

## Abstract

Intrinsically disordered regions in proteins are known to mediate macromolecular interactions. Despite recent advances, it is yet not very clear how these regions recognize the specific targets without any defined three dimensional conformation. Here, we report a repertoire of evolutionary conserved de novo peptides in long disordered regions, which have strong tendency to retain certain preferred conformations. The peptides exhibit distinct amino acid propensities, particularly the enrichment of Gly, Ala, Arg, Lys, when compared to other known short linear motifs. The peptides show significantly conserved Ramachandran conformations, secondary structures and the three dimensional folds when compared to neighboring regions. Significant enrichment of DNA-binding and extracellular matrix binding functions of proteins with these peptides further explains the abundance of positively charged amino acids like Arginine and Lysine. DNA-binding function of motifs was further confirmed through prediction of DNA binding residues. Importantly, the non-synonymous single nucleotide mutations in the peptides are predicted to be highly intolerant for the protein function when compared to neighbouring regions, hinting at their indispensable function in proteins. Preferred left handed bridge-conformation of enriched Glycines in CoREs suggest that Gly-to-nonGly mutations within CoRE can alter the backbone conformation and consequently the function, a hypothesis that we reconciled using mutation data. Overall, our observations uncover an evolutionary strategy wherein certain set of peptides, which have strong tendency to retain their conformations are recognized and utilized in disordered regions for molecular recognition. We therefore, propose that these peptides could serve as anchors for initial recognition, followed by other binding and folding events, during macromolecular interactions of disordered proteins. Structured nature of CoREs suggests possibilities to inhibit the molecular interactions using small molecules mimicking CoRE conformations.