Journal of Magnetic Resonance 241 (2014) 103–114

Contents lists available at ScienceDirect

Journal of Magnetic Resonance

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

Information-driven modeling of large macromolecular assemblies using NMR data

Hugo van Ingen *, Alexandre M.J.J. Bonvin *

NMR Spectroscopy Research Group, Bijvoet Center for Biomolecular Research, Utrecht University, Faculty of Science – Chemistry, Padulaan 8, 3854 CH Utrecht, The Netherlands

ARTICLE INFO

Keywords:
Biomolecular complexes
Modeling
Docking
Integrative structural biology
TROSY
Methyl TROSY

ABSTRACT

Availability of high-resolution atomic structures is one of the prerequisites for a mechanistic understanding of biomolecular function. This atomic information can, however, be difficult to acquire for interesting systems such as high molecular weight and multi-subunit complexes. For these, low-resolution and/or sparse data from a variety of sources including NMR are often available to define the interaction between the subunits. To make best use of all the available information and shed light on these challenging systems, integrative computational tools are required that can judiciously combine and accurately translate the sparse experimental data into structural information. In this Perspective we discuss NMR techniques and data sources available for the modeling of large and multi-subunit complexes. Recent developments are illustrated by particularly challenging application examples taken from the literature. Within this context, we also position our data-driven docking approach, HADDOCK, which can integrate a variety of information sources to drive the modeling of biomolecular complexes. It is the synergy between experimentation and computational modeling that will provides us with detailed views on the machinery of life and lead to a mechanistic understanding of biomolecular function.

1. Introduction

Proteins and their intricate network of interactions are one of the cornerstones of life, performing and regulating nearly all critically important processes in the cell. Not surprisingly, understanding protein function has been a longstanding goal of biochemists and structural biologists alike. In particular, relating function to protein structure and dynamics is key in order to develop a mechanistic understanding of biological function. This hinges on the ability to determine three-dimensional (3D) high-resolution atomic structures of proteins and their complexes, either by X-ray crystallography or solution- and solid-state nuclear magnetic resonance (NMR) spectroscopy. These classical techniques are, however, faced with many challenges, especially when the macromolecular systems under study become very large and consist of multiple subunits that may interact only transiently or are membrane-embedded or associated. In practice, complexes with molecular weight above 50–100 kDa are too large for conventional, *de novo* NMR structure determination relying on an extensive network of short-range inter-proton distance. However, in many cases it is still possible to determine 3D-structures of isolated subunits either by NMR or crystallography, and to acquire structural information on their organization in the complex, although less complete and precise. In addition, complementary information might be available from other types of biochemical and biophysical experiments. The resulting collections of sparse data, of different experimental origins and information content, call for integrative computational tools to judiciously combine and translate them into meaningful atomic structures or models. These can be interrogated to test existing hypothesis or generate new ones, which can then be probed experimentally. In this Perspective, we briefly review NMR-based approaches for the integrative modeling of large and multi-subunit complexes. We warn the reader that the goal here is not to be comprehensive, nor to provide a thorough review of the current literature. We describe the NMR techniques available to characterize soluble high molecular weight complexes, the types of data that can be extracted from these, and the sources of complementary data. We then outline the general procedure for integrative modeling and illustrate all this with a number of challenging cases from the literature. Finally, we dissect current bottlenecks and present an outlook to the future of integrative modeling of large multi-subunit complexes and the role of NMR in it.

2. NMR techniques for soluble high molecular weight complexes

Both the sensitivity and resolution of solution NMR spectra deteriorate significantly with increasing molecular weight due to

* Corresponding authors.
E-mail addresses: h.vaningen@uu.nl (H. van Ingen), a.m.j.j.bonvin@uu.nl (A.M.J.J. Bonvin).

1090-7807
http://dx.doi.org/10.1016/j.jmr.2013.10.021

the line broadening of peaks. This broadening is due to long rotational tumbling correlation times τ_c, which enhance transverse relaxation. The key break-through to circumvent these deleterious relaxation effects has been the development of transverse relaxation-optimized spectroscopy (TROSY [1]), in which slowly relaxing multiplet components are selectively observed in highly deuterated proteins. In the context of the characterization of large multi-subunit protein complexes, TROSY comes in basically two flavors (Table 1). The first type is aimed at the sensitive detection of backbone amide signals (TROSY, CRIPT/CRINEPT-TROSY [2]), while the second aims specifically at the detection of methyl groups (MeTROSY [3]).

Backbone-amide detection allows monitoring of all non-proline residues, making it an excellent tool for identifying binding surfaces. However, for single-chain proteins beyond 50–100 kDa the sheer amount of backbone signals complicates the spectra, and assignment becomes increasingly difficult. In such systems, methyl-based experiments offer a very attractive alternative. MeTROSY experiments have high sensitivity due to the threefold degeneracy of the methyl-group and because its fast rotation effectively decouples the methyl group from overall molecular tumbling, thus reducing the line broadening effect [3]. The obvious drawback of the MeTROSY approach is that it is not applicable to 14 out of 20 amino acids. While typically only methyl groups in Ile, Leu, Val are observed [4], specific isotope labeling strategies have also been developed for Met, Ala (reviewed in [5]) and Thr [6]. The limited sequence coverage of MeTROSY can be alleviated to some extent by site-specific introduction of $^{13}CH_3$ groups at desired positions, for example by site-directed mutagenesis, if the structure allows for it. Such MeTROSY-based methionine scanning of solvent exposed residues has recently been proposed to map binding interfaces [7]. Alternatively, a single methyl probe may be introduced by di-sulfide bond formation with a $^{13}CH_3$–S group from methylmethanethiosulfonate resulting in the methione-mimic S-methylthiocysteine [8].

Both backbone amide-based TROSY and MeTROSY experiments have proven to allow studies of protein structure, dynamics and interaction in systems as large as 1 MDa (Table 1). In addition, other approaches such ^{13}C direct detection [9,10] or stereo-selective amino acid labeling [11,12] can help to study large molecular systems. Yet, despite these advances, low molecular tumbling rates inherently limit the applicability of solution-state NMR. In contrast, the resonance line width in magic-angle spinning (MAS) solid-state NMR (ssNMR) is independent of the protein molecular weight. Recently, Reif and co-workers as well as Bertini et al. have shown that also soluble protein complexes can be investigated by ssNMR in an approach referred to as FROSTY [13] or sedNMR (sedimented NMR) [14]. Strong centrifugal forces during MAS lead to reversible protein sedimentation at the inner wall of the MAS rotor for protein complexes above 100 kDa, effectively creating a solid. Complexes can also be sedimented into the rotor by conventional ultracentrifugation using a dedicated filling-device [15,16]. Sedimented ssNMR is thus a promising method to overcome the size barrier in solution NMR.

3. NMR data available for high molecular weight complexes

Various types of NMR experiments can provide low-resolution structural information even for large systems. Assuming that the stoichiometry and composition of the macromolecular complex under study are known, these can provide useful insights into binding sites, distances between specific pairs or groups of atoms, and relative orientation of subunits. The most frequently used data and their information content are summarized in Table 2.

3.1. Chemical shift perturbations

The workhorse of NMR for interaction studies is chemical shift perturbations (CSP) mapping, a simple comparison of peak positions in spectra before and after adding a (unlabeled) binding partner. Ligand binding induces changes in the chemical environment of the observed protein, which can conveniently be monitored by NMR (Fig. 1). Typically, these changes are followed for the ^{15}N–^{1}H backbone amide resonance that are particularly sensitive reporters of (subtle) structural changes, as was recognized already early on [17,18]. Note that chemical shifts are also sensitive to conformational changes and, as such, the observed changes do not exclusively report on the binding site, but might indicate for example allosteric changes. Similar methods can also be applied in ssNMR [19,20], making CSP-based interaction mapping an universally applicable tool. In context of protein–protein interactions in solution, 2D TROSY spectra are excellent for this purpose up to 50–100 kDa as they offer both high sensitivity and resolution. For larger systems, CRINEPT-TROSY can enable backbone-based CSP study of complex formation, as was demonstrated on the 900 kDa GroEL–GroES complex [21]. Signal overlap and the presence of unsuppressed multiplet components may complicate spectral analysis in such large systems. MeTROSY-based studies offer an excellent alternative as they follow shift changes of a reduced set of resonances. The standard deviations (σ) of chemical shifts deposited in the Biological Magnetic Resonance Databank BMRB [22] ($\sigma \sim 0.30/1.6$ ppm $^{1}H_{Me}/^{13}C$; $\sim 0.64/3.8$ ppm $^{1}H_{N}/^{15}N$) suggest that in MeTROSY spectra smaller chemical shift changes will be observed compared to backbone TROSY spectra.

In case of large protein–protein complexes, it is also important to minimize transverse relaxation in the bound state to prevent gradual bleaching of the spectrum upon titration of the ligand. In

Table 1
Overview of the main NMR techniques to study high molecular weight complexes.

NMR technique	Observables	Size limit	Key complexes studied
TROSY	N, H backbone	50–100 kDa	Too many to list!
CRIPT/CRINEPT	N, H backbone	~1 MDa	900 kDa GroEL–GroES [21] 360 kDa F1 ATPase complex [84] 200 kDa Hsp90-p53 [85] 150 kDa HasA-HasR/micelle [86]
MeTROSY	CH_3 ILVMAT	~1 MDa	1 MDa proteasome-11S [87] 220 kDa nucleosome-HMGN2 [79] 210 kDa SecA-signal peptide [88] 180 kDa p53-DNA [89] 120 kDa CheA–CheW [23]
MAS-ssNMR	All	None	1 MDa proteasome-11S [20] 600 kDa aB-crystallin [76]

Table 2
Overview of NMR data types available for HMW complexes.

Data type	Resolution	Information content	Remarks
CSP	Residues	Interface	Can give false positives; binding affinity and kinetics
NOE	Atom pairs	Distance <5–10 Å	Can be difficult across interface due to exchange broadening; in MAS-ssNMR similar through-space correlations can be obtained
PRE	Residues/atom pairs	Distance 8–35 Å	Linker flexibility and exchange dynamics can complicate quantitative interpretation
PCS	Atoms pairs	Distance 8–40 Å angles	
RDC	Bond-vector to common frame	Relative subunit orientation	Relies on anisotropic medium (complex stability) or natural alignment
Saturation transfer	Residues	Interface	Needs differential labeling, does not suffer from indirect effects

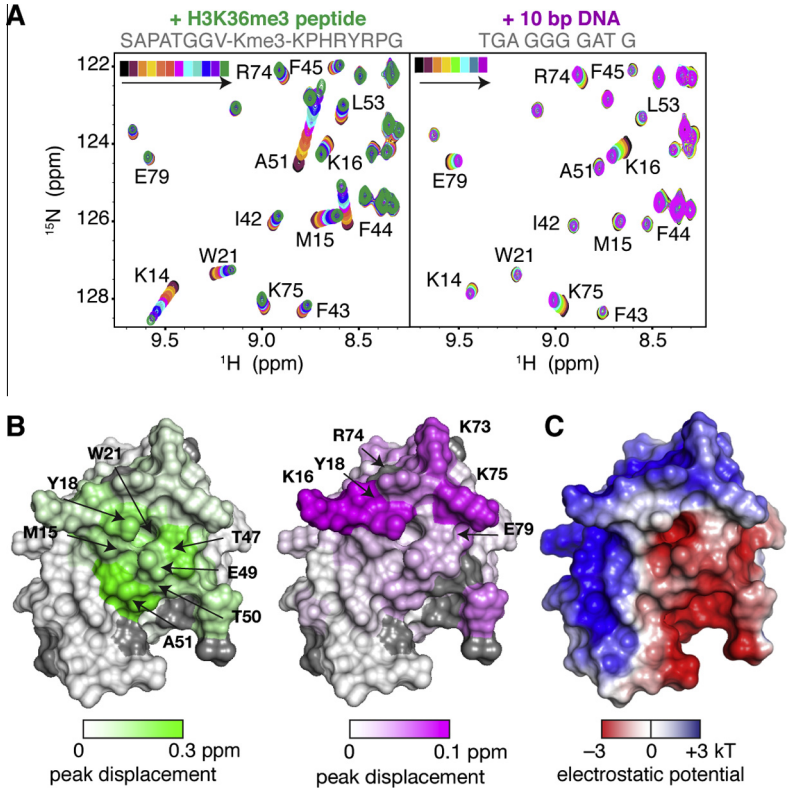

Fig. 1. Example of CSP-based interaction surface mapping, taken from Van Nuland et al. [80]. (A) Sections of the 2D ^{1}H–^{15}N HSQC spectrum of the PSIP1-PWWP domain during the titration with H3K36me3 peptide (left panel) and dsDNA fragment (right panel). Free PWWP spectrum in black; resonances of interest are labeled. (B) Interaction surfaces for the histone peptide (left) and dsDNA (right), coded on the van der Waals surface. Grey is used for residues w/o data; residues with shifts larger than 10% trimmed mean + 2σ are labeled. (C) Electrostatic potential on the solvent accessible surface color-coded on the Van der Waals surface. Originally published by BioMedCentral [80].

such case, it is advantageous to use a perdeuterated binding partner to avoid spurious relaxation of the TROSY coherence due to spin-flips caused by the external spins of the ligand [23].

As CSPs are usually monitored via 2D spectra, the CSP for both ^{1}H ($\Delta\delta_H$) and the heteronucleus ($\Delta\delta_X$) are obtained simultaneously and usually combined into a single score. It can be expressed as the geometric peak displacement in Hz or as a weighted average CSP expressed in ppm:

$$\mathrm{CSP} = \sqrt{\frac{1}{2}\left(\frac{\Delta\delta_H}{\alpha}\right)^2 + \left(\frac{\Delta\delta_X}{\beta}\right)^2}$$

The weighting factors α and β are usually taken to be 1 and 0.2 in case of backbone amides to account for the difference in spectral widths available for ^{15}N (~25 ppm) and ^{1}H (~5 ppm). An objective alternative is to weigh $\Delta\delta$ with the standard deviation of that particular resonance as taken from the BMRB database, thereby calculating a CSP "*Z*-score". For backbone amides, this will correspond to a setting of 1 and 0.17. Having a final list of CSP values, a threshold needs to be determined to identify the interface residues. As the observed CSPs typically form a continuous profile, no objective *a priori* threshold can be set. A common method is to set the threshold at 1 or 2 standard deviations σ above the mean CSP calculated on a 10% trimmed set in which the 10% largest values are excluded. In all cases, the significance must be judged on the basis of whether the identified residues form a contiguous patch on the surface of the molecule or not (Fig. 1B/C). We recently developed an algorithm (SAMPLEX) to identify the binding surface with minimal bias, taking structural neighbors into account [24]. Nevertheless, whatever procedure is taken, there will be falsely identified interface residues for which the observed CSP is in fact an indirect effect of binding.

In addition to indirect effects, chemical shift changes may be also be caused by slight changes in pH, salt concentrations upon addition of the binding partner. To minimize these effects great care must taken to have both molecules in exactly the same buffer

conditions, preferably by extensive simultaneous dialysis. This is especially important when the expected shifts are small, as for example when too little material is available to saturate the binding site. Under these conditions, very small changes in chemical shift (much less than the line width) can reliably be measured, as illustrated in a recent study on binding of a substrate to GroEL [25]. Finally, it should be noted that quantitative analysis of CSP can also be used to determine binding affinity and kinetics and dissect ligand binding modes. For further discussion of chemical shift perturbation mapping, see the excellent recent review by Williamson [26].

3.2. NOEs

Intermolecular NOEs have very high information content, provided they can be assigned unambiguously. Given a sufficient number of short-range distances between specific pairs of atoms, typically <5–6 Å (minimum of three independent ones distributed across the interface), two molecules can be unambiguously docked [27]. In the case of large complexes, NOEs can be measured efficiently and up to ~10 Å, provided the proteins are highly deuterated to suppress unwanted spin diffusion and transverse relaxation [28,29]. Measurement of intermolecular NOEs may still be complicated, however, due for example to exchange kinetics resulting in broadened lines at the interface or residual mobility in the complex. In addition, verification of the intermolecular nature of NOEs requires isotope-filtered experiments that have inherent lower sensitivity and their interpretation necessitates assignment of both interacting partners.

3.3. Paramagnetic mapping (PRE/PCS)

A robust alternative to measure intermolecular distances relies on paramagnetic relaxation enhancement (PRE) of protein ^{1}H resonances caused by the interaction of the magnetic dipole with unpaired electrons in a near-by paramagnetic center [30]. Because of the strong magnetic moment associated with electrons, PREs can be used to identify long-range distances up to 20–35 Å, depending on the paramagnetic species used [31]. The unpaired electron can be site-specifically introduced in a metal binding site or attached to the protein via a tag. For an overview of the available methods, the reader is referred to excellent recent reviews [32,33]. Commonly used tags are the nitroxide spinlabel MTSL [34] and Mn^{2+}–EDTA derivatives [35], which are introduced via cysteine mutants. The PRE effect can be quantified in straight-forward fashion by taking the ratio of peak intensities in paramagnetic and diamagnetic conditions [36] or from two-point T_2-measurements for both para- and diamagnetic states [37]. The latter approach has as advantages that precise calibration of protein concentrations in the two samples is not required, no long interscan delays are needed to ensure equilibrium, and no Lorentzian line shapes are required. Precise treatment of the intermolecular PRE effect as distance restraints necessitates knowledge of the exchange kinetics between free and bound states that averages the PRE effect. In addition, the tag might need to be explicitly modeled in the docking process and its flexibility accounted for, either by increasing the error bounds, or, more properly, by ensemble averaging [38,39]. Alternatively, intermolecular PREs can be used in a more qualitative manner to map the binding interface [40].

An alternative method without the need for covalent attachment of the paramagnetic center is to use solvent PREs. Here, chemically inert paramagnetic probes are added as co-solvents and cause relaxation and thus signal attenuation of solvent accessible protons [41]. Applied to protein complexes, solvent PREs can be used to quantitatively describe the distance of the observed nucleus to the molecular surface of the complex [42].

Paramagnetic lanthanide ions attached to a protein can also give rise to chemical shift changes, the so-called pseudocontact shifts (PCS). These depend on both the distance and relative orientation to the unpaired electron and may give long-range information up to 40 Å from the paramagnetic center [43]. Using a rigid, two-point anchored lanthanide tag, the possibility of obtaining both distance and angular information between subunits has been shown to allow for efficient docking [44–46].

3.4. Other NMR information sources

Information on the relative orientation of subunits can also be obtained from residual dipolar couplings (RDCs) caused by incomplete averaging of dipolar interactions in anisotropic conditions [47]. Finally, cross-saturation methods can effectively be used to map binding interfaces by saturating protons in one subunit and observing the transfer of saturation to non-overlapping protons in the deuterated observed subunit [48].

Here, we have focused on methods that provide information on the intermolecular interface within large complexes. It should be noted that complementary information on the bound-state conformation of the subunits may also be acquired using either backbone chemical shift prediction of dihedral angles [49], transferred NOEs [50] or cross-correlation experiments [51,52]. Overall, NMR provides the experimentalist with many options to obtain site-specific data, either at the atom or residue specific level, on the binding interfaces and structure of a complex.

4. Sources of complementary data

Other biophysical or biochemical sources of structural information that the experimentalist may turn to are listed in Table 3. They will be briefly reviewed here.

4.1. Interface information

Interface information can be obtained using site-directed mutagenesis, in combination with a binding assay, to identify specific

Table 3
Sources of complementary data.

Information content	Resolution	Data types	Remarks
Interface	Residues	Mutagenesis	Readout by any binding essay; can give false positives
	Residues	H/D exchange	Read-out by MS or NMR
Distance	Atom pairs	Crosslinking/MS	Long-range; linker length/flexibility
	Label pairs	FRET	Long-range, <80–100 Å
	Label pairs	EPR	Long-range, <60–80 Å
Shape	Complex	Cryo-EM	Overall structure, >10 Å usually
		SAXS/SANS	Molecular envelope, Rg
		IM-MS	Collision cross-section

residues that are critical for binding. Other options are the use of hydrogen/deuterium exchange to compare the solvent accessibility of surface residues in free and bound states, monitored by either mass-spectrometry (MS) [53] or NMR [54]. Bioinformatics approaches based on sequence conservation and/or surface properties can also help to identify interaction surfaces [55].

4.2. Long-range distance information

Other sources of long-range distance information include, among others, chemical cross-linking experiments, in which typically Lys side-chains are cross-linked and identified via mass-spectrometry (MS) [56], FRET, in which the measured distances depend on the separation of the fluorescently labeled residues of the complex [57], and EPR in which the distance between two paramagnetic center can be measured up to 60–80 Å [58,59].

4.3. Molecular shape information

In recent years, small angle X-ray and neutron scattering (SAXS/SANS) have become important complementary techniques to study complexes in solution that can provide radius of gyration (Rg), an indicator of the structure compactness, and low-resolution 3D molecular envelopes from the scattering intensity at very low angles [60]. SANS can be used on subunit-selectively deuterated samples to provide valuable additional information on the overall shape and positioning of subunits within a complex. It relies on matching the scattering intensity of a protonated subunit to the background scattering from the solvent in a particular H_2O/D_2O mixture, thus masking that particular subunit. Considering, that for large complexes subunit-selectively deuterated samples have to be used for NMR studies in any case, the acquisition of SANS data comes in principle at no additional costs [61].

Cryo-electron microscopy (cryo-EM) experiments provide an electron density map with a resolution range typically between 8 Å and 20 Å [62], into which individual subunit structures can be fitted [63]. Finally, ion mobility mass spectrometry (IM-MS) experiments also provide shape-related information in the form of collision cross-sections (CCS). The CCS corresponds to the rotationally-averaged molecular area to which the buffer gas can collide; it can thus offer information on the overall size and conformation of the complex [64].

5. General workflow of integrative modeling

Integrative modeling of complexes essentially revolves around placing atomic structures of the subunits together and refining them, guided by diverse sets of experimental data (Fig. 2). The required structures of the constituents may be available from the Protein Data Bank (PDB) or should be determined experimentally, or generated by homology modeling.

Several approaches for integrative modeling have been developed, one good example of which is the Integrative Modeling Platform IMP [65]. Here, we focus on our in-house developed HADDOCK (high-ambiguity data driven docking) program [66,67]. HADDOCK allows the inclusion of (sparse) data coming from various experimental sources and can deal simultaneously with molecules of different nature, i.e. proteins, small molecules, oligosaccharides, and nucleic acids. It allows incorporation of both ambiguous and unambiguous spatial information to drive the

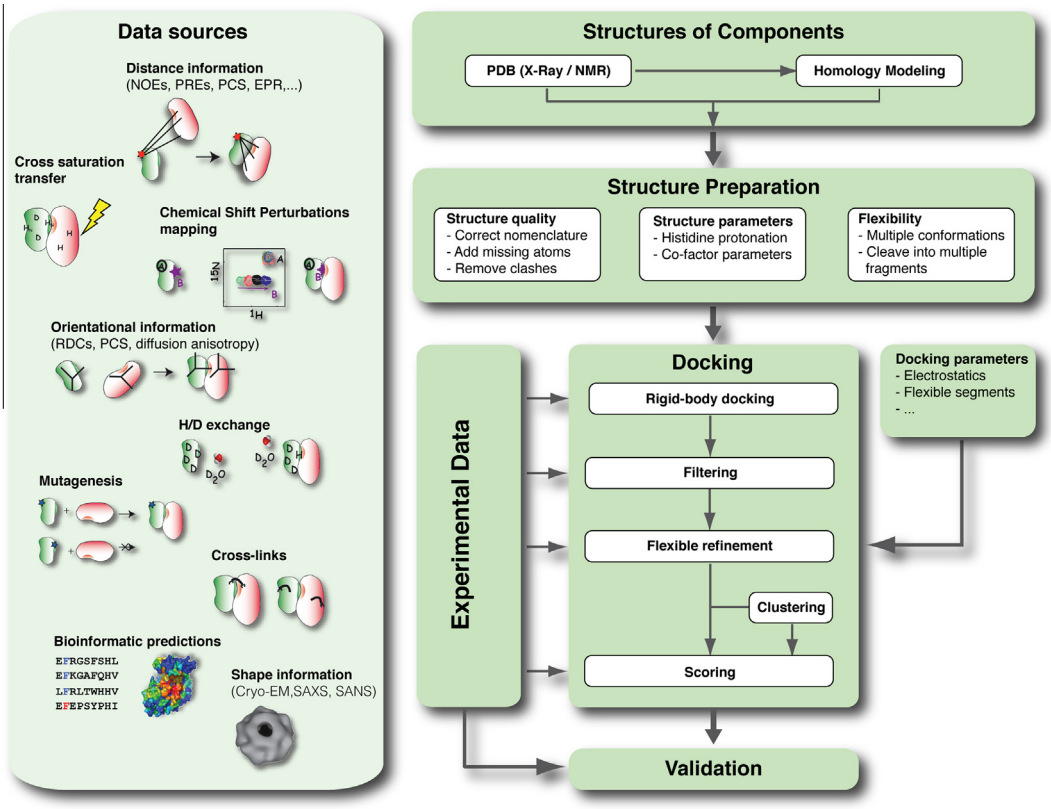

Fig. 2. Data sources (left) and overview of the docking workflow (right) as implemented in HADDOCK. Ambiguous interaction data from a variety of sources can be used to drive the docking (or score the resulting models, e.g. SAXS) of up to six separate subunits. Atomic structures can be taken from the PDB databank or obtained by homology modeling. They are pre-processed to correct any errors, define histidine protonation states and handle flexibility (some of these steps are automated, some do require user definitions). After an initial rigid-body docking stage, the best solutions are subjected to a flexible refinement (in two stages: in torsion angle space in vacuum followed by a refinement in Cartesian space in explicit solvent), after which the solutions are clustered and scored. Ideally, independent experimental data should be used to validate the resulting models.

simultaneous docking of up to 6 subunits. HADDOCK is essentially a collection of shell, Python and CNS scripts that control a customized, staged structure calculation within CNS [68], evaluating at each stage which structures are best in terms of interaction energies (van der Waals, electrostatics and desolvation energies), properties (buried surface area), and correspondence with the imposed restraints. The conformational space available to the complex is searched by minimizing a target function E_{target} that includes the experimental and/or bioinformatics data:

$$E_{target} = E_{FF} + E_{restr}$$

Minimization of E_{target} ensures that the computed model simultaneously agrees with *a priori* encoded empirical knowledge on covalent and non-bonded interactions (E_{FF}, i.e. bonds, angles, dihedrals, chirality, electrostatics and van der Waals), as well as the observed data, described by E_{restr}. While minimization/optimization methods are often not exhaustive, the experimental information restrains the conformational search space, thus resulting in an often more homogenous set of solutions. HADDOCK uses a flat-bottom, "soft-square" potential [69] to impose restraints. This potential behaves harmonically up to violations of 2 Å, after which it switches smoothly to a linear one. Such a modification avoids enormous forces due to large violations that can result in instabilities of the calculations. The flat-bottom potential, enables the incorporation of restraints with upper and lower limits to account for the uncertainty of the measurements.

Information about interfaces (but not the specific contacts made) is converted into Ambiguous Interaction Restraints (AIRs). AIRs are composed of active (residues that are known to make contact) and passive (residues that potentially make contact – usually the surface neighbors of active's) residues. Those residues are used to define a network of ambiguous distance restraints, which ensures that an active residue on the surface of a biomolecule should be in close vicinity to any active or passive residues on the partner biomolecule. If the list of interacting residues is not very accurate then a user-defined percentage of the restraints can be discarded at random during docking and refinement (50% by default). Another key advantage of HADDOCK is its flexibility in imposing the restraints. Users can impose different combination of restraints at different stages of the docking protocol and can change the weights assigned to each of them depending on the data accuracy and confidence in the data.

The docking procedure is composed of three stages: (i) initial docking by rigid body energy minimization (it0), (ii) semi-flexible refinement in torsion angle space (it1) and (iii) final refinement in explicit solvent (water). The binding mode of the complex is roughly determined during it0 and then a pre-defined percentage of the top-ranking solutions according to the HADDOCK-score (a weighted sum of electrostatics, van der Waals, restraint energies, buried surface area and an empirical desolvation term), are selected for further refinement. The consecutive refinement steps allow for small- to medium-range conformational changes while improving the overall score of the models. The final structures are clustered based on their pairwise ligand interface RMSD (the root-mean-square-deviation of the atomic coordinates, considering only heavy backbone atoms, of interface residues belonging to the 'ligand' subunit(s) when all models are superimposed on the interface residues of the first subunit) and the average cluster scores are calculated over the top 4 members of each cluster.

HADDOCK was originally developed to make use of NMR data, and in particular of chemical shift perturbation data. Currently, it can translate most of the information sources listed in Tables 1 and 2 into structural restraints (or additional scoring terms in the case of SAXS and CCS data [70]), except for cryo-EM data, although work in this direction is ongoing. All of these features are also offered via HADDOCK's user-friendly web server interface [71] at http://haddock.science.uu.nl.

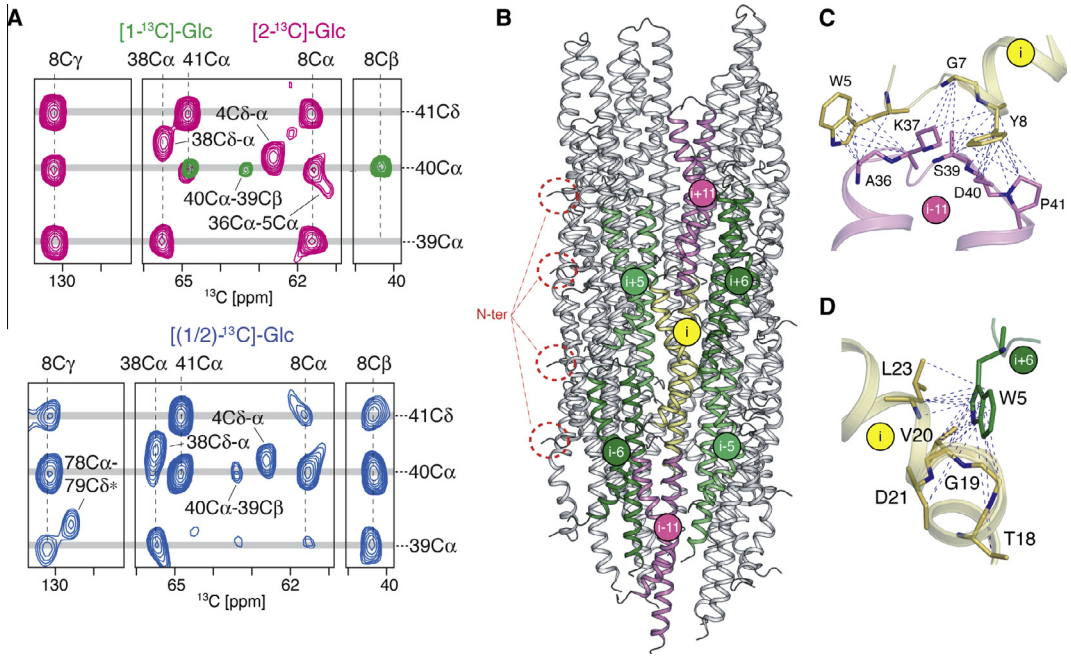

Fig. 3. ssNMR-based modeling of the Type III secretion system needle by Loquet et al. [72]. (A) Identification of inter-subunit contacts from comparison of ssNMR ^{13}C–^{13}C spin diffusion experiments on differentially labeled samples. Needles assembled from pure [1-^{13}C]-glucose (green), and [2-^{13}C]-glucose (magenta) labeled monomers (top) and assembled from a mixture of such monomers (bottom). Intersubunit restraints are encoded are Pro 41Cδ–Tyr 8Cβ and Ser 39Cα–Tyr 8Cβ. (B) Complete atomic model of the T3SS needle in ribbon representation viewed from the side. Subunit at positions *i* makes axial contacts to *i* ± 11, and lateral contacts to *i* ± 5, *i* ± 6. (C) Detailed view on the axial (top) and lateral (bottom) subunit–subunit interfaces. Blue dashed lines represent ssNMR restraints. Reprinted by permission from Macmillan Publishers Ltd: Nature, 486 (7402), Loquet et al., Atomic Model of the type III secretion system needle, copyright 2012.

6. Highlighted examples

We divide our discussion of applications of modeling large assemblies in two categories, based on the molecular topology, in symmetrical and non-symmetrical complexes.

6.1. Symmetrical complexes

Many supramolecular assemblies exist as symmetrical oligomers. The symmetry in these systems combined with knowledge of the subunit structures can be used to guide the modeling of these assemblies. An inspiring application has come from Loquet et al. who focused on the needle of the Type III secretion system, an insoluble symmetric homomeric complex consisting of 80-residue monomers, resistant to crystallization [72]. Using ssNMR experiments on recombinant, selectively isotope labeled Type III needle, they were able to define unambiguous intra- and intersubunit distance restraints (Fig. 3). Needles assembled from differentially labeled monomers were used to unambiguously identify inter-subunit contacts. EM measurements showed that the needle is formed as a helix with ~11 subunits per two turns. Starting from helically arrayed set of 29 monomers with an extended backbone conformation, they applied the fold-and-dock protocol of Rosetta [73], using the ssNMR chemical shifts, together with intra- and inter subunit distance restraints and the EM-based radius of the needle. In contrast to previous suggestions, the resulting structure revealed that N-terminal part of the subunit is located on the outside of the needle and mediates important inter-subunit interactions.

Modeling symmetric oligomers when the oligomeric state is variable is extra challenging. Baldwin et al. tackled this problem by integration of MS, NMR, and IM-MS data [74] to characterize αB-crystallin, a small heat shock protein (Fig. 4). MS data indicated that this system exists in a dynamic equilibrium of differently sized oligomers. NMR spectra revealed that each monomer exists in a symmetrical environment. A range of candidate structures was constructed formed by either series of regular polyhedra or rings. Computed collision cross-sections (CCS) of these models were compared to those obtained experimentally. Using the observed trends in CCS, consistent models of the dominant αB-crystallin 24-, 26- and 28-mer oligomers were identified as polyhedral architectures. These arrangements provide a structural rationale for the interconversion of these oligomers via loss and addition of a subunit. In a similar integrated approach atomic structures of 24-mer αB-crystallin complexes have been derived [75,76].

6.2. Non-symmetrical complexes

Lack of symmetry in a complex also means a significant loss of information to drive the modeling. Thus studies on non-symmetrical complexes are typically limited to dock two subunits together,

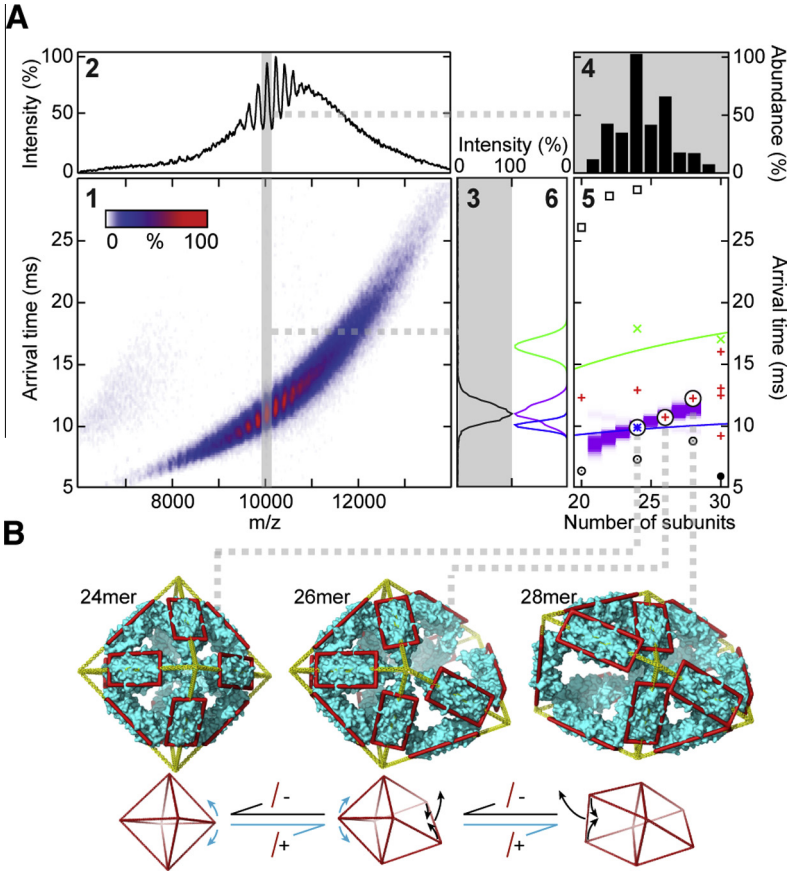

Fig. 4. Ion-mobility mass-spectrometry (IM-MS) based modeling of αB-crystallin by Baldwin et al. [74]. (A) Step-wise analysis and interpretation of IM-MS data of αB-crystallin. The IM-MS spectrum (1), and the corresponding summed mass spectrum (2), are characteristic of a broad range of oligomeric masses and sizes. The distribution of arrival times (ATD) of the peak at 10,100m/z can be extracted (3) and the identity and relative abundances of the underlying oligomers determined from tandem-MS (4). These experimental observations are then compared to the arrival times calculated for candidate oligomers based on their theoretical collision-cross sections (CCS) (5), and by combination with their relative abundances (4) to yield a theoretical ATD (6). Comparison between the experimental (3) and calculated data (6) allows the determination of the best fit of candidate structures to the IM-MS data. (B) Three structural models for the 24-mer, 26-mer, and 28-mer that fit the experimentally determined CCS measurements (top). They can interconvert simply by the insertion or removal of a single "dimeric" edge (bottom). Reprinted from Structure, 19 (12), Baldwin et al., The polydispersity of aB-crystallin is rationalized by an interconverting polyhedral architecture, pp. 1855–1863, copyright 2011, with permission from Elsevier.

of which one may be a known, multi-subunit complex itself. Recent work of the Kay lab focused on the interaction between the 70 kDa DnaK and the 580 kDa hexameric ClpB in protein disaggregation [77]. Using an impressive, and pragmatic, combination of backbone and methyl-group based TROSY and complexes with hexameric and monomeric variants of ClpB, the authors could define the binding surfaces on both proteins from CSP measurements and identify a 1:6 stoichiometry of the DnaK:ClpB complex. PRE measurements were performed on complexes of ILVM-labeled DnaK nucleotide binding domain bound to monomeric ClpB, labeled with MTSL at five different positions. The resulting 29 distance restraints were combined with CSP-derived ambiguous interaction restraints to dock the DnaK-NBD to a ClpB monomer (Fig. 5). The models were validated by mutagenesis and used to devise functional test of ClpB–DnaK function in protein disaggregation, revealing that the DnaK–ClpB interaction stimulates ClpB activity on the substrate.

A nice example of how different types of NMR data can be used comes from the docking of a nuclear export signal (NES) peptide to the 150 kDa exportin CRM1/RanGTP complex [42,78]. Using an intricate combination of ^{13}C-direct detection, CRINEPT-TROSY, several ambiguous and unambiguous intermolecular NOEs and solvent PREs, the peptide was docked precisely and in a well-defined conformation to its binding site. The resulting structures were consistent with the crystal structure of the complex based on a NES-fusion protein and explained structural basis of NES recognition.

As a large DNA–protein complex, nucleosomes present an additional challenge in modeling of their complexes with other chromatin factors. In a collaborative effort between the Kay and Bai labs a crucial break-through was realized in the near-complete assignment of ILV methyl groups in the nucleosome (200 kDa) [79]. The architecture of its complex with chromatin factor HMGN2 was derived on the basis of CSP, PRE and mutagenesis data,

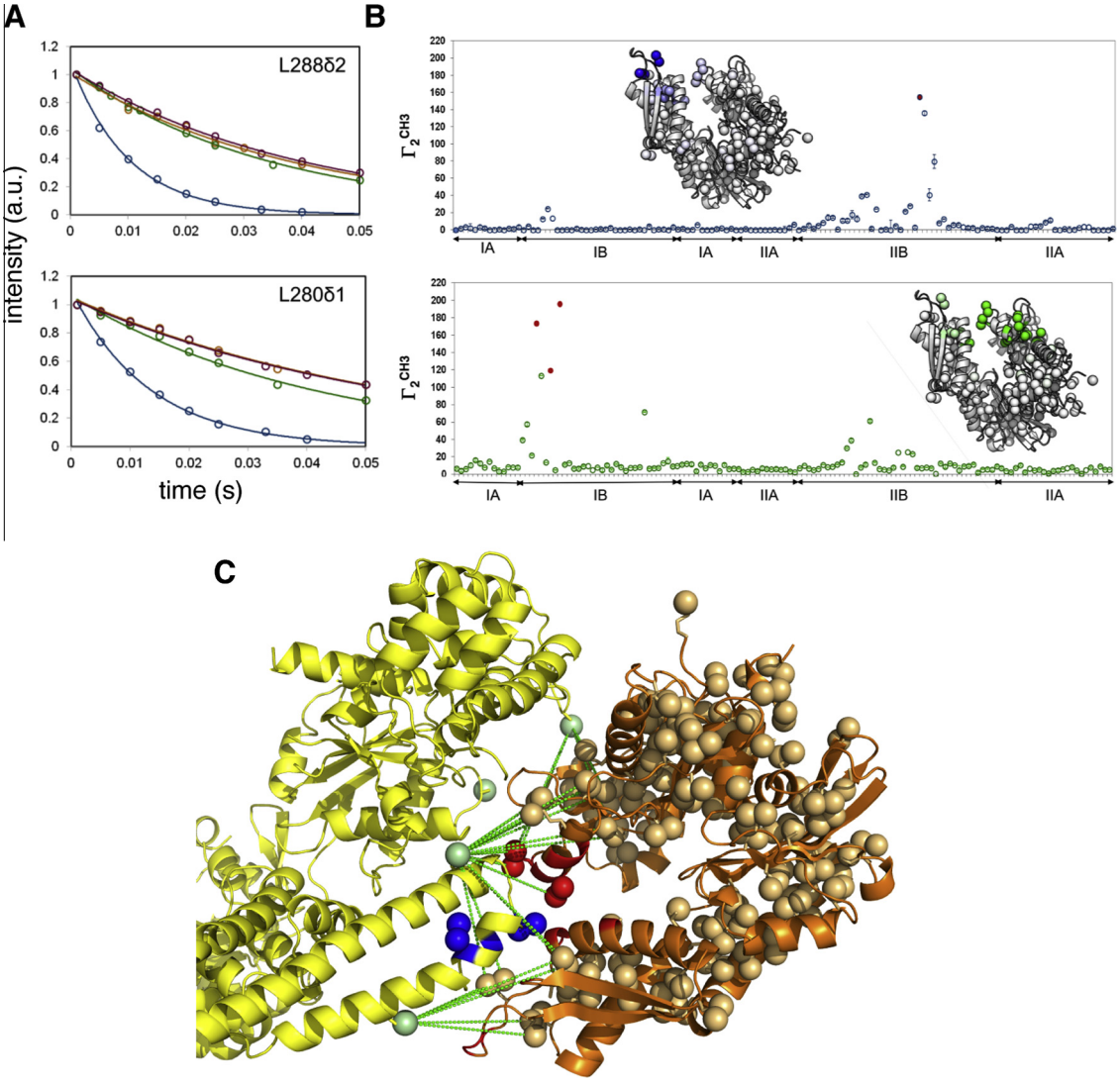

Fig. 5. Modeling of the ClpB–DnaK complex using paramagnetic relaxation enhancement (PRE) based distance restraints by Rosenzweig et al. [77]. (A) Experimental ^{1}H R_2 relaxation curves measured for selected ILVM residues on DnaK for different MTSL positions as indicated by the different colors. (B) Extracted PRE obtained for ILVM-DnaK nucleotide binding domain (NBD) for two MTSL spin label positions. Red filled circles indicate residues where cross-peaks in HMQC spectra are broadened beyond the limits of detection. Sub-domain boundaries for DnaK-NBD are indicated underneath each plot. Cartoon representations of DnaK NBD with methyl residues shown as balls and color coded according to their PRE. (C) Model of the DnaK–ClpB complex. DnaK-NBD in orange, ClpB in yellow. ILVM methyl groups show as balls, active residues on DnaK in red, on ClpB in blue. PRE derived distance restraints are shown as green dashed lines. From Science, 339 (6123), Rosenzweig et al. Unraveling the mechanism of protein disaggregation through a ClpB–DnaK interaction, copyright 2013. Reprinted with permission from AAAS.

demonstrating the feasibility of modeling nucleosome–protein complexes using solution NMR. Recently, for the first time a structural model for the read-out of an epigenetically modified nucleosome was determined in our lab [80], characterizing the interaction between the PSIP1-PWWP domain and a nucleosome trimethylated at H3K36 (H3K36me) (Fig. 6). Comparing the interactions of the PWWP domain with isolated H3K36me-peptides, DNA and H3K36me-nucleosomes, revealed that the nucleosomal DNA plays an important role in the specific recognition of this modification, boosting the affinity by more than 10,000-fold. The complex was modeled using HADDOCK and AIRs based on an extensive mutagenesis analysis and observed CSPs. Acknowledging the flexibility of the H3 N-terminal tail, the flexible multi-domain docking protocol was adapted [81]. First, the H3K36me3 peptide was docked to the aromatic cage of the PWWP domain on the basis of CSP and homology derived AIRs. Second, the resulting complex was docked back to nucleosome, guided by the identified DNA interaction surface and covalent restraints for the H3-tail. In this step, a threading approach was taken to systematically sample the binding site of PWWP on the nucleosomal DNA. The DNA surrounding the H3 N-terminal tail exit point was divided in 10 patches of each 5 bp. For each docked structure one of these patches were defined as passive residues. The resulting structures were cross-validated against mutagenesis data, leaving a single cluster of solutions. The solutions show how the arrangement of aromatic cage and basic patches on the surface PWWP domain matches perfectly to its nucleosomal substrate. Particularly, the solutions reveal a network of extensive electrostatic interactions between PWWP Lys and Arg residues and the DNA phosphate backbone. Subsequent modeling of other H3K36me3-readers showed that the relative configuration of aromatic cage and basic patches is conserved, suggesting conserved role of the nucleosomal DNA in H3K36me recognition.

Modeling of non-symmetrical complexes with three or more subunits is especially challenging, because of the increase in degrees-of-freedom and the requirement of obtaining experimental restraints for all mutual interactions. NMR data can be used to determine binding interface on all subunits, thus positioning the subunits. Restraints on the overall shape of the whole or part of the complex can be extremely useful to improve the quality of the models. Recent work of the Sattler group on a ternary protein–protein–RNA complex [61], systematically explored how SAXS/SANS-derived molecular envelops could help to refine structural models obtained from CSP-driven HADDOCK-models (Fig. 7). First, the RNA binding surfaces on the two proteins were mapped using TROSY experiments on perdeuterated proteins. Then, the protein–protein interface was deduced from comparison of the chemical shift of the ternary complex and the two protein–RNA

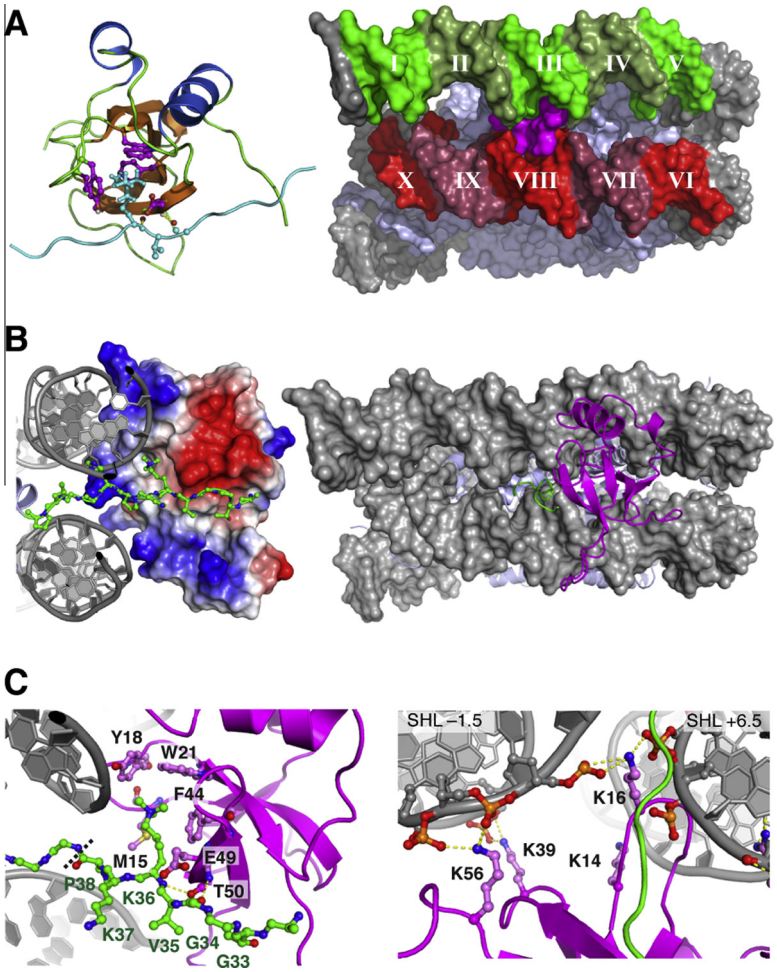

Fig. 6. Modeling of the PSIP1-PWWP H3K36me3-nucleosome complex using a flexible multi-domain approach by Van Nuland et al. [80]. (A) The H3K36me3 peptide bound to PSIP1-PWWP domain obtained in the first docking stage (left) is docked to one of the ten DNA patches on the nucleosome, indicated by roman numerals I–X, guided by CSPs and loose restraints to between the H3 peptide and the H3 tail exit site, shown in magenta. (B) Overall (right) and detailed view emphasizing the matching electrostatics (left) of the lowest energy structure of the cross-validated cluster of solutions. (C) Close-up on the intermolecular interactions that define H3K36me3 specificity (left) and nucleosome affinity (right). Originally published by BioMedCentral [80].

complexes. The resulting HADDOCK models clustered in 7 groups. Scoring them using the SAXS/SANS data led to a unique solution. In particular, the SANS data on subunit-selectively perdeuterated complexes at 70% D_2O, in which the RNA was masked from the scattering curve, provided strong restraints for the respective arrangement the protein components.

7. Critical outlook on the state-of-the-art and future prospects

Improvements in NMR methodology has broadened its scope into the range of large molecular assemblies where traditional structure determination approaches fail. Data-driven computational modeling has become a powerful complementary tool to obtain some atomistic insight into the structure–function relationships of such complexes. Nevertheless, the risk associated with modeling is that the resulting models are biased by the input structures, by the particular nature of the experimental restraints, and/or by the choices made during the modeling. It is the task of the modeling community to minimize the potential for bias by providing robust and well-balanced methods for integrative modeling. At the same time, users should be aware of the potential pitfalls and adjust their strategy of data collection and modeling accordingly.

Bias from the input structures can play a role when those are derived from homology models. Users should in particular assess the reliability of the binding interface structure from the sequence identity to the template structure. Another modeling challenge is dealing with the large structural changes in the subunits that can occur upon binding. Current protocols can typically deal with small to medium conformational changes, but new methodologies will be needed to deal with large-scale changes and folding-upon-binding events. For symmetric complexes, a number of attractive options already exist, provided sufficient data is available to drive the folding of monomers [73,82]. In other cases, a promising way forward is to use coarse-grained representations, in which groups of atoms (or even residues) are represented by a single particle, thereby reducing the degrees-of-freedom allowing greater sampling of conformational space. Such approach should be especially useful in modeling of very large systems, but comes at the price of a lower information content due to the reduced resolution.

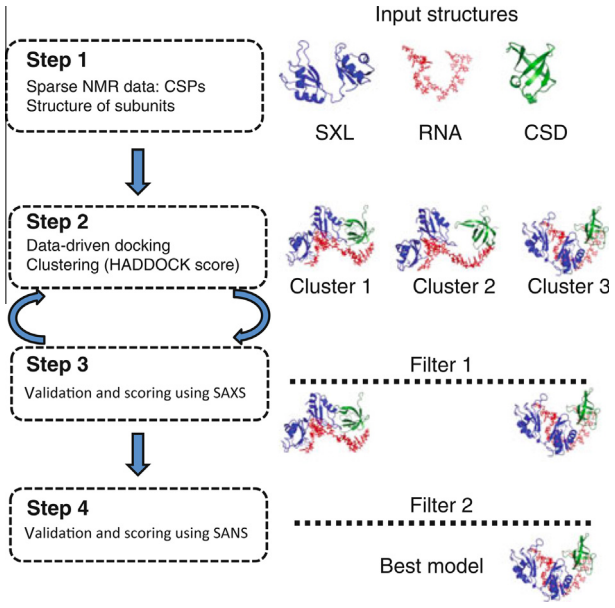

Fig. 7. Overview of NMR/SAXS/SANS-based modeling data of a ternary protein–protein–RNA complex by Hennig et al. [61]. Docking of the two RRM motifs of Sex-lethal (SXL), the cold shock domain 1 of UNR (CSD) and an 18-mer RNA based solely on NMR chemical shift pertubations resulted in several clusters of equally probable solutions. These clusters were scored against their correspondence with the experimental SAXS data resulting in the rejection of some clusters. Further scoring against SANS data recorded with different matching conditions resulted in a final single cluster of solutions that is simultaneously compatible with all experimental data. From Springer, Journal of Biomolecular NMR, 56 (1), Hennig et al., Combining NMR and small-angle X-ray and neutron scattering in the structural analysis of a ternary protein–RNA complex, copyright 2013. With kind permission from Springer Science and Business Media.

The ambiguity, lack, incompatibility or false-positive nature of experimental restraints may also be sources of bias. Considering integrative modeling, defining a robust protocol for integration of different data sets, dealing with false positives (wrong data, or data that represent indirect effects of the binding), deciding on the relative weights attributed to the various data in the restraining or scoring terms, as well as identifying the best combinations of data sources, are important tasks for the modeling community. Ongoing efforts using probabilistic approaches can potentially solve a number of these issues [83]. For the end-user it is clear that as much as data as possible should be acquired, including distance, orientational, and/or shape restraints, wherever possible. In addition, it also strongly advised to keep part of the data for cross-validation purposes or perform directed mutagenesis to confirm the validity of the obtained models.

Structures obtained from modeling are useful for the research community and as such open-access to these models should be warranted. Whether such models should be deposited in the protein data bank PDB is debatable, given their intrinsic ambiguity. However, the level of ambiguity is data-dependent. In particular, given enough unambiguous distance restraints, the modeled structure of the complex will be effectively the same as a traditional NMR structure. The difficulty is to assess the relation between the amount, type and precision of the data as well as the quality of the input structure on the one hand, and the resolution and ambiguity of the resulting models on the other hand. Thus, a grey area arises between 'models' and 'structures'. It should be noted that there are several smaller protein–protein complexes deposited in the PDB that are solely based on CSPs AIR restraints. For larger systems this is clearly not advisable, still these models should be made available. Currently, there are a handful of NMR-based structures of large complexes (>100 kDa) in the PDB in which a large part of the structure is either modeled or taken from an existing crystal-structure. In all cases, unambiguous distance restraints either from PRE or NOE were used to drive the modeling, sometimes in combination with CSPs. The PDB faces the difficult task to formulate a deposition policy on such structures that are based on sparse data. We advocate that researchers provide their models, associated statistics, and the restraint lists as supplementary material. In addition, one could envisage a 'PDB' for data-driven, integrative models of complexes where such data would be made freely available in a central repository.

8. Conclusions

In recent years NMR has established itself as a prime source of quantitative, site-specific structural information for large and multi-subunit assemblies. Combined with complementary data from other sources, these sparse data can be used to create atomic structures of such assemblies using integrative modeling approaches. We have reviewed and highlighted the NMR techniques and data sources available, the integrated modeling workflow from the perspective of the HADDOCK software, together with a number of recent standout applications. The synergy between experimentation and computational modeling will provide us in the future increasingly detailed views on the machinery of life, leading to a mechanistic understanding of biomolecular function.

References

[1] K. Pervushin, R. Riek, G. Wider, K. Wüthrich, Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution, Proc. Natl. Acad. Sci. U. S. A. 94 (1997) 12366–12371.

[2] R. Riek, G. Wider, K. Pervushin, K. Wüthrich, Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules, Proc. Natl. Acad. Sci. U. S. A. 96 (1999) 4918–4923.

[3] V. Tugarinov, P.M. Hwang, J.E. Ollerenshaw, L.E. Kay, Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes, J. Am. Chem. Soc. 125 (2003) 10420–10428.

[4] V. Tugarinov, L.E. Kay, An isotope labeling strategy for methyl TROSY spectroscopy, J. Biomol. NMR 28 (2004) 165–172.

[5] A.M. Ruschak, L.E. Kay, Methyl groups as probes of supra-molecular structure, dynamics and function, J. Biomol. NMR 46 (2010) 75–87.

[6] A. Velyvis, A.M. Ruschak, L.E. Kay, An economical method for production of 2H,13CH3-threonine for solution NMR studies of large protein complexes: application to the 670 kDa proteasome, PLoS ONE 7 (2012) e43725.

[7] M.C. Stoffregen, M.M. Schwer, F.A. Renschler, S. Wiesner, Methionine scanning as an NMR tool for detecting and analyzing biomolecular interaction surfaces, Structure 20 (2012) 573–581.

[8] T.L. Religa, A.M. Ruschak, R. Rosenzweig, L.E. Kay, Site-directed methyl group labeling as an NMR probe of structure and dynamics in supramolecular protein systems: applications to the proteasome and to the ClpP protease, J. Am. Chem. Soc. 133 (2011) 9063–9068.

[9] N. Nishida, F. Motojima, M. Idota, H. Fujikawa, M. Yoshida, I. Shimada, et al., Probing dynamics and conformational change of the GroEL–GroES complex by ^{13}C NMR spectroscopy, J. Biochem. 140 (2006) 591–598.

[10] M. Matzapetakis, P. Turano, E.C. Theil, I. Bertini, ^{13}C–^{13}C NOESY spectra of a 480 kDa protein: solution NMR of ferritin, J. Biomol. NMR 38 (2007) 237–242.

[11] M. Kainosho, T. Torizawa, Y. Iwashita, T. Terauchi, A. Mei Ono, P. Güntert, Optimal isotope labelling for NMR protein structure determinations, Nature 440 (2006) 52–57.

[12] T. Ikeya, T. Terauchi, P. Güntert, M. Kainosho, Evaluation of stereo-array isotope labeling (SAIL) patterns for automated structural analysis of proteins with CYANA, Magn. Reson. Chem. 44 (2006) S152–S157.

[13] A. Mainz, S. Jehle, B.J. van Rossum, H. Oschkinat, B. Reif, Large protein complexes with extreme rotational correlation times investigated in solution by magic-angle-spinning NMR spectroscopy, J. Am. Chem. Soc. 131 (2009) 15968–15969.

[14] I. Bertini, C. Luchinat, G. Parigi, E. Ravera, B. Reif, P. Turano, Solid-state NMR of proteins sedimented by ultracentrifugation, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 10396–10399.

[15] C. Gardiennet, A.K. Schütz, A. Hunkeler, B. Kunert, L. Terradot, A. Böckmann, et al., A sedimented sample of a 59 kDa dodecameric helicase yields high-resolution solid-state NMR spectra, Angew. Chem., Int. Ed. 51 (2012) 7855–7858.

[16] I. Bertini, F. Engelke, L. Gonnelli, B. Knott, C. Luchinat, D. Osen, et al., On the use of ultracentrifugal devices for sedimented solute NMR, J. Biomol. NMR 22 (2012) 11.

[17] N.A. van Nuland, G.J. Kroon, K. Dijkstra, G.K. Wolters, R.M. Scheek, G.T. Robillard, The NMR determination of the IIA(mtl) binding site on HPr of the *Escherichia coli* phosphoenol pyruvate-dependent phosphotransferase system, FEBS Lett. 315 (1993) 11–15.

[18] Y. Chen, J. Reizer, M.H. Saier, W.J. Fairbrother, P.E. Wright, Mapping of the binding interfaces of the proteins of the bacterial phosphotransferase system, HPr and IIAglc, Biochemistry 32 (1993) 32–37.

[19] S.G. Zech, E. Olejniczak, P. Hajduk, J. Mack, A.E. McDermott, Characterization of protein–ligand interactions by high-resolution solid-state NMR spectroscopy, J. Am. Chem. Soc. 126 (2004) 13948–13953.

[20] A. Mainz, T.L. Religa, R. Sprangers, R. Linser, L.E. Kay, B. Reif, NMR spectroscopy of soluble protein complexes at one mega-Dalton and beyond, Angew. Chem., Int. Ed. 52 (2013) 8746–8751.

[21] J. Fiaux, E.B. Bertelsen, A.L. Horwich, K. Wüthrich, NMR analysis of a 900 K GroEL–GroES complex, Nature 418 (2002) 207–211.

[22] E.L. Ulrich, H. Akutsu, J.F. Doreleijers, Y. Harano, Y.E. Ioannidis, J. Lin, et al., BioMagResBank 36 (2008) D402–D408.

[23] D.J. Hamel, F.W. Dahlquist, The contact interface of a 120 kD CheA–CheW complex by methyl TROSY interaction spectroscopy, J. Am. Chem. Soc. 127 (2005) 9676–9677.

[24] M. Krzeminski, K. Loth, R. Boelens, A.M.J.J. Bonvin, SAMPLEX: automatic mapping of perturbed and unperturbed regions of proteins and complexes, BMC Bioinform. 11 (2010) 51.

[25] D.S. Libich, N.L. Fawzi, J. Ying, G.M. Clore, Probing the transient dark state of substrate binding to GroEL by relaxation-based solution NMR, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 11361–11366.

[26] M.P. Williamson, Using chemical shift perturbation to characterise ligand binding, Prog. Nucl. Magn. Reson. Spectrosc. 73 (2013) 1–16.

[27] M. Ikura, G.M. Clore, A.M. Gronenborn, G. Zhu, C.B. Klee, A. Bax, Solution structure of a calmodulin-target peptide complex by multidimensional NMR, Science 256 (1992) 632–638.

[28] R. Horst, G. Wider, J. Fiaux, E.B. Bertelsen, A.L. Horwich, K. Wüthrich, Proton-proton Overhauser NMR spectroscopy with polypeptide chains in large structures, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 15445–15450.

[29] R. Sounier, L. Blanchard, Z. Wu, J. Boisbouvier, High-accuracy distance measurement between remote methyls in specifically protonated proteins, J. Am. Chem. Soc. 129 (2007) 472–473.

[30] N. Bloembergen, L.O. Morgan, Proton relaxation times in paramagnetic solutions, effects of electron spin relaxation, J. Chem. Phys. 34 (1961) 842.

[31] G.M. Clore, C. Tang, J. Iwahara, Elucidating transient macromolecular interactions using paramagnetic relaxation enhancement, Curr. Opin. Struct. Biol. 17 (2007) 603–616.

[32] J. Koehler, J. Meiler, Expanding the utility of NMR restraints with paramagnetic compounds: background and practical aspects, Prog. Nucl. Magn. Reson. Spectrosc. 59 (2011) 360–389.

[33] P.H.J. Keizers, M. Ubbink, Paramagnetic tagging for protein structure and dynamics analysis, Prog. Nucl. Magn. Reson. Spectrosc. 58 (2011) 88–96.

[34] V. Gaponenko, J.W. Howarth, L. Columbus, G. Gasmi-Seabrook, J. Yuan, W.L. Hubbell, et al., Protein global fold determination using site-directed spin and isotope labeling, Protein Sci. 9 (2000) 302–309.

[35] A. Dvoretsky, V. Gaponenko, P.R. Rosevear, Derivation of structural restraints using a thiol-reactive chelator, FEBS Lett. 528 (2002) 189–192.

[36] J. Battiste, G. Wagner, Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data, Biochemistry 39 (2000) 5355–5365.

[37] J. Iwahara, C. Tang, G.M. Clore, Practical aspects of 1H transverse paramagnetic relaxation enhancement measurements on macromolecules, J. Magn. Res. 184 (2007) 185.

[38] A.N. Volkov, J.A.R. Worrall, E. Holtzmann, M. Ubbink, Solution structure and dynamics of the complex between cytochrome c and cytochrome c peroxidase determined by paramagnetic NMR, Proc. Natl. Acad. Sci. U. S. A. 103 (2006) 18945–18950.

[39] J. Iwahara, C.D. Schwieters, G.M. Clore, Ensemble approach for NMR structure refinement against (1)H paramagnetic relaxation enhancement data arising from a flexible paramagnetic group attached to a macromolecule, J. Am. Chem. Soc. 126 (2004) 5879–5896.

[40] A. Impagliazzo, M. Ubbink, Mapping of the binding site on pseudoazurin in the transient 152 kDa complex with nitrite reductase, J. Am. Chem. Soc. 126 (2004) 5658–5659.

[41] G. Pintacuda, G. Otting, Identification of protein surfaces by NMR measurements with a paramagnetic Gd(III) chelate, J. Am. Chem. Soc. 124 (2002) 372–373.

[42] T. Madl, T. Güttler, D. Görlich, M. Sattler, Structural analysis of large protein complexes using solvent paramagnetic relaxation enhancements, Angew. Chem., Int. Ed. 50 (2011) 3993–3997.

[43] M. Allegrozzi, I. Bertini, M.B.L. Janik, Y.-M. Lee, G. Liu, C. Luchinat, Lanthanide-induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 Å from the metal ion, J. Am. Chem. Soc. 122 (2000) 4154–4161.

[44] P.H.J. Keizers, B. Mersinli, W. Reinle, J. Donauer, Y. Hiruma, F. Hannemann, et al., A solution model of the complex formed by adrenodoxin and adrenodoxin reductase determined by paramagnetic NMR spectroscopy, Biochemistry 49 (2010) 6846–6855.

[45] T. Saio, M. Yokochi, H. Kumeta, F. Inagaki, PCS-based structure determination of protein–protein complexes, J. Biomol. NMR 46 (2010) 271–280.

[46] C. Schmitz, A.M.J.J. Bonvin, Protein–protein HADDocking using exclusively pseudocontact shifts, J. Biomol. NMR 50 (2011) 263–266.

[47] A. Bax, G. Kontaxis, N. Tjandra, Dipolar couplings in macromolecular structure determination, Methods Enzymol. 339 (2001) 127–174.

[48] I. Shimada, NMR techniques for identifying the interface of a larger protein–protein complex: cross-saturation and transferred cross-saturation experiments, Methods Enzymol. (2005) 483–506.

[49] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, J. Biomol. NMR 44 (2009) 213–223.

[50] F. Ni, Recent developments in transferred NOE methods, Prog. Nucl. Magn. Reson. Spectrosc. 26 (1994) 517–606.

[51] T. Carlomagno, I. Felli, M. Czech, R. Fischer, M. Sprinzl, C. Griesinger, Transferred cross-correlated relaxation: Application to the determination of sugar pucker in an aminoacylated tRNA-mimetic weakly bound to EF-Tu, J. Am. Chem. Soc. 121 (1999) 1945–1948.

[52] T. Carlomagno, M.J.J. Blommers, J. Meiler, W. Jahnke, T. Schupp, F. Petersen, et al., The high-resolution solution structure of epothilone A bound to tubulin: an understanding of the structure-activity relationships for a powerful class of antitumor agents, Angew. Chem., Int. Ed. Engl. 42 (2003) 2511–2515.

[53] A.J. Percy, M. Rey, K.M. Burns, D.C. Schriemer, Probing protein interactions with hydrogen/deuterium exchange and mass spectrometry – a review, Anal. Chim. Acta 721 (2012) 7–21.

[54] Y. Paterson, S.W. Englander, H. Roder, An antibody binding site on cytochrome c defined by hydrogen exchange and two-dimensional NMR, Science 249 (1990) 755–759.

[55] H.X. Zhou, S. Qin, Interaction-site prediction for protein complexes: a critical assessment, Bioinformatics 23 (2007) 2203–2209.

[56] M. Trester-Zedlitz, K. Kamada, S.K. Burley, D. Fenyö, B.T. Chait, T.W. Muir, A modular cross-linking approach for exploring protein interactions, J. Am. Chem. Soc. 125 (2003) 2416–2425.
[57] U.B. Choi, P. Strop, M. Vrljic, S. Chu, A.T. Brunger, K.R. Weninger, Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex, Natl. Struct. Mol. Biol. 17 (2010) 318–324.
[58] G.F. White, L. Ottignon, T. Georgiou, C. Kleanthous, G.R. Moore, A.J. Thomson, et al., Analysis of nitroxide spin label motion in a protein–protein complex using multiple frequency EPR spectroscopy, J. Magn. Res. 185 (2007) 191–203.
[59] H. Yagi, D. Banerjee, B. Graham, T. Huber, D. Goldfarb, G. Otting, Gadolinium tagging for high-precision measurements of 6 nm distances in protein assemblies by EPR, J. Am. Chem. Soc. 133 (2011) 10418–10421.
[60] C.D. Putnam, M. Hammel, G.L. Hura, J.A. Tainer, X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution, Q. Rev. Biophys. 40 (2007) 191–285.
[61] J. Hennig, I. Wang, M. Sonntag, F. Gabel, M. Sattler, Combining NMR and small angle X-ray and neutron scattering in the structural analysis of a ternary protein–RNA complex, J. Biomol. NMR 56 (2013) 17–30.
[62] G.C. Lander, H.R. Saibil, E. Nogales, Go hybrid: EM, crystallography, and beyond, Curr. Opin. Struct. Biol. 22 (2012) 627–635.
[63] A. Fotin, Y. Cheng, P. Sliz, N. Grigorieff, S.C. Harrison, T. Kirchhausen, et al., Molecular model for a complete clathrin lattice from electron cryomicroscopy, Nature 432 (2004) 573–579.
[64] C. Uetrecht, R.J. Rose, E. van Duijn, K. Lorenzen, A.J.R. Heck, Ion mobility mass spectrometry of proteins and protein assemblies, Chem. Soc. Rev. 39 (2010) 1633–1655.
[65] F. Alber, F. Förster, D. Korkin, M. Topf, A. Sali, Integrating diverse data for structure determination of macromolecular assemblies, Annu. Rev. Biochem. 77 (2008) 443–477.
[66] C. Dominguez, R. Boelens, A.M.J.J. Bonvin, HADDOCK: a protein−protein docking approach based on biochemical or biophysical information, J. Am. Chem. Soc. 125 (2003) 1731–1737.
[67] S.J. de Vries, A.D.J. van Dijk, M. Krzeminski, M. van Dijk, A. Thureau, V. Hsu, et al., HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets, Proteins 69 (2007) 726–733.
[68] A.T. Brünger, P.D. Adams, G.M. Clore, W.L. DeLano, P. Gros, R.W. Grosse-Kunstleve, et al., Crystallography & NMR system: a new software suite for macromolecular structure determination, Acta Crystallogr., Sect. D: Biol. Crystallogr. 54 (1998) 905–921.
[69] M. Nilges, A.M. Gronenborn, A.T. Brünger, G.M. Clore, Determination of three-dimensional structures of proteins by simulated annealing with interproton distance restraints. application to crambin, potato carboxypeptidase inhibitor and barley serine proteinase inhibitor 2, Protein Eng. 2 (1988) 27–38.
[70] E. Karaca, A.M.J.J. Bonvin, On the usefulness of ion-mobility mass spectrometry and SAXS data in scoring docking decoys, Acta Crystallogr., Sect. D: Biol. Crystallogr. 69 (2013) 683–694.
[71] S.J. de Vries, M. van Dijk, A.M.J.J. Bonvin, The HADDOCK web server for data-driven biomolecular docking, Nat. Protocols 5 (2010) 883–897.
[72] A. Loquet, N.G. Sgourakis, R. Gupta, K. Giller, D. Riedel, C. Goosmann, et al., Atomic model of the type III secretion system needle, Nature 486 (2012) 276–279.
[73] R. Das, I. André, Y. Shen, Y. Wu, A. Lemak, S. Bansal, et al., Simultaneous prediction of protein folding and docking at high resolution, Proc. Natl. Acad. Sci. U. S. A. 106 (2009) 18978–18983.
[74] A.J. Baldwin, H. Lioe, G.R. Hilton, L.A. Baker, J.L. Rubinstein, L.E. Kay, et al., The polydispersity of αB-crystallin is rationalized by an interconverting polyhedral architecture, Structure 19 (2011) 1855–1863.
[75] N. Braun, M. Zacharias, J. Peschek, A. Kastenmüller, J. Zou, M. Hanzlik, et al., Multiple molecular architectures of the eye lens chaperone αB-crystallin elucidated by a triple hybrid approach, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 20491–20496.
[76] S. Jehle, B.S. Vollmar, B. Bardiaux, K.K. Dove, P. Rajagopal, T. Gonen, et al., N-terminal domain of alphaB-crystallin provides a conformational switch for multimerization and structural heterogeneity, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 6409–6414.
[77] R. Rosenzweig, S. Moradi, A. Zarrine-Afsar, J.R. Glover, L.E. Kay, Unraveling the mechanism of protein disaggregation through a ClpB–DnaK interaction, Science 339 (2013) 1080–1083.
[78] T. Güttler, T. Madl, P. Neumann, D. Deichsel, L. Corsini, T. Monecke, et al., NES consensus redefined by structures of PKI-type and Rev-type nuclear export signals bound to CRM1, Nat. Struct. Mol. Biol. 17 (2010) 1367–1376.
[79] H. Kato, H. van Ingen, B.-R. Zhou, H. Feng, M. Bustin, L.E. Kay, et al., Architecture of the high mobility group nucleosomal protein 2-nucleosome complex as revealed by methyl-based NMR, Proc. Natl. Acad. Sci. U. S. A. 108 (2011) 12283–12288.
[80] R. van Nuland, F.M. van Schaik, M. Simonis, S. van Heesch, E. Cuppen, R. Boelens, et al., Nucleosomal DNA binding drives the recognition of H3K36-methylated nucleosomes by the PSIP1-PWWP domain, Epigenet. Chromatin 6 (2013) 12.
[81] E. Karaca, A.M.J.J. Bonvin, A multidomain flexible docking approach to deal with large conformational changes in the modeling of biomolecular complexes, Structure 19 (2011) 555–565.
[82] B. Bardiaux, B.-J. van Rossum, M. Nilges, H. Oschkinat, Efficient modeling of symmetric protein aggregates from NMR data, Angew. Chem., Int. Ed. 51 (2012) 6916–6919.
[83] W. Rieping, M. Nilges, M. Habeck, ISD: a software package for Bayesian NMR structure calculation, Bioinformatics 24 (2008) 1104–1105.
[84] M. Kobayashi, H. Akutsu, T. Suzuki, M. Yoshida, H. Yagi, Analysis of the open and closed conformations of the beta subunits in thermophilic F1-ATPase by solution NMR, J. Mol. Biol. 398 (2010) 189–199.
[85] S. Rudiger, CRINEPT-TROSY NMR reveals p53 core domain bound in an unfolded form to the chaperone Hsp90, Proc. Natl. Acad. Sci. 99 (2002) 11085–11090.
[86] C. Caillet-Saguy, M. Piccioli, P. Turano, N. Izadi-Pruneyre, M. Delepierre, I. Bertini, et al., Mapping the interaction between the hemophore HasA and its outer membrane receptor HasR using CRINEPT-TROSY NMR spectroscopy, J. Am. Chem. Soc. 131 (2009) 1736–1744.
[87] R. Sprangers, L.E. Kay, Quantitative dynamics and binding studies of the 20S proteasome by NMR, Nature 445 (2007) 618–622.
[88] I. Gelis, A.M.J.J. Bonvin, D. Keramisanou, M. Koukaki, G. Gouridis, S. Karamanou, et al., Structural basis for signal-sequence recognition by the translocase motor SecA as determined by NMR, Cell 131 (2007) 756–769.
[89] M. Bista, S.M. Freund, A.R. Fersht, Domain–domain interactions in full-length p53 and a specific DNA complex probed by methyl NMR spectroscopy, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 15752–15756.