MMCG: A Monte Carlo and metropolis based conformer generation tool for flexible docking of small molecules

A Major Project thesis submitted

In partial fulfilment of the requirement for the degree of

Master of Technology

In

Bioinformatics

Submitted by

Ravi Shankar

(DTU/11/M.Tech/252)

Delhi Technological University, Delhi, India

Under the supervision of

Dr. Asmita Das

Department of Biotechnology
Delhi Technological University
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042
INDIA

CERTIFICATE

This is to certify that the M. Tech. dissertation entitled **"MMCG: Flexible docking of small molecules using monte-carlo metropolis approach "**, submitted by **Ravi Shankar (DTU/11/M.Tech/252)** in partial fulfilment of the requirement for the award of the degree of Master of Engineering, Delhi Technological University (Formerly Delhi College of Engineering, University of Delhi), is an authentic record of the candidate's own work carried out by him under my guidance.

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

Date:

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

ACKNOWLEDGEMENT

I express my warm regards to all those who have contributed directly or indirectly towards the fulfillment of this project. I am highly indebted to my mentor, **Dr. Asmita Das,** who has guided me through thick and thin. Her genuine and positive attitude towards research and zest for high quality work has inspired me and prompted me for the timely completion of my project.

I am thankful to **Prof. P. B. Sharma** (Honorable Vice Chancellor, DTU), for according me permission to undertake the project work at IIT, Delhi for partial fulfillment for the degree of Master of technology.

I express my gratitude to **Prof. B.jayaram** (Head of Department) and **Ms. Tanya Singh** , Department of Chemistry, IIT Delhi for their encouragement and support

Last but not the least my heartfelt thanks to my colleagues and friends for their belief and love.

RAVI SHANKAR

(2K11/BIO/16)

CONTENTS

LIST OF FIGURES AND TABLES

configurations considering 6 degrees of freedom and best energy structure is selected.

Figure 16: A: 1Com (Chosrimate mutase) B: 1ivf (Neuraminidase) alignment of the MMCG translational resultant complex with the structures bound in native form.

Figure 17: Alignment of the minimized docked structure of Neuraminidase (1ivf) with the native bound form of the complex.

Table 1. RMSD Analaysis Using self docking for 122 Protein-Ligand Complexes database

LIST OF ABBREVIATIONS

RMSD	:	Root Mean Square Deviation
HTS	:	High Throughput Screening
VS	:	Virtual Screening
QSAR	:	Quantitaive Structure Activity Relationship
MD	:	Molecular Dynamics
IC	:	Incremental Construction algorithm
GA	:	Genetic Algorithm
MCM	:	Monte Carlo Methods
NMR	:	Nuclear Magnetic Resonance
GAFF	:	General Amber Force Fields
AM1 BCC	:	Semi-empirical (AM1) with bond charge correction (BCC)
PDB	:	Protein Data Bank

MMCG: A Monte Carlo and metropolis based conformer generation tool for flexible docking of small molecules

Ravi shankar

Delhi Technological University, Delhi, India

1. ABSTRACT

Drug designing is one of the major thrust area of research these days as it deals with discovery of inhibitors or leads for proteins or targets responsible for various medical conditions. In quest of search for new leads for targets- *in sillico* drug designing has emerged as very efficient system. Molecular flexibility is one of the well known problems in computer aided drug designing as while mimicking biological system using in sillico approach we have to take in consideration that the molecules do not act as rigid structures moreover there is fluidity of system also which provides some degree of flexibility to the molecules. Up till know many docking softwares have been developed which accounts for molecular flexibility but the needed accuracy that is close to bound form of protein is achieved in very limited cases. Here we introduce a new approach to incorporate ligand flexibility in molecular docking system using monte-carlo metropolis simulations. This system produces 100 conformational decoys from a starting structure by random translational and rotational moves and deciding on their acceptance using the Metropolis criterion. The configurational search space is described using ParDOCK- all atom energy-based Monte Carlo, protein-ligand docking algorithm for rigid docking. Structures with most appropriate conformations and respective configurations are picked through our system to produce results as output. The results produced are further refined by minimization of complex using AMBER10 module. Through this approach we are able to capture conformers having RMSD < 2 Å as compared to the bound form of complex.

Keywords : molecular docking, monte-carlo simulations, metropolis criterion, ligand flexibility, minimization, RMSD.

2. INTRODUCTION

The number of drug-discovery projects that have a high-resolution crystal structure of the receptor available has increased in recent years and is expected to continue to rise because of the human genome project and high-throughput crystallography efforts. A common computational strategy in such a case is to dock molecules from a physical or virtual database into the receptor and to use a suitable scoring function to evaluate the binding affinity. But due to high complexity of the biological systems and limitations of the computational power it's not easy to mimic the biological process of protein and ligand interaction. Ever since the discovery on computer aided drug designing (CADD) the main focus is put on the development of efficient and more robust docking algorithms. Various methodologies of docking systems are used most common among them is rigid docking. But the main problem with rigid docking is that it does not produce biologically relevant results in most of the cases. The main reason behind inefficiency of rigid docking is ignorance of the flexibility of proteins and ligands in actual biological interactions. Here we introduce a different docking methodology- MMCG which takes flexibility of the ligand in consideration during molecular docking and also requires relatively less computational power.

MMCG program first pick up the flexible bonds within a ligand then based on monte-carlo simulations it produces a large number of decoys but among these decoys the relevant conformers are chosen based on metropolis approach. The docking process involves two basic steps: prediction of the ligand conformation as well as its position and orientation within these sites which is referred as binding pose and assessment of the binding affinity. These two steps are related to sampling methods and scoring schemes, respectively. The configurational search space is assessed through a rigid protein-ligand docking program-ParDock. The scoring function used is - an all atom energy based empirical scoring function comprising electrostatics, van der Waals, desolvation and loss of conformational entropy of protein side chains upon ligand binding. MMCG produces five possible binding poses and conformations of a protein-ligand complex. These resulting complexes are further refined by minimization of the complex using AMBER 10 force field methods. MMCG algorithm has been validated on 119 complexes of a database and it has been observed that this algorithm can attain nearly 80 percent accuracy in producing rmsd<2.0 Å as compared to bound form of the protein present in the database. programs run on linux clusters having infiniband network resources which facilitate a high through put distribution of the data across the various nodes. On an average, the total time taken by the complete docking and scoring protocol ranges from 15-40 minutes depending on the size of the protein and the ligand. The above time frames reported correspond to performance on a 32 processors cluster. The automated version of MMCG runs on atleast 18 processors for a single job . Memory consumption and I/O issues are minimal during program execution. The time taken also depends on the load on the server.

3. REVIEW OF LITERATURE

3.1 In-sillico Drug discovery

Use of computational techniques in drug discovery and development process is rapidly gaining in popularity, implementation and appreciation. Different terms are being applied to this area, including computer-aided drug design (CADD), computational drug design, computer-aided molecular design (CAMD), computer-aided molecular modeling (CAMM), rational drug design, *in silico* drug design, computer-aided rational drug design. Term Computer-Aided Drug Discovery and Development (CADDD) will be employed in this overview of the area to cover the entire process. Both computational and experimental techniques have important roles in drug discovery and development and represent complementary approaches. CADDD entails:
1. Use of computing power to streamline drug discovery and development process.
2. Leverage of chemical and biological information about ligands and/or targets to identify and optimize new drugs.
3. Design of *in silico* filters to eliminate compounds with undesirable properties (poor activity and/or poor Absorption, Distribution, Metabolism, Excretion and Toxicity, ADMET) and select the most promising candidates.(I.M. Kapetanovic *et al.,* 2006).

The completion of the human genome project has resulted in an increasing number of new therapeutic targets for drug discovery. At the same time, high-throughput protein purification, crystallography and nuclear magnetic resonance spectroscopy techniques have been developed and contributed to many structural details of proteins and protein–ligand complexes. These advances allow the computational strategies to permeate all aspects of drug discovery today (Jorgensen WL *et al.,* 2004; Bajorath J *et al.,* 2002; Kitchen DB *et al.,* 2004), such as the virtual screening (VS) techniques for hit identification and methods for lead optimization. Compared with traditional experimental high-throughput screening (HTS), VS is a more direct and rational drug discovery approach and has the advantage of low cost and effective screening (Moitessier N *et al.,* 2008; Shoichet BK *et al.,* 2002; Bailey D *et al.,* 2001). VS can be classified into ligand-based and structure-based methods. When a set of active ligand molecules is known and little or no structural information is available for targets, the ligand-based methods, such as pharmacophore modeling and quantitative structure activity relationship (QSAR) methods can be employed. As to structure-based drug design, molecular docking is the most common method which has been widely used ever since the early 1980s. (Kuntz ID *et al.,* 1982)

3.2 Molecular Docking

In the field of molecular modeling, docking is a method which predicts the preferred orientation of one molecule to a second when bound to each other to form a stable complex. Knowledge of the preferred orientation in turn may be used to predict the strength of association or binding affinity between two molecules Different algorithms for structure-based design can be divided into roughly two classes: de novo design, which builds ligands tailored to the target, and

docking, which searches for existing compounds with good complementarity to the target. In both these paradigms, the enzyme or receptor has traditionally been treated as a rigid body and only one conformation of the enzyme is considered. (Examples of de novo design include Lewis (Lewis, 1992) and Miranker (Miranker & Karplus, 1995) and the program LUDI (BoÈhm, 1992a,b);examples of molecular docking include the works of (Kuntz *et al*., 1982).

3.3 Docking methodologies

Programs based on different algorithms were developed to perform molecular docking studies, which have made docking an increasingly important tool in pharmaceutical research. Molecular docking using computational systems incorporates different docking methodologies which are – rigid ligand rigid receptor docking, flexible ligand rigid receptor docking and flexible ligand flexible receptor docking (Xuan-Yu Meng *et al.,* 2011).

3.3.1 Rigid ligand and rigid receptor docking

Primitive molecular docking algorithms uses rigid ligand and rigid receptor methodology due to limitation of computational power. In this case the search space is very limited, considering only three translational and three rotational degrees of freedom. In this case, ligand flexibility could be addressed by using a pre-computed a set of ligand conformations, or by allowing for a degree of atom–atom overlap between the protein and ligand. DOCK (Kuntz ID *et al.,* 1982) is the first automated procedure for docking a molecule into a receptor site and is being continuously developed. It characterizes the ligand and receptor as sets of spheres which could be overlaid by means of a clique detection procedure (Bron C *et al.,* 1973). Geometrical and chemical matching algorithms are used, and the ligand-receptor complexes can be scored by accounting for steric fit, chemical complementation or pharmacophore similarity.

3.3.2 Flexible ligand and rigid receptor docking

The main problem in molecular docking arises due to consideration of flexibility of ligands when present in biological system. As stated through induced–Fit mechanism (Koshland DE Jr. *et al.,*1963; Hammes GG *et al.,* 2002) the active site of the protein is continually reshaped by interactions with the ligands as the ligands interact with the protein. This theory suggests that the ligand and receptor should be treated as flexible during docking. However, the computational cost is very high when the receptor is also flexible. Thus the common approach, also a trade-off between accuracy and computational time, is treating the ligand as flexible while the receptor is kept rigid during docking. Almost all the docking programs have adopted this methodology, such as Auto Dock (Morris GM *et al.,* 1998), FlexX (Rarey M *et al.,* 1996). AutoDock 3.0 incorporates Monte Carlo simulated annealing, evolutionary, genetic and Lamarckian genetic algorithm methods to model the ligand flexibility while keeping the receptor rigid. The scoring function is based on the AMBER force field, including van der Waals, hydrogen bonding, electrostatic interactions, conformational entropy and desolvation terms.

3.3.3 Flexible ligand and flexible receptor docking

The intrinsic mobility of proteins has been proved to be closely related to ligand binding behavior and it has been reviewed. (Teague *et al.,* 2003). Incorporating the receptor flexibility is significant challenge in the field of docking. Ideally, using MD simulations could model all the degrees of freedom in the ligand-receptor complex. But MD has the problem of inadequate sampling that we mentioned earlier. Another hurdle is its high computational expense, which prevents this method from being used in the screening of large chemical database. Various methods are currently available to implement the receptor flexibility . The simplest one is so-called "soft-docking" (Jiang F *et al.,* 1991), decreases the van der Waals repulsion energy term in the scoring function to allow for a degree of atom-atom overlap between the receptor and ligand.

3.4 Docking algorithms

Various sampling algorithms have been developed and widely used in molecular docking software. Matching algorithms (MA) (Brint AT *et al.,* 1987; Fischer D *et al.,* 1993) based on molecular shape map a ligand into an active site of a protein in terms of shape features and chemical information. The protein and the ligand are represented as pharmacophores. Each distance of the pharmacophore within the protein and ligand is calculated for a match; new ligand conformations are governed by the distance matrix between the pharmacophore and the corresponding ligand atoms. Chemical properties, like hydrogen-bond donors and acceptors, can be taken into account during the match. Matching algorithms have the advantage of speed; thus they may be used for the enrichment of active compounds from large libraries (Moitessier N *et al.,* 2008). Stochastic methods search the conformational space by randomly modifying a ligand conformation or a population of ligands.

3.4.1 Incremental construction (IC)

These methods put the ligand into an active site in a fragmental and incremental fashion. The ligand is divided into several fragments by breaking its rotatable bonds and then one of these fragments is selected to dock into the active site first. This anchor is usually the largest fragment or the piece which may have significant functional role or interaction with protein. The remaining fragments can be added incrementally. Different orientations are generated to fit in the active site, which realizes the flexibility of the ligand. The incremental construction method has been used in DOCK 4.0 (Ewing T.J *et al.,* 2001), FlexX (Rarey M *et al.,* 1996).

3.4.2 Genetic algorithm (GA)

Genetic algorithms (GA) (Morris GM *et al.,* 1998; Jones G *et al.,* 1997; Oshiro CM *et al.,* 1995) form a class of well-known stochastic methods. The idea of the GA stems from Darwin's theory of evolution. Degrees of freedom of the ligand are encoded as binary strings called genes. These genes make up the 'chromosome' which actually represents the pose of the ligand. Mutation and

crossover are two kinds of genetic operators in GA. Mutation makes random changes to the genes; crossover exchanges genes between two chromosomes. When the genetic operators affect the genes, the result is a new ligand structure. New structures will be assessed by scoring function, and the ones that survived (i.e., exceeded a threshold) can be used for the next generation. Genetic algorithms have been used in AutoDock, GOLD (Verdonk ML *et al.,* 2003) and DARWIN (Taylor JS *et al.,* 2000).

3.4.3 Monte Carlo methods (MCM)

Experimental studies have demonstrated that a protein is not a static structure but instead undergoes fluctuations. Based on photo dissociation studies of carbon monoxide bound to myoglobin, it has been suggested that a protein can exist in a large number of conformational substrates separated by barriers, with transitions among substrates constituting equilibrium fluctuations (Goodsell DS *et al.,* 1993). A recent molecular dynamics study of myoglobin (Hart TN *et al.,* 1992) reported the existence of many minima in the vicinity of the native protein; these corresponded to relative reorientations of the a-helices coupled with rearrangements of the side chains, as a consequence of the internal dynamics of the protein. It follows, as a necessary condition that a structure be stable, that the native conformation of a protein must be stable not only against small disturbances but also against larger-scale thermal fluctuations; i.e., the native structure must be able to recover from any thermal impulse, even though the latter may (temporarily) lead to a different local minimum-energy even during docking complex formation. Monte Carlo (MC) (Goodsell DS *et al.,* 1993) methods generate poses of the ligand through bond rotation, rigid-body translation or rotation. The conformation obtained by this transformation is tested with an energy- based selection criterion. If it passes the criterion, it will be saved and further modified to generate next conformation. The iterations will proceed until the predefined quantity of conformations is collected. The main advantage of MC is that the change can be quite large allowing the ligand to cross the energy barriers on the potential energy surface, a point that isn't achieved easily by molecular dynamics based simulation methods. Examples of applying the Monte Carlo methods include an earlier version of Auto Dock (Goodsell DS et al. 1993), ICM (Abagyan R *et al.,* 1994)

3.5 Metropolis criterion

This criterion generates configurations according to the desired statistical-mechanics distribution. There is no time; the method cannot be used to study evolution of the system. Equilibrium properties can be studied. The Monte Carlo metropolis criterion is used to test the acceptance of generated conformer so that it doesn't get struck in local minima instead of reaching global minima.

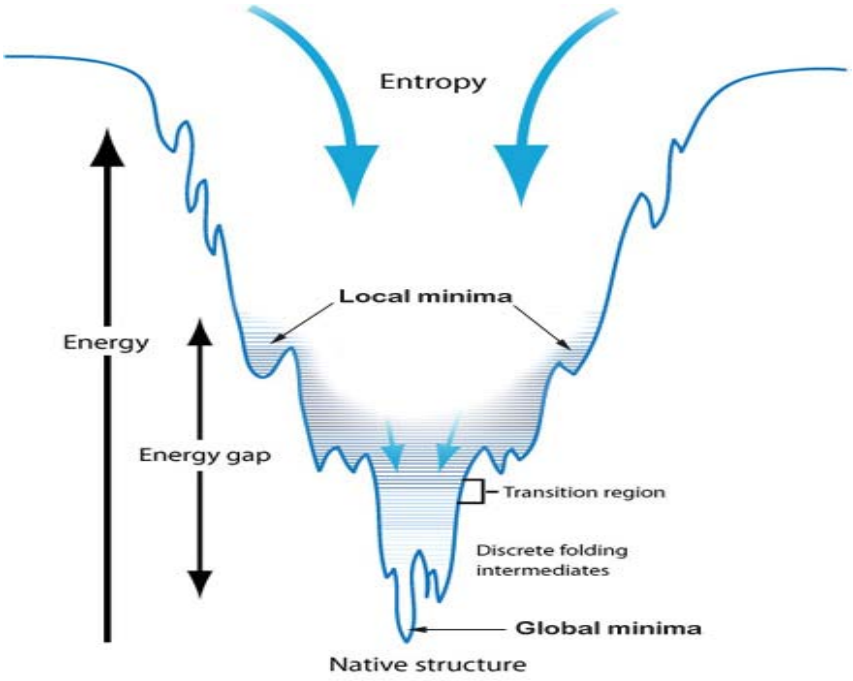


Fig.1 : Energy Funnel depicting descend of structures to local minima and global minima

The steps performed leads us to minimum energy structure but there is a much higher possibility that instead of attaining global minima the structure get strucked in local minima. To eradicate this problem an improved monte-carlo simulations method is used which uses metropolis criterion to solve this local minima problem. Statistical mechanics tells us that the probability pi of finding a system at constant number N, volume V and temperature T in a microstate i with total energy E_i is proportional to

$$p_i = \frac{\exp\left[-\beta E_i\right]}{Q(N,V,T)}$$

where the inverse temperature $\beta = 1/k_BT$, and k_B is Boltzmann's constant. The partition function Q(N, V, T) is defined as the sum over all states:

$$Q(N,V,T) = \sum_i \exp\left[-\beta E_i\right]$$

the method we employ is actually a modified Monte Carlo scheme, where, instead of choosing configurations randomly, then weighing them with exp(-E/kT), we choose configurations with a probability exp(-E/kT) and weight them evenly.

3.5.1 Monte Carlo metropolis algorithm

The algorithm works on the basis of making move from one point to another and acceptance of that move is done through it. (Metropolis *et al.,* 1953)

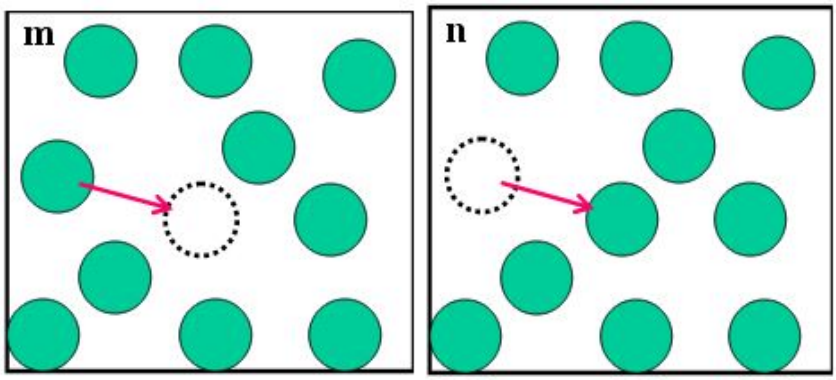


Fig. 2 : Monte Carlo moves m and n depicts the move of object from one point to another in a consecutive manner

1. Choose the initial configuration, calculate energy

2. Make a "move" (e.g., pick a random displacement). Calculate the energy for new "trail" configuration.

3. Decide whether to accept the move: if Unm = Un–Um < 0, then accept the new configuration,

$$W(m \rightarrow n) = \exp\left(-\frac{U_{nm}}{kT}\right)$$

Where, U is internal energy of the system, k is standard gas constant and T is temp at 273 k.

if Unm = Un–Um > 0, then calculate draw a random number R from 0 to 1. if W(m→n) > R then accept the new configuration, otherwise, stay at the same place.

4. Repeat from step 2, accumulating sums for averages (if atom is retained at its old position, the old configuration is recounted as a new state in the random walk).

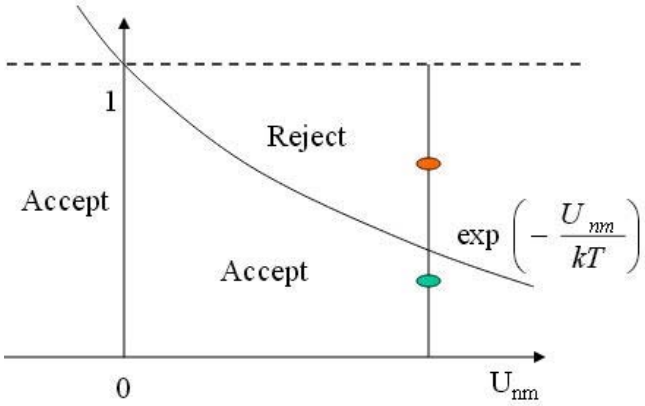


Fig.3 : Acceptance and rejection of new move on the basis of calculation of exponential value

3.6 Configurational search

Searching for the correct binding mode (pose prediction) of a molecule is typically carried out by performing a number of trials and keeping those poses that are energetically best. It involves finding the correct orientation and, as most ligand molecules are flexible, the correct conformation of the docked molecule. This implies that the degrees of freedom to be searched include translational and rotational degrees of freedom of the ligand as a whole, as well as its internal degrees of freedom, i.e., predominantly the rotatable bonds. The search stops once a certain number of trials have been carried out and/or a sufficient number of poses have been found for a molecule. In order to explore a large search space, algorithms have been developed that keep track of previously discovered minima and guide the search into new regions. The decision to keep a trial pose is based on the computed ligand–receptor interaction energy (score) of that pose. To identify and rank-order many different poses of a molecule during the search in a reasonable time, several programs calculate a 'dock score' (a crude score based on a simple energy function such as a force field with an electrostatic term and repulsive and attractive Van-der-Waals terms), which can be evaluated very rapidly during the docking process, while a more sophisticated function is used to calculate the final 'affinity score' for that molecule.

For the purpose of generation of effective docking system the search for correct binding pose is an essential step. an all atom energy based Monte Carlo docking protocol christened 'ParDOCK' (parallel dock) (B. jayaram *et al.*, 2007) implemented in a fully automated, parallel processing mode**.** In this the Monte Carlo method in six dimensional space is implemented to generate a large number of random configurations of the ligand in search of optimal location in the binding pocket of the target macromolecule. A combination of an all atom energy based scoring with a Monte Carlo search technique appears to provide a reliable method for protein ligand structure optimization and binding affinity prediction as the results indicate.

3.7 Structural Minimization

A protein structure is flexible. Atoms in a protein have a certain amount of freedom and are moving constantly with respect to each other. The simple forces of nature are also valid for these atoms. For example, "atoms that are very close will repulse each other, but atoms that are further away will attract each other" and "positive and negative charges will attract each other, while equal charges repulse each other". Using this knowledge and a whole bunch of other rules we can calculate the energy of all atoms, and thus of the total protein. A protein in its most favourable situation has the lowest possible energy. Computer programs can be used to find this lowest energy conformation. Every atom is moved in very small steps and following each step the total energy is calculated. So, during the energy minimization process a protein is moved towards its lowest energy conformation which is most favourable conformation. For this purpose AMBER program module is used. Amber is the collective name for a suite of programs that allow users to carry out molecular dynamics simulations, particularly on biomolecules. None of

the individual programs carries this name, but the various parts work reasonably well together, and provide a powerful framework for many common calculations. (Pearlman D.A *et al* , 1995; Case D.A *et al* , 2005). The term amber is also sometimes used to refer to the empirical force fields that are implemented here. (Ponder J.A, 2003; Cheatham T.E, 2005) It should be recognized however, that the code and force field are separate: several other computer packages have implemented the amber force fields, and other force fields can be implemented with the amber programs. Further, the force fields are in the public domain, whereas the codes are distributed under a license agreement. The Amber 10 programs mainly use dynamic memory allocation, and do not generally need to be compiled for any specific size of problem. Some sizes related to NMR refinements are pre-defined in the files, you may need to edit them, then recompile. If you get a "Killed" (or similar) message immediately upon starting a program (particularly if this happens with no arguments), you may not have enough memory to run the program. The "size" command will show you the size of the executable. Also check the limits of your shell; you may need to increase these (especially stacksize, which is sometimes set to quite small values).

4. METHODOLOGY

MMCG system uses a hierarchical setup comprising of various steps, starting from searching flexible bonds in the ligand to docking the ligand of appropriate conformer in right binding pose or configuration. Whole MMCG program works on a Linux based server utilizing at least 18 processors per job. All the codes are written using C-platform and shell scripting which makes program easy to progress step by step with high processing speed.

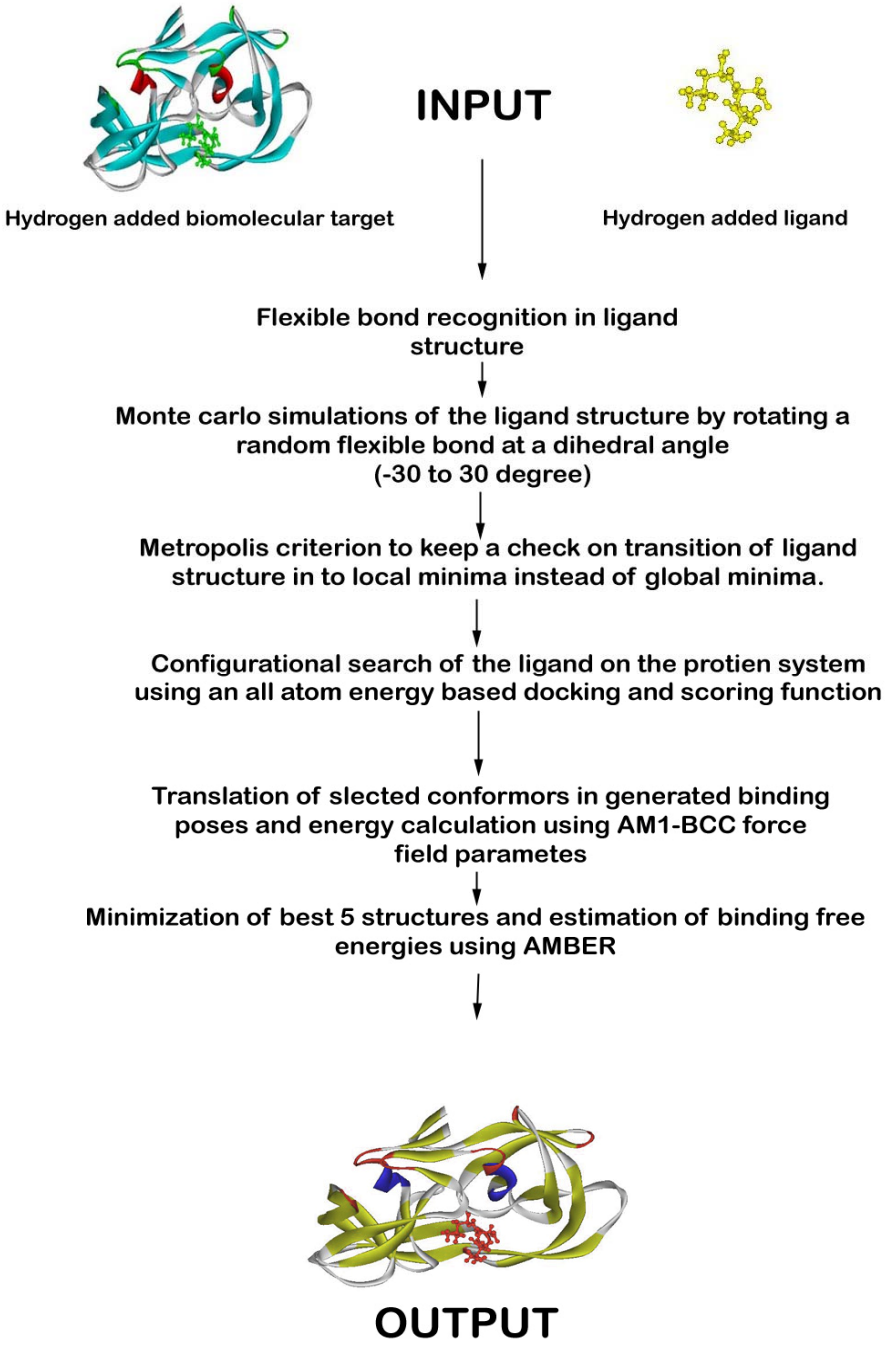


Fig.4 : Flow chart for the flexible ligand docking using MMCG

4.1 Preparation of proteins and ligand for docking

The complexes chosen for study are adapted from RCSB and prepared in a force field compatible manner. Initially the crystallographic water molecules are removed and the ligand coordinates are extracted from the protein-ligand complex. Hydrogen atoms are added keeping the ionization states of the atoms in the ligand as specified in the literature. The ligand is then geometry optimized through AM1 procedure followed by calculation of partial charges of the ligand by AM1-BCC procedure (Arazjakalian *et al.,* 2000). GAFF force field (Wang J. *et al.*,2004) is used to assign atom types (Cornell W.D. *et al.,* 1995), bond angle, dihedral and van der Waals parameters for the ligand.

4.2 Flexible bond recognition

The first and most important step in introduction of flexibility in molecular docking is to pick all the flexible bonds present in the ligand structure accurately. As the flexibility of whole ligand structure depends on number of rotatable bonds present in the structure. This is step is also important for further Monte Carlo simulations of ligands to generate conformers as if didn't pick right flexible bonds than the conformers generated will have distorted structures which is not-acceptable.

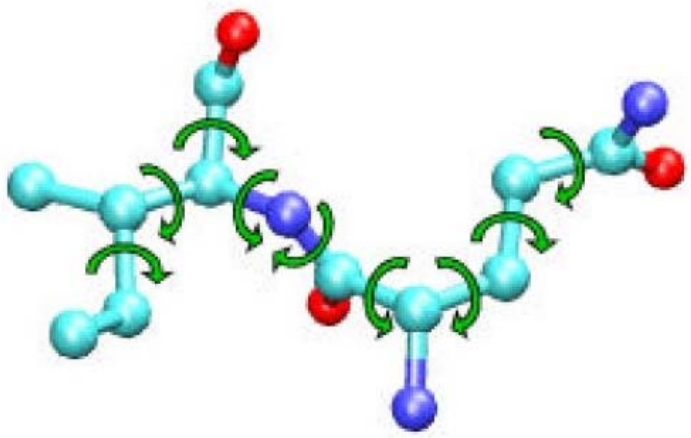


Fig.5: A drug molecule. Sphere represents atoms and bonds connecting them are represented by sticks. Curved arrows represent the rotatable degrees of freedom around bonds.

For this purpose MMCG uses a set of programs which recognises the bond interactions present within the ligand structure and accordingly picks up the bonds as rotatable or non- rotatable. The flexible bond recognition program first reads the PDB file of input ligand structure. It translates the PDB file to define the bond types and bonding patterns that is connect information of all the elements present in the ligand. It first checks for all the ring structures as these are most prominently seen in ligands and there are no rotatable bonds in the ringed structure. After checking for the number of rings present in the ligand structure, program defines the interaction

pattern of the ligand elements. It is generally considered that all the single bonds allow a considerable amount of flexibility in the structure making them as rotatable bonds whereas all the double and triple bonds incorporates rigidity in the structure such bonds are marked as non-rotatable.

4.3 Monte Carlo simulations

The Monte Carlo simulations has proven to be much efficient approach to search a high dimensional conformational space rather than discreet states. This system stimulates the natural thermodynamic process by taking in account both random fluctuation and energetic considerations, it is well applicable in molecular docking systems also. Monte Carlo (MC) methods generate poses of the ligand through bond rotation, rigid-body translation or rotation. MMCG system contains a monte-carlo simulation program to generate a decoy library among which the appropriate conformers are selected.

For this purpose MMCG takes in account all the flexible or rotatable bonds identified by flexible bond recognition program. Now for monte - carlo simulations a random flexible bond is picked by the system and this random bond is than rotated to a random dihedral angle of

$(-30° \leq \theta \leq 30°)$ and the energy of conformer is calculated by a program of AMBER 10 module. These steps are repeated several times , the low energy conformer is to be accepted and a high energy conformation is rejected. But with this approach there is a higher possibility that instead of reaching global minima the structure get struck in local minima that's why the conformation obtained by this transformation is tested with an energy- based selection criterion. If it passes the criterion, it will be saved and further modified to generate next conformation. The iterations will proceed until the predefined quantities of conformations are generated.

The criterion used by MMCG system is Metropolis criterion which simulates natural thermal processes, by taking into account both random fluctuations and energetic considerations, it might be applicable to protein folding. The successful application of the simulated annealing method which is essentially a Metropolis Monte Carlo simulation technique with an artificial "temperature," to the computationally difficult problem considering presence of multiple local minima. This local minimum is examined by the Metropolis criterion to compare it with the previously accepted local minimum to update the current conformation. As a consequence, the transition probabilities of the series of local minima generated in the Monte Carlo simulations satisfy the Boltzmann distribution (Abagyan R *et al.,* 1994). is repeated to continue the iteration process, which generates a Markov sequence with Boltzmann probabilities.

Metropolis Monte Carlo algorithm: implementation

1. Choose the initial configuration of the system, calculate energy

2. Loop through all the particles

   For each particle pick a random displacement d = (random # - 0.5)*d_{max} for x, y and z coordinates. Here d_{max} is the maximum displacement, and random # is from 0 to 1.

   Calculate the energy change ΔU due to the displacement.

   Decide whether to accept the move based on the Metropolis criterion:

   if $\Delta U < 0$, then accept the new configuration,

   if $\Delta U > 0$, then calculate $W = \exp\left(-\frac{\Delta U}{kT}\right)$

   draw a random number R from 0 to 1

   if W > R then accept the new configuration, otherwise, keep the old one

   Accumulate sums for averages (if atom is retained at its old position, the old configuration is recounted as a new state).
   Pick the next particle…

3. If # of MC cycles is < then maximum # of cycles, Go back to step 2.

Fig.6 : Implementation methodology of metropolis criterion with Monte Carlo simulations.

4.4 Docking and scoring

The docking module of MMCG comprises of ParDOCK: An All Atom Energy Based Monte Carlo Docking Protocol for Protein-Ligand Complexes. Here the Monte Carlo method in six dimensional space is implemented to generate a large number of random configurations of the ligand in search of optimal location in the binding pocket of the target macromolecule. A combination of an all atom energy based scoring with a Monte Carlo search technique appears to provide a reliable method for protein ligand structure optimization and binding affinity prediction as the results indicate. The module requires a reference protein-ligand complex (target protein bound to a reference ligand at its binding site) as an input along with the candidate molecule to be docked. The ParDOCK protocol consists of four main steps:

(a) Identification of the best possible grid/ translational points in radius of 3Å around the reference point (centre of mass)
(b) Generation of protein grid and preparation of energy grid in and around the active site of the protein to pre-calculate the energy of each atom in the candidate ligand
(c) Monte Carlo docking and intensive configurational search of the ligand inside the active site
(d) Identification of the best docked structures on an energy criterion and prediction of the binding free energy of the complex.

The algorithm docks the ligand molecule to the reference protein and outputs five docked structures representing different poses of ligand molecule along with the predicted binding free energies of the docked poses using a unique scoring function. The ligand configurations generated are ranked based on an all atom energy function, which calculates non-bonded interactions of protein-ligand complexes as described (Jain, T *et al.,* 2005) in equation.

$$E = \sum E_{el} + E_{vdw} + E_{hpb}$$

E is the total non-bonded energy, Eel is the electrostatic contribution to the energy, E_{vdw} is the van der Waals term, E_{hpb} is the hydrophobic term and the summation runs over all the atoms of the protein-ligand complex. Electrostatic contribution is calculated by Coulomb's law with sigmoidal dielectric function, van der Waals term is computed using a (Inbal *et al.,* 2002; Lengauer *et al.,*1996) Lennard-Jones potential (Ajay *et al.,* 1995) between the atoms of protein and ligand and hydrophobic interactions are calculated by Gurney parameter approach (Ramanathan P.S. *et al.,* 1971; Friedman,H.L *et al.,* 1973)

4.5 Translation of conformers in binding poses

Conformers generated through monte-carlo simulations are selected on the basis of least binding energies . The calculation of the energies of conformers is done using AMBER 10 force field module. Selected conformers are translated on the binding poses as derived through rigid docking system using a translational program. This program now incorporates the energetically appropriate conformers in to the binding pose resulting in a new complex which would represent the flexible docking phenomenon produced by MMCG system. Translation of each selected conformer is done on all the feasible configurations generated by ParDOCK system producing a large set of generated structures. Among this set the configurationally appropriate conformers top 5 complex structures are selected on the basis of translational energy calculated through translational program of MMCG system.

4.6 Energy minimization of docked structures and protein- ligand binding free energy estimations

The selected docked complexes are energy minimized in vacuum by AMBER (Pearlman, D.A *et al.,* 1995). For vacuum minimizations, 1000 steps of steepest descent and 1500 steps of conjugate gradient are carried out. The minimization procedure was repeated using explicit solvent, without much difference in the calculated energetics. Hence the vacuum minimization protocol was retained due to its expeditious nature. The energy minimized structure is employed in computing the binding affinity by a scoring function, BAPPL (Jain, T. *et al.,* 2005) developed in ScfBio IIT, Delhi. The energy function employed in BAPPL includes contributions of electrostatics, van der Waals, hydrophobicity and loss of conformational entropy of protein side chains upon ligand binding.

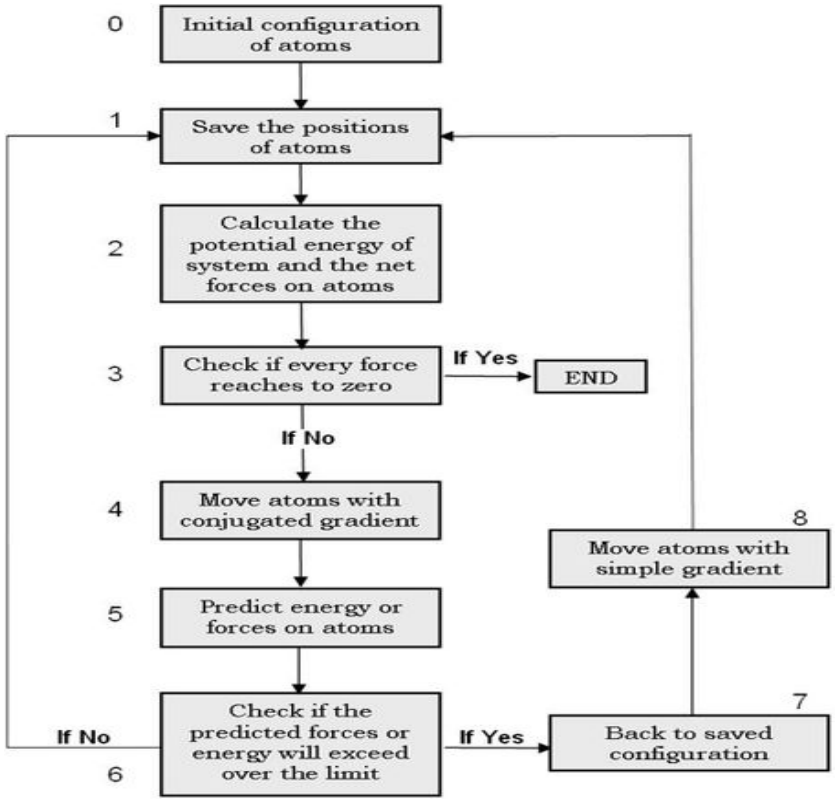


Fig.7: Methodology flow chart for energy minimization of structures

For the minimization of structure through AMBER following parameters are used as for in vacuo minimization.

Maxcyc — The maximum number of cycles of minimization. Default = 1 but for our program we used 1000

ncyc — If NTMIN is 1 then the method of minimization will be switched from steepest descent to conjugate gradient after NCYC cycles. Default 10. But for minimization of whole complex 250 cyc is used.

ntmin — Flag for the method of minimization.
= 0 Full conjugate gradient minimization. The first 4 cycles are steepest descent at the start of the run and after every nonbonded pairlist update.
= 1 For NCYC cycles the steepest descent method is used then conjugate Gradient is switched on (default).
= 2 Only the steepest descent method is used.

dx0 — The initial step length. If the initial step length is too big then will give a huge energy; however the minimizer is smart enough to adjust itself. Default 0.01.

drms — The convergence criterion for the energy gradient: minimization will halt when the root-mean-square of the Cartesian elements of the gradient is less than DRMS. Default 1.0E-4 kcal/mole-A

The Minimization and scoring function produces five flexibly docked complexes as final output of the MMCG system.

5. RESULTS

The MMCG program runs on a linux based server using shell scripts. A single job requires 18 processors to complete the docking process in nearly 30 min. This program is validated on 122 protein ligand complexes of ParDOCK database through self docking that is, the ligands were separated from the complexes and each ligand was randomly displaced in space and docked to respective target protein. The accuracy of the results is based on the Root Mean Square Deviations of docked structure as compared with that of experimental or native bound form. It is observed that the flexible ligand docking is much effective approach in attaining protein-ligand structure close to its native state.

Table 1. RMSD Analysis using self docking for 122 Protein-Ligand Complexes database

S.No.	PDB ID	NAME	STARTING RMSD	MININMUM RMSD
1	1a4q	Neuraminidase	2.59848	1.44699
2	1a4w	Alpha Thrombin	2.87077	1.5708
3	1a9m	HIV-1 Protease	4.21191	1.9601
4	1aco	Aconitase	0.6221	0.39592
5	1ae8	Alpha Thrombin	2.79614	1.258
6	1ajv	HIV-1 Protease	3.88118	1.99834
7	1ajx	HIV-1 Protease	3.81965	2.11699
8	1apb	L-arabinose - binding protein	0.90965	0.90926
9	1apt	Penicillopepsin	1.86758	1.56995
10	1apu	Penicillopepsin	1.82543	1.31082
11	1apv	Penicillopepsin	2.06575	1.49762
12	1apw	Penicillopepsin	3.48435	1.46984
13	1b5g	Thrombin	2.07857	1.57084
14	1b6j	HIV-1 Protease	3.50297	1.84161
15	1b6k	HIV-1 Protease	3.25006	1.95692
16	1b6l	HIV-1 Protease	3.75889	1.7971
17	1b9s	Neuraminidase	1.75555	0.36924
18	1b9t	Neuraminidase	1.65375	1.42111
19	1b9v	Neuraminidase	1.78236	1.4746
20	1bb0	Thrombin	2.74842	1.92217
21	1bdr	HIV-1 Protease	2.7108	1.28931
22	1bil	Renin	3.9604	1.60684
23	1bim	Renin	3.67655	1.69247
24	1bmm	Alpha Thrombin	2.63563	1.73552
25	1bmn	Alpha Thrombin	3.47591	1.85208

26	1bv7	HIV-1 Protease	4.70704	1.92514
27	1bv9	HIV-1 Protease	6.25811	1.73738
28	1c83	Tyrosine Phosphatase 1B	0.78905	0.73899
29	1c85	Tyrosine Phosphatase	0.8448	0.79548
30	1c86	Tyrosine Phosphatase	0.3003	0.13832
31	1c87	Tyrosine Phosphatase 1B	0.29332	0.14272
32	1c88	Tyrosine Phosphatase 1B	0.76418	0.87621
33	1c8k	Glycogen Phosphorylase	0.94095	0.78528
34	1cbs	Cellular Retinoic Acid Binding Protein Type II	1.44878	0.72715
35	1cdg	Cyclodextrin glycosyltransferase	2.23502	1.52641
36	1cf8	Catalytic Antibody 19A4	2.55738	0.99419
37	1com	Chorismate Mutase	1.613221	0.46924
38	1cpi	HIV-1 Protease	3.76428	2.04865
39	1cqp	Antigen CD11A	1.83909	1.40799
40	1ctr	Calmodulin	2.00312	0.86742
41	1cvu	Cyclooxygenase 2	2.75677	1.19704
42	1d3h	Dihydroorate Dehydrogenase	1.49874	0.90234
43	1d3t	Alpha thrombin	2.05184	1.28922
44	1d4l	HIV-1 Protease	2.51814	1.85201
45	1d4p	Alpha Thrombin	2.00322	0.2703
46	1dg5	Dihydrofolate Reductase	0.70731	0.48963
47	1dmp	HIV-1 Protease	3.02656	2.11756
48	1dog	Glycoamylase 471	1.00532	0.99242
49	1dr1	Dihydrofolate Reductase	2.08689	1.82419
50	1dwd	Alpha Thrombin	2.76754	1.59695
51	1ezq	Human coagulation factor XA	2.51601	0.88744
52	1f0t	Human coagulation factor XA	2.70136	1.50394
53	1f0u	Trypsin	2.38178	1.16888
54	1fax	Human coagulation	3.3962	1.80829

		factor XA		
55	1fkg	FK506 Binding Protein	2.57155	1.08228
56	1flr	FAB Fragment	0.6551	0.65523
57	1g2k	HIV-1 Protease	2.77248	1.46076
58	1gno	HIV-1 Protease	3.22546	1.46857
59	1hbv	HIV-1 Protease	2.20332	1.9073
60	1hdc	3 - alpha, 20 - beta - hydroxysteroid dehydrogenase	2.97714	1.9536
61	1hdt	Alpha Thrombin	3.71053	1.7184
62	1hew	Lysozyme	2.96518	1.76728
63	1hgi	Hemagglutinin	2.330047	1.75412
64	1hgj	Hemagglutinin	1.65603	0.49874
65	1hih	HIV-1 Protease	2.30579	1.575
66	1hii	HIV-2 Protease	4.0749	1.27079
67	1hiv	HIV-1 Protease	3.13621	1.85379
68	1hos	HIV-1 Protease	5.5084	1.9339
69	1hps	HIV-1 Protease	3.4219	1.76779
70	1hpv	HIV-1 Protease	2.92437	1.54537
71	1hpx	HIV-1 Protease	3.11089	1.6971
72	1hri	Human Rhinovirus	2.14947	0.71249
73	1hrn	Renin	2.75959	1.414332
74	1hsg	HIV-1 Protease	3.60659	1.35774
75	1hsh	HIV-1 Protease	2.27521	1.63198
76	1hte	HIV-1 Protease	2.07956	1.39829
77	1htf	HIV-1 Protease	3.42037	1.50573
78	1htg	HIV-1 Protease	3.26785	1.8087
79	1hvi	HIV-1 Protease	4.09128	2.17234
80	1hvj	HIV-1 Protease	2.932205	1.60322
81	1hvk	HIV-1 Protease	5.02845	2.25647
82	1hvr	HIV-1 Protease	4.29061	3.59035
83	1hxb	HIV-1 Protease	2.92984	1.72713
84	1hxw	HIV-1 Protease	3.50553	1.85734
85	1ida	HIV-2 Protease	3.07585	2.04092
86	1ivf	Neuraminidase	1.90798	1.51519
87	1k1n	Trypsin	2.59704	1.34185
88	1lyb	Cathepsin D	3.36363	1.88526
89	1mcf	Immunoglobulin	4.90002	1.93893
90	1mch	Immunoglobulin	4.06081	2.26026
91	1mcj	Immunoglobulin	2.61608	1.17776

92	1mrk	Alpha trichosanthin	1.6604	1.35031
93	1mtw	Trypsin	3.06695	1.52131
94	1nnb	Neuraminidase	1.61909	1.41591
95	1nsc	Neuraminidase	1.46101	1.27684
96	1pgp	6-Phosphogluconate Dehydrogenase	1.84061	0.97982
97	1pph	p-Hydroxybenzoate Hydroxylase	2.7861	1.82474
98	1qbt	HIV-1 Protease	4.95517	1.79158
99	1qbu	HIV-1 Protease	4.18488	2.13944
100	1rbp	Retinol Binding Protein	0.92851	0.57349
101	1rne	Renin	4.29109	1.91633
102	1sre	Streptavidin	2.1619	2.03844
103	1tlc	Thymidylate Synthase	2.46638	1.34263
104	1tng	Trypsin	0.58897	0.57729
105	1tnh	Trypsin	0.08629	0.02258
106	1tni	Trypsin	0.90309	0.07175
107	1tnj	Trypsin	0.45433	0.03362
108	1tnk	Trypsin	1.31947	0.17191
109	1tnl	Trypsin	1.10548	0.69652
110	1uvs	Alpha Thrombin	2.5784	1.43256
111	2cgr	IGG 2B Fab Fragment	2.25109	0.89903
112	2cmd	Malate Dehydrogenase	1.42855	0.59001
113	2gbp	D - Galactose D . Glucose binding protein	1.19223	1.14721
114	2ifb	Intestinal Fatty Acid Binding Protein	2.9584	1.04392
115	2pk4	Human plasminogen	1.63437	0.41044
116	2r04	Rhinovirus 14	3.52826	1.13534
117	2sim	Sialidase	1.71567	1.42798
118	2upj	HIV-1 Protease	2.96093	1.54408
119	2wea	Penicillopepsin	3.49859	1.89102
120	2wec	Penicillopepsin	3.84107	1.62137
121	4er2	Endothiapepsin	3.5685	1.98213
122	4est	Elastase	2.9022	1.25102

It was observed that after random displacement of ligand in space for self docking the starting RMSD really deviates from the bound form RMSD. After flexible ligand docking using MMCG

protocol the RMSD of the docked structure changes evidently and the resultant complexes have value really close to that of bound form.

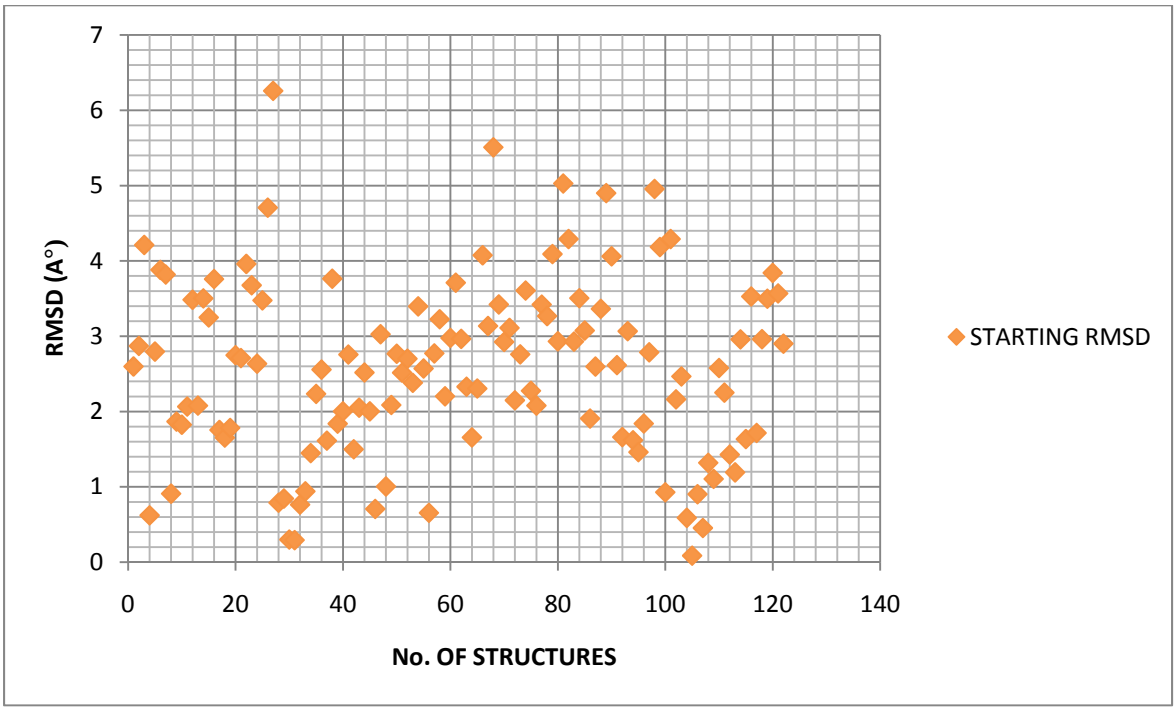


Fig.8: The initial RMSD of the dataset structures after random displacement as compared to bound form

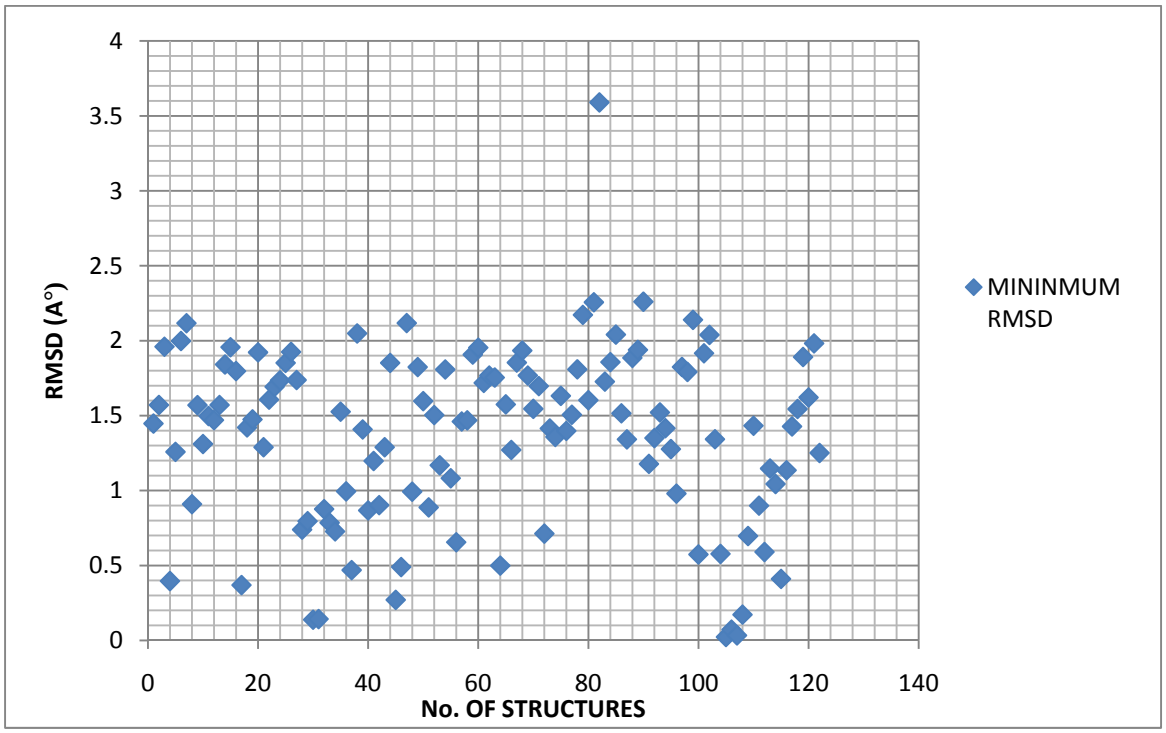


Fig.9: Resultant RMSD of the dataset structures after using MMCG protocol

The validation of program shows that out of 122 complexes above 80 percent structures were able to attain RMSD > 1.5 Å. The protocol shows that there is a correlation of r = 0.72 in RMSD between the docked structure and native bound form.

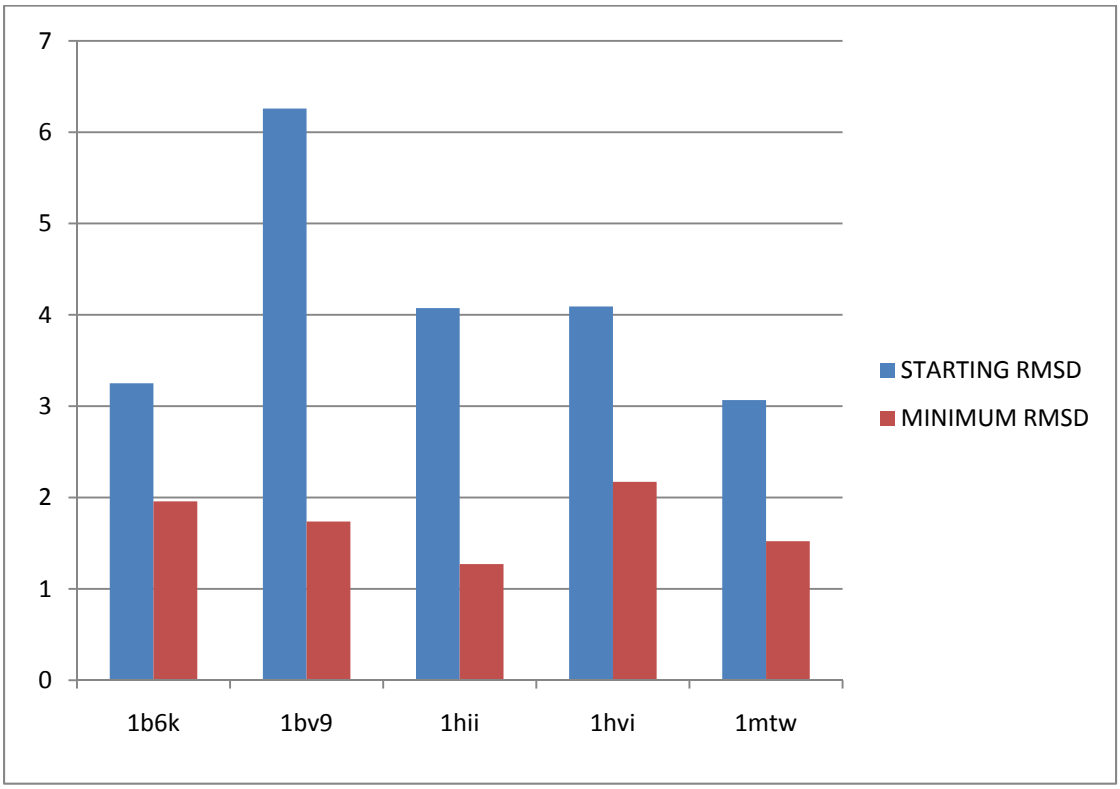


Fig. 10: Staring and minimum RMSD comparison of the structures as calculated with reference from bound form.

5.1 Flexible bonds recognition

The MMCG system taking protein-complex target and ligand as input first recognises all the flexible bonds present in the ligand structure. For the recognition of flexible bonds MMCG system reads the PDB format file of ligand and produces connect information of the ligand elements and also defines the bond types within the structure.

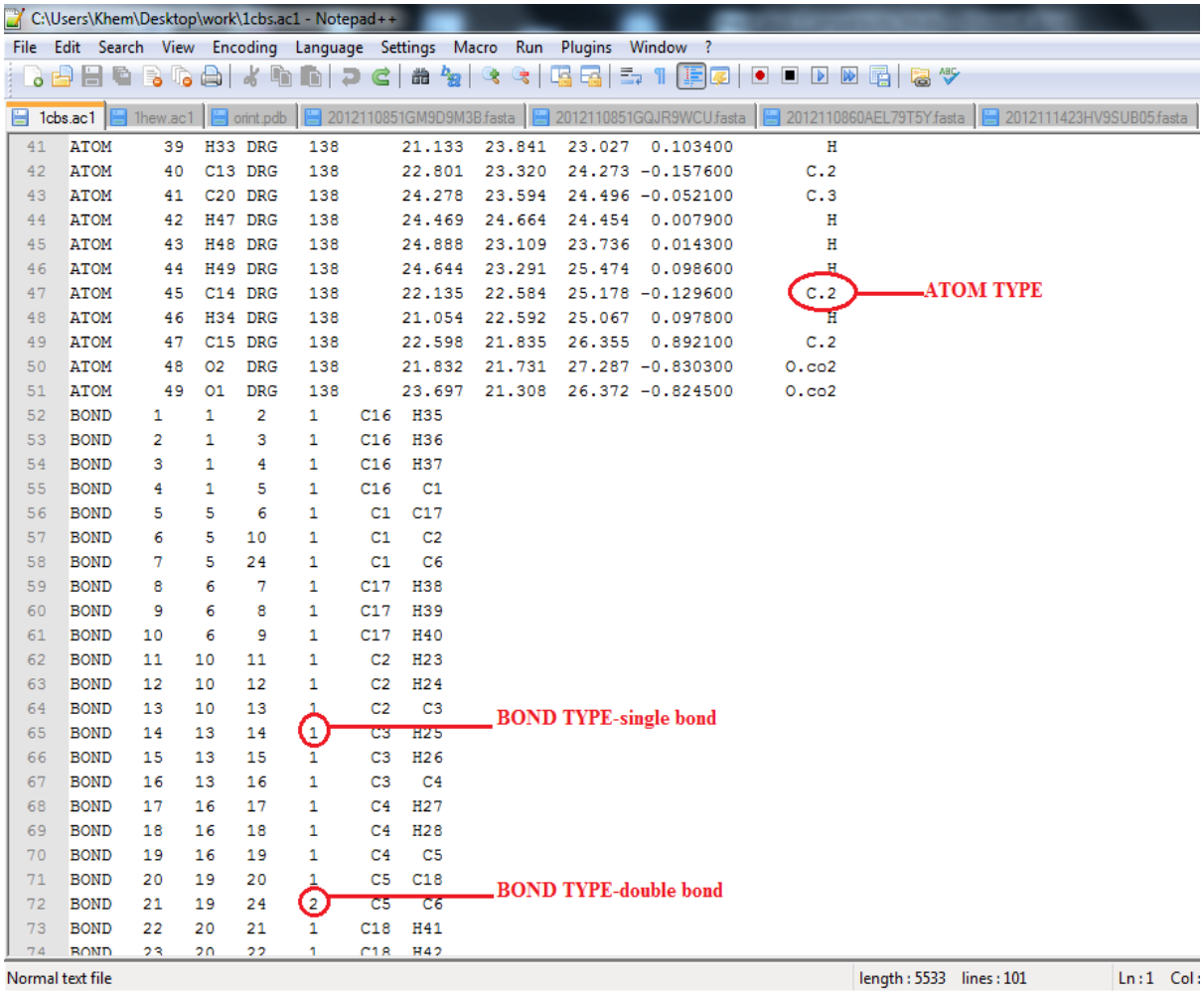


Fig. 11: Connect file generated using AM1 BCC force field parameters

This connect information is generated using AM1 BCC force field parameters. Using this information and bond type of the system, the number of flexible bonds i.e single bonded atoms within the structure are listed in a separate file. And also describe the candidate atoms with their atom types involved in the flexible bond.

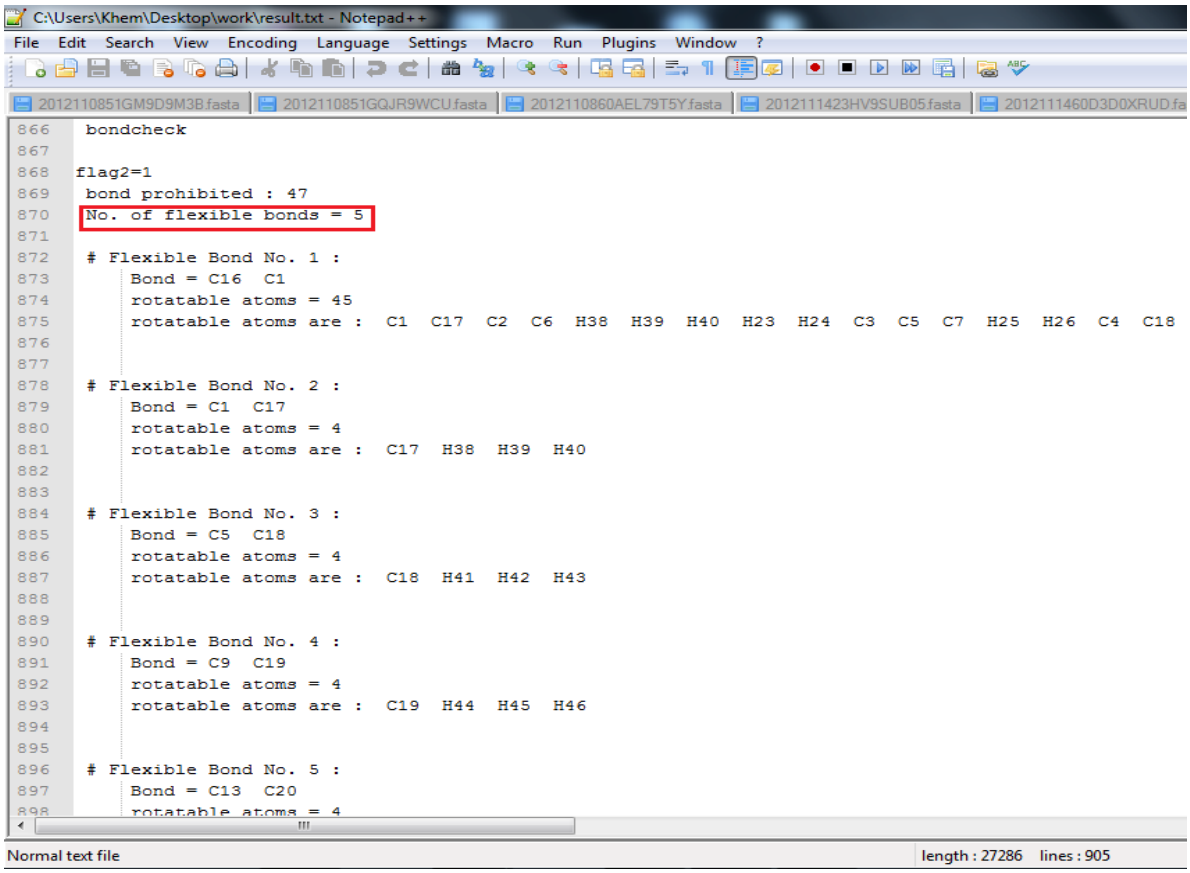


Fig.12: Flexible bond recognition result file containing information about the number of flexible bonds with their candidate atom types.

The Flexible bond recognition step is very important as picking of non-rotatable bond as rotatable would result in complete distortion of the ligand structure during Monte Carlo simulations.

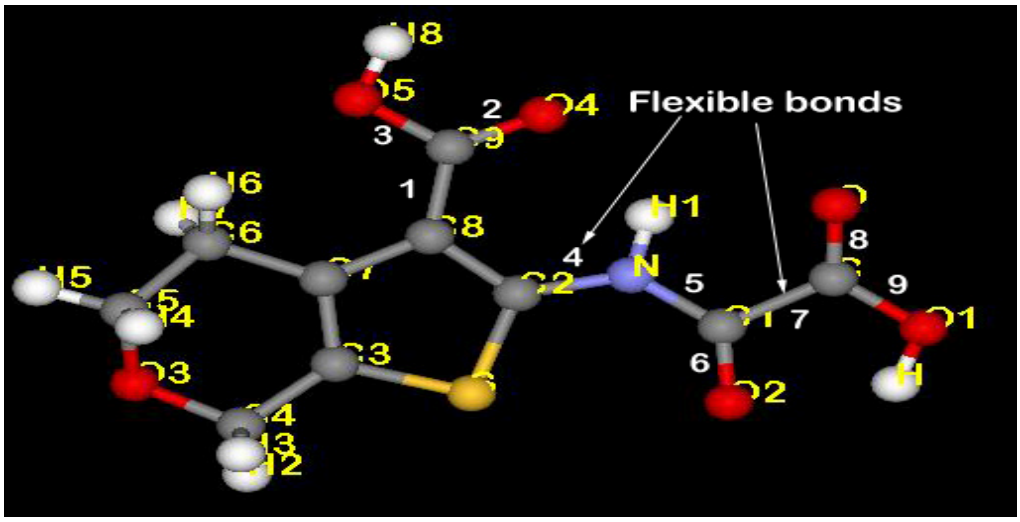


Fig.13 : Flexible bond recognition in 1c86 (Tyrosine phosphatase). The numbers of flexible bonds recognised over here are 9.

5.2 Flexible conformer generation

The input ligand after recognition of flexible bonds is subjected to monte-carlo simulations to generate a large set of decoys among which energetically most suitable conformations are selected. MMCG system picks up a random flexible bond reading information from the bond recognition output file. For the selected bond the dihedral angle (θ) is changed to a random value between -30° to +30°. The freedom of rotation is restricted to such a range as keeping in consideration that within a biological system also a very large change in the dihedral angles of ligand doesn't take place. The rotation program of MMCG on generating random value of θ rotates randomly picked flexible bond and subsequent connected flexible bonds to that value resulting in a new generated conformation.

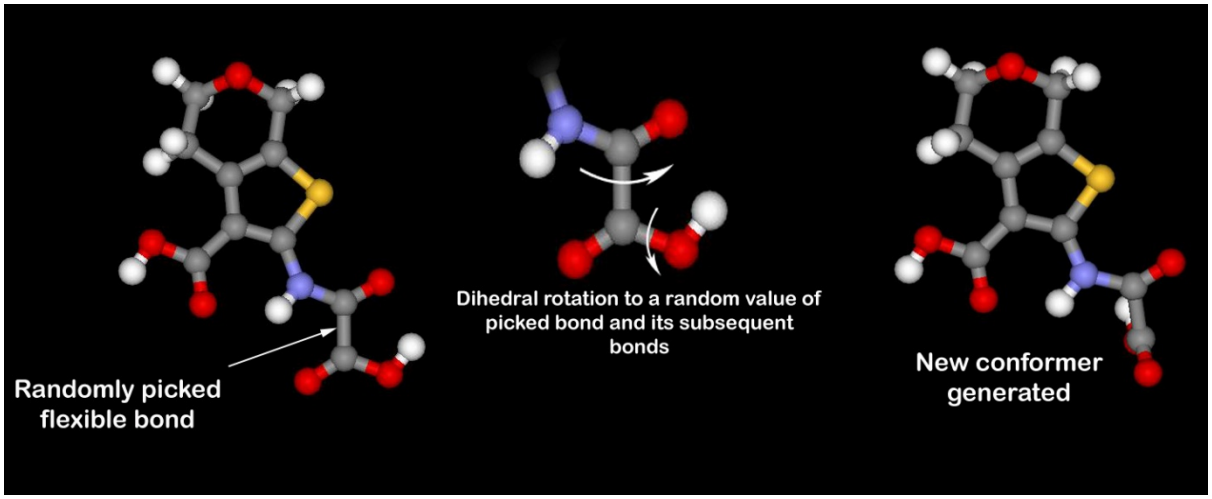


Fig. 14 : Monte-carlo simulation of ligand for the generation of low energy conformers.

The MMCG system runs these Monte Carlo steps repeatedly to generate 100 conformers for input ligands. Each conformer generated is first checked by metropolis criterion for the local minima. According to energy based evaluation calculated through AM1 BCC force fields parameters the conformers are selected or rejected.

$$P_i = \exp[Energy_2\text{-}Energy_1/R*T]$$

Where P_i is the probability of finding the system at the microstate (i). $Energy_1$ is the energy calculated of previous conformer generated and $Energy_2$ is the energy of new generated conformer. R is standard gas constant 1.98 at temperature T which is 298 k. A random number Z is generated between 0 to 1.

- If Pi < Z then in this case the new conformation is rejected and the previous structure again goes in to Monte Carlo step to generate new conformer.
- If Pi > Z then in this case the new conformation is accepted and the Monte Carlo steps proceeds on this structure for further generation of conformers.

-

5.3 Configuration generation and translation of conformers

After generation of the conformers through montecarlo metropolis approach energetically least value conformers are selected and considered to b closest to native state of the ligand docked in protein. To search the conformational space ParDOCK system is used. The ParDOCK performs an all atom based docking and scoring of the ligand in different configurations.

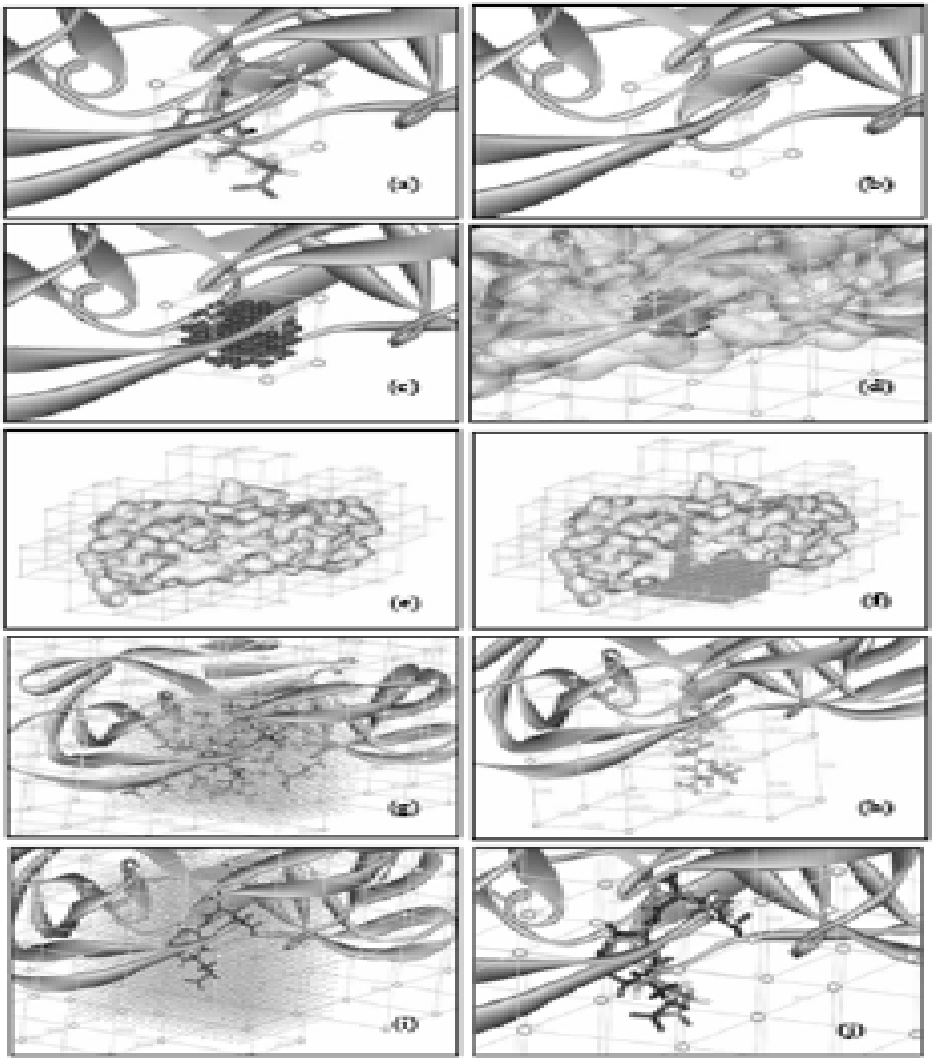
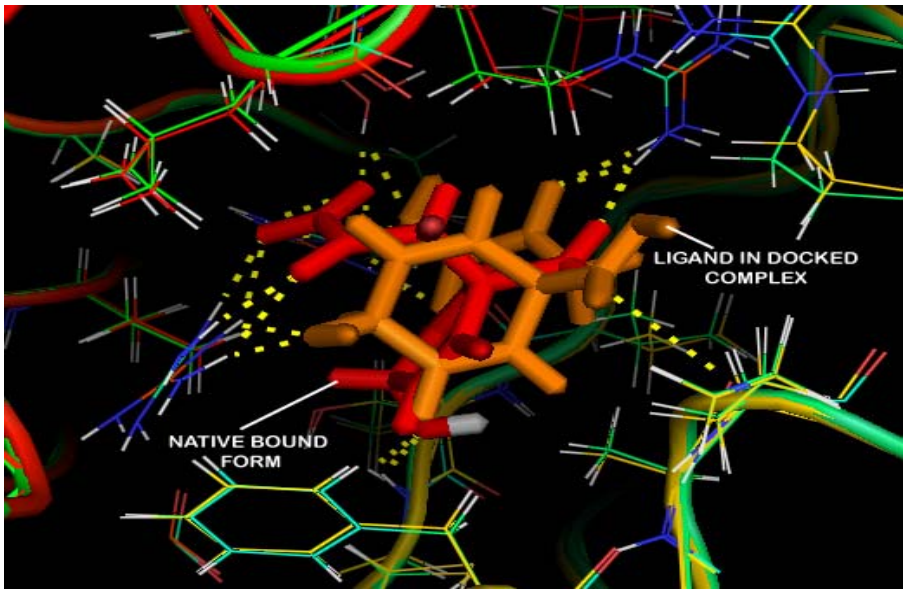
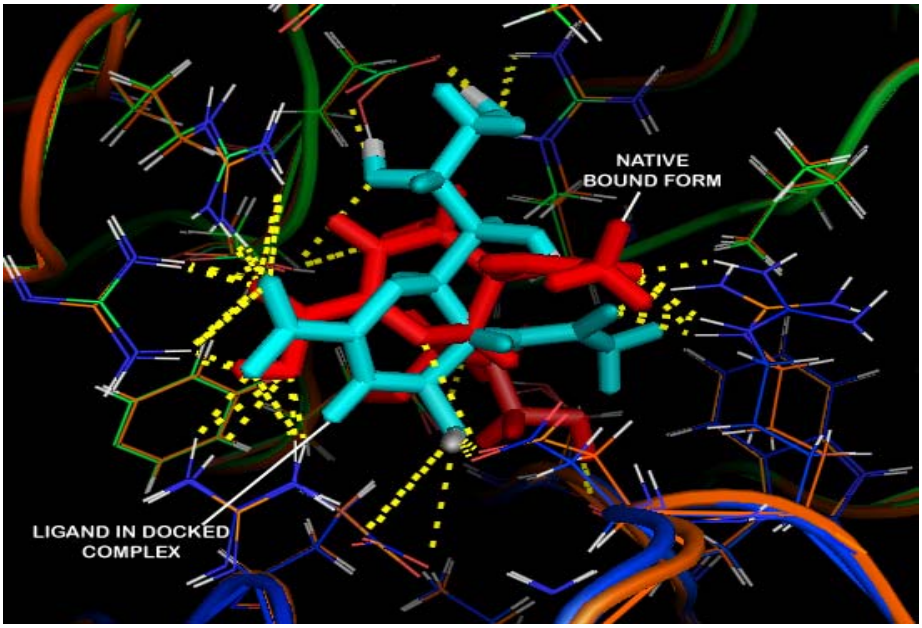


Fig. 15 :- Based on selected grid points energy grid is prepared at energy of each ligand atom at each energy point is pre-calculated. Generation of Monte Carlo configurations considering 6 degrees of freedom and best energy structure is selected.

ParDOCK produces fine least energy configurations as output with the ligand docked to respective poses within the complex. MMCG uses information from PDB file of the complex and causes translation of the least energy conformers produced from monte-carlo simulations to the binding pose of ligand in it. This translation step results in a generation of new docked complex which would consist of low energy conformer in to the apt binding pose.



A



B

Fig.16:- A: 1Com (Chosrimate mutase) B: 1ivf (Neuraminidase) alignment of the MMCG translational resultant complex with the structures bound in native form.

After the translation of conformers on to the binding pose the ligand gets flexibly docked at the active of the protein to mimic the native state of bound form. But these resultant complexes still are diverging from the native state to some extent. Hence, to refine the structures the energy calculation and minimization is done using AMBER module. The final scoring of the flexibly docked complex takes place using BAPPL scoring function which produces 5 flexibly docked complexes with their energy values and ranking.

5.4 Minimization

The minimization process reduces the energy conflicts in the structure and through validation on such a large data set it is observed that it really helps the system to mimic the native bound form the complex. For example if we take case of Neuraminidase (1ivf) after the translation it was observed that the RMSD between the formed structure and the native bound form is 1.51519 Å though the RMSD value is less than 2 Å but it still deviate from the native structure as can be seen from Fig B. So minimization of the complex is done in vacuum using AMBER with max 1000 cycles.

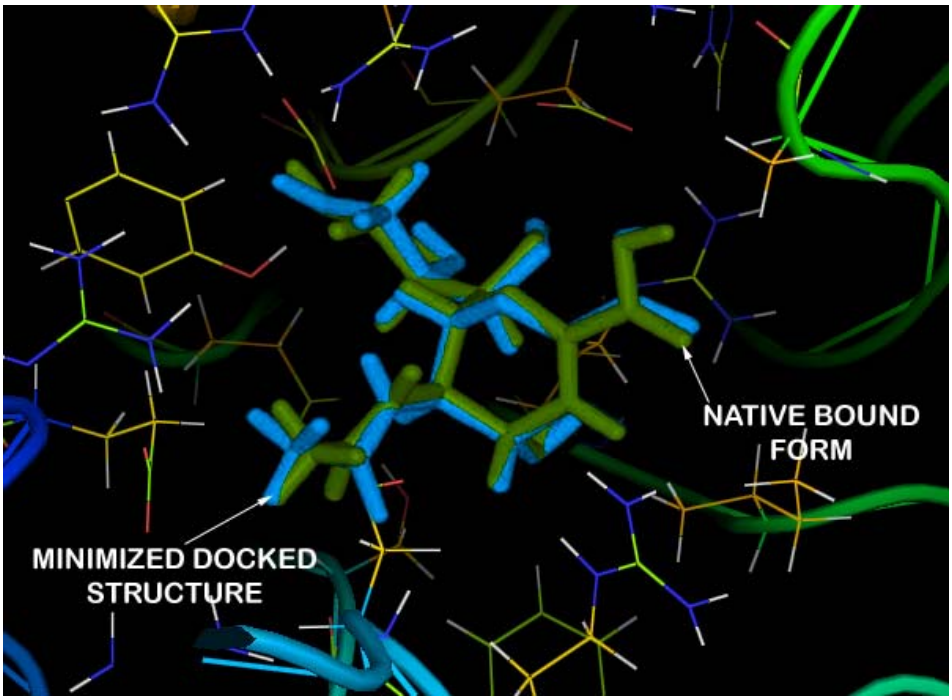


Fig. 17: Alignment of the minimized docked structure of Neuraminidase (1ivf) with the native bound form of the complex.

The minimization process calculates the energy of structure and accordingly minimizes it in step by step process. After minimization it was observed that the RMSD between docked structure and the native bound form gets reduced to 0.4285 Å that is even less than 0.5 Å which shows that the docked structure now resembles the native bound form of the ligand and protein.

6. CONCLUSION

The flexibility of ligands in semi-flexible docking has remained a biggest problem for bioinformatics scientists for the purpose of drug designing. As a solution to this problem here we have presented MMCG, A Monte Carlo metropolis approach based docking and scoring system. MMCG is observed to be computationally feasible and highly efficient as it is able to attain the native bound conformations with RMSD of less than 1.5 Å in nearly 80 percent of the cases. The MMCG system is validated on a highly diverse data set of 122 protein-ligand complexes. The success of this system is based on various factors like, pioneer step of flexible bond recognition which is very important for incorporation of ligand flexibility. This system uses monte-carlo simulations to generate large number of conformers and at each step the conformers are accepted on the basis of metropolis criterion. Metropolis criterion helps system to avoid getting strucked in local minima instead of global minima and conformers are selected on the basis of energy calculation. Configurational search is performed through an all atom energy based program which precalculates the energy of receptor to the binding energy by defining grid on whole protein which decreases the computational time. The selected conformers are translated to the resultant binding poses of configurational search, this translation results in complex formation and incorporate ligand flexibility in the docking procedure. Further refinement of the docked complexes is done using minimization process which helps the complex to attain native bound form to great extent. Fully automated version of MMCG runs on 18 processors and requires very less runtime which is about 30 min. to complete the whole docking process. Leads which are searched on the basis of rigid docking may have less RMSD that is around 2Å but these structures are biologically less efficient as the flexibility criterion is not considered. With the use of MMCG the potential ligands searched would be more appropriate as compared to rigid systems and it would really help in decreasing less relevant hit molecules for a protein which is an advantage from the point of view of wet lab testing also. This approach of flexible docking can act as a milestone in the field of computational drug designing as though this protocol we are able to mimic the actual docking of drugs in biological systems up to much extent using high performance computing environment.

7. DISCUSSION AND FUTURE PERSPECTIVE

Since the advent of computational drug designing molecular docking has remained its most important feature. The idea behind the use of molecular docking for lead invent was to mimic the natural system of interaction of drugs with protein. First type of molecular docking which is still considered effective one is rigid docking, where both the protein and ligand are considered to be rigid. The search space is very limited, considering only three translational and three rotational degrees of freedom. In this case, ligand flexibility could be addressed by using a pre-computed a set of ligand conformations, or by allowing for a degree of atom–atom overlap between the protein and ligand. Taking flexibility of both the ligand and molecule has been a mammoth task for the researchers and on the other hand computationally also its way difficult. So as to trade off with accuracy and limitations of computational power semi-flexible docking is introduced that provides flexibility to the ligands only but the protein molecule is kept rigid. With six degrees of translational and rotational freedom as well as the conformational degrees of freedom of both the ligand and protein, there are a huge number of possible binding modes between two molecules. Unfortunately, it would be too expensive to computationally generate all the possible conformations. Various sampling algorithms have been developed and widely used in molecular docking software which are matching algorithms, incremental constructions(IC), genetic algorithm(GA) but the most widely used in number of docking programs is monte-carlo simulations. The Monte Carlo approach used in case of MMCG is an improved algorithm which uses meteropolis criterion to keep a check on attaining local minima for the ligand structure instead of global minima. The combination of minimum energy conformation generation and all atom based configurational search is backbone of MMCG program and the results produced by this combination forms a docked complex which is really close to the native bound form.

The validation of program on such a large and diverse dataset is sufficient to test the efficiency of the MMCG system but it's still not able to capture the structures really close to the bound form. To improve the efficiency protein flexibility can also be implemented and this could be done through using flexibility of binding pockets only which would require less computational time as compared to other online flexible docking tools available

8. REFERENCES

Abagyan R, Totrov M, Kuznetsov D (1994). ICM-A new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. J. Comput. Chem. **15**:488–506.

Ajay and Murcko, M.A. (1995) J.Med. Chem., **38**, 4953 -4967.

Arazjakalian, Bruce,L. Bush, David B.J. and Christopher I.B (2000) *J. Comput. Chem., **21***,132 146.

Bailey D, Brown D (2001). High-throughput chemistry and structure-based design: survival of the smartest. Drug Discov Today; **6**(2):57–59.

Bajorath J (2002). Integration of virtual and high-throughput screening. Nat Rev Drug Discov.1(**11**):882–894.

BoÈhm, H. J. (1992a). The computer program LUDI: a new method for the de novo design of enzyme inhibitors. J. Comput.-aided Mol. Design, **6**, 61-78.

BoÈhm, H. J. (1992b). LUDI: Rule-based automatic design of new substituents for enzyme inhibitor leads. J. Comput-aided-Mol. Design, **6**, 593-606.

Brint AT, Willett P (1987). Algorithms for the Identification of Three-Dimensional Maximal Common Substructures. J. Chem. Inf. Comput. Sci. **27**:152–158.

Bron C, Kerbosch J. Algorithm (1973): Finding All Cliques of an Undirected Graph. Communications of the ACM. ; **16**(9):575–576.

Cornell,W.D. Cieplak,P. Bayly,C.I. Gould,I.R. and Merz, K.M.,*et al.*(1995) J. Am.Chem.Soc.,**117**, 5179 -5197.

Ewing TJ, Makino S, Skillman AG, Kuntz ID (2004) . DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. J Comput Aided Mol Des.; **15**(5):411– 428.

Fischer D, Norel R, Wolfson H, Nussinov R (1993). Surface motifs by a computer vision technique: searches, detection, and implications for protein-ligand recognition. Proteins. **16**(3):278–292.

Friedman,H.L and Krishnan, C.V**.** (1973) *J. Soln. Chem.*, ***2***, 119.

Goodsell DS, Lauble H, Stout CD, Olson AJ (1993). Automated docking in crystallography: analysis of the substrates of aconitase. Proteins. **17**(1):1–10.

Goodsell DS, Olson AJ (1990) . Automated docking of substrates to proteins by simulated annealing. Proteins. **8**(3):195–202.

Gupta A, Gandhimathi P, Sharma P, Jayaram B (2007): ParDOCK: An all atom energy based Monte Carlo docking protocol for protein-ligand complexes. Protein Pept Lett, **14**:632-46.

Hammes GG (2002). Multiple conformational changes in enzyme catalysis. Biochemistry ; **41**(26): 8221–8228.

Hart TN, Read RJ (1992). A multiple-start Monte Carlo docking method. Proteins. **13**(3):206–222.

I.M. Kapetanovic, (2006). Computer-Aided Drug Discovery and Development (CADDD): *in silico*-chemico-biological approach. Chem Biol Interact. **171**(2): 165–176.

Inbal, H. Buyong, M. Haim, W.and Ruth.N (2002) PROTEINS: Structure, Function, and Genetics, **47**,409–443.

Jain T, Jayaram B (2005): An all atom energy based computational protocol for predicting binding affinities of protein-ligand complexes. FEBS Letters, **579**:6659-6666.

Jain, T. and Jayaram, B. (2005), FEBS Letters, **579**, 6659-6666.

Jiang F, Kim SH (1991). "Soft docking": matching of molecular surface cubes. J Mol Biol.; **219**(1):79–102.

Jones G, Willett P, Glen RC, Leach AR, Taylor R (1997). Development and validation of a genetic algorithm for flexible docking. J Mol Biol. 267(3):727–748.

Jorgensen WL(2004). The many roles of computation in drug discovery. Science; **303**(5665): 1813–1818.

Kitchen DB, Decornez H, Furr JR, Bajorath J (2004). Docking and scoring in virtual screening for drug discovery: methods and applications. Nat Rev Drug Discov; 3(**11**):935 949.

Koshland DE Jr. (1963) Correlation of Structure and Function in Enzyme Action. Science ; **142**:1533–1541.

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (1982). A geometric approach to macromolecule-ligand interactions. J Mol Biol. ; **161**(2):269–288.

Kuntz ID, Blaney JM, Oatley SJ, Langridge R, Ferrin TE (2006). A geometric approach to macromolecule-ligand interactions. J Mol Biol.; **161**(2):269–288.

Kuntz, I. D. (1992). Structure-based strategies for drug design and discovery. Science, **257**, 1078-1082.

Lengauer, T. Rarey, M. (1996) Curr Opin Struct Biol., **6**,402-406.

Lewis, R. A. (1992). Automated site-directed drug design using molecular lattices. J. Mol. Graph. **10**, 66-78.

Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. (1953). "Equations of State Calculations by Fast Computing Machines". *Journal of Chemical Physics* 21 (6): 1087–1092.

Mirankar, A. & Karplus, M. (1995). An automated method for dynamic ligand design. Proteins: Struct. Funct. Genet. **23**, 472-490.

Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Br J Pharmacol. ; **153**(Suppl 1):S7–26.

Moitessier N, Englebienne P, Lee D, Lawandi J, Corbeil CR (2008). Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. Br J Pharmacol. **153**(Suppl 1):S7–26.

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998). Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry. **19**(14):1639–1662.

Morris GM, Goodsell DS, Halliday RS, Huey R, Hart WE, Belew RK, Olson AJ (1998) . Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. Journal of Computational Chemistry. **19**(14):1639–1662.

Oshiro CM, Kuntz ID, Dixon JS (1995). Flexible ligand docking using a genetic algorithm. J Comput Aided Mol Des. **9**(2):113–130.

Pearlman, D.A. Case, D.A. Caldwell, J.W. Ross, W.S. and Cheathem III, J.E. *et al.* (1995) *Comput. Phys Commun*., ***91***, 1- 41.

Ramanathan,P.S and Friedman, H.L. (1971) *J. Chem. Phys*., ***54*** ,1086.

Rarey M, Kramer B, Lengauer T, Klebe G (1996). A fast flexible docking method using an incremental construction algorithm. J Mol Biol. **261**(3):470–489.

Shoichet, BK.; McGovern, SL.; Wei, B.; Irwin, JJ (2002). Hits, leads and artifacts from virtual and high throughput screening. Molecular Informatics: Confronting Complexity.

Taylor JS, Burnett RM (2000). DARWIN: a program for docking flexible molecules. Proteins. **41**(2):173–191.

Teague SJ.(2003) Implications of protein flexibility for drug discovery. Nat Rev Drug Discov. ; **2**(7):527–541.

Verdonk ML, Cole JC, Hartshorn MJ, Murray CW, Taylor RD (2003). Improved protein ligand docking using GOLD. Proteins. **52**(4):609–623.

Wang,J. Wolf,R.M. Caldwell,J.W. Kollman,P.A.and Case, D.A.(2004) *J.Comput.Chem*., ***25***,1157-1174.

Xuan-Yu Meng, Hong-Xing Zhang, Mihaly Mezei, and Meng Cui,(2011). Molecular Docking: A powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des. ; **7**(2): 146–157.