International Journal of scientific research and management (IJSRM)
||Volume||2||Issue||10||Pages||1498-1501||2014||
Website: www.ijsrm.in ISSN (e): 2321-3418

A Comparative Study of Methodologies of Protein Secondary Structure

M.Rithvik

Aditya Institute of Technology and Management, Tekkali
rithvikmadugula@email.com

Abstract: *All living organisms are made up of cells and each cell in its turn consists of certain protein consequences which exercise an important role in catalyzing the chemical reactions. So, a study of a protein structure becomes a search lamp in the diagnosis of a disease .when the percent identity between two protein sequences falls below 33%,it necessities to carry out the analysis of protein secondary structure. Of the several methodologies developed to analyze the protein secondary structure, two methods proved to be sound-DSSP(Dictionary of Secondary Structure of Proteins) and GOR(Garnier ,Osguthrope and Robson),Even though the prediction accuracy of GORV is 73.5% due to hazards in its implementation ,GORIV is generally used in spite of its accuracy being only to 64.4%*

Keywords:protein,residue,motif,DSSP,GOR

1. Introduction

Every human body is consistuted with a certain amount of cells.The functioning of the human body depends upon the functioning of cells.Every cell consists of certain protein sequences.Proteins play a vital role in catalyzing the chemical reactions in all the living organisims. Proteins are formed by the combination of several amino acids.the information flow from a DNA sequence to the protein structure is as follows:
The information flow from DNA to RNA first.From RNA we acquire a protein sequence .This protein sequence helps us in predicting a protein structure.So,protein sequence plays a key role in predicting the structure of a protein.By knowing the structure we can find the function of the protein.

Necessity to predict the protein structure
Predicted structures can be used in structure based drug design.It can help us understand the effects of mutations on structure or function.Structural knowledge brings understanding of function and mechanism of an action.It Can help in prediction of function

2. Background

Protein structure is classified into 4 categories[1]:

1.Primary Structure: This structure is formed by the linear sequence of amino acids A single change in the amino acid sequence of hemoglobin will result in a disease
2.secondary structure: This structure is formed by the combination of hydrogen bonds between amino acids from two particular elements alpha helices and beta sheets
3.Teritary structure: This is a more compact globular shape with carbon rich amino acids that is inside far away from the surrounding water
4.Quaternary Structure: This structure is formed by the combination of two or more polypeptide chains into one module with several subunits
In this paper we are going to predict the secondary structure of a protein with the help of two algorithms 1.DSSP 2.GOR
We are also visualize that structure with the help of tools.
This kind of prediction gives us an accuracy of 70% when compared to other normal algorithms.

3. Methodology

The secondary structure of protein mainly constitutes of two elements:
i)Alpha helix ii)Beta sheet.The regular patterns of the H bonds are formed between neighboring amino acids and the amino acids have similar φ and ψ bonds
This secondary structure has three regular forms 1.Alpha helix 2.Beta sheet 3.loops.We give the input as the amino acid sequences and the results of output will be the predicted structure.A protein consists of about 32% alpha (□) helices 21% beta (□) sheets and 47% loops or non regular structure that is it may be a motif or a fold.In the above figure the rotation angle around the N-c□ is called ¢ and the rotation angle between c-cα is called psy.The proteins that are evolved from a common ancestor are called homologous proteins.Generally we used to predict the protein structure for a non homologous proteins from the protein data bank
***α* Helix:** we have to note an important point that here everything is related to residue and its position. The following are the various residues for a protein sequence[7].It consists of 3.6 residues per turn and the angles between -60 degrees and -50 degrees respectively.The hydrogen bond that exists between the NH group and CO group of residues is called a α helices.It consists of a helical structure with a turn and an average of ten residues of length

β-sheets: β sheet is built from a combination of several polypeptide chains that constitute of 5 to 10 residues long.These form a bond between CO and NH groups and form three kinds of chains 1.Parallel 2.antiparallel
In this paper we are going to predict the protein structure by using the two methods
1.DSSP 2.GOR

3.1.The DSSP

DSSP stands for Dictionary of secondary structures.We generally take a non homologous protein structure to calculate the protein structure.Dssp is a program that is used to convert or transfer the number of states that is it is used to decrease the number of states of residues from eight to three[3].The reason behind the decrease of the number of states is to predict the secondary structure that consist of α helices, β sheet, coil and fold.It was first designed by Wolfgang Kabsch and Chris Sander to standardize the secondary structure assignment.DSSP is a database of secondary structure assignments for all protein entries in the Protein Data Bank(PDB)

Brief history: The original DSSP application was written between 1983 and 1988 in the programming language PASCALThis pascal code is automatically converted but maintaining this proved to be a lot of trouble-taking

Working: The Dssp program works by calculating the most likely secondary structure assignment that was given in the 3D structure of a protein[7]It does this by reading the position of the atoms in a protein followed by calculation of the H bond energy between all atoms.The best two H-bonds for each atom are then used to determine the most likely class of secondary structure for each residue in the protein

Structure: The structure of the DSSP is as follows:

G = 3-turn helix (3_{10} helix). Min length 3 residues.
H = 4-turn helix (α helix). Min length 4 residues.
I = 5-turn helix (π helix). Min length 5 residues.
T = hydrogen bonded turn (3, 4 or 5 turn)
E = extended strand in parallel and/or anti-parallel β-sheet conformation. Min length 2 residues.
B = residue in isolated β-bridge (single pair β-sheet hydrogen bond formation)
S = bend (the only non-hydrogen-bond based assignment)

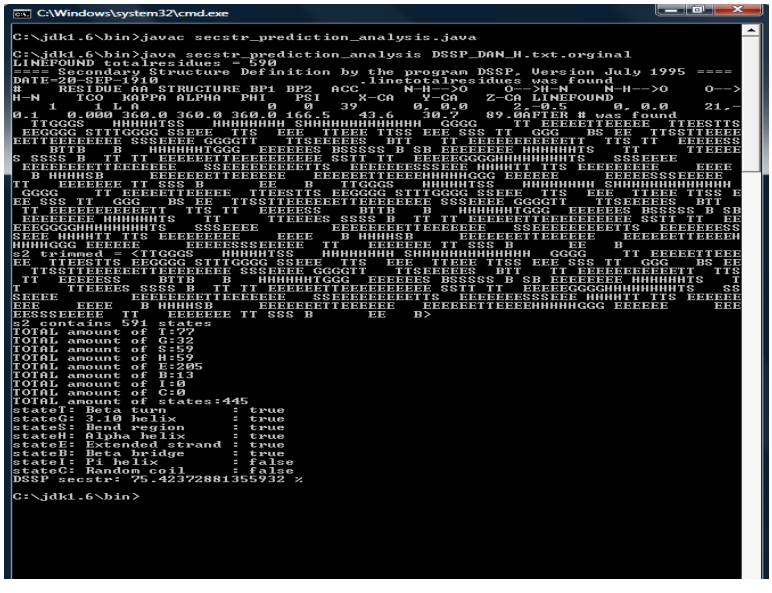
DSSP obtained from non redundant PDB_select dataset,secondary structure assigned by DSSP into eight conformational states to three states H=HGI,E=EB,C=STC
3.2.GOR: The GOR method was first developed by Garnier,Osguthrope and Robson[5].It is an information theory based method that uses a more powerful probabilistic techniques of Bayseian inference.This method not only takes into account the probability of each amino acid having a particular secondary structure but also the conditional probability of the amino acid assuming each structure given the contributions of its neighbors.This method assumes that amino acids upto eight residues on each side that influence the secondary structure of the central residue.This program has many versions but here we are going to write the program in the fourth version.The algorithm makes use a sliding window of 17 amino acids.All possible pairs of amino acids in this window are checked for their information content as to predict the central amino acid by comparing the set of 266 other proteins of known structure.In this version we are going to find the certain pairwise combinations of amino acids that are present in the region called as Flanking Region.If a particular amino acid is surrounded by residues that prefer to be a helics it is likely to be a helix.Even if its individual helical preference is low.This method helps in considering a propenties of a single residue position dependent propensities for helics,sheet and turn has been calculated for all types of residues

GORV:
This method is a bit improved method when compared to GOR IV method[8].The improvement was made in the algorithm.It was inclusion of evolutionary information using PSI-BLASTMultiple alignments are gennerated using PSI-BLAST after five iteratioms based on the non redundant database.The GOR V method prediction takes a longer time for the multiple alignments that is this kind of alignments takes much more time for hits

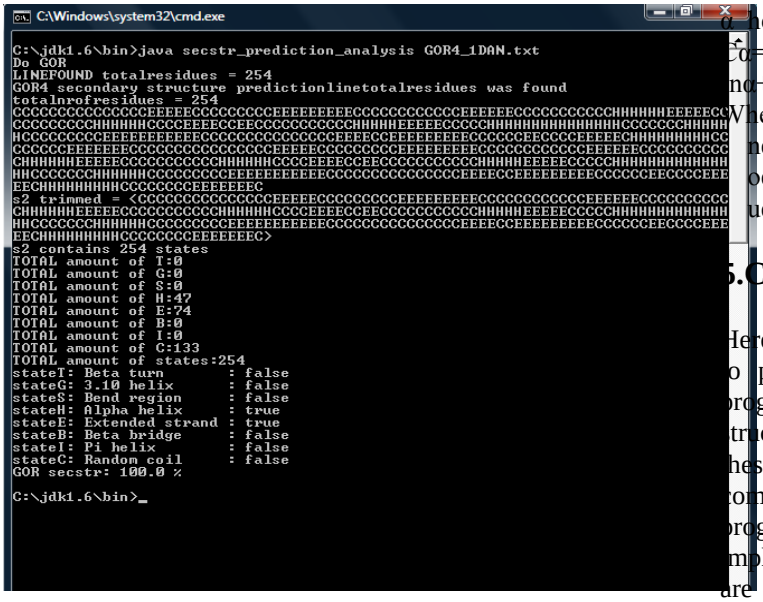
Accuracy of the two methods: The prediction accuracy of the GOR IV method is 64.4% and the prediction accuracy of GORV method is73.5%.Eventhough GOR V method has more prediction among accuracy people generally use the GOR IV methodThis means that GORV method is difficult to implement

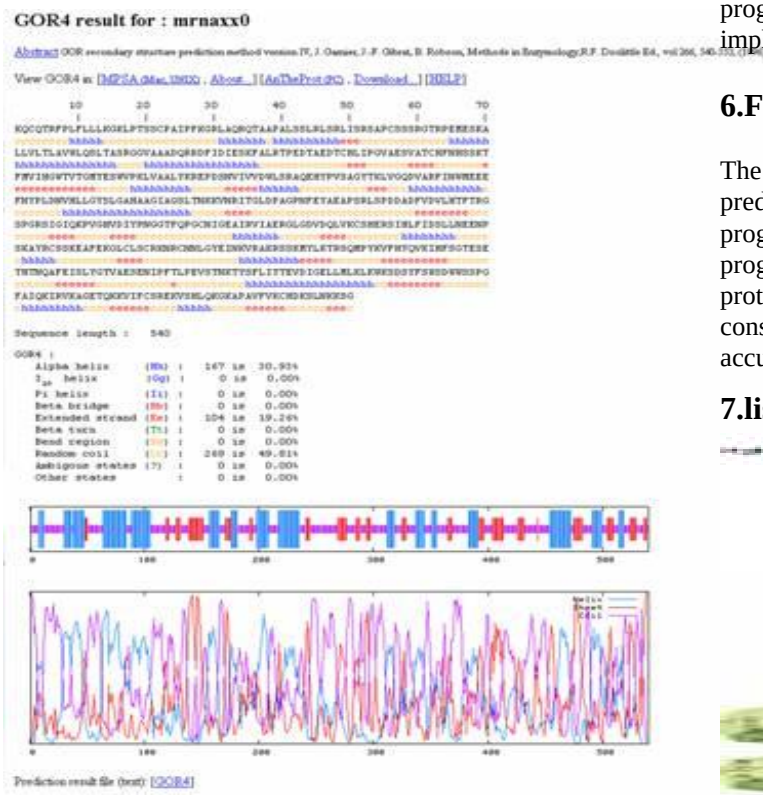
4.Results and Discussion
Here we are going to display the results of DSSP and GOR

Output of DSSP



4.1Output of GOR





Represents the Results of GORIV

These results show that here we are going to predict the secondary structure of a protein by using these two methods
The protein structure can be visualized by using the various tools

SWISS-PDB VIEWER: This is an apllicarion that provides a uer friendly interface allowing to analyze several proteins at the same time.This viewer helps us to visualize a protein secondary structure that was out when performing the DSSP and GOR programs

Accuracy of Prediction of Protein Structure: The accuracy of protein structure prediction depends upon the percentage of correctly predicted residues in sequences of known structure called Q3

A method to calculate the accuracy is to calculate a correlation coefficient for each type of predicted secondary structure. The coefficient indicating the success of predicting residues in the α helical configuration cα is given by:

cα=(pαnα-uαoα)/square root of [nα+uα][nα+oα][pα+uα][pα+oα])

Where pα is the correct prediction
nα is the negative prediction
oα is the overpredicted positive prediction
uα is the number of unpredicted residues

5.Conclusion

Here we are writing a program using DSSP and GOR methods to predict the protein secondary structure.Apart from the programs we are going to visualize the protein secondary structure by using the SWISS-PDB viewer.The performance of these two programs gives us an accuracy of 60-70% when compared to the other programs..Eventhough these two programs gives us less accuracy they are easy to write and implement in java.Eventhough new versions of the programs are giving more accuracy people generally prefer these programs .Because these programs are easy in their implementation and robust in nature

6.Future scope

The further development of these programs helps us in predicting the more accurate protein secondary structures.The programs that will be developed in future must consider these programs in order to predict the secondary structure of the protein.The further applications of these programs must consider the Ramchandran plot in mind for their prediction accuracy

7.list of figures

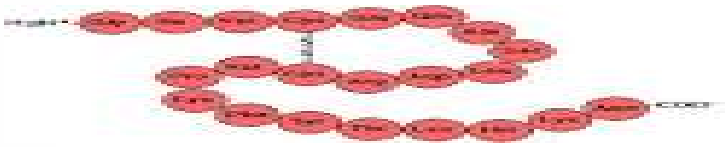
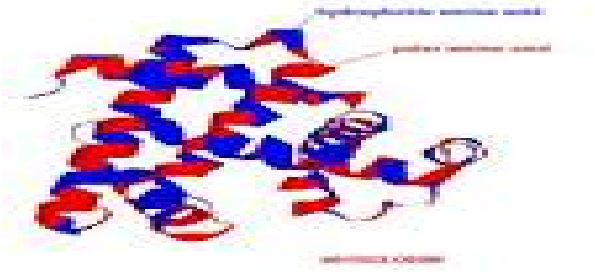


fig1: This figure represents the primary structure of a protein

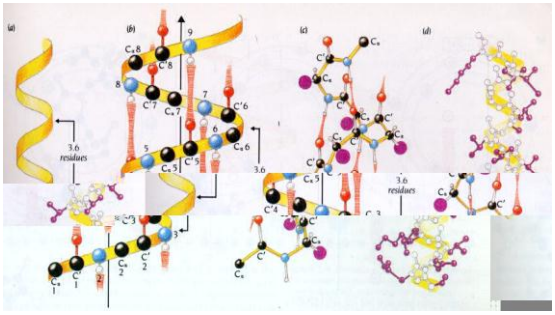
fig2(a) fig2(b)

fig2(a)**:** Represents the alpha helix and fig2(b) represents the beta sheet of secondary structure of a protein



fig(3) represents the tertiary structure of a protein

fig(4) represents the quaternary structure of a protein



fig(5) represents the helical structure

References

[1] Frishman D., Argos P. (1995). "Knowledge-based protein secondary structure assignment". Proteins 23 (4): 566–579. doi:10.1002/prot.340230412. PMID 8749853.

[2] Richards F. M., Kundrot C. E. (1988). "Identification of structural motifs from protein coordinate data: secondary structure and first-level supersecondary structure". Proteins 3 (2): 71–84. doi:10.1002/prot.340030202. PMID 3399495.

[3] Kabsch W., Sander C. (1983). "Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features". Biopolymers 22 (12): 2577–2637. doi:10.1002/bip.360221211. PMID 6667333.

[4]" Prediction of the Number of Residue Contacts in Proteins Piero Fariselli and Rita Casadio CIRB Biocomputing Unit and Lab. of Biophysics," Dept. of Biology University of Bologna via Irnerio 42, 40126 Bologna Italy

[5]. "A Java Based Protein Secondary Structure Prediction" ProgramG. Nageswara Rao, Allam Appa Rao

[6]. "literature survey of prediction of protein"

[7]. "DSSPcont: continuous secondary structure assignments for proteins" Phil Carter1,2,4,*, Claus A. F. Andersen1,3 and Burkhard Rost1,2,5 Nucleic Acids Research, 2003, Vol. 31, No. 13 3293–3295 DOI: 10.1093/nar/gkg626

[8]." GOR V server for protein secondary structure prediction", Taner Z. Sen,1,2 Robert L. Jernigan,1,2 Jean Garnier3 ,France Received on February 23, 2005; revised on March 21, 2005; accepted on March 22, 2005Advance Access publication March 29, 2005