Int J Pharm Bio Sci 2013 Apr; 4(2): (B) 51 - 61

Research Article **Bioinformatics**

International Journal of Pharma and Bio Sciences **ISSN 0975-6299**

PROTEIN-LIGAND BINDING DATA INTEGRATION FOR "DIABETES DISEASE RELATED STUDIES" IN DIFFERENT ORGANISM

D.PONMARY PUSHPA LATHA[*1] AND D.JEYA SUNDARA SHARMILA[2]

[1]Department of Computer Applications, Karunya University, Coimbatore, India.
[2]Department of Bioinformatics, Karunya University, Coimbatore, India.

ABSTRACT

Protein-Ligand Binding Data of Diabetes Disease Related Studies is fashioned to provide the amino acid compositions of the following organisms in the ligand binding environment: *Escherichia coli, Homo sapiens, Mus musculus, Oryctolagus cuniculus, Rattus norvegicus, Sus scrofa, Saccharomyces cerevisiae* and *Spodoptera frugiperda.* The physico-chemical properties and quantum mechanical descriptors of the diabetes disease related ligands are depicted. Keyword search for the chemical name and related information are incorporated in the database. The average tightness and co-crystal contact of ligand binding in different organisms are also calculated. Atomic coordinates for ligands are annexed in XML format.
(http://www.ponmary.com/LandingPage/dia/Diamain.html)

KEYWORDS:Protein, Ligand, Database, Organism, Diabetes.

D.PONMARY PUSHPA LATHA
Department of Computer Applications, Karunya University, Coimbatore, India.

Corresponding author

INTRODUCTION

The specificity as well as stability of ligand binding is commonly referred to as "ligand-protein interaction" or "molecular recognition". Molecular recognition is used to determine whether a substance is used as a drug in the advancement of active substances. The purpose of ligand is able to bind and form a complex biomolecule[1]. A living thing is called as an Organism. Animals and plant are organisms. Air, water, nutrients, energy and living place are the basic needs of organism. Organisms which contain millions of cells are called as multicellular organisms. Microorganisms are not visible by the naked eye. Organisms which are having a single cell are called as unicellular organisms or single celled organisms. A disease is a deviant state of an organism that weakens bodily functions. It is suggested to incorporate disease-specific approach in the administrative databases to enhance the effectiveness of organizational databases[2]. Normal metabolism of carbohydrates, lipids, proteins, water and nucleic acids are altered by metabolic disorder[3]. Diabetes is one of the metabolic disorders. Insulin in the human body is mandatory to digest the sugar, starches and other food particles into energy. A person is affected by diabetes when insulin does not produce in their body. Overweight, lack of exercise, family history and stress increases the composition levels of diabetes disease[4]. Diabetes is the reason for the following diseases: Adult blindness, End-Stage Renal Disease (ESRD), gangrene, amputations, kidney failure, cardiovascular problems, heart problem, stroke and neuropathy. Polyuria (frequent urination), polydipsia (increased thirst) and polyphagia (increased hunger) are the symptoms of diabetes disease. Diabetes disease is controlled to a greater extent, but it is not totally curable. Diet, exercise and medication are used to control the diabetes disease[5]. Diabetes is an enduring disease which is caused by elevated levels of sugar in the blood[4, 5]. DiaMedBase have the information of Medicinal plants for diabetes[6]. Plants used for the treatment of diabetes with their chemical constituents are reported in Phyto-Mellitus. Antioxidant and free radical scavenging are the active principles of these plants[7]. The following are the organism in which precise analyses are carried out in the present study. Animals have the capacity to capture their food and shelter. Fungi and bacteria consume the animalwastes and dead bodies. Animalia comprises of animals[8]. *Oryctolagus cuniculus* is small mammal which is found in countries like Europe and Japan. The male, female and young rabbits are called by the following names: buck, doe and kit[9]. *Sus scrofa* is a species in the family Suidae which exists in countries like Europe, North Africa, America, Australia and Indonesia[10]. Humans are called as *Homo sapiens*. They are proficient in the following skills: reasoning, language, introspection and problem solving[11]. Long-tailed rodents resembling mice in larger size are called as *Rattus norvegicus* [12], which comes under the species of Rattus. *Mus musculus* is smilar to *Rattus norvegicus* and they have small ears and hairless tails[13]. *Spodoptera frugiperda* is an insect larva that demolishes grains and grasses which is found in America[14]. Fungi are unicellular, multicellular or syncytial spore-producing organisms such as yeast, molds and mushrooms. *Baker's yeast* is the general name for the yeast *Saccharomyces cerevisiae*[15] that translates the fermentable sugars which are available in the dough into carbon dioxide and ethanol[16]. *Escherichia coli* are a gram-negative bacterium prokaryote. Some *Escherichia coli* are not dangerous and various *Eschericia coli* cause food toxicity in humans[17]. This importance of this study is to support effective computer based drug design.

MATERIALS AND METHODS

Metabolic disease is a disorder that affects the function of the metabolic action of the body,

ensuing in loss of control of homeostasis[18]. In the package, diabetes disease is considered and the system provides Protein Ligand interaction, 2-Dimension View, 3-Dimension View, Individual Ligand Binding Model 3-Dimension View, amino acid residues compositions in active site, average tightness and co-crystal contact percentage of binding site and cross link of other Databases. West Brook *et al*., (2005) has developed PDB in XML named PDBML[19].Data conversion from PDB to XML is carried out via PDB Exchange Dictionary. Since the PDBML require hierarchical structure characteristics, there is a requirement for a database which translates the PDB data into XML format through hierarchy[19]. Present package also provides the XML translation of ligand details. The ligand binds and forms a complex with a biomolecule. Characteristics of drug molecules, such as solubility, dissolution rate and stability are referred as physico chemical properties[20]. This system provides protein ligand interaction for diabetes disease[21].

(i) Data Extraction

In house tools are developed using java to extract and separate the protein and ligand atomic coordinates in XML format and the relational database for the diabetes disease. The Extractor has the capacity to parse any number of PDB files using FileDialog Class and StringTokenizer Class. ANHetatm, ANatm, ANhetnam, ANligand, ANtitle are the tables created in the system. ANHetatm table contains ligand details while protein informations are maintained in ANatm table. ANHetnam table has ligand names from protein data bank and ANTitle contains . For defining the binding site, Soga *et al*., (2007) used a cutoff of 4.5 Å[22], Malik *et al*., (2007) took a cutoff of 3.5 Å[23] and Reddy *et al*., (2008) applied a cutoff of 4Å in their graph[1]. With a broad spectrum, the present study has the cutoff distance of 6 Å for the statistical analysis of ligand interaction and accordingly. ANligand table is created to increase the performance of the system. A user interface is developed in the in-house tool to create table and to insert information into the table from the PDB. XML properties are taken using the column oriented PDB format. Marvin beans (Version 5.0.6.1) is used to generate mol files for the ligands and the Mol files are given as input to Chem Sketch (Version 11.0) to calculate the following physico chemical properties: molecular formula, formula weight, composition, molar refractivity, molar volume, parachor, index of refraction, surface tension, density, dielectric constant, polarizability, monoisotopic mass, nominal mass and average mass. The inhouse tool also calculates properties of the ligand, namely, the fraction of contact (*f*) and the average tightness (*g*) of the ligand. Fraction of contact (*f*) =*N*a/*N*; average tightness (*g*) =*N*p/*N*a; where, *N*p = total number of protein atoms in the binding environment, *N*= total number of ligand atoms, *N*a = total number of ligand atoms in the binding environment.

(ii) System Architecture

The system has been developed as a three tiered (Fig 1) web based application.

System Architecture

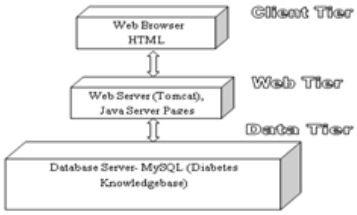


Figure 1
shows the system architecture of Diabetes Disease Related Studies

1. Entire diabetes disease ligand informations are presented in a relational database implementation in MySQL.
2. Web browser is acting as a client, where as HTML is serving as a front end.
3. Tomcat6.0 is used as a web server and the processing logic happens in the Tomcat application server and it acts as a middle tier. Java Server Pages creates the dynamic web content.

Chime plug-in (www.mdli.com) is used for 3- Dimensional Structure display.

(iii) Data Access

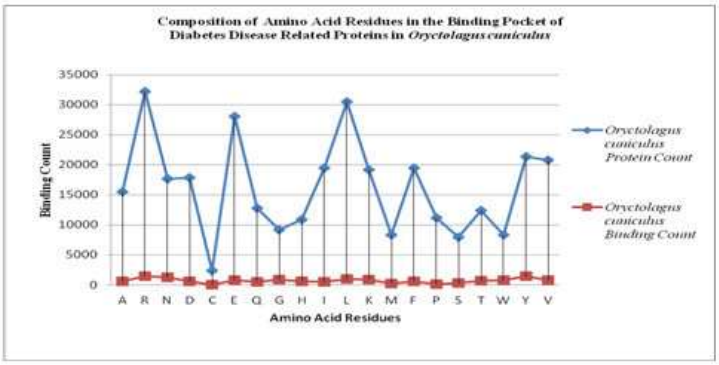
Protein-Navigation screen provides the PDB ID, title and the state of the protein whether mutated or native according to the chemical name specified by the user. XML translations of the system have the followin options: 1.Ligand coordinates in XML. 2. Chemical formula and name of the ligands in XML. Next, PDB screen lists the links of all the PDB files which have ligand interaction for the specified requirements (*Oryctolagus cuniculus, Sus scrofa*, *Homo sapien, Rattus norvegicus, Mus musculus, Spodoptera frugiperda, Saccharomyces cerevisiae and Escherichia coli*).“Binding pocket of individual protein” of the system provides the following details related to the ligand interaction for the selected protein within the cut off range specified by the user: Ligand details, Protein information and Protein-Ligand distance value. The user can specify the cut off value between 0Å to 6Å. “Binding site cluster” screen tabulates the 3-dimensional view of the atomic interactions of all the binding pockets for the entire specified category within 0Å to 6Å. 2-Dimensional and 3-Dimensional ligand structure views are displayed. In addition to this, the 3-Dimension views of the binding sites of all the models for a particular ligand are demonstrated. Individual Model Binding Site and clustered view of a ligand of 3-Dimension Structure is also displayed. Entry statistics is also provided. The following physico chemical properties: molecular formula, formula weight, composition, molar refractivity, molar volume, parachor, index of refraction, surface tension, density, dielectric constant, polarizability, monoisotopic mass, nominal mass, average mass and smiles notations are displayed for the ligands which are available in the user specified PDB ID [24]. The following methodology is carried out to favor of results and discussion. The protein data bank data are stored in structured format and retrieved from the relational database. The 2-Dimension, 3-Dimension, XML and text data are displayed in the web site (http://www.ponmary.com/LandingPage/dia/Diamain.html). The composition of amino acid residues are displayed in graph for every organism by taking the x-axis as amino acid

residues and y axis as binding counts which were taken from protein ligand interaction information (Graph 1).The protein count and binding count of amino acid residues are shown. Co-crystal contact percentages are depicted in the graph for every organism by taking the x-axis as co-crystal contact percentage and y-axis as binding count (Graph 2). Average Tightness was given in number for every organism in Table 1.

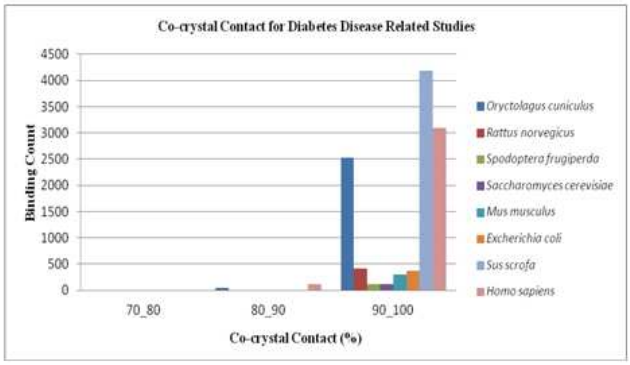
RESULTS AND DISCUSSIONS

1. Rabbit (Oryctolagus cuniculus)

Graph 1
Composition of Amino Acid Residues in Oryctolagus cuniculus



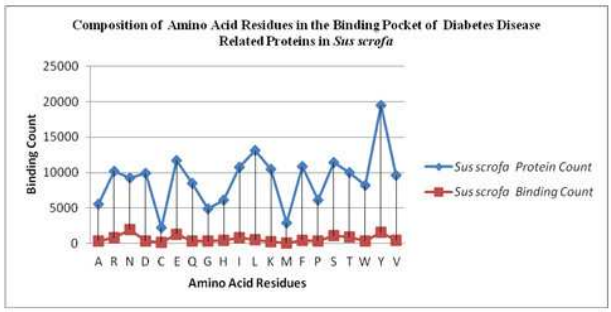
Graph 2
Co-Crystal Contact Percentage for Diabetes Disease Related Studies



In *Oryctolagus cuniculus*, 48 PDB records are discovered with protein ligand interaction; however, 182 ligand models are exposed for Diabetes Disease[25] Related Protein-Ligand Studies in the *Oryctolagus cuniculus* PDB records. 66,626 protein ligand interactions comprise the binding environment. Polar residues (Fig 2) Arginine (1464), Asparagine (1302), Tyrosine (1407) and non polar residue Leucine (965) are the predominantly occurring residues in binding environment (Graph 1). Totally there are 1,956 and 543 ligand binding counts afford for average tightness in the range between 4 to 6 and 6 to 8 respectively in the binding pocket (Graph 2). Co-crystal contacts of Diabetes Disease Related Studies in the *Oryctolagus cuniculus* (Table 1) binding pockets are notably well established (above 90%).

2. *Pig (Sus scrofa)*

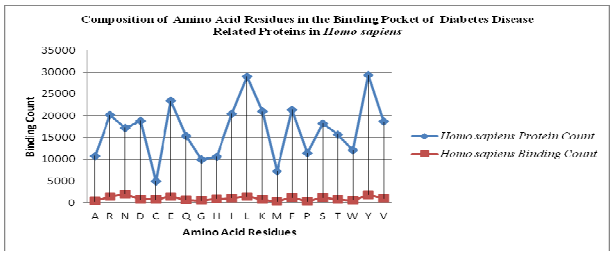
Graph 3
Composition of Amino Acid Residues in Sus scrofa



In *Sus scrofa*, 13 PDB Records are noted with protein ligand interaction. However, 111 ligand models are presented for Diabetes Disease Related Binding Studies in the *Sus scrofa* PDB Records. 65,661 protein ligand interactions are in the active site. The rate of occurance of polar residues asparagine (1,996), glutamic acid (1,331), tyrosine (1,552), serine (1,089), threonine (870) and non polar residue isoleucine (844) are high in active site (Graph 3). Totally there are 3890, 227 and 68 ligand atoms are bestowed for the average tightness of 2 to 4, 4 to 6 and 8 to 10 for the Diabetes Disease Related Studies in the *Sus scrofa* binding pocket (Graph 2). Co-crystal contacts of Diabetes Disease Related Studies in the Sus scrofa active sites are high (above 90%) (Table 1).

3. *Homo sapiens*

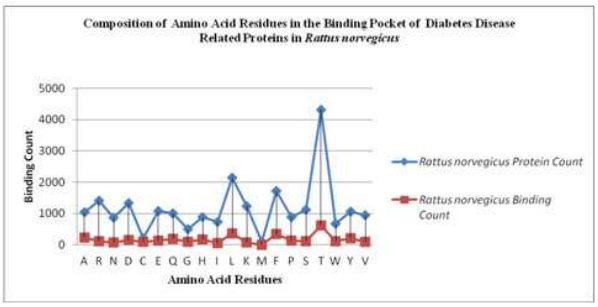
Graph 4
Composition of Amino Acid Residues in Homo sapiens



In *Homo sapiens*, 64 PDB Records are obtained with protein ligand interaction. However, 182 ligand models are found for Diabetes Disease Related Protein-Ligand Studies in the *Homo sapiens* PDB Records. 92,338 protein ligand interactions are in the active site. Frequent binding site non polar residues Leucine (1432) and polar residues Arginine (1450), Asparagine (1930), Glutamic acid (1472) and Tyrosine (1854) are the predominantly occurring residues in binding environment (Graph 4). It is noted that 3545, 959 and 208 are the ligand binding counts placed in the interval of 2 to 4, 4 to 6 and 6 to 8 respectively for the average tightness (Graph 2). A bifurcation was monitored for the co-crystal contact of Diabetes Disease Related Studies in the *Homo sapiens* ligand complex leads to above 80% (Table 1).

4. Rat (Rattus norvegicus)

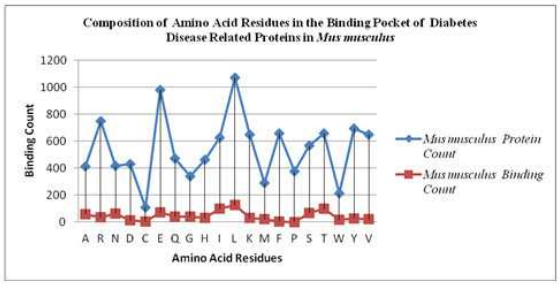
Graph 5
Composition of Amino Acid Residues in Rattus norvegicus



In *Rattus norvegicus*, three PDB Records are obtained with protein ligand interaction. However, 42 ligand models are found for Diabetes Disease Related Protein-Ligand Studies in the *Rattus norvegicus* PDB Records. 12,198 protein ligand interactions are in the active site. The frequency of occurrence of polar residue Threonine (620) and non polar residues Phenylalanine (338), Leucine (370) are found high in binding environment (Graph 5). Totally there are 323, 96 ligand binding counts afford for average tightness in the range between 6 to 8 and 8 to 10 (Graph 2). It is revealed that in the entire pocket, the ligand contact percentage is high (above 90%) (Table 1).

5. Mouse (Mus musculus)

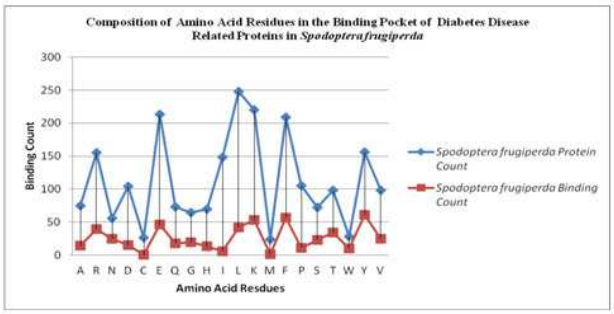
Graph 6
Composition of Amino Acid Residues in *Mus musculus*



In *Mus musculus*, three PDB Records are found with protein ligand interaction. However, 12 ligand models are available for Diabetes Disease Related Binding Environment Studies in the *Mus musculus* PDB Records. 6,191 protein ligand interactions are in the binding environment. Polar residue threonine (98) and non polar residues isoleucine (97) and leucine (125) are the predominantly occurring residues in binding environment (Graph 6). Totally there are 84 and 196 ligand binding counts afford for average tightness in the range between 0 to 2 and 2 to 4 respectively (Graph 2). It is observed that in the entire pocket, the ligand contact percentage is maximum (100%) (Table 1).

6. Fallarmy Worm (Spodoptera frugiperda)

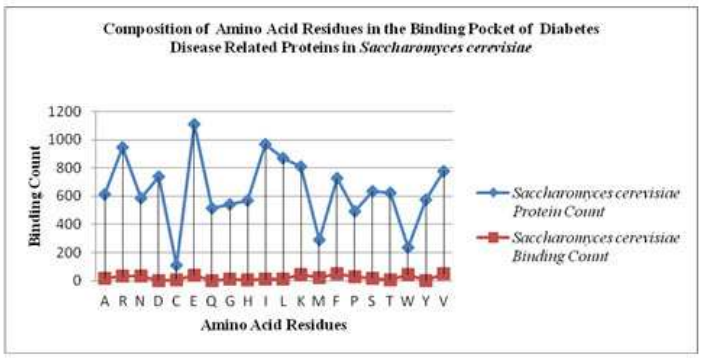
Graph 7
Composition of Amino Acid Residues in Spodoptera frugiperda



For *Spodoptera frugiperda*, a single PDB Record is found with protein ligand interaction. However, 15 ligand models are available for Diabetes Disease Related Protein-Ligand Studies in the *Spodoptera frugiperda* PDB Records. 2307 protein ligand interactions are observed in the binding site. The frequency of nonpolar residue Phenylalanine (57) and polar residues Lysine (53), Tyrosine (61) are high in the active site (Graph 7). Average tightness for Diabetes Disease Related Studies in the *Spodoptera frugiperda* ligand binding environment is calculated as 4.5 (Graph 2). Co-crystal contacts of Diabetes Disease Related Studies in the Spodoptera frugiperda active sites are maximum (100%) (Table 1)

7. Baker's Yeast (Saccharomyces cerevisiae)

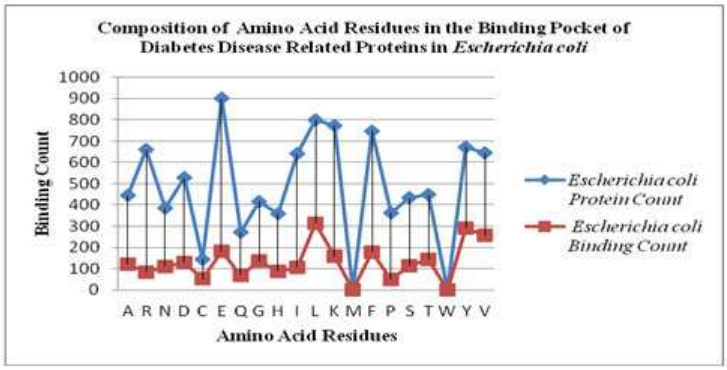
Graph 8
Composition of Amino Acid Residues in Saccharomyces cerevisiae



In *Saccharomyces cerevisiae, a* single PDB Record is found with protein ligand interaction. However, three ligand models are available for Diabetes Disease Related Studies in the *Saccharomyces cerevisiae* PDB Records. 2,307 protein ligand interactions are in the binding environment. The frequency of polar residue Lysine (45) and non polar residues Tryptophan (42), Valine (48), Phenylalanine (51) are the most populated polar residues in binding environment (Graph 8). It is noted that 111 are the ligand binding counts are placed for the average tightness of (2 to 4) (Graph 2). It is observed that in the entire pocket, the ligand contact percentage is maximum (100%) (Table 1).

8. Bacteria (Escherichia coli)

Graph 9
Composition of Amino Acid Residues in Escherichia coli



In *Escherichia coli*, three PDB Records are noted with protein ligand interaction. However, 32 ligand models are presented for Diabetes Disease Related Studies in the *Escherichia coli* PDB Records. 10,880 protein ligand interactions are in the active site. The rates of occurance of polar residues tyrosine (292) and non polar residues leucine (314), valine (255) are high in active site (Graph 9). 372 ligand atoms are bestowed for the average tightness of (6 to 8) for Diabetes Disease Related Studies in the *Escherichia coli* binding pocket (Graph 2). Co-crystal contacts of Diabetes Disease Related Studies in the *Escherichia coli* active sites are maximum (100%) (Table 1).

Table 1
Average Tightness for Diabetes Disease Related Studies

Average Tightness	0_2	2_4	4_6	6_8	8_10
Oryctolagus cuniculus	-	40	1956	543	-
Rattus norvegicus	-	-	-	323	96
Spodoptera frugiperda	-	-	112	-	-
Saccharomyces cerevisiae	-	111	-	-	-
Mus musculus	84	196	24	-	-
Escherichia coli	-	-	-	372	-
Sus scrofa	-	3890	227	-	68
Homo sapiens	-	3545	959	208	163

CONCLUSION

The present study integrates the protein ligand interaction data for diabetes disease related studies using organism based classification. The current work may be useful for the researcher who aspires for effective drug design against diabetes disease.

REFERENCES

1. Reddy AS, Amarnath HS, Bapi RS, Sastry GM and Sastry GN. Protein ligand interaction database (PLID). Comput Bio Chem, 32: 387–390, (2008).
2. Liang MH and Shadick N. Feasibility and Utility of Adding Disease-Specific Outcome Measures to Administrative Databases to improve Disease Management. J Qual Improv, 127(8): 739-742, (1997).
3. Gupta R, Deedwania PC, Gupta A, Rastogi S, Panwar RB and Kothari K. Prevalence of metabolic syndrome in an Indian urban population. Int J Cardiol, 97(2): 257-261, (2004).
4. King H and Rewers M. Global estimates for prevalence of diabetes mellitus and impaired glucose tolerance in adults. Diabetes Care, 16:157-177, (1993).
5. Mohan V, Sandeep S, Deepa R, Shah B and Varghese C. Epidemiology of type 2 diabetes: Inidan scenario. Indian J Med Res, 125: 217-230, (2007).
6. Babu PA, Suneetha G, Boddepalli R, Lakshmi VV, Rani TS, Rambabu Y and Srinivas K, A database of 389 medicinal plants for diabetes. Bioinformation, 1(4): 130-131, (2006).
7. Middha SK, Mittal Y, Usha T, Kumar D. Srinivasan R, Vashisth L, Bhattacahrgae B and Nagaveni MB. Phyto-mellitus: A phyto-chemical database for diabetes. Bioinformation, 4(2): 78-79, (2009).
8. Whittaker and Robert H. On the broad classification of organisms. Quart Rev Biol, 34: 210-66, (1959).
9. Daly JC, Effects of Social Organization and Environmental Diversity on Determining the Genetic Structure of a Population of the Wild Rabbit, Oryctolagus cuniculus. Evolution, 35(4): 689-706, (1981).
10. Grossi SM, Lui JF, Garcia JE and Meirelles FV. Genetic diversity in wild (*Sus scrofa scrofa*) and domestic (*Sus scrofa domestica*) pigs and their hybrids based on polymorphism of a fragment of the D-loop region in the mitochondrial DNA. Genet Mol Re, 5-4: 564-568, (2006).
11. Leakey REF, Butzer KW and Day MH. Early Homo sapiens remains from the Omo River region of Southwest Ethiopia, Nature, 222: 1132-1138, (1969).
12. Abdelkrim J, Pascal M, Calmet C and Samadi S. Importance of Assessing Population Genetic Structure before Eradication of Invasive Species: Examples from insular Norway Rat Populations, Con Biol, 19(5): 1509-1518, (2005).
13. Sage R, Atchley W and Capanna E. House mice as models in systematic biology. Syst Biol, 42(4): 523-561, (1993).
14. Ashley TR, Wiseman BR, Davis FM and Andrews KL. The fall armyworm: a bibliography. Flor Entomol, 1989; 72: 152-202.
15. Moore RT. Taxonomic proposals for the classification of marine yeasts and other yeast-like fungi including the smuts, Bot Mar, 23: 361–73, (1980).
16. Aa E, Townsend JP, Adams RI, Nielsen KM and Taylor JW. Population structure and gene evolution in *Saccharomyces cerevisiae*, FEMS, 702–715, (2006).
17. Hudault S, Guignot J and Servin AL. *Escherichia coli* strains colonising the gastrointestinal tract protect germfree mice against *Salmonella typhimurium* infection, Gut, 49(1): 47–55, (2001).
18. Shetty, KT. Metabolomics: impact on diagnosis and monitoring of 'inborn'/'acquired' metabolic disorders, Indian J Clin Biochem, 22(2): 3-5, (2007).
19. Westbrook J, Ito N, Nakamura H, Henrick K and Berman HM. PDBML: the representation of archival macromolecular structure data in XML, Bioinformatics, 21(7): 988-992, (2005).
20. Feng Z, Chen Li, Maddula H, Akcan O, Oughtred R, Berman HM and Westbrook J. Ligand Depot: a data warehouse for

ligands bound to macromolecules, Bioinformatics, 20: 2153–2155, (2004).
21. Latha DPP and Sharmila DJS. Ligand binding site analysis package for AB5 toxins and Neurotoxins Related Studies, Online journal of computer technology, MM Publication, 2(1): 12-19, (2010).
22. Soga *S*, Shirai H, Kobori M and Hirayama N. Use of amino acid composition to predict ligand-binding sites. J Chem Inf Model, 47: 400–406, (2007).
23. Malik A and Ahmad S. Sequence and structural features of a neural network, BMC Struct Biol*,* 7: 1–14, (2007).
24. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN and Bourne PE, The Protein Data Bank, Nucl Acids Res, 28 (1): 235-242, (2000).
25. Rose ML and Easwaran M, Insilico study and drug target for type II diabetes from a natural compound, International journal of Pharma and Bio sciences, 3(4):795-801 (2012).