ISSN 2250-3153

Prediction of Possible Cluster in Carbon Atom Rich and Low Amino Acids Using Permutation Method

R.Senthil[*] and M.Vijaya[**]

[*] Department of Bioinformatics, Marudupandiyar College, Thanjavur, Tamilnadu
[**] Department of Mathematics, Marudupandiyar College, Thanjavur, Tamilnadu

Abstract- Finding different clusters in sequence plays major role in the computational biology, which could help as evidence of structural and functional conservation, as well as of evolutionary relations among the sequence. Functionally related genes often appear in other's neighborhood on the genome. However, the order of the genes may not be at the same. These groups or clusters of genes may have an ancient evolutionary origin or may signify some other critical phenomenon and may also serve in function prediction. Similarly, cluster of protein domains as appearing in different orders in the protein sequence and suggests more functional parts. Proteins are the large organic compounds made of amino acid in a linear fashion. The side chains of these amino acids are chemically different from one another in some respect can be classified broadly two ways hydrophobic and hydrophilic. Carbon content in these side chains makes the amino acids. It is easily demonstrable that the linear amino acid sequence completely specifies the three dimensional shapes. The study is carried out for prediction of possible cluster in carbon atom rich and low amino acids using permutation method. It is interesting to observe that permutations involving as many as large hydrophobic residues or clusters.

Index Terms- Computational Biology**;** Clustering; Pattern Prediction; Permutation Method; C & Perl Program

I. INTRODUCTION

Computational methods of biological sequence analysis have become an indispensable part of the modern scientist's research arsenal. In protein studies, the results of sequence similarity searches in databases help generate reasonable hypotheses concerning structural and functional properties of proteins, as well as their evolutionary relationships. On the DNA level, sequence analysis techniques make it possible to identify genes and functional elements in newly sequenced genomes. The abundance of biomolecular sequence information, together with relatively high cost of laboratory experimentation, calls for powerful and efficient computational tools as primary means for "omics" investigation. Hence sequence comparison and motif analysis methods can be used to predict protein-protein interaction and interactions in transcriptional regularity networks. Computational sequence analysis also provides a basis for the rapidly developing field of systems biology. However, the evolution of DNA and proteins in living organisms is influenced by a number of random factors, and observed amino acid or nucleotide patterns may result from the action of these factors, rather than from the selective pressure maintaining a certain function [1].Domains are portions of the coding gene that correspond to that correspond to a functional subunit of protein. Those would be found by conserved nucleic acid sequences and amino acid sequences. Motif discovery tools are used to be identification of conservation. In addition, domain may appear in different order. In nature they are functionally related due to the common domains [5, 7].Proteins fold up to form particular structure which gives them a specific function. The hydrophobic interactions are the dominant force that determines the biomolecular association. Carbon is only element that contributes towards the hydrophobic interactions in proteins [2, 3, 4, 6 & 8].

Permutation and Clustering Method

The notion of permutation is used with several slightly different meaning, all related to the act of rearranging objects or values. ***n! = n [(n-1)!]*---------- (1)**

The principle of multiplication and addition could help to understand permutation and combination and form the base for permutations and combinations. ***(m * n) and (m + n)* ---------- (2)** The word permutations mean arrangement of alphabets. ***n P r = n! / (n-1)!* ---------- (3)**

Identification of Permutation

```
for(i=start;i<=n;i++)

{       d=c(start),c(start)=c(i),c(i)=t;
                Permutation(s,start+1,n);
                h=c(start),c(start)=c(i),c(i)=t;      }
```

Identification of Different Cluster

```
if($search=~ /DEFHILPWYR/)
    {     print("Carbon atom rich aa cluster 'DEFHILPWYR' is found.\n");
    }

if($search=~ / ACGKMNQRST /)
    {              print("Carbon atom low aa cluster 'ACGKMNQRST' is found.\n");
    }

if($search=~ / FILMV /)
    {       print("Large Hydrophobic cluster 'FILMV' is found.\n");
    }
```

```
if($search=~ / AGCWYPTS /)
    {    print("small Hydrophobic cluster ' AGCWYPTS ' is found.\n");
    }

if($search=~ / HENQDKR /)
    {    print("Hydrophilic cluster ' HENQDKR ' is found.\n");
    }
```

Permutation and Clustering Result

The theory of discrete signal processing pays a special attention to permutations of information elements. It is intended to identify that all the possible permutations of positions (Sergey V. Petoukhov, 2010 & 2011). Possible permutations of positions can be predicted in simultaneously. Those are example for cyclic transformation: **1-2-3** into different order **{123,132,213,231,312,321}**, the triplet **A-T-G** into different order {**ATG-AGT-TAG-TGA-GTA-GAT}**.Here, it is known that the molecular-genetic system of living matter includes the following alphabets each of which can be considered as a separate alphabet or as a part of complex alphabetic system: 20-Letter alphabet of amino acids; 5-Letter alphabet "Large Hydrophobic"; 8- Letter alphabet "Less Hydrophobic"; 7- Letter alphabet "Hydrophilic".

Table.1.Automatically discovered cluster of carbon rich and low amino acids.

S.NO	AMINO ACIDS	PERMUTATION
1	**Large Hydrophobic Residues-carbon rich (FILMV)**	**FILMV FILVM FIMLV FIMVL FIVML FIVLM FLIMV FLIVM FLMIV FLMVI FLVMI FLVIM FMLIV FMLVI FMILV FMIVL FMVIL FMVLI FVLMI FVLIM FVMLI FVMIL FVIML FVILM IFLMV IFLVM IFMLV IFMVL IFVML IFVLM ILFMV ILFVM ILMFV ILMVF ILVMF ILVFM IMLFV IMLVF IMFLV IMFVL IMVFL IMVLF IVLMF IVLFM IVMLF IVMFL IVFML IVFLM LIFMV LIFVM LIMFV LIMVF LIVMF LIVFM LFIMV LFIVM LFMIV LFMVI LFVMI LFVIM LMFIV LMFVI LMIFV LMIVF LMVIF LMVFI LVFMI LVFIM LVMFI LVMIF LVIMF LVIFM MILFV MILVF MIFLV MIFVL MIVFL MIVLF MLIFV MLIVF MLFIV MLFVI MLVFI MLVIF MFLIV MFLVI MFILV MFIVL MFVIL MFVLI MVLFI MVLIF MVFLI MVFIL MVIFL MVILF VILMF VILFM VIMLF VIMFL VIFML VIFLM VLIMF VLIFM VLMIF VLMFI VLFMI VLFIM VMLIF VMLFI VMILF VMIFL VMFIL VMFLI VFLMI VFLIM VFMLI VFMIL VFIML VFILM**
2	**Less Hydrophobic residues and Hydrophilic- Carbon low (CKRS)**	**CKRS CKSR CRKS CRSK CSRK CSKR KCRS KCSR KRCS KRSC KSRC KSCR RKCS RKSC RCKS RCSK RSCK RSKC SKRC SKCR SRKC SRCK SCRK SCKR**

In the paper, prediction of possible cluster is carried out for carbon rich and low amino acids (Table.1), as to know about the protein stability and order of subunits. There is a known factor that identification of hydrophobicity and hydrophilicity is important to address major issues in molecular association. It is essential to know what factor is responsible. In order to study these details, carbon distribution in proteins is helped. Therefore, identification of possible clusters is taken up in amino acids by permutation method.

II. Conclusion

As it was already used for dinucleotide and trinucleotide for DNA. Now it is possible to implement randomness by applying shuffling algorithm. Another approach is to select biomolecular sequences at random from available databases. In the field of molecular genetics whole families of proteins should work to provide physiological process in an active. Although it's biological significance is yet to be established, nevertheless, it appears to be an interesting phenomenon.

References

[1] Alexander Yu. Mitrophanov and Mark Borodovsky, 2005. Statistical significance in biological Sequence analysis, *Briefings in Bioinformatics*. **7(1)**:2-24.

[2] Jayaraj .V, Suhanya .R, Vijayasarathy .M, Anandagopu .P and Rajasekaran .E, 2009, Role of Large Hydrophobic Residues in Proteins, *Bioinformation*, **3(9)**:409-412.

[3] Jayaram .B, Aneja .N and Rajasekaran .E. 1994, Modeling DNA in Aqueous Solutions, *J. Sci. Ind. Res*., **53**:88.

[4] Rajasekaran. E, Vinobha C.S, Vijayasarathy.M and Senthil.R and Shankarganesh.P, 2009, The Nature of Proteins, IACSIT-SC, *IEEE*: 452-453.

[5] Revital Eres, Gad M. Landau, and Laxmi Parida, 2004. Permutation Pattern Discovery in Biosequences. *Journal of computational biology*, **11(6)**:1050-60.

[6] Senthil .R and Rajasekaran .E, 2009, Comparative Analysis of Carbon Distribution and Hydropathy Plot, *J.Adv. Biotech*. **8(9)**:30-31.

[7] Sergey V. Petoukhov, 2011. The genetic code, 8-dimensional hypercomplex numbers and dyadic shifts, *International Conference on Bioinformatics & Computational Biology*, Las Vegas, USA, July 18-21.

[8] Suhanya .R, Anandagopu .P and Rajasekaran .E, 2008, Significance of Carbon Distribution in the Functional Site of Mitochondrial Protein PINK1, *Int.J. Bioinfo*. **1 (1)**: 9-12.

[9] www.rajasekaran.net.in/tools/carana.html

AUTHORS

First Author – Dr.R.Senthil, Department of Bioinformatics, Marudupandiyar College, Thanjavur, Tamilnadu, renganathansenthil@gmail.com

Second Author – M. Vijaya, Department of Mathematics, Marudupandiyar College, Thanjavur, Tamilnadu, mathvijaya79@yahoo.com