Chapter 16

NMR-Based Modeling and Refinement of Protein 3D Structures

Wim F. Vranken, Geerten W. Vuister, and Alexandre M.J.J. Bonvin

Abstract

NMR is a well-established method to characterize the structure and dynamics of biomolecules in solution. High-quality structures can now be produced thanks to both experimental advances and computational developments that incorporate new NMR parameters and improved protocols and force fields in the structure calculation and refinement process. In this chapter, we give a short overview of the various types of NMR data that can provide structural information, and then focus on the structure calculation methodology itself. We discuss and illustrate with tutorial examples "classical" structure calculation, refinement, and structure validation approaches.

Key words NMR, Structure calculation, Structure refinement, Structure validation

1 Introduction

The first step of a structure determination by NMR spectroscopy consists of the acquisition of NMR data, typically using heteronuclear multidimensional experiments, that allow the assignment of all atoms/spins of a molecule (^{1}H, ^{15}N, ^{13}C) to their chemical shift values (Fig. 1). Once this chemical shift assignment step is completed, ^{13}C- and ^{15}N-edited 3D NOESY spectra are generally used to obtain inter-atomic distances from nuclear Overhauser effects (NOE). These NOESY spectra provide the most detailed structural information that can be obtained from NMR and are still the most common core data used to define the 3D structure of the protein [1, 2]. In addition to distance information, other parameters, such as J-couplings [3], residual dipolar couplings (RDCs) [4], paramagnetic relaxation enhancements (PRE) and pseudo-contact shifts [5] can be measured, providing additional information to define the protein structure. Recent developments have enabled the calculation of the structures of relatively small
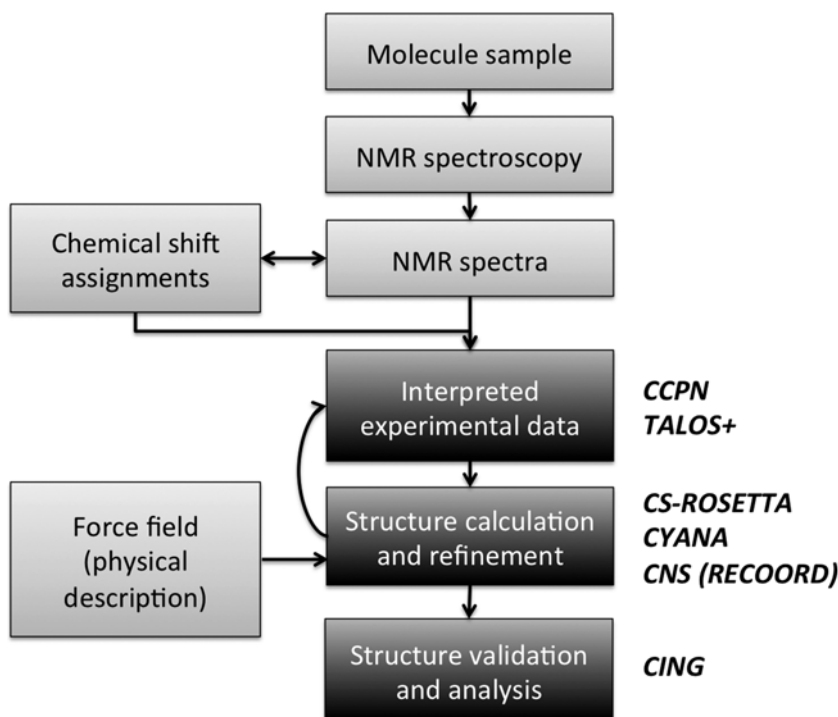
Fig. 1 Schematic overview of the structure calculation process in NMR. This chapter deals with the boxes in *black*, with indication of the software covered

proteins (less than ~12 kDa) from chemical shift values alone [6–8], a procedure that will be briefly outlined in this paper.

The experimental NMR parameters are converted to in silico restraints and 3D structures are generated from restrained molecular dynamics simulations following usually some form of molecular dynamics simulated annealing scheme (MD/SA) [9]. Multiple structures are calculated in this way, starting from the same experimental data but different random starting conditions. Provided that enough data of sufficient quality are available, the structures will converge onto the same overall fold. These structures are nowadays often further refined in explicit solvent (water), which has been shown to significantly improve their quality [10, 11]. Finally, the structures that best satisfy as many experimental restraints as possible, together with proper general chemical properties of proteins (such as bond lengths and angles), are then selected to form an ensemble of structures that represents the definitive solution of the structure calculation process.

In this chapter we will discuss the "classical" NMR structure calculation, refinement, and validation methods, with some reference to new chemical shift-based approaches. These will be illustrated with tutorial examples making use of the programs CYANA

[12] and CNS [13, 14] using the RECOORD [11] approach for water refinement [10], and the CS-ROSETTA protocol [7] based on chemical shift data. These are followed by a description of structure validation with the program CING [15].

2 Theory

This section gives a very brief overview of the NMR data relevant in the determination of 3D protein structures.

2.1 NMR Structural Information Sources

Several NMR parameters providing structural information can be measured for use in structure calculations and refinement; these are briefly described here.

2.1.1 Chemical Shifts

The first step in analyzing NMR experiments is to assign the observed resonances, each of which has a particular chemical shift value, to atoms in the molecule being studied. Although chemical shifts are very sensitive probes of the chemical environment of a spin, their dependency on the 3D structure is complex and their usage in 3D structure calculation is still evolving. A common use is to determine per-amino acid residue secondary structure preferences using the deviations of actual chemical shifts values from those of random coil peptides. This approach can be used to restrict the local conformation of a residue to a given region of the Ramachandran plot, either through torsion angle restraints [16] or by special database potential functions [17]. The last years have also seen the development of methods that can calculate protein structures from chemical shifts only [6–8]. In those, the chemical shifts are typically used to select peptide fragments that are then assembled to produce a 3D model. One of these, CS-ROSETTA [7], will be briefly illustrated in this chapter.

Many software programs are now available that perform the reverse operation, i.e. predict chemical shifts from in silico 3D structures (ShiftX [18], ShiftX2 [19], SHIFTS [20], SHIFTCALC [21], PROSHIFT [22] and SPARTA+ [23]). They can be used for example for structure validation.

2.1.2 NOEs

Classical protein structure determination by NMR relies on obtaining a dense network of distance restraints derived from nuclear Overhauser effects (NOEs) between nearby hydrogen atoms in a protein [1, 2]. Together, these restraints provide the essential information for defining the tertiary structure of a protein.

The NOE originates from cross-relaxation between dipolar coupled spins as a result of through-space spin–spin interactions that result in the transfer of magnetization from one spin to another. The NOE approximately scales with the distance r between the two spins as $1/r^6$. Because of this $1/r^6$ dependency, NOEs are

only detected between protons less than 5–6 Å away in space and are thus strongly biased towards the shorter distances. In addition, they are sensitive to dynamics averaging and suffer from so-called spin diffusion effects. Nevertheless, the large number of potential NOE-derived restraints renders them very valuable when treated appropriately.

2.1.3 J-Couplings

Scalar or J-couplings are mediated through chemical bonds connecting two spins. Particularly informative are the vicinal ^{3}J scalar coupling constants between atoms separated by three covalent bonds, which are correlated to the enclosed torsion angle, Θ, by the empirical Karplus equation [24]. In particular, ^{3}J(HN–Hα) and ^{3}J(Hα–Hβ) give information about the backbone φ-angle and the side-chain χ_1 angle in an amino acid, respectively. For amino acid with diastereotopic protons (Hβ2/Hβ3), the use of ^{3}J(Hα–Hβ) coupling does require their stereospecific assignments.

In contrast to the NOEs, scalar coupling constants only provide information on the local conformation of a polypeptide chain. J-couplings are commonly converted into dihedral angle restraints [1] or directly used as J-coupling restraints [25, 26] in NMR structure calculations. Nowadays, the less accurate dihedral angle restraints predicted from chemical shift values are in more common use.

2.1.4 Hydrogen Bonds

Some hydrogens, such as those present in the backbone amide groups, can chemically exchange with hydrogens from the water (H_2O) solvent. Experiments where the H_2O solvent is replaced by the proton–NMR-inactive D_2O can determine which of these hydrogens only exchange slowly; such hydrogens are assumed to be protected from the solvent and/or involved in a hydrogen bond [27]. Identification of the acceptor atom requires either additional experimentation or implicit assumptions. The latter is based on NOEs and/or the regularity of secondary structures and the inferred hydrogen bond restraints should be used with caution. Hydrogen bonds can also be detected directly from cross-hydrogen bond scalar coupling measured from constant time HNCO spectra [28, 29], thus providing accurate restraints for structure calculations [30]. Hydrogen bond restraints are usually introduced into the structure calculation protocol as distance restraints, typically by confining the donor-hydrogen/acceptor distance to a given range.

2.1.5 Residual Dipolar Couplings

Residual dipolar couplings (RDCs) are now a well-established source of structural information [31, 32]. They can be measured in solution by weakly aligning the molecule using a variety of methods [33]. RDCs provide orientational information of the internuclear vector of the two atoms for which the RDC is measured relative to three globally defined axes in the molecule, i.e. those of the alignment tensor. Note that if a RDC is measured between two

atoms that are not at a fixed distance from each other, there is also a distance dependence and hence usually only RDCs measured for inter-nuclear vectors with a fixed distance are used in the structure calculations. Residual dipolar couplings can be added as orientational restraints to the target function of the structure calculation algorithm [34].

2.1.6 Diffusion Anisotropy

For non-isotropical tumbling molecules, NMR relaxation data contain orientational information comparable to RDCs as result of the diffusion anisotropy [35]. NMR relaxation is characterized by relaxation times T_1 and T_2, and the T_1/T_2 ratio can be used to define diffusion anisotropy restraints in NMR structure calculations [36]. Again the orientation information comes from the angles of inter-nuclear vectors in an external frame, which, in the case of diffusion anisotropy data, corresponds to the orientational diffusion tensor frame. In practice, ^{15}N T_1 and T_{2-} relaxation data are most often used.

2.1.7 Paramagnetic Relaxation Effects

If a paramagnetic metal ion is present in a protein, or if it is introduced via for example a chelating agent chemically bound to the protein, the NMR signals of the nuclei in a shell around it will be affected [37] by several effects including contact and pseudo-contact shifts, relaxation rate enhancements, and cross-correlation effects. Analogously to RDC and diffusion anisotropy, these, depending on their type, can provide both distance and orientation information which can be converted into restraints to be used in various structure calculation softwares [38, 39].

2.2 Structure Calculation Software

The experimental information sources discussed above can be converted into restraints that can be used in the structure calculation process. Several computer programs have provisions for using the experimental NMR restraints; the most commonly used ones are CNS [13, 14], Xplor-NIH [40] and CYANA [12, 41], although many others are available, e.g. SCULPTOR [42], the SANDER module of AMBER [43], GROMACS [44] and YASARA [45].

Structure calculations in essence transform the experimental data (as restraints) into in silico atomic coordinate information. The calculations are usually based on some molecular dynamic simulated annealing protocol performed in torsion angle and/or Cartesian space, followed by a final refinement phase in explicit solvent (water). A general feature of all these protocols is the usage of a "target function": lower values of this function for a calculated structure indicates better agreement with the experimental data and with known molecular information. This molecular information is defined by a force field that contains physical energy terms for interactions such as van der Waals interactions and electrostatics, as well as terms describing the molecular geometry such as bond lengths, bond angles, etc. During the initial stages of a structure calculation

often the description of some of these terms is simplified to increase the computational speed and/or simplify the energy landscape to be search. For example, long-range nonbonded interactions are reduced to only repulsions between atoms and electrostatic interactions are at first neglected. A full nonbonded representation, including van der Waals (Lennard-Jones) and electrostatic (Coulomb) interactions, is then typically reintroduced for final refinement in explicit solvent [10].

2.3 Structure Selection and Structural Quality

Typically a large pool of structures is generated during the structure calculation process, from which a final ensemble of "best" structures is then selected. This choice of an ensemble of structures, rather than one single one, reflects the uncertainty in the experimental NMR data: often structures that agree with the experimental data equally well but differ locally (such as in loop regions) can be obtained. The most widely used structure selection procedure is based on the agreement with the experimental data (rather arbitrarily defined as a small number of restraint violations) and a low (overall) energy of the structures. Typically ensembles containing the 20 lowest energy models are selected, although this number is arbitrary. Ideally, the selected ensemble should represent the available conformational space accessible to the structure while simultaneously satisfying the experimental restraints. From this ensemble, a representative structure is usually defined; no real consensus exists, however, on how it should be selected. The wwPDB NMR validation taskforce recommends to select the structure that differs the least from all other structures within the ensemble, i.e. the mediod (the structure with the lowest atom coordinate RMSD from all other structures)[46].

The final ensemble is subsequently subjected to structure validation procedures in order to verify its quality. It is useful to distinguish the well-defined from the ill-defined regions of the ensemble during this process: these can be defined with, for example, CYRANGE [47], FindCore [48] or circular variance methods [49]. In practice, the quality indicators that are most commonly used to assess especially the well-defined regions of an NMR ensemble are [50]:

- the goodness of fit to the experimental data, by analyzing restraint violations;
- the precision of the ensemble, measured by positional root mean square deviation (RMSD);
- several physical and stereochemical quality indicators that assess the local and overall quality of protein structures, many of them based on knowledge from high-resolution X-ray structures.

Table 1 lists the most commonly used validation programs; some of which will be described later.

Table 1
Internet resources of NMR-related programs and databases mentioned in this chapter

Software	Internet address	Purpose
CNS	http://cns.csb.yale.edu/v1.3	Multilevel hierachical approach for the most commonly used algorithms in macromolecular structure determination (NMR, crystallography)
RECOORD	http://www.ebi.ac.uk/pdbe-apps/nmr/recoord/	Database of recalculated NMR structures with the CNS scripts used in the tutorial example
PDBe	http://www.pdbe.org/	An information portal to biological macromolecules structures
BMRB	http://www.bmrb.wisc.edu	Biological magnetic resonance data bank
CCPN	http://www.ccpn.ac.uk	A collaborative computing project for NMR
CING	http://nmr.cmbi.ru.nl/icing/	A server to validate the quality of NMR structures
PSVS	http://psvs-1_3.nesg.org/	A server to validate the quality of NMR structures
TALOS+	http://spin.niddk.nih.gov/bax/nmrserver/talos/	Protein backbone angle restraints from searching a database for chemical shift and sequence homology
CYANA	http://www.cyana.org	Structure calculation program (paid license required)
Below accessible via WeNMR framework		
CS-ROSETTA	http://haddock.science.uu.nl/enmr/services/CS-ROSETTA3	Structure calculation program using only chemical shift information
TALOS+	http://haddock.science.uu.nl/enmr/services/TALOS	Protein backbone angle restraints from searching a database for chemical shift and sequence homology
CYANA	http://www.enmr.eu/webportal/cyana.html	Structure calculation program (paid license required)
FormatConverter	http://haddock.chem.uu.nl/enmr/format-converter.html	Conversion between file formats (this web version limited compared to local install)

3 Methods

In this tutorial section we describe the procedures to generate various types of NMR restraints and how they can then be used in structure calculations using CYANA or CNS with the RECOORD scripts. The CS-ROSETTA chemical shift-only approach is also discussed, followed by a description of structure validation using the CING webserver. A classical, NOE distance-based structure

determination is recommended when possible since it typically leads to more accurate structures. The protocol to follow is however more complex than that of chemical shift-based methods like CS-ROSETTA. The latter is very simple to use and has been shown to generate accurate models for small systems [7, 51]. The commands to be executed are in bold Courier font with gray shading. Information about the various programs and web pages used in this tutorial can be found in Table 1. As an example project for these tutorials we provide a target from the CASD-NMR experiment [52, 53], OR36 (pdb 2LCI). This protein is 128 amino acids long. Besides backbone chemical shift data, NOE data are also provided that can be used in the other tutorials.

3.1 NMR Data and File Formats

A common problem in NMR is the abundance of file formats to store the NMR data (such as spectral data, chemical shifts, and peaks lists) and resulting restraint files for structure calculation. This makes data exchange between different programs a tedious process. The CCPN Data Model for macromolecular NMR [54] is intended to cover all data needed for macromolecular NMR spectroscopy from the initial experimental data to the final validation. The ccpNmr FormatConverter application allows consistent conversion between a large variety of data formats *via* import to and export from CCPN, and we recommend its local installation (*see* **Note 1**) and use for this purpose. The general steps to follow for importing data files into the FormatConverter from scratch are:

1. Start the FormatConverter on the command line:

```
formatConverter
```

2. Under the *Project* menu option, click **New** and enter a name for your project; you will not be able to use the FormatConverter until you have done this.

3. Under the *Import* menu option, go to *Single files*, select the type of data you want to import, then the name of the software the data comes from. A window will pop up, click on the **Select file** button, navigate to the file, and click **Select**. Click **IMPORT** to start the importing process.

4. Depending on the type of data, you might get popups that ask you to validate information or provide additional input. Press on the **?** button in these popups for additional information.

5. After successful import, repeat the process from **step 3**. Due to the way the CCPN framework stores information, you will need to import both the sequence of your molecule and experimental information to create a complete CCPN project. At this point you will get a prompt to initiate the linkResonances process. Click **Yes** when asked, and **Yes** again to perfom this

process automatically (if possible). For more information on this process, see the links in **Note 1**.

6. Under the *Project* menu option, click **Save** or **Save as** to store your project. The data will be saved in a directory with subdirectories containing a set of XML files; all these have to stay together to remain valid for the CCPN framework.

If you created a CCPN project or downloaded a publicly available one (*see* **Note 1**) you can export data using the following general steps:

1. Under the *Project* menu option, click **Open**, navigate to the main directory containing your CCPN project in the pop-up window, and click **Open**.

2. Under the *Export* menu option, select the format/software to which you want to export the data. In the pop-up window, select the type of data you want to export to file, then select the specific data item(s) from the CCPN project. Click on the **Select export file** button, identify the file name you want to export the data to, and click **Select**. Click **Export … file** to start the export process.

3. Depending on the type of format and data, you might get pop-ups that ask you to validate information or provide additional input. Press on the **?** button in these popups for additional information.

3.2 TALOS+: Chemical Shift-Derived Dihedral Angle Restraints

J-coupling and secondary chemical shifts can be used to define restraints on the torsion angles of the chemical bonds, typically ϕ, ψ, and χ_1 angles can be generated and included. They can be calculated applying the Karplus equation [24] to the measured J-couplings, or estimated from chemical shifts in programs such as TALOS+ [55], DANGLE [56] or CSI [57]. The chemical shift-based methods are now the most commonly used and have good accuracy; DANGLE can be run from the CCPN software, but we will describe here the more commonly used TALOS+ approach.

TALOS+ predicts ϕ and ψ backbone torsion angles from a combination of available chemical shifts (Hα, Cα, Cβ, CO, N) for a protein sequence; it is an empirical method supported by a database of existing information on chemical shifts and backbone torsion angles. Before executing TALOS+, it is essential to ensure that your chemical shifts are correctly referenced (*see* **Note 2**), and that you have the correct input file (*see* **Note 3**). We here describe the use of TALOS + from the web-based server; other options are available (local installation or the WeNMR server, *see* Subheading 3.3).

1. Connect to the TALOS+ Web server: http://spin.niddk.nih.gov/bax/nmrserver/talos.

2. Under the *Chemical Shift Input* tab, click **Browse** and select your chemical shift input file (*see* **Note 3**).

3. Under the *Prediction Options* tab, you can deselect the **Apply Offset Correction** tick mark if you are sure your chemical shifts are correctly referenced (*see* **Note 2**).

4. Under the *Submission Details* tab, enter your email address (twice) and click **Submit** to start TALOS+. The screen will change, and you will after a few minutes receive an email from the "NMR Server Agent" with details on where to retrieve the results.

5. Create a directory where you want to store the results, open the email, and save all six files to this directory (detailed information on these files is available from the TALOS+ server).

6. In order to convert the TALOS+ predictions to accurate dihedral angle restraints, they have to be manually inspected. Connect to the jRAMA+ online viewer: http://spin.niddk. nih.gov/bax/software/TALOS+/JRAMA+.

   The first time you access this page you have to explicitly state that you trust this server; click the tick mark box followed by **Run**.

7. Click on the TALOS+ image in the top left corner; a pop-up will appear. In this popup, click on the **File** menu option (top left) and select **Open Prediction Files**. In the file window that appears, navigate to the directory you created to store the TALOS+ results, and select the pred.tab file. A set of other windows will now appear; the *RCI-S2 and Secondary Structure Plot* gives an overview of the chemical shift predicted backbone dynamics (from RCI [58]) and secondary structure for your protein.

8. Go to the window that contains your protein sequence with a color-coded box for each amino acid. The color coding indicates the following:

   (a) **Green**: the prediction is "Good" for this residue and can be used to create a dihedral angle restraint.

   (b) **Yellow**: the prediction is "Ambiguous" and should be manually inspected before use.

   (c) **Red**: the prediction is "Bad" and should not be used.

   (d) **Blue**: the residue is highly "Dynamic" and a prediction is not possible.

   If you click on a box, the other windows will update to show detailed information about the prediction for this residue, and you can examine the ϕ/ψ distributions of the detected matches and override the TALOS+ decisions on which residues should be included in the prediction and which ones are outliers. You can change the prediction status of this residue selecting a box at the bottom of the main window—only residues with "Good" status will be written out the dihedral restraint file as discussed in the next point.

9. When you are satisfied that only well-predicted residues have "Good" status, you can export the dihedral restraints from the **Tools** menu option. Write out both Xplor and Cyana angle restraint files.

3.3 CS-ROSETTA: Chemical Shift-Based Structure Calculation

The CS-ROSETTA3 protocol [51] makes use of the original CS-ROSETTA protocol [7], using a new fragment selection method (Vernont et al. personal communication) and Rosetta 3.3, and is powered by the WeNMR Grid [59]; *see* **Note 4** for more information on the WeNMR project and how you can access its resources. This tutorial guides you through the steps for setting up a basic CS-ROSETTA run, and how to interpret the results. First create a working directory, download the supplementary data associated with this chapter from the *Springer extra* Website at http://extras.springer.com, and unpack it with the following commands:

```
mkdir OR36
cd OR36
tar xvfz OR36_blind.tar.gz
```

Before submitting your chemical shift list to CS-ROSETTA, it is recommended to check the flexibility of your protein. If there are flexible ends, we suggest you truncate them from your chemical shift list. CS-ROSETTA predicts structure, so performs best for parts of a protein that have structure. To identify flexible residues, run the TALOS+ server as described before or via the WeNMR infrastructure:

1. Connect to the WeNMR TALOS+ Web server: http://haddock.science.uu.nl/enmr/services/TALOS.

2. Give a name to your run.

3. Select as file type to submit: BMRB3.1.

4. Select as chemical shift file to submit the "`originalData/bmrb31.str`" file from the OR36 directory you created previously.

5. Click on **Submit.**

6. You will be presented with a link to the result page. The result page for this system (Fig. 2) shows the predicted phi and psi angles for the residues, and the uncertainty in the prediction. If the respective bar is colored green, then the prediction is reliable. If the bar is red, the prediction is ambiguous. If the bar is absent, no prediction is made, or the respective residue is dynamic. Several consecutive dynamic residues suggest the presence of flexible parts. TALOS+ does not predict the torsion angles for the first and last residue in the sequence. Our target in this tutorial case seems thus to be ordered (i.e. does
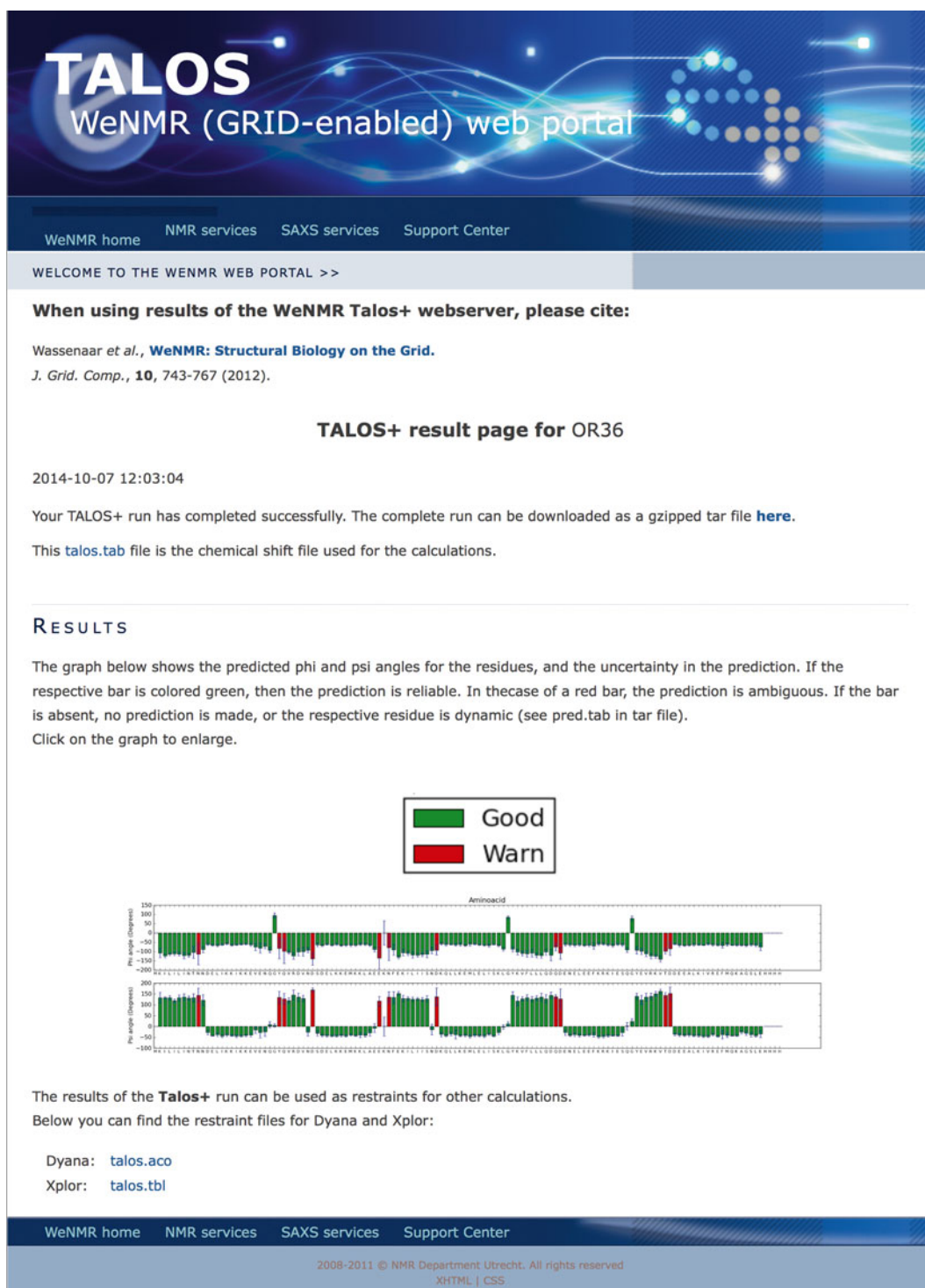
Fig. 2 Example of TALOS output from the WeNMR server based on the chemical shifts for entry 2lci. (The server is accessible via the "NMR" ->"Chemical Shifts" menu of www.wenmr.eu)

not seem to contain flexible ends). The last Histidines in the sequence are not predicted because no chemical shifts are available for those.

7. Before proceeding save the "`talos.tab`" file provided on the results page. We will use it for the CS-ROSETTA server submission. Edit this file and remove from the sequence the last six histidines since no chemical shifts are available for them. Also delete the chemical shift entries for residue 131 (one of the last histidines). It does not make sense to include them for CS-ROSETTA calculations.

Now set up and run the CS-ROSETTA calculations:

1. Connect to the WeNMR CS-Rosetta3 web portal:

   http://haddock.science.uu.nl/enmr/services/CS-ROSETTA3.

2. Figure 3 shows the web form you have to fill in. There are seven fields you can fill in:

   (a) The run name: has to be unique, has to be alpha numerical, and maximally 20 characters long.

   (b) The data format of the chemical shift list: you can supply your chemical shift list in three data formats: TALOS, BMRB2.1, and BMRB3.1.

   (c) The chemical shift list: the location of the chemical shift file on your local computer.

   (d) The number of models to generate: 10,000 is the maximum for a default account.

   (e) Option to automatically exclude flexible tails: we suggest you to do this manually as explained above.

   (f) Rescoring options: After generating your models, they have to be scored. ROSETTA3.3 does this with an all atom energy score (raw score). There are two additional rescoring algorithms: chemical shift (CS) rescoring, and DP rescoring (based on unassigned NOE peak lists) [60]. For DP rescoring different types of NOE data can be supplied (outside the scope of this tutorial).

3. To set up your tutorial run, choose the following options for the above:

   (a) Run name: OR36.

   (b) Type of file to submit: TALOS.

   (c) Chemical shift list: use the "`talos.tab`" file you just saved and edited.

   (d) Leave the number of models to their default value.

Fig. 3 CS-ROSETTA web form on the WeNMR server. (The server is accessible via the "NMR" -> "Structure Calculation" menu of www.wenmr.eu)

(e) Leave the remove flexible parts box unticked.

(f) Chemical shift rescoring is selected by default; leave this.

(g) Enter your username and password and submit the calculations.

After a successful submission you get a link to your personal result page (which is also mailed to you), on which you can track your run, and find the results when the run will have completed (which can take a few days depending on the system size and load on the server). The result page present the top five models based on various scoring scheme and also indicates the reliability of the predictions based on convergence and energetics criteria.

You can directly validate the CS-ROSETTA calculation results using iCING (*see* Subheading 3.6).

3.4 CYANA Structure Calculation from NOESY Peak Lists

A cross-peak in a NOESY spectrum indicates spatial proximity between two atoms; under idealized asumptions the intensity of the NOE peak (I) is proportional to the interatomic distance (r) as $I \sim 1/r^6$. Each NOESY peak can therefore be converted into an interatomic distance after determination of the proportionality constant (a process called "distance calibration"). Because the intensity–distance relationship is only approximate in practice, due to multiple complicating effects such as dynamics and the presence of many interacting spins, and as result of limitations in most structure calculation protocols, a distance range is usually assumed when converting the NOE peak intensity into a distance restraint. This range tends to be between 1.8 and 6.0 angstroms, with various ranges depending on the intensity of the NOE peaks.

An NOE peak can only be used as a distance restraint if it can be assigned to atoms in the molecule based on their chemical shift values: this step is of crucial importance, as the structure calculation process may not be able to deal with too many wrong or highly ambiguous (when many atoms with similar chemical shift values can be assigned to an NOE) assignments. The manual assignment of NOEs is an intensive and time-consuming job: many protocols are available to do this automatically starting from spectra [61] or peak lists [12, 62, 63]. We will here describe the procedure starting from peak lists as implemented in CYANA 2.1: note that CYANA requires a license, but because of its speed and ease of implementation it is in our view the most convenient software to use at this stage.

1. Create a directory for the structure calculation with CYANA. The following files have to be present in this directory:

   (a) A sequence file with the protein amino acid sequence in XEASY/CYANA format (*see* **Note 5**). We will refer to this file as `protein.seq`, and assume it has 100 residues.

   (b) A chemical shift file in XEASY/CYANA format (*see* **Note 5**). We will refer to this file as `shifts.prot`.

   (c) A dihedral angle file (e.g. from TALOS+ in Subheading 3.2). We will refer to this file as `dihedrals.aco`.

(d) One or more peak lists from NOESY spectra in XEASY/CYANA format (*see* **Note 5**). We will refer to these files as `peaks1.xpk, peaks2.xpk, …`.

(e) A CYANA initialization file (`init.cya`) containing the lines:

```
rmsdrange:=1..100
cyanalib
read seq protein.seq
```

Note that you can adjust the RMSD range to only include well-defined regions of the protein structure (if this information is known).

(f) A CYANA file with information for the structure calculation run (`AUTO.cya`) containing the lines:

```
peaks       := peaks1.xpk,peaks2.xpk # NOESY peak lists
prot        := shifts.prot           # Chemical shifts
constraints := dihedrals.aco         # Restraints
tolerance   := 0.030,0.040,0.25      # Tolerances for assignment
structures  := 100,20                # Initial/final structures
steps       := 10000                 # Calculation steps
randomseed  := 34983434              # Calculation seed

noeassign peaks=$peaks prot=$prot autoaco
```

The tolerances determine which atoms (based on their individual chemical shift values) are assigned to the NOESY peaks; wider tolerances will result in CYANA detecting more possible assignments for the NOESY peaks. If the tolerances are too narrow, less assignment possibilities will be found, but correct assignments might be missed. The above protocol assumes that the NOESY peaks are not yet assigned; additional protocols are available from the CYANA Web site and other resources.

2. Start CYANA with the command:

```
cyana
```

The program should automatically pick up all the relevant files and will assign the peak lists and calculate structures based on its internal procedures.

3. If CYANA successfully finished, you will see a set of files with *final* in the name that contain the following information:

(a) `final.pdb`: The atom coordinates of all models in the final structure ensemble.

(b) `final.upl`: An overview file containing information on the structure calculation. Particularly relevant here is the list of violated constraints: if they are violated in most structures (full sequence of+ or * signs) you should go back to the original peak lists with the program you were using for spectrum analysis and check whether the NOE peak and/or assignments make sense.

(c) final.aco, final.upl: Constraint files used in final calculation round.

(d) protein-final.prot**:** An XEASY file containing the final chemical shift assignments.

You should also check for "******* WARNING" messages in the AUTO.out file to make sure there were no problems during the calculations. Finally, the files ending in –cycle7.peaks and –cycle7-ref.peaks contain the original peak lists with information adapted based on final information from the structure calculation.

4. You can directly validate the CYANA calculation results using iCING (*see* Subheading 3.6). As CYANA uses a limited force field, we also recommend to water-refine the final structures with the protocol from Subheading 3.5 (*see* **Notes 5** and 7 on how to generate the CNS input via CCPN).

3.5   NMR Structure Refinement with CNS/ RECOORD

For the structure calculation part we are going to describe the use of the program CNS [13, 14] with a simulated annealing protocol derived from ARIA [64] followed by refinement in explicit solvent [10]. All the scripts mentioned in this section can be downloaded from the RECOORD [11] Webpage (*see* Table 1).

1. **Download**: Create a folder where you will run the calculations, download there the tar file containing the RECOORD scripts and decompress it:

```
mkdir struct-calc
cd struct-calc/
wget http://www.ebi.ac.uk/pdbe-
apps/nmr/data/recoord/RECOORDscripts-cns1.3.tgz
tar xzfv RECOORDscripts-cns1.3.tgz
```

In case the wget command does not work, use a Web browser to download the scripts manually from the RECOORD webpage (*see* Table 1). This tutorial uses CNS version 1.3.

2. **Initialise**: Before starting the calculations, you need to set up your current path for the scripts to work. In order to do this, you need to edit the file "`changeScriptsDir.sh`" located in RECOORDscripts-cns1.3 and change the directory path for `newDir` in line 8 with your current path and execute it:

```
./changeScriptsDir.sh
```

3. **Get data**: Make sure you have a working version of CNS set up; the last step is then to create a working directory and assigning a project name for the protein you are working on. This project name will be used to generate the file names at the different stages of the protocol. We will use as example the OR36 data for the PDB 2LCI structure with the corresponding NMR restraints available for this entry from the BioMagResBank (BMRB) [65]. First download and rename the PDB structure file:

```
mkdir 2lci
cd 2lci
wget http://www.ebi.ac.uk/pdbe-srv/view/files/2lci.ent.gz
gunzip pdb2lci.ent.gz
mv pdb2lci.ent 2lci.pdb
```

Then get the NMR restraints from this link:

```
http://www.bmrb.wisc.edu/servlets/MRGridServlet?pdb_id=2l
ci&min_items=0&block_text_type=3-converted-DOCR
```

In the result table, click on the number in the "distance" row under the "XPLOR/CNS" column, then click on the link in the "mrblock_id" column. Copy and paste these restraints in a text file called unambig.tbl in the 2lci/ directory (*see* **Note 6**). Alternatively you can export CNS restraints from CCPN projects (*see* **Note 7**) or use the data for OR36 directly from the "restraints/cns/" directory of the example project.

4. **Generation of molecular topology files**: We can generate the molecular topology either from the primary sequence or from a PDB coordinate file, depending on availability (*see* **Note 8**). We will use here the downloaded PDB file:

```
../RECOORDscripts-cns1.3/generate.sh 2lci.pdb
```

A topology file called 2lci_cns.mtf will be generated (note that this name is based on the name of the input file and will be different when you try other examples). You should check the ERRORS_generate file created inside the 2lci/ folder: in this particular case, you can see that the script reported many nomenclature errors which can be ignored at this stage.

A new pdb file called `2lci_cns.pdb` is also generated with the proper CNS nomenclature. You can display this structure in your favorite molecule viewer to verify it is correct.

5. **Generation of extended starting structure:** The next step is the generation of an extended starting conformation which will be used as input in the simulated annealing protocol:

```
../RECOORDscripts-cns1.3/generate_extended.sh 2lci_cns.mtf
```

The extended structure is in the file `2lci_cns_extended.pdb`. You should check the `ERRORS_generate_extended` file for errors, and again check the generated file in your favourite molecule viewer.

6. **Simulated annealing stage**: For the structure calculation itself, we can use three different types of restraints (if available): `unambig.tbl` (NOE distance restraints), `hbonds.tbl` (hydrogen bond restraints) and `dihedrals.tbl` (dihedral angle restraints). The script "`annealing.sh`" will generate a CNS parameter file (`run.cns`) with all details and specifications for the structure calculation protocol, and will start the calculation. This script should be run from a higher level than the previous two:

```
cd ..
RECOORDscripts-cns1.3/annealing.sh 2lci
```

Individual job files will be generated and executed for each model you want to calculate. By default two models will be generated in the created `str/` folder, with name similar to `2lci_cns_[1-2].pdb`. The CNS input and output files can be found in the directory `cnsRef/`, together with possible error files (*see* **Note 9**). The header of every PDB file generated contains information about violations and energy values.

7. **Water refinement stage:** Once the simulated annealing phase is finished and all resulting structures have been written into the `str/` directory, we can proceed to water refinement:

```
RECOORDscripts-cns1.3/re_h2o.sh 2lci
```

In the `str/` directory, a new directory called `wt/` will be created, the best energy structures will be copied there and subsequently refined (*see* **Note 10**).

3.6 Structure Validation and Quality Assessment

The 3D protein structures are the end result of a highly complex procedure involving interpretation of experimental data and transformation of this data to in silico atomic coordinates. It is therefore essential that these structures are well-validated and analyzed to ensure their correctness before they can be used as models that represent the conformation of the protein in solution.

This validation and analysis happens on two levels: comparison of the structures to original experimental data such as restraints and chemical shifts, and assessment of the physical quality of the in silico molecule description in terms of known physicochemical properties, such as atom bonds, angles, dihedrals, packing, hydrogen-bond, and electrostatics. Many programs are available to do parts of these validations (*see* Vuister et al. [49] for a recent overview); we will here focus on CING [15], a recently developed package that assembles the results of different validation programs and presents them in a joint, interactive, Web-2.0-based validation report. As an equivalent alternative, PSVS can be used [66].

The CING package is free to download and install; however, its large number of dependencies on external software makes this a nontrivial exercise. A virtual linux image with all public-domain software installed can be requested from the authors, requiring only the installation of the external licensed packages. Here, we will use the recommended iCing web server to validate the structures.

The iCing server can accept structure coordinates in the form of PDB files. However, it is highly advisable to also supply the server with the experimental data. The prefered input format of the iCing server is a self-contained CCPN project with the structure, restraints, chemical shifts, peaks, etc, compressed as a .tgz file. Such a file can be generated and submitted directly from the ccpNmr Analysis program or can be generated using the FormatConverter (*see* **Notes 1, 5,** and 7).

1. **Access the iCING server**: To upload the CCPN project data, first create an archive file of your project:

```
tar cvfz myCcpnProjectArchive.tgz myCcpnProjectDirectory/
```

Point your browser to http://nmr.le.ac.uk and select the iCing option. In the upload panel (Fig. 4a) use the "Choose File" button, select your ccpn.tgz file, which will then be uploaded. For "Program" we select CCPN (the default). Note that iCing also natively accepts PDB formatted coordinate files, multiple files in Cyana format as a .tgz file (*see* **Note 11**) or a CING project as a .tgz file. Selecting "Forward" allows one to proceed to the next panel with "Criteria". Leave these
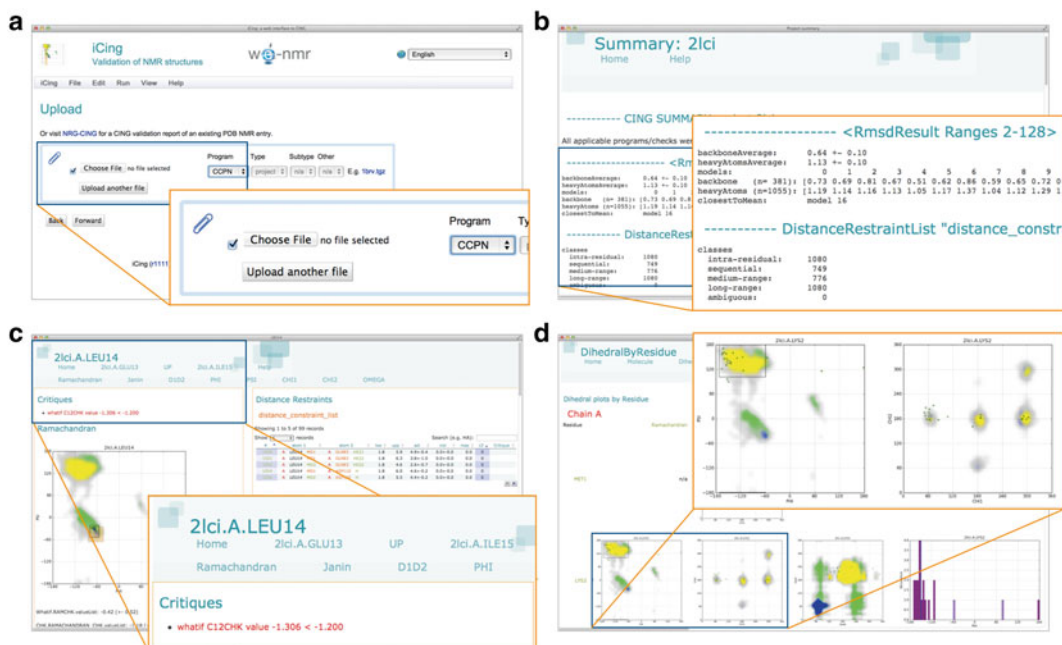
Fig. 4 iCing server and results. (**a**) iCing server upload page. Selection area is indicated by the *red box*. (**b**) Summary page for entry 2lci. (**c**) Residue-specific page for Leu14 of entry 2lci. The residue has a red ROG score as result of poor sidechain $\chi 1$-$\chi 2$ dihedral angle conformation (Janin plot not visible), Many related elements (previous and next residues, restraints, chemical shifts, etc.) can directly be accessed through the links on this page. (**d**) DihedralByResidue plots. A quick overview is obtained by scrolling down. Individual residue pages can be accessed through the links

set to their default values. Selecting "Forward" again, yields the "Options" panel that allows for selection of residues, or models in the ensemble. You can use this to set a range of residues to consider if you want to override the CING routines for determining this automatically. Generally, this is not required. Pressing "Forward" once more yield the Run-panel.

2. **Running CING**: Press "Submit" to start the CING validation analysis. The Output panel will appear, which can be manually updated to display the progress in the analysis. A typical CING validation analysis will take ~10–15 min. This is caused by the many programs that are run and the extensive graphical output that will be generated. When the program is finished, a link is presented to the on-line webpages with the results. The whole CING project file that includes all the data, the results from the individual programs and the all the webpages can also be downloaded. The results are stored anonymously on the iCing server and are automatically removed after 2 days.

3. **Using the CING results**: CING generates easily accessible, comprehensive, interactive HTML/Javascript-based validation

reports and directs the NMR spectroscopist to areas of the structure that have problems. Hyperlinks connect all elements at the different levels; e.g. the full molecule, chains, residues, atoms, as well as restraints, peaks, and chemical shifts. Lists are interactive entities that can be reordered and filtered by query. CING uses at all data levels a simple Red–Orange–Green (ROG) scoring, which depends upon the combined analysis of all results and allows CING to direct the user to important issues:

(a) **Red**: Potentially serious issues with the structures.

(b) **Orange**: Potential issues with the structures.

(c) **Green**: No issues detected.

In addition to this classification mechanism, CING displays the validation results in direct relation to the experimental data.

4. **Structure level results**: The link to the on-line webpages with the results first presents an overview of the various sections of the CING report. A link to the *summary* is best examined first (Fig. 4b). Here, an overview is given of the rmsd values within the ensemble, the overall CING [15], WHATIF [67] and PROCHECK_NMR [68] scores, as well as the results of the restraints analysis. CING also analyses, based upon circular variance criteria of the backbone dihedral angles ϕ and ψ, which residues should be considered well-defined (*see* **Note 12**); only these regions are used for the superposition of the structures and the related rmsd values, as are the PROCHECK_NMR scores and one of the two overal CING ROG scores; the other pertains to the full molecule. An ensemble of structures that has low green ($< \sim 20$ %) and high red ($> \sim 50$ %) overall ROG score should be labeled as highly troublesome.

5. **Residue level results**: Within the CING philosophy, a large emphasis is placed upon the validation of the individual residues. NMR assignment strategies are almost exclusively residue-based, NMR related parameters are residue dependent and the local nature of the NMR-derived restraints also correlates well with a residue-based approach. Structural properties can also be conveniently summarized at the residue level, and hence all residue-specifc elements are specified on the residue pages. Residue-specific Ramachandran plots (Fig. 4c) are generated for each residue with all the individual conformers of the ensemble marked and ϕ, ψ restraints, if present, indicated. Inspection of the sidechain $\chi 1$–$\chi 2$ dihedral angle distributions of the conformers in the ensemble is possible by means of the Janin plot, as also automatically generated by CING for each residue. The D1D2 plots display the conformation of the residue relative to the preceding and following residue in the chain. A residue page also contains all the experimental restraint

data involving the featured residue, as well as the "cri-tiques" defining its ROG scores. Thus, a residue page presents a comprehensive account of all relevant information pertaining to a specific residue. Hyperlinks connect the page to all other relevant pages; e.g. the previous and next residue in the chain, but also all residues connected through restraints or all its atoms and their chemical shifts. The analysis of the conformation of residues and identification of potential problems can also conveniently be done using the "Dihedral plots per residue" page (Fig. 4d), which displays the relevant Ramachandran, Janin and D1D2 plots of all residues sequentially, in one scrollable interface from which the relevant residue can also be selected.

6. **Other information**: It is useful to display the residue-specific ROG scores mapped onto a structure of the molecule. For this, macros for the structure visualization programs JMOL [69], YASARA [45], PyMol (The PyMOL Molecular Graphics System, Schrödinger, LLC.) and MOLMOL [70] are provided. They can be accessed by following the "Programs flat->Macros" link from the home page. Closely clustered red/orange residues in the structure are highly suspect and warrant further investigation.

The CING results are based, in part, upon the results of the programs PROCHECK_NMR [68] and WHATIF [67]. Direct links to the output of these programs are also provided.

4   Notes

1. To use the FormatConverter, install the latest stable version of the CCPN software, which can be downloaded from http://www.ccpn.ac.uk/downloads/stable. A less versatile but easier to use alternative is the FormatConverter web version at http://haddock.chem.uu.nl/enmr/format-converter.html. Further information on the FormatConverter is available from http://www.ccpn.ac.uk/software/fcfolder, as well as a detailed tutorial at http://www.ccpn.ac.uk/software/tutorials/intro. Examples of full CCPN projects that can be used for structure calculation are available from the RECOORD project at http://www.ebi.ac.uk/pdbe-apps/nmr/recoord/ by clicking the *CCPN projects* link near the bottom of the page.

2. Chemical shifts are calculated from absolute frequencies relative to a reference frequency; this way the positions of NMR resonances can be expressed independently of the magnetic field strength. If this reference frequency is not correctly set all chemical shift values calculated with it will be equally offset from their true value. There are a host of programs available to check if

chemical shifts are correctly referenced, and to suggest a correction to obtain their true value ("re-referencing" of the chemical shifts): AVS [71], LACS [72, 73], SPARTA+ [23], PANAV [74], CheckShift [75], ShiftX2 [19], and VASCO [76].

3. Load a CCPN project with the FormatConverter or import a sequence and chemical shift list. Go to the *Export* menu and select *TALOS*. Open the "*project export menu*", click on **Select export file** and define the file you want to write to. The chemical shift data to export can be selected next to *Select shift list to export:* (in the example OR36 project this will be "`1:bmrb21.str`", indicating a shift list with CCPN serial number `1` and name `bmrb21.str`). Click **Export project file**, then **OK** in the popup window (you can here change the sequence numbering should you wish to do so), then **OK** again if the export finished successfully. The exported TALOS file should contain sequence as well as chemical shift information and can be uploaded to the TALOS+ server. The procedure for CS-ROSETTA is the same, although in this case it is useful to first run TALOS+ and then, in the input file to be uploaded to the CS-ROSETTA server, first manually remove N- and C-terminal regions from the sequence that have no chemical shifts, as well as residues that are labeled as "Dynamic" by TALOS.

4. WeNMR [59] provides an extensive infrastructure that allows users to run up-to-date and computationally demanding software online. In order to use it, you will have to obtain a personal X509 certificate and register with the WeNMR project and virtual organization. Instructions for this can be found at the following site: http://www.wenmr.eu/wenmr/access/registration. Relevant software for this protocol available on WeNMR is indicated in Table 1.

5. (a) Export of sequence, peak list and chemical shift data for CYANA. First load a CCPN project with these data with the FormatConverter (*e.g.* the example OR36 project). (i) Sequence export: Go to the *Export* menu and select *Cyana*. Open the "*sequence export menu*" and select the molecule for which you want to export the sequence next to *Select chains to export:* (in the example OR36 project this will be "`Molecularsystem:'A'`", indicating a molecule with chain code "A" in the molecular system `Molecularsystem`). Click on **Select export file** and define the file you want to write to, then select the CYANA version next to *Cyana version:* Click **Export sequence file**, then **OK** in the popup window (you can here change the sequence numbering should you wish to do so), then **OK** again if the export finished successfully. (ii) Chemical shift list export: Go to the *Export* menu and select *XEasy*. Open the "*shift export menu*" and select the chemical shift list you want to export next to *Select shift list to*

export: (in the example OR36 project this is "`1:bmrb21.str`"). Click on **Select export file** and define the file you want to write to, then **select the button** next to *Use CYANA2.1 atom names:* if you are using the CYANA 2.1 version or higher. Click **Export shifts file**, then **OK** in the popup window, and **OK** again if the export finished successfully. (iii) Peak list export: Go to the *Export* menu and select *XEasy*. Open the "*peaks export menu*" and select the peak list you want to export next to *Select peak list to export:* (in the example OR36 project there will be three peak lists, ending in "`nnoe_raw`", "`alinoe_raw`" in "`aronoe_raw`"; the text in front of this indicates the CCPN experiment and processing information). Click on **Select export file** and define the file you want to write to, then **select the button** next to *Write as CYANA format:* Click **Export peaks file**, then **OK** in the popup window. In the next window you can define the order of the peak list dimensions in the output file, press **OK** when this is fine. Finally, press **OK** again if the export finished successfully. Note that for reasons particular to the XEASY format assignments in your peak list can only be written out if you first exported a chemical shift list.

(b) Import of coordinates, distance and dihedral restraint lists from CYANA. Create a new CCPN project or load an existing one in the FormatConverter. (i) Coordinate import: Go to the *Import* menu, *Single files*, *Coordinates* and select *PseudoPdb* (note that selecting *Pdb* will only import official files from the PDB). Click on **Select files**, select the file (for OR36, try "`restraints/coordinates.pdb`") and click **Select**. Click **IMPORT**. Then two situations are possible. (1) If your sequence was already loaded into the CCPN previously, click **OK** in the popup window after selecting the correct CCPN molecular system, **OK** again after successful import. (2) If you have no sequence in your project, you will first have to create one matching the molecule from the coordinates. In the first popup window you can change the sequence information and/or molecule name. Click **OK** when done, in the next popup give a name to your molecular system and press **OK**, then click **No** in the next popup unless you have a homomultimer. You might then get additional popups asking you to disambiguate atom names from the coordinate data. Finally click **OK** again after successful import (ii) Distance restraints: Go to the *Import* menu, *Single files*, *Distance constraints* and select *Cyana*. Click on **Select file**, select the file (for OR36, try "`restraints/cyana/distances.upl`") and click **Select**. You can also select the matching lower distance limits "`distances.lol`" file at **Select file (optional)**. Select the correct CYANA version (use 2.1 for the most recent versions), and click **IMPORT**. In the next popup you will be

able to select the name of an existing CCPN structure generation, or create a new one. In the next popup give your distance restraint list a name; another popup will appear asking you to read in a lower distance limit file in case you haven't specified it. Click **OK** again after successful import. Finally, click **No** for the popup asking you to run linkResonances—we will do this after the next step. (iii) Dihedral restraints: Go to the *Import* menu, *Single files*, *Dihedral constraints* and select *Cyana*. Click on **Select file**, select the file (for OR36, try "`restraints/cyana/dihedrals.aco`") and click **Select**. Select the correct CYANA version (use 2.1 for the most recent versions), and click **IMPORT**. In the next popup you will be able to select the name of an existing CCPN structure generation, or create a new one. In the next popup give your dihedral restraint list a name; click **OK** again after successful import. When you are done importing all connected restraint files, click **Yes** to the linkResonances question, then **Yes** again (*see* Subheading 3.1, **step 5**).

6. If dihedrals or any other types of restraints are available, they can be obtained in a similar way. The names assigned will be dihedrals.tbl and hbonds.tbl.

7. Export of distance and dihedral restraint lists as well as coordinates (PDB-like files) for CNS/XPLOR. First load a CCPN project with these data with the FormatConverter (e.g. the example OR36_withRestraints project). (i) Distance restraints export: Go to the *Export* menu and select *Cns*. Open the "*distanceConstraints export menu*" and select the distance constraint list for which you want to export the sequence next to *Select distance constraint list to export:* (in the example OR36 project this will be "`1:1:distance constraint list`", indicating a distance constraint list with serial number "1" in the structure generation with serial number "1"). Click on **Select export file** and define the file you want to write to. Click **Export distanceConstraints file**, then **OK** in the popup window (you can here change the chain codes and sequence numbering should you wish to do so), then **OK** again if the export finished successfully. (ii) Dihedral restraints export: This is the same procedure as for distance restraints except use the *dihedralConstraints* menu, and in the example OR36 project select "`1:2:dihedral constraint list`"). (iii) Coordinates export: Open the "*coordinates export menu*" and select the structure(s) you want to export next to *Select structures to export:* in the example OR36 project there will be one structure, "`MolecularSystem:1:model_1`", indicating a set of coordinates for "`model_1`" relating to molecular system "`MolecularSystem`" in structure ensemble with serial "1". Click on **Select export file** and define the file you want to write to, then click **Export coordinates file**.

8. This only works if a PDB coordinates file is available. Otherwise use generate_seq.inp and generate_template.inp from CNS to create such a PDB.

9. Once everything is set up in a proper way, you can edit the script and change the protocol parameters where necessary. You can for example change the number of models to generate. It is set to 2 by default, but more common numbers would be 100 or 200. For more complex systems, you can switch to a longer annealing protocol, by doubling the number of steps to be carried out. Depending if you are going to use a cluster or your own computer, you should change the submit command. Remember also to change the sleep time between submitting jobs, especially if you are not using a cluster and you do not want to have 100 jobs running in your computer at the same time! In such a case choose a sleep time that matches the time needed for one structure calculation.

10. You should also edit this script and change the number of structures to refine since by default it is set to only 1. Increase this number to 25 models. They will be assigned names like `2lci_cns_w_[1-25].pdb`. The CNS input and output will be directed to the directory `cnsWtRef/`.

11. Preparing Cyana input to iCing. The Cyana `myMolecule.cyana.tgz` of `myMolecule` should contain `myMolecule.seq` (the cyana sequence file), `myMolecule.prot` (the corresponding cyana prot file), `myMolecule.pdb` (the Cyana generated structure), `myMolecule.aco` (optional dihedral restraints), `myMolecule.lol` (optional lower bounds), `myMolecul.upl` (optional upper bounds).

12. The precision of a structure can be estimated by measuring the conformational variance over an ensemble of models. Usually, this variance has been expressed as the positional root-mean-square deviation (rmsd) of the individual models from the mean structure. This parameter is useful for estimating the precision of the calculation, but does not report on the accuracy. The later can only be calculated if a standard reference is available.

References

1. Wuthrich K (1986) Nmr of proteins and nucleic acids. Wiley, New York, NY

2. Neuhaus D, Williamson MP (2000) The nuclear overhauser effect in structural and conformational analysis. Wiley, New York, NY

3. Altona C (1996) Vicinal coupling constants and conformation of biomolecules. In Harris DMG, a K R (eds) Encyclopedia of nuclear magnetic resonance. Wiley, London. pp 4909–4922

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

5. Bertini I, Luchinat C, Parigi G (2012) Towards mechanistic systems biology. Wiley-VCH Verlag GmbH, Weinheim, Germany. pp 154–171

6. Cavalli A, Salvatella X, Dobson CM, Vendruscolo M (2007) Protein structure determination from NMR chemical shifts. Proc Natl Acad Sci U S A 104:9615–9620

7. Shen Y et al (2008) Consistent blind protein structure generation from NMR chemical shift data. Proc Natl Acad Sci U S A 105:4685–4690

8. Wishart DS et al (2008) CS23D: a web server for rapid protein structure generation using NMR chemical shifts and sequence data. Nucleic Acids Res 36:W496–W502

9. Güntert P (1998) Structure calculation of biological macromolecules from NMR data. Q Rev Biophys 31:145–237

10. Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M (2003) Refinement of protein structures in explicit solvent. Proteins 50:496–506

11. Nederveen AJ et al (2005) RECOORD: a recalculated coordinate database of 500+ proteins from the PDB using restraints from the BioMagResBank. Proteins 59:662–672

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

13. Brünger AT et al (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. Acta Crystallogr D Biol Crystallogr 54:905–921

14. Brunger AT (2007) Version 1.2 of the Crystallography and NMR system. Nat Protoc 2:2728–2733

15. Doreleijers JF et al (2012) CING: an integrated residue-based structure validation program suite. J Biomol NMR 54:267–283

16. Luginbühl P, Szyperski T, Wüthrich K (1995) Statistical basis for the use of ^{13}cα chemical shifts in protein structure determination. J Magn Res 109:92

17. Kuszewski J, Qin J, Gronenborn AM, Clore GM (1995) The impact of direct refinement against 13C alpha and 13C beta chemical shifts on protein structure determination by NMR. J Magn Reson B 106:92–96

18. Neal S, Nip AM, Zhang H, Wishart DS (2003) Rapid and accurate calculation of protein 1H, 13C and 15N chemical shifts. J Biomol NMR 26:215–240

19. Han B, Liu Y, Ginzinger SW, Wishart DS (2011) SHIFTX2: significantly improved protein chemical shift prediction. J Biomol NMR 50:43–57

20. Xu XP, Case DA (2001) Automated prediction of 15N, 13Calpha, 13Cbeta and 13C' chemical shifts in proteins using a density functional database. J Biomol NMR 21:321–333

21. Williamson MP, Kikuchi J, Asakura T (1995) Application of 1H NMR chemical shifts to measure the quality of protein structures. J Mol Biol 247:541–546

22. Meiler J (2003) PROSHIFT: protein chemical shift prediction using artificial neural networks. J Biomol NMR 26:25–37

23. Shen Y, Bax A (2010) SPARTA+: a modest improvement in empirical NMR chemical shift prediction by means of an artificial neural network. J Biomol NMR 48:13–22

24. Karplus M (1963) Vicinal proton coupling in nuclear magnetic resonance. J Am Chem Soc 85:2870–2871

25. Kim Y, Prestegard JH (1990) Refinement of the NMR structures for acyl carrier protein with scalar coupling data. Proteins 8:377–385

26. Torda AE, Brunne RM, Huber T, Kessler H, Van Gunsteren WF (1993) Structure refinement using time-averaged J-coupling constant restraints. J Biomol NMR 3:55–66

27. Wagner G, Wüthrich K (1982) Amide protein exchange and surface conformation of the basic pancreatic trypsin inhibitor in solution. Studies with two-dimensional nuclear magnetic resonance. J Mol Biol 160:343–361

28. Pervushin K et al (1998) NMR scalar couplings across Watson-Crick base pair hydrogen bonds in DNA observed by transverse relaxation-optimized spectroscopy. Proc Natl Acad Sci U S A 95:14147–14151

29. Cordier F, Rogowski M, Grzesiek S, Bax A (1999) Observation of through-hydrogen-bond 2hJHC' in a perdeuterated protein. J Magnet Res 140:510–512

30. Bonvin AMJJ, Houben K, Guenneugues M, Kaptein R, Boelens R (2001) Rapid protein fold determination using secondary chemical shifts and cross-hydrogen bond 15N-13C" scalar couplings (3hbJNC"). J Biomol NMR 21:221–233

31. Bax A (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics. Protein Sci 12:1–16

32. Bax A, Grishaev A (2005) Weak alignment NMR: a hawk-eyed view of biomolecular structure. Curr Opin Struct Biol 15:563–570

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

34. Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A (1997) Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. Nat Struct Biol 4:732–738

35. Fushman D, Varadan R, Assfalg M (2004) Determining domain orientation in macromolecules by using spin-relaxation and residual

dipolar coupling measurements. Prog Nucl Magn Reson Spectrosc 44:189–214

36. Tjandra N, Garrett DS, Gronenborn AM, Bax A, Clore GM (1997) Defining long range order in NMR structure determination from the dependence of heteronuclear relaxation times on rotational diffusion anisotropy. Nat Struct Biol 4:443–449

37. Bertini I, Luchinat C, Parigi G, Pierattelli R (2005) NMR spectroscopy of paramagnetic metalloproteins. ChemBioChem 6:1536–1549

38. Banci L et al (2004) Paramagnetism-based restraints for Xplor-NIH. J Biomol NMR 28:249–261

39. Bertini I, Luchinat C, Parigi G (2002) Paramagnetic constraints: an aid for quick solution structure determination of paramagnetic metalloproteins. Concepts Magn Reson 14:259–286

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

41. Güntert P, Mumenthaler C, Wüthrich K (1997) Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 273:283–298

42. Hus JC, Marion D, Blackledge M (2000) De novo determination of protein structure by NMR using orientational and long-range order restraints. J Mol Biol 298:927–936

43. Case DA et al (2005) The Amber biomolecular simulation programs. J Comput Chem 26:1668–1688

44. van der Spoel D et al (2005) GROMACS: fast, flexible, and free. J Comput Chem 26:1701–1718

45. Krieger E, Koraimann G, Vriend G (2002) Increasing the precision of comparative models with YASARA NOVA–a self-parameterizing force field. Proteins 47:393–402

46. Montelione GT et al (2013) Recommendations of the wwPDB NMR validation task force. Structure 21(9):1563–1570

47. Kirchner DK, Güntert P (2011) Objective identification of residue ranges for the superposition of protein structures. BMC Bioinformatics 12:170

48. Mao B, Guan R, Montelione GT (2011) Improved technologies now routinely provide protein NMR structures useful for molecular replacement. Structure 19:757–766

49. Vuister GW, Fogh RH, Hendrickx PMS, Doreleijers JF, Gutmanas A (2013) An overview of tools for the validation of protein NMR structures. J Biomol NMR 58(4):259–285

50. Spronk C, Nabuurs SB, Krieger E (2004) Validation of protein structures derived by NMR spectroscopy. Prog Nucl Magn Reson Spectrosc 45:315–337

51. van der Schot G et al (2013) Improving 3D structure prediction from chemical shift data. J Biomol NMR 57:27–35

52. Rosato A et al (2009) CASD-NMR: critical assessment of automated structure determination by NMR. Nat Methods 6:625–626

53. Rosato A et al (2012) Blind testing of routine, fully automated determination of protein structures from NMR data. Structure 20:227–236

54. Vranken WF et al (2005) The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 59:687–696

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

56. Cheung M-S, Maguire ML, Stevens TJ, Broadhurst RW (2010) DANGLE: A Bayesian inferential method for predicting protein backbone dihedral angles and secondary structure. J Magn Reson 202:223–233

57. Wishart DS, Sykes BD (1994) Chemical shifts as a tool for structure determination. Methods Enzymol 239:363–392

58. Berjanskii MV, Wishart DS (2005) A simple method to predict protein flexibility using secondary chemical shifts. J Am Chem Soc 127:14970–14971

59. Wassenaar TA, et al (2012) WeNMR: structural biology on the grid. J Grid Comp 10:743–767

60. Huang YJ, Powers R, Montelione GT (2005) Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. J Am Chem Soc 127:1665–1674

61. Guerry P, Herrmann T (2012) Comprehensive automation for NMR structure determination of proteins. Methods Mol Biol 831:429–451

62. Linge JP, Habeck M, Rieping W, Nilges M (2003) ARIA: automated NOE assignment and NMR structure calculation. Bioinformatics 19:315–316

63. Raman S et al (2010) Accurate automated protein NMR structure determination using unassigned NOESY data. J Am Chem Soc 132:202–207

64. Linge JP, O'Donoghue SI, Nilges M (2001) Automated assignment of ambiguous nuclear overhauser effects with ARIA. Methods Enzymol 339:71–90

65. Seavey BR, Farr EA, Westler WM, Markley JL (1991) A relational database for sequence-specific protein NMR data. J Biomol NMR 1:217–236

66. Bhattacharya A, Tejero R, Montelione GT (2007) Evaluating protein structures determined by structural genomics consortia. Proteins 66:778–795

67. Vriend G (1990) WHAT IF: a molecular modeling and drug design program. J Mol Graph 8:52–56

68. Laskowski RA, Rullmann J, MacArthur MW (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. J Biomol NMR 8:477–486

69. Herráez A (2006) Biomolecules in the computer: Jmol to the rescue. Biochem Mol Biol Educ 34:255–261

70. Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14(51–5):29–32

71. Moseley HNB, Sahota G, Montelione GT (2004) Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. J Biomol NMR 28:341–355

72. Wang L, Eghbalnia HR, Bahrami A, Markley JL (2005) Linear analysis of carbon-13 chemical shift differences and its application to the detection and correction of errors in referencing and spin system identifications. J Biomol NMR 32:13–22

73. Wang L, Markley JL (2009) Empirical correlation between protein backbone 15N and 13C secondary chemical shifts and its application to nitrogen chemical shift re-referencing. J Biomol NMR 44:95–99

74. Wang B, Wang Y, Wishart DS (2010) A probabilistic approach for validating protein NMR chemical shift assignments. J Biomol NMR 47:85–99

75. Ginzinger SW, Gerick F, Coles M, Heun V (2007) CheckShift: automatic correction of inconsistent chemical shift referencing. J Biomol NMR 39:223–227

76. Rieping W, Vranken WF (2010) Validation of archived chemical shifts through atomic coordinates. Proteins 78:2482–2489