In Silico Prediction of Host-Pathogen Protein Interactions in Malaria and Docking Analysis of Identified Potential Target with Modified Inhibitors

A Major Project dissertation submitted

in partial fulfilment of the requirement for the degree of

Master of Technology

In

Bioinformatics

Submitted by

Monika Samant

(2K12/BIO/15)

Delhi Technological University, Delhi, India

Under the supervision of

Dr. Yasha Hasija

Department of Biotechnology
Delhi Technological University
(Formerly Delhi College of Engineering)
Shahbad Daulatpur, Main Bawana Road,

CERTIFICATE

This is to certify that the M. Tech. dissertation entitled **"*IN SILICO* PREDICTION OF HOST-PATHOGEN INTERACTIONS IN MALARIA AND DOCKING ANALYSIS OF IDENTIFIED POTENTIAL TARGET WITH MODIFIED INHIBITORS"**, submitted by **MONIKA(2K12/BIO/15)** in partial fulfilment of the requirement for the award of the degree of Master of Engineering, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), is an authentic record of the candidate's own work carried out by her under my guidance.

The information and data enclosed in this dissertation is original and has not been submitted elsewhere for honouring of any other degree.

Date: 30.7.2014

Dr. Yasha Hasija
(Project Mentor)
Department of Bio-Technology
Delhi Technological University
(Formerly Delhi College of Engineering, University of Delhi)

DECLARATION

I, **Monika Samant**, hereby declare that the work entitled **"*In Silico* Prediction of Host-Pathogen Protein Interactions in Malaria and Docking Analysis of Identified Potential Target with Modified Inhibitors"** has been carried out by me under the guidance of Dr. Yasha Hasija, in Delhi Technological University, Delhi.

This dissertation is part of partial fulfillment of requirement for the degree of M.Tech in Bioinformatics. This is the original work and has not been submitted for any other degree in any other university.

Monika Samant
Roll No.: 2K12/BIO/15

ACKNOWLEDGEMENT

I want to express my profound sense of gratitude and respect to **Dr. P. B. Sharma, Vice Chancellor, Delhi Technological University** and **Professor B. D. Malhotra, Head, Department Of Biotechnology**, who had created all necessary conditions conductive to the initiation and completion of the project.

I am also thankful to **Dr. Yasha Hasija, Assistant Professor, Department of Biotechnology, Delhi Technological University, Delhi-110042** for her exceptional insight and for her exemplary guidance, monitoring and constant encouragement throughout the M. Tech course. I would also like to thank her for sparing the efforts in compiling the work presented here.

I express my sincere gratitude to **Dr. Anjani K. Tiwari, Scientist 'D', Division of Cyclotron and Radiopharmaceutical Sciences, Institute of Nuclear Medicine and Allied Sciences** for believing in me and giving me the opportunity to work in his lab as a trainee. I would like to thank him for his constant support during the project.

I am deep-heartedly thankful to **Ms. Nidhi Chadha, Senior Research Fellow, INMAS, DRDO** for guiding me with her immense support and encouragement throughout my project work. I am extremely thankful to her for bearing me with patience during the silliest mistakes in the project duration.

I would like to thank Division of Cyclotron and Radiopharmaceutical Sciences, INMAS, DRDO for providing their facilities and sound academic environment.

At last, I am extremely thankful to my parents, family members and friends whose blessings and support were always with me.

MONIKA SAMANT
2K12/BIO/15

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

SA	sub-Saharan Africa
WHO	World Health Organisation
AA	Amino Acids
PPI	Protein Protein interactions
Y2H	Yeast Two Hybrid
TAP-MS	Tandem Affinity Purification combined with Mass Spectrometry
RBC	Red blood cells
GTP	Guanosine Tri-Phosphate
APM	Amiprophos methyl
XP	Extra Precision Docking
BIPS	BIANA Interolog Prediction Server

In Silico Prediction of Host-Pathogen Protein Interactions in Malaria and Docking Analysis of Identified Potential Target with Modified Inhibitors

Monika Samant

Delhi Technological University, Delhi

ABSTRACT

Infectious diseases such as malaria have been a major concern in the field of healthcare. One of the most severe form of malaria is caused by *Plasmodium falciparum* which kills millions of people annually. The malarial infection is initiated by the protein interactions between pathogen and host proteins and severity of the disease is related to the extent to which the biological processes are shifted due to these interactions. The outcome of this disturbance also depends on the immune response of the host and how the pathogen interaction with the host, which is governed by the protein-protein interactions. Therefore, It is essential to analyze the protein-protein interactions among the host and pathogen proteins for understanding the process and characterizing specific molecular mechanisms involved in pathogen persistence and survival.

In this study, we intend to generate a complete protein-protein interaction network of human host and pathogen *Plasmodium falciparum* by integrating experimental data and computationally predicted interactions using interolog method. The highly interacting protein which was found to have important role in infection was considered as a potential target for drug development against malaria. According to the analysis α-tubulin, a pathogen protein is found to be essential for infection and is identified as a potential drug target. Inhibitors were designed by modifying amiprophos methyl by adding functional group at benzene ring and adding piperdine ring on its side chain. With the results of docking and binding affinity analysis, two modified inhibitors were found which showed better docking scores of -10.5 and -10.43 and has better binding affinity of -83.80 and -98.16 with target. This inhibitor can be can be considered as a potential drug molecule and can be further tested in vivo for its property as an anti-malarial drug.

Keywords: Malaria, α-tubulin, Amiprophos methyl, Docking, Interolog

INTRODUCTION

Despite clinical researches in the field of infectious diseases, they remain to be the major problem in the worldwide health issues (Snow, Guerra *et al*. 2005, O'Meara, Mangeni *et al*. 2010). Exploring the infection process in detail can help to decipher the mechanisms that govern infection and controls infection process. Pathogen have evolved infection mechanism whereas human have evolved immune responses as defense mechanisms. A majority of host-pathogen interactions are governed by specific protein-protein interactions (Barnes, Durrheim *et al*. 2005).

To obtain a deep understanding of the infection process, the specific interactions between host and pathogen needs to be studied (Mufunda, Nyarango *et al*. 2007). Host pathogen protein interactions are typically studied using conventional small-scale methods which focuses on single protein at a time. Few methods for large-scale discovery have also been discovered such as yeast two-hybrid experiments which allows more comprehensive identification but at the cost of high cost and time (Dyer, Murali *et al*. 2007, Mendez-Rios and Uetz 2010).

Malaria is one of the most devastating disease which is caused by parasitic protozoan *Plasmodium falciparum*. It causes millions of death every year and this rate is increasing with each growing year. According to WHO's Factsheet on the World Malaria Report 2013, 1.2 billion people out of a total of an estimated 3.4 billion are at high risk of malaria. Malaria is highly prevalent in sub-Saharan Africa where 90% of all malaria death occurs (Organisation, W.H., 2013).

There has been a lot of research going on in the field of malarial therapeutics. Presently there is a wide variety of anti-malarial drugs and significant progress has been made in the strategies for improvement and control of malaria but it remains to be the major health problem worldwide. The already present therapies are showing resistance to the treatment (Dow, O'Hara *et al*. 2000, Mara, Dempsey *et al*. 2013).

Here we have computationally predicted protein interactions between human host and *Plasmodium falciparum*. Then these predicted interactions are filtered on the basis of cellular to identify the feasible interactions and functional annotation on the basis of functional process and pathway involved was carried out. Out of all the interactions, highly interacting pathogen proteins were shortlisted and studied in literature for their functional significance in the infection process. It was observed that structural and assembly proteins are the most important proteins involved in the infection process. After the analysis of predicted protein interactions, we found that α-tubulin was one of the highest interacting proteins in malarial infection. It is also a validated target in the malarial infection.

In this study, we have tried to design an efficient drug molecule for the target. Several mitotic inhibitors are already present in the literature which interact with these protein and hinder the infection process. Amiprophos methyl is a validated tubulin inhibitor in reference studies and is found to have least mammalian toxicity. Therefore, we have designed derivatives of Amiprophos methyl at preferred locations with several functional group to find a molecule which has better binding affinity than the reference molecule and similar drug properties.

3. REVIEW OF LITERATURE

3.1. Malaria

Malaria is one of the serious infectious diseases which causes millions of death annually all over the world (Dyer, Murali *et al*. 2007). More than half of the countries in the world are malaria endemic countries (Figure 1). It is an important health problem and major cause of morbidity and mortality specifically in pregnant women and children below age of five years. In last few decades, it is estimated that 400-1000 million cases and 1-2 million cases are reported annually due to malaria, out of which 90% of the cases occur in sub-Saharan Africa (SA) (Snow, Guerra *et al*. 2005).

In 2013, approximately more than 200 million cases and more than 6 million deaths occurred due to malaria. It was reported from the statistics that 1300 children or one child every minute die every day due to malaria (Organisation, W.H., 2013). World also suffers with economic losses in poor countries as a result of public expenditure in malaria treatments. The cost of prevention and care are worsening conditions even more; the fact that no efficient vaccine for treatment currently exists (Foster and Phillips 1998) and the acquired parasite resistance has superseded numerous drugs (Kooij, Janse *et al*. 2006) necessitate urgent attention to malaria research. Consequently, studies on discovering a vaccine or better, less costly prevention methods have become critical.

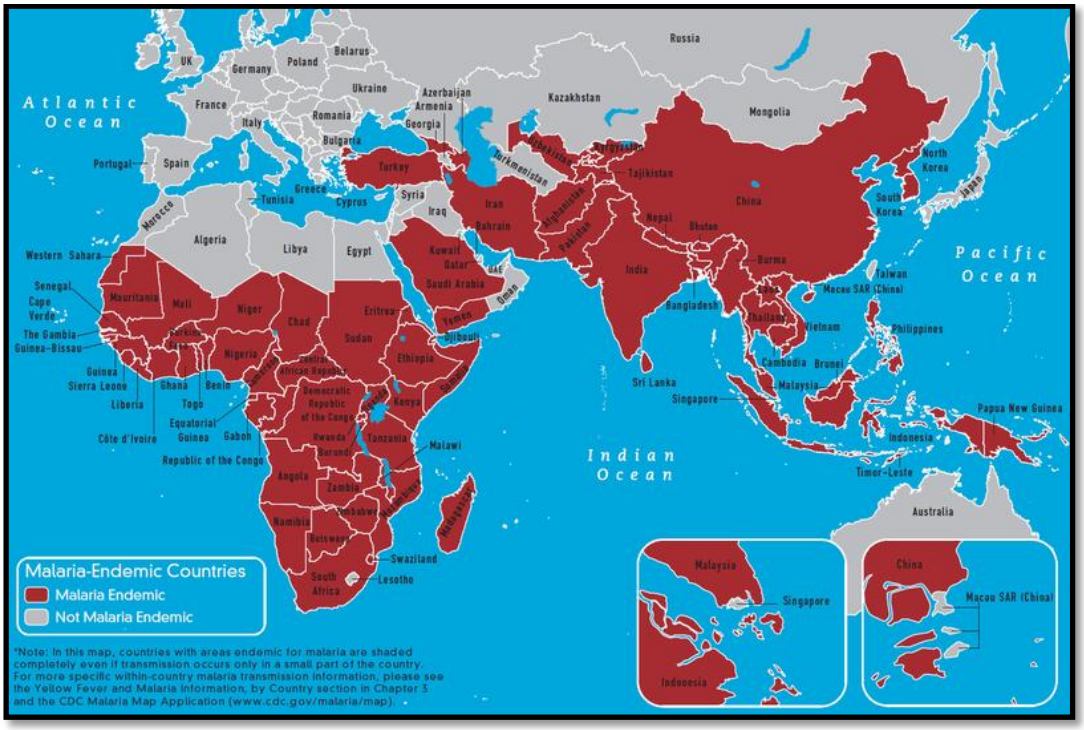


Figure 1: This map shows the malaria endemic countries in the Eastern hemisphere (Source -: http://wwwnc.cdc.gov/travel/yellowbook/2012/chapter-3-infectious-diseases-related-to-travel/malaria.html)

Malaria is a transmittable disease caused by protozoan parasites from the genus *Plasmodium* Female anopheles mosquito act as vector or carriers of the *Plasmodium* parasite, which enables cross infection between humans. Approximately two hundred known species of Plasmodium exist; about eleven of these species infect human. In most of the studies on malaria, five species are found to be involved in malarial infection i.e. *P. Falciparum, P. vivax, P. malariae, P. ovale, and P. knowlesi* and the most severe form of malaria is caused by *P. falciparum* (Cai, Zhou *et al*. 2012). In this study, only *Plasmodium falciparum* is included, which is extremely pathogenic and causes progressive illness that frequently result in a coma or death.

3.1.1 Prevention and Control of malaria

Several methods which focuses on either at host, vector or parasite are employed to prevent the spreading of the disease. Early diagnosis and effective drug treatments is considered to be the most effective strategies for malaria control since it not only palliates the disease severity but also interrupt malaria transmission. Current methods of control are directed at controlling of breeding sites and early diagnosis of disease condition using chemotherapy(Barnes, Durrheim *et al*. 2005). Nowadays, enormous challenges have been created due to development of resistance among malaria vector to common insecticides and anti-malarial drugs (Mufunda, Nyarango *et al*. 2007).

3.1.2. Malaria in general

3.1.2.1 Life cycle of malaria parasite

Plasmodium falciparum has a complex life cycle with multiple stages and it is dependent on two host i.e. mosquito vector Anopheles and the human host for its complete life cycle.

There are four stages in the life cycle of *Plasmodium falciparum*

- Male gamete development
- Sporozites formation
- Liver-stage development
- Blood-stage asexual reproduction.

The first two processes takes place in the mosquito vector whereas the remaining two processes take place in the vertebrate host.

1. Male gametes are prepared for sexual reproduction when mosquito ingests a microgamete during a blood meal which is a haploid cell of parasite. Then microgamete undergoes three rapid rounds of DNA synthesis and mitosis and forms a cell with 8n genomic complement. In the next 3 min, these complement separates from each other and eight new haploid male gamete begin to assemble.

2. In the midgut of mosquito, male gametes fuses with female gametes to create a diploid zygote. These zygotes the develop and become motile which moves and gets embedded in the basal lamina of midgut epithelial wall. It then undergoes several rounds of DNA synthesis and mitosis and eventually a massive cytokinesis event

occurs which produces thousands of haploid sporozites. These sporozites assemble form mother cell surface and migrate to the salivary gland of mosquito

3. The life cycle begins when an infected female mosquito penetrates the skin of host to obtain a blood meal. During penetration, saliva along with elongated sporozites is inoculated in the bloodstream of human host. The sporozites now travel via bloodstream to the liver, where a process called schizogony (rapid asexual reproduction) occurs without any clinical symptoms. During schizogony, mature schizonts are produced. At the end of liver stage, these mature schizonts gets ruptured and release a large number of merozoites. These merozites then either infect other liver cells or invade red blood cells(RBCs, erythrocytes). Inside RBC, the merozites develop further, entering either a sexual phase or a asexual phase.

4. The main pathology of malaria is governed by RBC stage like recurring fever due to lysis of the infected RBCs. During asexual phase, a merozite enlarges in erythrocyte forming an uni-nucleate ring trophozite . The ring trophozite then develops into schizonts with multiple nuclei through mitosis of the nucleus. These schionts then divide into multiple nuclei merozite, which causes the erythrocyte to rupture. After these merozites exits erythrocytes, they release toxins into the blood stream which causes fever and chills and other known symptoms of malaria(Rowe, Claessens *et al*. 2009, Gerald, Mahajan *et al*. 2011).

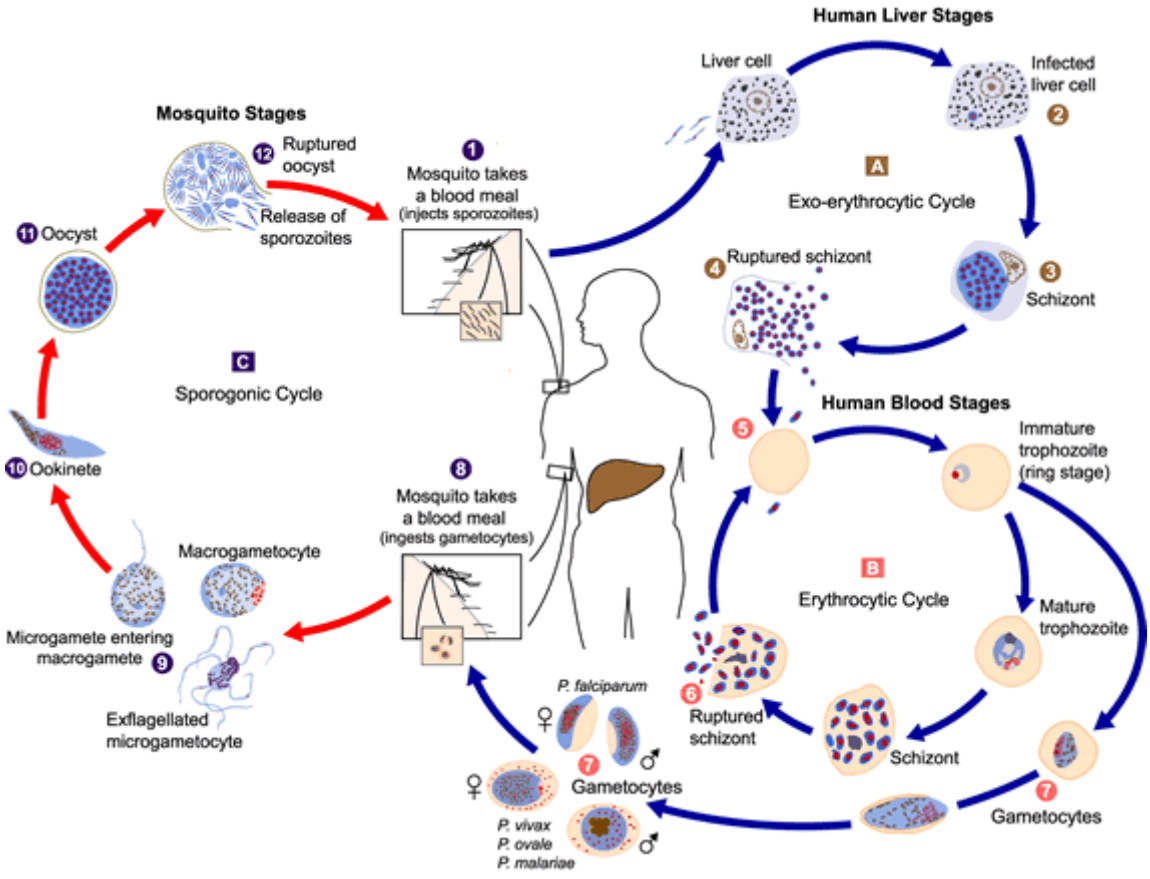


Figure 2: Life cycle of *Plasmodium falciparum* (Source-: http://www.cdc.gov/dpdx/)

Risk Of Malaria Around the World

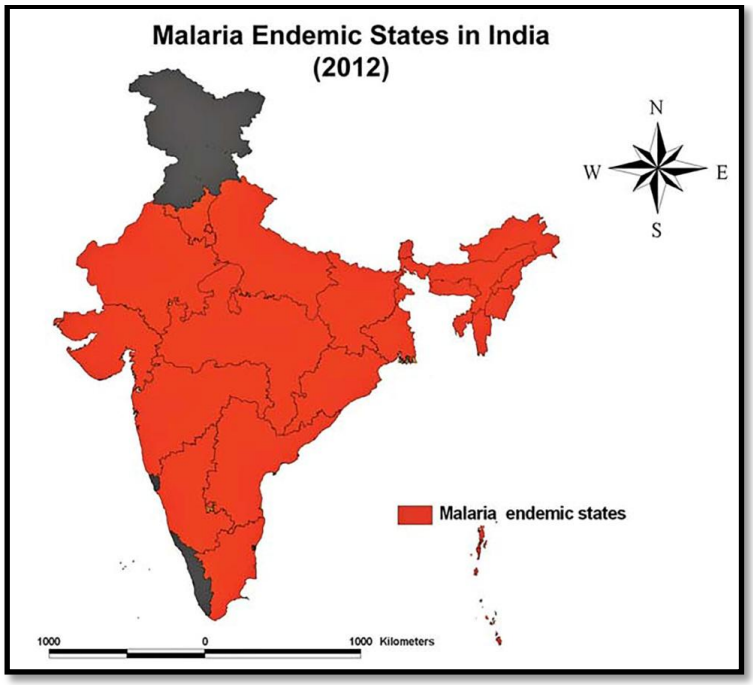


Figure 3: This map shows the Malaria endemic states in India. Source:(Palaniyandi 2012)

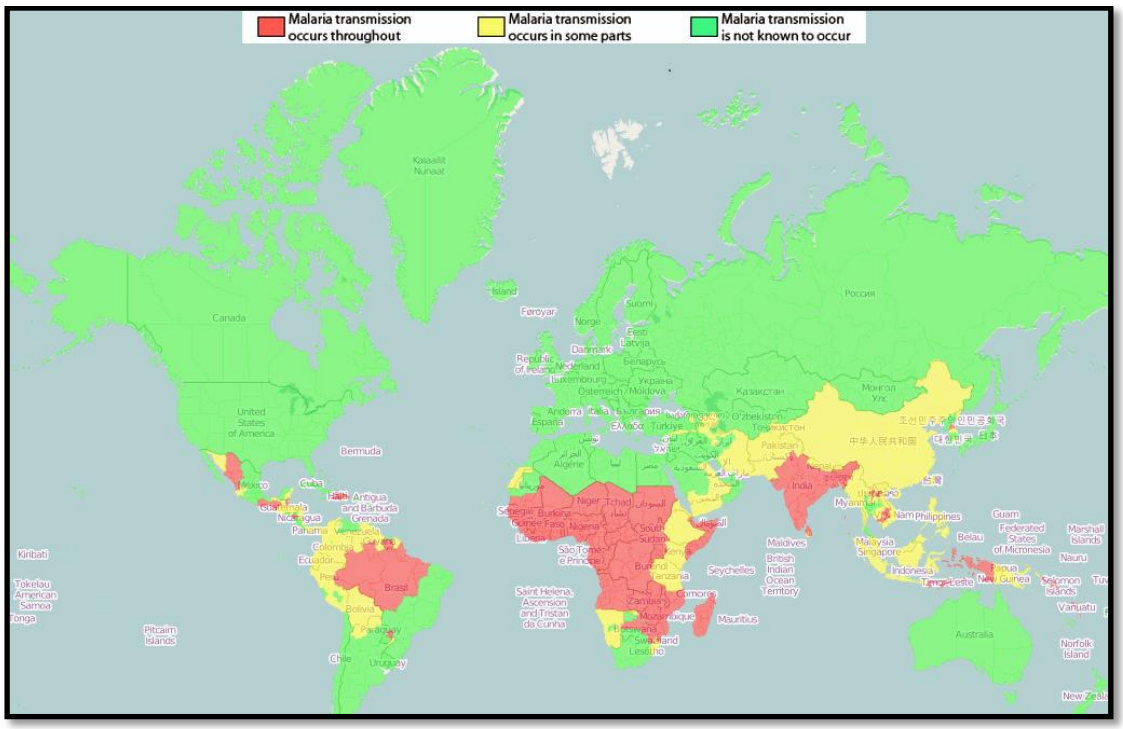


Figure 4: Global map showing the risk distribution of malaria.
(Source -: http://www.cdc.gov/malaria/map/index.html)

3.2. Protein-protein Interactions

Proteins are made up of 20 amino acids (AA) arranged in different combinations. The specific arrangement of the amino acids makes up a specific pattern which govern for specific interactions and specific functions. These specific patterns can be considered as reusable functional blocks called motifs which are conserved in species and during evolution. These interactions are evolved into functional pathways that transmit signals in the cell and produces a cellular response to an outside stimulus. These functional blocks, motifs are also observed in pathogen proteins which facilitate the disturbance in host's cellular functions.

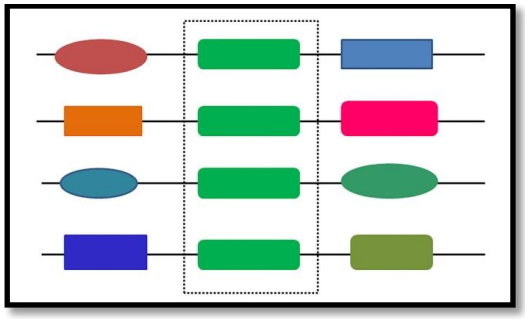


Figure 5: Demonstration of common functional unit i.e. motifs in the protein having similar functions

Protein-protein interactions (Yosef, Kupiec *et al*.) play a very significant role in the biological processes because they take part in almost all the cellular processes such as transcription, translation and almost all signalling cascades. Cellular processes depend greatly on the proteins for their interactions with other proteins to carry out specific functions as catalysts, signalling molecules, or building blocks in cells. Proteins need to bind together via domain interfaces to make the corresponding chemical reactions happen.. Specific interactions are needed by the proteins for some specific function, therefore some special relationship must exist for protein interactions (Mendez-Rios and Uetz 2010).

Hence it is important to understand these relationship. The importance of understanding for this relationship arouse the need of several experimental techniques to identify and analyse the protein-protein interactions (Zahiri, Bozorgmehr *et al*. 2013). Thus, a critical step towards understanding the inner workings of cellular machinery is to build a complete map of protein-to-protein physical interactions, which is called the interactome.

As efforts to get a complete image of the interactome, many high-throughput techniques have been developed over the last decade to detect protein interactions on a genome-wide level not only in yeast, two typical techniques among them are: Yeast two hybrid (Y2H) (Uetz, Giot *et al*. 2000, Ito, Chiba *et al*. 2001) and Tandem affinity purification combined with mass spectrometry (TAP-MS) (Gavin, Aloy *et al*. 2006, Krogan, Cagney *et al*. 2006).

Current protein complex detection methods rarely have 100% match for each detected complex, this hinders the comparisons between any two detected complexes from two species to identify the conserved pairs. Due to the above obstacles, protein complex detection from original PPI networks are still not an optimal approach for identifying conserved protein complexes among species.

3.2.1 Need of comparative interactomics and conserved protein complexes

One of the most important reasons behind the searching for conserved biological entities between species is that: conservation implies functional significance. This accounts for the birth of *comparative genomics* to identify proteins whose functions are conserved among species.

While sequence-conserved proteins form the basis of *comparative genomics*, it is also very important to consider the conserved patterns of interactions between proteins themselves, which can be referred to as *comparative interactomics (*Kiemer and Cesareni 2007). The reason here is that comparing interactomes among different species helps to transfer biological knowledge and function annotation at a higher level than comparing only protein sequences.

Conserved protein complexes and functional modules is one of the main outcomes from solving comparative interactomics problems. Identifying conserved complexes between species is a fundamental step towards identification of conserved mechanisms from model organisms to higher level organisms, such as protein translation, DNA transcription, cell cycle, etc. These mechanisms, at the same time, are considered as back-bones for a unit living system as cell. Therefore, conserved protein complexes are highly related to core cellular processes and critical to be studied carefully.

Another advantage supporting the comparative interactomics approach is that despite the noises in data, comparative analysis helps us to use the cross-species conservation criteria to focus on the more reliable parts of protein interaction networks and infer likely functional components. Once the number of well-studied species increases, we can use this approach to guide the search for protein complexes in newly-sequenced species, thereby increase the precision of current computational protein complex predicting methods.

Identifying conserved protein complexes can also help to understand the evolutionary mechanisms of protein complexes and protein interaction networks between multiple species, such as deriving evolutionary rate and age measures for protein complexes (Yosef, Kupiec *et al*. 2009).

In summary, the generalization from finding orthologous proteins to orthologous protein complexes is a significant extension.

3.2.2. Experimental methods

Many high-throughput techniques have been developed over the last decade to detect protein interactions on a genome-wide level not only in yeast, the following are the two typical techniques among them:

Yeast two hybrid (Y2H) is a screening technique for physical protein-protein and protein-DNA interactions which takes place in a living cell of yeast (*in vivo*). The two proteins of interest are injected into a genetically engineered strain of yeast. If they physically interact, a reporter is transcriptionally activated and we get a colour reaction on specific media. This technique is low-cost but can be degraded by a high number of false positive (as well as false negative) detections (Uetz, Giot *et al*. 2000, Ito, Chiba *et al*. 2001).

Tandem affinity purification combined with **mass spectrometry (TAP-MS)** is an *in vitro* technique, which has two steps: in the TAP stage, the protein of interest is embedded in a cell lysate to act as a bait for its interact-able proteins (prey) to bind, then together they will be identified by mass spectrometry after washing out the contaminants. Although TAP-MS technique still has a large number of false positive interactions and miss a lot of known interactions as Y2H, it can report higher-order interactions as protein complexes while Y2H has an advantage of detecting transient interactions (Gavin, Aloy *et al*. 2006, Krogan, Cagney *et al*. 2006).

As an inherent weakness of high-throughput techniques, protein interaction data generated by these techniques contains a large number of false positives. For this reason, PPI scoring methods are invented to assess the reliability of each interaction in the PPI network. Some typical PPI scoring methods are: FSweight (Chua, Sung *et al*. 2006), which use solely the PPI network topology to evaluate the reliability of PPIs and predict new interactions (Liu, Stenger *et al*. 2006).

3.2.3. Computational methods

Available computational prediction methods for PPI prediction can be divided into four categories on the basis of prediction.

1. **Homology-based approach**- It is an conventional method for prediction of protein interactions. The basic hypothesis of the homology-based approach is that the interaction between a pair of proteins in one species is expected to be conserved in related species (Matthews, Vaglio *et al*. 2001). This hypothesis is very reasonable as proteins descended from the ancestral pair of interacting proteins are expected to have same function and thus same interacting pairs.

2. **Domain based approach**- Domains are basic building blocks determining the structure and function of proteins and they play specialized role in mediating the interaction of proteins with other molecules (Itzhaki, Akiva *et al*. 2010). Some studies have proposed predicting host-pathogen PPI based on domain-domain interaction (DDI) (Dyer, Murali *et al*. 2007) and motif-domain interaction (Evans, Dampier *et al*. 2009).

3. **Machine learning methods** utilizing genomic and proteomic attributes-: Both supervised (Tastan, Qi *et al*. 2009, Dyer, Murali *et al*. 2011) and semi-supervised (Qi, Tastan *et al*. 2010) learning frameworks have also been used in predicting host-pathogen PPIs. A considerable amount of interacting and non-interacting pairs are usually needed by these machine learning algorithms to produce good classifiers.

3.3. Host-pathogen Protein interactions

For a pathogen to infect a host, it is important that the pathogen proteins must be present in the host and interfere with the host's usual biological processes. Viruses do not survive independently, therefore they enter their host and uses host's machinery to express their own proteins. Unlike viruses, bacteria inject only some effector proteins into the host cells (Mendez-Rios and Uetz 2010, Franzosa, Garamszegi *et al*. 2012).

The knowledge of host-pathogen protein interactions are very critical to the understanding of infection mechanism. Host-pathogen protein interactome guides the investigation on the essential PPIs involved in infection mechanism and therefore it aids in development of better treatment for prevention of disease. But prediction of protein-protein interactions poses many unique limitations. Already present methods of prediction of intra-species protein do not apply on the prediction of inter-species protein interaction. Hence, new computational methods are required for the study of inter-species protein interactions (Wuchty 2011).

3.3.1 Origin of host pathogen interactions

Most of the terms used for host-pathogen interactions have existed for approximately a century. Initially microbes were seen as the invader that causes disease. Further studies on the characteristics of microbes revealed that host-pathogen interactions do not always result in negative effects or disease. This meant that not all microbes were pathogens. Attention was shifted to the identification of harmless microbes and the definition of the different circumstances in which microbes exists without causing disease (Casadevall and Pirofski 1999, Casadevall and Pirofski 2000).

Terms like commensal and opportunists were suggested for describing this strange occurrence between microbe and hosts. These terms initially originated to describe microbe characteristics, rather than host-pathogen interactions. Thus, it became important to reconsider the definition of each term. Subsequently, studies developed towards a holistic perspective which includes both host and pathogen characteristics. in a framework for studying host-pathogen interactions (Casadevall and Pirofski 2000).

At the beginning of 20th century, it became clearer that pathogenicity was not a stable or consistent definition of microbes, because pathogens do not always causes disease. But from the studies, it was identified that host could influence a pathogen's ability to infect. Development of vaccines originated from this discovery. Later studies on infection identified

some hosts which were carrier of pathogen, but not as carrier of disease. This led to the hypothesis that certain hosts were more susceptible to pathogen than others.

There seems to be a lot of uncertainty around the theme of host-pathogen interactions. It is therefore necessary to recognize the need to study host- pathogen interactions as an integrated whole. Only then the interactions can be completely understood. The knowledge about the infection helps to guide effective drug discovery and development of new vaccines.

3.3.2 Protein interaction networks as a useful evidence to identify novel drug targets

Major problem in healthcare is resistance of the pathogen towards already present drugs. The new challenge is to develop new drugs against these resistant strains. Current drugs are based on specific proteins irrespective of their role in cellular network and their interaction with host proteins. But these proteins rarely act in isolation and they are involved in many integrated biological process.

Protein-protein interaction networks offers a complete knowledge of the interactions with other proteins and their involvement in the biological process. Hence, protein–protein interaction (Yosef, Kupiec *et al*.) networks has a lot of unexplored potential for identification of new generation drugs (Csermely, Korcsmaros *et al*. 2013, Zoraghi and Reiner 2013).

3.4 Docking

Structure information about interaction events among proteins could facilitate the new drug discovery and therapeutics. The success of molecules identified through these events against infectious diseases has enhanced the interest in this field of study (Rzychon, Chmiel *et al*. 2004). Several experimental methods are available for study of protein-protein interactions but has many limitations. Important binding interactions can be deduced by docking results which provides valuable information for development of drug discovery (Hillisch, Pineda *et al*. 2004). Earlier molecular docking was only used for virtual screening of compound libraries to identify lead compound for further drug designing (Desai, Patny *et al*. 2006).

Molecular docking is a computational method which studies protein-protein or protein-ligand interactions at many conformations and searches for the best conformation at which ligand binds to the receptor in most favourable manner in terms of geometry and binding energy (Gschwend, Good *et al*. 1996). This approach is very useful for rational structure based molecular drug design. Docking can be applied to whole protein in two ways i.e. Blind docking and docking at predicted sites. Blind docking is carried out at whole protein and finds the most favourable binding sites according to correct ligand binding orientation. Docking at predicted sites allows for virtual screening of ligand libraries at only high binding affinity sites (Laurie and Jackson 2006, Kumar and Zhang 2012). The latter binding approach is faster and cheaper than experimental methods.

There are two main types of protein docking, rigid and flexible.

3.4.1 Rigid docking

The basis of rigid docking approach is lock and key hypothesis of Fischer in 1890. In rigid docking, ligand undergoes several changes in its 3D conformation to identify the best spatial and energetically favourable conformation to fit into the receptor site (Sullivan and Holyoak 2008). This method is biased as it poses restrictions in conformational modification of receptor protein.

3.4.2 Flexible docking

Induced fit model is a more feasible protein-ligand binding approach was proposed by Koshland in 1958. In this approach of binding, both ligand and proteins are allowed to carry out conformational changes during interaction and form a complex with minimum binding energy. This approach increases conformational space for flexible binding. Partial flexible docking can also be done by selecting some residues to be flexible. Flexible docking has been proved to be more effective and efficient than rigid docking, but is computationally and time expensive (Lexa and Carlson 2012).

3.5 Relevance of Microtubule in malaria drug discovery

In *Plasmodium falciparum* cell cycle , microtubule is found to have a significant role. Microtubule is present at the tip of the merozites which are found to have a important role in cell division and infection. They are found to have a role RBC invasion because they disappear after invasion. In experimental studies it was confirmed when invasion was decreased and stopped completely when merozites were exposed to tubulin inhibitors.

Experimental studies have also demonstrated that microtubules were disrupted on exposure to anti-tubulin agent indicating the role of intact microtubule in merozite invasion. Microtubule is found in many stages of malaria parasite validating it as a potential drug target. As microtubule is found in several stages of malaria and it is used in cellular movement (Rawlings, Fujioka *et al*. 1992). Detailed examination of merozites in erythrocyte invasion identified that use of tubulin inhibitors is a potential approach for malaria therapy.

According to the theory, when tubulin inhibitors are added to sporozites, they got removed from blood circulation and become unable to invade liver cells. If they can evade liver cell, they are targeted at later stages (Fujioka and Aikawa 2002).

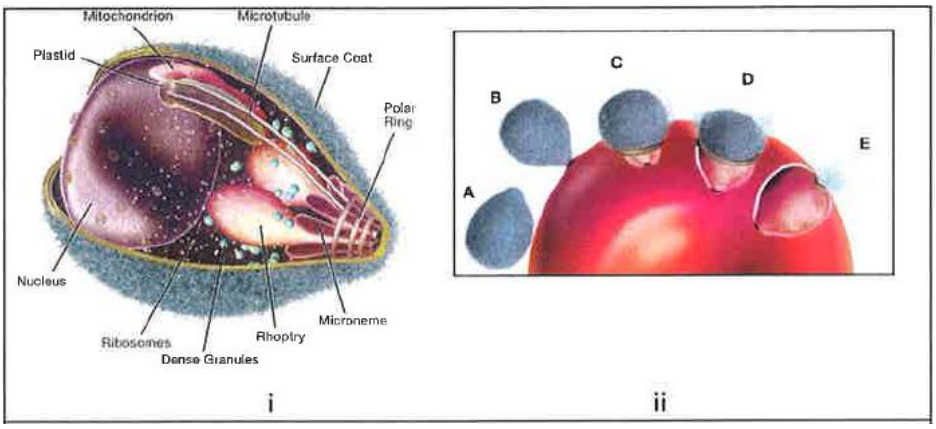


Figure 6: (i) Merozite with major organelle and cell structure (ii)Invasion of merozites into erythrocytes

From the above figure, we can see that merozites possess a band of microtubule at its tip called polar ring. Merozites also contain rhoptries and micronemes which are important during parasite invasion. The figure shows merozite invasion into erythrocytes. The merozite secretes proteins and enzyme upon recognition of which the merozites reorients itself so that the tip is attached to cell membrane of erythrocyte. The cell membrane become thicker and a tight junction is formed between cell receptor and ligand. A serine protease sheds the surface coat of merozites before entering the vacuole. The merozites divides several times inside the erythrocyte. Newly formed erythrocytes leave the erythrocyte , travel into bloodstream and invade new erythrocytes.

3.6 Microtubule as a drug target

Microtubule is a hetero-dimer consisting of two subunits i.e. α-tubulin and β-tubulin. α-tubulin has a irreversibly bound GTP and β-tubulin has a exchangeable GTP bound to it. α-tubulin and β-tubulin binds to each other and make a smallest subunit which polymerizes to make complete microtubule. Hence, it is a very important for structural integrity of the cell. Microtubule is an evolutionary conserved protein and has been used in many cancer therapies. While it is conserved in evolution, it shows a significant difference in protozoa and mammals. Therefore it is a potential target for drug discovery as it will show low mammalian toxicity (Rawlings, Fujioka *et al*. 1992, Martin, Robertson *et al*. 1997, Anthony, Waldin *et al*. 1998, Armson, Menon *et al*. 2002, Fujioka and Aikawa 2002).

Microtubule play an important role in cell division and other biological processes of malarial parasites. Microtubule depolymerising agents inhibits parasitic development and also kills cells in some cases. Microtubule depolymerising agents are of two types i.e. those which bind at vinblastine site and at colchicine binding site.

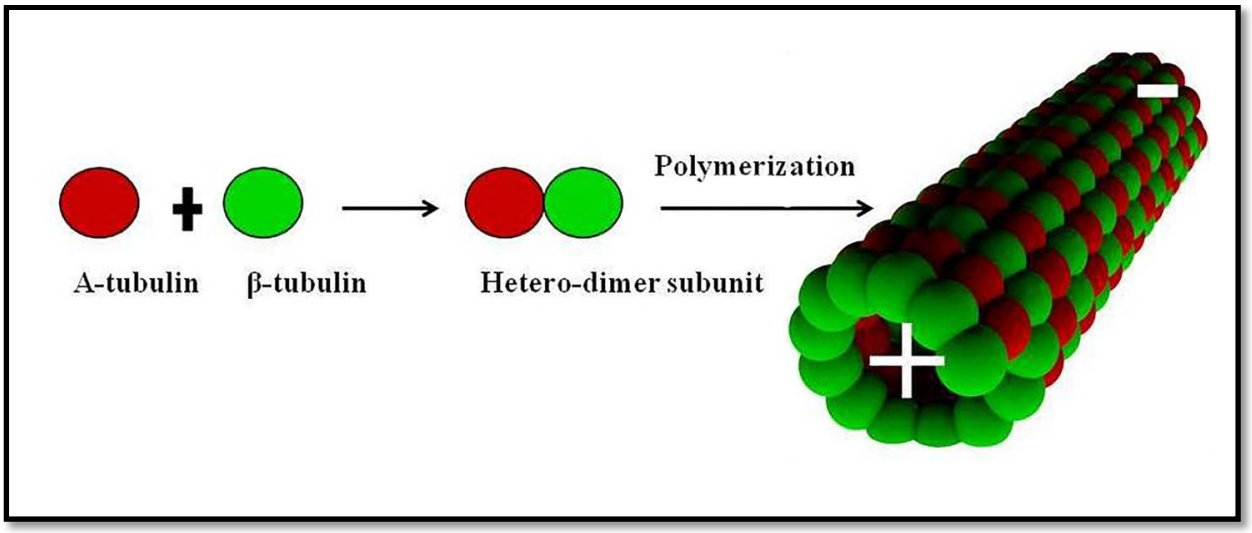


Figure 7: Assembly of microtubule form a continuous polymerization of hetero-dimer unit composed of α-tubulin and β-tubulin

3.7 New tubulin inhibitors and their specificities

There are few tubulin inhibitors already present and are being used for parasitic infections and cancer therapies.

Benzimidazole-:

Benzimidazole is a low-dose broad-spectrum anthelmintic having high therapeutic effect. Their mechanism of infection is interaction with β-tubulin and inhibiting polymerization. It has been found to be effective in other infectious diseases such as *Trichomonas vignalis*. But unfortunately, some of benzimidazole were found to have side effects (Dow, O'Hara *et al*. 2000, MacDonald, Armson *et al*. 2004).

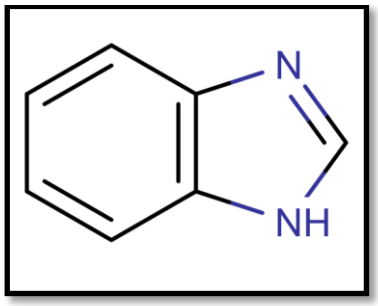


Figure 8: Chemical structure of Benzimidazole

Dinitroanilines-:

First reported in 1960, is used as a selective weed control in cotton agriculture. Treatment of weeds with dinitroanilines were found to have no microtubules and are shown to affect mitosis as a result of their interaction with microtubule. Several types of microtubules were found to have activity against microtubule. They were found to be interacting with both α and β tubulins as their mechanism of action. This group of compounds do not show any direct interaction with mammalian microtubule suggesting it as a potential drug agent (Armson, Menon *et al*. 2002).

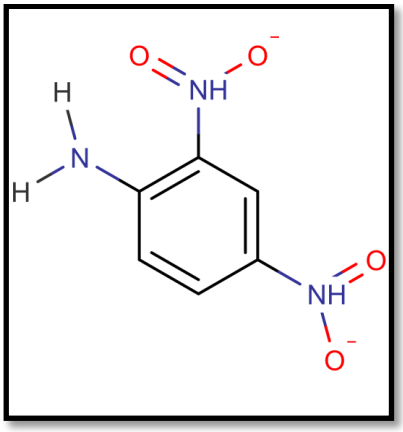


Figure 9: Chemical structure of Dinitroanilines

3.8 α-Tubulin as an antimalarial agent

In *Plasmodium falciparum*, tubulin family consists of two α tubulins i.e. αI and αII and one β subunit. αII is found to be specifically expressing in males (Rawlings, Fujioka *et al*. 1992). *Plasmodium* infection can be prevented by targeting microtubules by substances such as benzimidazole and dinitroaniline. This process involves capping and prevents polymerization.

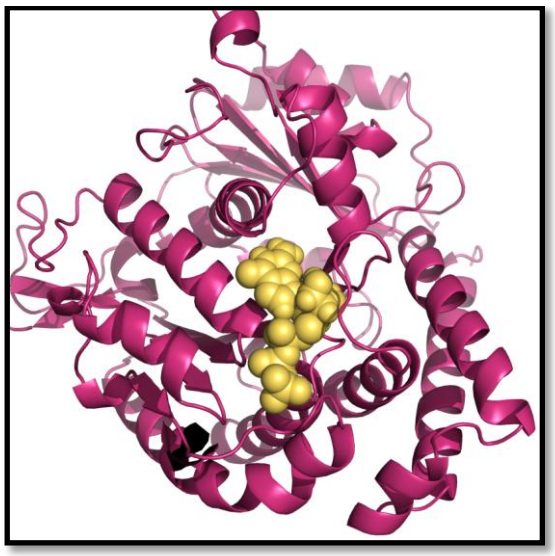


Figure 10: GTP bound α- tubulin

3.9 Amiprophos methyl as an inhibitor of α-Tubulin and an antimalarial agent

Amiprophos methyl (APM) is an antimitotic herbicide and is already known inhibitor for α-tubulin. It is found to be a promising molecule because of its low mammalian toxicity. It was reported in studies that amiprophos methyl has better specificity for pathogen proteins and has no binding site in human tubulin protein (Mara, Dempsey *et al*. 2011).

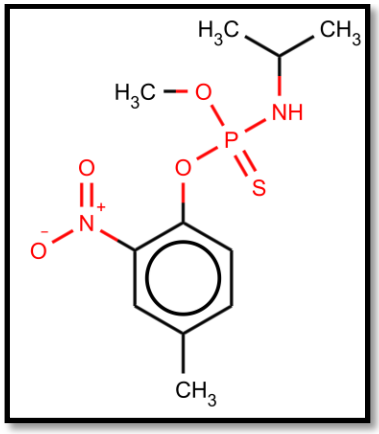


Figure 11: Molecular structure of Amiprophos methyl

Although tubulin is a ubiquitous protein, but still there is a significant difference in amino acids sequence of mammalian and parasitic tubulins. Dinitroanilines were approved as great anti tubulins and proved to be very good tubulin inhibitors. APM bind to tubulin in the same way as dinitroanilines. They are found to shown better inhibition in lower concentrations. With molecular studies, it was shown that these phosphorothiomidate compounds have similar electrostatic surfaces as dinitroanilines with similar shape and electronegative domains. APM also has low mammalian toxicity.

It was observed that APM prevents erythrocytic shizogony and blocks mitosis in *Plasmodium falciparum* infection and results in abnormal microtubule accumulation. This suggests that APM is worthy of investigation for its antimalarial potential (Fennell, Naughton *et al*. 2006, Mara, Dempsey *et al*. 2011, Mara, Dempsey *et al*. 2013).

3.10 Interolog approach

In this study, we have used interolog approach for prediction of protein interactions between Human host and pathogen of interest. This method is based on homology of the proteins. This method works on the basis on evolutionary conservation of interactions.

Interolog method predicts two protein to be interacting if their respective homologous protein are found to be interacting in the same species. We used BIPS (BIANA interolog prediction Server) for prediction server for prediction of protein interactions in Human host and pathogen of our interest (Garcia-Garcia, Schleker *et al*. 2012).

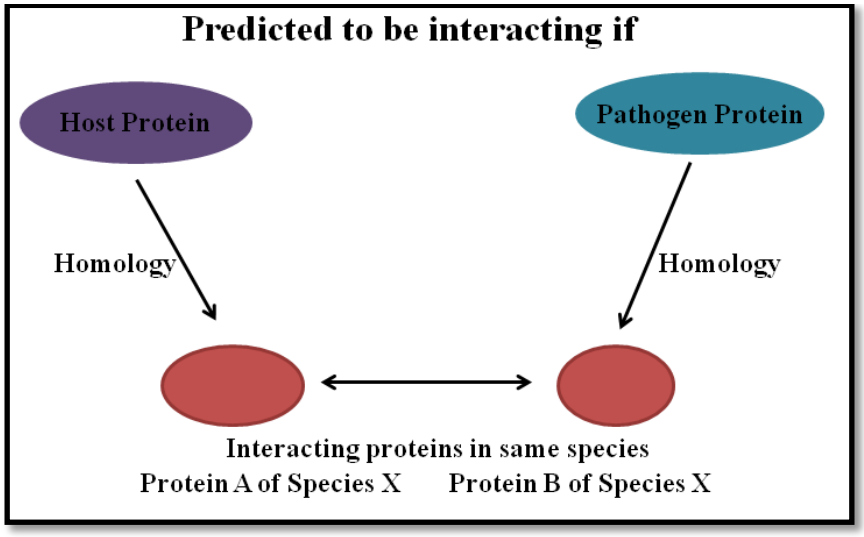


Figure 12: Pictorial description of Interolog approach

Work Plan of the Methodology- I

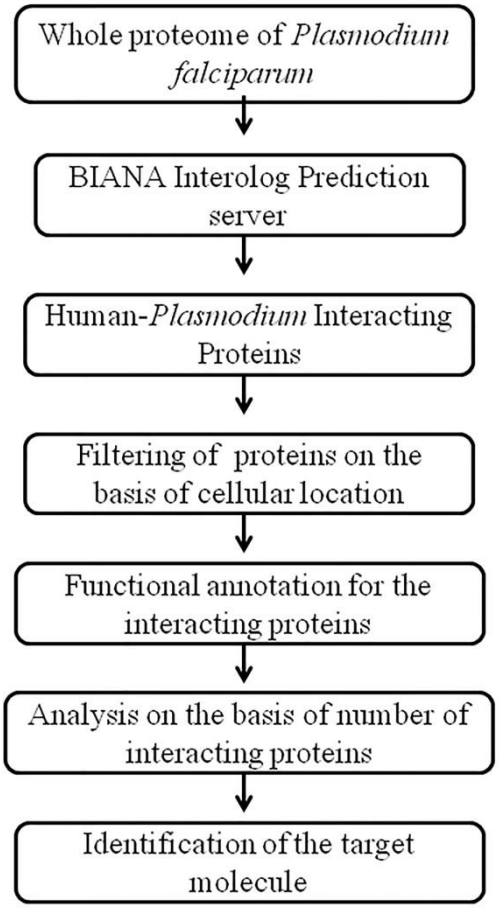
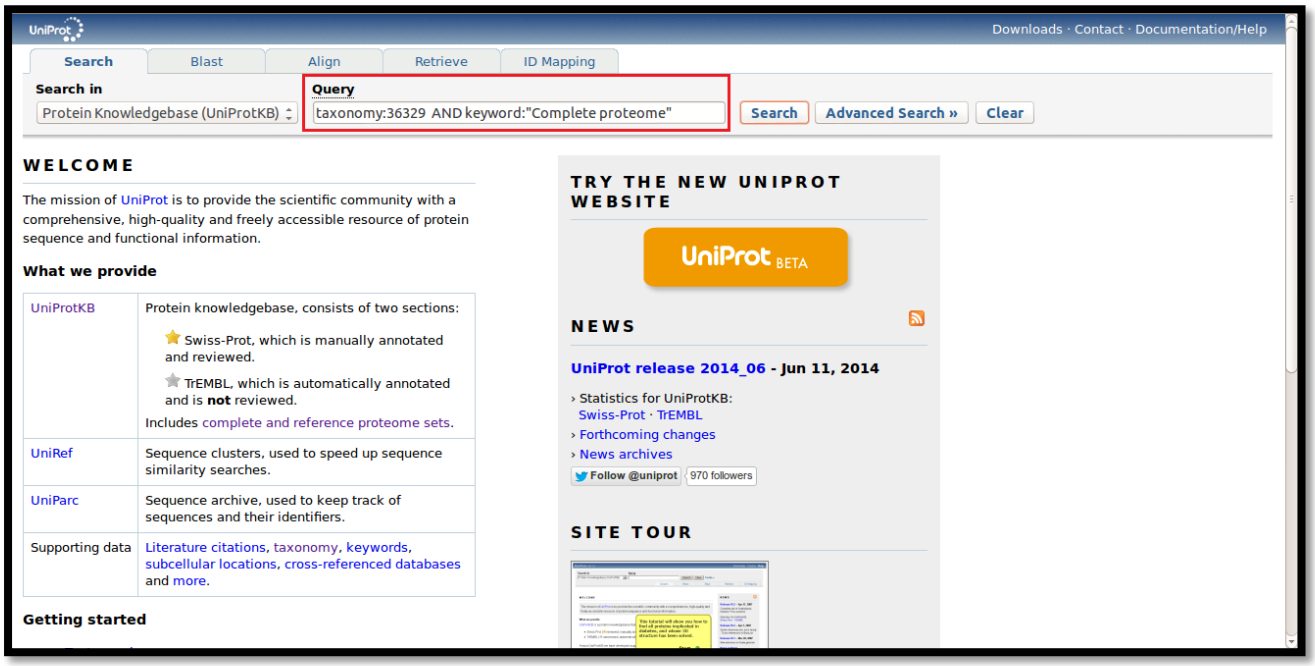


Figure 13: Work plan of the project to shortlist the protein for target. The whole proteome of pathogen is downloaded and entered into BIPS for generating *Human-Plasmodium* protein interaction predictions. The predicted protein pairs were then filtered, functionally annotated and analysed to identify a significant and potential protein to be used as a drug target

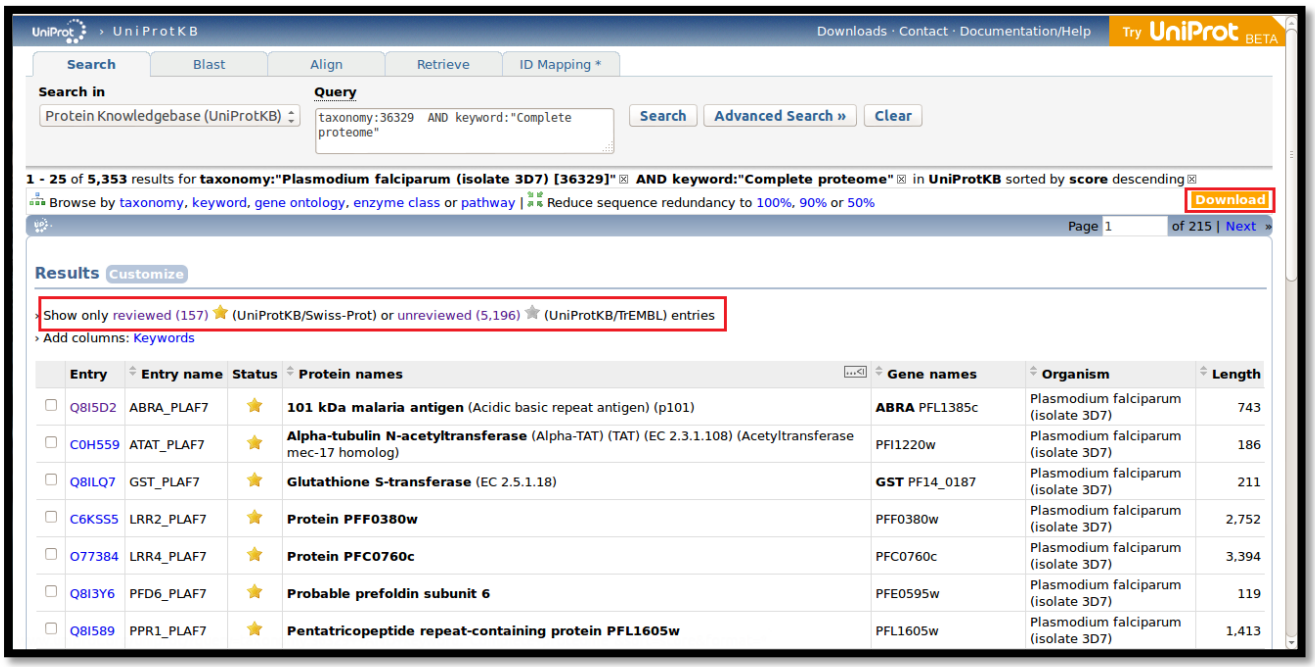
4. METHODOLOGY

4.1 Retrieval of Proteome -:

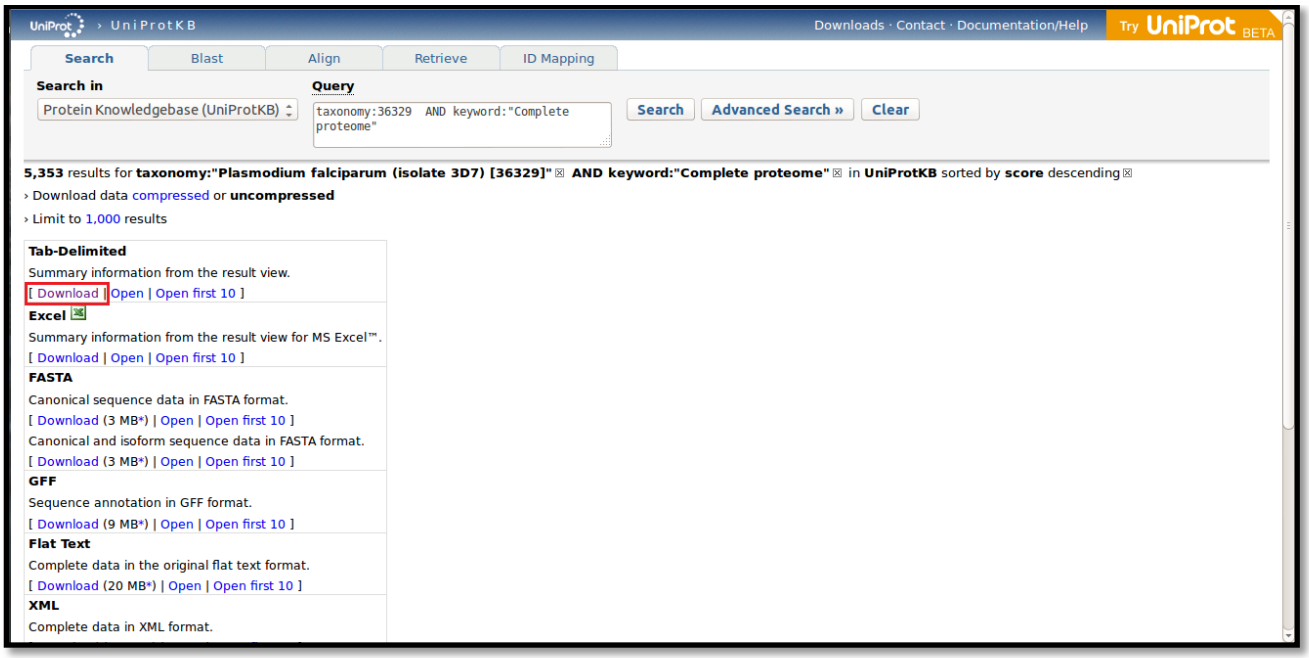
The whole proteome was downloaded from Uniprot database by entering in the query section. " t**axonomy:36329 AND keyword:"Complete proteome"**



4.1.2 It shows all the proteins with the desired taxonomy. The result is as shown below. Clicking on the "Download" option on the right will direct to the download options.

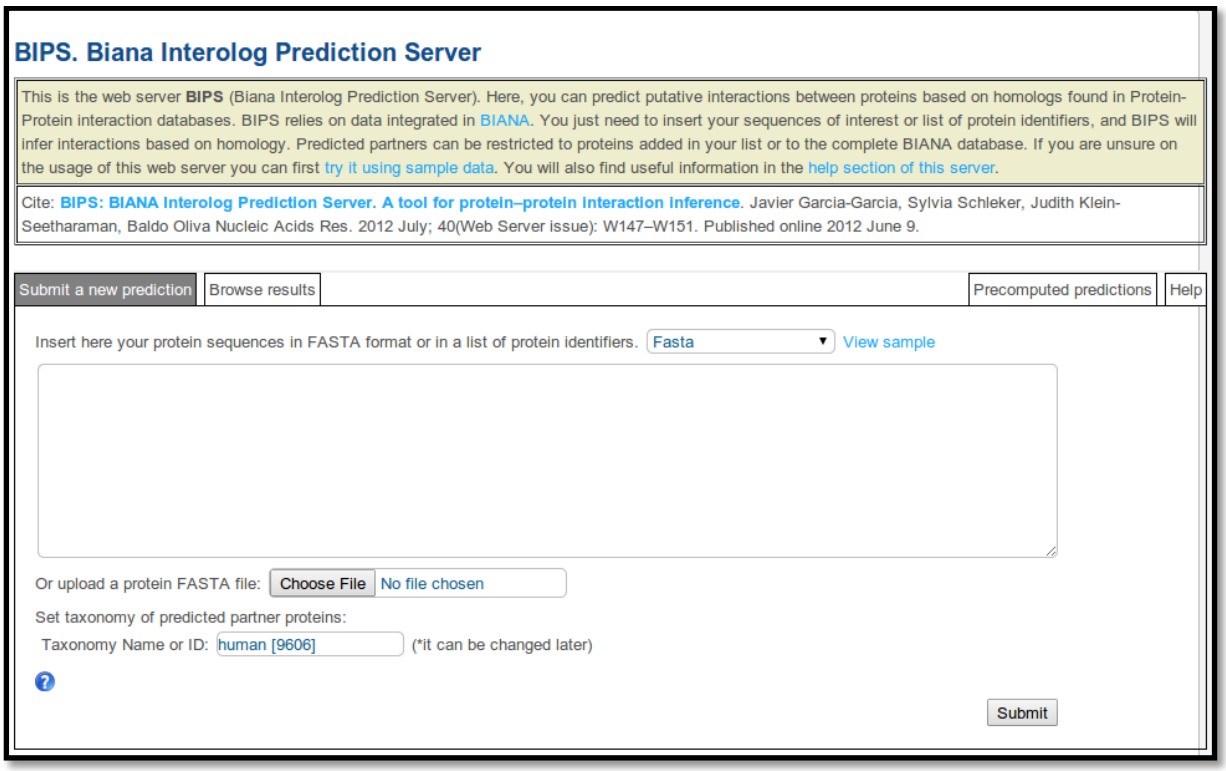


4.1.3 Clicking on the "Download" link will automatically start downloading the proteome in the desired format.

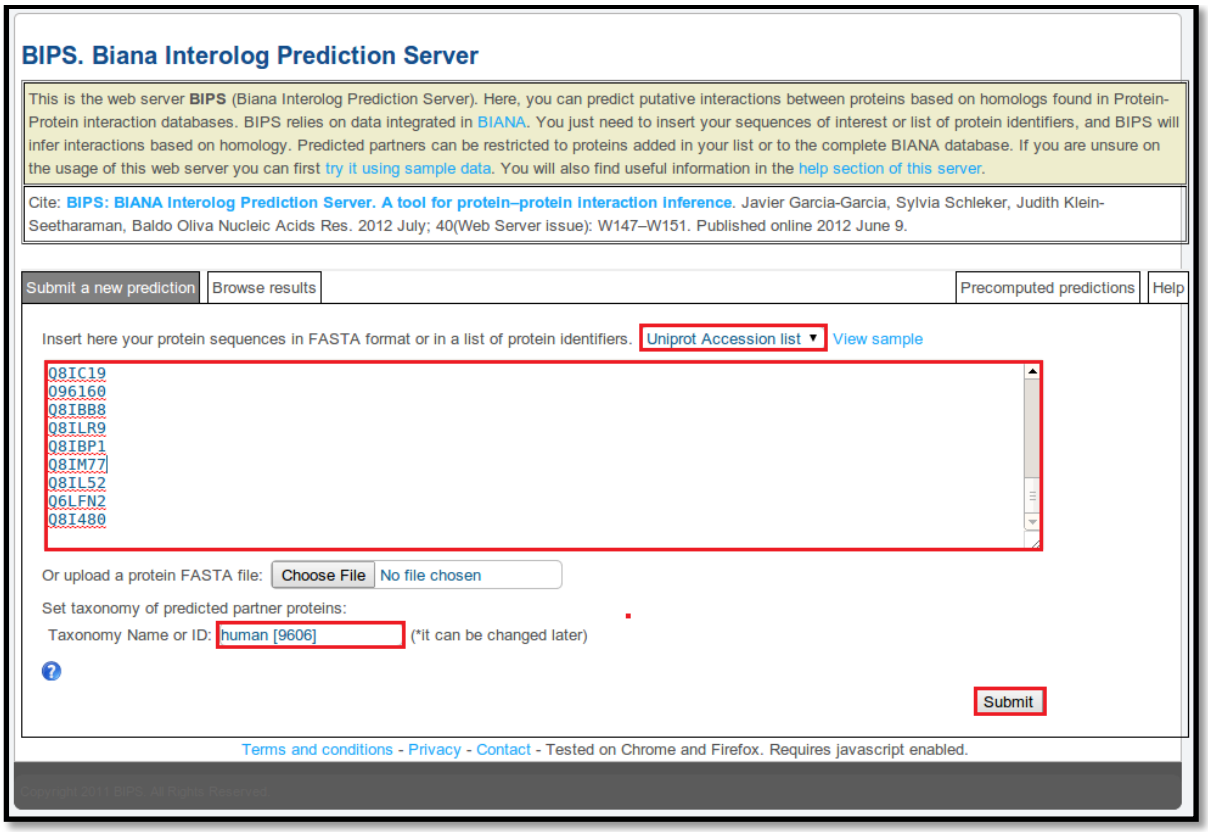


4.2. BIPS Prediction

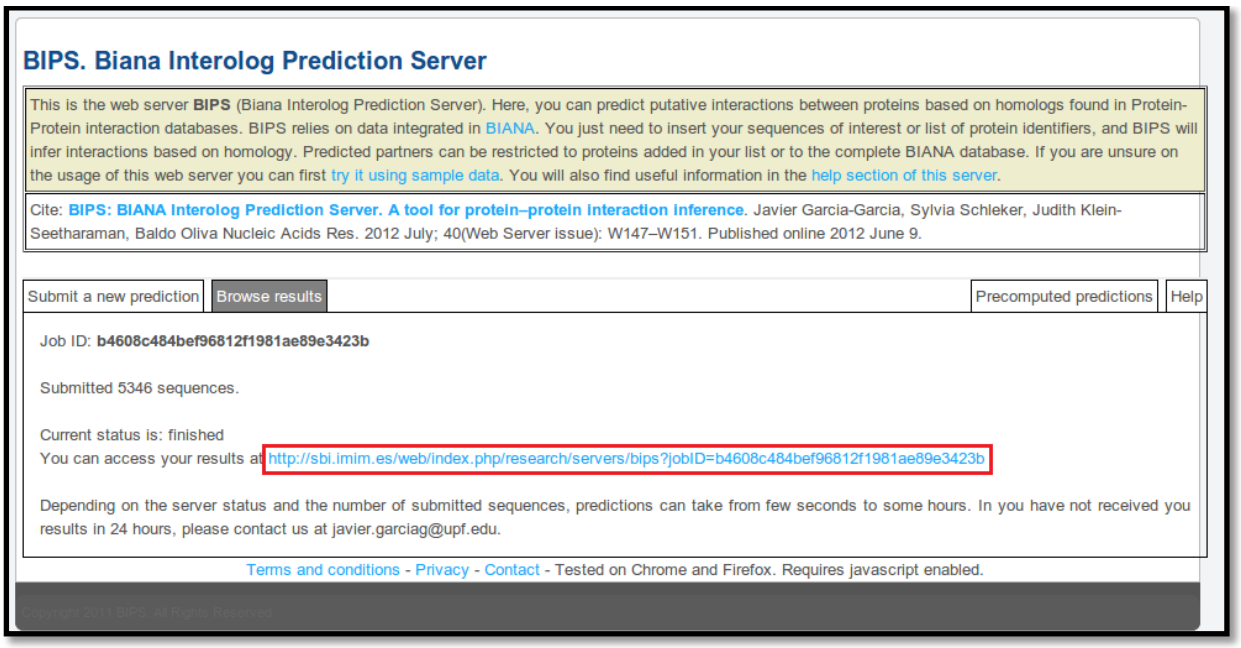
4.2.1 BIPS server is accessed by http://sbi.imim.es/web/index.php/research/servers/bips. This is web-page of BIPS server.



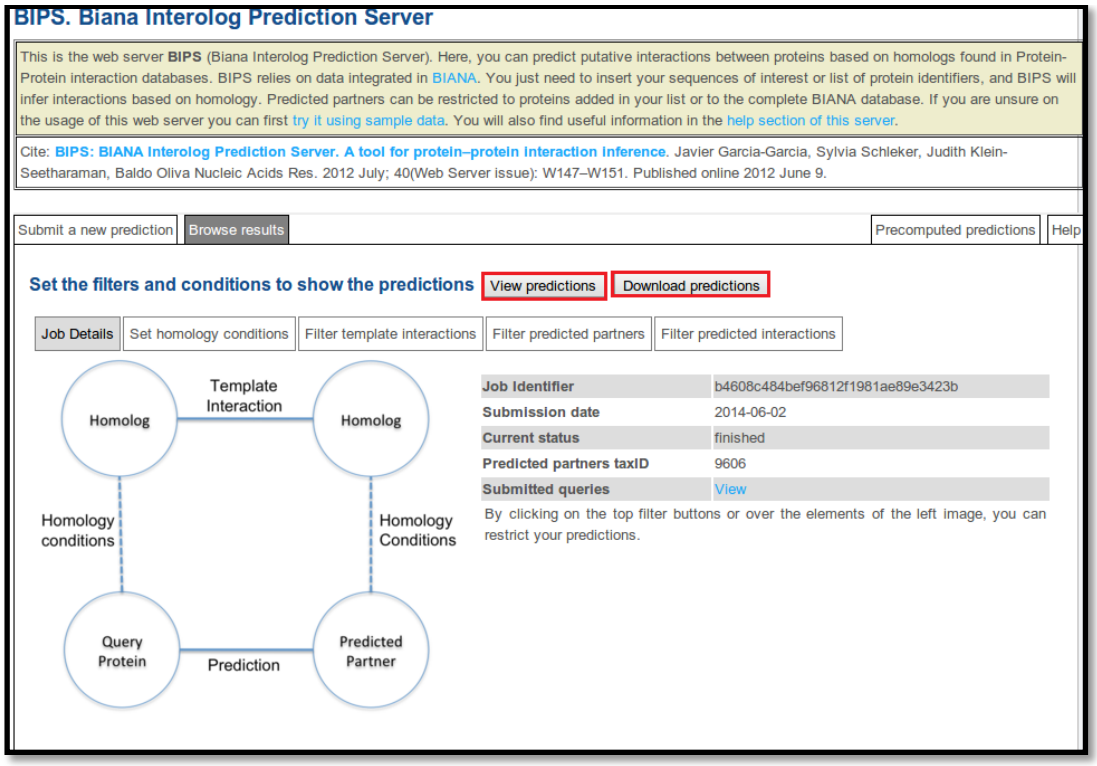
4.2.2 The whole proteome dataset was entered into BIPS in the text box. The predictions can be restricted by setting the taxonomy ID of the partner species. Click on "Submit".



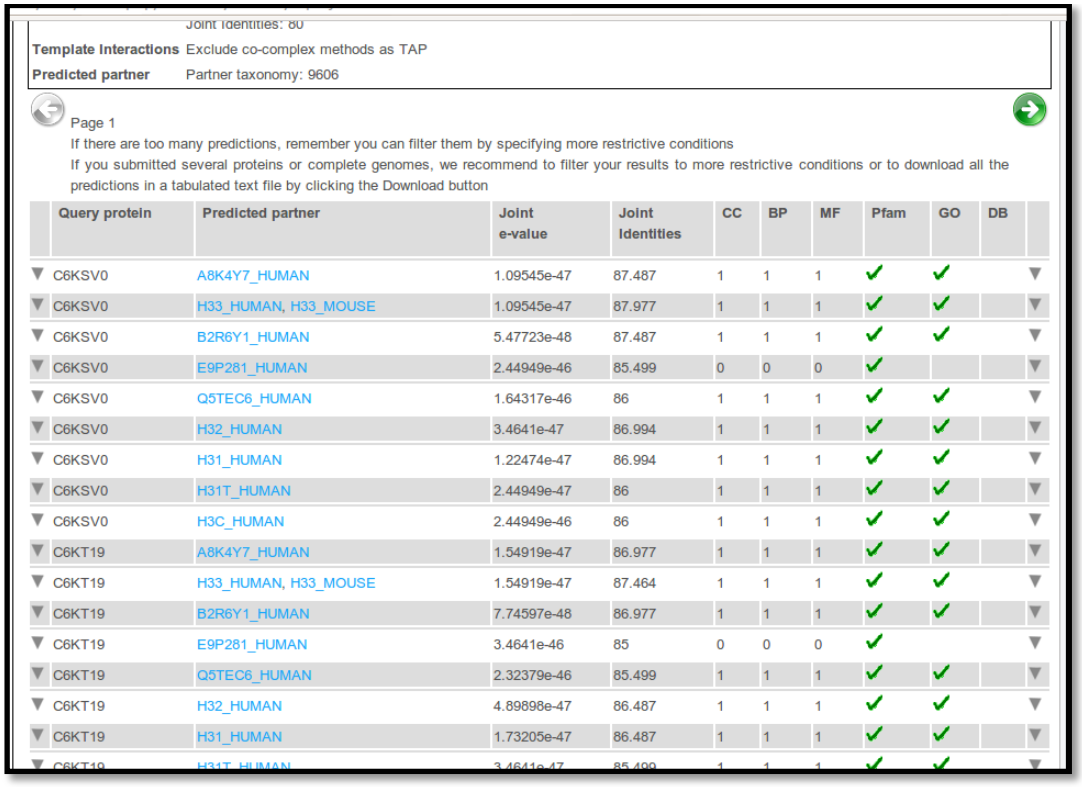
4.2.3 Clicking on submit button will redirect the server web-page to the result page where the link for the prediction result is provided.



4.2.4 Clicking on the result link will redirect the web-page where you can view and download predictions.



4.2.5 Clicking on "View Predictions" will redirect the page to the prediction page.



Query protein	Predicted partner	Joint e-value	Joint identities	CC	BP	MF	Pfam	GO	DB
C6KSV0	A8K4Y7_HUMAN	1.09545e-47	87.487	1	1	1	✓	✓	
C6KSV0	H33_HUMAN, H33_MOUSE	1.09545e-47	87.977	1	1	1	✓	✓	
C6KSV0	B2R6Y1_HUMAN	5.47723e-48	87.487	1	1	1	✓	✓	
C6KSV0	E9P281_HUMAN	2.44949e-46	85.499	0	0	0	✓		
C6KSV0	Q5TEC6_HUMAN	1.64317e-46	86	1	1	1	✓	✓	
C6KSV0	H32_HUMAN	3.4641e-47	86.994	1	1	1	✓	✓	
C6KSV0	H31_HUMAN	1.22474e-47	86.994	1	1	1	✓	✓	
C6KSV0	H31T_HUMAN	2.44949e-47	86	1	1	1	✓	✓	
C6KSV0	H3C_HUMAN	2.44949e-46	86	1	1	1	✓	✓	
C6KT19	A8K4Y7_HUMAN	1.54919e-47	86.977	1	1	1	✓	✓	
C6KT19	H33_HUMAN, H33_MOUSE	1.54919e-47	87.464	1	1	1	✓	✓	
C6KT19	B2R6Y1_HUMAN	7.74597e-48	86.977	1	1	1	✓	✓	
C6KT19	E9P281_HUMAN	3.4641e-46	85	0	0	0	✓		
C6KT19	Q5TEC6_HUMAN	2.32379e-46	85.499	1	1	1	✓	✓	
C6KT19	H32_HUMAN	4.89898e-47	86.487	1	1	1	✓	✓	
C6KT19	H31_HUMAN	1.73205e-47	86.487	1	1	1	✓	✓	
C6KT19	H31T_HUMAN	3.4641e-47	85.499	1	1	1	✓	✓	

4.3 Annotation of the prediction

The predictions were then annotated manually for the cellular location, cellular function and the biological process involved.

4.4 Filtering of the BIPS predicted interactions

The annotated interacting partners were then filtered for their biological significance. The proteins pairs with cellular location in which the interaction is not possible were discarded.

4.5 Functional annotation of the filtered interactions

The functional annotation of the predicted proteins was carried out using DAVID (http://david.abcc.ncifcrf.gov/) and Panther (http://www.pantherdb.org/).

4.6 Analysis of interactions and prioritization

Then the pathogen proteins were analysed for the number of interacting host proteins and the proteins with highest interacting partners were listed. The listed pathogen proteins were then studied in literature for their functional significance in the infection process and credibility as a drug target. One protein was finalized as a potential drug target on the basis of literature study. Already available inhibitors were then searched and listed from databases and literature. The listed inhibitors were then analysed for their advantages and limitations.

4.7 Selection of Inhibitor skeleton for target protein

One of the inhibitor was chosen as a final molecule to be used as a skeleton for further docking analysis.

Work Plan of the Methodology- II

The final protein is shortlisted out of all the pathogen proteins in host-pathogen protein interactions and is considered as a potential drug target. The shortlisted protein is an important protein for the pathogen's biological process. Hence, the idea is to inhibit the pathogen protein so that the disease condition can be avoided. The work plan of the project is as follows.

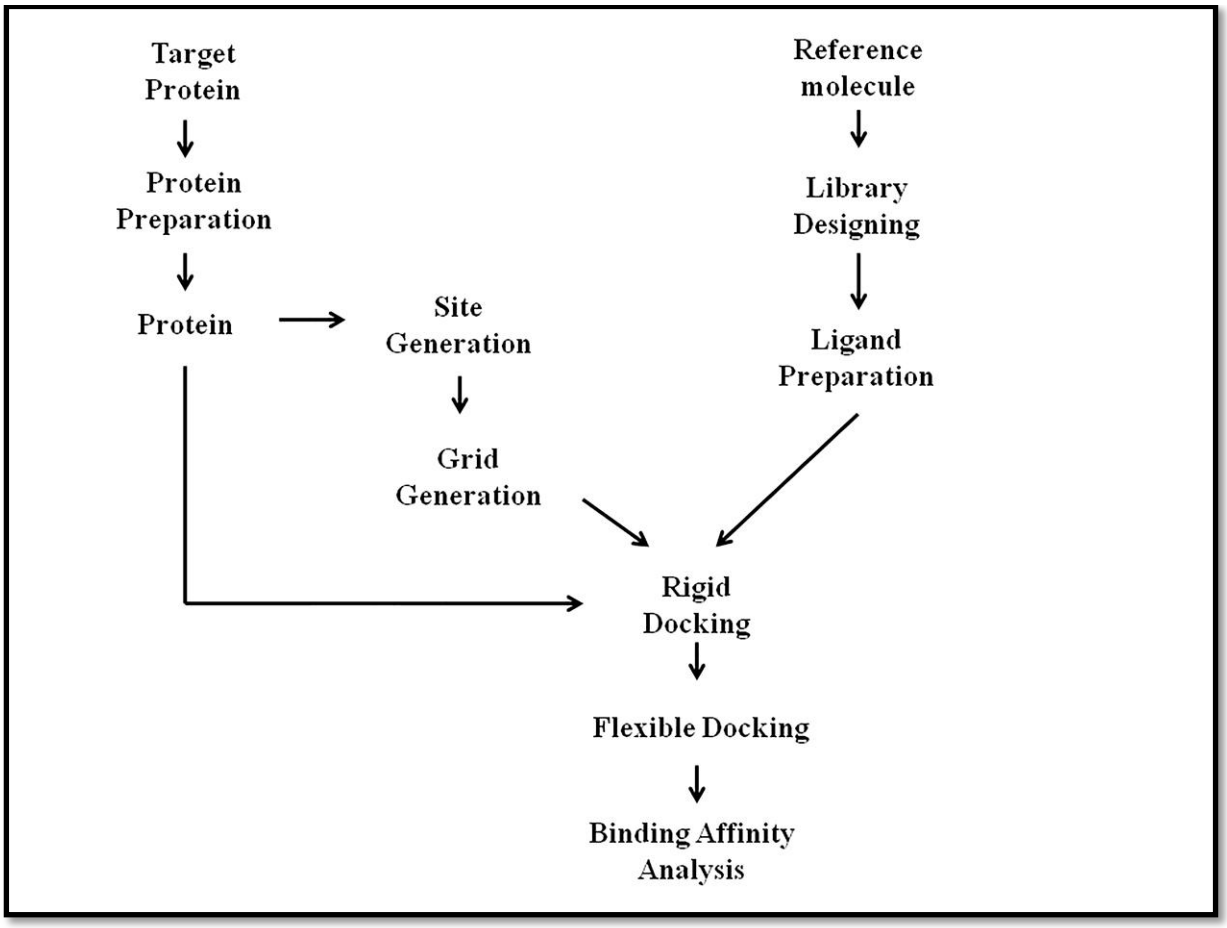
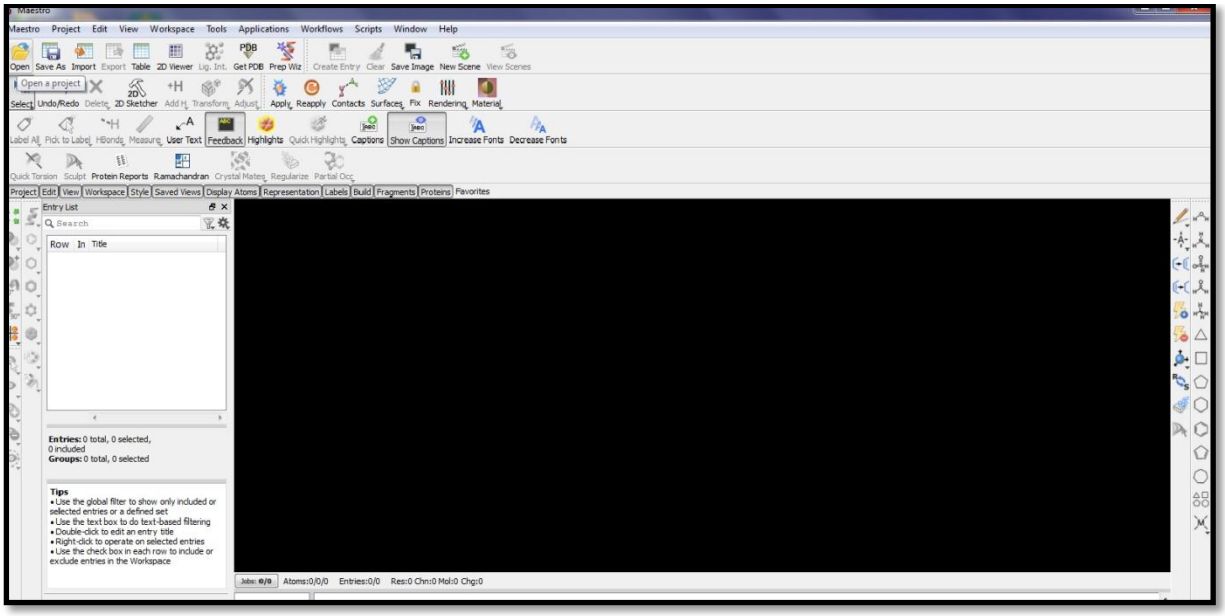


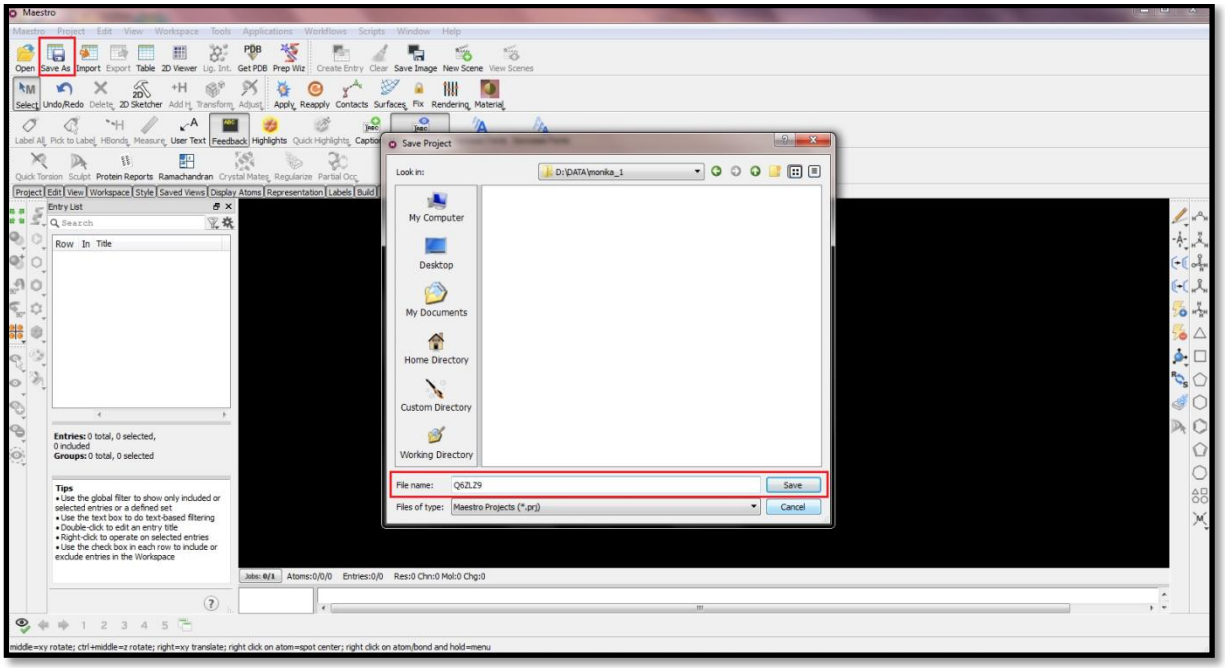
Figure 14: Work plan of the project after short listing the target protein. The target protein is prepared with PrepWiz and sites are predicted with Sitemap of Schrödinger. The grids are generated for all the sites. The inhibitor skeleton is taken and ligand library is designed by carrying out several modifications in it. Then ligand preparation is carried out by Ligprep of Schrödinger. Then prepared protein and ligands are used for rigid docking analysis at all the sites. Then Flexible docking is carried out. The ligand molecules with highest docking score are taken for binding affinity analysis

4.8. Steps involved while working with Schrödinger Software

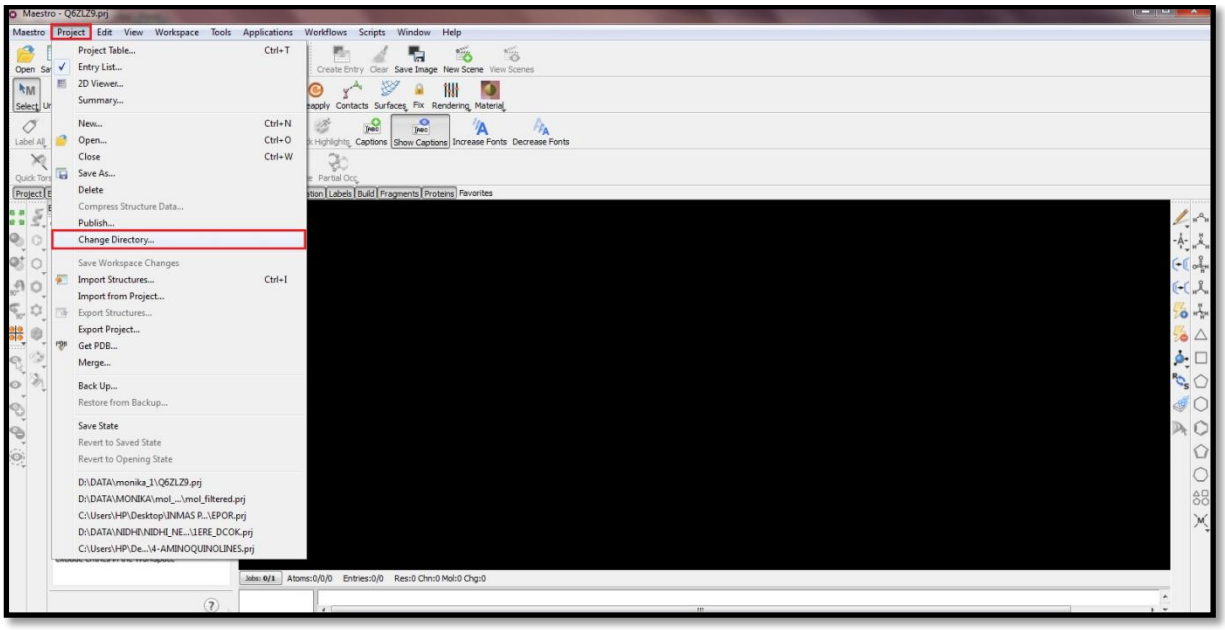
4.8.1 The workspace of Maestro is as shown below.



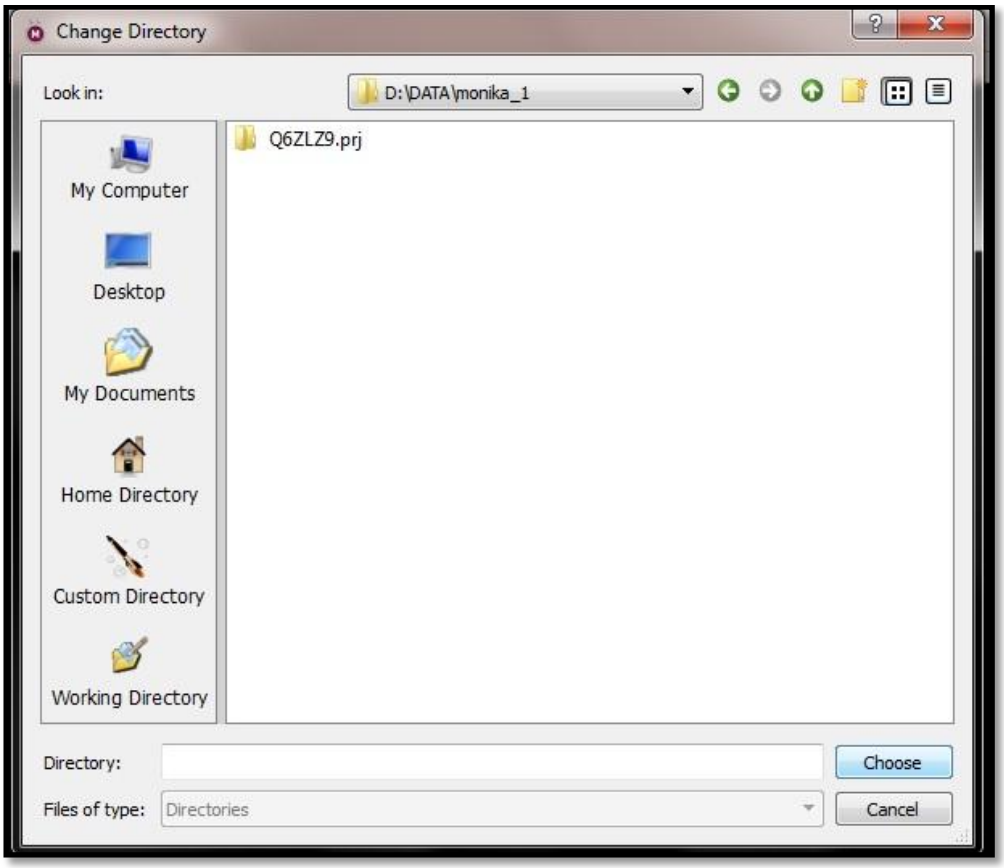
4.8.2 Before we start any activity, we should save a project in the desired folder. Click on "Save as" option to open the window to save the name of your current project.



4.8.3 Then we should Change the working directory by clicking on "Project" Tab and clicking "Change Directory" option.



4.8.4 It opens a small window where you can choose your desired folder to save your work.



4.9 Designing of Library

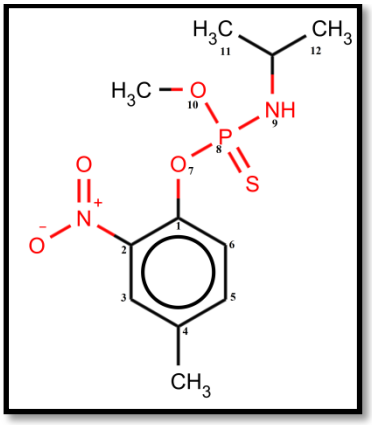


Figure 15: Molecular structure of Amiprophos methyl

Several modifications were carried out in the reference molecule by adding electron withdrawing and donating groups at positions 3, 5, 6,11, 12 and 13.

Type of Functional Groups added

Electron Withdrawing Group
OH, Cl, Br, I, F, CN,COOH, CH_3Cl

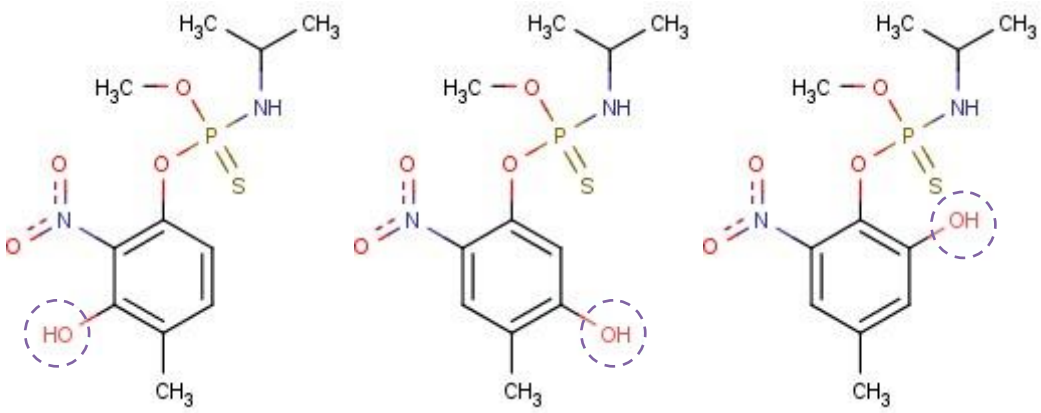
C_2H_5Cl, NO_2, C_2H_5OH, CH_3OH

Electron Donating Group
CH_3, NH_2, C_2H_5, C_3H_7

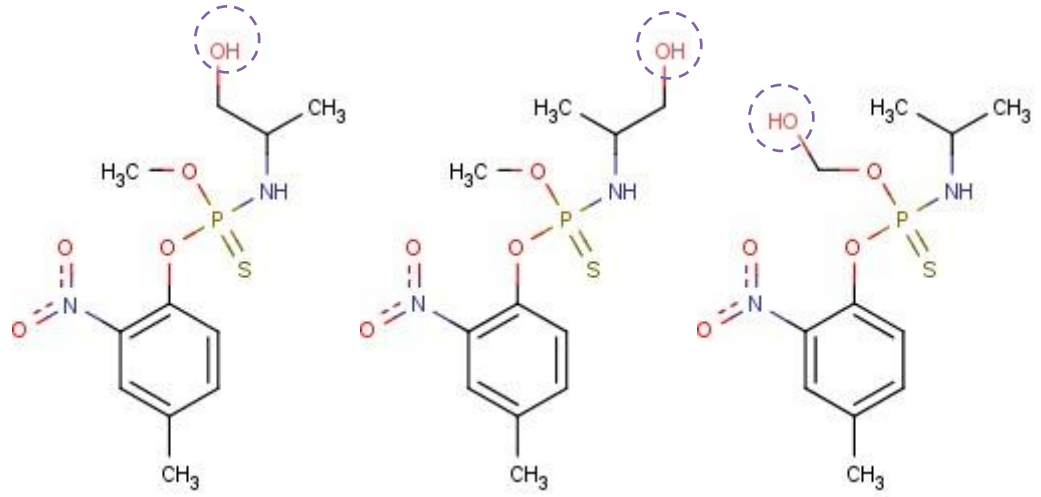
These modifications were carried out at individual sites or more than one site in different combination of functional groups. For e.g. Modification at one site per molecule by OH group is as follows. Similarly, molecules were modified for all the functional groups at individual sites.

S.no.	Pos 3	Pos 5	Pos 6	Pos 11	Pos 12	Pos 13
1	OH					
2		OH				
3			OH			
4				OH		
5					OH	
6						OH

Table 1-: Modification carried out at individual position by OH group.



1 **2** **3**



4 **5** **6**

Figure 16: Molecular structure of modified molecules at 6 individual positions by hydrooxyl (Uetz, Giot *et al*.) group

Modifications with two sites per molecules were also done by keeping one site constant with a molecule and adding different groups at other site. As shown below, modification at two sites in a molecule is done by keeping position 6 constant as CH_3 and changing position 3 and 5 with different functional groups.

S.no	Pos 6	Pos 5		S.no	Pos 6	Pos 3
1	CH_3	Cl		5	CH_3	Cl
2	CH_3	Br		6	CH_3	Br
3	CH_3	I		7	CH_3	I
4	CH_3	F		8	CH_3	F

Table 2: Modification carried out simultaneously at 2 position by keeping constant group at one position and varying group at other position

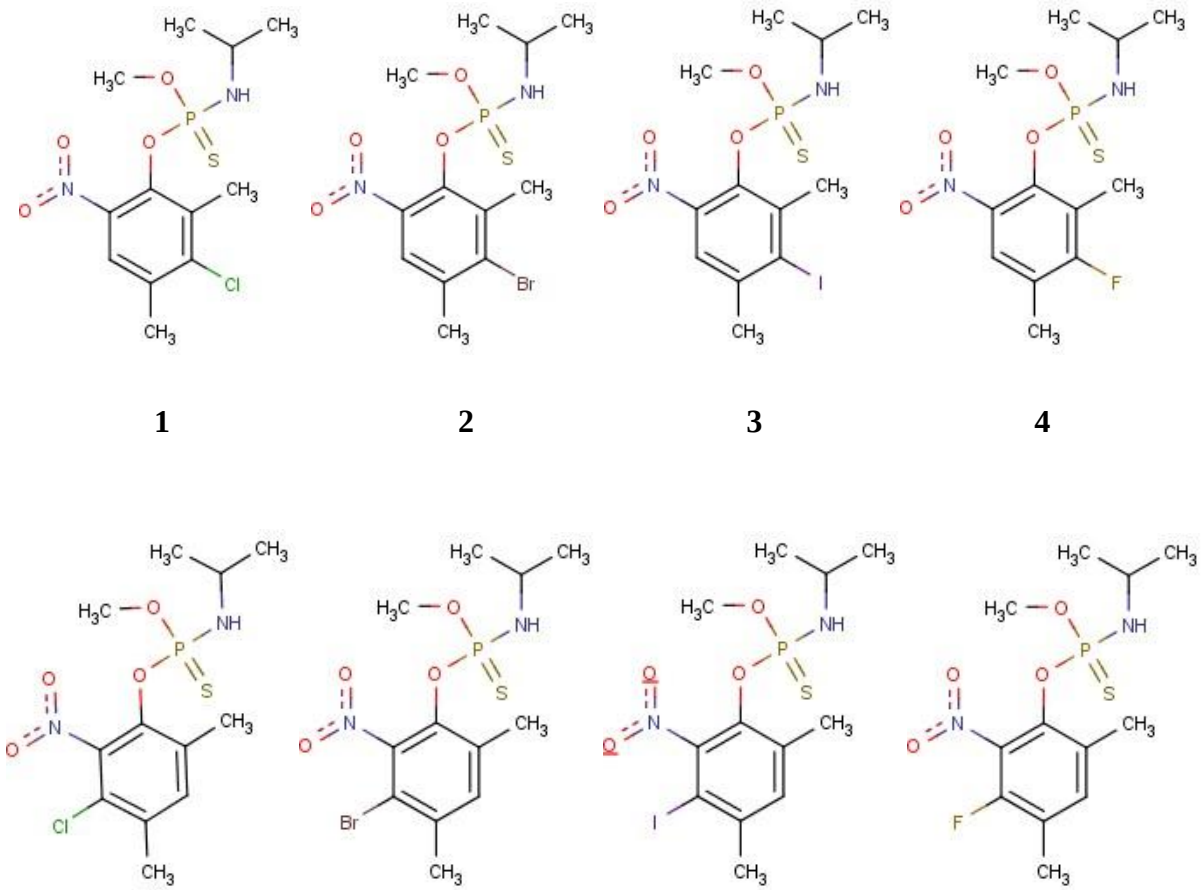


Figure 17: Molecular structure of modified molecules at 2 sites by keeping position 5 constant as CH_3 and changing groups at position 5 and 6

Modifications at simultaneously 3 positions was carried out in an ordered fashion. First two positions were kept constant and one site is changed and then the order for other sites id repeated. i.e. if earlier position 3 and5 were kept constant with CH_3 and position 6 was variable for different functional groups. Then in next round of modification, position 3 and 6 will be constant with CH_3 position 5 will be variable with different functional groups.

S.no.	Pos 3	Pos 5	Pos 6
1	CH_3	CH_3	CH_3
2	CH_3	CH_3	NH_2
3	CH_3	CH_3	OH
4	CH_3	CH_3	NO_2
5	CH_3	CH_3	Cl
6	CH_3	CH_3	Br
7	CH_3	CH_3	I
8	CH_3	CH_3	F
9	CH_3	CH_3	COOH
10	CH_3	CH_3	CN
11	CH_3	NH_2	CH_3

12	CH_3	OH	CH_3
13	CH_3	NO_2	CH_3
14	CH_3	Cl	CH_3
15	CH_3	Br	CH_3
16	CH_3	I	CH_3
17	CH_3	F	CH_3
18	CH_3	COOH	CH_3
19	CH_3	CN	CH_3
20	NH_2	CH_3	CH_3
21	OH	CH_3	CH_3
22	NO_2	CH_3	CH_3
23	Cl	CH_3	CH_3
24	Br	CH_3	CH_3
25	I	CH_3	CH_3
26	F	CH_3	CH_3
27	COOH	CH_3	CH_3
28	CN	CH_3	CH_3

Table 3: Modifications at 3 positions

A library of total 302 molecules was prepared by doing above mentioned modifications in the reference skeleton by using Marvin sketch. Now we have to prepared ligands from this library. We used Ligprep from Schrödinger to generate different conformers as ligands.

4.10. Ligand Preparation

LigPrep is a rich collection of tools designed to prepare high quality, all-atom 3D structures for large numbers of drug-like molecules, starting with 2D or 3D structures in SD or Maestro format. The resulting structures can be saved in either SD or Maestro format. The LigPrep produces a single, low-energy, 3D structure with correct chiralities for each successfully processed input structure. LigPrep can also produce a number of structures from each input structure with various ionization states, tautomers, stereo-chemistries, and ring conformations, and eliminate molecules using various criteria including molecular weight or specified numbers and types of functional groups present.

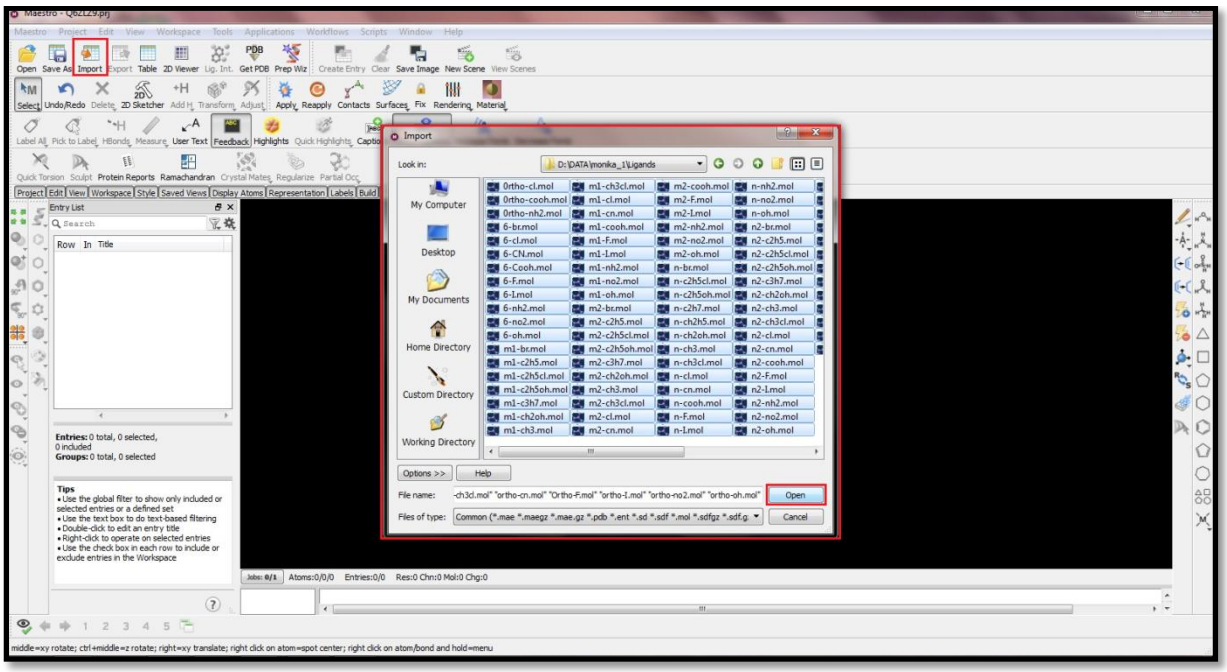
The ligand preparation involves the following tasks:

1. Addition of hydrogen atoms.
2. Filtering out unsuitable molecules based on their properties.
3. Removal of unwanted molecules, such as water, small ions.
4. Neutralization charged groups, then generation of ionization and tautomeric states with Epik
5. Generation of stereoisomer, particularly if stereo-chemical information is missing.
6. Generation of low-energy ring conformations.
7. Removal of any badly prepared structures.
8. Optimization of the geometries.

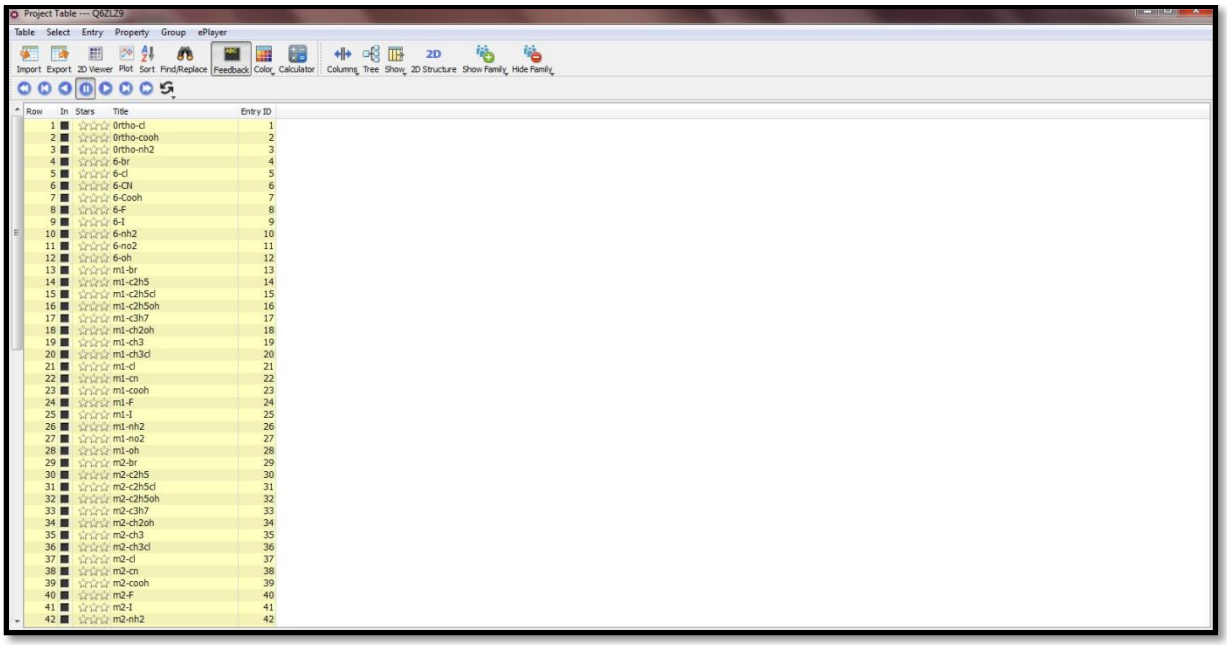
All the structures were imported into the project by using import option in Maestro.

LigPrep was chosen from "Applications tab". The Source of the structures was set as Project table. The output file format was chosen as Maestro and job was started. As a Result, Ligprep generated 751 conformers.

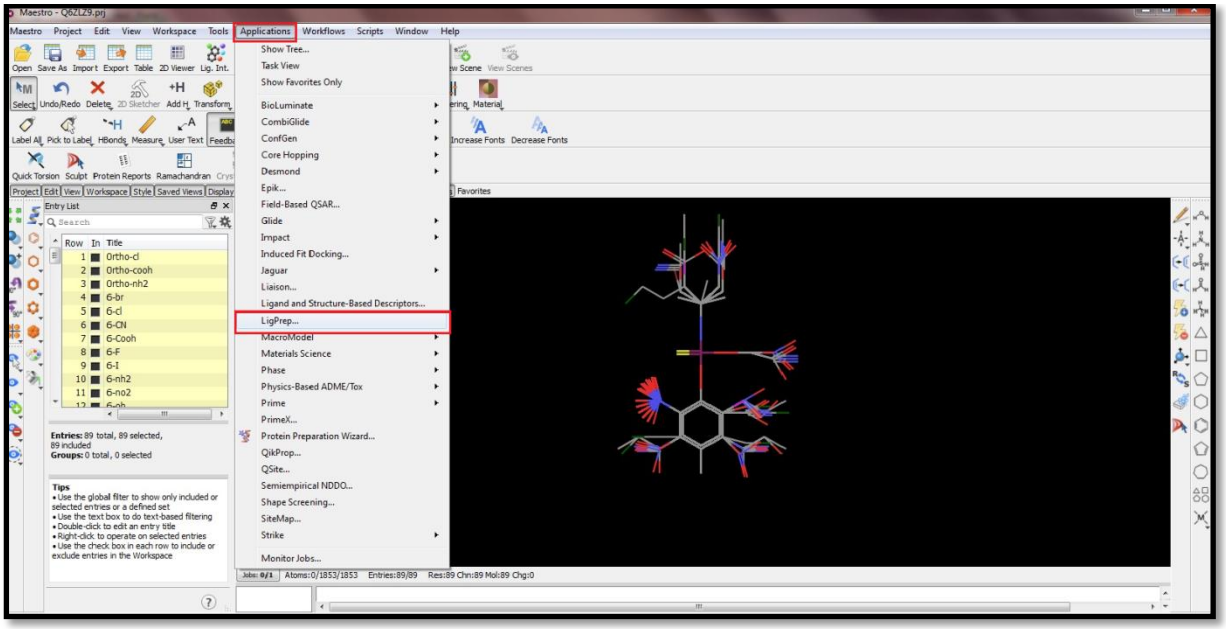
4.10.1. LigPrep takes the chemical structures in .mol format, so we have to import the chemical structures in .mol format. Click on "Import", select all the structures to import and Click "Open".



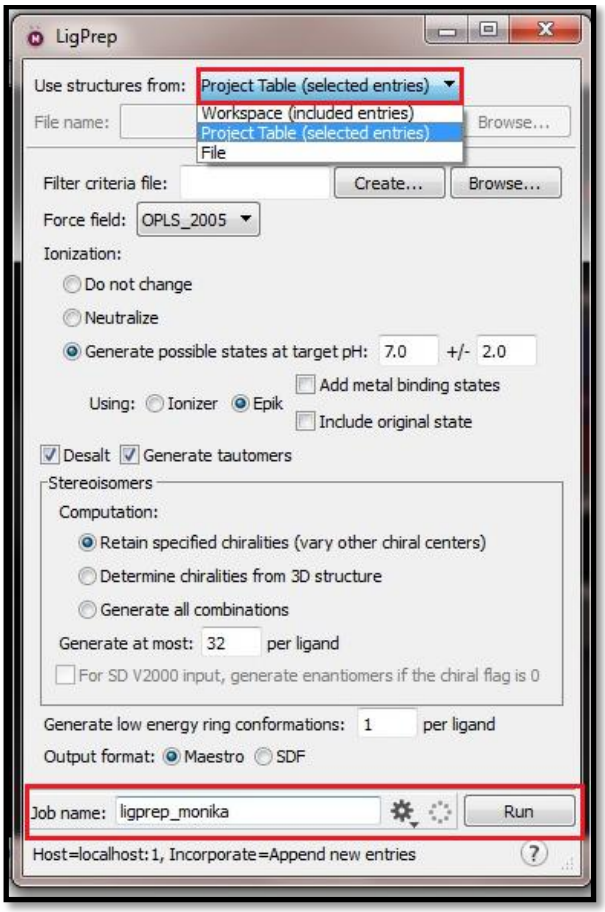
4.10.2. Select all the molecules in Project table.



4.10.3 Click on " Applications" tab and select "Ligprep".



4.10.4 Clicking on "Ligprep" opens a small window where you can set desired parameters and Click on "Run".



4.11. Protein Preparation

Protein preparation was carried out using PrepWiz in Maestro.

Protein preparation involves 3 steps.

1. Pre-process

2. Review and Modify

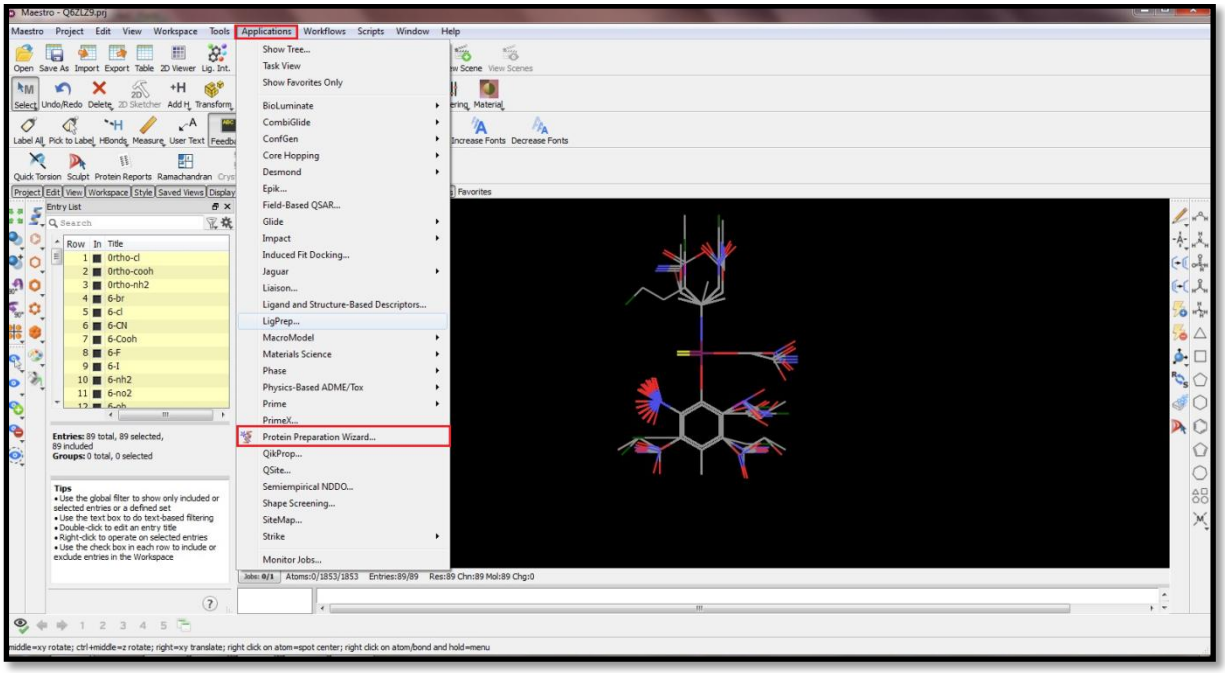
3. Refinement- Optimize and Minimization.

The protein was imported into the project. Protein Preparation Wizard was opened from the "Applications" tab or Tasks tab or with the toolbar button. The protein was pre-processed to remove all the water molecules.

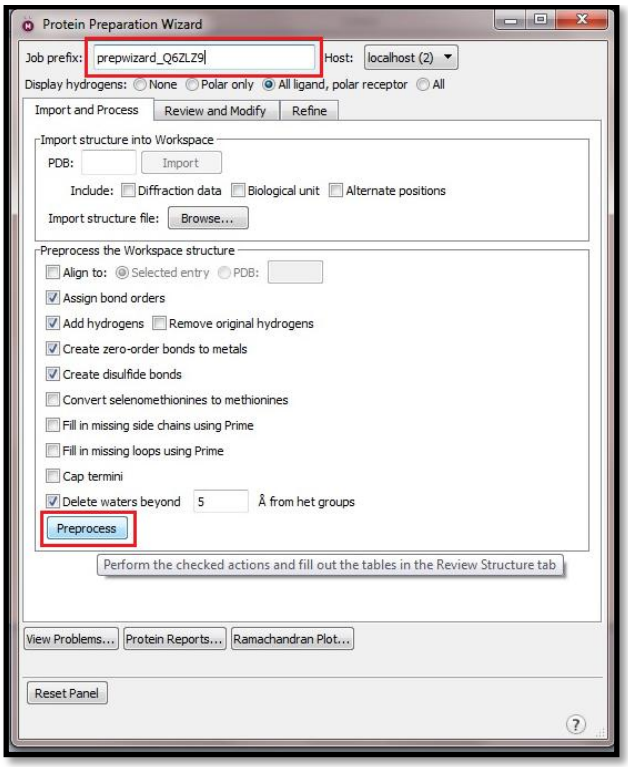
Then pre-processed protein is reviewed to delete part of protein that we do not want to use for modelling. In our protein, we did not have any such case.

In the last step of refinement, H-bond optimization and bond minimization is carried out.

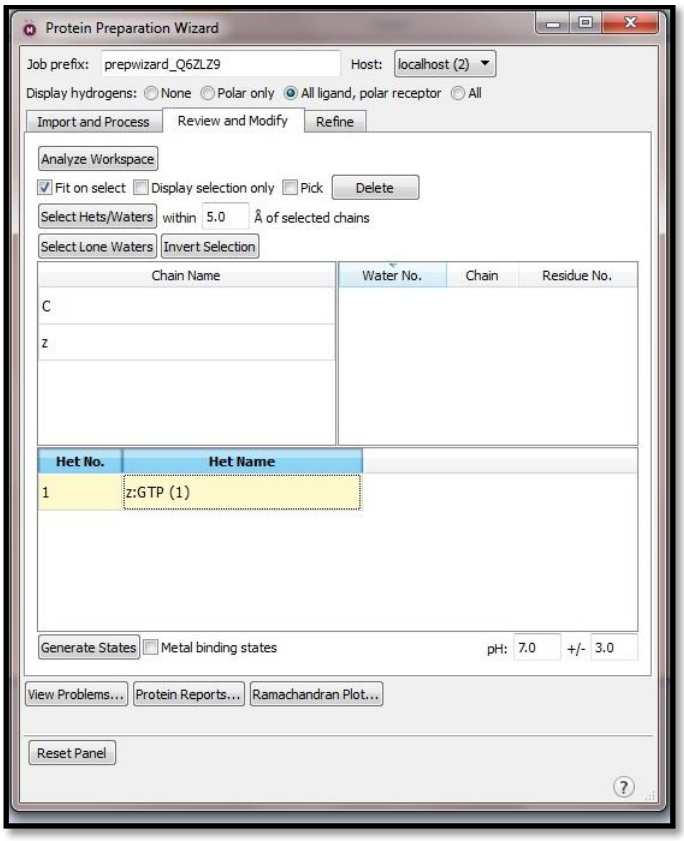
4.11.1 Now to prepare protein, we use Protein preparation wizard. First of all, Import protein PDB file and select it on workspace. Click on "Applications" and select "Protein Preparation Wizard".



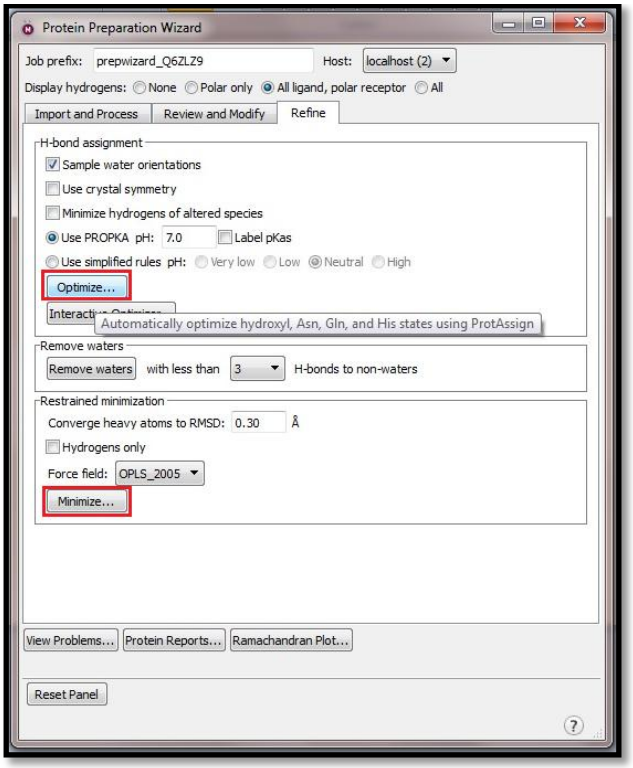
4.11.2 It opens a small window of "PrepWiz". Change the Job name and Click on "Preprocess".



4.11.3 Then review and modify the protein to add or remove any ligand or chain if needed. In our protein we did not need any modification. So we can skip this step.

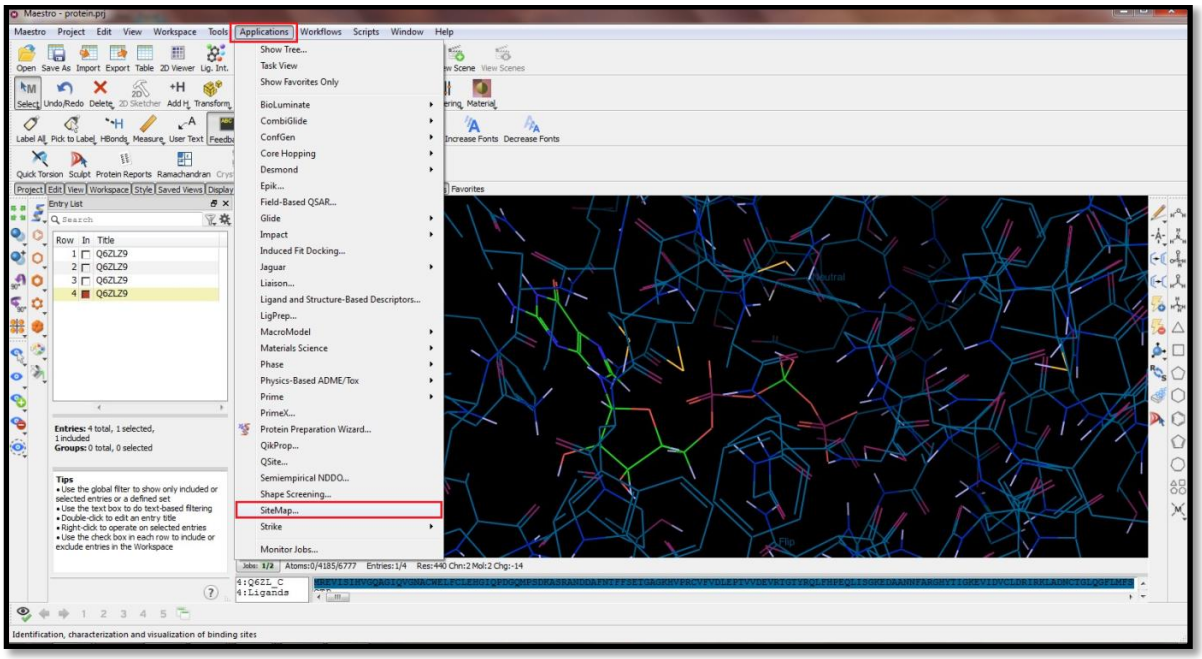


4.11.4 Then next step is Refinement. Click on "Optimize" and then "Minimize".

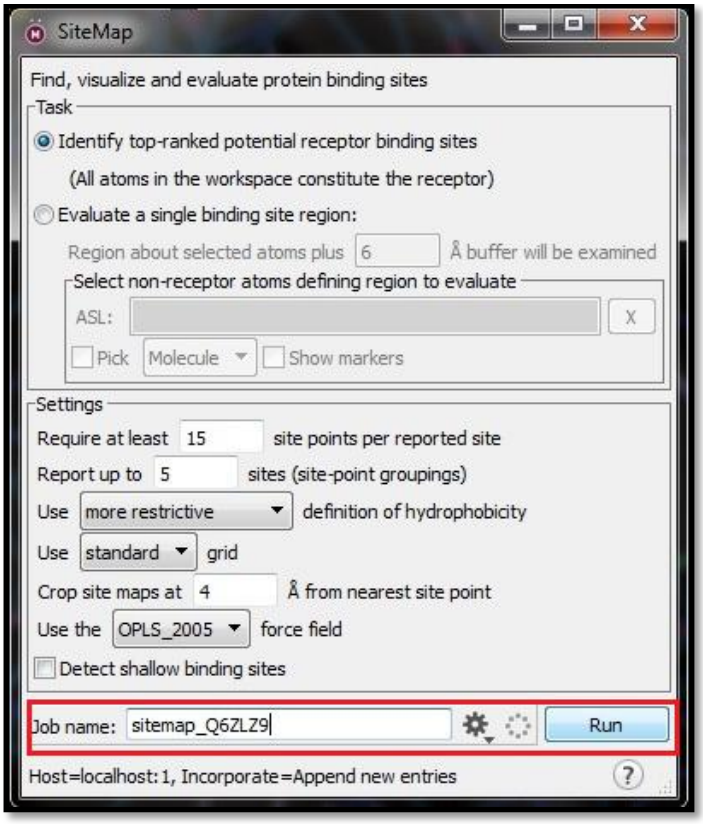


4.12. Site Generation

4.12.1 Next step is to find potential sites in the protein molecule. SiteMap is used to find potential sites in the protein of interest. Click on "Applications" tab and select "Sitemap".



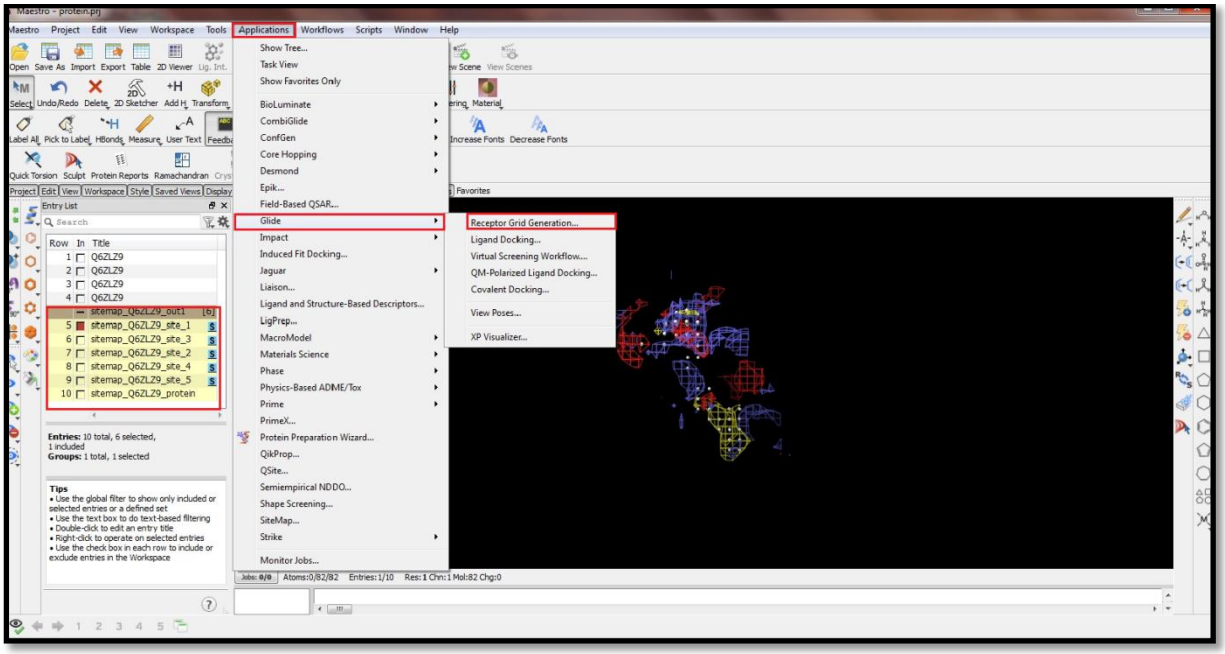
4.12.2 Change the parameters, change the job name and Click "Run".



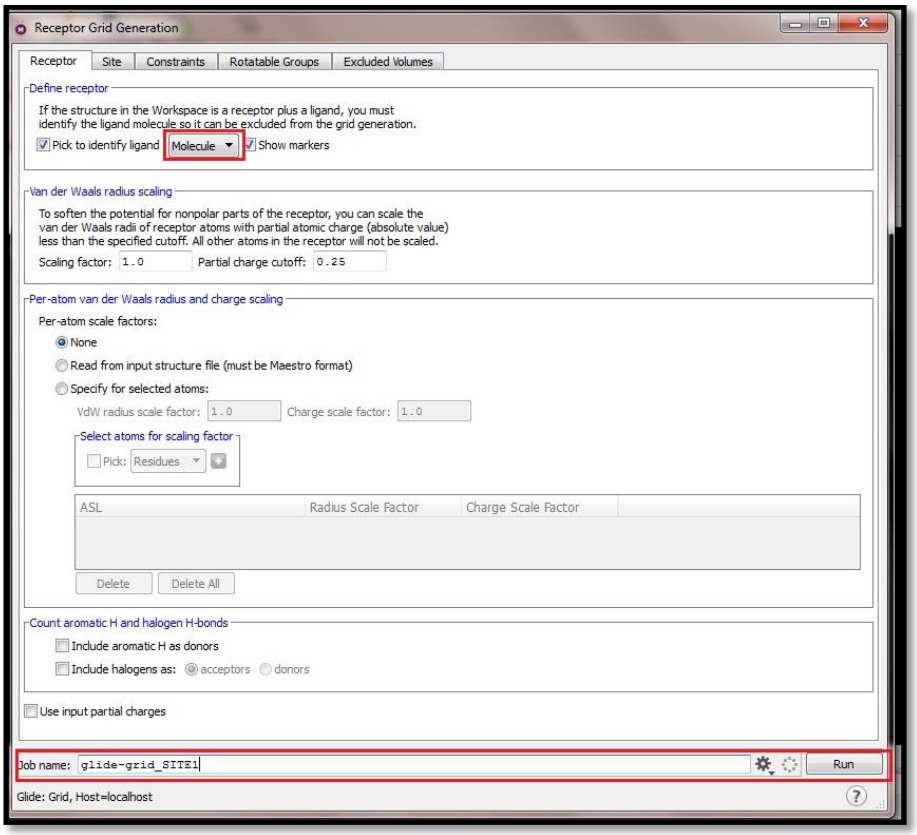
The result generated 5 sites and now Grid is to be generated.

4.13. Grid Generation

4.13.1 From the "Applications" tab click on Glide-> Receptor grid generation.



4.13.2 It opens a new window as shown in the snapshot.

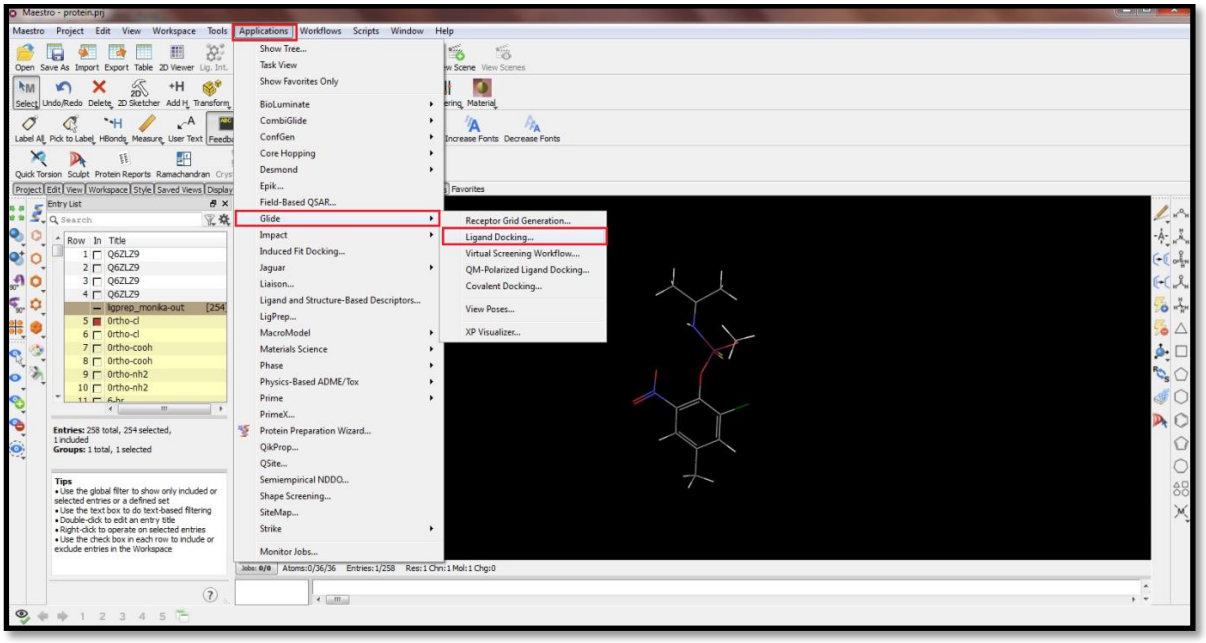


4.14. Ligand Docking

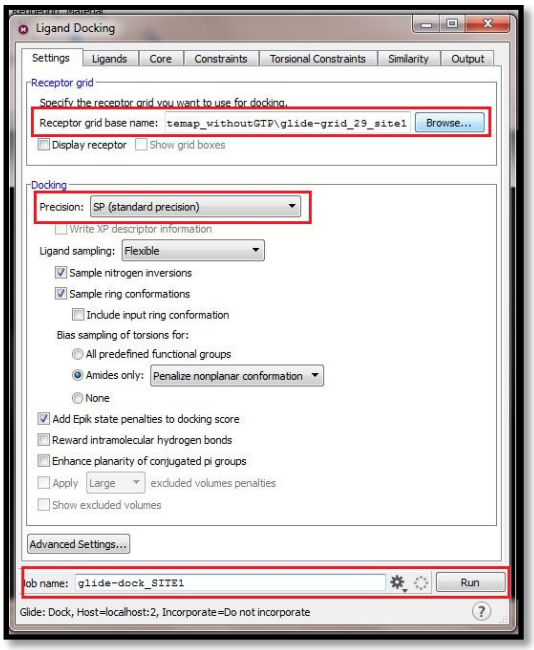
4.14.1 Rigid Docking

The result of protein preparation is used as a target during Ligand Docking.

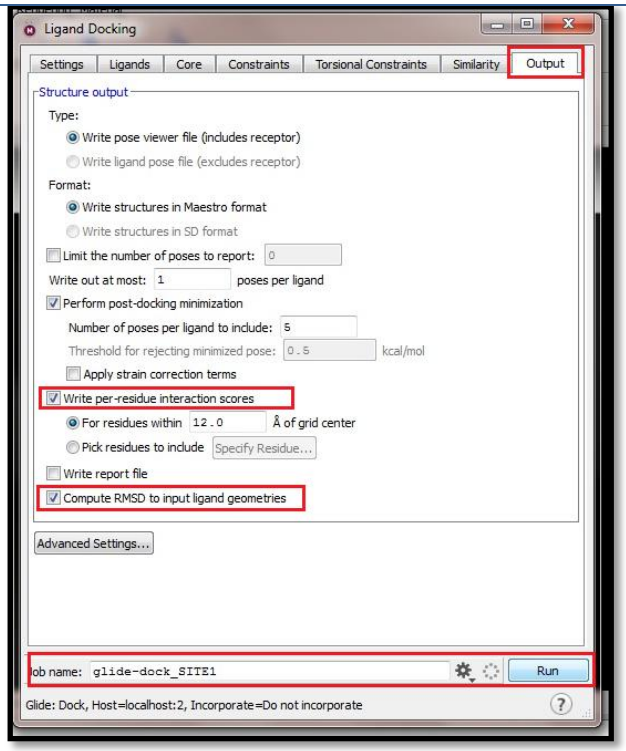
4.14.1.1 For Ligand Docking, Click on "Applications" and select Glide ----> Ligand Docking.



4.14.1.2 This opens a small window of "Ligand Docking" where we have to browse for the Grid of site where Docking is to be performed. Precision is change to XP precision.

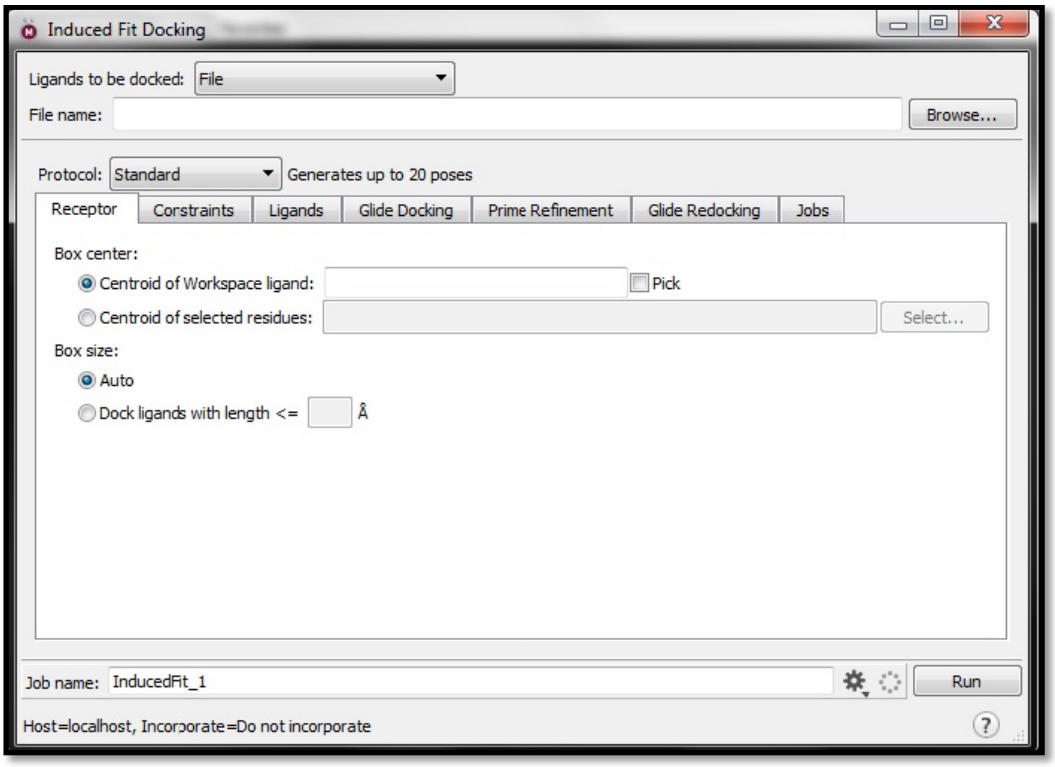


4.14.1.3 Select the output tab and select the output parameters and click "Run".

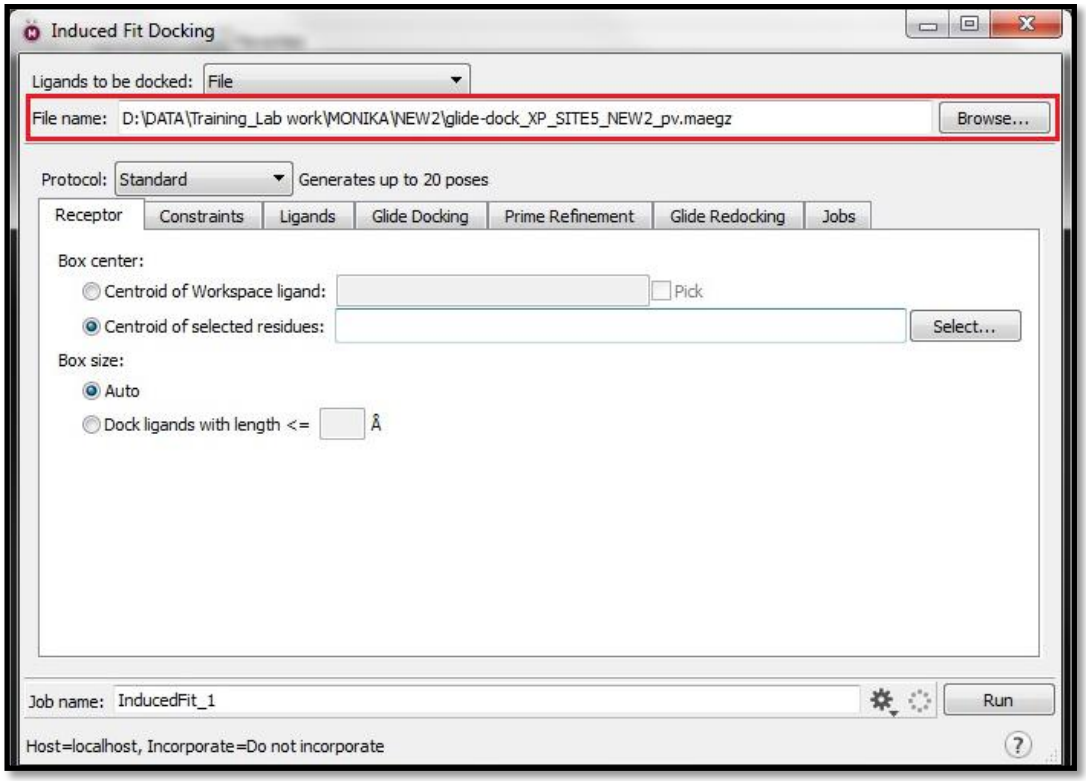


4.14.2 Flexible Docking (Induced Fit Docking)

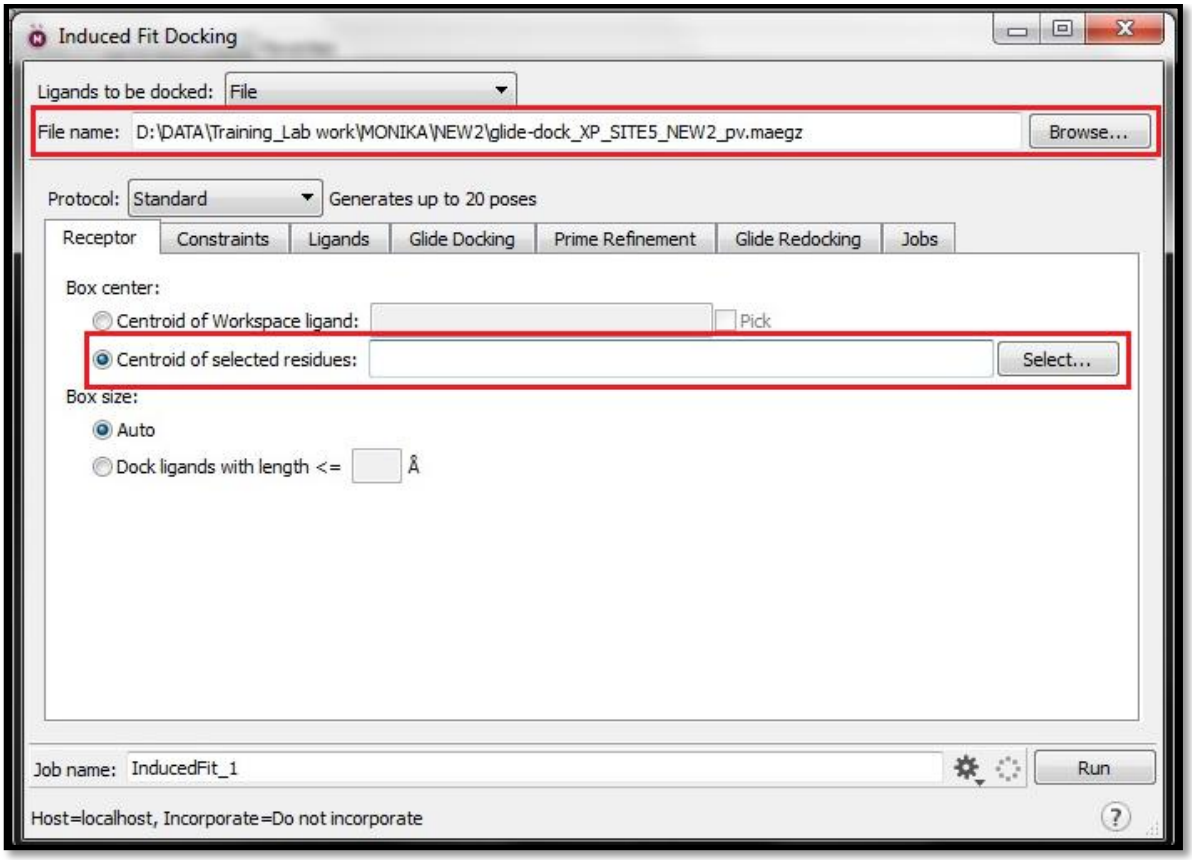
4.14.2.1 For Induced Fit Docking, Click on "Applications" tab and select "Induced Fit Docking".



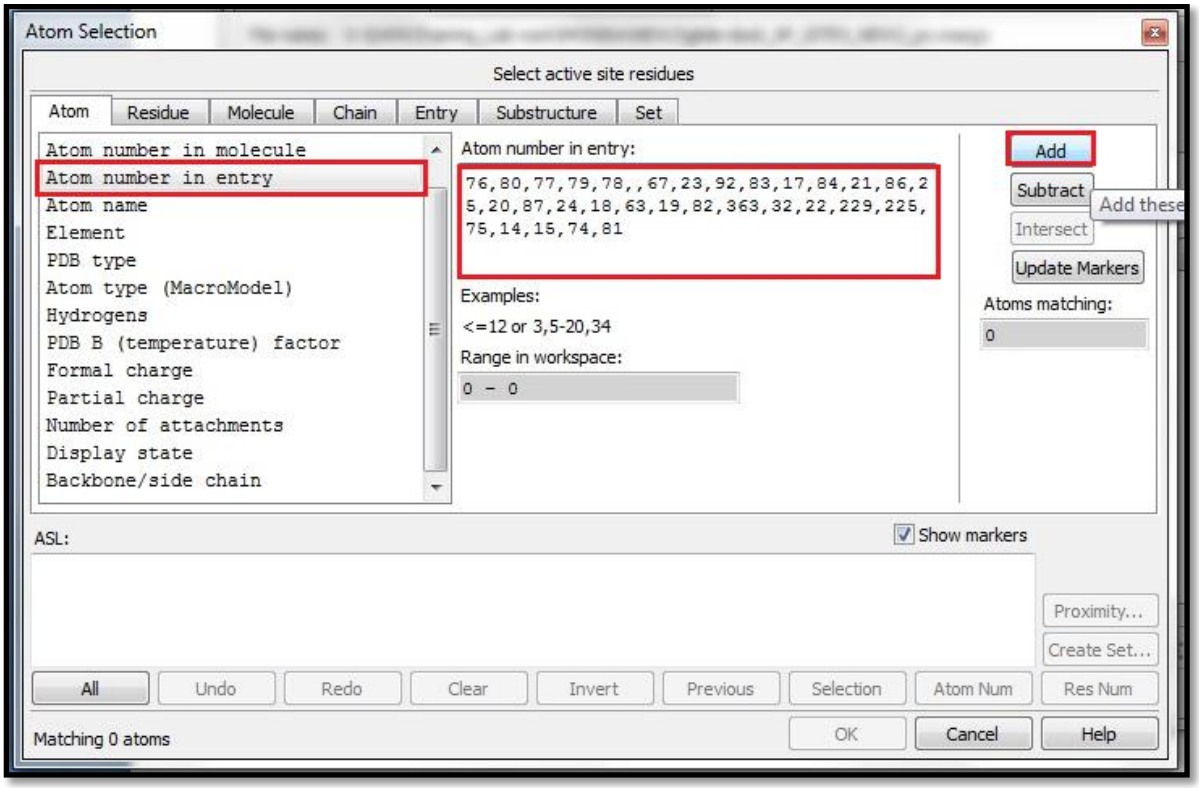
4.14.2.2 Select the molecules one by one from the result of Glide ligand docking.



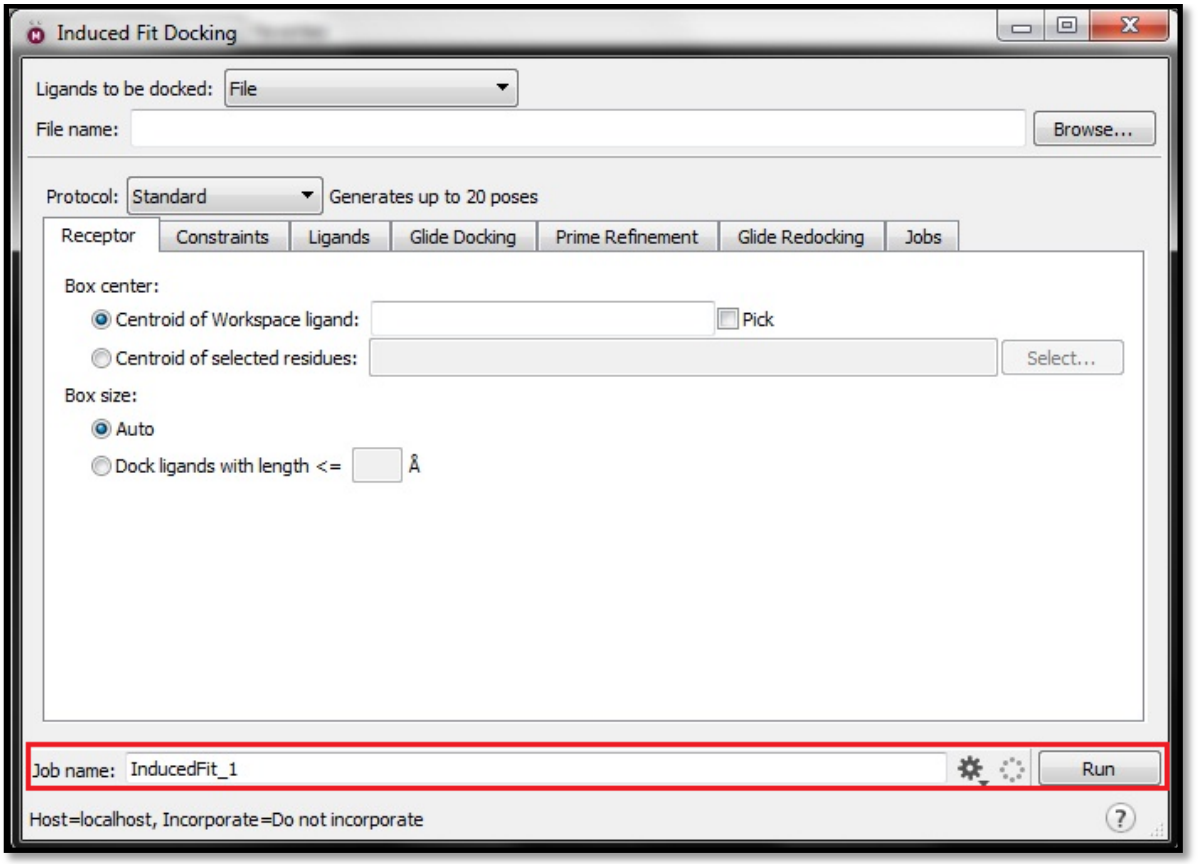
4.14.2.3 Select the residues for Induced Fit Docking.



4.14.2.4 Click on "Run" to perform induced fit docking.

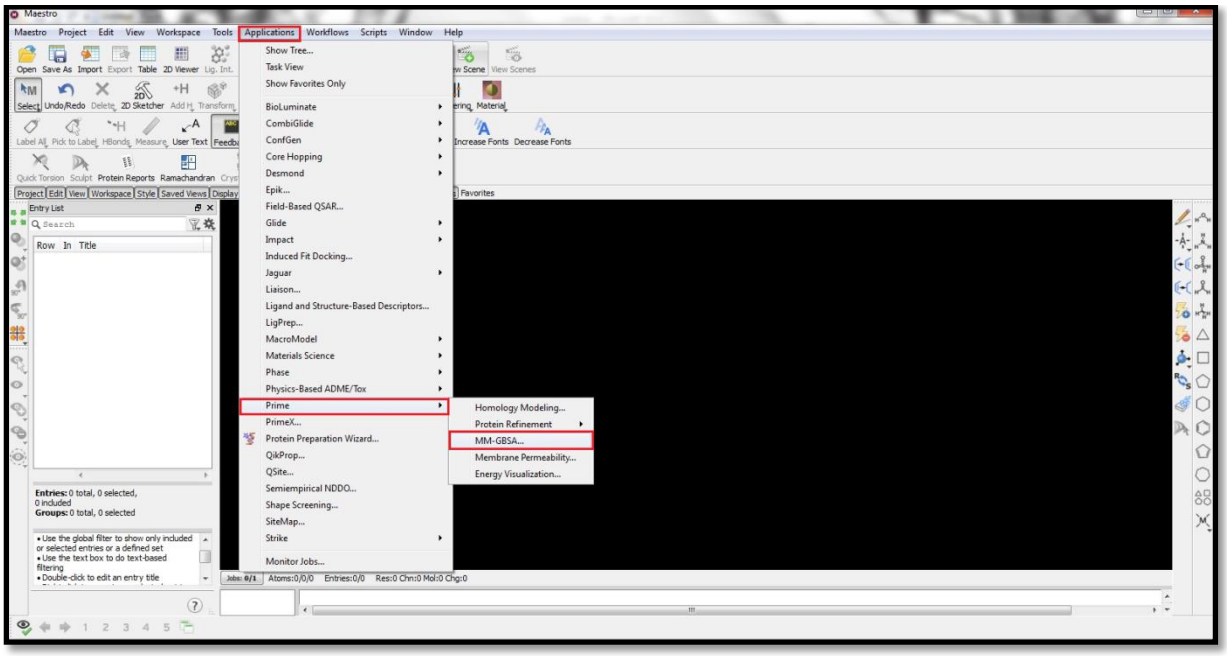


4.14.2.5 Click on "Run" to perform induced fit docking.

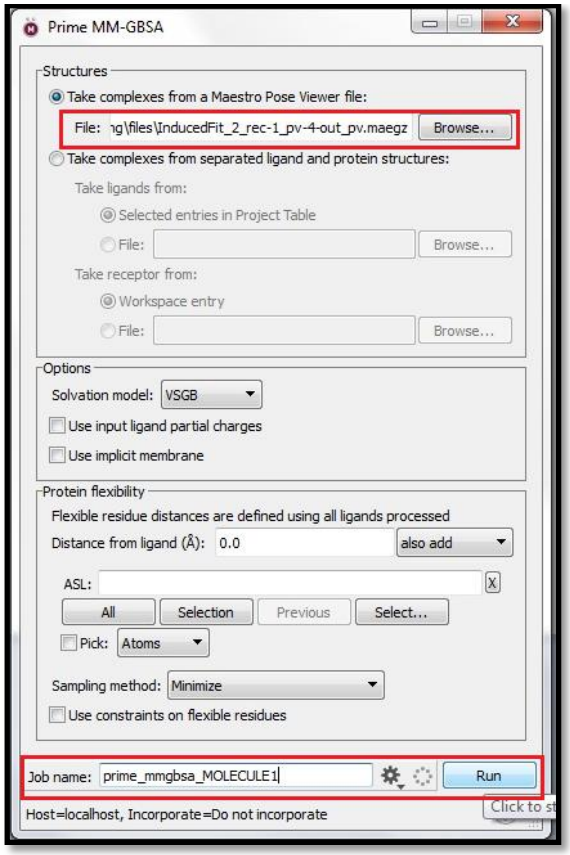


4.15. Binding Affinity Analysis

4.15.1 Now we have to find the Bind Affinities of the molecules. For this purpose we have to use Prime-MM GBSA. Click on "Applications-> Prime-> MM-GBSA".



4.15.2 It opens a small window where we have to select the molecules to be used to find binding energies. Click "Run".



5. RESULTS

5.1 Retrieval of whole proteome

The whole proteome of the pathogen species *Plasmodium falciparum* was downloaded from Uniprot database (http://www.uniprot.org/). It contained a total of 5353 proteins i.e. 157 reviewed and 5196 unreviewed proteins.

5.2 BIPS Prediction

From BIPS (Garcia-Garcia, Schleker *et al*. 2012), a result a total of 2381 inter-species interactions were obtained. BIPS also allows to browse the data related with the predicted partners.

5.3 Annotation of the prediction

The predicted partners were then annotated for their cellular locations, their cellular function and biological process involved.

5.4 Filtering of the BIPS predicted interactions

The annotated interacting partners were then filtered for their biological significance. The proteins pairs with cellular location in which the interaction is not possible were discarded. Filtered predictions were then used for making a protein interaction network in Cytoscape.

5.5 Functional annotation of the filtered interactions

The filtered proteins were then annotated for their functional role. Functional annotation was carried out by DAVID (http://david.abcc.ncifcrf.gov/) and Panther (http://www.pantherdb.org/)

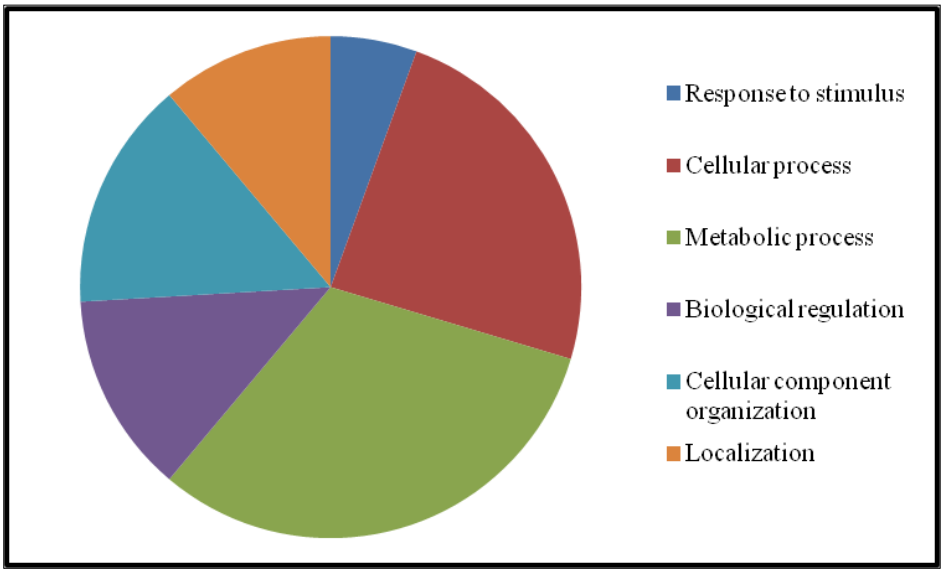


Figure 18: Pie chart representation of pathogen protein classified on the basis of biological process

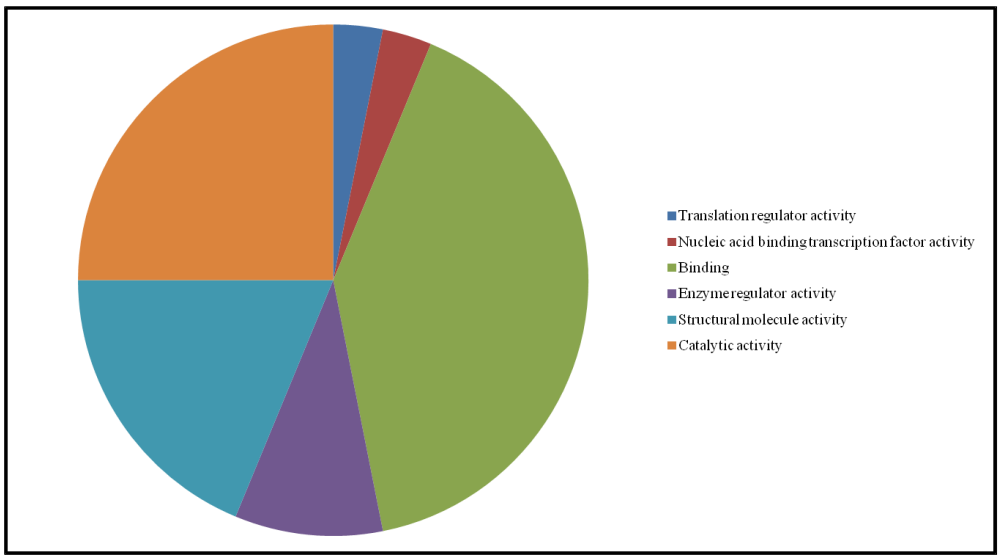


Figure 19: Pie chart representation of pathogen protein classified on the basis of molecular function

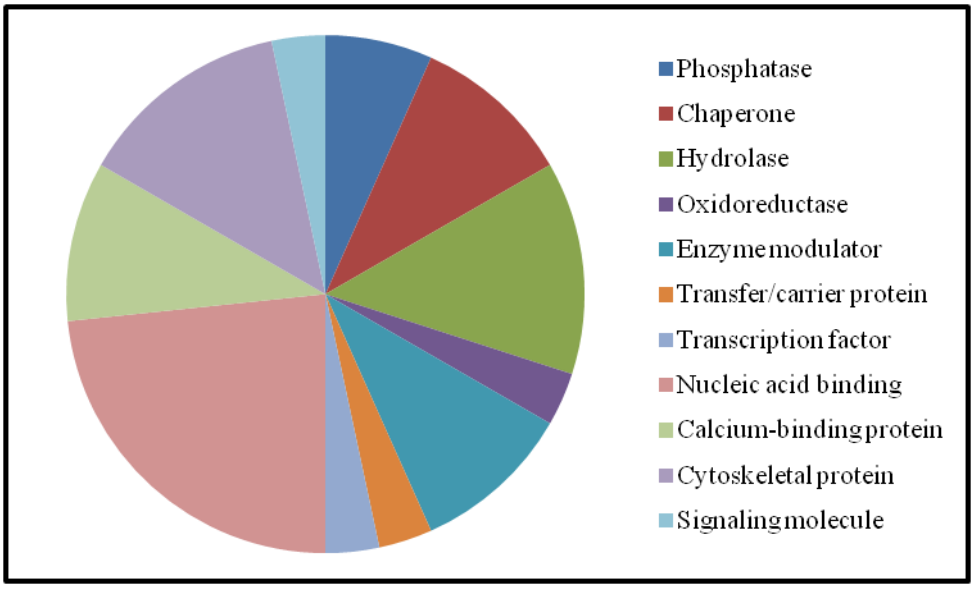


Figure 20: Pie chart representation of pathogen protein classified on the basis of protein class

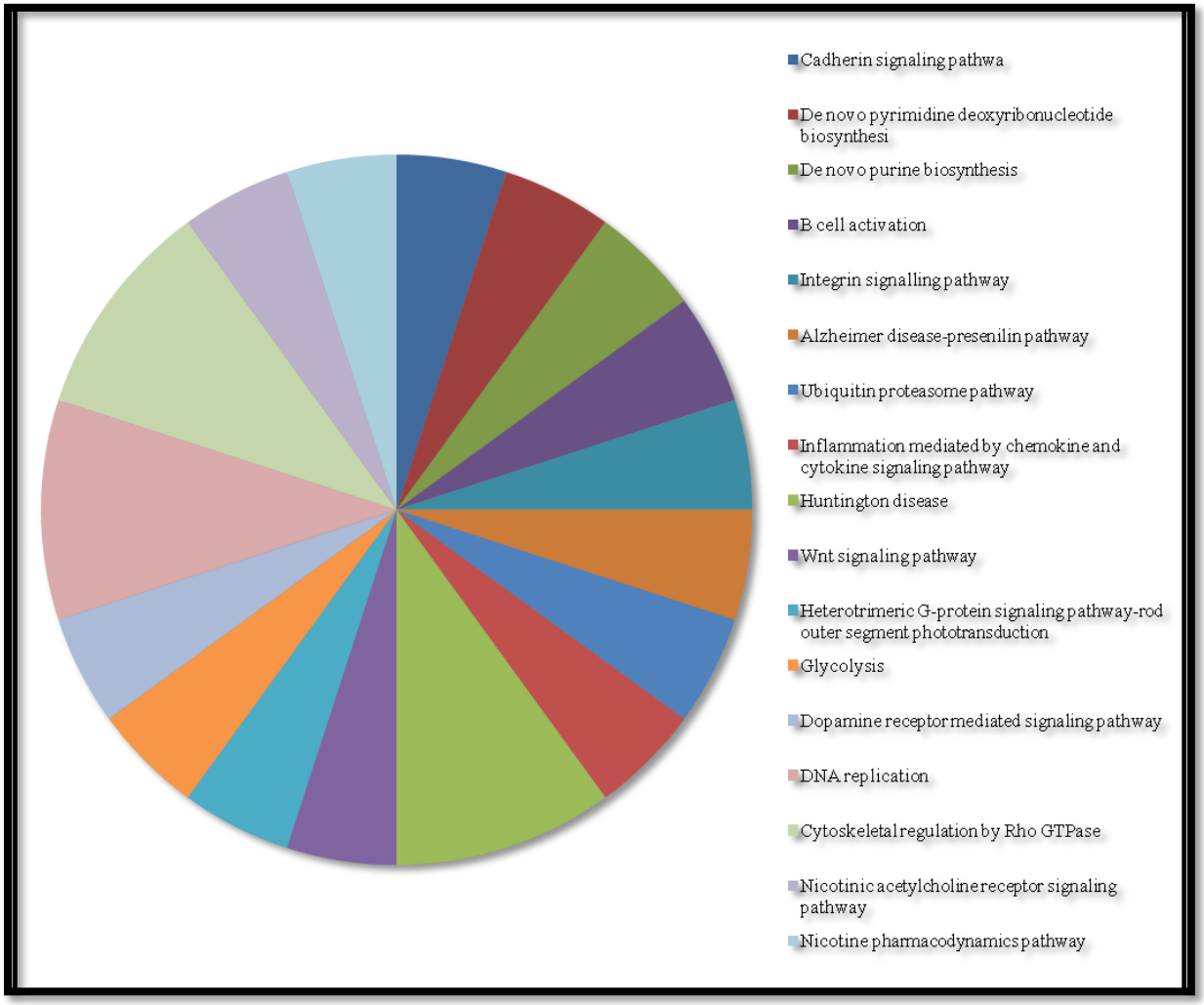


Figure 21: Pie chart representation of pathogen protein classified on the basis of pathway

5.6 Analysis of interactions and prioritization

The predicted interactions were then analysed for the number of interacting partners of pathogen. Out of all the protein interactions, highly interacting proteins were shortlisted. It was observed that most of the highly interacting proteins were those involved in structural assembly of the pathogen such as actin, tubulin and histone.

As a result of this analysis, we finalized α-tubulin as an important protein required for the infection process. It polymerizes to carry out several critically important roles throughout entire parasite life cycle. In parasite, they form mitotic spindle during cell division and even slight disruption of microtubule causes a severe impact on viability of parasite. *Plasmodium falciparum* infects host and initial contact occurs between merozite and erythrocyte. The tip of the merozite is surrounded by a band of microtubules which is responsible during infection. Hence this protein can be taken as a target for inhibitor designing.

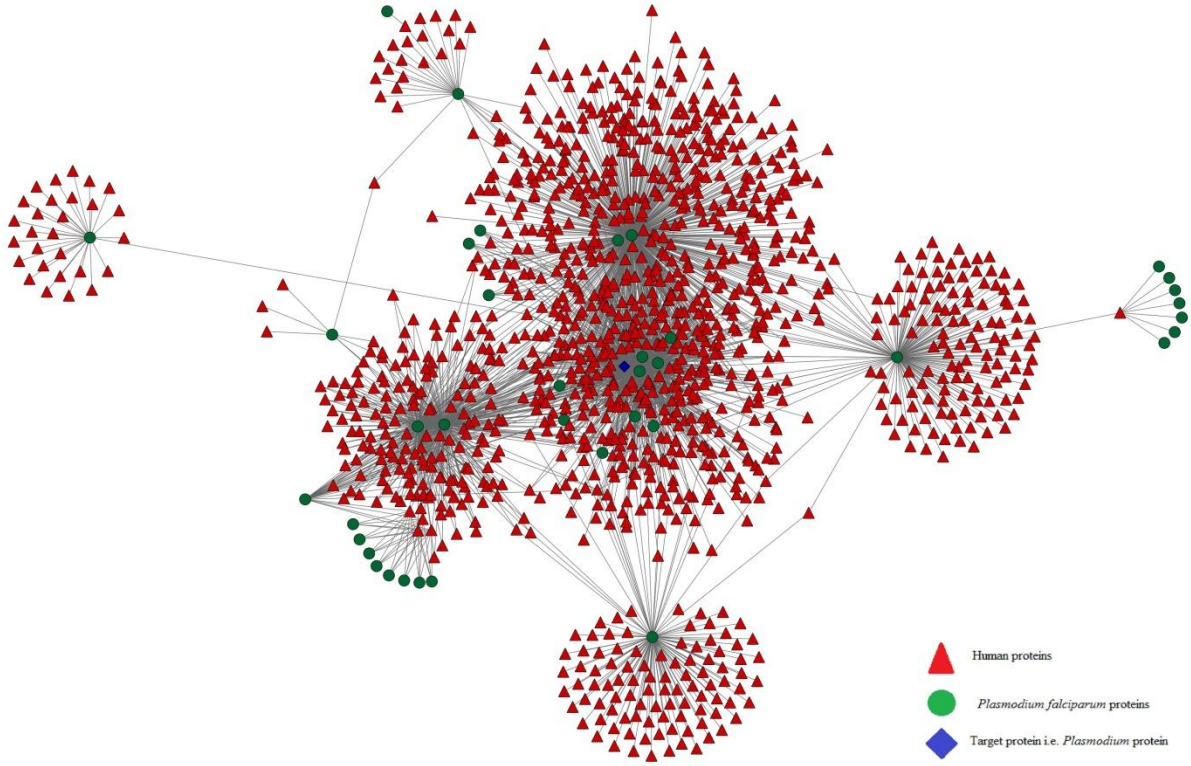


Figure 22: Protein-protein interaction network of Human host and *Plasmodium falciparum.* Different colours and shapes are representing proteins of different species.

5.7 Protein Preparation-:

Protein preparation was carried out by PrepWiz of **Schrödinger (Schrödinger Release 2014-2).** Prepwiz generated a minimized and optimized protein structure that can be used for further processing and analysis.

5.8 Prediction of sitemap site and Receptor Grid Generation-:

Sitemap **(Schrödinger Release 2014-2)** was used to generate prospective sites which gave 5 sites with following parameters as a result.

Site Name	Site Score	Size	Dscore
Site I	0.96	82	0.99
Site II	0.92	78	0.93
Site III	0.89	74	0.88
Site IV	0.72	48	0.71
Site V	0.71	48	0.67

Table 4-: Site score and Druggability score for all the predicted sites

From the site score and Dscore, Site I and Site II were prioritized for docking screening analysis.

Reference molecule: **Amiprophos methyl**

Amiprophos methyl is an anti-mitotic herbicide and is already known inhibitor for α-tubulin. It is found to be a promising molecule because of its low mammalian toxicity. It was reported in ref studies that amiprophos methyl has better specificity for pathogen proteins and has no binding site in human tubulin protein.

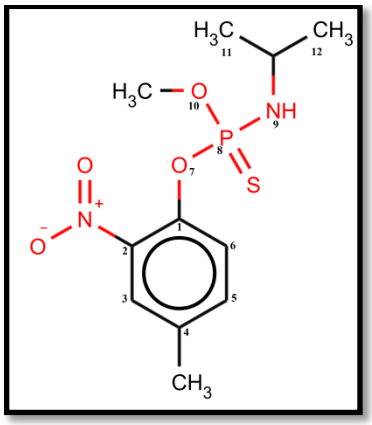


Figure 23: Molecular structure of Amiprophos methyl

5.9 Library Designing-: A total of 302 molecules were prepared by carrying out modification at individual and multiple positions simultaneously by using Marvin sketch **(Schrödinger Release 2014-2)**.

5.10 Ligand Preparation-: LigPrep generated 751 conformers as an output from the library of 302 ligands.

5.11. Ligand Docking- :

5.11.1 Rigid Docking

a) **Docking with reference molecule-: The reference molecule was docked at all the 5 sites.**

Site Name	Docking Score
Site I	-4.26521
Site II	-4.4262
Site III	-3.0137
Site IV	-2.34355
Site V	-4.17417

Table 5-: Docking score for reference molecule at all the predicted sites

According to the site docking score, we prioritized site I, II and V.

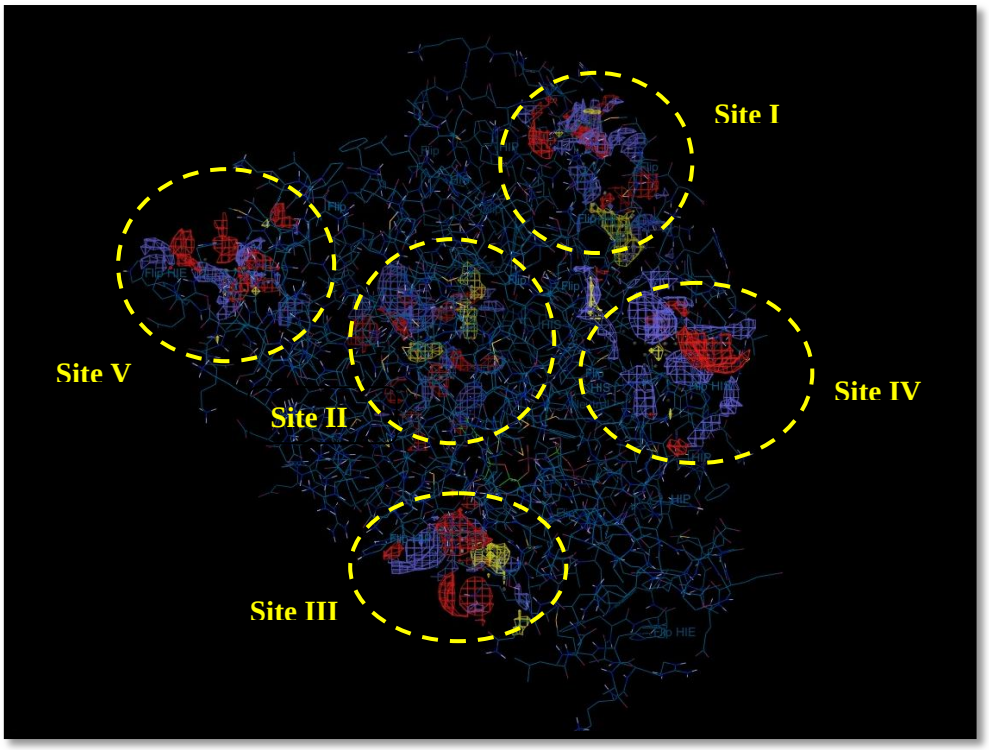


Figure 24: Representation of all the sites predicted are shown in protein structure

5.11.2 Rigid Glide XP docking with Designed ligand molecules-:

Then docking was carried between protein and ligands at all the 5 sites by GLIDE.

Initial Docking Results-: The results of Docking per site is as follows

Site I -:

Pos 2	Pos 3	Pos 4	Pos 5	Pos 6	Pos 8	Pos 9	Pos 10	Pos 11	Pos 12	Glide Score
NO_2	CH_3	CH_3	COOH	OH	S		CH_3			-5.45
NO_2		CH_3		C_2H_5OH	S		CH_3			-5.20
NO_2		CH_3			S		CH_3	C_2H_5OH		-5.12
		CF_3			O	Cyc-Pentane	C_2H_5			-5.08
NO_2	CH_3	CH_3	NO_2	OH	S		CH_3			-5.06
		CF_3			S	Cyc-Pentane	C_2H_5			-5.00
		CF_3			O		C_2H_5			-4.92
NO_2		CH_3			S		CH_3		C_2H_5OH	-4.85
NO_2	CH_3	CH_3	OH	OH	S		CH_3			-4.82
NO_2	OH	CH_3	CH_3	OH	S		CH_3			-4.79
		CF_3			O	Piperidine	C_2H_5			-4.74
NO_2		CH_3		CH_3OH	S		CH_3			-4.71
		CH_3			S		CH_3		OH	-4.70
NO_2	Br	CH_3	OH	CH_3	S		CH_3			-4.66
	CH_3	CH_3	NH_2	OH	S		CH_3			-4.60

Table 6: Glide docking score for the 15 High docking score molecules at site I

Site II -:

Pos 2	Pos 3	Pos 4	Pos 5	Pos 6	Pos 8	Pos 9	Pos 10	Pos 11	Pos 12	Glide Score
NO_2	CH_3OH	CH_3			S		CH_3			-5.85
NO_2		CH_3			S		NH_2			-5.81
NO_2	OH	CH_3	CH_3	OH	S		CH_3			-5.80
NO_2		CH_3			S		CH_3		NH_2	-5.78
NO_2		CH_3			S		CH_3	NH_2		-5.73
NO_2		CH_3			S		CH_3	NO_2		-5.73
NO_2		CH_3			S		CH_3	CH_3OH		-5.53
NO_2		CH_3			S		CH_3	OH		-5.51
NO_2		CH_3			S		CH_3	C_2H_5OH		-5.48
		CF_3			O	C_5H_{11}	C_2H_5			-5.44
NO_2		CH_3	NH_2	CH_3	S		CH_3			-5.41
		CF_3			O	Morpholine	C_2H_5			-5.40
NO_2		CH_3			S		CH_3			-5.35
NO_2	CH_3	CH_3	OH	OH	S		CH_3			-5.33
	CN	CH_3	OH	CH_3	S		CH_3			-5.14

Table 7: Glide docking score for the 15 best docking molecules at site II

Site III -:

Pos 2	Pos 3	Pos 4	Pos 5	Pos 6	Pos 8	Pos 9	Pos 10	Glide Score
		CF_3			O		C_2H_5	-5.42
NO_2		CH_3			S		NH_2	-5.19
		CF_3			O	Piperidine	C_2H_5	-4.97
NO_2	CH_3	CH_3	OH	OH	S		CH_3	-4.79
	Cl	CH_3			O		C_2H_5	-4.76
		Br			O	Cyc-Pentane	C_2H_5	-4.60
NO_2	CH_3	CH_3	CH_3	OH	S			-4.56
		Br			S	Cyc-Pentane	C_2H_5	-4.55
NO_2	CH_3	CH_3	NO_2	OH	O		CH_3	-4.53
		CF_3			O	C_5H_{11}	C_2H_5	-4.51
		CF_3			O	C_7H_{15}	C_2H_5	-4.50
		CF_3			S	Cyc-Pentane	C_2H_5	-4.48
		CF_3			O	Cyc-Butane	C_2H_5	-4.34
NO_2	CH_3	CH_3	NO_2	OH	S		CH_3	-4.31
NO_2	NH_2	CH_3	CH_3	OH	S			-4.31

Table 8: Glide docking score for the 15 best docking molecules at site III

Site IV -:

Pos 2	Pos 3	Pos 4	Pos 5	Pos 6	Pos 8	Pos 9	Pos 10	Pos 11	Pos 12	Glide Score
NO_2		CH_3			S		CH_3		NH_2	-5.88
NO_2		CH_3			S		CH_3	NH_2		-5.70
NO_2		CH_3			S		CH_3	NO_2		-5.70
NO_2	CH_3	CH_3	OH	OH	S		CH_3			-5.05
NO_2	CH_3	CH_3	CH_3	OH	S		CH_3			-4.90
		CF_3			O	Piperidine	C_2H_5			-4.49
NO_2	NH2	CH_3	CH_3	Br	S		CH_3			-4.34
NO_2		CH_3		OH	S		CH_3			-4.26
NO_2		CH_3		C_2H_5OH	S		CH_3			-4.12
NO_2	Br	CH_3	NH_2	CH_3	S		CH_3			-4.11
NO_2	COOH	CH_3	OH	CH_3	S		CH_3			-4.07
NO_2	I	CH_3	CH_3	CH_3	S		CH_3			-4.06
NO_2	COOH	CH_3	CH_3	NH_2	S		CH_3			-4.04
NO_2	NO_2	CH_3	OH	CH_3	S		CH_3			-4.03
NO_2	COOH	CH_3			S		CH_3			-4.01

Table 9: Glide docking score for the 15 best docking molecules at site IV

Site V

Pos 2	Pos 3	Pos 4	Pos 5	Pos 6	Pos 8	Pos 9	Pos 10	Pos 11	Pos 12	Glide Score
NO_2	CH_3	CH_3	OH	OH	S		CH_3			-5.98
NO_2		CH_3			S		CH_3		C_2H_5OH	-5.40
NO_2	OH	CH_3	OH	CH_3	S		CH_3			-5.40
NO_2		CH_3			S		CH_3	CH_3OH		-5.10
NO_2		CH_3			S		CH_3	OH		-5.08
NO_2		CH_3			S		CH_3		OH	-4.99
		CF_3			O	Morpholine	C_2H_5			-4.91
NO_2		CH_3			S		CH_3	COOH		-4.83
NO_2	Br	CH_3	OH	CH_3	S		CH_3			-4.79
NO_2		CH_3			S		CH_3	COOH		-4.76
NO_2	OH	CH_3			S		CH_3			-4.73
NO_2		CH_3			S		CH_3		CN	-4.70
NO_2		CH_3			S		CH_3		CH_3Cl	-4.69
NO_2		CH_3			S		CH_3	CN		-4.56
NO_2		CH_3			S		CH_3	Cl		-4.56

Table 10: This table shows Glide docking score for the 15 best docking molecules at site V

Analysis of Docking score of all the molecules at all the sites.

Pos 2	Pos 3	Pos 4	Pos 5	Pos 6	Pos 8	Pos 9	Pos 10	Pos 11	Pos 12	Glide Score
NO_2	CH_3	CH_3	OH	OH	S		CH_3			-5.98
NO_2		CH_3			S		CH_3		NH_2	-5.88
NO_2	CH_3OH	CH_3			S		CH_3			-5.85
NO_2		CH_3			S		NH_2			-5.81
NO_2	OH	CH_3	CH_3	OH	S		CH_3			-5.80
NO_2		CH_3			S		CH_3		NH_2	-5.78
NO_2		CH_3			S		CH_3	NH_2		-5.73
NO_2		CH_3			S		CH_3	NO_2		-5.73
NO_2		CH_3			S		CH_3	NH_2		-5.70
NO_2		CH_3			S		CH_3	NO_2		-5.70
NO_2		CH_3			S		CH_3	CH_3OH		-5.53
NO_2		CH_3			S		CH_3	OH		-5.50
NO_2		CH_3			S		CH_3	C_2H_5OH		-5.48
NO_2	CH_3	CH_3	COOH	OH	S		CH_3			-5.45
		CF_3			S	C_5H_{11}	C_2H_5			-5.44
		CF_3			S		C_2H_5			-5.42
NO_2		CH_3	NH_2	CH_3	S		CH_3			-5.41
NO_2		CH_3			S		CH_3		C_2H_5OH	-5.40
NO_2	OH	CH_3	OH	CH_3	S		CH_3			-5.39
		CF_3			O	Morpholine	C_2H_5			-5.39

Table 11: Glide docking score for the 20 best docking molecules at all sites

It was evidently seen that site II, V and I were very important sites for the molecule because it has better docking score and further Modifications were carried out on the basis of the docking results of these molecules. Further modifications were increase in chain length of preferred functional groups at preferred locations in several combinations.

Modification 1

Pos 3	Pos 5	Pos 6	Pos 11	Pos 12
OH	OH	OH		
				CH_2NH_2
				$C_2H_4NH_2$
			CH_2NH_2	
			CH_2NO_2	
C_2H_5OH				
CH_3OH	CH_3	OH		
CH_3	COOH	OH		
CH_3	COOH	CH_3OH		

Table 12: Modifications in the reference molecule

In some of the molecules, NO_2 was removed from the molecules and following modifications were carried out.

Pos 3	Pos 4	Pos 5	Pos 6	Pos 9	Pos 10
	CF_3		CH_3OH	C_2H_5	C_2H_5
OH	CF_3	OH	OH	Piperidine	C_2H_5
CH_3	CF_3	OH	OH	Piperidine	C_2H_5
OH	CF_3	CH_3	OH	Piperidine	C_2H_5
OH	CF_3	OH	CH_3	Piperidine	C_2H_5

Table 13: This table show modifications in the reference molecule

The modified molecules were then docked in site I, II and V.

Rigid Docking analysis at three sites after 1st Modification

Site I

Pos 2	Pos 3	Pos 4	Pos 5	Pos 6	Pos 8	Pos 9	Pos 10	Pos 11	Glide Score
		CF_3		C_2H_5OH	O		C_2H_5		-5.43
NO_2	OH	CH_3	OH	OH	S		CH_3		-4.96
	OH	CF_3	OH	OH	O	Piperidine	C_2H_5		-4.91
NO_2		CH_3			S		CH_3	NH_2	-4.61
	CH_3	CF_3	OH	OH	O	Piperidine	C_2H_5		-4.61

Table 14: Glide docking score for the 5 best docking molecules at site I

Site II

Pos 2	Pos 3	Pos 4	Pos 5	Pos 6	Pos 8	Pos 9	Pos 10	Pos 11	Glide Score
NO_2	OH	CH_3	OH	OH	S		CH_3		-7.30
	OH	CF_3	OH	OH	O	Piperidine	C_2H_5		-6.98
	OH	CF_3	CH_3	OH	O	Piperidine	C_2H_5		-6.11
NO_2		CH_3			S		CH_3	NH_2	-5.91
	OH	CF_3	OH	OH	O	Piperidine	C_2H_5		-5.77

Table 15: This table shows Glide docking score for the 5 best docking molecules at site II

Site V

Pos 2	Pos 3	Pos 4	Pos 5	Pos 6	Pos 8	Pos 9	Pos 10	Glide Score
	CH_3	CF_3	OH	OH	O	Piperidine	C_2H_5	-6.31
NO_2	CH_3	CH_3	COOH	OH	S		CH_3	-5.24
NO_2	OH	CH_3	OH	OH	S		CH3	-5.24
	OH	CF_3	OH	OH	O	Piperidine	C_2H_5	-5.11
	OH	CF_3	CH3	OH	O	Piperidine	C_2H_5	-4.56

Table 16: This table shows Glide docking score for the 5 best docking molecules at site V

Further modifications were carried out in the molecules and their docking was further observed at Site II.

2nd Modification

Pos 3	Pos 5	Pos 6	Pos 10	Pos 11	Pos 12
			CH_2NH_2		
		OH	CH_2NH_2		
	OH	OH	CH_2NH_2		
OH	OH	OH	CH_2NH_2		
CH_3OH	OH	OH			
OH	CH_3OH	OH			
OH	OH	CH_3OH			
CH_3OH	CH_3OH	CH_3OH			
CH_3OH	CH_3OH	CH_3OH			NH_2
CH_3OH	CH_3OH	CH_3OH		NH_2	NH_2
CH_3OH	CH_3OH	CH_3OH			CH_2NH_2
CH_3OH	CH_3OH	CH_3OH			$C_2H_4NH_2$
	CH_3OH	CH_3OH			$C_2H_4NH_2$
		CH_3OH			$C_2H_4NH_2$
		OH			NH_2
		OH			$C_2H_4NH_2$
	OH	OH			$C_2H_4NH_2$
OH	OH	OH			$C_2H_4NH_2$

Pos 3	Pos 4	Pos 5	Pos 6	Pos 8	Pos 9	Pos 10
	CF_3		OH	O	Piperidine	C_2H_5
	CF_3	OH		O	Piperidine	C_2H_5
OH	CF_3			O	Piperidine	C_2H_5
	CF_3	OH	OH	O	Piperidine	C_2H_5
OH	CF_3	OH	OH	O	Piperidine	C_2H_5
	CF_3		CH_3OH	O	Piperidine	C_2H_5
	CF_3	OH	CH_3OH	O	Piperidine	C_2H_5
OH	CF_3	OH	CH_3OH	O	Piperidine	C_2H_5
OH	CF_3	CH_3OH	OH	O	Piperidine	C_2H_5
CH_3OH	CF_3	OH	OH	O	Piperidine	C_2H_5
	CF_3		NH_2	O	Piperidine	C_2H_5
	CF_3	NH_2		O	Piperidine	C_2H_5
NH_2	CF_3			O	Piperidine	C_2H_5
	CF_3	NH_2	NH_2	O	Piperidine	C_2H_5
NH_2	CF_3	NH_2	NH_2	O	Piperidine	C_2H_5
NH_2	CF_3	NH_2		O	Piperidine	C_2H_5
NH_2	CF_3		NH_2	O	Piperidine	C_2H_5

Table 17: Modifications in the reference molecule

Rigid Docking Analysis at site II after 2nd round of Modification

Pos 2	Pos 3	Pos 4	Pos 5	Pos 6	Pos 8	Pos 9	Pos 10	Pos 11	Pos 12	Glide Score
	OH	CF_3	CH_3	CH_3OH	O	Piperidine	C_2H_5			-8.14
		CF_3	CH_3	CH_3OH	O	Piperidine	C_2H_5			-7.53
NO_2	CH_3OH	CH_3	CH_3OH	CH_3OH	S		CH_3	NH_2	NH_2	-7.52
NO_2	OH	CH_3	OH	OH	S		CH_2NH_2			-7.43
NO_2		CH_3	OH	OH	S		CH_3		$C_2H_4NH_2$	-7.17

Table 18: This table shows Glide docking score for the 5 best docking molecules at site II

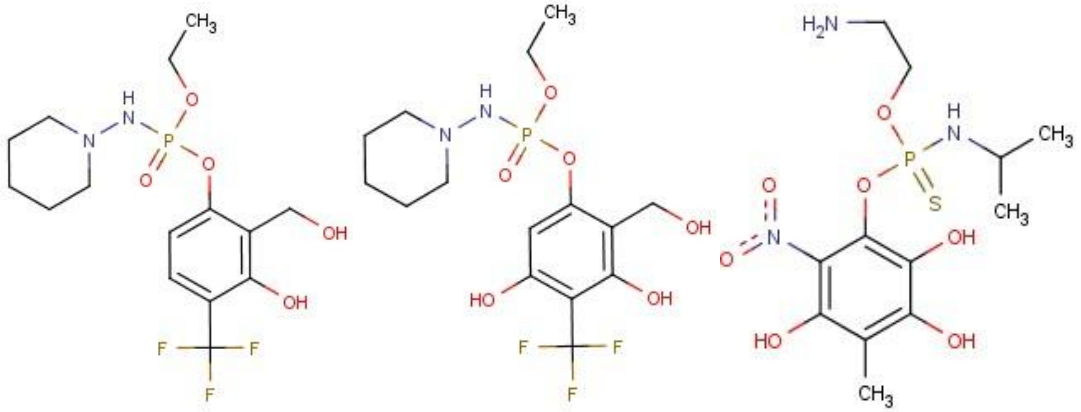
5.11.2 Flexible docking (Induced Fit Docking)

The molecules which were observed to have best docking score in the rigid docking analysis were the shortlisted for flexible docking analysis.

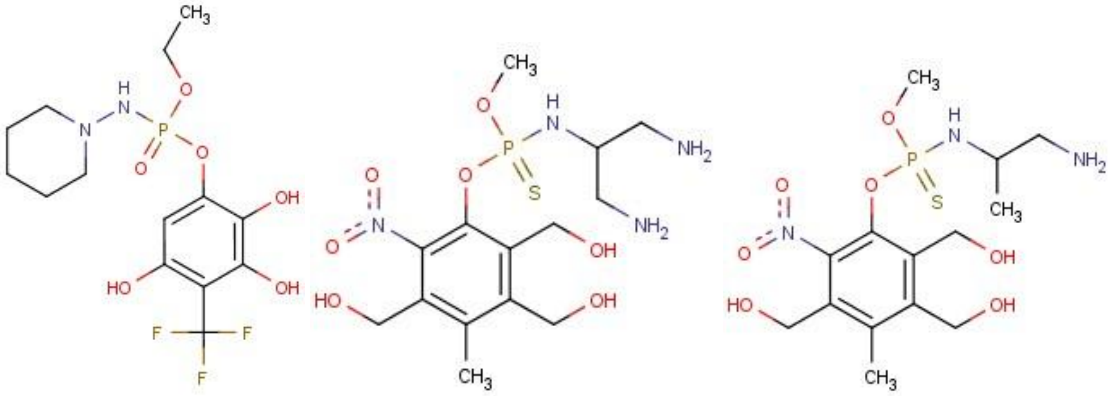
Mol no.	Pos2	Pos 3	Pos 4	Pos 5	Pos 6	Pos8	Pos 9	Pos 10	Pos11	Pos12
1			CF_3	CH_3	CH_3OH	O	Piperidine	C_2H_5		
2		OH	CF_3	CH_3	CH_3OH	O	Piperidine	C_2H_5		
3	NO_2	OH	CH_3	OH	OH	S		CH_2NH_2		
4		OH	CF_3	OH	OH	O	Piperidine	C_2H_5		
5	NO_2	CH_3OH	CH_3	CH_3OH	CH_3OH	S		CH_3	NH_2	NH_2
6	NO_2	CH_3OH	CH_3	CH_3OH	CH_3OH	S		CH_3		NH_2
7	NO_2	OH	CH_3	OH	OH	S		CH_3		
8	NO_2		CH_3			S		NH_2		
9	NO_2	CH_3OH	CH_3			S		CH_3	NH_2	NH_2
10	NO_2	CH_3	CH_3	OH	OH	S		CH_3		

Table 19: List of molecules taken for flexible docking analysis and binding affinity analysis

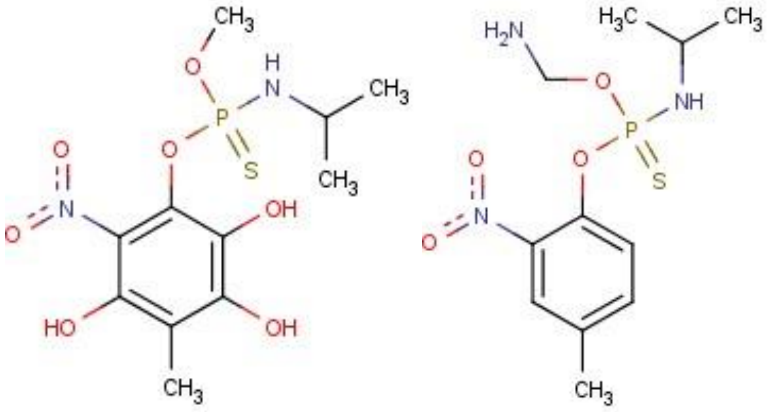
Molecules taken for Induced Fit Docking and Binding Affinity Analysis



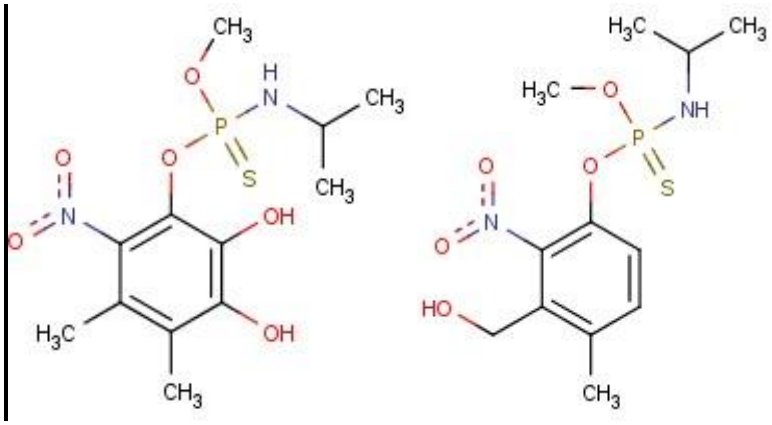
Molecule 1 **Molecule 2** **Molecule 3**



Molecule 4 **Molecule 5** **Molecule 6**



Molecule 7 **Molecule 8**



Molecule 9 **Molecule 10**

Figure 25: Molecular structure of the molecules shortlisted for flexible docking and binding affinity analysis

Induced Fit Docking Results

Mol no	Glide Score	IFD Score
1	-10.5	-762.72
2	-10.43	-762.29
3	-10.24	-766.93
4	-9.65	-768.15
5	-9.25	-761.87
6	-8.99	-760.97
7	-8.46	-764.92
8	-7.61	-767.01
9	-6.68	-757.47
10	-6.65	-763.28

Table 20: Glide scores and IFD scores for the shortlisted molecules from Induced fit docking

5.12 Binding Affinity Results-:

These results give H-bonding energy ,binding energy values and other energy values which should be less so that to have a stable binding structure. Lower the binding energy, more stable the bound structure.

Mol no-:	H-bonding energy	Coulomb energy	Covalent binding energy	Pi-pi packing energy	Lipophilic energy	Binding Energy	Generalized Born electrostatic solvation energy	Van der Waals energy
Ref mol	-0.329517	3.52449	4.771766	-0.16985	-42.4599	-80.9623	5.047434	-51.3467
1	-2.304962	-19.229285	2.092319	-3.63481	-37.4579	-83.7972	21.07959	-44.3422
2	-3.78345	-66.204085	5.945252	0	-32.5033	-98.1564	37.68769	-39.2986
3	-3.232112	-75.557023	6.867832	-0.96362	-26.9224	-75.7671	64.66182	-40.6217
4	-3.182465	-20.741459	5.036323	0	-34.65	-80.6318	15.84323	-42.9374
5	-3.081603	-86.815099	5.007722	-0.53329	-29.8421	-72.9671	76.36207	-34.0648
6	-3.012932	-63.984506	7.639102	-0.73442	-35.1454	-88.3183	47.4411	-40.5212
7	-3.079679	-7.945783	8.933063	-8E-06	-25.8327	-65.6209	4.586511	-42.2823
8	-1.655473	-28.592043	6.44118	-0.34908	-25.0286	-62.7406	30.33547	-43.8921
9	-1.029088	-14.79427	1.183564	-0.64003	-28.9749	-70.9146	10.23543	-36.8953
10	-1.136611	-2.340538	2.461681	-2.7176	-25.6116	-55.924	11.2985	-37.8779

Table 21: Prime MM-GBSA scores for analysis of Binding Affinity. The coloured rows contain molecules having better binding energy scores than the reference molecule

5.13 QikProp Results for ADME property analysis.

Mol. no.	#stars	MW	Human Oral Absorption	Percent Human Oral Absorption	Rule Of Five	Rule Of Three
Ref mol	0	304.3	3	100	0	0
1	0	398.318	3	80.594	0	0
2	0	414.318	3	69.874	0	0
3	1	381.34	2	36.938	1	2
4	0	400.291	3	68.09	0	0
5	5	424.408	1	0.127	2	2
6	1	409.393	2	37.642	1	2
7	0	352.298	3	77.948	0	0
8	1	319.315	3	76.897	0	0
9	0	350.326	3	91.736	0	0
10	0	334.326	3	100	0	0

Table 22: Qikprop scores for analysis of ADME properties. The coloured rows contain molecules having gud ADME properties

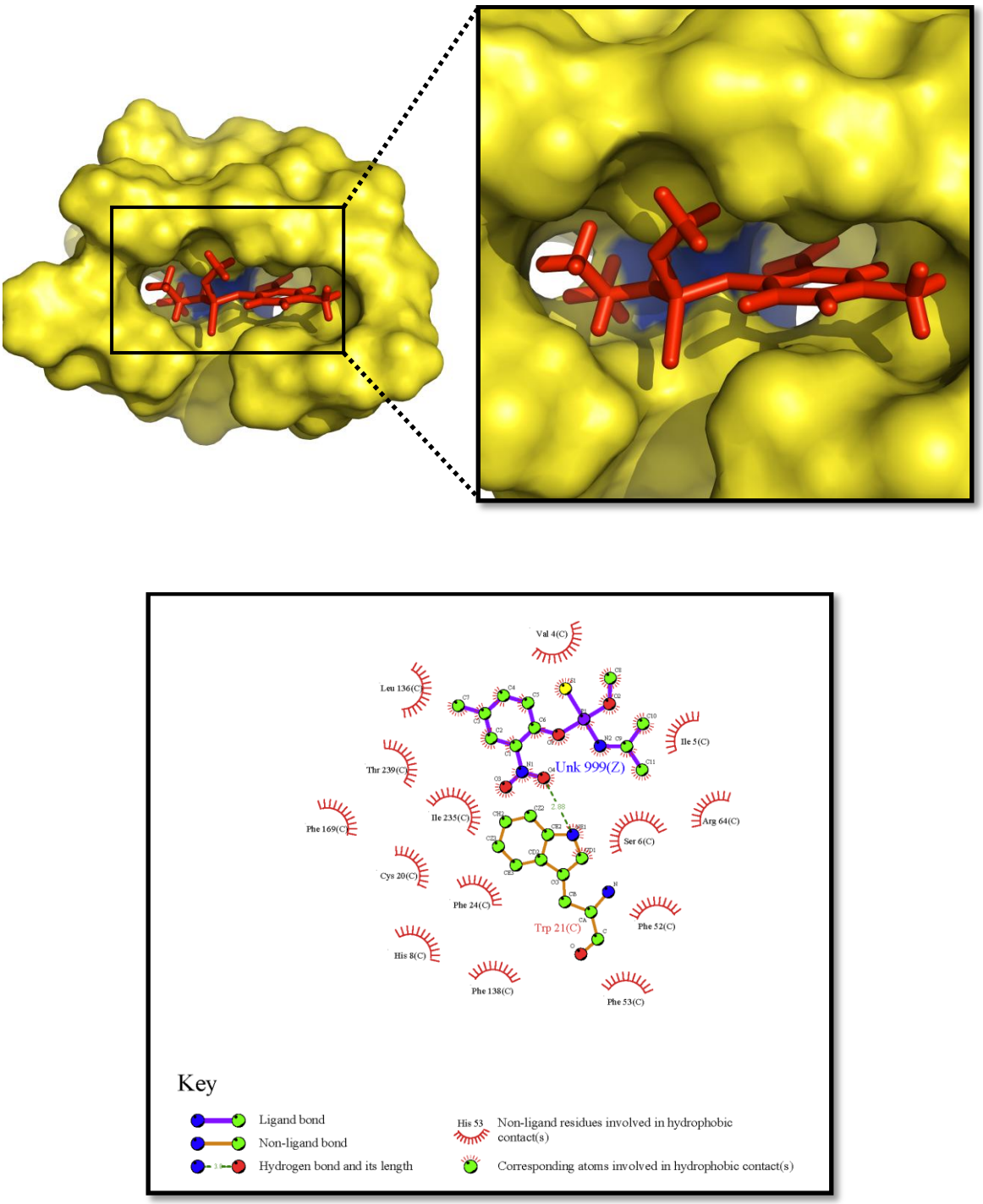
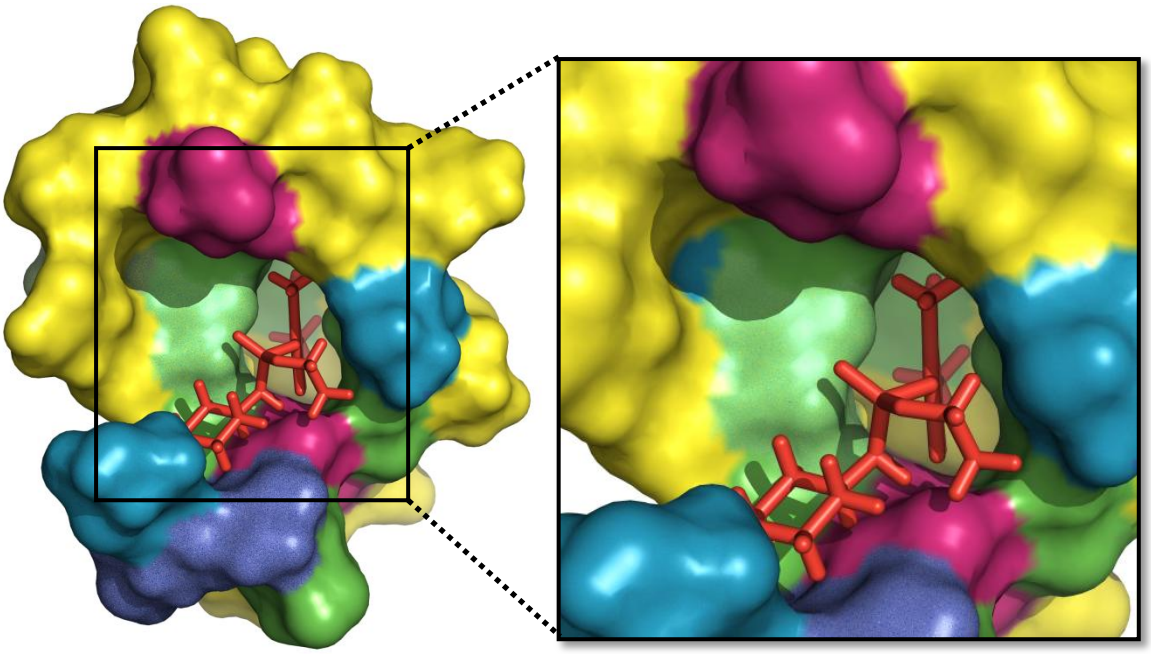
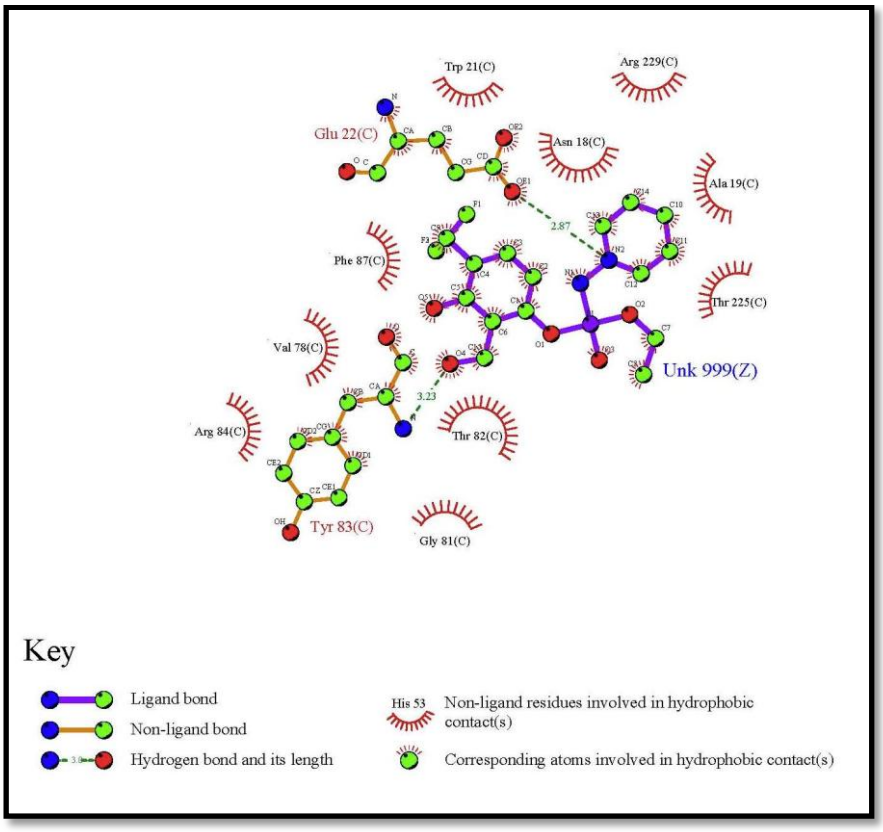


Figure 26: Interaction of amiprophos methyl with target protein at site II (i) Docked reference molecule in the target protein at the site II with blue colour showing hydrogen bind between ligand and Tyr(21) of protein (ii) Ligplot analysis of protein and ligand interactions

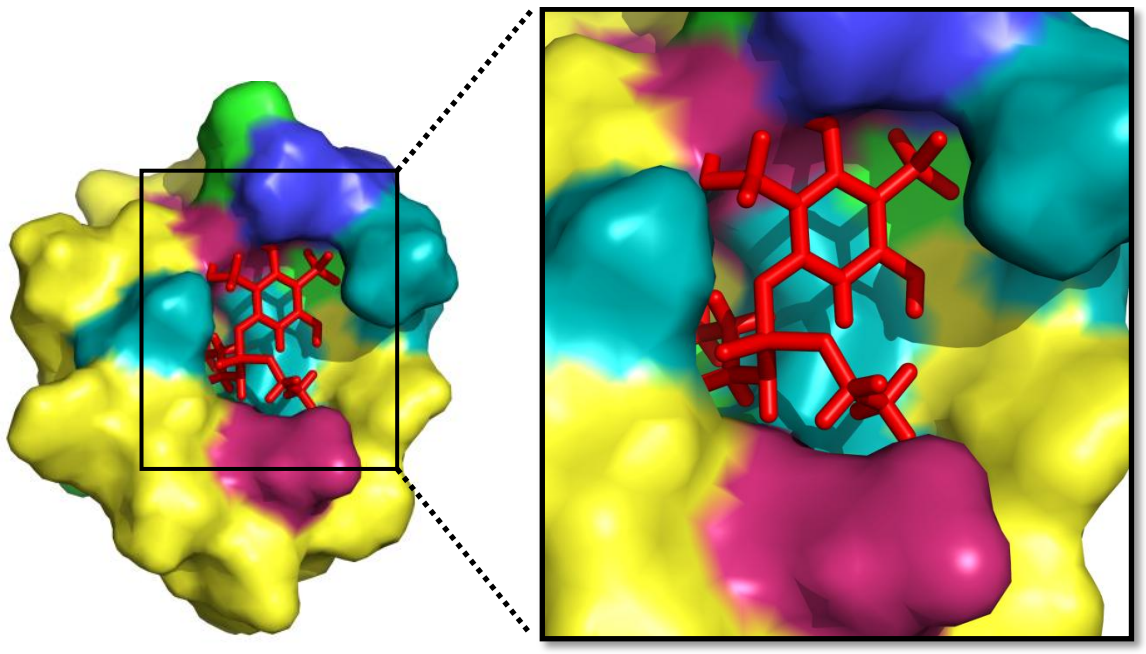


(i)

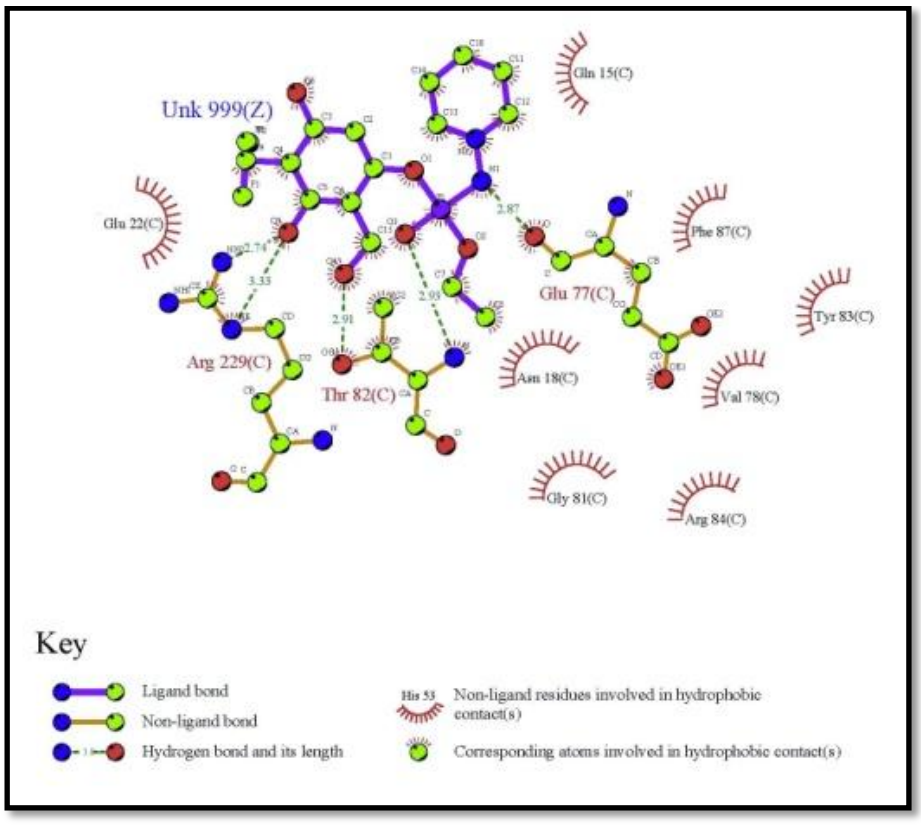


(ii)

Figure 27: Interaction of 3-({ethoxy[(piperidin-1-yl)amino]phosphoryl} oxy)-2-(hydroxymethyl)-6-(trifluoromethyl)phenol with target protein at site II (i) Docked molecule in the target protein at the site II with different colours showing different polarity and charges. Pink, blue, green and cyan represents negative charge, positive charge hydrophobic and polar residues respectively. (ii) Ligplot analysis of protein and ligand interactions



(i)



(ii)

Figure 28: Interaction of 5-({ethoxy[(piperidin-1-yl)amino]phosphoryl} oxy)-4-(hydroxymethyl)-2-(trifluoromethyl) benzene-1,3-diol with target protein at site II (i) Docked molecule in the target protein at the site II with different colours showing different polarity and charges. Pink, blue, green and cyan represents negative charge, positive charge hydrophobic and polar residues respectively (ii) Ligplot analysis of protein and ligand interactions

CONCLUSION

From the predicted Host-pathogen PPI , the present study concludes that most of the host proteins with which pathogen protein interacts are structural proteins such as actin, tubulin and histone. Most of the pathogen proteins involved in infection process are structural and assembly proteins and most of the host proteins are either structural proteins or nuclear assembly proteins. Hence, the pathogen caused infection by targeting nuclear assembly proteins and thereby inhibiting the host cell to function properly. α-tubulin of pathogen is targeted for development of anti-malarial agent for malarial treatment. Derivatives of a herbicide having anti-malarial property were developed and molecule with better binding affinity and ADME property was obtained. It was observed that molecules with electronegative groups have better binding properties than original molecule. Two final molecules with CF_3 at position 4, piperidine ring at position 9 and OH at position 3, 5 and 6 were top best molecules which can be considered as a drug molecules for in vivo analysis and validation.

DISCUSSION AND FUTURE PERSPECTIVE

Complete proteome of *Plasmodium falciparum* was downloaded from Uniprot and protein-protein interactions with its human host were predicted through BIPS. BIPS works on the concept of interologs which works on the evolutionary conservation of interactions. The server hosts a tool which predicts that two proteins from two different species should be interacting if their homologs are found to be interacting in the same species.

The predicted interactions were then annotated for their cellular location, Biological process in which they are involved and their gene ontologies. Then the interactions were filtered on the basis of their cellular location and biological process included. The filtering removes the proteins which lie in cellular locations where they could not interact. The filtered lists of interactions were annotated for their functional role to provide an insight into the interactions. Panther tool classifies pathogen proteins on the basis of their molecular function, biological process, pathways and their protein classes.

It was observed that most of the interacting pathogen proteins are involved in metabolic process and have binding as their molecular function. These proteins fall into nucleic acid binding class of proteins and thus we conclude that they have an important function in DNA replication and hence in cell survival. There was no majority seen in the pathway classification of these proteins but there was a similarity that they all were involved in signalling and disease pathways.

After filtering of protein interactions between *Plasmodium falciparum* and *Homo Sapiens*, they were analysed on the basis of their interacting Partners. We analysed the interacting proteins from the predicted interactions in Human and *Plasmodium* and most highly interacting proteins of pathogen were listed. The significance of this analysis is that the pathogen protein with highest interacting partner can be considered to be the most highly involved protein in the infection and may be a critical protein affecting the infection process. Thus this protein can be targeted as a potential target.

We didn't consider human protein for target because that protein might have a role as essential component in the biological processes. So targeting human proteins have a high risk of toxicity. Human proteins which were interacting with more than three pathogen proteins were usually structural and assembly proteins such as actin, tubulin and histone. Most of these human proteins were histone proteins. This shows that parasite infection affects mostly nuclear and cell assembly proteins in humans. Therefore, we left out these proteins because these are important proteins in human cells. If we will target human proteins or proteins similar to human protein, the problem of toxicity in human after drug delivery will persist. Hence, We targeted pathogen proteins to remove the possibility of negative effects in host. We short-listed following three proteins that are highly interacting on the basis of predicted interactions. The Uniprot IDs of the finally filtered pathogen proteins which are interacting with a large number of human proteins and hence they can be used as a potential target are as follows.

1.Q6ZLZ9

2.QZIFB3

3.Q8ILV1

Out of these following proteins, Q6ZLZ9 and QZIFB3 are alpha-tubulins and Q8ILV1is a serine/threonine protein phosphatase. Out of these three prioritized proteins, Q6ZLZ9 was found to be the most characterized and functionally important protein in pathogen. Q6ZLZ9 is a tubulin protein which has an important role as assembly protein in pathogen. It polymerizes to carry out several critically important roles throughout entire parasite life cycle. In parasite, they form mitotic spindle during cell division and even slight disruption of microtubule causes a severe impact on viability of parasite. *Plasmodium falciparum* infects host and initial contact occurs between merozite and erythrocyte.

In *Plasmodium falciparum* cell cycle ,microtubule is found to have a significant role. Microtubule is present at the tip of the merozites which are found to have a important role in cell division and infection. They are found to have a role RBC invasion because they disappear after invasion. In experimental studies it was confirmed when invasion was decreased and stopped completely when merozites were exposed to tubulin inhibitors. Experimental studies have also demonstrated that microtubules were disrupted on exposure to anti-tubulin agent indicating the role of intact microtubule in merozite invasion. Microtubule is found in many stages of malaria parasite validating it as a potential drug target. As microtubule is found in several stages of malaria and it is used in cellular movement (Rawlings, Fujioka *et al*. 1992). Detailed examination of merozites in erythrocyte invasion identified that targeting α-tubulin is potential approach for malaria therapy.

Amiprophos methyl (APM) is an antimitotic herbicide and is already known inhibitor for α-tubulin. It is found to be a promising molecule because of its low mammalian toxicity. It was reported in ref studies that amiprophios methyl has better specificity for pathogen proteins and has no binding site in human tubulin protein.

Although tubulin is a ubiquitous protein, but still there is a significant difference in amino acids sequence of mammalian and parasitic tubulins. Dinitroanilines were approved as great anti-tubulins and proved to be very good tubulin inhibitors. APM bind to tubulin in the same way as dinitroanilines. They are found to shown better inhibition in lower concentrations. With molecular studies, it was shown that these phosphorothiomidate compounds have similar electrostatic surfaces as dinitroanilines with similar shape and electronegative domains. APM also has low mammalian toxicity. It was observed that APM prevents erythrocytic schizogony and blocks mitosis in *Plasmodium falciparum* infection and results in abnormal microtubule accumulation. This suggests that APM is worthy of investigation for its anti-malarial potential (Fennell, Naughton *et al*. 2006, Mara, Dempsey *et al*. 2011, Mara, Dempsey *et al*. 2013).

We carried out docking analysis of target and amiprophos methyl so that we can obtain its binding score at all the sites which helps us to prioritize one site. This molecule is then used as a reference molecule for further analysis. The maximum docking score of reference molecule i.e. amiprophos methyl with target is -4.4262 at site II. Out of the five sites predicted by SiteMap, Site I and II showed best Site score and Druggability score. Hence we prioritized these sites.

The reference molecule docking to the target and site score results, we prioritized 3 sites i.e. site I, II and V. We have designed the derivatives of amiprophos methyl through modification by adding different groups at individual positions and in combinations. Docking analysis was then carried out.

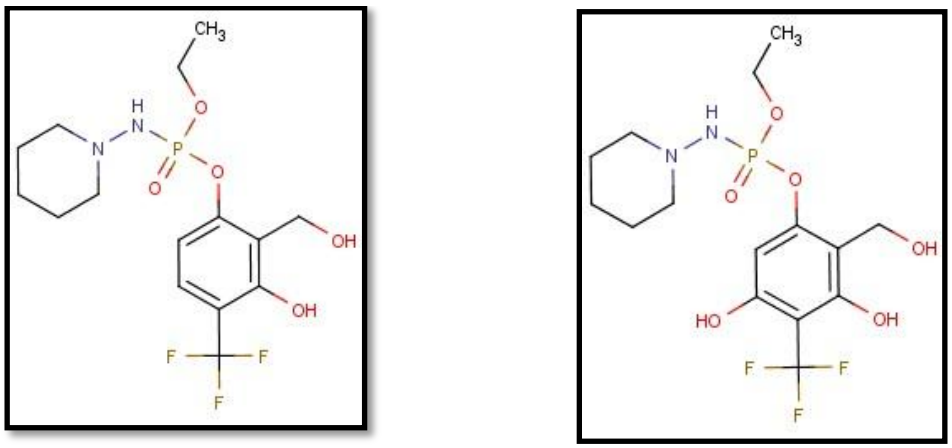
Then docking analysis of ligands and target was carried out at all the sites and variation of the docking score with respect to the functional group was observed. We analysed that the molecules with electronegative group OH at position 3,5 and 6, CH_3 replaced by CF_3 and molecules with piperidine at position 9 showed better docking score.

Then modified molecules were analysed for their docking property at site II. The molecule with OH at all 3 positions i.e. 3,5 and 6 showed better docking score than OH at any of the one and two positions. The docking score also increased when the chain length of NH_2 was increased at position 11 and 12.

Further Modifications were carried out according to the preferred functional groups at preferred sites according to the best docking scores analysed. Modified molecules with the functional groups showing better docking property. Repeated modification and docking score analysis was carried out at site I, II and V. The final molecules with best docking score were prioritized and analysis of these molecules for their docking property at site II was carried out. When oxygen group is replaced by sulphur at position 8, no significant change in docking score was observed. When CH_3 is replaced by CF_3 at position 4, docking score showed significant increase.

Molecules with piperidine ring at position 9 also showed better docking property than other molecules. Further OH group is added at position 3, 5 and 6 to the molecule with piperidine ring at position 9 which showed increase in docking property.

A molecule with CF_3 at position 4, piperidine at position 9 and OH at position 3 showed the best docking score of -8.13757 at site II. This molecule has 83% better docking score than the reference molecule.Molecules with best binding scores were further taken for flexible docking analysis. Ten molecules were shortlisted for flexible docking analysis with highest docking scores.



(a) (b)

Figure 28: Shortlisted molecules for in vivo drug validation.

These final molecules are then analysed for their interaction with residues at site II. Molecule a forms 2 hydrogen bonds with Glu(22) and Tyr(83). Molecule b forms 4 hydrogen bonds with surrounding residue i.e. 2 with Arg(229), 2 with Thr(82) and one with Glu(77). The reference molecule was forming only a single hydrogen bond with Trp(21). These hydrogen bonds of these molecules with surrounding residues governs their stability and hence new molecules have better stability than reference molecule.

These molecules were analysed for their ADME properties. The ADME analysis of these molecules helped us to compare these molecules for their drug like properties. It was essential to check if these molecules fall in the ranges assigned for the molecule for their drug like properties.

The final molecules shown in the figures have docking scores -10.5 and -10.43. These molecules have characteristic features of having electronegative functional groups at position 3,4, 5 and 6 and piperdine ring at position 9.

These shortlisted molecules were then analysed for their binding affinities. It was observed that final two molecules have low values of binding energy than reference molecule which is favourable to binding. Hence binding affinity of the final two molecules is better than the reference molecule.

These molecules were further analysed for their ADME properties. They were found to have high value of human oral absorption. These molecules show 0 violations in Lipinski's rule of five. It also shows zero violation in Jorgensen’s rule of three and all the properties lie under range.

In future work, we wish to extend this study by considering the two final molecules as potential drug molecules. We will study these *in vivo* activity of these molecules and their potential as anti-malarial compounds. We also wish to check the toxicity analysis of these molecules and hence confirm its drug like property in treatment of the disease.

References-:

1. Anthony, R. G., T. R. Waldin, J. A. Ray, S. W. Bright and P. J. Hussey (1998). "Herbicide resistance caused by spontaneous mutation of the cytoskeletal protein tubulin." Nature **393**(6682): 260-263.

2. Armson, A., K. Menon, A. O'Hara, L. M. MacDonald, C. M. Read, K. Sargent, R. C. Thompson and J. A. Reynoldson (2002). "Efficacy of oryzalin and associated histological changes in Cryptosporidium-infected neonatal rats." Parasitology **125**(Pt 2): 113-117.

3. Barnes, K. I., D. N. Durrheim, F. Little, A. Jackson, U. Mehta, E. Allen, S. S. Dlamini, J. Tsoka, B. Bredenkamp, D. J. Mthembu, N. J. White and B. L. Sharp (2005). "Effect of artemether-lumefantrine policy and improved vector control on malaria burden in KwaZulu-Natal, South Africa." PLoS Med **2**(11): e330.
4. Cai, H., Z. Zhou, J. Gu and Y. Wang (2012). "Comparative Genomics and Systems Biology of Malaria Parasites." Curr Bioinform **7**(4).

5. Casadevall, A. and L. A. Pirofski (1999). "Host-pathogen interactions: redefining the basic concepts of virulence and pathogenicity." Infect Immun **67**(8): 3703-3713.
   \
6. Casadevall, A. and L. A. Pirofski (2000). "Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease." Infect Immun **68**(12): 6511-6518.

7. Chua, H. N., W. K. Sung and L. Wong (2006). "Exploiting indirect neighbours and topological weight to predict protein function from protein-protein interactions." Bioinformatics **22**(13): 1623-1630.

8. Csermely, P., T. Korcsmaros, H. J. Kiss, G. London and R. Nussinov (2013). "Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review." Pharmacol Ther **138**(3): 333-408.

9. Desai, P. V., A. Patny, J. Gut, P. J. Rosenthal, B. Tekwani, A. Srivastava and M. Avery (2006). "Identification of novel parasitic cysteine protease inhibitors by use of virtual screening. 2. The available chemical directory." J Med Chem **49**(5): 1576-1584.

10. Dow, G. S., A. J. O'Hara, S. C. Newton, J. A. Reynoldson and R. C. Thompson (2000). "Plasmodium berghei: the antimalarial activity of albendazole in rats is mediated via effects on the hematopoietic system." Exp Parasitol **94**(4): 259-263.

11. Dyer, M. D., T. M. Murali and B. W. Sobral (2007). "Computational prediction of host-pathogen protein-protein interactions." Bioinformatics **23**(13): i159-166.

12. Dyer, M. D., T. M. Murali and B. W. Sobral (2011). "Supervised learning and prediction of physical interactions between human and HIV proteins." Infect Genet Evol **11**(5): 917-923.
13. Evans, P., W. Dampier, L. Ungar and A. Tozeren (2009). "Prediction of HIV-1 virus-host protein interactions using virus and host sequence motifs." BMC Med Genomics **2**: 27.

14. Fennell, B. J., J. A. Naughton, E. Dempsey and A. Bell (2006). "Cellular and molecular actions of dinitroaniline and phosphorothioamidate herbicides on Plasmodium falciparum: tubulin as a specific antimalarial target." Mol Biochem Parasitol **145**(2): 226-238.

15. Foster, S. and M. Phillips (1998). "Economics and its contribution to the fight against malaria." Ann Trop Med Parasitol **92**(4): 391-398.

16. Franzosa, E. A., S. Garamszegi and Y. Xia (2012). "Toward a three-dimensional view of protein networks between species." Front Microbiol **3**: 428.

17. Fujioka, H. and M. Aikawa (2002). "Structure and life cycle." Chem Immunol **80**: 1-26.

18. Garcia-Garcia, J., S. Schleker, J. Klein-Seetharaman and B. Oliva (2012). "BIPS: BIANA Interolog Prediction Server. A tool for protein-protein interaction inference." Nucleic Acids Res **40**(Web Server issue): W147-151.

19. Gavin, A. C., P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dumpelfeld, A. Edelmann, M. A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A. M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell and G. Superti-Furga (2006). "Proteome survey reveals modularity of the yeast cell machinery." Nature **440**(7084): 631-636.

20. Gerald, N., B. Mahajan and S. Kumar (2011). "Mitosis in the human malaria parasite Plasmodium falciparum." Eukaryot Cell **10**(4): 474-482.

21. Gschwend, D. A., A. C. Good and I. D. Kuntz (1996). "Molecular docking towards drug discovery." J Mol Recognit **9**(2): 175-186.

22. Hillisch, A., L. F. Pineda and R. Hilgenfeld (2004). "Utility of homology models in the drug discovery process." Drug Discov Today **9**(15): 659-669.

23. Ito, T., T. Chiba, R. Ozawa, M. Yoshida, M. Hattori and Y. Sakaki (2001). "A comprehensive two-hybrid analysis to explore the yeast protein interactome." Proc Natl Acad Sci U S A **98**(8): 4569-4574.

24. Itzhaki, Z., E. Akiva and H. Margalit (2010). "Preferential use of protein domain pairs as interaction mediators: order and transitivity." Bioinformatics **26**(20): 2564-2570.

25. Kiemer, L. and G. Cesareni (2007). "Comparative interactomics: comparing apples and pears?" Trends Biotechnol **25**(10): 448-454.

26. Kooij, T. W., C. J. Janse and A. P. Waters (2006). "Plasmodium post-genomics: better the bug you know?" Nat Rev Microbiol **4**(5): 344-357.

27. Krogan, N. J., G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrin-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St Onge, S. Ghanny, M. H. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili and J. F. Greenblatt (2006). "Global landscape of protein complexes in the yeast Saccharomyces cerevisiae." Nature **440**(7084): 637-643.

28. Kumar, A. and K. Y. Zhang (2012). "Computational fragment-based screening using RosettaLigand: the SAMPL3 challenge." J Comput Aided Mol Des **26**(5): 603-616.

29. Laurie, A. T. and R. M. Jackson (2006). "Methods for the prediction of protein-ligand binding sites for structure-based drug design and virtual ligand screening." Curr Protein Pept Sci **7**(5): 395-406.

30. Lexa, K. W. and H. A. Carlson (2012). "Protein flexibility in docking and surface mapping." Q Rev Biophys **45**(3): 301-343.

31. Liu, P. T., S. Stenger, H. Li, L. Wenzel, B. H. Tan, S. R. Krutzik, M. T. Ochoa, J. Schauber, K. Wu, C. Meinken, D. L. Kamen, M. Wagner, R. Bals, A. Steinmeyer, U. Zugel, R. L. Gallo, D. Eisenberg, M. Hewison, B. W. Hollis, J. S. Adams, B. R. Bloom and R. L. Modlin (2006). "Toll-like receptor triggering of a vitamin D-mediated human antimicrobial response." Science **311**(5768): 1770-1773.

32. MacDonald, L. M., A. Armson, A. R. Thompson and J. A. Reynoldson (2004). "Characterisation of benzimidazole binding with recombinant tubulin from Giardia duodenalis, Encephalitozoon intestinalis, and Cryptosporidium parvum." Mol Biochem Parasitol **138**(1): 89-96.

33. Mara, C., E. Dempsey, A. Bell and J. W. Barlow (2011). "Synthesis and evaluation of phosphoramidate and phosphorothioamidate analogues of amiprophos methyl as potential antimalarial agents." Bioorg Med Chem Lett **21**(20): 6180-6183.

34. Mara, C., E. Dempsey, A. Bell and J. W. Barlow (2013). "Synthesis and evaluation of phenoxyoxazaphospholidine, phenoxyoxazaphosphinane, and benzodioxaphosphininamine sulfides and related compounds as potential anti-malarial agents." Bioorg Med Chem Lett **23**(12): 3580-3583.

35. Martin, R. J., A. P. Robertson and H. Bjorn (1997). "Target sites of anthelmintics." Parasitology **114 Suppl**: S111-124.

36. Matthews, L. R., P. Vaglio, J. Reboul, H. Ge, B. P. Davis, J. Garrels, S. Vincent and M. Vidal (2001). "Identification of potential interaction networks using sequence-based searches for conserved protein-protein interactions or "interologs"." Genome Res **11**(12): 2120-2126.

37. Mendez-Rios, J. and P. Uetz (2010). "Global approaches to study protein-protein interactions among viruses and hosts." Future Microbiol **5**(2): 289-301.

38. Mufunda, J., P. Nyarango, A. Usman, T. Gebremeskel, G. Mebrahtu, A. Ogbamariam, A. Kosia, Y. Ghebrat, S. Gebresillosie, S. Goitom, E. Araya, G. Andemichael and A. Gebremichael (2007). "Roll back malaria--an African success story in Eritrea." S Afr Med J **97**(1): 46-50.

39. O'Meara, W. P., J. N. Mangeni, R. Steketee and B. Greenwood (2010). "Changes in the burden of malaria in sub-Saharan Africa." Lancet Infect Dis **10**(8): 545-555.

40. Palaniyandi, M. (2012). "The role of remote sensing and GIS for spatial prediction of vector-borne diseases transmission: a systematic review." J Vector Borne Dis **49**(4): 197-204.

41. Qi, Y., O. Tastan, J. G. Carbonell, J. Klein-Seetharaman and J. Weston (2010). "Semi-supervised multi-task learning for predicting interactions between HIV-1 and human proteins." Bioinformatics **26**(18): i645-652.

42. Rawlings, D. J., H. Fujioka, M. Fried, D. B. Keister, M. Aikawa and D. C. Kaslow (1992). "Alpha-tubulin II is a male-specific protein in Plasmodium falciparum." Mol Biochem Parasitol **56**(2): 239-250.

43. Rowe, J. A., A. Claessens, R. A. Corrigan and M. Arman (2009). "Adhesion of Plasmodium falciparum-infected erythrocytes to human cells: molecular mechanisms and therapeutic implications." Expert Rev Mol Med **11**: e16.

44. Rzychon, M., D. Chmiel and J. Stec-Niemczyk (2004). "Modes of inhibition of cysteine proteases." Acta Biochim Pol **51**(4): 861-873.

45. Snow, R. W., C. A. Guerra, A. M. Noor, H. Y. Myint and S. I. Hay (2005). "The global distribution of clinical episodes of Plasmodium falciparum malaria." Nature **434**(7030): 214-217.

46. Sullivan, S. M. and T. Holyoak (2008). "Enzymes with lid-gated active sites must operate by an induced fit mechanism instead of conformational selection." Proc Natl Acad Sci U S A **105**(37): 13829-13834.

47. Tastan, O., Y. Qi, J. G. Carbonell and J. Klein-Seetharaman (2009). "Prediction of interactions between HIV-1 and human proteins by information integration." Pac Symp Biocomput: 516-527.

48. Uetz, P., L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields and J. M. Rothberg (2000). "A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae." Nature **403**(6770): 623-627.

49. Wuchty, S. (2011). "Computational prediction of host-parasite protein interactions between P. falciparum and H. sapiens." PLoS One **6**(11): e26960.

50. Yosef, N., M. Kupiec, E. Ruppin and R. Sharan (2009). "A complex-centric view of protein network evolution." Nucleic Acids Res **37**(12): e88.

51. Zahiri, J., J. H. Bozorgmehr and A. Masoudi-Nejad (2013). "Computational Prediction of Protein-Protein Interaction Networks: Algo-rithms and Resources." Curr Genomics **14**(6): 397-414.

52. Zoraghi, R. and N. E. Reiner (2013). "Protein interaction networks as starting points to identify novel antimicrobial drug targets." Curr Opin Microbiol **16**(5): 566-572.