JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

In silico Comparison of Antimycobacterial Natural Products with Known Antituberculosis Drugs

Marlene Espinoza-Moraga,[†] Nicholas M. Njuguna,[†] Grace Mugumbate,[†] Julio Caballero,[‡] and Kelly Chibale*[,†,§]

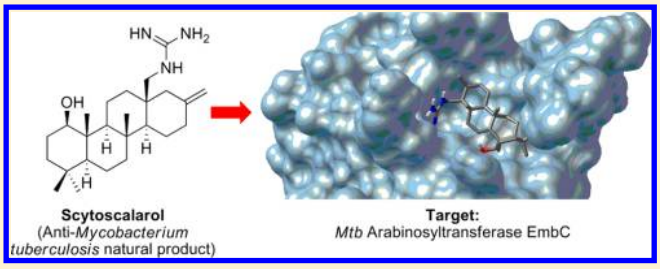
[†]Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

[‡]Centro de Bioinformática y Simulación Molecular, Facultad de Ingeniería, Universidad de Talca, 2 Norte 685, Casilla 721, Talca, Chile

[§]Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Rondebosch 7701, South Africa

Supporting Information

ABSTRACT: The chemical space based on physicochemical properties and structural features of a diverse group of natural products with reported in vitro activity against different *Mycobacterium tuberculosis* strains is investigated using in silico tools. This is compared to the chemical space occupied by drugs currently recommended for the treatment of various forms of tuberculosis as well as compounds in preclinical and clinical development. Docking studies exploring possible binding affinities and modes of two main clusters of natural products on two different mycobacterial targets are also reported. Our docking results suggest that scytoscalarol, an antibacterial and antifungal guanidine-bearing sesterterpene, can inhibit arabinosyltransferase *Mtb* EmbC, and the *β*-carboline alkaloids 8-hydroxymanzamine A and manzamine A can bind to the oxidoreductase of *Mtb* INHA. On this basis, these products showing high binding affinities to the two targets in silico could be rationally selected for in vitro testing against these targets and/or semisynthetic modification.



INTRODUCTION

It is estimated that roughly a third of the world's population is infected with *Mycobacterium tuberculosis* (*Mtb*), the pathogenic microorganism that causes tuberculosis (TB).[1] Globally, tuberculosis is second only to HIV/AIDS as the leading cause of death due to an infectious disease. In 2010, approximately 8.5−9.2 million new cases and 1.4 million deaths were reported worldwide as a result of TB.[2] Active TB is treated using combinations of different chemotherapeutic agents typically administered over a duration of 6−12 months.[3] First-line uncomplicated TB chemotherapy involves the use of four main drugs: rifampicin (**1**), isoniazid (**2**), ethambutol (**3**), and pyrazinamide (**4**) (Figure 1) administered over a 2 month intensive phase. This is followed by a further 4 month continuation phase treatment using only rifampicin and isoniazid.[4] The long duration of treatment is partly due to the difficulty in attaining therapeutic drug concentrations at the site of action because of low permeability across the *Mtb* cell wall.

In recent years, drug resistant strains of *Mtb* have been reported to cause multidrug resistant (MDR) and extensively drug resistant (XDR) forms of TB. For the treatment of these forms of the disease, a broad selection of approximately 25 different conventional drugs (CDs) has been used in different combinations. These include fluoroquinolones such as ofloxacin (**5**) and moxifloxacin (**6**), aminoglycosides such as amikacin (**7**) and kanamycin (**8**), and *β*-lactam antibiotics such as imipenem (**9**) and coamoxicillin-clavulanate (**10**, **11**) as well as peptide antibiotics such as capreomycin (**12**) (Figure 2).[5]

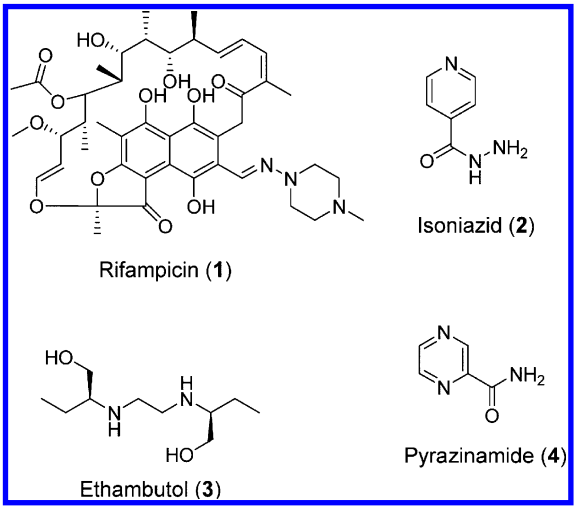


Figure 1. Chemical structures of first-line antituberculosis drugs.

In spite of the worldwide problem caused by TB, it has been more than 30 years since a drug having a novel mechanism of

Received: October 1, 2012

   dx.doi.org/10.1021/ci300467b | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

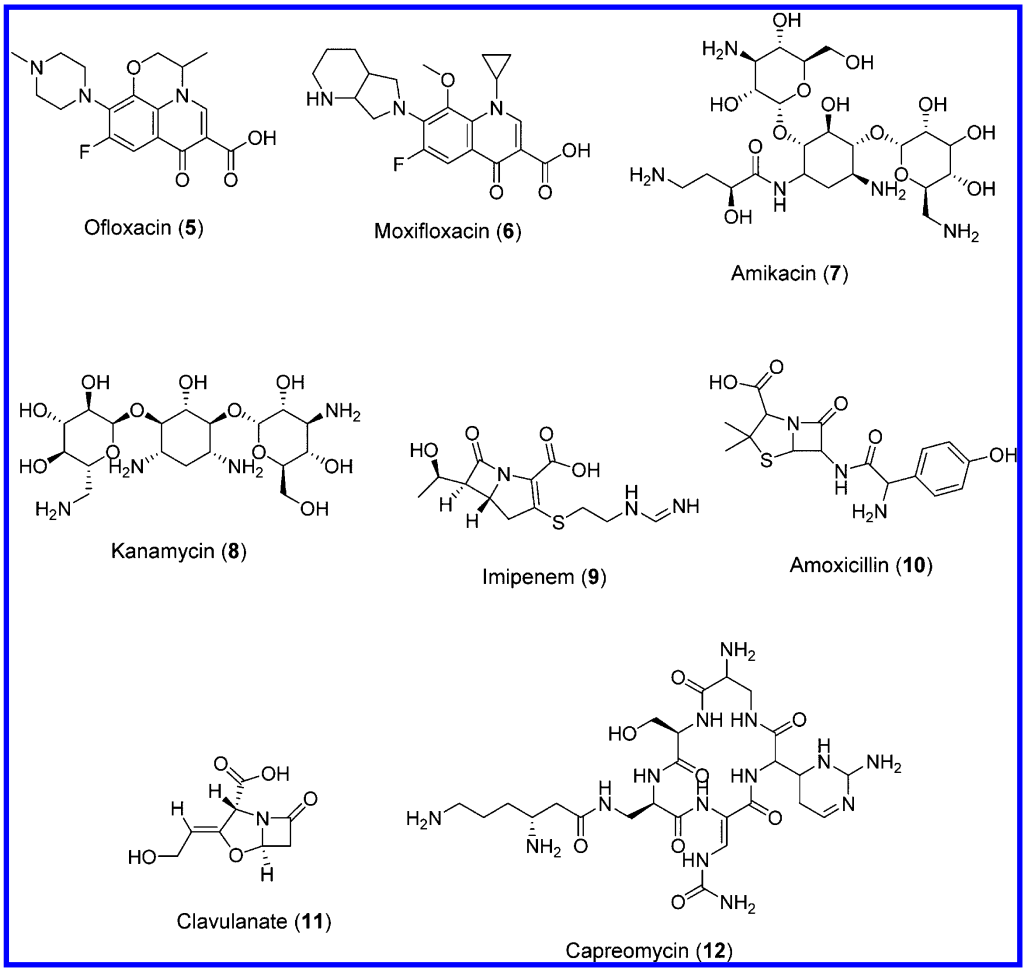


Figure 2. Drugs used to treat multidrug-resistant (MDR) and extensively drug resistant (XDR) tuberculosis.

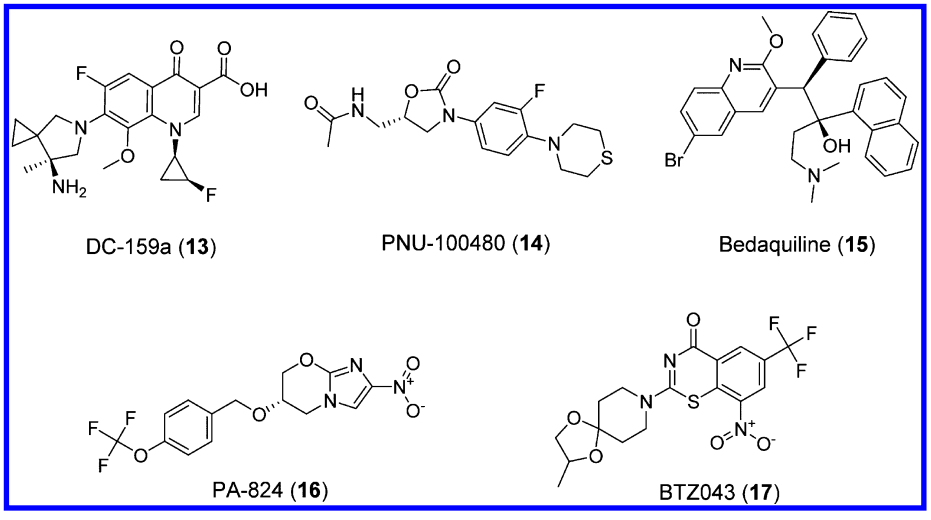


Figure 3. Examples of antituberculosis compounds in preclinical and clinical development.

action was introduced to treat the disease. The search for drugs having fewer side effects, administered over a shorter duration than current treatments and with novel mechanisms of action that circumvent the ability of the *Mycobacterium* to develop resistance and thus eliminate MDR and XDR tuberculosis has been ongoing for many years.[1] By March 2012, the number of preclinical and clinical development compounds (DCs) as anti-TB drugs stood at 6 and 11, respectively. These included the following: fluoroquinolones (DC-159a (**13**), gatifloxacin and moxifloxacin), rifapentine, oxazolidinones (AZD5847, linezolid and PNU-100480 (**14**)), the diarylquinoline TMC207 or bedaquiline (**15**), nitroimidazoles (PA-824 (**16**) and OPC-67683), and the enzothiazinone BTZ043 (**17**) (Figure 3).[6]

Traditionally, natural products (NPs) have played a key role in the discovery and development of new drugs.[7] In recent studies, NPs and some of their derivatives have shown promising

dx.doi.org/10.1021/ci300467b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

inhibitory activity against *Mtb*, highlighting their potential as a powerful source of new anti-TB drugs.

In order to take advantage of the biological relevance of NPs, and their diverse structural features, a systematic approach to chart, navigate, and analyze their chemical space for drug discovery is needed. Isolation and structural characterization of NPs is normally time-consuming, and their total synthesis often difficult and expensive. In addition, with a large data set of NPs, it is often challenging to obtain actual samples for experimental evaluation because of their diverse sources including plants, microorganisms, and marine organisms as well as terrestrial vertebrates and invertebrates.[8] Currently, in silico methods are routinely incorporated into drug discovery projects by allowing the prediction of physicochemical properties based on the chemical structures of large compound sets and their subsequent analysis for "drug-likeness" prior to more intensive experimental evaluation of the same. Recent analyses have shown that clinically approved natural products can be divided into two equal subsets namely, the "Lipinski universe" which complies with the Rule of Five and the "parallel universe" that violates the Lipinski Rule of Five.[9]

In this paper, the chemical space based on molecular properties (e.g., physicochemical properties and structural features) of a diverse group of NPs with reported in vitro activity against different *Mtb* strains is investigated. This is compared to the chemical space of CDs currently recommended for the treatment of various forms of TB as well as preclinical and clinical DCs for the same by employing in silico techniques. Docking studies to explore possible binding modes and affinities of two NP compound clusters on reported targets of CDs and/or DCs are also reported.

METHODOLOGY AND COMPUTATIONAL DETAILS

Data Set Characteristics and Physicochemical Properties. NPs used in this work were extracted from literature sources and a total of 397 compounds with reported antimycobacterial MIC < 50 μg/mL or 60 μM were selected.[10] For each compound, a molecular structure file was generated using ChemBioDraw Ultra 11.0.1.[11] In addition, structure files of 25 conventional anti-TB drugs and 13 compounds in preclinical and clinical development were similarly prepared. For all compounds, SMILES generated using ChemBioDraw Ultra 11.0.1 were used as the primary data entry format for property predictions using the other software programs.

In silico prediction of physicochemical and structural properties of the NPs, conventional drugs and development compounds was performed using three different software programs. Volsurf+ 1.0.4[12] was used to calculate a total of 126 molecular descriptors including cLogP and polar surface area from GRID generated molecular interaction fields (MIFs)[13] and to predict solubility and Caco-2 permeability of all compounds based on in-built PLS models. The incorporated builder tool in VIDA 4.0.3[14] was used to calculate the number of rotatable bonds, hydrogen bond (H-bond) donors and acceptors in each molecule. The same properties were also calculated using SciTegic Pipeline Pilot 8.5.[15] A separate protocol was designed in Pipeline Pilot to identify compounds that complied or deviated from the druggability compliant and noncompliant sets.

Similarity Analysis and Docking Studies. Cluster analysis based on common functional groups of all the 435 compounds was determined by computation of Tanimoto scores using the Pipeline Pilot FCFP4 property set. Docking calculations of two clusters with more than 20 compounds to two validated anti-TB targets were performed to determine the roles of their functional groups in binding and to determine their binding energies and orientations. To identify the suitable docking tool for the NPs, preliminary docking calculations were carried out using both Glide 5.7,[16] and the commonly used Autodock 4.2.[17] Glide 5.7 could not generate acceptable conformations for most of the NPs as some of the compounds had long aliphatic chains, complex rings, and high numbers of rotatable bonds. On the other hand, Autodock 4.2 managed to produce meaningful binding conformations for all NPs and regenerated the conformations of the ligands in the crystal structures within acceptable root-mean-square deviation of less than 3.0 Å. Thus Autodock 4.2 was the docking tool of choice in this work. This software uses the Lamarkian Genetic Algorithm (LGA) search function, a hybrid genetic algorithm with local optimization that uses a parametrized free-energy scoring function to estimate the binding energy.[18]

First, compounds from cluster 4 (denoted as EMB cluster in this manuscript) consisting of 19 NPs, ethambutol (CD), and SQ109 (DC) were docked in complex with AFO to the C-terminal extracellular domain of arabinosyltransferase *Mtb* EmbC (PDB ID 3PTY) binding site. This protein domain is a target for ethambutol (CD), SQ109 (DC), and SQ609 (DC). To provide enough space for free movements of the ligands, the grid box was constructed to cover the whole AFO active site using Autodock tools and it was defined using AutoGrid4. The grid points were set to 42 × 42 × 46, at a grid center of (*xyz*) 93.9, 10.7, 3.5 with spacing of 0.4 Å. Similarly, 24 compounds from the cluster 12 (denoted as ETH cluster in this manuscript), including prothionamide, isoniazid, pyrazinamide, and ethionamide (all CDs), were docked to the oxidoreductase of *Mtb* INHA bound to ETH-NAD adduct (PDB ID 2H9I). INHA is a known target for activated thionamides. Although pyrazinamide has a similar functional group (e.g., amide) to the thionamides, it is inactive toward INHA. For this system, the grid box of size 56 × 56 × 56 was centered at *xyz* coordinates −11.2, 38.8, 12.4 with a spacing of 0.4 Å. This grid box covered both the hydrophobic binding site for the activated thionamide moiety and the nucleotide-binding site. In both calculations a maximum possible torsional degrees of freedom for each molecule was utilized. For a population of 150 individuals, 5×10^6 maximum energy evaluations and operations were applied. All other parameters were set to Autodock 4.2 default values including the rate of gene mutation (0.02) and rate of crossover (0.8). A total of 50 generic algorithm (GA) trials were executed for each ligand. The generated 50 docked conformations were clustered at a tolerance of 2.0 Å and ranked based on their binding energies. Autodock tools and Chimera 1.5.2[19] were used to view and analyze the best docked conformations.

RESULTS AND DISCUSSION

NPs have recently been reported as having significant in vitro antimycobacterial activity, with minimal inhibitory concentrations (MICs) of less than 50 μg/mL or 60 μM. Therefore, a total of 397 natural chemical constituents were extracted from three different literature sources.[10] Computational tools were used to investigate the chemical space spanned by these NPs in comparison with that for 25 anti-TB CDs and 13 DCs currently in preclinical and clinical evaluations. Here chemical space includes "drug-like" properties, solubility, permeability as well as structural and functional similarity. In addition, the binding affinities of selected compounds to validated anti-TB targets were determined.

dx.doi.org/10.1021/ci300467b | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

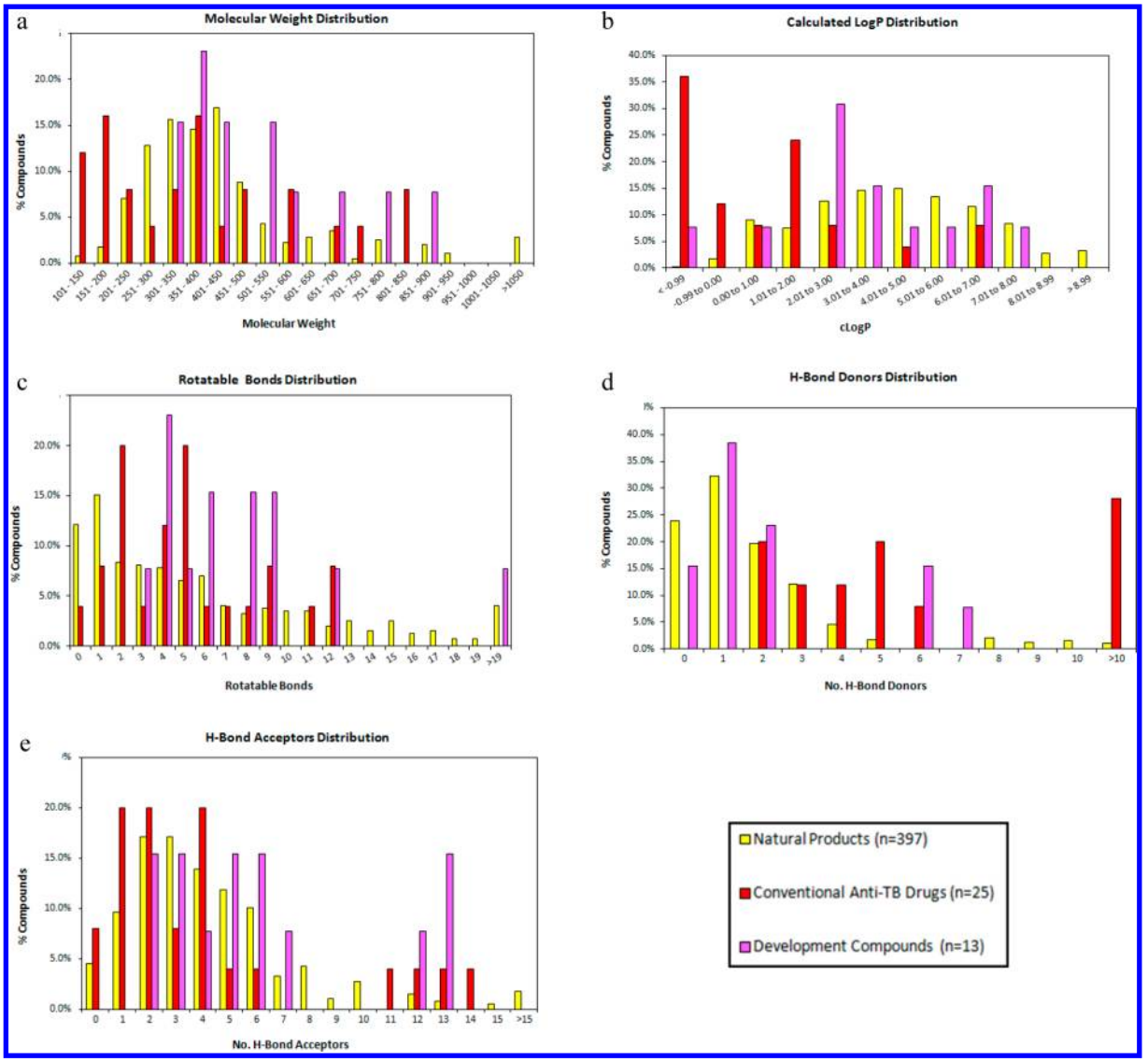


Figure 4. Frequency distribution histograms of physicochemical properties of natural product hits against conventional drugs and development compounds as antitubercular agents: (a) molecular weight; (b) cLogP; (c) rotatable bonds; (d) H-bond donors; (e) H-bond acceptors.

Drug-like Properties. In silico profiling of all NPs, CDs, and DCs to predict their drug-like properties, as summarized in Lipinski's "Rule of Five" for orally administered drugs,[20a] which are molecular weight (M_{wt}) ≤ 500, clogP ≤ 5, H-bond donors (HBD) ≤ 5, and acceptors (HBA) ≤ 10, was performed. We also considered the rotatable bonds that have to be less than 10 as proposed by Veber et al.[20b] The histograms in Figure 4 show the percent frequency distribution of the predicted or calculated values for these parameters.

From Figure 4a, more than 70.0% of the NPs and CDs had M_{wt} less than 500 with most having values ranging from 200 to 450. In both sets, only slightly more than 6% had weights greater than 750 while no compound had a value below 100. In contrast, 53.8% of the DCs had M_{wt} less than 500 mainly with M_{wt} ranging from 300 to 450. No compound had M_{wt} less than 300 while 15.4% had M_{wt} in excess of 750.

The analysis of Figure 4 indicates that distributions of M_{wt} are not normal. Medians of M_{wt} values for NPs, CDs, and DCs were calculated and are reported in Table 1. To compare the medians of M_{wt} for NPs, CDs, and DCs, we used the Wilcoxon rank-sum test.[21] This is a nonparametric statistical hypothesis test for assessing whether two not normally distributed samples of independent observations tend to have larger values than each other. The results of the application of this test considering M_{wt} are summarized in Table 2. According to the results of the test, it is not possible to guarantee that M_{wt} values in the analyzed sets are significantly different.

Table 1. Median Values of Predicted Physicochemical Parameters

	median calculated values		
	natural products (n = 397)	conventional anti-TB drugs (n = 25)	development compounds (n = 13)
molecular weight	388.5	358.5	431.4
cLogP	4.3	0.4	3.5
rotatable bonds	4	5	6
H-bond acceptors	4	3	5
H-bond donors	1	5	1

Estimated cLogP (Figure 4b) values for the NPs revealed a distinctly normal distribution ranging from −4.9 to 9.0 with a median value of 4.3. About 91.9% of NPs were within the 0−8.0 cLogP range and 61.0% of the compounds in this set had

dx.doi.org/10.1021/ci300467b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 2. Wilcoxon Rank-Sum Test Results for Differences between the Predicted Physicochemical Parameters for NPs, CDs, and DCs[a]

	M_{wt}		cLogP		rot. bonds		HBA		HBD		RO5 dev	
	p-value	*h*	*p*-value	*h*	*p*-value	*h*	*p*-value	*h*	*p*-value	*h*	*p*-value	*h*
NP vs CD	0.093	0	0.000	1	0.976	0	0.186	0	0.000	1	0.865	0
NP vs DC	0.076	0	0.344	0	0.089	0	0.055	0	0.595	0	0.106	0
CD vs DC	0.056	0	0.002	1	0.080	0	0.032	1	0.002	1	0.181	0

[a]The null hypothesis means that data in the compared sets are samples from continuous distributions with equal medians. The significance level was 0.05. If *p*-value is less than 0.05, then $h = 1$, and this indicates rejection of the null hypothesis at the significance level. If *p*-value is higher than 0.05, then $h = 0$, and this indicates a failure to reject the null hypothesis at the significance level.

cLogP values lower than 5.0. On the other hand, CDs presented a not normal distribution with cLogP values ranging from −9.0 to +6.8 with the range between −9.0 and +3.0 accounting for 88.0% of the total. In total, 92.0% of CDs had cLogP values within the acceptable Rule of Five range. Finally, DCs presented a distribution of cLogP values ranging from −2.7 to +7.7 with a median value of 3.5, and 69.2% of the compounds having values lower than 5.0. The Wilcoxon rank-sum test considering cLogP values indicates that there is no difference between median cLogP values of NPs and DCs (Table 2). However, the median cLogP value of 0.4 for CDs (Table 1) was notably lower than that for the NPs and DCs indicating a higher proportion of hydrophilic compounds in this collection. The Wilcoxon rank-sum test indicates that this difference is significant (Table 2).

According to Veber et al.[20b] most orally active drugs tend to possess at most 10 rotatable bonds. The NP, CD, and DC molecules exhibited similar distributions in this property with 79.6%, 88.0%, and 84.6%, respectively having 10 or less rotatable bonds (Figure 4c). NPs and CDs exhibited similar profiles with 51.4% and 48% displaying four or fewer rotatable bonds whereas 12.1% and 4.0%, correspondingly lacked any at all. The analysis of median values and Wilcoxon rank-sum test considering rotatable bonds (Tables 1 and 2) indicates that there is no difference in this property between the studied sets.

The ability to form H-bonds is a key requirement for drug-like compounds to be orally active, since such bonds enhance aqueous solubility and also partly because they facilitate substrate−target interactions at the subcellular active sites of most drugs. The Rule of Five proposes that the ideal number of HBD should be five or less, while HBA must not exceed 10. In general, the majority of NPs (94.2%), CDs (64.0%), and DCs (76.9%) complied with the rule regarding HBD (Figure 4d). On the other hand, 95.5% of the NPs had 10 or fewer HBA, while the CDs and DCs had 84.0% and 76.9%, correspondingly (Figure 4e). All three compound groups presented a distribution skewed to the right indicating higher incidence of fewer HBA in general, with this characteristic being markedly pronounced with the NPs and CDs molecules in which 62.2% and 76.0% had four or less acceptors, respectively. The analysis of median values and Wilcoxon rank-sum test considering HBA and HBD parameters reveal several differences between the studied sets. The most important difference is that the median HBD value for CDs was five, and the ones for NPs and DCs were one for both sets (Table 1). The Wilcoxon rank-sum test indicated that this difference is significant (Table 2). This means that CDs include molecules with more groups able to donate protons, but NPs and DCs include molecules lacking HBD which indicated molecules devoid of hydroxyl, amine, or carboxylic acid functional groups in these data sets. On the other hand DCs include molecules with higher HBA values with respect to these values for CDs according to the medians and the significance test (Tables 1 and 2).

It has been reported that about 50% small molecule natural products recently developed into drugs violate druggability rules.[9b] In the same way, all the analyzed compounds fall within the druggability compliant and noncompliant spaces. Here, 207 out of the 397 NPs (52.1%) failed to comply with at least one of the considered parameters (Table 3).

Table 3. Compounds' Lipinski Parameters Compliance

	natural products (n = 397)	conventional anti-TB drugs (n = 25)	development compounds (n = 13)
compliant	190 (47.9%)	15 (60.0%)	4 (30.8%)
noncompliant	207 (52.1%)	10 (40.0%)	9 (69.2%)

The most frequently violated rule for NPs was cLogP where 155 compounds (39.0%) exceeded the +5.0 limit followed by M_{wt} and rotatable bonds where 86 (21.7%) and 81 (20.4%) compounds fell outside the prescribed limits. Likewise, 40.0% of CDs (10 of the 25 compounds) violated one or more of the considered parameters and the most frequent deviation occurred in the limit for HBD where 36.0% of the total compounds violated the rule. M_{wt} and HBA were violated by a similar proportion of the compounds (24.0% and 16.0%, respectively) while 12.0% and 8.0% flouted the rule on number of rotatable bonds and cLogP, correspondingly. In contrast, 69.2% of DCs (9 of 13) violated at least one of the rules. M_{wt} and HBA were the most prone to deviations (both 46.2%). 30.8%, 23.1% and 15.4% of the deviating compounds exceeded the limits for cLogP, HBD, and rotatable bonds in that order. It is important to note that violation of Lipinski's Rule of Five or the number of rotatable bonds does not necessarily preclude development of a compound into novel drugs, as indicated by the high proportion of CDs and DCs that violate these rules. Therefore, more characteristics of the NPs need to be explored.

Solubility and Permeability. Physicochemical properties of a drug, such as solubility and permeability, also govern its oral bioavailability since they are important for absorption, distribution, metabolism, and excretion. The ability of an anti-TB drug to reach its target site is greatly hampered by the highly impermeable *Mtb* cell wall. Qualitative predictions of permeability and solubility of the NPs, CDs, and DCs were performed using Volsurf+ 1.0.4. The compounds were projected onto the in-built two-dimensional partial least-squares (PLS) models used to predict Caco-2 permeability and solubility. PLS t/t score plots for the permeability and solubility models in Figures 5 and 6 show the projected compounds and the compounds in the model's training set (gray dots).

dx.doi.org/10.1021/ci300467b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

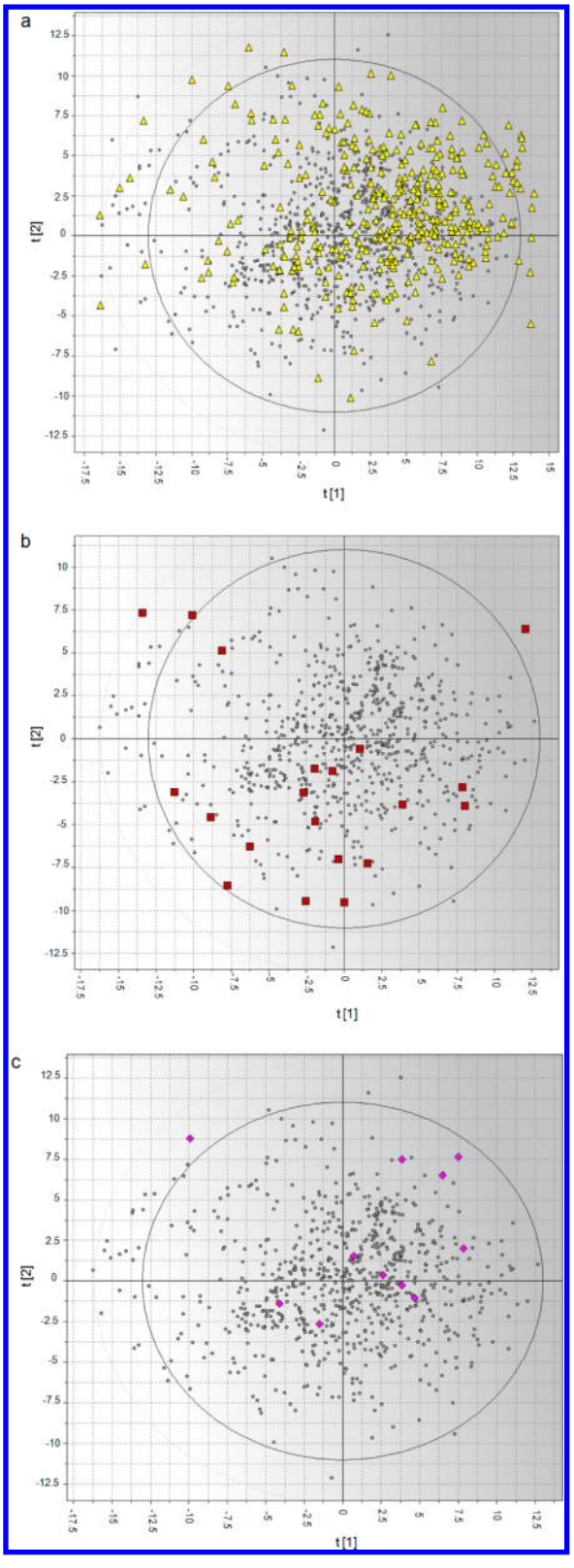


Figure 5. Partial least square t/t score plots of Caco-2 permeability model for (a) natural products; (b) conventional drugs; (c) development compounds.

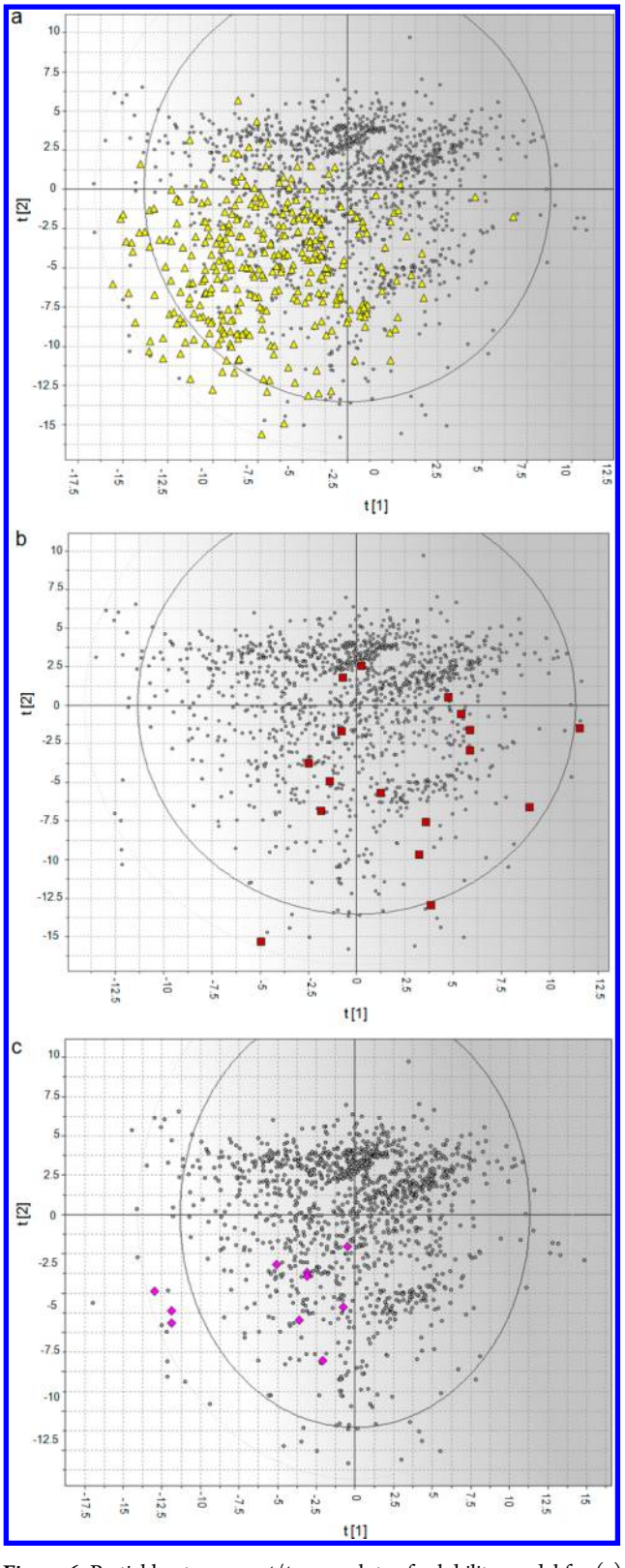


Figure 6. Partial least-squares t/t score plots of solubility model for (a) natural products; (b) conventional drugs; (c) development compounds.

In the plots in Figures 5 and 6, the gray regions on the right represent high permeability and solubility, respectively. More than 75.0% of all compounds fell within the acceptable permeation regions of the plots (i.e., within 95.0% confidence interval region represented by the solid line). Most of the NPs and the DCs occupied the high permeation space. Conversely, the CDs displayed moderate to low permeation. Moreover, about 70.0% of CDs were within the moderate acceptable solubility region. Outliers (compounds outside the 99.0%

dx.doi.org/10.1021/ci300467b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

confidence region (dotted line)) that included rifampicin, kanamycin, and amikacin are not shown in the plots since no reliable information about them could be predicted by the program. It was observed that most of these outliers were injectable drugs with typically very high solubility. On the contrary, more than 70.0% of NPs and DCs occupied low but acceptable solubility space.

The chemical space of several compounds with reported activity against *Mycobacterium tuberculosis* has been previously investigated using physicochemical properties.[22] There are reports where compounds from databases or other sources are explored to find potential active ligands using virtual screening methods and Lipinski's Rule of Five as a filter.[23,24] On the other hand, there are previous reports that verified that many antitubercular compounds do not obey Lipinski's Rule of Five.[25] In the current work we found similar results since many compounds in CD and DC sets do not comply with Lipinski's rule. Therefore, it is expected that compounds in the NP set can be considered as promising candidates in the treatment of tuberculosis although they do not comply with Lipinski's rules.

To have a broader view of the comparison between NPs, CDs, and DCs, we reported a summary of all the properties we analyzed in Table 4. This view can help in the analysis of the NP characteristics and their relation with the characteristics of CDs and DCs. It is important to remember when performing this analysis that there are 397 compounds in the NP set, but there are only 25 and 13 compounds respectively in CD and DC sets. From the analysis of Table 4, we observed that the NP set is similar to the CD set when M_{wt} and HBA properties are considered, and the NP set is similar to the DC set when clogP, HBD, permeability, and solubility properties are considered, while the analysis of the rotatable bonds showed no significant differences.

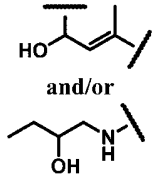
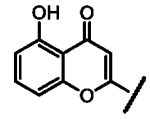
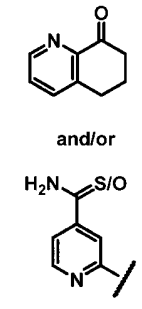
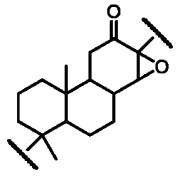
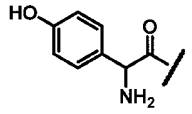
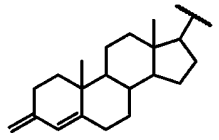
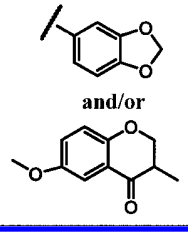
Similarity analysis. The similar property principle (SPP) basically states that similar molecules should have similar biological activities.[26] On the basis of the SPP, antitubercular NPs were subjected to 2D molecular similarity analysis, against CDs and DCs using Pipeline Pilot 8.5[15] and the functional class fingerprint (FCFP4) to determine similarity scores based on Tanimoto coefficient. Even though the NPs had showed high antimycobacterial activity, low 2D molecular similarity was displayed with both CDs and DCs. The highest similarity score of 0.54 was obtained between EJMC237 and amikacin. This suggests that NPs occupy a different 2D structural space which might call for structural modifications of the molecules or development of new anti-TB drugs with new modes of action. Upon cluster analysis of all NPs, CDs and DCs based on functionality of the atoms in their molecules using a Pipeline Pilot protocol and the FCFP4, a total of 435 compounds generated 45 clusters, 14 of which contained CD and/or DC compounds. The molecular characteristics of these clusters are reported in section D of the Supporting Information, and the more important clusters according to the number of compounds included are in Table 5. Three of the clusters (Table 5) had more than 15 NPs, and the compounds were distributed as follows: cluster 4 (EMB cluster) had 2 CD/DCs (e.g., ethambutol) and 19 NPs, cluster 12 (ETH cluster) had 4 CDs/DCs (e.g., thionamides) and 20 NPs, and cluster 28 had 7 CD/DCs (e.g., amoxicillin) and 15 NPs. On the other hand, the clusters containing rifamycins and amikacin had 6 and 5 NPs, respectively. In total, only 21.0% of the NPs (85 compounds) possess similar functional groups as the CDs and DCs. Thus, about 79.0% of anti-TB NPs occupies different functional

Table 4. Comparison between the Characteristics of NPs, CDs, and DCs

	M_{wt}	clogP	rotatable bonds	HBD and HBA	violated rules	permeability	solubility
NP (397)	more than 70% have M_{wt} less than 500 (most having values ranging from 200 to 450)	values ranging from −4.9 to 9.0.	79.6% have 10 or less	94.2% have HBD below 5	39% violate clogP rule	most compounds occupy high permeation space	more than 70% of compounds occupy low but acceptable solubility space
	only 6% have M_{wt} greater than 750	91.9% within 0−8.0	51.4% have 4 or less	95.5% have HBA below 10	21.7% violate the M_{wt} rule		
	no compound has values below 100	61.0% lower than 5.0	12.1% have no rotatable bonds	62.2% have 4 or less HBA	20.4% violate rotatable bonds rule		
		median value ≈ 4		median value for HBD = 1			
CD (25)	more than 70% have M_{wt} less than 500 (most having values ranging from 200 to 450)	values ranging from −9.0 to 6.8.	88.0% have 10 or less	64.0% have HBD below 5	40% violate one rule	most compounds occupy moderate to low permeation space	more than 70% of compounds were within moderate acceptable solubility region
	only 6% have M_{wt} greater than 750	88.0% within −9.0−3.0.	48.0% have 4 or less	84.0% have HBA below 10	36% violate the HBD rule		
	no compound has values below 100	92.0% lower than 5.0	4.0% have no rotatable bonds	76.0% have 4 or less HBA	24% violate the M_{wt} rule		
		median value = 0.4		median value for HBD = 5	16% violate the HBA rule		
DC (13)	53.8% have M_{wt} less than 500 ranging from 300 to 450	values ranging from −2.7 to 7.7.	84.6% have 10 or less	76.9% have HBD below 5	69.2% violate one rule	most compounds occupy high permeation space	more than 70% occupies low but acceptable solubility space
	no compound has values less than 300	69.2% lower than 5.0	30% have values 4 or less	76.9% have HBA below 10	46.2% violate the M_{wt} rule		
	15.4% has values above 750	median value ≈ 4	0 compound has no rotatable bonds	median value for HBD = 1 higher HBA values	46.2% violate HBA rule		
					30.8% violate clogP the rule		
					23.1% violate HBD rule		
					15.4% violate rotatable bonds rule		

dx.doi.org/10.1021/ci300467b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Table 5. Most Important Clusters According to the Number of Compounds Included from Similarity Analysis

Cluster	Example of common functional group	CDs and/or DCs in cluster	Cluster size	Number of NPs
4		Ethambutol, SQ109	21	19
5		-	40	40
12		Prothionamide, Isoniazid, Pyrazinamide, Ethionamide	24	20
17		-	24	24
28		Amoxicillin, Clavulanate, Moxifloxacin, Thioacetazone, Imipenem, *p*-Aminosalicylic acid, Gatifloxacin	22	15
32		-	29	29
40		-	23	23

space from the CDs and DCs. Therefore, NPs active against *Mtb* could be a rich source of potential drugs in novel functionality space and presumably have different modes of action.

The results of similarity analysis and the identified clusters suggest that compounds that are in the same clusters could have the same mechanism of action. This hypothesis should be tested experimentally.

Docking Calculations. The drugs ethambutol and thionamide have been used for many years for the treatment of TB.[1] These drugs inhibit the biosynthesis of mycolic acid which is essential for the structure of the impermeable cell wall of the

dx.doi.org/10.1021/ci300467b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

Mycobacterium. We investigated whether NPs from clusters containing ethambutol (EMB cluster) and thionamides (ETH cluster) respectively possess a similar mechanism of action to these drugs. For this, we docked the compounds contained in these clusters inside two important anti-TB targets to predict their binding energies and possible binding modes using Autodock 4.2.[17,27] Docked conformations of each ligand were clustered and ranked using their binding energies. The best docked conformations were selected, visualized, and analyzed using Autodock tools and Chimera 1.5.2.[19]

Compounds from the EMB cluster were docked to the octyl alpha-D-arabinofuranoside (AFO) binding site in the C-terminal extracellular domain of arabinosyltransferase *Mtb* EmbC (PDB ID 3PTY). In addition, compounds outside this cluster were randomly selected and were also docked to this protein to evaluate whether or not there is a preference for compounds inside the EMB cluster. Analysis of the docked binding energies (BEs) revealed that there is no preference for compounds of the selected cluster which indicates that many compounds in this cluster should not have the same mechanism of action. According to BEs, the best binding affinity was displayed by NP scytoscalarol (Table 6) (BE ∼ −8.3 kcal/mol) which suggests

Table 6. Calculated Binding Energies of Compounds in Cluster 4 to Target Arabinosyltransferase EmbC of *Mtb*

compound	binding energy (kcal/mol)
tetrahydroxysqualene	−3.4
(E)-phytol	−4.8
squalene	−5.0
EJMC51	−4.3
α-humulene	−5.2
SQ109[a]	−6.6
globiferin	−5.3
NPMC18c[b]	−6.1
lincomolide B	−5.0
litseakolide L	−4.1
imberbic acid	−6.3
lecheronol A	−6.0
scytoscalarol	−8.3
oleanolic acid	−7.5
cyanthiwigin C	−6.2
manoalide 25-acetate	−7.3
ethambutol[a]	−4.2
EJMC67[b]	−6.1
NPMC3[b]	−4.3
oplopandiol	−4.1
falcarindiol	−4.2
AFO[c]	−3.9

[a]Known anti-TB drugs. [b]Generic name assigned to differentiate unnamed compounds reported in the literature.[10] [c]AFO, octyl alpha-D-arabinofuranoside.

that this compound should inhibit EmbC. Scytoscalarol is larger than the native ligand AFO molecule that is cocrystallized with the C-terminal domain of EmbC and occupied an upper and smaller position in the binding pocket.[28] For scytoscalarol, the bound conformation was stabilized by strong interactions with polar and nonpolar side chains of the amino acids in the binding pocket. Owing to the presence of H-bond donors in scytoscalarol, it is anchored by two H-bonds involving Asp1014 (Figure 7b). In addition, hydrophobic interactions between Val987, Leu751, and Trp1057, and the aromatic rings positioned the molecule in a way that encourages binding. Asn740, which forms an H-bond with AFO, appears to have a steric effect that keeps scytoscalarol in position, despite the absence of an H-bond.

The ETH cluster contained four currently used anti-TB drugs ethionamide, prothionamide, isoniazid, and pyrazinamide. These four compounds are prodrugs that are activated by different enzymes: catalase-peroxidase for isoniazid, flavin-dependent monooxygenase for both ethionamide and prothionamide, and pyrazinamidase for pyrazinamide. The activated adducts, e.g. ETH-NAD, inhibit the enzyme oxidoreductase of *Mtb*, INHA (PDB ID 2H9I). The binding pocket of this enzyme consists of a hydrophobic site where the thionamide end of the adduct binds and an adjoining fatty substrate-binding site that accommodates the nucleotide end.[29] Validation of the docking calculations was performed by redocking the ETH-NAD adducts. The docked conformation had the same orientation as the original crystal structure. A comparison of the two structures gave a root-mean-square distance of 2.3 Å indicating minimal deviations.

Compounds from the ETH cluster were docked to the INHA binding site. In addition, compounds outside this cluster were randomly selected and docked to the protein to evaluate whether there is a preference for compounds inside the ETH cluster. As in the analysis of the EMB cluster, there is no preference for compounds of the ETH cluster for INHA, which indicates that many compounds in this cluster should not have the same mechanism of action. The best docked conformations were exhibited by the enantiomers of 8-hydroxymanzamine A (BE ∼ −11.6 and −11.1 kcal/mol) and manzamine A (BE ∼ −11.4 kcal/mol) (Table 7). The position of (−)-8-hydroxymanzamine A in the binding pocket is shown in Figure 7d. Due to steric hindrance, this molecule could not be accommodated into the deep and narrow hydrophobic site that houses the ethionamide end of the ETH-adduct and occupied the nucleotide-binding site of the enzyme. (−)-8-Hydroxymanzamine A forms two H-bonds with Gly14 and Thr39 with its hydrophobic end directed toward the hydrophobic binding site bringing the compound close to some of the important residues like Met155. Although no H-bonds were formed with the conserved Ser94, the distance (4.4 Å) between its carbonyl oxygen atom (O) and the second aromatic hydroxyl group in (−)-8-hydroxymanzamine A could indicate formation of a water mediated H-bond, similar to H-bond with phosphate ion in the ETH-NAD adduct.[29]

It is important to point out that the binding energy (BE) values obtained using docking allow scoring of the best orientations, as well as proposing good poses and mechanisms of action for the studied compounds. However, these values are not reliable for the purpose of deriving accurate inhibition constants.[30] Docking programs reproduce the lock-and-key model, where the three-dimensional structure of the ligand and the receptor complement each other, but they do not consider the dynamic of the complexes, the solvent effects, and the change in entropy due to the binding process. Docking methods calculate BE values by using simple scoring functions, which operate very fast. This characteristic allows the use of docking in virtual screening.

It is noteworthy that our docking results indicate that there is no guarantee that compounds in the same cluster (obtained by similarity analysis) have the same mechanism of action. In view of the diverse data set studied, a few clusters were obtained by considering the similarity of molecular characteristics. Other criteria to select clusters, such as the largest common substructures, could lead to a large number of clusters containing compounds with greater similitude (and perhaps clearly the same

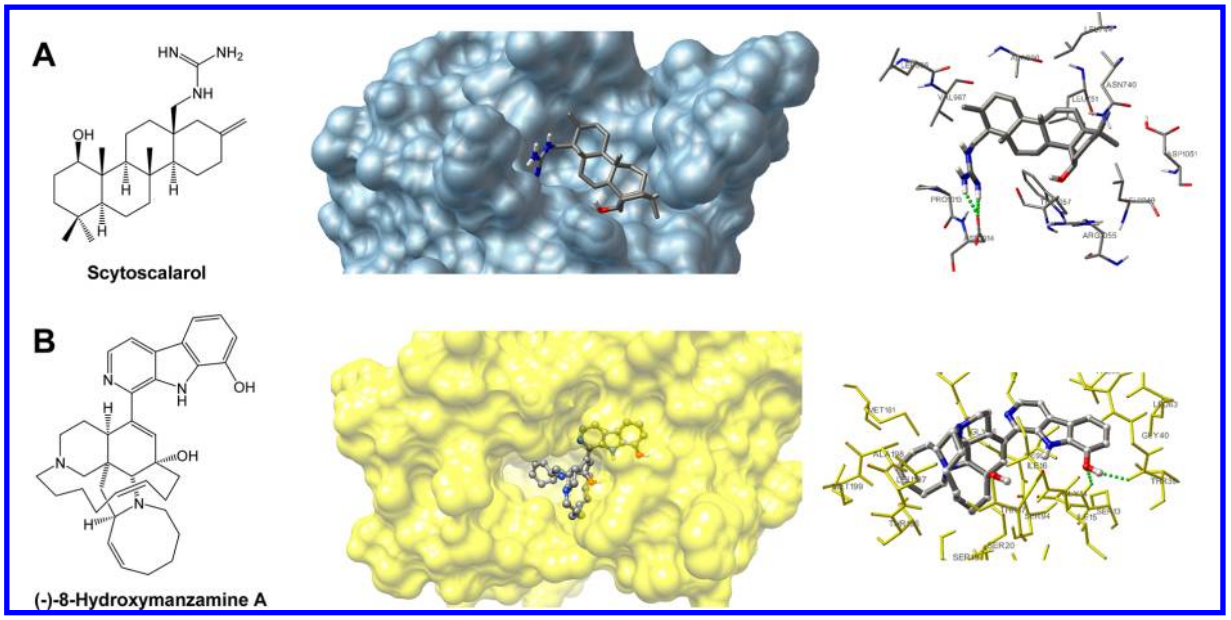


Figure 7. Location of scytoscalarol (A) and (−)-8-hydroxymanzamine A (B) in the binding pocket of *Mtb* arabinosyltransferase EmbC and *Mtb* oxido-reductase INHA, respectively. (left to right) structures of compounds; structures of the complexes; and predicted binding conformations with H-bonds represented in green indicating interaction of residues with NPs in the pocket of the targets.

Table 7. Calculated Binding Energies of Compounds in Cluster 12 to Target *Mtb* Oxido-Reductase Enzyme

compound	binding energy (kcal/mol)
diplamine	−10.3
isodiplamine	−11.0
ascididemin	−9.1
kuanoniamine A	−9.0
11-hydroascididemin	−9.3
shermilamine B	−10.3
NPMC34b[b]	−8.8
ethionamide[a]	−5.9
(−)-manzamine F	−10.9
prothionamide[a]	−6.3
NPMC8a[b]	−7.5
manzamine E	−10.4
NPMC34c[b]	−9.1
EJMC235[b]	−7.1
isoniazid[a]	−6.0
EJMC236[b]	−7.0
Marine46[b]	−9.6
(−)-8-hydroxymanzamine A	−11.6
(+)-8-hydroxymanzamine A	−11.1
manzamine A	−11.4
Marine45[b]	−10.1
hydroxymanzamine E	−10.9
NPMC7a[b]	−6.8
pyrazinamide	−4.8
ethionamide-NAD[c]	−12.7

[a]Known drugs ethionamide, prothionamide, and isoniazid prodrug forms (without the cofactor). [b]Generic name assigned to differentiate unnamed compounds reported in the literature.[10] [c]Ethionamide-NAD activated form.

mechanism of action), but the large number of clusters could greatly complicate the analysis.

In structure-based virtual screening, docking experiments are used to identify potential inhibitors for validated targets and to study their interactions. Similarly, potential targets for biologically active NPs can be identified based on the binding affinities of the compounds and their interactions with important residues in the target's binding pocket. NPs that have shown high binding affinities, e.g. (−)-8-hydroxymanzamine A, manzamine A, and scytoscalarol, with the two targets could be selected for chemical modification through semisynthesis and their inhibitory activities against the target validated experimentally.

CONCLUSIONS

Since ancient times, NPs have been successfully used as therapeutic agents and later, as leads for drug development. In many cases, their amenability to semisynthetic chemical modification has made it possible to determine their structure−activity relationship and such knowledge used to develop even more active drugs.[31]

In this study, NPs with reported antimycobacterial activity did not significantly differ from CDs in their deviation from the Rule of Five and had a better compliance profile than DCs. Conversely, NPs presented similar solubility and permeability profiles as DCs. From cluster analysis of the NPs, 21.0% were found to possess similar functional groups as DCs and CDs. Docking studies on representative *Mtb* targets of drugs in this subset suggested that scytoscalarol can inhibit arabinosyltransferase *Mtb* EmbC and 8-hydroxymanzamine A and manzamine A can bind to INHA. To confirm this, the mechanism of these compounds should be tested experimentally.

ASSOCIATED CONTENT

Supporting Information

Additional details of the predicted physicochemical properties, cluster analysis, and SMILES of the natural products, conventional drugs, and development compounds. This material is available free of charge via the Internet at http://pubs.acs.org.

dx.doi.org/10.1021/ci300467b | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

AUTHOR INFORMATION

Corresponding Author

*Phone: +27-21-6502553. Fax: +27-21-6505195. E-mail: Kelly.Chibale@uct.ac.za.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The University of Cape Town, South African National Research Foundation (NRF), South African Medical Research Council, and South African Research Chairs Initiative of the Department of Science and Technology administered through the NRF are gratefully acknowledged for support.

ABBREVIATIONS USED

AFO, octyl alpha-D-arabinofuranoside; BE, binding energy; CDs, conventional anti-TB drugs; DCs, development compounds: FCFP4, functional class fingerprint; EMB, ethambutol; ETH, ethionamide; HBD, H-bond donors; HBA, H-bond acceptors; LGA, Lamarkian genetic algorithm; MDR, multidrug resistant; MICs, minimum inhibitory concentrations; MIFs, molecular interaction fields; *Mtb*, *Mycobacterium tuberculosis*; M_{wt}, molecular weight; NPs, natural products; PLS, partial least-squares; SPP, similar property principle; TB, tuberculosis; XDR, extensively drug resistant

REFERENCES

(1) Koul, A.; Arnoult, E.; Lounis, N.; Guillemont, J.; Andries, K. The Challenge of New Drug Discovery for Tuberculosis. *Nature* **2011**, *469*, 483−490.

(2) World Health Organization. *Global Tuberculosis Control Report 2011*; World Health Organization: Geneva, Switzerland, 2011; ISBN 9789241564380.

(3) (a) Mitchison, D. A. Role of Individual Drugs in the Chemotherapy of Tuberculosis. *Int. J. Tuberc. Lung Dis.* **2000**, *4*, 796−806. (b) World Health Organization. *Treatment of Tuberculosis: Guidelines*, 4th ed.; World Health Organization: Geneva, Switzerland, 2009; WHO/HTM/TB/2009.420.

(4) Dover, L. G.; Coxon, G. D. Current Status and Research Strategies in Tuberculosis Drug Development. *J. Med. Chem.* **2011**, *54*, 6157−6165.

(5) Caminero, J. A. Treatment of Multidrug-Resistant Tuberculosis: Evidence and Controversies. *Int. J. Tuberc. Lung Dis.* **2006**, *10*, 829−837.

(6) Lienhardt, C.; Raviglione, M.; Spigelman, M.; Hafner, R.; Jaramillo, E.; Hoelscher, M.; Zumla, A.; Gheuens, J. New Drugs for the Treatment of Tuberculosis: Needs, Challenges, Promise, and Prospects for the Future. *J. Infect. Dis.* **2012**, *205*, S241−S249.

(7) (a) Li, J. W.-H.; Vederas, J. C. Drug Discovery and Natural Products: End of an Endless Frontier? *Science* **2009**, *325*, 161−165. (b) Newman, D. J.; Cragg, G. M.; Snader, K. M. The Influence of Natural Products upon Drug Discovery. *Nat. Prod. Rep.* **2000**, *17*, 215−234. (c) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the 30 Years from 1981 to 2010. *J. Nat. Prod.* **2012**, *75*, 311−355. (d) Lachance, H.; Wetzel, S.; Kumar, K.; Waldmann, H. Charting, Navigating, and Populating Natural Product Chemical Space for Drug Discovery. *J. Med. Chem.* **2012**, *55*, 5989−6001.

(8) Chin, Y. -W.; Balunas, M. J.; Chai, H. B.; Kinghorn, A. D. Drug Discovery From Natural Sources. *AAPS J.* **2006**, *8*, E239−E253.

(9) (a) Ganesan, A. The Impact of Natural Products Upon Modern Drug Discovery. *Curr. Op. Chem. Biol.* **2008**, *12*, 306−317. (b) Harvey, A. L. Natural Products in Drug Discovery. *Drug Discov. Today* **2008**, *13*, 894−901.

(10) (a) García, A.; Bocanegra-García, V.; Palma-Nicolás, J. P.; Rivera, G.; Vennerstrom, J. L. Recent Advances in Antitubercular Natural Products. *Eur. J. Med. Chem.* **2012**, *49*, 1−23. (b) Rogoza, L. N.; Salakhutdinov, N. F.; Tolstikov, G. A. Anti-tubercular Activity of Natural Products: Recent Developments. In *Opportunity, Challenge and Scope of Natural Products in Medicinal Chemistry*; Tiwari, V. K., Ed.; Research Signpost: India, 2011; pp 103−120. (c) De Souza, M. V. N. Marine Natural Products Against Tuberculosis. *Sci. World J.* **2006**, *6*, 847−861.

(11) *ChemBioDraw Ultra*, version 11.0.1; CambridgeSoft, Cambridge, MA, 2007 (http://www.cambridgesoft.com).

(12) *Volsurf+*, version 1.0.4; Molecular Discovery Ltd., Perugia, Italy, 2009 (http://www.moldiscovery.com).

(13) Cruciani, G.; Crivori, P.; Carrupt, P.-A.; Testa, B. Molecular fields in quantitative structure-permeation relationships: the VolSurf approach. *J. Mol. Struc. (Theochem)* **2000**, *503*, 17−30.

(14) *VIDA*, version 4.0.3; OpenEye Scientific Software, Santa Fe, NM, 2010 (http://www.eyesopen.com).

(15) *Pipeline Pilot*, version 8.5; Accelrys Software Inc., San Diego, 2011 (http://www.accelrys.com).

(16) *Glide*, version 5.7; Schrödinger L.L.C., New York, 2011 (http://www.schrodinger.com).

(17) *Autodock*, version 4.2; Molecular Graphics Laboratory, The Scripps Research Institute, La Jolla, CA, 2009 (http://autodock.scripps.edu).

(18) Umamaheswari, M.; Madeswaran, A.; Asokkumar, K.; Sivashanmugam, A. T.; Subhadradevi, V.; Jagannath, P. Docking Studies: Search for Possible Phytoconstituents for the Treatment of Gout. *Int. J. Biol. Pharm. Res.* **2012**, *3*, 6−11.

(19) *Chimera*, version 1.5.2; UCSF Resource for Biocomputing, Visualization, and Informatics, San Francisco, 2011 (http://www.cgl.ucsf.edu/chimera).

(20) (a) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3−25. (b) Veber, D. F.; Johnson, S. R.; Cheng, H.-Y.; Smith, B. R.; Ward, K. W.; Kopple, K. D. Molecular Properties That Influence the Oral Bioavailability of Drug Candidates. *J. Med. Chem.* **2002**, *45*, 2615−2623.

(21) Wilcoxon, F. Individual comparisons by ranking methods. *Biometrics Bull.* **1945**, *1*, 80−83.

(22) Ananthan, S.; Faaleolea, E. R.; Goldman, R. C.; Hobrath, J. V.; Kwong, C. D.; Laughon, B. E.; Maddry, J. A.; Mehta, A.; Rasmussen, L.; Reynolds, R. C.; Secrist, J. A., III; Shindo, N.; Showe, D. N.; Sosa, M. I.; Suling, W. J.; White, E. L. High-throughput screening for inhibitors of Mycobacterium tuberculosis H37Rv. *Tuberculosis (Edinb)* **2009**, *89*, 334−353.

(23) Sharma, A.; Subbias, K. K.; Robine, O.; Chaturvedi, I.; Nigam, A.; Sharma, N.; Chaudhary, P. P. Computational finding of potential inhibitor for Cytochrome P450 Mono-oxygenases Enzyme of *Mycobacterium tuberculosis*. *Bioinformation* **2012**, *8*, 931−937.

(24) Cho, Y.; Ioerger, T. R.; Sacchettini, J. C. Discovery of novel nitrobenzothiazole inhibitors for Mycobacterium tuberculosis ATP phosphoribosyl transferase (HisG) through virtual screening. *J. Med. Chem.* **2008**, *51*, 5984−5992.

(25) Ekins, S.; Freundlich, J. S.; Choi, I.; Sarker, M.; Talcott, C. Computational databases, pathway and cheminformatics tools for tuberculosis drug discovery. *Trends Microbiol.* **2011**, *19*, 65−74.

(26) Eckert, H.; Bajorath, J. Molecular Similarity Analysis in virtual Screening: Foundation, Limitation and Novel Approaches. *Drug Discov. Today* **2007**, *12*, 225−233.

(27) Cosconati, S.; Forli, S.; Perryman, A. L.; Harris, R.; Goodsell, D. S.; Olson, A. J. Virtual Screening with AutoDock: Theory and Practice. *Expert Opin. Drug Discov.* **2010**, *5*, 597−607.

(28) Alderwick, L. J.; Lloyd, G. S.; Ghadbane, H.; May, J. W.; Bhatt, A.; Eggeling, L.; Fütterer, K.; Besra, G. S. The C-Terminal Domain of the Arabinosyltransferase *Mycobacterium tuberculosis* EmbC is a Lectin-Like Carbohydrate Binding Module. *PLoS Pathog.* **2011**, *7*, 1−12.

(29) Wang, F.; Langley, R.; Gulten, G.; Dover, L. G.; Besra, G. S.; Jacobs, W. R., Jr.; Sacchettini, J. C. Mechanism of Thioamide Drug Action Against Tuberculosis and Leprosy. *J. Exper. Med.* **2007**, *204*, 73−78.

(30) Caballero, J.; Alzate-Morales, J. H. Molecular dynamics of protein kinase-inhibitor complexes: a valid structural information. *Curr. Pharm. Des.* **2012**, *18*, 2946−2963.
(31) Ganesan, A. Natural Products as Hunting Ground for Combinatorial Chemistry. *Curr. Opin. Biotech.* **2004**, *15*, 584−590.