Available online at www.pelagiaresearchlibrary.com

Pelagia Research Library

Asian Journal of Plant Science and Research, 2012, 2 (3):290-298

ISSN : 2249-7412
CODEN (USA): AJPSKY

In-silico Identification of Anti-fertility Proteins based on Sequence and Structural Similarity

Shriddha Shukla[1], Savita Dixit[2] and K. R. Pardasani[3]

[1]*Department of Bioinformatics, Maulana Azad National Institute of Technology, Bhopal (M.P.)*
[2]*Department of Chemistry, Maulana Azad National Institute of Technology, Bhopal (M.P.)*
[3]*Department of Mathematics, Maulana Azad National Institute of Technology, Bhopal (M.P.)*

ABSTRACT

The hiking population is one of the major problems of concern its negative impact can be seen on welfare economics of developing and even developed nations. In order to abolish this hike there is need to synthesize effective anti-fertility compounds. In this paper the Ribosomal Inactivating Proteins of Momordica charantia are used showing anti-fertility property. Even if the studies conducted on phyto-proteins serves to produce compounds with novel structures, excellent activity, and negligible side effects, but still a very less number of phyto-proteins are identified along with their medicinal importance. In this paper an in-silico approach is described for modeling the three-dimensional structure of a protein from the homologous protein structures and their amino acid sequences. A method is developed for the simultaneous superimposition of several protein molecules and for the calculation of an 'average structure' or 'framework' by the RMSD values in order to obtain most similar structure of the phyto-proteins showing anti-fertility property. Consequently, we obtained nine phyto-proteins on the basis of the family of homologous proteins. This alignment provides a basis for model building the tertiary structure for the unknown phyto-proteins.

Keywords: Anti-fertility, Phyto-proteins, Phylogenetic analysis, Modeling and Superimposition.

INTRODUCTION

The overwhelming impact of population growth can be seen on the economy as well as the environment of a country consequently the declining natural resources creating severe environmental uncertainties. Due to advancement of science and technology there has been tremendous decline in death rate resulting in higher birth rate at present it has been expected the population to reach 9.2 billion by 2050. The graph given depicts the total world population from the year 1750-2050 in Fig.1 [1].

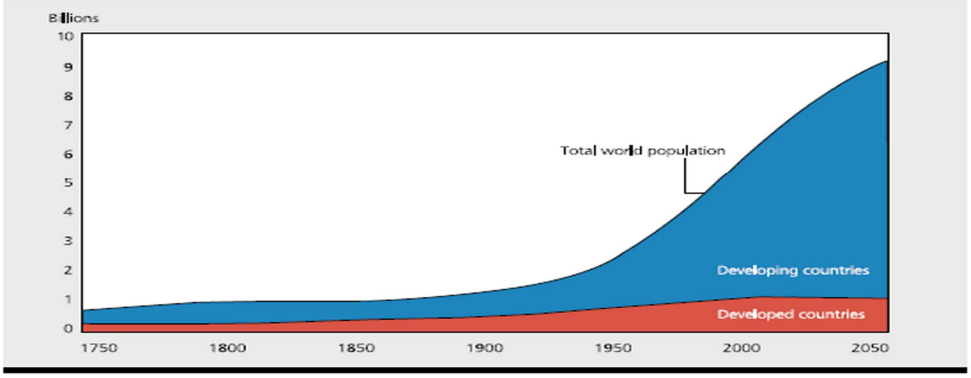

Fig.1 The graph depicts the rise in World Population from the year 1750-2050 in the developing and developed countries.

The negative impact of the growing population accelerates disease and malnutrition problems mainly in the rural areas of third world and so the poverty, declining natural resources, fossil fuel depletion, pollution and stunted economies of developing and even developed nations [2]. To be a prosperous nation there is a need to stabilize and reduce the population, falling birth rate may be an important factor in this context [3]. The World's birth rate i.e. births per 1,000 population has been depicted in the Fig.2.

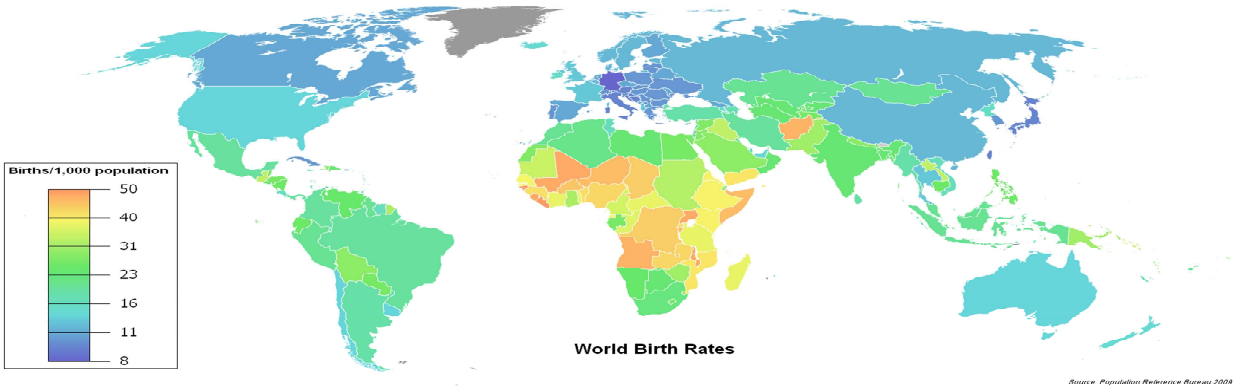

Fig.2 World Birth Rate(Birth/1000 Population).

As depicted in the above graph the major population explosion is in the third world countries of South Africa and many parts of Asian continent. Therefore, to have check on this hike in population there is a need to develop more effective anti-fertility drugs with negligible side effects which can provide less expensive alternatives to existing expensive drugs. Various synthetic drugs are available which are used for fertility regulation but due to their side effects their usage is suspicious even prohibited among common people. From the very ancient times the ayurvedic and unani medicines have been proven themselves most effective cure against any kind of ailment. More than 80% of the population rely on the medicines of plant origin as these are more effective and have negligible side effects as compared to the synthetic ones [4].The use of phyto-proteins as drug alternatives is rare and is a novel approach to cure several severe diseases. Since the RIP's are of natural origin and are mostly found in plants and from ancient times natural sources of drugs are more effective. RIP's have been proven to play a vital role in development of novel drugs as these are precursors for several synthetic drugs many of them act as lead compounds for several synthetic ones. The synthetic drugs are mostly of chemical origin and therefore are less preferred over those of natural origin as these are with negligible side effects and are also cost effective .There are several drugs of natural origin for instance anticancer drugs, Vincristine and Vinblastine complexes obtained from the Rosey Periwinkle (*Catharanthus rosea*) and many more drugs of natural origin are being effectively used for the treatment of many dreadful diseases [5].

The Ribosome Inactivating Proteins (RIP's) inactivates the eukaryotic ribosomes these are generally the plant proteins which depurinates the rRNA and therefore are the inhibitors for protein synthesis. The RIP's shows several pharmacological properties mainly include Anti-HIV, Antitumor and Abortifacient properties [6]. Ribosome Inactivating Proteins (RIP's) induces apoptosis by decreasing the action of anti-apoptotic factors they have antiviral and anti-parasitic properties and have proved to be a very effective drug against AIDS by acting directly on HIV infected cells by depurinating the RNA. The RIP's are better cure for certain allergies but are also having allergenic properties as they are raw eaten in the form of vegetables. RIP's as immunotoxins are potentially used to treat the tumor cells. The abortifacient property of RIP's has been reported in various plants species as they inhibit the protein synthesis [7]. Since the drugs of natural origin are more efficient and cost effective over the synthetic drugs. Therefore new drug alternatives from plants should be identified and designed in order to obtain drugs with negligible side effects. In this paper phyto-protein mainly the ribosome inactivating proteins (RIP's) are taken for analysis. RIP's of *Momordica charantia* (Bitter gourd) plant are taken namely Momordin-I, α-momocharin and β-momocharin showing the abortifacient property which has been proved experimentally [8]. The phylogenetic analysis was done and several homologous protein sequences were identified. Since the sequence similarity shows resemblance to functional similarity, and the amino acid sequences determines the protein three dimensional structures and the structural similarity between proteins is very good predictor of functional similarity. Therefore on structural analysis the protein structures identified showed anti-fertility properties similar to those of Momordica charantia. The structural analysis involves the comparative structure modeling the closely related proteins present in the same cluster i.e. the homologous (orthologues and paralogues) protein 3-D structures were modeled. These protein structures were then superimposed and the sequence alignments, structure alignments, PDB (Protein Data Bank) coordinates and RMSD statistics were generated [9, 11]. As a result of our analysis the structural homologues of *Momordica charantia* RIP's were identified which may also show the abortifacient property. On studying molecular insights of the three dimensional structures of proteins can help to study molecular mechanisms- such as site-directed mutagenesis, mapping of disease-related mutations, and the structure-based design of specific inhibitors and rational drug design approaches. For predicting the 3D structures of proteins the computational methods are of high degree of interest and are the focus of research and development activities. The prediction of the 3D structure of a protein from its amino acid sequence is always a problem to be resolved by identifying the novel protein structures [16].

MATERIALS AND METHODS

The present analysis involves the three ribosome inactivating proteins (RIP's) from the plant *Momordica charantia* (Bitter gourd, Balsam pear), namely Momordin I, α-momocharin, β-momocharin. The protein sequences were retrieved from Uniprot database and the 3-D structures were taken from RCSB Protein Data Bank (PDB) as shown in Table I.

Table I. Ribosomal Inactivating Proteins taken for study

Phyto-Protein	No. of Amino acid residues	PDB ID
Momordin I	264	1MOM
α-momocharin	263	1AHC
β-momocharin	286	1CF5

Further with the help of various Bioinformatics software tools the insilico analysis was done using the following softwares: BLASTp, Clustal-X, Geneious, Swiss-Model, Superpose etc.

The methodology involves following steps:

Step 2.1. Sequence search across the Uniprot Knowledge Base Database:
The sequences of the ribosome inactivating proteins namely: Momordin-I, α-momocharin and β-momocharin were retrieved in the FASTA format from the Uniprot Knowledgebase (UniProtKB).

Step 2.2. The similarity search done across Ribosome inactivating protein (RIP) sequences of Momordica charantia:
The similarity search was done with the help of Protein BLAST (Basic Local Alignment Search Tool) software available at NCBI (National Centre for Biotechnological Information) blastp which involves the pairwise alignment

algorithm. The sequences showing more than 80% similarity were taken for further analysis. There were fifteen such sequences which were 80% similar to the three RIP's of *Momordica charantia*.

Step 2.3. Performing the Multiple Squence Alignment for the similar RIP sequences:
CLUSTAL-X performs a global multiple sequence alignment and the similar sequences were aligned.

Step 2.4. Generating Phylogenetic tree by using UPGMA algorithm:
In this step after aligning all the protein sequences three different Phylogenetic tree were generated for three different RIP's of *Momordica charantia*: Tree-I for Momordin-I, Tree-II for α-momocharin and Tree-III for β-momocharin. Using the Genious Basic 5.4.2 software the distance matrix was calculated and the orthologues and paralogues were identified respectively.

Step 2.5. Retrieval of protein 3-D structures from Protein Data Bank (PDB):
The three dimensional protein structures were retrieved from PDB. The RIP's found on PDB were namely, β-luffin, Bryodin I, Cucurmosin, Luffaculin I, Trichosanthin and the rest protein structure were modeled which were not available at PDB.

Step 2.6. Modeling protein structure for the novel Ribosome Inactivating Proteins structure:
The protein sequences were then searched across the UniProtKB were modeled by using SWISS-MODEL (http://swissmodel.expasy.org). It is a server for automated comparative modeling of three-dimensional (3D) protein structures which is used for modeling the proteins whose structures were not available at PDB taken for analysis.

Step 2.7. Structure alignment of the closely related sequences:
To have structural similarity it has been found for protein structure after superimposition should have the RMSD values less than or equal to 2 Å. Using SUPERPOSE web server the proteins 3-D structures were superposed against the *Momordica charantia* RIP's and the proteins with RMSD 2 Å or less than that were predicted as the proteins homologous to the Momordin I, α-momocharin and β-momocharin showing anti-fertility property.

RESULTS

As a result of similarity search fifteen RIP sequences were found similar to the three RIP's of *Momordica charantia*, Momordin- I, α-momocharin, β-momocharin. A total of eleven plant species were taken for the analysis from them eighteen phyto-proteins including those of *Momordica charantia*. The fifteen phyto-proteins obtained as a result of similarity search are as under in Table II.

Table II. Ribosome Inactivating Proteins from various plant species obtained as a result of similarity search Consequently, on Phylogenetic analysis three trees were generated for Momordin-I, α-momocharin and β-momocharin respectively shown as under:

Phyo-proteins	Plant species
B-luffin	*Luffa cylindrica* (Smooth loofah, Sponge gourd)
Bryodin I	*Bryonia dioica* (Red bryony)
Cucurmosin	*Cucurbita moschata* (Winter crookneck squash, Cucurbita pepo var. moschata)
Gynostemmin	*Gynostemma pentaphyllum* (Jiaogulan)
Luffaculin I	*Luffa acutangula* (Ridged gourd, Cucumis acutangulus)
Karasurin-H	*Trichosanthes kirilowii* (Chinese snake gourd, Chinese cucumber)
Trichobakin	*Trichosanthes kirilowii*
Trichomislin	*Trichosanthes kirilowii*
Trichosanthin	*Trichosanthes kirilowii*
Type 2 Ribosome-Inactivating protein	*Camellia sinensis* (Tea)
Cinnamomin I	*Cinnamomum camphora* (Camphor tree)
Nigrin I	*Sambucus nigra* (European elder)
Nigrin-b	*Sambucus nigra* (European elder)
Ribosome-Inactivating protein	*Sambucus nigra* (European elder)
Preproricin	*Ricinus communis* (Castor bean)

Tree I:

In Fig.3 from Tree-I it has been found that the phyto proteins β-luffin and Luffaculin-I are found to be paralogs to Momordin-I i.e. the genes related by duplication within a genome; may evolve same or new functions in the course of evolution. As these are more closely related as depicted in the tree and can be shown as per the distances between these two from Momordin-I is very less. The β-luffin and Luffaculin-I are closely related to Momordin-I. While other species are also found to be distantly related in the same cluster to Momordin-I i.e. those RIP's (Bryodin-I, Karasurin-H, Trichobakin, Trichomislin, Trichosanthin, Curcumosin and Gynostemmin) are orthologs to Momordin-I. The orthologs are the genes in different species that evolved from a common ancestral gene by speciation, orthologs retain the same function in the course of evolution.

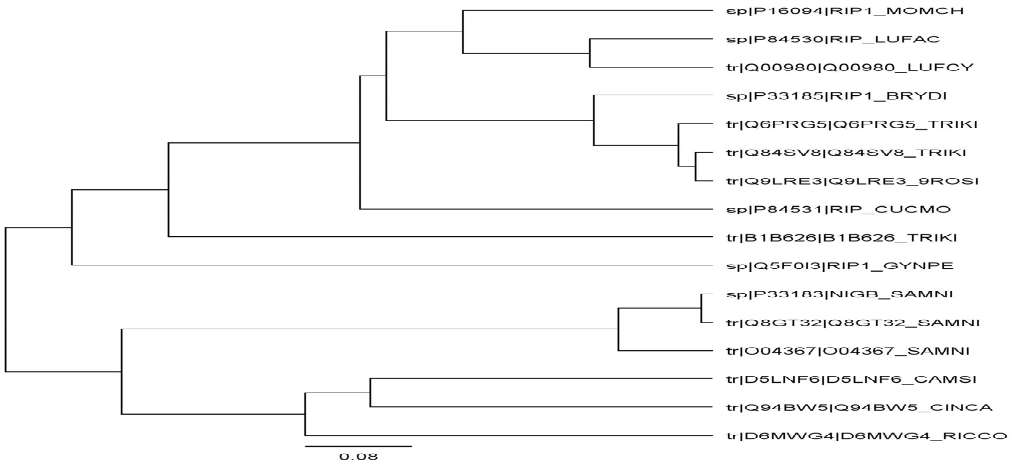

Fig.3 Tree-I depicts evolutionary relationships of Momordin-I (Ribosome Inactivating Protein of *Momordica charantia*) with other Ribosome Inactivating Proteins.

In the above tree the β-luffin and luffaculin RIP's are the paralogs of Momordin-I.

Tree II:

As shown in Fig.4 in Tree-II both the β-luffin and luffaculin are paralogs to α-momocharin. Here also the remaining RIP's (Bryodin-I, Karasurin-H, Trichobakin, Trichomislin, Trichosanthin,Curcumosin and Gynostemmin) in the cluster are orthologs to α-momocharin.

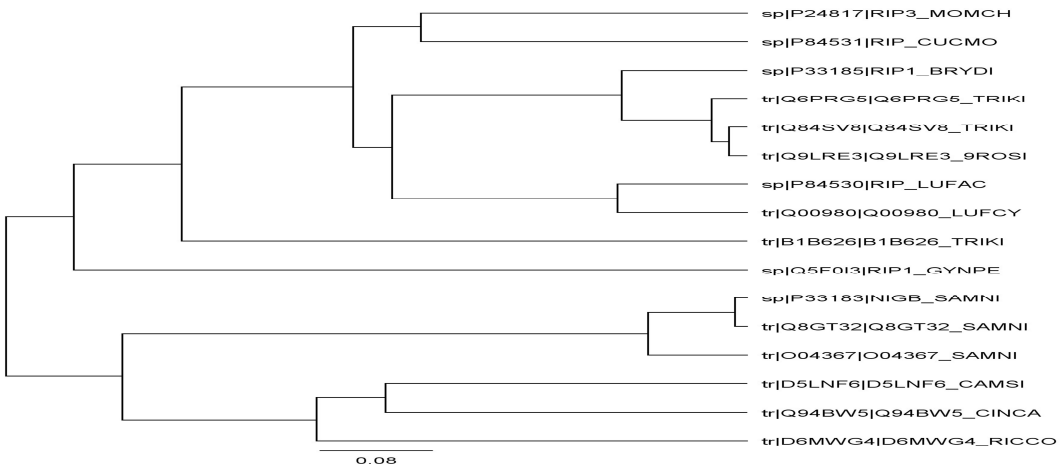

Fig.4 Tree-II depicts evolutionary relationships of α-momocharin (Ribosome Inactivating Protein of *Momordica charantia*) with other Ribosome Inactivating Proteins.

Tree III:

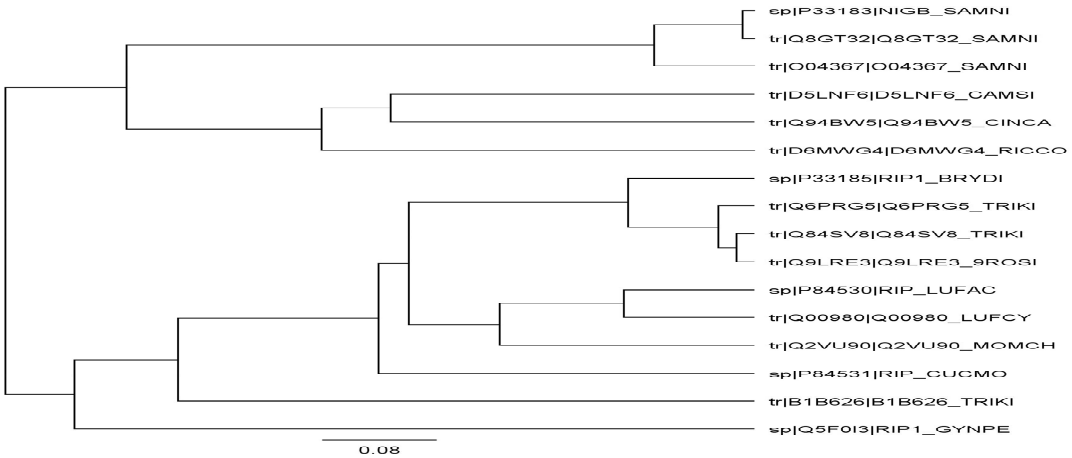

Fig.5 Tree-III depicts evolutionary relationships of β-momocharin (Ribosome Inactivating Protein of *Momordica charantia*) with other Ribosome Inactivating Proteins.

In Fig.5 while on analysing Tree-III it has been found that curcumosin is ortholog to β-momocharin and the remaining RIP's (Bryodin-I, Karasurin-H, Luffaculin I, β-luffin, Trichobakin, Trichomislin, Trichosanthin and Gynostemmin) in the cluster are paralogs to β-momocharin. In the above depicted Phylogenetic trees each tree comprises of sixteen proteins out of which nine proteins were found in one cluster which were taken for further structural analysis and out of those nine proteins only for five the 3-D structures were found in PDB their PDB ID along with their amino acid residues shown in Table III.

Table III. Phytoproteins with aminoacid residues and PDB ID

Phytoprotein	No. of amino acid residue	PDB ID
B-luffin	278	1NIO
Bryodin I	290	1BRY
Cucurmosin	244	3BWH
Luffaculin I	241	2OQA
Trichosanthin	289	1J4G

And the rest four proteins were modeled with SWISS-MODEL are shown in Table IV and the protein structures modeled are shown in Fig. 6, 7, 8 and 9 respectively.

Table IV. Phytoproteins with aminoacid residues and assigned PDB ID

Phytoprotein	No. of amino acid residue	PDB ID
Karasurin-H	282	Model I
Trichobakin	247	Model II
Trichomislin	270	Model III
Gynostemmin	300	Model IV

Next, after superimposing the structures we got the eight proteins having the structure homologous to the ribosome inactivating proteins of *Momordica charantia*. It has been predicted that only curcumosin is the orthologue for β-momocharin protein i.e. both of them retain same function in the course of evolution and are more closely related protein structures and are homologous in structure and function. The rest RIP's: β-luffin, Bryodin, Gynostemmin, Luffaculin, Karausin-H, Trichosanthin, Trichomislin, Trichobakin are paralogues to the Momordin-I, α-momocharin and β-momocharin respectively. In the Table 5 the phyto-proteins homologous to Momordica charantia RIP's along with their Root Mean Square Deviation (RMSD) values are depicted in Table V.

Fig.6

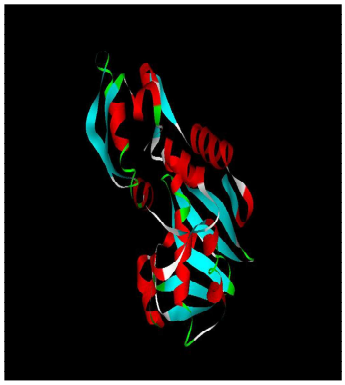

Fig.7

Fig.8

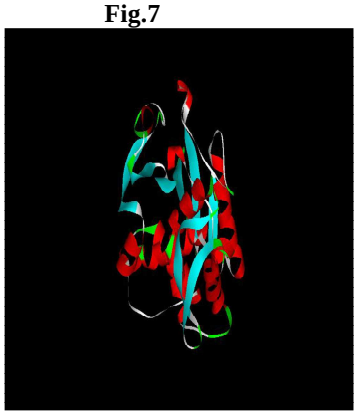

Fig.9

Table V. RMSD values of Phyto-proteins corresponding to Momordin-I, α-momocharin and β-momocharin.

Phyto-proteins	Root Mean Square Deviation (RMSD) in Å		
	Momordin-I	α-momocharin	β-momocharin
β-luffin	1.27 Å	1.10 Å	0.973 Å
Bryodin	0.627 Å	0.603 Å	0.91 Å
Curcumosin	1.47 Å	1.37 Å	0.833 Å
Gynostemmin	2.43 Å	2.39 Å	1.183 Å
Karausin-H	1.96 Å	1.88 Å	0.733 Å
Luffaculin	0.55 Å	0.527 Å	0.828 Å
Trichosanthin	0.084 Å	0.812 Å	0.908 Å
Trichobakin	1.55 Å	1.54 Å	0.393 Å
Trichomislin	1.12 Å	1.07 Å	0.943 Å

As a result of superimposition it has been found that the protein structures with RMSD less than 2 Å are found to be structurally similar to the RIP's of *Momordica charantia*. Out of these nine phytoproteins Gynostemmin shows no significant structural similarity with Momordin-I and α-momocharin as its RMSD value is more than 2 Å i.e. 2.43Å and 2.39 Å respectively. While it has been found similar to β-momocharin with RMSD 1.183 Å and therefore can show the abortifacient property as β-momocharin.

DISCUSSION

The impetus for research to identify novel chemical templates from plants for the synthesis of active anti-fertility compounds is very rare. Even if the studies conducted on phyto-proteins serves to produce compounds with novel structures, excellent activity, and negligible side effects, new molecular models will have been identified for further synthetic studies. The medicinal plants derived drugs are being effectively used as a cure for a variety of severe diseases and disorders. The ribosome inactivating proteins are mostly of plant origin. The activity of RIP varies from substrate to substrate aside from depurinating the ribosomes at α-sarcin loop. RIP's are found to be most effective against both animal and plant ribosomes still there are certain speculations to be done on activity of ribosome

inactivating proteins [10]. Although the prescription market for drugs made from plants is lucrative, plant-derived anti-fertility agents have not been investigated extensively. Nowadays with the advent of Proteomics and Genomics the use high throughput techniques enhanced sequencing of many genomes is yielding a large number of highly curated protein sequences [11-13]. The use of ribosome activating proteins as drug alternatives is a novel approach towards the development of drugs. The functions of these phyto-proteins now need to be characterized, elaborated, understood and manipulated. It is generally useful to know the three dimensional structures of the proteins for the identification of functional similarities of homologous phyto-proteins. In the absence of an experimentally determined structure, comparative or homology modeling can sometimes provide a useful model of a protein (target) that is related to at least one known protein structure (template) [14,15]. To ameliorate this problem we have introduced an *in-silico* approach for the identification and modeling of novel phyto-proteins showing the anti-fertility property. The ribosome inactivating proteins (RIP's) from plant *Momordica charantia* are taken for the present analysis shows the abortifacient property. As a result of similarity search similar proteins were identified by phylogenetic analysis it was found that some of them showed similar functional and structural homology to the RIP's of *Momordica charantia* and some showed no significant similarity. Some of protein structures similar to the RIP's of *Momordica charantia* were modeled further on superimposing these protein structures of different RIP's from plants we found certain protein structures which were more similar to the RIP's of *Momordica charantia* were predicted to show similar properties on the basis of their RMSD statistics. As for accuracy in structural similarity the proteins which deviate (more than or equal) from RMSD 2Å were discarded. Consequently, nine phyto-proteins were predicted out of which Gynostemmin showed structural similarity only to β-momocharin since its RMSD value falls within limits i.e. 1.183Å.While for the rest two its RMSD value exceeds more than 2Å. Therefore on the basis of structural similarity these all proteins can be predicted as novel proteins showing the anti-fertility properties. The comparative protein structure modeling is relevant to structure-based functional annotation of proteins and thus enhances the impact of genome sequencing, structural genomics and functional genomics on biology and medicine [16].

CONCLUSION

The impact of tremendously rising population gives rise to economic decline, stagnation, global poverty, hunger, environmental devastation, political unrest etc. Therefore there exists a need to abolish this hiking population. In-spite of the various effective synthetic anti-fertility drugs available there is a need to design new drugs of plant origin with negligible side effects as compared to synthetic ones. There are several approaches introduced for the proteome analysis here we have used the protein sequences in order to get the alternatives for existing anti-fertility proteins. The Ribosome Inactivating Proteins sequences were taken for study these proteins show the anti-fertility properties as reported in *Momordica charantia* taken here for study we have come up with nine structural homologues of this protein as a result of our analysis these phyto-proteins can also show the anti-fertility property as reported in Ribosome Inactivating Proteins of *Momordica charantia.*

REFERENCES

[1] Tatyana P, Soubbotina Katherine AS, The World Bank Washington D.C. U.S.A. , **2000** , pp. 16-21.
[2] Pimentel D, Huang X, Cordova A, Pimentel M, *Population and Environment: A journal of Interdisciplinary Studies*, **1997**, 19,1, 9-14.
[3] Anthony Cassils J, *Population and Environment*, **2004,** 25, 3, 171-194.
[4] Patwardhan B, Vaidya ADB, Chorghade M, *Current Science*, **2004**, 86,6, 789-799.
[5] Topliss JG, Clark AM, Ernst E, Hufford CD, Johnston GAR, Rimoldi JM, Weimann BJ, *Pure and Applied Chemistry*, **2002,** 74, 10, 1957–1985.
[6] Pang-Chui Shaw, Ka-Ming Lee, Kam-Bo Wong, *Toxicon*, **2005,** 45, 6, 683-689.
[7] Stirpe F, Battelli MG, *Cell. Mol. Life Science*, **2006,** 63, 1850–1866.
[8] Kumar DS , Sharathnath KV, Yogeswaran P, Harani A, Sudhakar K, Sudha P, David B, *International Journal of Pharmaceutical Sciences Review and Research*, **2010,** 1, 2, 95-100.
[9] Maiti R, Van Domselaar GH, Zhang H, Wishart DS, *Nucleic Acids Research*, **2004 ,** 32, w590-w594.
[10] Nielsen K, Boston RS, *Annu. Rev. Plant Physiol. Plant Mol.* Biol, **2001**, 52, 785–816.
[11] John B, Sali A, *Nucleic Acids Research*, **2003**, 31,14, 3982-3992.
[12] Cantor CR, Little DP, *Nature Genet*, **1998**, 20, 5-6.
[13] Grunenfelder B, Winzeler EA, *Nature Rev Genet*, **2002**, 3, 653-661.
[14] Sali A, Blundell TL, *J Mol Biol*, **1993**, 234, 779-815.

[15] Marti RMA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A, *Annu Rev Biophys Biomol Struct* **2000,** 29, 291-325.
[16] John B, Sali A, *Nucleic Acids Research,* **2003**, 31, 3982-3992.