Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

Hotspot Identification System for identification of core residues in Diabetic Proteins

Dr. P. V. S. Lakshmi Jagaamba[1]

[1] Jawarlal Nehru Technological University- Kakinada

Received: 12 April 2012 Accepted: 1 May 2012 Published: 15 May 2012

Abstract

Data on genome structural and functional features for various organisms are being accumulated and analyzed in laboratories all over the world. The data are stored and analyzed on a large variety of expert systems. The public access to most of these data offers to scientists around the world an unprecedented chance to data mine and explores in depth this extraordinary information repository, trying to convert data into knowledge. The DNA and RNA molecules are symbolic sequences of amino acids in the corresponding proteins has definite advantages in what concerns storage, search, and retrieval of genomic information. In this study an attempt is made to develop an algorithm for aligning multiple DNA / protein sequences. In this process hotspots are located in a protein sequence using the multiple sequence alignment.

Index terms— Symbolic sequences, DNA, RNA, Protein sequence, Multiple Sequence alignment.

1 INTRODUCTION

n Bioinformatics, sequence alignment is a prominent method of arranging the sequences of DNA, RNA or protein to identify regions of similarity. Similarity may be functional, structural or evolutionary relationships between the sequences. Aligned sequences of nucleotide, amino acid residues are represented in a row form of a matrix. Identical or similar characters are aligned in successive columns by inserting gaps between the residues. There is a storm of revolution in the areas of Genomics and Bioinformatics in recent years. Bioinformatics is widely used for computational usage and processing of molecular and genetic data. The biologists considered Bioinformatics for the use of computational methods and tools to handle large amounts of data and make the data more understandable and useful. On the other hand, others view Bioinformatics as an area of developing algorithms and tools and to use mathematical and computational approaches to address theoretical and experimental questions in biology. As genomic data is rapidly exposed to increasing research, knowledge based expert system is becoming indispensable for the emerging studies in Bioinformatics. Hence validation and analysis of mass experimental and predicted data to identify relevant biological patterns and to extract the hidden knowledge are becoming important.

In recent years, semantic web based methods are introduced and are designed in such a way that meaning is added to the raw data by using formal descriptions of concepts, terms and relationships encoded within the data. To analyze and understand the data, today's information rich environment developed and designed a number of software tools. These tools provide powerful computational platforms for performing Insilco experiments (8). As there is much complexity and diversity in the analysis of tools, the need is for an intelligent computer system for automated processing. Present researches in Bioinformatics need the use of integrated expert systems to extract more efficient knowledge. In the biological process proteins undergo some interactions. These protein-protein interactions are mediated molecular mechanisms. During this interaction, a small set of residues play a critical role. These residues are called hot spots. The ability to identify the hot spots from sequence accurately and efficiently as expert system that enables and analysis of protein-protein interaction hot spots. This analysis may

benefit function prediction and drug development. At present there is a strong need for methods to obtain an accurate description of protein interfaces. Many scientists try to extract protein interaction information from protein data bank.

Alignment Methods Used: In general the hot spots are identified as active sites in protein structures as binding is done using structures. The researcher tried to find the hotspots in protein sequence rather than structure. In this process, taking into consideration the evolutionary history, the families of sequences are aligned using multiple sequence alignment.

In the process of alignment two methods are used Standard method using dynamic programming and A proposed alternative-MSAPSO (Multiple Sequence alignment using Particle Swarm Optimization) method in which alignment is performed using PSO technique. A comparison of these two methods also made. If the sequences are very short or similar they can be aligned by hand. But lengthy and highly variable numerous sequences cannot be aligned manually. To produce high quality sequence alignments, construction of algorithms and application of human knowledge are necessary. Computational approaches to sequence alignments are of two types-Global alignments and local alignments. Global alignment is the alignment to span the entire length of sequences whereas local alignments identify regions of similarity within the long sequences. Then mature mRNA is used as a template for protein synthesis, which is known as translation onto a ribosome. Then read three nucleotides at a time by matching each codon to its base pairing anticodon to form transfer RNA (tRNA). Then tRNA recognizes the amino acid corresponding to the codon. The sequence thus obtained is protein sequence.

The amino acids in a protein sequence are shown in the following table.

The overall structure and function of a protein is determined by the amino sequence. Most proteins fold into 3-dimensional structures and its shape is known as its native state. There are four levels in a protein structure.

? G GLY Glycine W TRP Tryptopham A ALA Alanine Y TYR Threonine V VAL Valine N ASN Asparagine L LEU Leucine Q GLN Glutamine I ILE Lsoleucnie D ASP Asparatic Acid F PHE Phenylalanine E GLU Glutamic Acid P PRO Proline K LYS Lysine S SER Serine R ARG Arginine T THR Threonine H HIS Histidine C CYS Cyctenie M MET Methinine

? Enzymes: Enzyme is one of the functions of the protein which carries out most of the reactions involved in metabolic activities. Enzymes are proteins that increase the rate of chemical reaction. Adding or participation of the substance called catalyst does the change in the rate of chemical reaction. Catalysts that speed the reaction are called positive catalysts. Substances that interact with catalysts to slow the reaction are called inhibitors (or negative catalysts). Substances that increase the activity of catalysts are called promoters, and substances that deactivate catalysts are called catalytic poisons.

helix, beta sheet and turns.

? Active Sites in Proteins: An Active site is a part of an enzyme where substrates bind and undergo a chemical reaction. The substrate which is a molecule binds with the enzyme active site and then an enzymesubstrate complex is formed. It is then transformed into one or more products, which are released from the active site. The active site is now free to accept another substrate molecule. In the case of more than one substrate, these may bind in a particular order to the active site, before reacting together to produce products. A product is something ”manufactured” by an enzyme from its substrate. For example the products of Lactase are Galactose and Glucose, which are produced from the substrate Lactose. Two models-the lock and key model and induced fit model are the two models proposed to describe how the enzymes work. In the lock and key model the active site perfectly fits for a specific substrate. If once the substrate binds to the enzyme no further modification is necessary. On the other hand in the induced fit model, an active site is more flexible and the presence of certain residues (amino acids) of the active site the enzyme is encouraged to locate the correct substrate. Once the substrate is gone conformational changes may occur. Hot spots are a set of residues recognized or bound in the process of interacting with other proteins. These are the residues in the active site.

2 II.

3 RESULTS & DISCUSSION

Insulin is one of the important protein sequences which cause diabetes. So we tried to identify the hotspots in this protein sequence using the following methodology.

?

4 CONCLUSION

Hot spots are of residues comprising only a small fraction of interfaces of the binding energy. We present a new and efficient method to determine computational hot spots based on pair wiser technique using potentials and solvent accessibility of interface residues. The conservation does not have significant effect in hot spot prediction as a single feature. Residue occlusions from solvent and pair wise potentials are found to be the main discriminative features in hot spot prediction. The predicted hotspots are observed to match with the experimental hot spots with an accuracy of 70%. The solvent is a necessary factor to define a hot spot, but not sufficient itself. This

1

Figure 1: 1 .

```
AAB24882        TYHMCQFHCRYVNNHSGEKLYECNERSKAFSCPSHLQCHKRRQIGEKTHEHNQCGKAFPT 60
AAB24881        --------------------YECNQCGKAFAQHSSLKCHYRTHIGEKPYECNQCGKAFSK 40
                                     ****: .***:  * *:** * :****.:* *******..

AAB24882        PSHLQYHERTHTGEKPYECHQCGQAFKKCSLLQRHKRTHTGEKPYE-CNQCGKAFAQ- 116
AAB24881        HSHLQCHKRTHTGEKPYECNQCGKAFSQHGLLQRHKRTHTGEKPYMNVINMVKPLHNS 98
                 **** *:***********:***:**.: .***************    :  *.: :
```

Figure 2:

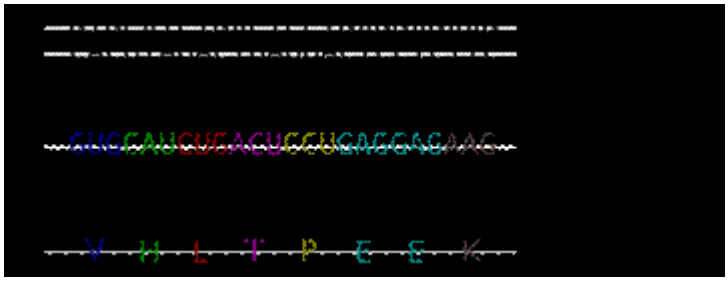

Figure 3: Global

1

One Letter	Three Letter	Full Name	One Letter	Three Letter	Full Name

Figure 4: Table 1

		1ai0	J	1	30
	13	1ai0	K	1	21
	14	1ai0	L	1	30
	15	1aiy	A	1	21
	16	1aiy	B	1	30
III.	17	1aiy	C	1	21
	18	1aiy	D	1	30
	19	1aiy	E	1	21
	20	1aiy	F	1	30
	21	1aiy	G	1	21
	22	1aiy	H	1	30
	23	1aiy	I	1	21
	24	1aiy	J	1	30
	25	1aiy	K	1	21

? Then identify the protein-protein interactions for each of these protein structures shown in the following table.

SNO	SNO	PDB Code	PDB Code	Chain	Chain	First PDB residue	Last PDB residue	Ch
	1	1a7f		A		1	21	
1	2	1a7f	1a7f	B	A	1	B	29
2	3	1ai0	1ai0	A	A	1	B	21
3	4	1ai0	1ai0	B	B	1	D	30
4	5	1ai0	1ai0	C	C	1	D	21
5	6	1ai0	1ai0	D	E	1	F	30
6	7	1ai0	1ai0	E	F	1	H	21
7	8	1ai0	1ai0	F	G	1	H	30
8	9	1ai0	1ai0	G	I	1	J	21
9	10	1ai0	1ai0	H	J	1	L	30
10	11	1ai0	1ai0	I	K	1	L	21
11	12	1ai0	1aiy	J	A	1	B	30
12		1aiy		B		D		

Figure 5:

is also compared our methods and other hot spot prediction methods. Our method outperforms them with its high performance expert system. [1] [2]

[1] January 2012© 2012 Global Journals Inc. (US)

[2] © 2012 Global Journals Inc. (US) Global Journal of Computer Science and Technology Volume XII Issue I Version I

[Roos ()] 'Bioinformaticstrying to swim in a seaof data'. D S Roos . *Science* 2001. 291 p. . (Computational biology)

[Chao-Yie Yng and Wang ()] 'Computational Analysis of Protein Hotspots'. Shaomeng Chao-Yie Yng , Wang . *ACS Medicinal Chemistry Letters* 2010. 1 (3) p. .

[Hajduk ()] 'Druggability induces for protein targets derived from NMR-based Screening Data'. P Hajduk . *J Med. Chem* 2005. 48 p. .

[Engelmopre and Feigenbaum ()] *Expert Sstems and Artificial Intelligence*, Robert S Engelmopre , Edward Feigenbaum . 1993. WTEC Hyper Librarian.

[Miller ()] 'Expert System An Introduction'. Hintze Miller , B . *PC AI where Intelligent technology meets the real world* 1988. 2 (3) p. 26.

[Aniba and Thompson ()] *Knowledge Based Expert Systems in Bioinformatics*, Mohamed Radhouene Aniba , Julie D Thompson . 2010. Petric? Vizureanu. p. . (published in Expert Systems)

[Dobson ()] 'The nature and significance of protein folding'. C M Dobson . *Mechanisms of Protein Folding 2nd ed. Ed. RH Pain. Frontiers in Molecular Biology series*, (New York, NY) 2000. OxfordUniversity Press.