INACTIVATION OF CU-ZN SUPEROXIDE DISMUTASE OF MYCOBACTERIUM TUBERCULOSIS H37RV

Zahra M. Al-Khafaji1, Aaisha B. Mahmood2 Marium B. Mahmood3

1Institute of Genetic Engineering and Biotechnology for Postgraduate Studies / University of Baghdad, Iraq.  
2Ministry of Agriculture, Veterinary Directorate, Baghdad Veterinary Hospital, Al-Dora Hospital, Iraq.  
3University of Baghdad / Financial Affairs Dept./ Computer Science.

*Corresponding Author: Dr. Zahra M. Al-Khafaji  
Institute of Genetic Engineering and Biotechnology for Postgraduate Studies / University of Baghdad – Iraq.

ABSTRACT

**Background:** Tuberculosis is one of serious disease caused by *Mycobacterium tuberculosis* which resides in macrophages, where exposed to toxic forms of reactive oxygen intermediates (ROS). The bacteria can overcome the intracellular defenses by producing different types of Superoxide dismutases (SODs), one of them Cu-Zn superoxide dismutase (SODc) with peripheral localization to protect the bacteria against oxidative stress and contribute to its pathogenicity and virulence. **Aim:** of this study was to find safe new inhibitors to SODc using QSAR modeling (MLR regression). **Results:** one QSAR model was satisfied the validation criteria, was used to find out new molecules, the latter were checked for different characters, two molecules 5,8-dihydroxy-1H-quinolin-4-one (zinc_34198840) and 1,3,6,8-Tetrahydroxynaphthalene (zinc_901903), were able to dock with the target protein (1PZS) strongly.

INTRODUCTION

Tuberculosis (TB) remains a major health problem worldwide. *Mycobacterium tuberculosis* (*M. tuberculosis*) the causative agent is very efficient and successful as it express numerous genes to evade the host immune responses to suit the intracellular life style and resist various drugs.[1] The mycobacteria are readily phagocytized by macrophages due, in part, to their very high cell surface hydrophobicity. Inside the macrophages they are exposed to toxic forms of oxygen compounds such as peroxides, superoxide and other reactive oxygen intermediates (ROS).[2] It has been found that *M. tuberculosis* virulence linked to resistance of H2O2, the low virulent isolates were found to be more susceptible to peroxide compared to high virulent isolates.[3] It has been found that number of factors such as catalases, peroxidasises and superoxide dismutesases permitting mycobacteria to overcome the intracellular defenses of macrophages. Superoxide dismutesases (SODs) paly an important role in protection against oxidative stress and contributed to the pathogenicity and virulence of many bacterial species.[4] *M. tuberculosis* produces SODs employ metal cofactor,[2] among these Cu-Zn superoxide dismutase (SODc), which is considered as virulence factor, as SODc-null mutants were more susceptible to superoxide compared to their SODc-wild type parents.[4]

Tuberculosis needs development of new drugs with new mechanisms of action and not interfere with other drugs as in people infected with HIV [5], or interfere with human targets. Different approaches are used to discover/design new drugs at early steps. Among these Quantitative Structure-Activity Relationship (QSAR), which is one of important approach for Computer Addie Drug Design (CADD).[6-9] QSAR models are multiple regression (MLR) models to relate a set of predictor variables (Descriptors) to the response variable (Biological Activity), this approach alongside with molecular docking were used to predict activities of various inhibitor compounds.[10-12] The relationship represented by.

**Activity (Response) = f(Physiochemical properties /or Structural Properties)**

When entirely the components are numerical a mathematical equation can be developed. The use of computational screening methods can potentially reduce the time, cost and effort needed for initial screening of candidate drug compounds with pharmacological activity against TB, and helping in prioritize compounds testing and minimizing randomization in the laboratory.[13]

The aim of this study was to develop validated QSAR model for estimation of inhibitors for SODc of *M. tuberculosis* H37Rv and using it for finding more molecules, and docking them with the target protein.

**MATERIALS AND METHODS**

Different databases and software were used, for different purposes.
Databases

**Binding DB:** https://www.bindingdb.org/bind/index.jsp.
Used to find out the inhibitors of SODc. Other source for inhibitors were used as well.

**Uniprot database:** https://www.uniprot.org/
To find out some information about the target (SODc)

**Zinc database:** http://zinc.docking.org/
Used to download different chemical format, and information about compounds.

Software

**Molinspiration:** https://www.molinspiration.com/cgibin/properties?textMode=1
Used for finding some molecules descriptors.

**SwissADME:** http://www.swissadme.ch/ [14]
Used for finding characters of molecules.

**SwissSimilarity:** http://www.swisssimilarity.ch/[15]
Used for virtual screening of new molecules.

**NCBI/BLASTp**
Used to find out similar proteins to the target.

**T.E.S.T. software:**
https://www.epa.gov/chemicalresearch/toxicity-estimation-software-tool-test. To find out the safety of molecules. [16]

**PyRx software v.8**
https://pyrx.sourceforge.io/
Used for docking.

**PyMOL software**
https://pymol.org/2/
Used for docking vitalization.

**DoGSiteScorer software**
Used to find binding pockets of the protein

**OriginPro2016**
Used for graphing and calculation of some results.

**CarcinoPred-EL**
http://ccisipb.lnu.edu.cn/toxicity/CarcinoPred-EL/
Used for estimation of compound carcinogenicity. [17]

**Model validation:** The robustness, applicability and stability of the generated QSAR model have been established by validation. The model was externally validated by using compounds of Test set, and calculated the R² pred.

**Docking:** This was done using PyRx package, after preparing the ligands (compounds), which were optimized to its lowest stable energy state. [18] The minimization was done until the energy change is less than (0.1) kcal/mol, using OpenBabel software incorporated in PyRx, the ligands were updated using PyRx software, and transformed into pdbqt format. The target macromolecule SODc was prepared to get pdbqt format, was docked after let the search space to its maximum. The results recorded as binding affinity (kcal/mol) with RMSD value of zero.

**RESULTS AND DISCUSSION**

*Mycobacterium tuberculosis* invades and replicates within the host macrophages, as an immune response, macrophages initiates a respiratory burst and produces high level of ROS to counter and kill the invade bacteria. [19] The important point is that the survivability of *M. tuberculosis* is highly dependent on the level of ROS produces by host immune cells, if the ROS levels are overwhelmed by *M. tuberculosis* antioxidant system, the bacteria will continue to survive and replicate in the host. [20] This situation is critical during an early time point of infection. [21] *M. tuberculosis* could utilize (SODc, EC 1.15.1.1) and catalase (Hydroxyperoxydase, EC 1.11.1.6) enzymes to transform the toxic ROS such as superoxide and H₂O₂ into water and oxygen. [22] *M. tuberculosis* has different SODs, this study concerns the Cu-Zn superoxide dismutase (SODc), with pdb ID 1PZS, (EC 1.15.1.1), gene number Rv0432 and Uniprot ID P9WGE9. The chosen enzyme has the crystal structure 1PZS with resolution 1.63 Å. The unique feature of SODc from *M. tuberculosis* is a fully functional copper-containing enzyme lacking Zinc in the active site, this is due to deletion or mutant of metal binding residues His115 and His123. [23] The enzyme has a lipoprotein binding motif which suggests that it may anchored in the membrane to protect *M. tuberculosis* from ROS at the bacterial surface and eventually contributes in the resistance of *M. tuberculosis* against oxidative burst products generated by activated macrophages, SODc was reported in most mycobacterial species. [24] The predicted protein sequence contains 240 amino acids with a putative signal peptide at the N-terminus, and is located in the periphery of *M. tuberculosis*. The enzymatic activity and subcellular localization of novel SODc suggests that it may play a role in determining virulence of *M. tuberculosis*. [25] The protein was checked for Ramachandran plot and found that 97.6% of residues are in favored region, in addition binding pockets and druggability were estimated using DoGSiteScorer software [26] as shown in Figure 1.
Al-Khafaji et al.  European Journal of Pharmaceutical and Medical Research

<table>
<thead>
<tr>
<th>Name</th>
<th>Volume $\AA^3$</th>
<th>Surface $\AA^2$</th>
<th>Drug Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>P_0</td>
<td>678.66</td>
<td>1331.54</td>
<td>0.85</td>
</tr>
<tr>
<td>P_1</td>
<td>201.6</td>
<td>479.17</td>
<td>0.56</td>
</tr>
<tr>
<td>P_2</td>
<td>151.04</td>
<td>344.17</td>
<td>0.48</td>
</tr>
<tr>
<td>P_3</td>
<td>138.43</td>
<td>357.53</td>
<td>0.24</td>
</tr>
<tr>
<td>P_4</td>
<td>111.23</td>
<td>262.53</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Figure 1: The binding pockets of SODc enzyme.

The active site of the enzyme contains copper as a cofactor, which coordinate with His47, His49 and His 126 as shown in Figure 2.

Choosing of Inhibitors: Many studies showed that SODc inhibitors have potential potency as drugs. These inhibitors exert their activity with different mechanisms. Among these polyphenolic compounds such as flavonols were found to diminish the effect of SODc.[27,28] Some inhibitors of SODc are already used as drugs, for example Nitoprusside appears to inhibit the known types of SODs irrespective of their metal prosthetic groups and regardless of their sources.[29] Disulfiram is a copper-chelator has potential anti-cancer activity,[30,31] another heavy metal chelator is 8-hydroxyquinoline which is proved to have antimicrobial activity[32], in addition to other inhibitors. This variation of molecules classes and different mechanisms of action exert their effect on developing a suitable QSAR models, since their descriptors are not in harmony. This is true as the basis for mechanistic interpretation states that the properties of biological interactions of chemicals are inherited to their molecular structure.[33] And throughout the previous studies, the majority of QSAR modeling applications to design new anti-TB agents have been used to modify previously discovered congeneric group of chemicals.[34]

In this study about 43 inhibitor molecules were collected from databases and literatures, and 22 descriptors form different classes were used. The inhibitory activity was expressed as IC50 values, this subjected to data transformation using logarithmic value to base 10, to ensure that are more uniformly distributed, in addition normality test was carried out which indicated that the values are under normal distribution space as shown in Figure 3.

Figure 2: The structure of SODc enzyme (1PZS).

The mechanism of activity depends on shifting of copper atom valence between +1 and +2 at the catalytic center.

The molecules were divided into Training set and Test set using different criteria and combinations to develop hundreds of QSAR models, most of them failed to satisfy the validation criteria. One model was acceptable: $Y = 0.413895 \times \text{Acceptor} - 0.185653 \times \text{Donor} + 0$

Descriptors used were orthogonal i.e., not related to each other[14] as shown in Figure 4.

Figure 3: Normal distribution of inhibitor IC50 values.

They are correlated to LogIC50 to some extent as in Figure 5A and B.
The statistics of the model are shown in Table 1.

Table 1: Statistics of developed QSAR model.

<table>
<thead>
<tr>
<th>Validation Parameters</th>
<th>Name</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>r</td>
<td>Correlation coefficient</td>
<td>0.859789</td>
</tr>
<tr>
<td>R²</td>
<td>Coefficient of determination</td>
<td>0.739237</td>
</tr>
<tr>
<td>R²_adj</td>
<td>Adjusted R²</td>
<td>0.660439</td>
</tr>
<tr>
<td>Q²</td>
<td>Squared Validation correlation</td>
<td>0.708026</td>
</tr>
<tr>
<td>R²-Q² cv</td>
<td>Cross validation coefficient</td>
<td>0.031211</td>
</tr>
<tr>
<td>R²_pred</td>
<td>Predicted coefficient</td>
<td>0.5516</td>
</tr>
<tr>
<td>P(95%)</td>
<td>Confidence interval at 95% confidence level</td>
<td>&lt;0.05</td>
</tr>
<tr>
<td>F-value</td>
<td>Significance of regression F-value</td>
<td>22.67919</td>
</tr>
<tr>
<td>Tabulate F-value</td>
<td>Critical Significance of regression F-value (95%)</td>
<td>2.92E-05</td>
</tr>
<tr>
<td>s</td>
<td>Standard error</td>
<td>0.805452</td>
</tr>
</tbody>
</table>

Model validation: The purpose of validation is to provide a statistically reliable model with selected descriptors as a consequence of cause-effect and not only of pure chemical relationship obtained by chance. Table 1 shows high correlation coefficient ($r = 0.859789$), $R^2$ (0.739237) which explains that 74% of variance in biological activity of tested compounds. $R^2_{adj}$ was calculated (0.660439) which adjusted for the number of explanatory terms of model and takes in consideration the number of degrees of freedom, it decreases upon addition of new descriptor (variable), and dose not reduce unexplained variance. The statistical significance of regression model can be assessed as well by means of F value, which represents the ratio between explained and unexplained variance for given number of degree of freedom, the higher the F value the greater probability is that the model is significant, when the value is greater than a tabulated value for chosen level of significance (typically 95%). The regression coefficient should be at $P<0.05$ which is checked by t-test. The difference of squared correlation $Q^2$ for internal validation (i.e. using the value of Training set) showed that $R^2-Q^2$ not exceed 0.2-0.3 and this indicates that the model not suffering from overfitting. Other model qualification is the value of standard error, for good model should be low, since it measures the dispersion of observed values about the regression line. It has been suggested that the external validation might be the only way to estimate the predictive power of QSAR model and considered the most rigorous validation power, because the compounds in the external Test set do not affect the model development which should be supported by $Q^2$ value. This indicated by $R^2_{pred}$ and it has been stated that the value of $R^2_{pred}$ above stipulated value of 0.5 are considered to be well predictive, the results shown in Figure 6.
Figure. 7: The values of IC50 predicted using the generated QSAR model (A), and their probability (B).

The Training set molecules were used for virtual screening using SwissSimilarity server, under FP2 Fingerprints/Drug item and Zinc DrugLike as it contains the largest collection of molecules (10,639,400). Top fifteen molecules similar to each Training set molecule were used for further studies and filtration. This step resulted 238 molecules. The latter were checked for their mutagenicity and teratogenicity (Developmental Toxicity) using TEST software. Carcinogenicity was checked using CarcinoPred-EL ADME properties of molecules were checked using SwissADME server to check some characters qualified them as drugs, such as PAINS, GI absorption, P-gp substrate, solubility, Bioavailability Score and Synthetic accessibility. Two molecules zinc_34198840 and zinc_901903 were passed all the filtration steps, and 14 molecules passed the filtration steps, but, with unknown mutagenicity and development toxicity. The IC50 values are shown in Table 2.

Table 2: The predicted IC50 values of new molecules obtained by virtual screening

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Log IC50</th>
<th>IC50 µM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZINC39080060</td>
<td>0.270831</td>
<td>1.865654</td>
</tr>
<tr>
<td>ZINC00066039</td>
<td>0.456484</td>
<td>2.860777</td>
</tr>
<tr>
<td>ZINC38860896</td>
<td>0.499073</td>
<td>3.155535</td>
</tr>
<tr>
<td>ZINC49474633</td>
<td>0.499073</td>
<td>3.155535</td>
</tr>
<tr>
<td>ZINC49474797</td>
<td>0.499073</td>
<td>3.155535</td>
</tr>
<tr>
<td>ZINC19366504</td>
<td>0.684726</td>
<td>4.83867</td>
</tr>
<tr>
<td>ZINC19395685</td>
<td>0.684726</td>
<td>4.83867</td>
</tr>
<tr>
<td>ZINC19685483</td>
<td>0.684726</td>
<td>4.83867</td>
</tr>
<tr>
<td>ZINC19901168</td>
<td>0.870379</td>
<td>7.419575</td>
</tr>
<tr>
<td>ZINC19919685</td>
<td>0.870379</td>
<td>7.419575</td>
</tr>
<tr>
<td>ZINC39113438</td>
<td>0.870379</td>
<td>7.419575</td>
</tr>
<tr>
<td>ZINC00901903</td>
<td>0.912968</td>
<td>8.184045</td>
</tr>
<tr>
<td>ZINC19735929</td>
<td>1.056032</td>
<td>11.3771</td>
</tr>
<tr>
<td>ZINC32918840</td>
<td>1.284274</td>
<td>19.24305</td>
</tr>
<tr>
<td>ZINC21992187</td>
<td>2.154653</td>
<td>142.7753</td>
</tr>
</tbody>
</table>

Docking studies: Since statistics can never replace chemistry and biology, non-statistical validation is required, so docking process was performed for the candidate molecules of Table 2 with target protein. The molecules/ligands were prepared and optimized for their lowest stable energy stats. The prepared molecules/ligands were docked to target protein (all of them in pdbqt format) using AutoDock vina of PyRx software v.8. Docking results evaluated by binding affinity and RMSD value of zero, this ranged from -3.1 to -6.6 kcal/mol, shown in Table 3.

Table 3: The binding affinity of molecules obtained by virtual screening

<table>
<thead>
<tr>
<th>Molecule</th>
<th>Energy kcal/mol</th>
</tr>
</thead>
<tbody>
<tr>
<td>zinc_21992187</td>
<td>-6.6</td>
</tr>
<tr>
<td>zinc_34198840</td>
<td>-5.5</td>
</tr>
<tr>
<td>zinc_901903</td>
<td>-5.3</td>
</tr>
<tr>
<td>zinc_66039</td>
<td>-5.2</td>
</tr>
<tr>
<td>zinc_19901168</td>
<td>-4.8</td>
</tr>
<tr>
<td>zinc_19919685</td>
<td>-4.4</td>
</tr>
<tr>
<td>zinc_49474797</td>
<td>-4.4</td>
</tr>
<tr>
<td>zinc_19395685</td>
<td>-4.3</td>
</tr>
<tr>
<td>zinc_38860896</td>
<td>-4.2</td>
</tr>
<tr>
<td>zinc_49474633</td>
<td>-4.1</td>
</tr>
<tr>
<td>zinc_19685483</td>
<td>-4.0</td>
</tr>
<tr>
<td>zinc_19735929</td>
<td>-4.0</td>
</tr>
<tr>
<td>zinc_19366504</td>
<td>-3.8</td>
</tr>
<tr>
<td>zinc_39113438</td>
<td>-3.8</td>
</tr>
<tr>
<td>zinc_19366286</td>
<td>-3.5</td>
</tr>
<tr>
<td>zinc_39080060</td>
<td>-3.1</td>
</tr>
</tbody>
</table>

The visualization of candidate ligands (zinc_34198840 and zinc_901903) was done using PyMOL. These are shown in Figure 8A for zinc_34198840 and Figure 8B for zinc_901903. The Figure shows that ligand zinc_34198840 attached to the protein away from the active site containing Cu, and there is some contacts with...
significant values i.e. less than 3Å°. The same scenario occurs with the molecule zinc_901903.

The rest 14 molecules showed more or less similar behavior, in that interacting with the protein away from the active/catalytic center.

**Characters of candidate molecules**

**Molecule zinc_34198840**

5,8-dihydroxy-1H-quinolin-4-one.

Is one of Quinoline derivatives which are used for treatment of different classes of diseases caused by pathogens, among them they have been tested as anti-TB.\(^{[38]}\)

**Molecule zinc_901903**

1,3,6,8-Tetrahydroxynaphthalene.

The molecule is type III polyketide, it is usually an intermediate in synthesis of dark pigments produced by some organisms such as *Wangiella dermatitidis, Colletotrichum lagenarium, Streptomyces coelicolor, Streptomyces peucetius, Nocardia sp.*\(^{[39-43]}\)

In conclusion the introduced candidate molecules in this study might be a promising inhibitors and future drugs, as their relatives are in use for treatment. And especially SODc is in peripheral area of *M. tuberculosis* cells, and chelators were estranged from the active site to be away from interfering with metalloproteins/enzymes of human body that are essential for the latter and contain copper in their active sites.

**REFERENCES**
