Università degli Studi di Ferrara

DOTTORATO DI RICERCA IN
SCIENZE CHIMICHE

CICLO XXXI

COORDINATORE Prof. Carlo Alberto Bignozzi

In silico protein modelling applied to the identification of new therapeutic agents

Settore Scientifico Disciplinare CHIM/08

Dottoranda/o
Dott. Maccesi Martina

Tutore
Prof. Salvadori Severo
Prof.ssa Silvia Rivara

_____ _____
(firma) *(firma)*

Anni 2015/2018

To my beloved
grandmother

Summary

Acknowledgements

Firstly, I would like to express my sincere gratitude to my advisor, Prof. Silvia Rivara of University of Parma. During these three years she taught me a lot about computational chemistry and she helped me advancing with my research. Thanks to her I've learnt to be more precise and more critic.
I would like to acknowledge Prof. Mor and all the members of his research group. He provided me an opportunity to join his team first as master student and then as PhD student, and he gave me access to the laboratory and the research facilities. Without his precious support it would not have been possible to conduct this research. I also would like to thank him for his precious advices and suggestions.
A huge thank also to Prof. Alessio Lodola for allowing me to call him just "Ale" and for the relaxing and nice time spent in the lab.
My sincere thanks also to my two supervisors of University of Ferrara, in particular Prof. Daniele Simoni for allowing me to take part to conferences and Prof. Severo Salvadori.
I'm particularly grateful to Prof. Abagyan for giving me the opportunity to spend three amazing months in the sunny San Diego. I wouldn't forget the nice time spent in his lab and the constructive Tuesday meetings. The experience at UCSD couldn't have been so perfect without him. I would also like to thank Prof. Conor Caffrey of UCSD for gave me the opportunity to work on one of his projects and for the stimulating discussions.
Special thanks go to my partners in crime during these three years, "le Shampiste" Laura and Dona. I want to thank them for their support and for the happy time spent together: my tough days were definitely better with them.
Last but not least, I would like to thank my family for supporting me spiritually throughout these years and my life in general.

This thesis is dedicated to my grandmother, who passed away last October. She has been a truly inspiration for me. I will never forget her kindness and generosity but most of all her strength. She has been my rock for 28 years of my life and I will keep her with me forever.

1. Introduction

The process of drug discovery and development is a long and exhaustive journey that starts usually from target identification or phenotypic screening of molecule libraries ending in drug approval for marketing. Analyses across different therapeutic areas indicate that the development of a new medicine takes over 12 years and the cost of developing a New Molecular Entity (NME) is estimated to be over a billion of dollars[1].

Drug discovery processes have considerably changed through the years and computational approaches are more and more applied in this field. As a matter of fact, the use of computers and computational techniques today permeates all aspects of drug discovery and the application of computational tools should bring the advantage to lead to new drug candidates more quickly and at lower cost compared to other competitors in both academia or industry[2]. Moreover, in the past decade, the rapid development of supercomputers and better algorithms allowed to perform intensive calculations in a reasonable time, allowing computationally driven drug design to further increase its impact in medicinal chemistry[3].

In-silico drug design methods can be divided in two main groups: the structure-based and the ligand-based methods. Structure-based drug design (SBDD) is based on the availability of the target structure, which is usually determined by X-ray, NMR (nuclear magnetic resonance) or electron microscopy. When the 3D structure of the target is not available, computational techniques like homology modelling can be used to predict its spatial arrangement. Docking is the best represented technique of this group, aimed at identifying the binding pose of the molecule and revealing the key interactions with the target that are needed for compound activity. The affinity of a molecule for its target is predicted by the docking scoring function, which estimates the energetic contributions of the interaction between target and ligand[4].

Another approach frequently used in SBDD is represented by Molecular Dynamics (MD) simulations. This method could be useful to give insights into ligand binding mode towards its target, accounting also for complex flexibility, compared to the docking approach where the structure of the target is usually maintained fixed. MD simulations can also be exploited to understand the behavior of the target, taking into account conformational changes and structural fluctuations that might happen in different environments (e.g. water, organic solvents, membrane)[5]. Under physiological conditions, proteins constantly move from one structural state to another and their function and regulation depend on these conformational changes. MD simulations are often used to identify novel conformations, to capture the transitional pathways between conformations and to determine equilibrium distributions among different conformations[6].

On the other hand, ligand-based methods are useful to capture information from known ligands. This approach relies on the assumption that compounds with similar structures also

---

[1] DiMasi, J. A.; Grabowski, H. G.; Hansen, R. W. Innovation in the Pharmaceutical Industry: New Estimates of R&D Costs. *J Health Econ* **2016**, *47*, 20–33.

[2] Jorgensen, W. L. The Many Roles of Computation in Drug Discovery. *Science* **2004**, *303* (5665), 1813–1818.

[3] Zhang, L.; Tan, J.; Han, D.; Zhu, H. From Machine Learning to Deep Learning: Progress in Machine Intelligence for Rational Drug Discovery. *Drug Discov. Today* **2017**, *22* (11), 1680–1685.

[4] Schneider, G.; Böhm, H.-J. Virtual Screening and Fast Automated Docking Methods. *Drug Discovery Today* **2002**, *7*, 64–70.

[5] Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nature Structural & Molecular Biology* **2002**, *9* (9), 646–652.

[6] Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H.; Shaw, D. E. Biomolecular Simulation: A Computational Microscope for Molecular Biology. *Annual Review of Biophysics* **2012**, *41* (1), 429–452.

share a similar biological activity[7]. Pharmacophore analysis, 2D similarity and shape screening are the most exploited techniques of this approach. A pharmacophore model describes the essential, steric and electronic features necessary for an optimal interaction with a relevant pharmacological target. It does not represent a real molecule or a real association of functional groups, but a purely abstract concept that accounts for the common molecular interaction capacities of a group of compounds towards their target structure. This approach could be used in both ligand-based and structure-based drug design, depending on the availability of ligands or target protein structures[8].
2D similarity and shape screening are useful when structural information about active molecules is available. They are based on the computation of similarity scores between compounds, based on the 2D or 3D structures of the ligands. These approaches are particularly applied when the binding partner is unknown or only little information is available.
Another important improvement carried out by the introduction of computers in drug design is the possibility to easily manage large database of data thanks to the development of data-mining softwares. During the last years, the growth of protein structures, genomic and biological data repositories, especially in the public domain, has increased, representing an incredible resource for drug design in both academia and industries. Computers are useful in handling these large amount of data, helping in extracting the relevant information[9].

In the projects carried on during my thesis period, different computational approaches were used, aimed at the identification of molecules potentially active as therapeutic agents and at the characterization of their mechanism of action. In the second chapter, a virtual screening that combined structure-based and ligand-based techniques was performed to identify compounds able to disrupt a specific protein-protein interaction in bacterial RNA polymerase. The most promising compounds have been tested in experimental assays, highlighting their activity on the target and their antimicrobial potential. Further computational studies were performed to rationalize their activity and to identify a common binding mode to the target.
Docking studies were applied to the first identified small-molecule inhibitor of the enzyme NAPE-PLD, in the third chapter. This enzyme is involved in the synthesis of specific bioactive lipids that affect physiological and pathological conditions. The binding pose identified by docking was supported by mutagenesis experiments and it represents a starting point towards the identification of new modulators of NAPE-PLD. Molecular dynamics simulations were also performed to investigate for NAPE-PLD a typical phenomenon of lipase enzymes called “interfacial activation” and gain insight into the catalytic mechanism of the protein.
In the fourth chapter of this thesis, a statistical analysis was carried out on a large database of biological data to identify the most promising compounds active against schistosomiasis, a neglected disease affecting millions of people living in tropical areas of the world.

---

[7] Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do Structurally Similar Molecules Have Similar Biological Activity? *Journal of Medicinal Chemistry* **2002**, *45* (19), 4350–4358.
[8] Wolber, G.; Seidel, T.; Bendix, F.; Langer, T. Molecule-Pharmacophore Superpositioning and Pattern Matching in Computational Drug Design. *Drug Discovery Today* **2008**, *13* (1), 23–29.
[9] Bajorath, J. Compound Data Mining for Drug Discovery. In *Bioinformatics: Volume II: Structure, Function, and Applications*; Keith, J. M., Ed.; Methods in Molecular Biology; Springer New York: New York, NY, 2017; pp 247–256.

2. Identification of β'-σ^{70} interaction inhibitors in bacterial RNA polymerase

2.1 Introduction

2.1.1 Antibiotic resistance

The discovery and introduction of antibiotics in therapeutics about 90 years ago (the first antibiotic, penicillin, was discovered by A. Fleming in 1928) represented a real milestone in the medicine field worldwide. Unfortunately, antibiotics success has revealed its weak side during the last years since the emergency of antibiotic resistance outbreaks. According to WHO, the abuse and overuse of antibiotics in human medicine, agriculture and animal agriculture has brought to the development of MDR (multi-drug resistant) bacteria capable of resisting to broad-spectrum antibiotics being the cause of one of the most important public health issues of the 21st century[10].

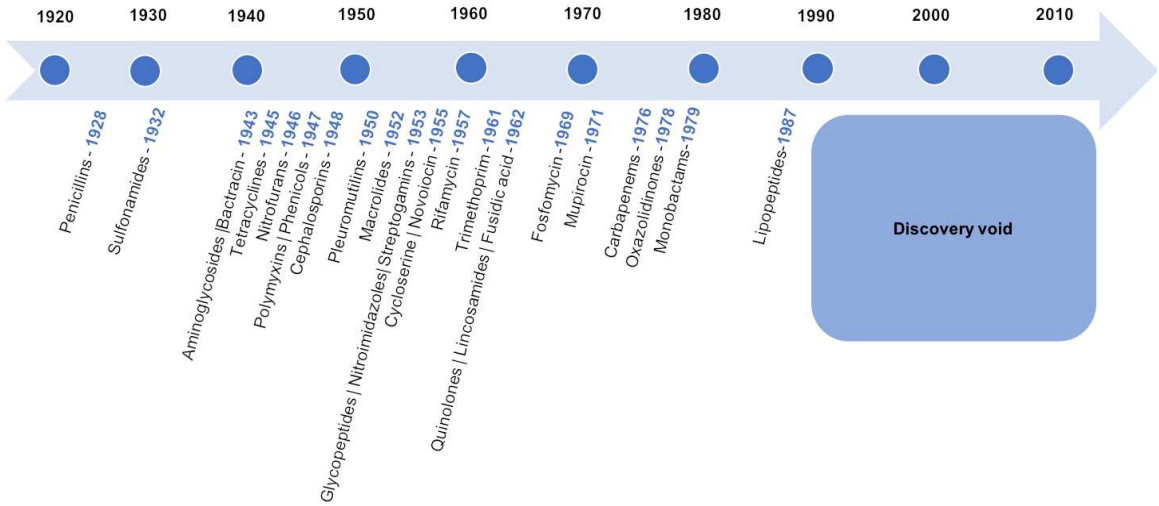


Figure 1: *the timeline of the discovery of antibiotics in the last 30 years is reported. Since 1990 no new classes of antibiotics have been discovered so far, the so called 'antibiotic discovery void'.*

Infections caused by superbugs are associated with increased mortality: it is estimated that 17 million people died each year by bacterial infection all over the world[11]. The CDC (Centers for Disease Control and Prevention) reports that each year in the United States approximately 2 million people are infected with bacteria resistant to antibiotics and 23 000 of them will subsequently die as a result of these infections[12]. Another aspect to take into account is the economic burden for the healthcare system: the cost of hospitalizations and pharmacological treatments as consequences of resistant infections is estimated to be billions of dollars per year in the US[13].

[10]WHO Antimicrobial resistance: global report on surveillance 2014. World Health Organization, Geneva; 2014: 257

[11] Martens, E.; Demain, A. L. The Antibiotic Resistance Crisis, with a Focus on the United States. *The Journal of Antibiotics* **2017**, *70* (5), 520–526.

[12]Centers for Disease Control and Prevention. Antimicrobial resistance. https://www.cdc.gov/drugresistance/index.html. Page last reviewed: August 18, 2017

[13] Cosgrove, S. E. The Relationship between Antimicrobial Resistance and Patient Outcomes: Mortality, Length of Hospital Stay, and Health Care Costs. *Clin. Infect. Dis.* **2006**, *42 Suppl 2*, S82-89.

This situation is worsened by the decrease in the development of new antibacterial agents by pharmaceutical companies that have stopped the economic investments in the antibiotic field causing a significant drop in the identification of new classes of antibacterial agents, thus originating the so called "antibiotic void". Unfortunately, this results in the emergence of infections that are almost untreatable, leaving clinicians with no reliable alternatives to treat patients. This apocalyptical scenario highlights the importance of antibiotic research and the necessity to find new molecules able to treat bacterial infections.

2.1.2 Antibiotic resistance mechanisms

Antimicrobial resistance is the result of the interaction between bacteria and environment and it's a very ancient process. In fact, the majority of antibacterial drugs derives from metabolites produced by microorganisms and bacteria have evolved different mechanisms to survive to the action of these natural products.
The development of resistance in bacteria can have two different bases:

A. *Genetic basis:* bacteria display a remarkable genetic plasticity that allows them to act against a large number of environmental threats, including antibiotics.
B. *Mechanistic basis:* microorganisms have evolved complex mechanisms of drug resistance during millions of years of evolution.

A. GENETIC BASIS OF RESISTANCE

Bacteria adopt two major genetic strategies to survive to antibiotics:

- *Mutational resistance*: mutations in bacteria gene(s) which is often linked to the mechanism of action of the compound. These mutations nullify the antibiotic action using different strategies: ***i)*** modification of the bacterial target ***ii)*** activation of efflux pumps that avoid reaching effective concentrations of antibiotics ***iii)*** modification of metabolic pathways. These mutations drive the selection of a population of resistant bacteria that can overcome the action of the antibiotic, while the susceptible microorganisms succumb by the antibiotic action.

- *Horizontal gene transfer (HGT):* mechanism that helps bacteria to acquire foreign DNA that code for resistance genes. Three main mechanisms of HGT are reported: ***i***) transformation, in which foreign DNA passes through the medium and it is incorporated into the recipient cell, ***ii***) phage mediated transduction, in which external DNA is injected by a bacteriophage virus in the recipient cell and, ***iii***) conjugation, which is the transfer of genetic material between two bacterial cells in direct contact.

B. MECHANISTIC BASIS OF RESISTANCE
In millions of years of evolution, bacteria have developed complex mechanisms to survive the antimicrobial activity and different species of bacteria display some preferential resistance mechanisms over others. Gram-negative bacteria can produce peculiar enzymes called β-lactamases that inactivate antibiotics with β-lactam nucleus. On the other hand, Gram-positive bacteria mechanisms of resistance are mainly characterized by target site modifications.
The mechanistic basis of microbial resistance can be summarized in:

- *Modification of antibiotics molecules*

- *Chemical alterations of the antibiotic:* bacteria generate enzymes that chemically modify the antibiotic molecule. Acetylation, phosphorylation or adenylation are the most frequently catalyzed reactions. An example is represented by the modification of the drug chloramphenicol by CAT (chloramphenicol acetyltransferase) enzyme[14]. This enzyme is expressed in both Gram-positive and -negative bacteria and it is classified in type A which results in high chloramphenicol resistance and B that gives low chloramphenicol resistance.
- *Destruction of the antibiotic* represents a very well-known mechanism of resistance. A classic example is given by β-lactamases, bacterial enzymes that disrupt the amide bond of the β-lactam ring inactivating β-lactam antibiotics. Different types of β-lactamases have been discovered up to date as new generation of β-lactams had been marketed, displaying how bacteria are effective in the adaptive evolution process[15].

o *Decreased Antibiotic Penetration and Efflux*

- *Decreased permeability:* many antibiotics target a bacterial intracellular site; thus, they have to penetrate the outer and/or the inner membrane to exert their action. Bacteria have developed strategies to prevent the antibiotics reaching the intracellular target by modifying the expression of porins, which are proteins that regulate the influx/efflux of compounds. Modifications of porins include the expression of different type of porins or the deterioration of porin functions. This type of resistance mechanism is usually associated with other resistance mechanisms such as the expression of efflux pumps.
- *Efflux Pumps* can be produced by bacteria to extrude a toxic compound out of the cell. This mechanism can result in antimicrobial resistance in both Gram-positive and -negative bacteria. Five families of efflux pumps have been identified so far and they differ for the type of substrate they can extrude, their structure, the source of energy and the microbial species in which they are found. The main classes of efflux pumps are ***i)*** the ATP-binding cassette family (ABC)*,* ***ii)*** the multidrug and toxic compound extrusion family (MATE), ***iii)*** the major facilitator superfamily (MFS), ***iv)*** the resistance-nodulation-cell- division family (RND) and ***v)*** the small multidrug resistance family (SMR)[16].

o *Changes in target sites*

- *Target protection:* in this case bacterial proteins can prevent the interaction between the antibiotic molecule and the target site avoiding the antibacterial action. One of the best studied examples of this mechanism is represented by the tetracycline resistance proteins Tet(M) and Tet(O)[17]. These proteins belong to the GTPase superfamily and they interface with the ribosome displacing the tetracycline from its binding site. Moreover, this interaction can modify the conformation of the ribosome preventing the tetracycline to rebind its target.
- *Modification of the target site:* this is the most common mechanism of resistance that affects almost all families of antibiotic molecules and results in a loss of affinity between the compound and the target site. These target modifications consist of ***i)***

[14] Schwarz, S.; Kehrenberg, C.; Doublet, B.; Cloeckaert, A. Molecular Basis of Bacterial Resistance to Chloramphenicol and Florfenicol. *FEMS Microbiology Reviews* **2004**, *28* (5), 519–542.
[15] Bush, K. Past and Present Perspectives on β-Lactamases. *Antimicrob. Agents Chemother.* **2018**, *62* (10).
[16] Duraes, F. A. P. M.; Pinto, M. M. M.; de Sousa, M. E. S. P. Medicinal Chemistry Updates on Bacterial Efflux Pump Modulators. *Curr. Med. Chem.* **2018**.
[17] Connell, S. R.; Tracz, D. M.; Nierhaus, K. H.; Taylor, D. E. Ribosomal Protection Proteins and Their Mechanism of Tetracycline Resistance. *Antimicrob Agents Chemother* **2003**, *47* (12), 3675–3681.

point mutations in genes that codify for the target site (one example is the resistance to rifampin in the RNA polymerase[18]), ***ii)*** enzymatic modifications of the binding site (methylation of the ribosome catalyzed by specific bacterial enzymes that results in macrolide resistance[19]), and/or ***iii)*** replacement or bypass of the original target. This strategy is accomplished by the fact that bacteria can synthetize target proteins with functions similar to the native one, but resistant to the antibiotic molecules. One example is given by methicillin resistant *S. aureus*. Its *mecA* gene codifies for the PBP2 protein which is very similar to PBP, the enzymes responsible for the synthesis of peptidoglycan, but is not susceptible to β-lactams.

- *Resistance Due to Global Cell Adaptations:* in millions of years of evolution bacteria had to fight for their survival and they had evolved different mechanisms to adapt to new and hostile environments. Inside the infected host microbes are constantly under the pressure of the immune system. Moreover, bacteria have to compete for food, thus they have to overcome very stressful situations. These mechanisms are mostly related to the maintenance of cell wall synthesis and membrane homeostasis.

2.1.3 Bacterial RNA polymerase

RNA polymerase (ribonucleic acid polymerase) is an essential enzyme for all the organisms since it catalyzes the transcription reaction. This process is necessary to convert DNA into RNA which will be then translated into proteins. In prokaryotes there is just one RNAP involved in gene transcription, while eukaryotes possess multiple types of RNAP, each one responsible for the synthesis of a specific subset of RNAs:

- RNAP I responsible for the synthesis of ribosomal RNA;
- RNAP II responsible for the synthesis of mRNA;
- RNAP III synthetizes tRNA (transfer RNA), rRNA (ribosomal RNA) and other small RNAs found in nucleus and cytosol;
- RNAP IV and V are only found in plants, where they synthetize specific types of RNAs.

The enzyme was first discovered by Hurwitz in early 1960s[20] and since then a lot of knowledge has been acquired. Thanks to his studies on RNA polymerases, in particular the discovery of the molecular basis of transcription in eukaryotic organisms, Roger D. Kornberg was awarded the Nobel Prize in Chemistry in 2006.
The bacterial RNA polymerase is an enzyme of about 400 kDa composed by five different subunits that form the so called "core enzyme": subunits α_I, α_{II}, β, β'and ω. The core enzyme is responsible for binding to template DNA and synthetize RNA, but for the reaction to occur is necessary a transcription factor called σ. Core enzyme and the transcription factor form the holoenzyme which recognizes the promoter sequence and start the transcription process.

Core enzyme
Bacterial RNAP core enzyme is composed by five different subunits which are conserved in all the three domains of life (Archaea, Bacteria and Eukarya). Subunit α_I and its copy, α_{II}, are characterized by two domains: the N-terminal domain or αNTD and the C-terminal

[18] Campbell, E. A.; Korzheva, N.; Mustaev, A.; Murakami, K.; Nair, S.; Goldfarb, A.; Darst, S. A. Structural Mechanism for Rifampicin Inhibition of Bacterial RNA Polymerase. *Cell* **2001**, *104* (6), 901–912.
[19] Weisblum, B. Erythromycin Resistance by Ribosome Modification. *Antimicrob Agents Chemother* **1995**, *39* (3), 577–585.
[20] Hurwitz, J. The Discovery of RNA Polymerase. *J. Biol. Chem.* **2005**, *280* (52), 42477-42485.

domain or αCTD which are connected by a linker sequence. The αNTD is necessary for the assembly of β and β' subunits while αCTD is needed for the interaction with promoter DNA and regulatory factors. β' is the largest subunit of RNAP and together with subunit β they contain the active site in which DNA template is bound and RNA is synthetized. ω is the smallest subunit and it seems to have no direct role in transcription, but it could help the assembly of RNAP and the folding of β' subunit[21].

Transcription factors σ in bacteria[22]

Transcription factors σ are needed in order to start the transcription machinery: they provide the determinants necessary for promoter sequence recognition and DNA melting, but these proteins works only when assembled to the RNAP. All bacteria have a primary σ factor which initiate the transcription of housekeeping genes under normal conditions. In *E. coli*, the primary sigma factor is σ^{70}, reflecting its molecular mass of approximately 70000 Da. In many other bacteria, in particular Gram-positive, the analogous primary σ factor is designated as σ^{A}. Most bacteria also have alternative σ factors (six were found in *E. coli*) that activate transcription under specific stress conditions: in *E. coli* σ^{54} also called σ^{N} is responsible for the expression of genes involved in nitrogen utilization or σ^{E} which is a σ factor that accumulates in response to heat-shock stress enabling RNAP to recognize promoters of genes helping the cell to survive at elevated temperatures. σ proteins are characterized by four domains as reported in Figure 2.

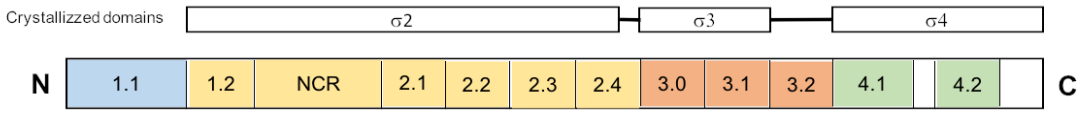


Figure 2: *σ^{70} domains are represented. The autoinhibitory domain 1 in blue, domain 2 in yellow, domain 3 in orange and domain 4 in green. Only domains 2, 3 and 4 have been crystallized.*

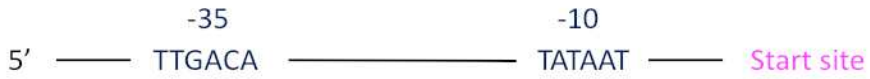
Domain 1 contains region 1.1, which is an autoinhibitory domain that masks DNA binding region in free σ^{70}. This domain hadn't been crystallized so far. Domain 2 includes regions 1.2 to 2.4 and comprises a non-conserved region (NCR). Residues in 2.3 are implicated in the melting process of DNA. Domain 3 comprises regions 3.0-3.2 and domain 4 includes region 4.1 and 4.2. These domains are connected via flexible linkers. Regions 2.4 and 3.0 can recognize the -10 and -10 extended motifs of DNA, while region 4.2 can bind the -35 motif on the DNA, starting the transcription process.

Transcription reaction[23]

1. Initiation stage

RNAP initiates transcription binding to specific DNA sequences known as promoters. Two common motifs are present on the 5′ (upstream) side of the start site. They are known as the *-10 sequence* and the *-35 sequence* because they are centered at about 10 and 35 nucleotides upstream of the start site. These sequences are each 6 base pairs long.

[21] Browning, D. F.; Busby, S. J. W. The Regulation of Bacterial Transcription Initiation. *Nature Reviews Microbiology* **2004**, *2* (1), 57–65.
[22] Gruber, T. M.; Gross, C. A. Multiple Sigma Subunits and the Partitioning of Bacterial Transcription Space. *Annual Review of Microbiology* **2003**, *57* (1), 441–466.
[23] Berg, J. M.; Tymoczko, J. L.; Stryer, L. Transcription Is Catalyzed by RNA Polymerase. *Biochemistry. 5th edition* **2002**.



The *σ* subunit contributes to specific initiation in two ways. First, it decreases the affinity of RNA polymerase for non-promoter regions of DNA by a factor of 10^4. In its absence, the core enzyme binds DNA indiscriminately and tightly. Second, the σ subunit enables RNA polymerase to recognize promoter sites. A large fragment of a σ subunit was found to have an α-helix on its surface implicated in recognizing the 5′-TATAAT sequence of the -10 region.

The holoenzyme binds to duplex DNA and moves rapidly along the double helix in search for a promoter to initiate transcription. As mentioned in the previous section, there are different σ factors, each of which recognize a distinct set of promoters that activates transcription of different genes.

After binding to DNA, the RNA polymerase switches from a closed complex to an open complex: this change involves the separation of the DNA duplex strand to form an unwound section of DNA. The transition from the *closed promoter* to the *open promoter complex* is an essential event in transcription and is necessary for the transcription to begin. The stage is now set for the formation of the first phosphodiester bond of the new RNA chain.

2. Elongation stage

This step is the process of adding nucleotides to the RNA, which is complementary to the template DNA strand. RNA synthesis can start *de novo,* without the necessity of a primer. Most newly synthesized RNA chains carry a highly distinctive tag on the 5′ end: the first base at that end is either *pppG* or *pppA*.

The synthesis of RNA follows the 5'→ 3'direction as the DNA synthesis.

The complex between RNAP, DNA and newborn RNA is referred to as the "transcription bubble" (the term bubble is referred to the melted bubble of DNA present). The newly synthesized RNA forms a hybrid helix with the template DNA strand. The 3′-hydroxyl group of the RNA in this hybrid helix is positioned so that it can attack the α-phosphorus atom of an incoming ribonucleoside triphosphate. RNA polymerase does not correct the nascent polynucleotide chain. Consequently, the fidelity of transcription is much lower than that of replication. The error rate of RNA synthesis is of the order of one mistake per 10^4 or 10^5 nucleotides, about 10^5 times as higher as that of DNA synthesis. The much lower fidelity of RNA synthesis can be tolerated because mistakes are not transmitted to progeny. For most genes, many RNA transcripts are synthesized; a few defective transcripts are unlikely to be harmful.

3.Termination stage

The termination of transcription is as precisely controlled as its initiation. In the termination phase of transcription, the formation of phosphodiester bonds ceases, the RNA-DNA hybrid dissociates, the melted region of DNA rewinds and RNA polymerase releases the DNA. The transcribed regions of DNA templates contain stop signals which determine where transcription is terminated. The simplest stop signal or terminator is a palindromic GC-rich region followed by an AT-rich region. This palindromic sequence often displays a hairpin conformation followed by a sequence of four or more uracil residues, which also are crucial for termination. The RNA transcript ends within or just after them. In some cases, Rho protein specifically binds RNA and helps terminating the transcription of some genes.

2.1.4 Crystal structure of bacterial RNA polymerase

Up to date about 60 structures of RNAP belonging to different bacterial species have been solved using X-ray crystallography, NMR spectroscopy or electron microscopy. The first crystal structure of RNAP core enzyme was obtained from the thermophilic bacterium *T. aquaticus* (PDB: 1HQM)[24].

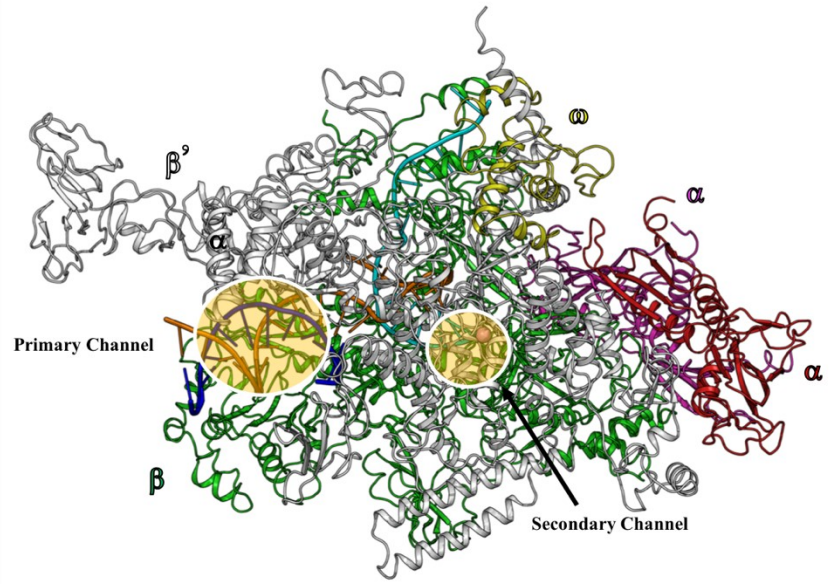


Figure 3: *RNAP from T. thermophilus (PDB:2O5I) is represented in complex with DNA and RNA during the elongation stage of transcription. Each subunit composing the core enzyme is represented with different colors: α subunits are red and magenta, β subunit is green, β' is white and ω subunit is yellow. DNA is located in the primary channel and the template strand is depicted in orange, while the non-template strand in blue. RNA single strand is depicted in cyan. A Mg^2 ion of the active site is shown as pink sphere. The secondary channel is located near the active site.*

Solving the first crystal structure of RNAP was a real milestone in the study of bacterial transcription and years later other crystal structures of bacterial RNA polymerase in complex with transcription factors and DNA were solved, providing further insights into the transcription mechanism.

The structure of RNAP revealed a unique crab claw-shaped molecule composed by the two subunits β and β′ that form the two pincers of the claw. The cleft inside the claw displays a positive charge to help the melting of DNA double strand, while the surface of RNAP is negatively charged. Some of the most relevant structural features of RNAP enzyme are:

- the primary channel is located in the cleft between the two subunits β and β**'** and it is large enough to accommodate the double stranded DNA template for RNA synthesis;
- the switch region is placed at the top of the RNAP clamp and it mediates the opening and closing of the clamp to load the DNA in the active site. This region is composed by five segments, called from "switch 1" to "switch 5";
- the β′ bridge helix separates the deep groove of RNAP into a DNA or active channel and a NTP entry secondary channel;
- the secondary channel is located near the active site and mediates NTP entrance;

[24] Minakhin, L.; Bhagat, S.; Brunning, A.; Campbell, E. A.; Darst, S. A.; Ebright, R. H.; Severinov, K. Bacterial RNA Polymerase Subunit ω and Eukaryotic RNA Polymerase Subunit RPB6 Are Sequence, Structural, and Functional Homologs and Promote RNA Polymerase Assembly. *PNAS* **2001**, *98* (3), 892–897.

- The β′ subunit trigger loop/helix (TLH) plays a critical role in the nucleotide addition cycle. The front edge of the TLH is highly flexible, but it becomes a rigid "trigger helix" structure when an incoming nucleotide is present at the active site;
- The β-flap region is placed on the upstream of RNAP and form the RNA exit channel;
- The clamp-helix region is located in the β' subunit, far from the active site of the enzyme and it interacts with region 2.2. of the transcription factor σ.

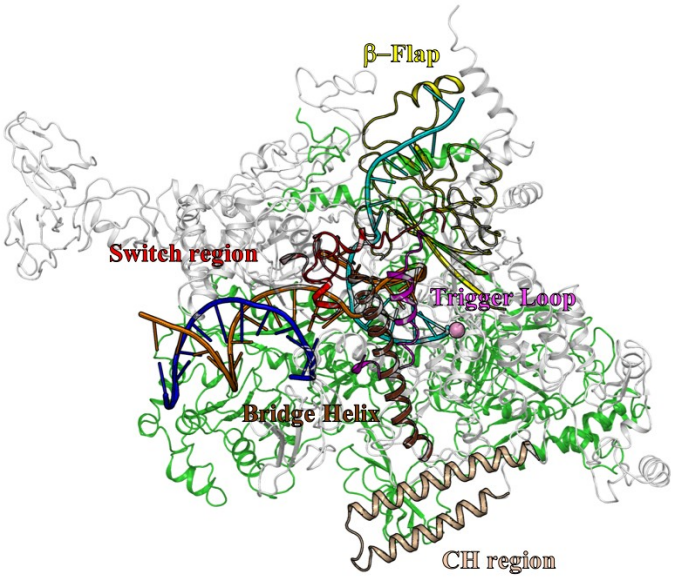


Figure 4: *β subunit in green and β' subunit in white of RNAP from T. thermophilus (PDB:2O5I) are represented in complex with DNA (template strand in orange and non-template strand in blue) and RNA (single strand in cyan) during the elongation stage of transcription. A Mg^{2+} ion of the active site is shown as pink sphere. Some structural elements of RNAP are reported: clamp-helix region in light brown, trigger loop in magenta, flap region in yellow, switch region in red and bridge helix in brown.*

RNAP active site

The active site of the enzyme is highly conserved among RNAP. It usually contains two Mg^{2+} ions that are responsible for phosphodiester bond formation: the Mg^{2+}A is bound to three aspartate residues in the β′ catalytic loop which is characterized by the conserved sequence NADFDGD. This ion is responsible for the activation of the RNA 3'-terminal hydroxyl group and for the coordination of the α-phosphate of the incoming nucleotide. The Mg^{2+}B is located not far from the other ion and it stabilizes the transition state, neutralizing the negative charge of the pyrophosphate group.

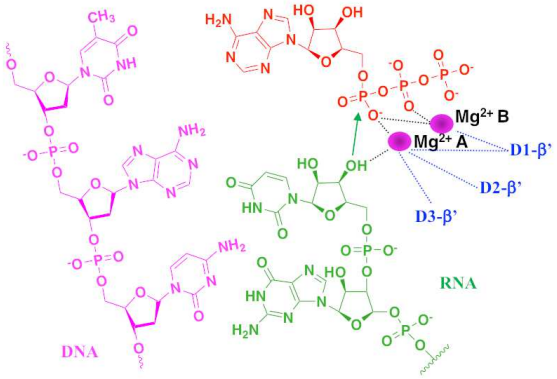


Figure 5: *a 2D representation of RNAP active site with the two Mg^{2+} ions coordinated by the three aspartic residues (in blue) of the conserved catalytic loop NADFDGD of β' subunit. The DNA template strand is reported in magenta, the new RNA strand in green and the incoming nucleotide in red. Mg ions are involved in the coordination of NTP phosphates, ensuring their optimal alignment for the nucleophilic attack of the activated 3'-OH group of the RNA on the α-phosphate.*

2.1.5 ß-∂ interaction in bacterial RNA polymerase as target for the discovery of new antibiotics

The availability of X-ray structures of RNAP holoenzyme of different bacterial species in public databases has provided accurate information about intermolecular interactions occurring between the core enzyme and the transcription factor σ revealing a unique protein-protein interaction (PPI) between the β' subunit and σ factor which is very well conserved across different bacterial species. This PPI could be exploited as a target for the discovery of new antibacterial agents. The hotspot is located at the interacting surface between two highly conserved regions: the clamp-helix (C-H) region of β' subunit and region 2.2 of σ factor. The crystal structure of *E. coli* RNAP holoenzyme (PDB: 4YG2)[25] shows that the clamp-helix region of the β' subunit interacts with region 2.2 of the σ^{70} factor undertaking both polar and hydrophobic interactions. In particular, arginine residues R275, R278 and R281 belonging to the β' subunit undertake salt-bridges mediated interactions with acidic residues D403 and E407 of the σ protein, while hydrophobic contacts, stabilizing the protein-protein interaction, involve residues L402 and I410 of the σ^{70} factor and L282 and M298 from the β' subunit, as reported in Figure 6b**.**

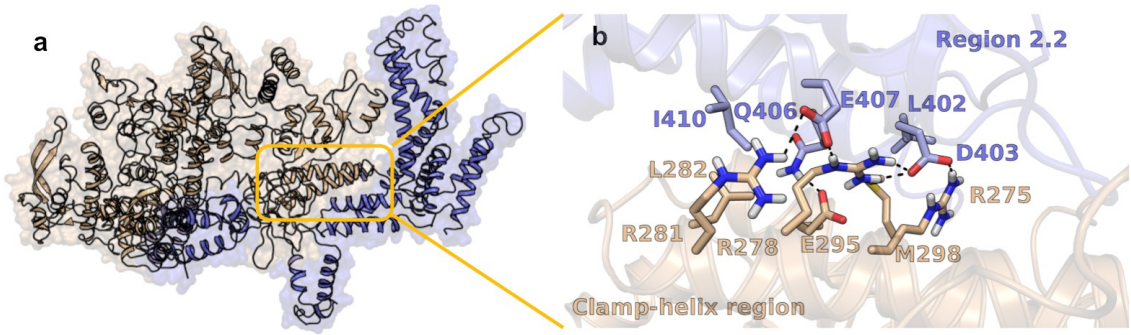


Figure 6: *subunit β' and σ^{70} of E. coli RNA polymerase (PDBID: 4YG2) are represented in light brown and light blue, respectively.* ***The protein-protein interaction between the clamp-helix region of β' subunit and region 2.2 of σ^{70} is circled in yellow (a). Close up view of amino acids involved in β'-σ^{70} binding. Polar interactions between the two subunits are indicated with black dashed lines (b).***

Mutagenesis experiments have been performed on both β' clamp-helix region and σ factor 2.2 region of different bacterial RNA polymerases, giving insights into the amino acids necessary for the protein-protein interaction to occur.
Alanine-scanning experiments were carried out on the β' clamp–helix region of *B. subtilis* highlighting the importance of arginine residues R264, R267 and R270: mutations of these residues have critical effects on σ binding, resulting in total disruption of the protein-protein interaction. Mutation of amino acids L271, I280, N283 and M287 also display severe effects on σ^{70} binding[26]. Change of charge mutations on the clamp-helix region of *E. coli,* demonstrated that R275Q and E295K avoid holoenzyme formation[27]. Mutagenesis experiments were also carried out on the 2.2 region of the σ factor and many residues were described as critical for interaction with β': E166 of *B. subtilis* was reported to be the most

[25] Murakami, K. S. X-Ray Crystal Structure of *Escherichia Coli* RNA Polymerase σ [70] Holoenzyme. *Journal of Biological Chemistry* **2013**, *288* (13), 9126–9134.
[26] Ma, C. *et al.* Inhibitors of Bacterial Transcription Initiation Complex Formation. *ACS Chemical Biology* **8,** 1972–1980 (2013).
[27] Arthur, T. M., Anthony, L. C. & Burgess, R. R. Mutational Analysis of '260-309, a 70 Binding Site Located on Escherichia coli Core RNA Polymerase. *Journal of Biological Chemistry* **275,** 23113–23119 (2000).

critical residue[28]. Mutations Q165K and Q165A decrease binding capacity[29]. Alanine scanning analysis was also carried out on *E. coli* σ^{70} subunit showing that E407A (equivalent residue of *B. subtilis* E166), Q406A (equivalent residue of *B. subtilis* Q165) and D403N mutations (equivalent residue of *B. subtilis* D162)[30] interfere with β' binding. Altogether, mutagenesis results prove the existence of a defined hotspot between the two proteins: this interaction could be an attractive target for the discovery of small molecules that can strongly obstruct the transcription initiation complex formation acting as antibacterial drugs.

Table 1: *amino acids for which mutagenesis data are available. Cells are color-coded from green to red from least to most critical residues* for β'-σ *interaction. Residues are numbered following the sequences of E. coli and B. subtilis.*

Coiled-coil region β'	
E.Coli	*B. Subtilis*
R275	R264
R278	R267
R281	R270
M298	M287
N294	N283
L282	L271
I291	I280
E295	E284

Region 2.2 σ Factor	
E.Coli	*B. Subtilis*
E407	E166
Q406	Q165
I410	M169
D403	D162
M413	M172

2.1.6 Molecules affecting bacterial RNA polymerase activity

There are several molecules known to inhibit bacterial RNA polymerase activity which can bind different sites of the enzyme and can act with different mechanism of actions. The majority of them has been isolated from microorganisms, but synthetic compounds have also been identified.
Up to date the only molecules approved for clinical use that act as inhibitor of transcription are rifamycins and fidaxomicin/lipiarmycin.

[28] Johnston, E. B, The RNA Polymerase-sigma interaction as a target for potential novel antimicrobials, in School of Environmental and Life Sciences, University of Newcastle, Newcastle, Australia (2010).
[29] Johnston, E. B., Lewis, P. J. & Griffith, R. The interaction of *Bacillus subtilis* σ^{A} with RNA polymerase. *Protein Science* **18,** 2287–2297 (2009).
[30] Sharp, M. M. *et al.* The interface of ς with core RNA polymerase is extensive, conserved, and functionally specialized. *Genes & development* **13,** 3015–3026 (1999).

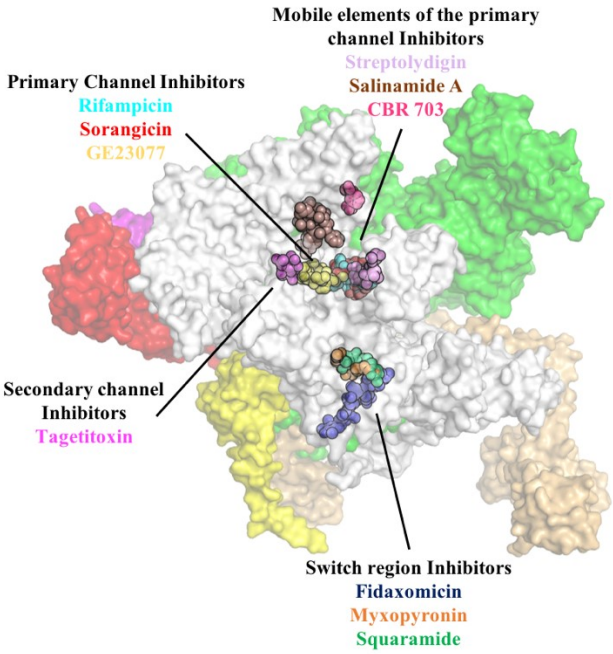


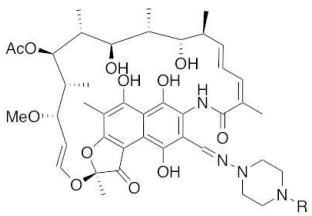
Figure 7: *surface representation of RNAP E. coli holoenzyme with different small-molecule RNAP inhibitors bound. Each subunit of RNAP is represented as colored surface: the two α subunits are represented in red and magenta, subunit β in green, subunit β' in white, subunit ω in yellow and σ transcription factor in light orange. Small molecule inhibitors bound to RNAP are represented as colored spheres and the name and the inhibitor classes are reported.*

In 2016 Lewis and co-workers have divided bacterial transcription inhibitors in different groups based on their target site on RNAP[31]:

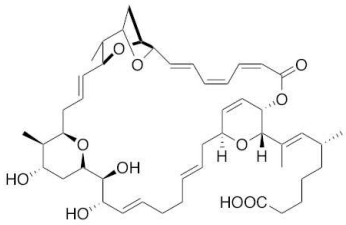
Primary channel inhibitors

The primary channel is the large cleft between β and β' subunits and it is highly conserved among RNAPs. The cleft displays a positive charge to help the melting of DNA double strand. The active site of the enzyme is located at the end of the channel, where transcription reaction occurs.

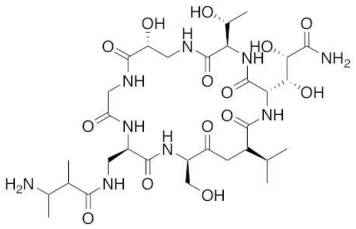
Primary Channel Inhibitors



Rifampicin R=Me
Rifapentine R= cyclopentyl



Sorangicin



GE23077

- **Rifamycins**

They were the first group of antimicrobials targeting bacterial RNAP discovered and they were identified by two microbiologists, working for the Italian drug company Lepetit SpA, as metabolites of *Amycolatopsis mediterranei* (known also as *Streptomyces mediterranei*)

[31] Ma, C.; Yang, X.; Lewis, P. J. Bacterial Transcription as a Target for Antibacterial Drug Development. *Microbiology and Molecular Biology Reviews* **2016**, *80* (1), 139–160.

in 1957. Cocrystal structure of rifamycins[32] with RNA polymerases from different bacteria revealed that their binding site is adjacent to the active site of the enzyme, sterically hindering the growth of RNA. RIF are part of first line therapy against tuberculosis, but their antimicrobial activity suffers from resistance which is easily acquired by bacteria in contact with the drug. Mutations can develop at different sites of the enzyme very easily and efforts to minimize drug resistance are represented by combinatorial therapy: today the treatment of tuberculosis is based on a combination of RIF, ethambutol, isoniazid and pirazinamide.

- **Sorangicin A**

Sorangicins are a class of macrolide antibiotics discovered in 1985 from the fermentation broth derived from the myxobacteria *Sorangium cellulosum*. Sorangicin A, the prevalent compound, displayed extraordinary antibiotic activity against both Gram-positive and Gram-negative bacteria and a cocrystal structure of the molecule with RNAP from *T. thermophilus* revealed that it binds to the same site of RIF, adjacent to the active site, likely sharing the same mechanism of action. Even if the cyclic structure of sorangicin is similar to that of RIF, this molecule displays a more flexible skeleton that helps to adapt to the binding site. Studies on resistant strains of bacteria revealed that sorangicin is susceptible to a narrower range of mutants compared to RIF, probably due to its more flexible structure.

- **GE23077**

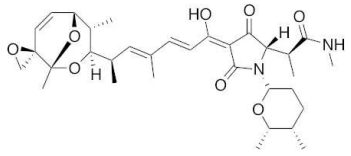
This molecule was isolated from the fermentation broth of an *Actinomadura* sp. It is a cyclic heptapeptide which shows strong inhibitory activity on both rifampin-sensitive and resistant polymerases. Unfortunately, this compound exhibits poor antimicrobial activity due to the lack of penetration of the molecule across the bacterial membrane, owing to its strong hydrophilic character. Crystal structures of GE23077 in complex with *T. termophilus* RNAP were published highlighting its binding site: the molecule binds to the i and i+1 site of the active center, near the catalytic Mg^{2+} ion and close to the RIF binding site. RNAP resistance against GE2377 is lower compared to the RIF and is determined by mutation of fewer amino acids. The combination of GE2377 and rifamycin SV displays good activity against GE2377 and RIF-resistant RNAP.

Mobile elements of the primary channel

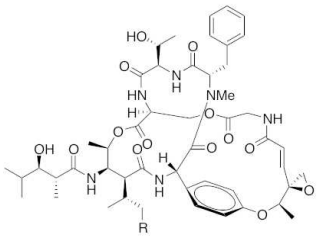
Transcription cycle is a highly dynamic and complex process and all the transcription stages are strictly regulated via mobile elements adjacent to the active site of the protein. The conserved bridge helix (BH) in the β' subunit of RNAP (Figure 4) makes contact with the mobile trigger loop (TL) and these two elements are involved in substrate loading and nucleotide addition by switching DNA binding cleft between open and closed conformation.

[32] Lin, W.; Mandal, S.; Degen, D.; Liu, Y.; Ebright, Y. W.; Li, S.; Feng, Y.; Zhang, Y.; Mandal, S.; Jiang, Y.; et al. Structural Basis of Mycobacterium Tuberculosis Transcription and Transcription Inhibition. *Molecular Cell* **2017**, *66* (2), 169-179.e8.

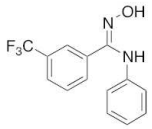
Mobile Elements of Primary Channel Inhibitors



Streptolydigin



Salinamide A R=Me
Salinamide B R=H



CBR703

- **Streptolydigin**

It is an antibiotic isolate from *Streptomyces lydicus* that works by inhibiting RNA elongation by binding to RNA polymerase: it decreases the rate of nucleotide addition during RNA synthesis. Streptolydigin binds to a region which is adjacent to the active site and it is involved in the nucleotide addition cycle. Bacterial resistance phenomena toward this molecule were reported in literature[33] and streptolydigin was no further developed as a drug.

- **Salinamides**

Salinamides A to E are byciclic polypeptide molecules isolated from both marine and soil *Streptomyces* spp. The crystal structure of salinamide A with *E. coli* RNAP revealed that this molecule binds to a region between the BH and secondary channel, making interactions with both β' and β chain residues. These compounds displayed good antibacterial activity and selectivity, since their binding site is not conserved in eukaryotic RNAP, but they showed poor membrane crossing properties due to their hydrophilic peptide structure.

- **CBR compounds**

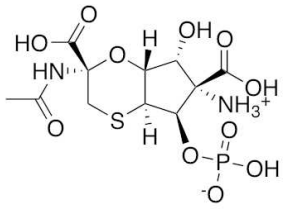
This series of small molecules displays two linked aromatic rings and they were discovered by screening of a compound library that assessed transcription inhibition in *E. coli.* These compounds showed the ability to inhibit RNAP transcription by slowing down nucleotides translocation and catalysis, weakening the interactions between BH and TL. Moreover, they showed the capability to inhibit biofilm formation which is the main source of nosocomial infections. Cocrystal structures show that CBR703 and its derivatives bind at the N-terminal of the bridge-helix of RNAP. These compounds displayed good antimicrobial activity against pathogen bacteria but unfortunately, they showed cytotoxicity and their hydrophobic character can lead to serum components binding and precipitation.

Secondary channel inhibitors

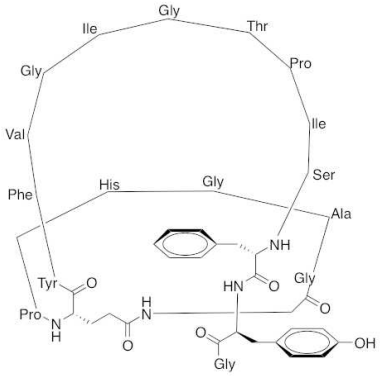
The secondary channel is a small channel in RNAP that allow nucleotide passage to reach the buried active site allowing RNA growth. The diameter of the channel allows only one nucleotide at a time to pass. The channel displays a negative charge that constrains NTP diffusion.

[33] Halling, S. M.; Burtis, K. C.; Doi, R. H. B′ Subunit of Bacterial RNA Polymerase Is Responsible for Streptolydigin Resistance in Bacillus Subtilis. *Nature* **1978**, *272* (5656), 837–839.

Secondary Channel Inhibitors



Tagetitoxin



Microcin J25

- **Tagetitoxin**

This molecule was isolated from the plant pathogen *Pseudomonas syringae* pv. *tagetis* and studies showed that it could target both chloroplast RNAP and RNAPIII but could also target bacterial RNAP, inhibiting the transcription elongation step. Its low cell permeability hampers antibacterial activity against live cells *in vitro.* The crystal structure of tagetitoxin with *T. thermophilus* RNAP shows that the compound binds to the secondary channel.

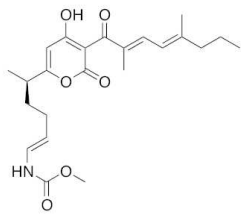
- **Microcin J25**

This is a 21 amino acids peptide produced by *E. coli* AY25 active against Gram-negative bacteria. Its mechanism of action seems related to its ability to block initiation and elongation stages by competitively preventing NPT uptake or binding. This peptide displayed low potency and it is susceptible to resistance phenomena thus it was no further developed.

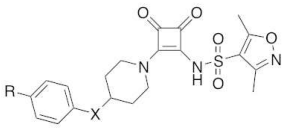
Switch region inhibitors

The claw structure of RNAP undergoes a series of conformational changes that allow DNA access to the catalytic site. In particular, the clamp part of the β' subunit displays large structural flexibility: in an open conformation it helps DNA loading into the active site, while the close conformation helps in retaining the DNA in the transcription bubble during the elongation stage of transcription. The swing motion of the clamp is regulated by the switch region, located at the base of the clamp domain. Some residues of this region interact with template DNA and newborn RNA transcript having regulatory roles in initiation and elongation. There are five segments of the switch regions named "switch 1", "switch 2" and so on.

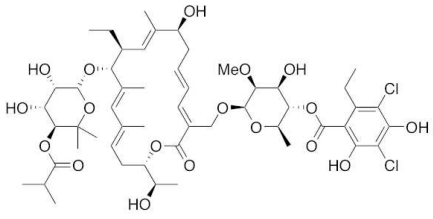
Switch Region Inhibitors



Myxopronin



Squaramides
X=CH_2, $NHCH_2$
R=Me,CF_3



Fidaxomicin

- **Fidaxomicin** or **lipiarmycin**

This is a macrocyclic drug approved by FDA in 2011 for the treatment of *Clostridium difficile* infections, a Gram-positive bacterium. Crystal structure of fidaxomicin in complex

with *M. tuberculosis* RNAP was published in 2018[34] highlighting structural elements of interaction between the compound and RNAP. Fidaxomicin binds to the "switch region" of bacterial RNA polymerase preventing its movement, acting like a doorstop to jam the enzyme in an open state, preventing the motions necessary to secure promoter DNA region in the active site.

- **Myxopironins**

These antibiotics were isolated from cultures of *Myxococcus fulvus* and inhibit transcription initiation displaying broad activity spectrum against both Gram-positive and -negative bacteria. They bind to the switch region in a hydrophobic pocket formed by "switch 1" and "switch 2" locking the clamp and switching the enzyme in a closed conformation, preventing the clamp from opening and loading DNA strands during transcription initiation.

- **Corallopyronins and Ripostatins**

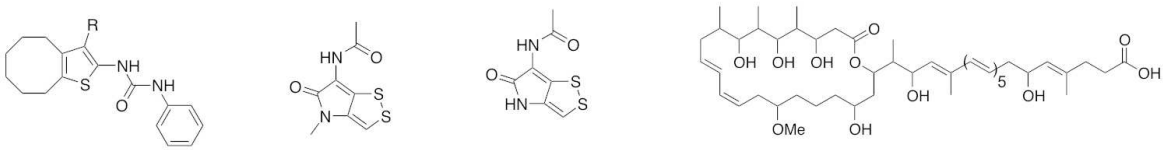
Corallopyronins are structurally similar to myxopyronins and, together with ripostatins, they bound the switch region.

- **Squaramides**

Synthetic small molecules discovered by an HTS. Their binding site is similar to that of myxopronins, and it was found that these compounds pushed "switch 2" into the DNA binding channel preventing the correct positioning of the template DNA.

Inhibitors with unknown targets

Inhibitors with unknown targets



- **Ureidothiophene**

It was discovered through a chemical library screening against *S. aureus* RNAP and demonstrated good antibacterial activity on Gram-positive bacteria.

- **Thiolutin and Holomycin**

They are natural products isolated from *Streptomyces luteosporeus* and *Streptomyces clavuligerus*, respectively. They demonstrated good activity against both bacterial and yeast RNAP, inhibiting bacterial transcription elongation and yeast transcription initiation.

- **Etnangien**

Antibiotic isolated from *Sorangium cellulosum* with activity against Gram-positive bacteria. *In vitro* inhibitory activity against *E. coli* RNAP was observed.

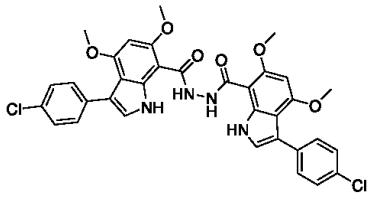
[34] Boyaci, H.; Chen, J.; Lilic, M.; Palka, M.; Mooney, R. A.; Landick, R.; Darst, S. A.; Campbell, E. A. Fidaxomicin Jams Mycobacterium Tuberculosis RNA Polymerase Motions Needed for Initiation via RbpA Contacts. *eLife* 7.

Interaction sites between RNAP and ∂ transcription factor as targets

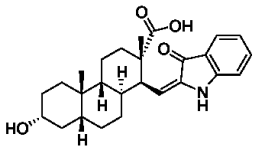
Transcription cycle is highly regulated by specific proteins called transcription factors that activate during normal cell growth or in response to environmental signals. For transcription initiation to begin, RNAP must be associated to a specific initiation factor: most bacteria have a primary σ factor which is responsible for the transcription of housekeeping genes (σ^{70} on Gram-negative and σ^{A} in Gram-positive bacteria), while other specific factors are needed in cases of environmental stimuli.

When the elongation stage begins, the initiation factor dissociates from RNAP and elongation factors became associated, regulating transcription activity, e.g., they can modify transcription rate or modulate the response to pause and termination signals. The transcription factors are essential for bacterial replication and they are usually not conserved in eukaryotes becoming good targets for antibiotic discovery.

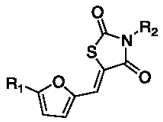
Transcription Factor Targets



GKL003



DSHS00507



SB series
R_1=*p*-PhCO$_2$H, *m*-PhCl, NO$_2$
R_2=H, Allyl,(CH$_2$)$_3$CO$_2$H

- **GKL series**

These compounds are characterized by an indole structure and they were discovered following a structure-based drug design approach: a virtual screening campaign of an in-house compound library was performed, using a pharmacophore model built on σ key residues. GKL003[35] was proven to bind the β' clamp-helix region of RNAP, inhibiting transcription reaction *in vitro*. Unfortunately, this molecule shows poor antibacterial activity, probably due to its poor aqueous solubility and its low capacity to cross bacterial membranes. No crystal structure of the complex with RNAP has been reported yet. Various series of GKL003 derivatives have been synthetized[36,37,38].

- **DSHS00507 or C5**[39]

[35] Ma, C.; Yang, X.; Kandemir, H.; Mielczarek, M.; Johnston, E. B.; Griffith, R.; Kumar, N.; Lewis, P. J. Inhibitors of Bacterial Transcription Initiation Complex Formation. *ACS Chemical Biology* **2013**, *8* (9), 1972–1980.

[36] Mielczarek, M.; Thomas, R. V.; Ma, C.; Kandemir, H.; Yang, X.; Bhadbhade, M.; Black, D. S.; Griffith, R.; Lewis, P. J.; Kumar, N. Synthesis and Biological Activity of Novel Mono-Indole and Mono-Benzofuran Inhibitors of Bacterial Transcription Initiation Complex Formation. *Bioorganic & Medicinal Chemistry* **2015**, *23* (8), 1763–1775.

[37] Mielczarek, M.; Devakaram, R. V.; Ma, C.; Yang, X.; Kandemir, H.; Purwono, B.; Black, D. S.; Griffith, R.; Lewis, P. J.; Kumar, N. Synthesis and Biological Activity of Novel Bis-Indole Inhibitors of Bacterial Transcription Initiation Complex Formation. *Org. Biomol. Chem.* **2014**, *12* (18), 2882–2894.

[38] Thach, O.; Mielczarek, M.; Ma, C.; Kutty, S. K.; Yang, X.; Black, D. S.; Griffith, R.; Lewis, P. J.; Kumar, N. From Indole to Pyrrole, Furan, Thiophene and Pyridine: Search for Novel Small Molecule Inhibitors of Bacterial Transcription Initiation Complex Formation. *Bioorganic & Medicinal Chemistry* **2016**, *24* (6), 1171–1182.

[39] Ma, C.; Yang, X.; Lewis, P. J. Bacterial Transcription Inhibitor of RNA Polymerase Holoenzyme Formation by Structure-Based Drug Design: From in Silico Screening to Validation. *ACS Infectious Diseases* **2016**, *2* (1), 39–46.

This compound was identified by an *in-silico* screening of the mini-Maybridge library choosing the compounds that fitted a pharmacophore model built on β' clamp-helix region (CH region), which is known to be in contact with initiation factor σ. The compound showed good inhibitory activity towards the β'-σ interaction specifically binding subunit β'. The compound could inhibit bacterial RNAP *in vitro* without affecting the eukaryotic enzyme and showed modest antibacterial activity on Gram-positive bacteria.

- **SB series**[40,41]

These compounds were discovered following an ELISA screening assay against *E. coli* holoenzyme formation. The binding site is unknown, but they are capable to inhibit σ^{70} binding to *E. coli* RNAP. The rhodanine structure of these compounds produced some doubts concerning the mechanism of action: they could act on multiple targets instead of acting as specific transcription inhibitors.

2.2 Aim of the work

Antibiotic resistance is becoming a major health problem worldwide, worsened by the fact that few compounds are in clinical development. Bacterial RNA polymerase is a strategic target in the search for new antimicrobial agents: as a matter of fact, the inhibition of this enzyme results in loss of gene transcription, which is essential for bacteria survival and replication. Rifamycins are the most known RNAP inhibitors: they act sterically hindering the catalytic site of the enzyme, blocking the transcription machinery. In 2011, FDA approved another drug that act on bacterial RNAP, called fidaxomicin, which is used to treat *Clostridium difficile* infections. This molecule acts in the so called "switch region" blocking RNAP in the closed and inactive conformation. In the last years, a peculiar protein-protein interaction, occurring between the clamp-helix region of subunit β' of bacterial RNAP and transcription factor σ, has been identified. Some molecules were discovered and characterized as disruptors of this interaction. Unfortunately, their potential activity as antibacterial agents was limited by their poor permeability in crossing bacterial membrane.
In this project, a virtual screening campaign, aiming at the identification of novel antimicrobial molecules, acting disrupting β'-σ interaction, was carried out. The newly identified compounds were experimentally tested by Prof. Barbara Montanini and her research group at the Laboratory of Biochemistry and Molecular Biology (Department of Chemistry, Life Sciences and Environmental Sustainability) at the University of Parma to confirm and characterized their activity. Modelling studies were then applied to rationalize their binding mode.

2.3 Results and discussion

[40] André, E.; Bastide, L.; Michaux-Charachon, S.; Gouby, A.; Villain-Guillot, P.; Latouche, J.; Bouchet, A.; Gualtiéri, M.; Leonetti, J.-P. Novel Synthetic Molecules Targeting the Bacterial RNA Polymerase Assembly. *Journal of Antimicrobial Chemotherapy* **2006**, *57* (2), 245–251.
[41] Villain-Guillot, P.; Gualtieri, M.; Bastide, L.; Roquet, F.; Martinez, J.; Amblard, M.; Pugniere, M.; Leonetti, J.-P. Structure−Activity Relationships of Phenyl-Furanyl-Rhodanines as Inhibitors of RNA Polymerase with Antibacterial Activity on Biofilms. *Journal of Medicinal Chemistry* **2007**, *50* (17), 4195–4204.

2.3.1 *Virtual screening campaign for the identification of β'-σ^{70} interaction inhibitors of bacterial RNA polymerase*

Virtual screening (VS) is a computational technique well established in the drug discovery field used to identify, among a molecule library, compounds which are potentially active against a specific target. This approach was chosen to select potential inhibitors of β'-σ^{70} interaction by screening a commercial library called 'Open Collection Scaffolds' of Compound Australia belonging to the Griffith University[42]. This library contains 33,999 molecules from ChemDiv and Enamine vendors. The aim of this procedure was to select about 5000 compounds and test these molecules in a yBRET screening assay, to increase the probability of finding active compounds compared to a random picking from the library. The compound library was initially filtered by molecular weight (MW) to remove small compounds which are reported to less likely inhibit protein-protein interactions[43,44] and retaining only compounds with MW≥ 250.

***Table 2**:summary of the number of compounds belonging to the "Open Collection Scaffold" library used for virtual screening. The number of starting structures of the two subsets, the number of retained structures after MW filtering and after the three-dimensional conversion are reported.*

Library	Starting structures	After MW filtering	2D → 3D conversion
ChemDiv	20000	19684	38886
Enamine	13999	13595	41985

The MW distribution of the filtered dataset was then analyzed and compared to the MW shown by the two β'-σ interaction inhibitors reported in literature, GKL003 and C5. Inhibitors GKL003 (compound **1**) and C5 (compound **2**) display a MW of 659.52 Da and 423.54 Da, respectively.

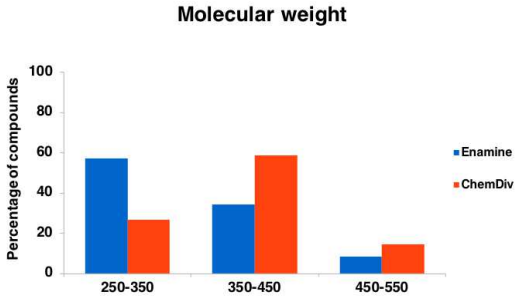


***Figure 8:** MW distribution of the two subsets composing the "Open Collection Scaffold" dataset.*

The analysis reported in Figure 8 revealed that more than a half of the Enamine subset has a MW below 350, while ChemDiv subset shows a greater MW. Overall, the database is composed by smaller compounds than the average MW reported for PPI inhibitors[35,36] and none of the compounds of the database shows a MW similar to the reference inhibitor GKL003.

[42] https://www.griffith.edu.au/griffith-sciences/compounds-australia/our-libraries
[43] Villoutreix, B. O.; Labbé, C.; Lagorce, D.; Laconde, G.; Sperandio, O. A Leap into the Chemical Space of Protein-Protein Interaction Inhibitors. 44.
[44] Sperandio, O.; Reynès, C. H.; Camproux, A.-C.; Villoutreix, B. O. Rationalizing the Chemical Space of Protein–Protein Interaction Inhibitors. *Drug Discovery Today* **2010**, *15* (5–6), 220–229.

The filtered library was then prepared to convert the 2D compounds in three-dimensional molecules with a suitable protonation state. Tautomeric forms as well as possible stereoisomers were built resulting in 808071 final structures. Each compound was then submitted to VS. The screening was performed using parallel approaches applying both ligand-based and structure-based selection methods according to the workflow reported in Figure 9. The decision to apply different selection methods was chosen to improve the possibility to retrieve active compounds given the very limited information on the structural elements important to disrupt β'-σ interaction.

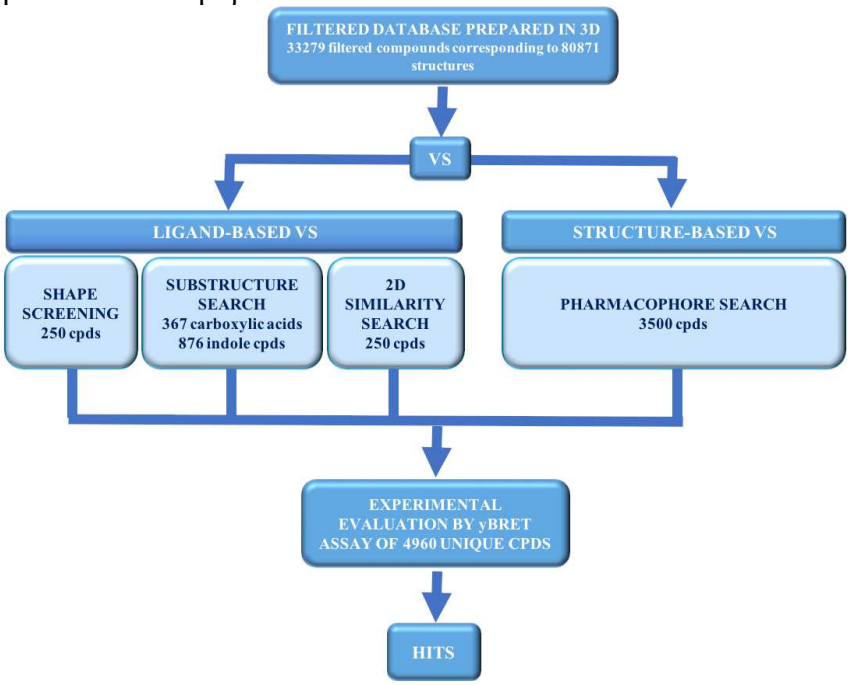


***Figure 9**:VS workflow describing the methodology used to identify potential β'-σ inhibitors.*

Structure-based pharmacophore built on σ^{70} subunit

The fitting of the compounds to a pharmacophore model built on region 2.2 of σ^{70} protein of *E. coli* RNAP (PDBID: 4YG2)[45] was chosen as structure-based selection method. The construction of the pharmacophore involved the selection of critical residues for β'-σ interaction identified through mutagenesis studies and thanks to the structural information retrieved by the RNAP holoenzyme crystal structure. The pharmacophore model used for virtual screening was composed by seven pharmacophore sites and it is reported in Figure 10a. Two negative sites were placed on the acidic groups of E407 and D403. Two hydrogen bond donor sites were placed on the two hydrogen atoms of the ammine group and an acceptor group was placed on the oxygen of the carbonyl side chain of Q406. Two hydrophobic sites were placed on the side chains of residues I410 and L402. Pharmacophore features placed on residues Q406, E407 and I410 were also part of a pharmacophore model reported by Lewis[46] which led to the identification of β'-σ inhibitor GKL003. Additional pharmacophore sites placed on σ^{70} residues Q406 and L402 were also included in our pharmacophore model since their side chains display peculiar interactions with β', as reported in the X-ray structure of *E. coli* (PDBID: 4YG2). The negative feature

[45] Murakami, K. S. X-ray Crystal Structure of Escherichia coli RNA Polymerase 70 Holoenzyme. *Journal of Biological Chemistry* **288,** 9126–9134 (2013).
[46] Ma, C. *et al.* Inhibitors of Bacterial Transcription Initiation Complex Formation. *ACS Chemical Biology* **8,** 1972–1980 (2013).

on the side chain of D403 was included since mutagenesis data on *E. coli*[47] confirmed that he substitution of this residue with alanine caused binding defects with β'subunit. Moreover, another difference is that the pharmacophore model published by Lewis[22] derived from a homology model of *B. subtilis* σ factor, while our pharmacophore model was built on an experimentally solved structure of *E. coli* (PDBID: 4YG2). Screened compounds were ranked according to Phase[48] scoring function, called *Fitness score.* This score measures how well functional groups of the screened compounds match with the pharmacophore sites of the hypothesis and how well the vector features (acceptor, donor and ring sites) superimpose with the one of the hypothesis.

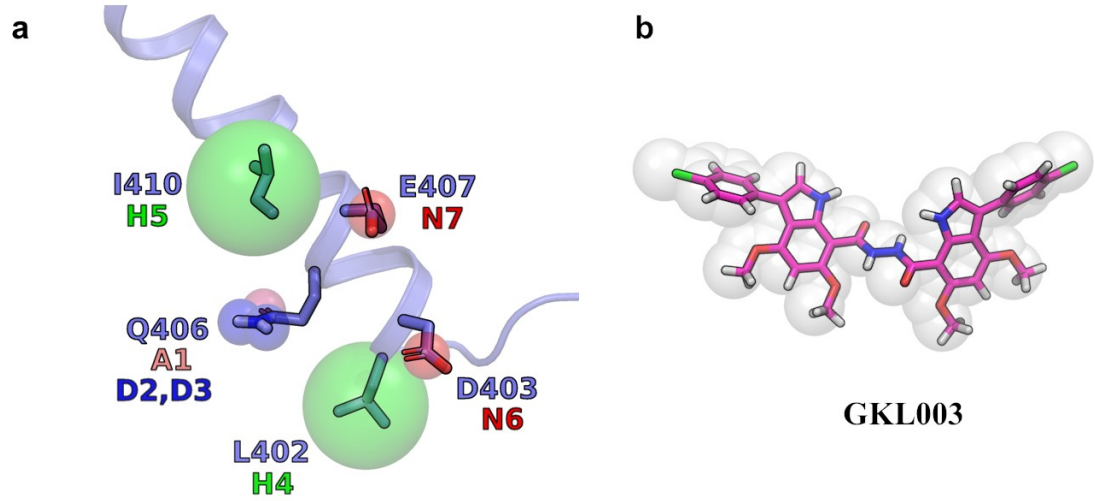


Figure 10: *structure-based pharmacophore model and reference inhibitor GKL003 conformation used for VS are represented. Pharmacophore was built on residues of σ^{70} subunit of E. coli RNAP (PDBID: 4YG2) reported to be critical for interaction with β'. Each pharmacophore site in the model is represented as a colored sphere: red spheres represent negatively charged groups, blue spheres represent donor groups, green spheres represent hydrophobic sites. The pink site represents a hydrogen bond acceptor group.* ***The dimension of the pharmacophore sphere is proportional to the feature tolerances used during pharmacophore screening*** ***(a).*** *The reference inhibitor GKL003 was used for shape screening in its extended conformation* ***(b).***

Ligand-based selection of compounds was performed using different computational screening approaches, i.e., like shape screening, 2D similarity and substructure search.

Shape screening

Shape screening was performed using compound **1** as query molecule. The 3D structure of the query was a low energy minimum in an extended conformation, as shown in Figure 10b. The first 250 compounds ordered by shape similarity were selected.

2D similarity

2D similarity was obtained by calculating linear fingerprints of reference compounds GKL003 and C5 and screened compounds, with Canvas v2.6[49]. The Tanimoto similarity was then used to rank the compounds. The similarity index spanned from 0.002 to 0.068 when compound C5 was taken as the reference compound, while it was comprised between 0.02 and 0.86 for compound GKL003 suggestive of a much greater similarity of the compounds in the database with the latter inhibitor. 250 compounds with highest similarity to compound GKL003 were therefore selected for experimental validation.

Substructure search

[47] Sharp, M. M. *et al.* The interface of ς with core RNA polymerase is extensive, conserved, and functionally specialized. *Genes & development* **13,** 3015–3026 (1999).
[48] Phase, version 4.5. Schrödinger, LLC, New York, NY (2015).
[49] Canvas, version 2.6. Schrödinger, LLC, New York, NY (2015).

As reported in section 2.1.5, acidic residues E166[50]/E407[14] of σ factor were reported as critical for the interaction with β' subunit by mutagenesis studies performed both on *B. subtilis* and *E. coli*. Moreover, reference inhibitor compound C5 contains a carboxylic group. Therefore, a substructure search was performed to isolate the 367 compounds carrying a carboxylic acid which were selected for yBRET screening. Moreover, given the presence of indole rings in compound GKL003 and other literature inhibitors of the same series (see section 2.1.6), the 876 indole compounds present in the Griffith library were also selected.
Combining the compounds selected by means of the different methods, we came up with 4960 compounds for in vitro yBRET assay.

2.3.2 Validation of yBRET screening platform

yBRET stays for yeast Bioluminescence Resonance Energy Transfer and it represents a well-established assay to identify protein-protein interaction inhibitors (PPII)[51]. This test is based on a naturally occurring phenomenon, called *resonance energy transfer*, that occurs between a luminescent donor, usually a luciferase, and a fluorescent acceptor, usually a fluorescent protein. The transfer of energy between the donor and acceptor depends on the distance between the two interacting proteins and their orientation: if the two proteins are in close contact a signal will be emitted by the fluorescent protein, otherwise if the two partners are distant no BRET signal will be emitted. The advantages of performing the assay in yeast cells is based on ***i)*** the possibility to evaluate the capacity of the compounds in crossing the cell membrane and also ***ii)*** to evaluate toxicity against eukaryotic cells.
In this test the two interacting proteins, σ^{70} (expressed as full-lenght sequence) and β' (expressed from residues 1 to 334) from *E. coli*, were linked to the luciferase donor protein NanoLuc and the acceptor Yellow Fluorescent Protein (YFP). To identify the combination with the highest BRET signals, the coding sequences of σ^{70} and β' have been cloned in-frame with the donor and the acceptor protein, respectively, in either C-terminal or N-terminal orientation. The system that gave the highest BRET signal was NLuc-σ^{70} and β'-YFP and was therefore chosen for the screening.
To validate the screening platform, the known β'- σ^{70} inhibitors GKL003 and C5 were tested in yBRET assay by Prof. Montanini and her research group, but the results were proven to be unreliable: as a matter of fact, GKL003 displayed significant solubility issues in the yeast medium, while C5 absorbed at the same wavelength of both NLuc and YFP spectra, making yBRET measures untrustworthy.The yBRET platform was thus checked using an in-house mini library of 27 indole derivatives: 10 mono-indoles and 17 bis-

[50] Johnston, E. B, The RNA Polymerase-sigma interaction as a target for potential novel antimicrobials, in School of Environmental and Life Sciences, University of Newcastle, Newcastle, Australia (2010).
[51] Corbel, C.; Sartini, S.; Levati, E.; Colas, P.; Maillet, L.; Couturier, C.; Montanini, B.; Bach, S. Screening for Protein-Protein Interaction Inhibitors Using a Bioluminescence Resonance Energy Transfer (BRET)-Based Assay in Yeast. *SLAS Discov* **2017**, *22* (6), 751–759.

indoles[52,53,54,55,56,57,58,59,60]. The choice of screening compounds with this structure came from the fact that containing-indole molecules had already proven their efficacy as inhibitors of β'- σ^{70} interaction[61,62]. These indoles were tested at 20 μM and the yBRET results are shown in Table 3.

[52] Di Giacomo, B.; Bedini, A.; Spadoni, G.; Tarzia, G.; Fraschini, F.; Pannacci, M.; Lucini, V. Synthesis and Biological Activity of New Melatonin Dimeric Derivatives. *Bioorganic & Medicinal Chemistry* **2007**, *15* (13), 4643–4650.

[53] Rivara, S.; Lorenzi, S.; Mor, M.; Plazzi, P. V.; Spadoni, G.; Bedini, A.; Tarzia, G. Analysis of Structure-Activity Relationships for MT2 Selective Antagonists by Melatonin MT1 and MT2 Receptor Models. *J. Med. Chem.* **2005**, *48* (12), 4049–4060.

[54] Chalaye-Mauger, H.; Denis, J.-N.; Averbuch-Pouchot, M.-T.; Vallée, Y. The Reactions of Nitrones with Indoles. *Tetrahedron* **2000**, *56* (5), 791–804.

[55] Mahboobi, S.; Teller, S.; Pongratz, H.; Hufsky, H.; Sellmer, A.; Botzki, A.; Uecker, A.; Beckers, T.; Baasner, S.; Schächtele, C.; et al. Bis(1H-2-Indolyl)Methanones as a Novel Class of Inhibitors of the Platelet-Derived Growth Factor Receptor Kinase. *J. Med. Chem.* **2002**, *45* (5), 1002–1018.

[56]Salucci, S.; Burattini, S.; Buontempo, F.; Orsini, E.; Furiassi, L.; Mari, M.; Lucarini, S.; Martelli, A. M.; Falcieri, E. Marine Bisindole Alkaloid: A Potential Apoptotic Inducer in Human Cancer Cells. *Eur J Histochem* **2018**, *62* (2), 2881.

[57] Spadoni, G.; Balsamini, C.; Bedini, A.; Diamantini, G.; Di Giacomo, B.; Tontini, A.; Tarzia, G.; Mor, M.; Plazzi, P. V.; Rivara, S.; et al. 2-[N-Acylamino(C1-C3)Alkyl]Indoles as MT1 Melatonin Receptor Partial Agonists, Antagonists, and Putative Inverse Agonists. *J. Med. Chem.* **1998**, *41* (19), 3624–3634.

[58] Mantenuto, S.; Cayuelas, A.; Favi, G.; Attanasi, O. A.; Mantellini, F.; Nájera, C.; Sansano, J. M. Reactivity of 1,2-Diaza-1,3-Dienes with Azomethine Ylides: [3+4] versus [3+2] Cycloadditions. *European Journal of Organic Chemistry* **2016**, *2016* (24), 4144–4151.

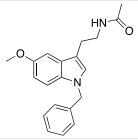
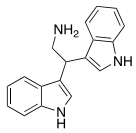
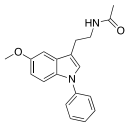
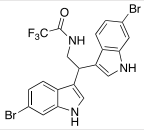
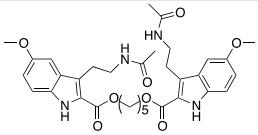
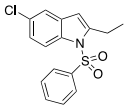
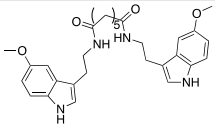
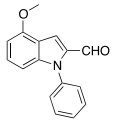
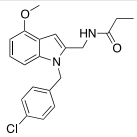
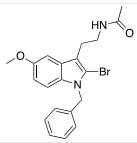
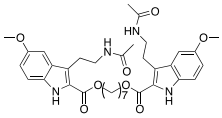
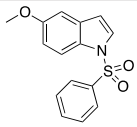
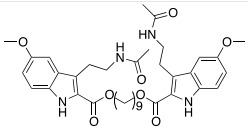
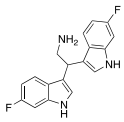
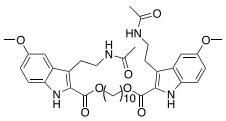
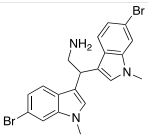
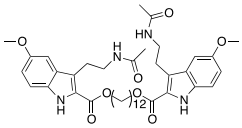
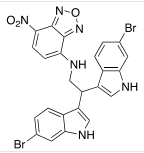
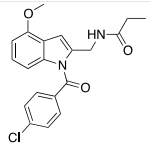
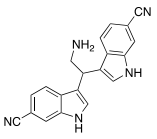
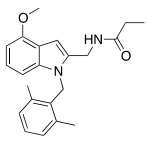
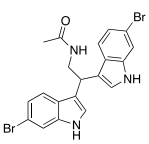
[59] Di Cesare, M. A.; Minetti, P.; Tarzia, G.; Spadoni, G. 5-Halo-tryptamine derivatives used as ligands of the 5-HT6 and/or 5-HT7 serotonin receptors: preparation, and therapeutic use. WO2003000252A1, 2003.

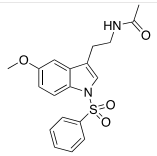
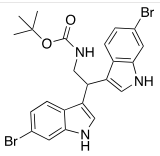
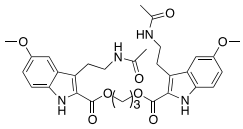
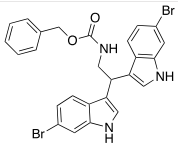
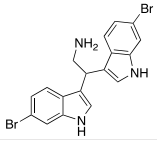
[60]Mari, M.; Tassoni, A.; Lucarini, S.; Fanelli, M.; Piersanti, G.; Spadoni, G. Brønsted Acid Catalyzed Bisindolization of α-Amido Acetals: Synthesis and Anticancer Activity of Bis(Indolyl)Ethanamino Derivatives. *European Journal of Organic Chemistry* **2014**, *2014* (18), 3822–3830.

[61] Ma, C.; Yang, X.; Kandemir, H.; Mielczarek, M.; Johnston, E. B.; Griffith, R.; Kumar, N.; Lewis, P. J. Inhibitors of Bacterial Transcription Initiation Complex Formation. *ACS Chemical Biology* **2013**, *8* (9), 1972–1980.

[62] Mielczarek, M.; Thomas, R. V.; Ma, C.; Kandemir, H.; Yang, X.; Bhadbhade, M.; Black, D. S.; Griffith, R.; Lewis, P. J.; Kumar, N. Synthesis and Biological Activity of Novel Mono-Indole and Mono-Benzofuran Inhibitors of Bacterial Transcription Initiation Complex Formation. *Bioorganic & Medicinal Chemistry* **2015**, *23* (8), 1763–1775.

Table 3: *yBRET results for the in-house indoles library used for platform validation. The percentage of yBRET inhibition is reported.*

Cmpd	Structure	% yBRET inhibition 20 µM	Cmpd	Structure	% yBRET inhibition 20 µM
3		0	**17**		3
4		0	**18**		1
5		33	**19**		2
6		4	**20**		17
7		48	**21**		16
8		30	**22**		8
9		42	**23**		3
10		40	**24**		4
11		37	**25**		0
12		3	**26**		0
13		5	**27**		0

14		5	**28**		0
15		0	**29**		0
16		1			

The most active compound of the series in the yBRET assay was compound **7**, followed by **9**, **10** and **11.** Compounds **12** and **13** are structurally similar to **7**, but they resulted inactive in the same assay. Compound **7** displays a structure which is characterized by an indole scaffold functionalized in position 1 with a *para*-chloro benzyl group, in position 2 with a propionylaminomethyl side chain and in position 4 with a methoxy group. The two inactive compounds **12** and **13** share the same structure of compound **7,** but they differ for the substituent in position 1: compound **12** displays a carbonyl linker attached to a *para*-chloro phenyl ring, while compound **13** shows a benzylic group with two methyl groups in *orto* position. These three compounds are structurally similar, but they showed a different activity in the yBRET assay and this could mean that compound **7** can specifically interact with the target. These results indicate that the yBRET system composed by NLuc-σ^{70} + β'-YFP strain is specific, interacting with compounds with a defined structure. In fact, even slightly differences in the chemical structure of the molecules (see for example compound **7** vs compounds **12** and **13**) resulted in different yBRET activity. The two mono-indoles **20** and **21**, were not further characterized due to their toxic effect on yeast cells, measured by a significant decrease in luciferase signal.

2.3.3 in vitro ELISA assay

An *in vitro* ELISA assay was also performed by Prof. Montanini and her research group to evaluate the capacity of the active compounds to disrupt β'- σ^{70} interaction at a molecular level. Molecules that displayed an inhibitory activity against the protein-protein interaction in yBRET, along with two inactive molecules of the same mini-library (the mono-indole **3** and the bis-indole **15**) and the two reference molecules **1** and **2** were tested at a concentration of 500 μM, to confirm ***i)*** their efficacy as β'- σ^{70} disruptors and ***ii)*** to validate the yBRET assay as a reliable drug discovery platform.

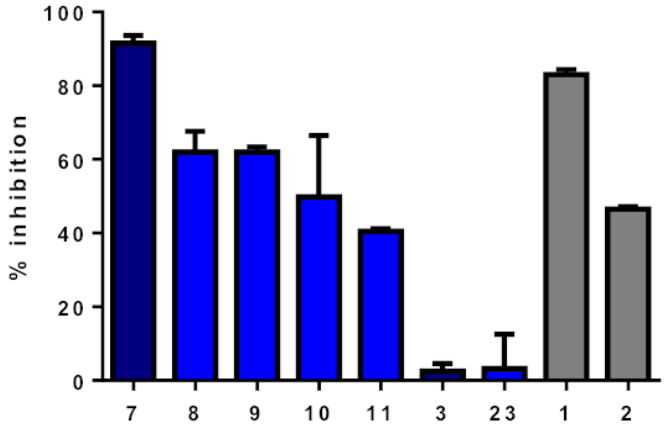


Figure 11: *in vitro ELISA assay performed for yBRET positive monoindoles (in blue) and, bis-indoles (in light blue) together with reference compounds **1** and **2** (in grey). Molecules **3** and **23** were tested as negative controls.*

As reported in Figure 11, the ELISA assay confirmed the results obtained in yBRET: active compounds found by yBRET were reconfirmed as actives together with the two reference compounds **1** and **2**. Inactive compounds **3** and **23** confirmed their inactivity in disrupting β'- σ^{70} interaction.
A dose-response ELISA assay was also performed to compare potency of compounds **7, 8** and the reference molecule **1**, as reported in Figure 12. Molecules from the indole in-house-library appeared to be three time less potent than the reference compound, with IC_{50} values of 43.8 μM (**7**) and 54.5 μM (**8**), while reference compound **1** showed an IC_{50} of 17.6 μM; the efficacy of **8** was lower (about 60%) than that of the other two molecules tested (about 80%).

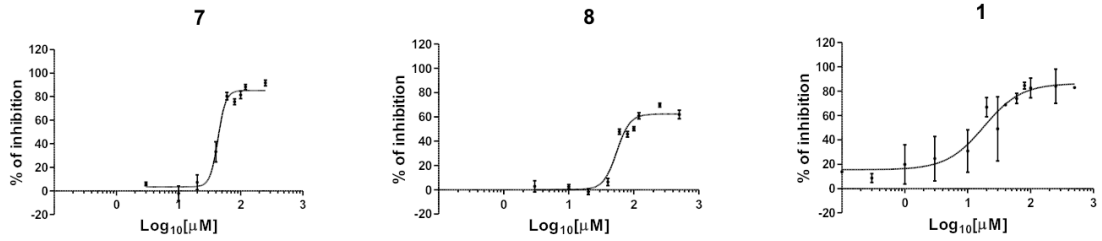
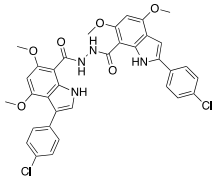
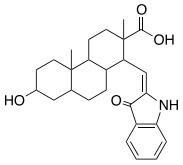
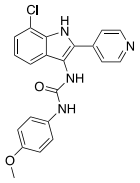
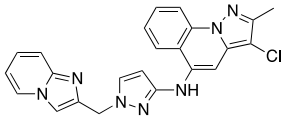
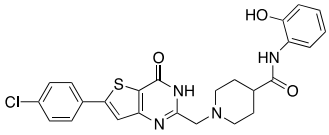
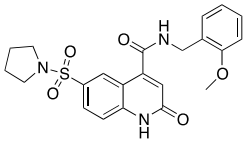
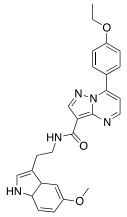
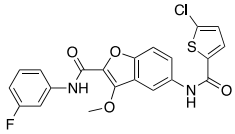
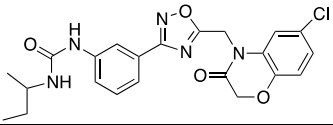


Figure 12**:** *a dose-response ELISA assay was performed to evaluate the inhibitory activity on β'-σ^{70} interaction of the active mono (**7**) and bis-indole (**8**) of the in-house mini library and the reference compound **1**. The percentage of β'-σ^{70} interaction inhibition referred to DMSO is reported (mean ±SD; error bars represent SD, n=3). The IC_{50} for each molecule was calculated using a non-linear regression curve fitting, as a measure of molecule potency. **7**: IC_{50} = 43.8 ± 1.0 μM; **8**: IC_{50} = 54.5 ± 1.1 μM; **1**: IC_{50} = 17.6 ± 1.3 μM.*

2.3.4 yBRET and ELISA assays on Griffith Library

After yBRET screening platform validation and the identification of new hit compounds from the in-house library of mono and bis-indoles, Prof. Montanini and her research group then tested the 4960 molecules selected by VS from the Griffith Library at 20 μM with the same yBRET system. Compounds able to reduce the BRET signal of at least 15% were re-tested at the same concentration and at 10 μM, and seven new hit molecules were identified (Table 4). The new molecules displayed diverse chemical structures, for both scaffold and for the nature and position of the substituents, and this result demonstrates that yBRET screening platform could be a useful and a reliable tool to identify new potential β'-σ^{70} interaction inhibitors characterized by novel structures.

Table 4: *the percentage of inhibition in yBRET and ELISA assay and the in vitro inhibitory activity (IC_{50}) of the hit compounds identified through VS are reported together with reference compounds GKL003 and C5. The "-" sign means that the compound was not tested.*

Cmpd	Structure	% yBRET inhibition 20 µM	%ELISA inhibition 250 µM	IC_{50} (µM)
1 GKL003		-	84±14	17.6±1.3
2 C5		-	46±1	-
30		28	87±4	16.7±1.1
31		21	99±4	11.4±1.1
32		28	93±1	11±1.1
33		16	82±6	91.3±1.2
34		18	-	-
35		23	-	-
36		22	-	-

The seven hits compounds were tested in an *in vitro* ELISA assay to validate their inhibitory activity at 250 μM and four of them (**30**, **31**, **32** and **33**) exhibited a certain efficacy in disrupting β'-σ^{70} interaction. Unfortunately, compounds **34, 35** and **36** resulted insoluble in the ELISA medium and were not further characterized. The inhibitory activity of the four soluble hits was assessed in a dose-response ELISA assay, highlighting that compound **33** was the least potent (IC_{50} = 91.3 μM), while the other three molecules were characterized by a potency similar to the reference GKL003 (IC_{50} spans from 11.0 μM and 16.7 μM) and by a slightly higher efficiency (98-100 % vs 87 %) (see Figure 13).

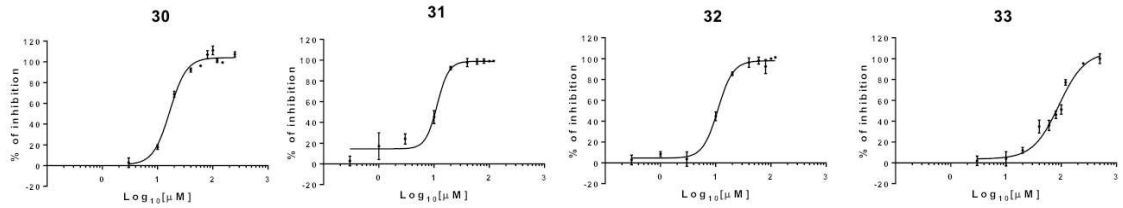
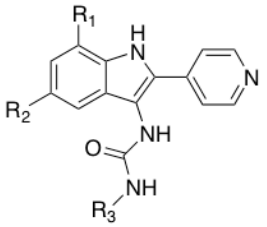


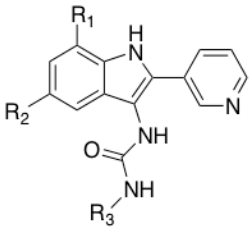
Figure 13: *dose-response ELISA assay for hit molecules **30**, **31**, **32**, and **33**. The percentage of β'-σ^{70} interaction inhibition referred to DMSO is reported (mean ±SD; error bars represent SD, n=3). The IC_{50} for each molecule was calculated using a non-linear regression curve fitting.*

Among the three most potent hits, it was decided to test some derivatives of the indol-3-yl-urea, compound **30**, to further explore this series (Table 5). Some commercially available analogs of compound **30** were tested in the ELISA assay to ***i)*** confirm the activity of the series, ***ii)*** to try to increase the potency and ***iii)*** obtain preliminary information on structure-activity relationships.

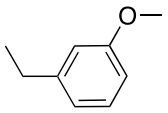
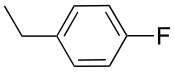
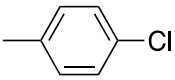
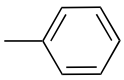
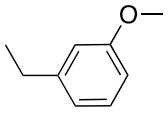
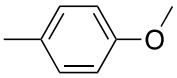
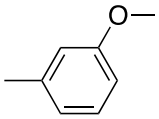
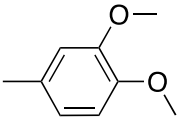
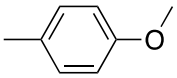
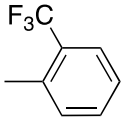
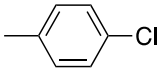
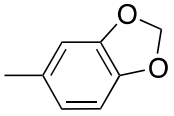
Table 5: *percentage of inhibition of β'-σ^{70} interaction in the ELISA assay and IC_{50} values of indol-3-yl urea commercial analogues of compound* ***30****. The "-" sign means that the compound was not tested.*



Cmpd 37-41



Cmpd 42-48

Cmpd	R_1	R_2	R_3	% ELISA inhibition 250 µM	IC_{50} (µM)
37	H	Cl		30±10	-
38	H	Cl		45±6	-
39	H	F		99±1	12.17±1.2
40	Cl	H		61±7	-
41	Cl	H		89±3	-
42	H	H		89±10	-
43	H	Cl		84±3	-
44	H	Cl		87±10	-
45	H	Cl		98±2	
46	H	Cl		98±2	20.47±1.0
47	H	F		98±2	8.35±1.1
48	H	F		98±2	-

Twelve analogues were tested in the ELISA assay at 250 µM and nine of them displayed an inhibitory activity above 80%. Collected data suggested that a phenyl substituent on the

amide linker is preferred to a benzyl one and that the phenyl ring tolerates lipophilic substituents. Position 2 of the indole ring can be substituted with a 3- or 4-pyridyl ring without affecting potency.
Three of these molecules were also tested in a dose-response ELISA assay to verify their potency: compound **39** was the most potent (IC_{50} = 8.3 μM) and all these molecules had high efficiency (96-100 %) (see Figure 14).

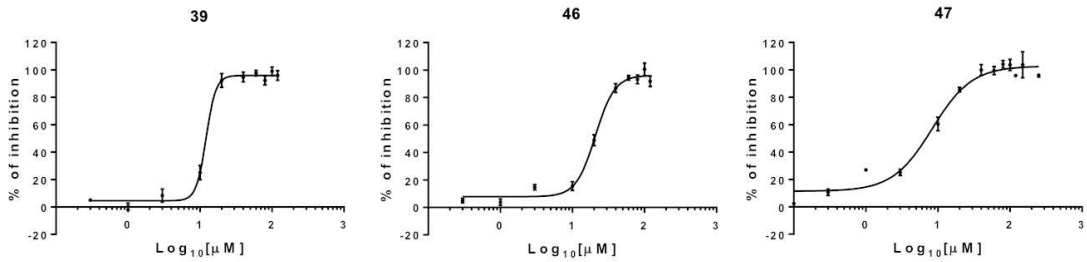


Figure 14: *a dose-response ELISA assay was performed to evaluate the inhibitory activity on β'-σ^{70} interaction of the hit molecules (**39**, **46** and **47**) at different compounds concentrations. The percentage of β'-σ^{70} interaction inhibition referred to DMSO is reported (mean ±SD; error bars represent SD, n=3). The IC_{50} for each molecule was calculated, as a measure of molecule potency, using a non-linear regression curve fitting.*

An ELISA assay was also performed to deduce whether selected hit molecules targeted subunit β' or σ^{70}. Seven of the active compounds identified in yBRET and confirmed in ELISA were incubated with σ^{70}, instead of β'subunit (protocol used for the previous, standard ELISA assays), prior to σ^{70}-β' complex formation. GKL003, known to target β' by calorimetric studies[63] was used as internal control. Results showed a significant decrease in the percentage of inhibition in comparison to the data obtained with β' incubation, suggesting that the selected compounds are most likely to bind β' (see Figure 15).

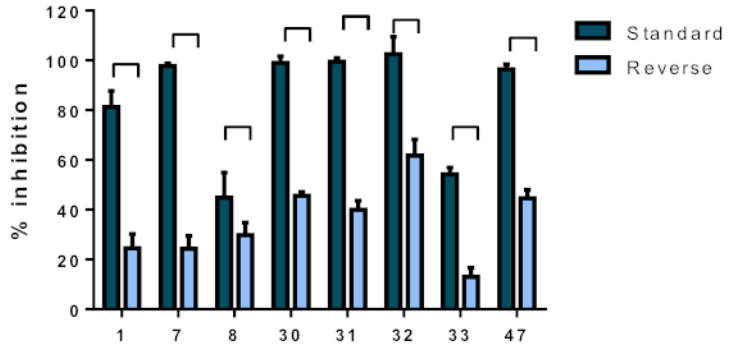


Figure 15:*ELISA binding assay suggested a preferential binding to β'subunit. Compounds were first incubated with His-tag-σ^{70} (cyan bars) or GST-β' (blue bars), before the complex formation. Compound final concentration was 100 μM. The percentage of interaction inhibition relative to DMSO is reported (mean ±SD; error bars represent SD, n=3).*

2.3.5 In vitro transcription inhibition assay

To verify whether these hit compounds, identified for their ability to inhibit β'-σ^{70} interaction, were also able to interfere with the RNAP holoenzyme assembly, mono-indole **7** and bis-indole **8** from the in house-mini library and the three characterized compounds

[63] Ma, C.; Yang, X.; Kandemir, H.; Mielczarek, M.; Johnston, E. B.; Griffith, R.; Kumar, N.; Lewis, P. J. Inhibitors of Bacterial Transcription Initiation Complex Formation. *ACS Chemical Biology* **2013**, *8* (9), 1972–1980.

from the indol-3-yl-urea class **39**, **46** and **47** were selected for the transcriptional assay. The *in vitro* transcription inhibition assay was initially performed with compound concentrations of 25 µM and 50 µM. The obtained results were compared with those of the best characterized inhibitor of β'- σ^{70} interaction, GKL003 and rifampicin, an approved drug used to treat tuberculosis that acts as inhibitor of RNAP active site. Except to molecule **8**, all the compounds caused an almost complete loss of transcription at 50 µM and only **46** and **47** retained a certain activity at 25 µM. A dose-response curve was obtained for compounds **46** and **47**, along with the reference **1**. Compounds **46** and **47** brought to a coherent sigmoid curve, with IC_{50} values close to 30 µM for both compounds, approximately three times higher than that of GKL003 (see Figure 16).

Compound	% inhibition at 25 µM	% inhibition at 50 µM
Rifampicin	97.2	100
1-GKL003	97.1	91.0
7	0.0	77.9
8	7.6	32.2
39	-	100
46	23.1	100
47	27.9	100

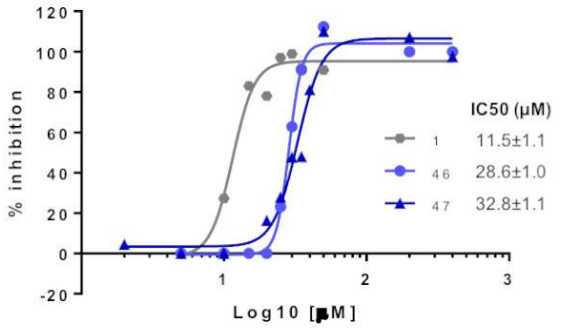


***Figure 16**:the table on the left reports the percentage of in vitro transcription inhibition. The assay was performed at two different concentrations (25 and 50 µM). The "–" sign means that the compound wasn't tested at that concentration. Transcription inhibition dose-response curves are reported on the right for reference compound **1**(in grey and spheres) and hit compounds **46** (in light blue and spheres) and **47** (in blue and triangles).*

2.3.6 Antimicrobial activity evaluation

Compounds **7**, **8**, **30**, **31**, **32**, **33**, **39**, **46** and **47**, which were characterized in ELISA assay, were tested for their antimicrobial activity by monitoring growth inhibition of non-pathogen bacteria. The experimental assays were carried out in *B. subtilis (*Gram-positive) and *E. coli* (Gram-negative), in the presence of the molecules previously dissolved in DMSO, while an equivalent quantity of DMSO was used as a negative control. Preliminary tests on *E. coli* strain showed that the activity of the compounds was increased by the addition of 2 µg/ml of the permeabilizing agent Polymixin B (PMBN), suggesting that the molecules cannot easily cross the Gram-negative bacterial outer membrane. To verify molecule's toxicity in eukaryotic cells, the growth of wild type W303 yeast strain in the presence of the compounds was also monitored. Bacterial growth was measured at 7 h, in the presence of molecules at 200 µM.

All compounds tested displayed greater inhibition of *E. coli* growth in the presence of PMBN as seen in preliminary tests. The most active molecules capable of inhibiting bacterial growth were compounds **39**, **46** and **47**. In particular, **47** showed a good activity in both bacterial strains, **46** displayed better activity against Gram-positive *B. subtilis* than on Gram-negative *E. coli,* while **39** was more active against *E. coli.* These three molecules were unable to inhibit *S. cerevisiae* cell growth.

Compounds **7**, **8**, **30**, **32** and **33** were unable to inhibit the growth of both strains or some of them were seen to mildly inhibit *E. coli* in presence of PMBN. **31** displayed an inhibitory activity on the yeast cells highlighting possible toxicity issues for eukaryotic organisms (see Table 6).

Table 6: *percentage of growth inhibition for different bacterial strains and yeast. The incubation time was of 7 hours, except for the percentage with "*" sign which were incubated for 20 hours.*

Compound	E. coli	E. coli + PMBN	B. subtilis	S. cerevisiae
7	2	24	0	-
8	0	0	0	-
30	19	27	12	0
31	0	40	22	33
32	0	0	0	-
33	0	0	0	-
39	36	81/42 *	59/33 *	0
46	28	46/45 *	99/50 *	5
47	36	100/60 *	88/38 *	0

Compounds **46** and **47** revealed to be the most active molecules capable to inhibit bacterial growth and they were tested at different concentrations every 30 min for 7 hours on both bacterial strains (see Figure 17). Compound **46** displayed a moderate inhibition against *E. coli* culture, while it showed a high inhibitory activity against *B. subtilis*, completely inhibiting bacterial growth down to 50 µM. The compound was also re-tested at lower concentration maintaining a low inhibitory activity at 25 µM. **47** was the most active compound against *E. coli*: as reported in Figure 17, it completely inhibited bacterial growth down to 100 µM and it maintained this ability at 50 µM. Compound **47** displayed a similar trend also against *B. subtilis*.

The antimicrobial activity of these molecules was also compared to the two reference compounds **1** and **2**, using the positive control rifampicin with compound concentration at 100 µM. Rifampicin was found to completely block bacterial growth at this concentration, as reported in literature, while **1** and **2** had no inhibitory activity on *E. coli* and only **2** showed a modest inhibition on *B. subtilis* growth (see Figure 17). These results confirmed that our newly identified compounds display higher potency compared to the reference σ-β' interaction disruptors compounds **1** and **2** in inhibiting bacterial strains.

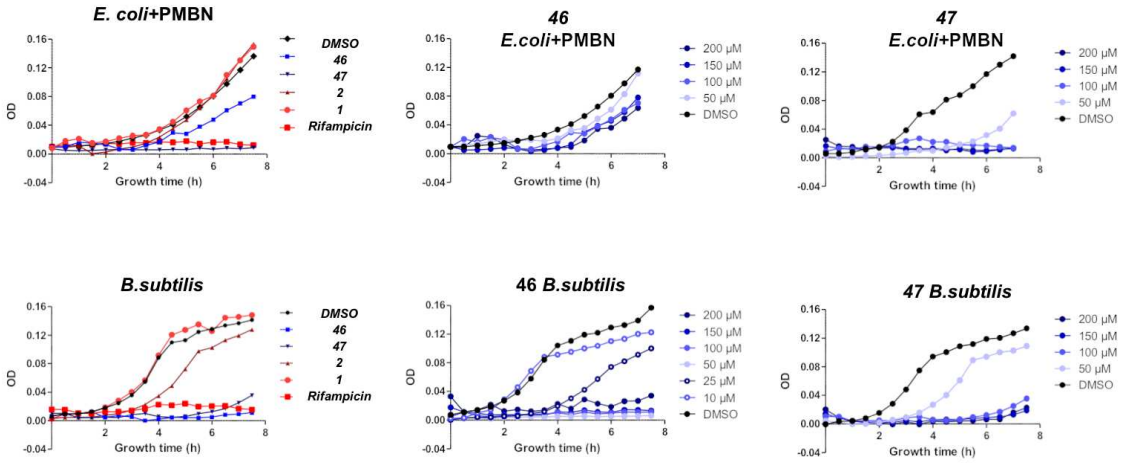


Figure 17: *growth inhibition assay of reference compounds* ***1, 2*** *and rifampicin and compounds* ***46*** *and* ***47*** *in B. subtilis and in E. coli+PMBN on the left. DMSO was used as control. Growth inhibition assay in the presence of increasing concentrations of* ***46*** *and* ***47*** *in B. subtilis and in E. coli+PMBN on the right.*

2.3.7 Ligand-based pharmacophore model

Active compounds **7**, **8**, **31**, **33** and **39** were representative of the structural variety of compounds able to inhibit β' and σ^{70} interaction found through this yBRET screening and together with reference compound **1**, were used to build a pharmacophore model, aiming at rationalize their activity against a common protein-protein interaction.
These compounds share the presence of aromatic portions variably combined with polar groups not ionizable at physiological pH. In fact, active compound **32** was excluded from the pharmacophore generation since it contains a basic functional group (a piperidine ring) which was not found in other compounds and which could lead to a different binding mode. The resulting pharmacophore model was characterized by 5 pharmacophore sites: two ring sites which accommodate cyclic portions of the compounds, a hydrophobic site, a hydrogen acceptor group and a donor group. In Figure 18 the pharmacophore model is displayed with spheres identifying the position of each site. All the six β' -σ^{70} inhibitors fulfilled the requirements of the pharmacophore model, with functional groups consistent with the nature of the pharmacophore sites and their spatial arrangement.

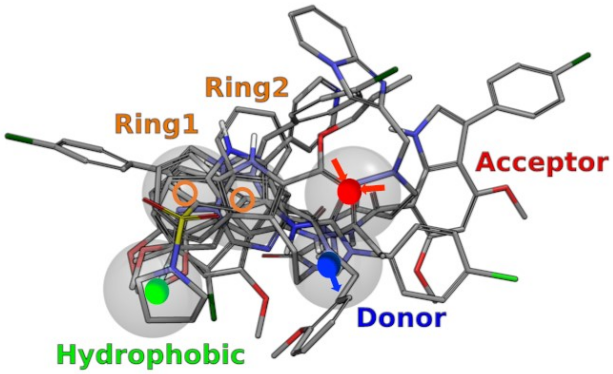


Figure 18: *ligand-based pharmacophore obtained by hit compound alignment is represented. The model is composed by a hydrophobic site (in green), two ring sites (in orange), one acceptor site (in red) and a donor site (in blue). Arrows in acceptor and donor sites represent the directionality of the features.*

Figure 18 reports reference compounds **1**, together with actives compounds **7**, **8**, **31**, **33** and **39** aligned to the selected ligand-based pharmacophore model. Figure 19 reports the fitting of each compound to the pharmacophore model. The indole ring of compound **7** matches the two ring sites and the methoxy group in position 4 occupies the hydrophobic site. The acetamidoethyl chain matches the donor and acceptor sites thanks to its amine and carbonyl groups, respectively. **8** displays a symmetric structure and only a half molecule was used to generate the pharmacophore model, hypothesizing that only one part could be responsible for the primary binding to the target protein. As for **7,** the indole nucleus occupies the two ring sites, and the methoxy group in 4 matches with the hydrophobic site. In this case the scaffold is functionalized with an acetamidoethyl chain in position 3 and only the NH group can match the donor site. The acceptor site is occupied by the carbonyl group of the ester linker.
Concerning the fitting of reference compound **1**, one indole ring occupies the two ring sites, the methoxy group is aligned onto the hydrophobic site while a NH group and a carbonyl group of the idrazide function occupy the acceptor and donor sites of the pharmacophore. In compound **31** the pyrazolo-quinazoline scaffold matches the two ring sites, the methyl group in position 2 is placed near the hydrophobic site, the amine group of the linker is placed onto the donor site, while the electronegative nitrogen atom of the pyrazole ring matches the acceptor site. Quinoline ring of **33** matches the two ring sites, the hydrophobic

site is occupied by the pyrrolidine ring, the acceptor and donor groups are occupied by the amide function. In compound **39** the indole scaffold matches the two ring sites, the fluorine atom in position 5 is accommodated in the hydrophobic site and the acceptor and donor groups correspond to the urea present in the side chain in position 3 of the indole ring.

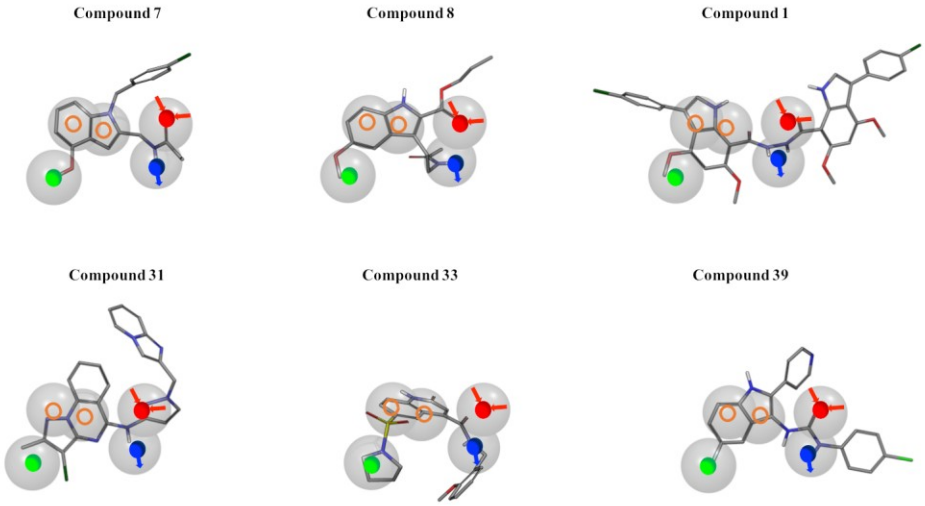


Figure 19: *reference compounds **1**, and active compounds **7, 8, 31, 33** and **39** are represented in their 3D conformation obtained by ligand-based pharmacophore search. Each compound is aligned on the pharmacophore points.*

2.3.8 Binding hypothesis

The next step of the work was the attempt to identify a common binding pose for the actives to the β' subunit. The first approach used was through standard automated docking of the compounds on the surface of β' clamp-helix region, but this approach did not provide useful results. Unfortunately, multiple solutions were identified for each ligand, with different orientations and amino acid counterparts. This is likely due to different reasons comprising 1) the flat shape of β' clamp-helix region in which no cavities or shaped structural motifs were identified that could accommodate the ligands, and 2) the crystal structure used for docking experiments in which β' was crystallized with σ^{70}protein. As a matter of fact, it can be hypothesized that binding of small molecules to the clamp-helix region could cause conformational changes of the receptor and, up to now, no structural information is available on the interaction of β' with an inhibitor.

The ligand-based pharmacophore model was then exploited to propose a binding hypothesis for the inhibitors to the β' subunit. We searched for a binding site on the surface of the β' subunit that could provide binding interactions for the pharmacophore elements previously identified for each compound, that could also allow a proper arrangement of the portions of each compound not comprised in the pharmacophore model.

The compounds share a polar region, in which a donor and an acceptor group are placed, and a hydrophobic region comprising hydrophobic/ring features. We identified residues E295 and N294 on the β' subunit as those geometrically suitable to act as donor and acceptor sites, respectively, for the corresponding acceptor and donor features of the inhibitors. These two residues had been analyzed by mutagenesis studies and mutation N283A in *B. subtilis* (corresponding to *E. coli* N294) reduced binding efficiency of about 50%[64], whereas mutation E295K in *E. coli* resulted in the inability of β'subunit to bind

[64] Ma, C.; Yang, X.; Kandemir, H.; Mielczarek, M.; Johnston, E. B.; Griffith, R.; Kumar, N.; Lewis, P. J. Inhibitors of Bacterial Transcription Initiation Complex Formation. *ACS Chemical Biology* **2013**, *8* (9), 1972–1980.

protein σ[65]. Moreover, at the top of the clamp helix region, not far from residues N294 and E295, a hydrophobic region could be identified as the portion capable of accommodating the ring features of inhibitors (Figure 20). Compounds were therefore manually docked on the surface of the β' protein in the same conformation as proposed by the pharmacophore model. Those portions of the compound structures not involved in pharmacophore definition were adapted to the protein counterpart. The complexes were then energy-minimized to remove steric clashes. Figure 20 reports the binding pose of the 2.2 region of the σ subunit as it is crystallized in the structure 4YG2 and the binding mode proposed for the most potent compound **39**. The inhibitor occupies the same region of the σ helix region 2.2, forming two H-bonds between its carbonyl group and the amide group of N294 and between its urea nitrogen and the acidic side chain of E295. The lipophilic 5-fluoro-indole portion is accommodated in a hydrophobic cavity on the top of β' subunit. The area occupied by σ factor in the *E. coli* crystal structure is now occupied by the inhibitor that can disrupt the protein-protein interaction and hamper RNA polymerase transcription.

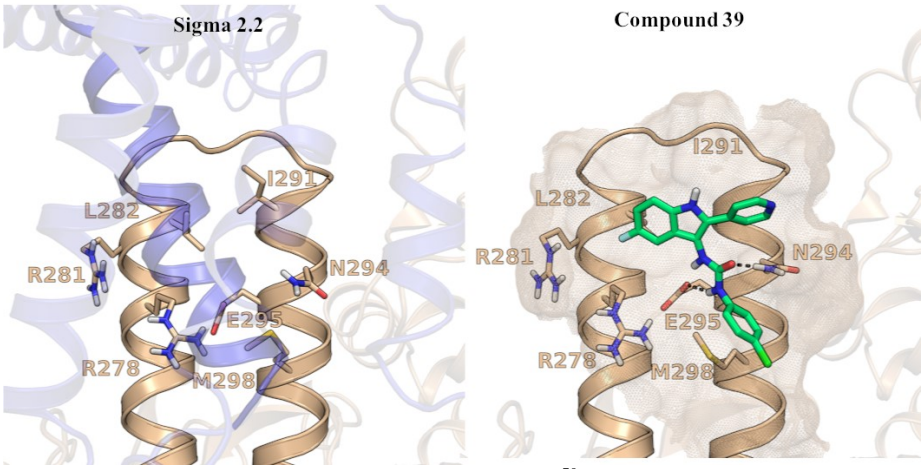


Figure 20: *crystal structure (4YG2) displays the helix domain 2.2 of σ[70] (on the left in blue cartoon) interacting with the clamp-helix region of subunit β'. The same region of subunit β' is occupied by the binding of hit compound **39** (on the right in green sticks. The target protein is shown as cartoon in light brown and polar interactions are depicted as black dashed lines.*

The other hit compounds share a similar binding pose to that of compound **39**, occupying the same region as σ factor and interacting with the same portion of β' subunit. Reference compound **1** has a large structure that accommodates on β' surface: the carbonyl of the idrazide group of the inhibitor forms a H-bond with the amide of N294, while the NH function of the same idrazide interacts with the acidic chain of E295. The pyrazolo-quinazoline scaffold of **31**, functionalized in position 2 by a methyl group and in position 3 by a chlorine atom, occupies the hydrophobic region of β' subunit. This compound undertakes H-bond interactions with the amide group of N294 via its pyrazole ring and with the acidic group of E295 via its amine linker. The quinolinone scaffold functionalized in position 6 by a pyrrolidin-sulfonyl group of **33** accommodates in the upper portion of β' subunit, while its amide linker makes H-bonds interactions with the acidic group of E295 and the amide of N294 of β' subunit.

[65] Arthur, T. M.; Anthony, L. C.; Burgess, R. R. Mutational Analysis of B′ $_{260–309}$, a ς ^{70} Binding Site Located on *Escherichia Coli* Core RNA Polymerase. *Journal of Biological Chemistry* **2000**, *275* (30), 23113–23119.

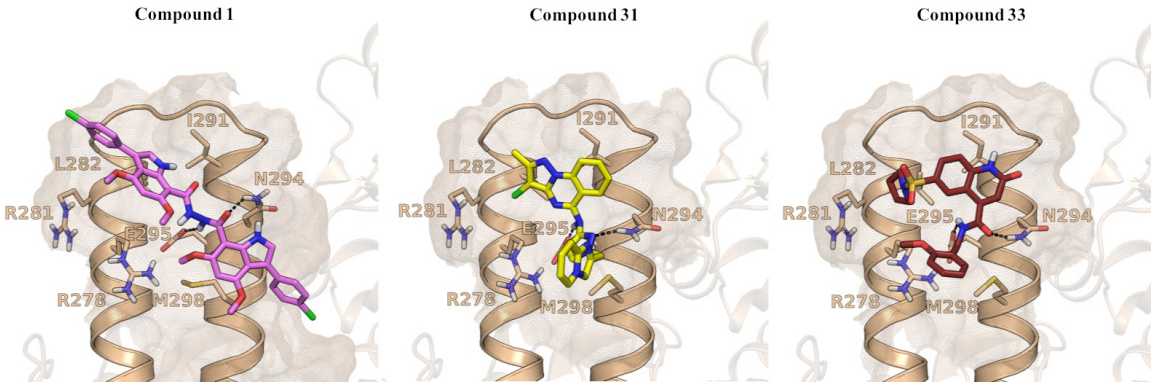


Figure 21: *binding pose of reference compound* ***1*** *(in purple) and hit compounds* ***31*** *(in yellow) and* ***33*** *(in brown) onto subunit β' are represented. The target protein is shown as cartoon in light brown and polar interactions are depicted as black dashed lines.*

As for the other inhibitors reported, the planar and rigid scaffold of compound **7** occupies the hydrophobic region of the protein with the chlorobenzyl portion exposed to the solvent. The amide group undertakes H-bond interactions with the acidic group of E295 and the NH group of N294. Bis-indole **8** is a symmetric compound characterized by two indole moieties decorated with an acetamidoethyl group in position 2 of the indole scaffold, connected through a seven-methylene linker. The first half of the molecule was manually docked to the target protein revealing the possibility to form H-bonds between the NH function of the acetamidoethyl group with the acid functional group of E295, while the carbonyl group of the ester linker could undertake a H-bond with the NH group of N294. The second half of the molecule was placed on the *β'* surface through automated docking (for the complete docking procedure see section 2.5.8) and the result obtained revealed the possibility to form a H-bond between the basic group of R297 and the carbonyl function of the acetamidoethyl group.

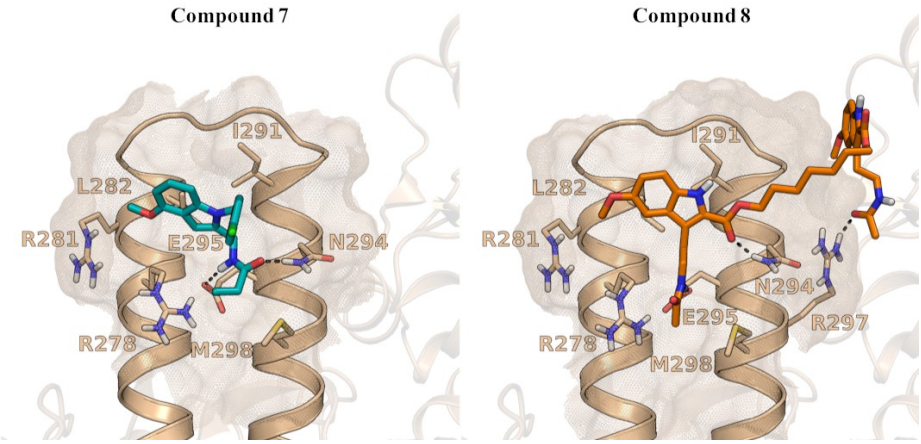


Figure 22: *binding poses of mono-indole 7 (in light blue) and bis-indole* ***8*** *(in orange) with β'subunit are represented. The target protein is shown as cartoon in light brown and polar interactions are depicted as black dashed lines.*

2.4 Conclusions

Today antibiotic resistance represents a challenging problem and many efforts are being made towards the development of new drugs active as antimicrobial agents without suffering resistance issues. Most current antibiotics inhibit the target by binding to its active site, while protein-protein interactions represent a less exploited target in the antibiotic field. In this work, a novel discovery pipeline that combines virtual screening, *in vivo*

(yeast) and *in vitro* assays resulted in the identification of new small molecules acting as inhibitors of β'-σ^{70} interaction of bacterial RNA polymerase.
Virtual screening was carried out on a commercial library characterized by small-molecules with lower molecular weight (MW<550) compared to classical PPI inhibitors. Multiple VS approaches were used to maximize the possibility of retrieving active compounds and the following experimental results confirmed the activity of 4 compounds (**30**, **31**, **32** and **33**) that showed an inhibitory activity >80% in the *in vitro* ELISA assay, displaying also comparable potency to reference inhibitor **1**. The hit compounds retrieved from the virtual screening, together with indoles **7** and **8** from the in-house mini library belong to different chemical classes, differing in the nature of the cyclic scaffold and in the nature and position of the substituents. However, it is possible to highlight some structural elements shared by many of the selected compounds. An indole ring is present in compounds **7**, **8** and **30** as well as in the most potent inhibitors of β'-σ^{70} interaction reported to date (i.e., **1** and its derivatives) and potential indole bioisoster can be seen in the compound **31**, which has one additional nitrogen atom in the six-membered ring of its indole-like structure. The majority of compounds have side chains characterized by the presence of polar groups not ionizable at physiological pH, mostly amide (**7**, **8**, **31** and **33**) and urea (**30**) groups.
Commercially available analogues of compound **30** were then retrieved and tested to confirm the activity of the indol-3-yl-urea class and to highlight some SAR. The *in vitro* ELISA results confirmed that compounds **39**, **46** and **47** were active in the low micromolar range and they were able to disrupt β'-σ^{70} interaction. A transcription inhibition assay was performed to confirm the capacity of inhibiting RNAP holoenzyme formation and only compounds **46** and **47** were capable to cause an almost complete loss of transcription also at 25 μM, with an IC_{50} close to 30 μM for both compounds, approximately three times higher than that of reference GKL003. Antimicrobial activity of the hit compounds was also evaluated on both Gram-negative and -positive bacteria, choosing *E. coli* and *B. subtilis* as non-pathogenic strains. Even if our compounds displayed some issues in crossing the outer membrane of Gram-negative *E. coli*, **46** and **47** showed to be the most active molecules capable to inhibit bacterial growth, displaying higher potency compared to the reference compounds **1** and **2** in inhibiting bacterial strains.
With the aim to rationalize the activity of the hit compounds against a common target site, compounds **7**, **8**, **31**, **33** and **39**, representative of the structural diversity of compounds able to inhibit β' and σ^{70} interaction together with reference compound **1**, were used to build a pharmacophore model. The ligand-based pharmacophore model obtained was characterized by five sites: two ring sites accommodating the rigid scaffold of the compounds, a hydrophobic site matching the different scaffold functionalizations and a donor and acceptor site that were occupied by the different polar groups.
Even though further tests are needed to verify the binding partner of these compounds (e.g. Surface Plasmon Resonance or X-rays), the molecules described in this work appear to preferentially bind β' over to σ^{70}, as suggested by ELISA binding assays performed on scaffold-representative molecules. A binding hypothesis was thus generated to identify a common binding pose on β' surface, based on the ligand-based pharmacophore conformations previously found for the hit compounds.
With this work new potential antimicrobials agents were identified, highlighting the antibiotic potential of new chemical classes of compounds. Even if VS was performed on a library of compounds that was not optimized for PPI inhibitors discovery, we were able to identify low micromolar compounds with comparable activity to the references reported in literature up to now. It has also to be pointed out that the biological results collected come from commercially available compounds that haven't been properly modified or

optimized. The binding hypothesis generated could be exploited for the design of derivatives with improved properties.

2.5 Computational Protocol

2.5.1 Database preparation

The virtual screening campaign was performed on the 'Open collection scaffolds' compound library of the Griffith University (Australia) containing 33,999 compounds[66]. The Ligfilter utility implemented in Maestro v10.4[67] was used to initially filter the database excluding compounds with molecular weight ≤ 250. The filtering yielded 33,279 compounds. LigPrep v3.6[68] was used to convert 2D structures in 3D ones. Protonation states of the compounds were generated at target pH 7.4±2 using Epik v3.4[69]. Chirality of asymmetric centers was assigned according to information available in the database: if no information was available, all possible stereoisomers were built. At the end of the procedure 80,871 3D structures were generated.

2.5.2 Protein preparation

X-ray structure of *E. coli* RNA polymerase holoenzyme was retrieved from Protein Databank (PDB: 4YG2)[70]. Subunit β' and σ were isolated and prepared using the Protein Preparation Wizard of the Schrödinger suite 2015-4[71]. Firstly, hydrogen atoms were added, and bond orders were assigned and checked. PROPKA[72] was then used to ***i)*** properly orient thiol and hydroxyl groups and to ***ii)*** adjust conformation of asparagine, glutamine and histidine residues to optimize the overall hydrogen bonding network. A first energy minimization of the complex, where only hydrogen atoms were free to move, was performed. A second minimization run was then performed, restraining the positions of the heavy atoms to an RMSD value of 0.3Å.

2.5.3 Structure-based pharmacophore model generation and library screening

Starting from the prepared protein chain σ, a customized pharmacophore model was generated using Phase v4.5[73] and then applied for virtual screening. Pharmacophore sites were manually chosen on the region 2.2 of σ^{70} subunit: a negative site was placed on the acidic group of E407, two hydrogen bond donor sites were placed on the two hydrogen atoms of the amine group and an acceptor group was placed on the oxygen of the carbonyl side chain of Q406. Two hydrophobic sites were placed on the side chains of residues I410 and L402. A negative site was placed on D403 carboxylic group.
A conformational search was performed using Confgen for each screened compound. Up to 1000 conformers per compound were generated and they were aligned on the query pharmacophore model and ranked according to the *Phase Fitness score*. Aromatic rings

[66] https://www.griffith.edu.au/science-aviation/compounds-australia/our-libraries
[67] Maestro, version 10.4. Schrödinger, LLC, New York, NY (2015).
[68] LigPrep, version 3.6. Schrödinger, LLC, New York, NY (2015).
[69] Epik, version 3.4. Schrödinger, LLC, New York, NY (2015).
[70] Murakami, K. S. X-ray Crystal Structure of Escherichia coli RNA Polymerase 70 Holoenzyme. *Journal of Biological Chemistry* **288,** 9126–9134 (2013).
[71] Protein Preparation Wizard 2015-4. Schrödinger LLC, New York, NY (2015).
[72] Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **7,** 525–537 (2011).
[73] Phase, version 4.5. Schrödinger, LLC, New York, NY (2015).

were allowed to match hydrophobic sites. Negative functional groups were allowed to match acceptor sites. Default weights were applied to the components of the Phase scoring function. Compounds were forced to match at least three pharmacophore sites and tolerances were set for each pharmacophore site: 1 Å tolerance was set for the acceptor, donor and negative sites to obtain a more precise alignment of functional groups onto these pharmacophore features. A 3 Å tolerance was set for the hydrophobic sites placed on residues I410 and L402 to facilitate the alignment of functional groups onto these sites since they were placed on the pharmacophore model extremity. Distances between pharmacophore features are reported in Table 7.

Table 7: *intersite distances between structure-based pharmacophore features built on region 2.2 of σ^{70} (PDB:4YG2)*

Site 1	Site 2	Distance (Å)	Site 1	Site 2	Distance (Å)
A1	D2	2.501	**D3**	H4	8.169
A1	D3	3.138	**D3**	H5	7.793
A1	H4	7.497	**D3**	N6	8.231
A1	H5	7.881	**D3**	N7	8.120
A1	N6	9.736	**H4**	H5	14.001
A1	N7	9.169	**H4**	N6	6.998
D2	D3	1.762	**H4**	N7	12.144
D2	H4	8.875	**H5**	N6	11.981
D2	H5	7.781	**H5**	N7	5.914
D2	N6	9.845	**N6**	N7	7.516
D2	N7	9.259			

2.5.4 Shape screening

Phase v4.2[74] was used to performed shape screening and GKL003 was chosen as reference compound. A conformational search was conducted using the mixed torsional/low-mode sampling approach (MCMM/LMOD) implemented in Macromodel v11[75], with default settings to obtain different conformations of the screened structures. The query molecule GKL003 was used as query in a minimum-energy extended conformation.
Confgen was chosen as engine to perform the conformational search on the ligand database, and up to 1000 conformers were generated for each screened structure. Shape screening was conducted using the *pharm* mode, thus each conformation is treated as a set of hard spheres that are flexibly superposed to the query and the similarity score is computed according to the matching of pharmacophore features between the query and the screened compounds. During compound alignment, acceptor pharmacophore sites, corresponding to acceptor functional groups of GKL003, could accommodate also anionic groups of screened structures. Compounds were ranked according to their *shape-similarity score*[76].

[74] Phase, version 4.2. Schrödinger, LLC, New York, NY (2011).
[75] Macromodel, version 11.0. Schrödinger, LLC, New York, NY (2015).
[76] Sastry, M.; Dixon, S. L.; Sherman, W. Rapid Shape-Based Ligand Alignment and Virtual Screening Method Based on Atom/Feature-Pair Similarities and Volume. *J. Chem. Inf. Model. 2011, 51(10), 2455-2466.*

2.5.5 Substructure search

Canvas v2.6[77] was used to carry out a substructure search to isolate compounds having a carboxylic acid or an indole ring.

2.5.6 2D similarity search

Canvas v2.6[22] was used to perform a 2D similarity search. Hashed linear fingerprints[78] were used to calculate the Tanimoto similarity index with reference inhibitors GKL003 and C5.

2.5.7 Ligand-based pharmacophore

The ligand-based pharmacophore model was generated using Phase v4.9[79]. The model included compounds **1**, **7**, **8**, **31**, **33** and **39**. The structure of the dimeric compound **8** was truncated three carbons after the ester group, hypothesizing that one half of the molecule was responsible for the primary binding to the target protein. A first conformational search for each compound was performed to choose the lowest energy conformers. The mixed torsional/low-mode sampling approach (MCMM/LMOD) implemented in Macromodel v11.4[80] was chosen using default settings. The conformers with the lowest potential energy where chosen as starting points for the pharmacophore model creation.
During pharmacophore search Confgen was used to generate 50 conformers for each input structure. Pharmacophore search was performed imposing that all the compounds had to be included in the model. A minimum of 4 pharmacophore features were required. Site tolerance was set to 2 Å. The option "minimize output conformers" was chosen to obtain energy minima for the final compounds. Twenty ligand-based pharmacophore hypotheses were generated by Phase and they were classified based on their *PhaseHypoScore*. Model including all the compounds had a maximum of five pharmacophore features and the model with best score was selected. The model showed a *PhaseHyposcore* of 0.81 and the *Fitness scores* for each compound used to generate the pharmacophore model are reported in Table 8.

Table 8: *Fitness scores values for each compound used to generate the pharmacophore model.*

Compound	*Fitness*
7	3.00
39	1.83
8	1.53
31	1.47
1	1.23
33	1.04

2.5.8 Binding hypothesis generation

The ligand conformations proposed by the ligand-based pharmacophore model were used as starting point to manually dock on the surface of β' clamp-helix region. Portions of each molecule not involved in the definition of the ligand-based pharmacophore model were

[77] Canvas, version 2.6. Schrödinger, LLC, New York, NY (2015).
[78] Duan, J., Dixon, S. L., Lowrie, J. F. & Sherman, W. Analysis and comparison of 2D fingerprints: Insights into database screening performance using eight fingerprint methods. *Journal of Molecular Graphics and Modelling* **29,** 157–170 (2010).
[79] Phase, version 4.9. Schrödinger. LLC, New York, NY (2016).
[80] Macromodel, version 11.4. Schrödinger, LLC, New York, NY (2016).

adapted to the protein surface. Macromodel v11-4[16] was used to minimize each complex using OPLS3[81]. Minimizations were carried out maintaining the backbone of the protein fixed, while residue side chains and ligands were free to move. Distance constraints were first applied between the oxygen atom of the acidic group of E295 and the donor atom of each ligand, and one hydrogen atom of the amine group of N294 and the acceptor atom of each compound.
Glide v7.3[82] software was used to optimize the position of **8**. This molecule is a symmetric compound characterized by two indole moieties decorated with an acetamidoethyl group in position 2 of the indole scaffold, connected through a seven-methylene linker. To optimize the position of the half moiety of the compound not part of the pharmacophore model, a docking run was performed using the SP scoring function. Glide grid was built on β' subunit of PDB 4YG2 and centered on residues N294 and E295. The dimension of the inner box was set to 15, 15, 20 Å, while the dimension of the outer box was set to 30.35, 30.35 and 35.35 Å, respectively. No Van der Waals scaling was applied to nonpolar protein atoms. Docking runs were performed using default settings and core constraints were applied to the indole-amide portion of compound **8** used to generate the pharmacophore model to maintain the native position of the indole scaffold of the compound. The docking run resulted in five docking poses showing different positions of the second indole moiety on β' surface. The fourth pose in terms of docking score was chosen since the carbonyl group of the amide of the second moiety displays an additional H-bond with residue R297.

[81] Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296.
[82] Glide, version 7.3. Schrödinger, LLC, New York, NY (2016).

3. Molecular modelling studies on NAPE-PLD

3.1 Introduction

3.1.1 Lipid signaling

Cell communication represents one of the most complex and specialized functions in biological systems. In multicellular organisms, these interactions are mediated either through cell to cell contacts or through transmitter molecules which are able to pass cell membranes. Intracellular compartments and organelles of a cell must also communicate with each other and for this reason chemical messengers are needed[83].
Among transmitters, lipids play important roles in signaling events since they are able to cross biological membranes. Lipids are involved in most cellular signaling cascades in a wide variety of tissues and cell types, playing an important role in the physiological homeostasis maintenance but also in pathological conditions. The source of lipids involved in signal transduction is mainly represented by components of cell membranes. The cell membrane is composed of a lipid bilayer with polar hydrophilic head groups that face the cytoplasmic and extracellular spaces and hydrophobic tails that face each other in the inner part of the bilayer. Phospholipids are the most abundant components of membrane lipids and they have a structural role as membrane constituents. Among this class, *glycerophospholipids* are the most represented class of lipids in membranes. A typical phosphoglyceride molecule consists of a hydrophobic tail composed of two fatty acyl chains esterified to two hydroxyl groups of a glycerol phosphate unit which represents the polar head of the lipid. A second class of membrane lipids is composed by *sphingolipids*. These molecules are derived from sphingosine, an amino alcohol with a long hydrocarbon chain and they contain a long-chain fatty acid attached to the sphingosine amino group. *Cholesterol* and its derivatives constitute a third important class of membrane lipids, the sterols. Basic structure of sterols is a four-ring hydrocarbon. Cholesterol, major steroidal constituent of animal tissues, has a hydroxyl substituent on one ring which makes it amphipathic[84].
As already mentioned, lipids play a key role in the structural organization of cell membranes but they can also act as bioactive molecules, capable of affecting cell functions, signaling and cell regulation. For example, the already mentioned sphingolipids, also known as ceramides, are best known for their structural role in skin, but recent experimental evidence has proven that they can affect cancer cells signaling[85]. Diacylglycerol (DAG) is a well-known lipid messenger which is essential in physiological pathways (role in activating PKC), while steroid hormones and eicosanoids are known to play a role in inflammation. Among all bioactive lipids, ethanolamides of long-chain fatty acids (FAEs) also known as N-acylethanolamines (NAEs) have attracted attention of researchers because of their wide variety of biological activities[86]. The most representative members of this class are *N*-palmitoylethanolamine (PEA), N-oleoylethanolamine (OEA) and N-arachidonoylethanolamine or anandamide (AEA).

---

[83] Alberts, B.; Johnson, A.; Lewis, J.; Raff, M.; Roberts, K.; Walter, P. General Principles of Cell Communication. *Molecular Biology of the Cell. 4th edition* **2002**.
[84] Harayama, T.; Riezman, H. Understanding the Diversity of Membrane Lipid Composition. *Nature Reviews Molecular Cell Biology* **2018**, *19* (5), 281–296.
[85] Morad, S. A. F.; Cabot, M. C. Ceramide-Orchestrated Signalling in Cancer Cells. *Nature Reviews Cancer* **2013**, *13* (1), 51–65.
[86] Wymann, M. P.; Schneiter, R. Lipid Signalling in Disease. *Nature Reviews Molecular Cell Biology* **2008**, *9* (2), 162–176.

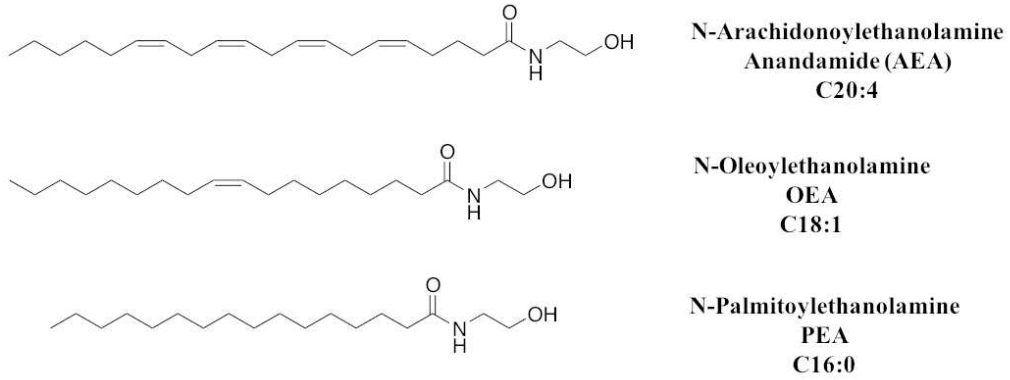


***Figure 23**:chemical structures of the most representative bioactive ethanolamides of long-chain fatty acids (FAEs) also known as N-acylethanolamines (NAEs).*

3.1.2 NAEs: a balance between synthesis and degradation

As mentioned in the previous section, NAEs act as lipid mediators and they are involved in several biological functions, including inflammation, regulation of food intake, analgesia and anxiety[87]. Their action depends on the binding to the respective effector protein but also on their production and degradation which regulate their concentration.

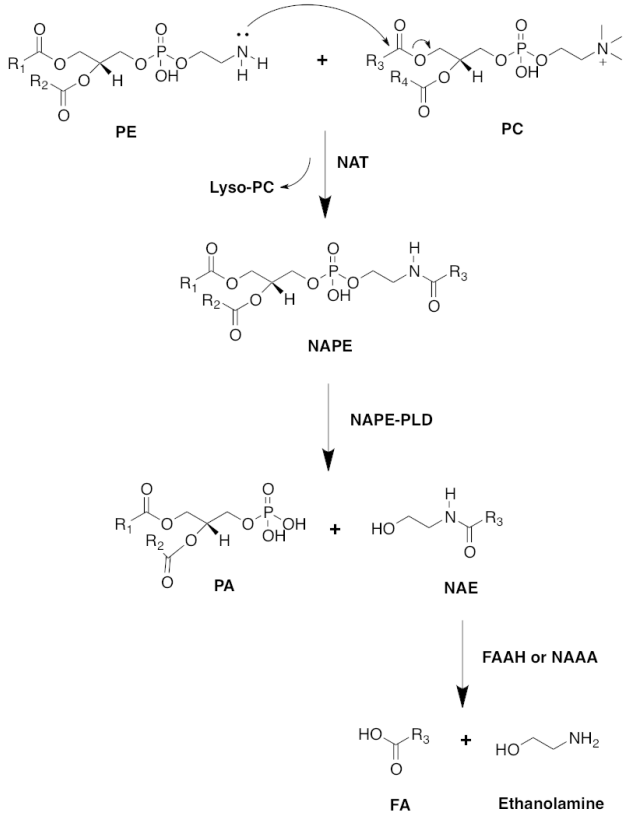


***Figure 24:** biosynthesis and degradation pathway of NAEs. N-acyltransferase enzyme (NAT) is responsible for NAPEs synthesis starting from phosphatidylethanolamine (PE) and phosphatidylcholine (PC). N- acylphophatidylethanolamine*

[87] Rahman, I. A. S.; Tsuboi, K.; Uyama, T.; Ueda, N. New Players in the Fatty Acyl Ethanolamide Metabolism. *Pharmacological Research* **2014**, *86*, 1–10.

phospholipase D (NAPE-PLD) hydrolyzes NAPE in NAEs producing phosphatidic acid. Fatty acid amide hydrolase (FAAH) and NAE acid amidase (NAAA) are responsible for degradation of NAEs in ethanolamine and fatty acids (FA).

The first step in NAEs synthesis is performed by N-acyltransferase enzyme (NAT). This calcium dependent enzyme catalyzes the transfer of a fatty acyl group from the sn-1 position of phosphatidylcholine to the amine group of phosphatidylethanolamine, producing a family of N-acylphosphatidylethanolamine (NAPE) species[88]. NAPEs are subsequently hydrolyzed by N-acyl phosphatidylethanolamine-hydrolyzing phospholipase D (NAPE), that cleaves the distal phosphodiester bond of NAPEs generating phosphatidic acid and different NAEs, depending on the nature of the substrate[89]. The structure and function of this enzyme will be discussed hereinafter. NAEs degradation is mainly mediated by two enzymes: N-acylethanolamine acid amidase (NAAA) and fatty acid amide hydrolase (FAAH) that regulate their concentration. FAAH[90] is a serine hydrolase highly expressed in CNS, liver and small intestine. Two isoforms of FAAH have been reported so far: FAAH-1 and FAAH-2, the last present only in placental mammals. This enzyme is an integral membrane protein. NAAA is a cysteine hydrolase particularly expressed in macrophages and PEA is the favorite substrate[91].

3.1.3 Anandamide and the endocannabinoid system

Anandamide was the first discovered endocannabinoid molecule in 1992[92] and, together with 2-arachidonoyl-*sn*-glycerol (2-AG), they act as agonists of the cannabinoid receptors. Endocannabinoids are hydrophobic neurotransmitters that are not stored in synaptic vesicles, but they are synthesized on-demand through the cleavage of membrane phospholipid precursors. They bind to the cannabinoid receptors CB_1 or CB_2 which are the most abundant GPCRs found in mammalian brain and coupled to G_i proteins. CB_1 receptors are expressed especially in the CNS where they mediate the effects of cannabinoid binding in the brain. They are localized in synapses, including glutamatergic and GABAergic neurons of different brain districts like hippocampus, basal ganglia, cerebellum and brainstem, which are the cerebral regions mainly involved in the behavioral effects of cannabinoid drugs[93].CB_2 receptors are localized mostly in peripheral tissues even if a moderate number is also present in the brain. These receptors are expressed on immune

[88] Schmid, H. H. O. Pathways and Mechanisms of N-Acylethanolamine Biosynthesis: Can Anandamide Be Generated Selectively? *Chemistry and Physics of Lipids* **2000**, *108* (1), 71–87.
[89] Okamoto, Y.; Morishita, J.; Tsuboi, K.; Tonai, T.; Ueda, N. Molecular Characterization of a Phospholipase D Generating Anandamide and Its Congeners. *J. Biol. Chem.* **2004**, *279* (7), 5298–5305.
[90] Cravatt, B. F.; Giang, D. K.; Mayfield, S. P.; Boger, D. L.; Lerner, R. A.; Gilula, N. B. Molecular Characterization of an Enzyme That Degrades Neuromodulatory Fatty-Acid Amides. *Nature* **1996**, *384* (6604), 83–87.
[91] Tsuboi, K.; Sun, Y.-X.; Okamoto, Y.; Araki, N.; Tonai, T.; Ueda, N. Molecular Characterization of N-Acylethanolamine-Hydrolyzing Acid Amidase, a Novel Member of the Choloylglycine Hydrolase Family with Structural and Functional Similarity to Acid Ceramidase. *J. Biol. Chem.* **2005**, *280* (12), 11082–11092.
[92] Devane, W. A.; Hanus, L.; Breuer, A.; Pertwee, R. G.; Stevenson, L. A.; Griffin, G.; Gibson, D.; Mandelbaum, A.; Etinger, A.; Mechoulam, R. Isolation and Structure of a Brain Constituent That Binds to the Cannabinoid Receptor. *Science* **1992**, *258* (5090), 1946–1949.
[93] Piomelli, D. The Molecular Logic of Endocannabinoid Signalling. *Nature Reviews Neuroscience* **2003**, *4* (11), 873–884.

cells and on activated microglia, suggesting their role in modulating the inflammatory response typical of neurodegenerative diseases[94,95].

Retrograde signaling[96]

Cannabinoid receptors can modulate neuronal activity regulating neurotransmitter release with a peculiar mechanism called retrograde signaling, in which anandamide and 2-AG are involved. Usually, after the release of neurotransmitters from the presynaptic cell, these molecules bind to post synaptic receptors causing an increase in intracellular Ca^{2+} levels. This mechanism triggers endocannabinoid synthesis that may rapidly diffuse in the extracellular space from the post synaptic cell acting to the presynaptic one in which cannabinoid receptors are located. The following activation of cannabinoid receptors CB_1 reduces the amount of neurotransmitters released, with the effects depending on the nature of the specific neurotransmitter.

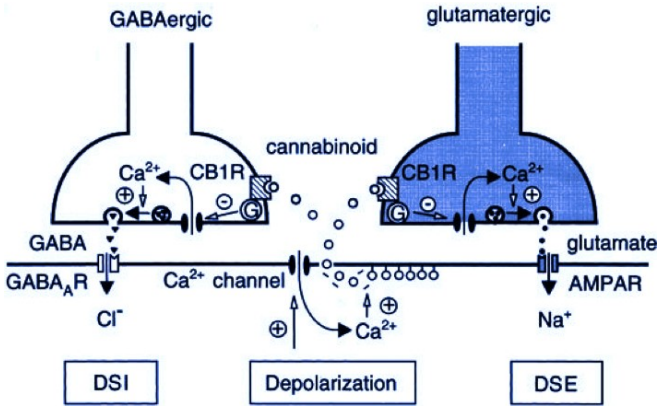


Figure 25: *a model for DSI and DSE retrograde signaling mechanisms. Upon release of neurotransmitter, postsynaptic receptors activate and trigger the mechanism which lead to Ca^{2+} influx. The postsynaptic depolarization stimulates endocannabinoid biosynthesis and their release. Endocannabinoids then migrate from postsynaptic neurons to CB_1 receptors located in presynaptic neurons suppressing the release of GABA (left) or glutamate (right) by inhibiting Ca^{2+} channels* (adapted by Kano, M.; Ohno-Shosaku, T.; Maejima, T. Retrograde Signaling at Central Synapses via Endogenous Cannabinoids. *Molecular Psychiatry* **2002**, *7*, 234–235).

As previously reported, cannabinoid receptors are mainly localized in glutamatergic and GABAergic neurons and retrograde signaling is thus characterized by two different phenomena:

- depolarization-induced suppression of inhibition (DSI) that target GABAergic neurons: if the release of the inhibitory neurotransmitter GABA is reduced the effect will be an increase of excitability of the endocannabinoid-releasing cell;
- depolarization-induced suppression of excitation (DSE) that target glutamatergic neurons: if the release of the excitatory neurotransmitter glutamate is reduced, the effect will be a decrease in the excitability of the post synaptic cell.

[94] Van Sickle, M. D.; Duncan, M.; Kingsley, P. J.; Mouihate, A.; Urbani, P.; Mackie, K.; Stella, N.; Makriyannis, A.; Piomelli, D.; Davison, J. S.; et al. Identification and Functional Characterization of Brainstem Cannabinoid CB2 Receptors. *Science* **2005**, *310* (5746), 329–332.

[95] Maresz, K.; Pryce, G.; Ponomarev, E. D.; Marsicano, G.; Croxford, J. L.; Shriver, L. P.; Ledent, C.; Cheng, X.; Carrier, E. J.; Mann, M. K.; et al. Direct Suppression of CNS Autoimmune Inflammation via the Cannabinoid Receptor CB1 on Neurons and CB2 on Autoreactive T Cells. *Nat. Med.* **2007**, *13* (4), 492–497.

[96] Kreitzer, A. C.; Regehr, W. G. Retrograde Signaling by Endocannabinoids. *Curr. Opin. Neurobiol.* **2002**, *12* (3), 324–330.

CB_1-mediated retrograde signaling inhibits GABA and glutamate release but it has been proven to affect also the neurotransmitter acetylcholine: in fact, *in vitro* and *in vivo* experiments suggested that stimulation of CB_1 receptor reduces acetylcholine release[97].

The endocannabinoid system mediates a broad range of physiological functions and modulating its activity could represent a therapeutic strategy against pain, anxiety, depression, obesity and neuronal disorders[98]. Since antiquity, cannabis extracts were used to reduce pain and several studies confirmed the analgesic properties of cannabinoid agonists. CB_1 receptors are located in brain regions implicated in the modulation of emotional response and it is known that cannabis can induce significant changes in mood and anxiety, increasing sociability and rewarding feelings. Cannabinoid agonists were also proven to be effective in nausea and increasing cannabinoid levels could be useful in the stimulation of appetite and food intake. On the other hand, rimonabant, a CB_1 antagonist now retired from the market, was used as anti-obesity drug. As previously reported, the endocannabinoid system plays a role in neurodegenerative diseases, like Huntington's disease[99], multiple sclerosis[100] and Alzheimer's disease[101], producing neuroprotective effects and hindering the progression of these pathologies.
Unfortunately, the modulation of cannabinoid receptors has revealed severe side effects that are caused by their direct activation. In particular, it has been proven that chronic administration of CB agonists can lead to tolerance, physical dependence[102], psychoactive effects[103,104] and impairment of motor recognition[105], raising doubts about their clinical development.

3.1.4 PEA and OEA as PPAR-α agonists

Despite anandamide, which is the endogenous binder of cannabinoid receptors, the other two NAEs representatives, PEA and OEA, act principally on the peroxisome proliferator

---

[97] Gifford, A. N.; Ashby, C. R. Electrically Evoked Acetylcholine Release from Hippocampal Slices Is Inhibited by the Cannabinoid Receptor Agonist, WIN 55212-2, and Is Potentiated by the Cannabinoid Antagonist, SR 141716A. *J. Pharmacol. Exp. Ther.* **1996**, *277* (3), 1431–1436.
[98] Wang, J.; Ueda, N. Biology of Endocannabinoid Synthesis System. *Prostaglandins & Other Lipid Mediators* **2009**, *89* (3), 112–119.
[99] Fernández-Ruiz, J.; Romero, J.; Ramos, J. A. Endocannabinoids and Neurodegenerative Disorders: Parkinson's Disease, Huntington's Chorea, Alzheimer's Disease, and Others. *Handb Exp Pharmacol* **2015**, *231*, 233–259.
[100] Migliore, M.; Pontis, S.; Fuentes de Arriba, A. L.; Realini, N.; Torrente, E.; Armirotti, A.; Romeo, E.; Di Martino, S.; Russo, D.; Pizzirani, D.; et al. Second-Generation Non-Covalent NAAA Inhibitors Are Protective in a Model of Multiple Sclerosis. *Angewandte Chemie International Edition* **2016**, *55* (37), 11193–11197.
[101] Bedse, G.; Romano, A.; Lavecchia, A. M.; Cassano, T.; Gaetani, S. The Role of Endocannabinoid Signaling in the Molecular Mechanisms of Neurodegeneration in Alzheimer's Disease. *J. Alzheimers Dis.* **2015**, *43* (4), 1115–1136.
[102] Desai, R. I.; Thakur, G. A.; Vemuri, V. K.; Bajaj, S.; Makriyannis, A.; Bergman, J. Analysis of Tolerance and Behavioral/Physical Dependence during Chronic CB1 Agonist Treatment: Effects of CB1 Agonists, Antagonists, and Noncannabinoid Drugs. *J Pharmacol Exp Ther* **2013**, *344* (2), 319–328.
[103] McLaughlin, P. J.; Thakur, G. A.; Vemuri, V. K.; McClure, E. D.; Brown, C. M.; Winston, K. M.; Wood, J. T.; Makriyannis, A.; Salamone, J. D. Behavioral Effects of the Novel Potent Cannabinoid CB1 Agonist AM 4054. *Pharmacol. Biochem. Behav.* **2013**, *109*, 16–22.
[104] Rubino, T.; Realini, N.; Castiglioni, C.; Guidali, C.; Viganó, D.; Marras, E.; Petrosino, S.; Perletti, G.; Maccarrone, M.; Di Marzo, V.; et al. Role in Anxiety Behavior of the Endocannabinoid System in the Prefrontal Cortex. *Cereb. Cortex* **2008**, *18* (6), 1292–1301.
[105] Hajós, M.; Hoffmann, W. E.; Kocsis, B. Activation of Cannabinoid-1 Receptors Disrupts Sensory Gating and Neuronal Oscillation: Relevance to Schizophrenia. *Biol. Psychiatry* **2008**, *63* (11), 1075–1083.

activated receptor α known as PPAR-α. This protein is a nuclear receptor together with PPAR-δ and PPAR-γ.

PEA as modulator of inflammation

N-palmitoylethanolamine (PEA) is an endogenous NAE synthetized on-demand starting from membrane phospholipids especially found in immune cells, neurons and microglia. PEA is the favorite substrate of N-acylethanolamine-hydrolyzing acid amidase (NAAA) and it is hydrolyzed especially in inflammatory status[106].

This lipid preferentially acts as an anti-inflammatory, neuroprotective and analgesic agent. In the past PEA was considered a CB_2 agonist, but studies carried out by Lo Verme[107] showed no effects of PEA in knockout CB_2 mice. PEA is now known to preferentially bind the nuclear transcription factor PPAR-α but it can act on other receptors, such as GPR5. It was also found that PEA could potentiate the effects of AEA on CB receptors, producing the so-called "entourage effect"[107,108]. This effect is mediated by PEA competitive inhibition of AEA hydrolysis on FAAH, leading to an increase in AEA levels that bind to CB_1 receptors. This effect was shown only in *in vitro* experiments, while PEA decreased levels of AEA *in vivo*.

The nuclear receptor PPAR-α is involved in transcription of genes involved in inflammation and pain and PEA was shown to directly activate this receptor with a half-maximal effective concentration (EC_{50}) of approximately 3 μM, sustaining it anti-inflammatory properties[107]. Reduction of PEA content was observed in inflammatory conditions in humans: for example, subjects with rheumatoid arthritis or osteoarthritis have low levels of PEA in the synovial fluid compared to healthy subjects[109]. Treatments with NAAA inhibitors seemed to block the reduction of PEA levels in macrophages stimulated by LPS, a bacterial endotoxin, suggesting that inhibitors of this PEA catabolizing enzyme could be useful as anti-inflammatory agents[110]. Concerning the neuroprotective activity of PEA, the molecule was found to counteract inflammation in an animal model of Alzheimer's Disease (AD), diminishing astrocytosis, an abnormal increase in astrocytes induced by beta amyloid peptides typically found in AD[111].

The inhibition of PEA degradation could be considered a therapeutic strategy to resolve inflammatory conditions and improve neurodegenerative diseases. NAAA is the enzyme that preferentially hydrolyzes PEA and different NAAA inhibitors are reported in pre-clinical stages of drug development. Another strategy that could increase PEA levels, not yet explored, is the activation of NAPE-PLD, the enzyme responsible for PEA synthesis.

---

[106] Mattace Raso, G.; Russo, R.; Calignano, A.; Meli, R. Palmitoylethanolamide in CNS Health and Disease. *Pharmacol. Res.* **2014**, *86*, 32–41.

[107] Lo Verme, J.; Fu, J.; Astarita, G.; La Rana, G.; Russo, R.; Calignano, A.; Piomelli, D. The Nuclear Receptor Peroxisome Proliferator-Activated Receptor-Alpha Mediates the Anti-Inflammatory Actions of Palmitoylethanolamide. *Mol. Pharmacol.* **2005**, *67* (1), 15–19.

[108] LoVerme, J.; La Rana, G.; Russo, R.; Calignano, A.; Piomelli, D. The Search for the Palmitoylethanolamide Receptor. *Life Sciences* **2005**, *77* (14), 1685–1698.

[109] Richardson, D.; Pearson, R. G.; Kurian, N.; Latif, M. L.; Garle, M. J.; Barrett, D. A.; Kendall, D. A.; Scammell, B. E.; Reeve, A. J.; Chapman, V. Characterisation of the Cannabinoid Receptor System in Synovial Tissue and Fluid in Patients with Osteoarthritis and Rheumatoid Arthritis. *Arthritis Res. Ther.* **2008**, *10* (2), R43.

[110] Pontis, S.; Ribeiro, A.; Sasso, O.; Piomelli, D. Macrophage-Derived Lipid Agonists of PPAR-α as Intrinsic Controllers of Inflammation. *Critical Reviews in Biochemistry and Molecular Biology* **2016**, *51* (1), 7–14.

[111] Scuderi, C.; Esposito, G.; Blasio, A.; Valenza, M.; Arietti, P.; Steardo, L.; Carnuccio, R.; De Filippis, D.; Petrosino, S.; Iuvone, T.; et al. Palmitoylethanolamide Counteracts Reactive Astrogliosis Induced by β-Amyloid Peptide. *J. Cell. Mol. Med.* **2011**, *15* (12), 2664–2674.

OEA and food intake[112]

Oleoylethanolamide (OEA) is a monounsaturated analogue of anandamide and, among all NAEs, is a key lipid mediator for the control of food intake, regulating feeding and body weight. This molecule is produced mainly in the epithelium of the small intestine upon fat adsorption. The exact mechanism by which OEA inhibits eating is still poorly understood but it seems to depend on the feeding state: in fact, levels of OEA in the small intestine are increased in the fed state, while they are decreased during a period of fasting. Different biochemical mechanisms that trigger this effect were reported: the enhanced production of NAPEs (OEA precursors), the activation of NAPE-PLD (the enzyme responsible for OEA production) and the inhibition of FAAH activity increase OEA levels. The pathway involved in OEA mobilization seems to start during dietary lipid digestion, where free oleic acid is generated in the gut lumen. The mucosal uptake of oleic acid in enterocytes is promoted by receptor CD36: its role in OEA mobilization is not clear, but it was proven that CD36-deleted mice were unable to produce food-induced OEA, highlighting its important role as a biosensor for oleic acids. After oleic acid internalization and CD36 activation, enterocytes will produce NAPEs and subsequently OEA will be released. CD36 could also regulate NAPE-PLD and FAAH activities, by engaging tyrosine kinases that are known effectors of the receptor.

The control of appetite is extremely important for the correct homeostasis of the organism and the gastrointestinal tract plays a critical role by releasing anorexigenic mediators such as peptide YY (PYY) and cholecystokinin (CCK). As a matter of fact, administration of OEA causes a long-lasting inhibition of food intake in rats and mice. The hypophagic action of OEA is different from the one caused by CCK or PYY and, as mentioned previously, seems to depend from the feeding state: if rats can have free access to their food, OEA decreases meal frequency without altering meal size, but in food deprived animals OEA acts reducing the two parameters. OEA seems to induce satiety thanks to its agonist activity on the nuclear receptor PPAR-α[113]. It was experimentally proven that OEA binds with high affinity to the purified ligand-binding domain of PPAR-α with a Kd~40 nM and stimulates the transcriptional activity of this nuclear receptor. Several lines of evidence suggested that PPAR-α receptor is responsible in mediating OEA anorexic action: ***i)*** OEA did not inhibit food intake in mice lacking PPAR-α ***ii)*** the hypophagic action of OEA was simulated by two PPAR-α agonists that were non-agonists of other PPAR subtype receptors, ***iii)*** OEA stimulated the expression of PPAR-α itself and CD36 **iv)** after feeding, levels of endogenous OEA were sufficient for PPAR-α activation and **v)** disruption of PPAR genes in mice abrogates fat-induced satiety response. In addition to PPAR-α, the capsaicin receptor TRPV1 (transient receptor potential cation channel vanilloid-1) seems to mediate OEA biological effects when it is previously phosphorylated by protein kinase C. OEA acts also as agonist of the cannabinoid receptor GPR119 (EC_{50} ~3 μM), particularly expressed in intestinal endocrine L-cells that secrete glucagon-like peptide-1 (GLP-1)[114].

Experimental evidences suggested that OEA mediates its anorexigenic effects by connecting gastrointestinal district and CNS through the sympathetic and parasympathetic

[112] Piomelli, D. A Fatty Gut Feeling. *Trends Endocrinol. Metab.* **2013**, *24* (7), 332–341.

[113] Fu, J.; Gaetani, S.; Oveisi, F.; Lo Verme, J.; Serrano, A.; Rodríguez de Fonseca, F.; Rosengarth, A.; Luecke, H.; Di Giacomo, B.; Tarzia, G.; et al. Oleylethanolamide Regulates Feeding and Body Weight through Activation of the Nuclear Receptor PPAR-α. *Nature* **2003**, *425* (6953), 90–93.

[114] Moss, C. E.; Glass, L. L.; Diakogiannaki, E.; Pais, R.; Lenaghan, C.; Smith, D. M.; Wedin, M.; Bohlooly-Y, M.; Gribble, F. M.; Reimann, F. Lipid Derivatives Activate GPR119 and Trigger GLP-1 Secretion in Primary Murine L-Cells. *Peptides* **2016**, *77*, 16–20.

nervous system with a complex mechanism that has to be fully elucidated yet[115]. OEA signaling is a key component for the monitoring of dietary fat intake and disfunctions of this system lies at the heart of obesity. OEA has also been reported to modulate cancer cell proliferation in leukemia[116]. Hence, selective modulation of OEA levels might provide us with new therapeutic strategies against different pathologies.
AEA, PEA and OEA exert different functions but they share in common the fact that NAPE-PLD is the enzyme involved in their synthesis.

3.1.5 NAPE-PLD synthetizes NAEs

NAPE-PLD is the enzyme responsible for the synthesis of NAEs. This enzyme belongs to the class of phospholipases D and catalyzes the hydrolysis reaction of the distal phosphodiester bond of specific membrane phospholipids called NAPEs (N-acyl-phosphatidylethanolamines). The resulting products are NAEs (depending on the nature of the substrate) and phosphatidic acid (PA) as reported in Figure 26.

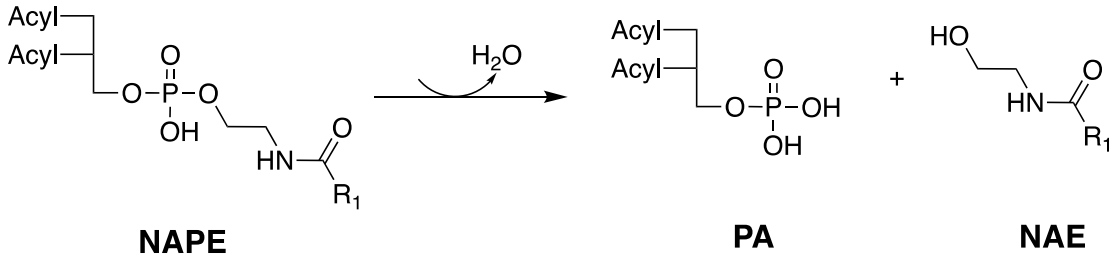


Figure 26: *the hydrolysis reaction catalyzed by NAPE-PLD. The membrane enzyme cleaves the distal phosphodiester bond of the NAPE substrate resulting in phosphatidic acid and N-acyletanolamine release.*

Phospholipase D (PLD) enzymes were firstly described in plants. After this discovery, other enzymes belonging to this class were identified in viruses, prokaryotic and eukaryotic organisms. Today the NCBI GenBank contains more than 4000 PLD enzymes that hydrolyze phosphodiester bonds, neutral lipids or polynucleotides. Two main families of PLDs are reported up to now: the largest PLD superfamily displays a conserved histidine-lysine-aspartate (HKD) motif in the active site that is responsible for the catalytic activity. Enzymes belonging to the second group are non-HDK PLD enzymes and they show different catalytic mechanisms and different structures compared to the HKD-PLD enzymes[117]. NAPE-PLD belongs to the non-HDK PLD family, as it is a metalloenzyme and a member of the β-lactamase family given its folding and the presence of a binuclear active site. Further experimental analysis revealed that this enzyme was associated to cell membranes. NAPE-PLD belongs to the class of lipase enzymes and these proteins are characterized by a peculiar phenomenon called "interfacial activation" which was reported

[115] Romano, A.; Tempesta, B.; Provensi, G.; Passani, M. B.; Gaetani, S. Central Mechanisms Mediating the Hypophagic Effects of Oleoylethanolamide and N-Acylphosphatidylethanolamines: Different Lipid Signals? *Front Pharmacol* **2015**, *6*.
[116] Masoodi, M.; Lee, E.; Eiden, M.; Bahlo, A.; Shi, Y.; Ceddia, R. B.; Baccei, C.; Prasit, P.; Spaner, D. E. A Role for Oleoylethanolamide in Chronic Lymphocytic Leukemia. *Leukemia* **2014**, *28* (7), 1381–138
[117] Selvy, P. E.; Lavieri, R. R.; Lindsley, C. W.; Brown, H. A. Phospholipase D: Enzymology, Functionality, and Chemical Modulation. *Chemical Reviews* **2011**, *111* (10), 6064–6119.

in literature for some member of the class[118,119,120]. This phenomenon suggests that lipases can exist in equilibrium between two distinct conformations: the "open" conformation which is compatible with the access to the active site by the substrate and the "closed" conformation in which the catalytic site is shielded and not accessible by the substrate. Since lipases are hydrophobic enzymes, the opening of the active site should be stabilized by the presence of the membrane environment (or organic solvents in *in vitro* experimental conditions), while the presence of water molecules should stabilize a closed conformation of their hydrophobic active site.

3.1.6 The crystal structure of NAPE-PLD

The first crystal structure of human NAPE-PLD was published in 2015 by Magotti and co-workers[121], giving insights into the 3D conformation of the protein and allowing to postulate some hypothesis about the protein machinery. The X-ray structure displays a resolution of 2.65Å and the sequence was expressed in *E. coli*. The protein was solved from residue 57 to residue 388 lacking an eleven amino acid loop (from residue 80 to 90).
Lithium sulfate and detergent sodium deoxycholate (DC) molecules were used to help the crystallization process and they are found in the published structure. The asymmetric unit shows the protein as a dimer and the two subunits are separated by an internal channel as reported in Figure 27. Each monomer displays a peculiar folding based on an αββα core which is typical of proteins belonging to the metallo-β-lactamase (MβL) superfamily. However, compared to this class of enzymes, NAPE-PLD is able to interact with membrane phospholipids. Loops L0 and L1 are rigid structures, thanks to their proline-rich sequences, and they mediate most of the subunit interactions. Together with helices α0 and α2, they form the internal channel at the dimer interface. This region is positively charged thanks to the presence of several basic residues such as lysine and arginine. Loop L1 displays a critical role in dimer assembly and catalysis. Experimental assays revealed that mutation of two of its residues, Q158S and Y159S, disrupts the ability of the monomers to dimerize and hydrolyze NAPEs, even without affecting the overall protein folding.

---

[118] Reis, P.; Holmberg, K.; Watzke, H.; Leser, M. E.; Miller, R. Lipases at Interfaces: A Review. *Adv Colloid Interface Sci* **2009**, *147–148*, 237–250

[119] Aloulou, A.; Rodriguez, J. A.; Fernandez, S.; van Oosterhout, D.; Puccinelli, D.; Carrière, F. Exploring the Specific Features of Interfacial Enzymology Based on Lipase Studies. *Biochim. Biophys. Acta* **2006**, *1761* (9), 995–1013.

[120] Scalvini, L.; Vacondio, F.; Bassi, M.; Pala, D.; Lodola, A.; Rivara, S.; Jung, K.-M.; Piomelli, D.; Mor, M. Free-Energy Studies Reveal a Possible Mechanism for Oxidation-Dependent Inhibition of MGL. *Scientific Reports* **2016**, *6*, 31046.

[121] Magotti, P.; Bauer, I.; Igarashi, M.; Babagoli, M.; Marotta, R.; Piomelli, D.; Garau, G. Structure of Human N -Acylphosphatidylethanolamine-Hydrolyzing Phospholipase D: Regulation of Fatty Acid Ethanolamide Biosynthesis by Bile Acids. *Structure* **2015**, *23* (3), 598–604.

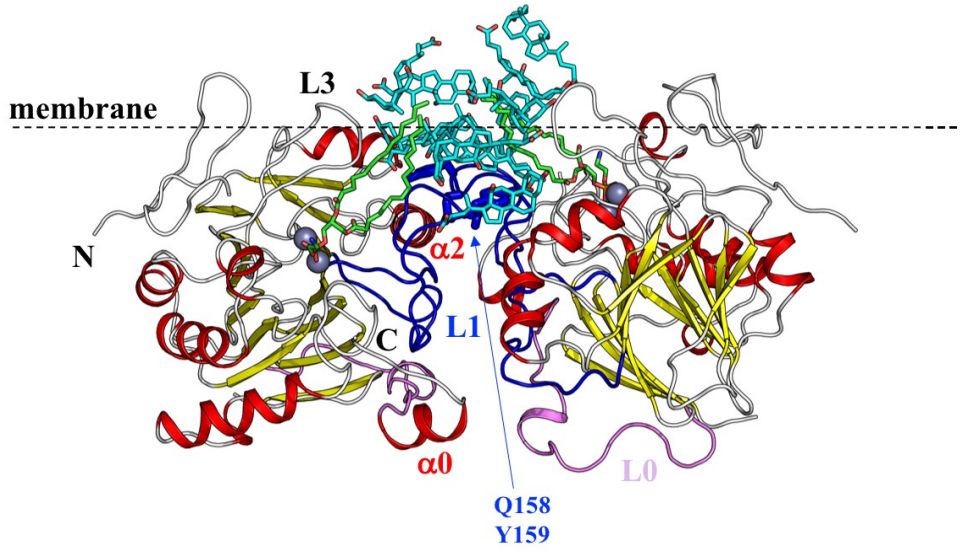


Figure 27: *crystal structure of human NAPE-PLD (PDBID: 4QN9). The protein is solved as a dimer and the αββα core folding of each monomer is displayed: beta-sheets are represented in yellow and alpha-helices in red. N and C terminal of one monomer are marked. PE inhibitor is represented as green sticks, bile acid molecules are represented in cyan and zinc atoms in the active site as grey spheres. The two subunits interact thanks to loop 0 and loop 1 which are represented in violet and blue, respectively. Residues Q158 and Y159 of loop L1 are represented as blue sticks.*

The active site of the enzyme is characterized by the presence of a molecule of phosphatidyletanolamine (PE) which is structurally similar to the substrate of NAPE-PLD but it's in fact an inhibitor derived from the membrane of *E. coli,* which was used to produce the recombinant protein for crystallization. PE was reported to inhibit detergent-stimulated NAPE-PLD activity with an $IC_{50} \approx 30$ μM. The pose of the crystallized inhibitor inside its binding site could suggest the binding mode of NAPE-PLD substrates. Moreover, the long fatty chains of the inhibitor protrude towards a region which is embedded into the membrane, the so called "hydrophobic hook".

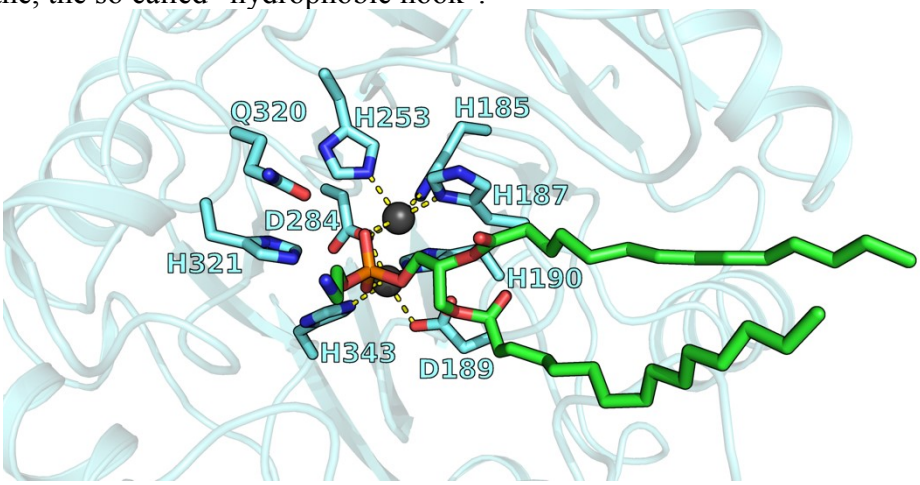


Figure 28: *active site of NAPE-PLD (PDB:4QN9). The two zinc atoms of the active site are coordinated by protein residues H132, H185, H187, D189, H190, H253, D284 and H343 and the two oxygen atoms of the PE inhibitor. Coordination bonds are displayed in yellow.*

Two zinc atoms are located in proximity of the polar head of PE and they are coordinated by a specific pattern of amino acids, like aspartic acid and histidine residues. In particular, residues involved in zinc coordination are H132, H185, H187, D189, H190, H253, D284 and H343. The authors of the paper suggest that the coordination of the two zinc atoms is trigonal bipyramidal. Other residues involved in the stabilization of the active site are H321

that forms an H-bond with the ethanolamine oxygen of PE and the adjacent residue Q320 which can bind the carboxamide oxygen of NAPE.
Loop L1 together with loop L3 define the hydrophobic surface in which PE is accommodated and, in proximity of the membrane, the hydrophobic hook is broadened to act ***i)*** as entry point for NAPEs into the active site and ***ii)*** as exit path for products of NAPE hydrolysis towards the membrane. The active site of the enzyme is large enough to accommodate different NAPEs with different N-acyl substituents.

Hydrolysis reaction: proposed mechanism

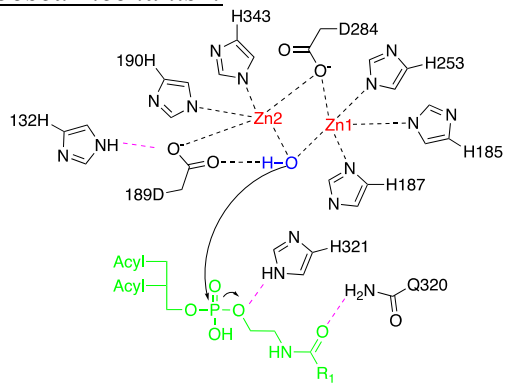


Figure 29: *proposed mechanism of NAPE-PLD hydrolysis reaction. The two zinc atoms in the active sites represented in red are coordinated by histidine and aspartic acid residues (coordination bonds are represented as black dashed lines). A water molecule, represented in blue, is necessary for the reaction to take place and even if it was not crystallized we can hypothesize that is placed near the zinc atoms. The enzyme substrate is represented in green and its position in the active site is stabilized by H-bonds with Q320 and H321 represented in magenta dashed lines. Residue H132 helps the stabilization of D189 through a H-bond.*

A proposed reaction mechanism is represented in Figure 29. Residue D189 is reported to be responsible for the activation of a water molecule of the active site. NAPE substrate is anchored to the active site thanks to hydrogen bonds with Q320 and H321. Nucleophilic attack is performed by the activated water molecule to the distal bond of the phosphate group that results in the release of NAE as leaving group, which is protonated by H321.

3.1.7 Role of bile acids in NAPE-PLD modulation

The crystal structure of NAPE-PLD contains some molecules of deoxycholic acid (DC) used to help the crystallization process. The bile acids molecules are grouped at the top of the hydrophobic hook, in the same region that is supposed to be occupied by the cell membrane. These molecules form a primary micelle and they expose their hydrophobic scaffolds to the membrane interface while their carboxylic acid groups make polar interactions with some protein residues as Y159, W218 and R257. The stability of the bile acids micelle is also reinforced by a dense network of water-mediated hydrogen bonds thanks to the hydroxyl groups of the steroid rings. These interactions are reported in Figure 30. Hydrophobic interactions are also made between bile acids and the tail end of the fatty acyl chains of PE. These specific and symmetric interactions between each bile acid molecule and residues R257 and Y159 might suggest that bile acid molecules could stabilize the conformation of loops L1 and L3, which have a critical role in dimerization. Moreover, another pair of DC molecules is reported forming a Van der Waals contact with helix α3 of one subunit and with M160 of the other subunit. Taken together, the structural information revealed by the crystal structure suggests a precise role for bile acids, that can mediate the anchoring of the enzyme to the membrane affecting dimer assembly and NAPE

admission to the active site. Magotti and co-workers[122] found that NAPE-PLD mutant R257A, which lacks a fundamental residue for DC binding, has a lower melting temperature (T_m = 71°C) than wild-type enzyme (T_m = 77°C) which is comparable to the Q158/Y159 mutant (T_m = 70°C), highlighting the important role of bile acids in favoring enzyme dimerization and stabilization.

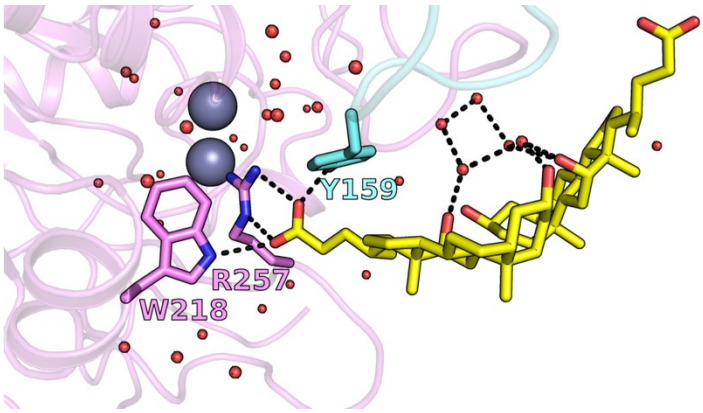


Figure 30: *hydrogen-bonds between deoxycholate (DC in yellow sticks) molecules and amino acids at the dimer interface. Waters molecules are indicated as red spheres. Chain A of NAPE-PLD dimer is represented in cyan and chain B in violet.*

Moreover, they discovered that DC could regulate enzyme's activity: DC bound NAPE-PLD with a K_D of 38 μM in a rapid and reversible manner resulting in enzyme's activation (EC_{50} = 186 μM). These results were not found for mutants R257A and Q158S/Y159S form of the enzyme. DC molecules were also capable of inhibiting detergent stimulated NAPE-PLD activity (Triton X-100, 0.1%) with an IC_{50} of 8.8 mM. This concentration could be reached *in vivo,* suggesting that bile acids could modulate NAPE-PLD activity in a concentration-dependent manner. In a following work by the same research group[123], the allosteric modulation of NAPE-PLD by bile acids was further investigated, highlighting the structural determinants of bile acids responsible for enzyme recognition and activation. For example, deoxycholic acid (DCA) was able to activate NAPE-PLD, while the secondary bile acid lithocholic acid (LCA) inhibited enzyme activation in the low micromolar range, suggesting a crucial role of the hydroxyl group in position 7 of the steroidal ring.
Bile acids binding seem also to promote the accommodation of NAPE substrates in the active site of the enzyme, favoring catalysis. The complex bile acids-NAPE-PLD increases the reaction rate with N-arachidonoyl-PE suggesting that the interaction with bile acids could favor the selective production of certain NAEs as anandamide and other unsaturated long NAEs[41].

[122] Magotti, P.; Bauer, I.; Igarashi, M.; Babagoli, M.; Marotta, R.; Piomelli, D.; Garau, G. Structure of Human N -Acylphosphatidylethanolamine-Hydrolyzing Phospholipase D: Regulation of Fatty Acid Ethanolamide Biosynthesis by Bile Acids. *Structure* **2015**, *23* (3), 598–604.
[123] Margheritis, E.; Castellani, B.; Magotti, P.; Peruzzi, S.; Romeo, E.; Natali, F.; Mostarda, S.; Gioiello, A.; Piomelli, D.; Garau, G. Bile Acid Recognition by NAPE-PLD. *ACS Chemical Biology* **2016**, *11* (10), 2908–2914.

3.1.8 Identification and characterization of ARN19874: the first NAPE-PLD inhibitor

A screening of a library of compounds was carried out at IIT (Italian Institute of Technology) leading to the identification of compound **50**, which inhibited human recombinant NAPE-PLD activity by 58% at 75 μM. The molecule is characterized by a quinazolinedione ring linked to an aromatic side chains by a sulfonamide linker.

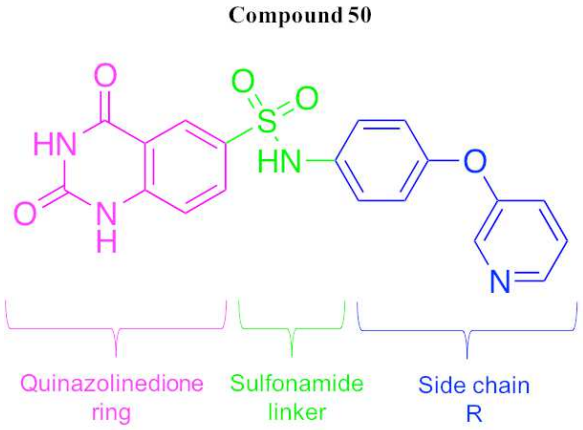
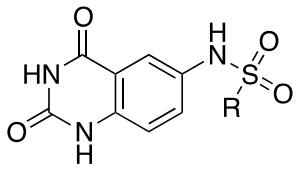
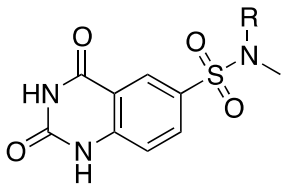
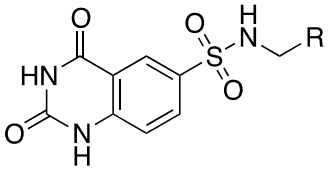
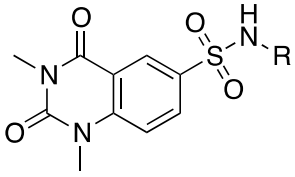
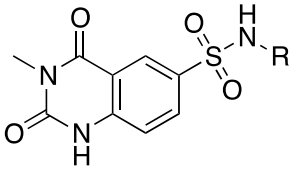
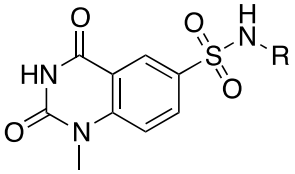
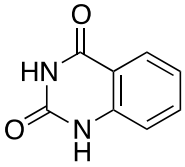


Figure 31: *compound **50** identified through the screening of a chemical library at IIT. The structure can be divided into three regions: the quinazolinedione ring in magenta, the sulfonamide linker in green and the side chain R in blue.*

Despite the moderate activity of this compound, it was hypothesized that the sulfonamide linker could coordinate the two zinc atoms in the NAPE-PLD active site and chemical modifications of the different region of the molecule could lead to an improved activity. With this in mind, a SAR (structure-activity relationship) study was then conducted to identify the structural elements responsible for NAPE-PLD inhibition and with the aim to increase potency.

Table 9: *chemical structure and NAPE-PLD inhibition of compounds 51–57. Compounds were tested at 75 μM on purified human recombinant NAPE-PLD.*

Modifications	Structure	Compound ID	% inhibition @75 μM
Sulfonamide linker		**51**	**16 ± 15**
		**52**	**67 ± 1**
		**53**	**16 ± 1**

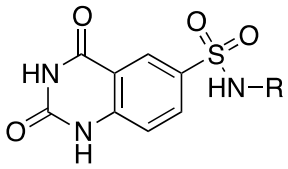
Quinazolinedione ring		**54**	**44 ± 3**
		**55**	**21 ± 4**
		**56**	**30 ± 9**
		**57**	**13 ± 7**

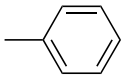
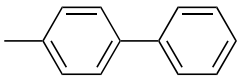
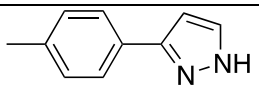
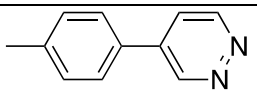
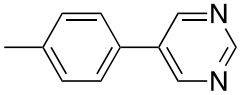
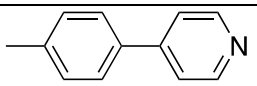
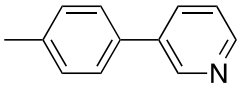
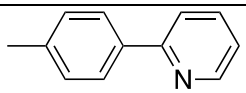
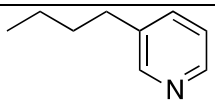
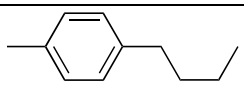
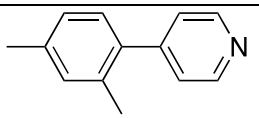
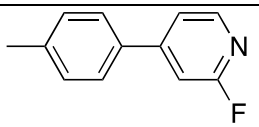
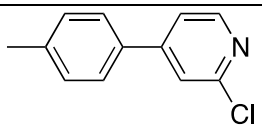
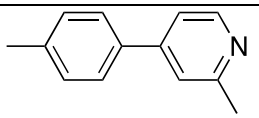
The sulfonamide linker was the first component to be investigated in SAR studies. Derivative **51** with reversed sulfonamide was characterized by a loss of inhibitory activity. The introduction of a methylene bridge in the linker of compound **53** also brought to a reduced inhibition of the enzyme. On the other hand, compound **52** displayed a N-methylated sulfonamide that retained the inhibitory activity. The drop in activity of compounds **51** and **53** suggested that the sulfonamide linker played a fundamental role in NAPE-PLD inhibition.

The quinazolinedione moiety was then modified, as this scaffold is reported in literature as chelating agent for metal ions and it is present in other metallo-enzyme inhibitors[124]. Methylation of the quinazolinedione nitrogens yielded to single and double-methylated derivatives **54**, **55** and **56** that retained the inhibitory activity, even though the inhibition was weaker than that of the parent molecule **50**. These results suggested that the quinazolinedione ring is an important component for NAPE-PLD inhibition, but less critical than the sulfonamide linker.

[124] Falsini, M.; Squarcialupi, L.; Catarzi, D.; Varano, F.; Betti, M.; Di Cesare Mannelli, L.; Tenci, B.; Ghelardini, C.; Tanc, M.; Angeli, A.; et al. 3-Hydroxy-1H-Quinazoline-2,4-Dione as a New Scaffold To Develop Potent and Selective Inhibitors of the Tumor-Associated Carbonic Anhydrases IX and XII. *J. Med. Chem.* **2017**, *60* (14), 6428–6439.

Table 10: *structures of compounds 58-71 and NAPE-PLD inhibition expressed as IC_{50}. Compounds were tested on human recombinant NAPE-PLD. The sign "*" means that the inhibition value is expressed as percentage of inhibition at 75 μM.*



Side chain R	Compound ID	Inhibition (IC_{50})
	**58**	**26.2 ± 0.7***
	**59**	**23.4 ± 7.3***
	**60**	**254.0 ± 1.3**
	**61**	**124.0 ± 1.1**
	**62**	**52.4 ± 1.2**
	**63**	**33.7 ± 9.8**
	**64**	**25.9 ± 1.2**
	**65**	**120.9 ± 1.2**
	**66**	**Not active**
	**67**	**Not active**
	**68**	**39.9 ± 1.2**
	**69**	**36.3 ± 1.2**
	**70**	**36.9 ± 2.3**
	**71**	**50.3 ± 1.3**

The side chain influence was then explored, maintaining intact both the quinazolinedione ring and the sulfonamide linker of compound **50,** while exploring R functionalizations (Table 10).

Compounds **58** and **59**, characterized by a phenyl or biphenyl side chain, showed only moderate inhibition of NAPE-PLD at 75μM while having some solubility issues. To overcome this problem, a series of derivatives, with a nitrogen-containing heterocycle that replaces the phenyl ring, were synthesized and tested. Compound **60** displayed a pyrazole ring that negatively affected the inhibitory activity. The introduction of a nitrogen-containing six-membered ring, like pyridazine **61** and pyrimidine **62,** helped to increase the potency. The hypothesis that nitrogen atoms of the inhibitor could undertake polar interactions with residues in the active site of NAPE-PLD was explored by synthesizing new pyridyl-phenyl derivatives with nitrogen atoms in different positions. Compounds characterized by a pyridino-phenyl side chain were assayed and **63** and **64** revealed to be the most potent NAPE-PLD inhibitors, while compound **65** displayed an activity comparable to that of **61**. These compounds differed in the position of the nitrogen atoms in the pyridine ring: in fact, nitrogen in position 2 was less tolerated compared to nitrogen in positions 3 or 4 and this difference suggested a precise pattern of interactions between the nitrogen atom of the pyridine ring and residues of NAPE-PLD active site. Moreover, the replacement of the phenyl ring with a more flexible n-propyl chain (**66)** or the replacement of the pyridyl ring with a n-butyl chain (**67**) resulted in a complete loss of activity. It was then decided to maintain fixed the 4-pyridil-phenyl side chain, trying to explore the effects of different substitutions. Compound **68** was characterized by an addition of a methyl group in *meta* position of the phenyl ring, retaining a good inhibitory activity. Compounds **69, 70** and **71** were assayed to test different substitutions in the side chain of the inhibitor. The compounds were functionalized in position alpha to the nitrogen atom: the first two molecules displayed halogens, like the small fluorine (**69**) and the bigger chlorine (**70)** atoms, while **71** displayed a methyl group. These functionalizations were tolerated but they didn't lead to an increase of potency.
In conclusion, this SAR analysis suggested that the sulfonamide group of the inhibitor should be appropriately positioned to determine NAPE-PLD inhibition and another essential component is the rigid aromatic side chain, that should have the chance to form H-bonds with the protein thanks to the pyridyl nitrogen.
Further experimental studies were performed for compound **63 (ARN19874),** which displayed good solubility, to better understand it pharmacological profile.

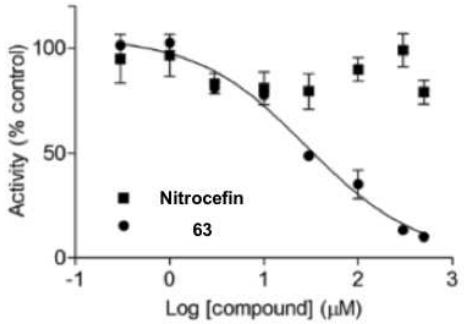
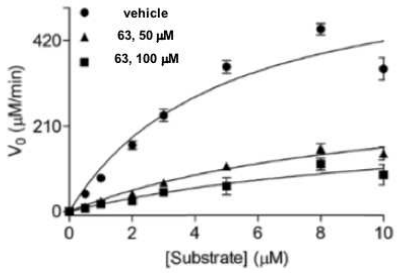


Figure 32: *inhibition of NAPE-PLD by nitrocefin and compound **63**. Nitrocefin resulted inactive in this assay.*

Petersen and coworkers tested the effect of the β-lactamase inhibitor nitrocefin against NAPE- PLD. Nitrocefin belongs to the class of cephalosporins and it is a β-lactam antibiotic, substrate for β-lactamase enzymes. The compound was reported to weak inhibit

NAPE-PLD with an IC_{50} of approximately 1 mM[125]. Figure 32 reports the effects of nitrocefin and compound **63** on human recombinant NAPE-PLD. The assay was unable to replicate the results obtained by Petersen and colleagues and nitrocefin was found inactive in this assay, while **63** was able to inhibit NAPE-PLD enzyme.



	Vehicle	63, 50 μM	63, 100 μM
Km (μM)	4.8 ± 2.1	10.1 ± 3.7	12.1 ± 8.9
Vmax (μM/min)	619.1 ± 123.8	314.5 ± 69.0	232.7 ± 109.2

Figure 33: *Michaelis–Menten kinetics of NAPE-PLD in the presence of vehicle (1% DMSO), or **63** at 50 μM or 100 μM. Kinetic parameters Km and Vmax were measured.*

Kinetic analysis of the compound **63** suggested that it acts with an uncompetitive mechanism of action. As reported in Figure 33, the K_m of the enzyme was 4.8 ± 2.1 μM and the V_{max} was 619.1 ± 123.8 μM/min. On the other hand, in presence of compound **63**, enzyme K_m increased, while V_{max} decreased.
Results shown in Figure 34 suggest that inhibition exerted by compound **63** was rapid and dilution of the complex readily reversed the effects of the inhibitor.

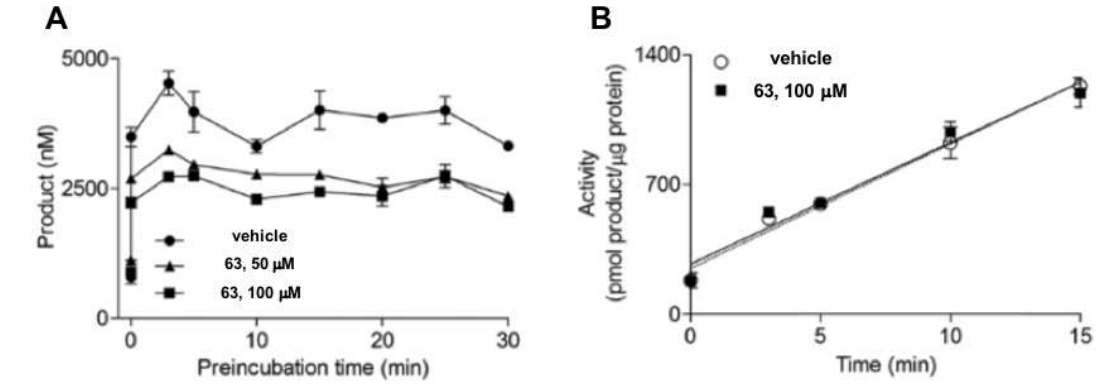


Figure 34: *time-course of NAPE-PLD in the presence of vehicle (1% DMSO), or **63** at 50 μM or 100 μM **(A)**. Rapid dilution assay of NAPE-PLD preincubated with vehicle (1% DMSO) or **63** (100 μM) **(B)**.*

In sum, these assays revealed that compound **63** could act as NAPE-PLD inhibitor through a reversible and uncompetitive mechanism. Since it was hypothesized that the hit compound could interact with the two zinc ions of the catalytic site of NAPE-PLD, its activity against other zinc-dependent enzymes was evaluated to determine its selectivity profile. The activity of **63** was tested against carbonic anhydrase (CA), neutral endopeptidase and angiotensin-converting enzyme (ACE) and it resulted inactive, as reported in Table 11.

Table 11: *compound **63** activity was tested against three different metallo-enzymes. The screening was performed by Eurofins Cerep Panlabs using compound **63** at 50 μM. Results are expressed as mean of percent inhibition of control*

[125] Petersen, G.; Pedersen, A. H.; Pickering, D. S.; Begtrup, M.; Hansen, H. S. Effect of Synthetic and Natural Phospholipids on N-Acylphosphatidylethanolamine-Hydrolyzing Phospholipase D Activity. *Chem. Phys. Lipids* **2009**, *162* (1–2), 53–61.

enzyme activity in three independent experiments. Moderate negative values are indicative of the variability of the signal around the control level.

Enzyme tested	Classification	**Inhibition** (% control)
Carbonic Anhydrase II	*Hydro-Lyase*	12
Neutral Endopeptidase	*Metalloendopeptidase*	-5.2
Angiotensin-converting enzyme	*Peptidyl-Dipeptidase*	-15.6

Moreover, it was found that compound **63** had no effects on the two enzymes involved in NAEs catabolism, FAAH and NAAA, as reported in Figure 35.

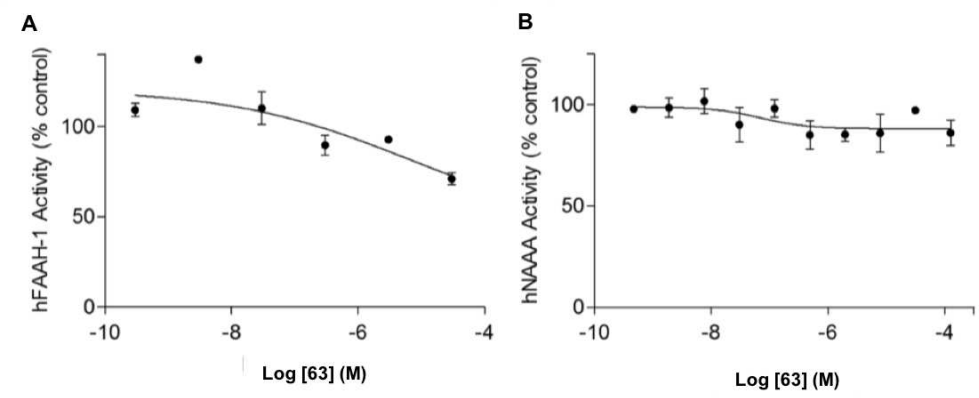


Figure 35: *activity of **63** was evaluated on two NAEs-hydrolyzing enzymes: human fatty acid amide hydrolase (FAAH) and N-acylethanolamine acid amidase (NAAA).*

Human HEK-293 cells, that constitutively express NAPE-PLD enzyme, were also incubated with compound **63** and different NAPEs, and NAE levels were measured by LCMS (liquid chromatography-mass spectrometry) to evaluate if the compound could be useful as probe to evaluate enzyme's activity. The incubation with **63** resulted in accumulation of non-metabolized NAPE species and, in absence of any other stimulation, the endogenous levels of OEA and PEA remained unmodified while significant changes were observed in stearoylethanolamide (SEA) levels.

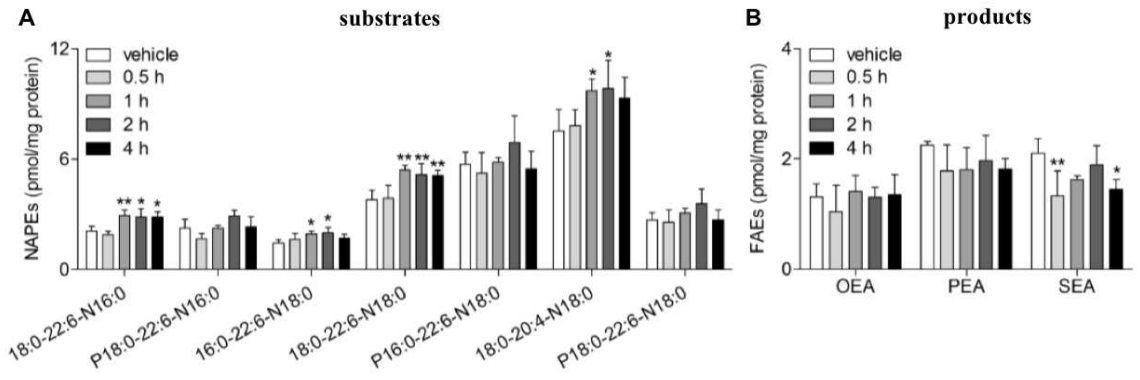


Figure 36: *effects of compound **63**(50 µM, 0 - 4 h) or its vehicle (0.5% DMSO) on NAPE and NAE levels in human HEK-293 cells. NAPE species differing in sn-1, sn-2 and N-acyl chains structure (A). NAE species, oleoylethanolamide (OEA), palmitoylethanolamide (PEA) and stearoylethanolamide (SEA) (B).*

3.2 Aim of the work

NAPE-PLD is the enzyme responsible for NAEs synthesis and its modulation could regulate the levels of these bioactive lipids, which mediate several physiological and pathological conditions. Anandamide, palmitoylethanolamide and oleoylethanolamide are involved in the regulation of anxiety, pain, inflammation, food intake, obesity, neurological disorders and cancer and the possibility to affect their concentration in particular districts of the organism represents a promising therapeutic strategy for a wide range of diseases. Despite the important role played by NAPE-PLD in regulating NAEs levels, only nitrocefin was reported as modulator of this enzyme so far. Up to now, molecules affecting NAE levels act against effector proteins (e.g., cannabinoid receptors) and the two enzymes involved in NAEs metabolism: FAAH and NAAA. These molecules have shown great potential as therapeutic agents, but they have displayed also some issues, in particular concerning undesired side effects.
In this chapter, docking studies were applied to the first small molecule inhibitor of NAPE-PLD with a novel chemical structure, ARN19874. Despite its low potency, this molecule displayed selectivity against the enzyme and its activity as molecular probe can help in better characterizing NAPE-PLD mechanisms and its role in pathologic and physiological processes.
As already mentioned NAPE-PLD is a phospholipase and, together with other lipases, is characterized by a mechanism called "interfacial activation", which affects enzyme conformation and its catalytic machinery. The understanding of this phenomenon could be crucial to gain insights into enzyme's behavior, highlighting aspects of the hydrolytic mechanism that couldn't be revealed by the X-ray structure alone. Molecular dynamics simulations were thus performed on the apo form of NAPE-PLD to better understand the conformational changes of the enzyme in both aqueous and membrane environment and to gain insight into its catalytic mechanism.

3.3 Results and discussion

3.3.1 Docking studies of ARN19874

We postulated that **ARN19874** could bind the active site of the enzyme, in particular, the sulfonamide function could coordinate the metal center of the enzyme characterized by the presence of two zinc ions. Docking studies of the hit compounds in the active site of NAPE-PLD were thus performed to gain insight into its binding pose and to evaluate the possible interactions with the protein residues. The best docking pose obtained in terms of *Gscore* resulted in **ARN19874** positioned in the active site with the sulfonamide group interacting with both zinc atoms, mimicking the phosphate group of the co-crystallized inhibitor phosphatidylethanolamine (PE) that probably occupies the same position of the substrate.

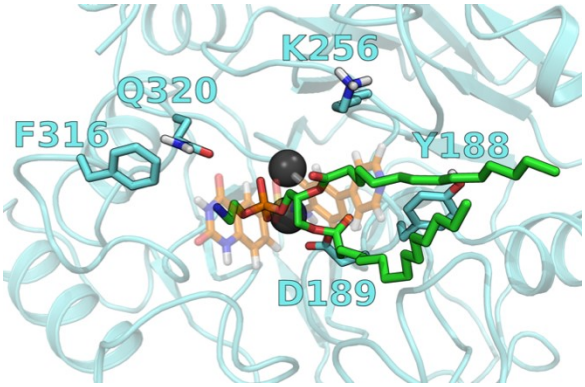


Figure 37: *superposition of the co-crystallized PE molecule (in green sticks) to the docking pose of compound* ***63*** *(in orange sticks) in the active site of NAPE-PLD enzyme.*

Additional polar interactions help in inhibitor stabilization: the carbonyl oxygen of D189 is H-bonded to the NH group of the sulfonamide linker and the oxygen of the quinazolinedione ring makes another H-bond with the side chain of Q320. From the crystal structure of NAPE-PLD (PDB 4QN9) this residue is thought to interact with the carboxyamide function of NAPEs. The 4-pyridyl-phenyl side chain of the inhibitor is located into the hydrophobic cavity occupied by the fatty acid chain of PE in the X-ray. The pyridine nucleus makes a pi-pi interaction with the aromatic ring of Y188. The pyridyl nitrogen of ARN19874 is not far from the flexible basic chain of K256 that can act as an additional point to anchor the hit compound to the active site.

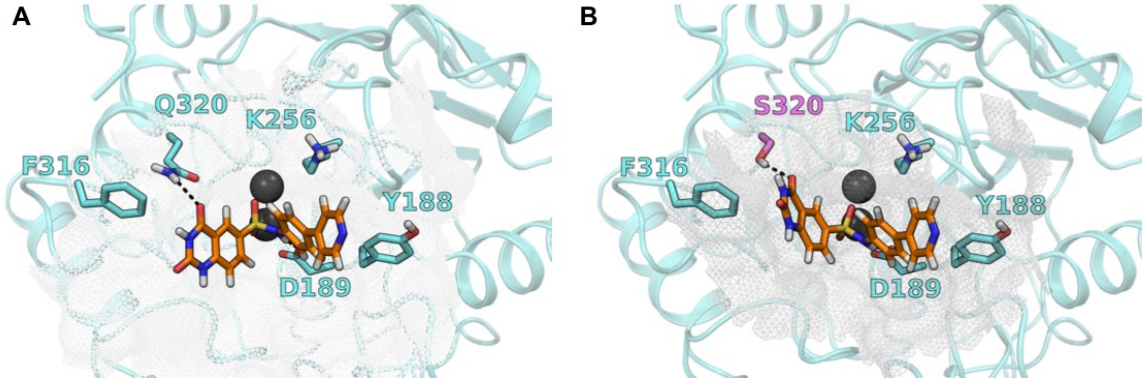


Figure 38*:best docking pose obtained for compound* ***63*** *in wild-type and mutated Q320S NAPE-PLD. The protein is shown in cyan ribbons and mutated residue S320 is displayed in purple.* ***ARN19874*** *is shown in orange sticks and zinc atoms in grey spheres. H-bonds are represented as black dashed lines.*

Mutagenesis studies were then performed at IIT to test the validity of the binding hypothesis generated by our docking studies. Two NAPE-PLD mutants were generated at residue Q320:

- Q320S that retained the capacity of glutamine to undertake an H-bond as both donor (hydrogen of the hydroxyl group) and acceptor (oxygen of the hydroxyl group) with the inhibitor;
- Q320A that abolished completely the possibility of glutamine to undertake an H-bond with the inhibitor.

Experimental analysis revealed that both Q320S and Q320A mutants displayed reduced activity compared to wild-type enzyme, confirming the important role of Q320 in catalysis. The reduced enzyme activity was however sufficient to investigate the effect of the inhibitor **63**. Observing the results obtained for the Q320A mutant, we deduced that the

compound was significantly less potent on the mutant form (IC_{50} =89.2 ± 2.3μM) compared to the wild-type NAPE-PLD (IC_{50} =46.2 ± 1.6μM). These results demonstrated that Q320 could play a role in supporting the inhibitory effects of **63**. On the other hand, mutation Q320S didn't significantly affect the inhibition capacity of the hit compound (IC_{50} =36.7 ± 1.6 μM) suggesting that serine could replace Q320 in the interaction with **63**.
These results support the hypothesis that our hit compound **63** binds the metal center of the enzyme thanks to its sulfonamide moiety, and the residue Q320 directly interacts with the inhibitor.

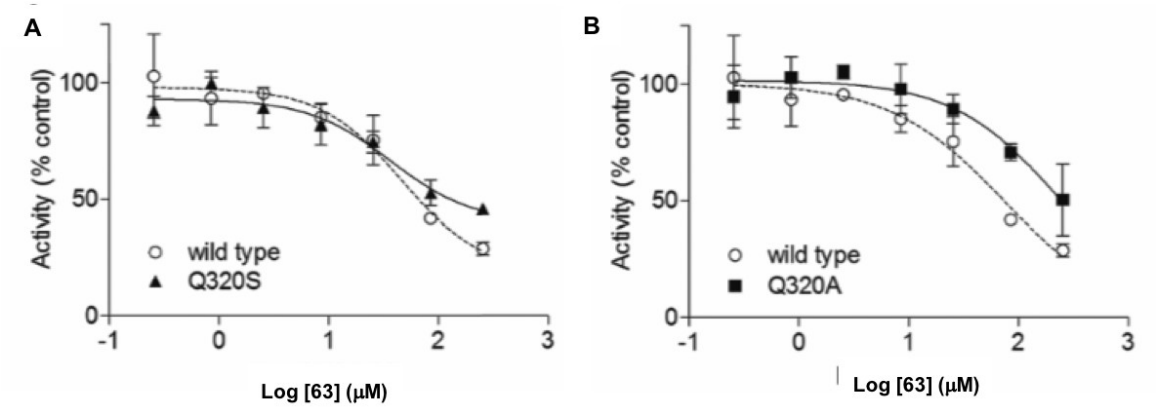


***Figure 39**:evaluation of compound **63** activity in wild-type and mutated form of the enzyme. Activity was retained in Q320S while reduced activity was observed in Q320A.*

These findings were collected in the following article: "Synthesis and characterization of the first inhibitor of N-acylphosphatidylethanolamine phospholipase D (NAPE-PLD)" by Castellani B., Diamanti E., Pizzirani D., Tardia P., Maccesi M., Realini N., Magotti P., Garau G., Bakkum T., Rivara S., Mor M. and Piomelli D published in Chemical Communication 2017 Nov 28;53(95):12814-12817.

3.3.2 Loop modelling and docking validation

The X-ray structure of human NAPE-PLD (PDBID: 4QN9) lacks two portions, as mentioned in section 3.1.6:

- the first 56 residues at the N-terminal (from residue 1 to residue 57);
- a eleven amino acid segment, from residue 80 to residue 90.

These two portions were not solved in the crystal structure probably because of their high flexibility, even if high B factors were reported for each residue in the protein in the PDB file. We decided to build the three-dimensional structure of the eleven amino acids segment because the loop was located in proximity of the active site of the enzyme and it could play a role in the catalytic mechanism and/or substrate accommodation, while the N-terminal portion was located far from the active center and no information were available about a possible role in enzyme activity. The crystal structure of NAPE-PLD from the OPM (Orientation of Proteins in Membrane) database[126,127] was also retrieved and we noted that the lacking 11 amino acid segment could interact with the membrane to anchor the protein to the lipid bilayer since it was located in its proximity.

[126] https://opm.phar.umich.edu
[127] Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: Orientations of Proteins in Membranes Database. *Bioinformatics* **2006**, *22* (5), 623–625.

The segment was thus rebuilt using Modeller software[128]. The lacking sequence of eleven residues was in proximity of the membrane interface, as seen in the 4QN9 structure deposited in the OPM database, thus we hypothesized that this loop could act as an anchor, helping to fix the protein on the lipid bilayer. Many membrane-associated proteins are characterized by a helical conformation in sequences that help anchoring the protein to the membrane[129] and we found some crystal structures of proteins displaying five residues of the lacking sequence as the first spiral of an alpha-helix. In the end, part of the missing sequence was modelled as a loop, while five residues of the sequence were rebuilt with a helix conformation (see section 3.5.3 for details).

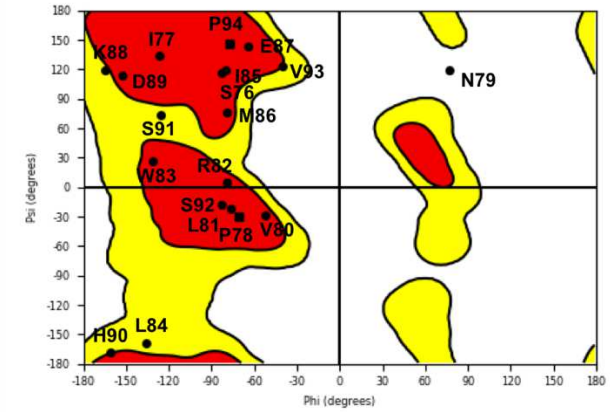


Figure 40: *Ramachandran plot of the sequence 76-94 obtained by Modeller. Each black dot represents a single residue: proline is plotted as squares and all other residues are plotted as circles. This plot is divided in different regions characterized by allowed and disallowed values of dihedral angles for the protein residues. The red regions are the "favored" regions, the yellow regions are the "allowed" regions, and the white regions are the "disallowed" regions.*

Ramachandran plot of the loop 76-94 was subsequently analyzed to check the dihedral angles of the modelled sequence (residues 80-90). The only residue placed in the "disallowed region" of the plot was N79 but this residue was reported with these dihedrals angles in the crystallographic structure and it was then optimized during the preparation and refinement of the model. Rebuilt residues were placed in both "allowed" and "favored" regions of the plot, with residues belonging to the helical segments (V80, L81, R82 and W83) placed in the region of the plot displaying dihedrals values typical of an alpha-helix (the red region found between -60 and 30 of the psi dihedral and -120 and -30 of the phi dihedral) confirming the geometrical goodness of our model.

[128] Fiser, A.; Do, R. K.; Sali, A. Modeling of Loops in Protein Structures. *Protein Sci.* **2000**, *9* (9), 1753–1773.
[129] Taylor, P. D.; Toseland, C. P.; Attwood, T. K.; Flower, D. R. Alpha Helical Trans-Membrane Proteins: Enhanced Prediction Using a Bayesian Approach. *Bioinformation* **2006**, *1* (6), 234–236.

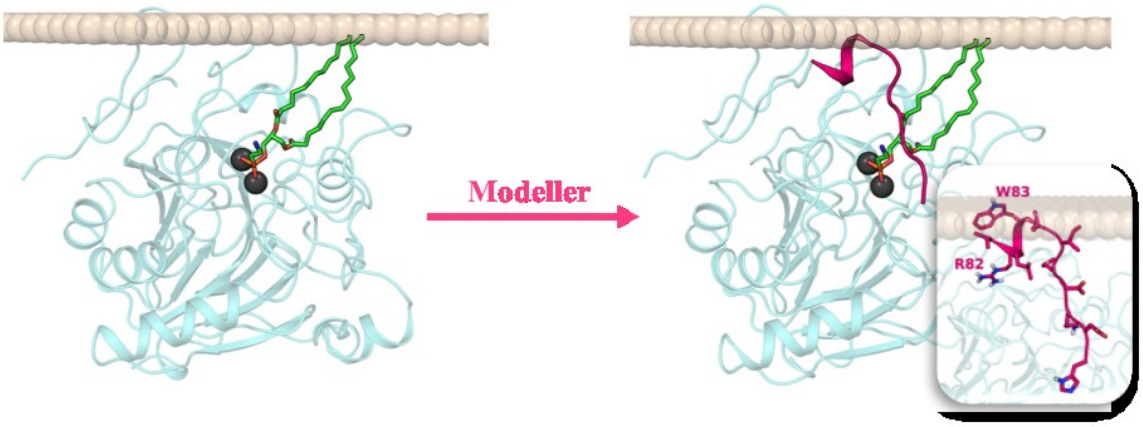


Figure 41: *crystal structure of human NAPE-PLD lacks an eleven amino acid segment that was rebuilt using Modeller. The loop is shown in magenta ribbon and it is part of a longer loop that flanks a channel that connects the active site of the enzyme to the membrane bilayer. Membrane lipids are shown as light brown spheres.*

Figure 41 displays the full model of the protein (from residue 57 to residue 388) obtained and it is possible to appreciate that residues R82 and W83 of the rebuilt helix show different orientations consistent with their polarity. R82 side chain, in fact, is exposed to the solvent, while the indole ring of W83 is inserted into the lipid bilayer.
Docking studies of inhibitor **63** were also performed in the full model in both wild-type and Q320S enzyme and the resulting poses displayed the same orientation and they reproduced the same interactions of the poses obtained in the crystal structure.
In particular, we observed that the eleven amino acid segment rebuilt using Modeller, belonged to a longer loop which formed the external side of the channel that connects the membrane bilayer to the active site of the protein and, given its high flexibility, we hypothesized that it could play a role in interfacial activation. We decided to submit the full model to molecular dynamics simulations to investigate potential conformational changes of this loop that could affect substrate recruitment.

3.3.3 Molecular dynamics simulations

The NAPE-PLD model obtained previously was submitted to molecular dynamics (MD) simulations to highlight conformational changes of the enzyme in two different environments: in water and in the presence of a membrane model to test the hypothesis of interfacial activation.
Five replicas of 100 ns MD simulation were performed, and the first parameter analyzed was the RMSD (root mean square deviation) of the backbone atoms to evaluate the overall stability of the protein during the simulation time.

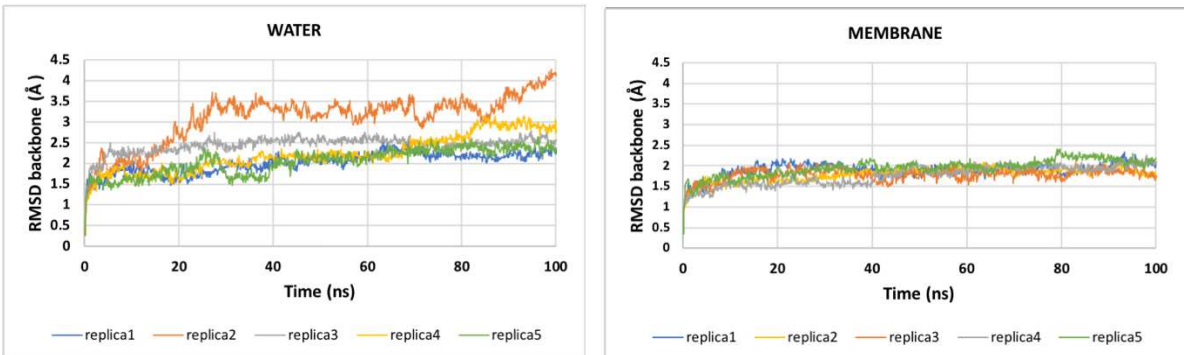


Figure 42*:RMSD was computed on the backbone atoms of NAPE-PLD during the simulations.*

As displayed in Figure 42, the simulations in water environment were characterized by a higher RMSD compared to those in membrane. The five replicas in water displayed an RMSD between 1 and 3 Å, except for replica 2 that showed a higher RMSD, between 3 and 3.5 Å, that reached 4 Å in the last 20 ns of the simulation. In this simulation residues from 67 to 78 and residues 86-90 (belonging to the modelled region) moved away from the starting positions without stabilizing in an equilibrium conformation. On the other hand, RMSD of the five replicas in membrane displayed a value of about 2 Å for all the simulations, suggesting that the membrane bilayer could play a role in stabilizing the enzyme structure. It has to be pointed out that the presence of a lipid bilayer interacting with the enzyme naturally decreases the conformational space that can be sampled by the protein.
Since all the simulations (except for replica 2 in water) showed reasonable RMSD values, we further analyzed them, calculating the RMSF (root mean square fluctuation) on alpha carbons. This value measures the average fluctuation of each alpha carbon of the protein during the simulation time.

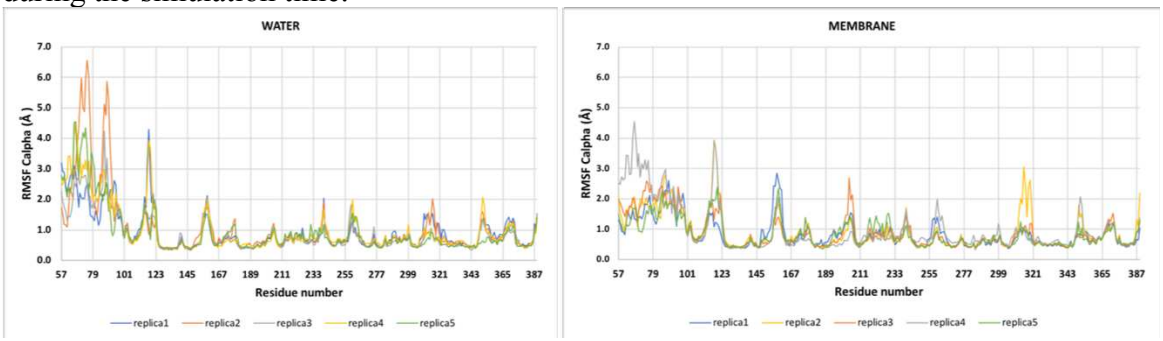


***Figure 43**:RMSF was computed on the alpha carbons of NAPE-PLD during the simulations.*

As Figure 43 revealed, high RMSF values were captured for the N-terminal sequence of NAPE-PLD both in water and membrane environment (except for the first five residues which were restrained during simulations). In water, alpha carbons of this region reached RMSF values of 6 Å for replica 2 (this replica displayed the highest RMSD), but the same loop displayed high mobility also in membrane even if the values were relatively lower. The alpha carbons of the rebuilt sequence (residue 80 to residue 90) displayed a RMSF below 3 Å suggesting a good stability of the modelled structure, except for replica 2 in water where RMSF of residues of this region was higher, as previously reported. The sequence modelled as alpha-helix retained its conformation for all the 100 ns of simulations in membrane environment thanks to the interaction with membrane phospholipids that probably helped in stabilizing the conformation of the segment. In water environment, all the replicas showed some frames in which the helix tended to unfold suggesting a specific role of the membrane in stabilizing NAPE-PLD conformation.
The N-terminal region of NAPE-PLD showed high flexibility during the simulations: this region flanks the channel that connects the hydrophobic hook to the active site of the protein and it is in contact with the membrane as suggested by the structure retrieved in the OPBM database. We postulated that, given its high fluctuation, it could be involved in interfacial activation mechanism.

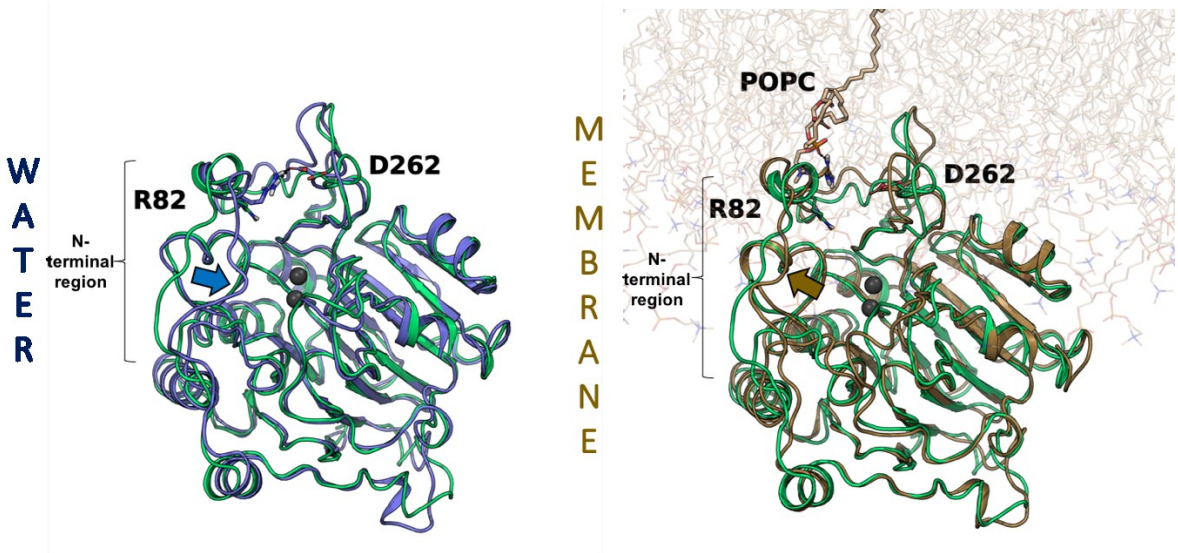


Figure 44:*representative frames of NAPE-PLD conformations assumed in water (blue ribbons) and membrane (brown ribbons) environments. The starting NAPE-PLD structure is represented in green ribbons. In water the N-terminal loop moves, closing the active site channel of the enzyme, as indicated by the blue arrow. On the other hand, the presence of the membrane stabilizes an open conformation of the protein, as indicated by the brown arrow.*

Figure 44 reports two representative frames of the MD simulations carried out in water and in membrane environment. In water, the loop that we postulated could be involved in interfacial activation, slightly moves towards the active site, reducing the size of the channel in three replicas out of five. On the other hand, the presence of the membrane maintains the channel in an open conformation for all the replicas. At the edge of the channel that connects the membrane bilayer to the active site of the protein there are two residues, R82 and D262 which are located at the opposite side of the gorge. The two residues were not interacting in the starting structure, but in water simulations the formation of a salt bridge between these two residues was favored compared to the simulations in membrane environment.

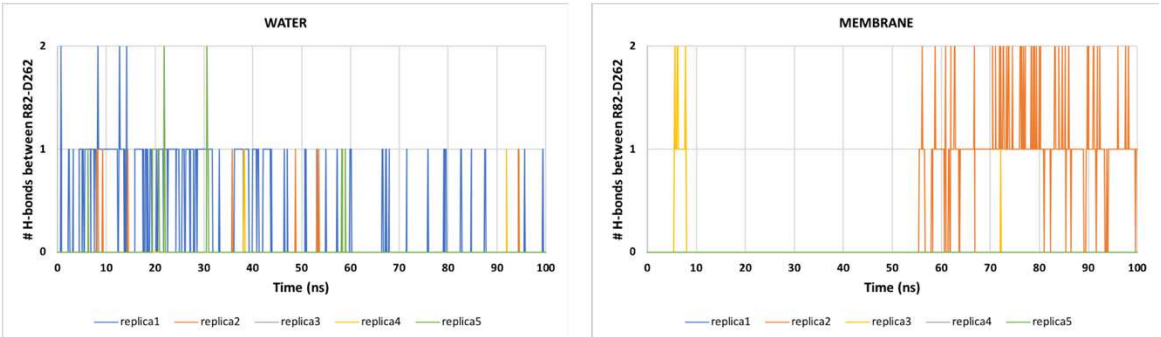


Figure 45: *the number of H-bonds between residues R82 and D262 was registered during 100 ns of all the simulations for each replica. The number of H-bonds between the two residues could be 1 or 2 because one or two guanidine nitrogens of R82 could interact at the same time with the acidic counterpart.*

As reported in Figure 45, all the simulations in water displayed the formation of one or two salt bridges between R82 and D262, except for replica 3. This event was rarer in membrane simulations where only replica 3 and some frames of replica 2 registered this phenomenon. We found that, in membrane environment, R82 preferred to interact with the polar head group of POPC residues in all the five replicas, as reported in Figure 46.

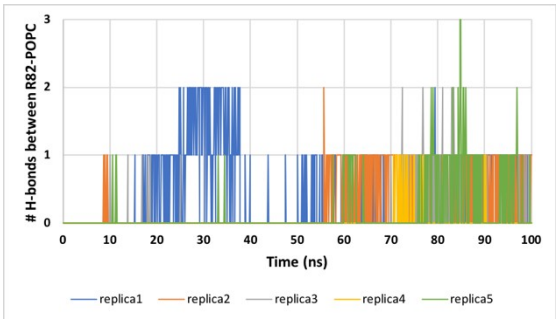


Figure 46: *number of H-bonds occurring between R82 and POPC residues of the membrane bilayer in the five replicas.*

NAPE-PLD hydrolyzes the distal bond of membrane phospholipids and the results of the MD simulations in membrane environment suggest that R82 and D262 act as a "gate" that allow the access of lipid substrates to the active site of the enzyme. In water environment the interaction between R82 and D262 avoid the access of the substrate to the active site, originating a closed conformation. In presence of the membrane, the gate is "open", with no salt-bridge between R82 and D262. Moreover, R82 seems to play an active role in substrate recruitment, as it could directly interact with membrane phospholipids that will be subsequently hydrolyzed in the active site of the enzyme.

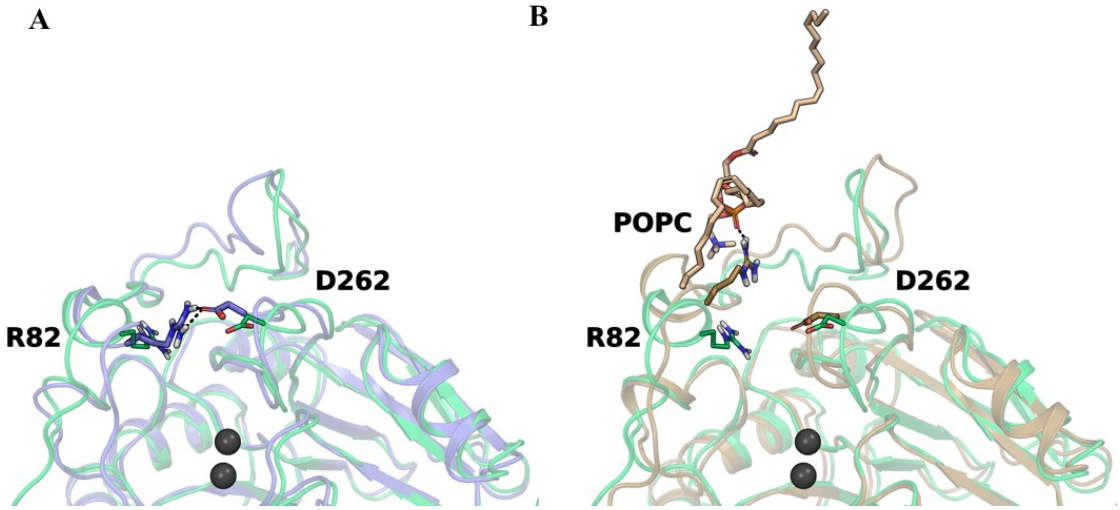


Figure 47: *a close-view of residues R82 and D262 proposed to act as the gate of the channel that connects the hydrophobic hook to the active site of NAPE-PLD. In water environment, formation of a salt-bridge between R82 and D262 is favored, hampering the access of the substrate to the active site* ***(A).*** *Formation of the salt bridge is prevented in membrane, as the gate is maintained in an open conformation, compatible with the access of the substrate; residue R82 preferentially interacts with POPC residues of the lipid bilayer* ***(B).***

3.4 Conclusions

NAPE-hydrolyzing phospholipase D (NAPE-PLD) is a membrane-associated enzyme belonging to the class of phospholipases. This enzyme is involved in the transacylation pathway that hydrolyze specific membrane phospholipids called N-acyl phosphatidylethanolamines (NAPEs), generating N-acyl ethanolamines (NAEs) and phosphatidic acid. NAEs are bioactive lipids, which include the endocannabinoid anandamide, the anti-inflammatory agent, palmitoylethanolamide, and oleoylethanolamide, which regulates several important functions, such as food intake, cancer cell proliferation, inflammation processes and cellular signaling.

The crystal structure of human NAPE-PLD has been recently solved[130] providing insights into the structure and the mechanism of catalysis of the protein. The enzyme belongs to the metallo-ß-lactamase superfamily, presenting an αβ/βα fold with the active site containing two metal (Zn) ions, which are probably involved in the hydrolysis reaction. In 2009 Petersen and colleagues tested a beta-lactamase substrate, the cephalosporin nitrocefin, highlighting its inhibitory activity against NAPE-PLD enzyme *in vitro*[131]. After that experiment no new NAPE-PLD inhibitors were discovered, but the development of enzyme modulators could be useful in the characterization and investigation of NAPE-PLD mechanism and of its role in several pathologic or physiological processes. Here, the identification and characterization of ARN19874 was describes. This compound represents the first small molecule with a novel structure targeting human NAPE-PLD so far discovered and, despite its low potency (IC_{50}=46.2 ± 1.6μM), it can be considered as a starting point towards the development of new NAPE-PLD inhibitors. This molecule acts with a reversible and uncompetitive mechanism of action with good selectivity towards the enzyme. Experimental assays have also revealed its capacity to affect NAEs levels in cells, highlighting its potential as NAPE-PLD probe. Docking studies of the newly identified NAPE-PLD inhibitor ARN19874 were performed on the active site of the enzyme to rationalize its binding mode and the binding mode of its derivatives. The computational studies were then validated by mutagenesis studies. In the end, the identification of the binding pose of the first small molecule inhibitor of NAPE-PLD provided thus the basis for a rational design of novel and more potent NAPE-PLD modulators.
Molecular dynamics simulations were performed to gain insights into the catalytic mechanism of the enzyme, testing the theory of interfacial activation, which is a common phenomenon for lipases. Five replicas were run on the apo enzyme in both aqueous solvent and in presence of a membrane bilayer. The results support the theory of interfacial activation, since two different conformations of the protein were identified in the two media. In water the formation of a salt bridge between residues R82 and D262, proposed to act as the gate of access to the catalytic center, promoted the "closed" conformation of the enzyme, not compatible with the access of the substrate to the active site. On the other hand, membrane environment allows the protein to maintain an "open" conformation, compatible with catalysis. Moreover, R82 seemed to play an active role in substrate recruitment. Even if these results have to be confirmed experimentally, they provide us a rationale on the molecular basis of NAPE-PLD catalysis that couldn't be perceived with the X-ray structure.

3.5 Computational protocol

3.5.1 Protein preparation for docking studies

Chain A of the x-ray structure of human NAPE-PLD was retrieved from Protein Data Bank (PDBID: 4QN9)[132] and used for docking studies. The protein was prepared using the

[130] Magotti, P.; Bauer, I.; Igarashi, M.; Babagoli, M.; Marotta, R.; Piomelli, D.; Garau, G. Structure of Human N -Acylphosphatidylethanolamine-Hydrolyzing Phospholipase D: Regulation of Fatty Acid Ethanolamide Biosynthesis by Bile Acids. *Structure* **2015**, *23* (3), 598–604.
[131] Petersen, G.; Pedersen, A. H.; Pickering, D. S.; Begtrup, M.; Hansen, H. S. Effect of Synthetic and Natural Phospholipids on N-Acylphosphatidylethanolamine-Hydrolyzing Phospholipase D Activity. *Chem. Phys. Lipids* **2009**, *162* (1–2), 53–61.
[132] Magotti, P.; Bauer, I.; Igarashi, M.; Babagoli, M.; Marotta, R.; Piomelli, D.; Garau, G. Structure of Human N -Acylphosphatidylethanolamine-Hydrolyzing Phospholipase D: Regulation of Fatty Acid Ethanolamide Biosynthesis by Bile Acids. *Structure* **2015**, *23* (3), 598–604.

Protein Preparation Wizard of the Schrödinger suite 2015–4[133]. Firstly, sulfate ions, deoxycholic acid molecules and PE inhibitor co-crystallized with the enzyme were removed. Hydrogen atoms were added to the structure and bond orders were assigned and checked: zero-order bonds connecting the two zinc ions to their coordinating atoms were added. PROPKA[134] was then used to ***i)*** properly orient thiol and hydroxyl groups and to ***ii)*** adjust conformation of asparagine, glutamine and histidine residues to optimize the overall hydrogen bonding network. A first energy minimization of the complex, where only hydrogen atoms were free to move, was performed. A second minimization run was then performed, restraining the position of the heavy atoms to an RMSD value of 0.3Å.
All crystallized water molecules were removed before docking studies. The Q320S NAPE-PLD mutant was modeled replacing glutamine 320 with serine.

3.5.2 Docking calculations

Glide software version 6.9[135,136] was used for docking studies with the SP scoring function and OPLS 2005 force field[137]. Glide grid was built on the prepared crystal structure and it was centered on the two zinc ions of the active site. The dimensions of the inner and outer boxes were set to 20 and 40 Å, respectively. No Van der Waals scaling was applied to nonpolar receptor atoms. Inhibitor **63** was built in Maestro 10.4[138] and prepared with LigPrep 3.6[139] using default settings. Only neutral species of the inhibitor were subjected to docking studies. Docking run was performed using default settings and 100 poses were collected for each run and subsequently ranked according to their G-score values. The top-ranked pose was retained and analyzed.

3.5.3 Loop building

The eleven amino acid segment from residue 80 to residue 90 was rebuilt using Modeller software v 9.16[140,141].
Since the lacking segment was in proximity of the membrane interface, as seen in the 4QN9 structure deposited in the OPM database, we hypothesized that this loop could act as an anchor, helping to fix the protein on the lipid bilayer: many membrane-associated proteins are characterized by a helical conformation in sequences that help anchoring the protein to the membrane[142], so we supposed that this lacking segment could display an alpha-helix

---

[133] Protein Preparation Wizard 2015-4. Schrödinger LLC, New York, NY (2015).
[134] Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M. & Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **7,** 525–537 (2011).
[135] Glide, version 6.9. Schrödinger, LLC, New York, NY (2015).
[136] Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S., Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem* **2004,** *47* (7), 1739-49.
[137] Banks, J. L.; Beard, H. S.; Cao, Y.; Cho, A. E.; Damm, W.; Farid, R.; Felts, A. K.; Halgren, T. A.; Mainz, D. T.; Maple, J. R.; et al. Integrated Modeling Program, Applied Chemical Theory (IMPACT). *J Comput Chem* **2005**, *26* (16), 1752–1780.
[138] Maestro, version 10.4, Schrödinger, LLC: New York, NY, 2015.
[139] LigPrep, version 3.6. Schrödinger, LLC, New York, NY (2015).
[140] Fiser, A.; Do, R. K.; Sali, A. Modeling of Loops in Protein Structures. *Protein Sci.* **2000**, *9* (9), 1753–1773.
[141] Sali, A.; Blundell, T. L. Comparative Protein Modelling by Satisfaction of Spatial Restraints. *J. Mol. Biol.* **1993**, *234* (3), 779–815.
[142] Taylor, P. D.; Toseland, C. P.; Attwood, T. K.; Flower, D. R. Alpha Helical Trans-Membrane Proteins: Enhanced Prediction Using a Bayesian Approach. *Bioinformation* **2006**, *1* (6), 234–236.

conformation. To confirm this hypothesis, a search with BLAST server[143] for proteins containing the sequence -PNVLRWLIMEKDHSS- was performed and crystal structures in the Protein Data Bank containing the sequence or part of it were retrieved to analyze the conformation. Several crystal structures (PDBID: 2V5G, 3EG9, 5KYN and 2NUP to name few of them) were identified containing sequence VLRWL in an alpha-helix conformation.

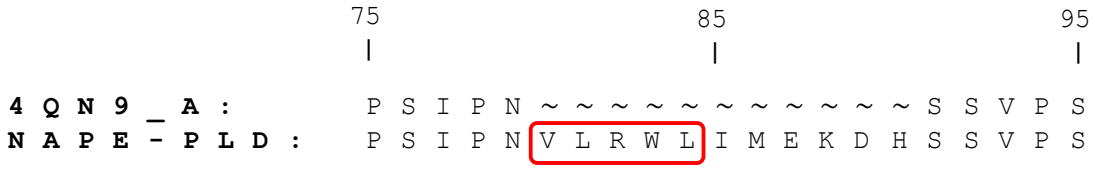


Figure 48: *crystal structure of human NAPE-PLD 4QN9 lacks a ten amino acid segment, from residue 80 to residue 90. The five amino acids circled in red were built as alpha helix using Modeller.*

The lacking loop of NAPE-PLD was thus rebuilt forcing the sequence VLRWL to display an alpha-helix conformation, while the remaining sequence was rebuilt as a loop. The model with the lowest *molpdf* (one of the Modeller scoring functions used to rank resulting models from an energetic perspective) was chosen.

3.5.4 Protein preparation for molecular dynamics simulations

The full protein model obtained with Modeller was then submitted to the Protein Preparation Wizard of the Schrödinger suite 2016–3[144]. Sulfate ions, deoxycholic acid molecules and PE inhibitor co-crystallized with the enzyme were removed. We hypothesized that a water molecule between the two zinc atoms was missed in the crystal structure: the involvement of a water molecule is, in fact, necessary for the hydrolysis reaction to happen. A water molecule was thus added in proximity of the two zinc atoms and coordination bonds were placed between the water molecule and the two zinc atoms of the active site. All the crystallized water molecules found in the original crystal structure were kept.
The procedure to prepare and refine the protein followed the same steps described in section 3.5.1.

3.5.5 Molecular dynamics simulations

Molecular dynamics simulations were performed using Desmond 5.2[145] with OPLS3[146] as force field.
For the MD simulations in water, the protein was solvated by TIP3P water molecules in an orthorhombic simulation box with boundaries10 Å distant from the protein in every direction. The net charge of the system was neutralized by adding five Na^+ atom. The total number of atoms in the system was 40238.
For simulations of the enzyme in presence of the membrane, the NAPE-PLD structure was embedded in a POPC lipid bilayer model according to the coordinates of the 4QN9 crystal structure deposited into the Orientation of Protein in Membrane (OPM)

[143] https://blast.ncbi.nlm.nih.gov/Blast.cgi
[144] Protein Preparation Wizard 2016-3. Schrödinger LLC, New York, NY (2016).
[145] Desmond, version 5.2. Schrödinger, LLC, New York, NY (2017).
[146] Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; et al. OPLS3: A Force Field Providing Broad Coverage of Drug-like Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12* (1), 281–296.

database[147]. The protein-membrane complex was then solvated by TIP3P water molecules, in a simulation box of 10 Å × 10 Å × 20 Å. The net charge of the system was neutralized by adding five Na^+ atom. The total number of atoms in the system was 77019.
All bond lengths to hydrogen atoms were constrained using the M-SHAKE algorithm. Short-range electrostatic interactions were cut off at 9 Å, whereas long-range electrostatic interactions were computed using the Smooth Particle Mesh Ewald method. A RESPA integrator was used with a time step of 2 fs, and long-range electrostatics were computed every 6 fs. Each MD simulation was carried out in the NPT ensemble, coupling to the Langevin thermostat method. Five replicas for each system were performed modifying the seed parameter. As the protein lacks the first 56 residues and to help the stabilization of the protein during MD simulations, the production stage of the dynamics was carried out with alpha carbons of residues 57-61 and 385-389 restrained by 2 kcal mol^{-1} $Å^{-2}$.

Equilibration protocol of Molecular Dynamics simulations performed in water

Each simulation performed was carried out in the NPT ensemble, coupling to the Langevin thermostat method, at a temperature of 300 K. The system was previously equilibrated by applying the following protocol:

1. 500 ps of Brownian dynamics in the NVT ensemble with:
    - Simulation Temperature of 10 K
    - 50 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms
    - Small timesteps: 1, 1, 3 fs
2. 500 ps simulation in the NVT ensemble, coupling the Langevin thermostat with:
    - Simulation Temperature of 10 K
    - 50 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms
    - Small timesteps: 1, 1, 3 fs
3. 500 ps simulation in the NVT ensemble, coupling the Langevin thermostat with:
    - Simulation Temperature of 100 K
    - 50 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms
    - Small timesteps: 1, 1, 3 fs
4. 500 ps simulation in the NVT ensemble, coupling the Langevin thermostat with:
    - Simulation Temperature of 200 K
    - 50 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms
    - Small timesteps: 1, 1, 3 fs
5. 1000 ps simulation in the NVT ensemble, coupling the Langevin thermostat with:
    - Simulation Temperature of 300 K
    - 50 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms
    - Small timesteps: 1, 1, 3 fs
6. 1000 ps simulation in the NVT ensemble, coupling the Langevin thermostat with:
    - Simulation Temperature of 300 K
    - 50 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms
    - Timesteps: 2, 2, 6 fs
7. 1000 ps simulation in the NPT ensemble, coupling the Langevin thermostat with:

[147] Lomize, M. A.; Lomize, A. L.; Pogozheva, I. D.; Mosberg, H. I. OPM: Orientations of Proteins in Membranes Database. *Bioinformatics* **2006**, *22* (5), 623–625.

- Simulation Temperature of 300 K
- 50 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms

8. 2000 ps simulation in the NPT ensemble, coupling the Langevin thermostat with:
    - Simulation Temperature of 300 K
    - 25 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms
9. 2000 ps simulation in the NPT ensemble, coupling the Langevin thermostat with:
    - Simulation Temperature of 300 K
    - Progressive decrease of restraints from 25 kcal mol^{-1} $Å^{-2}$ to 5 kcal mol^{-1} $Å^{-2}$ on solute heavy atoms
10. 2000 ps simulation in the NPT ensemble, coupling the Langevin thermostat with:
    - Simulation Temperature of 300 K
    - 5 kcal mol^{-1} $Å^{-2}$ restraints on protein backbone, zinc atoms and residues 185, 187, 189, 190, 253, 284, 343 and the water molecule 679 of the active site that coordinates the zinc atoms
11. 5000 ps simulation in the NPT ensemble, coupling the Langevin thermostat with:
    - Simulation Temperature of 300 K
    - Progressive decrease of restraints from 5 kcal mol^{-1} $Å^{-2}$ to 2 kcal mol^{-1} $Å^{-2}$ on backbone, zinc atoms and residues 185, 187, 189, 190, 253, 284, 343, and the water molecule 679 of the active site that coordinates the zinc atoms

Equilibration protocol of Molecular Dynamics simulations in presence of the membrane model

Each simulation performed was carried out in the NPT ensemble, coupling to the Langevin thermostat method, at a temperature of 300 K. The membrane system was previously equilibrated using the Desmond default protocol to relax water and the membrane bilayer:

1. 500 ps of Brownian dynamics in the NVT ensemble with:
    - Simulation Temperature of 10 K
    - 50 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms
    - Small timesteps: 1, 1, 3 fs
2. 5000 ps simulation in the NVT ensemble, coupling the Nose-Hover thermostat with:
    - Gradual heating from 10 K to 300 K
    - 50 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms
    - Small timesteps: 1, 1, 3 fs
3. 2000 ps simulation in the NPT ensemble, coupling with Langevin thermostat with:
    - Simulation Temperature of 300 K
    - 50 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms
    - Small timesteps: 1, 1, 3 fs
4. 500 ps simulation in the NPT ensemble, coupling the Langevin thermostat with:
    - Simulation Temperature of 300 K
    - 50 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms
    - Timesteps: 2, 2, 6 fs
5. 1000 ps simulation in the NVT ensemble, coupling the Langevin thermostat with:
    - Simulation Temperature of 300 K
    - 25 kcal mol^{-1} $Å^{-2}$ restraints on solute heavy atoms
    - Timesteps: 2, 2, 6 fs

After the membrane equilibration stages, the entire system was equilibrated using the same procedure of the water simulations starting from step 9.

4. Screening of the Pathogen Box against Schistosoma mansoni: data analysis and drug evaluation

4.1 Introduction

4.1.1 The challenge of Neglected Tropical Diseases

According to WHO, Neglected Tropical Diseases (NTDs) are a group of infectious diseases caused by parasites and bacteria that affect 149 countries and more than one billion people living in tropical and subtropical areas all around the globe[148]. These diseases are common in poor countries and territories where the population lives without adequate sanitation and in close contact with the sources of infection. The term "neglected" refers to the fact that these group of diseases have been eradicated in the most developed parts of the world and they only persist in the poorest and most marginalized parts, especially in conflict areas. These illnesses usually impair physical and cognitive development, causing disability and disfigurement, affecting person's ability to work and earn money for living, limiting the productivity and preventing children's education. These diseases can have also deadly consequences: WHO reports that half million people die each year for NTDs[149]. As a result, countries affected by NTDs remain blocked in a cycle of poverty and disease from which is difficult to escape. Another reason which makes these diseases as "neglected" is the fact that pharmaceutical companies are not interested in investing money for the research and development of drugs to treat NTDs, since there is little economic return.
NTDs are caused by different types of pathogens, mostly bacteria, viruses, protozoa and helminths. Table 12 reports the main NTDs classified according to the causative type of pathogen.

***Table 12**: overview of neglected tropical diseases grouped by classes of pathogen reported by WHO.*

Protozoa	Viruses	Helmints	Bacteria
Chagas Disease	*Dengue*	*Echinoccosis*	*Buruli ulcer*
Leishmaniasis	*Chikungunya*	*Mycetoma*	*Leprosy*
Human African Trypanosomiasis	*Rabies*	*Schistosomiasis*	*Trachoma*
		Soil-transmitted helminthiases	*Yaws*
		Lymohatic FIlariasis	
		Guinea-worm disease	
		Foodborne trematodes	
		Onchocerciasis	
		Taeniasis and neurocysticercosis	

[148] http://www.who.int/neglected_diseases/diseases/en/
[149] Sustaining the drive to overcome the global impact of neglected tropical diseases: Second WHO report on neglected tropical diseases. Geneva: World Health Organization; 2013.

In 2012 WHO published the NTDs roadmap[150]. This plan aims to prevent, control, and eradicate 17 neglected tropical diseases by 2020: among them we found dengue, schistosomiasis, rabies, onchocerciasis and leishmaniasis, just to name a few. Besides the intervention of WHO, other private and public organizations are helping in financing and developing programs for control and prevention of NTDs. For example, the Bill & Melinda Gates foundation[151] and the USA Agency for International Development have implemented a health program aimed at supporting poor countries affected by NTDs[152], while the Department for International Development of United Kingdom[153] declared to invest approximately 449 million dollars in the next 5 years against NTDs.

4.1.2 Schistosomiasis: an overview

Schistosomiasis, also called bilharzia, is the second most diffuse NTD after malaria[154]. It is an acute and chronic disease caused by the contact and following infection by parasitic blood flukes (trematode worms) called schistosomes.
Schistosomiasis affects about 240 million people especially in tropical and sub-tropical areas of the world, mostly in poor countries with no adequate hygiene conditions and no potable water and it is estimated that 89 million people have been treated against the disease[155]. Figure 49 reports the geographical distribution of schistosomiasis.

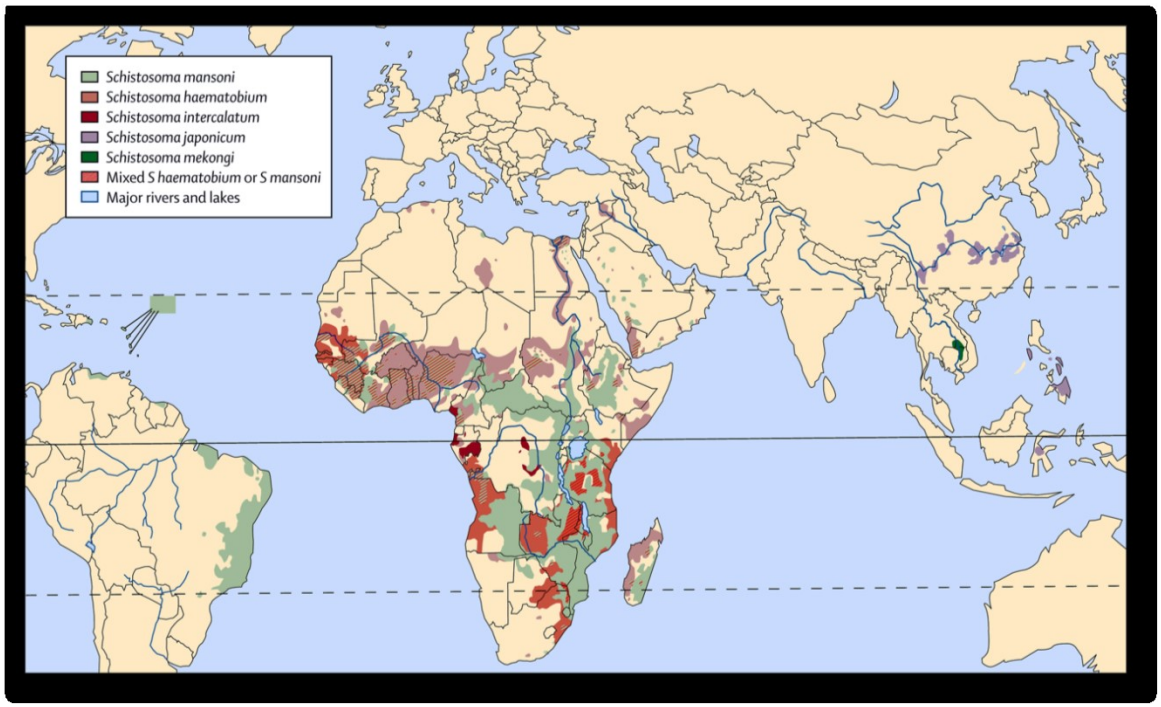


Figure 49: *geographical distribution of schistosomiasis* (adapted by Colley, D. G.; Bustinduy, A. L.; Secor, W. E.; King, C. H. Human Schistosomiasis. *The Lancet* **2014**, *383* (9936), 2253–2264.

[150] https://unitingtocombatntds.org/resources/who-roadmap-ntds/
[151] https://www.gatesfoundation.org/What-We-Do/Global-Health/Neglected-Tropical-Diseases
[152] https://www.neglecteddiseases.gov/about/what-we-do
[153] https://www.gov.uk/government/organisations/department-for-international-development
[154] https://www.cdc.gov/globalhealth/ntd/diseases/schisto_burden.html
[155] http://www.who.int/en/news-room/fact-sheets/detail/schistosomiasis

There are two main forms of schistosomiasis, intestinal and urogenital, caused by five different species of trematode worms that are diffused in different countries around the globe (see Table 13).

***Table 13**: geographical distribution of schistosomiasis.*

	Species	Geographical distribution
Intestinal schistosomiasis	*Schistosoma mansoni*	Africa, the Middle East, the Caribbean, Brazil, Venezuela and Suriname
	Schistosoma japonicum	China, Indonesia, the Philippines
	Schistosoma mekongi	Several districts of Cambodia and the Lao People's Democratic Republic
	Schistosoma guineensis and related *S. intercalatum*	Rain forest areas of central Africa
Urogenital schistosomiasis	*Schistosoma haematobium*	Africa, the Middle East, Corsica (France)

Life cycle of Schistosoma

Schistosomiasis, as other infectious diseases, is deeply affected by the pathogen life cycle (see Figure 50).

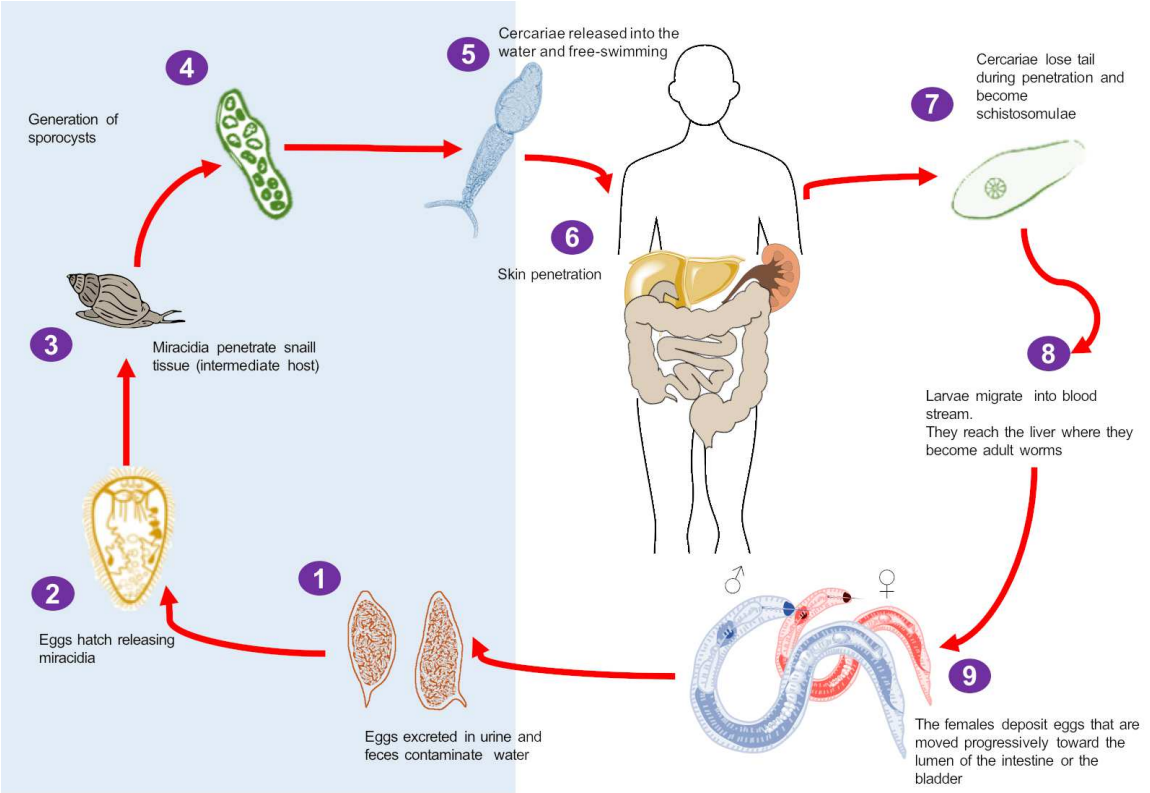


***Figure 50:** Schistosoma life cycle.*

Schistosoma eggs are eliminated in feces and urine of infected people. In poor countries with poor hygienic conditions and inappropriate sewage systems, infected water is released in the environment where people could be easily in contact with. In optimal conditions, eggs hatch releasing *miracidia* that can swim and penetrate the tissue of the intermediate host which is represented by specific snails. *Sporocysts* and then *cercariae* are produced in snails, where the last are released in water where they can swim. This step is crucial as

cercariae can infect humans simply by tissue contact; cercariae can penetrate human skin, losing their tail and becoming *schistosomulae.* They can enter the blood circulation reaching different organs: while *S. mansoni* and *japonicum* are found mainly in mesenteric veins of the intestine, the *S. hematobium* occurs mostly in the venous plexus of the bladder. In human body *schistosomulae* develop in adult worms and females can deposit eggs: some of the eggs will be excreted with feces and urine while others will be trapped in body tissues causing immune reactions and damage to the organs.
This disease is typical of rural communities and affects both adults and children. Women can be infected easily simply by doing houseworks (for example washing clothes with infected water) and they can develop a genital schistosomiasis which is considered a risk factor for HIV infection. Children can also be easily in contact with trematode worms by playing and swimming in the infested water. Symptoms of the disease are associated with the body's reaction to worms eggs, rather than with the presence of the worm itself. Eggs that are not excreted accumulate in the intestine or bladder causing inflammation and after years of infections also other organs can be affected by the parasite such as liver, spleen and lungs. Infected children can develop anemia, malnutrition and learning difficulties. If not treated, schistosomiasis can persist for years and develop chronic symptoms like abdominal pain, enlarged liver, blood in stool or urine and in case of chronic infections there is an increased risk of bladder cancer. Diagnosis of schistosomiasis can be made searching for eggs in stool or urine samples, but in less advanced infections a blood (serological) test is required[156].

4.1.3 Treatments available against schistosomiasis

Up to now no vaccines are available to treat schistosomiasis and the first line treatment is based on the drug praziquantel.

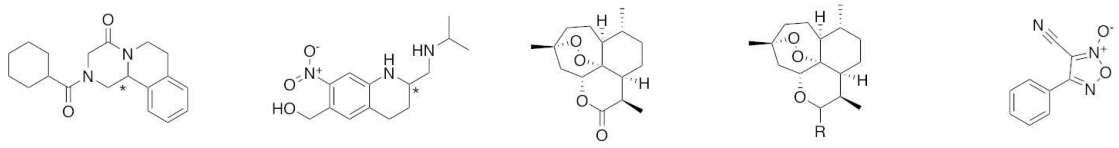


Figure 51: *molecules used to treat schistosomiasis.*

- Praziquantel

Praziquantel was firstly developed by scientists at Merck in 1970s who were interested in pyrazinoquinoline derivatives active as tranquillizers. The newly developed compounds achieved the desired effect at too high dosage, so they were no further advanced. The compounds were transferred at Bayer where they were tested in a veterinary screening. Among the 400 compounds tested, the most promising one was found to be praziquantel, which showed anthelmintic activity against broad-spectrum parasites.
It was only in 1977 that praziquantel activity against *S. mansoni* was discovered and after several tests it was approved for the treatment of human schistosomiasis by the early 1980s[157] with the name Biltricide®. In 1983 the Korean company Shin Poong developed a new synthetic patent for praziquantel and the drug's price decreased becoming an affordable treatment available also in poor countries.

[156] Colley, D. G.; Bustinduy, A. L.; Secor, W. E.; King, C. H. Human Schistosomiasis. *The Lancet* **2014**, *383* (9936), 2253–2264
[157] Andrews, P.; Thomas, H.; Pohlke, R.; Seubert, J. Praziquantel. *Med Res Rev* **1983**, *3* (2), 147–200.

Many factors have determined praziquantel's fortune: the affordable price, the fact that it is active against all species of Schistosoma, the oral delivery and its safety and efficacy that made it the primary treatment of schistosomiasis around the globe[158]. Praziquantel is included on the World Health Organization's List of Essential Medicines, which contains the most important medications needed in a basic health system. Praziquantel is effective against adult forms of all species of Schistosoma but, for unknown reasons, it is not so active toward young forms of the trematode. Today is the main drug to treat schistosomiasis but also taeniasis and cysticercosis. Praziquantel is administered as racemate: the eutomer has R configuration, while the S enantiomer is associated with side effects and it is primary responsible for the bitter taste of the pills, which can cause some problems especially for children. Strategies to synthetize enantiopure praziquantel in large scale to avoid compliance's issues have been reported by Todd et al.[159]

Mechanism of action

Even if praziquantel is the main treatment against schistosomiasis its precise mechanism of action has not been totally elucidated yet. Up to now two distinct mechanisms of action have been reported: ***i)*** muscular paralysis induction by Ca^{2+} influx and ***ii)*** tegument disruption, but since the molecular target of the drug is unknown it's difficult to understand if these effects are directly linked to the interaction with the target or if they are secondary effects[160].

The trematode muscle contraction is the first effect that can be observed after drug administration: worms lose their capacity of binding to tissues and they end up into the liver where an inflammatory reaction occurs, and they are eliminated. The other important effect of praziquantel is tegument disruption: this phenomenon is known as "vacuolation" or "blebbing" of the cutaneous tissues of the worm, allowing surface antigens exposition and successive host immune system response. It has been hypothesized that praziquantel acts on membrane protein complexes called voltage-operated Ca^{2+} channels (VOCC), altering their activity. These proteins regulate the intracellular levels of Ca^{2+} and they are essentials for contraction of the parasites among the other.

- Oxamniquine[161]

This drug is active only against *S. mansoni* and mostly against male than female parasites. It is common in Brazil where it is sold by Pfizer under the name of Mansil®. It is normally used as second choice treatment after praziquantel because of its central nervous system side effects, like hallucinations and convulsions. Oxamniquine causes worms to shift from the mesenteric veins to the liver where the male worms are retained; the female worms return to the mesentery but can no longer release eggs.

Oxamniquine is a prodrug and, after the adsorption by the parasite, is sulfonated by a sulfotransferase which is present only in the worms. The resulting ester is unstable, and it decomposes forming the reactive electrophilic molecule that acts alkylating DNA, proteins and other nucleophilic macromolecules causing the death of the parasite (see Figure 52).

---

[158] Caffrey, C. R. Schistosomiasis and Its Treatment. *Future Medicinal Chemistry* **2015**, *7* (6), 675–676.

[159] Woelfle, M.; Seerden, J.-P.; de Gooijer, J.; Pouwer, K.; Olliaro, P.; Todd, M. H. Resolution of Praziquantel. *PLoS Negl Trop Dis* **2011**, *5* (9).

[160] Cupit, P. M.; Cunningham, C. What Is the Mechanism of Action of Praziquantel and How Might Resistance Strike? *Future Medicinal Chemistry* **2015**, *7* (6), 701–705.

[161] da Silva, V. B. R.; Campos, B. R. K. L.; de Oliveira, J. F.; Decout, J.-L.; do Carmo Alves de Lima, M. Medicinal Chemistry of Antischistosomal Drugs: Praziquantel and Oxamniquine. *Bioorganic & Medicinal Chemistry* **2017**, *25* (13), 3259–3277.

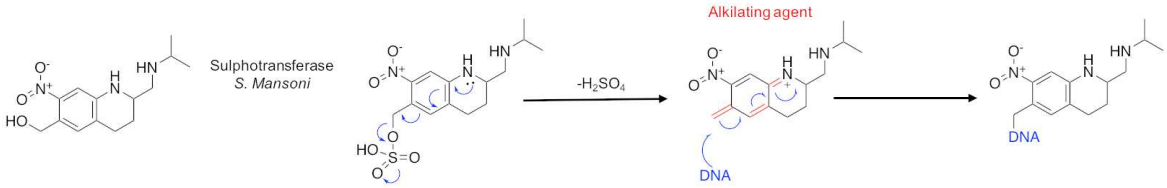


Figure 52: *oxamniquine mechanism of action.*

- Artemisinin derivatives and antimalaric compounds

Artemisinin and its derivatives (in particular artemether) are used as treatment against malaria, but they were found to be active also against schistosomiasis. The mechanism of action against the two targets is still unknown but it seems to involve a product of hemoglobin degradation that causes the formation of free radicals with alkylating properties[162]. Artemisinins are more active against immature forms of Schistosoma compared to the adult forms (the opposite of praziquantel) and attempts were made to combine the two drugs achieving more efficacy than using praziquantel alone. A limitation for the use of artemisinin is the possibility of developing resistance against plasmodia[163]. Mefloquine, another antimalarial compound, has shown to be also active against schistosomes[164]. Its mechanism of action seems to involve the inhibition of hemozoin formation. Hemozoin is a pigment derived by the digestive process of the parasites and essential for their survival.

- Furoxans

In vertebrates, enzymes glutathione reductase and thioredoxin reductase are essential for antioxidant defenses. Schistosomes, instead, rely on a single enzyme called thioredoxin-glutathione reductase (TGR) which is essential for parasite survival and it has been the target of several screening campaigns that led to the identification of oxadiazoles. Between them, furoxan showed activity both in vitro and in vivo[165].

- Vaccines

Even if no vaccine is available against schistosomiasis up to now, several efforts are made to try to develop a safe and effective one, and today different vaccines are in different stages of clinical trials. Series of proof-of-concept studies have demonstrated that mice and non-human primates immunized with radiation-attenuated cercariae were protected against Schistosoma percutaneous infections with an efficacy over 80%[166].Given these promising results, different companies and organizations are trying to develop vaccines, which mostly act preventing the schistosome infection or interrupting the parasite reproductive phase. There are four major vaccine candidates which are undergoing different stages of the clinical trials[13,167] (see Table 14).

---

[162] Caffrey, C. R. Chemotherapy of Schistosomiasis: Present and Future. *Current Opinion in Chemical Biology* **2007**, *11* (4), 433–439.

[163] Siqueira, L. da P.; Fontes, D. A. F.; Aguilera, C. S. B.; Timóteo, T. R. R.; Ângelos, M. A.; Silva, L. C. P. B. B.; de Melo, C. G.; Rolim, L. A.; da Silva, R. M. F.; Neto, P. J. R. Schistosomiasis: Drugs Used and Treatment Strategies. *Acta Tropica* **2017**, *176*, 179–187.

[164] Van Nassauw, L.; Toovey, S.; Van Op den Bosch, J.; Timmermans, J.-P.; Vercruysse, J. Schistosomicidal Activity of the Antimalarial Drug, Mefloquine, in Schistosoma Mansoni-Infected Mice. *Travel Med Infect Dis* **2008**, *6* (5), 253–258.

[165] Sayed, A. A.; Simeonov, A.; Thomas, C. J.; Inglese, J.; Austin, C. P.; Williams, D. L. Identification of Oxadiazoles as New Drug Leads for the Control of Schistosomiasis. *Nat Med* **2008**, *14* (4), 407–412

[166] Stutzer, C.; Richards, S. A.; Ferreira, M.; Baron, S.; Maritz-Olivier, C. Metazoan Parasite Vaccines: Present Status and Future Prospects. *Front. Cell. Infect. Microbiol.* **2018**, *8*.

[167] Merrifield, M.; Hotez, P. J.; Beaumier, C. M.; Gillespie, P.; Strych, U.; Hayward, T.; Bottazzi, M. E. Advancing a Vaccine to Prevent Human Schistosomiasis. *Vaccine* **2016**, *34* (26), 2988–2991.

Table 14: *candidate vaccines against schistosomiasis.*

Vaccine	Species targeted	Antigens	Sponsor	Current status
Bilhvax	*S. haematobium*	*Sh*28GST (28 kDa recombinant glutathione-S-transferase)	Institut Pasteur and INSERM (France)	Phase I: Completed (ClinicalTrials.gov Identifier: NCT01512277) Phase II: Completed (not published) Phase III: Completed, results pending (ClinicalTrials.gov Identifier: NCT00870649)
Sm-TSP-2	*S. mansoni*	9 kDa recombinant tetraspanin	**Sponsor**: NIH/Baylor College of Medicine Vaccine and Treatment Evaluation Unit **Production**: Sabin Vaccine Institute Product Development	Phase Ia: Completed (ClinicalTrials.gov Identifier: NCT02337855) Phase Ib: In preliminary stages (ClinicalTrials.gov Identifier: NCT03110757)
Sm-14	*S. mansoni*	*Sm14*-FABP (14 kDa recombinant fatty acid binding protein)	Oswaldo Cruz Foundation	Phase I: Completed (ClinicalTrials.gov Identifier: NCT01154049) Phase II: In progress (ClinicalTrials.gov Identifier: NCT03041766)
SchistoShield®	*S. mansoni*	*Sm-p80* (Large subunit of calpain-neutral cysteine protease)	Texas Tech University Health Sciences Centre	Pre-clinical completed Phase I: on-going

Institut Pasteur and INSERM (France) have developed Bilhvax, a vaccine against *S. haematobium* which is responsible for urinary schistosomiasis and it has completed phase III of clinical trials. The molecule is a schistosome glutathione S-transferase common to both schistosomula and adult stages of the trematode and it seems to be involved in parasite cellular detoxification and host immune regulation. The phase III trial studied the efficacy of the vaccine given in association with praziquantel in infected children, but no results were published up to now.

Three vaccines are being studied against *S. mansoni: Sm-TSP-2* is the extracellular domain of a *S. mansoni* membrane surface protein and it is currently on stage Ib showing promising results (the end of the study is expected in October 2018).

Sm-14 is a recombinant fatty acid binding protein currently on the final stages of phase II to assess the safety of the vaccine components (antigen and adjuvant) and the immunogenicity in adults.

SchistoShield® is another candidate vaccine which has completed the preclinical trials and is currently under stage I.

4.1.4 Resistance phenomena

Up to now no cases of resistant schistosomiasis have been reported in clinic. Resistance to praziquantel has been induced in laboratory in infected mice treated with sub-doses of the drug, but the mechanism of resistance is unknown[168]. Experiments have shown that some proteins involved in the transport of xenobiotics and toxins, called ABC transport proteins, are expressed in young worms exposed to praziquantel *in vitro,* giving some possible explanations about the molecular basis of resistance. The fact that no clinical resistance phenomena have been reported up to date could suggest that schistosomes have not reached a high evolutionary pressure yet allowing resistance to develop, but chances of resistance remains high. The worry on the insurgence of resistance has solid rational justifications and the search for new drugs is becoming more urgent[169]. On the other hand, some Schistosoma strains have been found resistant to oxamniquine: resistance to this drug is controlled by the absence of the enzyme sulfotransferase, which is necessary for the activation of the drug.

4.1.5 Pathogen Box

Pathogen Box is a collection of 400 diverse and drug-like molecules active against a variety of neglected diseases:

- Ascariasis caused by the infection of a parasitic worm *Ascaris lumbricoides*
- Buruli ulcer caused by *Mycobacterium ulcerans*
- Chagas disease caused by *Trypanosoma Cruzi*
- Cryptosporidiosis
- Hookworm
- Human African Trypanosomiasis (sleeping sickness)
- Visceral and cutaneous Leishmaniasis
- Lymphatic Filariasis caused by roundworms
- Malaria caused by *Plasmodium* family
- Onchocerciasis (river blindness)
- Schistosomiasis
- Trichuriasis
- Tuberculosis

The Pathogen Box is developed by MMV (Medicines for Malaria Venture), a group of scientists all around the globe with the aim of reducing the burden of malaria by discovering new drugs. In 2015 they launched an initiative with the aim to discover drugs for different neglected diseases and they developed the Pathogen Box.
The molecules belonging to the Pathogen Box were chosen by experienced medicinal chemists to ensure a collection of promising compounds suitable for drug discovery. The Pathogen Box is available free of charge, with the condition that all the data collected from screening would be published and shared in the public domain to boost neglected tropical disease research of new drugs. The aim of this project is the collection of the largest amount of data for these compounds aiming to contribute to the eradication of neglected diseases all around the globe[170].

---

[168] Fallon, P. G. Schistosome Resistance to Praziquantel. *Drug Resistance Updates* **1998**, *1* (4), 236–241.
[169] Cioli, D.; Pica-Mattoccia, L.; Basso, A.; Guidi, A. Schistosomiasis Control: Praziquantel Forever? *Molecular and Biochemical Parasitology* **2014**, *195* (1), 23–29.
[170] https://www.pathogenbox.org

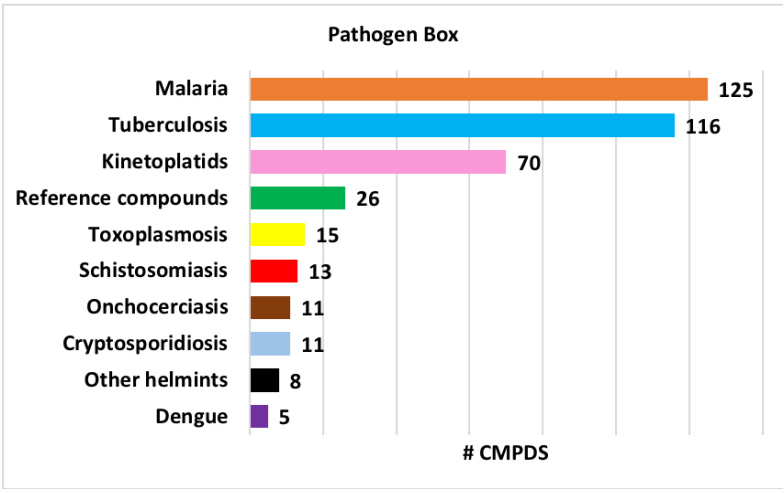


***Figure 53**:Pathogen Box composition. Compounds are classified according to the affected pathology.*

As reported in Figure 53, different classes of compounds are contained in the Pathogen Box. The largest disease set is the malaria one, with 125 compounds that have been reported active against this disease. A similar number has been identified for tuberculosis. Kinetoplastids disease set is composed by 70 compounds while the helmints disease set, formed by compounds known to be active against schistosomiasis, onchocerciasis, hookworm, lymphatic filariasis and trichuriusiasis, counts 32 compounds. Dengue virus set comprises 5 compounds, compounds known to be active against toxoplasmosis are 15 and cryptosporidiosis active compounds counts 11 molecules. The reference compounds disease set comprises molecules that are already used in therapy: some members of this group are mefloquine, a well know antimalaric compound, the antitubercular drug rifampicin, the anthelmintic drug praziquantel, the antibiotic streptomycin and the antifungal agent posaconazole, just to name few.

4.2 Aim of the work

Schistosomiasis is a tropical neglected disease that affects millions of people all over the world. Praziquantel represents the first line treatment against this parasitosis but resistance phenomena could be easily developed. Given the urgent need of new drugs able to treat schistosomiasis, an experimental screening of the Pathogen Box against *Schistosoma Mansoni*, one of the most pathogenic species of schistosomes, was carried out in three organizations: University of California San Diego (UCSD), Swiss Tropical and Public Health Institute (STPH) and Fiocruz Foundation in Brazil. A statistical analysis on the collection of biological data obtained by the screening was then performed to gain insight on the coherence of the activity data and to identify the most promising compounds.

4.3 Results and discussion

4.3.1 Experimental screening of the Pathogen Box

The 400 known bioactive compounds belonging to the Pathogen Box were tested for their activity against *Schistosoma mansoni.* Experimental assays were carried out by three organizations: University of California San Diego (UCSD), Swiss Tropical and Public Health Institute (STPH) and Fiocruz Foundation in Brazil. Compounds were given at the single dose of 10 µM into schistosomule growth medium and in a range of three days the effects on the pathogen motility were registered. UCSD and STPH assess worm response every 24 h, while Fiocruz only after 72 hours.

The experimental assays were performed using different protocols: UCSD and STPH assessed worm motility using an observational assay, while FIOCRUZ measured the % of cell viability at 72 h using a tetrazolic salt as metabolic indicator called XTT. In this assay only viable worms metabolize XTT through a reduction reaction into a water soluble product producing a change in color that can be measured spectrophotometrically.
Assays results were registered on different scales:

- UCSD used an observational-based scoring system from 0 to 4 (0 is a NON-hit while 4 is the strongest hit);
- STPH used an observational-based scoring system with fractional values from 0 to 1, where a value ≥0.5 is considered a hit compound;
- FIOCRUZ used a % value and a cell viability ≤50% was considered as a hit compound.

The activity data obtained by screening of the Pathogen Box by University of California San Diego (UCSD), Swiss Tropical and Public Health Institute (STPH) and Fiocruz Foundation are presented in the Appendix Table A (section 4.6.1). Data analysis were performed to collect as many information as possible about the biological activity of the dataset against schistosomes. Different aspects were evaluated, and they are presented in the following sections.

4.3.2 Degree of coherence between the different screening organizations

Since three different organizations (UCSD, STPH and Fiocruz) have tested the same set of compounds against the same target using different experimental assays (motility assay for UCSD and STPH and viability assay for Fiocruz) the evaluation of the consistency of the results obtained, meaning the reproducibility of the activity data collected for the same compounds by UCSD, STPH and Fiocruz, was performed. To answer this question, two different approaches were exploited, based on the different types of data used to perform the analysis: raw activity data and transformed data.

Raw data
In Table 15 the number and relative percentage of compounds sharing the same behavior, thus considered as active or inactive by the three organizations, are displayed. The calculation was based on a reference scale of activity decided by the same organizations (see section 4.5.2). Over three days 70.5 % of the compounds shared the same behavior between the three organizations, meaning that they were found active or inactive by both UCSD, STPH and Fiocruz. The percentage reached 85.25% taking into account the results shared by UCSD and STPH only. A lower percentage was obtained when UCSD and STPH results were combined with Fiocruz.

Table 15: *degree of accordance between UCSD, STPH and Fiocruz compared to the biological activity results calculated at 24h, 48h, 72h and over the three days is reported. UCSD is labelled as U, STPH is labelled as S and Fiocruz organization is labelled as B. The table reports the number of compounds and the relative percentage for which according results between the organizations were found.*

Agree		**U+S+B**	**U+S**	**U+B**	**S+B**
Over 3 days	#cmpds	282	341	316	307
	%	**70.5**	**85.25**	**79**	**76.75**
@72h	#cmpds	295	347	331	312
	%	**73.75**	**86.75**	**82.75**	**78**
@48h	#cmpds	-	348	-	-
	%	-	**87**	-	-
@24h	#cmpds	-	350	-	-
	%	-	**87.5**	-	-

The comparison was also made at 72 hours, where 73.75% of compounds displayed similar results between the three organizations. UCSD results combined with STPH and Fiocruz activity reached more than 80% of the agreement, while STPH with Fiocruz displayed only 78% of coherence between data. A very high percentage of compounds that behave the same was found between UCSD and STPH at 24 and 48 hours (87.5% and 87%, respectively).

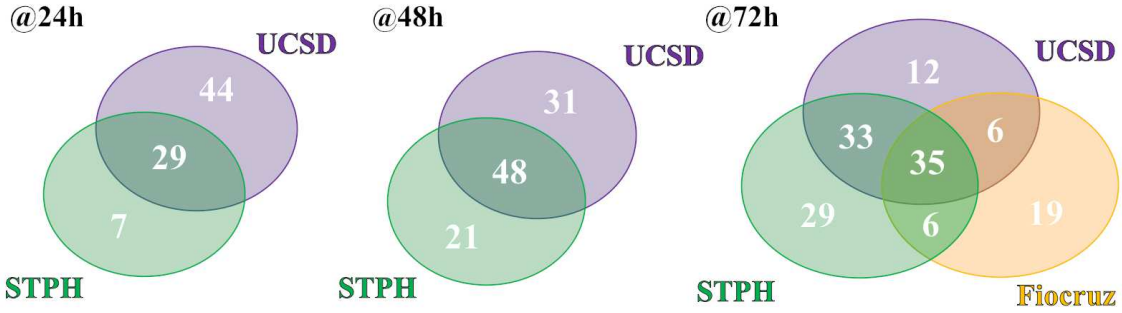


***Figure 54**:the number of actives shared and not between the three organizations is reported in the diagrams. UCSD is labelled as U and is represented by a purple circle, STPH is labelled as S and is represented by a green circle and Fiocruz organization is labelled as B and is represented by a yellow circle.*

In Figure 54, Venn diagrams display the common and non-common active compounds found between the three organizations at different times. At 72 hours only 35 compounds out of 400 were found as common actives between UCSD, STPH and Fiocruz. Additional six compounds were found in common between UCSD and Fiocruz, and STPH and Fiocruz, while more compounds (33 compounds) were found as common actives between UCSD and STPH. The number of inactive compounds shared by the three organization was 260, which is the 65% of the total number of compounds of the dataset (see Table 16).
The same diagrams were also built for data obtained at 24 and 48 hours by UCSD and STPH, highlighting a high number of compound with similar behavior (29 at 24 hours and 48 at 48 hours).

***Table 16:** the number of actives and inactives shared and not between the three organizations at different times are displayed.*

	@24h		@48h	
	# of cmpds	**% of cmpds**	**# of cmpds**	**% of cmpds**
Actives shared U+S	29	7.25	48	12
Actives U	44	11	31	7.75
Actives S	7	1.75	21	5.25
Inactives shared U+S	320	80	300	75

	@72h	
	# of cmpds	**% of cmpds**
Actives shared U+S+B	35	8.75
Actives shared U+S	33	8.25
Actives shared U+B	6	1.5
Actives shared S+B	6	1.5
Actives U	12	3
Actives S	29	7.25
Actives B	19	4.75
Inactives shared U+S+B	260	65

Transformed data

Transformed data were used to calculate the Pearson correlation and the respective value of the F distribution at different times. Table 17 reports a strong correlation between UCSD and STPH activity data with a correlation value of 0.68 at 72 hours: this means that the majority of the compounds showed a comparable activity in the tests performed by the two organizations at day three. A good correlation was also found between data collected by the same organizations at 24 and 48 hours. On the other hand, a moderate correlation was found taking into account UCSD and STPH data against Fiocruz (0.46 and 0.50, respectively).

Table 17: *Pearson coefficient and F distribution values calculated for UCSD, STPH and Fiocruz activity data.*

	Time	U vs S	U vs B	S vs B
Pearson coefficient	***72h***	0.68	0.46	0.50
	48h	0.63	-	-
	24h	0.61	-	-
F distribution	***72h***	6.39E-**56**	5.29E-**22**	5.57E-**27**
	48h	2.13E-**46**	-	-
	24h	8.69E-**42**	-	-

The value of the F distribution for each couple of data series was also calculated to evaluate the probability that the computed correlation occurred by chance. A very low probability that activity data are accidentally correlated was calculated for the couple UCSD and STPH biological data at different times: at 72 hour the value of the F distribution was power -56, while at 24 and 48 hours was -42 and -46, respectively. The probability that the correlation was obtained by chance increased looking at the result from UCSD and STPH together with Fiocruz: F distribution value for the couple UCSD and Fiocruz was power -22 and for the couple STPH and Fiocruz was power -27.

A PCA analysis was also performed to gain insight into the activity data collected by the three organizations by simplifying the variable space. The two first components, PC1 and PC2, accounted for almost the 90% of the variance of all quantified variables (70 and 19.5%, respectively).

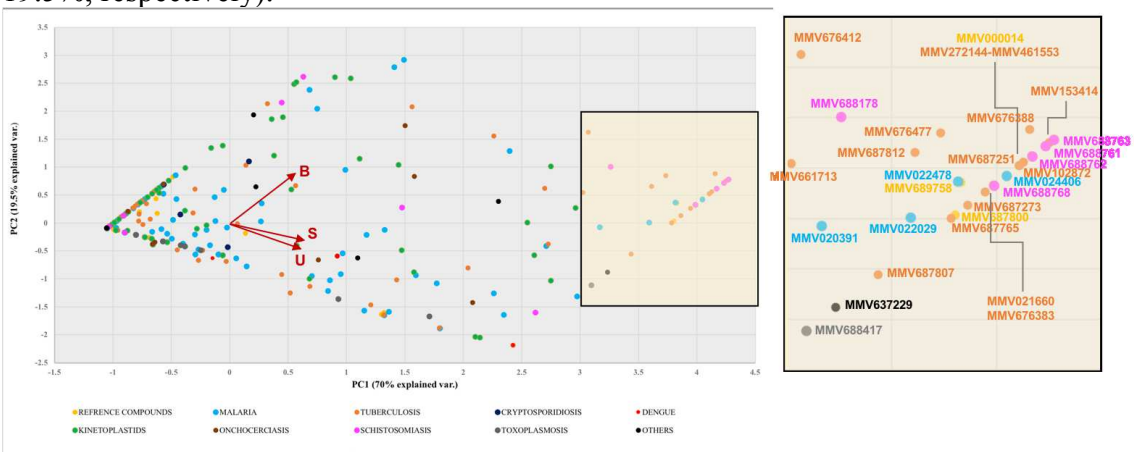


Figure 55: *projections of the scores of the 400 compounds of the Pathogen Box on the two principal components (PC1 and PC2). The three variables U, S and F are represented as red arrows. Each compound is represented by a dot colored according to the disease set. The yellow square highlights the distribution in the space of the new variables of the most active compounds of the dataset. The disease set called 'others' comprises compounds belonging to the Hookworm, Lymphatic Filariasis, Wolbachia LF and Trichuriusiasis set.*

Figure 55 shows the results of the PCA performed on the transformed data. Each compound is represented by a new set of values called 'scores' which place the compound in the space of the new variables. On the left side of the plot the majority of the compounds are represented: they are the inactive compounds. The most active compounds are displayed on the right part of the plot and reported in the yellow square. The most active compounds belong to different disease sets. In particular, a large number of anti-tubercular molecules were found active against schistosomes (orange dots), followed by compounds active against malaria (cyan dots) and schistosomiasis (magenta dots). Three reference compounds are reported in this plot (yellow dots): MMVV000014 (mefloquine, an antimalarial drug), MMV687800 (clofazimine, used to treat leprosy) and MMV689758 (bedaquiline, an anti-tubercular drug).
The three variables, UCSD, STPH and Fiocruz are plotted in the new cartesian space as red arrows starting from the origin of the scores plot (coordinates 0,0). The directions of the arrows explain how similar or dissimilar the variables are. UCSD and STPH variables share the same direction, while Fiocruz variable is almost orthogonal to the first two. This means that UCSD and STPH carry similar information which is different from the information carried by Fiocruz.

Taken together, these results showed that activity data collected by UCSD and STPH displayed a higher similarity compared to the data collected by Fiocruz. The number of compounds found in common between UCSD and STPH was higher compared to the compounds shared with Fiocruz by the same two organizations. Statistical analysis, in particular the calculation of the Pearson correlation, highlighted a good correlation between activity data collected by UCSD and STPH, with a low probability that the result was obtained by chance. On the other hand, low values of correlation were found when UCSD and STPH activity data were combined with the one collected by Fiocruz. Principal component analysis confirmed the previous results.
The fact that activity data collected by UCSD and STPH were more coherent compared to those obtained by Fiocruz highlighted that different experimental assays can give different results. In fact, UCSD and STPH evaluated the motility of schistosomes exposed to the Pathogen Box compounds, while Fiocruz evaluated the vitality of the trematodes using a colorimetric assay. This doesn't mean that Fiocruz results were wrong, but only that the activity data were collected in a different way. Moreover, it could be argued that observational assays made by UCSD and STPH is more biased than viability assays. In fact, in an observational assay the scientist observes worms motility assigning a score arbitrarily, while the results of a viability assay are collected recording a specific numerical value from the spectrophotometer.

4.3.3 Reference compound set analysis

Reference compound disease set of the Pathogen Box contains 26 molecules already known and used as drugs to treat different pathogens. The activity data obtained by the three organizations for this set were used to compare the results obtained by the three organizations, highlighting similarities and dissimilarities. Transformed activity data were used for the comparison.

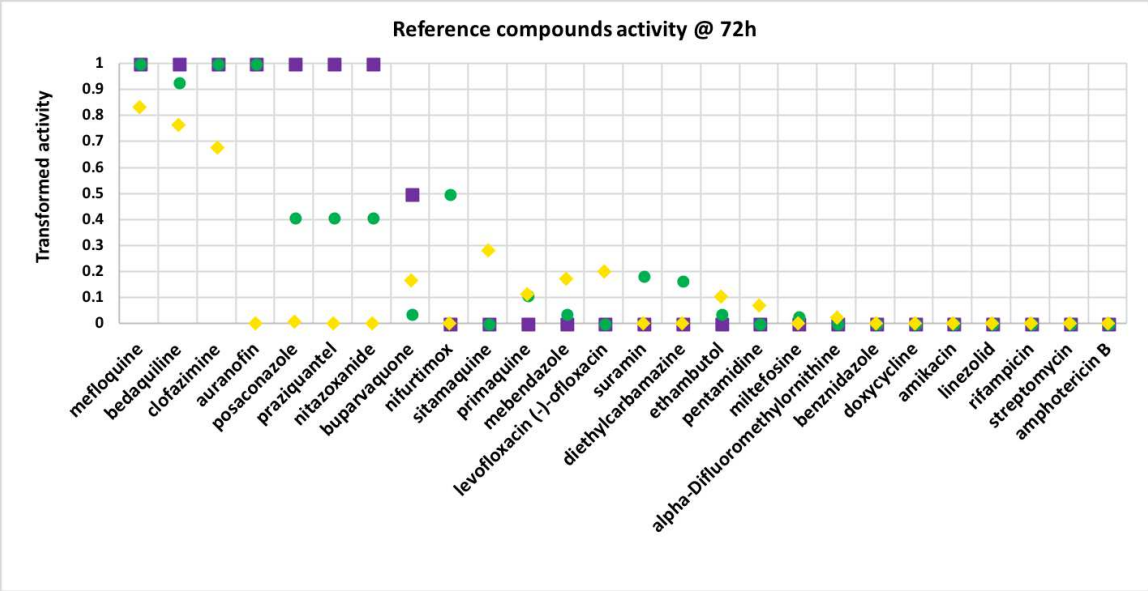


***Figure 56**: this graph shows the transformed activity values for each reference compound of the Pathogen Box collected by the three organizations at 72 hours. The activity value range spans from 0, meaning inactive compound, to 1, meaning hit compound. Each reference compound shows three different activity values: UCSD activity is represented by a purple square, STPH activity by a green circle and Fiocruz activity is represented by a yellow diamond.*

In Figure 56 it can be observed that the only reference compounds found active by all the three organizations were mefloquine, bedaquiline and clofazimine. Mefloquine is an antimalarial compound which was already found active against schistosomes in 2008[171], bedaquiline is an anti-tubercular drug and clofazimine is used against leprosy.
Auranofin, praziquantel, posaconazole and nitazoxanide were found active by UCSD. STPH recorded some activity for these compounds, while they were found inactive by Fiocruz. Auranofin is a gold complex compound used to treat arthritis, but in the last years scientific evidence has demonstrated its potential against other pathologies. Now in Brazil a clinical trial is ongoing to verify its activity against HIV (ClinicalTrials.gov Identifier: NCT02961829). In 2012 a screening campaign identified auranofin as a new lead compound to treat amebiasis[172] while in 2015 its anti-bacterial properties[173] were reported. In 2014 Park et al.[174] highlighted the anticancer properties of the same compound in ovarian cancer.
Praziquantel is the common treatment against schistosomiasis and it's curious the fact that it was found active by just two organization, UCSD and STPH, and not by Fiocruz. This could be due to the different assay conditions and endpoints used by the three organizations: in fact, Fiocruz assay verify worms vitality, while UCSD and STPH worms motility. Even

[171] Van Nassauw, L.; Toovey, S.; Van Op den Bosch, J.; Timmermans, J.-P.; Vercruysse, J. Schistosomicidal Activity of the Antimalarial Drug, Mefloquine, in Schistosoma Mansoni-Infected Mice. *Travel Med Infect Dis* **2008**, *6* (5), 253–258.
[172] Debnath, A.; Parsonage, D.; Andrade, R. M.; He, C.; Cobo, E. R.; Hirata, K.; Chen, S.; García-Rivera, G.; Orozco, E.; Martínez, M. B.; et al. A High Throughput Drug Screen for Entamoeba Histolytica Identifies a New Lead and Target. *Nat Med* **2012**, *18* (6), 956–960.
[173] Harbut, M. B.; Vilchèze, C.; Luo, X.; Hensler, M. E.; Guo, H.; Yang, B.; Chatterjee, A. K.; Nizet, V.; Jacobs, W. R.; Schultz, P. G.; et al. Auranofin Exerts Broad-Spectrum Bactericidal Activities by Targeting Thiol-Redox Homeostasis. *Proc Natl Acad Sci U S A* **2015**, *112* (14), 4453–4458.
[174] Park, S.-H.; Lee, J. H.; Berek, J. S.; Hu, M. C.-T. Auranofin Displays Anticancer Activity against Ovarian Cancer Cells through FOXO3 Activation Independent of P53. *Int. J. Oncol.* **2014**, *45* (4), 1691–1698.

if the mechanism of praziquantel has not been completely elucidated yet, it is known that it can act paralyzing the worm in the contracted state without killing it, so the biological result of the assay can reflect this mechanism of action. UCSD and STPH motility assay recorded an activity of the compound, since worms motility was negatively affected, but probably the worm was still alive thus leading to the classification as inactive by Fiocruz.
Posaconazole is a triazole anti-fungal drug that act by blocking ergosterol synthesis, a component of cell membranes in fungi and protozoa, while nitazoxanide is an anti-protozoal drug used to treat infections by *Cryptosporidium parvum* and *Giardia lamblia*.
The anti-protozoal drug buparvaquone was found active only by UCSD and nifurtimox, used to treat Chagas disease, was found active only by STPH.
Mebendazole, an anti-helminthic compound, was found inactive by all the three organizations.
The only anti-bacterial drug that was found active against schistosomes was bedaquiline which is known for its inhibitory activity against ATP-synthase. Antibacterial molecules known to inhibits protein synthesis (doxycycline, amikacin, linezolid and streptomycin) were found as inactive compounds against schistosomes together with rifampicin, an RNA polymerase inhibitor), ethambutol (inhibitor of an enzyme involved in the synthesis of mycobacterium cell wall) and levofloxacin (fluoroquinolone active against topoisomerase).

4.3.4 Schistosomiasis disease set

The schistosomiasis disease set comprises 13 compounds reported to be active against schistosomes in literature. The screening performed by the three organizations found that for some of these compounds the activity was confirmed: compound ID MMV688768, MMV688761, MMV688762, MMV688763 and MMV688178 resulted active against schistosomes in this screening campaign.
Four compounds were reported to be inactive against trematodes: compound ID MMV1198433, MMV676536, MMV688552 and MMV688771.
MMV676382 was reported active by UCSD and STPH, but not by Fiocruz.
MMV688270 was reported active by UCSD and FIOCRUZ, but not by STPH.
MMV688313 and MMV688766 were reported to be active only by Fiocruz.
The chemical structures of these compounds are reported in Figure 57. Looking at the structures it is difficult to note similarities between them that can help to rationalize their activity or inactivity in the assays.

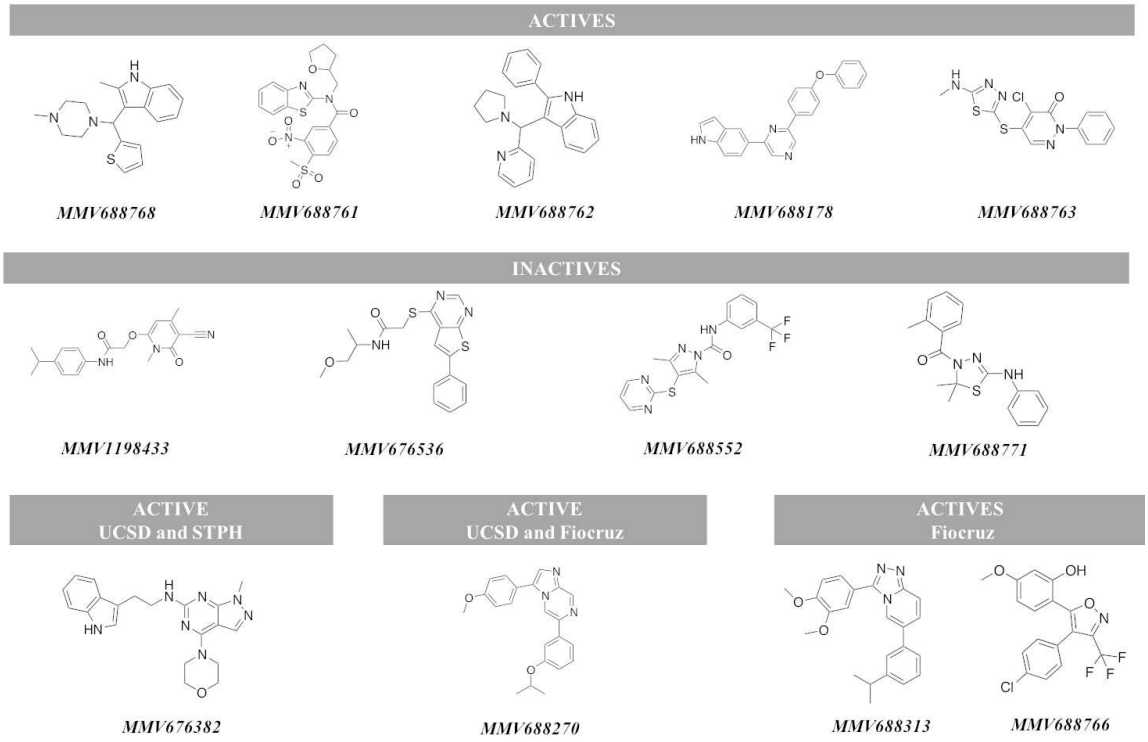


Figure 57:*schistosomiasis disease set compounds.*

In conclusion the assays performed by the three organizations have failed to identify 4 compounds reported to be active against schistosomes in previous tests, and other compounds have been reported as actives only by UCSD and Fiocruz.
It has to be pointed out that these compounds could have been previously tested in parasites at different stages of their life cycle (larval and adult) and it is experimentally proven that testing the same compounds on different parasite' life stages could affect the results. Some compounds that display activity only at the adult stage of the parasite could be inactive when the trematode is in its larval or juvenile stage. The same could be true for parasites in their juvenile stage and so on.

4.3.5 Hit compounds evaluation

The average activity was calculated for each of the 400 compounds of the Pathogen Box using transformed activity data at 72 hours. The average activity was used to rank the compounds. Figure 58 displays the first 50 compounds classified starting from the most active ones.
A clustering by structural similarity was also performed to highlight the presence of pattern of chemical classes of compounds that display activity against the target (see section 4.5.5). The 50 most active compounds of the dataset belong to different disease sets: a large number of them is part of the tuberculosis disease set (orange dots) highlighting the fact that this class of compounds in some way can inhibit schistosomes activity. Similarity calculations and cluster analysis revealed some structural occurrences between the most active compounds of the dataset. Compounds MMV676388 and MMV272144 share the same cluster displaying a phenyl tetrazole ring attached to a sulfone group (see Figure 59 cyan square). Compounds MMV153413 and MMV461553 share the same benzamido-tiofene scaffold and they were clustered together as reported in Figure 59 (yellow square). The schistosome disease set of the Pathogen Box contained 32 compounds known to be active against schistosomes and other helminths and some of them were confirmed to be

actives against schistosomes (magenta dots). In particular compound MMV688761, MMV688763, MMV688762 and MMV688768.
Some compounds active against malaria displayed also some activity against schistosomes (cyan dots): mefloquine was already known having an antischistosomal activity and it was confirmed as a good inhibitor by the three organizations. Other antimalarial compounds displayed some activity like MMV024406, MMV022478, MMV022029, MMV020391 and MMV023985. Moreover, compounds MMV022478 and MMV023985 share the same pyrazolopyrimidine scaffold and were clustered together, while MMV022029 and MMV020391 show a benzene sulfonamide moiety (see Figure 59 pink and green square). Some compounds belonging to the kinetoplastids disease set were found also as actives (green dots). Compounds belonging to the cryptosporidiosis disease display a very low activity recorded by the three organizations.

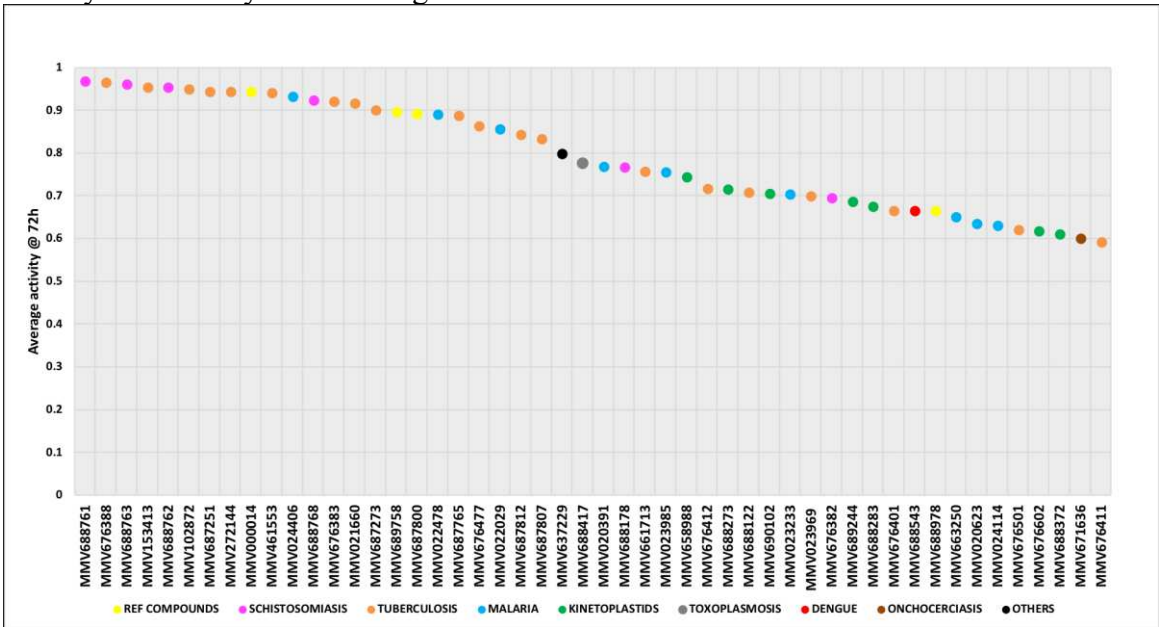


Figure 58: *the plot shows the first 50 compounds of the Pathogen Box in decreasing order of average transformed activity calculated at 72 hours. Compounds are displayed using different colors basing on the different disease set they belong to. The name for the reference compounds present in the plot is also displayed.*

Cluster analysis revealed also that some actives belonging to different disease sets displayed some structural similarities too: for example, compound ID MMV688762 and reference compound mefloquine were found in the same cluster. Compounds MMV021660 and MMV024406 were clustered together probably because of the presence of nitrogen atoms in their structure.

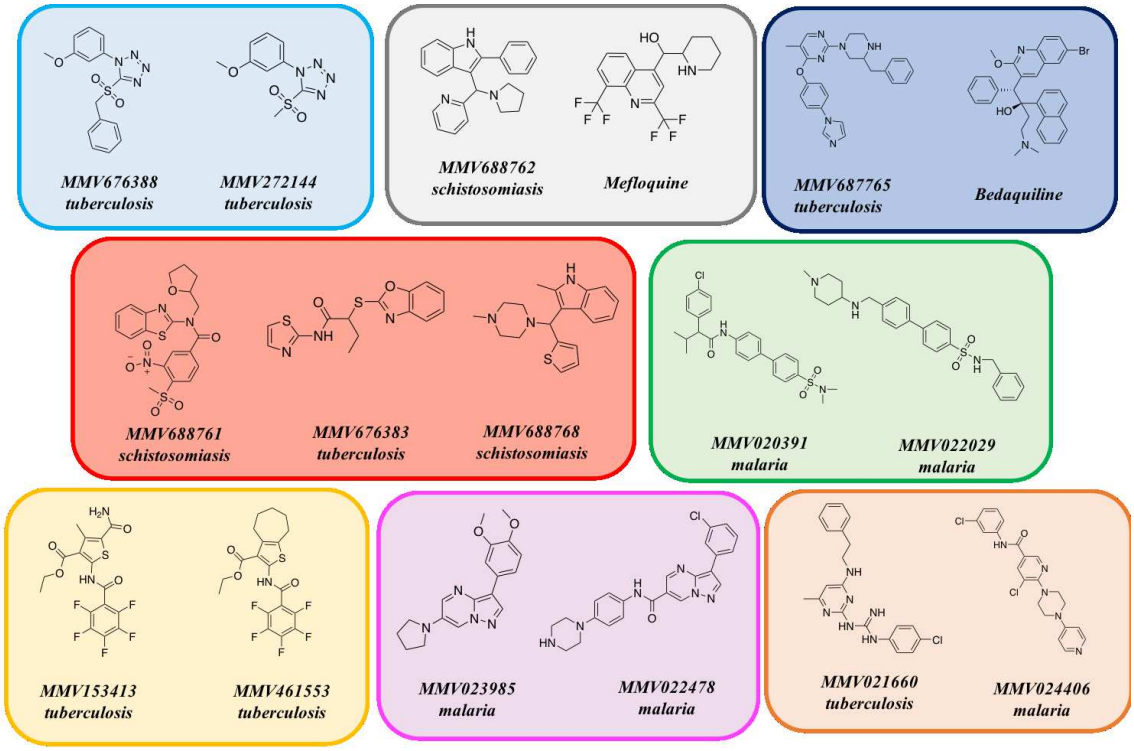


Figure 59: *some hit compounds grouped on the basis of structural similarities.*

Compounds MMV688761, MMV676383 and MMV688767 share the presence of thiazole or tiofene rings in their structure. Compounds MMV687765 and bedaquiline were also found in the same cluster.

Some hit compounds that were active at 72 hours were found highly active already in the first day (activity data from UCSD and STPH), like MMV688761, MMV688763 and MMV688762 of the schistosomiasis disease set, and MMV676388, MMV153413, MMV687251, MMV272144 of the tuberculosis disease set (see appendix Table A).

It is important also to highlight that some compounds were found active only by Fiocruz, but they were considered as inactives for both UCSD and STPH on each of the three days of the experiment. This result points out again the fact that different experimental conditions can affect the results of the screening and it could be interesting to investigate in detail the mechanism of these compounds.

4.3.6 Time-dependent activity

UCSD and STPH evaluated compound activity against schistosomes every 24 hours for three days, collecting for each compound three activity data at 24, 48 and 72 hours. The raw activity data collected allow us to investigate the time-dependency activity for each compound (see section 4.5.4).

Table 18: *the left table shows the number of active, inactive and transient compounds found by UCSD and STPH evaluating activity result scores every 24 hours. The transient compounds have been then classified into three activity categories: increasing activity compounds, decreasing activity compounds and peak activity compounds and the relative number of compounds for each category is displayed on the right.*

3 days	#U cmpds	#S cmpds	COMMON
Actives	56	36	27
Inactives	296	290	265
Transients	48	74	16

Transient compounds	#U cmpds	#S cmpds	COMMON
Increasing A	26	67	16
Decreasing A	13	-	-
Peak activity	9	7	-

Table 18 shows that UCSD found 56 compounds that were active for all the three days, while STPH 36. 27 of these compounds were found as common actives in all the three days by both UCSD and STPH. Concerning the transient compounds, i.e., molecules which have shown a combination of activity and inactivity in the three days, UCSD counted 48 compounds, STPH 74 compounds, and 16 of them were found in common between the two organizations.
Transients compounds have been re-classified into three different categories: compounds that displayed an increasing activity over the three days (an example is given in Figure 60A), compounds that display a decreasing activity over the three days (an example is given in Figure 60B), and compounds that display a different pattern of activity/inactivity during the three days (an example is given in Figure 60C). The majority of the transient compounds found by STPH display an increasing activity (67 compounds), only 7 compounds display a peak activity and none of them recorded a decreasing activity. UCSD found 26 increasing activity compounds, 13 decreasing activity compounds and 9 with a peak activity.

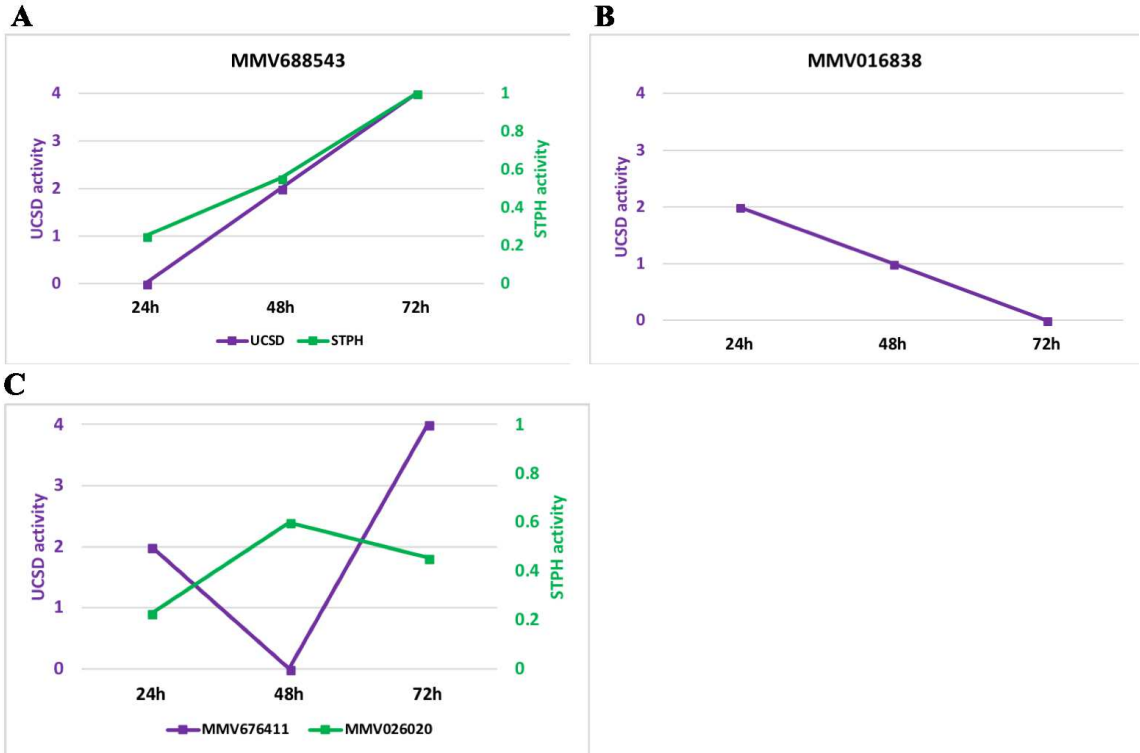


Figure 60: *these graphs shows three examples of transient activity compounds. Graph A shows compound MMV688543 which displays an increasing activity for both UCSD (purple line) and STPH (green line over the three days). Graph B shows compound MMV016838 that displays a decreasing activity as observed for UCSD. Graph C shows a peak activity for two different compounds: MMV676411 for UCSD (purple line) and MMV026020 for STPH (green line).*

The majority of the transient compounds display an increasing activity, likely due to the fact that the continue contact between the compound and the trematode over the three days, increases the probability of inhibiting trematode's motility Only UCSD recorded some compounds showing a decreasing activity: these compounds could undergo a metabolic decomposition that decrease their concentration in the plates, thus reducing their inhibitory capacity.
Compounds that were found having a transient activity could display a peculiar mechanism of action against the worms or they could be ineffective at a certain point, giving to schistosomes the possibility to recover their vitality.

4.3.7 DMPK properties analysis

Drug metabolism and pharmacokinetics properties (DMPK) are fundamental for a compound that is developed as a drug and they have to be taken into account when a new active compound is identified. A good drug is not only active against the target, but it has to display other properties that help the compound reaching the desired effect. For example, solubility and lipophilicity parameters should sustain proper drug distribution, effects of liver metabolism and renal excretion should also be optimal to avoid rapid degradation and excretion.
Pathogen Box provided some experimental DMPK properties for the compounds listed and an heatmap was generated to help the analysis of the data, in particular for the new hits. The complete heatmap with numerical values can be found in the Appendix table C.

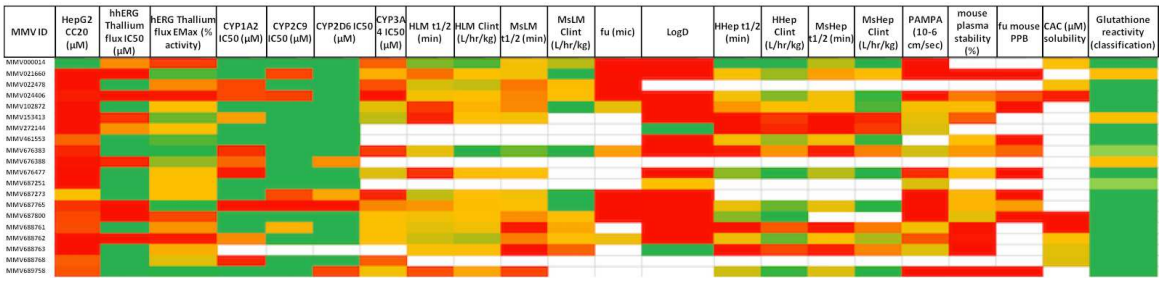


Figure 61: *this heatmap shows specific DMPK properties of 20 hits. Cells are color-coded from red (non-optimal DMPK property) to green (optimal DMPK property). White cells are non-available data.*

Cytotoxicity of the Pathogen Box compounds was tested against a human liver cancer cell line and the CC20 (concentration that reduced cell viability by 20%) was reported. Almost all the hits reported in Figure 61 displayed a low CC20 value, going from 30 nM (compound ID MMV676477) to 20 μM (MMV687273), highlighting the probability that these compounds can have some chances of toxicity in humans besides schistosomes, comprising the reference compounds bedaquiline (MMV689758 and CC20= 7.53 μM) and clofazimine (MMV687800 and CC20=6.59 μM). On the other hand, the reference compound mefloquine (MMV000014) displayed less chances of cytotoxicity with a CC20 in the high micromolar range (CC20=80 μM).
Activity on hERG channels was also provided since this protein is often the off-target of various drugs, including some antibacterials. The majority of the compounds display a high micromolar activity on the channel while only compounds MMV021660 and MMV024406 display an IC_{50} in the high nanomolar range. The reference compounds mefloquine and bedaquiline displayed low potency on hERG, while clofazimine retained a certain activity (IC_{50}=1.9 μM).
The half maximal inhibitory concentrations against different isoforms of the cytochrome family were also provided. Almost all the hits display low inhibitory activity on CYP2C9 and CYP2D6 (around 20 μM) with the exception of MMV687765 that display a nanomolar activity. The same compound shows also a certain activity on CYP1A2. Compounds display low activity against CYP3A4, except for MMV024406. The three reference compounds, mefloquine, bedaquiline and clofazimine were reported to have low activity on all CYP isoforms.
Microsomal and hepatic clearance and t ½ values, obtained both in mouse and human, were also reported. A low clearance and a high t½ are desirable properties to avoid a rapid elimination of the compound. The microsomal stability assay provided good results for

MMV676383 both in mouse and human. Compounds MMV676477 display good results in the hepatocyte stability assay.
Fraction of unbound compounds were also reported for both human microsomal proteins and mouse plasma. In this case the unbound fraction of compound is responsible for the biological effect: hit compounds tend to be heavily bound to plasma proteins in mouse, resulting in a low fraction of unbound compounds, while in human microsome the values of unbound fraction are slightly higher.
Plasma stability data were also provided by Pathogen Box allowing to know if compound concentration decreases when the molecule reaches the blood stream. The percentage of compound left in plasma after 4 hours was reported and some hit compounds showed complete degradation (MMV021660, MMV688761, MMV688763 and bedaquiline).
Lipophilicity values were also reported as LogD. In this case we were interested in compounds showing moderate lipophilicity to avoid BBB penetration with a LogD comprised between 1 and 3. The majority of the compounds displayed a slightly higher LogD with values of 4 or 5, except for compounds MMV272144 and MMV688763 which show a LogD in the desired range.
The results obtained by the critical aggregation concentration assay (CAC) were also provided to estimate the compound solubility. The CAC is the concentration at which the compound starts to aggregate: the poorly soluble compound was found to be MMV688761 (0.5 μM) and the most soluble compounds were MMV688763 and MMV688768 which display a CAC around 94 μM. Clofazimine displayed a low solubility.
Parallel artificial membrane permeability assay or PAMPA defines the capacity of a compound to passively cross an artificial membrane, simulating the passive transportation of the molecule across cell membranes. The value of apparent permeability (P_{app}) obtained from this assay could discriminate between low (< 10x10-6 cm/sec) and high permeability compounds (> 10x10-6 cm/sec). Hit compounds don't display a very high permeability, except for some compounds like MMV272144, MMV676383 and MMV687251. Their P_{app} is reported to be around 20x10-6 cm/sec. The three references mefloquine, clofazimine and bedaquiline displayed low permeability values in this assay.
All the hit compounds were reported to not react with glutathione or having just slightly chances.
In the end DMPK data provided us with some information about the metabolic and pharmacokinetics properties of the hit molecules found in the screening campaign. All the compounds, including mefloquine, bedaquiline and clofazimine of the reference disease set, seemed to display some criticisms. In particular, cytotoxicity and high lipophilicity were the two common non-optimal properties found for the hits. However, it has to be kept in mind that mefloquine, bedaquiline and clofazimine are commonly used as drugs to treat malaria, tuberculosis and leprosy, respectively, even if some DMPK data here reported are sub-optimal. The treatment of parasitic infections often involves drugs that are chosen as effective in killing pathogens, nonetheless they can show some deficiencies concerning their pharmacokinetics profile.
Praziquantel (MMV002529) is the first-line drug used to treat schistosomiasis and it displayed better DMPK properties (see appendix Table C) compared to the hit compounds: in particular, it showed lower cytotoxic effects (equivalent to that of mefloquine), a good hERG and CYP profile, a lipophilicity of 3.25, a good solubility and permeability, while its metabolic properties were in line with the other hit compounds.
This analysis confirmed that the new actives found could be used as new agents to treat schistosomiasis, as they showed comparable DMPK properties to approved drugs clofazimine, bedaquiline and mefloquine. Praziquantel DMPK profile revealed a similar

DMPK profile, slightly better in some properties, showing that DMPK properties for hit compounds could be improved to reach a similar profile.

4.4 Conclusions

Schistosomiasis is the second most diffuse parasitic disease after malaria[175] and the current treatment is based principally on just one drug, praziquantel, and in some cases oxamniquine which is used only for *S. mansoni* infections. Experimental evidences in in vitro tests have shown the possibility that trematodes could develop resistance towards these two drugs and new treatments are necessary.
The Pathogen Box represents a great opportunity for the identification of hit compounds and for progression into hit-to-lead research for a variety of pathogens.
In this work the 400 compounds contained in the Pathogen Box were screened against a juvenile stage of *S. mansoni*, one of the most pathogenic species across schistosomes, by three different organizations: University of California San Diego, Swiss Tropical and Public Health Institute and brasilian Fiocruz Foundation. The three organizations performed two different assays: UCSD and STPH used an observational assay that evaluates the effects on trematode motility. On the other hand, Fiocruz used a vitality assay, which counted the number of viable trematodes in presence of the compound.
The first aspect that was evaluated in this work was the coherence between the activity data collected by the three organizations that performed the screening campaign. Since the experimental protocols used were different and since the assays were carried out in three different environments it could be useful to investigate how the biological results were affected by these variables. Our statistical analysis reported more coherence between the data collected by UCSD and STPH compared to Fiocruz which collected different activity data. The number of common actives and inactives between UCSD and STPH was higher compared to the compounds shared with Fiocruz by the same two organizations. Moreover, the Pearson correlation highlighted a good correlation between activity data collected by UCSD and STPH while lower values of correlation were found when UCSD and STPH activity data were combined with the one collected by Fiocruz. A Principal Component Analysis was also performed, and the obtained results confirmed the previous observations. Taken together these results highlighted how screening results are affected by the choice of the experimental assay. Moreover, it has to be kept in mind that observational assays could be more biased than vitality assay.
The principal aim of this project was the discovery of compounds active against schistosomes. Three reference compounds were reported as actives against schistosomes by all the three organizations: mefloquine that was already known to act against trematodes, clofazimine and bedaquiline. Auranofin, praziquantel, posaconazole and nitazoxanide were found active by UCSD. STPH recorded some activity for these compounds while they were found inactive by Fiocruz.
The schistosomiasis disease set comprises 13 compounds and 5 were found active by all the three organizations, while 4 were reported as inactive. It should be noted that the antischistosomal activity of a compound can change between young and adult stage of the trematode and this could affect the assay results: in this case all the assays were performed on the schistosomulae stage of the trematode.
The screening campaign highlighted also good antischistosomal activity for compounds known to be active against other pathogens, especially belonging to the tuberculosis and malaria disease set. Some molecules known to be active as antimalarial compounds were

[175] https://www.cdc.gov/globalhealth/ntd/diseases/schisto_burden.html

already known to have some activity against schistosomes, for example mefloquine, but no molecules with anti-tubercular activity have been reported active against schistosomes up to now. Pathogen Box compounds known as antitubercular drugs reported to be active against schistosomes were MMV676388, MMV154313, MMV102872, MMV687251 and MMV272144.
Drug metabolism and pharmacokinetics properties provided by the Pathogen Box were also analyzed, especially for the hit compounds, to highlight non-optimal properties that should be optimized. The analysis reported that, even if some criticisms were reported for lipophilicity and cytotoxicity, the hit compounds displayed a similar DMPK profile to approved drugs mefloquine, bedaquiline and clofazimine not far from the DMPK profile of the first line antischistosomiasis drug Praziquantel.
Time-dependent activity was also evaluated taking into account the data collected by UCSD and STPH every 24 hour for three days. A certain number of compounds displayed a transient activity, meaning that their capacity to inhibit trematode worms could vary with time. 16 transient compounds were found in common between the two organizations. The majority of the transient compounds displayed an increasing activity, registered especially by STPH. Only 13 compounds displayed a decreasing activity and they were registered only by UCSD. Some compounds displayed a peak activity. This analysis could suggest a different mechanism of action for transient compounds and it could also highlight some metabolic implications. Some compounds showed a decreasing activity over time and this could be due to a metabolic degradation that affect these molecules.
In conclusion the Pathogen Box represented a good starting point for the identification of compounds active against schistosomiasis and the new hits found could represent attractive candidates for further drug development. The identification of their molecular target and the understanding of their mechanism of action could help in unraveling some of their vulnerabilities to achieve better activity.

4.5 Computational protocol

The experimental screening of the Pathogen Box by three different organizations (UCSD, STPH and Fiocruz) allowed to collect a large number of biological data for each of the 400 compounds listed (see Appendix table A).
Two different methodologies were used to analyze these data: in the first approach raw activity data were used for the analysis. The second approach considered “transformed” activity data: in this case data were put on the same scale, spanning from 0 to 1, using a mathematical transformation for a better comparison.
Data analysis was performed using Microsoft Office Excel and Molsoft ICM software[176,177]

4.5.1 Data transformation

The biological data collected by the three organizations, University of California San Diego (UCSD), Swiss Tropical and Public Health Institute of Health (STPH) and Fiocruz Foundation were mathematically transformed to obtain activity values placed in the same

[176] Abagyan, R., Totrov, M. & Kuznetsov, D. "ICM: a new method for structure modeling and design: Applications to docking and structure prediction from the distorted native conformation"*J. Comp. Chem.* **15,** 488-506 (1994)
[177] Abagyan, R. A., et al. "ICM Manual, 3.8" *MolSoft LLC, La Jolla, CA* (2009)

range for further analysis: in this case the values were set in a range that spans from 0 (inactive) to 1 (most active compound) as decimal numbers.
The raw data were firstly visually inspected to search for lacking values or inappropriate numbers. For the 24 and 48 hours of the STPH series, some raw data displayed negative activity: in the 24 hours series compounds MMV407539 and MMV024195 displayed an activity of -0.05, while in the 48 hours series compound MMV084864 showed an activity of -0.03 and compounds MMV676182 and MMV1019989 displayed an activity of -0.06. The activity value of these compounds was transformed to 0, meaning inactive compounds. Since the Fiocruz series spans from negative values, meaning highly active compounds, to high positive values, which means inactive compounds, the sign of the activity values was inverted. Sign inversion was necessary to obtain a series of numerical values with an increasing activity, consistent with that of the other two organizations.
Data were transformed to span a range from 0 to 1 using the following equation:

$$Transformeddata\frac{(x-median)}{(max(x)-median)}$$

Median was used given that the majority of compounds of the dataset were inactive; thus it was selected as threshold below which all the compounds were considered inactive. After the transformation the activity values for compounds showing negative numbers was transformed to 0 as they were considered inactive compounds.
The transformed data can be found in Appendix table A.

4.5.2 Degree of coherence between the different organizations

Coherence between data was investigated using both raw data and transformed data.

Raw data
The three organizations evaluated the activity of the compounds using different numerical ranges and they decided to define if a compound was active or not on the basis of the following reference scale:

Organization	Inactives	Actives
UCSD	<2	≥2
STPH	<0.5	≥0.5
FIOCRUZ	>50%	≤50%

The degree of accordance between the three different organizations was measured counting if a compound was considered active or inactive by all the three or two organizations at different times. The «agree» statement counts only compounds that are considered active or inactive by all the three (for 72 hours and over 3 days) or two organizations (for 24 and 48 hours). The number of compounds and the relative percentage was computed.
Another analysis was performed to evaluate the number of active and inactive compounds found in common or not between the three organizations.

Transformed data
Pearson correlation was calculated between couples of data series between the different organizations at different times to evaluate the degree of correlation. Transformed data were used for the analysis.
The value of F distribution was also calculated with the FDIST function in Excel to determine if the correlation values occurred by chance or not. A low value of F distribution

means that there is a low probability that the correlation between activity data is occurred by chance.
A Principal Component Analysis (PCA) was also performed on the three series (UCSD, STPH and Fiocruz) of transformed activity data at 72 hours using R software[178]. The R function *prcomp* was used to perform the calculation, and scaling of the data was applied to reach a unit variance. PCA scores for each compound are reported in Appendix table B. Practically, the original dataset (in this case the transformed activity data at 72 hours for all the compounds and for the three organizations) is mathematically transformed by an orthogonal transformation, to convert a set of possibly correlated variables in a set of uncorrelated variables called *principal components or eigenvectors*. For each eigenvector an eigenvalue is defined. Eigenvector represents a direction in the new space of variables, while the corresponding *eigenvalue* is a number expressing how much variance is present in the data in that direction. The first principal component is the eigenvector with the highest value of eigenvalues and so on. Eigenvector directions are orthogonal, and they correspond to variables not correlated with one another. Eigenvectors are useful to understand intuitively a complex group of data: they are the directions in which a bigger variance is present and consequently more information should be present. PCA helps to reduce the dimensionality of a dataset in its principal components, not considering non-significant or noisy parts. Principal components are correlated with originals variables through coefficients called *loadings*. Once that the dataset is transformed, it can be projected into a new cartesian space in which the variable with the highest variance (principal component 1 or PC1) is projected on the first axis, the variable with the second highest variance (PC2) is projected on the second axis, and so on. Each datum is represented by a new set of values called *scores*.

4.5.3 Ranking of the compounds in the Pathogen Box at 72 hours

The average activity was also calculated for each compound of the Pathogen Box at different times. An average activity was calculated for each compound at 72 hours using the transformed activity data: at 72 hours all the three series of transformed activity data were used for the calculation. The calculation of the average activity allows to rank all the compounds of the Pathogen Box.

4.5.4 Time-dependent activity

UCSD and STPH have collected activity data each 24 hours for three consecutive days and the time-dependent activity for each compound was evaluated using raw data. The compounds were first divided into three different groups on the basis of the reference scale of activity described in the section 4.2:

- *Actives*: compounds that retain activity over the three days (UCSD score of ≥2 and STPH score of ≥ 0.5).
- *Inactives*: compounds that display no activity over the three days (UCSD score of <2 and STPH score of <0.5).
- *Transients*: all the other compounds.

Transient compounds were then grouped into three different classes on the basis of their time-dependent behavior:

[178] R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

- *Increasing activity compounds*: they display an increasing activity over time (from inactives to actives).
- *Decreasing activity compounds*: they display a decreasing activity over time (from actives to inactives).
- *Peak activity compounds*: compounds that display a fluctuating behavior between activity and inactivity.

4.5.5 Chemical clustering

Chemical clustering was performed using ICM and 2D fingerprints were used as descriptors. Clustering is a procedure that groups molecule with a similar structure. Fingerprints are binary strings that identify a compound using a succession of "0" and "1" values reflecting the absence or presence of specific atoms or substructures in the molecule. A clustering threshold of 0.6 was applied.
Clustering is useful to highlight common chemical patterns in molecules especially when dealing with a large number of compounds: in the same cluster compounds with similar structures are grouped. If a cluster contains active molecules it could be useful to identify other molecules sharing the same scaffold, since it could be determinant for the activity on that specific target.

4.5.6 DMPK Heatmap

A heatmap was built using the DMPK properties provided by the Pathogen Box (see https://www.pathogenbox.org for details on the DMPK assays). Excel software was used to color heatmap cells. For each DMPK properties cell color changes on the basis of the relative range of each property. The color range spans from red (non-optimal DMPK property) to green (optimal DMPK property).
DMPK values can be found in the Appendix table C.

4.6 Appendix

4.6.1 Table A- data

This table contains the activity data collected for the 400 compounds of the Pathogen Box including compound ID, smiles, trivial name, if present, and disease set.
The raw activity data and the transformed data are reported: UCSD and STPH activity data are reported for 24, 48 and 72 hours. Fiocruz data are reported only at 72 hours. The raw data cells are color-coded using a scale from red (most active compounds) to green (inactive compounds) following the reference scale in section 4.5.2 of the computational protocol. Transformed data display a numerical range that goes from 0 (inactive compounds) to 1(hit compounds).

			RAW DATA							TRANSFORMED DATA						
			UCSD			STPH			FIOCRUZ	UCSD			STPH			FIOCRUZ
COMPOUND ID	TRIVIAL NAME	DISEASE SET	24 h	48 h	72 h	24 h	48 h	72 h	% VIABILITY XTT	24 h	48 h	72 h	24 h	48 h	72 h	% VIABILITY XTT
MMV000011	*Doxycycline*	REFERENCE COMPOUNDS	0	0	0	0.20	0.30	0.22	164.6	0.00	0.00	0.00	0.07	0.12	0.00	0.00
MMV000014	*Mefloquine*	REFERENCE COMPOUNDS	4	4	4	0.95	1.00	1.00	-0.1	1.00	1.00	1.00	0.95	1.00	1.00	0.83
MMV000023	*Primaquine*	REFERENCE COMPOUNDS	0	0	0	0.00	0.14	0.33	77.4	0.00	0.00	0.00	0.00	0.00	0.11	0.11
MMV000062	*Pentamidine*	REFERENCE COMPOUNDS	0	0	0	0.00	0.00	0.00	81.9	0.00	0.00	0.00	0.00	0.00	0.00	0.07
MMV000063	*Sitamaquine*	REFERENCE COMPOUNDS	0	0	0	0.10	0.14	0.17	59.3	0.00	0.00	0.00	0.00	0.00	0.00	0.28
MMV000858		MALARIA	1	0	0	0.05	0.08	0.11	75.2	0.25	0.00	0.00	0.00	0.00	0.00	0.13
MMV000907		MALARIA	0	0	0	0.18	0.20	0.13	93.3	0.00	0.00	0.00	0.05	0.00	0.00	0.00
MMV001059		MALARIA	2	4	4	0.24	0.28	0.75	95.7	0.50	1.00	1.00	0.12	0.10	0.67	0.00
MMV001493	*Isradipine*	ONCHOCERCIASIS	2	0	0	0.15	0.19	0.11	63.2	0.50	0.00	0.00	0.02	0.00	0.00	0.24
MMV001499	*Nifurtimox*	REFERENCE COMPOUNDS	0	0	0	0.30	0.28	0.63	223.9	0.00	0.00	0.00	0.19	0.10	0.50	0.00
MMV001561	*Fluoxetine*	KINETOPLASTIDS	0	0	0	0.05	0.08	0.33	20.4	0.00	0.00	0.00	0.00	0.00	0.11	0.64
MMV001625	*alpha-Difluoromethylornithine*	REFERENCE COMPOUNDS	0	0	0	0.15	0.14	0.22	87.1	0.00	0.00	0.00	0.02	0.00	0.00	0.02
MMV002529	*Praziquantel*	REFERENCE COMPOUNDS	4	4	4	0.30	0.35	0.56	97.3	1.00	1.00	1.00	0.19	0.19	0.41	0.00
MMV002816	*Diethylcarbamazine*	REFERENCE COMPOUNDS	0	0	0	0.15	0.22	0.38	95.5	0.00	0.00	0.00	0.02	0.03	0.17	0.00
MMV002817	*Iodoquinol*	ONCHOCERCIASIS	2	1	2	0.10	0.14	0.33	6.6	0.50	0.25	0.50	0.00	0.00	0.11	0.77
MMV003152	*Mebendazole*	REFERENCE COMPOUNDS	0	0	0	0.15	0.19	0.28	71.1	0.00	0.00	0.00	0.02	0.00	0.04	0.17
MMV003270	*Zoxazolamine*	HOOKWORM	0	0	0	1.00	1.00	1.00	76.5	0.00	0.00	0.00	1.00	1.00	1.00	0.12
MMV004168		KINETOPLASTIDS	0	0	0	0.45	0.05	0.26	67.9	0.00	0.00	0.00	0.37	0.00	0.02	0.20
MMV006239		MALARIA	0	0	0	0.05	0.08	0.17	109.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV006372		MALARIA	1	0	1	0.25	0.22	0.33	72.5	0.25	0.00	0.25	0.13	0.03	0.11	0.16
MMV006741		MALARIA	0	0	0	0.10	0.14	0.06	86.5	0.00	0.00	0.00	0.00	0.00	0.00	0.03
MMV006833		MALARIA	0	0	0	0.10	0.19	0.21	111.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV006901		MALARIA	1	0	0	0.18	0.35	0.45	-13.7	0.25	0.00	0.00	0.05	0.19	0.27	0.96
MMV007133		MALARIA	1	0	0	0.14	0.44	0.25	106.4	0.25	0.00	0.00	0.00	0.31	0.00	0.00
MMV007471		MALARIA	0	0	0	0.10	0.19	0.15	152.8	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV007625		MALARIA	0	0	0	0.10	0.14	0.15	135.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV007638		MALARIA	0	0	1	0.10	0.08	0.11	61.8	0.00	0.00	0.25	0.00	0.00	0.00	0.26
MMV007803		MALARIA	1	0	1	0.41	0.44	0.56	113.9	0.25	0.00	0.25	0.31	0.31	0.42	0.00
MMV007920		MALARIA	0	0	0	0.09	0.05	0.21	64.2	0.00	0.00	0.00	0.00	0.00	0.00	0.24
MMV008439		MALARIA	0	0	1	0.10	0.06	0.00	119.9	0.00	0.00	0.25	0.00	0.00	0.00	0.00
MMV009054		MALARIA	0	0	0	0.18	0.10	0.13	91.7	0.00	0.00	0.00	0.05	0.00	0.00	0.00
MMV009135		MALARIA	0	0	0	0.14	0.11	0.25	116.8	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV010545		MALARIA	0	0	0	0.15	0.22	0.31	125.4	0.00	0.00	0.00	0.02	0.03	0.08	0.00

MMV010576	MALARIA	2	1	1	0.10	0.19	0.15	139.2	0.50	0.25	0.25	0.00	0.00	0.00	0.00
MMV010764	MALARIA	0	0	0	0.18	0.55	0.35	95.7	0.00	0.00	0.00	0.05	0.44	0.13	0.00
MMV011229	MALARIA	0	0	0	0.09	0.15	0.16	71.7	0.00	0.00	0.00	0.00	0.00	0.00	0.16
MMV011511	MALARIA	0	0	0	0.00	0.03	0.03	98.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV011691	MALARIA	0	0	2	0.30	0.44	0.69	128.6	0.00	0.00	0.50	0.19	0.31	0.58	0.00
MMV011765	MALARIA	0	1	2	0.46	0.78	0.94	74.0	0.00	0.25	0.50	0.37	0.72	0.92	0.14
MMV011903	MALARIA	3	4	4	0.27	0.40	0.28	50.8	0.75	1.00	1.00	0.16	0.25	0.04	0.36
MMV012074	TUBERCULOSIS	0	0	0	0.10	0.19	0.33	81.0	0.00	0.00	0.00	0.00	0.00	0.11	0.08
MMV016136	MALARIA	0	0	0	0.10	0.14	0.17	64.9	0.00	0.00	0.00	0.00	0.00	0.00	0.23
MMV016838	MALARIA	2	1	0	0.50	0.67	1.00	145.7	0.50	0.25	0.00	0.42	0.58	1.00	0.00
MMV019087	MALARIA	0	0	0	0.05	0.14	0.22	73.3	0.00	0.00	0.00	0.00	0.00	0.00	0.15
MMV019189	MALARIA	2	0	0	0.00	0.14	0.22	125.5	0.50	0.00	0.00	0.00	0.00	0.00	0.00
MMV019234	MALARIA	0	0	0	0.10	0.14	0.17	65.7	0.00	0.00	0.00	0.00	0.00	0.00	0.22
MMV019551	MALARIA	0	1	1	0.20	0.17	0.25	82.9	0.00	0.25	0.25	0.07	0.00	0.00	0.06
MMV019721	MALARIA	1	0	0	0.05	0.03	0.11	87.9	0.25	0.00	0.00	0.00	0.00	0.00	0.01
MMV019742	MALARIA	2	2	1	0.45	0.75	0.84	82.6	0.50	0.50	0.25	0.37	0.69	0.78	0.06
MMV019790	MALARIA	0	0	0	0.08	0.39	0.63	106.8	0.00	0.00	0.00	0.00	0.24	0.50	0.00
MMV019807	MALARIA	0	0	1	0.05	0.08	0.11	99.0	0.00	0.00	0.25	0.00	0.00	0.00	0.00
MMV019838	MALARIA	0	0	0	0.10	0.08	0.11	91.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV019993	MALARIA	0	0	0	0.14	0.10	0.11	77.5	0.00	0.00	0.00	0.00	0.00	0.00	0.11
MMV020081	MALARIA	1	1	1	0.10	0.41	0.39	115.8	0.25	0.25	0.25	0.00	0.26	0.19	0.00
MMV020120	MALARIA	1	0	0	0.00	0.08	0.22	82.4	0.25	0.00	0.00	0.00	0.00	0.00	0.07
MMV020136	MALARIA	0	0	0	0.10	0.14	0.11	73.4	0.00	0.00	0.00	0.00	0.00	0.00	0.15
MMV020152	MALARIA	0	0	0	0.10	0.08	0.11	90.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV020165	MALARIA	0	0	0	0.20	0.17	0.38	86.7	0.00	0.00	0.00	0.07	0.00	0.17	0.03
MMV020289	MALARIA	0	0	0	0.20	0.22	0.38	90.2	0.00	0.00	0.00	0.07	0.03	0.17	0.00
MMV020291	MALARIA	1	0	0	0.05	0.03	0.09	108.9	0.25	0.00	0.00	0.00	0.00	0.00	0.00
MMV020320	MALARIA	1	1	1	0.00	0.08	0.11	56.8	0.25	0.25	0.25	0.00	0.00	0.00	0.30
MMV020321	MALARIA	0	0	0	0.15	0.14	0.28	67.7	0.00	0.00	0.00	0.02	0.00	0.04	0.20
MMV020388	MALARIA	0	0	0	0.15	0.22	0.38	137.3	0.00	0.00	0.00	0.02	0.03	0.17	0.00
MMV020391	MALARIA	2	4	3	0.95	0.85	1.00	29.5	0.50	1.00	0.75	0.95	0.81	1.00	0.56
MMV020512	MALARIA	2	2	0	0.20	0.33	0.56	77.0	0.50	0.50	0.00	0.07	0.17	0.41	0.12
MMV020517	MALARIA	0	0	0	0.20	0.46	0.67	38.4	0.00	0.00	0.00	0.07	0.32	0.56	0.48
MMV020520	MALARIA	0	1	0	0.05	0.03	0.11	71.5	0.00	0.25	0.00	0.00	0.00	0.00	0.17
MMV020537	MALARIA	0	0	0	0.35	0.46	0.50	83.5	0.00	0.00	0.00	0.25	0.32	0.33	0.06
MMV020591	MALARIA	0	0	0	0.15	0.22	0.33	68.8	0.00	0.00	0.00	0.02	0.03	0.11	0.19

MMV020623		MALARIA	4	4	4	0.35	0.50	0.78	67.5	1.00	1.00	1.00	0.25	0.38	0.70	0.20
MMV020670		MALARIA	1	0	0	0.00	0.03	0.03	121.3	0.25	0.00	0.00	0.00	0.00	0.00	0.00
MMV020710		MALARIA	0	0	0	0.20	0.35	0.39	78.4	0.00	0.00	0.00	0.07	0.19	0.19	0.10
MMV020982		MALARIA	0	0	0	0.00	0.08	0.06	79.8	0.00	0.00	0.00	0.00	0.00	0.00	0.09
MMV021013		TUBERCULOSIS	0	0	0	0.00	0.08	0.15	16.4	0.00	0.00	0.00	0.00	0.00	0.00	0.68
MMV021057	*Azoxystrobin*	MALARIA	0	0	0	0.05	0.03	0.11	79.5	0.00	0.00	0.00	0.00	0.00	0.00	0.09
MMV021375		MALARIA	0	0	0	0.11	0.11	0.06	92.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV021660		TUBERCULOSIS	4	4	4	1.00	1.00	1.00	8.5	1.00	1.00	1.00	1.00	1.00	1.00	0.75
MMV022029		MALARIA	0	2	4	0.60	0.78	0.94	19.2	0.00	0.50	1.00	0.54	0.73	0.92	0.65
MMV022236		MALARIA	0	0	1	0.20	0.11	0.25	97.3	0.00	0.00	0.25	0.07	0.00	0.00	0.00
MMV022478		MALARIA	2	4	4	0.73	0.78	0.94	7.7	0.50	1.00	1.00	0.69	0.72	0.92	0.76
MMV023183		MALARIA	0	0	0	0.00	0.10	0.21	58.0	0.00	0.00	0.00	0.00	0.00	0.00	0.29
MMV023227		MALARIA	0	0	0	0.10	0.11	0.25	103.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV023233		MALARIA	4	4	4	0.57	0.78	0.75	41.1	1.00	1.00	1.00	0.50	0.72	0.67	0.45
MMV023370		MALARIA	0	0	0	0.00	0.00	0.06	94.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV023388		MALARIA	0	0	0	0.14	0.56	0.94	103.3	0.00	0.00	0.00	0.00	0.44	0.92	0.00
MMV023860		MALARIA	0	0	0	0.15	0.19	0.45	116.5	0.00	0.00	0.00	0.02	0.00	0.27	0.00
MMV023949		MALARIA	0	0	1	0.00	0.03	0.03	119.9	0.00	0.00	0.25	0.00	0.00	0.00	0.00
MMV023953		MALARIA	1	2	4	0.20	0.46	0.52	67.8	0.25	0.50	1.00	0.07	0.32	0.35	0.20
MMV023969		TUBERCULOSIS	3	3	4	0.14	0.45	0.56	14.8	0.75	0.75	1.00	0.00	0.31	0.41	0.69
MMV023985		MALARIA	2	4	4	0.78	0.94	1.00	60.5	0.50	1.00	1.00	0.75	0.93	1.00	0.27
MMV024035		MALARIA	0	0	0	0.15	0.46	0.45	-18.1	0.00	0.00	0.00	0.02	0.32	0.27	1.00
MMV024101		MALARIA	0	0	0	0.14	0.11	0.31	98.4	0.00	0.00	0.00	0.00	0.00	0.08	0.00
MMV024114		MALARIA	2	3	3	0.45	0.57	0.52	4.5	0.50	0.75	0.75	0.36	0.46	0.35	0.79
MMV024195		MALARIA	0	0	0	-0.05	0.20	0.56	105.9	0.00	0.00	0.00	0.00	0.00	0.41	0.00
MMV024311		TUBERCULOSIS	0	0	0	0.00	0.00	0.11	75.2	0.00	0.00	0.00	0.00	0.00	0.00	0.13
MMV024397		MALARIA	0	0	0	0.10	0.08	0.17	91.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV024406		MALARIA	2	4	4	0.80	0.95	1.00	3.7	0.50	1.00	1.00	0.77	0.93	1.00	0.80
MMV024443		MALARIA	0	0	0	0.19	0.28	0.38	101.9	0.00	0.00	0.00	0.06	0.10	0.17	0.00
MMV024829		MALARIA	0	0	0	0.10	0.24	0.21	125.5	0.00	0.00	0.00	0.00	0.05	0.00	0.00
MMV024937		MALARIA	0	0	0	0.08	0.39	0.63	69.7	0.00	0.00	0.00	0.00	0.24	0.50	0.18
MMV026020		MALARIA	0	0	0	0.23	0.60	0.45	86.9	0.00	0.00	0.00	0.11	0.50	0.27	0.02
MMV026313		MALARIA	0	0	0	0.10	0.22	0.38	152.1	0.00	0.00	0.00	0.00	0.03	0.17	0.00
MMV026356		MALARIA	0	0	0	0.05	0.08	0.03	92.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV026468		MALARIA	0	0	0	0.10	0.08	0.15	99.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV026490		MALARIA	0	2	0	0.20	0.51	0.58	85.9	0.00	0.50	0.00	0.07	0.39	0.43	0.03

MMV026550	MALARIA	0	0	0	0.10	0.14	0.27	88.6	0.00	0.00	0.00	0.00	0.00	0.03	0.01
MMV028694	MALARIA	0	1	4	0.20	0.56	0.56	109.2	0.00	0.25	1.00	0.07	0.44	0.42	0.00
MMV030734	MALARIA	0	0	0	0.15	0.17	0.31	156.0	0.00	0.00	0.00	0.02	0.00	0.08	0.00
MMV031011	MALARIA	0	0	0	0.10	0.51	0.39	13.4	0.00	0.00	0.00	0.00	0.39	0.19	0.71
MMV032967	MALARIA	0	0	0	0.15	0.08	0.15	90.7	0.00	0.00	0.00	0.02	0.00	0.00	0.00
MMV032995	MALARIA	0	0	0	0.30	0.39	0.44	99.5	0.00	0.00	0.00	0.19	0.24	0.25	0.00
MMV045105	KINETOPLASTIDS	0	0	0	0.15	0.30	0.15	111.9	0.00	0.00	0.00	0.02	0.12	0.00	0.00
MMV047015	TUBERCULOSIS	0	0	0	0.14	0.05	0.05	79.3	0.00	0.00	0.00	0.00	0.00	0.00	0.09
MMV053220	TUBERCULOSIS	0	1	0	0.18	0.20	0.13	73.9	0.00	0.25	0.00	0.05	0.00	0.00	0.14
MMV054312	TUBERCULOSIS	0	0	0	0.20	0.11	0.06	111.8	0.00	0.00	0.00	0.07	0.00	0.00	0.00
MMV062221	MALARIA	2	1	4	0.05	0.28	0.25	55.6	0.50	0.25	1.00	0.00	0.10	0.00	0.32
MMV063404	TUBERCULOSIS	2	0	0	0.05	0.05	0.07	87.9	0.50	0.00	0.00	0.00	0.00	0.00	0.01
MMV069458	TUBERCULOSIS	1	0	0	0.20	0.19	0.33	84.0	0.25	0.00	0.00	0.07	0.00	0.11	0.05
MMV084603	MALARIA	0	0	0	0.18	0.30	0.24	87.2	0.00	0.00	0.00	0.05	0.13	0.00	0.02
MMV084864	MALARIA	0	0	0	0.00	-0.03	0.03	85.7	0.00	0.00	0.00	0.00	0.00	0.00	0.04
MMV085071	MALARIA	1	0	0	0.35	0.39	0.69	103.0	0.25	0.00	0.00	0.25	0.24	0.58	0.00
MMV085210	MALARIA	0	0	0	0.05	0.08	0.11	96.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV085230	MALARIA	0	0	0	0.10	0.15	0.11	101.8	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV085499	MALARIA	2	2	4	0.03	0.33	0.50	113.2	0.50	0.50	1.00	0.00	0.17	0.33	0.00
MMV090930	TUBERCULOSIS	0	0	0	0.09	0.20	0.18	133.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV099637	KINETOPLASTIDS	0	0	0	0.41	0.45	0.51	113.4	0.00	0.00	0.00	0.32	0.31	0.35	0.00
MMV1019989	MALARIA	0	0	0	0.00	-0.06	0.11	99.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV102872	TUBERCULOSIS	0	4	4	0.09	0.25	0.95	-10.0	0.00	1.00	1.00	0.00	0.06	0.93	0.92
MMV1028806	MALARIA	0	0	0	0.18	0.30	0.13	71.7	0.00	0.00	0.00	0.05	0.13	0.00	0.17
MMV1029203	MALARIA	0	0	0	0.11	0.11	0.13	90.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV1030799	MALARIA	1	0	0	0.20	0.28	0.31	5.7	0.25	0.00	0.00	0.07	0.10	0.08	0.78
MMV1037162	MALARIA	0	0	0	0.00	0.00	0.11	81.9	0.00	0.00	0.00	0.00	0.00	0.00	0.07
MMV1088520	MALARIA	0	1	0	0.05	0.22	0.39	84.4	0.00	0.25	0.00	0.00	0.03	0.19	0.05
MMV1110498	WOLBACHIA LF	0	0	0	0.18	0.20	0.13	95.7	0.00	0.00	0.00	0.05	0.00	0.00	0.00
MMV1198433	SCHISTOSOMIASIS	0	0	0	0.10	0.06	0.00	82.3	0.00	0.00	0.00	0.00	0.00	0.00	0.07
MMV1236379	KINETOPLASTIDS	0	0	0	0.05	0.08	0.09	94.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV146306	TUBERCULOSIS	0	0	0	0.05	0.03	0.03	90.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV153413	TUBERCULOSIS	4	4	4	1.00	1.00	1.00	-3.4	1.00	1.00	1.00	1.00	1.00	1.00	0.86
MMV161996	TUBERCULOSIS	0	0	0	0.20	0.28	0.38	107.1	0.00	0.00	0.00	0.07	0.10	0.17	0.00
MMV188296	KINETOPLASTIDS	0	0	0	0.23	0.20	0.24	84.2	0.00	0.00	0.00	0.11	0.00	0.00	0.05
MMV200748	TUBERCULOSIS	0	0	1	0.25	0.39	0.50	91.0	0.00	0.00	0.25	0.13	0.24	0.33	0.00

MMV202458		TUBERCULOSIS	0	0	0	0.18	0.35	0.35	97.2	0.00	0.00	0.00	0.05	0.19	0.13	0.00
MMV202553		KINETOPLASTIDS	1	0	0	0.14	0.30	0.24	87.3	0.25	0.00	0.00	0.00	0.13	0.00	0.02
MMV228911		TUBERCULOSIS	0	0	0	0.25	0.22	0.44	94.4	0.00	0.00	0.00	0.13	0.03	0.25	0.00
MMV272144		TUBERCULOSIS	4	4	4	1.00	1.00	1.00	-0.2	1.00	1.00	1.00	1.00	1.00	1.00	0.83
MMV392832		MALARIA	0	0	0	0.10	0.14	0.15	104.3	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV393144		MALARIA	0	2	4	0.09	0.10	0.16	66.3	0.00	0.50	1.00	0.00	0.00	0.00	0.22
MMV393995		TUBERCULOSIS	0	0	0	0.15	0.24	0.58	93.1	0.00	0.00	0.00	0.02	0.05	0.43	0.00
MMV407539		WOLBACHIA LF	0	0	0	-0.05	0.10	0.17	139.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV407834		MALARIA	0	0	0	0.15	0.17	0.31	96.0	0.00	0.00	0.00	0.02	0.00	0.08	0.00
MMV461553		TUBERCULOSIS	2	4	4	1.00	1.00	1.00	0.6	0.50	1.00	1.00	1.00	1.00	1.00	0.83
MMV495543		TUBERCULOSIS	0	1	1	1.00	1.00	1.00	79.9	0.00	0.25	0.25	1.00	1.00	1.00	0.09
MMV553002		TUBERCULOSIS	0	0	0	0.18	0.45	0.45	355.2	0.00	0.00	0.00	0.05	0.31	0.27	0.00
MMV560185		MALARIA	1	0	1	0.15	0.24	0.39	59.9	0.25	0.00	0.25	0.02	0.05	0.19	0.28
MMV595321		KINETOPLASTIDS	0	0	0	0.10	0.00	0.00	80.6	0.00	0.00	0.00	0.00	0.00	0.00	0.08
MMV611037		TUBERCULOSIS	0	0	0	0.10	0.11	0.31	82.8	0.00	0.00	0.00	0.00	0.00	0.08	0.06
MMV634140		MALARIA	0	0	0	0.15	0.22	0.44	146.5	0.00	0.00	0.00	0.02	0.03	0.25	0.00
MMV637229	*Clemastine*	TRICHURIASIS	2	4	4	0.90	1.00	1.00	46.6	0.50	1.00	1.00	0.88	1.00	1.00	0.40
MMV637953	*Suramin*	REFERENCE COMPOUNDS	0	0	0	0.15	0.30	0.39	94.8	0.00	0.00	0.00	0.02	0.12	0.19	0.00
MMV652003		KINETOPLASTIDS	0	0	0	0.10	0.06	0.17	71.1	0.00	0.00	0.00	0.00	0.00	0.00	0.17
MMV658988		KINETOPLASTIDS	1	4	4	0.00	0.19	0.70	20.7	0.25	1.00	1.00	0.00	0.00	0.60	0.64
MMV658993		KINETOPLASTIDS	0	4	4	0.14	0.17	0.50	66.9	0.00	1.00	1.00	0.00	0.00	0.33	0.21
MMV659004		KINETOPLASTIDS	0	0	0	0.15	0.10	0.33	-3.5	0.00	0.00	0.00	0.02	0.00	0.11	0.86
MMV659010		KINETOPLASTIDS	1	2	4	0.05	0.51	0.39	58.7	0.25	0.50	1.00	0.00	0.39	0.19	0.29
MMV661713		TUBERCULOSIS	1	4	4	0.27	0.65	0.67	12.9	0.25	1.00	1.00	0.16	0.56	0.56	0.71
MMV663250		MALARIA	1	1	4	0.41	0.67	0.88	76.1	0.25	0.25	1.00	0.31	0.58	0.83	0.12
MMV667494		MALARIA	2	4	4	0.30	0.39	0.56	86.9	0.50	1.00	1.00	0.19	0.24	0.42	0.02
MMV668727		ONCHOCERCIASIS	0	0	0	0.18	0.15	0.13	79.3	0.00	0.00	0.00	0.05	0.00	0.00	0.09
MMV671636		ONCHOCERCIASIS	2	4	4	0.45	0.78	0.75	74.6	0.50	1.00	1.00	0.36	0.72	0.67	0.14
MMV675968		CRYPTOSPORIDIOSIS	0	2	0	0.00	0.05	0.28	108.9	0.00	0.50	0.00	0.00	0.00	0.04	0.00
MMV675969		ONCHOCERCIASIS	0	0	0	0.00	0.00	0.11	117.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV675993		CRYPTOSPORIDIOSIS	0	0	0	0.05	0.06	0.06	95.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV675994		CRYPTOSPORIDIOSIS	0	0	0	0.11	0.39	0.31	103.0	0.00	0.00	0.00	0.00	0.24	0.08	0.00
MMV675995		ONCHOCERCIASIS	1	1	0	0.15	0.14	0.09	114.3	0.25	0.25	0.00	0.02	0.00	0.00	0.00
MMV675996		ONCHOCERCIASIS	0	2	1	0.15	0.20	0.11	88.7	0.00	0.50	0.25	0.02	0.00	0.00	0.01
MMV675997		KINETOPLASTIDS	0	0	0	0.08	0.11	0.13	118.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV675998		KINETOPLASTIDS	0	0	0	0.05	0.08	0.09	69.3	0.00	0.00	0.00	0.00	0.00	0.00	0.19

MMV676008	KINETOPLASTIDS	0	0	0	0.05	0.03	0.03	80.2	0.00	0.00	0.00	0.00	0.00	0.00	0.09
MMV676048	KINETOPLASTIDS	0	0	0	0.20	0.17	0.25	109.1	0.00	0.00	0.00	0.07	0.00	0.00	0.00
MMV676050	CRYPTOSPORIDIOSIS	0	0	0	0.10	0.57	0.64	122.4	0.00	0.00	0.00	0.00	0.46	0.52	0.00
MMV676053	CRYPTOSPORIDIOSIS	0	0	0	0.11	0.11	0.06	104.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV676057	KINETOPLASTIDS	0	0	0	0.05	0.17	0.25	66.1	0.00	0.00	0.00	0.00	0.00	0.00	0.22
MMV676063	ONCHOCERCIASIS	0	0	0	0.11	0.31	0.88	33.7	0.00	0.00	0.00	0.00	0.14	0.83	0.52
MMV676064	ONCHOCERCIASIS	1	0	0	0.10	0.03	0.09	116.7	0.25	0.00	0.00	0.00	0.00	0.00	0.00
MMV676159	KINETOPLASTIDS	3	4	4	0.00	0.10	0.16	80.7	0.75	1.00	1.00	0.00	0.00	0.00	0.08
MMV676161	KINETOPLASTIDS	0	0	0	0.05	0.05	0.16	85.9	0.00	0.00	0.00	0.00	0.00	0.00	0.03
MMV676162	KINETOPLASTIDS	0	0	0	0.05	0.03	0.03	54.1	0.00	0.00	0.00	0.00	0.00	0.00	0.33
MMV676182	CRYPTOSPORIDIOSIS	0	0	0	0.00	-0.06	0.00	96.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV676186	KINETOPLASTIDS	0	0	0	0.35	0.33	0.31	88.1	0.00	0.00	0.00	0.25	0.17	0.08	0.01
MMV676191	CRYPTOSPORIDIOSIS	0	0	0	0.00	0.00	0.06	107.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV676204	ONCHOCERCIASIS	1	0	1	0.14	0.17	0.25	126.2	0.25	0.00	0.25	0.00	0.00	0.00	0.00
MMV676260	MALARIA	0	0	0	0.00	0.10	0.17	109.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV676269	MALARIA	0	0	0	0.05	0.08	0.21	117.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV676270	MALARIA	0	0	0	0.20	0.44	0.50	88.1	0.00	0.00	0.00	0.07	0.31	0.33	0.01
MMV676350	MALARIA	1	1	0	0.14	0.45	0.51	79.7	0.25	0.25	0.00	0.00	0.31	0.35	0.09
MMV676358	MALARIA	0	0	0	0.10	0.11	0.19	130.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV676377	TUBERCULOSIS	0	0	0	0.18	0.20	0.13	99.9	0.00	0.00	0.00	0.05	0.00	0.00	0.00
MMV676379	TUBERCULOSIS	0	0	0	0.18	0.20	0.13	101.5	0.00	0.00	0.00	0.05	0.00	0.00	0.00
MMV676380	MALARIA	0	0	0	0.23	0.15	0.13	94.2	0.00	0.00	0.00	0.11	0.00	0.00	0.00
MMV676382	SCHISTOSOMIASIS	4	4	4	0.45	0.68	0.94	71.9	1.00	1.00	1.00	0.36	0.59	0.93	0.16
MMV676383	TUBERCULOSIS	2	4	4	0.18	1.00	1.00	7.1	0.50	1.00	1.00	0.05	1.00	1.00	0.77
MMV676384	TUBERCULOSIS	0	0	0	0.20	0.11	0.13	94.0	0.00	0.00	0.00	0.07	0.00	0.00	0.00
MMV676386	TUBERCULOSIS	0	0	0	0.06	0.13	0.00	93.2	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV676388	TUBERCULOSIS	4	4	4	1.00	1.00	1.00	-7.2	1.00	1.00	1.00	1.00	1.00	1.00	0.90
MMV676389	TUBERCULOSIS	0	0	0	0.09	0.25	0.24	89.6	0.00	0.00	0.00	0.00	0.06	0.00	0.00
MMV676395	TUBERCULOSIS	0	0	0	0.18	0.20	0.07	118.5	0.00	0.00	0.00	0.05	0.00	0.00	0.00
MMV676398	WOLBACHIA LF	0	1	0	0.20	0.33	0.50	55.5	0.00	0.25	0.00	0.07	0.17	0.33	0.32
MMV676401	TUBERCULOSIS	3	2	4	1.00	1.00	1.00	99.5	0.75	0.50	1.00	1.00	1.00	1.00	0.00
MMV676406	TUBERCULOSIS	2	0	1	0.27	0.50	0.40	71.6	0.50	0.00	0.25	0.16	0.38	0.20	0.17
MMV676409	TUBERCULOSIS	0	0	0	0.09	0.10	0.13	68.1	0.00	0.00	0.00	0.00	0.00	0.00	0.20
MMV676411	TUBERCULOSIS	2	0	4	0.30	0.50	0.63	59.4	0.50	0.00	1.00	0.19	0.38	0.50	0.28
MMV676412	TUBERCULOSIS	0	0	1	1.00	1.00	1.00	-7.6	0.00	0.00	0.25	1.00	1.00	1.00	0.90
MMV676431	TUBERCULOSIS	0	0	0	0.14	0.30	0.40	124.9	0.00	0.00	0.00	0.00	0.13	0.20	0.00

MMV676439	TUBERCULOSIS	0	0	0	0.18	0.20	0.07	83.2	0.00	0.00	0.00	0.05	0.00	0.00	0.06
MMV676442	MALARIA	0	0	0	0.05	0.08	0.11	94.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV676444	TUBERCULOSIS	0	0	0	0.05	0.20	0.13	88.6	0.00	0.00	0.00	0.00	0.00	0.00	0.01
MMV676445	TUBERCULOSIS	0	0	0	0.14	0.20	0.18	337.3	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV676449	TUBERCULOSIS	0	0	0	0.09	0.25	0.24	105.6	0.00	0.00	0.00	0.00	0.06	0.00	0.00
MMV676461	TUBERCULOSIS	1	1	0	0.18	0.30	0.29	95.6	0.25	0.25	0.00	0.05	0.13	0.05	0.00
MMV676468	TUBERCULOSIS	0	0	0	0.05	0.10	0.16	69.2	0.00	0.00	0.00	0.00	0.00	0.00	0.19
MMV676470	TUBERCULOSIS	0	0	0	0.09	0.05	0.16	67.0	0.00	0.00	0.00	0.00	0.00	0.00	0.21
MMV676472	TUBERCULOSIS	2	2	2	0.40	0.50	0.63	107.5	0.50	0.50	0.50	0.31	0.38	0.50	0.00
MMV676474	TUBERCULOSIS	0	0	0	0.14	0.20	0.18	76.0	0.00	0.00	0.00	0.00	0.00	0.00	0.13
MMV676476	TUBERCULOSIS	1	1	0	0.14	0.20	0.13	79.3	0.25	0.25	0.00	0.00	0.00	0.00	0.09
MMV676477	TUBERCULOSIS	0	1	3	0.27	0.95	1.00	-1.1	0.00	0.25	0.75	0.16	0.94	1.00	0.84
MMV676478	TUBERCULOSIS	1	1	0	0.05	0.11	0.00	103.4	0.25	0.25	0.00	0.00	0.00	0.00	0.00
MMV676480	ONCHOCERCIASIS	0	3	2	0.40	0.22	0.56	76.3	0.00	0.75	0.50	0.31	0.03	0.41	0.12
MMV676492	LYMPHATIC FILARIASIS	0	0	0	0.05	0.03	0.09	91.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV676501	TUBERCULOSIS	2	4	4	0.00	0.00	0.00	-3.4	0.50	1.00	1.00	0.00	0.00	0.00	0.86
MMV676509	TUBERCULOSIS	1	0	0	0.23	0.25	0.24	109.3	0.25	0.00	0.00	0.11	0.06	0.00	0.00
MMV676512	TUBERCULOSIS	0	0	0	0.09	0.15	0.24	68.7	0.00	0.00	0.00	0.00	0.00	0.00	0.19
MMV676520	TUBERCULOSIS	0	0	0	0.18	0.15	0.24	109.5	0.00	0.00	0.00	0.05	0.00	0.00	0.00
MMV676524	TUBERCULOSIS	0	0	0	0.15	0.28	0.38	157.3	0.00	0.00	0.00	0.02	0.10	0.17	0.00
MMV676526	TUBERCULOSIS	0	1	0	0.09	0.25	0.24	89.7	0.00	0.25	0.00	0.00	0.06	0.00	0.00
MMV676528	MALARIA	0	0	0	0.05	0.14	0.22	81.0	0.00	0.00	0.00	0.00	0.00	0.00	0.08
MMV676536	SCHISTOSOMIASIS	0	0	0	0.10	0.14	0.11	93.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV676539	TUBERCULOSIS	1	0	0	0.18	0.40	0.45	91.2	0.25	0.00	0.00	0.05	0.25	0.27	0.00
MMV676554	TUBERCULOSIS	0	0	0	0.18	0.45	0.40	176.2	0.00	0.00	0.00	0.05	0.31	0.20	0.00
MMV676555	TUBERCULOSIS	0	0	0	0.00	0.25	0.13	102.5	0.00	0.00	0.00	0.00	0.06	0.00	0.00
MMV676558	TUBERCULOSIS	0	0	0	0.09	0.10	0.13	61.8	0.00	0.00	0.00	0.00	0.00	0.00	0.26
MMV676571	TUBERCULOSIS	0	0	1	0.18	0.20	0.18	137.4	0.00	0.00	0.25	0.05	0.00	0.00	0.00
MMV676584	TUBERCULOSIS	0	0	0	0.09	0.10	0.02	81.0	0.00	0.00	0.00	0.00	0.00	0.00	0.08
MMV676588	TUBERCULOSIS	4	4	3	0.55	0.80	0.56	89.4	1.00	1.00	0.75	0.47	0.75	0.42	0.00
MMV676589	TUBERCULOSIS	2	4	4	0.18	0.20	0.17	144.9	0.50	1.00	1.00	0.05	0.00	0.00	0.00
MMV676597	TUBERCULOSIS	0	0	0	0.14	0.20	0.18	89.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV676599	CRYPTOSPORIDIOSIS	0	0	0	0.15	0.22	0.63	85.5	0.00	0.00	0.00	0.02	0.03	0.50	0.04
MMV676600	KINETOPLASTIDS	0	0	0	0.15	0.24	0.28	97.4	0.00	0.00	0.00	0.02	0.05	0.04	0.00
MMV676602	KINETOPLASTIDS	3	4	4	0.36	0.80	0.89	156.5	0.75	1.00	1.00	0.26	0.75	0.85	0.00
MMV676603	TUBERCULOSIS	1	0	0	0.14	0.25	0.29	80.0	0.25	0.00	0.00	0.00	0.06	0.05	0.09

MMV676604		KINETOPLASTIDS	0	0	0	0.15	0.14	0.17	66.9	0.00	0.00	0.00	0.02	0.00	0.00	0.21
MMV676605		MALARIA	0	0	0	0.05	0.03	0.11	64.0	0.00	0.00	0.00	0.00	0.00	0.00	0.24
MMV676877		MALARIA	0	0	0	0.14	0.28	0.63	93.3	0.00	0.00	0.00	0.00	0.10	0.50	0.00
MMV676881		MALARIA	4	4	4	0.35	0.89	0.75	97.7	1.00	1.00	1.00	0.25	0.86	0.67	0.00
MMV687138		TUBERCULOSIS	0	0	0	0.16	0.11	0.00	89.5	0.00	0.00	0.00	0.02	0.00	0.00	0.00
MMV687145		TUBERCULOSIS	0	0	0	0.08	0.11	0.06	141.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV687146		TUBERCULOSIS	0	1	0	0.32	0.33	0.44	123.9	0.00	0.25	0.00	0.21	0.17	0.25	0.00
MMV687170		TUBERCULOSIS	0	0	0	0.11	0.06	0.13	98.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV687172		TUBERCULOSIS	0	0	1	0.05	0.11	0.33	96.4	0.00	0.00	0.25	0.00	0.00	0.11	0.00
MMV687180		TUBERCULOSIS	1	1	0	0.10	0.11	0.06	99.5	0.25	0.25	0.00	0.00	0.00	0.00	0.00
MMV687188		TUBERCULOSIS	0	0	0	0.20	0.28	0.25	107.7	0.00	0.00	0.00	0.07	0.10	0.00	0.00
MMV687189		TUBERCULOSIS	0	0	0	0.05	0.06	0.00	74.7	0.00	0.00	0.00	0.00	0.00	0.00	0.14
MMV687239		TUBERCULOSIS	0	0	0	0.14	0.11	0.00	128.3	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV687243		TUBERCULOSIS	1	1	2	0.00	0.00	0.00	110.1	0.25	0.25	0.50	0.00	0.00	0.00	0.00
MMV687246		MALARIA	0	0	0	0.10	0.14	0.09	152.1	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV687248		TUBERCULOSIS	2	1	1	0.30	0.56	0.69	110.7	0.50	0.25	0.25	0.19	0.44	0.58	0.00
MMV687251		TUBERCULOSIS	4	4	4	1.00	1.00	1.00	-0.4	1.00	1.00	1.00	1.00	1.00	1.00	0.84
MMV687254		TUBERCULOSIS	1	0	1	0.10	0.17	0.19	94.2	0.25	0.00	0.25	0.00	0.00	0.00	0.00
MMV687273		TUBERCULOSIS	4	4	4	1.00	1.00	1.00	13.5	1.00	1.00	1.00	1.00	1.00	1.00	0.71
MMV687696		TUBERCULOSIS	1	0	0	0.11	0.06	0.00	66.8	0.25	0.00	0.00	0.00	0.00	0.00	0.21
MMV687699		TUBERCULOSIS	0	1	0	0.16	0.33	0.31	93.7	0.00	0.25	0.00	0.02	0.17	0.08	0.00
MMV687700		TUBERCULOSIS	0	0	0	0.10	0.11	0.13	113.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV687703		TUBERCULOSIS	4	4	4	0.35	0.78	0.75	158.3	1.00	1.00	1.00	0.25	0.72	0.67	0.00
MMV687706		KINETOPLASTIDS	2	2	2	0.05	0.15	0.44	25.2	0.50	0.50	0.50	0.00	0.00	0.26	0.60
MMV687729		TUBERCULOSIS	0	0	2	0.05	0.11	0.19	126.6	0.00	0.00	0.50	0.00	0.00	0.00	0.00
MMV687730		TUBERCULOSIS	3	4	4	0.15	0.39	0.56	131.6	0.75	1.00	1.00	0.02	0.24	0.42	0.00
MMV687747		TUBERCULOSIS	0	0	0	0.10	0.06	0.06	110.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV687749		TUBERCULOSIS	0	0	0	0.10	0.11	0.19	94.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV687762		KINETOPLASTIDS	0	0	0	0.18	0.35	0.13	98.4	0.00	0.00	0.00	0.05	0.19	0.00	0.00
MMV687765		TUBERCULOSIS	4	4	4	0.30	0.56	1.00	17.8	1.00	1.00	1.00	0.19	0.44	1.00	0.67
MMV687775		LYMPHATIC FILARIASIS	2	0	1	1.00	1.00	1.00	33.5	0.50	0.00	0.25	1.00	1.00	1.00	0.52
MMV687776		LYMPHATIC FILARIASIS	0	0	0	0.20	0.14	0.17	23.2	0.00	0.00	0.00	0.07	0.00	0.00	0.62
MMV687794		MALARIA	0	0	0	0.05	0.10	0.11	71.5	0.00	0.00	0.00	0.00	0.00	0.00	0.17
MMV687796	*Amikacin*	REFERENCE COMPOUNDS	0	0	0	0.00	0.06	0.00	93.8	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV687798	*Levofloxacin (-)-ofloxacin*	REFERENCE COMPOUNDS	0	0	0	0.05	0.05	0.05	67.9	0.00	0.00	0.00	0.00	0.00	0.00	0.20
MMV687800	*Clofazimine*	REFERENCE COMPOUNDS	4	4	4	1.00	1.00	1.00	16.6	1.00	1.00	1.00	1.00	1.00	1.00	0.68

MMV687801	*Ethambutol*	REFERENCE COMPOUNDS	0	0	0	0.15	0.14	0.28	78.3	0.00	0.00	0.00	0.02	0.00	0.04	0.10
MMV687803	*Linezolid*	REFERENCE COMPOUNDS	0	0	0	0.23	0.20	0.18	94.2	0.00	0.00	0.00	0.11	0.00	0.00	0.00
MMV687807		TUBERCULOSIS	4	4	4	1.00	1.00	1.00	35.4	1.00	1.00	1.00	1.00	1.00	1.00	0.50
MMV687812		TUBERCULOSIS	3	3	3	1.00	1.00	1.00	5.2	0.75	0.75	0.75	1.00	1.00	1.00	0.78
MMV687813		TUBERCULOSIS	0	0	0	0.10	0.17	0.25	72.9	0.00	0.00	0.00	0.00	0.00	0.00	0.15
MMV688122		TUBERCULOSIS	1	3	4	0.41	0.67	0.75	39.9	0.25	0.75	1.00	0.31	0.58	0.67	0.46
MMV688124		TUBERCULOSIS	1	2	4	0.25	0.44	0.50	85.8	0.25	0.50	1.00	0.13	0.31	0.33	0.03
MMV688125		TUBERCULOSIS	0	0	0	0.15	0.17	0.38	108.0	0.00	0.00	0.00	0.02	0.00	0.17	0.00
MMV688178		SCHISTOSOMIASIS	0	0	2	0.60	1.00	1.00	3.1	0.00	0.00	0.50	0.54	1.00	1.00	0.80
MMV688179		KINETOPLASTIDS	0	0	0	0.05	0.11	0.13	61.7	0.00	0.00	0.00	0.00	0.00	0.00	0.26
MMV688180		KINETOPLASTIDS	0	0	0	0.15	0.08	0.15	121.3	0.00	0.00	0.00	0.02	0.00	0.00	0.00
MMV688262	*Delamanid*	TUBERCULOSIS	0	0	0	0.16	0.11	0.06	110.5	0.00	0.00	0.00	0.02	0.00	0.00	0.00
MMV688270		SCHISTOSOMIASIS	2	4	4	0.05	0.00	0.11	39.0	0.50	1.00	1.00	0.00	0.00	0.00	0.47
MMV688271		KINETOPLASTIDS	0	0	0	0.00	0.00	0.06	89.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688273		KINETOPLASTIDS	4	4	4	0.30	0.62	0.88	55.9	1.00	1.00	1.00	0.19	0.53	0.84	0.31
MMV688274		KINETOPLASTIDS	0	0	0	0.30	0.25	0.22	124.8	0.00	0.00	0.00	0.19	0.06	0.00	0.00
MMV688279		KINETOPLASTIDS	0	0	0	0.00	0.19	0.38	-4.2	0.00	0.00	0.00	0.00	0.00	0.17	0.87
MMV688283		KINETOPLASTIDS	2	4	4	0.60	0.61	0.63	32.4	0.50	1.00	1.00	0.54	0.51	0.50	0.53
MMV688313		SCHISTOSOMIASIS	0	2	0	0.10	0.17	0.28	14.7	0.00	0.50	0.00	0.00	0.00	0.04	0.70
MMV688327	*Radezolid*	TUBERCULOSIS	0	0	0	0.05	0.06	0.00	107.9	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688330		TOXOPLASMOSIS	0	0	0	0.10	0.03	0.03	89.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688345		TOXOPLASMOSIS	0	0	0	0.15	0.28	0.44	131.0	0.00	0.00	0.00	0.02	0.10	0.25	0.00
MMV688350		DENGUE	1	1	1	0.41	0.45	0.45	89.1	0.25	0.25	0.25	0.32	0.31	0.27	0.00
MMV688352		DENGUE	3	4	4	0.14	0.10	0.16	68.1	0.75	1.00	1.00	0.00	0.00	0.00	0.20
MMV688360		KINETOPLASTIDS	1	0	0	0.27	0.40	0.40	91.8	0.25	0.00	0.00	0.16	0.25	0.20	0.00
MMV688361		KINETOPLASTIDS	0	0	0	0.10	0.11	0.13	115.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688362		KINETOPLASTIDS	0	0	0	0.10	0.03	0.03	4.9	0.00	0.00	0.00	0.00	0.00	0.00	0.79
MMV688364		TOXOPLASMOSIS	1	4	4	0.41	0.44	0.69	84.8	0.25	1.00	1.00	0.31	0.31	0.58	0.04
MMV688371		KINETOPLASTIDS	0	0	0	0.40	0.33	0.31	22.7	0.00	0.00	0.00	0.31	0.17	0.08	0.62
MMV688372		KINETOPLASTIDS	0	4	4	0.68	0.67	0.88	111.1	0.00	1.00	1.00	0.62	0.58	0.83	0.00
MMV688407		KINETOPLASTIDS	1	0	0	0.25	0.30	0.39	72.8	0.25	0.00	0.00	0.13	0.12	0.19	0.15
MMV688410		KINETOPLASTIDS	0	0	0	0.35	0.33	0.38	112.9	0.00	0.00	0.00	0.25	0.17	0.17	0.00
MMV688411		TOXOPLASMOSIS	0	0	0	0.20	0.28	0.50	168.7	0.00	0.00	0.00	0.07	0.10	0.33	0.00
MMV688415		KINETOPLASTIDS	0	0	0	0.00	0.00	0.11	90.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688416		DENGUE	0	0	0	0.18	0.15	0.18	93.7	0.00	0.00	0.00	0.05	0.00	0.00	0.00
MMV688417		TOXOPLASMOSIS	4	4	4	0.55	0.67	1.00	54.3	1.00	1.00	1.00	0.48	0.58	1.00	0.33

MMV688466		TUBERCULOSIS	0	0	0	0.10	0.06	0.13	113.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688467		KINETOPLASTIDS	0	0	0	0.15	0.14	0.27	40.5	0.00	0.00	0.00	0.02	0.00	0.03	0.46
MMV688469		TOXOPLASMOSIS	1	2	3	0.24	0.33	0.56	94.1	0.25	0.50	0.75	0.12	0.17	0.42	0.00
MMV688470		TOXOPLASMOSIS	0	0	0	0.27	0.45	0.51	90.7	0.00	0.00	0.00	0.16	0.31	0.35	0.00
MMV688471		TOXOPLASMOSIS	1	0	0	0.18	0.35	0.35	89.9	0.25	0.00	0.00	0.05	0.19	0.13	0.00
MMV688472		TOXOPLASMOSIS	1	0	0	0.32	0.45	0.56	108.8	0.25	0.00	0.00	0.21	0.31	0.42	0.00
MMV688474		KINETOPLASTIDS	1	0	0	0.24	0.28	0.50	77.2	0.25	0.00	0.00	0.12	0.10	0.33	0.11
MMV688508		TUBERCULOSIS	0	1	0	0.10	0.11	0.06	137.7	0.00	0.25	0.00	0.00	0.00	0.00	0.00
MMV688509		TOXOPLASMOSIS	0	0	0	0.10	0.22	0.19	124.2	0.00	0.00	0.00	0.00	0.03	0.00	0.00
MMV688514		KINETOPLASTIDS	2	1	0	0.18	0.65	0.45	91.6	0.50	0.25	0.00	0.05	0.56	0.27	0.00
MMV688543		DENGUE	0	2	4	0.25	0.56	1.00	121.6	0.00	0.50	1.00	0.13	0.44	1.00	0.00
MMV688547		KINETOPLASTIDS	0	0	0	0.10	0.11	0.25	85.9	0.00	0.00	0.00	0.00	0.00	0.00	0.03
MMV688548		TOXOPLASMOSIS	1	0	0	0.18	0.20	0.13	105.6	0.25	0.00	0.00	0.05	0.00	0.00	0.00
MMV688550		KINETOPLASTIDS	1	0	0	0.15	0.41	0.64	107.8	0.25	0.00	0.00	0.02	0.26	0.52	0.00
MMV688552		SCHISTOSOMIASIS	0	0	0	0.15	0.17	0.31	137.0	0.00	0.00	0.00	0.02	0.00	0.08	0.00
MMV688553		TUBERCULOSIS	0	0	0	0.14	0.20	0.18	76.2	0.00	0.00	0.00	0.00	0.00	0.00	0.12
MMV688554		TUBERCULOSIS	1	0	0	0.09	0.15	0.13	91.3	0.25	0.00	0.00	0.00	0.00	0.00	0.00
MMV688555		TUBERCULOSIS	0	0	0	0.14	0.30	0.35	93.6	0.00	0.00	0.00	0.00	0.13	0.13	0.00
MMV688557		TUBERCULOSIS	1	1	2	0.15	0.05	0.06	45.2	0.25	0.25	0.50	0.02	0.00	0.00	0.41
MMV688703		TOXOPLASMOSIS	0	0	0	0.00	0.06	0.11	94.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688704		TOXOPLASMOSIS	0	0	0	0.14	0.40	0.35	109.1	0.00	0.00	0.00	0.00	0.25	0.13	0.00
MMV688754	*Trifloxystrobin*	KINETOPLASTIDS	1	0	0	0.05	0.03	0.15	108.9	0.25	0.00	0.00	0.00	0.00	0.00	0.00
MMV688755		TUBERCULOSIS	0	0	0	0.20	0.17	0.31	158.3	0.00	0.00	0.00	0.07	0.00	0.08	0.00
MMV688756	*Sutezolid*	TUBERCULOSIS	0	0	0	0.14	0.20	0.24	144.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688761		SCHISTOSOMIASIS	4	4	4	1.00	1.00	1.00	-8.0	1.00	1.00	1.00	1.00	1.00	1.00	0.91
MMV688762		SCHISTOSOMIASIS	4	4	4	1.00	1.00	1.00	-3.4	1.00	1.00	1.00	1.00	1.00	1.00	0.86
MMV688763		SCHISTOSOMIASIS	4	4	4	1.00	1.00	1.00	-5.8	1.00	1.00	1.00	1.00	1.00	1.00	0.89
MMV688766		SCHISTOSOMIASIS	0	0	0	0.15	0.17	0.19	0.1	0.00	0.00	0.00	0.02	0.00	0.00	0.83
MMV688768		SCHISTOSOMIASIS	3	4	4	1.00	1.00	1.00	6.4	0.75	1.00	1.00	1.00	1.00	1.00	0.77
MMV688771		SCHISTOSOMIASIS	1	2	0	0.05	0.10	0.16	88.7	0.25	0.50	0.00	0.00	0.00	0.00	0.01
MMV688773	*Benznidazole*	REFERENCE COMPOUNDS	0	0	0	0.15	0.24	0.22	89.2	0.00	0.00	0.00	0.02	0.05	0.00	0.00
MMV688774	*Posaconazole*	REFERENCE COMPOUNDS	3	4	4	0.20	0.24	0.56	88.8	0.75	1.00	1.00	0.07	0.05	0.41	0.01
MMV688775	*Rifampicin*	REFERENCE COMPOUNDS	0	0	0	0.05	0.05	0.11	116.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688776		KINETOPLASTIDS	0	0	0	0.14	0.20	0.13	100.8	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688793		KINETOPLASTIDS	0	0	0	0.09	0.15	0.13	113.5	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688795		KINETOPLASTIDS	0	0	0	0.14	0.00	0.05	74.1	0.00	0.00	0.00	0.00	0.00	0.00	0.14

MMV688796		KINETOPLASTIDS	1	0	0	0.14	0.15	0.18	97.1	0.25	0.00	0.00	0.00	0.00	0.00	0.00
MMV688797		KINETOPLASTIDS	1	0	0	0.09	0.25	0.24	131.0	0.25	0.00	0.00	0.00	0.06	0.00	0.00
MMV688798		KINETOPLASTIDS	0	1	0	0.18	0.40	0.40	105.6	0.00	0.25	0.00	0.05	0.25	0.20	0.00
MMV688844		TUBERCULOSIS	0	0	0	0.05	0.06	0.11	77.2	0.00	0.00	0.00	0.00	0.00	0.00	0.11
MMV688845		TUBERCULOSIS	0	0	0	0.20	0.17	0.19	128.9	0.00	0.00	0.00	0.07	0.00	0.00	0.00
MMV688846		TUBERCULOSIS	2	0	1	0.10	0.06	0.19	71.1	0.50	0.00	0.25	0.00	0.00	0.00	0.17
MMV688852		TOXOPLASMOSIS	0	0	0	0.08	0.17	0.19	98.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688853		CRYPTOSPORIDIOSIS	0	0	0	0.20	0.24	0.39	75.2	0.00	0.00	0.00	0.07	0.05	0.19	0.13
MMV688854		CRYPTOSPORIDIOSIS	0	0	0	0.10	0.17	0.39	44.3	0.00	0.00	0.00	0.00	0.00	0.19	0.42
MMV688888		TUBERCULOSIS	0	0	0	0.18	0.35	0.35	63.0	0.00	0.00	0.00	0.05	0.19	0.13	0.25
MMV688889		TUBERCULOSIS	0	1	0	0.14	0.20	0.13	83.9	0.00	0.25	0.00	0.00	0.00	0.00	0.05
MMV688891		TUBERCULOSIS	0	0	0	0.10	0.06	0.06	90.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688921		DENGUE	0	0	0	0.16	0.11	0.00	93.9	0.00	0.00	0.00	0.02	0.00	0.00	0.00
MMV688934	*Tolfenpyrad*	KINETOPLASTIDS	1	1	2	0.36	0.55	0.45	72.8	0.25	0.25	0.50	0.26	0.44	0.27	0.16
MMV688936		TUBERCULOSIS	0	0	0	0.09	0.40	0.40	46.4	0.00	0.00	0.00	0.00	0.25	0.20	0.40
MMV688938		TUBERCULOSIS	0	0	0	0.27	0.40	0.63	2.3	0.00	0.00	0.00	0.16	0.25	0.51	0.81
MMV688939		TUBERCULOSIS	3	4	4	0.00	0.00	0.11	95.7	0.75	1.00	1.00	0.00	0.00	0.00	0.00
MMV688941		TUBERCULOSIS	1	1	0	0.10	0.14	0.27	73.9	0.25	0.25	0.00	0.00	0.00	0.03	0.14
MMV688942	*Bitertanol*	KINETOPLASTIDS	1	0	0	0.09	0.15	0.13	105.6	0.25	0.00	0.00	0.00	0.00	0.00	0.00
MMV688943	*Difenoconazol*	KINETOPLASTIDS	2	1	2	0.14	0.25	0.07	47.3	0.50	0.25	0.50	0.00	0.06	0.00	0.39
MMV688955		TOXOPLASMOSIS	0	0	0	0.10	0.00	0.11	99.6	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV688958		KINETOPLASTIDS	1	0	0	0.14	0.20	0.18	86.7	0.25	0.00	0.00	0.00	0.00	0.00	0.03
MMV688978	*Auranofin*	REFERENCE COMPOUNDS	4	4	4	1.00	1.00	1.00	203.0	1.00	1.00	1.00	1.00	1.00	1.00	0.00
MMV688980		MALARIA	0	0	1	0.14	0.11	0.00	115.4	0.00	0.00	0.25	0.00	0.00	0.00	0.00
MMV688990	*Miltefosine*	REFERENCE COMPOUNDS	0	0	0	0.05	0.24	0.27	126.6	0.00	0.00	0.00	0.00	0.05	0.03	0.00
MMV688991	*Nitazoxanide*	REFERENCE COMPOUNDS	2	4	4	0.20	0.30	0.56	102.0	0.50	1.00	1.00	0.07	0.12	0.41	0.00
MMV688994	*Streptomycin*	REFERENCE COMPOUNDS	0	0	0	0.18	0.10	0.07	93.0	0.00	0.00	0.00	0.05	0.00	0.00	0.00
MMV689000	*Amphotericin B*	REFERENCE COMPOUNDS	0	0	0	0.14	0.15	0.11	95.0	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV689028		KINETOPLASTIDS	0	1	1	0.05	0.06	0.06	130.8	0.00	0.25	0.25	0.00	0.00	0.00	0.00
MMV689029		KINETOPLASTIDS	0	0	1	0.10	0.11	0.19	143.6	0.00	0.00	0.25	0.00	0.00	0.00	0.00
MMV689060		KINETOPLASTIDS	1	0	0	0.10	0.06	0.13	81.9	0.25	0.00	0.00	0.00	0.00	0.00	0.07
MMV689061		KINETOPLASTIDS	0	0	0	0.10	0.06	0.19	117.4	0.00	0.00	0.00	0.00	0.00	0.00	0.00
MMV689243		KINETOPLASTIDS	0	0	1	0.20	0.22	0.31	77.5	0.00	0.00	0.25	0.07	0.03	0.08	0.11
MMV689244		KINETOPLASTIDS	0	2	4	0.75	0.72	0.75	46.8	0.00	0.50	1.00	0.71	0.65	0.67	0.40
MMV689255	*D-Eritadenine*	CRYPTOSPORIDIOSIS	0	0	0	0.10	0.19	0.22	63.7	0.00	0.00	0.00	0.00	0.00	0.00	0.24
MMV689437		KINETOPLASTIDS	0	0	0	0.00	0.00	0.17	42.2	0.00	0.00	0.00	0.00	0.00	0.00	0.44

MMV689480	*Buparvaquone*	REFERENCE COMPOUNDS	3	2	2	0.20	0.19	0.28	71.7	0.75	0.50	0.50	0.07	0.00	0.04	0.16
MMV689709		KINETOPLASTIDS	0	0	0	0.16	0.17	0.13	123.5	0.00	0.00	0.00	0.02	0.00	0.00	0.00
MMV689758	*Bedaquiline*	REFERENCE COMPOUNDS	2	4	4	0.32	0.80	0.95	7.2	0.50	1.00	1.00	0.21	0.75	0.93	0.76
MMV690027		KINETOPLASTIDS	2	0	0	0.35	0.57	0.67	31.0	0.50	0.00	0.00	0.25	0.46	0.56	0.54
MMV690028		KINETOPLASTIDS	1	0	0	0.18	0.20	0.13	3.5	0.25	0.00	0.00	0.05	0.00	0.00	0.80
MMV690102		KINETOPLASTIDS	4	4	4	0.53	0.50	0.50	4.7	1.00	1.00	1.00	0.45	0.38	0.33	0.79
MMV690103		KINETOPLASTIDS	0	0	0	0.20	0.28	0.44	39.1	0.00	0.00	0.00	0.07	0.10	0.25	0.47

4.6.2 Table B- PCA scores

This table reports the scores for each compound of the Pathogen Box projected on the principal components 1 and 2 (PC1 and PC2). These scores were used to generate the PCA plot.

Compound ID	Disease Set	PC1	PC2
MMV000011	REFERENCE COMPOUNDS	-1.054	-0.075
MMV000014	REFERENCE COMPOUNDS	4.104	0.536
MMV000023	REFERENCE COMPOUNDS	-0.622	0.181
MMV000062	REFERENCE COMPOUNDS	-0.913	0.153
MMV000063	REFERENCE COMPOUNDS	-0.487	0.839
MMV000858	MALARIA	-0.791	0.349
MMV000907	MALARIA	-1.054	-0.075
MMV001059	MALARIA	1.794	-1.867
MMV001493	ONCHOCERCIASIS	-0.568	0.708
MMV001499	REFERENCE COMPOUNDS	-0.101	-0.542
MMV001561	KINETOPLASTIDS	0.452	1.912
MMV001625	REFERENCE COMPOUNDS	-1.014	-0.010
MMV002529	REFERENCE COMPOUNDS	1.299	-1.624
MMV002816	REFERENCE COMPOUNDS	-0.730	-0.234
MMV002817	ONCHOCERCIASIS	1.501	1.753
MMV003152	REFERENCE COMPOUNDS	-0.634	0.442
MMV003270	HOOKWORM	1.095	-0.617
MMV004168	KINETOPLASTIDS	-0.611	0.559
MMV006239	MALARIA	-1.054	-0.075
MMV006372	MALARIA	-0.128	0.053
MMV006741	MALARIA	-0.994	0.023
MMV006833	MALARIA	-1.054	-0.075
MMV006901	MALARIA	1.405	2.808
MMV007133	MALARIA	-1.054	-0.075
MMV007471	MALARIA	-1.054	-0.075
MMV007625	MALARIA	-1.054	-0.075
MMV007638	MALARIA	-0.135	0.482
MMV007803	MALARIA	0.139	-0.759
MMV007920	MALARIA	-0.568	0.708
MMV008439	MALARIA	-0.661	-0.367
MMV009054	MALARIA	-1.054	-0.075
MMV009135	MALARIA	-1.054	-0.075
MMV010545	MALARIA	-0.902	-0.150
MMV010576	MALARIA	-0.661	-0.367
MMV010764	MALARIA	-0.806	-0.197
MMV011229	MALARIA	-0.730	0.447
MMV011511	MALARIA	-1.054	-0.075
MMV011691	MALARIA	0.837	-1.200
MMV011765	MALARIA	1.769	-1.060

MMV011903	MALARIA	1.323	-0.103
MMV012074	TUBERCULOSIS	-0.683	0.083
MMV016136	MALARIA	-0.588	0.676
MMV016838	MALARIA	0.852	-1.009
MMV019087	MALARIA	-0.750	0.414
MMV019189	MALARIA	-1.054	-0.075
MMV019234	MALARIA	-0.609	0.643
MMV019551	MALARIA	-0.540	-0.171
MMV019721	MALARIA	-1.034	-0.043
MMV019742	MALARIA	0.947	-0.899
MMV019790	MALARIA	-0.101	-0.542
MMV019807	MALARIA	-0.661	-0.367
MMV019838	MALARIA	-1.054	-0.075
MMV019993	MALARIA	-0.831	0.284
MMV020081	MALARIA	-0.299	-0.544
MMV020120	MALARIA	-0.913	0.153
MMV020136	MALARIA	-0.750	0.414
MMV020152	MALARIA	-1.054	-0.075
MMV020165	MALARIA	-0.669	-0.136
MMV020289	MALARIA	-0.730	-0.234
MMV020291	MALARIA	-1.054	-0.075
MMV020320	MALARIA	-0.054	0.613
MMV020321	MALARIA	-0.573	0.540
MMV020388	MALARIA	-0.730	-0.234
MMV020391	MALARIA	3.165	-0.055
MMV020512	MALARIA	-0.030	-0.066
MMV020517	MALARIA	0.985	0.970
MMV020520	MALARIA	-0.710	0.480
MMV020537	MALARIA	-0.304	-0.187
MMV020591	MALARIA	-0.460	0.442
MMV020623	MALARIA	2.257	-1.242
MMV020670	MALARIA	-1.054	-0.075
MMV020710	MALARIA	-0.490	0.074
MMV020982	MALARIA	-0.872	0.219
MMV021013	TUBERCULOSIS	0.323	2.145
MMV021057	MALARIA	-0.872	0.219
MMV021375	MALARIA	-1.054	-0.075
MMV021660	TUBERCULOSIS	3.942	0.274
MMV022029	MALARIA	3.587	0.023
MMV022236	MALARIA	-0.661	-0.367
MMV022478	MALARIA	3.810	0.382
MMV023183	MALARIA	-0.467	0.872
MMV023227	MALARIA	-1.054	-0.075
MMV023233	MALARIA	2.706	-0.397
MMV023370	MALARIA	-1.054	-0.075

MMV023388	MALARIA	0.699	-0.934
MMV023860	MALARIA	-0.540	-0.327
MMV023949	MALARIA	-0.661	-0.367
MMV023953	MALARIA	1.590	-0.915
MMV023969	TUBERCULOSIS	2.696	0.629
MMV023985	MALARIA	2.970	-1.293
MMV024035	MALARIA	1.486	2.938
MMV024101	MALARIA	-0.902	-0.150
MMV024114	MALARIA	2.392	1.303
MMV024195	MALARIA	-0.273	-0.458
MMV024311	TUBERCULOSIS	-0.791	0.349
MMV024397	MALARIA	-1.054	-0.075
MMV024406	MALARIA	4.044	0.438
MMV024443	MALARIA	-0.730	-0.234
MMV024829	MALARIA	-1.054	-0.075
MMV024937	MALARIA	0.263	0.046
MMV026020	MALARIA	-0.499	-0.262
MMV026313	MALARIA	-0.730	-0.234
MMV026356	MALARIA	-1.054	-0.075
MMV026468	MALARIA	-1.054	-0.075
MMV026490	MALARIA	-0.174	-0.379
MMV026550	MALARIA	-0.977	-0.071
MMV028694	MALARIA	1.318	-1.634
MMV030734	MALARIA	-0.902	-0.150
MMV031011	MALARIA	0.746	2.066
MMV032967	MALARIA	-1.054	-0.075
MMV032995	MALARIA	-0.578	-0.309
MMV045105	KINETOPLASTIDS	-1.054	-0.075
MMV047015	TUBERCULOSIS	-0.872	0.219
MMV053220	TUBERCULOSIS	-0.771	0.382
MMV054312	TUBERCULOSIS	-1.054	-0.075
MMV062221	MALARIA	1.165	-0.197
MMV063404	TUBERCULOSIS	-1.034	-0.043
MMV069458	TUBERCULOSIS	-0.743	-0.015
MMV084603	MALARIA	-1.014	-0.010
MMV084864	MALARIA	-0.973	0.055
MMV085071	MALARIA	0.051	-0.617
MMV085210	MALARIA	-1.054	-0.075
MMV085230	MALARIA	-1.054	-0.075
MMV085499	MALARIA	1.146	-1.550
MMV090930	TUBERCULOSIS	-1.054	-0.075
MMV099637	KINETOPLASTIDS	-0.387	-0.402
MMV101998	MALARIA	-1.05427853	-0.075
MMV102872	TUBERCULOSIS	4.153	0.895
MMV102880	MALARIA	-0.70996989	0.480

MMV102920	MALARIA	-1.05427853	-0.075
MMV103079	MALARIA	0.67798187	2.397
MMV103716	MALARIA	-0.91250439	0.153
MMV108852	MALARIA	-0.59084432	-0.089
MMV111049	WOLBACHIA LF	-1.05427853	-0.075
MMV119843	SCHISTOSOMIASIS	-0.91250439	0.153
MMV123637	KINETOPLASTIDS	-1.05427853	-0.075
MMV146306	TUBERCULOSIS	-1.054	-0.075
MMV153413	TUBERCULOSIS	4.165	0.634
MMV161996	TUBERCULOSIS	-0.730	-0.234
MMV188296	KINETOPLASTIDS	-0.953	0.088
MMV200748	TUBERCULOSIS	-0.032	-0.675
MMV202458	TUBERCULOSIS	-0.806	-0.197
MMV202553	KINETOPLASTIDS	-1.014	-0.010
MMV228911	TUBERCULOSIS	-0.578	-0.309
MMV272144	TUBERCULOSIS	4.104	0.536
MMV392832	MALARIA	-1.054	-0.075
MMV393144	MALARIA	0.963	-0.523
MMV393995	TUBERCULOSIS	-0.235	-0.477
MMV407539	WOLBACHIA LF	-1.054	-0.075
MMV407834	MALARIA	-0.902	-0.150
MMV461553	TUBERCULOSIS	4.104	0.536
MMV495543	TUBERCULOSIS	1.427	-1.006
MMV553002	TUBERCULOSIS	-0.540	-0.327
MMV560185	MALARIA	0.268	0.370
MMV595321	KINETOPLASTIDS	-0.892	0.186
MMV611037	TUBERCULOSIS	-0.780	0.046
MMV634140	MALARIA	-0.578	-0.309
MMV637229	TRICHURIASIS	3.234	-0.869
MMV637953	REFERENCE COMPOUNDS	-0.692	-0.253
MMV652003	KINETOPLASTIDS	-0.710	0.480
MMV658988	KINETOPLASTIDS	2.957	0.288
MMV658993	KINETOPLASTIDS	1.572	-0.864
MMV659004	KINETOPLASTIDS	0.897	2.630
MMV659010	KINETOPLASTIDS	1.467	-0.472
MMV661713	TUBERCULOSIS	3.023	0.554
MMV663250	MALARIA	2.342	-1.624
MMV667494	MALARIA	1.358	-1.568
MMV668727	ONCHOCERCIASIS	-0.872	0.219
MMV671636	ONCHOCERCIASIS	2.078	-1.410
MMV675968	CRYPTOSPORIDIOSIS	-0.978	-0.113
MMV675969	ONCHOCERCIASIS	-1.054	-0.075
MMV675993	CRYPTOSPORIDIOSIS	-1.054	-0.075
MMV675994	CRYPTOSPORIDIOSIS	-0.902	-0.150
MMV675995	ONCHOCERCIASIS	-1.054	-0.075

MMV675996	ONCHOCERCIASIS	-0.641	-0.334
MMV675997	KINETOPLASTIDS	-1.054	-0.075
MMV675998	KINETOPLASTIDS	-0.669	0.545
MMV676008	KINETOPLASTIDS	-0.872	0.219
MMV676048	KINETOPLASTIDS	-1.054	-0.075
MMV676050	CRYPTOSPORIDIOSIS	-0.063	-0.561
MMV676053	CRYPTOSPORIDIOSIS	-1.054	-0.075
MMV676057	KINETOPLASTIDS	-0.609	0.643
MMV676063	ONCHOCERCIASIS	1.581	0.848
MMV676064	ONCHOCERCIASIS	-1.054	-0.075
MMV676159	KINETOPLASTIDS	0.679	-0.980
MMV676161	KINETOPLASTIDS	-0.994	0.023
MMV676162	KINETOPLASTIDS	-0.386	1.002
MMV676182	CRYPTOSPORIDIOSIS	-1.054	-0.075
MMV676186	KINETOPLASTIDS	-0.882	-0.117
MMV676191	CRYPTOSPORIDIOSIS	-1.054	-0.075
MMV676204	ONCHOCERCIASIS	-0.661	-0.367
MMV676260	MALARIA	-1.054	-0.075
MMV676269	MALARIA	-1.054	-0.075
MMV676270	MALARIA	-0.405	-0.351
MMV676350	MALARIA	-0.205	-0.108
MMV676358	MALARIA	-1.054	-0.075
MMV676377	TUBERCULOSIS	-1.054	-0.075
MMV676379	TUBERCULOSIS	-1.054	-0.075
MMV676380	MALARIA	-1.054	-0.075
MMV676382	SCHISTOSOMIASIS	2.614	-1.587
MMV676383	TUBERCULOSIS	3.983	0.340
MMV676384	TUBERCULOSIS	-1.054	-0.075
MMV676386	TUBERCULOSIS	-1.054	-0.075
MMV676388	TUBERCULOSIS	4.246	0.764
MMV676389	TUBERCULOSIS	-1.054	-0.075
MMV676395	TUBERCULOSIS	-1.054	-0.075
MMV676398	WOLBACHIA LF	0.223	0.662
MMV676401	TUBERCULOSIS	2.423	-2.175
MMV676406	TUBERCULOSIS	0.064	0.002
MMV676409	TUBERCULOSIS	-0.649	0.578
MMV676411	TUBERCULOSIS	2.037	-0.794
MMV676412	TUBERCULOSIS	3.068	1.639
MMV676431	TUBERCULOSIS	-0.673	-0.262
MMV676439	TUBERCULOSIS	-0.933	0.121
MMV676442	MALARIA	-1.054	-0.075
MMV676444	TUBERCULOSIS	-1.034	-0.043
MMV676445	TUBERCULOSIS	-1.054	-0.075
MMV676449	TUBERCULOSIS	-1.054	-0.075
MMV676461	TUBERCULOSIS	-0.959	-0.122

MMV676468	TUBERCULOSIS	-0.669	0.545
MMV676470	TUBERCULOSIS	-0.629	0.610
MMV676472	TUBERCULOSIS	0.685	-1.125
MMV676474	TUBERCULOSIS	-0.791	0.349
MMV676476	TUBERCULOSIS	-0.872	0.219
MMV676477	TUBERCULOSIS	3.732	0.860
MMV676478	TUBERCULOSIS	-1.054	-0.075
MMV676480	ONCHOCERCIASIS	0.756	-0.649
MMV676492	LYMPHATIC FILARIASIS	-1.054	-0.075
MMV676501	TUBERCULOSIS	2.259	1.567
MMV676509	TUBERCULOSIS	-1.054	-0.075
MMV676512	TUBERCULOSIS	-0.669	0.545
MMV676520	TUBERCULOSIS	-1.054	-0.075
MMV676524	TUBERCULOSIS	-0.730	-0.234
MMV676526	TUBERCULOSIS	-1.054	-0.075
MMV676528	MALARIA	-0.892	0.186
MMV676536	SCHISTOSOMIASIS	-1.054	-0.075
MMV676539	TUBERCULOSIS	-0.540	-0.327
MMV676554	TUBERCULOSIS	-0.673	-0.262
MMV676555	TUBERCULOSIS	-1.054	-0.075
MMV676558	TUBERCULOSIS	-0.528	0.774
MMV676571	TUBERCULOSIS	-0.661	-0.367
MMV676584	TUBERCULOSIS	-0.892	0.186
MMV676588	TUBERCULOSIS	0.925	-1.342
MMV676589	TUBERCULOSIS	0.517	-1.242
MMV676597	TUBERCULOSIS	-1.054	-0.075
MMV676599	CRYPTOSPORIDIOSIS	-0.020	-0.411
MMV676600	KINETOPLASTIDS	-0.978	-0.113
MMV676602	KINETOPLASTIDS	2.138	-2.035
MMV676603	TUBERCULOSIS	-0.777	0.172
MMV676604	KINETOPLASTIDS	-0.629	0.610
MMV676605	MALARIA	-0.568	0.708
MMV676877	MALARIA	-0.101	-0.542
MMV676881	MALARIA	1.794	-1.867
MMV687138	TUBERCULOSIS	-1.054	-0.075
MMV687145	TUBERCULOSIS	-1.054	-0.075
MMV687146	TUBERCULOSIS	-0.578	-0.309
MMV687170	TUBERCULOSIS	-1.054	-0.075
MMV687172	TUBERCULOSIS	-0.452	-0.470
MMV687180	TUBERCULOSIS	-1.054	-0.075
MMV687188	TUBERCULOSIS	-1.054	-0.075
MMV687189	TUBERCULOSIS	-0.771	0.382
MMV687239	TUBERCULOSIS	-1.054	-0.075
MMV687243	TUBERCULOSIS	-0.268	-0.659
MMV687246	MALARIA	-1.054	-0.075

MMV687248	TUBERCULOSIS	0.444	-0.908
MMV687251	TUBERCULOSIS	4.125	0.568
MMV687254	TUBERCULOSIS	-0.661	-0.367
MMV687273	TUBERCULOSIS	3.861	0.144
MMV687696	TUBERCULOSIS	-0.629	0.610
MMV687699	TUBERCULOSIS	-0.902	-0.150
MMV687700	TUBERCULOSIS	-1.054	-0.075
MMV687703	TUBERCULOSIS	1.794	-1.867
MMV687706	KINETOPLASTIDS	1.442	1.058
MMV687729	TUBERCULOSIS	-0.268	-0.659
MMV687730	TUBERCULOSIS	1.318	-1.634
MMV687747	TUBERCULOSIS	-1.054	-0.075
MMV687749	TUBERCULOSIS	-1.054	-0.075
MMV687762	KINETOPLASTIDS	-1.054	-0.075
MMV687765	TUBERCULOSIS	3.780	0.013
MMV687775	LYMPHATIC FILARIASIS	2.298	0.398
MMV687776	LYMPHATIC FILARIASIS	0.201	1.949
MMV687794	MALARIA	-0.710	0.480
MMV687796	REFERENCE COMPOUNDS	-1.054	-0.075
MMV687798	REFERENCE COMPOUNDS	-0.649	0.578
MMV687800	REFERENCE COMPOUNDS	3.801	0.046
MMV687801	REFERENCE COMPOUNDS	-0.775	0.214
MMV687803	REFERENCE COMPOUNDS	-1.054	-0.075
MMV687807	TUBERCULOSIS	3.436	-0.542
MMV687812	TUBERCULOSIS	3.610	0.664
MMV687813	TUBERCULOSIS	-0.750	0.414
MMV688122	TUBERCULOSIS	2.726	-0.365
MMV688124	TUBERCULOSIS	1.207	-1.452
MMV688125	TUBERCULOSIS	-0.730	-0.234
MMV688178	SCHISTOSOMIASIS	3.258	1.021
MMV688179	KINETOPLASTIDS	-0.528	0.774
MMV688180	KINETOPLASTIDS	-1.054	-0.075
MMV688262	TUBERCULOSIS	-1.054	-0.075
MMV688270	SCHISTOSOMIASIS	1.469	0.293
MMV688271	KINETOPLASTIDS	-1.054	-0.075
MMV688273	KINETOPLASTIDS	2.746	-1.013
MMV688274	KINETOPLASTIDS	-1.054	-0.075
MMV688279	KINETOPLASTIDS	1.032	2.607
MMV688283	KINETOPLASTIDS	2.544	0.023
MMV688313	SCHISTOSOMIASIS	0.440	2.173
MMV688327	TUBERCULOSIS	-1.054	-0.075
MMV688330	TOXOPLASMOSIS	-1.054	-0.075
MMV688345	TOXOPLASMOSIS	-0.578	-0.309
MMV688350	DENGUE	-0.147	-0.619
MMV688352	DENGUE	0.922	-0.588

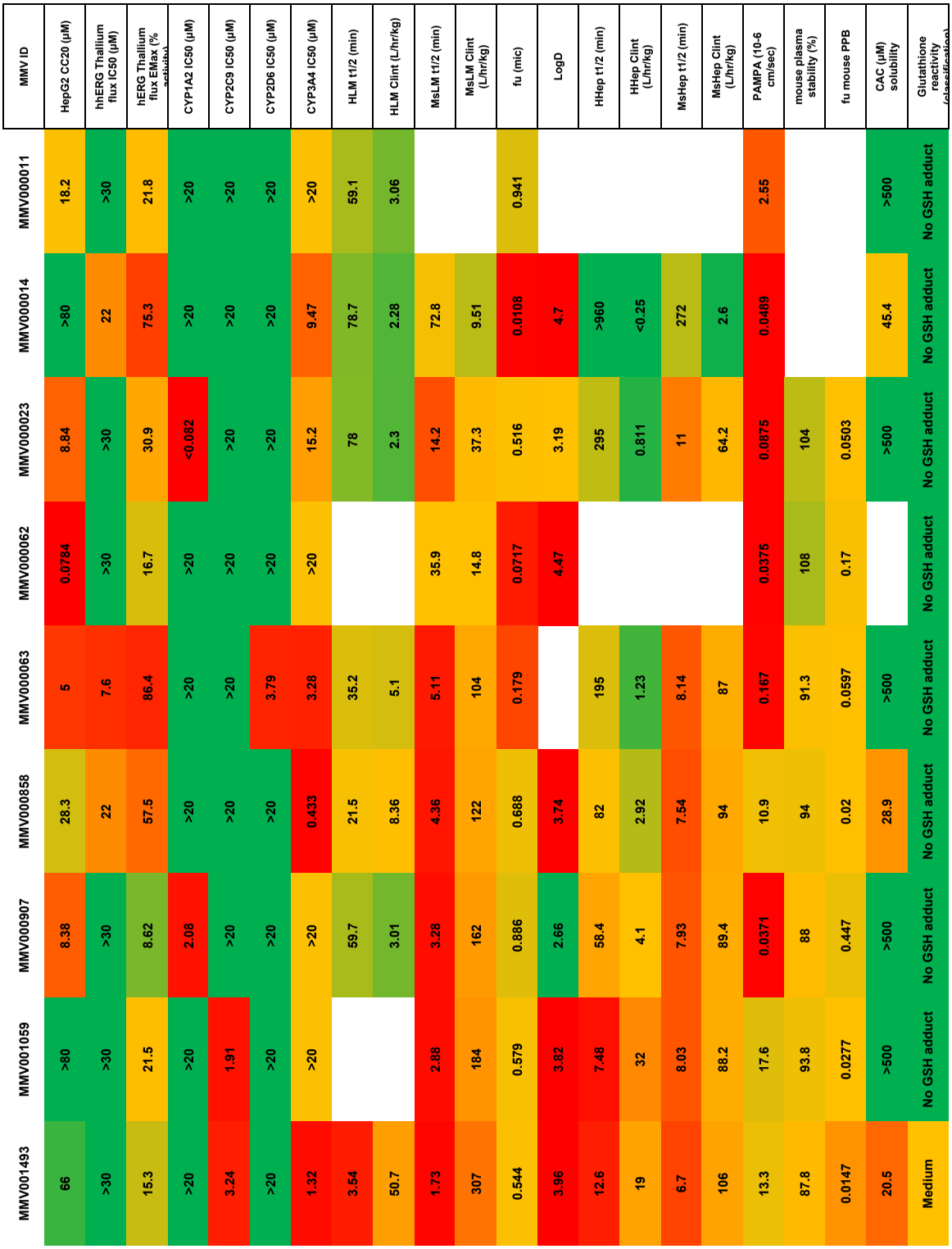
MMV688360	KINETOPLASTIDS	-0.673	-0.262
MMV688361	KINETOPLASTIDS	-1.054	-0.075
MMV688362	KINETOPLASTIDS	0.546	2.505
MMV688364	TOXOPLASMOSIS	1.704	-1.652
MMV688371	KINETOPLASTIDS	0.354	1.875
MMV688372	KINETOPLASTIDS	2.099	-2.016
MMV688407	KINETOPLASTIDS	-0.388	0.237
MMV688410	KINETOPLASTIDS	-0.730	-0.234
MMV688411	TOXOPLASMOSIS	-0.425	-0.383
MMV688415	KINETOPLASTIDS	-1.054	-0.075
MMV688416	DENGUE	-1.054	-0.075
MMV688417	TOXOPLASMOSIS	3.092	-1.097
MMV688466	TUBERCULOSIS	-1.054	-0.075
MMV688467	KINETOPLASTIDS	-0.065	1.399
MMV688469	TOXOPLASMOSIS	0.925	-1.342
MMV688470	TOXOPLASMOSIS	-0.387	-0.402
MMV688471	TOXOPLASMOSIS	-0.806	-0.197
MMV688472	TOXOPLASMOSIS	-0.254	-0.467
MMV688474	KINETOPLASTIDS	-0.202	-0.024
MMV688508	TUBERCULOSIS	-1.054	-0.075
MMV688509	TOXOPLASMOSIS	-1.054	-0.075
MMV688514	KINETOPLASTIDS	-0.540	-0.327
MMV688543	DENGUE	2.423	-2.175
MMV688547	KINETOPLASTIDS	-0.994	0.023
MMV688548	TOXOPLASMOSIS	-1.054	-0.075
MMV688550	KINETOPLASTIDS	-0.063	-0.561
MMV688552	SCHISTOSOMIASIS	-0.902	-0.150
MMV688553	TUBERCULOSIS	-0.811	0.316
MMV688554	TUBERCULOSIS	-1.054	-0.075
MMV688555	TUBERCULOSIS	-0.806	-0.197
MMV688557	TUBERCULOSIS	0.562	0.680
MMV688703	TOXOPLASMOSIS	-1.054	-0.075
MMV688704	TOXOPLASMOSIS	-0.806	-0.197
MMV688754	KINETOPLASTIDS	-1.054	-0.075
MMV688755	TUBERCULOSIS	-0.902	-0.150
MMV688756	TUBERCULOSIS	-1.054	-0.075
MMV688761	SCHISTOSOMIASIS	4.267	0.797
MMV688762	SCHISTOSOMIASIS	4.165	0.634
MMV688763	SCHISTOSOMIASIS	4.226	0.732
MMV688766	SCHISTOSOMIASIS	0.627	2.635
MMV688768	SCHISTOSOMIASIS	3.983	0.340
MMV688771	SCHISTOSOMIASIS	-1.034	-0.043
MMV688773	REFERENCE COMPOUNDS	-1.054	-0.075
MMV688774	REFERENCE COMPOUNDS	1.319	-1.592
MMV688775	REFERENCE COMPOUNDS	-1.054	-0.075

MMV688776	KINETOPLASTIDS	-1.054	-0.075
MMV688793	KINETOPLASTIDS	-1.054	-0.075
MMV688795	KINETOPLASTIDS	-0.771	0.382
MMV688796	KINETOPLASTIDS	-1.054	-0.075
MMV688797	KINETOPLASTIDS	-1.054	-0.075
MMV688798	KINETOPLASTIDS	-0.673	-0.262
MMV688844	TUBERCULOSIS	-0.831	0.284
MMV688845	TUBERCULOSIS	-1.054	-0.075
MMV688846	TUBERCULOSIS	-0.317	0.188
MMV688852	TOXOPLASMOSIS	-1.054	-0.075
MMV688853	CRYPTOSPORIDIOSIS	-0.429	0.172
MMV688854	CRYPTOSPORIDIOSIS	0.159	1.119
MMV688888	TUBERCULOSIS	-0.300	0.620
MMV688889	TUBERCULOSIS	-0.953	0.088
MMV688891	TUBERCULOSIS	-1.054	-0.075
MMV688921	DENGUE	-1.054	-0.075
MMV688934	KINETOPLASTIDS	0.570	-0.388
MMV688936	TUBERCULOSIS	0.137	1.044
MMV688938	TUBERCULOSIS	1.558	2.094
MMV688939	TUBERCULOSIS	0.517	-1.242
MMV688941	TUBERCULOSIS	-0.714	0.354
MMV688942	KINETOPLASTIDS	-1.054	-0.075
MMV688943	KINETOPLASTIDS	0.521	0.615
MMV688955	TOXOPLASMOSIS	-1.054	-0.075
MMV688958	KINETOPLASTIDS	-0.994	0.023
MMV688978	REFERENCE COMPOUNDS	2.423	-2.175
MMV688980	MALARIA	-0.661	-0.367
MMV688990	REFERENCE COMPOUNDS	-0.997	-0.103
MMV688991	REFERENCE COMPOUNDS	1.299	-1.624
MMV688994	REFERENCE COMPOUNDS	-1.054	-0.075
MMV689000	REFERENCE COMPOUNDS	-1.054	-0.075
MMV689028	KINETOPLASTIDS	-0.661	-0.367
MMV689029	KINETOPLASTIDS	-0.661	-0.367
MMV689060	KINETOPLASTIDS	-0.913	0.153
MMV689061	KINETOPLASTIDS	-1.054	-0.075
MMV689243	KINETOPLASTIDS	-0.286	-0.082
MMV689244	KINETOPLASTIDS	2.605	-0.561
MMV689255	CRYPTOSPORIDIOSIS	-0.568	0.708
MMV689437	KINETOPLASTIDS	-0.163	1.362
MMV689480	REFERENCE COMPOUNDS	0.132	-0.173
MMV689709	KINETOPLASTIDS	-1.054	-0.075
MMV689758	REFERENCE COMPOUNDS	3.829	0.372
MMV690027	KINETOPLASTIDS	1.107	1.165
MMV690028	KINETOPLASTIDS	0.566	2.537
MMV690102	KINETOPLASTIDS	2.746	1.030

MMV690103	KINETOPLASTIDS	0.374	1.226

4.6.3 *Table C- DMPK heatmap*

Drug metabolism and pharmacokinetic properties are reported as heatmap for all the compounds of the Pathogen Box. Cells are color-coded from red (non-optimal DMPK property) to green (optimal DMPK property). Cells are white when no DMPK data were provided.



MMV006741	MMV006372	MMV006239	MMV004168	MMV003270	MMV003152	MMV002817	MMV002816	MMV002529	MMV001625	MMV001561	MMV001499
26.4	21.9	19.9	8.7	>80	0.0134	2.53	>80	>80	>80	8.97	>80
6.8	10	30	2.1	>30	>30	>30	>30	>30	>30	7.5	>30
78.8	12.1	32.9	>20	-0.6	5.25	22	18.8	7.03	0.181	80.3	2.91
>20	>20	2.81	>20	15.1	>20	0.489		10.9	>20	>20	>20
>20	>20	>20	4.1	>20	>20	>20		>20	>20	>20	>20
>20	>20	>20	>20	>20	>20	>20		>20	>20	2.12	>20
>20	56.6	>20	50.7	>20	>20	>20		>20	>20	18.4	>20
100	3.18	18	3.54	44.2	67	16.4		38.4		41.5	
1.79	18.9	9.96	14.5	4.06	2.99	11		4.68		4.33	
11.9	28.1	18	191	13.5	45.8	6.79		1.3		40.8	>120
232	0.449	29.5	0.135	39.3	11.6	78.2		409		13	<4.43
	0.449	0.288	0.135	0.826	0.914	0.0661		0.815		0.0847	0.766
3.55	2.76	3.84	4.22	2.17	3.16		0.813	3.25		4.34	1.38
134	135	141	182	235	82.9	317		47.7		491	
1.79	1.77	1.69	1.23	1.02	2.89	0.756		5.02		0.488	
16.3	51	48.6	33.4	15.2	44.4	79		4.61		70.9	
43.6	13.9	14.6	21.2	46.6	16	8.97		154		9.99	
	0.0066	0.722	0.65	31.9	17.9		5.56	24.8		0.0828	
91.7	72	107	0	107	95.3	96.3	106	104		90.7	
0.133	0.34	0.014	0	0.24	0.089	0.0002	0.977	0.177		0.034	
>500		>500		>500	<0.5	1	36.5	>500	>500	418	247
No GSH adduct	No GSH adduct	No GSH adduct	No GSH adduct	Medium	Very Low	Very Low	No GSH adduct	No GSH adduct	Very Low	No GSH adduct	Very low

MMV010545	MMV009135	MMV009054	MMV008439	MMV007920	MMV007803	MMV007638	MMV007625	MMV007471	MMV007133	MMV006901	MMV006833
3.29	>80	30.6	17.4	18.2	31.6	20	31.1	14.5	>80	6.86	11.8
8.2	>30	>30	>30	>30	>30	7.3	>30	7.4	>30	2.1	>30
89.7	31.9	21.8	18.3	12.7	3.77	74.5	>20	3.12	>20	93.2	>20
>20	>20	>20	>20	>20	>20	>20	>20	1.04	>20	0.443	>20
15.8	>20	>20	>20	>20	>20	5.8	>20	0.78	>20	>20	>20
>20	>20	>20	>20	>20	18.8	>20	>20	0.811	>20	>20	>20
10.5	>20	>20	>20	5.63	4.61	6.53	35.7	2.73		>20	3.24
33.9	3.85	31.5		17	39.1	21.7	5.04	65.8		20.3	55.5
5.31	46.6	5.71		10.5	0.542	8.27	19	3.25		8.83	0.928
15	0.769	4.53		36.9	979	14.2	27.9	163		5.8	573
35.5	691	117		14.4		37.4	1			91.5	0.68
0.569	0.872	0.512		<0.001		0.182	1			0.22	0.68
3.13	3.62	3.99	5.3	4.31	3.85	4.23	3.31	4.26	4.18	4.53	4.08
223	21.9	21.9		12.1	4.67	75.7	181	68.4	19.7	67.9	10.5
1.07	10.9	10.9		19.8	51.7	3.16	1.675	3.9	12.6	3.53	24.5
44.9	5.46	5.12		59.2	3.03	16.1	15.6	14.2	5.18	7.79	7.07
15.8	130	138		12	234	44	45.5	49.8	137	91	100
0.464	17.5	18.2		3.52	22.5	4.28		6.76	7.07	6.38	13.1
104	93.3	19.3	100	83.6	84.3	84.8	8.33	90.3	0	106	72
0.0353			<0.0001	0.0004	0.051	0.0047		0.0086		0.0127	0.009
135	58.7	<0.5	84.3	10.5		11.7				7.6	
No GSH adduct	High	Medium	No GSH adduct	Very low	Medium	Very Low	No GSH adduct	No GSH adduct	Very low	No GSH adduct	High

MMV019189	MMV019087	MMV016838	MMV016136	MMV012074	MMV011903	MMV011765	MMV011691	MMV011511	MMV011229	MMV010764	MMV010576
>80	>80	>80	80	42.3	8.43	2.6	>80	13	>80	30	1.55
>30	30	>30	>30	>30	>30	>30	>30	4.1	>30	>30	21
11.4	43.7	>20	9.54	8.15	3.29	0.0321	33.6	>20	7	28	60.6
>20	>20	13.8	>20	>20	12.6	>20	>20	>20	>20	1.27	>20
>20	>20	>20	>20	>20	>20	7.26	>20	12.3	>20	2.62	18.5
>20	>20	7.64	>20	>20	>20	>20	>20	19.3	>20	>20	>20
>20	>20	19.5	>20		>20	47	1.81	3.86	>120	0.237	>20
	23.7	9.2	63.6		22.3	3.82	86.2	46.6	<1.5	16.5	>120
	7.57	3.34	2.83		8.05	9.39	2.08	14.9	20.2	11.1	<1.5
		159	>120		9.38	56.6	>120	35.7	26.4	86.1	>120
			<4.43		56.7		<4.43	0.314	0.57	6.17	<4.43
	0.868		0.729		0.0776		0.406	0.314	0.57	0.392	0.806
	4.09	5.18	3.4		4.72	3.75		3.55		4.19	2.7
	142		68.9	270	50.5	8.2	315	38.5	>960	35.9	11.8
	1.69		3.48	3.448	4.74	29.2	0.762	12.8	<0.25	6.67	20.3
	8.89		58.9	34.5	24.2	4.83	183	28.1	45	16.1	8.75
	79.7		12	20.5	29.2	147	3.88	27.2	15.7	44	81
	11.3		0.0284		3.53	25.5		0.0204	9.1	12.7	8.06
	0	90.3			109	17.7	92.6	94.3	98.7	89.7	90.8
		0.137			0.0038		0.113	0.273	5.47	0.025	0.055
>500	0.61		13.9		<0.5		16.1			41.3	>500
Very Low	No GSH adduct	No GSH adduct	No GSH adduct	Medium	No GSH adduct	No GSH adduct	Medium	No GSH adduct	No GSH adduct	Medium	Medium

MMV020152	MMV020136	MMV020120	MMV020081	MMV019993	MMV019838	MMV019807	MMV019790	MMV019742	MMV019721	MMV019551	MMV019234
>80	>80	>80	>80	39.3	>80	>80	>80	>80	10	>80	43.1
>30	>30	>30	>30	>30	>30	>30	>30	>30	>30	>30	17
4.05	8.3	25.4	30.6	2.06	2.38	4.37	3.33	30.3	30.6	>20	54.7
>20	>20	>20	>20	>20	0.145	>20	>20	8.75	>20	16.6	15
>20	>20	>20	>20	>20	>20	>20		9.57	6.38	>20	>20
>20	>20	0.765	>20	14.4	>20	>20	12	>20	>20	>20	1.6
>20	>20	>20	18.6		0.382	>20	11.3	5.28	8.8	7.28	>20
	62.5	5.04	8.77		17	18.9	15.9	22.9	14.2	24.7	13.5
	2.88	35.6	20.5		10.5	9.52	0.641	7.85	12.6	5.76	13.3
	75	1.69	5.38		17.8	1.01	829	4.01	7.51	92.2	3.88
	7.09	315	98.7		29.8	528		133	71.4		137
	0.846	0.569	0.707		0.272	0.663		0.414	0.697		0.418
4.44	2.37	4.75	3.36		5.03	3.32	3.7	3.81	3.85	4.24	3.68
5.04	50	24.8	25.9		35.7	49.7	16.1	25.8	58.5	3.38	120
47.6	4.79	9.64	9.25		6.71	4.82	16.2	9.29	4.24	74.9	1.99
3.51	25.8	7	10.3		34.4	4.51	3.97	8.39	12.2	4.2	9.83
202	27.5	101	69		20.6	157	179	84.5	58.3	169	72.1
	6.82	3.78	8.04		7.89	16.7	9.25	8.55	12.1	14	11.4
	92.3	86.7	111	86.7	91.1	101	67.7	67.3	84	0	92.7
	0.0397	0.0647	0.0953	0.047	0.0108	0.0073	0.0057	0.024	0.014		0.0113
4.8	>500	119	>500		9.8	7.5		<0.5	<0.5		5.2
Medium	Medium	Very Low	Medium	Very low	No GSH adduct	No GSH adduct	Medium	Very Low	No GSH adduct	Very low	No GSH adduct

MMV020591	MMV020537	MMV020520	MMV020517	MMV020512	MMV020391	MMV020388	MMV020321	MMV020320	MMV020291	MMV020289	MMV020165
14	16.4	21.8	11.3	17.5	5.35	51.8	33.7	0.917	>80	>80	11.5
>30	1.2	>30	17	27	>30	>30	>30	>30	>30	>30	10
27.5	0.407	40.5	71.8	56.2	31.2	>20	33.4	10.7	9.76	5.9	77.8
>20	>20	>20	3.19	>20	>20	>20	1.67	>20	>20	>20	>20
0.709		>20	>20	>20	7.2	>20	>20	>20	>20	>20	>20
>20	11.5	>20	16.2	>20	>20	>20	>20	>20	>20	>20	>20
19	21.6	1.76	>20	>20	>20	3.99	1.54	>20	>20	>20	>20
3.07	8.3	3.85	11.3	12.8	15.2	45	2.17	45.1	19.4		>120
58.5	1.37	46.7	15.9	14	11.8	2.07	82.9	3.98	9.25		<1.5
1.85	387		3.22		3.26	257	1.05	>120	4.23		12.7
287	0.351		165		163	0.416	508	<4.43	126		41.8
0.415	0.351	0.576	0.488	0.178	0.135	0.416		1	0.941		0.206
4.17	4.26	4.15	4.51	4.48	4.64	4.45	4.99	3.54	3.85		4.15
22.6	109	28.3	59	16.9	25.8	147	7.64	301	31.2		244
10.6	2.65	8.48	4.06	14.1	9.3	1.7	31.4	0.795	7.67		0.981
8.19	4.69	11.1	6.64	24	13.5	12.5	3.57	235	7.62		11.4
86.5	151	64.1	107	29.5	52.4	56.8	198	3.01	93		62.1
11.2	13.9	8.46	14	3.77	1.91	15.3	9.99	20	11.7		0.0366
83.9	100	110	91		96.3	94	0.078	96	91.7	99.3	37
0.0113	0.0113	0.01	0.0068		0.0032	0.0337		<0.0021	0.13	0.0523	
6.9		<0.5	<0.5	6.1	5.2		<0.5	13.4	>500	>500	81.7
Medium	High	Medium	Medium	Medium	No GSH adduct	Medium	No GSH adduct	No GSH adduct	Very low	Very low	No GSH adduct

MMV023183	MMV022478	MMV022236	MMV022029	MMV021660	MMV021375	MMV021057	MMV021013	MMV020982	MMV020710	MMV020670	MMV020623
6.88	0.764	>80	4.15	1.73	8.81	9.05	0.629	17.2	72.2	0.99	>80
2.3	30	>30	21	0.31	3.9	>30	3.3	>30	17	7.4	14
>20	44.2	10.3	64.1	5.19	1.52	>20	0.509	5.76	69	0.31	63.2
>20	7.65	>20	11.1	>20	>20	>20	>20	13.8	2.28	0.582	>20
0.0819	>20	>20	>20	7.69	0.864	>20	>20	10.3	>20		3.88
14.6	>20	>20	>20	>20	>20	>20	>20	>20	>20	0.335	>20
	7.88	>20	>20	16.8	12.1	26.2	23.2	>20	7.33	15.6	<0.082
	37.5	82.4	27	10.7	14.8	6.86	7.74	47.4	13.7	11.5	11
	4.79	2.18	6.65	6.36	5.95	0.862	1.78	3.79	13.1	5.37	16.4
	21.2	>120	17.5	83.5	89.4	617	299	24.1	0.958	98.9	10.4
0.259	25	<4.43	30.4	0.0135		0.626	0.339	22.1	555		51.3
0.259	0.0093	0.852	0.0656	0.0135		0.626	0.339	0.573	0.792		0.603
3.93		3.68	5.19	5.08	4.41	3.39	4.57	3.93	4.17	3.06	4.1
328		528		100	46.2	46	165	377		3.89	>960
0.73		0.454		2.58	7.64	5.24	1.9	0.636		66.7	<0.25
28.1		60		13.8	6.98	13.2	25.9	210		2.35	397
25.3		11.8		51.4	102	53.6	27.3	3.38		301	1.79
0.0559		0.00186	0.0199	0.0798	0.517	19.6	6.79	0.993		6.06	4.83
0		105		0	97	89	89	101		92.5	
0		<0.0001		0	0.017	0.0473	0.045	0.0207		0.0835	
	50.1	>500	12.1					<0.5	26.7		78.7
No GSH adduct	No GSH adduct	No GSH adduct	No GSH adduct	Medium	Very Low	Very Low	No GSH adduct	Medium	Very Low	Medium	No GSH adduct

MMV024114	MMV024101	MMV024035	MMV023985	MMV023969	MMV023953	MMV023949	MMV023860	MMV023388	MMV023370	MMV023233	MMV023227
5.28	0.489	3.65	4.95	6.12	17.2	22.9	7.61	5.82	2.82	5.41	27.7
2.5	>30	0.38	13	2.4	4.9	>30	>30	0.33	>30	10	2.7
97.2	27.9	99.7	78.7	11.3	91.3	25.6	41.7	95.9	21	69.1	99.5
9.34	0.198	17.4	<0.0820	>20	>20	>20	>20		1.01	4.22	1.72
19.1	>20	>20	14.1	0.235	>20	>20	>20	>20	>20	>20	>20
>20	>20	4.86		1.09	>20			>20		>20	7.02
7.17	>20	15.1	7.44	12.5	6.95	6.31	>20		2.91		4.13
16.9	26.7	26.5	20.6	14.4	61.9	3.41	90.8	51			13.8
10.6	6.73	6.79	8.74	8.19	2.9	52.6	1.98	3.52			13
2.9	1.67	14.1	4.22	64.9	16.9	1.11	23	27.2			1.93
183	318	37..6	126	0.0506	31.4	478	23.1	19.5			275
0.0428	0.842	0.0029		3.65	0.591	0.767	0.687	0.393			0.43
4.64	2.55	4.61	4.1	267	4.26	4.09	3	3.85	3.73	3.08	4.04
73.5	26.7	611	51.1	2.28	277	11.5	117	460	148		66.1
3.26	8.97	0.392	4.69	NV	0.866	20.9	2.04	0.521	1.62		3.62
11.1	25.8	5.87	10.4	NV	33.9	3.58	29.4	35.2	19.6		7.83
63.6	27.4	121	68.1	0.0637	20.9	198	117	20.1	36.2		90.5
	0.395		7.48	98.3	0.0113	14.5	8.58	0.0206	1.07		0.115
83.8	109	93.2		0.015	51.3	97.3	112				106
0.0017	0.0695	0.0063				0.0287	0.097				0.0307
	9.5	4.7	3.2		>500	69.1	80.1	>500	7.1	39.8	13.3
No GSH adduct	No GSH adduct	No GSH adduct	Medium	Medium	No GSH adduct	Medium	No GSH adduct	No GSH adduct	No GSH adduct	Medium	Medium

MMV026490	MMV026468	MMV026356	MMV026313	MMV026020	MMV024937	MMV024829	MMV024443	MMV024406	MMV024397	MMV024311	MMV024195
12.9	20.5	24.5	>80	>80	13.4	13.7	6.04	2.68	>80	5.28	6.21
8.9	11	>30	>30	>30	5.2	0.7	30	<0.124	>30	12	3.8
83.5	89.3	8.51	23.4	21	87.3	94.9	39.6	92.7	53.1	>20	84.4
1.08	0.43		<0.0820	>20	>20	9.32	0.896	7.75	>20	>20	>20
6.41	19	1.81	17.7	>20	15.5	>20	4.4	6.97	>20	0.0715	>20
	>20	>20	1.05	>20	>20	>20	>20	>20	>20	2.01	>20
1.07	4.3		4.3	>20	>20		9.87	0.775	>20		>20
10.7	8.4	53.9		79.2	103	31.8	10.8	21.9			>120
16.8	21.4	3.33		2.27	1.74	5.64	16.7	8.2			<1.5
1.99	4.48	36.1		>120	32.5	11	4.66	23.8	35.8		
267	119	14.7		<4.43	16.4	48.4	114	22.3	14.9		
0.318	0.219	0.85	1	1	0.153	0.487	0.567	0.0329	0.391		0.348
4.56	4.36	2.97	2.61	2.66	4.56	4.14	3.29	5	4.34	3.62	4.17
27	6.05	>960		22.5	258	528	35.9	105	62.6	403	677
8.86	39.6	<0.25		10.6	0.928	0.454	6.67	2.28	3.82	0.705	0.354
6.93	5.26	62		9.93	216	43	9	72.6	14.5	128	32.9
102	135	11.4		71.4	3.28	16.5	78.7	9.76	48.9	5.53	21.5
4.89	8.15	9.3		2.84	0.0809	0.152	24.4	0.0131	17	0.0724	0.223
78.3	80.6	88.5		89		87		57.7	94	85.7	102
0.0086	0.0098	0.0797		0.0187		0.103		0.0086	0.0061	0.0673	0.0223
5.8	9.8	>500	88.9	25.1	>500		98	6.3	31.9		>500
No GSH adduct	Medium	No GSH adduct	No GSH adduct	Medium	No GSH adduct	No GSH adduct	High	No GSH adduct	No GSH adduct	No GSH adduct	No GSH adduct

MMV063404	MMV062221	MMV054312	MMV053220	MMV047015	MMV045105	MMV032995	MMV032967	MMV031011	MMV030734	MMV028694	MMV026550
>80	10.7	>80	>80	>80	7.47	38.4	14.6	10	5.77	1.67	1.17
>30	14	>30	>30	>30	>30	14	2.7	0.37	>30	26	30
8.81	0.893	4.17	>20	0.191	25.2	75.1	96.7	102	-1.89	53	45.1
>20	>20	>20	>20	>20	12.5	14.7	0.146	>20	>20	3.88	>20
>20	>20	>20	>20	>20	>20	>20	0.878	>20	>20	3.99	>20
>20	9.92	>20	>20	>20	>20	>20	0.65	7.18	>20	>20	>20
46.9	14	12.4		22.4	>20	>20	0.269	>20	>20	>20	11
3.83	12.8	14.4		8.01	32.8		26.3	30		24.6	18.9
2.22	3.72	7.75		4.7	5.48		6.83	5.99		7.31	9.52
240	143	68.6		113	15.4		11.9	18		8.24	5.09
0.503	0.379		0.688	0.719	34.5		44.6	29.5		64.5	104
0.503	0.379		0.688	0.719	0.338		0.241	0.0649	0.898		0.23
4.34	4.36	2.58	2.68		4.23		4.26	4.2		2.99	4.09
83.6	57.7	38.4	10.7	307	42.3		56.3	102		33	61.5
7.11	6.47	7.6	27	0.788	5.66		4.25	2.34		7.26	3.9
24.5	7.69	24.9	2.39	15.6	23.7		32.5	153		12.9	10.2
28.9	92.1	28.5	296	45.3	29.9		21.8	4.64		55.1	69.4
6.81	6.65	12.9	17.6	28.8	12.9		0.742	0.0233		8.18	0.18
92	87.7	86.3	73.7	84.7	96.1		86.5	82.6	53.6	99.4	101
0.0193	0.007	0.283	0.167	0.0973	0.0102		0.0277	0.015	0.87	0.107	0.0237
					12.7	11.3	18	17.1	251	14.6	372
No GSH adduct	No GSH adduct	very low	No GSH adduct	Medium	No GSH adduct	Very low	No GSH adduct	No GSH adduct	Very low	Very low	No GSH adduct

MMV1028806	MMV102872	MMV1019989	MMV099637	MMV090930	MMV085499	MMV085230	MMV085210	MMV085071	MMV084864	MMV084603	MMV069458
1.8	1.07	25.2	>80	2.13	12	>80	21.9	4.65	>80	26.4	11.3301368
>30	>30	>30	>30	3.3	21	>30	>30	0.41	>30	13	30
0.85	>20	>20	>20	0.229	62.4	17.7	32.7	95.9	2.6	82.1	>20
3.94	>20	>20	>20	0.0929	>20	>20	>20	1.51	>20		>20
>20	>20	9.82	>20	>20	>20	>20	2.1	0.38	>20	>20	>20
>20	>20	>20	>20	11.1		>20	>20		>20	>20	>20
13.5	29.7		19.1	5.38	>20	>20	0.96	>20	>20		
13.3	6.05		9.41	33.4	52.6	55.4	2.69	32.8		12.4	
7.14	21.7		5.45	11.6	3.41	3.24	66.8	5.47		14.5	
74.5	24.5		97.5	45.7	>120		1.17	25.2	11.7	3.95	
	0.778		0.953		<4.43		455	21.1	45.4	135	
	0.778		0.953		0.703	0.83	0.313	0.905	0.848	0.399	
4.29	4.13		3.58	4.03	2.97	3.33	4.17	3.05	3.24	4.42	
46.7	53.6		71.4	13.9	949	67.7	14.6	399	136	44.4	240
10.1	4.6		6.2	17.8	0.252	3.54	16.4	0.6	1.76	5.39	1.11
11.4	50.6		10.9	75.5	604	7.44	6.47	66.2	13.9	12.6	27.3
62.4	14		64.9	9.39	1.17	95.2	109	10.7	51	56.2	26
4.96	9.44		7.86	20.3	4.43	22.1		0.51	4.81	7.44	5.97
85.7	88		80.7	88.3		102		88.7	83		116
0.0079	0.0016		0.0087	<0.0021		0.0913		0.197	0.069		0.57
					49.6	>500	16.8	113	10.1	12	<5
Very Low	No GSH adduct	No GSH adduct	Medium	Very Low	Very Low	Very low	Medium	Medium	Very low	Very Low	Medium

MMV200748	MMV188296	MMV161996	MMV153413	MMV146306	MMV1236379	MMV1198433	MMV1110498	MMV1088520	MMV1037162	MMV1030799	MMV1029203
20.3	>80	>80	1.75	9.16	12.9	>80	>80	1.84	>80	0.709	22
12	>30	>30	8.6	>30	>30	>30	>30	14	>30	2.8	>30
4.76	>20	>20	7.65	>20	>20	10.4	1.09	13.3	2.54	3.22	6.24
5.01	>20	10.5	16.6	>20	19.7	>20	>20	>20	7.44	>20	>20
>20	>20	16.3	>20	>20	>20	>20	>20	0.37	>20	>20	>20
5.7	>20	17.9	>20	>20	>20	>20	>20	9	>20	>20	>20
15.4	80.8	9.2	44	30.3	4.29	>20	>20	9.72	12.5	13	25.6
11.7	2.22	19.5	4.08	5.93	41.8	25.2	>120	18.5	14.4	13.8	7.01
1.74	25.9	1.92	7.84	5.57		7.13	<1.5	0.871	1.68	1.22	13.2
306	20.5	276	67.8	95.4		13	40.2	611	318	437	40.3
0.194				0.505		40.8	13.2	0.617	0.569		0.676
0.194				0.505		0.76	0.975	0.617	0.569		0.676
4.95	2.53	3.99	4.16	4.21	2.3	3.52	2.47	4.26	3.85	4.24	3.54
96.5	208	55.6	3.51	314	2.7	55.6	>960	45.5	63.7	11.1	116
3.92	1.27	4.35	73	2.6	96.8	4.31	<0.25	6.71	4.25	23.4	2.08
11.1	55.8	28.5	1.92	18.5	1.57	30.1	30.3	2.53	18.7	2.98	42.2
63.8	12.7	24.9	370	38.4	451	23.6	23.4	280	37.9	238	16.8
	15.8	9.59	14.4	17.8	23.4	7.03	23.2	1.86	19.6	7.26	19.7
80.3	104	60	44.3	88.7	0	64.5	107	85.7	0	45.7	0
<0.0163	0.45	0.026		0.0673		0.0597	0.393	0.044	0		0
						24.9	>500				
Medium	No GSH adduct	No GSH adduct	Medium	No GSH adduct	No GSH adduct	No GSH adduct	Medium	Medium	Medium	Medium	No GSH adduct

MMV553002	MMV495543	MMV461553	MMV407834	MMV407539	MMV393995	MMV393144	MMV392832	MMV272144	MMV228911	MMV202553	MMV202458
>80	5.28	9.15	29.8	18.6	4.64943007	10.4	2.29	0.764	13	>80	>80
>30	>30	>30	>30	10	30	2.4	>30	23	5.3	>30	>30
>20	4.22	5.61	32	98.3	>20	91.6	-0.562	>20	0.0508	8.64	2.04
>20	3.44	>20	2.77		>20	2.64	0.577	>20	3.15	>20	6.18
>20	>20	>20	4.2	4.5	>20	>20	13.1	>20		>20	>20
>20	>20	>20	>20	>20	>20	>20	13.6	>20	0.151	>20	11.5
	3.66		0.663	0.158	>120	3.54	1.33		7.34		
	49.1		2.2	1.03	<1.5	38.5	13.7		24.5		
	4.87		81.5	174	67.5	4.67	13.1		5.43		
	109		2.53	0.8	7.88	7.6	14.8		97.9		
	0.434		210	665	0.939	69.9	36				
	0.434		0.129	0.34	0.939	0.518					
	4.32	5.7	4.81	4.42	1.35	4.2	4.2	1.71	4.8		4.09
	8.14	114	27	20.5	>960		30.5	3.1	66.8	28.6	9.04
	33.5	2.6	8.88	11.7	<0.25		7.86	80.3	3.94	8.5	28.3
	7.36	74.4	10.6	3.24	85.2		38.7	2.16	59.6	5.79	3.33
	96.3	9.53	66.9	219	8.32		18.3	328	11.9	122	213
	11.8		1.91	1.79		0.294	0.122	18.8		6.38	7.95
0	87.7	91	76		94.3	87.5	91.1		82	23	0
0	0.0183	<0.002	0.0014		0.377	0.15	0.0117		0.0021		
			11.3	10		63	98.7				
No GSH adduct	No GSH adduct	No GSH adduct	High	Medium	No GSH adduct	Medium	High	No GSH adduct	No GSH adduct	No GSH adduct	Very Low

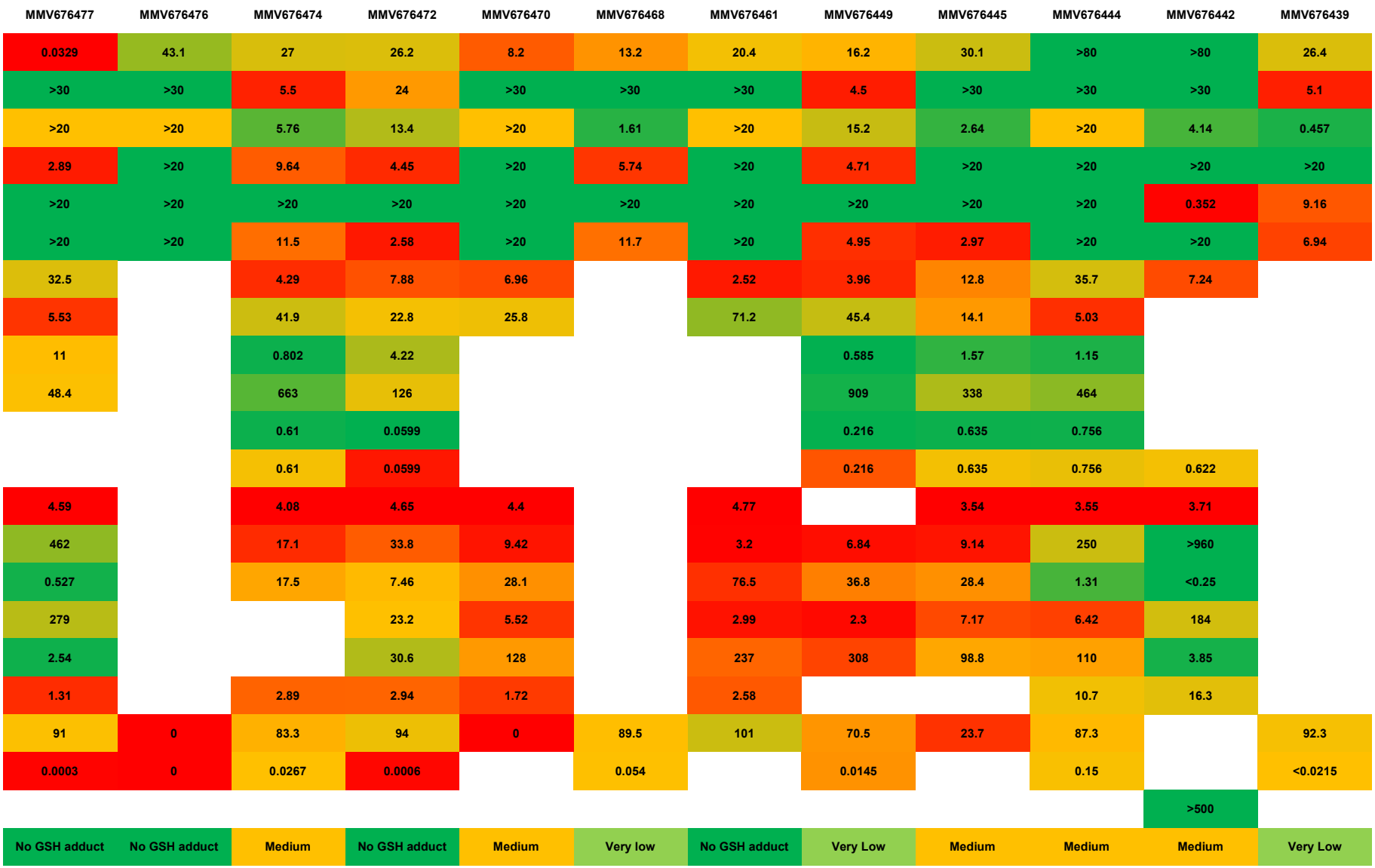
MMV661713	MMV659010	MMV659004	MMV658993	MMV658988	MMV652003	MMV637953	MMV637229	MMV634140	MMV611037	MMV595321	MMV560185
3.08	>80	10.2	27.9	11.3	0.413	>80	9.28	26.4	72.1	0.0401	22.9
2.1	30	7.1	>30	9.1	>30	>30	0.4	>30	>30	>30	4.6
0.0406	38.8	90.3	22	82.5	2.41	11	99.6	>20	>20	>20	95.2
0.695	2.55	1.37	13.2	0.101	>20	>20	>20	>20	1.58	5.56	0.459
>20	18.6	10.5	>20	15.7	>20	>20	<0.0820		>20	>20	>20
0.14	17.4	>20	>20	>20	>20	>20	>20	>20	>20	>20	>20
13.6	>20	2.56	>20		>20	>20	3.37	>120	3.53	22.9	12.7
13.2	15.6	11.6	31.6	4.74	75.4		57.5	<1.5	50.9	7.86	11
11.4	11.5	15.5	5.69	37.9	2.38		3.13	63.8	12.6	8.55	16.3
46.7	1.29	2.58		0.964			4.79	8.33	42.1	62.2	1.79
0.026	412	206		551			111	0.47	0.789	0.556	297
0.026	0.538		0.871	0.145	1		0.0685	0.47	0.789	0.556	0.177
4.86	4.5	5.32	3.18	4.86	2.53		4.78	3.14	3.75	4.06	4.85
111	30.8	47.9	467	9.61	>960		149	376	13.5	152	22.3
2.63	7.79	5	0.513	24.9	<0.25		1.61	0.744	19	2.19	10.8
31.7	7.83	38.6	17.1	7	250		8.51	140	13.5	12.5	7.59
22.4	90.5	18.4	41.5	101	2.84		83.3	5.05	52.5	56.5	93.4
0.0386	21	0.077	10.5	0.585	10.1		0.938	0.0183	40.1	2.12	0.672
114	108	102	92.4	105	123		78.8	88.7	74.3	0	73.8
0.0012	0.0097	0.0012	0.0603	0.0042	0.0123		0.0307	0.28	0.0117	0	0.0018
	52.4	5.3	>500	19.2	500	NA	>500				11.5
Very Low	Medium	Medium	Very low	Medium	No GSH adduct	No GSH adduct	Very Low	Very low	Medium	No GSH adduct	Very Low

MMV675998	MMV675997	MMV675996	MMV675995	MMV675994	MMV675993	MMV675969	MMV675968	MMV671636	MMV668727	MMV667494	MMV663250
5.47	14	>80	>80	13.6	1.63	35.9	3.44	>80	>80	5.63	15.2
>30	0.68	>30	>30	>30	>30	>30	11	>30	>30	>30	26
22.4	>20	24.8	8.31	28	5.44	26	71.7	4.44	19.5	>20	53.4
>20	>20	17	>20	>20	>20	>20	>20	>20	13.1	>20	>20
>20	3.01	>20	>20	>20	>20	>20	9.54	>20	>20		5.01
>20	6.61	>20	>20	>20	>20	>20	>20	>20	>20	2.95	0.929
>20		1.03	>20	>20	>20	12.2	10.4	>20	>20	20.4	0.592
		1.06	24.9	>120	65.9	14.6	39.8			8.83	48
		169	7.2	<1.5	2.72	12.3	4.52			6.94	3.74
31.5		0.706	0.933	23.4		0.74	3.5			76.6	55.4
16.9		753	570	22.7		718	152			0.0679	9.6
		0.637	1	0.677	0.702	0.504	0.441	0.0707		0.0679	
		4.22	2.67	3.16	2.7	4.61		5.05		4.41	3.71
	141	10.8		>960	134	32.9	88.6	>960	862	397	393
	1.68	22.1		<0.25	1.78	7.29	2.7	<0.25	0.278	0.924	0.609
		4.36		96.1	16.1	4.03	6.19	>960	462	58.8	187
		163		7.377	44.1	176	115	<0.738	1.53	12.1	3.79
		3.39	18.7	10.7	23.1	20.8	0.276	0.013		0.0098	
90.9	92	95.6	106	97.4	93	110	97		103	79.3	100
0.38	0.02	0.0223	0.603	0.0463	0.103	0.0195	0.0012		0.054	0.031	0.0963
68.6		>500	>500	37.4	486	50.3	95	8.3	68.2		454
No GSH adduct	Medium	Medium	Medium	Medium	Medium	No GSH adduct	Very low	No GSH adduct	No GSH adduct	No GSH adduct	No GSH adduct

MMV676186	MMV676182	MMV676162	MMV676161	MMV676159	MMV676064	MMV676063	MMV676057	MMV676053	MMV676050	MMV676048	MMV676008
>80	>80	>80	28.9	>80	39	3.59	4.03	>80	53	37.1	5.73
>30	24	>30	12	>30	>30	>30	>30	>30	>30	11	>30
24.6	54	4.32	74.4	29	8.55	17.4	32.1	9.43	36.7	77.7	24.7
>20		1.57	>20	3.32	>20	>20	>20	>20	>20	11.49	11.6
>20	>20	>20	>20	>20	>20	>20	10.5	>20	1.27	>20	>20
>20	>20		>20	>20	>20	>20	>20	>20	>20	10.4	>20
8.85		>20	>20	3.51	>20	>20	>20	>20	>20	<0.0820	
1.44	>120	>120	22.8	10.8	5.04	19.3		47.6	9.91	9.35	>120
125	<1.5	<1.5	7.88	16.6	35.7	9.33		3.78	18.1	19.2	<1.5
	63.9	20.9	1.12		1.37	30.5	38.7	8.81	3.09	2.56	2.25
	8.31	25.4	472		387	17.4	13.7	60.3	172	207	237
	0.742	0.795	0.328	0.677	0.837	0.0471		0.7	0.664	0.507	0.688
4.81	3.19	3.4	4.06	4.06	3.63	0.25	4.97	3.52	3.41	4.26	
3.75	>960	76.8	34.5	19.4	36.3	>960	35.9	131	39.2	17.1	65.3
64	<0.25	3.12	6.95	12.4	6.6	<0.25	6.68	1.82	6.1	14	3.67
2.21	33.9	60.1	9.89	16.8	7.69	>960	84.7	13.8	7.98	8.9	11.2
320	20.9	11.8	71.6	42.1	92.2	<0.738	8.37	51.2	88.8	79.7	63.3
	0.0528		2.23	14.7	22.9	0.00582	3.12	18.2	9.54	3.64	34
	96.1	93.8	101	89.8	89.2	103		87.4	103	96.8	93
	0.137	0.0002	0.0263	0.0096	0.055			0.039	0.0597	0.0163	0.002
<0.5	>500	363		31.8	52.2	1.5	24.8	13.7		>500	>500
No GSH adduct	No GSH adduct	No GSH adduct	Very low	No GSH adduct	Medium	No GSH adduct	No GSH adduct	Medium	Medium	No GSH adduct	Very low

MMV676383	MMV676382	MMV676380	MMV676379	MMV676377	MMV676358	MMV676350	MMV676270	MMV676269	MMV676260	MMV676204	MMV676191
3.43	4.63	>80	>80	>80	2.64	6.9	21.8	>80	10.3	18.5	>80
>30	30	>30	>30	>30	>30	2.4	25	>30	6.5	30	>30
0.442	38.1	7.42	1.04	0.838	0.061	97.5	57.2	28.6	80.6	40.6	0.5
5.35	1.85	>20	>20	>20	14.1	0.87	>20	>20	6.47	2.27	9.29
>20	4.27	>20	>20	>20		>20	>20	>20	>20	12.2	>20
>20	>20	>20	15.9	3.35	>20	>20	>20	>20	>20	>20	>20
6	13.1	>20		7		1.38	>20	7.94	8.5		
30	15.1	30.7		25.7		18.8	15.4	10.3	36.7	5.61	
0.923	11.9	6.21		4.71		9.54	11.7	17.5	4.91	32	
576	7.76			113		8.43	5.56	4.08	13.6	0.815	
0.401	68.5			0.701		63.1	95.8	130	39.1	652	
0.401	0.625			0.701		0.0789	0.132	0.514	0.619	0.704	
4.06	3.84	4.03	4.26	2.94	4.49	4.2	4.78	4.31	4.16	4.18	
11.8	24.2	33.6	37.3	124.5	169	114	35.7	38.2		10.2	>960
26.2	9.9	7.13	10.2	2.19	1.46	2.11	6.72	6.27		23.4	<0.25
3.9	11	28.3	10.1	26.1	45	17.4	12.6	14.2		4.96	182
179	64.3	25.1	70.4	27.2	15.7	40.8	56.1	50		143	3.89
22.1	12.1	0.529	5.88	12.3	14.2	0.0134	5.91		0.212		
72.7	96.3	47	93.3	79.7	88	65.8	97.5	71	96.7	104	102
0.0101	0.032		0.0293	0.0723	0.0273	0.015	0.0045	0.0533	0.22	0.012	0.031
	12.9	37				2.8	9.9	66.5	49.9	19.6	>500
Very Low	High	No GSH adduct	No GSH adduct	Medium	No GSH adduct	Medium	No GSH adduct	Very low	No GSH adduct	High	No GSH adduct

MMV676431	MMV676412	MMV676411	MMV676409	MMV676406	MMV676401	MMV676398	MMV676395	MMV676389	MMV676388	MMV676386	MMV676384
72.8	1.92	0.729	1.99	11.8	6.95	>80	>80	36.2	1.85	31.6	12.4
24	>30	22	4.8	11	0.37	>30	>30	>30	6.7	>30	>30
1.92	>20	1.25	0.346	>20	0.5	32.7	0.682	>20	10.8	9.86	0.228
18.6		5.02	16.7	8.83	>20	8.44	>20	>20	10.9	3.31	12.4
>20	>20	>20	>20	>20	>20	>20	>20	>20	>20	>20	>20
18.2	>20	>20	>20	>20	>20	>20	>20	>20	15	0.406	7.22
11.8		9.98	7.67	11	12.1	>20		22.4			
15.2		18	23.4	16.3	14.8	32.9		8.03			
1.95		0.74	0.884	6.49	3.02	5.47		3.53			
272		716	601	81.9	176	2.4		150			
0.222				0.132	0.245	221		0.824			
0.222				0.132	0.245	0.745		0.824			
4.73	4.78	3.78	3.8	4.83	4.21	3.3		3.37			
17.2		14.3	12.6	410	125	74.2	38	92.4		59	
14.5		17.1	19.8	0.655	2.25	3.23	6.3	4.49		4.47	
5.19		6.81	2.16	16.5	6.8	4.82	141	10.6		10.1	
137		104	329	42.9	104	147	5.04	66.8		70.4	
5.18		10.3	8.25	2.8		19.5	17.4	5			
39.7			5	90.3	87.7	101	86.3	85		53.7	82
				0.0029	0.0183	0.0467	0.0343	0.099		<0.0287	<0.023
						51.8					
Medium	No GSH adduct	Medium	No GSH adduct	No GSH adduct	No GSH adduct	Medium	No GSH adduct	Medium	Medium	Medium	High


MMV676477
MMV676476
MMV676474
MMV676472
MMV676470
MMV676468
MMV676461
MMV676449
MMV676445
MMV676444
MMV676442
MMV676439
0.0329 43.1 27 26.2 8.2 13.2 20.4 16.2 30.1 >80 >80 26.4
>30 >30 5.5 24 >30 >30 >30 4.5 >30 >30 >30 5.1
>20 >20 5.76 13.4 >20 1.61 >20 15.2 2.64 >20 4.14 0.457
2.89 >20 9.64 4.45 >20 5.74 >20 4.71 >20 >20 >20 >20
>20 >20 >20 >20 >20 >20 >20 >20 >20 >20 0.352 9.16
>20 >20 11.5 2.58 >20 11.7 >20 4.95 2.97 >20 >20 6.94
32.5 4.29 7.88 6.96 2.52 3.96 12.8 35.7 7.24
5.53 41.9 22.8 25.8 71.2 45.4 14.1 5.03
11 0.802 4.22 0.585 1.57 1.15
48.4 663 126 909 338 464
0.61 0.0599 0.216 0.635 0.756
0.61 0.0599 0.216 0.635 0.756 0.622
4.59 4.08 4.65 4.4 4.77 3.54 3.55 3.71
462 17.1 33.8 9.42 3.2 6.84 9.14 250 >960
0.527 17.5 7.46 28.1 76.5 36.8 28.4 1.31 <0.25
279 23.2 5.52 2.99 2.3 7.17 6.42 184
2.54 30.6 128 237 308 98.8 110 3.85
1.31 2.89 2.94 1.72 2.58 10.7 16.3
91 0 83.3 94 0 89.5 101 70.5 23.7 87.3 92.3
0.0003 0 0.0267 0.0006 0.054 0.0145 0.15 <0.0215
>500
No GSH adduct No GSH adduct Medium No GSH adduct Medium Very low No GSH adduct Very Low Medium Medium Medium Very Low

MMV676539	MMV676536	MMV676528	MMV676526	MMV676524	MMV676520	MMV676512	MMV676509	MMV676501	MMV676492	MMV676480	MMV676478
45.1	>80	26.4	30.5	3.31	28.3	0.0763	14.7	2.64	>80	12.3	>80
29	>30	15	7.4	>30	9.5	>30	>30	>30	>30	1.5	0.38
3.93	24.2	90.3	9.08	3.12	>20	2.66	6.81	>20	9.41	81.4	>20
>20	1.38	>20	13.1	>20	>20	14.8	2.95	>20	>20	>20	>20
>20	5.78	1.44	>20	>20	>20	>20	>20	>20	>20	>20	>20
17.4	>20		>20	>20	1.89	0.0912	>20	>20	>20	>20	>20
11.9		0.834	11.9			29.5	12.1	27.4	>20	5.7	1.96
15.1	5.31	4.97	15.1			6.1	14.9	6.55		33.4	91.7
4.29	33.9	36.1	0.829			8.25	4.84	20.8		5.39	3.33
124	0.752	1.08	641			64.4	110	25.6		5.78	160
0.539	707	492	0.319			0.647				92	
0.539	0.391	0.386	0.319			0.647				0.06	
3.83	4.22	4.03	4.36	2.98		3.19	5.19	3.6		5.33	4.45
8.43	19.3	18.8	36.1	76.5		60.4	83.9	25	59.1	75.3	11
28.4	12.4	12.8	6.67	3.44		3.98	2.9	9.6	4.06	3.18	24
5.32	3.69	5.74	5.29	6.3		23.7	10.2	32.3	32.3	15.7	29
133	192	123	134	112		29.9	69.8	21.9	21.9	45.1	24.5
11.2	19.9	8.98	1.96	7.24		15.5	0.376	28.3			14.6
100	83.3	96	98	94.3		91.3	89.5	84.3	121	75.2	96.5
0.04	0.0133	0.015	0.017	0.12		0.113		0.0015	0.0173	0.0003	0.0061
	<0.5	22.8							>500	6.3	
High	Medium	Very Low	Medium	Medium	No GSH adduct	Very Low	No GSH adduct	Very Low	No GSH adduct	No GSH adduct	Medium

MMV676603	MMV676602	MMV676600	MMV676599	MMV676597	MMV676589	MMV676588	MMV676584	MMV676571	MMV676558	MMV676555	MMV676554
>80	<0.264	4.83	19	>80	26.4	5.83	17.5	3.48	0.415	>80	4.82
>30	>30	>30	24	9.8	7.7	2.2	>30	>30	30	10.3	6.9
>20	30.8	6.41	50.6	>20	>20	0.225	1.25	>20	3.22	>20	>20
>20	>20	>20	6.01	>20	3.62	>20	>20	7.49	2.53	>20	>20
>20	>20	>20	6.46	>20	>20	11.3	>20	>20	>20	>20	>20
>20	16.5	>20	>20	>20	9.5	>20	>20	>20	15.3	4.82	7.13
	>20	>20	>20	1.28	5.16	6.85	2.09	14.6	4.96		>120
	56.6	36.1	15.1	140	34.8	26.2	86	12.3	36.2		<1.5
	3.25	4.98	12.6		1.41	1.41	3.52	10.3	41		11.7
	33.7	53.5	1.156		376	376	151	51.4	13		45.3
	15.8	9.93	407			0.445			0.316		0.818
	0.673	0.805	0.127			0.445			0.316		0.818
	4.2	3.47	4.6	4.68	5	4.01	4.08	4.91	4.21		2.63
		197	21.1	3	5.8	8.5	2.4	95	51.2		150
		1.22	11.4	79.4	41.5	33	100	4.29	4.75		1.6
	340	433	7.66		2.26	4.63	2.42	12.3	85.2		49.4
	2.17	1.64	92.7		313	153	292	57.7	8.32		14.4
	0.411	0.0298	2.99	5.15	9.29	6.44	12.7	2.66	0.66		2.4
	87	94.3	100		50.3	96.3	67	81	92	89.7	81.7
	0.13	0.303	0.049		0.0041	0.0018	0.0072	0.0015	0.0001	<0.0219	0.32
	>500	1.1	15.2								
No GSH adduct	High	No GSH adduct	Medium	No GSH adduct	No GSH adduct	No GSH adduct	Medium	Medium	No GSH adduct	Medium	Very Low

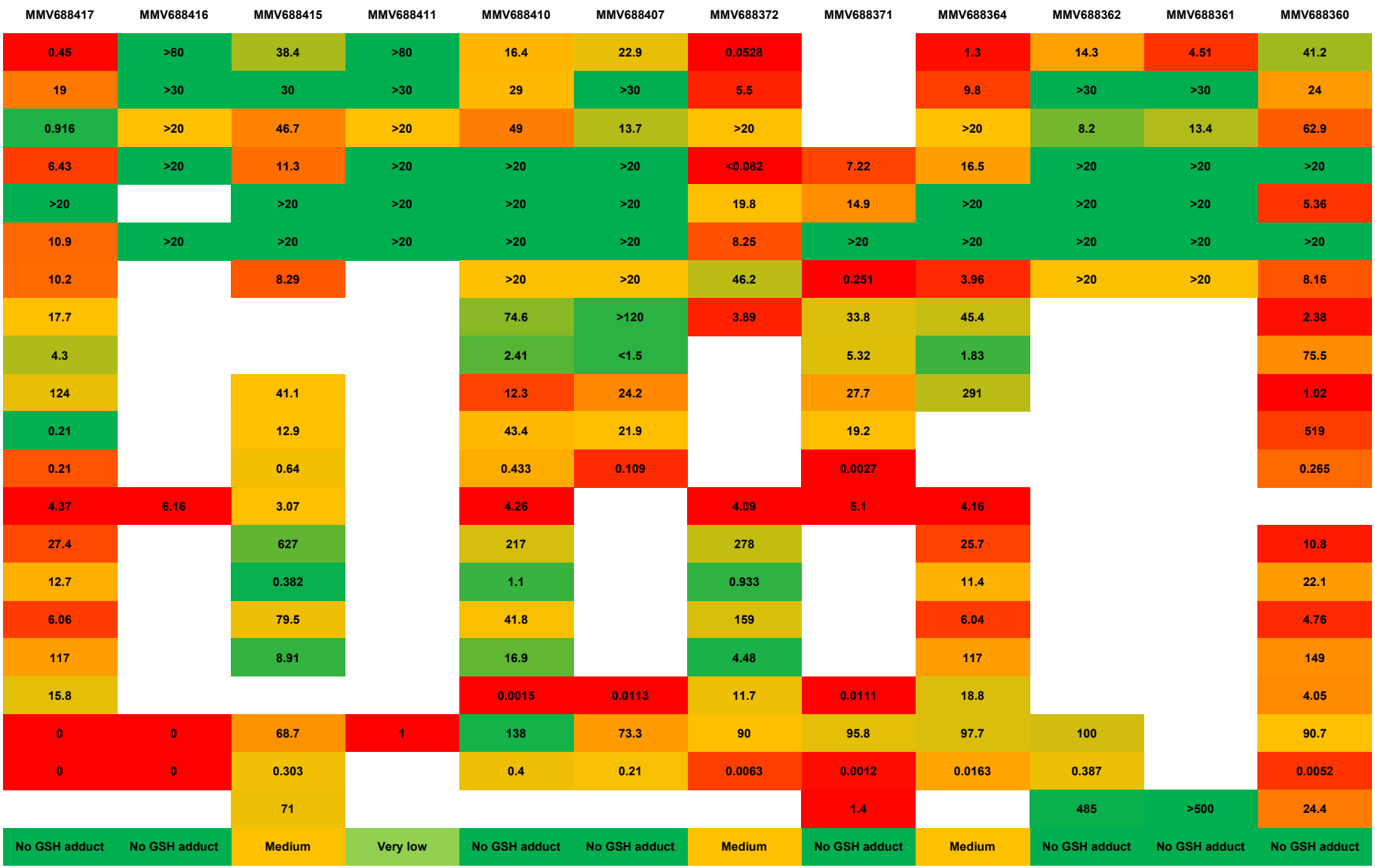
MMV687189	MMV687188	MMV687180	MMV687172	MMV687170	MMV687146	MMV687145	MMV687138	MMV676881	MMV676877	MMV676605	MMV676604
27	12.1	14.1	>80	22.1	>80	>80	30.7	22.5	12.8	4.67	0.179
14	2.8	>30	5.5	>30	>30	>30	>30	>30	30	1.2	26
3.35	0.769	>20	>20	10.4	12.4	>20	0.776	0.245	49.3	83	52.1
>20	5.97	16.8	>20	4.62	>20	>20	>20	>20	2.25	12.4	2.58
>20	>20	>20	>20	>20	>20	>20	>20	>20	6.46	3.72	10
5.47	9.18	3.1	>20	18.2	0.625	>20	>20	>20	>20		>20
16.1	11.5		14.8	5.03	15.3	55		16.1	11.1	>20	>20
11.1	15.6		12.1	35.7	11.7	3.27		11.1	22.6	107	>120
3.81	0.811		11.4	0.758	2.17	12.1		3.17	7.96	1.74	<1.5
139	655		46.4	701	245	43.9		168	15.2	38.3	35.4
0.721	0.4		0.492		0.0849	0.857		0.295	35	14.2	17.7
0.721	0.4		0.492		0.0849	0.857		0.295	0.274	0.408	0.896
3.4	4.04		4.03	3.83	5.18	2.52		4.45	4.25	3.8	2.7, 2.96
40.4	11.4		199	7.1	102	181		58.7	53.6	94	376
6.47	21.4		2.23	33.8	2.45	1.33		4.2	4.47	2.55	0.637
7.91	2.33		23.4	2.8	24.8	20.6		10.9	36.7	126	28.2
89.6	304		30.3	253	28.6	34.4		65	19.3	5.62	25.2
6	10.1		23.4	10.5		20.6		16.1	10.2	0.678	2.49
42.7	0	84.7	0	106	83.7	0	0	97.3		98.7	82.1
	0	<0.0022	0	0.0187	0.0019	0	0	0.0233		0.0065	0.117
									4.5	8.6	9.2
No GSH adduct	Medium	No GSH adduct	Medium	No GSH adduct	No GSH adduct	No GSH adduct	Medium	No GSH adduct	No GSH adduct	Very Low	No GSH adduct

MMV687706	MMV687703	MMV687700	MMV687699	MMV687696	MMV687273	MMV687254	MMV687251	MMV687248	MMV687246	MMV687243	MMV687239
6.43	5.15	>80	11.4	1.69	17.7	15.2	0.598	7.56	>80	>80	20.7
13	2.4	>30	>30	>30	30	2.7	>30	6.4	>30	>30	>30
79.3	0.538	>20	1.07	>20	>20	7.64	>20	7.08	13.7	2.07	12
17.2	>20	11.2	1.48	>20	>20	>20	>20	>20	7.25	15.9	6.12
>20		>20	>20	>20	7.87	>20	>20	>20	>20	17.1	>20
>20	8.19	>20	>20	>20	17.3	>20	>20	4.12	>20	>20	>20
>20	8.48	63.7	10.9		3.78	7.09		50.5			3.29
39.1	21.2	2.82	16.5		47.5	25.3		3.58	6.17		54.7
4.59	2.81	>120	4.77	63.3	6.05			36.4	29.1		0.772
8.94	189	4.43	112	8.44	87.8			14.6	11.1		689
59.4		0.906	0.295		<0.010			0.0747	47.9		
0.0554		0.906	0.295		<0.010			0.0747	0.137		
4.29	4.29	0.779	4.37	5.7	5.46	4.39	0.615	3.81	3.85	3.73	4.32
271	43.5		13.4	370		5.5		482	27.5	38.5	8.11
0.885	9.1		18	0.76		43.6		0.658	8.7	6.2	29.9
33.3	3.61		17	229		2.6		138	520	44	2.75
21.3	196		41.7	3.09		273		5.12	1.36	16.1	258
0.0192		0.0503	15.2	0.0083	0.0338		18.5	0.0411	0.0598		3.9
97.3	65.7	74.3	83	86.5	85.7	0		90.7	90.3		83.3
0.027	0.0193	0.283	0.014	<0.0001	0.0005	0		0.0157	0.0048		0.0123
216									0.6		
No GSH adduct	No GSH adduct	No GSH adduct	Medium	Very Low	No GSH adduct	No GSH adduct	Very Low	Medium	No GSH adduct	Medium	No GSH adduct

MMV687800	MMV687798	MMV687796	MMV687794	MMV687776	MMV687775	MMV687765	MMV687762	MMV687749	MMV687747	MMV687730	MMV687729
6.59	>80	>80	>80	10.9	15.4	4.72	10.8	8.66	3.79	8.32	6.67
1.9	>30	>30	>30	>30	>30	3	28	8.5	3.2	1.4	15
71	13.1	1.72	2.81	9.66	1.12	1.24	56.5	0.803	1.18	>20	>20
>20	>20	>20	>20		3.62	1.33	>20	0.977	>20	8.74	>20
>20	>20	>20	>20	>20	2.64	0.0355	>20	<0.01	>20	>20	>20
>20	>20	>20	>20	>20	>20	0.0146	>20	<0.01	13	18.6	>20
>20	>20	>20			3.39	13.1	>20	24.2	8.53	6.38	
24.5	72.5					13.7	20.2	7.44	21.1	28.2	
7.57	2.48					10.6	8.89	18.5	255	11.37	
25.2	>120					50	10.7	28.7	2.09	388	
21.1	<4.43					0.0232	49.8	0.185	0.42		
<0.001	0.795					0.0232	0.15	0.185	0.42		
5.44	1.63			5.36		4.06	4.5	3.68	4.11	5.18	
479						142	52.9	172	77.3	13.8	
0.5						1.85	4.53	1.4	8.72	17.8	
						106	21.7	28.8	7.2		
						6.69	32.7	24.6	98.4		
0.0351	0.795					0.0329	8.64	0.128	5.05		
96	113			91.9	91.9	85.3	94	0	71.7	0.054	
<0.0001				<0.0027	<0.0026	0.0157	0.003	0	0.0085		
1.2	>500	>500		<0.5	0.51		7.1				
No GSH adduct	Very low	Very low	No GSH adduct	No GSH adduct	No GSH adduct	No GSH adduct	Very Low	No GSH adduct	Medium	High	Very low

MMV688262	MMV688180	MMV688179	MMV688178	MMV688125	MMV688124	MMV688122	MMV687813	MMV687812	MMV687807	MMV687803	MMV687801
2.26	0.0454	6.75	25.7	7.03	5.31	6.86	3.82	2.41	0.658	>80	>80
>30	>30	>30	>30	1.4	4.1	>30	8.2	12	2	>30	>30
>20	31.8	31.7	4.84	>20	>20	1.2	11.8	6.72	10.8	-1.91	2.21
>20	11.1	2.93	>20	13.8	>20	>20	7.12	>20	>20	>20	>20
>20	>20	>20	>20	19.9	2.11		>20	>20	>20	>20	>20
>20	>20		>20	0.789	1.17	8.93	>20	17.5	>20	>20	>20
23	5.1	2.15	>20	12.4	4.86	16.9			40	>20	>20
7.81	55.6		32.7	14.5	37	10.6			4.5		
34	3.23		5.5	2.42	1.27	4.4			21.5		
15.6	>120		16.7	219	420	121			24.8		
	<4.43		31.8	0.15	0.188						
	0.487		0.496	0.15	0.188					0.7	0.939
4.85	3.27		5.63	4.01	4.42	3.09			5.14	1.32	
220	>960		138	416	152	82.3	474		22.7	>960	
1.33	<0.25		1.73	2.5	1.68	3.45	0.81		10.9	<0.25	
113	249		189	11.4	3.73	13	92.4		36.9	>960	
6.25	2.84		3.74	62.4	190	54.5	7.67		19.2	<0.738	
0.0409			0.0587	0.0673	0.395	27.1			0.208		
84	93.3			92	88.5	100	79.7	97	102	74.3	stable
0.0004	0.053			0.0457	0.016	0.0583	0.013	0.0547	0.0022	0.63	
		47.2	8.4							410	460
Medium	No GSH adduct	Medium	No GSH adduct	No GSH adduct	No GSH adduct	Medium	Very low	No GSH adduct	No GSH adduct	No GSH adduct	Very Low

MMV688352	MMV688350	MMV688345	MMV688330	MMV688327	MMV688313	MMV688283	MMV688279	MMV688274	MMV688273	MMV688271	MMV688270
	74.7	11.7	>80	8	7.5	8	10	10.9	9.72	13.5	24.3
	>30	16	>30	>30	6.4	6.4	4.3	5.4	17	>30	12
	>20	>20	>20	>20	89.1	98.1	65.8	80.2	72.2	13.6	82.2
	2.66	>20	>20	>20		>20	>20	>20	0.147	1.29	1.95
	>20		>20	>20	3.14	>20	>20	18.7	>20	10.2	6.55
	13.8	>20	>20	>20	>20	>20	16.2	>20			>20
			>120	71.4		0.61	>20	>20	3.17	2.05	4.76
			<1.5	2.52	13.2	13.7	16.5	13.6	23.2		8.07
			48.9	>120	13.6	13.1	10.9	13.2	7.75		22.3
			10.9	<4.43	6.87	13.7	3.72	16.1	111		3.91
			0.56		77.4	38.7	143	33.1	4.8		136
			0.56		0.126	0.098	0.0319	0.0229	0.0213		0.101
		2.98	3.67	2.22	4.58	4.4	4.85	4.81	4.69		4.89
		80	71.8	>960	27.3	44.3	127	26.1	47.1		19.7
		3.11	3.58	<0.25	8.78	5.4	1.88	9.19	5.09		12.2
		33	59	>960	17.6	34.7	10.1	55.6	59.6		21.9
		21.5	12	<0.738	40.2	20.4	70	12.8	11.9		32.4
			11.8	0.0969	5.6		0.0553		0.0312		
	26.7	85	91	98	94	104	103	85.3	91.7		95.5
		0.217	0.0553	0.13	0.0036	0.002	0.016	0.0011	<0.0017		0.0026
					3.8	<0.5	126	5.6	22.9		3.7
Very low	No GSH adduct	No GSH adduct	No GSH adduct	No GSH adduct	Medium	No GSH adduct	No GSH adduct	Medium	No GSH adduct	No GSH adduct	Medium



MMV688417	MMV688416	MMV688415	MMV688411	MMV688410	MMV688407	MMV688372	MMV688371	MMV688364	MMV688362	MMV688361	MMV688360
0.45	>80	38.4	>80	16.4	22.9	0.0528		1.3	14.3	4.51	41.2
19	>30	30	>30	29	>30	5.5		9.8	>30	>30	24
0.916	>20	46.7	>20	49	13.7	>20		>20	8.2	13.4	62.9
6.43	>20	11.3	>20	>20	>20	<0.082	7.22	16.5	>20	>20	>20
>20		>20	>20	>20	>20	19.8	14.9	>20	>20	>20	5.36
10.9	>20	>20	>20	>20	>20	8.25	>20	>20	>20	>20	>20
10.2		8.29		>20	>20	46.2	0.251	3.96	>20	>20	8.16
17.7				74.6	>120	3.89	33.8	45.4			2.38
4.3				2.41	<1.5		5.32	1.83			75.5
124		41.1		12.3	24.2		27.7	291			1.02
0.21		12.9		43.4	21.9		19.2				519
0.21		0.64		0.433	0.109		0.0027				0.265
4.37	6.16	3.07		4.26		4.09	5.1	4.16			
27.4		627		217		278		25.7			10.8
12.7		0.382		1.1		0.933		11.4			22.1
6.06		79.5		41.8		159		6.04			4.76
117		8.91		16.9		4.48		117			149
15.8				0.0015	0.0113	11.7	0.0111	18.8			4.05
0	0	68.7	1	138	73.3	90	95.8	97.7	100		90.7
0	0	0.303		0.4	0.21	0.0063	0.0012	0.0163	0.387		0.0052
		71					1.4		485	>500	24.4
No GSH adduct	No GSH adduct	Medium	Very low	No GSH adduct	No GSH adduct	Medium	No GSH adduct	Medium	No GSH adduct	No GSH adduct	No GSH adduct

MMV688547	MMV688543	MMV688514	MMV688509	MMV688508	MMV688474	MMV688472	MMV688471	MMV688470	MMV688469	MMV688467	MMV688466
>80	>80	2.64	2.19	5.87	>80	10.9	>80	10.9	>80	11.9	5.3
>30	>30	>30	0.59	>30	>30	10	>30	>30	>30	>30	1.4
17.5	>20	29.7	1.29	>20	13.1	7.28	14.2	>20	>20	6.59	>20
>20	>20	>20	0.716	>20	>20	3.64	10.3	3.95	>20	>20	>20
>20	>20	>20	7.1	>20	>20	13.1	19.1	>20	>20	5.74	>20
>20	12.9	>20	0.134	>20	>20	1.3	4.65	15.3	>20	>20	2.86
>20	14.2	>20	8	67.2	>20	16.4	35.5	4.22		8.21	
>120	12.6	43.5	22.4	2.67	>120	10.9	5.06	42.6			
<1.5	23.3	4.2	14.6	95.3	<1.5	6.91	36.5	3.42			
	22.8	78.7	36.5	5.58	26.3	76.9	14.6	155			
	0.506	6.75	0.0075	0.707	20.2		0.6	0.524			
0.334	0.506	0.0903	0.0075	0.707	0.434		0.6	0.524		0.036	
	4.14	4.04		2.42		4.41	3.96		2.65	5.01	
	359	>960	>960	18.9		11.7	95	265	242	13.7	180
	0.99	<0.250	<0.25	13.2		22.6	2.94	0.97	1.06	17.5	1.64
	20.3	129	89.3	7.98		5.56	32.3	14.2	30.6	9.74	
	34.9	5.49	7.93	88.8		127	21.9	49.8	23.1	72.8	
0.221	0.0435	0.0472	0.122	2.24	0.362	18.1	18.1	8.31	10.9		
142	84.7	83.4	82	84	94.5	30	84.7	92.7	99.3	90.6	0
0.603	0.14	0.019	0.0007	0.19	0.3		0.038	0.064	0.147	<0.001	0
>500		73.4			>500					2.5	
No GSH adduct	No GSH adduct	No GSH adduct	Very Low	No GSH adduct	No GSH adduct	No GSH adduct	No GSH adduct	Medium	Very low	Medium	very low


MMV688756
MMV688755
MMV688754
MMV688704
MMV688703
MMV688557
MMV688555
MMV688554
MMV688553
MMV688552
MMV688550
MMV688548
>80 >80 19.4 7.01 2.65 16.4 45.1 53.2 >80 30.7 44.3 16.4
>30 >30 >30 17 0.66 >30 >30 >30 >30 >30 >30 >30
>20 >20 14.9 >20 >20 >20 >20 >20 >20 10.4 14.1 5.18
>20 >20 >20 9.33 6.58 12.1 >20 >20 >20 7.56 >20 >20
>20 >20 >20 >20 14.1 >20 >20 >20 11.9 >20 1.29 >20
>20 11.2 >20 >20 14.8 1.05 15.5 4.36 6.91 >20 >20 >20
91.8 >120 1.93 43.7 >20 >20 21.3
1.96 <1.5 93.3 4.11 7.21 8.45
22.8 58.8 1.07 0.974 24.9 2.62
23.4 9.04 496 546 3.26 203
0.694 0.659 0.428 0.707 0.715 163 0.6
0.694 0.659 0.428 0.707 0.715 0.714 0.6
2.4 3.31 4.72 3.24 3.44 2.93 2.65 4.22 3.8
173 388 11.5 268 141 91 520 28.6 29.9
1.39 0.75 21.1 0.917 1.72 2.64 0.471 8.39 8.06
137 221 8.28 50.1 32.2 7.14 31.3 11.7 11.4
5.19 3.21 85.6 14.1 22 99.3 22.6 60.6 62
5.84 6.03 1.11 0.0107 22.9 17.2 4.89 17.8 23.5
92.3 102 0 0 92 92.7 85.3 96 96.7
0.47 0.12 0 0.064 0.227 0.357 0.0317 0.0323
101 18.5
No GSH adduct Medium No GSH adduct No GSH adduct Medium Medium No GSH adduct Very Low Medium No GSH adduct No GSH adduct Medium

MMV688795	MMV688793	MMV688776	MMV688775	MMV688774	MMV688773	MMV688771	MMV688768	MMV688766	MMV688763	MMV688762	MMV688761
>80	>80	>80	36.1	2.81	>80	8	8.46	9.64	1.55	1	6.94
>30	>30	>30	>30	>30	>30	>30	>30	>30	>30	3.1	30
8.71	>20	>20	16.5	23.4	3.42	32.8	17.5	1.02	27.8	90.9	38.9
>20	>20	>20	19.7	>20	>20	1.32	4.16	5.57		14.6	>20
>20	>20	>20	>20	>20	>20	0.836	>20	>20		>20	18.7
>20	>20	>20	>20	>20	>20	>20	>20	4.52		>20	>20
11.8	70.8	23.1	>20	<0.082	>20	0.772	9.18	18.5		>20	>20
15.2	2.54	7.78	65.3		>120	2.97		9.69	20.6	41.8	31.6
2.41	22.4	7.48	2.75		<1.5	60.5		1.17	8.7	4.3	5.68
220	23.7	71	68.3	50.6	>120	3.5		453	1.43	22.6	4.95
0.514	0.77	1	7.79	10.5	<23.1	152		0.0069	371	23.5	107
0.514	0.77	1	0.581	0.105	1	0.447		0.0069		0.706	
3.58	2.53	4.13	3.98	4.48	0.735	4.59		5.24	3	3.71	4.07
28.2	247	30	>960		>960	6.8		120	3.72	132	4.33
10.1	1.04	8.14	<0.25		<0.25	35.2		2.14	64.4	1.81	55.3
1.78	51.1	11.6	>960		>960	6.29		81.3	4.07	23.7	3.68
398	13.9	61.2	<0.738		<0.7380	113		8.71	174	29.9	192
22.6	19.2				3.84	17.5		0.211	16.8	2.51	10.3
84	90.7	98.3	72.5	114	98.7	101		106	0	11.7	0
0.0357	0.147	0.0082	0.0305	0.012	0.99	0.0044		<0.0015			
			>500	<0.5	>500	<0.5	95.5		93.7	46	<0.5
Medium	Very Low	Very Low	No GSH adduct	No GSH adduct	No GSH adduct	Medium	No GSH adduct	Medium	No GSH adduct	No GSH adduct	No GSH adduct

MMV688891	MMV688889	MMV688888	MMV688854	MMV688853	MMV688852	MMV688846	MMV688845	MMV688844	MMV688798	MMV688797	MMV688796
>80	3.63	>80	22.6	>80	26.1	5.33	27.6	>80	33.6	>80	>80
>30	4.1	0.63	1.1	>30	21	3.3	>30		>30	>30	>30
>20	3.63	5.8	92.7	6.08	8.05	>20	>20	>20	>20	15.8	>20
>20	8.94	4.34	>20	>20	13.1	>20	11.8	>20	>20	>20	>20
>20	>20	>20	>20	>20	>20	0.685	16.7	>20	>20	>20	>20
>20	>20	0.276	>20	>20	>20	>20	2.05	3.28	14.5	>20	>20
		17.7	>20	>20	15		1.39	12.6	6.18	11.5	81
		10.1	79	>120	11.9		129	14.2	29.1	15.6	2.22
		7.36	2.28	<1.5	1.15		0.668	2.85	1.8	0.955	11.4
		72.2	62.2	>120	463		795	186	295	556	46.5
	0.218	0.175	8.54	<4.43	0.187			0.0904	0.587	0.623	0.647
	0.218	0.175	0.262	0.912	0.187			0.0904	0.587	0.623	0.647
5.23	4.39	4.49	4.17	3.46	4.55		3.91	4.59	3.34	3.88	3.09
444.5		53.8	458	769	33.5		256		336	40	>960
0.628		4.8	0.523	0.311	7.3		1.21		0.72	6.1	<0.25
194		38.6	260	222	4.44		3.93		11.5	7.25	31.1
3.65		18.3	2.72	3.19	159		180		61.6	97.7	22.8
0.572	11.5	3.17	0.407	17.4	5.32		14.4	0.185	16.8	23.1	8.84
4.3	92.3	91.7	95.2	112	96.3	90	60	3	91.5	88.7	95.7
	0.0123	0.022	0.0973	0.113	0.0099	0.0065	0.024		0.0565	0.0487	0.16
			>500	221							
No GSH adduct	Medium	No GSH adduct	No GSH adduct	Very Low	Very Low	No GSH adduct	Medium	Low	Very Low	No GSH adduct	Very low

MMV688980	MMV688978	MMV688958	MMV688955	MMV688943	MMV688942	MMV688941	MMV688939	MMV688938	MMV688936	MMV688934	MMV688921
17	1.74	14.4	22.2	3.16	21	4.22	6.83	5	18.9	0.00547	
>30	>30	>30	>30	13	>30	3.7	3.6	1.1	1.5	>30	>30
>20	31.4	0.274	3.43	3.7	>20	>20	0.457	1.01	0.0216	>20	>20
14.3	>20	>20	5.12	2.34	7.14	>20	11.7	>20	0.449	>20	>20
>20	>20	>20	>20	>20	>20	0.639	>20	0.462	5.75	>20	
>20	>20	>20	>20	0.0379	3.44	4.95	11.5	0.873	<0.01	5.22	>20
43.2	0.288		16.4	9.21		7.37	8.19	8.49	17.2	5.05	>120
4.15			11	19.5		24.4	21.9	21.2	10.4	35.6	<1.5
84.6			4.52	7.38		0.984	1.43	3.05	5.46	11	8.82
6.29			117	72		540	371	174	97.4	48.2	60.3
0.781				0.221		0.157	0.418	0.0087		0.019	0.0022
0.781				0.221		0.157	0.418	0.0087		0.019	0.0022
3.29			3.8	4.59		4.52	4.11	4.7	4.41	5.09	4.98
285	44.2		22.5	72.8		61.4	86.5	23.4	8.57	70.4	103.5
0.934	5.42		10.7	8.35		4	4.68	11.2	28.2	3.54	2.38
23.5	31.3		34.5	24.2		9.31	5.57	26.6	21.2	80	152
30.1	22.7		20.5	29.3		76.1	127	26.6	33.4	8.85	4.67
27.5			8.91	6.97		2.29		0.991	0.516	0.218	0.0883
36		101	54.3	94.7		93	82	86	15	97.7	90.3
		0.0125	0.0052	0.0088		0.0163	0.027	0.004		0.0009	0.0004
	14.6										
Medium	No GSH adduct	No GSH adduct	No GSH adduct	No GSH adduct	No GSH adduct	Medium	Medium	Medium	No GSH adduct	No GSH adduct	Medium


MMV689437
MMV689255
MMV689244
MMV689243
MMV689061
MMV689060
MMV689029
MMV689028
MMV689000
MMV688994
MMV688991
MMV688990
2.56 >80 19 25.8 39.1 12.3 20.8 7.6 26.4 >80 6.35 >80
3.7 >30 8.2 2.5 18 20 20 8.9 >30 >30 1.7 >30
>20 9.4 78.3 57.2 >20 >20 >20 >20 14.3 11.9 48.7 6.28
14 >20 >20 >20 >20 >20 2.66 0.583 >20 >20 5.06 >20
>20 >20 7.38 12.7 >20 2.11 >20 >20 >20 >20 1.47 >20
>20 >20 14.3 >20 18.3 >20 >20 >20 >20 >20 >20 >20
47.2 >20 8.25 >20 13.5 4.2 12.1 >20 >20 2.55 >20
3.8 21.8 13.4 42.8 14.8 38.1
23.4 8.26 3.3 2.3 2.68 4.71
22.7 15.2 161 230 198 115
0.0981 35 0.478 0.636 0.524 4.62
0.0981 0.0036 0.478 0.636 0.524
4.31 5.14 5.08 3.54 3.27 3.45 5.82
451 278 >960 44.5 106 347 182
0.56 0.861 <0.25 5.4 2.61 0.74 2
29.3 305 364 3.16 59.1 14 11.6
24.2 2.32 1.95 224 12 50.5 61.1
0.0958 1.87 0.0158 1.8 1.56
94 100 90.7 0 97.7 0 94.7 119 0
0.0084 0.0017 <0.0001 0 0.177 0 0.0553
>500 49.9 205 37 NA >500 1.7
No GSH adduct No GSH adduct Very low No GSH adduct No GSH adduct No GSH adduct Very Low Very Low No GSH adduct No GSH adduct No GSH adduct No GSH adduct

MMV690103	MMV690102	MMV690028	MMV690027	MMV689758	MMV689709	MMV689480
2.33	2.87	8.71	7.76	7.53	23.2	3.79
8.3	2.2	12	>30	>30	6	>30
>20	>20	7.47	2.32	0.513	>20	11.6
>20	>20	3.32	5.93	>20	>20	>20
>20	>20	>20	>20	>20	8.61	>20
>20	6.37	5.09	0.596	7.81	>20	>20
11.8	46.7	12.8	8.57	22.9	>120	>20
15.2	3.85	14	21	7.83	<1.5	
1.03	22	6	17.6	40.6	7.31	
518	24.1	88.5	30.2	13.1	72.7	
0.0597	0.0248	0.099	0.0387			
0.0597	0.0248	0.099	0.0387			
4.52	4.39	4.34	4.59		2.3	5.27
16.7	304	111	81.4	241	334	
15.4	0.81	3.49	4.48	1.07	0.83	
3.7	61.2	93	61.3	149	42.2	
192	11.6	7.62	11.6	4.76	16.8	
0.0309	0.0386	0.345	0.432	0.124	0.0396	
90.7	90.3	96.7	88.3	0	104	
0.0062	0.0041	0.0067		0	0.443	
						21.7
Medium	Very Low	No GSH adduct	No GSH adduct	No GSH adduct	No GSH adduct	No GSH adduct