J Mol Model (2005) 11: 160–166
DOI 10.1007/s00894-005-0240-2

ORIGINAL PAPER

Nelson José Freitas da Silveira · Hugo Brandão Uchôa
José Henrique Pereira · Fernanda Canduri
Luiz Augusto Basso · Mário Sérgio Palma
Diógenes Santiago Santos
Walter Filgueira de Azevedo Jr.

Molecular models of protein targets from *Mycobacterium tuberculosis*

Received: 29 July 2004 / Accepted: 2 December 2004 / Published online: 10 March 2005

Abstract Structural characterization of enzymes that belong to microbial metabolic pathways is very important for structure-based drug design since some of these proteins may be present in the bacterial genome, but absent in humans. Thus, metabolic pathways became potential targets for drug design. The motivation of this work is the fact that *Mycobacterium tuberculosis* is the cause of the deaths of millions of people in the world, so that the structural characterization of protein targets to propose new drugs has become essential. DBMODELING is a relational database, created to highlight the importance of methods of molecular modeling applied to the *Mycobacterium tuberculosis* genome with the aim of proposing protein-ligand docking analysis. There are currently more than 300 models for proteins from *Mycobacterium tuberculosis* genome in the database. The database contains a detailed description of the reaction catalyzed by each enzyme and their atomic coordinates. Information about structures, a tool for animated gif image, a table with a specification of the metabolic pathway, modeled protein, inputs used in modeling, and analysis methods used in this project are available in the database for download. The search tool can be used for reseachers to find specific pathways or enzymes.

Keywords Bioinformatics · Mycobacterium tuberculosis · Drug design · Structural database

N. J. F. Silveira · H. B. Uchôa · J. H. Pereira · F. Canduri
W. F. de Azevedo Jr. (✉)
Programa de Pós-Graduação em Biofísica Molecular
Departamento de Física, UNESP,
São José do Rio Preto, SP, 15054-000, Brasil
E-mail: walterfa@ibilce.unesp.br
Tel.: +017-221-2463
Fax: +017-221-2247

L. A. Basso
Rede Brasileira de Pesquisas em Tuberculose,
Departamento de Biologia Molecular e Biotecnologia,
UFRGS, Porto Alegre, RS, 91501-970, Brasil

M. S. Palma
Departamento de Biologia, Instituto de Biociências
Laboratorio de Biologia Estrutural e Zooquímica,
UNESP, Rio Claro, SP, 13506-900, Brasil

F. Canduri · M. S. Palma · W. F. de Azevedo Jr.
Centro para Toxinologia Aplicada, Instituto Butantan,
São Paulo, SP, 05503-900, Brasil

D. S. Santos
Centro para Pesquisa e Desenvolvimento em Biologia
Molecular Estrutural e Funcional, PUCRS,
Porto Alegre, RS, 90619-900, Brasil

Introduction

Despite the availability of effective short-course chemotherapy (DOTS) and the Bacille Calmette-Guérin (BCG) vaccine, the tubercle bacillus continues to claim more lives than any other single infectious agent [1]. Recent years have seen increased incidence of tuberculosis in both developing and industrialized countries, the widespread emergence of drug-resistant strains and a deadly synergy with the human immunodeficiency virus (HIV). In 1993, the gravity of the situation led the World Health Organization (WHO) to declare tuberculosis a global emergency in an attempt to heighten public and political awareness [2]. The combination of genomics and bioinformatics has the potential to generate the information and knowledge that will enable the conception and development of new therapies and interventions needed to treat this airborne disease and to elucidate the unusual biology of its aetiological agent, *Mycobacterium tuberculosis*.

The treatment of tuberculosis is a special problem in the field of chemotherapy. Many of the drugs employed to treat the disease are used only for treating infections caused by mycobacteria. Treatment of the active case of *Mycobacterium tuberculosis* always includes simultaneous therapy with two or more of the frontline drugs: isoniazid, ethambutol, rifampicin and streptomycin, which are used to decrease the rate of emergence of resistant strains as well to increase the antibacterial ef-

fect[3, 4, 5]. Recent outbreaks of tuberculosis caused by multidrug-resistant (MDR) strains, mainly in individuals infected with HIV, have added a scaring element to the scenario, and also created a wordwide interest in expanding current programs for development of new drugs of different kinds to the existing ones, and based on the principle of selective toxicity on enzymes or structure of the bacillus, both to treat *Mt* strains resistant to existing drugs, and shorten the duration of treatment to improve patient compliance [6]. Because of the importance of the *Mt* genome, large-scale molecular modeling became necessary to model new structures for drug development. DBMODELING is a relational database of *Mt*, which contains models for segments of more than 300 proteins identified in the *Mt* genome (Table 1). The main interest in the study of metabolic pathways of some infectious agents is the fact that some of these pathways may not be present in humans, making them selective targets for drug design, decreasing impact of symptoms of drugs in the organism. This identification and characterization of extreme importance for drug development in the future.

One of the most important applications of molecular modeling techniques in structural biology is docking of a ligand molecule to a receptor, such as a protein [7, 8]. If the structure of the receptor is known, the application is essentially one of structure-based drug design. These methods have a number of related aims; they often seek to identify the location of the ligand-binding site and perhaps the geometry of the ligand in the active site. Another goal is the correct ranking of a series of related ligands in terms of their affinity, or to evaluate the absolute binding free energy as accurately as posssible [9].

DBMODELING is a bioinformatics database that describes metabolic pathaways and their component reactions, enzymes, and substrates [10]. Furthermore, since we aimed to use this database for drug design, special attention is devoted to potential protein targets.

Methods

Molecular modeling

Molecular modeling methods, first reported by Browne et al. [11] are able to predict the 3D structure of a protein sequence by using information derived from homologous protein of known structure [12, 13]. The utility of homology methods is evident when considering the vast numbers of the open reading frames (ORFs), which are potential protein coding sequences, produced as a result of genome sequencing projects. It has been estimated that of the order of 20—30% of these open reading frames can be assigned to a fold classification derived from structures in the PDB [14, 15].

If a 3D model of the protein of interest can be derived, it may be usable as the basis for a structure-based drug-design study. In addition, such models can be useful aid to the rational design of experiments such as site directed mutagenesis or in understanding protein stability and function.

For modeling of proteins identified in the genome from *Mt*, we used restrained-based modeling implemented in the program MODELLER [12], which was parallelized by the program Parmodel [16] on a Beowulf cluster with 16 nodes (AMD Athlon 2100+: BioComp. São José do Rio Preto, SP, Brazil). Parmodel was used to model the *Mt* genome. It controls the distribution of the MODELLER jobs on the Beowulf cluster using a library from C language, message passing interface (MPI). It allows parallel MODELLER execution and decreases the processing time of the modeling process.

Table 1 Contents of DBMODELING for the Mycobacterium tuberculosis genome

Mycobacterium tuberculosis	
Number of ORFs	3,924
Number of related structures	319
Number of solved protein structures	17
Number of modeled protein structures	302
Number of proteins with sterechemical quality worse than 85% (Percentage of residues in the allowed regions of Ramachandran plot)	8
Number of pathways	102
Number of enzymes	294

3D structure analysis

Difficult cases in homology modeling correspond to protein sequences that only posses distant homologues of known structure, where the level of sequence may be low. In such cases incorrect alignment can lead to regions of a model that have significant structural errors [9].

The quality of the predicted model determines the information that can be extracted from it. Thus, estimating the accuracy of 3D protein models is essential for interpreting them. The model can be evaluated as a whole as well as in the individual regions. There are many model evaluation programs and servers (Table 2) [17, 18].

A basic requirement for a model is to have good stereochemistry. Some useful programs for evaluating stereochemistry are PROCHECK [17, 19] and WHATCHECK [20]. The features of a model that are checked by these programs include bond lengths, bond angles, peptide bonds and side-chain ring planarities, chirality, main-chain and side-chain torsion angles, and clashes between non-bonded pair of atoms. The G-factor is essentially just log-odds score based on the observed distributions of the stereochemical parameters. There are also methods to test 3D models that take into account many spatial features compiled from high resolution protein structures implicitly. These methods are based on 3D profiles [21, 22] and statistical potentials of mean force [23, 24]. Programs implementing this approach include VERIFY3D [24], which measures the compatibility

Table 2 Programs and web servers useful in the evaluation of comparative modeling

Model valuation		
Name	Type	World Wide Web address
ANOLEA	S	http://www.fundp.ac.be/sciences/biologie/bms/CGI/anolea.html
AQUA	P	urchin.bmrb.wisc.edu/~jurgen/aqua/
BIOTECH	S	biotech.embl-heidelberg.de:8400
ERRAT	S	http://www.doe-mbi.ucla.edu/Services/ERRAT/
PROCHECK[a]	P	http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html
ProsaII	P	http://www.came.sbg.ac.at
PROVE	S	http://www.ucmb.ulb.ac.be/UCMB/PROVE
SQUID	P	http://www.ysbl.york.ac.uk/~oldfield/squid
VERIFY3D[a]	S	http://www.doe-mbi.ucla.edu/Services/Verify_3D/
WHATCHECK[a]	P	http://www.sander.embl-heidelberg.de/whatcheck/

S server; *P* program
[a]Used in evaluation of generated models by DBMODELING.

of a protein model with its sequence using a 3D profile [25, 26]. All models were evaluated by the above programs and the RMSD from ideal geometry extracted for each model with the program X-PLOR [27].

The purpose of PROCHECK and WHATCHECK is to (a) verify the syntax of the file, (b) check the consistency of an atomic model with the current library and identify outliers for further investigation, (c) detect gross errors in the structures, such as mistracing of the chain, (d) check for local abnormalities of stereochemistry, and (e) procedure global stereochemical quality criteria [28]. The atomic coordinates of the investigated enzyme are required to use VERIFY3D and as result it generates a profile window plot for the enzyme with the 3D-1D score for each amino acid, the 3D profile score S for its amino acid sequence, and the ideal score S_{ideal}, calculated from the length of the enzyme. These programs were used to assess the quality of the available models and can be performed by any researcher in the DBMODELING web page for each enzyme.

Database design

A MySQL (http://www.mysql.com) database based on the relational database management system (RDBMS) was developed to archive protein structures identified in the *Mt* genome. All supporting data related to the 3D structure modeling, such as protein name, atomic coordinates in PDB format from modeled proteins, fasta sequence, metabolic pathways, links to others databases and information about the protein were arranged in the MySQL [29] database under a master table. Four separate tables in MySQL were developed to contain path information for the image directory and text data related to the modeled protein, and their analysis for all structures. The database was queried from an html client using a Perl-CGI program, quick display of the records as dynamically generated web pages in different frames. Records were positioned, edited, retrieved, and displayed with various algorithms that were written in Perl [30]. The search engine was developed to display results of searches dynamically based on relational information from the database about 3D structures using input data such as keywords pertinent to the metabolic pathways or specific proteins. The design of the database tables is shown in Fig. 1. Due to the importance of protein modeling, several databases were created, such as MODBASE [31, 13] (http://salilab.org/modbase), GTOP [32] (http://spock.genes.nig.ac.jp/~genome/gtop.html) and others. However, the present database offers tools not available in the previously cited databases, such as structural quality analysis of the models, animated figures for the models, and a more robust modeling protocol (each model was chosen from a set of 1,000 models). The building of several models enhances the chance of obtaining good structures because the changes in the initial spatial restraints will generate different conformations in the final structures. Since several models are calculated for the same alignment, the best structure can be selected by picking that with the lowest value of the MODELLER objective function and stereochemical quality. The value of the objective function in MODELLER is not an absolute measure in the sense that it can only be used to rank models calculated from the same alignment[12].

Database contents and update

DBMODELING will be growing the number of organisms and metabolic pathways, once new templates become

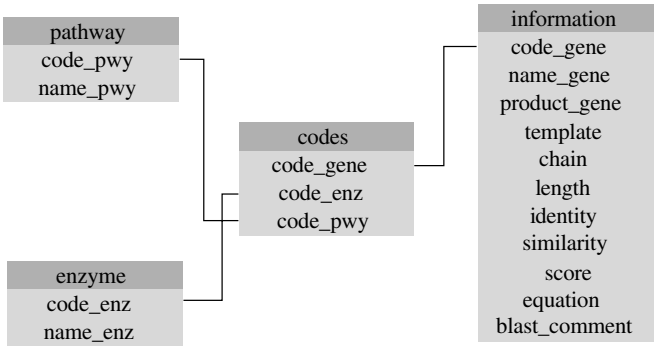


Fig. 1 This diagram describes the entity relationships for database tables. Each table is related to one code (code pwy, code enz and, code gene). For each selected pathway there is only one code pwy, which will select their non-redundant enzymes linking to the information results about each enzyme. The diagram reports the data set presentation in the web interface without redundancy

available, new models will be built. Several tests were performed with the contents of DBMODELING. The fingerprints of database results are shown in Fig. 1 and 2a, b. The current form of DBMODELING shows information about the statistical modeling such as the number of genome sequences, proteins, stereochemical restraints (proteins with $\leq 85\%$ of residues in allowed regions of Ramachandran plot are excluded), total number of protein from database and proteins already solved, which are not contained in the database (Table 1).

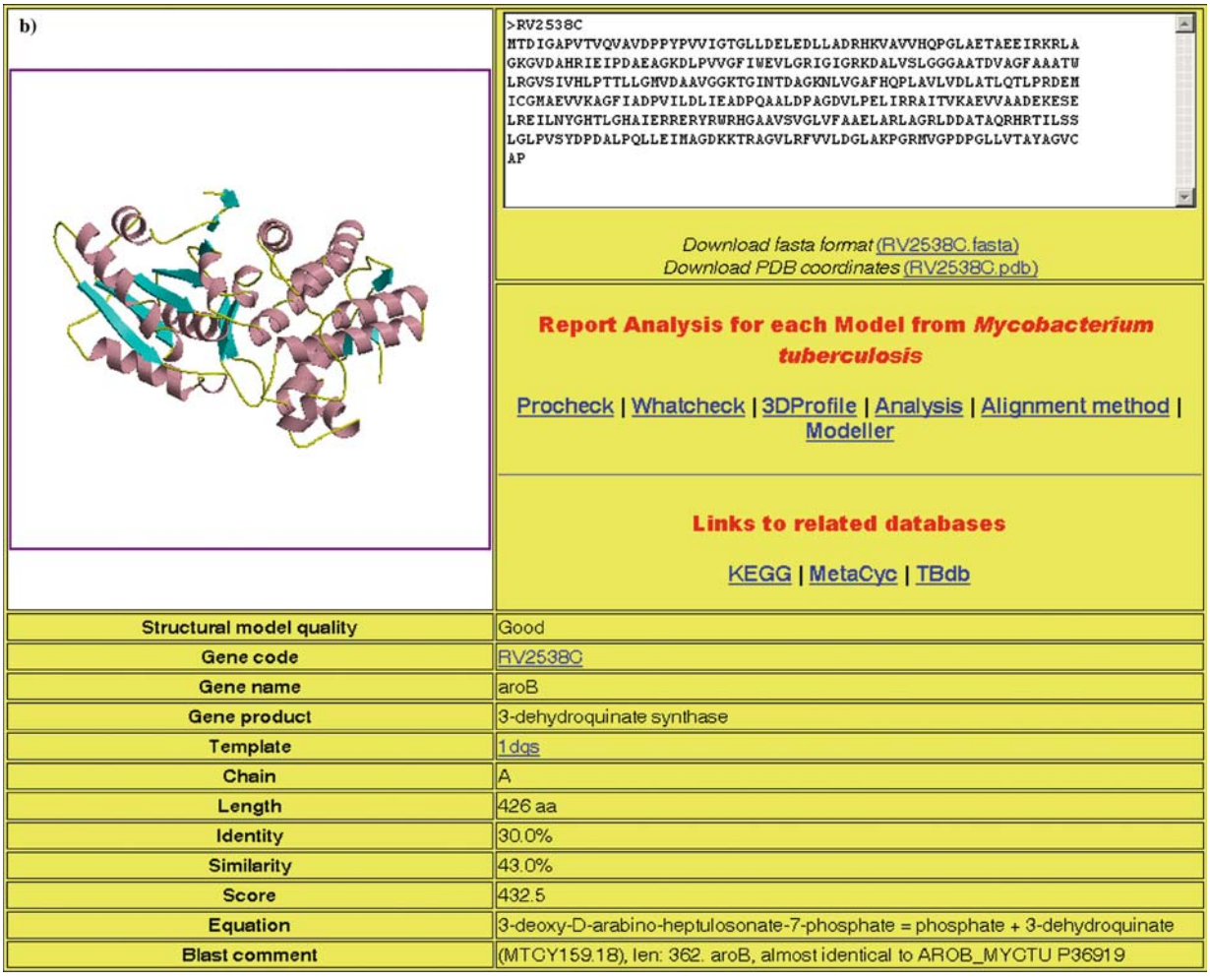
DBMODELING is updated regularly to take into account new sequences from others infectious agents, and new template structures released by the PDB that might allow the construction of models for previously unmodeled proteins, or might provide a better template for already existing model entries. The database will be updated at least monthly to reflect the growth of the sequence and structure databases, as well as improvements in the methods and utilization of new softwares used for analyzing the models.

Results and Discussion

Access and interface

The aim of the DBMODELING is to provide access to a collection of annotated models generated by automated homology modeling from *Mt*. All models in the database are publicly accessible via our interactive website at http://www.biocristalografia.df.ibilce.unesp.br/tools/index.php. The DBMODELING user interface provides user-friendly menus, so that all information can be printed in one step from any standard web browser (Figs. 2a, b). A small ribbon representation is included to obtain a first impression of the model structure. Model atomic coordinates can be downloaded in PDB format and their sequences in fasta format. The database can be queried for protein or metabolic pathway or gene name as keyword, selecting the database table and options as AND, OR, and only one keywords, to refine search. The search interface allows combining all these different descriptors to complex queries. For each model, DBMODELING provides frames in web browsers, which are adjusted into table form. The fields are defined with links to the target sequence entry in Swiss-Prot [33], the template structure

Fig. 2 **a** Description pathway-enzyme links to information about enzyme, and **b** Result frame of web page for 3-dehydroquinate synthase from Mt



Structural model quality	Good
Gene code	RV2538C
Gene name	aroB
Gene product	3-dehydroquinate synthase
Template	1dqs
Chain	A
Length	426 aa
Identity	30.0%
Similarity	43.0%
Score	432.5
Equation	3-deoxy-D-arabino-heptulosonate-7-phosphate = phosphate + 3-dehydroquinate
Blast comment	(MTCY159.18), len: 362. aroB, almost identical to AROB_MYCTU P36919

Fig. 2 (Contd.)

entries in PDB [34], structural information, analysis, and information about the modeling, such as the inputs used in the MODELLER for each model. The database is searchable by metabolic pathways and proteins. It also includes links to a more detailed description of the model, a summary of all models for a given protein, the PDB database [35] for a detailed description of the template structure used in modeling, the Swiss-Prot + TrEMBL protein sequence database [33], and link also to the KEGG [36] pathway database, which consists of graphical diagrams of biochemical pathways including most of the known metabolic pathways, and some of the known regulatory pathways, and MetaCyc [37] database, which is a metabolic-pathway database that describes pathways and enzymes occuriring in 158 organisms. In addition, it links to the model atomic coordinates in the PDB format, which is available for download, the target-template alignment used to obtain the model, visualization of the model performed by PROTGIF tool, which generates animated figures for all models.

Table 3 Structural model quality using Ramachandran plot report

Structural model quality	% of the residues in the most favored regions of the Ramachandran plot
Excellent	> 95
Good	90–95
Fair	85–90

All models that were rated below 85% were not deposited in the present version of this database

Accuracy of the models

To facilitate the evaluation of structure quality we devise a simple rating scheme for the deposited models, as shown in Table 3. The accuracy of comparative modeling is related to the percentage of sequence identity on which the model is based, correlating with the relationship between the structural and sequence similarities of two proteins [38– 40] High accuracy comparative models are based on > 50% sequence identity to their templates. They tend to have ~1 Å r.m.s. error for the main-chain atoms, which is comparable to the accuracy of a medium-resolution NMR structure or a low-relosution X-ray structure. Medium accuracy comparative models are based on 30–50% sequence identity. They tend to have ~90% of the main-chain modeled with 1.5 Å r.m.s. error. There are more frequent side-chain packing, core distortion and loop modeling errors, and there are

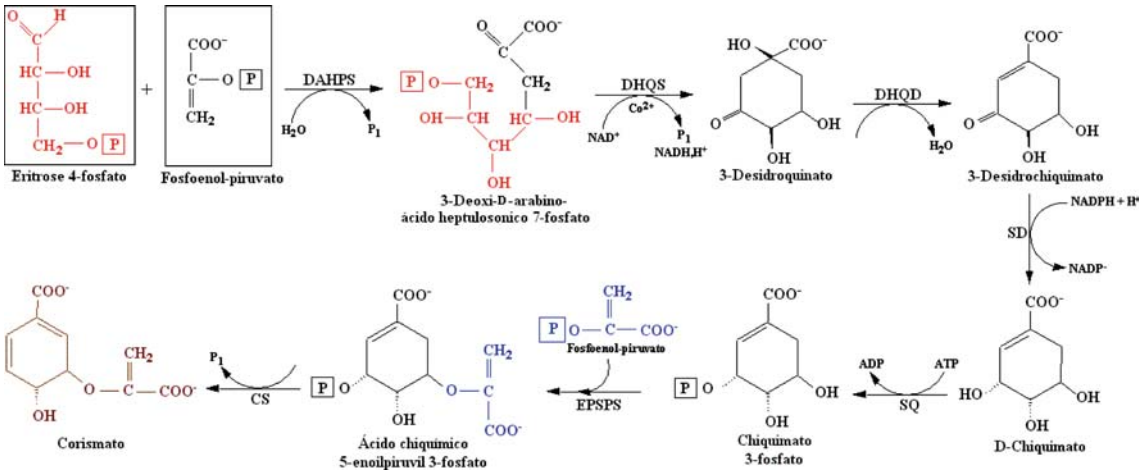


Fig. 3 The shikimate pathway in the sequence of seven metabolic steps, from phosphoenolpyruvate and erythrose-4-phosphate until the conversion to chorismate

occasional alignment mistakes. And finally, low accuracy comparative models are based on $<30\%$ sequence identity. The alignment errors increase rapidly below 30% sequence identity and become the most significant source of errors in comparative models. In addition, when a model is based on an almost insignificant alignment to a known structure, it may also have an entirely incorrect fold. Other factors such as template selection and alignment accuracy usually have a larger impact on the model accuracy, especially for models based on $<40\%$ sequence identity to the templates [31]. All structural models in the database were built using alignments with more than 30% sequence identity, which generated models with medium and high accuracy.

As described in the introduction, the main goal of the present database is to provide structural models to be used in docking simulations and drug design. However, since the accuracy of structural models is highly dependent on the sequence identity between template and target, it is necessary to make clear to the user that only models that show high structural quality should be used in such efforts. Therefore, we strongly recommend that any docking simulations should be focused on structural models that were rated as excelent quality.

Examples of molecular modeling using DBMODELING

There are several *Mt* enzymes whose structures have been determined using X-ray crystallography. The atomic coordinates of these enzymes are available at (http://www.doe-mbi.ucla.edu/TB) *Mt* Structural Genomics Consortium. The strategy of the Consortium is to determine the 3D structures of proteins from *Mt* and place them in the public domain. There are 170 structures of enzymes with released atomic coordinates in PDB available at the *Mt* Structural Genomics Consortium site. The focus of the database are structural models and metabolic pathways of these enzymes. In addition, attractive drug targets involve gene products from important metabolic pathways such as the shikimate pathway. Among several pathways present in DBMODELING, we will discuss one pathway as example of the power of DBMODELING database.

In plants and microorganisms, all the key aromatic compounds involved in primary metabolism, including the three aromatic amino acids found in proteins, are produced via the shikimate pathway. The shikimate pathway is essential in algae, higher plants, bacteria, and fungi, but absent from mammals, which depend on these compounds for their diet [41]. This pathway is an attractive target for drug development.

The shikimate pathway consists of seven enzymes that catalyze the sequential conversion of erythrose-4-phosphate and phosphoenolpyruvate to chorimate [42] (Fig. 3). All pathway intermediates can also be considered branch point compounds that may serve as substrates for other metabolic pathways [43]. The molecular organization and structure of the shikimate pathway enzymes varies considerably between microorganism groups. Bacteria have seven individual polypeptides, each possessing a single enzyme activity, which are encoded by different genes. Due to the importance of this pathway, we performed molecular modeling for two enzymes of this pathway, (1) shikimate kinase I (*Mt* SK, EC 2.7.1.71), whose encoding gene(*aroK*, Rv2539c) was proposed to be present by sequence homology. Shikimate kinase catalyzes a phosphate transfer from ATP to the carbon-3 hydroxyl group of shikimate resulting in the formation of shikimate-3-phosphate (S3P) and ADP [44], and (2) 5-enolpyruvylshikimate-3phosphate (*Mt* EPSP) synthase encoded by the *aroA* gene, which catalyzes the transfer of the enolpyruvyl moiety from phosphoenolpyruvate (PEP) and inorganic phosphate [45].

Comparison of the structural model of *Mt* SK with the recently solved crystallographic structure (PDB access code: 1WE2) [46, 47] generated an rmsd of 1.2 Å after superposition of C_α.*Mt* EPSP synthase has not yet been solved to perform structural comparison with homology model. This result illustrates the applicability of DBMODELING in generating reliable structural models for protein targets.

Conclusion

Large scale protein homology modeling, in which whole sequence databases or whole genomes are used as input into automated modeling algorithms, have been reported by several groups [9]. By using powerful computer systems with multiple processors, these efforts have allowed the creation of large databases of homology models for proteins. DBMODELING can be accessed by any researcher with access to the internet. This project emphasises that homology modeling is a useful tool in structural biology and that it can be very valuable in annotating genome sequence information and contributing to structural and functional genomics. Furthermore, superposition of a structural model present in DBMODELING showed good agreement with the crystallographic structure, validating the modeling protocol. The availability of accurate models has application in biologically significant problems such as substrate specificity and areas such as drug design [48]. In the future, DBMODELING will grow the number of the organisms to reflect importance of the study of proteins from metabolic pathways as potential targets for drug design such as shikimate pathway enzymes, and others potencial targets.

Acknowledgements This work was supported by grants from FAPESP (Process Numbers: 02/10239–6, SMOLBNet 01/07532–0, 02/04383–7, 04/00217–0), CNPq, CAPES and Instituto do Milênio (CNPq-MCT). WFA, LAB, MSP, and DSS are researchers for the Brazilian Council for Scientific and Technological Development.

References

1. Snider DE, Raviglione M, Kochi A (1994) In: Bloom BR (ed) Tuberculosis: pathogenesis, protection, and control, Vol 2(11). Am Soc Microbiol, Washington
2. Cole ST, Brosch R, Parkhill J, Garnier T, Churcher C, Harris D, Gordon SV, Eiglmeier K, Gas S, Barry III CE, Tekaia F, Badcock K, Basham D, Brown D, Chillingworth T, Connor R, Davies R, Devlin K, Feltwell T, Gentles S, Hamlin N, Holroyd S, Hornsby T, Jagels K, Barrell BG (1998) Nature 393:537–544
3. McKinney DJ, Jacobs Jr WR, Bloom BR (1998) Persisting problems in tuberculosis. In: Krause RM (ed) Emerging infections. Academic, New York, pp 51–146
4. Campos HS (1999) Boletim de Pneumologia Sanitária 7:51–64
5. Bloom BR, Murray CJL (1992) Science 257:1055–1064
6. Basso LA, Blanchard JS (1998) Adv Exp Med Biol 456:115–144
7. de Azevedo Jr WF, Mueller-Dieckmann HJ, Schulze-Gahmen U, Worland PJ, Sausville E, Kim SH (1996) Proc Natl Acad Sci USA 93:2735–2740
8. de Azevedo Jr WF, Leclerc S, Meijer L, Havlicek L, Strnad M, Kim SH (1997) Eur J Biochem 243:518–526
9. Foster MJ (2002) Micron 33:365–384
10. Karp PD (2001) Science 293:2040–2044
11. Browne WJ, North AC, Phillips DC, Drew K, Vanaman TC, Hill RL (1969) J Mol Biol 42:65–86
12. Sali A, Blundell TL (1993) J Mol Biol 234:779–815
13. Sánchez R, Sali A (1999) Bioinformatics 15:1060–1061
14. Gerstein M, Levitt M (1997) Proc Nat Acad Sci USA 94:11911–11916
15. Fischer D, Eisenberg D (1999) Curr Opin Struct Biol 9:208–211
16. Uchôa HB, Jorge GE, da Silveira NJF, Camera Jr JC, de Azevedo Jr WF (2004) Biochem Biophys Res Commun 325:1481–1486
17. Laskowski RA, MacArthur MW, Thornton JM (1998) Curr Opin Struct Biol 8:631–639
18. Wilson C, Gregoret LM, Agard DA (1993) J Mol Biol 229:996–1006
19. Laskowski RA, MacArthurm MW, SmithDK, Jones DT, Hutchinson EG, Morris AL, Naylor D, Moss DS, Thornton JM (1994) PROCHECK v.3.0—Program to check the stereochemistry quality of protein structures—Operating instructions
20. Hooft RW, Sander C, Vriend G (1996a) Proteins 26:363–376
21. Bowie JU, Luthy R, Eisenberg D (1991) Science 253:164–170
22. Kabsch W, Sander C (1983) Biopolymers 22:2577–2637
23. Sippl MJ (1990) J Mol Biol 213:859–883
24. Luthy R, Bowie JU, Eisenberg K (1992) Nature 356:83–85
25. Arcuri HA, Canduri F, Pereira JH, da Silveira NJF, Camera Jr JC, de Oliveira JS, Basso LA, Palma MS, Santos DS, de Azevedo Jr WF (2004) Biochem Biophys Res Commun 320:979–991
26. da Silveira NJF, Uchôa HB, Canduri F, Pereira JH, Camera Jr JC, Basso LA, Palma MS, Santos DS, de Azevedo Jr WF (2004) Biochem Biophys Res Commun 322:100–104
27. Brünger AT (1992) X-PLOR, a System for crystallography and NMR, Version 3.1.Yale University Press, New Haven
28. EU 3-D Validation Network (1998) J Mol Biol 276:417–436
29. DuBois P (2000) MySQL. New Riders, Indianapolis
30. Wall L, Christiansen T, Orwant J (2000) In: O'Reilly, Associates Inc (eds) Programming Perl, 3rd edn
31. Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, Stuart AC, Mirkovic N, Rossi A, Marti-Renom MA, Fiser A, Webb B, Greenblatt D, Huang CC, Ferrin TE, Sali A (2004) Nucleic Acids Res 32:D217-D222
32. Kawabata T, Fukuchi S, Homma K, Ota M, Araki J, Ito T, Ichiyoshi N, Nishikawa K (2002) Nucleic Acids Res 30:294–298
33. Bairoch A, Apweiler R (1999) Nucleic Acids Res 27:49–54
34. Westbrook J, Feng Z, Chen L, Yang H, Berman HM (2003) Nucleic Acids Res 31:489–491
35. Abola BB, Bernstein FC, Bryant SH, Koetzle T, Weng J (1987) Protein data bank. In: Allen FH, Bergerhoff G, Sievers R (eds) Crystallographic databases-information, content, software systems, scientific applications. Data Commission of the International Union of Crystallography, Bonn Cambridge Chester, pp 107–132
36. Ogata H, Goto S, Sato K, Fujibuchi W, Bono H, Kanehisa M (1999) Nucleic Acids Res 27:29–34
37. Karp PD, Riley M, Paley SM, Pellegrini-Toole A (2002) Nucleic Acids Res 30:59–61
38. Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000) Annu Rev Biophys Biomol Struct 29:291–325
39. Sanchez R, Sali A (1998) Proc Natl Acad Sci USA 95:13597–13602
40. Koehl P, Levitt M (1999) Nature Struct Biol 6:108–111
41. Coggins JR, Abell C, Evans LB, Frederickson M, Robinson DA, Roszak AW, Lapthorn AP (2003) Biochem Soc Trans 31:548–552
42. Campbell SA, Richards TA, Mui EJ, Samuel BV, Coggins JR, McLeod R, Roberts CW (2004) Int J Parasitol 34:5–13
43. Herman KM, Weaver LM (1999) Annu Rev Plant Mol Biol 50:473–503
44. de Azevedo Jr WF, Canduri F, de Oliveira JS, Basso LA, Palma MS, Pereira JH, Santos DS (2002) Biochem Biophys Res Commun 295:142–148
45. Pereira JH, Canduri F, de Oliveira JS, da Silveira NJF, Basso LA, Palma MS, de Azevedo Jr WF, Santos DS (2003) Biochem Biophys Res Commun 312:608–614
46. Pereira JH, Oliveira JS, Canduri F, Dias MVB, Palma MS, Basso LA, de Azevedo Jr WF, Santos DS (2004) Biochem Biophys Res Commun 325:10–17
47. Pereira JH, Oliveira JS, Canduri F, Dias MVB, Palma MS, Basso LA, Santos DS, de Azevedo Jr WF (2004) Acta Crystallogr Sect D-Biol Crystallogr 60:2310–2319
48. Chakrabarti S, John J, Sowdhamini R (2004) J Mol Model 10:69–75