Computational Analysis and Classification of p53 Mutants according to Primary Structure

Krishna Gopalakrishnan[1], R.H. Zadeh[2], Kayvan Najarian[1], Alireza Darvish[1]
[1]Department of Computer Science
[2]Department of Physics
University of North Carolina at Charlotte.
(kgopalak, rhakimza, knajaria, adarvish)@uncc.edu

Abstract

Widely used multiple alignment based techniques can give false results for single base mutation as the primary sequence of mutants and that of the wild types are very similar. We present a technique that uses signal processing methods along with biochemical properties of individual amino acids for the analysis of proteins. Each amino acid of mutant protein is replaced with the corresponding biochemical properties generates a set of biochemical signals. These signals are used to extract signal processing features like complexity, mobility, fractal dimension, and wavelet transformation. In an experimental study of p53 protein, mutants resulting from single mutation of eight residue of the β-strand 326-33 to alanine were analyzed for their ability to stimulate transcription, to inhibit the growth of Saos-2 cells, and to repress the promoter of multidrug resistance gene. Our classification results, merely based on the analysis of primary sequences, are matching with those of the experiential studies.

1. Introduction

There is a pressing need for fast and accurate methods for classification and analysis of mutated proteins. Single base mutation may or may not result in a change in structure and function of wild type. Multiple alignment based classification techniques can give a false result in such cases as the primary sequence of mutants and that of the wild types are very similar. In our previous studies we have used signal processing based techniques for classification of different families of proteins [1, 2]. In this paper we have applied a modified version of those techniques for clustering of proteins resulted from the mutation of the wild type.

The proposed technique uses signal processing methods along with biochemical properties of individual amino acids for clustering and further analysis. Each amino acid in protein produced by the mutation of wild type is replaced with the corresponding biochemical properties like isoelectric point value, hydrophobicity and hydrophilicity [3]. Some of these properties are known to be instrumental in predicting the function or structure of proteins [4]. Amino acid substitution with the corresponding biochemical properties generates a set of biochemical signals which are used to extract signal processing features like fractal dimension and wavelet transformation. These measures are shown to be suitable in distinguishing proteins of different functional families.

2. Proposed Method

The Proposed method consists of three main steps and they are forming signals based on biochemical encoding, extracting signal processing measures and cluster/classify the proteins. We have created 3 signals for each protein under study by replacing each amino acid with the values of above mentioned biochemical properties. These signals are used to extract signal processing features like fractal dimension and wavelet transformation.

Fractal dimension describes the complexity of a fundamental pattern that can generate the entire sequence by scaling and shifting by itself. Algorithm used is Higuchi's as it is the most accurate and efficient method for calculating fractal dimension [5]. From a time series with N points a set of k sub series are obtained each with a different step size or interval size (where m = 1,2,3…k): i.e.:

$X_k^m : X(m), X(m+k),..., X(m+[\frac{N-m}{k}]*k)$. Then

$$\left(\sum_{i=1}^{\lfloor N-m/k \rfloor} | x(m+ik) - x(m+i-1)k \mid N-1 \right) / (\lfloor N-m/k \rfloor)k$$

is l_k, length of the curve. Slope of the plot ln(L(k) versus ln(i/k) is the estimate of the fractal dimension.

Proceedings of the 2004 IEEE Computational Systems Bioinformatics Conference (CSB 2004)
0-7695-2194-0/04 $20.00 © 2004 IEEE

IEEE COMPUTER SOCIETY

Wavelet Transform uses a localized function called a mother wavelet and fits the scaled and shifted versions of this function to a sequence. This transform can also measure self-similarity at different scales of the sequence. In the proposed method, the measure of approximation of the wavelet transform to the original signal in each sub-band as well as the normalized power of coefficients in each sub-band are used as a feature for sequences. The wavelet measures, which are non-local features of the signal, are calculated based on the detail coefficients of the wavelet decomposition of the hydrophobicity signal on the 22nd and 24th level of decomposition using db2 (Daubechies 2) mother wavelet. The detail coefficients on the 22nd and 24th level of decomposition of the signal provide three coefficients which are used along with the fractal dimension measures to cluster the proteins.

3. Clustering Results

In an experimental study of p53 protein, mutants resulting from single mutation of eight residue of the β-strand 326-33 to alanine were analyzed for their ability to stimulate transcription, to inhibit the growth of Saos-2 cells, and to repress the promoter of multidrug resistance gene [6]. The results obtained by our computational techniques produce three clusters which are shown in figures 1 and 2. The first cluster contains mutants L330A and I332A, second contains F328A, and the last contains E326A, Y327A, T329A, Q331A, and R333A.

Figure 1. Clustering of p53 wild-type and mutants. The axes are Fractal dimension of hydrophobicity, Wavelet Feature #1 of Hydrophilicity, Wavelet Feature #2 of Hydrophilicity.

In the figures integer 7 represents mutants F328A where as 6 and 8 represents mutants L330A and I332A. Integers 1, 2, 3, 4, 5 and 9 correspond to mutants E326A, Y327A, T329A, Q331A, and R333A and the wild type p53 respectively. Members of the three clusters shown in the figures are in exact match with that of the experimental results obtained by Chène and Bechter.

Figure 2. Clustering of p53 wild-type and mutants. The axes are fractal dimension of hydrophobicity, and hydrophilicity, Wavelet Feature #1 of Hydrophilicity.

4. Conclusion

The obtained results show the ability of the proposed technique to cluster proteins based on its function even if there is a very high sequence similarity among the proteins that is being clustered. It also reveals the ability of the technique to group mutants in different cluster when the mutations are not in adjacent positions.

References

[1] K. Gopalakrishnan and K. Najarian, "Prediction of Protein Function Using Signal Processing of Biochemical Properties", Proceedings of The Computational Systems Bioinformatics Conference, Stanford, U.S.A., Aug 2003.

[2] D.S. Warren, and K. Najarian, "Neural Networks for Length-Dependent and Length-Independent Functional Classification and Prediction of Small Sequences of Lipase, Protease, and Isomerase", Proceedings of The Fourth International Conference on "Dynamical Systems and Differential Equations", May 24-27, 2002.

[3] J.Damborsk'y,"Quantitative Structure-Function Relationships of the Single Point Mutants of Haloalkane Dehalogenase: a Multivariate Approach", Quant. Struct.-Act. Relat., vol 16, pp. 126-135, 1997.

[4] S.C Bagley, L. Wei, C. Cheng, R.B Altman, "Characterizing oriented protein structural sites using biochemical properties", Proc. Int. Conf. Intell. Sys. Mol. Biol., vol 3, pp: 12-20, 1995.

[5] T. Higuchi, "Approach to an Irregular Time Series on the Basis of the Fractal Theory", Physica D, Vol 31, pp. 277-283, 1988.

[6] P.Chène and E. Bechter, "Cellular Characterization of P53 Mutants with a single missense mutation in the β strand 326-333 and Correlation of their Cellular Activities with in Vitro properties", J.Mol.Biol, vol 288, pp. 891-897, 1999.

0-7695-2194-0/04 $20.00 © 2004 IEEE

IEEE COMPUTER SOCIETY