

---

# Application of Nuclear Magnetic Resonance and Hybrid Methods to Structure Determination of Complex Systems

22

Filippo Prischi and Annalisa Pastore

---

## Abstract

The current main challenge of Structural Biology is to undertake the structure determination of increasingly complex systems in the attempt to better understand their biological function. As systems become more challenging, however, there is an increasing demand for the parallel use of more than one independent technique to allow pushing the frontiers of structure determination and, at the same time, obtaining independent structural validation. The combination of different Structural Biology methods has been named hybrid approaches. The aim of this review is to critically discuss the most recent examples and new developments that have allowed structure determination or experimentally-based modelling of various molecular complexes selecting them among those that combine the use of nuclear magnetic resonance and small angle scattering techniques. We provide a selective but focused account of some of the most exciting recent approaches and discuss their possible further developments.

---

## Keywords

Multidomain proteins • NMR • Protein-protein interactions • Structural biology • SAXS

---

F. Prischi  
School of Biological Sciences, University of Essex,  
Wivenhoe Park, Colchester CO4 3SQ, UK

A. Pastore (✉)  
Department of Clinical Neurosciences, King's  
College London, Denmark Hill Campus, London, UK  
e-mail: [Annalisa.Pastore@crick.ac.uk](mailto:Annalisa.Pastore@crick.ac.uk)

---

## 22.1 Introduction

Not so long ago researchers would be mainly experts in one field and carry out their research using a very specific expertise. Nowadays, the number of techniques at hand has enormously increased enhancing the possibility of independently validating results with more than one tool. This is particularly true in Structural Biology, a field that has gained enormous momentum in the

post-genomic era. The possibility of using more than one technique has also suggested new approaches, which may allow one to combine results and obtain a better and more complete picture thus moving further the frontiers of structure determination. In this review, we will focus on so-called ‘hybrid’ techniques developed for solution studies. We will first briefly overview what can by now be considered classical methods based on nuclear magnetic resonance (NMR) techniques to then overview hybrid methods combining NMR with small angle X-ray scattering (SAXS). This technique has proven well suited to provide information on the overall shape of a molecule and on the non-uniform distribution of the protein atomic density [1]. Because of their complementarity, NMR being a higher resolution technique but unable to deliver information on big assemblies, SAXS being low resolution but very effective in the reconstruction of the overall shapes, the two techniques are often exploited in combination, as detailed in the next pages.

There are currently three main applications of hybrid methods based on these two techniques: the definition of the relative orientation of multi-domain proteins, structure refinement of proteins having sparse distance restraints and the reconstruction of large molecular complexes [2]. We will review some examples, discuss the limitations encountered and suggest new directions.

---

## 22.2 Why NMR Cannot Do Everything

The time of structural studies of single domain proteins has rapidly had its sunset. Most of the forefront current structural biology projects deal with multi-domain proteins and large molecular assemblies, which can be as big as or larger than the ribosome [3, 4]. These changed priorities have revolutionized our perspective of the tools needed for structure determination. A well-

established way, which quickly approaches its 30th birthday, is to use the so-called cut-and-paste approach that relies on the minimalistic approach of cutting a protein/complex into isolated domains/components and solving their structures and their interactions in isolation. Only after these are studied, we will want to find ways to determine the relative orientation of the individual parts. Both the two main techniques traditionally used for structure determination, *i.e.*, X-ray crystallography and nuclear magnetic resonance (NMR) in solution, have strongly benefited from this concept although for different reasons.

The cut-and-paste approach is good in crystallography because it allows cutting away, at least in a first instance, flexible regions, which could be difficult to crystallize. It is also helpful in NMR studies: structure determination with classic NMR methods solely based on nuclear Overhauser effects (NOEs) can be very challenging with large proteins, because, due to slower rotational diffusion, the line widths increase, resulting in a decrease of the signal-to-noise ratio and an increase of resonance overlap up to the disappearance of the signals [5]. The exact limit is not simply a function of the molecular weight: proteins with similar molecular weights can be observed or not according to whether they are intrinsically unfolded or rigidly globular. The limit can anyway be extended by uniformly or selectively deuterating most of the molecule’s protons, which effectively ‘dilutes’ out the spin concentration and increases the  $T_2$  transversal relaxation with consequent decrease of the line widths [6]. Selective labeling, perdeuteration and the use of TROSY pulse sequences in high-field NMR spectrometers (900 MHz or higher fields) have dramatically increased the signal-to-noise ratio [7].

Besides the molecular weight, another limitation of NMR is encountered when wanting to determine the relative orientation of multi-domain proteins. The task is particularly prob-

lematic when systems do not have a rigid and extended interface. This is because NOEs are intrinsically short-range observables [5] (it could of course be argued that alternative techniques are not necessarily much better: X-ray crystallography can well obtain long-range information but the suspect can be that the result might be biased by the very process of crystallization). Residual dipolar couplings (RDCs) were introduced to resolve the problem. Internuclear magnetic dipole couplings contain a great deal of structural information, but they average to zero in isotropic solution as a result of rotational diffusion [8]. Tjandra and Bax developed a method in which alignment of proteins with the magnetic field can be achieved through the use of weak liquid crystalline (LC) media [9]. This induces an anisotropic distribution of orientations that allows accurately measurement of a wide array of RDCs [9] which provide a powerful source of orientational restraints by defining the angles between an inter-nuclear vector and the axis of an alignment frame. The alignment frame works as an external reference and is fixed to the molecular frame of the molecule. An advantage of this formalism is that, since RDCs are relative to the molecular frame, they are independent from the tumbling of the molecule, hence they pick up motions faster and slower than rotational tumbling correlation time of the molecule. This makes RDCs powerful tools to monitor protein dynamics.

The major disadvantage when using orientational restraints in the construction of oligomer models is that for a single set of RDCs and a structural model exists four combinations of relative orientation of the subunits in the complex [10]. This uncertainty can be resolved by collecting a second (or more) data set of RDCs from a different alignment medium [10]. Not always this is possible though as not all media are suitable for all proteins. Another way to get around these limitations is to combine RDCs with other NMR observable and builds models from these hybrid restraints [11–13]. These include not only NOEs, but also paramagnetic relaxation enhancement (PRE) and chemical shift perturbation (CSP) data. SAXS has been used as an alternative or in combination with these methods.

### 22.3 SAXS in Defining Multi-domain Relative Orientations

As illustrated by Mertens and Svergun, the analysis of flexible systems by SAXS has received a great boost by Bernado and collaborators [14]. A recent study [15] focused on the difficulties associated with the interpretation of SAXS curves for highly flexible proteins. These proteins are at times identified as rigid from dynamically averaged SAXS profiles unless several indicators are monitored. The best approach resides in a method called ensemble optimization method (EOM) [16] because it provides a reliable measure of the flexibility of the system under study. In the EOM approach it is necessary to generate first a large pool of random configurations and then select ensembles using a genetic algorithm [16]. When the Rg distribution of the models in the selected ensembles is as broad as that in the initial random pool, the protein is probably flexible, whereas a narrow Rg peak hints at a rigid system. The combination of SAXS and NMR spectroscopy provides averages of the entire ensembles of conformation [17–20]. However, it is very challenging to identify consistent ensembles, given the vast number of conformations that can potentially be adopted by flexible proteins. Several approaches have been developed.

Probably the first examples of using SAXS in combination with NMR information were published back in 1996–1997 in studies aiming at reconstructing modular proteins from the individual domain [21]. Sunnerhagen et al. [21] used this combination of techniques to study the relative orientation of Gla and EGF domains in the coagulation factor X. This is a serine protease containing three noncatalytic domains: an N-terminal gamma-carboxyglutamic acid (Gla) domain followed by two epidermal growth factor (EGF)-like domains. It was noticed that when linked to the Gla domain, the Ca<sup>2+</sup> affinity of the isolated N-terminal EGF domain is increased tenfold suggesting a cross-talk between the two domains. Through a study of the NMR solution structure of the factor X Gla-EGF domain pair (with Ca<sup>2+</sup> bound to the EGF domain),

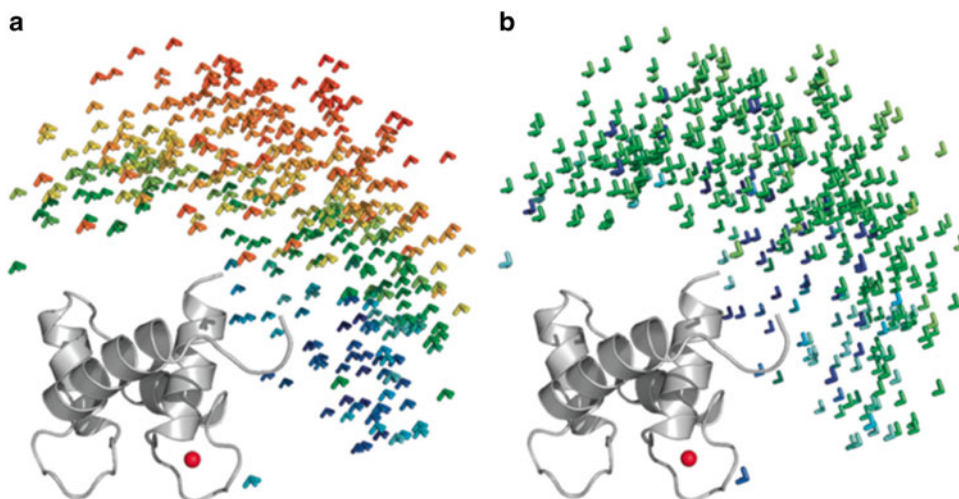
complemented by SAXS data on the Gla-EGF domain pair (with and without  $\text{Ca}^{2+}$ ), the authors showed that  $\text{Ca}^{2+}$  binding to the EGF domain makes the Gla and EGF domains fold toward each other using the  $\text{Ca}^{2+}$  site as a hinge. Presumably, a similar mechanism may be responsible for alterations in the relative orientation of protein domains in many other extracellular proteins containing EGF domains with the consensus for  $\text{Ca}^{2+}$  binding. Finally, this study demonstrated the powerful combination of NMR and SAXS in the study of modular proteins, since it combines reliable evaluation of short- (NMR) and long-range (SAXS) interactions.

Our group used NMR/SAXS shortly after to build a model of how the multidomain protein titin is assembled [22, 23]. Titin is a giant muscular protein and a prototype of a modular protein containing *ca.* 300 copies of two all- $\beta$  sequence motifs, the fibronectin type 3 and the immunoglobulin-like modules [23]. An important question was (and still to some extent is) whether titin modules interact with each other or are loosely connected without intermodule interactions. The question was addressed by assessing the extent of CSP between modules and measuring by SAXS the maximal distance of constructs of two- and four-modules. It was concluded that the linkers connecting the domains in the I-band are relatively rigid and dictate a total length of the multi-domain constructs shorter than the one expected for the sum of the individual domains.

Bertini et al. [19] developed an algorithm to determine the maximum occurrence (MO) of a given conformation, or the maximum percent of time a system spends in a given conformation. The program, publicly available using the grid computing infrastructure (<https://www.wenmr.eu/>), initially generates a pool of about  $10^5$  conformations using RANCH [20]. Theoretical NMR and SAXS data are generated for each conformation, using FANTASIAN [24] and CALCALL [19] to estimate respectively pseudo-contact shifts (PCS) and residual dipolar couplings (RDC) and using CRY SOL [25] to calculate SAXS intensities. At this point, a conformation (A) is selected and assigned a weight lower than 100%. A group of other conformations

(randomly selected from the initial pool) is added to this conformation with a weight that is adjusted to obtain the best fit between experimental and theoretical data. The program varies not only the weight of the different conformations, but also discards and substitutes other conformations from the pool, with the selection driven by a simulated annealing protocol. The procedure stops when the target function (TF) reaches a minimum value and the MO for A is determined (Fig. 22.1a, b). This methodology was applied to calmodulin (CaM), a classic model of a flexible two-domain protein [reviewed in 26]. The MO of the  $\text{Ca}^{2+}$  bound (PDB ID: 1CLN, 1CLL, 1PRW) [27–29] and of the closed peptide-bound forms (PDB ID: 1CDL, 1CDM, 1IQ5, 1NIW, 1YR5, 2BCX, 2XOG) was evaluated [30–35]. It was concluded that dumbbell-shape extended conformations as well as compact conformations have very low occupancy (MOs in the order of 5–15%). More expanded (or, implicitly, more flexible) conformations have MO as high as 35%, strongly suggesting that these conformations are most abundant in solution. These results mainly confirmed decades of previous studies that had already established the absence of extended conformations of calmodulin in solution [36–38].

A more sophisticated and interesting approach was implemented by Huang et al. [39] using a combination of PRE, RDC and SAXS data to study U2AF65. This protein is essential for spliceosome assembly [40] and is composed by three RNA Recognition Motif (RRM) domains connected by flexible linkers. PRE studies revealed a dynamic equilibrium between a predominant compact “closed” conformation and a less abundant “open” RNA-bound-like conformation [40]. This is a key feature in the recognition of a range of polypyrimidine (Py) tracts found in human pre-mRNA introns [40]. However, structures of the open and closed conformations do not completely fit SAXS data, indicating that the range of conformations sampled by U2AF65 in solution is much wider. To study these large-scale dynamic modes that are known to play key roles in a multitude of molecular recognition and signaling processes, the authors used the software ASTEROIDS [41, 42]. The software maps the



**Fig. 22.1** Orientation tensors centered in the center-of-mass of the C-terminal domain of CaM are color-coded with respect to the MO of the corresponding conformation, from *blue* (<5 %) to *red* (>40 %). Two different orientations (panels **a**, **b**) of the tensors are chosen to show

that MO depends on both the relative domain orientation and the position. A high- (**a**) and a low-MO orientation (**b**) are chosen (Reprinted with permission from Bertini et al. [19]. Copyright 2010 American Chemical Society)

conformational space adopted by the protein in an unbiased way using a sequence-dependent stochastic sampling algorithm. Experimental data are used to select from this pool ensembles of conformations. Instead of optimizing the weight of a conformer in the ensemble, the software uses the genetic algorithm to increase the number of copies of a specific conformer in the ensemble in order to obtain a better fitting of the experimental data.

This is conceptually different from the procedure adopted by Bertini et al. [19]. The difference in parameterization allows ASTEROIDS to perform a robust noise-based Monte Carlo error analysis, independently from the quality of the experimental data. The authors used this approach to map the conformational energy surface of the first two RRM domains of U2AF65, inputting RDC, PRE and SAXS data directly into ASTEROIDS. They concluded that the two domains are mainly in an extended conformation while the previously reported “closed” and “open” forms [40] are adopted only by one-quarter of conformers (mostly in the “closed” form). The authors rationalize their results by suggesting that, even if the structures of the “open” and “closed” forms dif-

fer appreciably, they lie within a continuous ensemble envelope and this could represent a possible “pathway” of available states that can flow between the two forms without major expensive energetic jumps.

## 22.4 Hidden Interdomain Information: Direct Structure Refinement

While reducing spin diffusion, perdeuteration also reduces the number of the resonances in the spectrum, including the majority of the resonances necessary for evaluating NOE effects between interdomain side-chains. In the attempt to obtaining experimental information that could compensate for this loss of information, Bax and co-workers [43] implemented SAXS data in NMR structure refinement. As in other SAXS applications,  $\chi^2$  statistics were used to evaluate back-calculated scattering curves during the molecular dynamic/energy minimization steps. The cycles of structure refinement were stopped when convergence, *i.e.*, the minimum of  $\chi^2$ , was reached. To make this procedure “slimmer” from a time and computational point of view, the

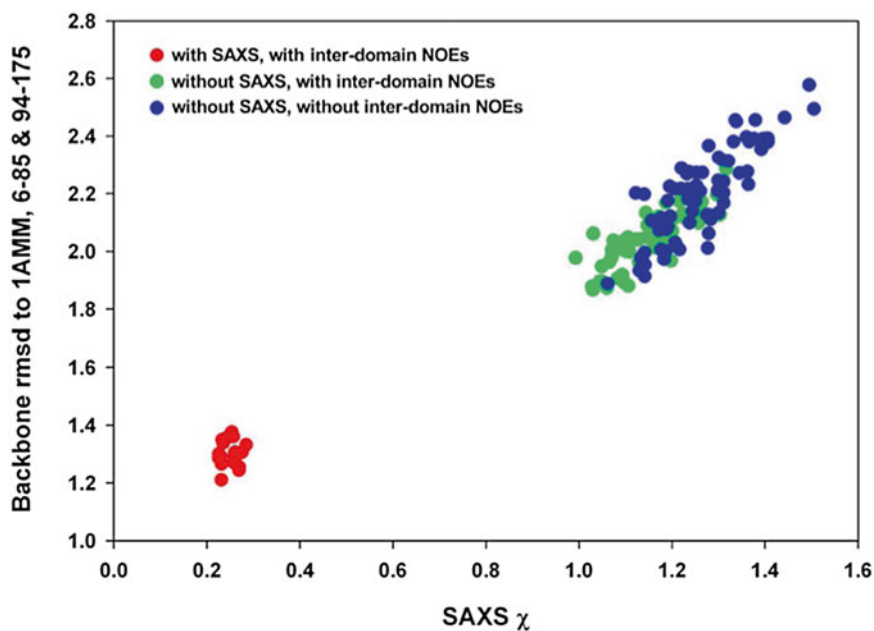
authors used a “glopic approximation”, *i.e.*, they represented the protein structure by small fragments of 3–9 heavy atoms (previously applied for X-ray crystallography [44] and SAXS [45, 46] structure determination) and the SAXS curve by a limited number of data points. The SAXS data-fitting module was implemented into the CNS structure refinement package [47]. Since the first publication, several structures have been solved following this approach (PDB ID: 2A5M, 2JQX, 2K4C, 2KX9, 2XDF, 2L5H) [43, 48–51]. Bax and co-workers tested this approach on  $\gamma$ S-Crystalline [43], malate synthase G (MSG) [48] and on tRNA<sup>Val</sup> [49].

$\gamma$ S-Crystalline is a two-domain protein of 177 residues. The NMR structure of the protein is very similar to that of other homologs (each domain consists of two four-strand  $\beta$ -sheets arranged in Greek key motifs) solved by X-ray crystallography. However, the orientation of the two domains in the NMR structure differs significantly due to lack of interdomain NOE restraints. Refinement including SAXS data allows a better agreement with the structures of other orthologs. The two domains pack closer to each other with a better backbone rmsd as compared to the  $\gamma$ B- and  $\gamma$ D-crystalline [52, 53] crystal structure (respectively without and with SAXS data 1.96–1.31 Å and 1.07–0.87 Å). No major translations are detected and only 5.5° rotation of the C-terminal domain is present when N-terminal domain is used for structure alignment (Fig. 22.2).

Malate synthase G (MSG) [48] is a challenging example, since it is an 82 kDa protein and currently the largest single chain protein solved by solution NMR. MSG catalyzes the chemical reaction between acetyl-CoA and glyoxylate to form malate and CoA. The structure of the enzyme was solved both by X-ray crystallography (PDB ID: 1D8C, 1N8I) [54, 55] and NMR (PDB ID: 1Y8B) [56]. The basic fold of the enzyme is that of a  $\beta$ 8/ $\alpha$ 8 (TIM) barrel with an N-terminal  $\alpha$ -helical domain flanking one side and a C-terminal  $\alpha$ -helical domain forming a plug which caps the active site and a  $\alpha/\beta$  domain with unknown function. Similarly to the  $\gamma$ S-crystalline structure, the inclusion of SAXS data improved the structure refinement and allowed an overall improvement

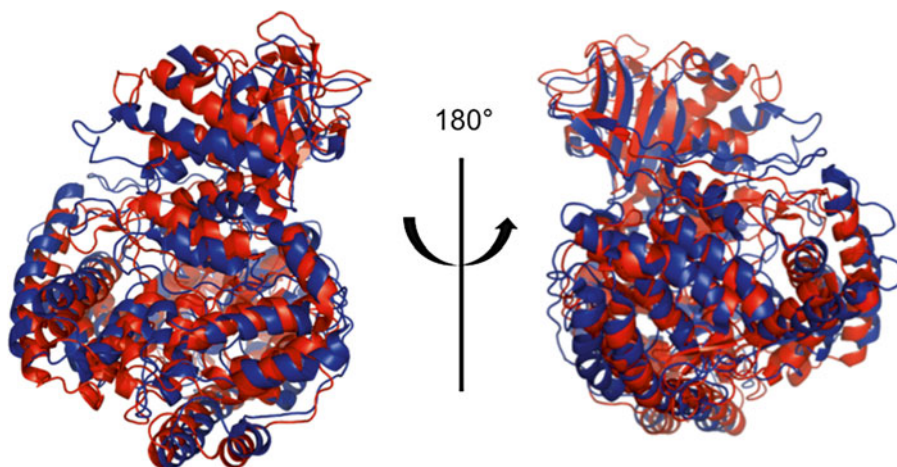
of the backbone root mean square deviation (rmsd) compared to the crystal structure (PDB ID: 1D8C) from 4.92 to 1.39 Å. A translation of  $\sim 4^\circ$ – $5^\circ$  for  $\alpha/\beta$  domain and  $\sim 3^\circ$ – $4^\circ$  for the C-terminal domain between the NMR-only and NMR/SAXS-refined structures was observed (Fig. 22.3). Globular domains like the  $\beta$ 8/ $\alpha$ 8 TIM barrel benefit the most from the introduction of SAXS data to counterbalance the reduction of information consequent to perdeuteration.

Clore and co-workers [57] used a combination of NMR and SAXS data for the study of the full length HIV-1 capsid protein, a challenging system that had given a hard time to several structural biologists. The HIV-1 capsid is a key component in viral infection. It is composed of N- and C-terminal domains connected by a flexible linker (Fig. 22.4a) with the N-terminal domain forming hexameric and pentameric rings (Fig. 22.4b) and the C-terminal domain forming homodimers that connect adjacent N-terminal domain rings [58–62] (Fig. 22.4c). The main problem encountered was caused by the backbone resonances of the linker residues and of residues at the dimer interface of full length HIV-1 capsid protein, which are broad because of monomer/dimer exchange. Similarly to the other two methodologies described above, the authors proceeded with mapping the conformational space sampled by the N-terminal domain relative to the C-terminal domain using a RDC and SAXS/WAXS-driven simulated annealing [50, 63]. It was noticed that the relative orientation of the N- and C-terminal domains does not overlap between the monomeric and dimeric forms and hence the authors postulated that oligomerization acts as a modulator of orientation equilibrium. Interestingly, intra-subunit interactions were detected in the monomeric form. These interactions were driven by the accessible hydrophobic dimerization helix (residues 179–192) of the C-terminal domain that makes contact with residues of the N-terminal domain (Glu29 and Ala31) and the linker region (highlighted by a circle in Fig. 22.4c). On the other hand, the HIV-1 capsid protein dimer is characterized by a single orientation of the C-terminal domain that is in agreement with the previously solved NMR



**Fig.22.2** Plot of the correlation between SAXS  $\chi$  and backbone rmsd to 1AMM (residues 6–85 and 94–175). These results show that calculation of a family of structures with the

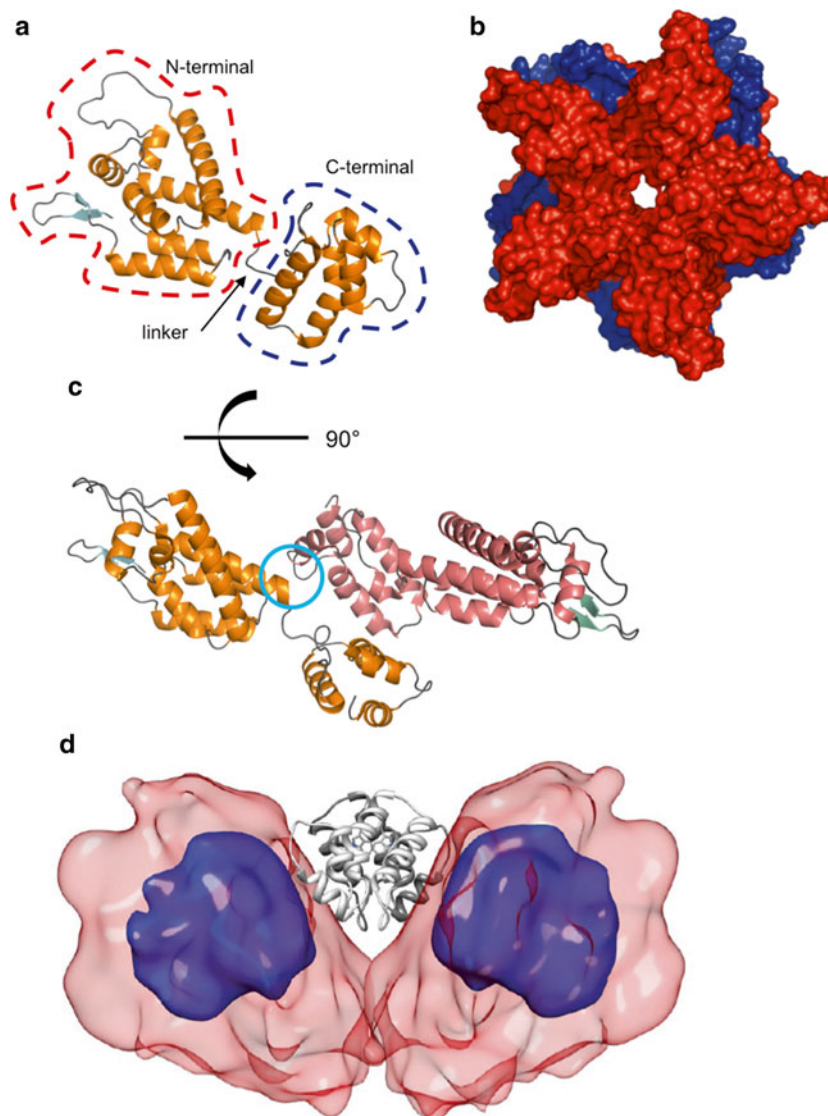
SAXS data produces an improvement in the structural accuracy [43] (Reprinted with permission from Grishaev et al. [43] Copyright 2005 American Chemical Society)



**Fig.22.3** Structural superposition of malate synthase G obtained by the joint fit of SAXS and NMR data (*red*, PDB ID: 2JQX) [48] and the NMR-only model (*blue*, PDB ID: 1Y8B) [56]

structure [64] and in contrast with the crystal structure [65, 66]. Dimerization of the HIV-1 capsid protein prevents the formation of intrasubunit interactions. This information is important for HIV treatment. Hydrophobic capsid assembly inhibitors [67, 68] stabilize, through

hydrophobic interactions, the interaction of the N-terminal domain with the C-terminal one shifting the monomer-dimer equilibrium in favor of the monomer preventing interaction between pentameric or hexameric assemblies and blocking capsid formation.



**Fig. 22.4** The example of the HIV-1 capsid protein. **(a)** Ribbon representation of the full-length monomer ( $\beta$ -strands,  $\alpha$  helices and loops are indicated in *cyan*, *orange* and *grey* respectively) (PDB ID: 2M8N) [57]. The N- and the C-terminal domains and the flexible linker region are highlighted. **(b)** Surface representation of the pentamer (the N-terminal and C-terminal domains are indicated in *red* and *blue* respectively) (PDB ID: 3PO5) [60]. **(c)** Ribbon representation of the full length dimeric protein ( $\beta$ -strands,  $\alpha$  helices and loops are indicated in *cyan*, *orange* and *grey* respectively in Chain A and in *pale*

*green*, *salmon* and *grey* in chain B) (PDB ID: 2M8L) [57]. The contact between the N-terminal domain of chain A and the C-terminal domain of chain B is highlighted. **(d)** Structural ensembles calculated for the full-length monomeric HIV-1 capsid protein. The overall distribution of the N-terminal domain relative to the C-terminal domain (*light* and *dark gray* ribbons) is displayed as a reweighted atomic probability plotted at 50% (*blue*) and 10% (*transparent red*) of the maximum value [57] (Panel **d** was reprinted with permission from Deshmukh et al. [57]. Copyright 2013 American Chemical Society)



Worth mentioning are two other similar approaches which use a combination of NMR and SAXS data. Sattler and co-workers [69] implemented in the CNS package an algorithm which performs a topology refinement of a complex using previously solved structures which are refined against SAXS and NMR RDC data. This approach is based on a first step where the radius of gyration of the complex is used to refine inter-domain distances and a second step in which SAXS data at higher angles are used to define domain positions. This procedure was tested on the barnase/barstar complex [69]. The authors developed further this approach, allowing a combination of SAXS data with any type of NMR restraints in a standard structure calculation set-up [70].

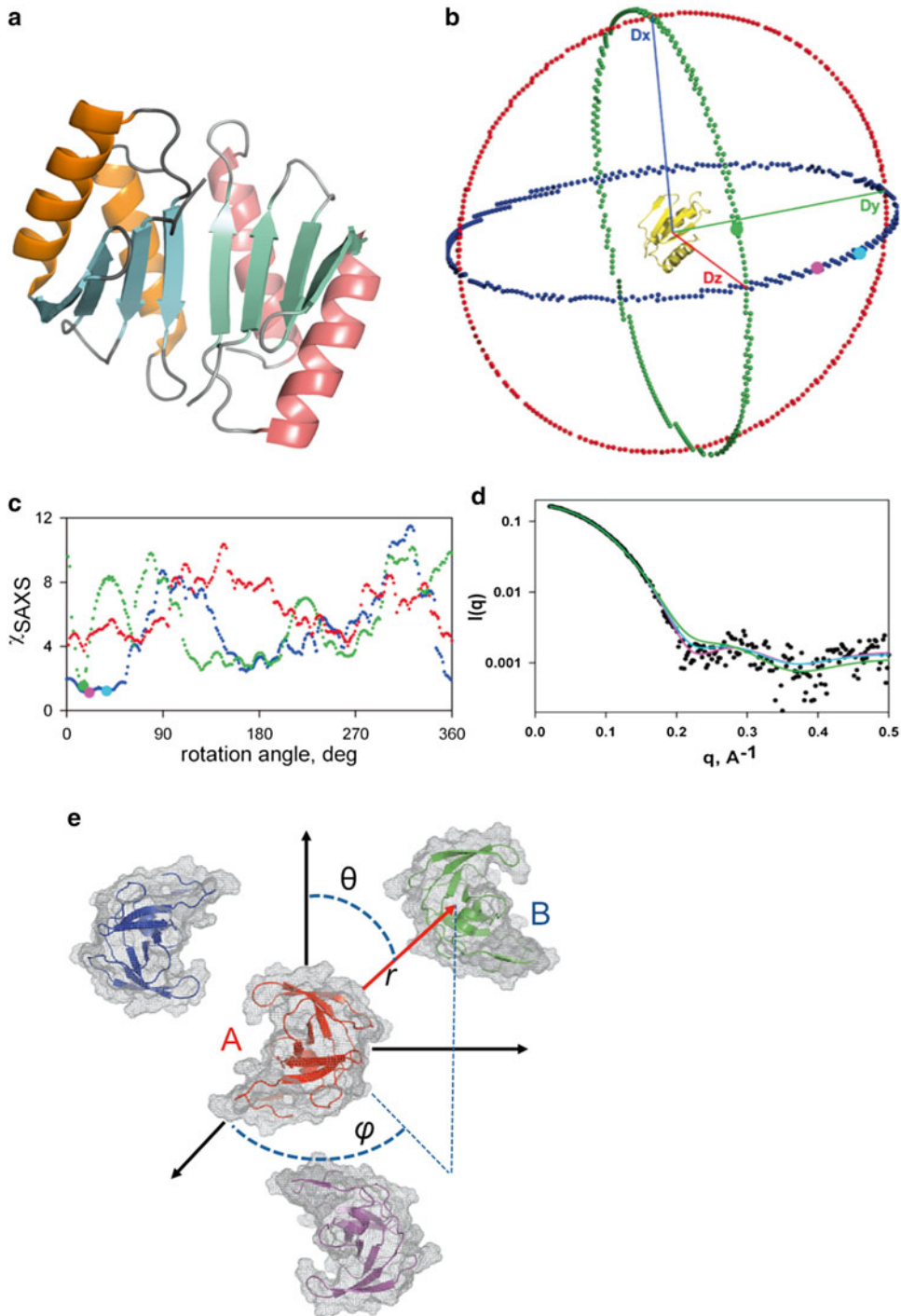
The DADIMODO software was initially born to optimize multidomain homology models using RDC and SAXS data [71]. The algorithm was enhanced and extended to allow refinement of proteins and molecular complexes using NMR derived distance and orientational restraints [72]. This program introduces “mutations” ( $\psi$  and  $\phi$  backbone torsion angles are modified by a random amount), and performs an energy minimization to amend backbone distortions. It then selects the conformations that converge in an energy minimum. Survivors of this first step are fitted versus experimental data, selected and used again in the mutation step. The number of cycles is determined by the user. DADIMODO was tested on the human spire protein (a two WH2 domains protein) and a two-domain fragment of the ribosomal S1 protein.

## 22.5 Multi-subunit Complexes

The combined use of SAXS and NMR was adopted by Parson et al. [73] for the study of TolR. This protein is part of the Pal/Tol system, which forms a five-member, membrane-spanning, multi-protein complex that is involved in several cellular processes (*e.g.*, bacterial outer membrane integrity [74], cell division [75]) and is a potential target for treatment against antibiotic-resistant bacteria. In this study the authors solved the NMR structure of periplasmic domain of

TolR from *Haemophilus influenzae* using conventional NOE restraints. According to gel filtration and light scattering, TolR is a dimer. The protein has a secondary structure  $\beta\beta\beta\alpha\beta\alpha$  with  $\beta 4$  pairing up with the other protomer in an antiparallel manner and forming an eight-stranded  $\beta$ -sheet (Fig. 22.5a). The  $\alpha$ -helices lie on the same side of the  $\beta$ -sheet, with  $\alpha 2$  from each monomer oriented in an antiparallel fashion. The two  $\beta$ -sheet planes of the monomers are twisted by about  $74^\circ$  respect to each other. The authors used only RDC and SAXS data to reconstitute the monomers' orientation in the dimer. Since only a single set of resonances is present in the  $^{15}\text{N}$ - $^1\text{H}$  HSQC, the authors concluded that the TolR dimer must have  $C_2$  symmetry. There are only three possible combinations due to the restriction imposed by this symmetry [76]. If one monomer (A) is fixed at the origin of the RDC alignment tensor frame, the orientation of the other protomer (B) must correspond to the orientation of A rotated by  $180^\circ$  around either of the x, y or z axes of the external frame (Fig. 22.5b). B was then translated along each of these orientations on a  $50 \text{ \AA}$  radius sphere, using a Fibonacci number-based vector grid, generating rings of spherical distribution of B. Arrangements that resulted having a backbone rmsd below  $0.25 \text{ \AA}$  [77] were selected, generating 800 possible dimers. At this point, B was translated towards A generating dimers with a  $2.8 \text{ \AA}$  minimal distance between the two protomers. Further selection was obtained by fitting every dot in panel B using CRYSOLO 2.6 [78] against experimental SAXS data (Fig. 22.5c, d). The best resulting model, with a  $\chi^2 = 1.127$ , has a rmsd of  $0.8 \text{ \AA}$  compared to the TolR dimer solved with conventional NMR methods. This proved that the methodology successfully identified the correct orientation of the two monomers, even if they have a slight translational shift between the two domains. The authors ascribed the success of their studies as compared to previous attempts using SAXS data alone on  $\gamma$ S-Crystalline [43] to the better signal-to noise of the SAXS data and to having a  $C_2$  symmetry.

Similarly, the Wang's group developed an algorithm named GASR (Global Architecture derived from SAXS and RDC) that uses RDC



**Fig. 22.5** (a) Ribbon representation of TolR (PDB ID: 2JWL) [73].  $\beta$ -strands,  $\alpha$  helices and loops are indicated in cyan, orange and grey respectively in Chain A and in pale green, salmon and grey in chain B. (b) Representation of the possible solutions obtained assuming  $C_2$  symmetry.

The centers of mass coordinates of domain B are shown as *solid dots*: blue dots correspond to the case in which Dx is the  $C_2$  axis (correct solution), green and red dots correspond to the cases in which the  $C_2$  axis is along the Dy and Dz axes, respectively [73]. (c) Plot of the  $\chi$  values from

and SAXS data to orient subunits and define the global shape of complexes [2]. They benchmarked the software using five different case studies, which included the HIV protease (homodimeric protein), L11 and  $\gamma$ D-Crystallin (two two-domain proteins, in which the two domains were treated as two independent interacting proteins), GB1 (weak affinity homodimer), and the ILK ankyrin repeat domain bound to the PINCH LIM1 domain (high affinity dimer). GASR uses a rigid body grid search, conceptually similar to the protocol used by Parson et al. [73] (Fig. 22.5e). Differently from the approach described before, GASR runs two grid searches, the second being a fine search on selected structures. The selection is based on  $R_g$ ,  $D_{max}$ , and  $D_{min}$ , which are, respectively, the radius of gyration, the maximum and the minimum linear dimensions between heavy atoms within the two subunits.  $R_g$  and  $D_{max}$  are easily estimated from experimental SAXS data.  $D_{min}$  is set to 1.5 Å for covalently (two domains proteins) linked proteins or to 3.0 Å for transiently interacting proteins, similarly to Parson et al. [73] that used 2.8 Å. Models generated during the second step are analyzed using a probability distribution.

A genuine *de novo* model generated with GASR was recently published by Hirano et al. [14]. In their NMR study of the conformational dynamics of Lys48-linked di-ubiquitin, the authors used GASR to determine the relative orientation and position of the two ubiquitin subunits in a cyclic Lys48-linked di-ubiquitin. The best model generated by the program revealed that the solution structure of cyclic Lys48-linked di-ubiquitin bears a close resemblance to previously reported crystal structures of the non-cyclic counterpart.

Wang et al. [15] applied a procedure similar to that implemented in GASR for the oligomeriza-

tion study of CCL5. This protein is a pro-inflammatory chemokine, which has a propensity for aggregation and is essential for migration *in vivo*, T cell activation and apoptosis, and HIV entry into cells. Previous structural studies had not explored the quaternary conformation of CCL5 higher order oligomers. Initial analysis of NMR, SAXS and DLS data suggested that CCL5 is mainly a tetramer in solution, with the presence of hexamer species. The model generated in this work was obtained through a simple grid search restrained by the symmetry and shape of the tetramer, using a procedure similar to that described by Wang et al. [2]. Model were selected using SAXS data and a favorable binding surface using a residue-pairing score [16]. Due to the presence of tetramer-hexamer equilibrium, the hexamer model was produced by adding an additional dimer unit to the tetramer, duplicating the initial dimer-dimer interface. The tetramer-hexamer ratio was adjusted using OLIGOMER [79] to find the best fit to the SAXS data. The best model fitted the SAXS data with a  $\chi^2 = 1.13$  when using 40 % tetramer and 60 % hexamer. This model forms a tetramer interface which pairs  $\beta 2$  of a protomer in dimer A with the  $\alpha$  helix of a protomer in dimer B. NMR cross saturation experiments were used to confirm the inter-dimer interface defined in the grid search.

The study by NMR of RNA-RNA complexes presents several difficulties mainly attributable to their elongated structures. The number of hydrogen atoms in RNA is small as compared to proteins leading to a considerably reduced proton spin density. NOE experiments are in general rather insensitive and lack of signal dispersion complicates resonance assignment. Although different RDC datasets should always be recorded using different alignment methods, different media give rise to similar alignment tensors for

←  
**Fig. 22.5** (continued) experimental SAXS data versus the rotation angle. The positions of the three best fitting geometries are shown in *magenta*, *cyan*, and *green* [73]. **(d)** Fit of the three dimer geometries to the experimental scattering data (*black dots*) with the color scheme matching panel **c** [73]. **(e)** Illustration of spatial search used in the GASR program for a two-subunit protein in a spherical polar axis system. Shown here are the two subunits of

the HIV-1 protease. Subunit A in *red* is fixed at the axis origin, while subunit B has three discrete possible orientations, depicted in *magenta*, *green* and *blue* “translated” around subunit A without “change in orientation relative to subunit A” [96] (Panels **a-c** were reprinted with permission from Parsons et al. [73]. Copyright 2008 American Chemical Society)

RNA and do not resolve orientation degeneracy [80]. SAXS-aided procedures thus play a major role in the study of RNA-RNA complexes. GASR was tested on a 30 kDa homodimeric tetraloop-receptor RNA complex [81], which is a commonly occurring RNA tertiary structural motif involved in helical packing [82]. This complex structure was previously solved (PDB ID: 2JYJ) using conventional NMR spectroscopy [83, 84]. The rmsd between the best SAXS-defined dimer and the PDB file 2JYJ was found to be 0.4 Å indicating a clear consistency between the two structures. The interaction interfaces were also almost identical including hydrogen bonds and base stacking. The authors stressed the importance of using SAXS data for structural refinement and the substantial difference of the final SAXS-refined model from the model generated without SAXS information, which is much shorter with an rmsd between the two models of 3.2 Å.

## 22.6 A Case Study: Frataxin and the Iron-Sulfur Cluster Machinery

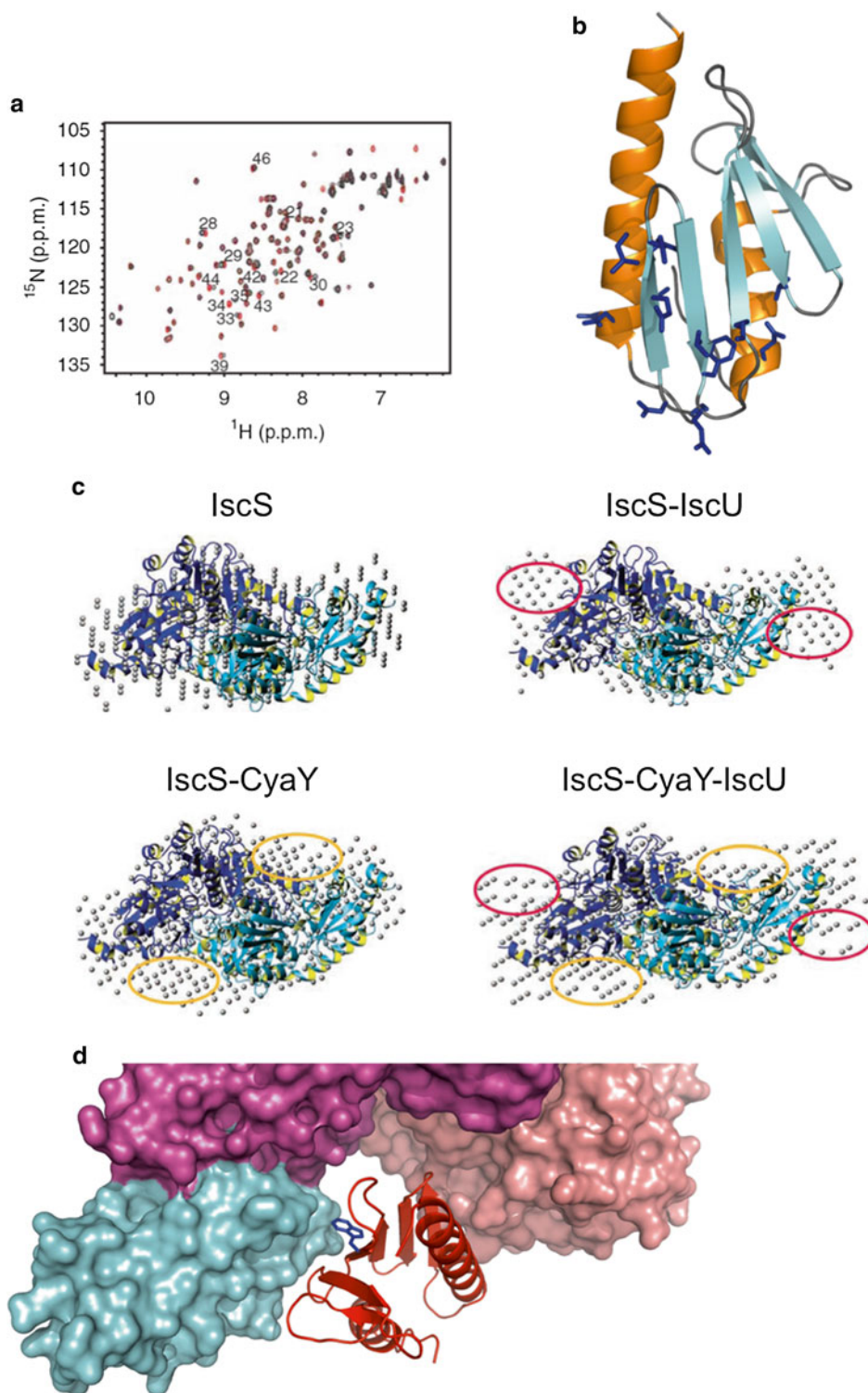
One of the main ongoing projects in our group is the study of Friedreich's ataxia, a relentless and currently incurable neurodegenerative disease [85]. This disease is caused by a reduced expression level of frataxin, an essential iron-binding protein highly conserved from bacteria to humans. Using the bacterial frataxin ortholog, CyaY, we showed that CyaY participates in iron-sulfur (Fe-S) cluster assembly as an iron-dependent inhibitor of cluster formation, through binding to the desulfurase IscS [86]. We proposed that frataxins are iron sensors that act as regulators of Fe-S cluster formation to fine-tune the quantity of Fe-S cluster formed to the concentration of the available

acceptors [86]. This is a highly conserved machine, which ensures the formation of these essential prosthetic groups and their transfer to the final acceptors. Central to the machine are the two components IscS and IscU (using the bacterial name or Nfs1 and Isu in eukaryotes). IscS is a PLP-dependent cysteine desulfurase, which delivers sulfur for Fe-S cluster synthesis to IscU, a Fe-S scaffold protein on which the Fe-S cluster is assembled [87, 88]. To support our enzymology studies with structural evidence, we resolved to model the IscS-IscU complex bound to frataxin using the bacterial proteins. To obtain a molecular description of the IscS-IscU-CyaY complex (where CyaY is the bacterial ortholog of frataxin), we first tried to crystallize the binary complexes obtaining good quality crystals under several different conditions. Unfortunately, they contained only IscS [89]. We thus used an alternative approach based on NMR restrained molecular docking simulations validated by experimental SAXS data. CSP data were used to identify the CyaY and IscU surfaces of interaction with IscS using  $^2\text{H}$ ,  $^{15}\text{N}$  double-labeled CyaY (or IscU) and titrating these proteins with IscS (Figs. 22.6a, b). The interaction surface of IscS was defined by titrating  $^2\text{H}$ ,  $^{15}\text{N}$  double-labelled CyaY (or IscU) with carefully designed IscS mutants, chosen to target residues that could potentially affect the interaction. The docking software HADDOCK [90], which allows the use of "protein interfaces in ambiguous interaction restraints" (AIRs) to drive the docking process, was used to model the ternary complex. Models were then scored and experimentally verified using SAXS data (Fig. 22.6c). Publication of the first high-resolution structure of the IscS-IscU complex gave us confidence in our procedure [91].

Our model also clarified whether frataxin interacts with IscS or with IscU. The question

**Fig. 22.6** (continued) (PDB ID: 1SOY) [96]. Helical and  $\beta$ -sheet regions are indicated in *orange* and *cyan* respectively. The side chains of the residues involved in IscS binding are explicitly shown in *blue*. (c) Comparison of the SAXS densities superposed to the crystal structures of IscS (PDB ID: 1P3W) [97] for IscS alone, IscS/IscU, CyaY/IscS and CyaY/IscS/IscU. Regions with additional

density in the binary and ternary complexes are highlighted in *red* or *yellow ovals*. (d) Final ternary model of the IscS, IscU and CyaY complex. CyaY is shown as a *red ribbon* while the two subunits of IscS homodimer and IscU are shown respectively as *violet*, *pink* and *cyan* molecular surfaces. The side chain of the conserved Trp61 of CyaY is shown in *blue*



**Fig. 22.6** Modeling the ternary complex of IcsS/IscU/CyaY. **(a)** Comparison of the NMR HSQC spectra of  $^{15}\text{N}$  labeled CyaY recorded at 25 °C and 800 MHz in the absence (red) and in the presence (black) of unlabeled IcsS (at a protein ratio of 1:0.8). Residues affected by the titration are marked. **(b)** Ribbon representation of the CyaY structure

was raised because while a direct interaction between human frataxin and the eukaryotic ortholog of IscU had been previously reported [92], experiments on the bacterial proteins had proved negative. We showed experimentally that CyaY packs mainly against IscS while limited interactions with IscU are possible but only in the context of the ternary complex [93] (Fig. 22.6d). Interestingly, the contact surface between CyaY and IscU involves Trp61 of CyaY, a highly conserved residue that is known to be indispensable for the binding of human frataxin with IscU [92]. Our model also indicates that formation of the IscS-IscU-CyaY complex does not require the presence of iron, in contrast to previously published data on the yeast frataxin [94]. On the opposite, the surface of interaction involves direct recognition of a highly negatively charged region of CyaY by a positively charged patch on IscS, thus strongly suggesting that an active role of iron in complex formation is unlikely.

More recently, we applied the same procedure to the study of the Fe-S cluster core machinery (IscS-IscU) in the presence of ferredoxin showing that this protein competes with the same binding site previously determined to accommodate CyaY [95].

## 22.7 Conclusions

It is clear that the future of Structural Biology relies on the combination of different techniques rather than the development of one unique methodology with the hope that this could solve the incredible complexity of Biology. Hybrid methodologies seem to provide a flexible and adaptable answer, which is worth expanding and potentiating. We thus hope that this review contributes to spreading the information and encourages always new groups to develop novel and more powerful approaches to the study of complex systems by NMR.

**Acknowledgments** AP was funded by MRC (Grant ref. U117584256) and the EFACTS EU consortium.

## References

1. Koch MH, Vachette P, Svergun DI (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys* 36(2):147–227
2. Wang J, Zuo X, Yu P, Byeon IJ, Jung J, Wang X, Dyba M, Seifert S, Schwieters CD, Qin J, Gronenborn AM, Wang YX (2009) Determination of multicomponent protein structures in solution using global orientation and shape restraints. *J Am Chem Soc* 131(30):10507–10515
3. Robinson CV, Sali A, Baumeister W (2007) The molecular sociology of the cell. *Nature* 450(7172):973–982
4. Ramakrishnan V (2011) Molecular biology. The eukaryotic ribosome. *Science* 331(6018):681–682
5. Neuhaus D, Williamson MP (2000) The nuclear overhauser effect in structural and conformational analysis, 2nd edn. Wiley, New York
6. Ernst RR, Bodenhausen G, Wokaun A (1997) Principles of nuclear magnetic resonance in one and two dimensions. The international series of monographs on chemistry, vol 14. Clarendon Press; Oxford University Press, Oxford
7. Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced  $^1\text{H}$ [bond] $^{13}\text{C}$  NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* 125(34):10420–10428
8. Clore GM, Gronenborn AM (1989) Determination of three-dimensional structures of proteins and nucleic acids in solution by nuclear magnetic resonance spectroscopy. *Crit Rev Biochem Mol Biol* 24(5):479–564
9. Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278(5340):1111–1114
10. Al-Hashimi HM, Valafar H, Terrell M, Zartler ER, Eidsness MK, Prestegard JH (2000) Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. *J Magn Reson* 143(2):402–406
11. Clore GM (2000) Accurate and rapid docking of protein-protein complexes on the basis of intermolecular nuclear overhauser enhancement data and dipolar couplings by rigid body minimization. *Proc Natl Acad Sci U S A* 97(16):9021–9025
12. Rumpel S, Becker S, Zweckstetter M (2008) High-resolution structure determination of the CylR2 homodimer using paramagnetic relaxation enhancement and structure-based prediction of molecular alignment. *J Biomol NMR* 40(1):1–13
13. Clore GM, Schwieters CD (2003) Docking of protein-protein complexes on the basis of highly ambiguous intermolecular distance restraints derived from

- <sup>1</sup>H/<sup>15</sup>N chemical shift mapping and backbone <sup>15</sup>N-<sup>1</sup>H residual dipolar couplings using conjoined rigid body/torsion angle dynamics. *J Am Chem Soc* 125(10):2902–2912
14. Hirano T, Serve O, Yagi-Utsumi M, Takemoto E, Hiromoto T, Satoh T, Mizushima T, Kato K (2011) Conformational dynamics of wild-type Lys-48-linked diubiquitin in solution. *J Biol Chem* 286(43):37496–37502
  15. Wang X, Watson C, Sharp JS, Handel TM, Prestegard JH (2011) Oligomeric structure of the chemokine CCL5/RANTES from NMR, MS, and SAXS data. *Structure* 19(8):1138–1148
  16. Moont G, Gabb HA, Sternberg MJ (1999) Use of pair potentials across protein interfaces in screening predicted docked complexes. *Proteins* 35(3):364–373
  17. Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D (2007) A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature* 450(7171):913–916
  18. Bernardo P, Blackledge M (2010) Structural biology: proteins in dynamic equilibrium. *Nature* 468(7327):1046–1048
  19. Bertini I, Giachetti A, Luchinat C, Parigi G, Petoukhov MV, Pierattelli R, Ravera E, Svergun DI (2010) Conformational space of flexible biological macromolecules from average data. *J Am Chem Soc* 132(38):13553–13558
  20. Bernardo P, Mylonas E, Petoukhov MV, Blackledge M, Svergun DI (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129(17):5656–5664
  21. Sunnerhagen M, Olah GA, Stenflo J, Forsen S, Drakenberg T, Trehwella J (1996) The relative orientation of Gla and EGF domains in coagulation factor X is altered by Ca<sup>2+</sup> binding to the first EGF domain. A combined NMR-small angle X-ray scattering study. *Biochemistry* 35(36):11547–11559
  22. Neylon C (2008) Small angle neutron and X-ray scattering in structural biology: recent examples from the literature. *Eur Biophys J* 37(5):531–541
  23. Improta S, Krueger JK, Gautel M, Atkinson RA, Lefevre JF, Moulton S, Trehwella J, Pastore A (1998) The assembly of immunoglobulin-like modules in titin: implications for muscle elasticity. *J Mol Biol* 284(3):761–777
  24. Banci L, Bertini I, Bren KL, Cremonini MA, Gray HB, Luchinat C, Turano P (1996) The use of pseudo-contact shifts to refine solution structures of paramagnetic metalloproteins: Met80Ala cytochrome c as an example. *J Biol Inorg Chem* 1:117–126
  25. Svergun D, Barberato C, Koch MH (1995) CRYSOLE – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Crystallogr* 28:768–773
  26. Yamniuk AP, Vogel HJ (2004) Calmodulin's flexibility allows for promiscuity in its interactions with target proteins and peptides. *Mol Biotechnol* 27(1):33–57
  27. Chattopadhyaya R, Meador WE, Means AR, Quioco FA (1992) Calmodulin structure refined at 1.7 Å resolution. *J Mol Biol* 228(4):1177–1192
  28. Babu YS, Bugg CE, Cook WJ (1988) Structure of calmodulin refined at 2.2 Å resolution. *J Mol Biol* 204(1):191–204
  29. Fallon JL, Quioco FA (2003) A closed compact structure of native Ca(2+)-calmodulin. *Structure* 11(10):1303–1307
  30. Maximciuc AA, Putkey JA, Shamoo Y, Mackenzie KR (2006) Complex of calmodulin with a ryanodine receptor target reveals a novel, flexible binding mode. *Structure* 14(10):1547–1556
  31. de Diego I, Kuper J, Bakalova N, Kursula P, Wilmanns M (2010) Molecular basis of the death-associated protein kinase-calcium/calmodulin regulator complex. *Sci Signal* 3(106):ra6
  32. Aoyagi M, Arvai AS, Tainer JA, Getzoff ED (2003) Structural basis for endothelial nitric oxide synthase binding to calmodulin. *EMBO J* 22(4):766–775
  33. Meador WE, Means AR, Quioco FA (1993) Modulation of calmodulin plasticity in molecular recognition on the basis of x-ray structures. *Science* 262(5140):1718–1721
  34. Meador WE, Means AR, Quioco FA (1992) Target enzyme recognition by calmodulin: 2.4 Å structure of a calmodulin-peptide complex. *Science* 257(5074):1251–1255
  35. Kurokawa H, Osawa M, Kurihara H, Katayama N, Tokumitsu H, Swindells MB, Kainosho M, Ikura M (2001) Target-induced conformational adaptation of calmodulin revealed by the crystal structure of a complex with nematode Ca(2+)/calmodulin-dependent kinase kinase peptide. *J Mol Biol* 312(1):59–68
  36. Chou JJ, Li S, Klee CB, Bax A (2001) Solution structure of Ca(2+)-calmodulin reveals flexible hand-like properties of its domains. *Nat Struct Biol* 8(11):990–997
  37. Ikura M, Clore GM, Gronenborn AM, Zhu G, Klee CB, Bax A (1992) Solution structure of a calmodulin-target peptide complex by multidimensional NMR. *Science* 256(5057):632–638
  38. Hoefflich KP, Ikura M (2002) Calmodulin in action: diversity in target recognition and activation mechanisms. *Cell* 108(6):739–742
  39. Huang JR, Warner LR, Sanchez C, Gabel F, Madl T, Mackereth CD, Sattler M, Blackledge M (2014) Transient electrostatic interactions dominate the conformational equilibrium sampled by multidomain splicing factor U2AF65: a combined NMR and SAXS study. *J Am Chem Soc* 136(19):7068–7076
  40. Mackereth CD, Madl T, Bonnal S, Simon B, Zanier K, Gasch A, Rybin V, Valcarcel J, Sattler M (2011) Multi-domain conformational selection underlies pre-mRNA splicing regulation by U2AF. *Nature* 475(7356):408–411

41. Salmon L, Nodet G, Ozenne V, Yin G, Jensen MR, Zweckstetter M, Blackledge M (2010) NMR characterization of long-range order in intrinsically disordered proteins. *J Am Chem Soc* 132(24):8407–8418
42. Guerry P, Salmon L, Mollica L, Ortega Roldan JL, Markwick P, van Nuland NA, McCammon JA, Blackledge M (2013) Mapping the population of protein conformational energy sub-states from NMR dipolar couplings. *Angew Chem Int Ed Engl* 52(11):3181–3185
43. Grishaev A, Wu J, Trehwella J, Bax A (2005) Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J Am Chem Soc* 127(47):16621–16628
44. Guo DY, Blessing RH, Langs DA (2000) Globbic approximation in low-resolution direct-methods phasing. *Acta Crystallogr D Biol Crystallogr* 56(Pt 9):1148–1155
45. Svergun DI, Petoukhov MV, Koch MH (2001) Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80(6):2946–2953
46. Chacon P, Moran F, Diaz JF, Pantos E, Andreu JM (1998) Low-resolution structures of proteins in solution retrieved from X-ray scattering with a genetic algorithm. *Biophys J* 74(6):2760–2775
47. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr* 54(Pt 5):905–921
48. Grishaev A, Tugarinov V, Kay LE, Trehwella J, Bax A (2008) Refined solution structure of the 82- kDa enzyme malate synthase G from joint NMR and synchrotron SAXS restraints. *J Biomol NMR* 40(2):95–106
49. Grishaev A, Ying J, Canny MD, Pardi A, Bax A (2008) Solution structure of tRNAVal from refinement of homology model against residual dipolar coupling and SAXS data. *J Biomol NMR* 42(2):99–109
50. Schwieters CD, Suh JY, Grishaev A, Ghirlando R, Takayama Y, Clore GM (2010) Solution structure of the 128 kDa enzyme I dimer from *Escherichia coli* and its 146 kDa complex with HPr using residual dipolar couplings and small- and wide-angle X-ray scattering. *J Am Chem Soc* 132(37):13026–13045
51. Takayama Y, Schwieters CD, Grishaev A, Ghirlando R, Clore GM (2011) Combined use of residual dipolar couplings and solution X-ray scattering to rapidly probe rigid-body conformational transitions in a non-phosphorylatable active-site mutant of the 128 kDa enzyme I dimer. *J Am Chem Soc* 133(3):424–427
52. Kumaraswamy VS, Lindley PF, Slingsby C, Glover ID (1996) An eye lens protein-water structure: 1.2 Å resolution structure of gammaB-crystallin at 150 K. *Acta Crystallogr D Biol Crystallogr* 52(Pt 4):611–622
53. Basak A, Bateman O, Slingsby C, Pande A, Asherie N, Ogun O, Benedek GB, Pande J (2003) High-resolution X-ray crystal structures of human gammaD crystallin (1.25 Å) and the R58H mutant (1.15 Å) associated with aculeiform cataract. *J Mol Biol* 328(5):1137–1147
54. Smith CV, Huang CC, Miczak A, Russell DG, Sacchettini JC, Honer zu Bentrup K (2003) Biochemical and structural studies of malate synthase from *Mycobacterium tuberculosis*. *J Biol Chem* 278(3):1735–1743
55. Howard BR, Endrizzi JA, Remington SJ (2000) Crystal structure of *Escherichia coli* malate synthase G complexed with magnesium and glyoxylate at 2.0 Å resolution: mechanistic implications. *Biochemistry* 39(11):3156–3168
56. Tugarinov V, Choy WY, Orekhov VY, Kay LE (2005) Solution NMR-derived global fold of a monomeric 82-kDa enzyme. *Proc Natl Acad Sci U S A* 102(3):622–627
57. Deshmukh L, Schwieters CD, Grishaev A, Ghirlando R, Baber JL, Clore GM (2013) Structure and dynamics of full-length HIV-1 capsid protein in solution. *J Am Chem Soc* 135(43):16133–16147
58. Ganser BK, Li S, Klishko VY, Finch JT, Sundquist WI (1999) Assembly and analysis of conical models for the HIV-1 core. *Science* 283(5398):80–83
59. Ganser-Pornillos BK, von Schwedler UK, Stray KM, Aiken C, Sundquist WI (2004) Assembly properties of the human immunodeficiency virus type 1 CA protein. *J Virol* 78(5):2545–2552
60. Pornillos O, Ganser-Pornillos BK, Yeager M (2011) Atomic-level modelling of the HIV capsid. *Nature* 469(7330):424–427
61. Borsetti A, Ohagen A, Gottlinger HG (1998) The C-terminal half of the human immunodeficiency virus type 1 Gag precursor is sufficient for efficient particle assembly. *J Virol* 72(11):9313–9317
62. Accola MA, Strack B, Gottlinger HG (2000) Efficient particle production by minimal Gag constructs which retain the carboxy-terminal domain of human immunodeficiency virus type 1 capsid-p2 and a late assembly domain. *J Virol* 74(12):5395–5402
63. Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160(1):65–73
64. Byeon IJ, Meng X, Jung J, Zhao G, Yang R, Ahn J, Shi J, Concel J, Aiken C, Zhang P, Gronenborn AM (2009) Structural convergence between Cryo-EM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell* 139(4):780–790
65. Gamble TR, Yoo S, Vajdos FF, von Schwedler UK, Worthylake DK, Wang H, McCutcheon JP, Sundquist WI, Hill CP (1997) Structure of the carboxyl-terminal dimerization domain of the HIV-1 capsid protein. *Science* 278(5339):849–853
66. Worthylake DK, Wang H, Yoo S, Sundquist WI, Hill CP (1999) Structures of the HIV-1 capsid protein dimerization domain at 2.6 Å resolution. *Acta Crystallogr D Biol Crystallogr* 55(Pt 1):85–92



67. Kelly BN, Kyere S, Kinde I, Tang C, Howard BR, Robinson H, Sundquist WI, Summers MF, Hill CP (2007) Structure of the antiviral assembly inhibitor CAP-1 complex with the HIV-1 CA protein. *J Mol Biol* 373(2):355–366
68. Lemke CT, Titolo S, von Schwedler U, Goudreau N, Mercier JF, Wardrop E, Faucher AM, Coulombe R, Banik SS, Fader L, Gagnon A, Kawai SH, Rancourt J, Tremblay M, Yoakim C, Simoneau B, Archambault J, Sundquist WI, Mason SW (2012) Distinct effects of two HIV-1 capsid assembly inhibitor families that bind the same site within the N-terminal domain of the viral CA protein. *J Virol* 86(12):6643–6655
69. Gabel F, Simon B, Sattler M (2006) A target function for quaternary structural refinement from small angle scattering and NMR orientational restraints. *Eur Biophys J* 35(4):313–327
70. Gabel F, Simon B, Nilges M, Petoukhov M, Svergun D, Sattler M (2008) A structure refinement protocol combining NMR residual dipolar couplings and small angle scattering restraints. *J Biomol NMR* 41(4):199–208
71. Mareuil F, Sizun C, Perez J, Schoenauer M, Lallemand JY, Bontems F (2007) A simple genetic algorithm for the optimization of multidomain protein homology models driven by NMR residual dipolar coupling and small angle X-ray scattering data. *Eur Biophys J* 37(1):95–104
72. Evrard G, Mareuil F, Bontems F, Sizun C, Perez J (2011) DADIMODO: a program for refining the structure of multidomain proteins and complexes against small-angle scattering data and NMR-derived restraints. *J Appl Crystallogr* 44:7
73. Parsons LM, Grishaev A, Bax A (2008) The periplasmic domain of TolR from *Haemophilus influenzae* forms a dimer with a large hydrophobic groove: NMR solution structure and comparison to SAXS data. *Biochemistry* 47(10):3131–3142
74. Cascales E, Bernadac A, Gavioli M, Lazzaroni JC, Llobes R (2002) Pal lipoprotein of *Escherichia coli* plays a major role in outer membrane integrity. *J Bacteriol* 184(3):754–759
75. Gerding MA, Ogata Y, Pecora ND, Niki H, de Boer PA (2007) The trans-envelope Tol-Pal complex is part of the cell division machinery and required for proper outer-membrane invagination during cell constriction in *E. coli*. *Mol Microbiol* 63(4):1008–1025
76. Al-Hashimi HM, Bolon PJ, Prestegard JH (2000) Molecular symmetry as an aid to geometry determination in ligand protein complexes. *J Magn Reson* 142(1):153–158
77. Dam J, Baber J, Grishaev A, Malchiodi EL, Schuck P, Bax A, Mariuzza RA (2006) Variable dimerization of the Ly49A natural killer cell receptor results in differential engagement of its MHC class I ligand. *J Mol Biol* 362(1):102–113
78. Konarev PV, Petoukhov MV, Volkov VV, Svergun DI (2006) ATSAS 2.1, a program package for small-angle scattering data analysis. *J Appl Crystallogr* 39:277–286
79. Latham MP, Hanson P, Brown DJ, Pardi A (2008) Comparison of alignment tensors generated for native tRNA(Val) using magnetic fields and liquid crystalline media. *J Biomol NMR* 40(2):83–94
80. Zuo X, Wang J, Foster TR, Schwieters CD, Tiede DM, Butcher SE, Wang YX (2008) Global molecular structure and interfaces: refining an RNA:RNA complex structure using solution X-ray scattering data. *J Am Chem Soc* 130(11):3292–3293
81. Costa M, Michel F (1995) Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J* 14(6):1276–1285
82. Davis JH, Tonelli M, Scott LG, Jaeger L, Williamson JR, Butcher SE (2005) RNA helical packing in solution: NMR structure of a 30 kDa GAAA tetraloop-receptor complex. *J Mol Biol* 351(2):371–382
83. Davis JH, Foster TR, Tonelli M, Butcher SE (2007) Role of metal ions in the tetraloop-receptor complex as analyzed by NMR. *RNA* 13(1):76–86
84. Pandolfo M, Pastore A (2009) The pathogenesis of Friedreich ataxia and the structure and function of frataxin. *J Neurol* 256(Suppl 1):9–17
85. Adinolfi S, Iannuzzi C, Prischi F, Pastore C, Iametti S, Martin SR, Bonomi F, Pastore A (2009) Bacterial frataxin CyaY is the gatekeeper of iron-sulfur cluster formation catalyzed by IscS. *Nat Struct Mol Biol* 16(4):390–396
86. Urbina HD, Silberg JJ, Hoff KG, Vickery LE (2001) Transfer of sulfur from IscS to IscU during Fe/S cluster assembly. *J Biol Chem* 276(48):44521–44526
87. Adinolfi S, Rizzo F, Masino L, Nair M, Martin SR, Pastore A, Temussi PA (2004) Bacterial IscU is a well folded and functional single domain protein. *Eur J Biochem* 271(11):2093–2100
88. Prischi F, Pastore C, Carroni M, Iannuzzi C, Adinolfi S, Temussi P, Pastore A (2010) Of the vulnerability of orphan complex proteins: the case study of the *E. coli* IscU and IscS proteins. *Protein Expr Purif* 73(2):161–166
89. Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 125(7):1731–1737
90. Shi R, Proteau A, Villarroya M, Moukadir I, Zhang L, Trempe JF, Matte A, Armengod ME, Cygler M (2010) Structural basis for Fe-S cluster assembly and tRNA thiolation mediated by IscS protein-protein interactions. *PLoS Biol* 8(4), e1000354
91. Yoon T, Cowan JA (2003) Iron-sulfur cluster biosynthesis. Characterization of frataxin as an iron donor for assembly of [2Fe-2S] clusters in ISU-type proteins. *J Am Chem Soc* 125(20):6078–6084
92. Prischi F, Konarev PV, Iannuzzi C, Pastore C, Adinolfi S, Martin SR, Svergun DI, Pastore A (2010) Structural bases for the interaction of frataxin with the central components of iron-sulphur cluster assembly. *Nat Commun* 1:95

93. Li H, Gakh O, Smith DY, Isaya G (2009) Oligomeric yeast frataxin drives assembly of core machinery for mitochondrial iron-sulfur cluster synthesis. *J Biol Chem* 284(33):21971–21980
94. Yan R, Konarev PV, Iannuzzi C, Adinolfi S, Roche B, Kelly G, Simon L, Martin SR, Py B, Barras F, Svergun DI, Pastore A (2013) Ferredoxin competes with bacterial frataxin in binding to the desulfurase IscS. *J Biol Chem* 288(34):24777–24787
95. Yamazaki T, Hinck AP, Wang YX, Nicholson LK, Torchia DA, Wingfield P, Stahl SJ, Kaufman JD, Chang CH, Domaille PJ, Lam PY (1996) Three-dimensional solution structure of the HIV-1 protease complexed with DMP323, a novel cyclic urea-type inhibitor, determined by nuclear magnetic resonance spectroscopy. *Protein Sci* 5(3):495–506
96. Nair M, Adinolfi S, Pastore C, Kelly G, Temussi P, Pastore A (2004) Solution structure of the bacterial frataxin ortholog, CyaY: mapping the iron binding sites. *Structure* 12(11):2037–2048
97. Cupp-Vickery JR, Urbina H, Vickery LE (2003) Crystal structure of IscS, a cysteine desulfurase from *Escherichia coli*. *J Mol Biol* 330(5):1049–1059