Advancement of Chagas Disease Treatment through the Identification of Potential Natural and Synthetic Product Targets in the *Trypanosoma Cruzi* Proteome

Jenny Jia Xin Liang
University of California, San Diego, Dr. Philip Bourne
Universiti Sains Malaysia, Dr. Habibah Wahab

Abstract

The use of a computational approach to drug discovery has paved way towards polypharmalogical advancement and the improvement of existing drugs. The methods of docking, virtual screening, and active site similarity search were used to identify possible natural and synthetic compounds from the NADI (Nature-Based Drug Discovery Resource) and the NCI Diversity Set II databases that may exhibit favorable characteristics in binding to the selected Trypanosoma Cruzi (*T. Cruzi)* proteins, Trans-sialidase and gp63. A binding pocket similarity search across the *T. Cruzi* proteome for the two selected proteins along with the results of virtual screening highlights the potential off target bindings of these compounds to similar proteins within the proteome. The data gathered provides a useful prediction of protein ligand interactions as well as the specificity of these compounds to targeted proteins.

Introduction

Chagas, American Trypanosomiasis, is a tropical disease caused by *Trypanosoma Cruzi*, a flagellated protozoan parasite spread via contact with triatomine insects' feces and contaminated bodily fluids. Estimated to have infected eight to ten million people, *T. Cruzi* is mainly prevalent in the rural regions of Latin and Central Americas.

In the acute phase of infection, parasites circulate in the host displaying mild symptoms. However in rare cases, inflammation of the brain or heart may occur. Following, the host

experiences an asymptomatic period with few to none parasites in the bloodstream. While some remain asymptomatic for life, others in the chronic phase develop heart and digestive abnormalities that may result in death.

T. Cruzi, the causative agent of Chagas, goes through a number of developmental stages in the host. Triatomine insects as known as the "kissing bug" become infected upon ingestion of the trypomastigote form of *T. Cruzi* and develop into noninvasive epimastigotes in the midgut. They develop into metacyclic trypomastigotes in the lower part of the digestive tract and are excreted through feces. Humans can become infected if trypomastigotes enter an open wound and penetrate surrounding cells transforming into amastigotes. These intracellular amastigotes develop into trypomastigotes and break out of the cell into the bloodstream where it can acquire sialic acid from host sialoglycoconjugates and invade essentially any host cell. The importance of the proteins associated with this life cycle is essential in understanding their role in cell infection and evasion from the host immune system (Centers for Disease Control and Prevention).

Two proteins that are of interest are *Trypanosoma cruzi*'s trans-sialidase (TcTS) and gp63. TcTS is glycosylphosphatidylinositol (GPI)-anchored surface enzyme that is expressed during the infective metacyclic trypomastigote stage. Upon entering the bloodstream, the parasite utilizes TcTS to catalyze a transglycosidase reaction that transfers host sialic acid onto parasite glycoconjugates (Weinkopff).This allows for evasion from the host immune system and penetration into the host cell which is crucial in its survival. The protein's role in cell to cell and cell to extracellular matrix interactions makes it an essential aspect of the life cycle. Sialic acid binding to TcTS donor substrate site triggers a conformational change that in turn modulates the affinity of the acceptor site for lactose. Under these conditions, the efficiency for

tranglycosylation is drastically increased. Important residues involved in the donor active site include Tyr342, Tyr119, Arg35, Arg245, Arg314, Arg59, Arg96 and Glu230 (Buchini et al., 2008).

Similarly, *T. Cruzi* gp63 (Tcgp63) is a surface GPI-anchored protein, but with Zn^{2+} metalloprotease activity and is predicted to have a role in cell infection. Genes of Tcgp63 show the highest homology to Leishmania guyanensis gp63. Compared to Tcgp63, research on Leishmania gp63 has been much more developed and there is significant data supporting involvement of gp63 in the establishment of infection. In Leishmania, the protein is mainly expressed in the promastigote and also in the amastigote stages. It plays a role in host-parasite interaction involving the attachment of the promastigote to host cell receptors and research suggests that the proteolytic activity of the metalloprotease is linked to the cleavage of host macromolecules for purposes of protection or nutrition whereby increasing the ability of the amastigotes to survive inside the macrophage (Cuevas, Cazzulo, and Sa´nchez, 2003). In contrast, Tcgp63 was found in all stages of the parasite and when incubated with tcgp63 antiserum, reduction in host infection was observed. Antibodies against Tcgp63 partially blocked trypomastigote infection highlighting its role in host infection (Kulkarni, Olson, Engman, and McGwire 2009). Currently, there is no evidence that can link Tcgp63 to Leishmania gp63 activity. However, Tcgp63-I group like Leishmania displays the same metalloprotease activity and is also bounded to the parasite surface via a GPI anchor. Further information on Tcgp63 is limited, but because of the close homology to Leishmania, they may have similar roles in the infection of the host cell. Important residues include His264, Glu265, His268, His334 with Glu265 being the main residue involved in catalysis (Macdonald, Morrison, and McMaster, 1995).

Currently studies on *T. Cruzi* are limited which calls for more research in this area for a better treatment to halt *T. Cruzi'*s role in the progression of Chagas disease. Nifurtimox and Benznidazole have been proven effective treatments for the acute phase of infection, but are limited in the chronic stages and display severe side effects on patients. Further research is needed to develop more effective, efficient and safe treatment for *T. Cruzi* infection. Because of the important roles of the two proteins TcTS and Tcgp63 in the parasite lifecycle, virtual screening of the NADI and NCI databases would identify compounds that may bind to the active site and possibly act as inhibitors. As of now, there are no known inhibitors for the two proteins and since natural products are highly stereodiversified and often have specific biological activity, any information on these proteins to ligand interactions would be valuable to the Chagas research community (Shingo et al., 2010).

Method

A search of the RCSB Protein Data Bank for *Trypanosoma Cruzi* Trans-sialidase and gp63 crystal structures gave PDB ID results of 1MS5, 1MS0, 1MS1,1MS3,1MS4,1MS5,1MS8, 1MS9,1S0I,1SOJ, 2AH2, 3B69, 30PZ 3PJQ for TcTS and 1LML for Leishmania gp63. Final structures were selected based on type, resolution, and location of missing residues. Further analysis of these PDB files showed that missing residues did not include those in the active site of the corresponding protein. Based on the protein mechanism, bounded/ unbounded structures and complexed inhibitors, PDB IDs were further analyzed. TcTS crystal structures in complex with sialidase inhibitor, 2, 3-dehydro-3-deoxy-N-acetylneuraminic acid (DANA), were favored since the goal of the virtual screening was to identify an inhibitor to the active site. As an inhibitor for bacterial and viral sialidases, DANA is a poor inhibitor for TcTS with low affinity to the binding pocket due to steric strains from hydrogen bonding protein- ligand interactions (Oppezzoa et al., 2002)

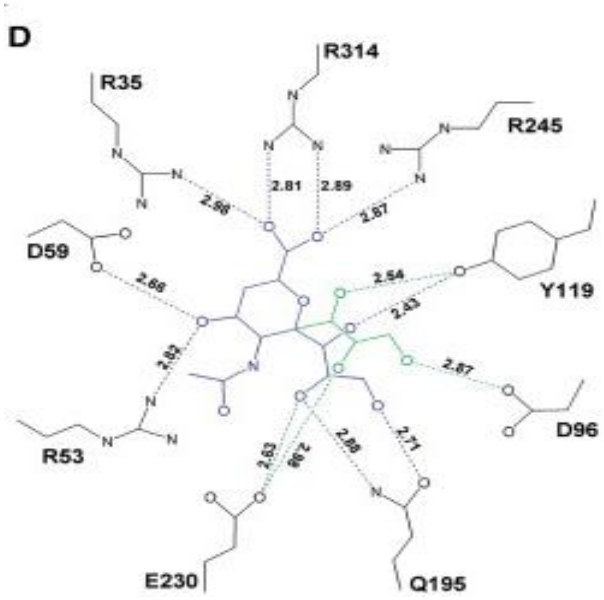


Figure 1:
2D Ligand View of hydrogen bonding interactions between important residues of 1MS8 and DANA. The same interactions are analyzed in the Autodock 3.0 virtual screening results which include Tyr342, Tyr119, Arg35, Arg245, Arg314, Arg59, Arg96 and Glu230.

Because TcTS is known to exhibit a conformational switch upon substrate binding, it was important to select bounded structures to exclude the possibility of residue interference when conducting docking and virtual screening. Bound lactose was unnecessary because without a substrate at the main active site, affinity for the sugar is lost and its binding pocket is not the target of interest. To measure the difference between bound lactose and unbound lactose structures, 1MS0 and 1MS8 were aligned using Visual Molecular Dynamics (VMD), a 3D molecular visualization program designed for displaying bimolecular systems. The resulting RMSD difference was 0.4. Due to resolution and the unbound lactose, screening results with 1MS8 would give a more accurate result. As for Tcgp63, 1LML was the only available gp63 structure of a closely related species (Schlagenhauf, Etges, and Metcalf 1035-46). The crystal structure for TcTS selected was 1MS8 obtained using X-ray diffraction with a resolution of 2.00 angstroms. For gp63, selected 1LML was obtained using X-ray diffraction with a resolution of 1.86 angstroms.

For docking and virtual screening, Autodock 3.0 a program created to predict how flexible ligands bind to designated 3D receptors, was used. A grid box of x, y, z dimensions was set to 60, 60, 60 angstroms with the default 0.375 angstroms grid point spacing at the center of the complexed ligand, DANA and Zn^{2+} binding site. As supported by literature and ProBiS, a tool used to detect protein binding sites, the active site of 1LML was close to the Zn^{2+} coordination site. Libraries of the NADI and NCI databases each consist of 3098 and 1364 compounds and all narrowed to satisfy Lipinski's Rule of Five. Ligand files for both databases were converted to PDBQ format and PDB files were edited to exclude water molecules while hydrogen coordinates were added. Because 1MS8 was composed of two identical chains, A and B, chain B was removed focusing our screening solely on chain A. In preparation of PDBQS files, partial charge and solvation parameters were set for each atom in the protein. Prior to control docking and screening, ligands DANA and Zn^{2+} were removed. A control docking was performed for 1MS8 with DANA to ensure that the ligand can dock with an acceptable RMSD score. Each compound was set to run 100 dockings and results were compiled by cluster in a histogram in output DLG files.

SMAP, a software designed for binding pocket similarity search, was used to compare TcTS and gp63 to other proteins within the *T. Cruzi* proteome. From the PDB, 141 crystal structures of the proteome were compiled and added to the library of T. Cruzi proteins via PuTTy, a SSH and telnet client. Chain A of each corresponding protein was used for the search.

For analysis of the virtual screening results, 1MS8 receptor residues Tyr342, Tyr119, Arg35, Arg245, Arg314, Arg59, Arg96, Glu230 and 1LML residues His264, Glu265, His268 and His334 where visualized along with compounds with the lowest free binding energies using VMD.

Results

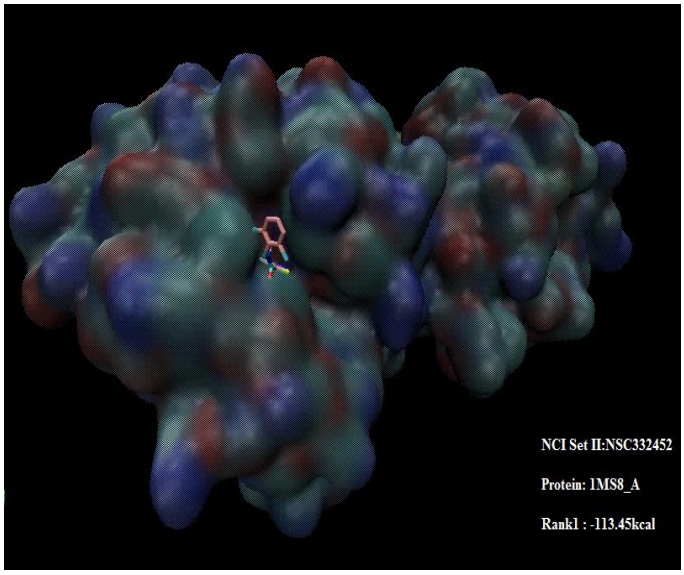
Results of the virtual screening were first ranked by free binding energy. Promising hits were chosen based on lowest free binding energy and hydrogen bonding interactions between the ligand and selected important residues. For 1MS8, ligand outputs of the virtual screening were visualized in the binding pocket and interactions with important residues were noted. These residues include Tyr119, Tyr342, Arg35, Arg245, Arg314, Arg59, Arg96 and Glu230, which are important in the catalysis and stabilization of the transition state. For 1LML, known residues surrounding the metal ion Zn included His 264, His 268, His 334 and Glu265, the residue responsible for catalytic activity were analyzed. Visualizations of binding pocket interactions with the residues are shown as followed.

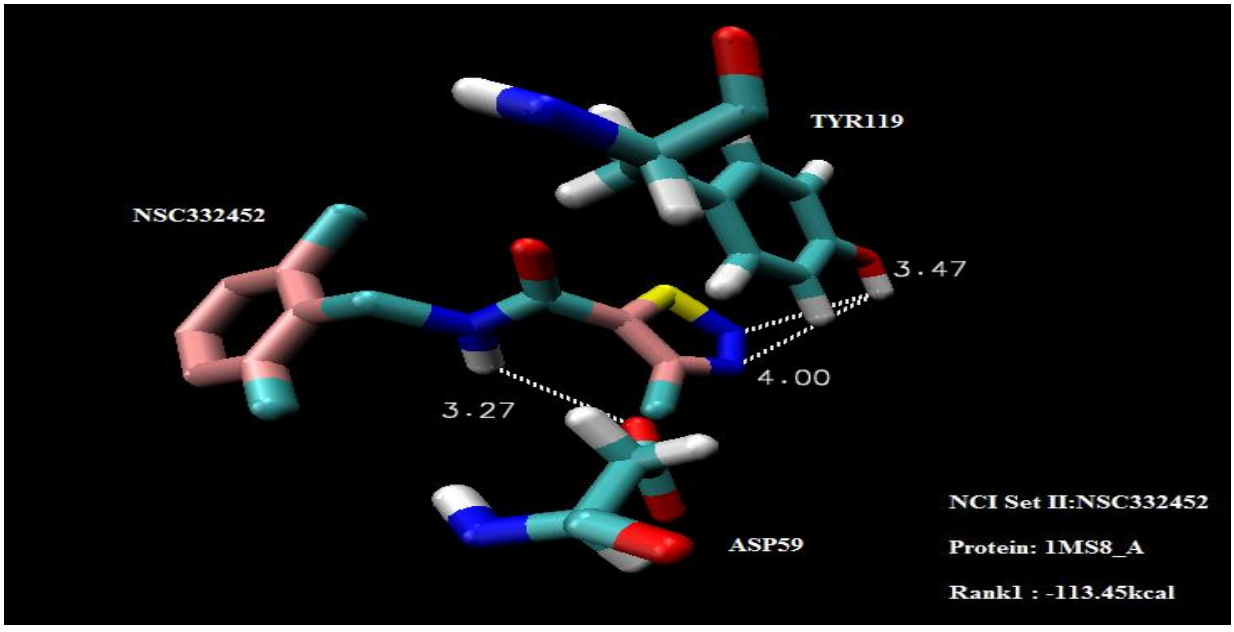
Promising Hits

1MS8

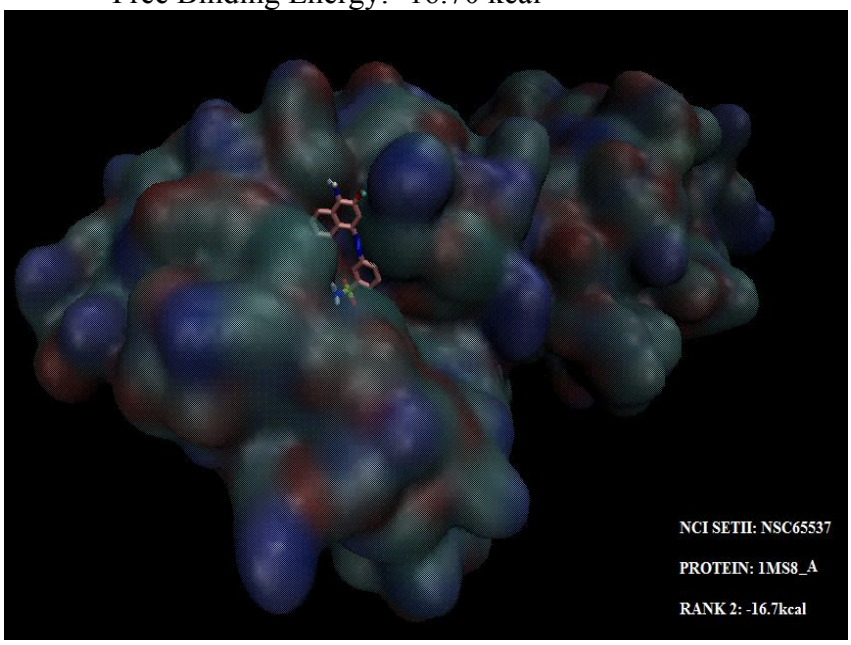
NCI Diversity Set II

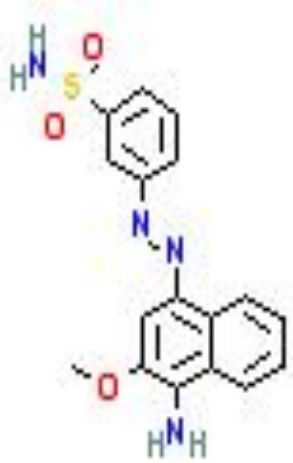
Rank 1: NSC332452 (N-(2-chloro-6-fluorobenzyl)-4-methyl-1, 2, 3-thiadiazole-5-carboxamide)
Free Binding Energy: -113.45kcal

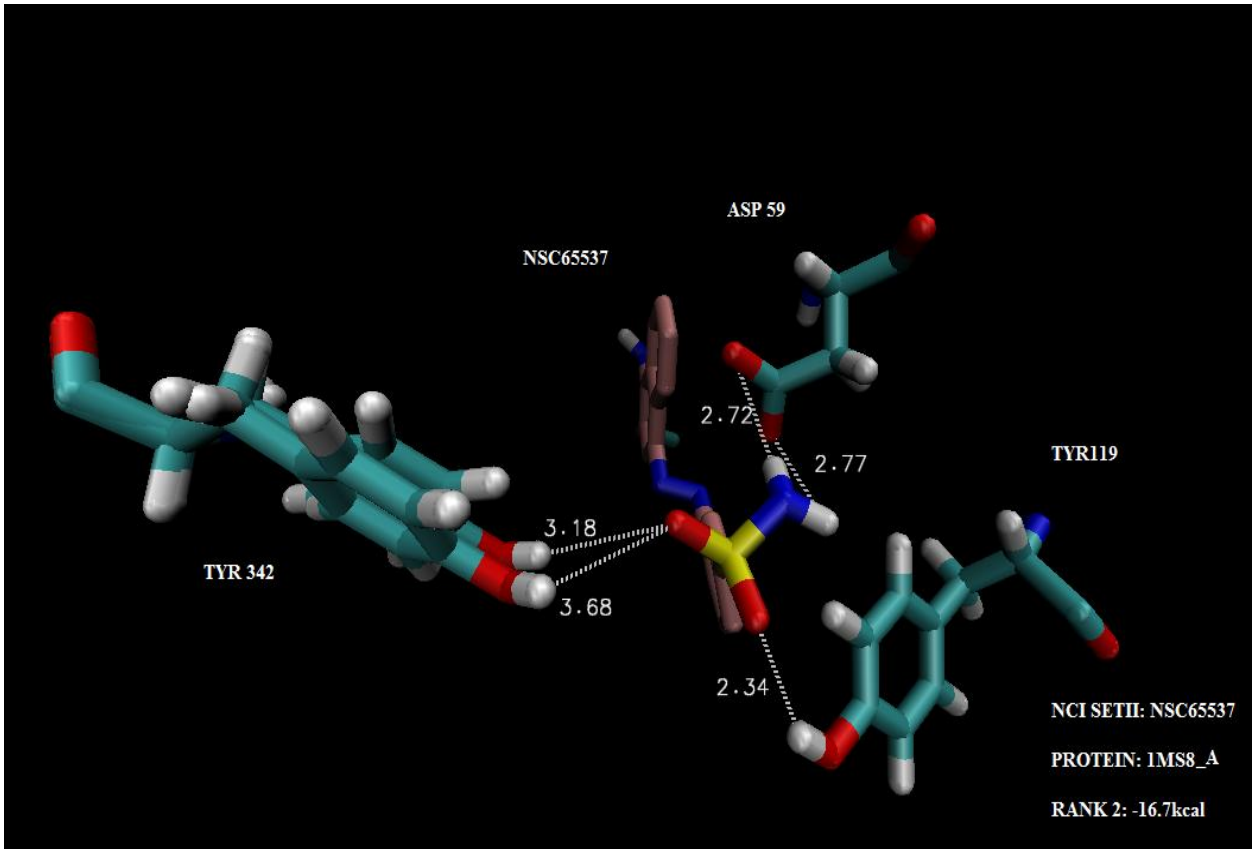




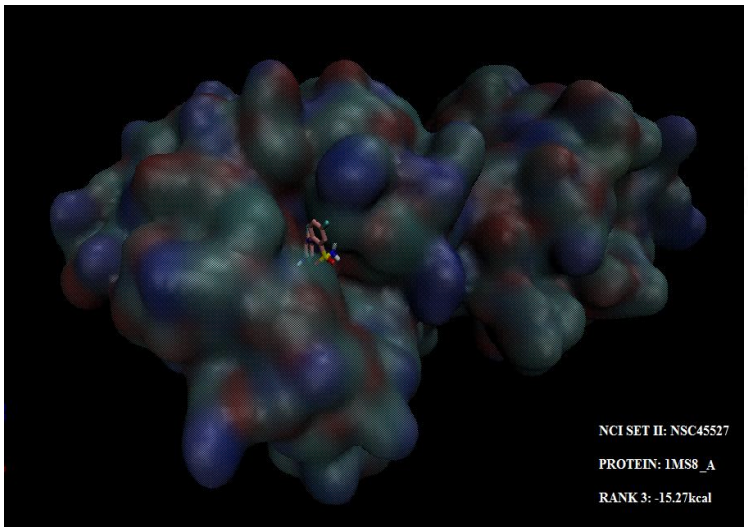
Rank 2: NSC65537 (3-((4-amino-3-methoxy-1-naphthyl)diazenyl)benzenesulfonamide)
Free Binding Energy: -16.70 kcal

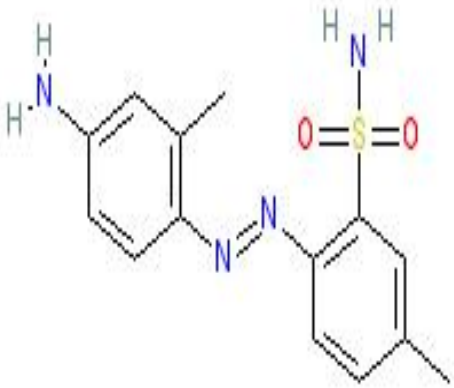


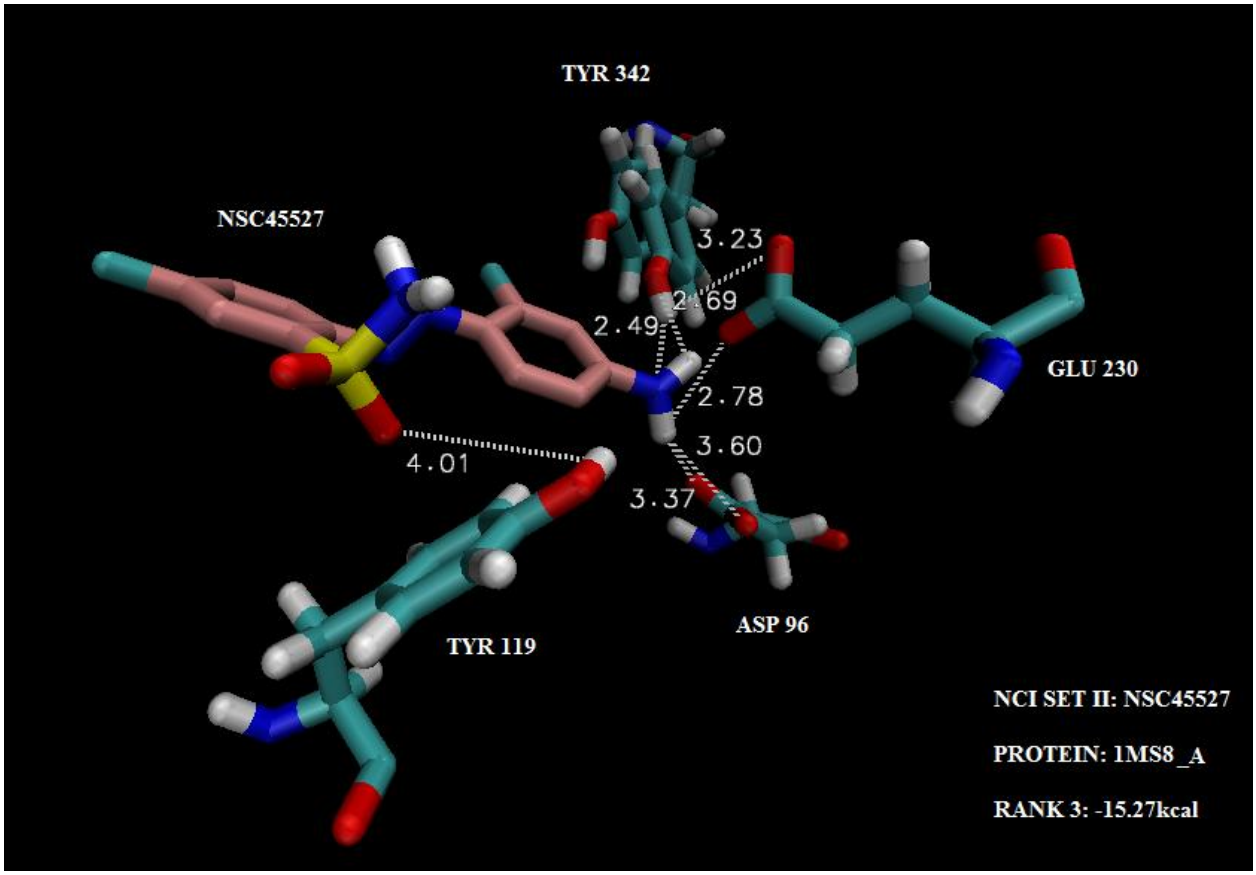




Rank 3: NSC45527 2-((4-amino-2-methylphenyl) diazenyl)-5-methylbenzenesulfonamide
Free Binding Energy: -15.27kcal

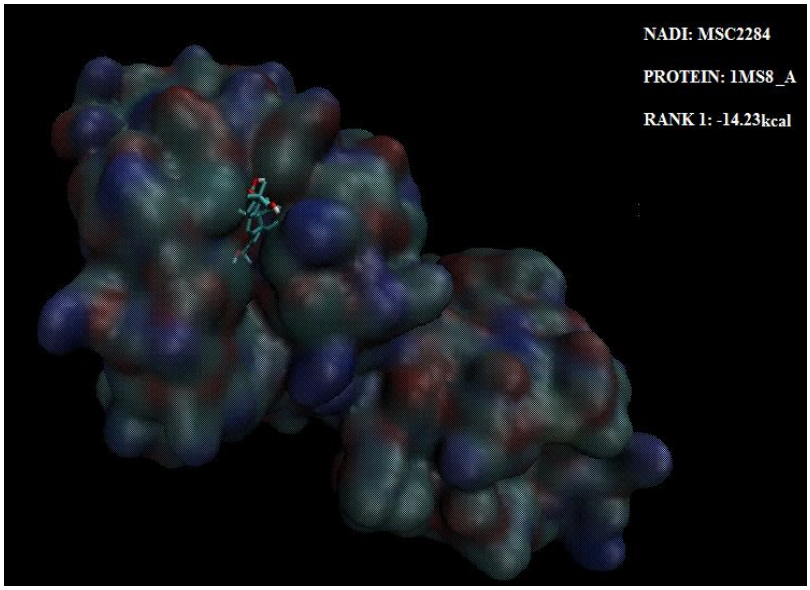


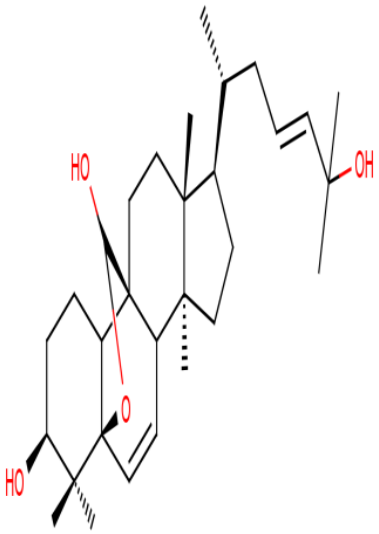


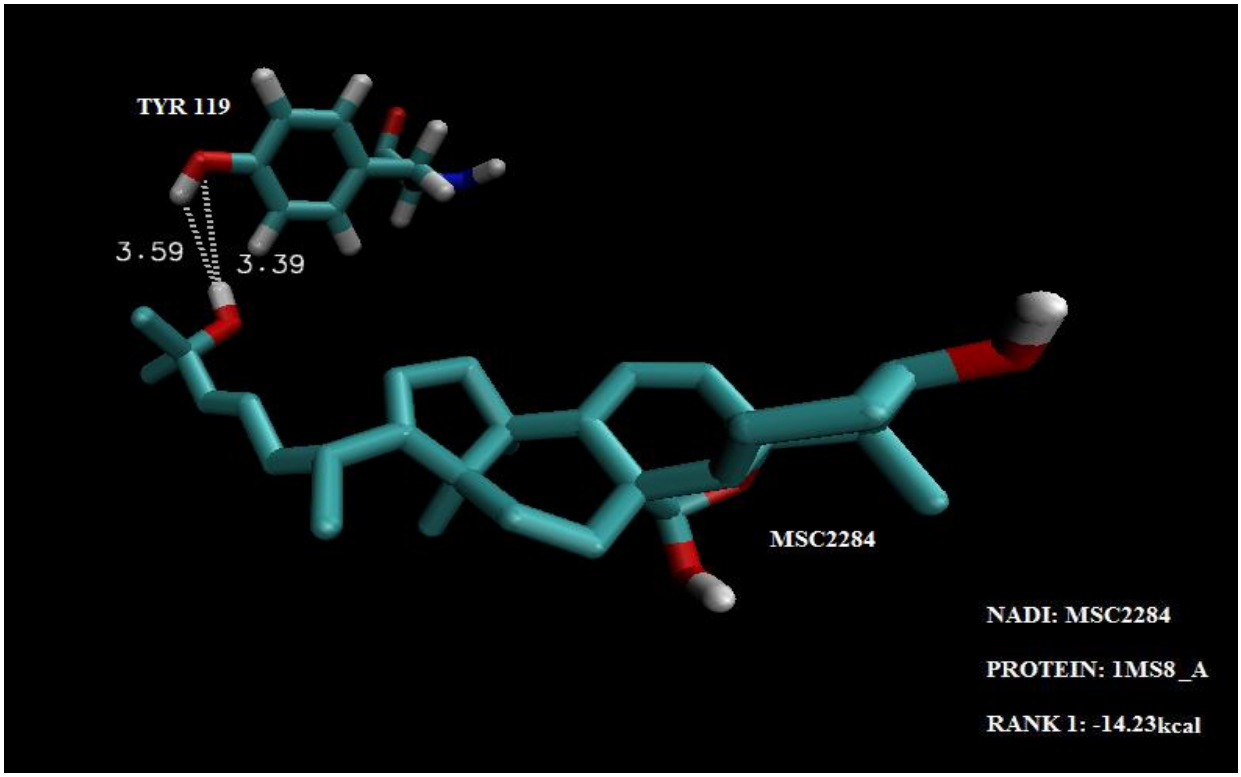


NADI

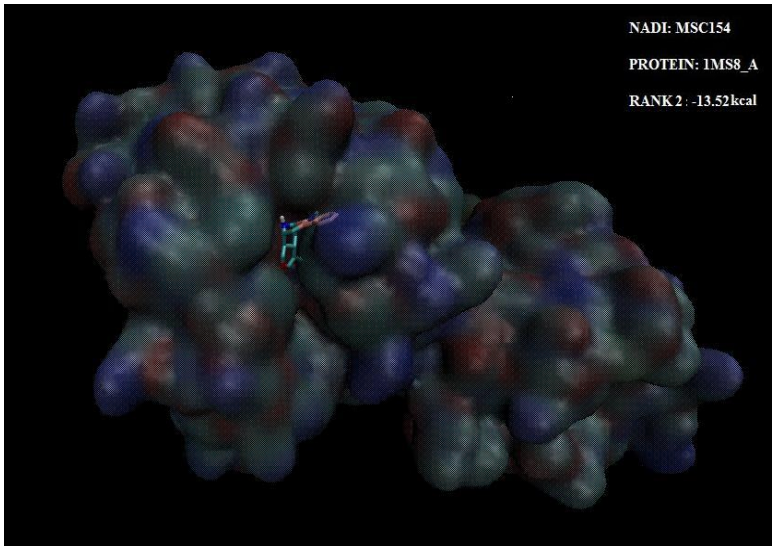
Rank 1:MSC2284, 5-beta-19-Epoxy-cucurbita-6-23(E)-diene-3-beta-19, 25-triol
Gourd of Momordica charantica
Free Binding Energy: -14.23kcal

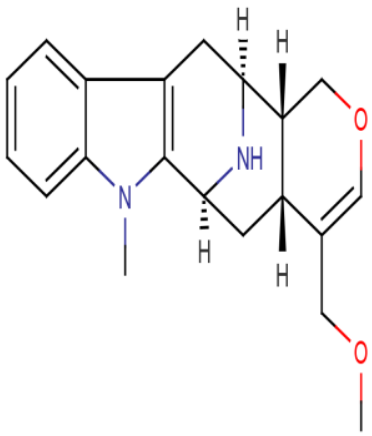


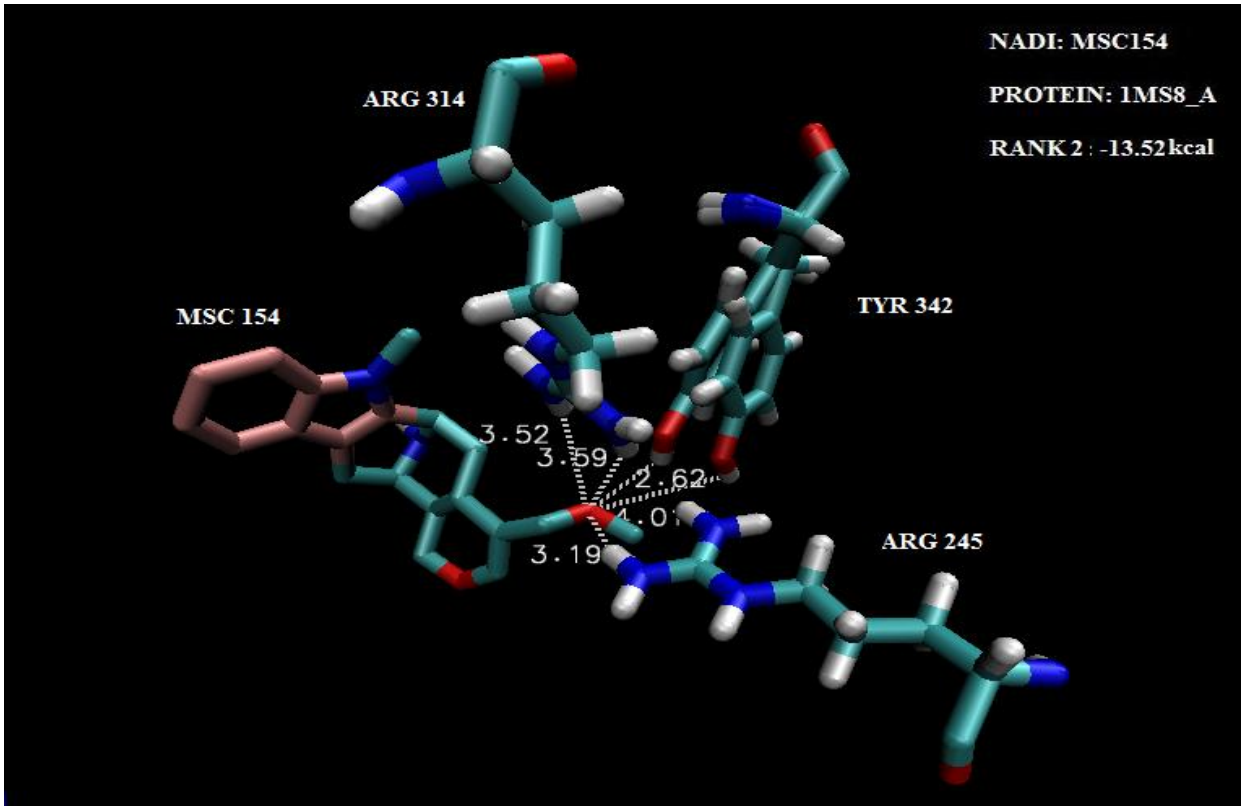




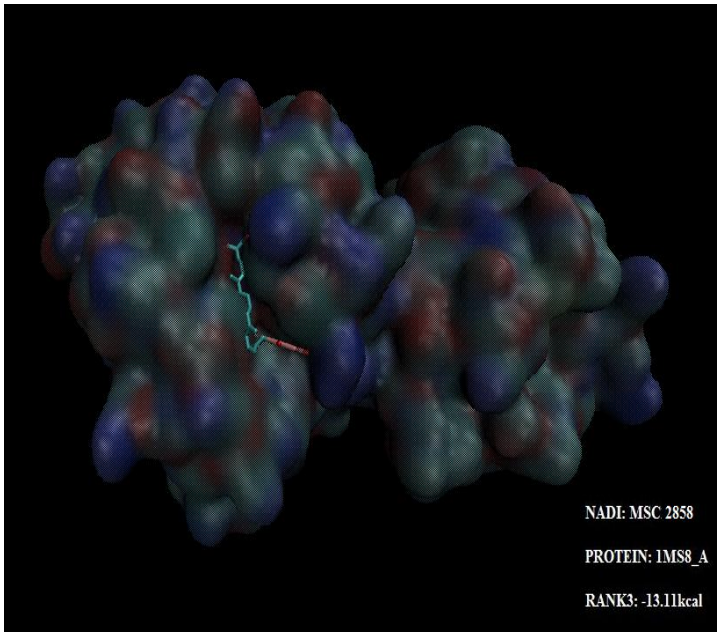
Rank 2: MSC154, N (4)-demethylalstonerine
Alstonia angustifolia leaf
Free Binding Energy:-13.52kcal

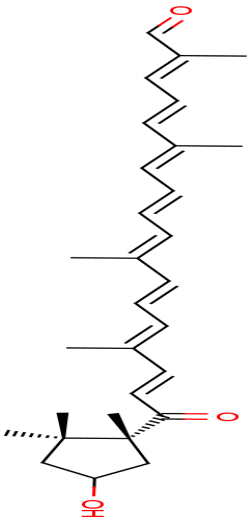


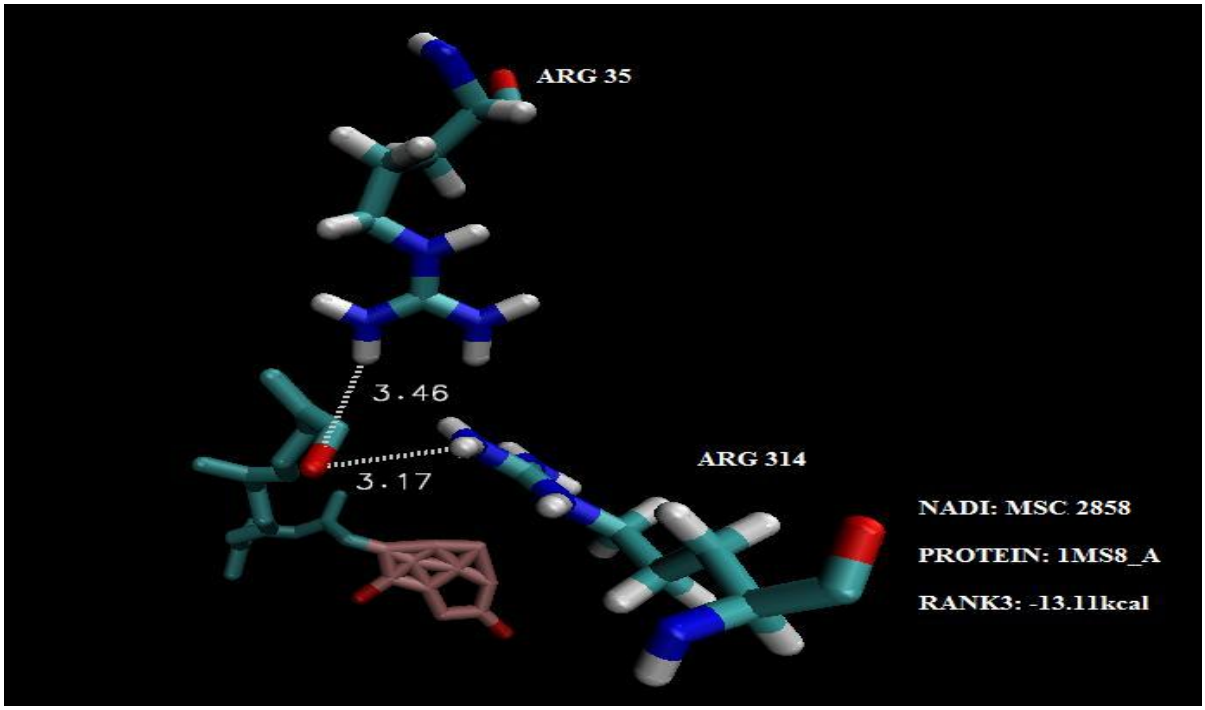




Rank 3: MSC2858, Capsorubinal
Fruit of Capsicum annum
Free Binding Energy:-13.11kcal



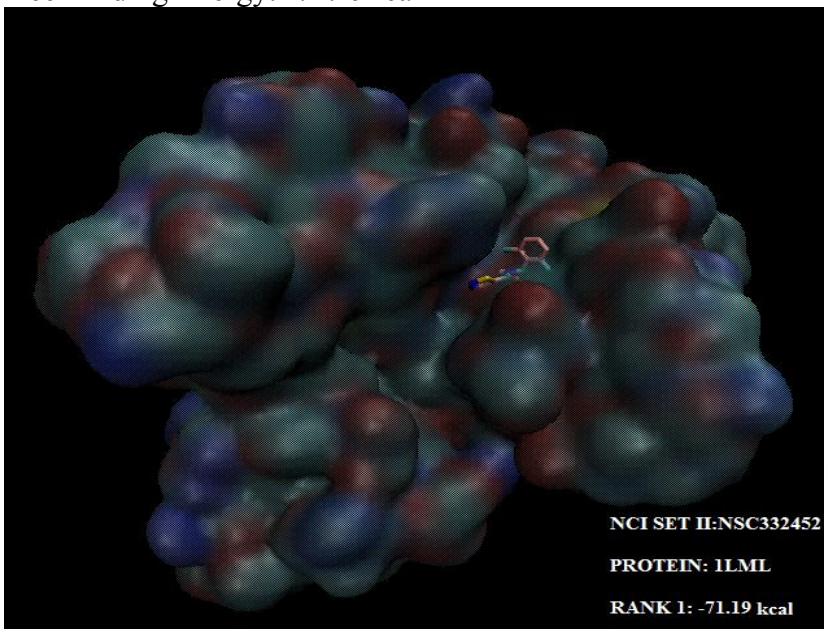


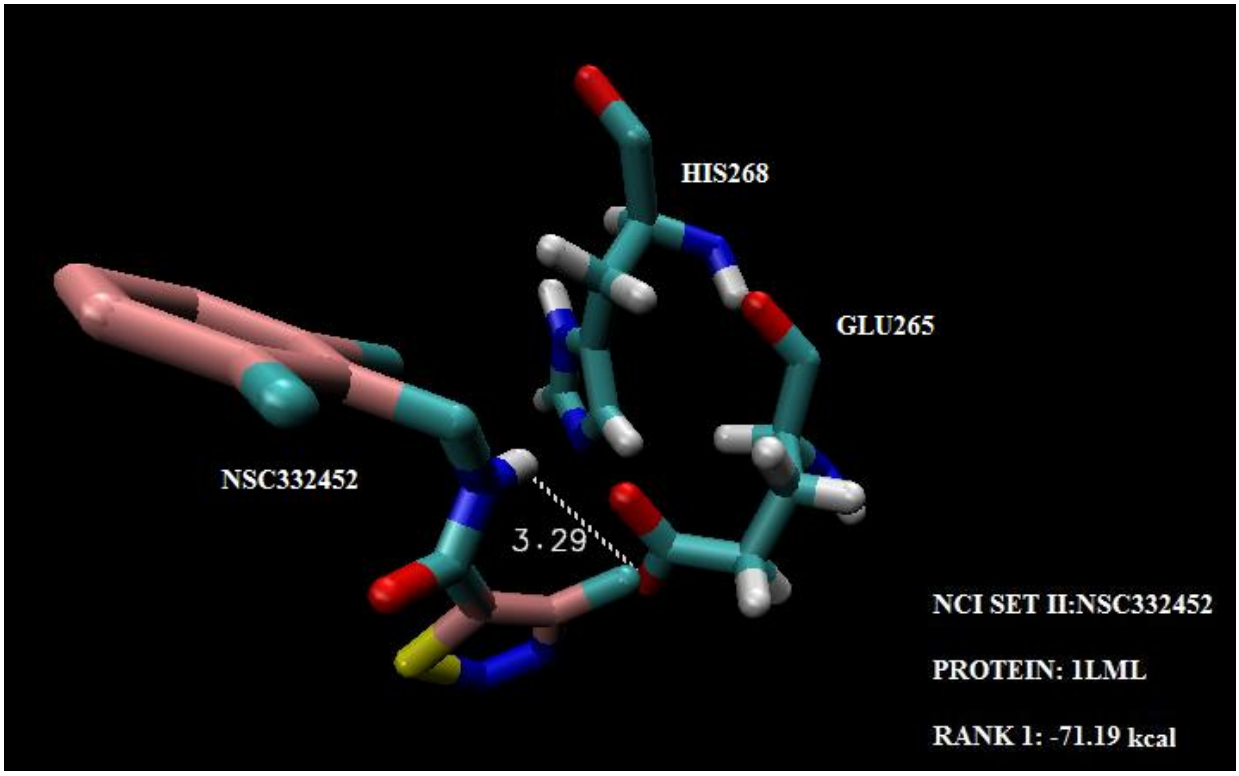


1LML

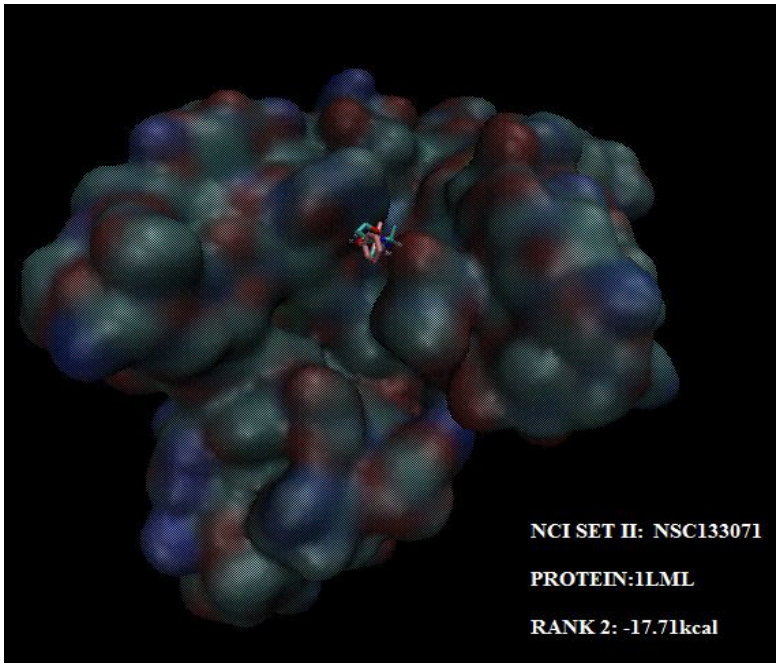
NCI Diversity Set II

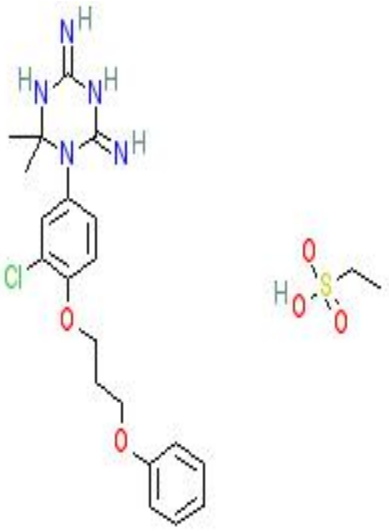
Rank 1: NSC332452, N-(2-chloro-6-fluorobenzyl)-4-methyl-1, 2, 3-thiadiazole-5-carboxamide
Free Binding Energy: -71.19kcal

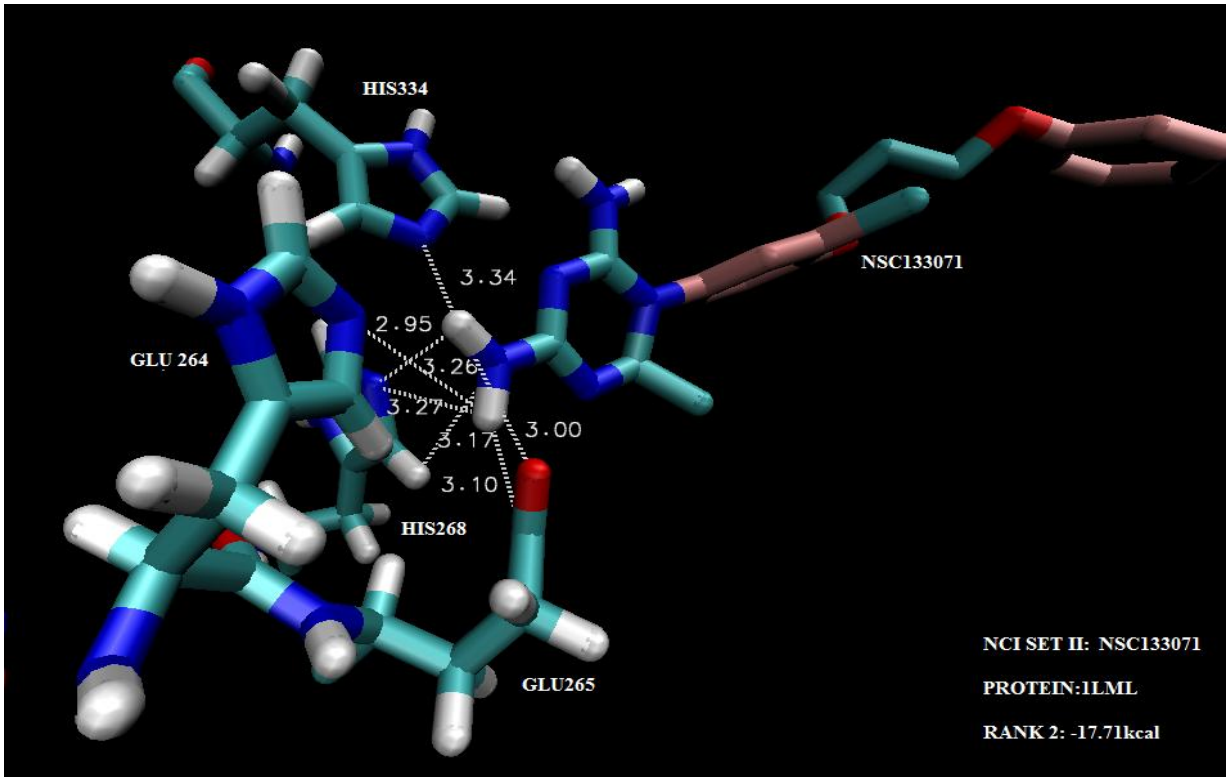




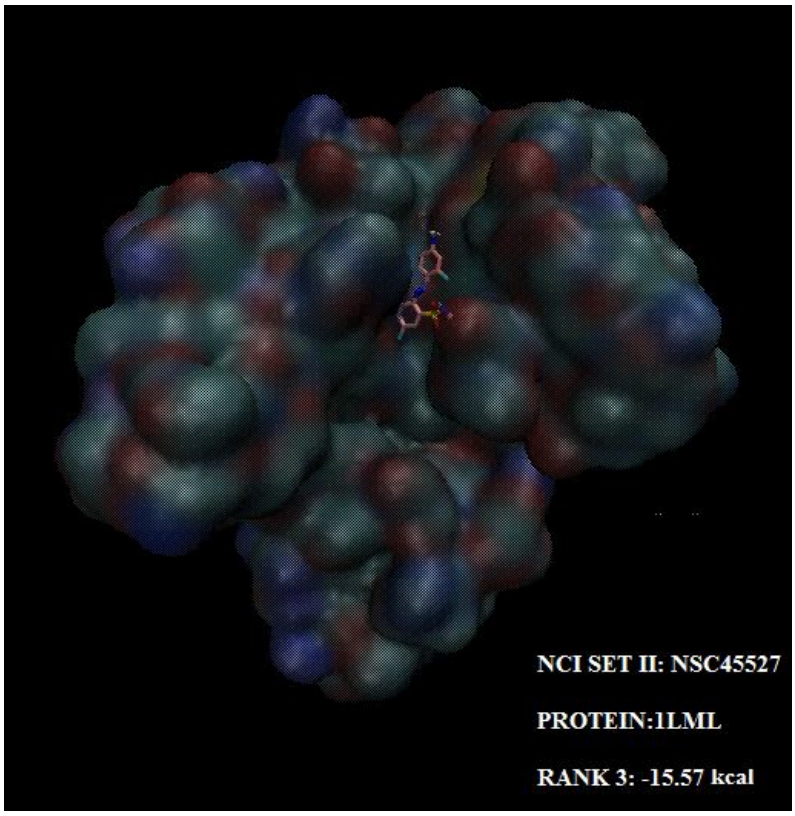
Rank 2: NSC133071, ethanesulfonic acid compound with 1-(3-chloro-4-(3-phenoxypropoxy) phenyl)-6, 6-dimethyl-1, 3, 5-triazinane-2, 4-diimine (1:1)
Free Binding Energy:-17.71kcal

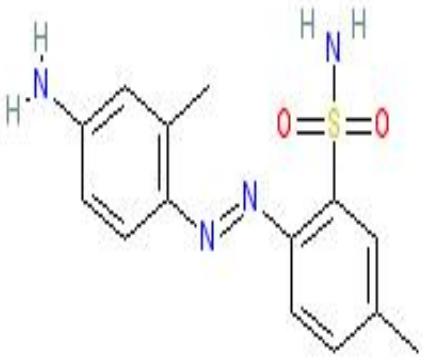


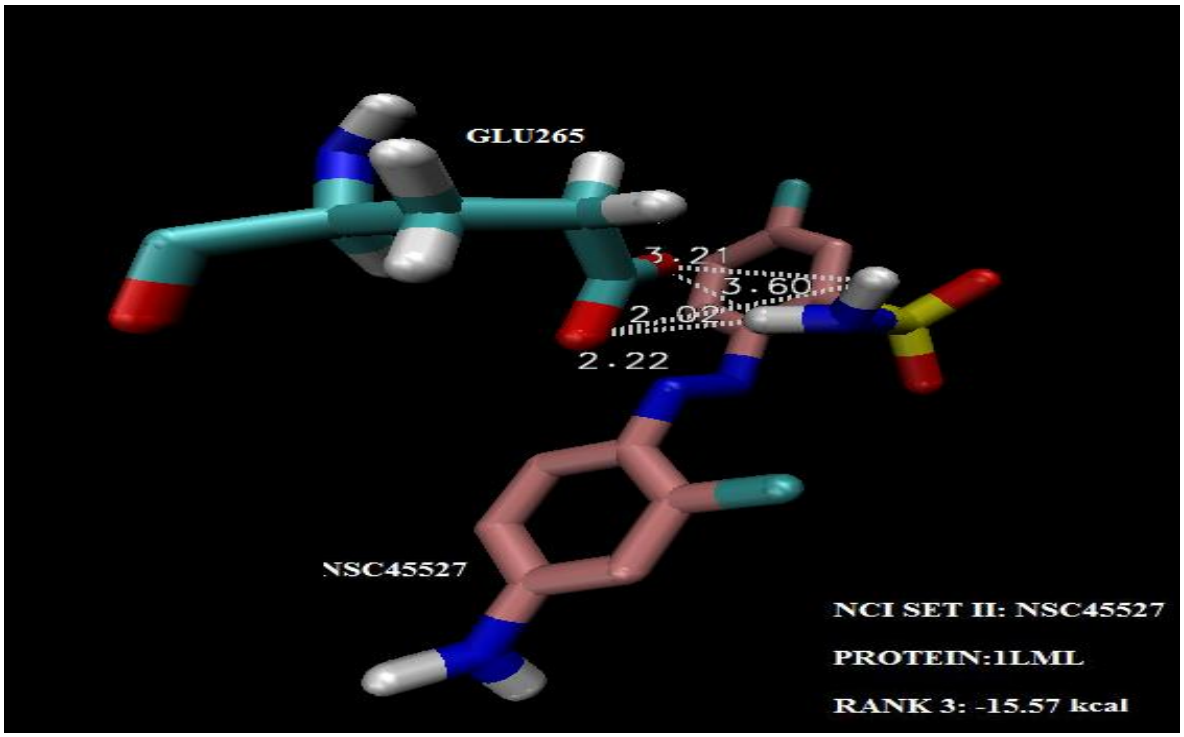




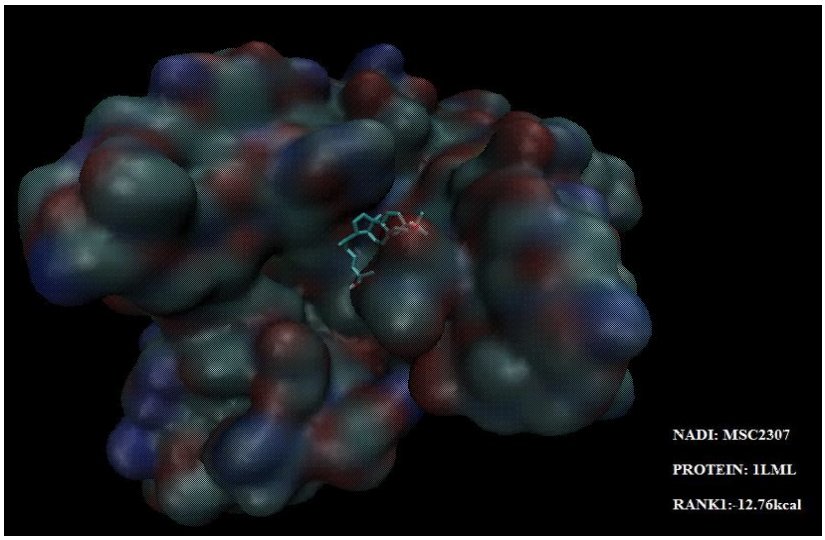
Rank 3: NSC45527, 2-((4-amino-2-methylphenyl)diazenyl)-5-methylbenzenesulfonamide
Free Binding Energy: -15.57kcal

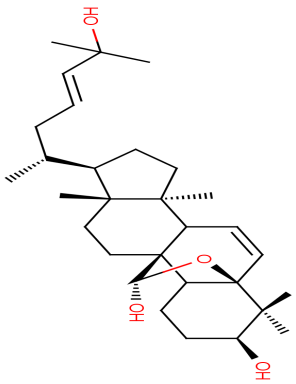


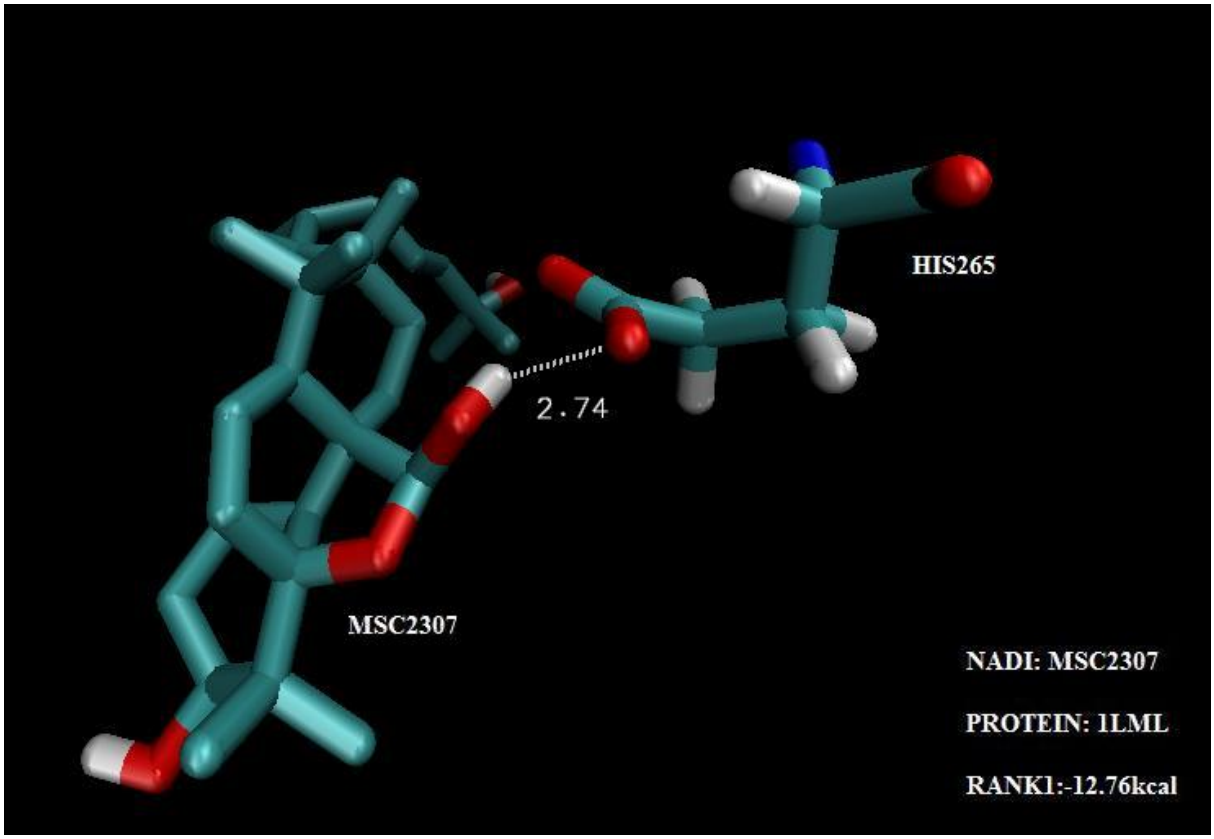




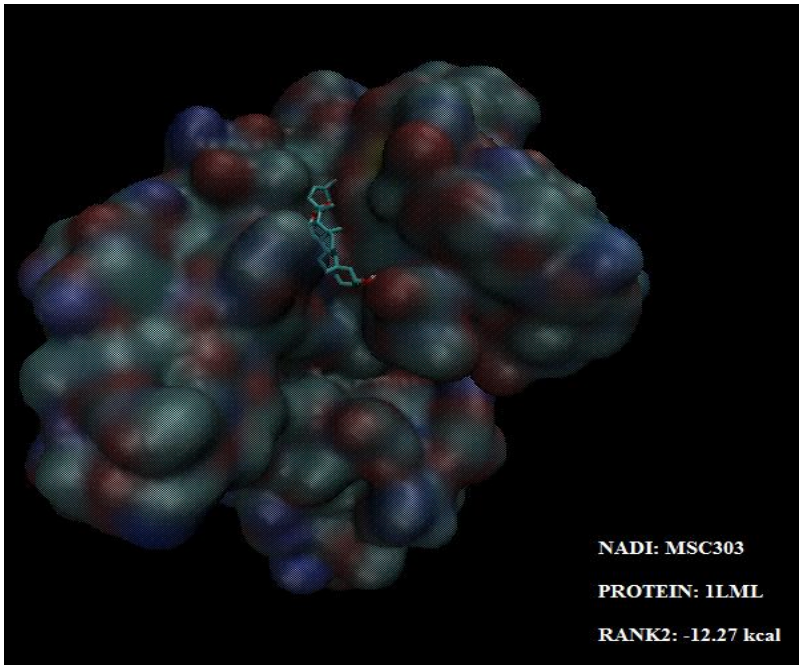
NADI
Rank 1: MSC2307, Kuguacin R
Leaf/Vine of Momordica charantica
Free Binding Energy: -12.76kcal

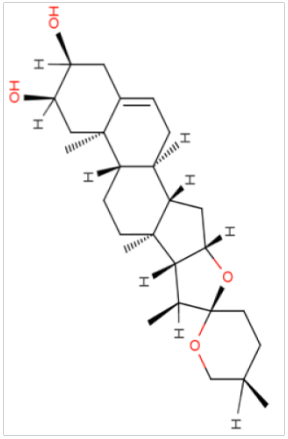


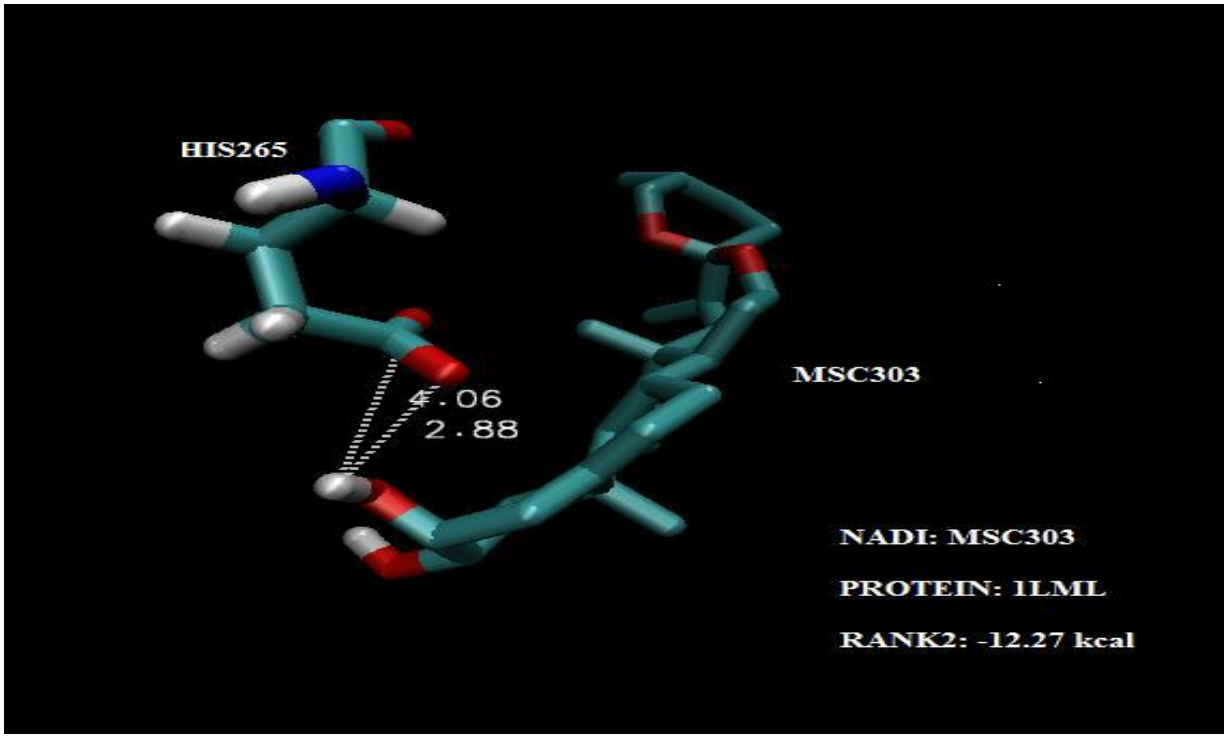




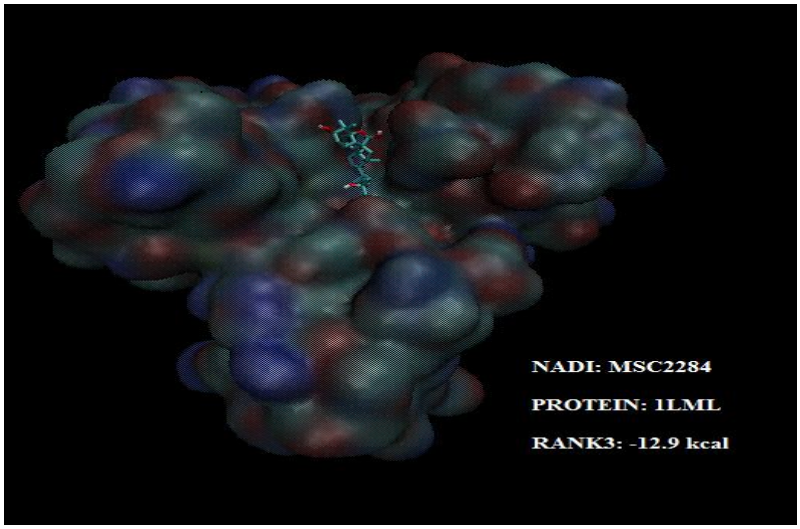
Rank 2: MSC303, Yuccagenin
Seed of Psophocarpus tetragonolobus
Free Binding Energy: -12.27kcal

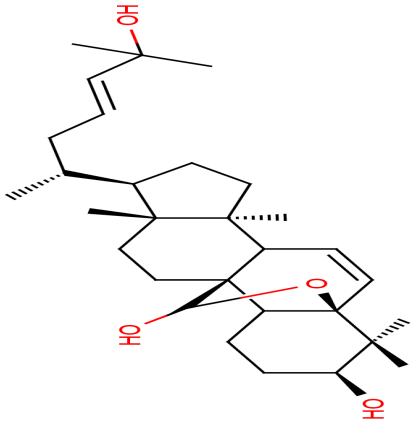






Rank 3:MSC2284, 5-beta-19-Epoxy-cucurbita-6-23(E)-diene-3-beta-19, 25-triol
Gourd of Momordica charantica
Free Binding Energy:-12.9kcal





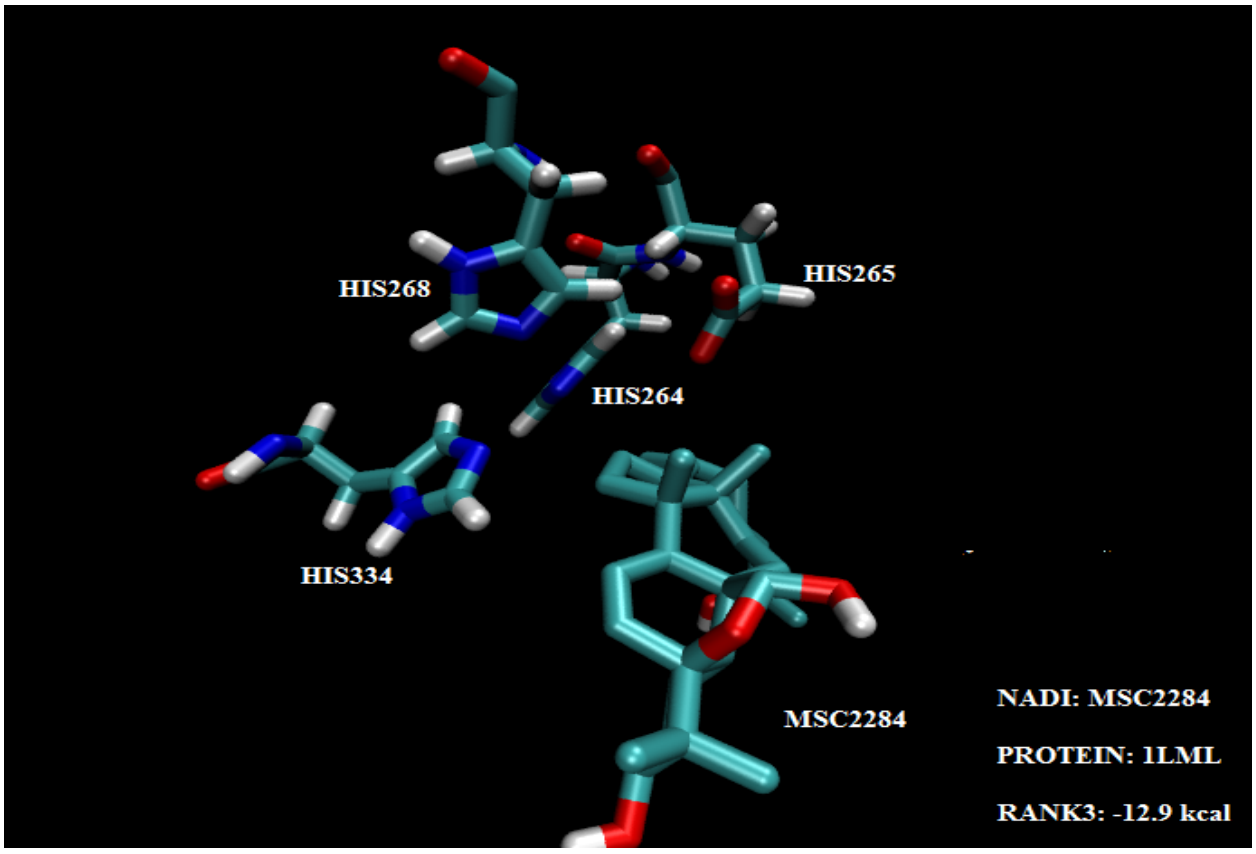


Table 1a. TOP Ten Ranking Results based on Free Binding Energy

	NCI	NADI
1MS8	1.NSC332452 -113.45 kcal 2.NSC65537 -16.70 kcal 3.NSC45527 -15.27 kcal 4.NSC114449 -14.88 kcal 5.NSC120290 -14.50 kcal 6.NSC3753 -13.88 kcal 7.NSC121861 -13.78 kcal 8.NSC49852 -13.71 kcal 9.NSC30663 -13.67 kcal 10.NSC133071 -13.66 kcal	1.MSC2284 -14.23 kcal 2.MSC154 -13.52 kcal 3.MSC155 -13.38 kcal 4.MSC2858 -13.11 kcal 5.MSC853 -13.03 kcal 6.MSC598 -13.02 kcal 7.MSC1645 -12.95 kcal 8.MSC385 -12.92 kcal 9.MSC2252 -12.89 kcal 10.MSC650 -12.87 kcal
1LML	1.NSC332452 -71.19 kcal 2.NSC133071 -17.71 kcal 3.NSC45527 -15.57 kcal 4.NSC114449 -15.35 kcal 5.NSC65537 -15.05 kcal 6.NSC120290 -14.85 kcal 7.NSC70959 -14.82 kcal 8.NSC71033 -14.41 kcal 9.NSC310365 -13.52 kcal 10.NSC96021 -13.48 kcal	1.MSC2284 -12.90 kcal 2.MSC2307 -12.76 kcal 3.MSC303 -12.27 kcal 4.MSC377 -11.97 kcal 5.MSC302 -11.81 kcal 6.MSC299 -11.52 kcal 7.MSC2246 -11.41 kcal 8.MSC2235 -11.31 kcal 9.MSC2283 -11.25 kcal 10.MSC91 -11.22 kcal

Table 1b. Final promising hits ranked by free binding energy and hydrogen bonding between protein and compounds

	NCI	NADI
1MS8	1.NSC332452 -113.45 kcal 2.NSC65537 -16.70 kcal 3.NSC45527 -15.27 kcal	1.MSC2284 -14.23 kcal 2.MSC154 -13.52 kcal 3.MSC2858 -13.11 kcal
1LML	1.NSC332452 -71.19 kcal 2.NSC133071 -17.71 kcal 3.NSC45527 -15.57 kcal	1.MSC2307 -12.76 kcal 2.MSC303 -12.27 kcal 3.MSC2284 -12.90 kcal

For the top three selected promising hits for each screening, an additional 1000 docking runs were performed to ensure the free binding energies are not dependent on number of runs and reported energies are accurate.

Table 1c. Results of 1000 runs with top three ligands of each database with 1MS8 and 1LML

	NCI	NADI
1MS8	1.NSC332452 -115.31 kcal 2.NSC65537 -17.49 kcal 3.NSC45527 -16.01 kcal	1.MSC2284 -15.25 kcal 2.MSC154 -13.87 kcal 3.MSC2858 -13.27 kcal
1LML	1.NSC332452 -74.04 kcal 2.NSC133071 -19.65 kcal 3.NSC45527 -16.19 kcal	1.MSC2307 -13.32 kcal 2.MSC303 -12.41 kcal 3.MSC2284 -13.70 kcal

For SMAP active site comparison, only 141 crystal structures of *T. Cruzi* proteins were reported by the Protein Data Bank. A run using 1MS8 protein and template ligand DANA was used for the comparison of proteins within the library. The cutoff P-value was set at 0.001. However, all significant results below this cutoff were other crystal structures of TcTS. The addition of *T.cruzi* homology models would give a more accurate result for off target binding of compounds.

Discussion

The top ten ranking results of each virtual screening with the NADI and NCI were analyzed for hydrogen bonding interactions and binding to the overall binding pocket. All results show good consensus binding. Promising hits were selected based on lowest free binding energy and compound interactions with surrounding important residues. Top rankings of the NCI diversity set II for 1MS8 and 1LML display abnormally low free binding energy and both lowest energy hits was the compound NSC332452 with energies of -113.45kcal and -71.19kcal. Because of these outliers, additional emphasis on binding location of the compound was especially important. Visualization with VMD showed that these compounds had good binding within the active site. Additional docking with Autodock 3.0 for the top three promising hits of each database with 1MS8 and 1LML was run 1000 dockings to ensure that the outliers were actual results. From this run, similar docking scores for the proteins were recorded with only ±1-3 kcal difference between 100 and 1000 runs. Additional analysis is needed to determine if the hydrogen interactions are valid based on angle and donor/ acceptor atoms in the bond proximity. Also, consideration of hydrophobic interactions within the pocket in relation to atoms of the compounds needs to be taken.

Based on these factors, promising hits for 1MS8 TcTS include NCI compounds NSC332452 (-113.45 kcal), NSC65537 (-16.70 kcal), NSC45527 (-15.27 kcal) and NADI

compounds MSC2284 (-14.23 kcal), MSC154 (-13.52 kcal), MSC2858 (-13.11 kcal). 1LML hits include NCI NSC332452 (-71.19 kcal), NSC133071 (-17.71 kcal), NSC45527 (-15.57 kcal), MSC2307 (-12.76 kcal), MSC303 (-12.27 kcal), and MSC2284 (-12.90 kcal).

With the promising compound hits, a literature search of therapeutic effects of these natural products in the NADI database found that top hits are used as herbal remedies for Type II diabetes. The plant, Momordica charantica, is used for glycaemic control and is known to have an insulin-like effect (Singh et al., 2011). Literature findings for MSC154 were closely related to protozoan infection treatment especially in cases of malaria and dysentery. Three dimeric alkaloids of nine found in the Alstonia angustifolia plant show significant inhibition of protozoan activity compared to the standard antiamoebic drug, emetine and antimalarial drug, chloroquine (Wright et al., 1992). However, studies on MSC154 are limited to only structure analysis and little information is found on effects.

As for the results of SMAP for 1MS8 screening, the 141 crystal structures in the set *T. Cruzi* proteome library only showed significant results for other PDB IDs of the same protein TcTS. Since the compiled *T. Cruzi* library only represented a small portion of proteins in the proteome, it was not surprising that the only results were that of TcTS. Added homology models from Modbase to the library would give a more accurate prediction of other proteins in the proteome that are not yet crystallized. Because of the limited proteome collection, a screening for 1LML against the *T. Cruzi* proteome will be conducted after a better representation of the proteome is gathered through homology models.

Conclusion

In conclusion, virtual screening with the natural and synthetic compound databases, NADI and NCI diversity set II, identified compounds that may target the selected 1MS8 TcTS and 1LML gp63 proteins. Due to factors of low free binding energy, binding location and

available hydrogen bond interactions between protein- ligand, results show promising hits for possible inhibition.

Future Work

Upon obtaining homology models for the *T. Cruzi* proteome, a more thorough screening is prompted for 1MS8 and 1LML to give a good prediction of off target binding possibilities. Promising hits of the screening could be docked to similar proteins, predicting other targets in the proteome. With these results, assays can be created to test if these predicted bindings occur in in vitro experiments and if any exhibit inhibitory effects. Results would contribute greatly to drug development for treatment of Chagas Disease and further *T. Cruzi* research.

Acknowledgements

The support from UCSD Pacific Rim Undergraduate Experiences (PRIME), National Science Foundation, IOSE-0710726 and the Tosh Nomura Eureka! Scholarship has contributed greatly to the development of the project. I truly appreciate the guidance and presence of Dr. Gabriele Wienhausen, Dr. Peter Arzberger, Teri Simas and Tricia Taylor in this working and cultural experience. Special thanks to my UCSD and USM mentors and partners, Dr. Philip Bourne, Dr. Habibah Wahab, Dr. Li Xie, Dr. SyBing Choi and Chirag Krishna for the detailed discussions throughout the program as well as the technical assistance while abroad.

References

1. Buchini, S., Buschiazzo, A. and Withers, Stephen G. (2008), A New Generation of Specific*Trypanosoma cruzi trans*-Sialidase Inhibitors. Angew. Chem. Int. Ed., 47: 2700–2703. doi: 10.1002/anie.200705435

2. Centers for Disease Control and Prevention, *Parasites - american trypanosomiasis (also known as chagas disease)*. N.p., 2010. Web. 25 Aug 2012. <http://www.cdc.gov/parasites/chagas/gen_info/index.html>.

3. Cuevas, Ileana C, Juan J Cazzulo, and Daniel O Sa´nchez. "Gp63 Homologues in Trypanosoma cruzi: Surface Antigens with Metalloprotease Activity and a Possible Role in Host Cell Infection." *Infection and Immunity.* Vol. 71.10 (2003): 5739–5749. Web. 25 Aug. 2012.

4. *Enhanced NCI Database*. Web. 25 Aug 2012. <http://cactus.nci.nih.gov/ncidb2.1/>.

5. Kulkarni, Manjusha M, Cheryl L Olson, David M Engman, and Bradford S McGwire. "Trypanosoma cruzi GP63 Proteins Undergo Stage-Specific Differential Posttranslational Modification and Are Important for Host Cell Infection." *Infection and Immunity*. 77.5 (2009): 2193–2200. Web. 25 Aug. 2012.

6. "Leishmanolysin." n.pag. *Hydrolases Catalytic Sites*. Web. 25 Aug 2012. <http://www.enzyme.chem.msu.ru/hcs/cgi/protein.cgi?id=P08148>.

7. Macdonald, Mary H, Charlotte J Morrison, and Robert McMaster. "Analysis of the active site and activation mechanism of the Leishmania surface metalloproteinase GP63." *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology*. 1253.2 (1995): 199-207. Print.

8. Oppezzoa, Pablo, , et al. "The Crystal Structure and Mode of Action of Trans-Sialidase, a Key Enzyme in Trypanosoma cruzi Pathogenesis." *Molecular Cell*. 10.4 (2002): 757–768. Web. 25 Aug. 2012.

9. Schlagenhauf , E, R Etges , and P Metcalf . "The crystal structure of the Leishmania major surface proteinase leishmanolysin (gp63)." *Structure.* 6.8 (1998): 1035-46. Print.

10. Shingo , Arioka, et al. "Potent inhibitor scaffold against Trypanosoma cruzi trans-sialidase." *Bioorganic & Medicinal Chemistry*. 18.4 (2010): 1633–1640. Web. 25 Aug. 2012.

11. Singh, Jaipaul , , et al. "Medicinal Chemistry of the Anti-Diabetic Effects of Momordica Charantia: Active Constituents and Modes of Actions." *Open Medicine Chemistry Journal*. 5 (2011): 70–77. Web. 25 Aug. 2012.

12. Weinkopff, Tiffany. "Trypanosoma cruzi trans-sialidase.”. N.p., n.d. Web. 25 Aug 2012. <http://tiffw.myweb.uga.edu/bcmb8010/report.pdf>.

13. Wright, C. W, D Allen, Ya Cai, and J. D Phillipson. "In Vitro Antiamoebic and Antiplasmodial Activities of Alkaloids Isolated from Alstonia Angustifolia Roots." *Phytotherapy Research*. 6. (1992): 121-124. Web. 25 Aug. 2012.