



Virtual screening against *Mycobacterium tuberculosis* isocitrate lyase and *in silico* ADME-Tox evaluation of Top Hits

Nina Abigail B. Clavio¹ and Junie B. Billones^{1,2*}

¹OVPAA-EIDR Program: Computer-aided Discovery of Compounds for the Treatment of Tuberculosis in the Philippines, Department of Physical Sciences and Mathematics, College of Arts and Sciences

²Institute of Pharmaceutical Sciences, National Institutes of Health, University of the Philippines Manila, Taft Avenue, Ermita, Manila, Philippines

ABSTRACT

Isocitrate lyase (ICL) is a key factor for the maintenance of latent tuberculosis infection. ICL catalyses the first committed step of the carbon-conserving glyoxylate bypass, the reversible cleavage of isocitrate into succinate and glyoxylate. Since *Mycobacterium tuberculosis* (MTb) has no back-up mechanism that can take over the role of ICL once it is inhibited, the ICL enzyme is an attractive target for inhibition and discovery of anti-TB drugs. Structure-based pharmacophore generation has been used here in virtual screening of database compounds and *de novo* evolution method was employed in subsequent hit optimization. Accordingly, the structure of ICL was modeled and a pharmacophore was generated based on the structure of the binding site. Subsequent pharmacophore-based screening of one million compounds yielded 17 hits with greater binding energies than that of the natural substrate. The derivative of each hit showed even much stronger binding affinities. Both set of original and modified hit compounds were evaluated *in silico* for their ADME-Tox properties. The results showed that Ligands O (242372) and P (Amb9999830), and derivatives of D (Compound2099^M), E (Compound3796^M), K (Compound556^M), N (STOCK1N-12208^M), and P (Amb9999830^M) possess promising drug-like properties and can be pursued as leads in the search for novel antitubercular agents.

Keywords: *Mycobacterium tuberculosis*; anti-tuberculosis compounds; isocitrate lyase; pharmacophore; ADMET; TOPKAT; computer-aided drug discovery

INTRODUCTION

Tuberculosis is one of the most widespread infections today than at any other time in human history [1]. Currently, one-third of the world's population is infected with *Mycobacterium tuberculosis* [2]. The latest estimates of the number of people in the world with TB are almost 9 million new cases in 2011 and 1.4 million deaths despite the availability of treatment that will cure most cases of TB [3].

In addition to the large number of tuberculosis cases, TB has become a global public health problem because of its resistance to frontline drugs such as the InhA-inhibitor isoniazid [4]. The emergence of multiple drug-resistant, extensive drug-resistant strains and its association with HIV has severely affected the fight against TB. If this continues, it is anticipated that there will be about 8.9 to 9.9 million new and relapse TB cases this year, more than any other in history [2]. Thus, there is a great need to develop new drugs that resist persistent TB infection.

Mycobacterium tuberculosis remains persistent in macrophages and gains energy through the glyoxylate pathway bypass to maintain its long-term existence in host cells. Therefore it is possible to stop persistent infections by interrupting the glyoxylate bypass in which the enzyme isocitrate lyase plays a main role [5]. Isocitrate lyase is a key rate-limiting enzyme in the glyoxylate cycle [6] and is a very important factor in the persistence of MTb [5], [7]. It is a

unique enzyme that plays a role in the glyoxylate cycle, an anaplerotic pathway of the tricarboxylic acid cycle[8]. This allows bacteria to grow on acetate or fatty acids as sole carbon sources[9]. ICL has been found to be essential for survival in the host [10], [11]. Since the glyoxylate pathway has not been observed in mammals[12], ICL is considered an appropriate target of new anti-tubercular drugs.

Experimental techniques used in identification of inhibitors of *Mtb* growth are very expensive, time-consuming, tedious, and requires sophisticated systems for controlling the risk of infection. This is where theoretical techniques such as *in silico* docking come of importance. Nowadays, rational drug discovery cannot be accomplished without the use of important disciplines such as chemoinformatics and bioinformatics. In fact, in the field related with the discovery of more effective anti-TB drugs, several families of compounds have been discovered with the application of computational approaches[13], [14], then synthesized and tested as anti-TB agents through inhibition of different targets proteins in *MTb* such as DNA gyrase subunit A, DNA gyrase subunit B, Enoyl-[acyl-carrier-protein] reductase (InhA), Fibonectin-binding protein C, Pantothenate synthetase, and Peptide deformylase[15]. *In silico* determination of compounds with potential anti-TB effects have indeed been very useful in establishing safer, less-expensive, shorter-time experiments. In this work, one million compounds were screened using a pharmacophore model generated from the structure of ICL. After performing rigid and flexible fitting, the high-scoring compounds were subsequently docked to the ICL model, then the top hits were rank-ordered based on calculated binding energies. The high-binding compounds were then subjected to structural modification using *de novo* evolution technique. The top hits and derivatives were finally evaluated *in silico* for their ADME (absorption, distribution, metabolism, excretion) and toxicity properties.

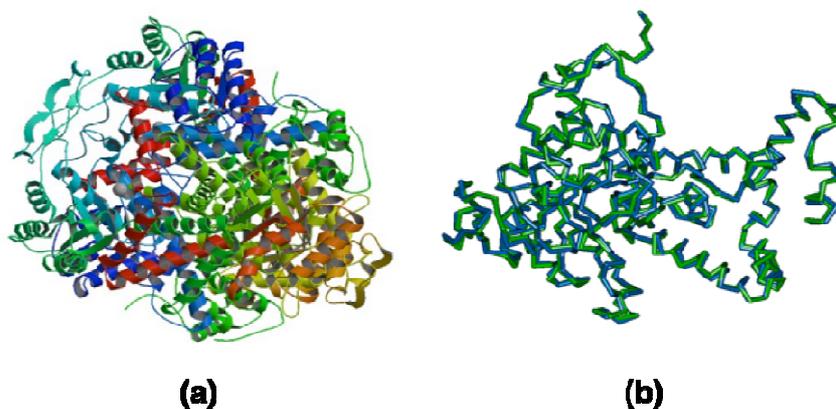


Figure1. (a) Ribbon diagram of ICL from *Mycobacterium tuberculosis*; (b) Molecular overlay representation of the original (green) and the prepared (blue) ICL structure (RMSD = 0.980Å)

EXPERIMENTAL SECTION

All calculations were done using the Discovery Studio v2.5.5 software [16].

Structural data of ICL protein and compound libraries

The 2.00Å-resolution 3D crystal structure of isocitrate lyase (PDB ID: 1F61) was retrieved from RCSB Protein Database (www.rcsb.org). The protein was prepared using the *Prepare Protein* protocol using the default parameters. Following preparation is the *Minimization* protocol using the default parameters. The RMSD of the prepared protein was calculated using the *Superimpose Proteins* tool.

The following databases were downloaded online: Ambinter (www.ambinter.com), AMRI (www.amriglobal.com), Analyticon MEGx and Analyticon NATx (www.ac-discovery.com), Drug Bank (www.drugbank.ca), InterBioScreen (www.ibscreen.com), Otava (www.otavachemicals.com), and Specs (www.specs.net). The compounds from the database were prepared using *Prepare Ligands* protocol. All parameter values were set as default except for the Lipinski filter, which was turned off.

Generation of Structure-based Pharmacophore Model

The pharmacophore was generated based on the binding site of ICL. The cavity located within the crystal structure agrees with the location of the catalytic site, thus, making it a favorable location for pharmacophore generation[17]. The selected binding site locates Cys191, a residue to which 3-bromopyruvate, a known inhibitor to ICL, binds covalently for effective inhibition. The binding site sphere was defined using the *Binding Site* tool. The *Interaction Generation* protocol was used to generate the pharmacophore model containing the hydrophobic, H-donor, and H-

acceptor features that are possible sources of interaction in the protein active site. The generated pharmacophore exhibits 28 features: 8 H-acceptors, 9 H-donors, 11 hydrophobes (Figure 1B).

Screening of Compound Libraries

The preparation of one million compounds was done using the *Prepare Ligands* protocol; generating conformations of each compound. The prepared ligands were then consolidated as one database using *Build 3D Database* protocol. The database compounds were subsequently screened using the *Screen Library* protocol, employing both rigid fitting and flexible fitting methods. The compounds with fit values better than that of 3-bromopyruvate (as reference) were subjected to further screening based on binding energies from docking calculations.

Molecular Docking

CDOCKER protocol was used for docking each hit to ICL. *Calculate Binding Energies* protocol was used for computing binding affinity of each complex. As control, the binding energy value of the 3-bromopyruvate-ICL complex was used as baseline. All compounds with better binding energy than 3-bromopyruvate were selected for further *in silico* ADMET screening.

In silico ADME-Toxicity

The compounds with significantly better binding energy calculations than 3-bromopyruvate were further screened through ADMET filters using the *ADMET* and the *TOPKAT* protocols in DS 2.5.

RESULTS AND DISCUSSION

The preparation of ICL crystal structure (Figure 1A) prior to virtual screening involves removal of bound water molecules, insertion of missing atoms in incomplete residues, optimization of side-chain conformation, modeling of missing loop regions, and removal of alternate conformations. Following protein preparation is minimization step, where the most stable protein conformation is calculated. Minimization relatively changes the conformation of the original protein structure. However, superimposition of the original and optimized structures revealed an RMSD of 0.980 Å (Figure 2A), which is well within the acceptable range [18].

The preparation of the ligands includes the removal of duplicate structures, generation of isomers and tautomers, and generation of 3D conformations [16]. The Lipinski filter was not applied because several studies have shown that some compounds, including many natural products, have become successful drug candidates despite violating the Lipinski rule.

Virtual screening allows rapid selection and testing of a small subset of compounds predicted to have significant interactions with the given biological target out of a large database of molecules. It is used to identify potential leads from a pool of compounds to reduce the number to be screened experimentally. The pharmacophore filtering method has been proven to perform better than traditional docking and scoring methods [19], [20], [21].

Two methods of screening were applied: rigid and flexible. In rigid fitting, the conformations are held rigid and the best fit is computed using the Kabsch algorithm. In flexible fitting, the conformations are manipulated within a specified energy threshold to minimize the distances between pharmacophore features and mapped atoms on the molecule [16].

In comparing a ligand and a pharmacophore, the quality of the mapping is indicated by the fit value. A higher fit value represents a better fit; a perfect mapping of features would result in a fit value equivalent to the sum of the weights of the features in the pharmacophore. The computed fit value depends on two parameters: the weights assigned to the pharmacophore features and how close the features in the molecule are to the centers of the corresponding location constraints of the pharmacophore.

Compounds that passed the first screening and the second screening were docked and the binding energies were determined. Docking small molecules into larger protein molecules is a complex and difficult task. A docking algorithm called *CDOCKER* was used for this. *CDOCKER* is a *CHARMm* (Chemistry at *HARvard* Macromolecular Mechanics) based docking tool that generates random ligand poses and places them in a rigid receptor [16]. This type of docking process has been applied and successful in developing novel drugs from past studies with a significant advance in algorithms that make possible the rapid docking of very large collections of small molecules into the chosen molecular target [21].

3-bromopyruvate, a known inhibitor of ICL, bonds covalently with Cys191, replacing bromine (Figure 2B). Polar and hydrogen bonding interactions are also observed with the side chains of His193, Asn313, Ser315, Ser317 and Thr347. *In silico* docking of the said inhibitor generates the following conserved interactions: polar and hydrogen bonding with His193, Asn313 and polar interaction with Thr347. The covalent interaction of the inhibitor with Cys191 is represented by the hydrogen bonding of the pair. The binding energy of the bromopyruvate-ICL complex was computed to be -241.37 kcal/mol.

Poses were evaluated by calculating their binding energies (BE) (Table 1). A negative value represents an exergonic process involved in formation of complexes. A higher magnitude of BE indicates a larger amount of energy released, meaning the complex is more stable. Thus, ligands with more negative binding energies are considered better hits than the others.

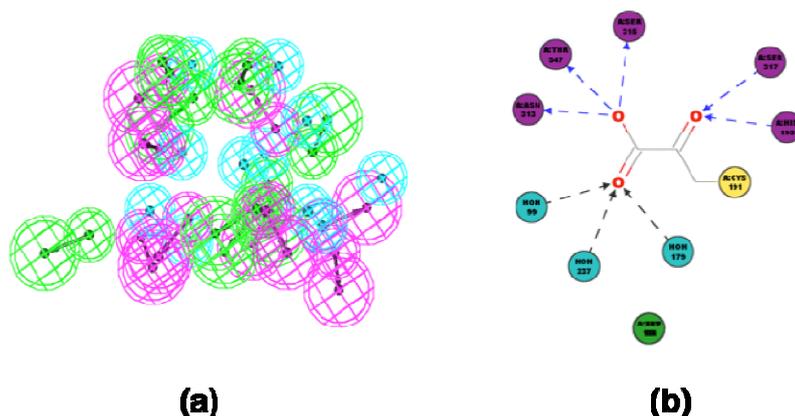


Figure 2. (a) Structure-based pharmacophore model having 28 features: 8 acceptors (green), 9 donors (magenta), and 11 hydrophobes (cyan); (b) Interaction diagrams of 3-bromopyruvate with ICL from experimental study [17]

Table 1. Fit values and binding energies of 1F61 with isocitrate, 3-bromopyruvate, and all compounds that passed the first and second screening tests

Compound ID	Fit Value (Rigid Fitting)	Fit Value (Flexible Fitting)	ΔG (kcal/mol)
Isocitrate	1.92604	1.99821	-220.64900
Bromopyruvate	1.97716	2.04992	-241.36700
A Compound3076	3.62451	4.29447	-470.11200
B Compound1596	3.58727	4.21420	-440.97200
C Compound2563	3.39119	4.24184	-419.80400
D Compound2099	3.57533	4.35659	-417.10300
E Compound3796	3.34673	4.54883	-407.94900
F Compound1842	3.39011	4.17372	-391.36800
G Compound3284	3.61701	4.47318	-381.42800
H Compound3092	3.41389	4.16772	-373.59000
I Compound2307	3.88637	4.14698	-373.28300
J Compound1134	4.11509	4.54553	-337.38327
K Compound556	3.06064	4.16101	-289.22700
L 242621	3.82169	4.05676	-255.26800
M Compound29	3.51158	4.16257	-253.95800
N STOCK1N-12208	3.19305	4.06212	-252.43300
O 242372	3.32685	4.18598	-243.30200
P Amb9999830	3.01088	4.04842	-226.22300
Q STOCK1N-57131	3.11350	4.00923	-221.37000
R Amb16506792	3.36358	4.20179	-207.51100
S Compound1249	3.40000	4.03323	-180.54665
T Amb16588329	3.82795	4.01267	-173.11400
U Amb7968888	3.82174	4.29916	-172.20600
V Amb1909647	3.85222	4.2066	-171.78100
W Amb16679279	3.63792	4.09489	-155.72400
X Amb16651149	3.56912	4.39881	-139.63900

Only 17 compounds out of the 24 that passed the screening tests are predicted to have inhibitory activity against ICL. These top 17 were considered for structural modification. The data on the number and type of interactions that the top 17 compounds make with ICL are summarized in Table 2. It can be seen that the top hit, Compound3076, has more polar and van der Waals interactions with ICL than the isocitrate-ICL complex. Most of the residues found from the isocitrate-ICL complex have been observed to have interactions with Compound3076 (Figure 4A). It has

hydrogen bonding interactions with the side chains of Glu155, Glu182, His393 and Gln394, van der Waals interactions with Pro107, Gln109, His180, Gln184, Leu185, Ala186, Lys189, Gly196, Ala234, Thr286, Trp283, Gly287, Asn313, Ser315, Pro316, Ser317, Phe318, Thr347, Leu348, charge interactions with Glu155, Glu182 and Lys197, and several polar interactions with Trp93, Asp108, Ser110, Asp153, Glu155, Ala156, Glu182, Gly195, Lys197, Arg228, Ala233, Thr235, Glu285, His393 and Gln394. A lot of the mentioned interactions are absent in isocitrate, which may be the cause of its lower binding energy compared with Compound3076. This observation is also evident with the other top hits. Structures of the top 17 hits are shown in Figure 3.

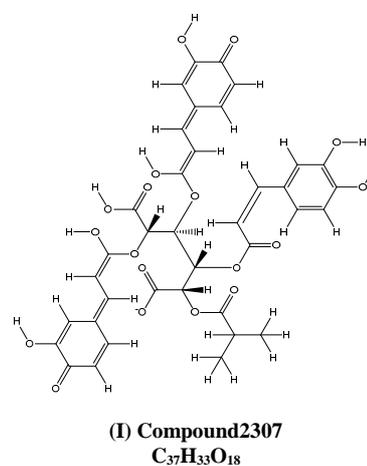
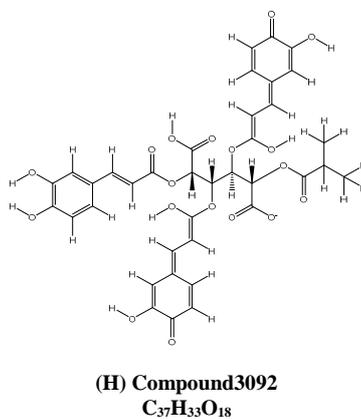
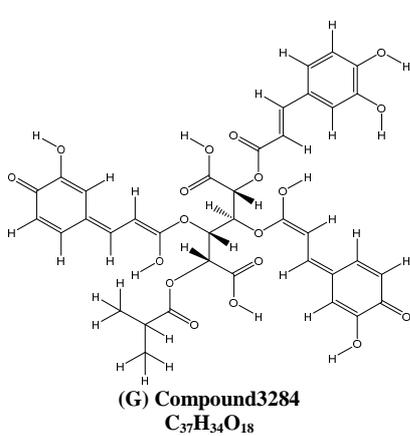
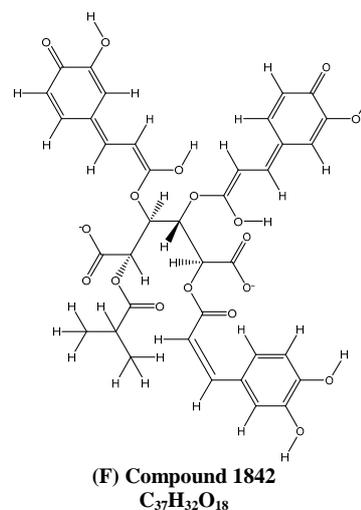
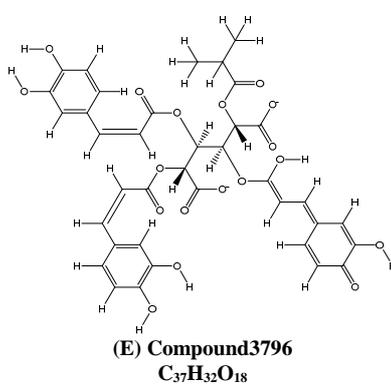
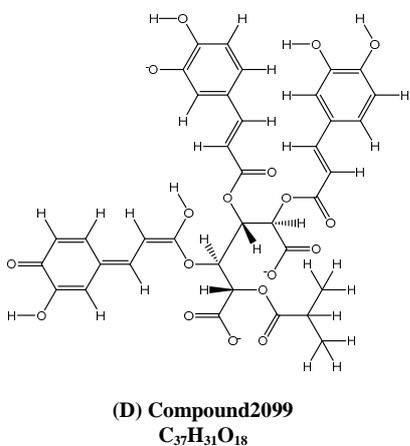
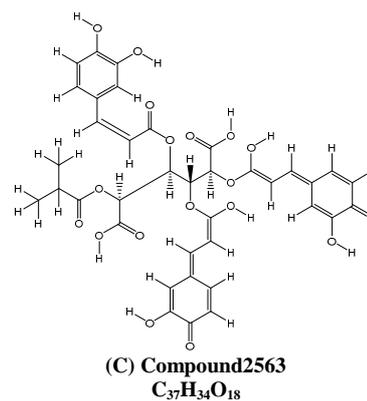
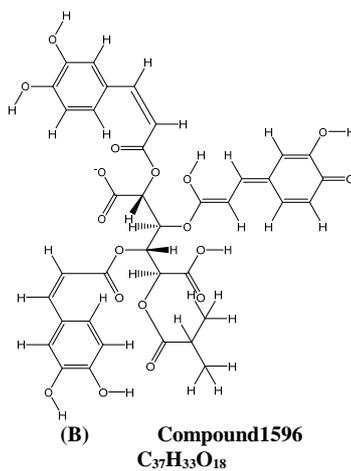
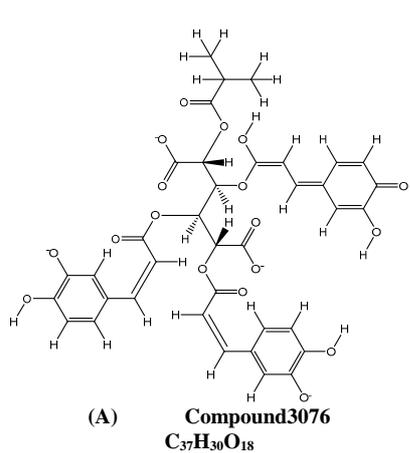
Furthermore, it was observed that most of the structures of the hit compounds consist of subunits similar to that of the structure of isocitrate. This is indicative of the predicted binding interactions with ICL and the superior binding affinities in comparison with its natural substrate.

Table 2. Number of interactions in each of the top ICL-ligand complexes

Compound	ΔG (kcal/mol)	Polar	VDW	H-bonds	Charged	Pi
Isocitrate	-220.649	14	2	5	8	0
Bromopyruvate	-241.367	14	2	3	8	0
A Compound3076	-470.112	15	19	4	3	0
B Compound1596	-440.972	13	15	5	0	1
C Compound2563	-419.804	14	16	5	0	1
D Compound2099	-417.103	14	14	5	2	0
E Compound3796	-407.949	17	11	4	1	0
F Compound1842	-391.368	19	12	7	4	0
G Compound3284	-381.428	14	13	5	0	1
H Compound3092	-373.590	12	15	3	0	3
I Compound2307	-373.283	21	11	3	1	1
J Compound1134	-337.383	14	13	3	0	0
K Compound556	-289.227	11	19	3	0	0
L 242621	-255.268	16	10	1	0	1
M Compound29	-253.958	14	10	5	0	1
N STOCK1N-12208	-252.433	10	19	1	0	1
O 242372	-243.302	13	16	5	0	0
P Amb9999830	-226.223	9	15	1	0	3
Q STOCK1N-57131	-221.370	7	20	3	0	1

Chemical modification. With the desire to seek out compounds with better binding energies, the top 17 compounds were modified using the *De Novo Evolution* protocol. This method allows for the generation of multiple molecules by adding one fragment at a time with only the fittest retained in the population[16]. An advantage of the *de novo* approach is that the program automatically maximizes binding by exhausting all the available chemical space. The top ranking molecules (in terms of binding energy) of each original molecule were considered.

The modifications to each ligand have resulted to better (more negative) binding energies (Table 3). Each complex has had an increase in the number of interactions, varying depending on which addition has been most helpful in giving better binding energies. Compound3076 still has the highest binding energy even after modification, while the rest of the top 17 have rearranged their rankings. Unmodified Compound3076 (Figure 4A) has a BE of -470.11 kcal/mol. It has increased to -540.623 kcal/mol upon addition of dimethylamine which causes an additional hydrogen interaction with Ala233 at the active site (Figure 4B). Structures of each modified compounds are shown in Figure 5.



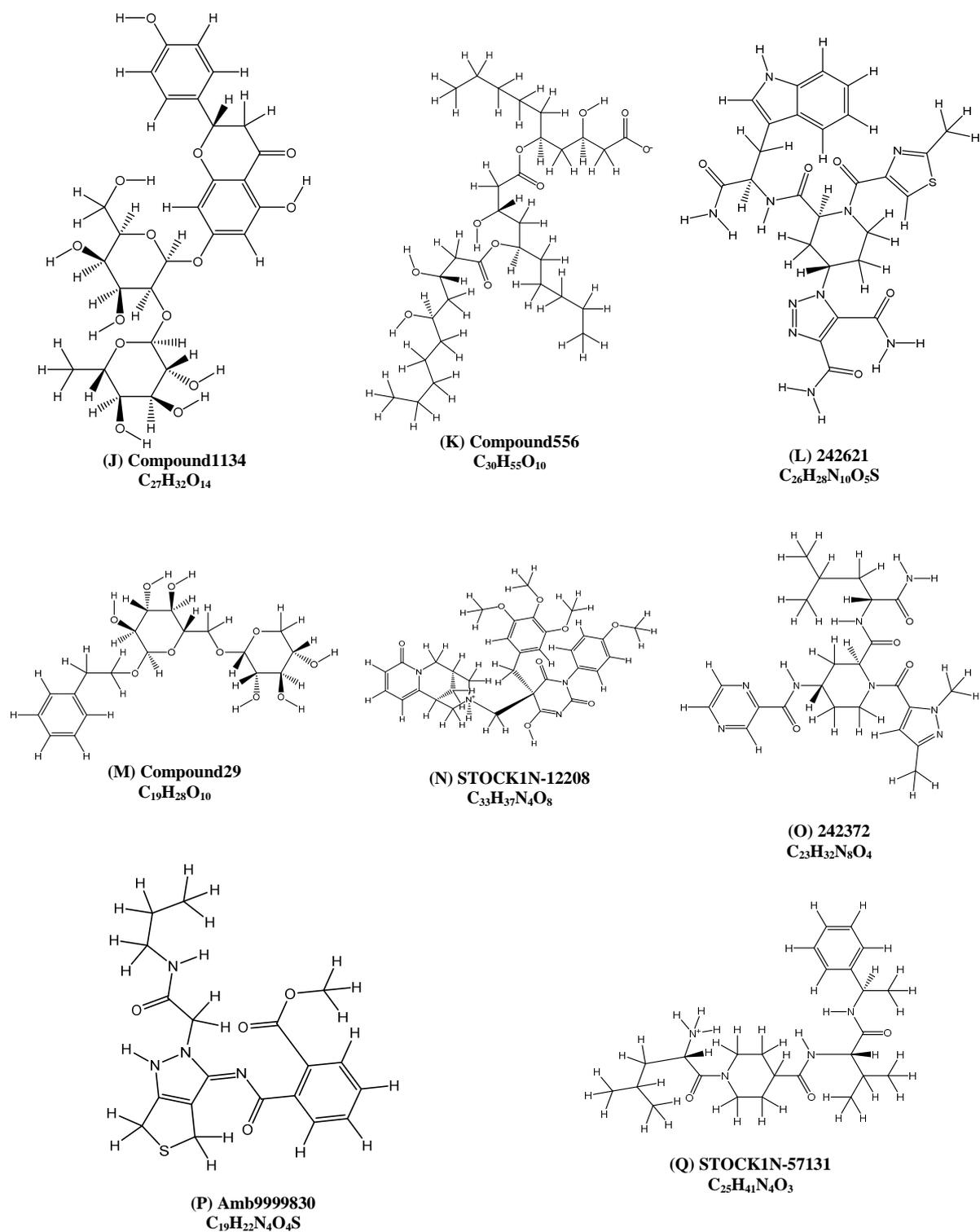


Figure 3. Structures of the top 17 hits after rigid and flexible screening, and molecular docking

Table 3. Binding energies of the best derivative of the top hits (superscript M = modified)

	Compound ID	Binding Energy (kcal/mol)
A	Compound3076 ^M	-540.623
B	Compound1596 ^M	-496.211
C	Compound2563 ^M	-486.759
D	Compound2099 ^M	-492.706
E	Compound3796 ^M	-523.500
F	Compound1842 ^M	-448.255
G	Compound3284 ^M	-445.778
H	Compound3092 ^M	-430.567
I	Compound2307 ^M	-434.805
J	Compound1134 ^M	-383.698
K	Compound556 ^M	-431.072
L	242621 ^M	-331.955
M	Compound29 ^M	-344.361
N	STOCK1N-12208 ^M	-345.253
O	242372 ^M	-332.275
P	Amb9999830 ^M	-326.073
Q	STOCK1N-57131 ^M	-351.776

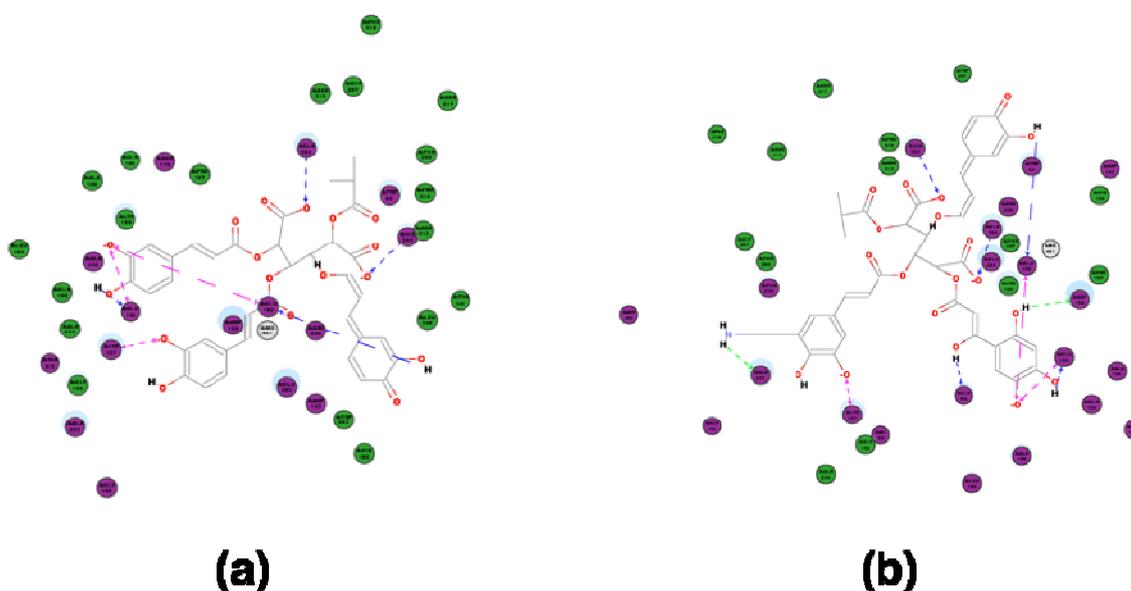
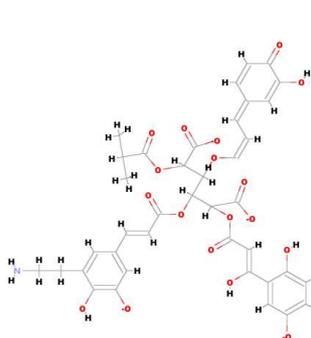
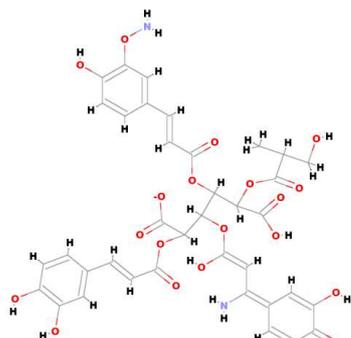


Figure 4. (a) Interaction diagram of Compound3076 and ICL complex; (b) Diagram of the interactions between ICL and the best derivative of Compound3076

Moreover, we have also performed ADMET calculations to further identify screening hits which have potential ADMET issues. Accordingly, we calculated the carcinogenicity, mutagenicity, developmental toxicity potential, and aerobic biodegradability properties of the top hits and their derivatives. The TOPKAT module in Discovery Studio utilizes statistical models to predict specific toxicological effects solely from chemical structure. It uses a patented algorithm that determines whether a structure lies within the Optimum Prediction Space (OPS) of a respective model. Within the OPS, the model is applicable. If the result of the toxicity assessment of a compound is well within the OPS limit, the prediction or computed probability is relatively accurate or has a high chance of being right. If the assessment is within OPS, but the assessment is that the compound is not toxic then there is a high chance that indeed it is not toxic. If the assessment is within OPS and the assessment is that the compound is toxic, then there is a high chance that the compound is toxic [16]. A good drug has low toxicity issues: is not hepatotoxic; has no cytochrome P450 induction liabilities; should be relatively stable while travelling in the system for it to reach the pathogen; and it should not be degraded by the enzymes of the human host. Furthermore, *ADMET Descriptors* protocol in DS help in predicting drug-like properties such as good human intestinal absorption, optimal aqueous solubility, non-CYP2D6 inhibitor, non-hepatotoxic, and has less than 90% plasma binding protein probability.

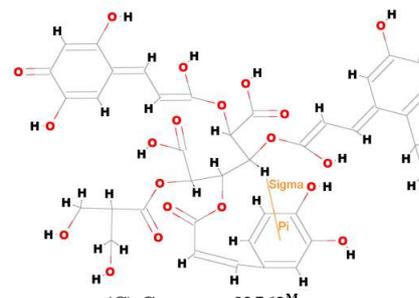


(A) Compound3076^M

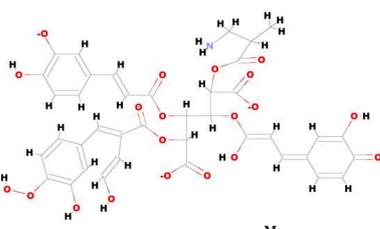


Compound1596^M

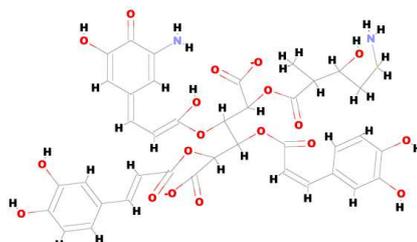
(B)



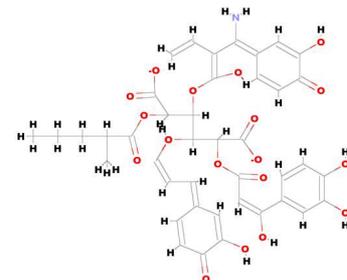
(C) Compound2563^M



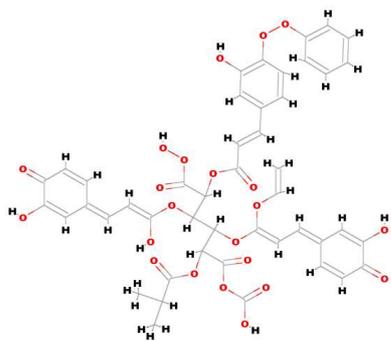
(D) Compound2099^M



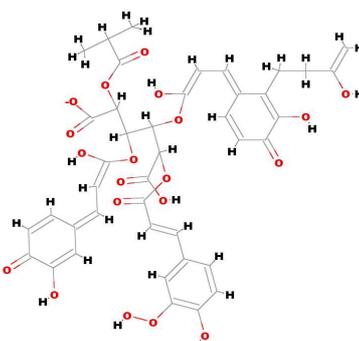
(E) Compound3796^M



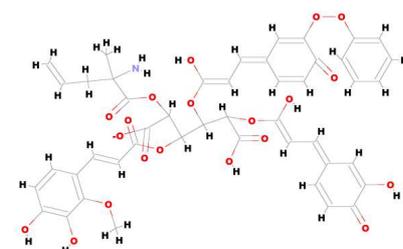
(F) Compound1842^M



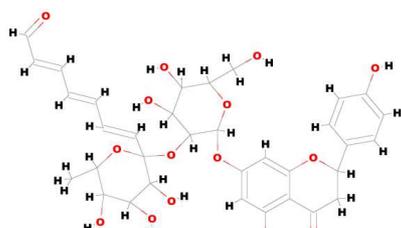
(G) Compound3284^M



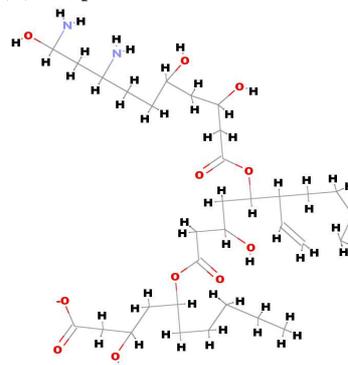
(H) Compound3092^M



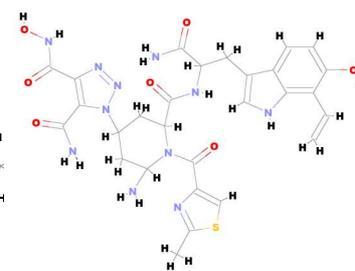
(I) Compound2307^M



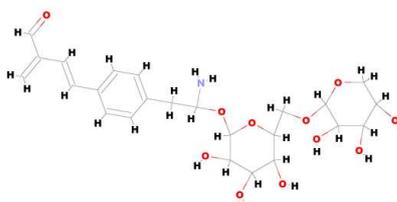
(J) Compound1134^M



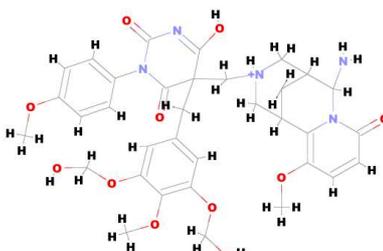
(K) Compound556^M



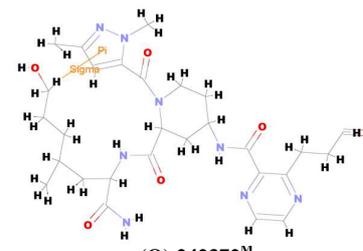
(L) 242621^M



(M) Compound29^M



(N) STOCKIN-12208^M



(O) 242372^M

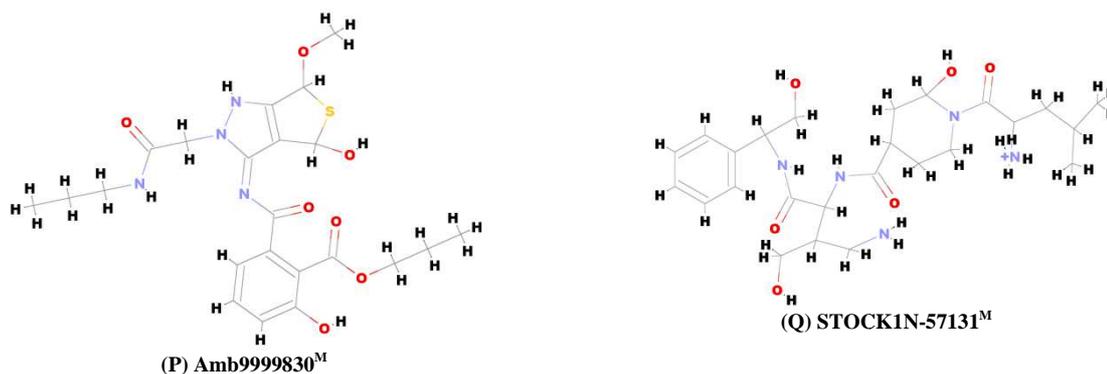
Figure 5. Structures of the best derivative of top 17 hits after *de novo* evolution optimization

Table 4. Predicted toxicity properties for the top hits

Compound ID	Weight of Evidence Carcinogenicity Call	Ames Mutagenicity	Developmental Toxicity Potential	Aerobic Biodegradability	
A	Compound3076	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
B	Compound1596	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
C	Compound2563	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
D	Compound2099	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
E	Compound3796	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
F	Compound1842	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
G	Compound3284	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
H	Compound3092	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
I	Compound2307	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
J	Compound1134	Non-carcinogenic	Non-mutagenic	Toxic	Biodegradable
K	Compound556	Non-carcinogenic	Non-mutagenic	Non-toxic	Biodegradable
L	242621	Non-carcinogenic	Non-mutagenic	Toxic	Biodegradable
M	Compound29	Non-carcinogenic	Non-mutagenic	Toxic	Biodegradable
N	STOCK1N-12208	Non-carcinogenic	Mutagenic	Toxic	Non-degradable
O	242372	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
P	Amb9999830	Non-carcinogenic	Non-mutagenic	Non-toxic	Non-degradable
Q	STOCK1N-57131	Carcinogenic	Non-mutagenic	Non-toxic	Non-degradable

Table 5. Predicted toxicity properties for the best derivative of the top hits. (superscript M = modified)

Compound ID	Weight of Evidence Carcinogenicity Call	Ames Mutagenicity	Developmental Toxicity Potential	Aerobic Biodegradability	
A	Compound3076 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Degradable
B	Compound1596 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
C	Compound2563 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Degradable
D	Compound2099 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
E	Compound3796 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
F	Compound1842 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Degradable
G	Compound3284 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
H	Compound3092 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Degradable
I	Compound2307 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
J	Compound1134 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Degradable
K	Compound556 ^M	Non-carcinogenic	Non-mutagenic	Non-toxic	Non-degradable
L	242621 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Degradable
M	Compound29 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Degradable
N	STOCK1N-12208 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable
O	242372 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Degradable
P	Amb9999830 ^M	Non-carcinogenic	Non-mutagenic	Non-toxic	Non-degradable
Q	STOCK1N-57131 ^M	Non-carcinogenic	Non-mutagenic	Toxic	Non-degradable

Table 6. Predicted ADMET properties for the top hits

Compound ID	Absorption	Solubility	CYP2D6 Binding	Hepatotoxicity	Plasma Protein Binding
A Compound3076	Very low	Extremely low	Non-inhibitor	Toxic	Binding > 90%
B Compound1596	Very low	Extremely low	Non-inhibitor	Toxic	Binding > 90%
C Compound2563	Very low	Extremely low	Non-inhibitor	Toxic	Binding > 90%
D Compound2099	Very low	Extremely low	Non-inhibitor	Toxic	Binding > 90%
E Compound3796	Very low	Extremely low	Non-inhibitor	Toxic	Binding > 90%
F Compound1842	Very low	Extremely low	Non-inhibitor	Toxic	Binding > 90%
G Compound3284	Very low	Extremely low	Non-inhibitor	Toxic	Binding > 90%
H Compound3092	Very low	Extremely low	Non-inhibitor	Toxic	Binding > 90%
I Compound2307	Very low	Extremely low	Non-inhibitor	Toxic	Binding > 90%
J Compound1134	Very low	Yes, low	Non-inhibitor	Toxic	Binding > 95%
K Compound556	Very low	Yes, optimal	Non-inhibitor	Non-toxic	Binding < 90%
L 242621	Very low	Yes, low	Non-inhibitor	Non-toxic	Binding < 90%
M Compound29	Very low	Yes, optimal	Non-inhibitor	Toxic	Binding > 90%
N STOCK1N-12208	Moderate	Yes, good	Non-inhibitor	Toxic	Binding < 90%
O 242372	Very low	Yes, optimal	Non-inhibitor	Non-toxic	Binding < 90%
P Amb9999830	Good	Yes, good	Inhibitor	Non-toxic	Binding < 90%
Q STOCK1N-57131	Good	Yes, optimal	Non-inhibitor	Non-toxic	Binding < 90%

Table 7. ADMET results for the best derivative of the top hits (superscript M = modified)

Compound ID	Absorption	Aqueous Solubility	CYP2D6 Binding	Hepatotoxicity	Plasma Protein Binding
A Compound3076 ^M	Very low	Yes, low	Non-inhibitor	Toxic	Binding < 90%
B Compound1596 ^M	Very low	Extremely low	Non-inhibitor	Toxic	Binding < 90%
C Compound2563 ^M	Very low	Extremely low	Non-inhibitor	Toxic	Binding < 90%
D Compound2099 ^M	Very low	Yes, low	Non-inhibitor	Toxic	Binding < 90%
E Compound3796 ^M	Very low	Yes, low	Non-inhibitor	Toxic	Binding < 90%
F Compound1842 ^M	Very low	Yes, low	Non-inhibitor	Toxic	Binding < 90%
G Compound3284 ^M	Very low	Very low	Non-inhibitor	Toxic	Binding > 95%
H Compound3092 ^M	Very low	Very low	Non-inhibitor	Toxic	Binding < 90%
I Compound2307 ^M	Very low	Very low	Non-inhibitor	Toxic	Binding > 90%
J Compound1134 ^M	Very low	Yes, low	Non-inhibitor	Toxic	Binding > 95%
K Compound556 ^M	Very low	Yes, good	Non-inhibitor	Non-toxic	Binding < 90%
L 242621 ^M	Very low	Very low	Non-inhibitor	Toxic	Binding < 90%
M Compound29 ^M	Very low	Yes, good	Non-inhibitor	Toxic	Binding < 90%
N STOCK1N-12208 ^M	Very low	Yes, low	Non-inhibitor	Toxic	Binding < 90%
O 242372 ^M	Very low	Yes, optimal	Non-inhibitor	Non-toxic	Binding < 90%
P Amb9999830 ^M	Very low	Yes, good	Non-inhibitor	Toxic	Binding < 90%
Q STOCK1N-57131 ^M	Very low	Too soluble	Non-inhibitor	Non-toxic	Binding < 90%

Of the 17 hits (*i.e.* original database compounds), Ligand P (Amb9999830) possesses all the favorable characteristics of a drug (Table 4). Ligands A to I and O also satisfy these characteristics except that they are predicted to be toxic in the developmental process. This means that they have the potential to be drugs but may not be administered to pregnant women. However, the OPS limits for all compounds indicate that the unfavorable predictions are outside the confidence level and that there is good chance that these compounds will still turn out to be non-toxic. The same toxicity calculations were done to the modified compounds (Table 5) to see if there would be any drug-likeness issues arising from structure elaboration. Ligands K and P have satisfied all favorable features while Ligands B, D, E, G, I, and N have satisfied all except that they tested positive to developmental toxicity potential. However, these compounds can still be further pursued as candidate compounds, albeit with possible undesirable effects for pregnant women. Again, upon consideration of the OPS limits, all compounds could still turn out to be non-toxic. Table 6 shows that among the 17 original hits, Ligand Q (STOCK1N-57131) is predicted to possess all five drug-like characteristics. Others that passed all, but one feature are Ligands K, L, N, O, and P. Moreover, the same ADMET calculations were done to the modified hits (Table 7). The results showed that the molecules were predicted to have very low absorption ability, which is not a favorable property for an oral drug. Nevertheless, they can still be delivered to the target site through alternative modes including the use of appropriate drug delivery systems. Finally, the best top hit derivative based on predicted ADMET properties are Ligands K and O.

CONCLUSION

Structure-based pharmacophore generation, virtual screening, rigid-body docking, *De Novo* optimization, and *in silico* ADME-Toxicity calculations were employed in this study to search for possible inhibitors of *Mycobacterium tuberculosis* Isocitrate Lyase. The pharmacophore consisting of 28 clustered features was used for virtual screening of several chemical databases that sums up to 1 million compounds. ICL-ligand binding energy calculations revealed 17 top hits with more favorable binding energies than the natural substrate, indicating possible inhibitory activities. Subsequently, the top hits were modified using the *de novo* method and gave even better binding energies

for each top hit. Calculation of toxicity and ADMET properties of the hit compounds have resulted to the prediction of possible leads, namely, unmodified Ligands O (242372) and P (Amb9999830), and derivatives of D (Compound2099^M), E (Compound3796^M), K (Compound556^M), N (STOCK1N-12208^M), and P (Amb9999830^M). The preparation and experimental bioactivity measurements on these candidate compounds are underway in our group.

Acknowledgement

This work is supported by the Office of the Vice President for Academic Affairs (OVPA) of the University of the Philippines (UP) System through the Emerging Inter-Disciplinary Research (EIDR) Program (OVPA-EIDR 12-001-121102).

REFERENCES

- [1] A Koul; E Arnoult; N Lounis, *Nature*, **2011**, 469, 483-490.
- [2] Y-H Li; H-G Fu; F Su; L-M Gao; S Tang; C-W Bi; Y-H Li; Y-X Wang; D-Q Song, *Chem. Cent. J.*, **2013**, 7, 117.
- [3] WHO. World Health Organization Global Tuberculosis Report 2013. http://www.who.int/tb/publications/global_report/en. (Accessed December 2013).
- [4] RC Hartkoorn; C Sala; J Neres; F Pojer; S Magnet; R Mukherjee; S Uplekar; S Boy-Rottger; K-H Altmann; ST Cole, *EMBO Mol. Med.*, **2012**, 4, 1032-1042.
- [5] Y-H Yin; X Niu; B Sun; G-S Teng; Y-H Zao; C-M Wu, *Chem. Res. Chinese Universities*, **2011**, 27, 635-640.
- [6] P Cordero; J Campion; F Milagro; F Marzo; J Martinez, *Lipids Health Dis.*, **2008**, 7, 49.
- [7] JL Flynn; J Chan, *Infect Immun.*, **2011**, 69, 4195-4201.
- [8] M Gengenbacher; SP Rao; K Pethe; T Dick, *Microbiology*, **2010**, 156, 81-87.
- [9] I Smith, *Clin. Microbiol. Rev.*, **2003**, 16, 463-496.
- [10] EJ Muñoz-Elias; JD McKinney, *Nat. Med.*, **2005**, 11, 638-644.
- [11] JM Li; N Li; D-Y Zhu; L-G Wan; Y-L He; C Yang, *Chinese Med. J.*, **2008**, 121, 1114-1119.
- [12] M Kratky; J Vinsova; E Novotna; J Mandikova; V Wsol; F Trejtnar; V Ulmann; J Stolarikova; S Fernandes; S Bhat; J Liu, *Tuberculosis-Edinburgh, Elsevier Science B.V., Amsterdam*, **2012**, 92, 5, 434-439.
- [13] KR Valasani; JR Vangavaragu; VW Day; SS Yan, *J. Chem. Inf. Model.*, **2014**, 54, 902-912.
- [14] A Vuorinen; R Engeli; A Meyer; F Bachmann; UJ Griesser; D Schuster; A Odermatt, *J. Med. Chem.*, **2014**, 57, 5995-6007.
- [15] A Speck-Planche; VV Kleandrova; F Luan; ND Cordeiro, *Comput. Chem. High Throughput Screening*, **2012**, 15, 666-673.
- [16] Accelrys Software Inc. Discovery Studio Modeling Environment, Release 2.5.5, San Diego, CA: Accelrys Software Inc., **2013**.
- [17] V Sharma; S Sharma; K Hoener zu Bentrup; JD McKinney; DG Russell; JWR Jacobs; JC Sacchettini, *Nature America*, **2000**, 7, 663-668.
- [18] R Dias; WF de Azevedo Jr, *Curr. Drug Targets*, **2008**, 9, 1040-1047.
- [19] R Griffith; T Luu; J Garner; PA Keller, *J. Mol. Graph. Model.*, **2005**, 23, 439-446.
- [20] ML Peach; MC Nicklaus, *J. Cheminformatics*, **2009**, 1, 6.
- [21] KJ Simmons; I Chopra; CW Fishwick, *Nature Rev.*, **2010**, 8, 501-510.