

Article

Evaluating Antimycobacterial Screening Schemes Using Chemical Global Positioning System-Natural Product Analysis

Muaaz Mutaz Alajlani ^{1,2,*} and Anders Backlund ¹

¹ Pharmacognosy Research Group, Department of Medicinal Chemistry—Faculty of Pharmacy, Uppsala University, BMC—Biomedical Center, Box 574, S-751 23 Uppsala, Sweden; anders.backlund@ilk.uu.se

² Department of Pharmaceutical Biology/Pharmacognosy, Institute of Pharmacy, University of Halle-Wittenberg, Hoher Weg 8, DE 06120 Halle (Saale), Germany

* Correspondence: muaaz.alajlani@pharmazie.uni-halle.de; Tel.: +46-73-8922722

Academic Editors: Marcello Locatelli, Simone Carradori, Andrei Mocan and Pascal Richomme
Received: 27 November 2019; Accepted: 14 February 2020; Published: 20 February 2020



Abstract: Most of the targeted discoveries in tuberculosis research have covered previously explored chemical structures but neglected physicochemical properties. Until now, no efficient prediction tools have been developed to discriminate the novelty of screened compounds at early stages. To overcome this deficit, a drastic novel approach must include physicochemical properties filters provided by Chemical Global Positioning System-Natural Product analysis (ChemGPS-NP). Three different screening schemes GSK, GVKBio, and NIAID provided 776, 2880, and 3779 compounds respectively and were evaluated based on their physicochemical properties and thereby proposed as deduction examples. Charting the physicochemical property spaces of these sets identified the merits and demerits of each screening scheme by simply observing the distribution over the chemical property space. We found that GSK screening set was confined to a certain space, losing potentially active compounds when compared with an in-house constructed 459 highly active compounds (active set), while the GVKBio and NIAID screening schemes were evenly distributed through space. The latter two sets had the advantage, as they have covered a larger space and presented compounds with additional variety of properties and activities. The in-house active set was cross-validated with MycPermCheck and SmartsFilter to be able to identify priority compounds. The model demonstrated undiscovered spaces when matched with Maybridge drug-like space, providing further potential targets. These undiscovered spaces should be considered in any future investigations. We have included the most active compounds along with permeability and toxicity filters as supplemented material.

Keywords: evaluation; screening; chemography; chemical property space; antimycobacteria; ChemGPS-NP; data evaluation; tuberculosis

1. Introduction

Mycobacterium tuberculosis, the causative agent of tuberculosis (TB), infects one-third of the world population with traces back in history [1]. Today, TB still represents one of the most challenging threats to global human health, with over 2 billion people harboring latent infection and more than 9 million new cases reported every year. More than half a million of these cases eventually develop multi-drug-resistant (MDR) TB, and in recent years, extensively drug-resistant TB (XDR-TB) has been more frequently reported [2]. Nearly 2 million deaths are estimated to occur yearly [3]. Despite continuous and massive screening for potential antituberculosis agents, no breakthrough has been reported for over 40 years. Only 22 approved drugs [4] are available, out of which four drugs are considered first-line therapy [5]. Most of these first-line drugs in TB treatment were discovered during

the 1950s and the 1960s, and only four drugs are today at phase three clinical trial. None are at phase one [6].

In late 2012, bedaquiline became the first novel TB drug approved in 40 years [7], due to the complexities that govern tuberculosis research. Bedaquiline itself is an orphan drug, and accelerated approval of the drug was sought, a type of temporary approval in the case of diseases lacking other viable treatment options. The challenge of meeting the expectations of this desired target product profile complicates drug discovery efforts. Considering how few drugs from the discovery stage successfully enter the TB clinical pipeline, an increased understanding of drug discovery hurdles should facilitate the development of novel intervention strategies [8]. Pharmaceutical companies are in constant search for new antituberculosis drugs with more favorable characteristics in terms of treatment period, toxicity reduction, targeting of MDR and even XDR, and possibly the ability to be coadministered with HIV medications.

Dobson has defined chemical space to consist of compounds that are characterized by a wide range of chemical descriptors [9]. From this perspective, chemical space is defined to be the total descriptor space that encompasses all the small carbon-based molecules that could in principle be created [10]. Such a definition provides a huge advantage over any existing library, as roughly more than 99.9% of all small molecules that can be imagined have not been prepared and tested [11]. Despite the urgency of tuberculosis, a total of 42 compounds are the only drugs present for treatment. Information includes mechanisms and activity profiles on all approved drugs used to treat tuberculosis, on drugs in clinical development for TB, and on some already approved drugs being investigated for potential use in TB [12,13].

In this paper, we aim to chart and investigate different screening schemes of antituberculosis and provide insights on their distribution over the physicochemical space. We have characterized the overall chemical property space of antituberculosis compounds provided by ChemGPS-NP to navigate, validate, and predict antituberculosis compounds by providing a model for comparison.

2. Results and Discussion

There are continuous attempts by several laboratories to search for global antituberculosis compounds and subfragments, but none has reported a breach using virtual screening. Most of the success was achieved from locally active subsets with drastic measurements; in other words, the database library must fit the purpose of the experiment before its selection for screening, and this is usually done either by ligand-based or structural-based screening. These strategies consist of quantitative structure–activity relations (QSAR), 3D-QSAR, and pharmacophores. The approach of the hit-and-trail method has limited the screening strategies to their known chemical moieties and with an extremely poor hit rate [14,15].

Chemical property space has no such disadvantages, as similarity or dissimilarity is based on physicochemical properties rather than topology of chemical structure alone. ChemGPS-NP mapping can be a key factor in shaping such space as it accounts for over 50% of data variance. It reflects 35 major properties and characters, and the results are produced as principle components values.

The chemical property space was charted by using the first three-dimensional values obtained from ChemGPS-NP for each compound. The first dimension (principal component one, PC1) represents size, shape, and polarizability; PC2 corresponds to aromatic and conjugation-related properties, while PC3 describes lipophilicity, polarity, and H-bond capacity. In that regard, each compound has a designated position in the chemical property space. Antituberculosis compounds from different data sets are scattered throughout the chemical property space, indicating a high diversity in form of physicochemical properties. This visualization method has constructed an easy way for data evaluation. The diversity in such properties is directly proportional to a higher discovery rate with potentially exclusive activities. In order to have more deducible indications, we have explored three screening schemes (GSK, GVKBio, and NIAID) from different suppliers using ChemGPS-NP. A great deal of variance was observed in accordance with the chemical property space of these sets but at

the same time a common volume amongst these three schemes (Negative PC1, Positive PC2, and Positive PC3) (Figure 1). The common volume was noted by a higher density in their chemical property space. The screening scheme that lacked distribution in the chemical property space will have the disadvantage of missing potential targets, and the assays used should be investigated to ensure removing any limitation in the procedures.

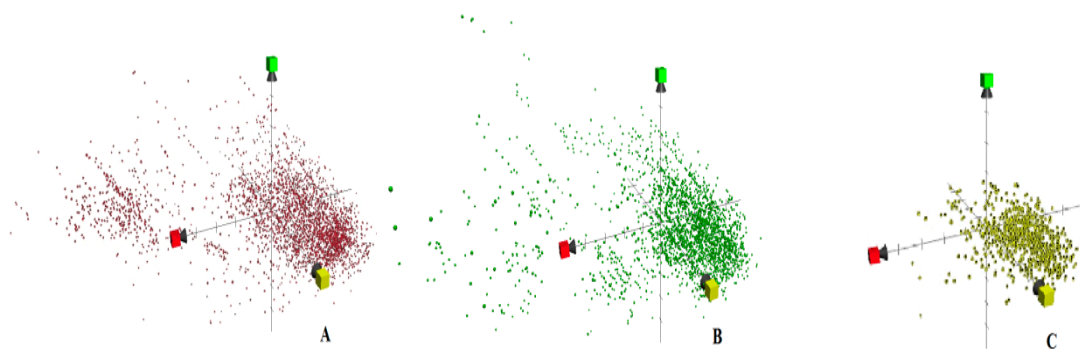


Figure 1. Different patterns revealed on plotting of (A) NIAID, (B) GVKBio, and (C) GSK on PC1 (x = red), PC2 (y = yellow) and PC3 (green) from ChemGPS-NP.

Obviously, variation is required to attain novel antituberculosis agents, but at the same time, we have observed that the GSK screening set was more confined to the common volume. This would indicate a safer or communal approach used by the company or a selective screening assay, especially that many active compounds were observed out in that volume. The proposed antimycobacterial chemical property space identified such merits and demerits alongside screening results and more usefully detected such problems and direct screening procedures towards their targets. High-throughput screening (HTS) and even Ultra HTS (uHTS) have become standard methods in the drug discovery [16]; unfortunately, they are still expensive techniques, as they require a library of semi-pure or preferably pure compounds at sufficient quantities, as well as more reliable assays that consider cost and validities. There are many compounds that have come from really expensive and time-consuming screen protocols, but they have produced very limited results, and no breakthrough was ever recorded in the field of tuberculosis research. Interestingly, this visualizing model presented in the chemical property space can indicate potential similarity that could be easily singled out from the proposed library, making screening enrichment a straightforward approach. The demerits in the GSK scheme were further demonstrated by incorporating it into the active set (459 compounds) and Maybridge screening collection (31,236 compounds) (Figure 2). Plotting these three sets together demonstrates the disadvantages of the confined GSK volume compared to NIAID and GVKBio schemes. As many highly active compounds (active set) were out of the range and far from GSK compounds. The 776 antituberculosis compounds of GSK have been proposed from GSK 2 million compounds yet missed potential active volumes in the highly active antituberculosis physicochemical space. The GSK example showed the importance of incorporating the physicochemical studies in screening programs, therefore, to cover all potential targets and to correct such deficiency at earlier stages.

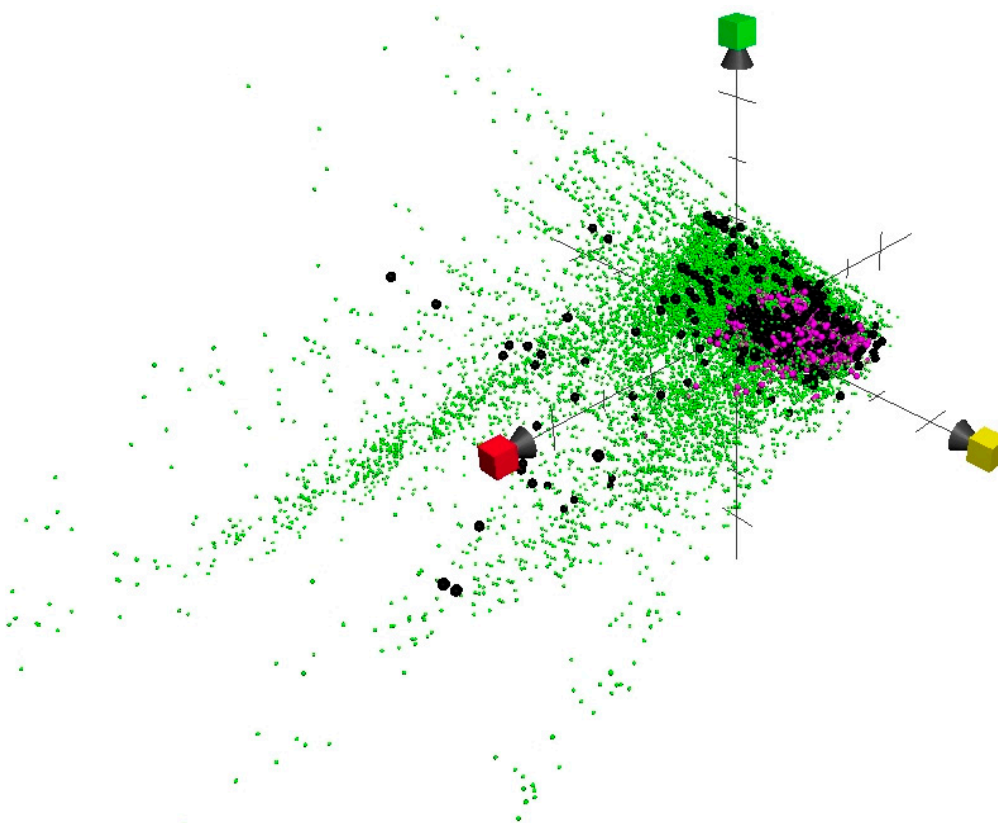


Figure 2. Chemical property space revealed as 3D Graphing (red axis = PC1, corresponding to size parameters; yellow axis = PC2, corresponding to aromaticity and conjugation related parameters; and green axis = PC 3, corresponding mainly to lipophilicity), compounds are contrasted by three colored sets; black representing antituberculosis compounds with highest activity, pink color represents GSK screening set, and green set represents the Maybridge screening collection. From this figure, it is obvious that the GSK did not cover potential targets presented by the active set (black spheres).

The concept of using chemical properties in virtual screening and data mining is becoming clearer to scientists. An extra dimension would not only enrich hit rate percentages but would also aid in distinguishing activities and contribute to the understanding of mechanisms of action by including the chemical properties [17]. Compounds tagged with the highest antituberculosis activity (black spheres) as plotted, along with Maybridge compound library, can be the bases of further selection. In this regard, we propose the 8-dimensional Euclidian distance calculation that further determines selection. Compounds with antituberculosis activity could be sorted according to Euclidean distances to determine the nearest highly active compound. The Maybridge screening collection (<http://www.maybridge.com>) is generally considered as consisting of drug-like compounds. These compounds have favorable drug properties and therefore may constitute a supreme database for screening. Two predictive models can be integrated, toxicity and permeability, using online application tools mentioned in material and methods. Potential antituberculosis compounds with higher activity can be identified from the Maybridge compound library by selecting those compounds that are closest to the active compounds (black spheres). We have used three computational models based on high throughput assays on *Mycobacterium tuberculosis* along with known and clinical trial drugs and the active set in addition we used a filter-based approach to filter out potential false positives/toxic molecules.

This straightforward approach is based on the fact that similarity in chemical properties is connected to the activity. The same principle can be employed for any database, simply by matching them with the compounds identified as most effective (black spheres in Figure 2). This approach will provide the chance to identify additional compounds with potential activities based on the similarity

in their physicochemical properties. The latter is of particular importance because antituberculosis compounds must have a favorable pharmacokinetic profile, lower toxicity, and permeability. It is well known that the mycobacteria have special anatomical barriers that prevent simpler treatment. Such properties are ultimately related to the physicochemical properties of any compounds. Another potential application of this method is reversible screening by affording a direct match of a compound of interest to match similarity of its physicochemical properties with other libraries or datasets available, for example, natural products. The selected similar compounds from the reversible screening would have the advantages of sharing similar physicochemical properties. These compounds can be of any chemical classes and therefore bypass any limitation presented by the structural screening methods.

Researchers can benefit from this study by adopting their databases to ChemGPS-NP model to monitor their screening schemes at earlier stages by visualizing the distribution pattern and resemblance to the active set of compounds provided as Supplementary Materials. Furthermore, the list of highly active compounds can act as a reference set of compounds that can be matched with any database for screening enrichment and potential identification of antituberculosis compounds.

3. Material and Methods

3.1. Data Collection and Sources

3.1.1. Screening Schemes

Publically available screening results from three different sources were used (collectively called screening sets). These are the GSK set (776 compounds) [18], the NIAID Set (3779 compounds) [19], and the GVK Bio Set (2880 compounds) [19]. All of these compounds showed antagonistic activities against tuberculosis, and their activities were averaged and categorized for the purpose of comparison.

3.1.2. In-House Active Set

An in-house active set was constructed from the different antimycobacterial screening schemes along with the 46 clinically proven drugs publicly available [18,19]. The selection was based on higher activities against tuberculosis. The active set consists of 38, 132, and 289 compounds, corresponding to GSK, GV KBio, and NIAID respectively. The top 20 compounds were piperazine and naphthalene derivatives. The list of compounds is provided as Supplementary Materials. The compounds were sorted according to the activity and provided with three ChemGPS-NP scores along with permeability and toxicity filters in the Supplementary Materials. This active set was used in the visualization model of ChemGPS-NP.

3.1.3. Database

MayBridge screening collection, with over 53,000 organic compounds of drug-like properties, was used in ChemGPS-NP visualization model. This collection was also used as an example of how to extract potential targets when compared to antimycobacterial screening schemes.

3.2. Compound Classification:

3.2.1. Classification According to Activity

The grouping was performed for most of the sets, dividing them into seven classes corresponding to AAA, AA, A, BBB, BB, B, and D, arranged logarithmically in order of the highest activity to the lowest activity. Classes BBB to D were regarded as ineffective and were eliminated from analysis. All activity classes were used in defining antituberculosis property space.

3.2.2. Classification According to Predicted Permeability

Probability predictions for permeability for all compounds were provided by MycPermCheck [20]. This logistic regression modeling of the physicochemical properties sorts compounds into a list of small compounds ranked by their permeability into *Mycobacterium tuberculosis*, ranked from highest to lowest probability.

3.2.3. Classification According to Toxicity Profile

The toxicities of compounds from different screening strategies were provided by the web application SmartsFilter (University of New Mexico; <http://pasilla.health.unm.edu/tomcat/biocomp/smartsfilter>). Each compound was marked pass or fail in the same regards.

3.3. Chemical Global Positioning System- Natural Product (ChemGPS-NP)

ChemGPS-NP [21] is a principal component analysis (PCA) [22] based global chemical positioning system tuned for exploration of biologically relevant chemical space. This is pursued through an eight-dimensional (8D) map with dimensions based on structure-derived physicochemical characteristics, for the ChemGPS-NP reference set of 1776 compounds. By calculating PCA prediction scores using a publicly available service (<http://chemgps.bmc.uu.se>), large sets of compounds can be mapped onto the 8D-map. Using this method, compounds from different screening sets were mapped together with reference sets with known antituberculosis mechanisms and activity degrees. Principal component and PCA score predictions were calculated, employing the software SIMCA-P+, with the training set ChemGPS-NP. Prior to PCA determination, all data were centered and scaled to unit variance.

Descriptors

Molecular and physicochemical properties were incorporated as chemically intuitive descriptors, described in detail by Larsson and co-authors [21]. The selected ChemGPS-NP descriptors were calculated for all sets of compounds using the software dragonX by Talete srl. (<https://chm.kode-solutions.ne>). Compounds were submitted on the basis of their structure information as simplified molecular input line entry specifications (SMILES; <http://www.daylight.com/smiles/>) and initially verified as described by Rosén and co-workers [23] to omit salts, hydration, stereochemical, and ionic information.

3.4. Data Analysis and Presentation:

3.4.1. Calculation of Euclidean Distances

Euclidean distance estimates based on ChemGPS-NP principles were calculated according to an Excel formula that was used to relate a compound and core set compounds. The Euclidean distance was calculated between points $P = (p_1, p_2, \dots, p_n)$ and $Q = (q_1, q_2, \dots, q_n)$ in Euclidean n -space, as defined by

$$d(p, q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2} \quad (1)$$

For the purpose of placing progressively greater weight on objects that were farther apart, the square of the Euclidean distances were also calculated as

$$d^2(p, q) = (p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_i - q_i)^2 + \dots + (p_n - q_n)^2 \quad (2)$$

3.4.2. Calculation of the Tanimoto Coefficient

$$T_{AB} = \frac{n_{AB}}{n_A + n_B - n_{AB}} \quad (3)$$

where n_{AB} is the number of common bits set to 1 in both ligands A and B , and n_A and n_B are the number of bits set to 1 in ligands A and B individually.

Supplementary Materials: The following are available online at <http://www.mdpi.com/1420-3049/25/4/945/s1>, Table S1: List of highly active compounds (Active set) along with their 3D positions on the chemical property space in addition to the permeability and toxicity filters.

Author Contributions: M.M.A. has performed and written this article with supervision and contribution of A.B. All authors have read and agreed to the published version of the manuscript.

Acknowledgments: Authors wish to pass humble gratitude to Jeff Elhi, Center for the Study of Biological Complexity, Virginia Commonwealth University, and the USA for English editing.

Conflicts of Interest: Authors declare that they have no competing interests.

Abbreviation

ChemGPS-NP, Chemical global positioning system–natural product; TB, tuberculosis; GSK, GlaxoSmithKline; GVKBio, GVK Biosciences; NIAID, National Institute for Allergy Infectious Diseases; PCA, principle component analysis; MDR, multidrug resistance; QSAR, quantitative structure–activity relations; HTS, high-throughput screening.

References

1. Cambau, E.; Drancourt, M. Steps towards the discovery of *Mycobacterium tuberculosis* by Robert Koch, 1882. *Clin. Microbio. Infec.* **2014**, *20*, 196–201. [CrossRef] [PubMed]
2. Dheda, K.; Migliori, G.B. The global rise of extensively drug resistant tuberculosis: is the time to bring back sanatoria now overdue? *Lancet* **2012**, *379*, 773–775. [CrossRef]
3. World Health Organization. *Global Tuberculosis Report 2012*; World Health Organization: Geneva, Switzerland, 2013. Available online: http://www.who.int/tb/publications/global_report/en/ (accessed on 1 February 2020).
4. Knox, C.; Law, V.; Jewison, T.; Liu, P.; Ly, S.; Frolkis, A.; Pon, A.; Banco, K.; Mak, C.; Neveu, V.; et al. DrugBank 3.0: A comprehensive resource for ‘Omics’ research on drugs. *Nucleic Acids Res.* **2014**, *39*, 1035–1041. [CrossRef] [PubMed]
5. World Health Organization. *Treatment of Tuberculosis Guidelines*, 4th ed.; WHO: Geneva, Switzerland, 2010. Available online: http://whqlibdoc.who.int/publications/2010/9789241547833_eng.pdf (accessed on 1 February 2020).
6. Global Alliance for TB Drug Development. Ongoing Projects without a Lead Compound Series. Available online: <https://www.newtbdrugs.org/pipeline/clinical> (accessed on 30 June 2018).
7. Osborne, R. First novel anti-tuberculosis drug in 40 years. *Nat. Biotech.* **2012**, *31*, 89–91. [CrossRef] [PubMed]
8. Kaneko, T.; Cooper, C.; Mdluli, K. Challenges and opportunities in developing novel drugs for TB. *Fut. Med. Chem.* **2011**, *3*, 1373–1400. [CrossRef] [PubMed]
9. Dobson, C.M. Chemical space and biology. *Nature* **2004**, *432*, 824–828. [CrossRef] [PubMed]
10. Pollock, S.N.; Coutsias, E.A.; Wester, M.J.; Oprea, T.I. Scaffold Topologies II: Analysis of Chemical Databases. *J. Chem. Inf. Model.* **2008**, *48*, 1304–1310. [CrossRef] [PubMed]
11. Reymond, J.L.; Awale, M. Exploring chemical space for drug discovery using the chemical universe database. *ACS Chem. Neurosci.* **2012**, *3*, 649–657. [CrossRef] [PubMed]
12. Brennan, P.J.; Young, B.D. Handbook of anti-tuberculosis agents. *Tuberculosis* **2008**, *88*, 85–170.
13. Sacchettini, J.C.; Rubin, E.J.; Freundlich, J.S. Drugs versus bugs: In pursuit of the persistent predator *Mycobacterium tuberculosis*. *Nat. Rev. Microbiol.* **2008**, *6*, 41–52. [CrossRef] [PubMed]
14. Maddry, J.A.; Ananthan, S.; Goldman, R.C.; Hobrath, J.V.; Kwong, C.D.; Maddox, C.; Rasmussen, L.; Reynolds, R.C.; Secrist, J.A.; Sosa, M.I.; et al. Antituberculosis activity of the molecular libraries screening center network library. *Tuberculosis* **2009**, *89*, 354–363. [CrossRef] [PubMed]

15. Macarron, R.; Banks, M.N.; Bojanic, D.; Burns, D.J.; Cirovic, D.A.; Garyantes, T.; Green, D.V.S.; Hertzberg, R.P.; Janzen, W.P.; Paslay, J.W.; et al. Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discov.* **2011**, *10*, 188–195. [[CrossRef](#)] [[PubMed](#)]
16. Ekins, S.; Bradford, J.; Dole, K.; Spektor, A.; Gregory, K.; Blondeau, D.; Hohman, M.; Bunin, B.A. A Collaborative database and computational models for tuberculosis drug discovery. *Mol. Biosyst.* **2010**, *6*, 840–851. [[CrossRef](#)] [[PubMed](#)]
17. Balcells, L.; Bates, R.H.; Young, R.J.; Alvarez-Gomez, D.; Alvarez-Ruiz, E.; Barroso, V.; Blanco, D.; Crespo, B.; Escribano, J.; González, R.; et al. Fueling Open-Source Drug Discovery: 177 Small-Molecule Leads against Tuberculosis. *Chem. Med. Chem.* **2013**, *8*, 313–321. [[CrossRef](#)] [[PubMed](#)]
18. Prathipati, P.; Ma, N.L.; Keller, T.H. Global Bayesian models for the prioritization of antitubercular agents. *J. Chem. Inf. Model.* **2008**, *48*, 2362–2370. [[CrossRef](#)] [[PubMed](#)]
19. Merget, B.; Zilian, D.; Muller, T.; Sotriffer, C.A. MycPermCheck: The *Mycobacterium tuberculosis* permeability prediction tool for small molecules. *Bioinformatics* **2013**, *29*, 62–68. [[CrossRef](#)] [[PubMed](#)]
20. Larsson, J.; Gottfries, J.; Muresan, S.; Backlund, A. ChemGPS-NP: Tuned for navigation in biologically relevant chemical space. *J. Nat. Prod.* **2007**, *70*, 789–794. [[CrossRef](#)] [[PubMed](#)]
21. Eriksson, L.; Johansson, E.; Kettaneh-Wold, N.; Trygg, J.; Wilkström, C.; Wold, S. *Multiand Megavariate Data Analysis Part I Basic Principles and Applications*, 2nd ed.; Utmetrics Inc.: Umeå, Sweden, 2006; pp. 1–425.
22. Rosén, J.; Lövgren, A.; Kogej, T.; Muresan, S.; Gottfries, J.; Backlund, A. ChemGPS-NPweb—Chemical space navigation online. *J. Comp. Aid. Mol. Design* **2009**, *23*, 253–259. [[CrossRef](#)] [[PubMed](#)]
23. Stanley, S.A.; Grant, S.S.; Kawate, T.; Iwase, N.; Shimizu, M.; Wivagg, C.; Silvis, M.; Kazyanskaya, E.; Aquadro, J.; Golas, A.; et al. Identification of novel inhibitors of *M. Tuberculosis* growth using whole cell based high-throughput screening. *ACS Chem Biol.* **2012**, *7*, 1377–1384. [[CrossRef](#)] [[PubMed](#)]

Sample Availability: Samples of the compounds are not available from the authors.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).