Structural representative of the protein family PF14466 has a new fold and establishes links with the C2 and PLAT domains from the widely distant Pfams PF00168 and PF01477

Pedro Serrano,^{1,2} Michael Geralt,^{1,2} Biswaranjan Mohanty,^{1,2,3} and Kurt Wüthrich^{1,2,3*}

¹Joint Center for Structural Genomics (<http://www.jcsg.org>), La Jolla, California

²Department of Integrative Structural and Computational Biology and

³The Skaggs Institute for Chemical Biology, The Scripps Research Institute, La Jolla 92037, California

Received 12 March 2013; Revised 29 April 2013; Accepted 1 May 2013

DOI: 10.1002/pro.2284

Published online 16 May 2013 proteinscience.org

Abstract: The domain of unknown function (DUF) YP_001302112.1, a protein secreted by the human intestinal microbota, has been determined by NMR and represents the first structure for the Pfam PF14466. Its NMR structure is classified as a new fold, which, nonetheless, shows limited similarities with representatives of the PLAT/LH2 domains from PF01477 and the C2 domains from PF00168, both of which bind Ca^{2+} for their physiological functions. Further experiments revealed affinity of YP_001302112.1 for Ca^{2+} , and the NMR structure in the presence of CaCl_2 was better defined than that of the apo-protein. Overall, these NMR structures establish a new connection between structural representatives from two widely different Pfams that include the calcium-binding domain of a sialidase from *Vibrio cholerae* and the α -toxin from *Clostridium perfringens*, whereby these two proteins have only 7% sequence identity. Furthermore, it provides information toward the functional annotation of YP_001302112.1, based on its capacity to bind Ca^{2+} , and thus adds to the structural and functional coverage of the protein sequence universe. © 2013 The Protein Society

Keywords: functional annotation of a DUF; calcium-binding protein; structural coverage of the protein sequence universe; human gut microbiome

Introduction

The protein YP_001302112.1 is encoded by the genome of *Parabacteroides distasonis* in the mammalian intestinal flora. It was targeted for structure determination by the Joint Center for Structural Genomics (JCSG) as the first structural representative of the Pfam PF14466 and, therefore, add to the

structural coverage of the protein sequence universe.^{1,2} The NMR structure of apo-YP_001302112.1 was found to share features with two well-established protein architectures in the SCOP database, i.e., C2 (calcium/lipid-binding domain) of the Pfam PF00168 and PLAT/LH2 (lipase/lipoxygenase domain) of the Pfam PF01477. The C2 and PLAT/LH2 domains bind Ca^{2+} in their functions as part of proteins interacting with membranes.^{3,4} Comparison of NMR spectra of YP_001302112.1 recorded after stepwise addition of CaCl_2 to the protein solution showed that this protein also binds Ca^{2+} . We therefore proceeded to determine its NMR structure in

Grant sponsor: NIH Protein Structure Initiative, National Institute of General Medical Sciences (www.nigms.nih.gov); Grant numbers: U54 GM074898 and U54 GM94586.

*Correspondence to: Kurt Wüthrich. E-mail: wuthrich@scripps.edu

Figure 1. NMR structure of apo-YP_001302112.1. (a) Amino acid sequence. Regular secondary structures in the apo-protein and its Ca^{2+} -complex are indicated in blue and gray, respectively, with arrows indicating β -strands and squares representing 3_{10} -helical turns as determined with MOLMOL. Green dots above the sequence identify residues for which the backbone ^{15}N - ^1H correlation NMR signal was observed only in the Ca^{2+} -complex. (b) Bundle of 20 energy-minimized CYANA conformers superimposed for best fit of residues 9–78, 91–112, and 125–129 (Table I). The complete polypeptide backbone with residues 1–129 is shown. (c) Ribbon presentation of the conformer closest to the mean coordinates of the Ca^{2+} -bound protein, with identification of the regular secondary structures. The structure of the Ca^{2+} -complex is shown because the secondary structural elements are somewhat better defined than in the apo-protein (see panel a). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

the presence of a saturating concentration of CaCl_2 . These data provided initial leads toward a functional annotation of YP_001302112.1 and novel insights into relations between the aforementioned Pfam families, which are very distant at the level of the amino acid sequences.

Results

Selection of YP_001302112.1 for NMR structure determination

To explore remote areas of the genomic protein universe, the JCSG targets proteins that are either members of as yet structurally uncharacterized Pfam families or not yet classified into a Pfam.⁵

YP_001302112.1 [Fig. 1(a)] belongs to an uncharacterized Pfam, and it is of interest also for the JCSG biological theme of investigating proteins secreted by organisms colonizing the human gut.⁶ After unsuccessful crystallization attempts by the JCSG crystallography pipeline,⁷ a microscale analysis of ^{15}N -labeled apo-YP_001302112.1 was performed to assess the feasibility of NMR structure determination on the basis of NMR profiles recorded with a microcoil probehead.⁸ These exploratory experiments showed that an incomplete set of backbone ^{15}N - ^1H correlation signals was observable in the temperature range 283 to 308 K [Fig. 1(a)] and that the NMR spectra were improved at the higher temperatures. The NMR structure determination was therefore performed at 308 K.

Table I. Input for the Structure Calculation and Characterization of the Ensemble of 20 Energy-Minimized CYANA Conformers Used to Represent the NMR Structures of Apo-YP_001302112.1 and Its Ca²⁺-Complex

Quantity	Value ^a	
	Apo-form (308 K)	Ca ²⁺ -complex (298 K)
NOE upper distance limits	1896	1840
Intraresidual	524	500
Short range	599	552
Medium range	250	183
Long range	523	605
Dihedral angle constraints	551	528
Residual target function value (Å ²)	1.79 ± 0.24	1.68 ± 0.26
Residual NOE violations		
Number ≥ 0.1 Å	4 ± 2	12 ± 4
Maximum (Å)	0.44 ± 0.18	0.11 ± 0.08
Residual dihedral angle violations		
Number ≥ 2.5°	0 ± 1	1 ± 1
Maximum (°)	4.0 ± 1.3	3.0 ± 0.5
Amber energies (kcal/mol)		
Total	−4441 ± 109	−4607 ± 152
Van der Waals	−300 ± 15	−322 ± 36
Electrostatic	−5122 ± 104	−5238 ± 176
RMSD from ideal geometry		
Bond lengths (Å)	0.0079 ± 0.0002	0.0078 ± 0.0002
Bond angles (°)	2.03 ± 0.04	1.91 ± 0.05
RMSD to the mean coordinates (Å) ^b		
bb (10–79, 92–112, 125–129)	0.69 ± 0.16	0.59 ± 0.11
ha (10–79, 92–112, 125–129)	1.61 ± 0.15	1.10 ± 0.12
Ramachandran plot statistics (%) ^c		
Most favored regions	70.7	71.7
Additional allowed regions	27.3	27.6
Generously allowed regions	1.4	0.5
Disallowed regions	0.6	0.2

^a Except for the top six entries, which describe the input generated in the final cycle of the ATNOS/CANDID/CYANA calculation, the entries refer to the 20 best CYANA conformers after energy minimization with OPALp (see text). Where applicable, the average value for the bundle of 20 conformers and the standard deviation are given.

^b bb indicates the backbone atoms N, C^α, and C'; ha stands for all heavy atoms. Numbers in parentheses indicate the residues for which the RMSD was calculated.

^c As determined by PROCHECK.²²

NMR structure of apo-YP_001302112.1

The structure determination was performed in 20 mM phosphate buffer at pH 6.0 containing 50 mM NaCl and 4.5 mM NaN₃, using the automated J-UNIO protocol.⁸ Details are described in Materials and Methods. Statistical data on the structure determination are listed in Table I and the bundle of 20 energy-minimized conformers is shown in Figure 1b.

The structure of apo-YP_001302112.1 can be approached by considering that there are two antiparallel β-sheets, which face each other in a sandwich-like fashion. One β-sheet consists of the strands β2', β2'', β4, β1, β6, and β7, and the other one of the short strands β3, β2, β5, and β8 [Figs. 1(c) and 2(b)].

NMR structure of apo-YP_001302112.1 represents a new fold

The NMR structure of apo-YP_001302112.1 was used as input for the algorithms DALI⁹ and FATCAT¹⁰ for two independent searches of similar structures in the Protein Data Bank (PDB). DALI identified similarity

with proteins from the PLAT/LH2 domain family, which contain an eight-stranded twisted β-barrel, as illustrated by the α-toxin from *Clostridium perfringens* [DALI z-score 4.7, Fig. 2(a)].¹¹ PLAT/LH2 domains are found in lipid- and membrane-associated proteins, where protein–lipid interactions are often mediated by Ca²⁺.⁴ The relative arrangement of the two β-sheets of YP_001302112.1 in a sandwich-like structure has only limited similarity to the twisted barrel observed in the α-toxin of *C. perfringens*.

FATCAT identified a number of proteins exhibiting a C2 fold, such as the calcium-binding domain of a sialidase from *Vibrio cholerae* [FATCAT P-value 3.41e-02, Fig. 2(c)].¹² Similar to PLAT/LH2 domains, C2 domains have been reported to undergo Ca²⁺-mediated interactions with cell membranes and lipids.³ YP_001302112.1 and the sialidase calcium-binding domain both exhibit two antiparallel β-sheets that face each other in a sandwich-type topology. However, while each of the two β-sheets is formed by four sequentially consecutive β-strands in C2, the β-strands are highly entangled in YP_001302112.1 [Fig. 2(b,c)].

Figure 2. Comparison of the NMR structure of Ca^{2+} -bound YP_001302112.1 with structural representatives of the PLAT/LH2 and C2 domain families. Ribbon representations, topology diagrams, and electrostatic potential surfaces are shown. Ca^{2+} atoms are represented as gray spheres. In the electrostatic potential surface rendering, positively and negatively charged residues are colored blue and red, respectively. (a) Crystal structure of the Ca^{2+} -complex of the C-terminal domain of the *Clostridium perfringens* α -toxin (PDB_id 1QMD). (b) Conformer closest to the mean coordinates of Apo-YP_001302112.1. The structure was rotated by 180° about a vertical axis relative to the orientation in Figure 1. (c) Crystal structure of the Ca^{2+} -binding domain of *Vibrio cholerae* sialidase (PDB_id 1W00). [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Overall, the limited similarity with PDB structures thus identified either by DALI or by FATCAT as well as the observation that the two algorithms identified different “closest structures,” suggested that YP_001302112.1 could represent a new fold. The assignment of a new fold was subsequently confirmed by the SCOP database (Dr. Alexey Murzin, personal communication).

YP_001302112.1 is a Ca^{2+} -binding protein

The two proteins from two widely distant Pfams identified as showing limited similarity to our protein by the algorithms FATCAT and DALI, respectively, both

have similar functions that are related to Ca^{2+} binding. For YP_001302112.1, the effect of adding CaCl_2 to the protein solution is shown in the superposition of the ^{15}N , ^1H -HSQC spectra of the apo-form and the Ca^{2+} -bound form [Fig. 3(a)]. A discrete set of resonances experience significant chemical shift differences, $\Delta\delta$, between the two forms, and 11 cross-peaks observed in the ^{15}N , ^1H -HSQC spectrum recorded in the presence of Ca^{2+} were broadened beyond detection in the apo-protein. Both the resonances with $\Delta\delta > 0.12$ ppm and the missing peaks in the apo-form are clustered in the loops in the lower right of the presentation in Figure 3(b). This location corresponds to the

Figure 3. NMR studies of Ca^{2+} -binding to YP_001302112.1. (a) Superposition of the 2D ^{15}N , ^1H -HSQC spectra of the apo-form (red) and the Ca^{2+} -bound form (black) of YP_001302112.1. Residues that experience $\Delta\delta > 0.12$ ppm chemical shifts upon addition of CaCl_2 are indicated in blue, and those that were observed only in Ca^{2+} -bound YP_001302112.1 are green. (b) Stereo-view of the conformer closest to the mean coordinates in the bundle of 20 NMR conformers representing the structure of Ca^{2+} -bound YP_001302112.1. Polypeptide segments of residues, which were only observed in the Ca^{2+} -complex are colored green, and those with chemical shift changes upon Ca^{2+} -binding of $\Delta\delta > 0.12$ ppm are blue. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Ca^{2+} -binding sites of the *C. perfringens* α -toxin and the *V. cholerae* sialidase [Fig. 2(a,c)].

NMR structure of Ca^{2+} -bound YP_001302112.1

The NMR structure of the Ca^{2+} -bound form of YP_001302112.1 was determined in the same buffer solution as the apo-form, i.e., 20 mM phosphate at pH = 6.0, 50 mM NaCl, 4.5 mM NaN_3 , after addition of 5 mM CaCl_2 (Table I, Fig. 4). Upon CaCl_2 addition, the stability with respect to irreversible thermal denaturation is increased from about 318 K to 350 K. Over the entire temperature range 283 to 308 K, the Ca^{2+} -complex yields higher-quality NMR spectra than the apo-form, with observation of a complete set of backbone ^{15}N - ^1H correlation signals. We therefore determined the structure at 298 K.

Near identity of the structures of apo-YP_001302112 and its Ca^{2+} -complex is apparent from comparison of the Figures 1 and 4. Differences are primarily located in the loops with residues 79–90 and 113–123, which are less well defined in the apo-form due to the absence of ^{15}N - ^1H correlation peaks for the amino acid residues 87 to 91, 93 to 94, and 115 to 118 [Fig. 1(a)]. In the absence of Ca^{2+} , these signals appear to be broadened beyond detection by slow conformational exchange of the polypeptide segments near the metal binding site.

Discussion

The present work is yet another illustration of the important role that structural genomics has assumed by extending the structural coverage of the

Figure 4. NMR structure of Ca^{2+} -YP_001302112.1. (a) Stereo-view of a bundle of 20 energy-minimized conformers superimposed for best fit of residues 9–78, 91–112, and 125–129 (Table I). Same orientation as the apo-protein structure in Figure 1. (b) Stereo-view of the all-heavy-atom presentation of the conformer closest to the mean coordinates. The backbone is represented in gray, and the amino acid side chains are color coded according to their displacements, with $D \leq 0.4$ Å green, 0.4 Å $< D \leq 0.8$ Å blue, and $D > 0.8$ Å red. [Color figure can be viewed in the online issue, which is available at [wileyonlinelibrary.com](http://www.interscience.wiley.com).]

genomic protein universe.^{1,2} Although searching the PDB with either DALI or FATCAT did not establish any relations between structural representatives of the widely distant Pfams PF00168 and PF01477, a C2 domain as well as a PLAT/LH domain were found to have limited structural similarity with the newly determined structure of YP_001302112.1, and a novel relation between the PF00168 and PF01477 families was thus established. The database search further resulted in a tentative functional annotation of YP_001302112.1 as a Ca^{2+} -binding protein interacting with lipids and/or membranes.

For future reference, it is noteworthy that a tentative functional annotation of the presently studied DUF was obtained from comparison with PDB structures, although it was subsequently discovered that its structure represents a new fold. The limited structural similarities with the well-characterized C2 and PLAT/LH2 domains were sufficient for obtaining this information. Comparison of the structures observed in the absence (Fig. 1) and presence (Fig. 4) of Ca^{2+} illustrates the unique ability of solution NMR to compare structures of the apo-form and ligand-bound forms of proteins in solution. The dynamic disorder observed near the metal-binding site of apo-YP_001302112.1

also provides a possible rationale for the failure to obtain diffracting crystals in the JCSG pipeline.⁷

Materials and Methods

Protein purification and NMR sample preparation

The pSpeedET-YP_001302112.1 vector for production of the uniformly ¹³C,¹⁵N-labeled protein was obtained from the JCSG Crystallomics Core based on *P. distasonis* genomic DNA originally obtained from the American Type Culture Collection. Expression was carried out by growing freshly transformed BL21(DE3) cells in M9 minimal medium containing 1 g/L ¹⁵NH₄Cl and 4 g/L [¹³C₆]-D-glucose as the sole nitrogen and carbon sources. Cell cultures were grown at 37°C with vigorous shaking to an optical density at 600 nm of 0.6, and the temperature was then reduced to 18°C. After induction with 1 mM isopropyl-β-D-1-thiogalactopyranoside, the cells were grown for 20 h, harvested by centrifugation, resuspended in buffer A (20 mM sodium phosphate at pH 7.5, 200 mM NaCl, 10 mM imidazole) supplemented with complete EDTA-free protease inhibitor cocktail tablets (Roche) and lysed by sonication. The cell debris was removed by centrifugation (20,000g for 30 min) and the supernatant loaded onto a Ni²⁺ affinity column (HisTrap HP; GE Healthcare) equilibrated with buffer A. The imidazole concentration was increased first to 30 mM in order to remove nonspecifically bound proteins and then to 500 mM to elute YP_001302112.1. After digestion with 0.08 mg/mL TEV protease overnight at room temperature, the sample was loaded onto a desalting column (HiPrep 26/10; GE Healthcare) and eluted with buffer A. The protein fractions were again loaded onto a Ni²⁺ affinity column (HisTrap HP; GE Healthcare) equilibrated with buffer A, in order to remove the Histagged TEV protease and the cleaved Histag. Fractions containing YP_001302112.1, as determined by SDS-PAGE, were pooled and further purified using a size exclusion column (Superdex 75 Hiload 26/60; GE Healthcare) equilibrated with buffer B (20 mM sodium phosphate at pH 6.0, 50 mM NaCl) and eluted with the same buffer. The fractions containing YP_001302112.1 were concentrated to 550 μL for a final protein concentration of 1.5 mM, using 3-kDa cutoff centrifugal filter devices (Millipore). The NMR sample was supplemented with 5% ²H₂O (vol/vol) and 4.5 mM NaN₃.

NMR spectroscopy

NMR data used for the structure determinations were recorded on Bruker AVANCE 600 MHz and AVANCE 800 MHz spectrometers equipped with 5 mm TXI-HCN z- or xyz-gradient probes. A 2D [¹⁵N,¹H]-HSQC spectrum, the experiments 4D APSY-HACANH, 5D APSY-HACACONH, and 5D

APSY-CBCACONH^{13,14} used for the polypeptide backbone assignments, and a 2D ¹⁵N{¹H}-NOE experiment¹⁵ were recorded at 600 MHz. 3D ¹⁵N-resolved, 3D ¹³C(aliphatic)-resolved, and 3D ¹³C(aromatic)-resolved [¹H,¹H]-NOESY experiments were recorded with a mixing time of $\tau_m = 65$ ms at 800 MHz. The 2D [¹⁵N,¹H]-HSQC spectra used to study binding of YP_001302112.1 to Ca²⁺ were recorded on a BRUKER DRX 700 MHz spectrometer equipped with a 1.7-mm microcoil probehead. Proton chemical shifts were referenced to internal 2,2-dimethyl-2-silapentane-5-sulfonic acid sodium salt (DSS). The ¹³C and ¹⁵N chemical shifts were referenced indirectly to DSS, using the absolute frequency ratios.¹⁶

NMR structure determinations

For the polypeptide backbone resonance assignment, we made use of a standard set of three APSY-NMR experiments (see above),^{13,14} which provided the input for the automated assignment software UNIO-MATCH.¹⁷ UNIO-MATCH yielded 85% and 96% of the backbone assignments for the apo-form and the Ca²⁺-bound form of YP_001302112, respectively. These backbone assignments were then interactively validated and extended with the aid of the ¹⁵N-resolved [¹H,¹H]-NOESY spectrum to 88% and 98% completeness, respectively. Side chain resonance assignments were obtained with the software UNIO-ATNOS/ASCAN,^{18,19} using input from the same three 3D heteronuclear-resolved [¹H,¹H]-NOESY data sets (see above) that subsequently provided the NOE distance constraints for the structure calculation. The UNIO-ATNOS/ASCAN^{18,19} outputs were interactively checked and extended to achieve assignments for 87% and 94%, respectively, of the side-chain hydrogen atoms in the apo- and Ca²⁺-bound forms of the protein. The lower assignment completeness for the apo-protein is primarily due to the fact that the amide moiety signals of 11 residues were broadened beyond detection [Fig. 1(a)] at the solution conditions used for the structure determination.

The automated seven-cycle ATNOS/CANDID/CYANA protocol^{18,20,21} implemented in the UNIO software package, which affords combined NOE assignment and protein structure calculation, was used for the NMR structure determination. This procedure yielded 1896 and 1840 upper limit distance constraints, respectively, as input for the final cycle of structure calculation (Table I). The larger number of long-range NOEs observed for the Ca²⁺-complex can be rationalized by more complete assignment of the backbone amide protons (see above). For shorter-range NOEs, the larger numbers observed in the apo-protein are a consequence of the higher temperature (Table I) and the fact that nearly complete side chain assignments were also obtained for the residues with missing backbone amide signals [Fig. 1(a)].

The quality of the molecular structures was validated based on in-house JCSG NMR Core criteria,⁸ in addition to the JCSG structure validation server (<http://smb.slac.stanford.edu/jcsg/QC/>), the psvs server (http://psus-1_4-dev.nesg.org), and the PDB validation server (<http://deposit.pdb.org/validate>).

Accession numbers

The chemical shifts of the apo-form and the Ca²⁺-bound form of YP_001302112.1 have been deposited in the Biological Magnetic Resonance Bank (<http://www.bmrb.wisc.edu/>) with the accession codes 17799 and 17806. The atomic coordinates of the bundles of 20 conformers used to represent the NMR structures of the apo-protein and the Ca²⁺-complex have been deposited to the PDB (<http://www.rcsb.org/pdb/>) under the accession codes 2LG7 and 2LGE, respectively.

Acknowledgments

Kurt Wüthrich is the Cecil H. and Ida M. Green Professor of Structural Biology at The Scripps Research Institute. We thank the members of the JCSG high-throughput structural biology pipeline for their contribution to this work, in particular Lukasz Jaroszewski for advice on Pfams and Marc-André Elsliger and Ian Wilson for helpful comments on the article. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

References

- Jaroszewski L, Li Z, Krishna SS, Bakolitsa C, Wooley J, Deacon AM, Wilson IA, Godzik A (2012) Exploration of uncharted regions of the protein universe. *PloS Biol* 7:e1000205.
- Dessailly BH, Nair R, Jaroszewski L, Fajardo JE, Kouranov A, Lee D, Fiser A, Godzik A, Rost B, Orengo C (2009) PSI-2: structural genomics to cover protein domain family space. *Structure* 17:869–881.
- Rizo J, Südhof TC (1998) C2-domains, structure and function of a universal Ca²⁺-binding domain. *J Biochem* 273:15879–15882.
- Harshman S, Sugg N (1985) Effect of calcium ions on *Staphylococcal* alpha-toxin-induced hemolysis of rabbit erythrocytes. *Infect Immun* 47:37–40.
- Finn RD, Mistry J, Tate J, Coghill P, Heger A, Pollington JE, Gavin OL, Gunasekaran P, Ceric G, Forslund K, Holm L, Sonnhammer ELL, Eddy SR, Bateman A (2010) The Pfam protein families database. *Nucl Acids Res* 38:D211–D222.
- Ellrott K, Jaroszewski L, Li W, Wooley JC, Godzik A (2010) Expansion of the protein repertoire in newly explored environments: human gut microbiome specific protein families. *PLoS Comput Biol* 6:e1000798.
- Lesley SA, Kuhn P, Godzik A, Deacon AM, Mathews I, Kreusch A, Spraggon G, Klock HE, McMullan D, Shin T, Vincent J, Robb A, Brinen LS, Miller MD, McPhillips TM, Miller MA, Scheibe D, Canaves JM, Guda C, Jaroszewski L, Selby TL, Elsliger MA, Wooley J, Taylor SS, Hodgson KO, Wilson IA, Schultz PG, Stevens RC (2002) Structural genomics of the *Thermotoga maritima* proteome implemented in a high-throughput structure determination pipeline. *Proc Natl Acad Sci USA* 99:11664–11669.
- 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.
- Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol* 233:123–138.
- Yuzhen Y, Godzik A (2003) Flexible structure alignment by chaining aligned fragment pairs following twists. *Bioinformatics* 19:246–255.
- Naylor C, Jepson M, Crane D, Titball R, Miller J, Basak A, Bolgiano B (1999) Characterisation of the calcium-binding C-terminal domain of *Clostridium perfringens* α-toxin. *J Mol Biol* 294:757–770.
- Moustafa I, Connaris H, Taylor M, Zaitsev V, Wilson J, Kiefel M, von Itzstein M, Taylor G (2004) Sialic acid recognition by *Vibrio cholerae* neuraminidase. *J Biol Chem* 279:40819–40824.
- 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.
- Zhu G, Youlin X, Nicholson LK, Sze KH (2000) Protein dynamics measurements by TROSY-based NMR experiments. *J Magn Reson* 143:423–426.
- Wishart D, Bigam C, Yao J, Abildgaard F, Dyson H, Oldfield E, Markley J, Sykes B (1995) ¹H, ¹³C and ¹⁵N chemical shift referencing in biomolecular NMR. *J Biomol NMR* 6:135–140.
- Volk J, Herrmann T, Wüthrich K (2008) Automated sequence-specific protein NMR assignment using the memetic algorithm MATCH. *J Biomol NMR* 41:127–138.
- Herrmann T, Güntert P, Wüthrich K (2002) Protein NMR structure determination with automated NOE identification in the NOESY spectra using the new software ATNOS. *J Biomol NMR* 24:171–189.
- 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.
- Herrmann T, Güntert P, Wüthrich K (2002) 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.
- 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.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM (1993) PROCHECK—a program to check the stereochemical quality of protein structures. *J App Cryst* 26:283–291.