Bioorganic & Medicinal Chemistry Letters 18 (2008) 3152–3157

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry Letters

journal homepage: www.elsevier.com/locate/bmcl

Structure-based prediction of *Mycobacterium tuberculosis* shikimate kinase inhibitors by high-throughput virtual screening

Aldo Segura-Cabrera *, Mario A. Rodríguez-Pérez

Laboratorio de Biomedicina Molecular, Centro de Biotecnología Genómica, Instituto Politécnico Nacional, Blvd. del Maestro esquina Elías Piña, Col. Narciso Mendoza, 88710, Cd. Reynosa, Tamaulipas, Mexico

ARTICLE INFO

Article history:
Received 25 March 2008
Revised 25 April 2008
Accepted 1 May 2008
Available online 4 May 2008

Keywords:
Structure-based drug design
Mycobacterium tuberculosis
Shikimate kinase inhibitors
Triazole
Tetrazole

ABSTRACT

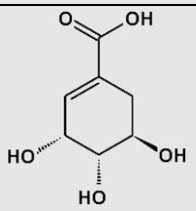
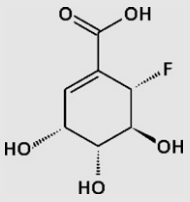
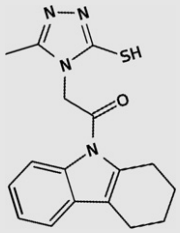
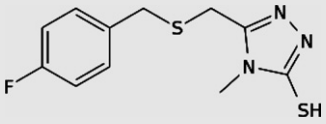
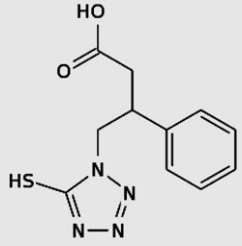
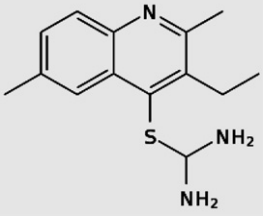
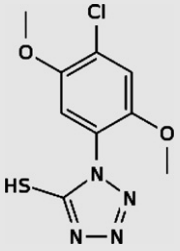
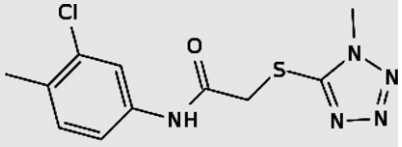
A structure-based virtual screening protocol was used to predict *Mycobacterium tuberculosis* shikimate kinase (MtSk) inhibitors. Docking simulations were performed using eHiTS® software and 644 drug-like compounds were identified as potential inhibitors. Forty-two percent of such inhibitors had a structural relationship to a triazole or a tetrazole heteroaromatic system which may provide a candidate lead for the discovery of MtSk inhibitors.

Tuberculosis (TB) is the leading cause of death in the world from a bacterial infectious disease. It has been estimated a prevalence of 1/3 the world's population, an incidence of 9 million cases each year, and 5% of the cases are bacteria resistant to anti-TB drugs.[1] The treatment of TB remains unsatisfactory. Multidrug-resistant *Mycobacterium tuberculosis* (MDR-MT) has become resistant to isoniazid and rifampicin. Likewise, the extensively drug-resistant MT (XDR-MT) has become resistant to any fluoroquinolone, kanamicyn, amikacin, and capreomycin.[34] This situation affects individuals in 45 countries.[1] In contrast to other bacteria, resistance in *M. tuberculosis* (Mt) results from chromosomal mutations.[35] Factors for emergence of resistance in Mt are multiple. Insufficient antibiotic coverage during a long anti-TB treatment period results in selection of refractory genotype bacteria. Antibiotic resistance is an inevitable process in bacterial evolution, and is a direct result of selective pressure by antimicrobial therapy.[36,37] The resistance delay may not be feasible by targeting enzyme catalytic sites. Emergence of resistance may possibly be prevented by using, rationally, antibiotic combinations, thus the prudent use of antibiotics will eventually result in a decline in the prevalence of drug resistance. There is an urgent need for discovering new drug targets for additional antimycobacterial drugs, particularly those effective at short treatment periods. The availability of the full genome sequence of Mt[2] has provided opportunities for the identification of new drug targets. Thus, it is expected that the development of novel, more effective, anti-TB drugs will result in improved therapeutic regimens aimed at preventing the emergence of MDR-TB, and reducing the costs of medical attendance and compliance. Genetic products of Mt have been identified which are essential for its survival as they regulate physiological processes such as metabolism, persistence, virulence, and cell signaling.[3–10] The understanding of the physiological processes of Mt is critical when designing new drugs anti-BT. For instance, it has been shown that the shikimate biosynthetic pathway of Mt, where the 3-deoxy-D-arabino-heptulosonate-7-phosphate (DAHP) is converted to chorismate, is essential for the synthesis of all aromatic amino acids, as well as other metabolites, such as folic acid and ubiquinone.[11] Given that the shikimate pathway is also essential for other parasites, bacteria, and fungi[12–15] but absent in mammals,[16] the enzymes involved in the shikimate pathway are considered as potential targets for the development of antimicrobial agents that would produce minimal side effects in non-target organisms such as mammals.[17] Shikimate kinase (Sk) is the fifth enzyme in the shikimate pathway that catalyzes the phosphate transfer from ATP to shikimate to generate shikimate 3-phosphate and ADP.[18] The shikimate kinase of *M. tuberculosis* (MtSk) has been extensively studied and characterized through X-ray crystallography techniques indicating that it belongs to the structural family of nucleoside monophosphate (NMP) kinases.[15,19–25] These functional and structural studies of MtSk provide a background for the discovery of MtSk inhibitors, but these remain to be revealed. This letter describes potential inhibitors of MtSk gathered by using an in silico high-throughput

* Corresponding author. Tel./fax: +52 899 924 36 27.
E-mail addresses: asegurac@ipn.mx, aldosegura@gmail.com (A. Segura-Cabrera).

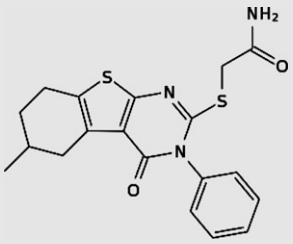
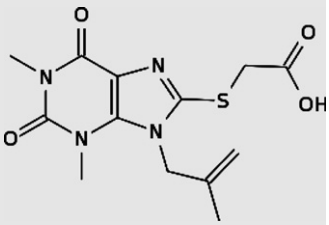
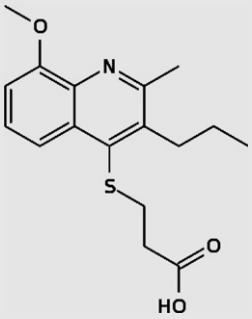
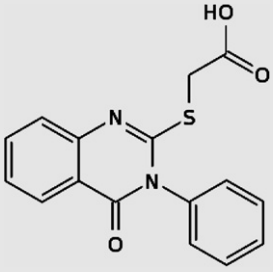
0960-894X/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.bmcl.2008.05.003

Table 1
The eHiTS score, and chemical structure identified using the structure-based screening protocol

Rank	Id	eHiTS score[26]	Structure
0	Control shikimate	−4.260	
0	Control 6-*S*-fluoroshikimate	−4.268	
1	asxe1	−7.252	
2	MW1	−6.595	
3	diver1	−6.453	
4	asxc1	−6.372	
5	spch1	−6.366	
6	asxc2	−6.319	

(*continued on next page*)

Table 1 (*continued*)

Rank	Id	eHiTS score[26]	Structure
7	spca1	−6.142	
8	spca2	−6.109	
9	kin1	−6.08	
10	asxb1	−6.041	

screening protocol. Although it will require further unbiased experimental validation, the finding of the SkMt inhibitors, for the first time reported herein, will be helpful in the rational development of novel drugs against Mt.

The eHiTS® 6.2 (Electronic High-Throughput Screening) software package[26] was used for flexible docking simulations into the shikimate binding site of MtSk. There is an association (884 PDB complexes, $R = 0.75$ and $q = 1.61$) between scoring function and experimental binding affinity using the eHiTS® software (data from SimBioSys Inc. at http://www.simbiosys.ca). eHiTS® provided comprehensive searching, conformational-space coverage, and generated all major docking modes that were compatible with the steric and chemistry constraints of the target cavity for each small molecule from the compounds database. The docking algorithm is divided into an exhaustive search algorithm based on graph matching, and a novel scoring function based on local surface point contact evaluation. The properties of surface included hydrogen bonding, hydrophobicity, electrostatic potential, van der Waals contact energy, ionic interactions, and pocket depth (considering changes in the dielectric constant). Complementary surface points that were within a threshold distance received a positive score. Points that have no ligand or repulsive surface points within the cutoff distance were assigned a penalty score. The sum of all receptor surface point scores was computed, and constituted the score for that pose of the ligand.[26] The eHiTS evaluated automatically all possible protonation states for ligands and receptor. Thus, in docking run each state was evaluated and scored systematically to achieve the best possible binding score, avoiding those

multiple functional groups with variable protonation states to produce a combinatorial effect. The crystal structure of SkMt bound with shikimate and ADP, access code PDB 2dfn,[27] was used as the template for docking simulations. The water molecules, ions, and the ADP were removed from the template, thus the SkMt-shikimate substrate complex was used alone. The input file containing ligands (compounds and controls) in multiple mol2 format and the receptor were set up, and pre-processed. Firstly, each ligand was divided into rigid fragments, and connecting flexible chains. The rigid fragments were each docked separately in the receptor site, and then a fast graph match algorithm found poses of each fragment, and reconnected to form the input ligand. The fully automated pre-processing phase included separation of receptor and ligand from a PDB complex file, conversion of the input receptor and ligands into native eHiTS file's format, and generation of steric grid and feature graph description of the cavity's characteristics. Secondly, the docked poses were refined by a local energy minimization in the receptor's cavity of the SkMt to produce the final ligand pose. Finally, eHiTS generated 3D coordinates of docked conformations of ligands in the receptor's cavity. The generated docked poses were scored, and saved into SDF files for further conformational visualization, and analysis.

The accuracy parameter was set to the highest level (sixth level) during the docking simulations. The validation training module in eHiTS® was used for enhancing the program's performance. The validation training module used PDB complexes to better predict the binding modes for a family of receptors. This methodology ensured that the scoring function was tailored so that ligands that dock in similar way to those found in the family of structures received the best score and ranking. The eHiTS training utility carried out optimization of the weights of the scoring functions used throughout the docking process, and these weights were stored in rkba file format. The rkba file was used as the parameter file when running eHiTS calculations by the "-rkb receptors .rkba" flag. Thus, all structures of MtSk, in complex with shikimate substrate, were downloaded from the RCSB PDB[31] and the train module was used to produce the .rkba file. The PDB ids were 1u8a, 1zyu, 2dfn, 2g1k, 2iyq, 2iyr, 2iys, 2iyx, and 2iyz.

Three-dimensional structures of compounds were extracted from collections with ADME/tox filters at the online service FAF-Drugs[28] which produced 214,492 compounds. These compounds satisfied Lipinski's rule of five[29] constraints (molecular weight < 500, partition coefficient log$P$ < 5, number of H-bond donor <5, and number of H-bond acceptor <10), and about 100 rules that eliminated compounds with undesirable chemical groups. Therefore, these compounds have physicochemical properties exhibited by orally bio-available drugs.

The shikimate substrate and 6-*S*-fluoroshikimate, a well-characterized competitive inhibitor in *Escherichia coli*,[17] were introduced as controls into the compound database. The docking screening protocol was able to detect correctly the inhibitor with a score higher than the shikimate substrate (Table 1). Thus, a total of 644 small molecules were identified with better docking scores than the controls (Table 1 and Supplemental Material). Therefore, these compounds exhibit better binding affinity for MtSk, with the best observed score being −7.25 (in logK_d units). Typically, compounds found with highly favorable scores (more negative values) had structural features that included a mercapto group, and a triazole or tetrazole heteroaromatic system. The potential hits identified were contained in these moieties in 86% and 42% of total cases, respectively. The score and chemical structures of the top 10 hits found in this study are summarized in Table 1. In addition, no significant association (r = −0.10, n = 50, and P > 0.05) was found in the top 50 compounds when their molecular weight was plotted against their score, indicating that predicted affinity is due to specificity and not simply to molecule's size.

Table 2
The interactions between MtSk residues with 200 top-scoring compounds

Position	Contact percent	Features of interaction[a]
R117	99	HF, HB
I45	93	HF
L119	89	HF, HB
P11	82	HF
R58	70	HB
G80	57	HF, HB
P118	54	HF
G81	44	HF, HB
R136	42	HB
V116	42	HF, HB
F57	40	HF
D34	28	HF, HB
F49	23	HF
G79	20	HF
K15	15	HF

[a] HF, hydrophobic forces and HB, hydrogen bond.

The 200 top-scoring compounds, that resulted from the docking screening protocol, were also graphically examined using the software Chevi 6.2® Chemical Visualizer (available from Simbiosys Inc.) to determine key interactions, and complementarity with residues of the shikimate binding site. The shikimate binding site is characterized by a hydrophobic surface together with a number of charged residues which project into the pocket. Several studies have revealed that some of these residues (D34, R58, G80, L119, and R136) in the shikimate binding site are essential for the positioning of the substrate and enzymatic catalysis.[20–24,27]

The 200 top-scoring compounds, from the eHiTS-predicted conformations, showed that these compounds interact frequently with the above-mentioned residues, and with others that are highly conserved into the sequences of the Sk.[25] It was the case for P11, I45, R58, G80, R117, and L119. A list of the interactions produced, among the 200 top-scoring compounds, along with the MtSk resi-

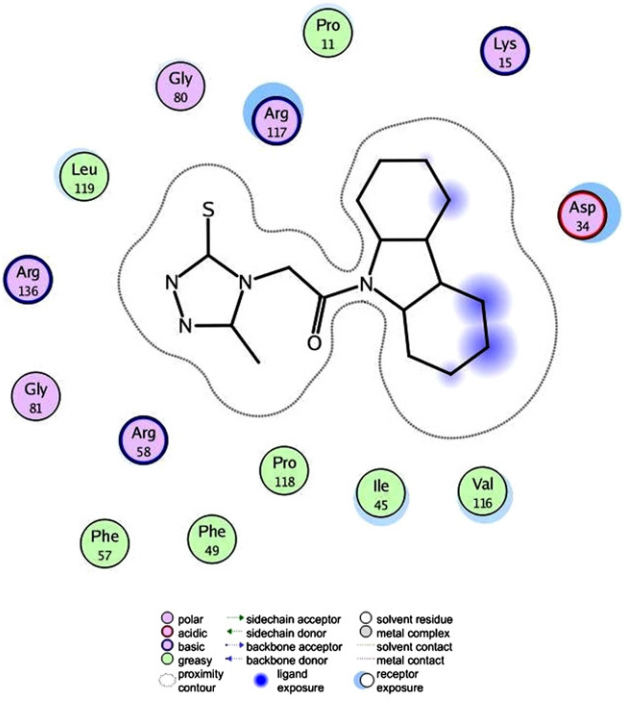


Figure 1. Interactions between MtSk residues with one top-scoring compound. The picture was inspired on LIGPLOT[32] with an algorithm implemented in the program Molecular Operating Environment (MOE)[30] at www.chemcomp.com.

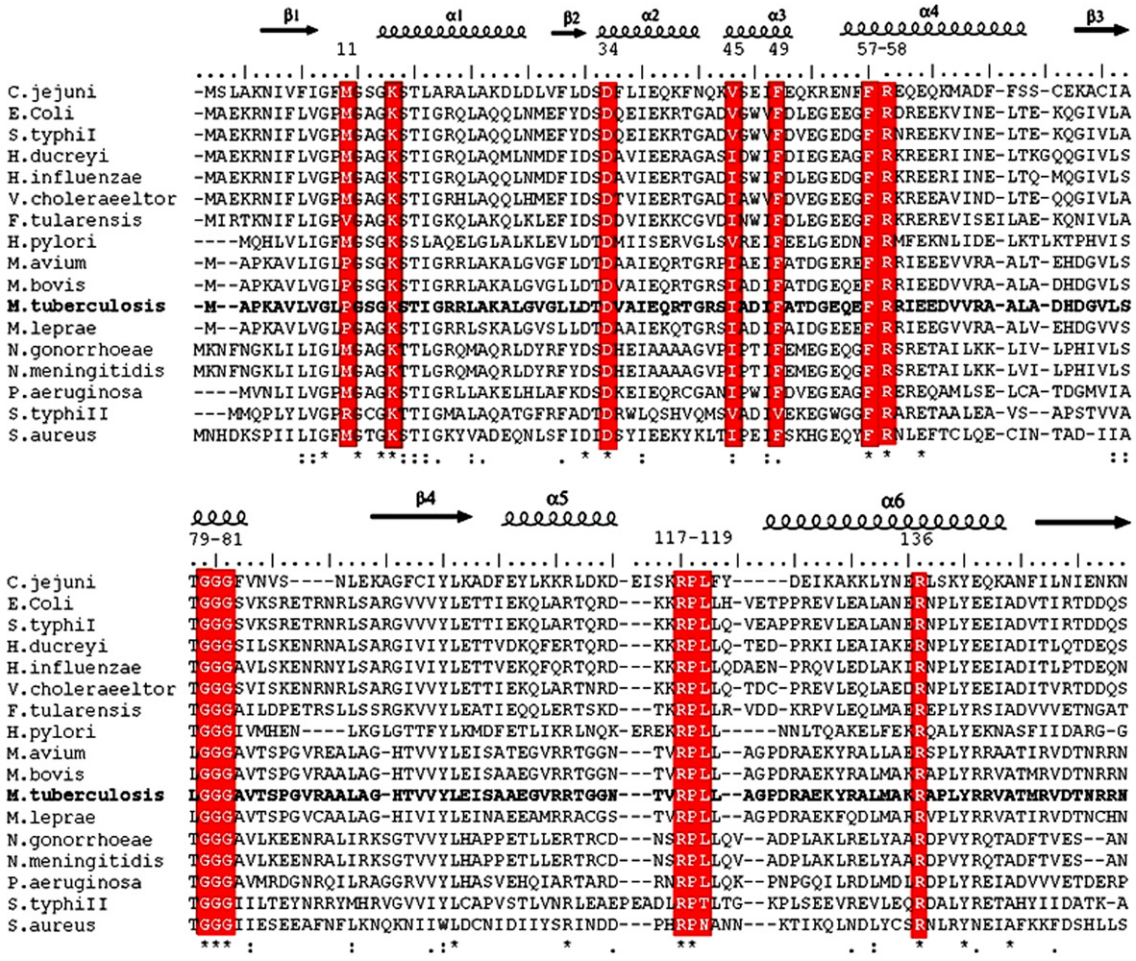


Figure 2. Structure-based sequence alignment of Sk sequences of Mt and other bacterial pathogens. Red boxes indicate the residues that interact with 200 top-scoring compounds, sequence of MtSk is highlighted in bold, and the secondary structural elements are shown above the sequences. The multiple sequence alignment was constructed using Expresso 3D-Coffee.[33]

dues, is shown in Table 2. Most interactions between these residues and the top compounds were driven mainly by hydrophobic forces. G80, R117, and L119 could also play a role as strong hydrogen bond donors as demonstrated elsewhere.[21] Similarly, G80 along with G79, and G81 (residues that have low frequency of interaction) conform an extended hydrogen bond donor patch. The K15 and D34 residues, although they do not interact frequently, along with R117 (a residue with high frequency of interaction) showed that they are key residues for the phosphoryl transfer reaction catalyzed by MtSk.[20,23,24]

The visual inspection of the docking poses of the 200 top-scoring compounds suggested that their triazole or tetrazole moieties were placed near the R58, L119, and R136 residues, and mimic key interactions of carboxyl group of the shikimate (Fig. 1) as reported elsewhere.[23] Two of the nitrogen atoms of the azole system were hydrogen bonded to R58 and R136, and the proximity of the L119 with the heteroaromatic ring suggested the possibility of stabilizing CH–pi electron interactions. These physicochemical features may explain why triazole or tetrazole moiety preferentially interacts with these residues.

A multiple alignment of Sk sequences from the members of tuberculosis complex and 11 bacterial pathogens (Fig. 2) reveals striking conservation at the residues that interact with 200 top-scoring compounds (e.g., the hydrophobic residues 45, 49, 57, 79, 80, 81, 118, 119, and charged residues 15, 34, 58, 117, 136). These residues are involved in the recognition and positioning of the shikimate substrate,[20,23,24] nevertheless *Mycobacteria* sequences bear some unique features; they differ from the others in the replacement of Pro11 mainly by methionine. Overall sequence identity between MtSk and the bacterial Sks is in the range from 25% to 100%. The high conservation in the above-mentioned residues shows that the Sks of some bacterial pathogens preserve similar interactions with the substrate. It should be kept in mind that Sk is absent in humans, thus, the compounds identified herein may also be relevant in the rational development of novel drugs specific against bacterial human pathogens of public health importance.

In conclusion, the structure-based virtual screening protocol described herein has proven to be useful for the discovery of potential inhibitors for MtSk. Although it will require further unbiased validation studies, a total of 644 potential hits with favorable drug-like properties were identified. Docking simulations, and validation training module in eHiTS, were used to rank, with confidence, the compounds that docked with high affinity to the receptor, and received the best score. Visual examination of the chemical match, binding modes, and interactions with the common amino acids, into the shikimate binding site from SkMt, were determined for 200 top hits. Structurally, the hits included mainly five-membered heteroaromatic systems (triazole or tetrazole). The triazole or tetrazole moieties identified were able to mimic key interactions involved in the complex enzyme–substrate, thus demonstrating their potential utility as powerful inhibitors of the MtSk.

These findings will contribute to the rational development and improvement of novel drugs against *M. tuberculosis*.

Acknowledgments

The authors thank SimBioSys Inc. for providing academic license of eHiTS 6.2® software, Arturo Rojo-Domínguez for critically reading the manuscript and very valuable comments, and Lilia Gómez-Caro for technical assistance. Mario A. Rodríguez-Pérez holds a scholarship from Comisión de Operación y Fomento de Actividades Académicas/Instituto Politécnico Nacional.

Supplementary data

Table 3 shows the 644 compounds, in a SMILES format, identified as potential MtSk inhibitors. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bmcl.2008.05.003.

References and notes

1. WHO. Anti-tuberculosis drug resistance in the world. Report No. 4. WHO, Geneva, Switzerland, 2008.
2. Cole, S. T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D. *Nature* **1998**, *393*, 537.
3. Lamichhane, G.; Zignol, M.; Blades, N. J.; Geiman, D. E.; Dougherty, A.; Grosset, J.; Broman, K. W.; Bishai, W. R. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 7213.
4. Lamichhane, G.; Tyagi, S.; Bishai, W. R. R. *Infect. Immun.* **2005**, *73*, 2533.
5. Sassetti, C. M.; Boyd, D. H.; Rubin, E. J. *Mol. Microbiol.* **2003**, *48*, 77.
6. Sassetti, C. M.; Rubin, E. J. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12989.
7. McKinney, J. D.; Höner zu Bentrup, K.; Muñoz-Elías, E. J.; Miczak, A.; Chen, B.; Chan, W. T.; Swenson, D.; Sacchettini, J. C.; Jacobs, W. R., Jr.; Russell, D. G. *Nature* **2000**, *406*, 735.
8. Glickman, M. S.; Cox, J. S.; Jacobs, W. R., Jr. *Mol. Cell* **2000**, *5*, 717.
9. Dahl, J. L.; Kraus, C. N.; Boshoff, H. I.; Doan, B.; Foley, K.; Avarbock, D.; Kaplan, G.; Mizrahi, V.; Rubin, H.; Barry, C. E., 3rd *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 10026.
10. Park, H. D.; Guinn, K. M.; Harrell, M. I.; Liao, R.; Voskuil, M. I.; Tompa, M.; Schoolnik, G. K.; Sherman, D. R. *Mol. Microbiol.* **2003**, *48*, 833.
11. Parish, T.; Stoker, N. G. *Microbiology* **2002**, *148*, 3069.
12. Kishore, G. M.; Shah, D. M. *Annu. Rev. Biochem.* **1988**, *57*, 627.
13. Roberts, F.; Roberts, C. W.; Johnson, J. J.; Kyle, D. E.; Krell, T.; Coggins, J. R.; Coombs, G. H.; Milhous, W. K.; Tzipori, S.; Ferguson, D. J.; Chakrabarti, D.; McLeod, R. *Nature* **1998**, *393*, 801.
14. Haslam, E. *Shikimic Acid: Metabolism and Metabolites*; John Wiley & Sons: Chichester, United Kingdom, 1993.
15. Herrmann, K. M.; Weaver, L. M. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1999**, *50*, 473.
16. Bentley, R. *Crit. Rev. Biochem. Mol. Biol.* **1990**, *25*, 307.
17. Davies, G. M.; Barrett-Bee, K. J.; Jude, D. A.; Lehan, M.; Nichols, W. W.; Pinder, P. E.; Thain, J. L.; Watkins, W. J.; Wilson, R. G. *Antimicrob. Agents Chemother.* **1994**, *38*, 403.
18. Berlyn, M. B.; Giles, N. H. *J. Bacteriol.* **1969**, *99*, 222.
19. Yan, H.; Ysai, M. D. *Adv. Enzymol. Relat. Areas Mol. Biol.* **1999**, *73*, 103.
20. Gu, Y.; Reshetnikova, L.; Li, Y.; Wu, Y.; Yan, H.; Singh, S.; Ji, X. J. *Mol. Biol.* **2002**, *319*, 779.
21. Dhaliwal, B.; Nichols, C. E.; Ren, J.; Lockyer, M.; Charles, I.; Hawkins, A. R.; Stammers, D. K. *FEBS Lett.* **2004**, *574*, 49.
22. Pereira, J. H.; de Oliveira, J. S.; Canduri, F.; Dias, M. V.; Palma, M. S.; Basso, L. A.; Santos, D. S.; de Azevedo, W. F., Jr. *Acta Crystallogr. D Biol. Crystallogr.* **2004**, *60*, 2310.
23. Gan, J.; Gu, Y.; Li, Y.; Yan, H.; Ji, X. *Biochemistry* **2006**, *45*, 8539.
24. Hartmann, M. D.; Bourenkov, G. P.; Oberschall, A.; Strizhov, N.; Bartunik, H. D. *J. Mol. Biol.* **2006**, *364*, 411.
25. Krell, T.; Maclean, J.; Boam, D. J.; Cooper, A.; Resmini, M.; Brocklehurst, K.; Kelly, S. M.; Price, N. C.; Lapthorn, A. J.; Coggins, J. R. *Protein Sci.* **2001**, *10*, 1137.
26. Zsoldos, Z.; Reid, D.; Simon, A.; Sadjad, S. B.; Johnson, A. P. *J. Mol. Graph. Model.* **2006**, *7*, 421. Available from: <http://www.simbiosys.ca/ehits/>.
27. Dias, M. V.; Faím, L. M.; Vasconcelos, I. B.; de Oliveira, J. S.; Basso, L. A.; Santos, D. S.; de Azevedo, W. F., Jr. *Acta Crystallogr. Sect. F Struct. Biol. Crystallogr. Commun.* **2007**, *63*, 1.
28. Miteva, M.; Violas, S.; Montes, M.; Gomez, D.; Tuffery, P.; Villoutreix, B. O. *Nucleic Acids Res.* **2006**, *34*, 738.
29. Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3.
30. Clark, A. M.; Labute, P. *J. Chem. Inf. Model.* **2007**, *47*, 1933.
31. Berman, H. M.; Battistuz, T.; Bhat, T. N.; Bluhm, W. F.; Bourne, P. E.; Burkhardt, K.; FENA, Z.; Gilliland, G. L.; Iype, L.; Jain, S.; Fagan, P.; Marvin, J.; Padilla, D.; Ravichandran, V.; Schneider, B.; Thanki, N.; Weissig, H.; Westbrook, J. D.; Zardecki, C. *Acta Crystallogr. D Biol. Crystallogr.* **2002**, *58*, 899.
32. Wallace, A. C.; Laskowski, R. A.; Thornton, J. M. *Protein Eng.* **1995**, *8*, 127.
33. Armougom, F.; Moretti, S.; Poirot, O.; Audic, S.; Dumas, P.; Schaeli, B.; Keduas, V.; Notredame, C. *Nucleic Acids Res.* **2006**, *34*, 604.
34. Raviglione, M.; Smith, I. *N. Engl. J. Med.* **2007**, *356*, 656.
35. Ramaswamy, S.; Musser, J. M. *Tuberculosis Lung Dis.* **1998**, *79*, 3.
36. Mouton, J. W. *Infection* **1999**, *27*, 24.
37. Rubinstein, E. *Infection* **1999**, *27*, 32.