Repeat proteins are made up of tandem arrays of similar 20–40 amino acid stretches that usually fold up in elongated architectures mainly stabilized by local interactions. Due to their simple and linear structure that lacks of sequence-distant contacts, these proteins represent a useful model to study protein folding, dynamics and function.

Ankyrin Repeat Proteins (ARPs) are widely distributed in nature. Their ‘biological function’ is usually attributed as mediating specific protein-protein interactions with versatility for recognition paralleled to that of antibodies. These proteins usually show sophisticated behaviours that are in contrast with their apparently simple structure.

We have calculated and analyzed several parameters in order to dissect the energetic contributions that underlie the conformational transitions that contribute to ARPs function. We have collected, processed and depurated all available ARPs sequences, structures and other publicly available data to build a relational database from which we aim to extract information to better understand how these proteins work. We focused in localizing, analyzing and quantifying frustration in ARPs structures. Local frustration patterns, that arise from the heterogeneities in the energy landscape when a protein folds, have been found to be related to protein recognition\(^1\), allostery\(^2\) and folding mechanisms\(^3\). We localized the degree of local frustration manifested by ARPs structures and searched for correlations with other structure and sequence measures (disorder predictions, secondary structure propensity, sequence and structural alignments and others) in order to weight in a proper manner the contributions to individual repeats, the entire array of repeats and the complete polypeptide sequences. We also analyzed those proteins that were crystallized forming complexes with other proteins and we classified their interaction and recognition mechanisms trying to determine if the observed behaviours could be traced back to sequences and their environment.

With this work we aim to establish correlations with experimental information and find simple descriptions that help us to interpret ARPs functional behaviour and repeat proteins in general.

