New and Notable

Mapping the Conformational Mobility of Multidomain Proteins

Martin Blackledge*
Protein Dynamics and Flexibility, Institute de Biologie Structurale Jean-Pierre Ebel, Centre d'Etudes Atomiques, Centre National de la Recherche Scientique, UJF UMR 5075, Grenoble, France

The central principle justifying continued investment in structural genomic projects has been founded upon the assumption that the resolution of static three-dimensional structures of a finite number of proteins will provide the key to understanding biological activity. Over the last decade, biophysical characterization, allied with computer-based prediction, has convincingly shown that >30% of proteins present in the human proteome may contain continuous regions of at least 30 amino acids in length that are intrinsically disordered. This realization highlights a detail that may in hindsight appear evident; that the determination of the three-dimensional structure of folded globular domains, while essential, represents only one aspect of the quest to complete the proteomic structural map. The role that disorder plays in such proteins and the advantages it confers in terms of function are subjects of intense interest and debate, whose resolution represents one of the major challenges for contemporary structural biology. (1) In some cases, disordered regions provide entropic linkers or spacers, whose role appears to be to position different, possibly folded domains in three-dimensional space. Such multidomain systems with folded domains separated by flexible linkers are particularly prevalent in transcription and replication, cellular signaling and adhesion, and in viral proteins. The behavior of linker regions and the relationship between primary sequence and different levels of flexibility encoded in their primary sequence remain open and poorly understood aspects of structural biology. (2) There is therefore a compelling need for novel methodology to develop a full molecular understanding of these complex proteins.

Due to their very high degree of conformational flexibility, this is of course far from trivial. The most common approaches evoke explicit ensembles of rapidly interconverting conformers, based on agreement with extensive sets of experimental data measured in solution. NMR spectroscopy is without doubt the tool of choice for studying both local and long-range structural propensities at atomic resolution in intrinsically disordered proteins, while small angle scattering (SAS) provides essential information about overall molecular dimensions (3). The combination of complementary experimental NMR and SAS data with ensemble molecular descriptions has indeed allowed the characterization of entire proteins involving significant folded and unfolded domains (4). NMR data report on population-weighted averages of all conformations exchanging on timescales faster than the millisecond range, but ensemble descriptions of disordered chains rarely address the important question of dynamic interconversion rates beyond this simple limit. The dynamics and thermodynamics of conformational change are, however, of great importance if we are to develop a true statistical mechanical representation of the protein in solution.

NMR spin relaxation can, in principle, provide an answer to this question, by characterizing the mobility of interatomic bonds throughout the protein on timescales in the hundreds of picosecond to tens of nanosecond range. The interpretation of relaxation rates in disordered chains, or in domains connected by flexible linkers, is complicated by the potential coupling between motion of individual domains and overall motion of the entire multidomain protein. It has been shown, however, that if a molecular description of the protein ensemble is available, combination with relaxation data can provide information about reorientational eigenmodes present in the ensemble. (5) Timescales relevant to conformational interconversion can thereby be derived directly from the relaxation data.

In this groundbreaking study, the authors combine many of these approaches to study ribosomal protein L12, a two-domain protein that dimerizes via its N-terminal domain (NTD). (6) NTD is separated from the C-terminal domain (CTD) via a flexible 20-residue chain such that the protein can be described in terms of three folded domains, connected by a more flexible linker. The flexibility of the CTD has been related to control of ribosomal function. The authors use molecular modeling to sample the conformational space available to the three domains as extensively as possible. This sampling is expressed in terms of 10,000 dimeric conformers (derived from 2 million original conformers) selected on the basis of agreement with experimental SAS data. This ensemble contains the conformational distributions of interdomain distances, and defines the volume space occupied by the protein. Although it contains no information about the dynamic timescales, it is assumed, reasonably, to contain the reorientational properties of the protein. Timescale information can be provided from the NMR relaxation data analysis of the individual domains within the ensemble. The combined analysis reveals anticorrelated motions of the CTD and NTD dimer and provides novel insight into the behavior of the linker regions in solution. The authors also mention an important issue concerning the resolution of this kind of ill-defined inverse problem, where the number of conformational degrees of freedom far outweighs the ability of

Submitted February 26, 2010, and accepted for publication March 2, 2010.

*Correspondence: martin.blackledge@ibs.r

Editor: Edward H. Egelman.

0006-3495/10/05/2043/2 $2.00

doi: 10.1016/j.bpj.2010.03.005

the experimental data to uniquely restrain the system: The approach relies on the ability of the initial analysis to define a relevant ensemble of structures that accurately represents the behavior of the protein in solution. The next challenge lies in the development of robust validation techniques (some kind of R-factor) that will allow the estimation of levels of confidence in the proposed models.

The study clearly demonstrates how the complementary dependences of SAS and NMR relaxation data can be combined to extract information that is far beyond the sum of their parts. Similar approaches can be used to combine different types of NMR data, for example dipolar couplings, chemical shifts, or paramagnetic enhancements (4). Developments such as these are providing the blueprints for future studies of multidomain, highly flexible, and intrinsically disordered proteins, and thereby establishing the basic tools for completing our understanding of the conformational proteome.

REFERENCES

1. Radivojac, P., L. M. Iakoucheva, …, A. K. Dunker. 2007. Intrinsic disorder and functional proteomics. *Biophys. J.* 92:1439–1456.
2. Pickford, A. R., and I. D. Campbell. 2004. NMR studies of modular protein structures and their interactions. *Chem. Rev.* 104:3557–3566.
3. Bernadó, P., E. Mylonas, …, D. I. Svergun. 2007. Structural characterization of flexible proteins using small-angle x-ray scattering. *J. Am. Chem. Soc.* 129:5656–5664.
4. Jensen, M. R., P. R. L. Markwick, …, M. Blackledge. 2009. Quantitative determination of the conformational properties of partially folded and intrinsically disordered proteins using NMR dipolar couplings. *Structure.* 17:1169–1185.
5. Prompers, J. J., and R. Brüschweiler. 2002. General framework for studying the dynamics of folded and nonfolded proteins by NMR relaxation spectroscopy and MD simulation. *J. Am. Chem. Soc.* 124:4522–4534.
6. Bernado, P., K. Modig, …, M. Akke. 2010. Structure and dynamics of ribosomal protein L12: an ensemble model based on SAXS and NMR relaxation. *Biophys. J.* 98:2374–2382.