Strategies and Challenges Involved in the Discovery of New Chemical Entities During Early-Stage Tuberculosis Drug Discovery

Geoffrey D. Coxon,1 Christopher B. Cooper,2 Stephen H. Gillespie,3 and Timothy D. McHugh4

1Strathclyde Institute for Pharmacy and Biomedical Sciences, University of Strathclyde, Glasgow, United Kingdom; 2The Global Alliance for TB Drug Development, New York, New York; 3School of Medicine, University of St Andrews; and 4Centre for Clinical Microbiology, University College London, United Kingdom

There is an increasing flow of new antituberculosis chemical entities entering the tuberculosis drug development pipeline. Although this is encouraging, the current number of compounds is too low to meet the demanding criteria required for registration, shorten treatment duration, treat drug-resistant infection, and address pediatric tuberculosis cases. More new chemical entities are needed urgently to supplement the pipeline and ensure that more drugs and regimens enter clinical practice. Most drug discovery projects under way exploit enzyme systems deemed essential in a specific Mycobacterium tuberculosis biosynthetic pathway or develop chemical scaffolds identified by phenotypic screening of compound libraries, specific pharmacophores or chemical clusters, and natural products. Because the development of a compound for treating tuberculosis is even longer than for treating other infection indications, the identification of selective, potent, and safe chemical entities early in the drug development process is essential to ensure that the pipeline is filled with new candidates that have the best chance to reach the clinic.

THE DRUG DEVELOPMENT PIPELINE

The overall process involved in identifying a new chemical entity and developing it as a new drug can take up to 15 years and cost from $800 million to $1 billion (Figure 1) [1]. The early discovery phase of drug development is arguably the most technically challenging but the least expensive part of the process. In the case of antituberculosis agents, at this early stage the aim is to identify specific, low-molecular-weight molecules called “leads” that are bactericidal against Mycobacterium tuberculosis, acting at a known site of action, and that have desirable physicochemical, pharmacokinetic, and safety properties that warrant further investigation in more costly preclinical development studies. In the preclinical phase, the compounds will typically be evaluated in 2 animal species with more detailed safety and drug metabolism investigations. The final clinical phases are very expensive, and an effective selection process is essential to identify the few compounds worth this significant investment. Of note, for drug-susceptible tuberculosis, despite the high costs involved in identifying and developing a new antituberculosis agent, such agents must also compete economically with current tuberculosis treatment regimens, which cost approximately $20 for standard 6-month treatment. This article briefly reviews the strategies and techniques available to the tuberculosis drug development community to discover new potential drug compounds and discusses their various merits in this role.

How does the search for a new lead chemical entity occur? For antibiotics, it can broadly be divided into 2 strategies. The first, referred to as either the classic, whole cell, empiric, or phenotypic approach, aims to identify compounds that kill the mycobacteria after screening of compounds or mixtures of compounds. The second strategy has been significantly exploited...
during the past decade after the availability of the genome sequences of *M. tuberculosis* [2] and related organisms [3, 4] and targets gene products directly with use of the expressed proteins as templates to design new inhibitors. Known as the target-based approach, the rationale is to develop inhibitors of essential and validated enzymes so that they will kill the mycobacteria through their specific action on the targeted enzyme. The aim of this perspective is to give the reader a broad overview of the methods used in both approaches and discuss their merits in the context of the specific challenges of tuberculosis drug development.

**TARGET-BASED APPROACHES**

The dawn of the genomic era in the late 1990s heralded the possibility of tailor-made and specifically targeted drugs designed against exquisite enzymes known to be essential for critical biosynthetic pathways of the mycobacteria. After a specific enzyme target has been validated and a robust practical assay has been identified to evaluate large numbers of compounds, molecular modeling coupled with structural biology has become the standard starting point. In a perfect scenario, the protein under investigation would be highly soluble and a high-resolution X-ray crystal structure with the natural substrate bound would be available. In general, structure-based molecular modeling techniques may be exploited in which key amino acid residues serve as a template that can be used to either identify new ligands or dock compounds to guide the medicinal chemist’s research program [5–7]. Pharmacophore (the fundamental 3-dimensional [3-D] molecular arrangement required for activity) models may also be generated to enable the virtual screening of hundreds of thousands of molecules, which, after refinement with predefined software parameters that enable the discarding of compounds with undesirable physicochemical properties, will lead to a relatively small number of compounds that can be accessed and assessed in the assay [8]. In both cases, a close working relationship with structural biologists may enable the co-crystal soaking of identified compounds with the protein from which the 3-D crystal structures can allow an iterative process to occur, in which ligands are optimized to achieve high-target affinity.

**Structure-Based Design**

As examples in tuberculosis drug discovery, structure-based design efforts have successfully identified inhibitors of pantothenate synthetase, a known target that may be useful for nonreplicating persistent forms of *M. tuberculosis*. After the screening of compound libraries and docking studies using the crystal structure of the enzyme (PDB: 2A88), the lead compound 1 was discovered that inhibited the enzyme with a half maximal inhibitory concentration (IC$_{50}$) of 90 nM (Figure 3) [9]. A similar approach was used to identify inhibitors of AccD5, an acyl-CoA carboxyltransferase that commits acyl-CoA fatty acids to the biosynthesis of essential cell wall lipids called mycolic acids. High-throughput library screening of compounds with use of a structure-based pharmacophore model based on the crystal structure of the enzyme (PDB: 2A7S), followed by structure-based design studies, have led to the identification of lead compound 2, which binds to the enzyme with a K$_i$ (the dissociation constant for the inhibitor binding to the enzyme) of 13.1 µM (Figure 3) [10].

When a key enzyme has been identified but the 3-D crystal structure has not been solved and a protein with similar function from a related organism exists, homology modeling may be exploited. This technique has been used in the identification of inhibitors of DevR from *M. tuberculosis* toward developing compounds against nonreplicating persistent forms of *M. tuberculosis* (Figure 3) [11]. In this investigation, the related crystal structures of NarL from *Escherichia coli* (PDB: 1AO4) and DosR from *M. tuberculosis* (PDB: 1ZLK) were used as templates for homology modeling. The identification of a pocket for docking, followed by molecular dynamics simulations and pharmacophore generation and clustering of structural classes, led to the selection of 11 structurally diverse compounds for testing. Evaluation in vitro led to the identification of compound 3 with an IC$_{50}$ <26.2 µg/mL as a lead template for further design.

**Ligand-Based Design**

In the absence of any available 3-D crystal structure or when a data set of IC$_{50}$ or K$_i$ values are known for a range of inhibitors, ligand-based pharmacophore models may be generated and used for the screening and design of inhibitors. In this strategy, the software builds data around diverse molecular functionality with the corresponding assay data to help guide and predict compound selection and optimization. Agrawal et al [12] recently reported the use of this method in the identification of initial enzyme hits against chorismate mutase (MtCM) from *M. tuberculosis*. This enzyme catalyzes the Claisen rearrangement of chorismate to prephenate in the shikimate biosynthetic pathway leading to the biosynthesis of the essential aromatic amino acids phenylalanine and tyrosine. Structures identified from these hits were then used in structure-based design with the crystal structure of MtCM (PDB:2F6L) to identify the lead compound 4 with a K$_i$ of 5.7 µM (Figure 3) [12].

**Fragment-Based Design**

When highly soluble and high diffraction quality enzymes are available for exploitation, a fragment-based approach may be used (Figure 2) [13–16]. In general, the principle with this technique is to soak soluble, low-molecular-weight (<250 Da) molecules at mM concentrations with the protein with the aim of finding weak binding molecular fragments. Because of their
inherently weak binding properties, these low-molecular-weight species need to be identified using nonstandard assay techniques, such as thermal shift, nuclear magnetic resonance spectroscopy, isothermal calorimetry, and X-ray crystallography. Once identified and using 3-D molecular modeling software, the fragments may then be optimized or tethered to other fragments bound in similar locations, to make stronger binding inhibitors. However, because this technique identifies more efficiently bound fragment starting points, compared with ligand optimization, the optimized inhibitors should possess more physicochemically acceptable properties than those discovered from conventional screening and optimization methods. To this end, this rationale has led to ligand efficiency, which is described as the free energy of ligand binding to an enzyme-binding site averaged for each nonhydrogen atom [13]. Thus, the emphasis with this technique is to prioritize molecules by the average binding efficiency rather than potency alone (Figure 2).

Examples of this technique have been used in tuberculosis drug discovery and may be revealed with the identification of inhibitors of the protein tyrosine phosphatases PtpA and PtpB. In this particular study, a fragment-based approach, substrate activity screening, was used to find low-molecular-weight potent inhibitors. From this research, compounds 5 and 6 were found to inhibit PtpA with \( K_i \) values of 1.4 and 1.6 \( \mu M \), respectively [17, 18]. Compound 7 was identified using substrate activity screening and was found to bind to PtpB with a \( K_i \) of 0.22 \( \mu M \) with subsequent structure-based design leading to the identification of the analogue 8, which bound to enzyme with a \( K_i \) value of 0.69 \( \mu M \) [19, 20].

The main advantage of the target-based approach is that the compound series identified as inhibitors of a particular enzyme can be optimized with the secure knowledge that, after genetic validation in vitro and in vivo, their action arises through the targeted protein and that molecular selectivity can be achieved through design. This is of enormous value when progressing a series from hit to lead and from lead to candidate to diminish polypharmacy potential or unwanted safety issues being encountered in later-stage drug development. The experience of this approach when applied to antibiotic and antituberculosis drug discovery in particular has not yet yielded a drug in clinical trials and, in this regard, can be regarded as largely unsuccessful [21, 22].

The limitation with the target-based approach may be attributed to the lack of known targets available for exploitation that have a positive therapeutic effect in patients after their progression from hit to lead and from lead to candidate to diminish polypharmacy potential or unwanted safety issues being encountered in later-stage drug development. The experience of this approach when applied to antibiotic and antituberculosis drug discovery in particular has not yet yielded a drug in clinical trials and, in this regard, can be regarded as largely unsuccessful [21, 22].

The limitation with the target-based approach may be attributed to the lack of known targets available for exploitation that have a positive therapeutic effect in patients after their progression from hit to lead and from lead to candidate to diminish polypharmacy potential or unwanted safety issues being encountered in later-stage drug development. The experience of this approach when applied to antibiotic and antituberculosis drug discovery in particular has not yet yielded a drug in clinical trials and, in this regard, can be regarded as largely unsuccessful [21, 22].
function has been inhibited by a drug molecule [21] and that, when a validated target has been exploited, there has been a lack of success in translating potent and selective enzyme activity into whole cell activity.

The target-based approach is heavily reliant on expensive equipment (e.g., nuclear magnetic resonance spectrometers and X-ray diffractometers), software licences, computational hardware, and skilled personnel, making it resource intensive. In particular, fragment based discovery is particularly dependent on these resources, limiting the ability of many research institutes, particularly in academic settings, to capitalize on the potential of this formidable method.

**Figure 3.** Lead compounds identified from target-based approaches exploiting *Mycobacterium tuberculosis* targets; 1–3 were identified using structure-based screening design against pantethenate synthase (PS), Acetyl-CoA carboxylase (AccD5) and DevR, respectively; 4 was identified using ligand-based screening followed by structure-based design against chorismate mutase (MtCM); 5–6 were identified using fragment-based techniques against protein tyrosine phosphates PtpA and 7 against PtpB; 8 was identified against PtpB using structure-based approaches based on the structure of 7.

**Figure 4.** Examples of compounds identified from phenotypic screening methods. In phase 2 clinical trials are TMC207 (bedaquiline) 9, PA-824 10, OPC-67683 (delamanid), and SQ109 12. In early-phase drug discovery (minimum inhibitory concentration [MIC] values showing activity against *Mycobacterium tuberculosis* H37Rv) 13 is BTZ043 (S enantiomer)/BTZ044 (R enantiomer); 14a,b are leading 2-aminothiazole-4-carboxylates (ATC), and 15 is a macrolide derivative.
PHENOTYPIC-BASED APPROACHES

Despite the advanced methodologies that have evolved from the genetic revolution, there is an increasing trend toward returning to phenotypic screening strategies that were so successful in the past [23]. Efforts here focus on the identification and subsequent exploration of specific compound classes to find compounds that exhibit activity against the whole organism rather than producing a highly engineered inhibitor for a specific target in a cell-free system.

Typically, investigators assess libraries of pure compounds or, in some cases, a mixture of compounds that may have been isolated from natural products, at known concentrations for their ability to kill the mycobacteria. The sufficiently active compounds are then assessed for cytotoxicity against human cell lines to identify leads early in the process that may prove to be problematic in later development with respect to safety. A number of in vitro techniques have been adopted by individual research groups or by focused programs, such as the National Institute of Allergy and Infectious Diseases–sponsored Tuberculosis Antimicrobial Acquisition and Coordination Facility or the Institute for Tuberculosis Research at the University of Illinois. Such methods include using agar plate or broth microdilution techniques [24] or the high-throughput percentage growth inhibition microplate Alamar Blue assay [25].

The success of this approach can be best shown with the discovery of the diarylquinoline TMC207 (bedaquiline) 9 (Figure 4) [26]. Discovered by Johnson & Johnson, this compound was originally identified from a library screening campaign against Mycobacterium smegmatis and then was developed as an antituberculosis agent. It is highly active against both drug-susceptible and drug-resistant strains of M. tuberculosis and is now in phase 2 clinical trials. Similar success has been seen with the development of the nitroimidazoles PA824 10 (Figure 4) [27] and OPC-67683 (delamanid) 11 (Figure 4) [28] and the diamine SQ-109 12 [29], which are also in phase 2 clinical trials.

In the early discovery phase of drug development, other recent examples of novel active series include the benzothiazinones (BTZ) 13 (Figure 4) [30], identified from the New Medicines for Tuberculosis program, and the 2-aminothiazole-4-carboxylates 14a–14b compounds from the Tuberculosis Drug Discovery UK consortium (Figure 4) [24]. In addition, a number of screening campaigns performed against M. tuberculosis H37Rv in recent years have identified potential new lead compounds with an indication of both their antituberculosis activity and selectivity against human cell lines [31, 32], such as reported by Franzblau et al (Figure 4) [33].

A significant advantage of this strategy is that drug activity confirms that the cell is permeable and that passive diffusion mechanisms and the complex cell wall structures and transport mechanisms that are present in M. tuberculosis are not significant barriers for the particular compound(s) of interest. The mycobacteria also possess a number of perceived and poorly understood metabolic states that are evident in both intercellular and extracellular environments and that may contribute to the difficulty of obtaining whole cell activity from target-based efforts.

Of course in this scenario, the site(s) of action may not be known. However, unlike the lack of success in translating target activity into whole cell activity, there exist a number of techniques available to identify the sites of action, with noted successful examples of this approach [26].

Challenges remain when selecting a newly discovered scaffold for further development. Although the site of action can be discovered after the initial identification, this is not a trivial process and means that whole cell assays may be the only method available to generate an iterative program for some time. It often leads to difficulties in generating structure activity relationships in which sometimes a shotgun approach is needed to advance. It is also difficult to ensure that activity and selectivity are on target when developing a structural class or identifying a backup series that is very different to lead progression in target-based drug development programs in other disease areas.

The major limitation of the classical approach, however, is the stringent biological containment level 3 safety requirements needed for laboratories that assess compounds against M. tuberculosis, the long duration of the growth phase of this mycobacterium, and the fact that simple and accurate predictable models to aid compound selection do not exist. These problems have contributed to low numbers of research groups and a limited expertise base in the field. Surrogate mycobacteria, such as Mycobacteria aurum and M. smegmatis [26, 34], are often used to address this problem, and as observed with the discovery of TMC207, there is strong evidence to support this approach. Although one could argue that the use of surrogates adds an unnecessary delay in the development process, because eventually all compounds must be assessed against M. tuberculosis and, importantly, activity against surrogates is not necessarily a predictor of efficacy against M. tuberculosis, the use of a surrogate screen can certainly facilitate the hit triaging process and thus allow a more focused research effort to proceed expeditiously.

The next stage in the process is to evaluate the antimycobacterial activity in animal models. This provides the opportunity to test safety and activity and to begin to understand the effective dose range that might be used in future clinical studies. Although it might be argued that the guinea pig model provides a better mirror of human pathology, the mouse is conventionally used to test activity. The propensity of M. tuberculosis to have a dormant or persistent state is a complication that is thought to be responsible, in part, for the need for a prolonged duration of treatment. This has been somewhat
addressed by the development of new persistence in animal models [35]. The mouse model is a cheap and readily affordable indicator used to assess in vivo activity, and the results generated appear to translate well to human activity studies [36]. It typically takes 2–4 months to evaluate a compound class in mouse models [36, 37]. As such, the iteration and development of a compound series can be difficult and inefficient, and thus, a close working relationship between the chemist and the biologist is essential to effectively select compounds for this type of investigation.

A final hurdle exists that is directly attributed to the clinical applicability of the new drugs. The usual way in which this is managed is for an initial monotherapy study [38] to be performed in which the drug is given alone for up to 2 weeks. Frequent assessment of the sputum viable count makes it possible to determine whether the novel agent is capable of killing M. tuberculosis in patients and may give an indication of the most appropriate doses that should be used [39]. For any new compound series, consideration must be given to understanding routes of metabolism, because these agents will be administered in combination with other drugs to help prevent resistance and relapse and to ensure complete sterilization of the patient. Importantly, because it is known that rifampicin induces the cytochrome P450 system (CYP3A4), new drugs must ensure that their metabolism occurs through other systems. Moreover, this problem is further complicated by the fact that new tuberculosis drugs will often require co-administration, with patients taking antiretroviral treatment for human immunodeficiency virus infection and AIDS, and thus, attention must be given to developing series that will not result in the reduction of the effective use of these agents. If monotherapy studies detect useful activity and relevant safety and interaction studies prove to be supportive, larger regimen-based studies can be initiated, but this stage requires a step-change in the financial investment.

**SUMMARY**

Both the target-based and classic approaches to drug development for tuberculosis treatment have advantages and disadvantages in their applicability. However, it has not yet been fully established which of these should be used. It is clear that direct comparisons cannot be made among the strategies adopted in programs for other disease indications, such as cancer, obesity, and cardiovascular disease. The poorly understood nature of the M. tuberculosis organism, its complex pathology in the host, and the practical challenges that working with the organism brings to the drug discovery process strongly suggest that tuberculosis drug discovery researchers should approach drug development as a separate research area that should be complemented by strategies used in other fields of drug development after advanced to an appropriate point in the research and development continuum. Close working partnerships between medicinal chemists and knowledgeable and skilled tuberculosis research biologists are therefore critical to ensuring that research programs in tuberculosis drug development have the greatest chance of success.

Taking all of the arguments reviewed in this article into account, the time has come for the global research community to adopt a different mind-set when both judging tuberculosis drug discovery programs in consideration for research funding and in key scientific decision making when under way. Because of the urgency for new tuberculosis drugs, the time for this change is now upon the research community.

**Note**

*Potential conflicts of interest.* All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

**References**


