



TECHNISCHE UNIVERSITÄT MÜNCHEN  
Fakultät für Chemie

# In Silico Structure-based Approaches to Design Mdmx Inhibitors

Zhonghua Xia

Vollständiger Abdruck der von der Fakultät für Chemie der Technischen  
Universität München zur Erlangung des akademischen Grades einer  
**Doktorin der Naturwissenschaften (Dr. rer. Nat.)**  
genehmigten Dissertation.

Vorsitzende: Prof. Dr. Angela Casini

Prüfer der Dissertation:

1. Dr. Igor V. Tetko
2. Prof. Dr. Michael Sattler

Die Dissertation wurde am 27.01.2022 bei der Technischen Universität München eingereicht  
und durch die Fakultät für Chemie am 16.02.2022 angenommen.

# Abstract

The rapid development of computer technologies has greatly increased their utility in drug research. Computational approaches accelerate many aspects of drug discovery and enable the exploration of chemical space beyond the traditional knowledge of experienced chemists. These methods have also been actively explored to develop small molecule-based drugs to fight many human diseases with strong medical needs. Mdmx, a promising anticancer target, is a negative regulator of the p53 tumour suppressor. Despite the high interest and potential as anticancer targets and decades of research, the design of Mdmx inhibitors is still poorly advanced compared to its homologous protein Mdm2.

This dissertation explored strategies to improve the design of small-molecule Mdmx inhibitors that target protein-protein interactions. The main goals were to (i) investigate the reasons for the difficulties in the design of Mdmx inhibitors, (ii) propose new putative inhibitors based on known active compounds and experimentally measured fragments, (iii) provide the possibility of *de novo* creation of new inhibitors by machine learning methods, and (iv) discover the features of ligands responsible for opening transient pockets to propose novel Mdmx inhibitors.

The first part of the thesis is dedicated to the comparison of differences in protein primary structures between Mdmx and Mdm2 by in-depth analyses of their experimental crystal structures using computer simulations. The study revealed that a methionine side chain located at different positions in the main binding pocket (Met50 in Mdm2 and Met53 in Mdmx) correlates with different orientations of a tyrosine, corresponding to ‘open’ and ‘closed’ states of the binding pocket. The ‘open’ state yields an enlarged pocket and provides a transient sub-pocket. By inspecting the conformations from crystal structures and molecular dynamics (MD) simulations, an interaction of the Met and Tyr residues with the ligand restricted the conformation of the tyrosine. Met50 interacts with Tyr100 in Mdm2 to adopt an open state, while the corresponding Met53 and Tyr99 in Mdmx promote the closed state. However, it could not be completely excluded that Mdmx was also able to bind ligands in a transiently sampled ‘open’ state, and thus both states were considered for further analysis.

In the second part, new Mdmx inhibitors were developed to mimic the binding mode of p53 from known active compounds using two structure-based strategies. One strategy was to optimise the structure of the known Mdmx inhibitor WK298; another approach was to rationally connect active fragments to fully occupy the binding pocket. The binding modes of the inhibitors were obtained and evaluated using MD simulations. Interestingly, the ligands show binding to both ‘open’ and ‘closed’ states of Mdmx. Four molecules were synthesised and tested to be inactive toward Mdm2. The lesson drawn for the future would be that a 10-ns trajectory was not sufficient to obtain a stable binding mode for an Mdmx-small molecule ligand system.

In the third project, state-of-the-art machine learning methods were used to design Mdmx inhibitors. The rapid progress of natural language processing (NLP) machine learning in chemistry has opened new avenues for designing compounds with desired properties. To contribute to these studies, an NLP model that generated *de novo* molecules was developed, referred to as the Gen-

erator model. Combined with a QSAR model to predict the  $IC_{50}$  of Mdmx inhibitors as well as models to predict physicochemical properties, the Generator generated potential Mdmx inhibitors with completely new chemical scaffolds but also re-discovered structures available in public databases. The most promising molecules had better predicted  $pIC_{50}$  than WK298 and constantly kept the ‘closed’ state of Mdmx. In contrast, WK298 induced Mdmx to present half ‘open’ and half ‘closed’ conformations. These findings highlight the importance of considering the presence of multiple conformations of the binding pocket for the discovery of inhibitors of Mdmx, but also for other drug targets in general.

Finally, the chemical features of ligands that are relevant to the induction of cryptic pockets were investigated, which could be important to design Mdmx inhibitor binding in the ‘open’ conformation. An automated program was developed to establish a holo-apo pair dataset, which accurately matched apo proteins for each holo (bound with a small-molecule ligand) protein. The volume changes between apo and holo proteins were used to differentiate the chemical features between inducers and non-inducers. Based on the holo-apo pair dataset, classification models were built to determine the optimum threshold. The model analysis indicated that inducers were more hydrophobic and aromatic and showed over-representation of phosphorus and halogen atoms. Fragment analysis showed that small changes in the structure of molecules could strongly affect the potential to induce a cryptic pocket. Most of the Mdmx inhibitors available in the literature were classified as inducers, thus indicating that the classical paradigm to design these inhibitors was targeting the ‘open’ state.

The thesis presented here provides new perspectives on structure-based approaches for designing Mdmx inhibitors, highlighting the importance of considering transient, alternative conformations and ‘closed’ states for drug design. New inhibitors have been proposed either directly optimised from known active compounds or de novo created by the Generator model, which will be promising for experimental validation of their inhibitory activities in the future. The proposed Inducer model provides novel opportunities to discover new inhibitors that could induce the opening of transient binding pockets. The Generator and Inducer models are not limited to Mdmx but are expected to provide useful tools for structure-based drug discovery in general.

## Zusammenfassung

Die rasche Entwicklung der Computertechnologien erhöht ihren Nutzen für die Arzneimittelforschung erheblich. Computergestützte Ansätze beschleunigen der Arzneimittelforschung in vielen Aspekten und ermöglichen die Erforschung des chemischen Raums über das traditionelle Wissen erfahrener Chemiker hinaus. Ebenso wird der Einsatz dieser Methoden zur aktiven Entwicklung von auf kleinen Molekülen basierender Arzneimittel zur Bekämpfung vieler medizinisch dringender menschlicher Krankheiten erkundet. Mdmx, ein vielversprechendes Zielmolekül gegen Krebs, ist ein negativer Regulator des Tumorsuppressors p53. Trotz des großen Interesses und Potenzials als Anti-Krebs Zielobjekt und jahrzehntelanger Forschung ist die Entwicklung von Mdmx-Inhibitoren im Vergleich zum homologen Protein Mdm2 noch nicht weit fortgeschritten.

In dieser Dissertation wurden Strategien zur Verbesserung des Designs von niedermolekularen Mdmx-Inhibitoren für die Protein-Protein-Interaktionen untersucht. Die Hauptziele waren, (i) die Gründe für die Schwierigkeiten bei der Entwicklung von Mdmx-Inhibitoren zu untersuchen, (ii) neue mutmaßliche Inhibitoren auf der Grundlage bekannter aktiver Verbindungen und experimentell gemessener Fragmente vorzuschlagen, (iii) eine Möglichkeit zu schaffen, neue Inhibitoren mit Hilfe von Methoden des maschinellen Lernens *de novo* zu entwickeln und (iv) Ligandeneigenschaften, die für die Öffnung von transienten Taschen verantwortlich sind, zu finden um neue Mdmx-Inhibitoren vorzuschlagen.

Der erste Teil der Arbeit widmet sich dem Vergleich der Unterschiede in den Proteinprimärstrukturen von Mdmx und Mdm2 durch eingehende Analysen ihrer experimentell ermittelten Kristallstrukturen mit Hilfe von Computersimulationen. Die Studie ergab, dass eine Methionin-Seitenkette, die sich an verschiedenen Positionen in der Hauptbindungstasche befindet (Met50 in Mdm2 und Met53 in Mdmx), mit verschiedenen Ausrichtungen eines Tyrosins korreliert, die einem „offenen“ und einem „geschlossenen“ Zustand der Bindungstasche entsprechen. Der „offene“ Zustand führt zu einer vergrößerten Tasche und bietet eine transiente Subtasche. Die Untersuchung der Konformationen aus Kristallstrukturen und Molekulardynamik (MD)-Simulationen ergab, dass eine Wechselwirkung der Met- und Tyr-Reste mit dem Liganden die Konformation des Tyrosins einschränkt. Met50 interagiert mit Tyr100 in Mdm2, um einen offenen Zustand einzunehmen, während das entsprechende Met53 und Tyr100 in Mdmx den geschlossenen Zustand fördern. Es konnte nicht völlig ausgeschlossen werden, dass Mdmx auch in der Lage ist, Liganden in einem vorübergehend abgetasteten „offenen“ Zustand zu binden, und so wurden beide Zustände für die weitere Analyse berücksichtigt.

Im zweiten Teil wurden neue Mdmx-Inhibitoren entwickelt, um den Bindungsmodus von p53 aus bekannten aktiven Verbindungen durch zwei strukturbasierte Strategien nachzuahmen. Ein Ansatz bestand darin, die Struktur des bekannten Mdmx-Inhibitors WK298 zu optimieren; ein anderer aktive Fragmente rational zu verbinden, um die Bindungstasche vollständig zu besetzen. Die Bindungsmodi der Inhibitoren wurden anhand von MD-Simulationen ermittelt und bewertet. Interessanterweise binden die Liganden sowohl an den „offenen“ als auch an den „geschlossenen“ Zustand von Mdmx. Mithilfe von vier synthetisierten Molekülen wurde in Tests



gezeigt, dass sie gegenüber Mdm2 inaktiv sind. Daraus ergibt sich für zukünftige Forschung, dass eine 10-ns-Trajektorie nicht ausreicht, um den stabilen Bindungsmodus für ein Mdmx-Ligandensystem mit kleinen Molekülen zu erhalten.

In einem dritten Projekt wurden modernste Methoden des maschinellen Lernens eingesetzt, um Mdmx-Inhibitoren zu entwickeln. Die rasanten Fortschritte im Bereich des maschinellen Lernens mit linguistischer Datenverarbeitung – natural language processing (NLP) – in der Chemie eröffnen neue Perspektiven für das Design von Verbindungen mit gewünschten Eigenschaften. Mit dieser Doktorarbeit wurde dazu beigetragen, indem ein NLP-Modell entwickelt wurde welches *de novo*-Moleküle generiert, das so genannte Generator-Modell. In Kombination mit einem QSAR-Modell zur Vorhersage der  $IC_{50}$  von Mdmx-Inhibitoren sowie mit Modellen zur Vorhersage physikalisch-chemischer Eigenschaften generierte der Generator potenzielle Mdmx-Inhibitoren mit völlig neuen chemischen als auch neu entdeckten Strukturen, aus öffentlich verfügbaren Datenbanken. Die vielversprechendsten Moleküle hatten einen besseren vorhergesagten  $pIC_{50}$  als WK298 und hielten den „geschlossenen“ Zustand von Mdmx konstant. WK298 veranlasste Mdmx im Gegensatz dazu, halb „offene“ und halb „geschlossene“ Konformationen zu zeigen. Diese Ergebnisse zeigen, wie wichtig es ist das Vorhandensein mehrerer Konformationen der Bindungstasche bei der Entdeckung von Hemmstoffen für Mdmx aber auch für andere Wirkstofftargets im Allgemeinen zu berücksichtigen.

Schließlich wurde untersucht welche chemischen Merkmale von Liganden für die induzierenden kryptischen Taschen relevant sind, da diese für die Entwicklung von Mdmx-Inhibitoren, die in der „offenen“ Umgebung binden, wichtig sein könnten. Es wurde ein automatisiertes Programm entwickelt, um einen Holo-Apo-Paar-Datensatz zu erstellen, der jedem Holo-Protein (das an einen niedermolekularen Liganden gebunden ist) genau passende Apo-Proteine zuordnete. Die Volumenänderungen zwischen Apo- und Holo-Protein wurden zur Unterscheidung chemischer Merkmale bei Induktoren und Nicht-Induktoren verwendet. Auf der Grundlage des Holo-Apo-Paar-Datensatzes wurden Klassifizierungsmodelle erstellt, um einen optimalen Schwellenwert zu bestimmen. Die Modellanalyse ergab, dass Induktoren hydrophober und aromatischer sind und Phosphor- und Halogenatome überrepräsentiert sind. Die Fragmentanalyse zeigte, dass kleine Änderungen in der Struktur von Molekülen das Potenzial zur Induktion einer kryptischen Tasche stark beeinflussen können. Die meisten in der Literatur verfügbaren Mdmx-Inhibitoren wurden als Induktoren klassifiziert, was darauf hindeutet, dass das klassische Paradigma zur Entwicklung dieser Inhibitoren auf den „offenen“ Zustand abzielte.

Die hier vorgestellte Arbeit bietet neue Perspektiven für strukturbasierte Ansätze zur Entwicklung von Mdmx-Inhibitoren und unterstreicht, wie wichtig es ist vorübergehende alternative Konformationen und „geschlossene“ Zustände bei der Entwicklung von Medikamenten zu berücksichtigen. Es werden neue Inhibitoren vorgeschlagen, die entweder direkt aus bekannten aktiven Verbindungen optimiert oder *de novo* durch das Generator-Modell erzeugt werden, was für die experimentelle Validierung ihrer inhibitorischen Aktivitäten in der Zukunft vielversprechend sein wird. Das vorgestellte Inducer-Modell bietet neue Möglichkeiten zur Entdeckung neuer Inhibitoren, welche die Öffnung von vorübergehenden Bindungstaschen induzieren könnten. Die Generator- und Inducer-Modelle sind nicht auf Mdmx beschränkt, sondern werden vo-

raussichtlich nützliche Werkzeuge für die strukturbasierte Arzneimittelforschung im Allgemeinen liefern.

# Table of Contents

Abstract .....	I
Zusammenfassung .....	III
Table of Contents .....	VI
Abbreviations .....	IX
1 Introduction .....	1
1.1 Background .....	1
1.2 Aims .....	1
2 Methods .....	3
2.1 Molecular docking .....	3
2.2 Conventional molecular dynamics (MD) simulation .....	4
2.3 Binding free energies .....	5
2.4 OCHEM .....	6
2.5 Schrodinger .....	6
3 Mdmx pocket dynamics .....	7
3.1 Introduction .....	7
3.2 Methods .....	8
3.2.1 Molecular docking .....	9
3.2.2 MD simulations .....	9
3.3 Results .....	10
3.3.1 Trajectory analysis .....	10
3.3.2 Key residues analysis .....	14
3.4 Summary and outlook .....	16
4 Developing new Mdmx inhibitors using structure-based drug discovery .....	17
4.1 Introduction .....	17
4.2 Methods .....	19
4.2.1 Molecular docking .....	19
4.2.2 MD simulation .....	20
4.2.3 Bioactivity test .....	20
4.3 Results .....	21
4.3.1 Strategy 1 .....	21
4.3.2 Strategy 2 .....	27

---

4.4	Summary and outlook.....	38
5	<i>De novo</i> design new Mdmx inhibitors by machine learning methods.....	40
5.1	Motivation .....	40
5.2	Methods .....	41
5.2.1	Datasets .....	41
5.2.2	SMILES generator neural network (SGNN).....	42
5.2.3	Molecular docking.....	43
5.2.4	QSAR models for IC <sub>50</sub> and solubility estimation.....	43
5.2.5	Tuning the Generator .....	43
5.2.6	Virtual screening .....	44
5.3	Results .....	45
5.3.1	The Generator.....	45
5.3.2	Putative Mdmx inhibitors via virtual screening workflow.....	47
5.3.3	Comprehensive analysis on ChEMBL compounds re-discovered by Generator....	51
5.3.4	Constraints during tuning .....	55
5.4	Summary and outlook.....	56
6	Development of a general model to characterize ligands with respect to opening transient pockets in drug targets.....	58
6.1	Motivation .....	58
6.2	Methods .....	60
6.2.1	Datasets .....	60
6.2.2	Development of models to separate inducers and non-inducers .....	61
6.2.3	Analysis of similarity of ligands .....	64
6.3	Results .....	64
6.3.1	Model performance .....	64
6.3.2	Interpretation of models .....	66
6.3.3	Analysis of functional groups .....	68
6.3.4	Fragment analysis on PDBbind set .....	68
6.3.5	Analysis of Mdmx inhibitors.....	70
6.4	Summary and outlook.....	72
7	Discussion .....	74
7.1	Difference in binding modes of ligands with Mdmx and Mdm2 .....	74
7.2	Explanation of inactivity of synthesised ligands .....	75

8	Conclusion .....	77
	Acknowledgements.....	79
	References.....	80
	Appendix.....	93

## Abbreviations

R&D	research and development
Mdmx	mouse double minute x (also known as Mdm4)
Mdm2	mouse double minute 2
p53	an important tumour-suppressor protein <sup>1</sup>
WK298	The first small-molecule inhibitor crystallised with Mdmx <sup>2</sup>
IC <sub>50</sub>	the half-maximal inhibitory concentration, which is a measure of the potency of a ligand in inhibiting 50% of a biological target
RMSD	root-mean-square deviation
PPI	protein-protein interaction
CADD	computer-aided drug design
MD	molecular dynamics
OCCHEM	Online Chemical Modeling Environment
QSAR	quantitative structure-activity relationship
MM/PB(GB)SA	Molecular Mechanics/ Poisson Boltzmann (or Generalised Born) Surface Area
YH300	a known and special Mdm2 inhibitor, which was proved to capture a transient state of Mdm2 via binding with the extension of Leu26 sub-pocket <sup>3</sup>
$\chi_1$ angle of Tyr99 (Tyr100 in Mdm2)	the chi1 angle of tyrosine, which is an indicator to designate the ‘open’ and ‘closed’ state of Mdmx and Mdm2 <sup>3-7</sup>
NT series compounds	putative Mdmx inhibitors optimised from WK298
LG series fragments	eight active fragments binding to Mdmx identified by NMR screening
NC series compounds	putative Mdmx inhibitors optimised from LG series fragments
GAFF2	general Amber force field, version 2
TIP3P	Transferable intermolecular potential 3-site water model <sup>8</sup>
FF14SB	one of the SB family force fields <sup>9</sup>
<i>LEaP</i>	an acronym constructed from the names of the older AMBER software modules it replaces: link, edit, and parm. It is the primary program to prepare input for Amber
AM1-BCC	Austin Model 1 <sup>10</sup> – Bond charge correction <sup>11,12</sup>

# 1 Introduction

## 1.1 Background

Protein-protein interactions (PPIs) are critical therapeutic targets for drug discovery. The challenge is to develop small molecules that target the large and flat interfaces with high affinity. To date, the application of artificial intelligence and incremental computational power has lent new vitality into drug discovery. These advances enable efficient and innovative molecule design, as well as the study of protein dynamics. Thus, computational methods can be used as an approach to rationally design PPI modulators.

The p53-Mdmx interacting surface is a widely studied PPI, where p53 is a tumour suppressor protein and Mdmx is its negative regulator<sup>13</sup>. The *TP53* gene mutates in about half of human cancers<sup>14</sup>; in the remainder, wild-type p53 is inactivated by its negative regulators, Mdm2 and Mdmx proteins<sup>15,16</sup>. The former is experimentally identified as the first major endogenous inhibitor of p53<sup>17</sup>, and the latter, which is highly homologous with Mdm2, influences p53 transcriptional activity<sup>13</sup> and is a vital independent regulator of p53<sup>18</sup>. Until now, however, studies on small-molecule Mdmx inhibitors have made slow progress. There are seven Mdm2 inhibitors in ongoing clinical trials<sup>19</sup>, but none are available for Mdmx. Development of SJ-172550, the first reported inhibitor of Mdmx<sup>20</sup>, was put on hold because of stability problems<sup>21</sup>, and other alternatives, like XI-011<sup>22</sup>, and NSC207895<sup>23</sup> met with difficulties in preclinical trials. Even though Mdm2 and Mdmx are structurally closely related, their slight differences lead to Mdm2 inhibitors or even the p53 transactivation domain being inferior in terms of binding affinity to Mdmx<sup>24</sup>. In some cases, the use of Mdm2 inhibitors is significantly limited because of their toxicity to normal cells<sup>25</sup>. Recently, it was reported that the loss of Mdmx induces p53 activation but has much less destructive effects *in vivo* than Mdm2 inhibition<sup>26</sup>. Therefore, Mdmx remains an interesting and unexplored target of great potential for cancer treatment.

## 1.2 Aims

Protein-protein interfaces are generally considered as ‘undruggable’ target, since the large and flat hydrophobic groove is difficult to adapt to small-molecule ligands with sufficient affinities. Although peptidomimetic ligands may provide similar interactions to the endogenous peptide, they used as drugs have disadvantages such as poor stability, easy aggregation, short half-life, high plasma clearance, and inconvenience in drug delivery. Hence, the goal of this doctoral research project was to develop small-molecule Mdmx inhibitors using advanced computational methods.

**Chapter 2** introduced the common methods used in this work.

The aim of the first phase of this project (**Chapter 3**) was to address the difficulties in the R&D of Mdmx inhibitors, starting from an in-depth analysis of the receptor structure. The initial emphasis was placed on the variant residues between the Mdm2 and Mdmx pockets. The differences in the primary structures resulted in different interactions between the residues and the ligand, further affecting the conformational changes in the protein pocket. By using molecular dynamics simulations, this study should provide a comprehensive understanding of the structures and dynamic changes of Mdmx/ Mdm2, and improve the strategies for designing Mdmx inhibitors.

The aim of the second phase (**Chapter 4**) was to obtain new compounds based on the known inhibitor WK298 and experimentally identified active fragments. One strategy was to optimise the structure of WK298 to mimic the three key residues (Phe19, Trp23, and Leu26) as well as Leu22 of p53, and the other was to connect active fragments with linkers to fully occupy the main functional pocket of Mdmx. The most promising compounds were tested for their inhibitory activities against Mdmx.

The aim of the third part of this dissertation (**Chapter 5**) was to adopt modern machine-learning methods to efficiently generate new Mdmx inhibitors. Wide-ranging successes in the field of natural language processing (NLP) prompted us to establish a language model that enabled *de novo* and automated creation of new molecules with predefined properties. A QSAR model to predict the IC<sub>50</sub> of Mdmx inhibitors was used to direct the output to converge on a certain drug target. Further models to predict physicochemical properties were used to tune the model to produce molecules with good pharmacological properties. The model would inherently explore and exploit the specific chemical space, proposing putative Mdmx inhibitors with new and rational structures.

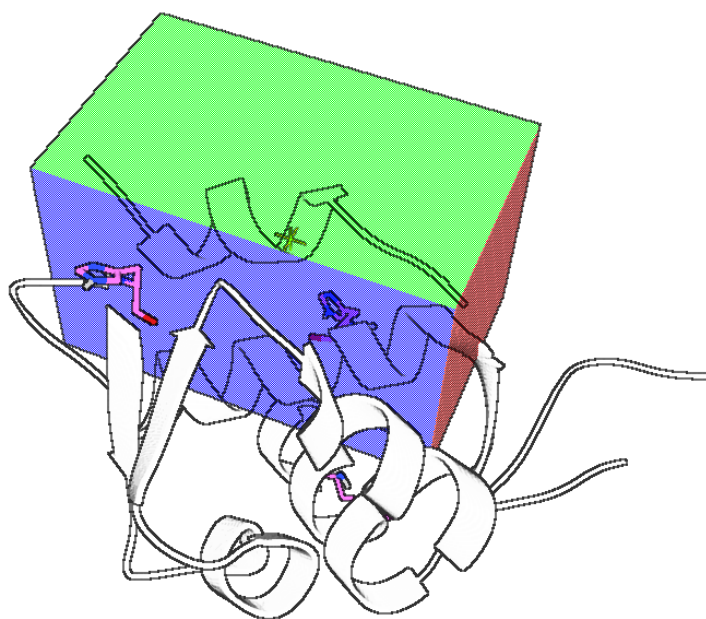
Intrigued with the new insights gained during this study, the aim of the final part (**Chapter 6**) was to assess whether it was possible to identify features in molecules that could induce the opening of cryptic pockets, which could be important in designing Mdmx inhibitor binding to transient conformations of the protein-protein complex. To establish a holo-apo pair dataset, an automated program was developed that accurately matched apo proteins for each holo (bound with a small-molecule ligand) protein. We sought to determine a metric to quantify the changes between apo and holo protein conformations to differentiate the chemical features between inducers and non-inducers. Based on the holo-apo pair dataset, classification models were built to determine the optimum threshold of the metric. The explanation of the model suggests some characteristics of inducers/non-inducers. This analysis and Inducer model could be useful for designing new inhibitors targeting the ‘open’ or ‘closed’ state of Mdmx.



## 2 Methods

### 2.1 Molecular docking

Molecular docking is one of the most important means of molecular modelling that is widely and frequently used in drug design<sup>27</sup>. The docking process simulates ligand and receptor mutual recognition via geometric and energetic matching. When an enzyme interacts with an activator/inhibitor or a pharmaceutical molecule exerts certain pharmacological functions, the ligand binds to the drug target in a manner that ought to represent a plausible approximation of the same interaction in the real world. The molecules tend to adopt appropriate orientations to integrate at a specific part with the necessary interactions. A stable ligand-receptor complex is formed when the ligand and receptor bind to their preferred orientations and conformations near the energy minimum. The rationale is to place the ligand near the active site of the receptor; then, based on geometric and energetic complements, the program searches for the optimum binding pose via real-time evaluation of the interaction between the ligand and receptor (Figure 2-1).



**Figure 2-1** The grid box of the active site in the receptor (e.g. Mdmx) was created with the ligand (e.g. p53) as the centre (PDB ID: 3dab). Putative ligands were put into the box one by one to search for the optimum binding pose.

Molecular docking originated from the ‘lock-key’ model introduced by Emil Fisher<sup>28</sup>. However, the identification between a ligand and a receptor is complicated, in addition to spatial shape matching. The conformation of molecules dynamically changes constantly, and it is also necessary to meet the energy matching required by the induced fit mechanism. The docking score is

usually generated by a function with empirical parameters that fit the experimental data.<sup>27</sup> Nevertheless, it is important to note that the scoring may not have physical significance, or may at least be different from the practical situation due in part to poor representation of complicated solvent and entropic effects.

Docking large numbers of molecules into a protein target is commonly used in virtual screening of receptor-based drug discovery, which allows the prioritisation of compounds for experimental screening and speeds up drug discovery. Comparing the binding poses of hit compounds with that of the known active compounds helps the user to modify them in order to increase the affinity of the hits. Such *in silico* analysis saves a great deal of time and effort for repetitive experiments and decreases the amounts of physical waste.

AutoDock Vina<sup>29</sup> was used in this project to dock a small-molecule ligand into the binding pocket of the drug target. This program is fast and enables the user to batch dock a large number of molecules using a command-line interface. It also ensures that the prediction accuracy is sufficient to provide an initial structure for further analysis and calculations, such as molecular dynamics simulations, particularly when the experimental holo structures are unavailable.

## 2.2 Conventional molecular dynamics (MD) simulation

MD simulation is a method for numerically solving Newton's equations of motion for a system of atoms and molecules in order to obtain insights into the conformational dynamics of the system at nano- to millisecond time scales at atomic resolution. The object given an initial position and velocity will move following the laws of mechanics in a fixed period of time; ideally, the system reaches a dynamic equilibrium governed by molecular mechanics force fields. By analysing the trajectories of atoms and molecules, one will know more about the interactions in a multibody system and the conformational change of each component. While the computation of an MD simulation is more complex than docking, this is offset by the chance that the results could be more accurate as conformational dynamics is considered. Virtual screening, which implements docking followed by MD simulations, brings more reliable hit compounds and assumptions of the binding mode between ligand and receptor so as to be more likely to be in good agreement with subsequent experimental validation.

### *General procedures for system setup*

#### (1) Creating topology and coordinate files for ligands

The *reduce* module of Amber suites<sup>30</sup> affords the addition of hydrogen to the ligands. The *antechamber* module was used with the general Amber force field (GAFF, version 2)<sup>31</sup>. This force field is specifically designed to cover most organic chemical spaces consisting of C, N, O, S, P, H, and halogens. The AM1-BCC charge model<sup>11</sup> can be used to calculate the atomic point charges. The *parmchk2* module checks if there are missing force-field parameters. Provided sufficient parameters, the LEaP program of Amber suites generates an Amber topology file and coordinate file for the ligand.

#### (2) Creating topology and coordinate files for Mdmx-ligand complexes

The LEaP program can mix the AMBER FF14SB force field<sup>9</sup> which is a protein force field with the GAFF force field to treat a complex. The *pdb4amber* program can add hydrogen and remove water molecules to preprocess the PDB files for use in LEaP.

### (3) Adding explicit solvent and counterions

The solute is an Mdmx-ligand complex. The solvent was a pre-equilibrated box of TIP3P<sup>8,32</sup> water, which was created with a buffer of 10 Å to immerse all atoms throughout the simulation. Counterions (e.g. sodium or chloride ions) are placed at the points of lowest or highest electrostatic potential to neutralise the entire system.

### Post-analyses on the trajectory

The side chain torsion angle ( $\chi_1$ ) of tyrosine is an indicator to designate the ‘open’ and ‘closed’ state of Tyr99 in Mdmx (Tyr100 in Mdm2).<sup>3,33</sup> The open state corresponds to  $\chi_1$  around 180°, and the closed state has  $\chi_1$  around 300°. The atoms involve N, C<sub>alpha</sub>, C<sub>beta</sub> and C<sub>gamma</sub>, as indicated in Figure 2-2.

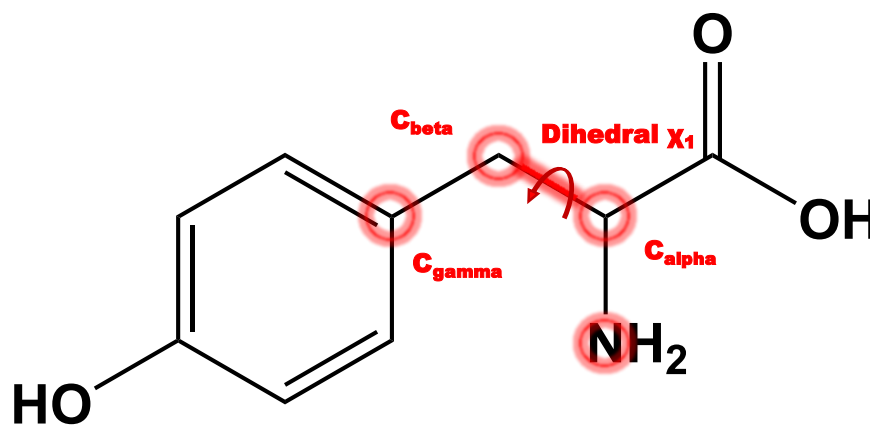


Figure 2-2 The diagram of the side chain torsion angle ( $\chi_1$ ) of Tyr99 in Mdmx (Tyr100 in Mdm2)

## 2.3 Binding free energies

MD simulations enable access to binding free energies with respect to every component of the system. Molecular Mechanics/Poisson Boltzmann (or Generalised Born) Surface Area calculations can be easily carried out using a Python script (MMPBSA.py)<sup>34</sup>. This is a post-processing end-state method in which snapshots selected from a trajectory of interest are used to calculate the free energy change between the bound and unbound states of the system ( $\Delta G$ ). Entropy contributions to the total binding free energy can be complemented as a further refinement, supported by sufficient computational power. The calculation follows the following equations:

$$\begin{aligned}\Delta G &= G_{\text{complex}} - G_{\text{receptor}} - G_{\text{ligand}} \\ &= \Delta H - T\Delta S = \Delta E_{MM} + \Delta G_{sol} - T\Delta S \\ \Delta E_{MM} &= \Delta E_{vdw} + \Delta E_{ele} + \Delta E_{inter} \\ \Delta G_{sol} &= \Delta G_{PB/GB} + \Delta G_{SA}\end{aligned}$$

where  $E_{MM}$  is the gas phase interaction energy constituted by van der Waals energy ( $E_{vdw}$ ), electrostatic energy ( $E_{ele}$ ), and internal energy ( $E_{inter}$ ) which is neglected with an assumption in the single trajectory protocol that the intramolecular energies of the ligand and receptor are unchanged upon binding<sup>35</sup>,  $G_{sol}$  is the sum of the electrostatic solvation free energy calculated by PB or GB equations ( $G_{PB/GB}$ ) and the nonpolar solvation free energy estimated by the solvent-accessible surface area ( $G_{SA}$ ); and  $S$  is the entropy contribution.

## 2.4 OCHEM

The online chemical modelling environment (OCHEM) is a web-based platform that aims to automate and simplify the typical steps required for QSAR modelling<sup>36</sup>. Along with the development of machine learning methods and the enhancement of computing power in the past decade, the OCHEM has introduced more state-of-the-art modelling methods<sup>37</sup> and is able to make models faster. The platform also includes an ever growing list of newer types of descriptors. The OCHEM is available online at <http://www.ochem.eu> and is convenient for users to carry out the work of either classification or regression models. Moreover, it is easy to apply well-trained models to predict unknown molecules with the help of the OCHEM.

## 2.5 Schrodinger

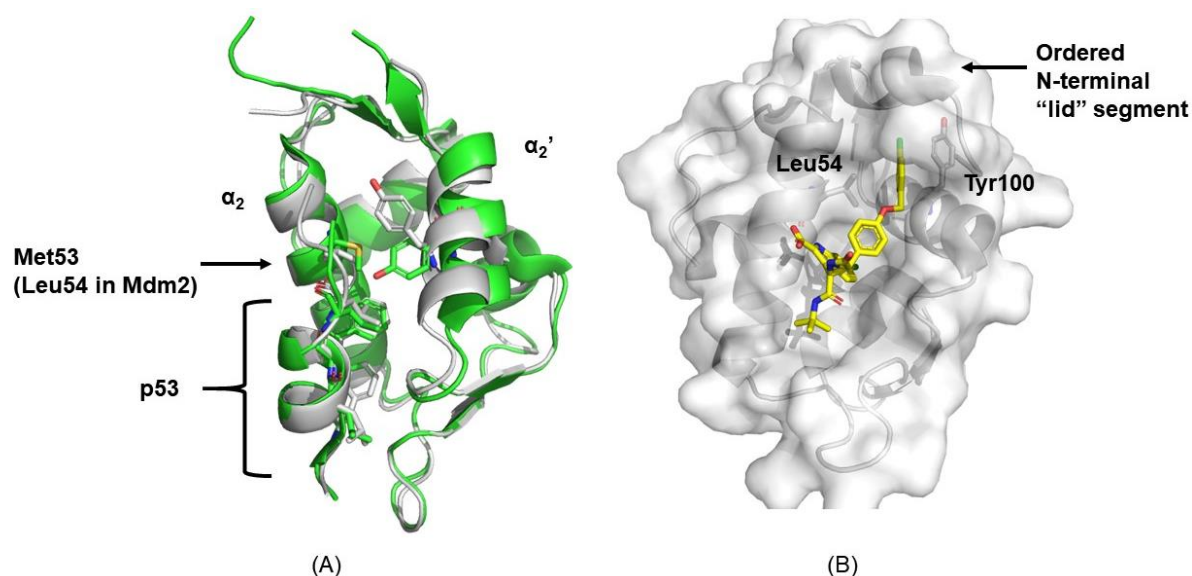
The Schrodinger suite of software provides a well-integrated environment for small-molecule drug discovery, including a variety of applications such as visualising (PyMOL<sup>38</sup>), editing (Maestro<sup>39</sup>), and analysis of molecules. The Schrodinger Python API allows users to flexibly deploy various functions available in the Schrodinger as requested and establish a pipeline to batch processing molecules via a custom Python script.

## 3 Mdmx pocket dynamics

### 3.1 Introduction

Popowicz et al.<sup>7,33</sup> dissected the differences in the details between Mdm2 and Mdmx, explaining why Mdm2 ligands do not have equivalent inhibition against Mdmx and emphasising the necessity of finding compounds specific to Mdmx. As shown in Figure 3-1, the binding pocket of Mdmx is smaller than that of Mdm2 because Met53 (Leu54 in Mdm2) and Tyr 99 (Tyr100 in Mdm2) protrude into the pocket as well as the larger size of methionine compared to leucine. The authors addressed the open and closed conformations of Tyr100 in Mdm2, measured by  $\chi_1$  (Tyr99 in Mdmx). The open conformation of Tyr100 corresponds to  $\chi_1$  around  $180^\circ$ , and the closed conformation has  $\chi_1$  around  $300^\circ$ . They believed that the closed Tyr99 conformation in Mdmx could be intrinsically caused by the distinct position of helix  $\alpha_2'$  which also significantly changed the shape of the binding pocket. Joseph et al.<sup>40</sup> tried to explain the differential binding of ligands to Mdm2/Mdmx with 15-ns trajectories of four systems: the Mdm2-p53 complex, Mdm2-Nutlin2 complex, Mdmx-p53 complex, and Mdmx-Nutlin2 complex. They obtained several interesting findings, but they were based on the premise that p53 has a higher affinity for Mdm2 than Mdmx. Based upon hydrogen bond analysis, they inferred that p53 complexed with Mdm2 had a more stable and longer helix than when complexed with Mdmx. The electrostatics results illustrated Mdm2 had more positive charges than Mdmx, which might make p53 more complementary to the binding groove of Mdm2 than it was to Mdmx. However, Chen et al.<sup>41</sup> theoretically showed that the closed conformation in Mdmx weakened the van der Waals contacts between the peptide-like ligand and several key residues, including Met53, which might be correlated with the lower affinity of ligands to Mdmx than to Mdm2. Despite whether electrostatic or van der Waals interactions dominated the differential binding to Mdm2/Mdmx, both Joseph and Chen assumed that p53 had a higher affinity for Mdm2 than Mdmx. They provided the biological argument that Mdm2 needed to transport p53 from the nucleus to the cytoplasm, but Mdmx did not require tight binding. Experimentally, this was not in agreement with the data in the review of Popowicz et al. ( $K_i = 0.89 \mu\text{M}$  for Mdm2- p53,  $K_i = 0.21 \mu\text{M}$  for Mdmx- p53).<sup>5</sup>

With the contribution of more crystal structures, Bista et al.<sup>3</sup> demonstrated the existence of a ligand-induced pocket between Tyr100 and Leu54 on the surface of the Mdm2 protein. The 4-chlorobenzyl group of the ligand YH300 lay in the cavity, which was an extension of the Leu26 pocket. Taken together, the Leu26 pocket in Mdm2 wrapped up the ligand more tightly, and YH300 induced wider opening of the Leu26 sub-pocket. This implied a higher binding affinity of inhibitors against Mdm2 than against Mdmx.<sup>42</sup> The crystallographic proofs above can be rationally used to create more potent and selective ligands. Given the high similarity of Mdm2 and Mdmx, the dynamic characteristics of the Mdmx binding pocket were studied in order to develop more potent Mdmx inhibitors.



**Figure 3-1. (A) The alignment of Mdmx and Mdm2 complex with p53, respectively. The  $\chi_1$  of Tyr99 in Mdmx is  $-68.9^\circ$  and that of Tyr100 in Mdm2 is  $162.3^\circ$ . Mdmx-p53 (PDB ID: 3dab<sup>33</sup>): green, cartoon; Mdm2-p53 (PDB ID: 1ycr<sup>43</sup>): white, cartoon. (B) Mdm2 complex with YH300 (PDB ID: 4mdn<sup>3</sup>). Mdm2: white, cartoon; YH300: yellow, sticks. Key residues are shown in sticks. (The structures were plotted and measured by PyMOL.)**

## 3.2 Methods

To check whether Mdmx also has an extension of the Leu26 sub-pocket, YH300 was utilised to investigate the binding mode with Mdmx by MD simulations. Docking was used because there was no crystal structure of the Mdmx-YH300 complex. Then, apo Mdmx, Mdmx-p53, and Mdmx-WK298 were used as reference systems with the same protocols (Figure 3-2). For comparison, I also investigated the crystal structures of typical Mdm2 complexes by visual inspection.

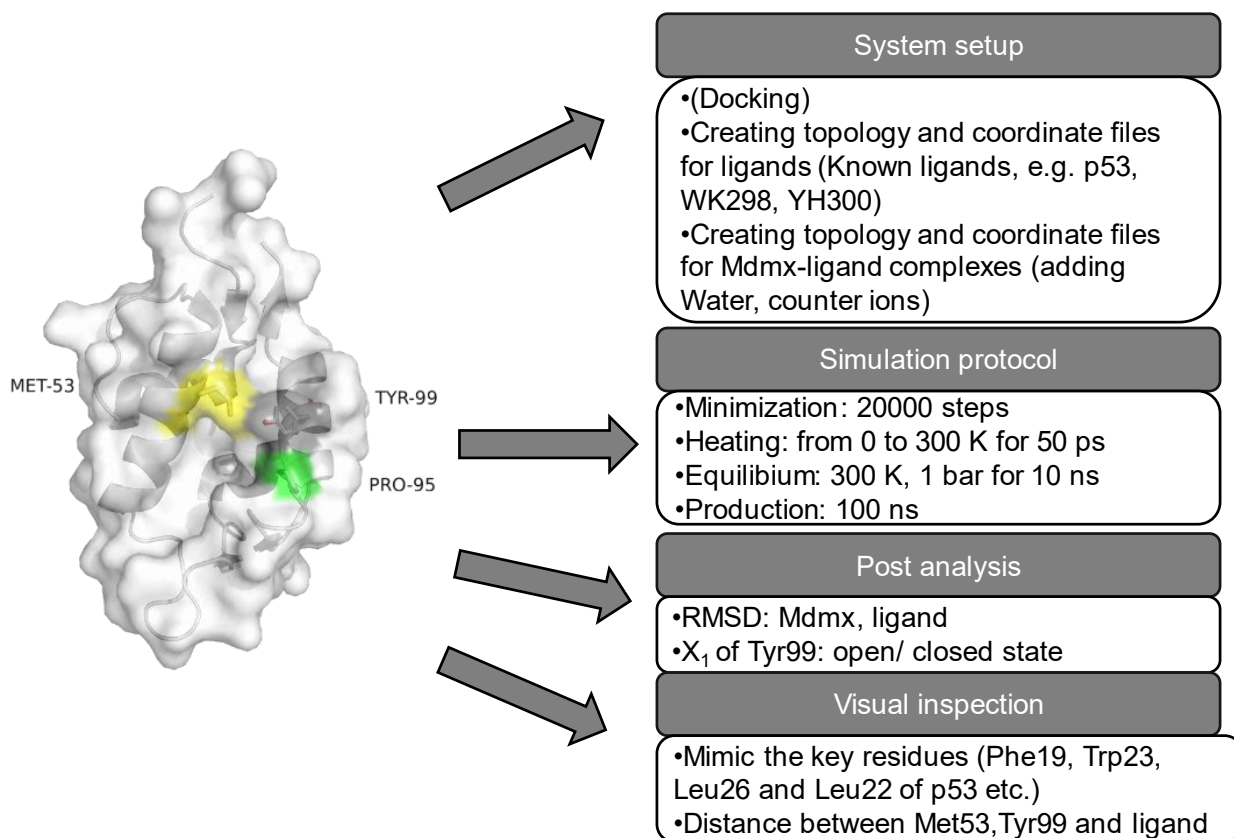


Figure 3-2 The general computational and analysis procedures

### 3.2.1 Molecular docking

It is interesting to learn how YH300 binds to Mdmx. I used the molecular structure of YH300 extracted from the crystal structure deposited with PDB ID 4mdn.<sup>3</sup> YH300 was first minimised to correct the bonds and atom positions because of the incorrect contacts in the formyl-substituted amino group (wrong bonds and angles). Then, it was docked into the binding pocket of Mdmx extracted from the crystal structure deposited with PDB ID 3dab<sup>33</sup> by AutoDock Vina.<sup>29</sup> The receptor was protonated at neutral pH with only polar hydrogens by AutoDockTools-1.5.6,<sup>44</sup> and Gasteiger charges were added. The centre was determined by the position of p53 in the Mdmx-p53 complex ( $x = 0.514$ ,  $y = -21.838$ ,  $z = 8.047$ ). A grid box size of  $24 \times 18 \times 22$  was generated with spacing of  $1.0 \text{ \AA}$  between the grid points. The other parameters of AutoDock Vina were used as the default values.

### 3.2.2 MD simulations

#### *System Setup for Molecular Dynamics (MD) Simulation*

To study the conformational change of the Mdmx binding pocket, all-atom MD simulations were implemented using AMBER 20 software running on CPUs and GPUs.<sup>30</sup> In addition to the Mdmx-YH300 complex selected from the docking results, all initial structures were directly re-

trieved from the crystal structures deposited in the RCSB Protein Data Bank (<http://www.rcsb.org/>).<sup>45</sup> The first system consisted of one copy of the Mdmx protein (PDB ID: 3dab<sup>33</sup>) surrounded by 5939 TIP3P water molecules, and one chloride ion neutralising the entire system. The second system consisted of one copy of Mdmx protein (PDB ID: 3dab), one copy of the transactivation domain of human p53 (PDB ID: 3dab), 5864 TIP3P water molecules, and one sodium ion neutralising the entire system. The third system consisted of one copy of the Mdmx-WK298 complex (PDB ID: 3ljb<sup>46</sup>), 5633 TIP3P water molecules, and one chloride ion neutralising the entire system. The fourth system consisted of one copy of the Mdmx-YH300 complex, surrounded by 5918 TIP3P water molecules, and one chloride ion neutralising the entire system. The partial atomic charges of the ligands were derived using the *antechamber* module implemented in the *AMBER* software package to calculate AM1-BCC charges. The other force field parameters of ligands derived from GAFF2 and the *AMBER* FF14SB force field were employed to model the Mdmx. *Antechamber* prepared residue topologies for ligands and *LEaP* for Mdmx.

### *MD Simulation Protocols*

For each system, the solvent of the MD system was first minimised using the XMIN method across 20000 steps.<sup>47</sup> All the solutes were restrained using a harmonic potential with a force constant of 100 (kcal/mol)/Å<sup>2</sup>.<sup>47</sup> The MD simulation consisted of three phases: the relaxation phase, equilibrium phase, and sampling phase. In the relaxation phase, the simulation system was heated progressively from 0 to 300 K in steps of 50 K and 5ps each with a force constant of 2 (kcal/mol)/Å<sup>2</sup>, followed by maintaining the temperature at 300 K in the last 20 ps. After heating steps, the system was equilibrated at 1 bar for 10 ns but without any restraints or constraints in the last 5 ns. Finally, a 100 ns MD simulation was performed for each system. In total, 10 000 snapshots were recorded during the production phase. Additional settings for constant volume and pressure MD simulations performed in this work are listed as follows: the temperature was regulated by the weak-coupling algorithm; pressure was regulated by the isotropic position scaling algorithm with the pressure relaxation time being set to 1.0 ps; integration of the equations of motion was conducted at a time step of 0.5 fs for the heating steps and 2 fs for others. All bonds involving hydrogen atoms were constrained using the SHAKE algorithm in the MD simulation stages. The particle-mesh Ewald (PME) procedure processed long-range electrostatic interactions.

The RMSD of backbone atoms or ligands was used to monitor the stability of the entire trajectory and to check if the selected parameters for the MD simulations were valid. The  $\chi_1$  angle of Tyr99 (Tyr100 in Mdm2) was used as an indicator of the shape change of the Leu26 pocket.

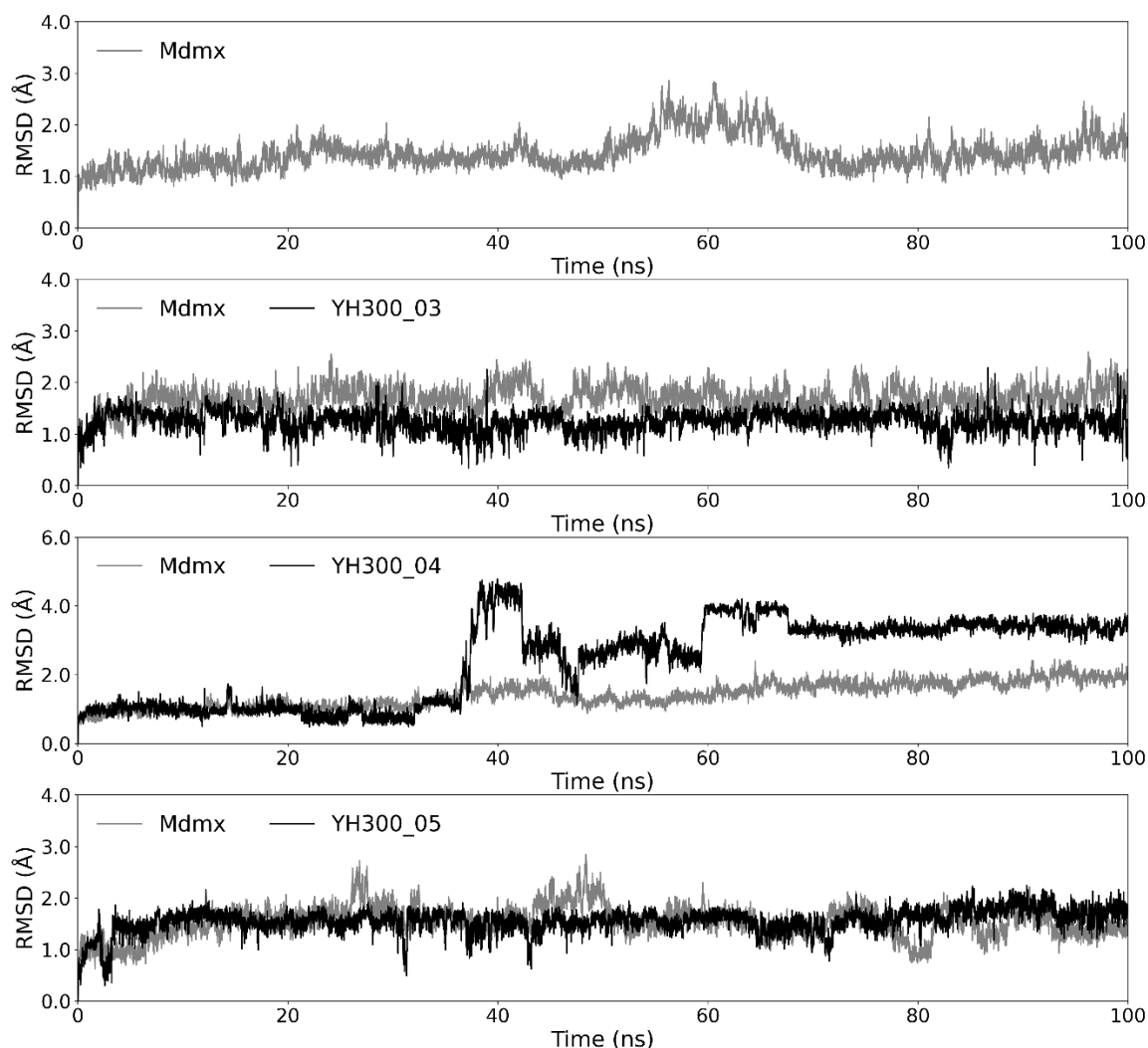
## **3.3 Results**

### **3.3.1 Trajectory analysis**

The docking results of YH300 and Mdmx did not mimic the Phe19, Trp23 and Leu26 of p53 very well and were quite different from the way it binds with Mdm2. Regardless of the docking



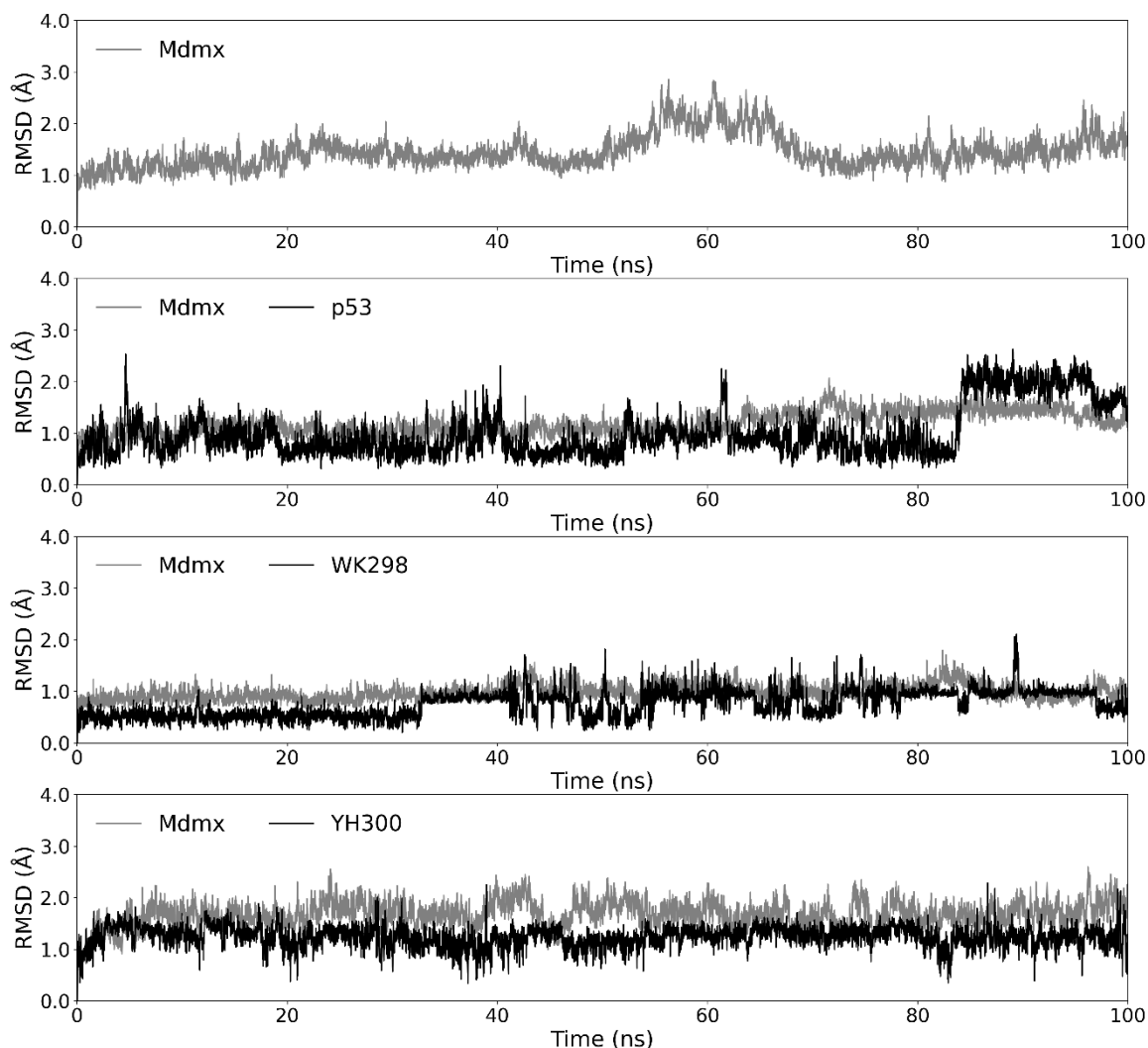
score, three binding poses for which the indole ring was the closest to the Trp23 sub-pocket were selected, referred to as YH300\_03, YH300\_04, and YH300\_05. Comparing the last frames extracted from these three trajectories, YH300 was finally located in the vicinity of the Leu26 sub-pocket (YH300\_03) or only occupied the Phe19 and Trp23 sub-pocket of p53 (YH300\_04 and YH300\_05). The trajectory of the former was more stable (Figure 3-3, Mdmx and YH300\_03).



**Figure 3-3** The RMSDs (backbone atoms) of three Mdmx-YH300 systems as a function of time were plotted through the 100-ns trajectories, each of which was calculated with the first frame as the reference. The docking and simulation indicated that YH300, as a promising Mdm2 inhibitor, could also bind with Mdmx to some extent because it did not escape from the pocket.

As shown in Figure 3-4, both Mdmx and ligands had RMSDs fluctuating within a certain range without sharp increments or decrements throughout the trajectories. This indicated that the parameters used for MD simulations were appropriate, and the trajectories were reliable for further analysis. The sharp increase in RMSD of p53 bound to Mdmx was caused by the flexible terminals of this peptide. Mdmx-WK298 had a stable trajectory compared to that of the Mdmx-p53 complex. Intrigued, Mdmx binding with YH300 had a larger RMSD of the main chain and was

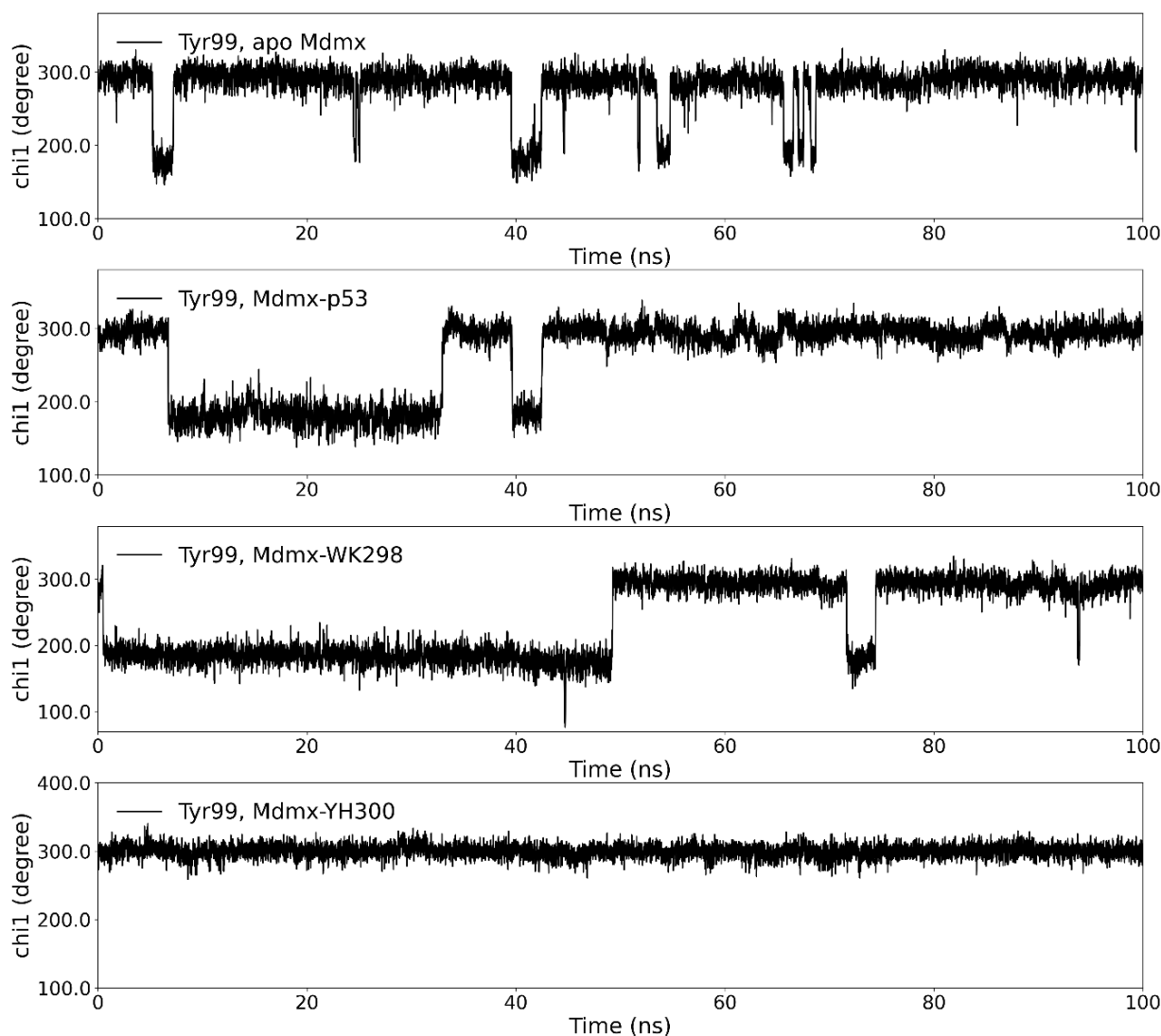
even less stable than the apo Mdmx, suggesting that YH300 does not have a stronger affinity to Mdmx than WK298 and p53.



**Figure 3-4** The RMSDs (backbone atoms) of apo Mdmx, Mdmx-p53, Mdmx-WK298 and Mdmx-YH300 as a function of time were plotted through the 100-ns trajectories, each of which was calculated with the first frame as the reference. The binding affinity was positively correlated to the stability of the trajectory. The Mdmx-YH300 complex presented relatively unstable trajectories compared to the other two complexes, indicating the binding affinity would be weaker.

The dynamic change of Tyr99  $\chi_1$  was designated as an indicator to study the shape change of the Leu26 pocket. In the apo Mdmx, Tyr99 almost always stayed in the closed conformation (average  $\chi_1 \approx 282^\circ$ , Figure 3-5), which agreed with the inference that Tyr99 of Mdmx was intrinsically closed in the free state<sup>4</sup>. However, YH300 kept Tyr99 closed throughout the entire trajectory, which was exactly opposite to the situation when it was binding with Mdm2<sup>3</sup>. In the other two bound states, Tyr99 of Mdmx presented more conformations in the open state, even though neither of them maintained this conformation until the end. It is worth mentioning that WK298 induced the opening of Tyr99 for nearly half of the simulation time, among which Tyr99  $\chi_1$  was as

low as  $77^\circ$  (Figure 3-5). Interestingly, Pro95-Ser96 of Mdmx lost their helical structure by binding with WK298 rather than p53. Compared to the equivalent His96-Arg97 in Mdm2, they were uncharged and shorter. These results suggest that WK298 could be a starting point for developing a more active Mdmx inhibitor that made the open state last longer and formed some contact with Pro95 to protect its helical conformation.



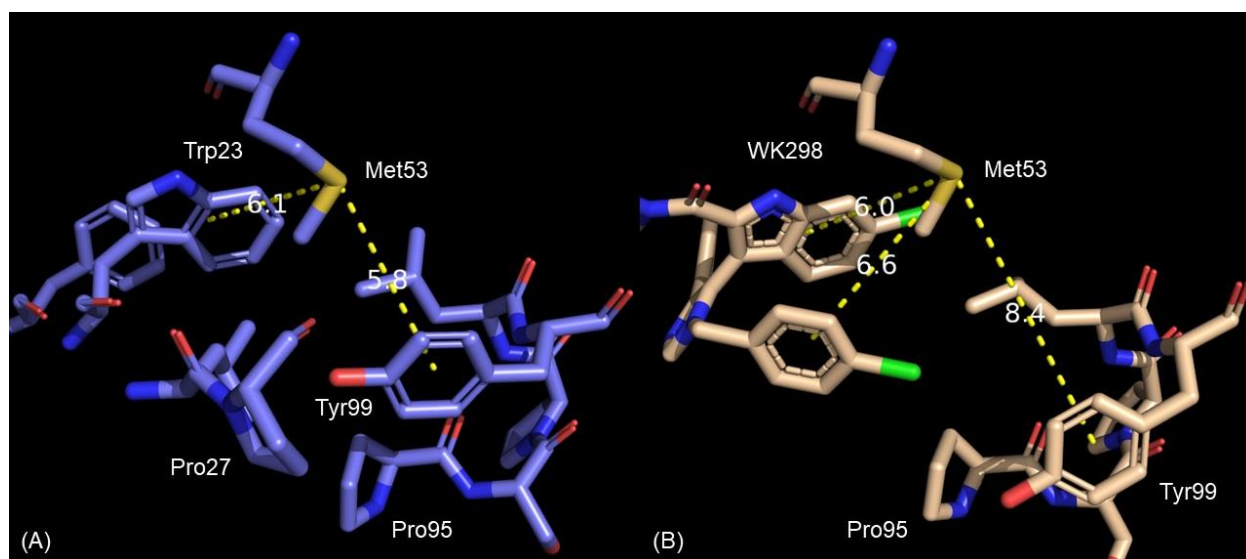
**Figure 3-5** The dynamic change of Tyr99  $\chi_1$  in apo Mdmx, Mdmx-p53, Mdmx-WK298 and Mdmx-YH300 systems through the 100-ns trajectories. The open conformation of Tyr100 corresponds to  $\chi_1$  around  $180^\circ$  and the closed conformation has  $\chi_1$  around  $300^\circ$ . The apo Mdmx inherently had a flexible Tyr99 but mostly stayed in the closed state; the binding of YH300 permanently made Tyr99 stuck in the closed state because YH300 occupied the same position as the open-state Tyr99; the other two complexes left the freedom for the rotation of Tyr99 and had more open states than the apo one.

### 3.3.2 Key residues analysis

The binding site in the Mdmx-WK298 complex involved the helix  $\alpha_2$  (Met53, Gly57, Ile60), the sheet  $\beta$  (Tyr66, Gln71), the loop (Val92, Lys93, Asp94, Pro95), and the helix  $\alpha_2'$  (Leu98, Tyr99). Among the differences between the Mdmx and Mdm2 pockets, Met53 and Pro95 (Leu54 and His96 in Mdm2) attracted my attention, as candidate residues that might lead to the different binding modes of p53 towards Mdmx and Mdm2.

First, proline is a special amino acid that connects with the main chain twice by forming a ring that complicates the adoption of a normal helical conformation. The hydrophobicity of proline may cause its propensity to interact with aromatic structures. However, histidine at the equivalent position in Mdm2 is the most common amino acid in the protein binding site, which is polar, aromatic and can be charged or neutral.<sup>48</sup> In the Mdm2 complex, the characteristics of His96 make it easily interact with various structures by forming hydrogen bonds or  $\pi$ - $\pi$  interactions. By analysing the hydrogen bonds in the trajectories of Mdmx, Pro95 only formed intramolecular hydrogen bonds with Tyr99 and Leu98 for approximately 30 ns and 5 ns (distance cutoff = 3.5 Å).

Second, the sulphur atom of methionine can mediate Met-aromatic interactions.<sup>49</sup> The aromatic residues could be tyrosine, tryptophan, and phenylalanine. In the Mdmx-p53 complex (Figure 3-6, A), Met53 interacted with Trp23 of p53 and Tyr99 of Mdmx, together with the hydrogen bonds formed between Tyr99 and Pro27, restricting Tyr99 in the closed conformation. In the Mdmx-WK298 complex (Figure 3-6, B), the 4-chlorobenzyl ring participated in the Met-aromatic interaction; however, Tyr99 was further to the Met53 and the Pro27 sub-pocket was unfilled, which might decrease the binding affinity to Mdmx. This implied that WK298 induced the opening of Tyr99 for a longer time, possibly due to a decrease in the Met-aromatic interaction between Met53 and Tyr99.



**Figure 3-6 (A) Mdmx-p53 complex (PDB ID: 3dab); (B) Mdmx-WK298 complex (PDB ID: 3lbj). The distance between the sulphur atom and aromatic ring (Å): yellow dot line with white label.**

Compared to endogenous p53, binding with WK298 made the Tyr99 further away from the Met53 of Mdmx, indicating the interaction between them was weakened.

Meanwhile, the Met-aromatic interaction also played an important role in the Mdm2 pocket (Figure 3-7). Met50 (Val49 in Mdmx) interacted with Tyr100 (Tyr99 in Mdmx) forming an open orientation. In the Mdm2-YH300 complex (PDB ID: 4mdn), the 4-chlorobenzyl of YH300 filled the Pro27 sub-pocket and pushed Tyr100 away, weakening the Met50-Tyr100 interaction. In Table 3-1, I used the distance between the sulphur atom of Met50 and the ring of Tyr100 to represent the strength of the Met50-Tyr100 interaction. The lower the  $\chi_1$  and distance, the stronger the ligand inhibited Mdm2. Moreover, the longer and more reactive His96 formed interactions with ligands independent of Tyr100. The His96-ligand interaction contributed much to the inhibition of Mdm2 (shown as distance in Figure 3-7).

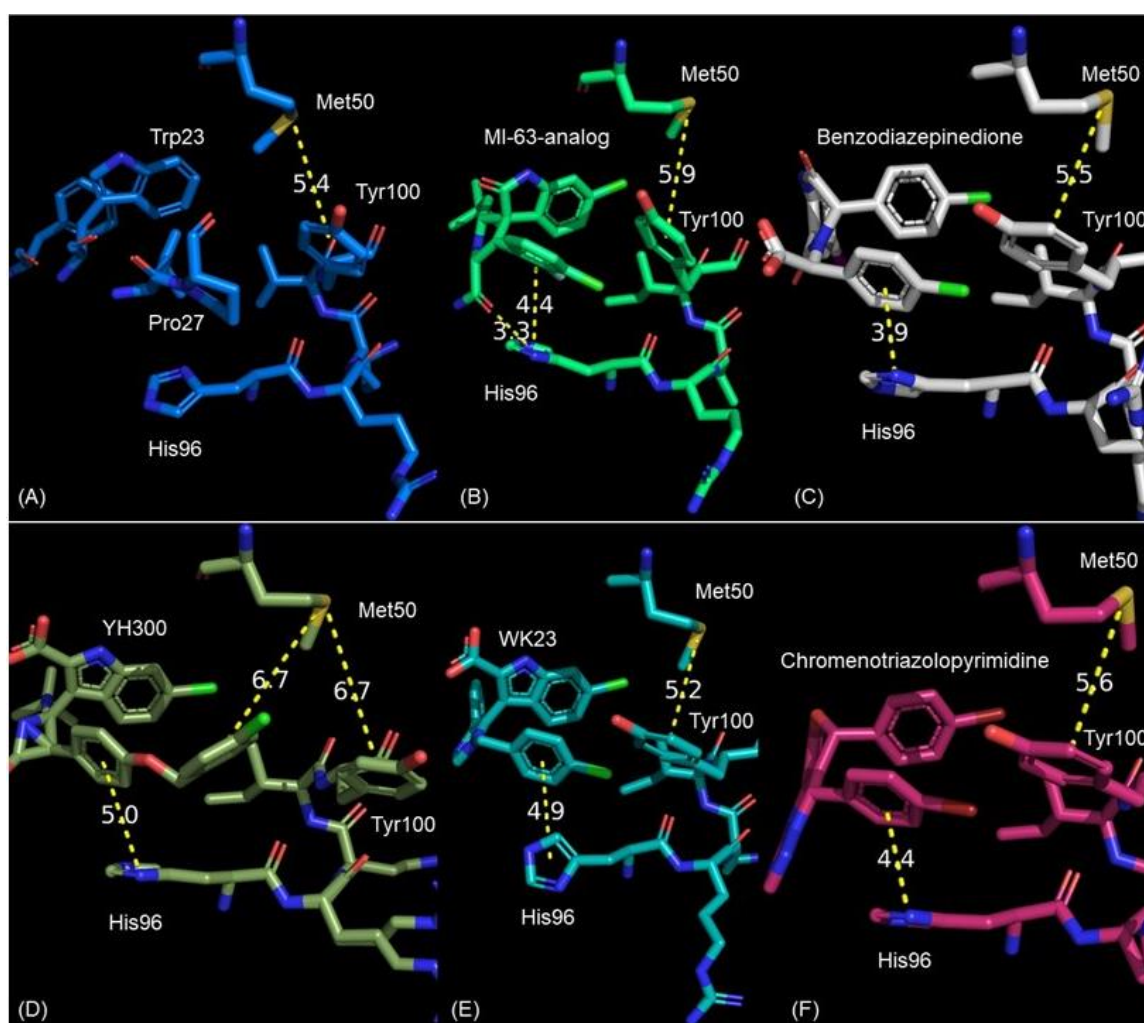


Figure 3-7 (A) Mdm2-p53 complex (PDB ID: 1ycr); (B) Mdm2-MI-63-analog complex (PDB ID: 3lbi); (C) Mdm2-Benzodiazepinedione complex (PDB ID: 1t4e); (D) Mdm2-YH300 complex (PDB ID: 4mdn); (E) Mdm2-WK23 complex (PDB ID: 3lbk); (F) Mdm2-Chromenotriazolopyrimidine complex (PDB ID: 3jzk). The distance between the sulphur atom and aromatic ring (Å): yellow dot line with white label. One can see from these complexes that the Met-aromatic interaction also played a great role in the binding between Mdm2 and ligand.

Ligand	K <sub>i</sub> (uM)	PDB ID	Tyr100 $\chi_1$ (°)	Distance (Met50 to Tyr100, Å)
p53	0.89	1ycr <sup>43</sup>	-162.3	5.4
MI-63-analog	0.036	3lbl <sup>50</sup>	-141.2	5.9
Benzodiazepinedione	0.08	1t4e <sup>51</sup>	-102.2	5.5
YH300	0.6	4mdn <sup>3</sup>	-174.5	6.7
WK23	0.916	3lbk <sup>50</sup>	-99.0	5.2
Chromenotriazolopyrimidine	1.23 (IC <sub>50</sub> )	3jzk <sup>52</sup>	-97.2	5.6

**Table 3-1 The inhibition constant (K<sub>i</sub>, μM) of Mdm2 ligands with crystal structures deposited in PDB.**

### 3.4 Summary and outlook

A comprehensive analysis was carried out on the structural differences between Mdmx and Mdm2 based on crystal structures and all-atom simulations. The initial emphasis was placed on differences in the protein primary structure. In particular, methionine was at different positions in the main binding pocket, namely Met50 in Mdm2 and Met53 in Mdmx. This methionine caused the different orientations of the tyrosine, referred to as the ‘open’ and ‘closed’ state of Tyr99 in Mdmx (Tyr100 in Mdm2). By inspecting the conformations from crystal structures and MD simulations, the Met-aromatic interaction (Met-Tyr-Ligand interaction in the case of Mdmx/Mdm2) restricted the rotation of tyrosine. As a result, the ‘open’ state of tyrosine made the main pocket larger than the ‘closed’ state of tyrosine. The enlarged part was regarded as a transient pocket which was reachable only when the tyrosine was in the ‘open’ state. In addition to methionine, other variant residues at equivalent positions, such as Pro95 in Mdmx (His96 in Mdm2), also contributed to the low affinity of small-molecule inhibitors against Mdmx. The findings in this section would differentiate the strategy for developing Mdmx inhibitors from those for Mdm2 inhibitors.

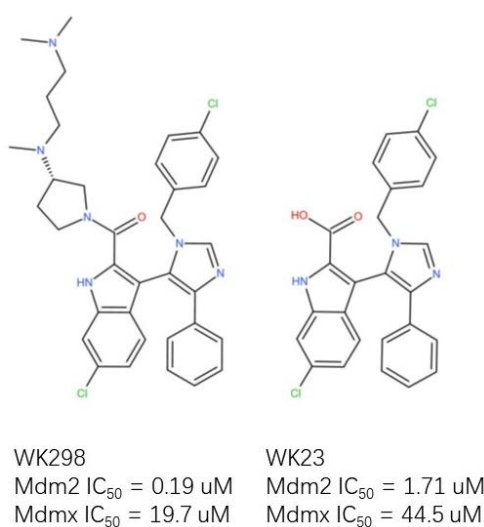
For Mdm2 ligands, forming stronger interactions with His96 and protecting the Met50-Tyr100 interaction (the ‘open’ conformation) would remarkably improve the binding affinity. The ‘open’ or ‘closed’ state of Tyr100 could be identified by the change of Tyr100  $\chi_1$ ; the distance between Met50 and Tyr100 could be used to monitor the strength of the Met-aromatic interaction. To develop potent Mdmx inhibitors, a promising direction would be to mimic the Trp23 (p53)-Met53-Tyr99 interactions and keep the helical conformation of Pro95 along with filling the Pro27 sub-pocket. Moreover, the substituent that occupied the Leu26 pocket should be small to avoid interference with the Met53-Tyr99 interaction.

The analysis in this chapter and the references<sup>3,4</sup> demonstrated that Mdm2 and Mdmx had different intrinsic conformations. This part of the work revealed that Met50 interacted with Tyr100 so that Mdm2 had its intrinsic state ‘open’, while Mdmx had its ‘closed’ because of the interaction between Met53 and Tyr99. However, I could not completely exclude Mdmx, which was also able to bind ligands in the ‘open’ state; thus, both states were considered later in the thesis.

## 4 Developing new Mdmx inhibitors using structure-based drug discovery

### 4.1 Introduction

The first small molecule crystallised with Mdmx is WK298, which sheds light on the binding mode between a small-molecule inhibitor and Mdmx.<sup>53</sup> The three key residues of p53 (Phe19, Trp23, and Leu26) were superimposed with the phenyl, 6-chloroindole, and 4-chlorobenzyl groups of WK298, respectively. WK23, which is an analogue of WK298, adopted almost the same conformation bound to Mdm2 as WK298 bound to Mdmx. While the extended side chain of the 2-substituted indole ring offered hydrophobic interactions and a transient hydrogen bond with Gln71 (650 ps of the trajectory) of benefit to the Mdmx inhibition, both affinities to Mdmx were much lower than to Mdm2 (Figure 4-1).



**Figure 4-1** The structures of Mdm2/Mdmx inhibitors.

The Leu22 of p53 was experimentally and computationally proven to be a key residue providing additional van der Waals contacts with Mdm2.<sup>43,54–57</sup> To mimic the Leu22 of p53 is a way to improve the binding affinity of ligands to Mdmx.<sup>58</sup> The 4-chlorophenyl group was not appropriate for the Leu26 sub-pocket of Mdmx as demonstrated in the previous chapter. Therefore, a combination of docking and MD simulations was used to reconstruct WK298 in order to have a side chain superimposed with Leu22 of p53 as well as a better fit for the Leu26 sub-pocket (**Strategy 1**, referred to as **NT series** compounds in the rest of this dissertation). In addition, eight small fragments were identified by NMR screening to be bound to Mdmx (Figure 4-2). Therefore, the second strategy is to develop active molecules based on the eight fragments (**Strategy 2**, referred to as **NC series** compounds in the remainder of this dissertation). Similarly,

they were first docked into the Mdmx-binding pocket. Next, the top-scoring binding mode was used to run the MD simulations. The representative conformation was selected from the trajectory. By visually inspecting the binding modes between ligands and receptors, a linker was added to connect fragments so as to create some molecules superimposed with the three key residues of p53 as well as possible. The larger molecules created entered the next round of the MD simulations. The molecules showing good alignment with the key residues of p53 were synthesised and their bioactivity (SN series compounds) was tested.

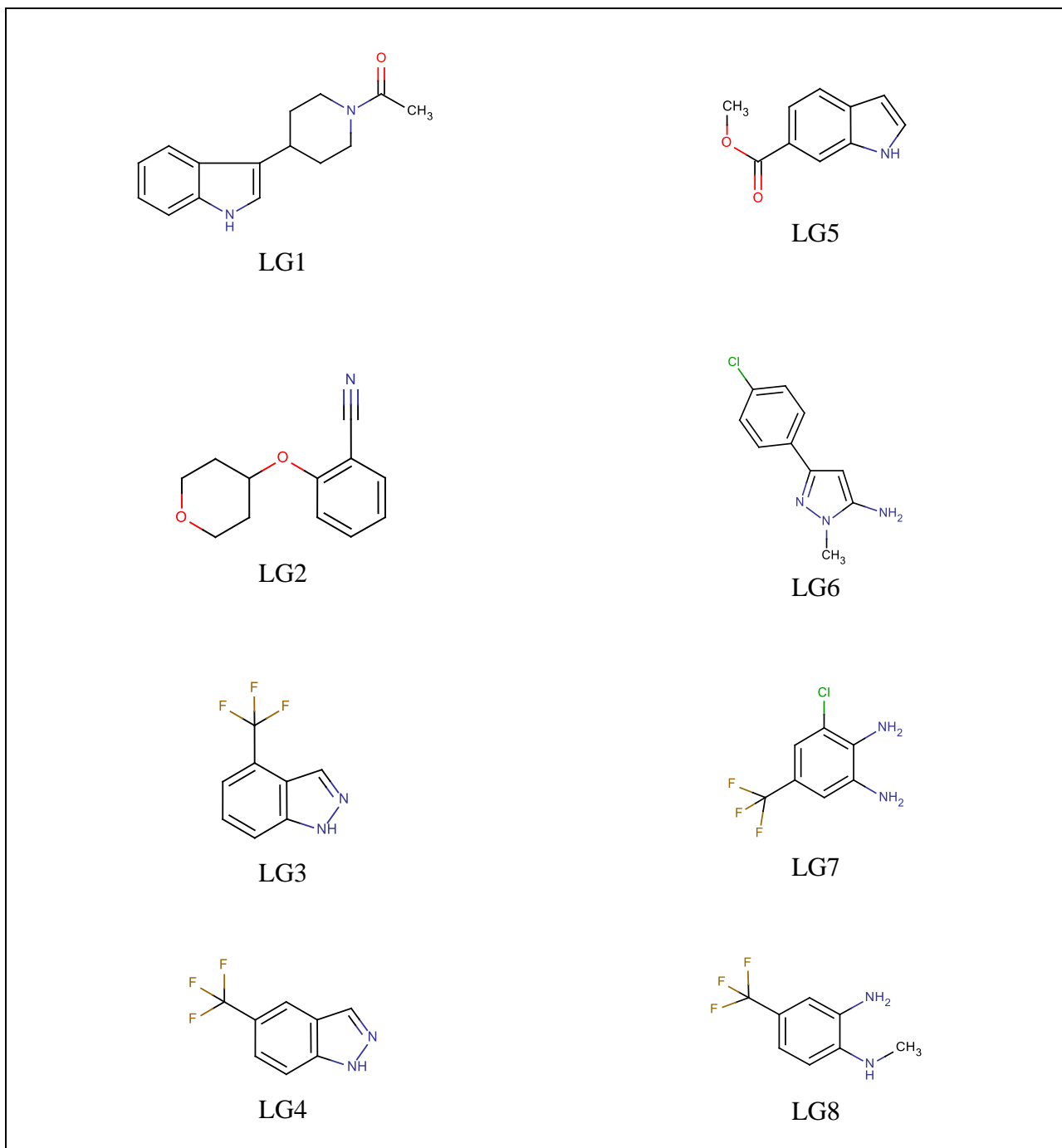


Figure 4-2 Eight active fragments binding to Mdmx identified by NMR screening.



## 4.2 Methods

Docking was first implemented for all the newly designed molecules. Then, apo Mdmx, Mdmx-p53, and Mdmx-WK298 were used as reference systems. The simulation protocol was similar to that described in the previous section (Figure 4-3).

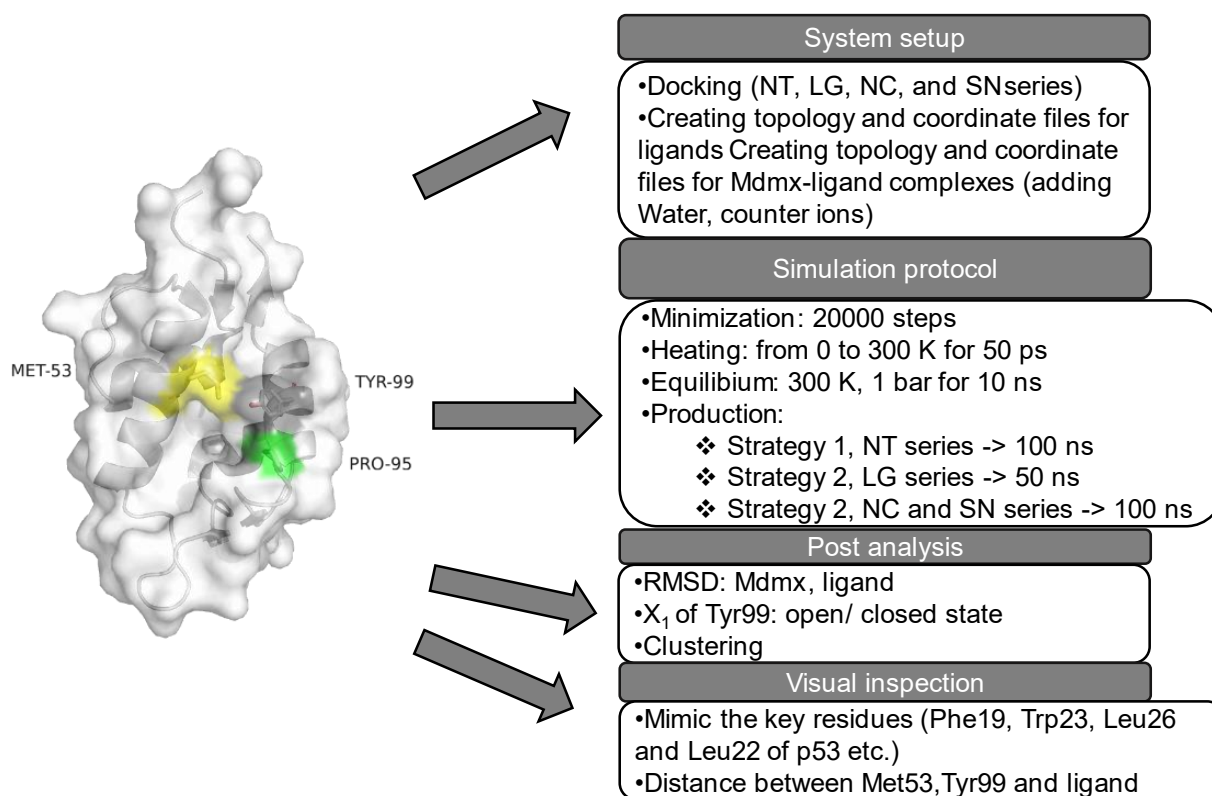


Figure 4-3 The general computational and analysis procedures

### 4.2.1 Molecular docking

Each fragment and new reconstructed molecule were docked into the binding pocket of Mdmx extracted from the crystal structure deposited with PDB ID of 3dab<sup>33</sup> by AutoDock Vina.<sup>29</sup> The receptor was protonated at neutral pH with only polar hydrogens by AutoDockTools-1.5.6,<sup>44</sup> and Gasteiger charges were added. The centre was determined by the position of p53 in the Mdmx-p53 complex ( $x = 0.514$ ,  $y = -21.838$ ,  $z = 8.047$ ). A grid box size of  $24 \times 18 \times 22$  was generated with spacing of  $1.0 \text{ \AA}$  between the grid points. The other parameters were left as the default values. Because of some bad contacts that might exist in the new molecules, the minimisation that could be done by Maestro<sup>39</sup> was necessary for each molecule to correct the bonds and atom positions. All 10 binding poses predicted for each molecule were visually inspected and one of them was selected according to the binding depth, and the occupation of the Trp23 sub-pocket was used for the MD simulation.

## 4.2.2 MD simulation

All-atom MD simulations were implemented using the AMBER 20 software on GPUs.<sup>30</sup> The initial structures were the binding poses selected from the docking results. NC series compounds were created based on the representative conformation of each fragment individually extracted from the trajectory, aiming to occupy three sub-pockets simultaneously as much as possible. Every two fragments that filled different sub-pockets were connected by single bonds with some saturated carbon or sulphur atoms and some removal of the ring structure if needed. Then, the energy of each NC molecule was minimised to remove bad contacts and optimise the structure. In addition, the *pdb-tools*<sup>59</sup> was used to correct the atom names, which is necessary for the system setup for the MD simulations. Each system consisted of one copy of the Mdmx protein (PDB ID: 3dab<sup>33</sup>) surrounded by TIP3P water molecules, and one chloride ion neutralising the whole system. The partial atomic charges of the ligands were derived using the *antechamber* module implemented in the AMBER software package to calculate AM1-BCC charges. The other force field parameters of ligands derived from GAFF2 and the AMBER FF14SB force field were employed to model the Mdmx. *Antechamber* prepared residue topologies for ligands and *LEaP* for Mdmx.

The MD simulation protocol was similar to that described in the previous chapter, and the differences are as follows. First, some systems of Mdmx complexes with newly created molecules (NT and NC series compounds) are required to reconstruct grid cells of the constant pressure and temperature simulation to reach the equilibrated density before the final sampling phase. Second, the simulations of eight fragments were run for 50 ns for each fragment. This duration was sufficient because the fragments were moderately small and had only a few degrees of freedom.

The RMSD was calculated based on the backbone atoms with the first frame used as the reference. The  $\chi_1$  angle of Tyr99 was an indicator of not only the open/closed state but also the shape change of the Leu26 pocket. The distance between Tyr99 and Met53 (the geometric centre of the ring to the sulphur atom) was mainly used to monitor the Met-aromatic interactions. The cluster analysis of the frames was carried out with a distance metric (RMSD of backbone atoms) using the K-means clustering algorithm. Both the representative and last frames were used to analyse the binding conformation.

## 4.2.3 Bioactivity test

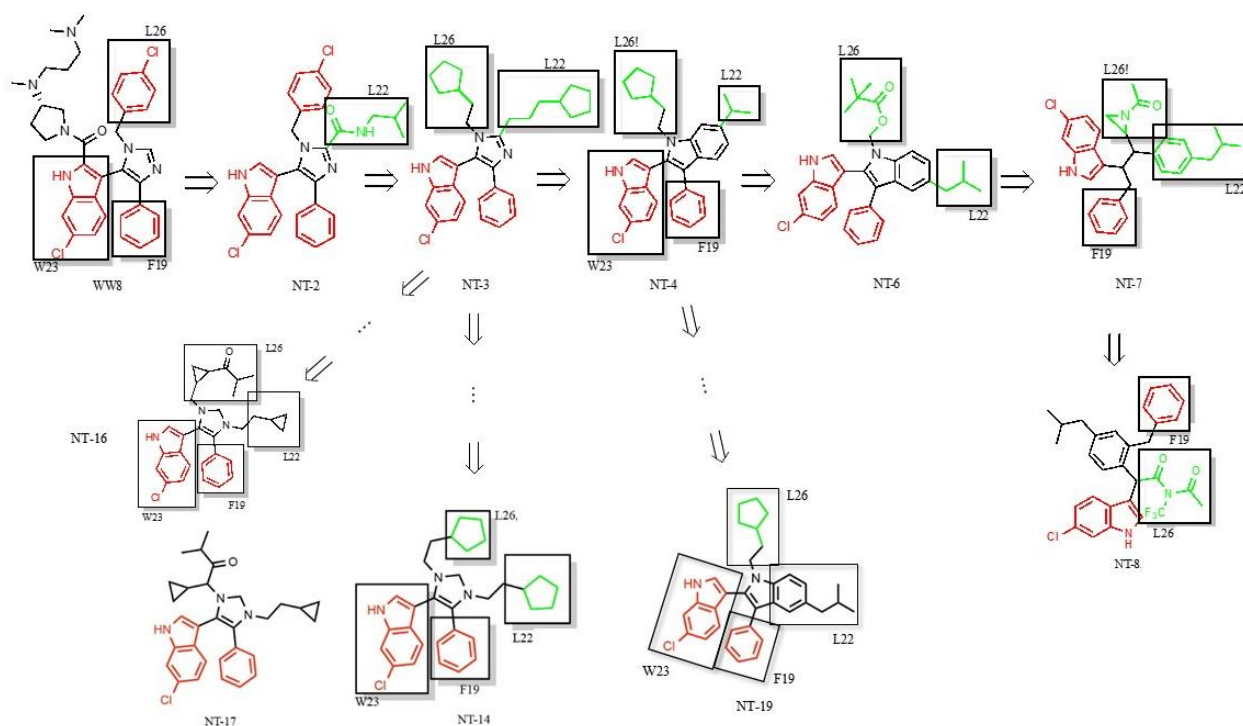
The inhibitory activity of SN compounds against Mdm2 was evaluated using a fluorescence polarisation (FP) assay (see Appendix I).

## 4.3 Results

### 4.3.1 Strategy 1

Since the goal was to obtain a molecule with four side chains superimposed with four key residues of p53, the first attempt was simply removing the aliphatic chain of the indole ring and adding a t-butyl group to the imidazole core connected by an amide group to mimic the Leu22 of p53. Next, a five-membered aliphatic ring was used alternatively to improve the occupancy of the Leu22 and Leu26 sub-pockets. The core from imidazole was changed to an indole ring and the position of substituent was also changed on the ring. Every change in the structure was guided and inspected by the most stable binding conformation selected from the trajectory. After more than 20 rounds of structural optimisation (Figure 4-4), the top eight compounds (NT-3, NT-4, NT-6, NT-7, NT-14, NT-16, NT-17, and NT-19) by ascending standard deviation of RMSD were selected for further analysis. As shown in Figure 4-5, all selected NT compounds entered into a stable phase at least by the end, except for NT-14. NT-3, NT-4, and NT-17 individually maintained stable binding conformations for a very long time, especially NT-3 with an RMSD of approximately 1. In addition, Mdmx binding with NT3 barely exhibited large conformational movements. According to the dynamic change of Tyr99  $\chi_1$  (Figure 4-6), NT-3, NT-16, and NT-19 stayed in the closed conformation for most of the simulation time; NT-4 and NT-7 maintained the open state shortly after starting the simulation. Combined with the distribution of distance between Tyr99 and Met53 through the trajectory (Figure 4-7), the Met-aromatic interaction was weakened when Tyr99  $\chi_1$  was in the open state. For example, the distance was mostly in the range of 7–8 Å for the NT-7 compound. However, binding with NT-4 maintained moderate Met-aromatic interactions and the open state simultaneously. Moreover, NT-16 strengthened the Met-aromatic interaction and maintained the native closed conformation of Tyr99. Because some trajectories had sharp conformational changes in the last frame, the representative structure of each NT compound was extracted from the last 70-ns simulation by cluster analysis, and their conformational characteristics are summarised in

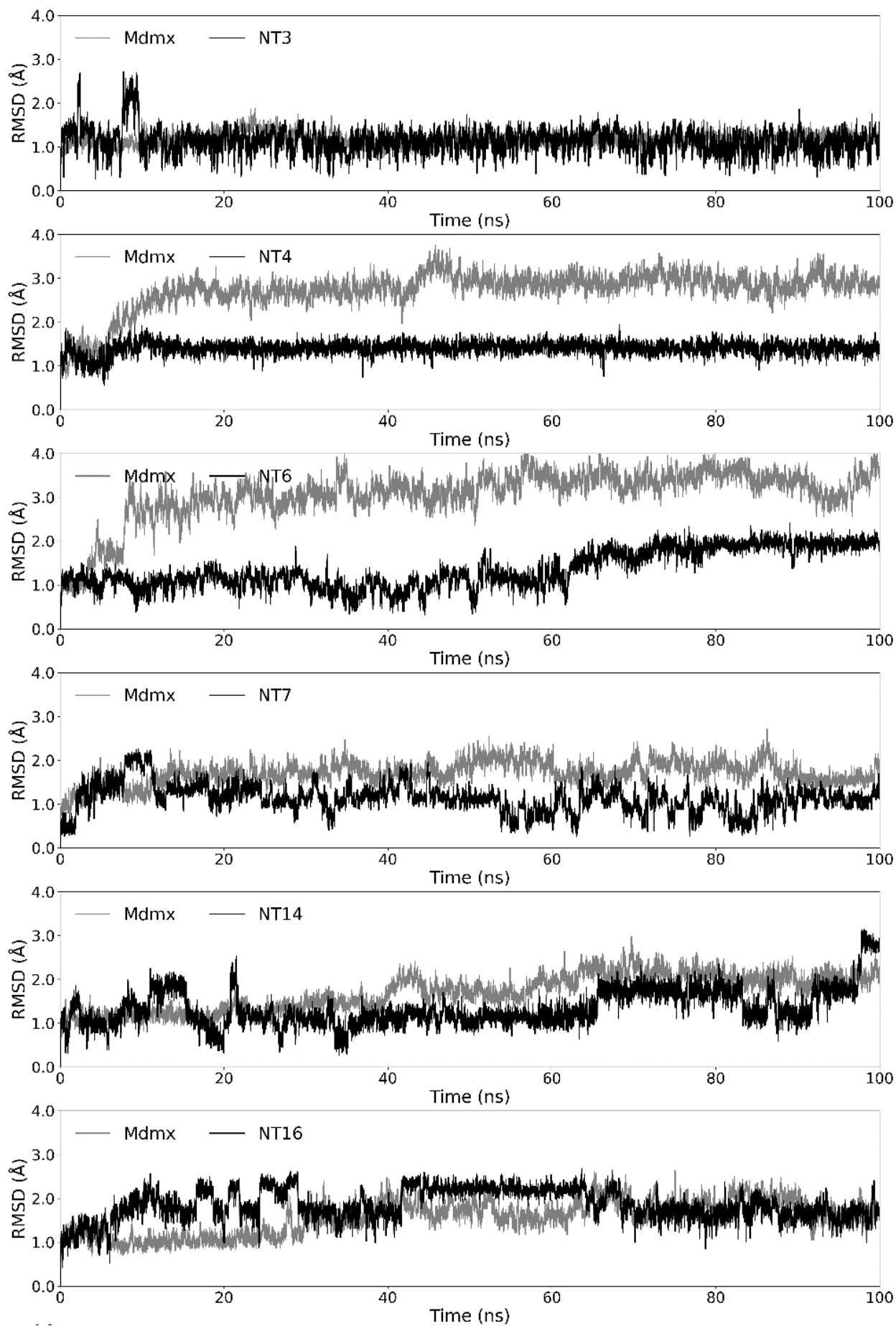
Table 4-1. Taken together, NT-3 and NT-4 had almost all the expected structural features.

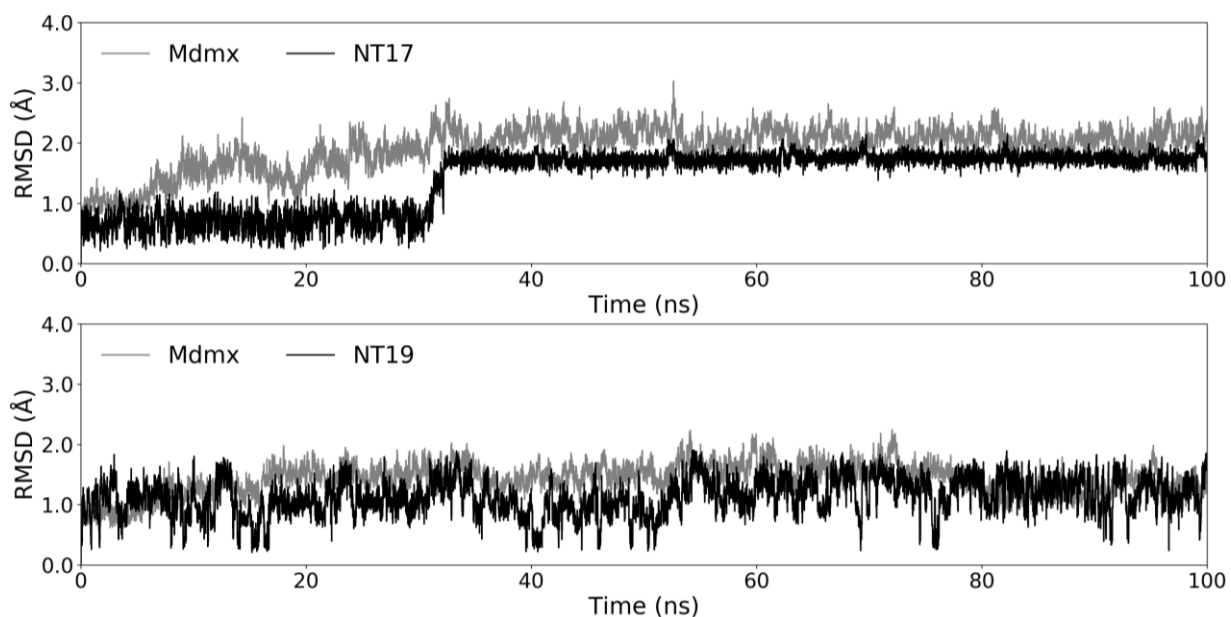


**Figure 4-4** The structural optimisation of WK298 to mimic the key residues of p53 including F19, L22, W23 and L26.

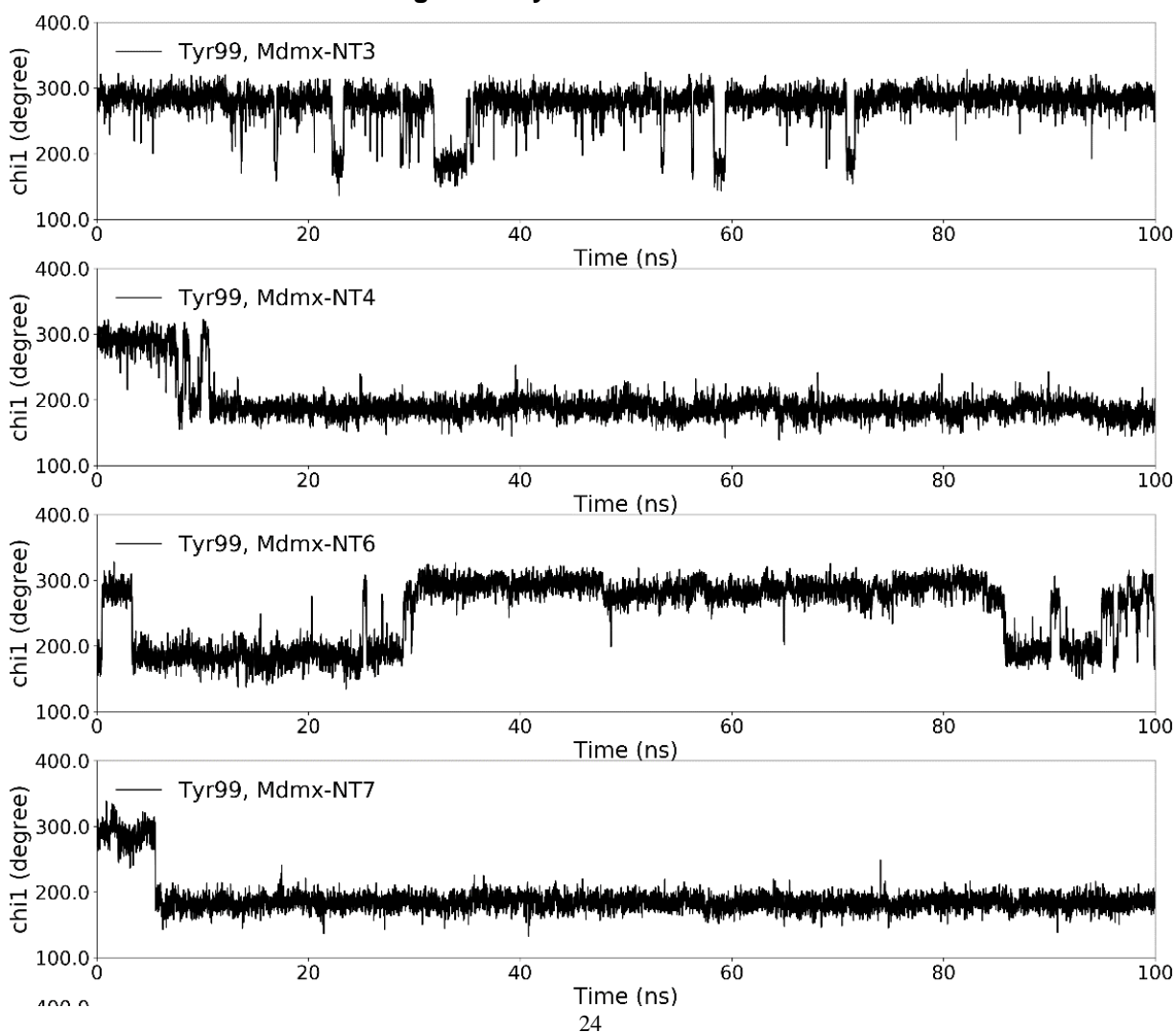
**Table 4-1** The characteristics of representative conformations of NT compounds

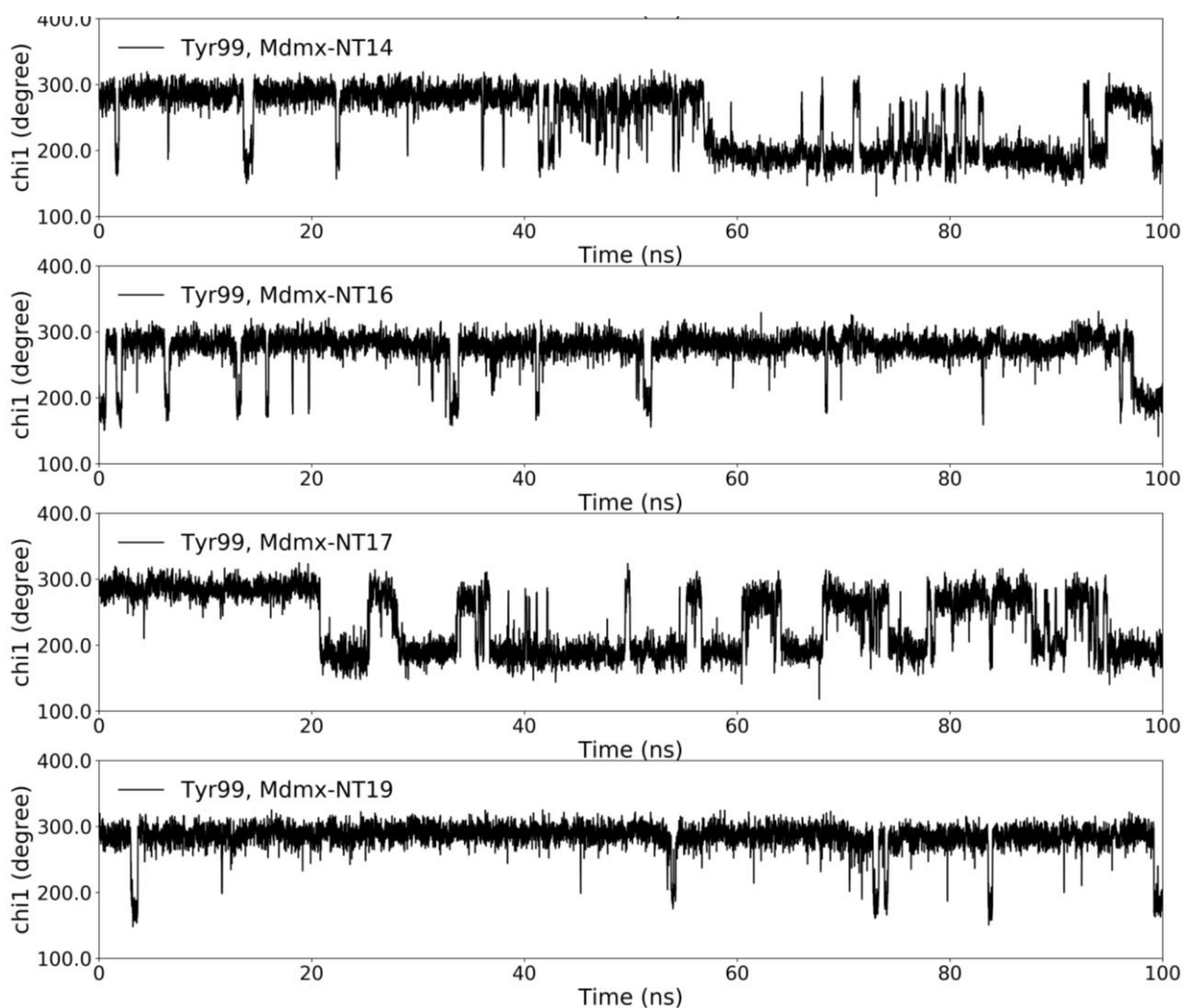
compound	Representative Frame No.	Tyr99 $\chi_1$ ( $^\circ$ )	Distance (Met53-Tyr99, $\text{\AA}$ )	Pro95
NT-3	5532	Closed	5.6640	Helical
NT-4	5470	Open	5.1773	Helical
NT-6	6624	Closed	5.1051	Loop
NT-7	3050	Open	6.9527	Helical
NT-14	8259	Open	5.8129	Loop
NT-16	5894	Closed	4.4204	Helical
NT-17	8284	Closed	4.9961	Helical
NT-19	4514	Closed	6.7319	Helical



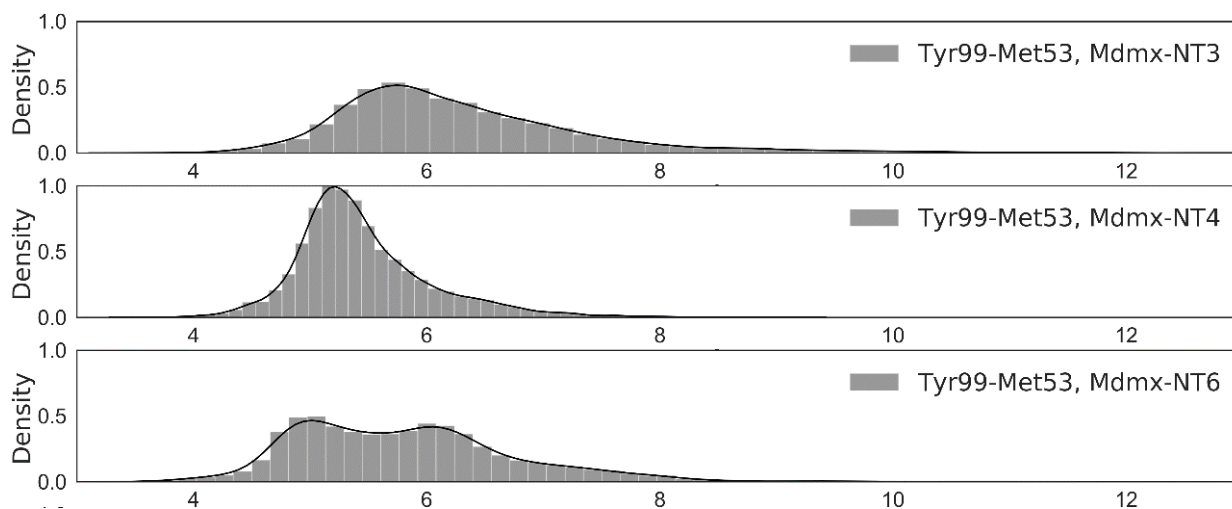


**Figure 4-5** The RMSDs (backbone atoms) of Mdmx complexes with selected NT compounds as a function of time were plotted through the 100-ns trajectories, each of which was calculated with the first frame as the reference. In addition to NT-14, the other NT compounds presented stably binding by the end. Particularly the Mdmx-NT-3 complex barely had large conformational movements due to significantly low fluctuations of RMSDs.

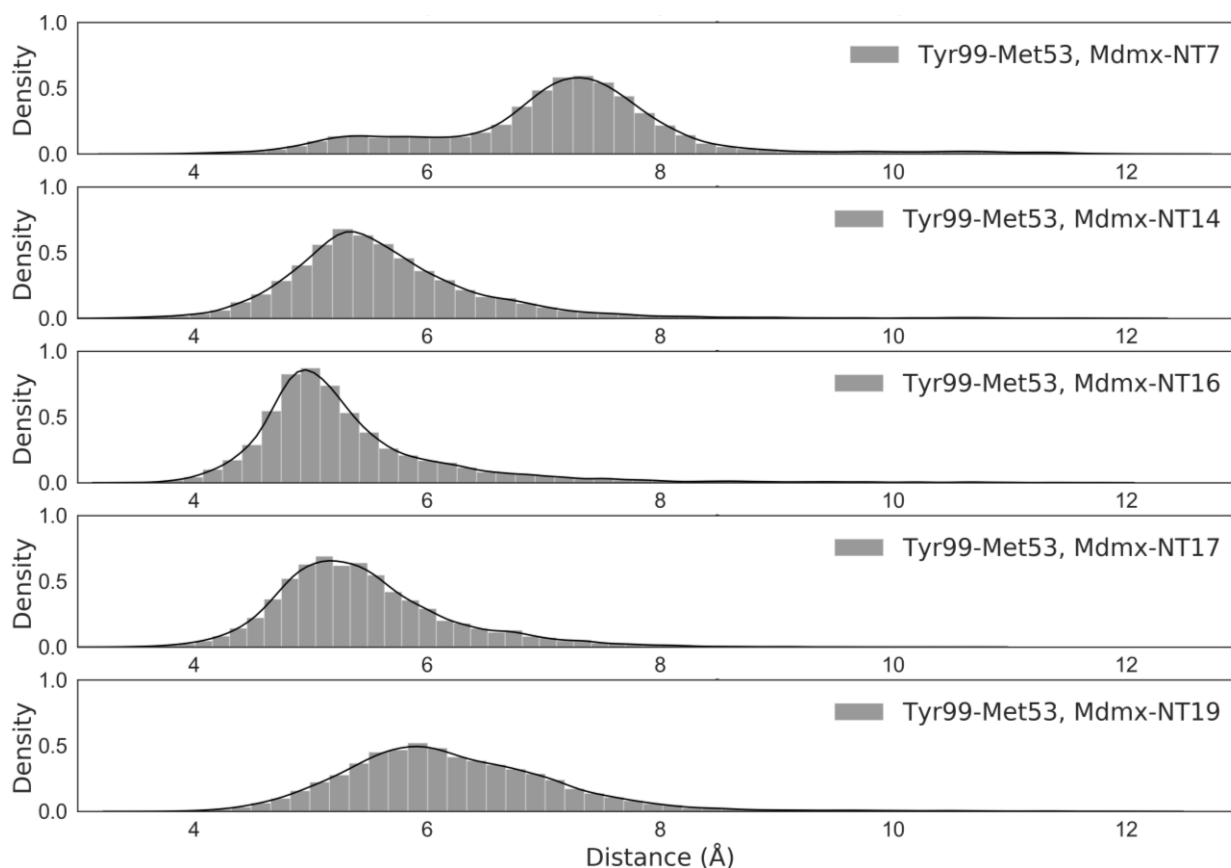




**Figure 4-6** The dynamic change of Tyr99  $\chi_1$  in Mdmx complexes with selected NT compounds through the 100-ns trajectories. NT-3, NT-16 and NT-19 stayed in the closed conformation most of the simulation time; NT-4 and NT-7 kept the open state shortly after starting the simulation.



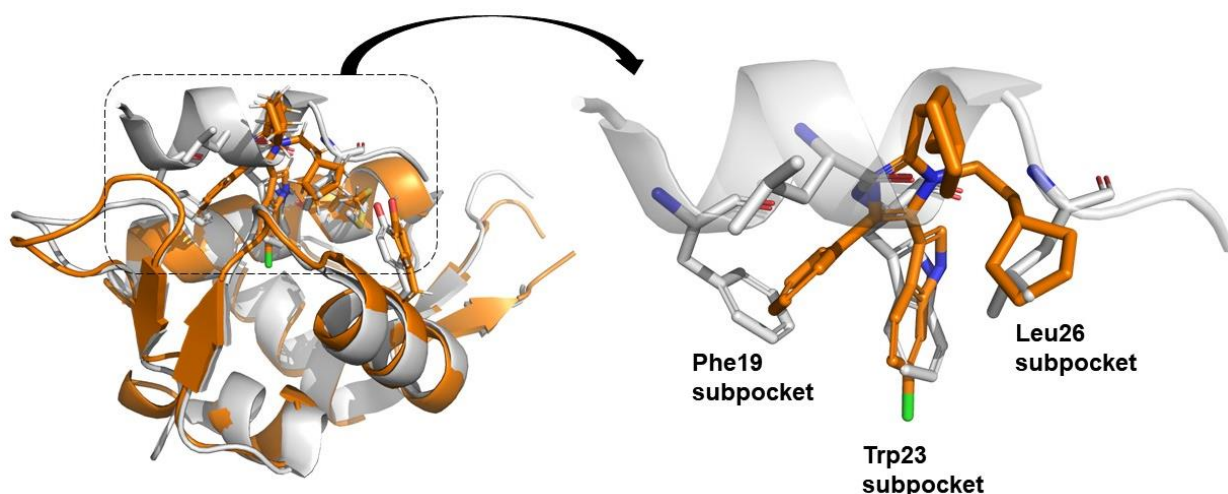




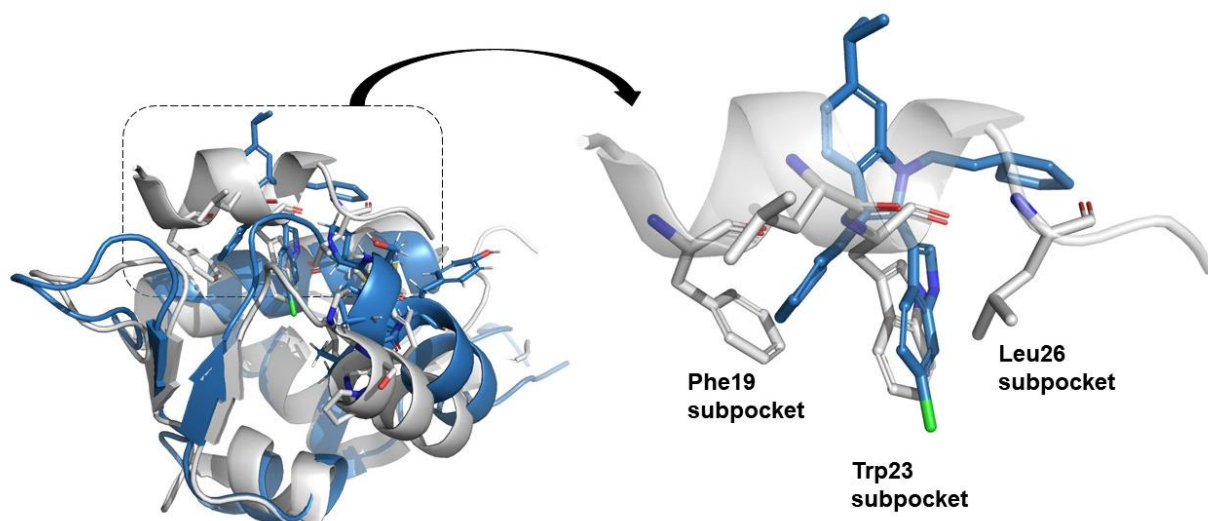
**Figure 4-7** The distribution of distance between Tyr99 and Met53 in Mdmx complexes with selected NT compounds through the 100-ns trajectories. Combined with Figure 4-6, the Met-aromatic interaction was weakened when the Tyr99  $\chi_1$  was in the open state, e.g. NT-7. However, NT-4 maintained moderate Met-aromatic interaction and the open state simultaneously, which would be further inspected for the binding mode.

Visual inspection revealed that the Mdmx-NT3 complex presented a similar conformation to the Mdmx-p53 complex, as shown in Figure 4-8. Consistent with WK298, the 6-chloroindole and phenyl ring still occupied the Trp23 and Phe19 sub-pockets, respectively. One of the five-membered rings was located in the Leu26 sub-pocket, while the other one was not superimposed well with Leu22 of p53, as expected. For NT-4 (Figure 4-9), only the 6-chloroindole group remained in the Trp23 sub-pocket, and the phenyl ring almost reached the Phe19 sub-pocket. The Leu26 sub-pocket was filled by Pro95 of Mdmx, and Tyr99 was largely open. Both movements made helix  $\alpha_2'$  closely approach the core of the binding pocket, which was narrower than the Mdm2 binding pocket. Furthermore, helix  $\alpha_2$  moved close to helix  $\alpha_2'$  in order to preserve the Met53-Tyr99 interaction.





**Figure 4-8** The alignment of Mdmx-NT-3 (the last frame of trajectory, orange) and Mdmx-p53 complex (PDB ID: 3dab<sup>33</sup>, white). This indicated NT-3 could perfectly mimic the classical binding mode of the key residues (Phe19, Trp23 and Leu26) except Leu22.



**Figure 4-9** The alignment of Mdmx-NT-4 (the last frame of trajectory, blue) and Mdmx-p53 complex (PDB ID: 3dab<sup>33</sup>, white). Although NT-4 kept the open state of Tyr99 and the short distance between Met53 and Tyr99, it was crucial that NT-4 only occupied the Trp23 subpocket, meaning it did not mimic the classical binding mode as well as NT-3.

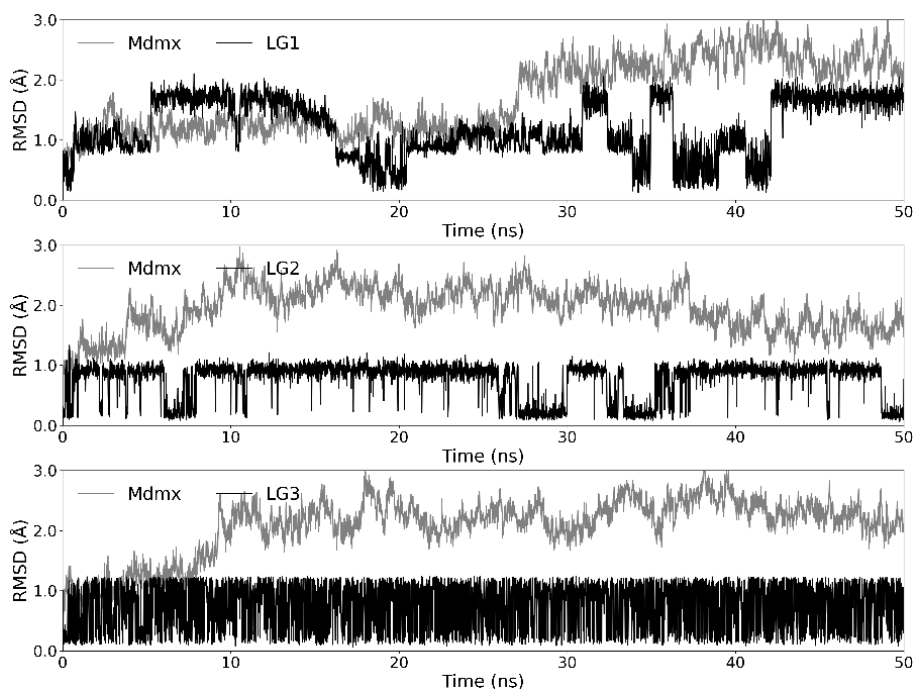
## 4.3.2 Strategy 2

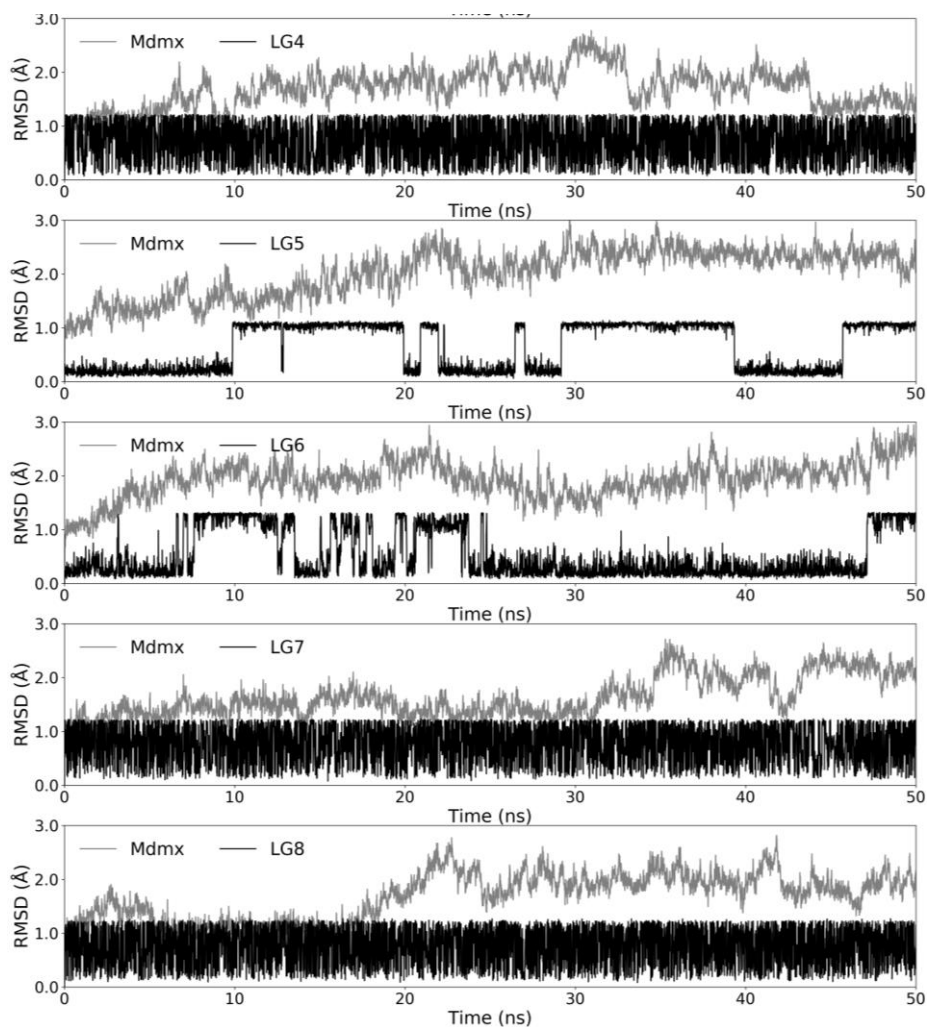
### *Analysis on the eight fragments*

For the eight active fragments, the selected best binding pose of each fragment was superimposed at least with Trp23 of p53, since they were small in size and could not occupy all the subpockets. Then a 50-ns MD simulation was run for each of them.

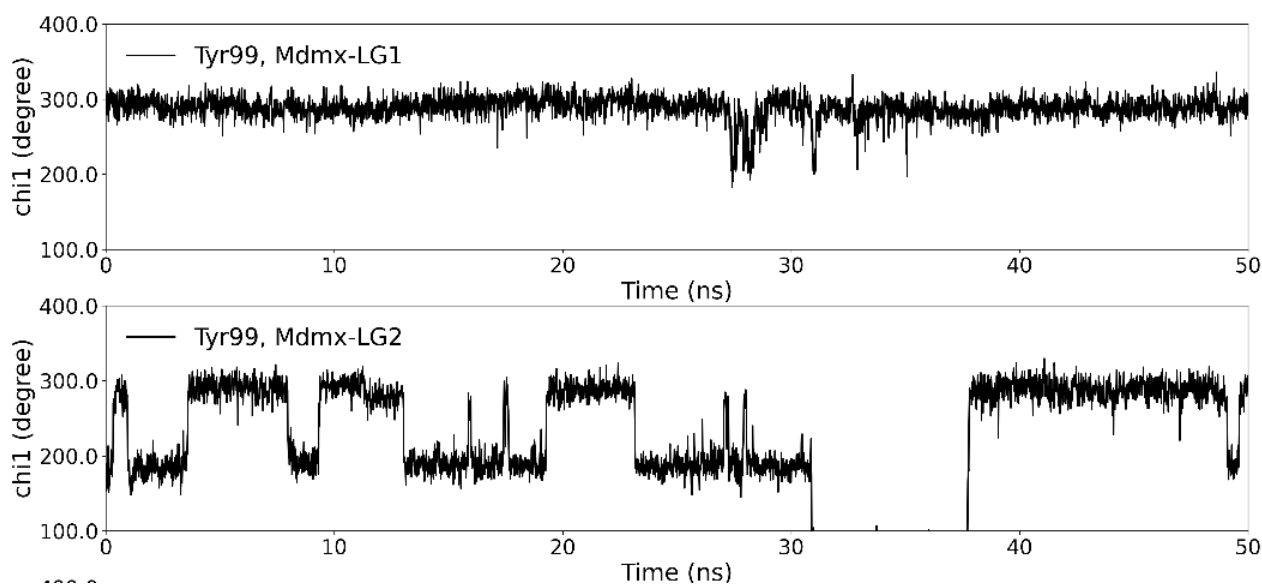
As shown in Figure 4-10, Mdmx was not stabilised by binding with these fragments compared to the apo Mdmx, and the LG2–8 fragments seemed to flexibly transform between two or more

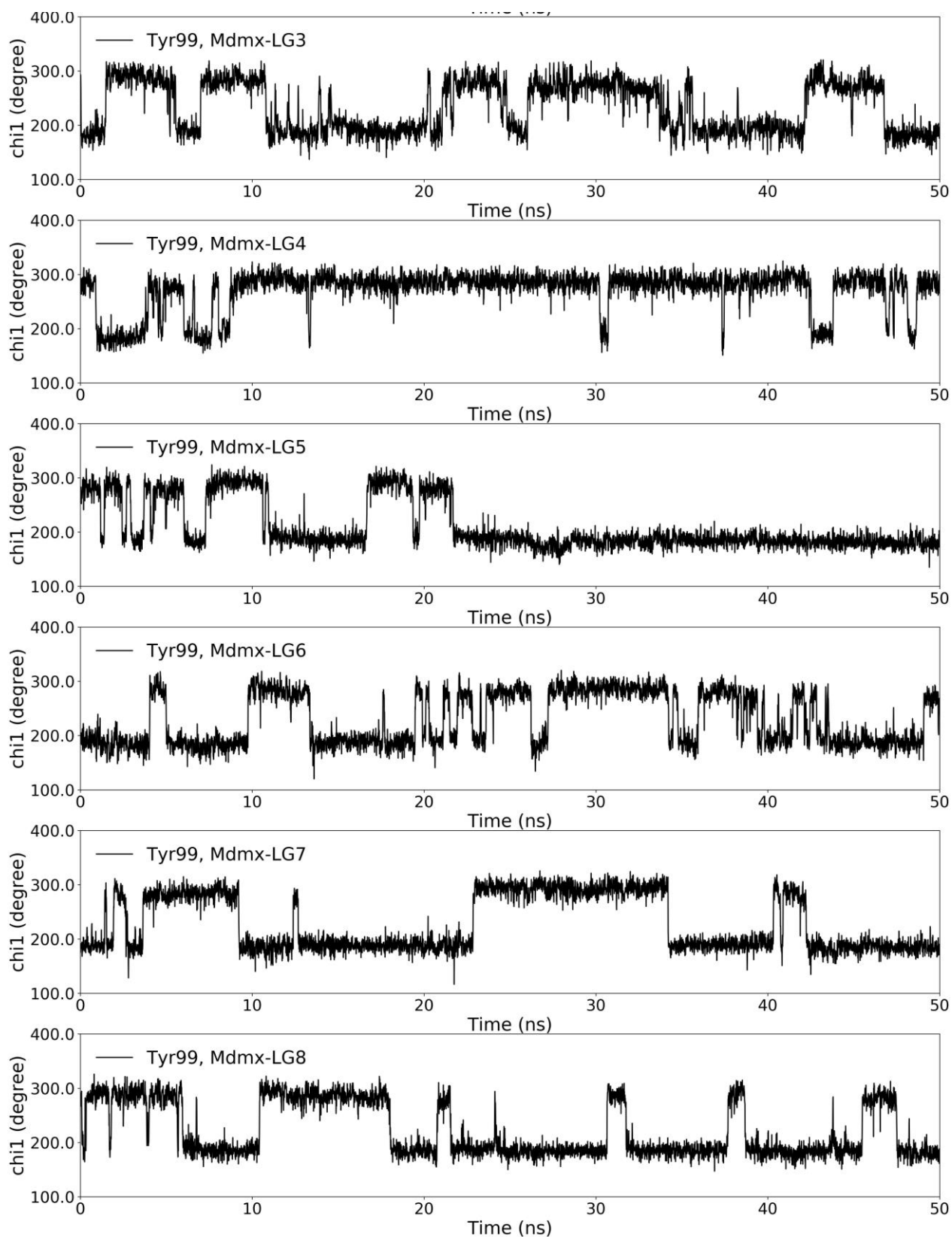
binding modes in the large hydrophobic cleft. However, LG1 presented various conformational changes but was likely to be steady during the last 7 ns. Moreover, the Tyr99 in the Mdmx-LG1 complex was nearly always closed, and the Met53-Tyr99 interaction was moderate over the entire 50-ns trajectory (Figure 4-11). In addition, when the time of Tyr99 being in the open state increased, the Met53-Tyr99 interaction weakened with the split peak of distance distribution shifting towards the right side, for example in the Mdmx-LG5 complex (Figure 4-12). Due to the short duration of the simulation and the frequent conformational changes in some cases, the representative structure was selected from the entire trajectory by cluster analysis. The characteristics of the representative conformations are summarised in Table 4-2. First, Pro95 retained its helical structure as the first residue of helix  $\alpha_2'$  in all cases. In the closed state, the Met53-Tyr99 interaction could be very strong, whereas this interaction would be lost in the open state along with relatively profound conformational changes. Intriguingly, Mdmx complexed with LG2 and LG4 closely resembled the Mdmx-p53 crystal structure, while these fragments left the Leu26 sub-pocket empty, inducing a few inward movements of helix  $\alpha_2$ . However, the unfilled Leu26 sub-pocket could also cause the inward bending of helix  $\alpha_2'$  (LG3, LG5, and LG8), especially in the case of LG5, where the Pro95 of Mdmx even reached the Pro27 sub-pocket passing by the Leu26 sub-pocket filled by Asp94 of Mdmx. Moreover, LG3 stopped in the core of the whole protein, which was the downward extension of Trp23 and Leu26 sub-pockets by the end of the simulation. As there were only two fragments occupying the Leu26 sub-pocket, LG6 and LG7 were necessary for the next fragment-based drug design and caused the Pro95-Ser96-Pro97 region to move inwardly. In the Mdmx-LG1 complex, the indole ring of LG1 interacted with Met53, at a distance of approximately 5.1 Å but LG1 failed to insert into the Trp23 sub-pocket and popped out of the pocket at the end of the simulation; thus, it would not be used as a synthon in the next step. It is worth mentioning that the pyrrolidine ring of Pro95 was very close to the pyrazole part of LG6 (4.1 Å), implying some interaction which could contribute to the binding affinity and be utilised later on.



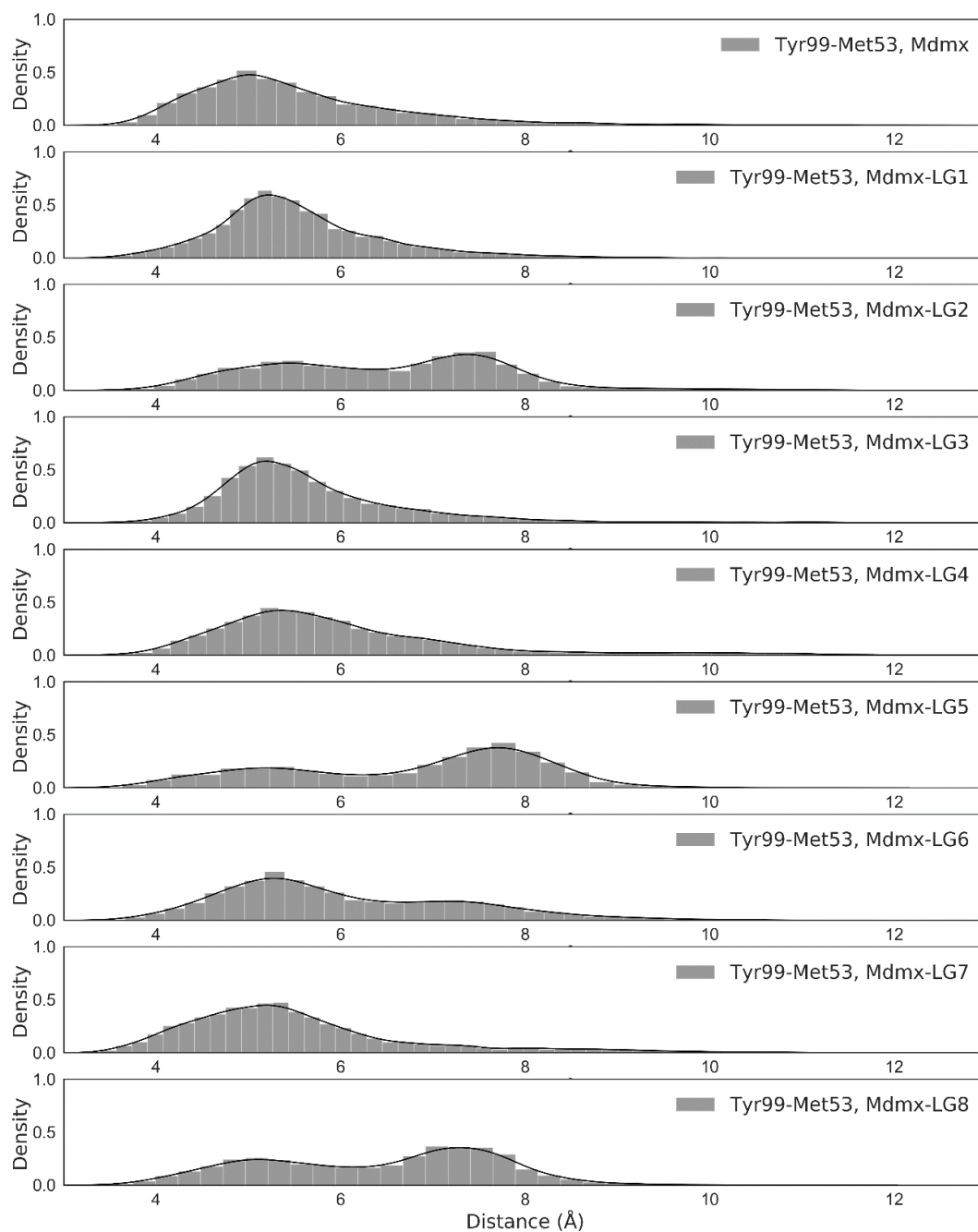


**Figure 4-10** The RMSDs (backbone atoms) of Mdmx complexes with 8 active fragments as a function of time were plotted through the 50-ns trajectories, each of which was calculated with the first frame as the reference. The simulation indicated Mdmx was not stabilised by binding with these fragments and the LG2 ~8 fragments seemed to flexibly transform among two or more binding modes in the large hydrophobic cleft.





**Figure 4-11** The dynamic change of Tyr99  $\chi_1$  in Mdmx complexes with 8 active fragments through the 50-ns trajectories. The Tyr99 frequently changed the conformations during the simulation for seven fragments except LG1 in which Tyr99 was nearly always closed.



**Figure 4-12** The distribution of distance between Tyr99 and Met53 in Mdmx complexes with 8 active fragments through the 50-ns trajectories. When the time of Tyr99 in the open state was increased, the Met53-Tyr99 interaction weakened with the split peak of distance distribution shifting towards the right side, e.g. as in the Mdmx-LG5 complex.

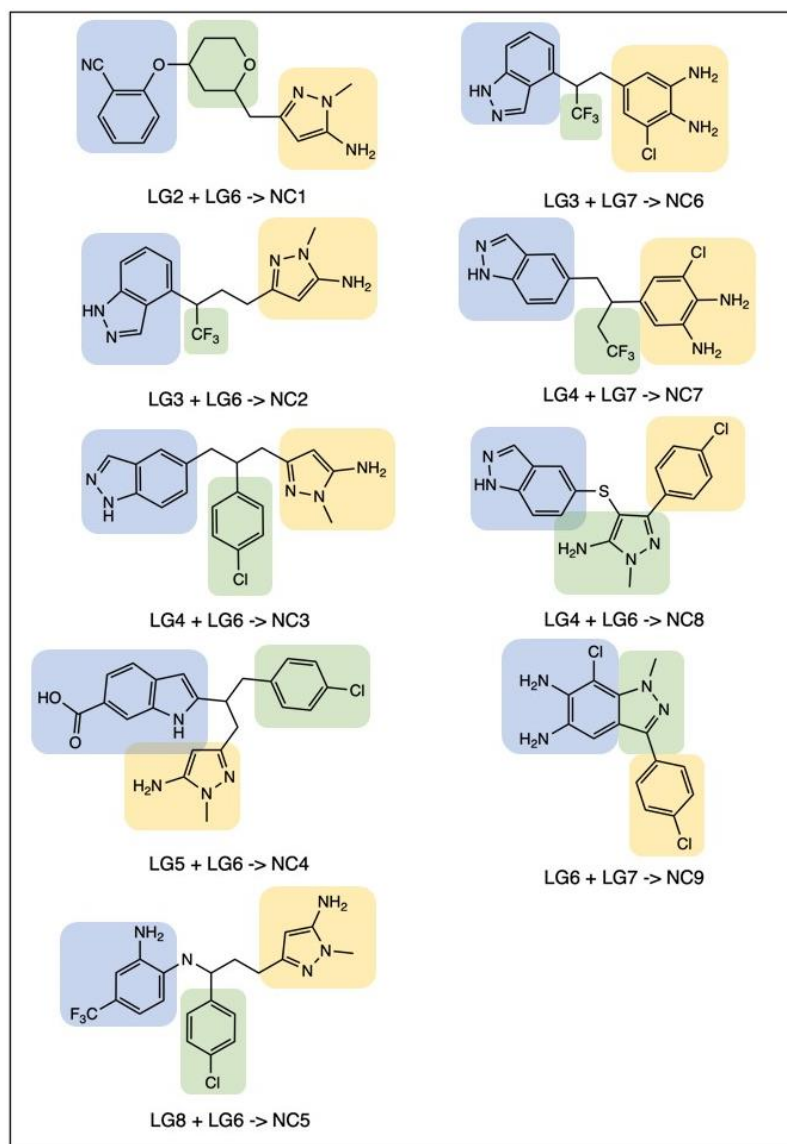
**Table 4-2** The characteristics of representative conformations of 8 active fragments. (a: the pocket occupation of a fragment; the main binding site is bold and italic.)

compound	Representative Frame No.	Tyr99 $\chi_1$ ( $^\circ$ )	Distance (Met53-Tyr99, Å)	Pocket occupation {P19:A,W23:B,L26:C} <sup>a</sup>
LG1	643	Closed	6.1523	-

LG2	4414	Closed	5.9581	<b>A + B</b>
LG3	3494	Open	5.7142	<b>B</b>
LG4	1929	Closed	6.3046	<b>A + B</b>
LG5	4806	Open	8.8951	<b>A + B</b>
LG6	3957	Open	7.6566	<b>B + C</b>
LG7	2598	Closed	6.1271	<b>B + C</b>
LG8	1532	Closed	4.7584	<b>A + B</b>

### Fragment-based drug design

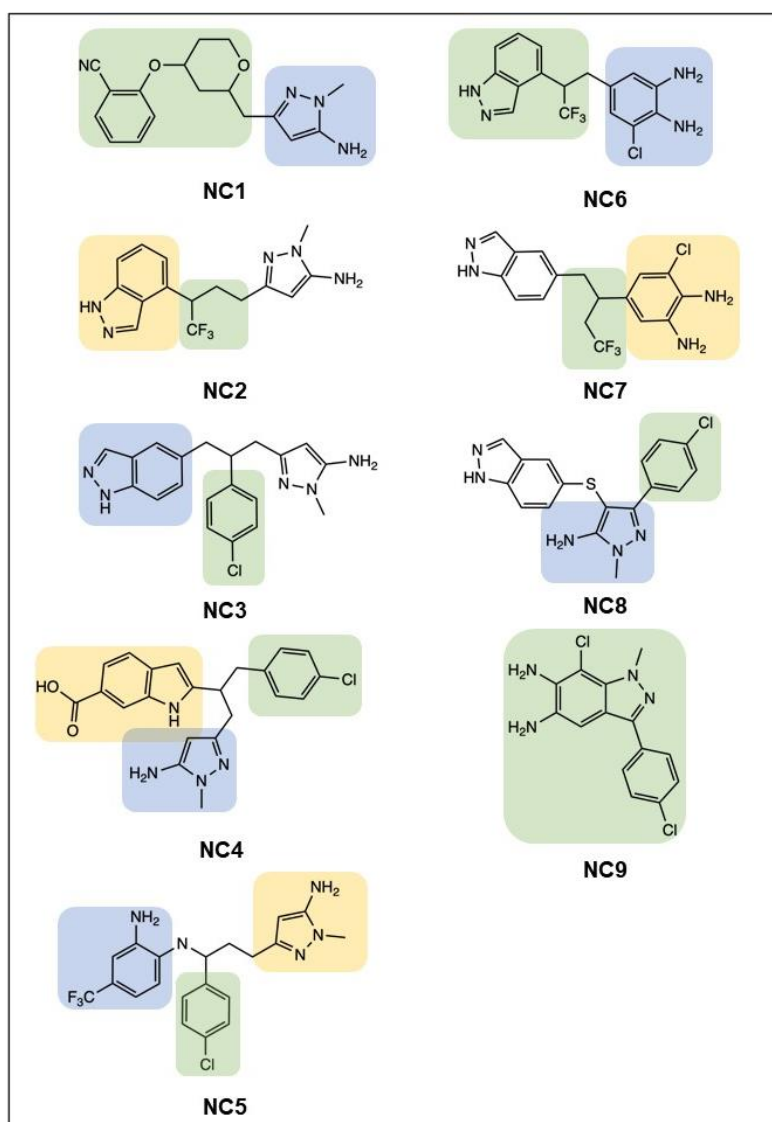
Based on the representative conformation of each fragment selected from trajectories, new and larger molecules were created consisting of groups expected to fill the three sub-pockets (Figure 4-13). MD simulations were implemented on these molecules for 100 ns (see Appendix II for RMSD, etc.). NC series compounds might transform a lot in the pocket because they did not have the typical three rings of WK298 that were like three fingers separately anchored in the three sub-pockets. Therefore, the frames extracted by cluster analysis based on the RMSD of the ligands were also used to visually inspect the binding modes.





**Figure 4-13 New created molecules (NC series) from eight active fragments (LG 1-8). The rounded rectangle marks a group expected to occupy one of the three sub-pockets; blue: Phe19 sub-pocket, green: Trp23 sub-pocket, yellow: Leu26 sub-pocket.**

As shown in Figure 4-14, all NC compounds did not run out of the pocket, but only NC5 had the same occupation of sub-pockets as expected. Moreover, NC4 occupied all sub-pockets with different substructures from the originally intended design. The pyrazole and trisubstituted benzene could occupy the Leu26 sub-pocket (NC5) as expected, but also the Phe19 sub-pocket (NC1, NC4, NC6, and NC8). For NC2 and NC3, the pyrazole ring extended over helix  $\alpha_2$  without filling any of the sub-pockets. The indazole ring could fill the Phe19 sub-pocket as expected but also the other two sub-pockets. In the case of NC7, the indazole was over helix  $\alpha_2$ . For NC8, indazole pointed towards the solvent. In most cases, the Trp23 sub-pocket was filled by trifluoromethyl and 6-chlorophenyl rings, as expected. Based on the above analysis, it was concluded that NC4 and NC5 would be promising for bioactivity testing as well as for further optimisation.

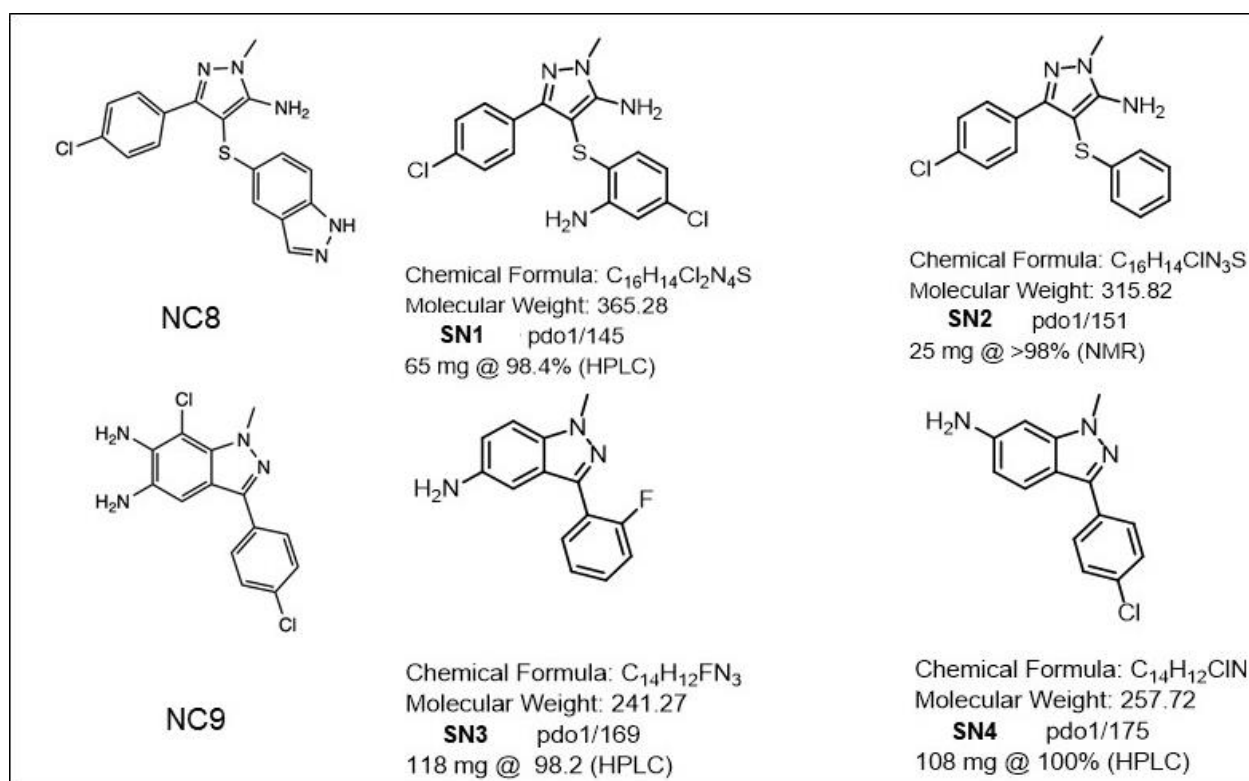


**Figure 4-14 NC series compounds. The rounded rectangle marks which sub-pocket is occupied by this part of molecule in the simulation; blue: Phe19 sub-pocket, green: Trp23 sub-**

**pocket, yellow: Leu26 sub-pocket. Here one can see NC4 and NC5 showed the possibility to occupy all the sub-pockets.**

#### *Analysis on the synthesised molecules*

Here, 10-ns trajectories were used to inspect the binding modes and decide which compound of the NC series would be promising. NC8 and NC9 had favourable binding modes with the Mdmx pocket and were used as templates for the synthesis process. Four synthesised derivatives (SN1-4, Figure 4-15) were tested for their inhibition of Mdm2, but all were negative. From a computational perspective, 100-ns MD simulations for each molecule were used to analyse the reason for the failure.



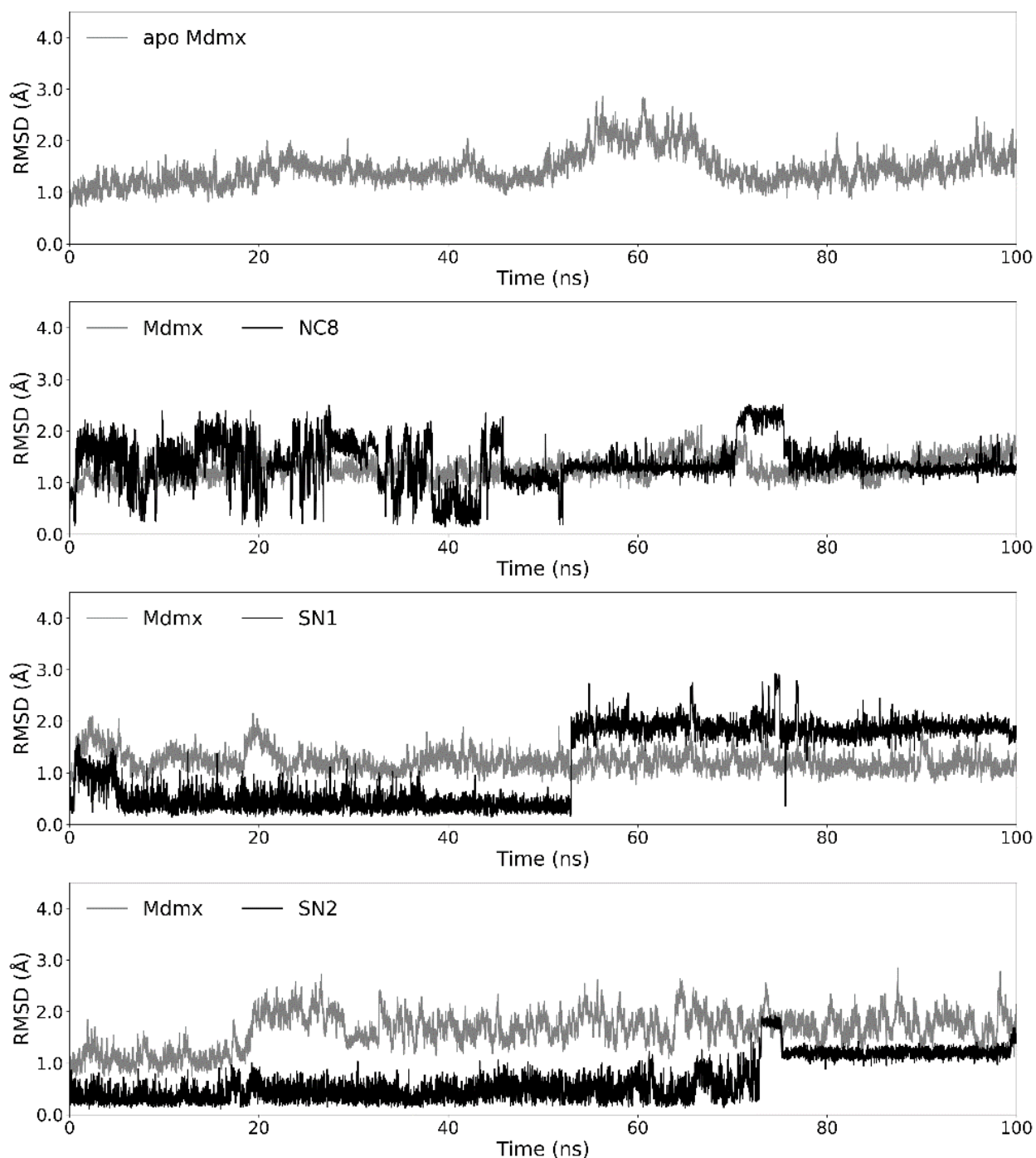
**Figure 4-15 Four derivatives from NC8 & NC9(SN1-4) were synthesised.**

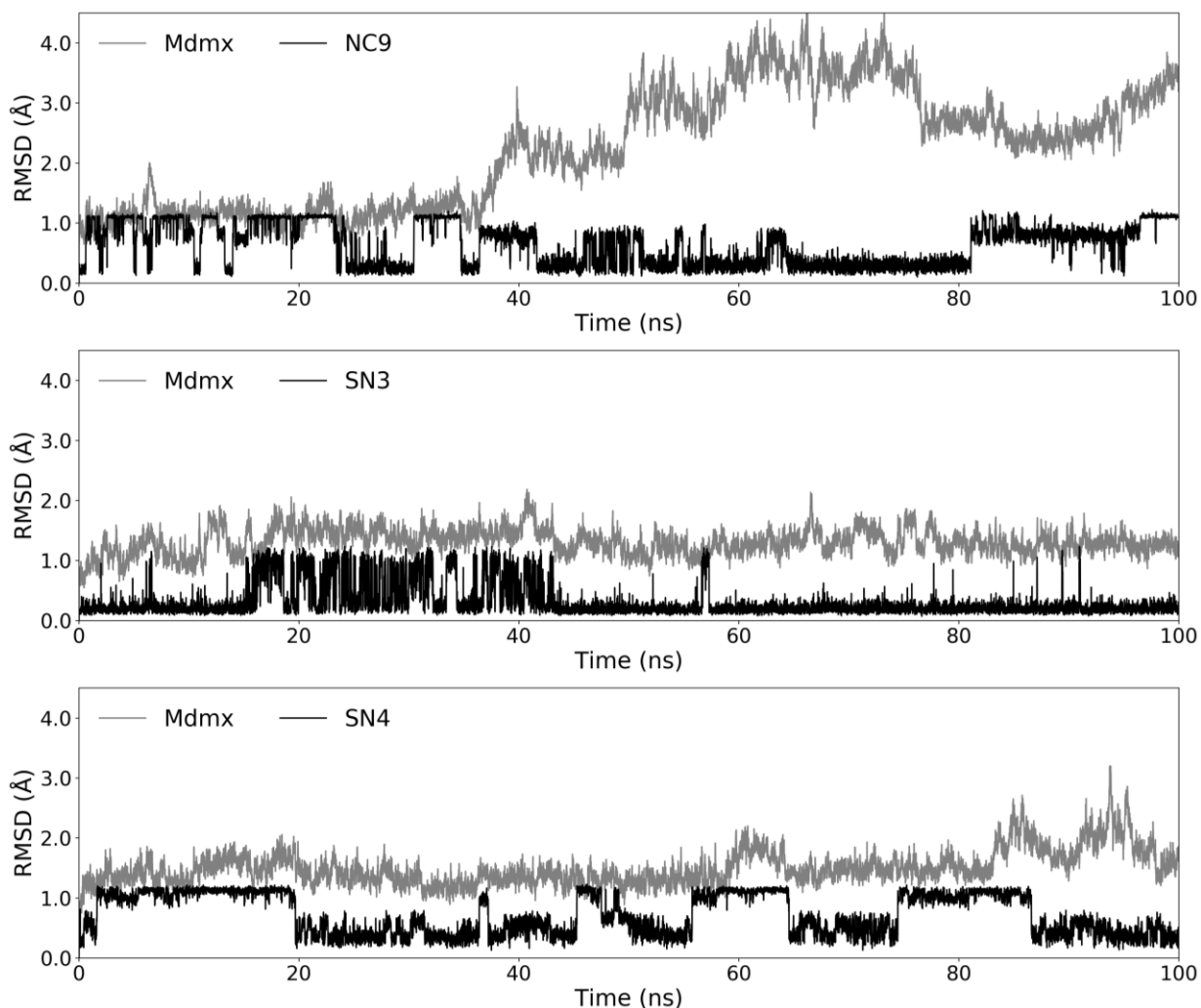
In Figure 4-16, Figure 4-16 the synthesised molecules presented stable conformations, at least in the last 10 ns. Mdmx also formed a dynamic equilibrium of conformations, which was especially improved compared to that of the template NC9. In Figure 4-17, Tyr99 of Mdmx was closed for most of the simulation time; in contrast, the template compounds showed more Tyr99 switches. Accordingly, both template compounds and derivatives maintained a relatively short distance between Met53 and Tyr99 (Figure 4-18). However, none of them had a similar distribution of Met-Tyr interactions to the Mdmx-p53 complex.

NC8 did not fill the Leu26 sub-pocket, and the Pro95-Ser96-Pro97 sequence bent inwardly. SN2 showed better alignments with Phe19 and Trp23 of p53 by its two phenyl rings, but still left the Leu26 sub-pocket empty. However, SN1 and NC9 had much weaker binding with Mdmx because they did not have a consistent binding mode according to their selected conformations (two centroidal frames from the cluster analysis on the protein and ligands as well as the last

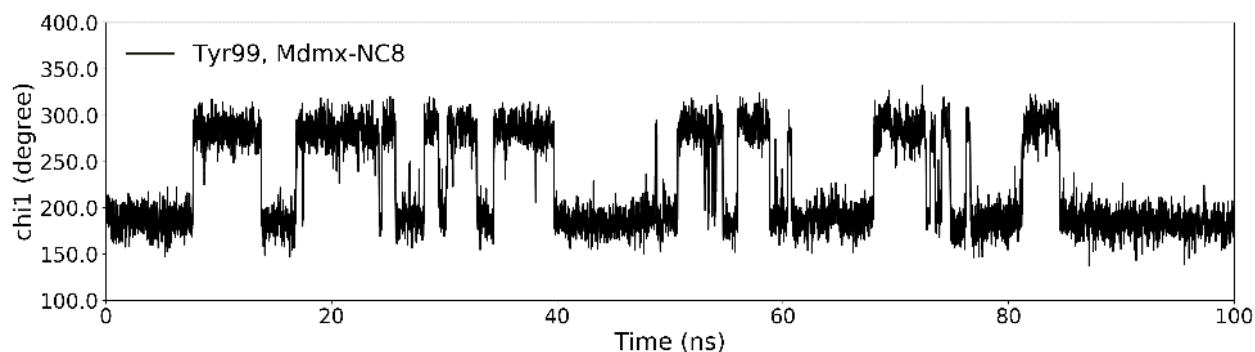


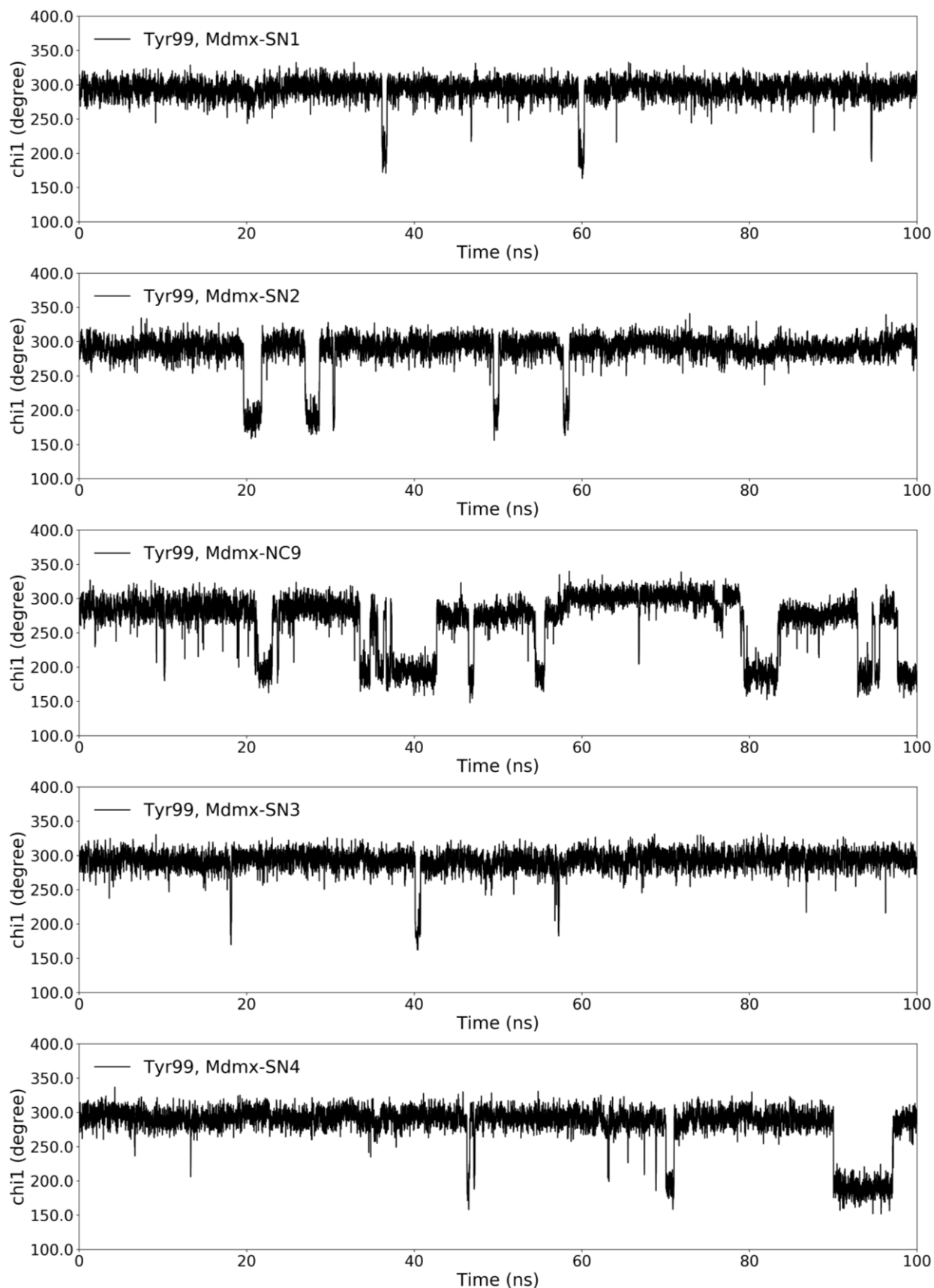
frame of the 100-ns simulation). Therefore, NC9 was not a good choice as a template. The indazole and phenyl ring of SN3 separately occupied the Phe19 and Trp23 sub-pockets. However, SN4 ran out of the pocket. Both the templates and the derivatives failed to fill the Leu26 sub-pocket and affect Tyr99, as well as its interaction with Met53. Figure 4-17 and Figure 4-18 show the natural movements of Tyr99 and Met53. While the template compounds selected from 10-ns trajectories remained in the pocket in the 100-ns simulation, one of the synthesised derivatives ran out of the pocket at the end. This indicated that a 10-ns simulation was not sufficient to acquire the stable binding mode of an Mdmx-small molecule complex.





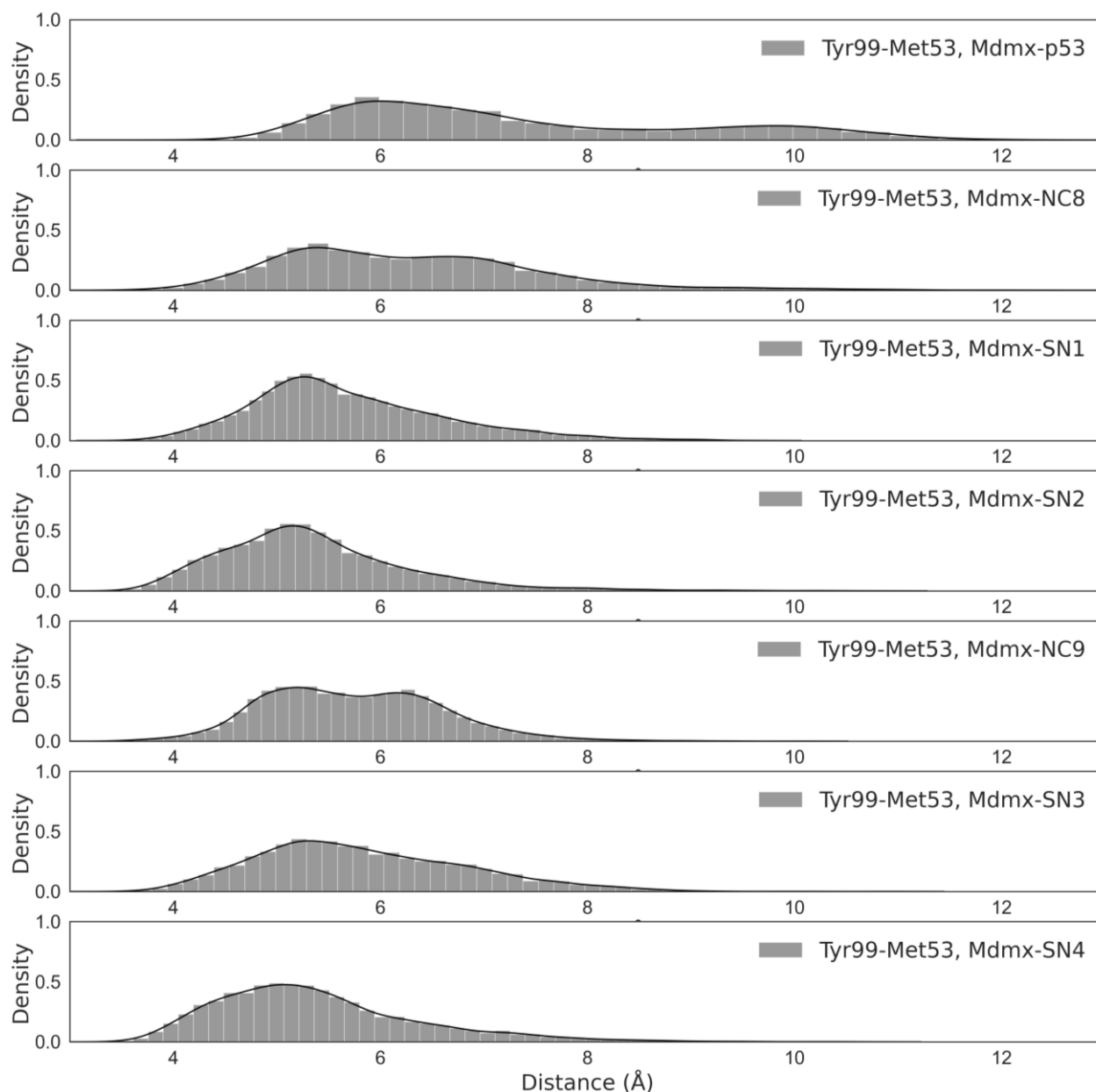
**Figure 4-16** The RMSDs (backbone atoms) of Mdmx complexes with the template compounds (NC8-9) and synthesised compounds (SN1-4) as a function of time were plotted through the 100-ns trajectories, each of which was calculated with the first frame as the reference. The synthesised molecules presented stable conformations at least in the last 10 ns. However, binding with the template NC9 did not stabilise the Mdmx in the simulation. NC9 and its derivative SN4 had frequent conformational changes.





**Figure 4-17** The dynamic change of Tyr99  $\chi_1$  in Mdmx complexes with the template compounds (NC8-9) and synthesised compounds (SN1-4) through the 100-ns trajectories. In all trajectories, the Tyr99 of Mdmx was closed most of time. By contrast, the template com-

pounds showed more conversions of Tyr99, indicating that Tyr99 did not form stable interaction with other residues.



**Figure 4-18** The distribution of distance between Tyr99 and Met53 in Mdmx complexes with the template compounds (NC8-9) and synthesised compounds (SN1-4) through the 100-ns trajectories. For all these compounds, Tyr99 and Met53 kept a short distance in between which was consistent with the general closed state of Tyr99 in their simulations. However, this common distribution was not observed in the trajectory of Mdmx-p53 complex.

#### 4.4 Summary and outlook

New Mdmx inhibitors were designed to mimic the binding mode of p53 from known active compounds using two structure-based strategies. The binding modes of the molecules were in-

investigated by MD simulations and evaluated by a series of post-analyses. Ligands binding both ‘open’ and ‘closed’ states of Mdmx were obtained.

The first strategy was to optimise the structure of the known Mdmx inhibitor, WK298. In this way, two compounds developed from WK298 were obtained that could bind with Mdmx, NT-3, and NT-4. The NT-3 compound formed a complex in which the conformation of Mdmx was very similar to that in the Mdmx-p53 crystal structure. The RMSD values were comparatively stable compared to those of the Mdmx-WK298 and Mdmx-p53 complexes. While NT-4 also had a stable trajectory after 10 ns simulation, it induced massive conformational changes in Mdmx. Interestingly, Tyr99 remained in the open state in the Mdmx-NT4 complex, along with an outstandingly strong interaction with the Met53 of Mdmx. In addition, extending the chain which connects the phenyl with the indole ring would be promising for improving the occupation of the Phe19 sub-pocket.

The second strategy was to rationally connect the active fragments to fully occupy the binding pocket. The bioactivity of NC series compounds developed from eight active fragments identified by NMR experiments was examined using a bioassay. The structures of the test compounds (SN series) were simplified during chemical synthesis. A bioassay was used to test the inhibitory activity of the test compounds against Mdm2. The negative results emphasised the importance of the duration of the simulation. At the beginning of the work, limited computational resources without GPU acceleration were available, so 10-ns trajectories were used to determine the synthesis template. Even though it was not necessarily the case that longer was better, in this instance a 10-ns trajectory did not have a high enough sampling of the entire dynamic range to acquire the stable binding mode for an Mdmx-small molecule ligand complex.

All the structural design and supposition in this chapter are worth further study and validation by wet experiments in the future.

# 5 *De novo* design new Mdmx inhibitors by machine learning methods

## 5.1 *Motivation*

Designing new chemical structures from scratch is a challenge. Given the almost endless number of possible organic compounds, simple combinatorial enumeration can do nothing valuable except for waste resources. Therefore, new approaches are being developed. One of the promising strategies is utilising so-called generative models, complex machine learning models that can mimic the original distribution, producing new samples in a Natural Language Processing (NLP) fashion. *In silico* structure production by generative models aims to efficiently design new molecules with desired properties against specific targets<sup>60,61</sup>. In this way, one can navigate into the unexplored area of chemical space not represented in existing libraries that are difficult to reach only by human cognition. It is believed that such compounds offer potentially better specificity, selectivity, and pharmacokinetic properties. They are also unlikely to be covered by existing intellectual property limitations.

The generative models are based on recent progress with sequence data processing methods. Examples of such data are natural language sentences or music where each subsequent piece of information (words or notes) depends on the previous parts. Since the first work on sequence generation revealed the possibility of recurrent neural networks to model the probability distribution of the next token in a sequence<sup>62</sup>, the scientific community immediately realised its power and opportunities. The chemical molecules can be presented as ordinary alphanumeric strings by means of a simplified molecular input line entry system (SMILES),<sup>63</sup> from which the model can learn the conditional probability distribution of the next symbol with respect to all previous tokens. Once trained, the model can generate a new and diverse set of valid SMILES that are similar to the training dataset<sup>64</sup>. Though the ability to generate an arbitrary compound from "nothing" looks promising and appealing, its practical application is not different from exploiting combinatorial libraries. Therefore, it is crucial to find a way to adapt the model for sampling new molecules from the desired region of chemical space. To this end, transfer learning techniques, where the pre-trained model continues training on a task-specific dataset, are widely used<sup>65</sup>. After several epochs, the model converges and begins producing SMILES, which are similar to such datasets.

A generative model was developed, which can computationally generate new molecules by using experimentally validated training sets. In this work, the generative model by the name of Generator was then used to discover potent Mdmx inhibitors.

## 5.2 Methods

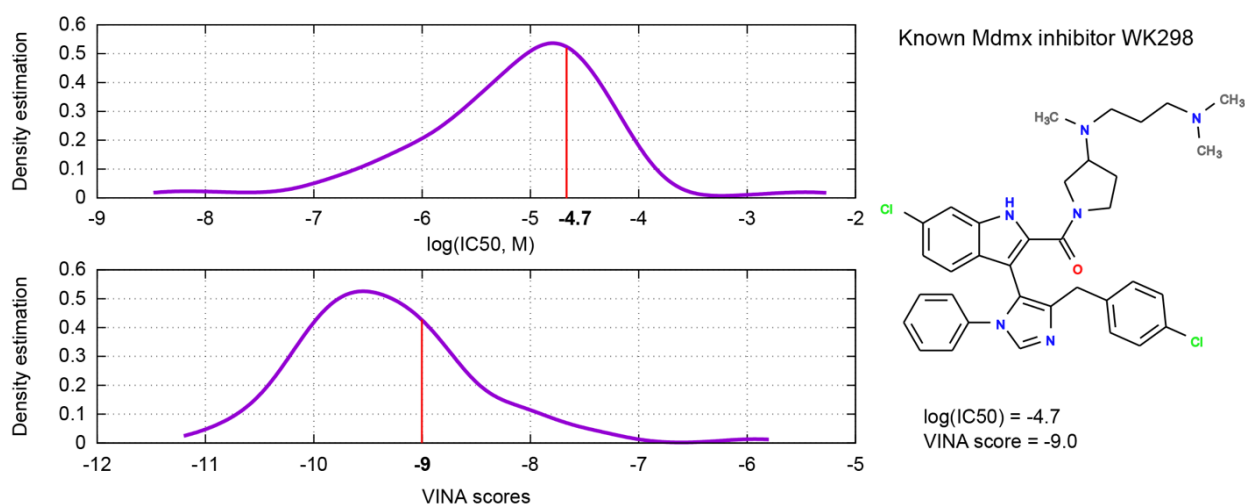
### 5.2.1 Datasets

#### ChEMBL library of SMILES

For our primary generative model, 1,727,112 compounds were selected from the ChEMBL library<sup>66</sup>. As distinct from all other works on SMILES generative models a pre-processing of the original data was not performed, for example, string size limitation, chirality removal, standardisation, and chemistry-relevant tokenisation. This was done with a goal of training a universal recurrent model that would be able to cover the whole ChEMBL database and then, finding weights (“embedding”) that could also be used separately from the original generative model. A total of 77,112 randomly selected SMILES were used as a validation set, while the remaining 1,650,000 SMILES were used as the training set.

#### Mdmx inhibitors

The training set of inhibitors (total 293) with experimental  $IC_{50}$  values,<sup>67–78</sup> which were tested with human Mdmx, were selected from the ChEMBL and BindingDB<sup>79</sup> databases. Their  $IC_{50}$  values were measured by several methods, such as the enzyme-linked immune sorbent assay (ELISA) (21 compounds), dose-response confirmation of inhibitors of Mdm2/Mdmx interaction in luminescent format (19 compounds), time-resolved fluorescence energy transfer (TR-FRET) assay (237 compounds), fluorescence polarisation assay (FPA) (2 compounds), and quantitative sandwich immune-enzymatic assay (14 compounds). Most of the inhibitors had  $\log(IC_{50})$  values less than  $-4.0$  (Figure 5-1).



**Figure 5-1** Distribution of  $\log(IC_{50})$  experimental values, and also VINA scores for the known Mdmx inhibitors

## 5.2.2 SMILES generator neural network (SGNN)

The generator neural network model consists of two long short-term memory (LSTM)<sup>80</sup> layers stacked upon each other with 512 internal units and a hyperbolic tangent as a nonlinearity activation function. Following the last LSTM layer, an ordinary dense layer with a SoftMax output was added (Figure 5-2). To facilitate batch computation, masking was applied to differentiate the positions in the input sequence as valid symbols. The input vocabulary involving stereochemistry and inorganic ions was fixed to 66 letters, including characters for the start (^) and the end (\$) of a string, as follows.

`$#%()+-./0123456789=@ABCFGHIKLMNOPRST- VXZ[\]abcdegilnoprstu^`

Each letter was one-hot encoded and then passed through the recurrent units. No tokenisation scheme was employed for the SMILES generation model because, as stated above, the primary goal was to train a universal model covering the entire ChEMBL database. Therefore, any restriction, even if it may seem chemistry-relevant, could lead to an under-representation of some classes of tokens in the training dataset, thus, worsening the quality of the entire model. The output of the SGNN is a probability distribution of the next symbol conditioned on all previous symbols in a string over the vocabulary. During decoding, the model randomly picks a symbol from the vocabulary according to this distribution. To balance the diversity and validity in sampling SMILES, the output of the Generator was adjusted to different temperatures (see details in Appendix III).

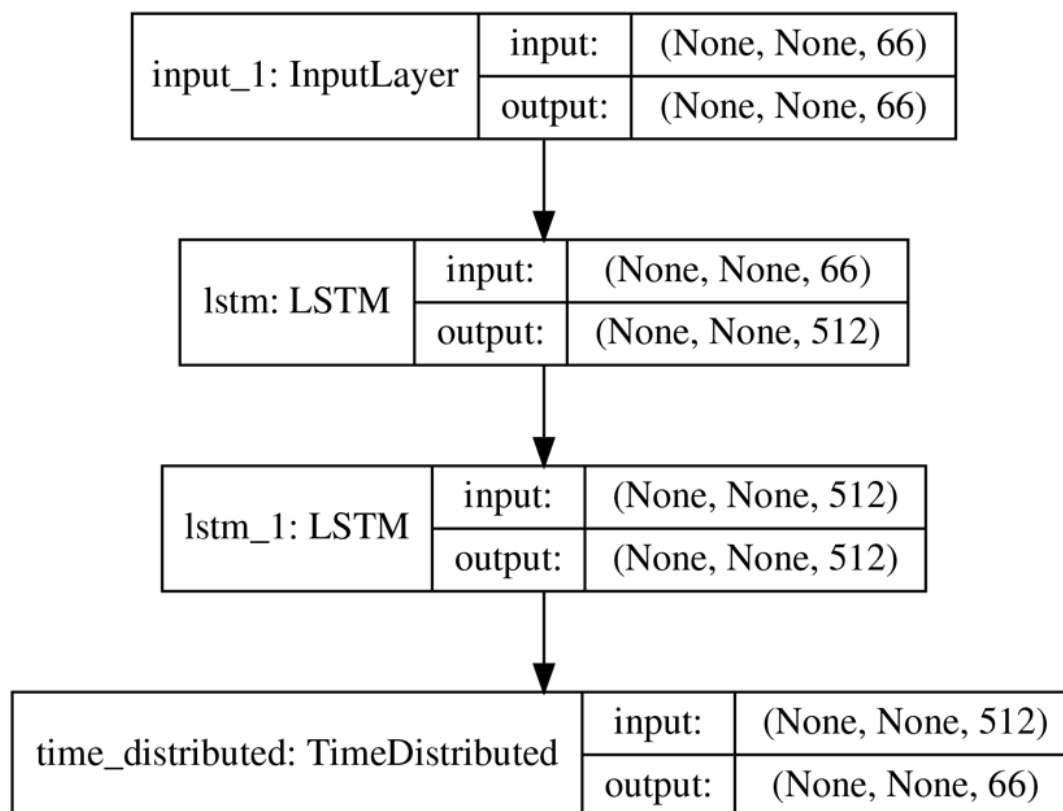


Figure 5-2 The architecture of the SGNN model



The SGNN was implemented in Lasagne<sup>81</sup> with Theano backend<sup>82</sup>, and SMILES validation was performed using the RDKit library<sup>83</sup>. The SGNN was trained for 100 epochs using the Adagrad optimisation algorithm<sup>84</sup> (constant learning rate 0.01, gradient clipping 100, batch size 2048). To avoid overfitting, the early-stopping technique with a validation dataset was used.

### 5.2.3 Molecular docking

The crystal structure of Mdmx in complex with p53 (PDB ID: 3dab<sup>33</sup>) at a resolution of 1.9 Å was used for the docking studies. The docking protocol was identical to that describes in the previous section. The ligands were docked using the AutoDock Vina<sup>29</sup> program. Ten predicted conformations were generated for each ligand. From the 10 binding poses, the optimal one was chosen in terms of docking score.

### 5.2.4 QSAR models for IC<sub>50</sub> and solubility estimation

To estimate the IC<sub>50</sub> values of the new Mdmx inhibitors, I built a regression model using the OCHEM platform. The OCHEM platform<sup>36</sup> automatically estimates the applicability domain (AD),<sup>85</sup> which is a crucial requirement for the current project. The Generator treats molecules as simple sequences of characters with predefined grammar rules. Although the final sequence may be a valid SMILES, it can result in a molecule that is chemically non-synthesisable or is very different from the training set for which IC<sub>50</sub> estimation may be unreliable. The QSAR model was used for additional filtering of the generated compounds by selecting only the molecules that are in its AD. Different machine learning methods and descriptors available on the OCHEM website were explored and the best model was based on a Transformer-CNN method<sup>37</sup>. The parameters of this model are the coefficient of determination ( $r^2 = 0.69$ ), root mean squared error (RMSE = 0.51), and mean absolute error MAE = 0.35 (n = 293). The model is publicly available at <https://ochem.eu/model/785>. All 293 compounds were used for modelling, although some values were measured using different methods because the number of points on individual assays was not sufficient to build and validate a model.

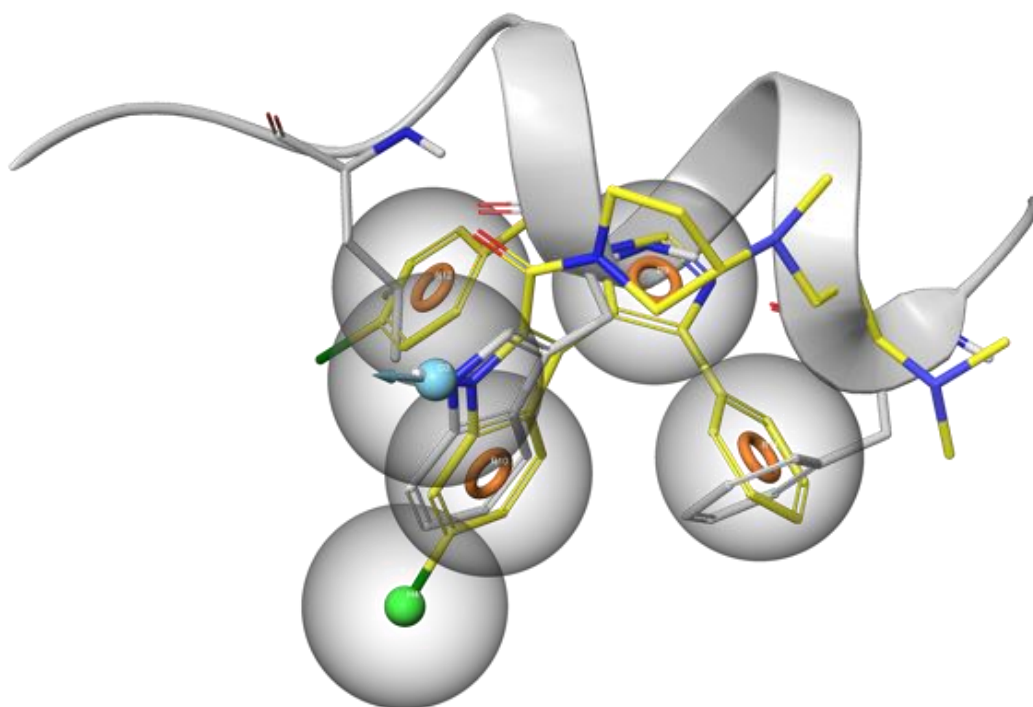
### 5.2.5 Tuning the Generator

After training the Generator on the ChEMBL database, the SGNN model was retrained on the transfer set, which in our case was the Mdmx dataset augmented (maximum 10 times) with non-canonical SMILES<sup>86,87</sup>. For example, toluene canonical SMILES is Cc1ccccc1, and non-canonical variants are c1cccc(C)c1, c1ccc(C)c1, c1c(cccc1)C, etc. The generated compounds which passed a Lipinski-like and a collection of Lilly filters<sup>88</sup> were then docked with AutoDock Vina.<sup>29</sup> Next, they were estimated the IC<sub>50</sub> values by the QSAR model. The new molecules were ranked by an integrated score from the IC<sub>50</sub> values and docking scores. A new training dataset was formed from the original data and highly scored putative ligands, keeping a ratio of known

and generated ligands 1:1. The SGNN model was retrained on this new dataset. By repeating the above tuning steps, promising inhibitors were selected based on the integrated score after finishing the default generation cycles.

## 5.2.6 Virtual screening

To perform virtual screening of the output of the Generator, a pharmacophore model was built using the Phase of Schrödinger suites<sup>89,90</sup> based on the Mdmx-WK298 complex. The complex was obtained by docking WK298 into Mdmx, which was extracted from crystal structures with PDB IDs of 3ljb<sup>50</sup> and 3dab<sup>33</sup>. The top-scoring binding pose was well superimposed with the crystal structure of the WK298-Mdmx complex, and when it was aligned with the crystal structure of the p53-Mdmx complex, one can see WK298 binds to Mdmx in a way that mimics the binding of the native p53 peptide. The Trp23 pocket is filled with the 6-chloroindole substituent, and the 4-chlorobenzyl ring inserts into the Leu26 pocket. Finally, the third key substituent of WK298, the phenyl ring, occupies the Phe19 pocket, although the plane of this ring is nearly perpendicular to the plane of the Phe19 side chain of p53. The pharmacophore model (Figure 5-3) was based on the aforementioned critical residues of p53 (Phe19, Trp23, and Leu26). The model was validated internally in the Schrödinger suite. In total, 204 of 274 active compounds were successfully retrieved, and the parameters were area under the curve (AUC) 0.72, EF1% 10.62. This pharmacophore model was used to screen all stereoisomers of the new molecules selected from the output of the Generator. Finally, the molecules were analysed using MD simulations.



**Figure 5-3** The pharmacophore model of Mdmx inhibitor used in this study (yellow: WK298, white: p53)

MD simulations were performed with AmberTools18<sup>91</sup>, including ligand reduction with *reduce*, atomic charge calculation with *antechamber* and AM1-BCC as the backend, and topology and coordinate file generation with *tleap*. Each MD simulation was run for 10 ns of production time under the NPT ensemble with a constant 0.987 atm with isotropic position scaling and 300 K using the weak-coupling and SHAKE algorithm; a maximum time step of 2 fs was used. The default value of the nonbonded cutoff was used along with the particle-mesh Ewald (PME) calculations. For the stable ones with a small standard deviation of the RMSD value (“std” in Table 5-2), the MMPBSA.py python script was further used to obtain the relative free energies of binding<sup>34</sup>. Post-processing trajectories according to the MM/PB(GB)SA approximation allows a better evaluation of the binding patterns of these ligands to Mdmx and, therefore, can help to identify the most potent hits<sup>92</sup>. One hundred frames were selected from the last nanosecond of simulation with 10 ps intervals to calculate  $\Delta G$  values. To reduce the computational time, the  $\Delta G$  calculation was not refined by the entropy contribution. Therefore, the calculated  $\Delta G$  was only an approximation of the true free energy, but could nevertheless be used to compare the generated ligands with similar Mdmx binding patterns.

The parameters and environment for MD simulations worked well on the WK298-Mdmx complex. Stable trajectories could be obtained by running 10 ns MD simulations on this complex, according to the RMSD (see Appendix III).

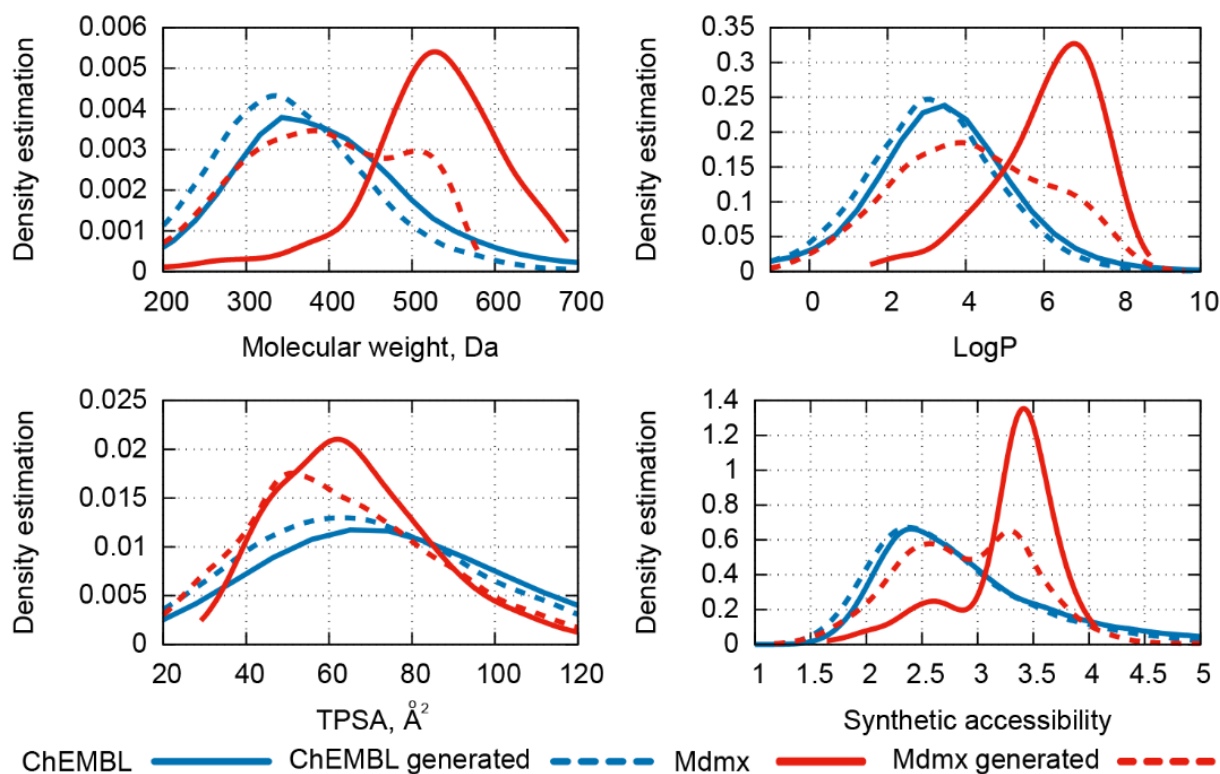
## 5.3 Results

### 5.3.1 The Generator

The ability of the Generator to produce chemical structures was examined by comparing the distributions of molecular weight, lipophilicity, topological surface area, and synthetic accessibility introduced in the work of Ertl et al<sup>64</sup>. For this purpose, approximately ten million SMILES were sampled from the SGNN, 79% of which were valid unique SMILES (see Table 5-1). The results in Figure 5-4 (blue dotted and solid lines labelled with "ChEMBL" and "ChEMBL (generated)") showed a good correlation between distribution shapes for different properties.

**Table 5-1 Parameters of generated molecules**

Strings	All	Valid SMILES	Unique SMILES
All	10,495,701	9,377,274 (89.3%)	8,297,705 (79.1%)
Stereo (with @)	1,677,294	1,388,454 (13.2%)	1,325,236 (12.6%)
Cis/trans (with / or \)	1,153,284	998,506 (9.5%)	905,536 (8.6%)



**Figure 5-4 Distribution of molecular weight, logP, TPSA, and synthetic accessibility for original ChEMBL database, generated molecules before and after tuning SGNN model with Mdmx inhibitors**

The Mdmx transfer dataset of known inhibitors consisted of 293 molecules. It was augmented to 2916 SMILES and then used to retrain the original ChEMBL model. While an early stopping technique was not required for training using ChEMBL, it was essential in this step because the number of parameters of the model was large compared to the amount of training data, and the model could easily overfit<sup>93</sup>. In the experiments, to 3-4 epochs were sufficient for transfer of the information. Total 5803 valid SMILES that passed drug-like Lilly filters were generated during all 10 cycles of tuning and 778 with VINA score less than  $-7.5$  were selected for further virtual screening. The visualisation of the chemical space of the generated inhibitors (Figure 5-5), based on the deep neural network model,<sup>94</sup> reveals an interesting landscape consisting of a valley with the known WK298 inhibitor in the middle and two ridges. Most of the compounds selected by MD simulations were located at the highest and most occupied ridge to the left of the valley. The landscape evidently shows that the Generator explored particular directions in the vast chemical space, and it was not a pure random walk.

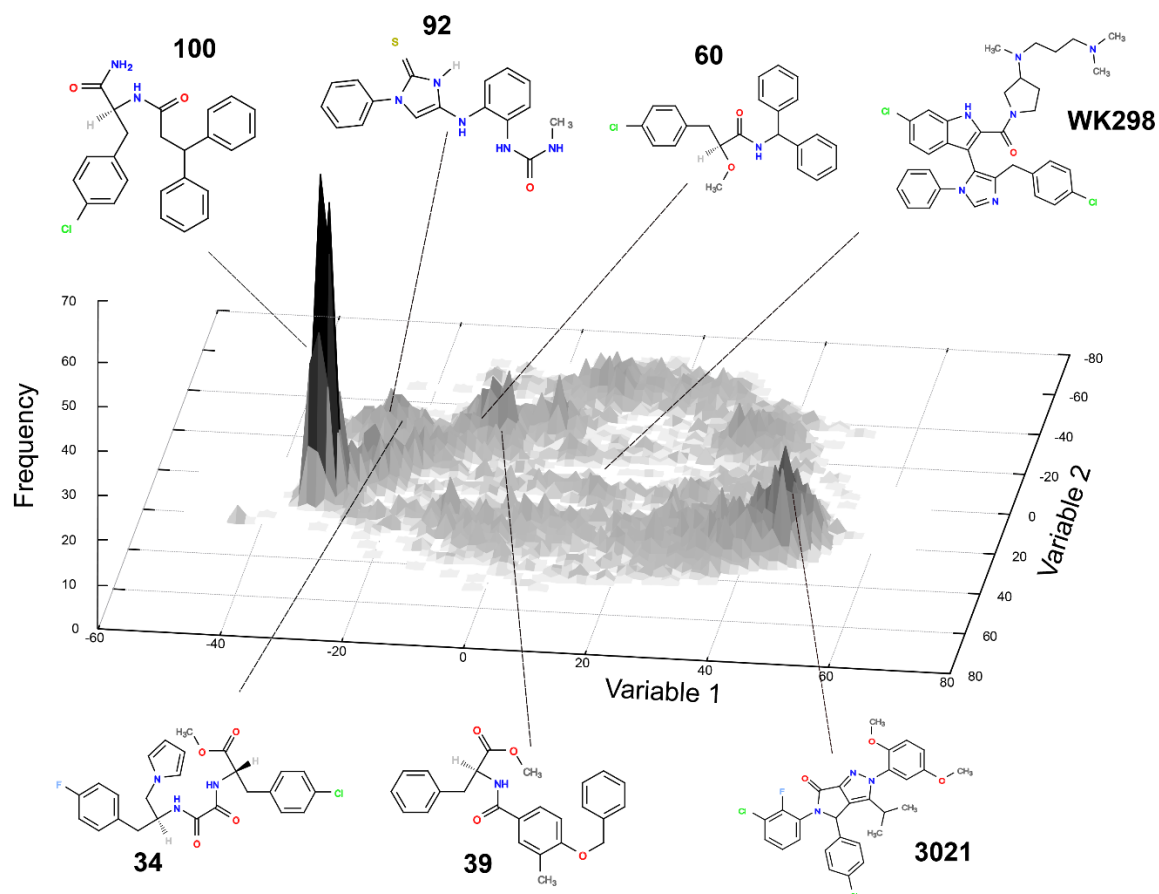


Figure 5-5 Visualisation of chemical space of Mdmx inhibitors

### 5.3.2 Putative Mdmx inhibitors via virtual screening workflow

The 5803 molecules, which passed advanced drug-likeness filters, were produced by the Generator and were ranked according to the sum of normalised VINA scores and  $pIC_{50}$  values (Figure 5-6). Then, the pharmacophore model filtered the top 102 molecules to 96. Mdmx complexes were simulated with these ligands, and 50 trajectories were obtained.

To investigate the stability of the ligand binding pose in the receptor pocket, RMSD values were analysed for the ligand of the Mdmx complex. For the known inhibitor WK298, which was subjected to the same process, the standard deviations of RMSD were in the range of 0.317–2.758. The threshold was set to 2.758 to filter relatively unstable trajectories. This filter further decreased the number of molecules to 49, for which the binding free energies were also estimated. Theoretical  $\Delta G_{\text{exp}}$  can be estimated from the experimental  $IC_{50}$  values with the formula  $\Delta G_{\text{exp}} \approx RT \ln(IC_{50})^{92,95}$ . For WK298 with  $IC_{50}$  value of  $19.7 \mu\text{M}^{50}$ , its  $\Delta G_{\text{exp}}$  is around  $-6.5$  kcal/mol. Below, five promising hits (Table 5-2) with  $\Delta G_{\text{exp}} < -6.5$  were analysed, which could thus be expected to have a higher activity than WK298.

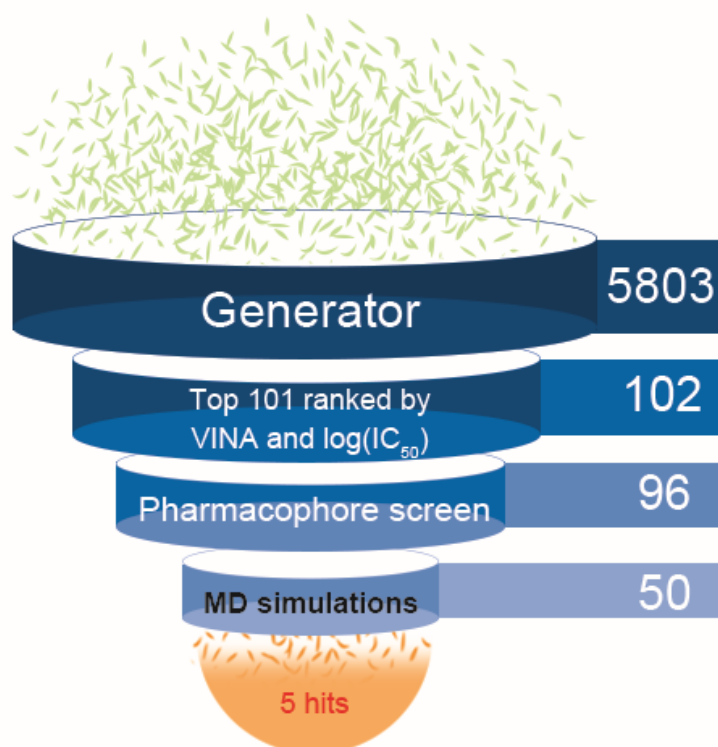


Figure 5-6 The scheme of the virtual screening workflow

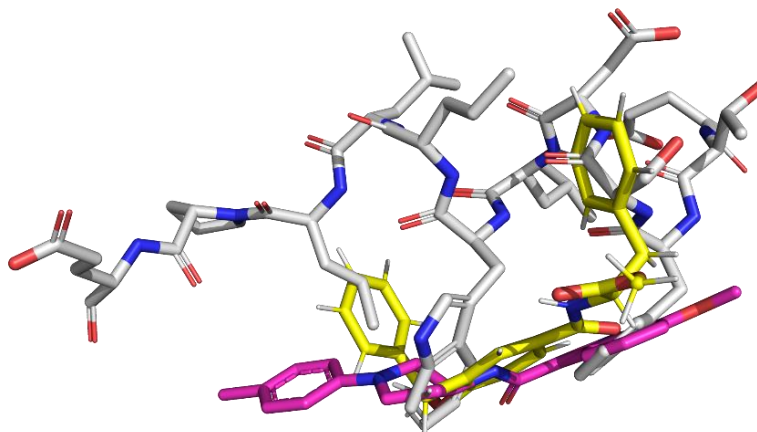
Table 5-2 RMSD and the binding free energy ( $\Delta G$ ) of the representative compounds and WK298 ascending by  $\Delta G$  (a. avg = the average; b. std = standard deviation; c. min = the minimum; d. max = the maximum)

Compound	RMSD				pIC <sub>50</sub>	$\Delta G$ (kcal/mol)
	avg <sup>a</sup>	std <sup>b</sup>	min <sup>c</sup>	max <sup>d</sup>		
WK298	1.5	0.3	0.52	2.9	4.7	- 9.4
3021	4.7	0.4	0.63	5.9	5.2	- 13.0
92	1.6	0.5	0.51	3.7	7.7	- 10.8
100	1.7	0.7	0.48	5.8	7.9	- 6.9
34	2.8	0.7	0.67	5.2	7.9	- 6.7
39	4	1	0.78	8.0	7.6	- 6.7

3,3-Diphenylpropanamide is a common substructure among the generated molecules, such as compound 100. 4-chlorophenyl is located in the Trp23 pocket. One of the two phenyl rings occupies the space of Lys24 of p53, and the other is exposed to the solvent. In the case of compound 34, one phenyl ring is substituted by its bioisosteres such as pyrrole and methyl ester. The latter appears more frequently to become another common fragment. Here, 4-fluorophenyl mimics Trp23 of p53 and 4-chlorophenyl mimics Phe19. In addition, 1H-pyrrole in the opposite direction was aligned with Lys24.



Outside the range of filtered molecules, compound 39 with  $\Delta G$  of  $-6.7$  kcal/mol also has an ester-substituted two-phenyl “umbrella”. Its long straight chain lies across the cleft of the protein, and the hydrophobic pocket shrinks significantly to fit around it. In the crystal structure of the Mdmx complex with the Mdmx selective inhibitor (PDB ID: 6q9y; compound 16,  $IC_{50} = 3.7$   $\mu M$ ), a similar binding conformation was observed<sup>96</sup>. Apart from filling the Trp23 and Leu26 pockets, the remaining phenyl group nicely overlapped with Ser20 of p53 (Figure 5-7).

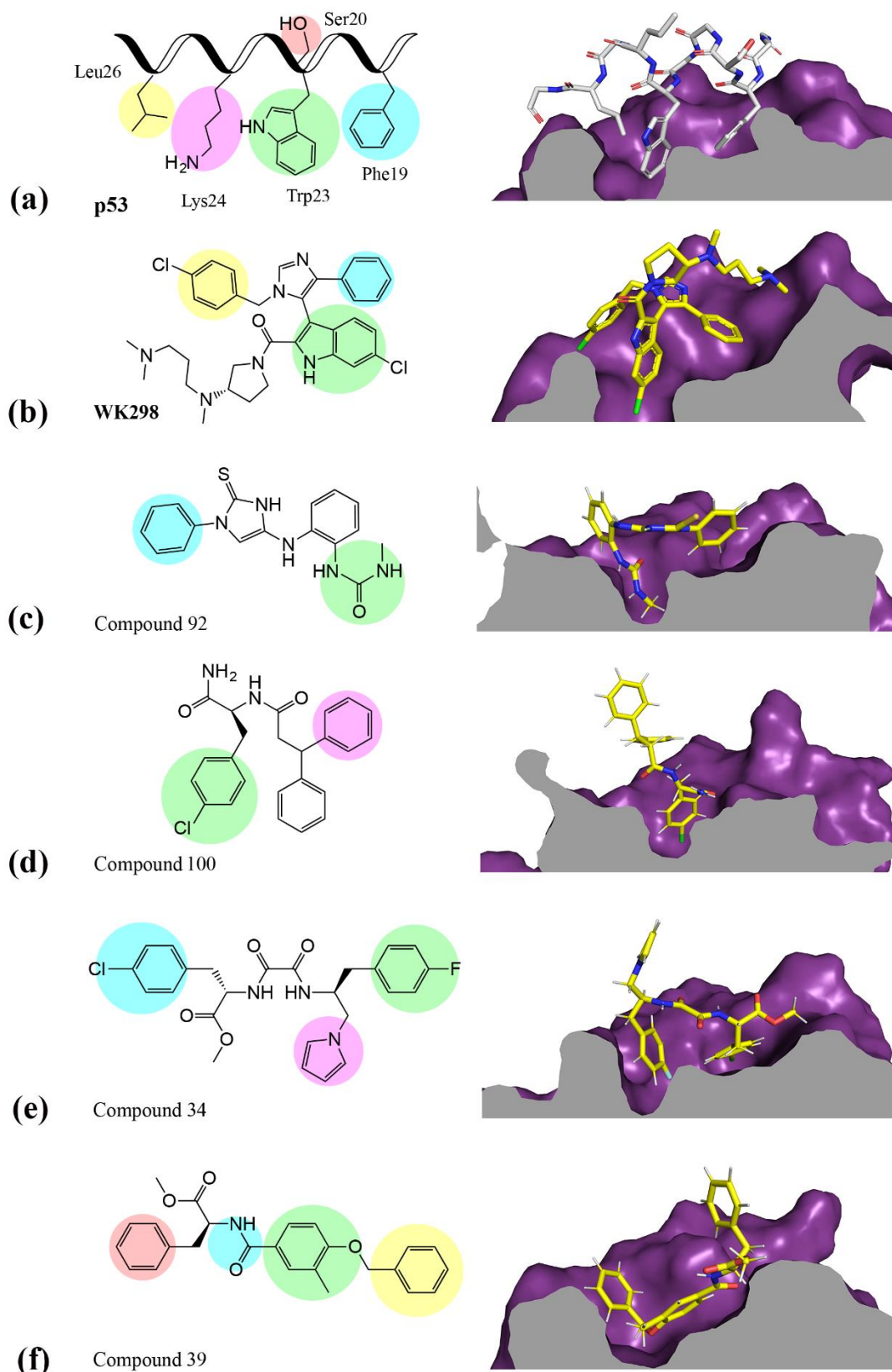


**Figure 5-7 Compound 39 and Mdmx complex aligned to known inhibitor in complex with Mdmx (PDB ID: 6q9y). (white: p53, magenta: 6q9y, yellow: compound 39)**

Compound 92 has a unique scaffold, 3,5-disubstituted-2-thioxo-2,3-dihydro-1H-imidazole. The Mdmx complex was aligned with the crystal structure of the p53-Mdmx complex (PDB ID: 3dab). As shown in Figure 5-8c, the urea group points towards the indole ring of the p53 residue Trp23, and both of its imino groups have H-bond interactions with Met32. 3-phenyl was fairly close to the Phe19 pocket. The three aromatic rings lie on the top of helix  $\alpha_2$  of Mdmx, and the imino group has H-bond interactions with His33. Moreover, Tyr99 of Mdmx kept the “closed” conformation, which was specially studied in this work<sup>50</sup>.

Comparing the binding patterns of these complexes, the extremities of the molecule can be well aligned to Phe19, Trp23 and Leu26 of p53. Ligand fragments are frequently inserted into the Trp23 hydrophobic pocket. Simultaneously, the pocket shrinks to accommodate the size of the small molecules. Interestingly, almost every compound analysed could mimic the three key interacting residues. Some of them can better align with p53 in more residues beyond these three residues. In particular, the surface over helix  $\alpha_2$  is a hot spot for fragments that occupy the Ser20 of p53. In most cases, Tyr99 remains in a “closed conformation”, pointing to the ligand. Therefore, this assume that it tends to interact with the ligand and flips only with the movement of the ligand. Selecting some of them that are typical and yet very different from known Mdmx inhibitors and testing by bioassay in the future would be a worthwhile endeavour. It is noted that some compounds with good  $\Delta G$  values are not similarly ordered at the top of the results from the Generator. This is because of the different computational principles involved. The scoring function integrates both empirical information and experimental affinity measurements. Meanwhile, the relative  $\Delta G$  in this work was calculated based on molecular mechanics and Poisson-Boltzmann

surface area (MM/PBSA) in the absence of entropic contribution. The evaluation criteria of the Generator also equivalently take the predicted  $IC_{50}$  values into account.



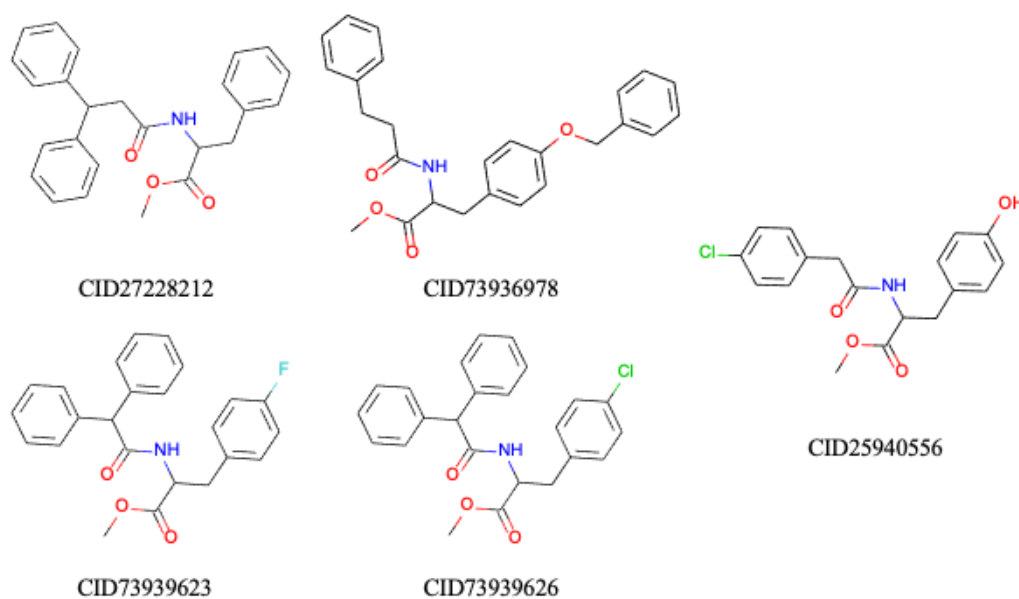
**Figure 5-8** (a) p53-Mdmx complex (PDB ID: 3dab). White colour indicates p53 and the surface of Mdmx is shown in purple colour; (b) WK298-Mdmx complex (PDB ID: 3lbj). Yellow colour



indicates WK298; (c-f) The binding patterns of respective compounds (shown in yellow) in complex with Mdmx are shown. Compounds 3021 and 92 had similar binding modes and thus only complex with compound 92 is shown.

### 5.3.3 Comprehensive analysis on ChEMBL compounds re-discovered by Generator

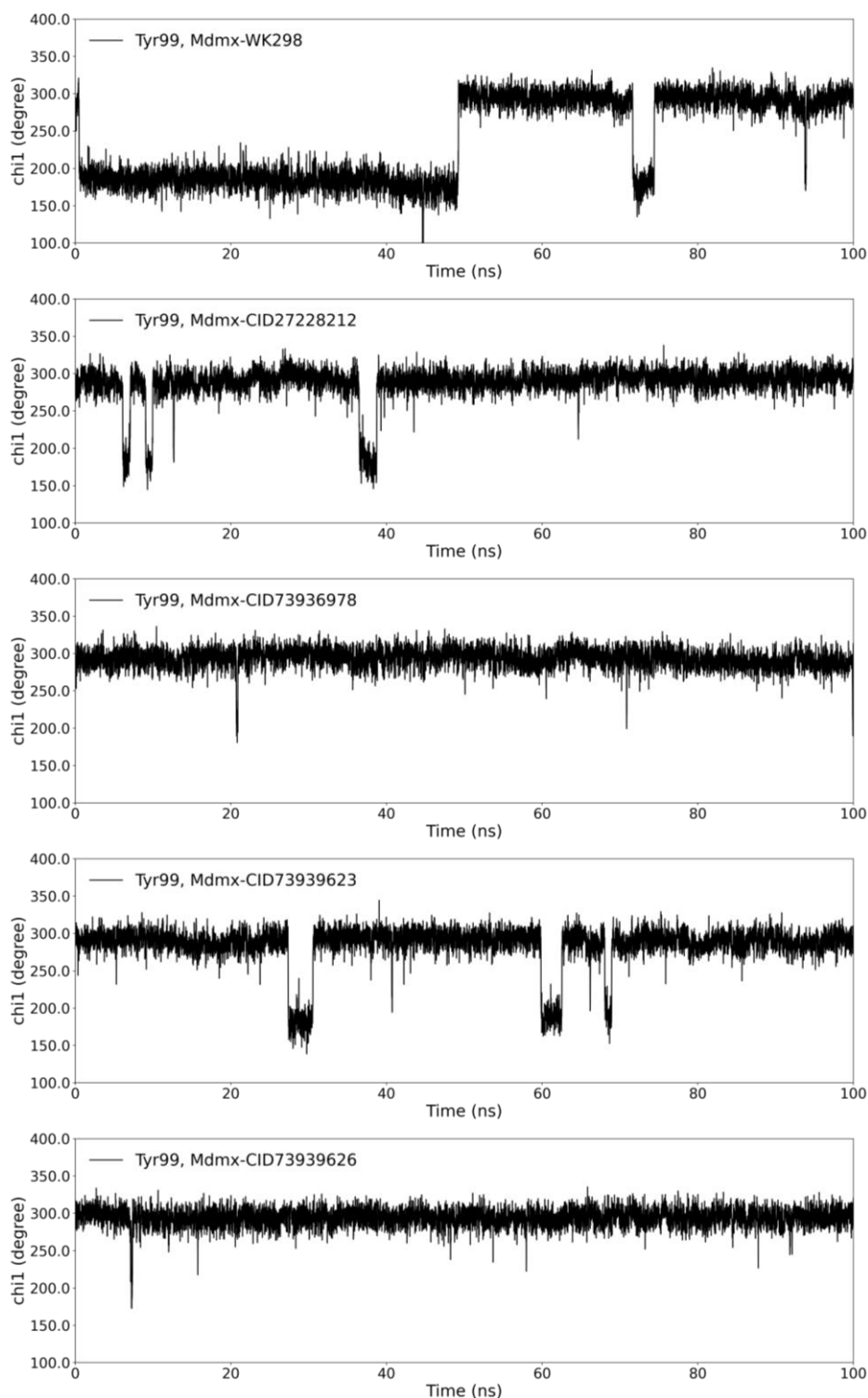
The molecules generated by the Generator included 49 compounds deposited in the ChEMBL database<sup>97</sup> (named the CID series, see Appendix III). The QSAR model of Mdmx inhibitors was used to rank the generated compounds. Within the applicability domain, the predicted pIC<sub>50</sub> (-lgM) of CID compounds were in the range of 6.41–7.55. Five CID compounds with the top predicted pIC<sub>50</sub> values (Figure 5-9) were selected to run MD simulations and estimate the binding free energy of the last 30-ns trajectory refined by the entropy approximation. Remarkably, five CID compounds had the same core structure, that is, methyl N-acetyl-d-phenylalaninate. Notably, p53 interacted with the Met53, Pro95, Leu98 and Tyr99 of Mdmx in the Mdmx-p53 complex according to the pairwise decomposition of binding free energy. Therefore, the energy decomposition was focused on these residues (Table 5-3).

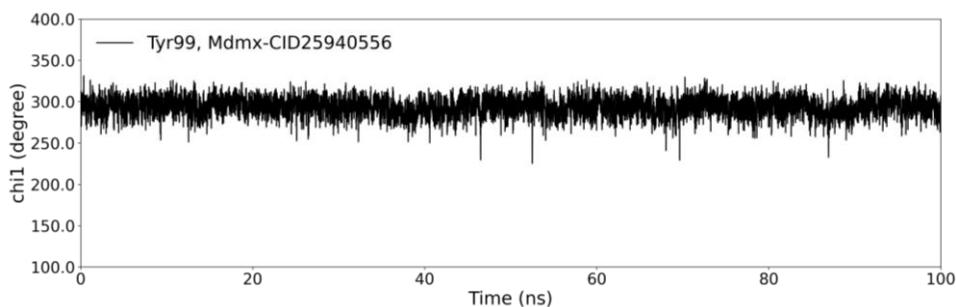


**Figure 5-9** Five CID compounds with high predicted pIC<sub>50</sub>

As shown in Figure 5-10, the binding of these five CID compounds maintained the ‘closed’ state of Mdmx for nearly the whole trajectory; the binding of WK298 had half ‘open’ and half ‘closed’ conformations, retaining the flexibility of Tyr99 and finally being stable in the ‘closed’ state. Even though all the CID molecules shown in Table 5-3 had predicted pIC<sub>50</sub> values greater than WK298, the binding free energy helped us further detect the potential of these compounds. However, CID27228212, CID73936978 and CID25940556 had positive binding free energies, indicating that they had little binding affinity to Mdmx. By comparison, CID73939626 not only

has good predicted  $pIC_{50}$  but also comparable binding free energy to WK298. Furthermore, the interactions between Met53 and Pro95-Ser96-Pro97-Leu98 resulted in the helices  $\alpha_2$  and  $\alpha_2'$  approaching each other and even losing the secondary structure. For example, Pro95-Ser96 lost its helical conformation in Mdmx complexed with CID73936978. Nevertheless, the binding free energy between Met53 and Tyr99 was positive for all observed ligands. This indicated that the Met-aromatic interaction could barely be determined in this way. In the future, it would be interesting to try quantum calculations or experimental methods (e.g. photoelectron spectra) to quantify this relatively weak interaction.





**Figure 5-10** The dynamic change of Tyr99  $\chi_1$  in Mdmx complexes with CID compounds through the 100-ns trajectories. All selected CID compounds stayed in the closed conformation for most of the simulation time.

**Table 5-3** The estimation of binding free energy ( $\Delta G$ ) and its pairwise decomposition for putative ligands (WK298 is the positive reference) as well as the predicted  $pIC_{50}$

Ligand ID	$\Delta G$ (kJ/mol)	M53 - P95	M53 - S96	M53 - P97	M53 - L98	Lig - M53	Lig - Y99	Lig - P95	Lig - S96	Lig - P97	Lig - L98	$pIC_{50}$ (pred)
WK298	-7.7	-	-	√	-	-5.7	√	-1.1	√	√	-1.8	4.7
CID2722 8212	7.0	-	√	-	-	-3.9	√	-1.6	√	√	-1.9	7.5
CID7393 6978	2.1	√	-	-	-	-4.4	-1.77	-1.3	√	√	-1.6	7.5
CID7393 9623	-2.0	-	-	√	-	-4.6	√	-1.2	√	√	√	7.4
<b>CID7393 9626</b>	<b>-6.0</b>	-	-	-	-	<b>-4.9</b>	√	<b>-1.3</b>	√	√	<b>-1.9</b>	<b>7.6</b>
CID2594 0556	3.8	-	-	√	-	-2.5	√	√	√	√	√	7.5

\* √ : mean of total energy decomposition is in the range of (-1, 0) kJ/mol; - : mean of total energy decomposition is greater than 0 kJ/mol.

Based on the data analysis above, it is interesting to investigate the binding mode between the Mdmx and CID73939626 that was estimated to have the binding free energy comparable to WK298. The initial and final conformation were obtained from the 100-ns trajectory of Mdmx-WK298 complex, as shown in Figure 5-11B and C. By comparison, the number of interactions between WK298 and Mdmx decreased by the end of the simulation. The distance between Val92 and WK298 increased from 3.4 to 4.0 Å and thus, the interaction weakened. The final conformation of CID73939626 in complex with Mdmx was shown in Figure 5-11D. This putative ligand formed more hydrophobic interactions with Mdmx compared to WK298. Apart from the hydrophobic interaction involving Val92, the ligand maintained such interactions presented in the Mdmx-WK298 as Ile60, Tyr66, and Phe90. Moreover, Met53 and Tyr99 also participated in the interaction network. Interestingly, there was a halogen bond exclusively, indicating the halogen atoms should be beneficial to the inhibition against Mdmx. The missing of hydrogen bonds is enlightening, so that to add substructures acting as hydrogen-bond donors/acceptors will be a promising direction for optimizing the structure of CID73939626 in the future work.

**(A)**

	Hydrophobic interactions	Hydrogen bonds	Halogen bonds
<b>WK298</b> (t = 0 ns)	Leu56, Ile60, Tyr66, Val92	His54, Gly57	
<b>WK298</b> (t = 100 ns)	Phe90, Val92	Met53	
<b>CID73939626</b> (t = 100 ns)	Met53, Ile60, Tyr66, Val74, Phe90, Val92, Leu98, Tyr99		Gln71

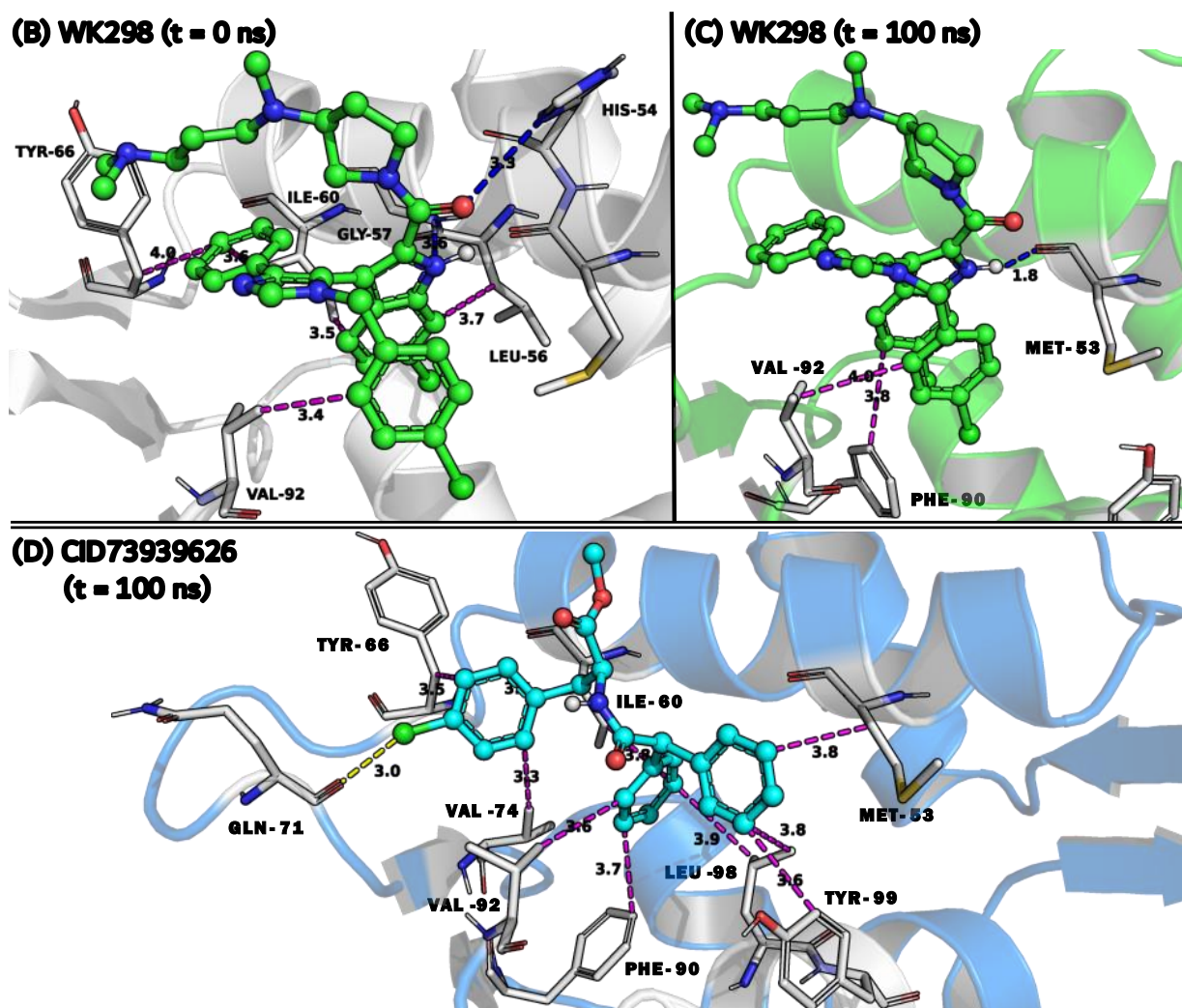
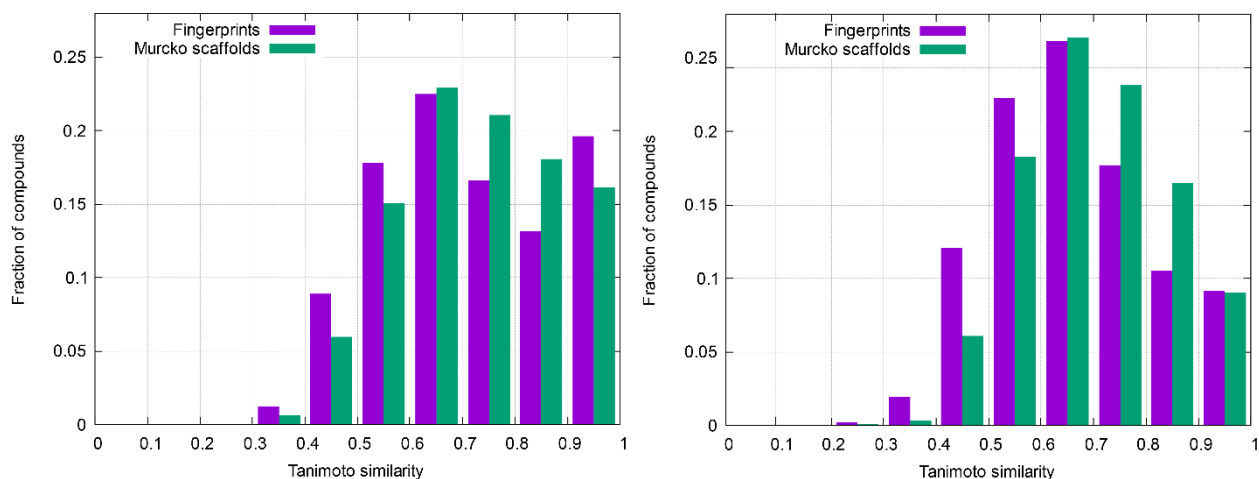


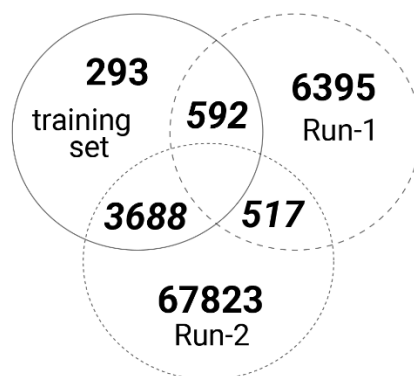
Figure 5-11 The binding mode between Mdmx and small-molecule ligands was analysed using MD simulation. (A) The summary of interactions between the ligands and Mdmx. (B) The crystal structure of Mdmx-WK298 complex, i.e., the snapshot at simulation time  $t = 0$  ns. (C) The snapshot of Mdmx-WK298 complex at simulation time  $t = 100$  ns. (D) The snapshot of Mdmx-CID73939626 complex at simulation time  $t = 100$  ns. The ligands and pocket residues are displayed as *sticks* and the protein is displayed as *cartoon*. The hydrophobic interactions, hydrogen bonds and halogen bonds are displayed in *magenta*, *blue* and *yellow*, respectively.

### 5.3.4 Constraints during tuning

The current implementation of the tuning workflow for the Generator allows the introduction of several constraints while generating new molecules. In this study, the IC<sub>50</sub> model was used to estimate the inhibitory potential of a compound and VINA scores to measure its binding affinity. Since solubility also druggability, a solubility model was added as another constraint on structure generation. The solubility model was built using the Transformer-CNN approach and resulted in  $r^2=0.92$ , RMSE=0.58, and MAE=0.41 (n=1311), which is available at <https://ochem.eu/model/784>, and another cycle of generation with this additional solubility filter was started. The distributions of the Tanimoto similarity of the generated molecules for all compounds and Murcko scaffolds<sup>98</sup> are depicted in Figure 5-12. It is clear that adding another filter shifts the Generator to produce a more diverse set of compounds, as well as chemical scaffolds. Utilising more QSAR filters at this stage should benefit the entire *de novo* design pipeline and further investigate the mutual influences of different filters on the overall scoring procedure. As shown in Figure 5-13, the overlap between the generated compounds during two different runs was less than 10% (517/6395). Note that the total number of generated compounds also includes those structures that overlap with the training dataset. For example, for run-1, the Generator produced a total of 6395 structures, but 592 of them were already in the training dataset (the Generator can produce the same molecule more than once). Thus, different filters guide the exploration of different regions of the chemical space and can be used complementarily to each other. However, the question of how to account for all of these filters in one workflow remains open.



**Figure 5-12 Comparison of similarity distributions for run-1 of generation (without solubility, left) and run-2 (with additional solubility parameter, right)**



**Figure 5-13** The number of compounds generated during run-1 and run-2 as compared to the original dataset

## 5.4 Summary and outlook

In this chapter, generative neural networks were applied to the design of putative Mdmx inhibitors *in silico*. The Generator model was developed to efficiently and automatically generate new molecules from scratch with the desired properties. The initial dataset comprised more than 1.7 million compounds curated from the ChEMBL library without pre-processing, making the model output molecules covering the entire diversity of chemical space. The training of the model on this large set resulted in nearly 79% of the entire output being valid unique molecules. The molecular properties of the generated molecules correlated well with those of the ChEMBL set. The second training set ( $n = 293$ ) consisted of all available Mdmx inhibitors collected from the literature. This set was augmented and used to retrain the previous model. This transfer learning process conducted the model to specifically generate molecules with potential inhibition against Mdmx, which was refined by docking scores. The other constraints further tuned the model to output molecules with good drug-likeness and solubility.

The validation of the focused library generated by the Generator was carried out through molecular dynamics simulations. The Generator generated potential Mdmx inhibitors with completely new structures but also re-discovered structures available in public databases, thus circumventing the difficulty of their synthesis. The promising molecules have better predicted  $pIC_{50}$  and similar or higher binding affinities compared to the reference inhibitor WK298. The compounds re-discovered from the ChEMBL database (CID series) were additionally investigated by 100-ns simulations. The MD analysis demonstrated that these compounds made Mdmx keep the ‘closed’ conformation for nearly entire simulated time. By contrast, WK298 had half ‘open’ and half ‘closed’ conformations. This suggests that the ‘closed’ state of Mdmx should not be neglected as a promising direction for developing new Mdmx inhibitors.

This work was published,<sup>99</sup> and the source code made available to the public (<https://github.com/bigchem/online-chem>).

The hydrophobic effect and desolvation penalties critically influence the drug activity. Adding hydrophobic groups can improve the binding affinity of a ligand with a hydrophobic pocket by decreasing the solvent-accessible surface. If the solubility of a ligand is increased, the binding

affinity may suffer a great loss because of the cost of desolvation. Moreover, adding modification groups has the risk of introducing higher flexibility, making the entropic cost of the binding process more significant. Future Generators should be penalised for the flexibility and desolvation of buried polar groups. Although the number and quality of filters applied during the generation phase were to be optimised, the whole algorithmic scheme would guide future drug development processes, yield new drug candidates for challenging targets, and contribute to the diversification of medicinal chemistry methodologies.

# 6 Development of a general model to characterize ligands with respect to opening transient pockets in drug targets

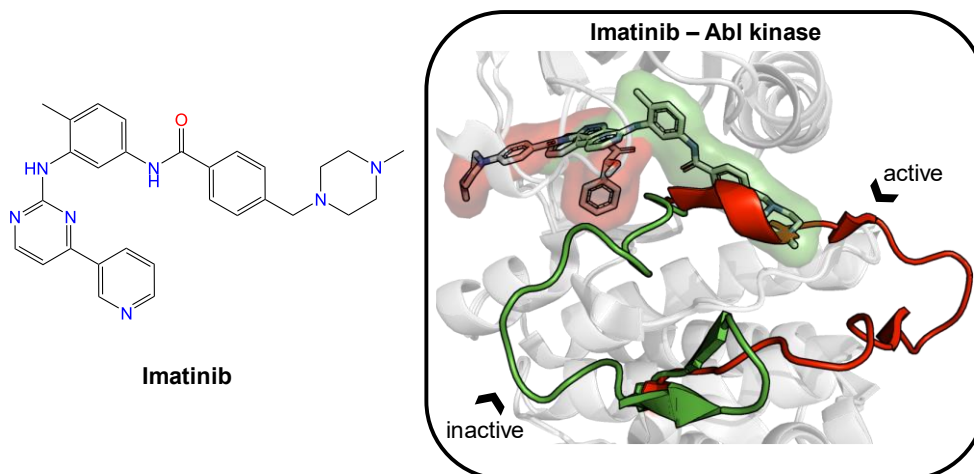
## 6.1 Motivation

Inspired by 1) the open and closed state of Tyr100 (Tyr99 in Mdmx), 2) Mdm2-YH300 complex where the 4-chlorobenzyl phenyl ether induced the opening of the fourth binding site in Mdm2, many researchers searched for the extension of the main pocket in Mdmx and designed new Mdmx inhibitors accordingly. However, it is not clear which features of ligands could contribute to the opening of cryptic pockets of proteins. The development of a general model that can predict the ability of ligands to open cryptic pockets could be useful for designing such ligands for Mdmx, as well as for other proteins.

The idea of this challenging project was to utilise the local structural flexibility of these interfaces to enable structure-based drug design. Imatinib (Figure 6-1) is an antineoplastic agent, which inhibits the Abelson (Abl) tyrosine kinase by capturing a transient state.<sup>100-103</sup> It specifically binds with the preformed ATP pocket and an adjacent allosteric pocket which likely appears induced by the ligand itself. Both experiments and molecular dynamics simulations of proteins revealed the existence of transient pockets, also known as cryptic sites<sup>104</sup>. Cryptic binding sites are absent or occluded in unbound proteins, but are present in ligand-bound structures. They can provide druggable pockets in cases where the main functional site of the protein cannot be targeted with sufficient potency or specificity. Mizukoshi et al.<sup>105</sup> developed a strategy to identify cryptic sites and manipulate their dynamic changes through nuclear magnetic resonance (NMR) combined with allosteric mutations. Kii et al.<sup>106</sup> identified an inhibitor (FINDY) targeting transient pockets in dual-specificity tyrosine-phosphorylation-regulated kinase 1A (DYRK1A). Cimermancic et al.<sup>107</sup> created the CryptoSite dataset of apo-holo pairs with cryptic sites and used machine learning to predict cryptic sites with relatively high accuracy (AUC = 0.83). They found that cryptic sites were less hydrophobic and more flexible than traditional binding pockets. Beglov et al.<sup>108</sup> explored the extended CryptoSite dataset by detecting hot spots using FTMap<sup>109</sup>. Their results showed that regions around cryptic sites had above-average flexibility. Moreover, they observed that cryptic sites, which were formed solely by side chain motion or less than 5 backbone residues' motion, showed weak ligand binding potential with limited use for drug discovery. Clark et al. investigated the structural flexibility of proteins by root mean square deviation (RMSD) calculations of backbone atoms and  $\chi_1$  angles of side-chain atoms.<sup>110</sup> The ligand-induced backbone flexibility across apo-holo pairs was slightly larger than the inherent backbone flexibility of the apo and holo states. Upon ligand binding, sidechain  $\chi_1$  angles were frequently



pushed to new orientations outside the range seen in the apo states. They found that the influences on binding-site variation could not be easily attributed to features such as ligand size or X-ray structure resolution. More recently, Evans et al.<sup>111</sup> developed a machine learning method (TACTICS) to predict cryptic sites. TACTICS was trained on an extended CryptoSite dataset and used molecular dynamics (MD) simulation data as input. This advanced method successfully detected the active sites and multiple allosteric sites in the SARS-CoV-2 main protease. The increasing use of MD simulations in conjunction with machine learning algorithms will likely contribute greatly to the identification of cryptic pockets and expand the scale of druggable proteomes in the foreseeable future.<sup>112</sup>



**Figure 6-1** The structure of Imatinib is given as an example of the inducers (*left*). Two crystal structures of Abelson (Abl) tyrosine kinase bound with small-molecule ligands were aligned (PDB ID: 1iep, Imatinib, *green*; PDB ID: 2v7a, PHA-739358, *red*). In the binding site, the major structural displacement was present in the rotation of the activation loop (residue 381 to 402 in Abl kinase). Imatinib (*green*) binds to the preformed pocket and the adjacent sub-pocket which is caused by the rotation of the loop.

It was interesting to identify the structural features of ligands that are important for the opening of cryptic sites. This information could be useful for designing new active compounds against currently undruggable proteins or for discovering new binding sites for underexplored proteins. In addition, it was interesting to know the extent to which this comprehensive issue involving both the characteristics of proteins and structure-activity relationships of ligands could be explained by the investigation of small molecules.

In this study, the ligands were classified into two groups: the first group consisted of the ligands that induced the opening of cryptic pockets, called ‘inducers’, and the second group included those that could not open the pockets, referred to as ‘non-inducers’. The classification criterion was based on the volume change of the pocket between the unbound state (apo) and the bound state (holo). The degree to which the enlargement of a pocket was considered as opening was discussed.

## 6.2 Methods

### 6.2.1 Datasets

#### *Dataset I: PDBbind-CN*

PDBbind-CN<sup>113</sup> v.2020 included all types of biomolecular complexes and their experimentally measured binding affinity data deposited in the Protein Data Bank (PDB)<sup>45</sup>. Starting from 5316 crystal structures of holo proteins, a dataset comprising apo (does not contain a bound ligand) and holo (complex with a ligand) protein pairs was established. The protein and ligand structures of the holo complexes were retrieved using their PDB IDs. The requirements were that only one ligand in the same pocket as identified by the ligand ID<sup>114</sup> and that the resolution of the crystal structure was better than 2.5 Å. The initial set of apo structures included PDB proteins with the same resolution requirements. The binding pocket was defined by the residues within 10 Å of the cognate ligand. Since I was only interested in drug-like small molecules, the holo structures whose identified ligands were peptide-like, cofactors, detergents and crystal additives, were excluded, as suggested in previous work<sup>115</sup>. Next, the sequence of each holo protein was analysed using the blastp program<sup>116</sup> to generate a list of PDB IDs descending by the similarity between the holo and apo protein sequences. It was determined that no ligands were bound in the same binding pockets of the apo structure, and the protein pair with the highest similarity score was retained for analysis. Both the retrieving and calculating processes were automatically carried out by the PyMOL<sup>38</sup>, blastp and SiteMap<sup>117–119</sup> programs with the help of custom scripts. Moreover, the CryptoSite dataset was combined with the PDBbind dataset as the training set, which has been commonly used in previous studies on cryptic pockets<sup>104,107,108,111</sup>. Originally, this dataset contained 93 holo proteins matched with multiple apo structures individually. Eight items binding to either multiple ligands or oligosaccharides in the pocket were removed. After the removal of duplicates from the CryptoSite dataset compared to the PDBbind dataset, eighty-four pairs of holo-apo proteins as well as their ligands were retrieved. All these ligands were labelled as inducers because the corresponding receptors had cryptic sites, which were manually curated and illustrated by the creators.

#### *Dataset II: NR-DBIND*

Nuclear receptors (NRs) are important drug targets and their ligand binding is generally accompanied by large conformational changes. The same workflow as the PDBbind dataset was implemented to build a dataset comprising apo and holo NR pairs. To this end, the nuclear receptor database (NR-DBIND)<sup>120</sup> was used as the starting point. These data indicated whether the PDB ID corresponded to a holo or apo structure, which helped in validating the developed in-house searching script.

Below, tools and procedures used to analyse the retrieved structures and create datasets were described.

### *Structure preparation and visualisation*

PyMOL<sup>38</sup> was used to retrieve structures from the PDB database to define the pocket and align the holo-apo pairs. It was also used to visually validate the matching results from in-house scripts.

The quality of starting structures is critical for computational drug design. Therefore, all protein structures were processed using the Schrodinger Protein Preparation Wizard<sup>121</sup>. This script is a tool for correcting common structural problems with experimentally derived structures to create reliable protein models. This tool was used 1) to assign bond orders in known residues and small molecule components based on their SMILES strings in the Chemical Component Dictionary<sup>114</sup>, 2) to add explicit hydrogens, 3) to create zero-order bonds to metals, and 4) to generate a favourable ligand protonation state based on the number of H-bonds and the Epik penalty score<sup>122</sup>.

### *Identification and characterisation of binding pockets*

The SiteMap<sup>117-119</sup> of Schrodinger suite was used to identify and calculate the pocket volumes of the pre-processed proteins. This program first detected all sites suitable for docking in the protein structure and then individually calculated the corresponding pocket volume. In this way, I obtained the volumes of the identical pocket in the holo and apo structures. Technically it was done using trajectory\_binding\_site\_volumes.py with -sitmap\_site\_asl option by specifying the residues constituting the pocket of interest. SiteMap determined the pockets suitable for docking using geometry and energy-based algorithms. As a result of this restriction, around one thousand input structures failed in this step, and thus were not further considered in this work.

The structures which were present in both databases were kept in the PDBbind set for model training and removed from NR-DBIND, which comprised the external validation set. Thus, neither datasets had overlapping structures. In total, 992 and 135 apo-holo pairs were obtained in PDBbind and NR-DBIND, respectively. Among them, 675 and 117 pairs increased  $V_{\text{holo}} - V_{\text{apo}} > 0 \text{ \AA}^3$  the pocket volume. Small changes in the volume could also be due to fluctuations in the volume or restricted conformational adjustments, and thus some threshold could be beneficial to use. A detailed procedure for defining the threshold to separate inducers and non-inducers is described in the Results section.

## **6.2.2 Development of models to separate inducers and non-inducers**

The OCHEM platform<sup>123</sup> was used to develop classification models for inducers and non-inducers. The models were developed with data from the PDBbind set, and the NR-DBIND set was used as an external validation. In practice, different machine-learning methods and descriptors available in OCHEM were explored. Random Forest (RF) and ALogPS and OEstate (AO) descriptors contributed to the models with the highest accuracy and were selected for all analyses. Below, they were briefly described.

Random forest (RF)<sup>124</sup> is an ensemble of several decision trees that are used to classify a new sample by a majority vote of individual decision trees. It is an effective tool for predictions and is widely used as a reference in classification and regression tasks. There are two important tuning parameters that significantly affect the performance of RF models: numFeatures and numTrees. For the first parameter, the recommendation of previous work<sup>125</sup> was followed and this parameter was set to the sqrt(p) of the number of descriptors, p, used to build the models. Increasing the number of trees above 512 did not lead to better models, as shown in other previous work<sup>126</sup>. Therefore, the fixed values of these both parameters were used.

The ALogPS and OEstate<sup>127</sup> (AO) descriptors contain the octanol/water partition coefficient (logP), solubility in water (logS),<sup>127-129</sup> and electrotopological state (E-state) indices that combine the electronic properties and topological context of skeleton atoms derived from the molecular hydrogen-depleted graph.<sup>130-133</sup> The E-state index is defined as follows:

$$S_i = I_i + \Delta I_i = I_i + \sum_{j=1}^A \frac{I_i - I_j}{(d_{ij} + 1)^k}$$

where  $S_i$  is the E-state of the  $i$ th atom,  $I_i$  is the intrinsic state of the  $i$ th atom,  $\Delta I_i$  is the field effect on the  $i$ th atom calculated as perturbation of the intrinsic state of the  $i$ th atom by all other atoms in the molecule,  $d_{ij}$  is the topological distance between the  $i$ th and  $j$ th atoms, and  $A$  is the number of atoms. The exponent  $k$  is a parameter that modifies the influence of distant or nearby atoms in particular studies. This is usually taken as  $k = 2$ . The intrinsic state of the  $i$ th atom is calculated as

$$I_i = \frac{\left(\frac{2}{L_i}\right)^2 \cdot \delta_i^v + 1}{\delta_i}$$

where  $L_i$  is the principal quantum number,  $\delta_i^v$  is the number of valence electrons, and  $\delta_i$  is the number of sigma electrons of an atom in the H-depleted molecular graph. The intrinsic state of an atom can be simply considered as the ratio of  $\pi$  and lone-pair electrons over the  $\sigma$  bond count in the molecular graph. Therefore, the E-state index is a measure of the electronic accessibility of an atom and can be interpreted as the probability of interaction with another molecule.

Atom-type E-state indices encode electronic and topological information related to particular atom types in molecules. They are calculated by summing the E-state values of all atoms of the same atom type in the molecule. The symbol of each atom-type E-state index is a composite of three parts. The first part is “S”, which refers to the sum of the E-states of all atoms of the same type. The second part is a string representing the bond types associated with the atom (“s”, “d”, “t”, “a” for single, double, triple, and aromatic bonds, respectively). The third part is the symbol identifying the chemical element and eventual bonded hydrogens, such as CH3, CH2, and F.

The bond E-state index is defined as:

$$BS_b = BI_b + \Delta BI_{bt} = \frac{(I_i + I_j)_b}{2} + \sum_{t \neq b} \frac{BI_b - BI_t}{(\bar{r}_{bt} + 1)^2}$$

where  $BS$  is a bond-state index,  $b$  is the bond formed by the atoms  $i$  and  $j$ ,  $t$  runs over all the remaining bonds different from bond  $b$ ,  $BI$  is the bond intrinsic state defined by the intrinsic states  $I$  of the adjacent vertices,  $\Delta BS$  is the perturbation term,  $\overline{r}_{bt}$  is the average bond length of the bonds  $b$  and  $t$ .

Some non E-state descriptors such as molecular weight, number of hydrogen bond acceptors/donors, rotatable bonds, etc., were also included.

#### *Model validation*

N-fold cross-validation was used to evaluate the robustness of the models and to avoid overfitting of the models. The method was performed as follows: (i) the initial dataset was randomly divided into  $n$  mutually disjoint subsets of the same size; (ii)  $(n-1)$  subsets were used to train the model and the remaining subset was used as a test set. The prediction results for the respective test sets were accumulated and used to evaluate the prediction performance of the final model, which was developed using the initial dataset. In this study, a 5-fold stratified cross-validation was used to account for data imbalance.

#### *Performance measurement*

The predictions of the models were evaluated based on prediction accuracy and balanced accuracy. They are calculated by:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Balanced\ Accuracy = \frac{\frac{TP}{TP + FN} + \frac{TN}{TN + FP}}{2} \times 100\%$$

Here, True Positives (TP) is the number of inducers predicted correctly by the model, True Negatives (TN) is the number of non-inducers predicted correctly, False Positives (FP) is the number of non-inducers predicted as inducers, and False Negatives (FN) is the number of inducers predicted as non-inducers.

Balanced accuracy was the average accuracy for the prediction of inducers and non-inducers. This measure better reflected the accuracy of models when using different thresholds to separate inducers and non-inducers, which contributed to their imbalances. In addition, AUC, the area under the receiver operating characteristic (ROC) curve, was used.

#### *Interpretation of models*

The Shapley value was introduced by Lloyd Shapley<sup>134</sup>, which provides a natural way to compute the features that contribute to a prediction. Shapley Additive exPlanations (SHAP)<sup>135</sup> is a unified framework for the class of additive feature attribution methods including LIME<sup>136</sup>, DeepLIT<sup>137</sup>, layer-wise relevance propagation (LRP)<sup>138</sup>, Shapley regression values<sup>139</sup>, Shapley sampling values<sup>140</sup>, and quantitative input influence<sup>141</sup>. SHAP combines all the characteristics of the six aforementioned methods and three additional desirable properties (local accuracy, missingness, and consistency). Features contributing to the predicted classification are evaluated via the absolute values as well as the plus/minus signs corresponding to the positive/negative impacts. TreeExplainer<sup>142</sup> is specifically implemented for tree-based models to compute SHAP

values for trees and ensembles of trees. Variables enter the machine learning model sequentially or repeatedly in the trees of the model. In every step of tree growth, the algorithms evaluate each of the variables equally to settle with the variable that contributes the most. In total, thousands of trees are constructed. It is possible that various permutations of the variables are available. Therefore, the marginal contribution of each variable is calculated. TreeExplainer computes local explanations based on exact Shapley values in polynomial time and extends local explanations to directly capture feature interactions so that it can interpret a model's global behaviour based on many local explanations.

The RF model to estimate the SHAP values was built using scikit-learn<sup>143</sup>. The same default feature filtering as for the OCHEM model was used using the same parameters for RF as described above to keep both models as similar as possible. The model was explained by TreeExplainer which allowed us to identify important descriptors in terms of structural features and properties.

### 6.2.3 Analysis of similarity of ligands

The RDKit which is an open-source toolkit for chemoinformatics was used<sup>83</sup>. It was used to load and convert the 3D structure files of ligands to 2D molecular graphs facilitating the presentation. It was also used to perform fragment analysis based on Functional-Class Fingerprints (FCFP)<sup>144</sup>. FCFP4 (radius = 2) was used to calculate the similarity of molecules based on the Tanimoto coefficient<sup>145</sup>. Tanimoto ( $S_{AB}$ ) is one of the most popular methods for comparing fingerprints, and is also known as the Jaccard coefficient. It is defined as:

$$S_{AB} = \frac{c}{a + b - c}$$

where  $a$  and  $b$  are the number of bits set in FCFP4 of  $A$  and  $B$ , respectively, and  $c$  is the number of bits shared in both.

## 6.3 Results

### 6.3.1 Model performance

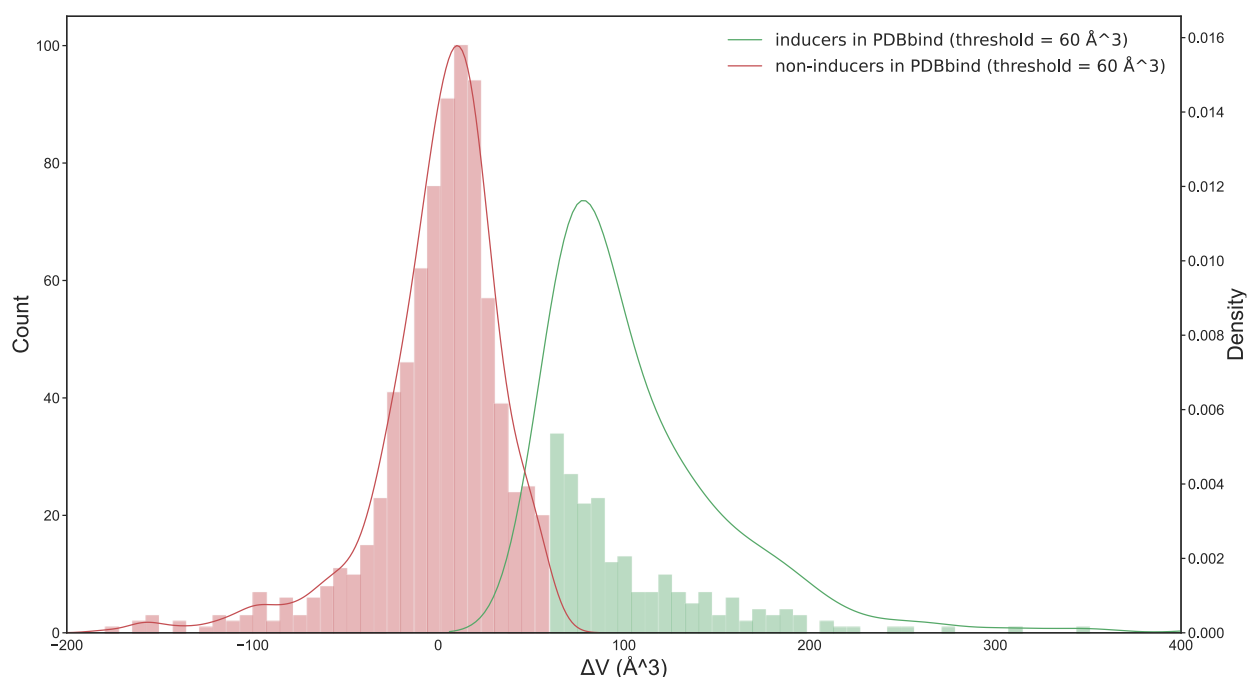
The PDBbind and NR-DBIND databases were processed using the protocol described in the Data section. The number of holo structures or apo structures considered for the same protein were not limited. Thus, the data covered the same protein binding with different ligands. Both agonists and antagonists were used in the analysis. In some cases, a protein had more than one apo structure resolved, providing diverse conformations during the movement of this protein in the ligand-free state. The 474 holo structures in the PDBbind set and 53 holo structures in the NR-DBIND set had only one matched apo structure. The remaining holo structures had multiple apo structures deposited in the PDB database, which were tentatively retrieved (at most ten apo structures per holo structure) to exactly determine the classification labels of the ligands. Among all apo structures that were considered as produced by the intrinsic flexibility of proteins, the one with the maximum pocket was used as a reference for comparison with respective holo structures.

To some extent, this analysis eliminated the cryptic pocket formed by the conformational selection.

Before modelling, both datasets were manually checked to ensure that the apo structure correctly matched the corresponding holo structure; the blastp identities for holo-apo pairs were greater than 85% for PDBbind and NR-DBIND sets. The OCHEM was used to build classification models to separate inducers and non-inducers.

For each protein, a pair with the largest differences in the volumes of apo and holo pockets was identified. The number of pairs that had  $V_{holo} - V_{apo} > 0 \text{ \AA}^3$  was more than twice the number of pairs that had a negative change in volume (Figure 6-2). However, a small change in pocket size should not be considered as a pocket opening. This could be simply pocket adaptation or pocket fluctuation. The minimum volume of the holo pocket identified by SiteMap was  $41 \text{ \AA}^3$  and the minimum ligand (pyruvic acid) was  $67 \text{ \AA}^3$  in size. An optimum threshold to separate inducers and non-inducers would be of similar size.

To determine an optimal threshold, different  $\Delta V$  values from 20 to  $100 \text{ \AA}^3$  at intervals of  $10 \text{ \AA}^3$  was investigated. The assumption was that if inducers and non-inducers have different features, then an optimal threshold should correspond to a model with the highest performance.



**Figure 6-2 The distribution of two classes in the PDBbind set ( $\Delta V = 0 \text{ \AA}^3$ )**

After analysing the performance of machine learning methods and descriptors available in OCHEM, random forest (RF)<sup>124</sup> combined with ALogPS<sup>129</sup> and OEstate<sup>127</sup> (AO) descriptors on average provided models with higher performances compared to other analysed methods and descriptors. The highest average AUC value for both sets was obtained with a threshold of  $60 \text{ \AA}^3$  (Table 6-1). For this threshold 132 and 93 ligands were classified as inducers for the PDBbind and NR-DBIND sets, respectively. The binding of NR ligands usually triggered large movements of NR, and thus, most of the ligands binding these proteins were likely to be inducers. Therefore, the change in the threshold practically did not change the number of inducers and non-inducers for this set. For the training set, the selection of the threshold strongly affected the ratios of inducers and non-inducers.

There are 95 types of proteins in the training set. The model was trained with the diverse proteins and was tested on the external validation set exclusively containing nuclear receptors. The model had a lower AUC accuracy for the test set as compared to that for the training one. Still the difference in the performances was not large (AUC 0.71 vs AUC 0.65).

**Table 6-1 Accuracy of classification models when using different thresholds**

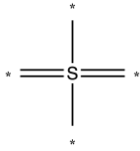
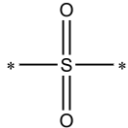
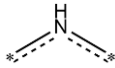
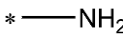

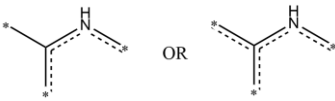
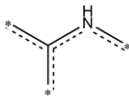
Threshold ( $\Delta V$ , Å <sup>3</sup> )	No. of inducers	No. of non-inducers	Training set			External validation set		
			AUC	Balanced Accuracy	Accuracy	AUC	Balanced Accuracy	Accuracy
0	759 (117)*	317 (18)*	0.61	68	72	0.56	56	81
20	498 (112)	578 (23)	0.67	69	69	0.51	51	76
30	405 (110)	671 (25)	0.69	73	75	0.59	59	78
40	356 (105)	720 (30)	0.68	73	76	0.54	54	67
50	319 (103)	757 (32)	0.7	74	78	0.58	58	75
<b>60</b>	<b>293 (100)</b>	<b>783 (35)</b>	<b>0.71</b>	<b>76</b>	<b>80</b>	<b>0.65</b>	<b>65</b>	<b>80</b>
70	251 (97)	825 (38)	0.73	77	81	0.60	60	73
80	216 (93)	860 (42)	0.68	75	81	0.67	67	79
90	185 (90)	891 (45)	0.64	73	83	0.57	57	64
100	170 (86)	906 (49)	0.68	77	86	0.60	60	64

\*The number of inducers/non-inducers was determined by the threshold used. Numbers in brackets correspond to the validation set.

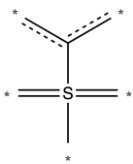
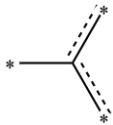
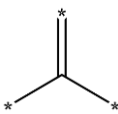
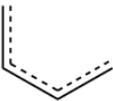
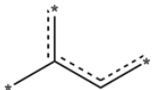
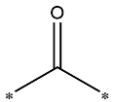
## 6.3.2 Interpretation of models

Table 6-2 contains the top-20 important descriptors identified by the SHAP algorithm.

**Table 6-2 The most influential descriptors to classify ligands as inducers.<sup>a</sup>**

Increases Probability of Being Inducers		Decreases Probability of Being Inducers	
Descriptor	Description	Descriptor	Description
SddssS		SdO (sulpho)	
ALogPS_logP	octanol/water partition coefficient	AlogPS_logS	solubility in water
HALOG	number of halogen atoms	SaaNH	
aCNOS	the sum of aromatic C, N, O, and S atoms	SsNH2	
SaaCH		SeaC3NHaa	 OR 



Se1C3S4ad		S	number of sulphur atoms
SaasC		SdssC	
SeaC2C2aa			
RBONDS	number of rotatable bonds		
P	number of phosphorus atoms		
SeaC2C3aa			
MW	Molecular weight		
SdO (acid)			

<sup>a</sup>see SHAP values calculated for the filtered descriptor in Figure S1

The top-20 AO descriptors were plotted in descending order of their respective importance in the model (see Appendix IV). There are several features related to heteroatoms. Sulphonyl (SddssS) and Se1C3S4ad had a positive impact, while the sulphones/sulphoxides [SdO (sulpho)] and sulphur atoms (S) were not beneficial as inducers. These results indicate that the influence of sulphur on the target property might be ambiguous and requires a deeper analysis to determine which fragment it participates in. Moreover, three nitrogen-containing substructures showed negative impacts, such as aromatic (SaaNH, SeaC3NHaa) and primary (SsNH2) amino groups. In addition, the molecular weight (MW), halogen (HALOG), phosphorus (P), and the double-bonded oxygen of carboxylic acids [SdO (acid)] also contribute to the probability that the ligand is an inducer. Interestingly, RBONDS descriptors suggested that sufficient flexibility is necessary for an inducer to adapt itself into a transient pocket, which is consistent with the conclusions of Cimerancic et al.<sup>107</sup> and Beglov et al.<sup>108</sup>

With regard to carbon atoms, descriptors such as SaaCH, SaasC, SeaC2C2aa, SeaC2C3aa, and number of aromatic atoms (aCNOS), were derived from aromatic atoms, indicating the importance of aromatic interactions between protein and ligand to the inducer properties of ligands.

The lipophilicity of molecules (ALogPS\_logP) increased the ability of molecules to act as inducers, while their solubility (ALogPS\_logS) had the opposite effect. The more hydrophobic a molecule was, the more likely it was to be an inducer.

### 6.3.3 Analysis of functional groups

The OCHEM SetCompare tool<sup>146–148</sup> uses a hyper-geometric distribution to identify the functional groups that are over-presented in inducers and non-inducers. As shown in Table 6-3, sulphonic acid derivatives appeared more often in non-inducers and sulphonamides accounted for 97% of these derivatives (179 out of 184 ligands). Note that, the primary amine was favourable for non-inducers, as indicated by the descriptor SsNH2 in the previous section. Here, inducers showed a propensity to have aromatic structures substituted by tertiary amines or halogens. In addition, phosphoric and carboxylic acids were over-presented in the inducer cohort. These features were consistent with the explanation of the model by SHAP, but also offered some new clues that were not shown before. The pyridine and five-membered heterocycles with two heteroatoms were two substructures, particularly identified by SetCompare, which could be used as building blocks for inducers.

**Table 6-3 Functional groups over-presented in inducers and non-inducers are listed as well as the p-value of the respective distribution.**

Functional group	The ratio in Inducers (%)	The ratio in Non-inducers (%)	<i>p</i> -value
Sulphonamides	6.3	23.4	$-4.8 \times 10^{-12}$
Sulphonic acid derivatives	7.3	24.1	$-4.03 \times 10^{-11}$
Halogens	44.1	28.7	$1.93 \times 10^{-6}$
Halogenated benzene	28.1	15.1	$1.83 \times 10^{-6}$
Aryl halides	36.5	24.3	$7.79 \times 10^{-5}$
Tertiary mixed amines	11.5	4.5	$6.66 \times 10^{-5}$
Five-membered heterocycles with two heteroatoms	17.4	9.9	$9.45 \times 10^{-4}$
Pyridine	11.8	5.6	$7.52 \times 10^{-4}$
Phosphorus	13.5	6.8	$6.28 \times 10^{-4}$
Phosphoric acids	10.1	4.3	$5.95 \times 10^{-4}$
Carboxylic acids	31.6	21.2	$3.62 \times 10^{-4}$
Aromatic halogen	33.3	22.1	$1.61 \times 10^{-4}$

### 6.3.4 Fragment analysis on PDBbind set

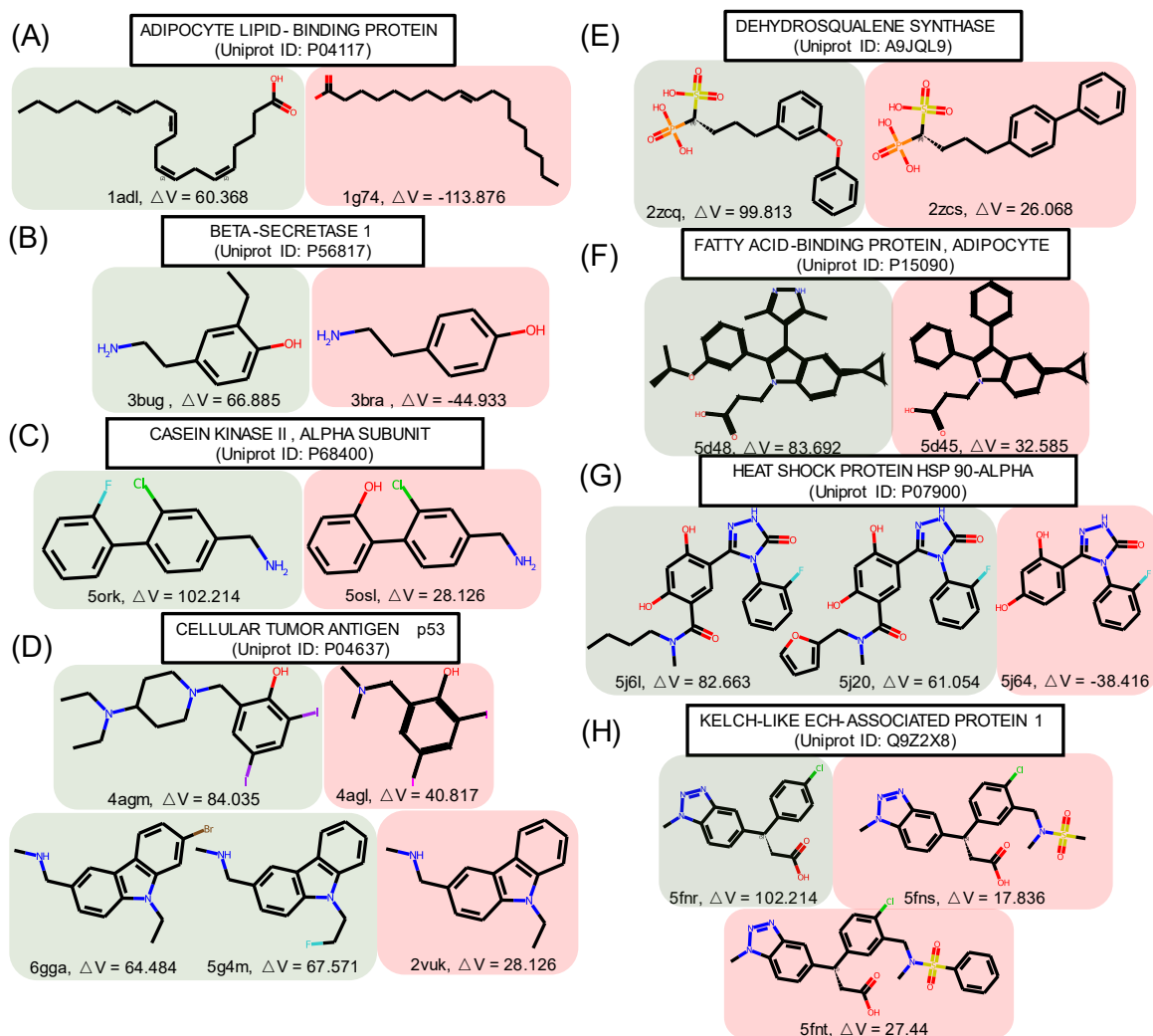
Some ligands targeting the identical protein had similar structures (similarity score  $\geq 0.9$ , calculated on FCFP4) but had different effects on cryptic pocket opening, as shown in Figure 6-3, where the structures were grouped according to the effect type.

In group A, inducers preferred to have a long fatty acid chain with unsaturated bonds. A short aliphatic chain or a halogen atom turned non-inducers into inducers, as shown in groups B, C, and D. In group D, the ligand in the crystal structure with PDB ID of 4agm<sup>149</sup> had additional tertiary amine and closed one of the tertiary amines forming a piperidine as compared to the similar non-inducer structure of PDB ID of 4agl<sup>149</sup>. This emphasises the importance of large hydro-

phobicity and a large number of tertiary amines for inducers. The addition of singly-bonded oxygen helped to obtain inducers in groups E and F.

In group F, replacement of the benzene ring by a pyrazole turned a non-inducer to an inducer, which was consistent with the SetCompare result, indicating that five-membered heterocycles with two heteroatoms were over-presented in the inducers. In group G, the difference between inducers and the non-inducer was the amide substructure substituted by alkyl chains or a furan ring. In group H, the inducer in the crystal structure with PDB ID of 5fnr<sup>150</sup> became a non-inducer with the addition of a sulphonamido group, which may suggest that sulphur atoms can decrease the propensity of a chemical structure to act as an inducer. In the crystal structure with PDB ID of 5fnt<sup>150</sup>, the addition of another benzene ring did not influence the property of the structure to act as an inducer.

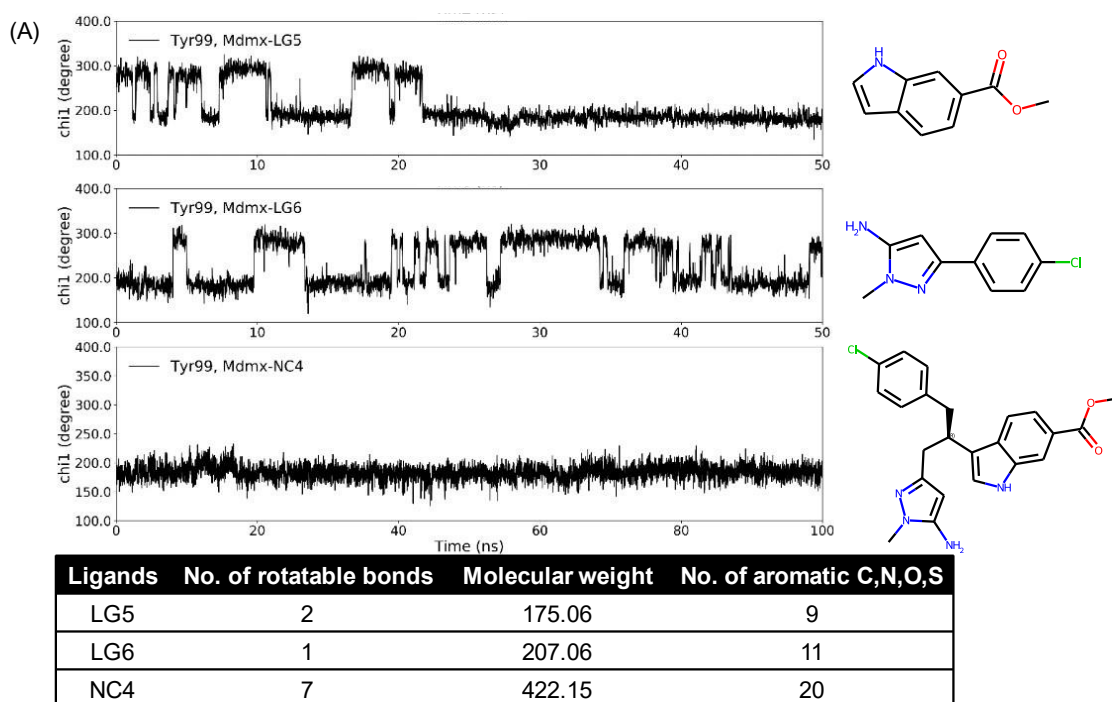
These examples illustrate that tiny changes in chemical structures can change their ability to act as inducers or non-inducers. They still follow the general tendencies identified using the SHAP and SetCompare algorithms.

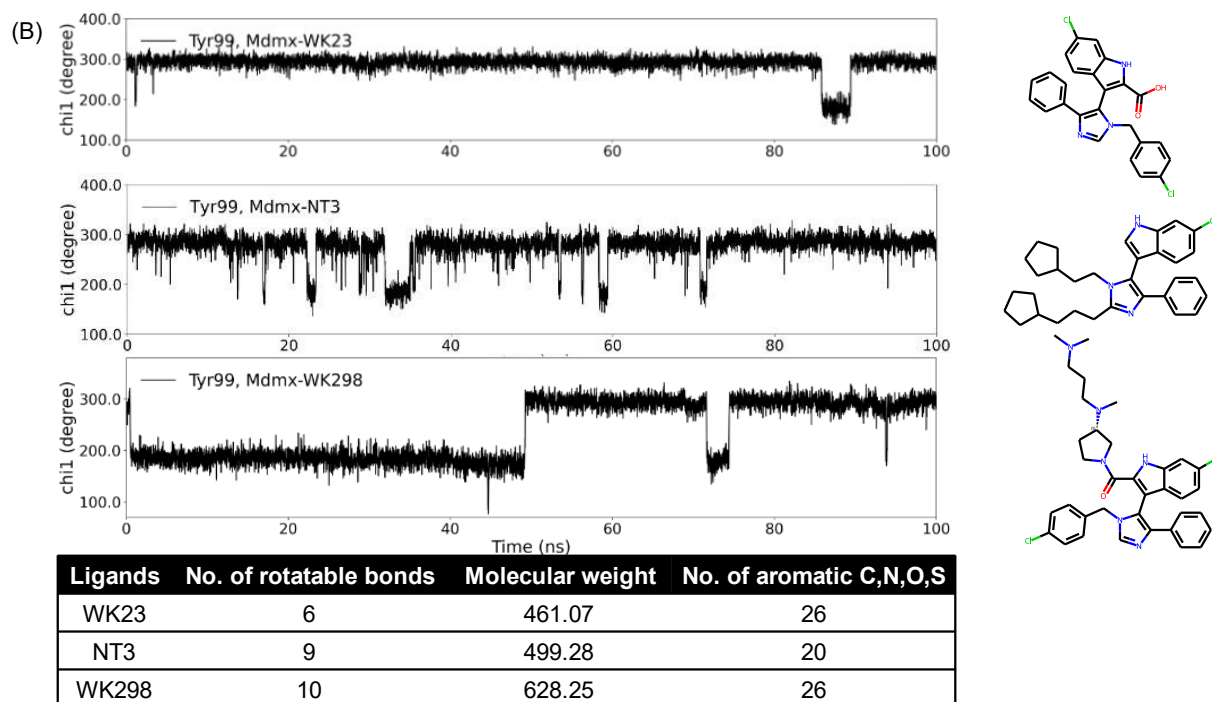


**Figure 6-3** Similar structures but which are identified as inducer and non-inducers in the PDBbind set. Green: inducer; red: non-inducer. The  $\Delta V$  represents the volume difference between the apo and holo pockets. The unit of  $\Delta V$  is  $\text{\AA}^3$ .

### 6.3.5 Analysis of Mdmx inhibitors

In the previous section of this dissertation, a series of Mdmx inhibitors were designed, i.e., NC and NT series compounds. Together with LG series fragments, all were studied using MD simulations. As seen from Figure 6-4, the increase in the number of rotatable bonds, molecular weight (MW) and aromatic atoms increased the probability of the ligands to bind the open state of Mdmx during the simulation time. Indeed, ligands LG5 and LG6 only partially bonded the Mdmx open state while NC4, which combined both ligands and thus had increased rotatable bonds and MW as well as aromatic atoms, continuously did it (Figure 6-4A). Figure 6-4B showed three ligands with common core structure consisting of single phenyl, indole and pyrazole rings. Compared to WK23<sup>6</sup> and NT3, WK298<sup>6</sup> had more rotatable bonds and MW. The addition of the tertiary amine sub-structure also promoted the inducing property of WK298 while WK23 and NT3 were both non-inducers. The changes of the inducing propensity of ligands were thus consistent with the analysis of the descriptors and functional groups described in sections 6.3.2 and 6.3.3.





**Figure 6-4 MD simulations and structural features of Mdmx inhibitors. WK23 (non-inducer) and WK298 (inducer) are known Mdmx inhibitors while the others are computationally predicted structures. The open conformation of Mdmx corresponds to Tyr99  $X_1$  around  $180^\circ$  while the closed one has  $X_1$  around  $300^\circ$ . The dynamic changes of  $X_1$  in Mdmx complexes with the analysed ligands show differences in binding properties of inducers (NC4 and WK298) as compared to the other analysed ligands. NC4 and WK298 have more data points of  $X_1$  around  $180^\circ$ , i.e. more open states; the others have more data points of  $X_1$  around  $300^\circ$ , i.e. more closed states.**

In addition, the known Mdmx inhibitors were retrieved from the ChEMBL database and the model was used to analyse these latest data. The search conditions were set as follows: i) Target: Mdmx, ii) Type: small molecule, iii) molecular weight < 700, iv) having exact experimental inhibitory activities (“=”). This set of molecules ( $n = 336$ ) was unique to the training set, and all of them were located in the applicability domain of the Inducer model. The prediction results showed that the majority of known Mdmx inhibitors ( $n = 324$ ) were classified as inducers, and only 19 of them were classified as non-inducers. This indicated that the classical paradigm to design these inhibitors targeted the ‘open’ state of Mdmx.

Mdmx has a transient pocket which is an extension of the Leu26 sub-pocket via the opening of Tyr99. The high flexibility of helices  $\alpha_2$  and  $\alpha_2'$  was observed in the MD trajectories (see Chapters 3 and 4). The characteristics of the corresponding receptor also affected the classification of inducers or non-inducers despite the molecular properties of the ligand itself. This, to some extent, explained from another perspective why most of the known Mdmx inhibitors were inducer-like due to the highly flexible Mdmx.

## 6.4 Summary and outlook

In this part of the work, the features of ligands that are relevant to the opening of cryptic pockets were investigated, which could be important in designing Mdmx inhibitor binding in the ‘open’ conformation. An automated program was developed to establish a holo-apo pair dataset, which accurately matched apo proteins for each holo (bound with a small-molecule ligand) protein. The training set comprised both the PDBbind and CryptoSite databases. NR-DBIND was used as the external validation set. The volume change between the holo and apo pockets was used as a metric to identify whether the ligand was an inducer. Classification models were developed by OCHEM and the optimum threshold to separate both types of molecules was determined as 60 Å<sup>3</sup>. The analysis of the developed machine learning model indicated that higher hydrophobicity and aromaticity increased the propensity of ligands to act as inducers. Inducers also tended to have tertiary amines, rather than primary or secondary amines. The impact of sulphur groups was ambiguous, and the analysis suggested that sulphones/sulphoxides substituents decreased the probability of molecules being inducers. The presence of phosphorus or halogen atoms increases the probability of molecules to be inducers, as identified by the SHAP and SetCompare methods. The five-membered heterocycles with two heteroatoms and pyridine were overrepresented in the class of inducers as identified by the SetCompare analysis. Pairs of similar molecules with opposite properties were also analysed and it was shown that small changes in the structures of molecules could result in a change in the class of compounds. We validated our findings about the inducing properties on the Mdmx inhibitors. Based on the results from MD simulations, Mdmx inhibitors that induced the opening of the transient pocket possess the same features as summarized from the previous analysis of the inducers. Last but not the least, the model was applied to a set of known Mdmx inhibitors collected from the ChEMBL database. Most of the Mdmx inhibitors available in the literature were classified as inducers, thus indicating that the classical paradigm to design these inhibitors was targeting the ‘open’ state.

Lipinski’s rule of five was implemented by Schrodinger software for analysing inducers. Majority of inducers (64%) were in agreement with the rules. A deep learning classification model<sup>151</sup> was used to evaluate the aggregation propensity of the analysed ligands. As a result, 92.7% of our dataset were predicted as non-aggregators and thus aggregation propensity did not affect the conclusion of this work.

The proposed approach is a general one and can be adopted to analyse, e.g., peptide ligands, identify inducers of enzymes and/or signaling proteins involved in extensive protein/protein interaction networks. Since this is a statistical approach a sufficiently large dataset should be collected from the literature to develop these models. Depending on the target class of compounds additional descriptors could be also considered, e.g. encoding the conformational information of peptide ligands.

This study revealed the structural features that are important for molecules to induce cryptic sites in a potential protein target, and the developed model can be used to provide such an analysis of potential ligands.

The calculated models do not have high accuracy, which can be attributed to the difficulty of the analysed property. Indeed, if the molecule has features that make it an inducer, it may not always induce a new pocket if there is already a suitable pocket for binding. Moreover, the protein should also have structural features that can enable cryptic pocket opening.

The Inducer classification model has the potential to be improved. The use of MD results as a filter to validate model predictions and/or their use as an additional input to the model could increase its accuracy. Furthermore, MD analysis can also improve the measurement of pocket volume. While MD simulations require a lot of computational time, a faster method could be to use the flexibility of the protein, especially of the binding site. Such an extension of the model should make it sensitive to the type of protein analysed. Still, the analysis provided and the features identified could be important for researchers in the design of new molecules which can open cryptic pockets. This topic is worth being further investigated together with wet-lab work in the future.

## 7 Discussion

### 7.1 *Difference in binding modes of ligands with Mdmx and Mdm2*

Previous studies have suggested that the open state of Tyr100 in Mdm2 leads to the high activity and selectivity of inhibitors against Mdm2. Therefore, the opening Tyr99 of Mdmx was considered a promising strategy and the side chain torsion angle  $\chi_1$  of Tyr99 was frequently used as an indicator to check whether it was open or not.

This work helped to understand the differences between the Mdmx and Mdm2 pockets in nature. As concluded in Chapter 3, the variant amino acids Met53/Pro95 and Leu54/His96 which were present in the pockets of Mdmx and Mdm2, respectively, could be important for selective targeting of Mdmx (Figure 7-1).

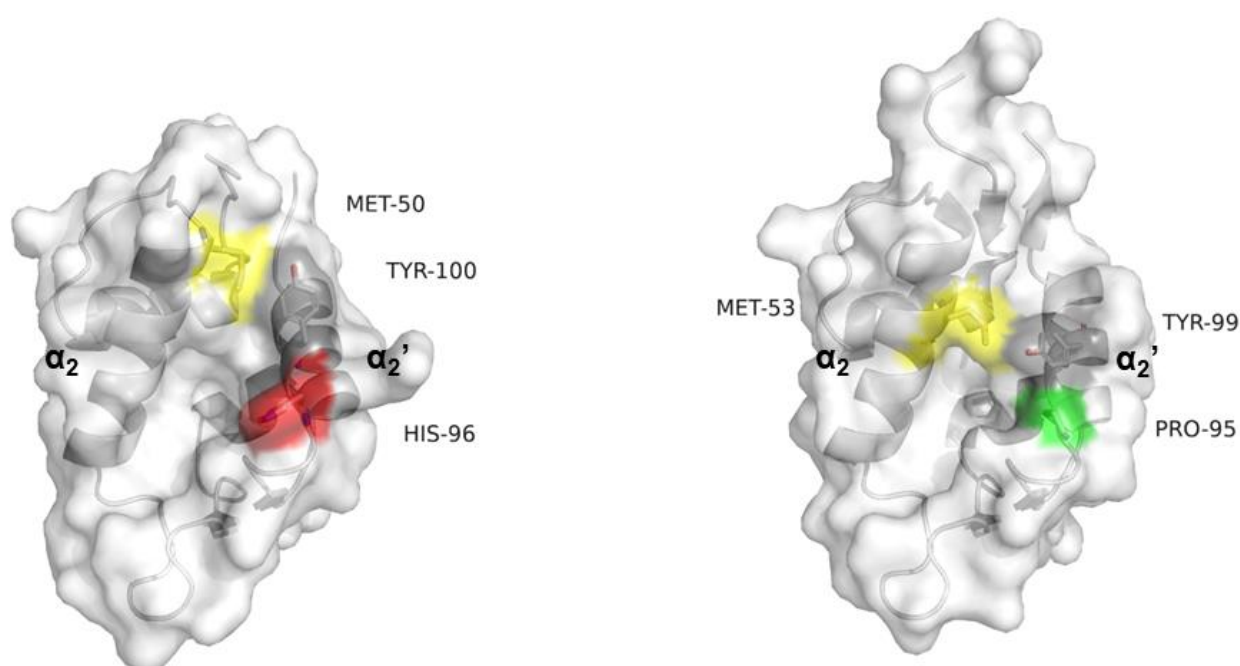
First, proline was undoubtedly a special amino acid because it was the only one that was connected to the main chain twice by forming a ring, which prevented itself from adopting a normal helical conformation. The hydrophobicity of proline may underlie its propensity to interact with aromatic structures. It is worth noting that Pro95 was the first residue of helix  $\alpha_2'$  in the crystal structure of the Mdmx-p53 complex (PDB ID: 3dab) but lost its helical conformation in the crystal structure of the Mdmx-WK298 complex (PDB ID: 3lbj). At the equivalent position, His96 in Mdm2 was the most common amino acid in the protein binding site, which was polar and free to be charged or neutral. Histidine also exhibits aromaticity. Therefore, this site of Mdm2 could be more active in forming interactions with ligands compared to the Pro95 of Mdmx. According to the crystal analysis in Chapter 3, His96 significantly contributed to the interaction with Mdm2 inhibitors.

Second, the particularity of methionine was that its sulphur atom could mediate the Met-aromatic interaction<sup>49</sup>. The aromatic residues could be tyrosine, tryptophan, and phenylalanine. In the Mdmx-p53 complex, Met53 interacted with Trp23 of p53 and Tyr99 of Mdmx, together with the hydrogen bonds formed between Tyr99 and Pro27, restricting Tyr99 in the closed conformation. This might determine the closed conformation as the intrinsic conformation of Mdmx. In the Mdmx-WK298 complex, the 4-chlorobenzyl ring participated in the Met-aromatic interaction; however, Tyr99 was farther to Met53 and the Pro27 sub-pocket was unfilled, which might decrease the binding affinity to Mdmx. This implied that WK298 induced the opening of Tyr99 for a longer time, possibly due to the reduction of the Met-aromatic interaction between Met53 and Tyr99. Moreover, the Met-aromatic interaction also plays an important role in the Mdm2 pocket. Met50 (Val49 in Mdmx) interacted with Tyr100 (Tyr99 in Mdmx), forming an open orientation. In the Mdm2-YH300 complex (PDB ID: 4mdn), the 4-chlorobenzyl of YH300 filled the Pro27 sub-pocket and blocked the Met50-Tyr100 interaction, but also possibly took part in this Met-aromatic interaction to some extent.



Based on the above considerations, it was suggested that the open state was intrinsic for Mdm2 and the closed state was native to Mdmx. In addition to the side chain torsion angle  $\chi_1$  of Tyr99 (Tyr100 in Mdm2), the distance between the sulphur atom of methionine and the ring of tyrosine could be used to evaluate the strength of the Met-aromatic interaction that mattered to the ligands binding with Mdmx/Mdm2. The Met53-Tyr99-ligand interaction is crucial for the post-analysis of the Mdmx-ligand complex.

From our point of view, the correlation between inhibitory activity and the transient state of Mdmx remains unclear. Stable binding with Mdmx should maintain the intrinsic state of the receptor ('closed'), which was proposed as an alternative strategy to design Mdmx inhibitors in this thesis. This assumption could be argued in the future with more experimental data targeting the 'closed' state of Mdmx.



**Figure 7-1** The key residues (stick) of Mdm2 (left, PDB ID: 1ycr) and Mdmx (right, PDB ID: 3dab); yellow surface: methionine, grey surface: tyrosine, red surface: histidine, green surface: proline.

## 7.2 Explanation of inactivity of synthesised ligands

In Chapter 4, several ligands were designed that were later found to be inactive during experimental testing in Mdm2 assays. As explained earlier, LG series were active fragments against Mdmx identified by NMR; NC series compounds were optimised from LG fragments; SN series were derivatives of two promising NC compounds (NC8 and NC9).

There are several possible explanations for why the SN1–4 compounds did not show activity in the wet experiments.

First, SN1–4 did not have identical structures as initially designed, especially SN1 and 2 lack of the indole-like substructure which was necessary for occupying the Trp23 sub-pocket as well as interacting with the Met53 of Mdmx.

Second, the duration of MD simulations was too short for our molecular system according to the subsequent longer trajectories, as explained in Chapter 4. At the beginning of the work, there were limited computational resources without GPU acceleration, so a 10-ns trajectory was used to determine the synthesis template. In the very first 10ns of the entire trajectories (Figure 4-16), all of them including the templates (NC8 and 9 compounds) entered into a stable binding state at the last moment. However, the binding conformation changed significantly during the rest of each trajectory. For instance, in the Mdmx-NC9 complex, the RMSD of NC9 fluctuated throughout and Mdmx significantly deviated from the initial stable binding conformation in the last half of the trajectory, illustrating that it was not a qualified template for carrying out the synthesis.

Third, the target protein was Mdm2 in the bioactivity test because, as stated by the scientists performing the assay, “if a molecule can bind with Mdmx, it will probably bind tightly with Mdm2” based on their practical experience. Nevertheless, this information was not known at the phase of molecular design. Therefore, the ligands were not optimised against Mdm2, which could be one of the reasons for these negative results. In the future, it would be better to start the simulation by targeting Mdm2 at first and then targeting Mdmx.

Fourth, only the RMSD data taken into account were insufficient when making decisions for the synthesis. Given the distance between Met53 and Tyr99 as an indicator for checking the binding situation, NC8 and NC9 would not have been selected as templates. As noted later, nearly half of their conformations had a distance of more than 6 Å (Figure 4-18); thus, they lost the Met-aromatic interactions, or these interactions were too weak, as discussed by Weber et al<sup>49</sup>. Unfortunately, this study was not available during this phase of the project.

Taken together, a reliable MD analysis should include the following: 1) RMSD of the ligand and the protein throughout the whole trajectory referring to the initial structure, which is used to check the stability of the binding; 2) the significant and specific interaction measured by distance, angle, etc., which is used to check the specificity of the binding. However, the priority is to ensure that the ligand does not come out of the pocket, which is difficult to know exactly from RMSD plots (e.g. LG1 and SN4 reside far away from the pocket; Figure 4-10 and Figure 4-16). Therefore, one should check the last frames of the trajectory to fulfil this requirement prior to any data analysis. It is also suggested that using cluster analysis of conformations to identify the meta-stable states, transition states, and dominant states. In addition, the estimation of the binding free energy is a good reference to determine the affinity between the protein and ligand before carrying out wet experiments.

In addition to the general analyses mentioned above, MD simulations and post-analysis should be set up in a case by case basis. With the hope of finding the ligands stably and selectively binding with the target protein as well as looking into the binding mode at the atom level, an exclusive analytic plan for the available data by monitoring a specific distance, angle, torsion, and H-bond through the whole trajectory will help to reduce the error probability in the wet experiment.

## 8 Conclusion

With the emergence of many new methods and rapid advances in computer technology, CADD has become an important tool for pharmaceutical companies and academia. Faster and cheaper computers and related hardware and software have made it possible to simulate complex biological systems at the atomic scale, which has greatly facilitated the development of CADD. The drugs designed by CADD, such as zanamivir and imatinib, have been marketed successfully. Mdmx is a promising anticancer target. The structural particularity of Mdmx causes difficulties in the design of inhibitors against Mdmx. Although several candidates targeting Mdm2 have entered the clinical stage, there is no such progress exclusively for Mdmx. This inspired us to seek new insights into the development of Mdmx inhibitors using advanced computational methods and exhaustively investigate all available data, including structural and activity information.

Starting from the variant residues in Mdmx and Mdm2 and a comprehensive survey of relevant crystal structures, methionine was found to be very important for determining the different shapes of binding pockets (**Chapter 3**). The Met-aromatic interaction restricted the rotation of Tyr99 in Mdmx (Tyr100 in Mdm2), leading to transient states in both proteins. The strength of the Met-Tyr-Ligand interaction was evaluated based on the distance between them. The ‘open’ state of tyrosine designated by the side chain torsion angle  $\chi_1$  made the main pocket larger than the ‘closed’ state of tyrosine. The enlarged part corresponded to a cryptic pocket which was reachable only when the tyrosine was in the ‘open’ state. Therefore, it was concluded that Tyr100 of Mdm2 was open, while Tyr99 of Mdmx was closed in the intrinsic state. However, that Mdmx was also able to bind ligands in the ‘open’ state could not be completely excluded; thus, both states were considered later in the thesis.

Optimised from known active compounds, new Mdmx inhibitors were developed to mimic the binding mode of p53 with Mdmx using ligand-based drug design methods (**Chapter 4**). One of the strategies was to optimise the structure of the known inhibitor WK298, and the other was to connect active fragments to make molecules fully occupy the binding pocket. Their binding modes were obtained by MD simulations and evaluated by RMSD, dynamic change of Tyr99  $\chi_1$ , and distance between Met53 and Tyr99. By visually inspecting conformations, new molecules were either binding separately with ‘open’ and ‘closed’ states of Mdmx, or transforming between the two states in a dynamic way through the whole simulation. Since experimental testing did not confirm the theoretical prediction of the activity of these molecules, a critical analysis of the used protocol pointed out several shortcomings that should be avoided in the future for similar studies.

A machine learning model named Generator (**Chapter 5**) was developed to efficiently create new molecules with predefined properties. Combined with a QSAR model of Mdmx inhibitors as well as models to predict physicochemical properties, the Generator generated potential Mdmx inhibitors with completely new structures, but also re-discovered existing molecules that

were available to be tested, circumventing the need for the synthesis. The promising molecules had better predicted  $pIC_{50}$  than WK298 and constantly kept the ‘closed’ state of Mdmx. In contrast, WK298 induced Mdmx to present half ‘open’ and half ‘closed’ conformations. This suggests that the ‘closed’ state should not be neglected for the R&D of Mdmx inhibitors. The overall approach is entirely artificial intelligence driven until the final inspection of the binding poses proved its ability, at least virtually, to design new compounds with machine learning methods and validate them with traditional molecular dynamics simulations in an automated manner. The study demonstrated that the Generator equipped with appropriate constraints could rationally navigate in the chemical space and generate promising inhibitors.

In the last part of this PhD research project, a general classification model based on the crystal data deposited in the PDB database was developed to learn the common features of molecules that could induce the opening of cryptic pockets (**Chapter 6**). An automated program was utilised to establish a holo-apo pair dataset, which accurately matched apo proteins for each holo protein. The volume changes between apo and holo proteins were used to differentiate the chemical features in inducers and non-inducers. Based on the holo-apo pair dataset, classification models were built to determine the optimum threshold. The model analysis suggested that inducers were more hydrophobic and aromatic, as well as an over-representation of phosphorus and halogen atoms. Fragment analysis showed that small changes in the structure of molecules could strongly affect the potential to induce a cryptic pocket. The model was used to classify a set of small-molecule Mdmx inhibitors collected from the ChEMBL database that were not included in the training set ( $n = 336$ ). Interestingly, most of Mdmx inhibitors available in the literature ( $n = 324$ ) were classified as inducers thus indicating the classical paradigm to design these inhibitors was targeting the ‘open’ state of Mdmx but our analysis suggested that ‘closed’ state should be also considered. The provided analysis and identified features of the Inducer classification model can be used to design new molecules which could open cryptic pockets.

**In the conclusion of the thesis**, new perspectives on the structure-based approaches for designing Mdmx inhibitors were summarised and the importance to consider also the ‘closed’ state of Mdmx was pinpointed. Through in-depth computational analyses, new insights into crucial interactions that caused the distinct binding modes of small-molecule ligands with Mdm2/Mdmx were uncovered. The classical and common strategy that either Mdm2 or Mdmx ligand should simultaneously occupy the three sub-pockets (Phe19, Trp23, and Leu26) is still inevitable for developing potent Mdm2/Mdmx inhibitors. The study produced putative ligands either directly optimised from known active compounds or *de novo* created by the Generator model, which would be promising for testing their inhibitory activities in the future. The use of the Inducer model would help discover novel ligands that precisely target the transient states. Moreover, both Generator and Inducer models are not limited to Mdmx, but can also be used for other drug targets.

## Acknowledgements

First and foremost, I would like to thank Dr. Igor V. Tetko, who has been the guide and a supervisor for me throughout the four-year study abroad. I am very grateful to have the opportunity to work in his group. Without his constant effort and help and patience, the project and my doctoral studies would never have succeeded. My sincere thanks to Prof. Dr. Michael Sattler for hosting me at the Institute of Structural Biology/ Department of Chemistry and been my academic supervisor. Despite his busy schedule, he always found time to give his valuable advice on my studies. I would also like to thank Dr. Grzegorz Popowicz, who was my third supervisor and the expert in the protein Mdmx. He provided valuable ideas for my project and practical comments on the scientific writing including my study reports, publications, and this dissertation.

A special thanks to Dr. Pavel Karpov, a talented scientist and good mentor, for accompanying from day one. He helped me during our first collaboration and move forward beyond my PhD study. A very big thanks to all the alumni and people in Dr. Tetko's group: Genny Cau, Elena Golosovskaia, Markus Antonios Grimm, Arvid Wartenberg, Varvara Voinarovska, and Peter Hartog, for being together during my stay at the lab and such great company. Also thanks to Dr. Dipan Ghosh for sharing his precious experience. He generously gave me supports and suggestions when I felt lost sometime. I would also like to thank Michael Withnall for proofreading my first publication in the group.

I sincerely appreciate my parents and anyone who loves me and whom I love, for their quietly concern and eternal support.

Last but not least, I want to thank the China Scholarship Council (No. CSC201706880010) for the PhD scholarship, thank Dr. Markus Drees who is the managing director of the Departmental Graduate Center for constantly patient and clear instructions about PhD regulations, thank HEL-ENA (Helmholtz Graduate School Environmental Health) for financial support of my attendance at the national academic meeting in 2019, and thank Prof. Dr. Angela Casini to chair my PhD exam.

## References

- (1) Vousden, K. H.; Lane, D. P. P53 in Health and Disease. *Nature Reviews Molecular Cell Biology* **2007**, *8* (4), 275–283. <https://doi.org/10.1038/nrm2147>.
- (2) US8053457B2 - 3-imidazolyl-indoles for the treatment of proliferative diseases - Google Patents <https://patents.google.com/patent/US8053457B2/en> (accessed 2021 -11 -09).
- (3) Bista, M.; Wolf, S.; Khoury, K.; Kowalska, K.; Huang, Y.; Wrona, E.; Arciniega, M.; Popowicz, G. M.; Holak, T. A.; Dömling, A. Transient Protein States in Designing Inhibitors of the MDM2-P53 Interaction. *Structure* **2013**, *21* (12), 2143–2151. <https://doi.org/10.1016/j.str.2013.09.006>.
- (4) Popowicz, G. M.; Czarna, A.; Holak, T. A. Structure of the Human Mdmx Protein Bound to the P53 Tumor Suppressor Transactivation Domain. *Cell cycle (Georgetown, Tex.)*. United States August 2008, pp 2441–2443. <https://doi.org/10.4161/cc.6365>.
- (5) Popowicz, G. M.; Dömling, A.; Holak, T. A. The Structure-Based Design of Mdm2/Mdmx-P53 Inhibitors Gets Serious. *Angewandte Chemie - International Edition* **2011**, *50* (12), 2680–2688. <https://doi.org/10.1002/anie.201003863>.
- (6) Popowicz, G. M.; Czarna, A.; Wolf, S.; Wang, K.; Wang, W.; Dömling, A.; Holak, T. A. Structures of Low Molecular Weight Inhibitors Bound to MDMX and MDM2 Reveal New Approaches for P53-MDMX/MDM2 Antagonist Drug Discovery. *Cell Cycle* **2010**, *9* (6), 1104–1111. <https://doi.org/10.4161/cc.9.6.10956>.
- (7) Popowicz, G. M.; Czarna, A.; Rothweiler, U.; Szwagierczak, A.; Krajewski, M.; Weber, L.; Holak, T. A. Molecular Basis for the Inhibition of P53 by Mdmx. *Cell Cycle* **2007**, *6* (19), 2386–2392. <https://doi.org/10.4161/cc.6.19.4740>.
- (8) Jorgensen, W. L.; Madura, J. D. Quantum and Statistical Mechanical Studies of Liquids. 25. Solvation and Conformation of Methanol in Water. *Journal of the American Chemical Society* **2002**, *105* (6), 1407–1413. <https://doi.org/10.1021/JA00344A001>.
- (9) 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. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3696–3713. [https://doi.org/10.1021/ACS.JCTC.5B00255/SUPPL\\_FILE/CT5B00255\\_SI\\_001.PDF](https://doi.org/10.1021/ACS.JCTC.5B00255/SUPPL_FILE/CT5B00255_SI_001.PDF).
- (10) Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model. *Journal of the American Chemical Society* **2002**, *107* (13), 3902–3909. <https://doi.org/10.1021/JA00299A024>.
- (11) Jakalian, A.; Bush, B. L.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: I. Method. *Journal of Computational Chemistry* **2000**, *21* (2), 132–146. [https://doi.org/10.1002/\(SICI\)1096-987X\(20000130\)21:2](https://doi.org/10.1002/(SICI)1096-987X(20000130)21:2).
- (12) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC Model: II. Parameterization and Validation. *Journal of computational chemistry* **2002**, *23* (16), 1623–1641. <https://doi.org/10.1002/JCC.10128>.
- (13) Shvarts, A.; Steegenga, W. T.; Riteco, N.; van Laar, T.; Dekker, P.; Bazuine, M.; van Ham, R. C.; van der Houven van Oordt, W.; Hateboer, G.; van der Eb, A. J.; Jochemsen,

- A. G. MDMX: A Novel P53-Binding Protein with Some Functional Properties of MDM2. *The EMBO journal* **1996**, *15* (19), 5349–5357.
- (14) Vousden, K. H.; Lu, X. Live or Let Die: The Cell's Response to P53. *Nature Reviews Cancer* **2002**, *2* (8), 594–604. <https://doi.org/10.1038/nrc864>.
- (15) Vogelstein, B.; Lane, D.; Levine, A. J. Surfing the P53 Network. *Nature* **2000**, *408* (6810), 307–310. <https://doi.org/10.1038/35042675>.
- (16) Toledo, F.; Wahl, G. M. Regulating the P53 Pathway: In Vitro Hypotheses, in Vivo Veritas. *Nature Reviews Cancer* **2006**, *6* (12), 909–923. <https://doi.org/10.1038/nrc2012>.
- (17) Momand, J.; Zambetti, G. P.; Olson, D. C.; George, D.; Levine, A. J. The *mdm-2* Oncogene Product Forms a Complex with the P53 Protein and Inhibits P53-Mediated Transactivation. *Cell* **1992**, *69* (7), 1237–1245. [https://doi.org/10.1016/0092-8674\(92\)90644-R](https://doi.org/10.1016/0092-8674(92)90644-R).
- (18) Haupt, S.; Mejía-Hernández, J. O.; Vijayakumaran, R.; Keam, S. P.; Haupt, Y. The Long and the Short of It: The MDM4 Tail so Far. *Journal of Molecular Cell Biology* **2019**, *11* (3), 231–244. <https://doi.org/10.1093/jmcb/mjz007>.
- (19) Konopleva, M.; Martinelli, G.; Daver, N.; Papayannidis, C.; Wei, A.; Higgins, B.; Ott, M.; Mascarenhas, J.; Andreeff, M. MDM2 Inhibition: An Important Step Forward in Cancer Therapy. *Leukemia* **2020**, *34* (11), 2858–2874. <https://doi.org/10.1038/s41375-020-0949-z>.
- (20) Reed, D.; Shen, Y.; Shelat, A. A.; Arnold, L. A.; Ferreira, A. M.; Zhu, F.; Mills, N.; Smithson, D. C.; Regni, C. A.; Bashford, D.; Cicero, S. A.; Schulman, B. A.; Jochemsen, A. G.; Guy, R. K.; Dyer, M. A. Identification and Characterization of the First Small Molecule Inhibitor of MDMX \*. *Journal of Biological Chemistry* **2010**, *285* (14), 10786–10796. <https://doi.org/10.1074/jbc.M109.056747>.
- (21) Bista, M.; Smithson, D.; Pecak, A.; Salinas, G.; Pustelny, K.; Min, J.; Pirog, A.; Finch, K.; Zdzalik, M.; Waddell, B.; Wladyka, B.; Kedracka-Krok, S.; Dyer, M. A.; Dubin, G.; Guy, R. K. On the Mechanism of Action of SJ-172550 in Inhibiting the Interaction of MDM4 and P53. *PLoS ONE* **2012**, *7* (6), e37518. <https://doi.org/10.1371/journal.pone.0037518>.
- (22) Roh, J.-L.; Park, J. Y.; Kim, E. H. XI-011 Enhances Cisplatin-Induced Apoptosis by Functional Restoration of P53 in Head and Neck Cancer. *Apoptosis : an international journal on programmed cell death* **2014**, *19* (11), 1594–1602. <https://doi.org/10.1007/s10495-014-1026-8>.
- (23) Wang, H.; Ma, X.; Ren, S.; Buolamwini, J. K.; Yan, C. A Small-Molecule Inhibitor of MDMX Activates P53 and Induces Apoptosis. *Molecular cancer therapeutics* **2011**, *10* (1), 69–79. <https://doi.org/10.1158/1535-7163.MCT-10-0581>.
- (24) Joseph, T. L.; Madhumalar, A.; Brown, C. J.; Lane, D. P.; Verma, C. Differential Binding of P53 and Nutlin to MDM2 and MDMX: Computational Studies. *Cell Cycle* **2010**, *9* (6), 1167–1181. <https://doi.org/10.4161/cc.9.6.11067>.
- (25) Li, Q.; Lozano, G. Molecular Pathways: Targeting Mdm2 and Mdm4 in Cancer Therapy. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2013**, *19* (1), 34–41. <https://doi.org/10.1158/1078-0432.CCR-12-0053>.
- (26) Kon, N.; Gu, W. Is MDMX the Better Target? *Aging* **2018**, *10* (6), 1184–1185. <https://doi.org/10.18632/aging.101479>.

- (27) Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and Scoring in Virtual Screening for Drug Discovery: Methods and Applications. *Nature Reviews Drug Discovery* **2004**, *3* (11), 935–949. <https://doi.org/10.1038/nrd1549>.
- (28) Berg, J. M.; Tymoczko, J. L.; Stryer, L. Biochemistry. *Biochemistry* **2002**, 319–344.
- (29) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *Journal of Computational Chemistry* **2010**, *31* (2), 455–461. <https://doi.org/https://doi.org/10.1002/jcc.21334>.
- (30) D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T, D. M. Y. and P. A. K. AMBER 2020, University of California, San Francisco. University of California, San Francisco 2020.
- (31) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *Journal of Computational Chemistry* **2004**, *25* (9), 1157–1174. <https://doi.org/10.1002/JCC.20035>.
- (32) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *The Journal of Chemical Physics* **1998**, *79* (2), 926. <https://doi.org/10.1063/1.445869>.
- (33) Popowicz, G. M.; Czarna, A.; Holak, T. A. Structure of the Human Mdmx Protein Bound to the P53 Tumor Suppressor Transactivation Domain. *Cell Cycle* **2008**, *7* (15), 2441–2443. <https://doi.org/10.4161/cc.6365>.
- (34) Miller, B. R.; McGee, T. D.; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.Py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation* **2012**, *8* (9), 3314–3321. <https://doi.org/10.1021/ct300418h>.
- (35) Wang, J.; Hou, T.; Xu, X. Recent Advances in Free Energy Calculations with a Combination of Molecular Mechanics and Continuum Models. *Current Computer Aided-Drug Design* **2006**, *2* (3), 287–306. <https://doi.org/10.2174/157340906778226454>.
- (36) Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A. K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V. V; Tanchuk, V. Y.; Todeschini, R.; Varnek, A.; Marcou, G.; Ertl, P.; Potemkin, V.; Grishina, M.; Gasteiger, J.; Schwab, C.; Baskin, I. I.; Palyulin, V. A.; Radchenko, E. V; Welsh, W. J.; Kholodovych, V.; Chekmarev, D.; Cherkasov, A.; Aires-de-Sousa, J.; Zhang, Q.-Y.; Bender, A.; Nigsch, F.; Patiny, L.; Williams, A.; Tkachenko, V.; Tetko, I. V. Online Chemical Modeling Environment (OCHEM): Web Platform for Data Storage, Model Development and Publishing of Chemical Information. *Journal of Computer-Aided Molecular Design* **2011**, *25* (6), 533–554. <https://doi.org/10.1007/s10822-011-9440-2>.
- (37) Karpov, P.; Godin, G.; Tetko, I. V. Transformer-CNN: Fast and Reliable Tool for QSAR. **2019**. <https://doi.org/10.1186/s13321-020-00423-w>.
- (38) Schrödinger, L. The PyMOL Molecular Graphics System. Schrödinger, LLC: New York 2021.
- (39) 2020-3, S. R. Maestro, Schrödinger, LLC, New York, NY, 2020.



- (40) Joseph, T. L.; Madhumalar, A.; Brown, C. J.; Lane, D. P.; Verma, C. Differential Binding of P53 and Nutlin to MDM2 and MDMX: Computational Studies. <https://doi.org/10.4161/cc.9.6.11067> **2010**, *9* (6), 1167–1181. <https://doi.org/10.4161/CC.9.6.11067>.
- (41) Chen, J.; Wang, J.; Zhang, Q.; Chen, K.; Zhu, W. Probing Origin of Binding Difference of Inhibitors to MDM2 and MDMX by Polarizable Molecular Dynamics Simulation and QM/MM-GBSA Calculation. *Scientific Reports 2015 5:1* **2015**, *5* (1), 1–16. <https://doi.org/10.1038/srep17421>.
- (42) Zhao, Y.; Bernard, D.; Wang, S. Small Molecule Inhibitors of MDM2-P53 and MDMX-P53 Interactions as New Cancer Therapeutics. *BioDiscovery* **2013**, *8* (8), 4. <https://doi.org/10.7750/BioDiscovery.2013.8.4>.
- (43) Kussie, P. H.; Gorina, S.; Marechal, V.; Elenbaas, B.; Moreau, J.; Levine, A. J.; Pavletich, N. P. Structure of the MDM2 Oncoprotein Bound to the P53 Tumor Suppressor Transactivation Domain. *Science (New York, N.Y.)* **1996**, *274* (5289), 948–953. <https://doi.org/10.1126/science.274.5289.948>.
- (44) Morris, G. M.; Ruth, H.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *Journal of computational chemistry* **2009**, *30* (16), 2785. <https://doi.org/10.1002/JCC.21256>.
- (45) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Research* **2000**, *28* (1), 235–242. <https://doi.org/10.1093/nar/28.1.235>.
- (46) Popowicz, G. M.; Czarna, A.; Wolf, S.; Wang, K.; Wang, W.; Dyming, A.; Holak, T. A. Structures of Low Molecular Weight Inhibitors Bound to MDMX and MDM2 Reveal New Approaches for P53-MDMX/MDM2 Antagonist Drug Discovery. *Cell Cycle* **2010**, *9* (6), 1104–1111. <https://doi.org/10.4161/cc.9.6.10956>.
- (47) D.A. Case et al. Amber 2020 Reference Manual. *Amber* **2020**.
- (48) Betts, M. J.; Russell, R. B. Amino Acid Properties and Consequences of Substitutions. *Bioinformatics for geneticists* **2003**, *317*, 289.
- (49) Weber, D. S.; Warren, J. J. The Interaction between Methionine and Two Aromatic Amino Acids Is an Abundant and Multifunctional Motif in Proteins. *Archives of Biochemistry and Biophysics* **2019**, *672*, 108053. <https://doi.org/https://doi.org/10.1016/j.abb.2019.07.018>.
- (50) Popowicz, G. M.; Czarna, A.; Wolf, S.; Wang, K.; Wang, W.; Dyming, A.; Holak, T. A. Structures of Low Molecular Weight Inhibitors Bound to MDMX and MDM2 Reveal New Approaches for P53-MDMX/MDM2 Antagonist Drug Discovery. *Cell Cycle* **2010**, *9* (6), 1104–1111. <https://doi.org/10.4161/cc.9.6.10956>.
- (51) L. Grasberger, B.; Lu, T.; Schubert, C.; J. Parks, D.; E. Carver, T.; K. Koblisch, H.; D. Cummings, M.; V. LaFrance, L.; L. Milkiewicz, K.; R. Calvo, R.; Maguire, D.; Lattanze, J.; F. Franks, C.; Zhao, S.; Ramachandren, K.; R. Bylebyl, G.; Zhang, M.; L. Manthey, C.; C. Petrella, E.; W. Pantoliano, M.; C. Deckman, I.; C. Spurlino, J.; C. Maroney, A.; E. Tomczuk, B.; J. Molloy, C.; F. Bone, R. Discovery and Cocrystal Structure of Benzodiaz-

- epinedione HDM2 Antagonists That Activate P53 in Cells. *Journal of Medicinal Chemistry* **2005**, 48 (4), 909–912. <https://doi.org/10.1021/jm049137g>.
- (52) G. Allen, J.; P. Bourbeau, M.; Erich Wohlhieter, G.; D. Bartberger, M.; Michelsen, K.; Hungate, R.; C. Gadwood, R.; D. Gaston, R.; Evans, B.; W. Mann, L.; E. Matison, M.; Schneider, S.; Huang, X.; Yu, D.; S. Andrews, P.; Reichelt, A.; M. Long, A.; Yakowec, P.; Y. Yang, E.; Ann Lee, T.; D. Oliner, J. Discovery and Optimization of Chromenotriazolopyrimidines as Potent Inhibitors of the Mouse Double Minute 2–Tumor Protein 53 Protein–Protein Interaction. *Journal of Medicinal Chemistry* **2009**, 52 (22), 7044–7053. <https://doi.org/10.1021/jm900681h>.
- (53) Popowicz, G. M.; Czarna, A.; Wolf, S.; Wang, K.; Wang, W.; Dömling, A.; Holak, T. A. Structures of Low Molecular Weight Inhibitors Bound to MDMX and MDM2 Reveal New Approaches for P53-MDMX/MDM2 Antagonist Drug Discovery. *Cell Cycle* **2010**, 9 (6), 1104–1111. <https://doi.org/10.4161/cc.9.6.10956>.
- (54) Picksley, S. M.; Vojtesek, B.; Sparks, A.; Lane, D. P. Immunochemical Analysis of the Interaction of P53 with MDM2;--Fine Mapping of the MDM2 Binding Site on P53 Using Synthetic Peptides. *Oncogene* **1994**, 9 (9), 2523–2529.
- (55) Lin, J.; Chen, J.; Elenbaas, B.; Levine, A. J. Several Hydrophobic Amino Acids in the P53 Amino-Terminal Domain Are Required for Transcriptional Activation, Binding to Mdm-2 and the Adenovirus 5 E1B 55-KD Protein. *Genes & development* **1994**, 8 (10), 1235–1246. <https://doi.org/10.1101/gad.8.10.1235>.
- (56) Shangary, S.; Wang, S. Small-Molecule Inhibitors of the MDM2-P53 Protein-Protein Interaction to Reactivate P53 Function: A Novel Approach for Cancer Therapy. *Annual review of pharmacology and toxicology* **2009**, 49, 223–241. <https://doi.org/10.1146/annurev.pharmtox.48.113006.094723>.
- (57) Massova, I.; A. Kollman, P. Computational Alanine Scanning To Probe Protein–Protein Interactions: A Novel Approach To Evaluate Binding Free Energies. *Journal of the American Chemical Society* **1999**, 121 (36), 8133–8143. <https://doi.org/10.1021/ja990935j>.
- (58) Grace, C. R.; Ban, D.; Min, J.; Mayasundari, A.; Min, L.; Finch, K. E.; Griffiths, L.; Bharatham, N.; Bashford, D.; Kiplin Guy, R.; Dyer, M. A.; Kriwacki, R. W. Monitoring Ligand-Induced Protein Ordering in Drug Discovery. *Journal of Molecular Biology* **2016**, 428 (6), 1290–1303. <https://doi.org/10.1016/j.jmb.2016.01.016>.
- (59) Rodrigues, J.; Teixeira, J. M. C.; Trellet, M.; Bonvin, A. Pdb-Tools: A Swiss Army Knife for Molecular Structures [Version 1; Peer Review: 2 Approved]. *F1000Research* **2018**, 7 (1961). <https://doi.org/10.12688/f1000research.17456.1>.
- (60) H. S. Segler, M.; Kogej, T.; Tyrchan, C.; P. Waller, M. Generating Focused Molecule Libraries for Drug Discovery with Recurrent Neural Networks. *ACS Central Science* **2017**, 4 (1), 120–131. <https://doi.org/10.1021/acscentsci.7b00512>.
- (61) Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular De-Novo Design through Deep Reinforcement Learning. *Journal of Cheminformatics* **2017**, 9 (1), 48. <https://doi.org/10.1186/s13321-017-0235-x>.
- (62) Graves, A. Generating Sequences With Recurrent Neural Networks. **2013**.

- (63) Weininger, D. SMILES, a Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *Journal of Chemical Information and Computer Sciences* **1988**, *28* (1), 31–36. <https://doi.org/10.1021/ci00057a005>.
- (64) Ertl, P.; Lewis, R.; Martin, E.; Polyakov, V. In Silico Generation of Novel, Drug-like Chemical Matter Using the LSTM Neural Network. **2017**.
- (65) Awale, M.; Sirockin, F.; Stiefl, N.; Reymond, J.-L. Drug Analogs from Fragment-Based Long Short-Term Memory Generative Neural Networks. *Journal of Chemical Information and Modeling* **2019**, *59* (4), 1347–1356. <https://doi.org/10.1021/acs.jcim.8b00902>.
- (66) Gaulton, A.; Bellis, L. J.; Bento, A. P.; Chambers, J.; Davies, M.; Hersey, A.; Light, Y.; McGlinchey, S.; Michalovich, D.; Al-Lazikani, B.; Overington, J. P. ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery. *Nucleic Acids Research* **2012**, *40* (D1), D1100–D1107. <https://doi.org/10.1093/nar/gkr777>.
- (67) Hayashi, R.; Wang, D.; Hara, T.; Iera, J. A.; Durell, S. R.; Appella, D. H. N-Acylpolyamine Inhibitors of HDM2 and HDMX Binding to P53. *Bioorganic & medicinal chemistry* **2009**, *17* (23), 7884–7893. <https://doi.org/10.1016/j.bmc.2009.10.032>.
- (68) Berghausen J, Buschmann N, Furet P, Gessier F, L. J.; Holzer P, Jacoby E, Kallen J, Masuya K, S. C. Substituted Isoquinolinones and Quinazolinones. US9051279B2, 2015.
- (69) Furet P, Guagnano V, Holzer P, Mah R, Masuya K, S.; A, Stutz S, V. A. Pyrazolopyrrolidine Compounds, 2015.
- (70) Furet P, Guagnano V, Holzer P, Kallen J, Masuya K, S. S. Cyclohexyl Isoquinolinone Compounds, 2014.
- (71) Furet P, Guagnano V, Holzer P, Kallen J, Lv L, Mah R, M. L.; Masuya K, Schlapbach A, S. S. Substituted Pyrrolo [3, 4-D] Imidazoles for the Treatment of MDM2/4 Mediated Diseases, 2014.
- (72) Madden, M. M.; Muppidi, A.; Li, Z.; Li, X.; Chen, J.; Lin, Q. Synthesis of Cell-Permeable Stapled Peptide Dual Inhibitors of the P53-Mdm2/Mdmx Interactions via Photoinduced Cycloaddition. *Bioorganic & medicinal chemistry letters* **2011**, *21* (5), 1472–1475. <https://doi.org/10.1016/j.bmcl.2011.01.004>.
- (73) Noguchi, T.; Oishi, S.; Honda, K.; Kondoh, Y.; Saito, T.; Kubo, T.; Kaneda, M.; Ohno, H.; Osada, H.; Fujii, N. Affinity-Based Screening of MDM2/MDMX–P53 Interaction Inhibitors by Chemical Array: Identification of Novel Peptidic Inhibitors. *Bioorganic & Medicinal Chemistry Letters* **2013**, *23* (13), 3802–3805. <https://doi.org/https://doi.org/10.1016/j.bmcl.2013.04.094>.
- (74) Blackburn, T. J.; Ahmed, S.; Coxon, C. R.; Liu, J.; Lu, X.; Golding, B. T.; Griffin, R. J.; Hutton, C.; Newell, D. R.; Ojo, S.; Watson, A. F.; Zaytzev, A.; Zhao, Y.; Lunec, J.; Hardcastle, I. R. Diaryl- and Triaryl-Pyrrole Derivatives: Inhibitors of the MDM2–P53 and MDMX–P53 Protein–Protein Interactions. *MedChemComm* **2013**, *4* (9), 1297–1304. <https://doi.org/10.1039/C3MD00161J>.
- (75) Macchiarulo, A.; Giacchè, N.; Carotti, A.; Moretti, F.; Pellicciari, R. Expanding the Horizon of Chemotherapeutic Targets: From MDM2 to MDMX (MDM4). *MedChemComm* **2011**, *2* (6), 455–465. <https://doi.org/10.1039/C0MD00238K>.
- (76) Holzer, P.; Masuya, K.; Furet, P.; Kallen, J.; Valat-Stachyra, T.; Ferretti, S.; Berghausen, J.; Bouisset-Leonard, M.; Buschmann, N.; Pissot-Soldermann, C.; Rynn, C.; Ruetz, S.;

- Stutz, S.; Chène, P.; Jeay, S.; Gessier, F. Discovery of a Dihydroisoquinolinone Derivative (NVP-CGM097): A Highly Potent and Selective MDM2 Inhibitor Undergoing Phase 1 Clinical Trials in P53wt Tumors. *Journal of Medicinal Chemistry* **2015**, *58* (16), 6348–6358. <https://doi.org/10.1021/acs.jmedchem.5b00810>.
- (77) Uesato, S.; Matsuura, Y.; Matsue, S.; Sumiyoshi, T.; Hirata, Y.; Takemoto, S.; Kawaratanani, Y.; Yamai, Y.; Ishida, K.; Sasaki, T.; Enari, M. Discovery of New Low-Molecular-Weight P53-Mdmx Disruptors and Their Anti-Cancer Activities. *Bioorganic & medicinal chemistry* **2016**, *24* (8), 1919–1926. <https://doi.org/10.1016/j.bmc.2016.03.021>.
- (78) Giustiniano, M.; Daniele, S.; Pelliccia, S.; La Pietra, V.; Pietrobono, D.; Brancaccio, D.; Cosconati, S.; Messere, A.; Giuntini, S.; Cerofolini, L.; Fragai, M.; Luchinat, C.; Taliani, S.; La Regina, G.; Da Settimo, F.; Silvestri, R.; Martini, C.; Novellino, E.; Marinelli, L. Computer-Aided Identification and Lead Optimization of Dual Murine Double Minute 2 and 4 Binders: Structure–Activity Relationship Studies and Pharmacological Activity. *Journal of Medicinal Chemistry* **2017**, *60* (19), 8115–8130. <https://doi.org/10.1021/acs.jmedchem.7b00912>.
- (79) Gilson, M. K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A Public Database for Medicinal Chemistry, Computational Chemistry and Systems Pharmacology. *Nucleic Acids Research* **2016**, *44* (D1), D1045–D1053. <https://doi.org/10.1093/nar/gkv1072>.
- (80) Hochreiter, S.; Uergen Schmidhuber, J. “. *Long Short-Term Memory*.
- (81) Dieleman, S.; Schlüter, J.; Raffel, C.; Olson, E.; Sønderby, S. K.; Nouri, D.; Maturana, D.; Thoma, M.; Battenberg, E.; Kelly, J.; Fauw, J. De; Heilman, M.; diogo149; McFee, B.; Weideman, H.; takacsg84; peterderivaz; Jon; instagibbs; Rasul, Dr. K.; CongLiu; Britefury; Degrave, J. Lasagne: First Release. **2015**. <https://doi.org/10.5281/ZENODO.27878>.
- (82) The Theano Development Team, T. T. D.; Al-Rfou, R.; Alain, G.; Almahairi, A.; Angermueller, C.; Bahdanau, D.; Ballas, N.; Bastien, F.; Bayer, J.; Belikov, A.; Belopolsky, A.; Bengio, Y.; Bergeron, A.; Bergstra, J.; Bisson, V.; Snyder, J. B.; Bouchard, N.; Boulanger-Lewandowski, N.; Bouthillier, X.; de Brébisson, A.; Breuleux, O.; Carrier, P.-L.; Cho, K.; Chorowski, J.; Christiano, P.; Cooijmans, T.; Côté, M.-A.; Côté, M.; Courville, A.; Dauphin, Y. N.; Delalleau, O.; Demouth, J.; Desjardins, G.; Dieleman, S.; Dinh, L.; Ducoffe, M.; Dumoulin, V.; Kahou, S. E.; Erhan, D.; Fan, Z.; Firat, O.; Germain, M.; Glorot, X.; Goodfellow, I.; Graham, M.; Gulcehre, C.; Hamel, P.; Harlouchet, I.; Heng, J.-P.; Hidasi, B.; Honari, S.; Jain, A.; Jean, S.; Jia, K.; Korobov, M.; Kulkarni, V.; Lamb, A.; Lamblin, P.; Larsen, E.; Laurent, C.; Lee, S.; Lefrancois, S.; Lemieux, S.; Léonard, N.; Lin, Z.; Livezey, J. A.; Lorenz, C.; Lowin, J.; Ma, Q.; Manzagol, P.-A.; Mastropietro, O.; McGibbon, R. T.; Memisevic, R.; van Merriënboer, B.; Michalski, V.; Mirza, M.; Orlandi, A.; Pal, C.; Pascanu, R.; Pezeshki, M.; Raffel, C.; Renshaw, D.; Rocklin, M.; Romero, A.; Roth, M.; Sadowski, P.; Salvatier, J.; Savard, F.; Schlüter, J.; Schulman, J.; Schwartz, G.; Serban, I. V.; Serdyuk, D.; Shabanian, S.; Simon, É.; Spieckermann, S.; Subramanyam, S. R.; Sygnowski, J.; Tanguay, J.; van Tulder, G.; Turian, J.; Urban, S.; Vincent, P.; Visin, F.; de Vries, H.; Warde-Farley, D.; Webb, D. J.; Willson, M.; Xu, K.; Xue, L.; Yao, L.;

- Zhang, S.; Zhang, Y. Theano: A Python Framework for Fast Computation of Mathematical Expressions. **2016**.
- (83) Landrum G. RDKit: Open-Source Cheminformatics. p <https://www.rdkit.org>.
- (84) Duchi, J.; Hazan, E.; Singer, Y. Adaptive Subgradient Methods for Online Learning and Stochastic Optimization. *Journal of machine learning research* **2011**, *12* (7).
- (85) Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A. K.; Kovalishyn, V. V.; Prokopenko, V. V.; Tetko, I. V. Applicability Domain for in Silico Models to Achieve Accuracy of Experimental Measurements. *Journal of Chemometrics* **2010**, *24* (3-4), 202–208. <https://doi.org/https://doi.org/10.1002/cem.1296>.
- (86) Bjerrum, E. J. SMILES Enumeration as Data Augmentation for Neural Network Modeling of Molecules. **2017**.
- (87) Tetko IV, Karpov P, Bruno E, Kimber TB, G. G. Augmentation Is What You Need! In *Tetko IV, Kůrková V, Karpov P, Theis F (eds) 28th international conference on artificial neural networks, Munich, Germany, September 17–19, 2019, proceedings. Artificial neural networks and machine learning—ICANN 2019: Workshop and special sessions*; Springer, Cham; pp 831–835.
- (88) F. Bruns, R.; A. Watson, I. Rules for Identifying Potentially Reactive or Promiscuous Compounds. *Journal of Medicinal Chemistry* **2012**, *55* (22), 9763–9772. <https://doi.org/10.1021/jm301008n>.
- (89) Dixon, S. L.; Smondirev, A. M.; Rao, S. N. PHASE: A Novel Approach to Pharmacophore Modeling and 3D Database Searching. *Chemical Biology & Drug Design* **2006**, *67* (5), 370–372. <https://doi.org/https://doi.org/10.1111/j.1747-0285.2006.00384.x>.
- (90) Dixon, S. L.; Smondirev, A. M.; Knoll, E. H.; Rao, S. N.; Shaw, D. E.; Friesner, R. A. PHASE: A New Engine for Pharmacophore Perception, 3D QSAR Model Development, and 3D Database Screening: 1. Methodology and Preliminary Results. *Journal of Computer-Aided Molecular Design* **2006**, *20* (10), 647–671. <https://doi.org/10.1007/s10822-006-9087-6>.
- (91) D.A. Case, K. Belfon, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, G. Giambasu, M.K. Gilson, H. Gohlke, A.W. Goetz, R. Harris, S. Izadi, S.A. Izmailov, K. Kasavajhala, A. Kovalenko, R. Krasny, T, D. M. Y. and P. A. K. AMBERTOOLS2018, University of California, San Francisco. 2018.
- (92) Hou, T.; Wang, J.; Li, Y.; Wang, W. Assessing the Performance of the MM/PBSA and MM/GBSA Methods. 1. The Accuracy of Binding Free Energy Calculations Based on Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **2010**, *51* (1), 69–82. <https://doi.org/10.1021/ci100275a>.
- (93) V. Tetko, I.; J. Livingstone, D.; I. Luik, A. Neural Network Studies. 1. Comparison of Overfitting and Overtraining. *Journal of Chemical Information and Computer Sciences* **2002**, *35* (5), 826–833. <https://doi.org/10.1021/ci00027a006>.
- (94) Karlov, D. S.; Sosnin, S.; Tetko, I. V.; Fedorov, M. V. Chemical Space Exploration Guided by Deep Neural Networks. *RSC Advances* **2019**, *9* (9), 5151–5157. <https://doi.org/10.1039/C8RA10182E>.
- (95) Chen, S.-K.; Zhao, P.; Shao, Y.-X.; Li, Z.; Zhang, C.; Liu, P.; He, X.; Luo, H.-B.; Hu, X. Moracin M from *Morus Alba* L. Is a Natural Phosphodiesterase-4 Inhibitor. *Bioorganic &*

- Medicinal Chemistry Letters* **2012**, *22* (9), 3261–3264.  
<https://doi.org/https://doi.org/10.1016/j.bmcl.2012.03.026>.
- (96) Kallen, J.; Izaac, A.; Chau, S.; Wirth, E.; Schoepfer, J.; Mah, R.; Schlapbach, A.; Stutz, S.; Vaupel, A.; Guagnano, V.; Masuya, K.; Stachyra, T.-M.; Salem, B.; Chene, P.; Gessier, F.; Holzer, P.; Furet, P. Structural States of Hdm2 and HdmX: X-Ray Elucidation of Adaptations and Binding Interactions for Different Chemical Compound Classes. *ChemMedChem* **2019**, *14* (14), 1305–1314.  
<https://doi.org/https://doi.org/10.1002/cmdc.201900201>.
- (97) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem in 2021: New Data Content and Improved Web Interfaces. *Nucleic Acids Research* **2021**, *49* (D1), D1388–D1395. <https://doi.org/10.1093/nar/gkaa971>.
- (98) W. Bemis, G.; A. Murcko, M. The Properties of Known Drugs. 1. Molecular Frameworks. *Journal of Medicinal Chemistry* **1996**, *39* (15), 2887–2893.  
<https://doi.org/10.1021/jm9602928>.
- (99) Xia, Z.; Karpov, P.; Popowicz, G.; Tetko, I. V. Focused Library Generator: Case of Mdmx Inhibitors. *Journal of Computer-Aided Molecular Design* **2020**, *34* (7), 769–782.  
<https://doi.org/10.1007/s10822-019-00242-8>.
- (100) Nagar, B.; Bornmann, W. G.; Pellicena, P.; Schindler, T.; Veach, D. R.; Miller, W. T.; Clarkson, B.; Kuriyan, J. Crystal Structures of the Kinase Domain of C-Abl in Complex with the Small Molecule Inhibitors PD173955 and Imatinib (STI-571). *Cancer Research* **2002**, *62* (15), 4236 LP – 4243.
- (101) Schindler, T.; Bornmann, W.; Pellicena, P.; Miller, W. T.; Clarkson, B.; Kuriyan, J. Structural Mechanism for STI-571 Inhibition of Abelson Tyrosine Kinase. *Science* **2000**, *289* (5486), 1938 LP – 1942. <https://doi.org/10.1126/science.289.5486.1938>.
- (102) Wodicka, L. M.; Ciceri, P.; Davis, M. I.; Hunt, J. P.; Floyd, M.; Salerno, S.; Hua, X. H.; Ford, J. M.; Armstrong, R. C.; Zarrinkar, P. P.; Treiber, D. K. Activation State-Dependent Binding of Small Molecule Kinase Inhibitors: Structural Insights from Biochemistry. *Chemistry & Biology* **2010**, *17* (11), 1241–1249.  
<https://doi.org/10.1016/J.CHEMBIOL.2010.09.010>.
- (103) Umezawa, K.; Kii, I. Druggable Transient Pockets in Protein Kinases. *Molecules* **2021**, *26* (3). <https://doi.org/10.3390/molecules26030651>.
- (104) Vajda, S.; Beglov, D.; Wakefield, A. E.; Egbert, M.; Whitty, A. Cryptic Binding Sites on Proteins: Definition, Detection, and Druggability. *Current Opinion in Chemical Biology* **2018**, *44*, 1–8. <https://doi.org/10.1016/J.CBPA.2018.05.003>.
- (105) Mizukoshi, Y.; Takeuchi, K.; Tokunaga, Y.; Matsuo, H.; Imai, M.; Fujisaki, M.; Kamoshida, H.; Takizawa, T.; Hanzawa, H.; Shimada, I. Targeting the Cryptic Sites: NMR-Based Strategy to Improve Protein Druggability by Controlling the Conformational Equilibrium. *Science Advances* **2020**, *6* (40).  
[https://doi.org/10.1126/SCIADV.ABD0480/SUPPL\\_FILE/ABD0480\\_SM.PDF](https://doi.org/10.1126/SCIADV.ABD0480/SUPPL_FILE/ABD0480_SM.PDF).
- (106) Kii, I.; Sumida, Y.; Goto, T.; Sonamoto, R.; Okuno, Y.; Yoshida, S.; Kato-Sumida, T.; Koike, Y.; Abe, M.; Nonaka, Y.; Ikura, T.; Ito, N.; Shibuya, H.; Hosoya, T.; Hagiwara, M.

- Selective Inhibition of the Kinase DYRK1A by Targeting Its Folding Process. *Nature Communications* 2016 7:1 **2016**, 7 (1), 1–14. <https://doi.org/10.1038/ncomms11391>.
- (107) Cimermancic, P.; Weinkam, P.; Rettenmaier, T. J.; Bichmann, L.; Keedy, D. A.; Woldeyes, R. A.; Schneidman-Duhovny, D.; Demerdash, O. N.; Mitchell, J. C.; Wells, J. A.; Fraser, J. S.; Sali, A. CryptoSite: Expanding the Druggable Proteome by Characterization and Prediction of Cryptic Binding Sites. *Journal of Molecular Biology* **2016**, 428 (4), 709–719. <https://doi.org/10.1016/J.JMB.2016.01.029>.
- (108) Beglov, D.; Hall, D. R.; Wakefield, A. E.; Luo, L.; Allen, K. N.; Kozakov, D.; Whitty, A.; Vajda, S. Exploring the Structural Origins of Cryptic Sites on Proteins. *Proceedings of the National Academy of Sciences* **2018**, 115 (15), E3416–E3425. <https://doi.org/10.1073/PNAS.1711490115>.
- (109) Kozakov, D.; Grove, L. E.; Hall, D. R.; Bohnuud, T.; Mottarella, S. E.; Luo, L.; Xia, B.; Beglov, D.; Vajda, S. The FTMap Family of Web Servers for Determining and Characterizing Ligand-Binding Hot Spots of Proteins. *Nature Protocols* 2015 10:5 **2015**, 10 (5), 733–755. <https://doi.org/10.1038/nprot.2015.043>.
- (110) Clark, J. J.; Benson, M. L.; Smith, R. D.; Carlson, H. A. Inherent versus Induced Protein Flexibility: Comparisons within and between Apo and Holo Structures. *PLOS Computational Biology* **2019**, 15 (1), e1006705.
- (111) Evans, D. J.; Yovanno, R. A.; Rahman, S.; Cao, D. W.; Beckett, M. Q.; Patel, M. H.; Bandak, A. F.; Lau, A. Y. Finding Druggable Sites in Proteins Using TACTICS. *Journal of Chemical Information and Modeling* **2021**, 61 (6), 2897–2910. [https://doi.org/10.1021/ACS.JCIM.1C00204/SUPPL\\_FILE/CI1C00204\\_SI\\_001.PDF](https://doi.org/10.1021/ACS.JCIM.1C00204/SUPPL_FILE/CI1C00204_SI_001.PDF).
- (112) Kuzmanic, A.; Bowman, G. R.; Juarez-Jimenez, J.; Michel, J.; Gervasio, F. L. Investigating Cryptic Binding Sites by Molecular Dynamics Simulations. *Accounts of Chemical Research* **2020**. <https://doi.org/10.1021/ACS.ACCOUNTS.9B00613>.
- (113) Wang, R.; Fang, X.; Lu, Y.; Wang, S. The PDBbind Database: Collection of Binding Affinities for Protein–Ligand Complexes with Known Three-Dimensional Structures. *Journal of Medicinal Chemistry* **2004**, 47 (12), 2977–2980. <https://doi.org/10.1021/JM030580L>.
- (114) Westbrook, J. D.; Shao, C.; Feng, Z.; Zhuravleva, M.; Velankar, S.; Young, J. The Chemical Component Dictionary: Complete Descriptions of Constituent Molecules in Experimentally Determined 3D Macromolecules in the Protein Data Bank. *Bioinformatics* **2015**, 31 (8), 1274–1278. <https://doi.org/10.1093/BIOINFORMATICS/BTU789>.
- (115) Clark, J. J.; Benson, M. L.; Smith, R. D.; Carlson, H. A. Inherent versus Induced Protein Flexibility: Comparisons within and between Apo and Holo Structures. *PLoS Computational Biology* **2019**, 15 (1). <https://doi.org/10.1371/JOURNAL.PCBI.1006705>.
- (116) Altschul, S. F.; Madden, T. L.; Schäffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. Gapped BLAST and PSI-BLAST: A New Generation of Protein Database Search Programs. *Nucleic Acids Research* **1997**, 25 (17), 3389–3402. <https://doi.org/10.1093/NAR/25.17.3389>.
- (117) Schrödinger Release 2020-3. Schrödinger, LLC, New York, NY, 2020.

- (118) Halgren, T. New Method for Fast and Accurate Binding-Site Identification and Analysis. *Chemical Biology & Drug Design* **2007**, *69* (2), 146–148.  
<https://doi.org/https://doi.org/10.1111/j.1747-0285.2007.00483.x>.
- (119) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *Journal of Chemical Information and Modeling* **2009**, *49* (2), 377–389.  
<https://doi.org/10.1021/ci800324m>.
- (120) Réau, M.; Lagarde, N.; Zagury, J. F.; Montes, M. Nuclear Receptors Database Including Negative Data (NR-DBIND): A Database Dedicated to Nuclear Receptors Binding Data Including Negative Data and Pharmacological Profile. *Journal of Medicinal Chemistry* **2019**, *62* (6), 2894–2904.  
[https://doi.org/10.1021/ACS.JMEDCHEM.8B01105/SUPPL\\_FILE/JM8B01105\\_SI\\_001.PDF](https://doi.org/10.1021/ACS.JMEDCHEM.8B01105/SUPPL_FILE/JM8B01105_SI_001.PDF).
- (121) Madhavi Sastry, G.; Adzhigirey, M.; Day, T.; Annabhimoju, R.; Sherman, W. Protein and Ligand Preparation: Parameters, Protocols, and Influence on Virtual Screening Enrichments. *Journal of Computer-Aided Molecular Design* **2013**, *27* (3), 221–234.  
<https://doi.org/10.1007/S10822-013-9644-8>.
- (122) Shelley, J. C.; Cholleti, A.; Frye, L. L.; Greenwood, J. R.; Timlin, M. R.; Uchimaya, M. Epik: A Software Program for PK prediction and Protonation State Generation for Drug-like Molecules. *Journal of Computer-Aided Molecular Design* **2007**, *21* (12), 681–691.  
<https://doi.org/10.1007/s10822-007-9133-z>.
- (123) Sushko, I.; Novotarskyi, S.; Körner, R.; Pandey, A. K.; Rupp, M.; Teetz, W.; Brandmaier, S.; Abdelaziz, A.; Prokopenko, V. v.; Tanchuk, V. Y.; Todeschini, R.; Varnek, A.; Marcou, G.; Ertl, P.; Potemkin, V.; Grishina, M.; Gasteiger, J.; Schwab, C.; Baskin, I. I.; Palyulin, V. A.; Radchenko, E. v.; Welsh, W. J.; Kholodovych, V.; Chekmarev, D.; Cherkasov, A.; Aires-De-Sousa, J.; Zhang, Q. Y.; Bender, A.; Nigsch, F.; Patiny, L.; Williams, A.; Tkachenko, V.; Tetko, I. v. Online Chemical Modeling Environment (OCHEM): Web Platform for Data Storage, Model Development and Publishing of Chemical Information. *Journal of Computer-Aided Molecular Design* **2011**, *25* (6), 533–554.  
<https://doi.org/10.1007/S10822-011-9440-2/FIGURES/12>.
- (124) Breiman, L. *Random Forests*; 2001; Vol. 45.
- (125) Svetnik, V.; Liaw, A.; Tong, C.; Christopher Culberson, J.; P. Sheridan, R.; P. Feuston, B. Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. *Journal of Chemical Information and Computer Sciences* **2003**, *43* (6), 1947–1958. <https://doi.org/10.1021/ci034160g>.
- (126) Oshiro, T. M.; Perez, P. S.; Baranauskas, J. A. How Many Trees in a Random Forest? In *Machine Learning and Data Mining in Pattern Recognition*; Perner, P., Ed.; Springer Berlin Heidelberg: Berlin, Heidelberg, 2012; pp 154–168.
- (127) v. Tetko, I.; Yu. Tanchuk, V.; E. P. Villa, A. Prediction of N-Octanol/Water Partition Coefficients from PHYSPROP Database Using Artificial Neural Networks and E-State Indices. *Journal of Chemical Information and Computer Sciences* **2001**, *41* (5), 1407–1421.  
<https://doi.org/10.1021/ci010368v>.



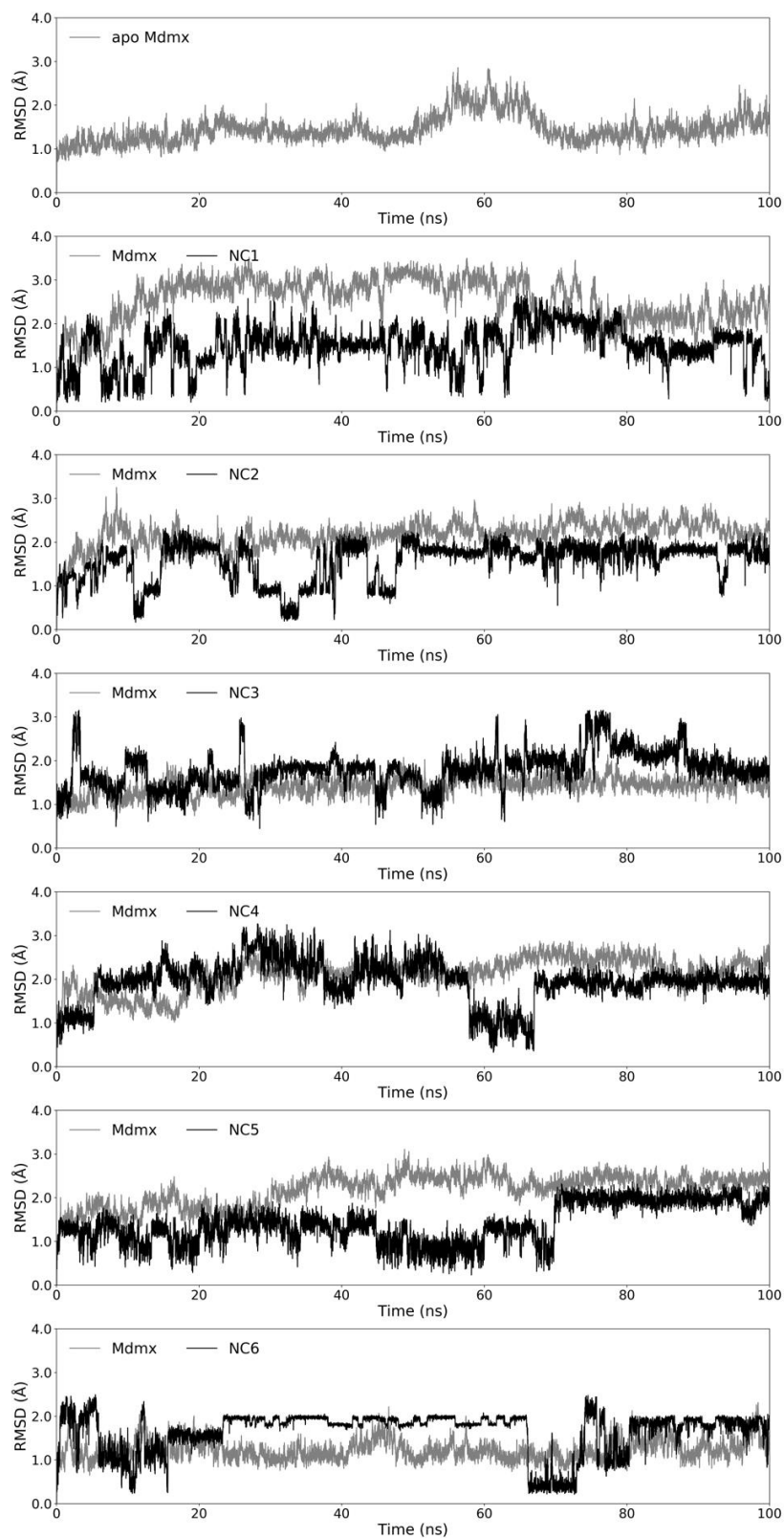
- (128) v. Tetko, I.; Yu. Tanchuk, V.; N. Kasheva, T.; E. P. Villa, A. Estimation of Aqueous Solubility of Chemical Compounds Using E-State Indices. *Journal of Chemical Information and Computer Sciences* **2001**, *41* (6), 1488–1493. <https://doi.org/10.1021/ci000392t>.
- (129) v. Tetko, I.; Yu. Tanchuk, V. Application of Associative Neural Networks for Prediction of Lipophilicity in ALOGPS 2.1 Program. *Journal of Chemical Information and Computer Sciences* **2002**, *42* (5), 1136–1145. <https://doi.org/10.1021/ci025515j>.
- (130) H. Hall, L.; B. Kier, L. Electrotopological State Indices for Atom Types: A Novel Combination of Electronic, Topological, and Valence State Information. *Journal of Chemical Information and Computer Sciences* **2002**, *35* (6), 1039–1045. <https://doi.org/10.1021/ci00028a014>.
- (131) Kier, L. B.; Hall, L. H. An Electrotopological-State Index for Atoms in Molecules. *Pharmaceutical Research* **1990**, *7* (8), 801–807. <https://doi.org/10.1023/A:1015952613760>.
- (132) Kier, L. B.; Hall, L. H. *Molecular Structure Description: The Electrotopological State*; Elsevier Science, 1999.
- (133) Timmerman, H.; Vacca, J.; van de Waterbeemd, H.; Wieland, T.; Smith, D. A.; Walker, D. K.; Ed, nd; Ottow, E.; Weinmann, H.; Testa, B. *Methods and Principles in Medicinal Chemistry Previous Volumes of This Series: Pharmacokinetics and Metabolism in Drug Design, Pharmacophores and Pharmacophore Searches Chirality in Drug Research Fragment-Based Approaches in Drug Discovery High-Throughput Screening in Drug Discovery Mass Spectrometry in Medicinal Chemistry Molecular Drug Properties Nuclear Receptors as Drug Targets*.
- (134) Shapley, L. A Value of  $n$ -Person Games. *Ann. Math. Study* **28**, *Contributions to the Theory of Games*, ed. by HW Kuhn, and AW Tucker **1953**, 307–317.
- (135) Lundberg, S. M.; Lee, S.-I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems*; Guyon, I., Luxburg, U. v., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc., 2017; Vol. 30.
- (136) Ribeiro, M. T.; Singh, S.; Guestrin, C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*; KDD '16; Association for Computing Machinery: New York, NY, USA, 2016; pp 1135–1144. <https://doi.org/10.1145/2939672.2939778>.
- (137) Shrikumar, A.; Greenside, P.; Kundaje, A. Learning Important Features Through Propagating Activation Differences. 2019.
- (138) Bach, S.; Binder, A.; Montavon, G.; Klauschen, F.; Müller, K.-R.; Samek, W. On Pixel-Wise Explanations for Non-Linear Classifier Decisions by Layer-Wise Relevance Propagation. *PLOS ONE* **2015**, *10* (7), e0130140. <https://doi.org/10.1371/journal.pone.0130140>.
- (139) Lipovetsky, S.; Conklin, M. Analysis of Regression in Game Theory Approach. *Applied Stochastic Models in Business and Industry* **2001**, *17* (4), 319–330. <https://doi.org/https://doi.org/10.1002/asmb.446>.
- (140) Štrumbelj, E.; Kononenko, I. Explaining Prediction Models and Individual Predictions with Feature Contributions. *Knowledge and Information Systems* **2014**, *41* (3), 647–665. <https://doi.org/10.1007/s10115-013-0679-x>.

- (141) Datta, A.; Sen, S.; Zick, Y. Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems. *2016 IEEE Symposium on Security and Privacy (SP)* **2016**, 598–617.
- (142) Lundberg, S. M.; Erion, G.; Chen, H.; DeGrave, A.; Prutkin, J. M.; Nair, B.; Katz, R.; Himmelfarb, J.; Bansal, N.; Lee, S.-I. From Local Explanations to Global Understanding with Explainable AI for Trees. *Nature Machine Intelligence* **2020**, 2 (1), 56–67. <https://doi.org/10.1038/s42256-019-0138-9>.
- (143) Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E. Scikit-Learn: Machine Learning in Python. *Journal of Machine Learning Research* **2011**, 12, 2825–2830.
- (144) Rogers, D.; Hahn, M. Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling* **2010**, 50 (5), 742–754. <https://doi.org/10.1021/ci100050t>.
- (145) Willett, P. Similarity-Based Virtual Screening Using 2D Fingerprints. *Drug Discovery Today* **2006**, 11 (23–24), 1046–1053. <https://doi.org/10.1016/J.DRUDIS.2006.10.005>.
- (146) Vorberg, S.; Tetko, I. v. Modeling the Biodegradability of Chemical Compounds Using the Online CHEmical Modeling Environment (OCHEM). *Molecular Informatics* **2014**, 33 (1), 73–85. <https://doi.org/https://doi.org/10.1002/minf.201300030>.
- (147) Tetko, I. v.; Lowe, D.; Williams, A. J. The Development of Models to Predict Melting and Pyrolysis Point Data Associated with Several Hundred Thousand Compounds Mined from PATENTS. *Journal of Cheminformatics* **2016**, 8 (1), 2. <https://doi.org/10.1186/s13321-016-0113-y>.
- (148) Tetko, I. v.; Novotarskyi, S.; Sushko, I.; Ivanov, V.; Petrenko, A. E.; Dieden, R.; Lebon, F.; Mathieu, B. Development of Dimethyl Sulfoxide Solubility Models Using 163 000 Molecules: Using a Domain Applicability Metric to Select More Reliable Predictions. *Journal of Chemical Information and Modeling* **2013**, 53 (8), 1990–2000. <https://doi.org/10.1021/CI400213D>.
- (149) Wilcken, R.; Liu, X.; O. Zimmermann, M.; J. Rutherford, T.; R. Fersht, A.; C. Joerger, A.; M. Boeckler, F. Halogen-Enriched Fragment Libraries as Leads for Drug Rescue of Mutant P53. *Journal of the American Chemical Society* **2012**, 134 (15), 6810–6818. <https://doi.org/10.1021/ja301056a>.
- (150) G. Davies, T.; E. Wixted, W.; E. Coyle, J.; Griffiths-Jones, C.; Hearn, K.; McMenamin, R.; Norton, D.; J. Rich, S.; Richardson, C.; Saxty, G.; M. G. Willems, H.; J.-A. Woolford, A.; E. Cottom, J.; Kou, J.-P.; G. Yonchuk, J.; G. Feldser, H.; Sanchez, Y.; P. Foley, J.; J. Bolognese, B.; Logan, G.; L. Podolin, P.; Yan, H.; F. Callahan, J.; D. Heightman, T.; K. Kerns, J. Monoacidic Inhibitors of the Kelch-like ECH-Associated Protein 1: Nuclear Factor Erythroid 2-Related Factor 2 (KEAP1:NRF2) Protein–Protein Interaction with High Cell Potency Identified by Fragment-Based Discovery. *Journal of Medicinal Chemistry* **2016**, 59 (8), 3991–4006. <https://doi.org/10.1021/acs.jmedchem.6b00228>.
- (151) Lee, K.; Yang, A.; Lin, Y. C.; Reker, D.; Bernardes, G. J. L.; Rodrigues, T. Combating Small-Molecule Aggregation with Machine Learning. *Cell Reports Physical Science* **2021**, 2 (9), 100573. <https://doi.org/10.1016/J.XCRP.2021.100573>.

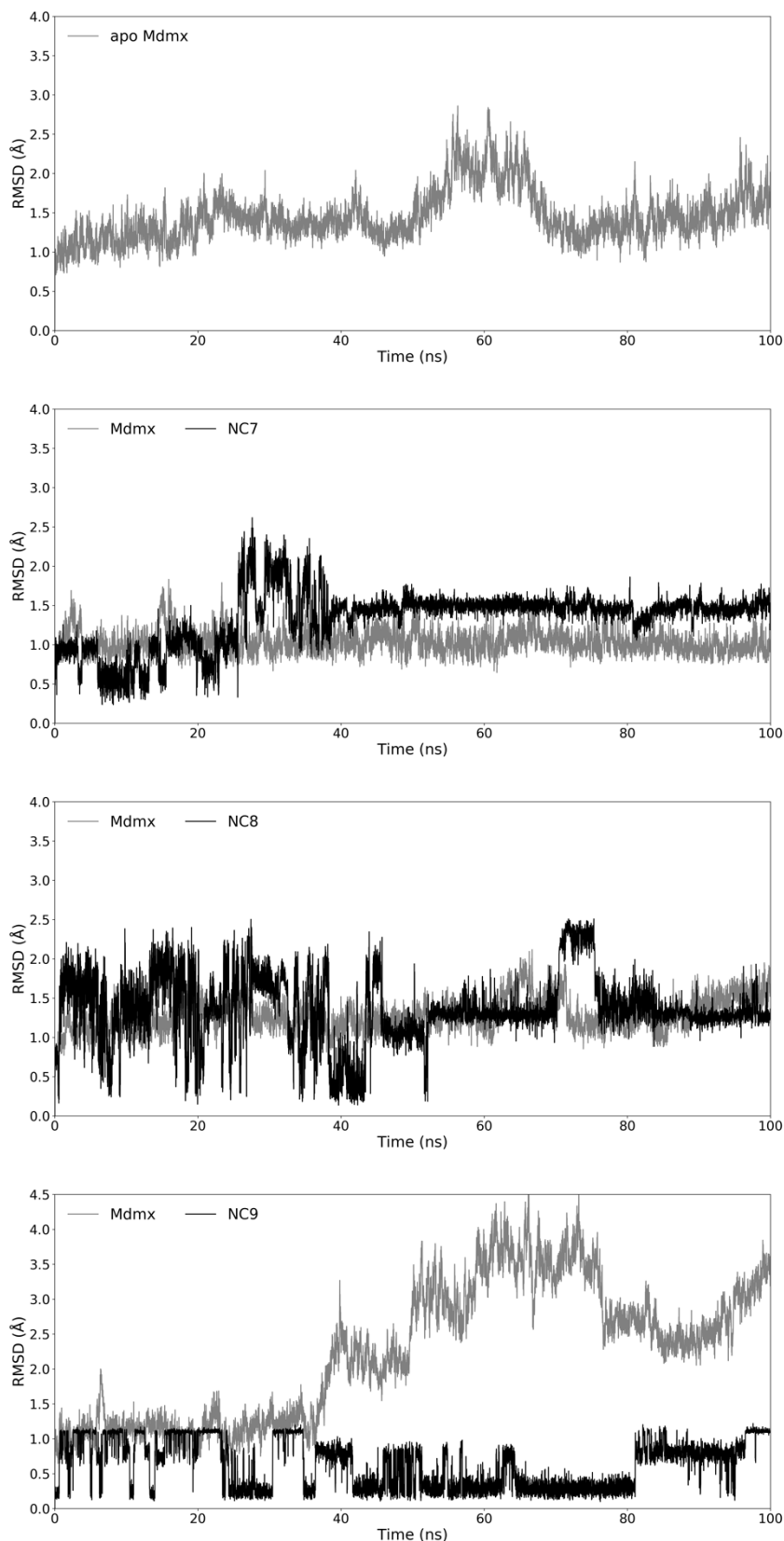
## Appendix

- I. The RMSD plots of NC series compounds (Chapter 4)
- II. Evaluation of inhibitory activity of SN compounds towards Mdm2 (Chapter 4)
- III. Supporting information of the Generator (Chapter 5)
- IV. Interpretation of models & Analysis on pocket residues (Chapter 6)

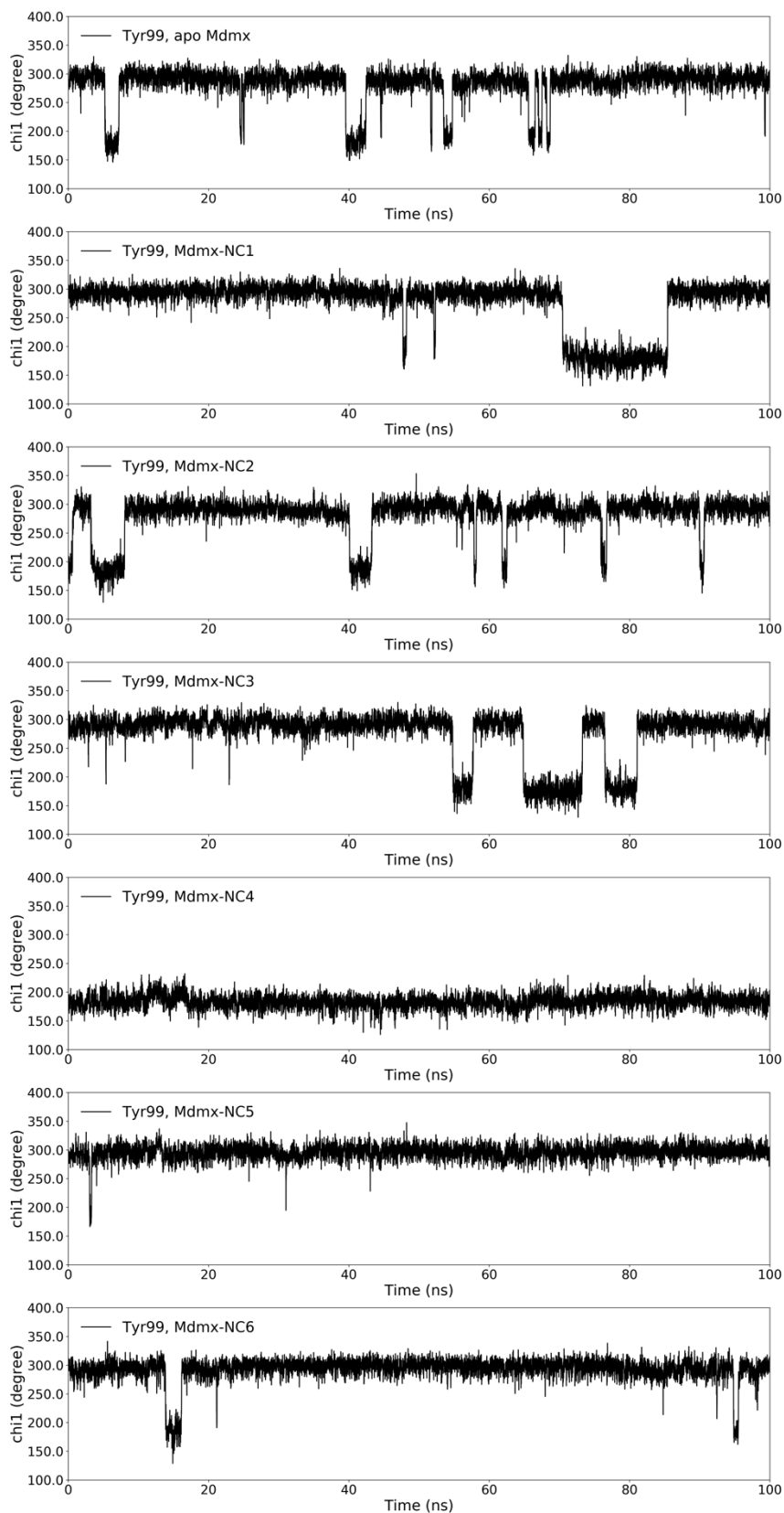
## Appendix I



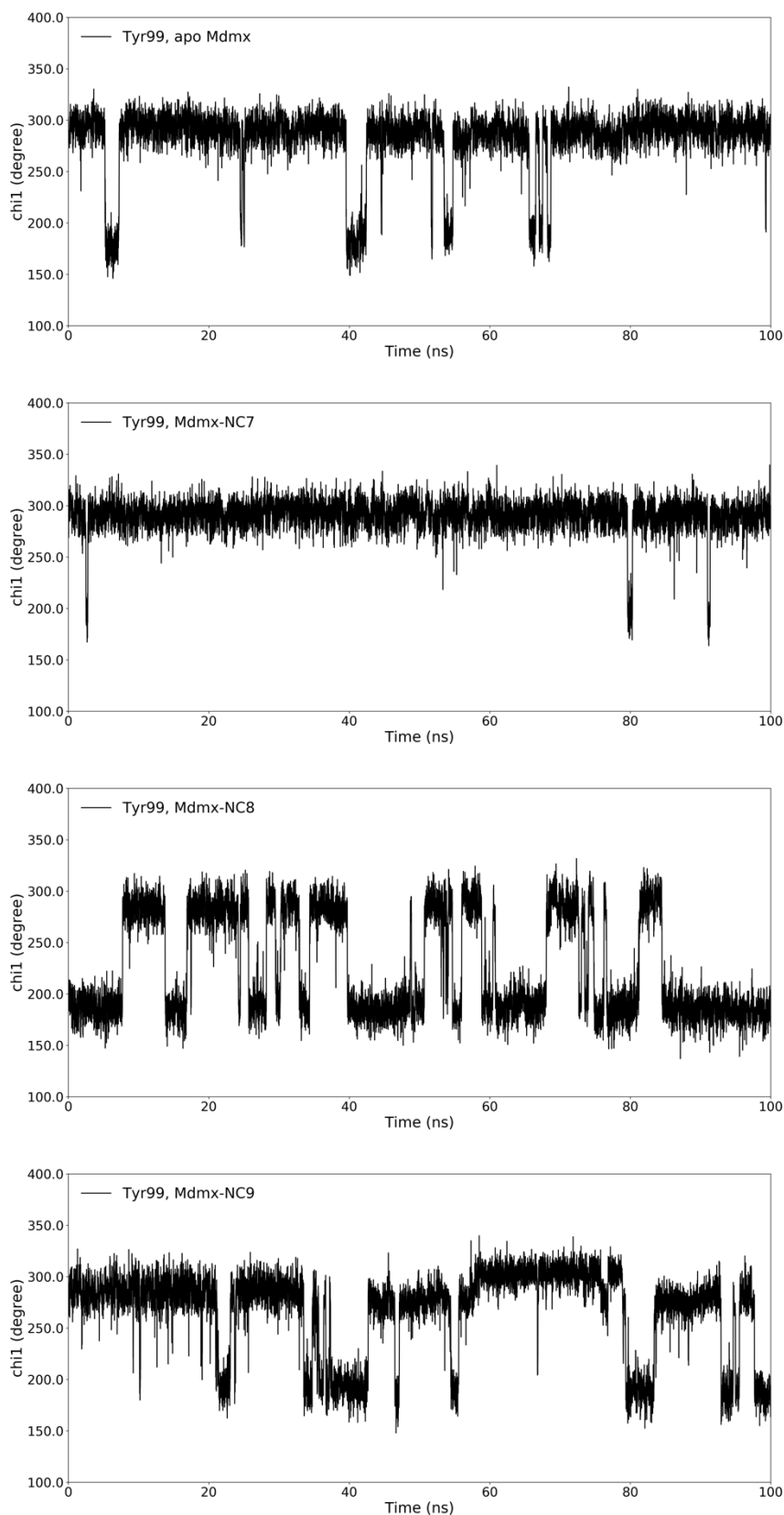
**Figure S 1** The RMSDs (backbone atoms) of Mdmx complexes with NC (1 - 6) compounds as a function of time were plotted through the 100-ns trajectories, each of which was calculated with the first frame as the reference.



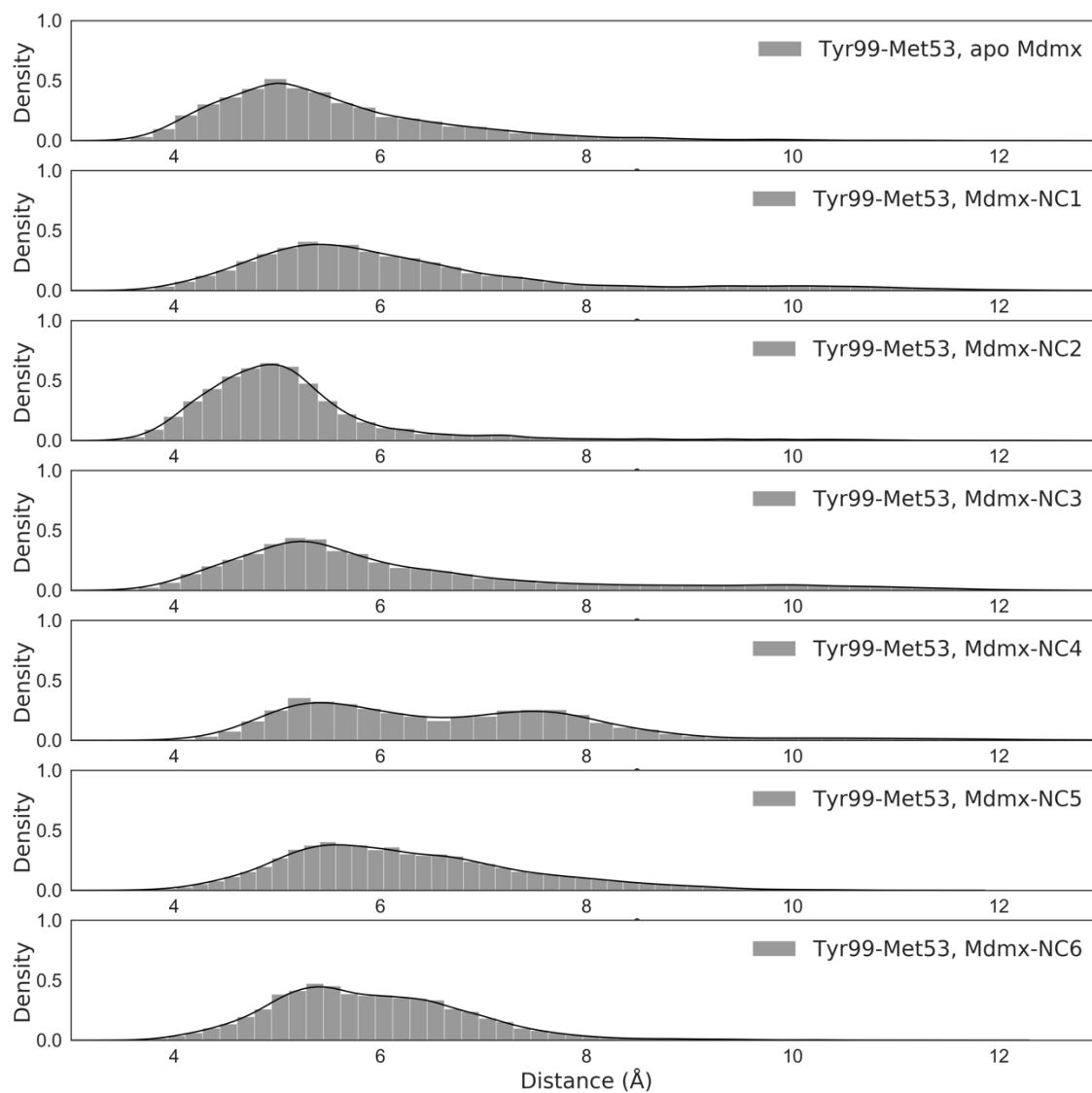
**Figure S 2** The RMSDs (backbone atoms) of Mdmx complexes with NC (7 - 9) compounds as a function of time were plotted through the 100-ns trajectories, each of which was calculated with the first frame as the reference.



**Figure S 3** The dynamic change of Tyr99  $\chi_1$  in Mdmx complexes with NC (1 - 6) compounds through the 100-ns trajectories.

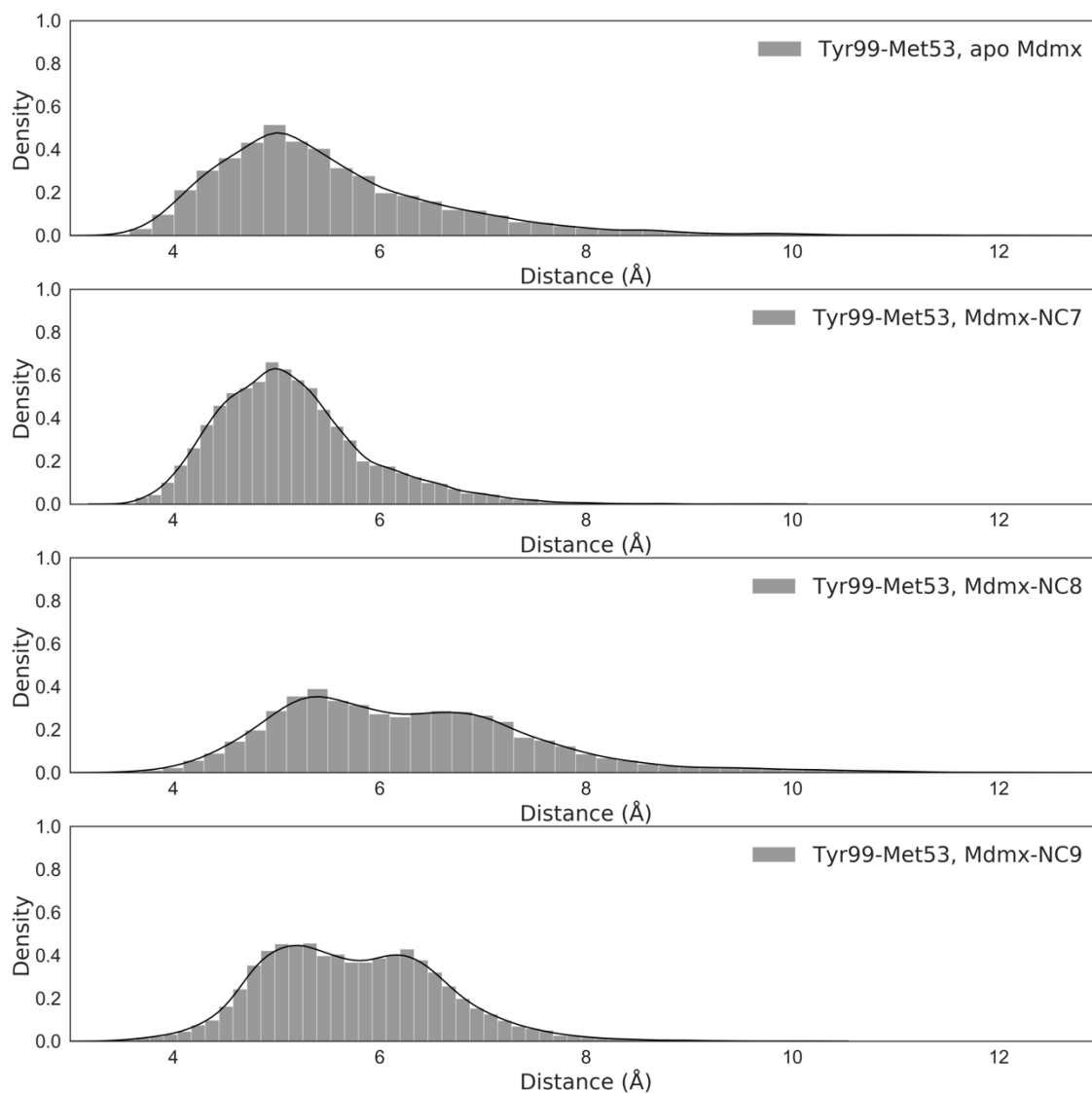


**Figure S 4** The dynamic change of Tyr99  $\chi_1$  in Mdmx complexes with NC (7 - 9) compounds through the 100-ns trajectories.



**Figure S 5** The distribution of distance between Tyr99 and Met53 in Mdmx complexes with NC (1 - 6) compounds through the 100-ns trajectories.





**Figure S 6** The distribution of distance between Tyr99 and Met53 in Mdmx complexes with NC (7 - 9) compounds through the 100-ns trajectories.

## Appendix II

Evaluation of inhibitory activity of SN compounds towards Mdm2

(Cited from the collaborator: Beata Łabuzek)

### 1. Determination of dissociation constant of labelled peptide P2 and Mdm2

New protein stock (constructs: Mdm2 1-118) was thawed and the concentration was measured using Bradford method. Dissociation constant ( $K_d$ ) of labelled peptide P2 – Mdm2 was determined for further evaluation of optimal protein concentration for inhibition constant determination. Peptide P2 was labelled with carboxyfluorescein and had the following sequence: LTFEHYWAQLTS.  $K_d$  was determined for each thawed new protein stock using fluorescence polarization (FP) assay.

FP assay was performed in duplicates on 96-well microplates (Corning NBS 3991) in final volume 100  $\mu$ l (70  $\mu$ l protein + 5  $\mu$ l DMSO + 25  $\mu$ l P2). Protein dilutions were made in FP buffer (10 mM Tris-HCL pH 8.0, 1 mM EDTA, 50 mM NaCl) to have to concentrations appropriate for  $K_d$  determination: 1,07  $\mu$ M for Mdm2. First column of the Corning NBS 3991 microplate was filled with protein, rest of the columns were filled with FP buffer. Next, serial dilutions of the protein (range from 750 to 0,012 nM final concentration on the plate for Mdm2) were prepared in columns 2 -11, last column (12) contained only FP buffer. The highest protein concentration in first column (0,75  $\mu$ M for Mdm2) corresponded to FP values of the peptide saturated with protein; whereas the FP buffer with no protein corresponded to FP values of the peptide alone. To all columns DMSO (Bioshop DMS555.500) previously distributed on Greiner PP 651201 microwell plate was added to protein dilutions on Corning NBS 3991 plate. After mixing and 15 min incubation at room temperature, 40 nM P2 peptide solution prepared in 15 ml Sarstedt Falcon and was added to every column (1-12) so that the final concentration of P2 peptide was 10 nM. FP measurements were made using BioTek Synergy H1 microplate reader.  $K_d$  was determined by fitting curve of *Equation S1*. to the experimental data.

*Equation S1.*

$$y = y_0 \frac{ax}{K_d + x}$$

where  $y_0$ :  $FP_{\min}$

a:  $FP_{\max} - FP_{\min}$

x: protein concentration

y: FP value measured at the desired concentration.

### 2. Preparation of stock solutions

The SN compounds were dissolved in DMSO (Sigma Aldrich 175943-10G 08828EJ) to obtain 50 mM concentration. Provided compounds were fully dissolved at room temperature at 50 mM and gave a transparent (SN1024) or slightly yellowish solutions (SN1025, SN1020, SN1021).

### 3. Determination of the inhibition constants of SN compounds

Optimal protein concentration for the measurement ( $f_0=0.8$ ) was calculated based on the determined  $K_d$  value (see point 1). FP assay was made using 96-well microplates (Corning NBS 3991) and final volumes of 100  $\mu$ l (70  $\mu$ l protein + 5  $\mu$ l inhibitor + 25  $\mu$ l P2).

Serial dilutions in DMSO (Bioshop DMS555.500) of the tested compounds were prepared in duplicates on 96-well Greiner 651201 microplates (wells A2-H12). The dilutions ranged from 50  $\mu$ M to 0.05  $\mu$ M (final concentrations on the plate). Wells A1-F1 were filled with DMSO to obtain the  $P_{max}$ ,  $P_{min}$  and  $P_{f0}$  values. Wells G1-H1 were controls filled with Nutlin 3 (Cayman Chemicals; control; final concentration 25  $\mu$ M). Protein at the optimal concentration was prepared in 15 ml Sarstedt Falcon and added into A2-H12 and E1- H1 wells on Corning NBS 3991 96-well microplates. Wells A1-B1 contained protein at final concentrations of 0.75  $\mu$ M for Mdm2 in order to determine  $P_{max}$ , whereas wells C1-D1- FP buffer only (to determine  $P_{min}$ ). Next, 5  $\mu$ l of each inhibitor dilution was transferred from Greiner 651201 to Corning NBS plate, mixed and incubated for 15 minutes at room temperature. After incubation, 40 nM P2 peptide solution was added (to final 10 nM concentration of the P2) and FP measurements were made using BioTek Synergy H1 microplate reader. Inhibition constants ( $K_i$ ) were determined by fitting curves into experimental values.

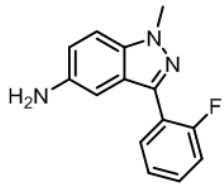
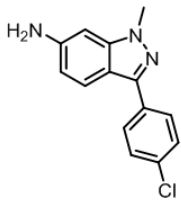
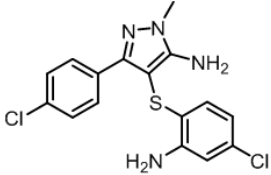
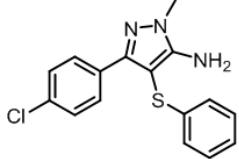
#### 4. Results

The results of the assay of inhibitory activity of tested compound are presented in Table S 1.

#### 5. Summary

Inhibitory activity of SN compounds against Mdm2 was evaluated using FP assay. The compounds were not active towards Mdm2.

**Table S 1 Results of the evaluation of inhibitory activity of SN compounds towards Mdm2.**

No	Compound	Obtained solubility [mM]	Ki Mdm2 [ $\mu$ M]
1	 <p>Chemical Formula: C<sub>14</sub>H<sub>12</sub>FN<sub>3</sub> Molecular Weight: 241.27 <b>SN1024.00</b> pdo1/169</p>	50	not active
2	 <p>Chemical Formula: C<sub>14</sub>H<sub>12</sub>ClN<sub>3</sub> Molecular Weight: 257.72 <b>SN1025.00</b> pdo1/175</p>	50	not active
2	 <p>Chemical Formula: C<sub>16</sub>H<sub>14</sub>Cl<sub>2</sub>N<sub>4</sub>S Molecular Weight: 365.28 <b>SN1020.00</b> pdo1/145</p>	50	not active
4	 <p>Chemical Formula: C<sub>16</sub>H<sub>14</sub>ClN<sub>3</sub>S Molecular Weight: 315.82 <b>SN1021.00</b> pdo1/151</p>	50	not active

## Supporting Information

**Focused Library Generator: Case of Mdmx Inhibitors**Zhonghua Xia<sup>1</sup>, Pavel Karpov<sup>1,2</sup>, Grzegorz Popowicz<sup>1</sup>, and Igor Tetko<sup>1,2</sup><sup>1</sup> Institute of Structural Biology, Helmholtz Zentrum München – Research Center for Environmental Health (GmbH)<sup>2</sup> BigChem GmbH, Germany, München

E-mail: zhonghua.xia, pavel.karpov, grzegorz.popowicz, i.tetko '@' helmholtz-muenchen.de

**Contents**

<b>1 Generator</b>	<b>1</b>
1.1 Temperature adjustment	1
1.2 Dependence of the valid SMILES on the temperature	2
1.3 Learning curves	2
<b>2 Molecular Dynamics validation</b>	<b>3</b>
<b>3 Putative ligands</b>	<b>4</b>
3.1 Top-500 compounds from the Run-1	4
3.2 Top-500 compounds from the Run-2	30

**1 Generator****1.1 Temperature adjustment**

The output of the Generator is a probability distribution of the next symbol over all vocabulary. Technically, this is the result of the softmax activation function.

$$y_i = \frac{e^{z_i}}{\sum e^{z_i}} \quad (1)$$

Here,  $z_i$  stands for logits,  $y_i$  is a normalized probability. To adjust this probability to a different temperature  $T$ , one needs to evaluate:

$$p_i = \frac{e^{\frac{\ln(y_i)}{T}}}{\sum e^{\frac{\ln(y_i)}{T}}} \quad (2)$$

After logarithm of the left and the right parts of the Eq. 1, and then dividing by  $T$ , we get:

$$\frac{\ln(y_i)}{T} = \frac{z_i}{T} - \frac{\ln(\sum e^{z_i})}{T} \quad (3)$$

Then after softmax using  $\frac{\ln(y_i)}{T}$  as logits:

$$e^{\frac{\ln(y_i)}{T}} = \frac{e^{\frac{z_i}{T}}}{\frac{1}{T} \sum e^{z_i}} \quad (4)$$

$$p_i = \frac{e^{\frac{\ln(y_i)}{T}}}{\sum e^{\frac{\ln(y_i)}{T}}} = \frac{e^{\frac{z_i}{T}}}{\frac{1}{T} \sum e^{z_i} \sum \frac{e^{\frac{z_i}{T}}}{\frac{1}{T} \sum e^{z_i}}} = \frac{e^{\frac{z_i}{T}}}{\sum e^{\frac{z_i}{T}}}. \quad (5)$$

## 1.2 Dependence of the valid SMILES on the temperature

At higher temperatures the probability landscape is smoother and the model is getting more erroneous in sampling SMILES. The fig. 1 shows this dependency. At 1.45 the model produces 34% of garbage.

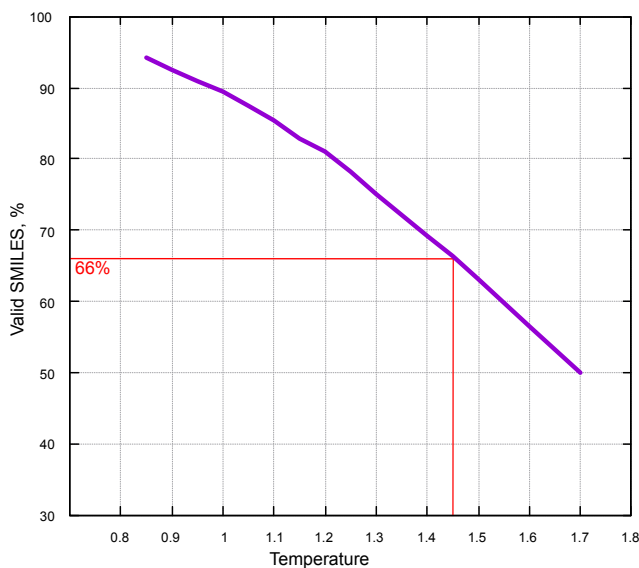


Figure 1: Dependence of the valid SMILES on the temperature.

## 1.3 Learning curves

Fig. 2 shows learning curves obtained during retraining of the Generator. Dash lines (1-6 cycle) have decrease of the validation loss during first epochs of training, whereas solid lines (8-10 cycle) show overfitting.

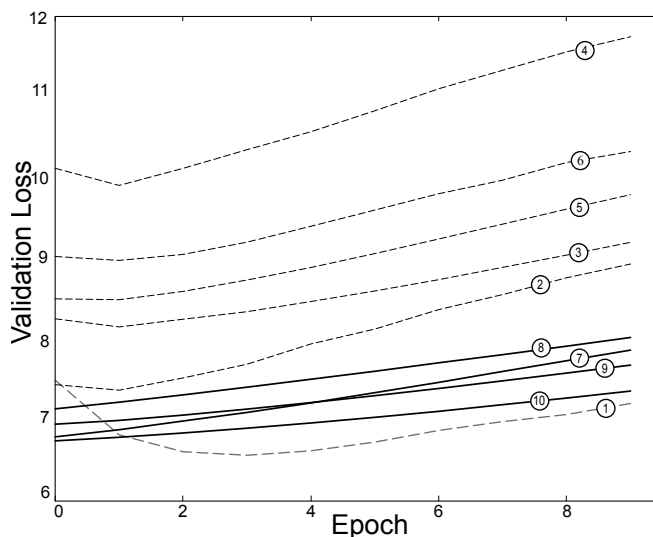


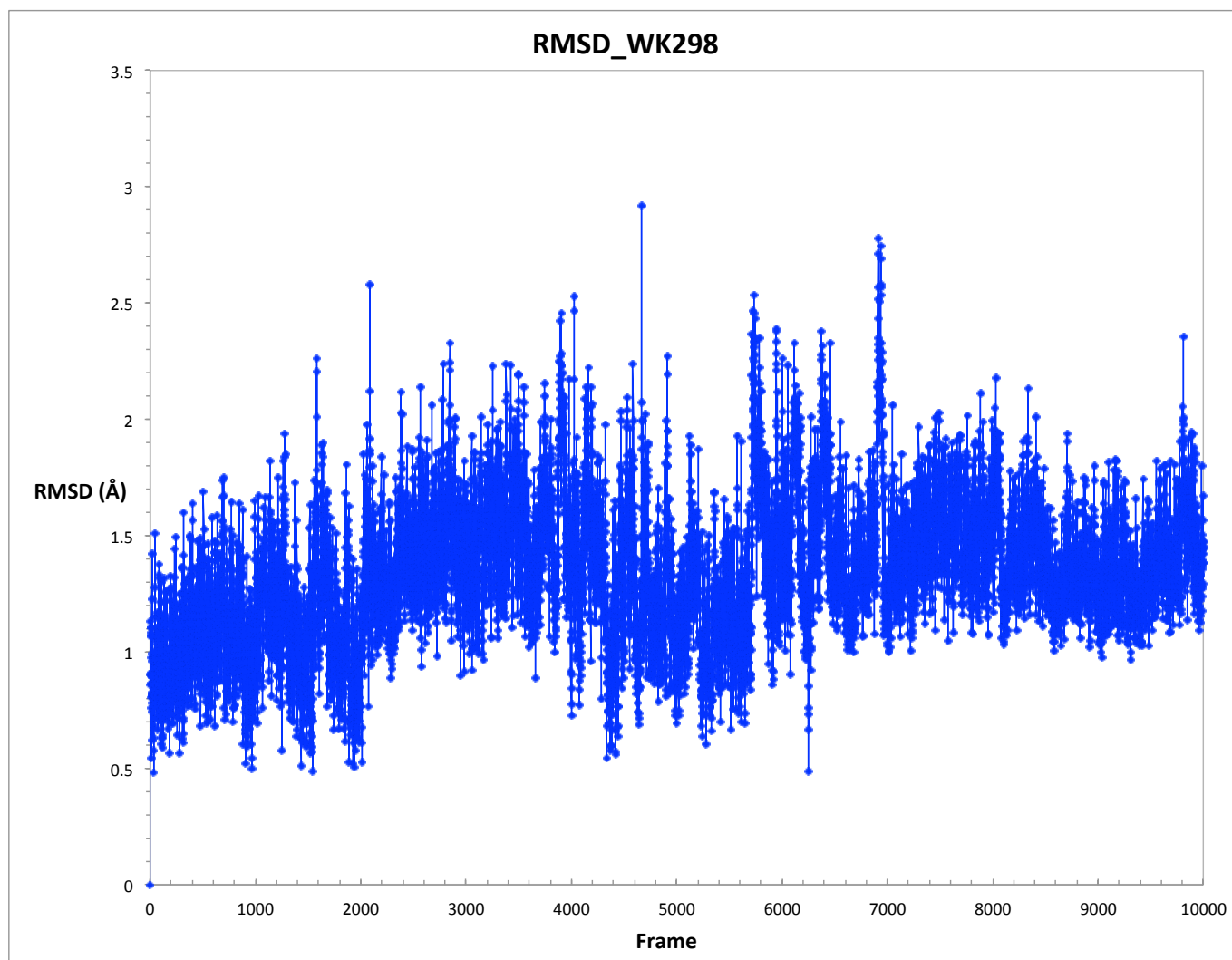
Figure 2: Learning curves. The numbers in circles correspond to cycle of the generation procedure.

## 2 Molecular Dynamics validation

For MD, the parameters and environment were set to run MD simulations work well on the WK298-Mdmx complex. We were able to get stable trajectories by running 10ns MD simulations on this complex, according to the RMSD (as shown in the table and the figure below).

Table 1: RMSD for trajectories of WK298-Mdmx

	std	avg	min	max
1	0.462	1.798	0.681	3.430
2	0.664	3.132	0.768	5.332
3	0.317	1.505	0.521	2.918
4	0.510	3.422	0.750	4.748
5	0.391	1.811	0.495	3.575
6	0.685	4.693	0.731	7.137
7	0.570	2.086	0.610	3.974
8	0.467	3.343	0.737	4.263
9	2.758	5.966	0.735	11.059
10	0.414	1.757	0.516	3.106



We also performed RMSD analysis based on the center of mass of the ligands.

Table 2: RMSD for trajectories of WK298-Mdmx (center of mass)

	std	avg	min	max
1	0.431	1.957	0.829	3.087
2	0.455	2.865	0.821	4.356
3	0.364	1.990	0.613	3.176
4	0.364	2.999	0.610	4.340
5	0.283	1.644	0.553	2.822
6	0.556	4.516	0.669	5.883
7	0.394	1.792	0.597	3.188
8	0.344	2.687	0.757	3.260
9	1.624	4.243	0.712	7.344
10	0.344	1.628	0.666	2.750

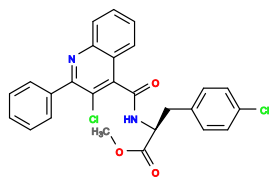
Table 3: RMSD for the most promising ligands (center of mass)

Compound	std	avg	min	max
WK298	0.364	1.990	0.613	3.176
3021	0.353	4.132	0.499	5.071
92	0.338	1.387	0.433	2.644
100	0.547	1.473	0.383	4.220
34	0.746	2.680	0.722	4.817
39	0.832	4.292	0.593	6.504

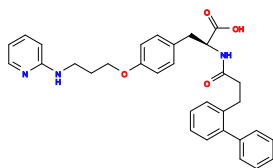
### 3 Putative ligands

#### 3.1 Top-500 compounds from the Run-1

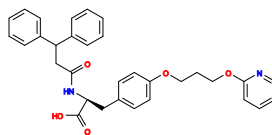
For each compound in the following tables we present at least two of the values:  $IC_{50}$  values (I), VINA scores (V), Solubility (log(mol/L)) (S), and CID (ChEMBL ID).



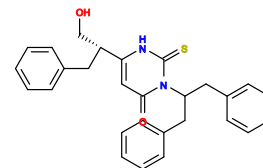
1 V= -7.4, I=-7.87



2 V= -6.8, I=-8.3

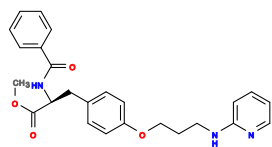


3 V= -6.8, I=-8.21

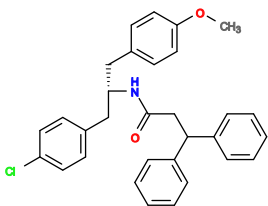


4 V= -6.9, I=-8.08

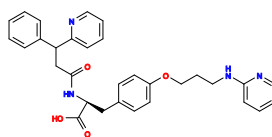




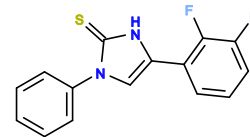
5 V= -6.7, I =-8.22



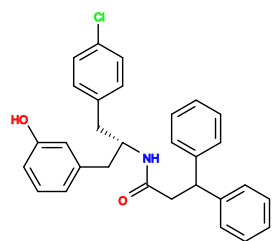
6 V= -7.3, I =-7.72



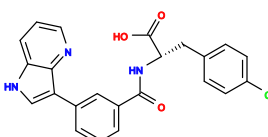
7 V= -6.4, I =-8.36



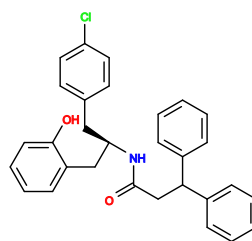
8 V= -6.7, I =-8.13



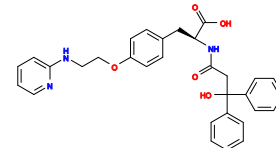
9 V= -7.3, I =-7.67



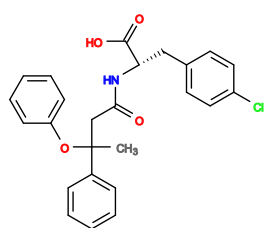
10 V= -7.8, I =-7.29



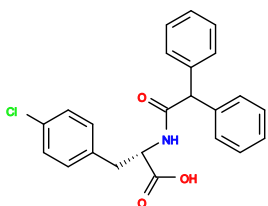
11 V= -7.2, I =-7.72



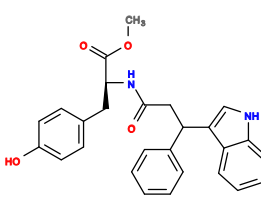
12 V= -6.3, I =-8.38



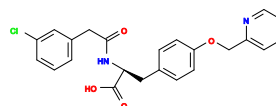
13 V= -6.7, I =-8.02



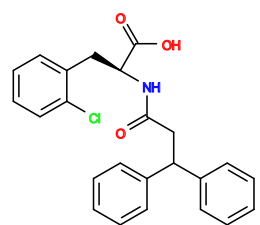
14 V= -6.9, I =-7.86,  
CID73918829



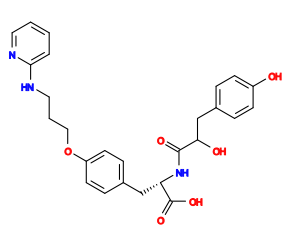
15 V= -7, I =-7.76



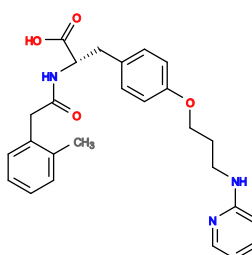
16 V= -6.6, I =-8.05



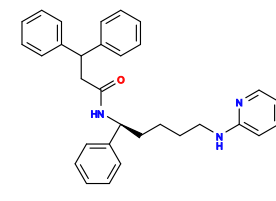
17 V= -7, I =-7.74



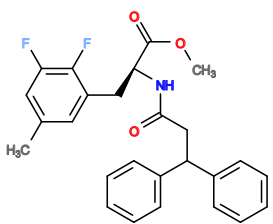
18 V= -6.3, I =-8.26



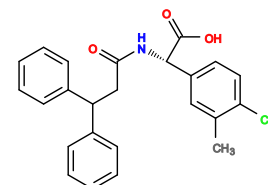
19 V= -6.1, I =-8.41



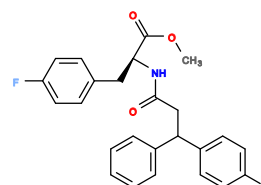
20 V= -7.2, I =-7.57



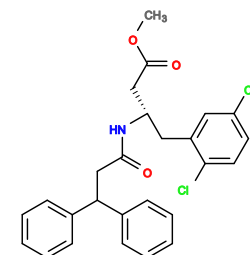
21 V= -7.1, I =-7.64



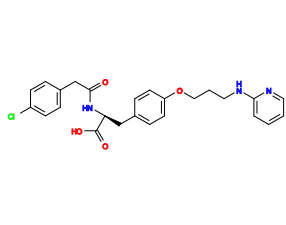
22 V= -7.2, I =-7.57



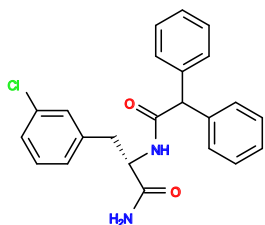
23 V= -6.9, I =-7.79



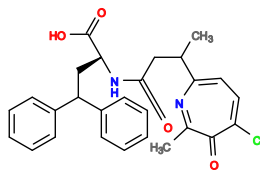
24 V= -6.8, I =-7.84



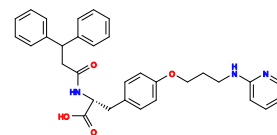
25  $V = -6.1, I = -8.35$



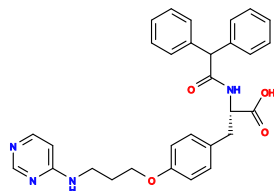
26  $V = -6.9, I = -7.73,$   
CID73976725



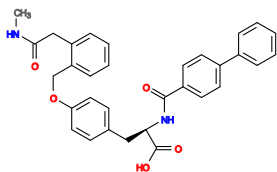
27  $V = -7.2, I = -7.5$



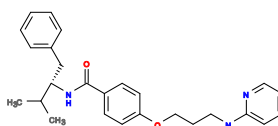
28  $V = -6.1, I = -8.31$



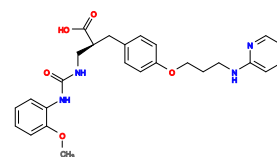
29  $V = -6.5, I = -7.99$



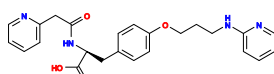
30  $V = -7.9, I = -6.94$



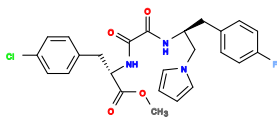
31  $V = -6.5, I = -7.98$



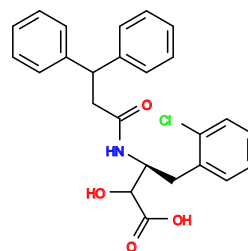
32  $V = -6.1, I = -8.27$



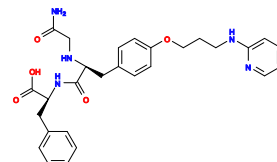
33  $V = -6, I = -8.34$



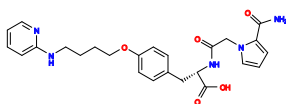
34  $V = -6.6, I = -7.88$



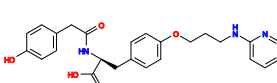
35  $V = -6.9, I = -7.65$



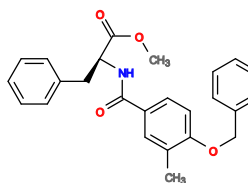
36  $V = -6.1, I = -8.25$



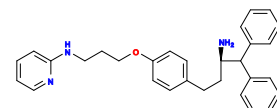
37  $V = -6.2, I = -8.17$



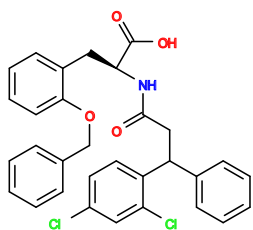
38  $V = -6.1, I = -8.24$



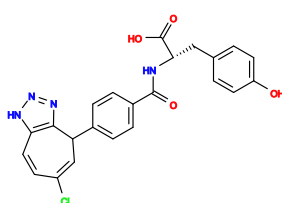
39  $V = -6.9, I = -7.62$



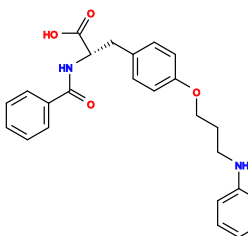
40  $V = -6.7, I = -7.77$



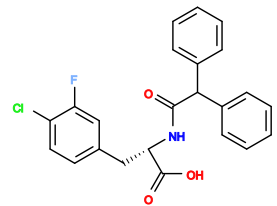
41  $V = -6.6, I = -7.83$



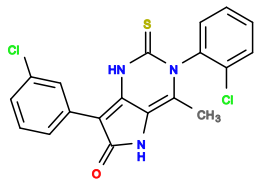
42  $V = -8.3, I = -6.55$



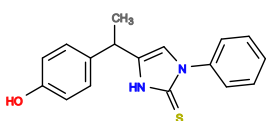
43  $V = -6, I = -8.27,$   
CID16665099



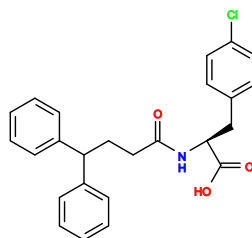
44  $V = -7, I = -7.52$



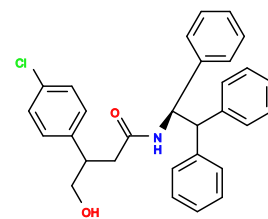
45 V= -6.8, I=-7.66



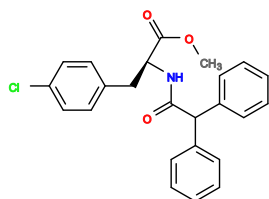
46 V= -6.3, I=-8.02



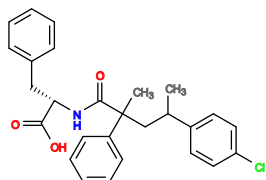
47 V= -6.6, I=-7.8



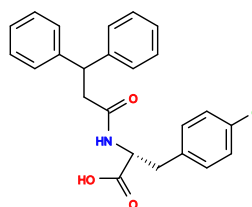
48 V= -7.3, I=-7.27



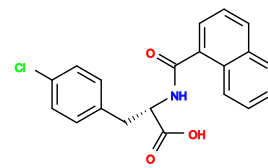
49 V= -6.4, I=-7.94,  
CID73939626



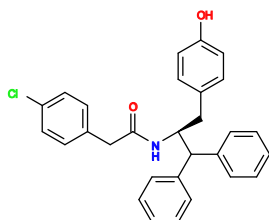
50 V= -6.9, I=-7.56



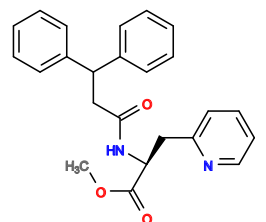
51 V= -6.4, I=-7.92



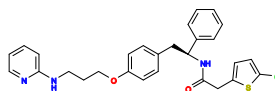
52 V= -6.7, I=-7.7,  
CID73918332



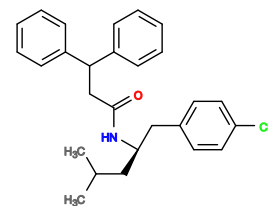
53 V= -6.9, I=-7.54



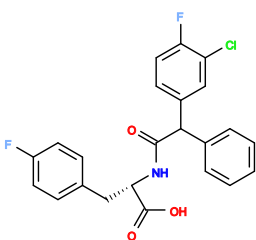
54 V= -6.6, I=-7.77



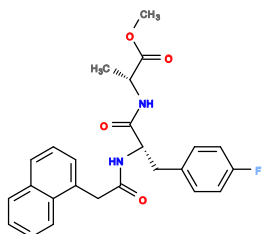
55 V= -6, I=-8.21



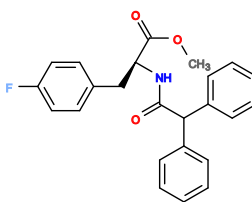
56 V= -6.8, I=-7.61



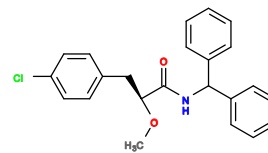
57 V= -7.4, I=-7.15



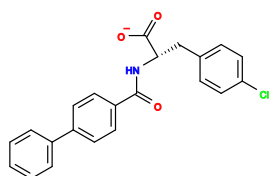
58 V= -6.8, I=-7.6



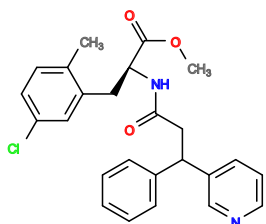
59 V= -6.7, I=-7.68,  
CID73939623



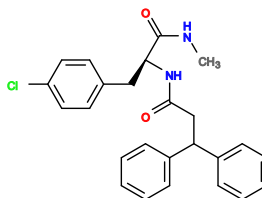
60 V= -7, I=-7.44



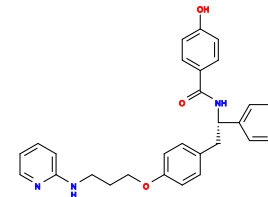
61 V= -7, I=-7.44



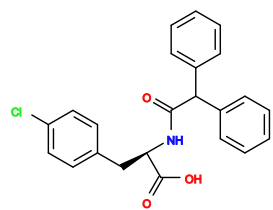
62 V= -6.7, I=-7.66



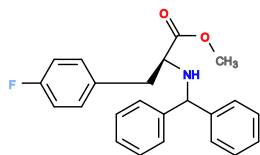
63 V= -6.4, I=-7.89



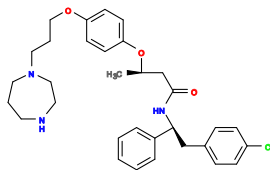
64 V= -6.3, I=-7.96



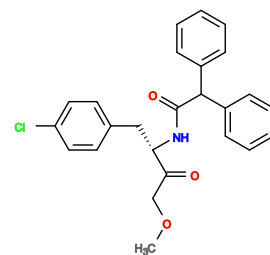
65  $V = -6.4$ ,  $I = -7.85$ ,  
CID73908087



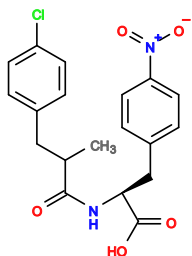
66  $V = -6.7$ ,  $I = -7.62$



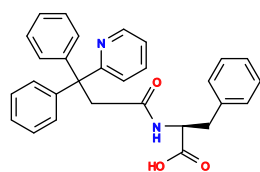
67  $V = -6.6$ ,  $I = -7.7$



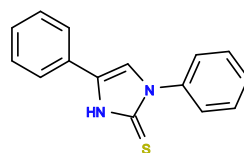
68  $V = -6.4$ ,  $I = -7.85$



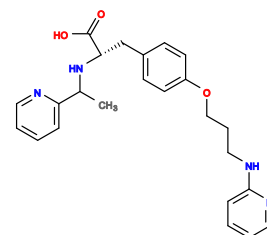
69  $V = -7$ ,  $I = -7.38$



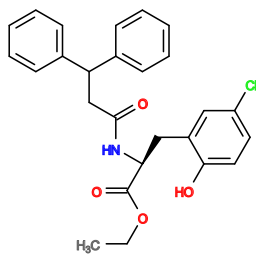
70  $V = -6.5$ ,  $I = -7.76$



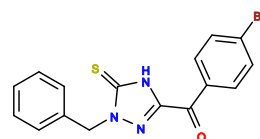
71  $V = -6.3$ ,  $I = -7.9$ ,  
CID736689



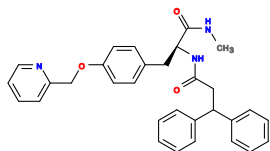
72  $V = -5.7$ ,  $I = -8.35$



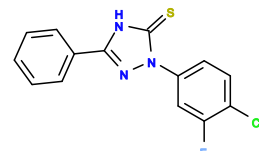
73  $V = -6.8$ ,  $I = -7.52$



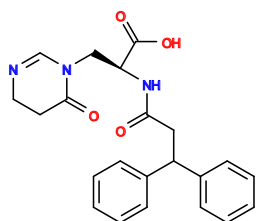
74  $V = -6.4$ ,  $I = -7.82$



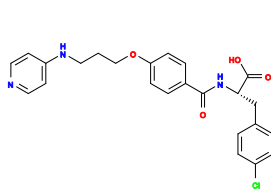
75  $V = -6.2$ ,  $I = -7.97$



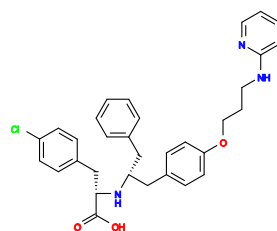
76  $V = -6.5$ ,  $I = -7.74$



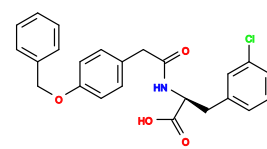
77  $V = -7$ ,  $I = -7.36$



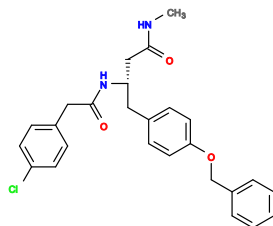
78  $V = -6.3$ ,  $I = -7.87$



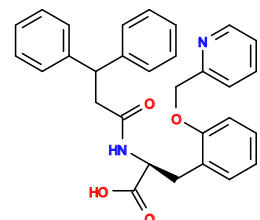
79  $V = -5.7$ ,  $I = -8.32$



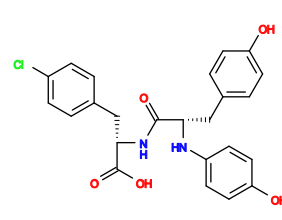
80  $V = -6.5$ ,  $I = -7.72$



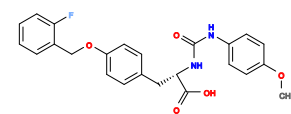
81  $V = -7.1$ ,  $I = -7.26$



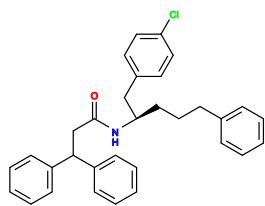
82  $V = -6.6$ ,  $I = -7.63$



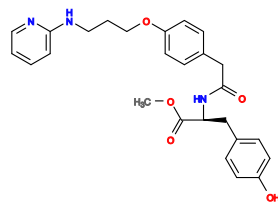
83  $V = -6.7$ ,  $I = -7.55$



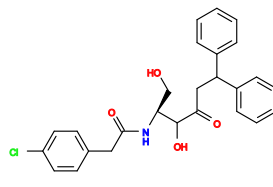
84  $V = -7.1$ ,  $I = -7.25$



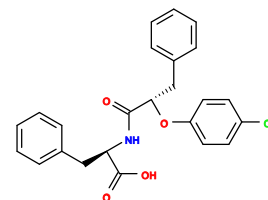
85  $V = -6.8$ ,  $I = -7.47$



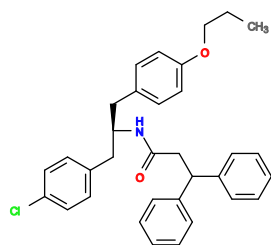
86  $V = -5.9$ ,  $I = -8.14$



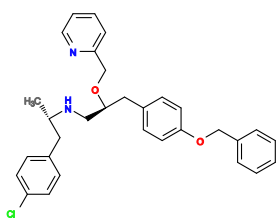
87  $V = -6.4$ ,  $I = -7.77$



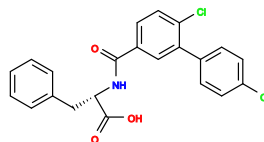
88  $V = -6.6$ ,  $I = -7.61$



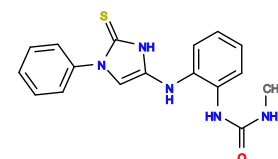
89  $V = -6.8$ ,  $I = -7.45$



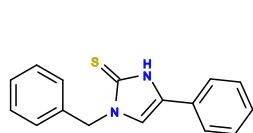
90  $V = -6.6$ ,  $I = -7.59$



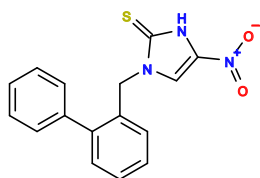
91  $V = -7$ ,  $I = -7.29$ ,  
CID44280454



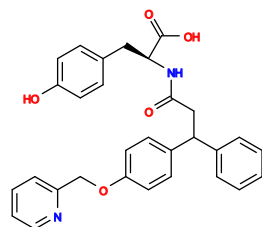
92  $V = -6.4$ ,  $I = -7.74$



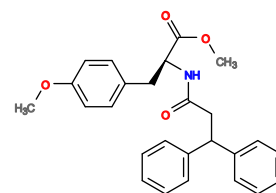
93  $V = -6.2$ ,  $I = -7.88$ ,  
CID24964671



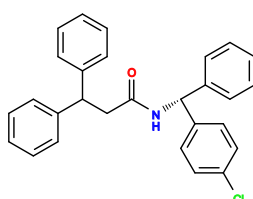
94  $V = -6.6$ ,  $I = -7.58$



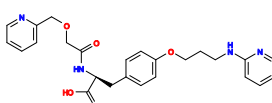
95  $V = -6.6$ ,  $I = -7.58$



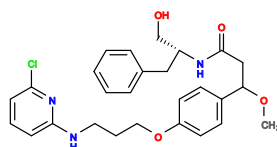
96  $V = -6.4$ ,  $I = -7.73$



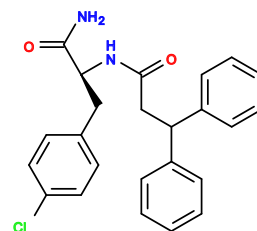
97  $V = -7.1$ ,  $I = -7.2$



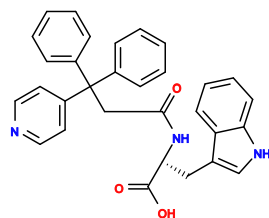
98  $V = -5.6$ ,  $I = -8.32$



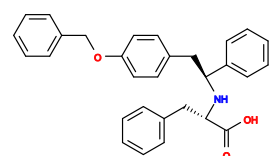
99  $V = -6.2$ ,  $I = -7.87$



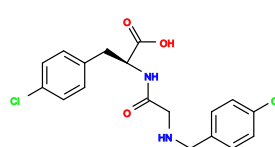
100  $V = -6.1$ ,  $I = -7.94$



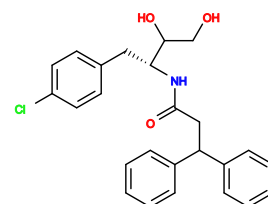
101  $V = -6.7$ ,  $I = -7.49$



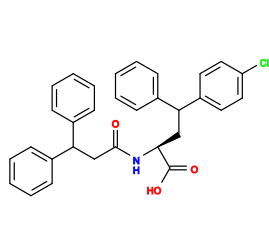
102  $V = -7$ ,  $I = -7.25$



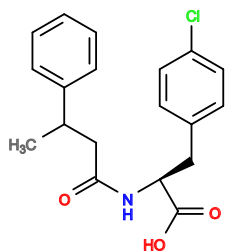
103  $V = -6.5$ ,  $I = -7.63$



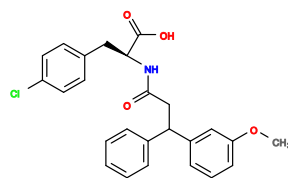
104  $V = -6.2$ ,  $I = -7.85$



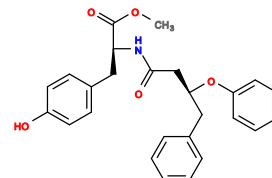
105 V= -6.2, I=-7.85



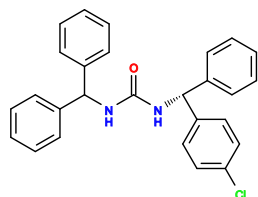
106 V= -6.1, I=-7.92



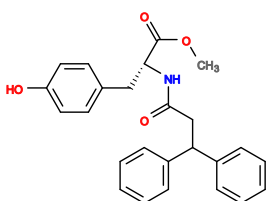
107 V= -6.5, I=-7.62



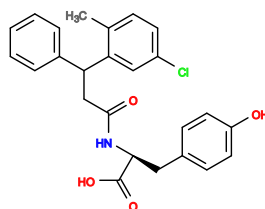
108 V= -6.4, I=-7.69



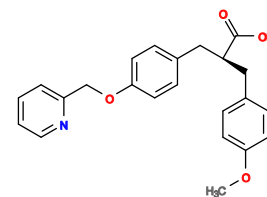
109 V= -7.2, I=-7.09



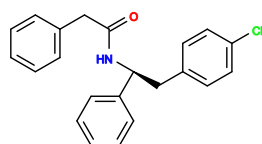
110 V= -6.4, I=-7.68



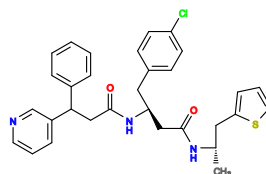
111 V= -6.8, I=-7.37



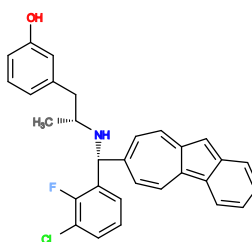
112 V= -6.7, I=-7.45



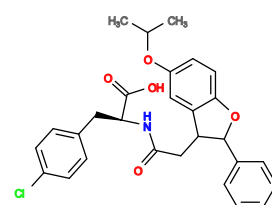
113 V= -6.9, I=-7.3



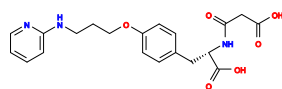
114 V= -6.7, I=-7.44



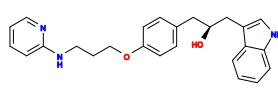
115 V= -8.1, I=-6.39



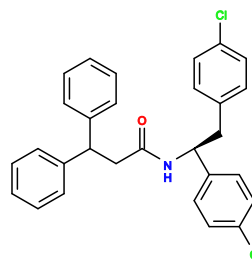
116 V= -7.2, I=-7.06



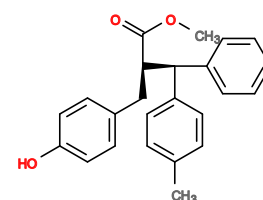
117 V= -5.5, I=-8.33



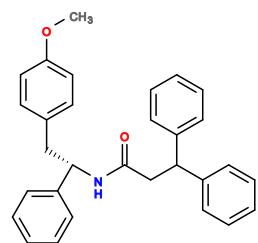
118 V= -6.3, I=-7.73



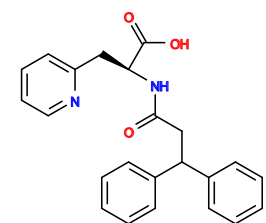
119 V= -6.5, I=-7.57



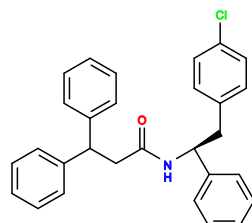
120 V= -7, I=-7.19



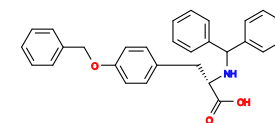
121 V= -7, I=-7.19



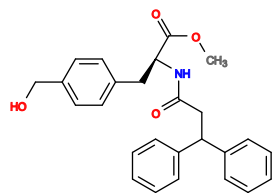
122 V= -6.8, I=-7.33



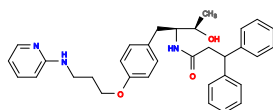
123 V= -6.4, I=-7.63



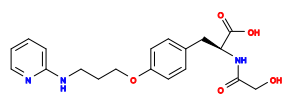
124 V= -7, I=-7.17



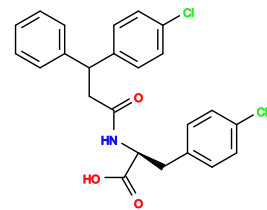
125 V= -6.3, I=-7.69



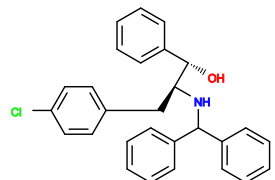
126 V= -5.6, I=-8.22



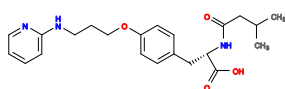
127 V= -5.7, I=-8.14



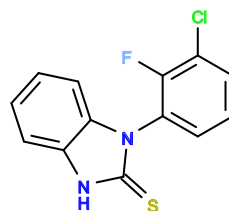
128 V= -6.2, I=-7.76



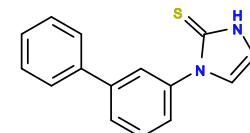
129 V= -6.4, I=-7.61



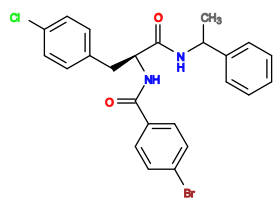
130 V= -5.4, I=-8.36



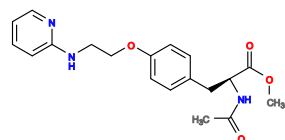
131 V= -6.4, I=-7.61,  
CID29067750



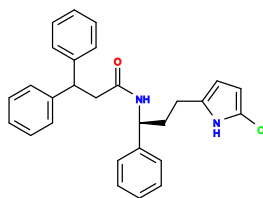
132 V= -6.8, I=-7.31



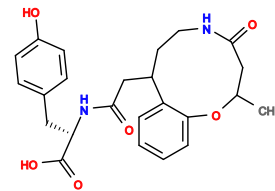
133 V= -6.7, I=-7.38



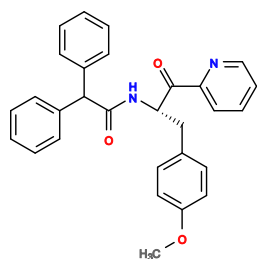
134 V= -5.7, I=-8.13



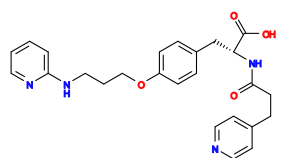
135 V= -6.8, I=-7.3



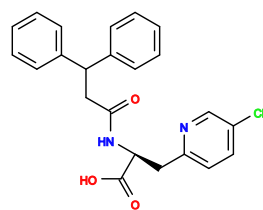
136 V= -7.6, I=-6.7



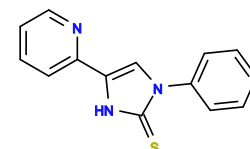
137 V= -6.9, I=-7.22



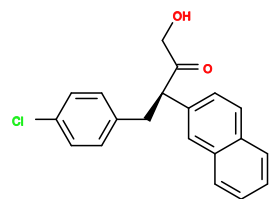
138 V= -6, I=-7.9



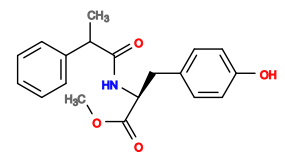
139 V= -6.9, I=-7.22



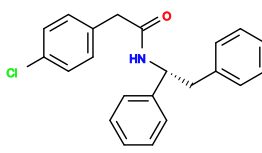
140 V= -5.9, I=-7.97



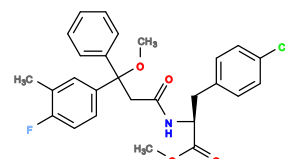
141 V= -6.9, I=-7.22



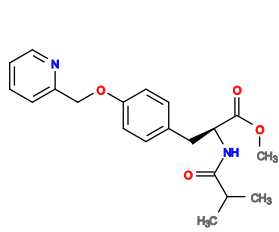
142 V= -6.4, I=-7.59



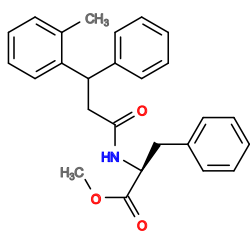
143 V= -6.9, I=-7.21



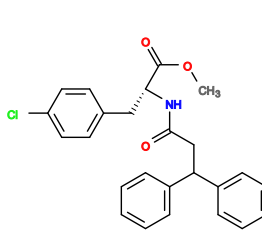
144 V= -5.9, I=-7.96



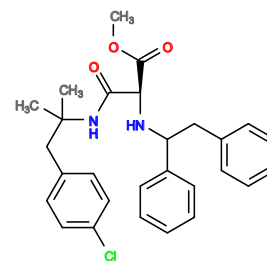
145 V= -6.1, I=-7.81



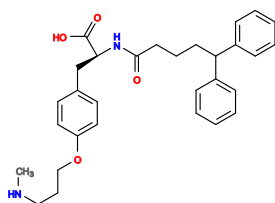
146 V= -6.2, I=-7.73



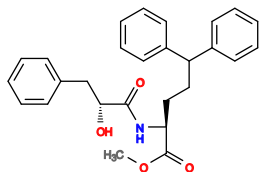
147 V= -5.9, I=-7.95



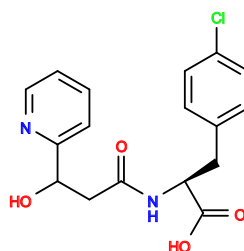
148 V= -6.5, I=-7.5



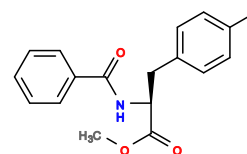
149 V= -6.4, I=-7.57



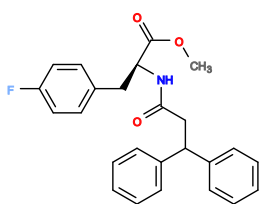
150 V= -6.3, I=-7.65



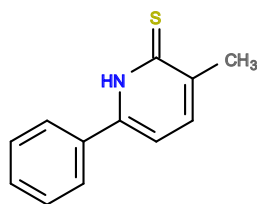
151 V= -6.2, I=-7.72



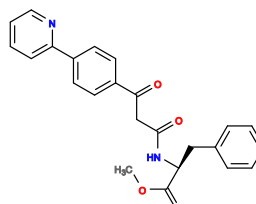
152 V= -6.4, I=-7.56,  
CID11312655



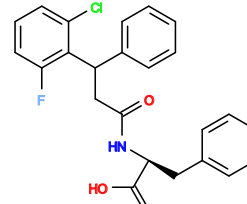
153 V= -6.2, I=-7.71



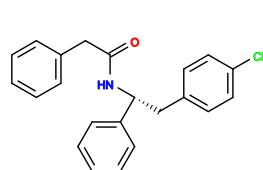
154 V= -6.4, I=-7.56



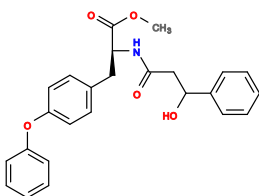
155 V= -6.1, I=-7.78



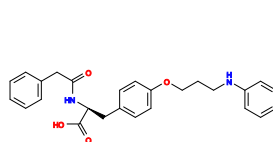
156 V= -6.8, I=-7.25



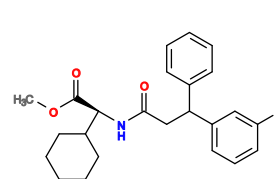
157 V= -6.8, I=-7.24



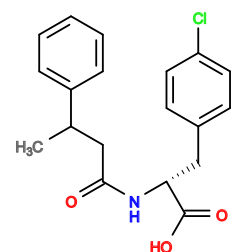
158 V= -6.1, I=-7.77



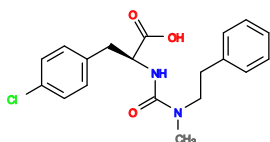
159 V= -5.9, I=-7.92



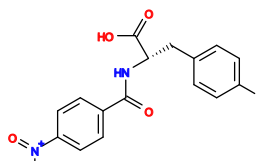
160 V= -6.6, I=-7.39



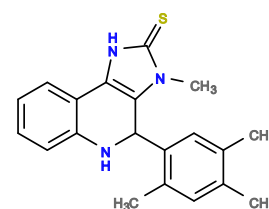
161 V= -5.9, I=-7.91



162 V= -6.4, I=-7.54

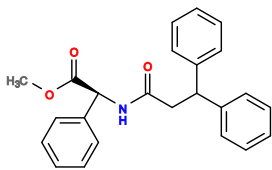


163 V= -6.8, I=-7.23

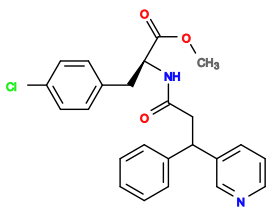


164 V= -7.3, I=-6.85

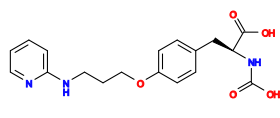




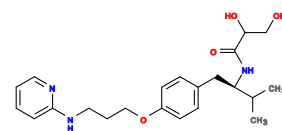
165 V= -6.2, I=-7.68



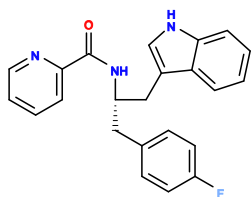
166 V= -5.9, I=-7.9



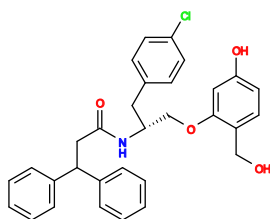
167 V= -5.5, I=-8.19



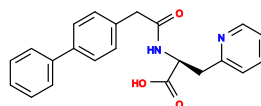
168 V= -5.7, I=-8.04



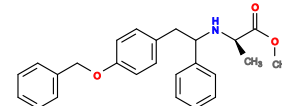
169 V= -7.2, I=-6.92



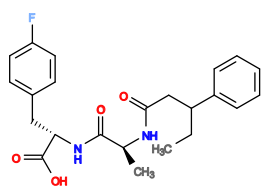
170 V= -6.4, I=-7.52



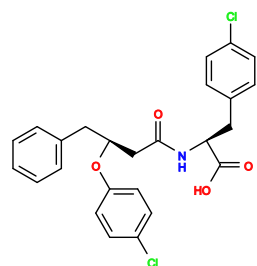
171 V= -6.7, I=-7.29



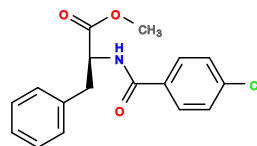
172 V= -6.4, I=-7.52



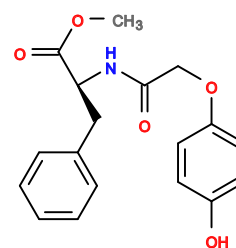
173 V= -6.8, I=-7.22



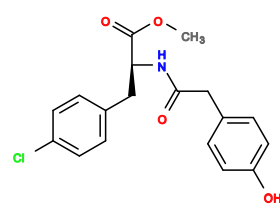
174 V= -6.2, I=-7.66



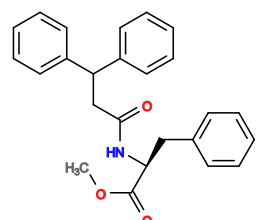
175 V= -6.2, I=-7.66,  
CID6935686



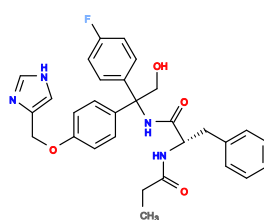
176 V= -6.3, I=-7.58



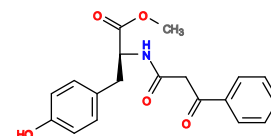
177 V= -6.1, I=-7.73



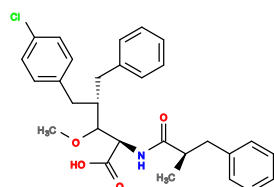
178 V= -6.1, I=-7.73,  
CID27228212



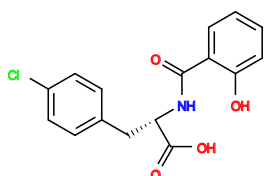
179 V= -7, I=-7.05



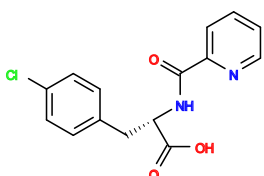
180 V= -6.3, I=-7.57



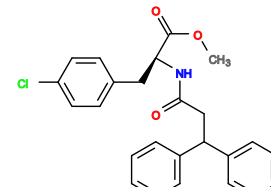
181 V= -6, I=-7.8



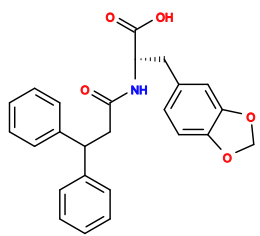
182 V= -6.3, I=-7.57



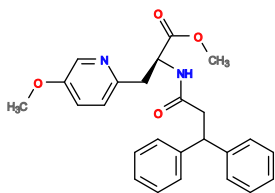
183 V= -6.4, I=-7.49,  
CID73912066



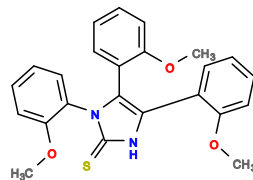
184 V= -5.8, I=-7.94



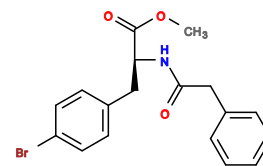
185  $V = -7.1$ ,  $I = -6.96$



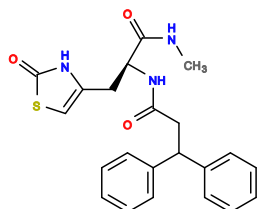
186  $V = -6.4$ ,  $I = -7.48$



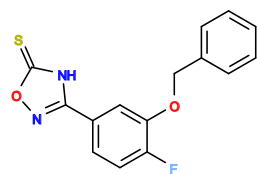
187  $V = -6$ ,  $I = -7.78$



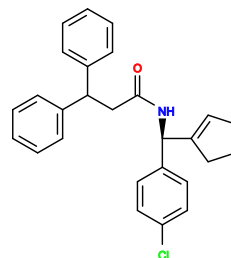
188  $V = -6.2$ ,  $I = -7.63$ ,  
CID73935591



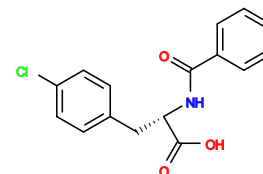
189  $V = -6.6$ ,  $I = -7.33$



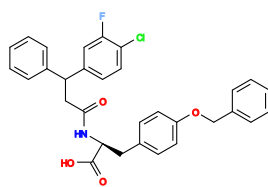
190  $V = -6.5$ ,  $I = -7.4$



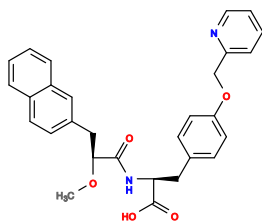
191  $V = -6.9$ ,  $I = -7.1$



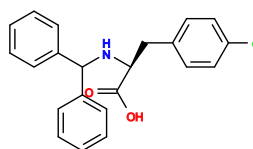
192  $V = -6.2$ ,  $I = -7.62$ ,  
CID21117456



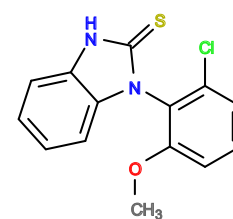
193  $V = -6.2$ ,  $I = -7.62$



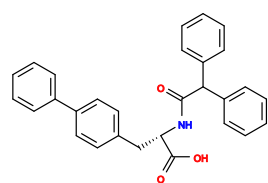
194  $V = -5.7$ ,  $I = -8.0$



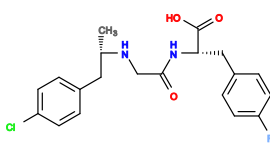
195  $V = -5.9$ ,  $I = -7.85$



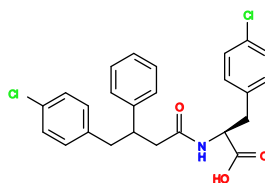
196  $V = -6.1$ ,  $I = -7.7$



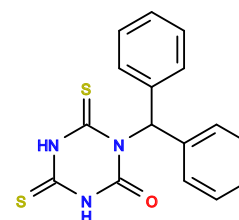
197  $V = -7.1$ ,  $I = -6.95$ ,  
CID73918862



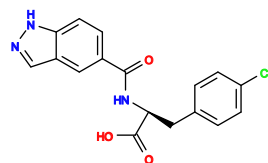
198  $V = -6.4$ ,  $I = -7.47$



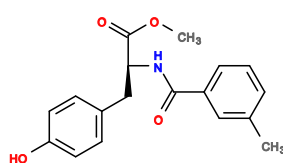
199  $V = -6.1$ ,  $I = -7.69$



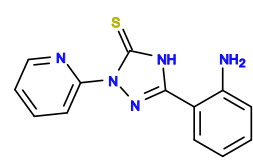
200  $V = -6.1$ ,  $I = -7.68$



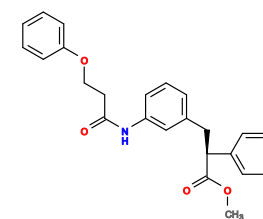
201  $V = -6.9$ ,  $I = -7.08$



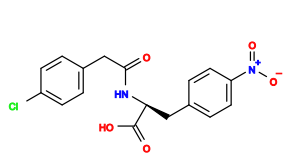
202  $V = -6.4$ ,  $I = -7.45$



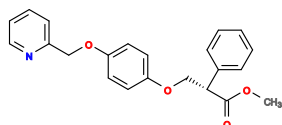
203  $V = -6.2$ ,  $I = -7.6$



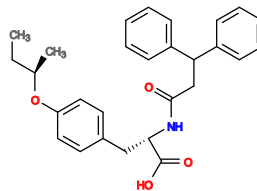
204  $V = -6.3$ ,  $I = -7.52$



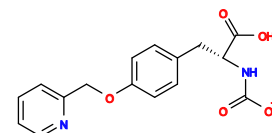
205 V= -6.6, I=-7.29



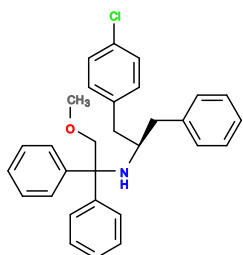
206 V= -6.1, I=-7.67



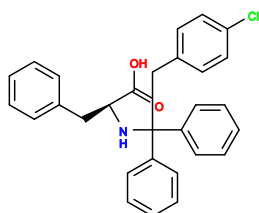
207 V= -6.7, I=-7.21



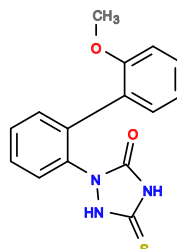
208 V= -6.3, I=-7.51



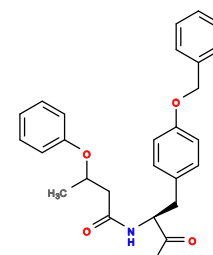
209 V= -6.1, I=-7.66



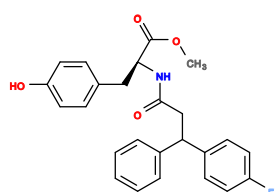
210 V= -5.7, I=-7.96



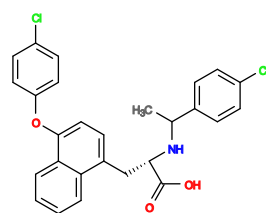
211 V= -6.1, I=-7.66



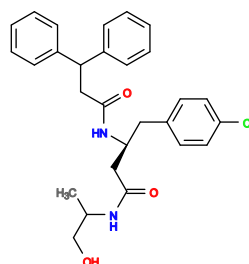
212 V= -6.1, I=-7.66



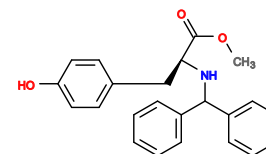
213 V= -6.1, I=-7.65



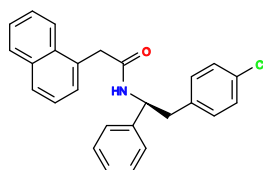
214 V= -6.4, I=-7.43



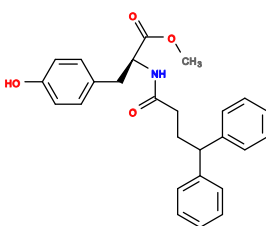
215 V= -6.1, I=-7.65



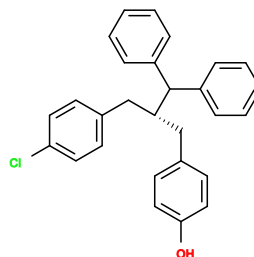
216 V= -6.2, I=-7.58



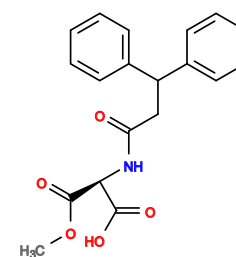
217 V= -6.5, I=-7.35



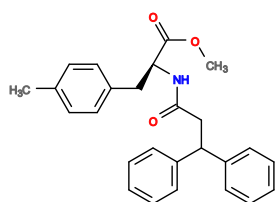
218 V= -6.1, I=-7.65



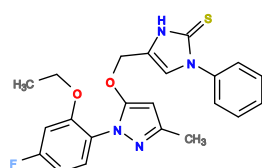
219 V= -6.8, I=-7.12



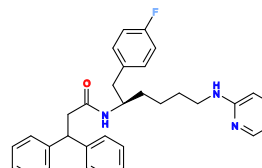
220 V= -6.1, I=-7.65



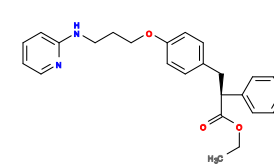
221 V= -6, I=-7.72



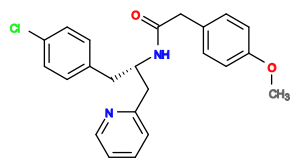
222 V= -6.5, I=-7.35



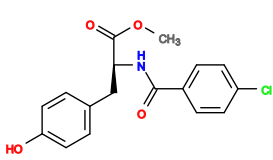
223 V= -5.8, I=-7.87



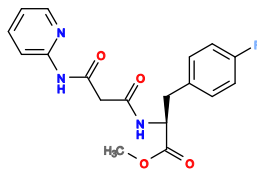
224 V= -5.8, I=-7.87



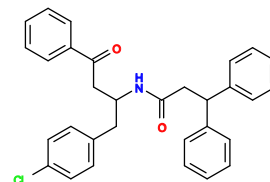
225  $V = -6.7$ ,  $I = -7.2$



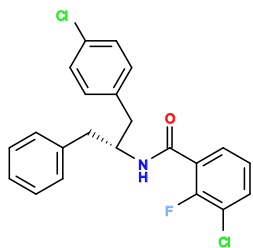
226  $V = -6.2$ ,  $I = -7.56$ ,  
CID678988



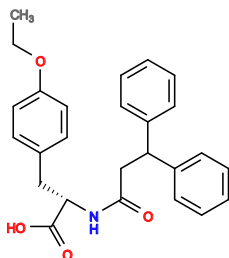
227  $V = -6.1$ ,  $I = -7.64$



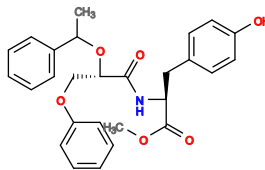
228  $V = -6.9$ ,  $I = -7.04$



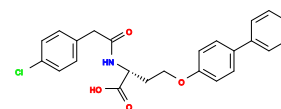
229  $V = -7.1$ ,  $I = -6.88$



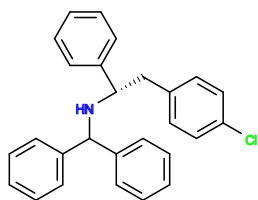
230  $V = -6.5$ ,  $I = -7.33$



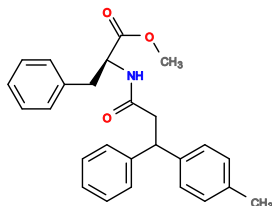
231  $V = -6$ ,  $I = -7.71$



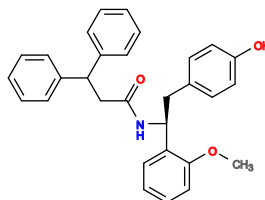
232  $V = -6$ ,  $I = -7.71$



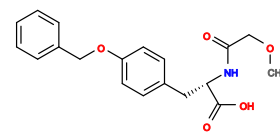
233  $V = -6.7$ ,  $I = -7.18$



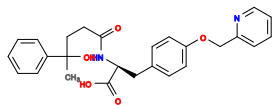
234  $V = -6.1$ ,  $I = -7.62$



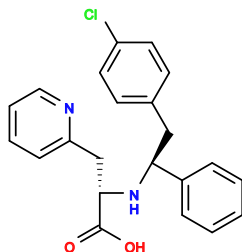
235  $V = -6.8$ ,  $I = -7.1$



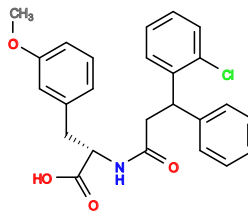
236  $V = -6.2$ ,  $I = -7.55$ ,  
CID73909070



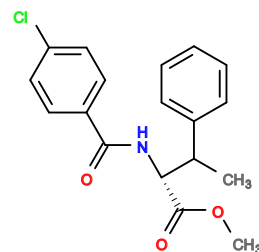
237  $V = -6.6$ ,  $I = -7.25$



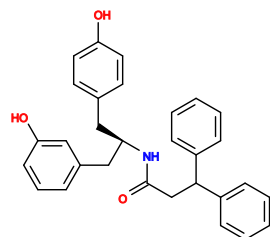
238  $V = -5.8$ ,  $I = -7.84$



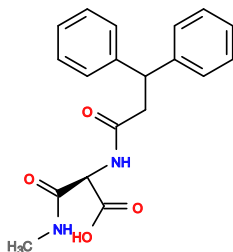
239  $V = -6.3$ ,  $I = -7.47$



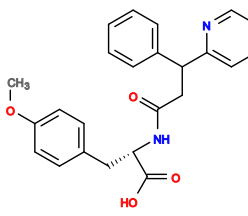
240  $V = -6.2$ ,  $I = -7.54$



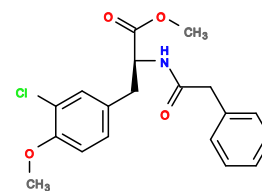
241  $V = -7.3$ ,  $I = -6.72$



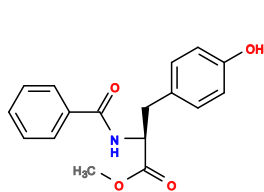
242  $V = -6.3$ ,  $I = -7.46$



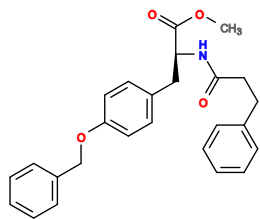
243  $V = -6.3$ ,  $I = -7.46$



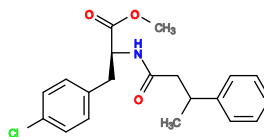
244  $V = -6$ ,  $I = -7.68$ ,  
CID73935588



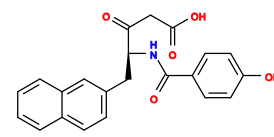
**245** V= -6.2, I =-7.53,  
CID691603



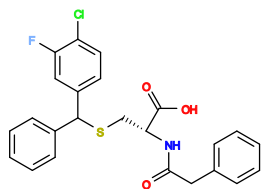
**246** V= -6, I =-7.68,  
CID73936978



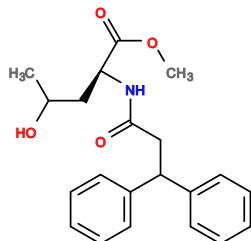
**247** V= -5.7, I =-7.9



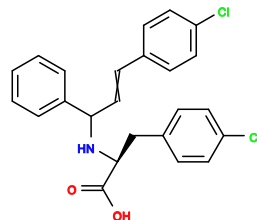
**248** V= -7.3, I =-6.7



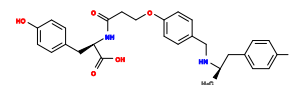
**249** V= -6.5, I =-7.3



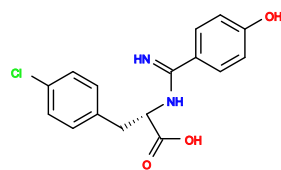
**250** V= -6.1, I =-7.6



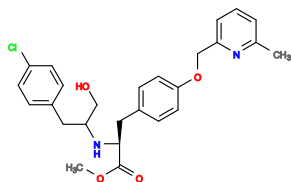
**251** V= -5.9, I =-7.75



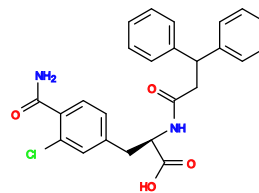
**252** V= -6.4, I =-7.37



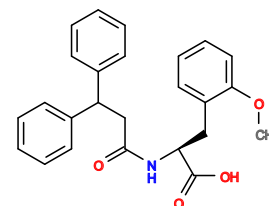
**253** V= -6.3, I =-7.44



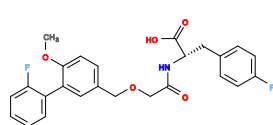
**254** V= -5.8, I =-7.82



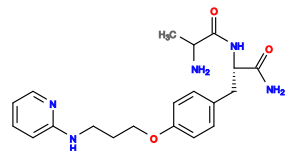
**255** V= -6.4, I =-7.35



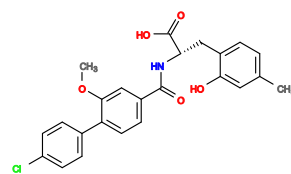
**256** V= -6.6, I =-7.2



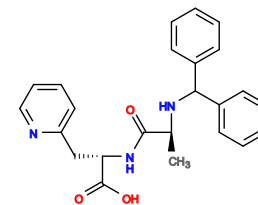
**257** V= -6.7, I =-7.12



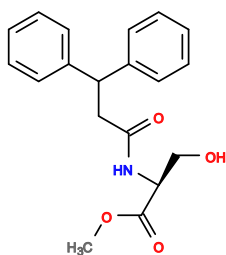
**258** V= -5.3, I =-8.17



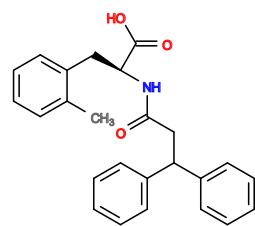
**259** V= -7.1, I =-6.82



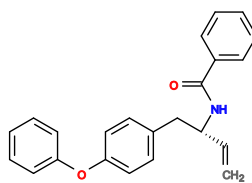
**260** V= -6.4, I =-7.34



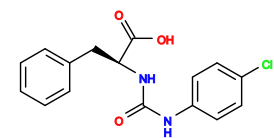
**261** V= -6.1, I =-7.57



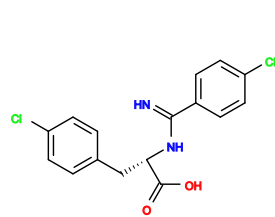
**262** V= -6.9, I =-6.96



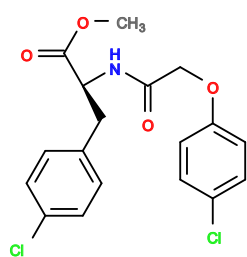
**263** V= -7.2, I =-6.74



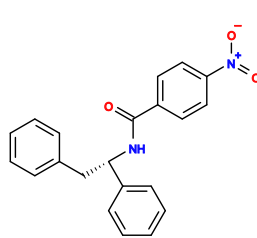
**264** V= -6.1, I =-7.56,  
CID941863



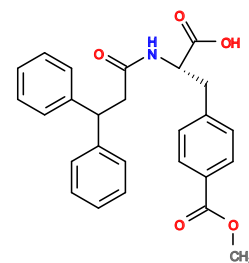
265  $V = -6.2$ ,  $I = -7.49$



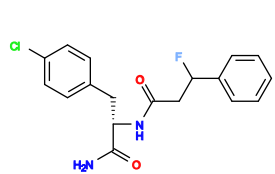
266  $V = -5.8$ ,  $I = -7.79$



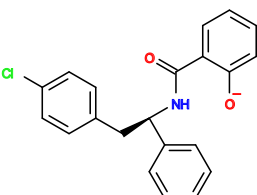
267  $V = -7.2$ ,  $I = -6.73$ ,  
CID29861853



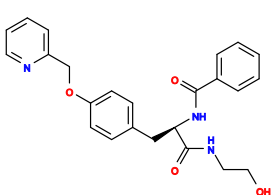
268  $V = -6.1$ ,  $I = -7.56$



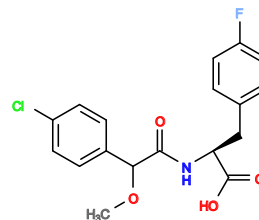
269  $V = -5.8$ ,  $I = -7.78$



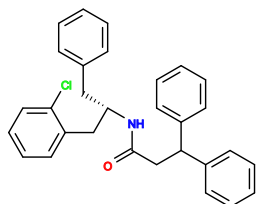
270  $V = -6.8$ ,  $I = -7.03$



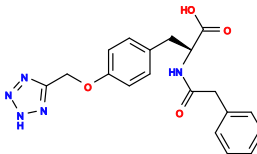
271  $V = -6.7$ ,  $I = -7.09$



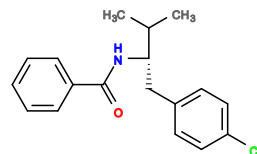
272  $V = -6.1$ ,  $I = -7.54$



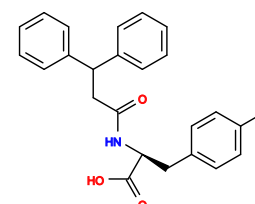
273  $V = -6.2$ ,  $I = -7.47$



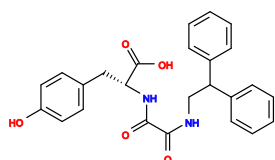
274  $V = -6.5$ ,  $I = -7.24$



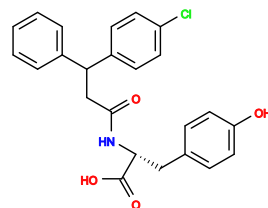
275  $V = -6.7$ ,  $I = -7.09$



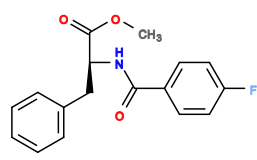
276  $V = -6.9$ ,  $I = -6.94$



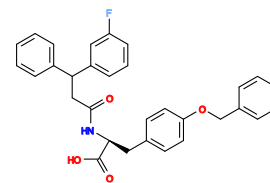
277  $V = -7$ ,  $I = -6.86$



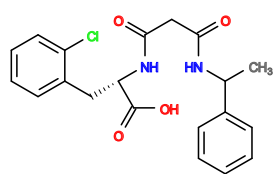
278  $V = -6.1$ ,  $I = -7.53$



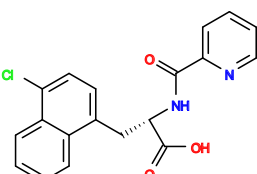
279  $V = -6.2$ ,  $I = -7.45$ ,  
CID6990864



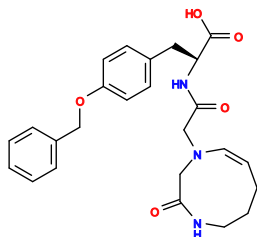
280  $V = -6.4$ ,  $I = -7.3$



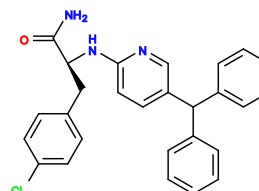
281  $V = -6.1$ ,  $I = -7.52$



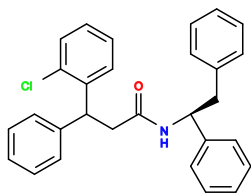
282  $V = -6.6$ ,  $I = -7.15$



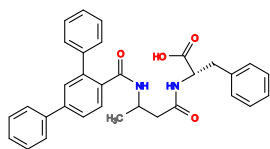
283  $V = -6.9$ ,  $I = -6.92$



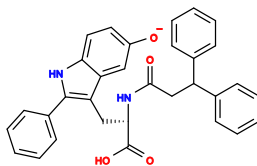
284  $V = -6.7$ ,  $I = -7.07$



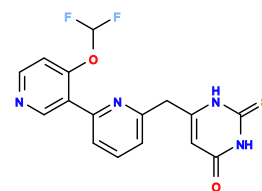
285 V= -6.6, I=-7.14



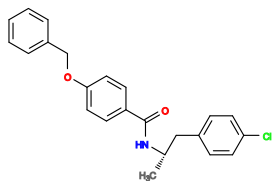
286 V= -6.8, I=-6.99



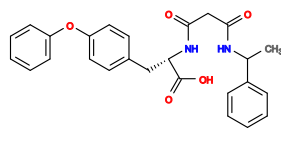
287 V= -6.7, I=-7.07



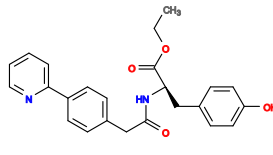
288 V= -7, I=-6.84



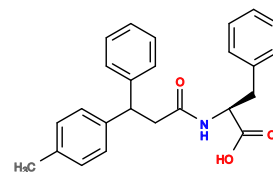
289 V= -6.9, I=-6.91



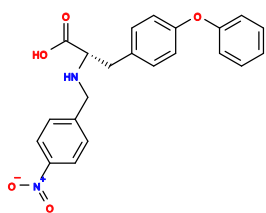
290 V= -6.3, I=-7.36



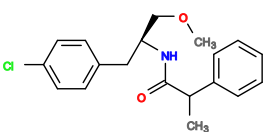
291 V= -6.5, I=-7.21



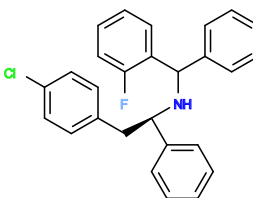
292 V= -6.9, I=-6.91



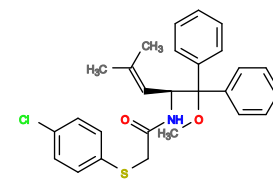
293 V= -6.8, I=-6.98



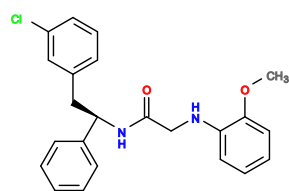
294 V= -6.3, I=-7.35



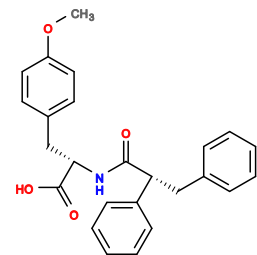
295 V= -6.8, I=-6.97



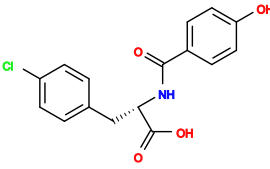
296 V= -6, I=-7.57



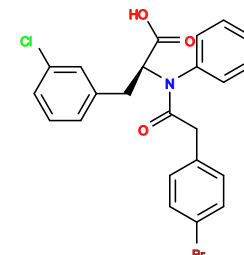
297 V= -6.9, I=-6.89



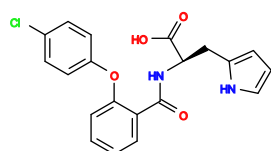
298 V= -6.5, I=-7.19



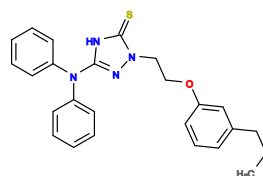
299 V= -6.4, I=-7.26



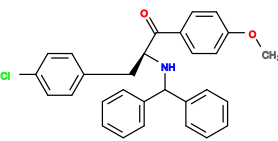
300 V= -6.3, I=-7.33



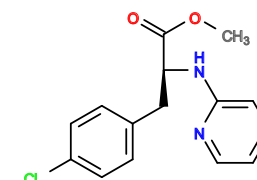
301 V= -6.3, I=-7.33



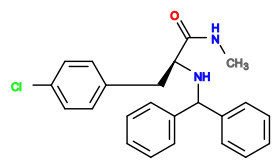
302 V= -6.6, I=-7.1



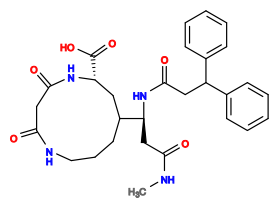
303 V= -6.3, I=-7.33



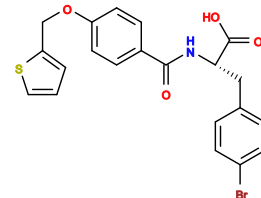
304 V= -5.8, I=-7.7



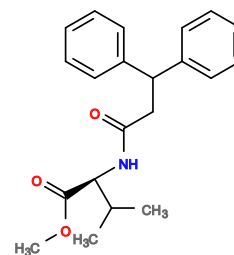
305  $V = -5.9$ ,  $I = -7.62$



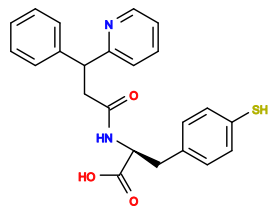
306  $V = -7$ ,  $I = -6.8$



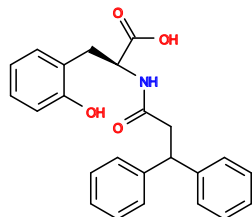
307  $V = -6.5$ ,  $I = -7.17$



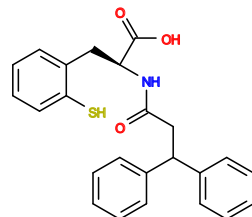
308  $V = -6$ ,  $I = -7.54$ ,  
CID52716123



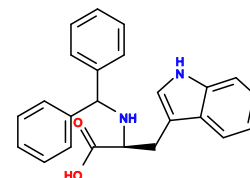
309  $V = -6.4$ ,  $I = -7.24$



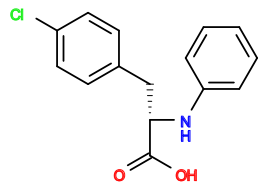
310  $V = -6.9$ ,  $I = -6.86$



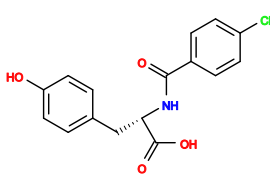
311  $V = -6.8$ ,  $I = -6.94$



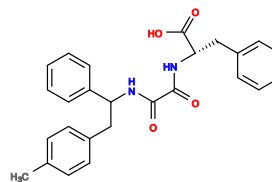
312  $V = -6.6$ ,  $I = -7.09$



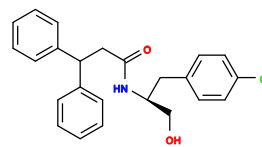
313  $V = -5.8$ ,  $I = -7.68$



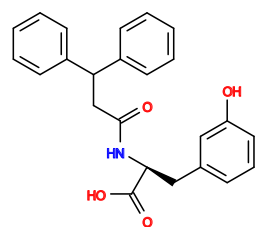
314  $V = -6.5$ ,  $I = -7.16$ ,  
CID57896336



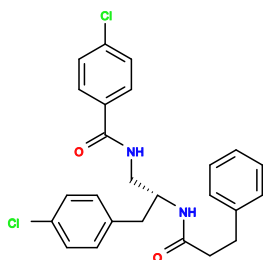
315  $V = -6.9$ ,  $I = -6.86$



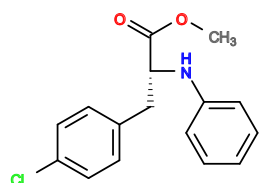
316  $V = -5.8$ ,  $I = -7.68$



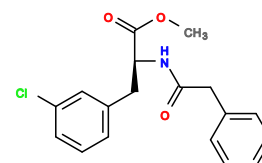
317  $V = -6.9$ ,  $I = -6.85$



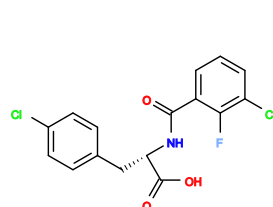
318  $V = -6.5$ ,  $I = -7.15$



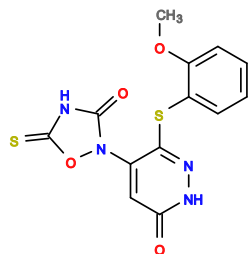
319  $V = -5.7$ ,  $I = -7.75$



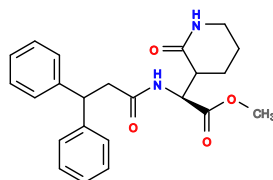
320  $V = -5.6$ ,  $I = -7.82$ ,  
CID73935584



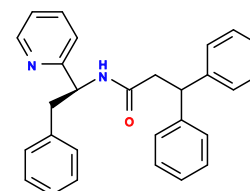
321  $V = -6.4$ ,  $I = -7.22$



322  $V = -5.9$ ,  $I = -7.59$

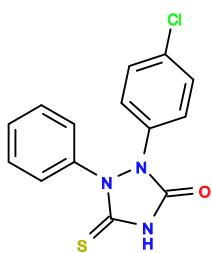


323  $V = -6.2$ ,  $I = -7.37$

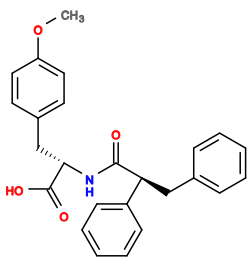


324  $V = -6.6$ ,  $I = -7.07$

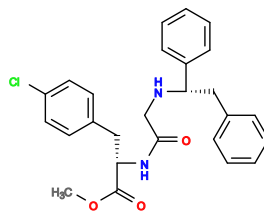




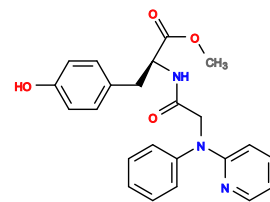
325 V= -5.9, I=-7.59



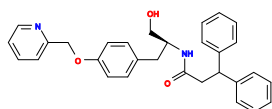
326 V= -6.4, I=-7.22



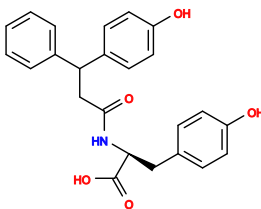
327 V= -5.4, I=-7.97



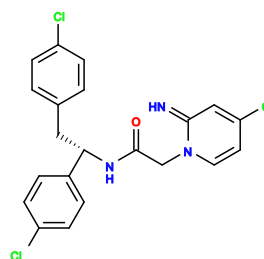
328 V= -5.7, I=-7.74



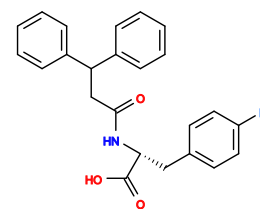
329 V= -5.7, I=-7.74



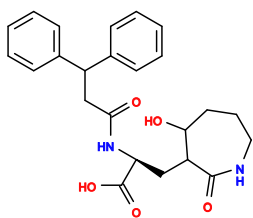
330 V= -6.9, I=-6.84



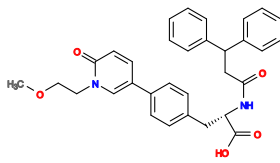
331 V= -7.1, I=-6.69



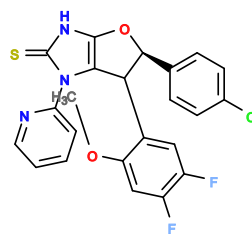
332 V= -6.8, I=-6.91



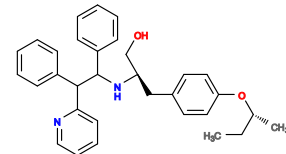
333 V= -7.1, I=-6.68



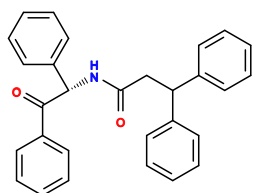
334 V= -6.2, I=-7.36



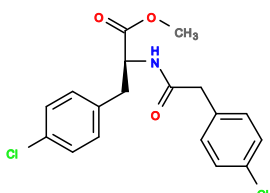
335 V= -6.6, I=-7.05



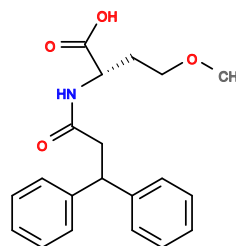
336 V= -6.5, I=-7.12



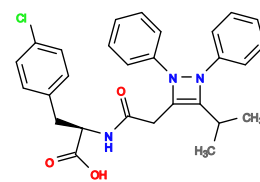
337 V= -7.2, I=-6.6



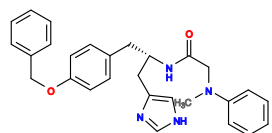
338 V= -5.7, I=-7.72



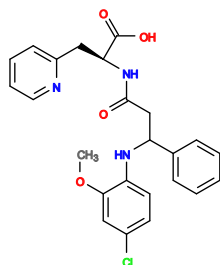
339 V= -6.1, I=-7.42



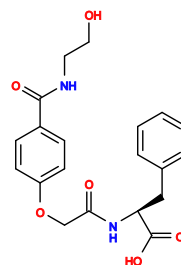
340 V= -6.6, I=-7.05



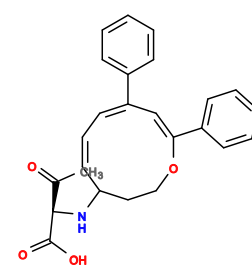
341 V= -6.6, I=-7.04



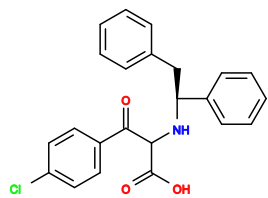
342 V= -6.2, I=-7.34



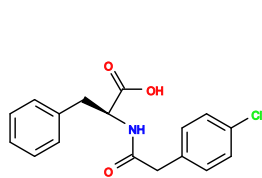
343 V= -6.6, I=-7.04



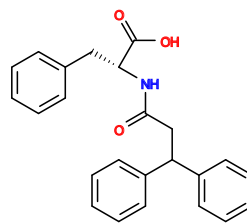
344 V= -7, I=-6.74



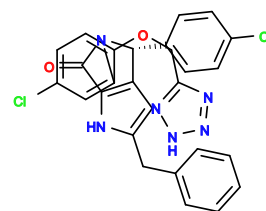
345 V= -6.2, I=-7.34



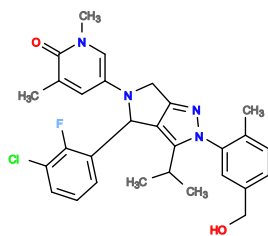
346 V= -5.7, I=-7.71



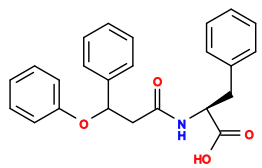
347 V= -6.8, I=-6.89



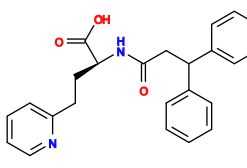
348 V= -7.7, I=-6.21



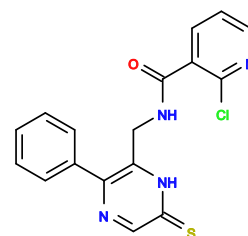
349 V= -8, I=-5.99



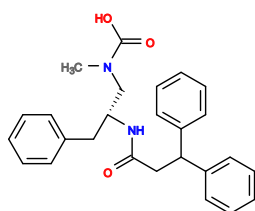
350 V= -6.7, I=-6.96



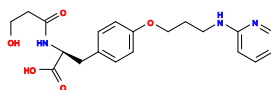
351 V= -6.3, I=-7.26



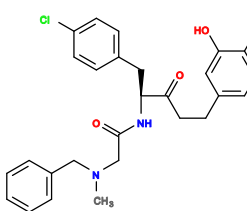
352 V= -6.5, I=-7.11



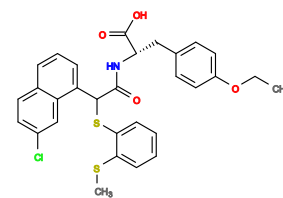
353 V= -6.6, I=-7.03



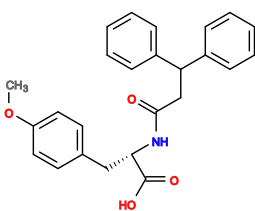
354 V= -5.3, I=-8.0



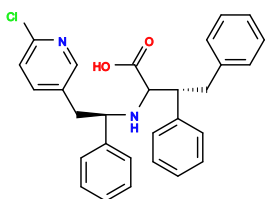
355 V= -6.4, I=-7.18



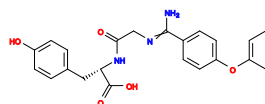
356 V= -6.8, I=-6.88



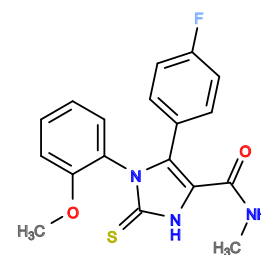
357 V= -6.3, I=-7.25



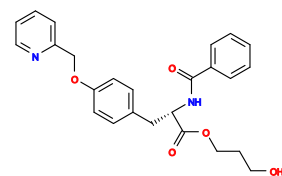
358 V= -6.6, I=-7.02



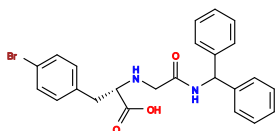
359 V= -6.7, I=-6.95



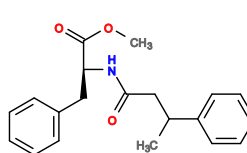
360 V= -6.2, I=-7.32



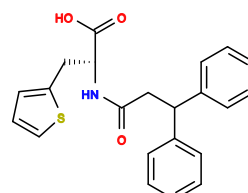
361 V= -6.3, I=-7.24



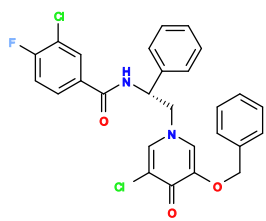
362 V= -6.6, I=-7.02



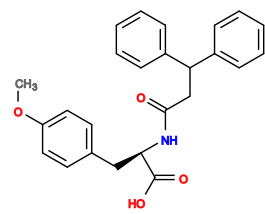
363 V= -5.7, I=-7.69



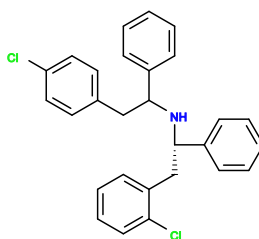
364 V= -6.5, I=-7.09



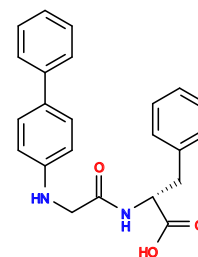
365 V= -7, I=-6.71



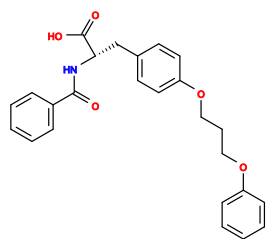
366 V= -6.3, I=-7.23



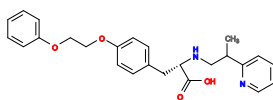
367 V= -6.5, I=-7.08



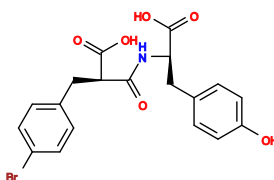
368 V= -6.7, I=-6.93



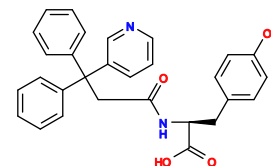
369 V= -5.8, I=-7.61



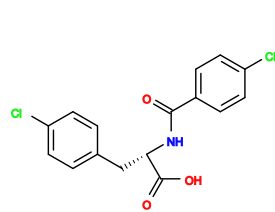
370 V= -5.7, I=-7.68



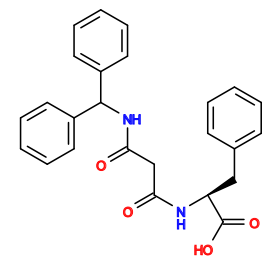
371 V= -6.7, I=-6.92



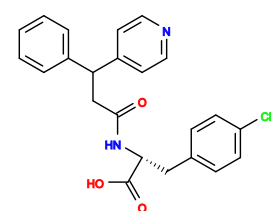
372 V= -6.4, I=-7.15



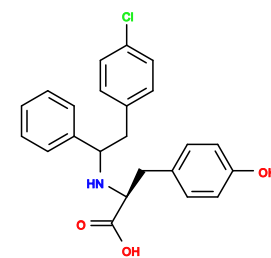
373 V= -6.2, I=-7.29,  
CID73914381



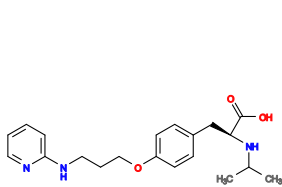
374 V= -6.6, I=-6.99



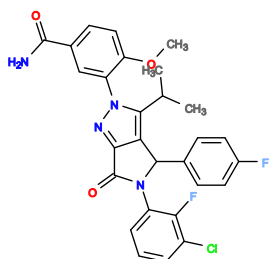
375 V= -6, I=-7.44



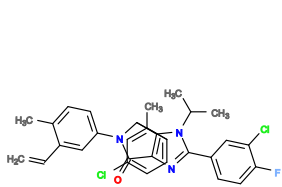
376 V= -5.9, I=-7.52



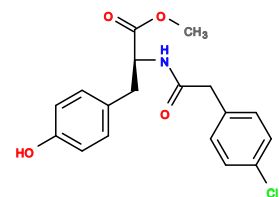
377 V= -5.2, I=-8.04



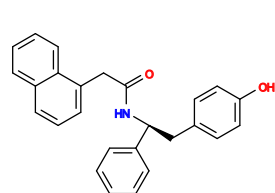
378 V= -7.9, I=-6.02



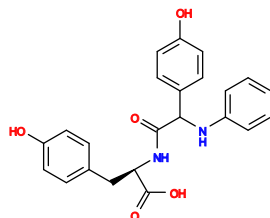
379 V= -8.1, I=-5.86



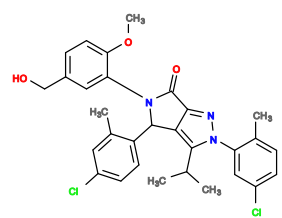
380 V= -5.7, I=-7.66,  
CID25940556



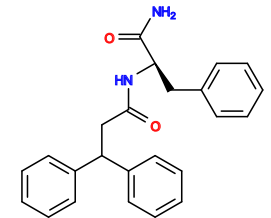
381 V= -7.2, I=-6.54



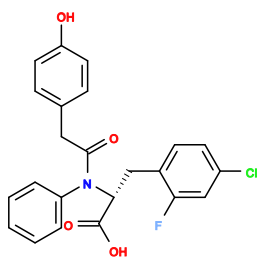
382 V= -7, I=-6.68



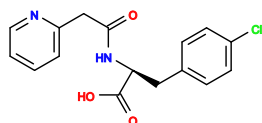
383 V= -8.1, I=-5.86



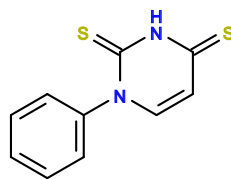
384 V= -6.8, I=-6.83



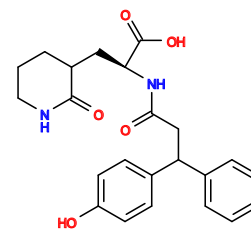
385  $V = -6.4$ ,  $I = -7.13$



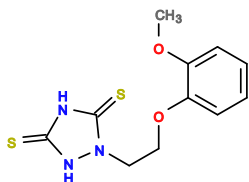
386  $V = -5.8$ ,  $I = -7.58$



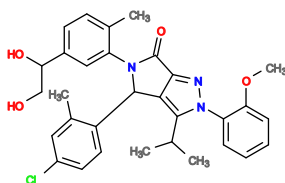
387  $V = -5.4$ ,  $I = -7.88$ ,  
CID19021655



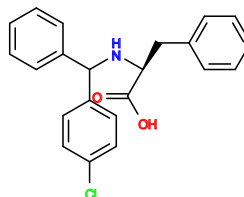
388  $V = -6.5$ ,  $I = -7.05$



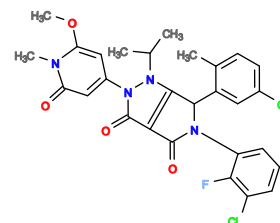
389  $V = -5.2$ ,  $I = -8.02$



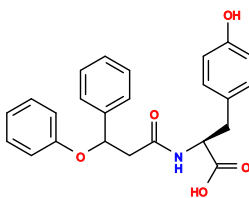
390  $V = -8$ ,  $I = -5.92$



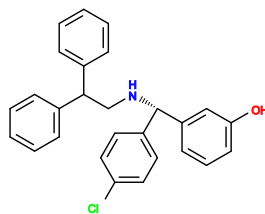
391  $V = -5.7$ ,  $I = -7.65$



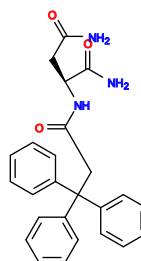
392  $V = -8.1$ ,  $I = -5.85$



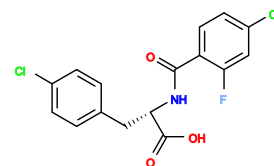
393  $V = -6.7$ ,  $I = -6.9$



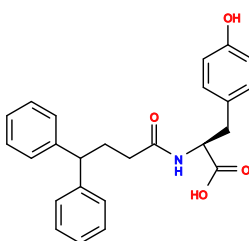
394  $V = -6.8$ ,  $I = -6.82$



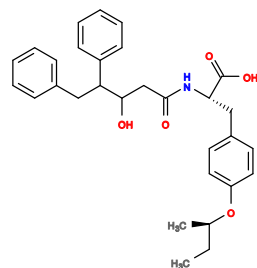
395  $V = -6.4$ ,  $I = -7.12$



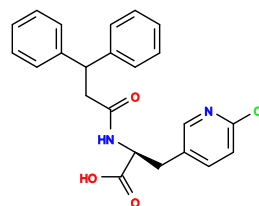
396  $V = -6.4$ ,  $I = -7.11$



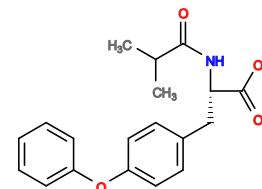
397  $V = -6.8$ ,  $I = -6.81$



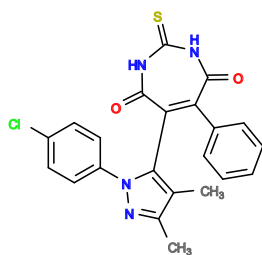
398  $V = -6.3$ ,  $I = -7.18$



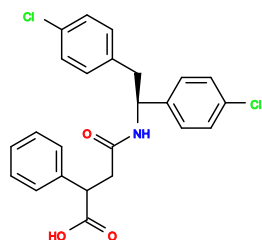
399  $V = -6.1$ ,  $I = -7.33$



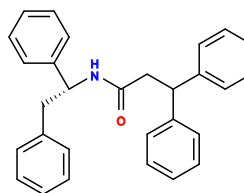
400  $V = -6.6$ ,  $I = -6.96$



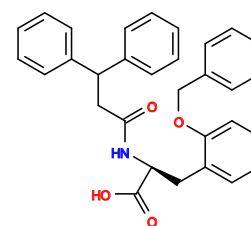
401  $V = -6.8$ ,  $I = -6.81$



402  $V = -6.1$ ,  $I = -7.33$

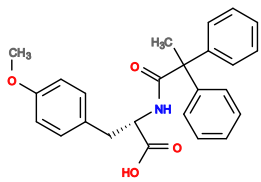


403  $V = -7.2$ ,  $I = -6.5$

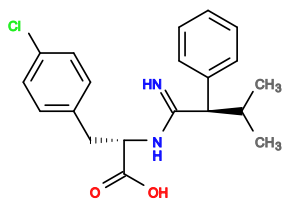


404  $V = -6.3$ ,  $I = -7.18$

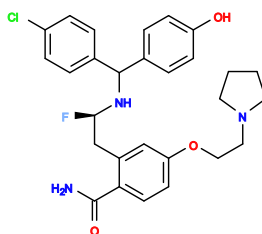
Appendix III: Supporting information of the Generator (Chapter 5)



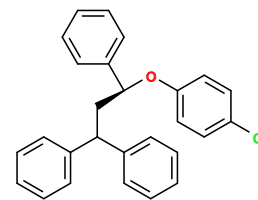
405 V= -6.2, I =-7.25,  
CID73919135



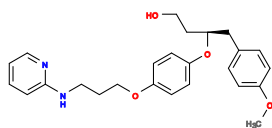
406 V= -5.5, I =-7.78



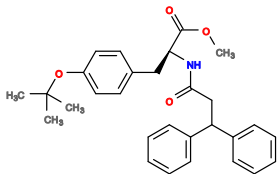
407 V= -6.8, I =-6.8



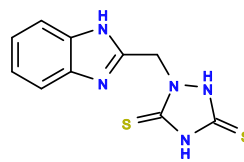
408 V= -6.6, I =-6.95



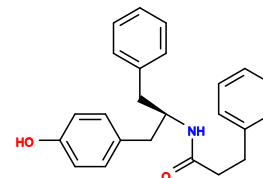
409 V= -6.2, I =-7.25



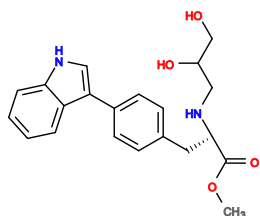
410 V= -6.7, I =-6.87



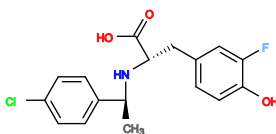
411 V= -5.8, I =-7.54



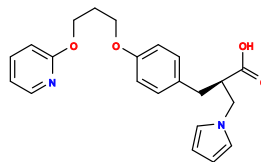
412 V= -7.2, I =-6.49



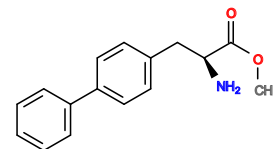
413 V= -5.9, I =-7.46



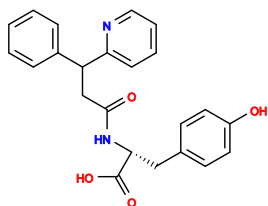
414 V= -6.5, I =-7.01



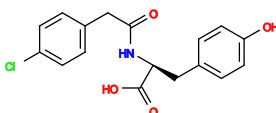
415 V= -5.3, I =-7.91



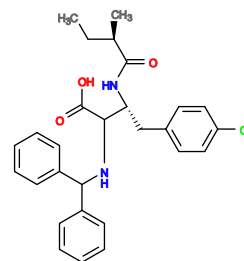
416 V= -6.2, I =-7.23,  
CID15714801



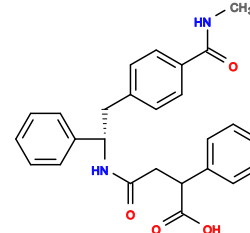
417 V= -6.3, I =-7.16



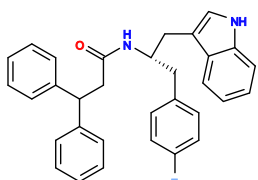
418 V= -5.9, I =-7.46



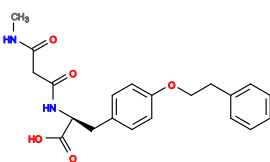
419 V= -5.9, I =-7.46



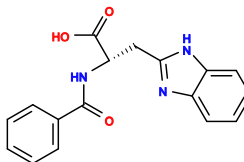
420 V= -6.4, I =-7.08



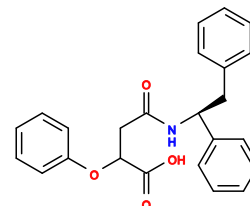
421 V= -6.3, I =-7.15



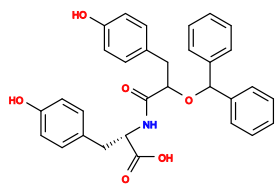
422 V= -6.2, I =-7.23



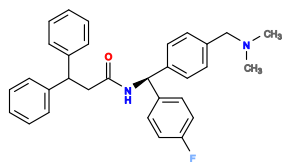
423 V= -6.4, I =-7.08



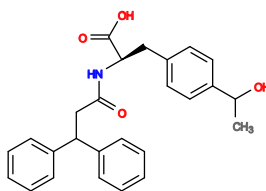
424 V= -6.8, I =-6.78



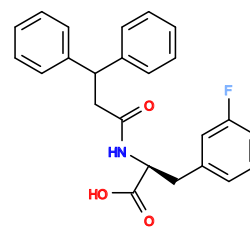
425 V= -6.7, I=-6.85



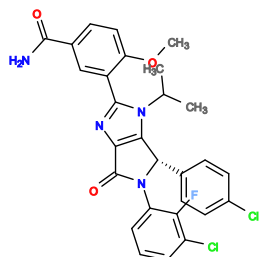
426 V= -7.1, I=-6.55



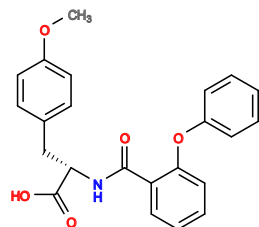
427 V= -6.7, I=-6.85



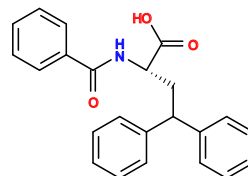
428 V= -6.6, I=-6.92



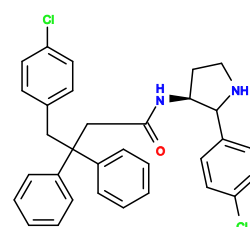
429 V= -7.5, I=-6.24



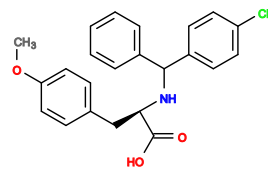
430 V= -6.4, I=-7.07



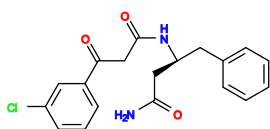
431 V= -6.8, I=-6.77



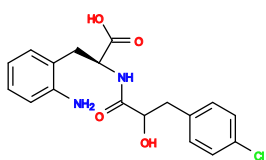
432 V= -7.3, I=-6.39



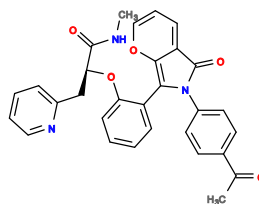
433 V= -6.1, I=-7.29



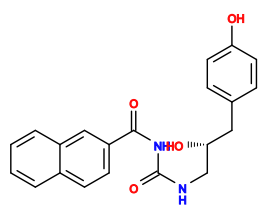
434 V= -6.4, I=-7.06



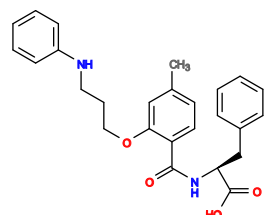
435 V= -6, I=-7.36



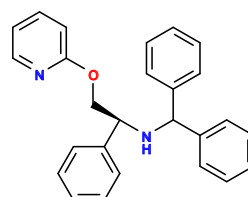
436 V= -7.2, I=-6.46



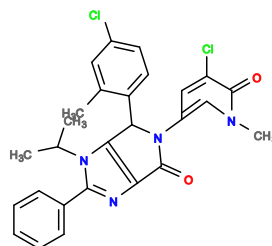
437 V= -7.2, I=-6.45



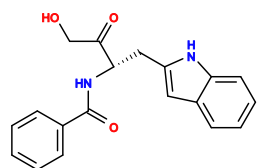
438 V= -5.7, I=-7.58



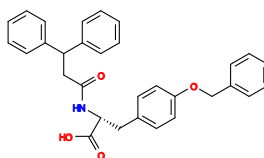
439 V= -6.7, I=-6.83



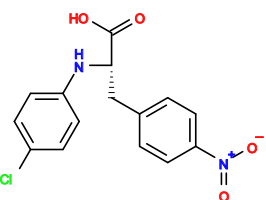
440 V= -7.6, I=-6.15



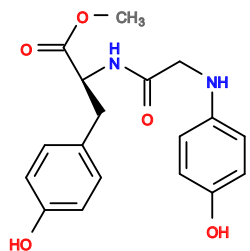
441 V= -6.5, I=-6.98



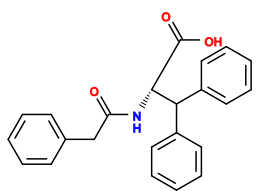
442 V= -6.1, I=-7.27



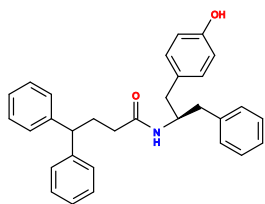
443 V= -6.2, I=-7.2



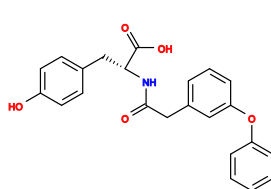
444 V= -5.7, I=-7.57



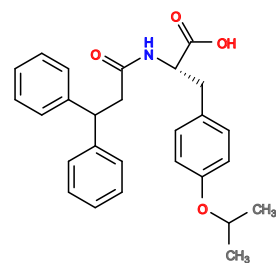
445 V= -6.7, I =-6.82,  
CID58759117



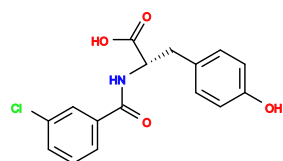
446 V= -7, I =-6.6



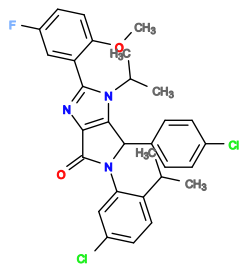
447 V= -6.7, I =-6.82



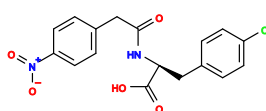
448 V= -6.5, I =-6.97



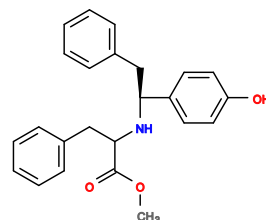
449 V= -6.3, I =-7.12



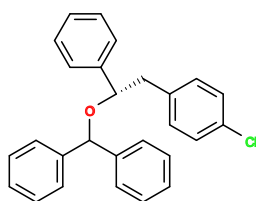
450 V= -7.7, I =-6.07



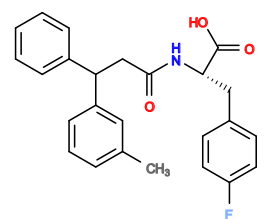
451 V= -6.1, I =-7.27,  
CID1286843



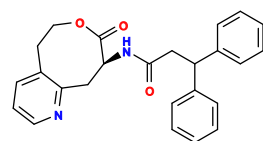
452 V= -6.4, I =-7.04



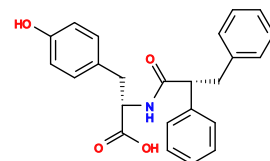
453 V= -6.4, I =-7.04



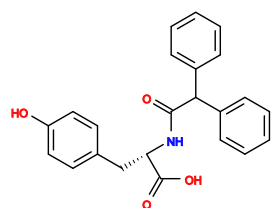
454 V= -6.4, I =-7.04



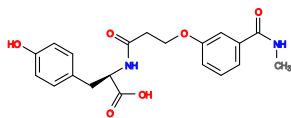
455 V= -7.5, I =-6.21



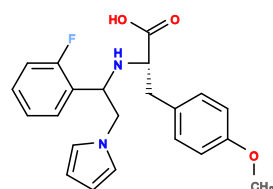
456 V= -6.8, I =-6.74



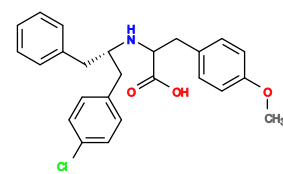
457 V= -6.8, I =-6.74,  
CID16066547



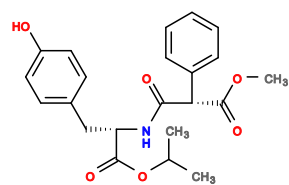
458 V= -6.5, I =-6.96



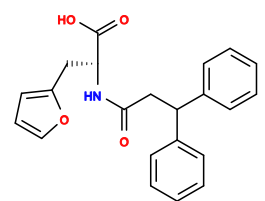
459 V= -6.4, I =-7.03



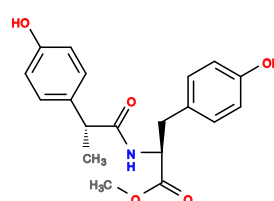
460 V= -6, I =-7.33



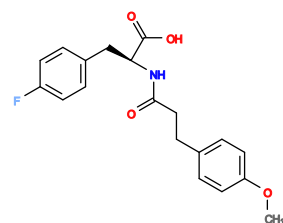
461 V= -5.8, I =-7.48



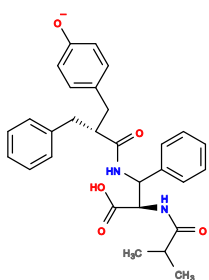
462 V= -6.3, I =-7.1



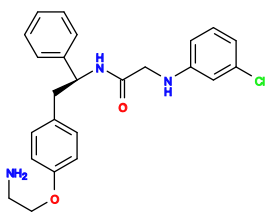
463 V= -5.8, I =-7.47



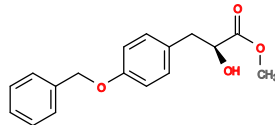
464 V= -6.4, I =-7.02



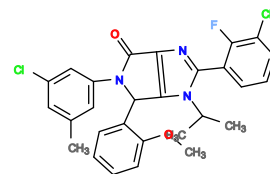
465 V= -6.5, I=-6.94



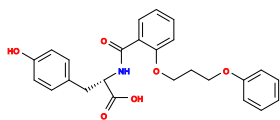
466 V= -6.9, I=-6.64



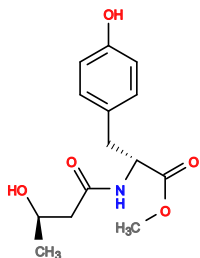
467 V= -5.9, I=-7.39,  
CID11277619



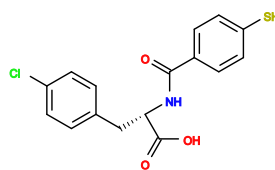
468 V= -7.6, I=-6.11



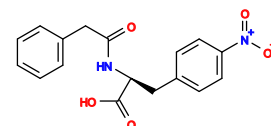
469 V= -6.1, I=-7.24



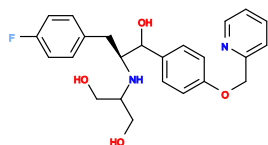
470 V= -5.8, I=-7.46



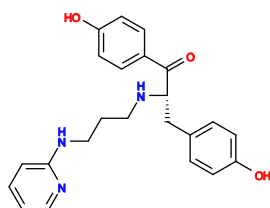
471 V= -6.2, I=-7.16



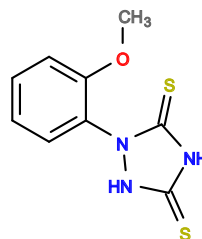
472 V= -6.5, I=-6.93,  
CID73914673



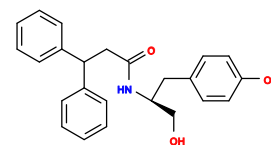
473 V= -5.9, I=-7.38



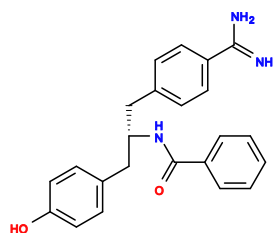
474 V= -5.7, I=-7.53



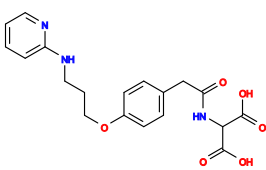
475 V= -5, I=-8.06



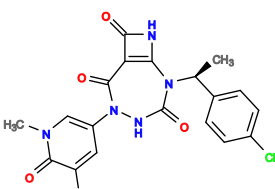
476 V= -6.7, I=-6.78



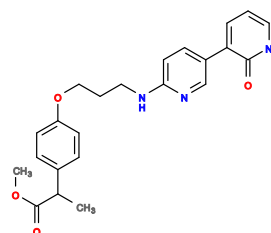
477 V= -6.9, I=-6.63



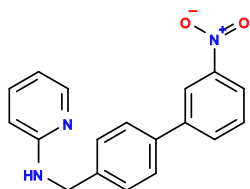
478 V= -5.5, I=-7.68



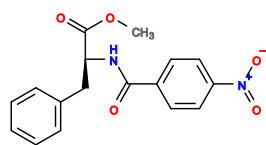
479 V= -6.9, I=-6.63



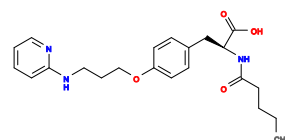
480 V= -7.3, I=-6.33



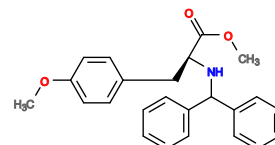
481 V= -6.8, I=-6.7



482 V= -6.1, I=-7.23,  
CID15687991

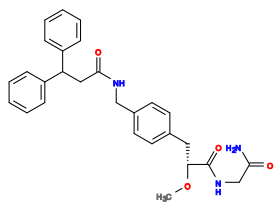


483 V= -5.2, I=-7.9

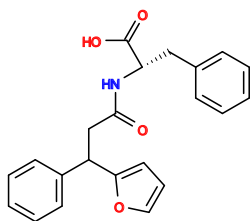


484 V= -5.6, I=-7.6

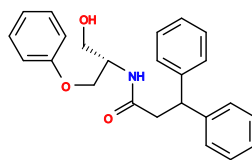




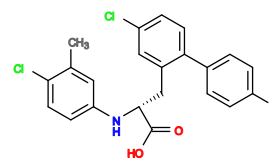
485 V= -6.4, I=-7.0



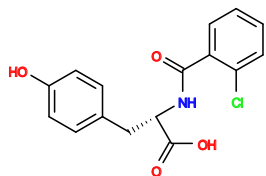
486 V= -6.3, I=-7.07



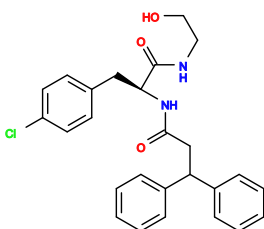
487 V= -6.5, I=-6.92



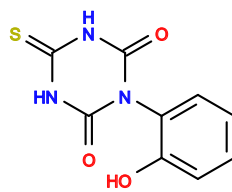
488 V= -6.5, I=-6.92



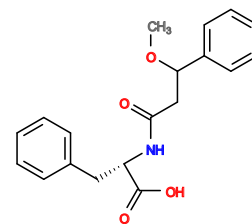
489 V= -6.4, I=-7.0



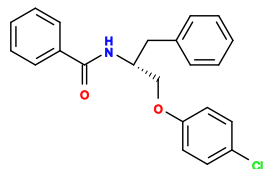
490 V= -5.8, I=-7.45



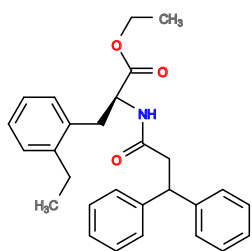
491 V= -5.5, I=-7.67



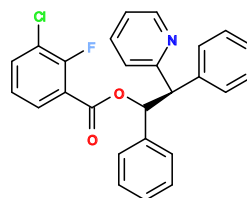
492 V= -5.9, I=-7.37



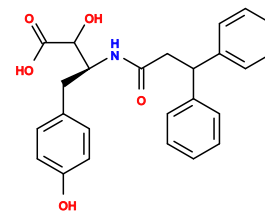
493 V= -6.1, I=-7.22



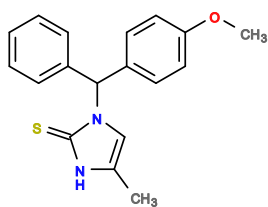
494 V= -5.9, I=-7.37



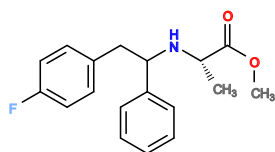
495 V= -7.3, I=-6.31



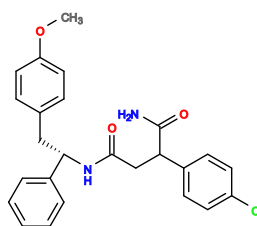
496 V= -6.6, I=-6.84



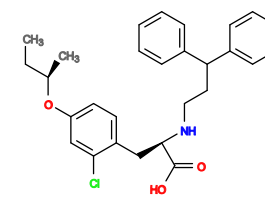
497 V= -5.9, I=-7.35



498 V= -5.9, I=-7.35

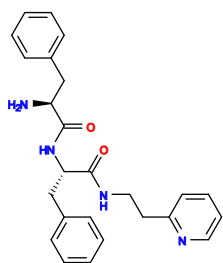


499 V= -6.8, I=-6.68

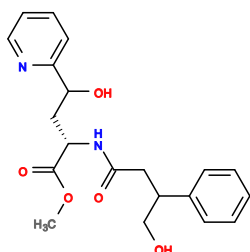


500 V= -6.2, I=-7.13

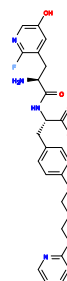
## 3.2 Top-500 compounds from the Run-2



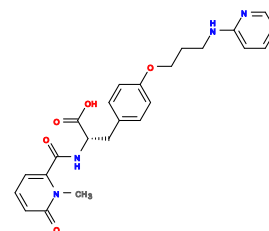
1 V= -6.8, I =-7.66, S =-2.14



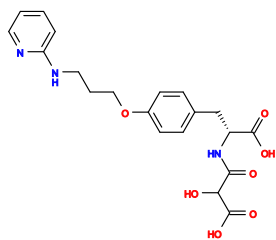
2 V= -6.1, I =-7.64, S =-1.44



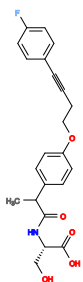
3 V= -6.3, I =-8.22, S =-2.44



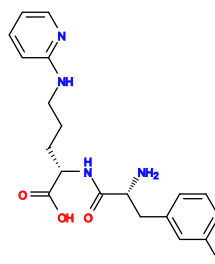
4 V= -6.2, I =-8.37, S =-2.54



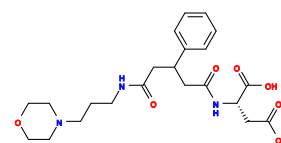
5 V= -5.7, I =-8.33, S =-2.0



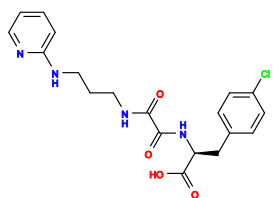
6 V= -6.5, I =-7.23, S =-1.36



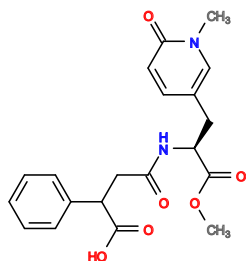
7 V= -6.2, I =-8.01, S =-2.17



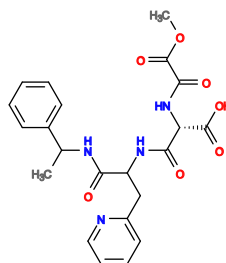
8 V= -6.2, I =-7.34, S =-1.29



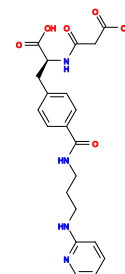
9 V= -6.4, I =-8.11, S =-2.51



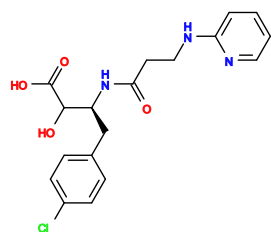
10 V= -6.2, I =-7.72, S =-1.81



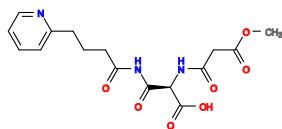
11 V= -6.4, I =-7.58, S =-1.82



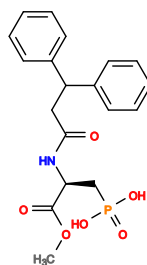
12 V= -5.9, I =-8.19, S =-2.16



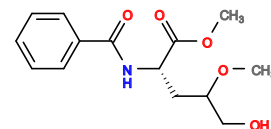
13 V= -6.4, I =-7.9, S =-2.29



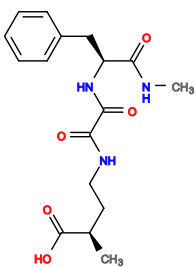
14 V= -5.9, I =-7.59, S =-1.37



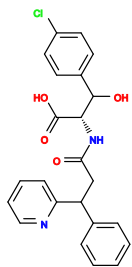
15 V= -6.2, I =-7.43, S =-1.49



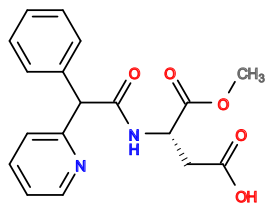
16 V= -5.7, I =-7.31, S =-0.83



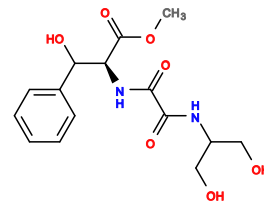
17 V= -6.2, I=-7.22, S =-1.22



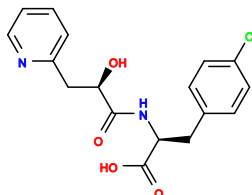
18 V= -6.9, I=-7.95, S =-2.91



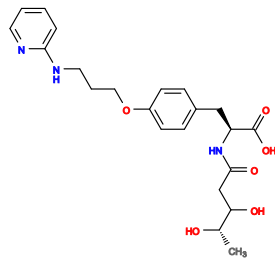
19 V= -6.1, I=-7.66, S =-1.72



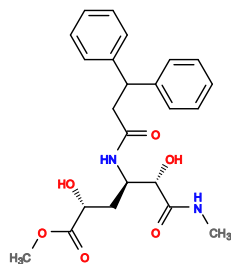
20 V= -5.7, I=-7.33, S =-0.89



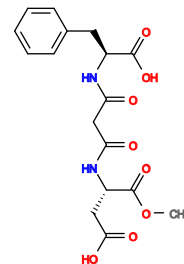
21 V= -6.3, I=-7.89, S =-2.25



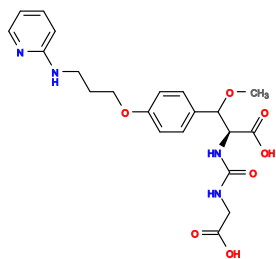
22 V= -5.6, I=-8.37, S =-2.21



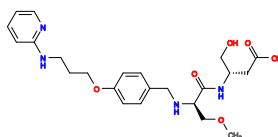
23 V= -6.1, I=-7.43, S =-1.47



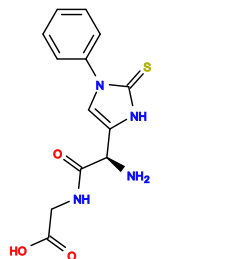
24 V= -5.9, I=-7.56, S =-1.44



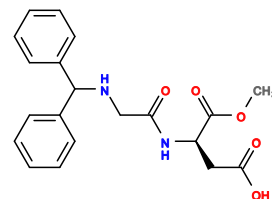
25 V= -6.5, I=-7.99, S =-2.62



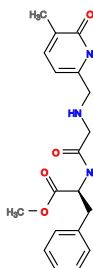
26 V= -5.2, I=-8.15, S =-1.54



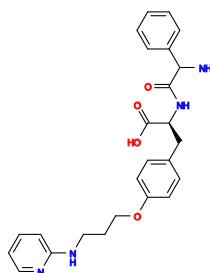
27 V= -5.9, I=-7.87, S =-1.86



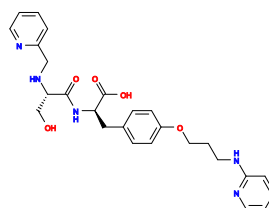
28 V= -6.1, I=-7.65, S =-1.76



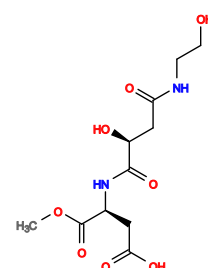
29 V= -6.7, I=-7.73, S =-2.48



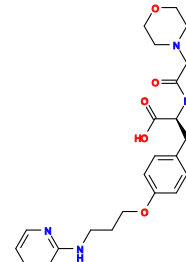
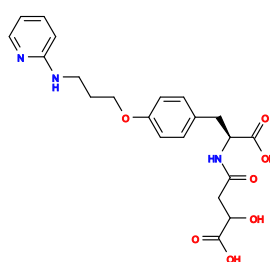
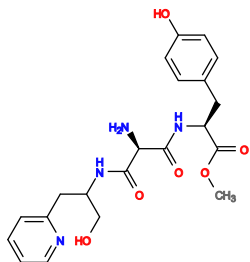
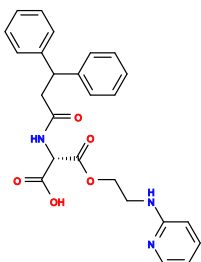
30 V= -5.9, I=-8.42, S =-2.61



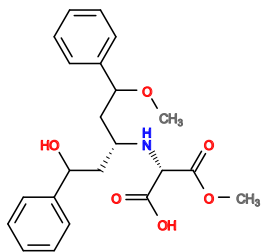
31 V= -5.8, I=-8.3, S =-2.34



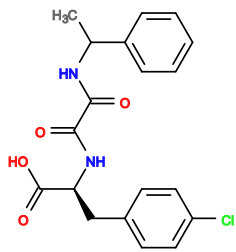
32 V= -4.9, I=-7.33, S =-0.16



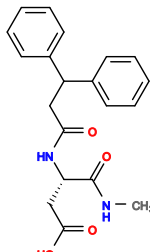
**33** V= -6.6, I =-7.99, S =- 2.74



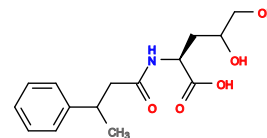
**34** V= -6.0, I =-7.66, S =- 1.73



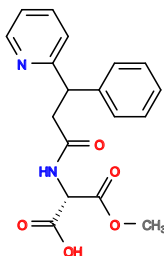
**35** V= -5.8, I =-8.31, S =- 2.4



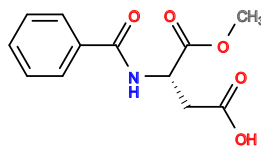
**36** V= -6.2, I =-7.94, S =- 2.32



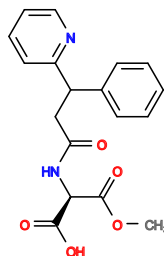
**37** V= -6.0, I =-7.69, S =- 1.79



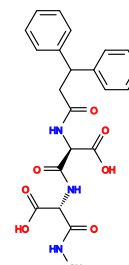
**38** V= -6.8, I =-7.64, S =- 2.52



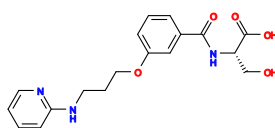
**39** V= -6.4, I =-7.48, S =- 1.92



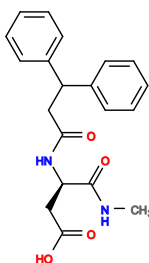
**40** V= -5.9, I =-7.06, S =- 0.87



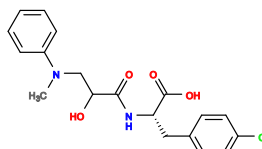
**41** V= -6.1, I =-7.76, S =- 2.0



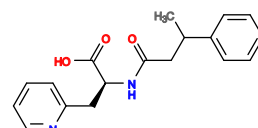
**42** V= -5.7, I =-7.41, S =- 1.14, [CID7019906](#)



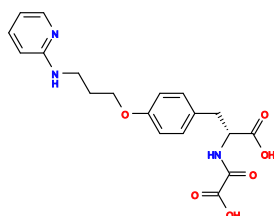
**43** V= -6.0, I =-7.78, S =- 1.94



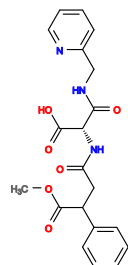
**44** V= -6.5, I =-7.52, S =- 2.09



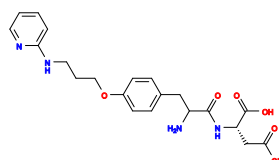
**45** V= -5.7, I =-7.95, S =- 1.87



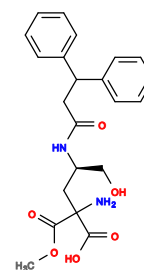
**46** V= -6.4, I =-7.49, S =- 1.95



**47** V= -6.3, I =-7.57, S =- 1.97



**48** V= -6.3, I =-7.49, S =- 1.87



**49** V= -5.7, I =-8.28, S =- 2.32



**50** V= -6.4, I =-7.6, S =- 2.12

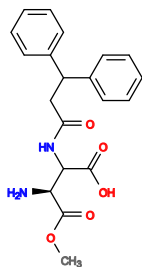


**51** V= -5.5, I =-8.16, S =- 1.98

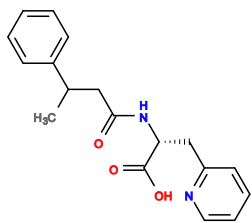


**52** V= -6.2, I =-7.43, S =- 1.71

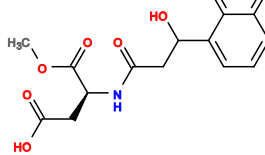




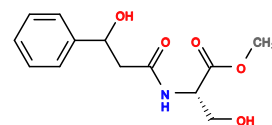
**53** V= -6.4, I =-7.55, S =- 2.06



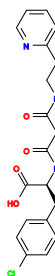
**54** V= -6.3, I =-7.46, S =- 1.84



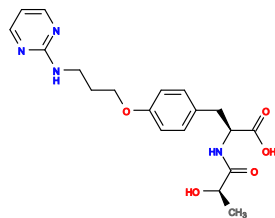
**55** V= -6.1, I =-7.58, S =- 1.8



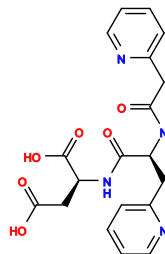
**56** V= -5.3, I =-7.4, S =- 0.76



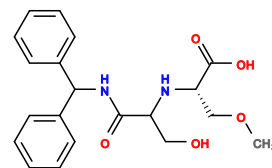
**57** V= -6.4, I =-7.87, S =- 2.5



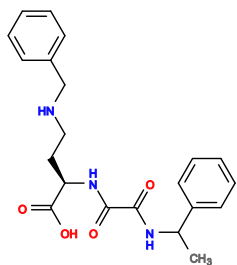
**58** V= -5.7, I =-8.24, S =- 2.3



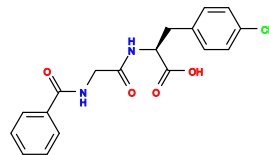
**59** V= -6.2, I =-7.25, S =- 1.47



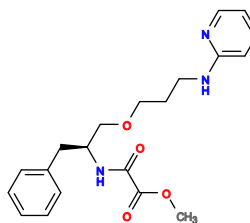
**60** V= -6.2, I =-7.16, S =- 1.35



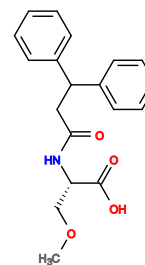
**61** V= -6.2, I =-7.34, S =- 1.59



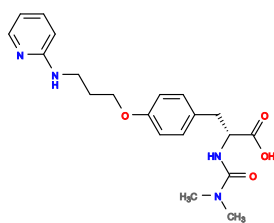
**62** V= -7.0, I =-7.6, S =- 2.75, [CID73908883](#)



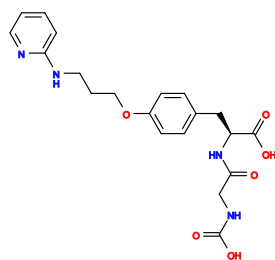
**63** V= -5.8, I =-8.09, S =- 2.19



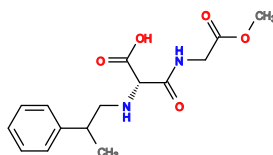
**64** V= -6.1, I =-7.46, S =- 1.67



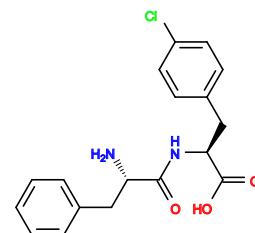
**65** V= -5.8, I =-8.2, S =- 2.36



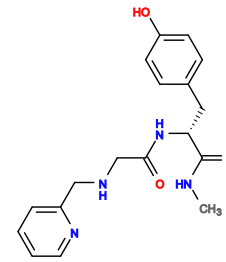
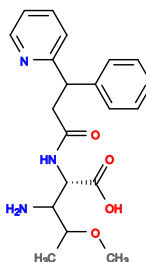
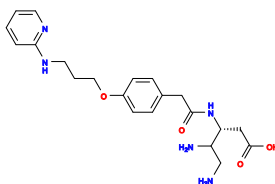
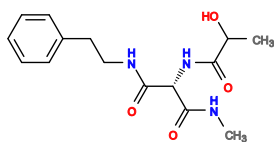
**66** V= -5.7, I =-8.28, S =- 2.38



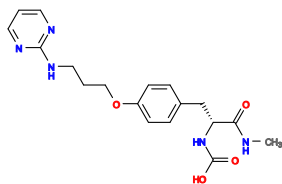
**67** V= -6.2, I =-7.07, S =- 1.27



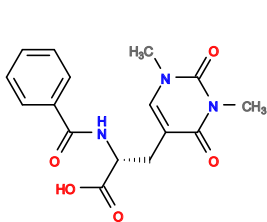
**68** V= -6.4, I =-7.66, S =- 2.25



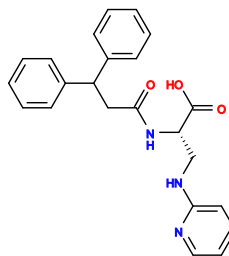
**69** V= -5.7, I =-7.2, S =-0.94



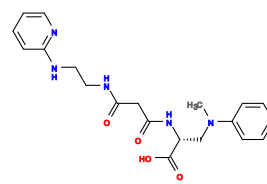
**70** V= -5.7, I =-7.92, S =-1.91



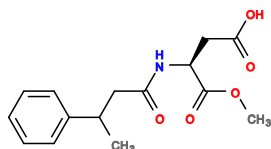
**71** V= -6.2, I =-7.34, S =-1.63



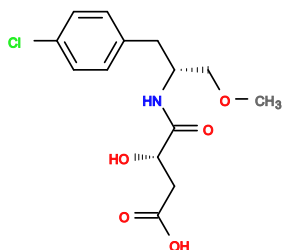
**72** V= -6.2, I =-7.41, S =-1.73



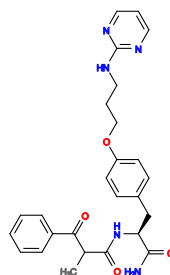
**73** V= -6.0, I =-7.96, S =-2.26



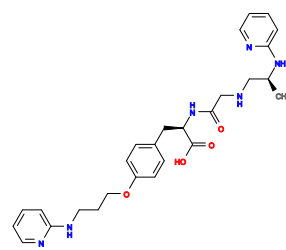
**74** V= -6.4, I =-6.79, S =-1.1



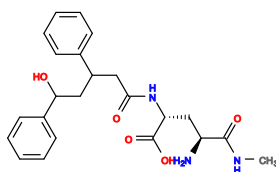
**75** V= -6.8, I =-7.47, S =-2.41



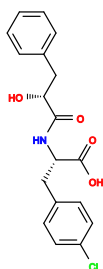
**76** V= -5.5, I =-7.87, S =-1.65



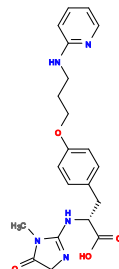
**77** V= -5.8, I =-7.58, S =-1.56



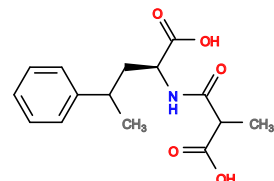
**78** V= -5.8, I =-7.26, S =-1.14



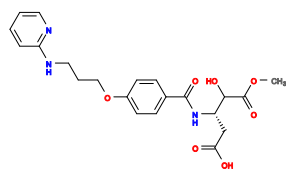
**79** V= -6.2, I =-8.32, S =-2.95



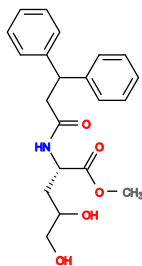
**80** V= -5.7, I =-8.29, S =-2.41



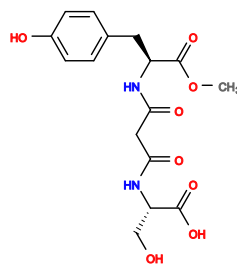
**81** V= -6.3, I =-7.43, S =-1.87



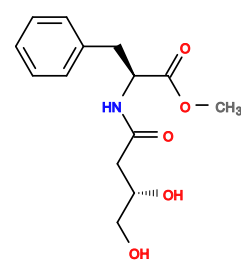
**82** V= -6.5, I =-7.84, S =-2.61



**83** V= -5.8, I =-8.17, S =-2.36



**84** V= -6.4, I =-6.91, S =-1.28



**85** V= -5.8, I =-8.17, S =-2.36



**86** V= -6.1, I =-7.56, S =-1.85

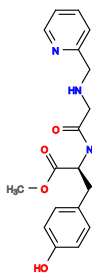


**87** V= -5.9, I =-7.44, S =-1.5

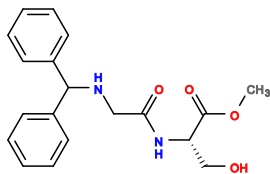


**88** V= -5.5, I =-7.45, S =-1.11

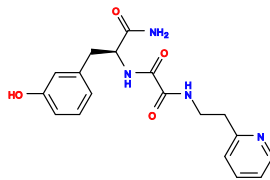




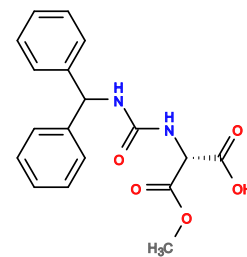
**89** V= -6.1, I=-7.69, S=-2.02



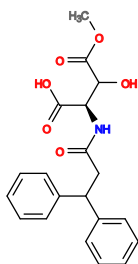
**90** V= -6.1, I=-7.43, S=-1.69



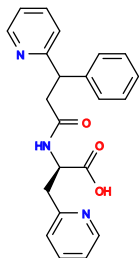
**91** V= -6.7, I=-7.04, S=-1.77



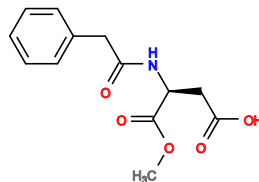
**92** V= -6.4, I=-7.59, S=-2.2



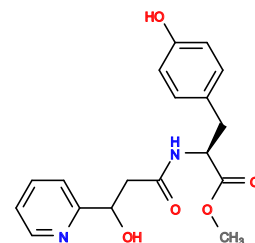
**93** V= -6.2, I=-7.54, S=-1.95



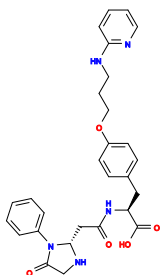
**94** V= -6.8, I=-7.36, S=-2.3



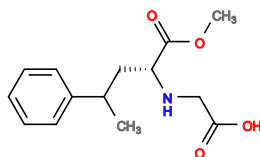
**95** V= -5.5, I=-7.51, S=-1.21



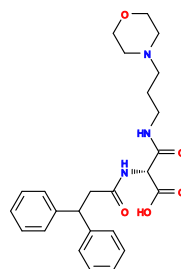
**96** V= -6.2, I=-7.66, S=-2.11



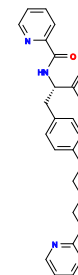
**97** V= -6.2, I=-8.21, S=-2.85



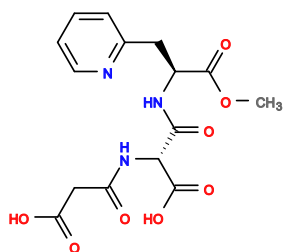
**98** V= -5.8, I=-7.44, S=-1.42



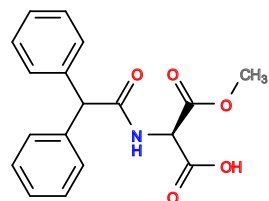
**99** V= -6.5, I=-7.39, S=-2.05



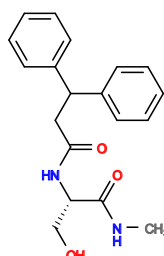
**100** V= -6.4, I=-8.0, S=-2.77



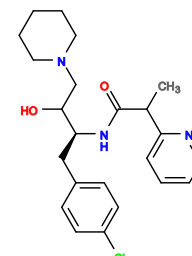
**101** V= -5.4, I=-7.67, S=-1.33



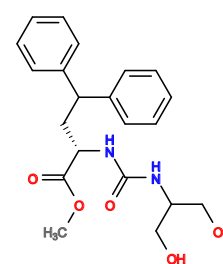
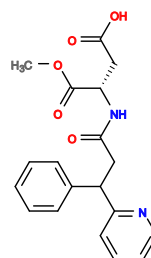
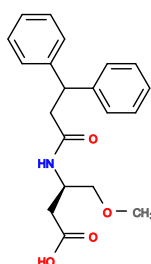
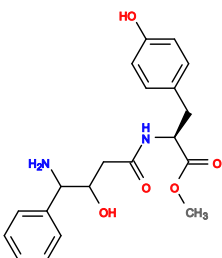
**102** V= -6.3, I=-7.57, S=-2.1



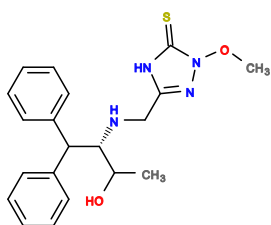
**103** V= -6.3, I=-7.29, S=-1.73



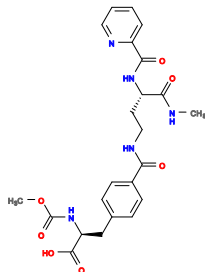
**104** V= -6.7, I=-7.4, S=-2.27



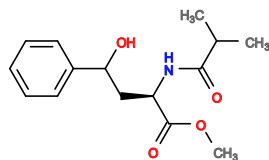
**105** V= -6.3, I=-7.6, S =-2.14



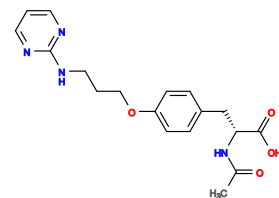
**106** V= -6.1, I=-7.4, S =-1.68



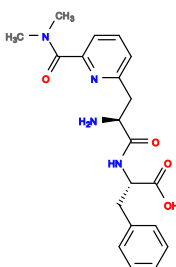
**107** V= -5.8, I=-7.76, S =-1.86



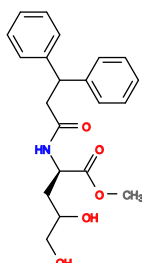
**108** V= -6.2, I=-7.59, S =-2.03



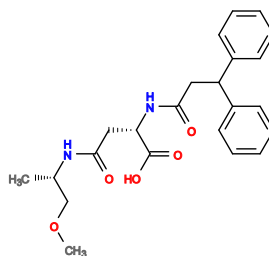
**109** V= -6.5, I=-7.82, S =-2.65



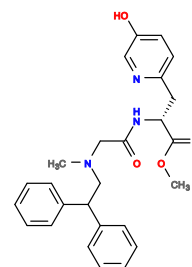
**110** V= -6.9, I=-7.21, S =-2.24



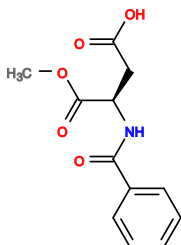
**111** V= -5.8, I=-7.44, S =-1.44



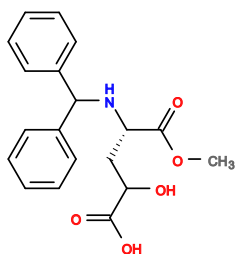
**112** V= -6.2, I=-7.86, S =-2.41



**113** V= -7.1, I=-6.29, S =-1.21



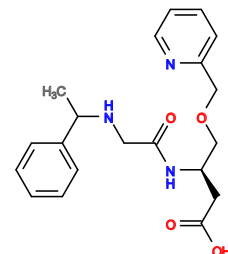
**114** V= -6.0, I=-7.58, S =-1.84



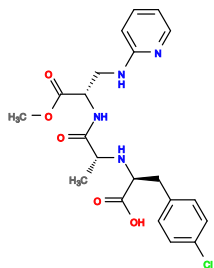
**115** V= -6.3, I=-7.51, S =-2.05



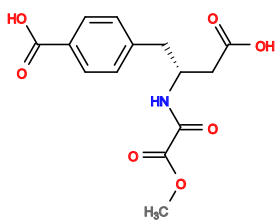
**116** V= -6.5, I=-7.37, S =-2.06



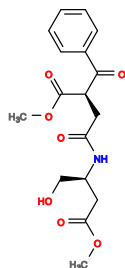
**117** V= -5.5, I=-7.4, S =-1.11, [CID124500287](#)



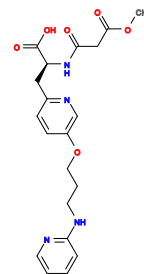
**118** V= -5.8, I=-7.61, S =-1.68



**119** V= -5.9, I=-7.44, S =-1.56



**120** V= -5.9, I=-7.4, S =-1.51



**121** V= -6.2, I=-7.88, S =-2.44



**122** V= -5.9, I=-7.19, S =-1.23



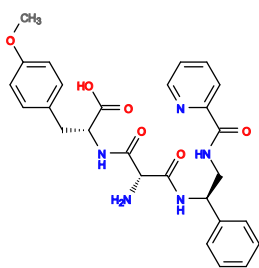
**123** V= -5.7, I=-7.3, S =-1.18



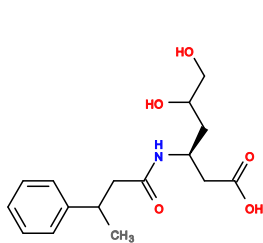
**124** V= -5.8, I=-8.19, S =-2.46



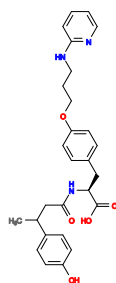




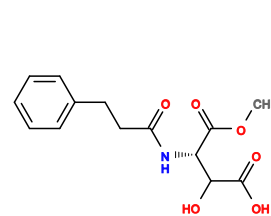
**125**  $V = -6.9$ ,  $I = -7.32$ ,  $S = -2.41$



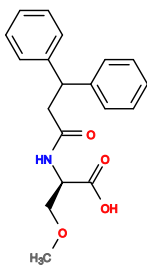
**126**  $V = -5.9$ ,  $I = -6.92$ ,  $S = -0.87$



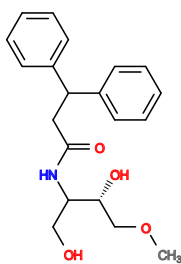
**127**  $V = -6.8$ ,  $I = -8.42$ ,  $S = -3.77$



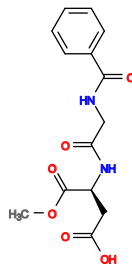
**128**  $V = -5.6$ ,  $I = -7.47$ ,  $S = -1.3$



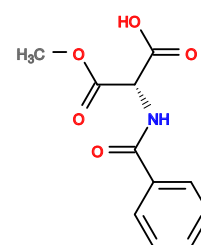
**129**  $V = -6.0$ ,  $I = -7.43$ ,  $S = -1.66$



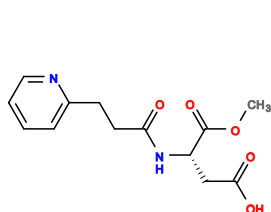
**130**  $V = -6.1$ ,  $I = -6.89$ ,  $S = -1.05$



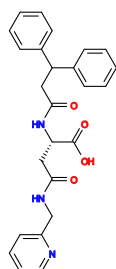
**131**  $V = -5.4$ ,  $I = -7.47$ ,  $S = -1.13$



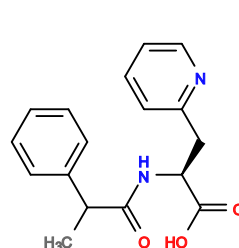
**132**  $V = -5.7$ ,  $I = -7.36$ ,  $S = -1.28$



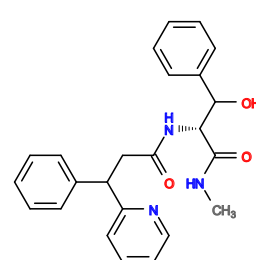
**133**  $V = -5.3$ ,  $I = -7.53$ ,  $S = -1.11$



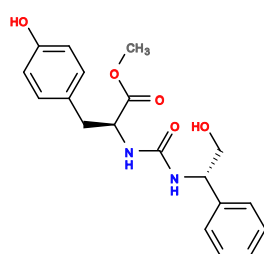
**134**  $V = -6.7$ ,  $I = -7.53$ ,  $S = -2.5$



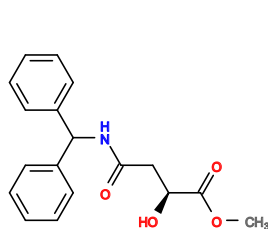
**135**  $V = -6.3$ ,  $I = -7.24$ ,  $S = -1.73$



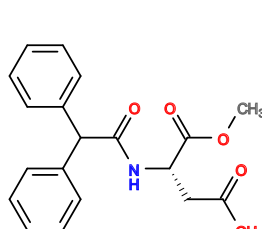
**136**  $V = -6.5$ ,  $I = -7.65$ ,  $S = -2.46$



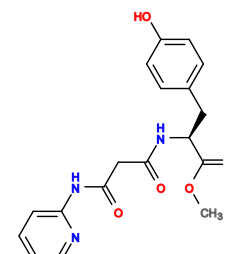
**137**  $V = -6.1$ ,  $I = -7.6$ ,  $S = -2.01$



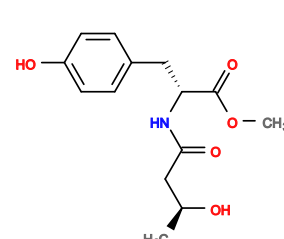
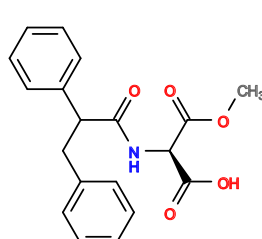
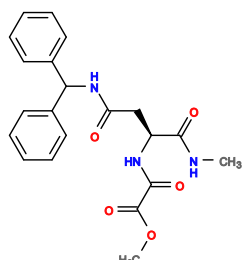
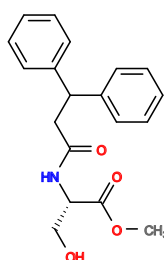
**138**  $V = -6.3$ ,  $I = -7.37$ ,  $S = -1.9$



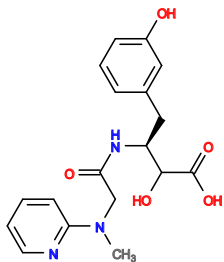
**139**  $V = -6.1$ ,  $I = -7.57$ ,  $S = -1.96$



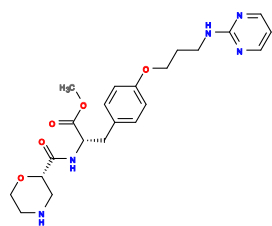
**140**  $V = -6.5$ ,  $I = -7.64$ ,  $S = -2.46$



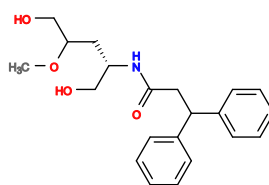
**141** V= -6.1, I=-7.57, S =-1.96



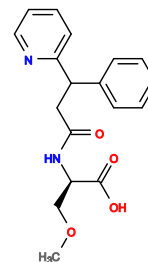
**142** V= -6.5, I=-7.35, S =-2.08



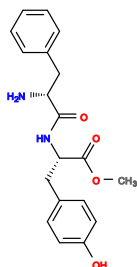
**143** V= -6.4, I=-7.58, S =-2.29



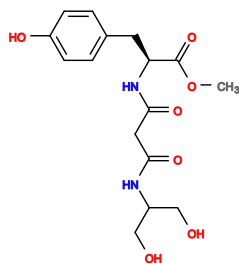
**144** V= -5.7, I=-7.46, S =-1.43



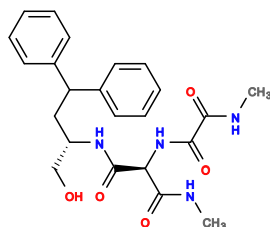
**145** V= -6.4, I=-7.13, S =-1.69



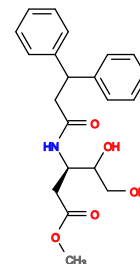
**146** V= -6.0, I=-8.06, S =-2.53



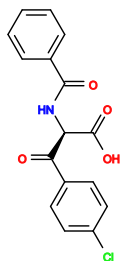
**147** V= -6.1, I=-7.16, S =-1.43



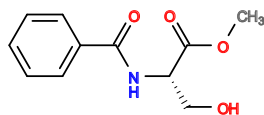
**148** V= -5.8, I=-7.44, S =-1.51



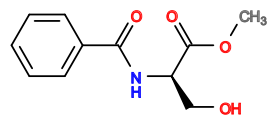
**149** V= -6.4, I=-7.61, S =-2.34



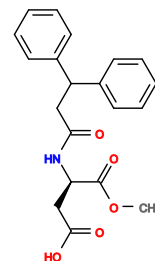
**150** V= -5.9, I=-7.4, S =-1.56



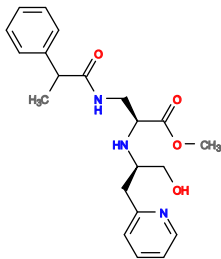
**151** V= -6.5, I=-7.02, S =-1.65



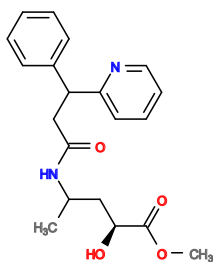
**152** V= -6.1, I=-7.47, S =-1.86



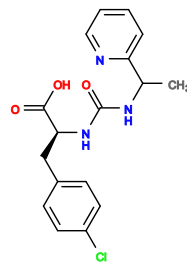
**153** V= -6.9, I=-7.39, S =-2.55



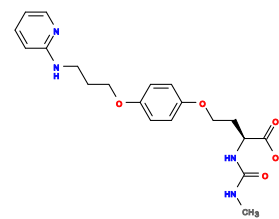
**154** V= -5.5, I=-7.11, S =-0.77, [CID2222571](#)



**155** V= -5.4, I=-7.1, S =-0.67, [CID2222570](#)



**156** V= -6.1, I=-7.66, S =-2.11



**157** V= -5.6, I=-7.6, S =-1.53



**158** V= -6.1, I=-7.51, S =-1.91

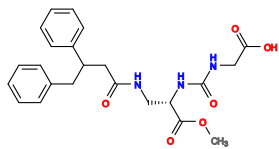


**159** V= -6.5, I=-7.68, S =-2.55

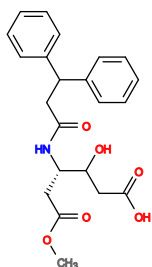


**160** V= -6.1, I=-8.0, S =-2.57

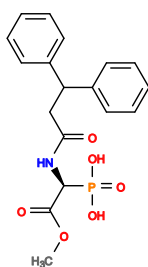




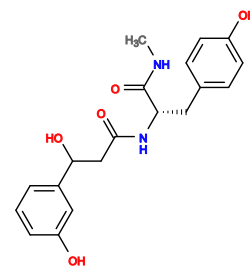
**161** V= -6.6, I=-7.57, S =- 2.5



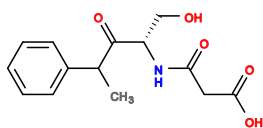
**162** V= -6.2, I=-7.53, S =- 2.05



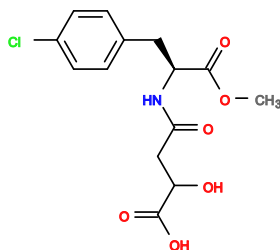
**163** V= -6.0, I=-7.34, S =- 1.59



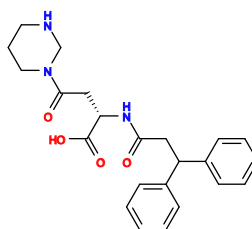
**164** V= -6.5, I=-7.33, S =- 2.09



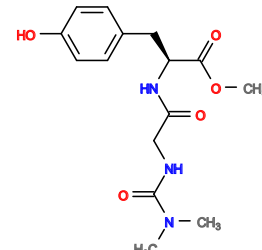
**165** V= -6.1, I=-6.77, S =- 0.94



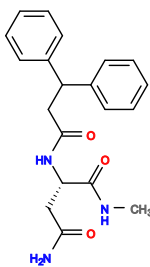
**166** V= -6.0, I=-7.73, S =- 2.11



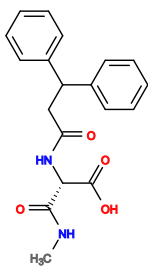
**167** V= -6.8, I=-7.18, S =- 2.19



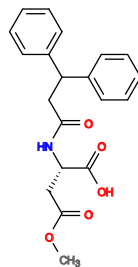
**168** V= -5.9, I=-7.26, S =- 1.39



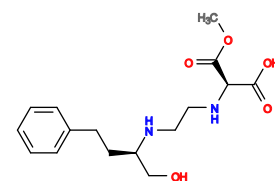
**169** V= -6.4, I=-7.36, S =- 2.03



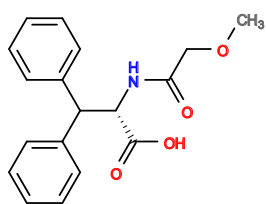
**170** V= -6.3, I=-7.48, S =- 2.08



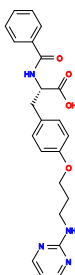
**171** V= -6.2, I=-7.59, S =- 2.14



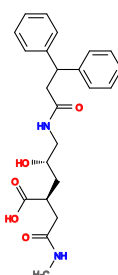
**172** V= -5.4, I=-7.47, S =- 1.18



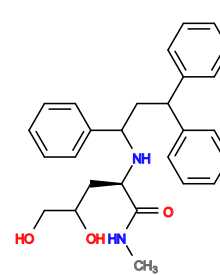
**173** V= -6.4, I=-7.33, S =- 1.99, [CID73909115](#)



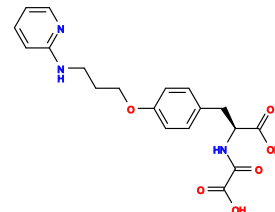
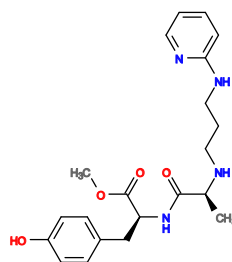
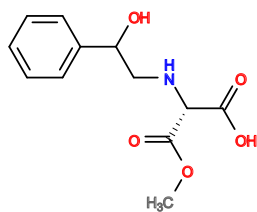
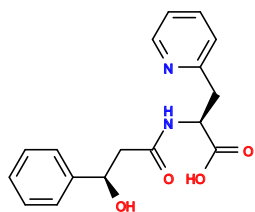
**174** V= -6.7, I=-8.02, S =- 3.22



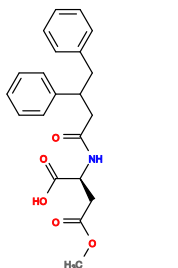
**175** V= -6.2, I=-7.27, S =- 1.71



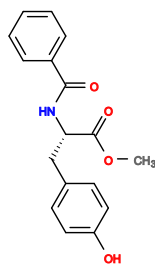
**176** V= -6.4, I=-7.46, S =- 2.16



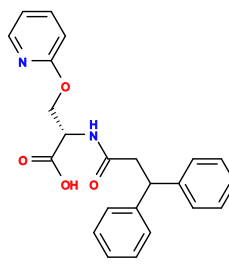
**177** V= -6.1, I=-7.13, S =-1.44



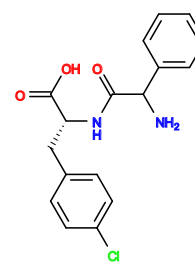
**178** V= -5.4, I=-7.41, S =-1.11



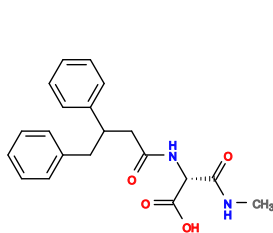
**179** V= -5.7, I=-8.09, S =-2.31



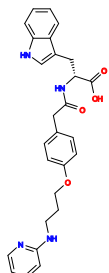
**180** V= -5.5, I=-8.29, S =-2.39



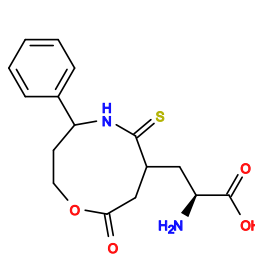
**181** V= -6.4, I=-7.58, S =-2.33



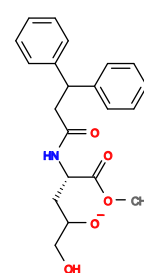
**182** V= -6.8, I=-7.53, S =-2.67, [CID691603](#)



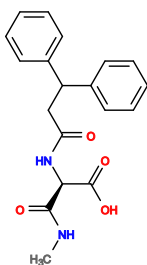
**183** V= -6.9, I=-7.46, S =-2.68



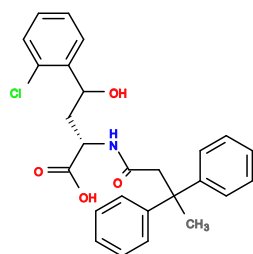
**184** V= -5.9, I=-7.79, S =-2.12



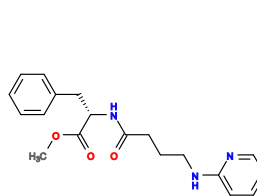
**185** V= -6.5, I=-7.41, S =-2.21



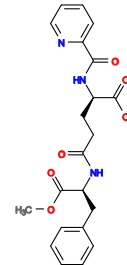
**186** V= -6.8, I=-8.22, S =-3.59



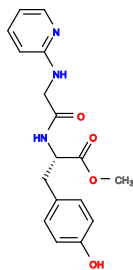
**187** V= -6.9, I=-7.22, S =-2.36



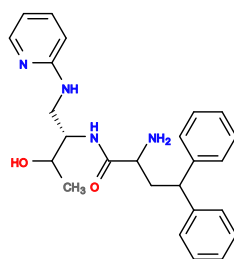
**188** V= -5.9, I=-7.54, S =-1.79



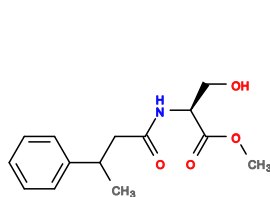
**189** V= -6.3, I=-7.45, S =-2.07



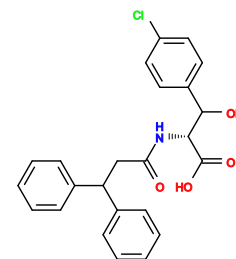
**190** V= -7.1, I=-7.83, S =-3.38



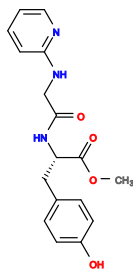
**191** V= -6.3, I=-7.98, S =-2.78



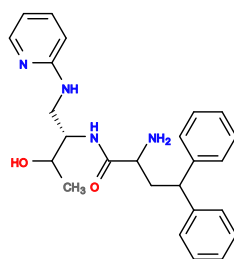
**192** V= -6.2, I=-7.74, S =-2.36



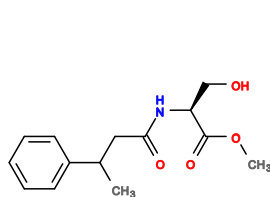
**193** V= -6.3, I=-7.61, S =-2.29



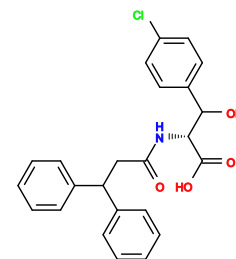
**194** V= -6.6, I=-6.96, S =-1.72

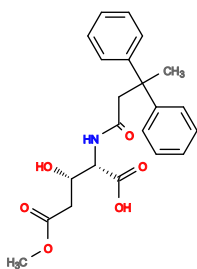


**195** V= -5.5, I=-7.42, S =-1.24

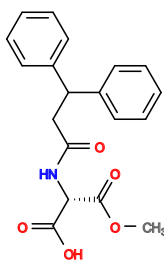


**196** V= -6.9, I=-7.86, S =-3.22

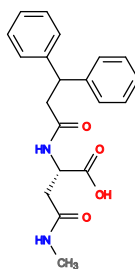




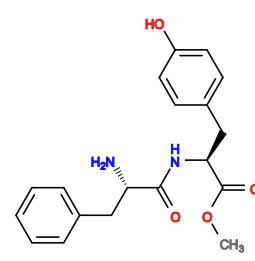
**197** V= -6.3, I=-7.69, S =- 2.41



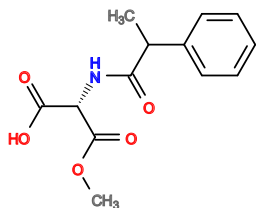
**198** V= -6.2, I=-7.65, S =- 2.25



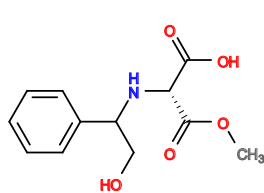
**199** V= -6.4, I=-7.33, S =- 2.02



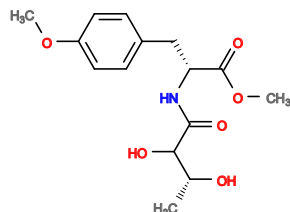
**200** V= -6.3, I=-7.62, S =- 2.3 , [CID13341189](#)



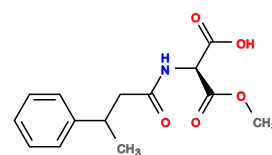
**201** V= -5.4, I=-7.48, S =- 1.22



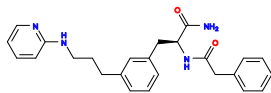
**202** V= -5.2, I=-7.44, S =- 0.97



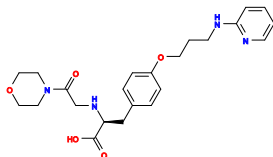
**203** V= -5.9, I=-7.41, S =- 1.64



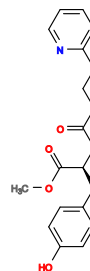
**204** V= -5.6, I=-7.58, S =- 1.56



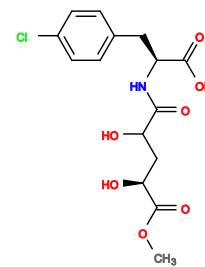
**205** V= -6.5, I=-8.05, S =- 3.09



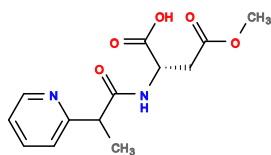
**206** V= -6.0, I=-7.84, S =- 2.32



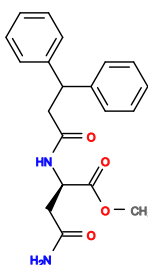
**207** V= -6.1, I=-7.82, S =- 2.39



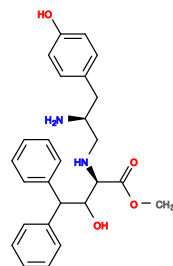
**208** V= -5.7, I=-7.48, S =- 1.54



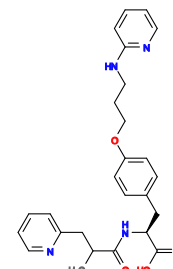
**209** V= -5.4, I=-7.15, S =- 0.8



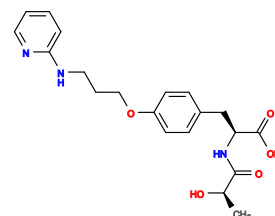
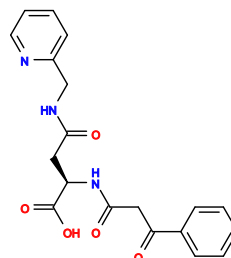
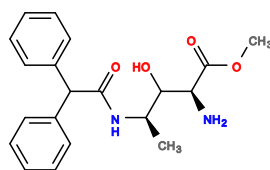
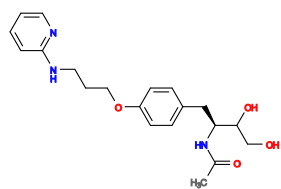
**210** V= -6.2, I=-7.61, S =- 2.21



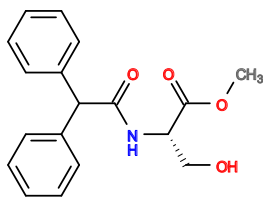
**211** V= -6.5, I=-7.55, S =- 2.43



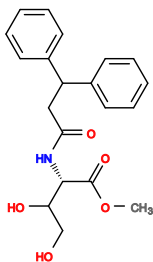
**212** V= -6.0, I=-8.44, S =- 3.12



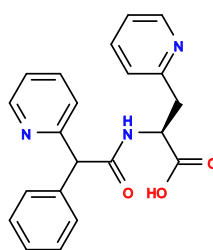
**213** V= -5.8, I=-8.06, S=-2.4



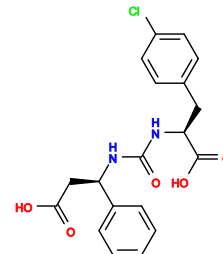
**214** V= -6.3, I=-7.4, S=-2.03



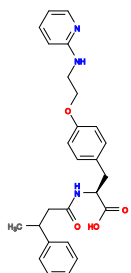
**215** V= -6.2, I=-7.35, S=-1.86



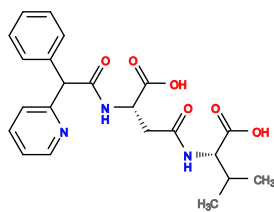
**216** V= -5.5, I=-8.33, S=-2.47



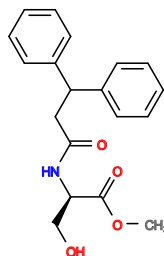
**217** V= -6.0, I=-7.42, S=-1.75, [CID22871490](#)



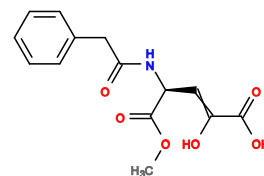
**218** V= -6.0, I=-7.55, S=-1.93



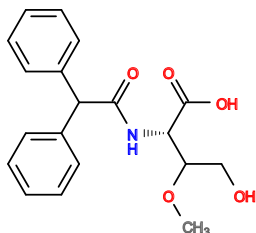
**219** V= -6.7, I=-7.19, S=-2.15



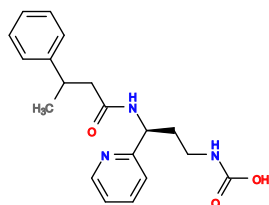
**220** V= -6.5, I=-7.73, S=-2.68



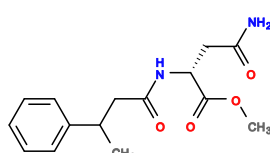
**221** V= -6.4, I=-8.46, S=-3.55



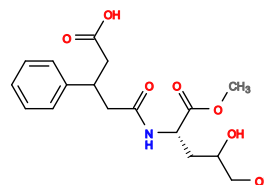
**222** V= -6.2, I=-7.28, S=-1.78



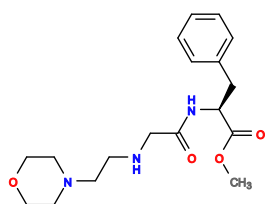
**223** V= -5.9, I=-7.57, S=-1.87



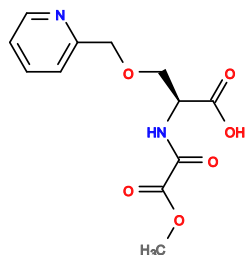
**224** V= -5.9, I=-7.5, S=-1.77



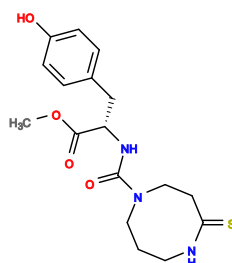
**225** V= -6.2, I=-7.24, S=-1.73



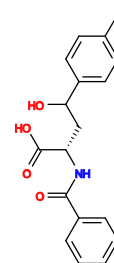
**226** V= -6.4, I=-7.15, S=-1.81



**227** V= -5.8, I=-7.51, S=-1.69



**228** V= -5.3, I=-7.46, S=-1.13



**229** V= -5.8, I=-7.43, S=-1.58



**230** V= -5.2, I=-7.39, S=-0.94

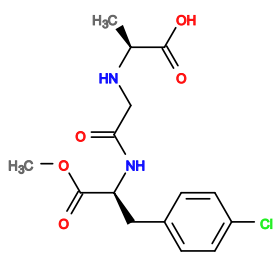


**231** V= -6.6, I=-7.59, S=-2.6

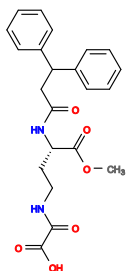


**232** V= -6.5, I=-7.44, S=-2.3

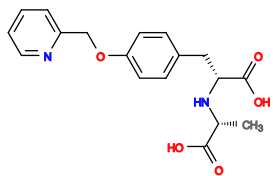




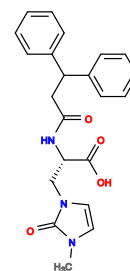
**233** V= -5.7, I=-7.68, S =-1.82



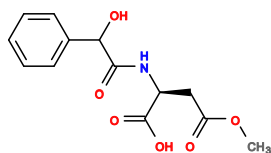
**234** V= -6.2, I=-7.66, S =-2.29



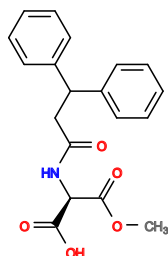
**235** V= -6.1, I=-7.65, S =-2.18



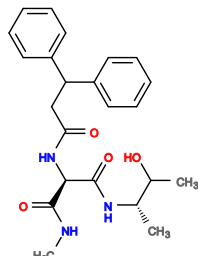
**236** V= -6.8, I=-6.97, S =-1.98



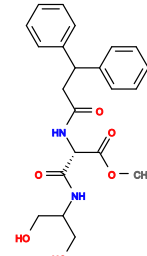
**237** V= -5.4, I=-7.28, S =-1.0, [CID63307126](#)



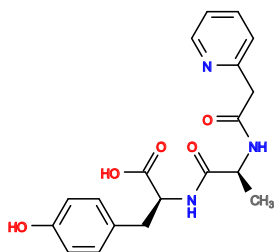
**238** V= -6.1, I=-7.66, S =-2.2



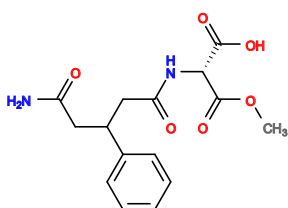
**239** V= -6.5, I=-7.46, S =-2.34



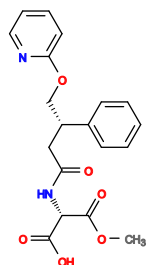
**240** V= -6.0, I=-7.64, S =-2.08



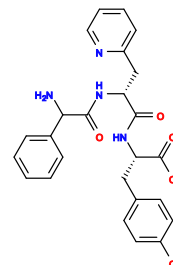
**241** V= -6.5, I=-7.17, S =-1.95



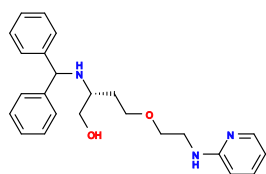
**242** V= -5.6, I=-7.43, S =-1.4



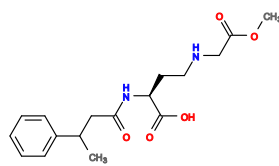
**243** V= -5.8, I=-7.83, S =-2.14



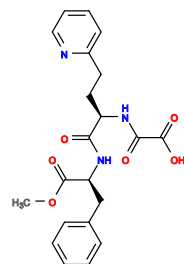
**244** V= -6.5, I=-7.26, S =-2.07



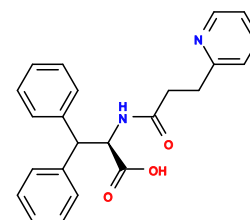
**245** V= -6.1, I=-7.35, S =-1.79



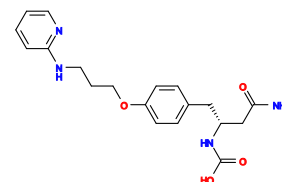
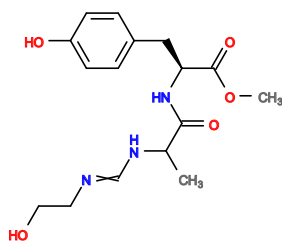
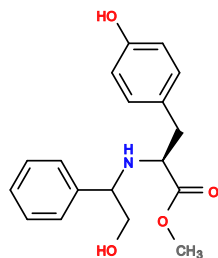
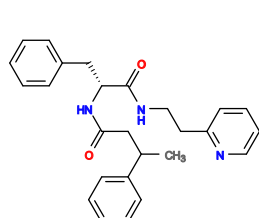
**246** V= -5.7, I=-7.36, S =-1.42



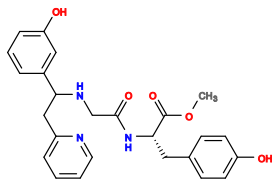
**247** V= -6.0, I=-7.76, S =-2.25



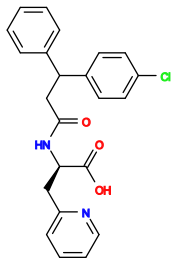
**248** V= -7.1, I=-7.12, S =-2.49



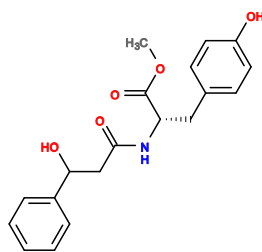
**249** V= -6.8, I=-7.92, S =-3.26



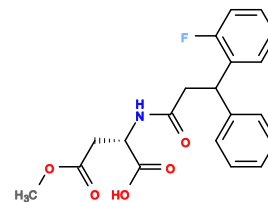
**250** V= -6.1, I=-7.56, S =-2.08



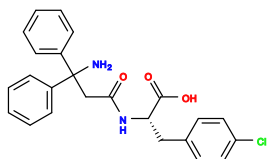
**251** V= -6.1, I=-7.33, S =-1.78



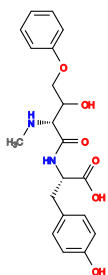
**252** V= -5.6, I=-8.14, S =-2.35



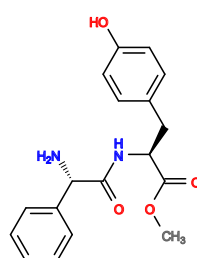
**253** V= -6.7, I=-7.75, S =-2.94



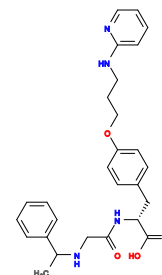
**254** V= -7.1, I=-7.87, S =-3.51



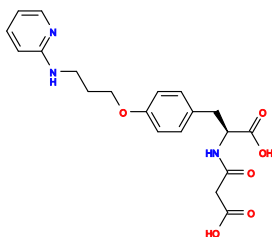
**255** V= -6.3, I=-7.65, S =-2.41



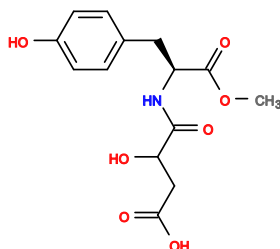
**256** V= -6.5, I=-7.54, S =-2.46



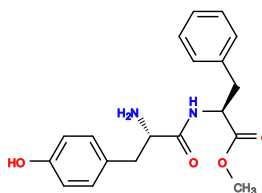
**257** V= -6.7, I=-8.02, S =-3.3



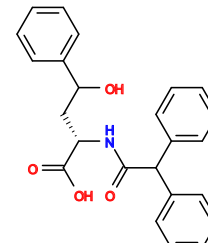
**258** V= -6.2, I=-7.05, S =-1.5



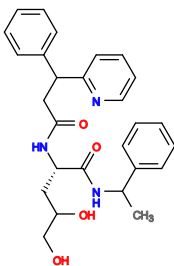
**259** V= -6.2, I=-7.56, S =-2.19



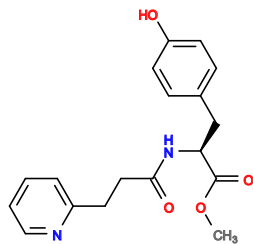
**260** V= -5.9, I=-8.41, S =-3.03



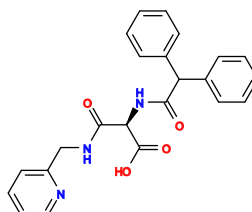
**261** V= -5.5, I=-8.33, S =-2.52



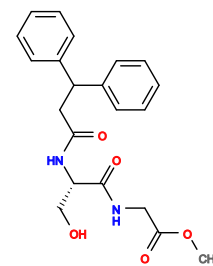
**262** V= -5.6, I=-7.44, S =-1.44



**263** V= -6.4, I=-7.56, S =-2.4, [CID14752371](#)



**264** V= -7.2, I=-6.89, S =-2.31



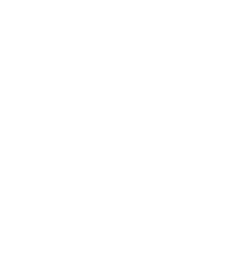
**265** V= -6.5, I=-7.47, S =-2.38



**266** V= -6.4, I=-7.63, S =-2.5



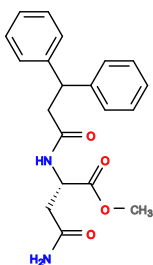
**267** V= -6.6, I=-7.38, S =-2.36



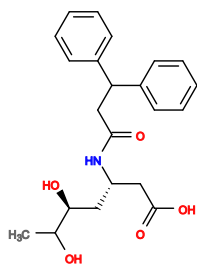
**268** V= -6.2, I=-7.39, S =-1.98



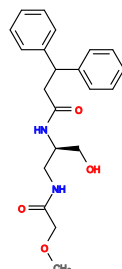




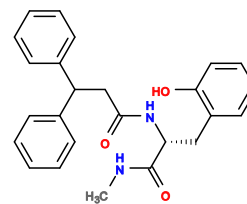
**269** V= -6.2, I=-7.59, S=-2.25



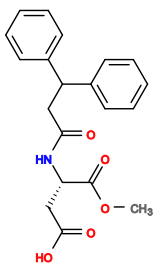
**270** V= -6.5, I=-6.82, S=-1.51



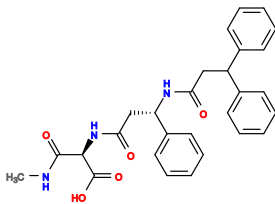
**271** V= -6.1, I=-7.26, S=-1.7



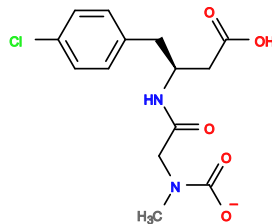
**272** V= -7.1, I=-7.54, S=-3.08



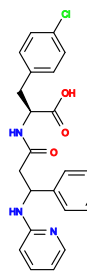
**273** V= -6.0, I=-7.65, S=-2.12



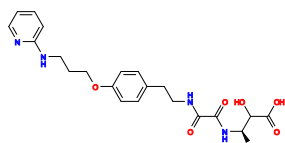
**274** V= -6.9, I=-7.54, S=-2.88



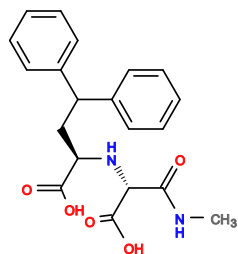
**275** V= -5.9, I=-7.02, S=-1.19



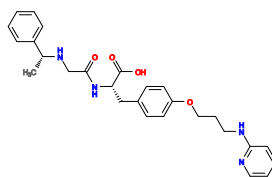
**276** V= -6.8, I=-8.11, S=-3.53



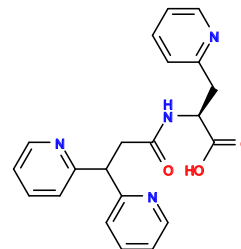
**277** V= -5.4, I=-8.02, S=-2.02



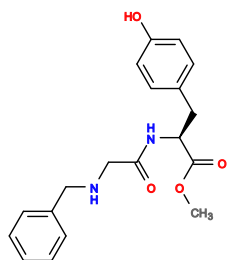
**278** V= -6.0, I=-7.49, S=-1.91



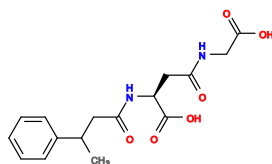
**279** V= -5.9, I=-8.39, S=-3.02



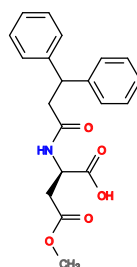
**280** V= -6.5, I=-7.15, S=-1.96



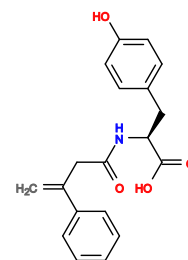
**281** V= -6.1, I=-7.67, S=-2.26



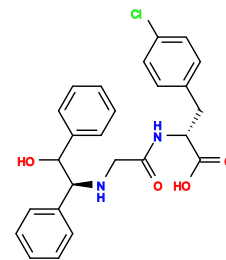
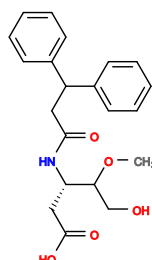
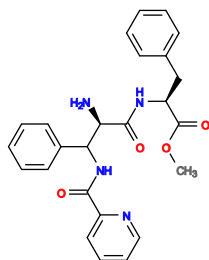
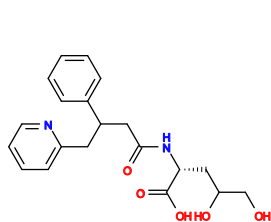
**282** V= -6.0, I=-7.01, S=-1.29



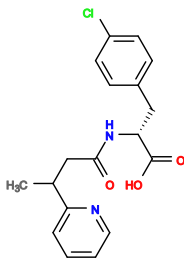
**283** V= -6.1, I=-7.59, S=-2.16



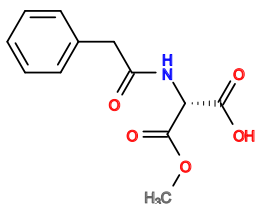
**284** V= -6.7, I=-7.18, S=-2.21



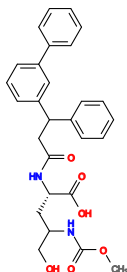
**285** V= -6.0, I=-7.15, S =-1.48



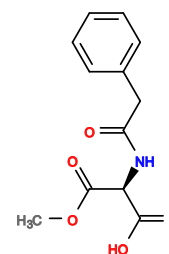
**286** V= -6.4, I=-7.83, S =-2.78



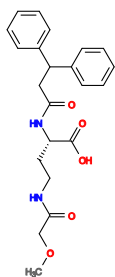
**287** V= -6.0, I=-7.41, S =-1.82



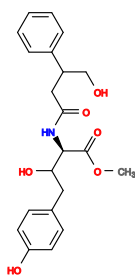
**288** V= -6.7, I=-7.82, S =-3.07



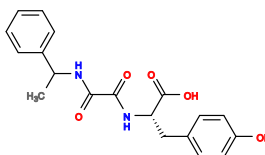
**289** V= -6.2, I=-7.83, S =-2.58



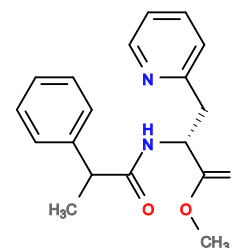
**290** V= -5.4, I=-7.49, S =-1.33



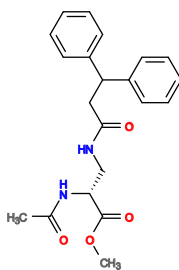
**291** V= -7.2, I=-7.46, S =-3.09



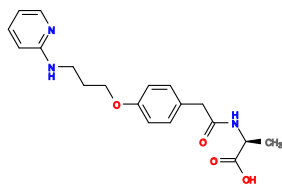
**292** V= -5.4, I=-7.49, S =-1.33



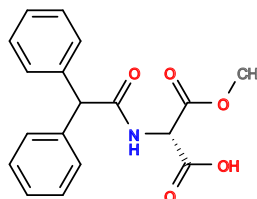
**293** V= -6.2, I=-7.42, S =-2.04



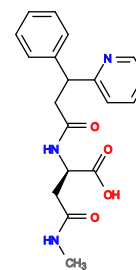
**294** V= -6.3, I=-7.54, S =-2.31



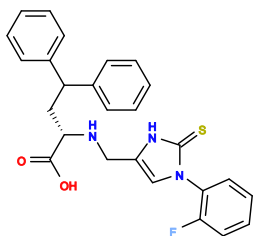
**295** V= -6.6, I=-6.9, S =-1.75



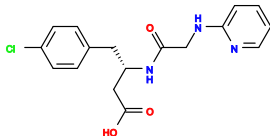
**296** V= -6.3, I=-7.53, S =-2.29



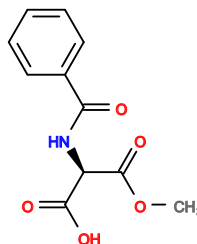
**297** V= -6.3, I=-7.6, S =-2.39



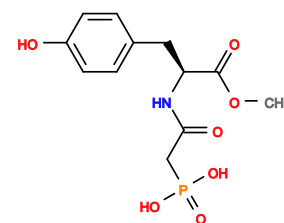
**298** V= -5.9, I=-8.07, S =-2.61



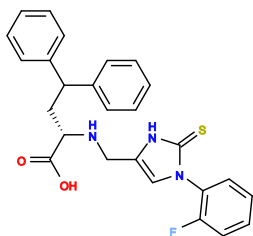
**299** V= -6.1, I=-7.56, S =-2.13



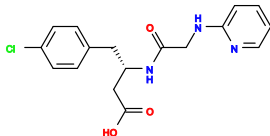
**300** V= -6.0, I=-7.47, S =-1.91



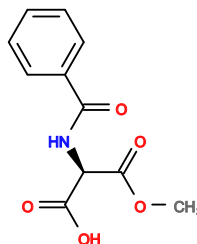
**301** V= -7.0, I=-8.11, S =-3.77



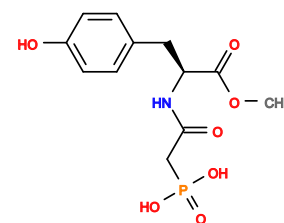
**302** V= -6.3, I=-7.71, S =-2.53

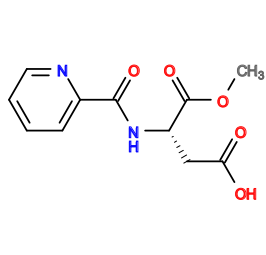


**303** V= -5.5, I=-7.35, S =-1.26

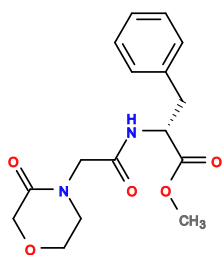


**304** V= -5.3, I=-7.22, S =-0.88

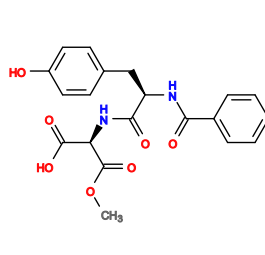




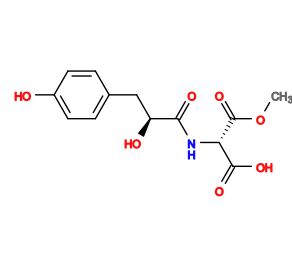
**305** V= -5.4, I=-7.16, S =-0.9



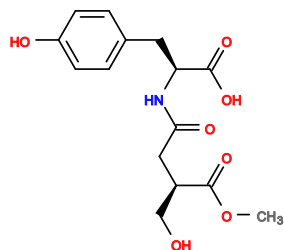
**306** V= -6.2, I=-7.26, S =-1.84



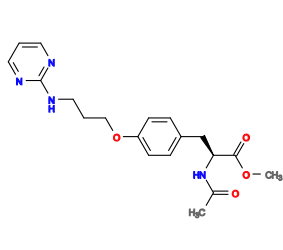
**307** V= -6.2, I=-7.57, S =-2.25



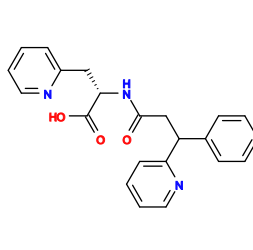
**308** V= -5.5, I=-7.41, S =-1.34



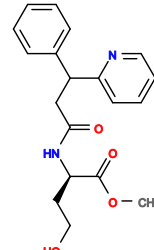
**309** V= -5.6, I=-7.32, S =-1.32



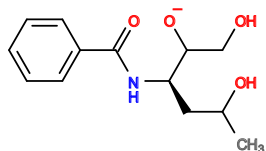
**310** V= -6.0, I=-8.09, S =-2.75



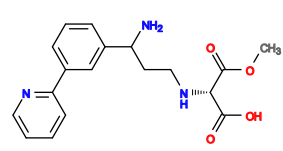
**311** V= -6.5, I=-7.39, S =-2.32



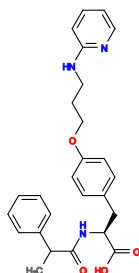
**312** V= -5.9, I=-7.38, S =-1.7



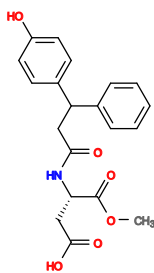
**313** V= -5.9, I=-6.35, S =-0.33



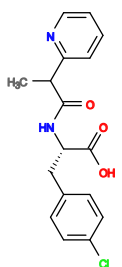
**314** V= -6.3, I=-7.29, S =-1.98



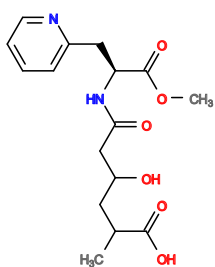
**315** V= -6.2, I=-8.44, S =-3.42



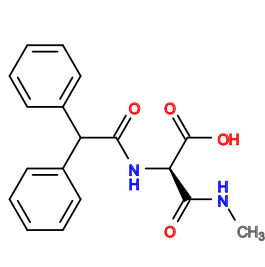
**316** V= -6.3, I=-7.58, S =-2.37



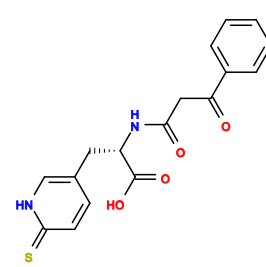
**317** V= -6.2, I=-7.68, S =-2.4



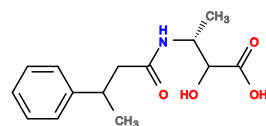
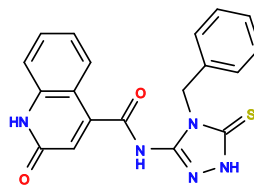
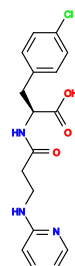
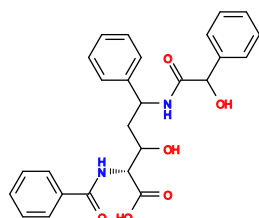
**318** V= -5.3, I=-7.53, S =-1.3



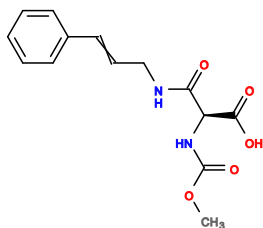
**319** V= -6.4, I=-7.26, S =-2.04



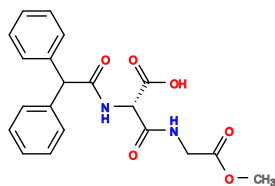
**320** V= -6.2, I=-7.76, S =-2.52



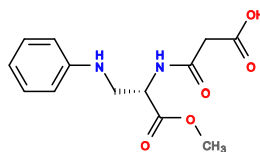
**321** V= -7.0, I=-6.86, S =-2.11



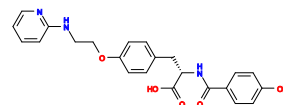
**322** V= -6.2, I=-7.87, S =-2.66



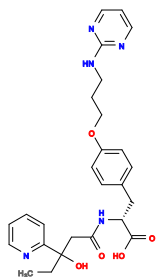
**323** V= -7.8, I=-7.66, S =-3.98



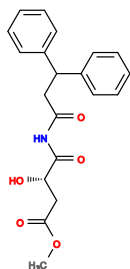
**324** V= -5.9, I=-6.92, S =-1.09



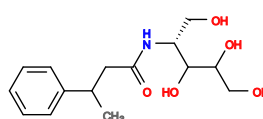
**325** V= -5.7, I=-7.46, S =-1.62



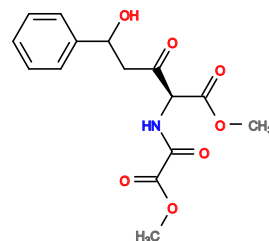
**326** V= -6.4, I=-7.21, S =-1.98



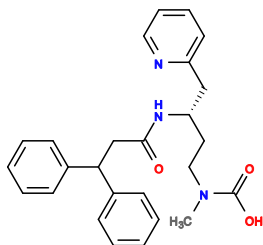
**327** V= -5.3, I=-7.55, S =-1.34



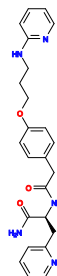
**328** V= -6.7, I=-8.1, S =-3.48



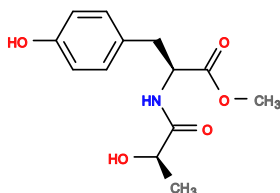
**329** V= -6.0, I=-8.18, S =-2.89



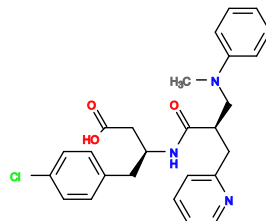
**330** V= -6.5, I=-7.32, S =-2.23



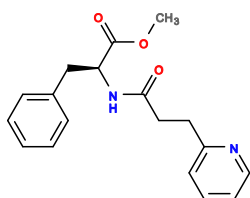
**331** V= -5.6, I=-6.83, S =-0.68



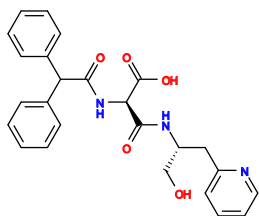
**332** V= -5.4, I=-7.33, S =-1.15



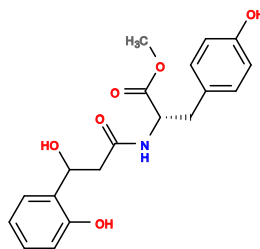
**333** V= -6.9, I=-7.02, S =-2.24



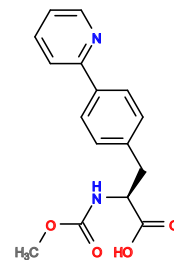
**334** V= -6.1, I=-8.09, S =-2.86



**335** V= -5.4, I=-7.44, S =-1.3



**336** V= -6.5, I=-7.86, S =-2.97



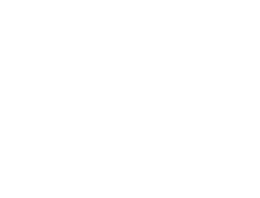
**337** V= -6.4, I=-7.67, S =-2.61



**338** V= -6.0, I=-7.41, S =-1.86

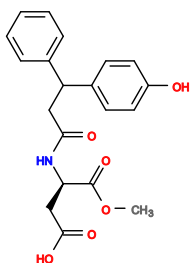


**339** V= -6.3, I=-7.63, S =-2.45

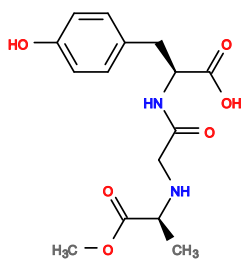


**340** V= -6.4, I=-7.64, S =-2.57

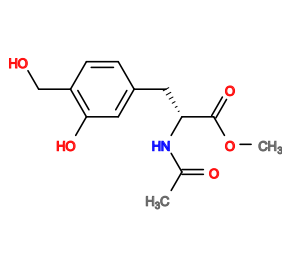




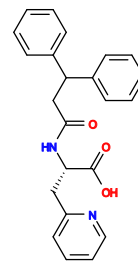
**341** V= -6.2, I=-7.59, S =- 2.3



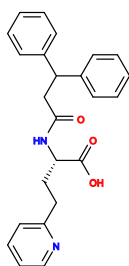
**342** V= -5.6, I=-7.44, S =- 1.5



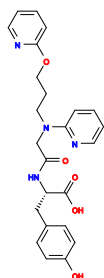
**343** V= -5.8, I=-7.14, S =- 1.31



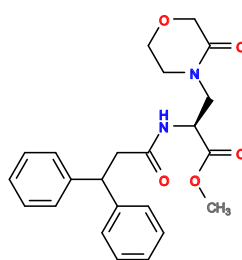
**344** V= -6.9, I=-7.32, S =- 2.65



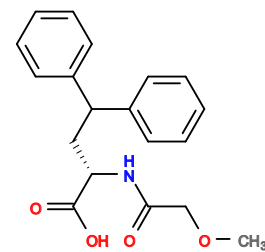
**345** V= -7.0, I=-7.27, S =- 2.69



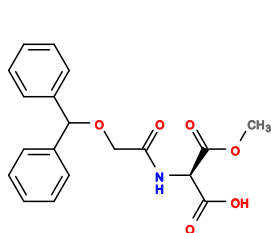
**346** V= -6.5, I=-7.67, S =- 2.72



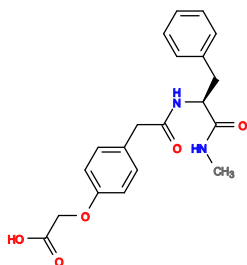
**347** V= -6.3, I=-7.46, S =- 2.24



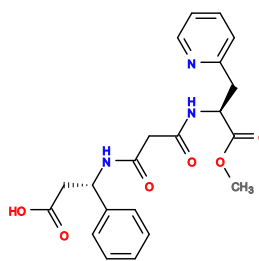
**348** V= -6.0, I=-7.47, S =- 1.96



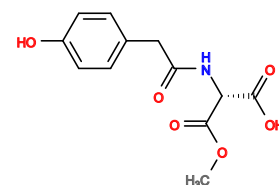
**349** V= -6.2, I=-7.65, S =- 2.4



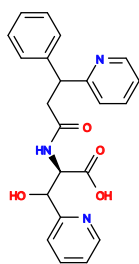
**350** V= -6.6, I=-7.35, S =- 2.4



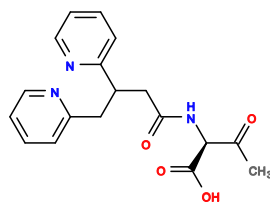
**351** V= -5.9, I=-7.71, S =- 2.18



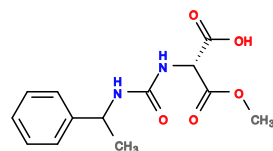
**352** V= -5.6, I=-7.4, S =- 1.47



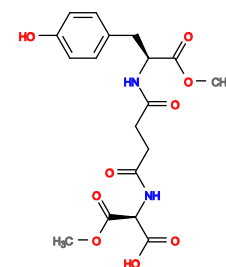
**353** V= -6.3, I=-7.2, S =- 1.91



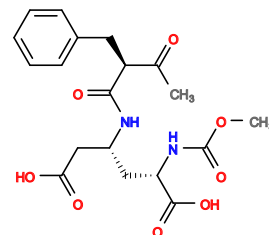
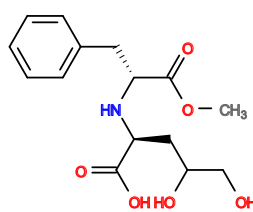
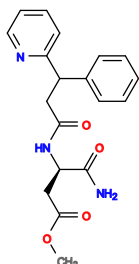
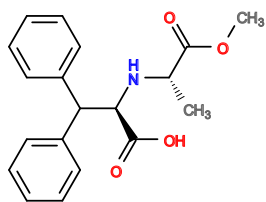
**354** V= -6.2, I=-7.11, S =- 1.69



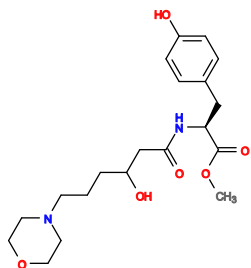
**355** V= -5.5, I=-7.44, S =- 1.43



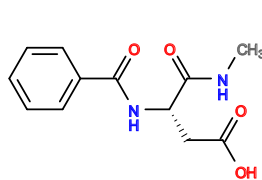
**356** V= -6.3, I=-7.41, S =- 2.18



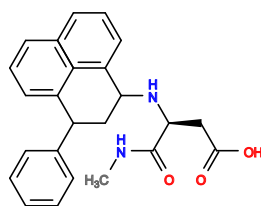
**357** V= -6.0, I=-7.58, S=-2.12



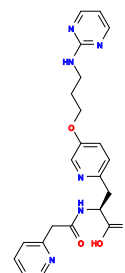
**358** V= -6.0, I=-7.46, S=-1.95



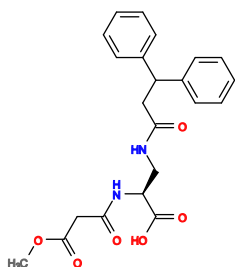
**359** V= -5.3, I=-7.42, S=-1.2



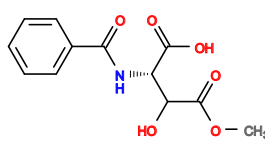
**360** V= -5.8, I=-7.44, S=-1.73



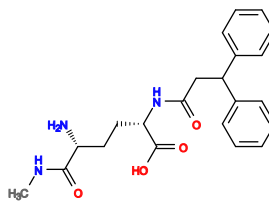
**361** V= -6.1, I=-7.35, S=-1.91



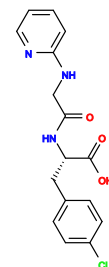
**362** V= -5.9, I=-6.95, S=-1.17



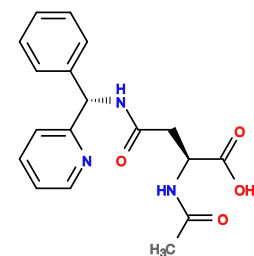
**363** V= -6.4, I=-7.48, S=-2.38



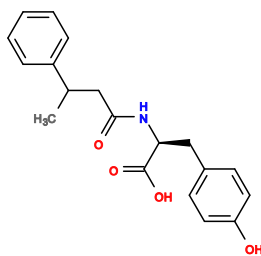
**364** V= -6.2, I=-7.67, S=-2.44



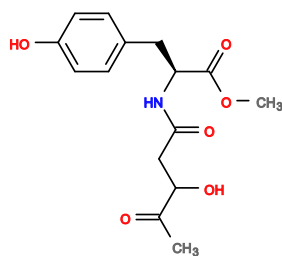
**365** V= -6.0, I=-7.58, S=-2.11



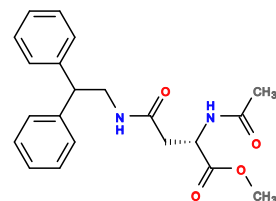
**366** V= -5.5, I=-7.13, S=-1.02



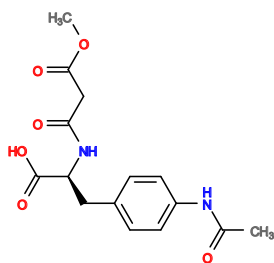
**367** V= -6.4, I=-7.25, S=-2.08



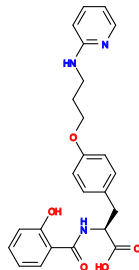
**368** V= -6.4, I=-7.65, S=-2.61



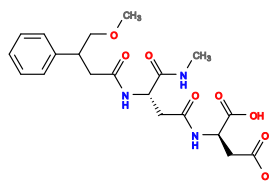
**369** V= -6.6, I=-6.98, S=-1.92



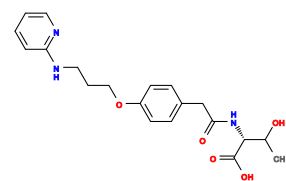
**370** V= -6.6, I=-7.15, S=-2.14



**371** V= -5.8, I=-7.49, S=-1.8



**372** V= -6.3, I=-7.57, S=-2.4



**373** V= -6.1, I=-7.12, S=-1.61



**374** V= -6.4, I=-8.3, S=-3.48

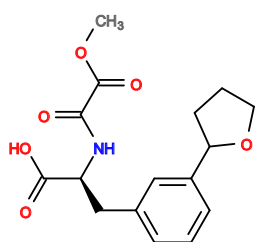


**375** V= -5.7, I=-7.31, S=-1.47

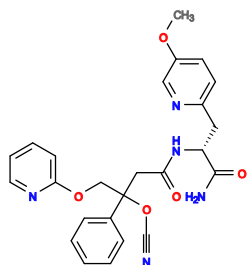


**376** V= -5.6, I=-8.22, S=-2.58

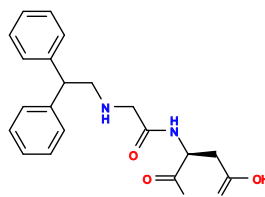




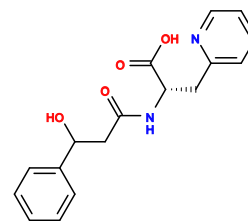
**377** V= -6.5, I=-7.11, S=-2.0



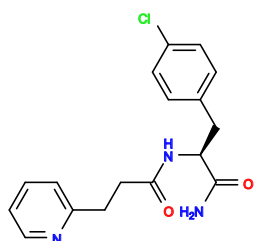
**378** V= -6.1, I=-7.47, S=-2.07



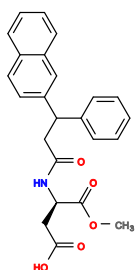
**379** V= -6.3, I=-6.92, S=-1.54



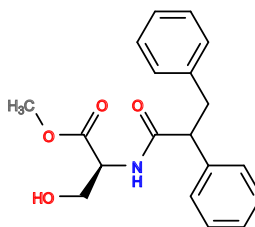
**380** V= -6.1, I=-7.15, S=-1.66



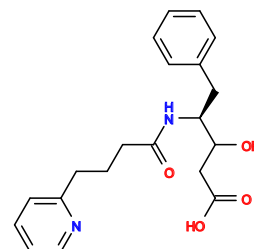
**381** V= -6.4, I=-7.6, S=-2.55



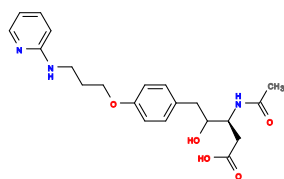
**382** V= -6.9, I=-7.72, S=-3.21



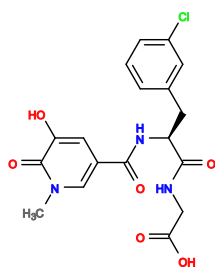
**383** V= -6.1, I=-7.43, S=-2.03



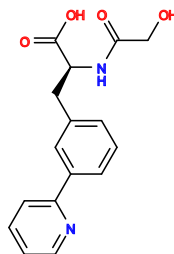
**384** V= -6.3, I=-7.38, S=-2.16



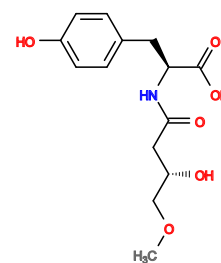
**385** V= -5.5, I=-8.03, S=-2.22



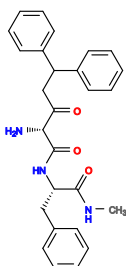
**386** V= -6.5, I=-7.02, S=-1.89



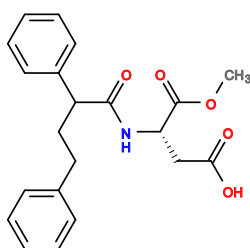
**387** V= -6.5, I=-7.05, S=-1.93



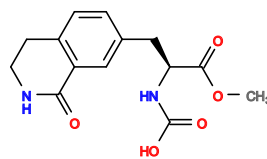
**388** V= -5.4, I=-7.06, S=-0.84



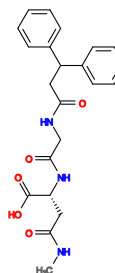
**389** V= -6.9, I=-7.57, S=-3.02



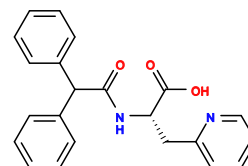
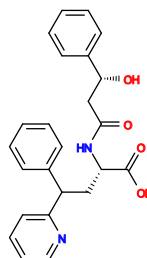
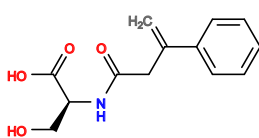
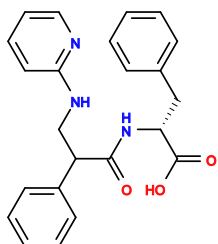
**390** V= -6.3, I=-7.51, S=-2.34



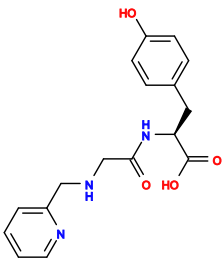
**391** V= -6.1, I=-7.39, S=-1.98



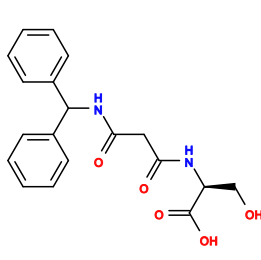
**392** V= -6.3, I=-7.34, S=-2.12



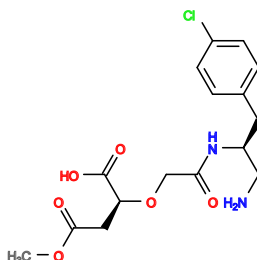
**393** V= -6.6, I=-7.39, S =- 2.49



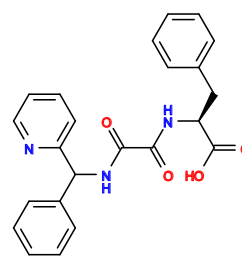
**394** V= -5.6, I=-7.06, S =- 1.04



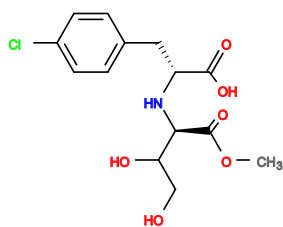
**395** V= -6.5, I=-7.32, S =- 2.29



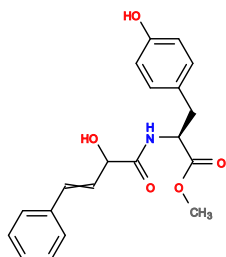
**396** V= -6.9, I=-7.15, S =- 2.46



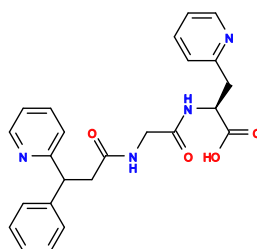
**397** V= -6.1, I=-7.11, S =- 1.62



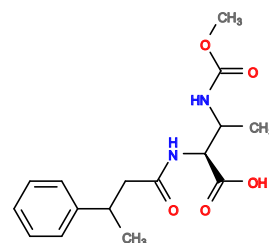
**398** V= -6.5, I=-6.86, S =- 1.68



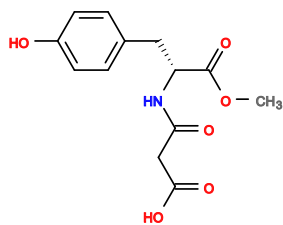
**399** V= -5.5, I=-7.38, S =- 1.38



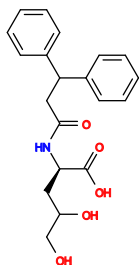
**400** V= -6.7, I=-7.29, S =- 2.46



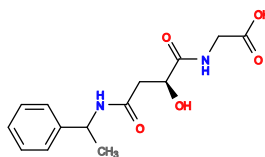
**401** V= -5.5, I=-7.52, S =- 1.56



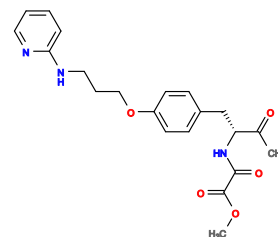
**402** V= -6.6, I=-7.53, S =- 2.69



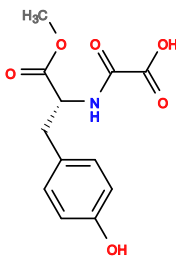
**403** V= -6.4, I=-7.49, S =- 2.43



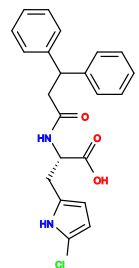
**404** V= -5.7, I=-7.42, S =- 1.63



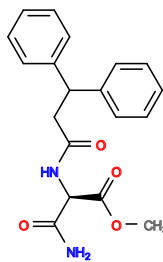
**405** V= -5.7, I=-7.44, S =- 1.67



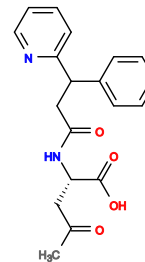
**406** V= -6.4, I=-6.88, S =- 1.61



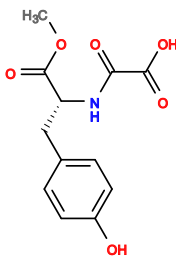
**407** V= -5.9, I=-6.62, S =- 0.77



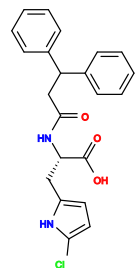
**408** V= -5.9, I=-8.29, S =- 2.99



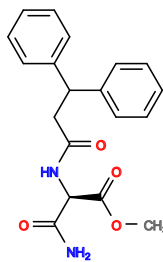
**409** V= -5.6, I=-7.42, S =- 1.54



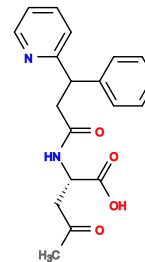
**410** V= -6.9, I=-7.94, S =- 3.53



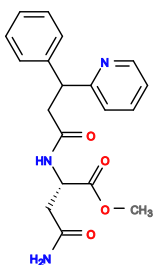
**411** V= -6.1, I=-7.56, S =- 2.23



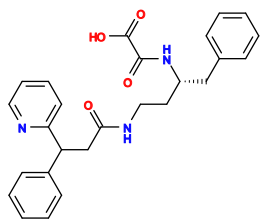
**412** V= -6.1, I=-7.33, S =- 1.92



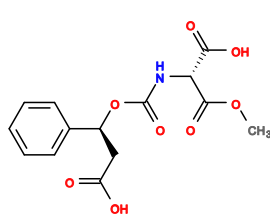




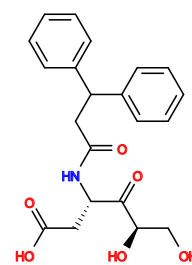
413 V= -5.8, I=-7.61, S =- 2.0



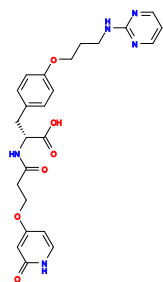
414 V= -6.8, I=-7.19, S =- 2.43



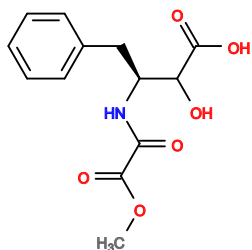
415 V= -5.5, I=-7.49, S =- 1.54



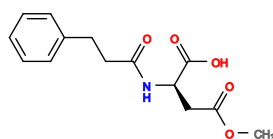
416 V= -6.3, I=-6.91, S =- 1.56



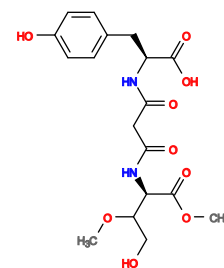
417 V= -6.2, I=-8.18, S =- 3.16



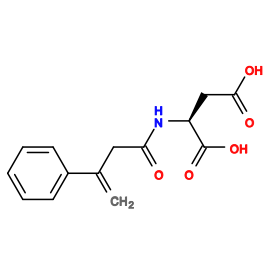
418 V= -5.5, I=-7.25, S =- 1.22



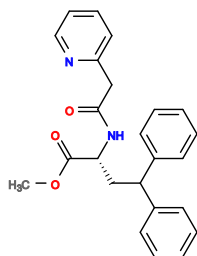
419 V= -5.6, I=-7.37, S =- 1.48



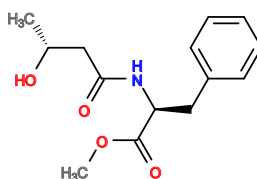
420 V= -5.7, I=-7.35, S =- 1.55



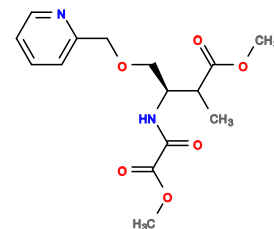
421 V= -5.7, I=-7.12, S =- 1.25



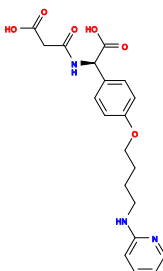
422 V= -6.6, I=-7.81, S =- 3.07



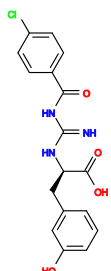
423 V= -5.4, I=-7.46, S =- 1.41



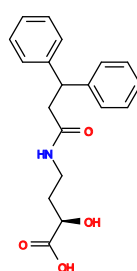
424 V= -5.3, I=-7.39, S =- 1.21



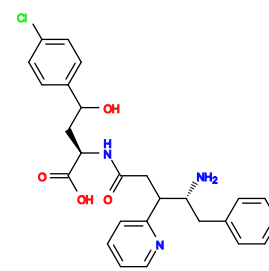
425 V= -5.8, I=-8.2, S =- 2.78



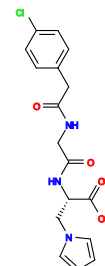
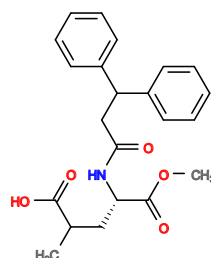
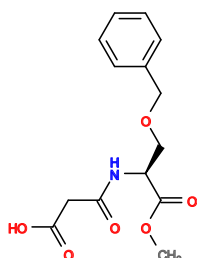
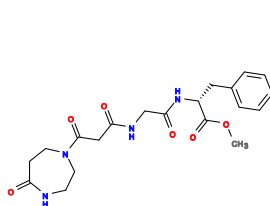
426 V= -6.6, I=-7.45, S =- 2.59



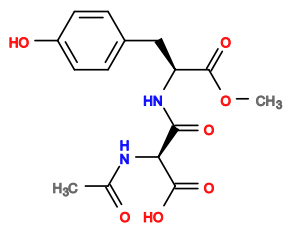
427 V= -6.3, I=-6.7, S =- 1.28



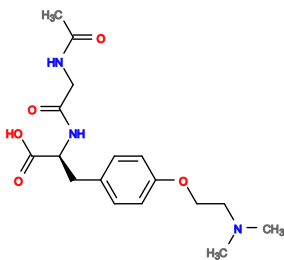
428 V= -6.4, I=-7.83, S =- 2.9



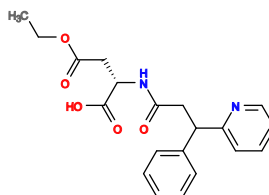
**429** V= -6.5, I=-7.12, S =- 2.05



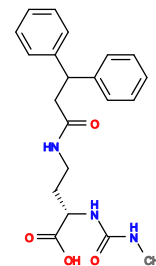
**430** V= -5.4, I=-7.57, S =- 1.56



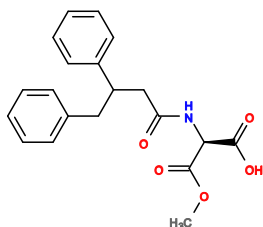
**431** V= -6.2, I=-7.71, S =- 2.54



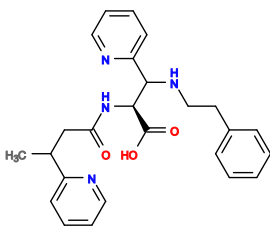
**432** V= -6.1, I=-7.67, S =- 2.38



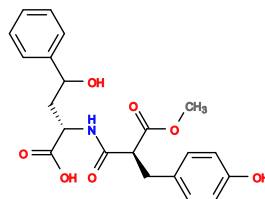
**433** V= -5.7, I=-7.4, S =- 1.62



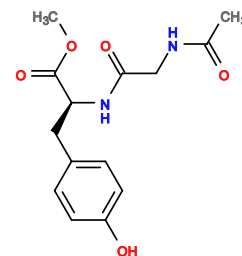
**434** V= -5.9, I=-7.11, S =- 1.44



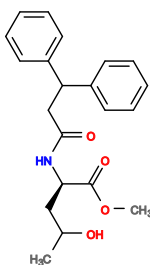
**435** V= -6.3, I=-7.29, S =- 2.08



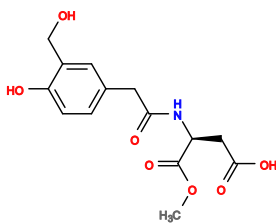
**436** V= -6.6, I=-7.1, S =- 2.12



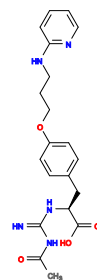
**437** V= -6.2, I=-7.65, S =- 2.46



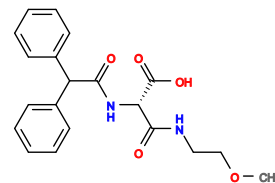
**438** V= -5.9, I=-7.75, S =- 2.3



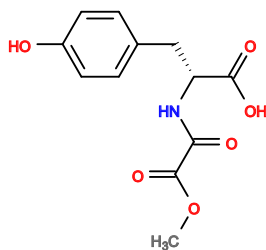
**439** V= -6.1, I=-7.66, S =- 2.38



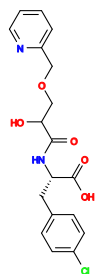
**440** V= -5.7, I=-7.36, S =- 1.58 , CID91978419



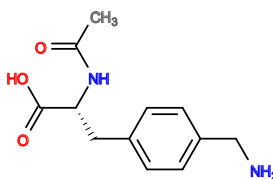
**441** V= -6.0, I=-7.62, S =- 2.22



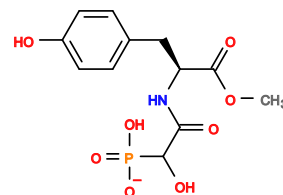
**442** V= -5.4, I=-7.32, S =- 1.23



**443** V= -5.4, I=-8.28, S =- 2.51



**444** V= -6.2, I=-7.23, S =- 1.91



**445** V= -5.7, I=-7.18, S =- 1.35



**446** V= -6.0, I=-7.74, S =- 2.38

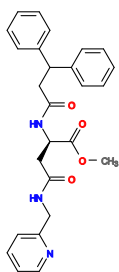


**447** V= -5.6, I=-6.72, S =- 0.62

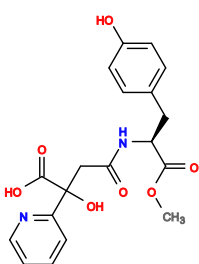


**448** V= -5.5, I=-7.08, S =- 1.01

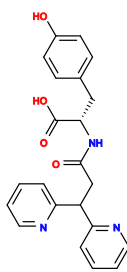




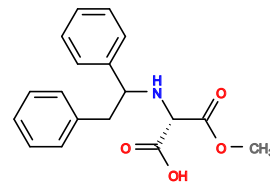
**449** V= -6.6, I=-7.89, S=-3.19



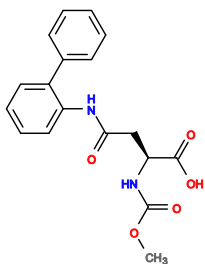
**450** V= -5.9, I=-7.64, S=-2.16



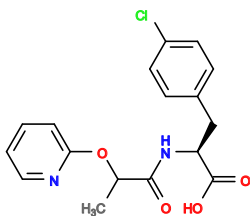
**451** V= -6.7, I=-7.16, S=-2.32



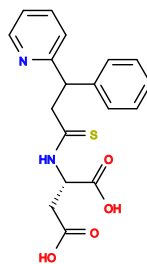
**452** V= -6.0, I=-7.59, S=-2.2



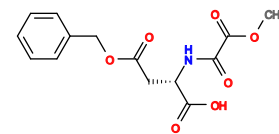
**453** V= -6.5, I=-7.38, S=-2.41



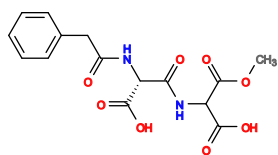
**454** V= -6.3, I=-7.76, S=-2.72



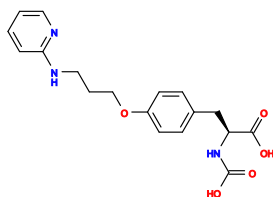
**455** V= -6.1, I=-7.66, S=-2.38



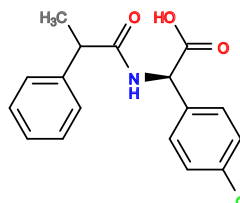
**456** V= -5.6, I=-7.44, S=-1.6



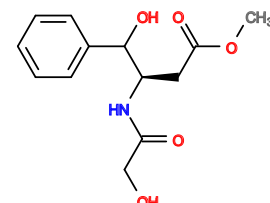
**457** V= -5.6, I=-7.24, S=-1.33



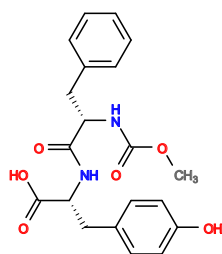
**458** V= -5.4, I=-8.19, S=-2.4



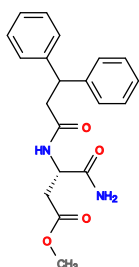
**459** V= -6.5, I=-7.61, S=-2.72



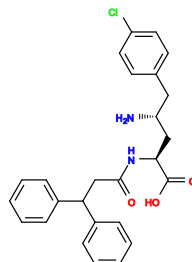
**460** V= -5.4, I=-7.1, S=-0.95



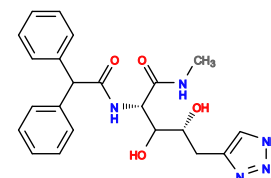
**461** V= -6.3, I=-7.52, S=-2.41



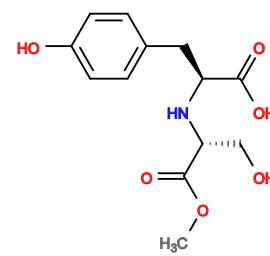
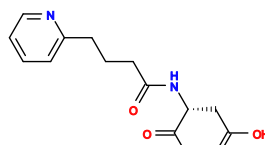
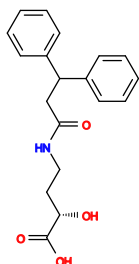
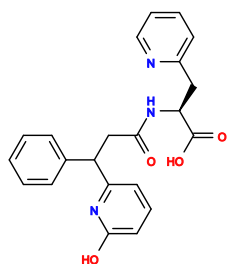
**462** V= -6.1, I=-7.47, S=-2.14



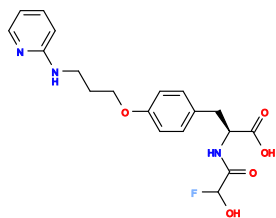
**463** V= -6.6, I=-7.93, S=-3.26



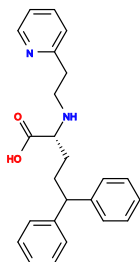
**464** V= -6.7, I=-6.72, S=-1.74



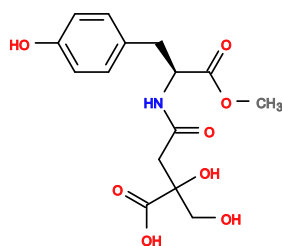
**465** V= -6.9, I=-7.07, S=-2.4



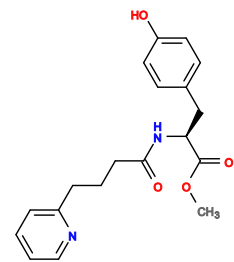
**466** V= -6.3, I=-6.68, S=-1.29



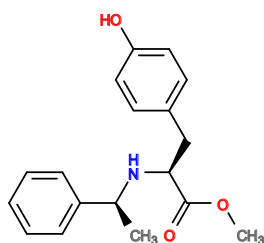
**467** V= -5.4, I=-7.08, S=-0.92



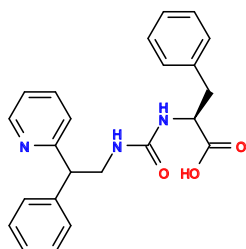
**468** V= -5.2, I=-7.36, S=-1.1



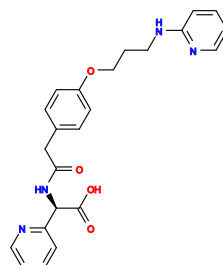
**469** V= -5.5, I=-8.28, S=-2.62



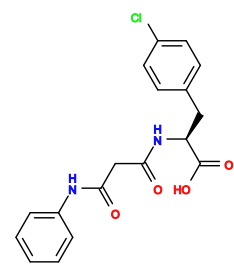
**470** V= -6.4, I=-7.35, S=-2.29



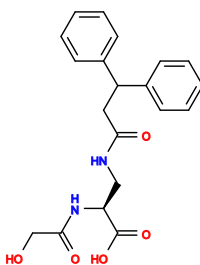
**471** V= -5.5, I=-7.36, S=-1.4



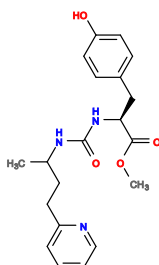
**472** V= -6.3, I=-7.73, S=-2.7



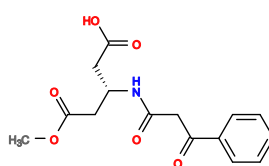
**473** V= -6.5, I=-7.52, S=-2.6



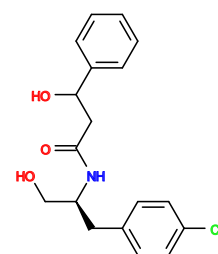
**474** V= -6.9, I=-7.31, S=-2.73



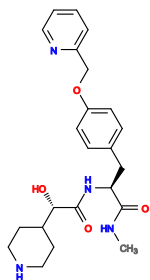
**475** V= -6.2, I=-8.16, S=-3.17



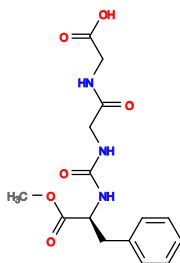
**476** V= -6.6, I=-7.71, S=-2.97



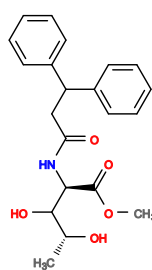
**477** V= -6.6, I=-6.89, S=-1.87



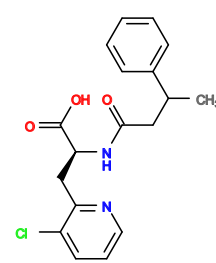
**478** V= -6.5, I=-7.58, S=-2.7



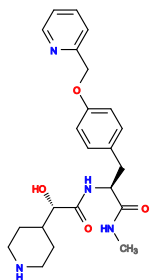
**479** V= -5.4, I=-7.29, S=-1.21



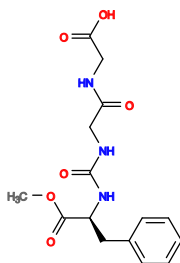
**480** V= -6.2, I=-7.44, S=-2.21



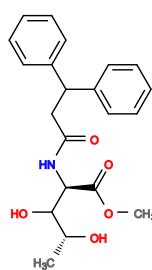
**481** V= -6.5, I=-7.35, S=-2.39



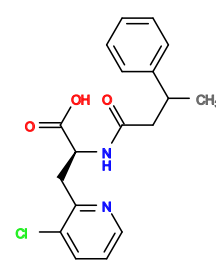
**482** V= -6.1, I=-7.26, S=-1.87

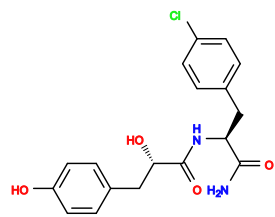


**483** V= -5.8, I=-7.64, S=-2.07

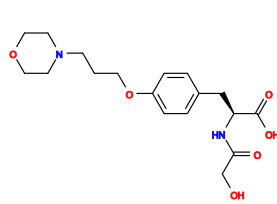


**484** V= -6.0, I=-7.58, S=-2.19

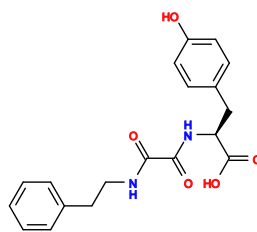




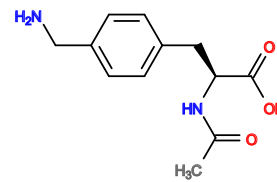
**485** V= -6.4, I=-7.42, S=-2.38



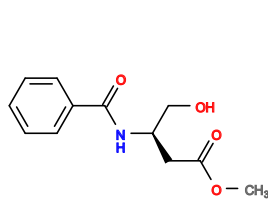
**486** V= -5.6, I=-7.47, S=-1.65



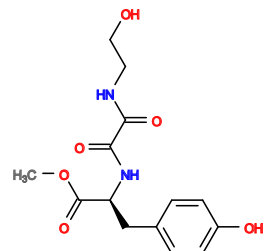
**487** V= -6.5, I=-6.95, S=-1.86



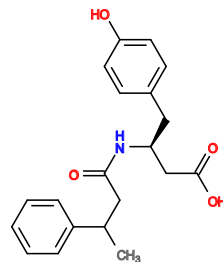
**488** V= -5.6, I=-6.73, S=-0.67, [CID68025972](#)



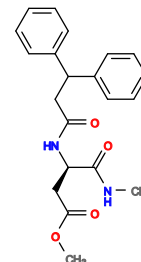
**489** V= -5.5, I=-6.84, S=-0.71



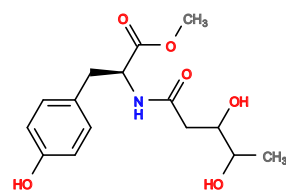
**490** V= -5.7, I=-7.32, S=-1.55



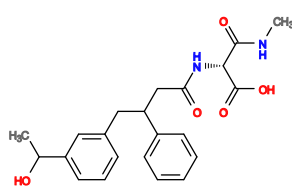
**491** V= -6.4, I=-7.09, S=-1.95



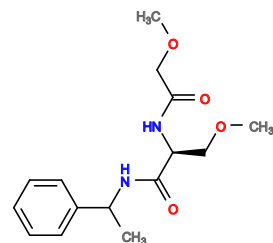
**492** V= -6.2, I=-7.5, S=-2.29



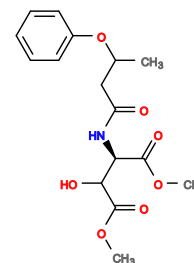
**493** V= -5.7, I=-7.42, S=-1.69



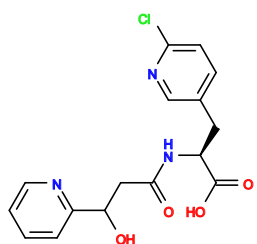
**494** V= -6.7, I=-7.19, S=-2.38



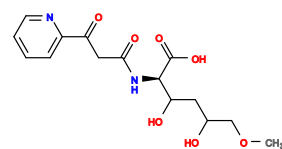
**495** V= -5.6, I=-7.05, S=-1.09



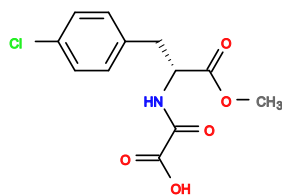
**496** V= -5.4, I=-7.43, S=-1.41



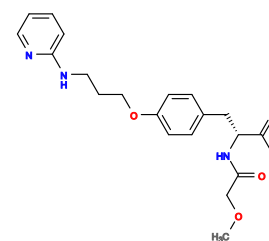
**497** V= -6.3, I=-6.92, S=-1.64



**498** V= -5.4, I=-6.9, S=-0.7



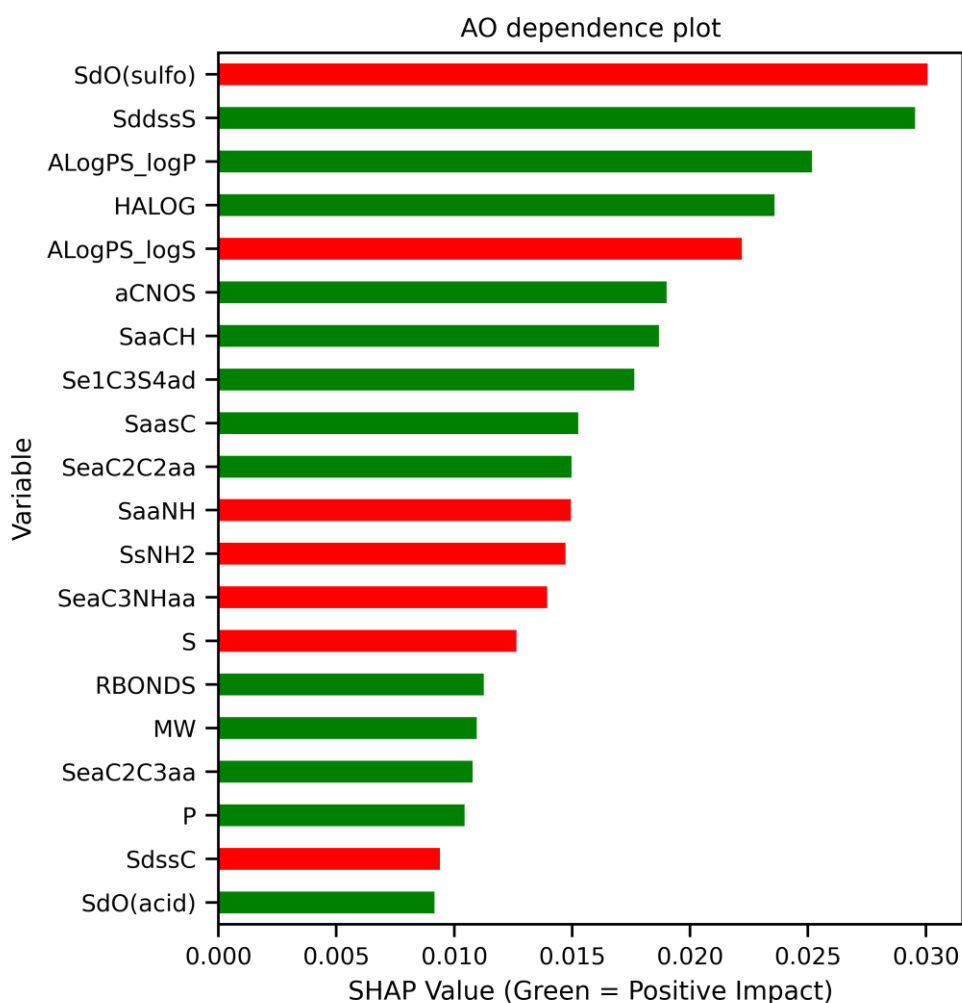
**499** V= -5.8, I=-7.56, S=-1.98



**500** V= -5.5, I=-8.24, S=-2.59

## Appendix IV

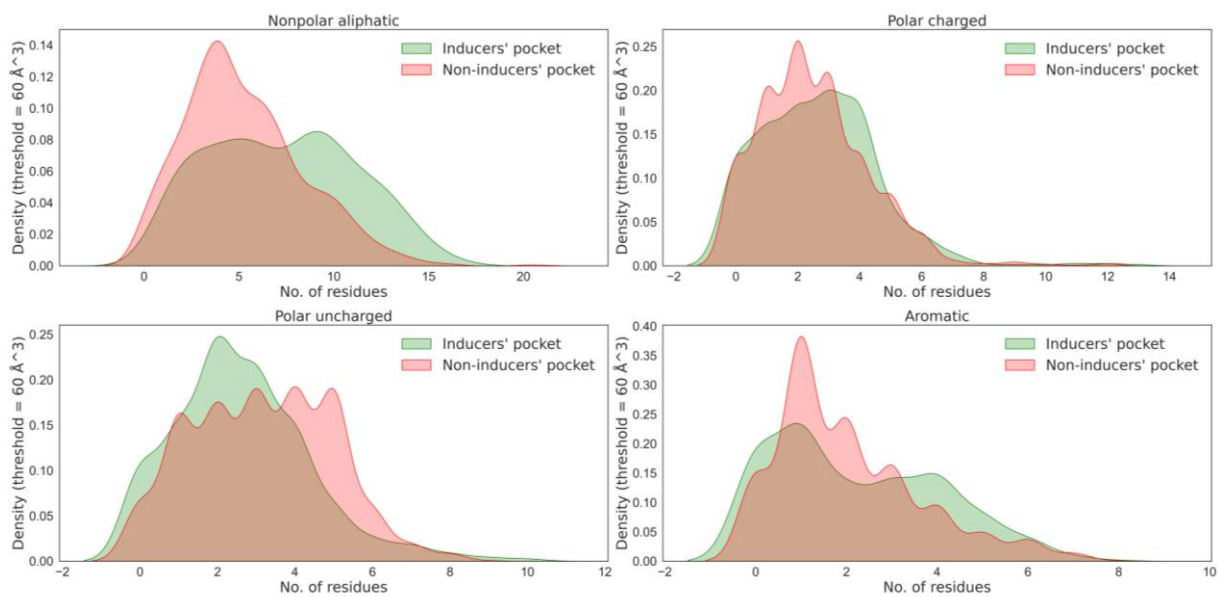
## 1. Interpretation of models



**Figure S1** The SHAP values calculated for the filtered descriptors. **Green:** the bit has a positive impact (increasing the propensity to be inducers) on the target property; **red:** the bit has a negative impact on the target property.

## 2. Analysis on pocket residues

The distribution of pocket residues between inducer and non-inducer groups was investigated into four categories of amino acids (nonpolar aliphatic, polar charged, polar uncharged, and aromatic) as shown in Figure S2. For inducers, the number of hydrophobic amino acids in holo pockets was always larger than that in apo pockets, which was inconsistent with the result of Cimermancic et al<sup>107</sup>. This observation explains why inducers tended to have strong hydrophobicity and aromaticity.



**Figure S2** The distribution of pocket residues compared between holo and apo pockets for both groups of ligands (based on the optimal threshold = 60 Å<sup>3</sup>).