

STRUCTURE NOTE

Solution NMR structure of the SOS response protein YnzC from *Bacillus subtilis*

James M. Aramini,^{1,2*} Seema Sharma,^{1,2} Yuanpeng J. Huang,^{1,2} G. V. T. Swapna,^{1,2} Chi Kent Ho,^{1,2} Karishma Shetty,^{1,2} Kellie Cunningham,^{1,2} Li-Chung Ma,^{1,2} Li Zhao,^{1,2} Leah A. Owens,^{1,2} Mei Jiang,^{1,2} Rong Xiao,^{1,2} Jinfeng Liu,^{2,3} Michael C. Baran,^{1,2} Thomas B. Acton,^{1,2} Burkhard Rost,^{2,3} and Gaetano T. Montelione^{1,2,4*}

¹ Department of Molecular Biology and Biochemistry, Center for Advanced Biotechnology and Medicine, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854

² Northeast Structural Genomics Consortium, Rutgers University, Piscataway, New Jersey 08854

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

⁴ Department of Biochemistry, Robert Wood Johnson Medical School, University of Medicine and Dentistry of New Jersey, Piscataway, New Jersey 08854

Key words: DUF896 family protein; solution NMR structure; SOS response protein; partially unfolded; Northeast Structural Genomics Consortium.

INTRODUCTION

To combat the mutagenic effects of exposure to DNA-damaging agents such as UV radiation and genotoxic chemicals, bacteria have evolved elaborate DNA repair mechanisms, collectively termed the SOS response.¹ The SOS response is triggered by the ssDNA-induced binding of RecA to the SOS regulon repressor LexA, and the subsequent activation of a plethora of SOS or damage-inducible (*din*) genes under its control. In both *Escherichia coli* and *Bacillus subtilis*, over 30 genes have been shown to be under the control of the LexA protein (formerly known as DinR in *B. subtilis*).^{2–4} In *B. subtilis*, one small SOS response operon under the control of LexA, the *yneA* operon, is comprised of three genes: *yneA*, *yneB*, and *ynzC*.⁵ Of the three gene products, YneA has been shown to suppress cell division during the SOS response,⁵ whereas the exact roles of YneB and YnzC are unknown.

The *ynzC* gene of *B. subtilis* encodes for a 77-residue basic protein [SWISS-PROT ID: YNZC_BACSU; NESG target ID: SR384] that is a member of the DUF896 protein domain family (Pfam identifier: PF05979). This family of small (<90 aa) proteins with unknown function is found in over 100 bacterial species (Pfam 22.0), almost

exclusively from the predominantly Gram-positive firmicutes. A multiple sequence alignment of *B. subtilis* YnzC with DUF896 protein domains from selected genera of this phylum is shown in Figure 1(A). In this article, we present the solution NMR structure of *B. subtilis* YnzC as well as a truncated form of the protein designed on the basis of ¹H/²H exchange mass spectrometry (DXMS) results.⁶ The N-terminal half of YnzC folds into an anti-parallel helix-loop-helix motif, which remains intact in the truncated construct, whereas the remainder of the protein is disordered in solution. The structure of *B. subtilis* YnzC constitutes the first structural representative of the DUF896 protein domain family.

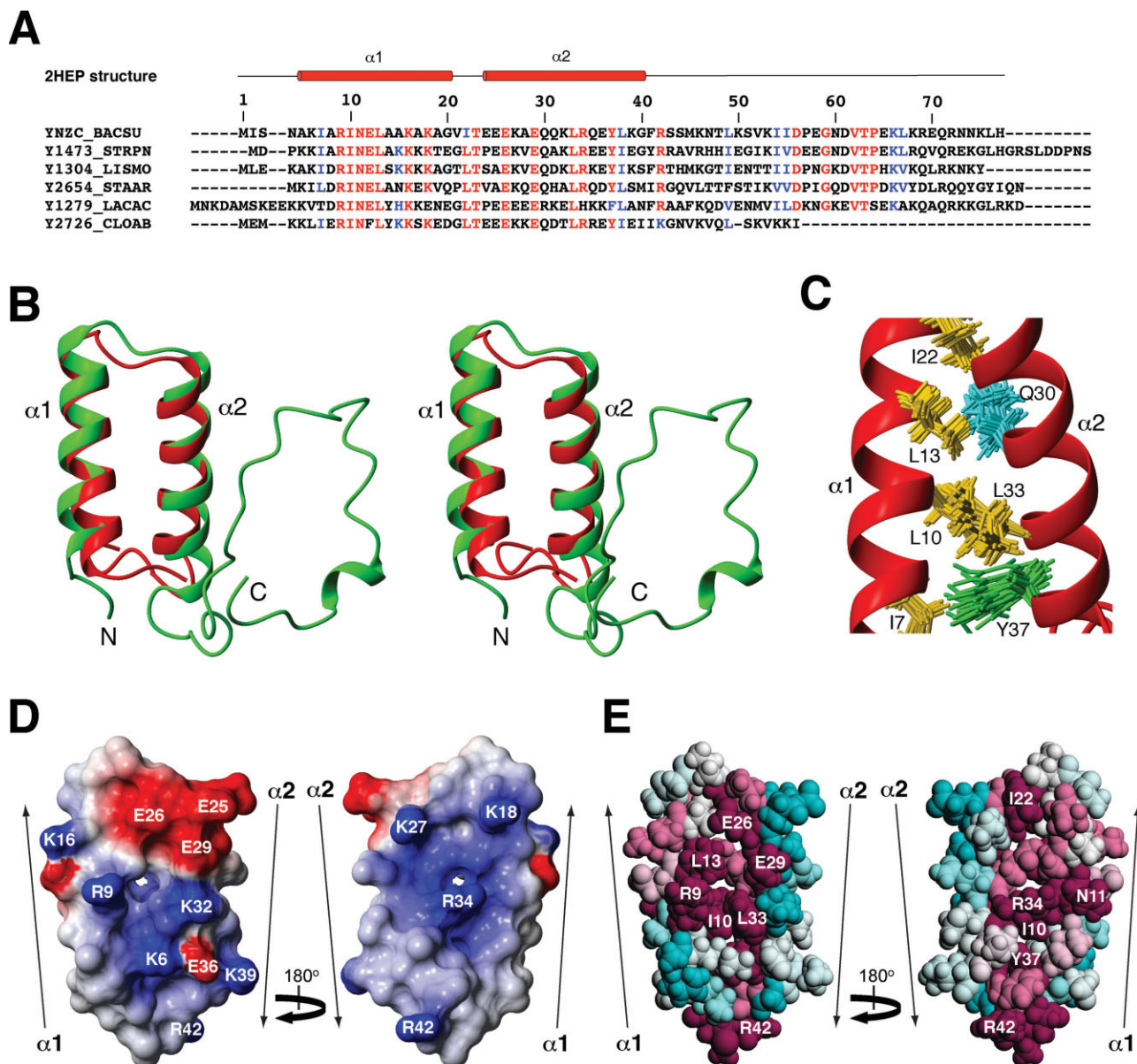
The Supplementary Material referred to in this article can be found online at <http://www.interscience.wiley.com/jpages/0887-3585/suppmat/>

Grant sponsor: National Institute of General Medical Sciences Protein Structure Initiative; Grant number: U54-GM074958.

*Correspondence to: James M. Aramini and Gaetano T. Montelione, CABM-Rutgers University, 679 Hoes Lane, Piscataway, NJ 08854. E-mail: guy@cabm.rutgers.edu or jma@cabm.rutgers.edu

Received 15 February 2008; Accepted 24 February 2008

Published online 22 April 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.22064

**Figure 1**

(A) A subset from the multiple sequence alignment of the entire DUF896 (PF05979) protein domain family (Pfam release 22.0) generated using Clustal X. The alignment includes YnzC from *Bacillus subtilis* plus representatives from the major genera of firmicutes that possess this protein (*Streptococcus*, *Listeria*, *Staphylococcus*, *Lactobacillus*, *Clostridium*). Amino acid residues identical or similar in 80% of the entire family are shown in red and blue, respectively. Complete protein sequences were used in the alignment and the conserved residues were colored using the BOXSHADE server. The sequence numbering for YnzC from *B. subtilis* and the secondary structural elements found in its solution NMR structure (PDB ID, 2HEP) are shown above the alignment. (B) Stereoview of the ribbon representation of the lowest energy conformers (lowest CNS energy) from the final solution NMR structures of full length YnzC (green) and the truncated YnzC-1-46 construct (red). The secondary structural elements are labeled. (C) A view into the core of the YnzC-1-46 structure showing key hydrophobic (gold), aromatic (green) and polar (cyan) side chains that form the interface between the two helices. (D) Electrostatic potential surface diagrams of the interhelical surfaces made by helix 1 and 2 in YnzC. For clarity, the unstructured C-terminal region of the protein has been omitted and only the structured residues (1-42) are shown. (E) ConSurf images of the same interhelical faces of YnzC based on the multiple sequence alignment of the entire DUF896 (PF05979) protein domain family. Residue coloring, reflecting the degree of residue conservation over the entire family, ranges from magenta (highly conserved) to cyan (variable).

METHODS

Uniformly ^{13}C , ^{15}N - and 5%- ^{13}C , U- ^{15}N -enriched *B. subtilis* YnzC and truncated YnzC-1-46 were cloned, expressed, and purified following standard protocols of

the NESG consortium⁷; see Supplementary Material for a complete description of the methods used in this work. Briefly, protein samples for NMR spectroscopy were concentrated to 1.1–1.4 mM in 95% H_2O /5% D_2O solution containing 20 mM MES, 100 mM NaCl, 10 mM DTT,

5 mM CaCl₂ at pH 6.5. Static light scattering data demonstrate that both the full length YnzC and truncated YnzC-1-46 proteins are monomeric in solution under the conditions used in the NMR studies (Supplementary Fig. S1). All NMR data were collected at 20°C on Varian INOVA 500 and 600 MHz and Bruker AVANCE 600 and 800 NMR spectrometers. Complete ¹H, ¹³C, and ¹⁵N resonance assignments for full length YnzC and truncated YnzC-1-46 were determined using GFT^{8,9} and conventional¹⁰ triple resonance NMR methods, respectively, and deposited in the BioMagResDB (BMRB accession numbers 7225 and 15476). Resonance assignments were validated using the Assignment Validation Suite (AVS) software package.¹¹ The full length YnzC structure was determined using AutoStructure 2.1.1¹² interfaced with XPLOR-NIH 2.11.2.¹³ The folded N-terminal residues (1–42) of the 20 lowest energy structures out of 100 calculated were deposited into the Protein Data Bank (PDB ID, 2HEP). The truncated YnzC(1-46) structure was calculated using CYANA 2.1^{14,15} and refined by restrained molecular dynamics in explicit water using CNS 1.2.^{16,17} The final ensemble of 20 models out of 100 calculated (excluding the C-terminal His₆) were deposited into the Protein Data Bank (PDB ID, 2JVD). Structural statistics and global structure quality factors, including Verify3D,¹⁸ ProsaII,¹⁹ PROCHECK,²⁰ and MolProbity²¹ raw and statistical Z-scores, were computed using the PSVS 1.3 software package.²² Values for the global goodness-of-fit of the final structure ensembles with the NOESY peak list data were determined using the RPF analysis program.²³

RESULTS AND DISCUSSION

The NMR solution structure of *B. subtilis* YnzC features two long (16–17 residue) α helices (α 1, A5-G20; α 2, E24-G40) followed by a disordered C-terminal tail. Stereo ribbon diagrams of representative structures of the superimposed full length YnzC and the YnzC-1-46 truncated construct are shown in Figure 1(B), and structural statistics are given in Table I. The α helices in the structured N-terminal half of the protein adopt an antiparallel helix-loop-helix motif in both full length and truncated YnzC that features several ordered side chains in its core [Fig. 1(C)]. The relative orientations of the helices in the two structures are quite similar, as indicated by a backbone RMSD of 0.84 Å between the average structures for the two ensembles (for residues 5–19 and 22–38). This is consistent with the minimal perturbations to the NMR resonance assignments observed for residues common to the full length and truncated proteins (Supplementary Fig. S2).⁶ Both structures exhibit excellent structure quality assessment scores (Table I). The reduced RPF scores for full length YnzC compared to the truncated construct can be attributed to poorer spectral quality and analysis

difficulties resulting from the intrinsically unstructured portion of the native protein. Electrostatic surface potential²⁴ [Fig. 1(D)] and ConSurf²⁵ [Fig. 1(E)] images of the two interhelical faces formed by the helix-loop-helix motif in YnzC, reveal that the core of the structure is stabilized by several strongly conserved hydrophobic residues (I10, L13, I22, L33, Y37), as well as potential interhelical electrostatic interactions between juxtaposed oppositely charged residues (K6:E36; R9:E29; K16:E26) on one face. Of these, the R9:E29 and K16:E26 salt bridges are strongly conserved across the DUF896 protein domain family [Fig. 1(A)]. The opposite interhelical face is strongly basic in nature, and overall, the helical interface is lined with numerous highly conserved residues. Contrary to secondary structure predictions, which include an additional short β -strand and a third α -helix near the C-terminus (Supplementary Fig. S3), the final \approx 35 residues are intrinsically unfolded in the structure of full length YnzC. This result is corroborated by a lack of NOE and chemical shift (CSI) data indicative of secondary structural elements, heteronuclear ¹⁵N NOE data (Supplementary Fig. S3), and DXMS results.⁶

The structures of full length and truncated YnzC presented here represent the first structures from the DUF896 protein domain family. In terms of uniqueness, YnzC shares no significant pairwise sequence identity (i.e., < 30%) to any protein deposited in the Protein Data Bank to date. The folded N-terminal half of the protein adopts an antiparallel helix-loop-helix motif. Helix-turn-helix (HTH) motifs are ubiquitous in nature, and often play roles as transcription factors.²⁶ However, HTH motifs commonly feature a tri-helical bundle, and variations on this theme, with the third helix involved in DNA binding.²⁶ Predictably, Dali²⁷ searches using both YnzC structures uncovered numerous examples of similar structural motifs ($Z < 5.5$) in other small proteins or within larger systems, many of which are involved in nucleic acid binding and protein transport. In *Bacillus subtilis*, a similar antiparallel helix motif with extremely low sequence identity compared to YnzC is employed by the 46-residue histidine kinase inhibitor Sda (PDB ID, 1PV0; Z , 3.5; C $^{\alpha}$ RMSD, 2.1 Å; 9% identity) to arrest the initiation of sporulation in response to DNA damage and replication defects.²⁸ Of relevance to DNA repair are Dali hits to the X-ray crystal structures of the isolated C-terminal helix-loop-helix motif from *E. coli* UvrB (PDB ID, 1QOJ; Z , 3.6; C $^{\alpha}$ RMSD, 2.6 Å; 17% identity) and a ternary complex involving *B. subtilis* UvrB (PDB ID, 2D7D; Z , 3.9; C $^{\alpha}$ RMSD, 2.3 Å; 16% identity), part of the UvrABC nucleotide excision repair pathway.^{29,30} In both of these structures, and a solution NMR structure of the *E. coli* UvrB C-terminal domain,³¹ the helix-loop-helix motif forms an antiparallel dimer featuring a small dimer interface comprising the interhelical loop and flanking residues in the helices, and the motif may be involved in mediating interactions with DNA and other

Table 1Summary of NMR and Structural Statistics for *B. subtilis* YnzC and YnzC-1-46^a

	YnzC		YnzC-1-46	
Completeness of resonance assignments ^b				
Backbone (%)	91.9		98.8	
Side chain (%)	78.3		96.9	
Aromatic (%)	100		100	
Stereospecific methyl (%)	87.5		100	
Conformationally-restricting constraints ^c				
Distance constraints				
Total	403		1022	
Intra-residue ($i = j$)	114		306	
Sequential ($ i - j = 1$)	118		206	
Medium range ($1 < i - j < 5$)	126		318	
Long range ($ i - j \geq 5$)	45		192	
Distance constraints per residue	9.8		23.2	
Dihedral angle constraints				
Total	67		68	
Hydrogen bond constraints				
Total	42		48	
Long range ($ i - j \geq 5$)	0		0	
Number of constraints per residue	12.5		25.9	
Number of long range constraints per residue	1.1		4.4	
Residual constraint violations ^c				
Average number of distance violations per structure				
0.1–0.2 Å	1.05		0.45	
0.2–0.5 Å	0		0	
>0.5 Å	0		0	
Average RMS distance violation/constraint (Å)	0.01		0.00	
Maximum distance violation (Å)	0.17		0.14	
Average number of dihedral angle violations per structure				
1–10°	0.75		0.1	
>10°	0		0	
Average RMS dihedral angle violation/constraint (degree)	0.19		0.05	
Maximum dihedral angle violation (degree)	2.1		1.2	
RMSD from average coordinates (Å) ^{c,d}				
Backbone atoms	0.9		0.3	
Heavy atoms	1.5		0.8	
Ramachandran plot statistics ^{c,d}				
Most favored regions (%)	96.4		98.6	
Additional allowed regions (%)	3.6		1.4	
Generously allowed (%)	0.0		0.0	
Disallowed regions (%)	0.0		0.0	
Global quality scores ^c				
	Raw	Z-score	Raw	Z-score
Verify3D	0.17	−4.65	0.24	−3.53
ProsaII	0.74	0.37	0.91	1.08
Procheck(phi-psi) ^d	0.11	0.75	0.53	2.40
Procheck(all) ^d	−0.12	−0.71	0.50	2.96
Molprobrity clash	14.69	−1.00	15.44	−1.12
RPF Scores ^e				
Recall	0.952		0.988	
Precision	0.839		0.944	
F-measure	0.892		0.966	
DP-score	0.628		0.822	

^aStructural statistics were computed for the ensemble of 20 deposited structures.^bComputed using AVS software¹¹ from the expected number of peaks, 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 tag.^cCalculated using PSVS 1.3 program.²² Average distance violations were calculated using the sum over r^{-6} . For YnzC (2HEP) only residues 1–42 are considered.^dOrdered residue ranges [$S(\phi) + S(\psi) > 1.8$]: YnzC: 5-20, 24-40; YnzC-1-46: 2–38.^eRPF scores^{2,3} reflecting the goodness-of-fit of the final ensemble of structures (including disordered residues) to the NMR data.

Uvr proteins. Although we have no evidence for dimer or oligomer formation of YnzC in solution, it is tempting to postulate that the largely basic interhelical face of YnzC featuring the conserved R34 represents a nucleic

acid binding site. Moreover, it is reasonable to suggest that the unfolded C-terminal region of full length YnzC, which features several conserved residues across the DUF896 protein domain family [Fig. 1(A)], becomes

structured upon binding to its biological target(s). Indeed, secondary structure prediction suggests that this disordered region has some tendency to form a third helix (residues P64-N73; Supplementary Fig. S3), which could be stabilized by intermolecular interactions. The confirmation of the exact role of YnzC in the *B. subtilis* SOS response awaits further structural and functional studies.

ACKNOWLEDGMENTS

We thank Paolo Rossi for valuable scientific discussions and Markus Fischer for insightful correspondence.

REFERENCES

- Friedberg EC, Walker GC, Siede W, Wood RD, Schultz RA, Ellenberger T. DNA repair and mutagenesis, 2nd ed. Washington, DC: ASM Press; 2006. 1089 p.
- Kelley WL. Lex marks the spot: the virulent side of SOS and a closer look at the LexA regulon. *Mol Microbiol* 2006;62:1228–1238.
- Fernandez de Henestrosa AR, Ogi T, Aoyagi S, Chaffin D, Hayes JJ, Ohmori H, Woodgate R. Identification of additional genes belonging to the LexA regulon in *Escherichia coli*. *Mol Microbiol* 2000;35:1560–1572.
- Au N, Kuester-Schoeck E, Mandava V, Bothwell LE, Canny SP, Chachu K, Colavito SA, Fuller SN, Groban ES, Hensley LA, O'Brien TC, Shah A, Tierney JT, Tomm LL, O'Gara TM, Goranov AI, Grossman AD, Lovett CM. Genetic composition of the *Bacillus subtilis* SOS system. *J Bacteriol* 2005;187:7655–7666.
- Kawai Y, Moriya S, Ogasawara N. Identification of a protein, Yne A, responsible for cell division suppression during the SOS response in *Bacillus subtilis*. *Mol Microbiol* 2003;47:1113–1122.
- Sharma S, Aramini JM, Zhang H, Huang YJ, Tejero R, Acton TB, Xiao R, Zhao L, Rossi P, Ma L-C, Swapna GVT, Lobel P, Montelione GT. Construct optimization for protein NMR structure analysis using amide $^1\text{H}/^2\text{H}$ exchange mass spectrometry. *Proteins*, submitted.
- Acton TB, Gunsalus KC, Xiao R, Ma L-C, Aramini JM, Baran MC, Chiang Y-W, Climent T, Cooper B, Denisova NG, Douglas SM, Everett JK, Ho CK, Macapagal D, Rajan PK, Shastry R, Shih L-Y, 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.
- Atreya HS, Szyperski T. G-matrix Fourier transform NMR spectroscopy for complete protein resonance assignment. *Proc Natl Acad Sci USA* 2004;101:9642–9647.
- Liu G, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran DK, Xiao R, Yee A, Lemak A, Bhattacharya A, Acton TB, Arrowsmith CH, Montelione GT, Szyperski T. NMR data collection and analysis protocol for high-throughput protein structure determination. *Proc Natl Acad Sci USA* 2005;102:10487–10492.
- Aramini JM, Huang YJ, Swapna GVT, Cort JR, Rajan PK, Xiao R, Shastry R, Acton TB, Liu J, Rost B, Kennedy MA, Montelione GT. Solution NMR structure of *Escherichia coli* ytfP expands the structural coverage of the UPF0131 protein domain family. *Proteins* 2007;68:789–795.
- 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.
- 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.
- Schwieters CD, Kuszewski JJ, Tjandra N, Clore GM. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 2003;160:65–73.
- Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. *J Mol Biol* 1997;273:283–298.
- Herrmann T, Güntert P, Wüthrich 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.
- Brünger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang J-S, 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 D* 1998;54:905–921.
- Linge JP, Williams MA, Spronk CAEM, Bonvin AMJJ, Nilges M. Refinement of protein structures in explicit solvent. *Proteins* 2003;50:496–506.
- Lüthy R, Bowie JU, Eisenberg D. Assessment of protein models with three-dimensional profiles. *Nature* 1992;356:83–85.
- Sippl MJ. Recognition of errors in three-dimensional structures of proteins. *Proteins* 1993;17:355–362.
- Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* 1993;26:283–291.
- Lovell SC, Davis IW, Arendall WB, III, de Bakker PIW, Word JM, Prisant MG, Richardson JS, Richardson DC. Structure validation by $\text{C}\alpha$ geometry: ϕ , ψ , and $\text{C}\beta$ deviation. *Proteins* 2003;50:437–450.
- Bhattacharya A, Tejero R, Montelione GT. Evaluating protein structures determined by structural genomics consortia. *Proteins* 2007;66:778–795.
- 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.
- Koradi R, Billeter M, Wüthrich K. MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 1996;14:51–55.
- Glaser F, Pupko T, Paz I, Bell RE, Bechor-Shental D, Martz E, Bentel N. ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics* 2003;19:163–164.
- Aravind L, Amantharaman V, Balaji S, Babu MM, Iyer LM. The many faces of the helix-turn-helix domain: transcription regulation and beyond. *FEMS Microbiol Rev* 2005;29:231–262.
- Holm L, Sander C. Protein structure comparison by alignment of distance matrices. *J Mol Biol* 1993;233:123–138.
- Rowland SL, Burkholder WF, Cunningham KA, Maciejewski MW, Grossman AD, King GE. Structure and mechanism of action of Sda, an inhibitor of the histidine kinases that regulate initiation of sporulation in *Bacillus subtilis*. *Mol Cell* 2004;13:689–701.
- Sohi M, Alexandrovich A, Moolenaar G, Visse R, Goosen N, Vernede X, Fontecilla-Camps JC, Champness J, Sanderson MR. Crystal structure of *Escherichia coli* UvrB C-terminal domain, and a model for UvrB-UvrC interaction. *FEBS Lett* 2000;465:161–164.
- Eryilmaz J, Ceschini S, Ryan J, Geddes S, Waters TR, Barrett TE. Structural insights into the cryptic DNA-dependent ATPase activity of UvrB. *J Mol Biol* 2006;357:62–72.
- Alexandrovich A, Czisch M, Frenkiel TA, Kelly GP, Goosen N, Moolenaar GE, Chowdhry BZ, Sanderson MR, Lane AN. Solution structure, hydrodynamics and thermodynamics of the UvrB C-terminal domain. *J Biomol Struct Dyn* 2001;19:219–236.