STRUCTURE-BASED MODELING AND DIRECT COUPLING ANALYSIS FOR THE PREDICTION OF DIMERIC MACROMOLECULAR SYSTEMS

Ricardo N. Santos[1,2 (PG)], **Faruck Morcos**[1 (PG)], **Biman Jana**[1 (PG)], **Adriano D. Andricopulo**[2 (PQ)], **José N. Onuchic**[1 (PQ)*]

[1] *The Center for Theoretical Biological Physics (CTBP), Biosciences Research Collaborative (BRC), Rice University, Houston, TX, USA.*

[2] *Laboratório de Química Medicinal e Computacional (LQMC), Centro de Pesquisa e Inovação em Biodiversidade e Fármacos (CIBFar), Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, SP, Brasil.*

Palavras Chave: *SBM, DCA, Structure Prediction, Coevolution, Dimers*

Introduction

Direct-Coupling Analysis (DCA) is a modern computational tool that enables the prediction of contacts between protein residues using only available amino-acid sequences of a target protein[1]. It is based on the hypothesis that mutations in residues that participate in specific interactions must be counterbalanced by other mutations to preserve protein structure and functionality[1]. A quantitative analysis of the direct correlation between residue mutations can provide interactions that are essential for protein function and gives fundamental clues to identify possible conformational states assumed by a protein.[1,2] A promising and powerful computational methodology to assist the understanding of target proteins behavior is the combination of a Structure-Based Models (SBM) and DCA. An SBM defines a specific energy potential that considers folded interactions from a particular experimental structure as parameters to guide molecular dynamics simulations. These approach allows the identification of hidden conformational states that occur in the functional landscape of certain proteins, but are not observed by crystallographic methods[1,3,4]. The combination of SBM and DCA has proven to be successful in predicting folded structures and functional conformational changes of several protein systems.

In this work, we have developed a procedure to expand the application of DCA/SBM methods to predict the association of protein structures into homodimers. Identification of dimerization contacts is more challenging than intradomain contacts since direct couplings can be confounded with internal contacts. Therefore, a systematic way to extract dimerization signals and consider them for molecular dynamics simulations is here proposed.

Results and Discussion

The dimeric interfaces of several protein systems with different sizes and complexity levels were successfully predicted using a combination of DCA and SBM methods. After separating identical monomers at a distance of 50 Å and reproducing an random orientation, the process of gradually reducing the equilibrium distance between DCA contacts in the simulation successfully recovers the correct orientation of both molecules upon binding (Fig.1). The analysis of the RMSD between the original and the predicted dimers (Fig.1) shows that the system gradually directs towards the correct homodimers followed by a decrease in the equilibrium distance of the gaussian potentials during SBM molecular dynamics simulation (Fig. 1).

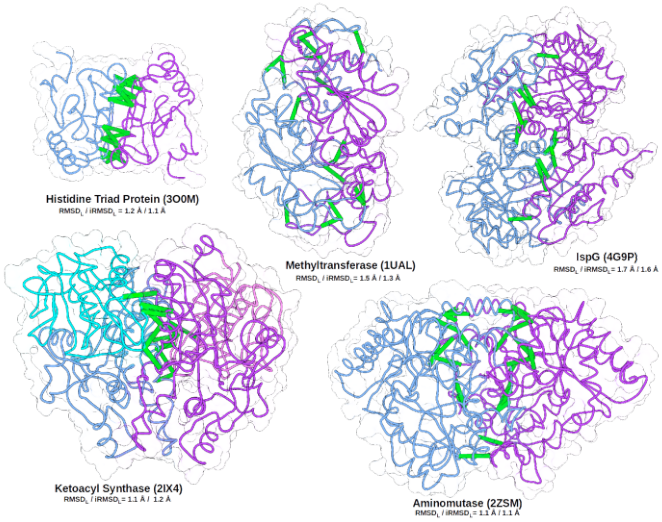

Figure 1. Predicted dimers for different systems using the DCA/SBM methodology.

Conclusions

Direct-Coupling Analysis (DCA) is a powerful computational tool for prediction of protein interactions. Here, we provide evidences that the prediction of homodimeric complexes using DCA is also possible with high accuracy. Furthermore, we presented a robust methodology that can be used to predict any homodimeric complex given enough data to infer coevolution. For six different systems with several complexity levels, the combination of DCA and SBM showed as a powerful method to predict interface contacts related to dimerization through molecular simulations. The identification of dimeric complexes can provide interesting molecular insights about protein function mechanism and also assist the design of novel compounds by elucidating inhibition mechanisms for protein interactions.

Acknowledgements

We gratefully acknowledge financial support from the National Science Foundation – NSF and FAPESP.

1 Morcos, F. *et al*. *PNAS*. **2011**, 108, E1293-301.

2 Sulkowska J. I. *et al*, *PNAS*. **2012**, 109, 10340-10345.

3 Whitford P. C. *et al*, *Proteins*. **2009,** 75, 430-441.

4 Morcos F. *et al*, *PNAS*. **2013**, 110, 20533-20538.