from the imidazole ring of His^{57} to allow entry of the hydrolytic water molecule.

We believe we have resolved the issues posed at the beginning. Because the course of hydrolysis is slow, the time resolution was sufficient to trap the hydrolytic water molecule. The Laue method, judiciously applied, revealed details as subtle as the rearrangement of water molecules. Finally, the GB-trypsin system provided firm structural evidence for a role played by water during ester hydrolysis that is similar to that played by the serine hydroxyl during amide cleavage. Further, the experiment defined the geometry of the incoming water molecule, His^{57}, and the GB group and suggests an approach route for this molecule as it attacks the carbonyl carbon atom, becoming more nucleophilic as its interaction with the catalytic His residue increases.

REFERENCES AND NOTES

1. D. M. Blow, J. J. Birktoft, B. S. Hartley, *Nature* **221**, 337 (1969).
2. T. Chase, Jr., and E. Shaw, *Biochem. Biophys. Res. Commun.* **29**, 508 (1967).
3. W. F. Mangel *et al.*, *Biochemistry* **29**, 8351 (1990); Brookhaven Protein Data Bank entry 1GBT.
4. H. D. Bartunik, L. J. Summers, H. H. Bartsch, *J. Mol. Biol.* **210**, 813 (1989).
5. P. T. Singer *et al.*, *Philos. Trans. R. Soc. London Ser. A* **340**, 285 (1992).
6. J. R. Helliwell *et al.*, *J. Appl. Crystallogr.* **22**, 483 (1989).
7. V. Luzatti, *Acta Crystallogr.* **5**, 802 (1952).
8. P. T. Singer *et al.*, in preparation.
9. R. J. Henderson, *J. Mol. Biol.* **54**, 341 (1970).
10. For example, R. M. Sweet *et al.*, *Biochemistry* **13**, 4212 (1974).
11. X-PLOR; A. Brünger, Polygen, Waltham, MA.
12. J. S. Sack, *J. Mol. Graphics* **6**, 24 (1988).

13. P.T.S. was supported by an Alexander Hollaender Fellowship during much of this work. We acknowledge support for this work by the Office of Health and Environmental Research of the U.S. Department of Energy and by the National Science Foundation.

9 July 1992; accepted 2 October 1992

Structure of the Regulatory Complex of *Escherichia coli* III^{Glc} with Glycerol Kinase

James H. Hurley,* H. Rick Faber, David Worthylake,† Norman D. Meadow, Saul Roseman, Donald W. Pettigrew, S. James Remington‡

The phosphocarrier protein III^{Glc} is an integral component of the bacterial phosphotransferase (PTS) system. Unphosphorylated III^{Glc} inhibits non-PTS carbohydrate transport systems by binding to diverse target proteins. The crystal structure at 2.6 Å resolution of one of the targets, glycerol kinase (GK), in complex with unphosphorylated III^{Glc}, glycerol, and adenosine diphosphate was determined. GK contains a region that is topologically identical to the adenosine triphosphate binding domains of hexokinase, the 70-kD heat shock cognate, and actin. III^{Glc} binds far from the catalytic site of GK, indicating that long-range conformational changes mediate the inhibition of GK by III^{Glc}. GK and III^{Glc} are bound by hydrophobic and electrostatic interactions, with only one hydrogen bond involving an uncharged group. The phosphorylation site of III^{Glc}, His^{90}, is buried in a hydrophobic environment formed by the active site region of III^{Glc} and a 3_{10} helix of GK, suggesting that phosphorylation prevents III^{Glc} binding to GK by directly disrupting protein-protein interactions.

In *Escherichia coli*, the phosphoenolpyruvate:glycose phosphotransferase system (PTS) functions both as a transport pathway for its sugar substrates and as a negative regulator of the transport of several energy sources that are not PTS substrates (*1*). The regulatory state of the cell is determined by the state of phosphorylation of III^{Glc}, a PTS component specific for the uptake and phosphorylation of glucose. In the presence of PTS sugars, unphosphorylated III^{Glc} binds and inhibits permeases and other enzymes that participate in uptake and metabolism of non-PTS carbon sources. Unphosphorylated III^{Glc} is an apparent noncompetitive inhibitor of glycerol kinase (GK) with respect to both glycerol and Mg^{2+}–adenosine triphosphate (ATP) (*2*). Inhibition is eliminated upon phosphorylation of III^{Glc} (*2*, *3*). To elucidate the structural basis for the regulatory action of III^{Glc}, we determined the structure of the complex of *E. coli* III^{Glc} with GK and glycerol in the presence and absence of adenosine diphosphate (ADP) by multiple isomorphous replacement, solvent flattening, and crystallographic refinement (Table 1). The model, containing 4991 nonhydrogen atoms, had a conventional crystallographic *R* factor of 0.191 for all data observed between 5 and 2.6 Å resolution. The root-mean-square (rms) deviation was 0.013 Å from ideal bond lengths, 3.1° from bond angles, and 2.3 $Å^2$ from restrained thermal parameters.

GK consists of two large domains, one on either side of a deep and narrow cleft (Figs. 1–3). Regions of the domains that form the cleft are topologically identical to the common ATPase core of hexokinase (*4*), the 70-kD heat shock cognate (HSC70) (*5*), and actin (*6*). We follow the nomenclature established for HSC70. The large domains I and II (Fig. 3A) are divided into smaller subdomains. The two subdomains of the common ATPase fold are denoted IA and IIA. The others, IB, IC, IIB, and IIC, are insertions into or extensions of the common ATPase core (*7*), with IC and IIC unique to GK. Domain IB of GK, consisting of two insertions, is larger than the other IB domains and contains a central five-stranded antiparallel β sheet not found in other members of the family. Although domain IIB is topologically identical to the IIB domains of HSC70 and actin, the tertiary structure of this domain is unrelated to hexokinase, HSC70 or actin.

Glycerol binds at the bottom of the interdomain cleft (Figs. 2A and 3B). The side chains of Arg^{83}, Glu^{84}, Tyr^{135}, and Asp^{245} and the amide NH of Arg^{83} form hydrogen bonds with the hydroxyl groups of glycerol. The side chains of Trp^{103} and Phe^{270} are in van der Waals contact with the carbon backbone of glycerol. Asp^{10} and Asp^{245} form a possible Mg^{2+} binding site. Asp^{10} has been proposed to function in ATP hydrolysis on the basis of its occurrence within a fingerprint sequence derived from the structures of the ATP binding sites of hexokinase, HSC70, and actin (*8*). This fingerprint sequence in GK also led to the prediction that GK would have a fold similar to these three proteins (*8*). Asp^{245} is structurally equivalent to Asp^{211} (*4*, *9*), the putative catalytic base of hexokinase, and is the only amino acid of GK that reacts with affinity labels dibromoacetone and dichloroacetone (*10*). Asp^{245} forms a hydrogen bond with the 3-hydroxyl of glycerol, sug-

J. H. Hurley and H. R. Faber, Institute of Molecular Biology, University of Oregon, Eugene, OR 97403.
D. Worthylake and S. J. Remington, Institute of Molecular Biology and Department of Physics, University of Oregon, Eugene, OR 97403.
N. D. Meadow and S. Roseman, Department of Biology and McCollum-Pratt Institute, The Johns Hopkins University, Baltimore, MD 21218.
D. W. Pettigrew, Department of Biochemistry and Biophysics, Texas A&M University, College Station, TX 77843.

*Present address: Laboratory of Molecular Biology, National Institute of Diabetes, Digestive and Kidney Diseases, National Institutes of Health, Bethesda, MD 20892.
†Present address: Molecular Biology Program, University of Utah, Salt Lake City, UT 84112.
‡To whom correspondence should be addressed.

gesting that it is the catalytic base in the phosphorylation reaction.

The ADP binding site is located directly above the glycerol binding site in the interdomain cleft (Figs. 2 and 3B). The only structural change on ADP binding is a rotation of the side chain of Arg^{17} and its interaction with the pyrophosphate moiety of ADP. The pyrophosphate moiety also interacts with main chain nitrogens of Gly^{266}, Thr^{267}, and Gly^{411}. The ribose forms hydrogen bonds with the side chain of Gln^{314} and the main chain at Gly^{310} and Ala^{412}, while the adenine base lies in a hydrophobic pocket formed by main chain atoms and the side chains of Ile^{313}, Ala^{326}, Leu^{381}, Ile^{384}, Ala^{412}, and Asn^{415}. A hydrogen bond may be formed between N-1 of adenine and the side chain of Asn^{415}.

GK exists at physiological concentrations in an equilibrium between functional dimers and tetramers (*11*). The crystal contains tetramers of GK with exact 222 point-group symmetry, with the tetramer located at the intersection of three twofold axes in the crystal lattice. Each GK monomer is related to the others by a twofold rotation about the X, Y, or Z axis. The original GK monomer and those related to it are designated O, X, Y, and Z, respectively. Domain IIB forms the most extensive intersubunit contact, primarily with the Y subunit. Residues 361 to 367 and 341 to 347 of the O and Y subunits form a four-stranded intersubunit antiparallel β sheet. Helices (residues 308 to 320 and 373 to 399) pack on one side of the intersubunit β sheet, and the COOH-terminal helix (residues 480 to 493) packs on the opposite side, with a central strand–to-helix angle of about −50°, reminiscent of the clasp-like subunit interfaces described for several proteins (*12*). Domain IIB also forms less extensive interactions with the Z subunit through a loop and an α helix comprising residues 321 to 333. Helices (residues 49 to 68) of domains IA of O and X pack in an antiparallel fashion to form the only interaction between these subunits (Fig. 2B). The O–Y interface buries 3521 $Å^2$ of solvent-accessible surface area (*13*) between both subunits, far more than the 381 $Å^2$ and 524 $Å^2$ buried by the O–X and O–Z interfaces, respectively. Thus, the dimer-tetramer equilibrium in solution probably involves dissociation of the tetramer into O–Y dimers. Although each active site comprises residues from only one monomer, the side chain of Arg^{369} of the Y subunit penetrates deeply into the O subunit, approaching within 7 Å of the ADP molecule, and may be involved in allosteric regulation of enzymatic activity.

Hexokinase undergoes a dramatic conformational transition from open to closed forms upon binding its substrate, glucose (*14*). GK is also thought to undergo a large conformational change upon binding glycerol (*15*). The GK structures were obtained with glycerol present in both instances. The ATPase core domains IA and IIA are in a closed conformation similar to that observed in the hexokinase-glucose complex and in the HSC70 and actin structures. One hundred eighty α carbons (Cα) of domains IA and IIA of HSC70 can be superimposed simultaneously on glycerol kinase with an rms deviation of only 2.9 Å. This deviation is only marginally greater than deviations of 2.4 Å and 2.7 Å rms for separate superimposition of 76 and 104 Cαs of the domains IA and IIA (*16*), suggesting that the relative orientations of domains IA and IIA are essentially identical in GK and HSC70. Binding of ADP to GK produces only very localized structural changes. By analogy with hexokinase, the observed closed conformation of GK is precisely what would be expected in the presence of substrate.

III^{Glc} consists of a β-sheet sandwich, with the active site residues His^{75} and His^{90} surrounded by hydrophobic and negatively charged residues on one face of the sandwich (*17, 18*). In the crystal of the complex, one molecule of III^{Glc} is bound to each GK monomer. There is little difference (0.3 Å rms deviation for Cα positions) between the structures of residues 19 to 168 of III^{Glc} and the same residues in uncomplexed III^{Glc} (*17*). The largest difference in the crystal of the complex is the location of residues 1 to 11. Nuclear magnetic resonance (NMR) studies of uncomplexed III^{Glc} in solution demonstrated that the NH_2-terminal 18 residues were unstructured (*19*), although they are important for phosphate transfer to glucose permease (*20*). Crystals of uncomplexed *E. coli* III^{Glc} (*17*) were only obtained for the fast form of the protein, lacking residues 1 to 7 (*21*), and residues 8 to 18 were disordered. In the present structure, residues 1 to 11 are bound

Fig. 1. Electron density from the final solvent-flattened multiple isomorphous replacement map, contoured at 1.0 σ. The refined model in the vicinity of $HgCl_2$ site 2, at Cys^{255} and Met^{260}, is presented.

Table 1. Crystallographic statistics. The III^{Glc}-GK complex was crystallized by hanging drop vapor diffusion from a 1:1 molar mixture of glycerol kinase and intact ("slow") III^{Glc} (*32*). Data were collected on a Xoung-Hamlin area detector (*33*) with the use of graphite monochromated Cu K_α from a Rigaku RU-200BH rotating anode x-ray generator. Initial data reduction was performed with the supplied detector software (*34*), but final intensities were scaled and merged with exponential scaling (*35*). Heavy atom parameters were refined by the origin-removed difference Patterson method (*36*). An improved phase set was obtained by solvent flattening (*37*), and these phases were used to improve heavy atom parameters by standard lack-of-closure refinement. Phases from a partial model, built with FRODO (*38*) and containing 75% of the total scattering matter, were used for a final round of lack-of-closure heavy atom parameter refinement, resulting in further improvement of the electron density map. Crystallographic refinement was carried out by simulated annealing (*39, 40*).

Data set	Completeness*	Observations (No.)		D_{min} (Å)	R_{merge}†	Sites (No.)	R_c‡	Phasing power§
		Total	Unique					
Native	0.76 (0.58)	81994	24169	2.6	0.056 (0.232)			
ADP	0.81 (0.59)	52729	26116	2.6	0.065 (0.235)			
$HgCl_2$	0.86 (0.70)	44277	17870	3.0	0.058 (0.169)	2	0.45	1.38
cis-Pt	0.89 (0.97)	20616	7754	4.0	0.052 (0.071)	3	0.73	0.42

*Completeness is the ratio of observed to theoretically possible reflections. †$R_{merge} = \Sigma|I_{obs} - I_{avg}| / \Sigma I_{avg}$. Values in parentheses are for the highest resolution shell used: 2.6 to 2.7 Å for native and ADP; 3.0 to 3.2 Å for the mercury derivative; and 4.0 to 4.2 Å for the platinum derivative. ‡$R_c = \Sigma \| F_{PH} \pm F_P| - F_H| / \Sigma |F_{PH} \pm F_P|$, where the summation is over centric reflections. §Phasing power is the ratio of the rms heavy atom scattering to the rms lack of closure of the phase triangles.

to GK in an extended conformation. No electron density was located for residues 12 to 18, which could span up to 28 Å. However, the proper connection between the NH_2-terminal region and the rest of III^{Glc} was inferred because residue 19 of the nearest III^{Glc} is 17 Å distant in the crystal, and the next nearest III^{Glc} is more than 50 Å distant.

Three contacts occur between III^{Glc} and two different GK tetramers in the crystal (Table 2). The largest and most hydrophobic of the contacts includes the active site region of III^{Glc}. Contact 1 comprises residues 38 to 46, 71, 78, 88, 90, and 94 to 97 of III^{Glc}, and residues 402 and 472 to 481 of GK. Contacts 2 and 3 are formed with a GK subunit of a different tetramer than the one involved in contact 1 (*22*). Contact 2 is also hydrophobic, but the composition of contact 3 is typical of protein surfaces in general (*23*).

Interfaces between functionally interacting proteins bury 1250 to 1950 $Å^2$ of solvent-accessible surface area (*23*). Only contact 1 is within this range, although contact 2 is only marginally less. Contact 1 is most likely to be relevant to the GK-III^{Glc} regulatory interaction. Because residues 1 to 11 of III^{Glc}, which participate in contact 2, are disordered for free III^{Glc}, the entropic cost of ordering these residues implies that this interaction is weak. Also, *Bacillus subtilis* III^{Glc}, lacking the seven NH_2-terminal residues of *E. coli* III^{Glc}, complements a defect in GK regulation by III^{Glc} in *E. coli* (*24*), indicating that these residues are not essential for regulation of GK. The small and relatively hydrophilic contact 3 probably exists only in the crystal.

The most dramatic feature of the major GK-III^{Glc} contact is the burial of the active site residues His^{75} and His^{90} of III^{Glc} (*25*) in the center of the interface. The side chain of His^{90} is completely buried in the complex but is 14% solvent accessible in uncomplexed III^{Glc} (*13*). There are no direct hydrogen bonds between the histidines and GK. The major III^{Glc} binding site on GK is within the IIC domain and consists of a two-turn segment of 3_{10} helix (Fig. 2A) linked to a longer α helix by a 45° bend. Residues 472 to 481 project directly into the concave active site of III^{Glc} (Fig. 4). The III^{Glc} binding site is part of the O–Y subunit interface, but there is no direct contact between the subunit O III^{Glc} and subunit Y. Salt bridges are formed between Arg^{402} and Arg^{479} of GK and Glu^{43} and Asp^{38} of III^{Glc}. The only hydrogen bond involving an uncharged group is made by Asn^{480} of GK and Asp^{94} of III^{Glc}, but it is long (3.35 Å). The side chains of Pro^{472}, Ile^{474}, Tyr^{481}, the aliphatic portions of Glu^{478} and Arg^{479}, and main chain atoms of GK form hydrophobic interactions with the side chains of Val^{40}, Phe^{41}, Ile^{45}, Val^{46}, Phe^{71}, Phe^{88}, and Val^{96} of III^{Glc}. These side chains of III^{Glc} form a hydrophobic ring around the active site of III^{Glc} and had previously been suggested to be involved in protein-protein interactions (*17*, *18*).

With burial of some 1300 $Å^2$ of molecular surface, the III^{Glc}-GK interface is smaller than most contact surfaces between functionally interacting proteins. There are also fewer polar interactions than in other classes of protein-protein interactions. The III^{Glc}-GK binding interaction in solution, with half-maximal inhibition at 4 to 10 μM depending on pH (*2*), is weaker than for other classes of protein-protein complexes (*23*). III^{Glc} is normally present at high (50 μM) concentrations in the cell (*2*); the interface may need to be relatively small in order to ensure that binding is reversible upon phosphorylation of III^{Glc}.

In order to carry out the multiple functions of phosphate transfer and regulation, III^{Glc} physically interacts with apparently unrelated proteins. The three-dimensional structure of one other III^{Glc}-binding protein, the histidine-containing phosphocarrier protein HPr, has been solved (*26*, *27*). HPr transfers a phosphate directly from His^{15} to His^{90} of III^{Glc}. There is no sequence similarity between the III^{Glc} binding regions of GK and HPr. While both structures contain helical secondary structure and an arginine two positions COOH-terminal to the position (Thr^{477} of GK, His^{15} of HPr) interacting with His^{90} of III^{Glc}, the structures have little in common. Thr^{477} of GK is near the center of a 3_{10} helix, while His^{15} of

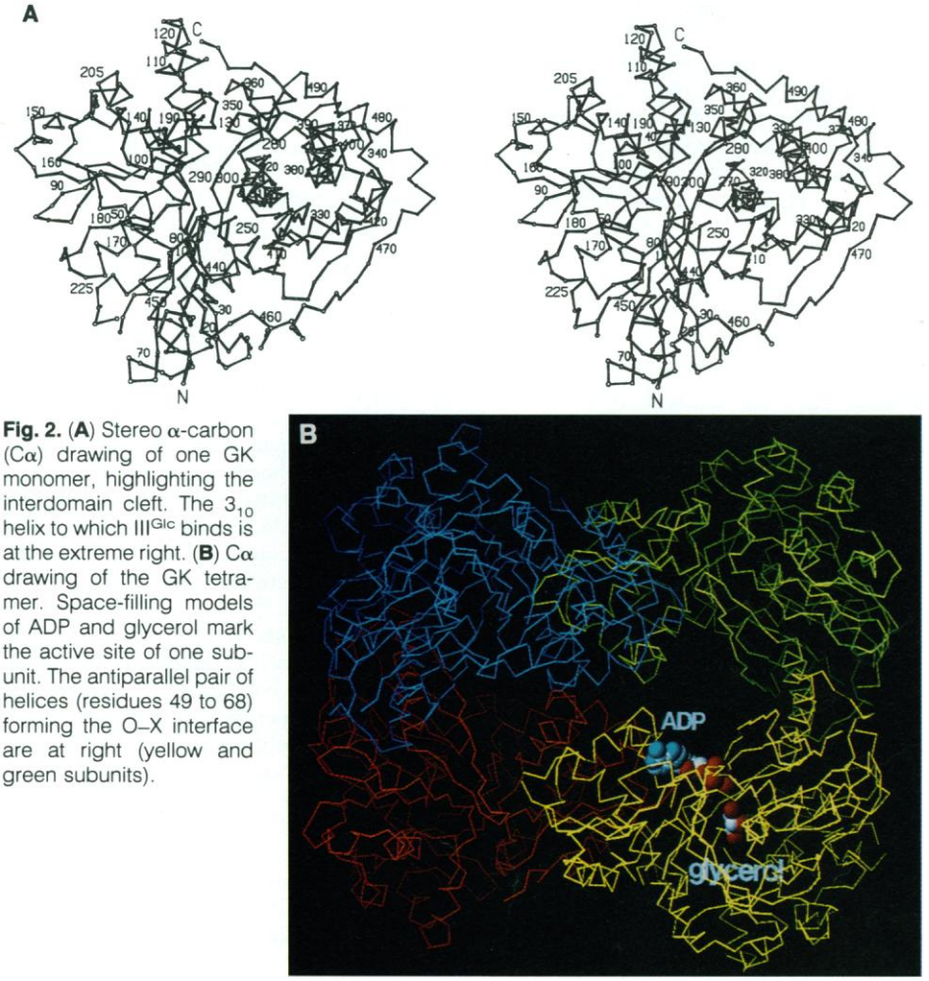

Fig. 2. (**A**) Stereo α-carbon (Cα) drawing of one GK monomer, highlighting the interdomain cleft. The 3_{10} helix to which III^{Glc} binds is at the extreme right. (**B**) Cα drawing of the GK tetramer. Space-filling models of ADP and glycerol mark the active site of one subunit. The antiparallel pair of helices (residues 49 to 68) forming the O–X interface are at right (yellow and green subunits).

Table 2. Glycerol kinase–III^{Glc} contacts. ΔA is the difference for both GK and III^{Glc} between the total surface area calculated for the isolated molecules and for the complex. Solvent-accessible surface areas were calculated with a probe radius of 1.4 Å and a grid-based algorithm with a step size of 0.8 Å. The nonpolar solvent accessible surface is defined as the surface area calculated for carbon atoms alone in the presence of all protein atoms.

Contact	ΔA ($Å^2$)		H bonds	Salt bridges
	Total	Nonpolar		
1	1302	895	1	2
2	1104	744	5	0
3	999	519	3	1

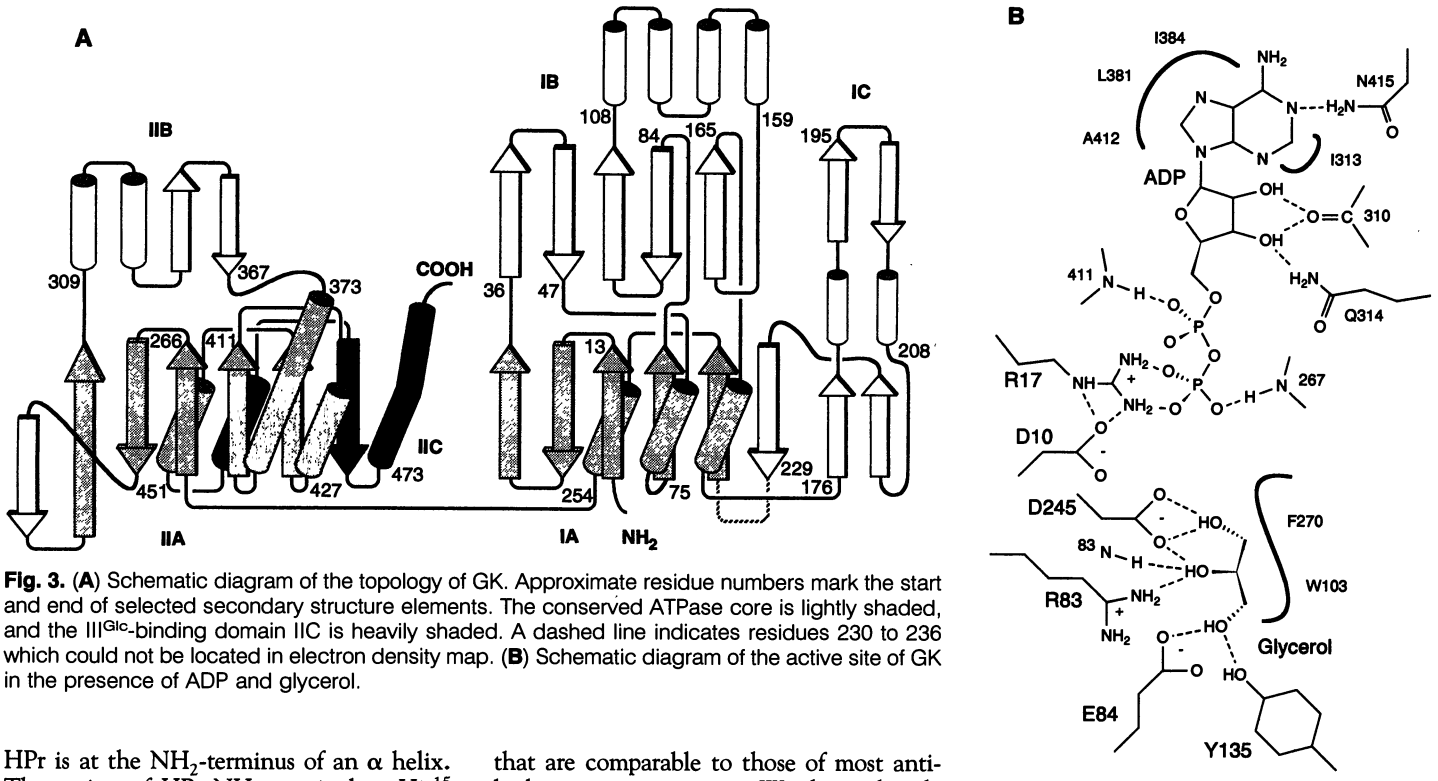

Fig. 3. (**A**) Schematic diagram of the topology of GK. Approximate residue numbers mark the start and end of selected secondary structure elements. The conserved ATPase core is lightly shaded, and the III^{Glc}-binding domain IIC is heavily shaded. A dashed line indicates residues 230 to 236 which could not be located in electron density map. (**B**) Schematic diagram of the active site of GK in the presence of ADP and glycerol.

HPr is at the NH_2-terminus of an α helix. The region of HPr NH_2-terminal to His^{15} consists of irregular secondary structure with no counterpart in GK.

A single binding site might recognize diverse sequences and structures by means of nondirectional intermolecular forces. Antigen-antibody interactions are highly specific and involve 10 to 23 hydrogen bonds in each complex (*23*). In contrast, the regulatory protein calmodulin recognizes a diverse family of proteins through hydrophobic and electrostatic interactions with no major role for hydrogen bonds (*28*). Given the 2.6 Å resolution of the GK-III^{Glc} complex, we have used a strict 3.4 Å cutoff in searching for potential hydrogen bonds, conditions that are comparable to those of most antibody-antigen structures. We located only one intermolecular hydrogen bond involving an uncharged group.

The structure of the III^{Glc}-GK complex suggests a mechanism whereby phosphorylation of III^{Glc} controls the interaction. The phosphorylation site of III^{Glc} at His^{90} is buried in a hydrophobic environment. An unfavorably close contact of approximately 2.2 Å between a phosphate oxygen and the side chain of Thr^{477} would occur in the absence of significant structural changes. Comparison of the three-dimensional NMR spectra of phosphorylated and unphosphorylated III^{Glc} indicates that phosphorylation produces only small structural perturbations (less than 1.5 Å) that are confined to the immediate neighborhood of His^{90} (*29*). It has also been shown by NMR that in phosphorylated III^{Glc} the active site His^{75} is not charged at cytoplasmic pH (*30*). The nearest charged groups to the His^{90} phosphate would be the side chains of Asp^{38} and Asp^{94} of III^{Glc} (6.5 and 4.9 Å distant, respectively) and Glu^{478} of GK (6.3 Å). The closest positively charged group would be Arg^{479} of GK (7.0 Å). Thus a phosphate covalently bound to His^{90} would be in a highly unfavorable environment, removed from solvation by bulk water, surrounded by negatively charged groups, without any compensating interactions with positively charged groups in the complex. *Escherichia coli* isocitrate dehydrogenase is inactivated by phosphorylation as a result of direct electrostatic and steric repulsion between the phosphate and the substrate isocitrate (*31*). For these cases, enzyme-substrate and protein-protein binding are both directly disrupted by phosphorylation.

How III^{Glc} binding inhibits the activity of GK is still unclear. Allosteric regulation of GK by III^{Glc} belongs to the rare class of purely noncompetitive inhibition with respect to both substrates (*2*). It is thus not surprising that bound glycerol and ADP are observed in the crystal in the presence of the inhibitor III^{Glc}. The III^{Glc} binding site is farther than 30 Å from the catalytic site of GK, which rules out a direct short-range effect of III^{Glc}. Presumably III^{Glc} destabilizes one or more of the conformations of GK in the reaction pathway.

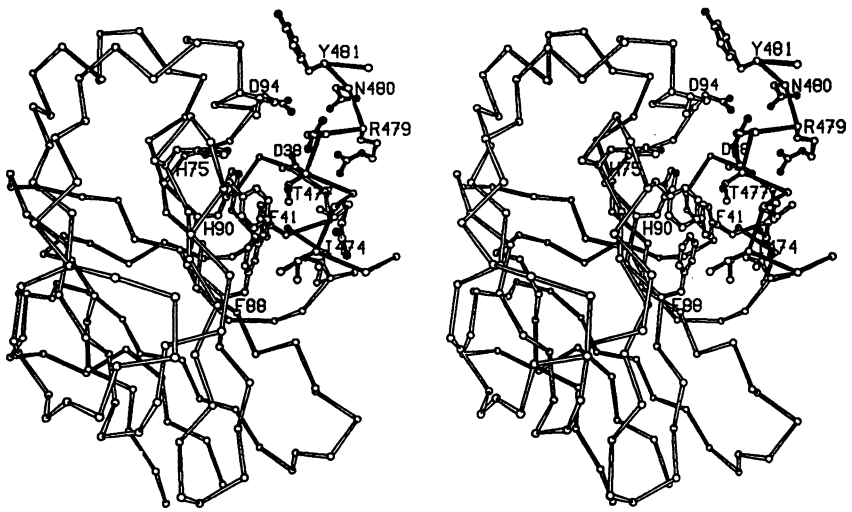

Fig. 4. Stereo Cα view, with selected side chains, of contact 1 between GK and the active site of III^{Glc}. GK (filled bonds), and III^{Glc} (open bonds) nitrogen and oxygen atoms (filled).

REFERENCES AND NOTES

1. P. W. Postma and J. Lengler, *Microbiol. Rev.* **49**, 232 (1985); M. H. Saier, Jr., *ibid.* **53**, 109 (1989); N. D. Meadow, D. K. Fox, S. Roseman, *Annu. Rev. Biochem.* **59**, 497 (1990); S. Roseman and N. D. Meadow, *J. Biol. Chem.* **265**, 2993 (1990).
2. M. J. Novotny *et al.*, *J. Bacteriol.* **162**, 810 (1985).
3. P. W. Postma *et al.*, *ibid.* **158**, 351 (1984).
4. C. M. Anderson, R. E. Stenkamp, R. C. McDonald, T. A. Steitz, *J. Mol. Biol.* **123**, 207 (1978).
5. K. M. Flaherty *et al.*, *Nature* **346**, 623 (1990).
6. W. Kabsch *et al.*, *ibid.* **347**, 37 (1990).
7. Domains of GK consist of IA, residues 1 to 35, 49 to 83, 165 to 173, and 221 to 253; IB, residues 36 to 48 and 82 to 164; IIA, residues 254 to 306 and 373 to 472; IIB, residues 307 to 372. Unique additions are IC, residues 174 to 220, and IIC, residues 456 to 501.
8. K. M. Flaherty *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 5041 (1991).
9. K.-U. Frohlich, K. D. Entian, D. Mecke, *Gene* **36**, 105 (1985). Residues numbers for hexokinase are quoted for the translated gene, not the crystallographic sequence.
10. D. W. Pettigrew and D. C. Thomas, unpublished results.
11. J. K. DeRiel and H. Paulus, *Biochemistry* **17**, 5134 (1978); *ibid.*, p. 5141; *ibid.*, p. 5146.
12. P. J. Bjorkman *et al.*, *Nature* **329**, 506 (1987); R. St. Charles, D. A. Walz, B. F. P. Edwards, *J. Biol. Chem.* **264**, 2092 (1989); G. M. Clore, E. Apella, M. Yamada, K. Matsushima, A. M. Gronenborn, *ibid.*, p. 18907; J. H. Hurley *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8635 (1989); K. Valegard *et al.*, *Nature* **345**, 36 (1990).
13. B. Lee and F. M. Richards, *J. Mol. Biol.* **55**, 379 (1971). Probe radius was 1.4 Å.
14. W. S. Bennett, Jr. and T. A. Steitz, *ibid.* **140**, 183 (1980).
15. J. W. Thorner and H. Paulus, *J. Biol. Chem.* **248**, 3922 (1973).
16. The superposition included residues 5 to 25, 48 to 68, 77 to 82, 164 to 174, 240 to 256, 260 to 266, 268 to 276, 302 to 319, 381 to 402, 407 to 436, and 437 to 454 of GK, and residues 5 to 25, 116 to 136, 142 to 147, 156 to 166, 170 to 186, 195 to 201, 204 to 212, 221 to 238, 306 to 327, 335 to 364, and 367 to 384 of HSC70.
17. D. Worthylake *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 10382 (1991).
18. D.-I. Liao *et al.*, *Biochemistry* **30**, 9583 (1991).
19. J. G. Pelton *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 3479 (1991).
20. N. D. Meadow and S. Roseman, *J. Biol. Chem.* **257**, 14526 (1982).
21. N. D. Meadow *et al.*, *J. Biol. Chem.* **261**, 13504 (1986).
22. In contact 2, residues 1 to 11 of III^{Glc} bind to residues 125 to 129, 195, 258 to 259, 274 to 281, and 399 to 401 of the GK subunit of the second tetramer. In contact 3, residues 27 to 29, 55 to 56, and 155 to 161 of III^{Glc} interact with residues 183, 195 to 197, 219, 277, and 298 to 300 of the same GK subunit as for contact 2.
23. D. R. Davies, E. A. Padlan, S. Sheriff, *Annu. Rev. Biochem.* **59**, 439 (1990); J. Janin and C. Chothia, *J. Biol. Chem.* **265**, 16027 (1990).
24. J. Reizer *et al.*, *J. Biol. Chem.* **267**, 9158 (1992).
25. K. A. Presper *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 4052 (1989).
26. R. E. Klevit and E. B. Waygood, *Biochemistry* **25**, 7774 (1986).
27. O. Herzberg *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **89**, 2499 (1992).
28. M. Ikura *et al.*, *Science* **256**, 632 (1992); W. E. Meador *et al.*, *ibid.* **257**, 1251 (1992).
29. J. G. Pelton, D. A. Torchia, N. D. Meadow, S. Roseman, *Biochemistry* **31**, 5215 (1992).
30. ———, *Protein Science*, in press.
31. J. H. Hurley, A. M. Dean, P. E. Thorsness, D. E. Koshland, Jr., R. M. Stroud, *J. Biol. Chem.* **265**, 3599 (1990); J. H. Hurley, A. M. Dean, J. L. Sohl, D. E. Koshland, Jr., R. M. Stroud, *Science* **249**, 1012 (1990); A. M. Dean and D. E. Koshland, Jr., *ibid.*, p. 1044.
32. GK and III^{Glc} were mixed at 1:1 stochiometry to yield a final protein concentration of approximately 20 mg/ml. A 5-μl drop of the well solution consisting of 0.5 to 0.8 M sodium acetate, 100 mM MES (pH 6.0) was mixed with 5 μl of the protein mixture. Crystals of up to 1.0 mm along the largest dimension were obtained in 1 week at room temperature. The space group is *I*222 with a = 123.4, b = 124.3, c = 133.6 Å. The packing parameter [B. W. Matthews, *J. Mol. Biol.* **33**, 491 (1968)] is V_m = 3.5 Å^3/D for one monomer each of GK and III^{Glc} in the asymmetric unit. The crystals diffracted to 2.4 Å. Heavy atom derivatization was performed in a storage solution of 16% (w/v) PEG 1550, 40 mM Pipes (pH 7.0). A 3-day soak in a 1:10 dilution of a saturated solution of *cis*-platinum(II) diamine dichloride (cis-Pt) and a 2-week soak of a 1:100 dilution of a saturated solution of $HgCl_2$ in the presence of 50 mM 2-mercaptoethylamine yielded useful derivatives. "ADP" refers to crystals soaked in storage solution containing 5 mM ADP, 5 mM $MgCl_2$, and 5 mM glycerol. Protein was purified as described [H. R. Faber, D. W. Pettigrew, S. J. Remington, *J. Mol. Biol.* **207**, 637 (1989); J. G. Pelton, D. A. Torchia, N. D. Meadow, C.-Y. Wong, S. Roseman, *Biochemistry* **30**, 10043 (1992)].
33. R. Hamlin, *Methods Enzymol.* **114**, 416 (1985).
34. A. J. Howard *et al.*, *ibid.*, p. 452.
35. M. G. Rossmann, *J. Appl. Crystallogr.* **12**, 225 (1979); M. F. Schmid *et al.*, *Acta Crystallogr. Sect. A* **37**, 701 (1981).
36. T. C. Terwilliger and D. Eisenberg, *Acta Crystallogr. Sect. A* **39**, 813 (1983).
37. B. C. Wang, *Methods Enzymol.* **115**, 90 (1985).
38. T. A. Jones, *ibid.*, p. 157.
39. A. T. Brünger, J. Kuriyan, M. Karplus, *Science* **235**, 458 (1987); A. T. Brunger, M. Karplus, G. A. Petsko, *Acta Crystallogr. Sect. A* **45**, 50 (1989).
40. In the first round of refinement, the partial model (75% of all atoms) was energy-minimized, annealed from 6000 K to 300 K, and minimized again. After group B-factor refinement, the R factor dropped from 51 to 32% for all data from 6.0 to 2.8 Å. After greater than 90% of all atoms had been modeled in electron density from the final solvent-flattened MIR map, refinement was continued with all data from 5.0 to 2.6 Å. Four rounds of manual rebuilding and simulated annealing from 2000 K to 300 K and two rounds of rebuilding with conventional conjugate gradient minimization led to the current model. A glycerol molecule was located in an $F_{obs} - F_{calc}$ Fourier synthesis following the third round of simulated annealing refinement. The final model contained 650 amino acid residues, one glycerol molecule, and no water molecules. About 2% of the main chain torsion angle (ϕ, ψ) combinations deviated from allowed regions of the Ramachandran diagram by more than 20°. Residues 1 to 3, 230 to 236, and 500 to 501 of GK and residues 12 to 18 of III^{Glc} as well as 18 side chains of GK could not be located in the electron density and have been omitted from the model. The ADP molecule was placed in a difference electron density map and the model refined.

41. We thank K. Kallio for technical assistance, R. Huber and his colleagues for calling our attention to their derivitization method [H. W. Hoeffken *et al.*, *J. Mol. Biol.* **204**, 417 (1988)]; D. B. McKay for providing complete coordinates of HSC70; and B. W. Matthews, F. W. Dahlquist, and W. Tulip for helpful discussions. Supported by the National Institutes of Health grants GM 42618-01A1 (S.J.R.), American Cancer Society ACS 1296 (S.J.R.), 5-R37 GM38759 (S.R.), a grant from the Lucille P. Markey Charitable Trust, and a postdoctoral fellowship from the American Cancer Society (J.H.H.). Crystallographic refinement was carried out in part at the Pittsburgh Supercomputing Center under grant DMB900038P (J.H.H. and B. W. Matthews). Coordinates have been deposited in the Protein Data Bank.

8 July 1992; accepted 6 November 1992

Patchiness and Correlations in DNA Sequences

Samuel Karlin and Volker Brendel

The highly nonrandom character of genomic DNA can confound attempts at modeling DNA sequence variation by standard stochastic processes (including random walk or fractal models). In particular, the mosaic character of DNA consisting of patches of different composition can fully account for apparent long-range correlations in DNA.

Genomic global and local compositional heterogeneity is widely recognized. The many examples of DNA heterogeneity in existence include isochore compartments [regions dominated by either G + C or A + T as determined by thermal-melting studies or density-gradient centrifugation; for example, the G + C– and A + T–rich halves of the bacteriophage lambda genome (*1*); see (*2*) for examples in mammalian species]; mobile insertion elements [such as Alu in human, Ty in yeast, and IS in *Escherichia coli* (*3*)]; characteristic satellite centromeric tandem repeats [such as the 171–base pair human alpha satellite DNA (*4*)]; characteristic telomeric sequences [such as AGGGTT tandem repeats in humans (*5*)]; HTF islands [vertebrate DNA sequences that occur generally upstream of genes and are abundant with nonmethylated CpG (*6*)]; repetitive extragenic palindromes (REPs) in *E. coli* and *Salmonella typhimurium* (*7*); repeat induced point (RIP) mutation in certain fungi (*8*); recombinational hot spots [such as chi elements in *E. coli* (*9*)]; universal underrepresentation of the dinucleotide TpA (*10*); suppression of the dinucleotide CpG in vertebrates (*11*); the pervasive rarity of the tetranucleotide CTAG (*12*); GNN periodicity in coding sequences (*13*); and methyltransferase modifications (*14*).

Thus, genome organization is complex and variegated. In general, genomic sequences are not homogeneous on any scale. For example, eukaryotic sequences are often endowed with tandem repeats accruing

Department of Mathematics, Stanford University, Stanford, CA 94305.