.37
148.84
141.58
137.83
131.40
131.35
131.31
125.58
123.05
123.02
120.82
120.78
120.72
120.68
120.64
120.46
39.91
15.41
28000
26000
24000
22000
20000
18000
16000
14000
12000
10000
8000
6000
4000
2000
0
-2000
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)



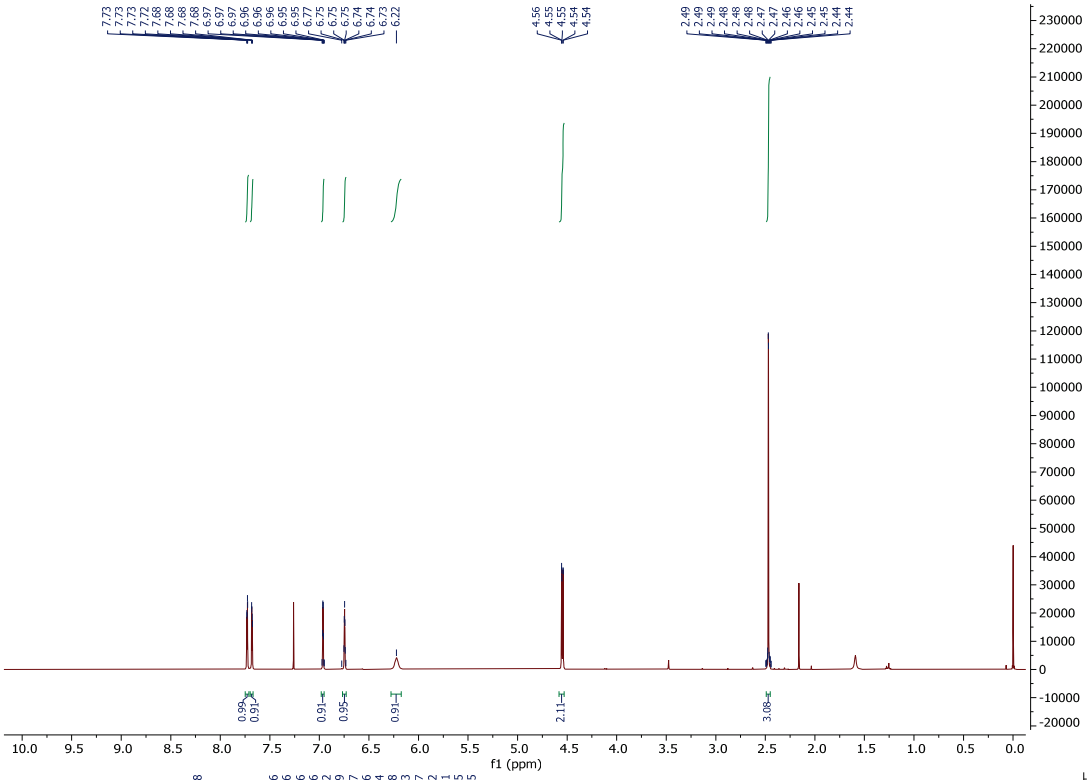
52

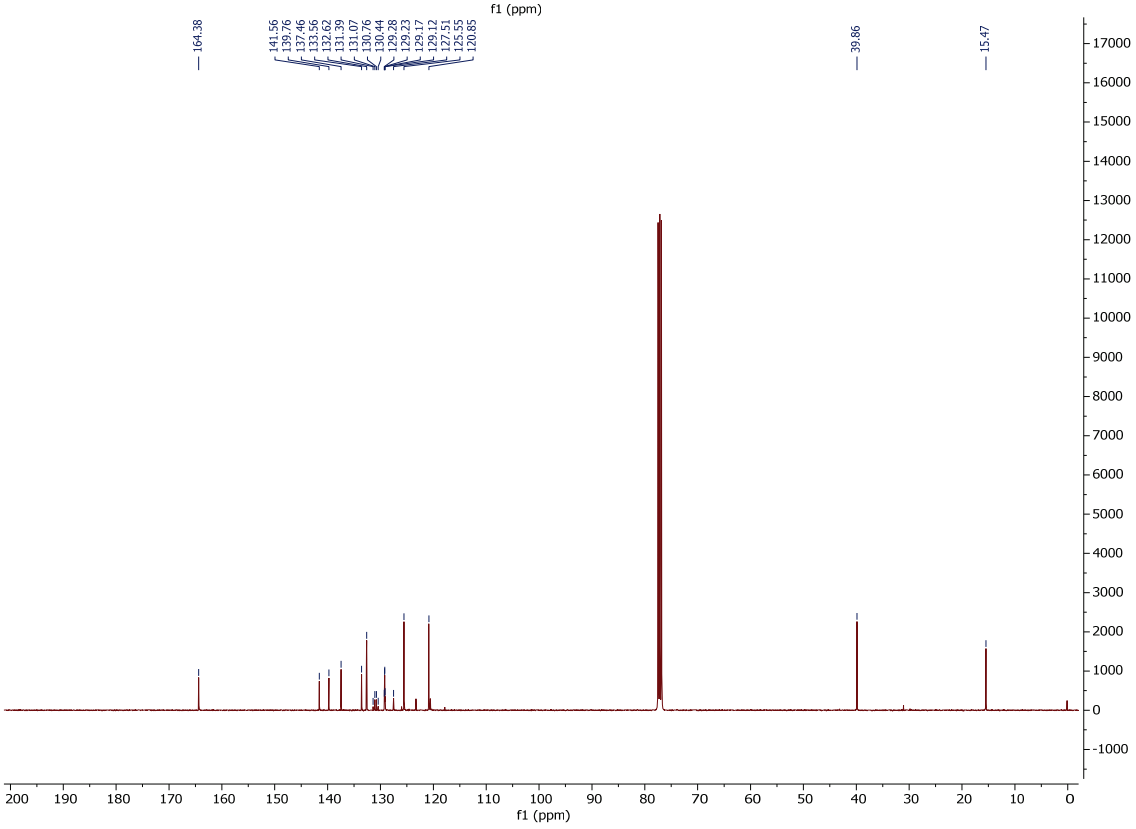


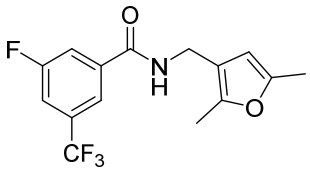



O
Cl
N
H
Cl
S
CF3

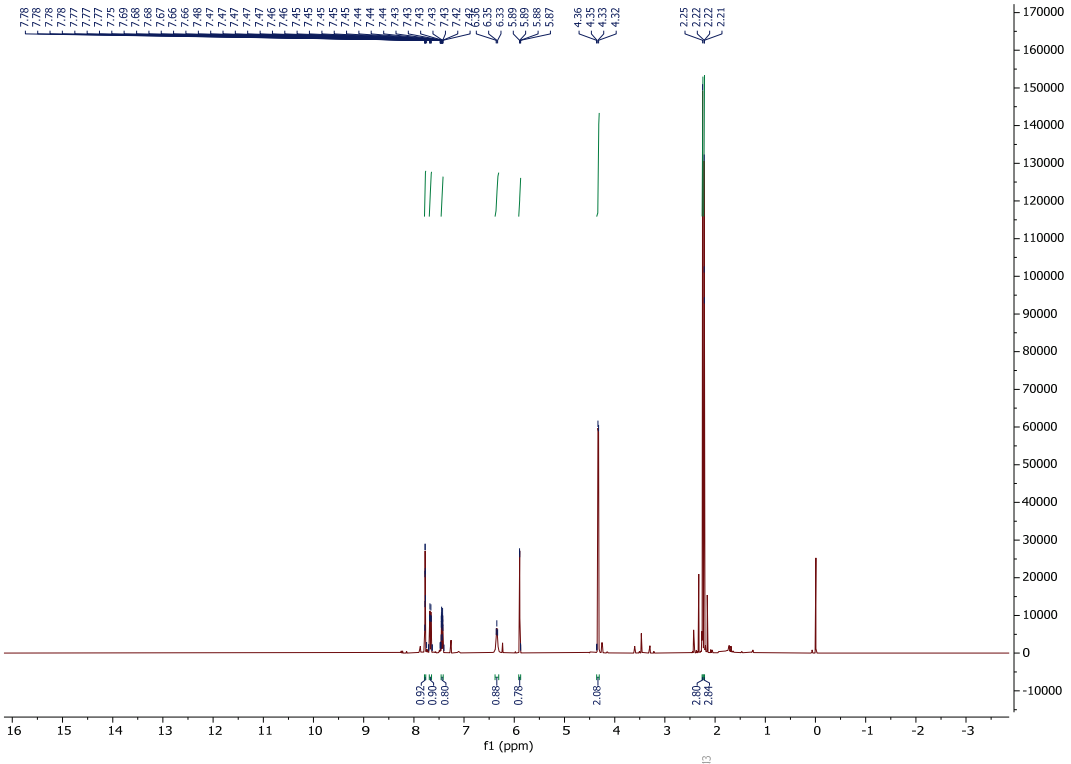
53

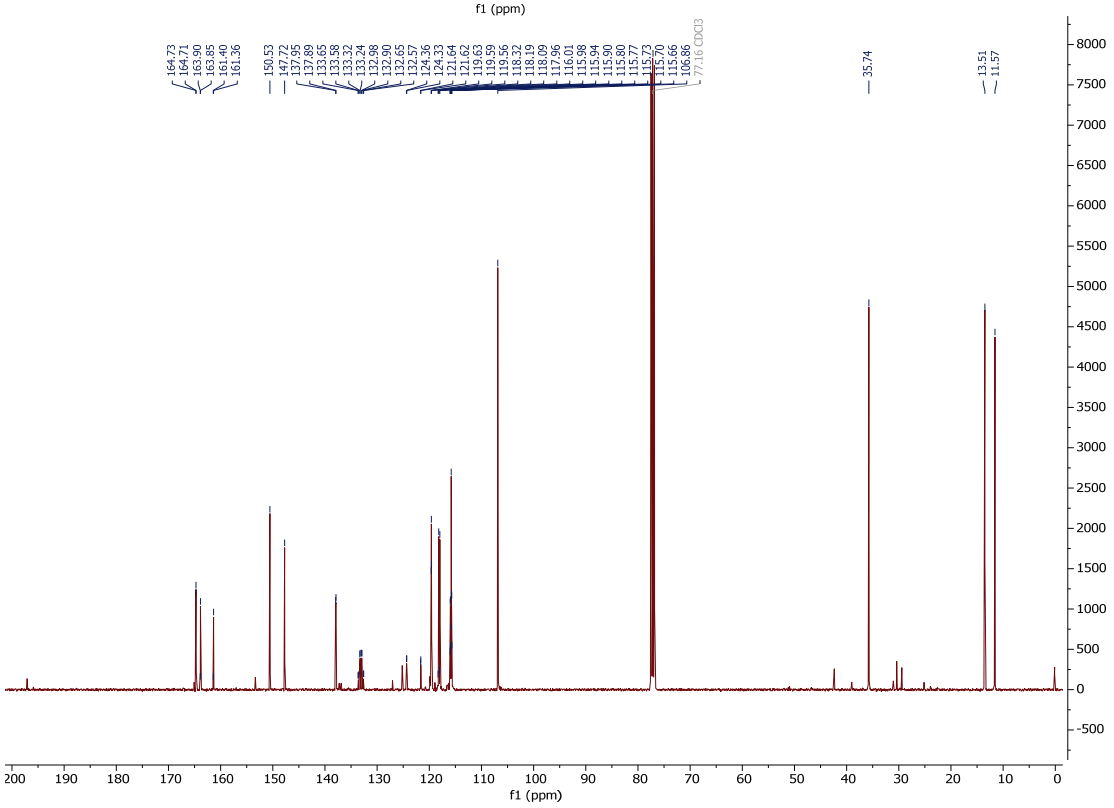

7.73
7.73
7.73
7.72
7.68
7.68
7.68
7.68
6.97
6.97
6.97
6.96
6.96
6.96
6.95
6.95
6.77
6.75
6.75
6.74
6.74
6.73
6.22
4.56
4.55
4.55
4.54
4.54
2.49
2.49
2.49
2.48
2.48
2.48
2.47
2.47
2.46
2.46
2.45
2.45
2.44
2.44
0.95
0.91
0.91
0.95
0.91
2.11
3.08
f1 (ppm)

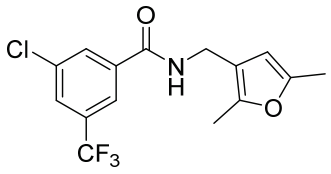

164.38
141.56
139.76
137.46
133.56
132.62
131.39
131.07
130.76
130.44
130.28
129.23
129.17
129.12
127.51
125.55
120.85
39.86
15.47
f1 (ppm)



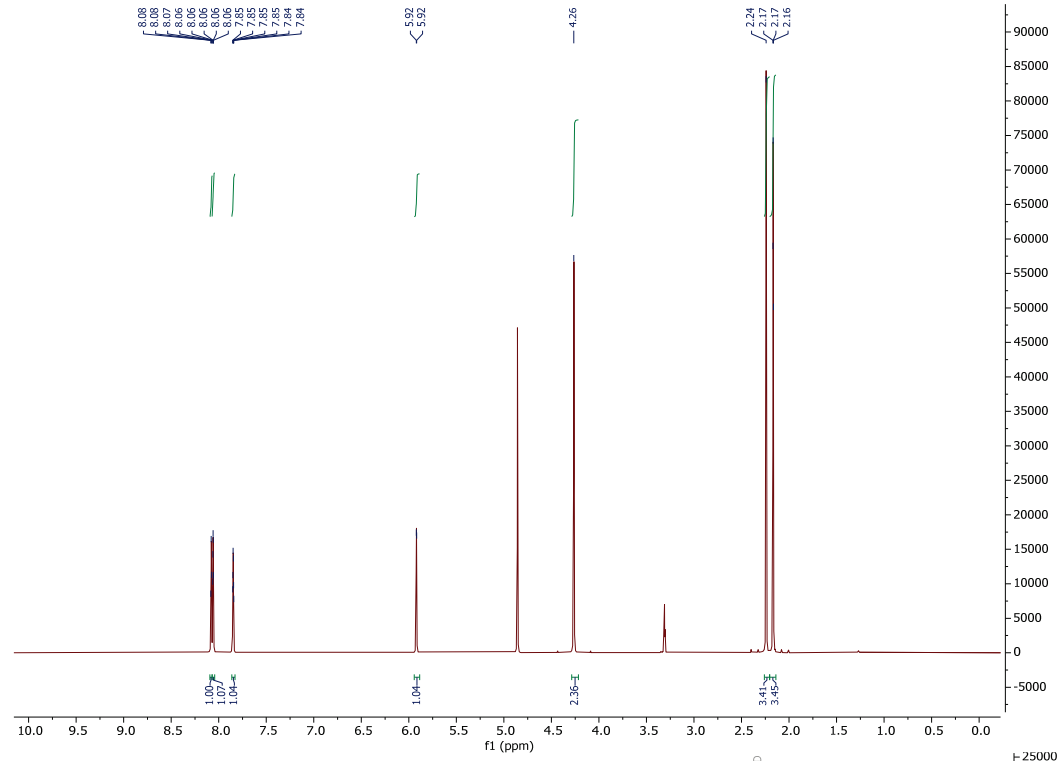
54

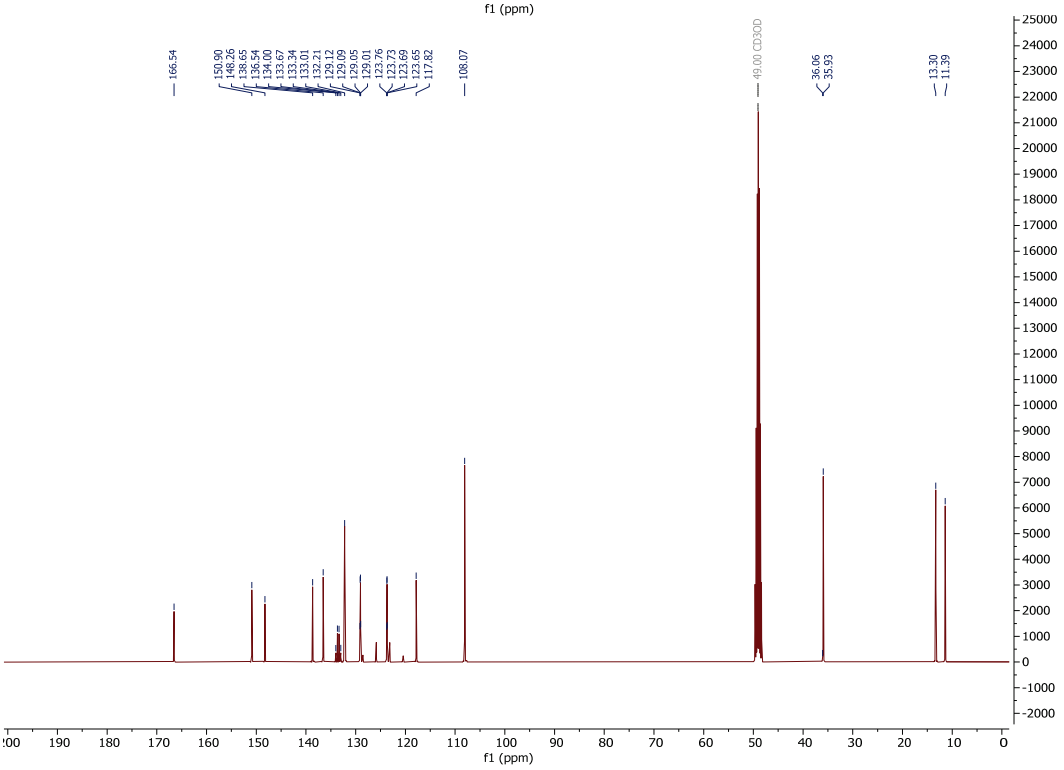


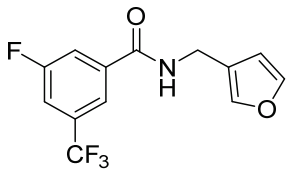



O
Cl
N
H
O
CF3

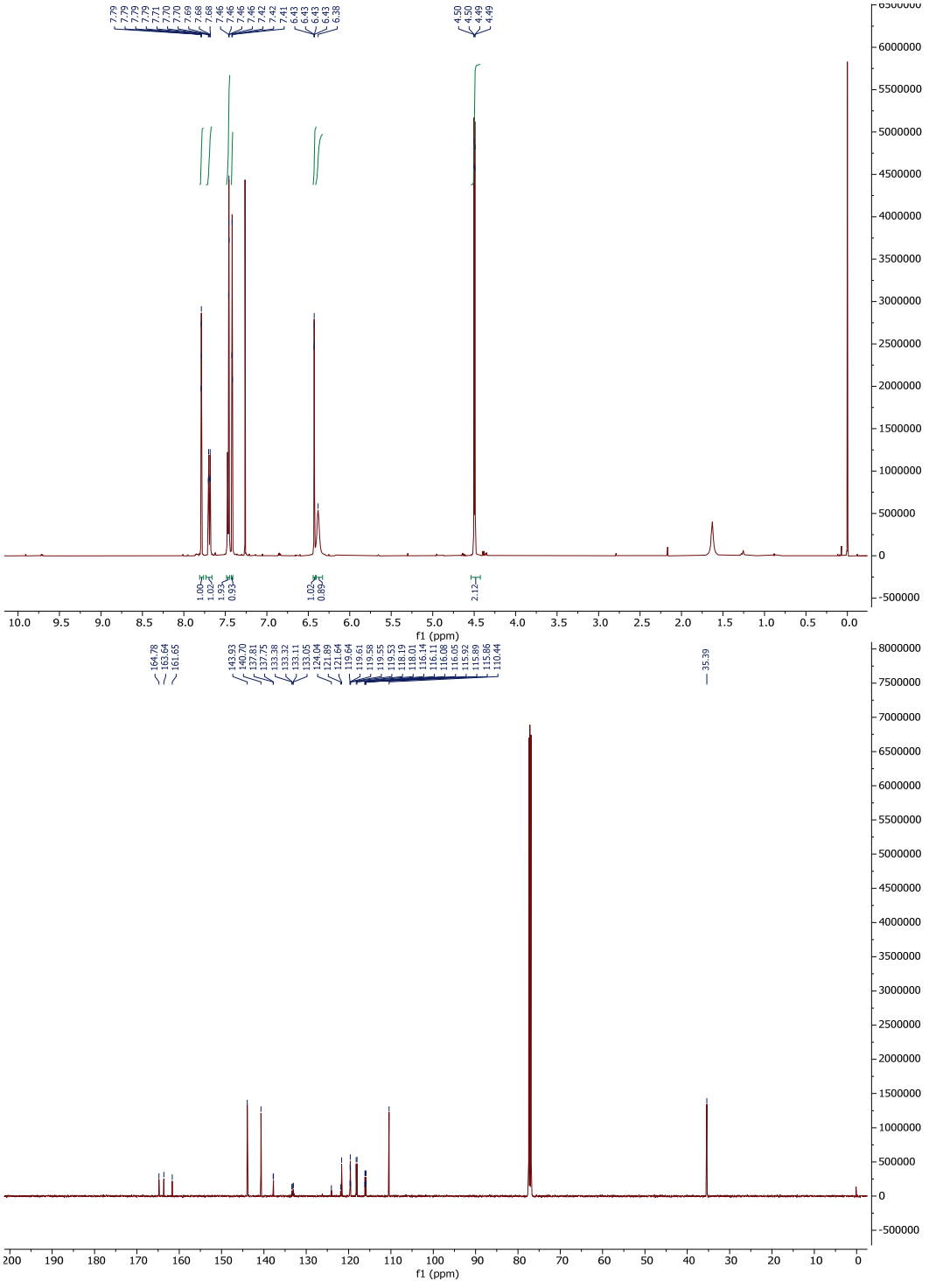
55

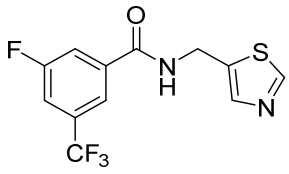

8.08
8.08
8.07
8.06
8.06
8.06
8.06
8.06
7.85
7.85
7.85
7.85
7.84
7.84
5.92
5.92
4.26
2.24
2.17
2.17
2.16
1.00
1.07
1.04
1.04
2.36
3.41
3.45
f1 (ppm)


166.54
150.90
148.26
138.65
136.54
134.00
133.67
133.34
133.01
132.21
129.12
129.09
129.05
129.01
123.76
123.73
123.69
123.65
117.82
108.07
49.00 CD3OD
36.06
35.93
13.30
11.39
f1 (ppm)

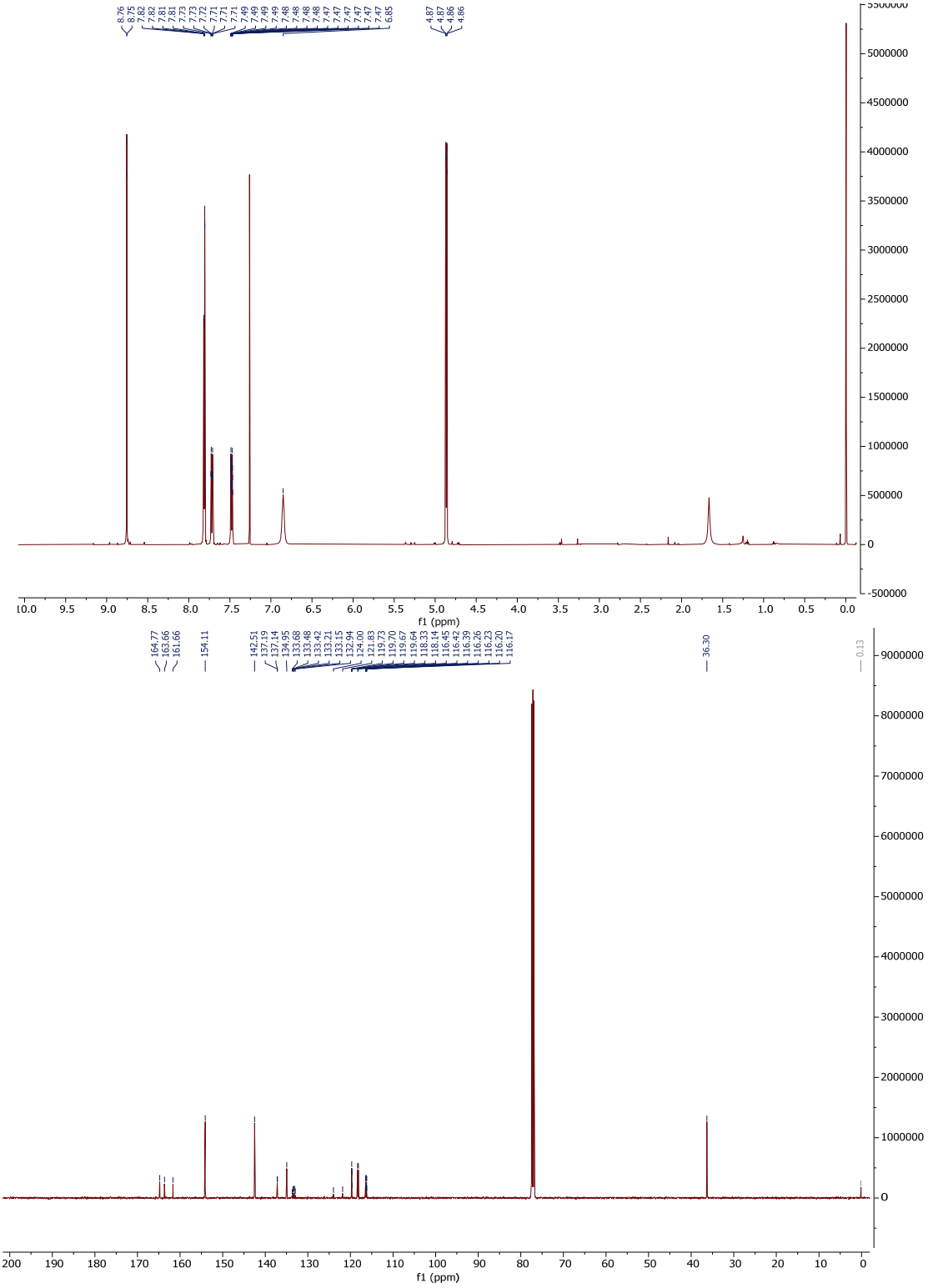

O
F
N
H
O
CF3

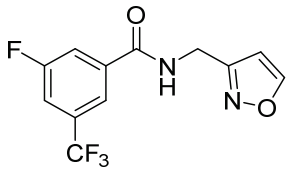
56


7.79
7.79
7.79
7.79
7.71
7.70
7.69
7.68
7.66
7.46
7.46
7.46
7.46
7.42
7.42
7.41
6.43
6.43
6.43
6.38
4.50
4.50
4.49
4.49
1.00
1.02
1.93
0.93
1.02
0.89
2.12
f1 (ppm)
164.78
163.64
161.65
143.93
140.70
137.81
137.75
133.38
133.32
133.11
133.05
124.04
121.89
121.64
119.64
119.61
119.58
119.55
119.53
118.19
118.01
116.14
116.11
116.08
116.05
115.92
115.89
115.86
110.44
35.39
f1 (ppm)

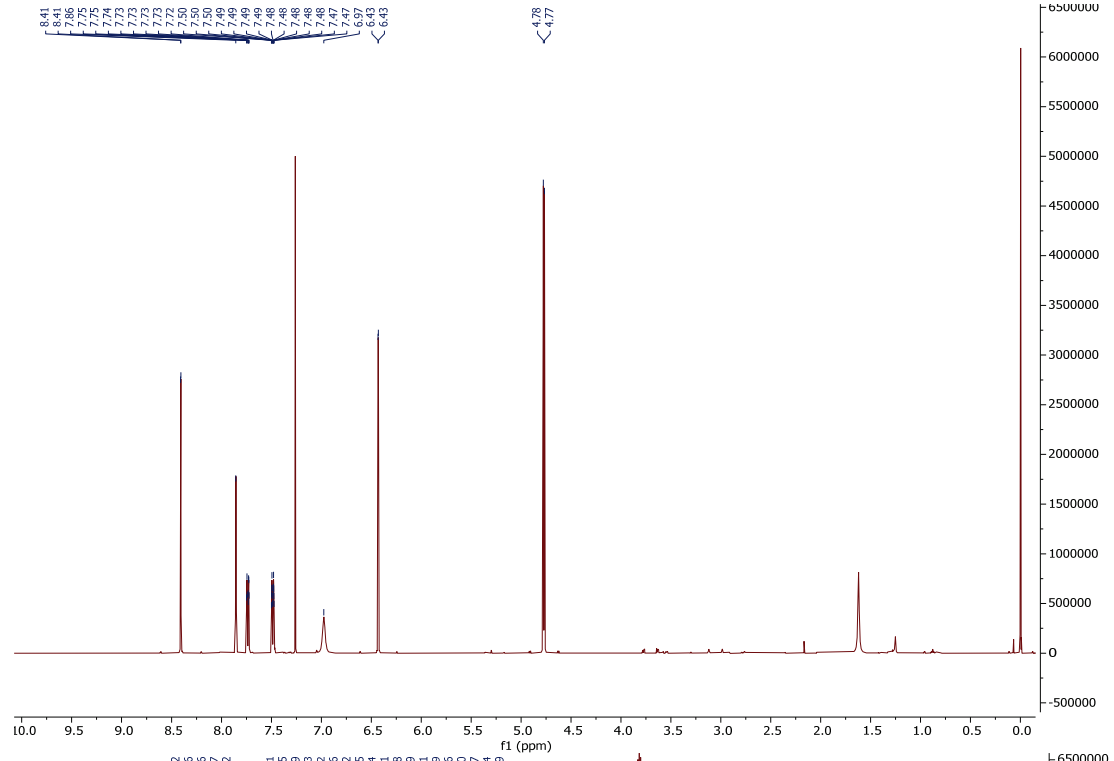

O
F
N
H
S
N
CF3

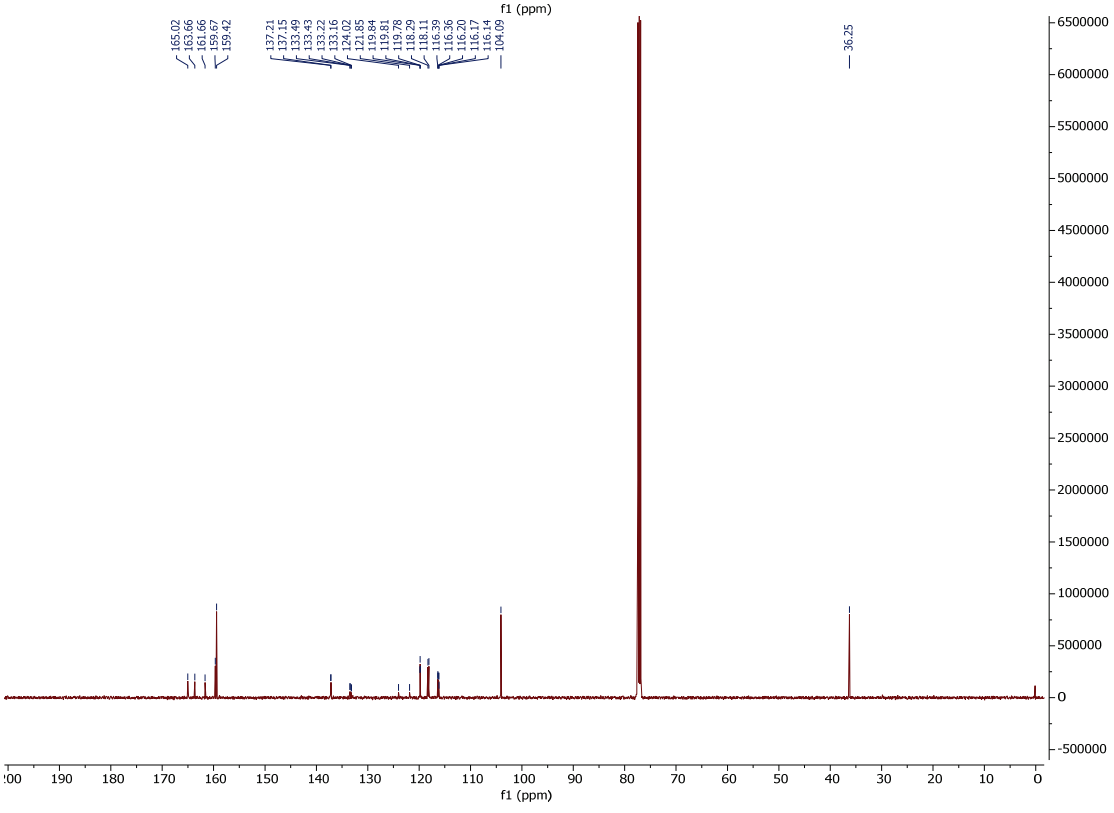
57

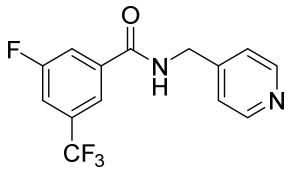

8.76
8.75
7.82
7.82
7.81
7.81
7.73
7.73
7.72
7.71
7.71
7.71
7.71
7.49
7.49
7.49
7.48
7.48
7.48
7.48
7.47
7.47
7.47
7.47
7.47
6.85
4.87
4.87
4.86
4.86
f1 (ppm)
164.77
163.66
161.66
154.11
142.51
137.19
137.14
134.95
133.68
133.48
133.42
133.21
133.15
132.94
124.00
121.83
119.73
119.70
119.67
119.64
118.33
118.14
116.45
116.42
116.39
116.26
116.23
116.20
116.17
36.30
0.13
f1 (ppm)


O
F
N
H
N
O
CF3

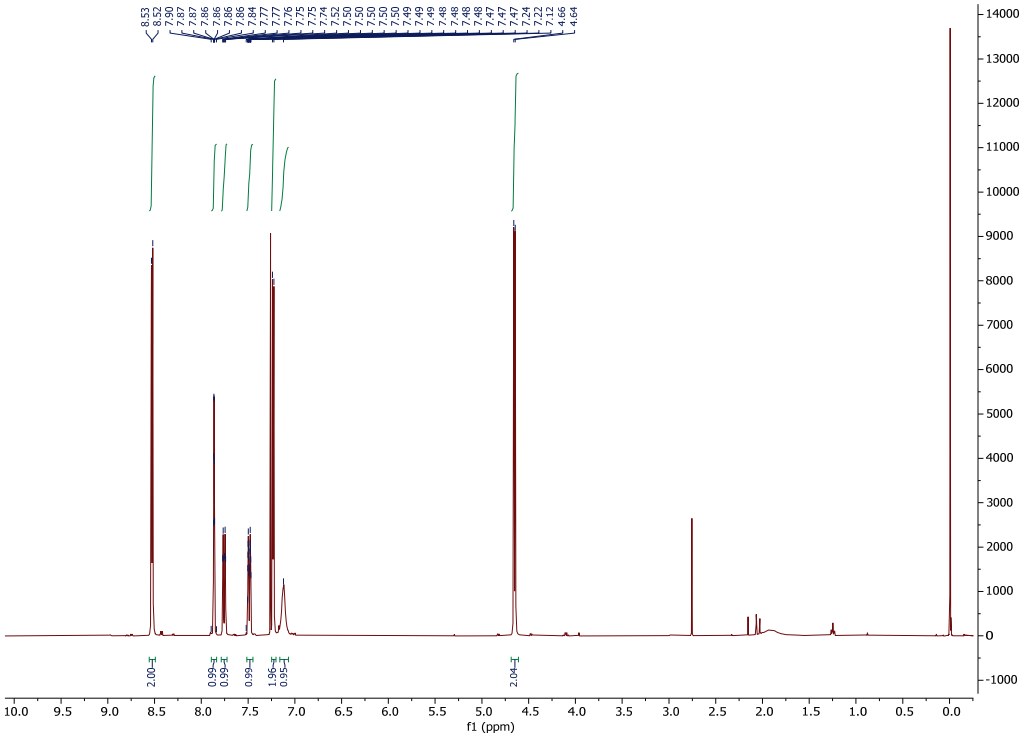
58

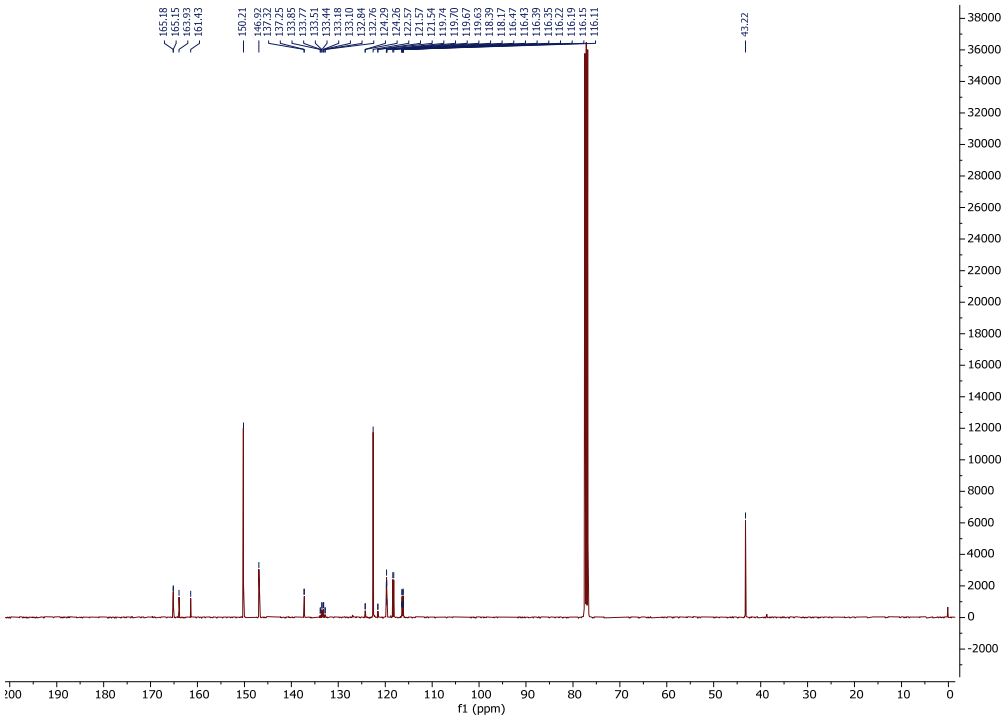

8.41
8.41
7.86
7.75
7.75
7.74
7.73
7.73
7.73
7.73
7.72
7.50
7.50
7.50
7.49
7.49
7.49
7.49
7.48
7.48
7.48
7.48
7.48
7.47
7.47
6.97
6.43
6.43
4.78
4.77
6500000
6000000
5500000
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
-500000
10.0
9.5
9.0
8.5
8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
f1 (ppm)

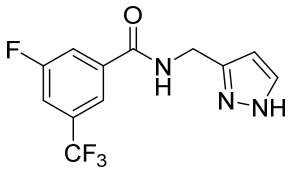

165.02
163.66
161.66
159.67
159.42
137.21
137.15
133.49
133.43
133.22
133.16
124.02
121.85
119.84
119.81
119.78
118.29
118.11
116.39
116.36
116.20
116.17
116.14
104.09
36.25
6500000
6000000
5500000
5000000
4500000
4000000
3500000
3000000
2500000
2000000
1500000
1000000
500000
0
-500000
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
f1 (ppm)



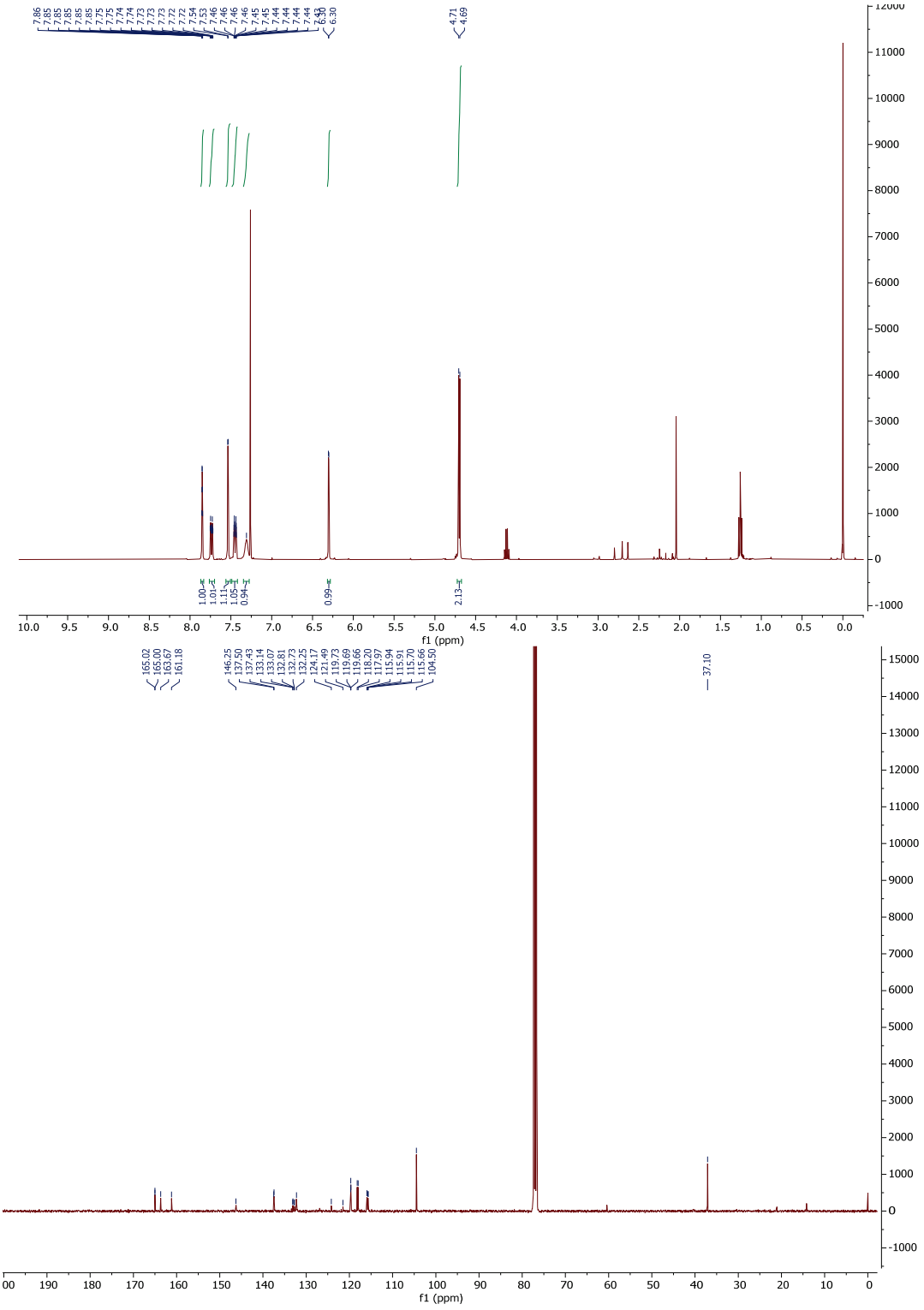
59

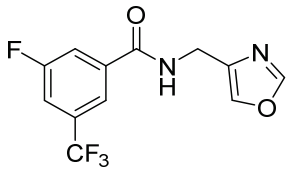




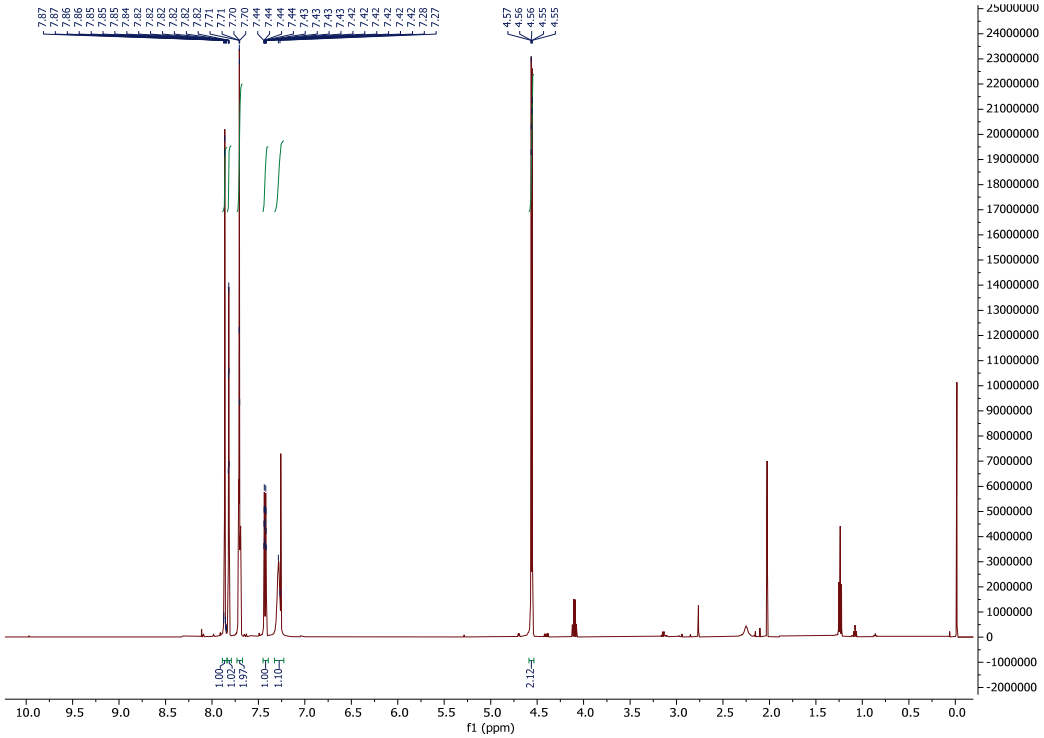

O
F
N
H
N
NH
CF3

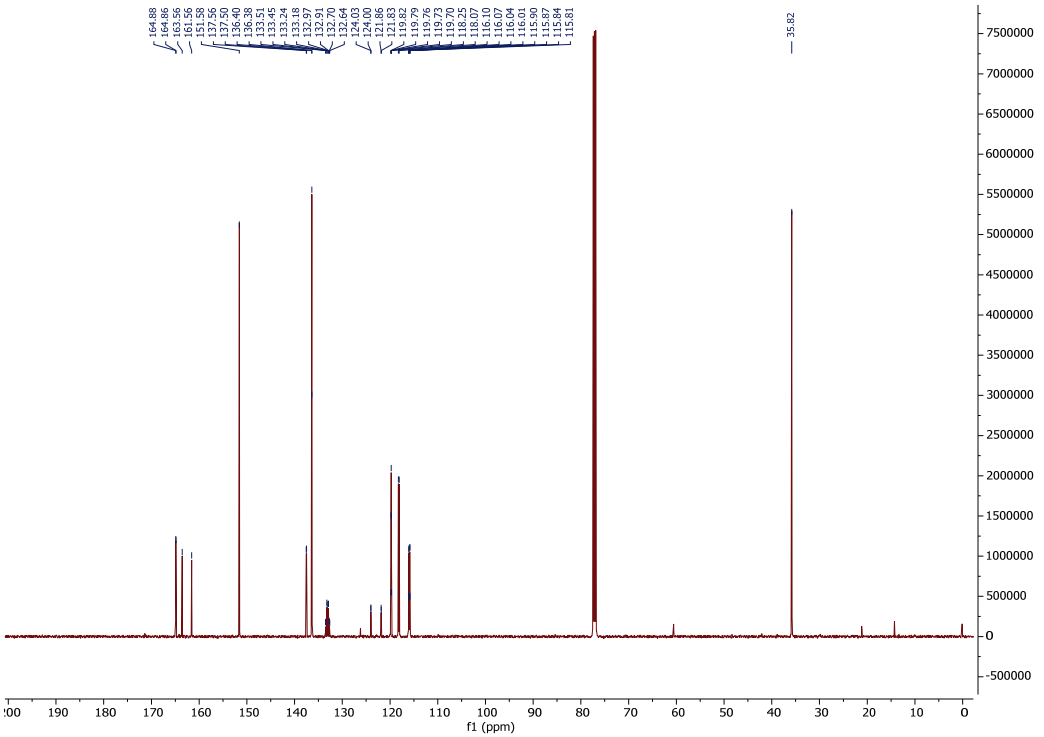
60

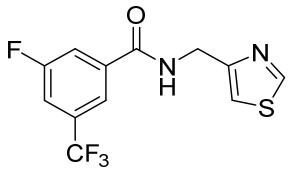

7.86
7.85
7.85
7.85
7.85
7.75
7.75
7.74
7.74
7.73
7.73
7.73
7.72
7.72
7.54
7.53
7.46
7.46
7.46
7.45
7.45
7.44
7.44
7.44
7.44
6.30
6.30
4.71
4.69
1.00
1.01
1.11
1.05
0.94
0.99
2.13
f1 (ppm)
165.02
165.00
163.67
161.18
146.25
137.50
137.43
133.14
133.07
132.81
132.73
132.25
124.17
121.49
119.73
119.69
119.66
118.20
117.97
115.94
115.91
115.70
115.66
104.50
37.10
f1 (ppm)



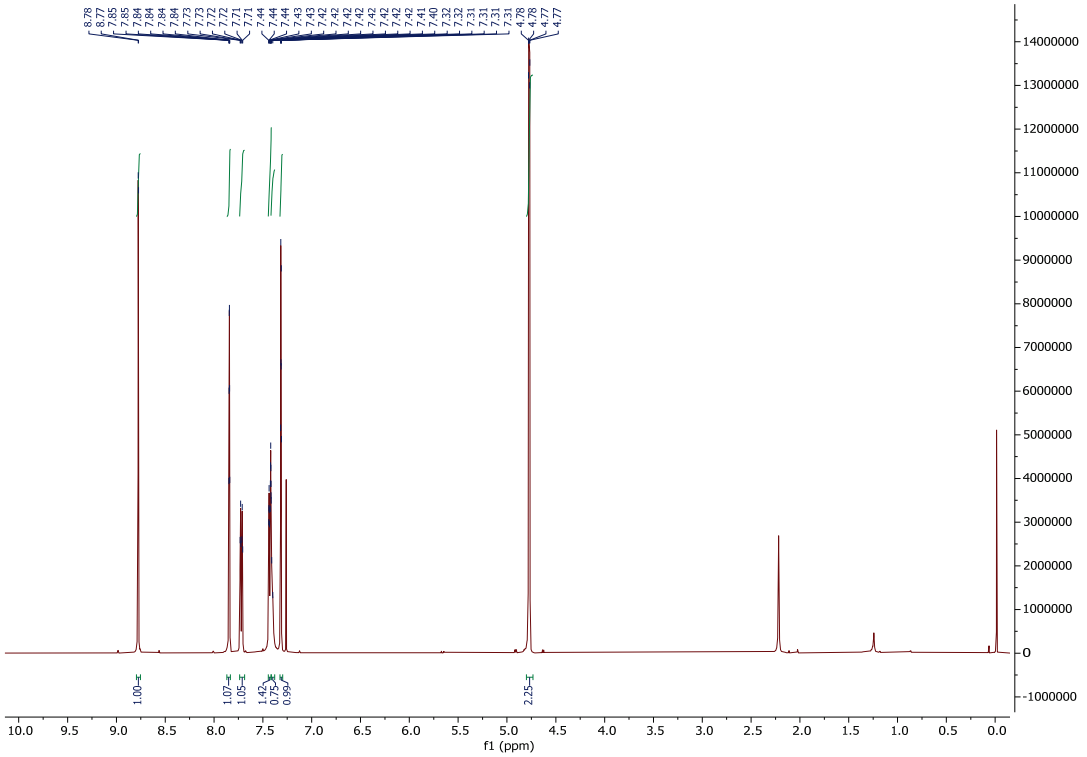
61

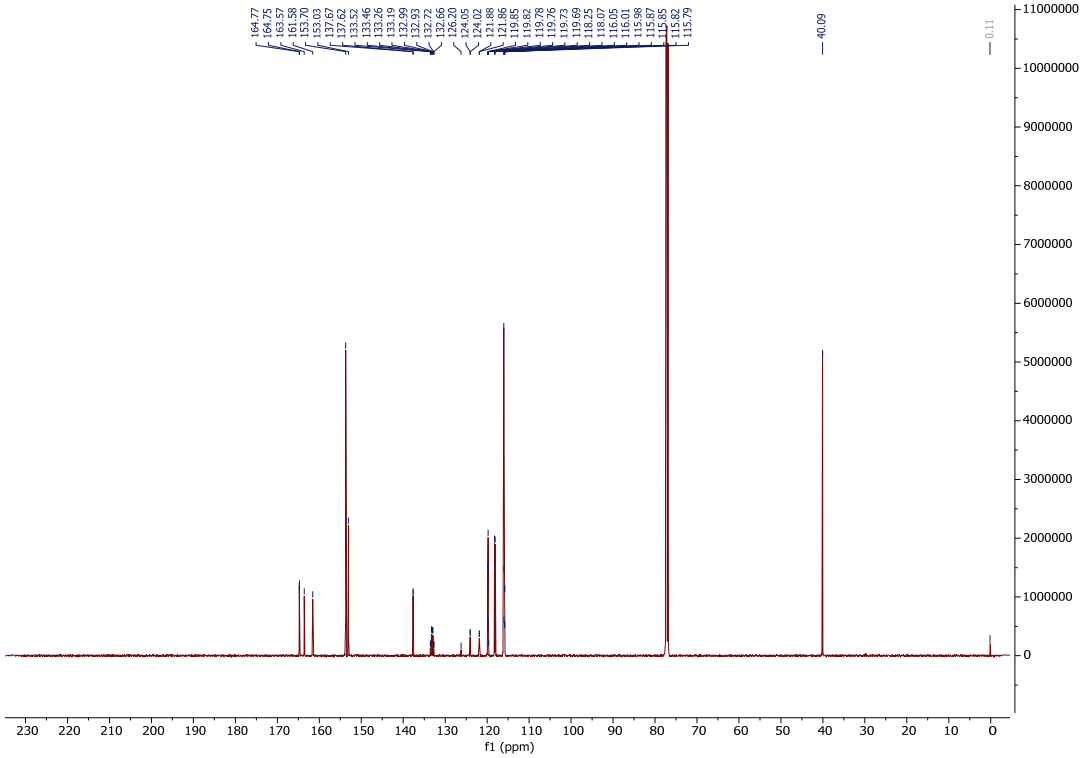


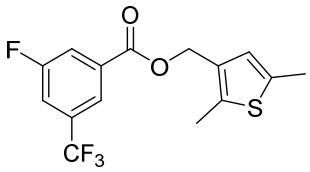




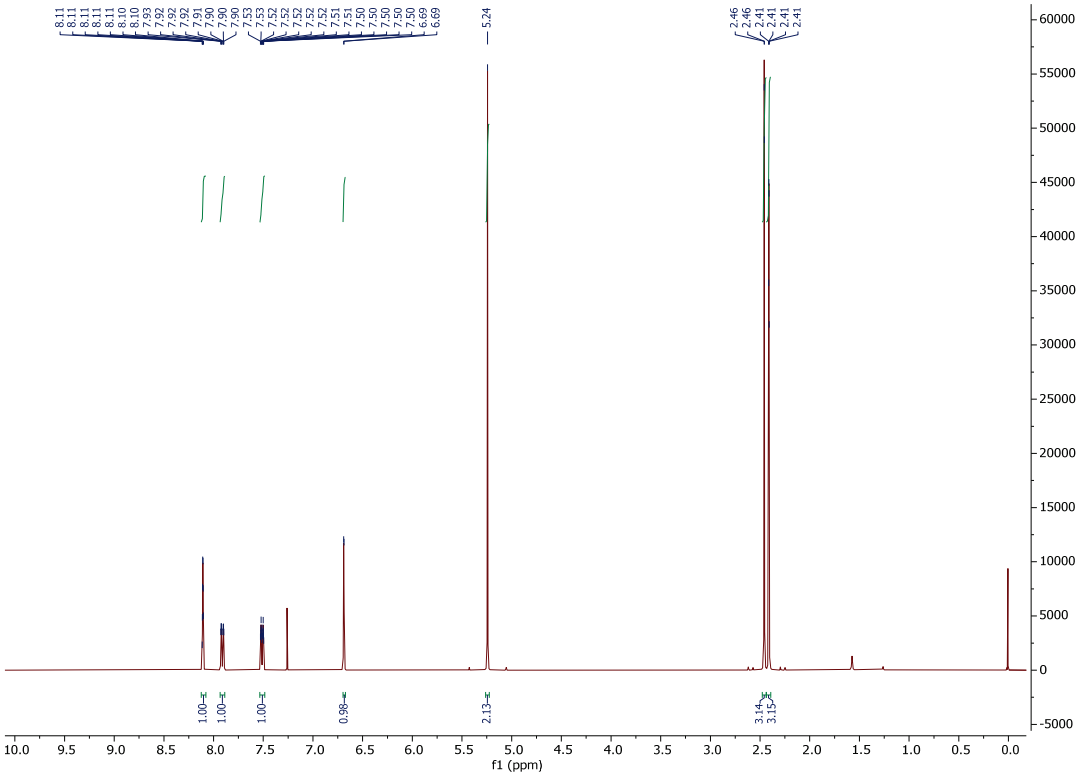
62

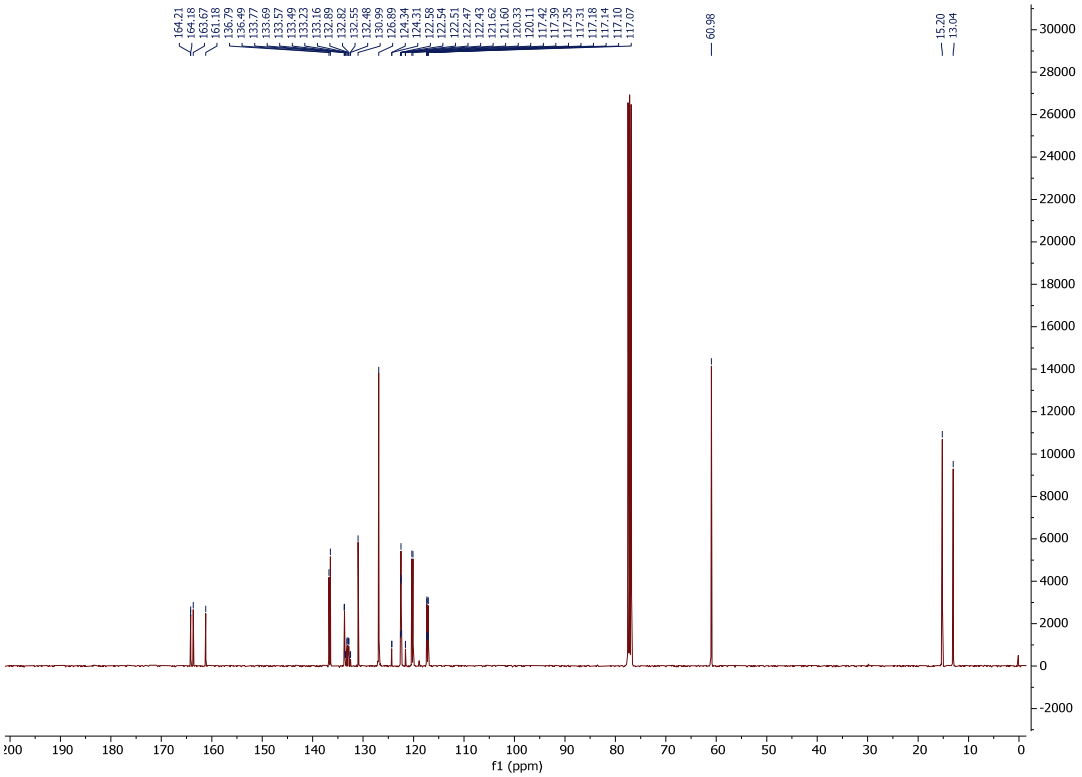


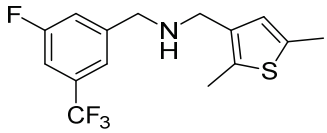



O
F
O
S
CF3

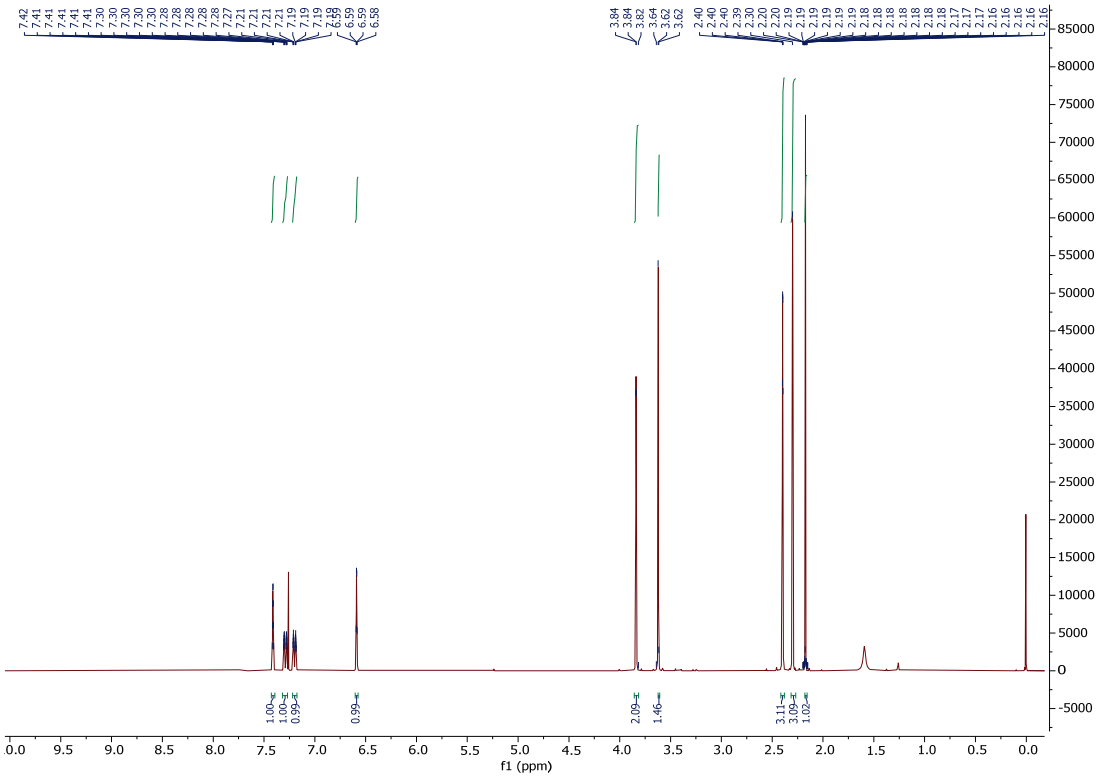
63

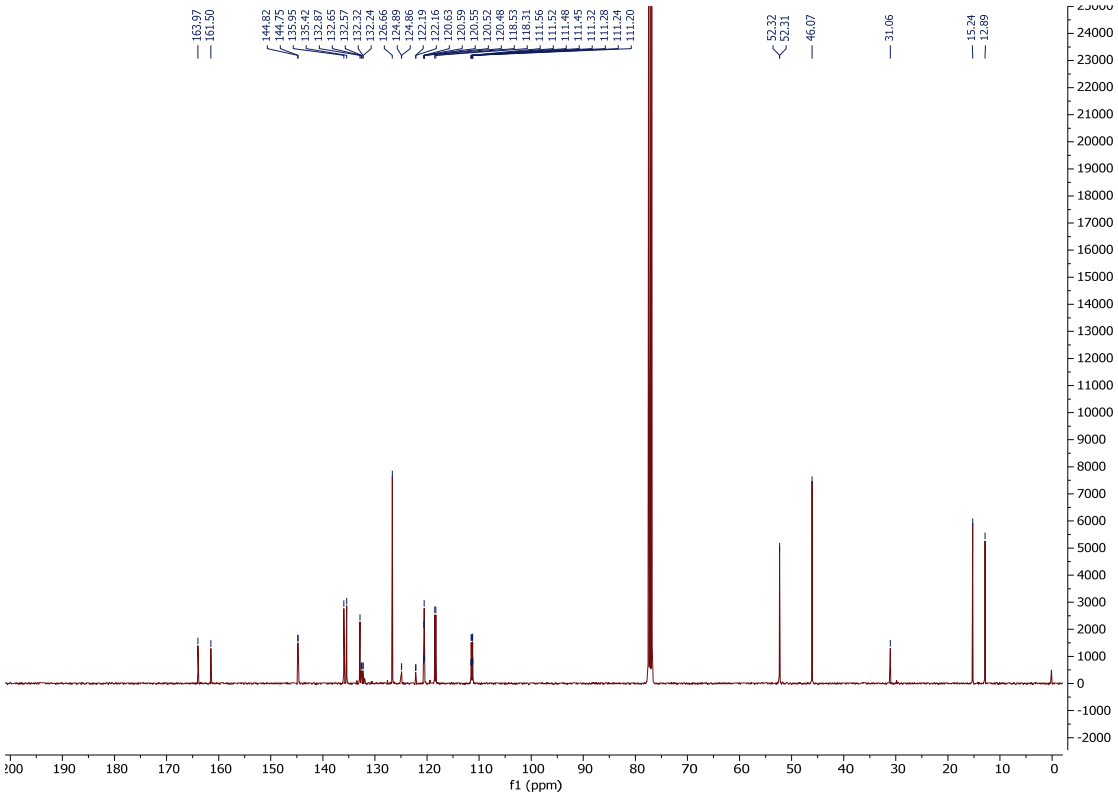

8.11
8.10
7.93
7.92
7.91
7.90
7.53
7.52
7.51
7.50
6.69
5.24
2.46
2.41
1.00
1.00
1.00
0.98
2.13
3.14
3.15
f1 (ppm)

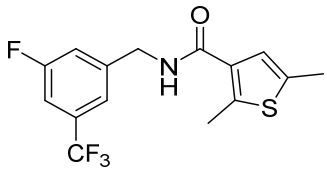

164.21
164.18
163.67
161.18
136.79
136.49
133.77
133.67
133.57
133.49
133.23
133.16
132.89
132.82
132.55
132.48
132.09
126.89
124.34
124.31
122.58
122.54
122.51
122.47
122.43
121.62
121.60
120.33
120.11
117.42
117.39
117.35
117.31
117.18
117.14
117.10
117.07
60.98
15.20
13.64
f1 (ppm)



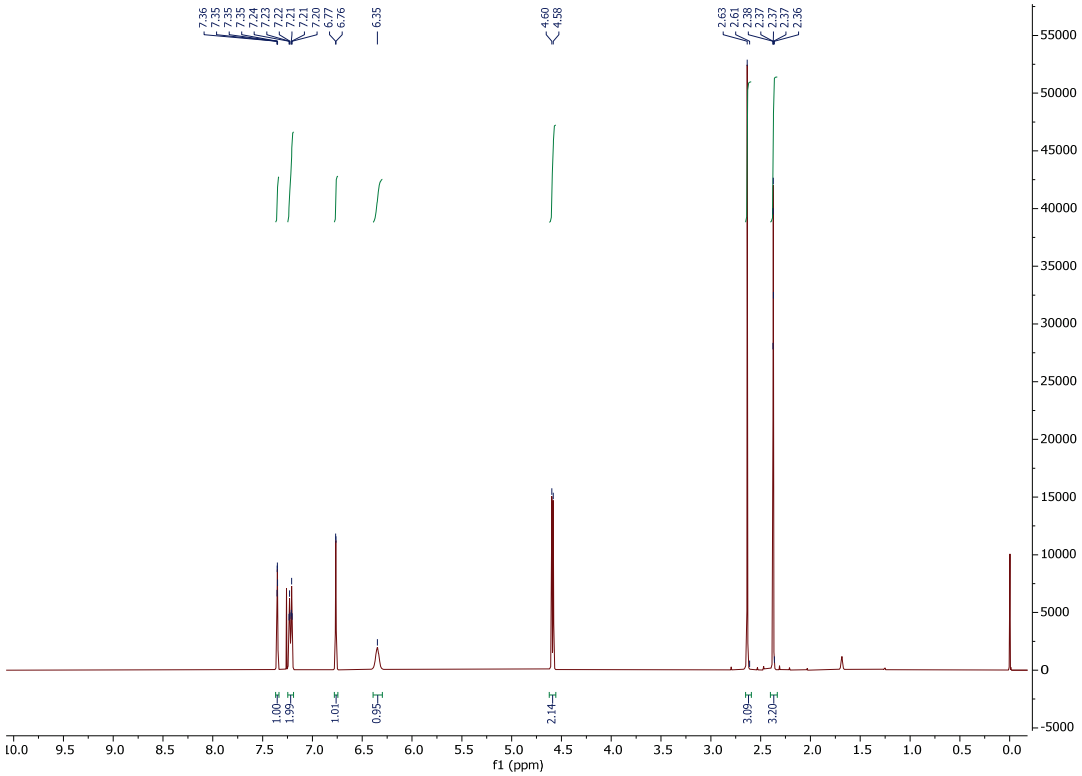
64

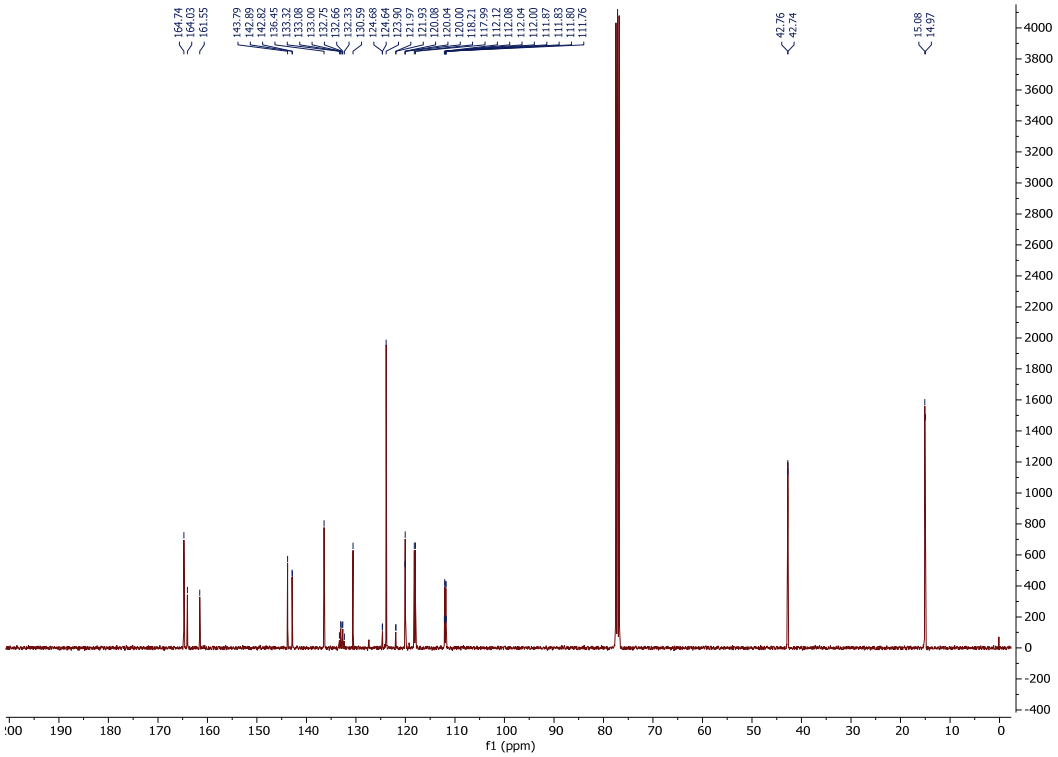





F
O
N
H
S
CF3

65


7.36
7.35
7.35
7.35
7.24
7.23
7.22
7.21
7.21
7.20
6.77
6.76
6.35
4.60
4.58
2.63
2.61
2.38
2.37
2.37
2.37
2.36
1.00
1.99
1.01
0.95
2.14
3.09
3.20
f1 (ppm)


164.74
164.03
161.55
143.79
142.89
142.82
136.45
133.32
133.08
133.00
132.75
132.66
130.53
130.50
124.68
124.64
123.90
121.97
121.93
120.08
120.04
118.21
117.99
112.12
112.08
112.04
112.00
111.87
111.83
111.80
111.76
42.76
42.74
15.08
14.97
f1 (ppm)

Compound	Molecular Formula	Calculated			Found		
		C	H	N	C	H	N
22	$C_{15}H_{13}F_4NOS$	54.38	3.95	4.23	54.56	4.13	4.13
23 (VF-57a)	$C_{15}H_{13}ClF_3NOS$	51.80	3.77	4.03	51.90	3.98	3.97
24	$C_{15}H_{13}BrF_3NOS$	45.93	3.34	3.57	45.98	3.52	3.53
25	$C_{16}H_{13}F_6NOS$	50.40	3.44	3.67	50.58	3.67	3.56
26	$C_{14}H_{13}F_2NOS \cdot 1H_2O$	56.18	5.05	4.68	56.27	4.65	4.56
27	$C_{14}H_{13}Cl_2NOS \cdot 1.25H_2O$	49.93	4.64	4.16	49.39	3.86	3.91
28	$C_{14}H_{13}Br_2NOS$	41.71	3.25	3.47	42.10	3.17	3.37
29	$C_{15}H_{16}ClNOS$	61.32	5.49	4.77	61.34	5.58	4.62
30	$C_{14}H_{14}FNOS$	63.86	5.36	5.32	63.87	5.46	5.22
31	$C_{15}H_{14}F_3NOS$	57.50	4.50	4.47	57.78	4.75	4.33
32	$C_{14}H_{14}ClNOS$	60.10	5.04	5.01	60.32	4.77	4.90
33	$C_{14}H_{14}BrNOS \cdot 2H_2O$	46.68	5.04	3.89	47.10	3.88	3.77
34	$C_{14}H_{14}N_2O_3S$	57.92	4.86	9.65	58.20	5.08	9.48
35	$C_{15}H_{12}F_5NOS \cdot 0.2\ C_3H_7NO$	51.57	3.46	4.01	51.60	3.42	4.85
36	$C_{15}H_{11}F_6NOS$	49.05	3.02	3.81	49.21	3.14	3.74
37	$C_{15}H_{12}ClF_4NOS$	49.25	3.31	3.83	49.40	3.21	3.66
38	$C_{15}H_{12}ClF_4NOS$	49.25	3.31	3.83	49.22	3.29	3.72
39	$C_{15}H_{12}Cl_2F_3NOS$	47.13	3.16	3.66	47.01	3.10	3.57
40	$C_{14}H_{15}NOS$	68.54	6.16	5.71	68.52	6.10	5.60
41	$C_{14}H_{14}ClNOS$	60.10	5.04	5.01	60.40	5.00	4.96
42	$C_{15}H_{17}NOS$	69.46	6.61	5.40	69.24	6.55	5.28
43	$C_{15}H_{17}NO_2S$	65.43	6.22	5.09	65.27	6.25	4.93
44	$C_{14}H_{13}Cl_2NOS$	53.51	4.17	4.46	53.61	4.08	4.35
45	$C_{14}H_{11}F_4NOS$	52.99	3.49	4.41	53.14	3.58	4.33
46	$C_{14}H_{11}F_4NOS$	53.00	3.49	4.41	53.05	3.41	4.34
47	$C_{13}H_9F_4NOS$	51.49	2.99	4.62	51.69	2.86	4.50
48	$C_{14}H_{11}ClF_3NOS$	50.38	3.32	4.20	50.50	3.26	4.03
49	$C_{14}H_{11}ClF_3NOS$	50.38	3.32	4.20	50.36	3.20	4.17
50	$C_{13}H_9ClF_3NOS$	48.84	2.84	4.38	49.08	2.83	4.16
51	$C_{14}H_{10}F_5NOS$	50.15	3.01	4.18	50.06	3.00	4.05
52	$C_{14}H_{10}ClF_4NOS$	47.81	2.87	3.98	47.81	2.86	3.85
53	$C_{14}H_{10}Cl_2F_3NOS$	45.67	2.74	3.80	45.68	2.67	3.74
54	$C_{15}H_{13}F_4NO_2$	57.15	4.16	4.44	57.06	4.30	4.31
55	$C_{15}H_{13}ClF_3NO_2$	54.31	3.95	4.22	54.29	4.09	4.03
56	$C_{13}H_9F_4NO_2 \cdot 0.35\ H_2O$	53.20	3.33	4.77	53.06	3.15	4.67
57	$C_{12}H_8F_4N_2OS$	47.37	2.65	9.21	47.36	2.67	9.05
58	$C_{12}H_8F_4N_2O_2$	50.01	2.80	9.72	50.16	2.64	9.55
59	$C_{14}H_{10}F_4N_2O$	56.38	3.38	9.39	55.35	3.41	9.08
60	$C_{12}H_9F_4N_3O$	50.18	3.16	14.63	50.26	3.23	14.32
61	$C_{12}H_8F_4N_2O_2$	50.01	2.80	9.72	50.24	2.67	9.50
62	$C_{12}H_8F_4N_2OS \cdot 0.3\ C_3H_6O$	47.65	2.80	9.03	47.85	2.54	9.05
63	$C_{15}H_{12}F_4O_2S$	54.22	-	3.64	54.44	-	3.54
64	$C_{15}H_{15}F_4NS$	56.77	4.76	4.41	56.85	4.82	4.63
65	$C_{15}H_{13}F_4NOS$	54.38	3.95	4.23	54.46	3.83	4.19

Table S1. Elemental analysis data.

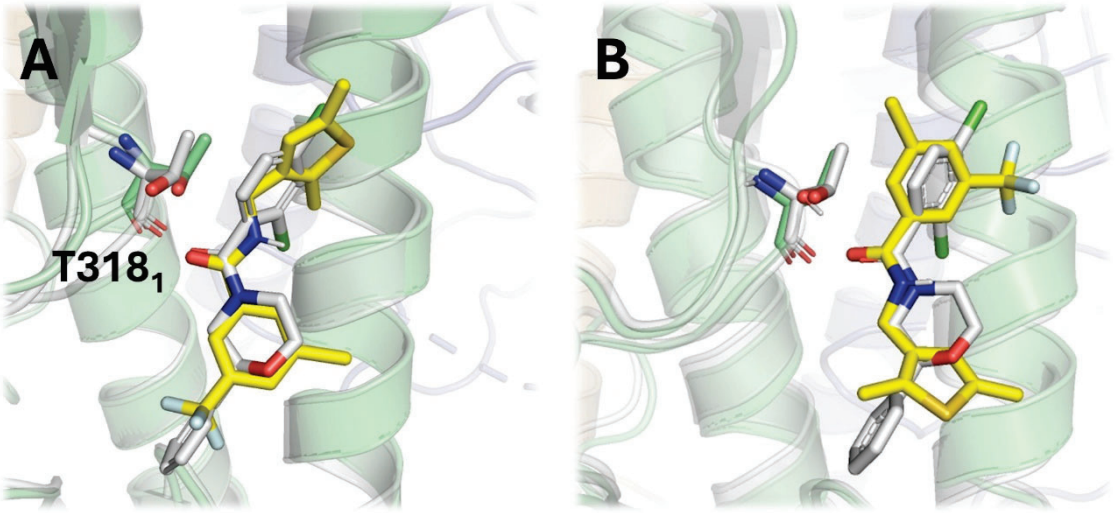


Figure S1. Representation of the two binding modes examined for VF-57a (yellow sticks) upon binding to site B in the H1 HA (A/Virginia/ATCC3/2009). The two binding modes differ in the relative position of the thiophene and phenyl rings, allowing the formation of the hydrogen bond with T318$_1$. The X-ray binding mode of (*S*)-F0045 bound to H1 HA (grey sticks).

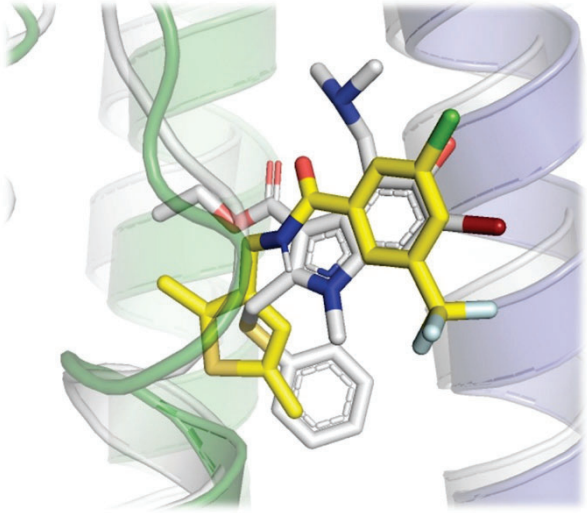


Figure S2. Representation of the binding mode examined for VF-57a upon binding to site A. Overlapped X-ray position of arbidol bound to H3 HA (A/Hong Kong/1/1968) (grey sticks) and initial position of VF-57a in our H1 HA model (A/Virginia/ATCC3/2009) (yellow) in site A binding pocket.

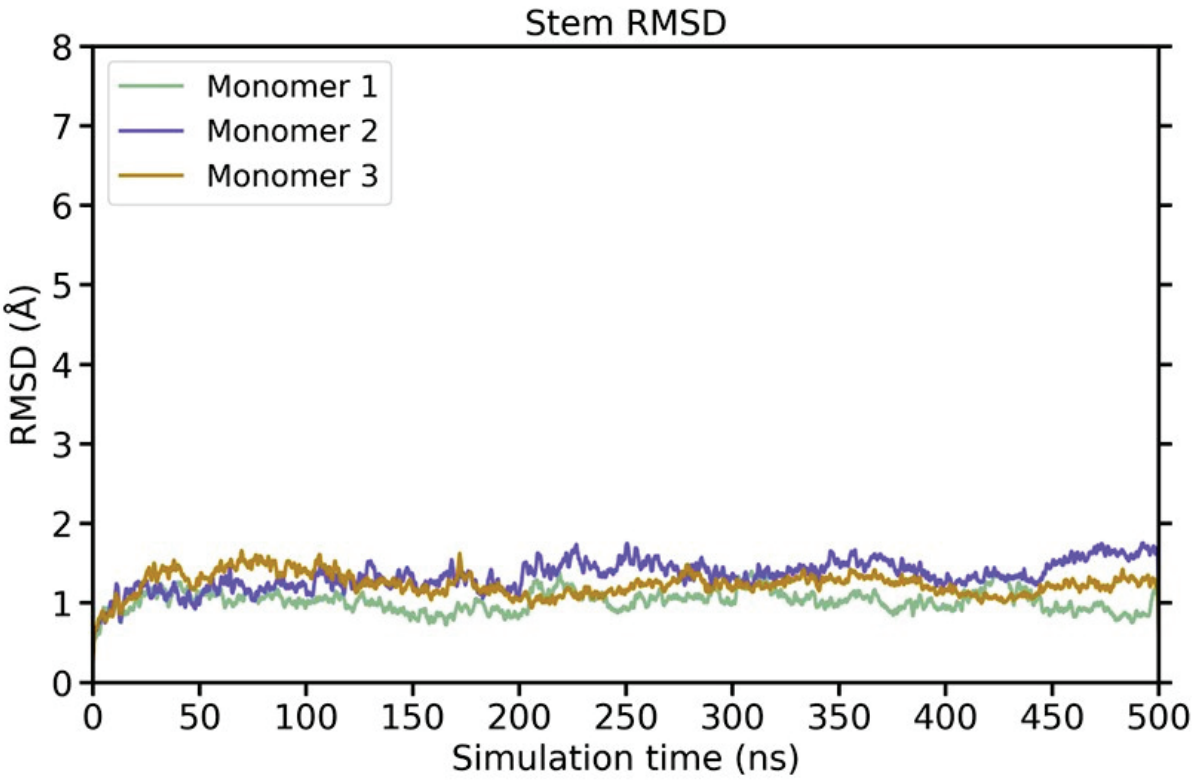


Figure S3. RMSD plot for each HA-stem backbone atoms for the simulation of VF-57a in site A binding pocket of the A/Virginia/ATCC3/2009 Influenza A strain (H1N1).



Figure S4. Sequence alignment for the studied Influenza A strains. The green line indicates the part of sequences corresponding to the HA1 subunit of HA, whereas the blue line the part of the sequence corresponding to the HA2 subunit. The cleavage point is highlighted in bold (black). The T residue involved in site B binding of (*S*)-F0045 is highlighted in violet. The proposed mutations affecting the binding mode of VF-57a in site A are highlighted in red.

Chapter 4

Discussion

"Now Garmr howls loudly before Gnipahelli,
the fetter will break, and the wolf will run free.
I know many things, and I see further ahead,
about the great fate of gods, the mighty gods.

A ship fares from the east, Muspell's forces
will come over the sea, and Loki steers.
All the giant's kindred travel with the wolf,
and with them the brother of Býleiptr goes."

Vǫlva (Vǫluspá, stanzas 47, 49)

The objectives pursued during this Doctoral Thesis focus on two main lines of research, both related to major public health problems highlighted by the WHO. Understanding the molecular interactions between the protein partners and the ligands, either substrates or inhibitors, is the basis of the proposed research questions. The first project addresses the molecular determinants of the substrate selectivity of the Niemann-Pick C1-like 1 (NPC1L1) transporter involved in the absorption of dietary cholesterol (**Project 1**, included in Chapter 3). The second and third projects aim to rationalise the binding modes and structure-activity relationships of several inhibitors of the Influenza A virus (IAV) Hemagglutinin (**Project 2** and **Project 3**, included in Chapter 3).

4.1 *"Binding of Cholesterol to the N-Terminal Domain of the NPC1L1 Transporter: Analysis of the Epimerization-Related Binding Selectivity and Loop Mutations"*

Cholesterol (CHL) homeostasis is tightly controlled by a balance between its biosynthesis and recycling by the liver and its intestinal absorption. NPC1L1 transporter is the major importer of free cholesterol in mammals,[97] but its exact mechanism of action is not completely understood.[82] It has a key role in the development of CVD disease, and thus it is exploited by small-molecules to treat hypercholesterolemia.[62,64] It is also the importer of other sterols, fatty acids and essential vitamins (vitamin K). In this project, we aimed to understand the substrate selectivity of NPC1L1.

NPC1L1 has three luminal domains called the N-terminal domain (NTD), the middle luminal domain, and the C-terminal domain, and a transmembrane domain composed of 13 transmembrane helices. The NTD selectively binds free cholesterol over other plant sterols, even in the absence of the other domains of the protein. Competitive binding assays have shown that epi-cholesterol is unable to displace labelled ^{3}H-cholesterol from the NTD, despite differing only in the absolute configuration of the hydroxyl moiety. Moreover, mutations in the binding cavity or removal of the NTD from the rest of the protein, drastically impair cholesterol uptake by NPC1L1.[108] Therefore, the role of NTD in selectively binding substrates is essential for the overall mechanism of cholesterol transport. Based on this information, I explored the molecular features behind the selective sterol binding to the isolated NTD and investigated the structural effects of single point mutations on the NTD that impair cholesterol uptake.

4.1.1 NPC1L1-NTD protein dynamics upon cholesterol binding

The comparison of the X-ray structures of the apo NTD of NPC1L1 and the NTD-CHL complex of its homologue NPC1 revealed relevant structural differences in the rearrangement of the α8/β7 loop (highlighted in blue in Figure 4.1, panels A, B and C) and the shape of the binding pocket entrance, the so-called S-opening (Figure 4.1, panel D). In NPC1L1, the S-opening is closed to the solvent by the side chain of L213 (highlighted in yellow in Figure 4.1, panels A, B and C) in the absence of a ligand. In NPC1, Q200 (highlighted in magenta in Figure 4.1, panel C) is found at the equivalent position of L213 in NPC1L1 and oriented to N86 (highlighted in magenta in Figure 4.1, panel C) in the α3 helix, thus keeping the S-opening open to the solvent and allowing the side chain of CHL to fit into this space.

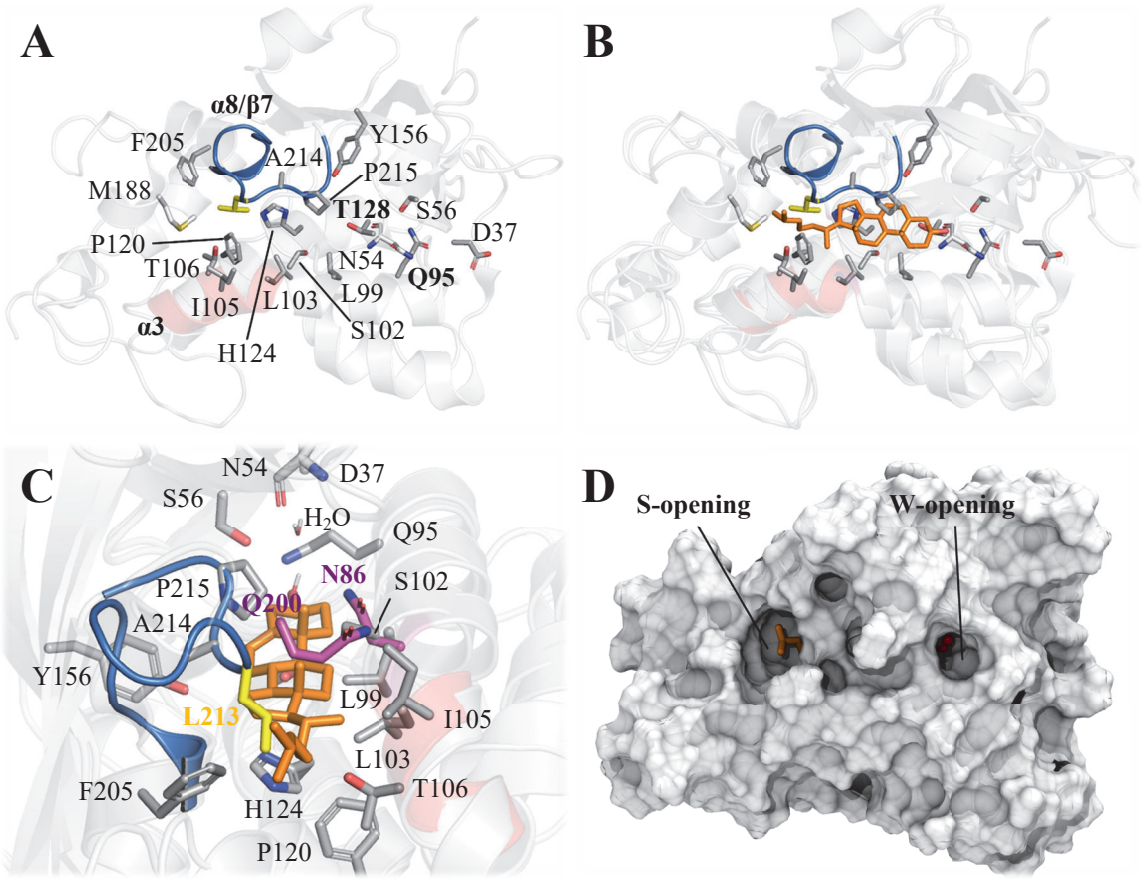


Figure 4.1 A) X-ray structure of the NTD of NPC1L1 (PDBid: 3QNT).[102] Relevant residues are highlighted in gray sticks, L213 is highlighted in yellow, α8/β7 loop in blue and the α3 helix in red; B) Superposition of the X-ray structures of the apo NTD of NPC1L1 and the NTD-CHL complex of NPC1 (PDBid: 3GKI).[107] CHL is highlighted in orange sticks; C) S-opening view for the superposition of X-ray structures. Q200 and N86 of NPC1 are highlighted in magenta sticks and the crystallographic water molecule on top of CHL's hydroxyl has been added for clarity; D) Surface representation of the NTD of NPC1L1 with CHL in the binding pocket.

To investigate how CHL molecules can fit into the binding site of the NTD of NPC1L1, I explored the dynamical behaviour of the apo NTD of NPC1L1 through MD simulations, which revealed a strong structural stability for the domain with the solvent-exposed loops exhibiting large motions. The side chain of L213 keeps the S-opening closed by forming a hydrophobic cluster interacting with the side chains of residues S102, L103, T106, F120 and F205 (Figure 4.1, panels A and C). The conformation was specifically stabilised by the interaction between the backbone of L213 and S102, which caused the α8/β7 loop to approach the α3 helix and reorient with respect to the X-ray structure, reducing the width of the S-opening. In the CHL-bound NTD (NTD-CHL) complex, L213 reoriented and mimicked the orientation of Q200 in the homologue NPC1-NTD. The aliphatic chain of CHL forced the open conformation of the S-opening by reorienting the α8/β7 loop. This behaviour was consistent with the hypothesis of Kwon et al.[102] of a gating mechanism involving L213 and multiple rearrangements at the NPC1L1-NTD S-opening to allow CHL binding. We have thus confirmed and provided the molecular features behind this gating mechanism, attributing it to the dynamic flexibility of the α8/β7 loop. Moreover, the binding stability of CHL was attributed to h-bond interactions between the CHL hydroxyl with the amide NH of Q95, the hydroxyl group of S56 and a water-mediated interaction with the main chain of N54 (Figure 4.1, panel C). The water molecule mediating the latter interaction, at the top of the CHL hydroxyl group, occupies the same local region as a co-crystallised water molecule in the homologue NPC1, suggesting a key role in the binding stability of sterols further discussed below.

4.1.2 Molecular features driving the sterol selectivity of the NPC1L1-NTD

According to experimental competitive binding assays performed by Kwon et al.[102] and Zhang et al.,[108] the NTD of NPC1L1 can selectively bind other sterols of animal origin such as epi-cholesterol (ECL), 25-hydroxycholesterol (HHL) and lanosterol (LNL). Among these, ECL was found to be the less effective compound in displacing CHL from the NTD, demonstrating its inability to compete. Through MD simulations of the NTD-ligand complexes for these compounds, we found that the overall behaviour of HHL and LNL complexes were comparable to that of the NTD-CHL system. The chemical differences of these sterols did not alter the characteristics of the binding mode observed for CHL. Furthermore, MMPBSA calculations revealed similar binding affinities for CHL, HHL and LNL, in agreement with the experimental evidence of binding competition.

MD simulations for the NTD-ECL complex revealed two distinct binding modes for ECL. First, the *Q95 binding mode* (Figure 4.2, panel A), was structurally close to the binding mode of CHL. A stable h-bond with Q95 was confirmed after a slight shift of ECL towards the W-opening, attributed to the α-configuration of the hydroxyl moiety. However, the new position of ECL displaced the structural water molecule present in the NTD-CHL simulations and in the X-ray structure of the homologue NPC1-NTD. The loss of the structural water molecule led to a decrease in the stability of the binding mode, as confirmed by relative binding free energy (RBFE) calculations, which showed that CHL binds more favourably to the *Q95 binding mode* than ECL by 1.8 kcal·mol^{-1}. The α-configuration of the hydroxyl group in ECL disrupts the ability to establish an effective and stable h-bond interaction with Q95, promoting the emergence of the *T128 binding mode* (Figure 4.2, panel B) and the displacement of the aforementioned structural water molecule. The orientation of the α8/β7 loop was also changed compared to the *Q95 binding mode* and L213 was reoriented towards F205 to avoid steric clashes with the hydrophobic ECL tail, thus slightly increasing the width of the S-opening.

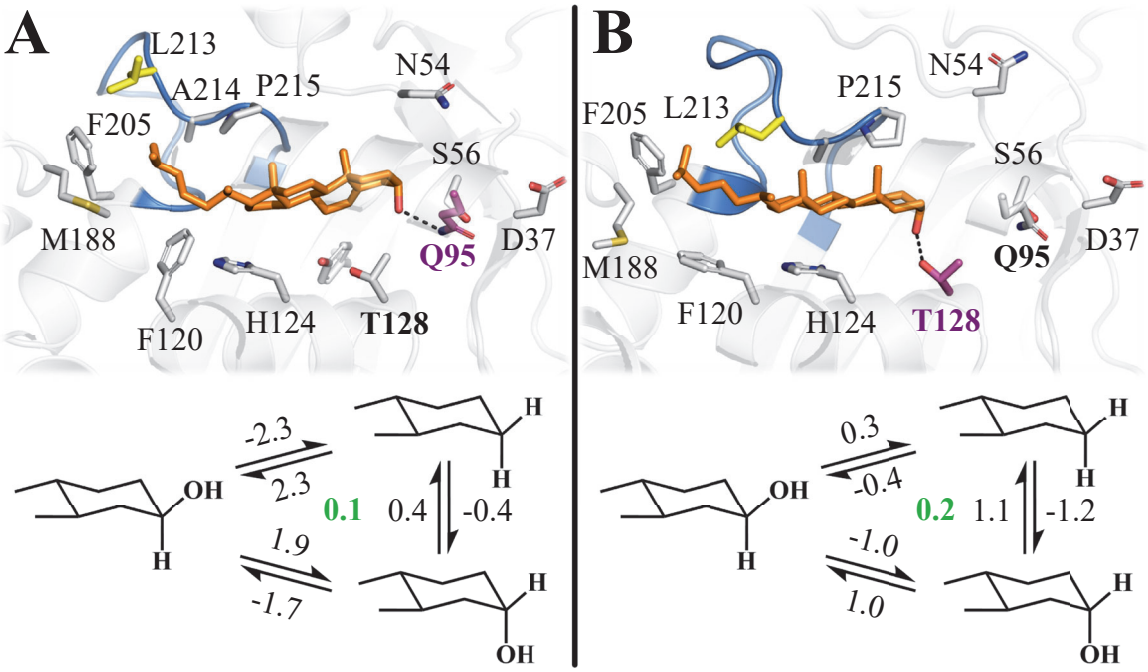


Figure 4.2 A) Representative snapshot of the *Q95 binding mode* for ECL. Q95 is highlighted in magenta. B) Representative snapshot of the *T128 binding mode* for ECL. T128 is highlighted in magenta; In both representations the α8/β7 loop is shown in blue, L213 is highlighted in yellow and the h-bond interaction between ECL and Q95/T128 is shown as a dashed line. Below each panel, the thermodynamic cycle considered for the RBFE calculations. The final averaged relative free energy values for each transformation are shown in kcal·mol^{-1} and the cycle closure error is shown in green.

Considering the h-bond formation/breakage between ECL and Q95, Umbrella Sampling simulations showed that the *Q95 binding mode* is destabilised with respect to the *T128 binding mode*. The energetic barrier for the T128→Q95 binding mode transition was measured to be approximately 1.1 kcal·mol^{-1} higher than that for the Q95→T128 binding mode transition. The energetic estimates for the binding mode transition were refined by Well-Tempered Metadynamics simulations, increasing the barrier difference up to 2.0 kcal·mol^{-1} when the conformational rearrangements of the α8/β7 loop were added as a second CV. This indicates a coupling between the two h-bond states of ECL and the conformational flexibility of the α8/β7 loop, revealing a dynamic response of the NTD of NPC1L1 to the epimerisation of CHL that is directly reflected in the selectivity mechanism.

Overall, the results provided a rationale for the experimentally observed binding stability of CHL, HHL and LNL and the molecular features behind the inability of ECL to compete with CHL. These results indicate that the absolute configuration of the hydroxyl moiety in sterols and the structural water molecule are key molecular features for the sterol selectivity mechanism of the NTD of NPC1L1.

4.1.3 Mutations impairing cholesterol transport alter the NTD dynamics

To understand the effect on the sterol affinity of single point mutations that reduce the CHL absorption of the NTD of NPC1L1, four mutations that led to a reduction ranging from 35% to 85% were selected: I105A (35%), P215A (52%), F205A (85%) and T128A (85%).[108] MD simulations of the apo mutated proteins and mutant-CHL (holo) complexes revealed comparable dynamic behaviours as the wild-type systems. MMPBSA calculations revealed a significant decrease of 5.1 kcal·mol^{-1} in the CHL binding affinity with respect the wild-type only for the P215A mutant, which was partly attributed to the loss of the P215 side chain contribution of -1.6 ± 0.3 kcal·mol^{-1} to the hydrophobic stabilisation of the CHL binding mode. This suggests that factors other than the binding affinity of CHL may influence the binding to the mutanted variants of the NPC1L1-NTD. To investigate the effect of α8/β7 loop dynamics on sterol affinity, a principal component analysis (PCA) of the mutant systems was compared with the wild-type systems in both apo and holo states.

For the I105A mutant, the overall dynamic arrangement of the α8/β7 loop was similar to that of the wild-type NTD for both apo and holo states, in qualitative agreement with our binding affinity estimates. Minor conformational differences favoured the appearance of a

distribution of conformations not sampled for the wild-type, resulting in a weaker distortion of the hydrophobic stabilisation near the S-opening. Overall, this agrees with the mild reduction in CHL absorption determined for this mutation.

The P215A mutant showed dynamic differences in the α8/β7 loop conformations in the apo state compared to the wild-type NTD, resulting in occluding rearrangements that reduce the cavity volume. On the other hand, no significant differences were observed for the holo species. This suggests that the P215A mutation impairs cholesterol absorption by distorting the dynamic equilibrium of the α8/β7 loop in the apo state, leading to a higher occlusion of the S-opening.

Regarding the F205A mutant, the apo form displayed a larger exploration of the conformational space of the α8/β7 loop compared to the wild type, as reflected in two substantially populated novel conformational states. As in the P215A mutant, this effect was also less pronounced in the holo state. Considering its residual contribution to the binding affinity, as predicted by MMPBSA calculations, we concluded that F205 plays a key role in stabilising the α8/β7 loop of the apo state of the NTD. Its mutation to alanine contributes to the disruption of the dynamic equilibrium, especially at the S-opening.

Finally, the T128A mutant was found to have little effect on the overall motions of the S-opening, especially for the apo state. The most relevant difference resulting from the population of a novel configuration state of the α8/β7 loop was attributed to a change in the cavity hydration pattern. This resulted in the loss of a water molecule located between L213 and S102, which bridges the interaction between L213 and H124 in the wild type apoprotein. On the other hand, the configuration of the α8/β7 loop in the holo state was similar to the wild-type NTD-CHL system.

Overall, the studied mutations revealed a small effect on the binding mode of CHL, except for the P215A mutant. On the other hand, we determined the structural impact of mutations on the conformational landscape in the apo state of the NTD, prior to the binding of CHL. The changes triggered by the mutations highlight the relevant impact reported in other enzyme regarding both the catalytic activity and the ability to accommodate other substrates.[184]

4.2 *"Mining Druggable Sites in Influenza A Hemagglutinin: Binding of the Pinanamine Based Inhibitor M090"*

The emergence of resistant strains and the high rate of immune adaptation of IAV has encouraged the development of small molecule antiviral treatments in recent decades. In this context, the work of Zhao et al.[183] showed that pinanamine derivatives are attractive candidates as anti-IAV compounds targeting HA showing sub-micromolar activity against both HA phylogenetic groups. These compounds were found to inhibit the HA-mediated fusion, hypothetically by stabilising the pre-fusion conformation and blocking the pH-induced structural rearrangements of HA in the endosome, thus preventing the release of the fusion peptide. The experimental assays performed by Zhao et al.[183] showed that the antiviral activity of M090, the most potent compound against all the tested strains of the virus, was independent of blocking the M2 proton channel, the original target of the study.

To explore the mechanism of action (MOA) of M090, the antiviral assays showed that the inhibitor acts during the first 2h of the viral cycle. This corresponds to the initial stages of IAV infection which are primarily associated with the M2 channel and HA and include the host cell recognition phase, the endocytosis process and membrane fusion.[160] As M090 was found to be independent of M2 channel blockage, the authors hypothesised that the target for M090 and its derivatives was HA. Resistance selection experiments revealed the occurrence of the $E74_2 \rightarrow D$ mutation in the presence of M090, located in the HA_2 domain of HA. With this information, the authors suggested a putative binding site for M090 based on molecular docking predictions.

In the absence of further mechanistic and structural information, I investigated the molecular features behind the MOA of M090 and provided an *in silico* validation of its binding mode by comparing the experimental SAR data for M090 and accurate binding affinity predictions.

4.2.1 M090 binding site and binding mode

The putative binding site of M090 proposed by Zhao et al.[183] was predicted to be located at the interface defined by the α-helix $L80_2$-$A96_2$ (H-BS) and by loops $P301_1$-$K318_1$ and $M59_2$-$L73_2$ (L1-BS and L2-BS, respectively) (Figure 4.3, panel A, from now on named BM-I). Based on the rigid docking model and on 150 ns of MD, the authors argue that the thiophene ring of M090 fills the hydrophobic cavity formed by $I308_1$, $Y316_1$, and $W92_2$, whereas the cyclopropyl

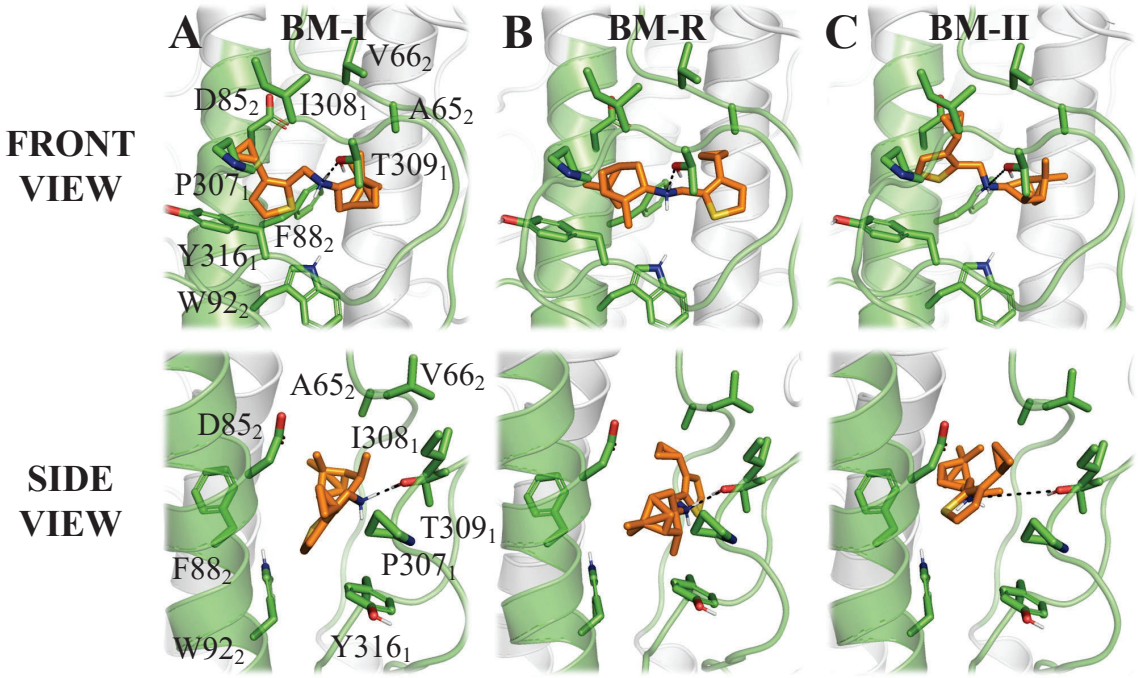


Figure 4.3 Initial conformation of M090 in the three different binding modes studied in this work. The front view is shown at the top of each panel, while the side view is shown at the bottom. The black dotted lines indicate the hydrogen bond between M090 and $T309_1$. The three explored binding modes are shown: BM-I (A), corresponds to the binding mode predicted by Zhao et al.,[183] BM-R (B) and BM-II (C).

ring interacts with $F88_2$ and $P307_1$. In addition, the pinane moiety would form hydrophobic interactions with $A65_2$ and $V66_2$, and a hydrogen bond between the N-H group and $T309_1$ would further stabilise the binding mode. It should be noted that in the work by Zhao et al.[183] there is no explicit mention, either in the text or in the figures, of the use of a monomeric or trimeric model of HA, nor of the protonation state of M090 in the simulations.

To explore the putative binding mode of M090 in more detail, I performed extensive MD simulations of the truncated HA trimer (without the transmembrane domain) in the apo state and in complex with M090 in a 1:3 stoichiometry, with a minimum sampling time of 1 μs and two independent replicates per system (i.e. enabling the binding mode assessment in the three pockets of HA in each simulation). With respect to the ligand model, we ensured that the N atom of M090 was protonated in all simulations, as a secondary amine is mainly present in this state at physiological pH and below (pka of 9.7 as predicted by MolGpka[185]). This protocol was applied to three different binding modes for M090 as a starting point: BM-I, BM-R and BM-II (Figure 4.3, panels A, B and C, respectively), using an HA trimeric model previously simulated for 2 μs in the apo state until structural equilibrium was reached.

Analysis of the two independent MD replicates performed for the BM-I system showed stable RMSD profiles for the amino acids forming the binding pocket (RMSD of 1.6−2.2 Å). However, the h-bond between M090 and $T309_1$ was lost in all cases in the early stages of the simulations and M090 was slightly shifted to the interior of the pocket in five out of six cases. One of the ligands left the binding pocket towards the bulk solvent and the measured RMSD profiles for the others ranged from 2.3 to 4.9 Å from the initial position. Therefore, we hypothesised that the binding mode for M090 proposed by Zhao et al.[183] might be biased due to the use of a rigid model in the molecular docking predictions omitting the overall dynamical plasticity of the loops forming the binding site. For the BM-R, where the positions of the thiophene ring and the pinane moiety were reversed, no clear binding mode was observed in either of the two independent simulations. Indeed, M090 was released from the binding cavity in two out of six cases, suggesting that this reversed orientation may not be feasible to achieve a stable binding mode.

Careful analysis of the first μs of the MD trajectory for the BM-II system revealed a highly stable binding mode for M090. To confirm this behaviour, I extended the replicate to 2 μs and performed three further independent simulations of 1 μs, obtaining the same overall behaviour, as reflected in the consistent RMSD profiles for the amino acids forming the binding pocket (RMSD of 1.2-2.0 Å). In this novel and stable binding mode, M090 approaches the main stem region of the HA trimer early in the simulation and establishes a stable salt bridge interaction between the amine nitrogen and $D85_2$. The thiophene ring carrying the cyclopropyl moiety is stabilised by van der Waals contacts with the side chains of $P307_1$, $I308_1$, and $L89_2$, and the pinane moiety by interactions with $F88_2$, $A65_2$ and $K83_2$ (from the adjacent protomer) (Figure 4.4, panel A). Despite the high stability of the binding mode, the average RMSD profiles for the ligands ranged from 1.0 to 3.6 Å. However, far from destabilising the binding mode, the higher values correspond to sudden reorientations of the pinane moiety through rotations around the N-C bond (Figure 4.4, panel B). Overall, we concluded that although the original binding mode proposed by Zhao et al.[183] may have been biased, the hypothesis arguing that the thiophene ring was a key feature responsible for the activity is supported by our results. In addition, the protonated amine of M090 was also critical to stabilise the binding mode.

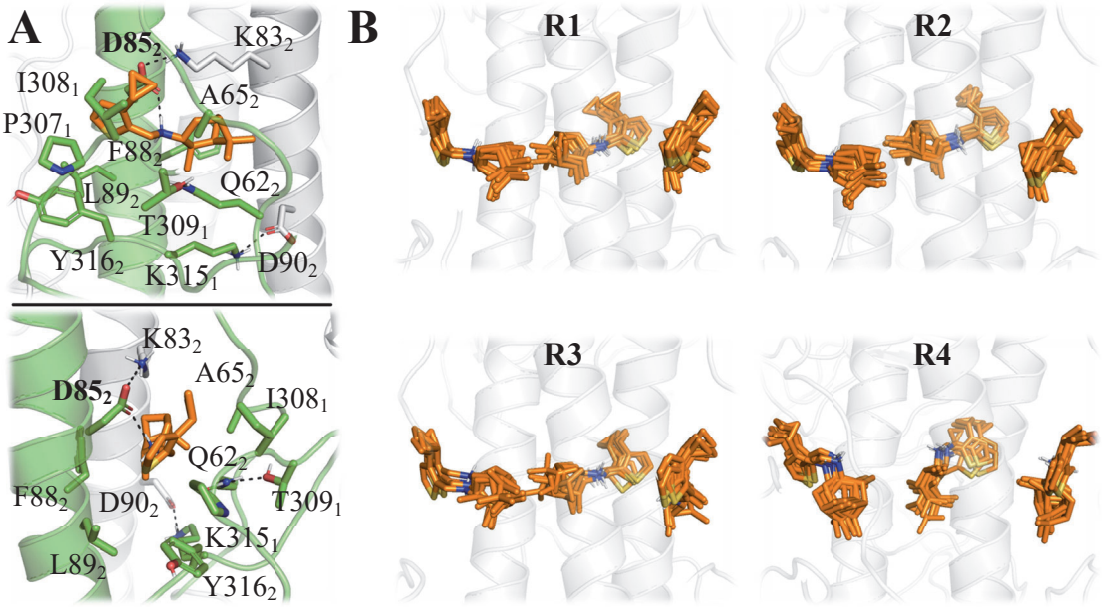


Figure 4.4 A) Representative snapshot of the final binding mode achieved for M090 in the BM-II systems. The front view of the pocket is shown above, while the side view is shown below. B) Superposition of snapshots along the simulation for the four independent replicates of the BM-II systems.

4.2.2 In silico validation of the binding mode using free energy predictions

To validate the novel binding mode of M090 in the BM-II system, I performed a series of relative binding free energy (RBFE) simulations using the experimental SAR data included in the work of Zhao et al.[183] for comparison. The predictions were done using the Thermodynamic Integration (TI) approach as described in Chapter 1 (subsection 1.2.4). The selected compounds exhibited experimental activities ranging from EC_{50} = 0.2 ± 0.0002 to 60.46 ± 1.75 μM, providing a reasonable range of activities suitable for this type of calculation. The closed thermodynamic cycle including all the compounds as well as the fitted correlation with the experimental SAR data is shown in Figure 4.5. The scheme includes all the predicted relative free energy changes in units of kcal·mol^{-1} for each alchemical transformation in both directions of the cycle as well as the closure error of each inner cycle in green. The fitted correlation was done with the averaged free energy changes considering all the possible reaction pathways and relative to M090 as the main compound (Figure 4.5).

The high degree of correlation between the relative free energy predictions and the experimental SAR data (Figure 4.5) confirmed that the predicted binding mode for M090 in the BM-II systems, represents the actual mode of action of M090 and its derivatives in HA.

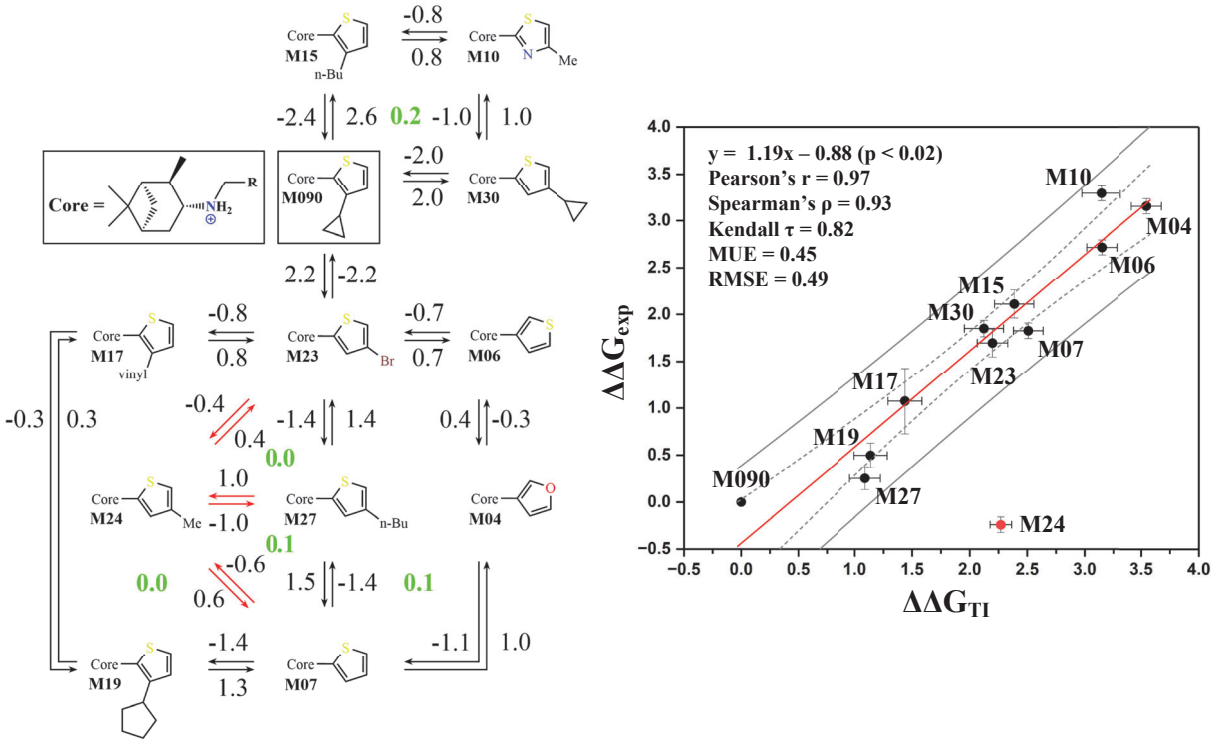


Figure 4.5 Final thermodynamic cycle used in the RBFE simulations. The numbering of the derivatives follows the notation used by Zhao et al. in ref 183. Green values denote the closure error of each inner cycle. The plot shows the comparison between the differences in binding affinity estimated from the experimental EC_{50} data and the relative free energy differences predicted from TI calculations. Bars denote the errors of calculated and experimental relative free energy changes. Dotted and dashed lines denote the 95% confidence and prediction intervals. All values (kcal·mol^{-1}) are relative to M090. MUE and RMSE stand for mean unsigned error and root-mean-square error, respectively.

Moreover, these results support the importance of both the protonated amine and the thiophene ring as a relevant pharmacophoric unit for the inhibitory capacity of these compounds as discussed in the previous section. Considering the structural features found for the binding mode, we hypothesise that this behaviour may be due to a local hydrophobic shielding effect. This shielding may make it more difficult for water molecules to access the interior of the binding pocket, thus preventing competition with the highly stable salt bridge observed in our MD simulations between the charged amine of M090 and $D85_2$. The replacement of the cyclopropyl moiety at this position by a linear *n*-butyl chain (M15) or the complete removal of the cyclopropyl moiety (M07), leads to a 33-fold and 21-fold reduction in activity, respectively. Moreover, the change of the cyclopropyl from position 2 to position 3 in the thiophene ring (M30) or the replacement of the thiophene ring by more polar moieties such as furan (M04) or

thiazole (M10) rings, leads to a drastic reduction in potency (~22-fold, ~200-fold and ~250-fold reduction, respectively). Altogether, the results support the specific binding mode predicted by the MD simulations and provides a rationale for the MOA of M090 and its derivatives.

Only one compound in the series, designated M24, showed a significant deviation from the predicted trend (Figure 4.5). With a methyl group in the meta position of the thiophene ring, M24 is the most active compound reported by Zhao et al.[183] with an EC_{50} = 0.2 ± 0.0002 μM and was not included in the final thermodynamic cycle or in the fitted correlation with the experimental data. It should be noted that in RBFE simulations an estimate of the relative binding affinities is obtained as the change in the relative free energy of binding between a pair of compounds. The approach is based on the approximation that any additional cell signalling process will interfere with the activity measurement in the experimental cell assay. In this case, this means that the measured EC_{50} values would only account for the binding of the compounds to HA. Considering that the initial target in the work of Zhao et al.[183] was the M2 channel of the IAV, I examined the crystallographic structure of the amantadine-resistant M2 channel S31N mutant in complex with N-[(5-bromothiophen-2-yl)methyl]adamantan-1-amine (compound 11 in ref. 185 as a derivative of WJ332, PDB id: 2MUV),[186] a compound showing a ~77% inhibition for both the M2 wild type and its S31N mutant and with a high structural similarity with the explored pinanamine-based compounds. Aligning the structures of M090 and M24 with the WJ332 derivative revealed that while M24 overlaps perfectly with the co-crystallised compound, whereas M090 exhibits steric clashes between the cyclopropyl moiety and the side chains of the amino acids forming the M2 cavity. In the absence of further experimental evidence and considering the high correlation between our binding affinity predictions and the SAR data for all the compounds in the series, we hypothesised that M24 targets either the M2 channel or both the M2 channel and the HA of the IAV.

4.2.3 Effect of the $E74_2 \rightarrow D$ drug-resistant mutation

Resistance selection experiments performed by Zhao et al.[183] disclosed the occurrence of a single point mutation in the presence of M090, $E74_2 \rightarrow D$, located in the HA_2 domain of HA (Figure 4.6). To assess the effect of this mutation on the mode of action of M090 discussed above, I performed further MD simulations for the mutant apo HA trimer (2 μs) and for the mutant holo HA-M090 complex (two replicates of 1 μs), preserving the predicted binding mode

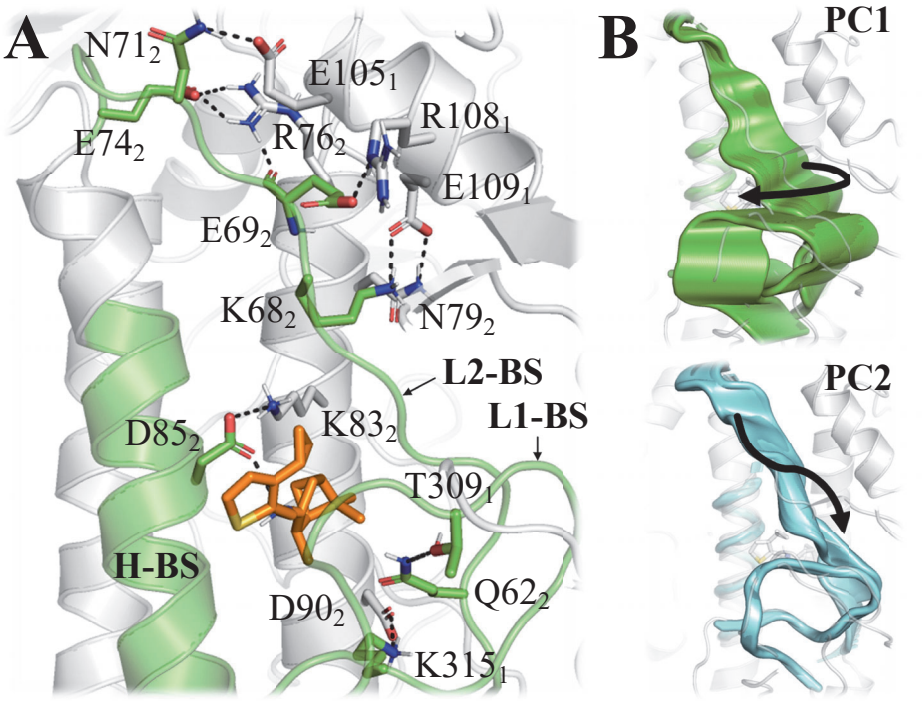


Figure 4.6 A) Representative snapshot of the predicted M090 binding mode (BM-II). H-BS, L1-BS and L2-BS regions are highlighted in green and the main amino acids forming the hydrogen bond network are shown in sticks. Green regions account for one protomer while grey regions account for the adjacent protomer. B) Representation of the two main essential motions of the backbone atoms that shape the binding pocket for both the wild type and the E74$_2$ → D mutant.

(BM-II). With the combined sampling for the wild type and the mutant systems, PCA was used to investigate the effect of the E74$_2$ → D mutation on the motions of the backbone atoms forming the binding site of M090, i.e. H-BS, L1-BS and L2-BS, as discussed previously and as shown in Figure 4.6.

The analysis revealed that for both the wild type and the mutant species, the first principal component (PC1) accounts for 30% and 37% of the structural variance for the apo and holo species, respectively. PC1 represents the rotational motion along the principal C3 axis of HA (Figure 4.6, panel B). However, the projection of the sampled conformations along the MD simulations revealed that the main difference between the wild type and the mutant species concerns the second principal component (PC2). This motion accounts for the conformational remodelling of the L2-BS loop and explains a 19% and an 18% of the structural variance for the apo and holo species, respectively. The E74$_2$ → D mutation alters the optimal h-bond network between several amino acids of the L2-BS loop and the neighbouring amino acids.

This network of interactions blocks the L2-BS conformation relative to the L1-BS and H-BS regions in the wild type species, allowing the binding of M090 in the correct position (Figure 4.6, panel A). This network is described by the following interactions: $E74_2 \cdots R76_2 \cdots E69_2 \cdots R108_1$, $N71_2 \cdots E105_1$, $K68_2 \cdots N79_2 \cdots E109_1$, $Q62_2 \cdots T309_1$ and $D90_2 \cdots K315_1$. In conclusion, the $E74_2 \rightarrow D$ mutation weakens the described h-bond network and increases the conformational flexibility of the L2-BS loop. The enhanced mobility of the L2-BS modifies the overall packing against the L1-BS and the H-BS regions and alters the proper accommodation of M090 in the pocket, resulting in the loss of inhibitory activity.

4.3 *"N-[(Thiophen-3-yl)methylbenzamides as Influenza Virus Fusion Inhibitors Acting on H1 and H5 Hemagglutinins"*

This project investigated a series of benzamide derivatives as HA-mediated fusion inhibitors against the IAV. In 1999, Plotch et al.[187] reported the promising submicromolar antiviral activity of an N-substituted piperidine, namely CL-385319, against A/H1N1 and A/H2N2 IAV strains. More than a decade later, Liu et al.[188] reported that CL-385319 was also effective against highly pathogenic A/H5N1 strains, despite of its reduced inhibitory capacity. The same authors then worked on several thiophene analogues of CL-385319, leading to a 2-(thiophenyl-2-yl)ethyl derivative,[189] which was 58-fold and 360-fold more potent against A/H1N1 and A/H5N1 virus-infected cell cultures, respectively, compared to CL-385319.[190]

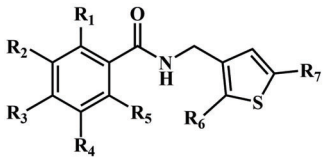
Based on the promising results for CL-385319 and its 2-(thiophenyl-2-yl)ethyl derivative as HA-mediated fusion inhibitor, a multidisciplinary collaboration with the groups of Prof. S. Vázquez (Departament de Farmacologia, Toxicologia i Química Terapèutica, UB) and Prof. L. Naesens (Department of Microbiology, Immunology and Transplantation, KUL) was established to further explore the chemotype of this compound. In this context, we investigated the SAR of a series of derivatives including chemical modifications on the benzamide and thiophene rings. The lead compound, VF-57a, was then tested against IAV-infected MDKC cells. My contribution was focused on the exploration of the MOA of VF-57a by combining MD simulations and binding affinity predictions.

4.3.1 Benzamides as HA-mediated fusion inhibitors: SAR study

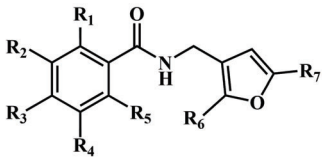
The chemical modifications studied by the group of Prof. Vázquez were designed to explore i) the optimal substitution pattern in the thiophene ring, ii) the effect of substitutions in the phenyl ring, and iii) the nature of the linker between the two aromatic moieties. With respect to the thiophene ring, the influence of the methyl groups in positions 2 and 5 was investigated. Substitution of the thiophene ring with more polar rings such as furan, thiazole, isoxazole, pyrazole and oxazoline were also examined. With respect to the benzene ring, modifications such as the replacement of fluorine by chlorine in *meta* position or the addition of additional electron withdrawing groups in *orto* and *para* positions were explored. Other linkers were also examined while maintaining the total length of the tether between the two aromatic rings. For comparison, RL-007, an aniline-based influenza virus fusion inhibitor identified before by the group,[191] was included as a reference compound (EC_{50} = 0.6 - 1.5 μM).

Anti-IV activity was determined using the cytopathic effect (CPE) reduction assay in Madin-Darby canine kidney (MDCK) cells against different strains: A/H1N1 strains (A/Puerto Rico/8/1934 and A/Virginia/ATCC3/2009), two A/H3N2 viruses (A/Victoria/361/11 and A/Hong Kong/7/87) and an influenza B strain (B/Hong Kong/5/1972). In addition, the cytotoxicity was assessed in mock-infected cells and both measures were evaluated by microscopic inspection. The results showed that none of the compounds tested were active against A/H3N2 or IAB strains of the virus. The active compounds shared the thiophene motif (non-methylated, monomethylated or dimethylated) and the presence of fluorine or chlorine and a trifluoromethyl group in *meta* positions. Notably, two compounds where the thiophene was replaced by furan also showed antiviral activity (Table 4.1). Based on the SAR evaluation, VF-57a was selected as the lead compound of the series and used for mechanistic studies.

Table 4.1 SAR for the antiviral (A/H1N1) activity of the explored benzamides in MDCK cells.



Series A: thiophene derivatives



Series B: furan derivatives

								Antiviral EC_{50} (µM) against A/H1N1 virus			
								A/Puerto Rico/8/1934		A/Virginia/ATCC3/2009	
Compound	**R_1**	**R_2**	**R_3**	**R_4**	**R_5**	**R_6**	**R_7**	MTS[a]	Microscopy[b]	MTS	Microscopy
Series A											
VF-32a	H	F	H	CF_3	H	Me	Me	1.1 ± 0.3	1.5 ± 0.6	0.44 ± 0.19	1.7 ± 1.1
VF-57a	H	Cl	H	CF_3	H	Me	Me	0.92 ± 0.19	1.1 ± 0.2	0.31 ± 0.12	0.71 ± 0.45
JMLv-05	H	F	H	CF_3	H	H	Me	0.68 ± 0.51	0.26 ± 0.02	0.39 ± 0.18	1.6 ± 0.5
JMLv-06	H	Cl	H	CF_3	H	H	Me	0.34 ± 0.18	0.46 ± 0.14	0.15 ± 0.12	1.9 ± 1.1
JMLv-70	H	F	H	CF_3	H	Me	H	0.23 ± 0.08	0.17 ± 0.04	6.8 ± 6.5	8.3 ± 5.1
JMLv-71	H	Cl	H	CF_3	H	Me	H	0.14 ± 0.04	0.14 ± 0.05	0.37 ± 0.16	0.48 ± 0.07
JMLv-72	H	F	H	CF_3	H	H	H	0.42 ± 0.10	0.44 ± 0.16	7.2 ± 4.1	20 ± 8
JMLv-73	H	Cl	H	CF_3	H	H	H	2.0 ± 1.0	1.0 ± 0.6	1.1 ± 0.5	2.1 ± 0.0
JMLv-08	F	F	F	CF_3	H	Me	Me	>100	>100	>100	>100
JMLv-09	F	Cl	H	CF_3	H	Me	Me	2.9	1.6 ± 0.8	>100	>100
JMLv-07	H	F	F	CF_3	H	Me	Me	0.22 ± 0.03	0.86 ± 0.41	0.20 ± 0.05	0.89 ± 0.13
JMLv-10	H	Cl	F	CF_3	H	Me	Me	0.27 ± 0.12	1.6 ± 0.8	0.18 ± 0.04	3.1 ± 0.5
JMLv-11	H	Cl	H	CF_3	Cl	Me	Me	5.9 ± 3.2	11 ± 6	0.82 ± 0.61	8.1 ± 0.5
JMLv-67	H	F	F	CF_3	H	H	Me	2.9 ± 1.1	3.5 ± 1.2	5.3 ± 0.8	>100
JMLv-68	H	Cl	F	CF_3	H	H	Me	8.5 ± 5.6	13 ± 8	>100	>100
JMLv-69	H	Cl	H	CF_3	Cl	H	Me	17 ± 8	>100	31 ± 3	>100
Series B											
JMLv-03	H	F	H	CF_3	H	Me	Me	0.47 ± 0.19	0.54 ± 0.17	0.54 ± 0.06	3.6 ± 2.8
JMLv-04	H	Cl	H	CF_3	H	Me	Me	0.24 ± 0.08	0.54 ± 0.37	1.1 ± 0.3	>100
Reference											
RL-007								1.4 ± 0.0	1.5 ± 0.1	0.61 ± 0.06	0.69 ± 0.09

[a] colorimetric cell viability assay; [b] cataphatic effect assay

4.3.2 Pseudovirus entry and cell-cell fusion assays: Validation of VF-57a

To confirm that VF-57a acts in the early stages of IAV infection, Prof. Naesens' group performed a luciferase-based pseudovirus entry assay in MDCK cells infected with A/H1N1 strains and A/H5N1 highly pathogenic strains (A/Virginia/ATCC3/2009 and A/bald eagle/FL/W22-114/2022, respectively). RL-007 and Arbidol were included in the assays for comparison. VF-57a showed strong inhibition of both viruses with EC_{50} = 0.30 and 0.81 µM for the A/H1N1 and A/H5N1 strains, respectively, confirming its activity in the early stages of infection (Figure 4.7, panel A). In addition, VF-57a proved to be ~20-fold more potent than Arbidol, demonstrating its potential as an anti-IV inhibitor against A/H1N1 strains and A/H5N1 highly pathogenic strains. Of note, RL-007 showed approximately 127-fold greater activity against A/H1N1 than A/H5N1 with EC_{50} = 0.079 and 10 µM, respectively.

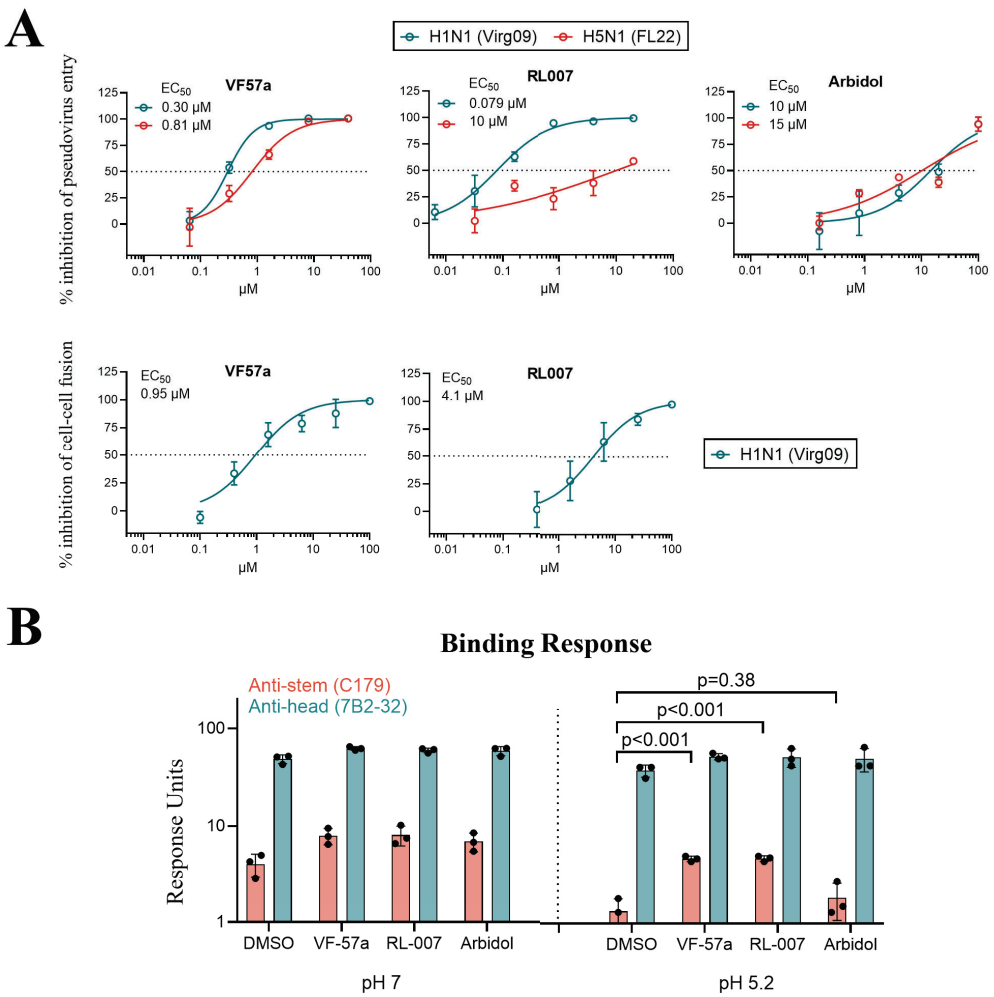


Figure 4.7 A) Dose-response curves for the pseudovirus entry and cell-cell fusion assays for measuring the inhibitory potency of VF-57a, RL-007 and Arbidol. B) Inhibitor binding response in the presence of HA-binding antibodies indicating whether HA is in the prefusion or in the postfusion conformation at physiological or acidic pH.

Cell-cell fusion assays were performed to confirm that VF-57a specifically inhibits the HA-mediated membrane fusion. The results showed that VF-57a is effective against H1 and H5 HA variants with activities of EC_{50} = 0.95 and 4.1 μM against A/H1N1 and A/H5N1 strains, respectively, the latter showing complete inhibition at 100 μM concentration (Figure 4.7, panel A). In addition, while RL-007 is partially effective at 25 μM and inactive at 6.25 μM concentrations, VF-57a is able to inhibit almost completely the fusion process at 25 μM and partially at 6.25 μM concentrations. Overall, the assays validate VF-57a as a potent HA-mediated fusion inhibitor against A/H1N1 and A/H5N1 strains of the IAV.

Finally, surface plasmon resonance (SPR) analysis with two anti-HA antibodies were performed. The anti-HA stem antibody C179 only binds when HA is in its prefusion conformation,[192] while the anti-HA head antibody 7B2-32 binds to both pre- and postfusion conformations.[193] When recombinant H1-HA is incubated in acidic media and DMSO as control, inducing the postfusion conformation, the protein loses its ability to bind to the C179-coated sensor (Figure 4.7, panel B). On the other hand, its binding to the 7B2-32 antibody is not affected. In the presence of VF-57a or RL-007, the binding response to the C179 antiobody is significantly higher compared to the DMSO control, demonstrating that both inhibitors can stabilise the prefusion conformation of HA during acidification.

4.3.3 VF-57a-resistant markers in H1 HA

To gain further insights into the MOA of VF-57a and RL-007, MDCK cells were infected with A/Virginia/ATCC3/2009 virus and exposed to increasing concentrations of the inhibitors until resistance was achieved. During this process, the different plaques were examined to measure the compound concentrations at which virus-induced CPE was visible.

As shown in Figure 4.8, the viral HA carried amino acid mutations in different regions depending on the inhibitor. It is worth noting that in a previous study,[175] Prof. Naesens' and coworkers found a similar behaviour for an H3 HA-specific fusion inhibitor where two different mechanistic substitutions were observed in the protein during the drug-resistance assay. Some mutations were found in the binding pocket of the compound, while others were located in more distant regions associated with higher acid-labile properties. In the latter case, the mutant viruses were able to evade the effect of the fusion inhibitor by undergoing fusion in less acidic endosomes.

Compound:mutation	Antiviral EC_{50} (CPE, µM) VF-57a	RL-007	Hemolysis pH (p vs control)
VF-57a: $F128_1L$	>40	>40	5.25 (p=0.81)
VF-57a: $T316_1I$	>40	>40	5.48 (p <0.0001)
VF-57a: $S54_2P$	>40	>40	5.30 (p =0.030)
RL-007: $A90_1V$	>40	>40	5.10 (p <0.0001)
RL-007: $D90_2G$	>40	>40	5.41 (p =0.030)
control: wild type	2.5 ± 1.7	0.94 ± 0.67	5.24

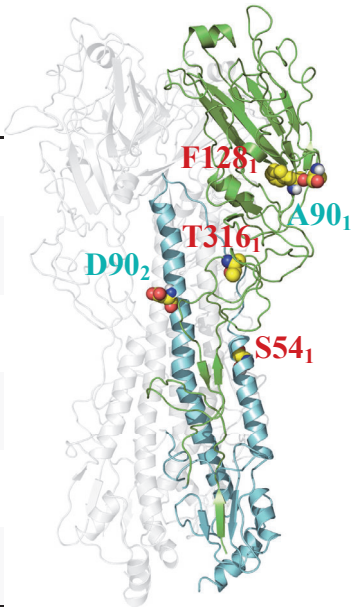


Figure 4.8 Impact of HA mutations on viral sensitivity to VF-57a and RL-007 and on acid-lability of HA. The hemolysis pH was defined as the pH where 50% of virus-induced hemolysis occurred relative to pH 5.0. Amino acid mutations are highlighted in yellow spheres in a model of A/Virginia/ATCC3/2009 HA. The mutations, emerging from the presence of VF-57a are highlighted in red, while the mutations emerging from the presence of RL-007 are highlighted in blue. The suffix refers to the HA1 (green) or HA2 (cyan) domain of HA.

On the basis of these results, a hemolysis assay was performed to determine the effect of the drug resistance mutations on the pH susceptibility of A/Virginia/ATCC3/2009 HA. The results revealed that the $T316_1I$ mutant, which occurs in the presence of VF-57a, and the $D90_2G$ mutant, which occurs in the presence of RL-007, showed the greater increase in the virus-induced hemolysis pH compared to the wild-type control. In the case of the $S54_1P$ mutant, which appears in the presence of VF-57a, only a slight increase in hemolysis pH was measured with respect to the control. The same behaviour was observed for the $F128_1L$ mutant, which also appears in the presence of VF-57a and has the same hemolysis pH as the control protein. On the other hand, the $A90_1V$ mutation, which appears in the presence of RL-007, is the only one that reduces the hemolysis pH compared to the control. Regarding VF-57a, the results suggest that its binding pocket might be close to the regions that surround either $S54_1P$ or $F128_1L$ mutations, which are located at >54 Å (Cα-Cα distance).

4.3.4 VF-57a binding site and binding mode

Previous studies have identified two major binding pockets for HA-mediated fusion inhibitors,[172,173,177–179] namely Site A and Site B, as discussed in Chapter 1 and shown in Figure 1.17 (subsection 1.3.2.3). Site A accounts for the binding pocket of the well-known broad-spectrum inhibitor Arbidol, which is located close to the drug-resistant mutation $S54_1P$ emerging in A/Virginia/ATCC3/2009 virus in the presence of VF-57a as discussed above. On the other hand, Site B accounts for the binding pocket for several inhibitors such as CBS1117, JNJ4776 or (*S*)-F0045, and the region in the HA structure where the anti-stem antibody C179 binds. For Site A, I performed MD simulations using the X-ray structure of H3 HA in complex with Arbidol as a control (PDBid: 5T6N,[173] resolution 2.54 Å). For Site B, I performed MD simulations using the X-ray structure of H1 HA in complex with (*S*)-F0045 as a control (PDBid: 6WCR,[178] resolution 2.68 Å).

For Site B, the potential binding of VF-57a was investigated considering two different binding modes, binding mode A (BMA) and B (BMB), which differ in the spatial arrangement of the benzene and thiophene rings along the groove (Figure 4.9, panels B and C). Analysis of the MD trajectories revealed that none of the investigated binding modes for VF-57a are stable

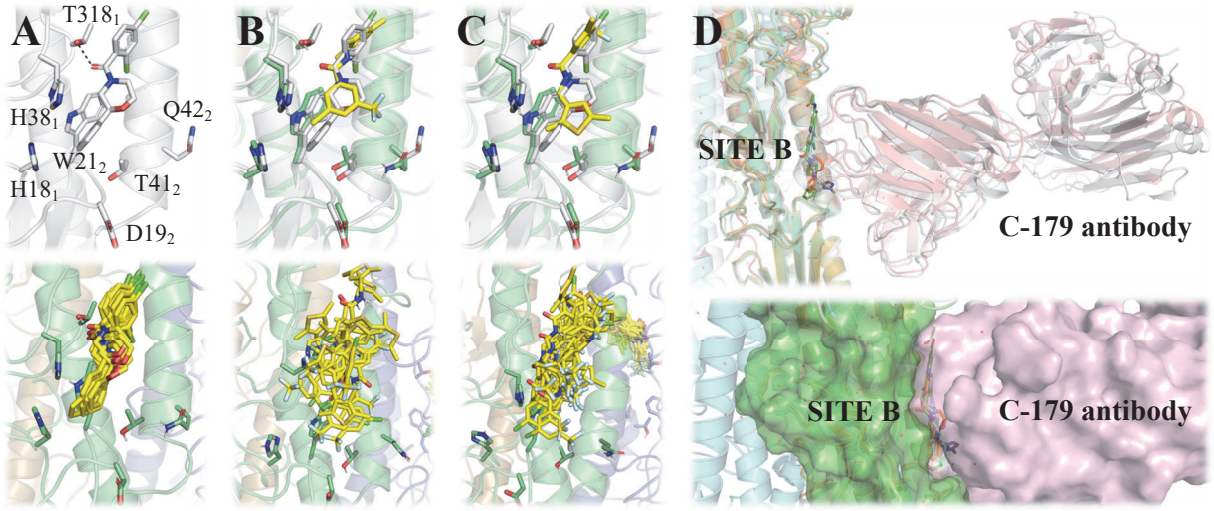


Figure 4.9 A) Initial conformation of (*S*)-F0045 and several snapshots along the MD simulation in one of the three pockets; B) Initial conformation of VF-57a in the BMA and several snapshots along the MD simulation in one of the three pockets; C) Initial conformation of VF-57a in the BMB and several snapshots along the MD simulation in one of the three pockets; D) H1 HA in complex with the anti-HA stem antibody C179 bound to Site B. X-ray structures of HA in complex with HA-mediated fusion inhibitors CBS1117, JNJ4776 and (*S*)-F0045 are included for clarity.

at Site B. Compared to (*S*)-F0045, the instability of VF-57a reflects the increased flexibility due to the methylene between the amide and thiophene moieties and the loss of the CH-π interactions formed by (*S*)-F0045 with $H18_1$, $H38_1$ and $W21_2$ due to the presence of CF_3 and CH_3 groups in the phenyl and thiophene rings (Figure 4.9, panels B and C). On the other hand, the inspection of an X-ray structure for H1 HA in complex with the anti-HA antibody C179 (PDBid: 5C0R,[194] resolution 3.19 Å) reveals that it binds directly to Site B (Figure 4.9, panel D). Taken together, VF-57a could not bind to Site B and stabilize the prefusion conformation of HA upon acidification as demonstrated by the SPR assay performed by the group of Prof. Naesens.

A similar approach was taken for site A. In this case, VF-57a is displaced deeply into the pocket compared to Arbidol and adopts a stable binding mode as indicated by averaged RMSD values of 2.3 ± 0.5, 2.3 ± 0.4 and 2.9 ± 1.0 Å and visual inspection of the trajectories. The thiophene ring is stabilised by hydrophobic interactions with $V36_1$, $L49_1$, $P308_1$, $F309_1$, $V55_2$, $L99_2$, $L98_2$ and $L102_2$. On the other hand, the ring containing the trifluoromethyl moiety is stabilised by π-stacking interactions with $Y94_2$ and $F309_1$ (Figure 4.10, panel A). It is worth noting that the binding pocket is close to the emergence of the $S54_1P$ mutation, as noted in the distance from the inhibitor to the Cα of this residue (8.0 ± 0.4 Å). In fact, the presence of proline instead of serine is expected to promote a structural rearrangement of the helix that shapes the entrance of the binding pocket potentially disrupting the proper accommodation of the inhibitor.

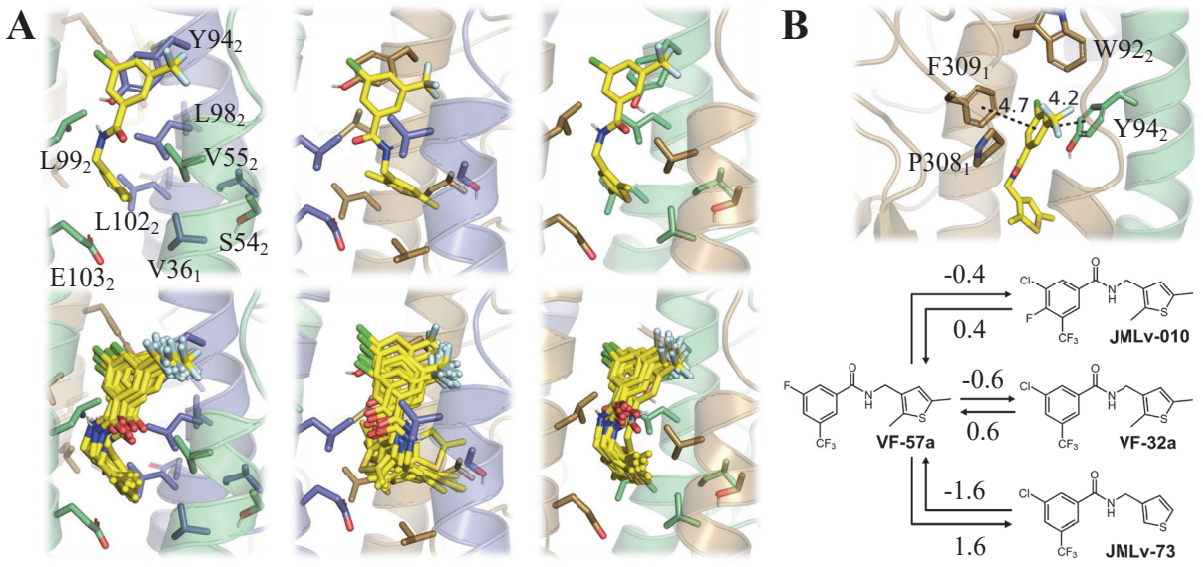


Figure 4.10 A) Final conformation of VF-57a in Site B and several snapshots along the MD trajectories for each of the three pockets of the HA trimer; B) results of the predicted change in free energy in kcal·mol^{-1} for both directions of each transformation.

To validate the predicted binding mode for VF-57a, a series of RBFE simulations using the experimental SAR data obtained by Prof. Vázquez's group (Table 4.1). From the SAR, we selected three compounds based on their cell viability activity against A/Virginia/ATCC3/2009 virus: compounds VF-32a, JML v-010 and JML v-73.

For the VF-57a → VF-32a transformation, a destabilisation of 0.6 kcal-mol^{-1} is predicted, which is consistent with the experimental 1.4-fold difference in inhibitory potency between the two compounds. The replacement of fluorine by chlorine occurs in the region surrounded by $N95_2$ and $W92_2$ (Figure 4.10, panel B). Although the substitution in the predicted binding mode should increase the local hydrophobicity of the inhibitor, the change may affect the solubility of the molecule and make proper accommodation in Site A more difficult.

The VF-57a → JML v-010 transformation implies the introduction of an additional fluorine atom in *para* position of the benzene ring of VF-57a (relative to the carbonyl group). This change is predicted to lead to a slight stabilisation of 0.4 kcal-mol^{-1}. This effect is consistent with the approximately 1.7-fold increase in potency over VF-57a in cell assays, which is attributed to an increase in local hydrophobicity due to the interaction of the introduced fluorine atom with $W92_2$ at this position.

Finally, for the VF-57a → JML v-73 transformation in which the two methyl groups in the thiophene ring are removed, a destabilisation of 1.6 kcal·mol^{-1} is predicted. This prediction is consistent with the 3.5-fold decrease in inhibitory potency relative to VF-57a, which is attributed to the loss of stabilising hydrophobic interactions between the dymethylthiophene moiety and the local hydrophobic environment that shapes the binding pocket in the predicted conformation for VF-57a.

Chapter 5

Conclusions

"Surtr travels from the south with the scourge of twigs,
the sun of the battle gods, shone from his sword.
Rocky cliffs collapse and witches wander,
the dead throng on the way to Hel, and the sky is cloven.

The sun turns black, land sinks into the sea,
the bright stars vanish from the sky,
steam rises with a roaring noise,
the high heat plays against the sky itself."

Vǫlva (Vǫluspá, stanzas 51, 55)

Project 1: *"Binding of Cholesterol to the N-Terminal Domain of the NPC1L1 Transporter: Analysis of the Epimerization-Related Binding Selectivity and Loop Mutations"*

- Unbiased molecular dynamics revealed that the side chain of L213 is involved in the gating mechanism of the NPC1L1-NTD. In the apo form, this residue interacts with S102 which causes a conformational shift of the α8β7 loop towards the α3 helix closing the S-opening to the solvent.
- The presence of cholesterol in the binding site of the NPC1L1-NTD forces the reorganisation of the sidechain of L213, mimicking the position of Q200 in the cholesterol bound state of its homologue NPC1.
- Cholesterol binds to the binding site of NPC1L1 by establishing hydrophobic interactions with the residues L99, L103, H124, A214 and P215 and three key h-bond interactions with Q95, S56 and a water-mediated interaction with N54. This water molecule is highly conserved in the available cholesterol-bound crystallographic structures of NPC1.
- These interactions are maintained for all the simulated complexes between sterol molecules with a β-configuration of C3 position: such as cholesterol, 25-hydroxycholesterol and lanosterol. The calculated binding affinities to the NPC1L1-NTD are in agreement with the experimental trends reported in the literature.
- Epi-cholesterol, which only differs from cholesterol in the α-configuration of C3 position, shows two different binding modes in the NPC1L1-NTD. The *Q95 binding mode* resembles of the binding mode of cholesterol, but the sterol core moves towards the W-opening to interact with the Q95 and S56 causing the loss of the structural water molecule. The *T128 binding mode* occurs when ECL loses the h-bond with Q95, retracts towards the S-opening and forms a new h-bond interaction with the sidechain of T128. The *Q95 binding mode* for ECL is energetically unfavourable compared to the *T128 binding mode* due to a weaker h-bond network and a larger structural flexibility of the α8β7 loop.
- Single point mutations of the NPC1L1 NTD that experimentally impair cholesterol binding distort the conformational landscape of the α8β7 loop in their respective apo state. This effect is not observed in the cholesterol bound states, and therefore the predicted binding affinities for cholesterol are not in line with the experimental insights. Altogether, these results suggest that the distortion of the S-entrance affects their ability to accommodate substrates.

Project 2: *"Mining Druggable Sites in Influenza A Hemagglutinin: Binding of the Pinanamine Based Inhibitor M090"*

- A thorough structural analysis revealed that M090 binds to the IAV-HA interface defined by the α-helix $L80_2$-$A96_2$ and by loops $P301_1$-$K318_1$ and $M59_2$-$L73_2$. In particular, M090 establishes a salt bridge interaction with the side chain of $D85_2$. In addition, the thiophene ring and pinane moieties are stabilised by van der Waals contacts with the side chains of $P307_1$, $I308_1$, and $L89_2$, and with the side chains of $F88_2$ and $A65_2$, respectively.
- The proposed binding mode for M090 and its derivatives has been further validated by the high degree of correlation between the performed binding free energy predictions and the experimental SAR data.
- We hypothesise that M24, another M090 derivative, may target both the M2 channel and the HA, given its particularly high activity in the cell assays performed by Zhao et al.[183] and the fact that it is the only compound to escape our binding affinity predictions. Our rationale is founded on a hydrophobic/philic-guided fitting overlap between M24 and the M2 inhibitor *N*-[(5-bromothiophen-2-yl)methyl]adamantan-1-amine, a compound that showed ~77% inhibition of the M2 channel and was solved by X-ray crystallography.[186]
- Also, the structural analysis of the unbiased molecular dynamics of the drug-resistant mutation $E74_2$ →D shows that the D sidechain is not able to stabilize the h-bond network locking the conformation of $M59_2$-$L73_2$ loop that forms the binding site of M090, justifying the loss of inhibitory activity.

Project 3: *"N-[(Thiophen-3-yl)methylbenzamides as Influenza Virus Fusion Inhibitors Acting on H1 and H5 Hemagglutinins"*

- A series of benzamide derivatives were synthesised and shown to be potent inhibitors of the HA-mediated fusion against A/H1N1 strains of the IAV. The lead compound, VF-57a showed inhibitory potencies of EC_{50} = 0.31 μM and EC_{50} = 0.92 μM against A/Virginia/ATCC3/2009 and A/Puerto Rico/8/1934, respectively.
- VF-57a was shown to act at early stages of infection in pseudovirus entry assays and to specifically inhibit the HA-mediated fusion mechanism in cell-cell fusion assays. In particular, VF-57a showed potent inhibition with EC_{50} = 0.30 and 0.81 μM for the A/H1N1 and highly pathogenic A/H5N1 strains of IAV, respectively.
- SPR analysis in the presence of two anti-HA antibodies (C179 anti-stem and 7B2-32 anti-head) revealed the ability of VF-57a to stabilise the prefusion conformation of HA upon acidification, further confirming its role as an HA-mediated fusion inhibitor.
- Drug-resistance assays revealed three mutations in the HA structure that occur in the presence of VF-57a: $F128_1L$, $T316_1I$ and $S54_2P$. The latter two, were identified as being in the potential binding regions of VF-57a, as the mutant viruses showed a shift in the hemolysis pH, indicating an impact on their ability to undergo fusion in the acidic endosome.
- VF-57a was predicted to bind in a hydrophobic pocket close to the binding pocket for the well-known anti-HA inhibitor Arbidol. The binding mode was validated by extensive MD simulations and the correlation of binding affinity predictions with the SAR data provided by the group of Prof. Vázquez.

Bibliography

"I see a second time
the earth rise, green again from the sea,
waterfalls tumble, an eagle flies above,
hunting fish on the mountain.

I see a hall standing, fairer than the sun,
thatched with gold, on Gimlé.
There shall the honourable hosts live
and enjoy delight during their lifedays."

Vǫlva (Vǫluspá, stanzas 57, 62)

1. Alberts, B., Heald, R., Johnson, A., Morgan, D., Raff, M., Roberts, K., & Walter, P. (2022). *Molecular biology of the cell*. 7th edn. W. W. Norton & Company.
2. Stick, R., & Williams, S. J. (2009). Carbohydrates: The Essential Molecules of Life. In *Carbohydrates: The Essential Molecules of Life*. Elsevier. doi: https://doi.org/10.1016/B978-0-240-52118-3.X0001-4
3. Slavin, J., & Carlson, J. (2014). Carbohydrates. *Advances in Nutrition*, *5*(6), 760–761. doi: https://doi.org/10.3945/an.114.006163
4. Dyall, S. C., Balas, L., Bazan, N. G., et al. (2022). Polyunsaturated fatty acids and fatty acid-derived lipid mediators: Recent advances in the understanding of their biosynthesis, structures, and functions. *Progress in Lipid Research*, *86*, 101165. doi: https://doi.org/10.1016/j.plipres.2022.101165
5. Minchin, S., & Lodge, J. (2019). Understanding biochemistry: Structure and function of nucleic acids. *Essays in Biochemistry*, *63*(4), 433–456. doi: https://doi.org/10.1042/EBC20180038
6. Koonin, E. V., & Starokadomskyy, P. (2016). Are viruses alive? The replicator paradigm sheds decisive light on an old but misguided question. *Studies in History and Philosophy of Science Part C :Studies in History and Philosophy of Biological and Biomedical Sciences*, *59*, 125–134. doi: https://doi.org/10.1016/j.shpsc.2016.02.016
7. Ouellette, R. J., & Rawn, J. D. (2015). Amino Acids, Peptides, and Proteins. *Principles of Organic Chemistry* (p. 371–396). Elsevier. doi: https://doi.org/10.1016/b978-0-12-802444-7.00014-8
8. Sanger, F., & Tuppy, H. (1951). The amino-acid sequence in the phenylalanyl chain of insulin. I. The identification of lower peptides from partial hydrolysates. *The Biochemical Journal*, *49*(4), 463–481. doi: https://doi.org/10.1042/bj0490463
9. Downing, A. K. (2001). Introduction to Protein Architecture: the structural biology of proteins. In *Journal of Cell Science* (Vol. 114, Issue 12). Oxford University Press. doi: https://doi.org/10.1242/jcs.114.12.2210
10. Donohue, J. (1953). Hydrogen Bonded Helical Configurations of the Polypeptide Chain. *Proceedings of the National Academy of Sciences*, *39*(6), 470–478. doi: https://doi.org/10.1073/pnas.39.6.470
11. Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, *181*(4096), 223–230. doi: https://doi.org/10.1126/science.181.4096.223
12. Gohlke, H., & Klebe, G. (2002). Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angewandte Chemie - International Edition*, *41*(15), 2644–2676. doi: https://doi.org/10.1002/1521-3773(20020802)41:15<2644::AID-ANIE2644>3.0.CO;2-O
13. Robinson, P. K. (2015). Enzymes: principles and biotechnological applications. *Essays in Biochemistry*, *59*, 1–41. doi: https://doi.org/10.1042/BSE0590001
14. Greco, T. M., & Cristea, I. M. (2017). Proteomics tracing the footsteps of infectious disease. *Molecular and Cellular Proteomics*, *16*(4), S5–S14. doi: https://doi.org/10.1074/mcp.O116.066001
15. Penders, B., Horstman, K., & Vos, R. (2008). Walking the line between lab and computation: The “moist” zone. *BioScience*, *58*(8), 747–755. doi: https://doi.org/10.1641/B580811
16. Silva, D. A., Yu, S., Ulge, U. Y., et al. (2019). De novo design of potent and selective mimics of IL-2 and IL-15. *Nature*, *565*(7738), 186–191. doi: https://doi.org/10.1038/s41586-018-0830-7
17. Mravic, M., Thomaston, J. L., Tucker, M., Solomon, P. E., Liu, L., & DeGrado, W. F. (2019). Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science*, *363*(6434), 1418–1423. doi: https://doi.org/10.1126/science.aav7541
18. Dou, J., Vorobieva, A. A., Sheffler, W., et al. (2018). De novo design of a fluorescence-activating β-barrel. *Nature*, *561*(7724), 485–491. doi: https://doi.org/10.1038/s41586-018-0509-0

19. Berman, H. M., Westbrook, J., Feng, Z., et al. (2000). The Protein Data Bank. *Nucleic Acids Research*, *28*(1), 235–242. doi: https://doi.org/10.1093/nar/28.1.235

20. Yu, L., Xue, L., Liu, F., Li, Y., Jing, R., & Luo, J. (2022). The applications of deep learning algorithms on in silico druggable proteins identification. *Journal of Advanced Research*, *41*, 219–231. doi: https://doi.org/10.1016/j.jare.2022.01.009

21. Senior, A. W., Evans, R., Jumper, J., et al. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, *577*(7792), 706–710. doi: https://doi.org/10.1038/s41586-019-1923-7

22. Jumper, J., Evans, R., Pritzel, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, *596*(7873), 583–589. doi: https://doi.org/10.1038/s41586-021-03819-2

23. Abramson, J., Adler, J., Dunger, J., et al. (2024). Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature*, *630*(8016), 493–500. doi: https://doi.org/10.1038/s41586-024-07487-w

24. Varadi, M., Bertoni, D., Magana, P., et al. (2024). AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Research*, *52*(D1), D368–D375. doi: https://doi.org/10.1093/nar/gkad1011

25. Baek, M., DiMaio, F., Anishchenko, I., et al. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. *Science*, *373*(6557), 871–876. doi: https://doi.org/10.1126/science.abj8754

26. Marchal, I. (2024). RoseTTAFold expands to all-atom for biomolecular prediction and design. *Nature Biotechnology*, *42*(4), 571. doi: https://doi.org/10.1038/s41587-024-02211-5

27. Lisanza, S. L., Gershon, J. M., Tipps, S. W. K., et al. (2024). Multistate and functional protein design using RoseTTAFold sequence space diffusion. *Nature Biotechnology*. Advance online publication. doi: https://doi.org/10.1038/s41587-024-02395-w

28. Vemula, D., Jayasurya, P., Sushmitha, V., Kumar, Y. N., & Bhandari, V. (2023). CADD, AI and ML in drug discovery: A comprehensive review. *European Journal of Pharmaceutical Sciences*, *181*, 106324. doi: https://doi.org/10.1016/j.ejps.2022.106324

29. Jorgensen, W. L. (2004). The Many Roles of Computation in Drug Discovery. *Science*, *303*(5665), 1813–1818. doi: https://doi.org/10.1126/science.1096361

30. Oliveira, T., Silva, M., Maia, E., Silva, A., & Taranto, A. (2023). Virtual Screening Algorithms in Drug Discovery: A Review Focused on Machine and Deep Learning Methods. In *Drugs and Drug Candidates, 2(2),* 311–334. doi: https://doi.org/10.3390/ddc2020017

31. Sahu, D., Rathor, L. S., Dwivedi, S. D., et al. (2024). A Review on Molecular Docking As an Interpretative Tool for Molecular Targets in Disease Management. *Assay and Drug Development Technologies*, *22*(1), 40–50. doi: https://doi.org/10.1089/adt.2023.060

32. Tang, Y., Moretti, R., & Meiler, J. (2024). Recent Advances in Automated Structure-Based De Novo Drug Design. *Journal of Chemical Information and Modeling*, *64*(6), 1794–1805. doi: https://doi.org/10.1021/acs.jcim.4c00247

33. Thomas, M., Bender, A., & de Graaf, C. (2023). Integrating structure-based approaches in generative molecular design. *Current Opinion in Structural Biology*, *79*, 102559. doi: https://doi.org/10.1016/j.sbi.2023.102559

34. Lewis-Atwell, T., Townsend, P. A., & Grayson, M. N. (2021). Comparisons of different force fields in conformational analysis and searching of organic molecules: A review. *Tetrahedron*, *79*, 131865. doi: https://doi.org/10.1016/j.tet.2020.131865

35. Sinha, S., Tam, B., & Wang, S. M. (2022). Applications of Molecular Dynamics Simulation in Protein Study. *Membranes*, *12*(9). doi: https://doi.org/10.3390/membranes12090844

36. He, X., Walker, B., Man, V. H., Ren, P., & Wang, J. (2022). Recent progress in general force fields of small molecules. *Current Opinion in Structural Biology*, *72*, 187–193. doi: https://doi.org/10.1016/j.sbi.2021.11.011

37. Ponder, J. W., & Case, D. A. (2003). Force fields for protein simulations. *Advances in Protein Chemistry*, *66*, 27–85. doi: https://doi.org/10.1016/S0065-3233(03)66002-X

38. Bayly, C. I., Merz, K. M., Ferguson, D. M., et al. (1995). A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *Journal of the American Chemical Society*, *117*(19), 5179–5197. doi: https://doi.org/10.1021/ja00124a002

39. Jones, J. E., & Chapman, S. (1924). On the determination of molecular fields.—I. From the variation of the viscosity of a gas with temperature. *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, *106*(738), 441–462. doi: https://doi.org/10.1098/rspa.1924.0081

40. Lennard-Jones, J. E. (1931). Cohesion. *Proceedings of the Physical Society*, *43*(5), 461–482. doi: https://doi.org/10.1088/0959-5309/43/5/301

41. Whitten, J. L. (1973). Coulombic potential energy integrals and approximations. *The Journal of Chemical Physics*, *58*(10), 4496–4501. doi: https://doi.org/10.1063/1.1679012

42. Galindo-Murillo, R., Robertson, J. C., Zgarbová, M., et al. (2016). Assessing the Current State of Amber Force Field Modifications for DNA. *Journal of Chemical Theory and Computation*, *12*(8), 4114–4127. doi: https://doi.org/10.1021/acs.jctc.6b00186

43. Kirschner, K. N., Yongye, A. B., Tschampel, S. M., et al. (2008). GLYCAM06: A generalizable biomolecular force field. carbohydrates. *Journal of Computational Chemistry*, *29*(4), 622–655. doi: https://doi.org/10.1002/jcc.20820

44. Dickson, C. J., Walker, R. C., & Gould, I. R. (2022). Lipid21: Complex Lipid Membrane Simulations with AMBER. *Journal of Chemical Theory and Computation*, *18*(3), 1726–1736. doi: https://doi.org/10.1021/acs.jctc.1c01217

45. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A., & Case, D. A. (2004). Development and testing of a general Amber force field. *Journal of Computational Chemistry*, *25*(9), 1157–1174. doi: https://doi.org/10.1002/jcc.20035

46. Joung, I. S., & Cheatham, T. E. (2008). Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations. *Journal of Physical Chemistry B*, *112*(30), 9020–9041. doi: https://doi.org/10.1021/jp8001614

47. Li, P., & Merz, K. M. (2017). Metal Ion Modeling Using Classical Mechanics. *Chemical Reviews*, *117*(3), 1564–1686. doi: https://doi.org/10.1021/acs.chemrev.6b00440

48. Yang, Y., Pan, L., Lightstone, F. C., & Merz, K. M. (2016). The Role of Molecular Dynamics Potential of Mean Force Calculations in the Investigation of Enzyme Catalysis. In G. A. B. T. M. in E. Voth (Ed.), *Methods in Enzymology, 577,* 1–29. Academic Press. doi: https://doi.org/10.1016/bs.mie.2016.05.040

49. Izrailev, S., Stepaniants, S., Isralewitz, B., et al. (1999). *Steered Molecular Dynamics BT - Computational Molecular Dynamics: Challenges, Methods, Ideas* (P. Deuflhard, J. Hermans, B. Leimkuhler, A. E. Mark, S. Reich, & R. D. Skeel, Eds.; p. 39–65). Springer Berlin Heidelberg.

50. Jarzynski, C. (1997). Nonequilibrium equality for free energy differences. *Physical Review Letters*, *78*(14), 2690–2693. doi: https://doi.org/10.1103/PhysRevLett.78.2690

51. Noh, S. Y., & Notman, R. (2020). Comparison of umbrella sampling and steered molecular dynamics methods for computing free energy profiles of aromatic substrates through phospholipid bilayers. *Journal of Chemical Physics*, *153*(3), 34115. doi: https://doi.org/10.1063/5.0016114

52. Kumar, S., Rosenberg, J. M., Bouzida, D., Swendsen, R. H., & Kollman, P. A. (1992). The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry*, *13*(8), 1011–1021. doi: https://doi.org/https://doi.org/10.1002/jcc.540130812

53. Kästner, J., & Thiel, W. (2005). Bridging the gap between thermodynamic integration and umbrella sampling provides a novel analysis method: “Umbrella integration.” *The Journal of Chemical Physics*, *123*(14), 144104. doi: https://doi.org/10.1063/1.2052648

54. Barducci, A., Bonomi, M., & Parrinello, M. (2011). Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, *1*(5), 826–843. doi: https://doi.org/10.1002/wcms.31

55. Laio, A., & Parrinello, M. (2002). Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(20), 12562–12566. doi: https://doi.org/10.1073/pnas.202427399

56. Barducci, A., Bussi, G., & Parrinello, M. (2008). Well-Tempered Metadynamics: A Smoothly Converging and Tunable Free-Energy Method. *Physical Review Letters*, *100*(2), 20603. doi: https://doi.org/10.1103/PhysRevLett.100.020603

57. Bussi, G., & Laio, A. (2020). Using metadynamics to explore complex free-energy landscapes. *Nature Reviews Physics*, *2*(4), 200–212. doi: https://doi.org/10.1038/s42254-020-0153-0

58. Adediwura, V. A., Koirala, K., Do, H. N., Wang, J., & Miao, Y. (2024). Understanding the impact of binding free energy and kinetics calculations in modern drug discovery. *Expert Opinion on Drug Discovery*, *19*(6), 671–682. doi: https://doi.org/10.1080/17460441.2024.2349149

59. Ott, J. B., & Boerio-Goates, J. (2000). Thermodynamic Relationships and Applications. In J. B. Ott & J. B. T. C. T. P. and A. Boerio-Goates (Eds.), *Chemical Thermodynamics: Principles and Applications* (p. 105–153). Academic Press. doi: https://doi.org/10.1016/b978-012530990-5/50004-3

60. Dcosta, V. M., King, C. E., Kalan, L., et al. (2011). Antibiotic resistance is ancient. *Nature*, *477*(7365), 457–461. doi: https://doi.org/10.1038/nature10388

61. Abbafati, C., Abbas, K. M., Abbasi, M., et al. (2020). Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The Lancet*, *396*(10258), 1204–1222. doi: https://doi.org/10.1016/S0140-6736(20)30925-9

62. Chong, B., Jayabaskaran, J., Jauhari, S. M., et al. (2024). Global burden of cardiovascular diseases: projections from 2025 to 2050. *European Journal of Preventive Cardiology*, zwae281. doi: https://doi.org/10.1093/eurjpc/zwae281

63. Wilcox, N. S., Amit, U., Reibel, J. B., Berlin, E., Howell, K., & Ky, B. (2024). Cardiovascular disease and cancer: shared risk factors and mechanisms. *Nature Reviews Cardiology*, *21*(9), 617–631. doi: https://doi.org/10.1038/s41569-024-01017-x

64. Mensah, G. A., Habtegiorgis Abate, Y., Abbasian, M., et al. (2023). Global Burden of Cardiovascular Diseases and Risks, 1990-2022. *Journal of the American College of Cardiology*, *82*(25), 2350–2473. doi: https://doi.org/10.1016/j.jacc.2023.11.007

65. Björkegren, J. L. M., & Lusis, A. J. (2022). Atherosclerosis: Recent developments. *Cell*, *185*(10), 1630–1645. doi: https://doi.org/10.1016/j.cell.2022.04.004

66. National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III). (2002). Third Report of the National Cholesterol Education Program (NCEP) Expert Panel on Detection, Evaluation, and Treatment of High Blood Cholesterol in Adults (Adult Treatment Panel III) final report. *Circulation*, *106*(25), 3143–3421. doi: https://doi.org/10.1161/circ.106.25.3143

67. Endo, A. (1992). The discovery and development of HMG-CoA reductase inhibitors. *Journal of Lipid Research*, *33*(11), 1569–1582. doi: https://doi.org/10.1016/s0022-2275(20)41379-3

68. Duan, Y., Gong, K., Xu, S., Zhang, F., Meng, X., & Han, J. (2022). Regulation of cholesterol homeostasis in health and diseases: from mechanisms to targeted therapeutics. *Signal Transduction and Targeted Therapy*, *7*(1), 265. doi: https://doi.org/10.1038/s41392-022-01125-5

69. Betters, J. L., & Yu, L. (2010). NPC1L1 and cholesterol transport. *FEBS Letters*, *584*(13), 2740–2747. doi: https://doi.org/10.1016/j.febslet.2010.03.030

70. Nakano, T., Inoue, I., & Murakoshi, T. (2019). A newly integrated model for intestinal cholesterol absorption and efflux reappraises how plant sterol intake reduces circulating cholesterol levels. *Nutrients*, *11*(2). doi: https://doi.org/10.3390/nu11020310

71. Brown, J., & Yu, L. (2009). Opposing Gatekeepers of Apical Sterol Transport: Niemann-Pick C1-Like 1 (NPC1L1) and ATP-Binding Cassette Transporters G5 and G8 (ABCG5/ABCG8). *Immunology, Endocrine & Metabolic Agents in Medicinal Chemistry*, *9*(1), 18–29. doi: https://doi.org/10.2174/187152209788009797

72. Rosenblum, S. B., Huynh, T., Afonso, A., et al. (1998). Discovery of 1-(4-fluorophenyl)-(3R)-[3-(4-fluorophenyl)-(3S)- hydroxypropyl]-(4S)-(4-hydroxyphenyl)-2-azetidinone (SCH 58235): A designed, potent, orally active inhibitor of cholesterol absorption. *Journal of Medicinal Chemistry*, *41*(6), 973–980. doi: https://doi.org/10.1021/jm970701f

73. Davis, H. R., Compton, D. S., Hoos, L., & Tetzloff, G. (2001). Ezetimibe, a potent cholesterol absorption inhibitor, inhibits the development of atherosclerosis in apoE knockout mice. *Arteriosclerosis, Thrombosis, and Vascular Biology*, *21*(12), 2032–2038. doi: https://doi.org/10.1161/hq1201.100260

74. Van Heek, M., Farley, C., Compton, D. S., et al. (2000). Comparison of the activity and disposition of the novel cholesterol absorption inhibitor, SCH58235, and its glucuronide, SCH60663. *British Journal of Pharmacology*, *129*(8), 1748–1754. doi: https://doi.org/10.1038/sj.bjp.0703235

75. Altmann, S. W., Davis, H. R., Zhu, L. J., et al. (2004). Niemann-Pick C1 Like 1 Protein Is Critical for Intestinal Cholesterol Absorption. *Science*, *303*(5661), 1201–1204. doi: https://doi.org/10.1126/science.1093131

76. Garcia-Calvo, M., Lisnock, J. M., Bull, H. G., et al. (2005). The target of ezetimibe is Niemann-Pick C1-Like 1 (NPC1L1). *Proceedings of the National Academy of Sciences of the United States of America*, *102*(23), 8132–8137. doi: https://doi.org/10.1073/pnas.0500269102

77. Schulthess, G., Compassi, S., Boffelli, D., Werder, M., Weber, F. E., & Hauser, H. (1996). A comparative study of sterol absorption in different small-intestinal brush border membrane models. *Journal of Lipid Research*, *37*(11), 2405–2419. doi: https://doi.org/10.1016/s0022-2275(20)37489-7

78. Turley, S. D., & Dietschy, J. M. (2003). Sterol absorption by the small intestine. *Current Opinion in Lipidology*, *14*(3), 233–240. doi: https://doi.org/10.1097/00041433-200306000-00002

79. Jia, L., Ma, Y., Rong, S., et al. (2010). Niemann-Pick C1-Like 1 deletion in mice prevents high-fat diet-induced fatty liver by reducing lipogenesis. *Journal of Lipid Research*, *51*(11), 3135–3144. doi: https://doi.org/10.1194/jlr.M006353

80. Ference, B. A., Kastelein, J. J. P., & Catapano, A. L. (2020). Lipids and lipoproteins in 2020. *JAMA - Journal of the American Medical Association*, *324*(6), 595–596. doi: https://doi.org/10.1001/jama.2020.5685

81. Xie, C., Li, N., Chen, Z. J., Li, B. L., & Song, B. L. (2011). The small GTPase Cdc42 interacts with Niemann-Pick C1-like 1 (NPC1L1) and controls its movement from endocytic recycling compartment to plasma membrane in a cholesterol-dependent manner. *The Journal of Biological Chemistry*, *286*(41), 35933–35942. doi: https://doi.org/10.1074/jbc.M111.270199

82. Xiao, J., Dong, L. W., Liu, S., et al. (2023). Bile acids-mediated intracellular cholesterol transport promotes intestinal cholesterol absorption and NPC1L1 recycling. *Nature Communications*, *14*(1), 6469. doi: https://doi.org/10.1038/s41467-023-42179-5

83. Ge, L., Wang, J., Qi, W., et al. (2008). The Cholesterol Absorption Inhibitor Ezetimibe Acts by Blocking the Sterol-Induced Internalization of NPC1L1. *Cell Metabolism*, *7*(6), 508–519. doi: https://doi.org/10.1016/j.cmet.2008.04.001

84. Xie, C., Zhou, Z. Sen, Li, N., et al. (2012). Ezetimibe blocks the internalization of NPC1L1 and cholesterol in mouse small intestine. *Journal of Lipid Research*, *53*(10), 2092–2101. doi: https://doi.org/10.1194/jlr.M027359

85. Li, P. S., Fu, Z. Y., Zhang, Y. Y., et al. (2014). The clathrin adaptor Numb regulates intestinal cholesterol absorption through dynamic interaction with NPC1L1. *Nature Medicine*, *20*(1), 80–86. doi: https://doi.org/10.1038/nm.3417

86. Winkler, M. B. L., Kidmose, R. T., Szomek, M., et al. (2019). Structural Insight into Eukaryotic Sterol Transport through Niemann-Pick Type C Proteins. *Cell*, *179*(2), 485-497.e18. doi: https://doi.org/10.1016/j.cell.2019.08.038

87. Qian, H., Wu, X., Du, X., et al. (2020). Structural Basis of Low-pH-Dependent Lysosomal Cholesterol Egress by NPC1 and NPC2. *Cell*, *182*(1), 98-111.e18. doi: https://doi.org/10.1016/j.cell.2020.05.020

88. Zhang, Y., Bulkley, D. P., Xin, Y., et al. (2018). Structural Basis for Cholesterol Transport-like Activity of the Hedgehog Receptor Patched. *Cell*, *175*(5), 1352-1364.e14. doi: https://doi.org/10.1016/j.cell.2018.10.026

89. Cannac, F., Qi, C., Falschlunger, J., Hausmann, G., Basler, K., & Korkhov, V. M. (2020). Cryo-EM structure of the Hedgehog release protein Dispatched. *Science Advances*, *6*(16), eaay7928. doi: https://doi.org/10.1126/sciadv.aay7928

90. Moon, Y. A. (2017). The SCAP/SREBP pathway: A mediator of hepatic steatosis. *Endocrinology and Metabolism*, *32*(1), 6–10. doi: https://doi.org/10.3803/EnM.2017.32.1.6

91. Istvan, E. S., Palnitkar, M., Buchanan, S. K., & Deisenhofer, J. (2000). Crystal structure of the catalytic portion of human HMG-CoA reductase: Insights into regulation of activity and catalysis. *EMBO Journal*, *19*(5), 819–830. doi: https://doi.org/10.1093/emboj/19.5.819

92. Ge, L., Qi, W., Wang, L. J., et al. (2011). Flotillins play an essential role in Niemann-Pick C1-like 1-mediated cholesterol uptake. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(2), 551–556. doi: https://doi.org/10.1073/pnas.1014434108

93. Zhang, Y. Y., Fu, Z. Y., Wei, J., et al. (2018). A LIMA1 variant promotes low plasma LDL cholesterol and decreases intestinal cholesterol absorption. *Science*, *360*(6393), 1087–1092. doi: https://doi.org/10.1126/science.aao6575

94. Butt, E., Günder, T., Stürzebecher, P., et al. (2024). Cholesterol uptake in the intestine is regulated by the LASP1-AKT-NPC1L1 signaling pathway. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, *327*(1), G25–G35. doi: https://doi.org/10.1152/ajpgi.00222.2023

95. Yamanashi, Y., Takada, T., Kurauchi, R., Tanaka, Y., Komine, T., & Suzuki, H. (2017). Transporters for the Intestinal Absorption of Cholesterol, Vitamin E, and Vitamin K. *Journal of Atherosclerosis and Thrombosis*, *24*(4), 347–359. doi: https://doi.org/10.5551/jat.RV16007

96. Davis, H. R., Zhu, L. J., Hoos, L. M., et al. (2004). Niemann-Pick C1 like 1 (NPC1L1) is the intestinal phytosterol and cholesterol transporter and a key modulator of whole-body cholesterol homeostasis. *Journal of Biological Chemistry*, *279*(32), 33586–33592. doi: https://doi.org/10.1074/jbc.M405817200

97. Davies, J. P., Levy, B., & Ioannou, Y. A. (2000). Evidence for a Niemann-Pick C (NPC) gene family: Identification and characterization of NPC1L1. *Genomics*, *65*(2), 137–145. doi: https://doi.org/10.1006/geno.2000.6151

98. Las Heras, M., Szenfeld, B., Ballout, R. A., et al. (2023). Understanding the phenotypic variability in Niemann-Pick disease type C (NPC): a need for precision medicine. *Npj Genomic Medicine*, *8*(1), 21. doi: https://doi.org/10.1038/s41525-023-00365-w

99. Yamanashi, Y., Takada, T., & Suzuki, H. (2007). Niemann-pick C1-like 1 overexpression facilitates ezetimibe-sensitive cholesterol and β-sitosterol uptake in CaCo-2 cells. *Journal of Pharmacology and Experimental Therapeutics*, *320*(2), 559–564. doi: https://doi.org/10.1124/jpet.106.114181

100. Jia, L., Betters, J. L., & Yu, L. (2011). Niemann-Pick C1-Like 1 (NPC1L1) protein in intestinal and hepatic cholesterol transport. *Annual Review of Physiology*, *73*, 239–259. doi: https://doi.org/10.1146/annurev-physiol-012110-142233

101. Huang, C. S., Yu, X., Fordstrom, P., et al. (2020). Cryo-EM structures of NPC1L1 reveal mechanisms of cholesterol transport and ezetimibe inhibition. *Science Advances*, *6*(25), eabb1989. doi: https://doi.org/10.1126/sciadv.abb1989

102. Kwon, H. J., Palnitkar, M., & Deisenhofer, J. (2011). The structure of the NPC1L1 N-terminal domain in a closed conformation. *PLoS ONE*, *6*(4), e18722. doi: https://doi.org/10.1371/journal.pone.0018722

103. Long, T., Liu, Y., Qin, Y., DeBose-Boyd, R. A., & Li, X. (2024). Structures of dimeric human NPC1L1 provide insight into mechanisms for cholesterol absorption. *Science Advances*, *7*(34), eabh3997. doi: https://doi.org/10.1126/sciadv.abh3997

104. Hu, M., Yang, F., Huang, Y., et al. (2021). Structural insights into the mechanism of human NPC1L1-mediated cholesterol uptake. *Science Advances*, *7*(29), eabg3188. doi: https://doi.org/10.1126/sciadv.abg3188

105. Li, X., Lu, F., Trinh, M. N., et al. (2017). 3.3 Å structure of Niemann–Pick C1 protein reveals insights into the function of the C-terminal luminal domain in cholesterol transport. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(34), 9116–9121. doi: https://doi.org/10.1073/pnas.1711716114

106. Long, T., Qi, X., Hassan, A., Liang, Q., De Brabander, J. K., & Li, X. (2020). Structural basis for itraconazole-mediated NPC1 inhibition. *Nature Communications*, *11*(1), 152. doi: https://doi.org/10.1038/s41467-019-13917-5

107. Kwon, H. J., Abi-Mosleh, L., Wang, M. L., et al. (2009). Structure of N-Terminal Domain of NPC1 Reveals Distinct Subdomains for Binding and Transfer of Cholesterol. *Cell*, *137*(7), 1213–1224. doi: https://doi.org/10.1016/j.cell.2009.03.049

108. Zhang, J. H., Ge, L., Qi, W., et al. (2011). The N-terminal domain of NPC1L1 protein binds cholesterol and plays essential roles in cholesterol uptake. *Journal of Biological Chemistry*, *286*(28), 25088–25097. doi: https://doi.org/10.1074/jbc.M111.244475

109. Dias, C. S., Shaywitz, A. J., Wasserman, S. M., et al. (2012). Effects of AMG 145 on low-density lipoprotein cholesterol levels: Results from 2 randomized, double-blind, placebo-controlled, ascending-dose phase 1 studies in healthy volunteers and hypercholesterolemic subjects on statins. *Journal of the American College of Cardiology*, *60*(19), 1888–1898. doi: https://doi.org/10.1016/j.jacc.2012.08.986

110. Xu, Q., Deng, Y., Xiao, J., et al. (2020). Three Musketeers for Lowering Cholesterol: Statins, Ezetimibe and Evolocumab. *Current Medicinal Chemistry*, *28*(5), 1025–1041. doi: https://doi.org/10.2174/0929867327666200505091738

111. Chalasani, N., Aljadhey, H., Kesterson, J., Murray, M. D., & Hall, S. D. (2004). Patients with Elevated Liver Enzymes Are Not at Higher Risk for Statin Hepatotoxicity. *Gastroenterology*, *126*(5), 1287–1292. doi: https://doi.org/10.1053/j.gastro.2004.02.015

112. Nekohashi, M., Ogawa, M., Ogihara, T., et al. (2014). Luteolin and quercetin affect the cholesterol absorption mediated by epithelial cholesterol transporter Niemann-Pick C1-Like 1 in Caco-2 cells and rats. *PLoS ONE*, *9*(5), e97901. doi: https://doi.org/10.1371/journal.pone.0097901

113. Ogawa, M., Yamanashi, Y., Takada, T., Abe, K., & Kobayashi, S. (2017). Effect of luteolin on the expression of intestinal cholesterol transporters. *Journal of Functional Foods*, *36*, 274–279. doi: https://doi.org/10.1016/j.jff.2017.07.008

114. Wong, T. Y., Tan, Y. Q., Lin, S. mei, & Leung, L. K. (2017). Apigenin and luteolin display differential hypocholesterolemic mechanisms in mice fed a high-fat diet. *Biomedicine and Pharmacotherapy*, *96*, 1000–1007. doi: https://doi.org/10.1016/j.biopha.2017.11.131

115. Feng, D., Ohlsson, L., & Duan, R. D. (2010). Curcumin inhibits cholesterol uptake in Caco-2 cells by down-regulation of NPC1L1 expression. *Lipids in Health and Disease*, *9*(1), 40. doi: https://doi.org/10.1186/1476-511X-9-40

116. Liu, S., You, L., Zhao, Y., & Chang, X. (2018). Wild Lonicera caerulea berry polyphenol extract reduces cholesterol accumulation and enhances antioxidant capacity in vitro and in vivo. *Food Research International*, *107*, 73–83. doi: https://doi.org/10.1016/j.foodres.2018.02.016

117. Feng, J., Yang, J., Chang, Y., et al. (2019). Caffeine-free hawk tea lowers cholesterol by reducing free cholesterol uptake and the production of very-low-density lipoprotein. *Communications Biology*, *2*(1), 173. doi: https://doi.org/10.1038/s42003-019-0396-4

118. Wang, L., Zhu, H., Zhao, Y., et al. (2018). Cranberry anthocyanin as an herbal medicine lowers plasma cholesterol by increasing excretion of fecal sterols. *Phytomedicine*, *38*, 98–106. doi: https://doi.org/10.1016/j.phymed.2017.11.008

119. Liang, Y., Chen, J., Zuo, Y., et al. (2013). Blueberry anthocyanins at doses of 0.5 and 1 % lowered plasma cholesterol by increasing fecal excretion of acidic and neutral sterols in hamsters fed a cholesterol-enriched diet. *European Journal of Nutrition*, *52*(3), 869–875. doi: https://doi.org/10.1007/s00394-012-0393-6

120. Qin, B., Dawson, H. D., Schoene, N. W., Polansky, M. M., & Anderson, R. A. (2012). Cinnamon polyphenols regulate multiple metabolic pathways involved in insulin signaling and intestinal lipoprotein metabolism of small intestinal enterocytes. *Nutrition*, *28*(11–12), 1172–1179. doi: https://doi.org/10.1016/j.nut.2012.03.020

121. ZiQi Zhou, J., Kou, X., & Stevenson, D. (1999). Rapid extraction and high-performance liquid chromatographic determination of parthenolide in feverfew (Tanacetum parthenium). *Journal of Agricultural and Food Chemistry*, *47*(3), 1018–1022. doi: https://doi.org/10.1021/jf980798u

122. Liu, W., Liang, B., Zeng, J., et al. (2022). First Discovery of Cholesterol-Lowering Activity of Parthenolide as NPC1L1 Inhibitor. In *Molecules* (Vol. 27, Issue 19). doi: https://doi.org/10.3390/molecules27196270

123. Chiba, T., Sakurada, T., Watanabe, R., et al. (2014). Fomiroid A, a novel compound from the mushroom Fomitopsis nigra, inhibits NPC1L1-mediated cholesterol uptake via a mode of action distinct from that of ezetimibe. *PloS One*, *9*(12), e116162. doi: https://doi.org/10.1371/journal.pone.0116162

124. Alqahtani, S., Qosa, H., Primeaux, B., & Kaddoumi, A. (2015). Orlistat limits cholesterol intestinal absorption by Niemann-pick C1-like 1 (NPC1L1) inhibition. *European Journal of Pharmacology*, *762*, 263–269. doi: https://doi.org/10.1016/j.ejphar.2015.05.060

125. McMasters, D. R., Garcia-Calvo, M., Maiorov, V., et al. (2009). Spiroimidazolidinone NPC1L1 inhibitors. 1: Discovery by 3D-similarity-based virtual screening. *Bioorganic and Medicinal Chemistry Letters*, *19*(11), 2965–2968. doi: https://doi.org/10.1016/j.bmcl.2009.04.031

126. Howell, K. L., Devita, R. J., Garcia-Calvo, M., et al. (2010). Spiroimidazolidinone NPC1L1 inhibitors. Part 2: Structure-activity studies and in vivo efficacy. *Bioorganic and Medicinal Chemistry Letters*, *20*(23), 6929–6932. doi: https://doi.org/10.1016/j.bmcl.2010.09.138

127. Pfefferkorn, J. A., Larsen, S. D., Huis, C. Van, et al. (2008). Substituted oxazolidinones as novel NPC1L1 ligands for the inhibition of cholesterol absorption. *Bioorganic and Medicinal Chemistry Letters*, *18*(2), 546–553. doi: https://doi.org/10.1016/j.bmcl.2007.11.083

128. Pappas, G., Kiriaze, I. J., & Falagas, M. E. (2008). Insights into infectious disease in the era of Hippocrates. *International Journal of Infectious Diseases*, *12*(4), 347–350. doi: https://doi.org/10.1016/j.ijid.2007.11.003

129. Kuszewski, K., & Brydak, L. (2000). The epidemiology and history of influenza. *Biomedicine and Pharmacotherapy*, *54*(4), 188–195. doi: https://doi.org/10.1016/S0753-3322(00)89025-3

130. Barberis, I., Myles, P., Ault, S. K., Bragazzi, N. L., & Martini, M. (2016). History and evolution of influenza control through vaccination: From the first monovalent vaccine to universal vaccines. *Journal of Preventive Medicine and Hygiene*, *57*(3), E115–E120. doi: https://doi.org/10.15167/2421-4248/jpmh2016.57.3.642

131. Smith, F. B. (1995). The russian influenza in the United Kingdom, 1889-1894. *Social History of Medicine*, *8*(1), 55–73. doi: https://doi.org/10.1093/shm/8.1.55

132. Spreeuwenberg, P., Kroneman, M., & Paget, J. (2018). Reassessing the global mortality burden of the 1918 influenza pandemic. *American Journal of Epidemiology*, *187*(12), 2561–2567. doi: https://doi.org/10.1093/aje/kwy191

133. Gibbs, M. J., Armstrong, J. S., & Gibbs, A. J. (2001). Recombination in the hemagglutinin gene of the 1918 "Spanish flu". *Science (New York, N.Y.)*, *293*(5536), 1842–1845. doi: https://doi.org/10.1126/science.1061662

134. Worobey, M., Rambaut, A., Pybus, O. G., & Robertson, D. L. (2002). Questioning the evidence for genetic recombination in the 1918 "Spanish flu" virus. *Science*, *296*(5566), 211. doi: https://doi.org/10.1126/science.296.5566.211a

135. Fanning, T. G., Slemons, R. D., Reid, A. H., Janczewski, T. A., Dean, J., & Taubenberger, J. K. (2002). 1917 Avian Influenza Virus Sequences Suggest that the 1918 Pandemic Virus Did Not Acquire Its Hemagglutinin Directly from Birds. *Journal of Virology*, *76*(15), 7860–7862. doi: https://doi.org/10.1128/jvi.76.15.7860-7862.2002

136. Zambon, M. C. (2001). The pathogenesis of influenza in humans. *Reviews in Medical Virology*, *11*(4), 227–241. doi: https://doi.org/10.1002/rmv.319

137. Iuliano, A. D., Roguski, K. M., Chang, H. H., et al. (2018). Estimates of global seasonal influenza-associated respiratory mortality: a modelling study. *The Lancet*, *391*(10127), 1285–1300. doi: https://doi.org/10.1016/S0140-6736(17)33293-2

138. Saunders-Hastings, P. R., & Krewski, D. (2016). Reviewing the history of pandemic influenza: Understanding patterns of emergence and transmission. *Pathogens*, *5*(4). doi: https://doi.org/10.3390/pathogens5040066

139. Hammond, A., Kim, J. J., Sadler, H., & Vandemaele, K. (2022). Influenza surveillance systems using traditional and alternative sources of data: A scoping review. *Influenza and Other Respiratory Viruses*, *16*(6), 965–974. doi: https://doi.org/10.1111/irv.13037

140. Harrington, W. N., Kackos, C. M., & Webby, R. J. (2021). The evolution and future of influenza pandemic preparedness. *Experimental and Molecular Medicine*, *53*(5), 737–749. doi: https://doi.org/10.1038/s12276-021-00603-0

141. Kilbourne, E. D. (1987). Taxonomy and Comparative Virology of the Influenza Viruses. In E. D. Kilbourne (Ed.), *Influenza* (p. 25–32). Springer US. doi: https://doi.org/10.1007/978-1-4684-5239-6_2

142. Chu, C. M., Dawson, I. M., & Elford, W. J. (1949). Filamentous Forms Associated With Newly Isolated Influenza Virus. *The Lancet*, *253*(6554), 602. doi: https://doi.org/10.1016/S0140-6736(49)91699-2

143. Kilbourne, E. D., & Murphy, J. S. (1960). Genetic studies of influenza viruses. I. Viral morphology and growth capacity as exchangeable genetic traits. Rapid in ovo adaptation of early passage Asian strain isolates by combination with PR8. *The Journal of Experimental Medicine*, *111*(3), 387–406. doi: https://doi.org/10.1084/jem.111.3.387

144. Harris, A., Cardone, G., Winkler, D. C., et al. (2006). Influenza virus pleiomorphy characterized by cryoelectron tomography. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(50), 19123–19127. doi: https://doi.org/10.1073/pnas.0607614103

145. O'Neill, R. E., Talon, J., & Palese, P. (1998). The influenza virus NEP (NS2 protein) mediates the nuclear export of viral ribonucleoproteins. *EMBO Journal*, *17*(1), 288–296. doi: https://doi.org/10.1093/emboj/17.1.288

146. Eisfeld, A. J., Neumann, G., & Kawaoka, Y. (2015). At the centre: influenza A virus ribonucleoproteins. *Nature Reviews Microbiology*, *13*(1), 28–41. doi: https://doi.org/10.1038/nrmicro3367

147. Pinto, R. M., Lycett, S., Gaunt, E., & Digard, P. (2021). Accessory Gene Products of Influenza A Virus. *Cold Spring Harbor Perspectives in Medicine*, *11*(12), a038380. doi: https://doi.org/10.1101/CSHPERSPECT.A038380

148. Chlanda, P., Schraidt, O., Kummer, S., et al. (2015). Structural Analysis of the Roles of Influenza A Virus Membrane-Associated Proteins in Assembly and Morphology. *Journal of Virology*, *89*(17), 8957–8966. doi: https://doi.org/10.1128/jvi.00592-15

149. Ghazimirsaeed, S. E., Khavani, M., & Mofrad, M. (2024). Exploring sialic acid derivative inhibitors targeting influenza virus hemagglutinin (HA) protein. *Biophysical Journal*, *123*(3), 471a–472a. doi: https://doi.org/10.1016/j.bpj.2023.11.2866

150. Chen, J., Lee, K. H., Steinhauer, D. A., Stevens, D. J., Skehel, J. J., & Wiley, D. C. (1998). Structure of the hemagglutinin precursor cleavage site, a determinant of influenza pathogenicity and the origin of the labile conformation. *Cell*, *95*(3), 409–417. doi: https://doi.org/10.1016/S0092-8674(00)81771-7

151. White, J. M., & Whittaker, G. R. (2016). Fusion of Enveloped Viruses in Endosomes. *Traffic*, *17*(6), 593–614. doi: https://doi.org/10.1111/tra.12389

152. Wallace, L. E., de Vries, E., van Kuppeveld, F. J. M., & de Haan, C. A. M. (2023). Neuraminidase-dependent entry of influenza A virus is determined by hemagglutinin receptor-binding specificity. *Journal of Virology*, *97*(10), e0060223. doi: https://doi.org/10.1128/jvi.00602-23

153. Peters, C. J., & Calisher, C. H. (2005). *Infectious Diseases from Nature: Mechanisms of Viral Emergence and Persistence*. Springer Vienna. doi: https://doi.org/10.1007/3-211-29981-5

154. Portela, A., & Digard, P. (2002). The influenza virus nucleoprotein: A multifunctional RNA-binding protein pivotal to virus replication. *Journal of General Virology*, *83*(4), 723–734. doi: https://doi.org/10.1099/0022-1317-83-4-723

155. Zebedee, S. L., & Lamb, R. A. (1988). Influenza A virus M2 protein: monoclonal antibody restriction of virus growth and detection of M2 in virions. *Journal of Virology*, *62*(8), 2762–2772. doi: https://doi.org/10.1128/jvi.62.8.2762-2772.1988

156. Lutz, M. M., Dunagan, M. M., Kurebayashi, Y., & Takimoto, T. (2020). Key role of the influenza A virus PA gene segment in the emergence of pandemic viruses. *Viruses*, *12*(4), 365. doi: https://doi.org/10.3390/v12040365

157. Hale, B. G., Randall, R. E., Ortin, J., & Jackson, D. (2008). The multifunctional NS1 protein of influenza A viruses. *Journal of General Virology*, *89*(10), 2359–2376. doi: https://doi.org/10.1099/vir.0.2008/004606-0

158. Jang, Y. H., & Seong, B. L. (2019). The Quest for a Truly Universal Influenza Vaccine. *Frontiers in Cellular and Infection Microbiology*, *9*. doi: https://doi.org/10.3389/fcimb.2019.00344

159. Fereidouni, S., Starick, E., Karamendin, K., et al. (2023). Genetic characterization of a new candidate hemagglutinin subtype of influenza A viruses. *Emerging Microbes and Infections*, *12*(2), 2225645. doi: https://doi.org/10.1080/22221751.2023.2225645

160. Zhu, Z., Fodor, E., & Keown, J. R. (2023). A structural understanding of influenza virus genome replication. *Trends in Microbiology*, *31*(3), 308–319. doi: https://doi.org/10.1016/j.tim.2022.09.015

161. Nayak, D. P., Balogun, R. A., Yamada, H., Zhou, Z. H., & Barman, S. (2009). Influenza virus morphogenesis and budding. *Virus Research*, *143*(2), 147–161. doi: https://doi.org/10.1016/j.virusres.2009.05.010

162. Hutchinson, E. C., von Kirchbach, J. C., Gog, J. R., & Digard, P. (2010). Genome packaging in influenza A virus. *Journal of General Virology*, *91*(2), 313–328. doi: https://doi.org/10.1099/vir.0.017608-0

163. Gerhard, W. (2001). The role of the antibody response in influenza virus infection. *Current Topics in Microbiology and Immunology*, *260*, 171–190. doi: https://doi.org/10.1007/978-3-662-05783-4_9

164. Henle, W., Henle, G., Hampil, B., Maris, E. P., & Stokes, J. (1946). Experiments on Vaccination of Human Beings against Epidemic Influenza. *The Journal of Immunology*, *53*(1), 75–93. doi: https://doi.org/10.4049/jimmunol.53.1.75

165. Caini, S., Meijer, A., Nunes, M. C., et al. (2024). Probable extinction of influenza B/Yamagata and its public health implications: a systematic literature review and assessment of global surveillance databases. *The Lancet Microbe*, *5*(8). doi: https://doi.org/10.1016/S2666-5247(24)00066-1

166. Jordan, K., Murchu, E. O., Comber, L., et al. (2023). Systematic review of the efficacy, effectiveness and safety of cell-based seasonal influenza vaccines for the prevention of laboratory-confirmed influenza in individuals ≥18 years of age. *Reviews in Medical Virology*, *33*(3), e2332. doi: https://doi.org/10.1002/rmv.2332

167. Ginex, T., & Luque, F. J. (2021). Searching for effective antiviral small molecules against influenza A virus: A patent review. *Expert Opinion on Therapeutic Patents*, *31*(1), 53–66. doi: https://doi.org/10.1080/13543776.2020.1831471

168. Aledavood, E., Selmi, B., Estarellas, C., Masetti, M., & Luque, F. J. (2022). From Acid Activation Mechanisms of Proton Conduction to Design of Inhibitors of the M2 Proton Channel of Influenza A Virus. *Frontiers in Molecular Biosciences*, *8*, 796229. doi: https://doi.org/10.3389/fmolb.2021.796229

169. Kumari, R., Sharma, S. D., Kumar, A., et al. (2023). Antiviral Approaches against Influenza Virus. *Clinical Microbiology Reviews*, *36*(1), e0004022. doi: https://doi.org/10.1128/cmr.00040-22

170. Aoki, F. Y. (2014). Antiviral Drugs for Influenza and Other Respiratory Virus Infections. In J. E. Bennett, R. Dolin, & M. J. B. T. M. Blaser Douglas, and Bennett's Principles and Practice of Infectious Diseases (Eighth Edition), *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases* (Vol. 1, p. 531–545). W.B. Saunders. doi: https://doi.org/10.1016/B978-1-4557-4801-3.00044-8

171. Takashita, E. (2021). Influenza polymerase inhibitors: Mechanisms of action and resistance. *Cold Spring Harbor Perspectives in Medicine*, *11*(5):a038687. doi: https://doi.org/10.1101/cshperspect.a038687

172. Russell, R. J., Kerry, P. S., Stevens, D. J., et al. (2008). Structure of influenza hemagglutinin in complex with an inhibitor of membrane fusion. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(46), 17736–17741. doi: https://doi.org/10.1073/pnas.0807142105

173. Kadam, R. U., & Wilson, I. A. (2017). Structural basis of influenza virus fusion inhibition by the antiviral drug Arbidol. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(2), 206–214. doi: https://doi.org/10.1073/pnas.1617020114

174. Blaising, J., Polyak, S. J., & Pécheur, E. I. (2014). Arbidol as a broad-spectrum antiviral: An update. *Antiviral Research*, *107*(1), 84–94. doi: https://doi.org/10.1016/j.antiviral.2014.04.006

175. Vanderlinden, E., Göktaş, F., Cesur, Z., et al. (2010). Novel Inhibitors of Influenza Virus Fusion: Structure-Activity Relationship and Interaction with the Viral Hemagglutinin. *Journal of Virology*, *84*(9), 4277–4288. doi: https://doi.org/10.1128/jvi.02325-09

176. Laursen, N. S., & Wilson, I. A. (2013). Broadly neutralizing antibodies against influenza viruses. *Antiviral Research*, *98*(3), 476–483. doi: https://doi.org/10.1016/j.antiviral.2013.03.021

177. Van Dongen, M. J. P., Kadam, R. U., Juraszek, J., et al. (2019). A small-molecule fusion inhibitor of influenza virus is orally active in mice. *Science*, *363*(6431), eaar6221. doi: https://doi.org/10.1126/science.aar6221

178. Yao, Y., Kadam, R. U., Lee, C. C. D., et al. (2020). An influenza A hemagglutinin small-molecule fusion inhibitor identified by a new high-throughput fluorescence polarization screen. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(31), 18431–18438. doi: https://doi.org/10.1073/pnas.2006893117

179. Antanasijevic, A., Durst, M. A., Cheng, H., et al. (2020). Structure of avian influenza hemagglutinin in complex with a small molecule entry inhibitor. *Life Science Alliance*, *3*(8), e202000724. doi: https://doi.org/10.26508/LSA.202000724

180. Wang, A., Li, Y., Lv, K., et al. (2021). Optimization and SAR research at the piperazine and phenyl rings of JNJ4796 as new anti-influenza A virus agents, part 1. *European Journal of Medicinal Chemistry*, *222*, 113591. doi: https://doi.org/10.1016/j.ejmech.2021.113591

181. Wu, W., Yan, H., Jiang, B., et al. (2023). Optimization and SAR research at the benzoxazole and tetrazole rings of JNJ4796 as new anti-influenza A virus agents, part 2. *European Journal of Medicinal Chemistry*, *245*(Pt 1), 114906. doi: https://doi.org/10.1016/j.ejmech.2022.114906

182. Hussein, A. F. A., Cheng, H., Tundup, S., et al. (2020). Identification of entry inhibitors with 4-aminopiperidine scaffold targeting group 1 influenza A virus. *Antiviral Research*, *177*, 104782. doi: https://doi.org/10.1016/j.antiviral.2020.104782

183. Zhao, X., Li, R., Zhou, Y., et al. (2018). Discovery of Highly Potent Pinanamine-Based Inhibitors against Amantadine- and Oseltamivir-Resistant Influenza A Viruses. *Journal of Medicinal Chemistry*, *61*(12), 5187–5198. doi: https://doi.org/10.1021/acs.jmedchem.8b00042

184. Corbella, M., Pinto, G. P., & Kamerlin, S. C. L. (2023). Loop dynamics and the evolution of enzyme activity. *Nature Reviews Chemistry*, *7*(8), 536–547. doi: https://doi.org/10.1038/s41570-023-00495-w

185. Pan, X., Wang, H., Li, C., Zhang, J. Z. H., & Ji, C. (2021). MolGpka: A Web Server for Small Molecule pKa Prediction Using a Graph-Convolutional Neural Network. *Journal of Chemical Information and Modeling*, *61*(7), 3159–3165. doi: https://doi.org/10.1021/acs.jcim.1c00075

186. Wu, Y., Canturk, B., Jo, H., et al. (2014). Flipping in the Pore: Discovery of Dual Inhibitors That Bind in Different Orientations to the Wild-Type versus the Amantadine-Resistant S31N Mutant of the Influenza A Virus M2 Proton Channel. *Journal of the American Chemical Society*, *136*(52), 17987–17995. doi: https://doi.org/10.1021/ja508461m

187. Plotch, S. J., O'Hara, B., Morin, J., et al. (1999). Inhibition of influenza A virus replication by compounds interfering with the fusogenic function of the viral hemagglutinin. *Journal of Virology*, *73*(1), 140–151. doi: https://doi.org/10.1128/JVI.73.1.140-151.1999

188. Liu, S., Li, R., Zhang, R., et al. (2011). CL-385319 inhibits H5N1 avian influenza A virus infection by blocking viral entry. *European Journal of Pharmacology*, *660*(2–3), 460. doi: https://doi.org/10.1016/j.ejphar.2011.04.013

189. Zhu, Z., Li, R., Xiao, G., et al. (2012). Design, synthesis and structure-activity relationship of novel inhibitors against H5N1 hemagglutinin-mediated membrane fusion. *European Journal of Medicinal Chemistry*, *57*, 211–216. doi: https://doi.org/10.1016/j.ejmech.2012.08.041

190. Zhu, Z., Yao, Z., Shen, X., et al. (2017). Oligothiophene compounds inhibit the membrane fusion between H5N1 avian influenza virus and the endosome of host cell. *European Journal of Medicinal Chemistry*, *130*, 185–194. doi: https://doi.org/10.1016/j.ejmech.2017.02.040

191. Leiva, R., Barniol-Xicota, M., Codony, S., et al. (2018). Aniline-Based Inhibitors of Influenza H1N1 Virus Acting on Hemagglutinin-Mediated Fusion. *Journal of Medicinal Chemistry*, *61*(1), 98–118. doi: https://doi.org/10.1021/acs.jmedchem.7b00908

192. Dreyfus, C., Ekiert, D. C., & Wilson, I. A. (2013). Structure of a classical broadly neutralizing stem antibody in complex with a pandemic H2 influenza virus hemagglutinin. *Journal of Virology*, *87*(12), 7149–7154. doi: https://doi.org/10.1128/JVI.02975-12

193. S., T. G., Florian, K., Dirk, E., Alita, K., M., M. T., & Peter, P. (2012). A Pan-H1 Anti-Hemagglutinin Monoclonal Antibody with Potent Broad-Spectrum Efficacy In Vivo . *Journal of Virology*, *86*(11), 6179–6188. doi: https://doi.org/10.1128/jvi.00469-12

194. Yassine, H. M., Boyington, J. C., McTamney, P. M., et al. (2015). Hemagglutinin-stem nanoparticles generate heterosubtypic influenza protection. *Nature Medicine*, *21*(9), 1065–1070. doi: https://doi.org/10.1038/nm.3927

Other Publications

"Influenza A virus hemagglutinin: from classical fusion inhibitors to proteolysis targeting chimera-based strategies in antiviral drug discovery"

Francisco Javier Hermoso Pinilla,[1] Aitor Valdivia,[2] Maria-José Camarasa,[3*] Tiziana Ginex,[4,*] and Francisco Javier Luque[1,2,*]

[1] Departament de Nutrició, Ciències de l'Alimentació I Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació - Campus Torribera, Institut de Biomedicina (IBUB), Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain

[2] Departament de Nutrició, Ciències de l'Alimentació I Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació - Campus Torribera, Institut de Química Teòrica i Computacional (IQTCUB), Universitat de Barcelona, 08921 Santa Coloma de Gramenet, Spain

[3] Instituto de Química Médica (IQM, CSIC), 28006 Madrid, Spain

[4] Pharmacelera, Parc Científic de Barcelona (PCB), Baldiri-Reixac 4-8, 08028 Barcelona, Spain

[*]**Correspondence:** Maria-José Camarasa, Instituto de Química Médica (IQM, CSIC), Juan de la Cierva 3, 28006 Madrid, Spain. mj.camarasa@iqm.csic.es; Tiziana Ginex, Pharmacelera, Parc Científic de Barcelona (PCB), Baldiri Reixac 4-8, 08028 Barcelona, Spain, 08028 Barcelona, Spain. tiziana.ginex@pharmacelera.com; Francisco Javier Luque, Departament de Nutrició, Ciències de l'Alimentació i Gastronomia, Facultat de Farmàcia i Ciències de l'Alimentació, Institut de Química Teòrica i Computacional (IQTCUB), Prat de la Riba 171, Universitat de Barcelona, 08021 Santa Coloma de Gramenet, Spain. fjluque@ub.edu

Published in: Exploration of Drug Science, **2024**, 2, 85-116.

Hermoso-Pinilla, F. J., Valdivia, A., Camarasa, M. J., Ginex, T., & Luque, F. J. (2024). Influenza A virus hemagglutinin: from classical fusion inhibitors to proteolysis targeting chimera-based strategies in antiviral drug discovery. Exploration of Drug Science, 2(1), 85–116. doi: https://doi.org/10.37349/eds.2024.00037



Now Hávi's sayings have been spoken in Hávi's hall,

very useful to the sons of men,

but useless to the sons of giants!

Hail he who spoke! Hail to him who knows!

May he who learned benefit!

Hail to those who heard!

Hár (Hávamál, stanza 164)

