Published in final edited form as:

J Struct Funct Genomics. 2013 December ; 14(4): . doi:10.1007/s10969-013-9164-8.

Solution NMR structure of CD1104B from pathogenic *Clostridium difficile* reveals a distinct α -helical architecture and provides first structural representative of protein domain family PF14203

Surya VSRK Pulavarti,

Department of Chemistry, The State University of New York at Buffalo, and Northeast Structural Genomics Consortium, Buffalo, NY 14260, USA

Alexander Eletsy,

Department of Chemistry, The State University of New York at Buffalo, and Northeast Structural Genomics Consortium, Buffalo, NY 14260, USA

Hsiau-Wei Lee,

Complex Carbohydrate Research Center, University at Georgia, and Northeast Structural Genomics Consortium, Athens, GA 30602, USA

Thomas B. Acton,

Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey and Northeast Structural Genomics Consortium, Piscataway, NJ 08854, USA

Rong Xiao,

Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey and Northeast Structural Genomics Consortium, Piscataway, NJ 08854, USA

John K. Everett,

Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey and Northeast Structural Genomics Consortium, Piscataway, NJ 08854, USA

James H. Prestegard,

Complex Carbohydrate Research Center, University at Georgia, and Northeast Structural Genomics Consortium, Athens, GA 30602, USA

Gaetano T. Montelione, and

Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers, The State University of New Jersey and Northeast Structural Genomics Consortium, Piscataway, NJ 08854, USA, Department of Biochemistry and Molecular Biology, Robert Wood Johnson Medical School, UMDNJ, Piscataway NJ 08854, USA

Thomas Szyperski

Department of Chemistry, The State University of New York at Buffalo, and Northeast Structural Genomics Consortium, Buffalo, NY 14260, USA

© Springer Science+Business Media B.V. 2012

Correspondence to: Thomas Szyperski, szypersk@buffalo.edu.

Hsiau-Wei Lee

Present address: Department of Chemistry and Biochemistry, University of California-Santa Cruz, Santa Cruz, CA 95064, USA

Thomas Szyperski: szypersk@buffalo.edu

Abstract

A high-quality structure of the 68-residue protein CD1104B from *Clostridium difficile* strain 630 exhibits a distinct all α -helical fold. The structure presented here is the first representative of bacterial protein domain family PF14203 (currently 180 members) of unknown function (DUF4319) and reveals that the side-chains of the only two strictly conserved residues (Glu 8 and Lys 48) form a salt bridge. Moreover, these two residues are located in the vicinity of the largest surface cleft which is predicted to contribute to a surface area involved in protein-protein interactions. This, along with its coding in transposon CTn4, suggests that CD1104B (and very likely all members of Pfam 14230) functions by interacting with other proteins required for the transfer of transposons between different bacterial species.

Keywords

CD1104B; PF14203; DUF4319; Transposon; Structural Genomics

Introduction

Clostridium difficile is a gram positive, spore forming anaerobic bacterium causing diseases such as antibiotic-associated diarrhea and lethal pseudomembranous colitis [1]. The genome of *C. difficile* is known to contain mobile genetic elements called ‘conjugative transposons’, which are capable of integrating in/excising from genomes. The conjugal transfer between bacterial genomes is mediated by proteins encoded in transposon operons [1]. The transposons can encode proteins conveying, antibiotic resistance: In the case of *C. difficile*, the genome contains transposons which provide tetracycline and erythromycin resistance [1]. Furthermore, protein CD1104B is part of the putative conjugative transposon CTn4 [2, 3] which is predicted to comprise genes of ABC transporters. Those aid the export of anti-microbial peptides which facilitate intestinal colonization by *C. difficile* [2].

The 68-residue protein CD1104B from *Clostridium difficile* strain 630 (UniProtKB/TrEMBL accession number Q18AW3; NESG ID Cfr130) belongs to Pfam [4] protein family PF14203 (Fig. 1a; Fig. S4) and is of unknown function (DUF4319). This family currently contains 180 bacterial sequences predominantly from class clostridia, erysielotrichi and bacilli. CD1104B does not exhibit significant sequence similarity to any protein with known atomic resolution structure. CD1104B was thus selected as a target of the Protein Structure Initiative and assigned to the Northeast Structural Genomics Consortium (NESG; <http://www.nesg.org>; target ID: Cfr130) as part of a cooperative inter-center effort aimed at providing structural coverage of novel proteins discovered by metagenomic sequencing of the human gut microbiome [5]. Since no atomic resolution structures were available when this target was chosen, high modeling leverage [6] could be anticipated. Here, we report the high-quality solution NMR structure of CD1104B which represents the first atomic resolution structures for protein domain family PF14203.

Material and Methods

CD1104B was cloned, expressed, and purified following protocols [7–9] established by the NESG (see Supplementary Material and <http://www.nmr2.buffalo.edu/nesc.wiki> for details). The protein included a C-terminal hexaHis tag (LEHHHHHH) to facilitate protein purification. The corresponding pET expression vector (Cfr130-21.6) has been deposited in the PSI Materials Repository (<http://psimr.asu.edu/>). The sample for NMR structure determination contained [U - ^{13}C , ^{15}N]-labeled CD1104B (concentration ~1.6 mM in 90%

H₂O/10% D₂O, 20 mM MES, 100 mM NaCl, 10 mM DTT, 5 mM CaCl₂, and 0.02% NaN₃ at pH 6.5.). An additional sample containing [5% ¹³C; U-¹⁵N]-labeled CD1104B (concentration ~1.6 mM in 90% H₂O/10% D₂O, 20 mM MES, 100 mM NaCl, 10 mM DTT, 5 mM CaCl₂, and 0.02% NaN₃ at pH 6.5.) enabled stereospecific assignment of the methyl groups of Val and Leu residues [10] and collection of residual dipolar couplings (RDCs). An isotropic overall rotational correlation time of about ~6 ns was inferred from average ¹⁵N spin relaxation times (see Supplementary Material and <http://www.nmr2.buffalo.edu/nesg.wiki> for details), indicating that the protein is monomeric in solution. This was confirmed with analytical gel-filtration and static light scattering (Supplementary Figs. S1 and S2).

NMR data were acquired at 25 C on Varian INOVA 500, 600 and 750 MHz spectrometers within ~5.5 days. Nearly complete sequence-specific ¹H, ¹⁵N and ¹³C resonance assignments (Table I; Supplementary Fig. S3) were obtained from G-matrix Fourier transform [11] and conventional triple-resonance NMR experiments (Supplementary Material) using the programs AutoAssign 2.3.0 [12, 13] followed by manual assignment of side-chain resonances. Assignments were validated using the AVS software suite [14]. Backbone ¹⁵N-¹H RDCs were measured for two alignment media (Supplementary Material). Chemical shifts, NOESY peak lists, and time domain NMR data were deposited in the BioMagResBank (accession number 17359).

Structure calculations were performed by consensus analysis [15, 16] of ¹H-¹H NOE-derived upper limit distance constraints from automated NOESY assignment programs CYANA [17, 18] and AutoStructure 2.2.1 [19]. Backbone dihedral angle constraints were derived from chemical shifts using the program TALOS+ [20] for residues located in well-defined regular structure elements, and orientational constraints were derived from polypeptide backbone ¹⁵N-¹H RDCs [21]. Stereospecific assignments of methylene protons were performed with the GLOMSA module of CYANA and the final structure calculation was performed with CYANA followed by refinement of selected conformers in an “explicit water bath” [22] using the program CNS 1.2 [23]. Validation of the resulting 20 refined conformers was performed with the Protein Structure Validation Software (PSVS) server 1.4 [24] and the agreement of structures and NOESY peak lists was verified using the AutoStructure/RPF 2.2.1[15] package.

The structural coverage provided by the CD1104B structure for PF14203 (DUF4319) was estimated using the program PSI-BLAST [25] as described in the Supplementary Material.

Results and Discussion

A high-quality solution NMR structure of CD1104B was obtained (Fig. 1b–d; Table 1). The coordinates were deposited in the Protein Data Bank [26] on 12/13/2010 (accession code 2L7K). CD1104B exhibits an α -helical fold (Fig. 1b) consisting of four α -helices: I (6–15), II (19–29), III (36–49) and IV (54–60). α -helix I is oriented perpendicular to a bundle formed from α -helices II and III, and α -helix IV is packed against the first three α -helices.

A search in the PDB [26] for structurally similar proteins using the DALI server [27] resulted in only one hit with a marginal Z-score of 4.3 (*Escherichia coli* protein YMGB; PDB ID: 2OXL, chain-B). Furthermore, visual inspection shows that α -helix IV (Fig. 1b) is not structurally aligned, indicating that CD1104B adopts a rather distinct α -helical architecture. As a result, the DALI-based search did not result in additional insight into the function of CD1104B.

Employment of the program ProFunc [28, 29] reveals the presence of two clefts ‘C1’ and ‘C2’ (C1: ~500 Å³ formed by Ile 6, Glu 7, Asn 10, Leu 11, Ser 13, Ile 14, Asp 37, Met 38,

Glu 40, Leu 41, Ala 65 and Val 68; C2: ~200 Å³ formed by Ile 27, Asn 28, Leu 31, Ala 42, Lys 43 and Leu 46), both located in spatial proximity between, respectively, α -helices 1, 3 and C-terminal loop, and between α -helices 2 and 3 (Fig. 1e). Moreover, ConSurf [30] analysis reveals higher conservation of residues around the larger cleft C1 and highlights the fact that only two residues (Glu 8 and Lys 48) are strictly conserved across the entire PF14230 family (Fig. S4). Intriguingly, our structure now reveals that these two residues are located in close spatial proximity and that their side chains form a salt bridge in 7 of the 20 conformers representing the NMR solution structure (Fig. 1f). These findings support the conjecture that Glu 8 and Lys 48, along with other conserved residues around cleft C1, are of functional importance [31].

Operon prediction using the ‘microbes-online’ server (<http://microbesonline.org>) shows that the gene encoding CD1104B belongs to an operon within transposon CTn4 which also contains the genes of a putative DNA primase, a DNA topoisomerase and conjugal transfer proteins (CD1105, CD1106 & CD1110) [3]. It may thus be that protein CD1104B interacts with either nucleic acids or with proteins required for the transfer of the transposon between genomes. Electrostatic surface potential calculations show that CD1104B exhibits on one side a largely negatively charged surface, while a mixed surface charge distribution is observed on the other side (Fig. 1h; Fig. S5). However, a search for DNA binding sites using ProFunc [32] did not result in any significant hits. In contrast, a structure based prediction for protein-protein interaction sites using the PredUs server [33, 34] suggests that a surface area (Fig. 1g) which is located around the largest cleft C1, and which overlaps with the conserved surface area identified using ConSurf [30] (Fig. 1e), is involved in such interactions. Taken together, our structure thus suggests that protein CD1104B (and very likely all members of Pfam 14230) functions by interacting with other proteins required for the transfer of transposon CTn4 between different bacterial species.

Finally, an important goal in the PSI program has been to provide novel modeling leverage [6, 35]. Hence, we estimated the number of proteins that can be reliably modeled using our newly determined experimental protein structure as a template. Based on the approach described by Liu et al. [6], the structure of CD1104B provides a structural coverage of ~75 % for PF14203.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by the National Institutes of Health, grant number: U54 GM094597 (T.S., J.H.P. and G.T.M.).

Abbreviations

ABC	ATP-binding cassette
CD	<i>Clostridium difficile</i>
CTn	Conjugative transposon
DSS	4,4-dimethyl-4-silapentane-1-sulfonate sodium salt
DTT	Dithiothreitol
DUF	Domain of unknown function

MES	2-(<i>N</i> -morpholino)ethanesulfonic acid
NESG	Northeast Structural Genomics Consortium
NOE	Nuclear Overhauser effect
PDB	Protein Data Bank
RDC	Residual dipolar coupling
RMSD	Root mean square deviation

References

1. He M, Miyajima F, Roberts P, Ellison L, Pickard DJ, Martin MJ, Connor TR, Harris SR, Fairley D, Bamford KB, D'Arc S, Brazier J, Brown D, Coia JE, Douce G, Gerding D, Kim HJ, Koh TH, Kato H, Senoh M, Louie T, Michell S, Butt E, Peacock SJ, Brown NM, Riley T, Songer G, Wilcox M, Pirmohamed M, Kuijper E, Hawkey P, Wren BW, Dougan G, Parkhill J, Lawley TD. Emergence and global spread of epidemic healthcare-associated *Clostridium difficile*. *Nat Genet.* 2013; 45:109–113. [PubMed: 23222960]
2. Brouwer MSM, Warburton PJ, Roberts AP, Mullany P, Allan E. Genetic organisation, mobility and predicted functions of genes on integrated, mobile genetic elements in sequenced strains of *Clostridium difficile*. *PLoS One.* 2011; 6:e23014. [PubMed: 21876735]
3. Sebaihia M, Wren BW, Mullany P, Fairweather NF, Minton N, Stabler R, Thomson NR, Roberts AP, Cerdano-Tarraga AM, Wang H, Holden MTG, Wright A, Churcher C, Quail MA, Baker S, Bason N, Brooks K, Chillingworth T, Cronin A, Davis P, Dowd L, Fraser A, Feltwell T, Hance Z, Holroyd S, Jagels K, Moule S, Mungall K, Price C, Rabinowitsch E, Sharp S, Simmonds M, Stevens K, Unwin L, Whithead S, Dupuy B, Dougan G, Barrell B, Parkhill J. The multidrug-resistant human pathogen *Clostridium difficile* has a highly mobile, mosaic genome. *Nat Genet.* 2006; 38:779–786. [PubMed: 16804543]
4. 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. The Pfam protein families database. *Nucleic Acids Res.* 2010; 38:D211–D222. [PubMed: 19920124]
5. Dessailly BH, Nair R, Jaroszewski L, Fajardo JE, Kouranov A, Lee D, Fiser A, Godzik A, Rost B, Orengo C. PSI-2: structural genomics to cover protein domain family space. *Structure.* 2009; 17:869–881. [PubMed: 19523904]
6. Liu JF, Montelione GT, Rost B. Novel leverage of structural genomics. *Nat Biotechnol.* 2007; 25:850–853.
7. Acton TB, Gunsalus KC, Xiao R, Ma LC, Aramini J, Baran MC, Chiang Y-W, Climent T, Cooper B, Denissova NG, Douglas SM, Everett JK, Ho CK, Macapagal D, Rajan PK, Shastry R, Shih LY, Swapna GVT, Wilson M, Wu M, Gerstein M, Inouye M, Hunt JF, Montelione GT. Robotic cloning and protein production platform of the Northeast Structural Genomics Consortium. *Methods Enzymol.* 2005; 394:210–243. [PubMed: 15808222]
8. Acton TB, Xiao R, Anderson S, Aramini J, Buchwald WA, Ciccocanti C, Conover K, Everett J, Hamilton K, Huang YJ, Janjua H, Kornhaber G, Lau J, Lee DY, Liu GH, Maglaqui M, Ma LC, Mao L, Patel D, Rossi P, Sahdev S, Shastry R, Swapna GVT, Tang YF, Tong SC, Wang DY, Wang H, Zhao L, Montelione GT. Preparation of protein samples for NMR structure, function, and small-molecule screening studies. *Fragment-Based Drug Design: Tools, Practical Approaches, and Examples. Methods Enzymol.* 2011; 493:21–60. [PubMed: 21371586]
9. Xiao R, Anderson S, Aramini J, Belote R, Buchwald WA, Ciccocanti C, Conover K, Everett JK, Hamilton K, Huang YJ, Janjua H, Jiang M, Kornhaber GJ, Lee DY, Locke JY, Ma L-C, Maglaqui M, Mitra S, Patel D, Rossi P, Sahdev S, Sharma S, Shastry R, Swapna GVT, Tong SN, Wang D, Wang H, Zhao L, Montelione GT, Acton TB. The high-throughput protein sample production platform of the Northeast Structural Genomics Consortium. *J Struct Biol.* 2010; 172:21–33. [PubMed: 20688167]

10. Neri D, Szyperski T, Otting G, Senn H, Wuthrich K. Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ¹³C labeling. *Biochemistry*. 1989; 28:7510–7516. [PubMed: 2692701]
11. Kim S, Szyperski T. GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J Am Chem Soc*. 2003; 125:1385–1393. [PubMed: 12553842]
12. Moseley HNB, Monleon D, Montelione GT. Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. *Nuclear Magnetic Resonance of Biological Macromolecules, Pt B. Methods Enzymol*. 2001; 339:91–108. [PubMed: 11462827]
13. Zimmerman DE, Kulikowski CA, Huang YP, Feng WQ, Tashiro M, Shimotakahara S, Chien CY, Powers R, Montelione GT. Automated analysis of protein NMR assignments using methods from artificial intelligence. *J Mol Biol*. 1997; 269:592–610. [PubMed: 9217263]
14. Moseley HNB, Sahota G, Montelione GT. Assignment validation software suite for the evaluation and presentation of protein resonance assignment data. *J Biomol NMR*. 2004; 28:341–355. [PubMed: 14872126]
15. Huang YJ, Powers R, Montelione GT. Protein NMR recall, precision, and F-measure scores (RPF scores): structure quality assessment measures based on information retrieval statistics. *J Am Chem Soc*. 2005; 127:1665–1674. [PubMed: 15701001]
16. Liu G, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran DK, Xiao R, Yee A, Lemak A, Bhattacharya A, Acton TA, Arrowsmith CH, Montelione GT, Szyperski T. NMR data collection and analysis protocol for high-throughput protein structure determination. *Proc Natl Acad Sci U S A*. 2005; 102:10487–10492. [PubMed: 16027363]
17. Guntert P, Mumenthaler C, Wuthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol*. 1997; 273:283–298. [PubMed: 9367762]
18. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. *J Mol Biol*. 2002; 319:209–227. [PubMed: 12051947]
19. Huang YJ, Tejero R, Powers R, Montelione GT. A topology-constrained distance network algorithm for protein structure determination from NOESY data. *Proteins*. 2006; 62:587–603. [PubMed: 16374783]
20. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR*. 1999; 13:289–302. [PubMed: 10212987]
21. Liu Y, Prestegard JH. Measurement of one and two bond N-C couplings in large proteins by TROSY-based J-modulation experiments. *J Magn Reson*. 2009; 200:109–118. [PubMed: 19581113]
22. Linge JP, Williams MA, Spronk C, Bonvin A, Nilges M. Refinement of protein structures in explicit solvent. *Proteins*. 2003; 50:496–506. [PubMed: 12557191]
23. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr Sect D Biol Crystallogr*. 1998; 54:905–921. [PubMed: 9757107]
24. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. *Proteins*. 2007; 66:778–795. [PubMed: 17186527]
25. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997; 25:3389–3402. [PubMed: 9254694]
26. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res*. 2000; 28:235–242. [PubMed: 10592235]
27. Holm L, Sander C. Dali: a network tool for protein structure comparison. *Trends Biochem Sci*. 1995; 20:478–480. [PubMed: 8578593]
28. Nayal M, Honig B. On the nature of cavities on protein surfaces: application to the identification of drug-binding sites. *Proteins*. 2006; 63:892–906. [PubMed: 16477622]

29. Petrey D, Fischer M, Honig B. Structural relationships among proteins with different global topologies and their implications for function annotation strategies. *Proc Natl Acad Sci U S A*. 2009; 106:17377–17382. [PubMed: 19805138]
30. Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. *Nucleic Acids Res*. 2010; 38:W529–W533. [PubMed: 20478830]
31. Xu D, Tsai CJ, Nussinov R. Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng*. 1997; 10:999–1012. [PubMed: 9464564]
32. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. *Nucleic Acids Res*. 2005; 33:W89–W93. [PubMed: 15980588]
33. Zhang QC, Deng L, Fisher M, Guan J, Honig B, Petrey D. PredUs: a web server for predicting protein interfaces using structural neighbors. *Nucleic Acids Res*. 2011; 39:W283–W287. [PubMed: 21609948]
34. Zhang QC, Petrey D, Norel R, Honig BH. Protein interface conservation across structure space. *Proc Natl Acad Sci U S A*. 2010; 107:10896–10901. [PubMed: 20534496]
35. Nair R, Liu J, Soong T-T, Acton TB, Everett JK, Kouranov A, Fiser A, Godzik A, Jaroszewski L, Orengo C, Montelione GT, Rost B. Structural genomics is the largest contributor of novel structural leverage. *J Struct Funct Genomics*. 2009; 10:181–191. [PubMed: 19194785]
36. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. *Proteins*. 2007; 66:778–795. [PubMed: 17186527]
37. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol Syst Biol*. 2011; 7
38. Gouet P, Robert X, Courcelle E. ESPript/ENDscript: extracting and rendering sequence and 3D information from atomic structures of proteins. *Nucleic Acids Res*. 2003; 31:3320–3323. [PubMed: 12824317]

Fig. 1. Solution NMR structure of CD1104B from *Clostridium difficile* (PDB ID, 2L7K). (a) Sequence alignment of a subset of the Pfam PF14203 protein domain family: *Clostridium difficile*, Q18AW3_CLOD6; *Faecalibacterium prausnitzii*, C7H731_9FIRM; *Ruminococcus obeum*, D4LS42_9FIRM; *Erysipelotrichaceae bacterium*, D6HRU5_9FIRM; *Ruminococcus sp.*, D4LNK2_9FIRM. Sequences were aligned using Clustal Omega [37] and the sequence alignment was rendered using ESPrpt [38]. Boxed residues represent identical (white font, highlighted in red) and similar (red font) amino acid conservation. Residue numbers for CD1104B and secondary structural elements from its solution NMR structure (PDB ID, 2L7K) are shown above the alignment. (b) Ribbon representation (only residues 2–68 are shown for clarity) of the lowest energy conformer of CD1104B. α -helices are shown in red and yellow while other polypeptide segments are shown in grey. The N- and C-termini are labeled with 'N' and 'C', respectively and the helices are indicated as I–IV. (c) Line representations of the polypeptide backbone depicting the ensemble of 20 structures after superposition of the backbone heavy atoms N, C and C of the α -helices for minimal RMSD. (d) Sausage representation of the backbone and superposition of the core side chains (Table I). A spline curve was drawn through the mean positions of C atoms with the thickness proportional to the mean global displacement of backbone heavy atoms in the 20 conformers superimposed in (c). (e) The degree of residue conservation within protein family PF14203 is color-coded according to the default ConSurf scheme on space filling representation of lowest energy conformer: burgundy for the strongest conservation and cyan for the highest variability. The image on the left has the same orientation as in Fig. 1b and the image on the right was obtained after a 180° rotation about the vertical axis. The two largest surface clefts are indicated as 'C1' and 'C2' (see text). (f) Sausage representation of backbone with the side chains of strictly conserved residues Glu 8 and Lys 48 taken from

lowest energy conformer shown in blue. The salt-bridge formed between the side chains is highlighted **(g)** Surface and space-filling representations of the lowest-energy conformer of CD1104B colored according scores from PredUs prediction of regions potentially involved in protein-protein interactions: the default PredUs color scheme for residues with scores larger than zero are shown from light red to red with increasing score. **(h)** Electrostatic surface potential map displayed for the lowest energy NMR conformer. Positively and negatively charged regions are colored in blue and red, respectively, and neutral regions are in grey.

Table 1

Summary of structural statistics^a

CD1104B		
<i>Completeness of resonance assignments^b</i>		
Backbone (%)		96.6
Side chain (%)		99.1
Aromatic (%)		100
Stereospecific methyl (%)		100
<i>Conformationally-restricting constraints^c</i>		
Distance constraints		
Total		1592
intra-residue ($i=j$)		277
sequential ($ i-j =1$)		403
medium range ($1 < i-j < 5$)		526
long range ($ i-j \geq 5$)		386
Dihedral angle constraints		72
RDC constraints ($^{15}\text{N}-^1\text{H}$)		79
No. of constraints per residue		22.9
No. of long range constraints per residue		5.9
<i>Residual constraint violations^c</i>		
Average no. of distance violations per structure:		
0.1 – 0.2 Å		7.3
0.2 – 0.5 Å		1.4
> 0.5 Å		0
Average no. of dihedral angle violations per structure:		
1 – 10°		6.85
> 10°		0
<i>Model Quality^c</i>		
RMSD backbone atoms (Å) ^d		0.5
RMSD heavy atoms (Å) ^d		1.1
RMSD bond lengths (Å)		0.009
RMSD bond angles (°)		1.0
MolProbity Ramachandran statistics ^{c,d}		
Most favored regions (%)		99.3
Allowed regions (%)		0.6
Disallowed regions (%)		0.1
Global quality scores (Raw/Z-score) ^c		
Verify3D	0.28	–2.89
ProsaII	0.66	0.04
Procheck (phi-psi) ^d	0.32	1.57

	CD1104B	
Procheck (all) ^d	0.11	0.65
MolProbity clash score	18.89	-1.72
RPF Scores ^e		
Recall/Precision	0.957	0.907
F-measure/DP-score	0.931	0.825
<i>BMRB accession number</i>	17359	
<i>PDB ID</i>	2L7K	

^aStructural statistics computed for the ensemble of 20 deposited structures.

^bComputed using AVS software[14] from the expected number of resonances, excluding: highly exchangeable protons (N-terminal, Lys, and Arg amino groups, hydroxyls of Ser, Thr, Tyr), carboxyls of Asp and Glu, non-protonated aromatic carbons, and the C-terminal His₆ tag.

^cCalculated using PSVS 1.4 [36]. Average distance violations were calculated using the sum over r^{-6} .

^dBased on ordered residue ranges [S(phi) + S(psi) > 1.8], 3–14, 20–32, 37–52, 55–60.

^eRPF scores [15] reflecting the goodness-of-fit of the final ensemble of structures (including disordered residues) to the NOESY data and resonance assignments.