In silico search for novel bacterial inhibitors targeting RNA polymerase switch region

Master's Thesis
Master's Degree Programme in Biomedical Science: Medicinal and Radiopharmaceutical Chemistry
Department of Chemistry
Faculty of Science
University of Turku

Author:
Aliaa Ali
Supervisor:
Adj. Prof. Outi Salo-Ahen

July 2022
Turku

The originality of this thesis has been checked in accordance with the University of Turku quality assurance system using the Turnitin Originality Check service.

Master's thesis
Medicinal Chemistry
Aliaa Ali
In silico search for novel bacterial inhibitors targeting RNA polymerase switch region
Supervisor: Adj. Prof. Outi Salo-Ahen
69 pages
Date: July 2022

Abstract:

The arise of antibiotic-resistant bacterial strains in an alarming rate has increased the interest in the discovery of novel antibiotics. The rifamycins are a valuable class of antibiotics that target bacterial Ribonucleic Acid polymerase (RNAP) and are considered the first-line treatment for tuberculosis. Consequently, bacterial strains resistant to rifamycin constitute a public health threat. RNAP switch region is an attractive target for the development of new antibacterial agents as it lies away from the rifamycin binding region and thus the compounds that target the switch region would not show cross-resistance with rifamycins.

In this work, we developed a virtual screening pipeline to identify new bacterial RNAP inhibitors that target the enzyme switch region. The screening pipeline involved docking of the designated libraries using the Maestro Glide docking tool, and the compounds with the best docking scores were submitted for binding free energy calculations using the molecular mechanics-generalised born surface area (MM-GBSA)-based method. Moreover, a quantitative structure-activity relationship (QSAR) model was developed, and it was applied to predict the biological activity of the compounds with the most favourable calculated binding free energies. Based on the results of docking, MM-GBSA binding free energies and the activities predicted by the QSAR model, the most promising compounds were chosen to be evaluated by molecular dynamics (MD) simulations. The results of the MD simulation of each docked candidate compound in the RNAP binding site were compared with the MD simulations carried out with the apo protein and with a reference co-crystallized ligand in the RNAP binding site. The candidate compounds showing comparable binding to the RNAP site to the reference ligand were selected for further biological testing.

Key words: antibacterial agents, RNA polymerase, QSAR, MM-GBSA, Molecular Dynamics simulation, myxopyronin, squaramide, structure-based drug design, switch region, virtual screening.

Acknowledgment

First, I wish to express my gratitude to my supervisor, Adj. Prof. Outi Salo-Ahen for accepting me as a master's student in her group and for giving me the chance to do my thesis in one of the drug design and development fields I am deeply passionate about. Additionally, I would like to thank her for her consistent guidance and advice.

I would like to especially express my sincerest thanks to Dr Rajendra Bhadane for his invaluable advice and insights that brought this thesis to reality.

I would also like to deeply thank Dr Parthiban Marimuthu for his assistance and support with my work and for his teaching that guided my initial hesitant steps into this field.

I want to extend my appreciation to all members of the Structural Bioinformatics Laboratory, Faculty of Science and Engineering at Åbo Akademi University for their help and suggestions.

Additionally, I have to thank Prof. Pasi Virta for his knowledgeable classes, supervision and guidance in project one, my first research experience, and specifically, for his advice, suggestions and help throughout my degree program.

I would take this opportunity to thank Prof. Jessica Rosenholm for enriching my knowledge, expertise, and research experience by accepting me as an intern in her research group, BioNanoMaterials.

I am also sincerely grateful to my supervisor and mentor in the BioNanoMaterials research group, Dr Kuldeep Bansal for his advice and suggestions. He always showed me the way to develop and improve but let me explore it on my own. Despite being challenging, I have learned a lot and everything I have learned has been a great asset to me while writing my thesis.

I wish to acknowledge CSC – IT Center for Science, Finland, for generous computational resources.

Last but not least, I would like to express my love and appreciation for my friends and for my parents who were there for me, supported me, encouraged me and withhold on with me when I was stressed and overwhelmed and made this thesis be possible.

Sincerely yours,

Aliaa Ali

Turku, Finland

July 2022

Table of content

LIST OF ABREVIATIONS

ADME	Absorption, distribution, metabolism, and excretion
ATP	Adenosine Triphosphate
AZT	Azidothymidine
BH	Bridge helix
CADD	Computer-aided drug design
cLogP	Calculated LogP
Cor	Corallopyronin
CTP	Cytidine Triphosphate
D-APP1	(D-Nα-aroyl-N-aryl-phenylalaninamides)
DPBB	double-psi beta-barrel
EC	Elongation complex
Fdx	Fidaxomicin
FL	Fork loop
GTP	Guanosine Triphophate
IC50	Half maximal inhibitory concentration
kDa	Kilo Dalton
LBDD	Ligand-based drug design
LBVS	Ligand-based virtual screening
MD	Molecular dynamics
MIC	minimum inhibitory concentrations
Mwt	Molecular weight
Myx	Myxopyronin
NA	Nucleoside analogue
NAC	Nucleotide addition cycle
NAI	Nucleoside analogue inhibitor
NMP	Nucleoside monophosphate
NTP	Nucleoside triphosphate
PDB	Protein Data Bank
PPi	Pyrophosphate
Prime/MM-GBSA	Prime/Molecular Mechanics-Generalized Born Surface Area
PUM	Pseudouridimycin
QSAR	Quantitative structure-activity relationship

RIFs	Rifamycins
Rip	Ripostatin
RMSD	Root-mean-square deviation
RMSE	Root-mean-squareerror
RMSF	Root-mean-square fluctuation
RNAP	RNA polymerase
Ro5	Rule of five
RPc	Closed RNAP-promoter complex
RPo	Open RNAP-promoter complex
SAL	Salinamide
SAR	Structure-activity relationship
SBDD	Structure-based drug design
SBVS	Structure-based virtual screening
SMILES	Simplified molecular input line entry system
SOR	Sorangicin
SP	Standard precision mode in Glide
SQ	Squaramide
STL	Streptolydigin
TL	Trigger loop
TTP	Thymidine Triphosphate
UTP	Uridine Triphosphate
XP	Extra precision mode in Glide
VS	Virtual screening

1 Introduction

The emergence of antibiotic-resistant bacterial strains in an alarming rate along with a diminished antibiotic pipeline has raised claims that a post-antibiotic era is eminent[1]. Hence, there has been an evolving interest in the identification of new compounds that carry antibacterial therapeutic value[2]. Ribonucleic Acid polymerase (RNAP) enzymes synthesize RNA from a DNA or RNA template through a process called transcription that renders them crucial for gene expression, cell growth, and viability. Although RNAP exists in most organisms, bacterial RNAP does not share extensive sequence homology with eukaryotic RNAP. This explains the selectivity of rifampicin, an antibiotic targeting bacterial RNAP, towards targeting bacteria while sparing eukaryotic RNAP[3]. Meanwhile, RNAP is highly conserved among different bacterial species, which permits broad spectrum activity. Therefore, RNAP inhibition is an attractive strategy for treating bacterial infections[4].

Several compounds are known to inhibit bacterial RNAP. However, only rifamycins and fidaxomicin (lipiarmycin) are currently approved for clinical use[4]. The rifamycins are effective against Gram-positive and Gram-negative bacterial infections. Besides, they are the first-line treatment for tuberculosis and thus they are of high clinical importance. Thus, the existence of bacterial strains resistant to rifamycins threatens their clinical utility and constitutes a public health threat. Resistance to rifamycins occurs through substitution of residues in or adjacent to the rifamycin binding site on bacterial RNAP that directly decrease the binding of rifamycins. Therefore, there is an urgent need for identifying new inhibitor classes that target bacterial RNAP but sites distinct from the rifamycin binding site. Such compounds would retain the same biochemical effect as rifamycins but would not show cross-resistance with them[5].

Herein, an attractive target region in bacterial RNAP for the development of new antibacterial agents is called the ‘switch region’. The ‘switch region’ is located at the clamp base of the enzyme and acts as a hinge for the clamp opening and closure. Importantly, it lies away from the binding region of the rifamycins and other previously characterized RNAP inhibitors. There are RNAP inhibitors previously reported in literature that target the switch region which are myxopyronin, corallopyronin, ripostatin and squaramides[6]. In addition, the increasing availability of macromolecular structures revealed by X-ray crystallography and cryoelectron microscopy allows structure-based drug design of inhibitors of large antibacterial targets such as the ribosome and RNAP[2]. Thus, and within the scope of this project, we tried to develop a pipeline that would enable an *in silico* search for RNAP inhibitors, targeting the enzyme's switch region.

1.1 Structure and function of bacterial RNAP

1.1.1 Structure of bacterial RNAP

RNAP is a multi-subunit enzyme in which the bacterial RNAP core enzyme consists of five subunits that are α (two copies), β, β′ and ω. The core enzyme synthesizes RNA after binding to template DNA. The structure is complemented by a sigma (σ) factor to form a holoenzyme[7]. The structure of RNAP exhibits the shape of a crab claw with two 'pincers' surrounding a cleft region that serves as the binding site for DNA and constitutes for the active site of the enzyme. The β′ subunit, which is the largest subunit, forms one pincer called 'the clamp'. While the latter is the largest, β subunit forms the other pincer. Through different crystal structures of RNAP, it can be observed that the RNAP clamp can adopt different conformational states, ranging from an open RNAP-promoter complex to a closed state RNAP-promoter complex. The two states differ by ≥ 20° swinging motion of the clamp about a hinge region, termed the 'switch region', that is located at the base of the clamp. They also differ by a ≥ 20 Å displacement of residues at the tip of the clamp (Figure 1)[8].

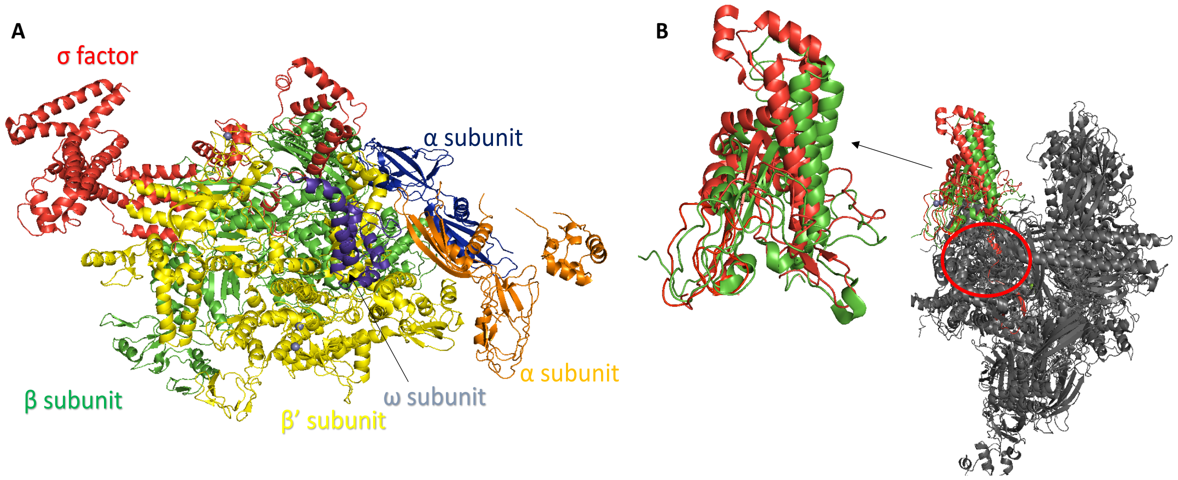


Figure 1: (A) Structure of RNA polymerase (RNAP) core enzyme. α, α', ω, β, β' subunits and the σ factor are shown in orange, blue, violet, green, yellow, and red respectively. (B) Structure of RNAP showing conformational states of the RNAP clamp, open (red), and closed (green) as observed in crystal structures (PDB ID: 6FBV, PDB ID: 4YFX). Red circle surrounds the switch region.

The α subunits initiate RNAP assembly by dimerizing into a platform that interacts with β and β′. In addition, they play key roles in transcription initiation and regulation[9][7]. The β and β′ subunits form the catalytic centre of RNA synthesis in which the active site is formed by two double-psi beta-barrel (DPBB) domains, both are β subunits. The β and β′ subunits form 80% of the total mass of the core enzyme in which each forms a claw arm of the crab claw structure of RNAP. This generates a cleft that allows the entry of the template DNA into the enzyme active site at the bottom of the cleft. They are responsible for carrying out nucleoside triphosphate (NTP) condensation and translocation during the nucleotide addition cycle (NAC)[7][10]. The role of the ω subunit is suggested to be correlated with maintaining the RNAP catalytic activity and/or protecting the DPBB domain against various damages[10]. The bacterial RNAP core alone is catalytically proficient. However, it cannot initiate transcription from the promoters and thus binding of the σ factor to the RNAP core forms the holoenzyme that can recognize and initiate transcription from the promoters[11].

1.1.2 Transcription

Transcription is a cyclic process, and it is the first step in gene expression where RNAP synthesizes the RNA chain complementary to the DNA template strand from nucleoside triphosphate (NTP) substrates. Transcription includes, roughly, three major steps: 1) promoter DNA binding and RNA chain initiation; 2) processive RNA chain elongation; 3) termination[12].

The initiation step starts with the formation of the holoenzyme as one of the promoter-specific σ subunits attaches to the core. The holoenzyme locates the promotor sequence via sliding along the DNA sequence and then binding to it. Thereafter, the DNA duplex around the transcriptional start-site melts forming the transcription bubble from which a stable open promoter complex is formed and RNA synthesis initiates. Short RNA strands are unstable and thus initially RNAP synthesizes short RNA products that are released and re-synthesized repeatedly in a cycle called abortive initiation. Once RNAP manages to successfully incorporate 8–12 nucleotide residues, it overcomes the abortive initiation phase and enters the elongation mode[12].

The RNAP transcription complex undergoes a promoter clearance where RNAP releases the sigma factor once RNA molecule of 13–15 nucleotide long has been synthesized, in which 7–9 nucleotides pair with the DNA template strand in an RNA/DNA hybrid. Thereafter, transition to the elongation phase happens. The enzyme then undergoes a significant conformational change; specifically, the sigma–core interface converts the initiation complex into an extremely stable and processive elongation complex (EC) that can uninterruptedly synthesize thousands of nucleotides long RNA chains. RNAP catalyses the DNA-directed addition of nucleotides to the growing RNA chain to the RNA $3'$-end and releases a pyrophosphate (PPi) ion. Then it is translocated along the DNA and the RNA to free the nucleotide addition site (denoted as $i + 1\ or\ A$ site), allowing for binding of the next NTP[12][13][14]. Once elongation complex (EC) reaches the termination signal at the terminator sequence, nascent transcript and the template DNA are both released and RNAP becomes available for rebinding to the sigma complex and starting a new cycle of transcription[12].

1.1.3 The nucleotide addition cycle

The EC repeatedly performs the nucleotide addition cycle (NAC) throughout the transcription elongation to attach a nucleotide to the hydroxyl at the $3'$-terminus of the growing messenger RNA (mRNA) chain by catalysing DNA template-directed formation of an RNA phosphodiester bond. Therefore, the catalytic centre of RNAP includes two sites, binding sites for the RNA $3'$-terminus (termed as i site) and the insertion site for the incoming NTP. After the formation of the phosphodiester bond, the newly formed 3 terminus is translocated from the A site to the i site with the release of the pyrophosphate (PPi) product[15][16].

There are several structural motifs that contribute to the key enzymatic processes of RNAPs and among them there are, the bridge helix (BH) and the trigger loop (TL). The bridge helix guides the template DNA strand into the active site and then positions the DNA-RNA hybrid relative to the catalytic site. Additionally, it plays a role in translocating the nucleic acid substrates through the active site at each single nucleotide step. The trigger loop influences the bridge helix conformations. Besides, it controls the catalytic functions of RNAP through conformation-specific contacts with the NTP in the nucleotide insertion site[17]. The interaction of the TL with the BH spans the main channel and secondary channel and the loop can adopt ‘open’ and ‘closed’ conformations (Figure 2)[18].

Substrate loading happens in two steps. Firstly, an NTP substrate binds to an open active centre conformation forming a pre-insertion state. Secondly, the NTP moves slightly to occupy the insertion site as the trigger loop folds and thus the complex isomerizes to the catalytically competent, closed ‘insertion’ state. The closure of the active site with NTP inside creates required contacts for the NTP selection and catalytic nucleotide incorporation with the release of pyrophosphate. Herein, TL unfolds and the active site opens. This is followed by a cooperative interaction between BH and TL to move the nascent hybrid base pair out of the active site and allowing a new template base to enter the active site as the culmination of NAC[19][15].

Therefore, the nucleotide addition cycle (NAC) could be summarized in three main steps: NTP entry and binding to EC, a chemistry step where incorporation of the nucleotide into the nascent transcript via phosphodiester bond takes place (NTP condensation, with PPi release), and finally translocation to free the NTP insertion site[16]. NTP condensation is thought to occur through a general two-metal-ion mechanism where two magnesium cations, coordinated by aspartate residues located in the active site is employed. Mg^{2+} ion A (Mg-A) is bound by three aspartate residues in the β' catalytic loop. It is primarily responsible for the activation of the RNA $3'$-terminal OH group by promoting deprotonation that facilitates $3'$ O$-$ attack on the NTP substrate α-phosphate forming a new phosphodiester bond and a leaving group, PPi. Meanwhile, the second Mg ion (Mg-B) is involved in the in stabilization of the penta covalent transition state. (Mg-B) coordinates the $\alpha-, \beta-$ and γ-phosphates of the NTP and $3'$-OH group on the α $-$phosphate[16][20].

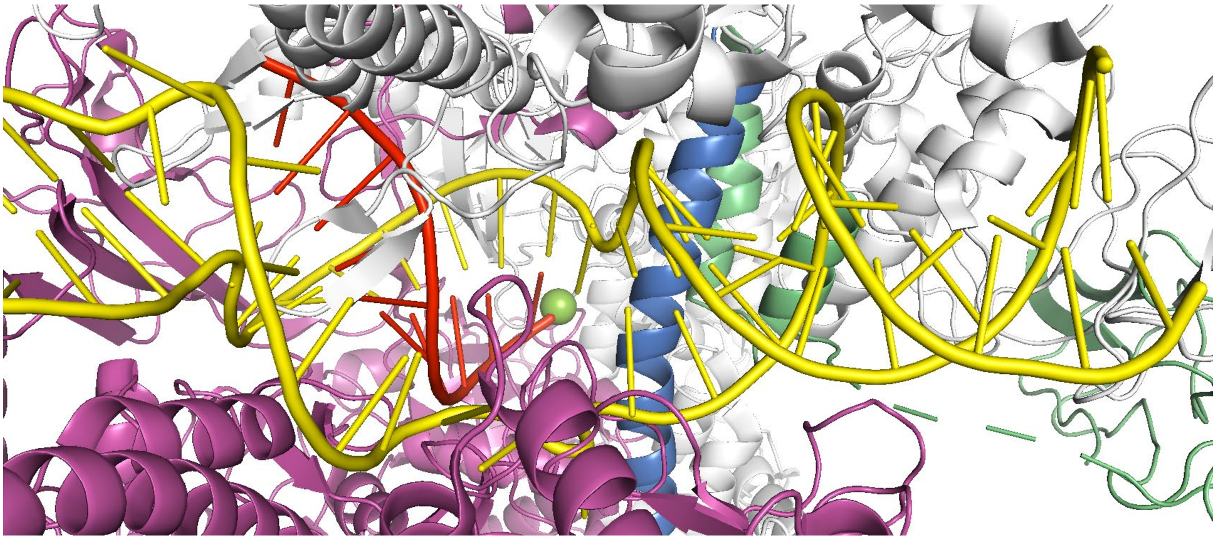


Figure 2: The active site of *E. coli* RNAP transcription elongation complex (PDB ID: 6ALG). DNA and RNA are shown in yellow and red, respectively. The β subunit is shown in pink, and the β' subunit is shown in white. Trigger loop (TL) is shown in light green, and bridge helix (BH) is shown in light blue. (TL) and (BH) are coloured differently from β' subunit for distinction. Catalytic Mg^{+2} is shown as a green sphere.

1.2 RNAP inhibitors and their mechanism of Inhibition

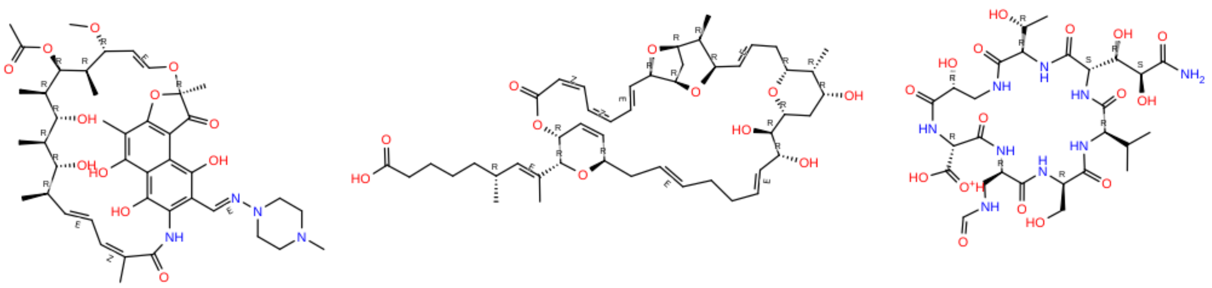
Although only a very few RNAP inhibitors are used clinically, there are many natural and synthetic compounds now known to target RNAP and inhibit different stages of the transcription cycle. The knowledge of these inhibitors and their mechanism of action (MOA) is important in the development of efficacious antibiotics in the future[21]. Thus, we will currently discuss those inhibitors and their MOA.

1.2.1 Antibiotics blocking the extension of the nascent RNA

Rifamycins (RIFs) are antibiotics derived from a natural source and belong to the ansamycin class. The structure of RIFs contains a polyketide chain, named ansa bridge. It spans a naphthalene moiety from their non-adjacent ends forming a cycle (Figure 3). Rifampicin (RMP), rifaximin (RXM), rifapentine (RPT) and rifabutin (RBT) are the RIFs that are currently utilized clinically. RMP is the first-line treatment for mycobacterial infections such as tuberculosis (TB). The MOA of RIFs relies on their strong binding to prokaryotic RNAPs leading to the inhibition of DNA-dependent RNA synthesis. Rifampicin (RMP) binds to the RIF-binding pocket within the DNA/RNA channel that lays 12 Å away from the catalytic Mg^{2+} ion in the β subunit of RNAP. Consequently, it hinders the formation of the second or third phosphodiester bond via steric blockage, which induces the release of short abortive RNAs preventing the extension of the RNA[21].

Sorangicin (SOR) is a naturally occurring macrolide polyether antibiotic. Similar to RIFS, SOR binds to the RIF-binding pocket and consequently inhibits transcription initiation of RNAP via the same mechanism as RMP. However, the structure of SOR lacks the naphthyl moiety. Thus, it shows more flexibility compared to RMP (Figure 3). This flexibility confers advantage to SOR over RMP as it makes it less sensitive to the conformational changes of the RIF-binding pocket. Hence, it retains activity against some of the RMP-resistant mutant strains[21].

GE23077 (GE) is a natural product antibiotic that specifically inhibits RNAP (Figure 3). GE binds to RNAP and occupies the RNAP i and $i + 1$ sites. This sterically blocks the binding of the initiating NTPs, thus inhibiting transcription initiation[21].



Rifampicin

Sorangicin A

GE23077

Figure 3: The two-dimentional (2D) structures of RNAP inhibitors that block the extension of the nascent RNA, rifampicin, sorangicin A, and GE23077.

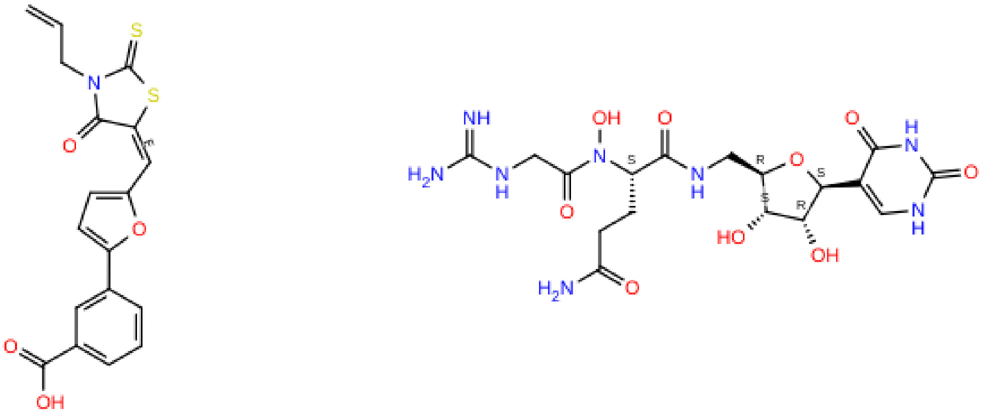
1.2.2 Inhibitors disrupting the holoenzyme assembly

Bacterial σ factors are crucial to initiate transcription. Moreover, they are divergent from the mammalian transcription factors while the interface between the σ factors and the core is highly conserved, which makes them a very attractive target for bacterial inhibition. SB-2 series is a low molecular weight synthetic compound discovered through high-throughput screening via a robust enzyme-linked immunosorbent assay (ELISA) technique (Figure 4). These compounds were found to be potent inhibitors of the *E. coli* RNAP by specifically inhibiting σ binding to the RNAP core. Additionally, the mechanism of action of SB compounds is suggested to be allosteric as they were able inhibit transcription even after the formation of the holoenzyme[22][23][21].

1.2.3 Nucleoside analogues

Nucleoside analogues' (NAs) significant potential has been demonstrated as early as the 1980–1990s after the introduction of various anti-HIV drugs such as azidothymidine (AZT). NAs was utilized as the backbone of the chemotherapy against several infectious viral diseases as HIV, herpes simplex, and hepatitis B and hepatitis C viruses[24][25]. NAs generally resemble the structure of endogenous nucleosides so that they would be incorporated into the DNA or RNA replication cycle by cellular or viral enzymes. Nevertheless, NAs contain one or more modifications compared with the original substrate. As a result, their incorporation into the growing RNA/DNA results in the disruption or the termination of replication[26]. NAs bind directly to the conserved active site of the viral polymerase after they are converted to their triphosphate active form[24].

Pseudouridimycin (PUM), a compound that was initially identified through the screening of microbial extracts (Figure 4). It is considered the first nucleoside analogue that can specifically inhibit bacterial RNAP[27]. PUM inhibits bacterial transcription by competing with uridine triphosphate (UTP) for the $i + 1$ site. PUM loses its activity with the DNA templates that lack the direct incorporation of UTP, which confirms that PUM can only prevent the incorporation of UTP but not adenosine triphosphate (ATP), guanosine triphosphate (GTP) or cytidine triphosphate (CTP). The crystal structures of *T. thermophilus* RNAP transcription initiation complexes that contain PUM confirmed that the compound targets the $i + 1$ site where it forms Watson–Crick interactions with the template strand. Besides, it interacts in a manner similar to NTPs with various residues within the $i + 1$ site[27][21].



SB-2

PUM

Figure 4: 2D structures of RNAP inhibitor disrupting the holoenzyme assembly, SB-2, and the nucleoside analogue RNAP inhibitor, pseudouridimycin.

1.2.4 Blocking the mobile elements in the primary channel of the active site

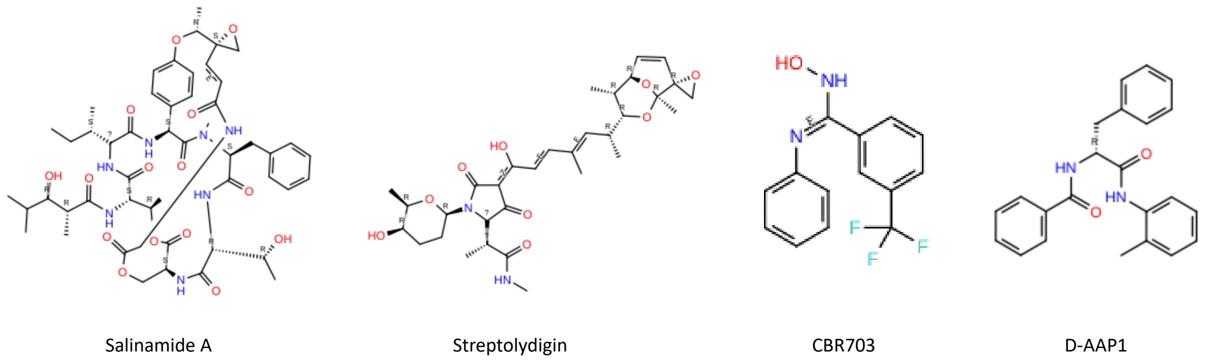
Salinamides (SALs) A – E are a group of depsipeptides derived from a marine *Streptomyces* sp., strain CNB-091 (Figure 5). Salinamides A, B and F were demonstrated to be effective inhibitors of bacterial RNAPs. SALs inhibit the transcription initiation and elongation through the interaction with the 'bridge-helix cap' in the RNAP active centre. The crystal structure of SAL A complexed within the *E. coli* RNAP holoenzyme shows that SAL binds to the unbent (straight) state of the RNAP BH N-terminal hinge (BH-HN) and makes direct interactions with the fork loop, BH-HN and the link region. The bent conformation of BH-HN is critical for the chemical reactions performed by RNAP such as bond formation and pyrophosphate-release. Thus, it is essential for the RNAP catalytic activity. Therefore, it is proposed that SALs stabilise the RNAP BH-NH, which prevents the conformational changes necessary for nucleotide addition. Consequently, they allosterically inhibit transcription[28][29][21].

Streptolydigin (STL) is an antibiotic with broad spectrum antibacterial activity derived from natural origin. It is comprised of a tetramic acid containing a sugar and a streptolol (Figure 5). The crystal structures of *T. thermophilus* RNAP containing STL demonstrate that STL binds in the vicinity of the active site, with its streptolol moiety directly interacting with two regions of β′ and with BH-NH. Meanwhile, its opposite side occupies a region close to the central part of BH and TL. Besides, the sugar moiety interacts with downstream DNA and TL. The MOA of STL is suggested to be locking the active site in a conformational state that disfavours substrate loading by trapping the BH at a straight conformation and the TL at an open conformation.

Hence, STL blocks the conformations that are essential for NTPs binding, phosphodiester bond catalysis and translocation[21][30].

CBR703 (N-hydroxy-N′ -phenyl-3- trifluoromethyl-benzamidine) series inhibitors (CBRs) are a synthetic class of antibiotics with a structure comprising two linked aromatic rings (Figure 5). CBRs' target site is suggested to be a hydrophobic two-pocket site between the N-terminal portion of the BH and a surrounding cap region that is believed to undergo conformational changes during NAC. Thus, CBRs inhibit the catalytic activities of RNAP. Specifically, CBRs inhibit TL folding, which disrupts nucleotide addition and the intrinsic hydrolysis of 3′-terminal dinucleotides in certain backtracked complexes. CBRs also have a TL independent mechanism that was speculated to involve the inhibition of BH movements necessary for the alignment of a reactive triphosphate and consequently affecting nucleotide addition and pyro phosphorolysis[21][31].

D-AAP1(D-Nα-aroyl-N-aryl-phenylalaninamides) binds to a pocket in the centre of BH-NH, having a similar binding pocket with CBRs. It is speculated that they share the same MOA. However, D-AAP1 is effective against *M. tuberculosis* RNAP but shows poor activity against other bacterial RNAPs. The structural differences between CBRs and D-AAP1 determine their selectivity against Gram-negative bacteria (i.e., *E. coli*) and *M. tuberculosis*, respectively. The two-ringed CBRs fit into the two-pocket site on *E. coli* RNAP while the three-ringed D-AAP1 accommodates into the three-pocket site in *M. tuberculosis* (Figure 5)[21][32]. D-AAP1 and RIFs are capable of binding simultaneously to RNAP, which confirms that D-AAP1 and RIFs binding sites do not overlap. Consequently D-AAP1 avoids cross resistance with RIFs. Besides, the ability of D-AAP1 and RIFs to bind simultaneously to RNAP indicates that their co-administration could offer additive antibacterial activity[32].



Salinamide A Streptolydigin CBR703 D-AAP1

Figure 5: The 2D structures of RNAP inhibitors that block the mobile elements in the active site, salinamide A, streptolydigin, CBR703, and D-AAP1.

1.2.5 Blocking the uptake of NTP

Microcin J25 (Mcc25) is a 21 amino-acid cyclic peptide antibiotic that inhibits both transcription initiation and elongation. Saturation mutagenesis experiments indicated that Mcc25 occupies the secondary channel that branches off from the main active site channel and acts as a path that allows the incoming NTP substrates to reach the catalytic centre. Therein, Mcc25 RNAP inhibitory mechanism is proposed to be blocking the secondary channel and thus also the uptake of NTP substrates into the active centre. In accordance with the proposed mechanism, it increases the Km of NTP binding[21][33]. Additionally, Mcc25 inhibits the RNAP catalysed pyrophosphorolysis. Nonetheless, some of Mcc25 resistance determinants (changes in amino acids that lead to resistance) overlap with those of the STL-binding pocket in the β subunit, suggesting that Mcc25 MOA might be more complex. Besides, Mcc25 and STL binding is mutually exclusive, which indicates that they may share the same MOA[34][21]

1.2.6 Inhibitors targeting the switch region

As discussed previously, the RNAP clamps exists in a range of distinct conformational states, from a fully open clamp conformation that allows the entry and the exit of DNA to a closed clamp conformation that prevents DNA from entry and exit. The clamp opens at the early stages of transcription initiation to allow DNA entry into the active site. Then, the clamp closes to retain DNA in the vicinity of the active centre during later stages of transcription initiation and transcription elongation. The switch region is a flexible hinge at the base of the clamp. It is composed of five discrete elements (SW1–SW5) of so called ‘switches’ and it allows a 30^{o} swinging motion of the clamp between both conformations. During open and close clamp conformational states, the switch region adopts different conformations. There is a direct contact between residues of switch 1, switch 2, and switch 3 of the switch region and loaded, unwound DNA phosphates in the transcription elongation complex. This direct contact is proposed to coordinate and mechanically couple clamp closure, DNA binding and DNA unwinding[35][2][21][36]. Switch 2 and switch 3 also form direct contacts with the nascent RNA in the transcription elongation complex. The switch region is highly conserved across both Gram-positive and Gram-negative bacteria. Thus, inhibitors targeting switch region will show broad spectrum activity. Besides, the switch region does not overlap with the rifamycin binding site and therefore, inhibitors targeting the switch region will not share cross-resistance with rifamycins[36].

Fidaxomicin (Fdx) is an RNAP inhibitor that binds to RNAP switch region and is used in the treatment of *Clostridium difficile*-associated diarrhoea (Figure 6). Normally, RNAP open RNAP-promoter complex (RPo) conformation depends on the recognition of −10 and −35 promoter elements by the σ factor substructure region 2 (σR2) and region 4 (σR4), respectively. Fdx is proposed to prevent the correct spatial orientation of σR2 and σR4 necessary for the simultaneous recognition of −10 and −35 core promoter elements and consequently locking the clamp in an open conformation. RNAP–Fdx complexes manage to bind to upstream promoter elements, however, they are unable to recognise the −10 element and fail to nucleate promoter melting. Thus, Fdx inhibits transcription by trapping an open-clamp conformation[37][21].

Myxopyronin (Myx), corallopyronin and ripostatin are natural product transcription inhibitors that target the switch region and bind in an adjacent pocket to that of Fdx (Figure 6) (Figure 7). Myx is the most extensively studied among the aforementioned compounds. Myx interacts with switch 1 and the opposing face of switch 2, locking the RNAP clamp in a closed conformation. Biochemical analysis of Myx suggests that it does not prevent nucleation or melting of the promoter DNA but instead blocks the promotor melting propagation from reaching transcription start site. Thus, it may target a late initiation intermediate en route to open state conformation. Regardless, Myx can inhibit transcription at artificially melted promoters, which suggests that Myx clamp locking interferes with critical interactions with the melted template DNA at the transcription start site[38][21]. Therein, there are two possible hypotheses regarding Myx inhibitory mechanism. In the hinge-jamming model, Myx interferes with the clamp conformational change and promoter DNA melting while in the steric clash model, it changes the conformation of switch 2, interfering with the enzyme's ability to further accommodate the template DNA into the RNAP active-centre[2]. Myx hydrophobic contacts probably have a dominant role in binding within the binding site. However, polar groups also make specific interactions with the protein[38].

Squaramides (SQs) are small synthetic RNAP inhibitors that also target the switch region and bind within the same pocket as Myx (Figure 6) (Figure 7). The middle part of the switch 2 (β′ residues 338−340) forms a short α-helix and occupies the squaramide binding pocket. Therefore, the binding of SQ causes these residues to be pushed into the DNA binding main channel. This displacement of switch 2 results in a clash with the correct positioning of template DNA at positions $i + 3$ and $i + 4$. Hence, SQ MOA is speculated to be similar to Myx through either locking the clamp in a closed conformation (hinge jamming model) and/or through sterically blocking template DNA loading (steric clash model)[2][21].

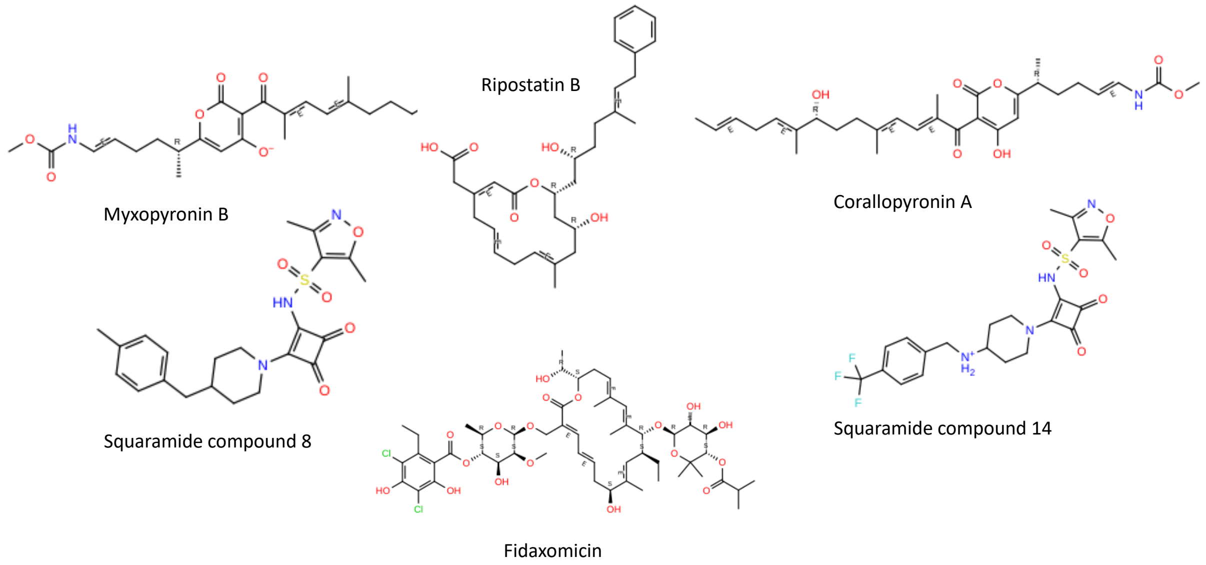


Figure 6: 2D structures of RNAP inhibitors targeting the ‘switch region’, myxopyronin B, ripostatin B, corallopyronin A, squaramides, compound 8 and compound 14 and fidaxomicin.

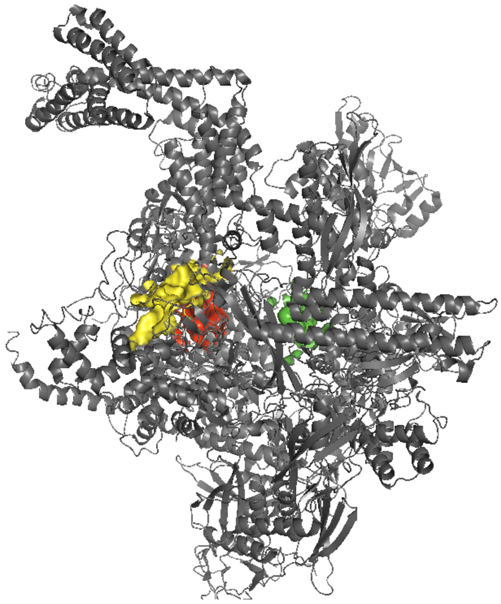


Figure 7: Target sites for RNAP inhibitors, rifamycins (in green), fidaxomicin (in yellow) and myxopyronin and squaramides (in red) as observed in crystal structures of (PDB IDs: 1YNN, 6FBV, 4YFX, 4YFK).

1.3 Computer-aided drug design (CADD)

The process of introducing a new drug in a market is very complicated, risky, and expensive; it may take 10-14 years and cost more than 1 billion dollars in capital. Thus, computational approaches in drug design, discovery and development process have been explored much and are today routinely implemented[39] to help reduce the drug development time and the associated costs and risk factors. It has been estimated that a successful the application of CADD approaches can reduce the costs of drug discovery and development even up to

50%[40]. CADD is defined as the use of any software/program-based process for establishing a standard to relate activity to structure and constitutes two major approaches that are structure-based drug design (SBDD) and ligand-based drug design (LBDD)[41][39].

1.3.1 Structure-based drug design

In SBDD, the structure of the target protein is known, and commonly, unknown compounds are docked to the target site and their interactions and/or affinity are determined to design a new molecule with better interactions with the protein target[39]. The first SBDD step starts with purification and structure determination of the target protein with experimental methods (e.g., X-ray crystallography or NMR), or modelling the target by theoretical methods (e.g., homology/comparative modelling). In a process called virtual screening (VS), compounds of different databases are docked into a selected region (active site) of the protein. The docking poses are scored and ranked, and the top scoring compounds will then be tested experimentally. Hereafter, in the second cycle, the structure of the protein in complex with the most optimistic lead from the first step with best *in vitro* results is determined. This elucidated structure of the lead-target complex then helps further to optimize the lead to achieve a marked increment in its target specificity and/or binding affinity[42]

1.3.2 Ligand based drug design

Contrary to SBDD, in LBDD the three-dimensional (3D) structure of the target protein is not known but the structures of ligands that bind to the desired target site are known. Through utilizing these available ligands, a pharmacophore model can be developed to understand the necessary structural features that need to exist in a molecule to enable it to bind to a target active site. Pharmacophore-based approaches and quantitative-structure activity relationships (QSARs) are usually the LBDD approaches applied. It is assumed that compounds that show similarity in their structure will also exhibit similar biological activity and interaction with the target protein[39].

1.3.3 Virtual Screening

Virtual screening (VS) is a convenient, detailed, knowledge-driven, compound database searching approach for finding novel compounds with favourable biological activity. VS is an attractive computational alternative for experimental high-throughput screening with the purpose of finding novel alternatives for an existing ligand or novel ligands for unexplored putative drug targets. VS approaches use the structural information available for either the protein target (structure-based virtual screening, SBVS) or known active ligands (ligand-based virtual screening, LBVS) (or both). In LBVS, an estimation of similarity between the known active and a database compound is calculated whereas in SBVS docking to the target site is commonly used[39][43].

A cascade of sequential filters is used to narrow down large libraries of compounds and in the end a set of hits with potential biological activity against the intended drug targets are selected. There are multiple electronic compound databases available for VS, either from commercial sources or public databases that contain millions of available compounds that can be screened in a VS experiment[39][43]. Medicinal chemistry groups in academia or companies may also have their own internal databases that can be used for VS.

1.4 Structure-activity relationship of RNAP inhibitors targeting Myx-binding pocket within the switch region

Structural studies of antibiotics bound to their target protein allow rational SBDD, which provides a shortcut to medicine by enhancing the process of drug discovery and development. Moreover, techniques such as X-ray crystallography permit the capture of snapshots of dynamic intermediates that are 'frozen' after inhibitor binding[38]. Herein, we will present studies that explored the interactions of RNAP inhibitors binding to the Myx-binding pocket within the switch region of bacterial RNAP. Investigating the interactions between the targeted binding site and its currently known inhibitors can convey additional insight into the nature of the binding pocket and the forces that drive the binding of these inhibitors. Additionally, the key residues essential for binding can be identified. Therefore, the knowledge gained from studying the binding site can be used as an additional basis for the evaluation of the identified hits from VS.

1.4.1 Myxopyronin

Myx is a polyketide-derived α-pyrone antibiotic produced by the myxobacterium *Myxococcus fulvus*. Myx inhibits growth of a broad spectrum of Gram-positive and Gram-negative bacteria by inhibiting bacterial RNAP (IC50 ~1 mM). Importantly, it does not inhibit eukaryotic RNAP II[35].

Mukhopadhyay et al. carried out an extensive mutagenesis study of *E. coli* RNAP β′ and β subunits genes *in vitro.* The authors isolated and sequenced Myx-resistant mutants to identify and study the Myx target site[35]. Specific amino acid residues at the Myx target site were deliberately mutated, and the susceptibility of the mutated strains to myxopyronin was then examined. The substitutions in residues that confer resistance to Myx indicate that these residues are essential for the binding of Myx. Myx binds within the RNAP switch region and interacts directly with switch 1 and switch 2 (β′ residues 1319–1328 and 330–347). Besides, it makes direct interactions with β′ residues 1346–1357 and β residues 1270–1292 and 1318–1328 that form the adjacent segments of the β′ and β subunits (Figure 8). The nitrogen atom and both oxygen atoms of the enecarbamate moiety of Myx are engaged in a network of hydrogen bonds (H-bonds) centred on an ordered bound water molecule that yields an unequivocal electron density. The ammonium sidechain of β′ Lys1348, the carboxyl sidechain of β′ Asp802, and the indole NH sidechain of β Trp1276 are involved within this H-bonds network. The carboxyl sidechain of β Glu1279 forms a hydrogen bond with the enecarbamate moiety carbonyl oxygen of Myx. The hydroxyl group of β Ser1322 sidechain forms a potential hydrogen bond with the α-pyrone ring C2 carbonyl oxygen of Myx. Also, the NH of β′ Gly620 backbone forms a potential hydrogen bond with the C4 hydroxyl of the α-pyrone ring of Myx. Additionally, the β subunit residues 1034, 1038, 1041, 1271, 1275, 1279, and 1291 and the β′ subunit residues 801, 805, and 1348 make van der Waals interactions with the enecarbamate sidechain of Myx. Meanwhile, the β subunit residue 1322 and the β′ subunit residues 344, 345, 346, and 1352 form van der Waals interactions with the α-pyrone ring of Myx. Furthermore, methylene groups of β′ Lys345 sidechain underlie and create a platform for the α-pyrone ring of Myx, whereas the β subunit residue 1326 and the β′ subunit residues interact extensively with the dienone sidechain of Myx via van der Waals interactions (Figure 9)[35]. In the presence of Myx, a highly conserved β′ switch 2 segment refolds where the α-helix is interrupted in the middle by four flipped-out residues and is straightened while its carboxy-terminal part, comprised of ~two-helical turns, unwinds and refolds into a loop (Figure 10)[38].

Mukhopadhyay et al. identified four substitutions that are predicted to disrupt RNAP-Myx H-bonds and may introduce steric clash and they confer high level of resistance (≥16-fold). These are: β1279Glu→Gly, β1279Glu→Lys, βSer1322→Pro, and βSer1322→Val. In addition, they identified five substitutions that confer high-level resistance and are predicted to disrupt favourable RNAP-Myx van der Waals interactions and also introduce steric clashes: β′345Lys→Arg, β′345Lys→Asn, β′345Lys→Thr, β1275Val→Met, β1279Val→Phe[35].

In previous studies, Doundoulakis et al. synthesized and biologically evaluated various myxopyronin analogues via modifying the dienone and enecarbamate side chains while keeping the pyrone ring constant[44]. Lira et al. conducted a similar study as well, but they made modifications to the pyrone core[45]. The study conducted by Mukhopadhyay et al. was consistent and supported the structure-activity relationships (SAR) for synthetic analogues of Myx reported by Doundoulakis et al. and Lira et al. with only one exception[44][44][35]. Lira et al. reported that O-methylation of the C4 hydroxyl of the α-pyrone ring does not have a major effect on the RNAP-Myx interaction. However, Mukhopadhyay et al. shows that this modification would disrupt an intramolecular hydrogen bond important for adjusting the orientation of the dienone sidechain relative to the α-pyrone ring and also introduce a steric clash profoundly impairing the RNAP-Myx interaction[45][35].

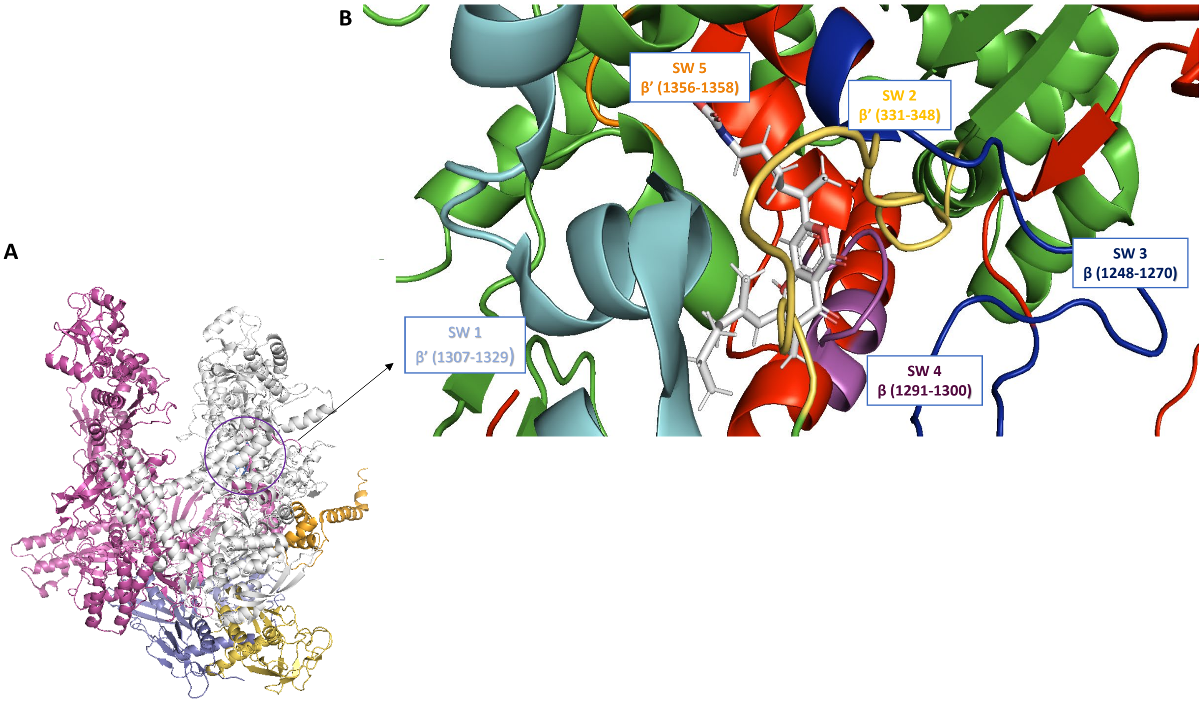


Figure 8: (A*) E. coli* RNAP complexed with myxopyronin B (PDB ID:4YFX). Violet circle: myxopyronin B binding pocket. (B) Magnified view of myxopyronin B binding pocket within the switch region. Myxopyronin B is depicted as a stick model in the interface between β (red) and β′ (green) subunits models; atom colour code: carbon atom (white), oxygen atoms (red), nitrogen atoms (blue). A unique colour is denoted to each of the five switches, and they are labelled correspondingly.

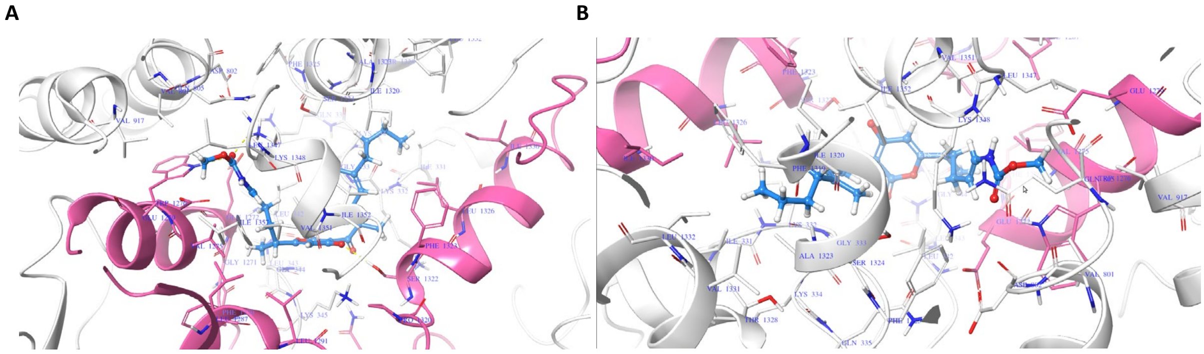


Figure 9: Binding pocket for Myx in *E. coli* RNAP (PDB ID. 4YFX). β and β' subunits depicted in pink and white, respectively. Myx is depicted in light blue. Oxygen atoms (red) and nitrogen atoms (blue). Side chains within proximity of 6 Å of the ligand, are shown in sticks and labelled. (A) Side view, (B) top view.

Corallopyronin (Cor) is another polyketide-derived α-pyrone antibiotic that is structurally related to Myx but with a seven-carbon sidechain extension[36][35]. Through the extensive mutagenesis study of the *E. coli* RNAP β′ and β subunits conducted by Mukhopadhyay et al., they demonstrated that the cross-resistance patterns of Myx and Cor are almost the same with the only exception of the β residue 1326 (β1326Leu→Trp). Therefore, they concluded that Cor interacts with a target site that overlaps the target site for Myx and that the β residue 1326 interacts with the seven-carbon sidechain extension present in Cor but not in Myx. β1326 is located in a close proximity to the ligand dienone sidechain terminus that is the point of attachment of the seven-carbon sidechain extension existing in Cor[35].

1.4.2 Ripostatin (Rip)

Rip is a polyketide-derived 14-membered macrocyclic-lactone antibiotic that is structurally unrelated to the other switch region-targeting inhibitors[35][36]. Despite, the lack of structural similarity between Rip and Myx or Cor, the cross-resistance patterns analysis done by Mukhopadhyay et al. indicated that all mutants that result in a high-level resistance to Myx and Cor cause also high-level resistance to Rip, including the β residue 1326 (β1326Leu→Trp). Nevertheless, Rip exhibits general similarity in size and hydrophobic character to Myx and Cor. Hence, the authors concluded that Rip most likely shares the same binding pocket as Myx and Cor[35].

1.4.3 Squaramides (SQs)

The squaramides (SQs) were identified in a high-throughput screen for novel inhibitors of a transcription-coupled translation assay using *Escherichia coli* S30 extracts. They are the first reported non-natural-product compounds targeting the switch region of RNA polymerase[46].

Molodtsov et al. determined the crystal structures of *E. coli* RNAP holoenzyme in complex with squaramide compounds 8 and 14, successfully confirming that the squaramides bind to the RNAP switches. The crystal structures revealed that the chemically distinct Myx and SQs bind to the same RNAP region (Figure 10). The SQs main interactions are with switch 1 and switch 2 while a few interactions are formed with switch 3 and switch 4. The positions of SQ and Myx within the *E. coli* RNAP are shown to mostly overlap. Regardless, they extend toward different sides within the binding pocket. The carbamate chain of Myx makes extensive contacts with the three α-helix bundle (the helix−turn−helix flanked by switches 3 and 4 of the β subunit and one helix from the β′ subunit), with its carbonyl group inserted deeply into the helix bundle, reaching the C-terminal end of the bridge helix. However, the squaramide contacts only the front edge of the three-helix bundle via its isoxazole ring. The left-hand side of squaramide interacts with switch 1 and the C-terminal α helix of the β subunit Thus, overlapping with the location of the dienone chain of Myx. The benzyl moieties of SQs, on the other hand, reach deeper, forming a larger interface than Myx. Although SQs induce structural changes to switch 2 like Myx, the switch 2 conformation is slightly different. The relatively narrow dienone chain of Myx interacts with its methyl group via van der Waals forces with β′K332, pulling switch 2 towards the Myx binding pocket. In SQ complexes, switch 2 is pushed out by the bulky piperidine that positions it deeper in the DNA binding main channel (Figure 10B)[2]. Additionally, Molodtsov et al. studied the SAR of SQ through a series of synthesized SQ analogues, revealing the key SAR information for these compounds. The preferred hydrophobic nature of SQ terminal substituents is compatible with the largely hydrophobic binding pocket of SQ. The terminal benzyl ring of SQ interacts with βL1326 through van der Waals contacts, playing a key role in the SQ and RNAP interaction (Figure 11). Also, the existence of the benzyl ring at a precise distance and orientation from the isoxazole is necessary. Introducing a trifluoro modification at the benzyl ring improved the biochemical potency by almost 10-fold. However, the cocrystal structure of the compound indicates that the binding pocket is still not fully occupied. Therefore, squaramide modifications that introduce hydrophobic group(s) at the benzyl ring are predicted to enhance the squaramide binding affinity. However, a balance between increasing hydrophobicity and the decreased free fraction as a result of the increased binding to plasma proteins, needs to be studied[2].

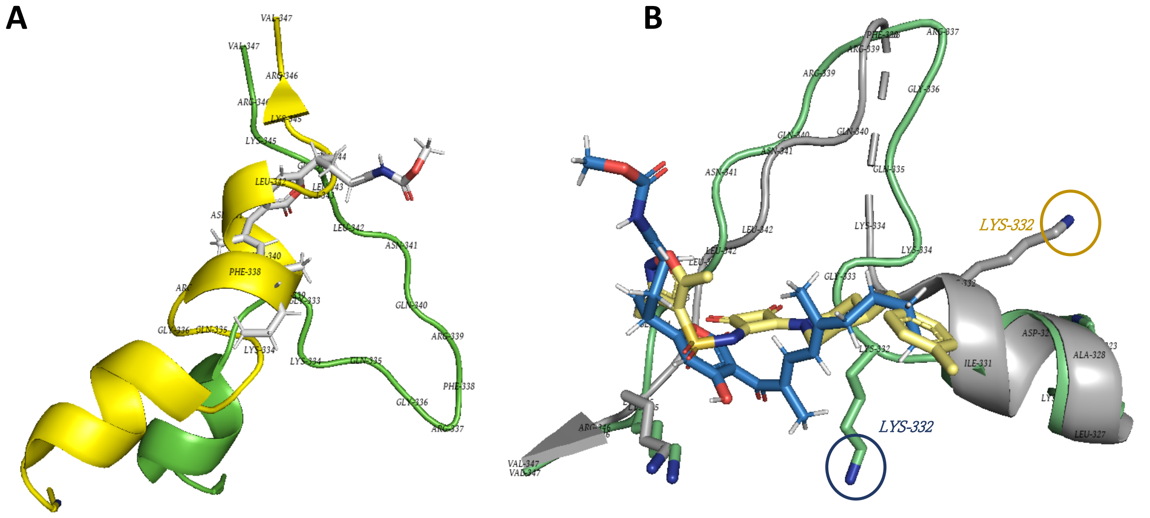


Figure 10: RNAP β’ switch 2 segment. (A) *E. coli* RNAP (PDB ID: 4LK1) (in yellow) superimposed with *E. coli* RNAP (in green) in complex with myxopyronin B (in white) (PDB ID: 4YFX) showing the α helix unwound and refolded into a loop. (B) *E. coli* RNAP (in green) in complex with myxopyronin B (in blue) (PDB ID: 4YFX) superimposed with *E. coli* RNAP (in grey) in complex with squaramide compound 8 (in yellow) (PDB ID: 4YFK) showing β′K332 (circled and labelled in blue and gold, respectively) pulling switch 2 towards the Myx binding pocket while pushing it out deeper in the DNA binding main channel in the SQ complex.

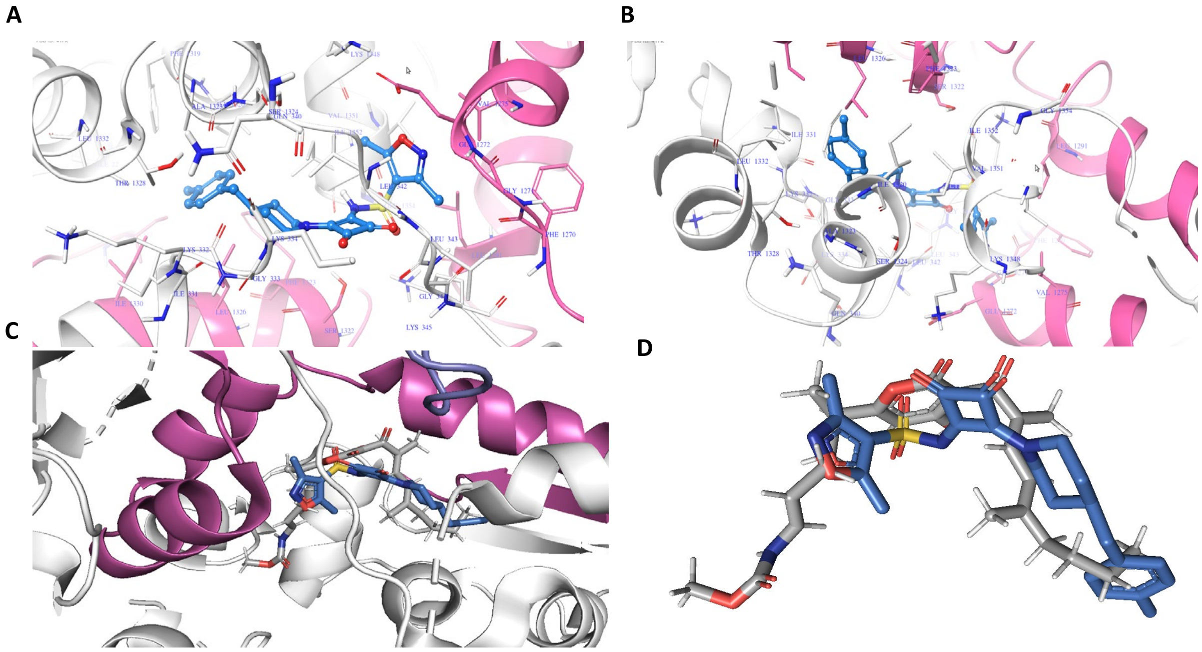


Figure 11: Binding pocket for squaramide in *E. coli* RNAP (PDB ID: 4YFK). β and β' subunits depicted in pink and white, respectively. Squaramide compound 8 is depicted in light blue. Atom colour code: oxygen atoms (red) and nitrogen atoms (blue). Side chains within 6 Å of the ligand, are shown and labelled. (A) Side view, (B) top view. (C) Squaramide compound 8 superimposed with myxopyronin B (grey) within the switch region (PDB IDs: 4YFK and 4YFX), side chains and labels are removed for clarity. (D) Protein from (C) is removed for clear illustration.

1.4.4 5-phenyl-3-ureidothiophene-2-carboxylic acids

In order to develop novel antibacterials with a broad-spectrum activity, Sahner et al. explored the SAR of a previously identified hit candidate, compound 3, targeting RNAP switch region via pharmacophore-based virtual screening (Figure 12). The hit showed an IC50 value of 75 mM and a minimum inhibitory concentration (MIC) value of 11 mg/mL in an *in vitro* assay against *E. coli* TolC, a TolC mutant strain of *E.coli.* TolC is a surface protein in *E. coli* involved in diverse cellular functions such as antibiotic efflux and alpha-hemolysin secretion and thus the multidrug AcrAB–TolC efflux system in the *E. coli* TolC strain is defective. The docking pose of the hit compound 3 suggests that it may bind in a tilted conformation similar to the myxopyronins. Myx adopted a U-shaped conformation that was observed to fill the 'switch region' binding pocket in the crystal structures. The thiophene core is anchored to Lys334 by H-bond or ion-pair interactions with its carboxylic acid moiety, placing it on the top of the entrance to the switch 2 binding cavity. The 4-chlorophenyl ring is shown to fit in the lower part of the enecarbamate-binding pocket of myxopyronins with its chloro atom fitted into a small pocket formed by Leu343, Gly344 and Lys345 (β′ subunit) from one side, and Phe1270, Gly1271, Val1275 and Leu1291 (β subunit) on the other side. Meanwhile, the ureido moiety is stabilized by an intramolecular hydrogen-bond with the carboxylic acid group overlapping the location of the dienone side chain of myxopyronins. This intramolecular hydrogen bond directs the ethyl and benzyl lipophilic substituents into the hydrophobic pocket formed by Leu1326 and Ile1337 (β subunit) and Phe1319, Ile1320, Ala1323, Thr1328, Ile1352 (β′ subunit)[47].

Herein, Sahner et al. synthesized a small library of 5-phenyl-3-ureidothiophene-2-carboxylic acids based on compound 3 to explore and optimize the SAR of the hit compound 3. The new compounds exhibited both good antibacterial activities against Gram-positive bacteria and Gram-negative *Escherichia coli* TolC and a reduced resistance compared to rifampicin. Compound 23 was synthesized by elongating the phenyl ring A with a myxopyronin enecarbamate chain mimicking group to occupy the enecarbamate chain pocket (Figure 12). Indeed, compound 23 possessed increased activity compared to the parent compound 3 with IC50 against *E. coli* RNAP of 20 mM and MIC against *E. coli* TolC of 25 mg/mL, which also supports the proposed binding mode. Substitution on the ureido motif B showed an increased activity in a compound series derived from primary alkyl amines with chain elongation. For example, mono substituted n-hexyl derivative displayed the highest potency compared to smaller substituents. Moreover, the introduction of a second chain at the nitrogen generally led to an increased activity, which is attributed to the additional hydrophobic contacts within the lipophilic binding pocket. These additional hydrophobic moieties better occupy the dienone-binding pocket, reaching into the pocket formed by Leu1326 and Ile1330 (β subunit), and Lys332, Ser1324, Thr1328, and Leu1332 (β′ subunit). The resulting improved activity could be explained by the fact that ureido motif B lies in a highly lipophilic pocket where hydrophobic contacts, van der Waals contacts, and CH-pi interactions dominate. The urea moiety seems not to form hydrogen bonds with the surrounding residues, but it is essential as a planar linker for the obligatory hydrophobic groups[47].

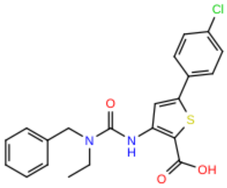
1.4.5 Hybrid-type myxopyronins

Dithiolopyrrolones are a class of antibiotics with a pyrrolinonodithiole (4H-[1,2] dithiolo [4,3-b] pyrrol-5-one) structure linked to two variable acyl groups. There is around 30 naturally occurring di-thiolopyrrolons such as holomycin (Figure 12). Dithiolopyrrolons exhibit relatively broad-spectrum antibacterial activity against both Gram-positive and Gram-negative bacteria[6]. Holomycin has antibacterial properties with moderate broad antibiotic spectrum. Besides, it was shown to inhibit RNA synthesis[48]. Since myxopyronins are very hydrophobic and exhibit a very high binding capacity to plasma proteins, Yakushiji et al. were inspired to hybridize myxopyronin with a holothin (deacetyl holomycin) molecule and designed and synthesized such novel hybrid-type inhibitors of bacterial RNAP. They designed and synthesized two types of hybrid derivatives; one is a right-hand amide derivative with an α-pyrone core at the centre and the holothin moiety positioned on the right-hand side, while the other is a left-hand amide derivative with holothin introduced at the left side of the molecule relative to the C3 side chain. A molecular docking study predicted that the switch region would be the binding site of the novel compounds. The left side of the myxopyronin binding site contains a useful space around the C20 hydrocarbon chain that seemed advantageous for the holothin moiety to fit in and consequently increase the interaction with RNAP. Thereafter, they synthesized a series of derivatives from both the right- and left-hand side derivatives that were tested for the antimicrobial activity. The results showed that the left-hand moiety of the α-pyrone component had a substantial effect on the antimicrobial activity, and the introduction of holothin enhanced this activity. Derivative 29, specifically, showed promising antibacterial activity against Gram-positive bacteria (Figure 12). Also, it showed *in vitro* inhibitory activity against *E. coli* RNAP with an IC50 of 14 ± 1.7 µM. However, the MIC value of derivative 29 showed that it is ineffective against *E. coli*, which could be a result of the low penetration ability in Gram-negative bacteria. The authors also concluded that a long hydrocarbon chain moiety at the C3 position of the α-pyrone appears to be critical for the expression of the antimicrobial activity[48].

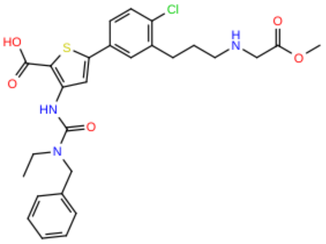
1.4.6 N-aryl pyrrothine-α-pyrone hybrids

Tan et el. [6] designed, synthesized, and assessed a series of novel hybrids of N-aryl pyrrothine-based α-pyrone as bacterial inhibitors of RNAP. Following the work of Yakushiji et al.[48], they used the pyrrothine core discovered in derivative 29 as a template for further optimization. Among this series, compound 13c showed strong antibacterial activity against antibiotic-resistant *S. aureus* with a MIC value in range of 1–4 µg/mL. Besides, it showed an IC50 value of 16.06 µM against *E. coli* RNAP (Figure 12).

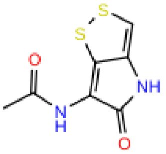
To further investigate the binding mode of the novel N-aryl pyrrothine-α-pyrone hybrids with bacterial RNAP, Tan et al.[6] docked all compounds into the switch region of bacterial RNAP (Protein Data Bank ID: 3DXJ; *T. thermophilus*). The authors observed that the 2, 4-dimethoxylphenyl at N-4 position of pyrrothine core formed a direct hydrogen bond with Trp1038 (*thermophilus* numbering). Also, it interacts with Trp1038, Asp1100 and Lys1463 indirectly through utilizing a water molecule (1539) as a bridge. The exact same amino acid interactions are seen with the myxopyronin enecarbamate moiety. The linker carbonyl group of compound 13c forms two further hydrogen bonds with Ser1084 and Lys621. The hydrogen bond with Ser1084 is also present with myxopyronin. Compound 13c forms hydrophobic interactions with a myriad of amino acids, including Leu1435, Ser1439, Phe614, Leu607, Glu1091, Glu1088, Val1037, Ile1467, Val1466, Leu1053 and Leu619, some of which interact also with myxopyronin. The docking pose of 13c shows its α-pyrone moiety adopting a similar position as the 'left-hand' side chain of myxopyronin A, which may indicate that they have the same interactions with the binding pocket.



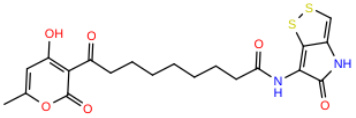
5-phenyl-3-ureidothiophene-2-carboxylic acids
Compound 3



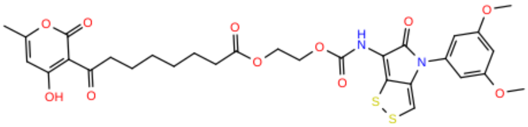
5-phenyl-3-ureidothiophene-2-carboxylic acids
Compound 23



Holomycin



Hybrid-type
Derivative 29



N-aryl pyrrothine-α-pyrone
Compound 13c

Figure 12: 2D structures of holomycin and miscellaneous RNAP inhibitor hits that target the switch region.

2 Aim and outline of the work

The objective of this computational study was to identify new bacterial RNAP inhibitors that target the enzyme switch region using a structure-based virtual screening approach. To enhance the drug discovery process, we developed a virtual screening pipeline that integrates several computational methodologies.

As the target protein for the *in-silico* studies, a crystal structure of *E. coli* RNAP in complex with a squaramide compound 8 was chosen. The ability of the docking algorithm to produce reliable docking poses was validated through self-docking studies.

The designated libraries were pre-filtered and docked, and a Molecular Mechanics – Generalized Born Surface Area (MM-GBSA) based method was used to estimate the binding free energy of the top scoring docking poses.

To aid in selecting the most promising candidate compounds for further analysis, a quantitative structure-activity (QSAR) model was developed based on the previously reported IC50 values of the compounds that bind to the myxopyronin-binding site[35][44][2][47][48]. The ligands with the most favourable MM-GBSA binding energies were then evaluated using the QSAR model. Consequently, the compounds predicted to have the best IC50 values by the model were then chosen to be analysed by molecular dynamics (MD) simulations for their stability and binding affinity throughout the simulation. The results of the MD simulation of 17 chosen candidate compounds were compared with the MD simulations carried out with the apo protein and with a reference co-crystallized ligand in the RNAP binding site. Eight candidate compounds that showed comparable binding to the RNAP site to the reference ligand, were selected for further biological testing (Scheme 1).

Scheme 1: Schematic outline for the virtual screening protocol applied in the study.

3 Material and Methods

3.1 Library preparation and filtration

We used the available MolPort molecular databases, the full database as of November 2021 (*fulldb*) and MolPort-Natural-Products (NP-db) in a structure-based virtual screening. The databases were downloaded in SMILES format in which the (NP-db) database includes 119,055 compounds of natural origin, and the *fulldb* includes 7,781,175 compounds. The databases were prepared with the Maestro (Schrödinger Release 2021-4: Maestro, Schrödinger, LLC, New York, NY, 2021) LigPrep tool (LigPrep, Schrödinger, LLC, New York, NY, 2021) converting the ligands into low-energy 3D structures. The default settings were applied, i.e., the OPLS4 force field[49] was used, the ionization states were set at a pH of 7.0+/-2.0 via Epik (Schrödinger Release 2021-4: Epik, Schrödinger, LLC, New York, NY, 2021), tautomers were generated, and a maximum of 4 stereoisomers per ligand. Altogether 15,007,020 compounds (including the tautomers and stereoisomers) were generated from the *fulldb.* As a result of technical issues while processing the (NP-db), some compounds have failed to be processed and only 55,982 compounds were available after the LigPrep step.

The generated molecules from LigPrep were then subjected to QikProp prediction tool to predict pharmaceutically relevant drug-properties (QikProp, Schrödinger, LLC, New York, NY, 2021).

Thereafter, the prepared ligands of *fulldb* were filtered using the Maestro Ligand Filtering Panel (Schrödinger Release 2021-4: Canvas, Schrödinger, LLC, New York, NY, 2021) where a compound that matches one or more of the following criteria was filtered out: 1) contains more than 5 hydrogen bond donors, 2) contains more than 10 hydrogen bond acceptors, 3) has clogP higher than 5.5 or less than 2, 4) has a molecular weight (Mwt) higher than 500 Daltons, 5) contains one or more reactive functional group. A somewhat less stringent filtering criteria was applied in case of the relatively small NP-db where a compound was filtered out if it meets one of the following criteria: 1) contains more than 6 hydrogen bond donors, 2) contains more than 11 hydrogen bond acceptors, 3) has cLogP higher than 6 or less than -1, 4) has a Mwt higher than 600 Daltons, 5) contains one or more reactive functional group.

The applied criteria highly conform with the Lipinski rule of 5[50] with the aim to increase the drug-likeness of the compounds to be screened. Also, removing reactive compounds decreases the likelihood of toxicity. 9,730,802 compounds from *full-db* and 29,181 compounds from NP-db passed the filtration process successfully.

3.2 Protein structure and preparation

3D structures of the target receptor RNAP were downloaded from the RCSB Protein Data Bank (PDB)[51]. Within PDB, there are several reported X-ray crystal structures with co-crystallized ligands bound at the myxopyronin binding pocket of either *E. coli* RNAP or *T. thermophilus* RNAP. The selected PDB structures were initially used for binding site analysis and self-docking studies. The PDB IDs of the structures were: 3DXJ with 3.000 Å resolution, 4YFX with 3.840 Å resolution, 4YFK with 3.570 Å resolution, and 4YFX with 3.820 Å resolution. 3DXJ is the crystal structure of *T. thermophilus* RNAP holoenzyme in complex with the myxopyronin, while 4YFX, 4YFK and 4YFN are crystal structures of *E. coli* RNAP in complex with myxopyronin B, squaramide compound 8 and squaramide compound 14, respectively[35][2].

The PyMOL Molecular Graphics System, version 2.0 (Schrödinger, LLC), and Maestro molecular modelling suite (Maestro, version 2021-4, Schrödinger, LLC, New York, 2021) were used for visualization and structure preparation. The protein chains not involved in the targeted binding site were removed from the crystal structures; thus, only chains C and D representing β and β′ subunits were left.

Myxopyronin binds within the same binding pocket in both *T. thermophilus* and *E. coli* RNAP switch region. However, the presence of minimal differences in amino acid sequences slightly modifies the position of enecarbamate chain of myxopyronin within the binding pocket. Histidine residue (β′ H1103) in *T. thermophilus* is present as glutamine (β′ Q803) in *E. coli*, which seems to sterically cause myxopyronin enecarbamate to slightly diverge from its position in *T. thermophilus* as was observed when 4YFX and 3DXJ were superimposed (Figure 13). Although this change does not seem to affect the binding of myxopyronin to *T. thermophilus* and *E. coli* RNAP, the chance of discrimination could happen with the computationally found new hits and since our future plan includes testing the resulting hits *in vitro* against *E. coli* RNAP, the 3DXJ crystal was excluded from the docking studies.

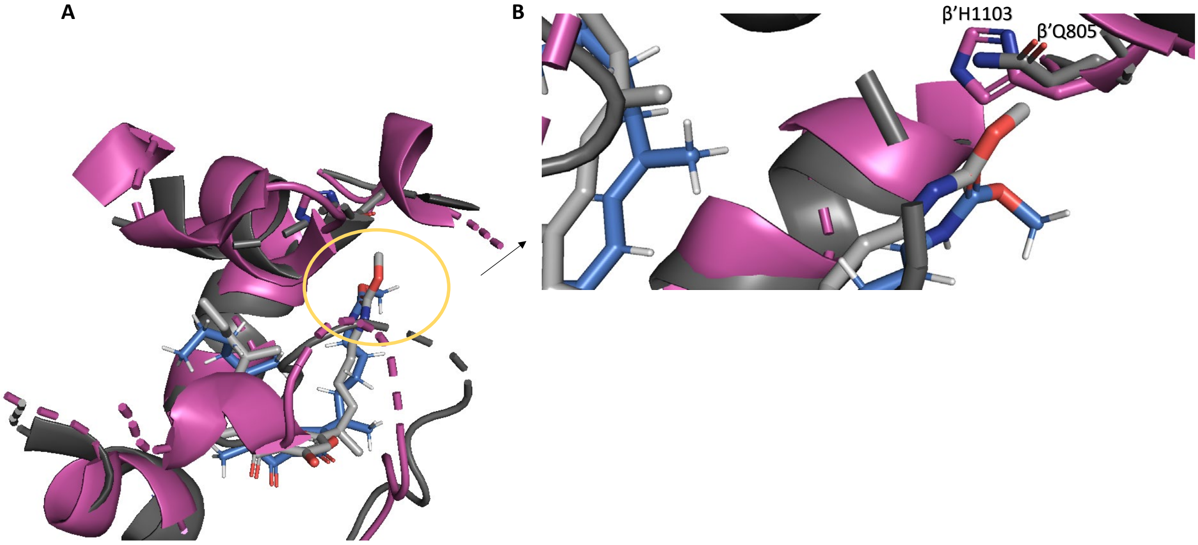


Figure 13: (A) Myxopyronin binding pocket in *T. thermophilus* RNAP (in pink) showing myxopyronin (in grey) (PDB ID:3DXJ) superimposed with myxopyronin binding pocket in *E. coli* RNAP (in grey) showing myxopyronin (in blue) (PDB ID:4YFX). (B) Zoom-in of the enecarbamate side chain in (A). The oxygen and nitrogen in ligands and side chain residues are coloured in red and blue, respectively.

The protein structures were the prepared using the Protein Preparation Wizard tool in Maestro (Schrödinger Release 2021-4: Protein Preparation Wizard; Epik, Schrödinger, LLC, New York, NY, 2021; Impact, Schrödinger, LLC, New York, NY; Prime, Schrödinger, LLC, New York, NY, 2021)[52]. The default options were used, in which hydrogen atoms, missing side chain residues and loops are added, the OPLS4 forcefield was used to minimize the protein structure and remove strain while allowing the convergence criteria of 0.30 Å RMSD for heavy atoms[49].

3.3 Docking and MM-GBSA

The docking site grid box was then generated for 4YFX. 4YFK and 4YFN structures using the Receptor Grid Generation tool of Maestro. The co-crystallized ligand in each crystal complex was used to locate the grid centre and the ligand was confined to an enclosing inner box of size 10 x 10 x10 A^3 and outer box of size 27.52 x 27.52 x 27.52 A^3.

Maestro Glide (Glide, Schrödinger, LLC, New York, NY, 2021) was employed for docking throughout the study, utilizing mainly the Glide standard precision (SP) and Glide extra precision (XP) modes[53][54]. Glide Score SP fitness function employs several terms which are, van der Waals energy (vdW), Coulomb energy (coul), lipophilic term (lipo-lipo), three hydrogen bonding energy terms (hbond-neut-neut, hbond-neut-charged, hbond-charged-charged), metal binding term (max-metal-ion), penalty for buried polar groups (polar-phob), penalty for freezing rotatable bonds (rotb) and solvation terms while terms are defined as in ChemScore with slight variations with equation as follows[53]:

$$\begin{aligned}\Delta G_{bind} = {} & C_{lipo-lipo}\sum f(r_{\mathrm{lr}}) + C_{hbond-neut-neut}\sum g(\Delta r)\,h(\Delta R)\\ & + C_{hbond-neut-charged}\sum g(\Delta r)\,h(\Delta R)\\ & + C_{hbond-charged-charged}\sum g(\Delta r)\,h(\Delta R)\\ & + C_{max-metal-ion}\sum f(r_{\mathrm{lm}}) + C_{rotb}H_{rotb} + C_{polar-phob}V_{polar-phob} + C_{coul}E_{coul}\\ & + C_{\mathrm{vd}W}E_{\mathrm{vd}W} + solvation\ terms\end{aligned}$$

f, g, and h functions give a full score (1.00) for distances or angles lying within the nominal limits and a partial score (1.00-0.00) for distances or angles lying outside these limits but inside larger threshold values.

Meanwhile, Glide Score XP fitness function is a more precise scoring function as it employs stricter desolvation penalties (desolv) and contact penalties (ligand-strain). Moreover, it includes more energy terms for hydrophobic enclosure (hyd enclosure) and pi-stacking (PI)[54]:

$$\text{Glide Score XP} = E_{\text{coul}} + E_{\text{vdW}} + E_{\text{bind}} + E_{\text{penalty}}$$

$$E_{\text{bind}} = E_{\text{hyd}}\ \text{enclosure} + E_{\text{hbond-neut-neut}} + E_{\text{hbond-charged-charged}} + E_{\text{PI}} + E_{\text{hb pair}} + E_{\text{phobic pair}}$$

$$E_{\text{penalty}} = E_{\text{desolv}} + E_{\text{ligand strain}}$$

Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) is a popular method to estimate the free energy of binding of a small ligand to a biological macromolecule. This approach combines the molecular mechanics energies and the generalized Born and surface area continuum solvation[55]. MM-GBSA is extensively utilized in structure-based drug design as it provides a higher accuracy than most docking scoring functions, achieving a good balance between computational efficiency and accuracy[56]. MM-GBSA binding free energy is calculated as follows:

$$\Delta G_{bind} = \Delta G_{Rec-Lig} - (\Delta G_{Receptor} + \Delta G_{ligand})$$

$$\Delta G_{bind} = \Delta G_{MM} + \Delta G_{solv} - T\Delta S$$

$$\Delta G_{MM} = \Delta G_{ele} + \Delta G_{vdw}$$

$$\Delta G_{solv} = \Delta G_{GB} + \Delta G_{SA}$$

in which, ΔGRec-Lig, ΔGReceptor and ΔGligand represent the change in Gibbs free energy of the receptor-ligand complex, the receptor protein, and the ligand, respectively. ΔG_{MM} stands for the molecular mechanics free energy change and it consists of electrostatic interaction (ΔG_{ele}) and van der Waals interaction (ΔG_{vdw}) terms. ΔG_{solv} term stands for solvation free energy change that constitutes the polar contributions of electrostatic solvation energy (ΔG_{GB}) and non-polar contributions of the non-electrostatic solvation (ΔG_{SA})[56]. Maestro's Prime was used to perform all the MM-GBSA calculations (Schrödinger Release 2022-1: Prime, Schrödinger, LLC, New York, NY, 2021).

3.3.1 Self-docking

Self-docking studies were performed to confirm the ability of the used software, Glide to reproduce the pose of the crystallized myxopyronin or squaramide in the active site and score it among the top poses. The co-crystallized ligands of 4YFX, 4YFK and 4YFN were docked each into its own respective protein binding site using the Glide standard precision (SP) mode, generating a maximum of 5 ligands, 'Enhance planarity of conjugated pi groups' option was chosen while other parameters were set to default[53]. Moreover, MM-GBSA was calculated for all generated poses and compared with the MM-GBSA value of the co-crystallized ligand.

3.3.2 Library docking

Since we were aiming to perform a virtual screening for large libraries, we had to choose one target receptor to proceed with. Protein structural models can have significant errors associated with them because of differences in the interpretation of regions with high mobility or multiple conformations. These protein regions can have significantly different conformations in the crystal from those in solution [57]. The three selected PDB structures 4YFX, 4YFN and 4YFK were assessed visually and through the Maestro Protein Preparation Wizard quality assessment criteria such as Ramachandran plot outliers and side chain outliers. Ramachandran plot is a method used for the evaluation of the stereochemical quality of a protein structure. The plot demonstrates the statistical distribution of the combinations of the backbone dihedral angles ϕ (Phi) and ψ (Psi) and predicts the structural stereochemical property of the residues. The distribution of the Phi/Psi values of the protein structure allows the visualization of the energetically favoured regions, allowed regions, generously allowed regions, and disallowed regions. Only up to 2% of residues are expected to belong to the allowed region while no residue should lie in the disallowed or outlier region. The presence of many dihedral angels in the outlier region of the Ramachandran plot indicates a poor-quality protein structure. The side chain outliers apply a similar concept as the Ramachandran outliers but for the angles of the side chains instead of the backbone[58] [59]

PDB protein structures of 4YFX, 4YFN and 4YFK showed 5.6%, 2.4% and 2.1% Ramachandran plot outliers, respectively. Side chain outliers were 16.8%, 9.5% and 9.6 % for PDB protein structures 4YFX, 4YFN AND 4YFK, respectively[2]. 4YFK was chosen for further studies since it showed less outliers and less missing loops. To minimize the size of the protein, only the chains involved in the binding pocket were kept (β1238-1342 and β’8-1376). By minimizing the size of the protein, we could minimize any unnecessary errors within the crystal structure that could affect calculations later on. Also, it is more efficient to use smaller structures for computationally costly calculations such as MD simulations.

MolPort compound library *fulldb* was docked to the 4YFK crystal structure following a structure-based virtual screening protocol executed in multiple stages using Glide within the docking section in Maestro’s Virtual Screening Workflow panel. The option ‘enhance planarity of pi-conjugated groups’ was used. The docking protocol was adjusted to initially dock the libraries using the Glide SP mode, after which the best scoring 1% of the compounds were redocked using the Glide XP mode where only top 10% of those were not filtered out.

The *fulldb* output was filtered using the Maestro Ligand Filtering Panel (Schrödinger Release 2021-4: Canvas, Schrödinger, LLC, New York, NY, 2021) in which compounds with a docking score more than -9 kcal/mol were filtered out (the more negative the score, the better the estimated binding affinity).

The filtered NP-db was also docked to the 4YFK crystal structure but using Glide XP docking directly, and compounds with docking score lower than -8 kcal/mol were submitted for MM-GBSA.

MM-GBSA binding free energies were then calculated for the remaining 9,978 compounds from the *fulldb* and for the 4,544 top compounds from the NP-db. Finally, compounds of both NP-db and *fulldb* with the MM-GBSA energy values more than -50 kcal/mol were filtered out. In total, 3,510 compounds passed the filter.

3.4 3D Field-based QSAR model

To further limit the number of candidate compounds taken to the next analysis step, a QSAR model was built and applied to predict the inhibitory activities of the remaining 3,510 compounds. 55 compounds that are known or predicted to bind to the myxopyronin-binding pocket within the RNAP switch region with available IC50 values were collected from literature[35][44][2][47][48] to build a QSAR model using the Field-based QSAR panel in Maestro (Schrödinger Release 2021-4: Field-based QSAR, Schrödinger, LLC, New York, NY, 2021). The ligands were aligned with respect to the co-crystallized ligand of the PDB structure 4YFK using the Ligand Alignment panel in Maestro. The training set was set randomly to 70% of the compounds, the Partial Least Square (PLS) factors were set to maximum of 4 and variables with |t-value| < 2 were eliminated, while the other parameters were set to default. Field-based QSAR employs a Gaussian field with steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fractions to infer how these fields correlate with the ligand's biological activity[60]. The top 100 compounds with the highest predicted biological activity by the QSAR model were selected. From these 100 compounds, 5-6 compounds from each of the following criteria were chosen: compounds with the lowest (most favourable) MM-GBSA energy values, compounds with the lowest docking score values and compounds with the highest QSAR-predicted inhibitory activity. A total of 17 compounds were then subjected to MD simulations. To ensure that we can test the best candidate compounds *in vitro* in the future, these 17 compounds were checked for their availability through the MolPort search, and any unavailable compounds were omitted and then next best in the particular category was taken instead.

3.5 Molecular dynamics simulations and MM-GBSA

MD simulations are a valuable computational technique for studying biological macromolecules, understanding the dynamic behaviour of proteins at different timescales. Additionally, it allows the study of protein structure and stability under explicit solvent molecules, providing time-averaged properties and variable thermodynamic parameters such as entropies and interaction energies. The current advances in computer power, methodologies such as force fields and algorithms have permitted simulations of realistic systems and events as they evolve in real time [61]. Nowadays, it is easy to achieve timescale of μs (10^{-6} s) with small proteins while with larger biomolecular systems of 10^4 to 10^6 atoms timescale of hundreds to thousands of ns (10^{-9} s) can be achieved with the state-of-the-art computational power[62].

After the system is built, force fields are used to calculate the forces acting on every atom. Force fields are complex equations that in a simple way represent the molecular features in bonded and non-bonded terms. The bonded terms represent bond lengths and bond angles in form of springs and bond rotations, while non-bonded terms represent van der Waals interactions as Lennard–Jones potentials and electrostatic interactions as Coulomb interactions between point charges. Thus, force fields deduce the potential energy from the molecular structure in an extremely fast fashion even for large systems. Thereafter, classical Newton's law of motion calculates accelerations and velocities of atoms and updates their positions[63]. The potential energy is calculated for each simulated atom as follows[63]:

$$E_{pot}\{x_i\}$$

$$\mathrm{F_i} = -\partial \mathrm{E_{pot}} / \partial \mathrm{x_i}$$

$$\mathrm{a_i} = \mathrm{F} / \mathrm{m_i}$$

$$\mathrm{v_i}(t + dt) = \mathrm{v}(t)_i + a_i dt$$

$$\mathrm{x_i}(t + dt) = \mathrm{x}(t)_i + \mathrm{v}_i dt$$

$\mathrm{E_{pot}}$ is the potential energy for a simulated atom i at an atom coordinate x. F is the forces component, a is the acceleration, m is the mass and v is the velocity. The simulation time is t and dt is the iteration time for each spatial coordinate of N simulated atoms.

The total force F acting on an individual particle includes several parameters and can be calculated using the following equation[63]:

$$Epot(total) = \sum_{i=1}^{N\ bond} Vbond + \sum_{i=1}^{N\ angle} V\ angle + \sum_{i=1}^{N\ dihedral} V\ dihedral + \sum_{i=1}^{N\ nonbond} V\ nonbonded$$

The apo protein (RNAP without any co-crystalized ligand), the protein with a co-crystalized reference ligand of PDB structure, 4YFK and the docking complexes of the 17 chosen compounds at RNAP SQ/Myx binding site were submitted for MD simulations using Maestro (Schrödinger Release 2022-2: Desmond Molecular Dynamics System, D. E. Shaw Research, New York, NY, 2021. Maestro-Desmond Interoperability Tools, Schrödinger, New York, NY, 2021)[64]. The purpose of the simulations was to assess the stability of the compounds in the binding pocket in a dynamic state. Each simulation system was prepared using the System Builder panel. The protein-ligand complex (or the apo protein) was placed inside the orthorhombic box measuring 10 x 10 x 10 Å in all directions from the protein and the system was solvated using the TIP3P solvent model[65]. The OPLS4 force field was applied, and the system was neutralized by adding counter ions of Cl^-.

The built systems were simulated for 250 ns. The recording interval for the trajectory and the energies was set as 250 ps, thus generating approximately 1000 snapshot frames. The NPT ensemble class was chosen, and the rest of the parameters were set to default. Three random seeds were generated for three parallel MD simulations that were carried out in each case, and the average of the 3 simulations' results are reported.

3.5.1 Molecular dynamics simulation analysis

MD simulation results are collected in the form of trajectories, a series of snapshots collected over a period of the simulation time using Maestro's automated Simulation Interactions Diagram (SID) tool. The values calculated by SID were Root-Mean-Square Deviation (RMSD) and Root-Mean-Square Fluctuation (RMSF).

RMSD measures the stability of the structure during the timespan of the whole simulation by showing the average variation in dislocation of the protein backbone Cα atoms. This change in the positions of the Cα atoms along the time frames of the trajectory is reported in reference to the Cα atoms of the initial frame. When the RMSD fluctuates only minimally, the simulation has reached an equilibrium. RMSD is calculated as follows:

$$RMSD = \sqrt{\sum_{i=1}^{N_{atoms}} (r_i(t_x) - r_i(\mathrm{t}_{ref}))^2 \ / N_{atoms}}$$

N atoms is the number of atoms whose positions are being compared to the reference frame, $ri(t)$ represents the position for an atom i at time t, t_{ref} is the reference frame at time $t = 0$ and t_x is the recorded frame at a certain time. The average is taken over the atoms and gives time specific values (Schrödinger Release 2022-2: Maestro, Schrödinger, LLC, New York, NY, 2021).

Root mean square fluctuation (RMSF) is also used to analyse the highest and the least fluctuating residues of Cα atoms along a protein chain. Thus, regions such as loops, N-terminus, and C-terminus show high RMSF values. On the other hand, secondary structures of protein such as helices and beta-sheets show less fluctuation due to their rigidity. RMSF is calculated as follows:

$$RMSF_i = \sqrt{\frac{1}{T}\sum_{t=1}^{T} < (r_i'(t)) - r_i(t_{ref}))^2 >}$$

RMSF is calculated for the trajectory time T, where the reference time is t_{ref}, r_i is the position of residue i, position of atoms in residue i after superimposition on the reference frame is r', and the angle brackets (<, >) specify that the average of the square distance is taken over the selection of atoms in the residue (Schrödinger Release 2022-2: Maestro, Schrödinger, LLC, New York, NY, 2021).

Finally, the hydrogen bonds between the protein-ligand complexes over the simulation time were measured. The analysed results from all assessed compounds were compared with the apo protein and the reference complex with the co-crystallized ligand.

3.5.2 Molecular Mechanics-Generalized Born Surface Area post Molecular Dynamic simulation

MM-GBSA energies were calculated for 100 snapshots of the complexes within the last 125 ns of simulation (1 snapshot every 5 snapshots from the last 500 frames of simulation) to assess the binding affinity of the compound during the end part of the simulation (Schrödinger Release 2022-1: Prime, Schrödinger, LLC, New York, Y, 2021). Graphs were performed using GraphPad Prism version 8.4.2 for Windows.

4 Results and discussion

In this study, we designed of a virtual screening pipeline to discover novel drug candidates that inhibit RNAP via targeting the RNAP 'switch region'. The approach of virtual screening produces, inevitably, false positives and negatives. Regardless, SBVS is a successful method in finding new drug candidates because it is considered mainly an enrichment process. The calculated scores and energies in virtual screening may not be used for accurate compound selection but for shortlisting large compound libraries into a few shortlisted candidates that have a higher probability of being active[66][67].

4.1 Library preparation and filtration

Firstly, the molecular databases *fulldb* and (NP-db) were prepared with Maestro's LigPrep tool converting the compounds into low-energy 3D structures whilst generating possible ionization states and tautomers. The process almost duplicated the initial number of the compounds in *fulldb* from 7,781,175 to 15,007,020 compounds. However, due to technical issues, some compounds from the NP-db could not be processed and only 55,982 compounds were generated successfully by LigPrep from the original 119,055 compounds.

The initial techniques applied early in virtual screening aim to reduce the size of large compound libraries. Thus, an initial stage of 'pre-filtering' was applied to eliminate compounds that are unlikely to be suitable drug candidates based on a series of physicochemical descriptors[61]. Lipinski's rule of 5 (Ro5) or the 'rule of thumb' was developed to evaluate the drug-like properties and to predict the 'drugability' of a compound. In drug discovery, Ro5 is an acceptable approach to predict the ADME (''absorption, distribution, metabolism, and excretion'') performance of the drugs. The Ro5 was developed through the retrospective analysis of 2,245 compounds at the entry to Phase II of development programs. Consequently, the physicochemical properties that are common within the selected compounds were identified and from which they concluded that poor absorption or permeation is more likely when the compound has more than 5 H-bond donors, 10 H-bond acceptors, molecular weight greater than 500, or calculated logP (cLogP) greater than 5[68][50]. Hence Ro5 is important in the selection of molecules during the screening of libraries[69]. It has been verified that almost 89% of successful drugs fulfil the thresholds of the Ro5, however, it is critical to highlight that it has limitations, and it is not uncommon for successful drugs to exhibit violations to Ro5[70]. Additionally, it is noteworthy that the decrease in lipophilicity can hinder (oral) bioavailability as BCS class III hydrophilic compounds show poor epithelial penetration and thus their mucosal uptake is insufficient and consequently their use as drugs is strongly limited[71]. Moreover, the outer membrane and the peptidoglycan (murein) layer of gram-negative bacteria act as a molecular filter for hydrophilic compounds[72]. In a literature review conducted by Fernandes *et al.,* they identified the most potent antitubercular compounds described in the literature during recent years with MIC values < 7 μM. They found that most of the active compounds have cLogP values in the range of 2.5 – 6. On the other hand, the targets of the drugs with cLogP values less than 2 are located in the membrane or the cytoplasm, and their uptake involves passive or active transport[73]. In addition to drug-likeness-based filtering, filtering out compounds with certain functional groups linked with toxicity provides a practical way for reducing large databases. More than half of the failures in clinical trial phases are blamed for poor pharmacokinetics and toxicity issues[74]. Therein, we used Schrödinger's QikProp prediction tool to predict the pharmaceutically relevant drug-properties and then we used the Maestro Ligand Filtering Panel to filter out compounds that are unlikely to be successful. 9,730,802 compounds from *full-db* and 29,181 compounds from NP-db passed the filtration process successfully.

4.2 Docking

Molecular docking is a popular computational method in medicinal chemistry used for developing drug design protocols including SBVS. Molecular docking is a valuable tool for identifying novel hits through predicting the binding modes of the ligands in a receptor-binding site and their molecular interactions[75]. Docking techniques explore a comprehensive set of conformations of the ligand-receptor complex and rank them based on their stabilities using scoring functions. The relatively short time needed for docking a compound and the reasonable low usage costs allow the scrutiny of large libraries, which makes docking a suitable technique for virtual screening[61].

4.2.1 Self-docking

The ability of the docking program to reproduce a ligand pose similar to that of the co-crystallized ligand in an X-ray complex is an essential determinant of the program's effectiveness for structure-based drug design purposes[76]. Herein, self-docking was performed to assess the effectiveness of Glide in reproducing the poses of co-crystallized ligands in myxopyronin binding pocket within the 'switch region'. The co-crystallized ligands myxopyronin B, squaramide compound 14 and squaramide compound 8 complexed within *E.*

coli RNAP 'switch region' from PDB structures 4YFX, 4YFN and 4YFK, respectively, were docked each into its own respective protein binding site using the Glide standard precision (SP) mode, generating a maximum of 5 docking poses[2]. The docking results were examined visually, and the MM-GBSA binding free energy values were calculated for all the generated poses and compared with the MM-GBSA values calculated for the crystallized ligand-receptor complexes that acted as a reference. Glide managed to reproduce the pose of the co-crystallized Myx and SQs within their binding pocket. More than 1 pose out of the 5 poses generated showed very similar poses to their relevant crystallized ligand, although the poses were not perfectly overlapping, and minimal differences could be noticed (Figure 14). The MM-GBSA binding free energy showed energetically favourable interactions of all generated poses with very close values to their co-crystallized ligand reference (Table 1).

Table 1 : The Prime/MM-GBSA binding free energy values for crystallized Myx, SQ compound 14 and SQ compound 8 and their three best-scored poses from the Glide self-docking experiment. The poses are arranged in an ascending order, based on their respective docking score values, i.e., the pose ranked number 1 had the lowest (best) docking score

Ligand	**Docking score (kcal/mol)**	**MM-GBSA_dG_bind (kcal/mol)**
4YFX co-crystallized ligand	**-**	**-51.82**
Pose 1	-9.081	-52.69
Pose 2	-8.584	-50.87
Pose 3	-8.152	-56.51
4YFN co-crystallized ligand	**-**	**-60.82**
Pose 1	-10.446	-60.68
Pose 2	-10.409	-61.59
Pose 3	-10.199	-60.75
4YFK co-crystallized ligand	**-**	**-55.74**
Pose 1	-8.621	-56.13
Pose 2	-8.150	-56.99
Pose 3	-6.208	-43.59

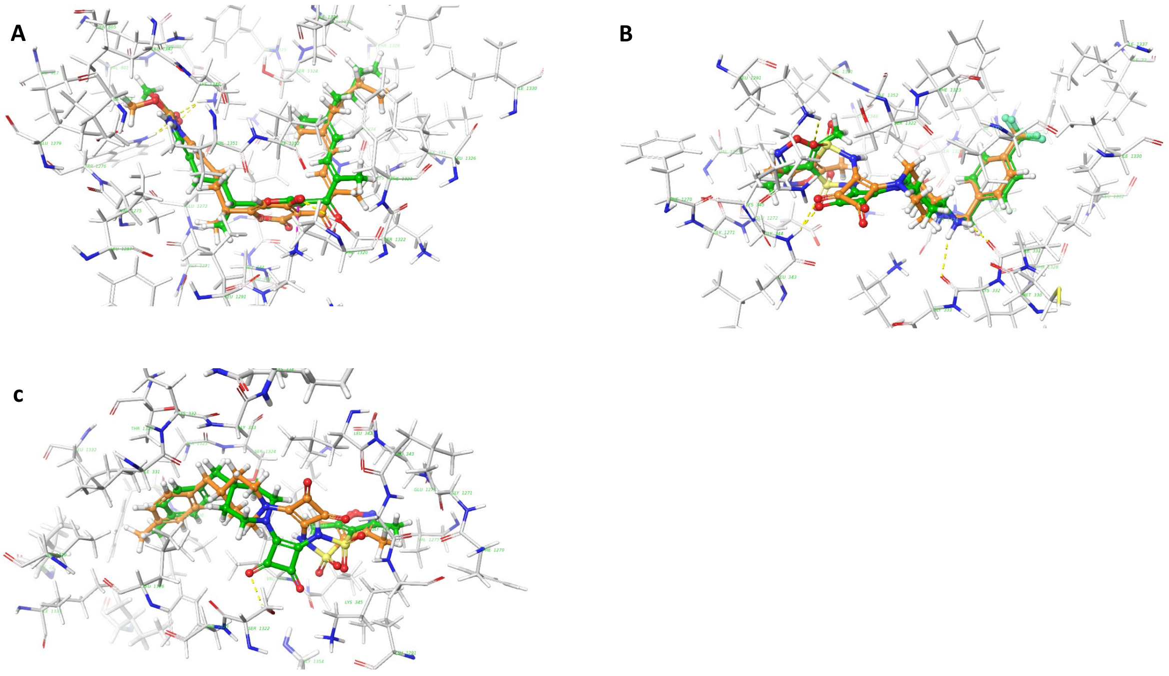


Figure 14: Redocked poses of co-crystallized ligands in *E. coli* RNAP switch region. (A) Myxopyronin B (PDB ID:4YFX); (B) squaramide compound 14 (PDB ID:4YFN); (C) squaramide compound 8 (PDB ID:4YFK) (in green) using Glide SP docking, each superimposed with their respective crystallized pose (in orange).

4.2.2 Library Docking

The main step in the virtual screening pipeline is the docking of the designated libraries. Herein different ligands are ranked with respect to their predicted binding affinity to the identified protein binding pocket[74].

At the first stage, the filtered *fulldb* was docked into the RNAP 'switch region' of the RNAP structure (PDB ID: 4YFK) using the Glide SP scoring function. The best 1% of the ranked ligands were then transferred to the next stage and docked using the more extensive scoring function, Glide XP, for an increased accuracy in scoring and ranking of the ligands. The top 10% of the ranked ligands were retained and the compounds with a docking score higher than -9 kcal/mol were filtered out. The filtered NP-db was also docked to the 4YFK crystal structure but using the Glide XP docking directly, and compounds with a docking score higher than -8 kcal/mol were filtered out. The remaining 9,978 compounds from *fulldb* and 4,544 compounds from NP-db were then submitted for Prime/MM-GBSA binding free energy estimation.

4.3 MM-GBSA

Although the docking algorithms are essential and efficient in prioritizing new hits and discriminating between binders and non-binders in a relatively rapid process, they are not particularly accurate, and the scores correlate poorly with the experimental results[55][77]. Alchemical perturbation methods like those based on extensive MD simulations, on the other hand, are in principle very accurate. However, and despite, the robustness of these methods, they are computationally intensive, which renders them inadequate for screening large number of compounds and thus, they have seldom been applied in VS protocols[55]. Molecular mechanics (MM) energies combined with the Poisson–Boltzmann (PB) or generalized Born (GB) and surface area (SA) continuum solvation (MM/PBSA and MM/GBSA) are considered as methods of intermediate performance. They can calculate free energy for only the end states, the complex and the free receptor and ligand. These methods require more computational time compared to docking, but they are less time-expensive than MD simulations and they provide more reliable results compared with scoring functions[55][77]. Therefore, we applied the Prime/MM-GBSA approach to re-score the poses with the best docking scores to enhance the accuracy of the virtual screening process by eliminating more false positive hits[77]. The compounds from both *fulldb* and NP-db with calculated MM-GBSA free energies more than -50 kcal/mol were filtered out, from which 3,510 compounds passed the filtration successfully.

4.4 3D Field-based QSAR model

Quantitative structure–activity relationship (QSAR) analysis is a ligand-based drug design method that builds mathematical models to find a statistically significant correlation between the chemical structure and a biological property using regression and classification techniques. Hence, the models built by QSAR can explain the correlation between activities and structure-based molecular descriptors. QSAR modelling has greatly evolved over the years so that nowadays it can be applied in the modelling and virtual screening of very large data sets comprising thousands of diverse compounds. QSAR currently plays a pivotal role in prioritizing compounds for synthesis and/or biological evaluation[78]. The number of descriptors that can be calculated is large and as a result, the number of combinations of the descriptors is large. Thus, QSAR modelling could be considered an optimization process with continuous refining to the metrices that relates to model performance on the training set[79]. The ability of the method to generate useful models to predict activity is well established[80][81][82][83].

By the end of the last stage, 3,510 compounds with both promising docking scores and MM-GBSA binding free energy estimates were available. However, the number of compounds was still too large to be evaluated via MD simulations or experimentally, *in vitro*. Thus, and as a further screening step, we built a 3D QSAR model in an attempt to choose only a few compounds with a higher probability of being biologically active. Fifty-five compounds known/predicted to bind to the myxopyronin-binding pocket within the switch region in literature with available IC50 values (pIC50 from 4.04 to 6.552)[35][44][2][47][48] were collected and utilized to build a QSAR model using Schrödinger's Field-based QSAR. Alignment of the compounds is a prerequisite for generating a highly predictive 3D QSAR model. Several alignments were tried, but the alignment of the compounds using the co-crystallized ligand as the template generated better models. In the Field-based QSAR method applied, the molecules are placed in a uniformly spaced grid lattice and a distant dependant Gaussian function is used to calculate their interactions with several probes: steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor. These interaction values are then correlated with the biological activity using partial least square (PLS) regression, a linear multivariate statistical method[84]. For the validation of the 3D QSAR model, the literature-derived compounds were split randomly into a training set with 70% of compounds and a test set with the remaining compounds. The model with the most favourable values for the statistical performance-evaluating metrics was chosen for the prediction of activities of the VS hits. The model had three PLS components, correlation coefficient R^2 of 0.7411 (the model can explain 74 % of the variance in the observed activity data) and standard deviation (SD) of 0.4256. Cross-validated R^2 (R^2 cv) value of 0.469 was derived from the leave-one-out (LOO) cross-validation method. The stability value of 0.762 is higher than R^2, which indicates that the model is not over-fitting but is relatively robust and not sensitive to omissions from the training set. The predictive power of the model seems to be relatively good as the Q^2 correlation coefficient of the test set predictions was 0.7901, root-mean-square error (RMSE) was 0.36 and the Pearson-r value for the correlation between the predicted and observed activity for the test set was 0.899 (Table 2) (Figure 15). The Gaussian field contributions were 49.6%, 5.7%, 17.9%, 16.8% and 9.8% for steric, electrostatic, hydrophobic, hydrogen bond donor and hydrogen bond acceptor fields, respectively (Table 2), which implies that steric contribution plays a major role in enhancing the activity followed by hydrophobic and hydrogen bond donor (HBD) contributions.

Table 2: Summary of results obtained with the Gaussian-based 3D-QSAR models

PLS statistics (PLS 3)	Description	Value
SD	standard deviation of the regression	0.4256
R^2	non-cross-validated correlation coefficient	0.7411

PLS statistics (PLS 3)	**Description**	**Value**
R^2 cv	cross-validated correlation coefficient	0.4694
Stability		0.762
Q^2	predictive correlation coefficient	0.7901
RMSE	root mean square error in the predictions of test set	0.36
Pearson - r	correlation between the test set predicted and observed activity	0.899
Gaussian field fraction		**Value (in percentage)**
Steric		49.6
Electrostatic		5.7
Hydrophobic		17.9
HBD	Hydrogen bond donor	16.8
HBA	Hydrogen bond acceptor	9.9

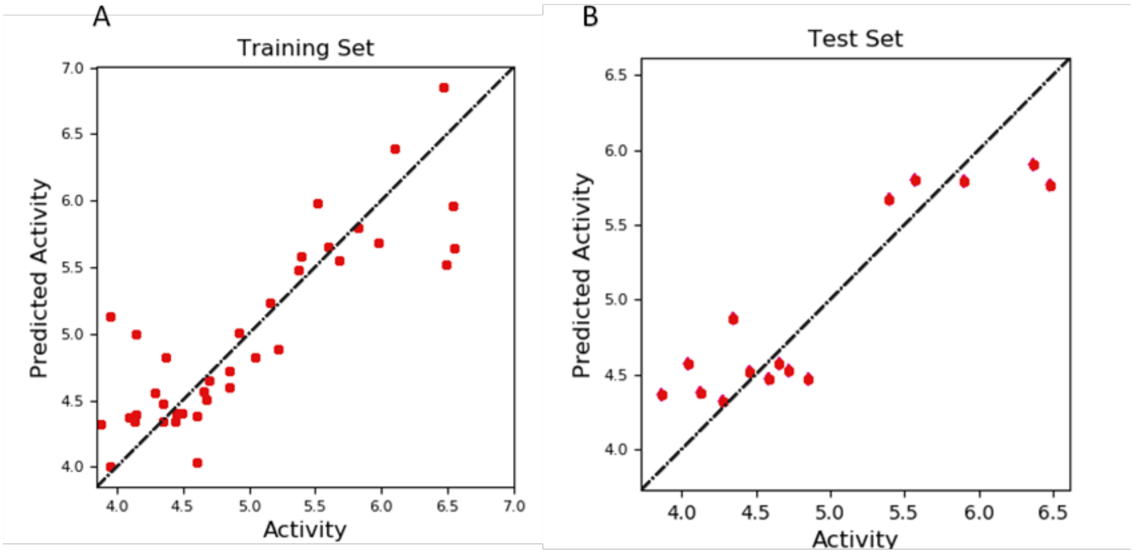


Figure 15: Predicted vs observed activities (pIC50) by the generated QSAR model for the compounds in (A) the training set, and (B) the test set.

The best 100 compounds were determined based on the predicted pIC50 values, and 17 compounds from those were chosen to be subjected to MD simulations (Figure 16) (Figure 17) (Table 3). The compounds were chosen according to one of the following criteria: 1) compounds with the lowest (most favourable) MM-GBSA energy values, 2) compounds with the lowest docking score values and 3) compounds with the highest QSAR-predicted inhibitory activity. The availability of the 17 compounds was checked through the MolPort search to ensure that we can test the best candidate compounds *in vitro* in the future and any unavailable compound was omitted and then next best in the particular category was taken instead.

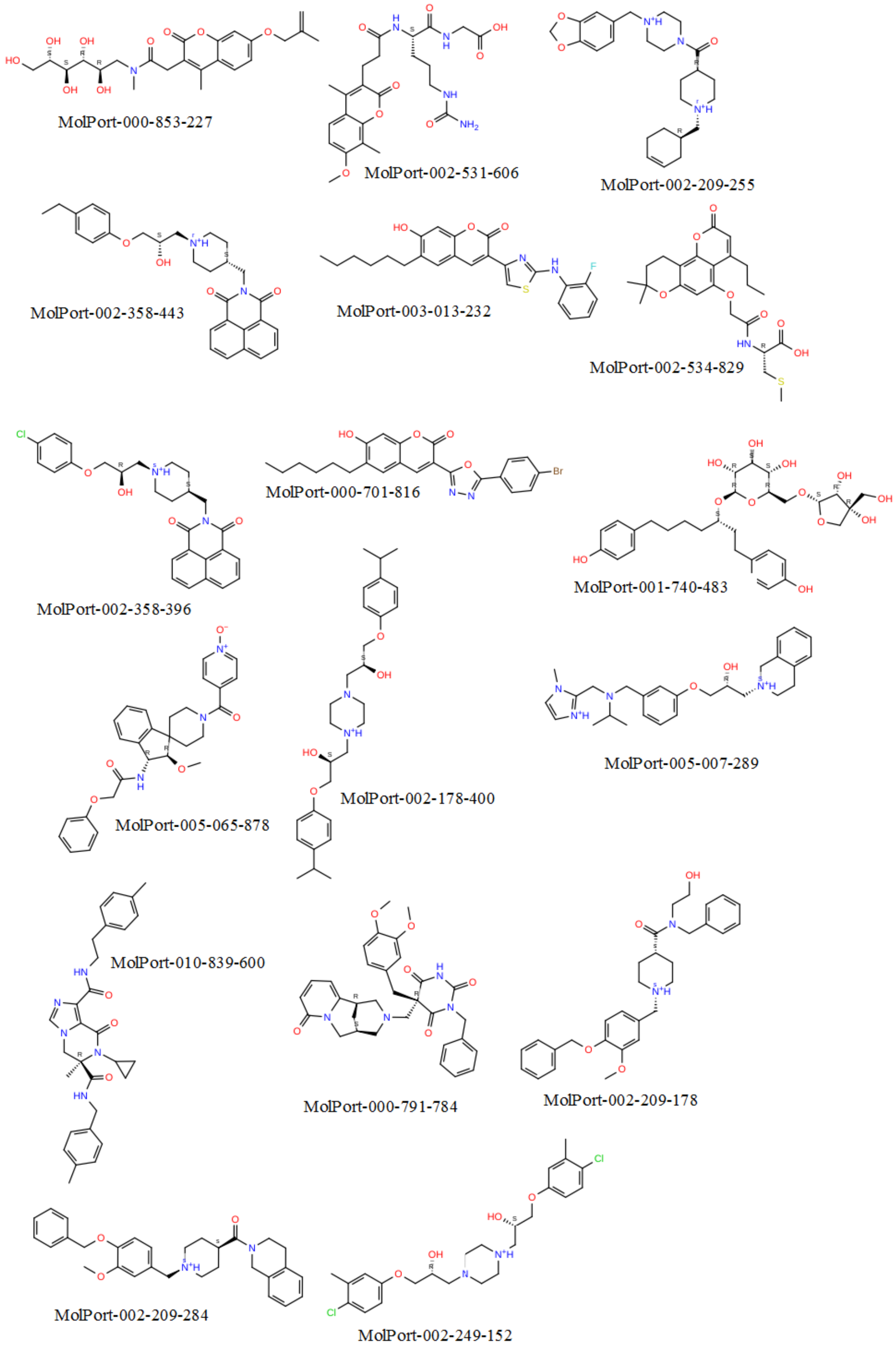


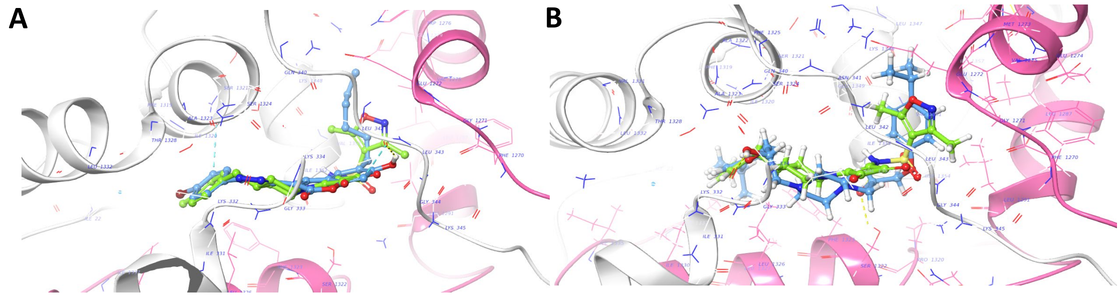
Figure 16: Structures of the 17 selected virtual hit compounds to be subjected to MD simulations with their MolPort database IDs.

Table 3: Docking score, MM-GBSA binding free energy and predicted pIC50 of the 17 candidate compounds from the MolPort database, arranged in a descending order according to their predicted pIC50

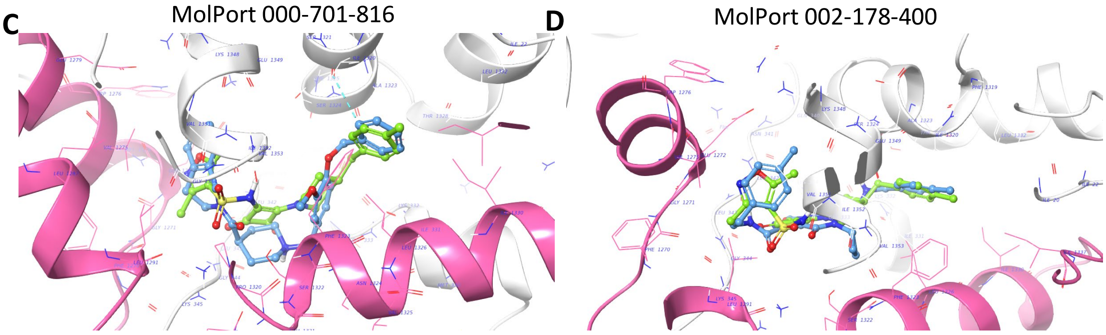
MolPort-ID	Docking Score (kcal/mol)	MM-GBSA (kcal/mol)	Predicted activity (pIC50)
Squaramide compound 8	-	-55.74	6.538[2]*[1]
MolPort-000-853-227	-8.383	-52.88	6.197
MolPort-002-531-606	-8.906	-60.69	6.158
MolPort-002-209-255	-9.93	-51.3	6.149
MolPort-002-358-443	-10.71	-59.39	6.129
MolPort-003-013-232	-9.825	-55.03	6.121
MolPort-002-534-829	-8.362	-59.13	6.121
MolPort-002-358-369	-10.521	-58.87	6.111
MolPort-000-701-816	-9.598	-55.45	6.071
MolPort-001-740-483	-15.836	-56.65	6.048
MolPort-005-065-878	-11.521	-54.29	5.952
MolPort-002-178-400	-12.821	-53.95	5.913
MolPort-005-007-289	-10.483	-65.71	5.88
MolPort-010-839-600	-11.541	-55.23	5.854
MolPort-000-791-784	-9.902	-60.05	5.816
MolPort-002-209-178	-10.877	-58.17	5.775
MolPort-002-209-284	-10.064	-61.78	5.764
MolPort-002-249-152	-10.089	-61.72	5.066

---

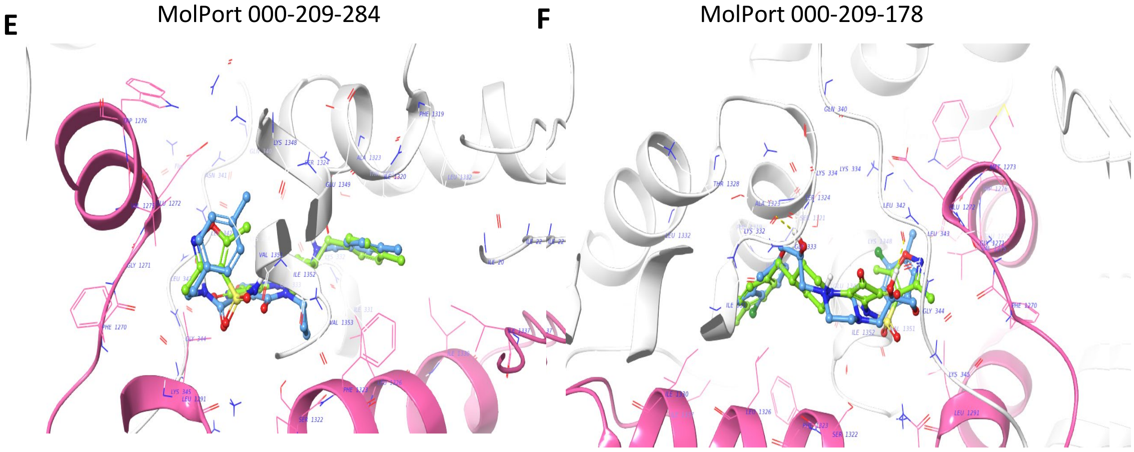
[1] *The experimental pIC50 of the co-crystallized ligand, squaramide compound 8 (PDB ID: 4YFK)



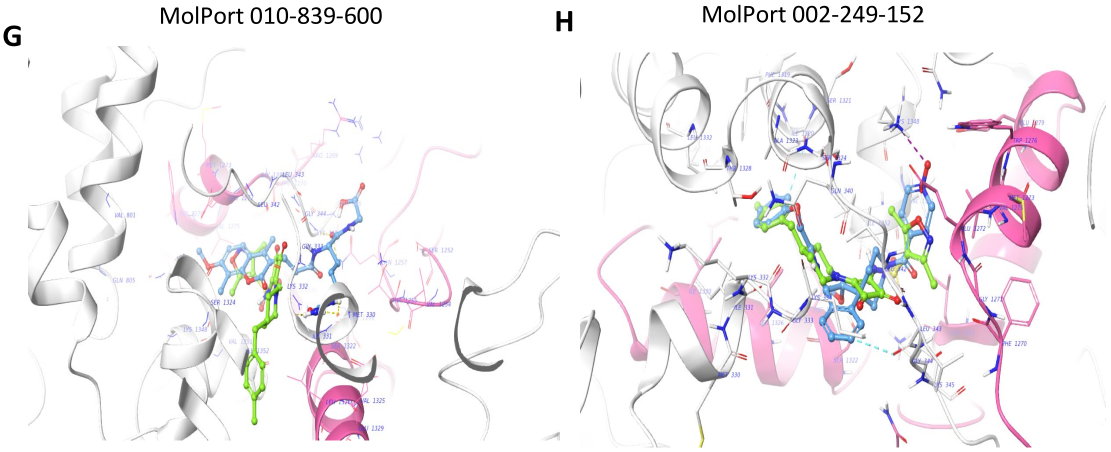
MolPort 000-701-816 MolPort 002-178-400



MolPort 000-209-284 MolPort 000-209-178



MolPort 010-839-600 MolPort 002-249-152



MolPort 002-531-606 MolPort 005-065-878

Figure 17: Docked poses of the most promising virtual screening hits (in light blue) in complex with RNAP switch region (PDB ID: 4YFK) superimposed with the co-crystallized ligand, squaramide 8 (in green). Oxygen and nitrogen in ligands and side chain residues are coloured in red and blue, respectively. Colour code of highlighted ligand-receptor main interactions: yellow – hydrogen bond, cyan – aromatic hydrogen bond, purple – halogen bond. β and β' subunits are presented in pink and white, respectively.

4.5 MD simulations

The integration of the inexpensive and fast docking protocols with the computationally expensive but accurate MD simulations can significantly improve the enrichment factor of virtual screening. In MD simulations, the ligand and protein are treated as flexible entities and the effect of explicit water molecules on the binding interaction can also be studied. Thus, MD simulations enable very accurate calculations of binding free energies. However, since MD simulations require extensive and time-consuming computational calculations, their application is limited to a few ligand–protein complexes. The strength of combining molecular docking with MD simulations lies in their complementary strengths to overcome their weaknesses. Thus, a fast docking protocol is applied as the first stage to screen large compound libraries and then only a few promising ligands are subjected to MD simulations[61][85][86]. Herein, 17 candidate compounds docked at RNAP were simulated for 250 ns, generating 1000 snapshot frames per trajectory for each ligand-protein complex. The entire trajectories were then analysed for the stability of the ligand-protein complexes. The binding affinities of the compounds throughout the second half of the simulation were estimated by calculating the Prime/MM-GBSA binding free energy using 100 frames per complex (every 5th frame from the last 500 frames) and the average MM-GBSA energy in kcal/mol (delta G, dG) was calculated for each compound. Herein, we report the MD analysis results for eight compounds that showed promising binding affinity *in silico,* comparable to the co-crystallized ligand, and thus have potential to be good RNAP inhibitors targeting the 'switch region'. These compounds were selected to be tested *in vitro.* The protein RMSD and the average MM-GBSA (dG) values of the rest of the 9 compounds that were not included within this section are reported in the appendix (Appendix, Figure A1 and Table A1).

4.5.1 Protein RMSD

RMSD is one of the techniques that gives insights into the undergoing structural changes throughout the simulation and defines the stability of the system in MD simulations. The protein conformational frames (the saved trajectory snapshots) are first aligned on the reference frame that is often the starting point (initial conformation) of the simulation. The RMSD of Cα atoms is calculated for all succeeding frames in reference to the starting point[87][88].

From the RMSD plots (Figure 18) it can be noticed that the apo protein is highly flexible throughout the whole simulation time with a sharp increase in RMSD in the second half of the simulation where the RMSD fluctuates between 6 and 7 Å. The protein RMSD of the co-crystallized ligand complex is shown to be fluctuating less than that of the apo protein, reaching a plateau at around 5.5 Å. All simulated virtual hit complexes except for those with MolPort-000-701-816 and MolPort-002-209-284 show a similar protein RMSD pattern to the co-crystallized ligand in which the protein seems to get stabilized upon ligand binding when compared with the ligand-free apo structure. MolPort-000-701-816 and MolPort-002-209-284 complexes show comparable RMSD values to the apo protein, although the fluctuations in the RMSD value are not so large as for the apo.

When RMSD is being computed, it is expected that there is an initial rapid rise in RMSD because of the thermal fluctuations. This precedes a long plateau that indicates that the system has reached a basin in the potential energy surface that corresponds with the equilibrium phase. If the steady state of RMSD is not achieved, it may indicate that the system is still equilibrating or that it is drifting away from the initial structure. Thus, as the system remains fluctuating significantly without reaching a steady state, it has not converged yet[89][90][87].

However, the methodology of RMSD calculation has been criticized several times and one of the reasons behind this criticism is that although the value of plateau should neither be zero or too large, the position of plateau for a specific simulation is unknown. The value of plateau will not only change between different macromolecules but will also change with varying the simulation conditions such as temperature and solvent[89]. Therein, performing reference simulations as a control is important for determining the plateau for a specific system under the same simulation conditions.

Although RMSD values of the order of 1-3 Å are generally acceptable, larger changes indicate that the protein is undergoing major changes during the simulation. Loops/turns and termini are very mobile sections within the protein and their increased mobility in MD simulations has been frequently reported[91][92]. Therein, and as previously stated, the RNAP 'switch region' is concluded to be a flexible region especially after the binding of Myx. Upon Myx binding, a highly conserved β′ switch 2 segment refolds, and the α-helix carboxy-terminal part comprised of ~two-helical turns unwind and refolds into a loop (Figure10)[38] Thus, the relatively high RMSD of Cα atoms observed for the apo structure and the ligand-bound structures may not necessarily mean that the systems did not converge but it could be attributed to protein flexibility. Although the systems of MolPort-000-701-816 and MolPort-002-209-284 did not seem to stabilize the protein within the 250 ns of the simulation time when compared with the apo, their average MM-GBSA energies (dG) are among the best of all the simulated hit compounds (Table 4). Additionally, the ligand RMSD of the two compounds that indicates how stable the ligand is with respect to the protein and its binding pocket, reaches plateau at a considerably lower level than that of the co-crystallized ligand (Appendix, Figure A2).

4.5.2 Protein RMSF

RMSF represents the fluctuations of each residue, and the peaks indicate areas of the protein that have the highest fluctuation during the simulation[93]. Consequently, tails (*N*- and *C*-termini) and loop regions fluctuate the most compared to the more rigid secondary structures. In Figure 19, it can be observed that residues within the loop regions of the ‘switch region’ and the N- and C-termini display higher flexibility as expected. There are no noticeable differences between the apo RMSF and the RMSF of the ligand complexes except for a highly flexible region at the end of the β' chain that is highly stabilized by the binding of ligands.

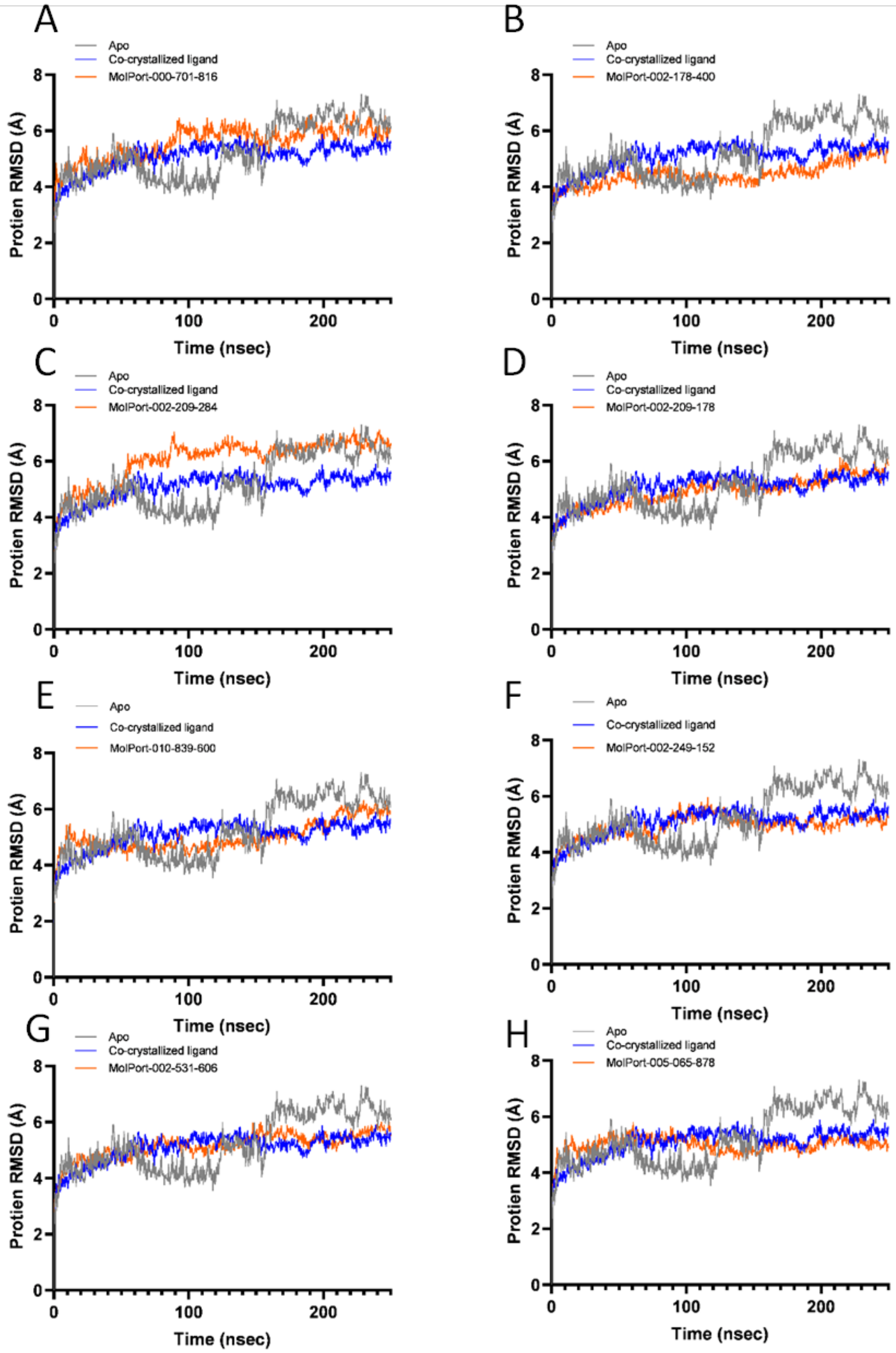


Figure 18: Protein RMSD of Cα atoms of RNAP in complex with docked virtual screening hits (in orange) during a 250-ns MD simulation compared with the apo protein (in grey) and the protein complexed with a co-crystallized ligand (in blue) (PDB ID: 4YFK). The compounds are arranged in an ascending order according to their average MM-GBSA energies (dG).

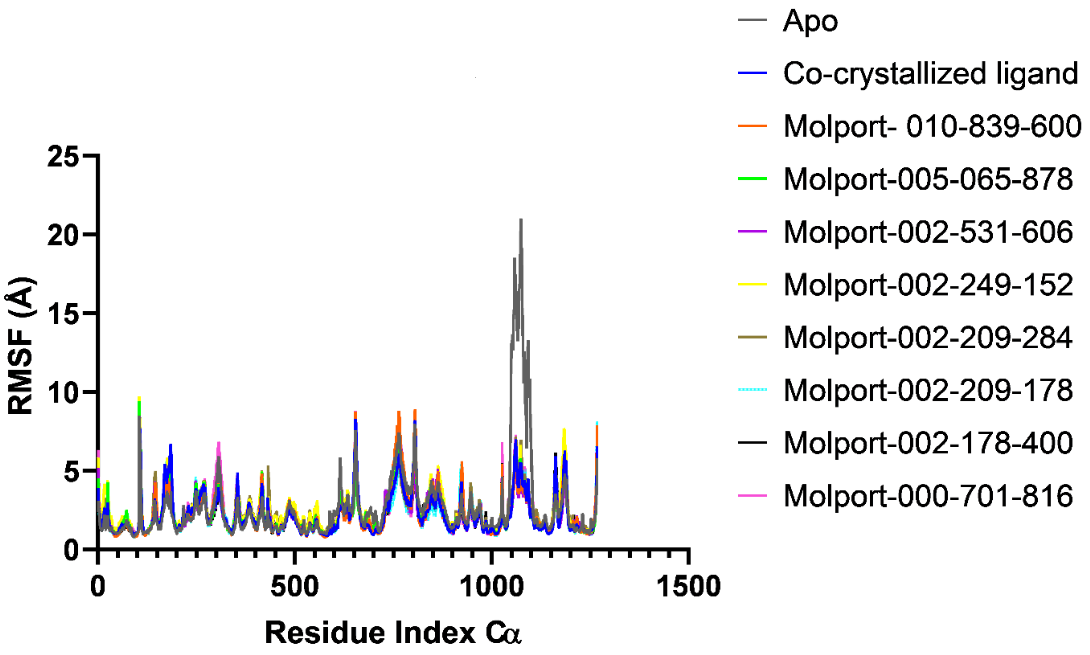


Figure 19: Cα RMSF values for each residue of the apo RNAP, the squaramide 8-RNAP crystal complex (PDB ID: 4YFK), and the ligand-RNAP complexes with the most promising virtual screening hits during a 250-ns MD simulation.

4.5.3 MM-GBSA

MM-GBSA binding free energy for complexes were estimated from snapshots within the last 125 ns of the MD trajectories simulated (Table 4) (Figure 20). Most compounds showed lower average MM-GBSA (dG) compared to their docked poses which indicates that under dynamic state, these compounds found more energy favourable poses.

Table 4: The Prime/MM-GBSA binding free energy values (dG) of the docked poses of the virtual screening hit compounds at RNAP and their average dG calculated from the last 125 ns of the 250-ns MD simulation

Compound	**MM-GBSA dG for docked poses (kcal/mol)**	**Average MM-GBSA dG after MD simulation ±SD**
Co-crystallized ligand	-55.74	-63.61 ± 5.14
MolPort-000-701-816	-55.45	-69.73 ± 7.05
MolPort-002-178-400	-53.95	-68.54 ± 6.79
MolPort-002-209-284	-61.78	-67.46 ± 4.83
MolPort-002-209-178	-58.17	-66.35 ± 6.2
MolPort-010-839-600	-55.23	-65.86 ± 7.14
MolPort-002-249-152	-61.72	-64.53 ± 5.32
MolPort-002-531-606	-60.69	-58.1 ± 8.05
MolPort-005-065-878	-54.29	-57.9 ± 4.5

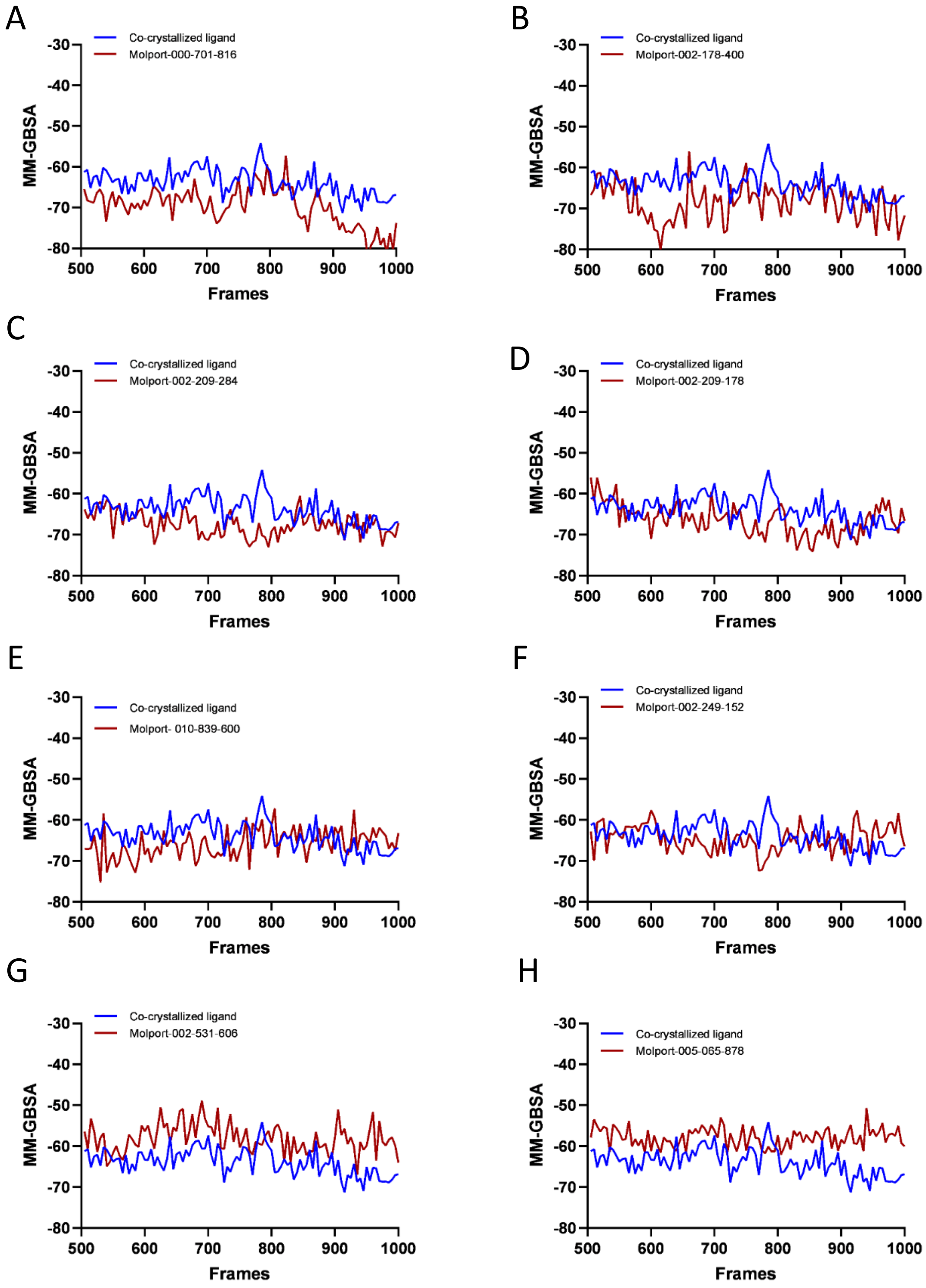
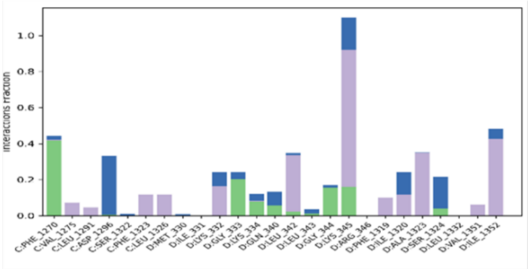


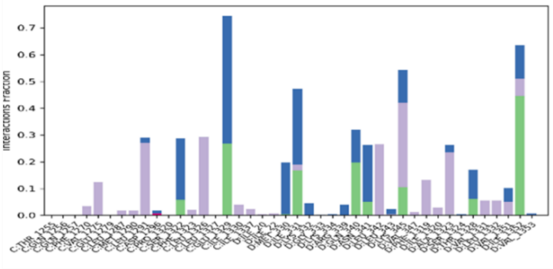
Figure 20: Prime/MM-GBSA binding free energy estimation for the most promising virtual screening hits during the second half of a 250-ns MD simulation (the snapshots of the ligand-RNAP complex structures were evaluated every 5 frames).

4.5.4 Protein-Ligand interactions

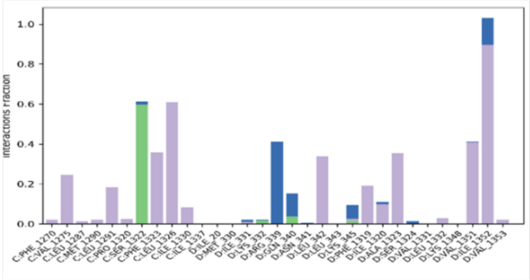
Ligand-receptor interactions were analysed over the MD simulation time (Figure 21). The co-crystallized ligand's interaction histogram shows hydrophobic interactions with several residues but long-term interactions mainly with βLeu1326, β′Ala1323 and β′Ile1352. Also, it is shown to form hydrogen bonds with βSer1322, β′Asn341 and β′Lys345. These interactions are consistent to a high degree with the SAR of the 'switch region' inhibitors discussed in section 1.4. Moreover, the virtual screening hit ligands also present some of these key interactions throughout the simulation time. For example, MolPort-002-178-400, Molport-002-209-284, Molport-010-839-600, Molport-002-249-152 and Molport-002-531-606 form a hydrogen bond with βSer1322. Furthermore, all compounds show major hydrophobic and/or hydrogen bonding interactions with the key residue β′Lys345. It can also be observed that all the simulated ligands show various hydrophobic contacts within the binding pocket, thus resembling Myx and SQs whose hydrophobic contacts have a dominant role in their binding. Specifically, all ligands seem to show major hydrophobic interactions with βLeu1326, β′Leu324, β′Phe1319, β′Ile1320, β′Ala1323 and β′Ile1352 except for Molport-002-531-606 that lacks some of these interactions. It is noteworthy that Molport-002-531-606 was the only candidate that had a higher (and non-converged) ligand RMSD than the co-crystallized ligand (Appendix, Figure A2). This suggests that it may lack some key interactions with the binding site and that renders its binding unstable and consequently, it may not be so good a candidate after all.



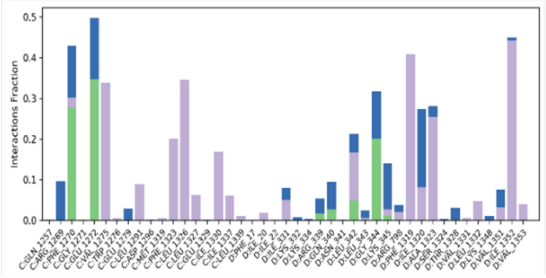
MolPort-000-701-816



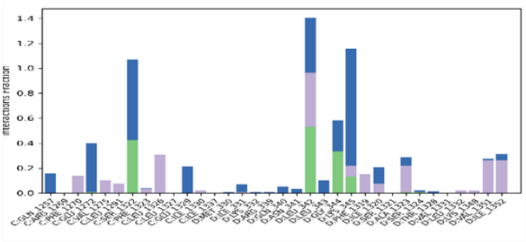
MolPort-002-178-400



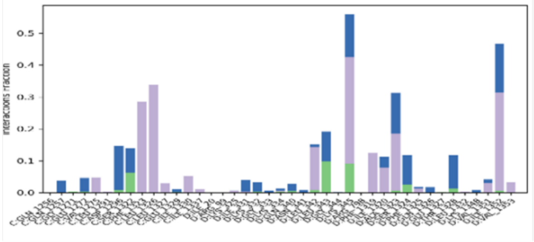
MolPort-002-209-284



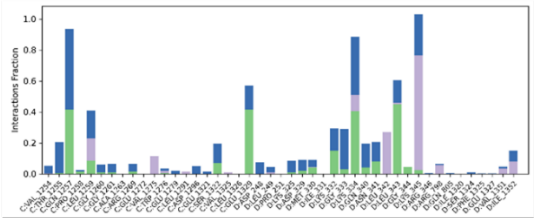
MolPort-002-209-178



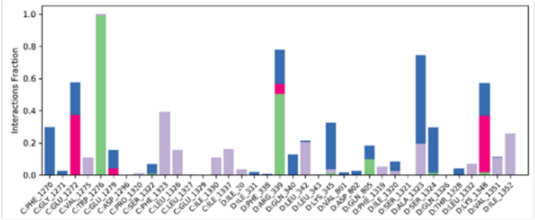
MolPort-010-839-600



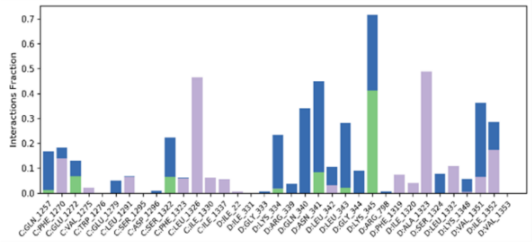
MolPort-002-249-152



MolPort-002-531-606



MolPort-005-065-878



Co-crystallized ligand

Figure 21: Ligand-receptor interaction histogram at the ‘switch region’ binding site of *E. coli* RNAP (PDB ID: 4YFK) of tested compounds arranged in ascending order according to their average MM-GBSA (dG) and the co-crystallized ligand histogram is presented at the end. The stacked bar charts are normalized over the course of a 250-ns molecular dynamics simulation trajectory; thus, the interactions fraction value denotes how long the interaction lasted with respect to the length of the total simulation. Colour code: green – hydrogen bond interactions, blue – water-mediated interactions, purple – hydrophobic interactions, red – ionic interactions. Only interactions with amino acid residues are shown.

4.6 Limitations of the study

Firstly, ‘pre-filtering’ is an essential step in virtual screening. However, if the selection criteria during the filtering steps is not carefully chosen, early exclusion of potential leads can happen[61]. The initial large size of the *fulldb* mandated stringent filtration at different stages for computational efficiency. For example, compounds with docking scores larger than -9 kcal/mol were filtered out before the MM-GBSA calculations. Compounds from the NP-db, which was subjected to a less strict filtration, showed good MM-GBSA binding free energy scores although they would have been filtered out if the same criteria of filtering as that for *fulldb* was applied. Moreover, the compounds from the NP-db, Molport-000-853-227, Molport-002-531-606 and Molport-002-534-829 with docking scores in the range of -8 to -9 kcal/mol were eligible for the MD simulation analysis and showed good post-MD binding free energies. Hence, the if the filtering criteria had been more stringent, this might have led to the loss of potential good hits.

Secondly, although the generated QSAR model showed good predictive power, QSAR models characterized by favourable values of the most common performance-evaluating metrics may not necessary perform well in virtual screening. These values characterize QSAR models performance within their applicability domain, however, during virtual screening the model is expected to perform beyond this[79]. The small number of compounds available to develop the QSAR model increases the risk of the model being over-fit despite the good statistical values. Additionally, the range of the pIC50 values of the compounds was relatively narrow. The computational success of QSAR models, generally, needs to be validated by experimental results[94]. Regardless, in our study, the QSAR model predictions were applied in the later stages of the screening on compounds with a higher probability of being drug candidates and hence even if the compounds showed good results *in vitro* it may not be enough to validate the model. Consequently, the reliability of the generated model might be doubtful. Testing a set of compounds with low predicted pIC50 despite their good docking scores and MM-GBSA binding free energy values and comparing the results with the originally selected most promising hit compounds might be a reasonable approach to confirm if there are any statistically significant differences between the two sets.

Finally, the 250 ns simulations might have been too short to confirm that the ligand-receptor complexes have reached equilibrium. Prolonging the simulation beyond 250 ns may give more insight into the stability of the binding of the proposed candidates within their binding pocket.

5 Conclusions

The continuous upsurge in drug-resistance bacteria diminishes the clinical utility of currently available antibiotics and creates a public health threat, which mandates the discovery of novel antibiotics. In this study, several computational methodologies were integrated to develop a structure-based virtual screening pipeline to find novel inhibitors that target the RNAP 'switch region'.

Firstly, the ability of the docking algorithm to produce reliable docking poses was validated through self-docking studies. Maestro's Glide was shown to effectively reproduce the crystal poses of three co-crystallized ligands in their respective sites within the *E. coli* RNAP switch region. Thereafter, the commercially available compound libraries used in the study were 'pre-filtered' and then docked with Glide to the ligand-binding site of the target protein, i.e., the squaramide compound 8 binding site in the crystal structure of *E. coli* RNAP. The best scoring ligands were submitted for Molecular Mechanics–Generalized Born Surface Area (MM-GBSA) calculations to estimate their binding free energy. Furthermore, a 3D QSAR model was developed based on the previously reported IC50 values of the compounds reported to bind to the myxopyronin-binding site to predict the biological activity of the compounds with the most favourable MM-GBSA binding free energies. The model was statistically good, however, the limited number of compounds used to develop the model and the narrow range of their pIC50 values increase the risk of the model being over-fit. Hence, further validation of the reliability of the 3D QSAR model might be needed.

Based on the results of docking, MM-GBSA binding free energies and the activities predicted by the 3D QSAR model, 17 compounds were submitted for MD simulations to further analyse the stability of the ligand-receptor complexes. The MD simulation results of the 17 most-promising ligands were then compared with the MD simulations carried out with the apo protein and with a reference complex of a co-crystallized ligand in the RNAP binding site. Finally, eight candidate compounds showing comparable RNAP binding to the reference ligand were selected for further biological testing. The protein RMSD for the studied compounds was relatively higher than generally recommended RMSD ranges, which may suggest that a 250-ns simulation time may have not been enough for the system to converge. Regardless, when compared with the apo protein and the reference complex with a co-crystallized ligand, six virtual screening hit candidates are shown to stabilize the protein and show comparable protein RMSD fluctuations to that of the reference complex. Additionally, the binding affinities estimated as average MM-GBSA (dG) values of all eight compounds were similar to the co-crystallized ligand throughout the simulation. Therefore, the high protein RMSD values could be attributed to the high flexibility of the target binding pocket.

In summary, the new promising hits discovered in this *in silico* study can be useful in the development of new antibacterial inhibitors. The potential of the hits will be further assessed *in vitro.* Besides, further research will be needed to assess their effectiveness and safety. Lead optimization can also be applied for the promising candidates from the *in vitro* study to enhance their target affinity and selectivity.

References

[1] R. J. Fair and Y. Tor, “Antibiotics and bacterial resistance in the 21st century,” *Perspectives in Medicinal Chemistry*, no. 6, pp. 25–64, 2014, doi: 10.4137/PMC.S14459.

[2] V. Molodtsov *et al.*, “X-ray Crystal Structures of *Escherichia coli* RNA Polymerase with Switch Region Binding Inhibitors Enable Rational Design of Squaramides with an Improved Fraction Unbound to Human Plasma Protein,” *Journal of Medicinal Chemistry*, vol. 58, no. 7, pp. 3156–3171, 2015, doi: 10.1021/acs.jmedchem.5b00050.

[3] P. Villain-Guillot, L. Bastide, M. Gualtieri, and J. P. Leonetti, “Progress in targeting bacterial transcription,” *Drug Discovery Today*, vol. 12, no. 5–6, pp. 200–208, 2007, doi: 10.1016/j.drudis.2007.01.005.

[4] S. Hinsberger, K. Hüsecken, M. Groh, M. Negri, J. Haupenthal, and R. W. Hartmann, “Discovery of novel bacterial RNA polymerase inhibitors: Pharmacophore-based virtual screening and hit optimization,” *Journal of Medicinal Chemistry*, vol. 56, no. 21, pp. 8332–8338, 2013, doi: 10.1021/jm400485e.

[5] J. Mukhopadhyay *et al.*, “HHMI Author Manuscript THE RNA POLYMERASE ‘ SWITCH REGION ’ IS A TARGET FOR INHIBITORS,” *October*, vol. 135, no. 2, pp. 295–307, 2008.

[6] X. Tan *et al.*, “Design, synthesis and biological evaluation of novel hybrids of N-aryl pyrrothine-base α-pyrone as bacterial RNA polymerase inhibitors,” *Bioorganic and Medicinal Chemistry Letters*, vol. 30, no. 11, p. 127146, 2020, doi: 10.1016/j.bmcl.2020.127146.

[7] K. S. Murakami, “Structural biology of bacterial RNA polymerase,” *Biomolecules*, vol. 5, no. 2, pp. 848–864, 2015, doi: 10.3390/biom5020848.

[8] A. Chakraborty *et al.*, “Opening and closing of the bacterial RNA polymerase clamp,” *Science (1979)*, vol. 337, no. 6094, pp. 591–595, 2012, doi: 10.1126/science.1218716.

[9] S. Lara-González, J. J. Birktoft, and C. L. Lawson, “Structure of the Escherichia coli RNA polymerase subunit C-terminal domain,” *Acta Crystallographica Section D: Biological Crystallography*, vol. 66, no. 7, pp. 806–812, 2010, doi: 10.1107/S0907444910018470.

[10] Benjamin Chun-Kit Tong, “乳鼠心肌提取 HHS Public Access,” *Physiol Behav*, vol. 176, no. 5, pp. 139–148, 2017, doi: 10.1128/ecosalplus.ESP-0004-2018.An.

[11] N. Zenkin and K. Severinov, “The role of RNA polymerase σ subunit in promoter-independent initiation of transcription,” *Proc Natl Acad Sci U S A*, vol. 101, no. 13, pp. 4396–4400, 2004, doi: 10.1073/pnas.0400886101.

[12] S. Borukhov and E. Nudler, “RNA polymerase: the vehicle of transcription,” *Trends in Microbiology*, vol. 16, no. 3, pp. 126–134, 2008, doi: 10.1016/j.tim.2007.12.006.

[13] D. G. Vassylyev, M. N. Vassylyeva, A. Perederina, T. H. Tahirov, and I. Artsimovitch, “Structural basis for transcription elongation by bacterial RNA polymerase,” *Nature*, vol. 448, no. 7150, pp. 157–162, 2007, doi: 10.1038/nature05932.

[14] F. W. Martinez-Rucobo and P. Cramer, “Structural basis of transcription elongation,” *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, vol. 1829, no. 1, pp. 9–19, 2013, doi: 10.1016/j.bbagrm.2012.09.002.

[15] A. C. M. Cheung and P. Cramer, “A movie of RNA polymerase II transcription,” *Cell*, vol. 149, no. 7, pp. 1431–1437, 2012, doi: 10.1016/j.cell.2012.06.006.

[16] E. Nudler, “RNA polymerase active center: The molecular engine of transcription,” *Annual Review of Biochemistry*, vol. 78, pp. 335–361, 2009, doi: 10.1146/annurev.biochem.76.052705.164655.

[17] L. Tan, S. Wiesler, D. Trzaska, H. C. Carney, and R. O. J. Weinzierl, “Bridge helix and trigger loop perturbations generate superactive RNA polymerases,” *Journal of Biology*, vol. 7, no. 10, 2008, doi: 10.1186/jbiol98.

[18] D. A. Erie and S. R. Kennedy, “Forks, pincers, and triggers: the tools for nucleotide incorporation and translocation in multi-subunit RNA polymerases,” *Current Opinion in Structural Biology*, vol. 19, no. 6, pp. 708–714, 2009, doi: 10.1016/j.sbi.2009.10.008.

[19] D. G. Vassylyev, “Elongation by RNA polymerase: a race through roadblocks,” *Current Opinion in Structural Biology*, vol. 19, no. 6, pp. 691–700, 2009, doi: 10.1016/j.sbi.2009.10.004.

[20] V. Svetlov and E. Nudler, “Basic mechanism of transcription by RNA polymerase II,” *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, vol. 1829, no. 1, pp. 20–28, 2013, doi: 10.1016/j.bbagrm.2012.08.009.

[21] H. Mosaei and J. Harbottle, “Mechanisms of antibiotics inhibiting bacterial RNA polymerase,” *Biochemical Society Transactions*, vol. 47, no. 1, pp. 339–350, 2019, doi: 10.1042/BST20180499.

[22] E. André, L. Bastide, P. Villain-Guillot, J. Latouche, J. Rouby, and J. P. Leonetti, “A multiwell assay to isolate compounds inhibiting the assembly of the prokaryotic RNA polymerase,” *Assay and Drug Development Technologies*, vol. 2, no. 6, pp. 629–635, 2004, doi: 10.1089/adt.2004.2.629.

[23] E. André *et al.*, “Novel synthetic molecules targeting the bacterial RNA polymerase assembly,” *Journal of Antimicrobial Chemotherapy*, vol. 57, no. 2, pp. 245–251, 2006, doi: 10.1093/jac/dki426.

[24] R. J. Geraghty, M. T. Aliota, and L. F. Bonnac, “Broad-spectrum antiviral strategies and nucleoside analogues,” *Viruses*, vol. 13, no. 4, 2021, doi: 10.3390/v13040667.

[25] L. Eyer, R. Nencka, E. de Clercq, K. Seley-Radtke, and D. Růžek, “Nucleoside analogs as a rich source of antiviral agents active against arthropod-borne flaviviruses,” *Antiviral Chemistry and Chemotherapy*, vol. 26, pp. 1–28, 2018, doi: 10.1177/2040206618761299.

[26] M. K. Yates and K. L. Seley-Radtke, “The evolution of antiviral nucleoside analogues: A review for chemists and non-chemists. Part II: Complex modifications to the nucleoside scaffold,” *Antiviral Research*, vol. 162, no. October 2018, pp. 5–21, 2019, doi: 10.1016/j.antiviral.2018.11.016.

[27] S. I. Maffioli *et al.*, “Antibacterial Nucleoside-Analog Inhibitor of Bacterial RNA Polymerase,” *Cell*, vol. 169, no. 7, pp. 1240-1248.e23, 2017, doi: 10.1016/j.cell.2017.05.042.

[28] H. M. Hassan, D. Degen, K. H. Jang, R. H. Ebright, and W. Fenical, “Salinamide F, new depsipeptide antibiotic and inhibitor of bacterial RNA polymerase from a marine-derived *Streptomyces* sp.,” *Journal of Antibiotics*, vol. 68, no. 3, pp. 206–209, 2015, doi: 10.1038/ja.2014.122.

[29] D. Degen *et al.*, “Transcription inhibition by the depsipeptide antibiotic salinamide A,” *Elife*, vol. 2014, no. 3, pp. 1–29, 2014, doi: 10.7554/eLife.02451.

[30] D. Temiakov *et al.*, “Structural basis of transcription inhibition by antibiotic streptolydigin,” *Molecular Cell*, vol. 19, no. 5, pp. 655–666, 2005, doi: 10.1016/j.molcel.2005.07.020.

[31] B. Bae, D. Nayak, A. Ray, A. Mustaev, R. Landick, and S. A. Darst, “CBR antimicrobials inhibit RNA polymerase via at least two bridge-helix cap-mediated effects on nucleotide addition,” *Proc Natl Acad Sci U S A*, vol. 112, no. 31, pp. E4178–E4187, 2015, doi: 10.1073/pnas.1502368112.

[32] W. Lin *et al.*, “Structural Basis of Mycobacterium tuberculosis Transcription and Transcription Inhibition,” *Molecular Cell*, vol. 66, no. 2, pp. 169-179.e8, 2017, doi: 10.1016/j.molcel.2017.03.001.

[33] K. Adelman *et al.*, “Molecular mechanism of transcription inhibition by peptide antibiotic Microcin J25,” *Molecular Cell*, vol. 14, no. 6, pp. 753–762, 2004, doi: 10.1016/j.molcel.2004.05.017.

[34] J. Yuzenkova *et al.*, “Mutations of bacterial RNA polymerase leading to resistance to microcin J25,” *Journal of Biological Chemistry*, vol. 277, no. 52, pp. 50867–50875, 2002, doi: 10.1074/jbc.M209425200.

[35] J. Mukhopadhyay *et al.*, “The RNA Polymerase ‘Switch Region’ Is a Target for Inhibitors,” *Cell*, vol. 135, no. 2, pp. 295–307, 2008, doi: 10.1016/j.cell.2008.09.033.

[36] A. Srivastava *et al.*, “New target for inhibition of bacterial RNA polymerase: ‘switch region,’” *Current Opinion in Microbiology*, vol. 14, no. 5, pp. 532–543, 2011, doi: 10.1016/j.mib.2011.07.030.

[37] W. Lin *et al.*, “Structural Basis of Transcription Inhibition by Fidaxomicin (Lipiarmycin A3),” *Molecular Cell*, vol. 70, no. 1, pp. 60-71.e15, 2018, doi: 10.1016/j.molcel.2018.02.026.

[38] G. A. Belogurov *et al.*, “Transcription inactivation through local refolding of the RNA polymerase structure,” *Nature*, vol. 457, no. 7227, pp. 332–335, 2009, doi: 10.1038/nature07510.

[39] A. Jain, “Computer aided drug design,” *Journal of Physics: Conference Series*, vol. 884, no. 1, pp. 504–509, 2017, doi: 10.1088/1742-6596/884/1/012072.

[40] M. Xiang, Y. Cao, W. Fan, L. Chen, and Y. Mo, “Computer-aided drug design: lead discovery and optimization.,” *Combinatorial chemistry \& high throughput screening*, vol. 15 4, pp. 328–337, 2012.

[41] A. J. Hopfinger, “Computer-assisted drug design,” *J Med Chem*, vol. 28, no. 9, pp. 1133–1139, 1985, doi: 10.1021/jm00147a001.

[42] A. C. Anderson, “The process of structure-based drug design.,” *Chem Biol*, vol. 10, no. 9, pp. 787–797, Sep. 2003, doi: 10.1016/j.chembiol.2003.09.002.

[43] C. McInnes, “Virtual screening strategies in drug discovery,” *Current Opinion in Chemical Biology*, vol. 11, no. 5, pp. 494–502, 2007, doi: 10.1016/j.cbpa.2007.08.033.

[44] T. Doundoulakis *et al.*, “Myxopyronin B analogs as inhibitors of RNA polymerase, synthesis and biological evaluation,” *Bioorganic and Medicinal Chemistry Letters*, vol. 14, no. 22, pp. 5667–5672, 2004, doi: 10.1016/j.bmcl.2004.08.045.

[45] R. Lira *et al.*, “Syntheses of novel myxopyronin B analogs as potential inhibitors of bacterial RNA polymerase,” *Bioorganic and Medicinal Chemistry Letters*, vol. 17, no. 24, pp. 6797–6800, 2007, doi: 10.1016/j.bmcl.2007.10.017.

[46] E. T. Buurman *et al.*, “Novel rapidly diversifiable antimicrobial RNA polymerase switch region inhibitors with confirmed mode of action in Haemophilus influenzae,” *Journal of Bacteriology*, vol. 194, no. 20, pp. 5504–5512, 2012, doi: 10.1128/JB.01103-12.

[47] J. H. Sahner, M. Groh, M. Negri, J. Haupenthal, and R. W. Hartmann, “Novel small molecule inhibitors targeting the ‘switch region’ of bacterial RNAP: Structure-based optimization of a virtual screening hit,” *European Journal of Medicinal Chemistry*, vol. 65, pp. 223–231, 2013, doi: 10.1016/j.ejmech.2013.04.060.

[48] F. Yakushiji *et al.*, “Novel hybrid-type antimicrobial agents targeting the switch region of bacterial RNA polymerase,” *ACS Medicinal Chemistry Letters*, vol. 4, no. 2, pp. 220–224, 2013, doi: 10.1021/ml300350p.

[49] W. L. Jorgensen and J. Tirado-Rives, “The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin.,” *J Am Chem Soc*, vol. 110, no. 6, pp. 1657–1666, Mar. 1988, doi: 10.1021/ja00214a001.

[50] C. A. Lipinski, F. Lombardo, B. W. Dominy, and P. J. Feeney, “Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings1PII of original article: S0169-409X(96)00423-1. The article was originally published in Advanced Drug Delivery Reviews 23 (1997) 3,” *Advanced Drug Delivery Reviews*, vol. 46, no. 1, pp. 3–26, 2001, doi: https://doi.org/10.1016/S0169-409X(00)00129-0.

[51] H. M. Berman *et al.*, “The Protein Data Bank,” *Nucleic Acids Research*, vol. 28, no. 1, pp. 235–242, 2000, doi: 10.1093/nar/28.1.235.

[52] G. Madhavi Sastry, M. Adzhigirey, T. Day, R. Annabhimoju, and W. Sherman, “Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments,” *Journal of Computer-Aided Molecular Design*, vol. 27, no. 3, pp. 221–234, 2013, doi: 10.1007/s10822-013-9644-8.

[53] R. A. Friesner *et al.*, “Glide: A New Approach for Rapid, Accurate Docking and Scoring. 1. Method and Assessment of Docking Accuracy,” *Journal of Medicinal Chemistry*, vol. 47, no. 7, pp. 1739–1749, Mar. 2004, doi: 10.1021/jm0306430.

[54] R. A. Friesner *et al.*, “Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein−Ligand Complexes,” *Journal of Medicinal Chemistry*, vol. 49, no. 21, pp. 6177–6196, Oct. 2006, doi: 10.1021/jm051256o.

[55] S. Genheden and U. Ryde, “The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities,” *Expert Opin Drug Discov*, vol. 10, no. 5, pp. 449–461, May 2015, doi: 10.1517/17460441.2015.1032936.

[56] E. Wang *et al.*, “End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design,” *Chemical Reviews*, vol. 119, no. 16, pp. 9478–9508, 2019, doi: 10.1021/acs.chemrev.9b00055.

[57] K. R. Acharya and M. D. Lloyd, “The advantages and limitations of protein crystal structures,” *Trends in Pharmacological Sciences*, vol. 26, no. 1, pp. 10–14, 2005, doi: https://doi.org/10.1016/j.tips.2004.10.011.

[58] R. A. Laskowski and G. J. Swaminathan, “3.24 - Problems of Protein Three-Dimensional Structures,” J. B. Taylor and D. J. B. T.-C. M. C. I. I. Triggle, Eds. Oxford: Elsevier, 2007, pp. 531–550. doi: https://doi.org/10.1016/B0-08-045044-X/00097-3.

[59] M. Wiltgen, “Algorithms for Structure Comparison and Analysis: Homology Modelling of Proteins,” S. Ranganathan, M. Gribskov, K. Nakai, and C. B. T.-E. of B. and C. B. Schönbach, Eds. Oxford: Academic Press, 2019, pp. 38–61. doi: https://doi.org/10.1016/B978-0-12-809633-8.20484-6.

[60] D. Cappel, S. L. Dixon, W. Sherman, and J. Duan, “Exploring conformational search protocols for ligand-based virtual screening and 3-D QSAR modeling,” *Journal of Computer-Aided Molecular Design*, vol. 29, no. 2, pp. 165–182, 2015, doi: 10.1007/s10822-014-9813-4.

[61] H. Alonso, A. A. Bliznyuk, and J. E. Gready, "Combining docking and molecular dynamic simulations in drug design," *Medicinal Research Reviews*, vol. 26, no. 5, pp. 531–568, 2006, doi: 10.1002/med.20067.

[62] O. M. H. Salo-Ahen *et al.*, "Molecular Dynamics Simulations in Drug Discovery and Pharmaceutical Development," *Processes* , vol. 9, no. 1. 2021. doi: 10.3390/pr9010071.

[63] A. Hospital, J. R. Goñi, M. Orozco, and J. L. Gelpí, "Molecular dynamics simulations: Advances and applications," *Advances and Applications in Bioinformatics and Chemistry*, vol. 8, no. 1, pp. 37–47, 2015, doi: 10.2147/AABC.S70333.

[64] K. J. Bowers *et al.*, "Scalable algorithms for molecular dynamics simulations on commodity clusters," *Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC'06*, no. May 2014, 2006, doi: 10.1145/1188455.1188544.

[65] P. Mark and L. Nilsson, "Structure and Dynamics of the TIP3P, SPC, and SPC/E Water Models at 298 K," *The Journal of Physical Chemistry A*, vol. 105, no. 43, pp. 9954–9960, Nov. 2001, doi: 10.1021/jp003020w.

[66] D. B. Kitchen, H. Decornez, J. R. Furr, and J. Bajorath, "Docking and scoring in virtual screening for drug discovery: Methods and applications," *Nature Reviews Drug Discovery*, vol. 3, no. 11, pp. 935–949, 2004, doi: 10.1038/nrd1549.

[67] T. N. Doman *et al.*, "Molecular docking and high-throughput screening for novel inhibitors of protein tyrosine phosphatase-1B," *Journal of Medicinal Chemistry*, vol. 45, no. 11, pp. 2213–2221, 2002, doi: 10.1021/jm010548w.

[68] T. K. Karami, S. Hailu, S. Feng, R. Graham, and H. J. Gukasyan, "Eyes on Lipinski's Rule of Five: A New 'Rule of Thumb' for Physicochemical Design Space of Ophthalmic Drugs," *Journal of Ocular Pharmacology and Therapeutics*, vol. 38, no. 1, pp. 43–55, 2022, doi: 10.1089/jop.2021.0069.

[69] L. Ioakimidis, L. Thoukydidis, A. Mirza, S. Naeem, and J. Reynisson, "Benchmarking the reliability of QikProp. correlation between experimental and predicted values," *QSAR and Combinatorial Science*, vol. 27, no. 4, pp. 445–456, 2008, doi: 10.1002/qsar.200730051.

[70] B. G. Giménez, M. S. Santos, M. Ferrarini, and J. P. Dos Santos Fernandes, "Evaluation of blockbuster drugs under the rule-of-five," *Pharmazie*, vol. 65, no. 2, pp. 148–152, 2010, doi: 10.1691/ph.2010.9733.

[71] W. G. Dai, L. C. Dong, and Y. Song, "Enhanced bioavailability of poorly absorbed hydrophilic compounds through drug complex/in situ gelling formulation," *International Journal of Pharmaceutics*, vol. 457, no. 1, pp. 63–70, 2013, doi: 10.1016/j.ijpharm.2013.07.066.

[72] R. BENZ and K. BAUER, "Permeation of hydrophilic molecules through the outer membrane of gram-negativ bacteria: Review of becterial porins," *European Journal of Biochemistry*, vol. 176, no. 1, pp. 1–19, 1988, doi: 10.1111/j.1432-1033.1988.tb14245.x.

[73] G. dos Santos Fernandes, D. Jornada, Paula, C. Souza, C. Chin, F. Pavan, and J. dos Santos, "Current Advances in Antitubercular Drug Discovery: Potent Prototypes and New Targets," *Current Medicinal Chemistry*, vol. 22, no. 27, pp. 3133–3161, 2015, doi: 10.2174/0929867322666150818103836.

[74] V. Vyas, A. Jain, A. Jain, and A. Gupta, "Virtual screening: A fast tool for drug design," *Scientia Pharmaceutica*, vol. 76, no. 3, pp. 333–360, 2008, doi: 10.3797/scipharm.0803-03.

[75] D. Ramírez and J. Caballero, "Is It Reliable to Take the Molecular Docking Top Scoring Position as the Best Solution without Considering Available Structural Data?," *Molecules*, vol. 23, no. 5, pp. 1–17, 2018, doi: 10.3390/molecules23051038.

[76] J. B. Cross *et al.*, "Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy," *Journal of Chemical Information and Modeling*, vol. 49, no. 6, pp. 1455–1474, Jun. 2009, doi: 10.1021/ci900056c.

[77] H. Sahakyan, "Improving virtual screening results with MM/GBSA and MM/PBSA rescoring," *Journal of Computer-Aided Molecular Design*, vol. 35, no. 6, pp. 731–736, 2021, doi: 10.1007/s10822-021-00389-3.

[78] B. J. Neves, R. C. Braga, C. C. Melo-Filho, J. T. Moreira-Filho, E. N. Muratov, and C. H. Andrade, "QSAR-based virtual screening: Advances and applications in drug discovery," *Frontiers in Pharmacology*, vol. 9, no. NOV, pp. 1–7, 2018, doi: 10.3389/fphar.2018.01275.

[79] J. Spiegel and H. Senderowitz, "Evaluation of qsar equations for virtual screening," *International Journal of Molecular Sciences*, vol. 21, no. 21, pp. 1–20, 2020, doi: 10.3390/ijms21217828.

[80] J. Singh *et al.*, “Successful shape-Based virtual screening: The discovery of a potent inhibitor of the type I TGF$β$ receptor kinase (T$β$RI),” *Bioorganic \& Medicinal Chemistry Letters*, vol. 13, pp. 4355–4359, 2003.

[81] J. Kirchmair *et al.*, “Fast and Efficient in Silico 3D Screening: Toward Maximum Computational Efficiency of Pharmacophore-Based and Shape-Based Approaches,” *J Chem Inf Model*, vol. 47 6, pp. 2182–2196, 2007.

[82] G. M. Sastry, V. S. S. Inakollu, and W. Sherman, “Boosting Virtual Screening Enrichments with Data Fusion: Coalescing Hits from Two-Dimensional Fingerprints, Shape, and Docking,” *J Chem Inf Model*, vol. 53 7, pp. 1531–1542, 2013.

[83] J. S. Mason, I. Morize, P. R. Menard, D. L. Cheney, C. Hulme, and R. F. Labaudiniere, “New 4-Point Pharmacophore Method for Molecular Similarity and Diversity Applications: Overview of the Method and Applications, Including a Novel Approach to the Design of Combinatorial Libraries Containing Privileged Substructures,” *Journal of Medicinal Chemistry*, vol. 42, no. 17, pp. 3251–3264, Aug. 1999, doi: 10.1021/jm9806998.

[84] G. Klebe, U. Abraham, and T. Mietzner, “Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity.,” *J Med Chem*, vol. 37, no. 24, pp. 4130–4146, Nov. 1994, doi: 10.1021/jm00050a010.

[85] E. Aghaee, J. B. Ghasemi, F. Manouchehri, and S. Balalaie, “Combined docking, molecular dynamics simulations and spectroscopic studies for the rational design of a dipeptide ligand for affinity chromatography separation of human serum albumin,” *Journal of Molecular Modeling*, vol. 20, no. 10, 2014, doi: 10.1007/s00894-014-2446-7.

[86] S. Rajasekhar, R. Karuppasamy, and K. Chanda, “Exploration of potential inhibitors for tuberculosis via structure-based drug design, molecular docking, and molecular dynamics simulation studies,” *Journal of Computational Chemistry*, vol. 42, no. 24, pp. 1736–1749, 2021, doi: 10.1002/jcc.26712.

[87] B. Knapp, S. Frantal, M. Cibena, W. Schreiner, and P. Bauer, “Is an intuitive convergence definition of molecular dynamics simulations solely based on the root mean square deviation possible?,” *Journal of Computational Biology*, vol. 18, no. 8, pp. 997–1005, 2011, doi: 10.1089/cmb.2010.0237.

[88] R. Senthilkumar *et al.*, “Plasma Protein Binding of Anisomelic Acid: Spectroscopy and Molecular Dynamic Simulations,” *Journal of Chemical Information and Modeling*, vol. 56, no. 12, pp. 2401–2412, 2016, doi: 10.1021/acs.jcim.6b00445.

[89] A. Grossfield and D. M. Zuckerman, “Quantifying uncertainty and sampling quality in biomolecular simulations.,” *Annu Rep Comput Chem*, vol. 5, pp. 23–48, Jan. 2009, doi: 10.1016/S1574-1400(09)00502-7.

[90] E. B. Walton and K. J. VanVliet, “Equilibration of experimentally determined protein structures for molecular dynamics simulation,” *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, vol. 74, no. 6, pp. 1–8, 2006, doi: 10.1103/PhysRevE.74.061901.

[91] R. J. Law *et al.*, “Membrane protein structure quality in molecular dynamics simulation,” *Journal of Molecular Graphics and Modelling*, vol. 24, no. 2, pp. 157–165, 2005, doi: 10.1016/j.jmgm.2005.05.006.

[92] Y. Zhao, C. Zeng, and M. A. Massiah, “Molecular dynamics simulation reveals insights into the mechanism of unfolding by the A130T/V mutations within the MID1 zinc-binding Bbox1 domain,” *PLoS ONE*, vol. 10, no. 4, pp. 1–11, 2015, doi: 10.1371/journal.pone.0124377.

[93] H. S. Lee, Y. Qi, and W. Im, “Effects of N-glycosylation on protein conformation and dynamics: Protein Data Bank analysis and molecular dynamics simulation study,” *Scientific Reports*, vol. 5, pp. 1–7, 2015, doi: 10.1038/srep08926.

[94] A. Tropsha, “Best practices for QSAR model development, validation, and exploitation,” *Molecular Informatics*, vol. 29, no. 6–7, pp. 476–488, 2010, doi: 10.1002/minf.201000061.

APPENDIX

Protein RMSD

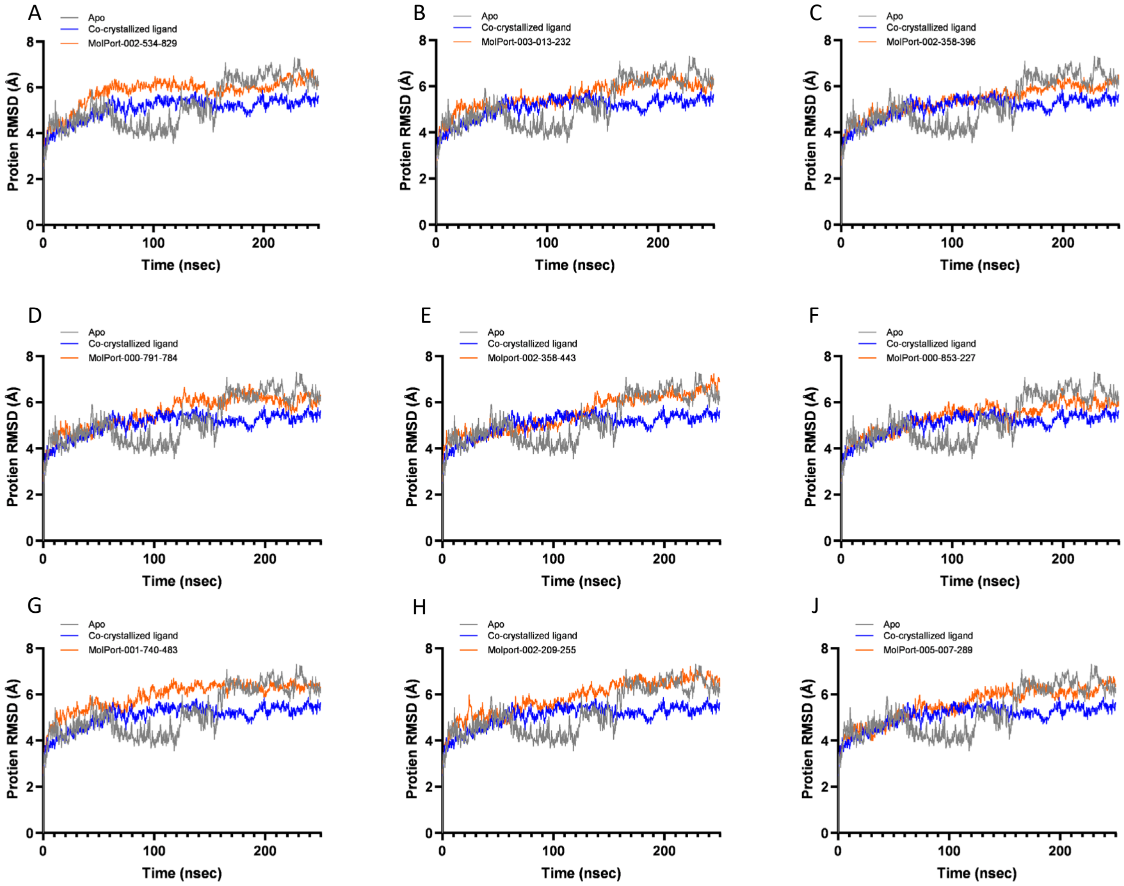


Figure A1: Protein RMSD of Cα atoms of RNAP in complex with docked virtual screening hit compounds (in orange) during a 250-ns MD simulation compared with the apo protein (in grey) and the protein complexed with a co-crystallized ligand (in blue) (PDB ID: 4YFK). The compounds are arranged in an ascending order according to their average MM-GBSA energies (dG).

MM-GBSA

Table A1: The average MM-GBSA energies (dG) calculated from the last 125 ns of the 250-ns MD simulation of the virtual screening hit compounds docked at RNAP

MolPort-ID	**Average MM-GBSA (dG)** ***after MD simulation ±SD* (kcal/mol)**
MolPort-002-534-829	-63.89 ± 6.92
MolPort-003-013-232	-63.7 ± 4.89
MolPort-002-358-369	-63.58 ± 6.73
MolPort-000-791-784	-60.42 ± 5.23
MolPort-002-358-443	-56.5 ± 7.5
MolPort-000-853-227	-56.08 ± 6.81
MolPort-001-740-483	-52.21 ± 8.4
MolPort-002-209-255	-51.15 ± 5.04
MolPort-005-007-289	-46.79 ± 7.59

Ligand RMSD

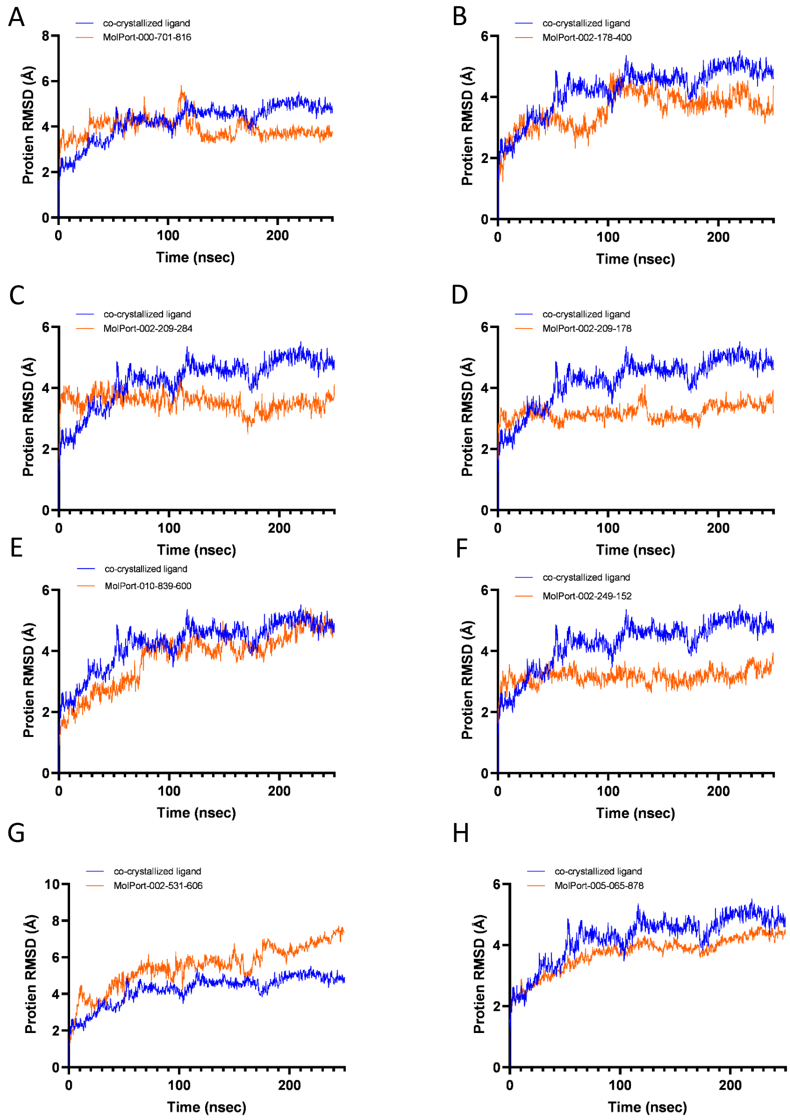


Figure A2: Ligand RMSD during a 250-ns MD simulation of the eight most promising virtual screening hits presented in the thesis arranged in an ascending order according to their average MM-GBSA (dG). The reference co-crystallized ligand RMSD is in blue.