J Biomol NMR (2015) 61:83–87
DOI 10.1007/s10858-014-9878-3

NMR STRUCTURE NOTE

NMR structure determination of the protein NP_344798.1 as the first representative of Pfam PF06042

Biswaranjan Mohanty · Pedro Serrano · Michael Geralt · Kurt Wüthrich

Received: 26 September 2014 / Accepted: 17 November 2014 / Published online: 28 November 2014

Biological context

The protein NP_344798.1 from *Streptococcus pneumoniae* (TIGR4) belongs to the Pfam protein family PF06042, which currently contains 786 sequences from 739 species (http://pfam.xfam.org/family/PF06042). Based on sequence analyses, NP_344798.1 has been identified as a member of the nucleotidyltransfererase (NTase)-fold superfamily and annotated as a protein domain of unknown function, DUF925 (Kuchta et al. 2009). The bioinformatics data further suggest that all PF06042 members contain a single domain and are active NTases, as they contain the characteristic functional catalytic residues (Kuchta et al. 2009). Nonetheless, due to scarcity of experimental data, no specific biological role of these enzymes could as yet be ascertained. The NMR core of the Joint Center for Structural Genomics (JCSG) targeted NP_344798.1 to obtain a first structure for the PF06042 family, and the J-UNIO protocol for automated NMR structure determination (Serrano et al. 2012) was applied to this 191-residue protein. The structure now presents a foundation for obtaining new insights into the function of proteins in this family by structure comparison with other related structures in the PDB, experimental studies of substrate binding and specificity, and homology modeling of other proteins in PF06042.

B. Mohanty · P. Serrano · M. Geralt · K. Wüthrich (✉)
Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA 92037, USA
e-mail: wuthrich@scripps.edu
URL: http://www.jcsg.org

B. Mohanty · P. Serrano · M. Geralt · K. Wüthrich
Joint Center for Structural Genomics, La Jolla, CA 92037, USA

B. Mohanty · K. Wüthrich
Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla, CA 92037, USA

Present Address:
B. Mohanty
Medicinal Chemistry, Monash Institute of Pharmaceutical Sciences, Monash University, 381 Royal Parade, Parkville, VIC 3052, Australia

Methods

Protein preparation

The protein NP_344798.1 was expressed in *E. coli*, using the pSpeedET-NP_344798.1 plasmid produced by the JCSG crystallomics core, and was purified following our standard protocol (Serrano et al. 2012). Micro-scale exploratory experiments were pursued with the [u-^{15}N]-protein (Pedrini et al. 2013; Serrano et al. 2012). For the NMR structure determination, a 1.2 mM solution of the [u-^{13}C,^{15}N]-labeled protein was prepared, with 20 mM sodium phosphate at pH 6.0, 50 mM sodium chloride and 4.5 mM NaN_3 in 5 % (v/v) D_2O/95 % H_2O. To remove oxygen prior to data collection, the sample was treated with Argon gas in the NMR tube.

NMR spectroscopy

2D [^{15}N,^{1}H]-HSQC spectra and APSY-NMR datasets (Hiller et al. 2005, 2008) were recorded on a Bruker Avance 600 MHz spectrometer equipped with a CPTCI HCN z-gradient cryoprobe. For the 5D APSY-HACACONH and 5D APSY-CBCACONH experiments, 24 2D projections were acquired, and a 4D APSY-HACANH experiment was acquired with 31 projections. 32 transients

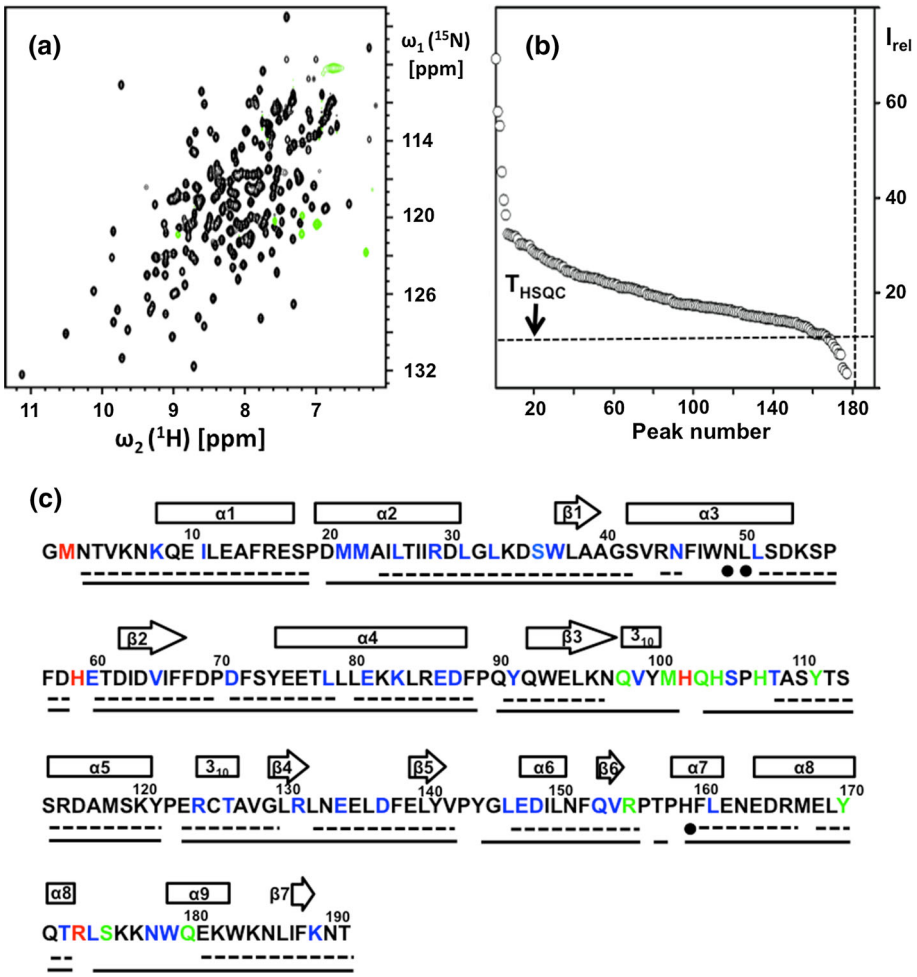

Fig. 1 Characterization of the NP_344798.1 solution used for the NMR structure determination and survey of the polypeptide backbone assignments. **a** 2D [^{15}N,^{1}H]-HSQC spectrum at 600 MHz collected with a 1.2 mM solution of [u-^{13}C,^{15}N]-labeled NP_344798.1, pH 6.0, T = 298 K. **b** NMR-Profile. The number of backbone amide cross peaks expected from the amino acid sequence and the intensity threshold, T_{HSQC} (Pedrini et al. 2013), are indicated by *dotted vertical and horizontal lines*, respectively. **c** Amino acid sequence, with the following color-coded information: Residues for which the backbone amide cross peaks in the [^{15}N,^{1}H]-HSQC spectra are missing (*red*), partially overlapped (*blue*), or have intensity below T_{HSQC} (*green*). The *dotted underline* indicates correct assignments obtained automatically with UNIO-MATCH. *Filled circles* denote three residues for which UNIO-MATCH yielded erroneous assignments, which were indentified during the interactive validation and replaced with the correct chemical shifts prior to the next steps of the structure determination (see text). The *solid horizontal line* represents the final assignments obtained after interactive validation and extension of the UNIO-MATCH results using the NOESY data (see text). Above the sequence, the locations of regular secondary structures are indicated by *rectangles* for the helices and *arrows* for the β-strands

were accumulated for each APSY projection. The total acquisition time for the three experiments was 84 h. The digital resolution of the 2D projections was 2,048 × 128 complex points for 4D APSY-HACANH, 2,048 × 96 points for 5D APSY-HACACONH, and 2,048 × 90 points for 5D APSY-CBCACONH. Prior to Fourier transformation, the time domains data were multiplied in both dimensions with a 45°-shifted sine bell (DeMarco and Wüthrich 1976) and zero-filled to 256 complex points in the indirect dimension.

3D [^{1}H,^{1}H]-NOESY-^{15}N-HSQC, 3D [^{1}H,^{1}H]-NOESY-$^{13}C_{ali}$-HSQC and 3D [^{1}H,^{1}H]-NOESY-$^{13}C_{aro}$-HSQC spectra were recorded on a Bruker Avance 800 MHz spectrometer equipped with a TXI HCN z-gradient room temperature probe. The mixing time was 60 ms and the relaxation delay was 1 s. The 3D [^{1}H,^{1}H]-NOESY-^{15}N-HSQC spectrum was acquired with 2,048 × 90 × 320 complex points, a spectral width of 30 ppm and with the carrier at 118 ppm. For the 3D [^{1}H,^{1}H]-NOESY-$^{13}C_{ali}$-HSQC spectrum, 2,048 × 100 × 330 complex points were acquired, with a spectral width of 32 ppm and the carrier at 33 ppm. The 3D [^{1}H,^{1}H]-NOESY-$^{13}C_{aro}$-HSQC spectrum was acquired with 2,048 × 80 × 330 complex points, a spectral width of 31 ppm and the carrier at 122 ppm. The NOESY data sets were zero-filled to 2,048 × 256 × 512 complex points, multiplied with a squared cosine window

Table 1 Input and calculation statistics for the NMR structure determination of the protein NP_344798.1 in aqueous solution at pH 6.0 and T = 298 K

Quantity	Value[a]
NOE upper distance limits	4,090
Intraresidual	993
Short-range	1,055
Medium-range	904
Long-range	1,138
Dihedral angle constraints	799
Residual target function value (Å^2)	3.9 ± 0.43
Residual NOE violations	
Number ≥0.1 Å	41 ± 6
Maximum (Å)	0.2 ± 0.19
Residual dihedral angle violations	
Number ≥ 2.5°	1 ± 1
Maximum (°)	4.79 ± 1.5
AMBER energies (kcal/mol)	
Total	−7,843 ± 144
van der Waals	−705 ± 35
Electrostatic	−8,779 ± 112
RMSD from the mean coordinates[b] (Å)	
Backbone (2–191)	0.66 ± 0.09
All heavy atoms (2–191)	1.05 ± 0.09
Ramachandran plot statistics[c]	
Most favoured regions (%)	75.1
Additional allowed regions (%)	23.0
Generously allowed regions (%)	1.6
Disallowed regions (%)	0.3

[a] Except for the top six entries, which represent the input generated for the final cycle of structure calculation with UNIO-ATNOS/CANDID and CYANA 3.0, average values and SD for the 20 energy-minimized conformers are given

[b] The numbers in parentheses indicate the residues for which the RMSD was calculated

[c] As determined by PROCHECK (Laskowski et al. 1993)

in both proton dimensions and with a 45°-shifted sine bell (DeMarco and Wüthrich 1976) in the ^{15}N or ^{13}C dimension, and processed using Topspin 2.1. The three NOESY experiments were acquired in 9 days. Chemical shifts were referenced to DSS.

NMR structure determination with J-UNIO

Automated chemical shift assignment with UNIO-MATCH 2.0.1 (Volk et al. 2008) and UNIO-ATNOS/ASCAN 2.0.1 (Fiorito et al. 2008) was interactively validated and extended based on the NOESY data, using CARA (Keller 2004). Structure calculation and validation was performed following the J-UNIO protocol (Serrano et al. 2012), using the software UNIO-ATNOS-CANDID (Herrmann et al. 2002a, b) and CYANA (Güntert et al. 1997). The 40 conformers with the smallest residual target function values from cycle 7 of the structure calculation were energy-minimized with OPALp (Luginbühl et al. 1996; Koradi et al. 2000). 20 energy-minimized conformers were selected on the basis of the J-UNIO validation criteria (Serrano et al. 2012) to represent the protein structure, which was analyzed with MOLMOL (Koradi et al. 1996).

Results

In exploratory micro-scale experiments, NP_344798.1 was assessed for protein structure determination from its NMR-Profile (Pedrini et al. 2013). 177 out of 181 backbone amide cross peaks expected from the amino acid sequence were observed, which included a few peaks with low intensity. Nonetheless, most of the peaks were well-resolved and exhibited quite uniform intensities, indicating that the protein was a promising target for structure determination by solution NMR, in spite of its rather large size. This was further supported by the fact that 167 of the 177 observed ^{15}N–^{1}H cross peaks had intensities above the T_{HSQC} threshold (Pedrini et al. 2013), so that good quality APSY-NMR data sets could be expected. For confirmation, a new NMR-Profile was recorded with the protein solution used for the structure determination, which confirmed the conclusions from the microscale experiments (Fig. 1a, b).

NMR structure determination

Initial automated polypeptide backbone assignment was based on the three experiments 4D APSY-HACANH, 5D APSY-HACACONH and 5D APSY-CBCACONH (Hiller et al. 2008; Serrano et al. 2012). Figure 1c shows that about 80 % of the $^{1}H^{N}$, ^{15}N, $^{13}C^{\alpha}$, $^{1}H^{\alpha}$ and $^{13}C^{\beta}$ signals were correctly assigned by the automated backbone assignment routine UNIO-MATCH (Volk et al. 2008). Missing assignments and three erroneous assignments were crowded in the polypeptide segments D20–M23, N43–N49, N97–H107, P156–H159 and R173–Q180. With regard to the structure characterization, it is of interest that the signal intensities of part of the backbone amide cross peaks were very weak or broadened beyond detection in the segments 97–107 and 173–180 (see the "Discussion" section). The automated UNIO-MATCH backbone assignments were interactively validated against the 3D heteronuclear-resolved [^{1}H,^{1}H]-NOESY data sets which were recorded for the subsequent collection of conformational constraints (Serrano et al. 2012). In addition to the identification and correction of the three erroneous assignments, this resulted in an extension of the backbone assignments to about 95 %. The remaining missing assignments include that

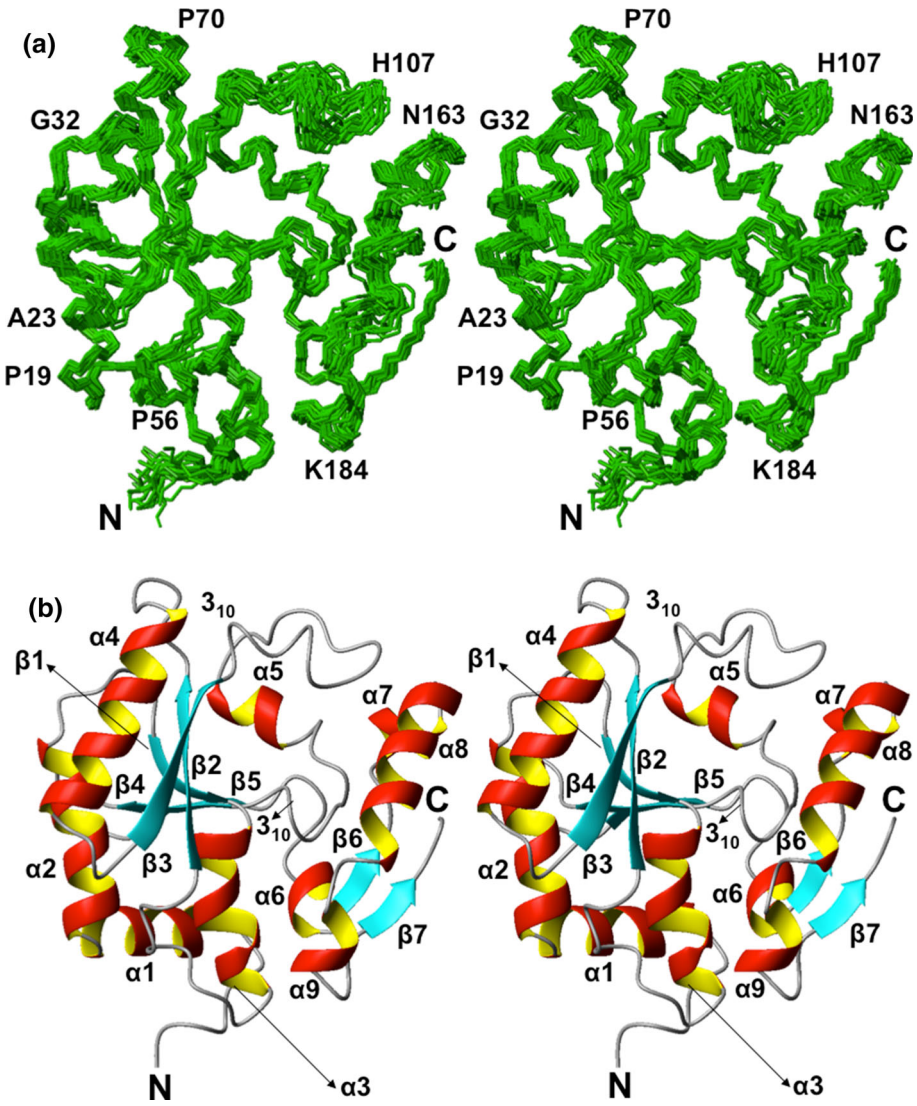

Fig. 2 NMR structure of NP_344798.1. **a** Stereo view of a bundle of 20 NMR conformers representing the NMR structure. The chain ends and some sequence positions are identified to guide the eye. **b** *Stereo ribbon* representation of the conformer closest to the mean coordinates of the bundle shown in a. β-strands are *cyan*, helices are *red/yellow*, and segments with non-regular secondary structure are *grey*. The chain ends and the regular secondary structures are identified. The figure was prepared using the program MOLMOL (Koradi et al. 1996)

only the chemical shifts of $^{13}C^{\alpha}$, $^{1}H^{\alpha}$ and $^{13}C^{\beta}$ were obtained for residues M2 and H59, and that H102, P156, P158, R173 and L174 remained unassigned (Fig. 1c).

The near-complete polypeptide backbone assignments (Fig. 1c) provided the foundation for the structure determination with the J-UNIO protocol. The automated UNIO-ANTOS/ASCAN routine, which uses as additional input the aforementioned three 3D heteronuclear-resolved [^{1}H,^{1}H]-NOESY spectra (Fiorito et al. 2008), yielded about 75 % of the expected assignments. Interactive validation of this result lead to assignments in the extent of 87 %. The missing assignments, in addition to the unassigned backbone chemical shifts (Fig. 1c), include exclusively peripheral side chain atoms of Met, Arg, Lys, His and Trp. Subsequent identification of ^{1}H–^{1}H NOEs with the program UNIO-ANTOS/CANDID (Herrmann et al. 2002a, b) in combination with the torsion angle dynamics program CYANA for the structure calculations (Güntert et al. 1997) yielded a final input of 4,090 distance constraints, including 1,138 long-range constraints. With a backbone RMSD of 0.66 Å and an all-heavy-atom RMSD of 1.05 Å (see Table 1 for the complete statistics of the structure calculation), high precision of the structure determination was obtained, which compares favorably with the results of interactive structure determinations.

NMR structure of the protein NP_344798.1

Two different presentations of the structure are shown in Fig. 2, i.e., a bundle of 20 NMR conformers and a ribbon diagram of the conformer closest to the mean coordinates of the bundle. NP_344798.1 exhibits an α/β-topology with seven β-strands, seven α-helices, and two 3_{10}-helices. The arrangement of the regular secondary structures along the amino acid sequence is shown in Fig. 1c. Strands β1–β5 form a strongly twisted antiparallel β-sheet, and the two

short strands β6 and β7 form a parallel β-sheet near the C-terminus of the protein. The larger β-sheet is well shielded from the solvent by the spatial arrangement of the helices α1–α5, whereas the two-stranded β-sheet near the C-terminus is partially solvent-accessible in spite of its association with the helices α6–α9. A homology search using DALI (Holm and Rosenstrom 2010) revealed that the fold is similar to the architecture of the catalytic head domain of class II CCA-adding enzymes (DALI Z-score > 9).

Discussion

The structure determination of NP_344798.1 expands the structural coverage of the genomic protein universe to a Pfam family with presently 786 members. Close similarities of the molecular architecture with the catalytic head domain of class II CCA-adding enzymes (Toh et al. 2009) provides a lead for functional studies, which has been followed up by NMR studies of substrate binding (to be published). Based on comparison with CCA-adding enzymes, there are indications that the two less precisely defined polypeptide segments of residues 97–107 and 173–180 are at or near the substrate-binding site (Toh et al. 2009). The indication of decreased structural order in these two segments motivated a detailed study of the structural dynamics in this molecular area and its possible functional role (to be published).

A recent determination of the structure of a 200-residue β-barrel protein with an integrative approach, "resolution-adapted structural recombination (RASREC) Rosetta", was considered to be a major technical advance (Lloyd and Wuttke 2014). This structure determination was based on the preparation of several differently isotope-labeled preparations of the protein and a large number of different NMR measurements (Sgourakis et al. 2014). In this context, it is remarkable that the present high-quality structure of a 191-residue α/β-protein was determined with the J-UNIO protocol, which uses a single, uniformly ^{13}C,^{15}N-labeled protein preparation and a total number of 7 NMR experiments, which were recorded with <2 weeks of instrument time.

Acknowledgments This work was funded by the Joint Center for Structural Genomics (JCSG) through the NIH Protein Structure Initiative (PSI) Grant Number U54 GM094586 from the National Institute of General Medical Sciences (www.nigms.nih.gov). BM received support from the Skaggs Institute of Chemical Biology. Kurt Wüthrich is the Cecil H. and Ida M. Green Professor of Structural Biology at TSRI. BM thanks Dr. Reto Horst for assistance in optimizing the setup of NMR experiments.

References

DeMarco A, Wüthrich K (1976) Digital filtering with a sinusoidal window function: an alternative technique for resolution enhancement in FT NMR. J Magn Reson 24:201–204

Fiorito F, Herrmann T, Damberger FF, Wüthrich K (2008) Automated amino acid side-chain NMR assignment of proteins using ^{13}C- and ^{15}N-resolved 3D [^{1}H,^{1}H]-NOESY. J Biomol NMR 42:23–33

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

Herrmann T, Güntert P, Wüthrich K (2002a) Protein NMR structure determination with automated NOE-identification in the NOESY spectra using the new software ATNOS. J Biomol NMR 24:171–189

Herrmann T, Güntert P, Wüthrich K (2002b) Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 319:209–227

Hiller S, Fiorito F, Wüthrich K, Wider G (2005) Automated projection spectroscopy (APSY). Proc Natl Acad Sci USA 102:10876–10881

Hiller S, Wider G, Wüthrich K (2008) APSY-NMR with proteins: practical aspects and backbone assignment. J Biomol NMR 42:179–195

Holm L, Rosenstrom P (2010) Dali server: conservation mapping in 3D. Nucleic Acids Res 38:W545–W549

Keller R (2004) CARA: computer aided resonance assignment. http://cara.nmr.ch/

Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 14:51–55

Koradi R, Billeter M, Güntert P (2000) Point-centered domain decomposition for parallel molecular dynamics simulation. Comput Phys Commun 124:139–147

Kuchta K, Knizewski L, Wyrwicz LS, Rychlewski L, Ginalski K (2009) Comprehensive classification of nucleotidyltransferase fold proteins: identification of novel families and their representatives in human. Nucleic Acids Res 37:7701–7714

Laskowski RA, Macarthur MW, Moss DS, Thornton JM (1993) PROCHECK—a program to check the stereochemical quality of protein structures. J Appl Crystallogr 26:283–291

Lloyd NR, Wuttke DS (2014) Less is more: structures of difficult targets with minimal constrains. Structure 22:1223–1224

Luginbühl P, Güntert P, Billeter M, Wüthrich K (1996) The new program OPAL for molecular dynamics simulations and energy refinements of biological macromolecules. J Biomol NMR 8:136–146

Pedrini B, Serrano P, Mohanty B, Geralt M, Wüthrich K (2013) NMR-Profiles of protein solutions. Biopolymers 99:825–831

Serrano P, Pedrini B, Mohanty B, Geralt M, Herrmann T, Wüthrich K (2012) The J-UNIO protocol for automated protein structure determination by NMR in solution. J Biomol NMR 53:341–354

Sgourakis NG, Natajaran K, Ying J, Vögeli B, Boyd LF, Margulies DH, Bax A (2014) The structure of mouse cytomegalovirus m04 protein obtained from sparse NMR data reveals a conserved fold of the m02–m06 viral immune modulator family. Structure 22:1263–1273

Toh Y, Takeshita D, Numata T, Fukai S, Nureki O, Tomita K (2009) Mechanism for the definition of elongation and termination by the class II CCA-adding enzyme. EMBO J 28:3353–3365

Volk J, Herrmann T, Wüthrich K (2008) Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. J Biomol NMR 41:127–138