Identification of Inhibitors of Fatty Acid Synthesis Enzymes in Mycobacterium Tuberculosis

Alexander Priest[1], E. Davis Oldham, Lynn Lewis
University of Mary Washington
1301 College Avenue
Fredericksburg, VA 22401
apriest,eoldham,llewis@umw.edu

David Toth[2]
Centre College
600 West Walnut Street
Danville, KY 40422
david.toth@centre.edu

ABSTRACT

Antibiotic-resistant strains of *Mycobacterium tuberculosis* have rendered some of the current treatments for tuberculosis ineffective, creating a need for new treatments. Today, the most efficient way to find new drugs to treat tuberculosis and other diseases is to use virtual screening to quickly consider millions of potential drug candidates and filter out all but the ones most likely to inhibit the disease. These top hits can then be tested in a traditional wet lab to determine their potential effectiveness. Using supercomputers, we screened over 4 million potential drug molecules against each of two enzymes that are critical to the survival of *Mycobacterium tuberculosis*. During this process, we determined the top candidate molecules to test in the wet lab.

Categories and Subject Descriptors

J.3 [**Life and Medical Sciences**]: Biology and genetics.

General Terms

Experimentation.

Keywords

Computational science education, drug discovery, virtual screening, parallel computing education.

1. INTRODUCTION

Since their discovery in the early 20th century, antibiotics have seen exponential growth in usage due to their unparalleled efficacy for the treatment of bacterial diseases [1]. Unfortunately, because this method of treatment is relatively new, we are only just now observing the ramifications of their ubiquity; widespread use and misuse of antibiotics has become a force of natural selection for bacteria, and as a result, these pathogens are evolving to resist them [2]. Antibiotic-resistant strains of many disease-causing bacteria have been observed, and among these is the causative agent of tuberculosis, *Mycobacterium tuberculosis* [2, 3]. Tuberculosis affects millions of people worldwide to this day, and a variety of reasons that have contributed to resistant strains of the disease have resulted in a critical need to search for novel drug treatments [3, 4]. In the past, this has been accomplished by taking soil samples and plating them to look for naturally occurring antibiotic producers, but as this research has gone on, it is more difficult to find novel antibiotic producers [5]. With this in mind, it is easy to see a need for new methodologies to come into

play. In the age of technology, there has been an increase in the use of computers to ease research processes like this. For example, there are several molecular docking programs which exist now that are designed to simulate the binding interactions of molecules with protein targets, including AutoDock Vina, DOCK, GOLD, and Glide [6, 7, 8, 9]. Screening molecules with a molecular docking program is much faster and more convenient than testing for inhibitors with *in vitro* methods. We can use this technology to investigate novel mechanisms for antibacterial compounds. Rather than waiting a week or more for a panel of bacterial plates to respond to exposure to potential drug candidates, these programs can give us an idea of how strong the interaction would be in a matter of minutes of compute time per compound. In this study, we used an *in silico* virtual drug screening process to comb through approximately 4.2 million ligands as potential drugs to target a critical enzyme in *M. tuberculosis*. To deal with the logistical issues of the sheer compute time this required, we decided to run the virtual screen on a supercomputer capable of running thousands of simulations at the same time, achieving a throughput unmatched by any *in vitro* assay method. However, while the results of a virtual screen indicate which molecules are likely to bind to a target protein, it does not necessarily mean the molecules will actually bind to the protein and even more importantly, inhibit the protein [10]. Because of this shortcoming, the virtual screening process is used as a first phase in the drug discovery process, filtering out the vast majority of molecules which likely will not bind to the protein [11]. After the virtual screening is completed, the top hits are screened with biological assays to test which molecules will actually work as treatments [11].

2. RELATED WORK

Using virtual screening to narrow down the list of compounds to test in a wet lab with biological assays has become accepted over the last number of years, and people from various research groups are using this method [12]. The corresponding author has worked with teams using virtual screening on several projects [13]. In one such study, the target was an essential enzyme found in *Plasmodium* sp., the causative agent of malaria [14]. The open-source docking simulation program AutoDock Vina, designed at the Scripps Research Institute, was used to screen the full_nci_ALL_TAUTOMERS_2011 library of about 320,000 chemical compounds from the ZINC database against the enzyme PfUCHL3 [6, 15]. The top scoring compounds were then re-screened against the human analog for this enzyme to determine which would be safest for human use; these were then screened in vitro in the lab to confirm their efficacy against *Plasmodium*. As a

---

[1] Undergraduate Student

[2] Corresponding Author

result, the authors determined two compounds with very high promise as novel malaria treatments which would be effective without causing side effects due to binding with the human orthologous protein. Franco et al. also screened DrpE1 in an attempt to find a novel cure for tuberculosis [16].

3. METHODS

The first step in our drug discovery process was to pick a target. We first decided on targeting *M. tuberculosis* because of its recent trend in antibacterial resistance; novel drugs for tuberculosis would be especially sought after as a result [3]. We chose two enzymes (called Target 1 and Target 2 in this paper) that are critical to the survival of *Mycobacterium tuberculosis*, and we understand the mechanisms of their action. Next, we found the structures for these proteins from the Protein Data Bank (PDB) and prepared them for docking using AutoDock Tools [17, 18]. The structures were derived from X-ray diffraction, and were specific to *Mycobacterium tuberculosis* [17]. Of the different structures available, we selected the wild-type structures including a ligand in the binding pocket. We also used AutoDock Tools to locate the coordinates of the binding site and noted these down. We then uploaded the prepared molecules and coordinates to the Blue Waters and Stampede supercomputers, along with the molecular docking program AutoDock Vina. We obtained about 4.2 million ligand files from the ZINC database and downloaded these as well, and created shell scripts to break the work into pieces. We created another shell script to run the program for each compound, and another to collect and package the results for download and analysis. Once we downloaded the data, we uploaded it to an SQL database and searched for the top hits.

4. RESULTS

The results of the virtual screens were grouped into bins based on the binding affinities of the compounds. These bins allow us to separate the most promising compounds from the rest and determine which compounds should be tested with assays. Table 1 and Table 2 show the number of compounds in each binding affinity range. For Target 1, 4,182,163 compounds were screened and for Target 2, 4,182,137 compounds were screened. Figure 1 and Figure 2 show the binding affinities for the compounds screened against Target 1 and Target 2, respectively. Figure 3 and Figure 4 break down the top hits for Target 1 and Target 2 into bins of narrower width. It is important to note that the best binding energies are the ones with the most negative values, so a compound with a binding energy of -13 is more likely to bind to the target than a compound with a binding energy of -12. The top hits for Targets 1 and 2 are given in Table 3 and Table 4.

5. CONCLUSIONS

Using the Blue Waters and Stampede supercomputers, we have screened over 4.1 million compounds against two enzymes that are critical to *Mycobacterium tuberculosis* surviving. The virtual screens have indicated 12 compounds with a binding affinity of < -13 that are likely to bind to *Mycobacterium tuberculosis*. If those compounds can indeed bind to the target enzymes in tuberculosis and inhibit the functioning of those enzymes, then the compounds may be useful in treating tuberculosis.

6. FUTURE WORK

For future work, we will test as many of the top hits as we can in the wet lab. The compounds that scored in the -13.0 to -13.9 range will be prioritized. High scoring compounds with different structures will also be prioritized to give a wide range of coverage of different types of compounds. We note that an entity with

Table 1 - Summary of Binding Affinities of Virtual Screen against Target 1

Binding Affinity Range (kcal/mol)	Number of Compounds in Range
-13 ≥ x > -14	8
-12 ≥ x > -13	139
-11 ≥ x > -12	3,576
-10 ≥ x > -11	55,866
-9 ≥ x > -10	413,115
-8 ≥ x > -9	1,377,570
-7 ≥ x > -8	1,607,582
-6 ≥ x > -7	606,245
-5 ≥ x > -6	94,722
-4 ≥ x > -5	20,151
-3 ≥ x > -4	3,082
-2 ≥ x > -3	94
-1 ≥ x > -2	6
0 ≥ x > -1	2
x > 0	5

Table 2 - Summary of Binding Affinities of Virtual Screen against Target 2

Binding Affinity Range (kcal/mol)	Number of Compounds in Range
-13 ≥ x > -14	4
-12 ≥ x > -13	91
-11 ≥ x > -12	3,756
-10 ≥ x > -11	71,393
-9 ≥ x > -10	571,938
-8 ≥ x > -9	1,453,342
-7 ≥ x > -8	1,101,984
-6 ≥ x > -7	443,499
-5 ≥ x > -6	192,355
-4 ≥ x > -5	108,482
-3 ≥ x > -4	68,534
-2 ≥ x > -3	47,881
-1 ≥ x > -2	33,959
0 ≥ x > -1	25,098
x > 0	59,821

Table 3 - Top Hits for Target 1 from the ZINC Database Libraries Screened

Score	Library	Folder	Compound
-13.5	full_nci_ALL_TAUTOMERS_2011	SetOf10k_0004	ZINC01588230.pdbqt
-13.3	ChemBridge_FullLibrary2011	SetOf10k_0037	ZINC02880067.pdbqt
-13.2	asinex_newMay2011_fixedForVinaInDec	SetOf10k_0000	ZINC06281466.pdbqt
-13.2	asinex_newMay2011_fixedForVinaInDec	SetOf10k_0014	ZINC13564997.pdbqt
-13.1	asinex_newMay2011_fixedForVinaInDec	SetOf10k_0019	ZINC13565797.pdbqt
-13.1	asinex_newMay2011_fixedForVinaInDec	SetOf10k_0014	ZINC13564995.pdbqt
-13.0	asinex_newMay2011_fixedForVinaInDec	SetOf10k_0014	ZINC13564992.pdbqt
-13.0	asinex_newMay2011_fixedForVinaInDec	SetOf10k_0018	ZINC13084337.pdbqt
-12.9	ChemBridge_FullLibrary2011	SetOf10k_0000	ZINC02833848.pdbqt
-12.9	asinex_newMay2011_fixedForVinaInDec	SetOf10k_0014	ZINC13565000.pdbqt
-12.9	asinex_newMay2011_fixedForVinaInDec	SetOf10k_0042	ZINC13564941.pdbqt

Table 4 - Top Hits for Target 2 from the ZINC Database Libraries Screened

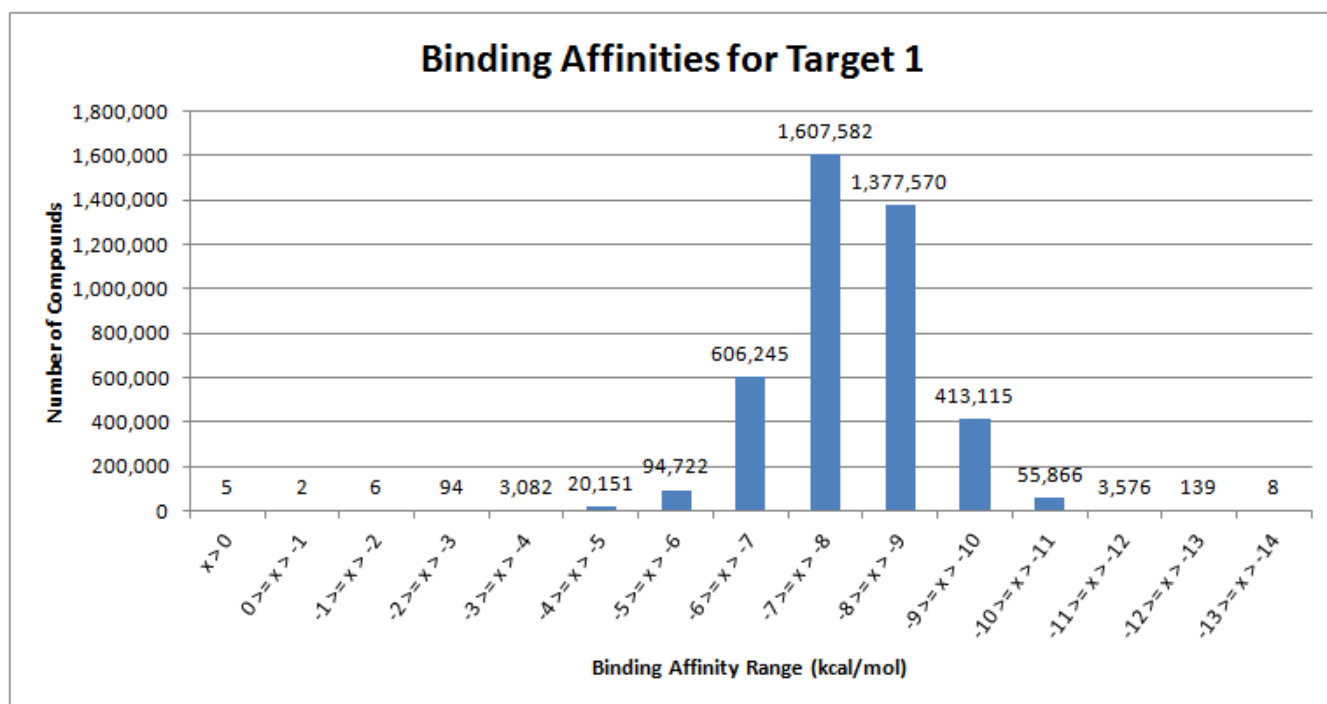| Score	Library	Folder	Compound
-13.9	full_nci_ALL_TAUTOMERS_2011	SetOf10k_0016	ZINC04824645.pdbqt
-13.1	asinex_newMay2011_fixedForVinaInDec	SetOf10k_0008	ZINC04838539.pdbqt
-13.1	ChemBridge_FullLibrary2011	SetOf10k_0069	ZINC19634897.pdbqt
-13.0	asinex_newMay2011_fixedForVinaInDec	SetOf10k_0025	ZINC06475337.pdbqt
-12.9	ChemBridge_FullLibrary2011	SetOf10k_0007	ZINC04980431.pdbqt
-12.8	full_nci_ALL_TAUTOMERS_2011	SetOf10k_0015	ZINC04428442.pdbqt
-12.7	ChemBridge_FullLibrary2011	SetOf10k_0087	ZINC16662786.pdbqt
-12.7	ChemBridge_FullLibrary2011	SetOf10k_0029	ZINC02893797.pdbqt
-12.6	ChemBridge_FullLibrary2011	SetOf10k_0074	ZINC19634255.pdbqt
-12.6	ChemBridge_FullLibrary2011	SetOf10k_0049	ZINC19632616.pdbqt
-12.6	ChemBridge_FullLibrary2011	SetOf10k_0074	ZINC23281397.pdbqt

Binding Affinities for Target 1

Number of Compounds (y-axis)

Binding Affinity Range	Value
x > 0	5
0 >= x > -1	2
-1 >= x > -2	6
-2 >= x > -3	94
-3 >= x > -4	3,082
-4 >= x > -5	20,151
-5 >= x > -6	94,722
-6 >= x > -7	606,245
-7 >= x > -8	1,607,582
-8 >= x > -9	1,377,570
-9 >= x > -10	413,115
-10 >= x > -11	55,866
-11 >= x > -12	3,576
-12 >= x > -13	139
-13 >= x > -14	8

Binding Affinity Range (kcal/mol) (x-axis)

Figure 1 - Results of Virtual Screen of Compounds against Target 1 with Binding Affinities Grouped in Ranges

Binding Affinities for Target 2

Number of Compounds (y-axis)

Binding Affinity Range	Value
x > 0	59,821
0 >= x > -1	25,098
-1 >= x > -2	33,959
-2 >= x > -3	47,881
-3 >= x > -4	68,534
-4 >= x > -5	108,482
-5 >= x > -6	192,355
-6 >= x > -7	443,499
-7 >= x > -8	1,101,984
-8 >= x > -9	1,453,342
-9 >= x > -10	571,938
-10 >= x > -11	71,393
-11 >= x > -12	3,756
-12 >= x > -13	91
-13 >= x > -14	4

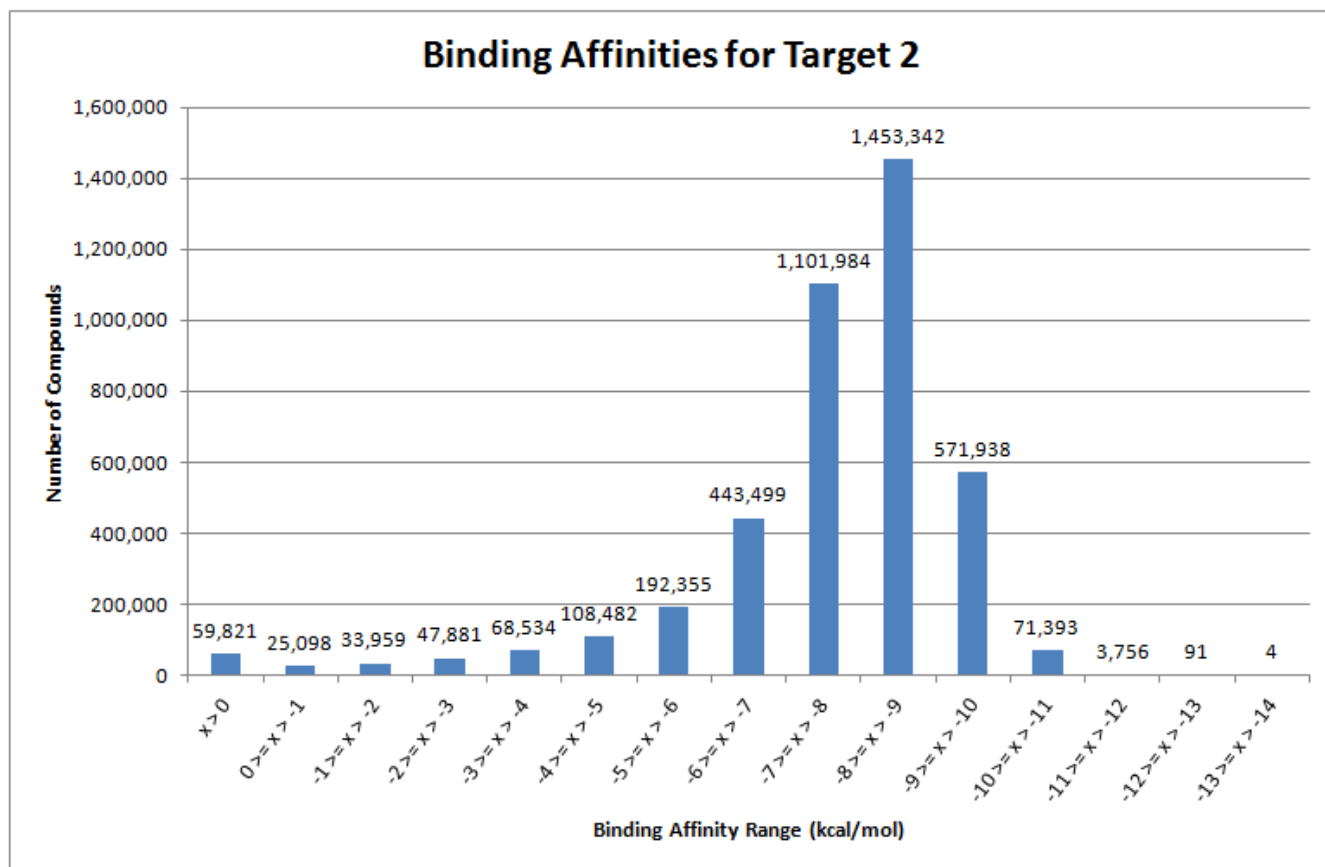**Binding Affinity Range (kcal/mol)** (x-axis)

Figure 2 - Results of Virtual Screen of Compounds against Target 2 with Binding Affinities Grouped in Ranges
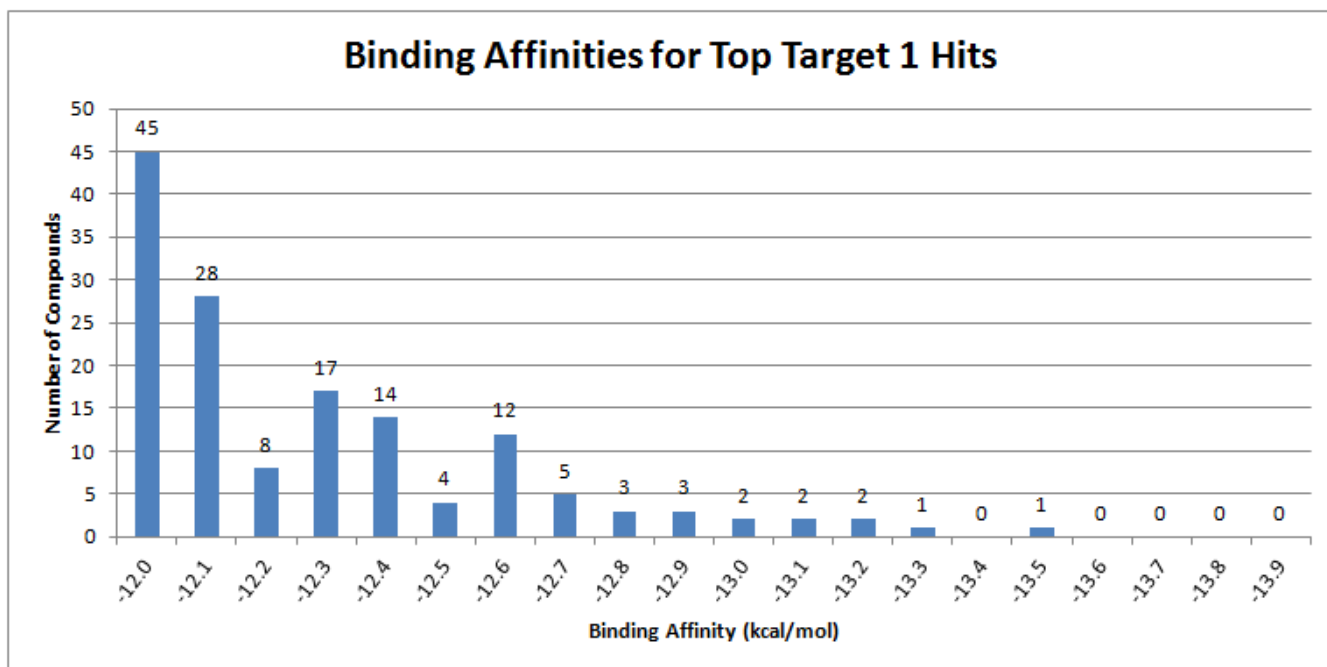
Figure 3 - The Top Binding Affinities for Virtual Screen of Compounds against Target 1 Grouped in Ranges
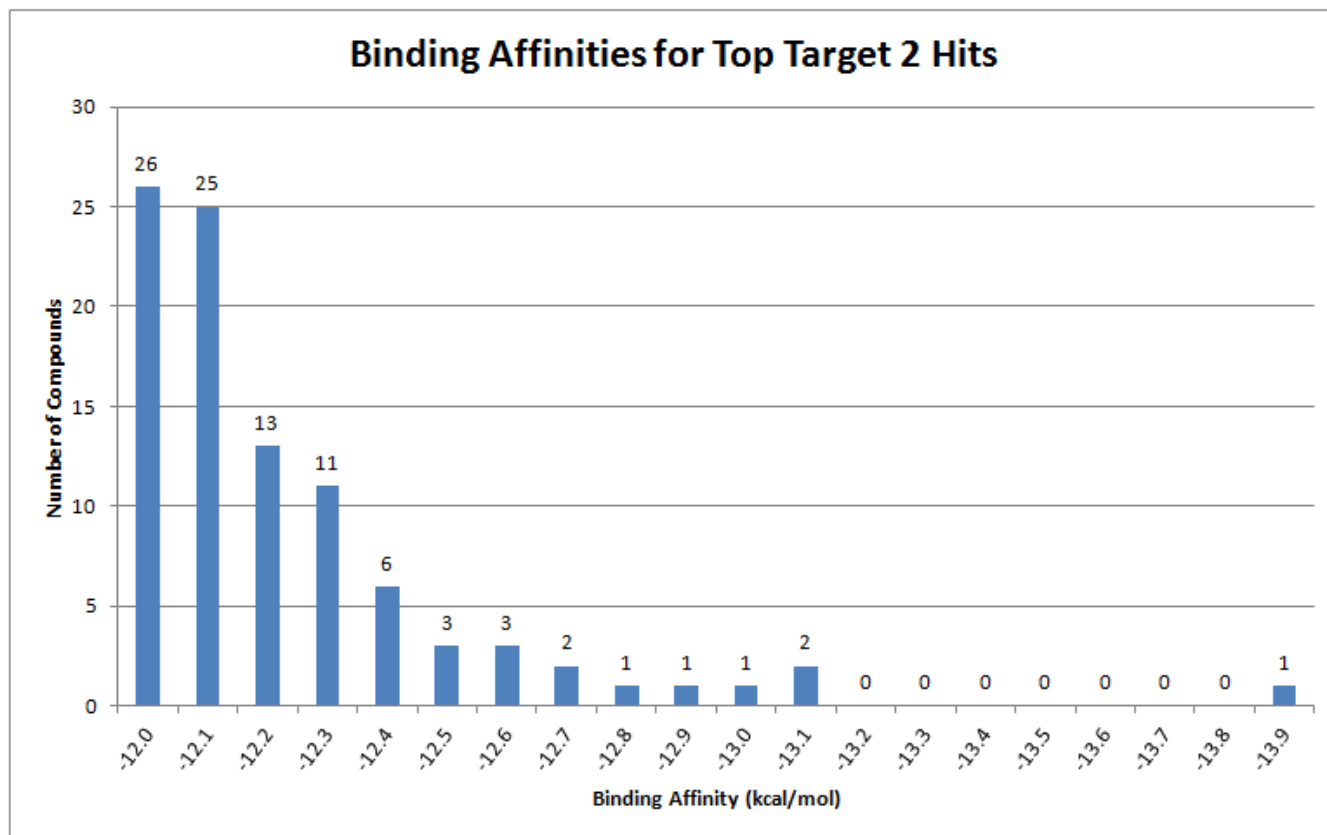

Figure 4 - The Top Binding Affinities for Virtual Screen of Compounds against Target 2 Grouped in Ranges

significant financial resources might test all the compounds scoring -12 or better.

We will apply the top scoring compounds to cultures of a model organism, *Mycobacterium bovis*, and determine if any of these compounds has an inhibitory effect on the growth of the bacteria [19]. We will use *Mycobacterium bovis* because unlike *Mycobacterium tuberculosis*, *Mycobacterium bovis* is not pathogenic [19]. Should one or more compounds prove effective at inhibiting the bacterial growth, the next step requires the resources of a larger organization. Further testing of the successful compounds would be necessary to confirm their action on the cells. Following this, research would be done to determine whether the targeted proteins have any human orthologs, or similar proteins which occur in the human body which may also be affected by the compound, resulting in unwanted side effects. *In vivo* testing with a model host would be the next step, as the compound would need to be proven safe for the consumption of the host organism.

7. REFLECTIONS

The project described in this paper was the first author's Blue Waters Student Internship project where he learned to incorporate computation and high-performance computing into his research. This section details the first author's reflections about his internship and the impact that it has had on his current and future academic endeavors: When I took my first course in computer science, I did not anticipate that it would give me the power to make a difference like this. At the time of beginning this project, I was a Biology major, with minors in Chemistry and Neuroscience. I could tell you a lot about how diseases like tuberculosis can ravage the human body. I could tell you how the increase in the prevalence of antibacterial soaps may have actually led to the rise of hyper-resistant superbugs. However, I could never have explained to you any way in which I could make a difference as an undergraduate student in any of these areas. Before I got involved with computer science, university was simply a place for learning, not for doing. I started with a single course on modeling and simulation which required no formal coding skill (we used drag-and-drop programming environments like Scratch), and grew into learning the basics of C++ in a week before attending the 2-week intensive high performance computing workshop for Blue Waters interns. At the workshop, I learned parts of the C and FORTRAN programming languages in order to learn the basics of the parallel computing libraries OpenMP, CUDA, MPI, and OpenACC. Having only taken a single introductory course in computer science before attending the workshop, I am proud of how much I was able to learn. Now, I am confident using a Linux command prompt and I can write some basic shell scripts. Having learned these skills, I am capable of using supercomputers for my research, which spans biology and chemistry. One of the most lasting impacts that this incredible experience has left me with, however, is my recent decision to stay an extra year at UMW in order to pursue a double major in Computer Science alongside my Biology major, and to add a Data Science minor. I am planning on finding a graduate school that will have the same zeal for interdisciplinary projects that I have now, and I believe that these experiences will make me very competitive in the application process. Having this Blue Waters Internship was genuinely a turning point in my college career and my life in general, and having this opportunity to do real research with real-world implications is a once-in-a-lifetime experience. This project represents the mixing of the disciplines that needs to happen if science as a whole is going to make new

discoveries this century to rival the marvels of the past. Computer modeling for scientific applications is certainly the way research will be conducted in the future, and the future is not so far away after all.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] Davies, Julian, and Davies, Dorothy. Origins and Evolution of Antibiotic Resistance. *Microbiol. Mol. Biol. Rev.* September 2010 vol. 74 no. 3 417-433. doi: 10.1128/MMBR.00016-10.

[2] Alanis, Alfonso. Resistance to Antibiotics: Are We in the Post-Antibiotic Era?, Archives of Medical Research, 36, (Nov.-Dec., 2005), 697-705, doi:10.1016/j.arcmed.2005.06.009.

[3] Rivers, Emma C. and Mancera, Ricardo L. New anti-tuberculosis drugs in clinical trials with novel mechanisms of action, Drug Discovery Today, 13, (Dec. 2008), 1090-1098, doi:10.1016/j.drudis.2008.09.004.

[4] World Health Organization, http://www.who.int/mediacentre/factsheets/fs104/en/.

[5] Ling et al. A new antibiotic kills pathogens without detectable resistance, *Nature* 517, 455–459 (Jan. 2015) doi:10.1038/nature14098.

[6] Trott, O. and Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *Computational Chemistry*, 31, (Jan. 2010), 455-461, DOI: 10.1002/jcc.21334.

[7] P. T. Lang, S. R. Brozell, S. Mukherjee, E. F. Petterson, E. C. Meng, V. Thomas, R. C. Rizzo, D. A. Case, T. L. James, and I. D. Kuntz, DOCK 6: Combining techniques to model RNA–small molecule complexes. RNA 2009, 15, 1219-1230.

[8] Verdonk ML, Cole JC, Hartshorn MJ, Murray CW. et al. Improved protein-ligand docking using GOLD. Proteins. 2003;52:609–623. doi: 10.1002/prot.10465.

[9] Friesner RA, Banks JL, Murphy RB, Halgren TA. et al. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. J Med Chem. 2004;47:1739–1749. doi: 10.1021/jm0306430.

[10] Shoichet, Brian. Virtual screening of chemical libraries. *Nature*. 2004 Dec 16; 432(7019): 862–865.

doi: 10.1038/nature03197

[11] Cheng, Tiejung, Li, Qingliang, Zhou, Zhigang, Wang, Yanli, and Bryant, Stephen H. Structure-Based Virtual Screening for Drug Discovery: a Problem-Centric Review. *AAPS J.* 2012 Mar; 14(1): 133–141. doi: 10.1208/s12248-012-9322-0

[12] Ellingson, Sally R., Smith, Jeremy C., and Baudry, Jerome. 2014. Polypharmacology and supercomputer-based docking: opportunities and challenges, *Molecular Simulation*, DOI: http://dx.doi.org/10.1080/08927022.2014.899699.

[13] Toth, David, Franco, Jimmy, and Berkes, Charlotte. Attacking HIV, Tuberculosis and Histoplasmosis with XSEDE Resources. In *Proceedings of the Conference on Extreme Science and Engineering Discovery Environment: Gateway to Discovery* (San Diego, CA, USA), July 22-25, 2013). ACM, New York, NY. DOI: http://dx.doi.org/10.1145/2484762.2484766.

[14] Franco, Jimmy, Blackie, Margaret A.L., Toth, David, Smith, Peter J., Capuano, Joseph, Fastnacht, Kurt, Berkes, Charlotte. A structural comparative approach to identifying novel antimalarial inhibitors. *Computational Biology and Chemistry*, 45, (Aug. 2013), 42-47. DOI: http://dx.doi.org/10.1016/j.compbiolchem.2013.04.002.

[15] Irwin, Sterling, Mysinger, Bolstad and Coleman, ZINC – A Free Database of Commercially Available Compounds for Virtual Screening. *Chem. Inf. Model.* 2012 DOI: 10.1021/ci3001277

[16] Wilsey, Claire, Gurka, Jessica, Toth, David, and Franco, Jimmy. A large scale virtual screen of DprE1. *Computational Biology and Chemistry*, 47, (Dec. 2013), 121-125, DOI: http://dx.doi.org/10.1016/j.compbiolchem.2013.08.006.

[17] RCSB Protein Data Bank. http://pdb.org/pdb/home/home.do.

[18] Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S. and Olson, A. J. Autodock4 and AutoDockTools4: automated docking with selective receptor flexiblity. *Computational Chemistry* 16 (2009), 2785-91.

[19] Altaf, Mudassar, Miller, Christopher H., Bellows, David S., and O'Toole, Ronan. Evaluation of the *Mycobacterium smegmatis* and BCG models for the discovery of Mycobacterium tuberculosis inhibitors. Tuberculosis, 90, (Nov. 2010), 333-337. http://dx.doi.org/10.1016/j.tube.2010.09.002