Solution Structure of the IIB Domain of the Glucose Transporter of *Escherichia coli*†

Matthias Eberstadt,§ Simona Golic Grdadolnik,§,|| Gerd Gemmecker,§ Horst Kessler,*,§ Andreas Buhr,⊥,▽ and Bernhard Erni*,⊥

Institut für Organische Chemie und Biochemie, Technische Universität München, D-85747 Garching, Germany, National Institute of Chemistry, SLO-61115 Ljubljana, Slovenia, and Institut für Biochemie, Universität Bern, CH-3012 Bern, Switzerland

Received February 28, 1996; Revised Manuscript Received June 6, 1996⊗

ABSTRACT: The structure of the IIB^{Glc} domain of the *Escherichia coli* transporter for glucose was determined by multidimensional heteronuclear NMR. The glucose transporter (IICB^{Glc}) belongs to the bacterial phosphotransferase system. It mediates uptake with concomittant phosphorylation of glucose. The N-terminal IIC^{Glc} domain spans the membrane, the C-terminal IIB^{Glc} domain (residues 386−477) contains the phosphorylation site, Cys421. The structure of the subclonal IIB domain was determined based on 927 conformational constraints, including 744 NOE derived upper bounds, 43 constraints of ranges of dihedral angles based on measurements of vicinal coupling constants, and 70 upper and lower bound constraints associated with 35 hydrogen bonds. The distance geometry interpretation of the NMR data is based on the previously published sequence-specific ^{1}H, ^{15}N, and ^{13}C resonance assignments [Golic Grdadolnik et al. (1994) *Eur. J. Biochem. 219,* 945−952]. The sequence of the secondary structure elements of IIB is $\alpha1\beta1\beta2\alpha2\beta3\beta4\alpha3$. The basic fold consists of a split α/β-sandwich composed of an antiparallel sheet with strand order $\beta1\beta2\beta4\beta3$ and three α-helices superimposed onto one side of the sheet. The hydrophobic helix α1 is packed against helices α2, α3, and the β-sheet. The phosphorylation site (Cys421) is at the end of $\beta1$ on the solvent-exposed face of the sheet surrounded by Asp419, Thr423 Arg424, Arg426, and Gln456 which are invariant in 15 homologous IIB domains from other PTS transporters.

The transporters of the bacterial phosphoenolpyruvate dependent phosphotransferase system (PTS)[1] act by a mechanism which couples uptake with phosphorylation of hexoses and hexitols. The transporters comprise three functional units (protein subunits or domains of a multidomain protein), which are termed IIA, IIB, and IIC. IIC is the membrane-spanning unit which contains the sugar binding site. The hydrophilic IIA and IIB domains mediate sequential phosphoryltransfer from the low molecular weight phosphoryl-carrier protein, HPr, to the transported hexose. Phosphoryltransfer proceeds through phosphohistidine and phosphocystein intermediates. There are at least thirteen PTS transporters of different but partly overlapping sugar specificity in *Escherichia coli* alone and many more, if non-enteric and gram positive bacteria are also considered. They have been grouped into four families according to sequence similarity [for reviews see Saier (1989), Meadow et al. (1990), Erni (1992), Postma et al. (1993), and Lengeler et al. (1994)].

The glucose specific transporter of *E. coli* consists of two subunits, IIA^{Glc} and IICB^{Glc}. The IIA subunit is a β-sheet sandwich with six antiparallel strands on either side, as shown by NMR and X-ray studies with IIA^{Glc} of *E. coli* and the homologous IIA^{Glc} domain of the *Bacillus subtilis* transporter (Worthylake et al., 1991; Pelton et al, 1991a,b; Liao et al., 1991; Fairbrother et al., 1991a,b, 1992). IIA^{Glc} of *E. coli* is phosphorylated by HPr on His90. Both the phosphorylated and dephosphorylated forms have been studied by heteronuclear multidimensional NMR (Pelton et al., 1992, 1993). IICB^{Glc} is a homodimer (Erni, 1986; Meins et al., 1988). The IICB^{Glc} subunit consists of two domains. The IIC domain spans the membrane eight times and contains the sugar binding site (Hummel et al., 1992; Buhr & Erni, 1993). The hydrophilic IIB domain mediates phosphoryltransfer from IIA^{Glc} to O-6′ of glucose. In this process IIB^{Glc} becomes transiently phosphorylated at Cys421 (Nuoffer et al., 1988; Meins et al., 1993). The IIB and IIC domains can be functionally expressed as separate polypeptides (Buhr et al., 1994).

The secondary structure of the subclonal monomeric IIB domain has been determined by heteronuclear NMR as $\alpha\beta\beta\alpha\beta\beta\alpha$ with the β-strands forming an antiparallel sheet of strand order $\beta1\beta2\beta4\beta3$ (Golic Grdadolnic et al., 1994). The three-dimensional structure determined by multidimensional NMR and distance geometry calculations is the subject of this report.

† This work was supported by grants from the Swiss National Science Foundation 31-29795.90 (B.E.) and from the Deutsche Forschungsgemeinschaft, the Sonderforschungsbereich 369, the Fonds der Chemischen Industrie, and the Dr.-Ing. Leonhard Lorenz-Stiftung. S.G.G. acknowledges support from the Ministry of Science and Technology of Slovenia.

* Authors to whom correspondence should be addressed.

‡ Atomic coordinates and NMR constraints of an ensemble of solution structures of the IIB^{Glc} domain have been deposited with the Protein Data Bank, Brookhaven National Laboratories, Long Island, NY 11973, under the identification code 1IBA.

§ Technische Universität München.

|| National Institute of Chemistry.

⊥ Universität Bern.

▽ Present address: Institut für Pharmakologie, Universität Bern, CH-3012 Bern, Switzerland.

⊗ Abstract published in *Advance ACS Abstracts,* August 1, 1996.

[1] Abbreviations: PTS, phosphoenolpyruvate−sugar phosphotransferase system; IICB^{Glc}, transmembrane subunit of the glucose transporter of the PTS; IIA^{Glc}, hydrophilic subunit of the glucose transporter; IIC^{Glc}, transmembrane domain of IICB^{Glc}; IIB^{Glc}, hydrophilic domain of IICB^{Glc}; HPr, histidine-containing phosphocarrier protein of the PTS; PEP, phosphoenolpyruvate, RMSD, root mean square deviation.

S0006-2960(96)00492-8 CCC: $12.00

EXPERIMENTAL PROCEDURES

Sample Preparation. The expression and purification by Ni^{2+} chelate affinity chromatography of subclonal ^{15}N- and $^{13}C,^{15}N$-labeled IIB^{Glc} have been described (Buhr et al., 1994; Golic Grdadolnic et al., 1994). For our NMR investigations we used two isotopically labeled protein samples of the IIB domain: an ^{15}N-labeled sample of 1.5 mM and a $^{13}C,^{15}N$-labeled sample of 0.9 mM. Both samples showed severe deterioration (especially line broadening) of the NMR spectra after some days, caused by the oxidative dimerization of IIB^{Glc} via the free sulfhydryl group of Cys35 in solution. However, the effect could be reversed by the addition of the reducing agent dithiothreitol, which was therefore used in excess to keep the sample stable over the measuring period.

NMR Experiments. All NMR measurements were performed at 300 K on a Bruker AMX 600 MHz NMR spectrometer equipped with a *z*-gradient unit. Quadrature detection in the indirectly detected dimensions was accomplished using the States-TPPI method (Marion et al., 1989). NMR spectra were processed using the UXNMR program (Bruker) and the software package SYBYL (TRIPOS Associates, Inc.).

The triple-resonance spectra used for the signal assignment have already been described elsewhere (Golic Grdadolnic et al., 1994). Structural parameters were extracted from 3D ^{15}N- and ^{13}C-edited NOESY-HSQC spectra, an HNHA experiment, and a series of $^{15}N,^{1}H$-COSY spectra, as described in the following.

The 3D ^{15}N-edited NOESY-HSQC experiment (Fesik & Zuiderweg, 1988; Marion et al., 1989) was performed with a mixing time of 50 ms. The data were collected as $48 \times 96 \times 512$ complex points over sweep widths of 1418 Hz (^{15}N, t_1), 7142 Hz (^{1}H, t_2), and 8333 Hz (^{1}H, t_3) in 16 scans per increment.

The 3D ^{13}C-edited HSQC-NOESY experiment (Fesik & Zuiderweg, 1988; Muhandiram et al., 1993) was collected with 50 ms mixing time. *z*-Gradients were used to suppress the residual water signal (Bax & Pochapsky, 1992). The data were collected as $93 \times 48 \times 512$ complex points using sweep widths of 7142 Hz (^{1}H, t_1), 5000 Hz (^{13}C, t_2), and 5208 Hz (^{1}H, t_3) and employing extensive folding of the ^{13}C-signals to increase the digital resolution in the carbon dimension (Schmieder et al., 1991).

$^{3}J_{HNH\alpha}$ coupling constants were obtained from a series of J-modulated $^{15}N,^{1}H$-COSY spectra (Billeter et al., 1992) and an HNHA experiment (Vuister & Bax, 1993). The HNHA experiment was acquired with 32 scans per increment with $32 \times 64 \times 512$ complex points using sweep widths of 1418 Hz (^{15}N, t_1), 7142 Hz (^{1}H, t_2), and 8333 Hz (^{1}H, t_3). Nine *J*-modulated $^{15}N,^{1}H$-COSY experiments (192×512 complex points) were recorded with modulation periods ranging over 10−120 ms. The cross-peak integrals were measured with the integration routine within the UXNMR program (Bruker).

Structure Calculations. For the determination of the solution structure of the IIB^{Glc} domain the distance geometry program DIANA was used, which allows the calculation of protein conformations on the basis of data about interatomic distances and dihedral angles (Güntert et al., 1991). The algorithm is based on the minimization of a variable target function (Braun & Go, 1985), where the degrees of freedom are the dihedral angles about single rotatable bonds of the protein molecule. Starting from random conformations, 60 structures were minimized. The minimization parameters were modified to 22 minimization steps and a maximal number of iterations of 14 000. To reduce the computation time for obtaining a group of acceptable conformers, redundant dihedral angle constraints (REDAC) derived from preliminary calculations were used (Güntert & Wüthrich, 1991). For residues 15−89, 744 distance restraints were derived from ^{15}N- and ^{13}C-resolved NOESY spectra acquired using 50 ms mixing time. To transform signal intensities into upper bound constraints the NOE cross-signals were classified in three classes: strong, medium, and weak, corresponding to distance limits of 2.5, 3.5, and 4.5 Å, respectively. Appropriate pseudoatom corrections were applied to the upper bound distance constraints according to usual rules (Wüthrich, 1986). A total of 35 hydrogen bonds were included for amide protons that did not show any exchange peaks in a 100 ms MEXICO spectrum (Gemmecker et al., 1994) and given bounds of 1.8−2.3 Å (H→O) and 2.7−3.3 Å (N→O). In addition, 43 ϕ angle constraints from $^{3}J_{HNH\alpha}$ coupling constants were included in the structure calculations. ϕ angles corresponding to $^{3}J_{HNH\alpha} \geq 8$ Hz were confined to the range −160° to −80°, and ϕ angles corresponding to 5.5 Hz $\geq {}^{3}J_{HNH\alpha}$ were confined to the range −90° to −40°. 927 constraints were used during structure calculations. From the resulting 60 structures, the ones with the lowest target function values were selected, after exclusion of structures with obvious faults (unusually high numbers of constraint violations, failure to be minimized). The remaining 11 structures were then subjected to ten steps of simplex and 100 steps of Powell minimization without experimental constraints in the Tripos force field of SYBYL to reduce van der Waals violations. The statistical evaluation of the resulting structural ensemble (cf. Table 1)

Table 1: Parameters and Results of the Structure Calculations of the IIBGlc Domain of *E. coli*

parameter		no. of constraints	
NOE		744	
H-bond		35	
dihedral angles		43	
		average	range
DIANA Statistics for 11 Structures (Residues 15−92) before Energy Minimization			
DIANA target function		85.19 Å^2	(77.72−94.38 Å^2)
upper limit violations (>0.20 Å)	number	96	(82−114)
	sum	53.6 Å	(49.2−60.3 Å)
	largest	1.25 Å	(0.98−1.50 Å)
lower limit violations (>0.20 Å)	number	12	(8−16)
	sum	5.9 Å	(4.6−6.9 Å)
	largest	0.48 Å	(0.43−0.56 Å)
van der Waals violations (>0.20 Å)	number	64	(54−80)
	sum	46.3 Å	(40.9−52.9 Å)
	largest	1.86 Å	(1.78−1.95 Å)
torsion angle violations (>5°)	number	2	(1−3)
	sum	30.5°	(23.4−41.5°)
	largest	8.03°	(5.4−11.5°)
Cartesian Coordinate RMSD to Mean Structure from MOLMOL (Koradi, 1996) for 11 Structures after Energy Minimization			
energy (TRIPOS force field)		175.0 ± 30.6 kcal/mol	(144.8−259.0 kcal/mol)
mean backbone RMSD (residues 15−89)		0.91 ± 0.21 Å	(0.45−1.30 Å)
mean heavy atom RMSD (residues 15−89)		1.69 ± 0.34 Å	(0.92−2.39 Å)

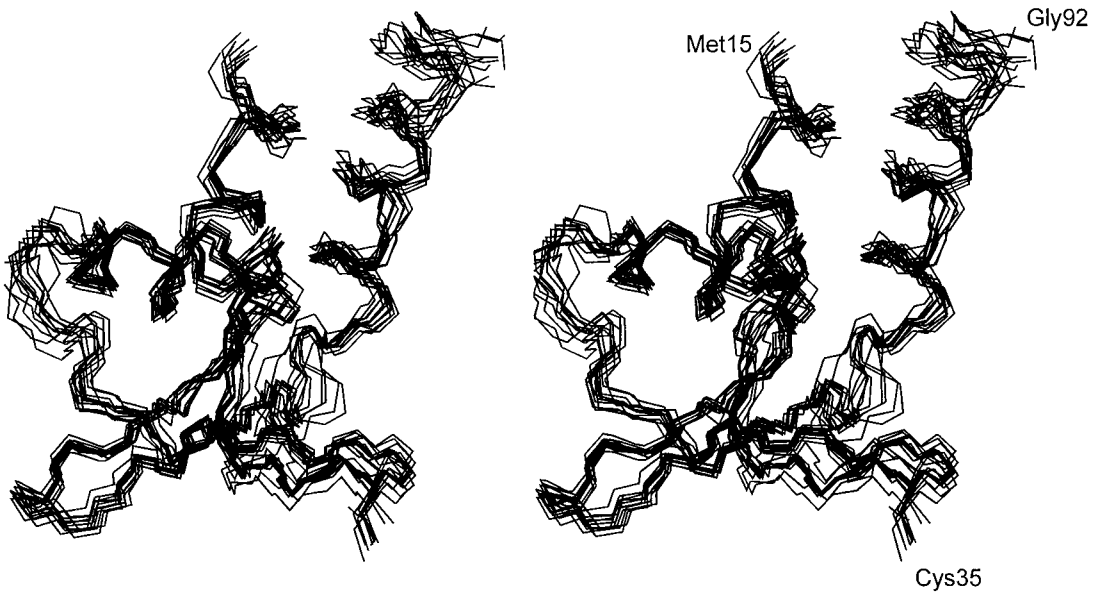

FIGURE 1: Stereoview of the backbone (N, C^{α}, C′) of 11 superimposed NMR-derived structures of the IIB^{Glc} domain of the glucose transporter of *E. coli* (residues 15−92). The side chain of the phosphorylation site Cys35 is labeled.

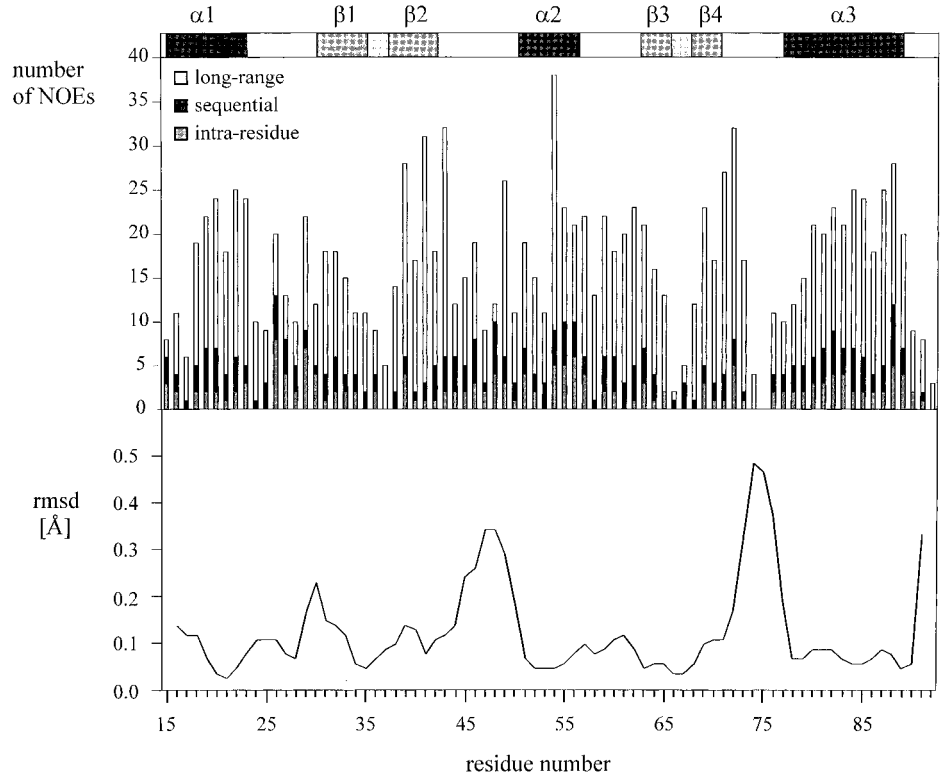

FIGURE 2: Upper panel: distribution of NOE constraints over the IIB^{Glc} protein sequence, intraresidue, sequential, and long-range NOEs are shown in different colors. Lower panel: distribution of backbone coordinate RMSDs over the protein sequence, based on a three-residue fit on the average structure with the program MOLMOL (Koradi et al., 1996). The locations of the secondary structure elements are displayed on top of the figure for reference.

was performed with the program MOLMOL, version 2.2.0 (Koradi, 1996).

RESULTS AND DISCUSSION

Assignments. The ^{1}H, ^{13}C, and ^{15}N spins had been assigned from analysis of a set of heteronuclear multidimensional spectra of uniformly ^{15}N- and ^{15}N,^{13}C-labeled protein (Golic Grdadolnic et al., 1994). The chemical shifts of the ^{1}H, $^{13}C^{\alpha}$, and $^{13}C^{\beta}$ spins were found to be consistent with the secondary structural elements determined from the NOE pattern.

Structure Determination. The three-dimensional structure of the IIB^{Glc} domain was determined by distance geometry calculations using the program DIANA. 744 interproton distance constraints were derived from NOE data (their distribution over the protein sequence is displayed in Figure 2). Hydrogen bonds were only used as constraints for amide protons if their exchange rates were significantly slower than the average and the acceptor carbonyl oxygen atom could be unambiguously identified from NOEs. This is the case for most amide protons within the four-stranded β-sheet and some of the amide protons located in α-helices. Together

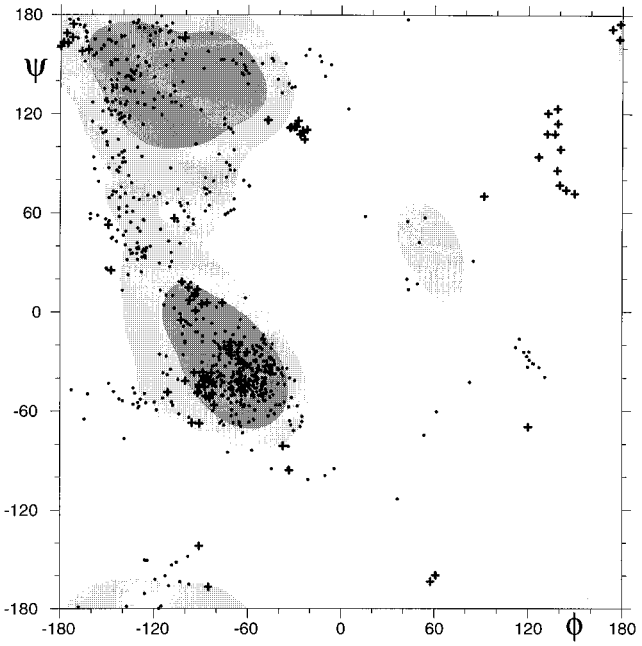

FIGURE 3: Ramachandran plot of the distribution of φ/ψ angles for all 11 structures, generated with the program MOLMOL (Koradi, 1996). Glycine residues are denoted by crosses, all other residues are denoted by dots (proline residues are omitted). Shaded regions correlate with the range containing 80% and 95% of all residues within a database comprising 109 crystal structures with a resolution <2.5 Å.

Table 2: Location of Secondary Structure Elements in the IIB^{Glc} Domain of *E. coli*

α-helices	α1	α2	α3	
residues	18−24	52−58	76−88	
β-strands	β1	β2	β3	β4
residues	30−35	38−43	60−65	68−73

with the dihedral angle constraints from *J* coupling data, this resulted in ca. 11 restraints per residue for the structure calculations, since only 78 amino acids contribute to the structured part of the protein.

After the calculations with NMR restraints and a final minimization run 11 low-energy structures resulted which are shown in Figure 1; structural statistics are given in Table 1. As shown in Figure 2, the structure of the backbone of residues 15−89 is well-defined by the NMR data; only the loop regions Asp45-Val49 and Phe73-Lys76 display an increased RMSD within the ensemble. The first 14 residues at the N-terminus (linker to the N-terminal IIC domain in the natural IIBC protein) and the last 12 residues at the C-terminus (linker and His_6-tag for purification) are omitted in this representation, since these residues are highly flexible and adopt no unique conformation. The atomic root-mean-square deviation (RMSD) about the mean coordinate positions for residues 15−89 is 0.91 ± 0.21 Å for the backbone atoms and 1.69 ± 0.34 Å for all heavy atoms. The Ramachandran plot of the φ and ψ angles for all 11 structures (Figure 3) shows a reasonable concentration within the limits of classical secondary structure elements.

Overall Fold and Secondary Structure. The protein core of IIB^{Glc} comprising 74 residues has a well-defined tertiary structure. The sequence of the secondary structure elements is α1β1β2α2β3β4α3 (Table 2). The basic fold consists of a split α/β sandwich composed of an antiparallel sheet with

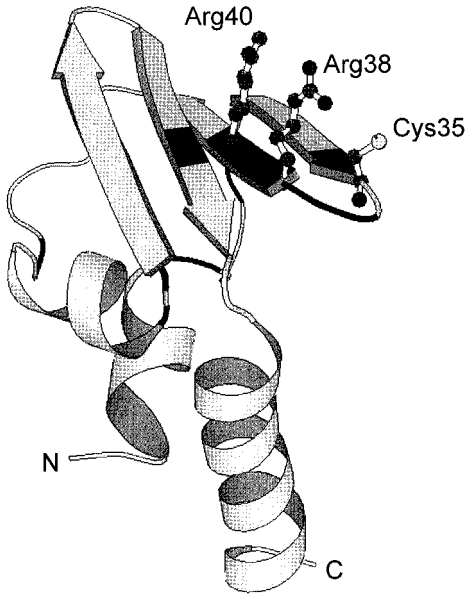

FIGURE 4: Ribbon drawing of the IIB^{Glc} domain. The side chains of Cys35, Arg38, and Arg40 are shown. The positions of invariant residues are indicated in black (cf. Figure 7). The drawing was generated using the program Molscript (Kraulis, 1991).

strand order β1β2β4β3 and three α-helices superimposed onto one side of the sheet (Figures 1 and 4). IIB^{Glc} has approximately the shape of a rectangular pyramid with the β-sheet as base. The strands β1 and β2 have amphiphilic periodicities of 2, characteristic of solvent-exposed strands, while strands β3 and β4 are formed of predominantly hydrophobic residues. Helix α1 is completely hydrophobic, and helices α2 and α3 show an amphiphilic residue distribution. Helix α1 which connects to the transmembrane IIC domain of the transporter is buried in the interior of the protein surrounded by helices α2 and α3 and the β-sheet. The hydrophobic core of the IIB domain is formed by residues L32, A34, I36, I39, V41, and V43 of β1 and β2, residues L19, V20, and F23 of α1, residues V46 and V49 from the loop between β2 and β3, residues L54 and L57 of α2, and residues L80, M84, and I88 of α3 (Figure 5).

As judged by visual inspection the IIB^{Glc} shares structural similarity with the small domain of the arginine repressor ArgR (van Duyne et al., 1996). The two proteins share a common β1β2α1β3β4α2 topology. However, the N-terminal helix α1 of IIB is missing in ArgR and the quaternary structures of IIB and ArgR are also different. IIB is monomeric in solution, while the ArgR domains trimerize to a 12-stranded antiparallel super-barrel, and two such super-barrels further dimerize by tail-to-tail contacts to a hexameric structure. A search through the database of protein structure families with common folding motifs did not produce additional hits (Holm et al., 1991). The β1β2α1β3β4α2 motif of IIB and ArgR appears to be a novel fold.

Structure of the Active Site. The $IICB^{Glc}$-catalyzed transphosphorylation reaction proceeds through a thioester intermediate. The active site cysteine (residue 35 corresponding to Cys421 of $IICB^{Glc}$) is located at the carboxy terminus of β1 on a protrusion formed by the edge of β1 and the reverse turn between β1 and β2. This highly exposed location of the thiol group explains the previously observed rapid formation of disulfide bridges between the subunits in the $IICB^{Glc}$ dimer (Meins et al., 1988). Residues D33 of

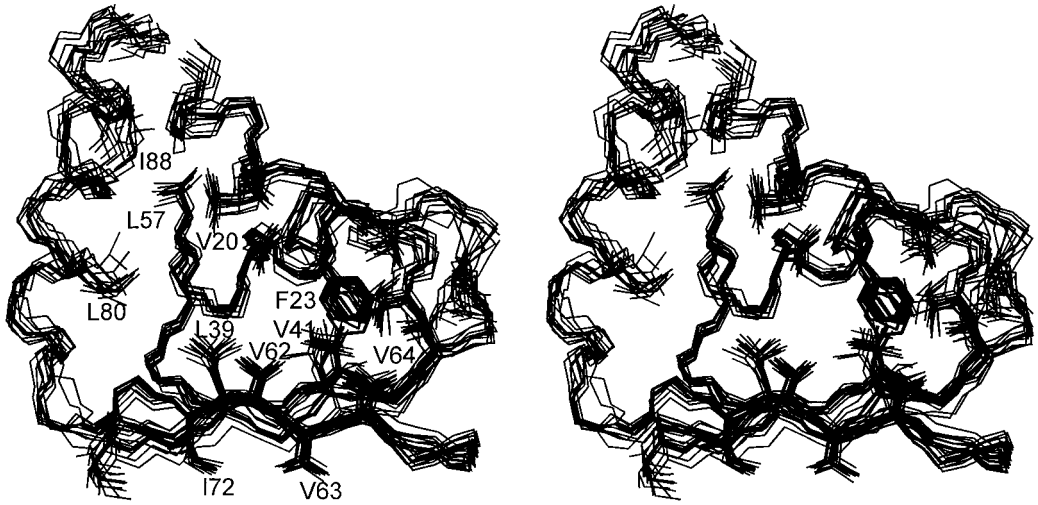

FIGURE 5: Stereoview of the backbone (N, C^{α}, C′) of 11 superimposed NMR-derived structures of the IIB^{Glc} domain. Side chain atoms are shown only for residues that form the hydrophobic core of the protein.

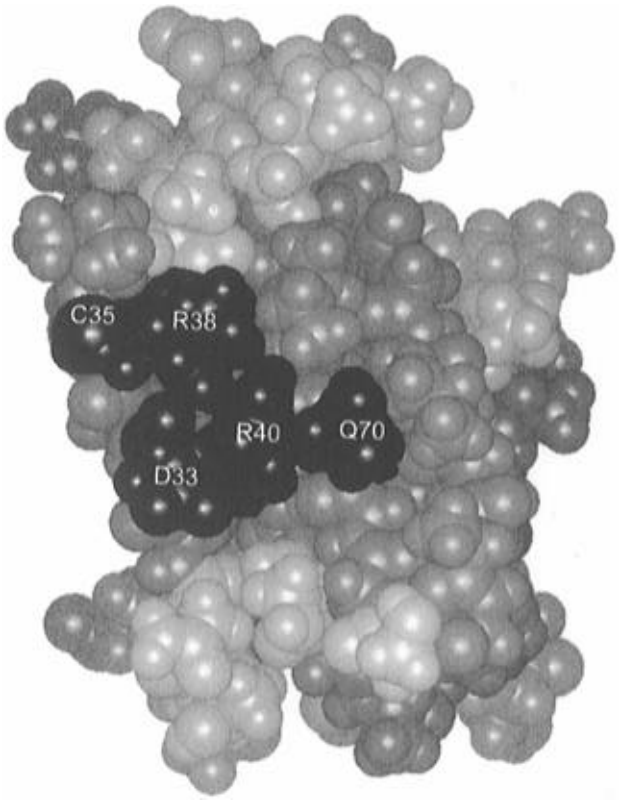

FIGURE 6: CPK model of the IIB^{Glc} domain. The strictly conserved residues D33, Cys35, R38, R40, and Q70 are marked in blue; hydrophobic residues are shown in brown. The conserved amino acids—containing the active site, Cys35—are clustered in the center of a hydrophobic surface area.

β1, R38 and R40 of β2, and Q70 of β4 form a polar semicircle around the active site cysteine. This polar patch on the solvent-exposed surface of the β-sheet is surrounded by a ring of hydrophobic residues (Figure 6).

R38 and R40 might be of particular interest, because their guanidino groups could interact with the phosphoryl group. Site-directed mutagenesis experiments have shown that R38 and R40 are both necessary for phosphoryltransfer from phospho-$IICB^{Glc}$ to glucose but not for the phosphorylation of $IICB^{Glc}$ by phospho-IIA^{Glc} (Lanz and B. Erni, unpublished data). One of the arginines could swing its positively charged guanidinium group close to the active site and thereby stabilize a negatively charged thiolate. One or both arginines might also be critical for stabilizing the negative charge of the bound phosphoryl residue. The guanidino groups could form hydrogen bonds to the phosphate oxygen and, by polarizing the phosphor−oxygen bonds, increase the electrophilicity of the phosphorous atom.

The IIB^{Glc} domain is phosphorylated by the IIA^{Glc} subunit of the glucose transporter. IIA^{Glc} in turn is phosphorylated by the phosphorylcarrier protein phospho-HPr. IIA^{Glc} interacts with both HPr and IIB^{Glc}, and it is likely that the latter two proteins share common features. Worthylake et al. (1991), Liao et al. (1991), and Herzberg (1992) pointed out the strong complementarity between the active sites of IIA^{Glc} and HPr. A similar complementarity with respect to shape and polarity exists between the active sites of IIA^{Glc} and IIB^{Glc}. The phosphorylation site Cys35 of IIB^{Glc}—like the phosphorylation site His15 of HPr—is located on a convex surface, whereas His90 of IIA^{Glc} is located in a shallow depression. In all three proteins, the active site residues are surrounded by a ring of hydrophobic residues which might mediate transient protein−protein contacts during phosphoryl transfer (Worthylake et al., 1991; Liao et al. 1991). There are additional structural similarities between the IIB^{Glc} domain and the phosphorylcarrier protein HPr. Both are open-faced β sandwiches of four antiparallel β-strands and three helices. However, HPr has a different topology ($\beta\alpha\beta\beta\alpha\beta\alpha$, strand order 1423) and the active site His15 is located in the loop between β1 and α1 (Herzberg et al., 1992; van Nuland et al., 1992; Wittekind et al., 1992; Jia et al., 1993a,b; Kalbitzer & Hengstenberg, 1993; Jia et al., 1994). Arg17 of HPr has been shown to play an important role in the active site of HPr where it stabilizes the phosphoryl group on His15 (Jia et al., 1993b; Rajagopal et al., 1994). On the basis of modeling of the complex between HPr and the IIA^{Glc} domain of *B. subtilis*, Arg17 of HPr has been postulated to switch between a salt bridge with the bound phosphate and one with two aspartyl residues associated with the active site of IIA^{Glc} (Herzberg, 1992). These invariant aspartyl residues of IIA^{Glc} could establish similar salt bridges to the conserved R38 and R40 of IIB^{Glc}.

Comparison with Other Phosphocysteine Proteins. Two classes of proteins are known to form phosphocysteine intermediates, the IIB domains of transporters belonging to

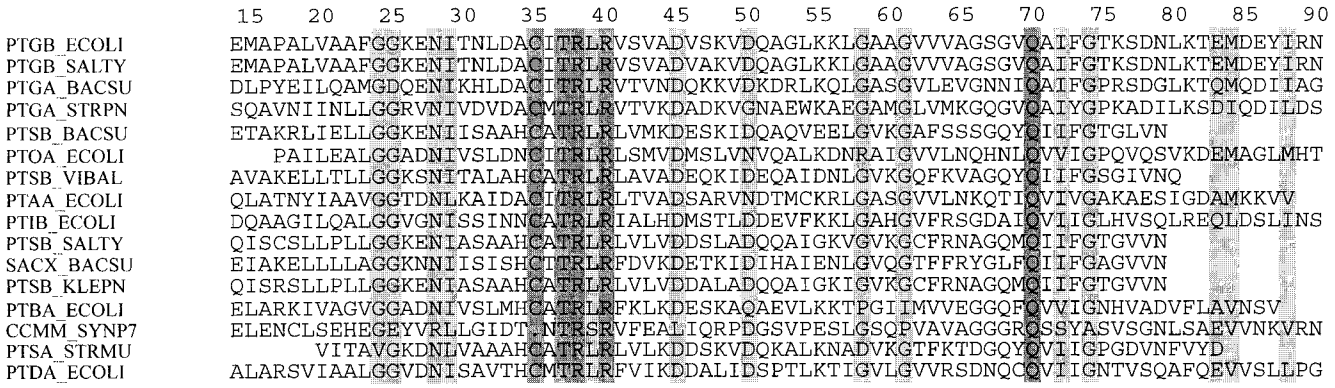

FIGURE 7: Sequence alignment of the homologous IIB domains of PTS transporters belonging to the glucose family, identified by their SWISS-PROT access codes (including two other highly homologous proteins, SACX_BACSU and CCMM_SYNP7). Invariant residues are highlighted by dark background [sequences from EMBL/SWISS-PROT Seqbase, Release 30.0, Sander and Schneider (1991)].

the bacterial phosphotransferase system and the protein tyrosine phosphatases (Stone & Dixon, 1994).

There are 15 known homologous IIB domains from gram negative and gram positive bacteria (Figure 7), which belong to the PTS transporters of the glucose family and share between 35% and 97 % sequence identity [see Sander and Schneider (1991), EMBL/SWISS-PROT Seqbase, Release 30.0]. In addition, they share 14 invariant residues, six of which are located close to the active site (C35, T37, R38, L39, R40, and Q70). Three invariant residues of as yet unknown function (N28, I29, D45) form a cluster located within the adjacent loops $\alpha1/\beta1$ and $\alpha2/\beta2$. Four invariant glycines (G24, G25, G61, G74) are located in turns and loops, and they may play a structural role. In view of the complete invariance of the active site residues, it is almost certain that all these domains have the same fold. However, there are two other families of PTS transporters which also have cysteines in the active site of their IIB domains but do not display amino acid sequence similarity. Of one representative, the IIB subunit of the cellobiose transporter, the secondary structure has been determined by three-dimensional NMR spectroscopy (Ab et al., 1994). IIB^{Cel} consists of a four-stranded parallel β-sheet and five α-helices ($\beta\alpha\beta\alpha\beta\alpha\alpha\alpha$, strand order 2134). It most likely has a doubly wound open α/β structure with the active site Cys10 in the first loop between $\beta1$ and $\alpha1$. It appears that the IIB domains from the different families of PTS transporters differ not only in their amino acid sequence but also in their structure.

Phosphotyrosine phosphatases (PTPases) are the second group of proteins which form phosphocysteine intermediates. They antagonize the activity of phosphotyrosine kinases and thus play a key role in signal transduction. Two types of PTPases exist, the high molecular weight and the low molecular weight PTPases [for a review see Stone and Dixon (1994)]. The X-ray structures of two high molecular weight and one low molecular weight PTPases are known (Barford et al., 1994; Stuckey et al., 1994; Su et al., 1994). The former have a twisted, mixed β-sheet structure flanked by α-helices, whereas the latter have a doubly wound open α/β structure ($\beta\alpha\beta\alpha\beta\alpha\beta\alpha$, strand order 2134). In both forms the catalytic cysteine is located in a loop between a β-strand and an α-helix. The active site sequences are HCXAGXRS/TG for the high molecular weight PTPase and CLGNICR for the low molecular weight PTPase. The common denominator between these and the DXCXTRLR signature sequence of IIB^{Glc} is the arginine residue in position $n+6$ or $n+5$ with respect to the active site cysteine. There is strong evidence from the X-ray structures that the positively charged guanidinium group of the invariant arginine stabilizes not only the negative charge of the bound phosphate but also the unphosphorylated thiolate (pK_a 4.67; Zhang & Dixon, 1993), thereby shifting the thiolate/thiole ratio toward the more nucleophilic ionic species (Barford et al., 1994; Stuckey et al., 1994).

Not only does the structure found for IIB^{Glc} display a novel protein fold, it is also quite different from other known phosphocysteine protein structures with respect to the location of its active site cystein residue. While in the known PTPases and in the IIB^{Cel} protein the cystein is found in a loop region between a β-sheet and an α-helix, Cys35 of IIB^{Glc} is located on a convex surface in the *i* position of a β-turn between strands $\beta1$ and $\beta2$. Nevertheless, the close neighborhood of essential arginine residues seems to be a common feature in these proteins, indicating an important role in the modulation of the nucleophilic activity of the thiolate group.

REFERENCES

Ab, E., Schuurman-Wolters, G. K., Saier, M. H., Reizer, J., Jacuinod, M., Roepstorff, P., Dijkstra, K., & Scheek, R. M. (1994) *Protein Sci. 3*, 282−290.

Barford, D., Flint, A. J., & Tonks, N. K. (1994) *Science 263*, 1397−1404.

Bax, A., & Pochapsky, S. (1992) *J. Magn. Reson. 99*, 638−643.

Billeter, M., Neri, D., Otting, G., Quian, Y. Q., & Wüthrich, K. (1992) *J. Biomol. NMR 2*, 257−274.

Braun, W., & Go, N. (1985) *J. Mol. Biol 186*, 611−626.

Buhr, A., & Erni, B. (1993) *J. Biol. Chem. 268*, 11599−11603.

Buhr, A., Flükiger, K., & Erni, B. (1994) *J. Biol. Chem. 269*, 23437−23443.

Erni, B. (1986) *Biochemistry 25*, 305−312.

Erni, B. (1992) *Int. Rev. Cytol. 137A*, 127−148.

Fairbrother, W. J., Cavanagh, J., Dyson, H. J., Palmer, A. G., Sutrina, S. L., Reizer, J., Saier, M. H., & Wright, P. E. (1991a) *Biochemistry 30*, 6896−6907.

Fairbrother, W. J., Gippert, G. P., Reizer, J., Saier, M. H., & Wright, P. E. (1991b) *FEBS Lett. 296*, 148−152.

Fairbrother, W. J., Palmer, A. G., Rance, M., Reizer, J., Saier, M. H., & Wright, P.E. (1992) *Biochemistry 31*, 4413−4425.

Fesik, S. W., & Zuiderweg, E. R. P. (1988) *J. Magn. Reson. 78*, 588−593.

Gemmecker, G., Jahnke, W., & Kessler, H. (1993) *J. Am. Chem. Soc. 115*, 11620−11621.

Golic Grdadolnik, S., Eberstadt, M., Gemmecker, G., Kessler, H., Buhr, A., & Erni, B. (1994) *Eur. J. Biochem. 219*, 945−952.

Güntert, P., & Wüthrich, K. (1991) *J. Biomol. NMR 1*, 446−456.

Güntert, P., Braun, W., & Wüthrich, K. (1991) *J. Mol. Biol. 217*, 517−530.

Herzberg, O. (1992) *J. Biol. Chem. 267*, 24819−24823.
Holm, L., Ouzounis, C., Tuparev, G., Vriend, G., & Sander, C. (1992) *Protein Sci. 1*, 1691−1698.
Hummel, U., Nuoffer, C., Zanolari, B., & Erni, B. (1992) *Protein Sci. 1*, 356−362.
Jia, Z., Quail, J. W., Waygood, E. B., & Delbaere, L. T. J. (1993a) *J. Biol. Chem. 268*, 22490−22501.
Jia, Z., Vandonselaar, M., Quail, J. W., & Delbaere, L. T. J. (1993b) *Nature 361*, 94−97.
Jia, Z., Vandonselaar, M., Hengstenberg, W., Quail, J. W., & Delbaere, L. T. J. (1994) *J. Mol. Biol. 236*, 1341−1355.
Johnson, L. N., & Barford, D. (1993) *Annu. Rev. Biophys. Biomol. Struct. 22*, 199−232.
Koradi, R., Billeter, M., & Wüthrich, K. (1996) *J. Mol. Graphics 14*, 51−55.
Kraulis, P. (1991) *J. Appl. Crystallogr. 24*, 946−950.
Lengeler, J. W., Jahreis, K., & Wehmeier, U. F. (1994) *Biochim. Biophys. Acta 1188*, 1−28.
Liao, D. I., Kapadia, G., Reddy, P., Saier, M. H., Reizer, J., & Herzberg, O. (1992) *Biochemistry 30*, 9583−9594.
Maas, W. K. (1994) *Microbiol. Rev. 58*, 631−640.
Marion, D., Kay, L. E., Sparks, S. W., Torchia, D. A., & Bax, A. (1989) *J. Am. Chem. Soc. 111*, 1515−1517.
Markovic-Housley, Z., Balbach, J., Stolz, B., & Génovésio-Taverne, J. C. (1994) *FEBS Lett. 340*, 202−206.
Meadow, N. D., Fox, D. K., & Roseman, S. (1990) *Annu. Rev. Biochem. 59*, 497−542.
Meins, M., Zanolari, B., Rosenbusch, J. P., & Erni, B. (1988) *J. Biol. Chem. 263*, 12986−12993.
Meins, M., Jenö, P., Müller, D., Richter, W. J., Rosenbusch, J. P., & Erni, B. (1993) *J. Biol. Chem. 268*, 11604−11609.
Muhandiram, D. R., Farrow, N. A., Xu, G. Y., Smallcombe, S. H., & Kay L. E. (1993) *J. Magn. Reson., Ser. B 102*, 317−321.
Nuoffer, C., Zanolari, B., & Erni, B. (1988) *J. Biol. Chem. 263*, 6647−6655.
Pelton, J. G., Torchia, D. A., Meadow, N. D., Wong, C. Y., & Roseman, S. (1991a) *Proc. Natl. Acad. Sci. U.S.A. 88*, 3479−3483.
Pelton, J. G., Torchia, D. A., Meadow, N. D., Wong, C. Y., & Roseman, S. (1991b) *Biochemistry 30*, 10043−10057.
Pelton, J. G., Torchia, D. A., Meadow, N. D., & Roseman, S. (1992) *Biochemistry 31*, 5215−5224.
Pelton, J. G., Torchia, D. A., Meadow, N. D., & Roseman, S. (1993) *Protein Sci. 2*, 543−558.
Postma, P. W., Lengeler, J. W., & Jacobson, G. R. (1993) *Microbiol. Rev. 57*, 543−594.
Rajagopal, P., Waygood, E. B., & Klevit, R. E. (1994) *Biochemistry 33*, 15271−15282.
Reizer, J., Saier, M. H., Deutscher, J., Grenier, F., Thompson, J., Hengstenberg, W. (1988) *CRC Crit. Rev. Microbiol. 15*, 297−338
Sander, C., & Schneider, R. (1991) *Proteins: Struct., Funct., Genet. 9*, 56−68.
Saier, M. H. (1989) *Microbiol. Rev. 53*, 109−120.
Schmieder, P., Zimmer, S., & Kessler, H. (1991) *Magn. Reson. Chem. 29*, 375−380.
Schnetz, K., Sutrina, S. L. Saier, M. H., & Rak, B. (1990) *J. Biol. Chem. 265*, 13464−13471.
Seip, S., Balbach, J., Behrens, S., Kessler, H., Flükiger, K., de Meyer, R., & Erni, B. (1994) *Biochemistry 33*, 7174−7183.
Stone, R. L., & Dixon, J. E. (1994) *J. Biol. Chem. 269*, 31323−31326.
Stuckey, J. A., Schubert, H. L., Fauman, E. B., Zhang, Z. Y., Dixon, J. E., & Saper, M.A. (1994) *Nature 370*, 571−575.
Su, X. D., Taddei, N., Stefani, M., Ramponi, G., & Nordlund, P. (1994) *Nature 370*, 575−578.
van Duyne, G. D., Ghosh, G., Maas, W. K., & Sigler, P. B. (1996) *J. Mol. Biol. 256*, 377−391.
van Nuland, N. A. J., Grötzinger, J., Dijkstra, K., Scheek, R. M., & Robillard, G. T. (1992) *Eur. J. Biochem. 210*, 881−891.
van Nuland, N. A. J., Kroon, G. J. A., Dijkstra, K., Wolters, G. K., Scheek, R. M., & Robillard, G. T. (1993) *FEBS Lett. 315*, 11−15.
van Nuland, N. A. J., Hangyi, I. W., van Schaik, R. C., Berendsen, H. J. C., van Gunsteren, W. F. G., Scheek, R. M., & Robillard, G. T. (1994) *J. Mol. Biol. 237*, 544−559.
van Nuland, N. A. J., Boelens, R., Scheek, R. M., & Robillard, G. T. (1995) *J. Mol. Biol. 246*, 180−193.
Vogler, A. P., Broekhuizen, C. P., Schuitema, A., Lengeler, J. W., Postma, P. W. (1988) *Mol. Microbiol. 2*, 719−726.
Vuister, G. W., & Bax, A. (1993) *J. Am. Chem. Soc. 115*, 7772−7777.
Worthylake, D., Meadow, N. D., Roseman, S., Liao, D. I., Herzberg, O., & Remington, S. J. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10382−10386.
Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Zhang, Z.-Y., & Dixon, J. E. (1993) *Biochemistry 32*, 9340−9345.

BI960492L