Published in final edited form as:
J Chem Theory Comput. 2012 September 11; 8(9): 3009–3014. doi:10.1021/ct300565f.

Detection of long-range concerted motions in protein by a distance covariance

Amitava Roy and **Carol Beth Post**
Medicinal Chemistry and Molecular Pharmacology, Purdue University, West Lafayette, USA

Amitava Roy: amitroy@purdue.edu

Abstract

We asses the ability of a distance correlation coefficient (DiCC), calculated from distance covariance, for detecting long-range concerted motion in proteins. We establish a set of criteria for ideal correlation coefficient values based on the coefficient of determination in multi-dimension, $\mathbf{R}^2$. We compare in detail DiCC and conventional coefficients against these criteria. We demonstrate that in contrast to conventional correlation coefficients, which capture long-distance correlation adequately only with certain restrictions in multi-dimension, DiCC reflects appropriate correlation in both one- and multi-dimension. Finally we demonstrate the usefulness of DiCC for assessing long-distance correlated fluctuation in protein dynamics.

1 Introduction

Concerted, low-frequency motions are inherent to large or multi-domain proteins, and can be essential for proteins to carry out their function; [1] particularly those involving allosteric processes. Large-scale, concerted motion implies correlated fluctuation of different parts of the protein separated by relatively long-distance. Atomistic molecular dynamics (MD) simulation of proteins has the potential for revealing concerted motions in great detail. Nonetheless the assessment of long-range correlated motion from simulations has so far been elusive except for a small number of cases.[2] A correlation coefficient (CC) can be defined to quantify correlation between two random variables, including atomic fluctuations in the case of proteins. The most widely used CC between scalar variables is Pearson's correlation coefficient (PCC). The displacement vector correlation coefficient (VCC) is an extension of PCC to quantify correlation between two positional vectors. Some recent insightful usages of VCC are reported in references. [3–5] VCC depends on the cosine of the angle between the vectors and is most sensitive when the vectors are parallel. [2,6,7] To overcome this shortcoming of VCC, a few studies have used the generalized correlation co-efficient (GCC)[7–9] or radial correlation coefficient (RCC) [2] to detect correlation between atomic fluctuations of proteins. In previous work, [2] we exploited the radial symmetry of icosahedral viral capsids and found long-range correlated motions between residues 55 Å apart in human rhinovirus using RCC, which is a PCC on the norm of position vectors. RCC is highly useful when applied to systems with radial symmetry, but is insensitive to azimuthal fluctuation. GCC is an excellent CC between scalar random variables, however in multi-dimensions GCC does not combine the one-dimensional CCs in a suitable way to

Correspondence to: Amitava Roy, amitroy@purdue.edu.

Supporting Information Available

A detailed discussion of relation between generalized correlation coefficient and canonical correlation coefficient, correlation between components of position vectors calculated form MD simulation of Src SH2, and comparison of VCC & DiCC, VCC & GCC of pairs of C_α atoms calculated from the MD simulations are given in Supporting information. This material is available free of charge via the Internet at http://pubs.acs.org/.

investigate concerted motions. In this article we asses the ability of a distance correlation coefficient (DiCC), [10,11] calculated from distance covariance, to capture correlation without imposing any assumption on the time series of the vectors. A comparison of DiCC with VCC, RCC and GCC elucidates the merit and weaknesses of each and the potential of DiCC for detecting long-range concerted motion in proteins.

2 Results

2.1 Correlation coefficients

DiCC between two vector series, {**A**} and {**B**}, is defined as

$$DiCC=\frac{\nu(\mathbf{A}.\mathbf{B})}{\sqrt{\nu(\mathbf{A}.\mathbf{A})\nu(\mathbf{B}.\mathbf{B})}}, \quad (1)$$

where $\nu(\mathbf{A}.\mathbf{B})$ is the distance covariance between the vectors. Let us assume that the vector series, {**A**} and {**B**} have n entries each and the ith entry in {**A**} is denoted by $\mathbf{A}^i$. If {**A**} and {**B**} are position vectors of two atoms from a simulation study then $\mathbf{A}^i$ is the ith saved position vector of one atom. Distance covariance is defined as

$$\begin{gathered}\nu(\mathbf{A}.\mathbf{B})=\sqrt{\frac{1}{n^2}\sum_{ij}\alpha_{ij}\beta_{ij}}\\ \text{where.}\\ \alpha_{ij}=a_{ij}-a_{i.}-a_{.j}+a_{..}\end{gathered} \quad (2)$$

The following steps are needed to calculate α_{ij} from {**A**}.

1. Build the $n \times n$ matrix, **a**, from {**A**}, where a_{ij} is the distance between the ith and jth entries of {**A**}: $a_{ij} = |\mathbf{A}^i - \mathbf{A}^j|$.
2. Average the rows of **a**: $a_{i.}=\frac{1}{n}\sum_j a_{ij}$.
3. Average the columns of **a**: $a_{.j}=\frac{1}{n}\sum_i a_{ij}$.
4. Average all elements of **a**: $a_{..}=\frac{1}{n^2}\sum_{ij} a_{ij}$.
5. Build the $n \times n$ matrix $\boldsymbol{\alpha}$ from **a** where $\alpha_{ij} = a_{ij} - a_{i.} - a_{.j} + a_{..}$.

VCC and GCC between the vector series {**A**} and {**B**}, and RCC between $\{A_r\}$ and $\{B_r\}$, the norms of **A** and **B** respectively, are defined as follows.

$$\begin{gathered}VCC=\frac{\langle(\mathbf{A}-\langle\mathbf{A}\rangle)\cdot(\mathbf{B}-\langle\mathbf{B}\rangle)\rangle}{\sqrt{\langle(\mathbf{A}-\langle\mathbf{A}\rangle)^2\rangle\langle(\mathbf{B}-\langle\mathbf{B}\rangle)^2\rangle}}\\ GCC=\sqrt{1-e^{\frac{-2I}{d}}}\\ RCC=\frac{\langle(A_r-\langle A_r\rangle)(B_r-\langle B_r\rangle)\rangle}{\sqrt{\langle(A_r-\langle A_r\rangle)^2\rangle\langle(B_r-\langle B_r\rangle)^2\rangle}},\end{gathered} \quad (3)$$

where $\langle\ldots\rangle$ is the ensemble average or average over all entries in {**A**} and {**B**}, I is the mutual information between {**A**} and {**B**}, calculated using the method developed by Kraskov, Stogbauer and Grassberger, [12,13] and d is the dimension of vectors **A** and **B**.

It should be noted that calculation of VCC and GCC require the dimensions of **A** and **B** to be the same, while the calculation of RCC and DiCC does not impose any such restriction.

2.2 Coefficient of determination

If the dependency between two scalar random variables is known, then the coefficient of determination,[14] R^2, can be considered a measure of correlation between the variables. In the case of a linear dependency R^2 is 1 when the variation in one of the variables can be determined exactly by the variation in the other, and R^2 is zero when the variation in one cannot be determined at all by the variation in the other. If the dependency is non-linear we can refer to Nagelkerke. [15]

R^2 between two scalar variables is a scalar quantity. In the case of two vectors of dimension m and n, we can define $\mathbf{R}^2$ as a $m \times n$ matrix where R^2_{ij}, the i jth component of the matrix, is the coefficient of determination between the ith component of one vector and the jth component of another vector. An example of such a matrix is given later in the article.

A CC indicates the strength of the relationship between random variables. For a CC to be practical, physically meaningful and robust in the context of atomic fluctuations, it should satisfy the following criteria.

- Be a scalar quantity.
- Equal 1 when $\mathbf{R}^2$ is a unity matrix, and the dependency between the random variables is linear.
- Equal 0 when $\mathbf{R}^2$ is a null matrix.
- If $\mathbf{R}^2$ between a pair of vectors is identical to $\mathbf{R}^2$ between another pair of vectors, then the CC should be similar in both cases.
- Be independent of coordinate system.

While we were developing this assessment based on R^2 of coefficients for detecting concerted motions in proteins, a study appeared [16] in which similar criteria were proposed to establish associations between scalar data sets. Reshef et al. showed that for non-linearly dependent random variables, none of the established CCs becomes 1 even when R^2 is 1, and the sensitivity of CC calculated with different methods depends on the specific functional form of the dependency. [16] Accordingly for practical purposes we demanded the second criteria stated above be true only for linearly dependent random variables, although we would like it to be true in general.

2.3 Correlation coefficients in multi-dimensions

To compare the performances of different CCs, we calculated the CC between the positions of two particles A and B specified by their two-dimensional position vectors, **A** and **B**, as shown in 1.

We can write,

$$\begin{aligned}\mathbf{A}&=A_r\widehat{r}+A_\theta\widehat{\theta}=A_x\widehat{i}+A_y\widehat{j}.\ \text{and.}\\ \mathbf{B}&=B_r\widehat{r}+B_\theta\widehat{\theta}=B_x\widehat{i}+B_y\widehat{j}.\end{aligned} \qquad (4)$$

where, $\hat{i}$ and $\hat{j}$ are unit vectors in Cartesian coordinate system and $\hat{r}$ and $\hat{\theta}$ are unit vectors in the spherical coordinate system. The value of the CC obtained from the different parameters are compared to the known coefficient of determination between the components of the vector. If B_r can be expressed as a linear function of A_r, $f(A_r)$, then the coefficient of determination, $R^2(B_r.A_r)$ is[14]

$$\begin{aligned} R^2(B_r.A_r) \equiv 1-\frac{\sigma^2_{err}}{\sigma^2_{tot}}. \quad \text{where.} \\ \sigma^2_{tot}=\langle B_r-\langle B_r\rangle\rangle^2 \quad \sigma^2_{err}=\langle B_r-f(A_r)\rangle^2. \end{aligned} \qquad (5)$$

In the two-dimensional model we define

$$\begin{aligned} B_r=f(A_r)=A_r+\Delta r+\delta r \\ B_\theta=f(A_\theta)=A_\theta+\Delta\theta; \end{aligned} \qquad (6)$$

where Δr and $\Delta\theta$ are constants and δr is a random variable normally distributed, with mean zero and variance σ^2_r.

To build a series {**A**} we generated 100,000 normally distributed values of A_r with mean value of 10 and variance of $\sigma^2_{A_r}=36$. We fixed the value of A_θ to π. 4. We independently generated another 100,000 normally distributed values, with a mean of 0 and variance of $\sigma^2_r=16$, to build $\{\delta r\}$. For a particular value of $\Delta\theta$ we built {**B**} from 6 with $\Delta r = 3.0$. We generated 90 such series of {**A**} and {**B**} while varying the value of $\Delta\theta$ from 0 to π. 2. In all 90 series $\sigma^2_{tot}=\sigma^2_{A_r}+\sigma^2_r$ and $\sigma^2_{err}=\sigma^2_r$ in B_r. Hence $R^2(B_r.A_r)=1-\sigma^2_r.(\sigma^2_{A_r}+\sigma^2_r)=0.69$. Variances of B_θ and A_θ are zero as their values are fixed. Also, σ^2_{err} is zero in B_θ. We can still define $R^2(B_\theta.A_\theta)$ in such a case from the limit $\sigma^2_{tot} \to 0$, $R^2(B_\theta. A_\theta)$ becomes 1 when $\sigma^2_{err}=0$. Since angular and radial components are independent of each other, $R^2(B_\theta.A_r)$ and $R^2(B_r.A_\theta)$ are zero. So the expected R^2 between $\hat{r}$ and $\hat{\theta}$ components are

$$\mathbf{R}^2(\mathbf{B}.\mathbf{A})=\left(\begin{array}{ccc} & A_r & A_\theta \\ B_r & 0.69 & 0 \\ B_\theta & 0 & 1 \end{array}\right)$$

We calculated the DiCC, VCC, and GCC of (**B.A**) and DiCC, RCC and GCC of $(B_r.A_r)$. The values of the correlation coefficients between (**B.A**) are plotted as a function of $\Delta\theta$ in 2. For reference, the PCC of two linearly dependent random scalars is equal to $\sqrt{R^2}$, the square root of the coefficient of determination between them, which is 0.83 here. DiCC of $(B_r.A_r)$ and (**B.A**), dotted and solid blue lines in 2 respectively, have identical values of 0.81. Uncertainty in determining B_r from A_r in one dimension and **B** from **A** in multi-dimension appears only due to the random variable δr and the DiCC values in one dimension and multi-dimension correctly reflects that. In 2, RCC of $(B_r.A_r)$ (red dotted line) and VCC when $\Delta\theta = 0$ (green solid line) become exactly $\sqrt{R^2}(B_r.A_r)$. VCC, however, decreases monotonically to 0 as $\Delta\theta$ increases from 0 to π. 2 and changes sign for π. 2 < $\Delta\theta$ < π. In multi-dimensions, VCC between two random vectors is the the VCC value when the vectors are parallel multiplied by the cosine of the angle between them.

RCC is independent of $\Delta\theta$ and reproduces R; however it depends on the position of the origin of the coordinate system as the definition of the radial component of motion depends on the position of the origin. To illustrate this limitation we used the series of {**A**} and {**B**} for $\Delta\theta = \pi$. 2 and calculated RCC of $(B_r.A_r)$ while moving the origin along the x-axis. The inset of 2 shows how RCC changes as a function of the position of the origin on the x-axis of original coordinate frame.

That a GCC-like quantity can be used to define correlation between Gaussian random scalars was first suggested by Joe. [17] For this case GCC of (B_r, A_r) becomes exactly $R(B_r, A_r)$ as evident in 2. However in multi-dimension GCC of (**B.A**) is much higher even though the source of uncertainty in determining B_r from A_r is the same in determining **B** from **A**. If $\mathbf{R}^2$ matrix is diagonal and λ_i^2 are it's diagonal elements then GCC= $\sqrt{1-(\prod_i (1-\lambda_i^2))^{\frac{1}{d}}}$, where d is the dimension of the vectors. The derivation and the physical meaning of the above relation is explained in the Supporting information Note 1.

Accordingly GCC of $(\mathbf{B}.\mathbf{A})=\sqrt{1-((1-0.69)\,(1-1))^{\frac{1}{2}}}=1$. Irrespective of the value of $R^2(B_r, A_r)$, GCC of (**B.A**) is 1 as $R^2(B_\theta, A_\theta)$ = 1. The scheme with which GCC combines one-dimensional PCC values is not suitable to find association between positional fluctuations.

In the model illustrated with 1, $\rho_{A_xA_y}$ and $\rho_{B_xB_y}$ are close to one. In protein dynamics however CCs between the components of a vector are usually much smaller. Distribution of CCs between the components of position vectors calculated from a protein dynamics simulation is given in the Supporting information Note 2 and Supporting information Figure S1. To check the performance of DiCC and GCC as the correlation between the components changes, we generated (**B.A**) with

$$\mathbf{R}^2(\mathbf{B}.\mathbf{A})=\left(\begin{array}{ccc} & A_x & A_y \\ B_x & 0.49 & 0.49 \\ B_y & 0 & 0 \end{array}\right)$$

while varying the correlation between $\rho_{A_xA_y}$ from 0.0 to 0.99.

DiCC of (**B**,**A**), shown in solid blue line in 3, changes very slightly as $\rho_{A_xA_y}$ varies from 0 to 0.99. GCC of (**B**,**A**), shown in solid red line in 3, is 0.93 for $\rho_{A_xA_y}$ equal to 0 and approaches DiCC of (**B**,**A**) as $\rho_{A_xA_y}$ tends toward 1. For the same $\mathbf{R}^2$ matrix in 3 DiCC of (**B**,**A**) varies slightly, from 0.63 to 0.52, while GCC by contrast varies greatly from 0.93 to 0.52.

2.4 Long-range correlated fluctuations in protein

We further compared the capability of the various CC parameters using the example of Src SH2 domain in complex with a conformationally constrained mimetic of a phosphotyrosyl tetra-peptide ligand pYEEI,[18,19] and show that DiCC detects long-range concerted motions that are underestimated by VCC. VCC and DiCC were calculated between 106 C_α atoms from the cumulative 80,000 conformations. In 4a, VCC and DiCC are plotted against the average i.stance between the C_α pairs. The DiCC values are overall much greater than VCC values. No C_α pairs, with average distances between them greater than 7.5 Å, have a VCC value greater than 0.6. On the other hand, there are more than 40 pairs of C_α pairs, shown as circles in 4a, separated by more than 7.5 Å and have DiCCs greater than 0.6. One C_α pair, shown as a diamond in 4a, with an average distance equal to 24.8 Å has a DiCC value of 0.58.

The GCC values are also much greater than VCC values, and the distribution is less disperse (4b, 5). The GCC value is also greater in general than the DiCC value (4c, 5). The increased value of GCC arises because GCC is dominated by the largest element of $\mathbf{R}^2$ calculated in a coordinate system where vector components are independent, as explained earlier and in the Supporting information Note 1. Accordingly GCC, without an effective scheme of combining one-dimensional PCC values, does not characterize correlated behavior in a manner suitable to find association between positional fluctuations. GCC reflects some kind of correlation between random vectors, but it is not clear given the tight distribution in 4b and the variation with respect to non-independent vector components (3), how useful it is to detect long-range concerted fluctuations in protein dynamics.

2.5 Convergence of correlation coefficient values

We investigated convergence behavior of different CCs of five pairs of C_α atoms whose DiCC values fall within 0.9–1.0, 0.8–0.9, 0.7–0.8, 0.6–0.7 & 0.5–0.6 respectively and have the highest intra-pair average distances among all C_α pairs with DiCC in their respective ranges. The five pairs of C_α atoms are from residues 164 & 165, residues 153 & 154, residues 206 & 215, residues 206 & 216 and residues 193 & 203 with average intra-pair distances 3.87 Å, 3.85 Å, 11.13 Å, 11.08 Å & 24.71 Å respectively. We combined n ps form the beginning of the 40 trajectories and calculated VCC, GCC and DiCC and their bootstrap standard deviation of the five C_α pairs from the combined data while varying n from 10 to 2000 with a step of 20. Mean values of CCs calculated from the combine data stabilizes with 40 × 400 ps of data (6). Standard deviations calculated from 400 bootstrap sample change by less than 0.01 for all CCs during last 40 × 1 ns of data (6). Convergence behavior of all CCs are similar and they converge well with 40 × 1 ns of data.

3 Discussion

A correlation coefficient should be able to characterize correlation between displacement vectors due to concerted motion in a protein regardless of the distance of separation. VCC depends on the angle between the position vectors and hence underestimates correlation when vectors are not parallel. While RCC is highly suitable for detecting radial motion in spherically symmetric system, it depends on the position of origin of the coordinates axis and is insensitive to azimuthal correlation. When radial symmetry does not dominate the concerted motion, RCC does not reflect the full correlation between positional vectors.

The CC best matching the criteria outlined above is DiCC calculated from distance covariance. DiCC was found here to capture the true correlation between positional vectors based on agreement with R^2, is insensitive to the angle between the displacement vectors and has limited sensitivity to the dependence between the vector components. Further DiCC reflects both linear and non-linear correlation.[16] Using DiCC we observe long-distance concerted motions in a protein that was not revealed by VCC. Detection of such collective motion, which has mostly been elusive in analyses of molecular dynamics simulation, can be insightful for understanding allosteric function and other long-distance effects in proteins.

4 Methods

4.1 Generating correlated Gaussian

We determine $\{\mathbf{B}\}$ and $\{\mathbf{A}\}$ with a specified covariance matrix $\mathbf{C}$ between A_x, A_y, B_x by defining

$$\mathbf{C}=\mathbf{W}\mathbf{W}^{\dagger}$$

$$\begin{pmatrix} A_x \\ A_y \\ B_x \end{pmatrix}=\mathbf{W}\begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix} \quad (7)$$

where v_1, v_2 and v_3 are three independent Gaussian random variables with variance one and $\mathbf{W}^{\dagger}$ is the transpose of $\mathbf{W}$. The actual $R^2(A_x.B_x)$ and $R^2(A_y.B_x)$ calculated from the generated {**B**} and {**A**} were 0.48 to 0.50.

4.2 Molecular dynamics of Src SH2

In the simulation system the natural phosphotyrosine (pY) residue was replaced by one with the main-chain amide nitrogen, C_α and C_β substituted by a cyclopropane moiety which effectively constrains the side-chain conformation of the residue to that of the protein-bound state. [18,19] The details of the MD simulations of the Src SH2 complex have been reported previously. [19] Briefly, five sets of initial coordinates for the complex in explicit water were obtained from the multiple copies of the complex in the crystallographic asymmetric unit (PDB identifier 1IS0 and 1SPS). Eight simulations were initiated from each conformation by varying the initial velocities, yielding a total of 40 independent simulations. Each simulation was equilibrated for 500 ps and extended for 2 ns of production MD under constant temperature (298 K) and pressure (1 atm). Coordinates were saved at 1 ps interval from the production period.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the Rosen Center for Advanced Computing at Purdue University for providing computing resources. We thank Dr. Joshua M. Ward for allowing us to analyze Src SH2 complex trajectories from MD simulation. We also thank Jyotishka Dutta for his insightful discussion about R^2. This work was supported by National Institutes of Health Grant AI039639.

References

1. Berendsen HJC, Hayward S. Collective protein dynamics in relation to function. Curr Opin Struct Biol. 2000; 10:165–169. [PubMed: 10753809]
2. Roy A, Post CB. Long-distance correlations of rhinovirus capsid dynamics contribute to uncoating and antiviral activity. Proc Natl Acad Sci US A. 2012; 109:5271–5276.
3. Wereszczynski J, McCammon JA. Simulations of the p97 complex suggest novel conformational states of hydrolysis intermediates. Prot Sci. 2012; 109:475–486.
4. Tan YS, Fuentes G, Verma C. A comparison of the dynamics of pantothenate synthetase from M. tuberculosis and E. coli: Computational studies. Proteins. 2011; 79:1715–1727. [PubMed: 21425349]
5. Mishra S, Caflisch A. Dynamics in the active site of β-secretase: A network analysis of atomistic simulations. Biochemistry. 2011; 50:9328–9339. [PubMed: 21942621]
6. Ichiye T, Karplus M. Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. Proteins. 1991; 11:205–217. [PubMed: 1749773]
7. Lange FO, Grubmüller H. Generalized Correlation for Biomolecular Dynamics. Proteins. 2006; 62:1053–1061. [PubMed: 16355416]
8. Olbrich C, Strümpfer J, Schulten K, Kleinekathöfer U. Quest for spatially correlated fluctuations in the FMO light-harvesting complex. J Phys Chem B. 2011; 115:758–764. [PubMed: 21142050]

9. Kamberaj H, van der Vaart A. Correlated Motions and Interactions at the Onset of the DNA-Induced Partial Unfolding of Ets-1. Biophys J. 2009; 97:1747–1755. [PubMed: 19751680]

10. Székely GJ, Rizzo ML, Bakirov NK. Measuring and testing dependence by correlation of distances. Ann Stat. 2007; 35:2769–2794.

11. Székely GJ, Rizzo ML. Brownian distance covariance. Ann Appl Stat. 2009; 4:1236–1265.

12. Kraskov A, Stogbauer H, Grassberger P. Estimating mutual information. Phys Rev E. 2004; 69:066138.

13. [accessed Jun 13th, 2012] MIxnyn. http://www.klab.caltech.edu/kraskov/MILCA/

14. Rao, CR. Linear statistical inference and its application. Wiley & Sons; New York: 1965. p. 220

15. Nagelkerke NJD. A note on a general definition of the coefficient of determination. Biometrika. 1991; 78:691–692.

16. Reshef DN, Reshef YA, Finucane HK, Grossman SR, McVean G, Turnbaugh PJ, Lander ES, Mitzenmacher M, Sabeti PC. Detecting novel associations in large data sets. Science. 2011; 334:1518–1524. [PubMed: 22174245]

17. Joe H. Relative entropy measures of multivariate dependence. J Amer Stat Assoc. 1989; 84:157–164.

18. Davidson JP, Lubman O, Rose T, Waksman G, Martin SF. Calorimetric and structural studies of 1, 2, 3-trisubstituted cyclopropanes as conformationally constrained peptide inhibitors of Src SH2 domain binding. J Am Chem Soc. 2002; 124:205–215. [PubMed: 11782172]

19. Ward JM, Gorenstein NM, Tian J, Martin SF, Post CB. Constraining Binding Hot Spots: NMR and Molecular Dynamics Simulations Provide a Structural Explanation for Enthalpy -Entropy Compensation in SH2 - Ligand Binding. J Am Chem Soc. 2010; 132:11058–11070. [PubMed: 20698672]

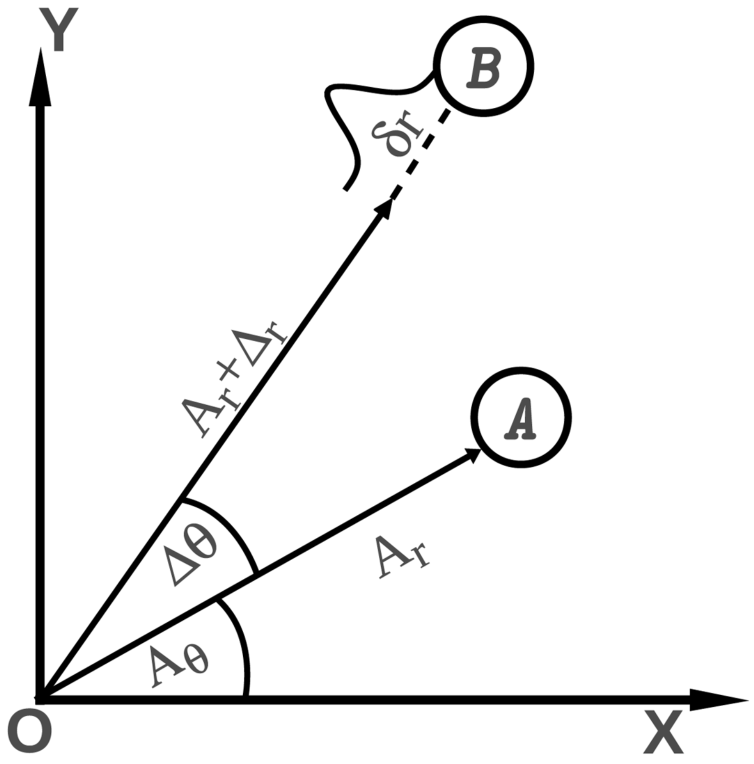

Figure 1.

A and *B* are two particles with their position vectors **A** and **B** respectively. $B_r = A_r + \Delta r + \delta r$ and $B_\theta = A_\theta + \Delta\theta$ where Δr and $\Delta\theta$ are two constants and δr is a normally distributed noise with mean zero and variance σ_r^2.

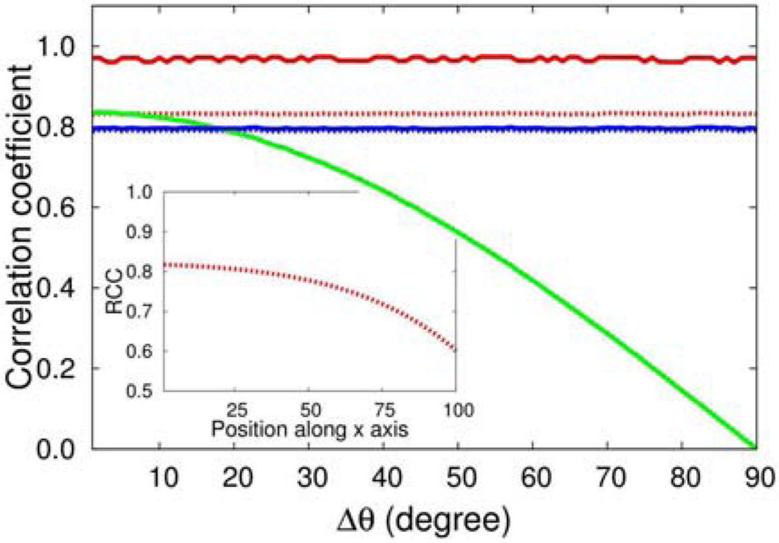

Figure 2.

Red dotted line represents GCC and RCC of ($B_r.A_r$). As both the values are very close to each other only one line is drawn for clarity. Blue solid and dotted lines represent DiCC of (**B.A**) and ($B_r.A_r$). Dotted blue line is hardly visible as it overlaps with the solid blue line. Solid red and green lines represent GCC and VCC of (**B.A**) respectively. VCC of (**B.A**) depends on $\Delta\theta$, the angle between **B** and **A**. The inset shows how the RCC of ($B_r.A_r$), when angle between (**B** and **A**) is π. 2, changes as the origin of the coordinate axis moves along x-axis. As the origin changes the radial component of the vectors decreases, as does RCC.

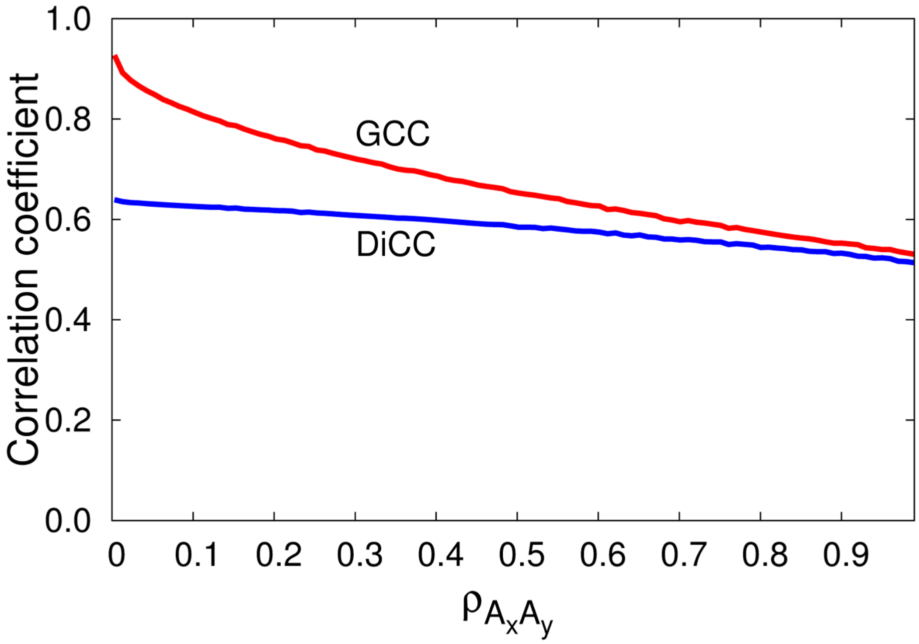

Figure 3.

Two random variables **A** and **B** are generated with $R^2(A_x.B_x)$=0.49, $R^2(A_y.B_x)$=0.49 and $R^2(A_x.B_y)$=0 and $R^2(A_y.B_y)$=0 while varying $\rho_{A_xA_y}$ from 0 to 0.99. Solid blue and red line represents DiCC and GCC of (**B**,**A**) respectively. DiCC of (**B**,**A**) changes very slightly as $\rho_{A_xA_y}$ varies from 0 to 0.99. GCC of (**B**,**A**) is 0.93 when $\rho_{A_xA_y}$ is equal to 0 and approaches DiCC of (**B**,**A**) as $\rho_{A_xA_y}$ tends toward 1.

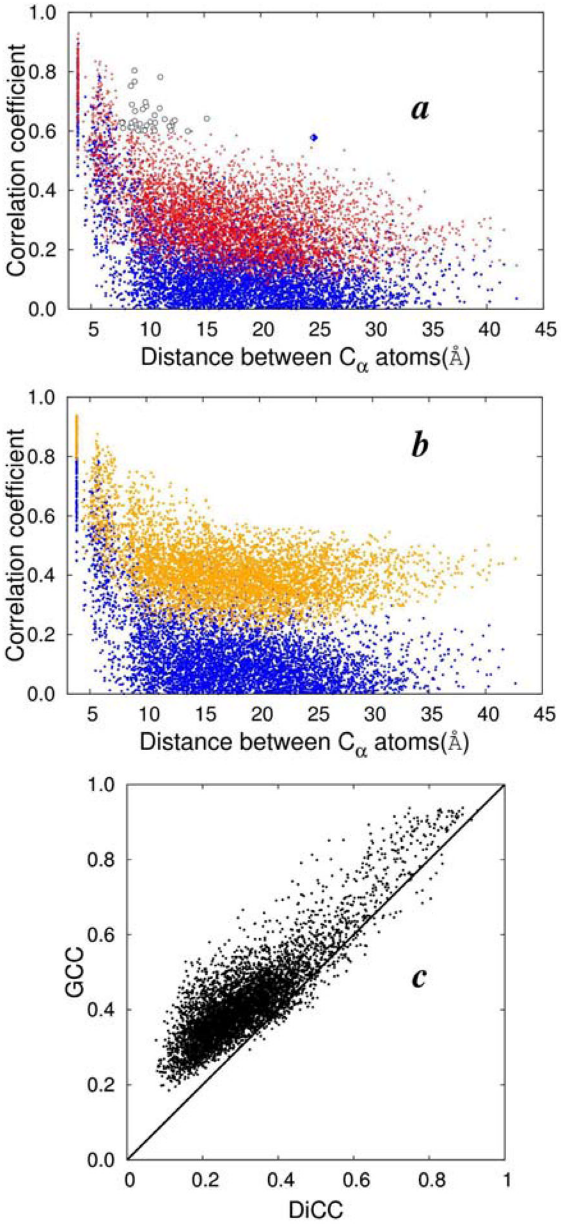

Figure 4.

(**a**) VCC (blue dots) and DiCC (red dots) between C_α atoms of from 40×2 ns long trajectory of Src SH2 domain in complex with the ligand pYEEI (see text for details) plotted against average distance between the C_α atoms. Distance correlation reveals more than 40 pairs of C_α atoms with average distances between them > 7.5 Å and DiCC > 0.6 (circles). One pair of C_α atoms, showed in diamond, has an average distance of 24.78 Å and DiCC of 0.58. DiCC reveals long distance correlations which are underestimated by VCC. (**b**) VCC (blue dots) and GCC (orange dots) of C_α atoms plotted against average distances between the C_α atoms. GCC does not reflect correlation suitable to investigate concerted fluctuation of positional vectors of C_α atoms. (**c**) DiCC and GCC of C_α atoms. Comparison between VCC & DiCC and VCC & GCC are given in Supporting information Figure S2.

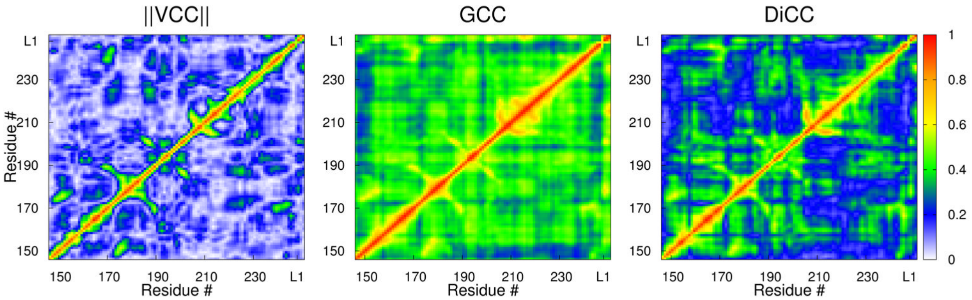

Figure 5.

||VCC|| (left), GCC (middle) and DiCC (right) between C_α atoms of Src SH2 domains complexed with ligand pYEEI. Last four residues, starting from L1, are from ligands.

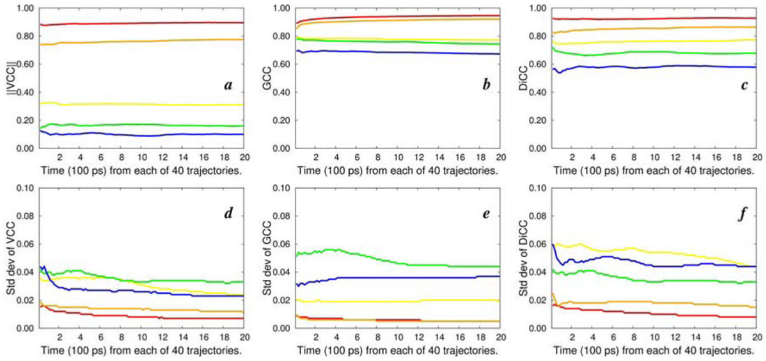

Figure 6.

Correlation coefficient of the position vectors of five pairs of C_α atoms were calculated using 10 to 2000 ps long trajectory from each of the 40 molecular dynamics simulation of Src SH2 domain. Five pairs of C_α atoms are from residues 164 & 165 (red), residues 153 & 154 (orange), residues 206 & 215 (yellow), residues 206 & 216 (green) and residues 193 & 203 (blue). Standard deviation of correlation coefficients were calculated from 400 bootstrap samples. Panel **a** and **d** show absolute mean value and standard deviation of VCC of five C_α pairs respectively as a function of time. Panel **b** and **e** show mean value and standard deviation of GCC of the same pairs respectively. Panel **c** and **f** show mean value and standard deviation of DiCC of the same pairs respectively.