A Dynamic 3D Clustering Algorithm for The Ligand Binding Disease Causing Targets

D. Ponmary Pushpa Latha[†] **and Dr. D. Jeya Sundara Sharmila**[††],
Assistant Professor Assistant Professor
Department of Computer Application Department of Bioinformatics
Karunya University
India-641 114, Tamil Nadu, INDIA.

Summary
The basic characteristics of ligand-protein interaction is commonly referred as "molecular recognition. It concerns the specificity as well as stability of ligand binding. Molecular recognition is used in the advancement of active substances, to determine whether a substance can be used as a drug. Our aim is to build up a dynamic clustering algorithm for the grouping ligand with slight involvement of a database administrator (DBA) and to preserve an satisfactory query response time at all conditions.

Key words:
Clustering, Ligand, Protein, Binding Site.

1. Introduction

A disease is a deviant state of an organism that weakens bodily functions. Brookhaven Protein Data Bank [4]; a database and format of files, which describes the 3D structure of a protein or nucleic acid, as determined by X-ray crystallography or nuclear magnetic resonance (NMR) imaging. A ligand [6] is distinguished as an addictive nearby in the crystal structure of a protein other than the obviously occurring amino acids, metal atoms and water. PDB bind [13] was developed to provide the details of an experimental binding data for the protein ligand complexes [1], [2]. In traditional database clustering, objects are grouped based on similarity in access patterns whereas, in data mining clustering, objects are clustered based on similarity in the actual data. The more similar two objects are, the more likely they belong to the same cluster. Data mining clustering algorithms use a distance measure to compute the distance between any two data objects values. Data objects are then assigned to clusters such that, the distance between objects within a cluster is less than a given threshold and the distance between objects in different clusters is greater than a given threshold. The traditional attribute clustering generates attribute clusters, which is also called vertical clusters. Many techniques have been evolved for record clustering in various databases like relational databases, object-oriented databases, attribute clustering and partitioning with record clustering, relations are broken down into groups of records, based on their affinity. PDB-Ligand [7] allows interactive clustering of ligand binding structures based on user specific rmsd values. This dynamic clustering algorithm gives effective clustering within the user specified models and within the precised rmsd values .

Manuscript received February 5, 2009
Manuscript revised February 20, 2009

2. Methodology

2.1 Data Extraction

In house tools are developed to extract and separate the protein and ligand coordinates into the relational database using Java for the disease causing ligand family. The Extractor of In house tools has the capacity to parse any number of PDB files using FileDialog Class and StringTokenizer Class. Hetatm, atm, hetnam, ligand are the tables created in the system. Hetatm table contains ligand details while protein informations are maintained in atm table. Hetnam table has ligand names from protein data bank. For defining the binding [14] site, Soga et al. [12] used a cutoff of 4.5 Å , Malik et al. [9] took a cutoff of 3.5 Å and A.S Reddy et al. [3] applied a cutoff of 4Å in their graph. With a broad spectrum, the cutoff distance of 6 Å has been taken for analysis.

2.2 System Architecture

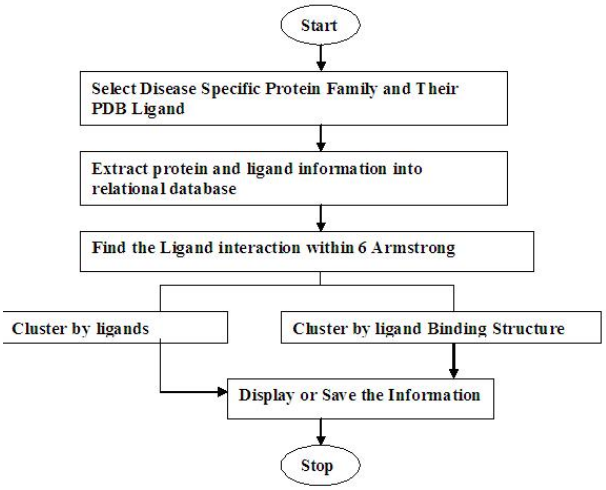

Fig. 1 System Archtiecture

2.3 Clustering

In proposed methodology (Fig 1), clustering is done for the ligand which is available in the relational database. Here the given support level threshold value is 0.5. Data mining clustering [7] typically looks for similarity in the actual data to group data objects based on relevant distance function. The system uses the rmsd attributes for clustering based on distance value and also gives 3D(Dimension) clustering [5] [8] for the user given RMSD value.

2.3.1 Algorithm for Clustering Ligand Name

1. Start
2. Find the distinct Item Iigand Name
3. Fix the reference ligand which serves as the origin for clustering.
4. Set the Default value of root mean square deviation (RMSD) from the origin (Reference) cut off value is 0.5.
5. For i=0 to n Step 0.5
6. Perform the iteration until n=3.5 RMSD and Find the total Number of Ligand binding structure.
7. End

2.3.1. Dynamic Ligand based 3D Clustering

1. Start
2. Select a Disease Specific Protein Target
3. Select a Ligand from the protein ligand binding site
4. Select two ligand pdb file, Fix one model as reference model, and the remaining model as a compared model
5. Find the RMSD value of the protein ligand binding site for the pair of model considered.
6. Translate the Ligand Models into origin.
7. Perform dynamics along it's φ, θ, ψ, axis into 10 degree intervals upto 360, 180, 360.
8. Filter out unfitting rotation and find the exact superimposing degrees φ1, θ1,ψ1.
9. Perform dynamics from φ1-5, θ1-5,ψ1-5 to φ1+5, θ1+5,ψ1+5 into 1 degree and find the exact superimposing degrees φ2, θ2,ψ2..
10. Translate the protein ligand binding model into origin
11. Apply the φ2, θ2,ψ2 for dynamics of protein ligand binding model.
12. Save the General Cluster files.
13. From the General Cluster files, read M files and translate into .pdb format for the the selected models within the RMSD cut off value specified by the user .
14. Display the 3D Clustering View for the user
15. End

2.4 Results and Discussion

In Fig. 2, Cholera Toxin Family [10], [11] Ligand Name GLA is taken as the example for ligand based clustering. Totally 15 models are available in the protein ligand binding site. 105 is taken as a reference model. 10 models are having the RMSD value less than 0.5 and the remaining 5 models are above 0.5. Fig. 2 gives the ligand based clustering of GLA. Fig 3 provides the user interface for the user specification of RMSD value and the selection of model. Fig 4 displays 10 model in 3 Dimension view for 0 to 0.5 rmsd value. Translation and Dynamics are done by java programming language.

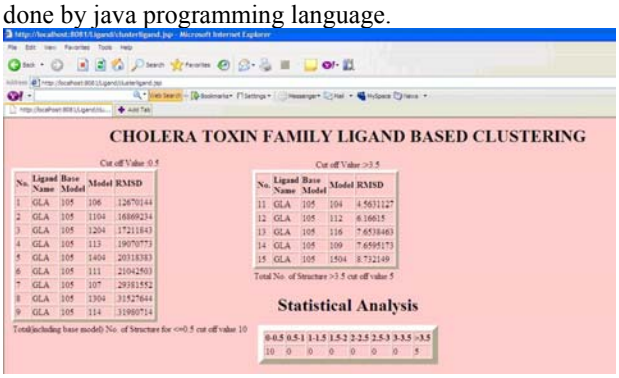

Fig.2 Clustering by Ligand Name in GLA

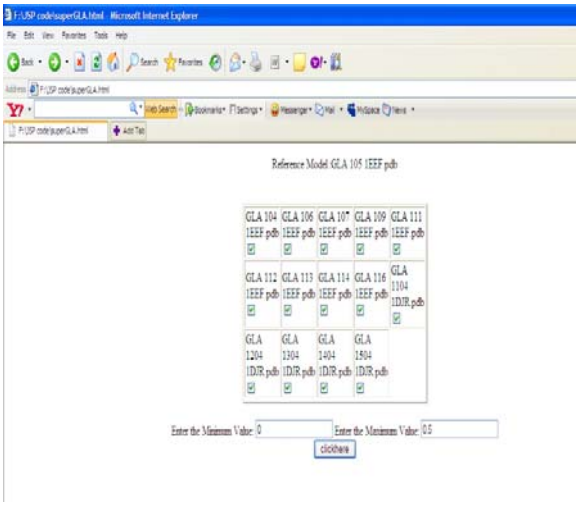

Fig.3 3D Clustering of GLA Model Selection

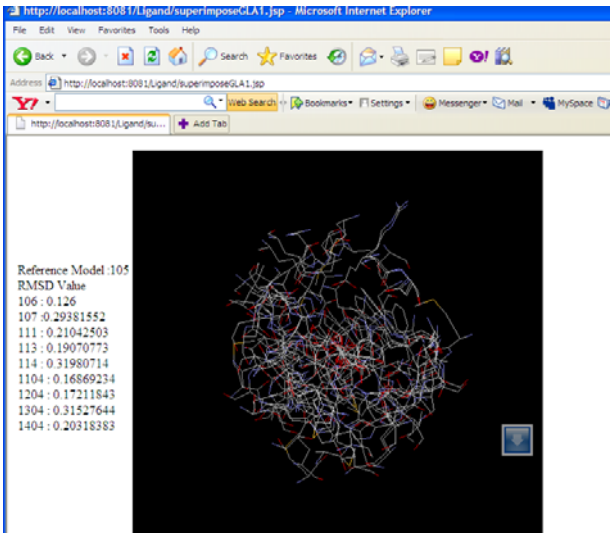

Fig.4 3D Clustering of GLA

3. Conclusion

This system will be helpful for effective drug design. 3D Clustering, Clustering by ligand name are exhibited. 3D Clustering is done by the system in a dynamic way with the minimum amount of time for the user specified RMSD ranges. Since the dynamics values were already calculated by the system, the speed was improved in the 3D view of Clustering.

References

[1] A. S. Reddy et al. (2006). From subtle to substantial: role of metal ions on interactions. J. Phys. Chem. B 110, 2479–2481.

[2] A. S. Reddy et al. (2007). Virtual screening in drug discovery—a computational perspective. Curr. Proteins Pept. Sci. 8, 329–351.

[3] A. S. Reddy et., al. (2008). Protein ligand interaction database (PLID). Computational Biology and Chemistry. 2008;

[4] Berman, H.M., et al. (2000) The protein data bank. Nucleic Acids Res. 28, 235–242.

[5] Lemmen et al., (1998). RigFit: A new approach to superimposing ligand molecules. Journal of Computer Aided Molecular Design. 12, 491-502.

[6] Gold, N.D., Jackson, R.M. (2006) A searchable database for comparing protein–ligand binding sites for the analysis of structure–function relationships. J. Chem. Inform. Model. 46. 736–742.

[7] Jae-Min Shin and Doo-Ho Cho (2005) PDB-Ligand: a ligand database based on PDB for the automated and customized classification of ligand-binding structures. Nuclic Acids Research. 33, D238-D241.

[8] Kabsch,W. (1978) A discussion of the solution for the best rotation to relate two sets of vectors. *Acta Crystallogr.*, ,1978, A34**,** , 827–828.

[9] Malik, A. and Ahmad, S. (2007) Sequence and structural features of a neural network. BMC Struct. Biol. 7, 1–14.

[10] Sharmila DJ and Veluraja K. (2006) conformations of higher gangliosides and their binding with cholera toxin investigation by molecular modeling, molecular mechanics and molecular dynamics. J.Biomolecular Structure and Dynamics. 23. 7, 641-56.

[11] Sharmila DJ, Veluraja K. (2004) Monosialogangliosides and their interaction with cholera toxin-investigation by molecular modeling and molecular mecanics. J.Biomolecular Structure and Dynamics. 21. 4, 591-614.

[12] Soga, S., et al, (2007) Use of amino acid composition to predict ligand-binding sites. J. Chem. Inform. Model, 47, 400–406.

[13] Wang, R., et al,. (2005) The PDBbind database: methodologies and updates. J. Med. Chem . 48, 4111–4119.

[14] Zukang Feng., et al,. (2004) Ligand Depot: a data warehouse for ligands bound to macromolecules. Bioinformatics. 20, 2153–2155