Structure-based Drug Design for Tuberculosis

Yuchen Wei[1]

June 2015

Abstract

This paper investigates structure-based drug design for developing novel therapeutic compounds against tuberculosis. The whole drug discovery process, based on ligand-protein interaction, is elucidated both theoretically and practically in this paper. *In silico* design, synthesis and evaluations are widely used as the standard method to identify potential inhibitors against tuberculosis. Different targets are chosen here to help get a better understanding of the design principle, by blocking a particular pathway in *Mycobacterium Tuberculosis*.

Contents

1 Introduction 2
2 Structure-based drug design 3
2.1 Molecular Modeling 4
2.1.1 Quantitative structure-activity relationship 4
2.1.2 Docking and scoring methodology 5
2.2 Assembly of compounds from fragments 6
2.3 Some criteria for potential drugs by SBDD 6
3 Recent Achievements 6
3.1 *In silico* SBDD of anti-TB drugs 6
3.1.1 *In silico* SBBD with a target's structure known 7
3.1.2 *In silico* SBDD with creation of a homology model 8
3.2 Synthesis and evaluation 11
4 Conclusions and discussions 14
5 Acknowledgement 14
Reference 14

[1] Nanoscience (Top Master Programme), Zernike Institute for Advanced Materials, University of Groningen. Email: Y.Wei.2@student.rug.nl. Supervisor: Dr. M. D. Witte.

1 Introduction

Tuberculosis (TB) is widely known as a fatal disease which may date back to thousands of years ago. The earliest evidence of TB's existence by now was discovered on the spine fossils of Neolithic times. Hippocrates of Cos (460-370 BC), an ancient Greek physician, recorded pulmonary TB and pointed out that TB was infectious [1]. In Roman times, Galenus (134-201 AD) put TB in specific recordation, including its symptoms and some simple treatments like placing patients at high altitude, giving them good rest and enough nutrient. Later, people gradually found out living with TB patients could also cause infections and some people even started to think that TB was caused by some invisible creatures. In 1865, a French scientist named Villemin experimented on rabbits by injecting patients' phlegm into their ear veins, and finally discovered tubercles in the rabbits' chests. Thus, TB's contagiousness was formally demonstrated. And in 1882, Koch used the acid-fast staining method to successfully isolate mycobacteria of TB, which won him the Nobel Prize in Physiology or Medicine in 1905 [2].

Mycobacterium Tuberculosis (*M.tb*) discovered by Koch is the main cause of TB. This mycobacterium has high lipid in its cell wall and it can protect them from the immune system of its host. Moreover, *M.tb* can even survive in weak disinfectants for weeks while maintaining its endospore (dormant) form. Besides being an effective protection, the cell wall can also selectively enable compounds in and out for the survival of the bacterium [3]. The cell wall of *M.tb* consists of mycolic acids, peptidoglycans, arabinogalactan, and lipoarabinomannan and so on. Due to the importance of *M.tb*'s cell wall, nowadays a lot of inhibitors are focused on targets participating in the generation of *M.tb*'s cell wall, i.e., glutamine synthetase. Other enzymes, for example, ATP synthase and DNA gyrase can also be promising targets. There are many *M.tb* strains, among which the most commonly known and virulent one is H37Rv strain. The genome sequencing of H37Rv strain was completed in 1998 by British and French scientists. Such an accomplishment, together with the genome sequencing of CDC1551 in 2001, offered a great chance for deep research into pathogenic genes of *M.tb.*

In the middle of 1900s, a series of effective medical therapies were developed. Streptomycin (SM) was the first effective chemotherapy discovered in 1944. It is a bactericidal antibiotic derived from actinobacterium Streptomyces griseus. SM inhibits the protein synthesis in *M.tb* because it can bind to 16S rRNA of the 30S subunit of *M.tb* ribosome, thus it interferes with the binding of formyl-methionyl-tRNA. The resulting codon-misreading and failure of protein synthesis eventually cause the death of the microbial cells [4]. SM is one of the five first-line drugs and the other four are isoniazid (INT, discovered in 1952)， Pyrazinamide (PZA, discovered in 1952), ethambutol (EMB, discovered in 1961) and rifampin (RF, discovered in 1966) [5]. The second-line drugs are more expensive and have more side effects, i.e. 4-aminosalicylic acid (PAS, discovered in 1946). Nonetheless, the control of TB received little accomplishment in the beginning, especially in some developing countries. To obtain effective control over TB, World Health Organization (WHO) proposed a standard chemotherapy treatment in collaboration with governments and foundations in 1970s.

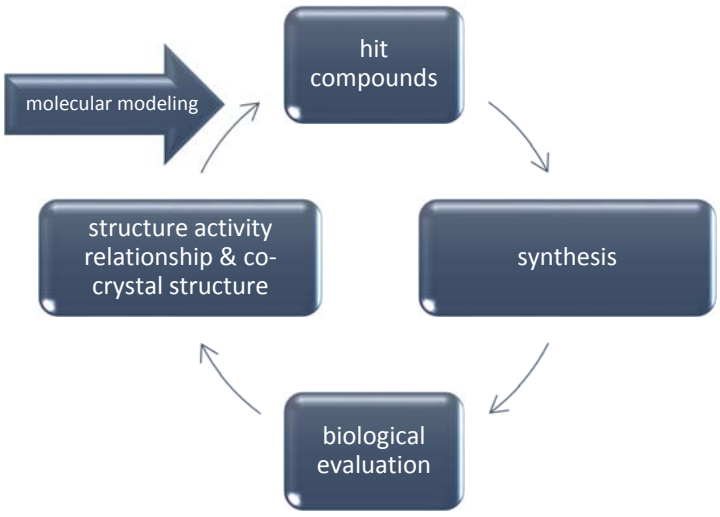
Nowadays, there are still a lot of patients suffering from TB disease. According to WHO's data, there were estimated 8.8 million of new cases happened in 2012 [6]. Although the mortality rate

dropped by approximately 45% since 1990, the rise of multiple drug-resistant and extensively-drug-resistant cases urged new effective therapies to be designed. In the following section, we will discuss on a new method of obtaining innovative and potential drugs for TB, as well as investigate some of new achievements in this area.

2 Structure-based drug design

Structure-based drug design (SBDD) involves the interaction between targets and inhibitors. A target is a molecule, usually an enzyme responsible for the functioning of a particular pathway (metabolic or signaling) of the pathogen. An inhibitor or a ligand is a molecule whose shape is complementary to the target. By binding to the active site of the target, the ligand therefore inhibits the target's function and causes a therapeutic effect to the patient. The binding forces are usually non-covalent, such as electrostatic interaction, hydrogen bonding, π-π stacking force and hydrophobic interaction.

SBDD exploits the three-dimensional structure of the target (or its homology) to develop ligands by computer modeling. The commonly used process is in a cyclic fashion as illustrated in Scheme 1. The 3D structure of the target can be obtained from databank, X-ray crystallography or NMR. With the target's structure and its active site in hand, scientists can visualize and model several potential ligands. Roughly, there are two categories to obtain ligands: the ligand-based design method and the receptor-based design method. The former one is referred as a database-searching method; the latter one is to construct new molecules to fit into the binding pocket. Receptor-based design is extraordinarily useful in novel drug design and will be mainly discussed in this paper.



Scheme 1. The basic procedure of SBDD. It starts form the design of the hit compound after modeling and run this circle to optimize the drug.

However, both ligand-based and receptor-based strategies have their defects. Searching database is usually time-consuming and there can be a large number of results; receptor-based strategy can also give thousands of ligands due to multiple active side residues in the binding pocket. To save time and money, further screening need to be conducted to reduce a library of ligands into a most promising hit. Molecular modeling has been used as the tool and it adopts computer programs to perform the screening [7-10]. After *in silico* design, expanded synthesis, biological evaluation and X-ray analysis can complete the whole design circle and also rationalize the whole design process.

2.1 Molecular Modeling

There are two different methods of molecular modeling. One is to build a regression model of ligands, namely quantitative structure-activity relationship (QSAR) modeling; the other one is to create a 3D model of the complexed ligand and the target, referred as molecular docking, which stresses their interaction. Both of these can be used separately or together.

2.1.1 Quantitative structure-activity relationship

QSAR modeling is a mathematic method aiming to establish a relationship between a ligand's structure and its biological activity. This relationship can be expressed with a function A=f(S), where S is responsible for a ligand's chemical and physical properties (melting point, geometry, hydrophobicity, electron density, HOMO-LUMO gap, etc.), and A is responsible for its biological activity (positive effectiveness or negative toxicity). Starting with a number of existing ligands, a regression model (the function) will be created, which will be used later to predict biological activities of potential ligands.

QSAR study has been widely used in SBDD recently. Traditional QSAR is 2D-QSAR, which means it doesn't use ligands' 3D structures. And a more precise method is 3D-QSAR, however, it is more time-consuming. Actually there is a methodology between 2D-QSAR and 3D-QSAR, and it is Hologram QSAR (HQSAR). HQSAR first cuts the existing ligands into fragments by a particular 2D hologram length. Next, it generates a hologram using these fragments by a series of calculation. HQSAR is a fast predicting approach and one paper in 2013 [11] could provide a better insight into it. This work chose M.tb thymidine monophosphate kinase (TMPKmt) as the target, because TMPKmt was essential for M.tb's DNA replication. 97 known inhibitors based on thymidine were used to develop a HQSAR model. Then the model was utilized to predict novel ligands' activities, which was further verified by quantum mechanics/molecular mechanics (QM/MM) docking.

Other QSARs, for example, multi-target methodologies focused on quantitative-structure activity relationships (mt-QSAR) and multitasking model based on quantitative-structure biological effect relationships (mtk-QSBER) are rising. Mt-QSAR predicts biological activities or toxicities separately against different targets, whereas, the extended mtk-QSBER method predicts the

activities as well as toxicity at the same time [12].

2.1.2 Docking and scoring methodology

Docking and scoring is a computer-based method to predict the interaction energy between a ligand and a target. It is a process reflecting the “lock and key” relationship. A target is a “lock” while its ligand is a “key”. The best fit “key” will get the lowest value of the interaction energy. However, the “key” is flexible in the binding pocket. Therefore, the interaction energy will be minimized by optimizing the conformation and orientation of/between the protein and the ligand.

The scoring is rather complex and there are several developed algorithms to perform it. An early algorithm, called rigid fitting, is a method losing a ligand’s flexibility. New algorithms, i.e., Monte Carlo algorithm, LUDI algorithm and genetic algorithm, are more accurate and will be discussed in this paper in sections 3.1.1 and 3.1.2. General Monte Carlo algorithm is a random sampling method. Random sampling is rather like gambling. It aims to obtaining a large number of results, of which a small probability can be even incorrect. In SBDD, Monte Carlo algorithm first fills the binding cavity with atoms of random types. Each atom has three properties, including its atom type, the hybridization and the number of implicit hydrogen atoms. And each atom can have 12 bonds at maximum. The interaction energy will be evaluated using an energy function, which consists of both intramolecular and intermolecular components, such as its geometry, shape and chemical affinity. By comparing the binding energies, computer can determine the atoms’ existence, types, bonding and so on.

LUDI algorithm was designed by a Germany company and used only in SBDD. It depends on the molecular-fragment relationship to find potential ligands. At first all the fragments will be identified via database searching. All fragments are balanced with four kinds of interactions at the binding site: lipophilic-aliphatic interaction, lipophilic-aromatic interaction, hydrogen donor, and hydrogen acceptor. All the hydrogen bonds are made and all the hydrophobic pockets are filled with hydrophobic fragments. The fragments are then optimized with proper 3D arrangement and ultimately be connected with standard linkers. After docking the ligand to the target where the arrangement and descriptors are maximally aligned, the interaction energy will be obtained by a build-in energy function.

Genetic algorithm (GA) is a biomimetic algorithm using the concept of natural selection. It starts with a population of candidate solutions and produces a new generation of solutions via the operations of selection, crossover and mutation. By repeating these genetic operations, an optimized population of solutions can be reached ultimately. In SBDD, the starting population is a set of state variables describing ligands’ translation, orientation, and conformation. Each of the variables corresponds to the genotype while the ligands’ atomic coordinates correspond to its phenotype. By crossover, random paired state variables can produce new individuals inheriting genes from either parent. Meanwhile some variables will mutate by random amount. Selections of the offspring will be conducted using energy function. By running repeatedly, it produces better solutions and finally reaches the approximate solutions. Recent GA has more complicated operators besides crossover and mutation, and they make the docking evaluation more precise.

2.2 Assembly of compounds from fragments

To construct novel ligands, scientists usually use receptor-based strategy by assembling fragments into a molecule. With the 3D structure of the target, computational programs will be run to fill the binding sites with suitable chemical groups. The algorithms mentioned in section 2.1.2 can be utilized in the selection of fragments. Then those groups will be connected by linkers to form a whole compound.

2.3 Some criteria for potential drugs by SBDD

Generally, there are some empirical criteria that can be applied in most situations in drug design. Commonly, the substituent groups of potent ligands must have structural diversity and all the design must be synthesis available. Moreover, there is a rule named Lipinski's rule of five applied to determine if a chemical compound is orally active. This rule states that an orally active drug cannot violate more than one of the following four criteria: firstly there should be no more than 5 H-bonding donors; secondly no more than 10 H-bonding acceptors are present in the molecule; the molecular weight should be less than 500; the octanol-water partition coefficient, log P, should be less than 5.

3 Recent Achievements

The completion of *M.tb*'s genome sequencing was a milestone of novel drugs discovery for TB. It helped scientists to look into the nano-scale world to understand how *M.tb* became pathogenic. As an important tool, SBDD works at the molecular level to develop novel drugs for TB. In recent years, many promising inhibitors against TB based on SBBD are published and some of the drugs are now under clinical experiments.

3.1 *In silico* SBDD of anti-TB drugs

SBDD is getting more and more popular in drug design. It is timesaving and easy to operate. Nowadays, the rapid growth of databank and computer programs have also facilitated the computer-based process quite a lot.

3.1.1 *In silico* SBBD with a target's structure known

Shikimate kinase (SK) of *M.tb* is an indispensable protein for the viability of the pathogen. It catalyzes the generation of chorismate in the fifth step, where it transfers a phosphoryl group from ATP to shikimate [13]. Meanwhile, it is absent in humans. These two factors make SK a perfect target. SK has four domains. While binding to shikimate, ATP or ADP, a conformational change of SK may take place, leaving SK's LID domain closed. However, whether a potential ligand is still able to cause the closure of LID domain is unknown. Hence, the ligand design will need the 3D structures of SK both in its open and closed LID conformations. The co-crystal structure of SK complexed with shikimate and ADP in the closed LID conformation, together with the structure of SK complexed with shikimate in its open LID conformation can be accessed in Protein Data Bank (PDB).

In 2010, one paper used SK as the target and *in silico* developed a promising dipeptide inhibitor [14]. At first, the receptor model was prepared by removing ligand shikimate from the co-crystal structure along with correction and optimization. Then, its binding site was identified, which was a hydrophobic pocket with several hydrophilic residues sticking out. The binding pocket had two phenyl rings oriented in a parallel manner, indicating its probability for π-π stacking. It also had several charged residues allowing for electrostatic interaction. Other factors were also taken into consideration: the binding site on the receptor, structural similarity with the original ligand, and the criteria mentioned in section 2.1.3.

A set of dipeptides were designed and docked into the receptor models. The best docking pose of each inhibitor was calculated by Monte Carlo-combined algorithm. Then LUDI 3 algorithm was employed to score the interaction energy. In the process, lipophilic-aliphatic interaction, lipophilic-aromatic interaction, hydrogen donor and hydrogen acceptor interaction, ionic interaction, the number of rotatable bonds in ligand and loss of entropy were the factors to be concerned to score the pose-refined ligand-protein complex. In their work, the docking process was validated by docking the substrate shikimate into the prepared SK models, of both the closed and open LID conformations, and then compared the docked complex with the already known co-crystal structures. Root mean square deviation describing the overlap of those two complexes could describe its validation. In the closed LID conformation, the result showed a value of 0.3 angstrom less than 2 angstrom (the standard for a good docking function). Also the binding affinity (K_d) of the complex was 12μM, which was in the same order of the experimental value 44μM. In the open LID conformation, it also gave reasonable results hence proved its validation.

The docking results revealed an Arg-Asp type dipeptide RD as the best candidate. RD possessed great affinity to SK which was even greater than the best inhibitor reported then. The docked complexes of SK with RD both in its closed conformation and open conformations are shown in Figure 1. The docking model showed generally RD has more hydrogen bonds than the substrate shikimate. Moreover, the guanidium group on Arg [D2 (P)] abstracted one extra proton and formed a salt thus increased its electrostatic interaction with the active residue. Meanwhile, this guanidium group simultaneously formed a bidentate hydrogen bond with the target. In Figure 1a, even though the closure of LID domain caused some affinity loss as a result of the loss of hydrogen bonds and also it caused some change in the docking pose, however, RD's affinity was still high in closed LID conformation.

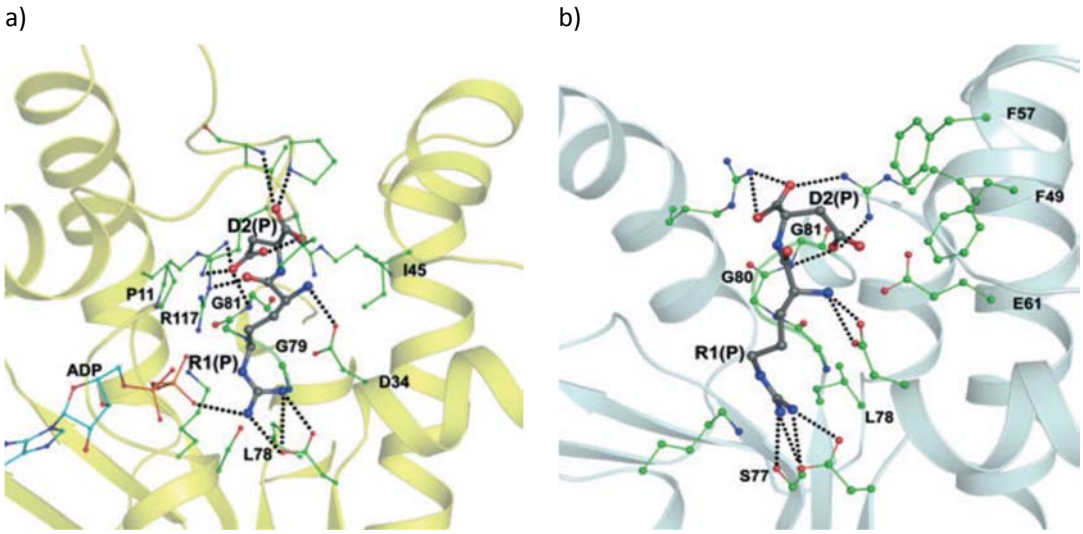


Figure 1. a) The docked complex of SK (yellow ribs, side residues are in green) in closed LID conformation with RD (ball-and-stick, different colors rendering different atoms) together with ADP (shown in cyan color, ball-and-stick); b) The complex of SK (grey ribs) in open LID conformation with RD.

To obtain an extensive analysis of the whole system, molecular dynamics simulation was conducted to measure the effect of the solvent (TIP3 water) on the docked complex. It allowed the complex interact with water for 150 picoseconds and then examined its stability. The results showed that most of the hydrogen bonds remained stable and its total interaction energy had only little variation. Therefore, the binding between RD and SK was anyway strong. To conclude, this paper exploited typical *in silico* SBDD procedures: a receptor model(s) preparation, ligands design by assembly of fragments, and molecular docking. It introduced a novel potential hit RD, a small dipeptide without any aromatic rings meaning its less toxicity and more orally activity. Their earlier work in 2009 also focused on small peptides as inhibitors, whereas it used another target Dihydrofolate reductase (DHFR) [15] (DHFR catalyzes the synthesis of tetrahydrofolate which is vital to the synthesis of DNA, RNA and protein of *M.tb*). The whole *in silico* design process was similar, except for one difference that the 2009 paper also docked the potential ligand—a tripeptide to DHFR in human body to make a comparison thus proved the inhibitor is more selective to DHFR of *M.tb.*

3.1.2 *In silico* SBDD with creation of a homology model

If the 3D structure of a target were not accessible from PDB, what should we do? To avoid time-consuming X-ray crystallography or NMR analysis, one may expect to create a homology model of the target. It is already known that a protein's structure usually resemble another if their sequences are very similar [16]. Hence an available structure of a similar protein can be the solution to do *in silico* drug design.

It is reported that DNA gyrase is the only type II topoisomerase in *M.tb*, however, it doesn't exist in eukaryotes [17]. So it is natural to treat DNA gyrase as the target to block the pathway of *M.tb*'s DNA winding, which further halts the DNA replication. A paper in 2010 reported that fluoroquinolone derivatives could be potential inhibitors against DNA gyrase in *M.tb* [18].

Quinolones were reported before as inhibitors for DNA gyrase in Escherichia coli by forming stable ternary complexes with DNA gyrase together with DNA, thus block DNA replication [19]. In addition, the binding site on DNA gyrase for fluoroquinolone was its N-terminal on A-subunit [20]. As a result, the focus of the drug design should be on A-subunit of the gyrase. However, whether quinolone could work on DNA gyrase of *M.tb* remained unknown. Moreover, the structure of *M.tb*'s DNA gyrase was inaccessible. To exam the bioactivity of quinolone, the paper in 2010 introduced DNA gyrases in *Mycobacterium fortuitum* and *Mycobacterium smegmatis* as the templates to produce a homology model. These two mycobacteria were usually used as comparative standards for the measurements of organism in *M.tb*. And their DNA gyrase structures could be accessed in PDB. In addition, A-subunit of DNA gyrase in *M.tb* exhibits more than 92% sequence similarity with that in *Mycobacterium smegmatis*.

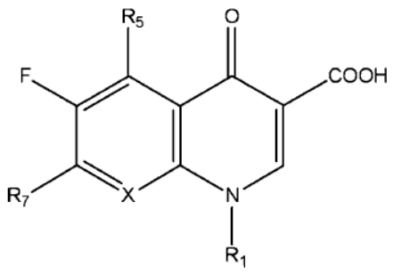


Figure 2. The template structure of fluoroquinolone which has four substituent sites R_1, R_5, R_7 and X.

The first step was the construction of ligands using fluoroquinolone backbone. As depicted in Figure 2, fluoroquinolone derivatives have four substituent sites. A library of ligands were generated by assembling different groups onto these four sites following the criteria in section 2.1.3. Then QSAR study was introduced to measure their activities. A QSAR model was developed based on data-searched ligands, whose ligand-protein bioactivities were already given against organisms in *Mycobacterium fortuitum* and *Mycobacterium smegmatis*. This process used genetic algorithm mentioned in section 2.1.2. All the ligands against *Mycobacterium fortuitum* and *Mycobacterium smegmatis* were separated into two training and test sets. The training set was used to set up the QSAR model while the test set was to prove the model's validation. The QSAR model was assessed[2] by calculation of the internal cross-validated number Q^2 and external predictive number R^2. Q^2 and R^2 of this QSAR model were in the range of good performance. Thus, it was used to fast screen the ligand library and several fluoroquinolone analogues stood out with high inhibitory activity.

[2] Q^2 is calculated by $Q^2 = 1 - \frac{\sum(Y_{pred} - Y)^2}{\sum(Y - \overline{Y})}$, Y_{pred} and Y are respectively predicted activity and observed activity while $\overline{Y}$ is the mean activity. $R^2 = 1 - \frac{\sum(Y_{pred\,(test)} - Y_{(test)})^2}{\sum(Y_{(test)} - \overline{Y}_{traing})^2}$ It is similar in the equation of Q^2. Both values should be more than 0.5 to indicate a good QSAR model.

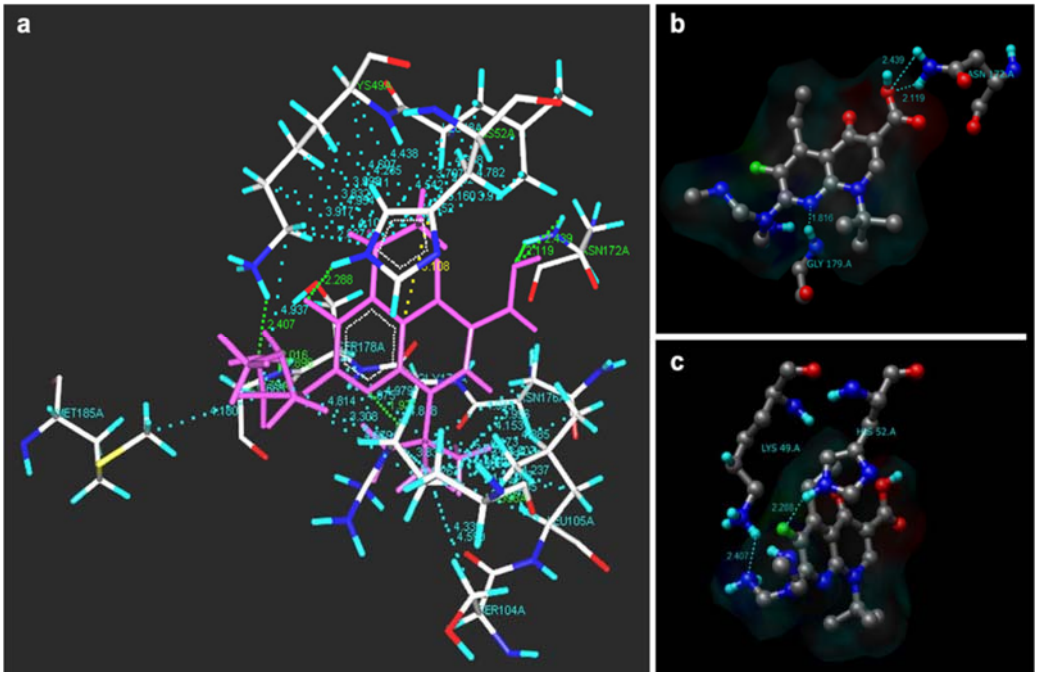


Figure 3. a) The docked complex of DNA gyrase A with the potential inhibitor (purple colored) with hydrophobic interaction (cyan dotted line), hydrogen bonding (green dotted line) and π-π stacking (yellow dotted line); b) H bonding between the inhibitor and Asn172(A) and Gly179(A); c) H bond between the inhibitor and Lys49(A) and His52(A) from another angle.

A second stage of screening was conducted by molecular docking using homology modeling. Here a homology model was developed based on DNA gyrase (A-subunit) of *Mycobacterium smegmatis* because of the sequence similarity mentioned before. Then GRIP docking algorithm (it uses Piecewise Linear Pairwise potential (PLP)-based scoring function, considering hydrogen bonding, repulsion and dispersion) was employed to score the interaction energy of all paired target and ligands. One inhibitor stood out as the most potential and the structure of ligand-protein complex is depicted in Figure 3. Hydrophobic interaction, π-π stacking force, and hydrogen bonding could be observed in the docked complex. And from the picture, we could also figure out the art of using fluoroquinolone instead of quinolone, for that the fluorine atom could form an extra hydrogen bonding with positively charged amino acid.

In this work, a homology model was developed. Nonetheless, the 3D structure of *M.tb*'s DNA gyrase was updated in PDB in the end of their work. As a result, they compared the homology docking results with real model, and found some similar numbers. This demonstrated the validation of homology modeling to some extent. One recent paper in 2014 also used homology modeling to find potent ligands for PknI by database searching [21].

PknI is one of the eleven Serine Threonine Protein Kinases (STPKs) in *M.tb* and it is very important for mycobacterial intracellular survival [22]. PknB was chosen as the template due to its 32% sequence identity with PknI. According to the sequence alignment of PknB, they modeled the residues of the homology model. Structure-based pharmacophore analysis was conducted using 21 co-crystal structures of kinase proteins bound with ATP to identify the critical residues on the target. The interaction between ATP and these 21 kinase proteins were specified, thus a comparative active site residues could be figured out on the PknI model. It was implied that Lys 41, Asp 96, Val92 and Asp 90 were the critical residues, among which Lys 41 was highly important

proved by site directed mutagenesis of Lys 41. After the identification of critical residues, a library of 2400 drugs against other kinase were docked into the ATP binding site of the homology model. The docking process used genetic algorithm. 6 hits were selected out and all of them form hydrogen bonds with Lys41.

More specified details of homology creation were illustrated in this work, although it only used ligand-based strategy by searching database instead of designing novel ligands. These two examples of homology modeling explained that this method could function well. In addition, a potential ligand developed by homology modeling can be used in X-ray crystallography, further solving the target' 3D structure.

3.2 Synthesis and evaluation

After *in silico* design, a hit compound will be selected out. To transfer a hit into a lead, synthesis must be employed to have the compound at hand, followed by a series of evaluation to test its biological activity practically. X-ray co-crystal structure of the ligand-protein complex will further verify the validation of the hit. The results can determine whether to move on with *in vivo* tests or go back to running another circle of design procedures to modify the hit.

One paper in 2012 started with the synthesis of a potential hit and made it feasible after evaluations [23]. It chose glutamine synthetase (GS) of *M.tb* as the target, and developed an imidazo [1, 2-a] pyridine inhibitor. Generally, GS catalyzes in the pathway of glutamine synthesis by binding with ATP. It usually exists in bacterial cytoplasm. However, in *M.tb* and other pathogenic mycobacteria, GS can be released extracellularly. What is more interesting, a kind of glutamine named poly-L-glutamate /glutamine component is quite unique in the cell wall of those pathogenic mycobacteria. Therefore, scientists correlated GS with the poly-L-glutamate /glutamine component, furthermore, with the virulence of the pathogenic mycobacteria [24]. An earlier reported potent inhibitor—MSO binds at the site where amino acid binds. Nevertheless, the site where MSO binds also exists in organisms of human bodies. Thus, it is better to find another binding site, for example, the ATP binding site on GS.

According to their previous work in 2009 [25], a potential ligand for the ATP-binding site on GS was 3-amino-imidazo [1, 2-a] pyridine analogues which had a halogen (Cl, Br or I) substituent on its sixth position, an aryl substituent with hydrogen bond donor at its *meta*- position, and also a cyclopentyl group at its R^1 position as drawn in Figure 4, compound **3**. In this work in 2012, they modified this imidazo pyridine ligand with expanded synthesis. Because there wasn't enough reference of ligands to the ATP-binding site on GS and also due to the multiple substituents on the ligand's phenyl ring, a designing strategy named focused hierarchical design of experiments (FHDoE) was utilized. FHDoE started with the synthesis of a library of compounds **3**. It chose R^1, R^2 (R^o, R^m and R^p) groups having different sizes and lipophilicity according to Principal Component Analysis (PCA)[3]. 16 compounds were synthesized by one-pot Ugi-type reaction as shown in Scheme 2 and then tested by the inhibition evaluation against GS. Compound **3g** (also shown in Figure 4) showed

[3] PCA uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of linearly uncorrelated variables.

the lowest IC_{50} (~3.0μM) value and was chosen as template for a second round of substituent modification and inhibition evaluation. R^2 was retained, where as R^1 group was varied to shorter linear chain, longer linear chain, smaller branched aliphatic group, aromatic group, or larger cycloalkane. As a result, another 5 compounds were synthesized and evaluated by their bioactive inhibitions. The result showed that **3g** still holds the lowest IC_{50} value and made butyl group the optimal R^1 substituent.

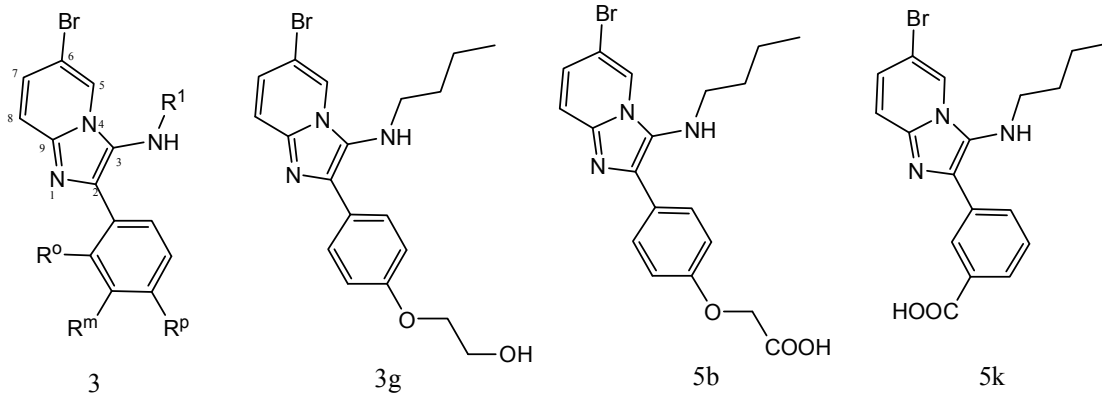
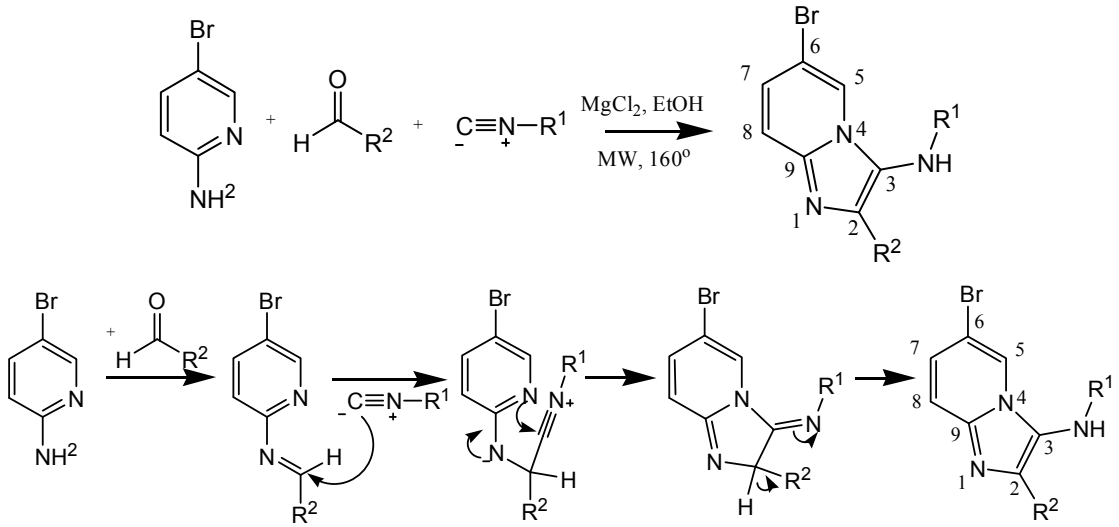


Figure 4. The chemical structure of compound imidazo [1, 2-a] pyridine derivatives.



Scheme 2. The preparation of imidazo [1, 2- a] pyridine derivatives using Ugi-type reaction and its plausible mechanism.

A third round of modification was conducted by comparing 12 compounds having different R^2 groups while keeping the same butyl R^1 group. By comparing a tertiary amine (IC_{50}=5.8μM) with carboxylic acidic (IC_{50}=1.6μM) substituent at the R^p position, it showed that there was no specific electrostatic interaction with GS. A branched substituent at R^p position had an IC_{50} value of 22μM, which was still relatively low, indicating that there was considerable space at the binding site. Moreover, the hydrogen bond donor at R^p position was not significant, because the IC_{50} values had little variation either by replacing it to R^m position or simply removing it. Therefore, compound **5k** displayed the highest inhibition. To conclude FHDoE strategy, it could be described as the following: first selected out the best inhibitor as the template from a roughly designed library, and modified the fragments from one to another.

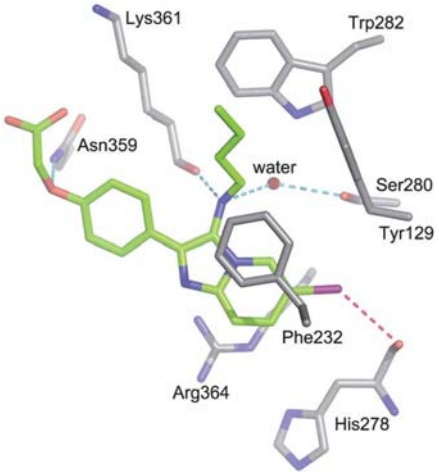


Figure 5. The complex formed by compound 5b (green carbon atoms) and GS of *M.tb* (grey carbon atoms).

Even though the aforementioned synthesis and inhibition evaluation showed that the compound **5k** was slightly more active (IC_{50}~0.6μM) than others, the following X-ray crystallography however used **5b** (IC_{50}~1.6μM) as the ligand (maybe for the consideration of its purity as well as its commercial cost). The co-crystal structure of GS complexed with **5b**, magnesium (impurity from the synthesis), and MSO-P (at the amino-acid binding site) was analyzed by X-ray. The ligand electron density was clear, except for the site at the R^2 phenyl group of **5b**. It resulted from the rotation of the whole phenyl group, which substantiated the earlier assumption of large space at that site. Part of **5b** and GS co-crystal structure is shown in Figure 5. In this figure, all the hydrogen bonds were drawn, including one at GS's side chain Asp359, one at Lys361 (where ADP binds), one at Asn359, and another one at Ser280 together with a bridge water molecule. Other binding forces like π-π stacking interaction and π-cation interaction could also be observed in the picture at Phe232 and at Arg364. The hydrophobic pocket formed by Tyr129, Phe232 and Trp282 was filled by n-butylamino chain.

The X-ray structure together with all the designed ligands were then used to expand a SAR (similar with QSAR in 3.1.2) model of ligands they built before [23], by docking all the new synthesized compounds into the target model and evaluating their interaction. This docking process reversely helped to rationalize previous results obtained during FHDoE synthesis. It suggested that R^2 substituent would point out of the pocket and that was why it didn't have strong electrostatic interaction with the protein. This also explained why there was large space to accommodate the bulky R^2 groups, just because it didn't fit into the pocket. Additionally, it didn't matter whether the hydrogen bond donor was at meta-position or para- position, because either of them could form hydrogen bonding with Asn359 and/or Lys361. The X-ray structure indicated a halogen bond between the bromine atom and the oxygen atom on His278 judging from their distance. Many phenomena were explained by this afterward-docking study, however, it didn't explain the inactivity of R^1-cyclopentyl substituent, whereas the model saying it was active. To optimize the docking process, other factors such as free energy of hydration, alternative binding modes and loss of entropy should be added, too.

In silico SBDD find a potent hit compound. While feasible, expanded synthesis and inhibition evaluation will be conducted to modify the hit. The best-performing inhibitor will be examined together with the target to form a co-crystal, followed by X-ray crystallography analysis. In one paper, X-ray crystallography analysis even facilitated finding another efficient binding site [26]. Till now, the complete SBDD process was illustrated. Although different targets were chosen, they still gave an insight into how a TB drug could be developed.

4 Conclusions and discussions

Structure-based drug design is a newly developed computer-based method and it has forced the progression of drug discoveries and refinements. Via SBDD, a serials of novel drugs have been developed to treat tuberculosis. In these papers, targets and their corresponding binding sites were identified. *In silico* design would find out the most potential hit compounds for those targets. Practically, the synthesized ligands were tested by their biological activities and X-ray co-crystal analysis. A combination of recent achievements were used in this paper to display the whole SBDD circle. It also introduced several potential lead compounds which could be applied for further tests and modifications. With the development of more complex and accurate computer programs, the outlook for SBDD is to focus more on the target discovery. Aiming to conquer tuberculosis one day, scientists are now endeavoring to figure out the extremely essential targets for the drug design.

5 Acknowledgement

How to design medical drugs in the atom and molecule scale is always intriguing to me. And one of my uncles was dead of TB. Though this practice I've got a deeper understanding of drug design by SBDD against tuberculosis. Herewith I'd like to thank my supervisor Martin for the treasure time for the discussing and guidance together with Anna for the lectures. I also would like to thank the course NS190 for giving me this chance to investigate on the topic I'm interested in.

Reference

[1] Hippocrates. *Aphorisms*.
[2] *Nobel Foundation*. The Nobel Prize in Physiology or Medicine **1905**.
[3] Rivers EC, Nancera RL. *Drug Discov Today*. **2008**, 13, 1090-1098.
[4] Singh B, Mitchison DA. *British Medical Journal*. **1954**, 4854, 130–132.
[5] a) Jones D. *Bulletin of the History of Medicine*. **2002**, 76(4), 749–790; b) *Centers for Disease Control and Prevention*. **2000,** MMWR 49 (RR–6), 31–32; c) Yendapally R, Lee RE. Bioorg. *Med. Chem. Lett*. **2008**, 18(5), 1607–1611; d) Long JW. *Essential Guide to Prescription Drugs.* **1992**.

[6] *Global tuberculosis report 2013*. World Health Organization. **2013**.
[7] *Quanta/Insight II/Cerius2*. Molecular Simulations Inc., 9685 Scranton Road, San Diego, CA 92121-3752.
[8] *Sybyl*. Tripos, Inc., 1699 South Hanley Road, St. Louis, MO 63144-2913.
[9] *MacroModel*. Department of Chemistry, Columbia University, New York, NY 10032.
[10] *Grasp*. Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032.
[11] Bueno RV, Toledo NR, Neves BJ, Braga RC, Andrade CH. *J Mol Model.* **2013**, 19, 179–192.
[12] Speck-Planche A *et al*. *European Journal of Pharmaceutical Sciences*. **2013**, 48, 812-818.
[13] Roberts F, Roberts CW, Johnson JJ, Kyle DE *et al. Nature*. **1998**, 393, 801-805.
[14] Kumar M, Verma S, Sharma S, Srinivasan A, Singh TP, Kaur P. *Chem Biol Drug Des*. **2010,** 76, 277-284.
[15] Kumar M, Vijayakrishnan R, Rao GS. *Mol Divers*. **2010**, 14, 595–604.
[16] a) Chothia D, Lesk AM. *EMBO J*. **1986**, 5(4), 823-826; b) Kaczanowski S, Zielenkiewcz P. *Theoretical Chemistry Accounts*. **2010**, 125, 543-550; c) Marti-Renom MA, Stuart AC, Fiser A *et al. Annu Rev Biophys Biomol Struct*. **2000,** 29, 291-325.
[17] a) Takiff HE, Salazar L, Guerrero C *et al*. *Antimicrob Agents Chemother*. **1994**, 38,773-780; b) Aubry A, Pan XS, Fisher LM *et al*. *Antimicrob Agents Chemother.* **2004**, 48, 1281-1288.
[18] Bagchi MC. *J Mol Model*. **2011**, 17, 1607-1620.
[19] a) Hiasa H, Shea ME. *J Bio Chem*. **2000**, 275, 34780-34786; b) Wentzell LM, Maxwell A. *J Mol Biol*. **2000**, 304:779-791.
[20] a) Madurga S, Belda I, Vila J *et al. Chem Bio Chem*. **2008**, 9(13), 2081-2086; b) Maxwell A. J *Antimicrob Chemother*. **1992**, 30, 409-416.
[21] Kandasamy S, Hassan S, Gopalaswamy R, Narayanan S. *Journal of Molecular Graphics and Modelling.* **2014**, 52, 11–19.
[22] a) Gopalaswamy R, Narayanan S, Chen B, Jacobs WR *et al*. *FEMS Microbiol. Lett.* **2009,** 295, 23–29; b) Kumar D, Palaniyandi K, Challu VK *et al. Arch. Microbiol.* **2013**, 195, 75–80; c) Walburger A, Koul A, Ferrari G, Nguyen L *et al*. *Science*. **2004**, 304, 1800–1804; d) Papavinasasundaram KG, Chan B, Chung JH, Colston MJ, Davis EO *et al. J. Bacteriol*. **2005**, 187, 5751–5760; e) Malhotra V, Arteaga-Cortes LT, Clay G, Clark-Curtiss JE. *Microbiology.* **2010**, 156, 2829–2841.
[23] Nordqvist A, Nilsson MT, Lagerlund O, Muthas D, Gising J, Yahiaoui S, Odell LR, Srinivasa BR, Larhed M, Mowbrey SL, Karlen A. *Med. Chem. Commun*. **2013**, 3, 620-626.
[24] a) Harth G, Clemens DL, Horwitz MA. *Proc. Acad. Sci. USA*. **1994**, 91, 9342-9346; b) Wietzerbin J, Lederer F, Petit JF. *Biochem. Biophys. Res. Commun*. **1975,** 62, 246-252; c) Hirschfield GR, McNeil M, Brennan PJ. *J. Bacteriol*. **1990,** 172, 1005-1013.
[25] Odell LR, Nilson MT, Gising J, Lagerlund O, Muthas D, Nordqvist A, Karlen A, Larhed M. *Bioorg. Med. Chem. Lett*. **2009**, 19, 4790-4793.
[26] Schoichet BK, Stroud RM, Santi DV, Kuntz ID, Perry KM. *Science.* **1993**, 259, 1445-1450.