

Summary

A recent, exciting development in structural biology is the recognition that many proteins (intrinsically disordered proteins, IDPs) or protein regions (intrinsically disordered regions, IDRs) exist and function without a well-defined structure. Being depleted in hydrophobic amino acids, these regions lack a stable core, which keeps them in a highly dynamic, heterogeneous structural state. Their conformational adaptability, extended interaction surface, various exposed binding and modification sites confer important regulatory, signalling and scaffolding roles on IDRs. Their structural freedom also provides increased tolerance against mutations, enabling fast evolutionary changes and exceptional adaptability. Studying IDRs is of prime importance as they play central roles in cellular interaction networks as well as in major human diseases, like cancer among others.

First, we provide an overview of the IDP literature, mainly focusing on the functional and evolutionary properties of IDRs, their binding modes, and the available bioinformatics tools to identify them. Next, we describe computational studies (Chapters 2-5), which address biological questions highlighting diverse aspects of the functional-evolutionary advantages disorder confers on proteins, pathways and proteomes.

Through a comprehensive analysis of the prevalence of IDRs in eukaryotic proteomes, we show that their abundance not only depends on the complexity of organisms, but also on the lifestyle. A decreased level of disorder seems to reflect adaptation to a static environment, whereas an elevated level indicates adaptation to a varied lifestyle with changing habitats. Overall, our results imply that unlike folded domains, structural disorder represents a rapidly evolving modality, enabling the fast appearance of novel functions.

After realizing the increased evolutionary plasticity of IDRs, we study how IDRs influence the functional-evolutionary properties of protein pathways. We demonstrate on the three main vesicle trafficking routes that the presence or absence of disordered protein segments can manifest in different evolutionary signatures even in pathways with very similar roles and architectures.

The functional innovative potential of disorder can also be tracked by analyzing the structural properties of protein extensions that result from stop codon readthrough in *Drosophila melanogaster*. We show that such extensions are enriched in disorder and interaction sites, indicating a high potential for the conditional rewiring of cellular interaction networks. We conclude that despite originating from a different molecular mechanism, readthrough extensions fulfil similar roles to protein segments encoded by tissue-specific exons.

We also demonstrate that IDRs not only excel in innovation owing to their protein-level functions, but due to their elevated mutation tolerance they enable the accommodation of nucleotide-level regulatory elements into coding regions, thereby contributing to complex eukaryotic gene regulation. This recognition stems from our analysis of genomic coding regions with diverse overlapping functions, showing that those tend to translate into protein regions rich in structural disorder, which are often of low compositional complexity.

Finally, in chapters 6 and 7 we describe the development and first application of our novel predictor, DynaMine. DynaMine is a linear regression approach trained on experimentally determined dynamics information on proteins in solution that can accurately predict protein backbone dynamics directly from protein sequence. We demonstrate that DynaMine has a great potential in distinguishing regions of different structural organization, such as folded domains, disordered linkers, molten globules, and pre-structured binding motifs. Also, it identifies protein disorder with an accuracy comparable to the most sophisticated existing predictors, but without

relying on prior knowledge on disorder. By doing so, it also firmly establishes the long anticipated link between dynamics of the polypeptide chain and structural disorder.

Lately, we assembled a comprehensive database of protein residues involved in early folding events based on dedicated experiments. Using these data, we demonstrate that residues predicted with a more rigid backbone by DynaMine correlate well with those initiating folding, and that this rigidity is conserved during evolution. Our results prove that backbone rigidity, as determined by local interactions within the protein chain, is a fundamental physical feature conserved by proteins, largely determining their folding mechanisms.

All these studies clearly and concisely demonstrate that structural disorder confers specific functional advantages on proteins, which seems critical in evolutionary innovation.