Structural Studies of the *Escherichia coli* Signal Transducing Protein IIA^{Glc}: Implications for Target Recognition[†,‡]

Michael D. Feese,[§,||] Luis Comolli,[§,⊥] Norman D. Meadow,[#] Saul Roseman,[#] and S. James Remington*[,§]

Institute of Molecular Biology and Departments of Physics and Chemistry, University of Oregon, Eugene, Oregon 97403, and McCollum-Pratt Institute, Department of Biology, The Johns Hopkins University, Baltimore, Maryland 21218

Received August 12, 1997; Revised Manuscript Received October 13, 1997[⊗]

ABSTRACT: In *Escherichia coli*, the glucose-specific phosphocarrier protein of the phosphotransferase system (PTS), IIA^{Glc} (III^{Glc} in older literature), is also the central regulatory protein of the PTS. Depending upon its state of phosphorylation, IIA^{Glc} binds to a number of different proteins that display no apparent sequence homology. Previous structural studies suggested that nonspecific hydrophobic interactions, specific salt bridges, and an intermolecular Zn(II) binding site contribute to the wide latitude in IIA^{Glc} binding sites. Two new crystal forms of IIA^{Glc} have been solved at high resolution, and the models were compared to those previously studied. The major intermolecular contacts in the crystals differ in detail, but all involve the hydrophobic active site of IIA^{Glc} interacting with a hydrophobic patch on a neighbor and all are shown to be surprisingly similar to the physiologically relevant regulatory interaction of IIA^{Glc} with glycerol kinase. In two crystal forms, a helix on one molecule interacts with the face of another, while in the other crystal form, the primary crystal contact consists of a strand of β-sheet that contributes to an intermolecular Zn(II) binding site with tetrahedral ligation identical to that of the zinc peptidase thermolysin. Thus, relatively nonspecific hydrophobic interactions combined with specific salt bridges and an intermolecular cation binding site (cation-promoted association) permit a regulatory protein to bind to target proteins that have little or no sequence or structural homology with one another. It is suggested that signal transduction by IIA^{Glc} is a binary switch in which phosphorylation at the active site directly controls binding to target molecules.

Specific macromolecular recognition is central to the function of signal transduction pathways. However, it is often the case that a signal transducing protein recognizes several different targets. In this paper we explore some determinants for target recognition by IIA^{Glc}, the central regulatory protein and glucose-specific phosphocarrier of the *Escherichia coli* phosphoenolpyruvate dependent phosphotransferase system (PTS)[1] (*1, 2*).

IIA^{Glc} (M_r 18 000, also referred to as III^{Glc} in older literature) is required for translocation and phosphorylation of glucose by the following sequence of reactions:

$$\text{P-enolpyruvate} \rightarrow \text{P-EI} \rightarrow \text{P-HPr} \rightarrow \text{P-IIA}^{Glc} \rightarrow \text{P-IICB}^{Glc} \rightarrow \text{P-glucose(in)}$$

EI and HPr are proteins common to all PTS pathways (although a larger variant of HPr is known). The EIIs are carbohydrate-specific permeases that consist of several covalently linked domains or unlinked subunits (A−C). In *E. coli*, the regulatory role of IIA^{Glc} is unique among the IIAs, and the level of phosphorylation of IIA^{Glc} reflects the availability of extracellular glucose. In its unphosphorylated state, IIA^{Glc} inhibits uptake of glycerol and other carbohydrates by binding to the permeases and/or glycerol kinase. IIA^{Glc} has been shown to bind to at least 10 different proteins [reviewed recently by Postma et al. (*2*)]. However, no consensus IIA^{Glc} binding motif can be identified from the amino acid sequences of its targets (*3*).

IIA^{Glc} is an antiparallel β-sandwich (*4*−*6*). Histidines 75 and 90, located off center on one face of the β-sandwich, are required for the phosphoryl transfer reaction. The roles of these two residues have recently been explored in great detail by a variety of techniques (*7*−*9*). His90 is covalently phosphorylated at Nϵ2 by HPr [M_r 9000 (*10, 11*)]. His75 is not essential for phosphoryl transfer from HPr, but substitution by Gln reduces the rate of phosphoryl transfer by a factor of about 200; thus it plays a catalytic role, probably by hydrogen bonding in the transition state (*8, 9*). The histidine residues are mostly buried within a shallow ring of solvent-exposed hydrophobic residues (*4*). These authors suggested that the hydrophobic patch would provide a binding site for regulatory targets, and the major crystal contacts did provide circumstantial evidence in support of this hypothesis.

The crystal structures of IIA^{Glc} in complex with glycerol kinase [one of its regulatory targets (*3, 12*)], are the only experimental models available at present for protein−protein interactions in the PTS. These show IIA^{Glc} bound to a short,

[†] This work was supported by NIH Grants 5R01GM42618 to S.J.R. and GM38759 to S.R. This paper is Publication 1514 from the McCollum-Pratt Institute.

[‡] Protein Data Bank codes: IIA^{Glc} form II, 1F3Z; IIA^{Glc} form III, 2F3G.

* Corresponding author. Phone: (541) 346-5190 (Institute of Molecular Biology). FAX: (541) 346-5870. E-mail: jim@uoxray.uoregon.edu.

[§] University of Oregon.

[||] Present address: Central Laboratories for Key Technology, 1-13-5 Fukuura, Kanazawa, Yokohama 236, Japan.

[⊥] Present address: Graduate Group for Biophysics, University of California at Berkeley, Berkeley, CA 94720.

[#] The Johns Hopkins University.

[⊗] Abstract published in *Advance ACS Abstracts,* December 1, 1997.

[1] Abbreviations: GK, glycerol kinase; HPr, histidine containing protein; NMR, nuclear magnetic resonance; PTS, phosphoenolpyruvate phosphotransferase system; SDS−PAGE, sodium dodecyl sulfate−polyacrylamide gel electrophoresis.

S0006-2960(97)01999-5 CCC: $14.00

hydrophobic solvent-exposed 3_{10} helix ~30 Å from the active site of glycerol kinase, suggesting that long-range or global conformational changes mediate inhibition (*12*). The models suggested the following basis for target recognition by IIA^{Glc}: (i) The large solvent-exposed hydrophobic surface surrounding the active site allows nonspecific binding to a nonpolar target surface. (ii) Salt bridges to conserved residues provide specificity and orientation and are presumably involved in the mechanism of inhibition. (iii) Formation of an intermolecular cation binding site specific for Zn(II) (cation-promoted association) (*3*) provides significant binding energy yet requires the contribution of only a single side chain from the target. Zn(II) ligands are His75, His90, and a water molecule bound to IIA^{Glc}, while glycerol kinase donates a fourth ligand (Glu478) to form tetrahedral ligation identical to that found on the zinc endopeptidase thermolysin.

In order to further explore the determinants for target recognition, we have determined the structures of IIA^{Glc} in two new crystal forms and compared the crystal packings to those of the previously determined model (crystal form I) (*4*). Intermolecular packing in each crystal form provides additional circumstantial evidence in support of several aspects of the protein−protein recognition model summarized above and strengthens the notion that one may often be able to generalize from crystal contacts to protein−protein interactions *in vivo* as other authors have previously suggested (*13, 14*).

MATERIALS AND METHODS

IIA^{Glc}_{fast}, which lacks the seven amino-terminal residues encoded by the *crr* gene, was isolated from an overproducing strain of *E. coli* by a procedure described previously (*4, 6*) and stored in 15 mM Tris-HCl (pH 7.0) and 1 mM NaEDTA (Sigma) at −80 °C. Crystals of IIA^{Glc} were obtained in two new crystal forms (II and III) using the hanging drop vapor diffusion method. IIA^{Glc} in 25 mM Tris-HCl buffer (pH 7.0) was mixed with equal volumes of precipitant solution on silanized cover slips (Fisher Scientific Co.), suspended over wells on tissue culture plates (Linbro) containing 0.5 mL of precipitant solution, and sealed with petroleum grease. The protein concentration was determined with the Bio-Rad protein assay.

Diffraction data were collected with Cu Kα radiation on a Rigaku RU-100 rotating anode X-ray generator equipped with a graphite monochromator and operated at 40 kV, 150 mA. Space group determination was done by precession photography, and three-dimensional data sets were collected with a Rigaku RAXIS IIC imaging plate system or a San Diego Multiwire Detector System Mark III area detector (*15*). The data were reduced using the HKL package (*16, 17*) for RAXIS data or the supplied software for the multiwire detector data (*18*).

The structures were solved by standard molecular replacement procedures. For crystal form II the existing model of IIA^{Glc} from the Brookhaven Protein Data Bank (entry 1F3G) was used for the search model, except that all solvent molecules were removed and lysine and arginine side chains were truncated to their α-carbons. For crystal form III the coordinates for IIA^{Glc} residues 19−168 from the IIA^{Glc}/glycerol kinase regulatory complex (Brookhaven Protein Data Bank entry 1GLC) were used as the search model.

The rotational correlation searches were performed using the programs FRF (unpublished fast rotation function program by Wolfgang Kabsch) and ROTFUN (*19*). A brute-force *R*-factor correlation search program (S. J. Remington, unpublished) was used for translational searches.

Atomic model refinement was performed with the TNT package (*20*). After several cycles of rigid body refinement, further refinement was done using the conjugate direction method (*20*) and correlated temperature factors. Temperature factors were restrained to the range of 6−75 $Å^2$. Electron density maps were examined and model building was performed with FRODO (*21*) and O (*22*).

Solvent-accessible surface areas of intermolecular interfaces were calculated by the method of Lee and Richards (*23*) as implemented in the program EDPDB (*24*). Gap volumes between interfaces were calculated using the program SURFNET (*25*) using gap sphere radii between 1.4 and 4.0 Å (*26*).

RESULTS

Crystal Form II. In order to investigate the ability of Zn(II) to serve as a bridging ligand in intermolecular interactions, crystallization conditions for IIA^{Glc}_{fast} (which lacks seven N-terminal residues of IIA^{Glc}) (*27*) in the presence of 1 mM Zn(II) were sought using a screen approach (*28*). A crystal form was obtained that was not reproducible if Zn(II) was omitted from the precipitant solution.

IIA^{Glc}_{fast} form II crystals were obtained at room temperature from 22% to 26% poly(ethylene glycol) M_r 4000 (SERVA or Hampton Research), 0.1−0.3 M sodium acetate, 1 mM zinc acetate, and 0.1 M Tris-HCl buffer (pH 8.5) with IIA^{Glc}_{fast} at ~6 mg/mL initial concentration in the drop. The crystals reached full size in about 2 weeks, diffracted to ~2.0 Å, and were fairly stable in the X-ray beam. The space group is $P4_32_12$ with $a = 47.7$ Å and $c = 144.6$ Å. The storage buffer for these crystals could be exchanged for storage buffer containing 100 mM Co(II) or Cu(II) without disrupting the crystal. The crystal with Cu(II) diffracted initially to high resolution but decayed very rapidly in the beam; no usable data were obtained. The crystal with Co(II) was stable in the beam.

Rotational and translational correlation searches for crystal form II were performed with data from 8.0 to 4.0 Å and gave a single unambiguous solution. One IIA^{Glc}_{fast} molecule per asymmetric unit was indicated, as the estimated specific volume, V_m (*29*), based on that assumption was 2.3 $Å^3$/Da. For the correct solution, the translational correlation search yielded an overall peak of 15.3σ above the mean and an initial *R* factor of 36.5% (8.0−4.0 Å data) which dropped to 28.8% after five cycles of rigid body refinement. $F_o - F_c$ difference electron density maps calculated during refinement showed a 14σ positive feature which was attributed to a bound Zn(II) atom. The Zn(II) was included in the atomic model with a fixed occupancy of 1.00; the final refined temperature factor was 38.9 $Å^2$. Statistics for the refined model are given in Table 1.

Zn(II) is ligated by the Nε2 atoms of His75 and His90 of one IIA^{Glc}_{fast} molecule and the Oε atoms of Glu148 of a molecule related by a crystallographic 2_1 axis. A fourth ligand is provided by a solvent molecule, presumed to be water or hydroxide (Figure 1). As in the IIA^{Glc}/glycerol kinase complex, this solvent molecule is stabilized by interactions with the amide nitrogens of residues 94−96.

Table 1: Diffraction Data and Atomic Model Statistics

data set	IIA$^{Glc}_{fast}$ + Zn(II)	IIA^{Glc}-(17−168)	IIA$^{Glc}_{fast}$ + Co(II)
crystal form	form II	form III	form II
data collection[a]	RAXIS IIC	SDMS Mark III	RAXIS IIC
no. of crystals	1	1	1
resolution (Å)	20.0−1.98	20.0−2.13	20.0−2.5
completeness[b] (%)	61	86	83
no. of observations	22910	44614	18007
no. of reflections	7578	15616	5226
R_{merge}[c] (%)	4.4	3.6	5.0
av *B* model[d] (Å^2)	41.5	MolA 20.4 MolB 23.5	48.8
R factor[e]	0.193	0.169	0.199
no. of solvents[f]	14	59	12
no. of atoms[g]	1123	2335	1121
RMS deviations[h]			
bonds (Å)	0.010	0.015	0.011
angles (deg)	1.9	2.5	2.0
B factor (Å^2)	5.7	6.3	6.4

[a] Data for form II crystals were collected in 15 frames by consecutive 4° oscillations with an exposure time of 60 min per degree and a detector position of 135 mm, $2\theta = 0°$. Data for form III were collected in 16 sweeps covering 234° using a step size of 0.09°, an exposure time of 60 s per frame, and detector positions of 609.2 mm, $2\theta = -28°$, and 557.6 mm, $2\theta = 16°$. All data sets were collected at ambient temperature. [b] Percentage of theoretically possible reflections observed. [c] Unweighted mean-squared *R* factor: $R_{merge} = \sum(I - \langle I\rangle)/\sum\langle I\rangle$. [d] Average *B* factor for all protein atoms. [e] Standard crystallographic *R* factor $= \sum||F_o| - |F_c||/\sum|F_o|$. [f] Criteria for solvent was a 4σ or larger peak in $(F_o - F_c)$ maps, a possible hydrogen bond partner within 3.0 Å, and a refined *B* factor of less than the restrained maximum. [g] Total number of atoms in the asymmetric unit, including solvents and metal ions. For both crystal forms residues 19−168 were located in the election density. Some atoms of residues 53, 72, 99, 104, 108, 114, 129, 132, and 167 in form II could not be located in the electron density and were omitted from the model. [h] Target RMS values: bonds, 0.02 Å; angles, 3.0°; *B* factors, 6.0 Å^2.

The coordination geometry of the Zn(II) site is a distorted tetrahedron, the same geometry as seen for the coordination of the intermolecular Zn(II) binding site at the regulatory interface of the IIA^{Glc}/glycerol kinase complex (*3*) and for the catalytic Zn(II) of the metalloproteinase thermolysin (*30*, *31*). The Zn(II) and its three liganding protein atoms of the IIA$^{Glc}_{fast}$ model were superimposed with those of a thermolysin structure from the Brookhaven Data Bank (entry 1TLP) with an RMS error of 0.28 Å. A similar overlay for the corresponding atoms of the IIA^{Glc}/glycerol kinase complex gave an RMS error of 0.43 Å. The data set of IIA$^{Glc}_{fast}$ for which the Zn(II) was exchanged for Co(II) was nearly identical to that with Zn(II). The RMS error for the overlay from the metal sites in the IIA$^{Glc}_{fast}$ models with Zn(II) and Co(II) is 0.27 Å whereas for the Cα atoms of the models it is 0.42 Å. The Co(II) was given a fixed occupancy of 1.00; the final refined *B* factor was 61.3 Å^2.

The most extensive crystal contact forms the 2_1 screw axis and includes the intermolecular Zn(II) binding site described above. A β-strand consisting of residues 147−154 of one molecule fits against the hydrophobic patch of another at roughly the same position as the IIA^{Glc} binding helix of glycerol kinase (Figures 2a and 3a). Ile150 and Leu152 of one molecule pack against the hydrophobic patch formed by Phe71, Phe88, Val40, Val46, and Ile45 of another. Lys147 and Glu97 form a salt bridge, and Ser153 hydrogen bonds to the backbone carbonyl of Phe41. A solvent molecule bridges the backbone nitrogen of Leu149 and Nζ of Lys69. Atom to atom distances for this interface are listed in Table 2.

Crystal Form III. Crystals of IIA$^{Glc}_{fast}$ form III were discovered in a hanging drop experiment that had been stored for 2.5 years at 4 °C and have not been reproduced. The crystals were obtained from what was initially 30% poly(ethylene glycol) M_r 4000, 0.2 M $MgCl_2$, and 0.1 M Tris-HCl buffer (pH 8.5) with IIA$^{Glc}_{fast}$ at ∼4 mg/mL concentration in the drop. Crystallization in this form may require limited proteolysis of IIA^{Glc}. Sequence analysis of a dis-

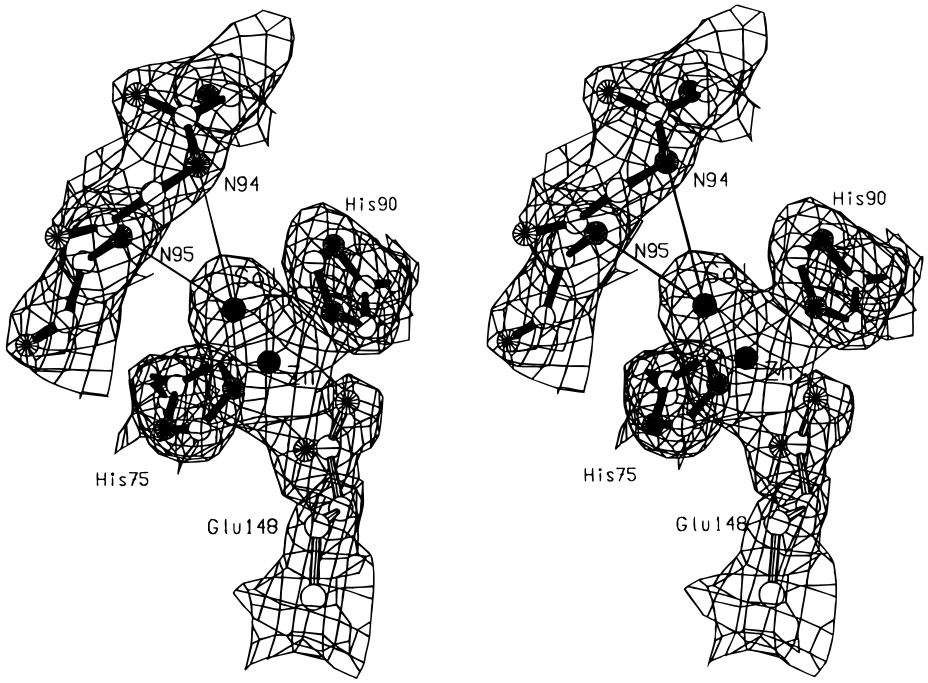

FIGURE 1: Stereo diagram of the intermolecular Zn(II) binding site of crystal form II. $2F_o - F_c$ difference electron density for the final refined model contoured at 1σ. Nitrogen atoms are shown with filled circles, oxygen with spoked circles, and carbon with open circles. Glu148 from one molecule (open bonds) binds to the Zn(II) atom, which in turn is liganded to His75 and His90 of another molecule (filled bonds). A fourth ligand is provided by a water (Sol) that makes hydrogen bonds to the backbone nitrogens of residues 94 and 95.

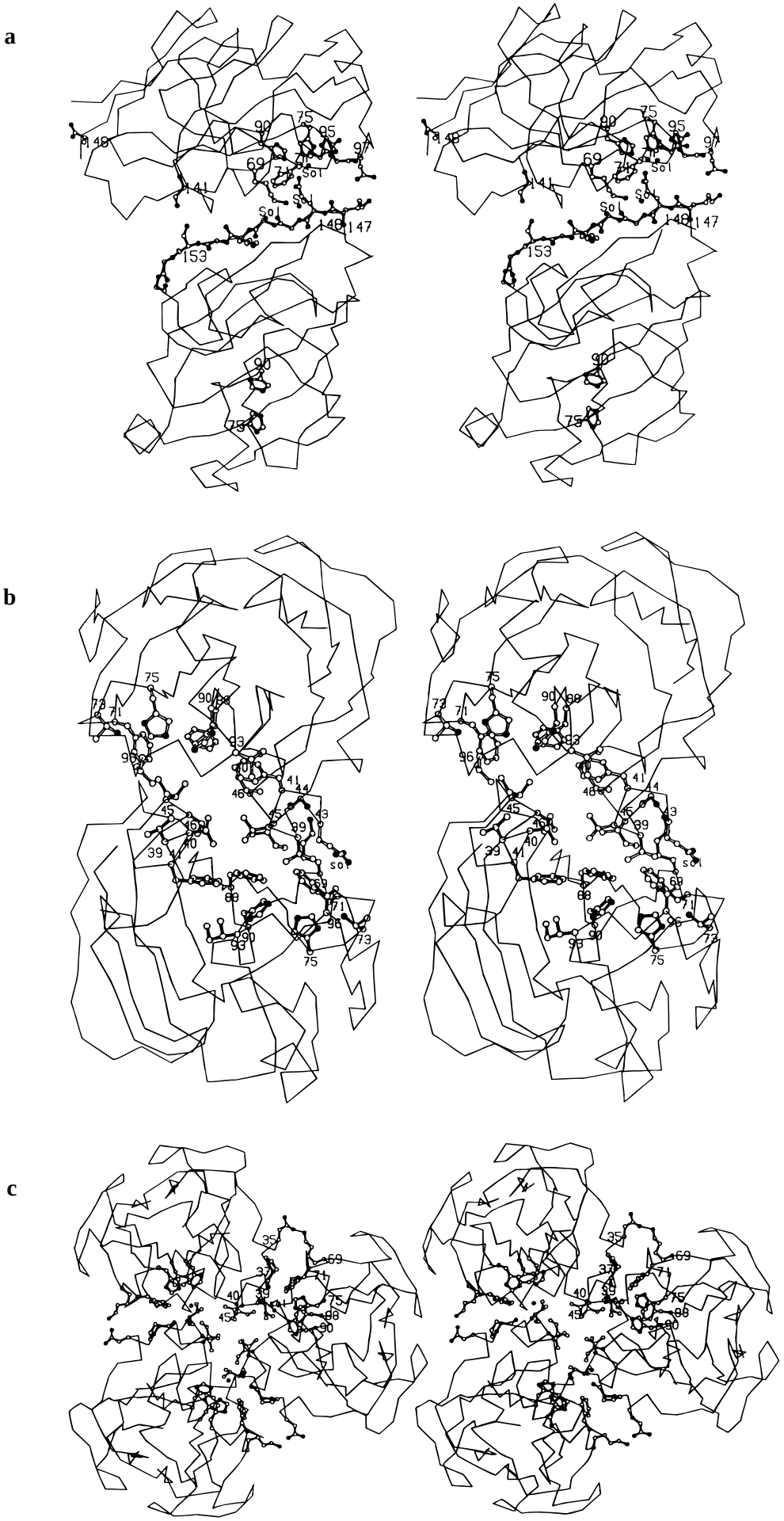
a
b
c

FIGURE 2: (a) Stereoview of the 2_1-related intermolecular contact in crystal form II that forms the Zn(II) binding site. Residues of one molecule are shown in open bonds and the other in filled bonds. This head-to-tail interaction forms continuous polymers through the crystal. (b) Stereo diagram of the IIA^{Glc} dimer that forms the asymmetric unit of crystal form III. The two molecules are related by a 174.9° rotation about an approximate 2-fold axis, forming a dimer with their active sites facing each other. The RMS overlay for their Cα atoms is 0.3 Å. (c) Stereo diagram of the IIA^{Glc} trimer formed by the 3-fold axis of the crystal form in space group *R*3 (form I). Pro37, Val39 and Val40, Ile45, and Phe71 and Phe88 form hydrophobic contacts. Asp35 and Lys69 form an intermolecular salt bridge, and a solvent molecule bridges the interface from the Nε2 of His90 to the backbone nitrogen of Val39.

Table 2: Active Site Contacts for Crystal Form II[a]

reference	neighbor	distance (Å)
Phe71 Cδ2	Glu148 Oε2	3.07
Phe71 Cε2	Glu148 Oε2	3.08
His75 Cε1	Glu148 Oε2	2.99
His75 Nε2	Glu148 Oε2	2.93
His75 Nε2	Zn(II)	2.01
His90 Nε2	Glu148 Oε2	3.41
His90 Nε2	Zn(II)	2.23
Ser141 O	Ser153 Og	3.17
Zn(II)	Glu148 Oε1	2.42
Zn(II)	Glu148 Oε2	2.35
Asp94 Cα	Sol1	3.37
Asp94 N	Sol1	3.39
Thr95 N	Sol1	3.29
Val96 N	Sol1	3.38
Sol1	Glu148 Oε1	3.39
Lys69 Cε	Sol5	3.20
Sol5	Leu149 O	3.15
Sol5	Lys151 Cε	3.37
Lys69 Nζ	Sol8	3.38
Sol8	Leu149 N	3.35

[a] Contacts less than 3.5 Å are listed. Sol refers to a solvent molecule modeled as oxygen.

Table 3: Active Site Contacts for Crystal Form III[a]

reference	neighbor	distance (Å)
Ile A45 Oδ1	Val B96 Cγ1	3.39
Ile A45 Cδ1	Phe B71 Cε1	3.41
Lys A69 Cδ	Ile B45 Cδ1	3.00
Lys A69 Nζ	Lys B44 O	2.99
Lys69 Cβ	Sol33	3.31
Sol33	Glu B43 Oε1	2.92

[a] Contacts less than 3.5 Å are listed. Sol refers to a solvent molecule modeled as oxygen.

solved crystal revealed the amino terminus to be Asp17 of the full-length protein (data not shown), and SDS−PAGE of the same material showed a single band at somewhat smaller M_r than IIA$^{Glc}_{fast}$, suggesting the material was otherwise intact.

The crystals diffracted to 2.0 Å and were stable in the X-ray beam. The space group is $P2_12_12_1$ with $a = 60.04$ Å, $b = 70.23$ Å, $c = 74.31$ Å. Data collection statistics are presented in Table 1. Specific volume calculations suggested that the asymmetric unit of crystal form III contained two IIA^{Glc} molecules ($V_m = 2.5$ Å^3/Da). The first molecule (molecule A) was located with a fast rotation function program using data from 20.0 to 5.0 Å. The translational correlation gave a peak of 6.9σ with data from 6.0 to 4.0 Å. In order to locate molecule B, the calculated Patterson coefficients for molecule A were subtracted from the observed Patterson coefficients during calculation of the rotational correlation, increasing the peak to background ratio. The rotation search for molecule B was performed with data from 12.0 to 4.0 Å. The translational correlation resulted in a peak of 2.9σ. The initial *R* factor for the model containing molecule A and molecule B was 38.8%, which dropped to 32.0% after five cycles of rigid body refinement with data from 20.0 to 5.0 Å. Residues 17 and 18 appear to be disordered and were not visible in the electron density maps. Statistics for the final refined model, for which residues 19−168 are included, are given in Table 1.

The helix comprised of residues 38−45 from each molecule packs against the hydrophobic active site of the other (Figure 2b). This helix occupies essentially the same location as the IIA^{Glc} binding helix in the IIA^{Glc}/glycerol kinase complex; however, the N to C direction of the helix is reversed (Figure 3b,c). It is interesting that the orientation and position of this helix are very similar to that proposed for the interaction of *Bacillus subtilis* HPr with *B. subtilis* IIA^{Glc} on the basis of modeling studies (*32*). In detail, Ile45 of molecule A (MolA) packs against Val96 and Phe71 molecule B (MolB). Phe88 and the aliphatic portion of Lys69 pack against MolB Ile45. Phe71 and Val96 pack against MolB Val39 and the aliphatic portion of Glu43, while the backbone amide of Lys69 forms a hydrogen bond to the backbone carbonyl of MolB Lys44. MolB Lys69, however, adopts a different rotamer and forms a salt bridge with Oε1 of Glu108 of MolA related by a crystallographic 2_1 axis. Atom to atom distances for this interface are listed in Table 3.

DISCUSSION

IIA^{Glc} is unusual among signal transducing proteins in its ability to recognize a wide range of targets with apparently unrelated sequences. Many signal transduction proteins bind a range of sequences that have a recognizable sequence motif [for example, major histocompatibility complexes and src homology domains (*33*−*35*)]. However, IIA^{Glc} binding sites, as identified by mutational analysis or as proposed on the basis of very limited sequence similarity (*36*−*38*), by computer modeling studies (*32*), and by NMR and crystallographic studies (*12*, *39*), do not reveal a consensus target sequence or structural motif. Therefore, the question arises as to what constitutes the determinants of specificity of target recognition by IIA^{Glc}.

Here, we summarize an analysis of intermolecular contacts between IIA^{Glc} and one of its targets (glycerol kinase) and compare these to contacts between symmetry-related molecules in three different crystal forms of IIA^{Glc}. The major intermolecular contacts are strikingly similar to each other, as illustrated by the dramatic color coded side-by-side comparison of Figure 4. In Figure 4, hydrophobic contacts to the active site of IIA^{Glc} are illustrated as white and polar contacts as red, while atoms that make no interaction with a partner are shown in blue.

Hydrophobic Interactions. The hydrophobic patch surrounding the active site of IIA^{Glc} is highly conserved among known IIA domains (*5*) and is comprised of several exposed valine, isoleucine, and phenylalanine residues. It has been proposed to drive relatively nonspecific protein−protein interactions with its targets (*4*). This is supported by NMR studies of the *B. subtilis* IIA^{Glc}/HPr phosphoryl transfer complex (*39*), which necessarily involves interaction with

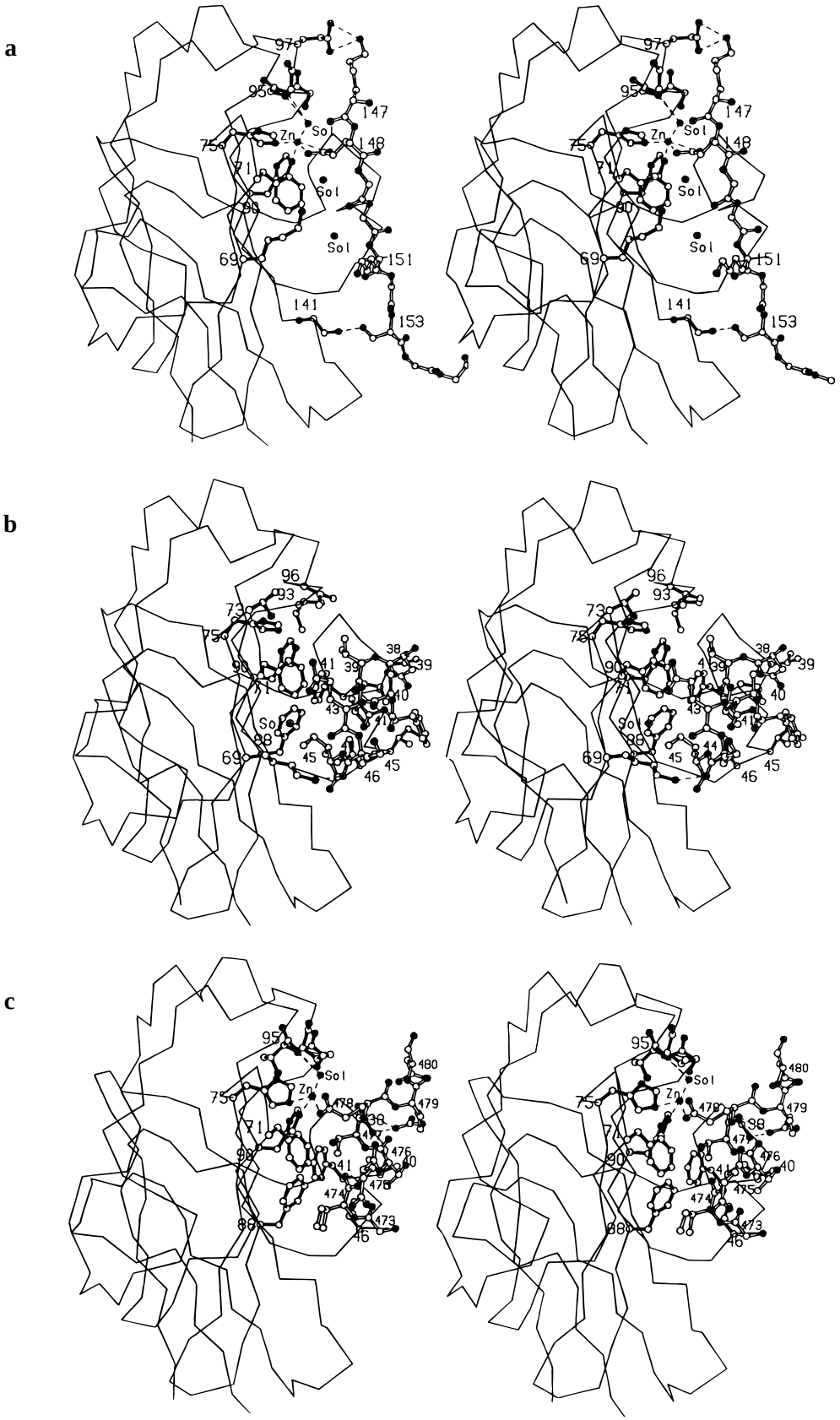

FIGURE 3: (a) Interactions with the hydrophobic patch of IIA^{Glc} in crystal form II. The β-sheet consisting of residues 147−155 is shown in open bonds, and the interacting residues around the active site are shown in filled bonds. Protein ligands that form the Zn(II) binding site are His75 and His90 and glutamate 148 from the 2_1 related molecule. Dashed lines show electrostatic interactions across the interface. (b) Interactions with the hydrophobic patch of IIA^{Glc} in crystal form III. Residues 38−45 of one IIA^{Glc} molecule are shown in open bonds. The intermolecular hydrogen bond between Lys69 and the backbone carbonyl of residue 44 is shown by a thin dashed line. (c) Interactions with the hydrophobic patch of IIA^{Glc} in the IIA^{Glc}/glycerol kinase interface. The 3_{10} helical segment of glycerol kinase consisting of residues 474−480 is shown in open bonds. Protein ligands that form the Zn(II) binding site are His75 and His90 from IIA^{Glc} and glutamate 478 from glycerol kinase. Sol1 binds to Zn(II) and is stabilized by the backbone nitrogens of residues 94−96. Dashed lines show electrostatic interactions at the Zn(II) site and the salt bridge between IIA^{Glc} Asp38 and glycerol kinase Arg479.

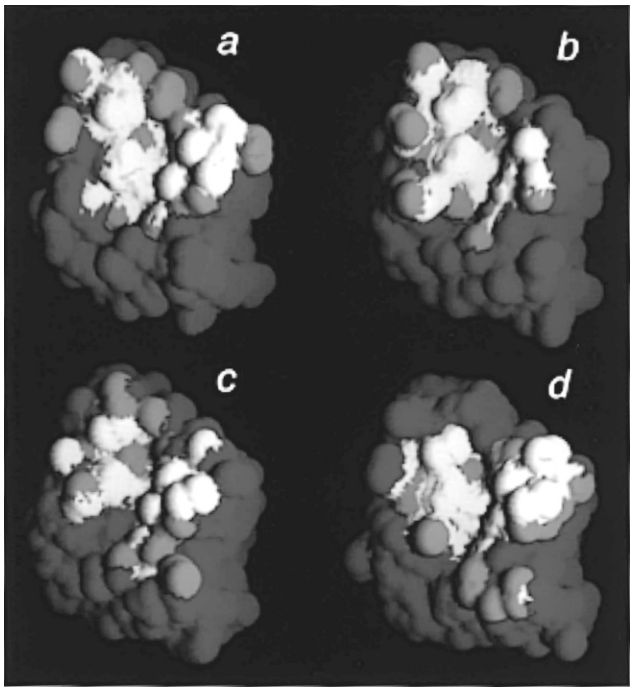

FIGURE 4: View into the active site of IIA^{Glc} in four different crystals in approximately the same orientation, showing the solvent-accessible surface area color coded according to the type of intermolecular contact that IIA^{Glc} makes with a partner; blue, no interaction; white, hydrophobic contact (defined as a carbon atom on IIA^{Glc} that is obscured upon complex formation); red, polar contact (defined as an oxygen or nitrogen atom that is obscured). Panels: (a) the regulatory complex with glycerol kinase; (b) crystal form I; (c) crystal form II; (d) crystal form III.

Table 4: Solvent-Inaccessible Surface Areas and Surface Complementarity for Active Site Interactions of IIA^{Glc}

crystal	space group	hydrophobic surface[a] (Å^2)	polar surface (Å^2)	gap index[b] (Å)
IIA$^{Glc}_{fast}$ + Zn	$P4_32_12$ (form II)	552	340	5.86
IIA^{Glc}(17−168)	$P2_12_12_1$ (form III)	893	322	6.04
IIA$^{Glc}_{fast}$	$R3$ (form I)	783	355	2.94
IIA$^{Glc}_{fast}$ + GK + Zn	$I222$	747	476	4.75
IIA$^{Glc}_{fast}$ + GK	$I222$	771	479	3.60

[a] The program EDPDB (*24*) was used to calculate solvent-accessible surface area (*23*) using a probe radius of 1.4 Å. Polar surface is considered to be nitrogen, oxygen, and sulfur and hydrophobic surface to be carbon. [b] The gap index is the ratio of the interfacial gap volume to the interfacial solvent-accessible surface area of one of the interacting partners (*25*).

the active site face of IIA^{Glc}. Crystallographic studies verified that the hydrophobic patch forms the regulatory interface with glycerol kinase and that phosphorylation at His90 would cause severe steric and electrostatic incompatibilities at the interface (*12*).

Furthermore, IIA^{Glc} aggregates in solution; dimeric, trimeric, and hexameric forms of IIA^{Glc} have been observed in gel filtration and immunoelectrophoretic studies (*40*). The dimeric and trimeric crystal contact interactions shown in Figure 2b,c are likely to represent the states observed in these studies. As shown in Table 4 and Figures 3 and 4, the hydrophobic patch makes major contacts in all crystal forms containing IIA^{Glc}. An average of 750 Å^2 of hydrophobic and 400 Å^2 of polar surface is buried at each of these interfaces. The hydrophobic patch presents ~570 Å^2 of solvent-accessible surface area (*23*).

Our crystallographic studies as well as a molecular modeling study of the IIA^{Glc}−HPr interaction (*32*) show that the hydrophobic patch of IIA^{Glc} is capable of making fairly complementary interactions with a variety of target sites of different sequence, orientation, and secondary structures. Close contacts between hydrophobic residues are relatively few (Tables 2 and 3), but the solvent-inaccessible surface area is fairly large (Table 4). Since hydrophobic binding is driven by the entropy of release of water molecules from hydrophobic contacts (*41*), the complementarity of the interacting surfaces need only be close enough to exclude solvent from the interface in order to contribute significantly to the free energy of binding.

The protein−protein contacts in crystal forms II and III described here, as well as in the previously determined form I crystal structure of IIA^{Glc} (*4*), are strikingly similar in many ways to the IIA^{Glc}/glycerol kinase regulatory interface, yet the target sites differ in secondary structure and/or orientation. The primary interactions appear to be driven by hydrophobic association, which is consistent with the notion that molecular recognition is not, in this instance, dictated by a given sequence or structural motif.

Hydrogen Bonds and Interfacial Solvent. The absence of intermolecular hydrogen bonds is one of the most striking features of the IIA^{Glc}/glycerol kinase regulatory interface (*12*), and this is generally true of the three different crystal packings of IIA^{Glc} discussed here. Molecular recognition in antigen−antibody and protease−inhibitor complexes generally involves 10 or more hydrogen bonds (*42*). Similarly, the major histocompatibility complexes (MHC) class I and class II bind a wide range of peptide sequences with a network of 12 conserved hydrogen bonds to the backbone of the peptide (*35*, *43*). Eliminating the vectorial restraint of donor and acceptor orientations in hydrogen-bonding networks is one possible key to relaxing the requirement for a consensus recognition motif.

Similarly, water molecules have been shown to be an important aspect of antibody−protein antigen binding sites and often fill cavities at the interface and the periphery, bridging polar groups on the antibody and antigen by hydrogen bonds and are suggested to contribute a substantial amount of binding energy [reviewed by Davies and Cohen (*44*); see also ref *45*]. However, a striking aspect of all of the IIA^{Glc} contacts, regardless of whether heteroprotein or homoprotein, is the *lack* of bound water molecules. In all the structures, only one water molecule makes a bridging interaction (Sol8 of crystal form II, Table 2), and there are never more than three found close to the interface in any of the structures. This underscores the hydrophobic nature of the interactions and provides a partial explanation for the general weakness of binding of IIA^{Glc} to its regulatory counterparts [e.g., K_i = 0.28 μM for glycerol kinase in the presence of 0.1 mM Zn(II)] (*3*).

Salt Bridges. Salt bridges are less constrained than hydrogen bonds and may compensate the lack of hydrogen bonding at the IIA^{Glc}/target interface. There are two aspartate residues (Asp38 and Asp94) at the edge of the IIA^{Glc} active site, conserved in nearly all of the 20-odd sequences of proteins homologous to IIA^{Glc} presently in the SwissProt data bank, and one (Asp38) does make a salt bridge to Arg479 in glycerol kinase (*12*). Arg479 has been shown by mutational analysis to be essential for regulation of activity of glycerol kinase by IIA^{Glc} (D. Pettigrew, personal com-

munication). A model building study of the *B. subtilis* IIA^{Glc}/HPr interaction predicted that Arg17 of HPr would interact with the residue corresponding to Asp94 of *E. coli* IIA^{Glc} (*32*), and Arg17 is important for phosphoryl transfer (*46*). Mutational analysis of the melibiose carrier of *Salmonella typhimurium* identified Arg441 as a critical residue for inhibition by IIA^{Glc} (*38*), and it seems likely that it interacts with one of the conserved aspartates.

In the new structures described here, there are no salt bridges with Asp38 and Asp94, which is not surprising as these complexes probably represent nonphysiological binding interactions. A possible exception concerns ligation of the intermolecular zinc in crystal form II, discussed in a following section.

Surface Complementarity. One indication of the strength of interaction of two macromolecules is the degree to which the mating surfaces are complementary in shape and charge. The "gap index", or ratio of the gap volume between two surfaces to the solvent-accessible surface area of the interface (of one partner), has been introduced as a means of quantifying this idea. As recently reviewed by Jones and Thornton (*47*), heteroprotein interactions tend to show a larger gap index than homoprotein interactions, which is not unexpected. Homoprotein contacts *in vivo* are usually between symmetric oligomers, are highly specific, and usually need to be tight and semipermanent (*26*), or they may form continuous filaments. Heteroprotein interactions, on the other hand, are more likely to be transient complexes that assemble for a specific purpose, as is the case for regulatory interactions with IIA^{Glc}. The crystal contacts discussed here are primarily homoprotein interactions; however, the gap indices calculated are quite high (Table 4, mean 4.60, maximum 6.04 Å) and more like heteroprotein interactions (mean 2.48, maximum 5.17 Å) than homodimer interactions (mean 2.20, maximum 4.43 Å) (*47*). Presumably the physiologically relevant intermolecular regulatory interactions are tighter as suggested by the lower gap index for the complex with glycerol kinase (Table 4), although this is still high and suggests a further explanation for the relatively weak binding of IIA^{Glc} to glycerol kinase.

Cation-Promoted Association. Cation-promoted association may also be a general feature of IIA^{Glc} target recognition. Since metals can be tightly bound to only a few side chains contributed by a variety of secondary structures, an intermolecular metal ion binding site can contribute significantly to the free energy of association without constraining the target sequence. We have shown that the IIA^{Glc}/glycerol kinase regulatory interface forms an intermolecular Zn(II) binding site that significantly enhances inhibition [0.1 mM Zn(II) decreases the apparent K_i by 60-fold (*3*)] but requires only one ligand (glutamate 478) from glycerol kinase. The Zn(II) is inaccessible to solvent, and its estimated dissociation constant is 10 nM (Pettigrew et al., submitted for publication). The IIA^{Glc}_{fast} + Zn(II) structure presented here contains an analogous intermolecular Zn(II) site. As in the IIA^{Glc}/glycerol kinase complex, the site is geometrically identical to that of the zinc peptidase thermolysin and the "target", a β-strand of another IIA^{Glc} molecule, contributes Glu148 to the tetrahedrally coordinated Zn(II). Zn(II) has been shown to promote self-association of IIA^{Glc} in solution (D. Torchia, personal communication), but the dissociation constant has not been determined. The dissociation constant of Zn(II) from IIA^{Glc}, however, has been estimated from kinetic analysis to be about 1 μM (Pettigrew et al., submitted for publication). The tight binding of this complex can be explained on the basis of the "chelate effect"; this phrase denotes the entropic advantage of polydentate ligands over monodentate ligands in forming complex ions. The active site histidines of IIA^{Glc} are rigidly held by its molecular architecture in ideal orientation to bind divalent transition metal cations. Additional stabilization is provided by the interaction of a liganding water molecule with several backbone nitrogens. Thus IIA^{Glc} essentially provides three of the ligands and only the fourth need be contributed by the target sequence. Glutamate, aspartate, histidine, and cysteine are potential target ligands that could interact strongly with Zn(II).

Although the biochemical effect of Zn(II) and other metal cations on inhibition of glycerol kinase by IIA^{Glc} has been demonstrated, a physiological role for zinc ions in PTS protein−protein interactions *in vivo* remains to be established. However, there are several other examples of the participation of Zn(II) in molecular recognition. The binding of human growth hormone (hGH) to the prolactin receptor ($hPRL_R$) is increased 8000-fold by addition of 50 μM $ZnCl_2$ (*48*). Crystallographic studies of the complex identified an intermolecularly bound Zn(II) coordinated by two histidines, a glutamate, and an aspartate (*49*). Also, the binding of the superantigen staphylococcal enterotoxin A to MHC class II binding sites has been shown to be Zn(II) dependent (*50*). Mutation of histidine 225 or aspartate 227 of the enterotoxin results in a 1000-fold decrease in affinity (*51*). A conserved histidine of the MHC has also been identified as a probable Zn(II) ligand (*52*). In both of these cases, as seen for the IIA^{Glc}/glycerol kinase complex (Pettigrew et al., submitted for publication), the enhancement is quite specific for Zn(II). The specificity for Zn(II) may result from its lack of ligand field stabilization which allows it to easily change from an octahedral ligand field in solution to the tetrahedral ligand field of the protein binding site (*53*). The Co(II)-containing form II crystal shows coordination geometry essentially identical to Zn(II) but is 170-fold less effective in enhancing inhibition of glycerol kinase. Thus the coordination properties of Zn(II) seem to be ideally suited for participation in transient ternary complexes with macromolecules.

Comparison with Other Signal Transducing Proteins. It is interesting to compare target recognition by IIA^{Glc} with that of another fairly well characterized system, the so-called two component response−regulator systems as best characterized in bacterial chemotaxis [for recent reviews see Stock et al. (*54*) and Volz (*55*)]. Target recognition and signal transduction by phosphorylatable response−regulators such as CheY, the signal transducing protein of chemotaxis, stand in contrast to the principles outlined here for the PTS for at least two reasons. First, CheY has been suggested to undergo conformational changes at sites distant to the active site upon phosphorylation at the active site Asp57 (*56*), which control its interaction with the flagellar motor assembly and ultimately the direction of rotation of the flagella. However, NMR studies of IIA^{Glc} reveal only small structural changes upon phosphorylation (*57*), strongly suggesting that the phosphorylation state of IIA^{Glc} is sensed by direct readout of a target at the active site of IIA^{Glc}.

Second, CheY is a member of a large family of proteins, involved in as many as 70 signal transduction pathways, with

at least 18 of these pathways, and more likely about 40 in a single *E. coli* (*55*, *58*). Interaction of CheY with the autophosphorylating kinase (CheA) and signal transduction target, the flagellar motor, must be limited to the signal pathway in question in order to minimize cross-talk between pathways. It is important to note that at least part of the binding site of CheA to CheY is distinct from the active site (*56*, *59*). However, the PTS appears to be a unique transport and signaling pathway in bacteria; therefore, cross-talk would not seem to be a problem. Again, this indirectly supports the idea that direct readout of the phosphorylation state of IIA^{Glc} by the target is the primary means of signal transduction. In the absence of high-resolution crystal structures of response−regulators with their cognate targets, whether the results of the present study can be generalized to other signal transduction pathways remains to be determined.

Generality of This Approach. One of the referees of this paper raised a number of thoughtful and helpful questions, for example, whether it might be more generally the case that contacts between neighboring molecules in a crystal can be informative concerning protein−protein interactions *in vivo*, as has been argued here. Two recent studies for which this was the case will be briefly discussed. Bullock et al. (*13*) proposed a plausible model for the motile filaments formed by the major sperm protein of the nematode *Ascaris suum* based on the observed crystallographic 2-fold axis relating neighboring dimeric molecules, combined with lattice translations in the crystal. The 2-fold axis generates a tightly organized eight-stranded β-sheet extending across two molecules, with interfacial accessible surface area comparable to the examples described here and nearly equal in area to the dimeric interface between the two monomers that form the asymmetric unit of that crystal. In another study, Shapiro et al. (*14*) described a structural analysis of the N-terminal domain of N-cadherin, which mediates cell−cell adhesion. In three different crystal lattices, a repeated set of dimer interfaces was found to generate a linear "cell adhesion zipper" that mirrors the observed dimensions of the intercellular filaments formed by these molecules, hence providing an atomic model for these interactions.

Thus, although crystallographers and other structural biologists often dismiss crystal contacts as irrelevant to *in vivo* protein−protein interactions, we suggest that there may often be more information in crystal interactions than one might think. Indeed, given the large number of crystal structures that have been solved, it seems possible that much biologically relevant information has actually been overlooked.

ACKNOWLEDGMENT

We thank Karen Kallio for expert technical assistance and Deborah McMillan for peptide sequencing of form III crystals.

REFERENCES

1. Meadow, N. D., Fox, D. K., and Roseman, S. (1990) *Annu. Rev. Biochem. 59*, 497.
2. Postma, P. W., Lengeler, J. W., and Jacobson, G. R. (1993) *Microbiol. Rev. 57*, 543−594.
3. Feese, M., Pettigrew, D. W., Meadow, N. D., Roseman, S., and Remington, S. J. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 3544−3548.
4. Worthylake, D., Meadow, N. D., Roseman, S., Liao, D.-I., Herzberg, O., and Remington, S. J. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 10382−10386.
5. Liao, D.-I., Kapadia, G., Reddy, P., Saier, M. H., Jr., Reizer, J., and Herzberg, O. (1991) *Biochemistry 30*, 9583−9594.
6. Pelton, J. G., Torchia, D. A., Meadow, N. D., Wong, C.-Y., and Roseman, S. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 3479.
7. Pelton, J. G., Torchia, D. A., Meadow, N. D., and Roseman, S. (1993) *Protein Sci. 2*, 543−558.
8. Pelton, J. G., Torchia, D. A., Remington, S. J., Murphy, K. P., Meadow, N. D., and Roseman, S. (1996) *J. Biol. Chem. 271*, 33445−33456.
9. Meadow, N. D., and Roseman, S. (1996) *J. Biol. Chem. 271*, 33440−33445.
10. Dorschug, M., Frank, R., Kalbitzer, R. H., Hengstenberg, W., and Deutscher, J. (1984) *Eur. J. Biochem. 144*, 113−119.
11. Meadow, N. D., and Roseman, S. (1982) *J. Biol. Chem. 257*, 14526−14537.
12. Hurley, J. H., Faber, H. R., Worthylake, D., Meadow, N. D., Roseman, S., Pettigrew, D. W., and Remington, S. J. (1993) *Science 259*, 673−677.
13. Bullock, T. L., Roberts, T. M. and Stewart, M. (1996) *J. Mol. Biol.* 263, 284−296.
14. Shapiro, L., Fannon, A. M., Kwong, P. D., Thompson, A., Lehmann, M. S., Gruebel, G., Legrand, J. F., Als-Nielsen, J., Colman, D. R., and Hendrickson, W. A. (1995) *Nature 374*, 327−337.
15. Hamlin, R. (1985) *Methods Enzymol. 114*, 416−452.
16. Minor, W. (1993) *XDisplayF computer program*, Purdue University, West Lafayette, IN.
17. Otwinowski, Z. (1993) in *Proceedings of the CCP4 Study Weekend* (Sawyer, L., Isaacs, N., and Bailey, S., Eds.) pp 56−63, SERC Daresbury Laboratory, Warrington, England.
18. Howard, A. J., Nielsen, C., and Xuong, N. H. (1985) *Methods Enzymol. 114*, 452−471.
19. Zhang, X.-J., and Matthews, B. W. (1994) *Acta Crystallogr. D50*, 675−686.
20. Tronrud, D. E., Ten Eyck, L. F., and Matthews, B. W. (1987) *Acta Crystallogr. A43*, 489−503.
21. Jones, T. A. (1985) *Methods Enzymol. 115*, 157−171.
22. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard, M. (1991) *Acta Crystallogr. A47*, 110−119.
23. Lee, B., and Richards, F. M. (1971) *J. Mol. Biol. 55*, 379−400.
24. Zhang, X.-J. (1992) Ph.D. Thesis, University of Oregon, Eugene, OR.
25. Laskowski, R. A. (1991) *SURFNET computer program*, University College, London, England.
26. Jones, S., and Thornton, J. M. (1995) *Prog. Biophys. Mol. Biol. 63*, 31−65.
27. Meadow, N. D., Coyle, P., Komoriya, A., Anfinsen, C. B., and Roseman, S. (1986) *J. Biol. Chem. 261*, 13504−13508.
28. Jancarik, J., and Kim, S.-H. (1991) *J. Appl. Crystallogr. 24*, 409−411.
29. Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491−497.
30. Colman, P. M., Jansonius, J. N., and Matthews, B. W. (1972) *J. Mol. Biol. 70*, 701−724.
31. Holmes, M. A., and Matthews, B. W. (1982) *J. Mol. Biol. 160*, 623−629.
32. Herzberg, O. (1992) *J. Biol. Chem. 267*, 24819−24823.
33. Feng, S., Chen, J. K., Yu, H., Simon, J., and Schreiber, S. L. (1994) *Science 266*, 1241−1247.
34. Musacchio, A., Saraste, M., and Wilmanns, M. (1994) *Nat. Struct. Biol. 1*, 546−551.
35. Stern, L. J., Brown, J. H., Jardetzky, T. S., Gorga, J. C., Urban, R. G., Strominger, J. L., and Wiley, D. C. (1994) *Nature 368*, 215−221.
36. Dean, D. A., Reizer, J., Nikaido, H., and Saier, M. H., Jr. (1990) *J. Biol. Chem. 265*, 21005−21010.
37. Kuehnau, S., Reyes, M., Sievertsen, A., Shuman, H. A., and Boos, W. (1991) *J. Bacteriol. 173*, 2180−2186.
38. Kuroda, M., de Waard, S., Mizushima, K., Tsuda, M., Postma, P., and Tsuchiya, T. (1992) *J. Biol. Chem. 267*, 18336−18341.

39. Chen, Y., Reizer, J., Saier, M. H., Jr., Fairbrother, W., and Wright, P. E. (1993) *Biochemistry 32*, 32−37.
40. Scholte, B. J., Schuitema, A. R. J., and Postma, P. W. (1982) *J. Bacteriol. 149*, 576−586.
41. Kauzmann, W. (1959) *Adv. Protein Chem. 16*, 1−63.
42. Janin, J., and Chothia, C. (1990) *J. Biol. Chem. 265*, 16027−16032.
43. Fremont, D. H., Stura, E. A., Matsumura, M., Peterson, P. A., and Wilson, I. A. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 2479−2483.
44. Davies, D. R., and Cohen, G. H. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 7−12.
45. Covell, D. G., and Wallqvist, A. (1997) *J. Mol. Biol. 269*, 281−297.
46. Anderson, J. W., Pullen, K., Georges, F., Klevit, R. E., and Waygood, E. B. (1993) *J. Biol. Chem. 268*, 12325−12333.
47. Jones, S., and Thornton, J. M. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 13−20.
48. Cunningham, B. C., Bass, S., Fuh, G., and Wells, J. A. (1990) *Science 250*, 1709−1712.
49. Somers, W., Ultsch, M., De Vos, A. M., and Kossiakoff, A. A. (1994) *Nature 372*, 478−481.
50. Fraser, J. D., Urban, R. G., Strominger, J. L., and Robinson, H. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 5507−5511.
51. Abrahmsén, L., Dohlsten, M., Segrén, S., Björk, P., Jonsson, E., and Kalland, T. (1995) *EMBO J. 14*, 2978−2986.
52. Herman, A., Labrecque, N., Thibodeau, J., Marrack, P., Kappler, J. W., and Sékaly, R. P. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 9954−9958.
53. Christianson, D. W. (1991) *Adv. Protein Chem. 42*, 281−355.
54. Stock, J. B., Surette, M. G., Levit, M., and Park, P. (1995) in *Two-component signal transduction* (Hoch, J. A., and Sihavy, T. J., Eds.) ASM Press, Washington, DC.
55. Volz, K. (1995) in *Two-component signal transduction* (Hoch, J. A., and Silhavy, T. J., Eds.) ASM Press, Washington, DC.
56. Lowry, D. F., Roth, A. F., Rupert, P. B., Dahlquist, F. W., Domaille, P. J. and Matsumura, P. (1994) *J. Biol. Chem. 269*, 26358−26362.
57. Pelton, J. G., Torchia, D. A., Meadow, N. D., and Roseman, S. (1992) *Biochemistry 31*, 5215−5224.
58. Parkinson, J. S., and Kofoid, E. C. (1992) *Annu. Rev. Genet. 26*, 71−112.
59. Swanson, R. V., Lowry, D. F., Matsumura, P., McEvoy, M. M., Simon, M. I., and Dahlquist, F. W. (1995) *Nat. Struct. Biol. 2*, 906−910.

BI971999E