MOLECULAR DOCKING USING MULTI-AGENT TECHNOLOGY

Mahabaduge Harindra Rasanjali Fernando

(139105C)

Degree of Master of Science in Artificial Intelligence

Department of Computational Mathematics

University of Moratuwa

Sri Lanka

December 2015

MOLECULAR DOCKING USING MULTI-AGENT TECHNOLOGY

Mahabaduge Harindra Rasanjali Fernando

(139105C)

Thesis submitted in
partial fulfillment of the requirements for the degree of MSc in Artificial Intelligence

Department of Computational Mathematics

University of Moratuwa

Sri Lanka

December 2015

Declaration

I declare that this dissertation does not incorporate, without acknowledgment, any material previously submitted for a Degree or a Diploma in any University and to the best of my knowledge and belief, it does not contain any material previously published or written by another person or myself except where due reference is made in the text. I also hereby give consent for my dissertation, if accepted, to be made available for photocopying and for interlibrary loans, and for the title and summary to be made available to outside organization.

Name of Student	Signature of Student
Mrs. M.Harindra.R.Fernando	Date:

Supervised by
Prof.Asoka Karunananda	Signature of Supervisor
Date:

Supervised by
Dr. Roshan P. Perera	Signature of Supervisor
Date:

Dedication

I dedicate this thesis to my parents
Mr.Shelton Fernando and Mrs.Hemamalani Fernando.

I hope this achievement will complete the dream that both of you had for me all those many years ago when you choose to give the best education you could.

Acknowledgements

Foremost, I would like to express my sincere gratitude to my advisor Prof. Asoka Somabandu Karunanada for the continuous support of my M.Sc. research, for his patience, motivation, enthusiasm, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis. I could not have imagined having a better advisor for my study. He helps me a lot to penetrate into the scientific research world with his excellent and valuable advice.

Besides my advisor, I would like to thank the rest of my thesis committee: Prof. Nalin Wickramarachchi and Prof. N G J Dias for their encouragement, insightful comments, and hard questions.

My sincere thanks also go to Dr.Roshan Perera for offering me chemistry related knowledge, leading me working on a diverse exciting project as the co-supervisor.

It is an honor for me to mention the name of Prof. Ranil Dassanayake, Department of Chemistry, University of Colombo, who gave me massive guidance for the evaluation of the software tool. Without him the value of this project will be a dream.

I would like to thank my family: my parents, for giving birth to me, in the first place and supporting me spiritually throughout my life. Last but not the least, my special thanks should go to my loving husband, to assist me, by giving his fullest cooperation to fulfill my research work.

Abstract

Traditional computer-based simulators for manual molecular docking for rational drug discovery have been very time consuming and a tedious task. It is evident from the literature that such computer-based solutions have been implemented merely with conventional software technologies. A large body of research publication has shown the power of Multi Agent technology for development of smart fully automated simulators.

In this research, a multi agent-based solution, named as NanoAgent, has been developed to automate the drug discovery process with little human intervention. In this solution, *ligands* and *proteins* are implemented as agents, who pose the knowledge of permitted connections with other agents to form new molecules. The system also includes several other agents for surface determination, cavity finding and energy calculation. These agents autonomously activate and communicate with each other to come up with a most probable structure over the ligands and proteins, which are participating in deliberation. Domain ontology is maintained to store the common knowledge of molecular bindings, whereas specific rules pertaining to the behavior of ligands and proteins are stored in their personal ontologies. Among other operational rules, agents are built with rules pertaining to theories of Poison Boltzmann, Vander Walls, and Monte Carlo, regarding ligands and proteins to calculate the optimal binding energy. Existing, Protein Data Bank (PDB) has also been used to calculate the space required by ligand to bond with the receptor. The drug discovery process of NanoAgent has exemplified exciting features of multi agent technology, including, communication, coordination, negotiation, butterfly effect, self-organizing and emergent behavior. Since agents consume fewer computing resources, NanoAgent has recorded optimal performance during the drug discovery process.

NanoAgent has been tested for the discovery of the known drugs for the known protein targets. It has 80% accuracy by considering the prediction of the correct actual existence of the docked molecules using energy calculations. By comparing the time taken for the manual docking process with the time taken for the molecular docking by NanoAgent, there has been 95% efficiency. The results suggest that the Multi-Agent Systems technology can be successfully applied to automate the manual molecular docking process, which is an inherently complex problem. Further work on this project can be identified as the development of automated solutions for protein-protein docking, which is a hot topic in biochemistry and allied disciplines.

Contents

List of Figures

Page

List of Tables

Page

List of Equations

Chapter 1

Introduction

1.1. Prolegomena

Increasing popularity and the penetration of Artificial Intelligence (AI) technologies into a wide spectrum of subject areas in the complex real world, has made AI as a distinct technology in the 21st century. Over last six decades Artificial Intelligence techniques were showing the unique capacity to solve various insoluble complex real world problems. In particular, the real world systems involving a large number of interconnected entities operating in a distributed environment under unpredictable uncertainty. Large volume of literature in AI has shown the power of it, as a hypothetical machine that exhibits behavior at least as skillful and flexible as humans do. With the increasing popularity in AI, numerous intelligent techniques including Fuzzy Logic [1], Artificial Neural Networks [2], Genetic Algorithms [3], Expert Systems [4] and also the Natural Language Processing [5] have the potential to mitigate complex and unsolvable problems, by carrying out automated intelligent processing of existing data or the information.

Among other AI techniques Multi-Agent System (MAS) has provided effective solutions to problem solving where extensive coordination and efficient resource usage among the molecular docking tools. Mutual interaction in between the agents that operate inside the multi-agent systems environment, enables the collaborative problem solving and it maximizes the utility of the resources. MAS enables the collaborative activities due to the autonomy of the agents .Exciting applications of Multi Agent Systems (MAS) Technology has been reported in the complex real world problems such as logistics management [6], aircraft maintenance [7] and industrial engineering system controlling [8] like areas.

The scientist needs to get the assistance from the various combinations of separate set of tools to generate molecular data such as geometries (torsion angles, bond angles, bond length), energies (activation energy, heat of formation) and also the properties (diffusion, volume, surface areas, viscosity). We can recognize the manual molecular docking process as an inherently complex system and this project has been conducted to develop Multi-Agent System (MAS) solution to automate molecular docking process. In this connection, this chapter present aim and objectives, background and motivation, problem in brief, novel approach to molecular docking and the structure of the overall thesis.

1.2. Aim and Objectives

The main aim of this project is to design and develop a computer based solution to automate the molecular docking process using Multi Agent Technology. So as to reach this, aim the following objectives are identified.

1. To critically study the molecular docking domain with a view to identify current practices and the issues in molecular docking.
2. Critically analyze and comprehensive evaluation of the existing software solutions in molecular docking with a view to define the research problem and possible technology.
3. In depth study about Multi-Agent Technology and its applications.
4. Design and implement Multi-Agent System for Molecular docking.
5. Evaluate the accuracy of the Multi-Agent System based molecular docking tool by entering known existing ligands and protein pairs.

1.3. Background and Motivation

In the structural molecular biology and the computer aided drug design, it is heavily used molecular docking as a predominant tool. This area is primarily concerned with binding a smaller molecules (ligand) with a target macromolecule (protein). Molecular docking is also performed to forecast the connecting mode of a protein with a small molecule or ligand, with the help of 3D structures of the protein and the ligand [9]. Before perform the docking process, it is important to prepare the input structures of the individual molecules, but sometimes when analyzing the ultimate docking result, it can be vague due to the erroneous status of the stochastic search methods [10].

The field of molecular docking has been momentously improved by the intense growth of the power of computers [11]. The easiness of the access to the online drugs (ligand or the small molecules) and the protein databases also helped for the improvement of the molecular docking field [12][13]. Automated molecular docking tools significantly help to understand the interaction in between the molecules such as predicting likely binding modes and energetically determine the binding modes [14].

Molecular docking is typically performing among a small molecule and a target protein (macromolecule). There are two types of molecular docking, such as ligand-protein docking and protein-protein docking. In this research, we have focused on ligand-protein docking. The proteins can be considered, such as DNA (Deoxyribonucleic acid), or RNA (Ribonucleic acid) as the macromolecule. The ligand or the smaller molecules can be considered as the drug that docked with the protein. The 3D molecular structure is the substances of the computer aided drug design. Structures are regularly accessible for the protein and the ligand independently. But it is harder to find the docked structures on the Internet [15].

Considering the major literature behind this research work, Molecular Theory and Computation Group [16] of Stanford University in USA is a popular research group which work on various types of computer aided simulations. They have used molecular dynamics by presenting their own theoretical methods using physics, chemistry and biology. They have emphasized the issue they have faced when simulating the molecular dynamics, that need large processing power even if for the Nano scale molecular processing using the computer.

In 1990 by David S. Goodsell [17] the most popular automated molecular docking program called AutoDock [18] was written in FORTRAN-77. But, AutoDock doesn’t have better search methods and better empirical free energy scoring functions. So it will give the false predictions of docking results [19]. It also needs the assistance from other external tools such as MGLTools [20] and FastGrid [21].

1.4. Problem in Brief

It is essential to use computer aided tools in the drug discovery procedure and it is becoming more easy in the present, because of the growth of X-ray and NMR (Nuclear magnetic resonance) [22] protein and ligand structures. Using manual methods, such as laboratory experiments, doing calculations on the papers are erroneous and economic cost is higher. For the laboratory experiments, most of the time the scientists are using living innocent animals. It is not ethical to harm animals' lives on behalf of the survival of the human beings.

For the extrapolation of protein-ligand relations and to assistance in picking successful molecules by querying the molecular structure databases, molecular docking is widely used in the present. But molecular docking has been seriously affected by limitations of tedious automated molecular docking tools that accessing various optimization tools in the molecular docking environment. It have resulted in inefficiency of the whole molecular docking process leading to irrational drug designs and more expensive clinical trials to test the validity of the discovered drugs.

All the configurations of the connections of the molecules and all the likely confirmations are included the molecular docking search space. It is an easy task to explore the entire search space for all the possible binding poses due to the increasing power of the computers. But still it needs time to find the correct pose. Molecular docking is compromised between the correctness and efficiency. A better docking tool is anticipated to preserve a respectable steadiness between the accuracy and the efficiency.

1.5. Novel Approach to Molecular Docking

Using Multi Agent Technology (MAS) to automate the molecular docking approach will be enhancing the efficiency and the accuracy of the entire process due to the features of MAS such as communication, collaboration and coordination. In this research, all the required tools such as energy minimization and orientation calculations can be done through the proposed all-in-one solution.

The proposed system gets the Protein Data Bank (PDB) files of the ligand and the proteins as the inputs, and output the discovered structures of the matched protein-ligand pairs. It also predicts the actual existence of the docked structure. Considering the design of the proposed system, ligands and proteins are programmed as the agents operate in the Multi Agent System and the rule set for the binding codified as domain ontology. Each agent operated according to their rule set. There are the other agents, who help for the docking of LigandAgents and ProteinAgent, such as

EnergyCalculationAgent, ActiveSiteFindingAgent, SurfaceMatchingAgent and CavityFindingAgent.

1.6. Outline of the Thesis

The remaining of the thesis organization is as follows:

Chapter 2 critically review the domain of molecular docking by highlighting current solutions, practices, technologies, limitations for defining the research problem. Chapter 3 describes the essentials of Multi-Agent Technology shows its relevance to solve the molecular docking domain. Chapter 4 presents our novel approach to molecular docking with Multi-Agent Technology. Chapter 5 is on the design of Multi-Agent System for Molecular Docking. Chapter 6 contains details of implementation of the MAS solution for molecular docking. Chapter 7 illustrates the real world application of the novel approach. Chapter 8 reports on evaluation of the novel solution by explaining evaluation strategy, participants, data collection, data representation and data analysis. Chapter 9 concludes the outcome of the research with a note on further work.

1.7. Summary

This chapter describes the full picture of the whole research project showing research problem, objectives, hypothesis and the novel solution. The next chapter will be on a literature review of molecular docking practices, technologies and issues with a view to define the research problem.

Chapter 2

State of the Art of Molecular Docking

2.1. Introduction

Previous chapter described the full picture of the whole research project showing research problem, objectives, hypothesis and the novel solution. In this chapter it will be present the literature review of molecular docking practices technologies and issues with a view to define the research problem.

There are two main types of molecular interactions. They are *protein-protein* docking interactions, *protein receptor-ligand* interactions. *Protein receptor-ligand* docking can be again divided into two types, called, *rigid receptor with flexible ligand and flexible receptor* with *rigid ligand* [23]. Protein-ligand docking is the most significant type of molecular docking due to the medical use of assembly based drug design. For the designing remedial interventions, calculations of the attraction and the binding mode between the molecules in recognition of correct binding modes of molecules is difficult [24]. The laboratory investigations for defining the molecular assemblies are tedious and the economic cost is higher. This is where the computer aided methods come forward to forecasting the molecular connections. The large number of potential poses of involvements are tried and assessed based on the protein structures in the molecular docking. The smallest energy result is prophesied as the correct pose.

2.2. Molecular Docking Process

There are two essential components called scoring and sampling, in a protein-ligand molecular docking program. Sampling calls for the identification of the ligand binding orientations adjacent to the likely binding region on the protein. It again classifies as, the protein flexibility and the ligand sampling. Scoring is calculation of the rigidity of the attachment for an orientation of the ligand using energy function. The orientation can be considered as the lowest energy score, is anticipated as the binding mode. The protein-ligand docking can be classifying in three ways, such as ligand sampling, scoring function and protein flexibility. That is shown in the following Figure 2.1.

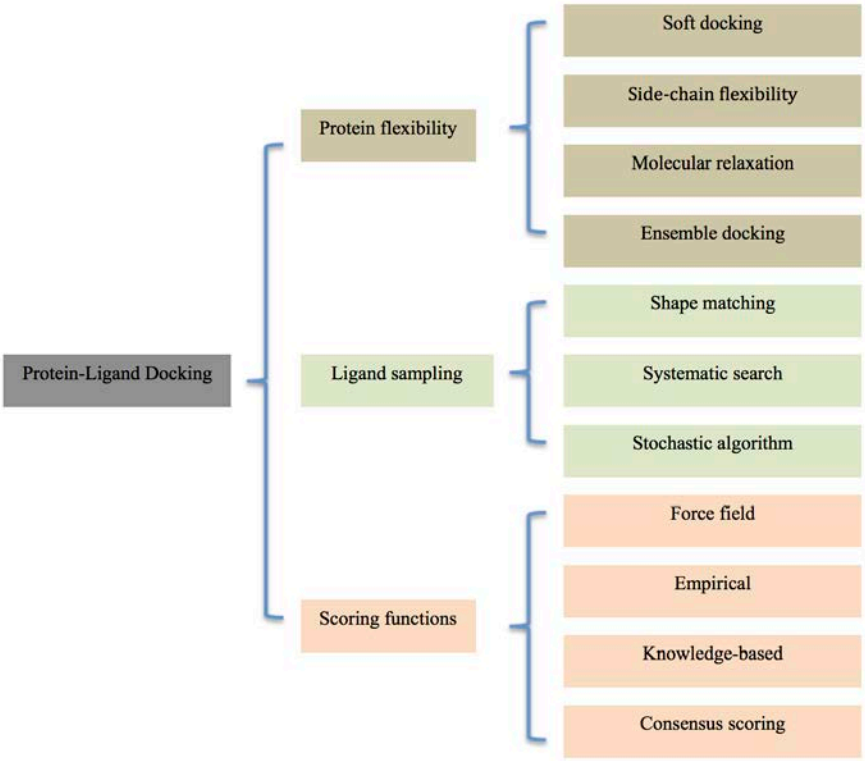


Figure 2.1 Protein-ligand docking methods classifications [24]

Protein Flexibility is the most challenging issue in molecular docking due to the larger size and many degrees of freedoms. Following Table 2.1 summaries the four methods account for protein flexibility.

	Method	Advantages	Disadvantages
1	Soft docking	Computationally efficient and easiness for implementation	Can account for only small conformational changes
2	Side chain flexibility	Enhanced computational techniques available	Not mentioned
3	Molecular relaxation	The inclusion of certain backbone flexibility instead of the side-chain conformational changes	More demanding for the scoring function and time consuming
4	Protein ensemble	Most widely uses method, FlexE [25], AutoDock [18] like docking tools use this method	The nature of the method may failure to the geometric precision of the protein

Table 2.1 Protein Flexibility: Four methods

In the primary stages of docking process *Shape matching* is the simplest sampling algorithm that is frequently used. It considered the six degrees of freedom which is called as three rotational and three translational, of the ligand permit for different probable ligand-binding poses. The shape-matching algorithm is used to predict the three dimensional space to connect with the effective region on a protein as fast as possible. It is computationally efficient. The shape of the ligand is usually stable throughout the shape matching process [24].

2.3. Free Chemistry Databases

A large variety of free online chemistry databases are available on the Internet. Some of the most important databases are shown in the following Table 2.2.

Database Name	No of records	Year	Features	License
PubChem [26]	over 8 million compounds	1999	Can be download for the local use [27]	Free
ZINC [28]	over 4.6 million compounds	2005	Can be download for the local use	Commercial
eMolecules [29]				
CHEBI [30]		2008	Can be download as ontology [31]	Free
EcoCyc/BioCyc [32]	About 3,500 compounds involved as enzyme substrates, products, inhibitors, and activators	2009	The data can be accessed as Chemical Markup Language format [33]	Free
DrugBank [34]	Over 4,300 drugs	2007	Currently service is unavailable	Unavailable

Table 2.2 Chemical Structure Databases available on the Internet

2.4. Computational Molecular Simulation Tools

There are lots of available computational molecular simulation tools such as FRED, DOCK, FLOG, LigandFit, EUDOC, Surflex, MDock and MS-DOCK, that can be used for molecular docking. The features of the available tools have been critically analyzed to find the techniques they have used for molecular docking.

2.4.1. Systematic Search

By exploring all possible degrees of freedom of the ligand, Systematic Search will generate all probable binding poses. Systematic Search can be further divided into three, as fragmentation, conformational ensemble and exhaustive search.

Following are the two search approaches used by the available docking tools.

- Hierarchical sampling methods - Glide and FRED
- Fragmentation methods - DOCK, LUDI, FlexX, ADAM and eHiTs

2.4.2. Stochastic Algorithms

By constructing arbitrary alterations to the ligand at each stage in the two ways, the translational and rotational angles of the ligand separately, Stochastic algorithms are used to test ligand binding orientations. Conferring to a probabilistic principle, the arbitrary alteration will be recognized or excluded. Following are the categories of stochastic algorithms:

- Evolutionary algorithms (EA)
- Tabu search methods
- Swarm optimization (SO) methods
- Monte Carlo (MC) methods

Monte Carlo method [35], is used to calculate the probability to admit a arbitrary alteration by applying the following Boltzmann probability function shown in Equation 2.4.2.1:

$$P \sim \exp\left[\frac{-(E_1 - E_0)}{k_\mathrm{B}T}\right]$$

Equation 2.4.2.1: Boltzmann Function

T - The systems absolute temperature

k_B - The constant of the Boltzmann

E_1 - The enthalpy of the ligand after the arbitrary alteration

E_0 - The enthalpy of the ligand before the arbitrary alteration

To find the optimum result *Swarm optimization (SO)* algorithm [36] is used in the search space. The engagements of a ligand in the search space are steered with the help of the knowledge of the finest locations of its nearest members.

In a docking algorithm, the *scoring function* [37] is a key component. The correctness of the algorithm is straightforwardly regulated by it.

Considering the most significant characteristics of a scoring function, the efficiency and the precision are the predominant. Computationally efficient and reliable scoring function can be considered as a perfect function.

Force field (FF) [37] is based on the breakdown of the ligand binding energy into separate molecular bindings. They are electrostatic energies, van der Waals (VDW) energies, and covalent bonding energies such as torsional, bending and stretching.

The easiest method is to consumption of the dielectric constant $\varepsilon(r_{ij})$ is used in DOCK, which is very similar to the force-field function, as following Equation 2.4.2.2:

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right)$$

Equation 2.4.2.2: Force Field Function used in DOCK

r_{ij} - The distance between ligand i and protein j
A_{ij} , B_{ij} - The van Der Wall parameters
q_j and q_i - The atomic charges
$\varepsilon(r_{ij})$ is typically considered as $4r_{ij}$, is the dielectric screening of the charge of water.

Several research groups have used geometric approach to the use of kinematics concepts from the robotics for the simulations. One had been called as the “haptic force-torque feedback device”. It’ll allow investigators to touch, control and connect a molecule in a virtual environment. A computer system has been used for the docking and Nano scale congregation. The techniques they are presented, have been used in the automated molecular object to offer the investigators a real time impression to cognize molecular connections. This tool can be used to estimate possible pharmaceutical drugs [38].

For the calculation of molecular interaction, potential energy and the collaborations among the atoms are characterized by the enthalpy generated among the atoms. While the protein and ligand atoms are malleable, it is commonly assumed that the protein is stiff while the ligand is a flexible-body [39]. The potential energy function has been used for the approximation of the VW terms and it is expressed as the following Equation 2.4.2.3.

$$E = \sum_{i=1}^{N_{lig}} \sum_{j=1}^{N_{rec}} \left[\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} \right]$$

Equation 2.4.2.3: Vander Waals Equation

Where N_{rec} and N_{lig} are the number of atoms in the receptor and the ligand, respectively, A_{ij}, B_{ij} are the van der Waals (VDW) attraction and repulsion parameters and r_{ij} is the space among the centers of atoms j and i. The VDW parameters for an atom is calculated based on the atom's VDW radius [40].

2.5. Applications of AI Techniques for Molecular Docking

Considering the molecular docking domain, the use of Artificial Intelligence Technologies is considerable. Robotics, Genetic algorithms and Artificial Neural Networks are the predominant among them.

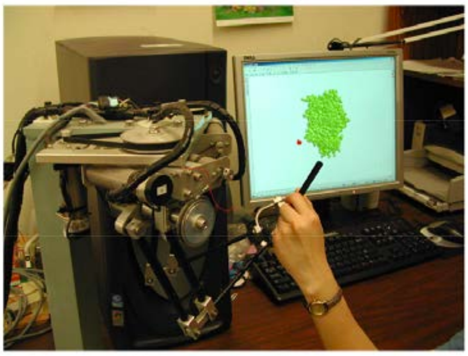


Figure 2.2: Haptic device and interface

Figure 2.2 shows a model of the computer based system for docking developed by the team. These techniques can use for the computer simulations to facilitate investigators a real-time artificial simulation for the visualization, manipulation and association of molecules in a virtual environment. In this approach they have used distance, geometric methods and energy minimization techniques using search algorithms based on van der Waals method like classical force field methods. But there are other techniques such as Distance geometry methods, Genetic Algorithms, Monte Carlo Methods, Molecular Dynamics, Point complementary methods and Fragment-based methods.

The Scripps Research Institute, Hewlett-Packard, Sandia National Laboratories and University of California collaboratively worked to find an innovative and strong simulated docking which forecasts the binding poses [41]. In Lamarckian model of genetics [42], this method is used. In this model, the object’s phenotype is reverse copied into the genotype to develop transmissible characters. They have considered these approaches called, the Lamarckian genetic algorithm, Evolutionary Algorithms and Monte Carlo. They could successfully equate the performance in dockings of random proteins-ligand test systems by knowing 3D structure. But the issue is it needed more computational resources, such as processing power and memory to do the calculations.

Machine learning and pattern recognition methodologies play a significant role in rational drug design [43]. Artificial Intelligence heavily incorporates statistical and machine learning theories. Also, neural networks, evolutionary computing and fuzzy modeling, like biologically motivated approaches have been used to increase the quality of the drug design.

The importance of the use of neural networks to train a classifier that is proficient of assigning beta, coil, helix or strand with the seventy percent of precession had been demonstrated by Rose and Sander. Their research had elaborated the applicability of AI techniques to improve the quality drug design. Find the poisonous properties of the drugs using laboratory experiments is not economically effective and efficient procedure and also harmful for the animals. It has opened the door to new researchers by showing the appropriateness of such AI techniques to design computer simulations rather than going for actual experiments.

Swarm Intelligence, Ant Colony Optimization, ANN, Fuzzy logic and Genetic Algorithm like different computational AI techniques can be used for drug discovery in silico. Drug discovery and design is an exhaustive process [44].

2.6. ParDOCK's Active Site Prediction Mechanism

ParDOCK is a tool to predict the active site of the proteins. One is an online tool called, ParDOCK - Automated Server for Protein Ligand Docking [45]. It has publish a known ligand and protein database as well [46].

Following Figure 4 is the flow chart for ParDOCK. It helps to understand the basic steps used for the molecular docking at a glace.

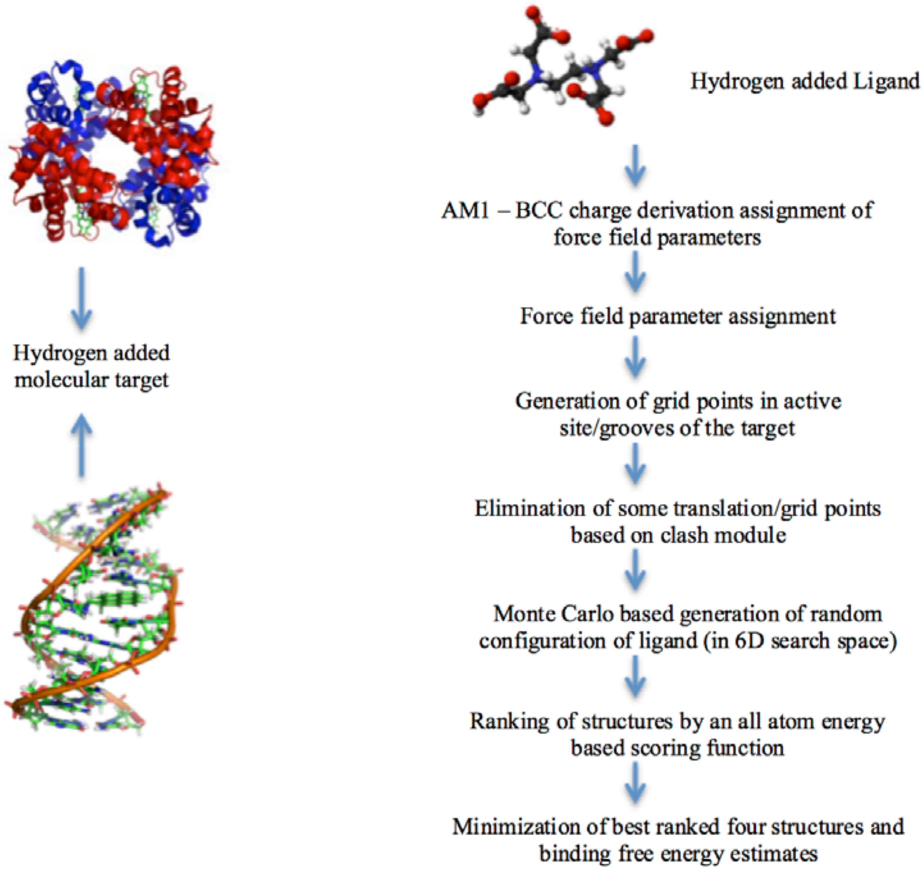

Hydrogen added Ligand
AM1 – BCC charge derivation assignment of force field parameters
Force field parameter assignment
Generation of grid points in active site/grooves of the target
Elimination of some translation/grid points based on clash module
Monte Carlo based generation of random configuration of ligand (in 6D search space)
Ranking of structures by an all atom energy based scoring function
Minimization of best ranked four structures and binding free energy estimates
Hydrogen added molecular target

Figure 2.3: ParDOCK docking flow chart [47]

2.7. Summary

In this chapter presents the literature of the existing molecular docking tools and their issues. And also in this chapter review and critically analyze their features to define the research problem clearly. Next chapter will describe the predominant theory, behind the research to solve the issues and discover novel approach to molecular docking which is the Multi-Agent Systems Technology.

Chapter 3

Multi-Agent Technology and Ontology in AI

3.1. Introduction

Previous chapter presented the literature for this research project. The second area of study from which the work presented here draws heavily can be broadly referred to as multi-agent systems technology combined with ontological engineering. Agents need set of rules to operate and these rules should store as the knowledge base of the agent. Ontology is the most famous technology currently used in agent related software tools [48]. It is a knowledge engineering technology and uses it to modeling the knowledge. Modeling knowledge consists in representing it in order to store it, to communicate it or to externally manipulate it.

A software agent is the one of the most latest trend is coming out of the Artificial Intelligence [49]. Agent technology uses to automate the software tools and the agent behavior trusts on the symbolic manipulation of proper models of knowledge fragments to perform meaningful operations that mimic intelligent competences.

3.2. Overview

Various research communities have developed different notions about what Multi-Agent Systems (MAS) technology area encompasses. This chapter will provide an overview over the MAS concepts and methods commonly employed in the research. It will begin by providing a short overview over the development of the MAS field over the past few decades and reviewing the basic terminology. A section will follow this on single agent and multi-agent decision-making processes, which form the basis for this work. The chapter will close with a summary of basic concepts of ontological engineering, which used to codify an agent behavioral rule set.

Ontologies can represent domain vocabulary, that convert the domain knowledge into concepts [50]. Ontologies are used in software development for the variety of purposes such sharing the knowledge and continuously update the knowledge [51].

3.3. What is Agent Technology?

Birth of Agent technology dates back to early 1990s [52]. Researchers believe that agents technology mimics the most natural way of problem solving by exploiting the team spirit [53]. The agent uses the message passing for solving problems over the algorithmic problem solving in conventional computing. Agents are necessarily tiny computational entities that require less computing power for their execution. As inspired from the nature, agents can be compared with small creates, such as bees and ants who demonstrate the marvel of team spirit in problem solving.

Agent is a computer system which has the competence of autonomously activates in dynamic, continuous and stochastic environments. The accuracy of the agents behavior is gauge using the environment it operates. These environments are classified as Deterministic, Episodic, Static, Observable and Discrete [54]. An agent perceives its environment through sensors and acts in this environment using actuators. It is represented in the Figure 3.3.

After the object oriented programming paradigm, agent technology become popular in the software industry [53]. The technology of an agent now has been applied in a miscellaneous areas of information technology such as e-commerce applications, computer networking systems, artificial intelligence and etc.

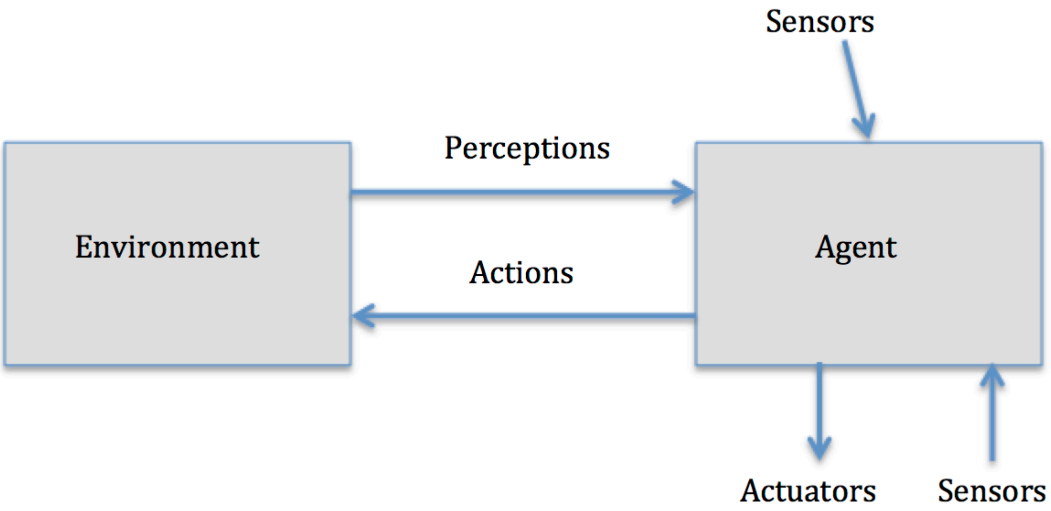


Figure 3.3: PEAS of agent

3.4. Characteristics of Software Agent

The multi-agent system definitely has some characteristics explained as follows. The below characteristics employed together to make multi-agent system more malleable to change.

1. Adaptability – The behavior of the agent can be dynamically changed according to the environment
2. Autonomy - An agent can behave according to its own rule set without any external intervention o a control
3. Collaboration – An agent interconnects and works supportively with other agents to work together on some task
4. Knowledgeable – Ability of the reasoning on some situations using the current available knowledge
5. Mobility – Ability to move from one location to another due to the size

3.5. Emergent Behaviour of Multi-Agent Systems

Emergent behavior is a common feature of all the complex systems. Individual components perform their actions and make decisions based on local information, while the entire system exhibits properties and behaviors that have durable global features [55]. In multi-agent systems, emergent behavior plays a significant role. Using the agent simulations, we can see the emergent behavior of the agents. In this research, I have reported an experimental study of the self organized agents' emergent behaviors in the molecular docking domain.

3.6. Need of Ontology in Agents

Agents want some shared understanding of the sense of the domain that they are operating. The knowledge of the domain provided by a specialized constituent called ontology. It postulates the entities and their associations in a specific domain. Ontology is typically constructed using a schema definition language, such as RDF [56] and OWL [57]. Designing ontologies is not an easy-going task so it is a tedious task. Lots of software tools already had been designed for exploratory and designing ontologies. Protégé [58] is the famous and frequently used tool, among those tools.

Ontologies can apprehend both the semantics (meaning) and structure of the domain. By apprehending the significant affiliations among the tables or the rows in a database table, it can consolidate database concepts and the keywords using ontologies [59]. Ontology is a terminology of concepts, tasks, their relationships and properties. Ontologies are heavily used by the agents for the understanding of its operational domain knowledge. Knowledge sharing and reuse is heavily encouraged to use ontologies [60].

3.7. Distributed Decision Making

Multi-agent systems (MAS) describe the association between numerous decision-making agents. They are intended to solve hard and complex problems in uniting the effort of each individual agent [61]. More specifically, MAS is also called distributed artificial intelligence [62] stressing the fact that complex problems are solved by a population of different agents each of them having its own skills, information, and preferences. Most of the discussion is still descriptive and experimental, considering, e.g. communication and coordination patterns.

Starting with a complex task, MAS typically decomposes the problem into several less complex jobs each of which being individually treated by a separate agent. These agents may roughly be described as depicted in Figure 3.7.

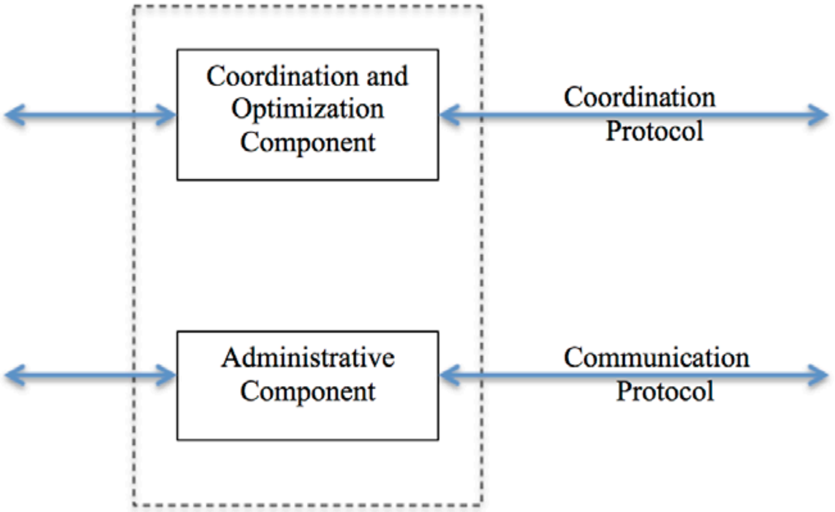


Figure 3.7: General Structure of Agent [61]

In the previous Figure 3.7 shows the main components of the structures of an agent as described below:

- Administrative component - It is responsible for the communication process
- Coordination and optimization component - It is providing the rules the agent employs to solve its specific problem.

3.8. Use of Ontologies in AI

According to the literature from Artificial Intelligence, agents can operates using the inference knowledge from the outside environment. But it needs an initial knowledge and some kind of rule set for their operations. Ontologies are the most efficient way to supply the outside knowledge for the operation of the agents. Ontologies are the models of the domain knowledge converted to entities, properties and relations.

Ontology is a machine readable model of some domain area. It is a some form of a semantic inferred network of a set of graphs its nodes are concepts and its arcs represent the relationships or the associations between the concepts [63]. Following Figure 3.8 is an example for the Ontology of a Bottle domain.

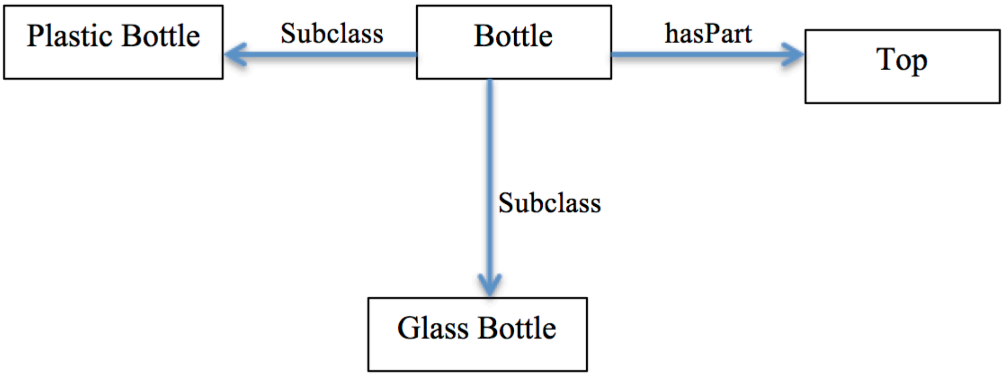


Figure 3.8: Structure of Ontology

Knowledge sharing and exchange is predominantly essential in multi-agent systems (MAS). An agent is typically described as an entity with some degree of autonomy that performs some set of processes based on what he perceives. An agent has intelligence at some level. So it has to have some knowledge about its target. The entire MAS is designed for the accomplishment of attainment of their goals which are hard to accomplish by an individual [64]. In multi-agent systems, an agent regularly collaborates with other agents. Because of these reason agents have some communicative skills.

An agent must be able to do the following tasks to communicate with the other agents:

- Deliver and receive messages - Agents have the capability to accept and deliver messages over the permitted environment limit
- Parse the messages – They are able to parse the content of the message and broken down into small parts of the sentence (syntactic level)
- Understand the messages – They are able to understand the parsed parts of the sentence [65]. The ontology has been used to understand the sentence and it is shared among all the agents sometimes (semantic level)

3.9. Knowledge Representation using Ontology in Chemistry

Processing of computational data is crucial for the scientists in Chemistry field [66]. Ontologies program expert domain knowledge in a structurally organized machine-process able format. The Royal Society of Chemistry (RSC) is a learned society in the United Kingdom with the ambition of enhancing the chemical sciences [67]. RSC started to build their own subject classifications covering preferred areas of chemistry using ontologies. The first three ontologies that they are making presented are:

- RXNO - name reaction ontology [68]
- CMO - chemical methods ontology
- MOP - molecular processes ontology [69]

Chemical Methods Ontology (CMO) [70], the chemical methods ontology that depicts methods used to gather data in chemical experiments, such as mass spectrometry and electron microscopy. It is planned to be corresponding to the Ontology for Biomedical Investigations (OBI) [71].

One such ontology for the chemical domain is ChEBI [72]. ChEBI has constantly grow progressively in content, and has added several new features. Furthermore, it is combining all users requested combinations, their explanation efforts have highlighted immunology, natural products and metabolites in many species [72]. Functional group ontology (FGO) [73] is another example for the chemical ontology.

3.10. Summary

There is a good trend for the MAS technology to address complex problems. According to the recent researches MAS has been proven its competencies to operates in various complex environments to achieve their targets. Although still there is no evidence of applications in the domain of molecular docking supported by MAS. In this chapter we have discussed the foundations of MAS and the general model and finally the possibility in using for molecular docking. Application of MAS will be discussed in detail in design and implementation chapters. In this chapter also presents the use of ontological engineering to the agent communication and the distributed decision making of MAS. Next chapter will describe the basic concepts of molecular docking that used for this research.

Chapter 4

Molecular Docking

4.1. Introduction

Previous chapter presented the use of ontological engineering to the agent communication and the distributed decision making of MAS. This chapter is meant to give a general overview of the molecular docking concept. A basic understanding of molecular docking will be presented in the simplest and easiest manner possible.

The three-dimensional structures known may be represented to show different views of the structures. With complex molecular mechanics' programs, it is possible to overlay one structure to another. The same approach is used to cover the three dimensional structure of a potential drug on its probable target site.

4.2. Steps of Molecular Docking

Current molecular docking tools follow the common steps as bellow: [74] [75]

1. Complex coordinates should be evaluated (i.e. from the PDB).
2. Remove unnecessary atoms from the complex (delete all the water and all non-interacting ions).
3. Insert the required hydrogen atoms to the complex.
4. Clean the complex
5. Isolate the minimized complex in protein and ligand.
6. Formulate the docking appropriate files (pdb files).
7. Analyze the docking results.

4.5. Lock and Key Theory

Emil Fischer suggested a model called the "lock-and-key model" that explained how biological systems function [76]. A ligand penetrates into the active site of a protein is similar to the key mounts into a lock. Biological 'locks' have unique stereo chemical features that are necessary to their function. Following Figure 4.3 depicts the Lock and Key concept.

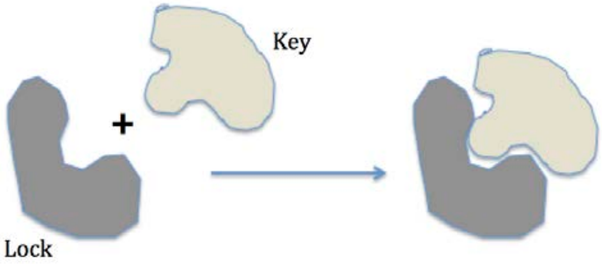


Figure 4.3: Lock and Key model

4.5 Different Types of Interactions

The interactions between particles to be the consequence of forces between the molecules contained by the particles are divided into four categories:

- Electrostatic forces - Forces with electrostatic origin is due to the charges residing in the matter (charge-charge, charge-dipole and dipole-dipole)
- Electro dynamics forces - The most widely known is probably the VDW interaction.
- Steric forces - Steric forces are produced by entropy
- Solvent forces – It is due to the physical variations of the solvent

4.5 Rule to Choose Appropriate Ligands

4.5.1 Lipinski's rule of five:

An orally active drug should not violates the following mention rules: [77]

- Should not exceed 5 hydrogen bond donors
  - The sum of the oxygen–hydrogen and nitrogen–hydrogen bonds
- Should not exceed 10 hydrogen bond acceptors
  - Total number of oxygen or nitrogen atoms
- A molecular mass should be fewer than 500 daltons
- An octanol-water partition coefficient log P should not not huger than 5

According to this rule, had chosen the appropriate ligands for the docking.

4.6. Energy Minimization

Most of the time the energy minimization analyses are made for the ligand. Ligand can exist in various conformations, which means the shapes or the structures. The conformation, which has minimized energy, will be the most stable conformation of the ligand [78].

There are two types of energies:

a) *Electrostatic Potential Energy* [79]: It is Electrostatic Potential Energy a pair wise total of columbic interactions as described in equilibrium. Please see the Equation 4.6.1 as follows:

$$E_{electrostatic} = \sum_{pairs-nonbonded} \frac{q_i q_k}{D_{r_{ik}}}$$

Equation 4.6.1: Electrostatic Potential Energy

q_i and q_k - the one point charges

r_{ik} – distance between the point charges

D - Coulomb's constant

b) *Vander Waals Potential Energy*: For general behavior of non-bonded interactions is regularly modeled by the following Equation 4.6.2.

$$E_{VanderWaals} = \sum_{pairs-nonbonded} \left(\frac{A_{ik}}{r_{ik}^{12}} - \frac{C_{ik}}{r_{ik}^{6}} \right)$$

Equation 4.6.2: Van der Waals Equation

r - The distance among two atoms having charges q_i and q_k

Vander Waals potential it expresses the bonding energy using the constants A and C that depends on the type of the atoms associates with the bonds. Values of A and C can be found by the diversity of methods.

4.7. Bond Energy

The energy E is depending on the atomic array of R as following Equation 4.7.1. R can be derived using the 3d structures of the PDB files.

E_{bonded} - sum of internal, or bonded

$E_{non\text{-}bonded}$ - sum of external or non bonded

$$V(R) = E_{bonded} + E_{nonbonded}$$

Equation 4.7.1 Formula to calculate bond energy

The E_{bonded} is the total of the below explained terms as the following Equation 4.7.2:

$$E_{bonded} = E_{strech} + E_{bend} + E_{rotate}$$

Equation 4.7.2 Formula to find bonded energy

There are three types of atom movements as shown in the following Figure 4.7.1.

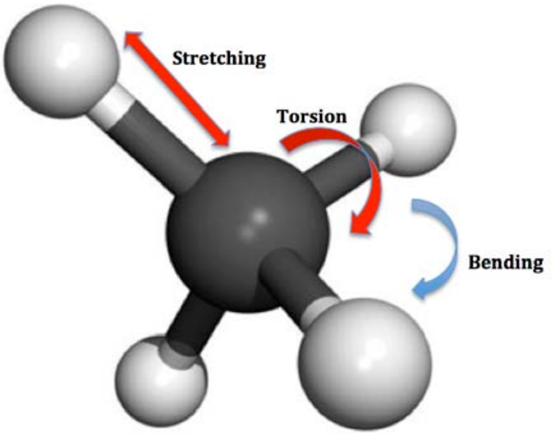


Figure 4.7.1: Bond interactions in Methane Molecule

4.8. Stretching

This is the approximation of bond energy as a function of b_0 and K_b, decides the strength of the connection as the following Equation 4.8.1.

b_0 - The ideal bond length

b – The bond length

K_b - The force constant

$$E_{streching} = \sum_{1,2\,pairs} K_b (b - b_0)^2$$

Equation 4.8.1: Bond energy for stretching

4.9. Bending

Bond energy for the bending can be calculated using the following Equation 4.9.1.

θ_0 – Ideal harmonic potential

θ- Harmonic potential

K_θ - Based on substance type of atoms

$$E_{streching} = \sum_{angles} K_\theta \left(\theta - \theta_0\right)^2$$

Equation 4.9.1: Bond Energy for Bending

These two expressions explain the deviation from an ideal geometry, efficiently. The summation of them should be approximately close to zero.

4.10. Rotating

The torsion angles possible function which represents the occurrence of steric obstructions among atoms disconnected by the below sections described three types of covalent bonds. The motion connected, which is explained by the coefficient of symmetry (n=1,2,3) and the dihedral angle. From the following Equation 4.10.1 can be used to calculate the energy of rotation of the bonds.

ϕ - Dihedral angle each atom axis

Coefficient of symmetry n=1,2,3

$$E_{rotation} = \sum_{1,4\,pairs} K_\phi \left(1 - \cos(n\phi)\right)$$

Equation 4.10.1: Bond energy for bending

4.11. Force Field Function

The fundamental functional form of potential energy in molecular mechanics contains bonded terms for interactions of atoms that are connected by the different types of bonds. Following Equation 4.11.1 is used to calculate the total of the above-mentioned energies:

$$E = \sum_i \sum_j \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}} + \frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right)$$

Equation 4.11.1: Bond energy for bending

r_{ij} - The space among ligand atom i and protein atom j

A_{ij}, B_{ij} - The Van der Walss parameters

q_i , q_j – Charges of the atoms

$\varepsilon(r_{ij})$ - screening effect of water

4.12. Summary

This chapter provided a conceptual background and discusses selected theoretical foundation for the molecular docking. The next chapter will be described the molecular docking approach using multi-agent technology.

Chapter 5

MAS Approach to Molecular Docking

5.1. Introduction

In the previous chapters we defined the research problem as the inefficiency in manual molecular docking and the existing solution by describing why Multi-Agent Technology should be a potential technology to develop novel solution for molecular docking. This chapter presents our approach by describing the inputs, outputs, process, features and users for novel solutions for Multi-Agent based molecular docking. Here we would postulate our hypotheses with the help of MAS. Our new intelligent, multi-agent systems technology based molecular docking solution has been named as 'NanoAgents'.

5.2. Hypothesis

We postulate the hypothesis that the manual molecule docking can be automated by multi-agent systems technology.

5.3. Building Blocks of Approach

Since we are to articulate the proposed model to testify the hypotheses, the sequence in the model happens to be the process in our approach. Let us discuss those steps in detail.

5.3 Inputs

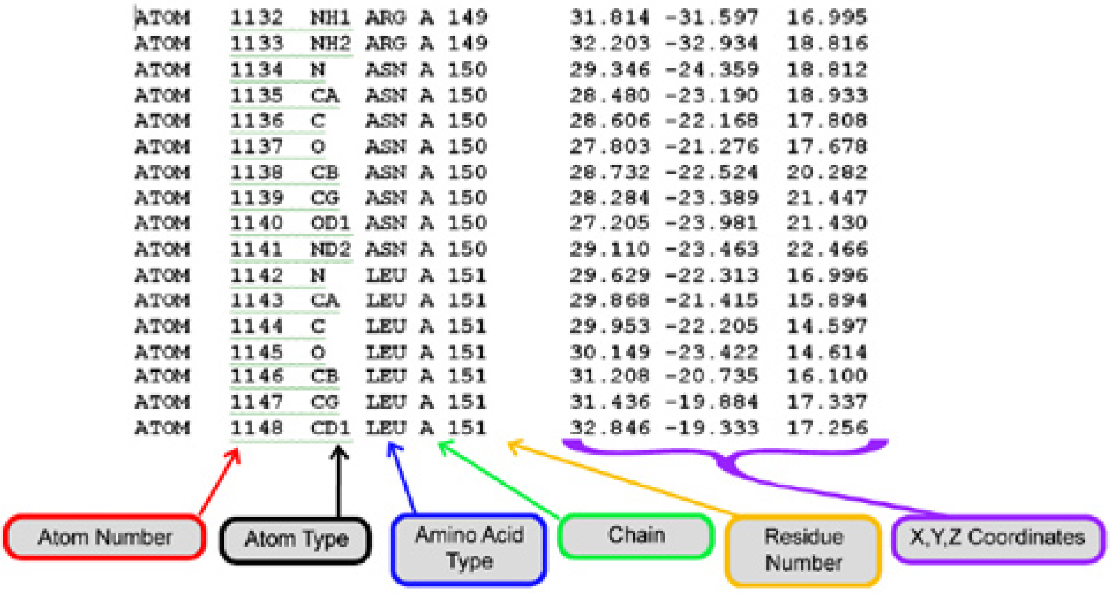
Multi-Agent system for Molecular Docking has been designed to accept multiple inputs coming from different entities of the Molecular Docking process. The following are the main inputs to the system. There are two major types of inputs to the system. Table 5.3.1 shows the inputs from corresponding entities.

Input	Entity (Agent)
Rules for the binding	Ontology
Description of protein	Protein Agent
Description of ligand	Ligand Agent
Description of bonds	Energy Calculation Agent
3D coordinated of the protein	Active Site Finding Agent
3D coordinated of the protein and the ligand	Surface Matching Agent
3D coordinated of the protein and the ligand	Cavity Finding Agent

Table 5.3.1: Inputs for the Molecular Docking System

5.4 Protein Data Bank (PDB) File

The Protein Data Bank (PDB) format specifies the standard depiction of the protein and ligand cordinate data derived from X-ray diffraction [80] and NMR studies [81]. The pdb format correspondingly specifies various components of the molecules such as nucleic acids, hydrogen bonds, water molecules and iorns. This atomic connectivity of the molecule can be described in the .pdb format. Following Figure 5.4.1 shows an example PDF file content.



```
ATOM    1132  NH1 ARG A 149       31.814 -31.597  16.995
ATOM    1133  NH2 ARG A 149       32.203 -32.934  18.816
ATOM    1134  N   ASN A 150       29.346 -24.359  18.812
ATOM    1135  CA  ASN A 150       28.480 -23.190  18.933
ATOM    1136  C   ASN A 150       28.606 -22.168  17.808
ATOM    1137  O   ASN A 150       27.803 -21.276  17.678
ATOM    1138  CB  ASN A 150       28.732 -22.524  20.282
ATOM    1139  CG  ASN A 150       28.284 -23.389  21.447
ATOM    1140  OD1 ASN A 150       27.205 -23.981  21.430
ATOM    1141  ND2 ASN A 150       29.110 -23.463  22.466
ATOM    1142  N   LEU A 151       29.629 -22.313  16.996
ATOM    1143  CA  LEU A 151       29.868 -21.415  15.894
ATOM    1144  C   LEU A 151       29.953 -22.205  14.597
ATOM    1145  O   LEU A 151       30.149 -23.422  14.614
ATOM    1146  CB  LEU A 151       31.208 -20.735  16.100
ATOM    1147  CG  LEU A 151       31.436 -19.884  17.337
ATOM    1148  CD1 LEU A 151       32.846 -19.333  17.256
```

Figure 5.4.1: Structure of a PDB file [82]

The rules for the operation of the Protein and Ligand agents are specified as the PDB file input to the system. For the calculation of the surface of the protein and the volume of the ligand, these coordinate values are used.

5.5 Chemical Ontologies

In this research continue to investigate on the illustration of molecular space structure and their attributes using entities and properties based methodology to gather knowledge of atomic connectivity and molecular dynamics. Dumontier Lab [83] provides chemical ontologies such as Atomic Ontology. Following 5.5.1 diagram shows the ontology used to retrieve details of the atoms for the molecular dynamics related calculations such as activation energies.

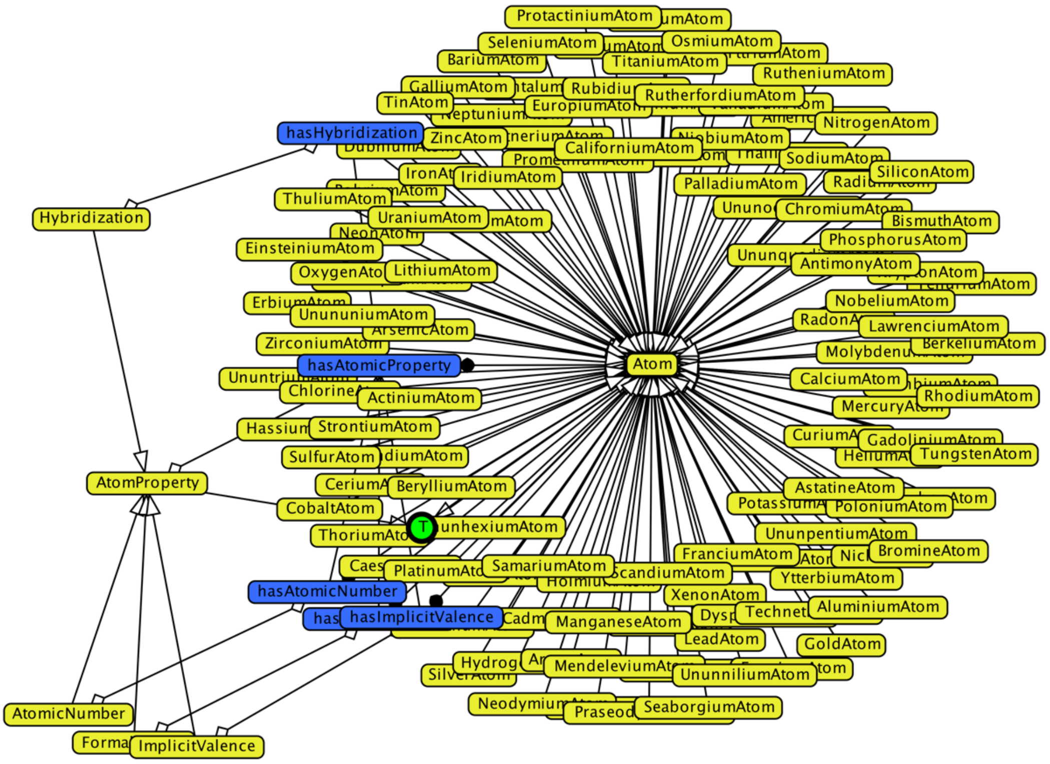


Figure 5.5.1: Atom ontology in Protégé [58]

5.6 Outputs

The outputs of the system will be the docked molecule, the matched ligand and the prediction of the actual existence of the docked molecule based on the calculated energy. These ligand outputs are coming as the rational drug. But that drug should be clinically evaluated before the actual consumption.

5.7 Process for the Molecular Docking

The system will use the PDB files of the database of molecules (both ligand and proteins) and the ontology as the inputs to generate the rational drugs. In this process, two main agents, namely protein agents and ligand agents are defined in the system. The knowledge requires to these agents to operate are stored in the common domain ontology and personal PDB files.

Get the atomic coordinates from the PDB as the input to the system and prepare the protein structure. As the next step, add the misplaced hydrogen bonds to the protein molecule. Then prepare the ligand as the protein by removing unnecessary atoms from the structure. Afterwards perform the docking and analyze the docking results.

There are separate ontologies to store the atom colors and charges for the atoms/elements. Agents get the help from these ontologies to find the best ligand agent to bind with the protein agent.

Following Figure 5.7.1 shows the entire process of the automated molecular docking.

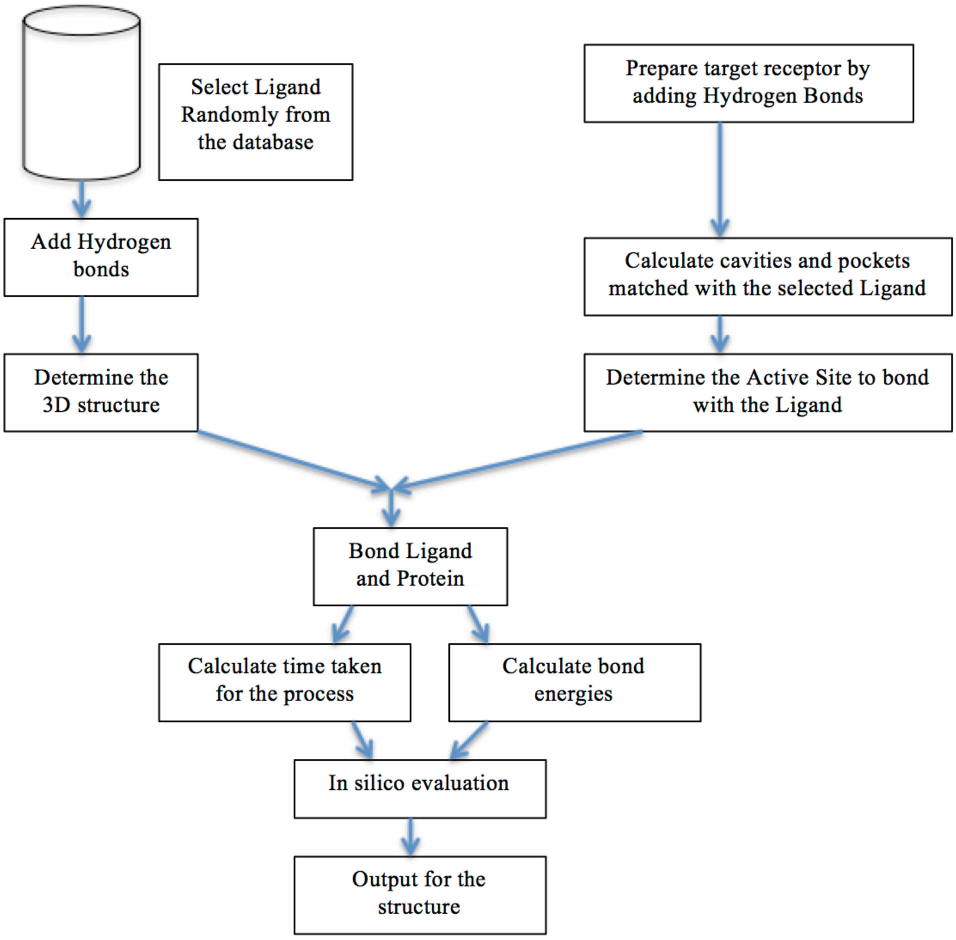

Select Ligand Randomly from the database
Prepare target receptor by adding Hydrogen Bonds
Add Hydrogen bonds
Calculate cavities and pockets matched with the selected Ligand
Determine the 3D structure
Determine the Active Site to bond with the Ligand
Bond Ligand and Protein
Calculate time taken for the process
Calculate bond energies
In silico evaluation
Output for the structure

Figure 5.7.1: Flow diagram for the drug design

5.8 JADE – Java Agent Development Framework

JADE (Java Agent DEvelopment Framework) [84] is a software Framework completely implemented in the Java language. It makes easy the accomplishment of multi-agent systems through a framework that compatible with the FIPA [85] specifications. JADE is implemented using the Java language.

5.9 Jena – A Semantic Web Framework for Java

Jena [86] is a free and open source framework to construction Semantic Software applications. It offers a development environment for RDFS [87], OWL [88], RDF [56] and SPARQL [89] and has a rule - based inference engine to derive knowledge from ontologies. Ontologies are built as classes and the relationships in between different concepts, as I described in section 3.2. Jena is a framework that enables ontologies to be aggregated and reused in different applications. It can be used to reasoned through the ontologies to derive meaningful data [90]. The Jena framework also helps to process data automatically and use by the agents. It is discovering new relationships in various concepts represented as classes [91].

5.10 SciPy - Scientific Computing Tools for Python

SciPy [92] is the most popular scientific programming language that is used for the mathematics, scientific and engineering applications..

5.11 OWL – Web Ontology Language

For the manipulation of the knowledge instead of displaying the extracted information, Web Ontology Language (OWL) is playing the better role in the current context [93]. OWL enables better machine readability of the information content. It is much better than the RDF and XML, by providing extra feature of revealing the meaning of the existing information and the reasoning on the information.

5.12 Jmol - An Open-Source Java Viewer for Chemical Structures

Jmol [94] is an open-source and free atomic viewer for three-dimensional molecular structures. It is used as a research tool in biochemistry and chemistry. Jmol has been implemented in the Java programming language. So Jmol can executes in any operating system.

5.13 MySQL - Database Management Tool

MySQL [95] is the world's second most widely used open-source relational database management system. It provides lots of graphical user interface to manage tables and database schemas easily.

5.14 Features – Non Functional Requirements

The following features are available in the system.

1. Rotate/translate molecules
2. Add/delete molecules
3. Minimal resource usage
4. Efficient and fast discovery of drugs
5. The development process is marginal special using JMol, Jade and Jena like frameworks
6. Load the ligand and the protein structure into the same window concurrently
7. User friendliness
8. Accurate and faster calculations of the surface and geometry

5.15 Users of the Molecular Docking

Researches, students in the Chemistry and Biology field can be utilized this tool to do their experiments. Pharmacologist can apply this tool to discover rational drugs.

5.16 Summary

In this chapter presented the novel approach to design **“NanoAgents”** software tool using its inputs, output, users and its fetures. Next chapter will describe the entire design of the software tool.

Chapter 6

MAS for Molecular Docking

6.1. Introduction

Previous chapter describes the approach of the NanoAgent. This chapter will describe the design of the multi agent system architecture which is incorporating suggested molecular docking tool. There it would initially discuss the relevant technologies that use for the realization of each component of our system altogether the conceptual basis behind the system. Then we would discuss the modules in system in detail.

The primary goal of this research is the automation of manual molecular docking to enable faster and more cost-effective molecular modeling with the help of Multi-Agent Technology. NanoAgents system design comprises of different system components for different tasks such as, find the effective site of the protein, find the shape of the hollow of the target, search matching ligand to the cavity, find the correct orientation of the cavity, calculate binding energies and energy minimization and evaluate the rationality of the selected ligand or the drug.

6.2. System Integration of the NanoAgents

In this section, we describe the system architecture and the integration for the NanoAgents for molecular docking and energy minimization. The overall system architecture and the control flow are similar for both the applications, except for the specific control and communication structures, which are discussed separately.

Molecular docking is a natural process, which occurs within seconds in a cell. Due to the limitations of current computational, experimental methods in molecular docking, 3D structures of docked molecular complexes are rarely available. But knowledge of the separated molecules in 3D (PDB files) such as ligands and proteins, are only weakly informative if we do not know how to assemble them each other.

To retrieve successful docking results, currently scientists have to use various combinations of tools. Following are the various properties need to be checked using different kind of tools:

- Geometries (torsion angles, bond length and bond angles)

- Energies (activation energy and heat of formation)
- Properties (diffusion, volume, viscosity and surface areas)

Manual molecular docking process can be considered as an inherently complex system because of the use of different kinds of software tools. There is no all-in-one software tool available for the molecular docking. Develop Multi-Agent System (MAS) solution to automate molecular docking process. Following are the basic steps for the manual molecular docking:

- STEP 1 – Preparation of ligands
- STEP 2 – Preparation of proteins
- STEP 3 – Setup ligand protein docking calculations
- STEP 4 – Evaluation of results

For every task there are different tools need to be used to achieve complete molecular docking results. And also need to install lots of software tools for manual molecular docking. Following are the software is used for the manual molecular docking.

- Python [96] – Scientific calculation
- AutoDock [18]– Edit bonds, shape of the molecule
- MGTools [20] – Graphic library for 3D object rendering
- Discovery Studio [97]– Docking tool
- PyMol [98] – Visualize the molecules

Using multi-agent system technology combined with ontological engineering enhances the development of all-in-one software tool development for molecular docking. For each task of the molecular docking I have specified the agent. Agents are little and have reduced set of rules for the operation. But they are working collaboratively to achieve the docking process.

6.2.1. System Architecture

The main architecture behind this solution has the Request Agents, Resource Agents, Message Space and the Domain Ontology. Following Figure 6.1 depicts the entire system architecture of the automated docking solution. This system is based on the blackboard system. It is an artificial intelligence design pattern based on the agent technology. AI researchers conceived the concept called the blackboard architecture [99] in the 1970's. Blackboard architectural model similarly has the components as the above-mentioned request, resource agents and the domain ontologies. A message space, which is the blackboard, has all the solutions updated by the agents involved in the system. The common knowledge base which is called as the "blackboard", is continuously restructured by the miscellaneous group of knowledgeable agents who is specialized in each areas of the entire process.

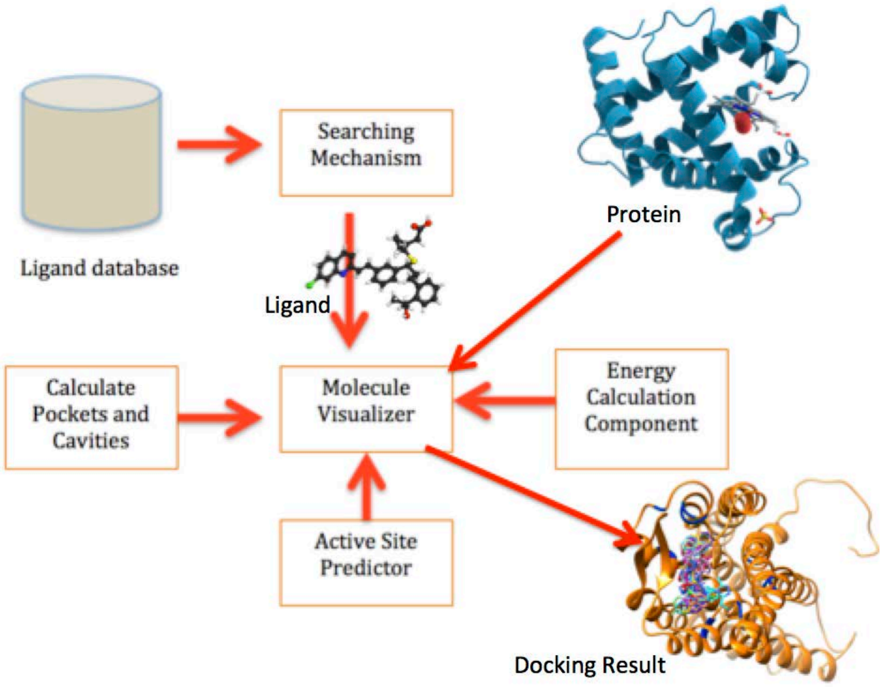


Figure 6.1: System Architecture

Typically, it initiates with a problem description and terminate with the appropriate solution. Each knowledge source or the specialist (agent) renovates the blackboard with an incomplete solution, when its internal restrictions match the blackboard state. It is very much similar as the experts work together to solve the problem. The blackboard model was designed to handle imprecise and complex problems, where the solution is the sum of its fragments of partial solutions. Molecular docking can be considered as a complex problem which the solution is the sum of different operations, such as energy calculations, active site finding and cavity finding.

6.3. Searching Mechanism

It is essential to choose optimum ligand for the docking process. There is a rule to choose ligands called, “Lipinski's rule of five” (see section 4.5 in the thesis for more information). Inside this tool there is a database of selected ligands that can be chosen for the docking. After selecting a ligand from the database, the relevant PDB file loaded into the LigandAgent to start the docking process. At that point LigandAgent get the knowledge to do the operation and also it gets the help from the external ontology as well.

In rigid-body docking, the search space is constrained to three translational degrees of freedom and three rotational. Following Figure 6.3 shows the three degrees of freedoms of the molecular motions.

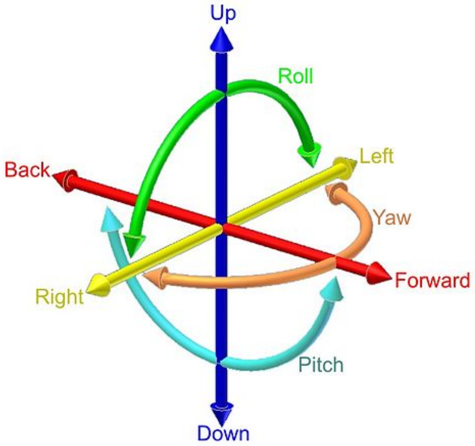


Figure 6.3: Three degrees of freedom

6.4. Calculate pocket and cavities

Identification of the volume of the cavities can be considered as the preliminary steps in molecular docking in drug design. There is a separate agent who is allocated for calculates pockets and cavities in the 3D structure of the Ligand and Protein. It explores the PDB files of Ligand and Protein to find the spaces on the surface. It automatically locates and gauging the hollows, is based on mathematical methods, such as alpha shape [100] when determination of the cavity in the macromolecules.

6.5. Active site predictor

Even if the cavity and pockets found need to find the actual effective binding site on the protein for the docking. The active site is an area in the receptor (protein) that the protein and ligand molecules bond and effect to a natural reaction. This happens at the time a ligand strikes with the protein and fits into the effective range of the protein. Active site made of chemically effective atoms such as hydrogen that can be interacted with ligand and dock. The active site is typically found in a cavity of the protein.

Ligand bind to the effective range or the active site of the protein through hydrogen bonds. Hydrogen bonds in the active site will perform as acceptors or the donors of protons on the ligand to enable the reaction. The active site predictor agent also uses the external ontologies for the knowledge base besides PDB file codified rules of the atoms. The active site predictor agent gets the help from pocket and the cavity finding agent to predict the best active site for the docking.

6.6. Molecular Visualizer

To enable the usability of the tool, need to visualize the ligand, protein and the docked molecule in 3D space. For that need to have separate component called “Molecular Visualizer”. It also has the knowledge for the docking of the ligand and the protein too. It uses all the knowledge shared by the different other agent modules. It uses for displaying the molecules. And also this act as the ‘Blackboard’ of the entire agent system.

6.7. Energy Calculation

The tool can come up with a structure of docked molecule using Ligand and Protein. But it is essential to check whether there can be a docked molecular structure in the nature. To check the stability and the actual existence of the docked molecule needs to perform some energy calculations. In this module it calculates the energies and it predicts the existence of the docked molecule.

The final free energy of binding will depend on the overall balance of these factors. The EnergyCalculation agent has all the knowledge for the energy calculations and also predicts the stability of the docked molecule.

The interaction forces between two molecules can be divided into following categories:

1. Electrostatic interactions
2. Hydrogen bond interactions
3. Van der Waals interactions
4. Hydrophobic forces

All the formulas to calculate the above energies are codified into this agent module.

6.8. Summary

In this chapter we have summarized the design of NanoAgents molecular docking tool based on the ‘Blackboard Architectural’ model. The detailed description of each agents have been described in this chapter. The next chapter will introduce the implementation related details.

Chapter 7

Implementation of MAS for Molecular Docking

7.1. Introduction

Previous chapter presented the design of the molecular docking system. In this chapter, let us discuss the software realization of our designed system. In this chapter the modules we discussed in the previous chapter will be further subdivided and analyzed at class level. The system flows and algorithm level realizations will be discussed. Implementation comprises of the implementation of the architecture of the multi-agent systems based molecular docking system.

7.5. How to find matching ligand to the protein

DOCK is another molecular docking program currently using in the drug discovery field. Following are the steps it follows to find the matching ligand, achieve the docking [101]. And also we have followed their approach similarly.

1. Inside the cavities of the protein that can be considered as the active site, a set of spheres is populated
2. The volume of the ligand is represented by the set of spheres
3. The volume of the small molecule is denoted by spheres in the ligand (see Figure 7.5.1)

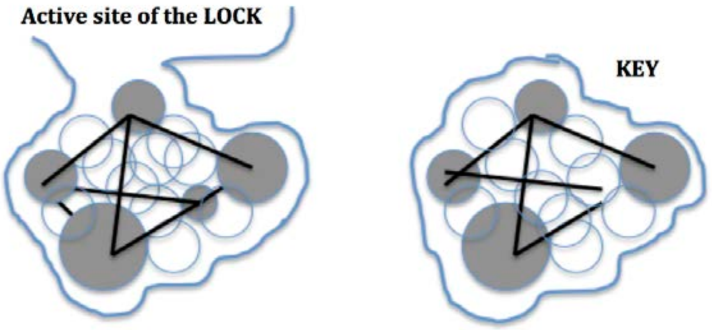


Figure 7.5.1: DOCK program algorithm: Match the spheres shown in gray (distances between the spheres are used for scoring)

7.3. Methods to find Active Sites of the Protein

7.3.1. Alpha Spheres

Alpha shape [102] of a set of weighted vertices. It is a theory from computational geometry and it is heavily used in computational chemistry. The sphere that associates some atoms on its edge and surrounds no internal atoms is called a contact sphere. An alpha sphere is a distinctive case of a contact sphere. An alpha sphere has the contacts four atoms on its boundary and contains no internal atoms. Following Figure 7.3.1 represents the Alpha sphere structure.

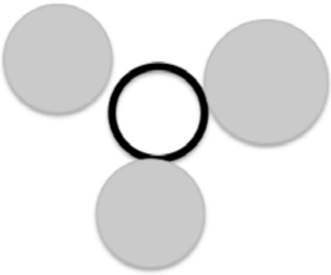


Figure 7.3.1: An Alpha sphere – In 2D space that contacts three atoms of dissimilar radius

7.4. Finding cavities in the protein

One of the many properties we can study is a molecular cavity, where the cavity is understood as a free space inside a molecule. After reviewing the literature, we are usually interested only in a certain subset of cavities. Voronoi diagram as shows in Figure 7.4.1 is then used to achieve the task of finding cavities for varying radiuses [103]

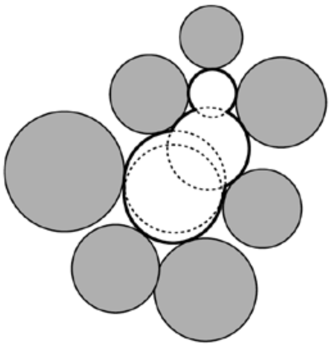
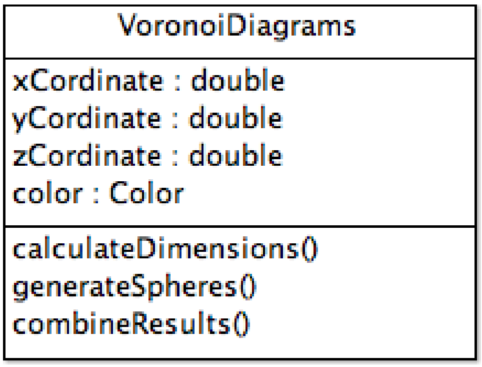


Figure 7.4.1: Cavities formed by Gray Atoms

Following Figure 7.4.2 shows the class diagram for the implementation to calculate Voronoi Diagrams.



VoronoiDiagrams

xCordinate : double
yCordinate : double
zCordinate : double
color : Color

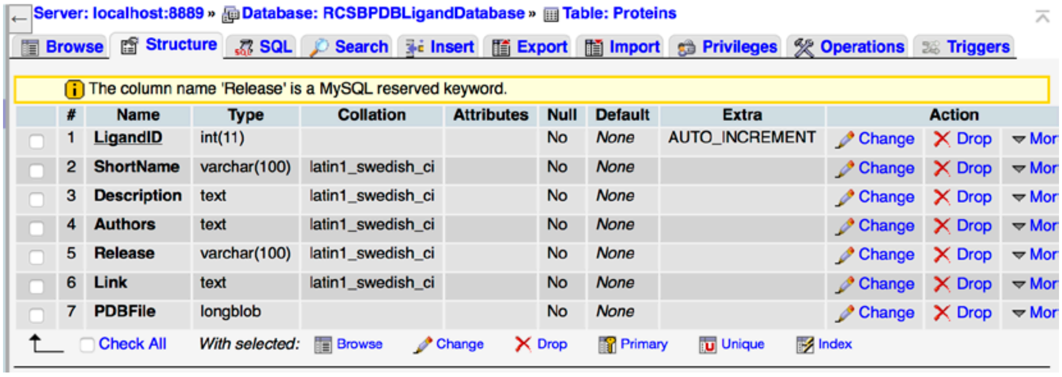
calculateDimensions()
generateSpheres()
combineResults()

Figure 7.4.2: Class diagram for Voronoi Diagrams

7.5. Data for Ligands and Proteins

The necessity of the offline data for the proteins and ligands is a must for a docking tool. For that AgentDock has features to insert proteins and ligand data into the tool. RCSB Protein Data Bank [104] was the main resource to retrieve correct information for the proteins. From RCSB database had retrieved the correct data for the docking tool.

Following Figure 7.5 shows the database table structure to be used to store protein data.



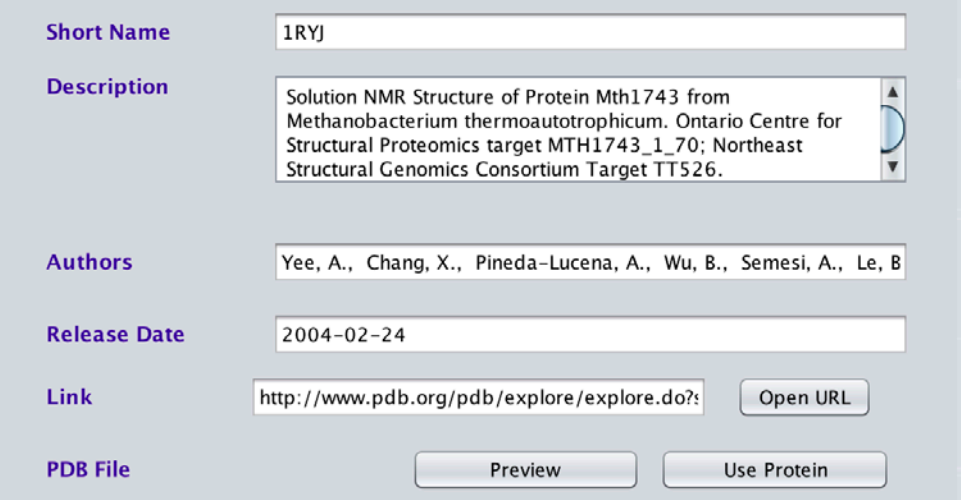
Server: localhost:8889 » Database: RCSBPDBLigandDatabase » Table: Proteins
Browse
Structure
SQL
Search
Insert
Export
Import
Privileges
Operations
Triggers

The column name 'Release' is a MySQL reserved keyword.

	#	Name	Type	Collation	Attributes	Null	Default	Extra	Action		
□	1	LigandID	int(11)			No	None	AUTO_INCREMENT	Change	Drop	Mor
□	2	ShortName	varchar(100)	latin1_swedish_ci		No	None		Change	Drop	Mor
□	3	Description	text	latin1_swedish_ci		No	None		Change	Drop	Mor
□	4	Authors	text	latin1_swedish_ci		No	None		Change	Drop	Mor
□	5	Release	varchar(100)	latin1_swedish_ci		No	None		Change	Drop	Mor
□	6	Link	text	latin1_swedish_ci		No	None		Change	Drop	Mor
□	7	PDBFile	longblob			No	None		Change	Drop	Mor

Figure 7.5.1. Protein table structure in phpMyAdmin [105] interface

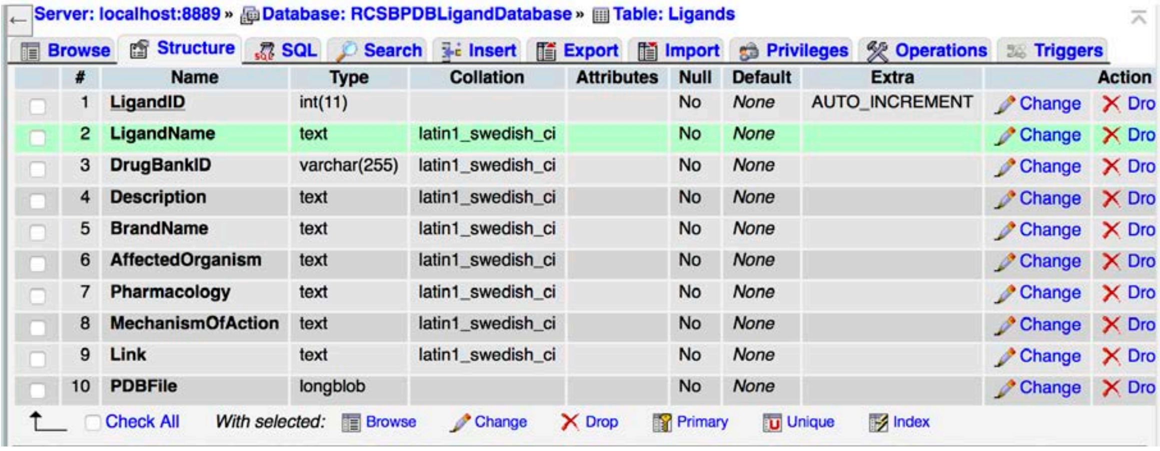
In the Protein table the PDB file structures stored as BLOB [106] file format. See Appendix A to view tool screens. Following Figure 7.5.2. shows data for a record from that table.



Short Name	1RYJ
Description	Solution NMR Structure of Protein Mth1743 from Methanobacterium thermoautotrophicum. Ontario Centre for Structural Proteomics target MTH1743_1_70; Northeast Structural Genomics Consortium Target TT526.
Authors	Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B
Release Date	2004-02-24
Link	http://www.pdb.org/pdb/explore/explore.do?s
PDB File	

Figure 7.5.2. Data for protein 1RYJ

The ligands or the data for the small drugs are retrieved from the popular web resource European Bioinformatics Institute (EBI) Database [107]. Following Figure 7.5.3 shows the database table structure to be used to store ligand data.

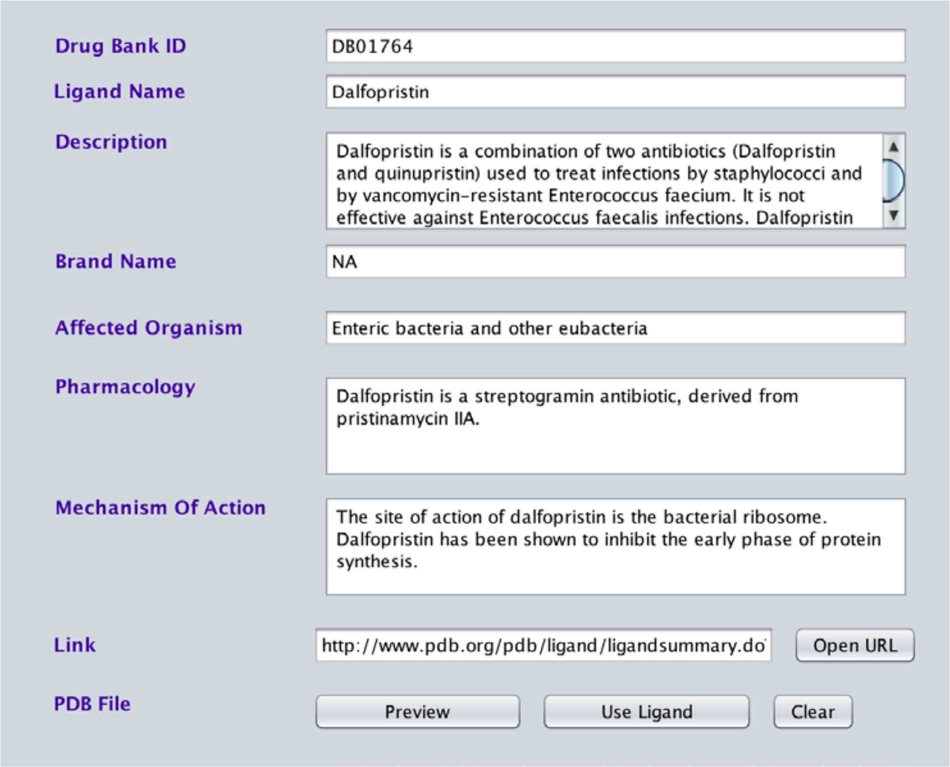


Server: localhost:8889 » Database: RCSBPDBLigandDatabase » Table: Ligands

	#	Name	Type	Collation	Attributes	Null	Default	Extra	Action	
□	1	LigandID	int(11)			No	None	AUTO_INCREMENT	Change	Dro
□	2	LigandName	text	latin1_swedish_ci		No	None		Change	Dro
□	3	DrugBankID	varchar(255)	latin1_swedish_ci		No	None		Change	Dro
□	4	Description	text	latin1_swedish_ci		No	None		Change	Dro
□	5	BrandName	text	latin1_swedish_ci		No	None		Change	Dro
□	6	AffectedOrganism	text	latin1_swedish_ci		No	None		Change	Dro
□	7	Pharmacology	text	latin1_swedish_ci		No	None		Change	Dro
□	8	MechanismOfAction	text	latin1_swedish_ci		No	None		Change	Dro
□	9	Link	text	latin1_swedish_ci		No	None		Change	Dro
□	10	PDBFile	longblob			No	None		Change	Dro

Figure 7.5.3. Ligand table structure in phpMyAdmin [105] interface

In the Ligand table the PDB file structures also stored as BLOB [106] file format. See Appendix A to view tool screens. Following Figure 7.5.4. shows data for a record from that table.



Drug Bank ID: DB01764

Ligand Name: Dalfopristin

Description: Dalfopristin is a combination of two antibiotics (Dalfopristin and quinupristin) used to treat infections by staphylococci and by vancomycin-resistant Enterococcus faecium. It is not effective against Enterococcus faecalis infections. Dalfopristin

Brand Name: NA

Affected Organism: Enteric bacteria and other eubacteria

Pharmacology: Dalfopristin is a streptogramin antibiotic, derived from pristinamycin IIA.

Mechanism Of Action: The site of action of dalfopristin is the bacterial ribosome. Dalfopristin has been shown to inhibit the early phase of protein synthesis.

Link: http://www.pdb.org/pdb/ligand/ligandsummary.do

PDB File

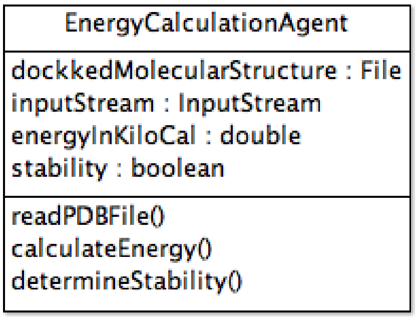
Figure 7.5.4. Data for ligand Dalfopristin [108]

7.5. Energy Calculation

It is a main requirement of this kind of molecular docking tool to predict the actual existence of the docked protein and ligand pair. To do that I have used, PDBTool [109]. It has implemented in python that is most popular scientific programming language in the world [110].

I have used following two components of PDBTool for the energy calculations:

- Determine coulomb energy using *pdb_coulomb.py* [111]
- Compute the dipole moment of the protein using *pdb_moment.py* [112]



EnergyCalculationAgent

dockkedMolecularStructure : File
inputStream : InputStream
energyInKiloCal : double
stability : boolean

readPDBFile()
calculateEnergy()
determineStability()

Figure 7.5: Class diagram - Energy Calculation Agent

7.6. Summary

In this chapter we have summarized the implementation of NanoAgents molecular docking. The next Chapter discusses about how this developed prototype has been used to prove that multi-agent approach is successful in simulating manual molecular docking process, by means of experimental results.

Chapter 8

Evaluation of MAS based Solution

8.1. Introduction

Previous chapter presented the implementation of the molecular docking system. And also we discussed the Voronoi diagrams to find Alpha Shape of the matching ligand and half plane intersection algorithm for the finding of Voronoi Diagrams of various computationally intensive steps of docking and binding site mapping. During conducting any research project, it is of high importance in terms of evaluating the validity of the proposed solution via some form of experimental basis. This Chapter presents the assessment of the suggested multi-agent based automated molecular docking discussed throughout this thesis. Evaluation is supported by performing a sample set of molecular docking simulations and matching the results with the real world manual molecular docking experiments.

8.2. Evaluation Methods

For the evaluation of the tool, compared known ligand-protein pair docking time taken in the usual manner with this software tool. Also analyzed the time difference for measuring the efficiency of the tool. The accuracy of the tool had been measured by the energy calculations. It will help to find the actual existence of the ligand-protein pairs in the environment found by the docking tool.

8.3. Known Receptor and Ligand Pairs

Following protein and ligand pairs had been used for the evaluation purpose of the software, to find out the time taken for the docking process. Some researches have contacted by the Department of Chemistry at University of Colombo, Sri Lanka [113] for the manual molecular docking. To compare the results, used that time taken for the manual molecular docking for each protein and ligand pairs. The next step was tracking the time taken for each docking process for the same protein-ligand pairs taken by the developed tool. To evaluate the accuracy of the software tool, also noted that the prediction of the actual existence of each docked pair using the tool’s energy calculation module. Following Table 8.2.1 shows the used ligand-protein pairs.

Protein	Ligand
1A52	OHT
1A52	EST
2AM9	B5R
1A52	GEN
1EWV	KAI
4PVU	BRL
4P6W	DEX
1EJF	R18
1DKF	REA
1A52	CME
1BY4	9RA
1A52	RAL
1HJ1	JJ3
1DSZ	TTB
1GDC	SNL
3KG2	BWD
1GDC	1CA
1BY4	REA

Table 8.2.1: Known Receptor and Ligand Pairs

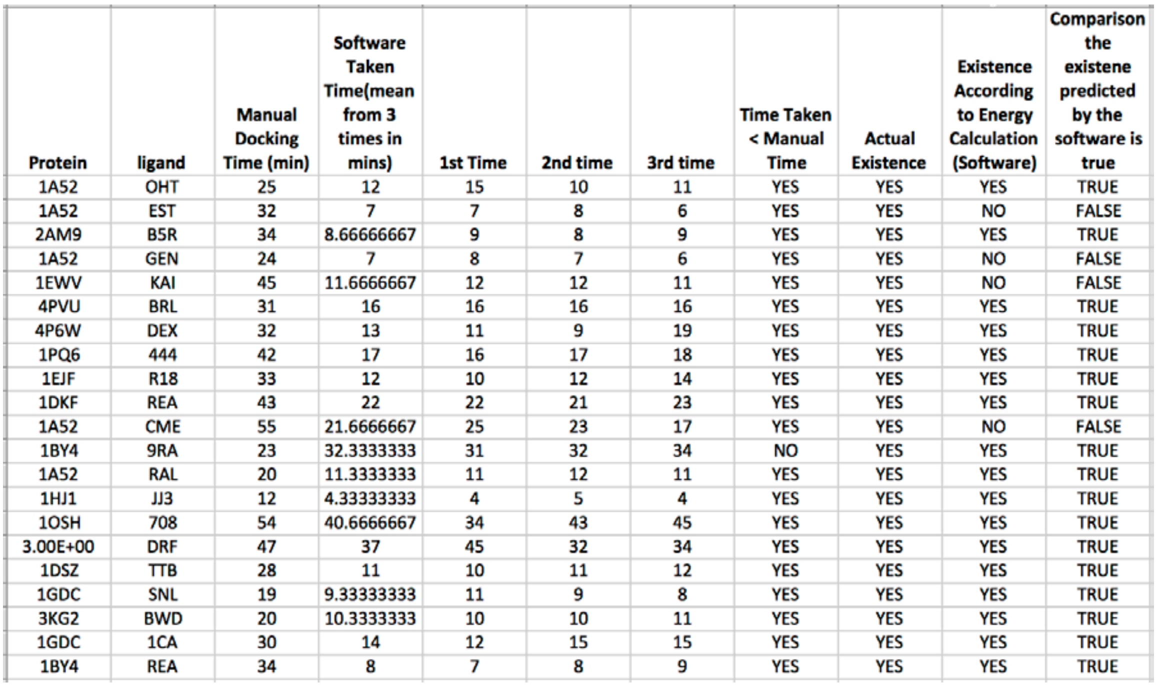
8.3. Test Results

Following are the two criteria used to evaluate the tool as explained the above paragraph:

Criteria 01:
The time taken for the molecular docking tool and the manual molecular docking
Criteria 02:
The prediction of the actual existence of the docked molecule

Following is the summary of the test:
Number of trials = 20
Number of trials that the time taken by the software is lesser than the actual manual molecular docking = 19
Number of trials that the actual existence predict the software is true = 16

Following Table 8.3.1 shows the actual results for each test:



Protein	ligand	Manual Docking Time (min)	Software Taken Time(mean from 3 times in mins)	1st Time	2nd time	3rd time	Time Taken < Manual Time	Actual Existence	Existence According to Energy Calculation (Software)	Comparison the existene predicted by the software is true
1A52	OHT	25	12	15	10	11	YES	YES	YES	TRUE
1A52	EST	32	7	7	8	6	YES	YES	NO	FALSE
2AM9	B5R	34	8.66666667	9	8	9	YES	YES	YES	TRUE
1A52	GEN	24	7	8	7	6	YES	YES	NO	FALSE
1EWV	KAI	45	11.6666667	12	12	11	YES	YES	NO	FALSE
4PVU	BRL	31	16	16	16	16	YES	YES	YES	TRUE
4P6W	DEX	32	13	11	9	19	YES	YES	YES	TRUE
1PQ6	444	42	17	16	17	18	YES	YES	YES	TRUE
1EJF	R18	33	12	10	12	14	YES	YES	YES	TRUE
1DKF	REA	43	22	22	21	23	YES	YES	YES	TRUE
1A52	CME	55	21.6666667	25	23	17	YES	YES	NO	FALSE
1BY4	9RA	23	32.3333333	31	32	34	NO	YES	YES	TRUE
1A52	RAL	20	11.3333333	11	12	11	YES	YES	YES	TRUE
1HJ1	JJ3	12	4.33333333	4	5	4	YES	YES	YES	TRUE
1OSH	708	54	40.6666667	34	43	45	YES	YES	YES	TRUE
3.00E+00	DRF	47	37	45	32	34	YES	YES	YES	TRUE
1DSZ	TTB	28	11	10	11	12	YES	YES	YES	TRUE
1GDC	SNL	19	9.33333333	11	9	8	YES	YES	YES	TRUE
3KG2	BWD	20	10.3333333	10	10	11	YES	YES	YES	TRUE
1GDC	1CA	30	14	12	15	15	YES	YES	YES	TRUE
1BY4	REA	34	8	7	8	9	YES	YES	YES	TRUE

Table 8.3.1: Known Protein-Ligand pairs and the results from the tool in Excel Sheet

8.4. Precision

Accuracy is a method that used as a statistical measure of how satisfactory a binary classification test accurately recognizes or rejects a condition [114].

$$\text{accuracy} = \frac{\text{number of true positives} + \text{number of true negatives}}{\text{number of true positives} + \text{false positives} + \text{false negatives} + \text{true negatives}}$$

Equation 8.4.1 Precision Equation 8.4 [115]

8.5. Precision for Criteria 01:

Considered the time taken for the molecular docking tool and the manual molecular docking for the evaluation of the tool.

Probability of success (time taken by the software is lesser than the actual manual molecular docking) = 19/20 = 0.95

Probability of failure (time taken by the software is higher than the actual manual molecular docking) = 1/20 = 0.05

Accuracy of the software = 9/20 = 0.95 = 95%

8.6. Precision for Criteria 02:

Considered the prediction accuracy of the actual existence of the docked molecules by the software tool by using energy calculation.

Probability of success (software predicts the actual existence of the docked molecule correctly) = 16/20 = 0.8

Probability of failure (software not predicts the actual existence of the docked molecule correctly) = 4/20 = 0.2

Accuracy of the software = 16/20 = 0.8 = 80%

8.4. Summary

In this chapter it is presented that the evaluation section of the Multi-Agent System technology based Automated Molecular Docking Tool. The next Chapter termed as Conclusion and Further Work winds up on what is finally to be discussed in this thesis after stating its hypothesis, designing an automated molecular docking tool, implementing a prototype and ultimately proving the hypothesis by means of experimental results obtained from evaluation.

Chapter 9

Conclusion and Further Work

9.1. Introduction

Based on the evaluation strategy and the recorded experimental results from previous Chapter; this Chapter discusses about the Conclusion of this thesis by interpreting the results derived from evaluation. The discussion continues further by describing the achievement of project objectives, solutions provided for the problems encountered along with possible further enhancements worth doing as a continuation of the project.

8.7. Conclusion

I have executed the molecular simulation 20 times for the known ligand-protein pairs and compared it to the actual time taken for the manual molecular docking. The experiment results derived based on two scenarios as follows:

- Time for the docking: By comparing the time taken for the manual docking process with the time taken for the automated molecular docking, it says the software tool can obtain 95% accuracy
- Accuracy: It has 80% accuracy by considering the prediction of the correct actual existence of the docked molecules according to the energy calculations

Unlike the established approaches following to develop molecular docking such as manual methods, the agent-based approach doesn't require developing algorithms to simulate the different steps of the molecular docking.

Further the information embedded in these rules for a given individual agent needs not to be completed. Each individual agent in the molecular docking simulation will always discover something which was earlier unidentified because of the knowledge sharing, with other individuals via message passing. The interesting thing is this tool can be used to discover new drugs for the selected protein targets. Some drugs can be effective for the selected proteins unexpected.

This tool could successfully exhibit multi-agent system characteristics such as autonomy, deliberation, communication and negotiation. Protein, Ligand, Shape

Determination and Energy Calculation agents are sharing their knowledge to come up with new docked molecule. NanoAgent system could exhibit self-organization as well as other control models and associated composite activities. The results suggest that the Multi-Agent Systems technology could be successfully applied to automate the manual molecular docking process, which is an inherently complex process.

8.8. Achievement of the Objectives

With respect to the aim and objectives listed for this project (under Section 1.2); core concepts behind the development of automated molecular docking tool and significant observations from real world scenarios related to manual molecular docking are discussed inside the literature review of this thesis (provided in Chapter 2).

The solution to the multi-agent system based automated molecular docking tool is described in detail under the design chapter of this thesis (refer to Chapter 5). Further implementing related details of the docking tool are discussed under the implementation chapter of this thesis (refer to Chapter 7). Results obtained from the simulation sessions of the molecular docking are validated against observations recorded from real world manual docking scenarios within the evaluation chapter (i.e. Chapter 8) of this thesis. Finally, Chapter 9 (i.e. the main Conclusion) managed to prove the hypothesis of this thesis (stated under section 5.2) while describing on how to use the proposed design of the multi-agent molecular docking system.

The results expressed that the Multi-Agent Systems technology can be successfully used to automate the manual molecular docking. So the aim of this research work could be successfully achieved with their mentioned objectives as follows:

1. Able to critically study the molecular docking domain with a view to identify current practices and the issues in molecular docking
2. Critically analyzed and did a comprehensive evaluation of the existing software solutions in molecular docking with a view to define the research problem and possible technology
3. Did in depth study about Multi-Agent Technology and its applications
4. Designed and implemented Multi-Agent System for Molecular Docking.

5. Evaluated the accuracy of the Multi-Agent System based molecular docking tool by entering known existing ligands and protein pairs

The results suggest that the Multi-Agent Systems technology could be successfully applied to automate the manual molecular docking. With this tool, it is better to have animated transition of the molecular docking process that associate with each step of the entire process. So this tool is lacking the presentation of the docking process step by step.

9.2. Further Work to Improve Molecular Docking Tool

The main focus throughout this thesis was about simulating automated molecular docking process by developing a computer simulation tool using multi-agent systems technology. Artificial Neural Network is amazing technology, which can be applied to improve this software tool. If we can implement this tool with a trained artificial neural network for the results of pre docking processes of known protein-ligand pairs, it will be helpful to improve the performance and the accuracy.

Even if it discovers the ligand-protein pairs it is essential to perform a test for identification of possible poisonous of drugs using experiments. But it is a not economically effective and not efficient procedure that needs actual living beigns testing. First, we can use this software tool and have to verify it by doing proper clinical trials.

The most significant types of docking systems are protein-ligand and protein-protein. In this research I have addressed only protein-ligand docking. But it can be improved for the other two categories too. However, the reduced efficacy of the existing scoring functions I have used for the docking tool, is the biggest barrier, which obstructs the improvement of the molecular docking method.

9.3. Summary

In this chapter presented the conclusion and the future work associated with this research work. There is clearly much work to be done in the area of automated molecular docking using multi-agent system technology to improve the quality and the accuracy of the software tool that I have developed for this research work.

This thesis addressed the application and benefits behind proposing a multi-agent based solution for automated molecular docking process. Further, it has been proven that the multi-agent based prototype is able to provide identical results during a simulation of ligand-protein binding process simulation, comparing with statistics recorded on real world manual molecular docking scenarios. The major advantage derived here over the conventional approaches is that the multi-agent based approach doesn't require waiting until complete one step in the docking process. Instead, through communication (i.e. passing messages) and unplanned knowledge sharing between the agents, the multi-agent based approach is able to emerge uncertain drug discoveries for the selected decease proteins.

References

[1] M. Hellmann, “Fuzzy logic introduction,” *Univ. Rennes*, vol. 1, 2001.

[2] C. Gershenson, “Artificial neural networks for beginners,” *ArXiv Prepr. Cs0308031*, 2003.

[3] M. Melanie, “An introduction to genetic algorithms,” *Camb. Mass. Lond. Engl. Fifth Print.*, vol. 3, 1999.

[4] D. Merritt, *Building expert systems in Prolog*. Springer-Verlag New York, USA, 1989.

[5] R. Collobert, J. Weston, L. Bottou, M. Karlen, K. Kavukcuoglu, and P. Kuksa, “Natural language processing (almost) from scratch,” *J. Mach. Learn. Res.*, vol. 12, pp. 2493–2537, 2011.

[6] T. Moyaux, B. Chaib-Draa, and S. D’Amours, “Supply chain management and multiagent systems: An overview,” in *Multiagent based supply chain management*, Springer, 2006, pp. 1–27.

[7] “Shehory O.,Sycara K.,Sukthankar G. and Mukherjee V.‘Agent aided aircraft maintenance’.1999.Accessed on15/09/2014.” .

[8] N. R. Jennings, “Commitments and conventions: The foundation of coordination in multi-agent systems,” *Knowl. Eng. Rev.*, vol. 8, no. 03, pp. 223–250, 1993.

[9] G. M. Morris and M. Lim-Wilby, “Molecular docking,” in *Molecular Modeling of Proteins*, Springer, 2008, pp. 365–382.

[10] N. A. Chowdhury and D. Dou, “Improving the accuracy of ontology alignment through ensemble fuzzy clustering,” in *On the Move to Meaningful Internet Systems: OTM 2011*, Springer, 2011, pp. 826–833.

[11] “UofT PSMDB.” [Online]. Available: http://compbio.cs.toronto.edu/psmdb/. [Accessed: 08-Sep-2014].

[12] “RCSB Protein Data Bank - RCSB PDB.” [Online]. Available: http://www.rcsb.org/pdb/static.do?p=general_information/web_links/databases.html. [Accessed: 08-Sep-2014].

[13] “SMPDB.” [Online]. Available: http://www.smpdb.ca/. [Accessed: 08-Sep-2014].

[14] C. de Graaf, P. Pospisil, W. Pos, G. Folkers, and N. P. E. Vermeulen, “Binding Mode Prediction of Cytochrome P450 and Thymidine Kinase Protein−Ligand Complexes by Consideration of Water and Rescoring in Automated Docking,” *J. Med. Chem.*, vol. 48, no. 7, pp. 2308–2318, Apr. 2005.

[15] C. A. Nicolaou and N. Brown, “Multi-objective optimization methods in drug design,” *Drug Discov. Today Technol.*, vol. 10, no. 3, pp. e427–e435, Sep. 2013.

[16] “Molecular Theory and Computation Group | Stanford University.” [Online]. Available: http://web.stanford.edu/group/mtc/members.html. [Accessed: 15-Sep-2014].

[17] “Home of David S. Goodsell.” [Online]. Available: http://mgl.scripps.edu/people/goodsell/. [Accessed: 15-Sep-2014].

[18] “AutoDock — AutoDock.” [Online]. Available: http://autodock.scripps.edu/. [Accessed: 15-Jul-2014].

[19] A. Rudnitskaya, B. Török, and M. Török, “Molecular docking of enzyme inhibitors: A COMPUTATIONAL TOOL FOR STRUCTURE-BASED DRUG DESIGN,” *Biochem. Mol. Biol. Educ.*, vol. 38, no. 4, pp. 261–265, Aug. 2010.

[20] “MGLTools Website - Welcome — MGLTools.” [Online]. Available: http://mgltools.scripps.edu/. [Accessed: 15-Sep-2014].

[21] M. Olšák, J. Filipovič, and M. Prokop, “FastGrid–The Accelerated AutoGrid Potential Maps Generation for Molecular Docking,” *Comput. Inform.*, vol. 29, no. 6+, pp. 1325–1336, 2012.

[22] A. A. Yee, A. Savchenko, A. Ignachenko, J. Lukin, X. Xu, T. Skarina, E. Evdokimova, C. S. Liu, A. Semesi, V. Guido, and others, “NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins,” *J. Am. Chem. Soc.*, vol. 127, no. 47, pp. 16512–16517, 2005.

[23] I. Wandzik, “Current molecular docking tools and comparisons thereof,” *MATCH*, vol. 55, no. 2, pp. 271–278, 2006.

[24] S.-Y. Huang and X. Zou, “Advances and Challenges in Protein-Ligand Docking,” *Int. J. Mol. Sci.*, vol. 11, no. 8, pp. 3016–3034, Aug. 2010.

[25] H. Claußen, C. Buning, M. Rarey, and T. Lengauer, “FlexE: efficient molecular docking considering protein structure variations,” *J. Mol. Biol.*, vol. 308, no. 2, pp. 377–395, Apr. 2001.

[26] “The PubChem Project.” [Online]. Available: http://pubchem.ncbi.nlm.nih.gov/. [Accessed: 15-Sep-2014].

[27] “pubchem compound download.” [Online]. Available: ftp://ftp.ncbi.nlm.nih.gov/pubchem/Compound/CURRENT-Full/SDF/. [Accessed: 15-Sep-2014].

[28] “Welcome to ZINC Is Not Commercial - A database of commercially-available compounds.” [Online]. Available: http://zinc.docking.org/. [Accessed: 15-Sep-2014].

[29] “Chemical Structure Drawing Search - eMolecules.” [Online]. Available: http://emolecules.com/. [Accessed: 15-Sep-2014].

[30] “Chemical Entities of Biological Interest (ChEBI).” [Online]. Available: http://www.ebi.ac.uk/chebi/. [Accessed: 15-Sep-2014].

[31] “chebi ontology download.” [Online]. Available: ftp://ftp.ebi.ac.uk/pub/databases/chebi/ontology/. [Accessed: 15-Sep-2014].

[32] “License BioCyc Data Files.” [Online]. Available: http://biocyc.org/download-flatfiles.shtml. [Accessed: 15-Sep-2014].

[33] “Download BioCyc Databases and Pathway Tools Software.” [Online]. Available: http://biocyc.org/download.shtml. [Accessed: 15-Sep-2014].

[34] “DrugBank.” [Online]. Available: http://www.drugbank.ca/. [Accessed: 15-Sep-2014].

[35] S. Raychaudhuri, “Introduction to monte carlo simulation,” in *Simulation Conference, 2008. WSC 2008. Winter*, 2008, pp. 91–100.

[36] E. Yang, A. Erdogan, T. Arslan, and N. Barton, “An Improved Particle Swarm Optimization Algorithm for Power-Efficient Wireless Sensor Networks,” 2007, pp. 76–82.

[37] S.-Y. Huang, S. Z. Grinter, and X. Zou, “Scoring functions and their evaluation methods for protein–ligand docking: recent advances and future directions,” *Phys. Chem. Chem. Phys.*, vol. 12, no. 40, p. 12899, 2010.

[38] S. K. Lai-Yuen and Y.-S. Lee, “Interactive computer-aided design for molecular docking and assembly,” *Comput.-Aided Des. Appl.*, vol. 3, no. 6, pp. 701–709, 2006.

[39] “FDS: flexible ligand and receptor docking with a continuum solvent model and soft-core energy function.” [Online]. Available: http://www.pubfacts.com/detail/12926007/FDS:-flexible-ligand-and-receptor-docking-with-a-continuum-solvent-model-and-soft-core-energy-functi. [Accessed: 15-Jul-2014].

[40] “CiteSeerX — A Randomized Kinematics-Based Approach to Pharmacophore-Constrained Conformational Search and Database Screening.” [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.19.43. [Accessed: 15-Jul-2014].

[41] “Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function - Morris - 1999 - Journal of Computational Chemistry - Wiley Online Library.” [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/(SICI)1096-987X(19981115)19:14%3C1639::AID-JCC10%3E3.0.CO;2-B/abstract. [Accessed: 15-Jul-2014].

[42] “IEEE Xplore Abstract - Ligand binding with OBPRM and user input.” [Online]. Available: http://ieeexplore.ieee.org/xpl/login.jsp?tp=&arnumber=932673&url=http%3A%2F%2Fieeexplore.ieee.org%2Fxpls%2Fabs_all.jsp%3Farnumber%3D932673. [Accessed: 15-Jul-2014].

[43] W. Duch, K. Swaminathan, and J. Meller, “Artificial intelligence approaches for rational drug design and discovery,” *Curr. Pharm. Des.*, vol. 13, no. 14, pp. 1497–1508, 2007.

[44] “Adaptation in Natural and Artificial Systems | The MIT Press.” [Online]. Available: http://mitpress.mit.edu/books/adaptation-natural-and-artificial-systems. [Accessed: 15-Jul-2014].

[45] “Protein Ligand Rigid Docking.” [Online]. Available: http://www.scfbio-iitd.res.in/dock/pardock.jsp. [Accessed: 02-Sep-2014].

[46] “ParDOCK - Data Set.” [Online]. Available: http://www.scfbio-iitd.res.in/dock/dockdata.jsp. [Accessed: 02-Sep-2014].

[47] “Docking Methodology Flowchart.” [Online]. Available: http://www.scfbio-iitd.res.in/dock/flowchart.html. [Accessed: 02-Sep-2014].

[48] G. Sakarkar and S. Upadhye, “A survey of software agent and ontology,” *Int. J. Comput. Appl.*, vol. 1, no. 7, pp. 18–23, 2010.

[49] N. Jennings and M. J. Wooldridge, *Agent technology: foundations, applications, and markets*. Springer Science & Business Media, 1998.

[50] B. Chandrasekaran, J. R. Josephson, and V. R. Benjamins, “What are ontologies, and why do we need them?,” *IEEE Intell. Syst.*, vol. 14, no. 1, pp. 20–26, 1999.

[51] A. H. André Filipe de Moraes Batista, Emerson Aguiar Noronha, Fábio Aragão da Silva, Maria das Graças Bruno Marietto, Robson dos Santos França Terry Lima Ruas, *Modeling Artificial Life Through Multi-Agent Based Simulation.* INTECH Open Access Publisher, 2011.

[52] M. University of Southampton Luck and AgentLink, *Agent technology: computing as interaction ; a roadmap for agent based computing*. Southampton: University of Southampton on behalf of AgentLink III, 2005.

[53] “‘Agent Technology : Computing as Interaction,A Roadmap for Agent Based Computing’.[Online]. Available: http://www.das.ufsc.br/~jomi/mas/leituras/agentlink-roadmap.pdf [Accessed: 16-Sep-2014].” .

[54] D. J. Cook, “Multi-agent smart environments,” *J. Ambient Intell. Smart Environ.*, vol. 1, no. 1, pp. 51–55, 2009.

[55] Z. Li, C. H. Sim, and M. Y. H. Low, “A survey of emergent behavior and its impacts in agent-based systems,” in *Industrial Informatics, 2006 IEEE International Conference on*, 2006, pp. 1295–1300.

[56] “RDF Current Status - W3C.” [Online]. Available: http://www.w3.org/standards/techs/rdf#w3c_all. [Accessed: 29-Jul-2014].

[57] P. Maret, J. Calmet, and others, “Agent-based knowledge communities,” *Int. J. Comput. Sci. Appl.*, vol. 6, no. 2, pp. 1–18, 2009.

[58] “protégé.” [Online]. Available: http://protege.stanford.edu/products.php. [Accessed: 25-Jul-2014].

[59] “‘Ontologies for Agents’ by Michael N. Huhns and Munindar P. Singh.” [Online]. Available: http://scholarcommons.sc.edu/csce_facpub/67/. [Accessed: 02-Jun-2015].

[60] Y. Ding and D. Fensel, “Ontology Library Systems: The key to successful Ontology Reuse.,” in *SWWS*, 2001, pp. 93–112.

[61] “Wooldridge M. and Jennings N. ‘Intelligent Agents: Theory and Practice’.1995.” .

[62] G. Weiss, *Multiagent systems a modern approach to distributed artificial intelligence*. Cambridge, Mass.: MIT Press, 1999.

[63] M. N. Huhns and M. P. Singh, “Ontologies for agents,” *Internet Comput. IEEE*, vol. 1, no. 6, pp. 81–83, 1997.

[64] L. Antunes and H. Coelho, “On how to conduct experimental research with self-motivated agents,” in *Regulated Agent-Based Social Systems*, Springer, 2004, pp. 31–47.

[65] E. J. Gonzalez, A. F. Hamilton, L. Moreno, R. L. Marichal, and J. Toledo, “Ontologies in a multi-agent system for automated scheduling,” *Comput. Inform.*, vol. 23, no. 2, pp. 157–177, 2012.

[66] K. Degtyarenko, “Chemical vocabularies and ontologies for bioinformatics,” in *Proceedings of the 2003 International Chemical Information Conference*, 2003.

[67] “The Royal Society of Chemistry.” [Online]. Available: http://www.rsc.org/. [Accessed: 02-Jun-2015].

[68] “Name Reaction Ontology (RXNO).” [Online]. Available: http://www.rsc.org/ontologies/RXNO/. [Accessed: 02-Jun-2015].

[69] “Molecular Processes Ontology (MOP).” [Online]. Available: http://www.rsc.org/ontologies/MOP/. [Accessed: 02-Jun-2015].

[70] “Chemical Methods Ontology (CMO).” [Online]. Available: http://www.rsc.org/ontologies/CMO/. [Accessed: 02-Jun-2015].

[71] “OBI Ontology.” [Online]. Available: http://obi-ontology.org/page/Main_Page. [Accessed: 02-Jun-2015].

[72] “The ChEBI reference database and ontology for biologically relevant chemistry: enhancements for 2013.” [Online]. Available: http://nar.oxfordjournals.org/content/early/2012/11/23/nar.gks1146.full. [Accessed: 02-Jun-2015].

[73] “FGO: A novel ontology for identification of ligand functional group.” [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2248449/. [Accessed: 02-Jun-2015].

[74] “Khaled H. Baraka, Molecular Docking, PharmaMatrix workshop in Computational Biophysic , 2009.” .

[75] “Machado_et_al_LNCS_4643_1-11_2007.pdf.” .

[76] “Carboxypeptidase.” [Online]. Available: http://www.chemistry.wustl.edu/~edudev/LabTutorials/Carboxypeptidase/carboxypeptidase.html. [Accessed: 17-Sep-2014].

[77] T. H. Keller, A. Pichota, and Z. Yin, “A practical view of ‘druggability,’” *Curr. Opin. Chem. Biol.*, vol. 10, no. 4, pp. 357–361, Aug. 2006.

[78] “Potential Energy Function.” [Online]. Available: http://www.ch.embnet.org/MD_tutorial/pages/MD.Part2.html. [Accessed: 17-Sep-2014].

[79] “Electric energy and potential.” [Online]. Available: http://physics.bu.edu/~duffy/PY106/Potential.html. [Accessed: 25-Oct-2014].

[80] “X-ray Powder Diffraction (XRD).” [Online]. Available: http://serc.carleton.edu/research_education/geochemsheets/techniques/XRD.html. [Accessed: 02-Jun-2015].

[81] “NMR studies of protein structure and dynamics. - PubMed - NCBI.” [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/15780912. [Accessed: 02-Jun-2015].

[82] “MSOE Center for BioMolecular Modeling - Jmol Training Guide.” [Online]. Available: http://cbm.msoe.edu/markMyweb/jmolWebsite/glossary.html. [Accessed: 02-Jun-2015].

[83] “Dumontier Lab - Laboratory for Biomedical Knowledge Discovery.” [Online]. Available: http://dumontierlab.com/?page=ontologies. [Accessed: 29-May-2015].

[84] “Jade Site | Java Agent DEvelopment Framework.” [Online]. Available: http://jade.tilab.com/. [Accessed: 17-Sep-2014].

[85] “FIPA ACL Message Structure Specification.” [Online]. Available: http://www.fipa.org/specs/fipa00061/SC00061G.html. [Accessed: 18-Jul-2014].

[86] “Apache Jena - Home.” [Online]. Available: https://jena.apache.org/. [Accessed: 09-Sep-2014].

[87] Y. Yan, C. Wang, A. Zhou, W. Qian, L. Ma, and Y. Pan, “Efficiently querying rdf data in triple stores,” in *Proceedings of the 17th international conference on World Wide Web*, 2008, pp. 1053–1054.

[88] J. C. Arpírez, A. Gómez-Pérez, A. Lozano-Tello, and H. S. A. N. Pinto, “Reference ontology and (ONTO) 2 agent: the ontology yellow pages,” *Knowl. Inf. Syst.*, vol. 2, no. 4, pp. 387–412, 2000.

[89] “SPARQL Query Language for RDF.” [Online]. Available: http://www.w3.org/TR/rdf-sparql-query/. [Accessed: 25-Jul-2014].

[90] G. Manjunath, C. Sayers, D. Reynolds, V. Ks, S. K. Mohalik, R. Badrinath, J. L. Recker, and M. Mesarina, “Semantic views for controlled access to the semantic web,” Tech. Rep. HPL-08, 2008.

[91] S. S. Ara, Z. U. Shamszaman, I. Chong, S. S. Ara, Z. U. Shamszaman, and I. Chong, “Web-of-Objects Based User-Centric Semantic Service Composition Methodology in the Internet of Things, Web-of-Objects Based User-Centric Semantic Service Composition Methodology in the Internet of Things,” *Int. J. Distrib. Sens. Netw. Int. J. Distrib. Sens. Netw.*, vol. 2014, 2014, p. e482873, May 2014.

[92] “SciPy.org — SciPy.org.” [Online]. Available: http://www.scipy.org/. [Accessed: 02-Jun-2015].

[93] “OWL Web Ontology Language Current Status - W3C.” [Online]. Available: http://www.w3.org/standards/techs/owl#w3c_all. [Accessed: 02-Jun-2015].

[94] “Jmol: an open-source Java viewer for chemical structures in 3D.” [Online]. Available: http://jmol.sourceforge.net/. [Accessed: 17-Sep-2014].

[95] “MySQL :: The world’s most popular open source database.” [Online]. Available: http://www.mysql.com/. [Accessed: 04-Feb-2015].

[96] “Welcome to Python.org.” [Online]. Available: https://www.python.org/. [Accessed: 02-Jun-2015].

[97] “Improve your drug discovery process with Accelrys Discovery Studio.” [Online]. Available: http://accelrys.com/products/discovery-studio/. [Accessed: 02-Jun-2015].

[98] “PyMOL | www.pymol.org.” [Online]. Available: https://www.pymol.org/. [Accessed: 02-Jun-2015].

[99] D. Rudenko and A. Borisov, “An overview of blackboard architecture application for real tasks,” in *Scientific Proceedings Of Riga Technical University, Ser*, 2007, vol. 5, pp. 50–57.

[100] H. Edelsbrunner, “Alpha shapes—a survey,” *Tessellations Sci.*, vol. 27, 2010.

[101] T. J. Ewing and I. D. Kuntz, “Critical evaluation of search algorithms for automated molecular docking and database screening,” *J. Comput. Chem.*, vol. 18, no. 9, pp. 1175–1189, 1997.

[102] “Locating Binding Sites in Protein Structures.” [Online]. Available: http://www.chemcomp.com/journal/sitefind.htm. [Accessed: 08-Sep-2014].

[103] L. Jirkovskỳ, S. M. Maňák, and I. Kolingerová, “Finding Cavities in a Molecule.”

[104] “RCSB Protein Data Bank - RCSB PDB.” [Online]. Available: http://www.rcsb.org/pdb/home/home.do. [Accessed: 15-Jul-2014].

[105] “phpMyAdmin.” [Online]. Available: http://www.phpmyadmin.net/home_page/index.php. [Accessed: 04-Feb-2015].

[106] “To BLOB or Not To BLOB: Large Object Storage in a Database or a Filesystem - Microsoft Research.” [Online]. Available: http://research.microsoft.com/apps/pubs/?id=64525. [Accessed: 04-Feb-2015].

[107] “Index of approved drugs and nutraceuticals in the PDB.” [Online]. Available: http://www.ebi.ac.uk/thornton-srv/databases/cgi-bin/drugport/GetPage.pl?template=drugindex_pdb.html. [Accessed: 04-Feb-2015].

[108] “Quinupristin/dalfopristin: the first available macrolide-lincosamide-streptogramin antibiotic.” [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1312217/. [Accessed: 01-May-2015].

[109] C. Pu, K. P. Sheka, J. Ong, L. Chang, and A. Chang, “PDBtool: A prototype object oriented toolkit for protein structure verification,” DTIC Document, 1992.

[110] M. Nosrati, “Python: An appropriate language for real world programming,” 2011.

[111] “pdb_coulomb.py - pdb-tools - A set of tools for manipulating and doing calculations on wwPDB macromolecule structure files - Google Project Hosting.” [Online]. Available: https://code.google.com/p/pdb-tools/source/browse/trunk/pdbTools/pdb_coulomb.py. [Accessed: 01-May-2015].

[112] “pdb_moment.py - pdb-tools - A set of tools for manipulating and doing calculations on wwPDB macromolecule structure files - Google Project Hosting.” [Online]. Available: https://code.google.com/p/pdb-tools/source/browse/trunk/pdbTools/pdb_moment.py. [Accessed: 01-May-2015].

[113] “University of Colombo » Prof. Ranil Dassanayake.” [Online]. Available: http://www.cmb.ac.lk/?page_id=3734. [Accessed: 01-May-2015].

[114] J. D. Brown, “Sample size and statistical precision,” *JALT Test. Eval. SIG Newsl.*, vol. 11, pp. 21–24, 2007.

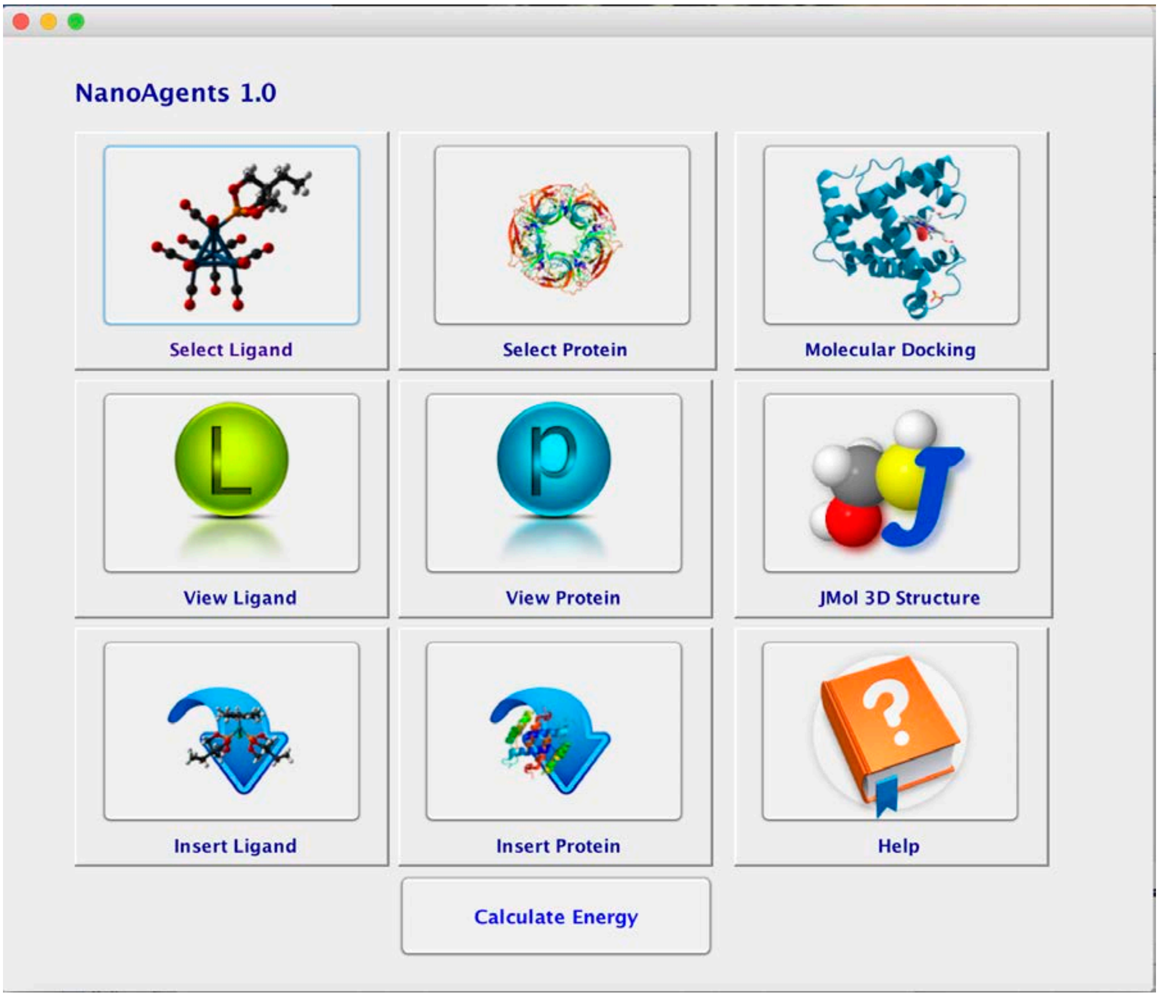
[115] “Basic Principals of ROC Analysis, Charles E. Metz,Seminars in Nuclera Medicine,Vol V111 ,No 4, October, 1978.” .

Appendix A

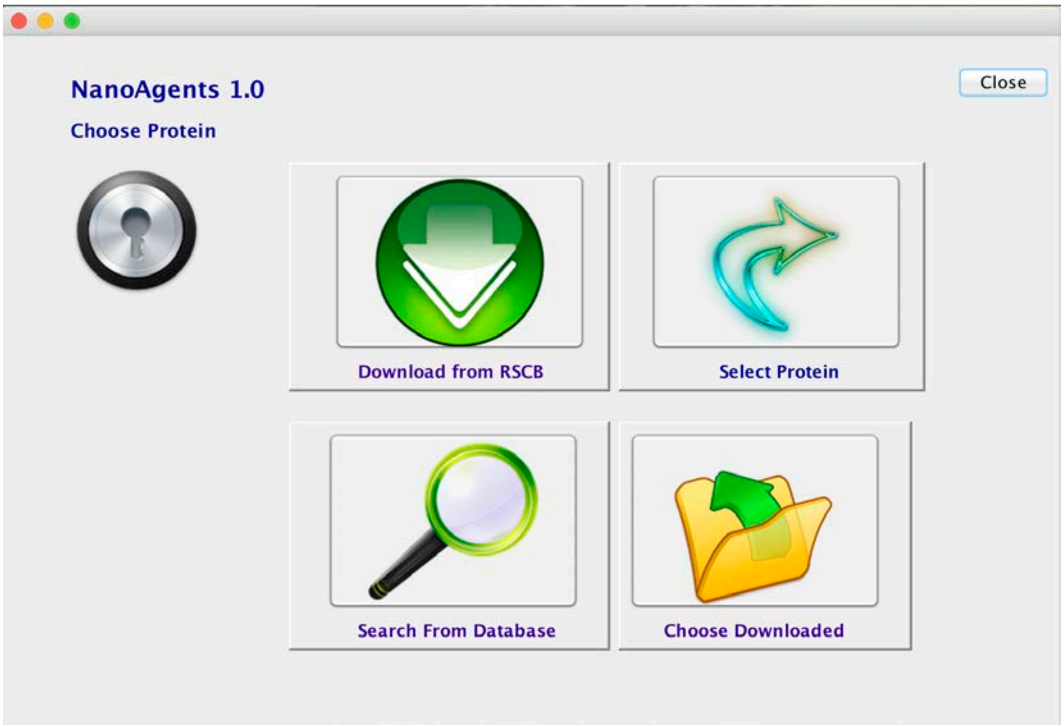
MAS Based Molecular Docking Tool

A.1 Screenshots

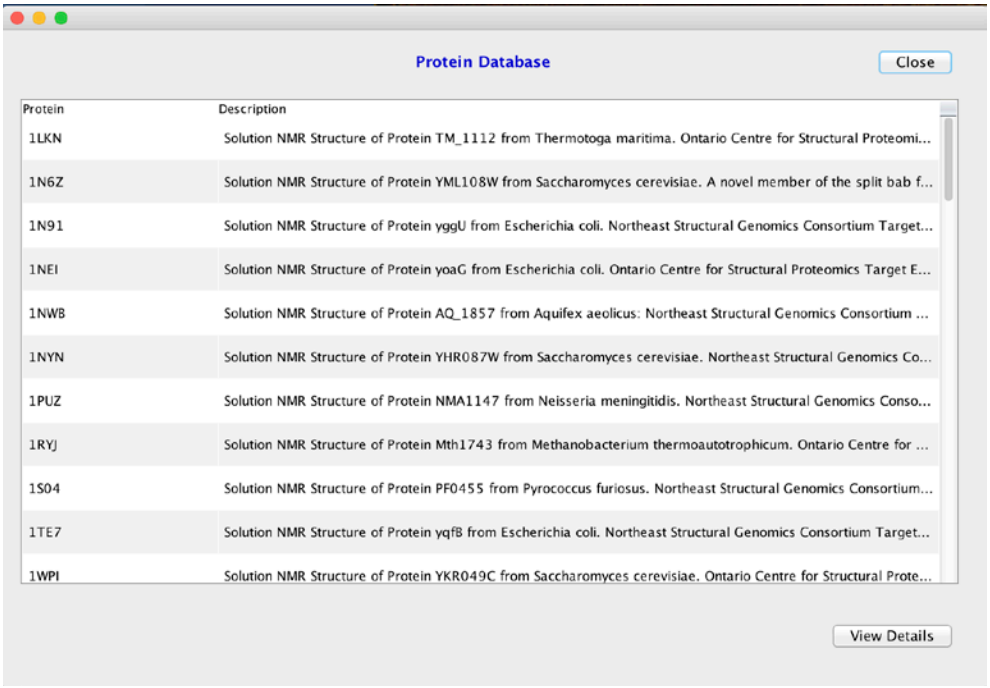
Following diagrams shows the main functions of the molecular docking tool.



Following diagram shows the dashboard for choosing the protein from the existing database, search protein if you know the name or download pdb files from the RCSB Protein Data Bank.



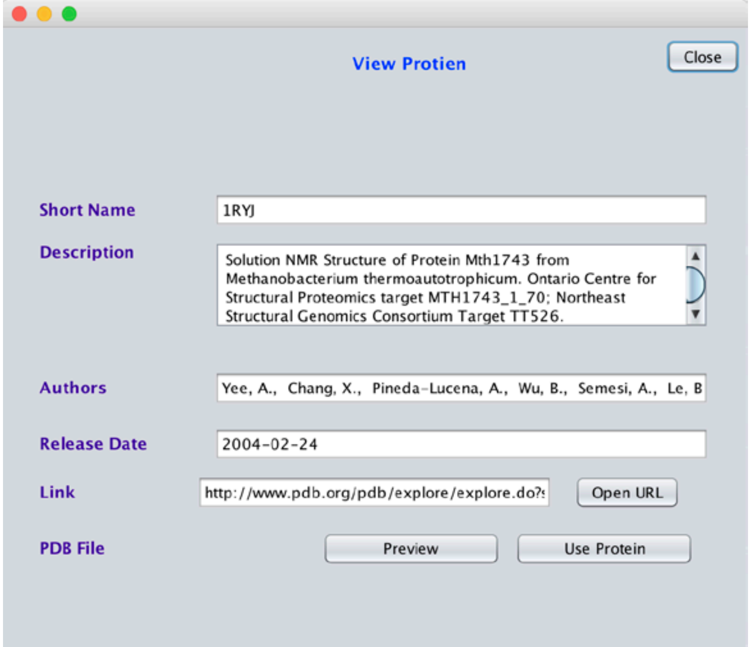
Following diagram shows the list of proteins retrieved from the local database.



Protein	Description
1LKN	Solution NMR Structure of Protein TM_1112 from Thermotoga maritima. Ontario Centre for Structural Proteomi...
1N6Z	Solution NMR Structure of Protein YML108W from Saccharomyces cerevisiae. A novel member of the split bab f...
1N91	Solution NMR Structure of Protein yggU from Escherichia coli. Northeast Structural Genomics Consortium Target...
1NEI	Solution NMR Structure of Protein yoaG from Escherichia coli. Ontario Centre for Structural Proteomics Target E...
1NWB	Solution NMR Structure of Protein AQ_1857 from Aquifex aeolicus: Northeast Structural Genomics Consortium ...
1NYN	Solution NMR Structure of Protein YHR087W from Saccharomyces cerevisiae. Northeast Structural Genomics Co...
1PUZ	Solution NMR Structure of Protein NMA1147 from Neisseria meningitidis. Northeast Structural Genomics Conso...
1RYJ	Solution NMR Structure of Protein Mth1743 from Methanobacterium thermoautotrophicum. Ontario Centre for ...
1S04	Solution NMR Structure of Protein PF0455 from Pyrococcus furiosus. Northeast Structural Genomics Consortium...
1TE7	Solution NMR Structure of Protein yqfB from Escherichia coli. Northeast Structural Genomics Consortium Target...
1WPI	Solution NMR Structure of Protein YKR049C from Saccharomyces cerevisiae. Ontario Centre for Structural Prote...

User can select one from the list and view the details and also can use it for the docking. It has the feature to preview the 3D structure before applying it for the docking.

User can see more details from the PDB online database by clicking on the “Open URL” button.



View Protien

Short Name: 1RYJ

Description: Solution NMR Structure of Protein Mth1743 from Methanobacterium thermoautotrophicum. Ontario Centre for Structural Proteomics target MTH1743_1_70; Northeast Structural Genomics Consortium Target TT526.

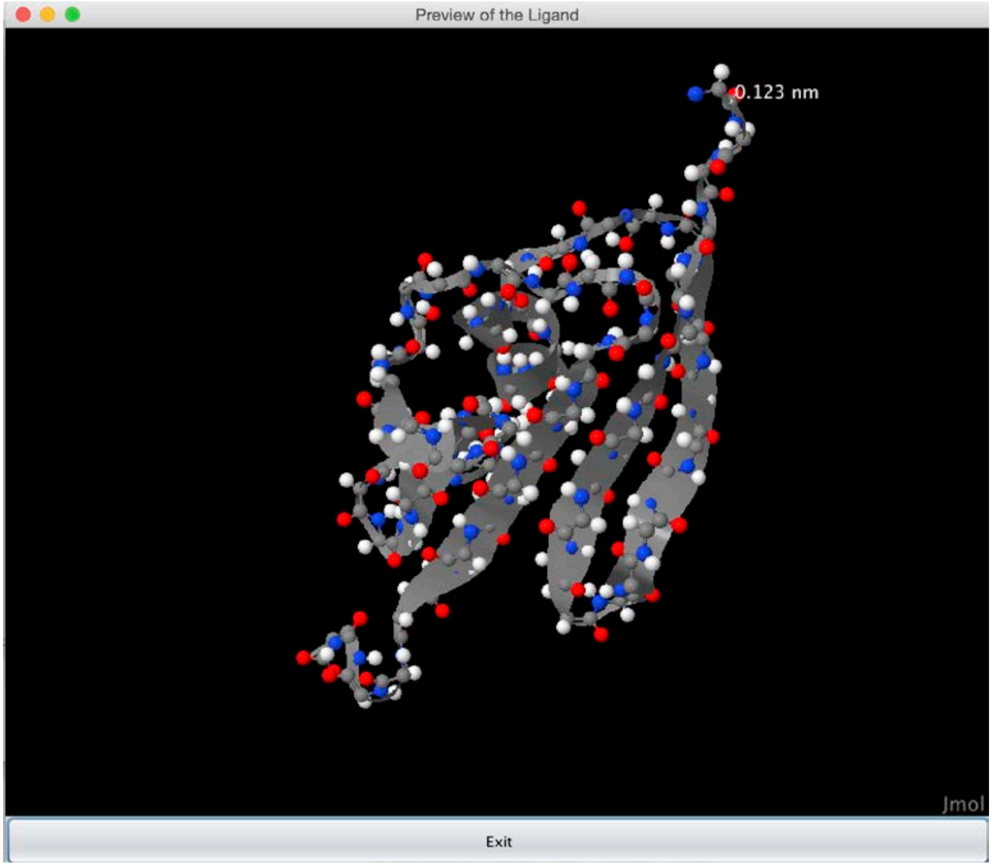
Authors: Yee, A., Chang, X., Pineda-Lucena, A., Wu, B., Semesi, A., Le, B

Release Date: 2004-02-24

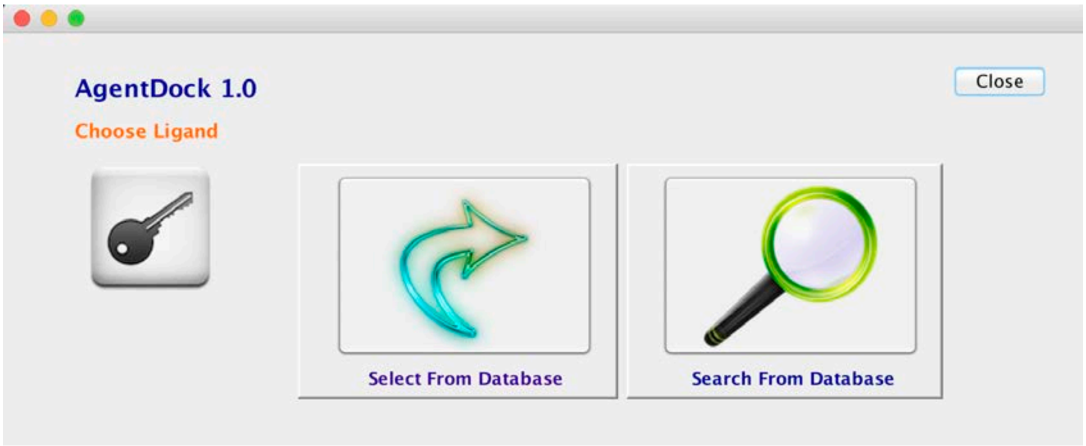
Link: http://www.pdb.org/pdb/explore/explore.do?s

PDB File

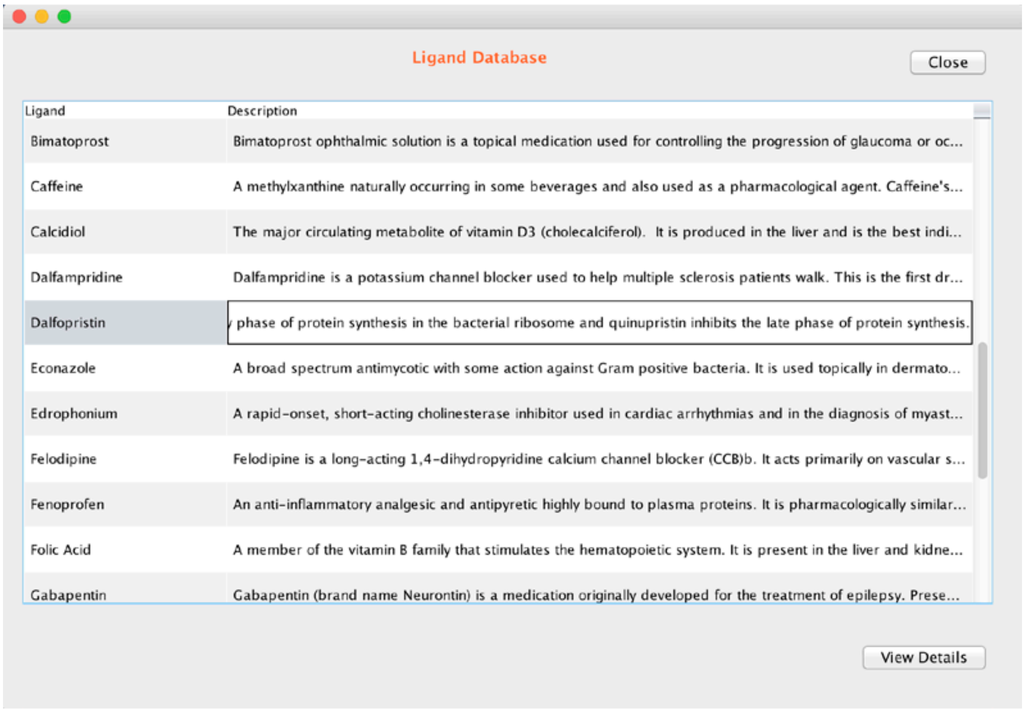
Following diagram shows the 3D preview of selected protein.



AgentDock has a dashboard to select a ligand from the database that the tool has.

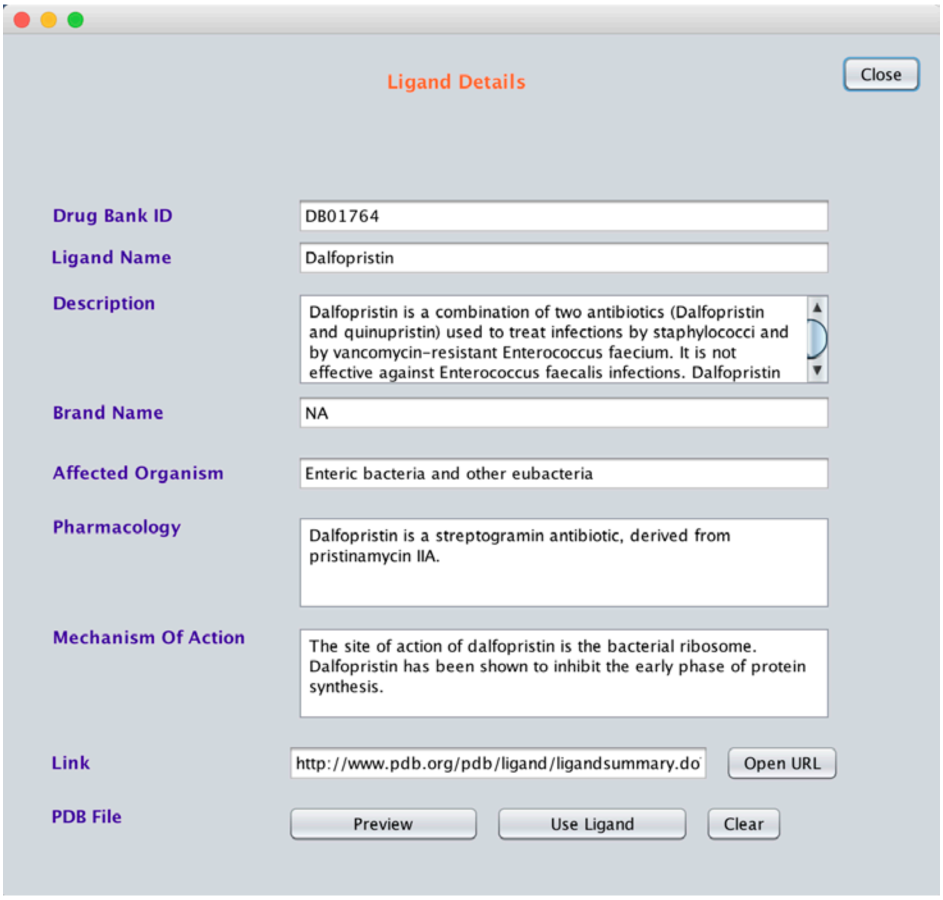


Following diagram shows the list of available ligands in the database.



Ligand	Description
Bimatoprost	Bimatoprost ophthalmic solution is a topical medication used for controlling the progression of glaucoma or oc...
Caffeine	A methylxanthine naturally occurring in some beverages and also used as a pharmacological agent. Caffeine's...
Calcidiol	The major circulating metabolite of vitamin D3 (cholecalciferol). It is produced in the liver and is the best indi...
Dalfampridine	Dalfampridine is a potassium channel blocker used to help multiple sclerosis patients walk. This is the first dr...
Dalfopristin	phase of protein synthesis in the bacterial ribosome and quinupristin inhibits the late phase of protein synthesis.
Econazole	A broad spectrum antimycotic with some action against Gram positive bacteria. It is used topically in dermato...
Edrophonium	A rapid-onset, short-acting cholinesterase inhibitor used in cardiac arrhythmias and in the diagnosis of myast...
Felodipine	Felodipine is a long-acting 1,4-dihydropyridine calcium channel blocker (CCB)b. It acts primarily on vascular s...
Fenoprofen	An anti-inflammatory analgesic and antipyretic highly bound to plasma proteins. It is pharmacologically similar...
Folic Acid	A member of the vitamin B family that stimulates the hematopoietic system. It is present in the liver and kidne...
Gabapentin	Gabapentin (brand name Neurontin) is a medication originally developed for the treatment of epilepsy. Prese...

User can view more details by selecting the required ligand and also can preview it before apply it for the docking.



Ligand Details

Close

Drug Bank ID: DB01764

Ligand Name: Dalfopristin

Description: Dalfopristin is a combination of two antibiotics (Dalfopristin and quinupristin) used to treat infections by staphylococci and by vancomycin-resistant Enterococcus faecium. It is not effective against Enterococcus faecalis infections. Dalfopristin

Brand Name: NA

Affected Organism: Enteric bacteria and other eubacteria

Pharmacology: Dalfopristin is a streptogramin antibiotic, derived from pristinamycin IIA.

Mechanism Of Action: The site of action of dalfopristin is the bacterial ribosome. Dalfopristin has been shown to inhibit the early phase of protein synthesis.

Link: http://www.pdb.org/pdb/ligand/ligandsummary.do

Open URL

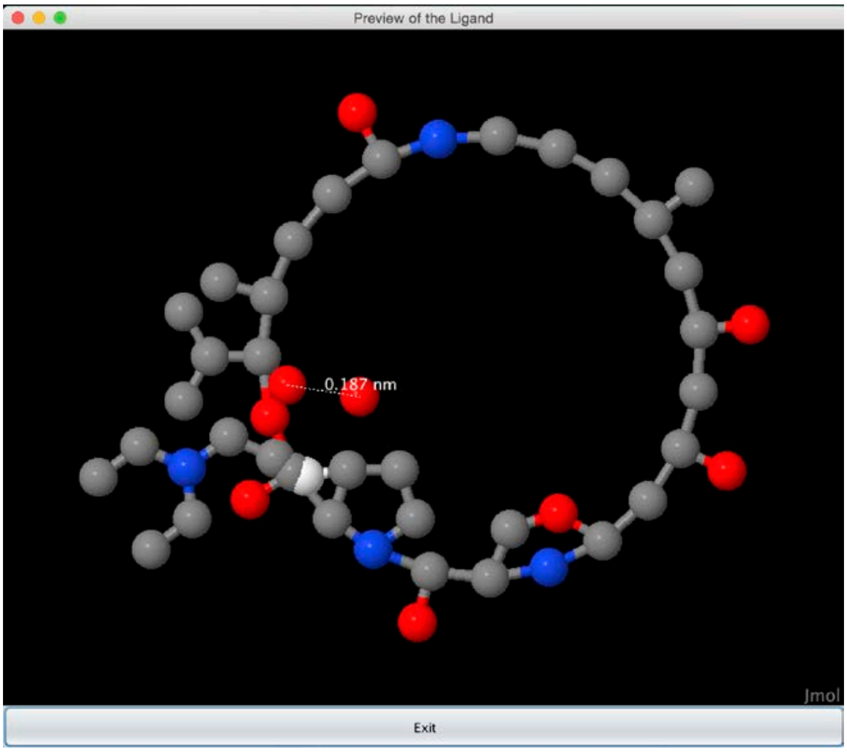
PDB File

Preview

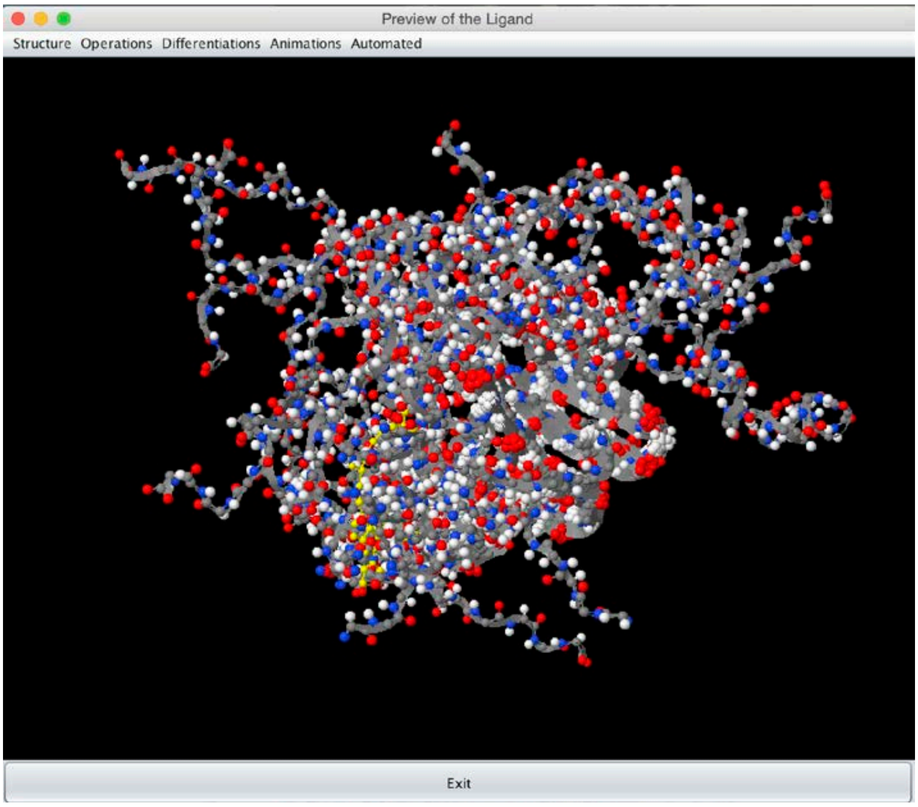
Use Ligand

Clear

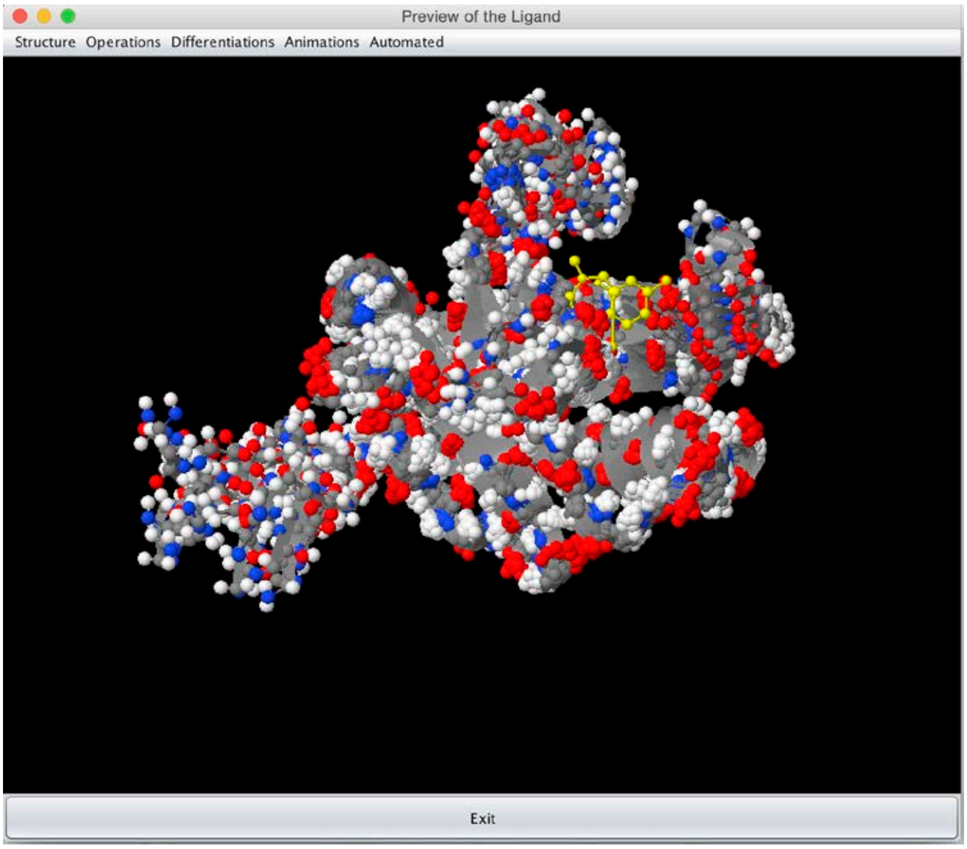
Following figure shows the selected ligand 3d structure preview.



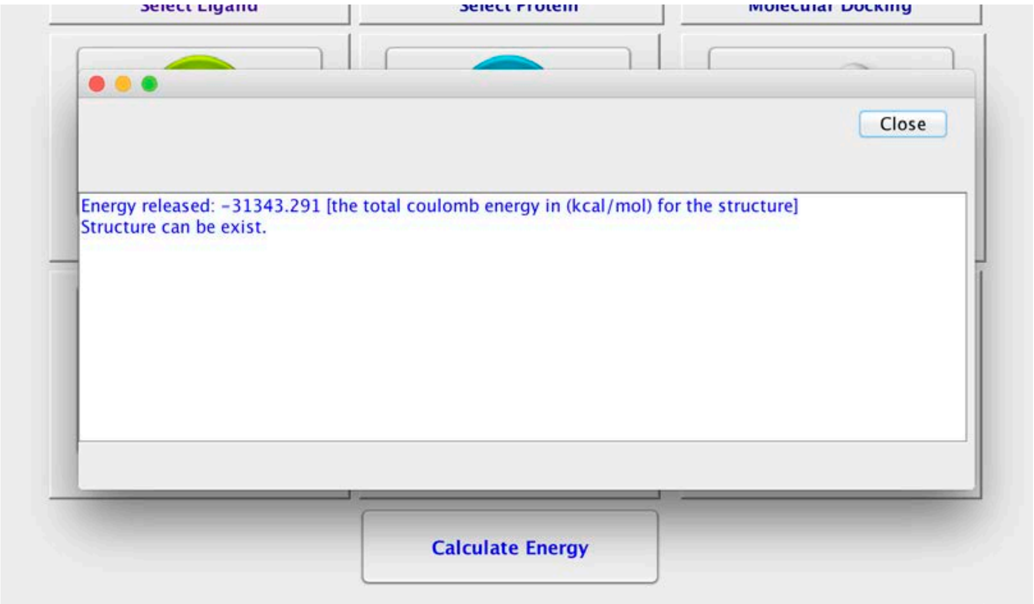
Following diagram shows the result after the molecular docking.



Another docked molecular result shows as follows (yellow color molecule represents the ligand)

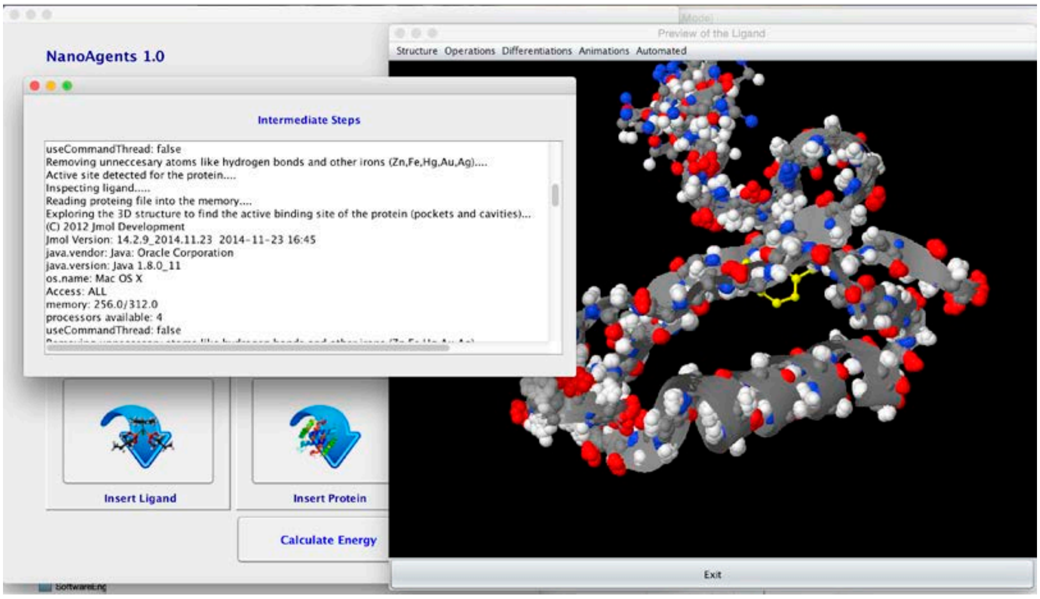


Energy calculation feature shows the actual existence of the docked molecule using Python component for the energy calculations.



```
Energy released: -31343.291 [the total coulomb energy in (kcal/mol) for the structure]
Structure can be exist.
```

Intermediate steps can be viewed in a separate window as logs.

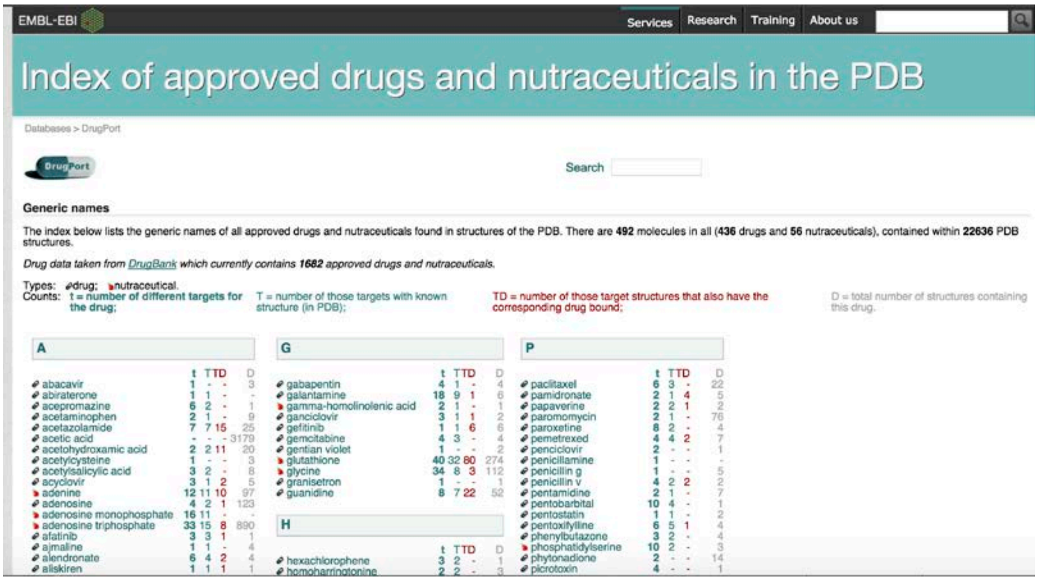


```
useCommandThread: false
Removing unneccesary atoms like hydrogen bonds and other irons (Zn,Fe,Hg,Au,Ag)....
Active site detected for the protein....
Inspecting ligand.....
Reading proteing file into the memory....
Exploring the 3D structure to find the active binding site of the protein (pockets and cavities)...
(C) 2012 Jmol Development
Jmol Version: 14.2.9_2014.11.23  2014-11-23 16:45
java.vendor: Java: Oracle Corporation
java.version: Java 1.8.0_11
os.name: Mac OS X
Access: ALL
memory: 256.0/312.0
processors available: 4
useCommandThread: false
```

Appendix B

Free Databases and Ontologies

B.1 Ligand Database



Index of approved drugs and nutraceuticals in the PDB

Generic names

The index below lists the generic names of all approved drugs and nutraceuticals found in structures of the PDB. There are **492** molecules in all (**436** drugs and **56** nutraceuticals), contained within **22636** PDB structures.

*Drug data taken from DrugBank which currently contains **1682** approved drugs and nutraceuticals.*

Types: drug; nutraceutical.
Counts: t = number of different targets for the drug; T = number of those targets with known structure (in PDB); TD = number of those target structures that also have the corresponding drug bound; D = total number of structures containing this drug.

A

	t	T	TD	D
abacavir	1	-	-	3
abiraterone	1	1	-	-
acepromazine	6	2	-	1
acetaminophen	2	1	-	9
acetazolamide	7	7	15	25
acetic acid	-	-	-	3179
acetohydroxamic acid	2	2	11	20
acetylcysteine	1	-	-	3
acetylsalicylic acid	3	2	-	8
acyclovir	3	1	2	5
adenine	12	11	10	97
adenosine	4	2	1	123
adenosine monophosphate	16	11	-	-
adenosine triphosphate	33	15	8	890
afatinib	3	3	1	1
ajmaline	1	1	-	4
alendronate	6	4	2	4
aliskiren	1	1	1	1

G

	t	T	TD	D
gabapentin	4	1	-	4
galantamine	18	9	1	6
gamma-homolinolenic acid	2	1	-	1
ganciclovir	3	1	1	2
gefitinib	1	1	6	6
gemcitabine	4	3	-	4
gentian violet	1	-	-	2
glutathione	40	32	80	274
glycine	34	8	3	112
granisetron	1	-	-	1
guanidine	8	7	22	52

H

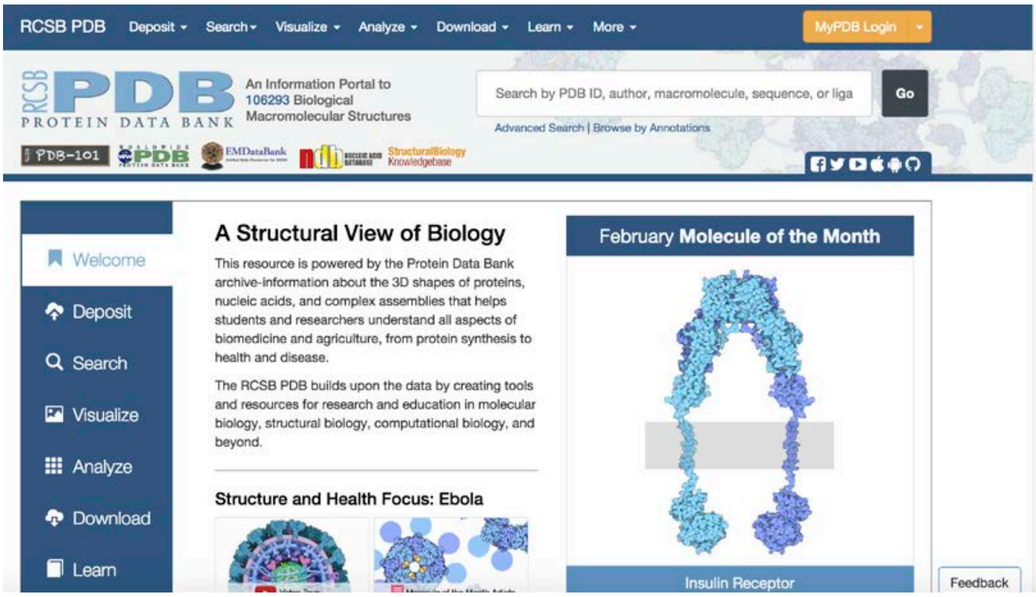
	t	T	TD	D
hexachlorophene	3	2	-	1
homoharringtonine	2	2	-	3

P

	t	T	TD	D
paclitaxel	6	3	-	22
pamidronate	2	1	4	5
papaverine	2	2	1	2
paromomycin	2	1	-	76
paroxetine	8	2	-	4
pemetrexed	4	4	2	7
penciclovir	2	-	-	1
penicillamine	1	-	-	-
penicillin g	1	-	-	5
penicillin v	4	2	2	2
pentamidine	2	1	-	7
pentobarbital	10	4	-	1
pentostatin	1	1	-	2
pentoxifylline	6	5	1	4
phenylbutazone	3	2	-	4
phosphatidylserine	10	2	-	3
phytonadione	2	-	-	14
picrotoxin	4	-	-	1

The index below lists the generic names of all approved drugs and Nutraceuticals found in the structures of the PDB. There are 493 molecules in all (437 drugs and 56 Nutraceuticals), contained within 22935 PDB structures. Drug data taken from DrugBank [34], which currently contains 1682 approved drugs and Nutraceuticals.

B.2 Protein Database


RCSB PDB
Deposit
Search
Visualize
Analyze
Download
Learn
More
MyPDB Login
RCSB PDB PROTEIN DATA BANK
Search by PDB ID, author, macromolecule, sequence, or liga
Go
Advanced Search | Browse by Annotations
PDB-101
WORLDWIDE PDB PROTEIN DATA BANK
EMDataBank
ndb NUCLEIC ACID DATABASE
StructuralBiology Knowledgebase
Welcome
Deposit
Search
Visualize
Analyze
Download
Learn
February Molecule of the Month
Insulin Receptor
Feedback

An Information Portal to 106293 Biological Macromolecular Structures

A Structural View of Biology

This resource is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

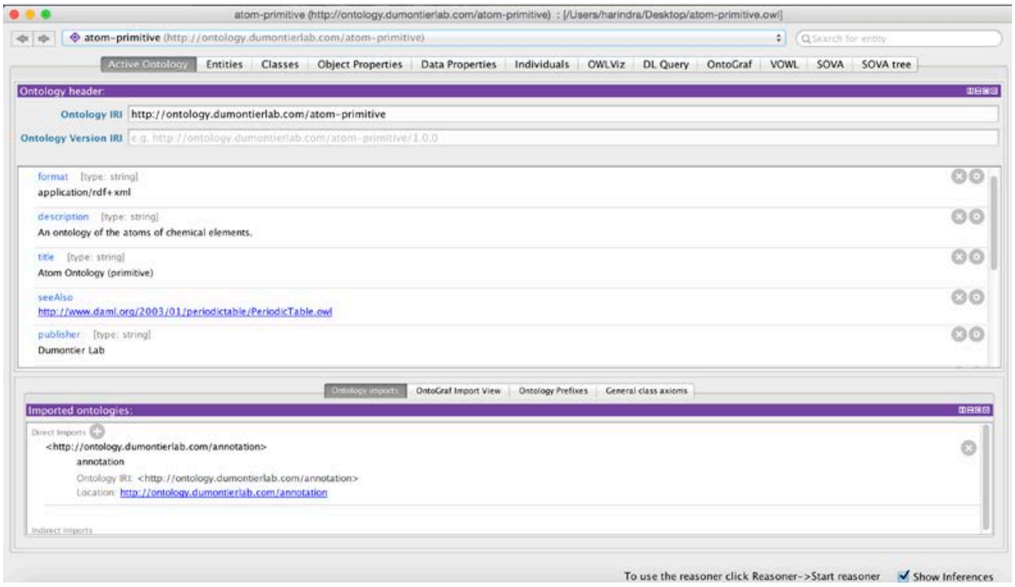
The RCSB PDB builds upon the data by creating tools and resources for research and education in molecular biology, structural biology, computational biology, and beyond.

Structure and Health Focus: Ebola

This resource, is powered by the Protein Data Bank archive-information about the 3D shapes of proteins, nucleic acids, and complex assemblies that helps students and researchers understand all aspects of biomedicine and agriculture, from protein synthesis to health and disease.

B.3 Atom Ontology

Dumontier Lab provides free chemistry ontologies. Their research has significant implications for basic science, drug discovery, and health care. Their investigations into the dynamics of biochemical transformations will lead to improved identification of drug leads, thereby reducing the time and cost of drug discovery [83]. Following diagram shows the ontology in Protégé.



atom-primitive (http://ontology.dumontierlab.com/atom-primitive) : [/Users/harindra/Desktop/atom-primitive.owl]

atom-primitive (http://ontology.dumontierlab.com/atom-primitive)

Ontology header:

Ontology IRI http://ontology.dumontierlab.com/atom-primitive

Ontology Version IRI e.g. http://ontology.dumontierlab.com/atom-primitive/1.0.0

format [type: string]
application/rdf+xml

description [type: string]
An ontology of the atoms of chemical elements.

title [type: string]
Atom Ontology (primitive)

seeAlso
http://www.daml.org/2003/01/periodictable/PeriodicTable.owl

publisher [type: string]
Dumontier Lab

Imported ontologies:

Direct Imports
<http://ontology.dumontierlab.com/annotation>
annotation
Ontology IRI: <http://ontology.dumontierlab.com/annotation>
Location: http://ontology.dumontierlab.com/annotation

Indirect Imports

To use the reasoner click Reasoner->Start reasoner ☑ Show Inferences