

## Proposal for a call for a Collaborative Project on Natively Unfolded Proteins

It has recently been recognized that conformational disorder may have an important functional role by allowing a protein to easily adapt to the different cellular stimuli. The extent of structural disorder may vary in a continuous way from one protein to another. At one extreme, there are proteins that have only a few specific regions experiencing conformational averaging while the remainder of the polypeptide chain has a stable three-dimensional structure. At the other extreme, there are proteins lacking an independently folded state. Several terms have been used to indicate these different situations: partially unfolded protein, molten globule, unfolded or unstructured protein. Indeed, a classification of unfolded protein states is difficult, as proteins span a continuum of properties such as compactness, secondary structure propensity and stability of tertiary contacts. Partially or fully unfolded proteins can be best described as ensembles of conformers in equilibrium.

The cellular role of natively unfolded proteins is linked to their structural features (or lack thereof) in a subtle way. So-called “intrinsically unfolded proteins” have been recently found to have a physiological function *in vivo* in a substantially unstructured state. In other cases, unfolded proteins experience a transition between a fully or partially unstructured conformation to a more defined three dimensional structure of a protein upon binding to physiological partners, leading to a fine regulation of their function. Knowledge of these interactions and of the associated structure-inducing mechanisms is important for the understanding of physiological roles of intrinsically unfolded proteins. In some cases also biological membranes play a role in these processes, which requires a proper simulation of the environment e.g. through detergent micelles. Contiguous to natively unfolded proteins are the flexible, structurally heterogeneous linkers connecting domains in multi-domain proteins. These linkers very often play an important physiological function by modulating the interaction between the domains and with the biological partners.

High-resolution NMR spectroscopy is the main technique to obtain information at atomic resolution on the structure and dynamics of the above-described systems. This information can be usefully supplemented by data obtained from other complementary biophysical techniques. It is important to note the features of the NMR spectra of intrinsically unfolded proteins (or protein regions such as linkers) are very different from those of proteins characterized by a well defined three-dimensional structure. Indeed, the interconversion between different conformers causes a large reduction in the chemical shift dispersion of different nuclei and the nuclear relaxation properties are significantly affected by the increased local mobility. Therefore, a thorough understanding of the cellular roles of the proteins addressed here, whose biological importance is just beginning to be recognized, will only be achieved after a significant development of the NMR-based methodology for their characterization.

We want to achieve these objectives by pursuing the following major areas of research:

*Protonless NMR methods, in combination with  $^1\text{H}$  based methods, to improve the available methodology for sequence specific assignment.*

Among the different nuclei suitable for NMR observation ( $^1\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ), the heteronuclei ( $^{13}\text{C}$ ,  $^{15}\text{N}$ ), are characterized by the largest residual chemical shift dispersion in unfolded systems. Moreover amide protons are absent in prolines, which are one of the most abundant residues in unfolded systems. Therefore the use of  $^{13}\text{C}$  direct detection NMR experiments in conjunction with  $^1\text{H}$  based

ones, will have a major impact in sequence specific assignment of unfolded systems where the  $^1\text{H}$  resonances are highly overlapped.

*Development of new NMR experiments to obtain “residual” structural and dynamic information.*

The experiments to obtain structural information for folded proteins can in principle be applied also to unfolded or partially folded proteins. However the observables will be averaged over the conformational ensemble characterizing the system, resulting in: a) small residual chemical shifts (a problem that becomes dramatic for  $^1\text{H}$  and for aliphatic  $^1\text{H}$  in particular), b) loss on information content of parameters reporting on local structure, such as  $^3J_s$  and intra-residues NOEs, 3) weakening or even quenching of NMR observables reporting on medium- and long-range structural information. We will thus develop new experiments, based either on novel detection schemes of “traditional” observables informative on structural properties or on the analysis of entirely new parameters (e.g. paramagnetic effects), to obtain insight into the residual structure and the structural propensities of (partly) unfolded systems.

*Development of software tools to interpret NMR observables on the basis of conformational ensembles rather than on a single structure*

No matter the experimental approach, the NMR parameters acquired will always be the result of an average over the different conformers in equilibrium. However, the availability of a wealth of observables of intrinsically different nature (local, long-range, global), will allow us to identify any key interatomic contacts and to define the structure propensity of the proteins investigated. To this end, specific computational tools must be developed.