Volume 296, number 2, 148–152 FEBS 10605 January 1992

Low resolution solution structure of the *Bacillus subtilis* glucose permease IIA domain derived from heteronuclear three-dimensional NMR spectroscopy

Wayne J. Fairbrother[a], Garry P. Gippert[a], Jonathan Reizer[b], Milton H. Saier Jr.[b] and Peter E. Wright[a]

[a]*Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA* and [b]*Department of Biology, University of California at San Diego, La Jolla, CA 92093-0116, USA*

Received 30 October 1991

A low resolution solution structure of the IIA domain of the *Bacillus subtilis* phospho*enol*pyruvate-sugar phosphotransferase system (PTS) glucose permease has been determined using 945 inter-residue and 724 intra-residue distance constraints derived from three-dimensional ^{15}N and ^{13}C edited NOESY spectra. A total of 15 structures was generated using distance geometry calculations. The protein is comprised of 13 β-strands forming an antiparallel β-barrel. The average backbone atomic RMS deviation about the average distance geometry structure for the β-sheet residues is 1.1 Å. The conformations of the loop regions between the β-strands are less well determined. Backbone distance constraints obtained during the process of sequential assignment were insufficient to correctly calculate the polypeptide fold.

Phospho*enol*pyruvate-sugar phosphotransferase system (PTS); Glucose permease IIA^{glc} domain; Protein structure; Nuclear magnetic resonance; Distance geometry; *Bacillus subtilis*

1. INTRODUCTION

The bacterial phospho*enol*pyruvate-sugar phosphotransferase system (PTS) phosphocarrier protein, enzyme IIA^{glc} (previously called enzyme III^{glc} or factor III^{glc}), plays a central role in the regulatory processes of the PTS, in addition to its functional role in sugar phosphorylation and transport (for reviews see [1–3]). In *Bacillus subtilis* the glucose-specific PTS consists of three proteins: enzyme I, HPr and the sugar-specific permease, enzyme II, which consists of a cytosolic C-terminal IIA domain linked to the membrane bound IICB domains via a Q-linker. The 162-residue IIA domain of the *B. subtilis* glucose permease has been overexpressed in *Escherichia coli*, and has been shown to assume both the functional and regulatory roles of the soluble *E. coli* IIA^{glc} protein when expressed in *crr* mutants (i.e. those defective for the gene encoding IIA^{glc}) [4–6]. Site-directed mutagenesis experiments have established that HPr phosphorylates His^{83} of the IIA^{glc} domain. Replacement of His^{68} by alanine results in a protein that accepts a phosphoryl group from HPr but cannot transfer it to the IIB domain of the glucose permease. These results were analogous to those reported previously for mutations of His^{90} and His^{75} in the *E. coli* IIA^{glc} protein [7].

Recently, we reported the ^1H and ^{15}N backbone resonance assignments for the *B. subtilis* IIA^{glc} domain and determination of the secondary structure using principally three-dimensional (3D) ^1H–^{15}N NMR spectroscopy [8]. It was shown that IIA^{glc} contains three antiparallel β-sheets comprised of eight, three and two β- strands. No regular helical structure was identified although two short regions of irregular 'helix-like' structure or helical turns were noted. Parallel NMR studies of the homologous *E. coli* IIA^{glc} protein yielded similar results [9]. We have since assigned the $^{13}C^{\alpha}$ and aliphatic ^1H and ^{13}C resonances of the *B. subtilis* IIA^{glc} domain using 3D triple-resonance HCA(CO)N [10–12], 3D HCCH-COSY [13–15] and 3D HCCH-TOCSY [16] experiments [17]. The assignments enabled analysis of 3D ^1H–^{15}N NOESY-HMQC and ^1H–^{13}C NOESY-HSQC spectra, from which approximate interproton distance constraints could be obtained. By using the unambiguous backbone–backbone and backbone–side chain distance constraints obtained from the ^1H–^{15}N NOESY-HMQC spectrum and a subset of the unambiguous backbone–backbone, backbone–side chain and side chain–side chain distance constraints obtained from the ^1H-^{13}C NOESY-HSQC spectrum as input to distance geometry calculations, we were able to establish the global fold of the protein. In this paper we report the low resolution solution structure of the *B. subtilis* IIA^{glc} domain.

Correspondence address: P.E. Wright, Department of Molecular Biology, The Scripps Research Institute, La Jolla, CA 92037, USA. Fax: (1) (619) 554 9822.

Published by Elsevier Science Publishers B.V.

2. MATERIALS AND METHODS

Distance constraints were derived from a 100 ms mixing time 3D ^{1}H–^{15}N NOESY-HMQC spectrum obtained from a 0.8 mM sample of uniformly ^{15}N-labelled IIA^{glc} domain in 10 mM potassium phosphate, pH 6.6 (92% H_2O/8% D_2O), recorded at 308 K [8], and a 100 ms mixing time 3D ^{1}H–^{13}C NOESY-HSQC spectrum of a 2.0 mM sample of uniformly ^{15}N, ^{13}C-labelled IIA^{glc} domain in 10 mM potassium phosphate, pH* 6.7 (99.9% D_2O) recorded at 308 K. The NOEs were characterised as strong, medium or weak on the basis of their total F_1–F_3 cross peak volumes in the 3D NOESY spectra. Total volumes were obtained by adding the volumes measured for a given cross peak in the two or three F_2 planes in which it appeared. Upper bounds were determined by calibrating volumes versus known distances in the regular β-sheets. The resulting upper bound distance constraints were 2.5, 3.5 and 5.0 Å for backbone–backbone NOEs and 3.0, 4.0 and 5.0 Å for NOEs involving side chain protons. Lower bounds between non-bonded atoms were set to the sum of their van der Waals radii. A total of 945 inter-residue and 724 intra-residue distance constraints were used for the distance geometry calculations (Table I). Pseudo-atom corrections were added to interproton distance constraints where necessary [18]. In addition to the NOE-derived distance constraints, 35 hydrogen bond constraints ($r_{NH\text{-}O}$ = 1.8 to 2.0 Å and $r_{N\text{-}O}$ = 2.7 to 3.0 Å), identified on the basis of long-range backbone–backbone NOE connectivities [$d_{NN}(i,j)$, $d_{\alpha N}(i,j)$ and $d_{\alpha\alpha}(i,j)$] and slow amide proton exchange data [8], were used. Structures were calculated on a Cray YMP using a modified version of the distance geometry program DISGEO [19], except the all-atom embed stage which was implemented on a Convex C240.

3. RESULTS AND DISCUSSION

A total of 15 distance geometry (DG) structures were calculated for residues 12–162, using the distance constraints described above. The eleven N-terminal residues, which form part of the Q-linker in the intact glucose permease, IICBA, were omitted from the calculation because they appear to be disordered in solution. Evidence for this includes a complete absence of medium- or long-range NOEs for these residues, fast amide proton exchange rates [8], and a close correspondence between the ^{1}H chemical shifts for these residues and the 'random coil' values. More direct evidence for the N-terminal region being disordered and highly mobile in solution has also been obtained from measurements of ^{15}N relaxation parameters and steady state NOEs [20]. Analysis of these data using the Lipari and Szabo model free formalism [21] gives order parameters (S^2) for the 13 N-terminal residues between 0.2 and 0.55, indicating relatively unrestrained internal motion. Most other regions of the protein have order parameters in the range 0.75 to 0.90.

The residual constraint violations and backbone RMS deviations from the average DG structure are summarized in Tables II and III, respectively. A best fit superposition of the 15 DG structures is shown in Fig. 1. Overall, the polypeptide backbone conformation is reasonably well determined and the global fold of the protein has been established. The topology of the protein is best described as an antiparallel β-barrel. The main eight stranded β-sheet has a 'hybrid' Greek-key/jellyroll topology. In general, the 13 β-strands, with an average backbone RMS deviation about the average structure of ~1.1 Å (Table III), are better defined than the loops between strands and the N-terminal (residues 12–15) and C-terminal (residues 161–162) regions. Note that although the C-terminal residues 161 and 162 appear slightly disordered they were included as part of β-strand XIII for the purpose of calculating RMSDs for β-sheet residues in Table III. There also appears to be some disorder in the vicinity of the β-bulges previously identified [8] in β-strands IV, VII and IX. In addition to the β-sheets, irregular helical structure is evident between residues 89–92 and 117–123, consistent with the previous secondary structure analysis [8]. A further helical turn is found between residues 33–36. The latter residues are part of a long loop (25–41) which folds back

Table I

NMR distance constraints used in distance geometry calculations

Bound (Å)	No.	INTRA	SEQ	MBB	LBB	LNG
2.5	117	9	92	0	16	0
3.0	207	152	20	0	6	29
3.5	274	127	73	15	59	0
4.0	823	394	67	13	32	317
5.0	248	42	37	19	35	115
total	1669	742	289	47	148	461

NOEs from residue i to residue j: INTRA, intra-residue (j=i); SEQ, sequential backbone–backbone or backbone–β, ($|j-i|$ = 1); MBB, medium-range backbone–backbone or backbone–β ($|j-i|$ = 2 to 4); LBB, long-range backbone–backbone or backbone–β, ($|j-i| \geq 5$); LNG, all other inter-residue NOEs. Pro δ protons were counted with the β protons for this categorisation.

Table II

Summary of residual constraint violations for 15 distance geometry structures

Range (Å)	Average no. of distance constraint violations
0.1–0.2	74
0.2–0.3	37
0.3–0.4	23
0.4–0.5	15
> 0.5	27

The average maximum distance constraint violation is 1.2 Å.

Table III

Average backbone RMS deviations from average distance geometry structure

Residues	RMS deviation (Å)
12–162	1.32
β-sheet residues	1.11

β-sheet residues: 19–25, 42–46, 51–54, 58–64, 69–74, 79–84, 96–99, 104–106, 110–115, 129–134, 140–144, 149–152 and 158–162.

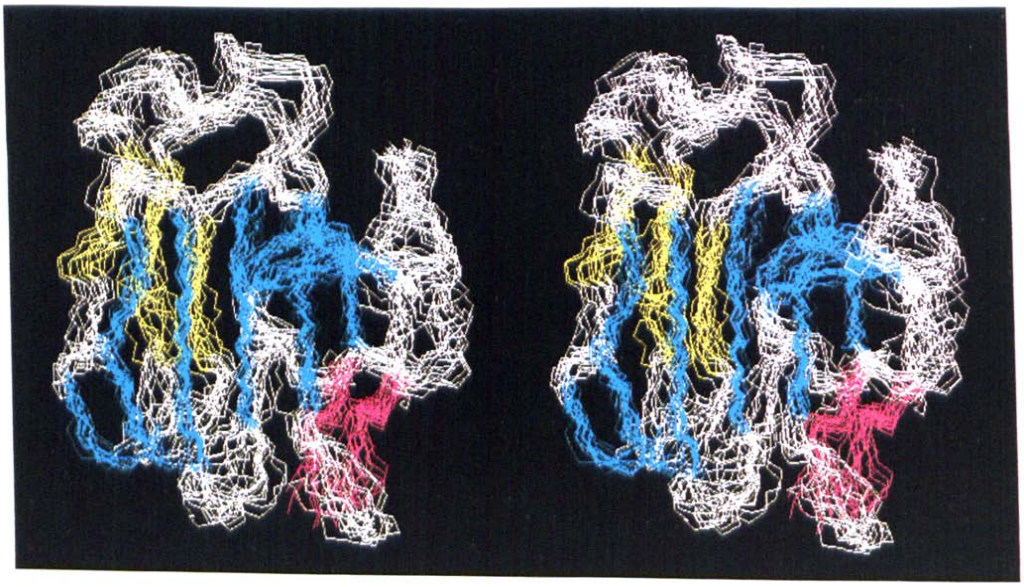

Fig. 1. Stereoview of the best-fit superposition of 15 DG structure of the IIA^{glc} domain. The three β-sheets are coloured according to the secondary structure assignment of [8]. Blue = strands I, II, IV, V, VI, VIII, X and XII; Yellow = strands III, VII and IX; Red = strands XI and XIII.

and packs onto the right side of the protein in the view of Fig. 1.

Attempts to calculate structures using only NOE constraints derived from the 3D ^1H–^{15}N NOESY-HMQC spectrum (d_{NN}, $d_{\alpha N}$ and $d_{\beta N}$ NOEs) and $d_{\alpha\alpha}$ NOEs obtained from a 2D ^1H NOESY spectrum of IIA^{glc} in D_2O (a total of 815 distance constraints; 432 inter-residue and 383 intra-residue) were largely unsuccessful. Although the secondary structure was reasonably well determined, the packing between the β-sheets was not, resulting in an open, clam shell-like structure with large hydrophobic surfaces exposed to the solvent. From these attempts it was clear that additional backbone–side chain and side chain–side chain distance constraints would be necessary to correctly pack the three β- sheets. Our experience with IIA^{glc} differs from that of Clore et al. [22] with IL-1β in that they were able to establish the global fold using only NOEs involving the amide, C^α and C^β protons. While it was obvious that the structures of IIA^{glc} calculated in this way were incorrect, this may not generally be the case and caution should therefore be used when calculating structures based only on the backbone NOE constraints obtained from ^{15}N edited spectra.

The NMR solution structure determination presented here is only the second example involving a protein of greater than 150 residues. The low resolution structure of IL-1β, mentioned above, has since been refined to a high resolution structure using additional constraints obtained from 3D and 4D NMR spectra of ^{15}N,^{13}C-labelled protein [23]. Although the present solution structure of the IIA^{glc} domain is of relatively low precision, it is sufficient to locate the active site and provide spatial information on residues which may be important for function. For instance, Fig. 2 shows the positions of the invariant amino acids from six sequenced PTS IIA domains [4,24–31]. There is an obvious cluster of conserved residues close to the site of phosphorylation, His^{83}, which may be important for either phosphoryl transfer or protein-protein recognition. His^{83} is located on the surface of the protein at the C-terminal end of β-strand VI, and is close in space to His^{68} (there are NOE connectivities between the side chain protons of His^{83} and His^{68}). The proximity of His^{68} to His^{83} was noted during our earlier determination of the secondary structure [8]. His^{68} has been implicated as being important for the function of the IIA^{glc} domain from site-directed mutagenesis experiments [6]. The low resolution structure presented here could be used to design other site-specific mutations in order to further investigate structure/function relationships. It is also worth noting that the N- and C-termini are situated on the opposite side of the protein to the active site.

At the same time as the low resolution solution structure determination was completed, the results of an independently determined X-ray crystal structure of the IIA^{glc} domain were published [32]. While we have not compared the coordinates of this 2.2 Å resolution crystal structure with our own structures, it is clear that the overall folds obtained by the X-ray and NMR methods are very similar. We are currently refining the solution structure of the IIA^{glc} domain by adding more distance constraints derived from both 3D and 4D heteronuclear edited NOESY experiments [23], and by the addition of dihedral angle constraints derived from J couplings measured in 2D and 3D heteronuclear edited experi-

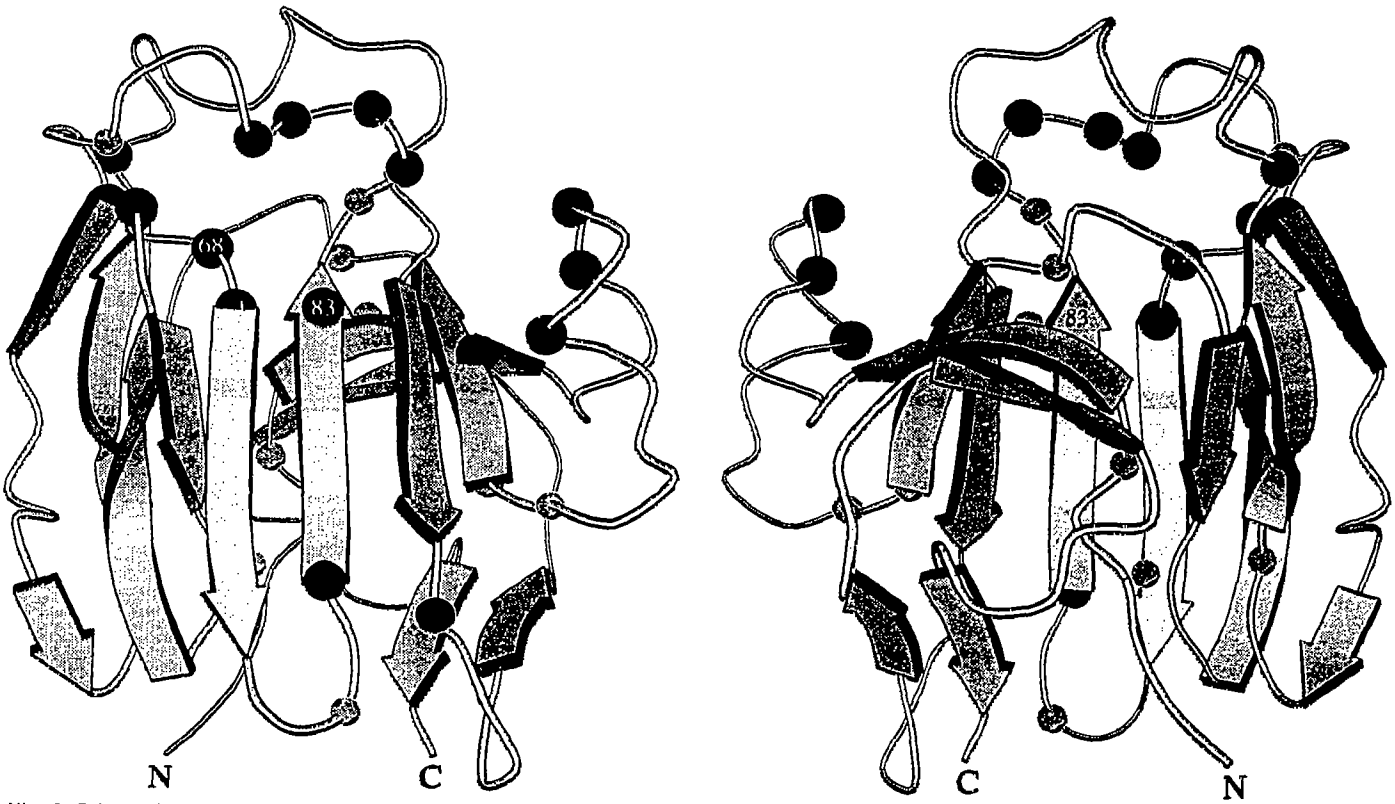

Fig. 2. Schematic representation of the solution structure of the IIA^{glc} domain. The two views shown differ by a 180° rotation. The positions of amino acids conserved in six sequences of PTS IIA domains are indicated by spheres at the C^{α} positions. Conserved glycines and prolines, which may be expected to fulfill a structural role, are represented by small grey spheres. All other totally conserved residues are represented by larger black spheres. The site of phosphorylation, His^{83}, is labelled as is the site of His^{68}, which has been implicated in the phosphoryl transfer mechanism by site-directed mutagenesis experiments. This diagram was produced with the program MOLSCRIPT [33].

ments. A detailed comparison of the solution and crystal structures will be made once the NMR structure refinement has been completed.

Acknowledgements: This work was supported by Grant GM-36643 from the National Institutes of Health (P.E.W.). W.J.F. was supported by a Damon Runyon–Walter Winchell Cancer Research Fund Fellowship, DRG-1059.

REFERENCES

[1] Reizer, J., Saier, M.H., Jr., Deutscher, J., Grenier, F., Thompson, J. and Hengstenberg, W. (1988) CRC Crit. Rev. Microbiol. 15, 297–338.
[2] Saier, M.H. Jr. (1989) Microbiol. Rev. 53, 109–120.
[3] Meadow, N.D., Fox, D.K. and Roseman, S. (1990) Annu. Rev. Biochem. 59, 497–542.
[4] Sutrina, S.L., Reddy, P., Saier, M.H., Jr. and Reizer, J. (1990) J. Biol. Chem. 265, 18581–18589.
[5] Dean, D.A., Reizer, J., Nikaido, H. and Saier, M.H., Jr. (1990) J. Biol. Chem. 265, 21005–21010.
[6] Reizer, J., Sutrina, S.L., Wu, L.-F., Deutscher, J. and Saier, M.H., Jr. (1991) J. Biol. Chem., in press.
[7] Presper, K.A., Wong, C.-Y., Liu, L., Meadow, N.D. and Roseman, S. (1989) Proc. Natl. Acad. Sci. USA 86, 4052–4055.
[8] Fairbrother, W.J., Cavanagh, J., Dyson, H.J., Palmer, A.G., III, Sutrina, S.L., Reizer, J., Saier, M.H., Jr. and Wright, P.E. (1991) Biochemistry 30, 6896–6907.
[9] Pelton, J.G., Torchia, D.A., Meadow, N.D., Wong, C.-Y. and Roseman, S. (1991) Proc. Natl. Acad. Sci. USA 88, 3479–3483.
[10] Kay, L.E., Ikura, M., Tschudin, R. and Bax, A. (1990) J. Magn. Reson. 89, 496–515.
[11] Powers, R., Gronenborn, A.M., Clore, G.M. and Bax, A. (1991) J. Magn. Reson. 94, 209–213.
[12] Palmer, A.G., III, Fairbrother, W.J., Cavanagh, J., Wright, P.E. and Rance, M. (1992) J. Biomol. NMR, submitted.
[13] Kay, L.E., Ikura, M. and Bax, A. (1990) J. Am. Chem. Soc. 112, 888–889.
[14] Bax, A., Clore, G.M., Driscoll, P.C., Gronenborn, A.M., Ikura, M. and Kay, L.E. (1990) J. Magn. Reson. 87, 620–627.
[15] Ikura, M., Kay, L.E. and Bax, A. (1991) J. Biomol. NMR 1, 299–304.
[16] Bax, A., Clore, G.M. and Gronenborn, A.M. (1990) J. Magn. Reson. 88, 425–431.
[17] Fairbrother, W.J., Palmer, A.G., III, Reizer, J., Saier, M.H., Jr. and Wright, P.E. (1992) Biochemistry, submitted.
[18] Wüthrich, K., Billeter, M. and Braun, W. (1983) J. Mol. Biol. 169, 949–961.
[19] Havel, T.F. and Wüthrich, K. (1984) Bull. Math. Biol. 46, 673–698.
[20] Stone, M.J., Fairbrother, W.J., Palmer, A.G., III, Reizer, J., Saier, M.H., Jr. and Wright, P.E. (1992) Biochemistry, submitted.
[21] Lipari, G. and Szabo, A. (1982) J. Am. Chem. Soc. 104, 4546–4559.
[22] Clore, G.M., Driscoll, P.C., Wingfield, P.T. and Gronenborn, A.M. (1990) J. Mol. Biol. 214, 811–817.
[23] Clore, G.M., Wingfield, P.T. and Gronenborn, A.M. (1991) Biochemistry 30, 2315–2323.
[24] Nelson, S.O., Schuitema, A.R.J., Benne, R., van der Ploeg, L.,

Pleiter, J.J., Aan, F. and Postma, P.W. (1984) EMBO J. 3, 1587–1593.
[25] Saffen, D.W., Presper, K.A., Doering, T.L. and Roseman, S. (1987) J. Biol. Chem. 262, 16241–16253.
[26] Schnetz, K., Toloczyki, C. and Rak, B. (1987) J. Bacteriol. 269, 2579–2590.
[27] Bramley, H.F. and Kornberg, H.L. (1987) J. Gen. Microbiol. 133, 563–573.
[28] De Reuse, H. and Danchin, A. (1988) J. Bacteriol. 170, 3827–3837.
[29] Peri, K.G. and Waygood, E.B. (1988) Biochemistry 27, 6054–6061.
[30] Rogers, M.J., Ohgi, T., Plumbridge, J. and Söll, D. (1988) Gene 62, 197–207.
[31] Sato, Y., Poy, F., Jacobson, G.R. and Kuramitsu, H.K. (1989) J. Bacteriol. 171, 263–271.
[32] Liao, D.-I., Kapadia, G., Reddy, P., Saier, M.H., Jr., Reizer, J. and Herzberg, O. (1991) Biochemistry 30, 9583–9594.
[33] Kraulis, P.J. (1991) J. Appl. Cryst., in press.