Journal of Structural Biology 173 (2011) 472–482

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

NMR and small-angle scattering-based structural analysis of protein complexes in solution

Tobias Madl [a,b,1], Frank Gabel [c,1], Michael Sattler [a,b,*]

[a] Institute of Structural Biology, Helmholtz Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany
[b] Munich Center for Integrated Protein Science, Department Chemie, Technische Universität München, Lichtenbergstr. 4, 85747 Garching, Germany
[c] Institut de Biologie Structurale IBS Jean-Pierre Ebel, CEA-CNRS-UJF, 41 rue Jules Horowitz, F-38027 Grenoble Cedex, France

ARTICLE INFO

Article history:
Available online 11 November 2010

Keywords:
NMR
SAXS
SANS
SAS
Structural biology
Protein complexes
Multi-domain proteins

ABSTRACT

Structural analysis of multi-domain protein complexes is a key challenge in current biology and a prerequisite for understanding the molecular basis of essential cellular processes. The use of solution techniques is important for characterizing the quaternary arrangements and dynamics of domains and subunits of these complexes. In this respect solution NMR is the only technique that allows atomic- or residue-resolution structure determination and investigation of dynamic properties of multi-domain -proteins and their complexes. As experimental NMR data for large protein complexes are sparse, it is advantageous to combine these data with additional information from other solution techniques. Here, the utility and computational approaches of combining solution state NMR with small-angle X-ray and Neutron scattering (SAXS/SANS) experiments for structural analysis of large protein complexes is reviewed. Recent progress in experimental and computational approaches of combining NMR and SAS are discussed and illustrated with recent examples from the literature. The complementary aspects of combining NMR and SAS data for studying multi-domain proteins, i.e. where weakly interacting domains are connected by flexible linkers, are illustrated with the structural analysis of the tandem RNA recognition motif (RRM) domains (RRM1-RRM2) of the human splicing factor U2AF65 bound to a nine-uridine (U9) RNA oligonucleotide.

1. Introduction

Networks of macromolecular interactions between proteins, nucleic acids and other ligands form the molecular basis of biological function. In particular, protein complexes are essential components in the regulation of almost every cellular process. In recent years, advances of the experimental methods available for structure determination of biological macromolecules and their complexes, such as X-ray crystallography, NMR-spectroscopy and electron microscopy techniques have pushed the limits and allow the investigation of bio-molecular structures of higher molecular weight and complexity. Nevertheless, structures of large (multi-domain) proteins and their complexes represented in the Protein Data Bank (PDB; www.rcsb.org) are sparse even though these complexes play critical roles in the regulation of gene expression, cell growth, cell cycle, metabolic pathways, signal transduction, protein folding and transport (Alberts, 1998; Nooren and Thornton, 2003; Gavin et al., 2006; Robinson et al., 2007).

Most eukaryotic proteins are composed of multiple structural modules of conserved domains that are connected by long and flexible linkers. These multi-domain proteins can be regulated by the formation of transient interactions, with dynamic rearrangements of intra- and intermolecular interfaces. Nowadays atomic resolution structures are often available for the individual domains, free or bound to their peptide or nucleic acid ligands. However, the arrangement of these multiple domains in the full-length proteins and in macromolecular complexes that eventually mediate their molecular functions is often unknown. Studying such dynamic multi-domain arrangements by X-ray crystallography is challenging as crystal packing forces often outcompete the weak domain interactions, which are required for the functional activity. In order to study the structure and dynamics of such multi-domain protein complexes it is important to complement crystallographic studies with solution techniques to capture the more native-like conformation in solution. In this respect solution NMR is the only technique that allows atomic- or residue-resolution structure

Abbreviations: NMR, nuclear magnetic resonance; SAS, small-angle scattering; SAXS, small-angle X-ray scattering; WAXS, wide angle X-ray scattering; SANS, small angle neutron scattering; RDC, residual dipolar coupling; PRE, paramagnetic relaxation enhancement; CSP, chemical shift perturbation; PCS, pseudo-contact shift; TALOS, torsion angle likelihood obtained from shifts and sequence similarity.

* Corresponding author. Fax: +49 89 289 13869.
E-mail address: sattler@helmholtz-muenchen.de (M. Sattler).

[1] These authors contributed equally.

1047-8477/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.jsb.2010.11.004

determination and investigation of dynamic properties of multi-domain proteins and their complexes. As structural analysis of large proteins by NMR is still challenging and experimental data are often sparse it is advantageous to complement NMR data with additional information from other solution techniques, such as small-angle scattering (SAS) or fluorescence experiments (see for example: Ciruela, 2008; Huang et al., 2009; Lamichhane et al., 2010). Here, we review and discuss the utility of combining solution state NMR with small-angle X-ray and neutron scattering (SAXS/SANS) experiments for structural analysis of large protein complexes, with an emphasis on multi-domain proteins where weakly interacting domains are connected by flexible linkers. The different aspects are illustrated with recent examples from the literature. Our own structural analysis of a protein–RNA complex serves as a model system for the development and application of these techniques to large multi-domain protein complexes.

2. NMR-based complex structure determination

2.1. Domain structures

To build the quaternary structure of multi-domain proteins and complexes, information about the structure of the individual components is combined with data restraining their respective distances and orientations within the complex (Fig. 1). If structures of the individual subunits are not already available from the PDB, they can be derived using standard NMR and X-ray crystallographic methods as well as homology modeling. By comparing NMR data (chemical shifts, residual dipolar couplings (RDCs)) of the available domain structures and those in the assembled complex it can be evaluated whether large conformational changes take place within the individual domains. If the overall fold is unchanged local refinement calculations can be done using the NMR data recorded for the complex (Chou et al., 2000; Sibille et al., 2001; Blackledge, 2005; Simon et al., 2010).

2.2. Mapping of subunit interfaces

In the next step, the interaction interfaces between the individual domains or subunits of the protein complex have to be determined. A number of powerful NMR techniques are available for this purpose: (1) chemical shift perturbations (CSP) (Roberts, 1993; Zuiderweg, 2002; Fielding, 2007) monitor the change in chemical environment upon ligand binding or assembly of a subunit into the complex. The presence of a (typically non NMR isotope-labeled) binding partner changes the chemical environment on the interacting subunit (e.g. ^{15}N-labeled domain/protein) and can be easily monitored using isotope-edited NMR fingerprint spectra (e.g. ^{1}H,^{15}N or ^{1}H,^{13}C^{methyl} correlation experiments) as a change in the chemical shift. As even slight changes in the chemical environment are picked up by the chemical shift, this methodology is powerful to detect even weak/transient interactions. Starting from the chemical shift assignments of the unbound subunit, weak binding (with fast binding kinetics relative to the NMR chemical shift time scale) allows transfer of assignments to the complex by tracking the shifting peaks during the titration. In the case of tight binding and high-molecular weight protein complexes, i.e. in multimeric enzymes, chemical shift assignments may require site-directed mutagenesis (Sprangers and Kay, 2007). (2) Paramagnetic relaxation enhancements (PREs, see below) by soluble, chemically inert paramagnetic probes that are used as co-solvents (Pintacuda and Otting, 2002; Madl et al., 2006, 2009; Bernini et al., 2009). These so-called solvent PREs allow the identification of a binding interface by comparing the surface accessibility of the isolated domains or subunits with those in the complex. NMR signals of surface residues located in or close to the binding interface of a complex are shielded and therefore experience reduced PRE, which provides qualitative information about the binding interface. (3) Measurement of intermolecular or inter-domain nuclear Overhauser effects (NOEs). Whereas chemical shift perturbations and PRE data yield a residue-resolved picture for the interacting surface patches, NOEs can give more precise unambiguous inter-atomic distance restraints for the interface. This methodology can be readily applied for small complexes using optimized pulse sequences to distinguish between inter- and intra-molecular NOEs (Otting and Wuthrich, 1990; Zwahlen et al., 1997; Sattler et al., 1999; Breeze, 2000). However, for larger complexes these isotope-filtered experiments are relatively insensitive and the unambiguous identification and assignment of such NOEs is difficult, rendering the use of NOEs often impractical.

2.3. Domain orientation from NMR residual dipolar couplings (RDCs)

To define the quaternary structure of the complex, information on the binding interfaces has to be complemented with data restraining the distance and relative orientation of the individual subunits. Orientation of bond vectors relative to an external

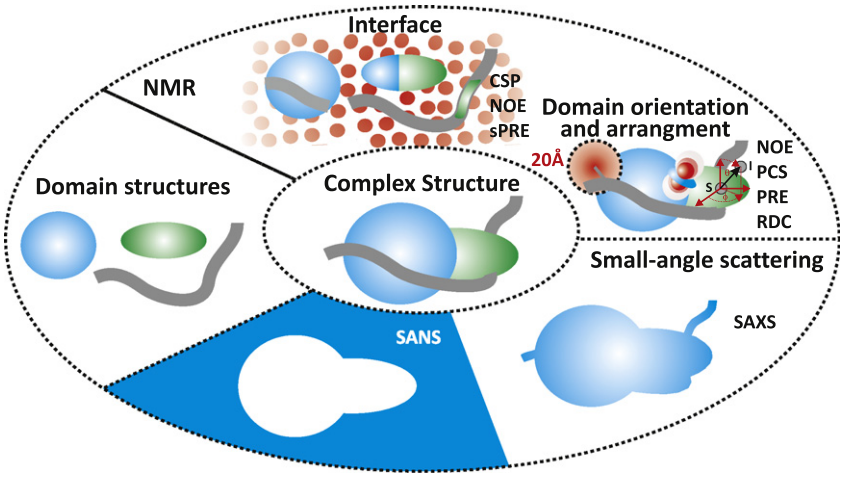

Fig. 1. Outline of the multidisciplinary approach for structure determination of protein complexes in solution by combining NMR and small-angle scattering. CSP, chemical shift perturbation; NOE, nuclear overhauser enhancement; PRE, paramagnetic relaxation enhancement; sPRE, solvent PRE; PCS, pseudo contact shift; RDC, residual dipolar coupling.