

## **Inferring direct couplings to unveil coevolutionary signals in protein 3D structure, interactions and recognition in signaling networks.**

Modern sequencing technologies provide us with a rich source of data about the evolutionary history of proteins. Inferring a joint probability distribution of amino acid sequences that are members of a protein family, signals from amino acid coevolution at different sequence positions in multiple sequence alignments can be exploited to infer spatial contacts within the three dimensional folded structure of proteins. We developed a computationally efficient algorithm termed mean field Direct Coupling Analysis (mfDCA) which has the ability to disentangle direct and indirect residue correlations for a large number of protein domains. Information uncovered by mfDCA allows us to reconstruct the structure of contact maps for many protein domains. Inferred contacts by mfDCA can be utilized as a reliable guide in high accuracy computational predictions of domain structure. Our results capture clear signals beyond intradomain residue contacts, for instance, interdomain interactions in macro molecular assemblies and diversity in conformational states. Beyond single proteins, we studied evolutionarily conserved protein-protein interactions, for which we developed a "direct information score" to quantify mutational changes in the interaction between phosphotransfer proteins in Two Component Signaling (TCS). Our score accurately correlates with mutagenesis studies that probe the mutational change in measured *in vitro* signaling. With our methodology we are able to isolate the determinants that give rise to interaction specificity and recognition among different TCS partners. Finally, our interprotein evolutionary couplings can also be used to reconstruct protein interaction complexes with high accuracy.