

**IN SILICO ANALYSIS, MODELING, DOCKING AND PHARMACOPHORE STUDIES OF PROTEINS INVOLVED IN AUTO IMMUNE MALADIES (RHEUMATOID ARTHRITIS)**G. Prathima^{1*}, Dr. A. Ravindernath², Dr. P. Raja Rao³, P. Sahithi⁴¹*M.Tech Biotechnology Final year, University College of Technology, Osmania University, Hyderabad, India² Professor and Head, University College of Technology, Osmania University, Hyderabad, India³ Associate Professor, University College of Technology, Osmania University, Hyderabad, India⁴ Assistant Professor, University College of Technology, Osmania University, Hyderabad, India***Corresponding author e-mail:** gujjulaprathima@gmail.com**ABSTRACT**

Cytokine networks participate with paracrine and autocrine loops maintaining cellular activation in the synovial intimal lining. In rheumatoid arthritis inflammatory changes occur throughout the connective tissues of the body. The most useful medications in relieving the pain and disability of rheumatoid arthritis with anti-inflammatory properties are aspirin and ibuprofen. The present project was focused on study of anti-rheumatoid arthritis activity of bioactive compounds. In the present study, homology modeling, mutagenesis, docking studies were carried out with some of the selected bioactive compounds. PDB latest version was used to identify the target protein, obtain sequence from protein sequence data bank and homology modeling for the target protein was done using modeler 9.14 version and MOE 2008. version. Docking studies using molecular operating environment program revealed that Ellagic acid, Curcuminoid and Methyl gallate possess anti rheumatoid arthritis activity. Further, pharmacophore mapping studies were performed using DISCOVERY STUDIO on these compounds in order to identify the pharmacophoric feature responsible for the observed activity of the compounds.

Keywords: Autoimmune disease, DISCOVERY STUDIO, Modeller, Pharmacophore, Rheumatoid Arthritis**INTRODUCTION**

Autoimmune diseases result from a dysfunction of the immune system in which the body attacks its own organs, tissues, and cells. Physicians and scientists have identified more than 80 clinically distinct autoimmune diseases. Several are well known, including rheumatoid arthritis, multiple sclerosis, type1 diabetes, and systemic lupus erythematosus; others are less familiar, including autoimmune hepatitis, autoimmune ear disease, and pemphigus. The social and financial burdens imposed by these chronic, debilitating diseases include poor quality of life, high healthcare costs, and substantial loss of productivity. Auto immune diseases are broadly classified into hemolytic autoimmune diseases,

localized autoimmune diseases and systemic autoimmune diseases. Hemolytic autoimmune disease is a clinical disorder due to destructions of blood components where in auto Ab are formed against one's own RBCs, Platelets or Leucocytes. Examples include Hemolytic anaemia, Leucopenia, Thrombocytopenia, etc. Localized autoimmune diseases are also called as Organ specific autoimmune diseases where a particular organ is affected due to auto Abs. For example, Thyroiditis, Multiple sclerosis, Type I DM. Systemic autoimmune diseases are also called as Non-organ specific autoimmune diseases where immune complexes accumulate in many tissues and cause inflammation and damage. Examples include, Systemic Lupus

Erythematosus (anti-nuclear Ab.), Rheumatoid Arthritis (anti-IgG antibodies).

Rheumatoid arthritis (RA), a chronic autoimmune inflammatory disorder of unknown etiology occurs in approximately 1% of the population [Albani and Carson 1997] where in it is more prevalent among women than men, and usually develops in the fourth and fifth decades of life with the primary presenting symptoms being pain, stiffness and swelling of the joints, accompanied by constitutional symptoms such as fever and malaise (Grassi et al. 1998)^[1,2]. Synovial inflammation underlies the cardinal manifestations of this disease, which include pain, swelling, and tenderness followed by cartilage destruction, bone erosion, and subsequent joint deformities. Despite intensive research, the precise cause of RA remains elusive.

Cytokine networks participate with paracrine and autocrine loops that maintain cellular activation in the synovial intimal lining. When cells are recruited to the synovium, or even perhaps at extrasynovial sites, antigens can be processed by dendritic cells (DCs). They can present antigen in synovial germinal centers or, more likely, migrate to central lymphoid organs, where they can activate naive T cells through interactions with the T cell receptor (TCR) and costimulatory signals^[3]. In rheumatoid arthritis inflammatory changes occur throughout the connective tissues of the body. The most useful medications in relieving the pain and disability of rheumatoid arthritis are aspirin and ibuprofen, which have anti-inflammatory properties as well and the most useful drugs which slow down the disease progression are Rheumatrex and Trexall.

Bioinformatics is an interdisciplinary field that combines computer science, statistics, mathematics, biological sciences and engineering to study and process biological data and develops methods and software tools for understanding biological data. It involves the development and application of data-analytical and theoretical methods, mathematical modeling and computational simulation techniques to the study of biological, behavioral, and social systems. Computational biology is different from biological computation, which is a subfield of computer science and computer engineering using bioengineering and biology to build computers, but is similar to bioinformatics^[4]. Computational Biology, sometimes referred to as bioinformatics, is the science of using biological data to develop algorithms and relations among various biological systems. The tools and techniques employed in the present study are homology modelling, docking studies and pharmacophore studies.

MATERIALS AND METHODS

Modeller:

Modeller 9.14 version is used in project for homology modeling of target protein. Modeller is a computer program used in producing homology models of protein tertiary structures as well as quaternary structures (rarer). It implements a technique inspired by nuclear magnetic resonance known as satisfaction of spatial restraints, by which a set of geometrical criteria are used to create a probability density function for the location of each atom in the protein. The method relies on an input sequence alignment between the target amino acid sequence to be modeled and a template protein whose structure has been solved. The program also incorporates limited functionality for ab initio structure prediction of loop regions of proteins, which are often highly variable even among homologous proteins and therefore difficult to predict by homology modeling. The ModWeb comparative protein structure modeling web server is based on modeller and other tools for automatic protein structure modeling, with an option to deposit the resulting models into ModBase^[5]. A new version of EasyModeller (EasyModeller 4.0) has been recently released and available for free download for Linux and Windows OS. There is also a simple interface to modeller in the program UCSF Chimera.

Discovery studio:

DISCOVERY STUDIO is a suite of software for simulating small molecule and macromolecule systems. It is developed and distributed by ACCELRY'S. It helps to:

- Investigate and test hypotheses in silico prior to costly experimental implementation, thus reducing the time and expense involved in bringing products to market.
- Drive scientific exploration from target identification to lead optimization with a wealth of trusted life science modeling and simulation tools.
- Leverage an open and scalable platform to automate processes, create and deploy custom workflows, and integrate data types, databases, and third-party or in-house tools.
- Enhance personal productivity and boost team collaboration by enabling researchers to share data and make better informed decisions.

The Molecular Operating Environment:

The Molecular Operating Environment, is a comprehensive software system for Life and Material Science developed by Chemical Computing Group

Inc. (CCG). MOE is a combined Applications Environment and Methodology Development Platform that integrates visualization, simulate ion and application development in one package. MOE strongly supports drug design through molecular simulation, protein structure analysis, data processing of small molecules, docking study of proteins and small molecules, and so on under the unified operations. SVL, the Scientific Vector Language, is the portable high performance programming language built in MOE. SVL is the vectrized command language, scripting language and applications programming language. One tenth reductions in code size over C and FORTRAN are routinely realized. SVL source code to MOE applications is distributed with MOE to end Users; this allows MOE applications to be freely customized or modified.

Swiss-PdbViewer:

Swiss-PdbViewer is an application that provides a user friendly interface allowing to analyze several proteins at the same time. The proteins can be superimposed in order to deduce structural alignments and compare their active sites or any other relevant parts. Amino acid mutations, H-bonds, angles and distances between atoms are easy to obtain because of the advances in the latest PDB version. Moreover, Swiss-PdbViewer is tightly linked to Swiss-Model, an automated homology modeling server developed within the Swiss Institute of Bioinformatics (SIB) in collaboration between GlaxoSmithKline R&D and the Structural Bioinformatics Group at the Biozentrum in Basel. Working with these two programs greatly reduces the amount of work necessary to generate models, as it is possible to thread a protein primary sequence onto a 3D template and get an immediate feedback of how well the threaded protein will be accepted by thereference structure before submitting a request to build missing loops and refine sidechain packing. Swiss-PdbViewer can also read electron density maps, and provides various tools to build into the density. In addition, various modeling tools are integrated and command files for popular energy minimization packages can be generated[spdbv.vital-it.ch/]

Homology modeling:

The term "homology modeling", also called comparative modeling or template based modeling (TBM), refers to modeling a protein 3D structure using a known experimentally determined structure of a homologous protein as a template [6]. A protein structure is always of great assistance in the study of protein function, dynamics, interactions with ligands

and other proteins, and even within pharmaceutical industry in structure-based drug discovery and drug design [7]. Homology modeling can provide the molecular biologists and biochemists with "low-resolution" structures, which will contain sufficient information about the spatial arrangement of important residues in the protein and which may guide the design of new experiments. For example, the design of site-directed mutagenesis experiments could be considerably improved if such "low-resolution" model structures could be used. Experimental elucidation of a protein structure may often be delayed by difficulties in obtaining a sufficient amount of protein (cloning, expression and purification of milligram quantities), by difficulties associated with crystallization, and even the protein crystallographic part may become a source of problems. In this context, it is not surprising that methods dealing with the prediction of protein structure have gained much interest. Among these methods, the method of homology modeling usually provides the most reliable result. The use of this method is based on the observation that two proteins belonging to the same family and sharing similar amino acid sequences will have similar three-dimensional structures. By other words, the problem is reduced to finding a modeling template. The steps in homology modeling are the following:

- Template identification;
- Amino acid sequence alignment;
- Alignment correction;
- Backbone generation;
- Generation of loops;
- Side chain generation & optimization;
- *Ab initio* loop building;
- Overall model optimization;
- Model verification. Quality criteria, model quality;

After finding a template it is an absolute requirement that before starting the modeling project, you make a multiple sequence alignment, which should include your sequence, the sequence of the template and some other sequences of proteins belonging to the same family. This will give you an overview of the general features of the protein family, the degree of conservation, the consensus sequence motifs, etc. It would also be very desirable to make a secondary structure prediction, discussed in the tutorial on sequence alignment. Most importantly, the positions of insertions and deletions should be correct, likewise the conservation of important residues, for example active site residues. When the sequence analysis is done and the alignment is corrected accordingly, we may proceed to the modeling. The modeling software will thread your sequence on the template structure,

thus creating a preliminary model of you protein (backbone generation). After that it will try to build missing parts, generate side chains for replaced residues and optimize side chain conformations, etc. At the last step the overall model needs to be optimized followed by verification of model quality. There are several servers which may be used for modelling. Here modeler 9.14 version is used.

Sequence Retrieval and Template Selection:

The protein sequence of IL-6R was retrieved from SWISSPROT database with known structures available in the PDB, NCBI BLAST (www.ncbi.nlm.nih.gov/blast/) tool was used to identify the template for homology modeling.

Sequence Alignment and Model Building:

The template sequence was extracted and aligned with the target sequence using Clustal W. "MODELLER 9.14v" was employed in generation of the 3D structure of the protein^[8]. CHARMM force fields and steepest descent algorithm followed by conjugant gradient algorithm were employed for energy minimization of the modeled structure until the convergence gradient is satisfied.

Mutant Construction:

Mutations are created in the modeled structure of IL-6R protein and mutant models MU-,MU-2,MU-3,MU-4, MU-5 and MU-6 were constructed. Energy minimization was carried out by applying CHARMM force fields in DS.

Model Validation:

i. RAMPAGE Server: Ramachandran Plot Analysis: A Ramachandran plot (also known as a Ramachandran diagram or a $[\phi, \psi]$ plot), originally developed in 1963 by G. N. Ramachandran, C. Ramakrishnan, and V. Sasisekharan, is a way to visualize backbone dihedral angles ψ against ϕ of amino acid residues in protein structure. The figure at left illustrates the definition of the ϕ and ψ backbone dihedral angles (called ϕ and ϕ' by Ramachandran). The ω angle at the peptide bond is normally 180° , since the partial-double-bond character keeps the peptide planar. The figure at top right shows the allowed ϕ, ψ backbone conformational regions from the Ramachandran et al. 1963 and 1968 hard-sphere calculations: full radius in solid outline, reduced radius in dashed, and relaxed tau (N-Calpha-C) angle in dotted lines. Because dihedral angle values are circular and 0° is the same as 360° , the edges of the Ramachandran plot "wrap" right-to-left and bottom-to-top. For instance, the small strip of allowed values along the lower-left edge of the plot are a

continuation of the large, extended-chain region at upper left.

ii. Prosa: Prosa compares Z scores between target and template structure. The Z score obtained for the modelled structure should be comparable to the Z score obtained for the template.

iii. RMSD: Root Mean Squared Deviation (RMSD) is used to represent the distance between two objects^[9]. Low RMSD value indicates more similarity between the structures. SPDBV program was used in calculating the RMSD value between the modelled structure and the template.

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 using, for example, scoring functions. The associations between biologically relevant molecules such as proteins, nucleic acids, carbohydrates, and lipids play a central role in signal transduction. Furthermore, the relative orientation of the two interacting partners may affect the type of signal produced (e.g., agonism vs antagonism). Therefore docking is useful for predicting both the strength and type of signal produced. Docking is frequently used to predict the binding orientation of small molecule drug candidates to their protein targets in order to in turn predict the affinity and activity of the small molecule. Hence docking plays an important role in the rational design of drugs. Given the biological and pharmaceutical significance of molecular docking, considerable efforts have been directed towards improving the methods used to predict docking.

Ligand Generation and Optimization:

ACD-Chemsketch was used in drawing the structure of the ligands taken for binding analysis. Catalyst algorithm in DS was used in ligand preparation with constraint parameters such as tautomer and isomer generation, removal of all the duplicate structures and generation of the 3D structure.

Calculating the active site sequence:

Active sites present in the protein were identified from the 3D atomic coordinates of the receptor using Q-SITE FINDER. It is an energy-based method for the prediction of protein-ligand binding sites^[10].

Docking:

The software package MOE program (Molecular Operating Environment) designed by the Chemical Computing Group was used in docking and analyzing

the binding of the ligand molecules with the protein molecule. The parameters used for the Docking were, Total Runs = 50, Cycle/Runs = 15, Iteration Limit=10 000, Potential Energy Grid: ON, Annealing Algorithm: Simulated Annealing.

Pharmacophore mapping:

A pharmacophore is an abstract description of molecular features which are necessary for molecular recognition of a ligand by a biological macromolecule ^[11].

Model development:

The process for developing a pharmacophore model generally involves the following steps:

1. Select training set of ligands– Choose a structurally diverse set of molecules that will be used for developing the pharmacophore model. As a pharmacophore model should be able to discriminate between molecules with and without bioactivity, the set of molecules should include both active and inactive compounds.
2. Conformational analysis– Generate a set of low energy conformations that is likely to contain the bioactive conformation for each of the selected molecules.
3. Molecular superimposition– Superimpose ("fit") all combinations of the low energy conformations of the molecules. Similar (bioisosteric) functional groups common to all molecules in the set might be fitted (e.g., phenyl rings or carboxylic acid groups). The set of conformations (one conformation from each active molecule) that results in the best fit is presumed to be the active conformation.
4. Abstraction– Transform the superimposed molecules into an abstract representation. For example, superimposed phenyl rings might be referred to more conceptually as an 'aromatic ring' pharmacophore element. Likewise, hydroxyl groups could be designated as a 'hydrogen-bond donor/acceptor' pharmacophore element.
5. Validation– A pharmacophore model is a *hypothesis* accounting for the observed biological activities of a set of molecules that bind to a common biological target. The model is only valid insofar as it is able to account for differences in biological activity of a range of molecules. As the biological activities of new molecules become available, the pharmacophore model can be updated to further refine it.

RESULTS

Docking was performed using the newly designed protein mutants with three ligand molecules (L-3, L-4 and L-8). From the overall docking scores we

identified that mutant 1 with L-4, mutant 5 with L-3 is having a high dock scores.

In the present study, Pharmacophore Mapping has been conducted using DISCOVERY STUDIO from ACCELRYYS to extract a 3D pharmacophore that reflects the important functional groups that are essential for inhibitor binding. Receptor-Ligand Pharmacophore Generation protocol produced pharmacophore models that were ranked according to their selectivity score, the higher the better. Thus, among the ligands selected from literature, pharmacophore mapping studies revealed that L-3, L-4, L-8 ligands possessed better pharmacophore feature.

DISCUSSION

Based on the fit value it is clear that three ligands molecules, ellagic acid, Curcuminoid and Methyl gallate are showing best pharmacophore features. Ligands namely ellagic acid, curcuminoid and methyl gallate are showing best values when performed docking with the target protein i.e, IL-6 with highest docking values: - 9.4494, -9.6214, and also showing good pharmacophoric features with fit values: 2.00, 1.991, 1.946. Thus the research scope for all the mathematical techniques and algorithms coupled with software programming languages, software development and deployment tools are to get a real boost. In addition, information technologies such as databases, middleware, graphical user interface (GUI) design, distributed object computing, storage area networks (SAN), data compression, network and communication and remote management are all set to play a very critical role in taking forward the goals for which the Bioinformatics field came into existence ^[12].

CONCLUSION

Because of one or other of side effects associated with synthetic drugs used for treatment of rheumatoid arthritis, now-a-days research is focused on development of drugs from bioactive compounds possessing anti rheumatoid arthritis activity. In the present project, insilico studies i.e. homology modeling using modeler 9.12 version, docking using molecular docking environment and pharmacophore features by DISCOVERY STUDIOs were carried out on some of the selected bioactive compounds which revealed that Ellagic acid, Curcuminoid and Methyl gallate found to possess anti rheumatoid arthritis activity and could be used as potent leads for development of IL-6R inhibitors with minimum side effects. Ligands namely ellagic acid, curcuminoid and methyl gallate are showing best values when

performed docking with the target protein i.e, IL-6 with highest docking values: - 9.4494,-9.6214, and also showing good pharmacophoric features with fit values: 2.00, 1.991, 1.946.

ACKNOWLEDGEMENT

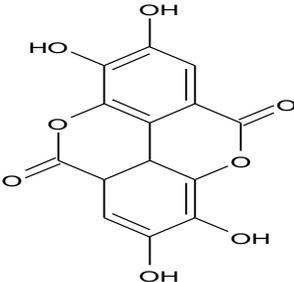
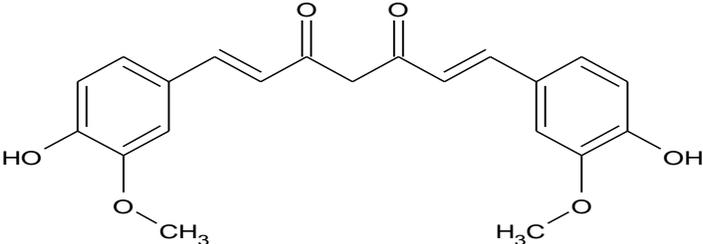
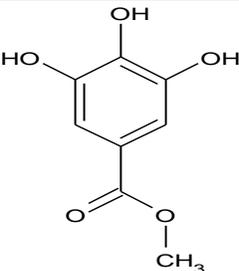
We would like to thank the Principal of University college of Technology, Osmania University for

providing the facilities and the necessary softwares for our work. We also thank the Director for The Centre for Distance Education for their support and extending access to the bioinformatics lab for the docking studies. We extend our sincere thanks to the technicians who supported our work in the Centre for Biotechnology, College of Technology, Osmania Univeristy, Hyderabad.

Table 1: Represents that the docking poses of both wild and Mutant protein structures of Modeled IL6R.

S.No.	Ligand	Wild type	Mutant-1 (His 288 to Tyr 288)	Mutant-2 (Asp 281 to Valine 281)	Mutant-3 (His 280 to Ser 280)	Mutant-4 (Ala 282 to Asp 282)	Mutant-5 (Gly 285 to Asp 285)	Mutant-6 (Val 289 to Leu 289)
1.	L-3	-12.1079	-11.0376	-10.7353	-10.7589	-11.7008	-9.6214	-12.9187
2.	L-4	-11.4690	-94494	-14.4766	-10.3042	-11.9232	-15.0658	-15.8036
3.	L-8	-13.2315	-10.7641	-13.1238	-11.6323	-11.9948	-13.9563	-13.1793

Table.2 .Ligand Pharmacophore Mapping results

S.NO	Name	Structure	Fit Value IL-6R
1.	Ellagic acid (L4)		1.946
2.	Curcuminoid (L8)		1.991
3.	Methyl gallate (L3)		2.00

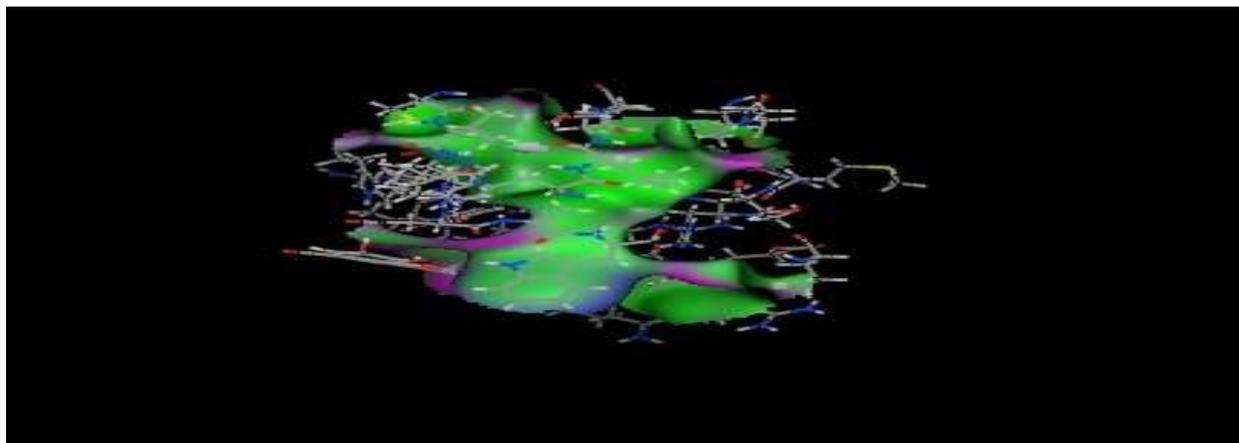


Figure 1: Docking of L3 with IL-6R

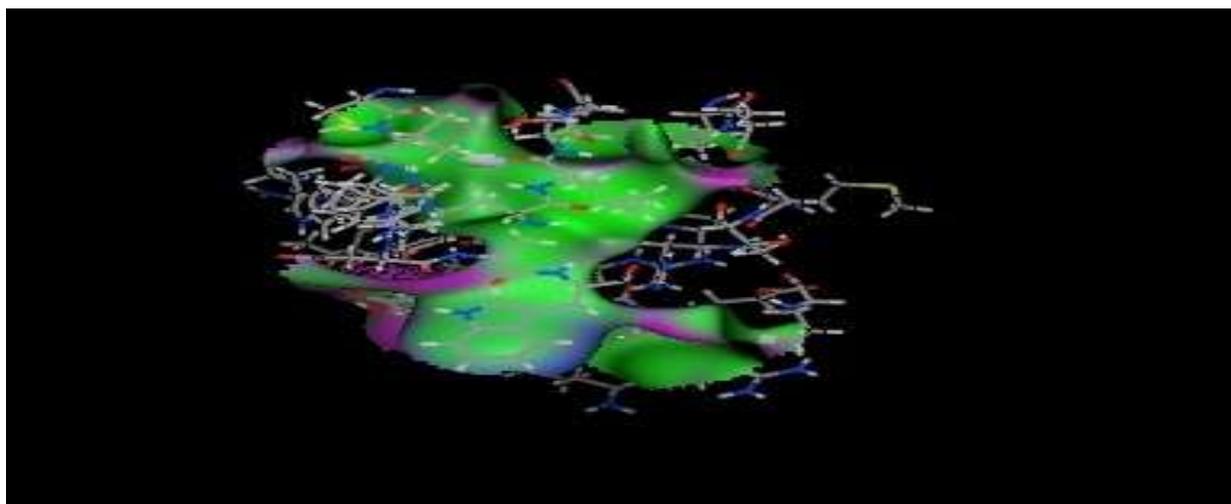


Figure 2: Docking of L4 with IL-6R

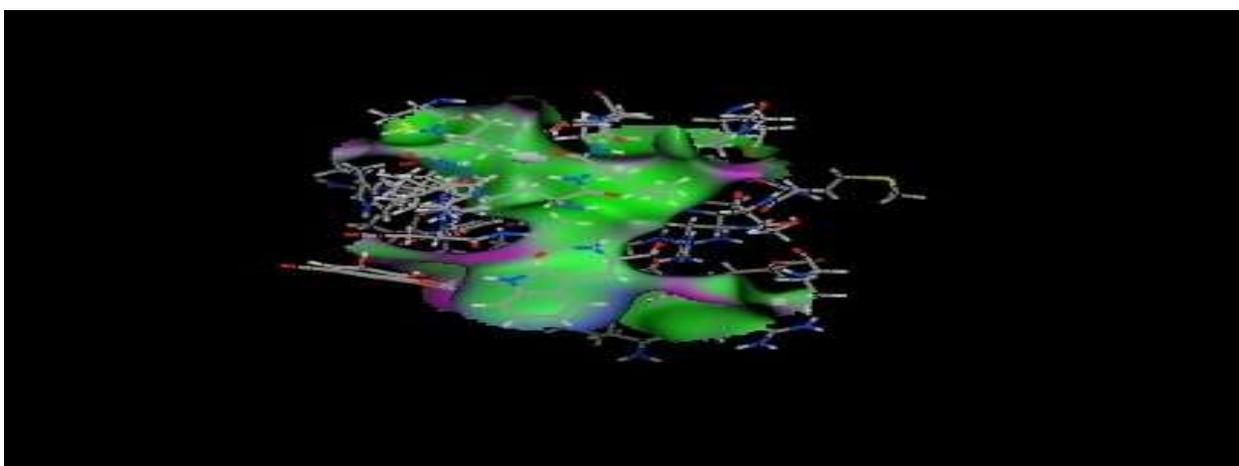


Figure 3: Docking of L8 with IL-6R

REFERENCES

1. Albani and Carson 1997 Etiology and pathogenesis of rheumatoid arthritis n: KoopmanWJ (ed) Arthritis and allied conditions. Lippincott Williams & Wilkins, Baltimore,MD, 979
2. Grassi et al. 1998 Grassi W, De Angelis R, Lamanna G, Cervini C (1998) Eur J Radiol 27(1):S18–S24.
3. Firestein GS, Zvaifler NJ. 2002; 46: 298–308. [PubMed]
4. Ankur Singh Bist *1, Babeesh Kumar 2 Singh 3 (7): July, 2014
5. Ursula Pieper, Narayanan Eswar, Hannes Braberg, M. S. Madhusudhan, Fred P. Davis, Ashley C. Stuart,¹ Nebojsa Mirkovic,¹ Andrea Rossi, Marc A. Marti-Renom, Andras Fiser,² Ben Webb, Daniel Greenblatt, Conrad C. Huang, Thomas E. Ferrin, and Andrej Sali* Nucleic Acids Res. 2004 Jan 1; 32(Database issue): D217–D222. PMID: PMC308829
6. Neena L. Summers, Martin Karplus; Volume 210, Issue 4, 20 December 1989, Pages 785–811
7. Vilar, Santiago; Cozza, Giorgio; Moro, Stefano; Volume 8, December 2008, pp. 1555-1572(18); Bentham Science Publishers
8. Pandrangi et al., Pharm Anal Acta 2015, 6:6
9. Plumed; RMSD http://plumed.github.io/doc-v2.1/user-doc/html/_r_m_s_d.html
10. Richard M Jackson; 02/2002; 16(1):43-57. DOI: 10.1023/A:1016307520660; [PubMed]
11. Riya Rane, Rahul Ravichandran, Pritam Kumar Panda and Shalaka Joshi*; Volume 4, Issue 07, 2012-2021.
12. Michael Lawrence, John Verzani; Programming Graphical User Interfaces in R; CRC Press, 08-Jun-2012.