

The discovery of intrinsically disordered proteins encouraged structural biologists to revisit their views on the protein structure-function paradigm. These proteins are ubiquitously involved in cellular processes, and are often implicated in human pathological conditions, yet investigating their characteristics and describing their structural behavior is far from trivial; in particular, inferring the function(s) of a disordered protein region remains a major challenge. Computational methods are particularly well suited for investigating these proteins: on the sequence level their dependence on distinct characteristics determined by the local amino acid context make sequence-based prediction algorithms viable and reliable tools for large scale analyses, while on the structure level the in silico integration of fundamentally different experimental data types is essential to describe the behavior of a flexible protein chain. During the years of my PhD I worked on developing computational tools and online resources that are aimed at connecting protein function to intrinsic disorder both on the sequence and the structural level.

My contribution to studying disordered proteins on the sequence level is DisCons (<http://pedb.vib.be/discons>), a novel pipelined tool that combines the quantification of sequence- and disorder conservation to classify disordered residue positions. Residues that show conservation of both the sequence and the property of disorder are associated mainly with specific binding functionalities (e.g., short, linear motifs, SLiMs), whereas protein segments where only disorder as a feature is important for function as opposed to the identity of the underlying sequence generally serve as entropic chains and linkers.

On the structural level, I have developed the Protein Ensemble Database (PED) (<http://pedb.vib.be>) which is the only international initiative to store and provide access to structural ensembles of intrinsically disordered proteins (IDPs) based on nuclear magnetic resonance (NMR) spectroscopy and small-angle X-ray scattering (SAXS) data measured in solution. Due to the inherent flexibility of IDPs, solution techniques are particularly appropriate for characterizing their biophysical properties, and structural ensembles in agreement with these data provide a convenient tool for describing the underlying conformational sampling. The PED is already widely accessed by the scientific community, and it is expected to promote the development of new modeling methods and lead to a better understanding of how function arises from disordered states.

De ontdekking van intrinsiek ongevouwen eiwitten heeft structureel biologen ertoe aangezet om het structuur-functie paradigma van eiwitten te herbekijken. Deze eiwitten zijn alomtegenwoordig betrokken in cellulaire processen en zijn vaak ook verwickeld in pathologische omstandigheden. Toch is het bestuderen van hun eigenschappen en de beschrijving van hun structureel gedrag verre van voor de hand liggend; meer bepaald blijft het afleiden van de functies van een ongevouwen eiwitregio een moeilijke opgave. Desalnietemin zijn computationele methodes bijzonder geschikt voor de studie van deze eiwitten: op het sequentieniveau zorgen verscheidene karakteristieken, die afhangen van de lokale aminozuurcontext, ervoor dat sequentie-gebaseerde algoritmes bruikbare en betrouwbare instrumenten geworden zijn voor grootschalige analyses, terwijl op het structurele niveau de in silico integratie van fundamenteel verschillende experimentele datasets essentieel is om het gedrag van een flexibele eiwitketen in kaart te brengen. Tijdens mijn doctoraatsstudie heb ik bijgedragen tot de ontwikkeling van computationele methodes en online hulpbronnen teneinde eiwitfunctie en intrinsieke structurele wanorde te linken, zowel op het sequentieniveau als op het structurele niveau.

Mijn bijdrage tot de studie van ongevouwen eiwitten op het sequentieniveau betreft DisCons (<http://pedb.vib.be/discons>), een nieuwe pijplijn die de kwantificatie van sequentie- en wanorde-conservatie combineert om de classificering van ongestructureerde residu-posities toe te laten. Residuen die zowel conservatie van zowel sequentie als intrinsieke wanorde vertonen, zijn voornamelijk geassocieerd met specifieke bindingsfunctionaliteiten (bv. korte lineaire motieven (zoals SLiMs)), terwijl eiwitsegmenten waar enkel wanorde van belang is voor de functie en niet de identiteit van de onderliggende sequentie, algemeen als entropische keten en linkers dienst doen.

Wat het structurele niveau betreft, heb ik een de Protein Ensemble Database (PED; <http://pedb.vib.be>) ontwikkeld, wat tot op heden een uniek internationaal initiatief is voor het opslaan van en het toegang verschaffen tot de structurele ensembles van intrinsiek ongevouwen eiwitten op basis van nucleaire magnetische resonantie (NMR) spectroscopie en kleine hoek x-stralen scattering (SAXS). Ingevolge de inherente flexibiliteit van intrinsiek ongevouwen eiwitten zijn technieken die berusten op het in oplossing meten van eiwitten bijzonder geschikt om de biofysische karakteristieken op te tekenen, terwijl de structurele ensembles die met deze data overeenstemmen een handig instrument verschaffen om de onderliggende conformationele sampling te beschrijven. De PED is reeds ruim geconsulteerd door de wetenschappelijke gemeenschap en het valt te verwachten dat het de ontwikkeling van nieuwe modelleringsmethodes zal bevorderen evenals zal leiden tot een beter begrip hoe de eiwitfunctie zijn oorsprong vindt in de ongevouwen toestand. "