Virtual Screening of CPP Conjugated Peptide as QcrB BCC Protein Inhibitors of *Mycobacterium tuberculosis* Bacteria

Muhammad Ihsan Muttaqin, Manaman Huang and Usman Sumo Friend Tambunan

Department of Chemistry, Universitas Indonesia, Depok, Indonesia

**Abstract:** Tuberculosis is a global leading cause of death, just below COVID-19 and ranked above HIV. This disease is caused by a pathogen called *Mycobacterium tuberculosis*, which spreads easily through the air and is known to remain latent in most people's bodies, about a quarter of the world's population. The current problems with TB also include two main concerns: The official vaccine is ineffective and the bacteria keep gaining resistance to drugs. In this research, we proposed a strategy to create a new drug to overcome this resistance by simulating it through in silico methods by running molecular docking through Molecular Operating Environment (MOE) software, dynamic simulation through the iMODS website, and pharmacological prediction by comparing it through multiple pharmacophore predictors (AdmetSAR, SwissAdme, and pkCSM). We proposed five peptide drugs (Noopept, Glycyl-L-Proline, Leuteonosticon, Alaptide, dan NNZ-2591) that were then conjugated with a Cell-Penetrating Peptide (CPP) known for its transference prowess, with the receptor of QcrB protein, a complex related to the enzyme responsible for respiration and electron transference in MTB, which can prevent the bacteria development when inhibited. All candidates were picked following the rule of five proposed by Lipinski along with ADME and toxicity evaluation, with models and structures gained from the PubChem database. This research hopes to propose a way of combating drug-resistant TB by using a specified target and CPP for breaking the bacteria's lines of defense.

**Keywords:** MTB, Molecular Docking, Peptide, CPP, ADME-Tox

**Introduction**

Tuberculosis (TB) is caused by the bacterium *Mycobacterium Tuberculosis* (MTB) and can attack any body part, with the lungs being the most common. There are two conditions related to TB: Latent TB Infection (LTBI) and TB disease, which, if left untreated, can be fatal (CDCP, 2016). TB is easily spread through the air when infected people cough, talk, or sing. It is a severe global health problem, with an estimated 1.6 million deaths in 2021 alone. Treatment for TB is long and expensive, making prevention through early detection and prompt treatment crucial (WHO, 2021).

Drug-resistant tuberculosis (DR-TB) occurs when bacteria become resistant to the drugs used to treat TB. Resistance can develop when TB drugs are misused or mismanaged, exacerbated by extensive and inappropriate use, which allows bacteria to mutate and evolve resistance mechanisms (CDCP, 2017). Multidrug-resistant TB (MDR-TB) is caused by strains of *Mycobacterium tuberculosis* that are resistant to at least rifampicin and isoniazid, two of the most effective first-line TB drugs. Treatment of DR-TB is prolonged, toxic, and complex, requiring expensive second-line drugs that often cause serious side effects. The burden of DR-TB disproportionately affects low and middle-income countries and is exacerbated by factors such as weak health systems, poor infection control, and inadequate provision of TB services (WHO, 2021).

The bcc cytochrome complex is an enzyme complex that transports electrons and participates in respiration. This complex catalyzes the electron transfer from ubiquinol to cytochrome (Bahuguna and Rawat, 2020). Its function is to catalyze electron transfer from quinol to another electron acceptor protein. This complex is encoded by the QcrCAB operon, a large dimer protein complex known as complex III, and is part of the respiratory super complex (Berry *et al*., 2000). One of its subunits is QcrB. Various structurally diverse compounds have targeted the QcrB subunit of the cytochrome bcc...
complex, with most inhibition occurring at the $Q_p$ site (Bahuguna et al., 2021).

Telacebec, previously known as Q203, is a novel first-in-class drug candidate that has shown its efficacy in the treatment of TB (De Jager et al., 2020). Its development and clinical trial have validated MTB's Electron Transport Chain (ETC) as a viable drug target. The target for the said novel drug was the bcc cytochrome believed to be specified as QcrB (Bahuguna et al., 2021). ETC in MTB helps move protons and create a proton motive force which assists in generating energy or ATP. Inhibiting this process will lead to a quick loss of cell viability and even death of the MTB (Imran et al., 2023). Targeting this metabolism process can also combat dormant infection and potentially shorten treatment and research duration (Bald et al., 2017).

Cell-Penetrating Peptides (CPPs) are small peptides that have the ability to cross the plasma membrane of cells and deliver a variety of cargo, such as proteins, drugs, and nucleic acids. CPPs are widely used in biomedicine for basic research and therapeutic applications. CPPs can interact with the cell membrane through various mechanisms, such as electrostatic interactions, hydrophobic interactions, and receptor-mediated endocytosis. Several studies have demonstrated the potential of CPPs as delivery vehicles in various therapeutic areas, such as cancer, genetic disorders, and infectious diseases. The use of CPP can help further enhance the therapeutic effect of MTB drugs with its role as a delivery vector (Pári et al., 2020). However, there is not yet enough research conducted for the use of CPP against gram-positive bacteria.

In this research, we conducted several predictions and simulations using the QcrB subunit as a receptor for drug candidates and peptides as the ligands. These candidates consisted of peptides conjugated with CPP, namely HIV-1 Tat, in the hope of challenging drug-resistant TB, by predicting the inhibition capability of each candidate.

Materials and Methods

This research was conducted using an in silico method and is purely computational. No animals or chemicals were yet to be used in the process. This research aims to provide a theory rather than to conclude a wet lab experiment.

Protein and Ligand Selection

The receptors used in this research were built and downloaded from SWISS-MODEL (Waterhouse et al., 2018) with code 7elv.1 for QcrB (Zhou et al., 2021) and 7np 7.1 for MTB's inner membrane (Bunduc et al., 2021), which both were obtained through cryogenic microscope electron. The Ramachandran Plot results are mentioned on SWISS-MODEL.

The ligands used in this research were taken from the PubChem database for about 8944 peptides. All of them were sorted using Osiris data warrior (Sander et al., 2015) by following Lipinski's Ro5 (Lipinski, 2004) which the structures must have: H-bond donors ≤5, H-bond acceptors ≤10, Molecular weight ≤500 DA and log $p$≤5.

While conducting Osiris DataWarrior, we also included the absence of tumorgenic, mutagenic, reproductive effective and irritant attributes to help better select the candidates.

Docking Simulation

All the docking simulations in this research were done using Molecular Operating Environment (MOE) 2014.09 (Chemical Computing Group ULC, 2024), along with all the model preparations for both receptors and ligands. The molecular mechanic calculation or methods used for protein preparation and docking were done using AMBER10: EHT force field and ligand preparation was done using MMFF94 forcefield.

The statistical analysis was then conducted by comparing the Root Mean Square Deviation (RMSD) values as well as the $S$ values.

ChemDraw Ultra 12.0 was used to draw and link between the peptide candidates and CPP (HIV-1 Tat).

Pharmacological Prediction

Alongside Osiris DataWarrior, the tests for ADME and Toxicity were done using admetSAR (Yang et al., 2019) and pkCSM (Pires et al., 2015). SwissAdme (Daina et al., 2017) was also used as a reference, with the drug-likeness written on its website also being considered when selecting candidates. The processes were conducted by inputting the candidate's Simplified Molecular-Input Line-Entry system (SMILE) for the website to process.

Dynamic Simulation

The dynamic simulations were conducted using the iMODS server (López-Blanco et al., 2014) by putting the molecular docking results obtained from MOE after successful simulations.

Results and Discussion

Docking Simulation

The first step of docking simulation is to prepare the models that will be used as receptors and ligands, all of which were conducted using MOE 2014.09 software. The preparation and docking were simulated in the gas phase since the process of converting an atom or molecule into an ion by adding or removing charged particles such as electrons or other ions can occur in the gas phase (Aubry and Holmes, 2000). However, for the calculations, the receptor preparation used Amber10: EHT, while the ligand used MMFF94 because the Amber molecular
mechan was designed for larger molecules like protein. All the peptides then get their hydrogen and charge optimized, and their energy minimized.

After all the models were prepared, the simulation then occurred. Three types of molecular docking were conducted in this research: Virtual Screening used to identify the candidates based on the MOE database, rigid docking to simulate rigid internal geometry, and Induced Fit to simulate flexible internal geometry.

QcrB model used in this research (7e1v.1) was retrieved from SWISS-MODEL with a Ramachandran Favoured value of 93.10%, as seen in the number of dots presented on the more greenish area in Fig. 1. This serve as the mean that the model was valid to be used for further testing. The site used for QcrB followed the example given by (Zhou et al., 2021) in their journal and model, which are: PHE39, PRO40, PHE45, GLY48, GLU49, ALA51, LEU52, TYR53, PHE121, ILE125, HIS128, LEU129, ILE132, ASN145, TRP146, GLY149, SER150, LEU152, LEU153, ILE228, HIS231, LEU232, VAL235 and HIS240, PHE262, SER266.

However, after conducting our simulations, we found the binding sites to be: GLY48, ARG131, ARG142, HIS231, TRP236, and HIS240 for molecular docking and PHE45, LEU47, GLY48, GLU49, ALA51, LEU52, ILE125, HIS128, ALA137, PHE138, ASN145, TRP146, HIS231, LEU232, LEU234, VAL235 and HIS240 for molecular dynamic.

As for the inner membrane, we simulated every atom of the model to check whether there was any potential or not.

For rigid docking and induced fit, we used Triangle Matcher as placement with a score function of affinity δG and Forcefield as refinement with a score function of GBVI/WSA δG, with the retain values of 100 and 1, respectively. The docking simulation for the inner membrane with HIV-1 Tat was only conducted using flexible docking. All docking simulations were run in the gas phase under the AMBER10: EHT forcefield. The main two parameters used for the results were the S score in kcal/mol and RMSD in Å.

Initially, 7np7.1 contained 27 models, but most were duplicates of the others, except for 7, which are made of different sequences.

The peptides in Table 1. Are written using their PubChem CID. Their names are Noopept (180496), Glycyl-L-Proline (3013625), Leuteonosticon (191115), Alaptide (119329), dan NZN-2591 (11735696).

To examine the results of a molecular docking simulation, we need to evaluate the value of the obtained RMSD. If the numbers are higher than 2.0 Å, then we can conclude that the simulation ran poorly (Castro-Alvarez et al., 2017). Furthermore, for the best candidates, we want it to be as close as possible to the value of 1.0, as it suggests a better accuracy and a more natural inhibition that might occur.

### Table 1: QcrB docking results with peptides (R = RMSD)

<table>
<thead>
<tr>
<th>Peptide (PubChem CID)</th>
<th>Virtual screening</th>
<th>Rigid 100-1</th>
<th>Flexible 100-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>180496</td>
<td>-11.020</td>
<td>-9.91</td>
<td>1.68</td>
</tr>
<tr>
<td>3013625</td>
<td>-10.832</td>
<td>-6.70</td>
<td>1.36</td>
</tr>
<tr>
<td>191115</td>
<td>-11.282</td>
<td>-9.96</td>
<td>2.48</td>
</tr>
<tr>
<td>119329</td>
<td>-9.129</td>
<td>-6.22</td>
<td>0.96</td>
</tr>
<tr>
<td>11735696</td>
<td>-8.978</td>
<td>-6.17</td>
<td>1.01</td>
</tr>
</tbody>
</table>

Meanwhile, the S score indicates the receptor-ligand binding affinity (Attique et al., 2019), equivalent to ∆G<sub>binding</sub>. This S score is also the total and concluded calculation that MOE made to indicate how good the model was and can be compared by seeing which of the candidates provide lower scores as they are better according to the calculation.

The interactions shown in Figs. 2-3 were the ligand-receptor interactions of the numerically best candidate with the binding sites.

From the obtained results, three prominent residues were identified in the docking simulation: GLY48, HIS231, and HIS240, with significant receptor exposure. Receptor and ligand exposure affects cellular interactions, which can influence numerous biological processes and, in the case of drugs, disrupt these processes through inhibitory actions. This exposure also indicates the precise interaction between residues and ligands, particularly in highly flexible conformations.

Glycyl-L-Proline (3013625), Leuteonosticon (191115), and Alaptide (119329) also exhibit hydrogen bonding interactions that enhance and stabilize the inhibition between the ligand and its receptor. Additionally, Glycyl-L-Proline (3013625) forms a unique charge-charge interaction through a salt bridge. However, the strength of these interactions is not as strong as the hydrogen bonding that occurs with electronegative acceptor and donor molecules.
Fig. 2: Ligand interaction of QcrB and 11735696

Fig. 3: 3D interaction of QcrB and 11735696

Pantano et al. (2004) have simulated the performance of HIV-1 Tat in dynamic simulations in aqueous solutions. However, this can only serve as a reference to how they will computationally perform. As of late, CPP has only shown a good enough result against gram-negative bacteria such as *E. coli* (Lee et al., 2021) and cyanobacteria (Liu et al., 2013). Yet, there is still not enough data for gram-positive bacteria that provide a satisfactory result.

Nevertheless, seeing how we can create a CPP using a virus that attacks human cells, we might someday be able to create a CPP using a virus that can attack MTB, also known as mycobacteriophage (Hatfull, 2018). However, since there is a proven suitable CPP for MTB drugs, the use of HIV-1 Tat CPP in this research was merely a model to see how much of the drug candidates’ performance gets affected after being conjugated to a CPP.

The objective of the docking simulation of HIV-1 Tat against MTB’s inner membrane was to see whether the peptide gets docked to the last layer of MTB’s cell wall or not, as we needed the CPP to act as a transporter and will not get stuck binding to the cell wall, instead of penetrating it further inside and the result at Table 2 showed how poorly the two interact.

### Table 2: Inner membrane induced fit results with HIV-1 Tat

<table>
<thead>
<tr>
<th>Peptide</th>
<th>S</th>
<th>RMSD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Secretion system ATPase EccB5 (model A)</td>
<td>-17.7653</td>
<td>3.7662</td>
</tr>
<tr>
<td>Secretion system ATPase EccB5 (model D)</td>
<td>-14.4234</td>
<td>5.1987</td>
</tr>
<tr>
<td>Secretion system protein EccC5 (model G)</td>
<td>-17.2348</td>
<td>2.2242</td>
</tr>
<tr>
<td>Secretion system protein EccC5 (model M)</td>
<td>-14.4042</td>
<td>3.6509</td>
</tr>
<tr>
<td>Secretion system protein EccC5 (Model N)</td>
<td>-16.1595</td>
<td>2.5645</td>
</tr>
<tr>
<td>Secretion system protein EccC5 (model P)</td>
<td>-15.0028</td>
<td>4.8283</td>
</tr>
<tr>
<td>Mycosin-5 (model Y)</td>
<td>-16.2355</td>
<td>6.9814</td>
</tr>
</tbody>
</table>

After obtaining the docking results, we proceeded to conjugate it with the HIV-1 Tat CPP using a (Gly-Ser-Gly) linker, as it is known to have good flexibility, hydrophilicity, and spacer length. This process was done using ChemDraw by following the peptide conjugation rule through its C-terminal and N-terminal (Tambunan et al., 2017). The result was then simulated once more using MOE, but only through an induced fit with retained values of 100 and 1, respectively.

Table 3 shows the increase in S value equivalent to $\Delta G_{\text{binding}}$. This increase indicates a better binding affinity of each candidate. Yet that alone will not guarantee whether they will retain their inhibition property or not with QcrB. This is why a dynamic simulation is needed to further validate the result, especially with the use of MMFFX94 force field for ligand preparation has become less reliable since it was a calculation made for smaller molecules.

### Table 3: S value of ligand before and after conjugation

<table>
<thead>
<tr>
<th>Peptide</th>
<th>S (before)</th>
<th>S (after)</th>
</tr>
</thead>
<tbody>
<tr>
<td>180496</td>
<td>-10.6879</td>
<td>-14.6046</td>
</tr>
<tr>
<td>3013625</td>
<td>-6.8412</td>
<td>-18.1533</td>
</tr>
<tr>
<td>191115</td>
<td>-10.5357</td>
<td>-14.3164</td>
</tr>
<tr>
<td>119329</td>
<td>-7.3629</td>
<td>-14.3164</td>
</tr>
<tr>
<td>11735696</td>
<td>-6.8621</td>
<td>-16.6155</td>
</tr>
</tbody>
</table>

### Pharmacological Prediction

Pharmacokinetics is the study of how the body interacts with administered substances for the entire duration of exposure. The five parameters generally examined by this field include Absorption, Distribution, Metabolism, Excretion (ADME), and toxicity (Spiehler and Levine, 2022).

The result in Table 4 is to be expected as it was included during the sorting of peptides obtained from PubChem. Theoretically, none of the drug candidates will cause tumors (tumorigenic), genetic mutation (mutagenic), reproductive effect, or be irritant, which can cause harm to the patient consuming the drugs.
 Absorption, which refers to the process through which orally administered drugs are absorbed from the gastrointestinal system into the human body's bloodstream. CYP (cytochrome 450) are enzymes that contribute to metabolism's first phase (Zanger and Schwab, 2013). Substrates are drugs metabolized by CYP, while inhibitors are drugs competing for more than 1 CYP enzyme. hERG is related to QT syndrome, which can cause sudden death. Hepatoxicity relates to the drug's potential toxicity to the liver.

Hepatoxicity is by far the only concern for the drug candidates, although some drugs in the market also contain this effect, such as paracetamol (Rotundo and Pyrsopoulos, 2020). However wet lab tests are still needed to assess this prediction's actual results further. Combining it with the theory provided could help determine the dose of each candidate if they ended up being used as drugs.

Dynamic Simulation

The dynamic simulations were done with iMODS using the atomic modeling or Coarse Groaning (CG) of C5. The measure of a molecule's capability to deform (Fig. 4) at each of its residues is known as main-chain deformability. The chain's 'hinges' location can be determined by identifying regions that exhibit high deformability.

Deformability is the ability of a molecule to change its conformation in order to adapt to the target protein. iMODS calculated all the inputted molecular structures and showed them through the atom index. So to see the main results of it, we need to observe the lines and spikes at the back of the graph.

At an atom index of 2500 and above, we can see the deformability of the ligand. The deformability of the lower graph showed a low value around that number since it calculated the conjugated CPP that remained stable with its conformation. As for the drug candidate, it has shown an increase in deformability which is desired for flexibility.

So to get a clearer picture and better data regarding their performance, we test it further using AdmetSar while also looking at a couple of additional scores and values.

From the prediction in Table 5, we get more information regarding their toxicity probabilities. Ames test is used to test the mutagenic property against sample bacteria; none of the candidates is carcinogenic, which means they do not theoretically cause cancer, but category III of oral acute toxicity labeled them as slightly toxic and slightly irritant, compared to category IV which categorized as undamaging. Lastly, the LD₅₀ shown in Table 5 tells each candidate's dose in which they will kill 50% of the lab rat population. These toxicities can be used to learn the harmful effects the drugs might possess, in which none are desired to affect the patient. They can also help in deciding the doses the patient should and can consume.

After predicting the toxicity, we conducted another test regarding the rest of the pharmacokinetics. In this research, we used pkCSM software to test the ADME of each drug candidate as shown in Table 6.

The HIA in pkCSM stands for Human Intestinal Absorption, which refers to the process through which orally administered drugs are absorbed from the gastrointestinal system into the human body's bloodstream. CYP (cytochrome 450) are enzymes that contribute to metabolism's first phase (Zanger and Schwab, 2013). Substrates are drugs metabolized by CYP, while inhibitors are drugs competing for more than 1 CYP
Fig. 5: B-factor of 11735696 before and after HIV-1 Tat conjugation, respectively

Similar to deformability, the ligand’s B-factor results (Fig. 5) are shown at the back of the atom index. Since the NMR models proved RMSD values, the desired outcome would be a lower number. However, the B-factor in PDB indicates the molecule’s conformation capability to handle thermal influence. Hence a higher number is desired.

The stiffness of the motion is represented by the eigenvalue (Fig. 6) linked with each normal mode and it is directly proportional to the energy needed to deform the structure. A lower eigenvalue indicates an easier deformation, whereas a higher eigenvalue indicates a more difficult deformation. Since the value got higher after the conjugation, this also indicates the better stability of the molecule.

The variance (Fig. 7) associated with each normal mode has an inverse relationship with the eigenvalue. The colored bars represent the individual (red) and cumulative (green) variances. However, not much has changed with both data.

The covariance matrix (Fig. 8) indicates the coupling between pairs of residues, which determines whether they experience correlated (red), uncorrelated (white), or anti-correlated (blue) motions. The data shown can then be studied further to understand the interaction behavior between ligands and receptors.

The elastic network (Fig. 9) model determines the pairs of atoms that are connected by springs and each dot in the graph represents a spring between the corresponding pair of atoms. The stiffness of the springs is indicated by the color of the dots, where darker grays indicate stiffer springs and lighter grays indicate less stiff springs. The lower result showed a lot more stiffer springs at a higher index due to the presence of CPP, which is more or less stiff and stable.

After conjugation, the drug candidate showed a better or similar result each time compared to their unmodified variant, especially at a higher index. This is to be expected since both are using the same receptor. However, the number being relatively stable also indicates that the candidates retain their inhibition value even after conjugation.
Fig. 8: Covariance map of 11735696 before and after HIV-1 Tat conjugation, respectively

Fig. 9: Elastic Network of 11735696 before and after HIV-1 Tat conjugation, respectively

Conclusion

Five drug candidates have been proposed with corresponding RMSD values, namely Noopept (0.971 Å), Glycyl-L-Proline (0.971 Å), Leuteonosticon (0.979 Å), Alaptide (0.996 Å) and NNZ-2591 (1.002 Å). These peptides were downloaded from the PubChem database and designed to inhibit the QcrB protein of Mycobacterium tuberculosis. The candidates have been conjugated with Cell-Penetrating Peptides (CPP) from the HIV-1 Tat protein, with a decrease in $\Delta G_{\text{binding}}$ value indicating stronger binding. Each proposed peptide candidate meets the criteria for a molecular drug as outlined in the Rule of Five (Ro5) and each has been predicted to have relatively low toxicity except for hepatotoxicity. Although since the main objective of bioinformatic research is to produce a constructed theory, a wet lab experiment and testing are needed to be further conducted.

Acknowledgment

The researchers are grateful to department of chemistry, faculty of mathematics and natural sciences, Universitas Indonesia for their support of this research.

Funding Information

The researchers are grateful to Universitas Indonesia for funding the research through the Directorate of Research and Community Service (DRPM UI) under grant contract NKB-292/UN2.RST/HKP.05.00/2022.

Author's Contributions

Muhammad Ihsan Muttaqin: Data collection, data implementation, manuscript written.

Manaman Huang: Data implementation, proofread the manuscript.

Usman Sumo Friend Tambunan: Research supervisor, manuscript written and proofread.

Ethics

We declare that there is no ethical violation involved in this manuscript.

References


