Published in final edited form as:
Proteins. 2009 September ; 76(4): 1037–1041. doi:10.1002/prot.22459.

NMR Structure of Protein YvyC from *Bacillus subtilis* Reveals Unexpected Structural Similarity between Two PFAM Families

Alexander Eletsky[1,4], **Dinesh K. Sukumaran**[1,4], **Rong Xiao**[2,4], **Tom Acton**[2,4], **Burkhard Rost**[3,4], **Gaetano T. Montelione**[2,4], and **Thomas Szyperski**[1,4,*]

[1]Department of Chemistry, State University of New York at Buffalo, Buffalo, New York 14260

[2]Center of Advanced Biotechnology and Medicine and Department of Molecular Biology and Biochemistry, Rutgers University, Piscataway, New Jersey 08854

[3]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, New York 10032

[4]Northeast Structural Genomics Consortium

Keywords

Structural genomics; GFT NMR; flagella; YvyC; chaperone

Introduction

109-residue protein YvyC (MW=13 kDa) from *B. subtilis* (gi|580862, SwissProt/TrEMBL ID YVYC_BACSU, accession number P39737) was selected as a target of the Protein Structure Initiative 2 and assigned to the Northeast Structural Genomics consortium (NESG; http://www.nesg.org) for structure determination (NESG Target ID SR482). YvyC belongs to the Pfam1 protein family FlaG (PF03646) and is required for proper assembly of bacterial flagella. The FlaG family contains 215 members (for sequence alignment of the FlaG family seeds see Fig. S1 in the Supporting Information). The NMR structure of YvyC presented here is the first atomic resolution structure available for this family.

Self-assembly of bacterial flagella has been studied extensively in recent decades and comprehensive reviews are available.2–4 The gene encoding YvyC is a part of the *fliD* operon comprising the genes *yvyC*, *fliD*, *fliS* and *fliT*.5 This operon is located immediately downstream of the gene *hag* (also called *fliC*) encoding flagellin – the major protein required for assembly of the flagellar filaments. During filament elongation, flagellin monomers are exported through the central channel of a flagellum and oligomerize at its distal end.2,4 FliD (also named 'hook-associated protein 2', or HAP26) forms a pentameric cap at the distal end of a flagellum and is essential for oligomerization of flagellin.2,4,7 FliS and FliT serve as flagellin-specific8 and FliD-specific9,10 chaperones, respectively. It was demonstrated that mutations in *yvyC* gene orthologs in *P. fluorescens* (*flaG*) and *V. anguillarum* (ORF3) lead to phenotypes with unusually long filaments,11,12 but the specific role of YvyC remains unknown.

*Correspondence to: Thomas Szyperski, Department of Chemistry, The State University of New York at Buffalo, Buffalo, New York 14260. E-mail: szypersk@chem.buffalo.edu, Phone: (716) 6456800 (ext. 2250), Fax: (716) 6457338.

Materials and Methods

YvyC was cloned, expressed and purified following standard automated protocols to produce a uniformly $^{13}C,^{15}N$-labeled protein sample.13,14 Briefly, the full length *yvyC* gene from *Bacillus subtlis* was cloned into a pET21d (Novagen) derivative, yielding the plasmid pSR482-21.1. The resulting construct contains eight nonnative residues at the C-terminus (LEHHHHHH) to facilitate protein purification and one residue insertion (L) following the initiation codon introduced by cloning. *Escherichia Coli* BL21 (DE3) pMGK cells, a rare codon enhanced strain, were transformed with pSR482-21.1, and cultured in MJ9 minimal medium containing $(^{15}NH_4)_2SO_4$ and *U*-^{13}C-glucose as sole nitrogen and carbon sources. *U*-$^{13}C,^{15}N$ YvyC was purified using an AKTA Express (GE Healthcare) based two step protocol consisting of IMAC (HisTrap HP) and gel filtration (HiLoad 26/60 Superdex 75) chromatography. The final yield of purified *U*-$^{13}C,^{15}N$ YvyC (> 97% homogeneous by SDS-PAGE; 15.0 kDa by MALDI-TOF mass spectrometry) was ~48 mg/L. In addition, a *U*-^{15}N and 5% biosynthetically directed fractionally ^{13}C-labeled sample15 was generated for stereo-specific assignment of isopropyl methyl groups. *U*-$^{13}C,^{15}N$ and 5%^{13}C,*U*-^{15}N YvyC were dissolved, respectively, at concentrations of ~1.0 mM and 1.1 mM in 95% H_2O/5% D_2O (20 mM MES, 100 mM NaCl, 10 mM DTT, 5 mM $CaCl_2$, 0.02% NaN_3) at pH 6.5. An isotropic overall rotational correlation time of ~8 ns was inferred from ^{15}N spin relaxation times, indicating that the protein is monomeric in solution under the conditions used for these NMR studies. This conclusion was further confirmed by analytic gel-filtration in 100 mM Tris, 100 mM NaCl, 250 ppm NaN_3, at pH 7.5, with detection using a combination of static light scattering and refractive index (as described in ref 12); under these conditions the sample was observed to be > 97% monomeric.

All NMR spectra were recorded at 25 °C. Five G-matrix Fourier transform (GFT) NMR experiments17 and a simultaneous 3D $^{15}N/^{13}C^{aliphatic}/^{13}C^{aromatic}$-resolved NOESY18 spectrum (mixing time 60 ms) were acquired on a Varian INOVA 750 MHz spectrometer equipped with a conventional probe. 2D constant-time [^{13}C, ^{1}H]-HSQC spectra with 28 ms and 56 ms constant-time delays were recorded for the 5% biosynthetically directed fractionally ^{13}C-labeled sample on a Varian INOVA 600 MHz spectrometer equipped with a cryogenic probe in order to obtain stereo-specific assignments for isopropyl groups of valines and leucines.15 Spectra were processed using the program PROSA19 and analyzed using the program CARA.20 Sequence-specific backbone (H^N, H^α, N, C^α) and H^β/C^β resonance assignments were obtained by using (4,3)D HNN$\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$ / $\underline{C}^{\alpha\beta}\underline{C}^{\alpha}$(CO)NHN and (4,3)D $\underline{H}^{\alpha\beta}\underline{C}^{\alpha\beta}$(CO)NHN along with the program AutoAssign program.21 Side-chain spin system identification was accomplished by using aliphatic16,17 and aromatic22 (4,3)D $\underline{HC}$CH. Assignments were obtained for 100% of backbone and side-chain chemical shifts assignable with the NMR experiments listed above (excluding N-terminal NH_3^+, Lys NH_3^+, Arg NH_2, OH of Ser, Thr and Tyr, $^{13}C^\varepsilon$ of Asp and Asn, $^{13}C^\delta$ of Glu and Gln, and aromatic $^{13}C^\gamma$ shifts; Table I). Stereo-specific assignments were obtained for all Val and Leu methyl groups and for 40% of the β-methylene groups exhibiting non-degenerate chemical shifts (Table I). Chemical shifts were deposited in the BioMagResBank on 06/15/2006 with accession code 7170.

^{1}H–^{1}H upper distance limit constraints for structure calculations were obtained from NOESY (Table I). In addition, backbone dihedral angle constraints were derived from chemical shifts using the program TALOS23 for residues located in well-defined secondary structure elements (Table I). The programs CYANA24,25 and AUTOSTRUCTURE26 were used in parallel to assign NOEs by consensus, and the remaining assignments were carrieMOVEd by interactive spectral analysis.27 The final structure calculation was performed with CYANA, and the 20 conformers with the lowest target function value were

refined in an 'explicit water bath'28 using the program CNS.29 The coordinates were deposited in the Protein Data Bank on 06/15/2006 (accession code 2HC5).

Results and Discussion

A high-quality NMR structure of protein YvyC (Table I) was obtained. The structure consists of three α-helices I-III (residues 9-22, 36-52 and 86-105) and three β-strands (residues 58-65, 68-75 and 81-85) forming one anti-parallel β-sheet with topology A(↑), B(↓), C(↑) (Fig. 1b). The helices form a three-helix bundle, which is attached to one side of the sheet. The secondary structure elements are locally and globally well-defined. The segment comprising residues 23-35 connecting helices I and II is largely disordered (Fig 1a,c), which is manifested by comparably narrow NMR lines and lack of medium- and long-range NOEs.

A search of the CATH database using the CATHEDRAL30 server did not yield a matching fold for protein YvyC, indicating that YvyC exhibits a distinct protein architecture. Moreover, a search of the PDB database for structurally similar proteins using both DALI31 and SSM32 identified protein YkfF from *E. coli* (NESG target ER397, PDB code 2HJJ) as the only significant (DALI Z-score = 3.3) hit. The structurally aligned fragment comprises 56 residues (r.m.s.d. of C^{α} atoms = 2.9 Å; sequence identity 5%) and contains five regular secondary structure elements (the β-sheet with strands A-C and α-helices II and III, Fig 1d). YkfF is the sole structural representative of Pfam family PF06006 comprising 38 proteins of unknown function. Hence, structure comparison of proteins YvyC from *B. Subtilis* and YkfF from *E. coli* indicates a possible distant homology between the thus far unrelated PFAM families PF03646 and PF06006.

Analysis of conserved surface features33 within the FlaG protein family PF03646 reveals a single cluster of highly conserved residues on the surface of YvyC (Fig. 1e). These residues belong to β-strands B and C. Furthermore, a search for protein surface cavities using the ProFunc34 server revealed three clefts (Fig. 1e,f) with volumes of ~1,800 Å3, ~800 Å3 and ~660 Å3. The calculation of the electrostatic protein surface potential calculated with the program MolMol35 shows that in protein YvyC cavity 2 is charged negatively while cavities 1 and 3 exhibit a mixed charge distribution (Fig. 1f). Considering (i) that residues Leu 58, Glu 75, Ile 82 in cavity 1 and Glu 83 and Pro 86 in cavity 3 are highly conserved and (ii) that functional sites on protein surfaces are mostly located in the largest cavities36, one may conclude that these two cavities are likely to be important for the molecular function.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The authors thank Dr. Hunjoong Lee and Prof. Diana Murray for helpful discussions. Supported was obtained from the National Institutes of Health (U54 GM074958-01) and the University at Buffalo's Center for Computational Research.

References

1. Finn RD, Mistry J, Schuster-Bockler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer ELL, Bateman A. Pfam: clans, web tools and services. Nucleic Acids Res. 2006; 34:D247–D251. [PubMed: 16381856]
2. Macnab RM. How bacteria assemble flagella. Annual Review of Microbiology. 2003; 57:77–100.

3. Chevance FFV, Hughes KT. Coordinating assembly of a bacterial macromolecular machine. Nat Rev Micro. 2008; 6:455–465.
4. Minamino T, Namba K. Self-assembly and type III protein export of the bacterial flagellum. J Mol Microbiol Biotechnol. 2004; 7:5–17. [PubMed: 15170399]
5. Chen L, Helmann JD. The Bacillus subtilis sigma D-dependent operon encoding the flagellar proteins FliD, FliS, and FliT. J Bacteriol. 1994; 176:3093–3101. [PubMed: 8195064]
6. Homma M, Derosier DJ, Macnab RM. Flagellar hook and hook-associated proteins of Salmonella typhimurium and their relationship to other axial components of the flagellum. J Mol Biol. 1990; 213:819–832. [PubMed: 2193164]
7. Homma M, Fujita H, Yamaguchi S, Iino T. Excretion of unassembled flagellin by salmonella-typhimurium mutants deficient in hook-associated proteins. J Bacteriol. 1984; 159:1056–1059. [PubMed: 6384179]
8. Auvray F, Thomas J, Fraser GM, Hughes C. Flagellin polymerisation control by a cytosolic export chaperone. J Mol Biol. 2001; 308:221–229. [PubMed: 11327763]
9. Bennett JCQ, Thomas J, Fraser GM, Hughes C. Substrate complexes and domain organization of the Salmonella flagellar export chaperones FlgN and FliT. Mol Microbiol. 2001; 39:781–791. [PubMed: 11169117]
10. Fraser GM, Bennett JCQ, Hughes C. Substrate-specific binding of hook-associated proteins by FlgN and FliT, putative chaperones for flagellum assembly. Mol Microbiol. 1999; 32:569–580. [PubMed: 10320579]
11. Capdevila S, Martinez-Granero FM, Sanchez-Contreras M, Rivilla R, Martin M. Analysis of Pseudomonas fluorescens F113 genes implicated in flagellar filament synthesis and their role in competitive root colonization. Microbiology-Sgm. 2004; 150:3889–3897.
12. McGee K, Horstedt P, Milton DL. Identification and characterization of additional flagellin genes from Vibrio anguillarum. J Bacteriol. 1996; 178:5188–5198. [PubMed: 8752337]
13. Acton, TB.; Gunsalus, KC.; Xiao, R.; Ma, LC.; Aramini, J.; Baran, MC.; Chiang, YW.; 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. Methods in Enzymology. Vol. 394. San Diego: Elsevier Academic Press Inc; 2005. Robotic cloning and protein production platform of the Northeast Structural Genomics Consortium. Nuclear Magnetic Resonance of Biological Macromolecules, Part C; p. 210-243.
14. Jansson M, Li YC, Jendeberg L, Anderson S, Montelione GT, Nilsson B. High-level production of uniformly N-15- and C-13-enriched fusion proteins in Escherichia coli. J Biomol NMR. 1996; 7:131–141. [PubMed: 8616269]
15. 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]
16. Szyperski T, Yeh DC, Sukumaran DK, Moseley HNB, Montelione GT. Reduced-dimensionality NMR spectroscopy for high-throughput protein resonance assignment. Proc Natl Acad Sci U S A. 2002; 99:8009–8014. [PubMed: 12060747]
17. Atreya HS, Szyperski T. G-matrix Fourier transform NMR spectroscopy for complete protein resonance assignment. Proc Natl Acad Sci U S A. 2004; 101:9642–9647. [PubMed: 15210958]
18. Shen Y, Atreya HS, Liu GH, Szyperski T. G-matrix Fourier transform NOESY-based protocol for high-quality protein structure determination. J Am Chem Soc. 2005; 127:9085–9099. [PubMed: 15969587]
19. Guntert P, Dotsch V, Wider G, Wuthrich K. Processing of multidimensional NMR data with the new software PROSA. J Biomol NMR. 1992; 2:619–629.
20. Keller, R. The Computer Aided Resonance Assignment Tutorial. Goldau: CANTINA Verlag; 2004. p. 81
21. 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]

22. Eletsky A, Atreya HS, Liu GH, Szyperski T. Probing structure and functional dynamics of (large) proteins with aromatic rings: L-GFT-TROSY (4,3)D HCCH NMR spectroscopy. J Am Chem Soc. 2005; 127:14578–14579. [PubMed: 16231903]
23. 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]
24. 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]
25. 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]
26. Huang YJ, Tejero R, Powers R, Montelione GT. A topology-constrained distance network algorithm for protein structure determination from NOESY data. Proteins-Structure Function and Bioinformatics. 2006; 62:587–603.
27. Liu GH, 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]
28. Linge JP, Williams MA, Spronk C, Bonvin A, Nilges M. Refinement of protein structures in explicit solvent. Proteins. 2003; 50:496–506. [PubMed: 12557191]
29. 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]
30. Redfern OC, Harrison A, Dallman T, Pearl FMG, Orengo CA. CATHEDRAL: A fast and effective algorithm to predict folds and domain boundaries from multidomain protein structures. PLoS Comput Biol. 2007; 3:2333–2347.
31. Holm L, Sander C. Dali - a network tool for protein structure comparison. Trends BiochemSci. 1995; 20:478–480.
32. Krissinel E, Henrick K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr Sect D-Biol Crystallogr. 2004; 60:2256–2268. [PubMed: 15572779]
33. Landau M, Mayrose I, Rosenberg Y, Glaser F, Martz E, Pupko T, Ben-Tal N. ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. Nucleic Acids Res. 2005; 33:W299–W302. [PubMed: 15980475]
34. Laskowski RA, Watson JD, Thornton JM. ProFunc: a server for predicting protein function from 3D structure. Nucl Acids Res. 2005; 33:W89–93. [PubMed: 15980588]
35. Koradi R, Billeter M, Wuthrich K. MOLMOL: A program for display and analysis of macromolecular structures. J Mol Graph. 1996; 14:51–55. [PubMed: 8744573]
36. Laskowski RA, Luscombe NM, Swindells MB, Thornton JM. Protein clefts in molecular recognition and function. Protein Sci. 1996; 5:2438–2452. [PubMed: 8976552]
37. Laskowski RA, Rullmann JAC, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J Biomol NMR. 1996; 8:477–486. [PubMed: 9008363]
38. Word JM, Bateman RC, Presley BK, Lovell SC, Richardson DC. Exploring steric constraints on protein mutations using MAGE/PROBE. Protein Sci. 2000; 9:2251–2259. [PubMed: 11152136]
39. 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]
40. Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. Proteins-Structure Function and Bioinformatics. 2007; 66:778–795.

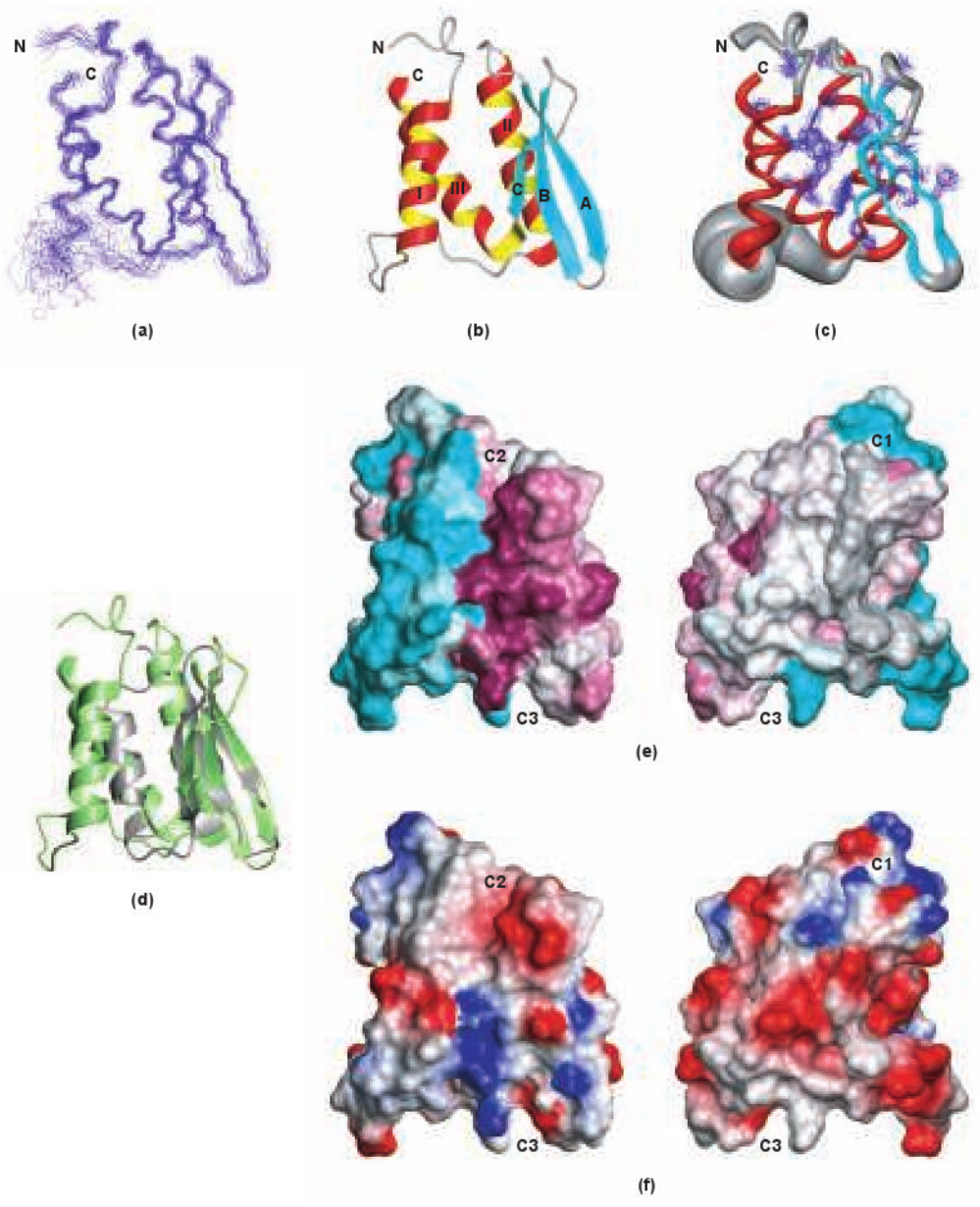

Fig. 1.

NMR structure of YvyC. **(a)** Backbone trace of residues 1-104 of the 20 representative CYANA conformers after superposition of backbone N, C^{α} and C′ atoms of the regular secondary structure elements for minimal root-mean-square deviation (RMSD). **(b)** Ribbon drawing of residues 1-104 of the conformer with the lowest CYANA target function. α-helices I-III are shown in red and yellow, β-strands A-C are shown in cyan, other polypeptide segments are shown in gray and the N- and C-termini are labeled as “N” and “C”. **(c)** Sausage representation of backbone and best defined side chains. A spline curve was drawn through the mean positions of C^{α} atoms of residues 1-104 with the thickness proportional to the mean global displacement of C^{α} atoms in the 20 conformers superimposed in (a). α-helices I-III are shown in red, β-strands A-C are shown in cyan, other polypeptide segments are shown in gray and a superposition of 34 side chains with the lowest global displacement is shown in blue. **(d)** Ribbon drawing of the conformers with the

lowest CYANA target function of YvyC (pale green) and YkfF (grey) after superposition of Ca atoms of residues 37-46, 60-63, 70-74, 81-85, 90-98 and 14-23, 27-30, 38-42, 48-52, 59-67, respectively. For YvyC only residues 1-104 are shown. **(e)** Surface representation of the conformer with the lowest CYANA target function. The structure shown on the left has the same orientation as in (a-c), while the one on the right is rotated by 180° about the vertical axis. Surface colors represent sequence conservation among the seed sequences of the FlaG protein family calculated with ConSurf,33 with burgundy corresponding to the highest conservation and cyan – to the highest variability. The cavities are labeled as C1, C2 and C3. **(f)** Same as (e), but with the colors according to the electrostatic potential. All figures were prepared with the program MOLMOL.35

TABLE I
Statistics of NMR Structure of Protein YvyC

Completeness of stereo-specific assignments[a] [%]	
$^{\alpha}CH_2$ of Gly	100 (1/1)
$^{\beta}CH_2$	40 (28/70)
Val and Leu methyl groups	100 (22/22)
Conformationally restricting distance constraints	
Intraresidue [$i = j$]	548
Sequential [$\|i - j\| = 1$]	728
Medium Range [$1 < \|i - j\| < 5$]	697
Long Range [$\|i - j\| \geq 5$]	916
Total	2889
Dihedral angle constraints	
φ	35
ψ	35
Average number of constraints per residue	25.8
Average number of long-range distance constraints per residue	8.2
CYANA target function [$Å^2$]	2.33±0.12
Average number of distance constraints violations per CYANA conformer	
0.2 – 0.5 Å	0
> 0.5 Å	0
Average number of dihedral-angle constraint violations per CYANA conformer	
> 5°	0
Average r.m.s.d. to the mean CYANA coordinates [Å]	
Regular secondary structure elements[b], backbone heavy atoms	0.60±0.10
Regular secondary structure elements[b], all heavy atoms	1.00±0.11
Ordered residues[c], backbone heavy atoms	0.97±0.36
Ordered residues[c], all heavy atoms	1.43±0.35
Heavy atoms of molecular core including best-defined side chains[d]	0.70±0.09
PROCHECK37 G-factors raw score (φ and ψ / all dihedral angles) [c]	-0.15/-0.20
PROCHECK37 G-factors Z-score (φ and ψ / all dihedral angles) [c]	-0.28/-1.18
MOLPROBITY38 clash score (raw / Z-score)[c]	24.03/-2.66
AutoQF R/P/DP scores39 [%]	95/95/85
Ramachandran plot summary (residues 2-17, 37-104) [c] [%]	
most favored regions	91.2
Additionally allowed regions	7.8
generously allowed regions	0.4
disallowed regions	0.5

[a] Relative to pairs with non-degenerate chemical shifts.

[b] Residues 10-19, 37-49, 58-65, 68-75, 80-85, 90-101.

[c] Residues 3-5, 8-21, 38-55, 58-64, 67-104, 108-111. Ordered residues were defined based on dihedral angle order parameters, and Z-scores were computed relative to corresponding structure quality measures for high resolution X-ray crystal structures, as described in reference.40

[d] Backbone and side-chain heavy atoms of residues 3, 6, 10-13, 16, 41, 43-45, 47, 48, 50, 51, 53, 54, 59, 62, 63, 71-74, 78, 81, 82, 85-87, 90, 94-97, 102, 103.