The potential impact of structural genomics on tuberculosis drug discovery

Vickery L. Arcus[1,2], J. Shaun Lott[2], Jodie M. Johnston[2] and Edward N. Baker[2]

[1]AgResearch Structural Biology Laboratory, School of Biological Sciences, University of Auckland, Private Bag 92-019, Auckland, New Zealand
[2]Laboratory for Structural Biology, School of Biological Sciences, University of Auckland, Private Bag 92-019, Auckland, New Zealand

***Mycobacterium tuberculosis*, the causative agent of tuberculosis (TB) in humans, is a devastating infectious organism that kills approximately two million people annually. The current suite of antibiotics used to treat TB faces two main difficulties: (i) the emergence of multidrug-resistant (MDR) strains of *M. tuberculosis*, and (ii) the persistent state of the bacterium, which is less susceptible to antibiotics and causes very long antibiotic treatment regimes. The complete genome sequences of a laboratory strain (H37Rv) and a clinical strain (CDC1551) of *M. tuberculosis* and the concurrent identification of all the open reading frames that encode proteins within this organism, present structural biologists with a wide array of protein targets for structure determination. Comparative genomics of the species that make up the *M. tuberculosis* complex has also added an array of genomic information to our understanding of these organisms. In response to this, structural genomics consortia have been established for targeting proteins from *M. tuberculosis*. This review looks at the progress of these major initiatives and the potential impact of large scale structure determination efforts on the development of inhibitors to many proteins. Increasing sophistication in structure-based drug design approaches, in combination with increasing numbers of protein structures and inhibitors for TB proteins, will have a significant impact on the downstream development of TB antibiotics.**

Tuberculosis (TB) is pandemic, and its causative agent – *Mycobacterium tuberculosis* – is one of the most prolific infectious agents affecting humans. A third of the world's population is thought to host *M. tuberculosis* and ~30 million people have died from the disease in the past decade [1]. Several antibiotics for the treatment of TB were discovered and developed in the 1940s and 1950s and in the subsequent decades the incidence of the disease declined, particularly in developed countries. This trend has been reversed in the past twenty years, and this reversal is thought to be driven by several factors, including poverty, overcrowding and the synergy between HIV and TB [2,3]. This synergy is particularly sinister, where active TB infection arises after immune suppression by HIV. Where there is co-infection with TB and HIV, the risk of death is twice that of a person infected with HIV alone [2]. These alarming statistics led to the World Health Organization (WHO) declaring TB 'a global emergency' in 1993 [4].

The emergence of drug resistant strains of *M. tuberculosis* [5] has mirrored that of emerging drug resistance in other pathogenic bacteria, such as *Staphylococcus aureus*, caused by the widespread use of antibiotics. What sets *M. tuberculosis* apart from other pathogenic bacteria is the long treatment time required to sterilize tissues using anti-TB antibiotics – typically six to eight months. This extended period of treatment is a direct result of the lifestyle of the bacilli, which enter a 'dormant' or 'persistent' phase after the initial infection, and the concomitant immune response. Under conditions of dormancy, the bacilli are particularly difficult to kill and thus require prolonged exposure to antibiotics. The lengthy treatment regime, which includes two or more antibiotics, exacerbates compliance issues and where antibiotics are abandoned mid-course, the possibility for selection of resistant strains is high.

Corresponding authors: Arcus, V.L. (v.arcus@auckland.ac.nz); Baker, E.N. (ted.baker@auckland.ac.nz)

1359-6446/06/$ – see front matter ©2006 Elsevier Ltd. All rights reserved. PII: S1359-6446(05)03667-6

No new antibiotics against TB have been developed in the past 30 years. There are three front line antibiotics, isoniazid, rifampin and pyrazinamide, and several second tier antibiotics, including ethambutol, streptomycin and para-aminosalicylic acid. There is an internationally recognized optimized treatment regime called DOTS (directly observed therapy – short course) which is effective in TB treatment. Indeed, recent reports suggest that DOTS is also effective in reducing the incidence of single and multiple drug resistant forms of TB [6]. Despite this, an urgent need for new antibiotics towards TB is widely acknowledged and a Global Alliance for TB Drug Development (www.tballiance.org) [7] has been established with goals that include increasing compliance with current therapies, shortening treatment times with the development of new adjuvant drugs, improved treatment for multidrug-resistant (MDR) TB and new treatments for persistent *M. tuberculosis*. There have been significant successes very recently, and four new drugs are currently in clinical trials.

Structure-based drug design

The incorporation of atomic-resolution protein structure information into the design of drugs – structure-based drug design – has been discussed as a concept since the first protein structures were determined. Despite this long gestation period, there have only been a few successful examples where drugs have reached the market via this route (notably the neuraminidase inhibitors Relenza [8] and Tamiflu [9] and the HIV-protease inhibitors, such as Viracept [10], Agenerase [11] and Aluviran [12]). However, new developments in experimental methods and bioinformatics have lead to an explosion of activity in this area over the past five years (for a recent general review see [13]). Experimental methods have now been developed using X-ray crystallography and high resolution NMR to identify lead compounds by fragment-based screening and then iteratively improve the specificity and affinity of those compounds for their target proteins. One of the original methods called 'SAR by NMR' [14] uses the perturbation of NMR protein amide signals to screen for ligand binding. A second NMR-based method to detect ligand-binding, called SHAPES [15], involves qualitative analysis of NMR spectra. Fragment-based screening has also been embraced by the protein X-ray crystallographers and methodological advances in this field have been rapid [16]. These experimental approaches have been combined with increasingly sophisticated bioinformatics tools like ligand docking to provide high-affinity lead molecules for drug discovery. It is estimated that there are 40 drugs – for a range of conditions – currently in clinical trials where structure-based drug design has had an important role in their identification and development [17].

The combination of high-throughput structure determination (structural genomics), new structure-based drug design approaches and comparative genomics between *M. tuberculosis* and closely related mycobacteria [*M. bovis, M. microti, M. canettii* and *M. africanum* (collectively called the '*M. tuberculosis* complex')] presents the possibility for rapid progress in drug discovery that is informed by protein structure and the development of new classes of inhibitors of *M. tuberculosis* proteins.

Genomics of the *M. tuberculosis* complex

There are now six species of mycobacteria whose genomes have been fully sequenced, including two strains of *M. tuberculosis* – a laboratory strain, H37Rv [18,19], and a clinical strain, CDC1551 [20]. Among these organisms, several are uniquely pathogenic to humans (*M. tuberculosis, M. africanum* and *M. canettii*), whereas *M. bovis* and *M. microti* have broad host ranges. These five mycobacterial species are thought to have a common evolutionary origin and to have undergone an evolutionary bottleneck ~20,000 years ago [21]. Detailed comparative genomics has recently overturned the commonly held view that TB in humans arose by a species leap from cattle [21]. It is now thought that after the bottleneck ~20,000 years ago, all species in the *M. tuberculosis* complex diverged from a single species which closely resembled *M. tuberculosis* or *M. canettii*. This ancestral mycobacterium might have been a human pathogen whose host range then expanded to encompass other mammals, including cattle [21,22].

This plethora of genomic sequence data presents a unique opportunity for structural biologists. Not only have all the genes in each of these species been identified, but comparative and molecular genomics techniques (such as whole genome transposon mutagenesis and microarray analysis) can reveal genes essential for growth, pathogenicity and persistence. Structural genomics can be added to this list as another weapon in the armoury of research whose final goal is the eradication of TB.

Structural genomics and *M. tuberculosis*

Structural genomics was first mooted approximately ten years ago and proposed the use of X-ray crystallography and high resolution NMR to solve protein structures in the discovery or genomics phase of research. One of the central ideas was that structure determination of a hypothetical protein from a genome sequence could provide clues to the biological function of the protein from distant structural homology not apparent from sequence comparison [23,24]. Protein structural topology and structural motifs are much more tightly conserved over evolutionary time when compared with sequence motifs. Another ambitious goal for structural genomics was gaining structural data for all proteins expressed in a particular organism, organelle or tissue. Indeed, the pinnacle of this endeavour would be to have complete coverage of protein structure space, so that any new sequence obtained from genome sequencing would be within homology modelling distance of an experimentally determined structure. It is generally accepted that the lower bound for modelling protein structures requires sequence identities of ≥30% between the query sequence and a known structure. For example, for accurate homology modelling suitable for drug design, sequence identities of >60% are necessary. It has been predicted that coverage of 'protein structure space' will require an additional 16,000 new carefully chosen and experimentally determined protein structures [25].

A more applied approach to structural genomics is to focus on pathogenic organisms and to target proteins that are considered to be good candidates for inhibitor design. These inhibitors are expected to act as lead compounds for drug discovery. This is the central goal for the *Mycobacterium tuberculosis* Structural Genomics Consortium (www.tbgenomics.org). 'A vision inspiring many members of the project is that if structures of many of the proteins from *M. tuberculosis* were known, then as key *M. tuberculosis* genes are identified by genetic means, high-throughput screening, or other approaches, the structural information that could help speed up drug discovery would already be in place' [26]. The

BOX 1

Descriptions of techniques in the structure determination pipeline

- **GATEWAY™-based cloning.** A method for cloning genes into plasmids which utilizes the recombination system from phage and allows inserts to be shuttled, in frame, between expression vectors. This is used instead of more traditional approaches that use restriction enzymes and ligase to clone and subclone genes.
- **Auto-inducible *E. coli* expression.** Overexpression of proteins in the host strain *E. coli* using a media that allows growth to high densities without induction of protein expression, and then facilitates induction by lactose automatically due to the depletion of factors which repress induction during growth.
- **Incorporation of selenomethionine.** The introduction of the heavy atom selenium into a protein during expression by using selenomethionine instead of methionine. This is useful to crystallographers in solving protein structures as the selenium atoms scatter X-rays in peculiar ways when compared with the smaller atoms normally found in proteins – carbon, nitrogen, hydrogen and oxygen.
- **Multiple wavelength anomalous dispersion (MAD).** A technique used to solve protein structures by recording X-ray diffraction data at different wavelengths. This is usually done at a synchrotron where the X-rays are 'tuneable' and often uses selenomethionine-substituted protein crystals. The scattering of X-rays at different wavelengths shows small variations which crystallographers can exploit to solve protein structures.
- **Multiple isomorphous replacement (MIR).** In contrast to MAD techniques, MIR uses the small differences in the scattering of X-rays when different heavy metals are introduced into the protein crystals. This is usually achieved by soaking the crystals in separate solutions of heavy atoms such as platinum, gold or lead and then recording the X-ray diffraction data for each different protein crystal containing a heavy atom derivative.

Mycobacterium tuberculosis Structural Genomics Consortium was established in 2000 [as part of the U.S. Protein Structure Initiative (PSI)] and now encompasses 230 researchers from 31 organisations working in 11 countries [26,27]. In addition, structural genomics efforts focussing on mycobacteria run in parallel in several European countries – these are associated with the SPinE (structural proteomics in Europe, www.spineurope.org) and X-MTB (www.xmtb.org) projects.

Technical advances in high-throughput structure determination

Whether the goal in a structural genomics effort is mapping protein-fold space, functional annotation of hypothetical proteins using structure, or assisting drug design, a common thread is a concerted effort to increase the number of protein structures determined and to decrease the time, effort and cost required for each experimentally determined 3D protein structure. This, in turn, has required significant technical advances in all stages of protein structure determination, from high-throughput cloning methods to protein expression, purification and crystallization methods, and methods for structure solution.

The PSI, launched by the US National Institute of General Medical Sciences, and parallel initiatives in Europe and Asia have had their greatest impact in this area of technological advance over their first five years of operation (since 2000). Very significant advances have been made in the number of proteins that can be cloned, overexpressed and purified in parallel. Robotics has been used at each step in an effort to automate as many steps as possible. Robotic crystallization trials have greatly increased the breadth of the experimental space that can be sampled to find conditions under which proteins will crystallize. The volume of protein solution required for crystallization experiments, and hence the amount of purified protein required, has been concomitantly reduced by an order of magnitude using robotics. There is anecdotal evidence that the use of nanolitre volumes in crystallization has also increased success rates [28]. Automated methods for structure solution from X-ray data have also had a significant impact on the number of structures that can be determined.

Examples of the techniques used in this structure determination 'pipeline' are: GATEWAY™-based cloning and expression; auto-inducible *Escherichia coli* expression strains and media [29]; parallel protein-purification using affinity tags; automated robotic crystallization; automated crystal handling at cryogenic temperatures; incorporation of selenomethionine into proteins for structure solution using multiple wavelength anomalous dispersion (MAD); automated structure solution for multiple isomorphous replacement (MIR) or MAD; and automated interpretation of electron density maps (for more detailed explanations see Box 1).

Potential drug targets among *M. tuberculosis* protein structures

Targets for 3D structure determination by the *M. tuberculosis* Structural Genomics Consortium have been chosen with two main goals: the characterisation of potential new drug targets and the investigation of important aspects of TB biology. Many of the drug targets are classic biosynthetic enzymes that are easily recognized in the genome sequence. Others are suggested by biological studies. For example, a genome-wide transposon mutagenesis study [30] has identified a large cohort of genes that are essential for the growth of *M. tuberculosis in vitro*; many of these correspond to biosynthetic proteins but others are of completely unknown function. Likewise, microarray studies have examined the upregulation or downregulation of gene expression in response to antibiotic challenge [31] or to conditions of hypoxia (thought to have a role in the onset of dormancy) [32] and have implicated other gene products of unknown function as potentially important in pathogenesis. These proteins are prime candidates for structural analysis to address function because of their evident (but unexplained) importance in *M. tuberculosis* biology.

Amino acid biosynthesis

The biosynthesis of many amino acids is essential for the growth of microorganisms and, by inference, the biosynthetic enzymes in these pathways are possible targets for inhibitors as lead compounds for drug discovery. Several structures have been determined for such *M. tuberculosis* enzymes, with typical examples including: LeuA (α-isopropylmalate synthase; Figure 1a) and LeuB (3-isopropylmalate dehydrogenase; Figure 1c), both from the leucine biosynthesis pathway [33,34]; LysA (meso-diaminopimelate decarboxylase, DAPDC; Figure 1b) from the lysine biosynthetic pathway [35]; and HisG (ATP phosphoribosyltransferase; Figure 1d) from the pathway for histidine biosynthesis [36]. In the case of leucine biosynthesis there is evidence to suggest that an inhibitor that

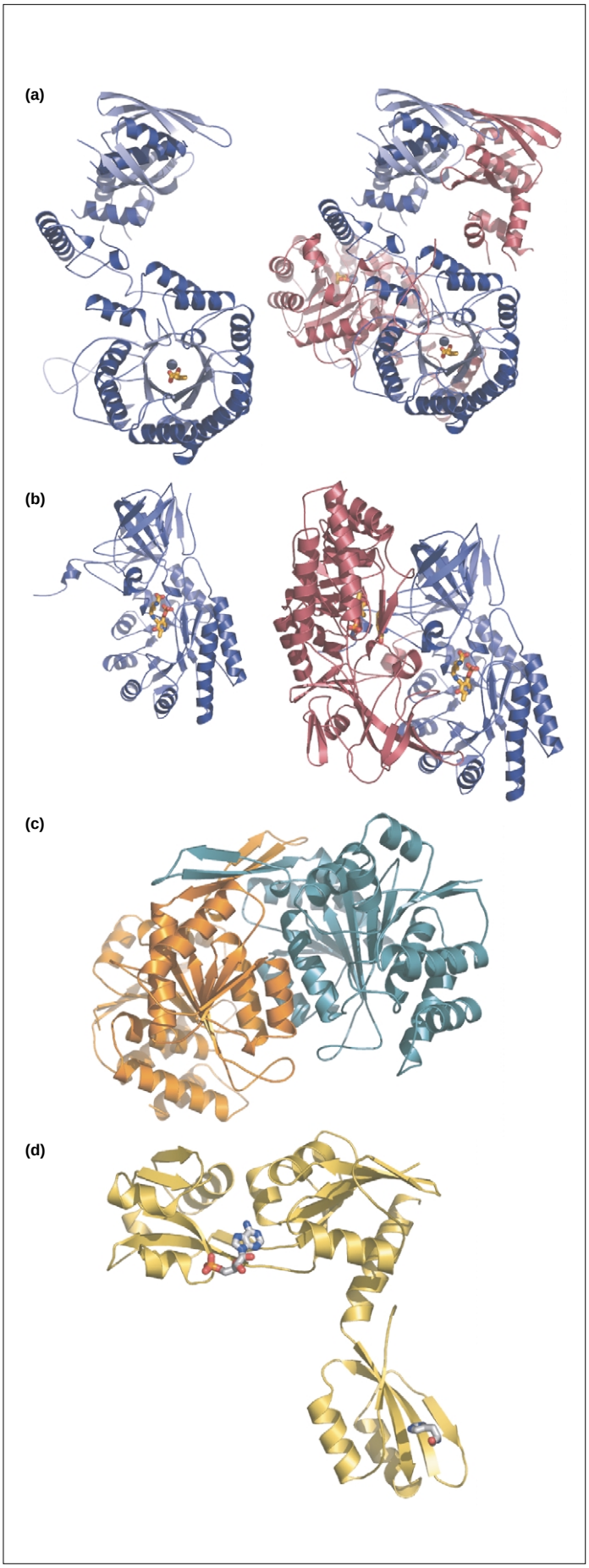

FIGURE 1

3D structures of selected amino acid biosynthesis enzymes from *M. tuberculosis*. **(a)** A ribbon representation of the structure of α-isopropylmalate synthase, LeuA [33]. On the left is the monomer showing the catalytic TIM-barrel domain and the active site containing a zinc ion (blue sphere) and α-ketoisovalerate (α-KIV, depicted as sticks). On the right is the biologically relevant dimer. The site for feedback regulation by leucine is between the two C-terminal domains at the top of the figure. Some loops are missing as these are disordered in the crystal. **(b)** The structure of meso-diaminopimelate decarboxylase, LysA [35]. This enzyme is also a dimer (shown on the right) and the monomer is shown on the left. The active site, which lies at one end of the TIM-barrel domain, is shown with bound lysine and pyridoxyl 5′-phosphate (PLP). **(c)** The structure of 3-isopropylmalate dehydrogenase, LeuB [34]. This enzyme is also a dimer (monomers are shown in orange and turquoise). **(d)** The structure of a monomer of ATP phosphoribosyltransferase, HisG [36], showing the active site which binds AMP and the regulatory domain which binds the product of the biosynthetic pathway, Histidine. This inhibited form of the enzyme is a hexamer in the crystal whereas the active enzyme is thought to exist primarily as a dimer in solution. This figure and subsequent figures were made using PyMOL (DeLano Scientific, www.pymol.org).

knocked out leucine biosynthesis could be a promising drug lead. For example, an *M. tuberculosis* leucine auxotroph, made by exchange of *leuD* for a truncated version of the gene, was shown to be unable to replicate in macrophages *in vitro* and to have a reduced capacity for infection in mice *in vivo* [37]. For LysA, structural homology between this enzyme and ornithine decarboxylase from *Trypanosoma brucei* (the causative agent of African sleeping sickness) revealed that inhibitors designed around α-difluoromethylornithine (DFMO) could be potential leads in drug discovery.

Cofactor biosynthesis

An example of the fulfilment of the complementary objectives of deriving functional information from structure and possible lead compound discovery is the recent structure of the open reading frame, designated as Rv1347c [38], from the laboratory strain of *M. tuberculosis*. This protein was originally annotated as an aminoglycoside *N*-acetyltransferase (AAC), with a supposed activity of acetylating amino groups on aminoglycoside antibiotics, such as streptomycin and kanamycin, thereby inactivating them. However, its sequence identity with known enzymes of this type is only 15%, and the postulated activity could not be demonstrated. Subsequently, the crystal structure showed that the protein fold clearly identified Rv1347c as belonging to the GCN5-related N-acetyltransferase (GNAT) family, to which the AACs belong (Figure 2). Rv1347c is an essential gene for the growth of *M. tuberculosis*, its expression is regulated by iron and its closest homologues are other bacterial proteins involved in the biosynthesis of siderophores (small molecule chelators used to acquire iron from the environment). Genes that flank Rv1347c in the genome are implicated in the biosynthesis of mycobactin, the siderophore used by *M. tuberculosis* to take up iron. Crucially, in the crystal structure, extra electron density attributed to a bound detergent molecule indicated a hydrophobic channel leading to the active site (Figure 2). The conclusion drawn from these structural observations, which has since been verified experimentally, was that Rv1347c is a 'missing' enzyme of mycobactin biosynthesis that adds a long-chain acyl group to the *N*-hydroxylysine side chain of mycobactin [38].

Cofactor biosynthesis is a fertile area for alternative antibiotic targets. Structures have been determined for the enzymes that catalyse the first and last steps in the biosynthesis of pantothenate

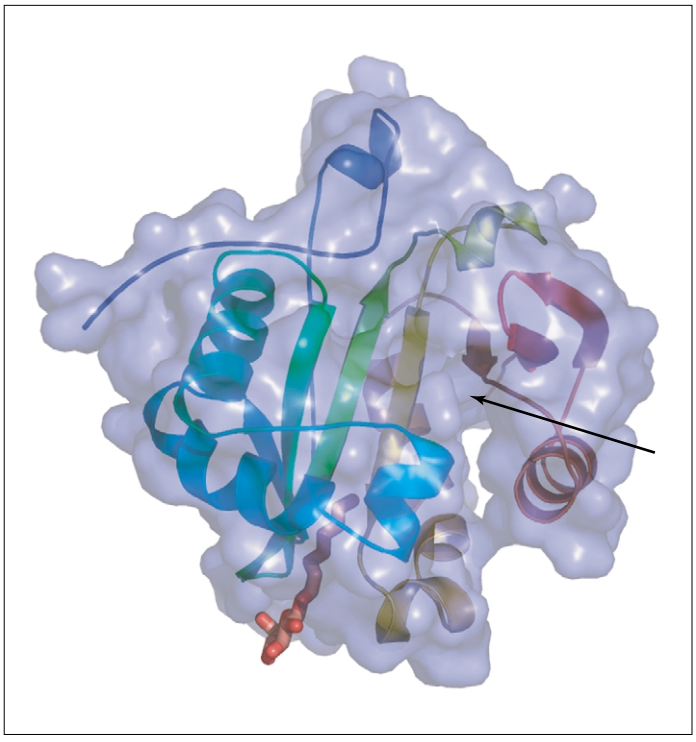

FIGURE 2

The structure of Rv1347c, a siderophore biosynthesis enzyme [38]. The structure of Rv1347c is depicted as a ribbon which is coloured blue-to-red from the N-terminus to the C-terminus. The surface of the protein is shown in blue and is partially transparent. The fortuitous binding of a molecule of detergent (β-octylglucoside) is depicted as a stick model in pink and red. This shows the location of a hydrophobic channel in the enzyme. At the bottom-right of the molecule, a second hydrophobic channel, which is thought to bind acyl-CoA is indicated by an arrow and a 'hole' or 'tunnel' beneath it.

(vitamin B5), an essential precursor for the biosynthesis of Coenzyme A: PanB (ketopantoate hydroxymethyltransferase) [39] and PanC (pantothenate synthase) [40]. These enzymes are attractive targets because they are not present in humans. However, this need not be a prerequisite for drug discovery because some classic drug targets are ubiquitous. For example, dihydrofolate reductase (DHFR, encoded by *dfrA*) is an NADPH-dependent enzyme whose product is a precursor for the synthesis of nucleotides and the amino acids methionine, serine and glycine. DHFR-knockouts are usually lethal but because this enzyme is present in prokaryotes and eukaryotes, including mammals, there is an added complexity in trying to design species-specific inhibitors to this enzyme. Nevertheless, DHFR has been a target for drug discovery in fungal and bacterial infections together with some autoimmune conditions [41]. The structure of *M. tuberculosis* DHFR [42] in complex with the drug methotrexate, the antimicrobial trimethoprim and an antimalarial compound revealed some important differences between the *M. tuberculosis* enzyme and its human homologue, including a pocket in the active site that contained a bound glycerol molecule from the crystallization matrix. The equivalent position in the active site of the human enzyme is occupied by amino acid side chains [42]. Subsequently, a screen using *Saccharomyces cerevisiae* whose survival has been engineered to be dependent on *M. tuberculosis* DHFR has shown that a group of potent anti-DHFR compounds are possible drug leads [43].

Secreted proteins

Secreted proteins form a group of important targets because of their likely role in host–pathogen interactions. For drug development, they have the major advantage of being extracellular. Examples for *M. tuberculosis* protein structures that have been determined in this class include the mycolyl transferases, also called antigens 85A, 85B and 85C [44–46], and an adaptin-like secreted protein of unknown function MPT63 [47].

Persistence

One of the principal goals of the Global Alliance for TB Drug Discovery [7] is the development of drugs that treat persistent TB infection. This is particularly challenging because the bacilli lie dormant, and protein targets that are typical for many antibiotics, such as the translational machinery of the cell, only operate at a basal level in this state. In an effort to specifically target bacilli in the persistent phase, several groups have determined protein structures for enzymes in the glyoxylate shunt pathway [48,49]. This pathway is thought to be crucial for *M. tuberculosis* survival in the persistent phase of infection. The enzymes isocitrate lyase (ICL, encoded by *icl*) and malate synthase are the first and second metabolic enzymes on the glyoxylate shunt pathway for the production of glyoxylate and succinate from isocitrate. In both cases, protein structures were solved at high resolution [50,51] with and without substrates (or inhibitors) and the attributes of the active sites were thus clearly defined. For example, the structure of isocitrate lyase (Figure 3) with its inhibitor 3-nitropropionate and glyoxylate showed a large conformational change upon substrate binding [50]. Using this detailed structural information, inhibitors for isocitrate lyase are in the discovery phase of drug development at GlaxoSmithKline [7] and recently ICL inhibitors have been shown to be effective against *M. tuberculosis* growth in macrophages [52]. Inhibitor screening in search of lead compounds for malate synthase is also ongoing.

Careful comparative genomics has revealed a second isocitrate lyase gene, *aceA*, in the clinical strain of TB (CDC1551) that, as a result of a mutation, is a pseudogene in the laboratory-strain (H37Rv) [49]. This second isocitrate lyase is considerably longer than is typical and its biological function is unknown. It has 23% amino acid sequence identity with *icl* and retains the sequence motifs which are characteristic of isocitrate lyase activity. The functions of the two genes, *icl* and *aceA,* are known not to be redundant. In this context, the structure of *aceA* will present atomic-resolution detail for a possible new drug target together with valuable insight into the biological function of this protein.

Mycolic acid biosynthesis

Mycolic acids are central to the lifestyle and survival of mycobacteria. These long chain fatty acids are a significant constituent of the cell envelope and confer properties, such as impermeability and resistance to dehydration and lysosomal processes, to the structure of the cell envelope. Indeed, one of the difficulties in the treatment of TB is the discovery of compounds that are able to traverse the unusual cell wall of *M. tuberculosis*. The front line antibiotic isoniazid targets one of the enzymes in mycolic acid biosynthesis, InhA, and thus this metabolic pathway is seen as a fruitful avenue for the development of new antibiotics. Structures have been determined for three mycolic acid cyclopropane

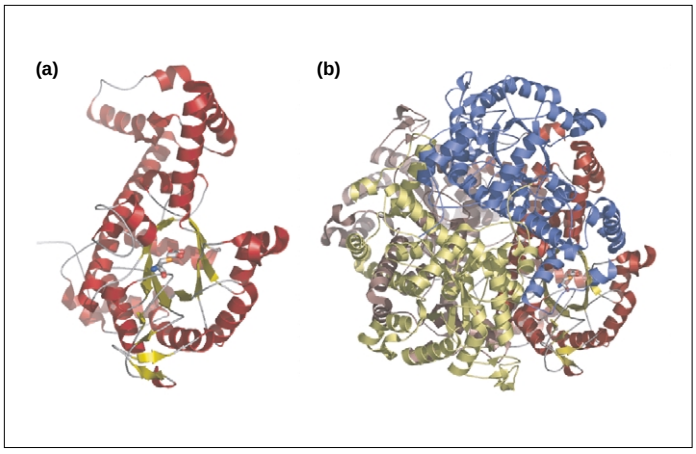

FIGURE 3

3D structure of isocitrate lyase from *M. tuberculosis* [50]. The monomer of isocitrate lyase is shown on the left as a ribbon diagram with α-helices in red and β-sheets in green. Similarly to LeuA and LysA from Figure 1, the catalytic domain is a TIM-barrel domain. The active site is shown by the covalent binding of the inhibitor 3-bromopyruvate to cysteine-191 (shown as sticks). The biologically active tetramer is shown on the right.

synthase enzymes, PcaA, CmaA1 and CmaA2 [53]. Each enzyme is specific for cyclopropanation at different points and/or in different configurations, along the hydrophobic chain of mycolic acids. In the case of PcaA, the gene encoding this protein has been identified as having a role in persistence and virulence in an animal model. The structures for all three show that they belong to the *S*-adenosyl-methionine-dependent methyltransferase superfamily. The structures have also revealed detergent molecules bound to the synthases, which points to the possible site and orientation of binding for the long hydrophobic mycolic acids. The elucidation of precise geometries and topologies for the binding pockets for these enzymes is the first step in designing inhibitors.

Other targets

Efforts are currently underway to obtain structural information for a group of 11 putative serine/threonine kinases that could have a key role in host–pathogen interactions [54,55]. A second group of proteins (unique to mycobacteria) that are more recalcitrant to structural biology are the large cohort of gene products referred to as PE and PPE proteins (for their Pro-Glu and Pro-Pro-Glu repeats) [18], none of which has yet been structurally elucidated. It is an exciting prospect that these proteins might have their structures determined in the near future, and that this could shed light on their biological function and their amenability to inhibition.

Conclusions

Structure-based drug design has a long history but, to date, has had only a moderate impact on bringing new drugs to market. This has been caused by the long lead times for successful transposition of structural knowledge to marketable drugs. Current estimates for the number of drugs that are derived from structure-based approaches are modest – between 10 and 40 drugs in clinical trials or in the marketplace [13,17]. However, evidence now suggests that the field has matured and that rapid progress is underway. The level of investment in structural biology and structure-oriented bioinformatics, by large and small pharma and biotech companies, supports this assertion. This has also had an impact on drugable *M. tuberculosis* protein structures. In 2000 only eight protein structures from *M. tuberculosis* were available. Today, more than 200 structures can be found in the Protein Data Bank and although some are duplicates or ligand complexes, structures are now available for more than 100 unique *M. tuberculosis* proteins. About two-thirds of these were determined within the *M. tuberculosis* Structural Genomics Consortium, and it is clear that genomic knowledge of the *M. tuberculosis* complex has stimulated many research groups worldwide, inside and outside the *M. tuberculosis* Structural Genomics Consortium. Many of the structures are for proteins that are potential drug targets, shown by gene knockout to be essential for *M. tuberculosis* growth. In one case, that of ICL, drug development based on the protein structure is in a commercial development pipeline.

The availability of more than 100 *M. tuberculosis* protein structures (a number that is increasing monthly), their cloned genes and protein production protocols, will be of huge benefit for combating this devastating pathogen. However, some practical challenges for structural genomics and structure-based drug development remain. Up to now, typically only 30–40% of *M. tuberculosis* proteins are obtained in soluble form when expressed in *E. coli*, even after the exclusion of membrane proteins from the targeted set of open reading frames. This severely limits the possibilities for structure determination. Some proteins could be rescued by using alternative expression hosts, or co-expressing with interacting partners. However, with increasingly sophisticated and automated methodologies (a legacy of global structural genomics efforts) many of these proteins are being successfully interrogated by structural biology, and drug development is poised to benefit dramatically from this structural information. A recent success story in the fight against tuberculosis is the discovery of a small molecule ATP synthase inhibitor which has potent antimycobacterial properties [56]. Although this drug was not discovered using a structure-based route, the marriage of recent successes in the structural studies of ATP synthase [57,58] together with the development of this new TB antibiotic has exciting possibilities for the successful treatment of this human pathogen in the near future.

References

1 Dye, C. *et al.* (1999) Global burden of tuberculosis: estimated incidence prevalence, and mortality by country. *JAMA* 282, 677–686
2 Havlir, D.V. and Barnes, P.F. (1999) Tuberculosis in patients with human immunodeficiency virus infection. *N. Engl. J. Med.* 340, 367–373
3 Sonnenberg, P. *et al.* (2005) How soon after infection with HIV does the risk of tuberculosis start to increase? A retrospective cohort study in South African gold miners. *J. Infect. Dis.* 191, 150–158
4 World Health Organisation, W.H.O. (1993) Tuberculosis on the increase. *World Health Forum* 14, 1–94
5 Dye, C. *et al.* (2002) Worldwide incidence of multidrug-resistant tuberculosis. *J. Infect. Dis.* 185, 1197–1202
6 DeRiemer, K. *et al.* (2005) Does DOTS work in populations with drug-resistant tuberculosis? *Lancet* 365, 1239–1245
7 Global Alliance for TB Drug Discovery (2003) *Annual Report 03/04*
8 von Itzstein, M. *et al.* (1993) Rational design of potent sialidase-based inhibitors of influenza virus replication. *Nature* 363, 418–423
9 Kim, C.U. *et al.* (1998) Structure–activity relationship studies of novel carbocyclic influenza neuraminidase inhibitors. *J. Med. Chem.* 41, 2451–2460
10 Kaldor, S.W. *et al.* (1997) Viracept (nelfinavir mesylate, AG1343): a potent, orally bioavailable inhibitor of HIV-1 protease. *J. Med. Chem.* 40, 3979–3985
11 Kim, E.E. *et al.* (1995) Crystal structure of HIV-1 protease in complex with VX-478, a potent and orally bioavailable inhibitor of the enzyme. *J. Am. Chem. Soc.* 117, 1181–1182
12 Sham, H.L. *et al.* (1998) ABT-378, a highly potent inhibitor of the human

immunodeficiency virus protease. *Antimicrob. Agents Chemother.* 42, 3218–3224
13 Congreve, M. *et al.* (2005) Structural biology and drug discovery. *Drug Discov. Today* 10, 895–907
14 Shuker, S.B. *et al.* (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274, 1531–1534
15 Fejzo, J. *et al.* (1999) The SHAPES strategy: an NMR-based approach for lead generation in drug discovery. *Chem. Biol.* 6, 755–769
16 Nienaber, V.L. *et al.* (2000) Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat. Biotechnol.* 18, 1105–1108
17 Hardy, L.W. and Malikayil, A. (2003) The impact of structure-guided drug design on clinical agents. *Curr. Drug Discov.* December, 15–20
18 Cole, S.T. *et al.* (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* 393, 537–544
19 Camus, J.C. *et al.* (2002) Re-annotation of the genome sequence of *Mycobacterium tuberculosis* H37Rv. *Microbiology* 148, 2967–2973
20 Fleischmann, R.D. *et al.* (2002) Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains. *J. Bacteriol.* 184, 5479–5490
21 Brosch, R. *et al.* (2002) A new evolutionary scenario for the *Mycobacterium tuberculosis* complex. *Proc. Natl. Acad. Sci. U. S. A.* 99, 3684–3689
22 Cole, S.T. (2002) Comparative and functional genomics of the *Mycobacterium tuberculosis* complex. *Microbiol.* 148, 2919–2928
23 Zarembinski, T.I. *et al.* (1998) Structure-based assignment of the biochemical function of a hypothetical protein: a test case of structural genomics. *Proc. Natl. Acad. Sci. U. S. A.* 95, 15189–15193
24 Kim, S.H. (1998) Shining a light on structural genomics. *Nat. Struct. Biol.* 5(*Suppl*), 643–645
25 Vitkup, D. *et al.* (2001) Completeness in structural genomics. *Nat. Struct. Biol.* 8, 559–566
26 Terwilliger, T.C. *et al.* (2003) The TB structural genomics consortium: a resource for *Mycobacterium tuberculosis* biology. *Tuberculosis (Edinb.)* 83, 223–249
27 Goulding, C.W. *et al.* (2003) Structural genomics of *Mycobacterium tuberculosis*: a preliminary report of progress at UCLA. *Biophys. Chem.* 105, 361–370
28 Moreland, N. *et al.* (2005) A flexible and economical medium-throughput strategy for protein production and crystallization. *Acta Crystallogr. D Biol. Crystallogr.* 61, 1378–1385
29 Studier, F.W. (2005) Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* 41, 207–234
30 Sassetti, C.M. *et al.* (2003) Genes required for mycobacterial growth defined by high density mutagenesis. *Mol. Microbiol.* 48, 77–84
31 Wilson, M. *et al.* (1999) Exploring drug-induced alterations in gene expression in *Mycobacterium tuberculosis* by microarray hybridization. *Proc. Natl. Acad. Sci. U. S. A.* 96, 12833–12838
32 Sherman, D.R. *et al.* (2001) Regulation of the *Mycobacterium tuberculosis* hypoxic response gene encoding alpha-crystallin. *Proc. Natl. Acad. Sci. U. S. A.* 98, 7534–7539
33 Koon, N. *et al.* (2004) Crystal structure of LeuA from Mycobacterium tuberculosis, a key enzyme in leucine biosynthesis. *Proc. Natl. Acad. Sci. U. S. A.* 101, 8295–8300
34 Singh, R.K. *et al.* (2005) The high-resolution structure of LeuB (Rv2995c) from *Mycobacterium tuberculosis*. *J. Mol. Biol.* 346, 1–11
35 Gokulan, K. *et al.* (2003) Crystal structure of *Mycobacterium tuberculosis* diaminopimelate decarboxylase, an essential enzyme in bacterial lysine biosynthesis. *J. Biol. Chem.* 278, 18588–18596
36 Cho, Y. *et al.* (2003) Crystal Structure of ATP Phosphoribosyltransferase from Mycobacterium tuberculosis. *J. Biol. Chem.* 278, 8333–8339
37 Hondalus, M.K. *et al.* (2000) Attenuation of and protection induced by a leucine auxotroph of *Mycobacterium tuberculosis*. *Infect. Immun.* 68, 2888–2898
38 Card, G.L. *et al.* (2005) The crystal structure of Rv1347c, a putative antibiotic resistance protein from *Mycobacterium tuberculosis*, reveals a GCN5-related fold and suggests an alternative function in siderophore biosynthesis. *J. Biol. Chem.* 280, 13978–13986
39 Chaudhuri, B.N. *et al.* (2003) The crystal structure of the first enzyme in the pantothenate biosynthetic pathway, ketopantoate hydroxymethyltransferase, from *M. tuberculosis*. *Structure* 11, 753–764
40 Wang, S.S. and Eisenberg, D. (2003) Crystal structures of a pantothenate synthetase from *M. tuberculosis* and its complexes with substrates and a reaction intermediate. *Protein Sci.* 12, 1097–1108
41 Schweitzer, B.I. *et al.* (1990) Dihydrofolate reductase as a therapeutic target. *FASEB J.* 4, 2441–2452
42 Li, R. *et al.* (2000) Three-dimensional structure of M-tuberculosis dihydrofolate reductase reveals opportunities for the design of novel tuberculosis drugs. *J. Mol. Biol.* 295, 307–323
43 Gerum, A.B. *et al.* (2002) Novel Saccharomyces cerevisiae Screen Identifies WR99210 Analogues That Inhibit *Mycobacterium tuberculosis* Dihydrofolate Reductase. *Antimicrob. Agents Chemother.* 46, 3362–3369
44 Ronning, D.R. *et al.* (2000) Crystal structure of the secreted form of antigen 85C reveals potential targets for mycobacterial drugs and vaccines. *Nat. Struct. Biol.* 7, 141–146
45 Ronning, D.R. *et al.* (2004) *Mycobacterium tuberculosis* antigen 85A and 85C structures confirm binding orientation and conserved substrate specificity. *J. Biol. Chem.* 279, 36771–36777
46 Anderson, D.H. *et al.* (2001) 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, 671–681
47 Goulding, C.W. *et al.* (2002) Crystal structure of a major secreted protein of *Mycobacterium tuberculosis* – MPT63 at 1.5-angstrom resolution. *Protein Sci.* 11, 2887–2893
48 McKinney, J.D. *et al.* (2000) Persistence of *Mycobacterium tuberculosis* in macrophages and mice requires the glyoxylate shunt enzyme isocitrate lyase. *Nature* 406, 735–738
49 Smith, C.V. *et al.* (2004) TB drug discovery: addressing issues of persistence and resistance. *Tuberculosis (Edinb.)* 84, 45–55
50 Sharma, V. *et al.* (2000) Structure of isocitrate lyase, a persistence factor of *Mycobacterium tuberculosis*. *Nat. Struct. Biol.* 7, 663–668
51 Smith, C.V. *et al.* (2003) Biochemical and structural studies of malate synthase from *Mycobacterium tuberculosis*. *J. Biol. Chem.* 278, 1735–1743
52 Munoz-Elias, E.J. and McKinney, J.D. (2005) *Mycobacterium tuberculosis* isocitrate lyases 1 and 2 are jointly required for *in vivo* growth and virulence. *Nat. Med.* 11, 638–644
53 Huang, C.C. *et al.* (2002) Crystal structures of mycolic acid cyclopropane synthases from *Mycobacterium tuberculosis*. *J. Biol. Chem.* 277, 11559–11569
54 Good, M.C. *et al.* (2004) Sensor domain of the *Mycobacterium tuberculosis* receptor Ser/Thr protein kinase, PknD, forms a highly symmetric beta propeller. *J. Mol. Biol.* 339, 459–469
55 Young, T.A. *et al.* (2003) Structure of *Mycobacterium tuberculosis* PknB supports a universal activation mechanism for Ser/Thr protein kinases. *Nat. Struct. Biol.* 10, 168–174
56 Andries, K. *et al.* (2005) A diarylquinoline drug active on the ATP synthase of *Mycobacterium tuberculosis*. *Science* 307, 223–227
57 Mueller, D.M. *et al.* (2004) Crystallization and preliminary crystallographic studies of the mitochondrial F1-ATPase from the yeast *Saccharomyces cerevisiae*. *Acta Crystallogr. D Biol. Crystallogr.* 60, 1441–1444
58 Leslie, A.G. *et al.* (1999) The structure of bovine mitochondrial F1-ATPase: an example of rotary catalysis. *Biochem. Soc. Trans.* 27, 37–42