Detecting protein atom correlations using correlation of probability of recurrence

Article *in* Proteins Structure Function and Bioinformatics · September 2014

DOI: 10.1002/prot.24574

Wael Karain
Birzeit University

Detecting Protein Atom Correlations Using Correlation of Probability of Recurrence

Hiba Fataftah and Wael Karain*

Department of Physics, Birzeit University, Birzeit, Palestine.

*Corresponding Author: wqaran@birzeit.edu

Abstract The Dynamic Cross-Correlation Map(DCCM) technique has been used extensively to study protein dynamics. In this work, we introduce the use of the method of Correlation of Probability of Recurrence (CPR) as a complementary method to detect correlations between protein residue atoms. Time series of the distances of the C_α atoms of the β-Lactamase Inhibitory protein(BLIP) from a reference position are analyzed using CPR and Mutual Information(MI). The results are compared to those provided by DCCM. In comparison to MI, CPR is found to detect more of the correlations present in DCCM. It is also able to detect a small number of significant correlations between distant residues that are not detected by DCCM.

KEY WORDS: BLIP; Mutual Information; Cross-Correlation; Molecular Dynamics.

Introduction:

The method of the dynamic cross-correlation map or matrix (DCCM) was introduced to study collective atomic motions in proteins [1,2]. Since then, it has been used extensively to study protein dynamics [3,4,5]. However, this method has drawbacks: First, it does not provide information about the size of the correlated motions. It only provides the value of correlation or anti-correlation between atomic motions. Second, for atoms moving in perfect correlated motion, but along perpendicular directions, the correlation value given will be zero. Thus only atoms moving along the same line will give an accurate correlation value. Third is the fact that the covariance matrix, upon which DCCM depends, only contains linear correlations. This is not suitable for protein motions that exhibit both linear and non-linear components. To deal with these limitations a generalized correlation approach based on mutual information(MI) was introduced. It was shown to identify more correlations than the DCCM method, and captured both linear and nonlinear correlations[6]. A method based on MI, but applied to time series of atom distances from a moving average, as opposed to their (x,y,z) coordinates, was recently used to study residue correlations[7,8]. The advantages of such an approach are:a) lower computational cost, which can be significant for large proteins, b) the independence of the correlations from the direction of motion, c)the ability to detect both linear and nonlinear correlations. In this work, a method based on the recurrences in time series of atom distances from a reference position will be applied to find correlations between the different residues of the protein system under study.

The formal concept of recurrence was introduced by Henri Poincaré in 1890 [9]. In 1987 the method of recurrence plots (RPs) was introduced as a visualization tool to highlight recurrences of dynamical systems [10]. This technique was later upgraded to give quantitative results and became known as recurrence quantification analysis (RQA) [11]. Since then, it has been used in many fields [12].

Recurrence plots (RP) and Recurrence Quantification Analysis (RQA) have been used extensively to study proteins over the years [13 – 47]. Most of the work investigated the sequence-structure relationship in proteins. A few works dealt with analyzing protein molecular dynamics simulations [13,14,20], and the Lennard-Jones systems of different fluid and solid states [36].

The RP is a non-linear analysis method. It visualizes graphically the recurrence of a state S_i to a former state S_j in the phase space trajectory of the system. For example, assume that a dynamical system, with an unknown time evolution law, is being investigated experimentally. If a scalar time-series $\{u_i\}$ representing one of the measurable quantities for this system is available, then the trajectory of this dynamical system can be reconstructed [48]. This reconstruction involves using time delays. The dynamics of the system are presumably stored in the time-series for the single measurable quantity, with the time delays approximating derivatives [49]. The m-dimensional orbit is re-constructed from the scalar u_i, such that $S_i= (u_i,u_{i+1},\ldots..u_{i+m-1})$. This m-dimensional vector in phase space, S_i, represents the state of the system at time i. The RP can be prepared by assigning a black dot at each point (i, j) whenever a point S_j lies within the ball of radius ε centered at S_i. In other words, if two vectors representing the state of the system are within a certain tolerance from each other, then this means that the system is in similar states at two different time instances, i and j. The mathematical expression of the RP is given by:

$$R_{i,j}(\varepsilon) = \Theta\,(\varepsilon - \|S_i - S_j\|) \qquad i,j = 1,\ldots..,N \qquad (1)$$

where N is the number of states, ε is a threshold distance, Θ is the Heaviside function (i.e. Θ(x) = 0 if x<0 and 1 otherwise), and ǁ.ǁ is the Euclidean norm.

The embedding dimension (m) represents the degrees of freedom (or the number of dominant operating variables) in the dynamical system of interest. It is estimated by the method of false nearest neighbors [50]. The delay parameter (d) determines the number of points to be skipped in the time-series between the numbers forming the m-dimensional vector S. It is set to a value that makes the interaction between the points of the measured time-series minimum, and is estimated by finding the first minimum in the mutual information function [51]. The norm parameter determines the size and the shape of the neighborhood surrounding each reference point. There are three types of norms: minimum, maximum, and Euclidean norms. In this work, the Euclidean norm will be used due to its intermediate value of neighbors between the maximum and minimum norms [12]. The threshold or radius parameter (ε) is the limit that transforms the distance matrix (DM) into a recurrence matrix (RM). It plays a role similar to that of the Heaviside function. Elements (i, j) in the DM with distances between states at or below the radius cutoff are included in the RM ($R_{i,j}$ = 1). Elements above the cutoff are excluded from RM ($R_{i,j}$ = 0). This threshold can be chosen using a number of different techniques. In this work, it will be chosen to give a recurrence ratio of 1% for the whole matrix [12].

For each diagonal line in the recurrence plot, a distance τ away from the main diagonal, a τ-recurrence rate can be defined [12]. This measure can be considered as a probability of recurrence, and simply gives the rate of dark points ($R_{i,j}$ =1) per line parallel to the main diagonal. It is given by

$$P(\tau) = \frac{1}{N-\tau} \sum_{i=1}^{N-\tau} R_{i,i+\tau} \qquad (2)$$

where N is the number of states. In other words, $P(\tau)$ gives an estimate of the probability that a system will return to a certain state after a delay of τ[Fig.1]. If we are interested in detecting synchronization between two systems x and y, a measure based on the probability of recurrence can be defined as the correlation between $p_x(\tau)$ and $p_y(\tau)$

$$CPR = \langle p_x \cdot p_y \rangle \qquad (3)$$

where the two probabilities of recurrence values are normalized to zero mean, and a standard deviation of 1,respectively. If the two systems are perfectly synchronized then CPR has a value of 1. If they are not synchronized at all, CPR will have a value of zero [52]. The synchronization value is independent of the amplitude of the motions. This technique has been used to investigate correlations in financial datasets [53] and brain dynamics [54]. In this work, the CPR value is calculated for τ larger than 5 time steps to remove the bias towards high values (P(0)=1)[54].

Figure 1 . The probability of recurrence ($P(\tau)$) for residue 49 in BLIP, as a function of the delay time steps.

Due to the fact that the underlying dynamics of the protein residues are not known exactly, one needs to test for the statistical significance of the results given by the CPR values. For this goal, the method of twin surrogates is used [55]. Assume that a CPR value, C_0, is calculated between the two system time series. The original time series representing one of the systems is not modified, while a number of independent, surrogate, copies of the other system time series are produced based on its recurrence properties. A set of CPR values between the first system original time series, and the surrogate copies of the second system time series, are then calculated. The z-test is then used to compute the statistical significance of the observed CPR value at a certain P-value.

Figure 2. The structure of the BLIP protein(PDB entry 3gmu).

In this work, the protein system to be analyzed is the β-Lactamase Inhibitory Protein (BLIP) produced by *Streptomyces clavuligerus* [Fig.2]. This 165 amino-acid protein inhibits a variety of β-Lactamase enzymes by clamping over their active sites, with residues 49 and 142 playing a major role in this process[56,57]. It consists of two similar domains. The first domain (D1) consists of the amino acids 1-76. The second domain (D2) consists of the amino acids 80-165. Each domain is made up of a helix-loop-helix motif packed against a four stranded β-sheet. Loops of varying lengths connect the four strands in the anti-parallel β-sheet. The protein has a slightly concave interaction interface [57]. In this work, helices will be represented by the letter H. The β strands will be represented by the letter S. There are two disulphide bridges between residues 30 and 42, and residues 109 and 131, respectively[57]. The secondary structure elements in D1 are: H1(5-12), H2(17-26), H3(33-37), S1(30-31), S2(40-47), S3(50-58), S4(67-73). The secondary structure elements in D2 are: H4(84-91), H5(96-105), S5(109-115),S6(126-132), S7(146-152), and S8(155-162).

Methods

The molecular dynamics simulation and related analysis are performed using the molecular dynamics computer programs NAMD[58] and VMD[59], respectively. The starting BLIP protein structure is downloaded from the protein data bank (PDB entry 3gmu)[57]. Periodic boundary conditions are used in an 80Å X 80Å X 80Å box. The protein is neutralized using 20 Cl^- ions and 22 Na^+ ions. The protein is solvated using 15264 TIP3P waters(0.15 M/ NaCl). The Particle-Mesh-Ewald method [60] is used to make the electrostatic calculations. A switching function is used for non-bonded interactions with a switch distance of 10 Å and a cutoff distance of 12 Å. A pair-list distance of 14 Å is used. The simulation is performed at constant pressure of 1.01325 atm with an integration step of 2fs. The protein is minimized using the conjugate gradient method for 2000 steps. This is followed by a gradual heating from an initial temperature of 100K in steps of 30K, with the simulation running for 10ps at each temperature, until reaching the final temperature of 310 K. The simulation then runs for a total of 16ns, with the atom positions saved every 1ps. The first 5ns of this simulation are considered the equilibration stage. The analysis is performed over the next 10ns. For each protein residue, a time series of the root mean square deviation (RMSD) value for the C_α atoms (backbone atom) is prepared. This is done after removing translational and rotational motions using least square fitting. Each series is 1000 points long, with the time interval of 10ps between the points. The protein structure used as a reference in the RMSD calculations is that at the end of the 5ns equilibration interval in the simulation.

The DCCM is calculated using the CARMA program over a time period of 10ns [61]. The MI matrix is calculated using the MILCA program [62]. This is performed using the k-nearest neighbor distance approach with k=6[63]. The MI matrix values are normalized to the range [0, 1] using the following transformation

$$MI_{norm} = (1 - \exp(-2|MI_{raw}|\ ))^{-1/2}$$

where MI_{norm} and MI_{raw} are the normalized and the raw mutual information values, respectively [6]. The recurrence parameters of embedding dimension and delay are calculated using the VRA program [64]. The threshold parameter for a recurrence value of 1% and the CPR matrix are calculated using the CRP toolbox[65]. All plots are prepared using MATLAB[66].

Results and Discussion

Figure 3 shows the inter-residue distance matrix for BLIP. Correlations between residues that are distant from each other should be of special interest [67]. For BLIP, most of these long distance correlations will be between inter-domain residues.

Figure 3. The inter-residue distance matrix for BLIP. The distances(in Å) are calculated between the C_α atoms.

Figure 4.The DCCM correlation map for BLIP. The secondary structure elements are shown. The color-map covers correlation values between -1 and 1.

Figure 5. The MI correlation map for BLIP. The secondary structure elements are shown. The color-map covers correlation values between 0 and 1. The diagonal values are set to zero to improve color contrast.

Figure 6. The CPR correlation map for BLIP. The secondary structure elements are shown. The color-map covers correlation values between 0 and 1. The diagonal values are set to zero to improve color contrast.

Figures 4,5, and 6 show the correlation results for BLIP by DCCM, MI, and CPR, respectively . All matrices are symmetric above and below the diagonal. The color scheme for DCCM is different from MI and CPR due to the presence of anti-correlations with negative values in DCCM. Weak, medium, and high correlations are chosen arbitrarily to refer to absolute values between 0 and .33, .34 and .66, and .67 to 1, respectively. For D1 (residues 1-76), DCCM shows that the whole domain moves in a highly correlated fashion except for a few pockets of weak correlations. These pockets are mainly between some of the loops connecting the secondary structure elements in this domain. The MI matrix also shows clear correlations between the different components of D1. Limited regions of a few residues form bands of weak correlations with other residues in D1. Similarly, CPR shows that D1 is self correlated, with a few bands of weak correlations. Interestingly MI and CPR are able to detect significant correlations in areas of D1 that DCCM fails to detect. For example, while DCCM gives weak correlation values between the residues 58-69 and 44-54, MI gives high correlation values between some of these residues, while CPR gives high correlation values between most of these residues. These correlations are especially noteworthy due to the large distances between the residues involved (Fig. 3).

DCCM shows that D2 (residues 80-165) is also self correlated. Residue groups 80-111, 114-127,129-149,150-165, show thick correlation lines at and near the diagonal. Thick lines parallel and perpendicular to the diagonal give correlations between S6 and S7, H4 and H5, S7 and S8, S6 and H4, S6 and H5, S7 and H4, S7 and H5, S8 and H4, and S8 and H5. These correlations are due to geometrical proximity as can be seen from Figure 2, and thus can be considered trivial. DCCM results also show large anti-correlation islands in D2, mostly between connecting loops. The first island consists mainly of the correlations between the loop joining S6 and S7, with the loop joining S5 and S6. The second island consists mainly of the correlations between the loop connecting S5 and S6, with the N-terminal end of D2 and part of H4. The third island consists mainly of correlations between the loop connecting H5 to S5, and the loop connecting S5 to S6. The fourth island consists of the correlations between the loop connecting S7 to S8, and the loop connecting S6 to S7. The fifth island consists of the correlations between the C-terminal end of D2, and the loop connecting S5 to S6. The sixth island contains the correlations between a part of the loop connecting S6 to S7 with a part of H5. All of these anti-correlation islands occur between residue groups that are geometrically distant from each other (Fig. 3).

The MI matrix shows a sparse correlation picture for D2. It shows a few thick regions of self-correlations along the diagonal, but not nearly as pronounced as those shown by DCCM.   The off-diagonal correlation clusters in MI are also considerably less than those shown by DCCM. The first cluster is between the loop connecting S6 to S7 and the N-terminal end of D2. The second cluster is between the loop connecting S6 to S7, and the loop connecting H4 to H5, and H5, respectively. The third cluster consists of the loop connecting S6 to S7, with the loop connecting S5 to S6. The fourth cluster is between the loop connecting S7 to S8, S8, and the loop connecting H4 to H5, and part of H5, respectively. The fifth cluster is between the loop connecting S7 to S8 and a part of

S8, with the loop connecting S5 to S6. The sixth cluster consists of the loop connecting S7 to S8 with a part of S8, and the loop connecting S6 to S7. MI detects completely three of the anti-correlation regions detected by DCCM between residues that are distant from each other, while it detects two of them partially, and misses one completely. The significant correlation landscape is mostly due to correlations between connecting loops. Correlations between secondary elements in D2, similar to those in DCCM, are not detected.

The CPR matrix shows larger self-correlation blocks along the diagonal for D2 than the MI matrix. However, these blocks are less pronounced than those in DCCM. There are nine major off-diagonal correlation clusters. These include correlations between the loop connecting S5 to S6, with the N terminal end, the loop connecting H4 to H5, the loop connecting S6 to S7, and the C terminal end, respectively. They also include the correlations between the loop connecting S6 to S7, with the N terminal end, the C terminal end, and the loop connecting H4 to H5, respectively. There are also correlations between the C-terminal end with the N-terminal end and the loop connecting H4 to H5 respectively. CPR completely detects five of the anti-correlation regions detected by DCCM, and partially detects the sixth region. CPR also detects a small region of high correlations for distant residues between residues 162-165 and 94-100. DCCM and MI give low correlation values between these residues. In contrast to MI, CPR detects correlations between secondary structure elements as well as correlations between connecting loops in D2.

DCCM shows a mixture of correlation and anti-correlation regions between D1 and D2, with the anti-correlation areas being clearly larger. The correlation areas correspond mostly to geometrically close residue groups, while the anti-correlation regions correspond mainly to geometrically distant residue groups (Fig. 3). We will therefore concentrate on the anti-correlation regions in our comparison with MI and CPR. H4, S6, S7, and S8 are each anti-correlated with H1 and H2, respectively. H5 is anti-correlated with H2, S1, H3, and S2 respectively. H4 and H5 are each anti-correlated with S3, S4 respectively. The loops connecting S5 to S6, and S7 to S8, are each anti-correlated with the loop connecting S2 to S3. The loop connecting S6 to S7, S7, and the C-terminal end of D2, are each anti-correlated to the loop connecting S3 to S4. Thus the inter-domain correlations consist of correlations between secondary elements as well as between loops. Basically, the correlations describe an increase in the concavity of the interaction surface. This correlation picture is in line with the reported inhibition action of BLIP which reportedly clamps over the active site of β-Lactamase[68].

The MI matrix shows clear correlation bands between D1 and D2. A strong correlation band is obvious between the loop connecting S6 to S7, with most of the residues in D1. Another clear band of correlations is between part of the loop connecting S5 to S6 with most of the residues in D1. A band consisting of the loop connecting S7 to S8, with most of the residues of D1 can also be seen. This band is not as homogeneous as the two previous bands. Two weak bands of correlations can also be seen between the loop connecting H4 to H5 and H5, with most of D1, and the C-terminal end of D2 with D1, respectively. A still weaker band can be seen between the N-terminal end of D2 with some of the residues in D1. The correlation bands in MI manage to detect most of the anti-correlations in DCCM. However, the large anti-correlation band between residues 80-110 and most of D1 is only weakly detected by MI.

Compared to MI, the CPR matrix gives larger and clearer correlation bands between D2 and D1. For example, the C-terminal end of D2 shows stronger correlations with D1. The correlations between the loop connecting S5 to S6 with D1, are also much more pronounced and exhibit correlations from all the residues forming the loop and not just some of them. Most noteworthy is that all the anti-correlation areas present in DCCM are detected by CPR, including the large anti-correlation island between residues 80-110 with D1.

In addition DCCM detects very low correlations between certain residues in D2 and D1, while MI and CPR were able to detect higher correlations for the same residues, with CPR giving larger values of correlation for these residues. These include residues 161-165 with residues 1-6, and residues 157-162 with residues 34-37. These residue groups are distant from each other, thus removing geometric proximity as a possible cause of correlations.

Figure 7. The CPR correlation results vs. BLIP inter-residue distance.

Figure 8. The MI correlation results vs. BLIP inter-residue distances.

Figure 9. DCCM correlation results vs. BLIP inter-residue distances.

Figures 7, 8, and 9 show the residue correlation values vs. the inter-residue distances for CPR, MI, and DCCM, respectively. The CPR and MI correlation results are uniformly distributed between close and distant residue pairs. However, DCCM shows that high correlation values occur between close residues, while large anti-correlation values occur mainly between distant residue pairs. Thus it is usually more interesting to study the anti-correlations between residues. A similar observation is reported in the literature[67].

To test for statistical significance of the CPR results, two groups of residues, 40-50 and 130-150, exhibiting weak, medium, and high CPR values are chosen. For each residue RMSD time series in the second group, 100 surrogate time series are produced using the method of twin surrogates. Figure 10 shows the z score results between the original CPR values and the surrogate distribution. Almost all of the observed CPR values between the residues in the two groups have a very small probability (P<0.01) of occurring by chance.

Figure 10. The zscore values for the CPR values between residues 40-50 and 130-150 in BLIP. The vertical dashed line denotes the test significance level, p=0.01.

Conclusion

The correlations between the C_α atoms of the BLIP protein are investigated using DCCM, CPR, and MI. The three methods show that D1 is self-correlated, and basically behaves as one unit. They also show a large number of correlations between residues that are distant from each other. This is especially obvious between residues in D1 and D2. The method of DCCM uses the x, y, and z coordinates for each one of these atoms. It has the ability to find correlations and anti-correlations. On the other hand, CPR and MI are applied to time series of the distances of the atoms from a reference position. This is especially useful when studying proteins with a large number of residues. While DCCM is limited to linear correlations, CPR and MI can detect both linear and non-linear correlations. CPR is able to detect more of the correlations furnished by DCCM than MI. It is also able to detect a small number of significant correlations that are not detected by either DCCM or MI. This is the first time that CPR is used in the field of protein dynamics, and the results in this work show that it complements DCCM and MI very well. In addition, while DCCM and MI provide correlations between atoms, CPR provides information about the amount of synchronization between atoms in terms of their dynamics. Its results are also easy to test for significance using the method of twin surrogates.

Acknowledgments

We would like to acknowledge Dr. Abdallah Sayyed Ahmed for supplying the molecular dynamics simulation for the BLIP protein.

References

1) Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins: Structure, Function, and Bioinformatics 1991; 11(3): 205-217.

2) Hünenberger PH, Mark AE, Van Gunsteren WF. Fluctuation and cross-correlation analysis of protein motions observed in nanosecond molecular dynamics simulations. Journal of Molecular Biology 1995; 252(4):492-503.

3) Arnold GE, Ornstein RL. Molecular dynamics study of time-correlated protein domain motions and molecular flexibility: cytochrome P450BM-3. Biophysical Journal 1997;73(3): 1147-1159.

4) Karplus M, McCammon JA. Molecular dynamics simulations of biomolecules. Nature Structural & Molecular Biology 2002; 9(9): 646-652.

5) Okan OB, Atilgan AR, Atilgan C. Nanosecond motions in proteins impose bounds on the timescale distributions of local dynamics. Biophysical Journal 2009; 97(7): 2080-2088.

6) Lange OF, Grubmüller H. Generalized correlation for biomolecular dynamics. Proteins: Structure, Function, and Bioinformatics 2006; 62(4): 1053-1061.

7) Brandman R, Lampe J N, Brandman Y, de Montellano PRO. Active-site residues move independently from the rest of the protein in a 200ns molecular dynamics

simulation of cytochrome P450 CYP119. Archives of biochemistry and biophysics 2011; 509(2):127-132.

8) Brandman R, Brandman Y, Pande VS. A-site residues move independently from P-site residues in all-atom molecular dynamics simulations of the 70S bacterial ribosome. PloS one 2012; 7(1): e29377.

9) Poincaré H. Sur le problème des trois corps et les équations de la dynamique . Acta Mathematica 1890; 13(1): 3-270.

10) Eckmann JP, Kamphorst SO, Ruelle D. Recurrence plots of dynamical systems. Europhysics Letters 1987; 4(9): 973-977.

11) Webber CL, Zbilut JP. Dynamical assessment of physiological systems and states using recurrence plot strategies. Journal of Applied Physiology 1994; 76(2): 965-973.

12) Marwan N, Romano MC, Thiel M, Kurths J. Recurrence plots for the analysis of complex systems . Physics Reports 2007; 438(5): 237-329.

13) Giuliani A, Manetti C. Hidden peculiarities in the potential energy time series of a tripeptide highlighted by a recurrence plot analysis: a molecular dynamics simulation. Physical Review E 1996; 53(6): 6336-6340.

14) Manetti C, Ceruso MA, Giuliani A, Webber CL, JP Zbilut. Recurrence quantification analysis as a tool for characterization of molecular dynamics simulations. Physical Review E 1999; 59(1): 992-998.

15) Giuliani A, Benigni R, Sirabella P, Zbilut JP, Colosimo A. Nonlinear methods in the analysis of protein sequences: a case study in rubredoxins. Biophysical journal 2000; 78(1): 136-149.

16) Zbilut JP, Webber CL, Colosimo A, Giuliani A. The role of hydrophobicity patterns in prion folding as revealed by recurrence quantification analysis of primary structure. Protein engineering 2000; 13(2): 99-104.

17) Giuliani A, Sirabella P, Benigni R, Colosimo A. Mapping protein sequence spaces by recurrence quantification analysis: a case study on chimeric structures. Protein engineering 2000; 13(10): 671-678.

18) Giuliani A, Colafranceschi M, Webber CL, Zbilut JP. A complexity score derived from principal components analysis of nonlinear order measures. Physica A: Statistical Mechanics and its Applications 2001; 301(1): 567-588.

19) Webber CL, Giuliani A, Zbilut JP, Colosimo A. Elucidating protein secondary structures using alpha-carbon recurrence quantifications. Proteins: Structure, Function, and Bioinformatics 2001; 44(3): 292-303.

20) Manetti C, Giuliani A, Ceruso MA, Webber CL, Zbilut JP. Recurrence analysis of hydration effects on nonlinear protein dynamics: multiplicative scaling and additive processes. Physics Letters A 2001; 281(5): 317-323.

21) Giuliani A, Benigni R, Zbilut JP, Webber CL, Sirabella P, Colosimo A. Nonlinear signal analysis methods in the elucidation of protein sequence-structure relationships. Chemical Reviews-Columbus 2002; 102(5): 1471-1492.

22) Zbilut JP, Sirabella P, Giuliani A, Manetti C, Colosimo A, Webber CL. Review of nonlinear analysis of proteins through recurrence quantification. Cell biochemistry and biophysics 2002; 36(1): 67-87.

23) Giuliani A, Tomasi M. Recurrence quantification analysis reveals interaction partners in paramyxoviridae envelope glycoproteins. Proteins: Structure, Function, and Bioinformatics 2002; 46(2): 171-176.

24) Giuliani A, Benigni R, Colafranceschi M, Chandrashekar I, Cowsik SM. Large contact surface interactions between proteins detected by time series analysis methods: Case study on C-phycocyanins. Proteins: Structure, Function, and Bioinformatics 2003; 51(2): 299-310.

25) Zbilut JP, Colosimo A, Conti F, Colafranceschi M, Manetti C, Valerio MC, Webber CL, Giuliani A. Protein Aggregation/Folding: The Role of Deterministic Singularities of Sequence Hydrophobicity as Determined by Nonlinear Signal Analysis of Acylphosphatase and Aβ(1–40). Biophysical journal 2003; 85(6): 3544-3557.

26) Zbilut JP, Giuliani A, Colosimo A, Mitchell JC, Colafranceschi M, Marwan N, Webber CL, Uversky VN. Charge and hydrophobicity patterning along the sequence predicts the folding mechanism and aggregation of proteins: a computational approach. Journal of proteome research 2004; 3(6): 1243-1253.

27) Porrello A, Soddu S, Zbilut JP, Crescenzi M, Giuliani A. Discrimination of single amino acid mutations of the p53 protein by means of deterministic singularities of recurrence quantification analysis. Proteins: Structure, Function, and Bioinformatics 2004; 55(3): 743-755.

28) Li M, Huang Y, Xu R, Xiao Y. Nonlinear analysis of sequence symmetry of beta-trefoil family proteins. Chaos, Solitons & Fractals 2005; 25(2): 491-497.

29) Ming-Feng L, Yan-Zhao H, Yi X. Nonlinear correlations of protein sequences and symmetries of their structures. Chinese Physics Letters 2005; 22(4): 1006.

30) Colafranceschi M, Colosimo A, Zbilut JP, Uversky VN, Giuliani A. Structure-related statistical singularities along protein sequences: a correlation study. Journal of Chemical Information and Modeling 2005; 45(1): 183-189.

31) Zbilut J P, Chua GH, Krishnan A, Bossa C, Colafranceschi M, Giuliani A. Entropic criteria for protein folding derived from recurrences: Six residues patch as the basic protein word. FEBS letters 2006; 580(20): 4861-4864.

32) Grover A, Dugar D, Kundu B. Predicting alternate structure attainment and amyloidogenesis: A nonlinear signal analysis approach. Biochemical and Biophysical Research Communications 2005; 338(3): 1410-1416.

33) Huang Y, Li M, Xiao Y. Nonlinear analysis of sequence repeats of multi-domain proteins. Chaos, Solitons & Fractals 2007; 34(3): 782-786.

34) Zhou Y, Yu Z, Anh V. Cluster protein structures using recurrence quantification analysis on coordinates of alpha-carbon atoms of proteins. Physics Letters A 2007; 368(3): 314-319.

35) Mitra J, Mundra P, Kulkarni BD, Jayaraman VK. Using recurrence quantification analysis descriptors for protein sequence classification with support vector machines. Journal of Biomolecular Structure and Dynamics 2007; 25(3): 289-297.

36) Karakasidis TE, Fragkou A, Liakopoulos A. System dynamics revealed by recurrence quantification analysis: Application to molecular dynamics simulations. Physical Review E 2007; 76(2): 021120.

37) Giuliani A, Krishnan A, Zbilut JP, Tomita M. Proteins as networks: usefulness of graph theory in protein science. Current Protein and Peptide Science 2008; **9**(1): 28-38.

38) Krishnan A, Giuliani A, Zbilut JP, Tomita M. Implications from a network-based topological analysis of ubiquitin unfolding simulations. PloS one 2008; 3(5): e2149.

39) Angadi S, Kulkarni A. Nonlinear signal analysis to understand the dynamics of the protein sequences. The European Physical Journal-Special Topics 2008; 164(1): 141-155.

40) Yang Y, Tantoso E, Li K. Remote protein homology detection using recurrence quantification analysis and amino acid physicochemical properties. Journal of Theoretical Biology 2008; 252(1): 145-154.

41) Karnik S, Prasad A, Diwevedi A, Sundararajan V, Jayaraman V. Identification of Defensins Employing Recurrence Quantification Analysis and Random Forest Classifiers. Pattern Recognition and Machine Intelligence 2009; 5909:152-157.

42) Yang JY, Peng ZL, Yu ZG, Zhang RJ, Anh V, Wang D. Prediction of protein structural classes by recurrence quantification analysis based on chaos game representation. Journal of Theoretical Biology 2009; 257(4): 618-626.

43) Namboodiri S, Verma C, Dhar PK, Giuliani A, Nair AS. Application of Recurrence Quantification Analysis (RQA) in Biosequence Pattern Recognition. Advances in Computing and Communications 2011; 190: 284-293.

44) Kulkarni, A., Shreyas Karnik, Savita Angadi. Analysis of intrinsically disordered regions in proteins using recurrence quantification analysis. International Journal of Bifurcation and Chaos 2011; 21(04): 1193-1202.

45) Han GS, Zu-Guo Yu, Vo Anh. Predicting the subcellular location of apoptosis proteins based on recurrence quantification analysis and the Hilbert-Huang transform. Chinese Physics B 2011; 20(10): 0504.

46) Namboodiri S, Alessandro Giuliani, Achuthsankar S. Nair, Pawan K. Dhar. Looking for a sequence based allostery definition: a statistical journey at different resolution scales. Journal of Theoretical Biology 2012; 304:211-218.

47) Shao G, Yuehui C. Predict the tertiary structure of protein with flexible neural tree. Intelligent Computing Theories and Applications 2012; 7390: 324-331.

48) Takens, F. Detecting strange attractors in turbulence . In: Dynamical systems and turbulence, Warwick 1980. Springer Berlin Heidelberg; 1981. P 366-381.

49) Anastasios AT. Reconstructing dynamics from observables: the issue of the delay parameter revisited. International Journal of Bifurcation and Chaos 2007; 17(12): 4229-4243.

50) Kennel MB, Reggie Brown, Henry DI Abarbanel. Determining embedding dimension for phase-space reconstruction using a geometrical construction. Physical review A 1992; 45(6): 3403-3411.

51) Grassberger P, Thomas Schreiber, Carsten Schaffrath. Nonlinear time sequence analysis. International Journal of Bifurcation and Chaos 1991; 1(03): 521-547.

52) Bandt Christoph, Groth A, Marwan N, Romano M C, Thiel M, Rosenblum M, Kurths J. Analysis of bivariate coupling by means of recurrence. In: Mathematical Methods in Signal Processing and Digital Image Analysis. Springer Berlin Heidelberg; 2008. p153-182.

53) Goswami B, Ambika G, Marwan N, Kurths J. On interrelations of recurrences and connectivity trends between stock indices. Physica A: Statistical Mechanics and its Applications 2012; 391(18): 4364-4376.

54) Rangaprakash D, Xiaoping Hu, and Gopikrishna Deshpande. Phase synchronization in brain networks derived from correlation between probabilities of recurrences in functional MRI data. International journal of neural systems 2013; 23(02).

55) Thiel M, Romano M C, Kurths J, Rolf M, Kliegl R. Twin surrogates to test for complex synchronization. EPL (Europhysics Letters) 2006; 75(4): 535-541.

56) Strynadka Natalie CJ, Jensen SE, Johns K, Blanchard H, Page M, Matagne A, Frere JM, James MNG. Structural and kinetic characterization of a -lactamase-inhibitor protein. Nature 1994; 368(6472): 657-659.

57) Gretes M, Lim DC, de Castro L, Jensen SE, Kang SG, Lee KJ, Strynadka NC. Insights into Positive and Negative Requirements for Protein–Protein Interactions by Crystallographic Analysis of the β-Lactamase Inhibitory Proteins BLIP, BLIP-I, and BLP. Journal of molecular biology 2009; 389(2): 289-305.

58) James CP, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. Journal of Computational Chemistry 2005; 26:1781-1802.

59) HumphreyW, Dalke A, Schulten K. VMD - Visual Molecular Dynamics. J. Molecular Graphics 1996; 14: 33-38.

60) Darden, T, Darrin Y, Pedersen L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. The Journal of chemical physics 1993; 98:10089-10092.

61) Glykos, NM. Software news and updates carma: A molecular dynamics analysis program. Journal of computational chemistry 2006; 27(14): 1765-1768.

62) Stögbauer H, Kraskov A, Astakhov S A, Grassberger P. (2004). Least-dependent-component analysis based on mutual information. Physical Review E, 2004; 70(6):066123.

63) Kraskov A, Stögbauer H, Grassberger P. Estimating mutual information. Physical Review E 2004; 69(6): 066138.

64) VRA. http://softadvice.informer.com/Vra_Eugene_Kononov.html

65) Marwan, N. (2010). "Cross Recurrence Plot Toolbox for Matlab, v 5.15."

66) MATLAB version 7.2.0.232. Natick, Massachusetts: The MathWorks Inc., 2006.

67) Luo J, Thomas CB. Ten-nanosecond molecular dynamics simulation of the motions of the horse liver alcohol dehydrogenase· PhCH2O− complex. Proceedings of the National Academy of Sciences 2002; 99(26): 16597-16600.

68) Strynadka NCJ, Jensen S E, Alzari PM, James MNG. Nat. Struct. Biol.1996; 3:290–297.

Figure Legends

Figure 1. The probability of recurrence (P(τ)) for residue 49 in BLIP, as a function of the delay time steps.

Figure 2. The structure of the BLIP protein(PDB entry 3gmu).

Figure 3. The inter-residue distance matrix for BLIP. The distances(in Å) are calculated between the C_α atoms.

Figure 4.The DCCM correlation map for BLIP. The secondary structure elements are shown. The color-map covers correlation values between -1 and 1.

Figure 5. The MI correlation map for BLIP. The secondary structure elements are shown. The color-map covers correlation values between 0 and 1. The diagonal values are set to zero to improve color contrast.

Figure 6. The CPR correlation map for BLIP. The secondary structure elements are shown. The color-map covers correlation values between 0 and 1. The diagonal values are set to zero to improve color contrast.

Figure 7. The CPR correlation results vs. BLIP inter-residue distance.

Figure 8. The MI correlation results vs. BLIP inter-residue distances.

Figure 9. DCCM correlation results vs. BLIP inter-residue distances.

Figure 10. The zscore values for the CPR values between residues 40-50 and 130-150 in BLIP. The vertical dashed line denotes the test significance level, p=0.01.