



## ***In-Silico* Drug Design through Protein-Ligand Interaction Studies of Snake Venom Peptides**

\*Ambika Iyer

Department of Biotechnology, B.V. Bhoomaraddi College of Engineering & Technology, Hubli-580031,

\*E-mail id: [ambikaiyer@yahoo.com](mailto:ambikaiyer@yahoo.com)

Received: 10<sup>th</sup> August 2015, Accepted: 14<sup>th</sup> August 2015, Published: 1<sup>st</sup> September 2015

### **Abstract**

Toxins have evolved in plants, animals and microbes multiple times as part of defensive and/or prey capture strategies. Peptide toxins are usually found in animal venom. Most venom comprises a highly complex mixture of peptides, often with diverse and selective pharmacology. The source of peptide is specifically snake venom which is usually neurotoxins. We have selected four peptides Aggretin, Crostamine, Mambalgin, Myotoxin. In-silico drug design is crucial step to reduce the number of steps in drug designing process. The present study was focused to design a drug for snake venom peptides. The peptides are causative proteins for diseases like blood clotting, necrosis, acute renal failure, and anaphylaxis. The peptide structure was retrieved from Uniprot database and multiple sequence alignment was performed on the retrieved sequences using Clustal W. Rasmol software was used for protein visualization. Pubchem database was used for screening the chemical ligands. The chemical structure of selected ligands was obtained from Dundee pro drg server. The ligands were screened based on log P value for selection of better ligands which was used for docking studies. Docking was performed to find the inhibition activity against snake venom using HEX software. For the disease Coagulopathy the protein aggretin and ligand ellagic acid showed highest negative E-value -232.51, for the disease necrosis the protein crotamine and ligand sterigmatocystin solution showed E-value of -270.78, for the disease acute renal failure the protein aggretin and ligand 1LGS5JRP31-isoquinoline derivative gave E-value -229.40, for the disease anaphylaxis the protein aggretin and the ligand ketotifenfumarate gave E-value -258.49. Vibrational analysis was performed to check the energy changes and it was found that the energy of docking in case of Hex after vibrational analysis was higher. Different conformational structures of aggretin were obtained and redocked with the ligand ellagic acid for the disease and the highest E-value obtained was -255.22, for the disease necrosis different conformational structures of mambalgin was used to redock and the highest negative E-value obtained was -255.25, for the disease acute renal failure and anaphylaxis the vibrational analysis was performed using the protein aggretin and the highest E-value obtained was -235.43 and -275.69 respectively.

Autodocking was done using vina for different conformations of ligand. For all the four diseases the protein aggretin showed higher affinity towards the ligand. For the disease coagulopathy the highest affinity towards ligand ellagic acid of value -5.8, for the disease necrosis the highest affinity towards ligand Sterigmatocystin solution of value -7.9, for the disease acute renal failure the highest affinity towards ligand Isoquinoline derivative of value -6.0, for the disease anaphylaxis the highest affinity towards ligand Ketotifenfumarate of value -5.6. Therefore, this in silico analysis provides rapid and potential approach for identification of drug target and designing of drug.

### **Keyword:**

Snake Venom Peptides, Protein-ligand interaction, Docking, Vibrational analysis, Autodocking

### **Introduction**

During drug design 'disease selection' is a potent step. The human body consists of four main molecules; carbohydrates, lipids, nucleic acid and proteins. As carbohydrates, lipids and nucleic acid don't have binding sites, the drug acts on protein. Drugs are either made of ligands or peptides. Toxins have evolved in plants, animals and microbes multiple times as part of defensive and/or prey capture strategies. Peptide toxins are usually found in animal venom associated with specialized Envenomation Apparatus that allows their delivery into the soft tissue of animals via subcutaneous, intramuscular or intravenous routes [Mackessy et al, 2010]. Most venom comprises a highly complex mixture of peptides, often with diverse and selective pharmacology. Despite this diversity, venom peptides seem to have evolved from a relatively small number of structural frameworks that are particularly well suited to addressing the crucial issues of potency and stability. It is this evolved biodiversity that makes venom peptides a unique source of leads and structural templates from which new therapeutic agents might be developed. Crude venoms contain a diverse array of different peptides many of which are bioactive [Saviola A.J. et al, 2013]. Snake venoms are complex mixtures of proteins, nucleotides and inorganic ions. These combinations confer a formidable array of toxic properties on the venom, the peptides and polypeptides being responsible for

a variety of toxic properties. The number of venom components in venomous animals like snake, scorpion or cone snail ranges from 50-200 toxins [Tan PT et al, 2003]. The source of peptide is specifically snake venom. We have selected four peptides Aggreitin, Crotamine, Mambalgin, and Myotoxin for drug designing. Snake venoms are important tools in toxicology, neuroscience, and pharmacology. The venom components are highly variable and functionally complex and they offer many research opportunities [Joyce P.Y.S et al, 2004]. These peptides cause diseases like blood clotting, necrosis, anaphylaxis, acute renal failure, neurological and psychiatric disorders. Proteins within the body are always stable. But outside the body the proteins have to be studied and stabilized. To stabilize the protein the structure of the protein is to be known hence we can stabilize it by optimizing the geometry of the protein. The geometry optimization can be done by minimizing the energy of the protein molecule. For the protein to bind with the target site both have to have equal energy levels for 100% binding. Screening of the ligand is necessary to eliminate the junk data. Screening can be done by Lipinski rule. Ambiguity study of the ligand is done for the chemical properties of the ligand. Docking is done to bind the receptor to the drug molecule to attain protein-ligand binding. The ability to predict the conformations and energies involved in the binding of small molecules to proteins is quite crucial in designing potential drugs that can interact favorably with their target proteins. Docking is also used to predict protein-protein complexes, which are helpful in determining the quaternary structures of intrinsically multimeric proteins as well as to gain an understanding of the protein interaction networks [Alex Mathew Jet al, 2009]. After docking the selected ligand binds with the receptor, the expected output is docking energy. Among the different docking pairs the best pair can be selected based upon the lowest energy value which is an indication of high stability. High negativity shows good peptide receptor pairing. High negativity leads to less energy leading to high stability.

#### **Materials & Methods**

**Phylogenetic Analysis:** Phylogenetic analysis was carried out using ClustalW2 software. ClustalW2 is a general purpose multiple sequence alignment program for DNA or proteins. Retrieve all the peptide sequences from Uni-Prot: [www.uniprot.org](http://www.uniprot.org) in fasta format and paste in notepad leaving one line space between each sequence. Open the url: <http://www.ebi.ac.uk/clustalw> which contains graphical user interface (GUI) of the ClustalW software. Copy the sequence from notepad and paste them in textbox provided in ClustalW page. Default Parameters are used for the analysis. Click

on the “submit” button. Result including alignment and tree appear in the window after the page refreshes. The alignment may be seen in colours by selecting the option “show colours”.

#### **Screening of Proteins:**

Protein screening is done to identify the causative peptides for the diseases. This was done by using the Genecards server. Url: <http://www.genecards.org/> Elimination of junk data by screening for the protein specific to blood clotting, necrosis, acute renal failure and anaphylaxis. Once the peptides are narrowed specific to the disease, identify the structure of the peptide. The structure of the peptide can be obtained from the server RCSB: <http://www.rcsb.org/pdb/home/home.do> the structure of the peptides can be viewed using the software RASMOL, by using the commands we can view the various atoms and make conformational changes.

#### **Structural Analysis & Verification Server:**

Structural analysis and verification server is also known as SAVES. For SAVES open the url: <http://services.mbi.ucla.edu/SAVES/> This metaserver runs 6 programs for checking and validating protein structures during and after model refinement. The programs are PROCHECK, WHAT\_CHECK, ERRAT, VERIFY\_3D, PROVE, Ramachandran plot.

#### **Screening of Ligands:**

Screening of the ligand is necessary to eliminate the junk data. Screening can be done by Lipinski rule. There are four basic principles of Lipinski rule, they are; Molecular weight should range between 300-600, Number of hydrogen donors should range from 1-5, Number of hydrogen acceptors should range from 5-10, The value of log P should be 5 or less. The pH of our blood is 7.4. Log P denotes the toxicity of the ligand. As the ligand is consumed and enters our blood stream its toxicity value should be less than or equal to 5. To study the ambiguity of the ligand we use software such as Dundee pro drg server. Open the url: <http://davapc1.bioch.dundee.ac.uk/cgi-bin/prodrg> Ambiguity study is the tertiary screening of the compound. This tool is used for the chemical properties of the ligand. Hence through primary, secondary and tertiary screening we can identify the target drug for the causative protein and obtain the structure of the ligand from the server DUNDEE PRO DRG under “all hydrogen (PDB)” option, and by saving the coding in the notepad under .pdb

#### **Docking Using Hex:**

Upload a pair of protein and ligand structures in PDB format in Hex server. Provide an e mail

address for notification of the status of their jobs. Default parameters are used for carrying out the jobs. To be able to analyze the docking, the e-values were obtained using the Hex software. Install the free Hex version 6.0 software available online suitable to the existing operating system. In the Hex window, go to File>Open>Receptor. The receptor and ligand molecules must be previously saved in the hex/examples directory of Hex. This done by opening the pdb files of receptor and ligand in SWISS Pdb Viewer and saving them in project mode using “magic fit” tool. The ligand file must be opened, File>Open>Ligand. The process of docking is started by using the option Controls>docking. There are five stages in the process. Once the five stages are completed it gives a docked structure indicating the corresponding e-values. More negative the e-value, the more efficient is the docking process. [D.W.Ritchie et al, 2000]

#### **Vibrational Analysis:**

The site used for this provides online servers for algorithms such as normal mode calculation, structural refinement, solvation and mutation and transition path calculation. Calculate the first major vibrational mode of the assigned protein using the web server NOMAD-REF: <http://lorenz.immstr.pasteur.fr/> Under “Normal Model Calculation”, go to, “Submit a Job (PDB)” and follow the online instructions to upload the protein pdb file. Follow the instructions to obtain the results. Re-dock the proteins obtained from vibrational calculations with ligand using HEX server.

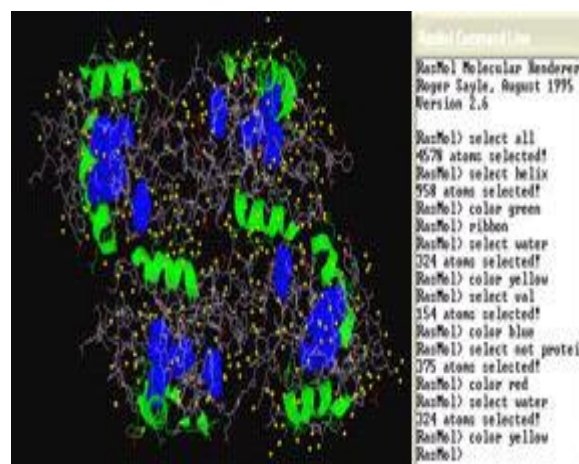
#### **Autodocking using Vina:**

MGL tool is used as the autodock tools. Click on file, read molecule and open the protein molecule in pdb format. To add hydrogen structures click on edit, hydrogen, add, polar only, no bond order, yes, ok. Hydrogen structures are required for the protein in pdbqt format. Use dashboard to display the molecular surface or hide it. Click on grid, macromolecule, choose, protein, select molecule, ok. The protein has to be saved in pdbqt format. Click on grid, gridbox, set values. We can choose the part of the receptor where the ligand can be docked, also known as search space. Set sizes of the search space, autodock tools measures the sizes in points instead of angstroms to convince it we set spacing (angstrom) to 1.0 is the scaling factor. Click on the ligand, input, open, select ligand molecule in pdb format. Using dashboard we can hide the protein to view the ligand properly. Click on ligand, torsion tree, choose torsions, to view the different bonds of the ligand. The ligand is saved by clicking ligand, output, save as pdbqt. The ligand is saved in pdbqt format. Load command prompt and change the drive to the location of the

saved structures. Vina is default saved in program files under the name The Scripps Research Institute. Help function is used to print all the other options. Type the command “\Program Files\The Scripps Research Institute\Vina\vina.exe”—help. The value used for setting the search space is to be saved in a notepad. Open notepad and type the command receptor = protein.pdbqt, ligand = ligand.pdbqt, out = all.pdbqt, center values and size values as same as gridbox. Vina produces one or more binding modes each goes to its own pdbqt file. For our convenience we create out = all.pdbqt all the binding conformations into one model. Vina doesn't require us to calculate the gridmap files or to determine the end type of the grid maps. Vina calculates its grid maps automatically. The modes described the binding modes that are available and the corresponding affinity values. According to vina the first binding mode is the best one. PyMOL software is used to visualize the output; the ligand can be viewed in the search space of the protein.

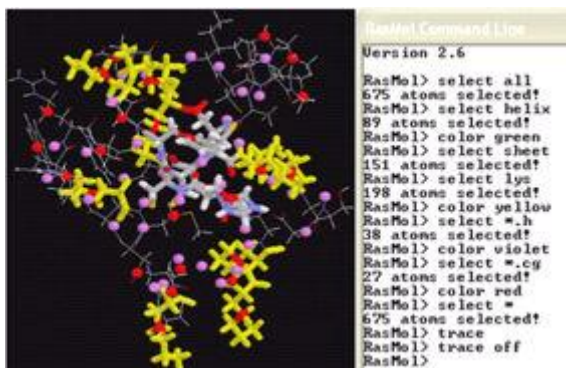
#### **Results & Discussion**

Visualization of protein structure was done using Rasmol software. Aggretin showed a total of 4578 atoms. The total energy values of the predicted 3-D model were calculated as 91.0% of Ramachandran plot. From Errat the Overall quality factor obtained was 94.781. Verify\_3D results shows 100% of the residues had an averaged 3D-1D score >0.2 and test passed. The total number of buried outlier protein atoms was 27(1.0) of scored atoms.



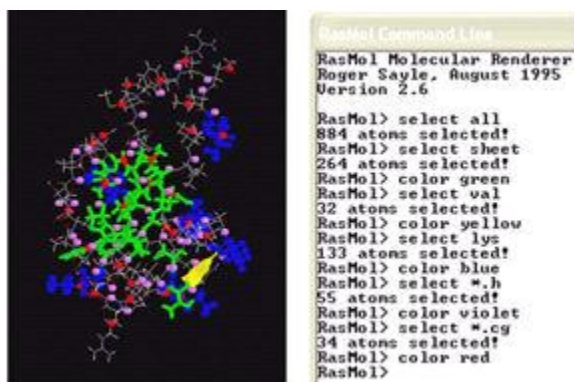
*Aggretin*

Crotamine showed a total of 675 atoms, the total energy values of the predicted 3-D model were calculated as 72.7% of Ramachandran plot. From Errat the Overall quality factor obtained was 87.500. Verify\_3D results shows 100% of the residues had an averaged 3D-1D score >0.2 and test passed. The total number of buried outlier protein atoms was 3(3.4) of scored atoms.



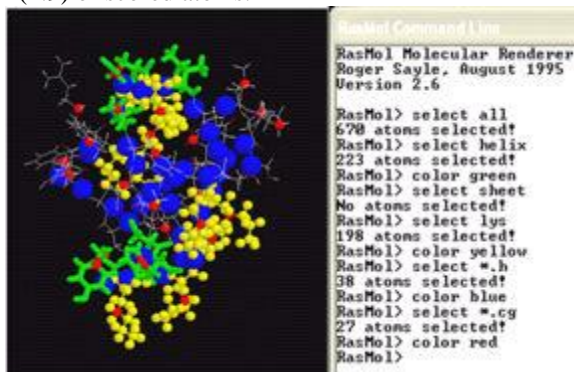
Crostamine

Mambalgins showed a total of 884 atoms. The total energy values of the predicted 3-D model were calculated as 82.4% of Ramachandran plot. From Errat the Overall quality factor obtained was 97.619 Verify\_3D results shows 70.18% of the residues had an averaged 3D-1D score >0.2 and test passed. The total number of buried outlier protein atoms was calculated to be 3(2.2) of scored atoms.



Mambalgins

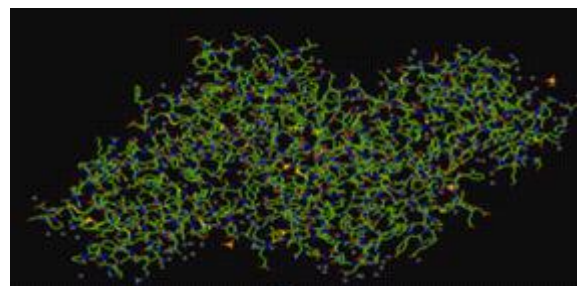
Myotoxin showed a total of 670 atoms. The total energy values of the predicted 3-D model were calculated as 60.6% of Ramachandran plot. From Errat the Overall quality factor obtained was 76.47 Verify\_3D results shows 100% of the residues had an averaged 3D-1D score >0.2 and test passed. The total number of buried outlier protein atoms was 2(1.9) of scored atoms.



Myotoxin

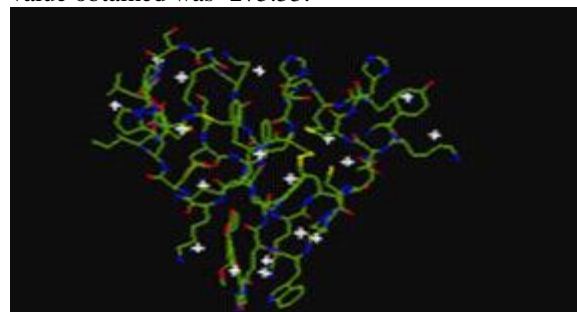
Ligands were screened through Lipinski rule for each disease. After primary, secondary and tertiary screening for the disease Coagulopathy the ligands screened were Ellagic acid, Gama-Carboxyglutamic acid, 3-Amino-1,1,3-propanetricarboxylic acid. Ligands for the disease Necrosis screened are Sterigmatocystin solution – 10048-13-2, PiperidinylIsoindole – 4Z8R60RS6L, Sterigmatocystin. For the disease Acute Renal Failure the ligands screened were Isoquinoline derivative - 1LGS5JRP31 and Nomifensine maleate. Ligands screened for the disease anaphylaxis is KetotifenFumarate.

For the disease Coagulopathy, docking using hex software the ligand ellagic acid showed better efficiency towards the peptide aggretrin. It gave an E-value of -232.51 which is the most negative among docking for all the other ligands. Further Vibrational analysis was done for different conformational structures of aggretrin. Further re-docking was done with the ligand ellagic acid. The E-value obtained was -255.22.



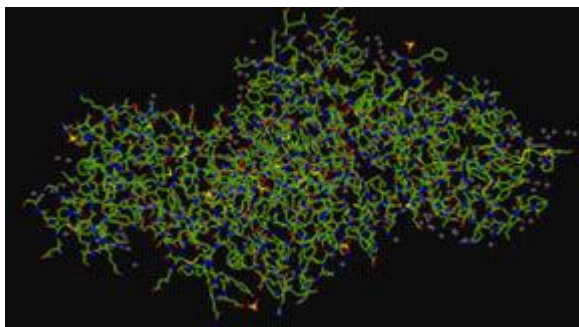
Docking Result for the disease Coagulopathy

In the case of Necrosis, sterigmatocystin solution showed higher efficiency towards the peptide crostamine, the E-value was -270.78. Further vibrational analysis was done for different conformational structures of the peptide. The E-value obtained was -275.55.



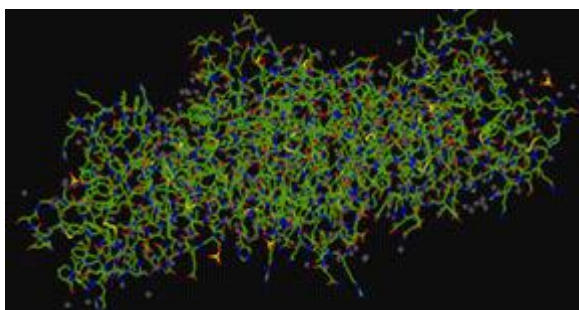
Docking Result for the disease Necrosis

For the disease Acute Renal Failure, the ligand 1LGS5JRP1-isoquinoline derivative showed higher efficiency towards the peptide aggretrin giving the E-value -229.40. Further vibrational analysis was done for different conformational structures of peptide aggretrin. The E-value obtained was -235.43.



*Docking Result for the disease Acute Renal Failure*

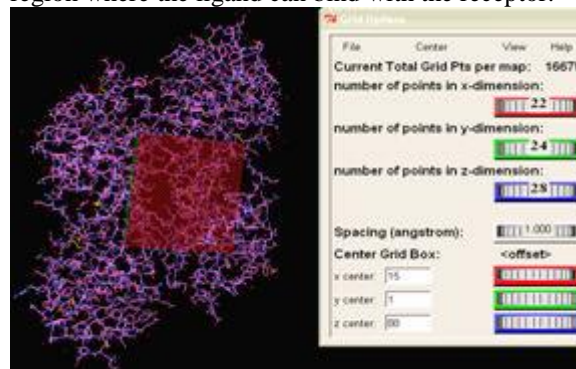
In the case of Anaphylaxis, ketotifen fumarate showed higher efficiency towards the peptide aggretrin, the evaluate was -258.49 Further vibrational analysis was done for diiferent conformational structures of aggretrin and the E-value obtained was -275.69



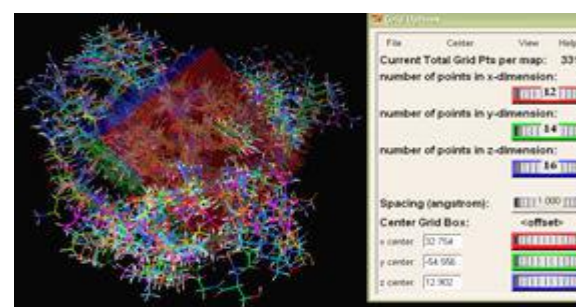
*Docking Result for the disease Anaphylaxis*

Autodocking was done using the software vina, AutoDockVina is an open-source program for doing molecular docking. It was designed and implemented by Dr. Oleg Trott in the Molecular Graphics Lab at The Scripps Research Institute. It is a system consisting of a receptor and a ligand. Collectively refer to their degree of freedom as conformation. For each conformation there is a corresponding potential energy value. However as we consider a solvent we have a energy like quantity. The global minimum corresponds to the negative conformation. The deeper or narrow the well is it more preferable because of the entropy. The quantity we are interested in is called the chemical potential. It can be used to calculate the free energy of binding and the affinity of the ligand to the receptor. However to calculate the chemical potential from the first principles is very expensive conformationally. Docking program generally uses a scoring function, which is a function tuned to recognize the promising conformations and to predict the affinity as much as possible. [O. Trott et al, 2010] MGL tools and PyMOL are tools developed by molecular graphics lab. Both are autodock tools to visualize the structures and to make conformational changes. Through MGL tools we can visualize the

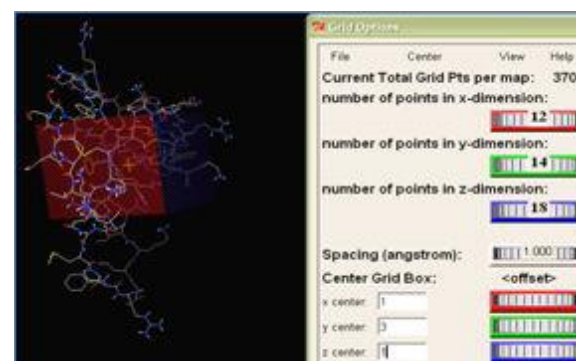
search space. With search space we can assign the region where the ligand can bind with the receptor.



*Visulization of search space: Aggretrin*



*Visulization of search space: Crotnamine*



*Visulization of search space: Mambalgins*



*Visulization of search space: Myotoxin*

For the disease coagulopathy the highest affinity of value -5.8 was shown by ligand ellagic acid towards the peptide aggretrin. For the disease necrosis the highest affinity of value -7.9 was shown by ligand Sterigmatocystin solution towards the peptide aggretrin. For the disease acute renal

failure the highest affinity of value -6.0 was shown by ligand Isoquinoline derivative towards peptide aggretrin. For the disease anaphylaxis the highest affinity of value -5.6 was shown by ligand Ketotifen fumarate towards the peptide aggretrin.

### Conclusion

The Phylogenetic Analysis was carried out using CLUSTAL W tool. This showed the evolutionary relationships of the different species which were selected for the study. The snake venom peptide structure was obtained from RCSB site in PDB format and the protein visualization and modeling were done using Rasmol software. The structure was analyzed and validated using SAVES server. The ligands, were screened using lipinski rule for each disease from pubchem server. The Hex software gives corresponding e-values for each docking. More negative the E-value more efficient is the docking. For the disease coagulopathy, the ligand ellagic acid showed better efficiency towards the peptide aggretrin. It gave an E-value of -232.51 which is the most negative among docking for all the other ligands. In the case of necrosis, sterigmatocystin solution showed higher efficiency towards the peptide crotamine, the e-value was -270.78. For the disease acute renal failure the ligand 1LGS5JRP1-isoquinoline derivative showed higher efficiency towards the peptide aggretrin giving the e-value -229.40. In the case of anaphylaxis, ketotifen fumarate showed higher efficiency towards the peptide aggretrin, the e-value was -258.49. Vibrational analysis was used to find vibrational modes of the receptor and docking was conducted on the most structurally different form with the ligand. The docking energy for all the diseases before was low and after the vibrational analysis was found to be the higher. In the case of coagulopathy the value was -255.22, for necrosis the e-value was -255.25, for acute renal failure the e-value was -235.43, for anaphylaxis the e-value was -275.69. The docking energy for all the diseases after the vibrational analysis was found to be the higher than the regular docking, Hence it was observed that there is change in the energy value for the docking of the proteins, obtained from the vibrational analysis, with the ligand. This shows that the conformational change achieved by vibrational tool is associated with the protein binding site. Further analysis was done through autodocking. The conformations were recognized using the scoring functions to predict the affinity. Autodocking was performed to find affinity values of the ligand to receptor. Autodocking was performed using vina software. Through autodocking tools the binding site/location can be viewed and optimized. Vina produces more than one binding mode, according to vina the first binding mode is the best one. After performing autodock it was observed that the ligands showed higher affinity towards the peptide

aggretrin for all the four disease. Hence the results of docking was confirmed with autodocking.

### Acknowledgement

I would like to acknowledge our Guide Prof. Zabin K. Bagewadi for her endeavor support & Dept of Biotechnology for providing facilities and the Principal Dr. Ashok Shettar for facilitating us to carry out our final year Project.

### Reference

1. Alex Mathew J, Nixon Raj N, Docking Studies on Anticancer Drugs for Breast Cancer Using Hex Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009 Vol I IMECS
2. Joyce P.Y.S, A.M. Khan, Paul T. J. Tan, J.L.Y. Koh, S.H Seah, C.Y Koo, S.C. Chai, A. Armugam, V. Brusica, and K. Jeyaseelan, Systematic analysis of snake neurotoxins functional classification using a data warehousing approach. *Bioinformatics*, (2004) **20**: 3466 – 3480
3. Mackessy, Stephen P. "Evolutionary trends in venom composition in the Western Rattlesnakes (*Crotalus viridissensulato*): Toxicity vs. tenderizers". (2010) *Toxicon* **55** (8): 1463–1474.
4. O. Trott, A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading, *Journal of Computational Chemistry* 31 (2010) 455-461
5. D.W. Ritchie & G.J.L. Kemp Protein Docking Using Spherical Polar Fourier Correlations", *PROTEINS: Structure Function Genet.* **39**, (2000) 178-194.
6. Saviola, A.J.; Chiszar, D.; Busch, C.; Mackessy, S.P. "Molecular basis for prey relocation in viperid snakes". *Bmc Biology* **11**. (2013)
7. Tan PT, Khan AM and Brusica V, *Bioinformatics for venom and toxin sciences. Brief Bioinform*, **4**(1):53-62, (2003).