**[BIOINFO 2025]**

# Statistical sequence coevolution analysis of Intrinsically Disordered Proteins and Regions for binding functionalities

Joongyu Heo[1], Minjae Seo[1], Wookyung Yu[1,*]

*[1]Department of Brain Sciences, DGIST*
**Corresponding author: wkyu@dgist.ac.kr*

Intrinsically Disordered Proteins and Regions (IDP/Rs) do not have a fixed native structure under physiological conditions, yet their functionalities and binding partners are consistent. We took notice that the IDP/R sequence itself does not change throughout the binding interaction process, and utilized Statistical Coupling Analysis (SCA) to discover the sequence-function relationships of the IDP/Rs. SCA leverages residue coevolution patterns on the multiple sequence alignment (MSA) of the target protein to find statistically significant residue relationships and clusters.

We compared the coevolution and conservation pattern with the functional classification of 20 well-studied IDP/Rs based on their binding patterns. Non-binding entropic chain regions were predicted to require neither statistically significant conservation nor a coevolutionary relationship. In contrast, transient binding regions require only significant coevolution for their sub-optimal binding function, and permanent binding regions require both significant coevolution and conservation to maintain binding interaction in the long term. There were also complex edge cases of single regions showing multiple binding patterns, such as p21 and p27. Finally, to test whether these coevolution patterns are a valid signal of interaction to IDP/R binding partners, we conducted SCA on the joint MSA between IDP/Rs and their binding partners. We confirmed a coevolution signal between IDP/R residues and their interaction counterparts.