An angle based variant of Dynamic cross correlation metric

Jyothish Soman  
Center for Security, Theory  
and Algorithmic Research  
International Institute of Information Technology  
Hyderabad, India 500032  
Email: jyothish@students.iiit.ac.in

Abhijit Mitra  
Center for Computational Natural Sciences  
and Bioinformatics  
International Institute of Information Technology  
Hyderabad, India 500032  
Email: abi_chem@iiit.ac.in

I. INTRODUCTION
Molecular dynamic simulation is a powerful tool for finding dependencies in biological molecules. Correlation matrices can map out the nature of correlated movements performed by different regions within the molecule during their solution state simulations and provide a critical tool in understanding the interactions between different residues in a structure. Here we present a methodology to extend a popular method for finding correlation by addressing to known problems with it. Dynamic cross correlation metric (DCCM) also known as Pearson’s coefficient [2] is the most commonly used method for finding correlation in data obtained from Molecular dynamics (MD) simulations. DCCM has well documented fallacies, which are fundamentally related to its closed definition of correlated motion, which focuses on linear correlation of motion. A description of these are found in [1], [4].

II. OUR EXTENSION
DCCM can be improved by a minor variations to the base method, that is an extension to the type of motions covered as correlated. Herein, DCCM can be extended to include a more generic definition of correlation, that is angular correlation. According to Angular correlation two points in motion have correlated motion if the variation in the angle made by the two vectors \( V^i_n \) and \( V^j_n \) across frames is low. Where \( V^i_n \) and \( V^j_n \) are the positional fluctuation vectors of residues i and j in the \( n^{th} \) frame.

A normalization based on the average angle, as well extreme values, provides a -1 to 1 range. The method requires a single, memory efficient, sweep of the data, hence is highly computationally efficient.

III. BASE ALGORITHM FOR ADCCM
The algorithm for finding Angular Dynamic cross correlation metric is presented in table 1.

Algorithm 1 Angular Cross correlation

- Find \( D_{ij}^n \), the angle between \( V^i_n \) and \( V^j_n \) in frame n using equation.

\[
D_{ij}^n = \cos^{-1} \frac{V^i_n \cdot V^j_n}{\det V^i_n \det V^j_n} 
\]  

- Find \( A_{ij} \), the average of \( D_{ij}^n \) across frames using equation.

\[
A_{ij} = \frac{\sum_{n=0}^{N} D_{ij}^n}{N} 
\]  

- Find \( C_{ij} \), the variance of \( D_{ij}^n \) across frames.

\[
C_{ij} = \frac{\sum_{n=0}^{N} (D_{ij}^n - A_{ij})^2}{N} 
\]  

- Find \( C_{s_{ij}} \), the signed correlation coefficient.

\[
C_{s_{ij}} = \frac{\cos A_{ij}}{\det(\cos A_{ij})} \cdot C_{ij} 
\]  

- Find max\_positive, the maximum positive value of \( C_{s_{ij}} \)

- Find max\_negative, the minimum negative value of \( C_{s_{ij}} \)

- Find \( C_{n_{ij}} \), the normalized correlation coefficient.

\[
C_{n_{ij}} = \begin{cases} 
1 - \frac{C_{s_{ij}}}{\max_{\text{positive}}} & \text{if } C_{s_{ij}} \text{ is positive} \\
\frac{C_{s_{ij}}}{\max_{\text{negative}}} - 1 & \text{if } C_{s_{ij}} \text{ is negative}
\end{cases}
\]  

deaminase A-riboswitch from Vibrio vulnificus. The cross-correlation coefficients are calculated for the centroids of nucleobases, as time averages over the simulation.

V. IMPROVEMENTS AND CONCLUSION
The algorithm was able to detect some major dependencies in the structure as compared to DCCM. For example, in closed
state, residue 13, which according to DCCM data was uncorrelated with residues 62 and 63, shows significant correlation in ADCCM map. In general, the boundaries between positively and negatively correlated regions in the ADCCM maps get sharper and conform far better with the expected scenario. A comparative picture of both DCCM and ADCCM are given in figure 1 and figure 2.

![Fig. 1](image1.png)

**Fig. 1.** Comparative plot for Angular DCCM (upper left triangle), and DCCM (lower right triangle) for aptamer domain of add A-riboswitch from Vibrio vulnificus for ligand-free (OPEN) state

![Fig. 2](image2.png)

**Fig. 2.** Comparative plot for Angular DCCM (upper left triangle), and DCCM (lower right triangle) for aptamer domain of add A-riboswitch from Vibrio vulnificus for ligand-bound (CLOSED) state

VI. Conclusion

We have presented here, a correlation metric which is an extension of pearson’s coefficient. It provides us enough flexibility to characterise and classify interactions from MD simulation data. The linear and non linear but correlated motions are accounted for.

VII. Acknowledgement

I thank Monika Sharma, Phd scholar, IIIT-Hyderabad for providing me with the data. Which helped me carry forward my work.

REFERENCES


