UNIVERSITÀ
DEGLI STUDI
FIRENZE

DOTTORATO DI RICERCA
INTERNATIONAL DOCTORATE IN STRUCTURAL BIOLOGY

CICLO XXIX

COORDINATORE Prof. Claudio Luchinat

Computational aspects of NMR in structural biology

Settore Scientifico Disciplinare CHIM/03

Dottorando	Tutore
Dott. *Azzurra Carlon*	Prof. Claudio Luchinat

Coordinatore
Prof. Claudio Luchinat

Novembre 2013 - 2016

Table of contents

1. Introduction 1

1.1 User-friendly web application for paramagnetic NMR data 6

1.2 Integration of paramagnetic NMR and X-ray data..8

1.3 Systematic characterization of multi-domain systems 12

2. Methodological Aspects 17

2.1 Paramagnetic and diamagnetic NMR-based restraints...................................17

2.1.1 Magnetic susceptibility and its anisotropy 17

2.1.2 Pseudo-contact shifts (PCSs)..18

2.1.3 Paramagnetic residual dipolar couplings (paramagnetic RDCs)...........19

2.1.4 Diamagnetic residual dipolar couplings (diamagnetic RDCs) 20

2.1.5 The Q-factor 21

2.2 Software tools for NMR data 22

2.2.1 FANTEN web-application..22

2.2.2 Constrained optimization in REFMAC-NMR 22

3. Results 23

3.1 FANTEN: a new web-based interface for the analysis of magnetic anisotropy-induced NMR data 23

3.2 Extending the possibilities of joint X-ray/NMR structure refinement to high molecular weight complexes...47

3.3 How to tackle protein structural data from solution and solid state: An integrated approach 73

3.4 Improved Accuracy from Joint X-ray and NMR Refinement of a Protein–RNA Complex Structure..93

3.5 Impact of experimental uncertainties in protein mobility studies 115

4. Conclusions 147

Bibliography 151

1. Introduction

Structural molecular biology is the study of the structure of biological macromolecules, such as proteins and nucleic acids. Such structural information is then linked to biomolecular functioning in the biological systems.

Over the years, several techniques have been developed for tackling the complexity of macromolecular systems and to improve our knowledge in many aspects of structural biology at different levels of resolution, i.e. from the atomistic details of individual proteins or protein domains up to the information related to their conformation, position and orientation in a macromolecular assembly or biological system, both in vitro and in their cellular context. These techniques include NMR spectroscopy, X-ray and neutron crystallography, electron microscopy and tomography, small angle X-ray and neutron scattering, mass spectroscopy and advanced light microscopy techniques (Figure 1). Each of these techniques provides very useful biological insights, but their combination is expected to yield more profound advancement: i.e.: the information provided by these techniques as a whole is much greater than the sum of each of them taken as stand-alone tools. Therefore, we can expect that hybrid methods will play an essential role in the near future of structural biology leading to a deeper and more complete understanding of many important biological problems.[1]

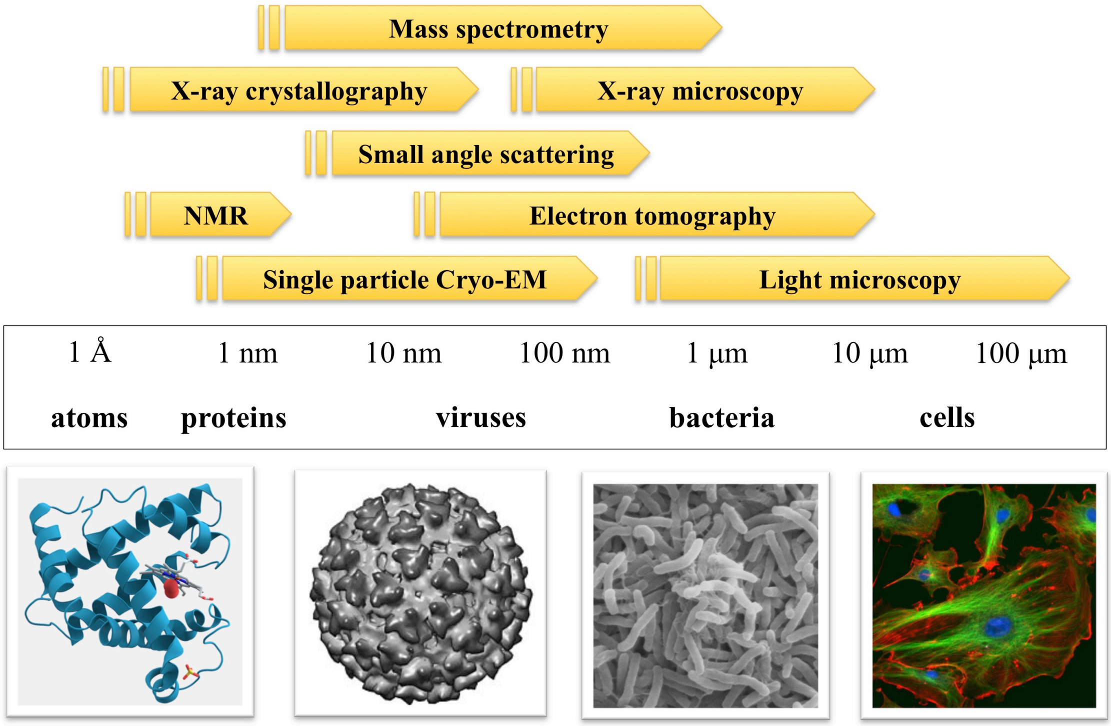


Figure 1: Overview of the most commonly used techniques in structural biology.

During the last decades, a number of biological systems have been investigated by the combination of a high- and low- resolution structural techniques achieving, in most of the cases, groundbreaking results. The study of HIV capsid formation is an important example where cryo-electron tomography has been used in combination with molecular dynamic simulations as well as previously solved X-ray and NMR structures to reconstruct the virus tubular assembly.[2,3] Through the integration of these techniques, Zhang P., Gronenborn A., and co-workers have been able to identify novel inter-molecular interactions that are considered responsible to provide the necessary plasticity for the assembly and disassembly of the virus capsid. Another remarkable example of integration of multiple structural methodologies has been presented by Sgourakis N.G., Lange A. and co-workers where cryo-electron microscopy is used as a constraint to aid the determination of solid state NMR structures.[4] Sattler M. and co-workers proposed an efficient protocol for the determination of protein complexes in solution, starting from individual domain obtained from X-ray crystallography, NMR spectroscopy or homology modeling and determining the relative position and orientation of the domains by taking advantage of the measurements in solution of residual dipolar couplings (RDCs) and paramagnetic relaxation effects (PREs), respectively.[5] Based on the increasing need of combining a multitude of different information, an Integrated Modeling Platform (IMP) software package[6–8] has been recently developed to facilitate the process of data integration for the determination of the structure of macromolecular assemblies, providing a frame in which the resulting models can be easily evaluated and re-refined when new information is gathered. IMP includes the use not only of the most widespread experimental techniques used in structural biology (X-ray crystallography, NMR spectroscopy, electron microscopy, chemical cross-linking, FRET spectroscopy, small angle X-ray scattering, proteomics data) but also of more theoretical sources of information, such as comparative modeling, statistical inference and physics-based energy functions. This integrative modeling approach has been already used to solve numerous structures [9,10,5,11,12], like in the case of 26S proteasome, where Baumeister W., Sali A. and co-workers were able to determine its molecular architecture based on data from cryo-electron microscopy, X-ray crystallography, residue-specific chemical cross-linking, and several proteomics techniques.[12]

Among all the available methodologies, NMR spectroscopy, X-ray crystallography and, to a lesser extent, single molecule cryo-electron microscopy are the only high-resolution techniques able to provide atomic-level structural information of biomolecular systems and the most commonly used among the deposited structure in the Protein Data Bank (Figure 2). X-ray diffraction is a very powerful and reliable technique and it is by far the most widely used methodology for solving protein structures, accounting as of August 2016 for 90% and 60% of all deposited protein and nucleic acid structures, respectively. The strength of X-ray crystallography lies in a precise determination of a unique macromolecular structure or of a unique set of different structures present in the asymmetric unit of the crystal. NMR spectroscopy has the unique capability to provide information on the internal dynamics of the biomolecular system in solution and over a wide range of timescales. In this way, by combining structural and dynamic aspects, NMR spectroscopy affords a complete picture of the behavior of proteins, which cannot be fulfilled by the use of any other technique. Nevertheless, neither X-ray nor NMR, if used as stand-alone tools, can provide a complete, precise and accurate picture of the biological system under investigation and of its interactions with other complexes or biomolecules. It should also be mentioned that X-ray crystallographic diffraction and NMR data are intrinsically different - the former gives information that progresses from the overall shape of the molecule up to individual atom positions as the resolution increases, whereas NMR provides immediate information about short-range inter-atomic distances and bond orientations, which progresses to overall shape of the molecule with increasing number and quality of restraints. Therefore, the combination of these two highly complementary techniques is mutually beneficial, yielding valuable information throughout the spectrum of distance scales, even in the presence of suboptimal X-ray and/or NMR data.

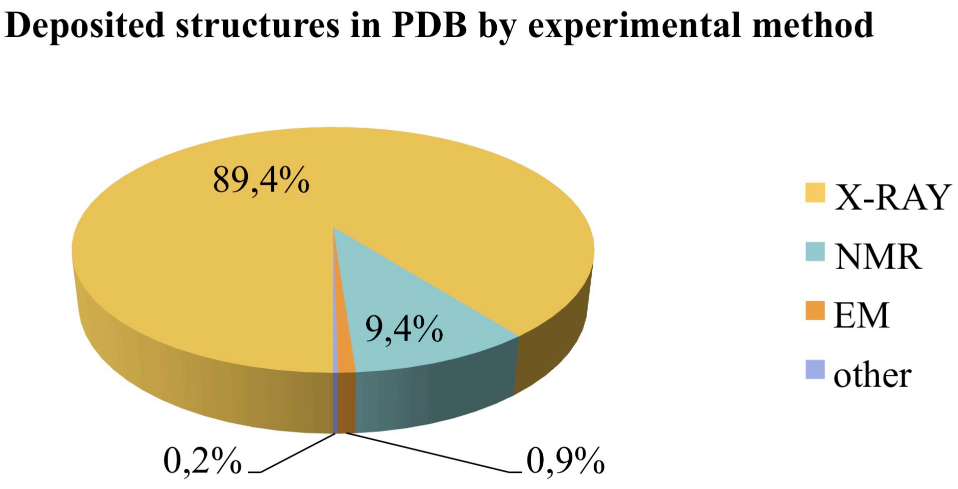


Figure 2: Statistics of the experimental methods used for solving the structures deposited in the Protein Data Bank (PDB) up to 2016.

Among all the structural restraints that can be obtained by NMR-based experiments, pseudo-contact shifts (PCSs)[13] and residual dipolar couplings (RDCs)[14] gained increasing popularity within the scientific community during the last decades due to their intrinsic capability of providing long-range information. PCSs and RDCs originate from the presence of a paramagnetic metal in the molecule, which results in an interaction of the unpaired electron with the nuclear spins of the molecule.[15] The unpaired electron spin, averaged by the fast electron relaxation results in distance- and/or orientation- dependent effects that can be exploited as structural restraints.[15–17] More specifically, PCSs result from chemical shifts perturbations due the dipolar interaction between paramagnetic centre and the nuclei belonging to the molecule, and can be exploited as structural restraints due to its distance - and orientation- dependence (Figure 3-A).[15–17] RDCs, instead, arise from the partial alignment of the molecule in the magnetic field due to the magnetic susceptibility anisotropy associated to the paramagnetic centre, which leads to an incomplete averaging of the nuclear dipole-

dipole coupling and thus results in a modulation of J-couplings between nuclear pairs that is orientation dependent (Figure 3-B). In the case the paramagnetic centre is not originally present in the molecule, the metal can be included by substitution of a diamagnetic metal ion[18–23] or, alternatively, rigidly attached through tags[24–38]. Partial alignment of the molecule in the magnetic field can alternatively be induced by external alignment media (i.e. bicelles[39], poly-acrylamide gels[40], bacteriophages[41]) to collect RDCs, defined "diamagnetic" in antithesis to the "paramagnetic" version.[40,42–50]

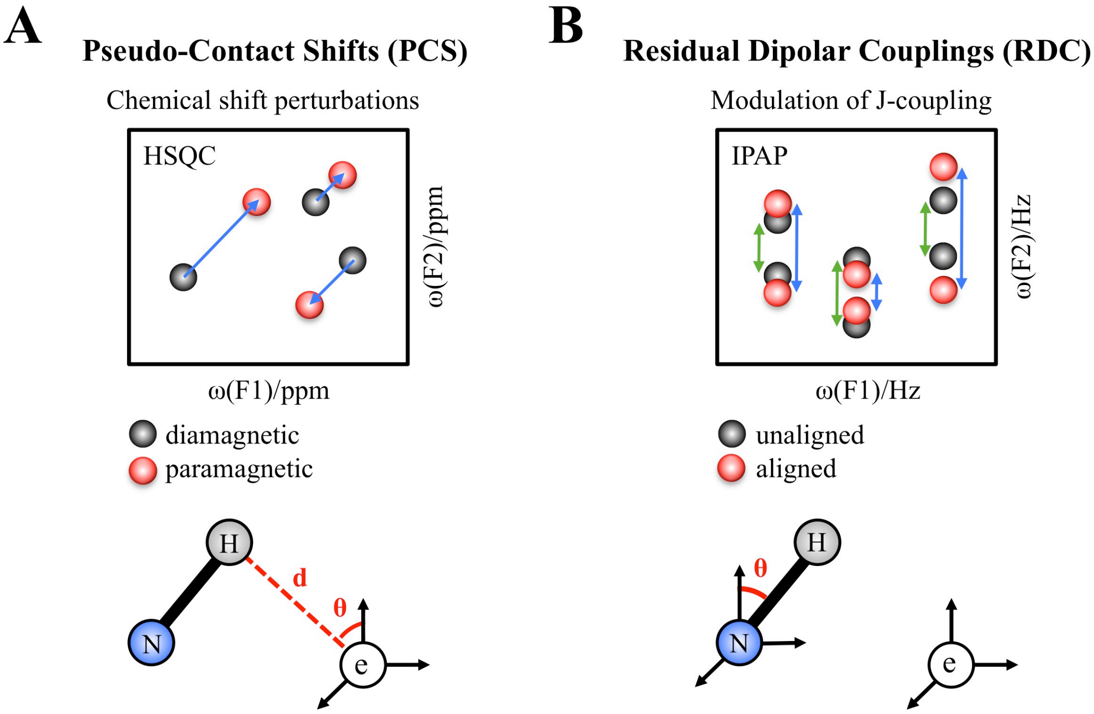


Figure 3: Schematics representing how paramagnetic restraints (PCSs and RDCs) are practically measured. (A) For PCSs the chemical shift differences of NMR resonances between a paramagnetic and a diamagnetic protein are measured in a HSQC experiment. (B) RDCs are extracted from the observed splitting in an IPAP experiment that is decoupled in one dimension. The lower panel displays the structural parameters that are measured for the different types of restraints. The frame of reference is the magnetic susceptibility tensor frame associated with the unpaired electron. A single NH vector is displayed in this reference frame and θ describes the displayed angle in polar coordinates. Modified figure taken from Koehler J. & Meiler J[51].

Nowadays, PCSs and RDCs are routinely measured through NMR experiments and used for solving protein structures. Various programs (i.e. Xplor-NIH[52–54], PARAMAGNETIC CYANA[55,56], PCS-HADDOCK[57], PCS-ROSETTA[58], PATIDOCK[59]) for protein structure determination from NMR data allow the use of PCSs and RDCs as structural restraints. PCSs and RDCs proved to be even more helpful for the characterization of very challenging systems, such as multi-domain proteins and protein complexes, because their long-range nature allows for retrieving the information about either static or dynamic conformational variability. Indeed, the comparison of the paramagnetic tensors calculated using PCSs and/or RDCs for individual domains or binding partners provide the direct detection and quantification of the conformational arrangements and dynamic effects present in the system.[27,51,60–62]

In paragraph 1.1, FANTEN will be presented: a user-friendly web application for the determination of the anisotropy and alignment tensors related to PCS and RDC data. In paragraph 1.2, a protocol will be described for the integration of NMR and X-ray crystallographic data and the latest developments that allow for the inclusion of constraints in the tensor calculations. Finally, in paragraph 1.3, it will be shown how the previous programs can be used and efficiently combined together to accomplish a systematic characterization of multi-domain proteins and complex systems.

1.1 User-friendly web application for paramagnetic NMR data

Many computational tools have been implemented in order to allow for the use of paramagnetic data and to aid the correct analysis and interpretation of the results (FANTASIA[55], NUMBAT[63], PALES[64], PATI[65], REDCAT[66]). Laboratory experience is of prior importance for understanding the set of needs and of problems that a scientist might experience when working with this kind of data. Even more essential is to make such knowledge available to the wider scientific community. For such a purpose, a web application for the treatment and the analysis of PCSs and RDCs has been implemented in the program FANTEN (Finding ANisotropy TENsor), available through the We-NMR portal[67] (http://abs.cerm.unifi.it:8080/). FANTEN has been developed not only to provide a tool easily accessible for the scientific community but it has been designed with particular focus on user-friendliness, which frequently represents one of the major hurdles for scientist non-expert in the field.

Contribution of my thesis

The program is organized in three main parts (also termed “actions”), which have been designed to perform different types of calculations and with different levels of complexity. The action *“PCS-RDC Fitting (Smart)”* has been implemented to perform the most commonly required types of calculations in very easy and quick way. This action allows for the calculation of tensors for multiple datasets of PCSs and RDCs and for multiple models present in the PDB file. In one single step, all the tensors are computed and the results of the fitting can be readily visualized and examined through the correlation plots. Moreover, this action provides some additional features, such as the generation and the visualization of the iso-PCS surfaces using an integrated JSmol application, and the estimation of the error on the tensor parameters using a Monte-Carlo bootstrap approach. The action *“PCS-RDC Fitting (Custom)”* allows for a fully customized calculation of the tensor where, for any single dataset, the user can define (if known) the metal position, the tensor orientation and the anisotropy parameters, or any combination thereof. Moreover, some global actions are available, such as the calculations using the same metal position for all the defined tensors or the same tensor for all the uploaded datasets. The first operation is very useful when the position of the paramagnetic centre is unknown but can be reasonably assumed to be the same for different metals. The second operation is important if the experimental data have been collected from different binding partners of a complex but using the same experimental conditions, therefore it can be reasonably assumed that the tensor is the same. Finally, the action *“Rigid Body Minimization”* has been realized to support the calculation of the optimal arrangement between two binding partners or protein domains according to PCSs and RDCs restraints. In this case, the rigid rototranslation to apply to the second structure provided by the user is computed to obtain the best agreement with the available structural restraints. For each action, all the results in terms of tensor parameters, data fittings and rototranslated structures are fully available for download. An Overview of some of the tools and functionalities available in FANTEN web application is shown in Figure 4.

FANTEN web application is presented in detail in *“FANTEN: a new web-based interface for the analysis of magnetic anisotropy-induced NMR data”* in Section 3.1.

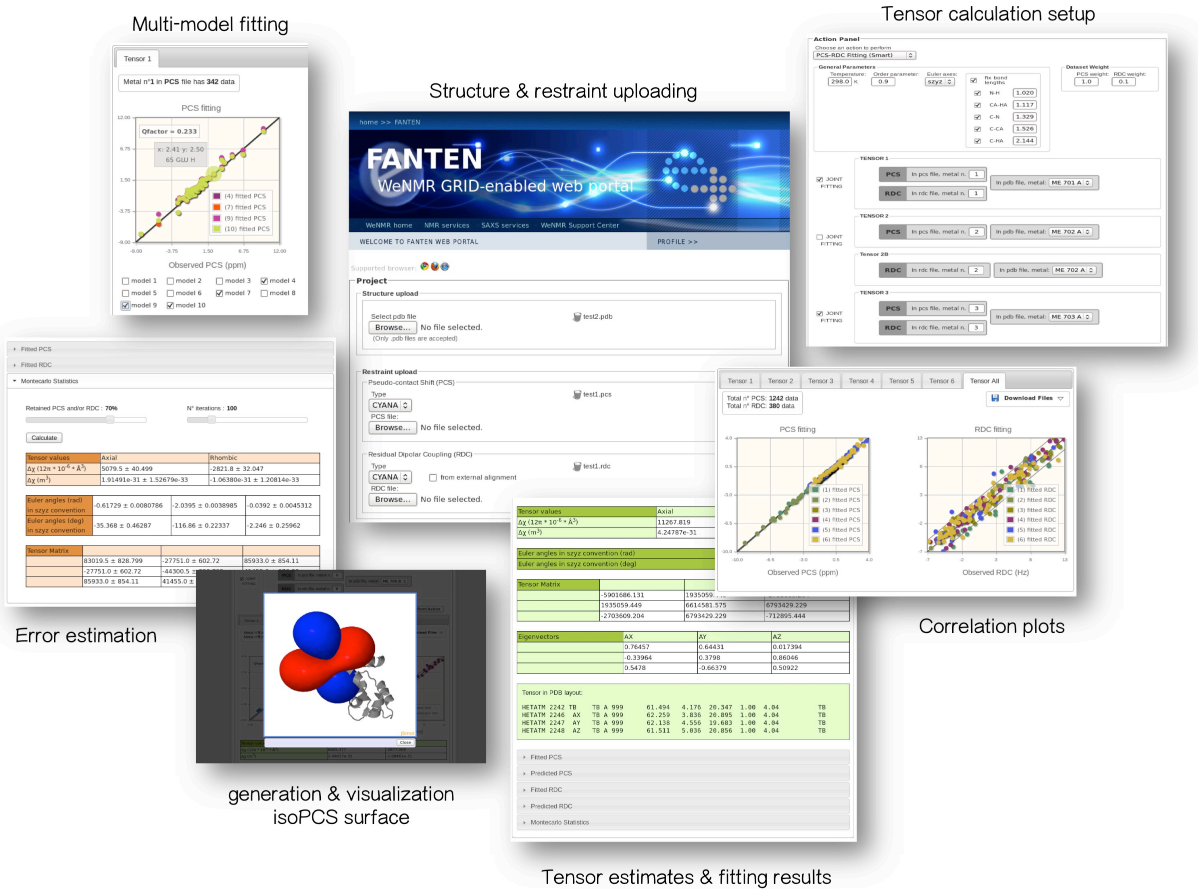


Figure 4: Overview of some of the tools and functionalities available in FANTEN web application.

1.2 Integration of paramagnetic NMR and X-ray data

NMR spectroscopy data are often used in comparison (or, ideally, in combination) with X-ray crystallographic data, whenever available. It has long been known that X-ray and NMR data provide complementary information, which can be profitably analysed together for a more accurate description of biomolecules. The complementarity of X-ray and NMR resides in the different types of information provided by these techniques, since X-ray relies mostly on the contribution given by the heavy atoms to the electron diffraction pattern, while for NMR the vast majority of restraints involve the hydrogen nuclei. Even more importantly, as anticipated above, at low and medium resolution, X-ray data contain information on the overall shape and long-range structural details, whereas short-range structural details, of the order of the interatomic distances, are accessible only at very high resolutions, which are not always achievable. In contrast, NMR data mainly provide direct information on local details, in the form of inter-atomic distances or orientations of vectors connecting chemically bound nuclei. Therefore, information from NMR and X-ray data is perfectly complementary.

During the past years, the combination of structural information derived from these two techniques opened up the way to an integrated elucidation of macromolecular structures. However, one potential pitfall is represented by the fact that, often, apparent discrepancies are observed between the models determined independently from these two methods. In general, NMR data frequently show limited compatibility with models obtained from crystallographic data but, intriguingly, the level of agreement often depending on the resolution of X-ray data.[68] Historically, such discrepancies have also occasionally led to correction of the X-ray structure on the basis of the NMR information.[69] However, such discrepancies may either reflect real differences between the true structures of the molecules in crystals and in solution, or may be due to the different but complementary information contained in the X-ray and NMR data, where the true structures may be the same but the process of deriving the models by analysing the different types of data may lead to inconsistencies. In a number of cases, refinements performed using combined X-ray and NOE-derived restraints revealed large consistency of the data, resulting in the improvement of the geometry of the model in terms of Ramachandran plot with respect to the structure calculated without NMR data.[70–74] The few violations were interpreted as real differences between the structures in crystals and in solution, mostly ascribable to a limitation of the conformational freedom for some flexible parts and to the presence of crystal packing forces in the solid state. As an example, some discrepancies were observed between a very large NMR dataset collected for ubiquitin, a protein frequently used as a model system for the development of new NMR methodologies, and its crystal structures.[75] In some cases, these discrepancies were interpreted as mobility effects and used to derive motional information on the system.[75–77] In other cases, the inconsistencies were explained by the intrinsic inaccuracy in the crystallographic models, so that the assumption of a single newly refined structure was considered sufficient to adequately explain the NMR data.[50,78] This points out some critical aspects in the interpretation of potential inconsistencies observed between solution data and crystal models, which require a careful analysis before any further interpretation.

For such purpose, a joint refinement using both X-ray and NMR data was implemented in the program REFMAC-NMR[78], part of the CCP4 suite[79], with the purpose of demonstrating that X-ray and NMR data can be profitably combined to produce models that are compatible with both experimental methods and to reliably identify the genuine differences between the solution and crystal structures. In particular, REFMAC-NMR performs a simultaneous refinement of macromolecular structures using X-ray diffractions together with structural restraints derived from PCSs and/or RDCs. Such refinement protocol has already been successfully applied to several single domain systems[78], and it has found even more interesting application in the study of multi-domain proteins and protein complexes, due to recent developments that allow for the inclusion of constraints to the tensors calculated for individual structural units.

Contribution of my thesis

In REFMAC-NMR, tensors are calculated at each refinement step and used as further restraints together with those encoded in the X-ray data and by ideal geometric values. A limiting aspect of this approach is the high number of degrees of freedom arising in the fit of the tensor, making the optimization problem under-constrained. Assumptions on the relations between the parameters describing the tensors (i.e. anisotropy values and orientation) referred to the different domains or binding partners can be profitably used and included as further restraints during the refinement. The use of constraints (or restraints, more in general) not only reduces the effective degree of freedom, but it stabilizes the refinement and ensures that the final model is consistent with prior knowledge. A-priori information can be mathematically translated into constraints that can be included in the optimization function for the estimation of the tensor. Imposing properties for the tensors estimated from individual structural units constituting multi-domain proteins and protein complexes, results of particular importance when:

a) Domain rearrangement occurs in the crystal with respect to the solution state – in this case, all the tensors calculated for the individual units must have the same anisotropy values, although their orientation can be different;

b) Data are provided by independent experimental datasets (this usually happens for RDCs collected through the use of external alignment media) – in this case, all the tensors calculated for the individual units must have the same orientation, although their anisotropy values can be different.

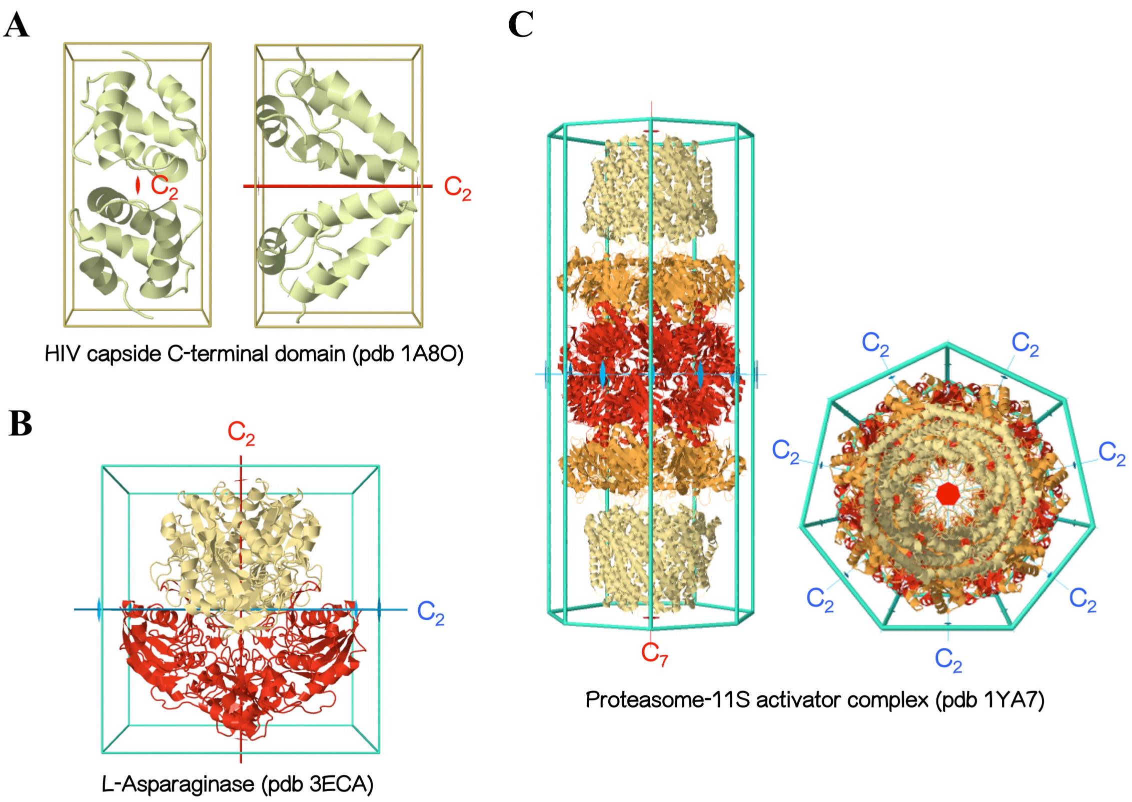
A special case of multi-domain systems arises when a number of identical structural units related by symmetry operations constitute the macromolecule. RDCs collected from external alignment media can be measured on these symmetric assemblies, resulting in a single set of data encoding both the structural information of the repeating units and of their overall organization. In this case, further properties related to the specific symmetry of the system must be taken into account during the calculation of the tensor. Estimated tensor must fulfill specific properties depending of the type of symmetry (Figure 5), which can be summarized as following:

a) C_2 symmetry – one axis of the tensor must be directed along the symmetry axis of the system (Figure 5-A);

b) D_2 symmetry - two axis of the tensor must be directed along the two symmetry axes of the system, thus defining completely its orientation (Figure 5-B);

c) C_n/D_n symmetry (with n greater of 2) – the tensor must be completely axial and one axis must be directed along the axis of cyclic symmetry (Figure 5-C).

The practical implementation of the algorithm together with the analytical derivation of the first and second derivatives used for solving the constrained optimization problems are described in detail in *"Extending the possibilities of joint X-ray/NMR structure refinement to high molecular weight complexes"* in Section 0. The latest version of REFMAC-NMR will be soon available in the CCP4 suite, with the additional possibility of performing the joint refinement with NMR data directly from the interface.



HIV capside C-terminal domain (pdb 1A8O)

L-Asparaginase (pdb 3ECA)

Proteasome-11S activator complex (pdb 1YA7)

Figure 5: Example of symmetric assemblies for which RDCs can be measured. The symmetries are classified according to the characteristic properties that the alignment tensor must assume in that condition.

1.3 Systematic characterization of multi-domain systems

Paramagnetic data are widely used in structure determination, refinement and validation. In particular, PCSs and RDCs due to their intrinsic nature of being long-range restraints can provide even more precious information in the investigation of proteins constituted by multiple domains and of protein-protein complexes. If the structural model of the molecule is already available, PCSs and RDCs can be used to determine the correct mutual orientation and positions of the domains or binding partners constituting the system.[80–83,31] PCSs and RDCs can also provide insights into the conformation variability experienced by the macromolecule in solution. RDCs, which are independent on the distance from the metal ion and which are sensitive to motions occurring on time scales faster than milliseconds, can be exploited to identify and quantify the extent of the mobility present in the system. Indeed, when dynamics affect the system, RDCs experience an averaging effect that is proportional to the extent of such motions.[62] This is not valid in principle for PCSs, which due to their dependence on the metal-nucleus distances cannot be described by the use of a averaged tensor in case of motion, unless all nuclei have a fixed distance from the metal ion during the motion[84], which is never the case in practice. Once the presence of significant mobility has been assessed for the system both PCSs and RDCs can be used to further characterize the resulting conformational variability. Although this is by definition an ill-posed problem, a multitude of methods and mathematical approaches have been developed to reconstruct the potential ensemble of conformations sampled by the system in solution.[75,85,77,76,61,86,87,88]

Contribution of my thesis

As already discussed, the local adjustments provided by NMR data reveal if discrepancies observed using X-ray -derived structural model can be reasonably explained within the uncertainty related to its atomic coordinates and, in case, to provide more reliable structural models. In the case in which significant differences are detected, a top-down approach can be used by performing independent refinement for different structural units of the original structure, or by discarding the residues responsible for the mismatch (if only few). Once adequate structural models for the single units are obtained, the comparison of tensor magnitudes help in detecting the presence of extensive mobility, whereas comparison of tensor alignments can be used to retrieve the relative arrangement of its constituting sub-units. The use of REFMAC-NMR and FANTEN can be easily combined in this frame to perform a systematic characterization of the biological system. A simple schematics of the discussed approach in reported in Figure 6.

This approach, discussed in detail in "*How to tackle protein structural data from solution and solid state: An integrated approach"* in Section 3.3, has been used for the investigation of multi-domain protein (Calmodulin-IQ peptide complex) and protein-protein complex systems (Sxl-Unr-msl2-mRNA regulatory complex).

Sxl-Unr-msl2-mRNA regulatory complex

Sxl-Unr-msl2-mRNA regulatory complex consists of two RNA recognition motifs (RRMs) of Sxl, the first of five cold shock domains of Unr (CSD1), and an 18-mer single stranded RNA derived from msl2-mRNA. Assembly of this complex is vital for female viability in fruit flies, as repression of msl2-mRNA by Sex-lethal (Sxl) and Upstream-of-N-

Ras (Unr) prevents the formation of the dosage compensation complex resulting in normal transcription of X-linked genes.[89–91] The X-ray structure of the Sxl-Unr complex was refined using RDCs induced by Pf1 phages alignment medium revealing that – despite the poor quality of the fit observed for the original structural model – the NMR data could be largely accommodated within the structural uncertainty of its primary X-ray data, with a decrease in the Q-factor of more than 50%. Moreover, joint refinement of X-ray and NMR data confirmed the overall domain arrangements and binding interfaces to be preserved in solution and highlighted local conformational differences, which provide additional information on specific features of the structure. For example, conformational dynamics and heterogeneity observed at the interface between the CSD1 and Sxl protein components in the ternary complex could be revealed only by the combination of NMR and crystallographic data.

The refinement of the Sxl-Unr-msl2-mRNA regulatory complex is described in detail in *"Improved Accuracy from Joint X-ray and NMR Refinement of a Protein–RNA Complex Structure"* in Section 3.4.

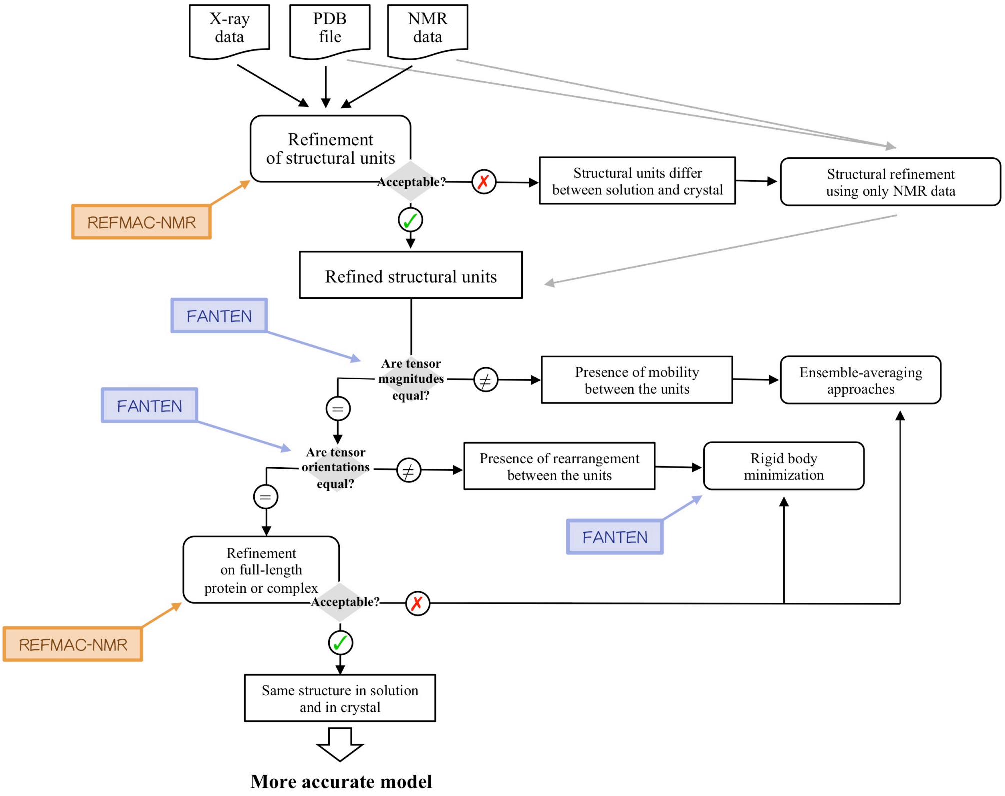


Figure 6: Simple schematics representing the approach used for the characterization of multi-domain and complex systems.

Calmodulin-IQ peptide complex

Calmodulin (CaM) comprises of two very similar domains linked together by a long central α-helix that occasionally unfolds, making the system very flexible. Peculiarity of CaM resides in its inter-domain mobility that allows the protein to sample a wide range of conformations, and that is highly related to its biological function of intermediate messenger.[92] The complex of CaM with the IQ-recognition motif was refined using H^N, N PCSs and H^N-N RDCs measured for six different lanthanides (Tb^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, Tm^{3+}, and Yb^{3+}).[93] A stepwise approach has been used to investigate the origin of the inconsistencies found for the NMR data with the different CaM conformations present in the asymmetric unit of the crystal. REFMAC-NMR was applied to perform the refinement of both the single domains and full-length protein, achieving very good agreement with all the available NMR data without increase in the crystallographic Rfree-factor or significant deviations of the structure from the ideal geometrical values. Beside the detection of the most compliant conformation, joint refinement by X-ray and NMR data showed that all the available data agree with the assumption of a single static conformation. This conclusion questions the application of an ensemble-based approach[93], previously used to achieve a satisfactory agreement of the data. This confirms that resolution-dependent inaccuracies present in X-ray models may lead to unreliable interpretation of the data and thus need to be always taken into account for obtaining significant conclusions.

The refinement of Calmodulin-IQ peptide complex is described in detail in *"Impact of experimental uncertainties in protein mobility studies"* in see Section 3.5.

2. Methodological Aspects

2.1 Paramagnetic and diamagnetic NMR-based restraints

Pseudo-contact shifts (PCSs) and residual dipolar couplings (RDCs) are long-range structural restraints that for their nature have been shown to be valuable restraints to help solving protein structures in the solution state. While PCSs give information about the relative positions of the paramagnetic metal and the interacting atoms, RDCs tell us how the vectors joining pairs of coupled nuclei are oriented in the space.

Both PCSs and self-orientation RDCs arise in the presence of paramagnetic metal ions coordinated to the protein[15] or purposely attached to it with a paramagnetic tag[24–38]. In the absence of a paramagnetic metal, diamagnetic RDCs can be induced by other sources of molecular magnetic anisotropy, i.e. by adding external orienting devices in solution. [40,42–50]

2.1.1 Magnetic susceptibility and its anisotropy

Magnetic susceptibility[51] is an inherent property of a molecule that represents how the molecule interacts with an external magnetic field and it is generally expressed by a scalar quantity χ. If the response is orientation-dependent, the magnetic susceptibility is anisotropic and is mathematically described by a symmetric second rank tensor in the following way:

$$\boldsymbol{\chi} = \begin{pmatrix} \chi_{xx} & \chi_{xy} & \chi_{xz} \\ \chi_{xy} & \chi_{yy} & \chi_{yz} \\ \chi_{xz} & \chi_{yz} & \chi_{zz} \end{pmatrix}$$

where (x, y, z) are the principal axes in a molecule-fixed coordinate system. The axial and the rhombic components of the magnetic anisotropy can be easily defined, respectively, as:

$$\Delta\chi_{\mathrm{ax}} = \tilde{\chi}_{\mathrm{zz}} - \frac{\tilde{\chi}_{\mathrm{xx}} + \tilde{\chi}_{\mathrm{yy}}}{2}$$

$$\Delta\chi_{\mathrm{rh}} = \tilde{\chi}_{\mathrm{xx}} - \tilde{\chi}_{\mathrm{yy}}$$

where $\tilde{\chi}_{xx}$, $\tilde{\chi}_{yy}$, $\tilde{\chi}_{zz}$ are the components of the magnetic susceptibility tensor in the frame in which it is diagonal.

2.1.2 *Pseudo-contact shifts (PCSs)*

The pseudo-contact shift is the rotational average of the electron-nuclear dipolar interaction when paramagnetic center has an anisotropic magnetic susceptibility and, as such, it can be regarded as an additional chemical shift term, which reflects an average “distortive” effect of the electron on the magnetic field.

The equation for PCS can be written in many ways, depending on the reference frame, on the coordinates used to describe the positions of atoms (Cartesian, polar or direction cosines) and on different combinations of the magnetic susceptibility parameters. The complete list of these expressions is available in Bertini et al.[15] In the frame given by the principal axes of the magnetic susceptibility tensor centered on the metal ion, the value of PCS is given by:

$$PCS = \frac{1}{12\pi r^3}\left[\Delta\chi_{ax}(3\cos^2\Theta - 1) + \frac{3}{2}\Delta\chi_{rh}(\sin^2\Theta\cos 2\Omega)\right] \tag{1}$$

where $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic components of the magnetic susceptibility anisotropy tensor $\boldsymbol{\chi}$, r is the metal-nucleus distance and Θ and Ω are the spherical angles describing the orientation of the metal-nucleus vector with respect to the principal axes of the χ tensor.

In Cartesian coordinates, the most suitable expression for computational purposes is:

$$PCS = \frac{1}{4\pi r^3}\left[(\chi_{zz} - \bar{\chi})\frac{2z^2 - x^2 - y^2}{2r^2} + (\chi_{xx} - \chi_{yy})\frac{x^2 - y^2}{2r^2} + \chi_{xy}\frac{2xy}{r^2} + \chi_{xz}\frac{2xz}{r^2} + \chi_{yz}\frac{2yz}{r^2}\right] \tag{2}$$

where $\bar{\chi}$ is the trace of χ divided by three. Of note, the equation is linear in the tensor variables.

2.1.3 Paramagnetic residual dipolar couplings (paramagnetic RDCs)

The dipolar coupling is a through-space interaction that arises between any two magnetically active nuclei. As a result of the effects of isotropic tumbling due to Brownian motion, dipolar couplings average to zero, but are different from zero when a preferential orientation is induced to the molecule, producing a modulation in the 1-bond splitting of coupled nuclei. This effect occurs, for instance, when a paramagnetic ion is introduced in the molecule, generating paramagnetic (self-orientation) RDCs. The RDC value on two coupled nuclei A and B thus depends on the paramagnetic susceptibility anisotropy tensor χ, and on the orientation of the vector connecting A and B with respect to the tensor.

Similarly to PCSs, the equation to determine RDCs can have many expressions, depending on the reference frame, on the coordinates used to describe the positions of atoms (Cartesian, polar or direction cosines) and on different combinations of the magnetic susceptibility components[15], where the most commonly used are:

$$RDC = k\left[\Delta\chi_{ax}(3\cos^2\Theta - 1) + \frac{3}{2}\Delta\chi_{rh}(\sin^2\Theta\cos 2\Omega)\right] \tag{3}$$

$$RDC = 3k\left[(\chi_{zz} - \bar{\chi})\frac{2z_{AB}^2 - x_{AB}^2 - y_{AB}^2}{2r_{AB}^2} + (\chi_{xx} - \chi_{yy})\frac{x_{AB}^2 - y_{AB}^2}{2r_{AB}^2} \right. \\ \left. + \chi_{xy}\frac{2x_{AB}y_{AB}}{r_{AB}^2} + \chi_{xz}\frac{2x_{AB}z_{AB}}{r_{AB}^2} + \chi_{yz}\frac{2y_{AB}z_{AB}}{r_{AB}^2}\right] \qquad (4)$$

with:

$$k = -\frac{S_{LS}}{4\pi}\frac{B_0^2}{15kT}\frac{\gamma_A\gamma_B\hbar}{2\pi r_{AB}^3}$$

$$x_{AB}^2 = (x_A - x_B)^2,\ y_{AB}^2 = (y_A - y_B)^2,\ z_{AB}^2 = (z_A - z_B)^2,\ \ r_{AB}^2 = x_{AB}^2 + y_{AB}^2 + z_{AB}^2$$

where r_{AB} is the distance between the two coupled nuclei A and B, S_{LS} is the Lipari-Szabo model-free order parameter, introduced to take into account some average local mobility of the nitrogen-amide proton vectors, γ_A and γ_B are the gyromagnetic ratios of the two nuclei and other symbols have the usual meaning. RDCs do not depend on the position of the coupled nuclei with respect to the metal ion. Of note, except for the constants reported in front of the expressions, equations for PCSs (1-2) and for RDCs (3-4) are very similar and if measured in the same structure and in presence of the same paramagnetic ions, they depends exactly from the same parameters describing the tensor anisotropy and orientation.

Once a structural model of the molecule is available and S_{LS} is estimated, RDCs depend only on the five parameters defining the paramagnetic susceptibility anisotropy tensor. If these five parameters are determined from the analysis of the PCSs, RDCs are directly dependent on the orientation of the vectors connecting the coupled nuclei in a common frame. Degeneracy in the solutions can be removed by measuring several sets of RDC data arising from different paramagnetic metal ions, with different principal frames of the susceptibility anisotropy tensor.

2.1.4 *Diamagnetic residual dipolar couplings (diamagnetic RDCs)*

For diamagnetic molecules the spatial anisotropy of the dipolar coupling can originate from external orienting media.

Diamagnetic RDCs are described by an equation of the same form of self-orientation RDCs:

$$RDC = -\frac{3\mu_0 S_{LS}}{4\pi}\frac{\gamma_A\gamma_B\hbar}{r_{AB}^3}\left[D_{zz}\frac{2z_{AB}^2 - x_{AB}^2 - y_{AB}^2}{2r_{AB}^2} + \left(D_{xx} - D_{yy}\right)\frac{x_{AB}^2 - y_{AB}^2}{2r_{AB}^2} + D_{xy}\frac{2x_{AB}y_{AB}}{r_{AB}^2} + D_{xz}\frac{2x_{AB}z_{AB}}{r_{AB}^2} + D_{yz}\frac{2y_{AB}z_{AB}}{r_{AB}^2}\right] \tag{5}$$

where D_i are the components of the molecular alignment tensor.

2.1.5 *Q-factor*

The agreement between experimental and back-calculated PCSs and RDCs against a given structural model is usually monitored by the Q-factors[94], defined as:

$$Q_{PCS} = \sqrt{\frac{\sum_i\left(PCS_i^{exp} - PCS_i^{calc}\right)^2}{\sum_i\left(PCS_i^{exp}\right)^2}}, \qquad Q_{RDC} = \sqrt{\frac{\sum_i\left(RDC_i^{exp} - RDC_i^{calc}\right)^2}{\sum_i\left(RDC_i^{exp}\right)^2}} \ .$$

Q-factors are independent from the weights used, so they provide an essential tool to evaluate the agreement of the NMR data with the calculated structures.

Sometimes, it is useful to determine an overall agreement of all the available restraints. This can be achieved by the use of a global Q-factor[93] that results from the combination of the PCS and RDC Q-factors, expressed as following:

$$Q_{PCS+RDC} = \sqrt{\frac{\left(N_{PCS}Q_{PCS}^2 + N_{RDC}Q_{RDC}^2\right)}{\left(N_{PCS} + N_{RDC}\right)}} \ .$$

2.2 Software tools for NMR data

2.2.1 *FANTEN web-application*

FANTEN is a web-application for the simultaneous fit of PCSs and RDCs to protein structures, to determine the anisotropy or alignment tensors (see Section 3.1). The underlying library is written in Python in an object-oriented fashion, which permits to organize the program in a clear and modular structure by definition of abstract datatypes (i.e. ParamagneticTensor, PCS, RDC etc.) and which makes the existing code easy to maintain and to modify. The web development exploits the Pylons framework and HTML, CSS, jQuery and JavaScript languages. For the initial reading of pdb files, the program makes use of a modified version of the p3d package.[95] For the fitting of PCSs and RDCs equations (1), (2) and (3), (4), have been used.

2.2.2 *Constrained optimization in REFMAC-NMR*

In the new version of REFMAC-NMR, the inclusion of constraints is fulfilled by the direct adaptation of the algorithm used for solving the unconstrained problem to the constrained case by re-parameterization (see Section 0). The generated constraints were mathematically translated in terms of additional contributions to the first and second derivatives of the target function with respect to the tensor parameters for the re-parameterized problem. Finally, first and second derivatives were used to iteratively solve the problem using a Gauss-Newton optimization approach. The main program and all the libraries have been entirely written in Fortran90.

3. Results

3.1 FANTEN: a new web-based interface for the analysis of magnetic anisotropy-induced NMR data

Mauro Rinaldelli[a], *Azzurra Carlon*[a], *Enrico Ravera*[a,b], *Giacomo Parigi*[a,b], *Claudio Luchinat*[a,b]

[a] Magnetic Resonance Center (CERM), University of Florence,
via Sacconi 6, 50019 Sesto Fiorentino (FI), Italy
[b] Department of Chemistry, University of Florence,
via della Lastruccia 3, 50019 Sesto Fiorentino (FI), Italy

Published in:
Journal of biomolecular NMR ***61*** *(1), 21-34*

J Biomol NMR (2015) 61:21–34
DOI 10.1007/s10858-014-9877-4

ARTICLE

FANTEN: a new web-based interface for the analysis of magnetic anisotropy-induced NMR data

Mauro Rinaldelli · Azzurra Carlon · Enrico Ravera · Giacomo Parigi · Claudio Luchinat

Received: 9 October 2014 / Accepted: 15 November 2014 / Published online: 22 November 2014

Abstract Pseudocontact shifts (PCSs) and residual dipolar couplings (RDCs) arising from the presence of paramagnetic metal ions in proteins as well as RDCs due to partial orientation induced by external orienting media are nowadays routinely measured as a part of the NMR characterization of biologically relevant systems. PCSs and RDCs are becoming more and more popular as restraints (1) to determine and/or refine protein structures in solution, (2) to monitor the extent of conformational heterogeneity in systems composed of rigid domains which can reorient with respect to one another, and (3) to obtain structural information in protein–protein complexes. The use of both PCSs and RDCs proceeds through the determination of the anisotropy tensors which are at the origin of these NMR observables. A new user-friendly web tool, called FANTEN (Finding ANisotropy TENsors), has been developed for the determination of the anisotropy tensors related to PCSs and RDCs and has been made freely available through the WeNMR (http://fanten-enmr.cerm.unifi.it:8080) gateway. The program has many new features not available in other existing programs, among which the possibility of a joint analysis of several sets of PCS and RDC data and the possibility to perform rigid body minimizations.

Keywords Paramagnetic NMR · Pseudocontact shift · Residual dipolar coupling · Magnetic susceptibility tensor · Web-based interface

Introduction

Pseudocontact shifts (PCSs) and residual dipolar couplings (RDCs) from self alignment arise in paramagnetic molecules in case the magnetic susceptibility tensor associated with the paramagnetic center is anisotropic. RDCs can also be induced in the absence of paramagnetic centers by the presence of an anisotropic alignment medium. PCSs and RDCs contain structural information that proved very helpful for solving protein structures (Gochin and Roder 1995; Banci et al. 1996, 1998; Bertini et al. 2001; Gaponenko et al. 2004; Diaz-Moreno et al. 2005; Jensen et al. 2006; Schmitz et al. 2012; Yagi et al. 2013b). They have been thus included as structural restraints (Balayssac et al. 2006; Banci et al. 2004) in the most commonly used programs for protein structure determination from NMR data [CYANA (Güntert 2004) and Xplor-NIH (Schwieters et al. 2003)] and more recently for structural refinement in combination with X-ray data (Rinaldelli et al. 2014) using the program REFMAC5 (Murshudov et al. 1997). PCS and RDC values depend on the coordinates of the nuclei and on the orientation of internuclear vectors between coupled nuclei, respectively, with respect to the unique magnetic susceptibility anisotropy tensor associated to the paramagnetic center present in the molecule (Bertini et al. 2002, 2011b).

In order to translate PCS and RDC data into structural information, the magnetic susceptibility anisotropy tensor

---

Mauro Rinaldelli and Azzurra Carlon have contributed equally to this article.

Electronic supplementary material The online version of this article (doi:10.1007/s10858-014-9877-4) contains supplementary material, which is available to authorized users.

M. Rinaldelli · A. Carlon · E. Ravera · G. Parigi (✉) ·
C. Luchinat (✉)
CERM and Department of Chemistry "Ugo Schiff", University of Florence, via Sacconi 6, Sesto Fiorentino, Florence, Italy
e-mail: parigi@cerm.unifi.it

C. Luchinat
e-mail: luchinat@cerm.unifi.it

should thus be determined. Conversely, once the protein structure is known, PCSs and RDCs can be used to calculate the tensor. Such a calculation can be performed by several available programs. In our lab, we have developed a series of programs (FANTASIA (Finding ANisotropy Tensors: A Simplex Approach), FANTALIN, FANTAORIENT) (Banci et al. 1996, 1998). Huber and co-workers have developed NUMBAT (Schmitz et al. 2008) to obtain the magnetic susceptibility anisotropy tensor from PCSs and protein structures. A python-based library, PyParaTool, has been recently developed, which encompasses the use of PCS and RDC (http://comp-bio.anu.edu.au/mscook/PPT/). RDCs may also arise in the absence of paramagnetic ions by molecular partial orientation achieved in the presence of orienting devices in solution (Tjandra and Bax 1997; Chou et al. 2001; Tolman et al. 2001; Prestegard et al. 2004; Chill et al. 2007; Lange et al. 2008; Grishaev et al. 2008), or by molecular anisotropy (Zhang et al. 2007), with the difference that they are not coupled to PCSs, as the latter are not present in diamagnetic systems. In this case the relevant tensor is the alignment tensor of the molecule, which depends both on the biomolecule and on the alignment medium. Command-line software packages (PALES and PATI) are available to fit RDCs arising from external alignment media and to give an estimate of the alignment tensor due to either steric effects or electrostatic effects or both (Zweckstetter and Bax 2000; Zweckstetter 2008; Berlin et al. 2009, 2010, 2011).

In practice, in structure calculations, PCSs and RDCs are used (1) to calculate the magnetic susceptibility anisotropy tensors in best agreement with a tentative protein structure, using either programs of the FANTASIA family or NUMBAT, and (2) to calculate the protein structures with minimal restraints violations in agreement with the tentative magnetic susceptibility anisotropy tensors, using appropriate versions of CYANA (Balayssac et al. 2006) or Xplor-NIH (Banci et al. 2004). These calculations are repeated iteratively, until convergence of the tensors is achieved, usually in few steps if the restraints are enough to reconstruct the protein fold. PCS-Rosetta have been also employed to obtain simultaneously the protein structure and an accurate fit of the tensor, without the use of an external software for tensor determination (Schmitz et al. 2012; Yagi et al. 2013b). PCSs and RDCs also proved useful for the refinement of protein structures starting from available models (Gochin and Roder 1995; Bertini et al. 2004, 2009, 2012a). In these cases, the models could be used to determine the first estimate of the magnetic susceptibility anisotropy tensors. The same procedure can of course be followed if RDCs from external alignment are used: the tensor can be fit (by PALES or PATI), and the iterations needed for structural determination can be as well performed with Xplor-NIH or CYANA. Also in this case an existing model (Chou et al. 2001) can be used to determine an estimate of the alignment tensor, of course to some degree of approximation.

It was also shown that RDCs and PCS can be used to screen the PDB for proteins which have a partial homology, based on the 3D arrangement of secondary structure elements as given by PCSs and/or RDCs (Meiler et al. 2000). Other most used programs available for the analysis of PCS and RDC data are reported in Table 1.

PCSs and/or RDCs are even more precious structural restraints in the case of proteins composed of multiple rigid domains, the structure of which is known but not the relative orientation, or of protein–protein complexes, when the structure of each protein is known but again their relative orientation is not known. In these cases, a rigid body minimization can locate the position of the different domains or molecules with respect to one another in order to reproduce the experimental PCS and RDC data (Al-

Table 1 Some of the most used programs available for the analysis of PCS and RDC data

$\Delta\chi$-tensor determination	
FANTEN	
FANTASIA	Banci et al. (1996)
FANTAORIENT	Banci et al. (1996)
NUMBAT	Schmitz et al. (2008)
RDCs analysis	
iDC	Wei and Werner (2006)
MSPIN	Navarro-Vazquez (2012)
PALES	Zweckstetter (2008)
PATI	Berlin et al. (2009)
REDCAT	Valafar and Prestegard (2004)
Protein assignment	
ECHIDNA	Schmitz et al. (2006)
PARASSIGN	Skinner et al. (2013)
PLATYPUS	Pintacuda et al. (2004)
POSSUM	John et al. (2007)
Protein structure determination	
DIPOCOUP	Meiler et al. (2000)
MODULE	Dosset et al. (2001)
PARAMAGNETIC CYANA	Banci et al. (1996); Balayssac et al. (2006)
PATIDOCK	Berlin et al. (2011)
PCS-HADDOCK	Schmitz and Bonvin (2011)
PCS-ROSETTA	Schmitz et al. (2012)
REFMAC5	Rinaldelli et al. (2014)
Xplor-NIH	Schwieters et al. (2003); Banci et al. (2004)
Conformational averaging	
FLEXIBLE MECCANO	Ozenne et al. (2012)
MaxOcc	Bertini et al. (2012b)
REDCRAFT	Bryson et al. (2008)

Hashimi et al. 2000; Clore 2000; Dosset et al. 2001; Valafar and Prestegard 2004; Pintacuda et al. 2006; Hulsker et al. 2008; Simon et al. 2010; Bertini et al. 2011b). A rigid body molecular docking program, PATIDOCK (Berlin et al. 2011), was developed to recover the structure of a molecular complex from the three-dimensional structure of the individual components and the experimental RDCs arising from the steric alignment, due to the anisotropy of the molecular shape, in the presence of external orienting devices.

Due to the quadratic mathematical form of these restraints, more than one set of PCSs and RDCs are needed to remove degeneracies in the solutions (Dosset et al. 2001; Longinetti et al. 2002; Bertini et al. 2002; Fragai et al. 2013). These multiple sets can be obtained by alternative replacement of the paramagnetic ion with other paramagnetic ions with different magnetic susceptibility anisotropy tensors or by attaching, once at a time, paramagnetic tags in different points (Bertini et al. 2001; Pintacuda et al. 2006; Otting 2010; Liu et al. 2012). Finally, PCSs and RDCs can provide information on interdomain mobility in multidomain proteins or internal mobility in protein–protein complexes, whenever internal consistency of all PCSs and RDCs arising from each paramagnetic metal with its own magnetic susceptibility anisotropy tensor cannot be achieved by assuming a single rigid structure (Gardner et al. 2005; Longinetti et al. 2006; Bertini et al. 2007, 2010, 2012b; Russo et al. 2013; Ravera et al. 2014; Andralojc et al. 2014).

The advantages offered by paramagnetic restraints are more and more exploited thanks to the development of paramagnetic tags that can be attached to diamagnetic proteins, so that PCSs and RDCs can be easily measured also for molecules without a natural metal binding site (Wöhnert et al. 2003; Rodriguez-Castañeda et al. 2006; Su et al. 2006; John and Otting 2007; Su et al. 2008a; Keizers et al. 2008; Su et al. 2008b; Zhuang et al. 2008; Häussinger et al. 2009; Su and Otting 2010; Hass et al. 2010; Man et al. 2010; Das Gupta et al. 2011; Liu et al. 2012; Cerofolini et al. 2013; Yagi et al. 2013a; Gempf et al. 2013; Kobashigawa et al. 2012; Saio et al. 2011; Watanabe et al. 2010; Loh et al. 2013; Swarbrick et al. 2011a, 2011b).

Due to the increasingly wide use of paramagnetic restraints and diamagnetic RDCs for structural purposes we found it useful to make available a new web-based interface, called FANTEN (Finding ANisotropy TENsor) for the simultaneous fit of PCSs and RDCs to protein structures in order to determine the anisotropy tensors. FANTEN accepts as experimental data both PCSs and RDCs, and can perform a minimization using several sets of data either independently or globally; it can also be used to perform a rigid body minimization, when data are available for different rigid domains. This user-friendly web interface is based on a framework (Bertini et al. 2011a) developed within the WeNMR project (Wassenaar et al. 2012), and can be accessed from the open access web site (http://fanten-enmr.cerm.unifi.it:8080). Data can be provided either in the Xplor-NIH format (also available through the WeNMR web pages) or in the PARAMAGNETIC CYANA format (Balayssac et al. 2006), and thus the program can be easily used in combination with these programs.

Methods

PCSs, arising in the presence of a paramagnetic metal, depend on the magnetic susceptibility anisotropy tensor and on the nuclear coordinates according to the following equation, written in the so-called linear form (McConnell and Robertson 1958; Kurland and McGarvey 1970; Kemple et al. 1988; Bertini et al. 2002):

$$PCS = \frac{1}{4\pi r^3}\left[\chi_{zz}\frac{2z^2 - x^2 - y^2}{2r^2} + \left(\chi_{xx} - \chi_{yy}\right)\frac{x^2 - y^2}{2r^2} + \chi_{xy}\frac{2xy}{r^2} + \chi_{xz}\frac{2xz}{r^2} + \chi_{yz}\frac{2yz}{r^2}\right] \tag{1}$$

where x, y, z are the coordinates of the nucleus when the metal ion sits at the origin of the coordinate system, r is the distance between the observed nucleus and the metal ion and χ_{ij} are the components of the symmetric susceptibility anisotropy tensor. Paramagnetic RDCs arising in the presence of the same metal depend on the same χ tensor components and on the orientation of the dipole–dipole coupled nuclei according the following equation (Tolman et al. 1995; Bertini et al. 2002):

$$RDC = 3k\left[\chi_{zz}\frac{2z_{AB}^2 - x_{AB}^2 - y_{AB}^2}{2r_{AB}^2} + \left(\chi_{xx} - \chi_{yy}\right)\frac{x_{AB}^2 - y_{AB}^2}{2r_{AB}^2} + \chi_{xy}\frac{2x_{AB}y_{AB}}{r_{AB}^2} + \chi_{xz}\frac{2x_{AB}z_{AB}}{r_{AB}^2} + \chi_{yz}\frac{2y_{AB}z_{AB}}{r_{AB}^2}\right] \tag{2}$$

where

$$w_{AB}^2 = (w_A - w_B)^2 \quad (w = x, y, z)$$

$$k = -\frac{S_{LS}}{4\pi}\frac{B_0^2}{15kT}\frac{\gamma_A\gamma_B\hbar}{2\pi r_{AB}^3}$$

r_{AB} is the distance between the two coupled nuclei A and B, and S_{LS} is the model-free order parameter, introduced to take into account some average local mobility of the coupled nuclei vectors. Other symbols have the usual meaning. RDCs do not depend on the position of the coupled nuclei with respect to the metal ion.

Diamagnetic RDCs are described by an equation with the same form as that for paramagnetic RDCs (Eq. 2):

$$RDC = -\frac{3\mu_0 S_{LS}}{8\pi^2}\frac{\gamma_A\gamma_B\hbar}{r_{AB}^3}\left[A_{zz}\frac{2z_{AB}^2 - x_{AB}^2 - y_{AB}^2}{2r_{AB}^2} + \left(A_{xx} - A_{yy}\right)\frac{x_{AB}^2 - y_{AB}^2}{2r_{AB}^2} + A_{xy}\frac{2x_{AB}y_{AB}}{r_{AB}^2} + A_{xz}\frac{2x_{AB}z_{AB}}{r_{AB}^2} + A_{yz}\frac{2y_{AB}z_{AB}}{r_{AB}^2}\right] \tag{3}$$

where A_{ij} are the components of the molecular alignment tensor.

In Eqs. 1–3, PCSs and RDCs are reported as linear functions of the five linearly independent components of the traceless χ or A tensors. However, it is common to describe these tensors by reporting two anisotropy values and the three Euler angles describing the rotations needed to pass from the reference frame in which the nuclear coordinates are provided to the frame in which the χ or A tensors are diagonal. These values can be determined after diagonalization of the χ or A tensors. The magnetic susceptibility anisotropy values are defined by the axial and rhombic components provided by

$$\Delta\chi_{ax} = \chi_{zz} - \frac{\chi_{xx} + \chi_{yy}}{2} = \frac{3}{2}\chi_{zz}$$

and

$$\Delta\chi_{rh} = \chi_{xx} - \chi_{yy},$$

where χ_{ii} are the components of the χ tensor in the frame where it is diagonal. Analogously, the anisotropies of the alignment tensor are described by the fraction of alignment along the z axis (A) and by the rhombicity (R), (Zweckstetter 2008)

$$A = \frac{3}{2}A_{zz}$$

$$R = \frac{A_{xx} - A_{yy}}{A},$$

where A_{ii} are the components of the A tensor in the frame where it is diagonal. Alternatively, the anisotropy values can be described by the maximum RDC induced for a specific nuclear pair, i.e. the N–^{N}H pair (D_{NH}) (Tjandra and Bax 1997),

$$D_{AB} = -\frac{S_{LS}\mu_0\gamma_A\gamma_B h}{16\pi^3 r_{AB}^3}A$$

and by the rhombicity R.

Due to the dependence of Eqs. (1) and (2) on the nuclear coordinates, the axes of the χ tensor are not uniquely determined. Using a standard convention, they are defined in such a way that $|\chi_{yy}| \leq |\chi_{xx}| \leq |\chi_{zz}|$ and $|\Delta\chi_{rh}| \leq 2/3\ |\Delta\chi_{ax}|$ (Banci et al. 1996).

The agreement of calculated and experimental PCSs/RDCs is described by the Q-factors (Cornilescu et al. 1998), defined as

$$Q_{PCS} = \sqrt{\frac{\sum_i \left(PCS_i^{\mathrm{exp}} - PCS_i^{\mathrm{calc}}\right)^2}{\sum_i \left(PCS_i^{\mathrm{exp}}\right)^2}},\ Q_{RDC} = \sqrt{\frac{\sum_i \left(RDC_i^{\mathrm{exp}} - RDC_i^{\mathrm{calc}}\right)^2}{\sum_i \left(RDC_i^{\mathrm{exp}}\right)^2}} \tag{4}$$

The estimation of the error of the values determined through the best fit procedures is implemented using the bootstrap Monte Carlo method, which consists in calculating the standard deviation of the values obtained after multiple removals of a user-defined number of randomly selected data.

The web interface is based on Pylons (www.pylonsproject.org) web application framework (Bertini et al. 2011a), and library functions has been implemented using Python 2.6 (www.python.org). The following Python packages were used: scipy, numpy (Oliphant 2007), lmfit (http://lmfit.github.io/lmfit-py/), and P3D (Fufezan and Specht 2009).

Results

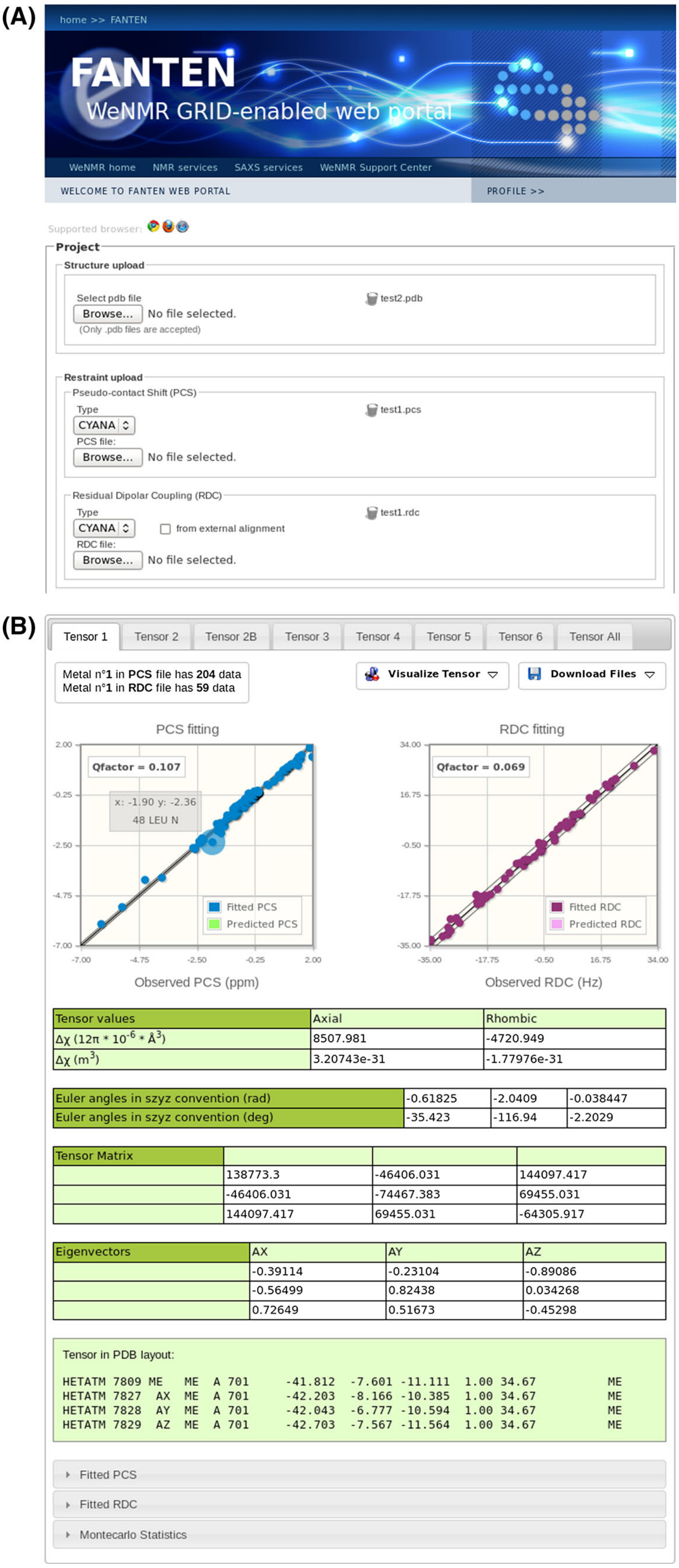
The program FANTEN performs a linear fit of the five independent components of the traceless χ tensor (χ_{xx}, χ_{zz}, χ_{xz}, χ_{yz}, χ_{xy}; with $\chi_{yy} = -\chi_{zz} - \chi_{xx}$) in order to obtain the best agreement of experimental PCSs and/or paramagnetic RDCs against the protein structure, when the coordinates of the metal are included in the structure (if not, the metal coordinates are three additional parameters to be determined in the minimization procedure, see later); it determines the five components of the A tensor in best agreement against the protein structure when diamagnetic RDCs are provided. Users can provide one protein structure or an ensemble thereof (in PDB format) and multiple sets of PCSs and/or paramagnetic RDCs or diamagnetic RDCs (both in PARAMAGNETIC CYANA format or in the format used in the paramagnetic routines for Xplor-NIH) referred to the same or different metals or orienting devices. The interface for data upload is shown in Fig. 1a.

The calculations can be performed using three different interfaces (Custom Interface, Smart Interface and Rigid Body Minimization), designed and optimized for specific applications with different levels of complexity.

Custom Interface

FANTEN allows for a fully customized calculation of the tensor: from the user-provided PDB file it is possible to

Fig. 1 **a** Interface for the upload of the PDB file and of the NMR restraints. **b** Overview of the results of the fit performed using six PCS and RDC datasets



Metal n°1 in **PCS** file has **204** data
Metal n°1 in **RDC** file has **59** data

Tensor values	Axial	Rhombic
$\Delta\chi$ ($12\pi * 10^{-6} * Å^3$)	8507.981	-4720.949
$\Delta\chi$ (m^3)	3.20743e-31	-1.77976e-31

Euler angles in szyz convention (rad)	-0.61825	-2.0409	-0.038447
Euler angles in szyz convention (deg)	-35.423	-116.94	-2.2029

Tensor Matrix			
	138773.3	-46406.031	144097.417
	-46406.031	-74467.383	69455.031
	144097.417	69455.031	-64305.917

Eigenvectors	AX	AY	AZ
	-0.39114	-0.23104	-0.89086
	-0.56499	0.82438	0.034268
	0.72649	0.51673	-0.45298

Tensor in PDB layout:

```
HETATM 7809 ME   ME  A 701     -41.812  -7.601 -11.111  1.00 34.67          ME
HETATM 7827  AX  ME  A 701     -42.203  -8.166 -10.385  1.00 34.67          ME
HETATM 7828  AY  ME  A 701     -42.043  -6.777 -10.594  1.00 34.67          ME
HETATM 7829  AZ  ME  A 701     -42.703  -7.567 -11.564  1.00 34.67          ME
```

select the model and the chain for which the calculation should be performed, and the associated PCS and/or RDC datasets (Fig. 2; Fig. S1). For any dataset, the user can define, if known, (1) metal position, (2) tensor orientation, and (3) anisotropy parameters, or any combination thereof.

1. In case the coordinates of the paramagnetic metal are provided in the PDB file, the metal can be simply selected from the list of the metals present in the PDB, with the possibility of keeping its coordinates fixed (by selecting "Fix position to this metal") or refining them (by selecting "Fit position around this metal") during the calculation. Information on the position of the paramagnetic metal can also be provided as distance restraints from any atom present in the PDB, so that during the fit the metal position can be restrained in a user-specified volume of space. These restraints may be available from knowledge of the metal binding site or protein tagging position and may be useful to increase the speed of the calculations. In case the paramagnetic metal position is completely unknown, a grid search can be performed starting from points in the proximity of the protein surface or from atoms constituting its backbone. The minimization is performed using eight different starting points for the metal position, corresponding to the vertices of a box containing the protein or randomly choosing eight atoms belonging to the protein backbone, respectively.
2. The tensor orientation can be fixed in three different ways: by providing the Euler angles (in the preferred convention), the eigenvectors of the tensor expressed in its non-diagonal matrix form, or by selecting the tensor origin and axes orientations as defined by pseudoatoms included in the uploaded PDB. The Euler angles convention is defined by a 4-character string, in which the first character defines if rotations are applied to static (s) or rotating (r) frame, whereas the

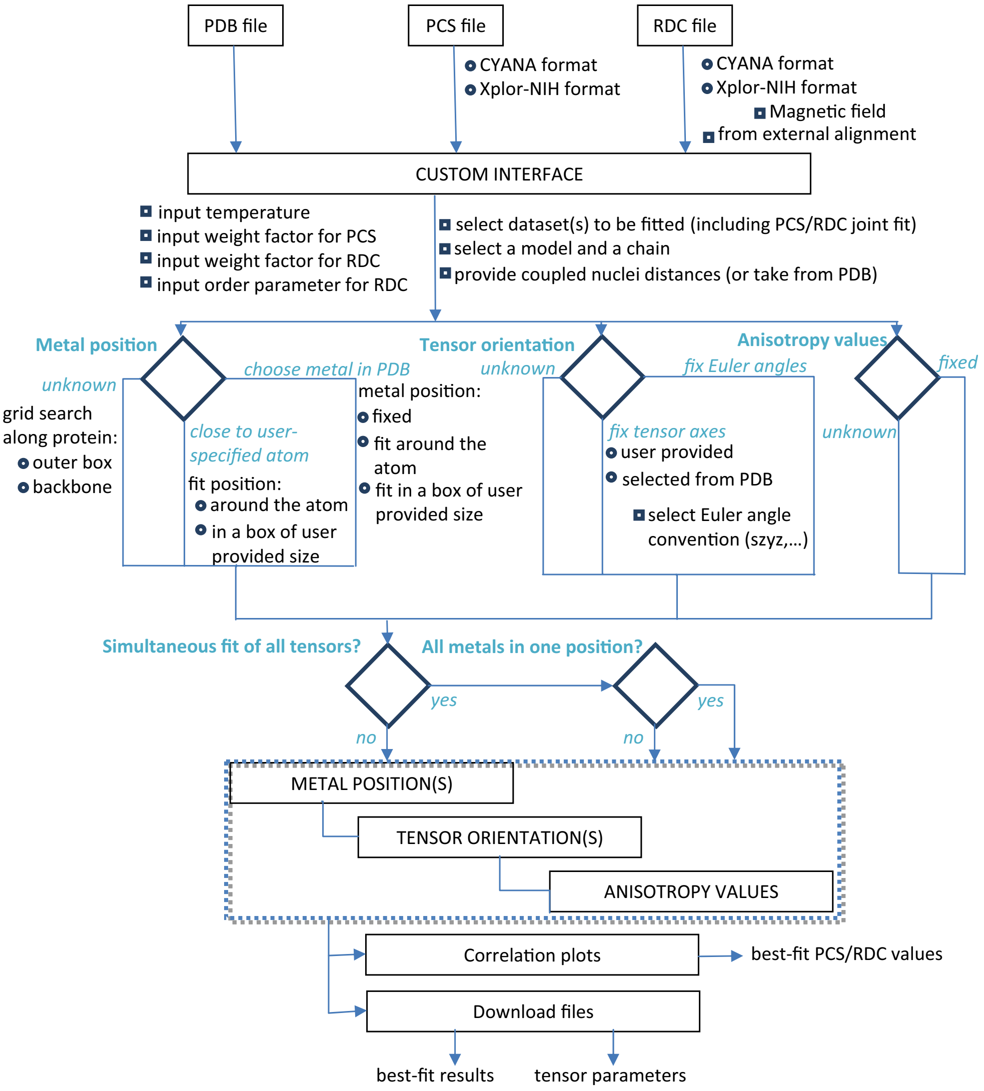


Fig. 2 Scheme of the workflow of the Custom Interface, showing the steps performed and the available options

remaining three characters specify the axes about which the three consecutive rotations occur. The resulting transformation matrix represents the rotation that brings the reference system to coincide with the principal axis system of the anisotropy tensor (passive rotation) (Dosset et al. 2013). As example, in the "szyz" convention, where three consecutive rotations of α, β and γ about the fixed initial z, y and z axes, respectively, of the reference frame occur, the rotation matrix is

$$R = \begin{pmatrix} \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & -\sin\alpha\cos\beta\cos\gamma - \cos\alpha\sin\gamma & \sin\beta\cos\gamma \\ \cos\alpha\cos\beta\sin\gamma + \sin\alpha\cos\gamma & -\sin\alpha\cos\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\beta\sin\gamma \\ -\cos\alpha\sin\beta & \sin\alpha\sin\beta & \cos\beta \end{pmatrix}$$

whereas in the "rzyz" convention, where three consecutive rotations of α, β and γ about the z axis, the new y' axis and the new z'' axis, respectively, occur, the rotation matrix is

$$R = \begin{pmatrix} \cos\alpha\cos\beta\cos\gamma - \sin\alpha\sin\gamma & -\cos\alpha\cos\beta\sin\gamma - \sin\alpha\cos\gamma & \cos\alpha\sin\beta \\ \sin\alpha\cos\beta\cos\gamma + \cos\alpha\sin\gamma & -\sin\alpha\cos\beta\sin\gamma + \cos\alpha\cos\gamma & \sin\alpha\sin\beta \\ -\sin\beta\cos\gamma & \sin\beta\sin\gamma & \cos\beta \end{pmatrix}$$

3. Users may also choose to fix the anisotropy values in case the latter are known from e.g. analogous constructs (Keizers and Ubbink 2011; Skinner et al. 2013) or previous calculations performed for the same paramagnetic metal and protein domain by providing them through the web interface. This can be useful to obtain a preliminary orientation of the tensor which can then be left free for refinement during structure calculations.

Multiple datasets of PCSs and paramagnetic RDCs referring to different metals can be provided for the minimization against a protein structure. In this case a different anisotropy tensor in best agreement with both PCSs and RDCs related to the same paramagnetic metal is calculated for each metal. The positions of the different metals can be independent or constrained to coincide. The different datasets may in fact arise from the presence of different paramagnetic ions alternatively substituted in the same position: in these cases, the different datasets must be associated to different metals (because the associated anisotropy tensors are different), and the different metals can be either constrained to have the same coordinates or restrained individually using the same distance restraints. In case that the metals are constrained to have the same coordinates, the restraints on the metal position selected for the first dataset are automatically applied to all other datasets, with the additional constraint that the best fit position must be unique for all metals.

The program can also calculate PCSs and/or RDCs, without performing any minimization, if the anisotropy parameters are provided as input together with a single PDB file containing the protein structure and the pseudoatoms defining the anisotropy tensor, as obtained from PARAMAGNETIC CYANA or Xplor-NIH.

The distances of coupled atoms, r_{AB} in Eqs. 2 and 3, can be (1) determined from the coordinates of the protein atoms, (2) fixed to default values, or (3) fixed to values preset by the user. Also the temperature and the model-free order parameter (S_{LS}) can be fixed by the user. In case calculations are performed using both PCSs and RDCs to determine the same anisotropy tensor, their relative contribution can be controlled through a weight constant.

The Custom Interface permits the optimization of metal coordinates, anisotropy parameters and/or Euler angles defining the orientation of the anisotropy tensor. The minimization is performed using the Levenberg–Marquardt algorithm, available from SciPy library, which also permits the inclusion of constraints on the parameters to be optimized in the minimization procedure. A least-squares approach is used to minimize the norm of the vector with components provided by the product of the differences between experimental and back-calculated data and the corresponding weights. The latter are provided by the product between the specific weights provided in the uploaded files (if present) and the global weights provided through the web interface in the case that both PCSs and RDCs are fitted simultaneously. The default values for the global weights for PCSs and RDCs are 1 and the ratio between the norm of vector with components equal to the experimental values of RDCs and of PCSs, respectively. The target function which is minimized is

$$TF = \left[\sum_l \sum_i \left(w'_l w_i \left(X_{i,obs} - X_{i,calc} \right) \right)^2 \right]^{1/2}$$

where the index l runs over the two classes of restraints (PCSs and RDCs), the index i on all experimental data of each class, w_i is the weight of the i-restraint and w'_l is the weight of the l-class restraints.

Once the anisotropy tensor has been calculated through the best fit minimization, the quality of the fit can be evaluated at a glance by inspection of the correlation plots, which show the agreement between experimental and back-calculated PCSs and RDCs, as shown in Fig. 1b. These interactive plots allow the user to identify outliers by positioning the pointer directly on the data point. Moreover, the agreement between experimental and back-calculated data is quantified and reported on the top left of the plots, in terms of Q-factors of PCSs and RDCs (Q_{PCS} and Q_{RDC}).

The calculation provides the anisotropy parameters ($\Delta\chi_{ax}$ and $\Delta\chi_{rh}$) and the Euler angles (in the "szyz" convention or in any user-selected convention) defining the main coordinate frame of the anisotropy tensor with respect to the frame of the protein coordinates. The tensor matrix and the eigenvectors providing the main axes of the tensor, i.e. the frame where χ is diagonal, are also provided, together with the tensor representation as a triad of pseudoatoms representing unit vectors in PDB format. The latter can be easily copied and pasted in the uploaded file for graphical visualization. Experimental and back-calculated data are also given at the bottom of the page in tabular form.

When RDCs arising from partial alignment induced by external orienting media are provided, the results of the calculation are given in terms of alignment tensors. As mentioned in the methods section, the size of the tensors is provided by their magnitude A and rhombicity R; for the sake of completeness also the maximum D_{NH} value is given (Fig. S2).

Simultaneous calculation of all tensors referring to several datasets can be performed after the set-up of each of them. The program automatically generates a summary of all calculated tensors, including the plots showing the agreement between experimental and calculated data for the different tensors, so that they can be easily compared (Fig. S3).

All results from the fit procedure and the associated parameters can be downloaded as text files.

Smart Interface

This interface can be either used as a simplified stand-alone interface or, after having run the custom interface, to exploit its additional graphical features or for the calculation of the Monte Carlo error (Fig. 3). The use of this simplified interface is recommended to obtain the best fit anisotropy tensor(s) when the position of the paramagnetic metals is known (or previously determined through the Custom Interface) or when diamagnetic RDCs should be analyzed. Since the metal position does not need to be determined, the Smart Interface permits a more efficient calculation of the anisotropy tensor, which can be estimated directly using Eqs. 1–3. In this case the first derivate can be easily computed and used in a Gauss–Newton optimization procedure. As for the Custom Interface, the procedure minimizes the squared residuals between experimental and back-calculated data, weighted by the product between the specific weights provided in the uploaded files (if present) and the global weights provided through the web interface in the case that both PCSs and RDCs are fitted simultaneously. Again, a model-free order parameter, the value of which can be adjusted in the General Parameters frame of the interface, is used as in Eqs. 2 and 3. Moreover, this interface allows the users to perform the minimization in a single step even if multiple datasets of PCSs and/or RDCs are provided in input, as described above (Fig. 4a).

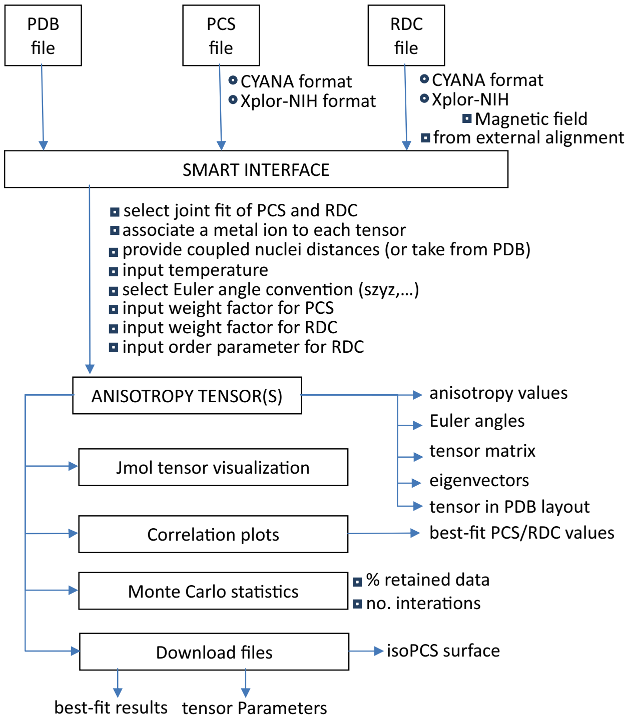
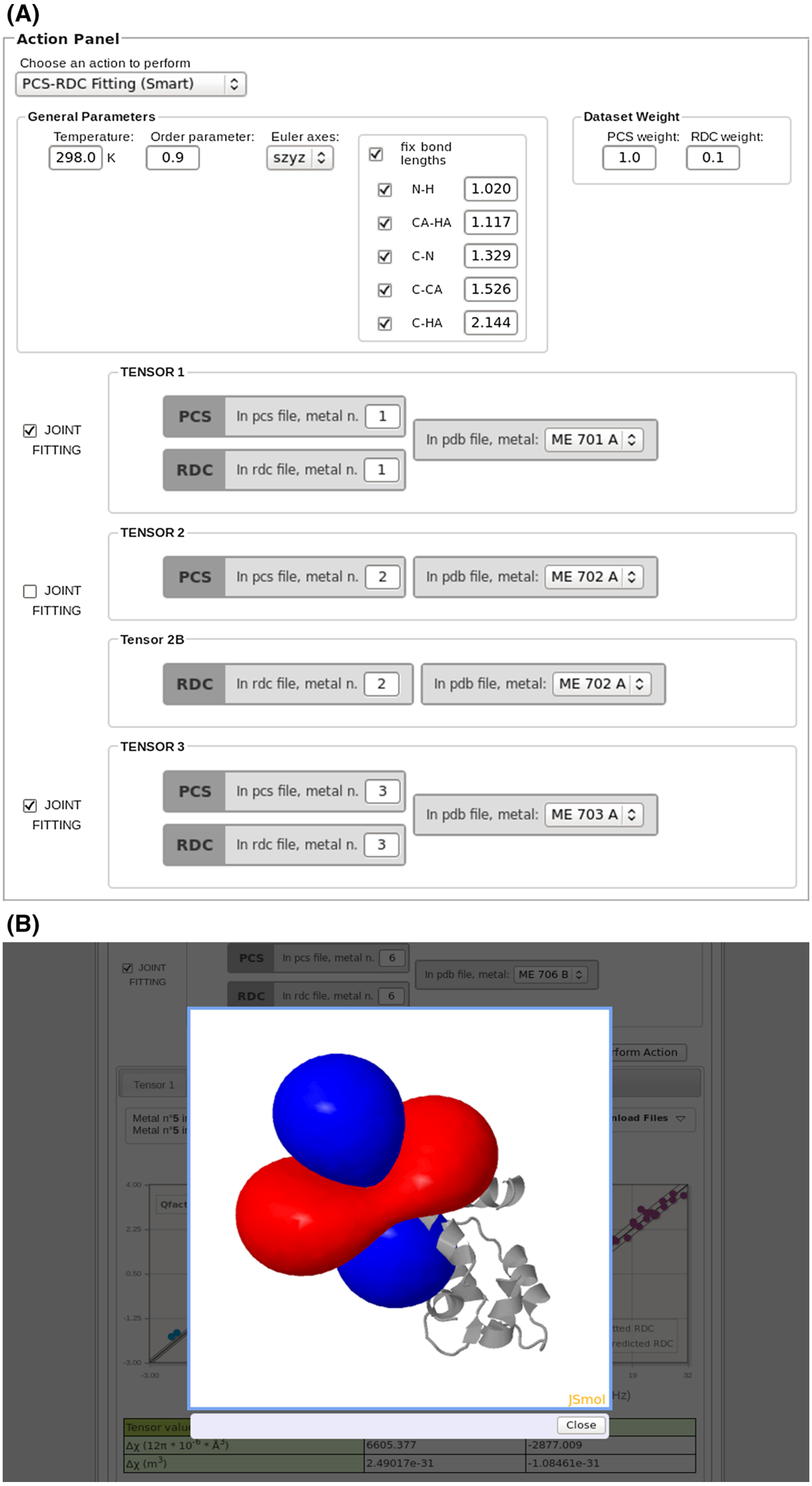


Fig. 3 Scheme of the workflow of the Smart Interface, showing the steps performed and the available options

If a single protein structure is provided as input, the errors on the best fit anisotropy values and on the values of the Euler angles defining the orientation of the anisotropy

Fig. 4 **a** The Smart Interface permits to upload multiple PCSs and/or RDCs datasets and determine the anisotropy tensors in a single step. A single anisotropy tensor in best agreement with both PCSs and RDCs can be determined, at will, through the "joint fitting" option. **b** The isoPCS surface (with threshold equal to 1 ppm) superimposed to the protein chain can be visualized through an integrated JSmol applet



Action Panel

Choose an action to perform: PCS-RDC Fitting (Smart)

General Parameters

Temperature: 298.0 K

Order parameter: 0.9

Euler axes: szyz

☑ fix bond lengths

☑ N-H 1.020

☑ CA-HA 1.117

☑ C-N 1.329

☑ C-CA 1.526

☑ C-HA 2.144

Dataset Weight

PCS weight: 1.0

RDC weight: 0.1

TENSOR 1

☑ JOINT FITTING

PCS In pcs file, metal n. 1

RDC In rdc file, metal n. 1

In pdb file, metal: ME 701 A

TENSOR 2

☐ JOINT FITTING

PCS In pcs file, metal n. 2

In pdb file, metal: ME 702 A

Tensor 2B

RDC In rdc file, metal n. 2

In pdb file, metal: ME 702 A

TENSOR 3

☑ JOINT FITTING

PCS In pcs file, metal n. 3

RDC In rdc file, metal n. 3

In pdb file, metal: ME 703 A

☑ JOINT FITTING

PCS In pcs file, metal n. 6

RDC In rdc file, metal n. 6

In pdb file, metal: ME 706 B

Tensor valu		
Δχ (12π * 10^{-6} * Å^3)	6605.377	-2877.009
Δχ (m^3)	2.49017e-31	-1.08461e-31

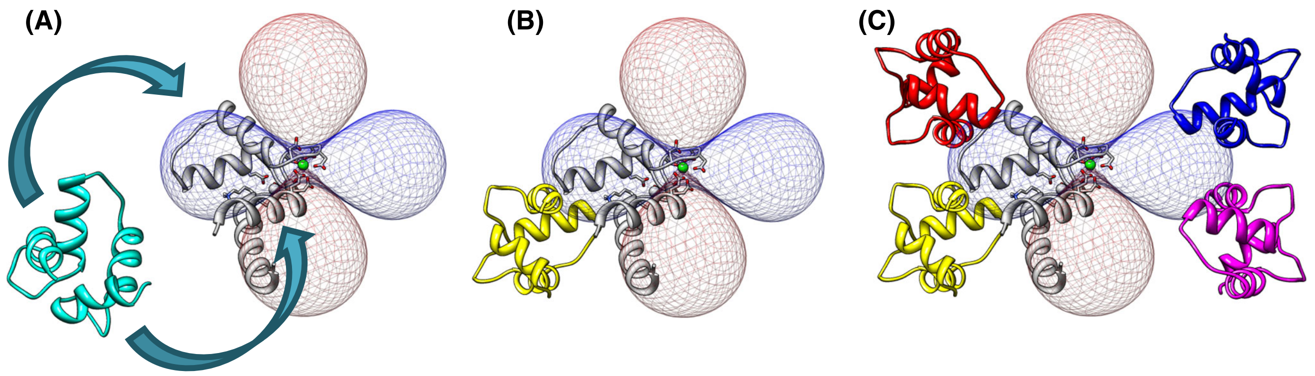


Fig. 5 A rigid body minimization can be performed after determination of the anisotropy tensor for the Subunit A, shown in *gray* (**a**). The position of the subunit B, in *yellow*, is uniquely determined if at least two datasets with non-parallel anisotropy axes are provided (**b**). If a single dataset is provided, the four degenerate solutions are shown (**c**)

tensors can be estimated through a bootstrap Monte Carlo approach, in which the user may decide the percentage of retained experimental data used during the calculation as well as the number of iterations (Fig. S4). When PCSs are used for the calculation, the isoPCS surface can be generated and visualized (with threshold equal to 1 ppm) superimposed to the protein chain selected for the calculation by an integrated JSmol applet, as shown in Fig. 4b. The isosurface is downloadable as a cube file that can be imported by PyMol and UCSF Chimera.

In case a structural ensemble is provided as input, a tensor for each model contained in the PDB is estimated, and $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ values, the Euler angles and the tensor matrices are reported in terms of average and standard deviation obtained from the different models. The results of the individual fits are also given and available for download. A plot showing the agreement between experimental and back-calculated PCSs and RDCs for all the models is provided in the interface, with the possibility of visualizing the data referred to the different models by simply clicking/unclicking of the corresponding checkbox (Fig. S5).

Rigid body minimization

In case of systems composed of multiple domains or subunits the structures of which are rigid and available but with unknown relative positions, PCSs and RDCs arising in the presence of paramagnetic metal(s) in one subunit can be used to determine the relative position of the other subunit(s). As already seen, PCSs and/or RDCs in one subunit can easily provide the anisotropy tensor(s); since PCSs and/or RDCs in another subunit should also yield the same tensor(s) (provided the overall system is rigid), the superposition of the tensors determined from the fit of the data measured for the two subunits permits to reconstruct their relative position (or orientation if only RDCs are used).

This interface has the same structure for data upload of the previous interfaces. After the upload of the first molecular structure (Subunit A) and the related experimental datasets, a second molecular structure (Subunit B) and the corresponding PCS/RDC datasets can be provided (Fig. S6). The datasets for the two subunits corresponding to the same tensors can be paired and associated to the corresponding metals, the coordinates of which must be provided in the PDB file related to Subunit A. After having determined the anisotropy tensors from the data referred to Subunit A (see Smart Interface for the algorithm used), the algorithm automatically performs a grid search to determine the metal position and tensor orientations with respect to the coordinates of Subunit B (see Custom Interface for the grid search algorithm). The algorithm maximizes the agreement between the experimental and the back-calculated data referred to Subunit B, using the anisotropy values previously determined for Subunit A. Subunit B is then translated and rotated rigidly with respect to Subunit A in order to superimpose the anisotropy tensors of the two subunits. In this way the relative position of the two subunits is recovered. A further estimate of the paramagnetic tensor based on the resulting structure, i.e. the structure comprising both Subunit A and the rototranslated Subunit B, is finally carried out. As in the previously described interfaces, FANTEN provides as output the anisotropy parameters ($\Delta\chi_{ax}$ and $\Delta\chi_{rh}$), the values of the Euler angles, the tensor matrices and the eigenvectors providing the main axes of the tensor together with the plots showing the agreement between experimental and back-calculated data for both Subunit A and Subunit B (Fig. S7). The applied transformation can be written as

$$X' = T1 + M(X - T2)$$

where X and X′ are the coordinates of Subunit B before and after the transformation, respectively, T2 is the vector of the metal coordinates, M is the rotation matrix, and T1 the translation vector. T1, T2 and M are provided as output.

The results of the fit, the parameters defining the calculated tensors as well as the structure comprising Subunit A and the rototranslated Subunit B are available for download.

When all PCS/RDC data refer to a single paramagnetic metal, the position of Subunit B is not univocally determined, but four degenerate solutions are possible (Tolman et al. 1995; Fragai et al. 2013). Indeed, once one of the four degenerate solutions is obtained, the other three are automatically generated by changing the sign of two out of three nuclear coordinates, when the latter are expressed in the anisotropy tensor frame centered on the metal ion. All the possible solutions are available for download.

Finally, if only RDCs are provided, only the relative orientation of the two subunits can be determined; in this case the relative position of Subunit B with respect to Subunit A is unchanged.

Figure 5 shows the results of the rigid body minimization performed for the two-domain protein calmodulin when bound to a peptide from the death-associated protein kinase. The position of the C-terminal domain of the protein with respect to the N-terminal domain is determined using PCSs and RDCs measured for Tb(III) or for three lanthanides (Tb(III), Tm(III) and Yb(III); Bertini et al. 2009).

FANTEN is hosted in the WeNMR gateway, which is a worldwide e-infrastructure for NMR and structural biology (Wassenaar et al. 2012) jointly maintained by the NMR facilities of Florence, Frankfurt and Utrecht.

Conclusions

A new, user-friendly, web-based interface, FANTEN, for the analysis of PCSs and RDCs against structural models has been made available. It allows the users to determine the anisotropy tensors through a joint analysis of these restraints. The agreement of the experimental data with the structural model can be monitored from the quality of the best fit of the experimental data.

In structural calculation protocols, the PCSs and RDCs files used in either PARAMAGNETIC CYANA or Xplor-NIH can be given as input to FANTEN together with the PDB file of the protein produced by these programs, so that the anisotropy tensors (and possibly the metal coordinates) can be obtained and used for further PARAMAGNETIC CYANA or Xplor-NIH calculations, until convergence is reached. Alternatively, FANTEN can be used to predict PCS and RDC data from the coordinates of the protein structure and the anisotropy tensor either provided as input or determined from the available data.

FANTEN can be run from the open access web site http://fanten-enmr.cerm.unifi.it:8080 through the WeNMR gateway.

References

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 143:402–406

Andralojc W, Luchinat C, Parigi G, Ravera E (2014) Exploring regions of conformational space occupied by two-domain proteins. J Phys Chem B 118:10576–10587

Balayssac S, Bertini I, Luchinat C, Parigi G, Piccioli M (2006) ^{13}C direct detected NMR increases the detectability of residual dipolar couplings. J Am Chem Soc 128:15042–15043

Banci L, Bertini I, Bren KL, Cremonini MA, Gray HB, Luchinat C, Turano P (1996) The use of pseudocontact shifts to refine solution structures of paramagnetic metalloproteins: Met80Ala cyano-cytochrome *c* as an example. J Biol Inorg Chem 1:117–126

Banci L, Bertini I, Huber JG, Luchinat C, Rosato A (1998) Partial orientation of oxidized and reduced cytochrome b5 at high magnetic fields: magnetic susceptibility anisotropy contributions and consequences for protein solution structure determination. J Am Chem Soc 120:12903–12909

Banci L, Bertini I, Cavallaro G, Giachetti A, Luchinat C, Parigi G (2004) Paramagnetism-based restraints for Xplor-NIH. J Biomol NMR 28:249–261

Berlin K, O'Leary DP, Fushman D (2009) Improvement and analysis of computational methods for prediction of residual dipolar couplings. J Magn Reson 201:25–33

Berlin K, O'Leary DP, Fushman D (2010) Structural assembly of molecular complexes based on residual dipolar couplings. J Am Chem Soc 132:8961–8972

Berlin K, O'Leary DP, Fushman D (2011) Fast approximations of the rotational diffusion tensor and their application to structural assembly of molecular complexes. Proteins 79:2268–2281

Bertini I, Donaire A, Jiménez B, Luchinat C, Parigi G, Piccioli M, Poggi L (2001) Paramagnetism-based versus classical constraints: an analysis of the solution structure of Ca Ln Calbindin D9k. J Biomol NMR 21:85–98

Bertini I, Luchinat C, Parigi G (2002) Magnetic susceptibility in paramagnetic NMR. Prog NMR Spectrosc 40:249–273

Bertini I, Faraone-Mennella J, Gray BH, Luchinat C, Parigi G, Winkler JR (2004) NMR-validated structural model for oxidized *Rhodopseudomonas palustris* cytochrome c556. J Biol Inorg Chem 9:224–230

Bertini I, Gupta YK, Luchinat C, Parigi G, Peana M, Sgheri L, Yuan J (2007) Paramagnetism-based NMR restraints provide maximum allowed probabilities for the different conformations of partially independent protein domains. J Am Chem Soc 129:12786–12794

Bertini I, Kursula P, Luchinat C, Parigi G, Vahokoski J, Willmans M, Yuan J (2009) Accurate solution structures of proteins from X-ray data and minimal set of NMR data: calmodulin peptide complexes as examples. J Am Chem Soc 131:5134–5144

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 132:13553–13558

Bertini I, Case DA, Ferella L, Giachetti A, Rosato A (2011a) A grid-enable web portal for NMR structure refinement with AMBER. Bioinformatics 27:2384–2390

Bertini I, Luchinat C, Parigi G (2011b) Moving the frontiers in solution solid state bioNMR. A celebration of Harry Gray's 75th birthday. Coord Chem Rev 255:649–663

Bertini I, Calderone V, Cerofolini L, Fragai M, Geraldes CFGC, Hermann P, Luchinat C, Parigi G, Teixeira JMC (2012a) The catalytic domain of MMP-1 studied through tagged lanthanides. Dedicated to Prof. A. V. Xavier. FEBS Lett 586:557–567

Bertini I, Ferella L, Luchinat C, Parigi G, Petoukhov MV, Ravera E, Rosato A, Svergun DI (2012b) MaxOcc: a web portal for maximum occurence analysis. J Biomol NMR 53:271–280

Bryson M, Tian F, Prestegard JH, Valafar H (2008) REDCRAFT: a tool for simultaneous characterization of protein backbone structure and motion from RDC data. J Magn Reson 191:322–334

Cerofolini L, Fields GB, Fragai M, Geraldes CFGC, Luchinat C, Parigi G, Ravera E, Svergun DI, Teixeira JMC (2013) Examination of matrix metalloproteinase-1 (MMP-1) in solution: a preference for the pre-collagenolysis state. J Biol Chem 288:30659–30671

Chill JH, Louis JM, Delaglio F, Bax A (2007) Local and global structure of the monomeric subunit of the potassium channel KcsA probed by NMR. Biochim Biophys Acta 1768:3260–3270

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 8:990–997

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 USA 97:9021–9025

Cornilescu G, Marquardt J, Ottiger M, Bax A (1998) Validation of protein structure from anisotropic carbonyl chemical shifts in a dilute liquid crystalline phase. J Am Chem Soc 120:6836–6837

Das Gupta S, Hu X, Keizers PHJ, Liu W-M, Luchinat C, Nagulapalli M, Overhand M, Parigi G, Sgheri L, Ubbink M (2011) Narrowing the conformational space sampled by two-domain proteins with paramagnetic probes in both domains. J Biomol NMR 51:253–263

Diaz-Moreno I, Diaz-Quintana A, De la Rosa MA, Ubbink M (2005) Structure of the complex between plastocyanin and cytochrome f from the cyanobacterium nostoc Sp. PCC 7119 as determined by paramagnetic NMR. J Biol Chem 280:18908–18915

Dosset P, Hus JC, Marion D, Blackledge M (2001) A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. J Biomol NMR 20:223–231

Dosset P, Barthe P, Cohen-Gonsaud M, Roumestand C, Déméné H (2013) Equivalence between Euler angle conventions for the description of tensorial interactions in liquid NMR: application to different software programs. J Biomol NMR 57:305–311

Fragai M, Luchinat C, Parigi G, Ravera E (2013) Conformational freedom of metalloproteins revealed by paramagnetism-assisted NMR. Coord Chem Rev 257:2652–2667

Fufezan C, Specht M (2009) p3d–Python module for structural bioinformatics. BMC bioinform 10:258

Gaponenko V, Sarma SP, Altieri AS, Horita DA, Li J, Byrd RA (2004) Improving the accuracy of NMR structures of large proteins using pseudocontact shifts as long/range restraints. J Biomol NMR 28:205–212

Gardner RJ, Longinetti M, Sgheri L (2005) Reconstruction of orientations of a moving protein domain from paramagnetic data. Inverse Probl 21:879–898

Gempf KL, Butler SJ, Funk AM, Parker D (2013) Direct and selective tagging of cysteine residues in peptides and proteins with 4-nitropyridyl lanthanide complexes. Chem Commun (Camb) 49:9104–9106

Gochin M, Roder H (1995) Protein structure refinement based on paramagnetic NMR shifts. applications to wild-type and mutants forms of cytochrome c. Protein Sci 4:296–305

Grishaev A, Tugarinov V, Kay LE, Trewhella 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 40:95–106

Güntert P (2004) Automated NMR structure calculation with CYANA. Methods Mol Biol 278:353–378

Hass MAS, Keizers PHJ, Blok A, Hiruma Y, Ubbink M (2010) Validation of a lanthanide tag for the analysis of protein dynamics by paramagnetic NMR spectroscopy. J Am Chem Soc 132:9952–9953

Häussinger D, Huang J, Grzesiek S (2009) DOTA-M8: an extremely rigid, high-affinity lanthanide chelating tag for PCS NMR spectroscopy. J Am Chem Soc 131:14761–14767

Hulsker R, Baranova MV, Bullerjahn GS, Ubbink M (2008) Dynamics in the transient complex of plastocyanin-cytochrome f from Prochlorothrix hollandica. J Am Chem Soc 130: 1985–1991

Jensen MR, Hansen DF, Ayna U, Dagil R, Hass MA, Christensen HE, Led JJ (2006) On the use of pseudocontact shifts in the structure determination of metalloproteins. Magn Reson Chem 44: 294–301

John M, Otting G (2007) Strategies for measurements of pseudocontact shifts in protein NMR spectroscopy. ChemPhysChem 8:2309–2313

John M, Schmitz C, Park AY, Dixon NE, Huber T, Otting G (2007) Sequence-specific and stereospecific assignment of methyl groups using paramagnetic lanthanides. J Am Chem Soc 129: 13749–13757

Keizers PHJ, Ubbink M (2011) Paramagnetic tagging for protein structure and dynamics analysis. Prog Nucl Magn Reson Spectrosc 58:88–96

Keizers PHJ, Saragliadis A, Hiruma Y, Overhand M, Ubbink M (2008) Design, synthesis, and evaluation of a lanthanide chelating protein probe: CLaNP-5 yields predictable paramagnetic effects independent of environment. J Am Chem Soc 130:14802–14812

Kemple MD, Ray BD, Lipkowitz KB, Prendergast FG, Rao BDN (1988) The use of lanthanides for solution structure determination of biomolecules by NMR: evaluation of methodology with EDTA derivatives as model systems. J Am Chem Soc 110: 8275–8287

Kobashigawa Y, Saio T, Ushio M, Sekiguchi M, Yokochi M, Ogura K, Inagaki F (2012) Convenient method for resolving degeneracies due to symmetry of the magnetic susceptibility tensor and its application to pseudo contact shift-based protein-protein complex structure determination. J Biomol NMR 53:53–63

Kurland RJ, McGarvey BR (1970) Isotropic NMR shifts in transition metal complexes: calculation of the Fermi contact and pseudocontact terms. J Magn Reson 2:286–301

Lange OF, Lakomek N-A, Farès C, Schröder GF, Walter KFA, Becker S, Meiler J, Grubmüller H, Griesinger C, de Groot BL (2008) Recognition dynamics up to microseconds revealed from an RDC-derived ubiquitin ensemble in solution. Science 320: 1471–1475

Liu WM, Keizers PH, Hass MA, Blok A, Timmer M, Sarris AJ, Overhand M, Ubbink M (2012) A pH-sensitive, colorful, lanthanide-chelating paramagnetic NMR probe. J Am Chem Soc 134:17306–17313

Loh CT, Ozawa K, Tuck KL, Barlow N, Huber T, Otting G, Graham B (2013) Lanthanide tags for site-specific ligation to an unnatural amino acid and generation of pseudocontact shifts in proteins. Bioconjugate Chem 24:260–268

Longinetti M, Parigi G, Sgheri L (2002) Uniqueness and degeneracy in the localization of rigid domains in paramagnetic proteins. J Phys A: Math Gen 35:8153–8169

Longinetti M, Luchinat C, Parigi G, Sgheri L (2006) Efficient determination of the most favored orientations of protein domains from paramagnetic NMR data. Inverse Probl 22:1485–1502

Man B, Su XC, Liang H, Simonsen S, Huber T, Messerle BA, Otting G (2010) 3-Mercapto-2,6-pyridinedicarboxylic acid: a small lanthanide-binding tag for protein studies by NMR spectroscopy. Chem Eur J 16:3827–3832

McConnell HM, Robertson RE (1958) Isotropic nuclear resonance shifts. J Chem Phys 29:1361–1365

Meiler J, Peti W, Griesinger C (2000) DipoCoup: a versatile program for 3D-structure homology comparison based on residual dipolar couplings and pseudocontact shifts. J Biomol NMR 17:283–294

Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53:240–255

Navarro-Vazquez A. (2012) MSpin-RDC. A program for the use of residual dipolar couplings for structure elucidation of small molecules. Magn Reson Chem 50:S73–S79

Oliphant TE (2007) Python for scientific computing. Comput Sci Eng 9:10–20

Otting G (2010) Protein NMR using paramagnetic ions. Annu Rev Biophys 39:387–405

Ozenne V, Bauer F, Salmon L, Huang JR, Jensen MR, Segard S, Bernadó P, Charavay C, Blackledge M (2012) Flexible-meccano: a tool for the generation of explicit ensemble descriptions of intrinsically disordered proteins and their associated experimental observables. Bioinformatics 28:1463–1470

Pintacuda G, Keniry MA, Huber T, Park AY, Dixon NE, Otting G (2004) Fast structure/based assignment of ^{15}N HSQC spectra of selectively ^{15}N labeled paramagnetic proteins. J Am Chem Soc 126:2963–2970

Pintacuda G, Park AY, Keniry MA, Dixonj NE, Otting G (2006) Lanthanide labeling offers fast NMR approach to 3D structure determinations of protein–protein complexes. J Am Chem Soc 128:3696–3702

Prestegard JH, Bougault CM, Kishore AI (2004) Residual dipolar couplings in structure determination of biomolecules. Chem Rev 104:3519–3540

Ravera E, Salmon L, Fragai M, Parigi G, Al-Hashimi HM, Luchinat C (2014) Insights into domain-domain motions in proteins and RNA from solution NMR. Acc Chem Res 47:3118–3126

Rinaldelli M, Ravera E, Calderone V, Parigi G, Murshudov GN, Luchinat C (2014) Simultaneous use of solution NMR and X-ray data REFMAC5 for joint refinement/detection of structural differences. Acta Crystallogr D D70:958–967

Rodriguez-Castañeda F, Haberz P, Leonov A, Griesinger C (2006) Paramagnetic tagging of diamagnetic proteins for solution NMR. Magn Reson Chem 44:S10–S16

Russo L, Maestre-Martinez M, Wolff S, Becker S, Griesinger C (2013) Interdomain dynamics explored by paramagnetic NMR. J Am Chem Soc 135:17111–17120

Saio T, Ogura K, Shimizu K, Yokochi M, Burke TR Jr, Inagaki F (2011) An NMR strategy for fragment-based ligand screening utilizing a paramagnetic lanthanide probe. J Biomol NMR 51:395–408

Schmitz C, Bonvin AM (2011) Protein–protein HADDocking using exclusively pseudocontact shifts. J Biomol NMR 50:263–266

Schmitz C, John M, Park AY, Dixon NE, Otting G, Pintacuda G, Huber T (2006) Efficient chi-tensor determination and NH assignment of paramagnetic proteins. J Biomol NMR 35:79–87

Schmitz C, Stanton-Cook MJ, Su XC, Otting G, Huber T (2008) Numbat: an interactive software tool for fitting $\Delta\chi$-tensors to molecular coordinates using pseudocontact shifts. J Biomol NMR 41:179–189

Schmitz C, Vernon R, Otting G, Baker D, Huber T (2012) Protein structure determination from pseudocontact shifts using ROSETTA. J Mol Biol 416:668–677

Schwieters CD, Kuszewski J, Tjandra N, Clore GM (2003) The Xplor-NIH NMR molecular structure determination package. J Magn Reson 160:65–73

Simon B, Madl T, Mackereth CD, Nilges M, Sattler M (2010) An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution. Angew Chem Int Ed 49:1967–1970

Skinner SP, Moshev M, Hass MAS, Ubbink M (2013) PARAssign—paramagnetic NMR assignments of protein nuclei on the basis of pseudocontact shifts. J Biomol NMR 55:379–389

Su XC, Otting G (2010) Paramagnetic labelling of proteins and oligonucleotides for NMR. J Biomol NMR 46:101–112

Su XC, Huber T, Dixon NE, Otting G (2006) Site-specific labelling of proteins with a rigid lanthanide-binding tag. ChemBioChem 7:1599–1604

Su XC, Man B, Beeren S, Liang H, Simonsen S, Schmitz C, Huber T, Messerle BA, Otting G (2008a) A dipicolinic acid tag for rigid lanthanide tagging of proteins and paramagnetic NMR spectroscopy. J Am Chem Soc 130:10486–10487

Su XC, McAndrew K, Huber T, Otting G (2008b) Lanthanide-binding peptides for NMR measurements of residual dipolar couplings and paramagnetic effects from multiple angles. J Am Chem Soc 130:1681–1687

Swarbrick JD, Ung P, Chhabra S, Graham B (2011a) An iminodiacetic acid based lanthanide binding tag for paramagnetic exchange NMR spectroscopy. Angew Chem Int Ed Engl 50:4403–4406

Swarbrick JD, Ung P, Su XC, Maleckis A, Chhabra S, Huber T, Otting G, Graham B (2011b) Engineering of a bis-chelator motif into a protein alpha-helix for rigid lanthanide binding and paramagnetic NMR spectroscopy. Chem Commun (Camb) 47:7368–7370

Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a diluite liquid crystalline medium. Science 278:1111–1114

Tolman JR, Flanagan JM, Kennedy MA, Prestegard JH (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution. Proc Natl Acad Sci USA 92:9279–9283

Tolman JR, Al-Hashimi HM, Kay LE, Prestegard JH (2001) Structural and dynamic analysis of residual dipolar coupling data for proteins. J Am Chem Soc 123:1416–1424

Valafar H, Prestegard JH (2004) REDCAT: a residual dipolar coupling analysis tool. J Magn Reson 167:228–241

Wassenaar TA, van Dijk M, Loureiro-Ferreira N, van der Schot G, de Vries SJ, Schmitz C, van der Zwan J, Boelens R, Giachetti A, Ferella L, Rosato A, Bertini I, Herrmann T, Jonker HRA, Bagaria A, Jaravine V, Guntert P, Schwalbe H, Vranken WF, Doreleijers JF, Vriend G, Vuister GW, Franke D, Kikhney A, Svergun DI, Fogh RH, Ionides J, Laue ED, Spronk C, Jurksa S, Verlato M, Badoer S, DalPra S, Mazzucato M, Frizziero E, Bonvin AMJJ (2012) WeNMR: structural Biology on the grid. J Grid Computing 10:743–767

Watanabe Y, Hiraoka W, Igarashi M, Ito K, Shimoyama Y, Horiuchi M, Yamamori T, Yasui H, Kuwabara M, Inagaki F, Inanami O (2010) A novel copper(II) coordination at His186 in full-length murine prion protein. Biochem Biophys Res Commun 394: 522–528

Wei Y, Werner MH (2006) iDC: a comprehensive toolkit for the analysis of residual dipolar couplings for macromolecular structure determination. J Biomol 35:17–25

Wöhnert J, Franz KJ, Nitz M, Imperiali B, Schwalbe H (2003) Protein alignment by a coexpressed lanthanide-binding tag for the measurement of residual dipolar couplings. J Am Chem Soc 125:13338–13339

Yagi H, Maleckis A, Otting G (2013a) A systematic study of labelling an alpha-helix in a protein with a lanthanide using IDA-SH or NTA-SH tags. J Biomol NMR 55:157–166

Yagi H, Pilla KB, Maleckis A, Graham B, Huber T, Otting G (2013b) Three-dimensional protein fold determination from backbone amide pseudocontact shifts generated by lanthanide tags at multiple sites. Structure 21:883–890

Zhang Q, Stelzer AC, Fisher CK, Al-Hashimi HM (2007) Visualizing spatially correlated dynamics that directs RNA conformational transitions. Nature 450:1263–1267

Zhuang T, Lee HS, Imperiali B, Prestegard JH (2008) Structure determination of a Galectin-3-carbohydrate complex using paramagnetism-based NMR constraints. Protein Sci 17:1220–1231

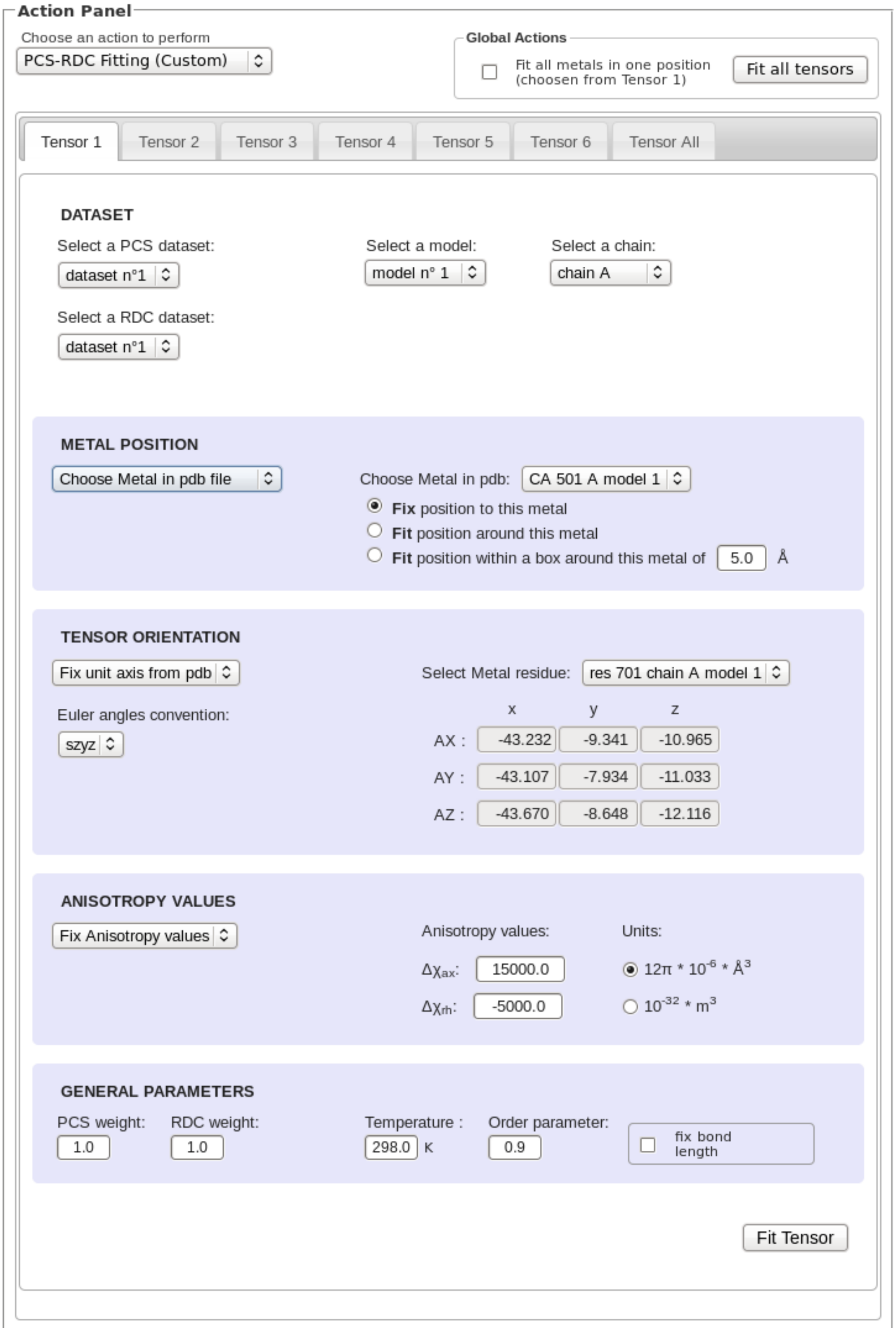
Zweckstetter M (2008) NMR: prediction of molecular alignment from structure using the PALES software. Nat Protoc 3:679–690

Zweckstetter M, Bax A (2000) Prediction of sterically induced alignment in a dilute liquid crystalline phase: aid to protein structure determination by NMR. J Am Chem Soc 122: 3791–3792

FANTEN: a new web-based interface for the analysis of magnetic anisotropy-induced NMR data

Mauro Rinaldelli#, Azzurra Carlon#, Enrico Ravera, Giacomo Parigi*, Claudio Luchinat*

SUPPLEMENTARY MATERIAL



Choose an action to perform: PCS-RDC Fitting (Custom)

☐ Fit all metals in one position (choosen from Tensor 1)

DATASET

Select a PCS dataset: dataset n°1

Select a model: model n° 1

Select a chain: chain A

Select a RDC dataset: dataset n°1

METAL POSITION

Choose Metal in pdb file

Choose Metal in pdb: CA 501 A model 1

● **Fix** position to this metal
○ **Fit** position around this metal
○ **Fit** position within a box around this metal of 5.0 Å

TENSOR ORIENTATION

Fix unit axis from pdb

Select Metal residue: res 701 chain A model 1

Euler angles convention: szyz

	x	y	z
AX :	-43.232	-9.341	-10.965
AY :	-43.107	-7.934	-11.033
AZ :	-43.670	-8.648	-12.116

ANISOTROPY VALUES

Fix Anisotropy values

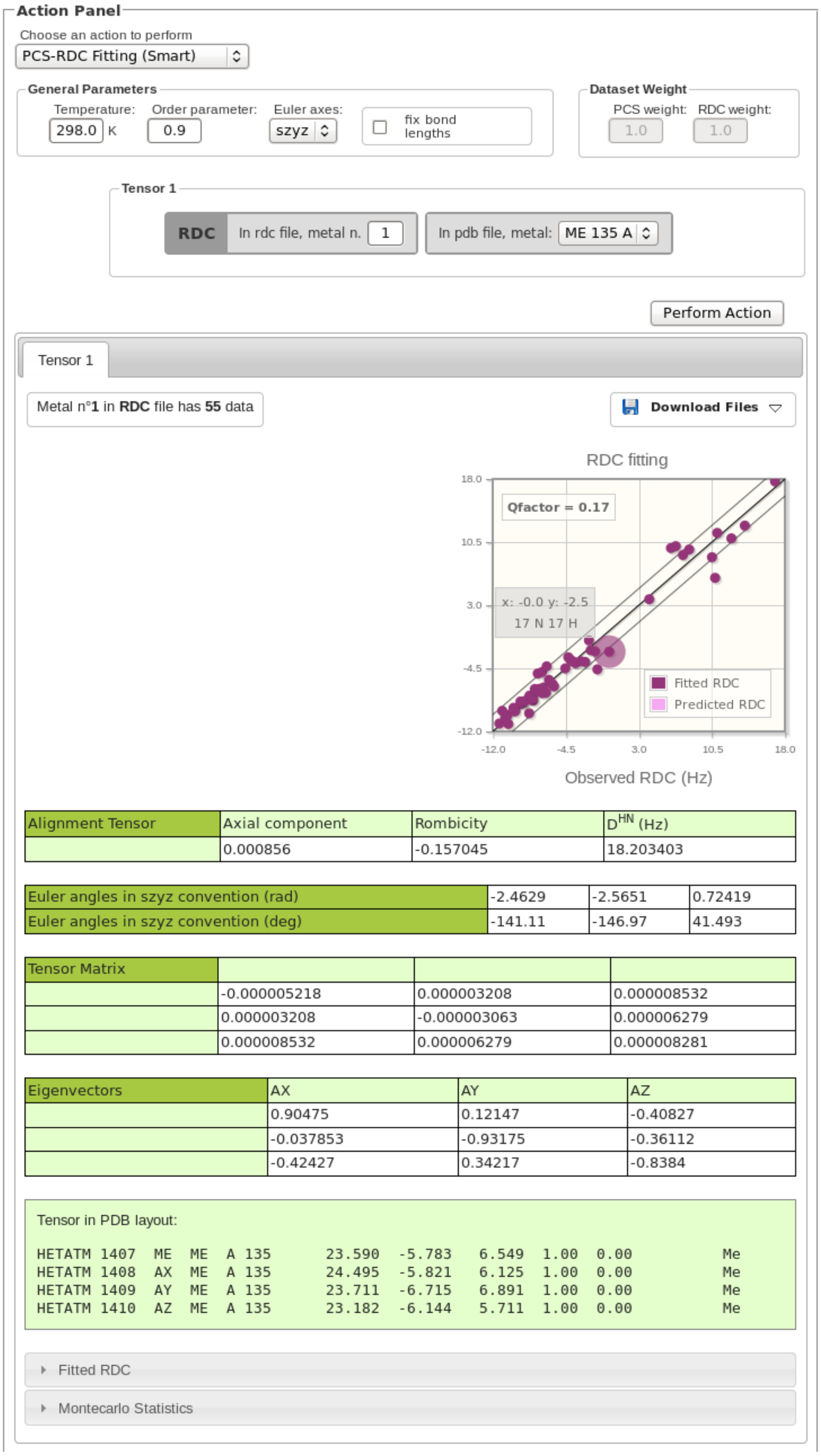
Anisotropy values:
$\Delta\chi_{ax}$: 15000.0
$\Delta\chi_{rh}$: -5000.0

Units:
● $12\pi * 10^{-6} * Å^3$
○ $10^{-32} * m^3$

GENERAL PARAMETERS

PCS weight: 1.0
RDC weight: 1.0
Temperature : 298.0 K
Order parameter: 0.9
☐ fix bond length

Figure S1. CUSTOM interface. It permits to fully customize the calculation of the tensor defining, if known, metal position, tensor orientation, and anisotropy parameters, or any combination thereof.



Choose an action to perform: PCS-RDC Fitting (Smart)

Temperature: 298.0 K
Order parameter: 0.9
Euler axes: szyz
PCS weight: 1.0
RDC weight: 1.0

In rdc file, metal n. 1
In pdb file, metal: ME 135 A

Metal n°**1** in **RDC** file has **55** data

Alignment Tensor	Axial component	Rombicity	D^{HN} (Hz)
	0.000856	-0.157045	18.203403

Euler angles in szyz convention (rad)	-2.4629	-2.5651	0.72419
Euler angles in szyz convention (deg)	-141.11	-146.97	41.493

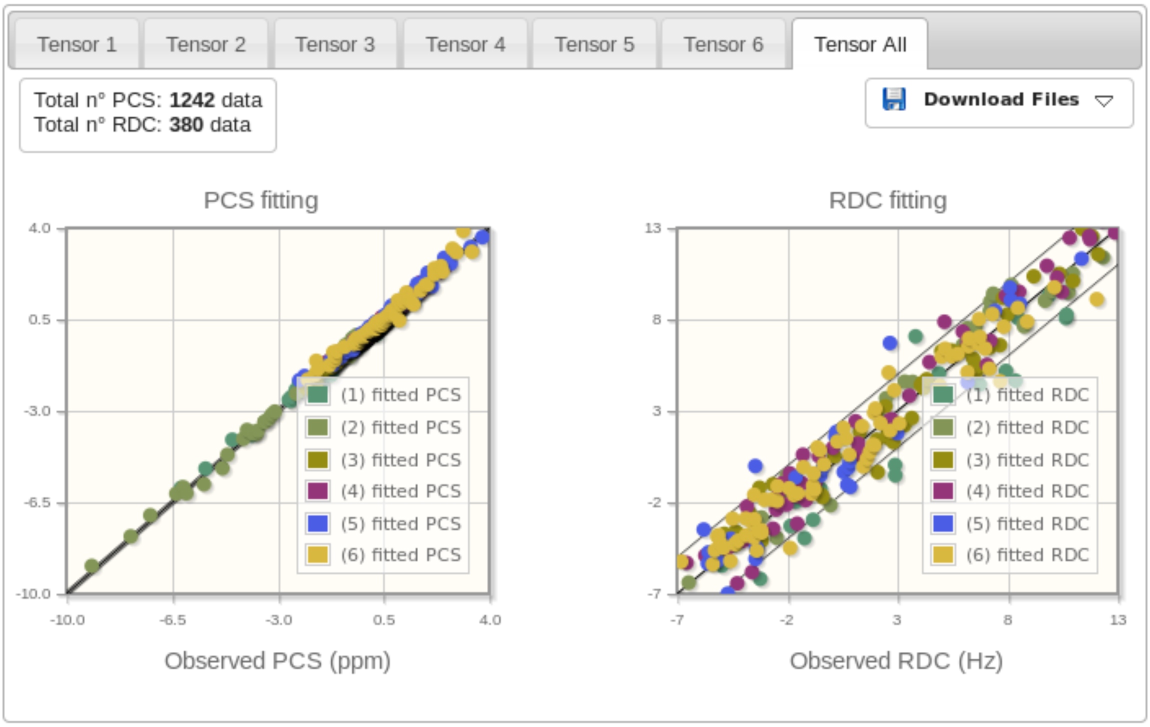
Tensor Matrix			
	-0.000005218	0.000003208	0.000008532
	0.000003208	-0.000003063	0.000006279
	0.000008532	0.000006279	0.000008281

Eigenvectors	AX	AY	AZ
	0.90475	0.12147	-0.40827
	-0.037853	-0.93175	-0.36112
	-0.42427	0.34217	-0.8384

Tensor in PDB layout:

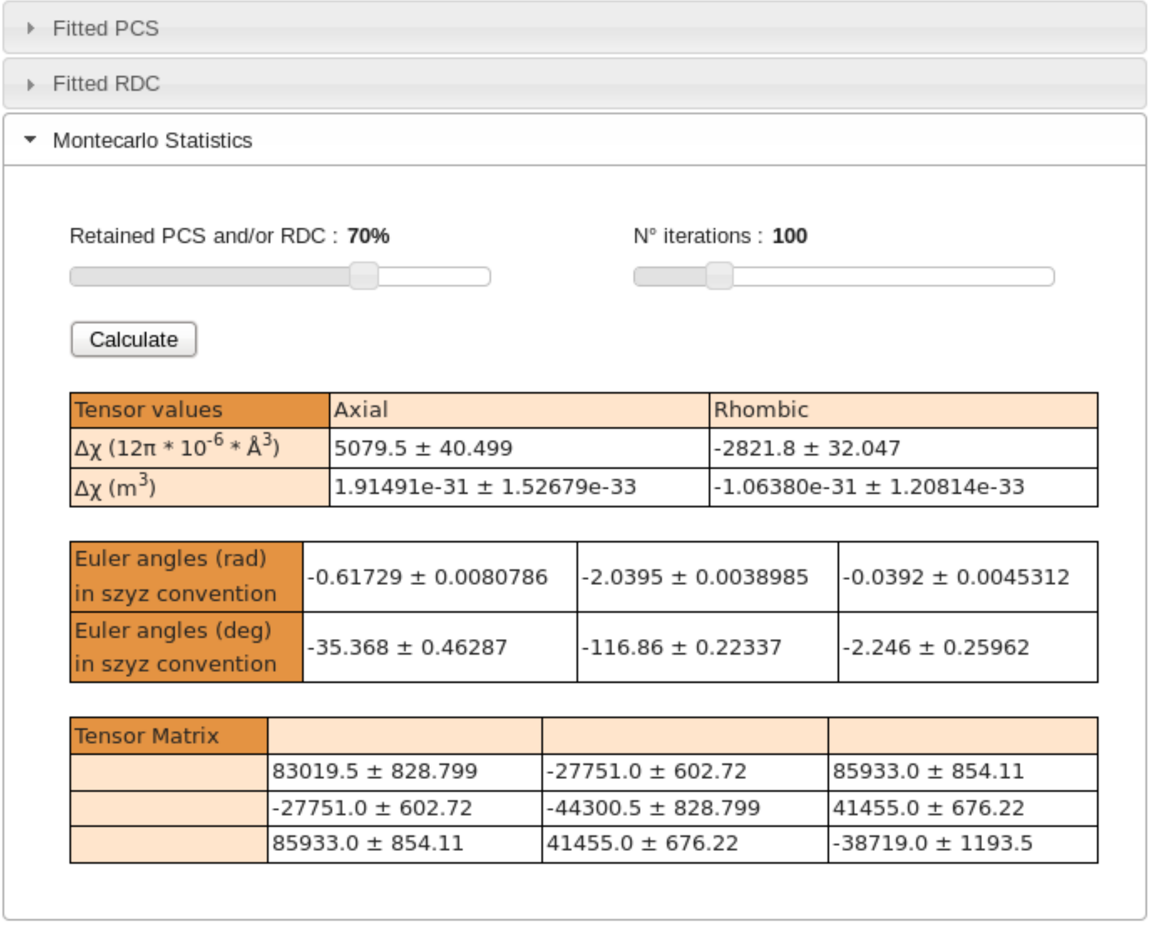
```
HETATM 1407  ME  ME  A 135      23.590  -5.783   6.549  1.00  0.00           Me
HETATM 1408  AX  ME  A 135      24.495  -5.821   6.125  1.00  0.00           Me
HETATM 1409  AY  ME  A 135      23.711  -6.715   6.891  1.00  0.00           Me
HETATM 1410  AZ  ME  A 135      23.182  -6.144   5.711  1.00  0.00           Me
```

Figure S2. SMART Interface. For RDCs arising from external alignment media the size of the tensors is provided by their magnitude A, rhombicity R, and maximum D_{NH} value.



Total n° PCS: **1242** data
Total n° RDC: **380** data

Figure S3. FANTEN automatically generates a summary of all calculated tensors, including the plots showing the agreement between experimental and calculated data for the different tensors.



Retained PCS and/or RDC : **70%**

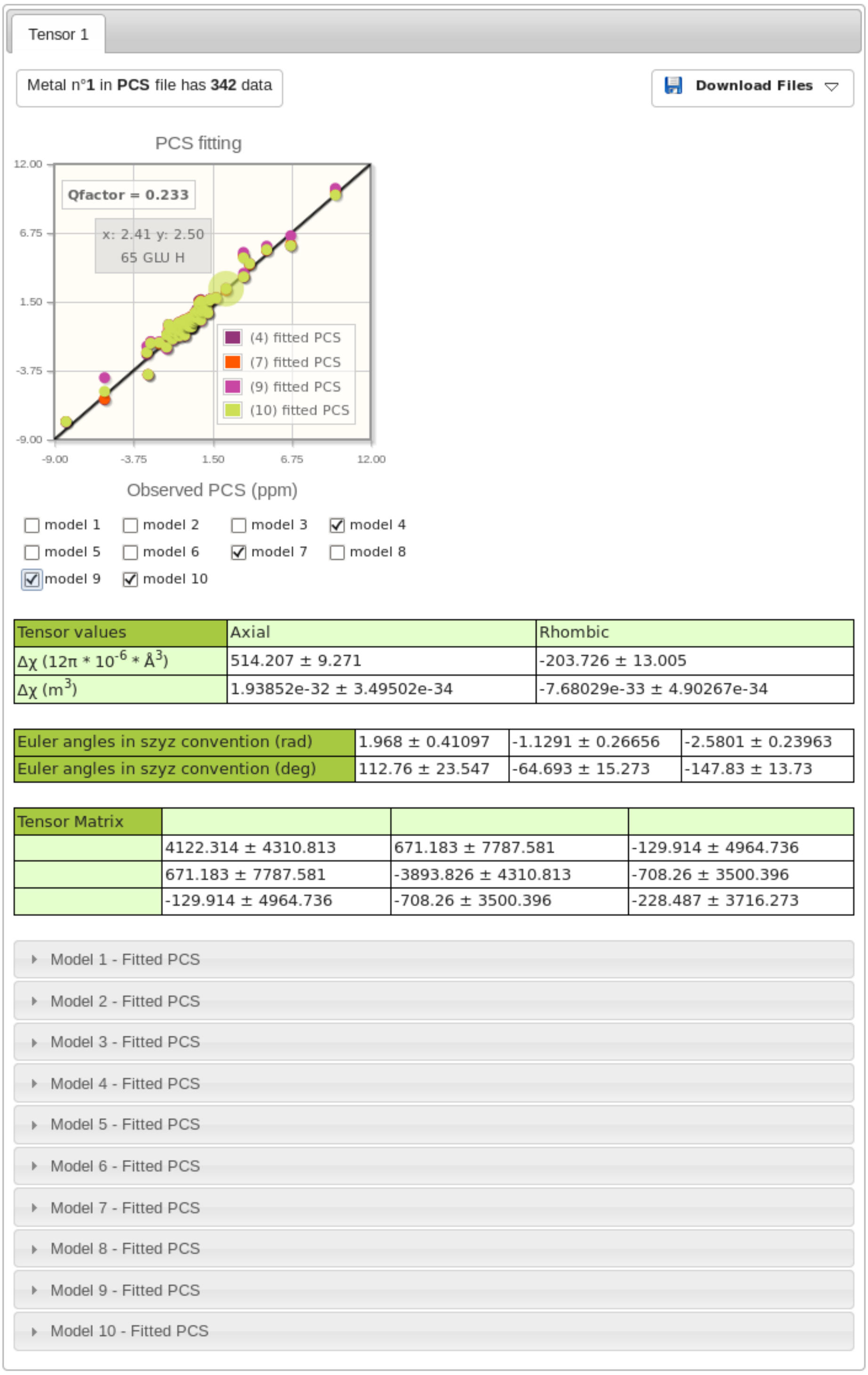
N° iterations : **100**

Tensor values	Axial	Rhombic
Δχ (12π * 10^{-6} * $Å^3$)	5079.5 ± 40.499	-2821.8 ± 32.047
Δχ (m^3)	1.91491e-31 ± 1.52679e-33	-1.06380e-31 ± 1.20814e-33

Euler angles (rad) in szyz convention	-0.61729 ± 0.0080786	-2.0395 ± 0.0038985	-0.0392 ± 0.0045312
Euler angles (deg) in szyz convention	-35.368 ± 0.46287	-116.86 ± 0.22337	-2.246 ± 0.25962

Tensor Matrix			
	83019.5 ± 828.799	-27751.0 ± 602.72	85933.0 ± 854.11
	-27751.0 ± 602.72	-44300.5 ± 828.799	41455.0 ± 676.22
	85933.0 ± 854.11	41455.0 ± 676.22	-38719.0 ± 1193.5

Figure S4. SMART interface. During the a bootstrap Monte Carlo approach for the estimation of the errors on the best fit parameters, the user may decide the percentage of retained experimental data used during the calculation and the number of iterations.



Metal n°**1** in **PCS** file has **342** data

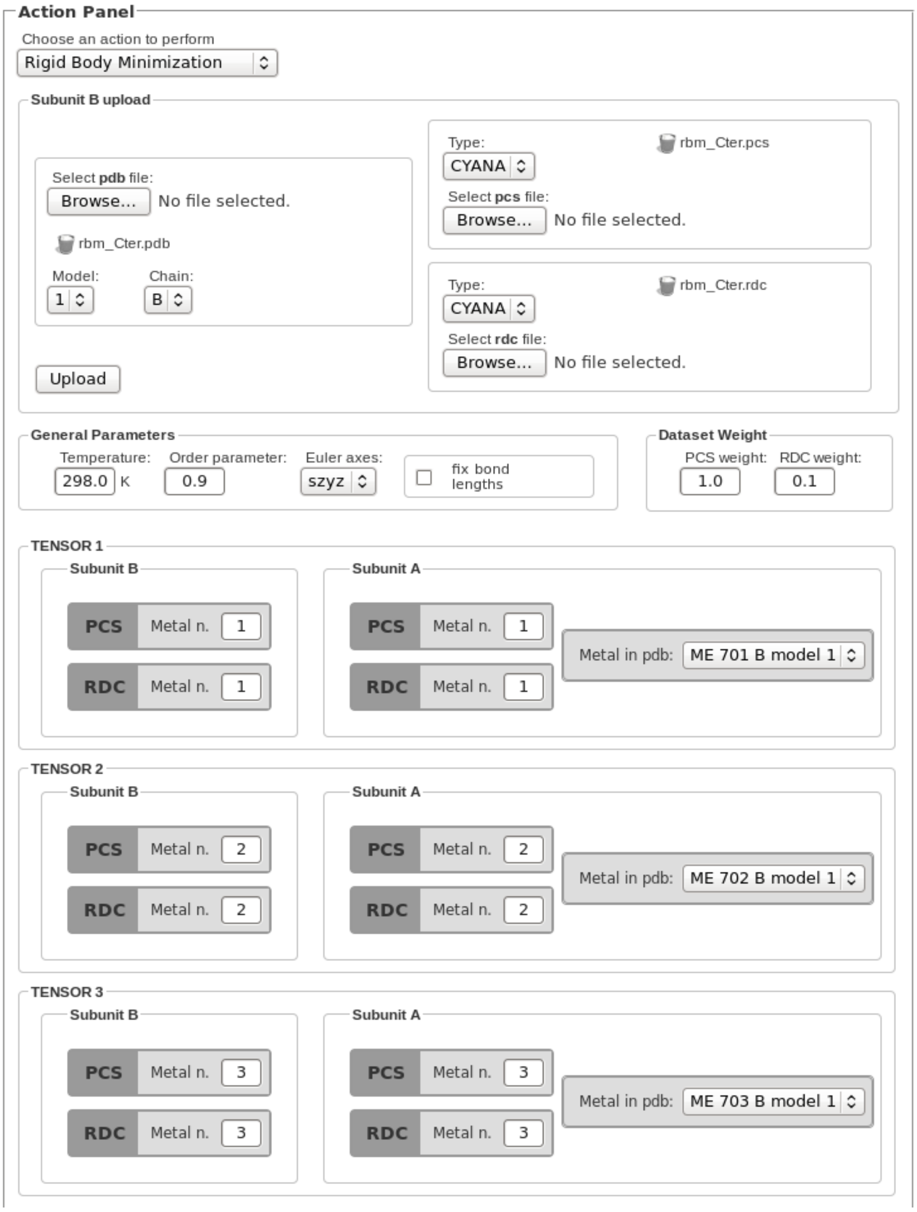
☐ model 1 ☐ model 2 ☐ model 3 ☑ model 4
☐ model 5 ☐ model 6 ☑ model 7 ☐ model 8
☑ model 9 ☑ model 10

Tensor values	Axial	Rhombic
Δχ (12π * 10^{-6} * $Å^3$)	514.207 ± 9.271	-203.726 ± 13.005
Δχ (m^3)	1.93852e-32 ± 3.49502e-34	-7.68029e-33 ± 4.90267e-34

Euler angles in szyz convention (rad)	1.968 ± 0.41097	-1.1291 ± 0.26656	-2.5801 ± 0.23963
Euler angles in szyz convention (deg)	112.76 ± 23.547	-64.693 ± 15.273	-147.83 ± 13.73

Tensor Matrix			
	4122.314 ± 4310.813	671.183 ± 7787.581	-129.914 ± 4964.736
	671.183 ± 7787.581	-3893.826 ± 4310.813	-708.26 ± 3500.396
	-129.914 ± 4964.736	-708.26 ± 3500.396	-228.487 ± 3716.273

Figure S5. SMART interface. When a structural ensemble is provided as input, a tensor for each model contained in the PDB is estimated, and $\Delta\chi_{ax}$ and $\Delta\chi_{rh,}$ values, the Euler angles and the tensor matrices are reported in terms of average and standard deviation obtained from the different models.



Action Panel

Choose an action to perform: Rigid Body Minimization

Subunit B upload

Select pdb file: No file selected.

rbm_Cter.pdb

Model: 1

Chain: B

Type: CYANA

rbm_Cter.pcs

Select pcs file: No file selected.

Type: CYANA

rbm_Cter.rdc

Select rdc file: No file selected.

General Parameters

Temperature: 298.0 K

Order parameter: 0.9

Euler axes: szyz

☐ fix bond lengths

Dataset Weight

PCS weight: 1.0

RDC weight: 0.1

TENSOR 1

Subunit B: PCS Metal n. 1; RDC Metal n. 1

Subunit A: PCS Metal n. 1; RDC Metal n. 1; Metal in pdb: ME 701 B model 1

TENSOR 2

Subunit B: PCS Metal n. 2; RDC Metal n. 2

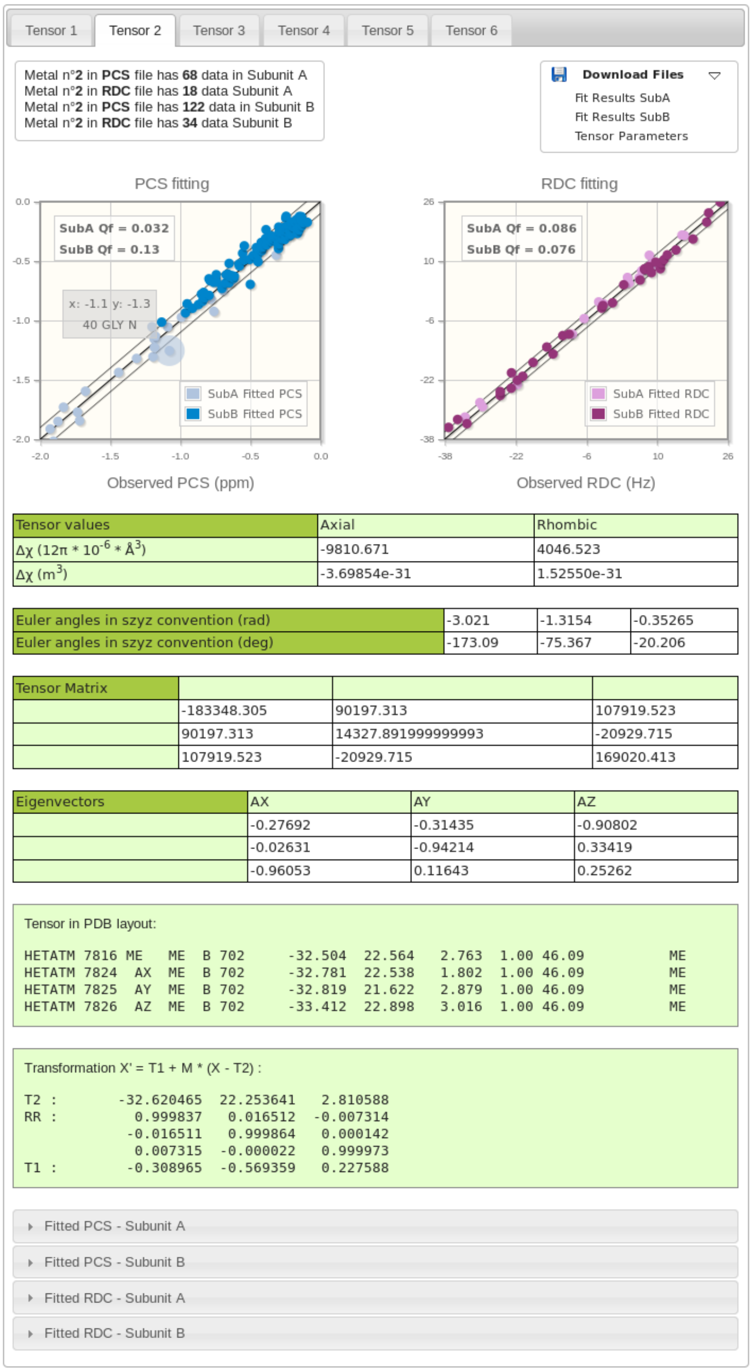
Subunit A: PCS Metal n. 2; RDC Metal n. 2; Metal in pdb: ME 702 B model 1

TENSOR 3

Subunit B: PCS Metal n. 3; RDC Metal n. 3

Subunit A: PCS Metal n. 3; RDC Metal n. 3; Metal in pdb: ME 703 B model 1

Figure S6. Rigid Body Minimization Interface. After the upload of the first molecular structure (Subunit A) and the related experimental datasets, a second molecular structure (Subunit B) and the corresponding PCS/RDC datasets can be provided. The datasets for the two subunits corresponding to the same tensors can be paired and associated to the corresponding metals, the coordinates of which must be provided in the PDB file related to the Subunit A.



Metal n°**2** in **PCS** file has **68** data in Subunit A
Metal n°**2** in **RDC** file has **18** data Subunit A
Metal n°**2** in **PCS** file has **122** data in Subunit B
Metal n°**2** in **RDC** file has **34** data Subunit B

Tensor values	Axial	Rhombic
Δχ (12π * 10^{-6} * $Å^3$)	-9810.671	4046.523
Δχ (m^3)	-3.69854e-31	1.52550e-31

Euler angles in szyz convention (rad)	-3.021	-1.3154	-0.35265
Euler angles in szyz convention (deg)	-173.09	-75.367	-20.206

Tensor Matrix			
	-183348.305	90197.313	107919.523
	90197.313	14327.891999999993	-20929.715
	107919.523	-20929.715	169020.413

Eigenvectors	AX	AY	AZ
	-0.27692	-0.31435	-0.90802
	-0.02631	-0.94214	0.33419
	-0.96053	0.11643	0.25262

Tensor in PDB layout:

```
HETATM 7816 ME   ME  B 702     -32.504  22.564   2.763  1.00 46.09          ME
HETATM 7824  AX  ME  B 702     -32.781  22.538   1.802  1.00 46.09          ME
HETATM 7825  AY  ME  B 702     -32.819  21.622   2.879  1.00 46.09          ME
HETATM 7826  AZ  ME  B 702     -33.412  22.898   3.016  1.00 46.09          ME
```

Transformation X' = T1 + M * (X - T2) :

```
T2 :       -32.620465  22.253641   2.810588
RR :         0.999837   0.016512  -0.007314
            -0.016511   0.999864   0.000142
             0.007315  -0.000022   0.999973
T1 :        -0.308965  -0.569359   0.227588
```

Figure S7. Rigid Body Minimization Interface. It provides as output the anisotropy parameter, the values of the Euler angles, the tensor matrices and the eigenvectors providing the main axes of the tensor together with the plots showing the agreement between experimental and back-calculated data for both Subunit A and Subunit B, and also the information about the applied transformation.

3.2 Extending the possibilities of joint X-ray/NMR structure refinement to high molecular weight complexes

Work in collaboration with: *Prof. Claudio Luchinat*[a,b]*, Dr. Garib N. Murshudov*[c] *and Prof. Lewis E. Kay*[d]

[a] Center for Magnetic Resonance (CERM), University of Florence,
via Sacconi 6, 50019 Sesto Fiorentino (FI), Italy
[b] Department of Chemistry, University of Florence,
via della Lastruccia 3, 50019 Sesto Fiorentino (FI), Italy
[c] MRC Laboratory of Molecular Biology,
Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0QH UK
[d] Departments of Biochemistry, Medical Genetics and Chemistry, The University of Toronto,
Toronto, Ontario M5S 1A8, Canada

Manuscript in preparation

Extending the possibilities of joint X-ray/NMR structure refinement to high molecular weight complexes

Abstract

Integrative approaches play an increasingly important role in structural biology, as it is acknowledged that the best of different methodologies can be brought together to achieve a unified result instead of competing on each other's turf. In particular, the advent of methyl-TROSY and related methodologies, which allow NMR to access systems larger than a MDa, yields direct information on the orientation of aminoacidic side chains. In this frame, combination with other primary data such as X-ray diffraction, which tend to lose handle on side chains for lower resolutions, is needed to maximize the attainable information.

We here present a set of new features of REFMAC-NMR that allow for improved handling of data for large proteins, including symmetry-related restraints for multimeric assemblies.

Keywords:

Structure refinement; integrated structural biology; RDC; symmetric assemblies

Introduction

Historically, solution NMR studies on biomolecules have been limited to molecular weights up to 60 kDa. However, the technological improvements and the new methodologies that have maturated upon the last 15 years have made it possible to approach systems exceeding MDa. The core idea is to detect the resonances of the methyl groups, which can be approached by the sequences tailored to exploit cross-relaxation to drastically reduce the transverse relaxation time and of the improvements in isotope labeling strategies. This strategy, later extended so as to exploit the resonances of aromatic residues has its main disadvantage in the intrinsically limited number of probes (methyl groups) that can be observed. This brings forth an increased need to rely upon libraries of idealized geometry so as to make the best possible use of the available restraints.

We have thus decided to extend the applicability of the REFMAC-NMR approach to the study of larger systems. To do this, we have introduced several new features in REFMAC. First of all, we have introduced the possibility of using residual dipolar couplings from methyl groups, which are likely to be the most common reporters for biomolecules larger than 60 kDa. More importantly, we have introduced symmetry-based restraints in the optimization of the NMR alignment tensors. Large assemblies are often made of several subunits, which may be related by symmetry operations, imposing restraints to the features of their alignment tensor. We also take onto account the massive use of differential labeling of the subunits that is common in biomolecular NMR studies to reduce the spectral complexity in larger systems: this corresponds to producing multiple samples, in which the extent of the external alignment medium may be slightly different depending on the sample preparation, but in which orientation of the alignment tensor can be safely assumed to be maintained.

As a test case, we have applied this newly-developed approach to the case of the 20S proteasome, a heptameric protein playing a crucial role in regulating some cellular processes, such as cell division, gene expression, signal transduction and apoptosis.

Program implementation

Description of the alignment tensor

In solution NMR, it is assumed that molecules reorient freely, sampling all the possible orientations with equal probability. If, for any reason, not all orientations have the same probability, two effects arise: dipolar couplings do not average to zero any more, and the chemical shifts average to values different from those measured in isotropic solutions.[1] Partial orientation of molecules in solution can be induced by the use of external alignment media in solution (i.e. bicelles, poly-acrylamide gels, bacteriophages), in which case the extent of the alignment depends on the shape of the molecule and/or on its charge distribution. Alignment also occurs when the magnetic susceptibility of the molecule is strongly anisotropic (also referred to as “self-alignment”), such as in the presence of paramagnetic centers, of extended aromatic systems or in nucleic acid stacks. In the most general form, the effect of residual dipolar coupling (RDC) is described as follows:

$$RDC = 3k\left[S_{zz}\frac{2z_{AB}^2 - x_{AB}^2 - y_{AB}^2}{2r_{AB}^2} + \left(S_{xx} - S_{yy}\right)\frac{x_{AB}^2 - y_{AB}^2}{2r_{AB}^2} + S_{xy}\frac{2x_{AB}y_{AB}}{r_{AB}^2} + S_{xz}\frac{2x_{AB}z_{AB}}{r_{AB}^2} + S_{yz}\frac{2y_{AB}z_{AB}}{r_{AB}^2}\right]$$

with

$$k = -\frac{\mu_0 S_{LS}}{4\pi}\frac{\gamma_A\gamma_B\hbar}{2\pi r_{AB}^3}$$

$$x_{AB}^2 = (x_A - x_B)^2,\ \ y_{AB}^2 = (y_A - y_B)^2,\ \ z_{AB}^2 = (z_A - z_B)^2,\ \ r_{AB}^2 = x_{AB}^2 + y_{AB}^2 + z_{AB}^2$$

where r_{AB} is the distance between the two coupled nuclei A and B, and S_{ij} are the components of the molecular alignment tensor. The anisotropies of the alignment tensor are described by the fraction of alignment along the z-axis (A) and by the rhombicity (R) as follows:

$$A = \frac{3}{2}\tilde{S}_{zz}$$

$$R = \frac{\tilde{S}_{xx} - \tilde{S}_{yy}}{A}$$

when $\tilde{S}_{ii}$ are the components of the alignment tensor in the frame in which it is diagonal.

In the case in which the alignment of the molecule is induced by magnetic susceptibility anisotropy, the alignment is dictated by the anisotropy of the magnetic susceptibility tensor $\boldsymbol{\chi}$ and the tensor components are expressed as follows:

$$\tilde{S}_{ii} = \frac{2}{3}\frac{B_0^2}{15\mu_0 kT}(\tilde{\chi}_{ii} - \bar{\chi})$$

where $\tilde{S}_{ii}$ are the principal components of the alignment tensor $\boldsymbol{S}$, $\tilde{\chi}_{ii}$ are the principal components of the magnetic susceptibility tensor $\boldsymbol{\chi}$, and $\bar{\boldsymbol{\chi}}$ is the trace of the tensor $\boldsymbol{\chi}$

divided by three. The subtraction is required as we consider only the anisotropic component of the tensor.

Constrained optimization in REFMAC-NMR

In the previous version of REFMAC-NMR[2], the alignment tensor $\boldsymbol{S}$ describing the RDCs was determined through a Gauss-Newton optimization approach of 5 tensor components S_{zz}, S_{xx}🎗S_{yy}, S_{xy}, S_{xz}, S_{yz}. Using these components as fitting parameters, constraints between tensors referring to different structural units could not be applied, neither in terms of magnitude nor in terms of orientation (see Section "Tensor orientation in rigid multisubunit systems"). We have thus re-parameterized the problem to allow the inclusion of such constraints.

Instead of using the 5 tensor components given above, the tensor $\boldsymbol{S}$ is reconstructed as follows: the orientation is provided by three variables describing the rotation that brings the tensor $\boldsymbol{S}$ from any arbitrary molecular frame to the frame where it is diagonal, and the magnitude is determined by two of the diagonal elements, $\tilde{S}_{xx}$ and $\tilde{S}_{yy}$ ($\tilde{S}_{zz} = -\tilde{S}_{xx} - \tilde{S}_{yy}$, as the tensor is defined as traceless). For a convenient sampling of the orientational space and to simplify the handling of the derivatives, the rotation is expressed in terms of quaternions:

$$\boldsymbol{q} = q_0 + q_1 + q_2 + q_3 = \cos\frac{\theta}{2} + (v_1\boldsymbol{i} + v_2\boldsymbol{j} + v_3\boldsymbol{z})\sin\frac{\theta}{2}$$

where $\boldsymbol{i}$, $\boldsymbol{j}$, $\boldsymbol{k}$ are unit vectors representing the three Cartesian axes, (v_1, v_2, v_3) are the components of the unit vector defining the axis of rotation, and θ is the angle of rotation. To represent a rotation, quaternions must satisfy the constraint $|\boldsymbol{q}| = 1$ (unit quaternions). This reduces the number of independent variables describing the rotation matrix to three. Thus, the new parameters used in the calculations are the following:

$$p = [q_0, q_1, q_2, q_3, \tilde{S}_{xx}, \tilde{S}_{yy}] \ with \ |\boldsymbol{q}| = 1 \ .$$

The minimization performed by REFMAC for the structure refinement requires the computation of the first and second derivatives with respect to the selected parameters of the target function f to be minimized, reported *Appendix A.1*.

In the presence of two structural units, two independent tensors can be defined, $\boldsymbol{S}_a$ and $\boldsymbol{S}_b$, depending on the two sets of the parameters p_a and p_b defined as:

$$p_a = [q_{a,i}, \tilde{S}_{a,ii}], \quad p_b = [q_{b,i}, \tilde{S}_{b,ii}] \ \ with \ i = 0, 1, 2, 3 \ and \ ii = xx, yy$$

The function f is thus the sum of the functions referring to each domain:

$$f(p') = f(p_a, p_b) = f_a(p_a) + f_b(p_b) = f_a(q_{a,i}, \tilde{S}_{a,ii}) + f_b(q_{b,i}, \tilde{S}_{b,ii})$$

where

$$p' = [p_a, p_b] = [q_{a,i}, \tilde{S}_{a,ii}, q_{b,i}, \tilde{S}_{b,ii}]$$

The first and second derivatives of f are given by the derivatives of f_a with respect to the p_a parameters and of f_b with respect to the p_b parameters, and are reported in the *Appendix A.2*.

In order to constrain the orientation of the two tensors to be the same ($q_{a,i} = q_{b,i}$ for i=0, 1, 2, 3), parameters and target function can be re-expressed in the following way:

$$p' = [q_i, \tilde{S}_{a,ii}, \tilde{S}_{b,ii}] \qquad f(p') = f_a(q_i, \tilde{S}_{a,ii}) + f_b(q_i, \tilde{S}_{b,ii})$$

where the non-null elements of the first and second derivatives of f for the present case are reported in the *Appendix A.2*.

In order to constrain the anisotropy values of the two tensors to be the same ($\tilde{S}_{a,ii} = \tilde{S}_{b,ii}$ for *ii*=*xx*, *yy*), parameters and target function can be re-expressed in the following way:

$$p' = \left[q_{a,i}, \tilde{S}_{ii}, q_{b,i}\right] \qquad f(p') = f_a\left(q_{a,i}, \tilde{S}_{ii}\right) + f_b\left(q_{b,i}, \tilde{S}_{ii}\right)$$

where the non-null elements of the first and second derivatives of f are given in the *Appendix A.2*.

The constraints described above can be applied to any generic set of tensors, indicated in form of list of pairs in the standard instruction file provided to REFMAC-NMR. The algorithm in REFMAC-NMR can now identify the tensors to be constrained (in orientation, anisotropy values, or both) according to user's instructions, and perform the appropriate minimizations.

In the presence of symmetric assemblies, the alignment tensor assumes particular properties strictly related to the symmetry of the system (automatically detected by the program) and that are taken into account during the calculations. Based on the symmetry of the system, the program automatically sets the constraints to be applied in the tensor calculation. These constraints are then "translated" in terms of additional contributions to the first and second derivatives of the target function for the re-parameterized problem.

In the presence of a cyclic 2-fold (C_2) symmetry, one of the tensor axes must be directed along the symmetry axis of the system. The directions of the two other axes can thus be determined by a rotation about the C_2 axis, so that in this case only three parameters $\left[q_0, \tilde{S}_{xx}, \tilde{S}_{yy}\right]$ need to be determined. The first and second derivate with the respect of q_0 are reported in the *Appendix A.3*.

In case a dihedral 2-fold (D_2) symmetry has been identified, two axes of the tensor must be directed along the two symmetry axes of the system, thus defining completely its orientation. This further reduces the number of parameters that need to be estimated during the calculation of the tensor to only two $[\tilde{S}_{xx}, \tilde{S}_{yy}]$.

In the remaining cases in which a cyclic or dihedral symmetry has been identified with higher multiplicity (C_n or D_n), one axis must be directed along the cyclic n-fold symmetry axis of the system, and the tensor must be assumed to be completely axial. Assuming $\tilde{S}_{zz}$ as the eigenvalue associated to the eigenvector directed along the symmetry axis, the last condition implies the other two eigenvalues $\tilde{S}_{xx}$ and $\tilde{S}_{yy}$ to be equal, making the tensor invariant with the respect to the direction of their associated eigenvectors. Therefore, since one axis of the tensor is defined by the symmetry of the system and the other two can be arbitrarily chosen, the only parameter that needs to be estimated is $\tilde{S}_{xx}$ ($\tilde{S}_{zz} = -2\tilde{S}_{xx}$), with further simplification of the problem.

The case of the proteasome

The PDB structure of the 20S proteasome at the highest resolution, 1YAR (1.9 Å resolution, D9S mutation) was refined against X-ray data together with available NMR data using the new version of REFMAC-NMR. The available NMR restraints were the ^{1}H-^{13}C RDCs, measured with a methyl-TROSY experiment on the Ile-δ_1-[$^{13}CH_3$], Leu, Val-[$^{13}CH_3$, $^{12}CD_3$]-labeled side chains of the α7α7 rings of the 20S proteasome (360 kDa), also referred to as “half-proteasome”. RDCs for the methyl groups were collected in presence of about 40mg/mL Pf1 phages.[3] Mobile residues were identified from relaxation measurements of both ^{2}H and ^{13}C spins[4], providing the order parameters S_{LS}^2, and treated accordingly. Indeed, RDCs contain information on both protein structure and dynamics. In the limit of axially symmetric internal motion about a mean axis, the measured RDC can be expressed as:

$$RDC_{measured} = RDC_{static} \times S_{LS} \qquad (1)$$

where RDC_{static} is the coupling that would be obtained for a bond along the time-averaged orientation in the absence of internal protein dynamics, and S_{LS} is the reduction in the size of the observed coupling due to internal motion. For simplicity, we assume that the S_{LS} determined by relaxation measurements (affected by mobility in the time scale up to ns) coincides with that of RDCs (which are affected by mobility in a much larger time scale, up to ms). It is important to remark that for structured parts of the protein backbone, these motions are usually small ($S_{LS} \approx 0.9$), and the dynamic component of the ^{1}H-^{15}N RDC is often assumed to be uniform. Side chains, on the other hand, are often differentially mobile, with dynamics that can be asymmetric, leading to complications in the interpretation of RDCs in terms of structure.[5]

Structural variability in the X-ray structure

Before proceeding with the combination of the different available experimental data, we have analyzed the structural variability among the chains as present in the PDB deposition. As expected, while the backbone conformation is mostly preserved in the different chains of the X-ray structure, the conformations of side chains may differ significantly (**Fig. 1**).

RDC agreement against X-ray and REFMAC-NMR structures

Residues resulting as mobile from relaxation measurements are expected to experience multiple side chain conformations. The fitting of experimental data was performed by fixing the direction of one axis of the ***S*** tensor to the main cyclic symmetry axis (**Fig. 2**), and by imposing it to be axial. The agreement between the X-ray structure and the observed methyl- RDCs, as obtained through the best fit of ***S*** tensor, was thus monitored by excluding these mobile (S_{LS}<0.77) residues. The remaining RDCs (with S_{LS}>0.77) are in a quite remarkable agreement with the X-ray structure, as shown by comparing the experimental data with the data back-calculated from the best fit ***S*** tensor (**Fig. 3**, left and **Table 1**). It is interesting to observe that the variability in the data back-calculated for the different chains of the X-ray structure is much larger among the residues identified as mobile than among those identified as rigid.

We then performed a joint refinement of the structure against both X-ray and RDC data. The S_{LS} for each residue was considered in the evaluation of methyl- RDCs as indicated by Eq. (1), i.e. assuming an axially symmetric motion within a cone. Due to the weakness of this assumption for mobile residues, the RDCs of the latter were given a half weight with the respect to those of the residues classified as rigid. As the structure has a D_7 symmetry, the ***S*** tensor was constrained to be axial, with the z-axis pointing along the axis of higher symmetry (**Fig. 2** - this choice is encoded in the software and does not require any a-priori user intervention). The agreement between experimental and back-calculated RDCs against the protein structure shows a significant improvement after the joint refinement (**Fig. 3** and **Table 1**). Although the relevant improvement in the agreement of the X-ray structure with the NMR data, some significant discrepancies between experimental and back-calculated data can be observed, interestingly, clustered at the interface between the individual units constituting the α7 ring (**Fig. 4**). The chain showing the best agreement with the NMR data is shown in detail in **Fig. 5**.

The side chains experiencing the largest structural variation after joint refinement undergo a conformational change as illustrated in **Fig. 6**, which is a non-trivial rearrangement of the position of a few atoms. The individual side chains undergoing the largest structural variations after the refinement are shown in detail in **Fig. 7**.

Using tensor restraints in other systems

Restraints on the tensors reciprocal orientation and/or magnitude can also be imposed by the user. We here describe a particular case that is representative of a situation that is easily encountered in real life: a rigid multisubunit system, where RDCs have been measured in different samples to reduce spectral complexity.

Tensor orientation in rigid multisubunit systems

Larger systems are hard to be approached by NMR, not only because size-dependent nuclear relaxation mechanisms may broaden resonances beyond detection, but also because spectral complexity is increased. In the case of multisubunit systems, where the different subunits can be separately expressed recombinantly and subsequently reconstituted in vitro, it is often the method of choice to label the individual subunits. This approach largely simplifies the NMR characterization, but it might introduce slight perturbation when recording RDCs with external alignment media: since the RDCs are collected in different samples, the concentration of the alignment medium may vary from one sample to the other, and this might yield impression of protein mobility. We have recently encountered such a situation[6] and here we re-examine the refinement results by imposing the tensors from the individual subunits (Sxl and CSD1) to be equally oriented. We found that the inclusion of this orientation constraints, not only permit a safest approach for the identification of scaling factors between different experimental conditions, but it also contributes positively to the refinement stabilizing the calculation (R_{free} is smaller when the constraints are applied). Refinements of the individual Sxl and CSD1 domains performed before and after the inclusion of the orientation constraints between the tensors are compared in terms of REFMAC-NMR output and tensor comparisons (**Tab. 2-3**).

Conclusions

We have implemented a new set of features in REFMAC-NMR that encompasses most of the needs that are encountered when analyzing NMR data, with particular reference to RDCs, in larger systems. Among those, we have implemented symmetry-based restraints that allowed for refinement of the 20S proteasome structure.

List of figures and tables

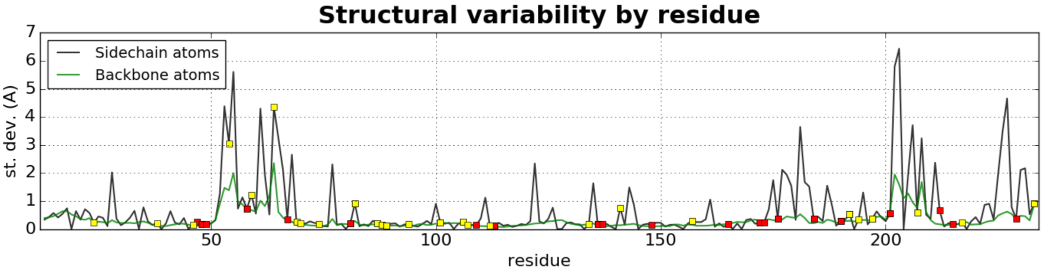


Figure 1: Structural variability among the individual chains in the X-ray structure (PDB: 1YAR). For each residue, the standard deviation among the different chains is reported for the backbone atoms (blue line), for the side chains (black line). Residues for which an RDC has been measured are indicated by red (S_{LS}>0.77) and yellow (S_{LS}<=0.77) circles.

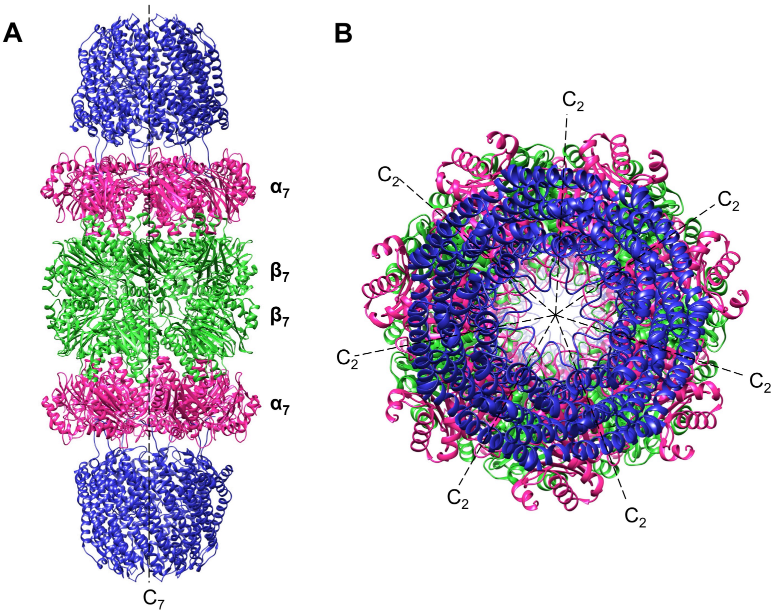


Figure 2: X-ray structure of the 20S proteasome (PDB: 1YAR). (A) Side view showing the axis of cyclic symmetry (C_7). (B) Top view showing the seven axis of dihedral symmetry (C_2).

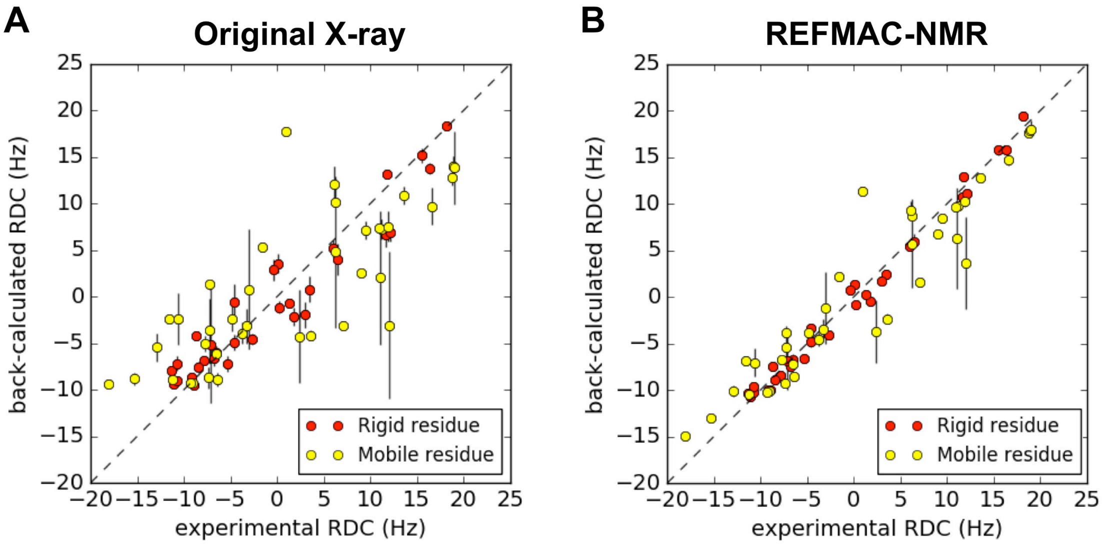


Figure 3: Agreement between experimental and back-calculated RDC for (A) the original X-ray (PDB: 1YAR) and (B) the REFMAC-NMR refined structures. RDC measured for rigid (S_{LS}>0.77) and mobile (S_{LS}<=0.77) residues are indicated, respectively, by red and yellow dots. Variability for the fit obtained from the different chains (A, B, C, D, E, F, G) are indicated by the error bar.

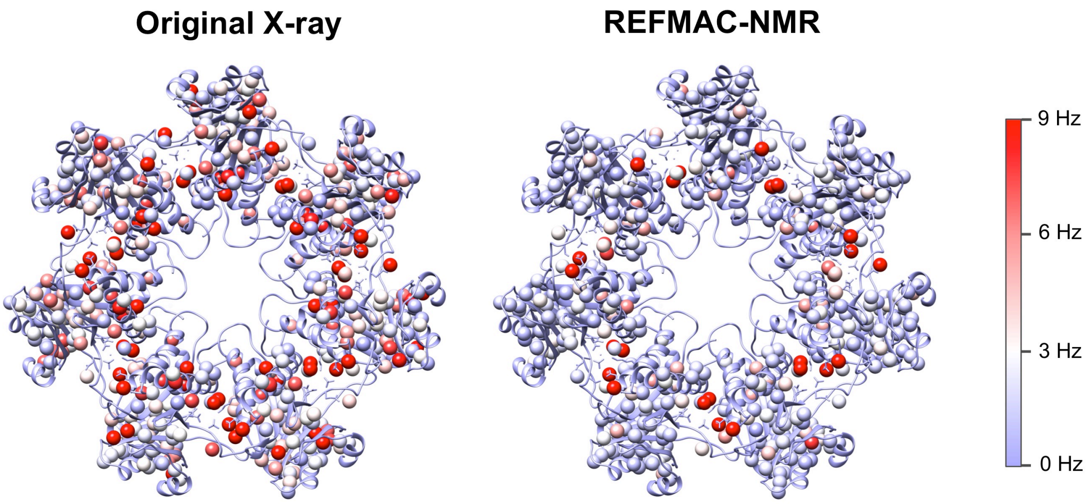


Figure 4: Agreement of predicted RDCs for the original X-ray (PDB: 1YAR) and REFMAC-NMR refined structures. Differences between experimental and predicted RDCs are reported in absolute value and colored according to their discrepancy.

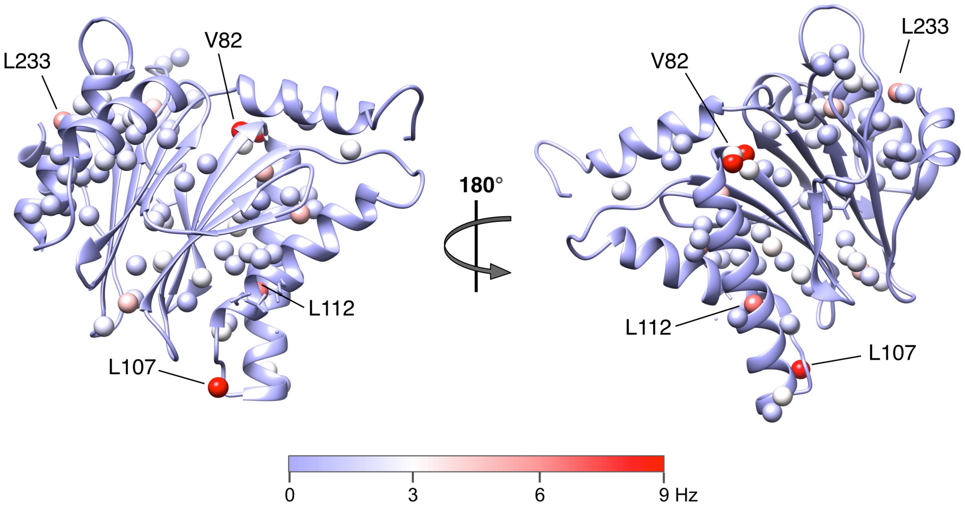


Figure 5: Agreement of predicted RDCs for REFMAC-NMR refined structure. Only the chain in the best agreement is shown (chain A). Differences between experimental and predicted RDCs are reported in absolute value and colored according to their discrepancy.

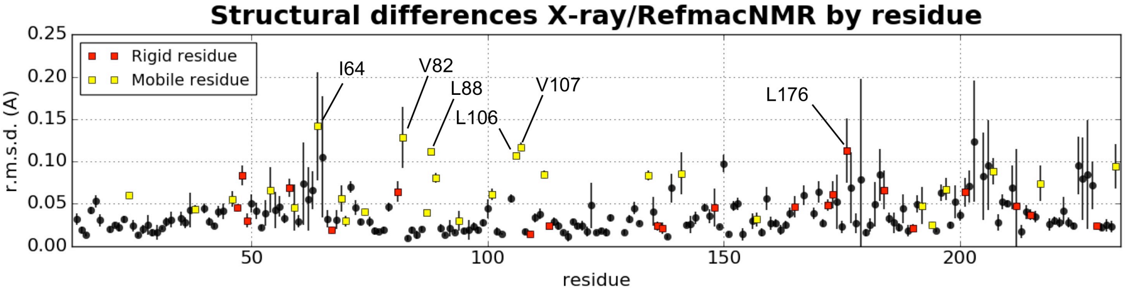


Figure 6: Structural differences after REFMAC-NMR refinement. The root mean square deviation (rmsd) between structures refined using REFMAC before and after the inclusion of the NMR data is reported in terms of mean and standard deviations of the values obtained from the individual chains. RDC measured for rigid (S>0.77) and mobile (S<=0.77) residues are indicated, respectively, by red and yellow dots.

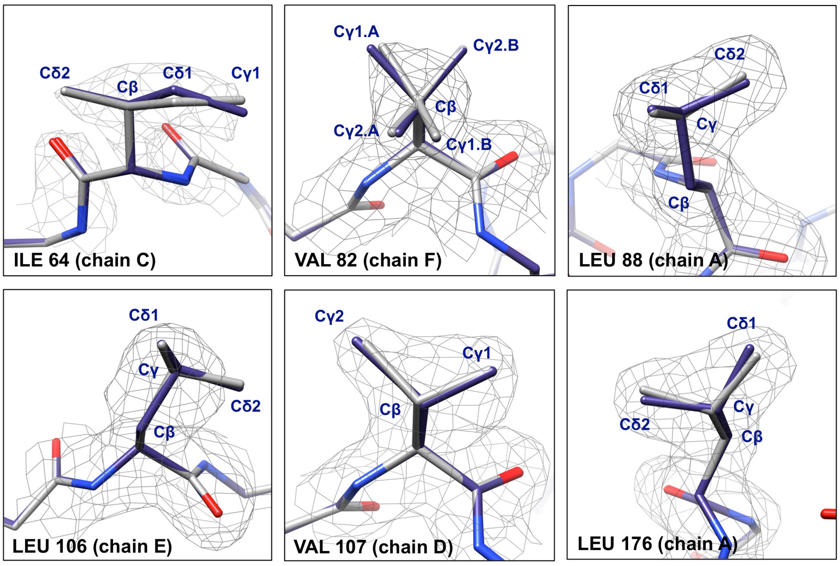


Figure 7: Detail of the side chains undergoing the largest conformational variations. Original and refined structures are colored, respectively, in white and in blue. Electron density map is plotted at 1σ threshold.

Chain	**X-ray**		**REFMAC-NMR**	
	Rigid residues	All residues	Rigid residues	All residues
A	0.309	0.540	0.123	0.283
B	0.287	0.506	0.111	0.267
C	0.286	0.629	0.113	0.318
D	0.471	0.600	0.192	0.315
E	0.312	0.586	0.117	0.312
F	0.328	0.573	0.120	0.328
G	0.349	0.585	0.130	0.296

Table 1: Q-factors calculated for the X-ray and the REFMAC-NMR refined structures (PDB: 1YAR). Q-factor is reported for the individual chains for both only rigid or all the residues.

PDB code: **4QQB** – Resolution: **2.80 Å**		
Parameters	**Unconstrained minimization**	**Constrained minimization**
R-value	0.199	0.199
R-free	**0.237**	**0.236**
RMSD bond length	0.010	0.010
RMSD bond angles	1.663	1.673
RMSD chiral volume	0.108	0.107

Table 2: REFMAC-NMR output for the independent refinement of Sxl and CSD1 domains, before and after the inclusion of orientation constraints for the tensors calculated for the individual units. Refinement are performed with REFAMC 5.9.000 version

Unconstrained minimization			
	Q-factor RDCs Q_{Sxl}, Q_{CSD1}	Axial component $\frac{A_{Sxl}}{A_{CSD1}}$	Orientation and shape $\frac{\boldsymbol{S}_{Sxl} \cdot \boldsymbol{S}_{CSD1}}{\|\boldsymbol{S}_{Sxl}\| \, \|\boldsymbol{S}_{CSD1}\|}$
Chain A, X	0.147,0.059	0.761	0.980
Chain B, Y	0.135, 0.062	0.759	0.989
Constrained minimization			
	Q-factor RDCs Q_{Sxl}, Q_{CSD1}	Axial component $\frac{A_{Sxl}}{A_{CSD1}}$	Orientation and shape $\frac{\boldsymbol{S}_{Sxl} \cdot \boldsymbol{S}_{CSD1}}{\|\boldsymbol{S}_{Sxl}\| \, \|\boldsymbol{S}_{CSD1}\|}$
Chains A, X	0.151, 0.071	0.751	0.997
Chains B, Y	0.143, 0.074	0.753	0.996

Table 3: Comparison of the tensors calculated independently for refined structured of Sxl and CSD1 before and after the inclusion of orientation constraints, in terms axial components and orientation/shape. Q-factors for the individual Sxl and CSD1 domains are also reported.

Appendix

A.1 Problem re-parameterization

To allow the inclusion of the discussed constraints, it results more convenient to re-express the problem in terms of eigenvectors and eigenvalues of the alignment tensor ***S***. Eigenvectors e_x, e_y, e_z and associated eigenvalues $\tilde{S}_{xx}$, $\tilde{S}_{yy}$, $\tilde{S}_{zz}$ encode, respectively, the information about the orientation and the magnitude/anisotropy of the tensor ***S***, and are related to the tensor matrix by the simple relationship:

$$\boldsymbol{S} = R\,A\,R^{-1} = \begin{pmatrix}\boldsymbol{e}_x & \boldsymbol{e}_y & \boldsymbol{e}_z\end{pmatrix}\begin{pmatrix}\tilde{S}_{xx} & 0 & 0\\ 0 & \tilde{S}_{yy} & 0\\ 0 & 0 & \tilde{S}_{zz}\end{pmatrix}\begin{pmatrix}\boldsymbol{e}_x & \boldsymbol{e}_y & \boldsymbol{e}_z\end{pmatrix}^{-1}$$

Eigenvectors basis constitutes the rotation matrix R that transforms the tensor from the molecular frame to its principal axis frame. This allows expressing each element S_{ij} of the alignment tensor ***S*** in terms of its principal components $\tilde{S}_{ii}$ and of element of the rotation matrix R as following:

$$\boldsymbol{S} = R\,A\,R^{T} = \begin{pmatrix}r_{11} & r_{12} & r_{13}\\ r_{21} & r_{22} & r_{23}\\ r_{31} & r_{32} & r_{33}\end{pmatrix}\begin{pmatrix}\tilde{S}_{xx} & 0 & 0\\ 0 & \tilde{S}_{yy} & 0\\ 0 & 0 & \tilde{S}_{zz}\end{pmatrix}\begin{pmatrix}r_{11} & r_{12} & r_{13}\\ r_{21} & r_{22} & r_{23}\\ r_{31} & r_{32} & r_{33}\end{pmatrix}^{T}$$

$$S_{ij} = \sum_{k=1}^{3} r_{ik}\tilde{S}_{kk}r_{kj} \ .$$

Each element of the rotation matrix R is then more conveniently expressed in terms of quaternions, permitting to reduce the number of parameters describing the rotation from nine to four.

Given f the target function to be minimized, x the set of parameters for the original problem, defined as:

$$x = \left[\mathrm{S}_{zz}, \mathrm{S}_{xx} - \mathrm{S}_{yy}, S_{xy}, S_{xz}, S_{yz}\right] \ .$$

and p the set of parameters for the re-parameterized problem, defined in this particular case as:

$$p = \left[q_0, q_1, q_2, q_3, \tilde{S}_{xx}, \tilde{S}_{yy}\right]$$

the first and second derivatives of the re-parameterized problem can then be directly calculated starting from the derivatives of the original problem by application of the chain rule, as following:

$$\frac{\partial f}{\partial p_j} = \sum_i \frac{\partial f}{\partial x_i}\frac{\partial x_i}{\partial p_j}$$

$$\frac{\partial^2 f}{\partial p_i \partial p_j} = \sum_{k,l} \frac{\partial x_k}{\partial p_i}\frac{\partial^2 f}{\partial x_k \partial x_l}\frac{\partial x_k}{\partial p_i} + \sum_k \frac{\partial f}{\partial x_k}\frac{\partial^2 x_k}{\partial p_i \partial p_j}$$

where the last term of the second derivative has not be considered since it involves the use of the first derivative with the respect of the original problem making the optimization unstable.

A.2 Imposing anisotropy and orientation constraints between tensors

In the presence of two structural units, two independent tensors can be defined, ***S**ₐ* and ***S**_b*, depending on the two sets of the parameters p_a and p_b defined as:

$$p_a = \left[q_{a,i}, \tilde{S}_{a,ii}\right], \quad p_b = \left[q_{b,i}, \tilde{S}_{b,ii}\right] \ \ with\ i = 0, 1, 2, 3\ and\ ii = xx, yy$$

The function f is thus the sum of the functions referring to each domain:

$$f(p') = f(p_a, p_b) = f_a(p_a) + f_b(p_b) = f_a\left(q_{a,i}, \tilde{S}_{a,ii}\right) + f_b\left(q_{b,i}, \tilde{S}_{b,ii}\right)$$

where

$$p' = [p_a, p_b] = \left[q_{a,i}, \tilde{S}_{a,ii}, q_{b,i}, \tilde{S}_{b,ii}\right]$$

The first and second derivatives of f are given by the derivatives of f_a with respect to the p_a parameters and of f_b with respect to the p_b parameters. The derivatives are reported, for simplicity, in the following matrix form where only the non-null elements are shown:

$$\frac{\partial f}{\partial p'} = \left[\frac{\partial f_a}{\partial q_{a,i}}; \frac{\partial f_a}{\partial \tilde{S}_{a,ii}}; \frac{\partial f_b}{\partial q_{b,i}}; \frac{\partial f_b}{\partial \tilde{S}_{b,ii}}\right]$$

$$\frac{\partial^2 f}{\partial p'^2} = \begin{bmatrix} \frac{\partial^2 f_a}{\partial q_{a,i} \partial q_{a,j}} & \frac{\partial^2 f_a}{\partial q_{a,i} \partial \tilde{S}_{a,jj}} & & \\ \frac{\partial^2 f_a}{\partial q_{a,i} \partial \tilde{S}_{a,jj}} & \frac{\partial^2 f_a}{\partial \tilde{S}_{a,ii} \partial \tilde{S}_{a,jj}} & & \\ & & \frac{\partial^2 f_b}{\partial q_{b,i} \partial q_{b,j}} & \frac{\partial^2 f_b}{\partial q_{b,i} \partial \tilde{S}_{b,jj}} \\ & & \frac{\partial^2 f_b}{\partial q_{b,i} \partial \tilde{S}_{b,jj}} & \frac{\partial^2 f_b}{\partial \tilde{S}_{b,ii} \partial \tilde{S}_{b,jj}} \end{bmatrix}$$

In order to constrain the orientation of the two tensors to be the same ($q_{a,i} = q_{b,i}$ for i=0, 1, 2, 3), parameters and target function are expressed in the following way:

$$p' = \left[q_i, \tilde{S}_{a,ii}, \tilde{S}_{b,ii}\right] \qquad f(p') = f_a\left(q_i, \tilde{S}_{a,ii}\right) + f_b\left(q_i, \tilde{S}_{b,ii}\right)$$

where the non-null elements of the first and second derivatives of f are given as following:

$$\frac{\partial f}{\partial p'} = \left[\frac{\partial f_a}{\partial q_i} + \frac{\partial f_b}{\partial q_i}; \frac{\partial f_a}{\partial \tilde{S}_{a,ii}}; \frac{\partial f_b}{\partial \tilde{S}_{b,ii}}\right]$$

$$\frac{\partial^2 f}{\partial p'^2} = \begin{bmatrix} \frac{\partial^2 f_a}{\partial q_i \partial q_j} + \frac{\partial^2 f_b}{\partial q_i \partial q_j} & \frac{\partial^2 f_a}{\partial q_i \partial \tilde{S}_{a,jj}} & \frac{\partial^2 f_b}{\partial q_i \partial \tilde{S}_{b,jj}} \\ \frac{\partial^2 f_a}{\partial q_i \partial \tilde{S}_{a,jj}} & \frac{\partial^2 f_a}{\partial \tilde{S}_{a,ii} \partial \tilde{S}_{a,jj}} & \\ \frac{\partial^2 f_b}{\partial q_i \partial \tilde{S}_{b,jj}} & & \frac{\partial^2 f_b}{\partial \tilde{S}_{b,ii} \partial \tilde{S}_{b,jj}} \end{bmatrix}$$

In order to constrain the anisotropy values of the two tensors to be the same ($\tilde{S}_{a,ii} = \tilde{S}_{b,ii}$ for *ii*=xx, yy), parameters and target function can be re-expressed in the following way:

$$p' = \left[q_{a,i}, \tilde{S}_{ii}, q_{b,i}\right] \qquad f(p') = f_a\left(q_{a,i}, \tilde{S}_{ii}\right) + f_b\left(q_{b,i}, \tilde{S}_{ii}\right)$$

where the non-null elements of the first and second derivatives of f are given as following:

$$\frac{\partial f}{\partial p'} = \left[\frac{\partial f_a}{\partial q_{a,i}}; \frac{\partial f_a}{\partial \tilde{S}_{ii}} + \frac{\partial f_b}{\partial \tilde{S}_{ii}}; \frac{\partial f_b}{\partial q_{b,i}}\right]$$

$$\frac{\partial^2 f}{\partial p'^2} = \begin{bmatrix} \dfrac{\partial^2 f_a}{\partial q_{a,i} \partial q_{a,j}} & \dfrac{\partial^2 f_a}{\partial q_{a,i} \partial \tilde{S}_{jj}} & \\ \dfrac{\partial^2 f_a}{\partial q_{a,i} \partial \tilde{S}_{jj}} & \dfrac{\partial^2 f_a}{\partial \tilde{S}_{,i} \partial \tilde{S}_{jj}} + \dfrac{\partial^2 f_b}{\partial \tilde{S}_{ii} \partial \tilde{S}_{jj}} & \dfrac{\partial^2 f_b}{\partial q_{b,i} \partial \tilde{S}_{jj}} \\ & \dfrac{\partial^2 f_b}{\partial q_{b,i} \partial \tilde{S}_{jj}} & \dfrac{\partial^2 f_b}{\partial q_{b,i} \partial q_{b,j}} \end{bmatrix}$$

This approach can be used to perform the minimization of a generic number of tensors, constrained to each other by their anisotropy, orientation or both.

A.3 Imposing symmetry-related constraints to tensor parameters

In case a cyclic 2-fold (*C2*) symmetry has been identified, one axis of the tensor must be directed along the symmetry axis of the system. To include this constraint in the calculation of the tensor, the interpretation of quaternion in term of axis-angle representation can be exploited. Indeed, according to the definition of quaternion, the unit vector (v_1, v_2, v_3) multiplying the three imaginary parts ***i***, ***j***, ***k*** define the axis of rotation. In this way, the tensor calculation is reduced to the estimation of only three parameters $\left[q_0, \tilde{\chi}_{xx}, \tilde{\chi}_{yy}\right]$, instead of initial six parameters. Taking into account the following relationship:

$$q_i = v_i\sqrt{1 - q_0^2} \qquad i = 1 \dots 3$$

the first derivative of the target function *f'* of the constrained problem with the respect to q_0 is given by:

$$\frac{\partial f'}{\partial q_0} = \frac{\partial f}{\partial q_0} + \sum_{i=1}^{3} \frac{\partial f}{\partial q_i}\frac{\partial q_i}{\partial q_0} = \frac{\partial f}{\partial q_0} - \frac{q_0}{\sqrt{1 - q_0^2}} \sum_{i=1}^{3} v_i \frac{\partial f}{\partial q_i}$$

and the auto- and cross- terms constituting the Hessian matrix of the second derivatives are, respectively:

$$\frac{\partial^2 f'}{\partial q_0^2} = \frac{\partial^2 f}{\partial q_0^2} + \left(\frac{q_0}{\sqrt{1 - q_0^2}}\right)^2 \sum_{i=1}^{3}\sum_{j=1}^{3} v_i \frac{\partial f}{\partial q_i} v_j \frac{\partial f}{\partial q_j} - \frac{1}{(1 - q_0^2)^{\frac{3}{2}}} \sum_{i=1}^{3} v_i \frac{\partial f}{\partial q_i}$$

$$- 2\frac{q_0}{\sqrt{1 - q_0^2}} \sum_{i=1}^{3} v_i \frac{\partial^2 f}{\partial q_0 \partial q_i}$$

$$\frac{\partial^2 f'}{\partial q_0 \partial \tilde{\chi}_{ii}} = \frac{\partial^2 f}{\partial q_0 \partial \tilde{\chi}_{ii}} - \frac{q_0}{\sqrt{1 - q_0^2}} \sum_{i=1}^{3} v_i \frac{\partial^2 f}{\partial q_i \partial \tilde{\chi}_{ii}} \quad .$$

References

(1) Bertini, I.; Luchinat, C.; Parigi, G. *Progress in Nuclear Magnetic Resonance Spectroscopy* **2002**, *40*, 249.

(2) Rinaldelli, M.; Ravera, E.; Calderone, V.; Parigi, G.; Murshudov, G. N.; Luchinat, C. *Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr* **2014**, *70* (4), 958.

(3) Sprangers, R.; Kay, L. E. *J. Am. Chem. Soc.* **2007**, *129* (42), 12668.

(4) Sprangers, R.; Kay, L. E. *Nature* **2007**, *445* (7128), 618.

(5) Ottiger, M.; Delaglio, F.; Bax, A. *Journal of Magnetic Resonance* **1998**, *131* (2), 373.

(6) Carlon, A.; Ravera, E.; Hennig, J.; Parigi, G.; Sattler, M.; Luchinat, C. *Journal of the American Chemical Society* **2016**, *138* (5), 1601.

3.3 How to tackle protein structural data from solution and solid state: An integrated approach

Azzurra Carlon[a], *Enrico Ravera*[a,b], *Witold Andrałojć*[a], *Giacomo Parigi*[a,b], *Garib N. Murshudov*[c], *Claudio Luchinat*[a,b]

[a] Center for Magnetic Resonance (CERM), University of Florence,
via Sacconi 6, 50019 Sesto Fiorentino (FI), Italy
[b] Department of Chemistry, University of Florence,
via della Lastruccia 3, 50019 Sesto Fiorentino (FI), Italy
[c] MRC Laboratory of Molecular Biology,
Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge, CB2 0QH UK

Published in:

Progress in nuclear magnetic resonance spectroscopy ***92****, 54-70*

Progress in Nuclear Magnetic Resonance Spectroscopy 92–93 (2016) 54–70

Contents lists available at ScienceDirect

Progress in Nuclear Magnetic Resonance Spectroscopy

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

How to tackle protein structural data from solution and solid state: An integrated approach

Azzurra Carlon[a], Enrico Ravera[a], Witold Andrałojć[a], Giacomo Parigi[a], Garib N. Murshudov[b], Claudio Luchinat[a],*

[a] Magnetic Resonance Center (CERM) and Department of Chemistry “Ugo Schiff”, University of Florence, Italy[1]
[b] MRC Laboratory for Molecular Biology, Francis Crick Ave, Cambridge CB2 0QH, UK

Edited by David Neuhaus and Gareth Morris

ARTICLE INFO

Article history:
Received 15 September 2015
Accepted 13 January 2016
Available online 21 January 2016

Keywords:
Paramagnetic restraints
Residual dipolar couplings
Pseudo-contact shifts
Structural refinement
Integrated structural biology

ABSTRACT

Long-range NMR restraints, such as diamagnetic residual dipolar couplings and paramagnetic data, can be used to determine 3D structures of macromolecules. They are also used to monitor, and potentially to improve, the accuracy of a macromolecular structure in solution by validating or “correcting” a crystal model. Since crystal structures suffer from crystal packing forces they may not be accurate models for the macromolecular structures in solution. However, the presence of real differences should be tested for by simultaneous refinement of the structure using both crystal and solution NMR data. To achieve this, the program REFMAC5 from CCP4 was modified to allow the simultaneous use of X-ray crystallographic and paramagnetic NMR data and/or diamagnetic residual dipolar couplings. Inconsistencies between crystal structures and solution NMR data, if any, may be due either to structural rearrangements occurring on passing from the solution to solid state, or to a greater degree of conformational heterogeneity in solution with respect to the crystal. In the case of multidomain proteins, paramagnetic restraints can provide the correct mutual orientations and positions of domains in solution, as well as information on the conformational variability experienced by the macromolecule.

Contents

1. Introduction 55
2. Combining X-ray and solution NMR data. 56
  2.1. General aspects 56
  2.2. Structure refinement using REFMAC-NMR 57
  2.3. Tensor calculation 58
  2.4. Motionally averaged data 58
  2.5. A step-by-step approach for the refinement of multidomain systems 59
    2.5.1. Check 1: Are tensor magnitudes equal? 59
    2.5.2. Check 2: Are tensor orientations equal? 59
    2.5.3. Check 3: Is the full-length protein refinement acceptable?. 60
  2.6. Averaged tensors from diamagnetic and paramagnetic RDCs 61
3. Case studies 62
  3.1. Single domain proteins 62
    3.1.1. The catalytic domain of matrix metalloproteinase 1 (MMP1). 62
    3.1.2. Ubiquitin 63

---

* Corresponding author at: Via L. Sacconi 6, 50019 Sesto Fiorentino (FI), Italy. Tel.: +39 0554574296; fax: +39 0554574924.
E-mail addresses: carlon@cerm.unifi.it (A. Carlon), ravera@cerm.unifi.it (E. Ravera), andralojc@cerm.unifi.it (W. Andrałojć), parigi@cerm.unifi.it (G. Parigi), garib@mrc-lmb.cam.ac.uk (G.N. Murshudov), claudioluchinat@cerm.unifi.it (C. Luchinat).

[1] URL: http://www.cerm.unifi.it/.

http://dx.doi.org/10.1016/j.pnmrs.2016.01.001

3.1.3. IgG-binding domain of protein G (GB3) 63
3.1.4. The N-terminal domain of calmodulin (CaM). 63
3.2. Multi-domain proteins and complexes 64
3.2.1. Full length matrix metalloproteinase 1 (MMP1) 64
3.2.2. Full length CaM 66
3.2.3. Calmodulin–IQ peptide complex (CaM–IQ). 66
3.2.4. Calmodulin–DAPK peptide complex (CaM–DAPKp) 67
3.2.5. Sxl-Unr translation complex 68
4. Conclusions. 68
Acknowledgements 68
References 68

1. Introduction

The most widely used techniques for elucidation of molecular structures at atomic resolution are X-ray crystallography and NMR spectroscopy, accounting as of July 2015 for 90% and 9.4% of all deposited protein structures, respectively, and 58% and 41% of all deposited nucleic acid structures. Besides new experimental techniques gaining more and more resounding success (e.g. cryo-EM, with more than 390 new entries in the last 3 years and resolution down to 2.2 Å [1]), X-ray and NMR still play a key role in answering many unresolved questions in the structural biology field. The unique importance of the integration of these two techniques has been recognized, taking advantage of the distinctive features of each. The strength of X-ray crystallography lies in a precise determination of a unique macromolecular structure (or a unique set of different structures present in the asymmetric unit of the crystal), whereas NMR spectroscopy has the power to probe the dynamics experienced in solution. On the other hand, neither X-ray nor NMR, if used as stand-alone tools, can provide a complete, precise and accurate picture of the biological system under investigation and of its interactions with other complexes or biomolecules. It also should be mentioned that X-ray crystallographic diffraction and NMR data are intrinsically different – the former gives information that progresses from the overall shape of the molecule up to individual atom positions as the resolution increases, whereas NMR provides immediate information about short-range inter-atom distances and bond orientations, which progresses to overall shape of the molecule with increasing number and quality of restraints. Therefore, the two techniques are highly complementary, because the combination of the two yields valuable information throughout the spectrum of distance scales, even in the presence of suboptimal X-ray and/or NMR data.

Despite its great success, there are some intrinsic limitations of X-ray crystallography: molecules in crystals experience crystal packing forces that may change their conformation and/or reduce conformational heterogeneity. NMR data are usually very accurate, but the collection of a large number of long-range interatomic distances is often very difficult, resulting in a lower precision of the NMR model with respect to the X-ray structure. Furthermore, NMR restraints are usually too few for solving molecular structures without strongly relying on prior knowledge defined by geometrical constraints based on covalent bonding. Therefore, it has long been known that X-ray and NMR data provide complementary information, which can be profitably analysed together for a more accurate description of biomolecules. Moreover, the complementarity of X-ray and NMR resides in the different types of information provided by these techniques, since X-ray relies mostly on the contribution given by the heavy atoms to the electron diffraction pattern, while for NMR the vast majority of restraints involve the hydrogen nuclei. Even more importantly, as anticipated above, at low and medium resolution, X-ray data contain information on overall shape and long-range structural details, whereas short-range structural details, of the order of the interatomic distances, are accessible only at very high resolutions, which are not always achievable. In contrast, NMR data mainly provide direct information on local details, in the form of interatomic distances or orientations of vectors connecting chemically bound nuclei. Therefore, information from NMR and X-ray data is perfectly complementary.

Among the structural restraints which can be obtained in NMR spectroscopy, pseudo-contact shifts (PCSs) [2] and residual dipolar coupling (RDCs) [3] have attracted increasing interest during the last decades for their intrinsic long-range nature. They can in fact provide structural information on the relative positions or orientations of pairs of atoms throughout the whole macromolecule or a large part of it. When the molecule is paramagnetic, dipolar interactions arise between the nuclei and the residual electron polarization, which is proportional to the magnetic susceptibility. If the magnetic susceptibility is anisotropic, these dipolar interactions do not average to zero upon rotation and PCSs arise (see later). An anisotropic magnetic susceptibility is usually associated with metal ions coordinated to the molecule [4–6] which, if not originally present, can be included by substitution of a diamagnetic metal ion [7–16] or, alternatively, rigidly attached through tags [17–41]. Other paramagnetic centres, such as organic radicals, have too little anisotropy to cause PCS or alignment effects. Magnetic susceptibility anisotropy also causes partial alignment of the molecule. In turn, partial alignment prevents internuclear dipolar interactions to be completely abolished by rotation, causing RDCs. This self-orientation is an alternate way to generate RDCs without using an external alignment medium [4,42–54]. PCSs and paramagnetic RDCs depend on the molecular nuclear coordinates in a common frame defined by the magnetic susceptibility anisotropy tensor associated with the paramagnetic metal. Self-orientation RDCs can also be obtained in the case of molecules for which the diamagnetic susceptibility is anisotropic [3,55–59] although in this case PCSs are not present.

It is interesting to observe that the presence of self-alignment also affects the chemical shifts of the observed species, if the chemical shielding of the nucleus is anisotropic. In the case of paramagnetic systems, the observed shift (not to be confused with the PCS) will be a combination of the effects of chemical shielding anisotropy and of the interaction with the electron average magnetic moment [60–62]; in diamagnetic systems it will reflect the chemical shielding anisotropy [63].

PCSs and RDCs contain structural information that has proved very helpful for solving protein structures [5,10,64–70], and they have therefore been included as structural restraints in the most commonly used programs for protein structure determination from NMR data [65,69,71–75].

PCSs and RDCs are even more precious restraints in the investigation of proteins constituted by multiple domains, and of protein–protein complexes. In the case of rigid systems, in which the structure of each single unit is known, PCSs and RDCs can be used to determine the relative arrangements of the individual units [76–83]. An easy way to recover information on the relative orientations of rigid units is to estimate the magnetic susceptibility tensors for each of the units and then superimpose them. The presence of degenerate solutions can be removed by the use of an adequate number of metal ions or orienting media [84,85].

PCSs and RDCs are even more useful for retrieving important information about the investigated biological system when mobility is present. Especially RDCs, due to their intrinsic sensitivity to small structural changes, are widely used to probe conformational rearrangements of proteins occurring on timescales up to 10^{-2} s. Also in this case, the comparison of the anisotropy tensors or alignment tensors estimated for the single units constituting a multi-domain protein or a protein–protein complex can provide interesting insights into the mobility of the system.

Finally, PCSs and RDCs have been used for validation of existing molecular models and for their refinement. Several protocols have been presented for calculating refined structures based on restraining the backbone dihedral angles and/or the nuclear coordinates to undergo minimal deviations from the values of crystal models and simultaneously to improve the agreement with the experimental PCSs and/or RDCs, or also by allowing for sizable global changes [35,64,86–93].

More recently, PCSs and RDCs have been used for structure refinement in combination with X-ray data, by implementing them as structural restraints in the program REFMAC5 [94]. Differently from the above-mentioned approaches, the refinement protocol implemented in REFMAC-NMR takes advantage of the structural information contained not only in the available crystallographic model, but also directly in the primary X-ray data (i.e. electron diffraction pattern) and in the NMR data. This permits the limitations of the direct use of crystallographic models, whose accuracy in the atomic coordinates is related to the resolution of X-ray data, to overcome, obtaining improved structural models with accuracy determined equally by the contributions of both X-ray and NMR data.

2. Combining X-ray and solution NMR data

2.1. General aspects

In several cases, a large number of violations of solution NMR data have been noticed with respect to structural models obtained from X-ray crystallography. Historically, such discrepancies have also occasionally led to correction of the X-ray structure on the basis of the NMR information [95]. However, such discrepancies may either reflect real differences between the true structures of the molecules in crystals and in solution, or may be due to the different but complementary information contained in the X-ray and NMR data, thus the true structures may be the same but the process of deriving the models by analysing the different types of data may lead to inconsistencies. In a number of cases, refinements performed using combined X-ray and NOE-derived restraints revealed large consistency of the data, resulting in the improvement of the geometry of the model in terms of Ramachandran plot with respect to the structure calculated without NMR data [96–100]. The few violations were interpreted as real differences between the structures in crystals and in solution, mostly ascribable to a limitation of the conformational freedom for some flexible parts and to the presence of crystal packing forces in the solid state. In other cases, the joint refinement provided more accurate models, in the sense that they better fulfil independent sets of experimental observables, especially in the presence of poorly determined regions of the electron density maps, due to packing disorder within the crystal [101], or in the presence of medium or low resolution diffraction data [102].

As an example, some discrepancies were observed between a very large NMR dataset collected for ubiquitin, a protein frequently used as a model system for the development of new NMR methodologies, and its crystal structures [103]. These discrepancies can be interpreted as mobility effects and used to derive motional information on the system [103–105]. On the other hand, they can also be explained by an intrinsic inaccuracy in the crystallographic models, so that a single refined structure could adequately explain the NMR data [94,106]. This points out some critical aspects in the interpretation of potential inconsistencies observed between solution data and crystal models, which requires a careful analysis before any further action.

Hence, a joint refinement of the crystal structural model with solution data may represent a first step that should be undertaken for the correct interpretation of any inconsistencies between X-ray and NMR data. In particular, the assessment of the significance of the possible discrepancies represents a crucial point, which must take into account the inaccuracies characteristic of the two types of data.

Zweckstetter and Bax [107] examined the effect of the presence of structural inaccuracy, also referred to as “structural noise”, demonstrating through a number of simulations and Monte Carlo analysis that the agreement between the experimental and back-calculated RDCs is strictly related to the amount of structural noise present in the model. Moreover, in many cases the agreement of the NMR data was observed to depend on the resolution of the X-ray structure, such that it improves significantly for crystal structures solved with higher resolution [90,108–110]. Indeed, the uncertainty in heavy atom positions necessarily affects the positions of ^{1}H nuclei; because of the small contribution hydrogens make to the diffraction pattern, information on their positions from the experimental data is usually insufficient, and the different computational tools used to position them indirectly on the basis of ideal geometrical values often give different results. On the other hand, ultrahigh resolution X-ray structures show that the positions of the resolved ^{1}H nuclei can deviate from ideal geometry [111]. Indeed, as extensively demonstrated by many authors [111–114], the geometries of molecules may often deviate from standard values for various types of protein bond length and bond angles, due to effects that are conformation-dependent, explained by avoidance of atomic clashes or optimization of favourable electrostatic interactions.

Regarding the positions of hydrogen atoms, attention should be paid to the fact that the hydrogen electrons (seen by X-ray data) are not centred on the positions of the nuclei (seen by NMR data) but are closer to the atoms to which they are attached. Therefore, the hydrogen “positions” must differ for the evaluation of the X-ray and NMR restraints, to take into account the different distances of hydrogen nuclei and their electron clouds from the atoms to which they are bound.

In summary, it is of primary importance to take advantage of X-ray and NMR data in an efficient and unbiased way, and to combine them carefully so as to avoid any possible misinterpretation of the data. As already mentioned, REFMAC-NMR performs a structural refinement by the simultaneous use of primary X-ray and NMR data, mostly PCSs and/or RDCs, and thus can be used to detect the presence of significant discrepancies, outside the uncertainty related to the experimental dataset, between solution and crystal structures; in cases where all data are consistent, it can provide a more reliable structural model.

In the sections that follow, a systematic approach to the investigation of possible structural differences between crystal and solution data, and possibly to the calculation of refined models, is proposed, and applied to a number of different proteins.

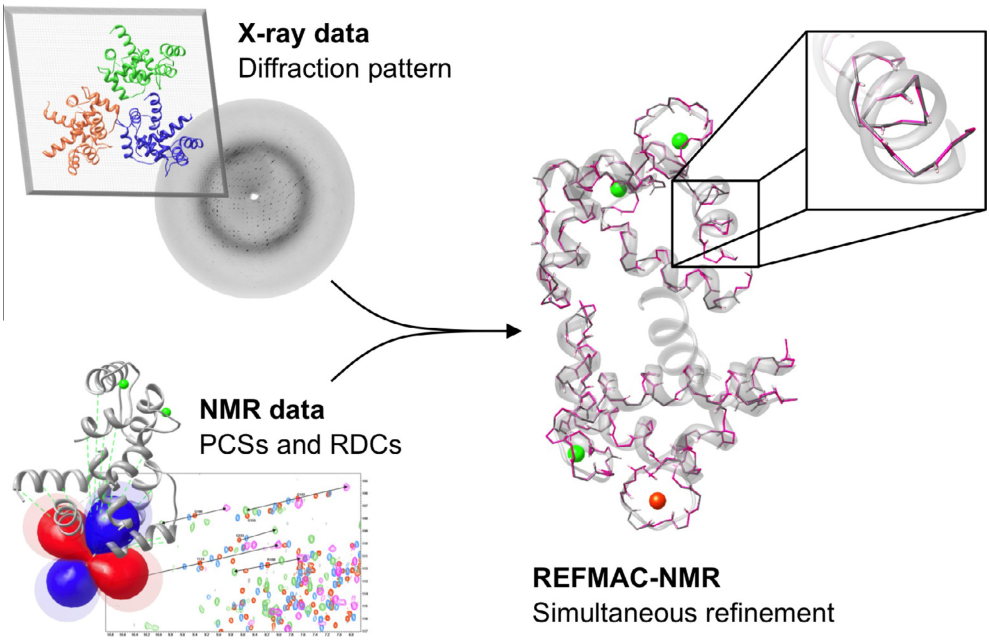


Fig. 1. Schematic representation of the information used in REFMAC-NMR structure refinement.

2.2. Structure refinement using REFMAC-NMR

REFMAC-NMR consists of an extension of the crystallographic refinement program REFMAC5 [115], a part of the CCP4 suite [116] and commonly used for medium-to-low resolution structures, by the inclusion of PCS and RDC NMR data as structural restraints in addition to the X-ray diffraction pattern (Fig. 1).

In the case of medium-to-low resolution X-ray data, the uncertainty related to the heavy atoms coordinates, as determined from X-ray data alone, may directly affect the positions of those remaining atoms which cannot be observed in the X-ray diffraction pattern (i.e. those with very weak electron density) and which are then included in the model according to the covalent geometry. This effect mostly involves the detection of hydrogen positions and often results in a non-negligible disagreement with experimental NMR data collected for them. Therefore, the inaccuracy related to the structural model can be taken into account through joint refinement against both NMR and X-ray data.

The general approach used by REFMAC-NMR for the inclusion of NMR data in the structure refinement consists of: (i) a first minimization against the X-ray data alone, with an automatic setting of relative weights of geometry and X-ray violations, possibly followed by manual adjustment of the weights to reduce the calculated rms deviations of bond lengths, bond angles, and chiral volumes, if too large; and (ii) a second minimization performed with the same weight settings including the NMR data (i.e. PCSs, RDCs), in order to decrease the discrepancy between experimental and back-calculated data, often expressed by means of the Q-factor value defined as:

$$Q = \sqrt{\frac{\sum_i (\mathrm{exp}_i - \mathrm{calc}_i)^2}{\sum_i \mathrm{exp}_i^2}}$$

As detailed in [94], in REFMAC-NMR the NMR restraint contribution (t) to the total minimized function was defined as:

$$t = k_{\mathrm{PCS}} \sum_i w_i \left[\max\left(|\mathrm{PCS}_i^{\mathrm{calc}} - \mathrm{PCS}_i^{\mathrm{obs}}| - T_i, 0 \right)^2 \right] + k_{\mathrm{RDC}} \sum_i w_i \left[\max\left(|\mathrm{RDC}_i^{\mathrm{calc}} - \mathrm{RDC}_i^{\mathrm{obs}}| - T_i, 0 \right)^2 \right]$$

where T_i is the tolerance on each PCS or RDC value, w_i is the weight, and k_{PCS} and k_{RDC} are the overall weighting factors for PCSs and RDCs respectively. Besides the already optimized weight matrix values, the second minimization requires optimization of the weights of the NMR data and of the additional torsion angle restraints *pep1*, *pep2* and ω. Three further torsion angles were in fact introduced in the REFMAC library to restrain the planarity of the proper ($\mathrm{O_i}$, $\mathrm{C_i}$, $\mathrm{N_{i+1}}$, $\mathrm{C^{\alpha}_{i+1}}$), of the improper ($\mathrm{C_{i-1}}$, $\mathrm{N_i}$, $\mathrm{C^{\alpha}_i}$, $\mathrm{H_i}$) (out of plane bending of HN), and of the proper $\mathrm{C^{\alpha}_i}$–$\mathrm{C_i}$–$\mathrm{N_{i+1}}$–$\mathrm{C^{\alpha}_{i+1}}$ dihedral angles (*pep1*, *pep2* and ω, respectively). This is needed to avoid departure of geometric parameters from ideality being caused by the inclusion of the NMR data in the refinement. Furthermore, overall weighting parameters are also introduced to enforce the contribution given by the ideal geometries on all the atoms involved ('*weight refined_atoms*') or not involved ('*weight other_atoms*') in the calculation of gradients and of the second derivatives corresponding to X-ray reflections.

In this second minimization, together with the refined structure, the best-fit magnetic susceptibility anisotropy tensors or the alignment tensors associated to the PCSs and paramagnetic RDCs or to the diamagnetic RDCs, respectively, are also calculated.

Comparison between the results of refinement calculations performed with and without the NMR restraints provides the first indication of consistency between X-ray and NMR datasets and, thus, whether significant differences are present between the crystal and solution structures.

If the inclusion of NMR data in the structure refinement does not produce evident worsening in the agreement of X-ray data with the structural model (indicated by the free R factor), and provides at the same time a good agreement of the PCS/RDC data (indicated by the Q-factors), it can be concluded that the molecule has essentially the same structure in the crystal and in solution. The local adjustments driven by the NMR data may then reveal whether or not the discrepancies initially observed with the crystal structures can be explained within the uncertainty related to the X-ray data, and in this case, REFMAC-NMR can provide more reliable structural models.

In cases where it is not possible to obtain a good agreement between NMR (high Q factors) and X-ray (high free R factor) data with the same refined structure, it can be concluded that there are significant differences between the structures of the molecule in solution and in the solid state. Still, a top-down approach can be used by performing independent refinement for different units/sub-units of the investigated system, or by discarding the residues responsible for the mismatch. Obviously, a limitation of this approach is the reduced number of NMR measurements available for the different units/sub-units, which may result in a larger uncertainty in the anisotropy tensors associated with the experimental PCSs and RDCs.

For multi-domain protein cases, the number of PCSs and RDCs available is usually sufficient to be used for an independent structure refinement of each domain. From these data, the anisotropy tensors are determined independently for each domain. It should be mentioned that, while RDCs provide valuable information both in the absence and in the presence of inter-domain mobility (see Section 2.4), PCSs collected for domains moving with respect to the metal ion also report on the variation of the metal-nucleus distance and, with some caveats (see Section 2.4), they are better used for calculation of the anisotropy tensors for the domain bearing the paramagnetic metal.

2.3. Tensor calculation

PCSs arising in the presence of a paramagnetic metal depend on the magnetic susceptibility anisotropy tensor and on the nuclear coordinates according to the following equation [2,61,117,118]:

$$\mathrm{PCS}=\frac{1}{4\pi r^3}\left[\chi_{zz}\frac{2z^2-x^2-y^2}{2r^2}+(\chi_{xx}-\chi_{yy})\frac{x^2-y^2}{2r^2}+\chi_{xy}\frac{2xy}{r^2}+\chi_{xz}\frac{2xz}{r^2}+\chi_{yz}\frac{2yz}{r^2}\right] \tag{1}$$

where x, y, z are the coordinates of the nucleus when the metal ion is defined to be at the origin of the coordinate system, r is the distance between the observed nucleus and the metal ion, and χ_{ij} are the components of the symmetric susceptibility anisotropy tensor.

Dipolar couplings are averaged in solution according to the following relation

$$\mathrm{RDC}=\frac{\mu_0\gamma_A\gamma_B\hbar}{4\pi^2 r^3_{AB}}\langle 3\cos^2\theta-1\rangle \tag{2}$$

where r_{AB} is the distance between the two coupled nuclei A and B and θ is the angle between the interspin vector $\boldsymbol{r_{AB}}$ and the external magnetic field; therefore, upon free rotation, they average to zero. However, when a partial orientation is imposed, the average is no longer equal to zero. For instance, when orientation arises because of the presence of a paramagnetic centre, and thus depends on the same χ tensor components present in Eq. (1), Eq. (2) is recast into the form [4,61]:

$$\mathrm{RDC}=3k\left[\chi_{zz}\frac{2z^2_{AB}-x^2_{AB}-y^2_{AB}}{2r^2_{AB}}+(\chi_{xx}-\chi_{yy})\frac{x^2_{AB}-y^2_{AB}}{2r^2_{AB}}\right.$$
$$\left.+\chi_{xy}\frac{2x_{AB}y_{AB}}{r^2_{AB}}+\chi_{xz}\frac{2x_{AB}z_{AB}}{r^2_{AB}}+\chi_{yz}\frac{2y_{AB}z_{AB}}{r^2_{AB}}\right] \tag{3}$$

where

$$w^2_{AB}=(w_A-w_B)^2 \quad (w=x,y,z)$$
$$K=-\frac{S_{LS}}{4\pi}\frac{B_0^2}{15kT}\frac{\gamma_A\gamma_B\hbar}{2\pi r^3_{AB}}$$

and S_{LS} is the model-free order parameter, introduced to take account of some average local mobility of the coupled nuclei vectors. RDCs do not depend on the position of the coupled nuclei with respect to the metal ion.

Diamagnetic RDCs are described by an equation of the same form as that for paramagnetic RDCs (Eq. (2)):

$$\mathrm{RDC}=-\frac{3\mu_0 S_{LS}}{8\pi^2}\frac{\gamma_a\gamma_b\hbar}{r^2_{AB}}\left[A_{zz}\frac{2z^2_{AB}-x^2_{AB}-y^2_{AB}}{2r^2_{AB}}+(A_{xx}-A_{yy})\frac{x^2_{AB}-y^2_{AB}}{2r^2_{AB}}\right.$$
$$\left.+A_{xy}\frac{2x_{AB}y_{AB}}{r^2_{AB}}+A_{xz}\frac{2x_{AB}z_{AB}}{r^2_{AB}}+A_{yz}\frac{2y_{AB}z_{AB}}{r^2_{AB}}\right] \tag{4}$$

where A_{ij} are the components of the molecular alignment tensor. The magnetic susceptibility anisotropy values are defined by axial and rhombic components given by

$$\Delta\chi_{ax}=\chi_{zz}-\frac{\chi_{xx}+\chi_{yy}}{2}=\frac{3}{2}\chi_{zz}$$

and

$$\Delta\chi_{rh}=\chi_{xx}-\chi_{yy}$$

where χ_{ii} are the components of the χ tensor in the frame in which it is diagonal. Analogously, the anisotropies of the alignment tensor are described by the fraction of alignment along the z axis (A) and by the rhombicity (R), [119]

$$A=\frac{3}{2}A_{zz}$$
$$R=\frac{A_{xx}-A_{yy}}{A}$$

where A_{ii} are the components of the A tensor, again in the frame in which it is diagonal.

Anisotropy tensors and alignment tensors can be calculated from the agreement between experimental and back-calculated PCS and/or RDC data against a structural model. Several programs are available that can do this, among them the web-based interface FANTEN [76].

Recently, a method for structure calculations was proposed that does not require the calculation of a tensor describing RDC (and PCS) data. This tensor-free method, called the theta-method, relies on determining the molecular structure from MD trajectories and replica averaging, to obtain the best agreement between the experimental RDCs and theoretical RDCs directly evaluated using Eq. (2) [120].

2.4. Motionally averaged data

For systems composed of multiple units or domains, RDCs, which are independent of the distance to the metal ion, can provide a first estimate of the presence and extent of mobility. Indeed, whatever the conformational sampling experienced by the system, RDCs collected from single units or domains can always be described by a tensor, though averaged and reduced with respect to the real susceptibility anisotropy tensor of the paramagnetic ion (Fig. 2c). In the case of proteins composed of two domains, only one of which contains a paramagnetic metal, experiencing mobility with respect to one another, the RDCs collected for the domains without a metal ion are reduced by averaging in proportion to the extent of the motion. Therefore, the anisotropies of the tensors back-calculated from these data (let us call them $\Delta\tilde{\chi}_{ax(\text{domain 2})}$) are reduced with respect to the anisotropies of the tensors back-calculated from the RDCs of the metal-bearing domain ($\Delta\chi_{ax(\text{domain 1})}$). The extent of the motion can be easily quantified using the ratio of the magnitudes of the anisotropies of the tensors calculated for the two domains, i.e. from $\Delta\tilde{\chi}_{ax(\text{domain 2})}/\Delta\chi_{ax(\text{domain 1})}$. This is not valid in principle for PCSs, which cannot be represented by any mean tensor in the case of motion, unless all nuclei have a fixed distance from the metal ion during the motion [121], which is never the case in practice.

Once it has been assessed that significant mobility affects the system, PCSs and RDCs can be used to further characterize the resulting conformational variability. In this respect, a multitude

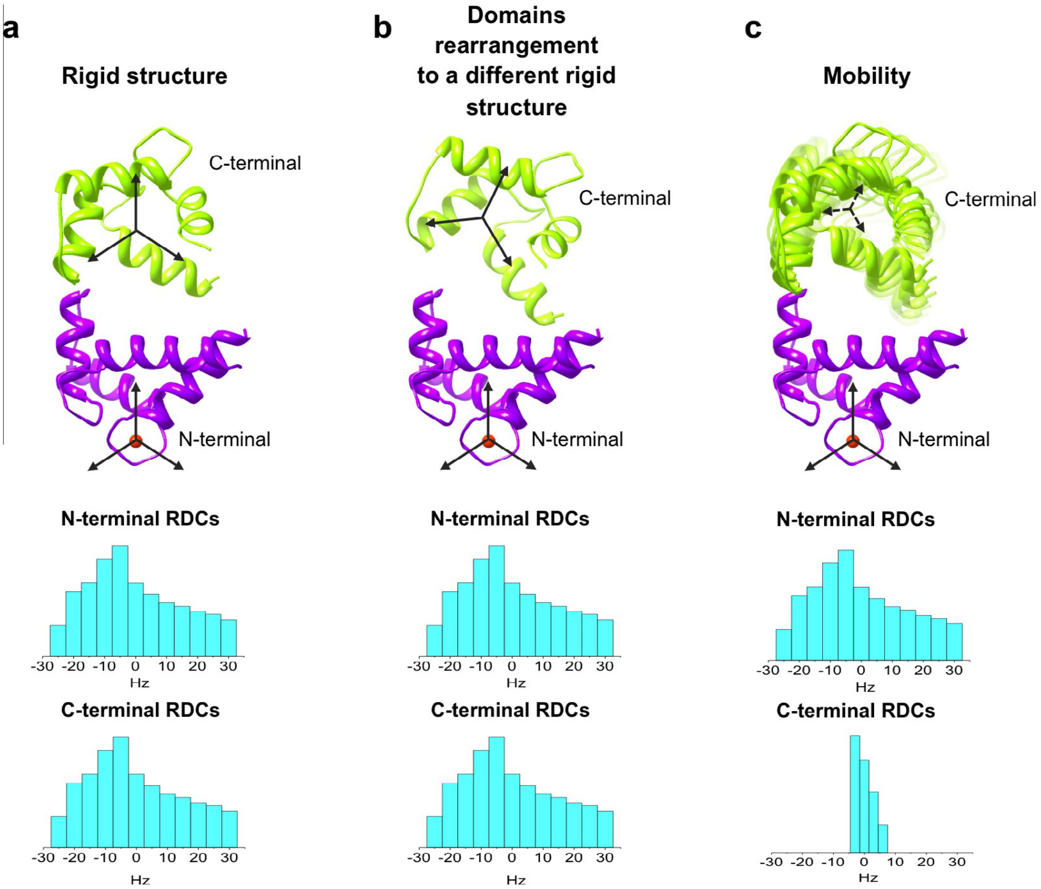


Fig. 2. The different effects of domain rearrangement and mobility on the range of observed RDCs.

of methods have been implemented to reconstruct the ensembles of putative conformations sampled by the system. This is, by definition, an ill-posed inverse problem, allowing for an infinite number of solutions (see Section 3.2.1) [57,81,103–105,122–152].

2.5. A step-by-step approach for the refinement of multidomain systems

When all domains composing a system have been satisfactorily refined against X-ray and NMR data, the best-fit tensors obtained for each domain should be compared (Fig. 2a–c). As already mentioned, this comparison might provide a very useful insight into the presence of inter-domain mobility or conformational rearrangements for the solution structure with respect to the crystal structure.

Scheme 1 reports a general approach that can be used to analyse the data and results that may be obtained under various different circumstances.

2.5.1. Check 1: Are tensor magnitudes equal?

As seen, mobility affects the measured RDCs, which are averaged over the values corresponding to all sampled conformations; this always results in a reduction of the observed values with respect to those that would have been observed in the absence of conformational variability. This, in turn, corresponds to a reduction of the associated anisotropy/alignment tensor (Fig. 2c).

In the case of paramagnetic RDCs, the averaged tensor obtained from the best fit of the experimental data against the structural model of a protein domain corresponds to an averaged magnetic susceptibility anisotropy tensor: the average is performed over all orientations sampled by the tensor with respect to the protein domain, and results in a reduced tensor.

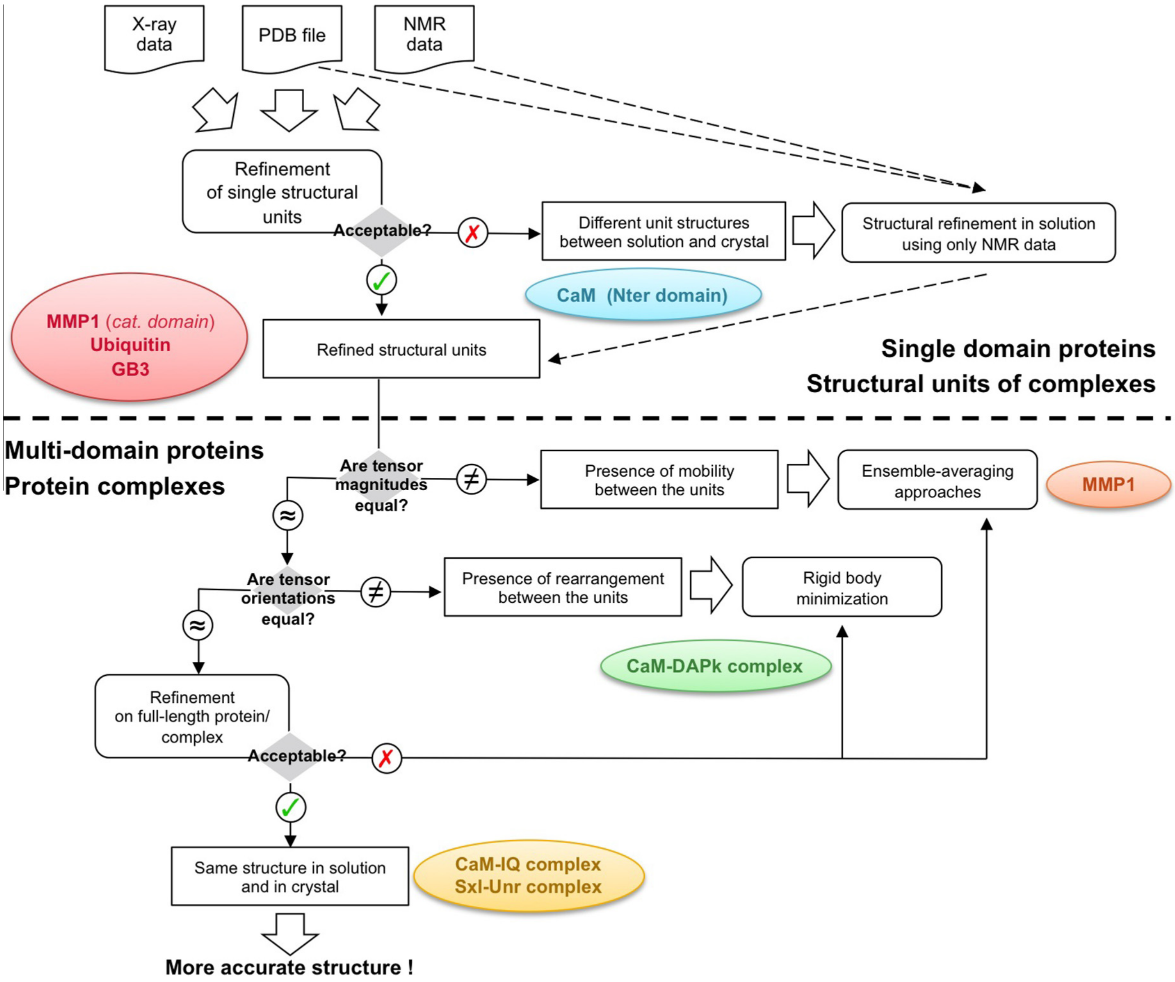
In the case of diamagnetic RDCs, an averaged tensor can still be calculated, although in this case it results from the averaging of alignment tensors, which differ from one another depending on the shape of the protein conformations (see Section 2.6). Therefore, a sizable difference in the RDC-determined tensors as calculated for the different domains is a clear indication of inter-domain mobility.

PCSs, which for rigid systems are usually employed to obtain a robust estimation of the tensor, are not scaled correctly in a dynamic system. For this reason, in principle they cannot be used in the structure refinement of the protein domain if the latter is mobile with respect to the position of the metal ion, unless the geometry of the system suggests that the variability in the metal position is small compared to the metal-nuclear distances, and some inaccuracy in the results can be tolerated [121].

Because of their sensitivity to global long-range conformational changes, RDCs and PCSs can be usefully used jointly to explore the conformational space sampled by the system in the case of conformational variability. Various approaches have been conceived for such a purpose (see Section 3.2.1), to generate and analyse ensembles of conformations in agreement with the averaged data.

2.5.2. Check 2: Are tensor orientations equal?

If the system can be stated to be rigid because the magnitudes of the anisotropy tensors calculated for the different domains are very similar, still the domains can be arranged differently in solution and in the solid state. This results in metal positions and tensor orientations that differ for the various domains (Fig. 2b).



Scheme 1.

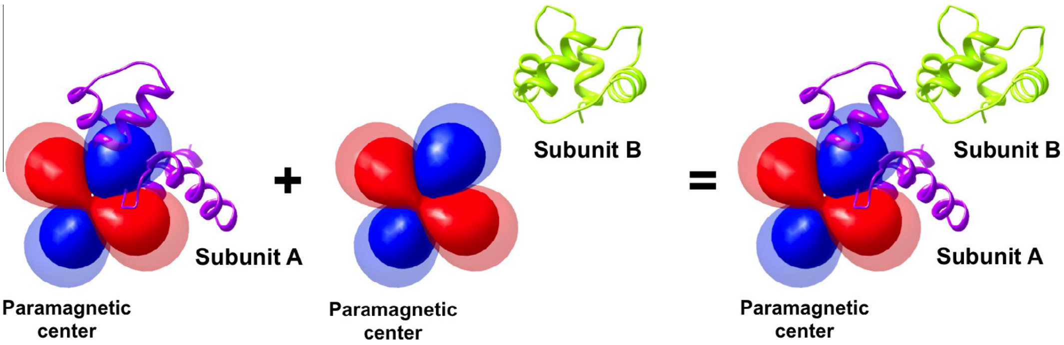


Fig. 3. Rigid body minimization based on the use of magnetic susceptibility anisotropy tensors. The tensors are first computed for each domain, then superimposed to reconstruct the reciprocal position of the two domains.

Therefore, the relative arrangements of the protein domains in solution can be retrieved by applying a rigid rototranslation to one domain in order to superimpose the sets of estimated tensors (i.e. the centre positions and axes defining their orientations) on those determined for the other domain (Fig. 3). This corresponds to recovering a structure in agreement with a set of tensors able to correctly describe the whole system.

2.5.3. Check 3: Is the full-length protein refinement acceptable?

If both the magnitudes and the orientations of the anisotropy/alignment tensors calculated for the protein domains are similar (Fig. 2a), as well as the metal positions, a full-length protein refinement can be attempted through REFMAC-NMR using a single tensor for each dataset. If this final refinement is also of good quality (low free *R*-factor and *Q*-factor), it can be stated that all data (X-ray and NMR) are in agreement with a single structural model and that the refined structure obtained represents a more accurate description of the protein. If not, as may happen in the presence of some differences in the anisotropy/alignment tensors obtained for the distinct protein domains, the discrepancies can find explanations in structural rearrangements or in conformational heterogeneity, which could not be clearly detected in the previous steps.

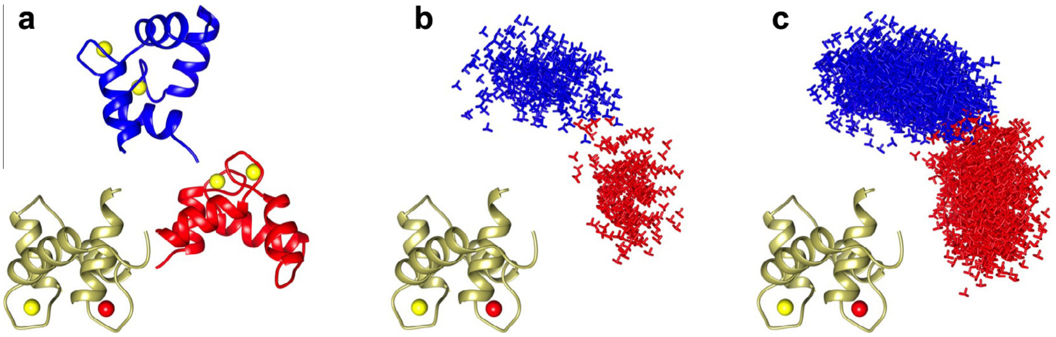


Fig. 4. (a) The two selected conformations of the two domain protein CaM used for the synthetic test: the N-terminal domain is shown in yellow, the C-terminal domain in blue (elongated conformation) or in red (compact conformation); (b and c) mobility schemes around the two centres: the C-terminal domain (replaced for each conformation with a 3-axes system centred on its centre-of-mass for a better visualization) is allowed to move within 20 Å and 20° (b) or within 20 Å and 40° (c) of translation and rotation, around the central conformations shown in (a).

Table 1
Axial components of the alignment tensors (A_a) and axial magnetic susceptibility anisotropy ($\Delta\chi_{ax}$)[a] calculated for ensembles of structures generated around the extended and compact conformations shown in Fig. 4a (blue and red, respectively).

d (Å)	δ (°)	# structures	N-terminal A_a[a]	C-terminal A_a[a]	N-terminal $\Delta\chi_{ax}$[a,d]	C-terminal $\Delta\chi_{ax}$[a,d]
Extended conformation						
0	0	1	1	1	1	1
20[b]	20[b]	321[b]	0.89	0.86	1	0.88
20[c]	40[c]	1418[c]	0.75	0.84	1	0.61
30	50	3872	0.73	0.69	1	0.39
Whole pool of sterically allowed conformations		32,723	0.52	0.41	1	0.025
Compact conformation						
0	0	1	1	1	1	1
20[b]	20[b]	247[b]	1.13	1.03	1	0.87
20[c]	40[c]	1264[c]	0.67	0.86	1	0.63
30	50	3393	0.81	0.62	1	0.43
Whole pool of sterically allowed conformations		32,723	0.63	0.49	1	0.025

[a] With respect to the single central conformation.
[b] Ensembles shown in Fig. 4b.
[c] Ensembles shown in Fig. 4c.
[d] With the metal in the N-terminal domain.

As already discussed, this approach can be applied with some flexibility in the definition of the subunits/domains composing the biological system investigated. The approach is limited by the number and variability of the experimental data available for the individual parts, and thus by the uncertainty in the corresponding estimated tensors. Moreover, the quality of the achieved agreements intrinsically depends on the experimental error as well as on the approximations in the applied model. Thus, a careful examination of all possible cases is generally recommended.

This step-by-step approach can also be applied in a top-down fashion: in this case, an overall refinement of the whole system is first attempted and, in case of unsatisfactory results (large *Q*-factor and/or *R*-free value), refinements of the individual subunits/domains can be performed.

2.6. Averaged tensors from diamagnetic and paramagnetic RDCs

As seen in Section 2.4, in case of paramagnetic (self-alignment) RDCs measured for two-domain proteins (or for other systems composed of two, or more, internally rigid domains), one of which contains a paramagnetic metal, conformational averaging greatly reduces the magnitude of the alignment tensor of the domain without the metal with respect to that of the metal-bearing domain. This reduction is dependent on the degree of interdomain mobility and, indeed, the ratio of the anisotropies of the alignment tensors for the two domains is sometimes used as a generalized order parameter describing interdomain mobility Fig. 2c [127,153]. The situation is very different for diamagnetic RDCs induced by external alignment media. In that case, none of the domains is fixed in the frame of the alignment tensor, and thus the RDCs measured for both domains are subject to motional averaging (except in the case where one domain is much bigger and/or more charged than the other and thus completely dominates the alignment properties of the system [154]). For this reason, the averaged tensors obtained for the two domains can have very similar magnitudes even in the presence of extensive interdomain mobility. Furthermore, as each conformer aligns to a different extent (according to its shape and charge distribution), different alignment tensors should be considered for the different conformations sampled by the system. This may have the result that, on passing from a single rigid state to an ensemble of conformations, an increase rather than a decrease of the alignment tensor can be observed, if some of the newly populated states align more strongly than the original structure.

It is instructive to illustrate the statements made above by means of some specific numerical simulations. For this purpose, two conformers (one extended and one compact, see Fig. 4) were chosen from a broad pool of sterically allowed conformations of

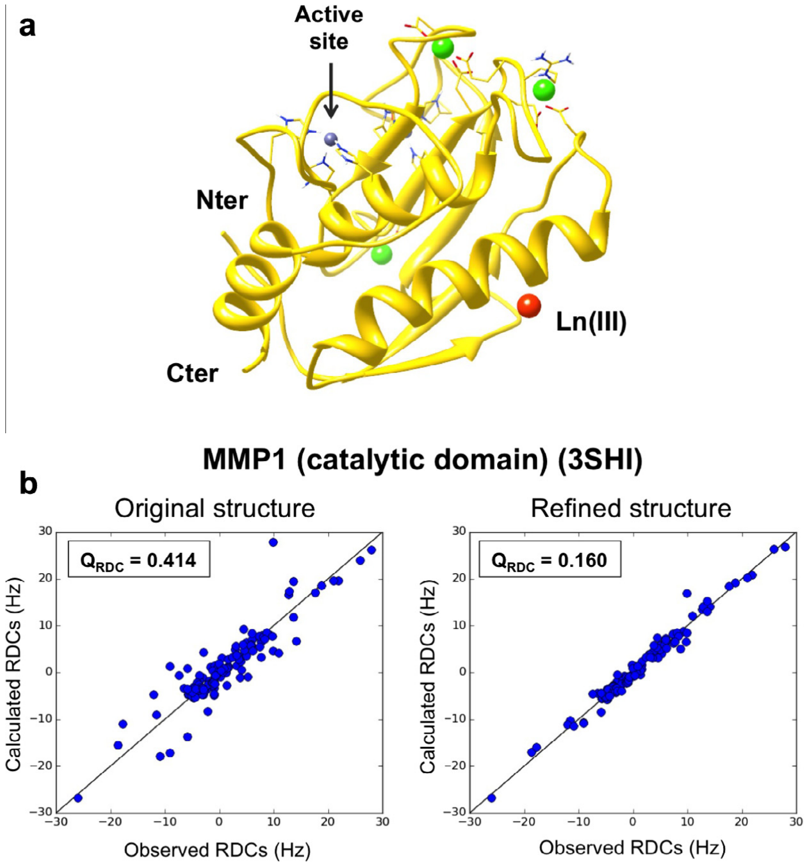


Fig. 5. (a) Structure of the catalytic domain of MMP1, showing the active site Zn(II) ion and the lanthanide tag position; (b) correlation plot between calculated and observed RDCs before and after the refinement.

the protein calmodulin (CaM, see Sections 3.1.4 and 3.2.2) [140]. The diamagnetic RDCs for the two conformers (and all the other conformers in the pool), induced by a fully steric alignment medium, were calculated using the program PALES [119,155]. The corresponding alignment tensors differ for the two conformers in such a way that the ratio of tensor magnitudes between the compact and the extended conformation is 0.83. The presence of conformational variability around a central structure was then simulated by constructing ensembles comprised of all the conformers with Cartesian and angular displacement from either the extended or the compact structure within defined thresholds (see Fig. 4 and Table 1). Averaged RDC values (and corresponding averaged tensors) were calculated by averaging (with equal weight) the PALES-derived RDCs.

Table 1 provides a summary of the obtained results in terms of the ratios of the magnitudes of the alignment tensors corresponding to the different ensembles with respect to those for the central (rigid) structures. Clearly, in both cases the magnitudes of the alignment tensors for the C-terminal and the N-terminal domains remain comparable even when a very large mobility is present. Therefore, it is not possible to exclude the presence of extensive conformational variability, or to quantify its extent, from the observation that different domains have alignment tensors of similar magnitude. This is in contrast with the average tensors determined from the paramagnetic RDCs, which are progressively reduced as conformational variability increases (last column of Table 1). It is also worth noting that for an ensemble around a compact conformer, the presence of limited mobility actually leads to an increase in the molecular alignment determined from diamagnetic RDCs, as discussed above.

3. Case studies

3.1. Single domain proteins

3.1.1. The catalytic domain of matrix metalloproteinase 1 (MMP1)

Matrix metalloproteinases are a family of multidomain proteins that mediate the hydrolysis of extracellular protein substrates such as elastin and collagen [156], and are putatively involved in cancer invasion. The interdomain orientation, which is crucial for catalysis, will be discussed in Section 3.2.1. These proteases have one catalytic Zn(II) ion in one domain, the catalytic (cat) domain. The structures of the cat domains of several MMPs have been solved by NMR and X-ray crystallography, mainly for the purpose of rational drug design [157].

REFMAC-NMR was used to refine the catalytic domain of MMP1 [94] (PDB code: 3SHI), for which PCSs of H^N nuclei and RDCs of H^N–N coupled nuclei for three paramagnetic lanthanides (Tb^{3+}, Tm^{3+}, and Yb^{3+}) bound to the protein through the CLaNP-5 tag were available [35]. Residues revealing extensive mobility through relaxation measurements were discarded from the refinement calculation. The original crystal structure (resolution: 2.2 Å) showed modest agreement for PCSs and RDCs, initially pointing to possible discrepancies between the solution and the crystal structure. The REFMAC-NMR refinement protocol was applied and a significant overall improvement in the agreement of NMR data was obtained, especially for RDCs, for which the *Q*-factor decreases from 0.414 to 0.160 (Fig. 5). The local adjustment to the crystallographic structure carried out by PCSs and RDCs resulted in a rms deviation for the backbone atoms with respect to the original model of only 0.039 Å. These small displacements in heavy atom positions, which

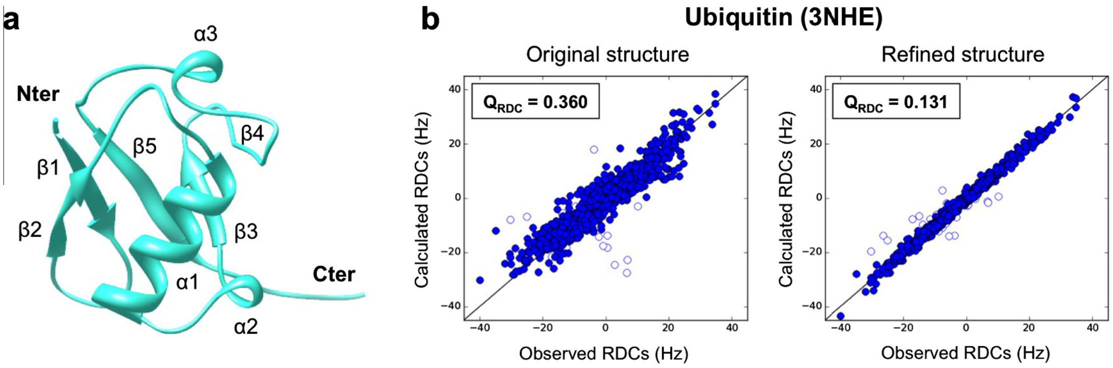


Fig. 6. (a) Structure of ubiquitin; (b) correlation plot between calculated and observed RDCs before and after the refinement (empty dots refer to residues 8 and 72).

are as compliant with the X-ray reflections as the original structure, are sufficient to improve the positioning of H^N nuclei in the structure and, thus, the agreement with the NMR data.

3.1.2. Ubiquitin

Ubiquitin is a small regulatory protein (76 residues, 8.5 kDa), which is found in almost all eukaryotic cells as well as in bacteria [158,159]. Attachment of this protein to other proteins exerts a signalling effect that depends on the location, the residue through which ubiquitin is coordinated, and how many ubiquitin molecules are bound [160]; a detailed discussion is given in Refs. [161,162]. A prototypical example is the covalent binding of ubiquitin to proteins targeted for proteasomal degradation [163,164].

The structure of ubiquitin has been solved as many as 178 times, alone or in complexes, by X-ray (129 PDB entries), NMR (46 solution and 2 solid-state entries) and electron microscopy (1 entry).

A sizable range of RDC datasets has been available for ubiquitin for a long time, and this wealth of RDC data has been interpreted differently by different groups. Prestegard and co-workers developed a methodology to extract simultaneously both structural and dynamic properties for ubiquitin, obtaining a solution structure (in agreement, on average, with the X-ray structure) and a characterization of its anisotropic internal motions, quantified by the generalized degree of order [153]. Analysis of the data was carried out by Clore and Schwieters looking for the smallest possible ensemble which could account for the experimental data, yielding a two-site jump model [104]. In the same paper it was also noted that the structural impact of residues showing the largest difference between the two states is minor, as they can be accounted for by compensatory movement of the backbone angles. Vendruscolo and co-workers developed a protocol for simultaneous determination of ubiquitin structure and dynamics by combination of molecular dynamics simulations and experimental information derived from NMR spectroscopy, finding a considerable conformational heterogeneity throughout the protein structure [105]. Along the same lines, Griesinger and co-workers postulated that ubiquitin follows mainly a single concerted motion, which allows for sampling many of the conformations observed in X-ray structures of complexes and at the same time explains the discrepancy between the experimental and the back-calculated data from any individual conformations [103]. In contrast, Bax and co-workers recently showed that it is possible to generate a single low-energy conformation that exhibits similar agreement to the experimental data (both the working set and the cross-validation set) as the previously discussed ensemble. This implies that the better fit of the ensemble to the experimental data with respect to any individual X-ray structure “is likely dominated by averaging of small errors in the backbone coordinates and not from true dynamics” [106]. It is worth noting that these conclusions do not to detract from the intrinsic ability of RDCs to reveal motions, but rather question the accuracy of reconstructed ensembles.

In this example, H^N–N RDCs measured from as many as 36 different alignment media [54,103] were used [94] for the refinement of an X-ray structure of ubiquitin (PDB code: 3NHE), which displayed a 1.26 Å resolution. The initial disagreement shown for the original structure decreases, in terms of Q_{RDC}, from 0.360 to 0.121 after performing REFMAC-NMR refinement (Fig. 6) [94]. Notably, a few violating RDCs were still found for the refined structure, mostly belonging to residues located in regions experiencing extensive mobility, as revealed by relaxation experiments [103,106], although not showing larger B-factors in the crystal structure. The possibility of achieving a similarly good agreement with the experimental data using different structures does not undermine the validity of the approach, but rather reflects the need to acquire as many experimental restraints as possible to reduce the uncertainty in the structural model.

3.1.3. IgG-binding domain of protein G (GB3)

Protein G is of particular interest, as it binds to a wide range of antibodies and its interactions have been well-characterized. Protein G has been shown to bind strongly to the Fc fragment and weakly to the Fab fragment of human immunoglobulin G (IgG) [165]. Protein–protein interactions between various protein G domains and isolated fragments of IgG have been studied by both solution NMR spectroscopy and X-ray crystallography [90,166–168].

For GB3 (PDB code: 1IGD), RDCs for five alignment media were collected for H^N–N, C^α–H^α, and C–C^α nuclear pairs [90]. Despite the acceptable agreement with experimental datasets already shown for the original model (with resolution 1.1 Å), REFMAC-NMR produced a further improvement in the Q_{RDC}, which decreased from 0.146 to 0.081.

A refinement of the same X-ray structure has been also performed by Bax and co-workers and used to draw interesting conclusions about the peptide bond H^N–N orientations as a result of pyramidalization of peptide N and fluctuations orthogonal to the peptide plane [90]. More recently, the same group presented a newly refined solution structure of GB3 using further C^α–C^β RDCs collected for three different alignment media, and validated the obtained structure using the prediction of three-bond *J* couplings by Karplus equations [168].

3.1.4. The N-terminal domain of calmodulin (CaM)

CaM is a 16 kDa, highly conserved protein, implicated in many biological functions. It is composed of two very similar domains. Each domain is composed of two EF-hand motifs, and can bind up to 2 calcium(II) ions, so that the full length protein can bind

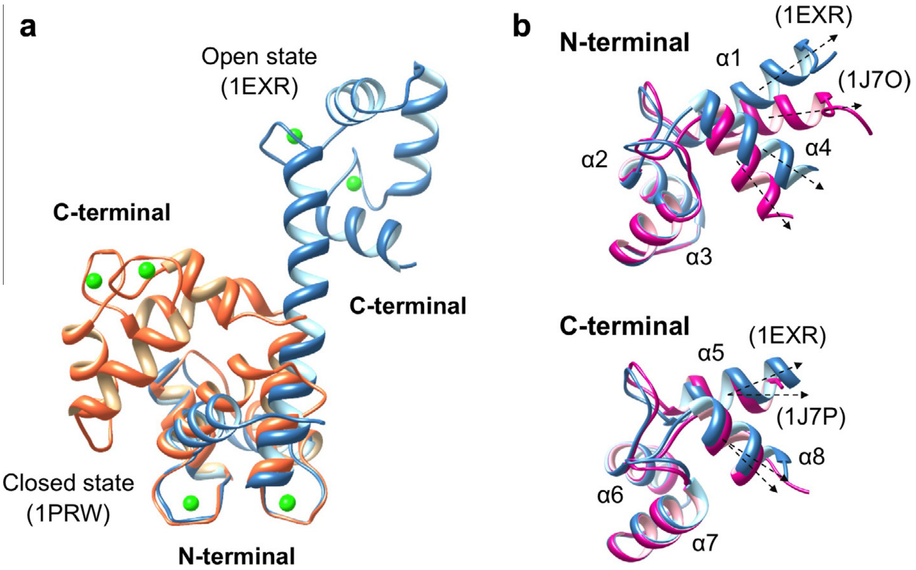


Fig. 7. (a) X-ray structures of free calmodulin, showing open and closed conformation; (b) differences between the solution structure and the X-ray structure of the individual domains [170].

up to 4 calcium(II) ions. In resting cells the protein binds one calcium(II) ion only. Upon increase of calcium concentration, all four sites are occupied, and the protein exposes two hydrophobic patches that can bind to several targets and is thus a prototypical intermediate messenger [169]. The interdomain mobility, which is relevant for the biological function, will be discussed in Section 3.2.1.

A set of diamagnetic H^N–N, C^α–H^α, C–C^α, and C–H^α RDCs [170] were used as input in an attempted refinement with REFMAC-NMR of the N-terminal domain of CaM (PDB code: 1EXR) [94]. However, in this case no satisfactory agreement was achieved for the NMR data without a significant increase in the *R*-free values and in the violations of geometrical parameters. This result confirms the findings by Bax and co-workers [170], showing substantial differences in the inter-helical orientations between the crystal and the solution structures of CaM N-terminal domain, mainly due to reorientation of helices 1 and 4. The REFMAC-NMR calculations indicate that the detected differences are outside the uncertainty of X-ray data (resolution: 1.0 Å) (Fig. 7).

3.2. Multi-domain proteins and complexes

3.2.1. Full length matrix metalloproteinase 1 (MMP1)

Matrix metalloproteinases (MMPs) are enzymes involved in extracellular matrix degradation, which is a fundamental step in many physiological processes such as tissue remodelling and repair [171,172]. Most MMPs comprise a prodomain that is removed upon activation, a catalytic domain responsible for hydrolytic activity, and a hemopexin-like domain that probably plays a role in substrate recognition. In several pathologies, overexpression of MMPs, or misregulation of their activity, is related to disease progression [173–175]. MMPs are thus validated pharmaceutical targets. Selectively targeting these proteins has turned out to be a difficult task [176,177], and structure-based drug design appears to be of great importance [178–180]. For this reason, the structural characterization of MMPs has received much attention [181].

After binding of the CLaNP-5 tag into the catalytic N-terminal domain of MMP-1 (see Section 3.1.1), PCS and paramagnetic RDCs induced by three different lanthanides (Tb^{3+}, Tm^{3+}, and Yb^{3+}) were also measured for the C-terminal hemopexin-like domain of the protein. The analysis of the collected RDCs reveals that the magnitudes of RDC-derived magnetic susceptibility anisotropy tensors for the catalytic domain are around 3–4 times higher than those calculated for the hemopexin domain. Such a reduction of the RDC-derived tensor observed for the metal-free domain is, as described above, a clear indication of conformational averaging. Therefore, a joint refinement against X-ray and NMR data was not performed. On the other hand, the system clearly samples only a limited subset of the sterically allowed conformations, as a uniform population of the entire conformational space would result in a much more drastic reduction of the tensor magnitudes (up to approximately 30 times).

As already mentioned, the problem of recovering the sampled conformers from averaged experimental observables is an ill-posed inverse problem. This means that an infinite number of, often very different, ensembles of structures can satisfy the experimental restraints equally well. Still the averaged experimental data clearly contain information about the conformational preferences of the system. In order to extract as much information as possible without the risk of over-interpreting the data, we developed an approach called Maximum Occurrence (MaxOcc) [134,140,182,183]. The method permits one to assign to each sterically allowed conformation a value, called MaxOcc, defined as the highest weight which the conformation can have in any ensemble in best agreement with the experimental observables. The MaxOcc of a conformation thus represents the upper bound that the averaged data impose on the statistical weight of that structure. Numerous synthetic tests have shown that in many cases the conformers with the highest MaxOcc correspond to those that are the most prominent in the actual conformational ensemble [31,85,140,151].

The MaxOcc analysis performed using averaged PCS and paramagnetic RDC data, complemented by the SAXS profile measured for the protein, could identify structures with MaxOcc values up to 47%. Interestingly, only a very small fraction of sterically allowed conformers have high MaxOcc (only 0.3% of the structures have MaxOcc > 40%), and all of them are clustered in a well-defined region of the conformational space of the protein, so that they are characterized by an interdomain orientation and position that can be defined relatively well (see Fig. 8). In these highest MaxOcc conformations the residues of the hemopexin domain responsible for collagen binding are well exposed to the solution; after collagen binding by the hemopexin domain, the catalytic domain is already

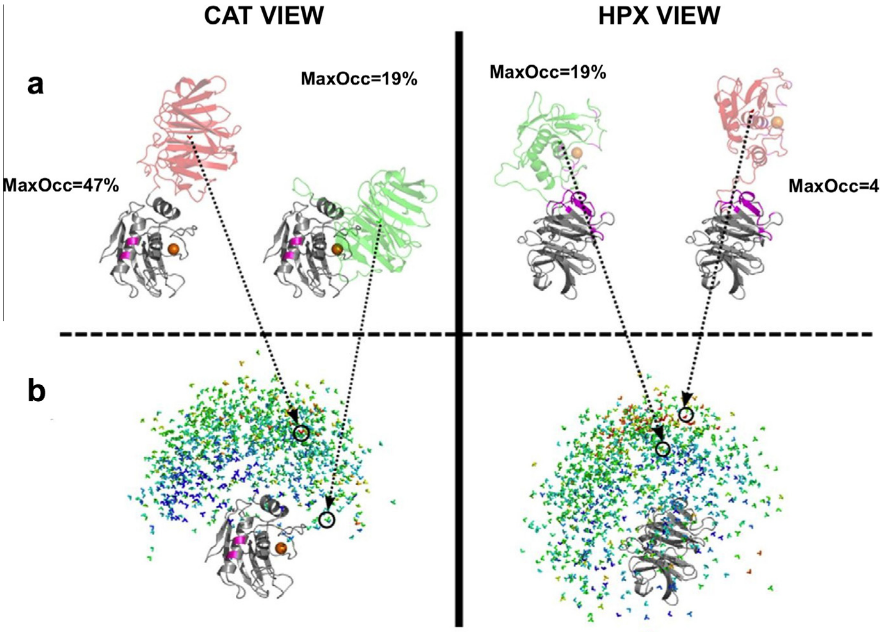


Fig. 8. Visualization of the results of MaxOcc calculations for 1000 randomly selected MMP-1 conformations. The conformations are displayed superimposed on the CAT domain (left) and on the HPX domain (right). The catalytic metal is represented as an orange sphere. Colours from blue (<5%) to red (47%) represent the MaxOcc values of the various structures. (A) The structure with the highest MaxOcc (47%) and the X-ray crystallographic structure 2CLT are coloured according to their MaxOcc values. (B) Each conformation is represented, for graphical simplicity, as a colour-coded 3-axis system, positioned at the centre of mass of the HPX (left column) or CAT (right column) domain. Reproduced with permission from Ref. [38] © of the American Society for Biochemistry and Molecular Biology.

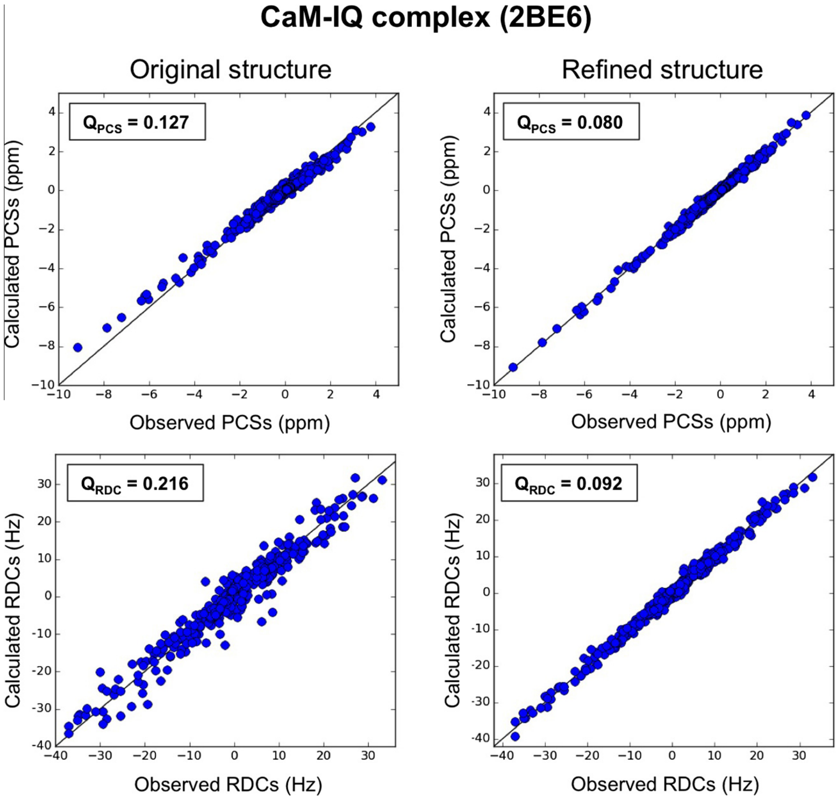


Fig. 9. Correlation plot of the PCS and RDCs calculated for chain B of the CaM–IQ 2BE6 structure before and after refinement.

preferentially oriented in such a way as to easily access the collagen as well, and a rotation by around 50° is sufficient to recover the previously proposed conformation responsible for the first step of collagenolysis [184]. Thus the MaxOcc analysis suggests that already in its free form in solution the full-length MMP-1 shows a preference for conformations which are poised for interaction with collagen and thus for catalytic activity [38]. Notably, the highest MaxOcc conformations differ largely from the closed MMP-1 structures obtained by X-ray crystallography, with a MaxOcc of about 20%.

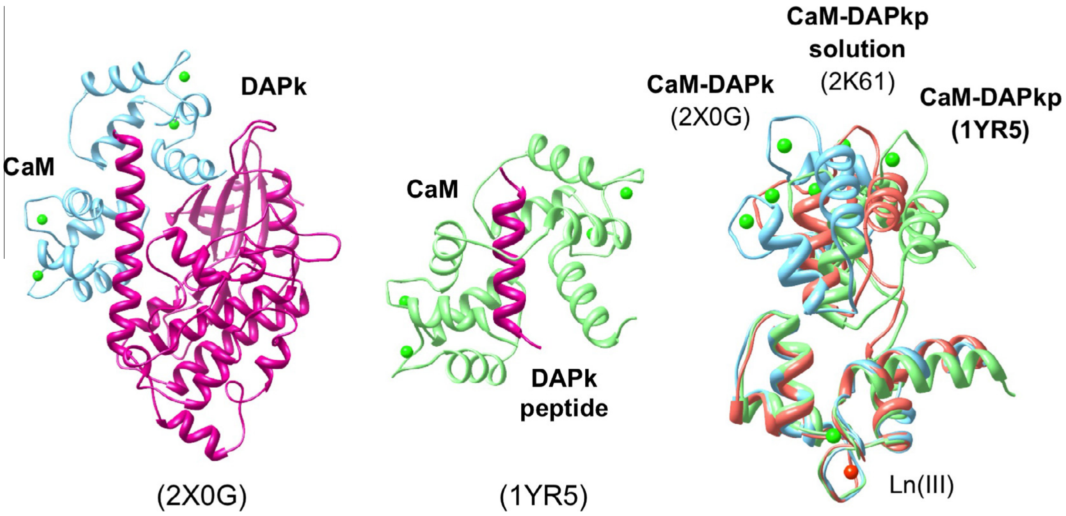


Fig. 10. Comparison between the X-ray structures of CaM bound to full-length DAPK (2X0G) [191], a peptide derived from DAPK–DAPKp– (1YR5) [92], and the solution structure in the CaM–DAPKp complex (2K61) [92].

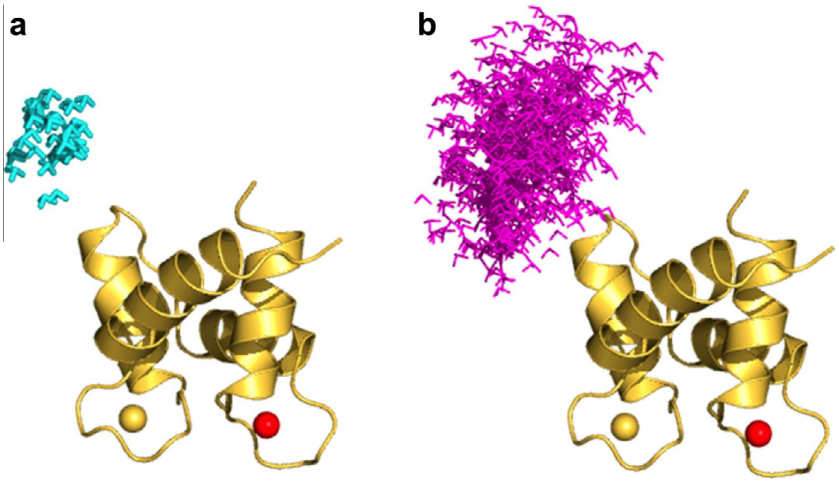


Fig. 11. (a) Smallest region with MaxOR = 1 and (b) smallest region with MinOR (Minimum Occurrence of Regions) = 0.54 of the CaM–DAPKp complex. Reproduced with permission from Ref. [151] © of the American Chemical Society.

3.2.2. Full length CaM

As already mentioned, CaM represents an interesting system as it comprises two almost rigid domains able to rearrange freely with respect to one another, when the protein is free in solution, because they are connected by a flexible linker. This high degree of flexibility is of course absent in the solid state, so that no joint refinement can be performed for the full-length structure using X-ray and NMR data. The protein has been engineered with the mutation N60D so as to make it possible for it to bind one lanthanide(III) ion selectively in the second calcium(II) binding site of the N-terminal domain [11].

The very different spread of the RDC data measured for the two domains when a paramagnetic metal is placed in one domain was earlier used to deduce the presence of a very large conformational variability [31,185–187]. Protein conformations can be ranked according to their compliance with PCSs and paramagnetic RDCs, using the MaxOcc analysis described for MMP-1 (see Section 3.2.1) [140,187,188]. This analysis takes advantage of the availability of the solution structures of the individual domains of the protein, previously obtained by refining the crystal structures using diamagnetic RDC data [170].

Interdomain flexibility is the basis of the functionally relevant role of CaM, for which a large number of binding partners have been identified. The formation of complexes induces CaM to assume a compact conformation, and depending on the specific binding interactions, the mobility between the two domains can be reduced to different extents, ranging from still very flexible [134,189] up to almost complete relative immobilization.

3.2.3. Calmodulin–IQ peptide complex (CaM–IQ)

Calmodulin is also involved in the regulation of voltage-gated calcium channels (Ca(V)s), and drives two self-regulatory calcium-dependent feedback processes that require interaction between Ca^{2+}–CaM and a Ca(V) channel consensus isoleucine–glutamine (IQ) motif: calcium-dependent inactivation (CDI) and calcium-dependent facilitation (CDF) [190].

The structure of CaM in complex with the IQ-recognition motif (PDB code: 2BE6) was refined using H^N, N PCSs and H^N–N RDCs measured for six different lanthanides (Tb^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, Tm^{3+}, and Yb^{3+}) [144]. No reasonably good fit was observed for any of the three models (chains A, B, and C) contained in the asymmetric unit of the crystal. The stepwise approach described in

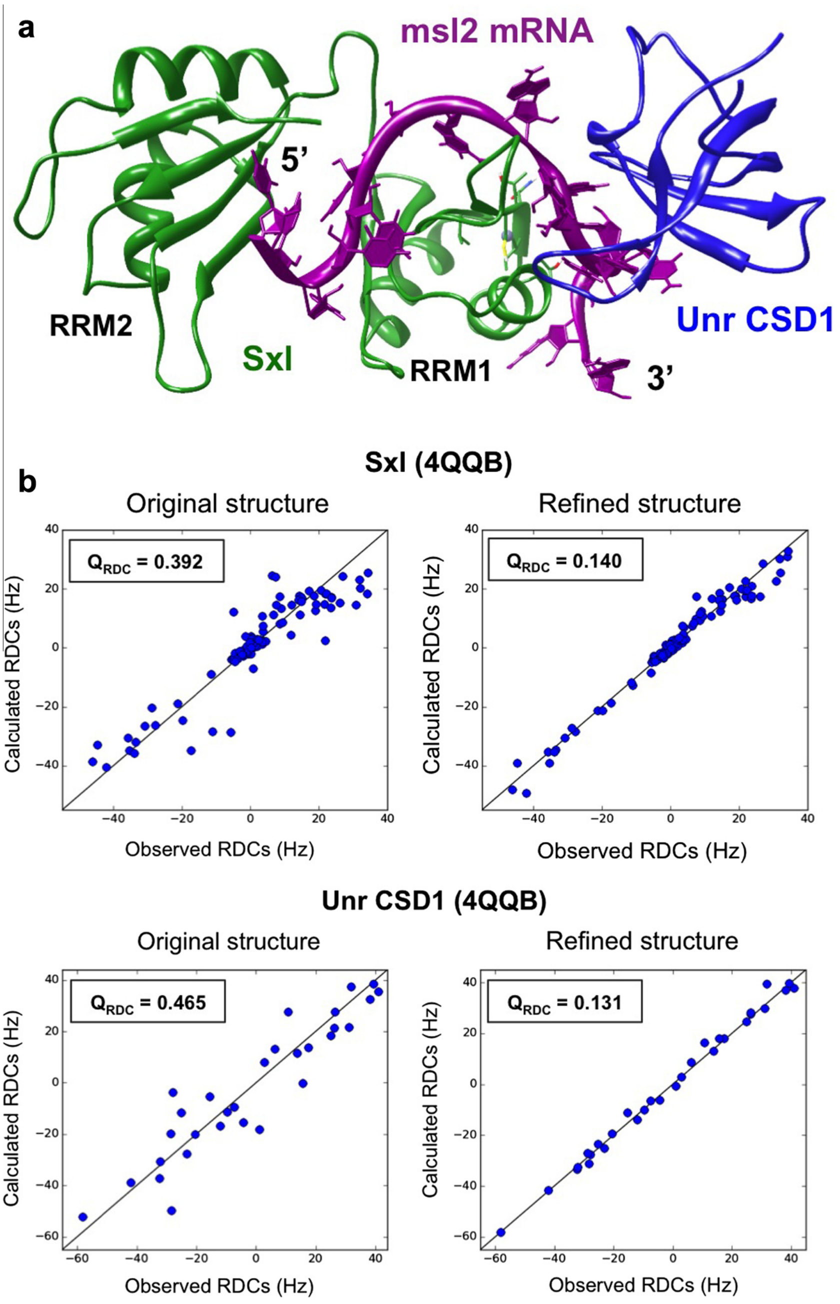


Fig. 12. (a) Structure of the ternary Sxl-Unr-*msl2*-mRNA regulatory complex; (b) correlation plot of the RDCs calculated for chains A and X of the 4QQB structure before and after refinement.

Scheme 1 was therefore applied so as to try to understand the origin of the discrepancies between the NMR data and the crystal structures. A REFMAC-NMR refinement of the individual CaM domains was first performed using both PCSs and RDCs for the N-terminal domain (the one bearing the metal ion), and RDCs only for the C-terminal domain. Note that the exclusive use of RDCs permits one to avoid any kind of assumption regarding the presence of interdomain mobility. Since the comparison of tensor magnitudes and alignments revealed the absence of extensive mobility or conformational rearrangement, a refinement of the full-length crystallographic structure was performed using both PCSs and RDCs. The refinement succeeded in reaching low *Q*-factors without increasing the free *R*-factor, thus providing a structure in agreement with both X-ray and NMR data (Fig. 9) (unpublished observations from the authors' laboratory).

This result should be compared with a previously performed analysis [144] where the initial discrepancies were explained in terms of mobility, so that the NMR data could be reproduced using a conformational ensemble. A careful assessment of the accuracy of the back-calculated tensors, and of their differences, outside the experimental error, is thus needed to reveal interdomain mobility.

3.2.4. Calmodulin–DAPK peptide complex (CaM–DAPKp)

Death-associated protein kinase (DAPK) provides a model for calcium-bound calmodulin (CaM)-dependent protein kinases (CaMKs). When DAPK is in a complex with CaM, the DAPK autoregulatory domain forms a seven-turn helix, through which it interacts with the DAPK catalytic domain. When bound to DAPK, CaM adopts a more extended conformation [191] than found in the crystal structures of CaM–CaMK peptide complexes.

In the case of CaM in complex with the DAPK peptide (DAPKp), neither the crystal structure of the complex (PDB code: 1YR5) [92] nor that with the full-length DAPK (PDB code: 2X0G) [191] exhibited satisfactory agreement with the NMR data. REFMAC-NMR refinement calculations could not provide any good agreement with PCSs and RDCs [94], even excluding residues affected by significant mobility (identified by relaxation experiments). Exclusion of the first helix of the N-terminal domain of CaM drastically reduced the disagreement, but not sufficiently to obtain an acceptable result. This indicates that in this complex the arrangement of the two CaM domains is globally different in solution compared to the crystal. Each CaM domain in the CaM–DAPKp complex could actually be refined using Xplor-NIH and NMR data only, by restraining the backbone dihedral angles to be as close as possible to those of the X-ray structure. In this case the REFMAC-NMR refinement would not be possible even for the single domains because of differences in the inter-helical orientations (see Section 3.1.4). The two domains could subsequently be positioned with respect to one another in order to superimpose the corresponding anisotropy tensors [92]. Interestingly, the final model shows a relative position of the two domains that is intermediate between the positions in the crystal structures determined for the full-length DAPK and the DAPK peptide (Fig. 10).

A more accurate analysis of the data reveals that the anisotropy tensors determined for the three metals from the C-terminal domain are marginally (a few percent) smaller than their N-terminal domain counterparts. The system was therefore subjected to an additional analysis using MaxOcc and its latest extension MaxOR (Maximum Occurrence of Regions) [151,192]. The previously refined structure was shown to have a MaxOcc of 92%, thus implying that other conformations, although with small weight, must be considered as contributors to a structural ensemble for a best fit of the data. The MaxOR analysis allowed us to identify the most compact structural ensemble (composed of rigid domain conformations) which can fully explain the experimental observables [151] (Fig. 11).

3.2.5. Sxl-Unr translation complex

An example of a system composed of multiple domains that has been successfully refined with REFMAC-NMR is provided by the ternary Sxl-Unr-*msl2*-mRNA regulatory complex, which consists of both RNA recognition motifs (RRMs) of Sxl, the first of five cold shock domains of Unr (CSD1), and an 18-mer single-stranded RNA derived from *msl2*-mRNA. Assembly of this complex is vital for female viability in fruit flies, as repression of *msl2*-mRNA by Sex-lethal (Sxl) and Upstream-of-N-Ras (Unr) prevents the formation of the dosage compensation complex, resulting in normal transcription of X-linked genes.

The crystal structure of the complex (PDB code: 4QQB) showed a modest agreement with the RDCs induced by Pf1 phage alignment medium [193]. This modest agreement is probably consistent with the extent of structural noise expected for a crystal structure at 2.8 Å resolution. The refinement revealed an almost invariant arrangement of the two domains constituting Sxl, and of the entire complex, with a relevant decrease in Q_{RDC} from 0.440 in the crystal structure to 0.144 in the refined structure. Moreover, the reduction in the structural noise of the Sxl-Unr complex by REFMAC-NMR allowed the identification of putatively mobile residues, and of minor discrepancies due to packing forces experienced in the crystal lattice [194] (Fig. 12).

Notably, the REFMAC-NMR refinement produced an effective improvement of the structural model with NMR data. As a proof that the major contribution to the improvement was not due to in-plane or out-of-plane distortions of the H^N–N bonds (even if within the standard limits), the protons were removed from the refined structure and added back using automatic methods available from common software (Molprobity [195]). Evaluation of the “reprotonated” structure showed that the agreement with NMR data was clearly maintained, whereas adding protons with the same program to the original structure did not provide any improvement.

4. Conclusions

The results of structural approaches based on the use of individual techniques such as X-ray and NMR have long provided the basis for the understanding of complex macromolecules. However, they increasingly appear somewhat limited; the results summarized here prove that an integrated approach based on the combination of data from these two radically different sources not only yields structures with improved quality (i.e., which agree with both sets of data), but also allows for a deeper understanding of the behaviour of the biomolecules in solution, revealing changes that may occur upon crystallization.

Acknowledgements

This work has been supported by Ente Cassa di Risparmio di Firenze, MIUR PRIN 2012SK7ASN, European Commission projects BioMedBridges No. 284209, pNMR No. 317127, and Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI) and supported by national member subscriptions. Specifically, we thank the EU ESFRI Instruct Core Centre CERM, Italy. GNM was funded by the Medical Research Council (grant No. MC_US_A025_0102).

References

[1] A. Bartesaghi, A. Merk, S. Banerjee, D. Matthies, X. Wu, J.L. Milne, S. Subramaniam, Science 348 (2015) 1147–1151.
[2] R.J. Kurland, B.R. McGarvey, J. Magn. Reson. 2 (1970) 286–301.
[3] J.A.B. Lohman, C. Maclean, Chem. Phys. 35 (1978) 269–274.
[4] J.R. Tolman, J.M. Flanagan, M.A. Kennedy, J.H. Prestegard, Proc. Natl. Acad. Sci. U.S.A. 92 (1995) 9279–9283.
[5] L. Banci, I. Bertini, K.L. Bren, M.A. Cremonini, H.B. Gray, C. Luchinat, P. Turano, J. Biol. Inorg. Chem. 1 (1996) 117–126.
[6] L. Banci, I. Bertini, G. Gori Savellini, A. Romagnoli, P. Turano, M.A. Cremonini, C. Luchinat, H.B. Gray, Proteins Struct. Funct. Genet. 29 (1997) 68–76.
[7] M. Allegrozzi, I. Bertini, M.B.L. Janik, Y.-M. Lee, G. Liu, C. Luchinat, J. Am. Chem. Soc. 122 (2000) 4154–4161.
[8] I. Bertini, M.B.L. Janik, G. Liu, C. Luchinat, A. Rosato, J. Magn. Reson. 148 (2001) 23–30.
[9] I. Bertini, M.B.L. Janik, Y.-M. Lee, C. Luchinat, A. Rosato, J. Am. Chem. Soc. 123 (2001) 4181–4188.
[10] I. Bertini, A. Donaire, B. Jiménez, C. Luchinat, G. Parigi, M. Piccioli, L. Poggi, J. Biomol. NMR 21 (2001) 85–98.
[11] I. Bertini, I. Gelis, N. Katsaros, C. Luchinat, A. Provenzani, Biochemistry 42 (2003) 8011–8021.
[12] I. Bertini, M. Fragai, Y.-M. Lee, C. Luchinat, B. Terni, Angew. Chem. Int. Ed. 43 (2004) 2254–2256.
[13] S. Balayssac, I. Bertini, A. Bhaumik, M. Lelli, C. Luchinat, Proc. Natl. Acad. Sci. U.S.A. 105 (2008) 17284–17289.
[14] I. Bertini, A. Bhaumik, G. De Paepe, R.G. Griffin, M. Lelli, J.R. Lewandowski, C. Luchinat, J. Am. Chem. Soc. 132 (2010) 1032–1040.
[15] C. Luchinat, G. Parigi, E. Ravera, M. Rinaldelli, J. Am. Chem. Soc. 134 (2012) 5006–5009.
[16] A. Bhaumik, C. Luchinat, G. Parigi, E. Ravera, M. Rinaldelli, Cryst. Eng. Commun. 15 (2013) 8639–8656.
[17] K. Barthelmes, A.M. Reynolds, E. Peisach, H.R.A. Jonker, N.J. DeNunzio, K.N. Allen, B. Imperiali, H. Schwalbe, J. Am. Chem. Soc. 133 (2011) 808–819.
[18] J. Wöhnert, K.J. Franz, M. Nitz, B. Imperiali, H. Schwalbe, J. Am. Chem. Soc. 125 (2003) 13338–13339.
[19] F. Rodriguez-Castañeda, P. Haberz, A. Leonov, C. Griesinger, Magn. Reson. Chem. 44 (2006) S10–S16.
[20] X.C. Su, T. Huber, N.E. Dixon, G. Otting, ChemBioChem 7 (2006) 1599–1604.
[21] M. John, G. Otting, ChemPhysChem 8 (2007) 2309–2313.
[22] G. Pintacuda, M. John, X.C. Su, G. Otting, Acc. Chem. Res. 40 (2007) 206–212.
[23] T. Zhuang, H.S. Lee, B. Imperiali, J.H. Prestegard, Protein Sci. 17 (2008) 1220–1231.
[24] X.C. Su, K. McAndrew, T. Huber, G. Otting, J. Am. Chem. Soc. 130 (2008) 1681–1687.

[25] X.C. Su, B. Man, S. Beeren, H. Liang, S. Simonsen, C. Schmitz, T. Huber, B.A. Messerle, G. Otting, J. Am. Chem. Soc. 130 (2008) 10486–10487.
[26] P.H.J. Keizers, A. Saragliadis, Y. Hiruma, M. Overhand, M. Ubbink, J. Am. Chem. Soc. 130 (2008) 14802–14812.
[27] D. Häussinger, J. Huang, S. Grzesiek, J. Am. Chem. Soc. 131 (2009) 14761–14767.
[28] X.C. Su, G. Otting, J. Biomol. NMR 46 (2010) 101–112.
[29] B. Man, X.C. Su, H. Liang, S. Simonsen, T. Huber, B.A. Messerle, G. Otting, Chem. Eur. J. 16 (2010) 3827–3832.
[30] M.A.S. Hass, P.H.J. Keizers, A. Blok, Y. Hiruma, M. Ubbink, J. Am. Chem. Soc. 132 (2010) 9952–9953.
[31] S. Das Gupta, X. Hu, P.H.J. Keizers, W.-M. Liu, C. Luchinat, M. Nagulapalli, M. Overhand, G. Parigi, L. Sgheri, M. Ubbink, J. Biomol. NMR 51 (2011) 253–263.
[32] T. Saio, K. Ogura, K. Shimizu, M. Yokochi, T.R. Burke Jr., F. Inagaki, J. Biomol. NMR 51 (2011) 395–408.
[33] J.D. Swarbrick, P. Ung, X.C. Su, A. Maleckis, S. Chhabra, T. Huber, G. Otting, B. Graham, Chem. Commun. (Camb.) 47 (2011) 7368–7370.
[34] J.D. Swarbrick, P. Ung, S. Chhabra, B. Graham, Angew. Chem. Int. Ed. Engl. 50 (2011) 4403–4406.
[35] I. Bertini, V. Calderone, L. Cerofolini, M. Fragai, C.F.G.C. Geraldes, P. Hermann, C. Luchinat, G. Parigi, J.M.C. Teixeira, FEBS Lett. 586 (2012) 557–567.
[36] W.M. Liu, P.H. Keizers, M.A. Hass, A. Blok, M. Timmer, A.J. Sarris, M. Overhand, M. Ubbink, J. Am. Chem. Soc. 134 (2012) 17306–17313.
[37] Y. Kobashigawa, T. Saio, M. Ushio, M. Sekiguchi, M. Yokochi, K. Ogura, F. Inagaki, J. Biomol. NMR 53 (2012) 53–63.
[38] L. Cerofolini, G.B. Fields, M. Fragai, C.F.G.C. Geraldes, C. Luchinat, G. Parigi, E. Ravera, D.I. Svergun, J.M.C. Teixeira, J. Biol. Chem. 288 (2013) 30659–30671.
[39] H. Yagi, A. Maleckis, G. Otting, J. Biomol. NMR 55 (2013) 157–166.
[40] K.L. Gempf, S.J. Butler, A.M. Funk, D. Parker, Chem. Commun. (Camb.) 49 (2013) 9104–9106.
[41] C.T. Loh, K. Ozawa, K.L. Tuck, N. Barlow, T. Huber, G. Otting, B. Graham, Bioconjug. Chem. 24 (2013) 260–268.
[42] A. Bax, N. Tjandra, Nat. Struct. Biol. 4 (1997) 254–256.
[43] A. Bax, N. Tjandra, J. Biomol. NMR 10 (1997) 289–292.
[44] N. Tjandra, J.G. Omichinski, A.M. Gronenborn, G.M. Clore, A. Bax, Nat. Struct. Biol. 4 (1997) 732–738.
[45] N. Tjandra, A. Bax, Science 278 (1997) 1111–1114.
[46] M.R. Hansen, L. Mueller, A. Pardi, Nat. Struct. Biol. 5 (1998) 1065–1074.
[47] J.A. Losonczi, J.H. Prestegard, J. Biomol. NMR 12 (1998) 447–451.
[48] P.S. Prosser, J.A. Losonczi, I.V. Shiyanovskaya, J. Am. Chem. Soc. 120 (1998) 11010–11011.
[49] B.E. Ramirez, A. Bax, J. Am. Chem. Soc. 120 (1998) 9106–9107.
[50] H. Wang, M. Eberstadt, E.T. Olejniczak, R.P. Meadows, S.W. Fesik, J. Biomol. NMR 12 (1998) 443–446.
[51] H.M. Al-Hashimi, H. Valafar, M. Terrell, E.R. Zartler, M.K. Eidsness, J.H. Prestegard, J. Magn. Reson. 143 (2000) 402–406.
[52] J.H. Prestegard, H.M. Al-Hashimi, J.R. Tolman, Q. Rev. Biophys. 33 (2000) 371–424.
[53] M. Zweckstetter, A. Bax, J. Biomol. NMR 20 (2001) 365–377.
[54] N.A. Lakomek, K.F. Walter, C. Fares, O.F. Lange, B.L. de Groot, H. Grubmuller, R. Bruschweiler, A. Munk, S. Becker, J. Meiler, C. Griesinger, J. Biomol. NMR 41 (2008) 139–155.
[55] Q. Zhang, R. Throolin, S.W. Pitt, A. Serganov, H.M. Al Hashimi, J. Am. Chem. Soc. 125 (2003) 10530–10531.
[56] M.P. Latham, P. Hanson, D.J. Brown, A. Pardi, J. Biomol. NMR 40 (2008) 83–94.
[57] E. Ravera, L. Salmon, M. Fragai, G. Parigi, H.M. Al-Hashimi, C. Luchinat, Acc. Chem. Res. 47 (2014) 3118–3126.
[58] L. Yao, J. Ying, A. Bax, J. Biomol. NMR 43 (2009) 161–170.
[59] M. Assfalg, I. Bertini, P. Turano, A.G. Mauk, J.R. Winkler, B.H. Gray, Biophys. J. 84 (2003) 3917–3923.
[60] I. Bertini, I.C. Felli, C. Luchinat, J. Magn. Reson. 134 (1998) 360–364.
[61] I. Bertini, C. Luchinat, G. Parigi, Progr. NMR Spectrosc. 40 (2002) 249–273.
[62] M. John, A.Y. Park, G. Pintacuda, N.E. Dixon, G. Otting, J. Am. Chem. Soc. 127 (2005) 17190–17191.
[63] G. Cornilescu, A. Bax, J. Am. Chem. Soc. 122 (2000) 10143–10154.
[64] M. Gochin, H. Roder, Protein Sci. 4 (1995) 296–305.
[65] L. Banci, I. Bertini, J.G. Huber, C. Luchinat, A. Rosato, J. Am. Chem. Soc. 120 (1998) 12903–12909.
[66] V. Gaponenko, S.P. Sarma, A.S. Altieri, D.A. Horita, J. Li, R.A. Byrd, J. Biomol. NMR 28 (2004) 205–212.
[67] I. Diaz-Moreno, A. Diaz-Quintana, M.A. De la Rosa, M. Ubbink, J. Biol. Chem. 280 (2005) 18908–18915.
[68] M.R. Jensen, D.F. Hansen, U. Ayna, R. Dagil, M.A. Hass, H.E. Christensen, J.J. Led, Magn. Reson. Chem. 44 (2006) 294–301.
[69] C. Schmitz, R. Vernon, G. Otting, D. Baker, T. Huber, J. Mol. Biol. 416 (2012) 668–677.
[70] H. Yagi, K.B. Pilla, A. Maleckis, B. Graham, T. Huber, G. Otting, Structure 21 (2013) 883–890.
[71] L. Banci, I. Bertini, M.A. Cremonini, G. Gori Savellini, C. Luchinat, K. Wüthrich, P. Güntert, J. Biomol. NMR 12 (1998) 553–557.
[72] L. Banci, I. Bertini, G. Cavallaro, A. Giachetti, C. Luchinat, G. Parigi, J. Biomol. NMR 28 (2004) 249–261.
[73] C. Schmitz, A.M. Bonvin, J. Biomol. NMR 50 (2011) 263–266.
[74] P. Güntert, Methods Mol. Biol. 278 (2004) 353–378.
[75] C.D. Schwieters, J. Kuszewski, N. Tjandra, G.M. Clore, J. Magn. Reson. 160 (2003) 65–73.
[76] M. Rinaldelli, A. Carlon, E. Ravera, G. Parigi, C. Luchinat, J. Biomol. NMR 61 (2015) 21–34.
[77] G.M. Clore, Proc. Natl. Acad. Sci. U.S.A. 97 (2000) 9021–9025.
[78] P. Dosset, J.C. Hus, D. Marion, M. Blackledge, J. Biomol. NMR 20 (2001) 223–231.
[79] H. Valafar, J.H. Prestegard, J. Magn. Reson. 167 (2004) 228–241.
[80] G. Pintacuda, A.Y. Park, M.A. Keniry, N.E. Dixonj, G. Otting, J. Am. Chem. Soc. 128 (2006) 3696–3702.
[81] R. Hulsker, M.V. Baranova, G.S. Bullerjahn, M. Ubbink, J. Am. Chem. Soc. 130 (2008) 1985–1991.
[82] B. Simon, T. Madl, C.D. Mackereth, M. Nilges, M. Sattler, Angew. Chem. Int. Ed. 49 (2010) 1967–1970.
[83] I. Bertini, C. Luchinat, G. Parigi, Coord. Chem. Rev. 255 (2011) 649–663.
[84] M. Longinetti, G. Parigi, L. Sgheri, J. Phys. A: Math. Gen. 35 (2002) 8153–8169.
[85] W. Andralojc, K. Berlin, D. Fushman, C. Luchinat, G. Parigi, E. Ravera, L. Sgheri, J. Biomol. NMR 62 (2015) 353–371.
[86] J.J. Chou, S. Li, A. Bax, J. Biomol. NMR 18 (2000) 217–227.
[87] N.R. Skrynnikov, N.K. Goto, D. Yang, W.-Y. Choy, J.R. Tolman, G.A. Mueller, L.E. Kay, J. Mol. Biol. 295 (2000) 1265–1273.
[88] F. Tian, H. Valafar, J.H. Prestegard, J. Am. Chem. Soc. 123 (2001) 11791–11796.
[89] J.J. Chou, D.A. Case, A. Bax, J. Am. Chem. Soc. 125 (2003) 8959–8966.
[90] T.S. Ulmer, B.E. Ramirez, F. Delaglio, A. Bax, J. Am. Chem. Soc. 125 (2003) 9179–9191.
[91] J.H. Prestegard, K.L. Mayer, H. Valafar, G.C. Benison, Methods Enzymol. 394 (2005) 175–209.
[92] I. Bertini, P. Kursula, C. Luchinat, G. Parigi, J. Vahokoski, M. Wilmanns, J. Yuan, J. Am. Chem. Soc. 131 (2009) 5134–5144.
[93] D. Gottstein, D.K. Kirchner, P. Güntert, J. Biomol. NMR 52 (2012) 351–364.
[94] M. Rinaldelli, E. Ravera, V. Calderone, G. Parigi, G.N. Murshudov, C. Luchinat, Acta Crystallogr. D D70 (2014) 958–967.
[95] A. Arseniev, P. Schultze, E. Wörgötter, W. Braun, G. Wagner, M. Vasak, J.H. Kägi, K. Wüthrich, J. Mol. Biol. 201 (1988) 637–657.
[96] B. Shaanan, A.M. Gronenborn, G.H. Cohen, G.L. Gilliland, B. Veerapandian, D.R. Davies, G.M. Clore, Science 257 (1992) 961–964.
[97] C.A. Schiffer, R. Huber, K. Wüthrich, W.F. Van Gunsteren, J. Mol. Biol. 241 (1994) 588–599.
[98] M. Miller, J. Lubkowski, J.K.M. Rao, A.T. Danishefsky, J.G. Omichinski, K. Sakaguchi, H. Sakamoto, E. Appella, A.M. Gronenborn, G.M. Clore, FEBS Lett. 399 (1996) 166–170.
[99] M.L. Raves, J.F. Doreleijers, H. Vis, C.E. Vorgias, K.S. Wilson, R. Kaptein, J. Biomol. NMR 21 (2001) 235–248.
[100] M. Tang, L.J. Sperling, D.A. Berthold, C.D. Schwieters, A.E. Nesbitt, A.J. Nieuwkoop, R.B. Gennis, C.M. Rienstra, J. Biomol. NMR 51 (2011) 227–233.
[101] D.W. Hoffman, C.S. Cameron, C. Davies, S.W. White, V. Ramakrishnan, J. Mol. Biol. 264 (1996) 1058–1071.
[102] J.A. Chao, J.R. Williamson, Structure 12 (2004) 1165–1176.
[103] O.F. Lange, N.-A. Lakomek, C. Farès, G.F. Schröder, K.F.A. Walter, S. Becker, J. Meiler, H. Grubmüller, C. Griesinger, B.L. de Groot, Science 320 (2008) 1471–1475.
[104] G.M. Clore, C.D. Schwieters, J. Am. Chem. Soc. 126 (2004) 2923–2938.
[105] K. Lindorff-Larsen, R.B. Best, M.A. DePristo, C.M. Dobson, M. Vendruscolo, Nature 433 (2005) 128–132.
[106] A.S. Maltsev, A. Grishaev, J. Roche, M. Zasloff, A. Bax, J. Am. Chem. Soc. 136 (2014) 3752–3755.
[107] M. Zweckstetter, A. Bax, J. Biomol. NMR 23 (2002) 127–137.
[108] S. Vijay-Kumar, C.E. Bugg, W.J. Cook, J. Mol. Biol. 194 (1987) 531–544.
[109] M. Ottiger, F. Delaglio, A. Bax, J. Magn. Reson. 131 (1998) 373–378.
[110] J.P. Derrick, D.B. Wigley, J. Mol. Biol. 243 (1994) 906–918.
[111] D.S. Berkholz, M.V. Shapovalov, R.L. Dunbrack Jr., P.A. Karplus, Structure 17 (2009) 1316–1325.
[112] M.W. MacArthur, J.M. Thornton, J. Mol. Biol. 264 (1996) 1180–1195.
[113] L. Esposito, S.A. De, A. Zagari, L. Vitagliano, J. Mol. Biol. 347 (2005) 483–487.
[114] D.S. Berkholz, C.M. Driggers, M.V. Shapovalov, R.L. Dunbrack Jr., P.A. Karplus, Proc. Natl. Acad. Sci. U.S.A. 109 (2012) 449–453.
[115] G.N. Murshudov, P. Skubàk, A.A. Lebedev, N.S. Pannu, R.A. Steiner, R.A. Nicholls, M.D. Winn, F. Long, A.A. Vagin, Acta Crystallogr. D 67 (2011) 355–367.
[116] M.D. Winn, C.C. Ballard, K.D. Cowtan, E.J. Dodson, P. Emsley, P.R. Evans, R.M. Keegan, E.B. Krissinel, A.G.W. Leslie, A. McCoy, S.J. McNicholas, G.N. Murshudov, N.S. Pannu, E.A. Potterton, H.R. Powell, R.J. Read, A.A. Vagin, K. S. Wilson, Acta Crystallogr. D 67 (2011) 235–242.
[117] H.M. McConnell, R.E. Robertson, J. Chem. Phys. 29 (1958) 1361–1365.
[118] M.D. Kemple, B.D. Ray, K.B. Lipkowitz, F.G. Prendergast, B.D.N. Rao, J. Am. Chem. Soc. 110 (1988) 8275–8287.
[119] M. Zweckstetter, Nat. Protoc. 3 (2008) 679–690.
[120] C. Camilloni, M. Vendruscolo, J. Phys. Chem. B 119 (2015) 653–661.
[121] D. Shishmarev, G. Otting, J. Biomol. NMR 56 (2013) 203–216.
[122] A.M. Bonvin, A.T. Brunger, J. Biomol. NMR 7 (1996) 72–76.
[123] W.-Y. Choy, J.D. Forman-Kay, J. Mol. Biol. 308 (2001) 1011–1032.
[124] D.I. Svergun, M.V. Petoukhov, M.H.J. Koch, Biophys. J. 80 (2001) 2946–2953.
[125] R. Burgi, J. Pitera, W.F. Van Gunsteren, J. Biomol. NMR 19 (2001) 305–320.
[126] J. Iwahara, C.D. Schwieters, G.M. Clore, J. Am. Chem. Soc. 126 (2004) 5879–5896.
[127] I. Bertini, C. Del Bianco, I. Gelis, N. Katsaros, C. Luchinat, G. Parigi, M. Peana, A. Provenzani, M.A. Zoroddu, Proc. Natl. Acad. Sci. U.S.A. 101 (2004) 6841–6846.
[128] M. Blackledge, Progr. NMR Spectrosc. 46 (2005) 23–61.

[129] M. Fragai, C. Luchinat, G. Parigi, Acc. Chem. Res. 39 (2006) 909–917.
[130] D.D. Boehr, R. Nussinov, P.E. Wright, Nat. Chem. Biol. 5 (2009) 954.
[131] Y.E. Ryabov, D. Fushman, Magn. Reson. Chem. 44 (2006) S143–S151.
[132] Y. Chen, S.L. Campbell, N.V. Dokholyan, Biophys. J. 93 (2007) 2300–2306.
[133] P. Bernadò, E. Mylonas, M.V. Petoukhov, M. Blackledge, D.I. Svergun, J. Am. Chem. Soc. 129 (2007) 5656–5664.
[134] I. Bertini, Y.K. Gupta, C. Luchinat, G. Parigi, M. Peana, L. Sgheri, J. Yuan, J. Am. Chem. Soc. 129 (2007) 12786–12794.
[135] L. Nodet, L. Salmon, V. Ozenne, S. Meier, M.R. Jensen, M. Blackledge, J. Am. Chem. Soc. 131 (2009) 17908–17918.
[136] A.C. Stelzer, A.T. Frank, M.H. Bailor, I. Andricioaei, H.M. Al Hashimi, Methods 49 (2009) 167–173.
[137] J. Huang, S. Grzesiek, J. Am. Chem. Soc. 132 (2010) 694–705.
[138] C.K. Fisher, A. Huang, C.M. Stultz, J. Am. Chem. Soc. 132 (2010) 14919–14927.
[139] Q. Bashir, A.N. Volkov, G.M. Ullmann, M. Ubbink, J. Am. Chem. Soc. 132 (2010) 241–247.
[140] I. Bertini, A. Giachetti, C. Luchinat, G. Parigi, M.V. Petoukhov, R. Pierattelli, E. Ravera, D.I. Svergun, J. Am. Chem. Soc. 132 (2010) 13553–13558.
[141] J. Rinnenthal, J. Buck, J. Ferner, A. Wacker, B. Furtig, H. Schwalbe, Acc. Chem. Res. 44 (2011) 1292–1301.
[142] J.R. Bothe, E.N. Nikolova, C.D. Eichhorn, J. Chugh, A.L. Hansen, H.M. Al Hashimi, Nat. Methods 8 (2011) 919–931.
[143] K. Berlin, C.A. Castañeda, D. Schneidman-Dohovny, A. Sali, A. Nava-Tudela, D. Fushman, J. Am. Chem. Soc. 135 (2013) 16595–16609.
[144] L. Russo, M. Maestre-Martinez, S. Wolff, S. Becker, C. Griesinger, J. Am. Chem. Soc. 135 (2013) 17111–17120.
[145] P. Guerry, L. Salmon, L. Mollica, J.L. Ortega Roldan, P. Markwick, N.A. van Nuland, J.A. McCammon, M. Blackledge, Angew. Chem. Int. Ed. Engl. 52 (2013) 3181–3185.
[146] L. Salmon, G. Bascom, I. Andricioaei, H.M. Al Hashimi, J. Am. Chem. Soc. 135 (2013) 5457–5466.
[147] A. Cavalli, C. Camilloni, M. Vendruscolo, J. Chem. Phys. 138 (2013) 094112.
[148] P. Kukic, C. Camilloni, A. Cavalli, M. Vendruscolo, J. Mol. Biol. 426 (2014) 1826–1838.
[149] M. Sanchez-Martinez, R. Crehuet, Phys. Chem. Chem. Phys. 16 (2014) 26030–26039.
[150] W. Boomsma, J. Ferkinghoff-Borg, K. Lindorff-Larsen, PLoS Comput. Biol. 10 (2014) e1003406.
[151] W. Andralojc, C. Luchinat, G. Parigi, E. Ravera, J. Phys. Chem. B 118 (2014) 10576–10587.
[152] D.A. Torchia, Prog. Nucl. Magn. Reson. Spectrosc. 84–85 (2015) 14–32.
[153] J.R. Tolman, H.M. Al-Hashimi, L.E. Kay, J.H. Prestegard, J. Am. Chem. Soc. 123 (2001) 1416–1424.
[154] Q. Zhang, H.M. Al-Hashimi, Nat. Methods 5 (2008) 243–245.
[155] M. Zweckstetter, A. Bax, J. Am. Chem. Soc. 122 (2000) 3791–3792.
[156] I. Bertini, M. Fragai, C. Luchinat, Curr. Pharm. Des. 15 (2009) 3592–3605.
[157] W. Bode, K. Maskos, Biol. Chem. 384 (2003) 863–872.
[158] G. Goldstein, M. Scheid, U. Hammerling, D.H. Schlesinger, H.D. Niall, E.A. Boyse, Proc. Natl. Acad. Sci. U.S.A. 72 (1975) 11–15.
[159] K.D. Wilkinson, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 15280–15282.
[160] D.C. Schwartz, M. Hochstrasser, Trends Biochem. Sci. 28 (2003) 321–328.
[161] D. Komander, M. Rape, Annu. Rev. Biochem. 81 (2012) 203–229.
[162] D. Fushman, K.D. Wilkinson, F1000 Biol. Rep. 3 (2011) 26.
[163] A. Hershko, A. Ciechanover, H. Heller, A.L. Haas, I.A. Rose, Proc. Natl. Acad. Sci. U.S.A. 77 (1980) 1783–1786.
[164] A. Ciechanover, H. Heller, S. Elias, A.L. Haas, A. Hershko, Proc. Natl. Acad. Sci. U.S.A. 77 (1980) 1365–1368.
[165] G.C. Stone, U. Sjobring, L. Bjorck, J. Sjoquist, C.V. Barber, F.A. Nardella, J. Immunol. 143 (1989) 565–570.
[166] A.M. Gronenborn, G.M. Clore, J. Mol. Biol. 233 (1993) 331–335.
[167] J.P. Derrick, D.B. Wigley, Nature 359 (1992) 752–754.
[168] F. Li, J.H. Lee, A. Grishaev, J. Ying, A. Bax, ChemPhysChem 16 (2015) 572–578.
[169] F.C. Stevens, Can. J. Biochem. Cell Biol. 61 (1983) 906–910.
[170] J.J. Chou, S. Li, C.B. Klee, A. Bax, Nat. Struct. Biol. 8 (2001) 990–997.
[171] H. Birkedal-Hansen, Curr. Opin. Cell Biol. 7 (1995) 728–735.
[172] G.J. Murphy, G. Murphy, J.J. Reynolds, FEBS Lett. 89 (1991) 4–7.
[173] W.G. Stetler-Stevenson, R. Hewitt, Sem. Cancer Biol. 7 (1996) 147–154.
[174] N. Borkakoti, J. Mol. Med. 78 (2000) 261–268.
[175] L. Steinman, Cell 85 (1996) 299–302.
[176] L.M. Coussens, B. Fingleton, L.M. Matrisian, Science 295 (2002) 2387–2392.
[177] M. Pavlaki, S. Zucker, Cancer Metastasis Rev. 22 (2003) 177–203.
[178] V. Lukacova, Y.F. Zhang, M. Mackov, P. Baricic, S. Raha, J.A. Calvo, S. Balaz, J. Biol. Chem. 279 (2004) 14194–14200.
[179] N. Borkakoti, Biochem. Soc. Trans. 32 (2004) 17–20.
[180] I. Bertini, V. Calderone, M. Cosenza, M. Fragai, Y.-M. Lee, C. Luchinat, S. Mangani, B. Terni, P. Turano, Proc. Natl. Acad. Sci. U.S.A. 102 (2005) 5334–5339.
[181] K. Maskos, W. Bode, Mol. Biotechnol. 25 (2003) 241–266.
[182] M. Longinetti, C. Luchinat, G. Parigi, L. Sgheri, Inv. Probl. 22 (2006) 1485–1502.
[183] M. Fragai, C. Luchinat, G. Parigi, E. Ravera, Coord. Chem. Rev. 257 (2013) 2652–2667.
[184] I. Bertini, M. Fragai, C. Luchinat, M. Melikian, M. Toccafondi, J.L. Lauer, G.B. Fields, J. Am. Chem. Soc. 134 (2012) 2100–2110.
[185] G. Barbato, M. Ikura, L.E. Kay, R.W. Pastor, A. Bax, Biochemistry 31 (1992) 5269–5278.
[186] J.L. Baber, A. Szabo, N. Tjandra, J. Am. Chem. Soc. 123 (2001) 3953–3959.
[187] I. Bertini, L. Ferella, C. Luchinat, G. Parigi, M.V. Petoukhov, E. Ravera, A. Rosato, D.I. Svergun, J. Biomol. NMR 53 (2012) 271–280.
[188] I. Bertini, C. Luchinat, M. Nagulapalli, G. Parigi, E. Ravera, Phys. Chem. Chem. Phys. 14 (2012) 9149–9156.
[189] M. Nagulapalli, G. Parigi, J. Yuan, J. Gsponer, S. Deraos, V.V. Bamm, G. Harauz, J. Matsoukas, M. de Planque, I.P. Gerothanassis, M.M. Babu, C. Luchinat, A.G. Tzakos, Structure 20 (2012) 522–533.
[190] F. Van Petegem, F.C. Chatelain, D.L. Minor Jr., Nat. Struct. Mol. Biol. 12 (2005) 1108–1115.
[191] D. de Diego, I.J. Kuper, N. Bakalova, P. Kursula, M. Wilmanns, Sci. Signal 3 (2010) ra6.
[192] W. Andralojc, E. Ravera, L. Salmon, G. Parigi, H.M. Al-Hashimi, C. Luchinat, Phys. Chem. Chem. Phys. (2015), http://dx.doi.org/10.1039/C5CP03993B.
[193] J. Hennig, C. Militti, G.M. Popowicz, I. Wang, M. Sonntag, A. Geerlof, F. Gabel, F. Gebaur, M. Sattler, Nature 515 (2014) 287–290.
[194] A. Carlon, E. Ravera, J. Hennig, G. Parigi, M. Sattler, C. Luchinat, J. Am. Chem. Soc. (2016), http://dx.doi.org/10.1021/jacs.5b11598.
[195] V.B. Chen, W.B. Arendall III, J.J. Headd, D.A. Keedy, R.M. Immormino, G.J. Kapral, L.W. Murray, J.S. Richardson, D.C. Richardson, Acta Crystallogr. D: Biol. Crystallogr. 66 (2010) 12–21.

Glossary of abbreviations

A: Alignment tensor
CaM: Calmodulin
CCP4: Collaborative computational project no. 4 Software for Macromolecular Crystallography (www.ccp4.ac.uk)
DAPk: Death-associated protein kinase
GB3: IgG-binding domain 3 of protein G
IQ: IQ-recognition motif of voltage-gated calcium channel
MaxOcc: Maximum Occurrence
MaxOR: Maximum Occurrence of Regions
MMP1: Matrix-metalloproteinase 1
NOE: Nuclear Overhauser Effect
PCS: Pseudocontact shift
PDB: RCSB Protein DataBank (www.rcsb.org/pdb/home/home.do)
RDC: Residual dipolar coupling
REFMAC5: REFinement of MACromolecular structures (software for crystallography, www2.mrc-lmb.cam.ac.uk/groups/murshudov/)
SAXS: Small-angle X-ray Scattering
Sxl-Unr: Sex-lethal and Upstream-of-N-Ras
χ: Magnetic susceptibility tensor in molecular coordinate frame

3.4 Improved Accuracy from Joint X-ray and NMR Refinement of a Protein–RNA Complex Structure

Azzurra Carlon[a], *Enrico Ravera*[a], *Janosch Hennig*[b,c,d], *Giacomo Parigi*[a], *Michael Sattler*[b,c],
Claudio Luchinat[a]

[a] Center for Magnetic Resonance (CERM), University of Florence,
via Sacconi 6, 50019 Sesto Fiorentino (FI), Italy
[b] *Center for Integrated Protein Science Munich (CIPSM) at Department Chemie, Technische Universität München, 85747 Garching, Germany*
[c] *Institute of Structural Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany*
[d] *Present address: Structural and Computational Biology Unit, EMBL Heidelberg, 69117 Heidelberg, Germany*

Published in:
Journal of the American Chemical Society ***138*** *(5), 1601-1*

JACS JOURNAL OF THE AMERICAN CHEMICAL SOCIETY
Article
pubs.acs.org/JACS

Improved Accuracy from Joint X-ray and NMR Refinement of a Protein−RNA Complex Structure

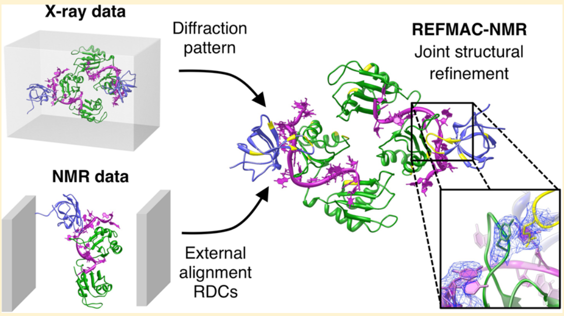
Azzurra Carlon,[†] Enrico Ravera,[†] Janosch Hennig,[‡,§,∥] Giacomo Parigi,[†] Michael Sattler,[*,‡,§] and Claudio Luchinat[*,†]

[†]Magnetic Resonance Center "CERM" and Department of Chemistry "Ugo Schiff", University of Florence and Magnetic Resonance Consortium (CIRMMP), Via L. Sacconi 6, 50019 Sesto Fiorentino, Firenze, Italy

[‡]Center for Integrated Protein Science Munich (CIPSM) at Department Chemie, Technische Universität München, 85747 Garching, Germany

[§]Institute of Structural Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany

Ⓢ *Supporting Information*



ABSTRACT: Integrated experimental approaches play an increasingly important role in structural biology, taking advantage of the complementary information provided by different techniques. In particular, the combination of NMR data with X-ray diffraction patterns may provide accurate and precise information about local conformations not available from average-resolution X-ray structures alone. Here, we refined the structure of a ternary protein−protein−RNA complex comprising three domains, Sxl and Unr, bound to a single-stranded region derived in the *msl2* mRNA. The joint X-ray and NMR refinement reveals that—despite the poor quality of the fit found for the original structural model—the NMR data can be largely accommodated within the uncertainty in the atom positioning (structural noise) from the primary X-ray data and that the overall domain arrangements and binding interfaces are preserved on passing from the crystalline state to the solution. The refinement highlights local conformational differences, which provide additional information on specific features of the structure. For example, conformational dynamics and heterogeneity observed at the interface between the CSD1 and the Sxl protein components in the ternary complex are revealed by the combination of NMR and crystallographic data. The joint refinement protocol offers unique opportunities to detect structural differences arising from various experimental conditions and reveals static or dynamic differences in the conformation of the biomolecule between the solution and the crystals.

■ INTRODUCTION

X-ray crystallography and NMR spectroscopy are the most popular techniques able to retrieve information at the atomic resolution level. The structural knowledge provided by these two techniques is complementary, since X-ray diffraction patterns are mainly derived from heavy atom contributions, whereas NMR structural restraints mostly involve hydrogen nuclei. Moreover, the crystalline and solution states are two distinct physical environments, which may influence the structural arrangement of macromolecular systems. Indeed, a number of studies document the presence of differences between solution and X-ray structures, where the crystalline state reports on structural snapshots or minor conformations that in some cases are not expected to exist in solution.[1−14] It is not hard to imagine that the presence of a crystal lattice may add additional constraints (e.g., crystal packing forces), which can induce changes in the intra- and intermolecular conformations and/or in the dynamic features of the system.

Obtaining a comprehensive data set for complete structural characterization by NMR spectroscopy is usually difficult and very time consuming, especially in the case of high molecular weight systems where extensive isotope labeling schemes need to be applied to enhance the spectral quality.[15−17] On the other hand, residual dipolar coupling (RDC) data can be collected, with relative ease, also for large systems and can be used to detect potential inconsistencies between solution and crystal states.[8,10,18] In case no detectable inconsistencies are found, a joint structural refinement using both NMR and X-ray data provides a method to obtain a more reliable structural model, which may disclose additional relevant information on its functional mechanisms.

Uncertainty related to the experimental measurements is an important issue that needs to be carefully analyzed to assess the significance of the inconsistencies found when NMR data are used in conjunction with an X-ray structural model. Extensive, long-standing, and controversial discussions have built on such inconsistencies.[11,19−25] In this regard, it is important to realize

Received: November 5, 2015
Published: January 13, 2016

DOI: 10.1021/jacs.5b11598

that besides the uncertainty related to NMR data measurements also the atomic coordinates in X-ray models may exhibit a non-negligible level of inaccuracy. Such inaccuracies, which may affect the positioning of different atomic moieties, mainly depend on the resolution of the X-ray reflections and on the structural refinement protocol employed.[11,26] Therefore, this so-called "structural noise" should be actively taken into account during the evaluation of inconsistencies, if any, between solution and crystal information.[26−28]

In the recent literature a number of approaches have been reported for refining X-ray structures with NMR data.[2,8,11−13,18,29−31] Most common refinement protocols consist in starting from an X-ray-derived structure and morphing the latter to achieve an acceptable agreement with the NMR data. This approach strongly relies on molecular libraries where the correct bonding geometry has to be kept (almost) completely rigid, because NMR measurements generally do not provide sufficient information to constrain the atom coordinates. Some of us have thus recently developed an approach (REFMAC-NMR) based on the simultaneous refinement of structural models against X-ray and NMR experimental data.[32] This allows for the joint use of the information about heavy atom positions, which often dominate X-ray reflections, together with the information about bond orientations for different nuclear pairs derived, in this particular case, from RDC data. Here, we show that REFMAC-NMR[32] refinement can be used to assess whether experimental NMR data can be explained by a structural model derived from X-ray crystallography within the accuracy of its diffraction pattern, and beyond the validation of the structure, RDCs can be used to provide a structure that complies with more than one data set and is thus more reliable. Moreover, we demonstrate that local conformational differences between the crystalline and solution states can be detected and exploited as useful hints on the functional mechanism of the system.

To provide a proof-of-principle of this workflow on a biologically relevant system, we assessed the recently reported crystal structure of the ternary Sxl−Unr-*msl2*-mRNA complex,[33] which consists of the two RNA recognition motifs (RRMs) of Sxl, the first of five cold shock domains of Unr (CSD1), and an 18-mer single-stranded RNA derived from *msl2*-mRNA. Assembly of this complex is vital for female viability in fruit flies, as translational repression of *msl2*-mRNA by Sex-lethal (Sxl) and Upstream-of-N-Ras (Unr) prevents the formation of the dosage compensation complex resulting in normal transcription of X-linked genes. The structure has unique protein−RNA and protein−protein interfaces that demonstrate how specificity and affinity for the cognate RNA is achieved by cooperative action of two distinct RNA binding proteins. The structure of this ternary complex constitutes an ideal test case for our purpose, as complementary NMR data are also available.[33,34] In the present manuscript we focus on the use of a set of residual dipolar couplings obtained for Pf1 phages alignment medium together with the available X-ray data at 2.8 Å resolution (PDB code 4QQB).

■ MATERIALS AND METHODS

Structural Refinement. Structural refinements were performed by the simultaneous use of the X-ray diffraction pattern and RDC data employing the recently developed program REFMAC-NMR.[32] The general approach consists of (i) a first minimization against the X-ray data alone, with an automatic setting of the *weight matrix* value (i.e., of the relative weights of geometry violations and X-ray violations), possibly followed by manual adjusting of the weight matrix to reduce the calculated rmsd of bond lengths, bond angles, and chiral volumes, if too large, and (ii) a second minimization performed including RDC restraints in order to decrease their Q factor. RDCs were included in the calculation only for residues which are not dynamic, most of which are part of secondary structure elements. In particular, the NMR restraints contribution (t) to the total minimized function is

$$t = k_{\mathrm{RDC}} \sum_i w_i[\max(|RDC_i^{\mathrm{calc}} - \mathrm{RDC}_i^{\mathrm{obs}}| - T_i, 0)]^2 \tag{1}$$

where T_i is the tolerance on each RDC value, w_i is its weight, and k_{RDC} is the overall weighting factor for RDCs. In tables the products of the k_{RDC} and w_i values will be indicated as RDC weight. The second minimization, besides optimization of the weight matrix value, requires optimization of the weights of the NMR data and of additional torsion angle restraints. Three further torsion angles were in fact introduced in the REFMAC library to restrain the planarity of the $\mathrm{O_i{-}C_i{-}N_{i+1}{-}C^{\alpha}_{i+1}}$, the $\mathrm{C_{i-1}{-}N_i{-}C^{\alpha}_i{-}H_i}$ (out of plane bending of $\mathrm{H^N{-}N}$ bonds), and the $\mathrm{C^{\alpha}_i{-}C_i{-}N_{i+1}{-}C^{\alpha}_{i+1}}$ dihedral angles (*pep1*, *pep2* and ω, respectively; force constants and tolerances used in the calculations are reported in Table S1). This was needed to avoid worsening of the deviations of geometric parameters from ideality by the inclusion of the NMR data in the refinement. Furthermore, overall weighting parameters over ideal geometries of all atoms involved (*weight refined_atoms*) or not involved (*weight other_atoms*) in the calculation of gradients and of the second derivatives corresponding to X-ray reflections were also introduced. Of note, bond distances of hydrogens in X-ray libraries are different from those in NMR libraries, because the hydrogen electron is not centered on the position of the nucleus but closer to the atom to which it is attached. Therefore, the coordinates of the hydrogens used for back-calculating the NMR restraints were recalculated by increasing the distances between the hydrogens and their binding nuclei to the values used in the AMBER[35,36] library ($\mathrm{H^N{-}N}$ distance of 1.020 Å). This correction for the evaluation of the NMR restraints does not affect the geometric restraints in the usual X-ray refinement, which considers hydrogen positions according to the standard crystallographic library.

Alignment Tensor Calculation. The alignment tensors and the agreement between experimental and back-calculated RDCs were computed using the FANTEN web application,[37] available in the WeNMR portal.[38] From the fit of the experimental RDCs to eq 2

$$\mathrm{RDC} = -\frac{3\mu_0 S_{\mathrm{LS}}}{4\pi^2}\frac{\gamma_{\mathrm{A}}\gamma_{\mathrm{B}}\hbar}{r_{\mathrm{AB}}^3}\left[D_{zz}\frac{2z_{\mathrm{AB}}^2 - x_{\mathrm{AB}}^2 - y_{\mathrm{AB}}^2}{2r_{\mathrm{AB}}^2} + (D_{xx} - D_{yy})\frac{x_{\mathrm{AB}}^2 - y_{\mathrm{AB}}^2}{2r_{\mathrm{AB}}^2} + D_{xy}\frac{2x_{\mathrm{AB}}y_{\mathrm{AB}}}{r_{\mathrm{AB}}^2} + D_{xz}\frac{2x_{\mathrm{AB}}z_{\mathrm{AB}}}{r_{\mathrm{AB}}^2} + D_{yz}\frac{2y_{\mathrm{AB}}z_{\mathrm{AB}}}{r_{\mathrm{AB}}^2}\right] \tag{2}$$

the program provides the five independent elements of the alignment tensor ($D_{xx} - D_{yy}$, D_{zz}, D_{xy}, D_{xz}, D_{yz}), from which the axial component of the tensor and its rhombicity, A and R, respectively, and the Euler angles defining the principal directions of the tensor can be derived (x_{AB}, y_{AB}, and z_{AB} are the components of the distance between the two coupled nuclei A and B). The similarity between tensors calculated from the best fit against different structures was assessed according to the following indicators: (1) the ratio of the axial components of the tensors; (2) the ratio of the tensor sizes, taking into account also their rhombicity; (3) the normalized dot product between the five independent elements of the alignment tensor. The first two criteria report on the similarity of the size of the tensors; the third criterion encodes information on their shape and orientation. In all cases, values close to 1 indicate good similarity between tensors.[19,39−41]

The experimental RDC data were taken from Hennig et al.;[33] X-ray data were taken from PDB accession code 4QQB.

^{15}N NMR Relaxation Measurements. ^{15}N R_1 and R_2 relaxation rates for Sxl at two different concentrations (0.25 and 1 mM) were measured at 800 MHz proton Larmor frequencies at 298 K.[42] $\{^1\mathrm{H}\}{-}^{15}\mathrm{N}$ heteronuclear NOE data for CSD1 were acquired at 0.3

mM protein concentration. For Sxl at 0.25 mM, R_1 relaxation rates were derived from measuring 14 different relaxation delays, including 3 duplicates for error estimation (21.6 (2×), 43.2, 86.4, 162, 345.6, 518.4 (2×), 669.6, 885.6 (2×), 1080, 1274.4, 1555.2, 1944, and 2376 ms). For $R_{1\rho}$ 12 different relaxation delays with two duplicates were measured (5 (2×), 10, 15, 25, 40, 55, 70, 80 (2×), 90, 110, and 140 ms). At 1 mM concentration of Sxl, R_1 was derived from 10 different relaxation delays recorded with two duplicates (21.6 (2×), 86.4, 162, 345.6, 518.4 (2×), 669.6, 885.6, 1274.4, 1728, and 2376 ms). R_2 was derived from measuring eight relaxation delays with two duplicates (10.88 (2×), 21.76, 32.64, 43.52 (2×), 54.4, 65.28, 76.16, and 87.04 ms). Data were fitted and analyzed using the software PINT.[43]

■ RESULTS

REFMAC-NMR Refinement. The crystal structure of the Sxl−Unr-*msl2*-mRNA complex was used to evaluate (a) the conformation and (b) the relative arrangement of the two Sxl RRM domains and CSD1 domain in solution against H^N−N and C−N RDC data.[33] From the best fit against a single alignment tensor, the agreement of the RDCs with the available X-ray structure (PDB code 4QQB) provides a *Q* factor of 0.440 (Tables 1 and S2). Despite the quality of the fit appearing to be rather modest, the result is in line with what can be expected from the X-ray data resolution for this system (2.8 Å).[26] The quality of the fit for the two complexes found in the asymmetric unit (chains A−P−X and B−C−Y) is slightly different, although the derived alignment tensor parameters are rather similar (Table 1).

Table 1. REFMAC-NMR Refinement Calculations Performed as for the Original Structure and Without (−NMR) and With (+NMR) the Inclusion of NMR Restraints[a]

PDB code 4QQB; resolution 2.80 Å					
		three tensors Sxl (Nter), Sxl (Cter), and CSD1	two tensors Sxl and CSD1	single-tensor Sxl−CSD1 complex	
parameters	original structure	−NMR	+NMR	+NMR	+NMR
R	0.198	0.198	0.201	0.201	0.201
R^{free}	0.236	0.234	0.236	0.236	0.235
RMSD bond length	0.006	0.006	0.009	0.009	0.009
RMSD bond angles	1.113	1.260	1.592	1.591	1.595
RMSD chiral volume	0.074	0.099	0.097	0.097	0.097
Q factor RDC	0.440		0.124	0.131	0.144

[a] Simultaneous refinement of X-ray and RDC data was performed using independent tensors for two domains of Sxl (RRM1, RRM2) and for CSD1 (three tensors) using independent tensor for full-length Sxl and CSD1 (two tensors) and using a unique tensor for the overall Sxl−CSD1 complex (single tensor).

To evaluate the presence of any intradomain conformational differences, RDC data were used to refine the conformations of the individual structural units of the protein components of the ternary complex (both RRM domains of Sxl and the CSD1 domain). REFMAC-NMR was employed for performing the structural refinement using the protocol previously described.[32] The peculiarity of this approach consists of taking into account the experimental uncertainty and coordinate precision of the X-ray data when RDC data are included as structural restraints (see Table 1). The joint refinement against both X-ray and RDC data allows for small but relevant changes of the atomic coordinates in order to satisfy the RDC data still being in agreement with the X-ray data. The joint refinement leads to an overall drop of the *Q* factor from 0.440 to 0.124, without any significant increase in the *R* or R^{free} values or in violations of geometrical constraints (rmsd for bond lengths, bond angles, and chiral volumes; see Table 1). These results suggest that the poor agreement of the RDC data initially observed for the original X-ray structure was mainly due to the presence of inaccuracy in atom positions and that no significant (i.e., outside the experimental error) structural differences in the conformation of the single domains constituting the complex exist between the crystalline and the solution states. Moreover, such improvement in the agreement of RDCs reveals that the additional information provided by these restraints assists in better defining the structure upon joint refinement, whereas the fact that the crystallographic R^{free} does not significantly increase with respect to *R* indicates the absence of over-refinement. Actually, the electron density map is even slightly changed after inclusion of NMR restraints (see Figure 4).

In order to test the presence of interdomain rearrangements, the tensors calculated for the individual units were compared with one another in terms of magnitude, alignment, and shape. These parameters turned out to be very similar for the two domains of Sxl (Table 2), pointing out that both domains could be refined by using the same tensor without any significant worsening in the agreement with RDC data. As expected, the refinement results obtained by imposing a single tensor for the two domains are satisfactory (with only a small increase in the *Q* factor from 0.124 to 0.131, see Table 1). This indicates that the Sxl domains in solution maintain the same relative rearrangement as observed in the crystal and that the presence of significant interdomain motion can be reasonably excluded. On the contrary, a notable difference in the magnitude of the alignment tensor was observed for CSD1. However, the shape and orientation of the tensor is almost indistinguishable from the tensor determined for Sxl (Table 2). The most likely explanation for this is a difference in the experimental conditions, e.g., a slight variation in alignment medium concentrations in the different samples used for the RDC

Table 2. Comparison between the Alignment Tensors Calculated Independently for the Two Domains of Sxl (RRM1, RRM2), and for Sxl and CSD1

comparison between tensors calculated for Sxl domains			
	magnitude of the axial component $A_{Sxl(RRM1)}/A_{Sxl(RRM2)}$	magnitude of the axial and rhombic components $(D_{zz} - D_{xx})_{Sxl(RRM1)}/(D_{zz} - D_{xx})_{Sxl(RRM2)}$	orientation and shape $(D_{Sxl(RRM1)}\cdot D_{Sxl(RRM2)})/(\|D_{Sxl(RRM1)}\|\|D_{Sxl(RRM2)}\|)$
chain A	0.98	1.02	0.95
chain B	1.02	1.07	0.98
comparison between tensors calculated for Sxl and CSD1			
	magnitude of the axial component A_{Sxl}/A_{CSD1}	magnitude of the axial and rhombic components $(D_{zz} - D_{xx})_{Sxl}/(D_{zz} - D_{xx})_{CSD1}$	orientation and shape $(D_{Sxl}\cdot D_{CSD1})/(\|D_{Sxl}\|\|D_{CSD1}\|)$
chains A, X	0.80	0.81	0.98
chains B, Y	0.79	0.80	0.99

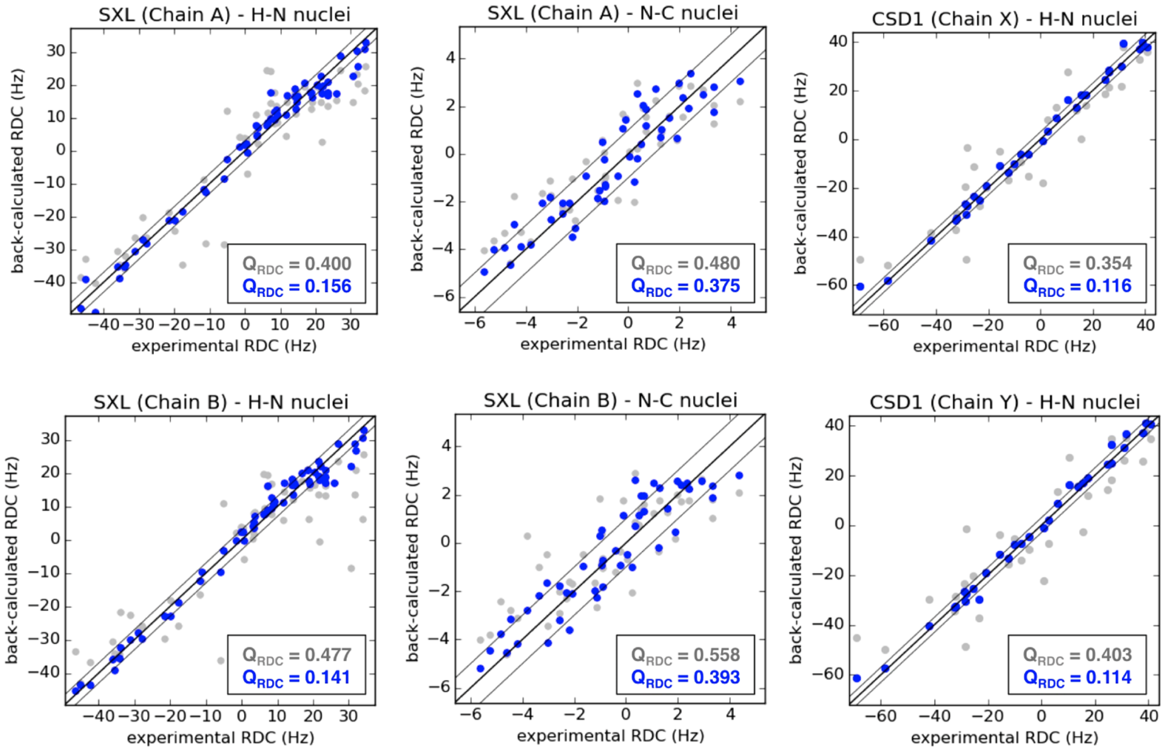


Figure 1. Correlation plots between experimental and back-calculated RDCs for the original structure (gray dots) and for the refined structure (blue dots).

measurements of the Sxl and CSD1 data involving Sxl- and CSD1-isotope-labeled complexes, respectively. This explanation is supported by the following considerations: (a) the tensor is reduced but has the same orientation and anisotropy, which is a hint of just a smaller alignment, and (b) it is unlikely that Sxl, which is larger, with a more anisotropic shape and intimately connected with the highly charged RNA molecule, could show smaller alignment than the smaller and more isotropic CSD1. Hence, a uniform scaling by an empirical factor 0.8 was applied to CSD1 RDC values in order to perform a new REFMAC-NMR refinement calculation using a single alignment tensor for the complete RDC data set. The refinement showed that the RDC Q factor increases only marginally (from 0.124 to 0.144, see Table 1) on passing from the use of three independent tensors for the three individual units of the Sxl−Unr complex to the use of a single tensor, remaining much smaller with respect to the Q factor of 0.440 calculated for the original X-ray model. No appreciable differences are observed for the structural statistics of the X-ray data (see Table 1).

In summary, these results indicate that the refined crystal structures provide a very good fit of the NMR data and thus represent also a good model of the Sxl−Unr complex in solution. The correlation plots reporting the agreement of the experimental RDCs with the refined structural model are shown in Figure 1. A good overall agreement is observed for both Sxl−Unr complexes (chains A−P−X and B−C−Y) found in the asymmetric unit of the crystal, with a slight preference for chains B and Y with respect to chains A and X. Taking into account the measurement errors (3 and 1 Hz for H^{N}−N and C−N, respectively), H^{N}−N RDCs collected for CSD1 and C−N RDCs for Sxl reveal optimal fit of the available data, with χ^2_{reduced} values of 1.003 (chain X), 0.97 (chain Y), 1.04 (chain A), and 1.12 (chain B); Sxl H^{N}−N RDCs show χ^2_{reduced} of 1.458 for chain A and 1.192 for chain B. Whereas C−N RDCs show agreement between calculated and observed data almost within the experimental error, it is immediately apparent that larger deviations exist in H^{N}−N RDCs. These deviations, which are discussed in detail in the Possible Differences between Crystal and Solution Structures section, show a peculiar effect of the joint refinement (see Table 3 and Figure S1). The refinement produces a sizable overall reduction of the differences between observed and calculated values, but still some of the values largely exceed the experimental uncertainty. Most notably, those residues that stand out after refinement were not violating more than others in the original structure,

Table 3. Residues Showing after Refinement a Deviation between Observed and Calculated Data Larger than the Standard Deviation Compared with Their Deviation before the Refinement (see Figure S1)

residue	original	refined
137	A, 0.300; B, 0.246	A, 1.353; B, 1.412
138	A, 0.261; B, 0.068	A, 0.989; B, 0.894
142	A, 1.364; B, 1.307	A, 1.920; B, 1.584
146	A, 2.234; B, 2.598	A, 1.260; B, 1.110
188	A, 0.404; B, 0.313	A, 1.317; B, 1.179
190	A, 0.506; B, 0.124	A, 0.966; B, 0.426
191	A, 1.859; B, 4.463	A, 2.384; B, 2.582
195	A, 0.208; B, 0.492	A, 0.875; B, 0.888
212	A, 0.812; B, 0.963	A, 2.925; B, 2.536
238	A, 1.138; B, 0.763	A, 1.967; B, 1.890
240	A, 0.762; B, 1.199	A, 1.743; B, 1.394
257	A, 1.264; B, 1.043	A, 2.560; B, 2.617

whereas they do after the violations of the others that have been successfully reduced by refinement.

Comparison of the refined structure with respect to the original model (Table 4) showed slightly improved fits of crystallographic data and Ramachandran scores with an increase in the percentage of residues belonging to the core (from 86.0% to 87.2%) and a decrease of those in the allowed (from 13.3% to 12.3%) and in the generously allowed regions (from 0.5% to 0.2%).

Table 4. Quality Evaluation for the Original Structure and for the REFMAC-NMR Refined Structure As Calculated by the wwPDB Validation Server (wwpdb-validation.wwpdb.org)

	original	refined
R^{free}	0.236	0.235
RSRZ outliers	5.4%	5.0%
Clashscore	4	5
Ramachandran outliers	core: 86.0%	core: 87.2%
	allowed: 13.3%	allowed: 12.3%
	generous: 0.5%	generous: 0.2%
	disallowed: 0.2%	disallowed: 0.2%
RNA backbone	0.35	0.34

Notably, REFMAC-NMR refinement produced an effective improvement of the structural model with NMR data. As a proof that the major contribution to the improvement is not due to simple in-plane or out-of-plane distortions of the H^N−N bonds (even if within the standard limits, see Figure S2a,b) the protons were removed from the refined structure and added back by using automatic methods available from common software (i.e., Molprobity[44] and MOLMOL 2K). Evaluation of the "reprotonated" structure shows that the agreement with NMR data is clearly maintained (Table 5), whereas adding protons with the same programs to the original structure does not provide any improvement. This demonstrates how much the uncertainty in heavy atom coordinates can affect the automatic positioning of hydrogen atoms and thus the orientation of H^N−N bonds to which RDCs mostly refer to.

Novel Insights from the Joint Refinement. REFMAC-NMR does not produce any global difference in the refined structure with respect to the original one (backbone coordinate rmsd of 0.066, 0.075, 0.093, and 0.094 Å for chains A, B, X, and Y, respectively). Indeed, complexes A−P−X and B−C−Y, which slightly differ in their relative positions (backbone coordinate rmsd of the secondary structure is 0.470 and 0.444 Å, respectively, between chains A and B and X and Y), get only slightly closer after the refinement (backbone coordinate rmsd of the secondary structure is 0.463 and 0.400 Å, respectively, between chains A and B and X and Y) (Table 6). However, by looking at the structural changes in more detail, it can be observed that small but significant local differences between the two complexes in the asymmetric unit exist (e.g., around residue 60 in CSD1, see Figure S3). Local differences are even more apparent in the backbone dihedral angles of the two complexes (Figure S4), which become more similar (average decrease of the rmsd in the secondary structure elements around 12% for both φ and ψ angles, see Table 6).

Table 6. RMSD between the Two Complexes in the Asymmetric Unit before and after Refinement

rmsd	original	refined
backbone atoms[a]	A−B, 0.470; X−Y, 0.444	A−B, 0.463; X−Y, 0.400
backbone angles[b]	φ A−B, 15.686; X−Y, 12.324	φ A−B, 13.015; X−Y, 8.156
	ψ A−B, 13.855; X−Y, 7.681	ψ A−B, 12.123; X−Y, 7.199
backbone atoms[c]	A−B, 0.508; X−Y, 0.488	A−B, 0.502; X−Y, 0.446

[a] rmsd calculated upon independent alignment of chains A and B and of chains X and Y. [b] rmsd calculated for the secondary structure elements. [c] rmsd calculated upon alignment of complexes A−P−X and B−C−Y.

Indeed, inclusion of RDC data mostly aids the local refinement of the two structures. Figures 2 and 3 report the backbone rmsd per residue between original and refined structures. Some differences of modest extent (never exceeding 0.2 Å) are present for Sxl (chains A and B). The most relevant changes involve residues N152, I230, and A271 for chain A and residues R146, N152, V185, T190, V191, Q239, and K240 for chain B, scattered in different regions of the two RRM domains. Slight differences are also detected in the recognition of secondary structure elements by the DSSP software,[45] the most significant involving a better definition of a new β-strand constituted by residues S285−L288 and of a helix-3 for residues E206−I208 (Figure 6). The improved definition of these structural elements is likely a consequence of additional information provided by the joint refinement.

In contrast, for the CSD1 domain (chains X and Y) structural variations are found to be all grouped in the well-defined loop region constituted by residues Y56−P63 (Figure 2), for which a conformational difference is observed upon joint refinement (Figures 3 and 6). Refinement of the loop conformation is reflected also in the slight rearrangement of some of the side chains, the most relevant one being R59 (Figure 4a and 4b). This residue is of particular importance as it forms an essential contact with Sxl Y164. Substitution of either residue by alanine severely impairs or abolishes the formation of the entire

Table 5. Comparison of the Q_{RDC} for the Sxl−CSD1 Complex for the Refined Structure when H^N Atoms Are Added by REFMAC-NMR by MolProbity and by MOLMOL 2K as Well as for the Original Structure

	refined			original
	H^N added by REFMAC-NMR Q_{RDC} (H^N−N, C−N)	H^N added by MolProbity Q_{RDC} (H^N−N, C−N)	H^N added by MOLMOL Q_{RDC} (H^N−N, C−N)	H^N added by MolProbity Q_{RDC} (H^N−N, C−N)
SXL (chain A) RRM1	0.144 (0.140, 0.390)	0.161 (0.157, 0.390)	0.197 (0.194, 0.389)	0.425 (0.425, 0.470)
SXL (chain A) RRM2	0.180 (0.176, 0.360)	0.195 (0.192, 0.360)	0.178 (0.175, 0.354)	0.366 (0.364, 0.491)
CSD1 (chain X)	0.116	0.135	0.189	0.354
SXL (chain B) RRM1	0.134 (0.129, 0.398)	0.155 (0.151, 0.398)	0.192 (0.189, 0.396)	0.560 (0.561, 0.478)
SXL (chain B) RRM2	0.161 (0.156, 0.379)	0.170 (0.165, 0.379)	0.182 (0.178, 0.376)	0.336 (0.331, 0.633)
CSD1 (chain Y)	0.114	0.136	0.200	0.403

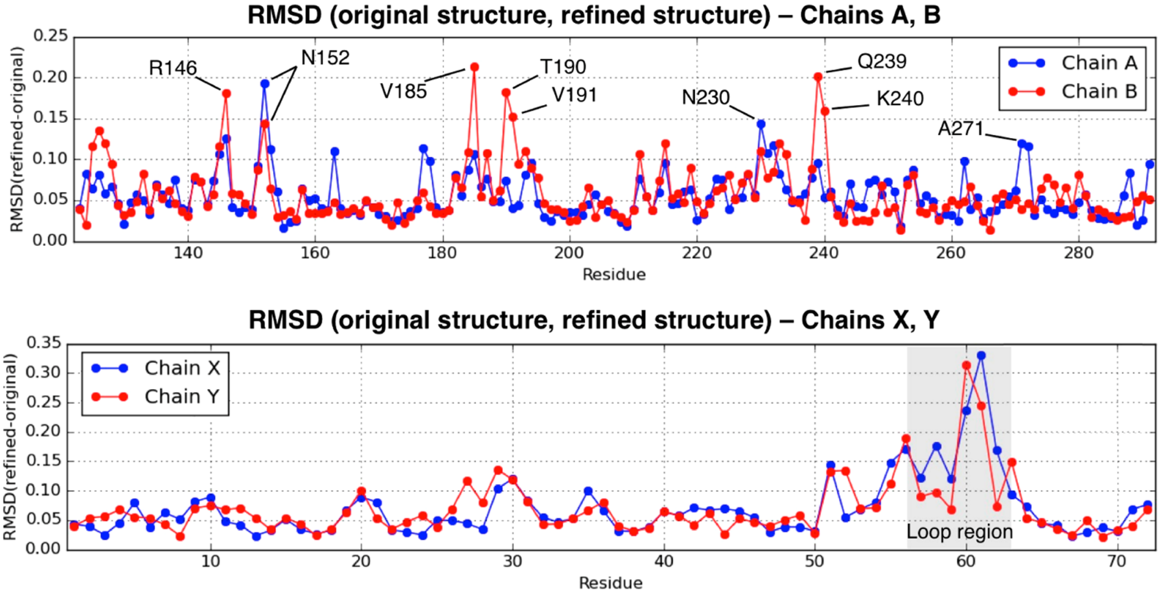


Figure 2. rmsd calculated from backbone C, C_α, and N nuclear positions between the original and the refined structure for chains A, B (top), and X, Y (bottom).

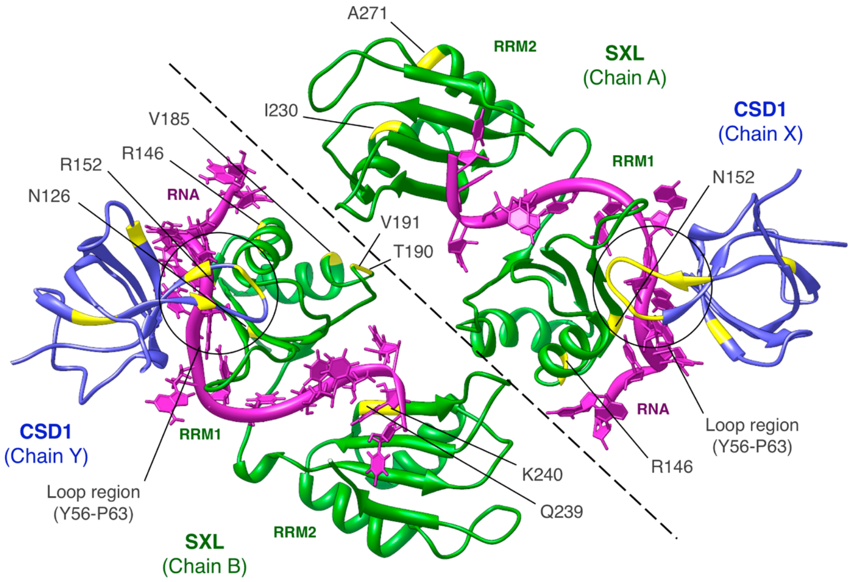


Figure 3. Two conformations of Sxl−Unr complex structure (chains A−P−X and B−C−Y) as found in the asymmetric crystal unit (pdb file 4QQB). Residues showing a rmsd in the backbone C, C_α, and N nuclear positions between the original and the refined structures above 0.12 Å are shown in yellow.

complex.[33] It is also interesting to observe that in the free CSD1 domain residues R58 and R59 exhibit slightly higher conformational flexibility on subnanosecond time scales as indicated by the low heteronuclear $\{^1H\}-^{15}N$ NOE data (Figure S5) when compared to other residues in the free state. This likely correlates with the observed conformational differences detected by REFMAC-NMR, suggesting that the refinement is more effective where electron density is less determined. Further indication of conformational rearrangement of CSD1 loop is provided by the fact that Y164 of Sxl, which interacts directly with R59 of CSD1, adopts two distinct conformations in the crystal by chains A−P−X and B−C−Y (Figure 4c and 4d). Indeed, while in the A−P−X complex the side chain of Y164 is stacked against the side chain of CSD1−R59, in the B−C−Y complex it is flipped to the other side and interacts with the RNA base of U7. The former is not observed in the absence of CSD1 but the latter while bound to *transformer* mRNA.[46] This indicates that both conformations of Y164 are energetically accessible from the solution conformation. Bringing together the information about the two partners, this suggests that the high flexibility of the CSD1 loop, coupled to the conformational heterogeneity detected through the side chain of Y164, may play an important role for complex formation. Of note, in this case REFMAC-NMR refinement gave access to useful information about the CSD1 loop, otherwise impossible to retrieve by NMR relaxation measure-

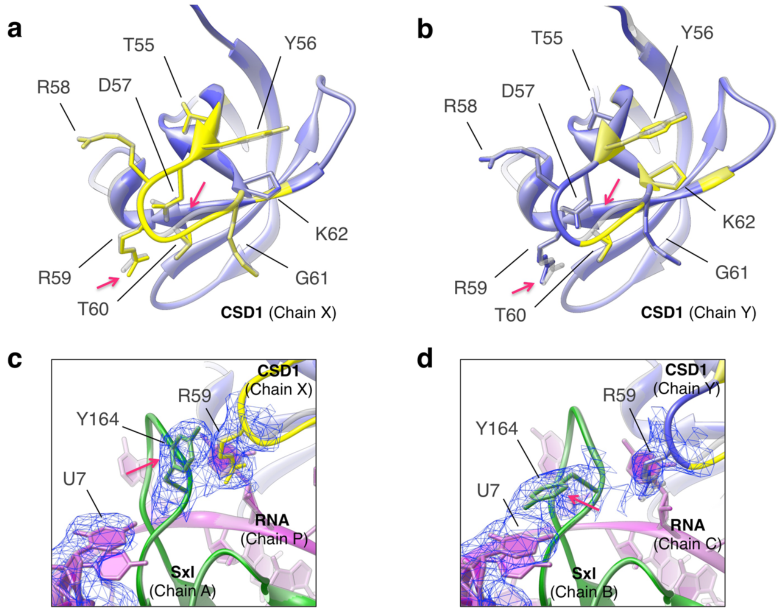


Figure 4. Detail of the CSD1 loop at the interface with RNA (a) for chains A–P–X and (b) B–C–Y. The original X-ray structure is shown in gray, whereas the refined structure is colored as in Figure 3. Relevant variations are indicated by red arrows. Zoomed view indicating conformational differences for CSD1 R59 and Sxl Y164 for chains (c) A–P–X and (d) B–C–Y. The flipping side chain of Y164 in the two conformations is indicated by a red arrow. Electron density for R59, Y164, and U7 and for neighboring residues/bases is shown in blue mesh lines at contour level of sigma 1.

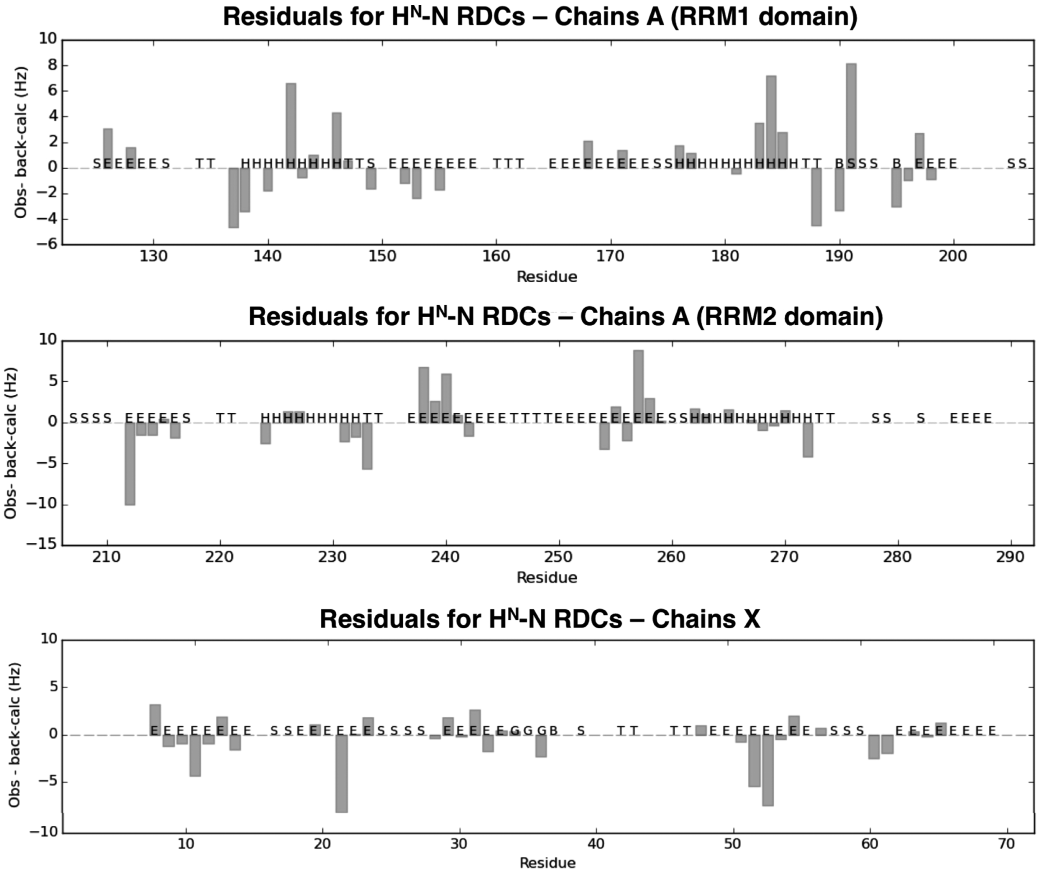


Figure 5. Residuals computed as difference between experimental and back-calculated RDCs for H^N–N nuclei of chains A and X.

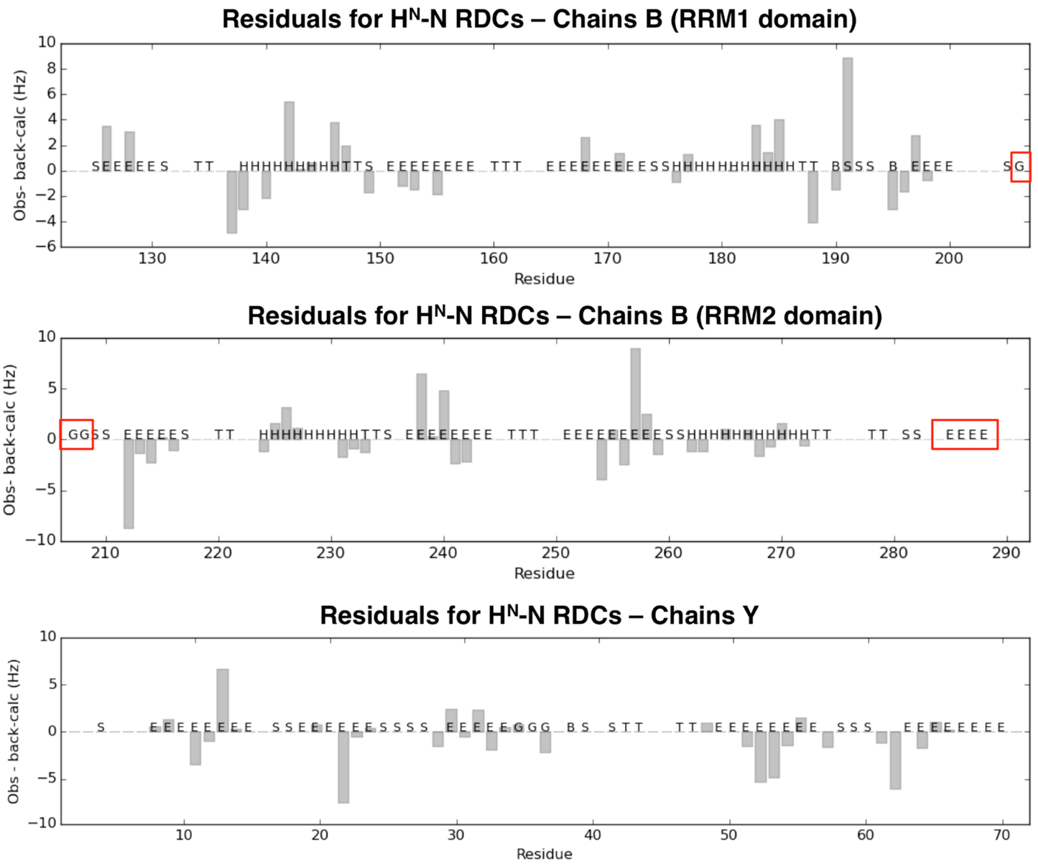


Figure 6. Residuals computed as difference between experimental and back-calculated RDCs for H^N−N nuclei of chains B and Y. Highlighted residues S285−L288 and E206−I208 are detected by DSSP software to change fold after REFMAC-NMR refinement.

ments of the complex due to line broadening of the amide resonance.

Possible Differences between Crystal and Solution Structures. Despite the refined structure presenting an overall very good fit for the available NMR data, a number of violations can still be observed, especially for some H^N−N RDC values belonging to Sxl (see section REFMAC-NMR Refinement). Figures 5−7 report the differences between experimental and back-calculated data, referred to as residuals, for the H^N−N RDC of both Sxl and CSD1. Interestingly, differences are found to be mainly clustered in two groups comprising residues T137, D138, Y142, and R146 (Figure 7b) and N212, V238, K240, and V257 (Figure 7c), which are located in the RRM1 and RRM2 domains of Sxl, respectively, and at the RNA binding interface. In particular, residues Y142 and R146 in RRM1 are of special significance, as chemical shift perturbations and mutational analyses confirmed their key role during RNA binding and complex formation[33] in vitro and during functional activity in vivo. This illustrates how the joint refinement provides novel structural insight for regions that are not well defined by the individual methods.

Minor but significant discrepancies are also observed for residues G188, T190, V191, and R195, located in the loop of Sxl (Figure 7d). This loop represents the region in which the two complexes A−P−X and B−C−Y are closer in space: the interaction of R192 in chain A with the loop K246−R250 of chain B results in the formation of two H bonds (R192−K246, R192−L247). Therefore, the presence of these interactions in the crystalline state may be at the origin of the structural differences observed for this region with respect to the solution structure. Of note, although at low concentrations this complex is monomeric, a concentration-dependent weak dimerization is observed for Sxl in solution, as indicated by the increased rotational correlation time (Figure S6). Consistently, residues around I189, including G188, T190, and V191 (relaxation rates of R195 could not be assessed), exhibit longer apparent rotational correlation times derived from ^{15}N R_2/R_1 relaxation rates at higher concentrations. During crystallization of the complex the concentration is even higher, and the same residues seem to be involved in the interaction between both complexes in the asymmetric unit. This interaction does not take place at lower concentrations at which RDCs were acquired and may thus explain the discrepancy observed. On the other hand, residues around K246 are more flexible at both concentrations. In the ternary complex studied here differences between crystal and solution data are observed for residues that interact with the 3′ region of the RNA. As in the cellular context the full-length RNA sequence extends beyond A20; additional factors might stabilize the RNA contacts.

CONCLUSIONS

Here, we show that joint X-ray and NMR refinement can be effectively used to probe if a crystal structure reflects the conformation found in solution at a residue-by-residue level (depending on the NMR data used). Moreover, the joint refinement can reveal local differences between solution and

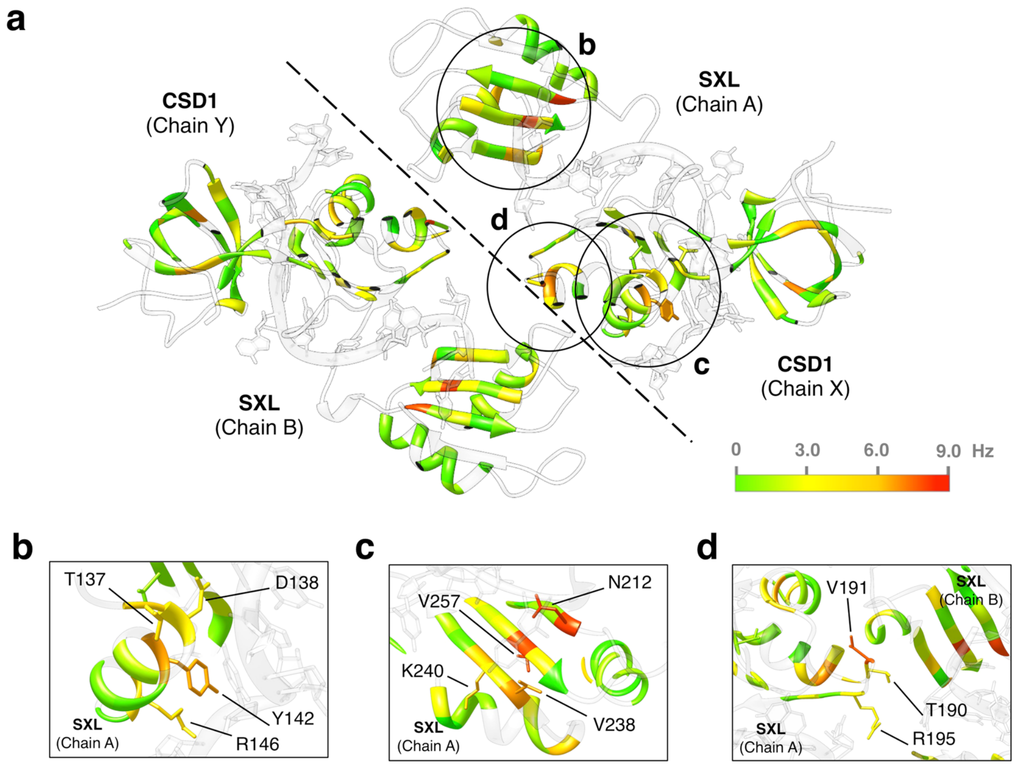


Figure 7. (a) Differences between experimental RDCs and values back-calculated from the original crystal structure for H^N–N interactions (in absolute value) are mapped onto two X-ray models of the Sxl–Unr complex according to the color code reported in the bar. (b–d) Zoomed views for distinct residues where large deviations (i.e., sizably exceeding the experimental uncertainty) are observed at the protein–RNA interface.

crystal state conformations and thus provide complementary information. Local conformational inaccuracies can arise from the uncertainty in the exact position of atoms within the electron density maps of X-rays, referred to as structural noise. These inaccuracies can be detected by NMR data, which provide complementary information about bond angles and moiety orientations. NMR data provide a direct experimental handle on these structural properties.

In the present example of Sxl–Unr translation regulatory complex, where an unsatisfactory fit of diamagnetic RDCs was obtained against the original X-ray structure, REFMAC-NMR refinement produced an effective improvement in the quality of the fit, with a drop of the Q factor from 0.440 to 0.144. This confirms that the solution and crystal structures of this complex are sufficiently close, and thus, the general conclusion drawn from analysis of the crystal structure are unaffected. On the other hand, the small but significant changes are not trivially due to adjustments of N–H bond vector orientation but rather reflect non-negligible backbone rearrangements. In turn, these backbone rearrangements, although small, imply some adjustments in the corresponding side chains. Comparison of the tensors calculated independently for the different subunits reasonably exclude interdomain motion effects. Moreover, the results from the step-by-step refinement confirm that both the intradomain conformations and the interdomain positions observed in the crystal are very good models also for the solution state. We also demonstrated that the reduction of the structural noise (i.e., of the uncertainty in the atomic positions within the electron density map, which imposes systematic errors in the calculated values of RDCs[26]) accomplished by the inclusion of RDCs in the structural refinement can reveal minor but interesting conformational differences between crystal and solution conformations, which help to better understand the structural biology of the studied complex. These few differences could only be spotted after successful reduction of the many differences that could be taken care of by a minor perturbation of the structure within the X-ray structure uncertainty. Indeed, joint refinement yields access to local structural differences that escaped detection when using the two methods separately and points to differences in the protein–RNA interface, which may be relevant for understanding the biological function of the complex. Detecting differences between solution and crystalline states can also help to rule out effects and artifacts from crystal packing and to identify interactions that may be important in the assembly of higher order complexes and thus guide follow-up studies. Thus, simultaneous refinement helps to understand phenomena observed in solution, which cannot be directly explained from the crystal structure alone.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b11598.

> Tables of force constants and tolerances used for the refinement calculations and of alignment tensors; figures of RDC residuals, dihedral angles, residue-by-residue backbone RMSD, heteronuclear $\{^1H\}$–^{15}N NOE values for the free CSD1, ^{15}N relaxation data for free Sxl (PDF)
> Refined PDB structure (PDB)

AUTHOR INFORMATION

Corresponding Authors

*sattler@helmholtz-muenchen.de
*claudioluchinat@cerm.unifi.it

Present Address

‖Structural and Computational Biology Unit, EMBL Heidelberg, 69117 Heidelberg, Germany

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by The Deutsche Forschungsgemeinschaft (DFG, SFB1035 and GRK1721 to M.S.), the Ente Cassa di Risparmio di Firenze (C.L.), MIUR PRIN 2012SK7ASN (C.L), European Commission (284209 to C.L., 317127 C.L. and G.P.), the EU ESFRI Instruct Core Centre CERM, and the Center for Integrated Protein Science Munich (CIPSM). J.H. acknowledges the EMBO for a long-term fellowship (ALTF-276-2010) and the Swedish Research Council (Vetenskapsrådet) for a postdoctoral fellowship. E.R. holds a FIRC fellowship "Guglielmina Locatello e Gino Mazzega" (17941).

REFERENCES

(1) Brunger, A. T. *Nat. Struct. Biol.* **1997**, *4* (Suppl), 862.
(2) Chou, J. J.; Li, S.; Klee, C. B.; Bax, A. *Nat. Struct. Biol.* **2001**, *8*, 990.
(3) Goto, N. K.; Skrynnikov, N. R.; Dahlquist, F. W.; Kay, L. E. *J. Mol. Biol.* **2001**, *308*, 745.
(4) Sikic, K.; Tomic, S.; Carugo, O. *Open Biochem. J.* **2010**, *4*, 83.
(5) Volkov, A. N.; Worrall, J. A. R.; Holtzmann, E.; Ubbink, M. *Proc. Natl. Acad. Sci. U. S. A.* **2006**, *103*, 18945.
(6) Ryabov, Y. E.; Fushman, D. *J. Am. Chem. Soc.* **2007**, *129*, 3315.
(7) Tang, C.; Schwieters, C. D.; Clore, G. M. *Nature* **2007**, *449*, 1078.
(8) Bertini, I.; Kursula, P.; Luchinat, C.; Parigi, G.; Vahokoski, J.; Wilmanns, M.; Yuan, J. *J. Am. Chem. Soc.* **2009**, *131*, 5134.
(9) Bashir, Q.; Volkov, A. N.; Ullmann, G. M.; Ubbink, M. *J. Am. Chem. Soc.* **2010**, *132*, 241.
(10) Mackereth, C. D.; Madl, T.; Bonnal, S.; Simon, B.; Zanier, K.; Gasch, A.; Rybin, V.; Valcárcel, J.; Sattler, M. *Nature* **2011**, *475*, 408.
(11) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125*, 9179.
(12) Tian, F.; Valafar, H.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 11791.
(13) Skrynnikov, N. R.; Goto, N. K.; Yang, D.; Choy, W.-Y.; Tolman, J. R.; Mueller, G. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *295*, 1265.
(14) Fraser, J. S.; Clarkson, M. W.; Degnan, S. C.; Erion, R.; Kern, D.; Alber, T. *Nature* **2009**, *462*, 669.
(15) Tugarinov, V.; Kay, L. E.; Ibraghimov, I.; Orekhov, V. Y. *J. Am. Chem. Soc.* **2005**, *127*, 2767.
(16) Riek, R.; Pervushin, K.; Wüthrich, K. *Trends Biochem. Sci.* **2000**, *25*, 462.
(17) Matzapetakis, M.; Turano, P.; Theil, E. C.; Bertini, I. *J. Biomol. NMR* **2007**, *38*, 237.
(18) Blackledge, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46*, 23.
(19) Tolman, J. R.; Al-Hashimi, H. M.; Kay, L. E.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123*, 1416.
(20) Meiler, J.; Prompers, J. J.; Peti, W.; Griesinger, C.; Bruschweiler, R. *J. Am. Chem. Soc.* **2001**, *123*, 6098.
(21) Meiler, J.; Peti, W.; Griesinger, C. *J. Am. Chem. Soc.* **2003**, *125*, 8072.
(22) Clore, G. M.; Schwieters, C. D. *J. Am. Chem. Soc.* **2004**, *126*, 2923.
(23) Lange, O. F.; Lakomek, N.-A.; Farès, C.; Schröder, G. F.; Walter, K. F. A.; Becker, S.; Meiler, J.; Grubmüller, H.; Griesinger, C.; de Groot, B. L. *Science* **2008**, *320*, 1471.
(24) Maltsev, A. S.; Grishaev, A.; Roche, J.; Zasloff, M.; Bax, A. *J. Am. Chem. Soc.* **2014**, *136*, 3752.
(25) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. *Nature* **2005**, *433*, 128.
(26) Zweckstetter, M.; Bax, A. *J. Biomol. NMR* **2002**, *23*, 127.
(27) Han, B.; Liu, Y.; Ginzinger, S.; Wishart, D. S. *J. Biomol. NMR* **2011**, *50*, 43.
(28) Li, F.; Lee, J. H.; Grishaev, A.; Ying, J.; Bax, A. *ChemPhysChem* **2015**, *16*, 572.
(29) Chou, J. J.; Li, S.; Bax, A. *J. Biomol. NMR* **2000**, *18*, 217.
(30) Prestegard, J. H.; Mayer, K. L.; Valafar, H.; Benison, G. C. *Methods Enzymol.* **2005**, *394*, 175.
(31) Simon, B.; Madl, T.; Mackereth, C. D.; Nilges, M.; Sattler, M. *Angew. Chem., Int. Ed.* **2010**, *49*, 1967.
(32) Rinaldelli, M.; Ravera, E.; Calderone, V.; Parigi, G.; Murshudov, G. N.; Luchinat, C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2014**, *70*, 958.
(33) Hennig, J.; Militti, C.; Popowicz, G. M.; Wang, I.; Sonntag, M.; Geerlof, A.; Gabel, F.; Gebaur, F.; Sattler, M. *Nature* **2014**, *515*, 287.
(34) Hennig, J.; Wang, I.; Sonntag, M.; Gabel, F.; Sattler, M. *J. Biomol. NMR* **2013**, *56*, 17.
(35) Bertini, I.; Case, D. A.; Ferella, L.; Giachetti, A.; Rosato, A. *Bioinformatics* **2011**, *27*, 2384.
(36) Case, D. A.; Darden, T. A.; Cheatham, T. E., III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kilossvary, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T., Gohlke, H.; Cai, Q.; Ye, X.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12*; University of California: San Francisco, CA, 2012.
(37) Rinaldelli, M.; Carlon, A.; Ravera, E.; Parigi, G.; Luchinat, C. *J. Biomol. NMR* **2015**, *61*, 21.
(38) Wassenaar, T. A.; van Dijk, M.; Loureiro-Ferreira, N.; van der Schot, G.; de Vries, S. J.; Schmitz, C.; van der Zwan, J.; Boelens, R.; Giachetti, A.; Ferella, L.; Rosato, A.; Bertini, I.; Herrmann, T.; Jonker, H. R. A.; Bagaria, A.; Jaravine, V.; Guntert, P.; Schwalbe, H.; Vranken, W. F.; Doreleijers, J. F.; Vriend, G.; Vuister, G. W.; Franke, D.; Kikhney, A.; Svergun, D. I.; Fogh, R. H.; Ionides, J.; Laue, E. D.; Spronk, C.; Jurksa, S.; Verlato, M.; Badoer, S.; Dal Pra, S.; Mazzucato, M.; Frizziero, E.; Bonvin, A. M. J. J. *J. Grid Computing* **2012**, *10*, 743.
(39) Losonczi, J. A.; Prestegard, J. H. *J. Biomol. NMR* **1998**, *12*, 447.
(40) Bertini, I.; Del Bianco, C.; Gelis, I.; Katsaros, N.; Luchinat, C.; Parigi, G.; Peana, M.; Provenzani, A.; Zoroddu, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 6841.
(41) Russo, L.; Maestre-Martinez, M.; Wolff, S.; Becker, S.; Griesinger, C. *J. Am. Chem. Soc.* **2013**, *135*, 17111.
(42) Farrow, N. A.; Muhandiram, R.; Singer, A. U.; Pascal, S. M.; Kay, C. M.; Gish, G.; Shoelson, S. E.; Pawson, T.; Forman-Kay, J. D.; Kay, L. E. *Biochemistry* **1994**, *33*, 5984.
(43) Ahlner, A.; Carlsson, M.; Jonsson, B. H.; Lundstrom, P. *J. Biomol. NMR* **2013**, *56*, 191.
(44) Chen, V. B.; Arendall, W. B., III; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr., Sect. D: Biol. Crystallogr.* **2010**, *66*, 12.
(45) Touw, W. G.; Baakman, C.; Black, J.; te Beek, T. A. H.; Krieger, E.; Joosten, R. P.; Vriend, G. *Nucleic Acids Res.* **2015**, *43*, D364−D368.
(46) Handa, N.; Nureki, O.; Kurimoto, K.; Kim, I.; Sakamoto, H.; Shin, D. I.; Muto, Y.; Yokoyama, S. *Nature* **1999**, *398*, 579.

DOI: 10.1021/jacs.5b11598

Improved accuracy from joint X-ray and NMR refinement of a protein-RNA complex structure

Azzurra Carlon,[1] Enrico Ravera,[1] Janosch Hennig,[2,3,4] Giacomo Parigi,[1] Michael Sattler,[2,3*] Claudio Luchinat[1*]

[1]Magnetic Resonance Center "CERM" and Department of Chemistry "Ugo Schiff", University of Florence and Magnetic Resonance Consortium (CIRMMP), Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy
[2]Center for Integrated Protein Science Munich (CIPSM) at Department Chemie, Technische Universität München, 85747 Garching, Germany
[3]Institute of Structural Biology, Helmholtz Zentrum München, 85764 Neuherberg, Germany
[4]Present address: Structural and Computational Biology Unit, EMBL Heidelberg, 69117 Heidelberg, Germany

Supporting Information

PDB code: 4QQB – Resolution: 2.80 Å					
Parameters	Original structure	**Three tensors** Sxl (Nter), Sxl (Cter), and CSD1		**Two tensors** Sxl and CSD1	**Single tensor** Sxl-CSD1 complex
		– NMR	+ NMR	+ NMR	+ NMR
R-value	0.198	0.198	0.201	0.201	0.201
R-free	**0.236**	**0.234**	**0.236**	**0.236**	**0.235**
RMSD bond length	0.006	0.006	0.009	0.009	0.009
RMSD bond angles	1.113	1.260	1.592	1.591	1.595
RMSD chiral volume	0.074	0.099	0.097	0.097	0.097
Q-factor RDC	**0.440**	**-**	**0.124**	**0.131**	**0.144**
RDC weight	-	-	Sxl H^N-N = 0.015 (tol = 1.0) Sxl C-N = 0.50 (tol = 2.0) CSD1 H^N-N = 0.009 (tol = 1.0)	Sxl H^N-N = 0.015 (tol = 1.0) Sxl C-N = 0.50 (tol = 2.0) CSD1 H^N-N = 0.009 (tol = 1.0)	Sxl H^N-N = 0.015 (tol = 1.0) Sxl C-N = 0.50 (tol = 2.0) CSD1 H^N-N = 0.009 (tol = 1.0)
Weight dihedral angles* (*pep1, pep2, ω*)	-	2.0, 0.8, 2.0	2.0, 0.8, 2.0	2.0, 0.8, 2.0	2.0, 0.8, 2.0
Weight matrix	0.01	0.002	0.002	0.002	0.002
Weight refined_atoms, Weight other_atoms	1.0, 1.0	1.0, 25.0	1.0, 25.0	1.0, 25.0	1.0, 25.0

[*] REFMAC-NMR additional restraints defined in the CCP4 library:

- *pep1* value of 180° with a tolerance of 1.0°
- *pep2* value of 180° with a tolerance of 1.0°
- *ω* value of 180° with a tolerance of 5.0°

Table S1: Force constants and tolerances used for the different refinement calculations performed for the Sxl-Unr complex.

Original structure – Tensor parameters				
	A	R	D^{HN} (Hz)	Q_{RDC}
SXL (chain A) and CSD1 (chain X)	-0.00189	-0.186	-21.644	0.392
SXL (chain B) and CSD1 (chain Y)	-0.00186	-0.188	-21.403	0.465

Table S2: Alignment tensors calculated for the original X-ray structures of the two complexes A-P-X and B-C-Y found in the asymmetric unit of the crystal.

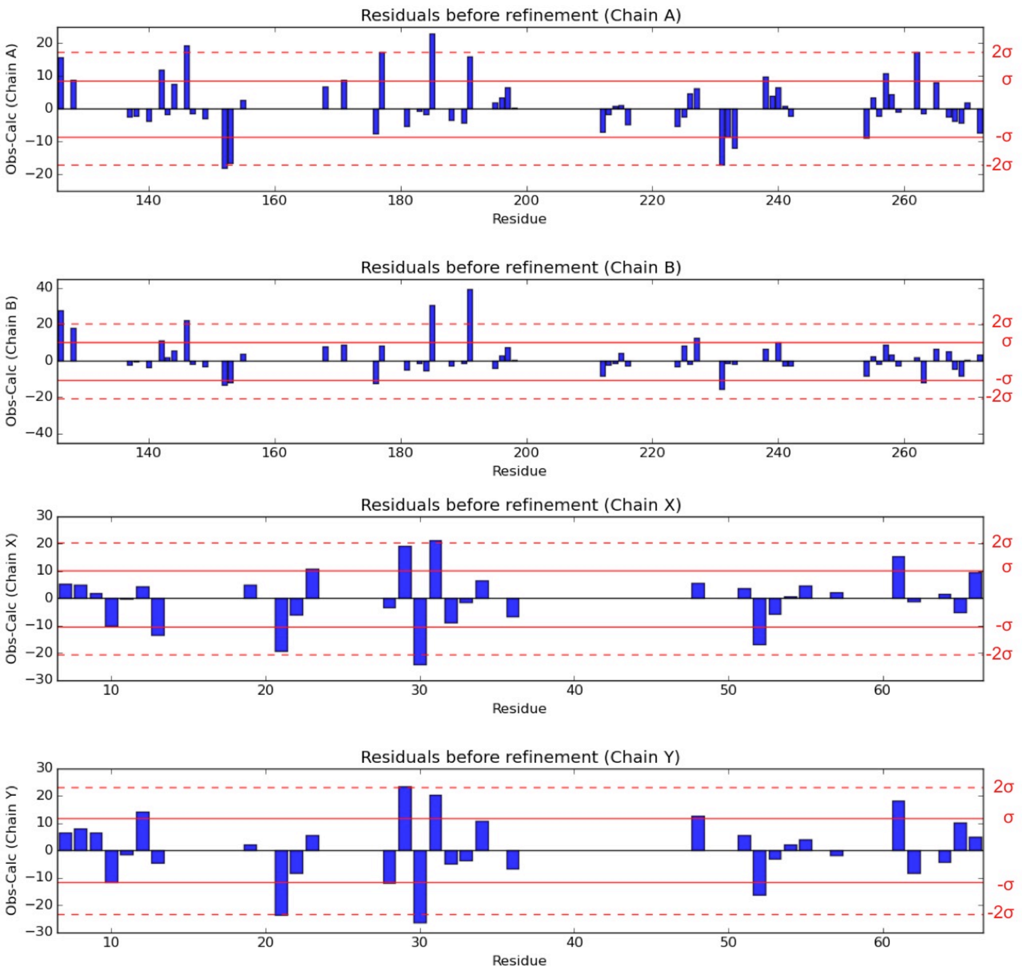


Fig.S1a: RDC residuals compared to their standard deviation before refinement.

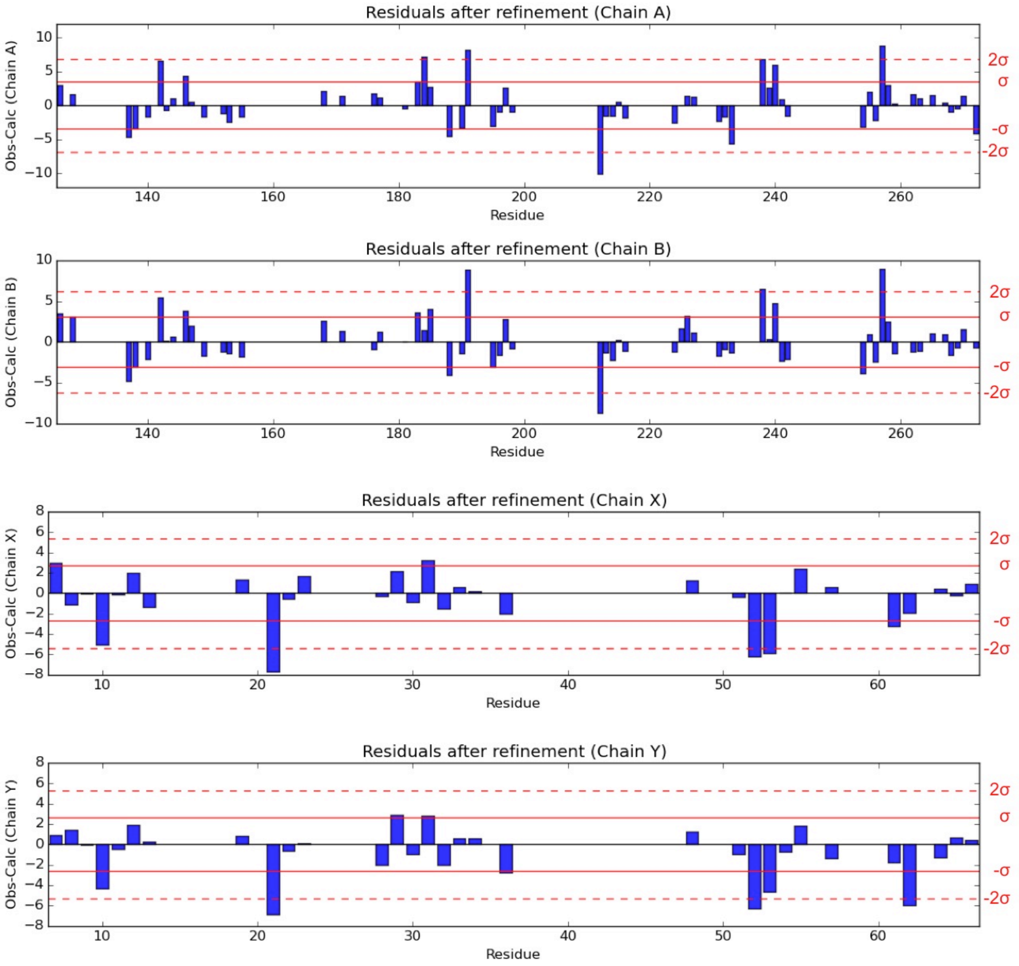
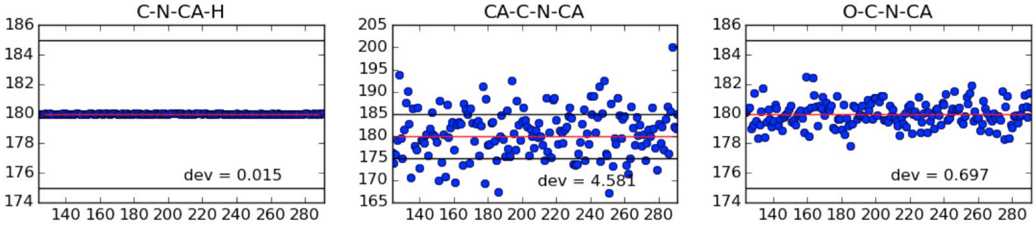
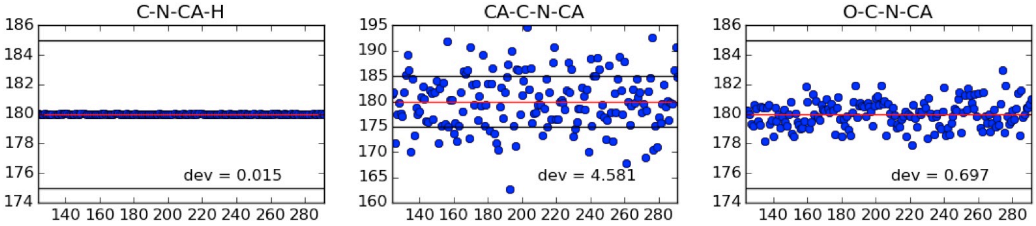


Fig.S1b: RDC residuals compared to their standard deviation after refinement.

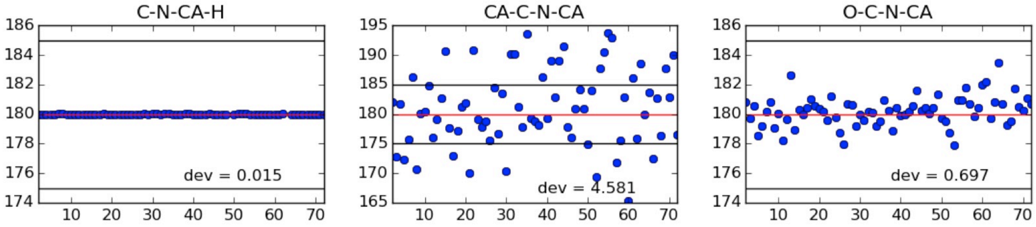
Sxl (Chain A) – Original structure



Sxl (Chain B) – Original structure



CSD1 (Chain X) – Original structure



CSD1 (Chain Y) – Original structure

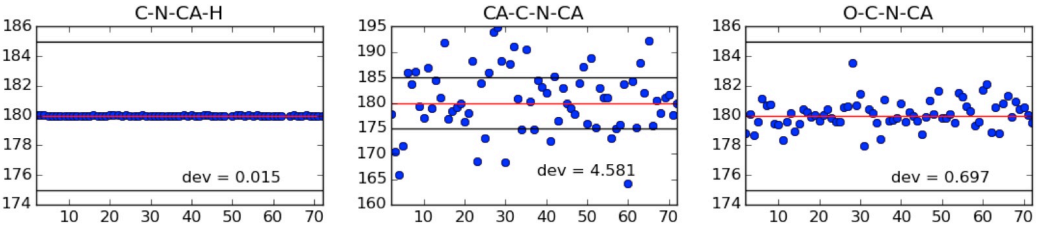
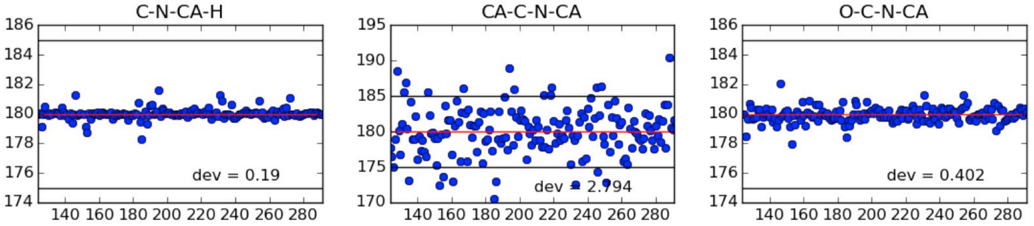
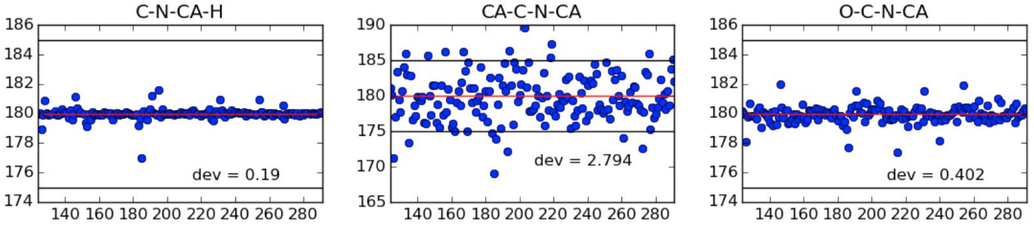


Fig.S2a: Reported values of pep2 (C_{i-1}-N_i- C^{α}_i-H_i), ω (C^{α}_i-C_i-N_{i+1}- C^{α}_{i+1}) and pep1 (O_i-C_i-N_{i+1}-C^{α}_i) for the original structures of Sxl (Chains A and B) and CSD1 (Chains X and Y).

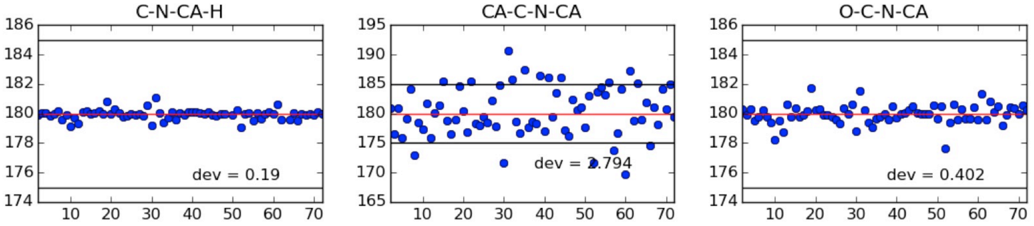
Sxl (Chain A) – Refined structure



Sxl (Chain B) – Refined structure



CSD1 (Chain X) – Refined structure



CSD1 (Chain Y) – Refined structure

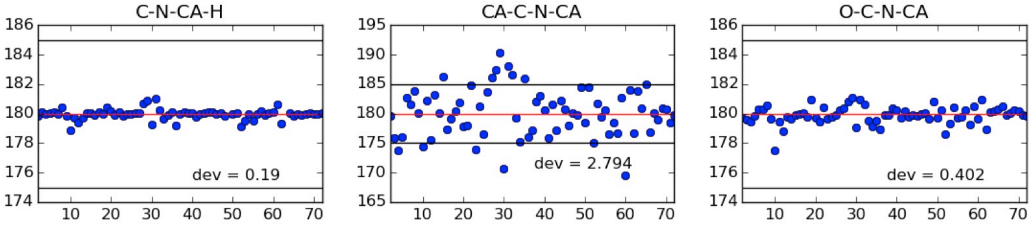


Fig.S2b: Reported values of pep2 (C_{i-1}-N_i- C^{α}_{i}-H_i), ω (C^{α}_{i}-C_i-N_{i+1}- C^{α}_{i+1}) and pep1 (O_i-C_i-N_{i+1}-C^{α}_{i}) for the REFMAC-NMR refined structures of Sxl (Chains A and B) and CSD1 (Chains X and Y).

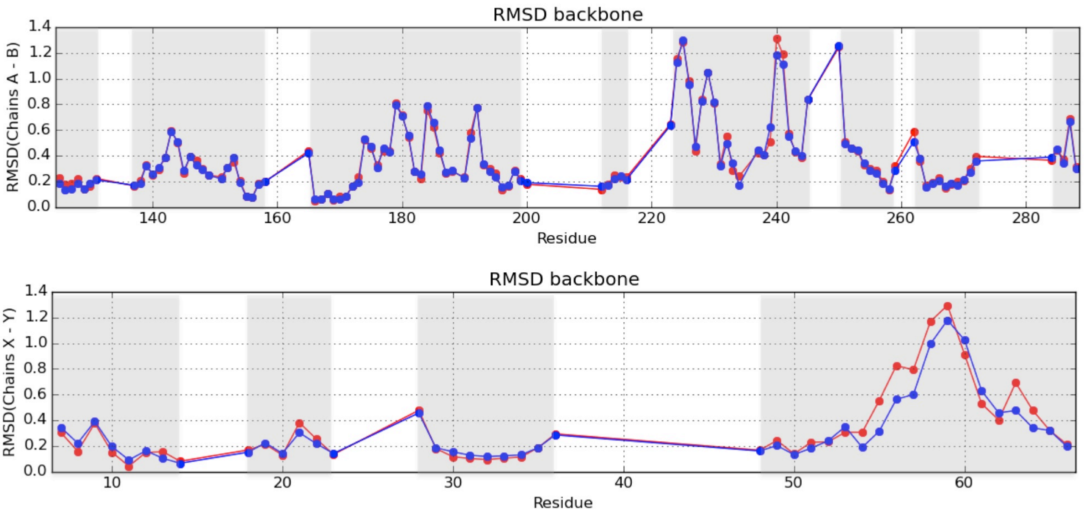


Fig.S3: Residue-by-residue backbone RMSD (for secondary structure elements or for residues not belonging to secondary structure elements for which RDCs were measured) between the two complexes in the asymmetric unit before (red) and after (blue) the refinement, for chains A and B (top), and chains X and Y (bottom) for the .

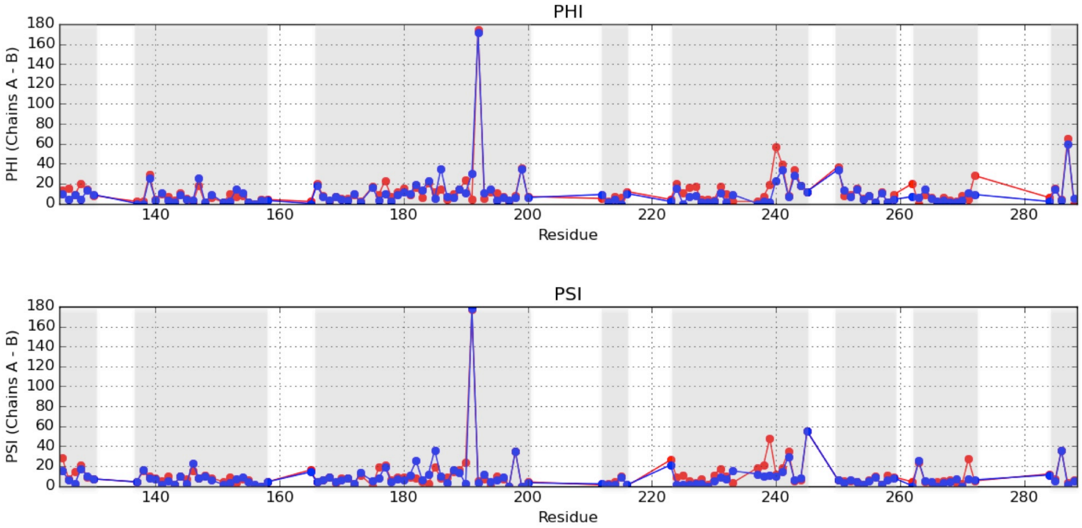


Fig.S4a: Residue-by-residue angle RMSD (for secondary structure elements or for residues not belonging to secondary structure elements for which RDCs were measured) between the chains A and B before (red) and after (blue) the refinement.

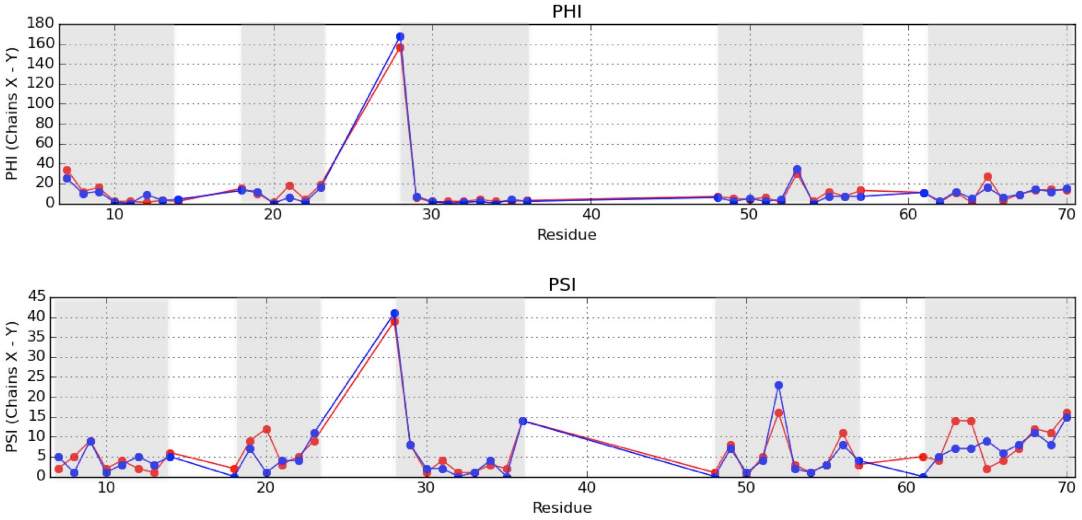


Fig.S4b: Residue-by-residue angle RMSD (for secondary structure elements or for residues not belonging to secondary structure elements for which RDCs were measured) between the chains X and Y before (red) and after (blue) the refinement.

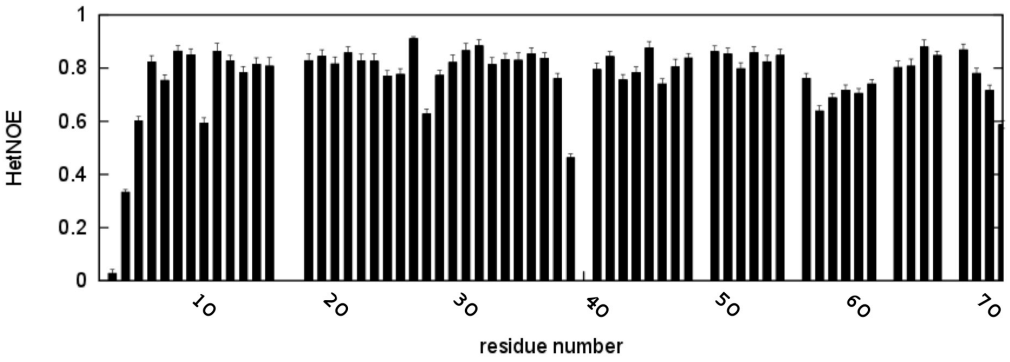


Fig.S5: Heteronuclear {^{1}H}-^{15}N NOE values for the free CSD1. The region around and including residues R58-R59 exhibits a lower NOE value, indicating a higher flexibility.

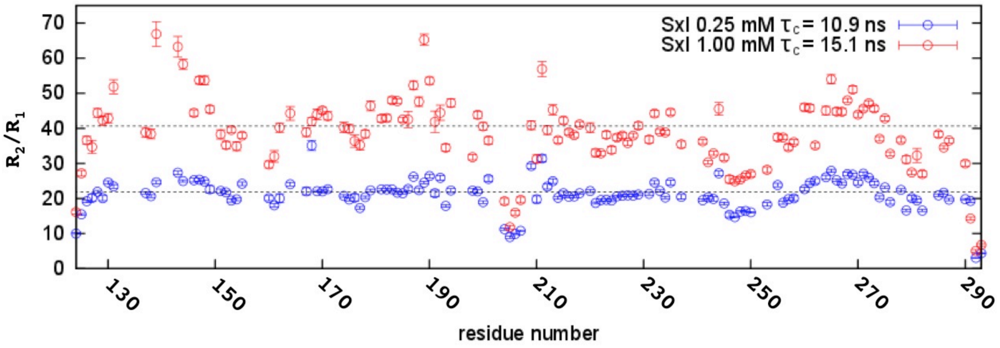


Fig.S6: ^{15}N relaxation data R_2/R_1) of free Sxl at 0.25 mM (blue) and 1 mM (red) concentrations. At higher concentration, Sxl has a higher rotational correlation time (τ_c = 15.1 ns at 1 mM compared to τ_c = 10.9 ns at 0.25 mM), indicating a weak dimerization propensity, which maybe mediated by residues around I189 and K246.

3.5 Impact of experimental uncertainties in protein mobility studies

Work in collaboration with: *Prof. Claudio Luchinat*[a,b] *and Prof. Christian Griesinger*[c]

[a] Center for Magnetic Resonance (CERM), University of Florence,
via Sacconi 6, 50019 Sesto Fiorentino (FI), Italy
[b] *Center for Integrated Protein Science Munich (CIPSM) at Department Chemie, Technische Universität München, 85747 Garching, Germany*
[c] *NMR Based Structural Biology, Max Planck Institute for Biophysical Chemistry, Am Fassberg 11 37077 Göttingen, Germany*

Manuscript in preparation

Impact of experimental uncertainties in protein mobility studies

Abstract

In the light of the recent developments of integrated structural biology (sometimes also referred to as “integrative”), there is an increasing need to bring together the information available from the different experimental techniques, such as X-ray and NMR. In particular, different structural information from the two techniques may help in characterizing protein mobility in solution. However, it is possible that the uncertainty in the determination of atom coordinates (also termed “structural noise”) from the primary X-ray data may affect the outcome of the mobility reconstruction. We have selected a calmodulin peptide complex, for which a 2 Å resolution X-ray structure as well as a large amount of accurate NMR data are available, as a relevant test case. We find that on average a single structure refined by simultaneous use of X-ray and NMR paramagnetic data scores similarly on three parameters able to detect domain motion as a previously published ensemble of structures.

Introduction

Biomolecules move, and their motions might be extensive in space and time and relevant for the accomplishment of their biological function. But, given the quality of the available experimental data, what is our ability to reconstruct such motions and their extent in solution – i.e. closer to physiological environment? The situation is even more complex when facing multi-domain proteins, for which both intra- and inter- domain motions may occur and need to be described. The two-domain protein Calmodulin (CaM) bound to a peptide is the system of choice to address these issues.[1–7] Extensive paramagnetically induced NMR data (residual dipolar couplings, RDCs, and pseudocontact shifts, PCSs) have been previously recorded for CaM bound to the IQ peptide.[8] PCSs and RDCs are the cleanest parameters to derive information concerning inter-domain motions in solution, and furthermore motional averaging reflects to different extents in these two types of restraints.[9,4,10–14]

These paramagnetic data can directly show the presence of inter-domain motions if intra-domain motions are negligible and the structures of the domains are accurately known. However, X-ray structures are affected by structural inaccuracy that needs to be considered: this inaccuracy is not limited to hydrogen atoms, which are often absent in electron density maps, but it also affects heavy atom coordinates, which in turn may further affect the positioning of hydrogen atoms. This inaccuracy, termed “structural noise”,[15] is larger the lower the resolution of the X-ray data, and imposes a limitation in the quality of the attainable fit for the NMR data. In this work, we aim at determining whether inter-domain motion is required to explain the available experimental NMR data after careful minimization of the structural noise.[16–19,3,20–22,6,23] To this end, we resort to the use of REFMAC[24,25], part of the CCP4 suite for crystallography[26], in which NMR data can be introduced at the stage of structural refinement (REFMAC-NMR).[27] We have already proven that this approach is able to reveal subtle differences between solution and crystal data even when the agreement seems already satisfactory[27,28] or, in some other cases, to largely reconcile the crystallographic model with the NMR data when they appeared sensibly different.[29]

Materials and Methods

Simultaneous refinement of X-ray and NMR data

The structural refinements of the different X-ray structures were performed by simultaneous use of X-ray structure factors and NMR data (PCSs and RDCs), using the REFMAC-NMR software, which is based on Refmac5[24,25] from the CCP4 crystallography suite[26]. The joint refinement is achieved by a minimization of a target function that balances the contribution of X-ray data, NMR data and ideal geometric constraints, as detailed in the Supporting Information.

Agreement of the experimental data

The agreement of the NMR data against the different structural models was evaluated by the calculation of the Q-factor. In particular, the Q-factor for individual PCS or RDC data (Q_{PCS}, Q_{RDC}) is defined as:[30]

$$Q = \sqrt{\frac{\sum_i^n (exp_i - calc_i)^2}{\sum_i^n exp_i^2}} \qquad (1)$$

where *exp* and *calc* indicate, respectively, experimental and back-calculated PCS or RDC values.

To balance the contributions of PCSs and RDCs, we have applied the composite Q-factor for PCSs and RDCs ($Q_{PCS+RDC}$) defined as:[8]

$$Q_{PCS+RDC} = \sqrt{\frac{N_{PCS}Q_{PCS}^2 + N_{RDC}Q_{RDC}^2}{N_{PCS} + N_{RDC}}} \qquad (2)$$

where N_{PCS} and N_{RDC} indicate, respectively, the number of experimental PCSs and RDCs.

The tensor parameters and the Q-factors for PCSs and RDCs against the original and the refined structures were obtained using the program FANTEN[31], available in the WeNMR portal[32]. In particular, for the N-terminal domain the calculation of the tensor was performed by fixing the metal position to the one optimized during REFMAC-NMR refinement and by fitting the tensor anisotropies and orientations by a Gauss-Newton optimization approach; for the C-terminal domain the fit of all tensor parameters as well as of the metal positions were obtained by using a non-linear least squares approach, minimizing the weighted differences between experimental and back-calculated data. As performed in [8], tensor calculations for each domain and each lanthanide were performed by minimizing the $Q_{PCS+RDC}$.

Error bars were calculated by using a Montecarlo bootstrap approach discarding 20% of the experimental data. 100 cycles were performed.

Tensor-based criteria

The tensors calculated for the individual N- and C- terminal domains were compared according to the three criteria proposed by Russo et al.:[8] tensor size ratio, tensor alignment and shape, and metal positioning. The tensor size ratio criterion reports on the deviation from 1 of the ratios between the anisotropies of the tensors calculated for the C-terminal and N-terminal domains; the tensor alignment and shape criterion reports on the discrepancies in alignment and rhombicity between the C-terminal and N-terminal domains tensors, and the metal positioning criterion reports on the differences in the coordinates of the paramagnetic metals as determined from C-terminal and N-terminal domains data *(see SI for more details*).

Results and Discussion

First of all, we observe that the agreement between the NMR data (RDCs and PCSs[8]) and each of the three structures of the three independent molecules in the asymmetric unit of the crystals of CaM when bound to the IQ peptide (PDB 2BE6) is already remarkable, the composite Q-factors being below 0.25 for all six lanthanides alternatively substituted to the calcium ion in the second binding site of the N-terminal domain of the protein. However, the back-calculation of the tensors for the N- and C-terminal domains derived from RDCs and PCSs shows some differences (*see SI*), which prompted a deeper investigation.[8] In the previous study,[8] it was found that, in solution, an exchange between the three conformations present in the asymmetric unit of the crystal explained the paramagnetic data for the individual domains better than any of the individual structures. By contrast to the study here, no attempts were undertaken to optimize the structures of the individual domains. A structural ensemble constituted by the three crystallographic structures compounded with few selected conformations extracted from molecular dynamics simulations was constructed.[8] The obtained ensemble shows a remarkable improvement in terms of Q-factors, but most importantly it almost completely reconciles the differences in the back-calculation of the tensors for the N- and C- terminal domains. In the present work, we investigate whether the NMR data and the X-ray data can also be reconciled, and to what extent, at the stage of the structural refinement.

Under the assumption of no mobility, we have refined each of the three structures against all the available PCS and RDC data. The three X-ray models show a much better agreement with the experimental data after refinement, with chains A and B providing the best results (*see SI*). We chose to pursue the investigation using chain A (Table 1) as it also scores slightly better in terms of the tensor size and tensor alignment criteria proposed by Russo et al.[8] The three panels in Figure 1 show the improvements of the refined structure (*Full-length refinement*) with respect to the X-ray structure (*X-ray structure*) in terms of these criteria (*see also SI for chains B and C*).

Of note, although the structural changes introduced by the simultaneous refinement against X-ray and NMR data are very small ($\mathrm{RMSD_{bb}}$ = 0.076 Å between the refined and the original X-ray structures), they result in a large improvement of the agreement with the NMR data, similar to that of the published ensemble of structures[8] (see Table 1 and Fig. S2). Such improvement is not due to trivial deviations of $\mathrm{H^N}$-N bonds with respect to the peptide plane, but rather reflects changes in heavy atom coordinates (*see SI*).

The agreement in terms of alignment between the tensors of N-terminal and C-terminal domains calculated for the refined structure is similarly good as that achieved for the ensemble solution (Fig. 1, *Ensemble approach*), whereas the agreement in the metal position is somewhat worse. The tensor size ratio in the refined structure is smaller than or equal to 1 for all metals, while the tensor size ratios for the ensemble scatter around 1 indicating that domain motion drives this parameter to values as would be expected for a faithful structural representation.

While an essentially random distribution of the differences between experimental and back-calculated RDCs (residuals) is obtained, the discrepancies of PCSs show a clear polarization (or bias), especially in the C-terminal domain (Figures

S3-S4). The worse reproduction of the metal position with respect to the ensemble approach could be thus a reporter of mobility, which is obviously not captured by the static simultaneous refinement approach. To check this, a rigid body minimization for the two subunits was performed using as starting point the refined structure. This procedure implies applying a rigid rototranslation to one of the two domains, to provide the best fit of the NMR data. The resulting structure is shown in Figure 2, and the transformation mainly consists of a rigid translation of the C-terminal with respect to the N-terminal domain by about 3 Å with a negligible rotation. This simple movement does produce an improvement in the residuals distribution (Figure S4), together with a slight improvement in the tensor alignment and a significant improvement in the metal position (Figure 1 - *Rigid body minimization*), which now results in a better agreement for the single structure than for the ensemble, taking all six metals into account. In the ensemble approach, only 4 metals were used for refinement and Yb and Ho used for cross validation. It is therefore not surprising that the metal position for those deviates the most in the ensemble.

Overall, the scoring parameters are differently well fulfilled for the single structure and the ensemble, but on average for all three criteria, the two representations score similarly. It can also be seen that tensor size and metal position are the most sensitive parameters and it might still be possible to improve the structural representation.

In conclusion, both a single refined structure and the previously selected ensemble of structures can fit the data similarly well and fulfill the tensor parameter criteria similarly well (Figure 1), one scoring a little better regarding the tensor sizes, the other scoring a little better regarding the metal position.

Given the high quality of the NMR data, the residual small discrepancy in the back-calculated tensor parameters (the tensor size ratio is still around 10% smaller than 1 for a couple of lanthanides, Figure 1) can still be significant. This residual discrepancy, which is similar for both cases, may originate from different reasons. While the spread of tensor sizes is similar for both approaches, the average ratio is closer to 1 in the ensemble approach than in the single structure approach. Thus, taking motion into account clearly moves this parameter closer to the value expected for an ideal description of the structure, i.e. 1. Yet, the spread is similar for both approaches.

The reasons could be different for the two approaches. Clearly in the ensemble approach, except for averaging the individual domains from the X-ray ensemble, no further attempts to reduce the structural noise were made. In the single structure approach, the structural noise is removed as much as possible, such that anisotropic motion of the domains might explain the scatter. However, it is possible that there are some other specific inconsistencies between the X-ray data and the NMR data, e.g. resulting from small local, intra-domain rearrangements that can be present between the crystal and the solution, as already observed for calmodulin-peptide complexes.[1,6,33] This explanation would be supported by the small polarization of several residuals of the PCSs of protons belonging to the first helix of the C-terminal domain, which is better observable when the rigid body minimization eliminates most of the other polarization effects (see Figure S4). Obviously, this hypothesis denies the initial assumption of rigidity of the two domains. We thus conclude that any further attempt to explain the small residual discrepancies would require a full re-determination of the structure, and this is beyond the scope of the present work.

In summary, we have shown that in this test case NMR data can be explained similarly well either as an average over an ensemble of crystallographic structures augmented with additional MD derived structures or by a single structure obtained by simultaneous refinement against NMR and X-ray data. This result underlines the importance of structural noise when using non-refined X-ray structures for NMR data interpretation. Structural noise may play a dual role: on the one hand it may conceal real structural differences between the crystal structure and the solution structure, but on the other hand it may introduce ambiguities in the analysis of solution data. Further refinement removing structural noise as much as possible and taking domain motion into account is therefore necessary.

Acknowledgements

This research was supported by Ente Cassa di Risparmio di Firenze, MIUR PRIN 2012SK7ASN, European Commission projects BioMedBridges No. 284209, pNMR No. 317127, and Instruct, part of the European Strategy Forum on Research Infrastructures (ESFRI). Specifically, we thank the EU ESFRI Instruct Core Centre CERM, Italy. ER holds a FIRC triennial fellowship "Gino Mazzega and Guglielmina Locatello" (17941).

List of figures and tables

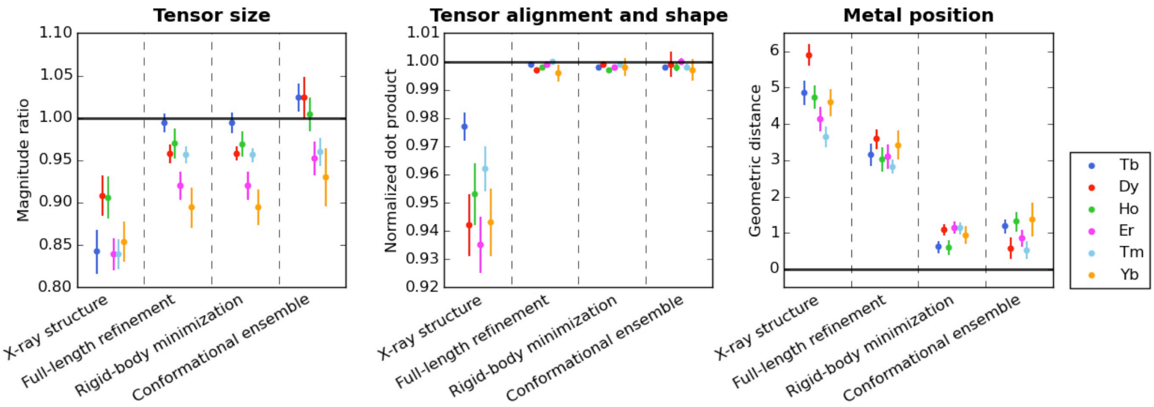


Figure 1: Comparison between the tensor calculated independently for the N- and C- terminal domains in terms of tensor sizes, tensor alignments and shapes, and metal positions for the original X-ray structure, for the structure refined by REFMAC-NMR (full-length refinement), for the refined structure after applying rigid body minimization, and for the ensemble-based approach[8].

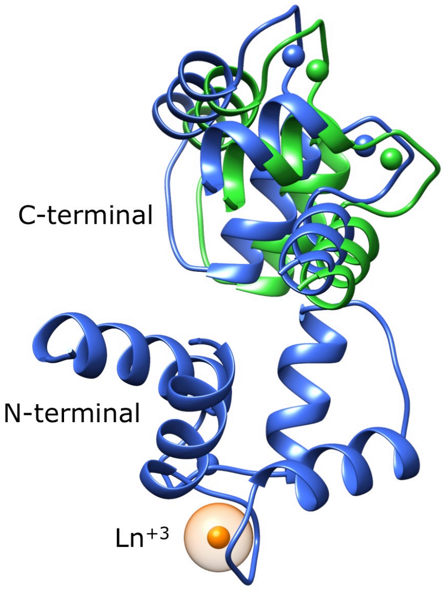
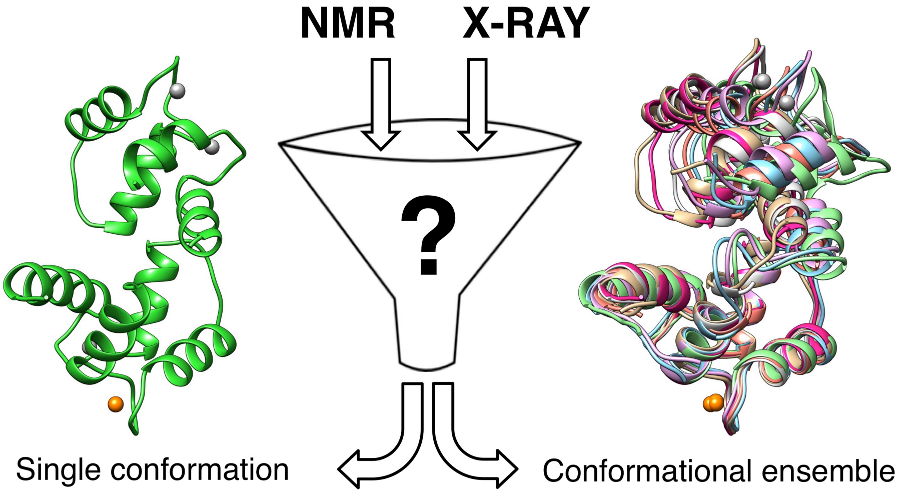


Figure 2: REFMAC-NMR structure before (blue) and after rigid body minimization (green).

	PCS N-ter	RDC N-ter	PCS+RDC N-ter	PCS C-ter	RDC C-ter	PCS+RDC C-ter
X-ray	0.176	0.251	0.200	0.181	0.350	0.232
Refined	0.090	0.124	0.100	0.154	0.156	0.156
RBM	0.077	0.122	0.090	0.136	0.130	0.136
Ensemble[8]	0.094	0.191	0.125	0.138	0.172	0.148

Table 1. Q-factors of PCSs and RDCs (calculated as average of the Q-factors calculated for the seven lanthanides) for the N- and C- terminal domains calculated for the original X-ray structure, for the refined structure, after further rigid body minimization, and for the ensemble-based approach.[8]

Graphical abstract



References

(1) Ikura, M.; Barbato, G.; Klee, C. B.; Bax, A. *Cell Calcium* **1992**, *13*, 391.

(2) Barbato, G.; Ikura, M.; Kay, L. E.; Pastor, R. W.; Bax, A. *Biochemistry (Mosc.)* **1992**, *31*, 5269.

(3) Chou, J. J.; Li, S.; Klee, C. B.; Bax, A. *Nat. Struct Mol Biol* **2001**, *8*, 990.

(4) Bertini, I.; Del Bianco, C.; Gelis, I.; Katsaros, N.; Luchinat, C.; Parigi, G.; Peana, M. ;.; Provenzani, A.; Zoroddu, M. A. *Proc.Natl.Acad.Sci.USA* **2004**, *101*, 6841.

(5) Bertini, I.; Gupta, Y. K.; Luchinat, C.; Parigi, G.; Peana, M.; Sgheri, L.; Yuan, J. *J. Am. Chem. Soc.* **2007**, *129*, 12786.

(6) Bertini, I.; Kursula, P.; Luchinat, C.; Parigi, G.; Vahokoski, J.; Wilmanns, M.; Yuan, J. *J. Am. Chem. Soc.* **2009**, *131* (14), 5134.

(7) Bertini, I.; Giachetti, A.; Luchinat, C.; Parigi, G.; Petoukhov, M. V.; Pierattelli, R.; Ravera, E.; Svergun, D. I. *J. Am. Chem. Soc.* **2010**, *132*, 13553.

(8) Russo, L.; Maestre-Martinez, M.; Wolff, S.; Becker, S.; Griesinger, C. *J. Am. Chem. Soc.* **2013**, *135* (45), 17111.

(9) Tolman, J. R.; Al-Hashimi, H. M.; Kay, L. E.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123* (7), 1416.

(10) Pintacuda, G.; John, M.; Su, X.-C.; Otting, G. *Acc. Chem. Res.* **2007**, *40* (3), 206.

(11) Koehler, J.; Meiler, J. *Prog. Nucl. Magn. Reson. Spectrosc.* **2011**, *59* (4), 360.

(12) Shishmarev, D.; Otting, G. *J. Biomol. NMR* **2013**, *56* (3), 203.

(13) Fragai, M.; Luchinat, C.; Parigi, G.; Ravera, E. *Coord. Chem. Rev.* **2013**.

(14) Andra\lojć, W.; Berlin, K.; Fushman, D.; Luchinat, C.; Parigi, G.; Ravera, E.; Sgheri, L. *J. Biomol. NMR* **2015**, 1.

(15) Zweckstetter, M.; Bax, A. *J. Biomol. NMR* **2002**, *23* (2), 127.

(16) Shaanan, B.; Gronenborn, A. M.; Cohen, G. H.; Gilliland, G. L.; Veerapandian, B.; Davies, D. R.; Clore, G. M. *Science* **1992**, *257* (5072), 961.

(17) Chou, J. J.; Li, S.; Bax, A. *J.Biomol.NMR* **2000**, *18* (3), 217.

(18) Skrynnikov, N. R.; Goto, N. K.; Yang, D.; Choy, W. Y.; Tolman, J. R.; Mueller, G. A.; Kay, L. E. *J. Mol. Biol.* **2000**, *295* (5), 1265.

(19) Tian, F.; Valafar, H.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123* (47), 11791.

(20) Ulmer, T. S.; Ramirez, B. E.; Delaglio, F.; Bax, A. *J. Am. Chem. Soc.* **2003**, *125* (30), 9179.

(21) Blackledge, M. *Prog. Nucl. Magn. Reson. Spectrosc.* **2005**, *46* (1), 23.

(22) Prestegard, J. H.; Mayer, K. L.; Valafar, H.; Benison, G. C. *Methods Enzymol.* **2005**, *394*, 175.

(23) Simon, B.; Madl, T.; Mackereth, C. D.; Nilges, M.; Sattler, M. *Angew. Chem. Int. Ed Engl.* **2010**, *49* (11), 1967.

(24) Murshudov, G. N.; Vagin, A. A.; Dodson, E. J. *Acta Crystallogr. D Biol. Crystallogr.* **1997**, *53* (3), 240.

(25) Murshudov, G. N.; Skubák, P.; Lebedev, A. A.; Pannu, N. S.; Steiner, R. A.; Nicholls, R. A.; Winn, M. D.; Long, F.; Vagin, A. A. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67* (4), 355.

(26) Winn, M. D.; Ballard, C. C.; Cowtan, K. D.; Dodson, E. J.; Emsley, P.; Evans, P. R.; Keegan, R. M.; Krissinel, E. B.; Leslie, A. G. W.; McCoy, A.; McNicholas, S. J.; Murshudov, G. N.; Pannu, N. S.; Potterton, E. A.; Powell, H. R.; Read, R. J.; Vagin, A.; Wilson, K. S. *Acta Crystallogr. D Biol. Crystallogr.* **2011**, *67* (4), 235.

(27) Rinaldelli, M.; Ravera, E.; Calderone, V.; Parigi, G.; Murshudov, G. N.; Luchinat, C. *Acta Crystallogr. D Biol. Crystallogr.* **2014**, *70* (4), 958.

(28) Carlon, A.; Ravera, E.; Andrałojć, W.; Parigi, G.; Murshudov, G. N.; Luchinat, C. *Prog. Nucl. Magn. Reson. Spectrosc.* **2016**, *92–93*, 54.

(29) Carlon, A.; Ravera, E.; Hennig, J.; Parigi, G.; Sattler, M.; Luchinat, C. *J. Am. Chem. Soc.* **2016**, *138* (5), 1601.

(30) Cornilescu, G.; Marquardt, J.; Ottiger, M.; Bax, A. *J. Am. Chem. Soc.* **1998**, *120* (27), 6836.

(31) Rinaldelli, M.; Carlon, A.; Ravera, E.; Parigi, G.; Luchinat, C. *J. Biomol. NMR* **2015**, *61* (1), 21.

(32) Wassenaar, T. A.; Dijk, M. van; Loureiro-Ferreira, N.; Schot, G. van der; Vries, S. J. de; Schmitz, C.; Zwan, J. van der; Boelens, R.; Giachetti, A.; Ferella, L.; Rosato, A.; Bertini, I.; Herrmann, T.; Jonker, H. R. A.; Bagaria, A.; Jaravine, V.; Güntert, P.; Schwalbe, H.; Vranken, W. F.; Doreleijers, J. F.; Vriend, G.; Vuister, G. W.; Franke, D.; Kikhney, A.; Svergun, D. I.; Fogh, R. H.; Ionides, J.; Laue, E. D.; Spronk, C.; Jurkša, S.; Verlato, M.; Badoer, S.; Pra, S. D.; Mazzucato, M.; Frizziero, E.; Bonvin, A. M. J. J. *J. Grid Comput.* **2012**, *10* (4), 743.

(33) Kursula, P. *Amino Acids* **2014**, *46* (10), 2295.

Supporting Information

Simultaneous refinement of X-ray and NMR data

The general approach used for the inclusion of the NMR data into the structural refinement consists of: i) a first minimization against the X-ray data alone, with an automatic setting of the "*weight matrix*" value (i.e. of the relative weights of geometry violations and X-ray violations), possibly followed by manual adjusting of the weight matrix to reduce the calculated RMSD of bond lengths, bond angles, and chiral volumes, if too large; and ii) a second minimization performed including the NMR data (i.e. PCSs, RDCs), in order to decrease their Q-factor. In particular, the NMR restraints contribution (t) to the total minimized function is:

$$t = k_{PCS} \sum_i w_i \left[max\left(\left| PCS_i^{calc} - PCS_i^{obs} \right| - T_i, 0 \right)^2 \right] + k_{RDC} \sum_i w_i \left[max\left(\left| RDC_i^{calc} - RDC_i^{obs} \right| - T_i, 0 \right)^2 \right]$$

where T_i is the tolerance on each PCS or RDC values, w_i is the weight, and k_{PCS}/k_{RDC} are the overall weighting factors for PCS and RDC. This term needs to be fine-balanced with the geometric term and with the X-ray data agreement term. For simplicity, the products of the k_{PCS} and w_i value is indicated as "PCS weight", and the products of the k_{RDC} and w_i value is indicated as "RDC weight" used in the calculations for the investigated systems. The second minimization, besides the optimization of the weight matrix value, requires the optimization of the weights of the NMR data and of the additional torsion angle restraints *pep1*, *pep2* and ω. Three further torsion angles have been introduced in the REFMAC library to restrain the planarity of the O_i-C_i-N_{i+1}-C^{α}_i, of the C_{i-1}-(N_i--H_i)-C^{α}_i (out of plane bending of HN), and of the C^{α}_i-C_i-N_{i+1}-C^{α}_{i+1} dihedral angles (*pep1*, *pep2* and ω, respectively). This is needed to avoid the worsening of geometric parameters from ideality by the inclusion of the NMR data in the refinement. Furthermore, overall weighting parameters over ideal geometries of all atoms involved (‘*weight refined_atoms*’) or not involved (‘*weight other_atoms*’) in the calculation of gradients and of the second derivatives corresponding to X-ray reflections has been also introduced. To guarantee comparable contributions of both paramagnetic restraints to the total minimized function, in all the fitting procedures RDCs were given a weight of 1.0, whereas PCSs were given weights calculated from the ratios between the maximum of values of RDCs and PCSs.

Calculation of the paramagnetic tensors

PCSs (in ppm) are described according to the following equation:[1,2]

$$\delta^{PCS} = \frac{1}{12\pi r^3}\left[\Delta\chi_{ax}(3\cos^2\vartheta - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\vartheta\cos 2\varphi\right]$$

where r, θ, and φ are the spherical coordinates defining the position of the nucleus in the frame where the magnetic susceptibility anisotropy tensor is diagonal, with origin on the metal ion, and $\Delta\chi_{ax}$ and $\Delta\chi_{rh}$ are the axial and rhombic magnetic susceptibility anisotropy parameters defined as:

$$\Delta\chi_{ax} = \chi_{zz} - \frac{\chi_{xx} + \chi_{yy}}{2} \qquad \Delta\chi_{rh} = \chi_{xx} - \chi_{yy}\,.$$

The x, y, and z axes of the principal frame of the tensor are defined using the convention $|\Delta\chi_{rh}| \leq 2|\Delta\chi_{ax}|/3$.

RDCs (in Hz) are described according to the following equation: [1,2]

$$\delta^{RDC} = k\left[\Delta\chi_{ax}(3\cos^2\alpha - 1) + \frac{3}{2}\Delta\chi_{rh}\sin^2\alpha\cos 2\beta\right]$$

with

$$k = -\frac{S_{LS}}{4\pi}\frac{B_0^2}{15k_BT}\frac{\gamma_A\gamma_B\hbar}{2\pi r_{AB}^3}$$

where r_{AB} is the distance between the two coupled nuclei A and B, and the α and β angles define the orientation of the vector connecting the coupled nuclei in the frame of the magnetic susceptibility anisotropy tensor (the same appearing in the equation of PCSs), B_0 is the magnetic field, T the absolute temperature, k_B the Boltzmann constant, γ the magnetogyric ratio, ħ the Plank constant divided by 2π, and S_{LS} is the model-free order parameter introduced to take into account the local mobility of the coupled nuclei. As performed in reference [8], tensor back-calculations for each domain and each lanthanide were performed by minimizing the $Q_{PCS+RDC}$ defined in Eq. 2 (see “Agreement of the experimental data” in the main text).

Tensor-based criteria

Comparison of the tensor sizes is performed by computing the ratio of their axial components according to the following equation:

$$ratio = \frac{\Delta\chi_{ax(Cter)}}{\Delta\chi_{ax(Nter)}}$$

where values close to 1 indicate very similar tensor size.

Comparison of the tensor alignment and shape is realized by normalized dot product between the five components of the Saupe matrices (S) for the N- and C-terminal domains defined as:

$$normDot = \frac{\overline{S_{Nter}} \cdot \overline{S_{Cter}}}{\|\overline{S_{Nter}}\| \cdot \|\overline{S_{Cter}}\|}$$

where values close to 1 still indicate that the tensors are very similar in terms of alignment and shape.

The distances between metal positions fixed and fitted, respectively, for the N-terminal and C-terminal domains are calculated for each lanthanide.

Refinement results using REFMAC-NMR

Parameters	**REFMAC**	**REFMAC-NMR**		
		Model A	**Model B**	**Model C**
R-factor	0.221	0.219	0.219	0.220
R-free	0.249	0.253	0.252	0.252
Rmsd Bond Length	0.014	0.016	0.016	0.017
Rmsd Bond Angle	1.817	1.898	1.918	1.917
Rmsd Chiral Volume	0.128	0.140	0.140	0.145
Q-factor PCS	Tab.S2	0.075	0.067	0.082
Q-factor RDC	Tab.S2	0.097	0.100	0.108

Table S1: Results obtained from the refinement of the original X-ray structures without NMR data (REFMAC) and after the inclusion of PCSs and RDCs in the calculation (REFMAC-NMR) for individual models A, B and C.

Agreement with the experimental data

X-ray structure	N-terminal			C-terminal		
	PCS	RDC	PCS+RDC	PCS	RDC	PCS+RDC
Model A	0.176	0.251	0.200	0.181	0.350	0.232
Model B	0.117	0.266	0.168	0.153	0.281	0.191
Model C	0.158	0.285	0.196	0.151	0.358	0.219

Table S2: Data agreement for the original X-ray structures. Q-factors of PCSs and RDCs (calculated as average of the Q-factors calculated for the six lanthanides) for the N- and C- terminal domains calculated for the original X-ray structure.

REFMAC-NMR structure	N-terminal			C-terminal		
	PCS	RDC	PCS+RDC	PCS	RDC	PCS+RDC
Model A	0.090	0.124	0.100	0.154	0.156	0.156
Model B	0.080	0.118	0.091	0.137	0.152	0.142
Model C	0.078	0.136	0.095	0.142	0.148	0.146

Table S3: Data agreement for the REFMAC-NMR refined structures. Q-factors of PCSs and RDCs (calculated as average of the Q-factors calculated for the six lanthanides) for the N- and C- terminal domains calculated for the full-length refinement structure.

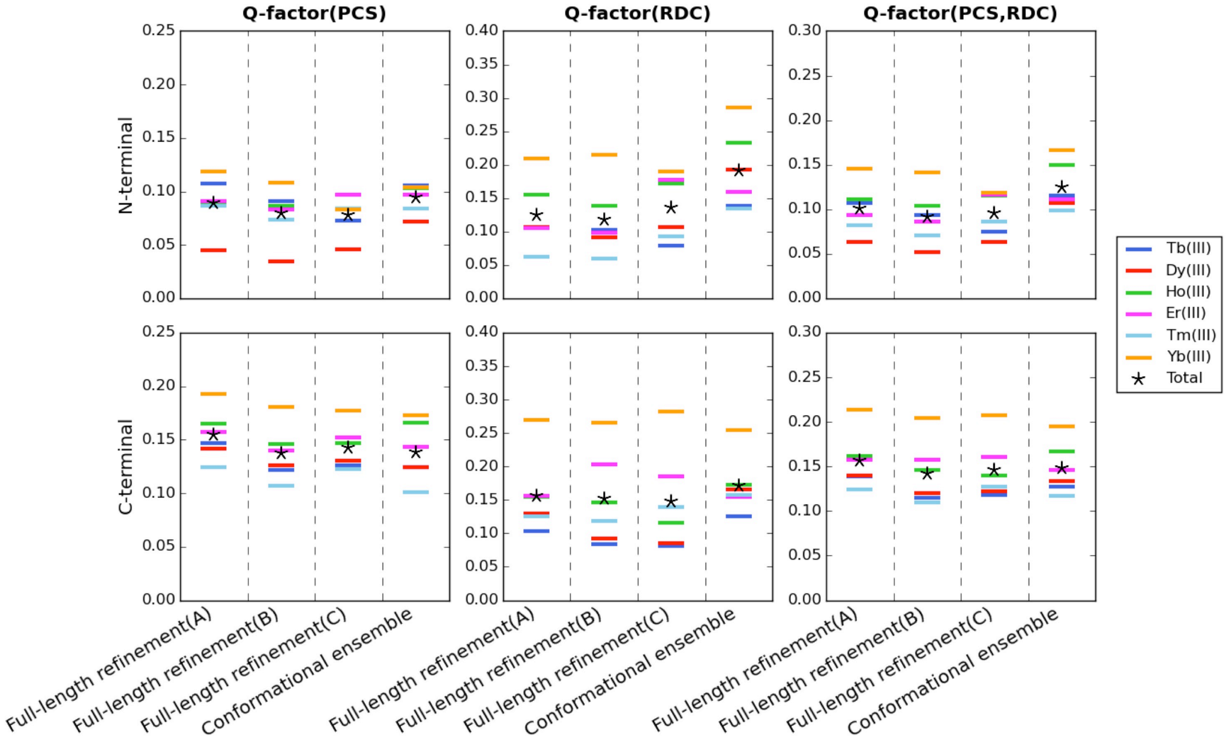


Figure S1: Data agreement for models A, B and C. Q-factors of PCSs and RDCs for the N- and C- terminal domains calculated for the structures refined by REFMAC-NMR and for the ensemble-based approach[8].

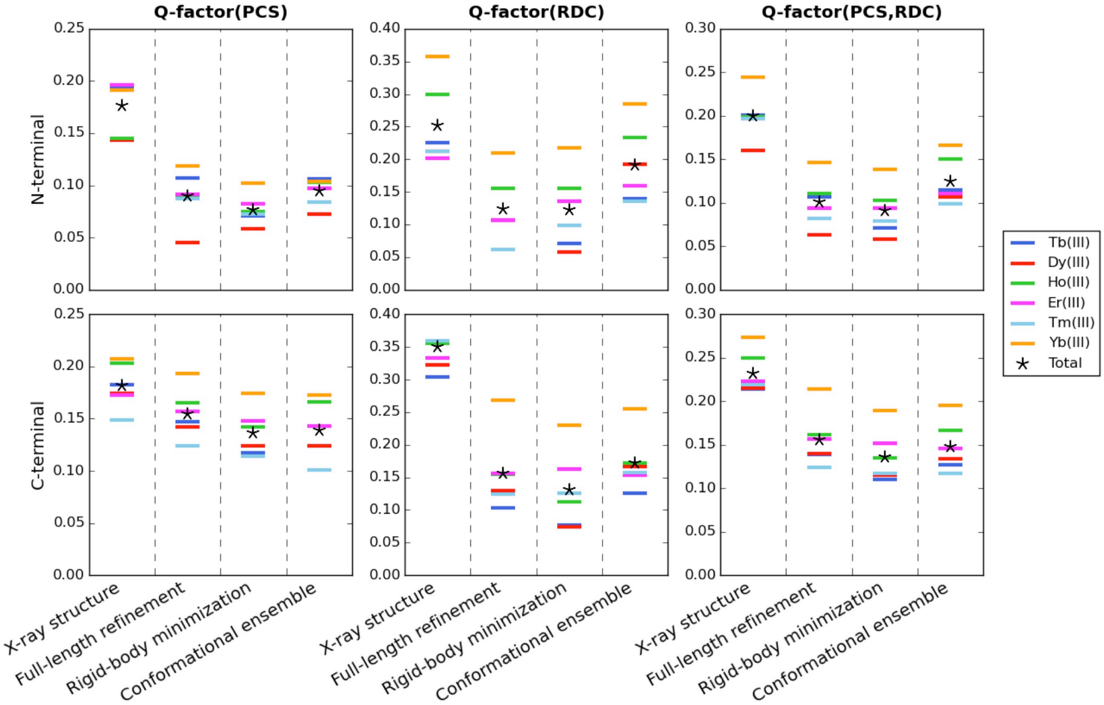


Figure S2: Data agreement for model A. Q-factors of PCSs and RDCs for the N- and C- terminal domains calculated from model A for the original X-ray structure, for the structures refined by REFMAC-NMR, for the refined structure after applying rigid body minimization, and for the ensemble-based approach[8].

Evaluation of the tensor-based criteria

X-ray structure	**Tensor size**		
Ln	Model A	Model B	Model C
Tb	0.842±0.026	0.915±0.034	1.074±0.026
Dy	0.908±0.024	0.923±0.024	0.977±0.027
Ho	0.906±0.025	0.950±0.021	1.010±0.026
Er	0.839±0.019	0.807±0.020	0.878±0.025
Tm	0.839±0.018	0.826±0.017	0.921±0.017
Yb	0.854±0.024	0.844±0.029	0.875±0.022

Table S4: Ratio between the axial anisotropies of the C-terminal and N-terminal domains, as calculated from the tensors determined individually for the two domains of each of the three models A, B and C in the original X-ray structures.

X-ray structure	**Tensor alignment and shape**		
Ln	Model A	Model B	Model C
Tb	0.977±0.005	0.980±0.005	0.892±0.014
Dy	0.942±0.011	0.985±0.007	0.943±0.015
Ho	0.953±0.011	0.990±0.003	0.885±0.017
Er	0.935±0.010	0.981±0.007	0.910±0.021
Tm	0.962±0.008	0.987±0.006	0.930±0.017
Yb	0.943±0.012	0.997±0.004	0.915±0.012

Table S5: Normalized scalar product between the independent components of the tensors calculated individually for the N- and C-terminal domains of each of the three models A, B and C in the original X-ray structures.

X-ray structure	**Metal position**		
Ln	Model A	Model B	Model C
Tb	4.858±0.326	2.432±0.173	4.113±0.321
Dy	5.903±0.302	2.555±0.250	4.140±0.403
Ho	4.730±0.323	2.800±0.191	4.303±0.361
Er	4.130±0.345	2.604±0.349	4.813±0.360
Tm	3.651±0.282	2.288±0.182	5.351±0.298
Yb	4.591±0.378	3.076±0.450	4.383±0.375

Table S6: Distance (in Å) between the metal coordinates calculated individually for the N- and C-terminal domains of each of the three models A, B and C in the original X-ray structures.

REFMAC-NMR structure	**Tensor size**		
Ln	Model A	Model B	Model C
Tb	0.994±0.011	1.000±0.011	0.989±0.014
Dy	0.958±0.011	0.972±0.011	0.989±0.013
Ho	0.979±0.018	0.987±0.017	0.987±0.017
Er	0.920±0.017	0.919±0.021	0.898±0.020
Tm	0.956±0.010	0.943±0.012	0.918±0.012
Yb	0.894±0.024	0.914±0.022	0.928±0.022

Table S7: Ratio between the axial anisotropies of the C-terminal and N-terminal domains, as calculated from the tensors determined individually for the two domains of each of the three models A, B and C in the REFMAC-NMR refined structures.

REFMAC-NMR structure	**Tensor alignment and shape**		
Ln	Model A	Model B	Model C
Tb	0.999±0.000	0.998±0.001	0.995±0.001
Dy	0.997±0.001	0.999±0.001	0.999±0.001
Ho	0.998±0.001	0.997±0.001	0.994±0.003
Er	0.999±0.001	0.999±0.002	0.986±0.006
Tm	1.000±0.000	0.999±0.000	0.989±0.003
Yb	0.996±0.003	0.995±0.003	0.985±0.005

Table S8: Normalized scalar product between the independent components of the tensors calculated individually for the N- and C-terminal domains of each of the three models A, B and C in the REFMAC-NMR refined structures.

REFMAC-NMR structure	**Metal position**		
Ln	Model A	Model B	Model C
Tb	3.152±0.300	1.508±0.189	2.082±0.220
Dy	3.579±0.281	1.514±0.149	1.581±0.228
Ho	3.016±0.337	1.632±0.192	1.797±0.252
Er	3.109±0.339	1.948±0.352	2.491±0.237
Tm	2.821±0.195	1.467±0.153	2.624±0.217
Yb	3.421±0.393	2.189±0.400	1.948±0.264

Table S9: Distance (in Å) between the metal coordinates calculated individually for the N- and C-terminal domains of each of the three models A, B and C in the REFMAC-NMR refined structures.

Structure validation using MolProbity

Protons were removed and added back to the refined structures by using Molprobity[9,10], to verify that such improvement is not due to trivial in-plane or out-of-plane deviations of H^N-N bonds with respect to the peptide plane. Re-protonated structure shows only minor changes in the agreement with the NMR data, suggesting that the backbone atoms refinement, although modest, is very significant for the correct placement of the protons.

	H^N added using REFMAC-NMR		H^N added using MolProbity	
Ln	Q_{PCS}	Q_{RDC}	Q_{PCS}	Q_{RDC}
Tb	0.090	0.071	0.092	0.088
Dy	0.053	0.077	0.055	0.094
Ho	0.097	0.133	0.098	0.145
Er	0.088	0.134	0.089	0.134
Tm	0.081	0.097	0.082	0.112
Yb	0.106	0.225	0.107	0.228
Total	**0.075**	**0.097**	**0.077**	**0.110**

Table S10: Q-factor for the structure of model A as refined by REFMAC-NMR and for the same structure where the protons were removed and added back using Molprobity[9,10].

Analysis of the residuals

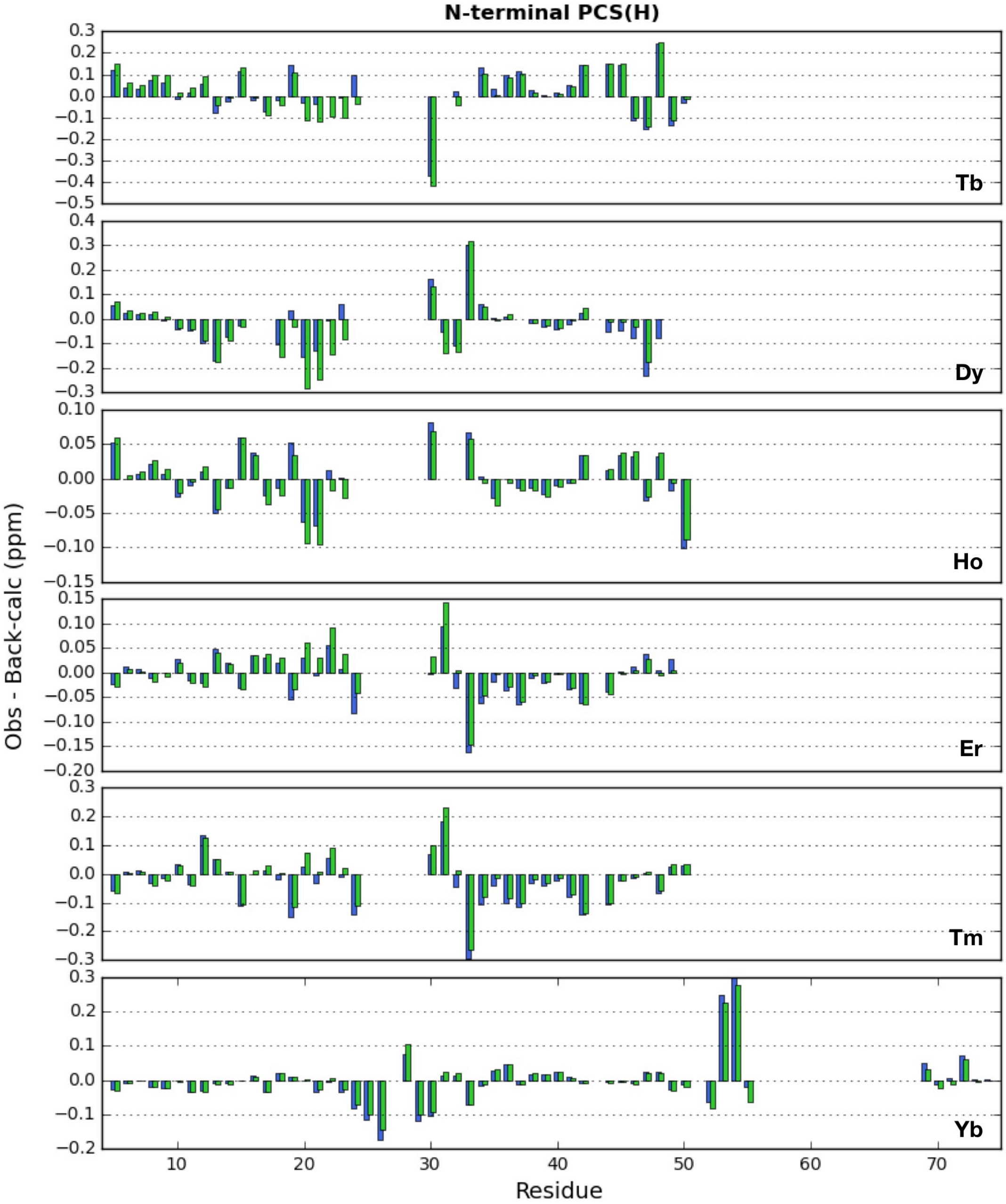


Figure S3: Discrepancies between experimental and back-calculated PCS data of the N-terminal domain for the REFMAC-NMR refined structure before (blue bar) and after (green bar) the rigid body minimization.

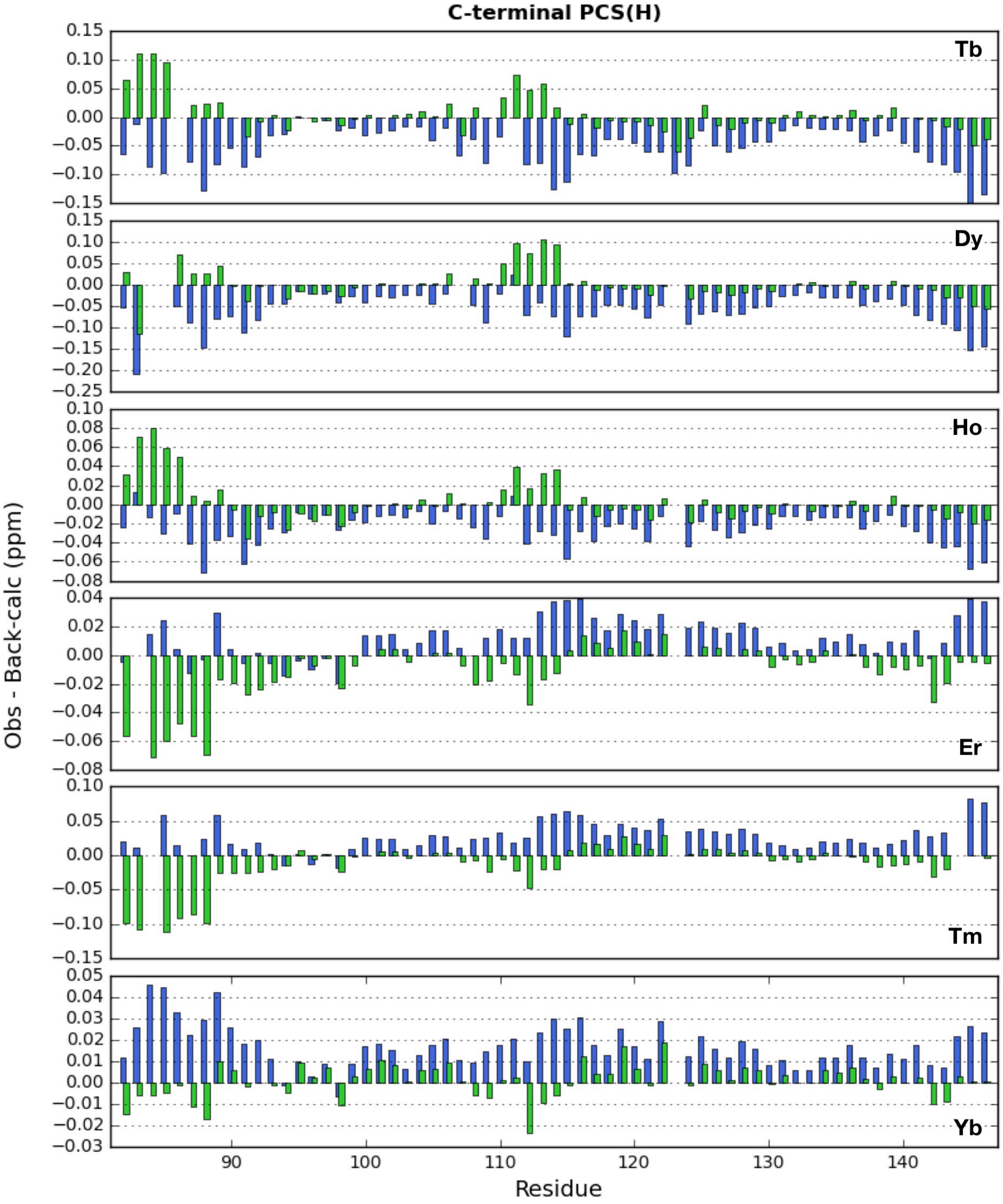


Figure S4: Discrepancies between experimental and back-calculated PCS data of the C-terminal domain for the REFMAC-NMR refined structure before (blue bar) and after (green bar) the rigid body minimization.

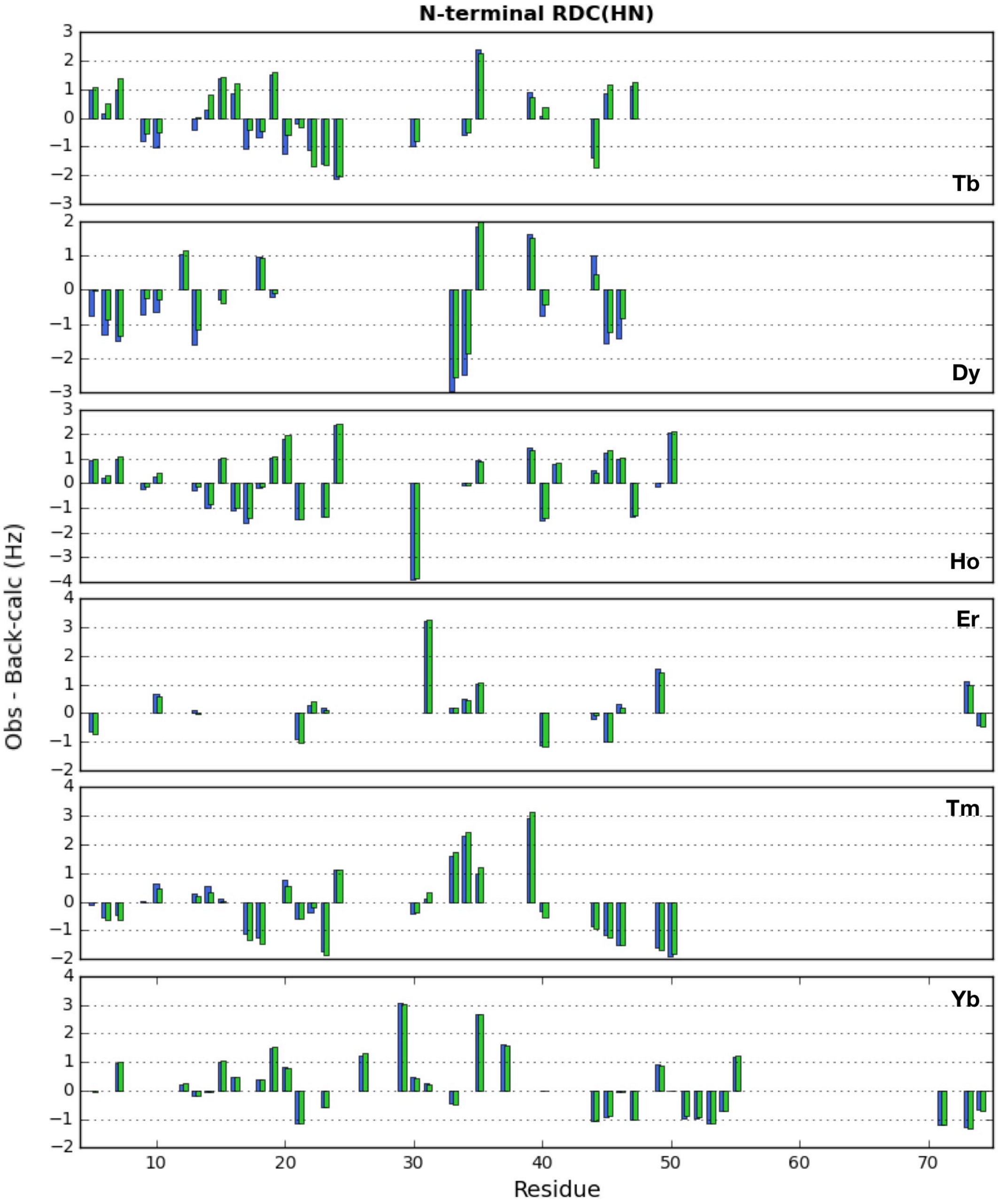


Figure S5: Discrepancies between experimental and back-calculated RDC data of the N-terminal domain for the REFMAC-NMR refined structure before (blue bar) and after (green bar) the rigid body minimization.

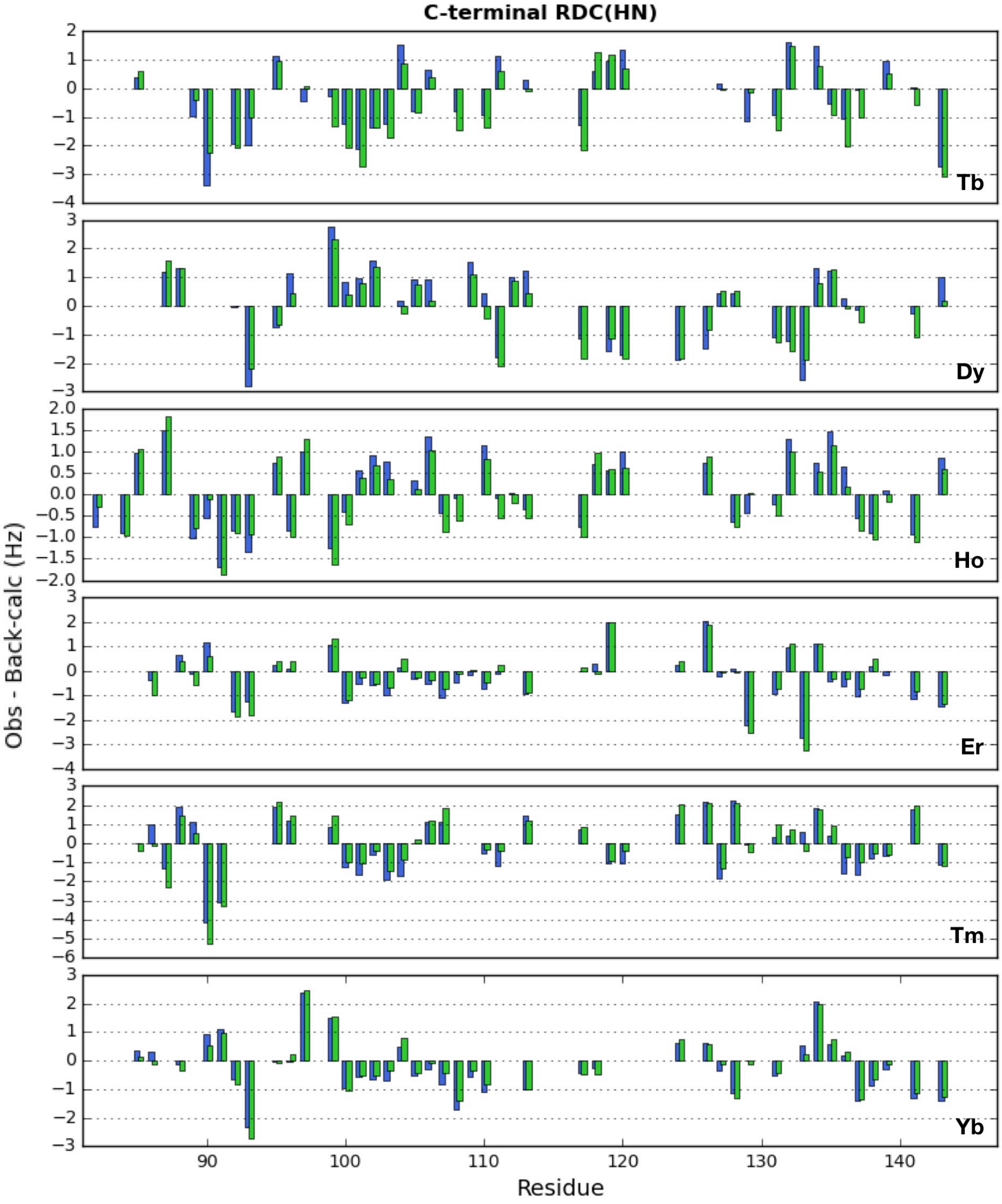


Figure S6: Discrepancies between experimental and back-calculated RDC data of the C-terminal domain for the REFMAC-NMR refined structure before (blue bar) and after (green bar) the rigid body minimization.

References

(1) Bertini, I.; Luchinat, C.; Parigi, G. *Prog. Nucl. Magn. Reson. Spectrosc.* **2002**, *40*, 249.

(2) Rinaldelli, M.; Carlon, A.; Ravera, E.; Parigi, G.; Luchinat, C. *J. Biomol. NMR* **2015**, *61* (1), 21.

(3) Tolman, J. R.; Al-Hashimi, H. M.; Kay, L. E.; Prestegard, J. H. *J. Am. Chem. Soc.* **2001**, *123* (7), 1416.

(4) Bertini, I.; Del Bianco, C.; Gelis, I.; Katsaros, N.; Luchinat, C.; Parigi, G.; Peana, M. ;.; Provenzani, A.; Zoroddu, M. A. *Proc.Natl.Acad.Sci.USA* **2004**, *101*, 6841.

(5) Longinetti, M.; Luchinat, C.; Parigi, G.; Sgheri, L. *Inverse Probl.* **2006**, *22* (4), 1485.

(6) Rodríguez-Castañeda, F.; Maestre-Martínez, M.; Coudevylle, N.; Dimova, K.; Junge, H.; Lipstein, N.; Lee, D.; Becker, S.; Brose, N.; Jahn, O.; Carlomagno, T.; Griesinger, C. *EMBO J.* **2010**, *29* (3), 680.

(7) Cerofolini, L.; Fields, G. B.; Fragai, M.; Geraldes, C. F. G. C.; Luchinat, C.; Parigi, G.; Ravera, E.; Svergun, D. I.; Teixeira, J. M. C. *J. Biol. Chem.* **2013**, *288* (42), 30659.

(8) Russo, L.; Maestre-Martinez, M.; Wolff, S.; Becker, S.; Griesinger, C. *J. Am. Chem. Soc.* **2013**, *135* (45), 17111.

(9) Davis, I. W.; Leaver-Fay, A.; Chen, V. B.; Block, J. N.; Kapral, G. J.; Wang, X.; Murray, L. W.; Arendall, W. B. I. I. I.; Snoeyink, J.; Richardson, J. S.; Richardson, D. C. *Nucleic Acids Res* **2007**, *35* (Web Server issue), W375.

(10) Chen, V. B.; Arendall, W. B.; Headd, J. J.; Keedy, D. A.; Immormino, R. M.; Kapral, G. J.; Murray, L. W.; Richardson, J. S.; Richardson, D. C. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (1), 12.

4. Conclusions

The efficient integration of many different physical techniques is required in order to maximize the reliability of structural biology studies. The combination of the data provided from these heterogeneous techniques is not as straightforward as it sometimes appears and the quality of the scientific results is often limited by the level of expertise of the scientist dealing with them. Hence, it is on the large research infrastructures to encode their specific high-level know-how into user-friendly computational tools for the combination of different informative sources that need to be made available to the wider scientific community. As members of the NMR community, we focused on giving our contribution to extend the applicability of paramagnetic-assisted solution NMR in the context of structural biology, exploring all the possibilities that all these data offer. During the last decades, pseudo-contact shifts (PCSs) and residual dipolar couplings (RDCs) arising from the paramagnetic centers present in proteins gained a relevant role in the characterization of biological systems and, nowadays, they are routinely measured in several laboratories. The increasing popularity of these restraints mostly resides on their wide applicability such as determination, refinement and validation of protein structures in solution. Besides this, PCSs and RDCs stand out among NMR-based restrains, for their intrinsic property of giving long-range structural information, which results of particular interest in the study of multi-domain proteins and protein-protein complexes, detecting the relative arrangement and/or the mobility effects between the different structural units. The measurements and application of paramagnetic restraints has been widely extended by the use of specific tags able to coordinate the paramagnetic ion that can be purposely attached to proteins. Another popular approach is the use of external media for inducing the alignment of the molecule, avoiding the use of paramagnetic ions for such measurements.

As a first goal, we developed a user-friendly wed-based application for the analysis of PCS and RDC data. FANTEN, standing for “Finding ANisotropy TENsor”, allows for the calculation of the anisotropy tensors through a joint analysis of these restraints against an available structural model, aiding the user step-by-step in the calculation setup and providing many additional features for the analysis and visualization of the fitting of the experimental data. Starting from the input data that can be provided in the most common format (i.e. CYANA and Xplor-NIH), the user is allowed to choose among three kinds of interfaces having different purposes and different level of complexity. The “PCS-RDC fitting (SMART)” interface summarizes the most commonly required types of calculations, i.e. tensor estimation using PCSs and/or RDCs by fixing the metal position, allowing multiple fittings from different dataset or structural models in one single click. Data loading and calculation setup is almost automatic, allowing for some modification from the user if necessary. All the results are reported in very clear and comprehensive way in the interface output and are all available for download. Inspection of the fitting of experimental data is fulfilled by using interactive correlation plots thanks to the use of JQuery widgets. Iso-PCS surfaces can be visualized through the use an integrated JSmol application, or downloaded in Gaussian cube format and used as input for other molecular visualization programs (i.e. UCSF Chimera and PyMol). Finally, the uncertainty related to the tensor parameters can be computed through a Monte-Carlo bootstrap approach where the number of cycles and the percentage of discarded data are defined by the user. The “PCS-RDC fitting (CUSTOM)” interface is instead required to tackle more complex types of calculations, allowing the user for fixing or fitting the position of the metal, the anisotropies and the orientation of the tensors. Also in this case, the user is helped by possibility of customizing the calculation in many different ways, depending on the prior knowledge. For instance, the position of the metal can be just slightly optimized, estimated starting from the information about neighboring atoms or calculated using a grid-search algorithm. Finally, “Rigid Body Minimization” allows for a quick determination of relative position/orientation of two structural units. The calculation exploits the singular value decomposition approach to derive the direct transformation that needs to be applied to the second unit to obtain the best agreement of the data and provide the rigidly rototranslated unit/s ready for download.

The second goal required the integration of the discussed NMR-based restraints with the most popular technique used for structural studies of biomolecules: X-ray crystallography. We focused on the reliability of the comparison between NMR- and X-ray- derived structural information and investigated the sources of discrepancies that usually affect these two techniques when used in an independent way. We used REFMAC-NMR to efficiently combine NMR data with X-ray reflections at the refinement stage permitting to take into account not only the experimental uncertainties related to NMR measurements, but also the structural uncertainty or “noise” in the crystallographic model, underestimated in most of the cases. This issue assumes relevant importance if NMR data are combined with X-ray models from medium to low resolution, where the uncertainties of the coordinates of the backbone atoms might have a relevant role in the determination of hydrogen atoms (almost absent in the electron diffraction pattern), on which most of the NMR information is related. As a first test case, we performed the refinement of medium/low resolution structure of a protein-protein-RNA complex proving that the addition of RDCs in the X-ray structural refinement is enough to largely reconcile the crystallographic model with the NMR data also when they appeared sensibly different. Moreover, the approach permitted to obtain more accurate conclusion on the preservation of the inter-domain arrangement passing from solution to the solid state but also the identification of relevant outliers in the fitting of the experimental data in specific location of the complex. In the second test case, we performed the refinement the two domains protein calmodulin, for which the initial disagreement of PCSs and RDCs data with the three conformations present tin the asymmetric unit of the crystal was interpreted in terms of inter-domain mobility in a previously published work. Each model was refined using REFMAC-NMR showing significant improvement in the fitting of NMR data. The comparison between the tensors calculated independently from the two domains in terms of tensor size, alignment and metal positions permitted to identify the most compliant conformation that was obtained from a rigid body optimization of the refined domains. This study shows how a single static structural representation optimized by the joint X-ray/NMR refinement is sufficient to explain the experimental data in the same way as an ensemble of conformations do, and makes the point on the reliability of the interpretation of the discrepancies between NMR and X-ray. The analysis of the these two cases prompt us to the development of a systematic approach for the characterization of multisubunit systems that can be used as a decision tree and that each scientist can follow and interrogate for choosing the right approach for investigating a biological system. The entire approach is based on the faithful representation of the individual structural units. This task can be achieved by the joint refinement using REFMAC-NMR, if the individual domains preserve their structure in the two states (solution and crystal) or by other refinement methods (CYANA, Xplor-NIH) if the intra-domain differences are substantial. The use of REFMAC-NMR for the refinement of multisubunit systems was improved by the implementation of a new algorithm for constrained optimization that permitted the inclusion of a-priori information for the tensor calculations. The considered constraints are related to the anisotropies and the orientations that the tensors must assume during the calculation, with specific focus on the treatment of symmetric systems. In REFMAC-NMR was also introduced the possibility of structural refinement using RDCs collected from methyl- sidechains, extending its applicability to high molecular weight systems: this lays the grounds for identifying discrepancies in sidechain arrangements, which are the determinant for an improved modeling of protein-protein and protein-ligand interactions.

Bibliography

1. Gierasch, L. M. & Gershenson, A. Post-reductionist protein science, or putting Humpty Dumpty back together again. *Nature Chemical Biology* **5,** 774–777 (2009).
2. Byeon, I.-J. L. *et al.* Structural Convergence between Cryo-EM and NMR Reveals Intersubunit Interactions Critical for HIV-1 Capsid Function. *Cell* **139,** 780–790 (2009).
3. Zhao, G. *et al.* Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. *Nature* **497,** 643–646 (2013).
4. Demers, J.-P. *et al.* High-resolution structure of the Shigella type-III secretion needle by solid-state NMR and cryo-electron microscopy. *Nature Communications* **5,** 4976 (2014).
5. Simon, B., Madl, T., Mackereth, C. D., Nilges, M. & Sattler, M. An efficient protocol for NMR-spectroscopy-based structure determination of protein complexes in solution. *Angew. Chem. Int. Ed. Engl.* **49,** 1967–1970 (2010).
6. Russel, D. *et al.* Putting the Pieces Together: Integrative Modeling Platform Software for Structure Determination of Macromolecular Assemblies. *PLoS Biology* **10,** e1001244 (2012).
7. Yang, Z. *et al.* UCSF Chimera, MODELLER, and IMP: An integrated modeling system. *Journal of Structural Biology* **179,** 269–278 (2012).
8. Ward, A. B., Sali, A. & Wilson, I. A. Integrative Structural Biology. *Science* **339,** 913–915 (2013).
9. Alber, F. *et al.* The molecular architecture of the nuclear pore complex. *Nature* **450,** 695–701 (2007).
10. Alber, F. *et al.* Determining the architectures of macromolecular assemblies. *Nature* **450,** 683–694 (2007).
11. Baù, D. *et al.* The three-dimensional folding of the α-globin gene domain reveals formation of chromatin globules. *Nature Structural & Molecular Biology* **18,** 107–114 (2011).

12. Lasker, K. *et al.* Molecular architecture of the 26S proteasome holocomplex determined by an integrative approach. *Proceedings of the National Academy of Sciences* **109,** 1380–1387 (2012).
13. Kurland, R. J. & McGarvey, B. R. Isotropic NMR shifts in transition metal complexes: calculation of the Fermi contact and pseudocontact terms. *J.Magn.Reson.* **2,** 286–301 (1970).
14. Lohman, J. A. B. & Maclean, C. Magnetic susceptibility anisotropies from quadrupolar magnetic field effects in high field 2H NMR. *Mol.Phys.* **38,** 1255–1261 (1979).
15. Bertini, I., Luchinat, C. & Parigi, G. Magnetic susceptibility in paramagnetic NMR. *Progress in Nuclear Magnetic Resonance Spectroscopy* **40,** 249–273 (2002).
16. Bertini, I., Luchinat, C. & Parigi, G. Paramagnetic constraints: an aid for quick solution structure determination of paramagnetic metalloproteins. *Concepts in Magnetic Resonance* **14,** 259–286 (2002).
17. Bertini, I., Luchinat, C., Parigi, G. & Pierattelli, R. NMR Spectroscopy of Paramagnetic Metalloproteins. *ChemBioChem* **6,** 1536–1549 (2005).
18. Allegrozzi, M. *et al.* Lanthanide induced pseudocontact shifts for solution structure refinements of macromolecules in shells up to 40 A from the metal ion. *Journal of the American Chemical Society* **122,** 4154–4161 (2000).
19. Bertini, I., Janik, M. B. L., Lee, Y.-M., Luchinat, C. & Rosato, A. Magnetic Susceptibility Tensor Anisotropies for a Lanthanide Ion Series in a Fixed Protein Matrix. *Journal of the American Chemical Society* **123,** 4181–4188 (2001).
20. Bertini, I., Janik, M. B. L., Liu, G., Luchinat, C. & Rosato, A. Solution structure calculations through self-orientation in a magnetic field of cerium (III) substituted calcium-binding protein. *J.Magn.Reson.* **148,** 23–30 (2001).
21. Bertini, I. *et al.* Paramagnetism-based Versus Classical Constraints: An Analysis of the Solution Structure of Ca Ln Calbindin D9k. *J.Biomol.NMR* **21,** 85–98 (2001).
22. Bertini, I., Gelis, I., Katsaros, N., Luchinat, C. & Provenzani, A. Tuning the Affinity for Lanthanides of Calcium Binding Proteins. *Biochemistry* **42,** 8011–8021 (2003).
23. Bertini, I., Fragai, M., Lee, Y.-M., Luchinat, C. & Terni, B. Paramagnetic metal ions in ligand screening: the CoII matrix metalloproteinase 12. *Angew.Chem.Int.Ed.* **43,** 2254–2256 (2004).
24. Wöhnert, J., Franz, K. J., Nitz, M., Imperiali, B. & Schwalbe, H. Protein Alignment by a Coexpressed Lanthanide-Binding Tag for the Measurement of Residual Dipolar Couplings. *J. Am. Chem. Soc.* **125,** 13338–13339 (2003).

25. Haberz, P. *et al.* Two new chiral EDTA-based metal chelates for weak alignment of proteins in solution. *Org.Lett.* **8,** 1275–1278 (2006).
26. John, M. & Otting, G. Strategies for Measurements of Pseudocontact Shifts in Protein NMR Spectroscopy. *ChemPhysChem* **8,** 2309–2313 (2007).
27. Pintacuda, G., John, M., Su, X.-C. & Otting, G. NMR Structure Determination of Protein−Ligand Complexes by Lanthanide Labeling. *Accounts of Chemical Research* **40,** 206–212 (2007).
28. Su, X.-C. *et al.* A Dipicolinic Acid Tag for Rigid Lanthanide Tagging of Proteins and Paramagnetic NMR Spectroscopy. *J. Am. Chem. Soc.* **130,** 10486–10487 (2008).
29. Keizers, P. H. J., Saragliadis, A., Hiruma, Y., Overhand, M. & Ubbink, M. Design, Synthesis, and Evaluation of a Lanthanide Chelating Protein Probe: CLaNP-5 Yields Predictable Paramagnetic Effects Independent of Environment. *Journal of the American Chemical Society* **130,** 14802–14812 (2008).
30. Häussinger, D., Huang, J. & Grzesiek, S. DOTA-M8: An extremely Rigid, High-Affinity Lanthanide Chelating Tag for PCS NMR Spectroscopy. *Journal of the American Chemical Society* **131,** 14761–14767 (2009).
31. Man, B. *et al.* 3-Mercapto-2,6-Pyridinedicarboxylic Acid: A Small Lanthanide-Binding Tag for Protein Studies by NMR Spectroscopy. *Chem.Eur.J.* **16,** 3827–3832 (2010).
32. Hass, M. A. S., Keizers, P. H. J., Blok, A., Hiruma, Y. & Ubbink, M. Validation of a Lanthanide Tag for the Analysis of Protein Dynamics by Paramagnetic NMR Spectroscopy. *Journal of the American Chemical Society* **132,** 9952–9953 (2010).
33. Saio, T. *et al.* An NMR strategy for fragment-based ligand screening utilizing a paramagnetic lanthanide probe. *Journal of Biomolecular NMR* **51,** 395–408 (2011).
34. Swarbrick, J. D., Ung, P., Chhabra, S. & Graham, B. An Iminodiacetic Acid Based Lanthanide Binding Tag for Paramagnetic Exchange NMR Spectroscopy. *Angewandte Chemie International Edition* **50,** 4403–4406 (2011).
35. Bertini, I. *et al.* The catalytic domain of MMP-1 studied through tagged lanthanides. *FEBS Lett.* **586,** 557–567 (2012).
36. Liu, W.-M. *et al.* A pH-Sensitive, Colorful, Lanthanide-Chelating Paramagnetic NMR Probe. *J. Am. Chem. Soc.* **134,** 17306–17313 (2012).

37. Yagi, H., Maleckis, A. & Otting, G. A systematic study of labelling an α-helix in a protein with a lanthanide using IDA-SH or NTA-SH tags. *Journal of Biomolecular NMR* **55,** 157–166 (2013).
38. Loh, C. T. *et al.* Lanthanide Tags for Site-Specific Ligation to an Unnatural Amino Acid and Generation of Pseudocontact Shifts in Proteins. *Bioconjugate Chemistry* **24,** 260–268 (2013).
39. Sanders, C. R. & Landis, G. C. Reconstitution of membrane proteins into lipid-rich bilayered mixed micelles for NMR studies. *Biochemistry* **34,** 4030–4040 (1995).
40. Hansen, M. R., Mueller, L. & Pardi, A. Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* **5,** 1065–1074 (1998).
41. Hus, J. C., Marion, D. & Blackledge, M. De novo determination of protein structure by NMR using orientational and long-range order restraints. *J Mol Biol* **298,** 927–936 (2000).
42. Bax, A. & Tjandra, N. Are proteins even floppier than we thought? *Nature Struct.Biol.* **4,** 254–256 (1997).
43. Tjandra, N., Omichinski, J. G., Gronenborn, A. M., Clore, G. M. & Bax, A. Use of dipolar 1H–15N and 1H–13C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat Struct Mol Biol* **4,** 732–738 (1997).
44. Tjandra, N. & Bax, A. Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium. *Science* **278,** 1111–1114 (1997).
45. Losonczi, J. A. & Prestegard, J. H. Improved dilute bicelle solutions for high-resolution NMR of biological macromolecules. *J.Biomol.NMR* **12,** 447–451 (1998).
46. Al-Hashimi, H. M. *et al.* Variation of molecular alignment as a means of resolving orientational ambiguities in protein structures from dipolar couplings. *J.Magn.Reson.* **143,** 402–406 (2000).
47. Prestegard, J. H., al-Hashimi, H. M. & Tolman, J. R. NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q. Rev. Biophys.* **33,** 371–424 (2000).
48. Zweckstetter, M. & Bax, A. Characterization of molecular alignment in aqueous suspensions of Pf1 bacteriophage. *J. Biomol. NMR* **20,** 365–377 (2001).
49. Lakomek, N.-A. *et al.* Self-consistent residual dipolar coupling based model-free analysis for the robust determination of nanosecond to microsecond protein dynamics. *J Biomol NMR* **41,** 139–155 (2008).

50. Maltsev, A. S., Grishaev, A., Roche, J., Zasloff, M. & Bax, A. Improved Cross Validation of a Static Ubiquitin Structure Derived from High Precision Residual Dipolar Couplings Measured in a Drug-Based Liquid Crystalline Phase. *J. Am. Chem. Soc.* (2014). doi:10.1021/ja4132642
51. Koehler, J. & Meiler, J. Expanding the utility of NMR restraints with paramagnetic compounds: Background and practical aspects. *Progress in Nuclear Magnetic Resonance Spectroscopy* **59,** 360–389 (2011).
52. Schwieters, C. D., Kuszewski, J. J., Tjandra, N. & Marius Clore, G. The Xplor-NIH NMR molecular structure determination package. *Journal of Magnetic Resonance* **160,** 65–73 (2003).
53. Banci, L. *et al.* Paramagnetism-based restraints for Xplor-NIH. *J.Biomol.NMR* **28,** 249–261 (2004).
54. Schwieters, C. D., Kuszewski, J. J. & Marius Clore, G. Using Xplor-NIH for NMR molecular structure determination. *Progress in Nuclear Magnetic Resonance Spectroscopy* **48,** 47–62 (2006).
55. Banci, L. *et al.* The use of pseudocontact shifts to refine solution structures of paramagnetic metalloproteins: Met80Ala cyano-cytochrome c as an example. *Journal of Biological Inorganic Chemistry* **1,** 117–126 (1996).
56. Balayssac, S., Bertini, I., Luchinat, C., Parigi, G. & Piccioli, M. 13C direct detected NMR increases the detectability of residual dipolar couplings. *Journal of the American Chemical Society* **128,** 15042–15043 (2006).
57. Schmitz, C. & Bonvin, A. M. J. J. Protein-protein HADDocking using exclusively pseudocontact shifts. *J Biomol NMR* **50,** 263–266 (2011).
58. Schmitz, C., Vernon, R., Otting, G., Baker, D. & Huber, T. Protein Structure Determination from Pseudocontact Shifts Using ROSETTA. *Journal of Molecular Biology* **416,** 668–677 (2012).
59. Berlin, K., O'Leary, D. P. & Fushman, D. Fast approximations of the rotational diffusion tensor and their application to structural assembly of molecular complexes. *Proteins: Structure, Function, and Bioinformatics* **79,** 2268–2281 (2011).
60. Tolman, J. R., Al-Hashimi, H. M., Kay, L. E. & Prestegard, J. H. Structural and dynamic analysis of residual dipolar coupling data for proteins. *J. Am. Chem. Soc.* **123,** 1416–1424 (2001).

61. Bertini, I. *et al.* Experimentally exploring the conformational space sampled by domain reorientation in calmodulin. *Proc.Natl.Acad.Sci.USA* **101,** 6841–6846 (2004).

62. Fragai, M., Luchinat, C., Parigi, G. & Ravera, E. Conformational freedom of metalloproteins revealed by paramagnetism-assisted NMR. *Coordination Chemistry Reviews* (2013). doi:10.1016/j.ccr.2013.02.009

63. Schmitz, C., Stanton-Cook, M. J., Su, X.-C., Otting, G. & Huber, T. Numbat: an interactive software tool for fitting $\Delta\chi$-tensors to molecular coordinates using pseudocontact shifts. *J Biomol NMR* **41,** 179–189 (2008).

64. Zweckstetter, M. NMR: prediction of molecular alignment from structure using the PALES software. *Nature Protocols* **3,** 679–690 (2008).

65. Berlin, K., O'Leary, D. P. & Fushman, D. Improvement and analysis of computational methods for prediction of residual dipolar couplings. *Journal of Magnetic Resonance* **201,** 25–33 (2009).

66. Valafar, H. & Prestegard, J. H. REDCAT: a residual dipolar coupling analysis tool. *J. Magn. Reson.* **167,** 228–241 (2004).

67. Wassenaar, T. A. *et al.* WeNMR: Structural Biology on the Grid. *J Grid Computing* **10,** 743–767 (2012).

68. Zweckstetter, M. & Bax, A. Evaluation of uncertainty in alignment tensors obtained from dipolar couplings. *J. Biomol. NMR* **23,** 127–137 (2002).

69. Arseniev, A. *et al.* Three-dimensional structure of rabbit liver [Cd7]-metallothionein-2a in aqueous solution determined by nuclear magnetic resonance. *J.Mol.Biol.* **201,** 637–657 (1988).

70. Shaanan, B. *et al.* Combining experimental information from crystal and solution studies: joint X-ray and NMR refinement. *Science* **257,** 961–964 (1992).

71. Schiffer, C. A., Huber, R., Wüthrich, K. & van Gunsteren, W. F. Simultaneous Refinement of the Structure of BPTI Against NMR Data Measured in Solution and X-ray Diffraction Data Measured in Single Crystals. *Journal of Molecular Biology* **241,** 588–599 (1994).

72. Miller, M. *et al.* The oligomerization domain of p53: crystal structure of the trigonal form. *FEBS Lett.* **399,** 166–170 (1996).

73. Raves, M. L. *et al.* Joint refinement as a tool for thorough comparison between NMR and X-ray data and structures of HU protein. *J. Biomol. NMR* **21,** 235–248 (2001).

74. Tang, M. *et al.* High-resolution membrane protein structure by joint calculations with solid-state NMR and X-ray experimental data. *Journal of Biomolecular NMR* **51,** 227–233 (2011).
75. Lange, O. F. *et al.* Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution. *Science* **320,** 1471–1475 (2008).
76. Clore, G. M. & Schwieters, C. D. How Much Backbone Motion in Ubiquitin Is Required To Account for Dipolar Coupling Data Measured in Multiple Alignment Media as Assessed by Independent Cross-Validation? *Journal of the American Chemical Society* **126,** 2923–2938 (2004).
77. Lindorff-Larsen, K., Best, R. B., DePristo, M. A., Dobson, C. M. & Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **433,** 128–132 (2005).
78. Rinaldelli, M. *et al.* Simultaneous use of solution NMR and X-ray data in REFMAC5 for joint refinement/detection of structural differences. *Acta Cryst D, Acta Cryst Sect D, Acta Crystallogr D, Acta Crystallogr Sect D, Acta Crystallogr D Biol Crystallogr, Acta Crystallogr Sect D Biol Crystallogr* **70,** 958–967 (2014).
79. Winn, M. D. *et al.* Overview of the *CCP* 4 suite and current developments. *Acta Crystallographica Section D Biological Crystallography* **67,** 235–242 (2011).
80. Clore, G. M. 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.USA* **97,** 9021–9025 (2000).
81. Dosset, P., Hus, J. C., Marion, D. & Blackledge, M. A novel interactive tool for rigid-body modeling of multi-domain macromolecules using residual dipolar couplings. *J.Biomol.NMR* **20,** 223–231 (2001).
82. Pintacuda, G., Park, A. Y., Keniry, M. A., Dixon, N. E. & Otting, G. Lanthanide Labeling Offers Fast NMR Approach to 3D Structure Determinations of Protein−Protein Complexes. *Journal of the American Chemical Society* **128,** 3696–3702 (2006).
83. Hulsker, R., Baranova, M. V., Bullerjahn, G. S. & Ubbink, M. Dynamics in the transient complex of plastocyanin-cytochrome f from Prochlorothrix hollandica. *Journal of the American Chemical Society* **130,** 1985–1991 (2008).

84. Shishmarev, D. & Otting, G. How reliable are pseudocontact shifts induced in proteins and ligands by mobile paramagnetic metal tags? A modelling study. *J. Biomol. NMR* **56,** 203–216 (2013).

85. Ravera, E. *et al.* Insights into Domain–Domain Motions in Proteins and RNA from Solution NMR. *Accounts of Chemical Research* 140822145226006 (2014). doi:10.1021/ar5002318

86. Bertini, I. *et al.* Conformational space of flexible biological macromolecules from average data. *Journal of the American Chemical Society* **132,** 13553–13558 (2010).

87. Berlin, K. *et al.* Recovering a Representative Conformational Ensemble from Underdetermined Macromolecular Structural Data. *Journal of the American Chemical Society* **135,** 16595–16609 (2013).

88. Blackledge, M. Recent progress in the study of biomolecular structure and dynamics in solution from residual dipolar couplings. *Progress in Nuclear Magnetic Resonance Spectroscopy* **46,** 23–61 (2005).

89. Gelbart, M. E. & Kuroda, M. I. Drosophila dosage compensation: a complex voyage to the X chromosome. *Development* **136,** 1399–1410 (2009).

90. Graindorge, A., Militti, C. & Gebauer, F. Posttranscriptional control of X-chromosome dosage compensation. *Wiley Interdisciplinary Reviews: RNA* **2,** 534–545 (2011).

91. Conrad, T. & Akhtar, A. Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. *Nature Reviews Genetics* **13,** 123–134 (2012).

92. Stevens, F. C. Calmodulin: an introduction. *Can. J. Biochem. Cell Biol.* **61,** 906–910 (1983).

93. Russo, L., Maestre-Martinez, M., Wolff, S., Becker, S. & Griesinger, C. Interdomain Dynamics Explored by Paramagnetic NMR. *J. Am. Chem. Soc.* **135,** 17111–17120 (2013).

94. Cornilescu, G., Marquardt, J., Ottiger, M. & Bax, A. Validation of Protein Structure from Anisotropic Carbonyl Chemical Shifts in a Dilute Liquid Crystalline Phase. *Journal of the American Chemical Society* **120,** 6836–6837 (1998).

95. Fufezan, C. & Specht, M. p3d–Python module for structural bioinformatics. *BMC.Bioinformatics.* **10,** 258 (2009).