

# We are IntechOpen, the world's leading publisher of Open Access books Built by scientists, for scientists

4,200

Open access books available

116,000

International authors and editors

125M

Downloads

Our authors are among the

154

Countries delivered to

TOP 1%

most cited scientists

12.2%

Contributors from top 500 universities



WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index  
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?  
Contact [book.department@intechopen.com](mailto:book.department@intechopen.com)

Numbers displayed above are based on latest data collected.  
For more information visit [www.intechopen.com](http://www.intechopen.com)



# Computational Deorphaning of *Mycobacterium tuberculosis* Targets

Lorraine Yamurai Bishi, Sundeep Chaitanya Vedithi,  
Tom L. Blundell and Grace Chitima Mugumbate

## Abstract

Tuberculosis (TB) continues to be a major health hazard worldwide due to the resurgence of drug discovery strains of *Mycobacterium tuberculosis* (*Mtb*) and co-infection. For decades drug discovery has concentrated on identifying ligands for ~10 *Mtb* targets, hence most of the identified essential proteins are not utilised in TB chemotherapy. Here computational techniques were used to identify ligands for the orphan *Mtb* proteins. These range from ligand-based and structure-based virtual screening modelling the proteome of the bacterium. Identification of ligands for most of the *Mtb* proteins will provide novel TB drugs and targets and hence address drug resistance, toxicity and the duration of TB treatment.

**Keywords:** *Mycobacterium tuberculosis*, target deorphaning, target deconvolution, proteome modelling, virtual screening

## 1. Introduction

Tuberculosis (TB) continues to be a major public health concern with over 2 billion people currently infected, 8.6 million new cases per year, and more than 1.3 million deaths annually [1]. The current drug-regimen combination for drug sensitive TB consists of isoniazid, rifampicin, ethambutol and pyrazinamide, administered over 6 months [2]. If this treatment fails, second-line drugs are used, such as para-aminosalicylate (PAS) and fluoroquinolones, which are usually either less effective or more toxic with serious side effects. Although this regimen has a high success rate, it is marred by compliance issues, which have resulted in the rise of multidrug resistant (MDR), extensively drug resistant (XDR) and totally drug resistant (TDR) strains of the causative agent, *Mycobacterium tuberculosis* (*Mtb*) [3, 4], in both immunocompetent and immunocompromised patients worldwide [5]. However, it took about 40 years for a new TB drug to be discovered and most of the current TB drugs target a total of only ~10 proteins, even though the complete genome of *Mtb* was published nearly 20 years ago [6]. Consequently, most of the essential proteins are orphans since their ligands are still to be identified. In our context, target deorphaning or deconvolution encompasses identification of ligands for *Mtb* proteins not currently exploited in TB chemotherapy and those of old TB targets. Targeting further essential proteins should allow the fight against drug resistance to be enhanced, and possibly lead to a reduction in the duration of TB treatment.



















## 6.2 Comparative 3D modelling of proteins

Comparative modelling proteins, based on the fold recognition and structural alignment with the closest homologues that have experimentally solved structures, began using interactive graphics in the 1970s [41–43]. The development of automated modelling software began in the 1980s, initially with Composer [44] and later developed with Comparer [45] and Modeller [46], based on satisfaction of 3D restraints derived from structurally aligned homologues. Modeller has now been cited ~10,500 times in the literature!

### 6.2.1 Computational modelling pipelines and structural proteome databases

Rapid progress in this and other related software coupled with increasing computing power has enabled genome scale prediction of protein structures, as a viable alternative to experimental determination. In order to construct computational models of all gene products, which we here refer to as the structural proteome, we identify templates by a sequence-structure homology search using Fugue [47], which uses local-structural-environment-specific substitution tables to predict the likelihood of a common 3D structure. We have incorporated Fugue into a pipeline (Vivace), in which templates are selected from TOCCATA (Ochoa Montañó and Blundell, unpublished), a database of consensus profiles built from CATH 3.5 [48] and SCOP 1.75A [49] based classification of proteins structures (PDB files). PDBs within each profile are clustered based on sequence similarity using CD-HIT [50] and structures are aligned using BATON, a modified version of COMPARE [45]. After further optimization of the clusters by discarding templates with more than 20% difference in sequence identity to the maximum hit, remaining templates are classified into states based on ligand binding and oligomerization. Five different states, known as “liganded-monomeric,” “liganded-complexed,” “apo-monomeric,” “apo-complexed” and “any,” are generated in each profile hit. Models are built in each of these states using Modeller 9.10 [46] and refined. Later NDOPE, GA341 [51] Molprobitry [52] and SSAG [53] are used to determine the quality of the models.

### 6.2.2 Mycobacterial proteome databases

The first application of this approach was to construct the Chopin Database (<http://mordred.bioc.cam.ac.uk/chopin/about>), a database of protein structures for H37Rv strain of *Mtb*. This has provided structures that are reasonably certain for around 65% of gene products. These have proved reliable indicators of the overall structures but may have some uncertainties especially in loop regions and domain-domain relationships. A further ~19% probably have correct folds while the remaining would unlikely to be correct. Nevertheless, compared to those structures defined experimentally by X-ray analysis, this represents a 6-fold increase of structural information available that might be useful in assessing druggability and the impacts of mutations.

Similar models of the structural proteome for *M. abscessus* (Skwark et al., unpublished) and *M. leprae* (Vedithi et al., unpublished) have been developed in the group. In *M. leprae*, of the 1615 gene products, templates were identified for 1429 gene products and we were able to model 1161 proteins with high confidence. A total of 36,408 models were built in different ligand bound and oligomeric states for the 1161 proteins. The distribution of Fugue Z score across models indicates that only 4% of the proteome has no hits and 15% has poor scores. ~80% of the proteome has acceptable and good hits, and the corresponding Z scores. Around 47% of the protein queries identified templates with identity and coverage greater than 40 and 67% of the models in the proteome are of best quality as estimated by NDOPE, GA341, Molprobitry and Secondary Structure Agreement (SSAG).

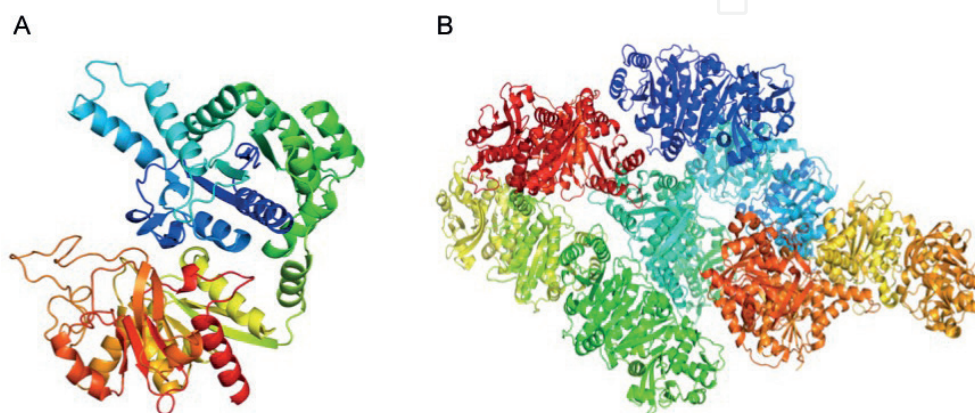
### 6.2.3 Oligomeric protein models

Current work on structural proteomes includes efforts to extend the modelling pipeline to homo-oligomeric (and eventually hetero-oligomeric) structures using comparative approaches (Malhotra et al., unpublished), extending models and improving models of small molecule complexes, and linking individual protein structures into the metabolic networks and interactions in the cell (Bannerman et al., unpublished). An example of an oligomeric structure is CTP-synthase, encoded by *PyrG*, which is an essential gene in *Mtb* identified by transposon saturation mutagenesis [54] and catalyses ATP-dependent amination of UTP to CTP with either L-glutamine or ammonia. The allosteric effector GTP functions by stabilising the protein conformation that binds to the tetrahedral intermediates formed during glutamine hydrolysis. Its closest homologue in *M. leprae* ML1363 is a target of choice and was modelled using *Vivace* during the proteome modelling exercise. We modelled the apomeric and ligand bound states of the model and oligomerized the protomer using our inhouse oligomerization pipeline. The protomeric and oligomeric states are depicted in **Figure 3A** and **B**.

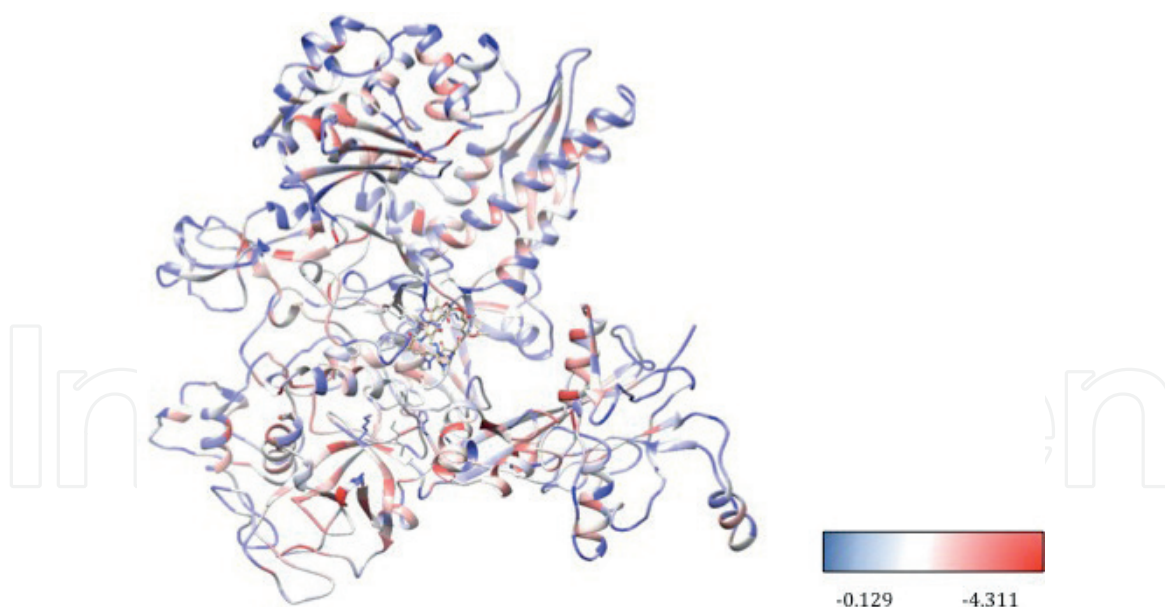
The models were built by using templates PDB-IDs: 4zdI and 4zdK for *PyrG* of *Mtb* [55]. Both the templates are 89% identical and 100% coverage to the query sequence. The superposition of the models with the templates indicated a root mean square deviation (RMSD) of 0.758.

### 6.3 Structural implications of mutations

We have also spent time over 2 decades analysing the impacts of mutations evident in the increasing wealth of available genome sequences for pathogenic mycobacteria and cancers. We originally developed SDM [56] in 1997, a method depending on statistical analysis of environment-dependent amino-acid substitution tables [57, 58]. In 2013 machine learning was introduced with the arrival of Douglas Pires in Cambridge, developing first mCSM for stability [59] followed by several “flavours” including mCSM-PPI for impacts on protein-protein interactions, mCSM-NA [60] for nucleic acid interactions and mCSM-lig for impacts on small-molecule ligand interactions useful for understanding drug resistance [61]. A critical part of using machine learning is to have an extensive database of experimentally-defined impacts of mutations on stability and interactions, such as Platinum by David Ascher when in Cambridge [62], a database of experimentally measured effects of mutations on structurally defined protein-ligand complexes that was developed for mCSM-lig. These two structural approaches to predicting the impacts of mutations (SDM & mCSM) have proved complementary and more reliable than most sequence-only



**Figure 3.**  
(A) Protomeric model of *PyrG* (CTP-Synthase) of *M. leprae* modelled with a quality of 4.25 (best).  
(B) Homo-8-mer of *PyrG* of *M. leprae* modelled with a quality of 4.25 (best).



**Figure 4.**

Indicates the maximum destabilising effect a mutation can induce on the stability of RNA-polymerase  $\beta$ -subunit of *M. leprae* (target for rifampin) measured by mCSM-stability.

methods. They also allow the application of saturation mutagenesis, facilitating *in silico* systematic analysis of mutations [63], an approach now being adopted to whole proteomes where every residue in each of the proteins in the proteome is mutated to all the other 19 amino acids and the effects of the mutations are measured using various methods mentioned above. In structure-guided fragment-based drug discovery, this provides comprehensive information on the regions of the protein that are less likely to lead to drug resistance and therefore can be probed by elaboration of fragments/small molecules. We performed saturation mutagenesis on the drug targets in *M. leprae* for leprosy and the average or highest impact a mutation can induce in each residue position is depicted on the structure (**Figure 4**).

#### 6.4 Active sites, cavities and fragment hotspot maps

Although comparative modelling of homologues in complex with ligands can often give clues about active sites, cofactor binding and substrate or other ligand binding sites, this is not always possible. In order to indicate putative binding sites in the absence of appropriate experimental data, we have exploited cavity-defining software such as VolSite [64] for novel binding site description together with an alignment and comparison tool (Shaper) [65]. We have used FuzCav, a novel alignment-free high-throughput algorithm to compute pairwise similarities between protein-ligand binding sites [66] and GHECOM [67], to study the small pockets that often characterise protein-protein and protein-peptide interactions.

Further to the identification of cavities and pockets, it is also useful to be able to identify hotspots, region(s) of the binding site defined as a major contributor to the binding free energy, and often characterised by their ability to bind fragment-sized organic molecules in well-defined orientations. The usual understanding is that the fragment, with a mixed polar and hydrophobic character, can displace an “unhappy water.” We have tried to mimic this *in silico* by using SuperStar [68] to generate atomic interaction propensities on a grid. We then carry out a search with three fragments, each having a six-membered carbon ring, but having a donor, acceptor or a non-polar substituent. The resulting map is convoluted with an estimate of the depth below the surface, which generally appears to correlate with favourable entropic gain on water release on binding of a ligand [69]. The hotspot maps, computed in this way and

indicating donor, acceptor and lipophilic interactions correlate well with experimental binding sites of fragments that can be elaborated in fragment-based discovery. For the ligand bound structures, lower contouring can provide “warm spots” for the binding sites, indicating possibilities for elaborating the fragment in the binding pocket.

The models of individual molecules of the modelled proteome can be individually decorated with the hotspot maps. They give a good indication of the known functional sites on experimentally defined structures of proteins, often demonstrating that a functional site comprises several hotspots involved in binding substrates and cofactors. They also provide a good indication of the location of allosteric sites [70].

## 7. Conclusion

In summary we can move from the study of individual targets to an understanding of the majority of targets coded by the genome. Indeed, we can build 3D structures for a majority of the genes, so providing a model of the “structural proteome”. Hotspots and cavities provide a basis for identification of the ligandability of putative binding sites and have been used in our group to predict pharmacophores that can be used in docking and virtual screening and so deorphaning of mycobacterial proteins.

To identify druggable proteins from the structural proteome, we have adopted a hierarchal selection process wherein chokepoint analysis is initially performed to identify metabolic reactions that are critical to cell survival. Gene products identified in this screen are later subjected to essentiality analysis using either flux balance analysis (FBA) based models or by data from the transposon saturation mutagenesis experiments in the literature. Genes that are essential are chosen at this stage and understanding of the gene expression profiles in different growth conditions is analysed. Genes whose expression is condition specific are excluded. Later for the selected genes, the structural information of the corresponding proteins is analysed in the context of prior knowledge and attempts in drug discovery, druggable pockets and fragment hotspots maps, small molecule bound states, non-human homologue, non-homologous to human microbiome, cellular localization and biochemical properties of the proteins. Structure-guided virtual screening is performed on the selected drug targets with a choice of fragment and compound libraries using CCDC Gold (The Cambridge Crystallographic Data Centre) [71]. Best poses with good scores lead the experimental process of structure-guided fragment-based drug discovery.

The challenge now is to test the computational methods outlined here for identifying ligands and understanding the druggability of the proteome—several thousand gene products from the whole genome of *Mtb*. We can then begin to assess the degree to which we can de-orphan the many *Mtb* proteins that have until now not featured as targets in the worldwide efforts to combat the global challenge of TB to the health and well-being of human kind.

## Acknowledgements

LYB and GCM are grateful to Chinhoyi University of Technology for their support in introducing computational drug discovery and development research work at the University and all our collaborators. TLB and SCV thank the Gates Foundation, the Cystic Fibrosis Trust and the American Leprosy Mission for their funding of computational and experimental work on approaches to combating disease from mycobacterial infections. They also thank colleagues in Cambridge and elsewhere who have contributed over the years to our efforts to develop new approaches to structural biology, computational bioinformatics and drug discovery.

IntechOpen

### **Author details**


Lorraine Yamurai Bishi<sup>1</sup>, Sundeep Chaitanya Vedithi<sup>2</sup>, Tom L. Blundell<sup>2</sup>  
and Grace Chitima Mugumbate<sup>1\*</sup>

1 Department of Chemistry, School of Natural Sciences and Mathematics, Chinhoyi University of Technology, Chinhoyi, Zimbabwe

2 Department of Biochemistry, University of Cambridge, Cambridge, United Kingdom

\*Address all correspondence to: gmugumbate@cut.ac.zw

### **IntechOpen**

© 2019 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

## References

- [1] Nguta JM, Appiah-Opong R, Nyarko AK, Yeboah-Manu D, Addo PGA. Medicinal plants used to treat TB in Ghana. *International Journal of Mycobacteriology*. 2015;4:116-123. DOI: 10.1016/j.ijmyco.2015.02.003
- [2] Jiang J, Gu J, Zhang L, Zhang C, Deng X, Dou T, et al. Comparing *Mycobacterium tuberculosis* genomes using genome topology networks. *BMC Genomics*. 2015;16:1-10. DOI: 10.1186/s12864-015-1259-0
- [3] Shah I. Drug Resistant Tuberculosis Children in India. *Pediatric Oncall Journal*. 2012;9(5). DOI: 10.7199/ped.oncall.2012.27
- [4] Nguta JM, Appiah-Opong R, Nyarko AK, Yeboah-Manu D, Addo PGA. Current perspectives in drug discovery against tuberculosis from natural products. *International Journal of Mycobacteriology*. 2015;4:165-183. DOI: 10.1016/j.ijmyco.2015.05.004
- [5] Janbaz KH, Qadir MI, Ahmad B, Sarwar A, Yaqoob N, Masood MI. Tuberculosis burning issues: Multidrug resistance and HIV-coinfection. *Critical Reviews in Microbiology*. 2012;38:267-275. DOI: 10.3109/1040841X.2012.664539
- [6] Anishetty S, Pulimi M, Pennathur G. Potential drug targets in *Mycobacterium tuberculosis* through metabolic pathway analysis. *Computational Biology and Chemistry*. 2005;29:368-378. DOI: 10.1016/j.compbiolchem.2005.07.001
- [7] Hart CP. Finding the target after screening the phenotype. *Drug Discovery Today*. 2005;10:513-519. DOI: 10.1016/S1359-6446(05)03415-X
- [8] Mugumbate G, Abrahams KA, Cox JAG, Papadatos G, van Westen G, Lelièvre J, et al. Mycobacterial dihydrofolate reductase inhibitors identified using chemogenomic methods and in vitro validation. *PLoS One*. 2015;10:e0121492. DOI: 10.1371/journal.pone.0121492
- [9] Bajorath J. Computer-aided drug discovery. *F1000 Research*. 2015;4:630. DOI: 10.12688/f1000research.6653.1
- [10] Bender A, Young D, Jenkins J, Serrano M, Mikhailov D, Clemons P, et al. Chemogenomic data analysis: Prediction of small-molecule targets and the advent of biological fingerprints. *Combinatorial Chemistry & High Throughput Screening*. 2007;10:719-731. DOI: 10.2174/138620707782507313
- [11] Mugumbate G, Mendes V, Blaszczyk M, Sabbah M, Papadatos G, Lelievre J, et al. Target identification of *Mycobacterium tuberculosis* phenotypic hits using a concerted chemogenomic, biophysical, and structural approach. *Frontiers in Pharmacology*. 2017;8:681. DOI: 10.3389/fphar.2017.00681
- [12] Cox JAG, Mugumbate G, Del Peral LVG, Jankute M, Abrahams KA, Jarvis P, et al. Novel inhibitors of *Mycobacterium tuberculosis* GvaB2 identified by a target based high-throughput phenotypic screen. *Scientific Reports*. 2016;6:1-10. DOI: 10.1038/srep38986
- [13] Mendes V, Blundell TL. Targeting tuberculosis using drug design. *Drug Discovery Today*. 2016;00:1-9. DOI: 10.1016/j.drudis.2016.10.003
- [14] Cole RBST, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, et al. Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature*. 1998;393:537-544. DOI: 10.1038/31159
- [15] Smith I. *Mycobacterium tuberculosis* pathogenesis and molecular determinants of virulence. *Clinical*



- Microbiology Reviews. 2003;**16**:463-496. DOI: 10.1128/CMR.16.3.463
- [16] Hoagland DT, Liu J, Lee RB, Lee RE. New agents for the treatment of drug-resistant *Mycobacterium tuberculosis* ☆. Advanced Drug Delivery Reviews. 2016;**102**:55-72. DOI: 10.1016/j.addr.2016.04.026
- [17] World Health Organisation. Global tuberculosis Report 2017. Geneva; 2017
- [18] Zhang Y, Post-Martens K, Denkin S. New drug candidates and therapeutic targets for tuberculosis therapy. Drug Discovery Today. 2006;**11**:21-27. DOI: 10.1016/S1359-6446(05)03626-3
- [19] Pawlowski A, Jansson M, Sköld M, Rottenberg ME, Källenius G. Tuberculosis and HIV co-infection. PLoS Pathogens. 2012;**8**(2):e1002464. <https://doi.org/10.1371/journal.ppat.1002464>
- [20] Metcalfe JZ, Porco TC, Westenhouse J, Damesyn M, Facer M, Hill J, et al. Tuberculosis and HIV co-infection, California, USA, 1993-2008. Emerging Infectious Diseases. 2013;**19**:400-406. DOI: 10.3201/eid1903.121521
- [21] Worley MV, Estrada SJ. Bedaquiline: A novel antitubercular agent for the treatment of multidrug-resistant tuberculosis. The Journal of Human Pharmacology and Drug Therapy. 2014;**34**(11):1187-1197
- [22] Gawad J, Bonde C. Current affairs, future perspectives of tuberculosis and antitubercular agents. The Indian Journal of Tuberculosis. 2018;**65**:15-22. DOI: 10.1016/j.ijtb.2017.08.011
- [23] Lamichhane G. Novel targets in M. tuberculosis: Search for new drugs. Trends in Molecular Medicine. 2011;**17**:25-33. DOI: 10.1016/j.molmed.2010.10.004
- [24] Lewis K. Platforms for antibiotic discovery. Nature Reviews. Drug Discovery. 2013;**12**:371-387. DOI: 10.1038/nrd3975
- [25] Kaneko T, Cooper C, Mdluli K. Challenges and opportunities in developing novel drugs for TB. Future Medicinal Chemistry. 2011;**3**:1373-1400. DOI: 10.4155/fmc.11.115
- [26] Campaniço A, Moreira R, Lopes F. Drug discovery in tuberculosis. New drug targets and antimycobacterial agents. European Journal of Medicinal Chemistry. 2018;**150**:525-545. DOI: 10.1016/j.ejmech.2018.03.020
- [27] Chiliza TE, Pillay M, Pillay B. Identification of unique essential proteins from a *Mycobacterium tuberculosis* F15/LAM4/KZN phage secretome library. Pathogens and Disease. 2017;**75**:1-10. DOI: 10.1093/femspd/ftx001
- [28] Kapetanovic IM. Computer-aided drug discovery and development (CADD): In silico-chemico-biological approach. Chemico-Biological Interactions. 2008;**171**:165-176. DOI: 10.1016/j.cbi.2006.12.006
- [29] Machado E, Cerdeira C, de Miranda AB, Catanho M. Web resources on tuberculosis: Information, research, and data analysis. In: Mycobacterium-research and development. IntechOpen
- [30] Gaulton A, Bellis LJ, Bento AP, Chambers J, Davies M, Hersey A, et al. {ChEMBL}: A large-scale bioactivity database for drug discovery. Nucleic Acids Research. 2012;**40**:D1100-D1107. DOI: 10.1093/nar/gkr777
- [31] Malhotra S, Mugumbate G, Blundell TL, Higuero AP. TIBLE: A web-based, freely accessible resource for small-molecule binding data for mycobacterial species. Database (Oxford). 2017;**2017**:1-7. DOI: 10.1093/database/bax041
- [32] Mugumbate G, Newton AS, Rosenthal PJ, Gut J, Moreira R, Chibale K, et al.

Novel anti-plasmodial hits identified by virtual screening of the ZINC database. *Journal of Computer-Aided Molecular Design*. 2013;**27**:859-871. DOI: 10.1007/s10822-013-9685-z

[33] Shoichet BK. Virtual screening of chemical libraries. *Nature*. 2004;**432**:862-865. DOI: 10.1038/nature03197

[34] Ferreira LG, Dos Santos RN, Oliva G, Andricopulo AD. Molecular docking and structure-based drug design strategies. 2015. DOI: 10.3390/molecules200713384

[35] Sastry GM, Adzhigirey M, Day T, Annabhimoju R, Sherman W. Protein and ligand preparation: Parameters, protocols, and influence on virtual screening enrichments. *Journal of Computer-Aided Molecular Design*. 2013;**27**:221-234. DOI: 10.1007/s10822-013-9644-8

[36] Morris G, Huey R. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry*. 2009;**30**:2785-2791. DOI: 10.1002/jcc.21256.AutoDock4

[37] Friesner RA, Banks JL, Murphy RB, Halgren TA, Klicic JJ, Mainz DT, et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *Journal of Medicinal Chemistry*. 2004;**47**:1739-1749. DOI: 10.1021/jm0306430

[38] Jubb HC, Higuero AP, Ochoa-Montaño B, Pitt WR, Ascher DB, Blundell TL. Arpeggio: A web server for calculating and visualising interatomic interactions in protein structures. *Journal of Molecular Biology*. 2017;**429**:365-371. DOI: 10.1016/j.jmb.2016.12.004

[39] Colwill K, Renewable Protein Binder Working Group, Gräslund S.

A roadmap to generate renewable protein binders to the human proteome. *Nature Methods*. 2011;**8**:551-558. DOI: 10.1038/nmeth.1607

[40] Chim N, Habel JE, Johnston JM, Krieger I, Miallau L, Sankaranarayanan R, et al. The TB structural genomics consortium: A decade of progress. *Tuberculosis*. 2011;**91**:155-172. DOI: 10.1016/j.tube.2010.11.009

[41] Bedarkar S, Blundell TL, Dockerill S, Tickle IJ, Wood SP. Polypeptide hormone-receptor interactions: The structure and receptor binding of insulin and glucagon. In: *Molecular interactions and activity in proteins*. Amsterdam, Oxford, New York: Excerpta Medica. 1978;**60**:105

[42] Blundell TL, Sibanda BL, Sternberg MJE, Thornton JM. Knowledge-based prediction of protein structures and the design of novel molecules. *Nature*. 1987;**326**:347-352. DOI: 10.1038/326347a0

[43] Blundell T, Sibanda BL, Pearl L. Three-dimensional structure, specificity and catalytic mechanism of renin. *Nature*. 1983;**304**:273-275. DOI: 10.1038/304273a0

[44] Sutcliffe MJ, Haneef I, Carney D, Blundell TL. Knowledge based modelling of homologous proteins, part I: Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Engineering, Design & Selection*. 1987;**1**:377-384. DOI: 10.1093/protein/1.5.377

[45] Šali A, Blundell TL. Definition of general topological equivalence in protein structures. A procedure involving comparison of properties and relationships through simulated annealing and dynamic programming. *Journal of Molecular Biology*. 1990;**212**:403-428. DOI: 10.1016/0022-2836(90)90134-8