ELSEVIER

Biophysical Chemistry 105 (2003) 361–370

Biophysical Chemistry

www.elsevier.com/locate/bpc

# Structural genomics of *Mycobacterium tuberculosis*: a preliminary report of progress at UCLA☆

Celia W. Goulding[1], L. Jeanne Perry[1] , Daniel Anderson, Michael R. Sawaya, Duilio Cascio, Marcin I. Apostol, Sum Chan, Angineh Parseghian, Shui-Shu Wang, Yim Wu, Vicente Cassano, Harindarpal S. Gill, David Eisenberg*

*UCLA-DOE Center for Genomics and Proteomics, UCLA Box 951570, Los Angeles, CA 90095-1570, USA*

Received 7 November 2002; accepted 18 November 2002

**Abstract**

The growing list of fully sequenced genomes, combined with innovations in the fields of structural biology and bioinformatics, provides a synergy for the discovery of new drug targets. With this background, the TB Structural Genomics Consortium has been formed. This international consortium is comprised of laboratories from 31 universities and institutes in 13 countries. The goal of the consortium is to determine the structures of over 400 potential drug targets from the genome of *Mycobacterium tuberculosis* and analyze their structures in the context of functional information. We summarize the efforts of the UCLA consortium members. Potential drug targets were selected using a variety of bioinformatics methods and screened for certain physical and species-specific properties to yield a starting group of protein targets for structure determination. Target determination methods include protein phylogenetic profiles and Rosetta Stone methods, and the use of related biochemical pathways to select genes linked to essential prokaryotic genes. Criteria imposed on target selection included potential protein solubility, protein or domain size, and targets that lack homologs in eukaryotic organisms. In addition, some protein targets were chosen that are specific to *M. tuberculosis*, such as PE and PPE domains. Thus far, the UCLA group has cloned 263 targets, expressed 171 proteins and purified 40 proteins, which are currently in crystallization trials. Our efforts have yielded 13 crystals and eight structures. Seven structures are summarized here. Four of the structures are secreted proteins: antigen 85B; MPT 63, which is one of the three major secreted proteins of *M. tuberculosis*; a thioredoxin derivative Rv2878c; and potentially secreted glutamate synthetase. We also report the structures of three proteins that are potentially essential to the survival of *M. tuberculosis*: a protein involved in the folate biosynthetic pathway (Rv3607c); a protein involved in the biosynthesis of vitamin B5 (Rv3602c); and a pyrophosphatase, Rv2697c. Our approach to the *M. tuberculosis*

*Abbreviations:* TB, tuberculosis; *Mtb*, *Mycobacterium tuberculosis*; TEV, Tobacco Etch Virus; TMAO, Trimethylamine n-oxide

☆ This contribution is dedicated to Professor Walter Kauzmann, a great teacher and researcher. Kauzmann's 1959 article on 'Some factors in the interpretation of protein denaturation [Adv. Protein Chem. XIV (1959) 1] started the modern era of thinking about the organization and stability of proteins.

*Corresponding author. Tel.: +1-310-825-3754; fax: +1-310-206-3914.

*E-mail address:* david@mbi.ucla.edu (D. Eisenberg).

[1] Joint first authors.

0301-4622/03/$ - see front matter 
doi:10.1016/S0301-4622(03)00101-7

structural genomics project will yield information for drug design and vaccine production against tuberculosis. In addition, this study will provide further insights into the mechanisms of mycobacterial pathogenesis.


*Keywords:* Structural genomics; Tuberculosis; Bioinformatics

## 1. Introduction

Structural genomics is the large-scale determination and analysis of protein structures from a single organism. This new field has emerged following the great number of completely sequenced genomes and is aided by major technological improvements in robotics and structure determination. One very promising application of large-scale structure determination is to provide a framework for drug discovery on a genomic scale. The discovery of many protein structures from *Mycobacterium tuberculosis* (*Mtb*) will yield information that will provide a foundation to elucidate the cellular biology of this organism.

Tuberculosis is caused by the bacterial pathogen *Mtb*, which kills 2–3 million people around the world each year, more than any other infectious disease. One-third of the world's population is infected with latent TB infection, and approximately 10% of them will develop active tuberculosis. The rise in TB incidence over the last two decades is partly due to TB infections of HIV patients and the emergence of multidrug-resistant strains of the bacteria. HIV-attributed deaths are on the increase, as the weakened immune systems of HIV-infected patients make TB the leading killer of AIDS patients.

Multidrug-resistant strains of *Mtb* arise due to poor administration of therapy. An intensive course of multiple antibiotics over a 6-month period or more is required to prevent reoccurrence of the disease. Patients tend to stop their treatment after several weeks as the symptoms disappear, which fuels the emergence of drug-resistant strains. The World Health Organization has recently funded a massive effort toward stopping this disease before it becomes a global epidemic [1].

In the last decade, molecular understanding of the novel pathogenic strategies of *Mtb* has greatly increased due to the development of efficient mutagenesis strategies to test for essential genes [2] and the completion of the sequence and annotation of the *Mtb* genome at the Sanger Center and Pasteur Institute [3]. With the development of new computational techniques, we have identified potential drug targets and predicted novel folds for the TB Structural Genomics Consortium [4]. This work focuses on the efforts of the UCLA members of the TB Structural Genomics Consortium in drug targeting, cloning, protein purification and structure determination. The structures are briefly discussed.

## 2. Results

### 2.1. Methods for targeting

Determining the function of proteins has traditionally relied on either the study of homologous proteins of known functions or, in the absence of informative homology, laborious experimental work. The availability of many complete genome sequences has made it possible to develop new strategies for computational determination of protein function. Several methods have been developed that can place proteins in their metabolic context by analyzing their functional linkages rather than sequence similarity. One method is the phylogenetic profile method, which examines the co-occurrence of functionally related proteins [5]. Its central assumption is that genes of functionally related proteins co-segregate during speciation, being gained or lost together. Therefore, it can be inferred that genes showing similar patterns of presence or absence across a set of genomes are functionally linked. The gene proximity method looks for positional conservation of gene clusters across genomes [6,7]. If two or more genes have remained close to each other throughout evolution, it can be inferred that there is some functional

significance to this conservation. The domain fusion method looks for pairs of proteins in an organism that are found as a single, fused protein in a different organism. The fused protein is termed the Rosetta Stone protein [8]. Since the functions of the two fused proteins are related, the corresponding pair of proteins can be assumed to be functionally related as well. Utilizing these three computational methods for the functional assignment of proteins [9], we have predicted potential drug targets. Proteins were targeted that are involved in the pathogenesis of *Mtb*, are essential to *Mtb*, belong to functional biochemical pathways known to be targets for anti-tuberculosis drugs, are specific to *Mtb*, and proteins with novel folds.

### 2.2. Drug targets and novel folds

The *Mtb* protein targets consist of five different classes.

1. Extracellular proteins that are potentially involved in the pathogenicity of *Mtb* are important drug targets [10]. Their importance as therapeutic targets is due to the complex composition and structure of the *Mtb* cell envelope, which is impermeable to many antibacterial agents. The ability to target a secreted and yet essential protein by a drug obviates the need to penetrate the cell. *Mtb* secretes a major repository of antigens, as well as proteins that are involved, to some extent, in virulence and persistence determinants.
2. Iron is essential to *Mtb* and plays a critical role in its pathophysiology; therefore, iron-regulatory proteins may prove to be critical protein drug targets [11]. *Mtb* produces in abundance at least two high-affinity iron siderophores, exochelins and mycobactins, which are secreted to scavenge iron from its host, which is essential for *Mtb* survival. Thus, enzymes involved in siderophore production are targeted, i.e. MbtA-H, and the ferric uptake-regulating proteins, FurA and FurB. Proteins functionally related to known proteins essential for bacterial survival (e.g. *E. coli*) [12] have also been targeted. Interestingly, genes in the folate biosynthetic pathway are linked to several essential bacterial genes, and therefore are good *Mtb* drug targets.
3. Proteins functionally related to known targets of anti-TB drugs [2] are of obvious importance; inhibiting these functionally related proteins should have a similar effect on the organism as the present drug, since the same processes or pathways would be disrupted. InhA and enoyl-ACP reductase are the known targets for *Mtb* drugs isoniazid and ethionamide. InhA is linked to polyketide synthases, which are in turn linked to GpsI and PepR, as well as MurE. All are implicated in the synthesis of diaminopicolinate, a component of the cell wall. Another anti-*Mtb* drug, ethambutol, is believed to interact with the EmbB protein, inhibiting the biosynthesis of a lipid component of the cell wall. The EmbB gene is functionally linked to MurF. Both InhA and EmbB are linked to genes that have unknown functions. The proteins these genes encode are of potential interest.
4. Proteins that are specific to mycobacteria are very important drug targets and may provide further information on the virulence and pathogenicity of mycobacteria. In particular, 10% of the *Mtb* genome consists of genes that encode for PE, PPE and PE-PGRS proteins [3]. The function or functions of these proteins are unknown, although they have been implicated in virulence [13].
5. Finally, proteins with predicted novel folds are included in our targets [14]. The approach for predicting novel folds proved to be successful in identifying novel folds in the organism *Pyrobaculum aerophilium*, in which homologs of 10 of the proteins predicted to have unique folds were found to be novel.

We included several criteria for our *Mtb* targets (Fig. 1). To strengthen our drug target strategy, we imposed the restraint that mammalian homologs are absent. For structure determination purposes, the proteins must contain no predicted transmembrane helices, must be smaller than 50 000 Da, all hypothetical proteins must have at least three homologs in three different organisms and must contain 1% methionine (excluding the N-terminal methionine).

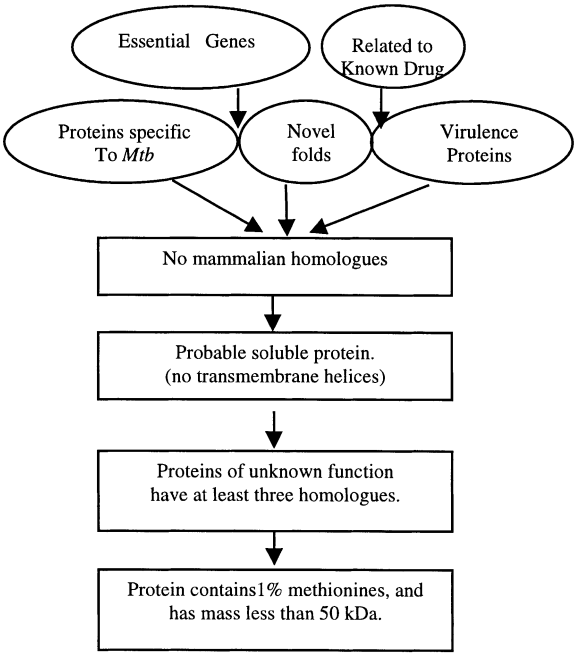


Fig. 1. Flow diagram for the process of selection of proteins from *Mycobacterium tuberculosis* for structural studies. Summary of the criteria imposed on potential drug targets.

### *2.3. Methods and results of cloning, expression, solubility, purification and crystallization trials*

Standard methods were used for the cloning and expression studies, solubility studies, purification and crystallization trials; see Fig. 2 for a flow diagram of the process from targeted gene to structure. Each major step is described below.

#### *2.3.1. Cloning*

*Mtb* ORFs were cloned using standard techniques (Sambrook). The sequence of each targeted gene was analyzed for restriction sites for cohesive ligation. Specifically, our vector of choice was pET22b (Novagen Inc), which requires a unique 5′ *Nde*I site, and contains a C-terminal His-tag. If an *Nde*I site is present within the gene, it is cloned into pETM11, which requires a unique 5′ *Nco*I site and contains an N-terminal His-tag followed by a sequence encoding a TEV cleavage site. PCR primers were designed to introduce unique restriction sites for ligation into expression vectors. For pET22b cloning, the 3′ primer also contained a sequence encoding a thrombin cleavage site. PCR cloning from *Mtb* H37Rv genomic DNA was carried out in a 96-well format. *Mtb* has a GC-rich genome, which often requires optimization of polymerase reactions. Most successful PCRs are achieved by addition of 0–4% dimethylsulfoxide (DMSO) or Q Solution (Qiagen). If further optimization was required, the gradient annealing temperature and magnesium concentration were varied. All PCR products were confirmed by gel electrophoresis. PCR products were rapidly cloned, in a 96-well format, into either TA or blunt PCR cloning vectors. PCR clones were first screened by restriction analysis of miniprep DNA by agarose gel electrophoresis. All PCR clones were both 3′ and 5′ sequenced to confirm ORF sequence and introduction of restriction sites.

#### *2.3.2. Bacterial expression cloning*

Expression cloning was carried out individually. Sequenced plasmid DNA was cleaved, and the DNA fragment was agarose gel-purified and ligated into a cleaved, dephosphorylated expression vector. Most ligations were carried out using the temperature-cycling method of Lund et al. [14a] For each protein, various expression constructs may need to be tested for optimal protein expression, solubility and crystallization. Expression constructs were screened by restriction analysis of miniprep DNA by agarose gel electrophoresis. All constructs were confirmed by 5′ DNA sequencing to check the expression vector 5′ hookup and 5′ ORF sequence. ORFs were cloned into various vectors (pET22b, pET23d, pETM11, pETM30, pETM40 or pQE2) that include a His-tag, MBP or GST fusion with either a TEV or an engineered thrombin cleavage site.

#### *2.3.3. Expression studies*

Once a gene was cloned into an expression vector, the process of determining the level of protein expressed and optimizing expression conditions was begun. Currently, we have streamlined our protocol for expression studies to enable samples to be taken during growth for both expression and solubility studies. Expression studies were carried out on a small scale (8–50 ml) to deter-

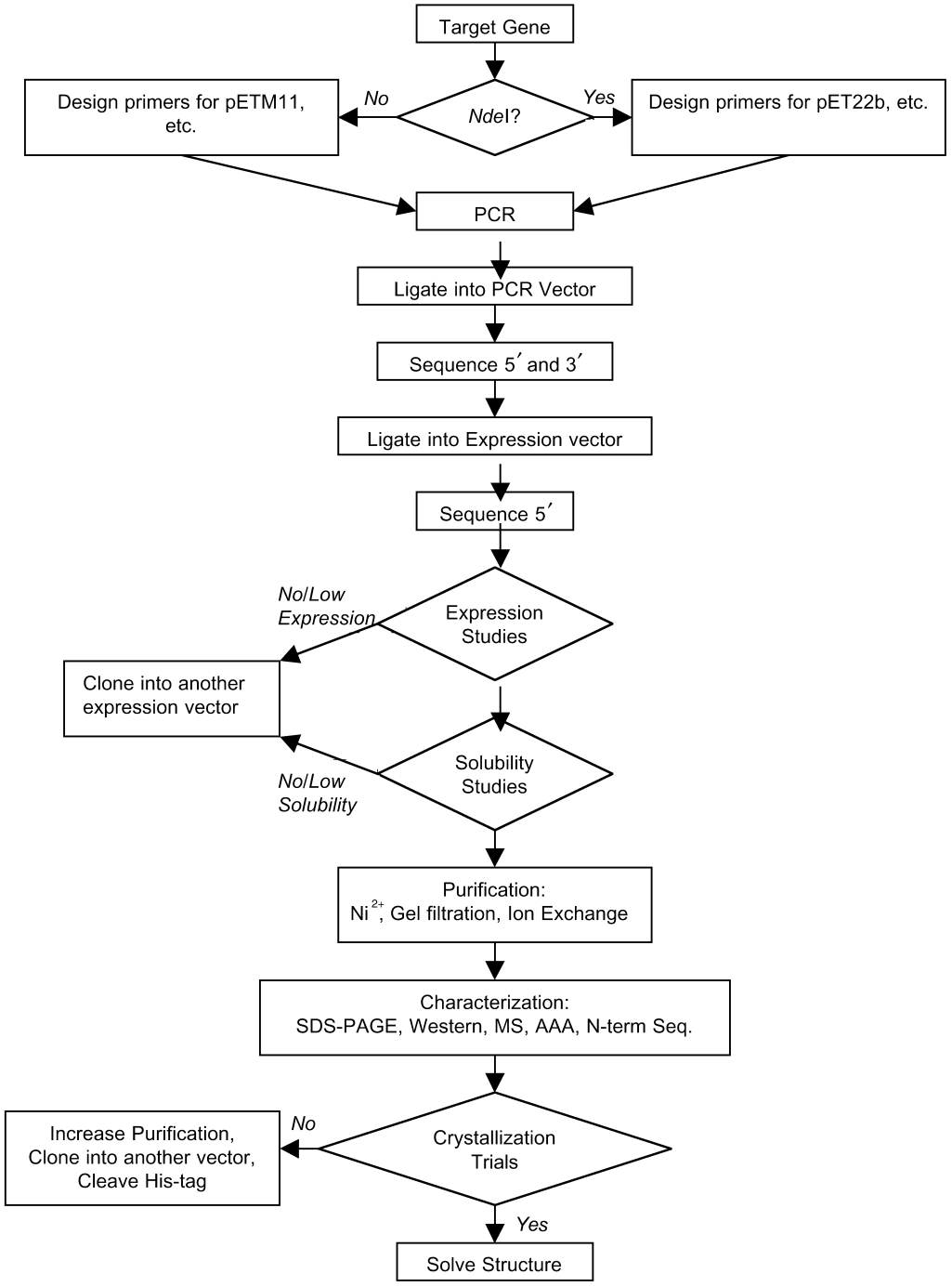


Fig. 2. Flowchart for experimental protocols for cloning, expression and purification of the UCLA component of the TB Consortium for Structural Genomics.

mine optimal *E. coli* host growth conditions and induction time. All constructs were first transformed into BL21(DE3) cell-line. Expression levels were determined by whole-cell lysis and analyzed by SDS-PAGE and Western blot. Based on these results, further expression testing with various codon-enhanced *E. coli* strains was then carried out. Typical expression levels were 5–20 mg/l OD. If expression was $<5$ mg/l OD in all strains and conditions, the targeted ORF was cloned into another expression vector in an attempt to increase yields.

*2.3.4. Solubility studies*

Following determination of expression conditions, evaluation of the soluble conditions was carried out. Solubility studies were conducted on a small-scale to determine optimal *E. coli* lysis and purification buffers. We have generated a simple protocol that includes a general study of buffers containing detergent, 0.5 M urea or high salt to compare these to a native buffer. Solubility levels were determined by separation of soluble and insoluble fractions, followed by analysis by SDS-PAGE and Western blot. If no soluble protein was found under these conditions, further screens were made surveying different detergents, pH values or additives, such as proline, betaine, trehalose and TMAO. In addition, microscopic examination of induced cultures was undertaken in an attempt to observe inclusion bodies. If inclusion bodies were found, the culture was treated according to the deaggregation protocol of Carrio and Villaverde [15]. Cell pellets from the deaggregation protocol were then subjected to solubility screens. Attempts are currently being made to screen for solubility using the 96-well format of Knaust and Nordlund [16]. Once soluble conditions were determined, the construct was grown in a 10-L fermenter. If solubility was $<10\%$, the targeted ORF was cloned into another expression vector, pETM40, which contains an MBP fusion, in an attempt to increase soluble yields.

*2.3.5. Fermentation*

We have the capacity to run 10-L fermentations (BioFlow3000, New Brunswick Inc). Fermentation optimization may include pH control, batch feeding, dissolved oxygen (DO) control, media and temperature shifts. Generally, 10-L fermentations are grown in a modified ECPM1 media or $4\times$ YT in 10 mM MOPS, batch fed at 37 °C, with a temperature shift to 30 °C after induction. Expression levels are measured by whole-cell lysis of time point samples and analyzed by SDS-PAGE and Western blot.

*2.3.6. Protein purification*

The optimal purification scheme is dependent upon the unique properties of each protein. Procedures may include, but are not limited to: French press; ammonium sulfate cuts; addition of protease inhibitors; affinity chromatography (GST, amylose, metal chelate, heparin, dye, etc.); ion exchange chromatography (anion or cation); hydrophobic interaction chromatography; hydroxyapatite, reverse phase and size exclusion chromatography. Since most of our expressed protein are His-tagged, we have streamlined the nickel affinity chromatographic step [17] by carrying out a quick, bench-top purification scheme. The protein purified in this manner is usually relatively (80–95%) pure and concentrated (at least 1 mg/ml). At this point the protein can be used directly for crystallization trials, or may require further concentration and/or purification. Protein concentration was determined by AAA, Bradford [18], or by SDS-PAGE. All proteins were subjected to mass spectroscopy to verify molecular weight. Generally, 1–2 mg of $>95\%$ protein was produced for each small-scale purification run (approx. 100 μl of 15 mg/ml protein).

*2.3.7. Fusion protein cleavage*

For some proteins, it was necessary to remove the tag or fusion partner for enhanced solubility or crystallization. Towards this aim, we often carried out studies designed to optimize cleavage of these fusion partners or tags. Investigation of conditions may involve exploration of temperature variation, additives (detergents, buffers, salts, cofactors, etc.), protease concentration, and timing of cleavage reactions.

*2.3.8. Protein crystallography*

Optimal conditions for crystallization are very difficult to predict. An effective way to search

Table 1
Progress in the structural genomics of *Mycobacterium tuberculosis* at UCLA: the first crystal structures

| Protein | Purity (%) | Diffraction limit (Å) | Phasing | Progress/PDB id code |
|---|---|---|---|---|
| Antigen 85 Rv1886c | 99.9 | 1.8 | MR | Solved/1FOP |
| Glutamine synthetase Rv2220 | >95 | 2.3 | MR | Solved/1HTO |
| 16.3-kDa antigen Rv1926c | >95 | 1.5 | MAD (SeMet) | Solved/1LM1 |
| MPT53-DsbE Rv2878c | >95 | 1.1 | MAD (SeMet) | Solved/1LU4 |
| FolB Rv3607c | >95 | 1.6 | MR | Solved/1NBU |
| dUPTase Rv2697c | >95 | 1.95 | MR | Solved/1MQ7 |
| PanC Rv3602c | >95 | 1.6 | MR | Solved/1MOP |

through a large number of variables (pH, precipitants, salts, concentration, additives, detergents, buffers, organics, etc.) that may affect crystal growth is to conduct a sparse matrix search of specific crystallization conditions. We began most of our initial screens with commercial crystal screening kits. This information was then used to develop expanded screens in subsequent trials. Currently, we are also employing the capillary crystallization method of Ng and co-workers [19], as well as the small-scale batch crystallization of Rayment [20]. We are currently using both hanging drop and sitting drop crystallization. Manual inspection of crystallization experiments is carried under a microscope.

### 2.4. Structures

Summaries of the *Mtb* protein crystals obtained, along with structures in refinement and those determined, are listed in Table 1. Two out of the seven structures were solved by multiwavelength anomalous diffraction from selenomethionine protein derivatives; the remaining five were solved by molecular replacement.

#### 2.4.1. Rv1886c

The 30-kDa major secreted protein (antigen 85B) is the most abundant protein exported by *Mtb*, a potent immunoprotective antigen and a leading drug target. A mycolyl transferase of 285 residues, it is closely related to two other mycolyl transferases, each of 32 000 Da in molecular mass: antigen 85A and antigen 85C. All three catalyze transfer of the fatty acid mycolate from one trehalose monomycolate to another, resulting in trehalose dimycolate and free trehalose, and thus helping to build the mycobacterial cell wall (Fig. 3a) [21]. Based on the trehalose-bound structure, we suggest a new class of antituberculous drugs, made by connecting two trehalose molecules by an amphipathic linker.

#### 2.4.2. Rv2220

*Mtb* has been shown to secrete the enzyme glutamine synthetase [22], which is apparently essential for infection. The structure consists of two dodecamers in the asymmetric unit, having both their six-fold and two-fold axes parallel to one another (Fig. 3b) [23]. As a highly regulated enzyme at the core of nitrogen metabolism, glutamine synthetase has been studied intensively.

#### 2.4.3. Rv1926c

Rv1926c is a major secreted protein of unknown function, which is specific to mycobacteria. Its structure is a beta-sandwich (Fig. 3c) with structural similarity to cell-surface binding proteins (i.e. arrestin, adaptin, invasin), some of which are involved in endocytosis [24–27]. Structural similarity implicates Rv1926c in possible host–bacterial interactions [31].

#### 2.4.4. Rv2878c

Rv2878c is also a secreted protein with sequence homology to disulfide bond isomerase proteins, in particular *E. coli* DsbE. It has a thioredoxin-like fold with active site cysteines in the reduced form (Fig. 3d) (Goulding, unpublished data). Rv2878c potentially protects mycobacteria from oxidative damage by macrophages and/or corrects incorrectly formed disulfide bonds in secreted proteins.

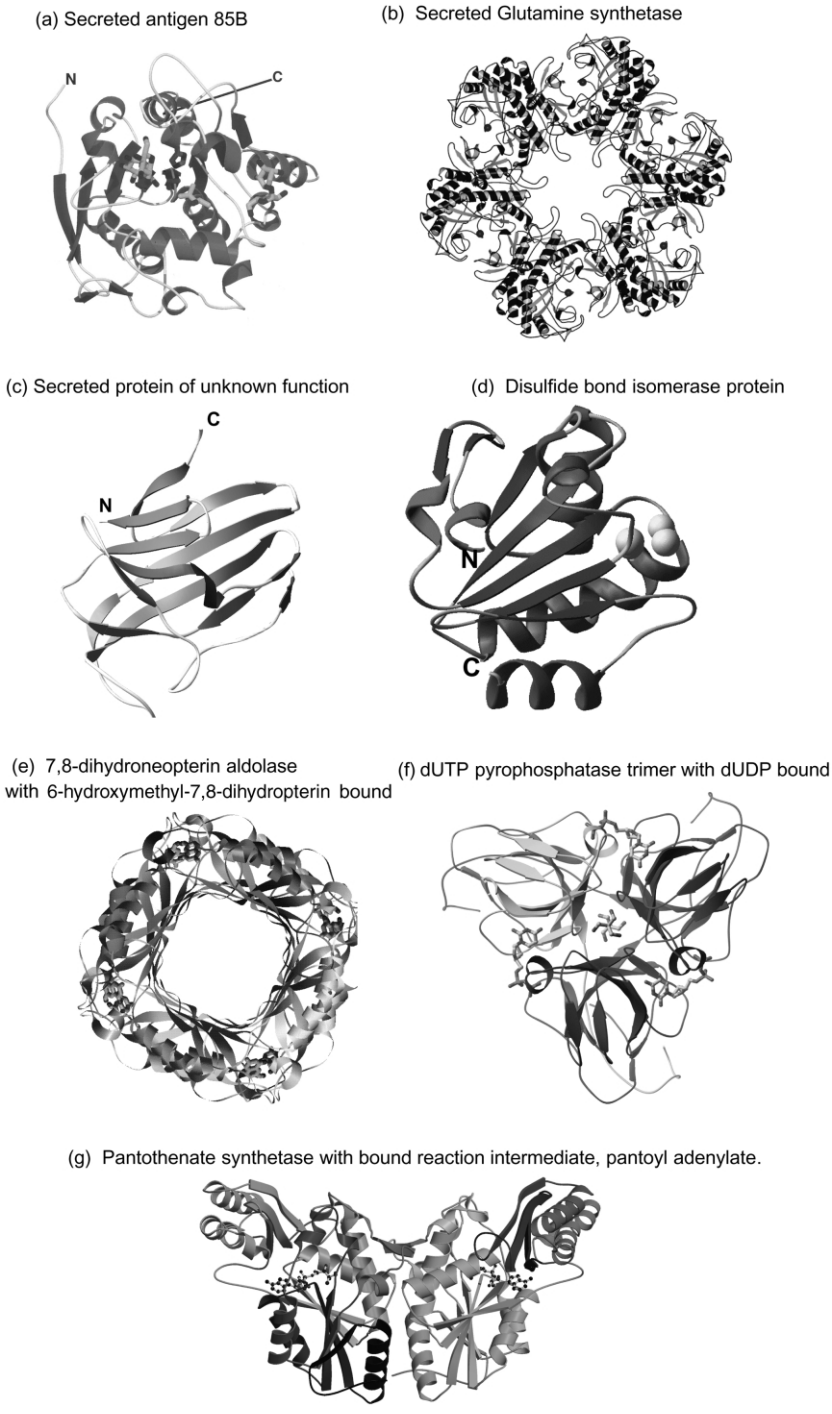


Fig. 3. The first structures of proteins from MTB determined at UCLA: (a) Rv1886c; (b) Rv2220; (c) Rv2878c; (d) Rv2878c; (e) Rv3607c; (f) Rv2697c; and (g) Rv3602c. (a) Secreted antigen 85B; (b) secreted glutamine synthetase; (c) secreted protein of unknown function; (d) disulfide bond isomerase protein; (e) 7,8-dihydroneopterin aldolase with 6-hydroxymethyl-7,8-dihydropterin bound; (f) dUTP pyrophosphatase trimer with dUDP bound; and (g) pantothenate synthetase with bound reaction intermediate, pantoyl adenylate.

*2.4.5. Rv3607c*

Rv3607c is a protein involved in the folate biosynthetic pathway with homology to *E. coli* FolB. The solution was found by molecular replacement and the structure is still being refined, although it appears to be a dimer of tetramers (Fig. 3e), which is consistent with the structure of *Staphylococcus aureus* FolB [28] (Goulding, unpublished data). Folate production is essential for survival of bacteria, whereas humans can absorb folate from their diet. This implies that *Mtb* FolB would be a potentially important, novel drug target.

*2.4.6. Rv2697c*

Rv2697c is a dUTP pyrophosphatase, essential for preventing the accumulation of dUTP and its cytotoxic incorporation into DNA. Inhibition of dUTPase has been recognized as a potential means of slowing viral replication and cancer growth [29]. Thus, it seems that a drug targeted to *Mtb* dUTPase may likewise inhibit the growth of *Mtb*. Comparison of the structure of Rv2697c from *Mtb* (Fig. 3f) with the human dUTPase reveals significant differences in the dUTP binding site that may be exploited in drug design (Chan, unpublished data).

*2.4.7. Rv3602c*

Rv3602c is a pantothenate synthetase (PS) involved in the last step of the pantothenate (vitamin B5) biosynthetic pathway. PS is a dimer, with each subunit consisting of two domains (Fig. 3g). The N-terminal domain forms a typical Rossmann fold, with a central, six-stranded, parallel β-sheet flanked by α-helices. The C-terminal domain forms a hinged lid over the active site cavity on the N-terminal domain. It has two layers: a three-stranded antiparallel beta sheet with two helices layered on one side [32]. Pantothenate synthetase is essential for the growth of *Mtb* [30], and is therefore a potential drug target.

## 3. Conclusions

The abundance of genomic sequence data enables the efficient identification of target genes. Protein production from *E. coli* expression clones is a valuable resource for structural and functional genomics, protein engineering and protein–protein interaction studies. It is clear that current methodologies are not sufficient to provide the necessary high-throughput protein structure determination to keep pace with the rate of gene discovery. We are at present attempting to streamline protocols and procedures, implementing new technologies, to provide a rate and scale of protein production for efficient crystallization and structural analyses.

## Acknowledgments

The authors thank Dr John T. Belisle, Colorado State University and NIH, NIAID Contract NO1 AI-75320 for the generous supply of TB H37Rv genomic DNA; Dr Marila Gennaro for the Rv1926c and Rv2878c expression clones; and the TB Structural Genomics Consortium for support and helpful discussions. We also thank Brookhaven National Laboratory for the use of beamline X8C of the National Sychrotron Light Source, in particular Joel Berendzen, Li Wei Hung and Leonid Flaks. This work has been supported by the National Institutes of Health for the TB Structural Genomics Consortium, DOE and HHMI.

## References

[1] World Health Organization, The World Health Report 1998: Life in the 21st Century—A Vision for All, WHO, Geneva, Switzerland, 1998.
[2] C. Guilhot, M. Jackson, B. Gicquel, in: C. Ragledge, J. Dale (Eds.), Mycobacteria: Molecular Biology and Virulence, Blackwell Science Ltd, 1999, p. 17.
[3] S.T. Cole, R. Brosch, J. Parkhill, et al., Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence, Nature 393 (1998) 537–544.
[4] http://www.doe-mbi.ucla.edu/TB/.
[5] M. Pellegrini, E.M. Marcotte, M.J. Thompson, D. Eisenberg, T.O. Yeates, Assigning protein functions by comparative genome analysis: protein phylogenetic profiles, Proc. Natl. Acad. Sci. USA 96 (1999) 4285–4288.
[6] T. Dandekar, B. Snel, M. Huynen, P. Bork, Conservation of gene order: a fingerprint of proteins that physically interact, Trends Biochem. Sci. 23 (1998) 324–328.
[7] R. Overbeek, M. Fonstein, M. D'Souza, G.D. Pusch, N. Maltsev, The use of gene clusters to infer functional coupling, Proc. Natl. Acad. Sci. USA 96 (1999) 2896–2901.
[8] E.M. Marcotte, M. Pellegrini, H.L. Ng, D.W. Rice, T.O. Yeates, D. Eisenberg, Detecting protein function and

protein-protein interactions from genome sequences, Science 285 (1999) 751–753.
[9] E.M. Marcotte, M. Pellegrini, M.J. Thompson, T.O. Yeates, D. Eisenberg, A combined algorithm for genome-wide prediction of protein function, Nature 402 (1999) 83–86.
[10] J. Thole, R. Janssen, D. Young, in: C. Ragledge, J. Dale (Eds.), Mycobacteria: Molecular Biology and Virulence, Blackwell Science Ltd, 1999, p. 356.
[11] V. Webb, J. Davis, in: C. Ragledge, J. Dale (Eds.), Mycobacteria: Molecular Biology and Virulence, Blackwell Science Ltd, 1999, p. 287.
[12] R.E. Bruccoleri, T.J. Dougherty, D.B. Davison, Concordance analysis of microbial genomes, Nucleic Acids Res. 26 (1998) 4482–4486.
[13] L. Ramakrishnan, N.A. Federspiel, S. Falkow, Granuloma-specific expression of *Mycobacterium* virulence proteins from the glycine-rich PE-PGRS family, Science 288 (2000) 1436–1439.
[14] P. Mallick, K.E. Goodwill, S. Fitz-Gibbon, J.H. Miller, D. Eisenberg, Selecting protein targets for structural genomics of *Pyrobaculum aerophilium*: validating automated fold assignment methods by using binary hypothesis testing, Proc. Natl. Acad. Sci. USA 97 (2000) 2450–2455.
[14a] A.H. Lund, M. Duch, F.S. Pedersen, Increased cloning efficiency by temperature-cycle ligation, Nucl. Acids Res. 24 (1996) 800–801.
[15] M.M. Carrio, A. Villaverde, Protein aggregation as bacterial inclusion bodies is reversible, FEBS Lett. 489 (2001) 29–33.
[16] R.K. Knaust, P. Nordlund, Screening for soluble expression of recombinant proteins in a 96-well format, Anal. Biochem. 297 (2001) 79–85.
[17] E. Hochuli, H. Dobeli, A. Schacher, New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues, J. Chromatogr. 411 (1987) 177–184.
[18] M.M. Bradford, A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein–dye binding, Anal. Biochem. 72 (1976) 248–254.
[19] J.A. Gavira, D. Toh, J. Lopez-Jaramillo, J.M. Garcia-Ruiz, J.D. Ng, Ab initio crystallographic structure determination of insulin from protein to electron density without crystal handling, Acta Crystallogr. D: Biol. Crystallogr. 58 (2002) 1147–1154.
[20] I. Rayment, Small-scale batch crystallization of proteins revisited: an underutilized way to grow large protein crystals, Structure (Camb.) 10 (2002) 147–151.
[21] D.H. Anderson, G. Harth, M.A. Horwitz, D. Eisenberg, An interfacial mechanism and a class of inhibitors inferred from two crystal structures of the *Mycobacterium tuberculosis* 30-kDa major secretory protein (antigen 85B), a mycolyl transferase, J. Mol. Biol. 307 (2001) 671–681.
[22] G. Harth, D.L. Clemens, M.A. Horwitz, Glutamine synthetase of *Mycobacterium tuberculosis*: extracellular release and characterization of its enzymatic activity, PNAS 91 (1994) 9342–9346.
[23] H.S. Gill, G.M. Pfluegl, D. Eisenberg, Multicopy crystallographic refinement of a relaxed glutamine synthetase from *Mycobacterium tuberculosis* highlights flexible loops in the enzymatic mechanism and its regulation, Biochemistry 41 (2002) 9863–9872.
[24] D.J. Owen, Y. Vallis, B.M. Pearse, H.T. McMahon, P.R. Evans, The structure and function of the beta 2-adaptin appendage domain, EMBO J. 19 (2000) 4216–4227.
[25] D.J. Owen, J.P. Luzio, Structural insights into clathrin-mediated endocytosis, Curr. Opin. Cell Biol. 12 (2000) 467–474.
[26] J.A. Hirsch, C. Schubert, V.V. Gurevich, P.B. Sigler, The 2.8-Å crystal structure of visual arrestin: a model for arrestin's regulation, Cell 97 (1999) 257–269.
[27] Z.A. Hamburger, M.S. Brown, R.R. Isberg, P.J. Bjorkman, Crystal structure of invasin: a bacterial integrin-binding protein, Science 286 (1999) 291–295.
[28] M. Hennig, A. D'Arcy, I.C. Hampele, M.G. Page, C. Oefner, G.E. Dale, Crystal structure and reaction mechanism of 7,8-dihydroneopterin aldolase from *Staphylococcus aureus*, Nat. Struct. Biol. 5 (1998) 357–362.
[29] J.M. Harris, E.M. McIntosh, G.E. Muscat, Structure/function analysis of a dUTPase: catalytic mechanism of a potential chemotherapeutic target, J. Mol. Biol. 288 (1999) 275–287.
[30] R. Zheng, J.S. Blanchard, Steady-state and pre-steady-state kinetic analysis of *Mycobacterium tuberculosis* pantothenate synthetase, Biochemistry 40 (2001) 12904–12912.
[31] C.W. Goulding, A. Parseghian, M.R. Sawaya, D. Cascio, M.I. Apostol, M.L. Gennaro, D. Eisenberg, Crystal structure of a major secreted protein of Mycobacterium tuberculosis-MPT63 at 1.5-A resolution', Protein Sci. 11 (2002) 2887–2893.
[32] S. Wang, D. Eisenberg, Crystal structures of a pantothenate synthetase from M. tuberculosis and its complexes with substrates and a reaction intermediate, Protein Sci. 13 (2003) 1097–1108.