Classification of Mer Proteins in a Quantitative Manner

Ranjeet Kumar Rout[1,a], Suvankar Ghosh[2,b], Pabitra Pal Choudhury[3,a]

[a]*Indian Statistical Institute, Kolkata-700018, India*
[b]*West Bengal University of Technology, Kolkata-700064, India*

[1]ranjeetkumarrout@gmail.com , [2]bubunxt@gmail.com, [3]pabitrapalchoudhury@gmail.com

***Abstract*— Mer protein has not been classified beyond the scope of its signaling activity. In this paper, the entire Mer protein sequences have been quantified, and a set of clusters have been made to explore the similar behaviors of Mer protein in a quantitative manner. Previous studies have established that Mer proteins have an important role in cancer biology, so for identifying potential strategies it is highly needed to understand the Mer proteins in a quantitative manner for throwing new light in the complex activity of cancer formation. Our results identify several classes of Mer proteins having some similar function.**

***Keywords*— Indicator matrix, Fractal Dimension, Mer proteins, Dendogram.**

I. INTRODUCTION

Mer is a proto-oncogene receptor tyrosine-protein kinase. It regulates many physiological processes including cell survival, migration, differentiation, and phagocytosis of apoptotic cells (efferocytosis). It coats the outside of the cancer cells and helps to transmitting signals inside the cells that aid their uncontrolled growth. Migdall and Graham think that Mer in the nucleus may influence "gene expression" – helping to decide which parts of the cells' DNA are printed or expressed into proteins [1]. If Mer is, in fact, altering genes within cells, it may be one way in which healthy cells become cancerous – with the wrong genes expressed, a good cell may go bad. Or perhaps Mer in the nucleus may help existing cancer cells survive and thrive despite chemotherapy treatment, as is commonly the case in patients who relapse. It is known that leukemic B and T cells have a lot of Mer on their surface, while normal lymphocytes have none, and that this protein promotes cancer cell survival, recent finding that Mer also resides in the nucleus suggests there may be additional ways that Mer is promoting cancer from *within* the cell. Mer receptor tyrosine kinases (RTKs) are increasingly being implicated in a host of discrete cellular responses including cell survival, proliferation, migration and phagocytosis [2]. The results in [3] show that Axl and Mer can function as oncogenes in a number of cancers; these genes have a protective role against the development of colitisassociated cancer. Several human cancers are caused over express Mer, including mantle cell lymphomas [4].

In the present study, a phylogenetic tree has been constructed and a mathematical quantification of Mer protein sequence has been deciphered by using *Fractal Geometry* [5, 6 and 7]. Also, a set of clusters have been made on using the quantitative results based on fractal dimension. So on using this proposed method; one can easily make probable classification of given amino acid sequence of Mer protein.

II. DATA SPECIFICATION AND METHODS

The Uniprot Data Bank (http://www.uniprot.org/uniprot) is the *largest* and most commonly used repository for any kind of information regarding Mer proteins. The quantitative details of Mer protein have been studied in the light of fractal dimension. The method of computation of features for Mer protein sequences are in the following.

A. Phylogenetic tree of Mer proteins

Phylogenetic tree is the evolutionary interrelations between biological species. Their phylogeny depends upon their physical and genetic characteristics [11, 12]. In a rooted phylogenetic tree, each node with descendants represents the inferred most recent common ancestor of the descendants. The distance of one group from the other group indicates the degree of relationship. Tree building methods can be assessed by the efficiency (how long does it take to compute), power (does it make good use of data), consistency (will it converge on the same answer), robustness (does it cope well with violations) and falsifiability (does it alert us when it is not good).

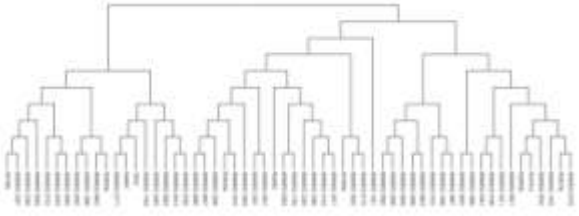


Fig. 01: Phylogenetic tree of 58 Mer Proteins

From the multiple sequence alignments data, we got this dendrogram which helps us how these 58 Mer proteins are sequentially related to each other. Like MSMEG 6703 & Rv3079c are closely related than MSMEG 1475 again Rv2161c & Rv0953c are closely related than MSMEG 5520, but we cannot conclude that these proteins are in a same group. So our aim is to form some sub groups in quantitative manner.

B. Generating indicator matrix and its quantification.

A protein sequence is composed of twenty amino acids namely Glycine=G, Alanini=A, Valine=V, Leucine=L, Isoleucine=I, Serine=S, Cysteine=C, Threonine=T, Methionine=M, Proline=P, Phenylalanine=F, Tyrosine=Y, Tryptophan=W, Histidine=H, Lysine=K, Arginine=R, Aspartate=D, Glutamate=E, Asparagine=N and Glutamine=Q. Let $\nu \stackrel{\text{def}}{=} \{G, A, V, L, I, S, C, T, M, P, D, E, N, H, F, K, Q, Y, R, W\}$ be the set of amino acids and $x \in \nu$ be any member of the alphabet. A protein sequence can be thought of as a finite symbolic string $\mathcal{S} = \mathbb{N} \times \nu$ so that $\mathcal{S} \stackrel{\text{def}}{=} x_i$, $i = 1, 2 \ldots N$ being $x_i \stackrel{\text{def}}{=} (i, x) = x(i), (i = 1, 2 \ldots N; x \in \nu$ the value of x at position i and N denote length of the string. The notion of indicator matrix and its characterization through fractal dimension was proposed by C. Cattani [8] as follows;

$$f: \mathcal{S} \times \mathcal{S} \rightarrow \{0, 1\} \text{ such}$$

$$\text{that} f(x_{\text{h}}, x_{\text{k}}) \stackrel{\text{def}}{=} \begin{cases} 1 \; if \; x_{\text{h}} = x_{\text{k}} \\ 0 \; if \; x_{\text{h}} \neq x_{\text{k}} \end{cases} x_{\text{h}}, x_{\text{k}} \in \mathcal{S}$$

Therefore, the indicator matrix of an N-length string can be easily described as N×N sparse symmetric, binary matrix which results from

$$M_{\text{hk}} = f_{x_h}(x_k) x_h, x_k \in \mathcal{S}, h, k = 1, 2, 3 \ldots, N$$

This definition of indicator matrix does not help us differentiate between zeros formed by distinct base pairs. A slightly modified definition of f is proposed as follows [8, 9]:

$$f: \mathcal{S} \times \mathcal{S} \rightarrow \{0, 1, 2, \ldots, 19\} \text{ such that} f(x_{\text{h}}, x_{\text{k}}) \stackrel{\text{def}}{=} \begin{cases} 0 \; if \; mod((x_{\text{h}}, x_{\text{k}}), 20) = 0 \; x_{\text{h}}; x_{\text{k}} \in \mathcal{S} \\ 1 \; if \; mod((x_{\text{h}}, x_{\text{k}}), 20) = 1 \; x_{\text{h}}; x_{\text{k}} \in \mathcal{S} \\ 2 \; if \; mod((x_{\text{h}}, x_{\text{k}}), 20) = 2 \; x_{\text{h}}; x_{\text{k}} \in \mathcal{S} \\ 3 \; if \; mod((x_{\text{h}}, x_{\text{k}}), 20) = 3 \; x_{\text{h}}; x_{\text{k}} \in \mathcal{S} \\ \vdots \\ 19 \; if \; mod((x_{\text{h}}, x_{\text{k}}), 20) = 19 \; x_{\text{h}}; x_{\text{k}} \in \mathcal{S} \end{cases}$$

From the indicator matrix we have an idea of fractal-like distribution of the amino acids in protein sequences. The corresponding fractal dimensions for the graphical representation of indicator matrices can be computed through *Box counting method* as given below:

Box-Counting Method: This method computes the number of cells required to entirely cover an object, with grids of cells of varying size. Practically, this is performed by superimposing regular grids over an object and by counting the number of occupied cells. The logarithm of N(r), the number of occupied cells, versus the logarithm of 1/r, where r is the size of one cell, gives a line whose gradient corresponds to the box dimension [6, 7]. To calculate the dimension for a fractal S, the Box-Counting dimension is defined as,

$$\text{Dim}_{\text{box}}(\text{S}) = \lim_{n \to 0} \frac{\log N(r)}{\log \frac{1}{r}}$$

Let us understand through an example considering the sequence MSMEG_3445 having the sequence MNSRLFNSHRVVAGCVTECGIVVTNEEFDM…… ……..... (Continuing). Consequently, the matrix M_{hk} corresponding to a given Amino acid sequence is a twenty -threshold matrix, namely 0, 1, 2,…, 19. Let us decompose the matrix M_{hk} into twenty binary matrices A0, A1, and A19 as follows:

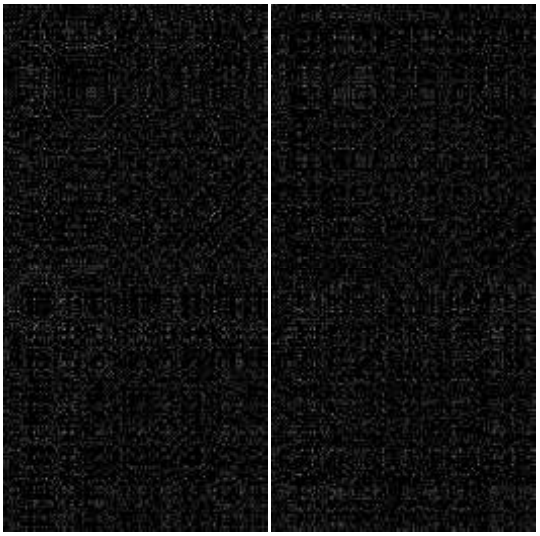
$$A0_{\text{hk}} = \begin{cases} 1 \; where \; mod((x_{\text{h}}, x_{\text{k}}), 20) = 0 \; x_{\text{h}}; x_{\text{k}} \in \mathcal{S} \\ 0 \; otherwise \end{cases}$$

$$A1_{\text{hk}} = \begin{cases} 1 \; where \; mod((x_{\text{h}}, x_{\text{k}}), 20) = 1 \; x_{\text{h}}; x_{\text{k}} \in \mathcal{S} \\ 0 \; otherwise \end{cases}$$

$$A2_{\text{hk}} = \begin{cases} 1 \; where \; mod((x_{\text{h}}, x_{\text{k}}), 20) = 2 \; x_{\text{h}}; x_{\text{k}} \in \mathcal{S} \\ 0 \; otherwise \end{cases}$$

$$A19_{\text{hk}} = \begin{cases} 1 \; where \; mod((x_{\text{h}}, x_{\text{k}}), 20) = 19 \; x_{\text{h}}; x_{\text{k}} \in \mathcal{S} \\ 0 \; otherwise \end{cases}$$

The indicator matrices A0, A1, A2, and A3 for the MSMEG_3445 are as given below.



A0 *A1*

A2 *A3*

Fig.02: Indicator Matrices of MSMEG_3445

We have calculated the fractal dimensions of the images using one of the well-known methods called *'Box-Counting method'*. The fractal dimensions of the indicator matrices A0, A1, A2 up to A19 for MSMEG_3445 are 1.24655, 1.22364, 1.18134, …, 1.15497 respectively [10,11]. In the similar fashion the fractal dimensions of the indicator matrices for all MER Protein amino acid strings have been computed.

1) Fractal Dimension of Indicator matrixes

Here we compute the indicator matrices A0, A1… and A19 for all the Mer protein sequences. Then we calculate the fractal dimension for each of those indicator matrices. The results are elucidated in the following. It is noted that the descriptive statistics for all the features are obtained using the software *Statistica*. The fractal dimensions (FD) for the entire Mer protein sequences (58) are ranging from of A0 indicator matrix is from 1.24655 to 1.52486. The detailed result is given in the *fig .03*. It is noted that the harmonic and geometric mean are almost same 1.43. Similarly the fractal dimension (FD) of 19 indicator matrix is calculated (as shown in *Fig. 03*) and it has been observed that the fractal dimension (FD) of each indicator matrix for all 58 MER proteins lies within the interval from 1.1 to 1.5.

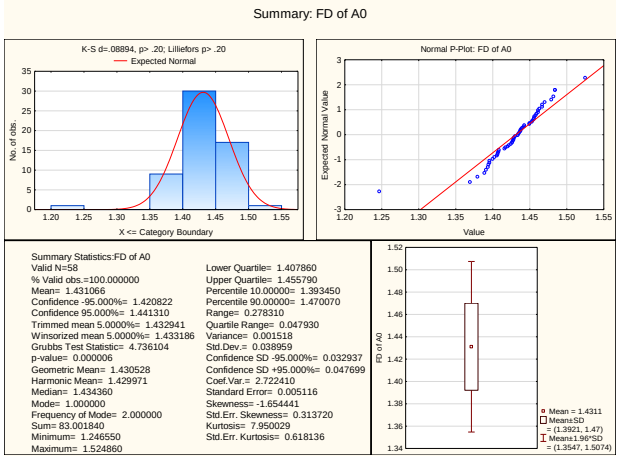


Fig.03: Descriptive Statistics of (FD) of Indicator matrix A0

III. Result and Discussion

Taking twenty amino acids as distinct parameter, we have clustered Mer proteins family into ten (10) different clusters using K-means clustering method [13]. Each cluster contains more than 1 and less than 12 Mer proteins. The clusters are given in the Table. I. The method of Generating Indicator matrix and its quantification of amino acid sequences is used to find out some relevant sub-groups, the observed classifications and the characteristics of the sub-groups will be rationalized in terms of three dimensional structure and functional implications, the observed classification will be used to develop hypothesis on new structural and functional characteristics that can be associated with the respective sub-groups and then these hypothesis will be biologically tested and justified.

TABLE I: CLUSTERS OF 58 MER PROTEINS

Clusters	Members
Cluster-1	MSMEG_3445, MSMEG_2516, Rv1855c, MSMEG_4967, Rv0953c, MSMEG_5520, MSMEG_6621, MSMEG_2896 and MSMEG_2037.
Cluster-2	MSMEG_5715, MSMEG_1749, MSMEG_3301, MSMEG_1996, MSMEG_0341, MSMEG_5920, MSMEG_0339, MSMEG_1794 and 1Rhc.
Cluster-3	Rv2893, MSMEG_4398, MSMEG_3609 and MSMEG_4879
Cluster-4	MSMEG_2249, MSMEG_2996, MSMEG_6454, MSMEG_2307, MSMEG_1027, Rv1360, MSMEG_2256, MSMEG_2456, MSMEG_2103, Rv0407 and MSMEG_0777
Cluster-5	MSMEG_5592, Rv0044c, MSMEG_6885, MSMEG_2906, MSMEG_1475 and Rv3079
Cluster-6	MSMEG_3977, MSMEG_4820, MSMEG_0196, MSMEG_6602, MSMEG_1566, MSMEG_4013 and MSMEG_3554
Cluster-7	MSMEG_6907, MSMEG_3620, MSMEG_0702, MSMEG_0654 and MSMEG_5732
Cluster-8	MSMEG_2904
Cluster-9	MSMEG_2811, MSMEG_6703 and Rv3093c
Cluster-10	Rv2161c and MSMEG_3545

IV. Conclusion

In this paper, we proposed a quantitative method to classify Mer proteins through which Mer proteins are classified based on amino acids distribution and the relationship between them (proteins) without using any biological experiment. This would help biologist for further research in the field of Mer proteins.

Acknowledgment

The authors are grateful to Mr. Sk. Sarif Hassan of International Centre for Theoretical Sciences, Tata Institute of Fundamental Research (TIFR), Bangalore, for their valuable suggestions.

References

[1] J. Migdall-Wilson, C. Bates, J. Schlegel, L. Brandão, RM. Linger, D. DeRyckere, DK. Graham, "Prolonged exposure to a Mer ligand in leukemia: Gas6 favors expression of a partial Mer glycoform and reveals a novel role for Mer in the nucleus, 7(2):e31635(2012). doi: 10.1371/journal.pone.0031635.

[2] S. Hafizi and B. Dahlback, " Signalling and functional diversity within the Axl subfamily of receptor tyrosine kinases", Cytokine & Growth Factor Reviews, Elsevier, Vol-17(2006),Issue-4 pp-295-304.

[3] L. Bosurgi, J. H. Bernink, V. D Cuevas, N. Gagliani, L. Joannas, E. T. Schmid, C. J. Booth, S. Ghosh, and C. V. Rothlin, "Paradoxical role of the proto-oncogene Axl and Mer

receptor tyrosine kinases in colon cancer" PNAS (2013): 1302507110v1-201302507.

[4] S. Ek, CM. Hogerkorp, M. Dictor, M. Ehinger, CA. Borrebaeck . Mantle cell lymphomas express a distinct genetic signature affecting lymphocyte trafficking and growth regulation as compared with subpopulations of normal human B cells. Cancer vol. 62 (2002), pp-4398-4405.

[5] B. B. Mandelbrot, "The fractal geometry of nature". New York,(1982) ISBN 0-7167-1186-9.

[6] D. Avnir "Is the geometry of Nature fractal", Science(1998) 279, 39.

[7] K Develi, T Babadagli "Quantification of natural fracture surfaces using fractal geometry" Math. Geology 30 (8) (1998), 971-998.

[8] C. Carlo,"Fractals and Hidden Symmetries in DNA", Math. Prblm. in Engng.(2010), 507056.

[9] Yu Zu-Guo, "Fractals in DNA sequence analysis", Chinese Physics, 11 (12) (2002), 1313-1318.

[10] Sk. Sarif Hassan, Pabitra Pal Choudhury and Aritra Bose," *A Quantitative Model for Human Olfactory Receptors* ", Nature Precedings, npre20126967-2 (2011).

[11] M. Nei, S. Kumar, Molecular evolution and phylogenetics. New York(2000.): Oxford University Press.

[12] J. Felsenstein, Inferring phylogenies. Sunderland (MA): (2004) Sinauer Associates.

[13] R. H. C. de Melo and A. Conci, "Succolarity: Defining a Method to calculate this Fractal Measure," (2008) ISBN: 978-80-227-2856-0 291-294.

[14] E. Purwantini , B. MukhopadhyayRv0132c of Mycobacterium tuberculosis Encodes a Coenzyme F420-Dependent Hydroxymycolic Acid Dehydrogenase. (2013) PLoS ONE 8(12): e81985. doi:10.1371/journal.pone.0081985