2008-08

Crystal structure of yeast hexokinase PI in complex with glucose: a classical "induced fit" example revised

Proteins,Hoboken, v. 72, n. 2, p. 731-740, Aug. 2008
http://www.producao.usp.br/handle/BDPI/49109

PROTEINS
STRUCTURE ■ FUNCTION ■ BIOINFORMATICS

Crystal structure of yeast hexokinase PI in complex with glucose: A classical "induced fit" example revised

Paula Kuser,[1] Fabio Cupri,[2] Lucas Bleicher,[3] and Igor Polikarpov[3]*

[1] EMBRAPA Informática Agropecuária, Av. Dr. André Tosello, s/n, Cidade Universitária "Zeferino Vaz," C. P. 6041, CEP 13083-970, Campinas, São Paulo, Brazil

[2] Laboratório Nacional de Luz Síncrotron, C.P. 6192, CEP 13083-970, Campinas, São Paulo, Brazil

[3] Grupo de Cristalografia, Instituto de Física em São Carlos, USP, Av. Trabalhador Saocarlense, 400 CEP 13560-970, São Carlos, SP, Brazil

ABSTRACT

Hexokinase is the first enzyme in the glycolytic pathway that catalyzes the transfer of a phosphoryl group from ATP to glucose to form glucose-6-phosphate and ADP. Two yeast hexokinase isozymes are known, namely PI and PII. Here we redetermined the crystal structure of yeast hexokinase PI from Saccharomyces cerevisiae as a complex with its substrate, glucose, and refined it at 2.95 Å resolution. Comparison of the holo-PI yeast hexokinase and apo-hexokinase structures shows in detail the rigid body domain closure and specific loop movements as glucose binds and sheds more light on structural basis of the "induced fit" mechanism of reaction in the HK enzymatic action. We also performed statistical coupling analysis of the hexokinase family, which reveals two co-evolved continuous clusters of amino acid residues and shows that the evolutionary coupled amino acid residues are mostly confined to the active site and the hinge region, further supporting the importance of these parts of the protein for the enzymatic catalysis.

Proteins 2008; 72:731–740.

Key words: hexokinase X-ray structure; statistical coupling analysis; Saccharomyces cerevisiae.

INTRODUCTION

Hexokinase (ATP: D-hexose 6-phosphotransferase, EC 2.7.1.1; HK) is a member of the kinase family of tissue-specific isoenzymes. *Saccharomyces cerevisiae* genome codify for two isoenzymes of hexokinase: PI and PII, which are about 76% similar in their amino-acid sequences.[1,2] Isoenzyme PII has a total of 486 amino acid residues, whereas PI has 485 residues. Hexokinase PII is the predominant hexose kinase in *Saccharomyces cerevisiae* in growth on glucose[3] and is required for catabolite repression by glucose of expression of other genes.[4,5] The yeast HKs are known to exist as phosphoproteins *in vitro*[6] and *in vivo*[7]; *in vivo* phosphorylation site was identified as Ser15.[8] It was demonstrated that the transformed cells with the hexokinase PII S15A mutant gene could not provide glucose repression of invertase, suggesting that the phosphorylation of HK PII is essential for glucose signal transduction,[9] however, it seems to be not the only factor involved.[10] Nuclear localization of HK PII was shown to be mediated by an internal N-terminal decapeptide (amino acid residues 7–16) that is necessary to direct the protein to the nucleus, where it seems to be involved in the formation of specific DNA-protein complexes during glucose-induced repression of the SUC2 gene.[11] HK PI can only partially substitute HK PII in triggering glucose repression, but it is equally important in triggering fructose repression.[5,12,13]

Hexokinase catalyses the formation of hexose-6-phosphate (mainly glucose-6-phosphate, Glc-6P) from hexose using ATP as a phosphoryl donor. The isoenzymes differ in their activity toward particular hexoses. The enzymatic activity of PI isoenzyme with fructose and mannose is 2.6 and 0.6 times than that with glucose, respectively, whereas with PII isoenzyme, the fructose:glucose ratio is only 1.3 and mannose:glucose ratio is 0.3.[14] Yeast HKes exist in solution as dimers and monomers. The dissociation of dimers into monomers is promoted by glucose binding and an increase in pH and/or ionic strength.[15–17]

HKs are found in practically all living beings. Mammals, for example, have four HK isoenzymes[18] three of which have molecular weight of about 100 kDa and display a considerable sequence identity between their N and C-terminal

Grant sponsor: Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP); Grant number: 06/00182-8; Grant sponsor: Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq); Grant number: 473875/2003-9; Grant sponsor: Coordenação de Aperfeiçoamento de Pessoal do Nível Superior (CAPES).
*Correspondence to: Igor Polikarpov, Grupo de Cristalografia, Instituto de Física em São Carlos, USP, Av. Trabalhador Saocarlense, 400 CEP 13560-970, São Carlos, SP, Brazil. E-mail: ipolikarpov@if.sc.usp.br
Received 21 June 2007; Revised 8 November 2007; Accepted 29 November 2007
Published online 7 February 2008 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.21956

halves.[16] The fourth mammalian HK isoenzyme has a molecular mass of about 50 kDa, similar to yeast and *Schistosoma mansoni* HKs. The sequence similarity between N-and C-terminal part of the mammalian isoenzymes I to II is believed to be a consequence of duplication and fusion of a primordial hexokinase gene.[16,19,20] Yeast HKs, unlike human HKs and HK from *Schistosoma mansoni*, are insensitive to inhibition by the product of reaction Glc-6P.

Yeast HK isoenzyme is known to be a classical example of the theory of "induced fit" mechanism of enzymatic reaction originally proposed by Koshland[21,22] The crystallographic structures of yeast HK isoenzymes PI complexed with glucose[23,24] and PII complexed with a competitive inhibitor, ortho-toluoylglucosamine[25,26] have been determined at the end of the 70's and used to illustrate the role of the "induced fit" in enzyme action and to portray structural changes of the enzyme associated to glucose binding.[22] However, at the time of structure determination, the primary sequences of the yeast HK isoenzymes were not available, and the identity of the side-chains was guessed from inspection of crystallographically computed electron density. The models of PI and PII isoenzyme contain only 458 residues each and about 17% of these residues have uncertain identity. A comparison of the HK amino acid sequences found from X-ray studies with the actual ones[1,2] revealed only 30% identity between them. This fact complicated detailed analysis of the conformational changes of the enzyme associated with the glucose binding and introduced uncertainty in quantitative parameters of the "induced fit" mechanism of reaction obtained from these structures.

One of the ways to assess and identify amino acid residues involved in the allosteric networks is offered by statistical coupling analysis (SCA). This technique, whose development started in 1999 by Rama Ranganathan's group,[27] uses the amino acid distribution in positions of protein sequences alignments and the interdependence of these distributions between positions to identify amino acids with important roles in structure and function in a protein family. A later improvement of the technique consisted in iterative clusterization of $\Delta\Delta G^{stat}$ matrices.[28] The results obtained using clusters analysis show that the relatively restricted groups of residues found with the method connect the functionally important regions of protein families, suggesting that they could represent the core of protein allostery.[28] The technique has a great potential to identify "hot spots" in protein families, with the advantage of using no structural information – only a multiple sequence alignment of the protein/domain family is needed.[27,28]

Here, we report the 2.95 Å resolution crystal structure of yeast holo-PI hexokinase (yhxPI) complexed with glucose in an attempt to present a better defined three-dimensional (3D) model of this yeast hexokinase isoenzyme, and to add more detailed structural information to the "induced fit" in yeast HK enzymatic mechanism. The model is based on the correct amino-acid sequence and the glucose binding site is described in detail. Comparing the present structure with the structure of apo-PII isoenzyme,[29] we were able to quantify conformational changes induced in HK by ligand binding more accurately. The mechanism of reaction, the lack of Glc-6P inhibition in yeast HKs, and the putative mode of ATP binding are also discussed. Furthermore, based on SCA we identified two clusters, which form a continuous 3D network of statistically coupled residues. The cluster with the strongest couplings is composed by amino acid residues of the active site and the hinge regions, further confirming cooperative function of these parts of the protein in the induced fit mechanism of the enzyme function.

MATERIALS AND METHODS

Crystallization

Hexokinase PI isoenzyme was purchased from Sigma-Aldrich and used for crystallization trials without any further purification. Crystals were grown in hanging drops at 277 K.[30] The protein solution was buffered with 100 m*M* potassium phosphate in the pH range 5.5–6.5 and 25–30% PEG as a precipitant. Hanging drops contained 3 μL of the protein at concentrations of 10 mg mL^{-1} and 3 μL of precipitant. Needle shaped crystals appeared between 1 and 4 weeks and with typical crystal dimensions of 0.5 × 0.05 × 0.05 mm^3. Data were collected from these crystals, which diffracted to 2.95 Å resolution.

X-ray diffraction data

Diffraction data were collected using a MAR image plate system at the dedicated protein crystallography beamline of the Brazilian synchrotron light source.[31,32] The data were integrated, scaled, and merged using the programs DENZO and SCALEPACK.[33] A total of 14,357 reflections were collected, with a completeness of 92.27% for all data to 2.95 Å (94.44% for the highest resolution shell). The crystal belonged to the orthorhombic space group $P2_12_12_1$, with cell constants of a = 62.12 Å, b = 78.87 Å, and c = 144.74 Å. Estimation of the solvent content indicated that the crystal contained one molecule per asymmetric unit. The Matthew's coefficient[34] assuming that the asymmetric unit contained a monomer was 3.1 $Å^3$ Da^{-1}, which corresponds to a solvent content of 60%. The crystal structure of the yeast hexokinase PI was determined by molecular replacement using a P152K mutant of yeast hexokinase as the starting model. The program AMoRe[35] was used in calculation of rotation and translation functions based on data collected up to 3 Å. A unique solution was found after rigid-body refine-

ment, with a correlation coefficient of 47.9% and an *R*-factor of 0.53.

Refinement, model building, and overall quality of the model

The structure was then refined using the programs REFMAC[36] and Phenix[37] with working and free (5%) sets of data extended to include all data from 12.9 to 2.95 Å. Thirteen thousand six hundred forty-two reflections were used for the refinement, while 715 reflections, randomly selected and set aside during refinement process, were used for *R*-free calculation. Rounds of positional and group *B*-factors refinement were alternated with model verification using the O program.[38] Given the medium resolution of the X-ray data and the limited number of collected reflections, only one *B*-factor per residue has been refined to improve the observations-to-model parameters ratio. The refinement, which was forced to maintain tight geometry of the structure, was monitored by the decline in *R*-free. Refinement and rebuilding of the model lowered the *R*-factor to a final value of 19.2% and *R*-free to 24.5%. Details of the final stage of the refinement are given in Table I.

The final model contains all but the first 14 residues of N-terminal for which no discernible density could be observed. The average thermal parameter for the model is 44.8 $Å^2$ and the final $(2F_o - F_c)$ electron density is well defined for most residues. The program LSQKAB of the CCP4 package was used for the analysis of root-mean-squares deviations in the positions of Cα atoms of different models. Ramachandran analysis[39] as implemented in PROCHECK[40] shows that 86.7% of the residues lie in the most favored regions, while 13.3% lie in the additionally and generously allowed regions. No residues are found in disallowed regions. The *G*-factor of the model is 0.2. The final model and structure factors were deposited in the Protein Data Bank under the PDB ID 3B8A.

Table I
Crystallographic Information

Resolution range (Å)	12.92–2.95 (3.03–2.95)
Space group	$P2_12_12_1$
Unit cell parameters (Å)	$a = 62.12$, $b = 78.87$, $c = 144.74$
Completeness (%)	92.27 (94.44)
Reflections used in refinement	13,642 (1007)
Total reflections/Reflections for R_{free}	14,357/715
I/σ	8.8 (2.49)
R_{merge} (%)	11.9 (47.7)
Multiplicity	2.8 (2.7)
R_{factor} (%)	19.19
R_{free} (%)	24.45
Bond length RMS deviation (Å)	0.002
Bond angle RMS deviation (°)	0.470
Mean overall *B*-factor	44.8
Overall *B*-factor from Wilson plot	47.4
Mean *B*-factor for protein	44.9
Mean *B*-factor for glucose	29.1
Mean *B*-factor for SO_4	48.71
Mean *B*-factor for water molecules	32.68

Values between parenthesis correspond to the value in the highest resolution bin (a total of 20 bins was used). $R_{factor} = \frac{\sum_{hkl} ||F_o| - |F_c||}{\sum_{hkl} |F_o|}$, R_{free} is defined the same as R_{factor}, except that the summation is over the 715 (about 5%) reflections not included in the refinement.

Statistical coupling analysis

A multiple sequence alignment of hexokinases was obtained from the PFAM database,[41] in a total of 317 sequences, including that of yeast hexokinase PI. From the alignment, it is possible to compute, in a parameter named ΔG^{stat}, the discrepancy between amino acid frequencies in a given position and the amino acid frequencies observed in nature, which shows how strongly this position is conserved in the protein family.[27] A second parameter, termed $\Delta\Delta G^{stat}$, quantifies how much the amino acid distribution of a certain position varies when a specific condition is imposed at another position (i.e., the later position should be occupied by a certain type of amino acid residue). If the amino acid distribution at the former position in the alignment varies (which is reflected in a high value for $\Delta\Delta G^{stat}$), this means that for some reason, normally related to the structure and/or function of the protein, these two positions coevolved together, as suggested by the observed statistical coupling.[27] Site conservation (ΔG^{stat}) and statistical coupling ($\Delta\Delta G^{stat}$) parameters between positions in the alignment were calculated as described in Lockless and Ranganathan, 1999.[27] ΔG^{stat} values for the 20 most conserved residues are shown in Table II. To determine the set of positions to be perturbed in order to calculate $\Delta\Delta G^{stat}$, we used the method used by Ranganathan's group, described in Süel *et al.*, 2003.[28] Monitoring the variance of mean $\Delta\Delta G^{stat}$ for the five least conserved positions of the alignment, which should, by definition, have no evolutionary restraint and therefore low coupling values, it is possible to define a limit for the size of the subalignment used to calculate $\Delta\Delta G^{stat}$. For our case, the subalignment should have at least 35% of the original sequences, resulting in 209 positions which could be perturbed. The initial matrix of $\Delta\Delta G^{stat}$ contained 209 columns (perturbed positions) and 485 lines (sequence size of HK PI). This matrix is sparse, suggesting that during evolution, few positions in a protein evolved together. The iterative analysis consists of clustering rows and columns with similar $\Delta\Delta G^{stat}$ distribution and eliminating low noise rows and columns. The matrix was subjected to successive rounds of clustering using the software package MATLAB and elimination of low-signal rows and columns. During the process, high signal positions grouped together in two clusters. The first cluster consists of a 2×29 matrix (30 amino acid residues positions), with mean $\Delta\Delta G^{stat} = 219$. The second

Table II
Statistical ΔG Parameter for 20 Most Conserved Residues

Ranking	Amino acid residue	ΔG^{stat}	Normalized ΔG^{stat}
1	Trp174	1510.28	100
2	Asn210	1141.41	75.58
3	Cys268	1128.30	74.71
4	Pro160	1114.98	73.83
5	Pro74	1091.75	72.29
6	Arg93	1043.57	69.10
7	Phe157	969.50	64.19
8	Ser158	965.34	63.92
9	Lys176	946.17	62.65
10	Phe240	932.62	61.75
11	Pro115	915.23	60.60
12	Asn237	894.36	59.22
13	Asp211	890.73	58.98
14	Gly307	888.12	58.80
15	Gly80	870.01	57.61
16	Gln163	865.02	57.28
17	Asn91	853.06	56.48
18	Gly88	848.74	56.20
19	Asp86	841.42	55.71
20	Trp128	841.12	55.69

$\Delta G_i^{stat} = \sqrt{\sum_x \left(\ln \frac{P_i^x}{P_{MSA}^x}\right)^2}$. P_i^x is the binomial probability of finding observed frequency of the amino acid x (x = alanine, cysteine, etc.) in the position i of the alignment, given the known frequency observed in nature, while P_{MSA}^x is the binomial probability of finding the frequency of the amino acid in all the multiple sequence alignment (MSA) in that position. As ΔG^{stat} is dependent of the number of sequences, its value has arbitrary units. The highest value for ΔG^{stat}, 1510.28, corresponds to a position which presents 77.3% of tryptophan residues. This is highly improbable to happen by chance, given that the expected frequency for tryptophan residue in proteins in nature is only 1%.

cluster consists in a 7 × 18 matrix (20 amino acid residues positions), with mean $\Delta\Delta G^{stat} = 182$. $\Delta\Delta G^{stat}$ values for all residues in the clusters are shown in Tables III and IV. Residue conservation is shown in Figure 1. This figure was rendered in Pymol[42] after substituting the *B*-factor column of the PDB file with ΔG^{stat} values after normalization and setting the highest *B*-factor to one hundred. The same normalized values were used for all ΔG^{stat} reports throughout the text.

RESULTS AND DISCUSSION

The overall structure

Yeast hexokinase PI is a palm shape molecule and exhibits the same α/β fold observed in other hexokinase structures (Fig. 1).[23–26,29,43–45] The polypeptide chain of 485 residues is distinctly folded into two domains of unequal size: the large domain (residues 17–76 and 212–458) and the small domain (residues 77–211 and 459–485). These are separated by a deep cleft containing the residues making the enzyme active site. The protein is found in closed conformation.[24] The most conserved residues, as revealed by SCA, are located in the protein core; near the glucose binding site (Fig. 1).

The crystallization of yeast HK isoenzyme PI was performed in the absence of a substrate. However, at the early stage of refinement a clear and continuous residual electron density appeared in both ($2F_{obs} - F_{calc}$) and ($F_{obs} - F_{calc}$) Fourier syntheses. This electron density could be readily interpreted in terms of a glucose molecule and a sulfate ion. Sulfate ions were present in 2*M* concentration in the crystallization conditions. The thermal parameters for ligands positioned in that density are comparable with the average thermal parameter for solvent molecules, inferring their substantial occupancy that was set to 1 during structure refinement.

Superposition of the previously determined yeast apo-HK PII model onto holo-HK PI+glucose complex gives an average rms deviation of 2.8 Å for Cα atom pairs with maximum deviation of over 8 Å for several parts of the polypeptide backbone (Fig. 2). Separately comparing the large and the small domain of the structures give an average rms deviations of 1.17 and 1.12 Å, respectively. Glucose-

Table III
Statistically Coupled Clusters of Residues in HK: 2 × 29 Cluster (Mean $\Delta\Delta G^{stat} = 219$)

Perturbation	210	174
G177	183.83	246
K176	208.42	258.82
D211	233.57	209.11
G185	173.03	286.07
D187	188.47	289.25
G169	173.56	242.36
L192	180.86	281.18
L171	184.38	308.2
F178	177.39	254.23
V189	151.94	285.2
V205	174.79	200.84
G214	196.79	320.99
W174	190.13	31.799
T175	184.08	385.97
V188	173.67	357.13
A207	208.62	233.29
T212	209.26	301.42
T215	212.64	254.05
I231	210.7	216.37
N237	222.91	201.56
T233	211.32	195.73
G235	216.96	188.49
N267	184.81	224.26
E269	183.42	202.55
S306	158.21	193.24
Y240	191.58	308.3
E242	186.93	295.07
N295	158.05	263.51
D353	144.84	216.15

Here $\Delta\Delta G_{ij}^{stat} = \sqrt{\sum_x \left(\ln \frac{P_{i\delta j}^x}{P_{MSA|\delta j}^x} - \ln \frac{P_i^x}{P_{MSA}^x}\right)^2}$, P_i^x and P_{MSA}^x as defined before. "Iδj" stands for "upon perturbation of residue j", i.e., the probabilities are calculated for the whole alignment and for a subalignment where a condition is imposed (e.g., j is an alanine). $\Delta\Delta G^{stat}$ is also dependent of alignment size and has arbitrary units. The highest $\Delta\Delta G^{stat}$ found (385.97) was due to the perturbation T175: the presence of a threonine in this position led to the increase of the tryptophan frequency in position 174 from 77.3% to 99.2%.

Table IV
Statistically Coupled Clusters of Residues in HK: 7 × 18 Cluster (Mean $\Delta\Delta G^{stat}$ = 182)

Perturbation	219	226	239	424	421	385	210
L465	100.97	118.83	79.204	154.05	129.92	171.57	106.61
N295	124.1	134.44	212.13	124.03	167.66	217.75	158.05
D353	136.62	130.63	192.62	156.36	194.83	283.3	144.84
G274	238.74	248.74	223.4	278.99	220.37	249.28	90.786
G278	247.85	252.62	237.38	285.86	224.07	252.17	85.223
L368	198.93	252.52	201.67	241.47	191.14	208.59	85.223
V314	195.51	200.59	200.98	233.58	199.33	249.85	97.938
D377	175.37	183.82	226.66	230.85	196.6	260.85	114.86
F342	208.89	213.25	238.89	202.8	196.91	233.71	116.45
L327	146.75	154.7	194.07	173.36	171.5	232.54	96.96
V388	129.51	153.95	190.52	166.66	197.68	248.2	120.39
S455	193.34	206.49	192.6	212.87	216.34	213.56	97.51
C226	291.42	47.623	268.32	223.99	197.66	193.39	104.18
C385	159.71	167.62	200.91	200.7	232.22	46.027	120.08
Y421	180.68	189.4	215.52	218.43	45.66	261.25	109.43
C239	219.48	228.78	43.006	181.8	193.85	201.36	134.1
C219	46.932	302.82	265.92	220.21	193.91	189.78	113.82
T413	150.56	163.33	241.96	141.19	155.37	154.99	86.437

Here $\Delta\Delta G_{ij}^{stat} = \sqrt{\sum_x \left(\ln \frac{P_{i|\delta j}^x}{P_{MSA|\delta j}^x} - \ln \frac{P_i^x}{P_{MSA}^x} \right)^2}$, P_i^x and P_{MSA}^x as defined before. "|δj" stands for "upon perturbation of residue j", i.e., the probabilities are calculated for the whole alignment and for a subalignment where a condition is imposed (e.g., j is an alanine). $\Delta\Delta G^{stat}$ is also dependent of alignment size and has arbitrary units. The highest $\Delta\Delta G^{stat}$ found (385.97) was due to the perturbation T175: the presence of a threonine in this position led to the increase of the tryptophan frequency in position 174 from 77.3% to 99.2%.

induced conformational fit can be described as a rigid-body rotation and nonrigid body conformational changes.

Glucose-induced conformational changes

The main part of the HK fit induced by glucose is a rigid body rotation of one part of the molecule with respect to the other. Following the analysis by Hayward and Berendsen,[46] apo-HK PII and holo-HK PI+glucose complex structures comparison permits identification of the dynamic domains of the protein. The whole structure can be divided into the fixed domain, comprising residues 20–59, 61–73, 104–106, 211, 214–259, 262–308, 310–458, 467–482 and a rotating domain made of residues 74–103, 107–210, 212–213, 459–466. The definition of the fixed and rotating domains approximately corresponds to the traditional definition of the large and small domain, respectively. Residues 73–74, 104–107, 211–212, 213–215, 458–

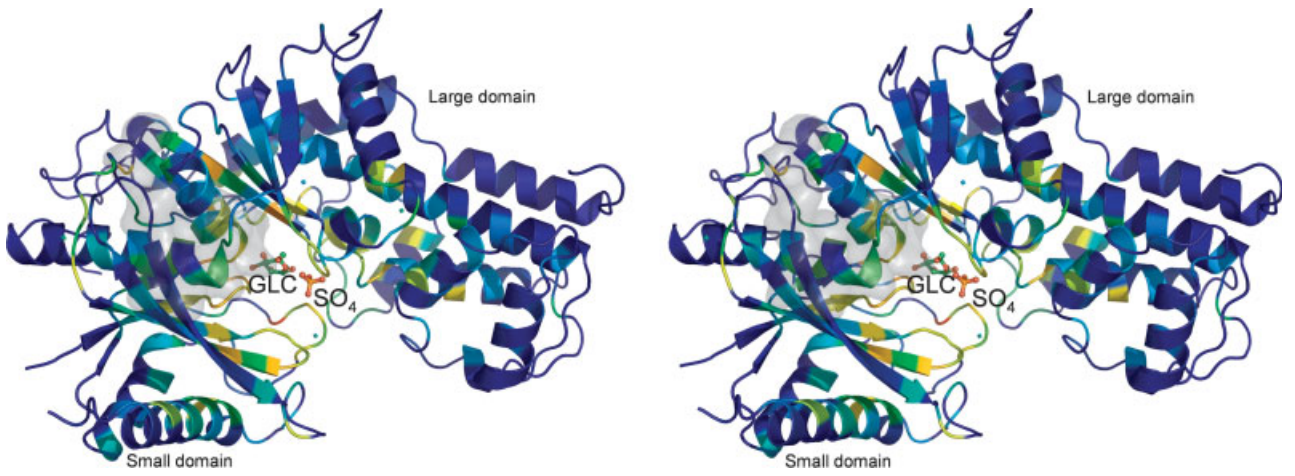

Figure 1
Overall fold of holo PI hexokinase. The protein is colored by residue conservation (ΔG): poorly conserved positions (low ΔG) are shown in "cold" colors, highly conserved positions (high ΔG) are shown in "hot" colors. The gray surface corresponds to the regions of the mechanical hinges, connecting the two domains.

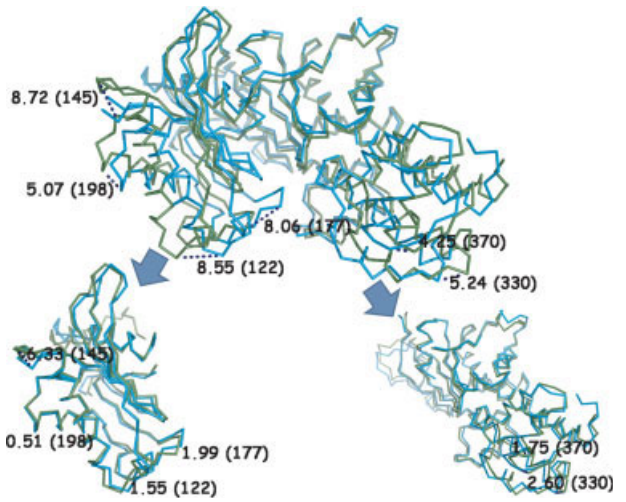

Figure 2

Superposition of the previously determined yeast apo-HK PII model onto holo-HK PI+glucose complex. The superposition of the whole structure leads to high Cα deviances in a considerable portion of the protein. These deviances fall dramatically when the two domains are superposed separately. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

459, and 466–467, noteworthy, highly conserved Asp211, Thr213, Gly215, and Gly458 (normalized ΔG^{stat}'s are, respectively, 58.9, 30.5, 51.4, and 32.9), act as "mechanical hinges." The moving domain rotates with respect to the axis that is approximately perpendicular to the straight line between the centers of masses of the two domains in the direction of the fixed domain by an angle of 17°. This is a surprisingly large rotation angle, taking into consideration that previous estimation, done on the basis of superposition of earlier HK models, was 12°.[22–26] Conformational changes result in movements of the polypeptide backbone of as much as 11.5 Å, which is also considerably larger than the maximum displacement of 8 Å reported before. To exclude possible computational differences related to the slightly different definition of the moving and fixed domains, we repeated the calculation of the rigid body rotation based on the earlier HK models.[22–26] The result of such calculation, a rotation angle of 13.6°, is close to the value reported by Steitz et al. and Bennett and Steitz.[23,24] The rest of the difference in the angle of HK domain closure, induced by binding of glucose, can be attributed to the fact that the early HK PI structure was determined as a complex with competitive inhibitor, ortho-toluoylglucosamine,[25,26] which induces a slight closure of the domains,[29] leading to underestimation of the glucose-induced conformational changes of HK.

In addition to the rigid-body domain rotation, several regions of the polypeptide chain are involved in nonrigid conformational changes. These changes, however, are mostly confined to the surface loop regions, such as loops 142–149 and 470–478 of the small domain and loops 245–262, 338–342, and 445–448 of the large domain and might be induced by crystallographic contacts or are result of the single amino acid insertion (loop 445–448) (Fig. 2).

Surprisingly, virtually all the conformational changes in the active site region, including loop regions, could be described as a rigid-body rotation of the small domain of the protein with respect to its large domain.

The ligand binding site

The glucose smugly fits into the cleft between two HK domains (Fig. 3) and interacts with the HK binding site

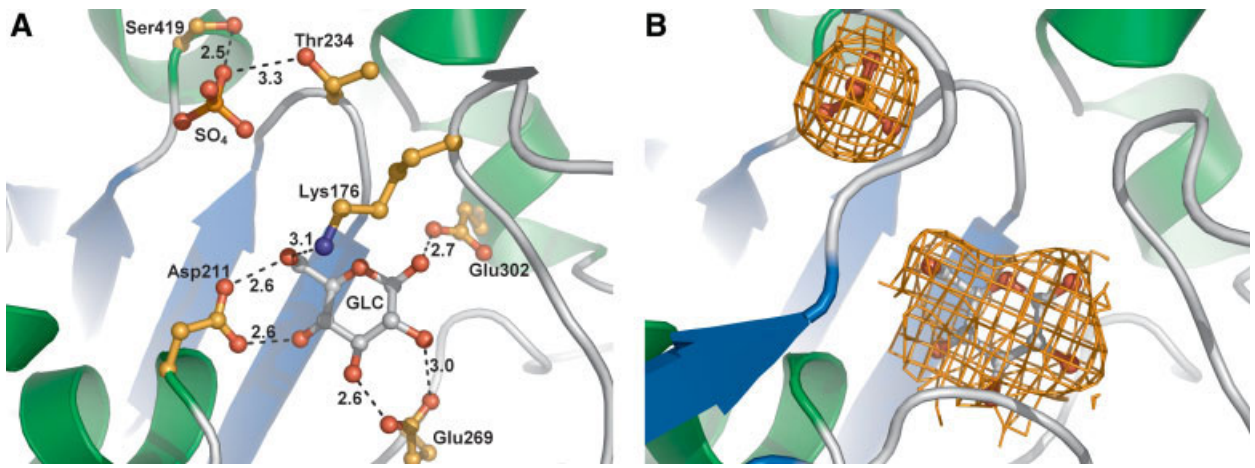

Figure 3

A: *Interactions between PI hexokinase, glucose and a sulfate anion in the binding site.* ***B:*** *The glucose molecule and the sulfate anion superimposed with the OMIT $2F_o - F_c$ maps, contoured at 1σ. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]*

via well-defined interactions at the positions of 1-, 3-, 4-, 5-, and 6-OH groups. This is in agreement with substrate specificity studies of Sols and Crane[47], which implicate the participation of all but the 2-OH group of glucose in interactions with the mammalian HK. The β-enantiomer of glucose is more consistent with the electron density and ligand interactions. Although glucose α-enantiomer does not produce a steric hindrance, the distance between Glu 302 Oε2 and the 1-OH group becomes too long for strong hydrogen bond interaction.

The highly conserved Asp211 (normalized ΔG = 58.9, 13th highest for the protein), implicated as a catalytic base,[24,25] makes a hydrogen bond with the hydroxyls at the positions 4 and 6 of glucose, whereas side chains of the amino acid residues Glu302, Glu269, and Lys176 are at the hydrogen bonding distances from the 1-OH, 3-OH, and 5-OH groups of glucose molecule, respectively. Lys176 makes an additional hydrogen bond with the 6-hydroxyl group of glucose as well. The absence of interactions with 2-OH group at the glucose-binding pocket is consistent with the fact that 2-subsituted derivatives of glucose serve as HK substrate. Site-directed mutations of Asp211 cause loss of catalytical activity, but high sugar binding affinity is retained.[48] This indicates that sugar binding is mainly coordinated by Glu302 and Glu269, whereas Asp211 is indispensable for promoting a nucleophilic attack of the glucose 6-OH group on the γ-phosphate of ATP.[25] Residues Glu302 and Glu269 make part of the large (or fixed) domain, Asp211 belongs to a hinge region, whereas Lys176 comes from a small (rotating) domain. Obviously, the glucose binding pocket becomes complete only upon the closure of HK domains. Therefore, one can speculate that initial glucose binding to HK, mediated by interactions with the side chains of amino acid residues Asp211, Glu302, and Glu269, triggers large-scale conformational rotation of the HK domains with respect one another, bringing Lys176 side chain into hydrogen bonding distance from the glucose hydroxyl groups 4 and 6 and thus creating the competent glucose binding site.

The glucose molecule binds in a manner structurally analogous to that observed for other hexokinase complexes, maintaining the glucose binding site on our model in close agreement with the binding site as previously elucidated.[16–19] However, several important differences exist. Strongly conserved Ser158 (normalized ΔG^{stat} = 63.9, 8th mostly conserved residue in the protein), previously described as a conserved glucose-binding residue,[24] is at the distance of 3.86 Å from the 6-hydroxyl group of glucose, which is somewhat too long for a strong hydrogen binding. At the same time, directed mutation studies identify Ser158 as a critical determinant of the catalytical activity and conformation of HK.[49,50] Furthermore, it was shown that binding of the certain nonphosphophorylatable inhibitors of glucose such as D-xylose or D-lyxose promotes strikingly the short-lived hydrolytic activity of the enzyme followed by inactivation of the enzyme via autophosphorylation of Ser158.[17,51–53] All these evidences taken together support the idea that additional conformational changes might occur upon binding of ATP[54–56] and that Ser158 might play in important role in the process of the phosphate transfer.

The sulfate group, observed in the active site of the holo-HK PI structure, interacts with conserved residues Ser419 and Thr234 (normalized ΔG^{stat}'s are 37.6 and 47.8) of the fixed HK domain (Fig. 3). Its position is identical to the one of sulfate anion in the HK PII structure.[30] The presence of a sulfate ion in the structure of HK both in open and in close conformations proves that, first, glucose binding is not required for sulfate/phosphate binding and, second, sulfate/phosphate binding alone is not sufficient to provoke significant conformational changes of the enzyme. Furthermore, the fact that the same sulfate/phosphate anion binding site was repeatedly found in yeast[23–26,29] and human[43–45] hexokinase models shows that HK ATP binding site is able to bind monophosphates. This might have functional importance. It is known, for example, that glucose binding was strongly promoted in the presence of 0.05*M* phosphate.[52] Phosphate binding might therefore somewhat restrict the conformations of the amino acid residues of the HK binding site and thus facilitating glucose binding.

Insights about the lack of HK PI inhibition by Glc-6P are obtained after comparison to the Glc-6P bound structure of human hexokinase I. For the human HK, Glc-6P contacts the loops 86–93 and also 454–460, which are on different conformations when compared with its structural analogs in yeast HK PI. The phosphate group of the Glc-6P interacts with the Thr536 OG1 and a main chain nitrogen atom of the same residue (loop 532–539) whereas in yeast HK interactions of the Glc-6P with the correspondent loop (residues 86–93). Correspondent Thr90 residue is more than 8 Å away. It has also been found that, for yeast HK PI, residue Asp211 is at a position which could form an hydrogen bond with the 4-OH group of Glc-6P, but the interaction between 2-OH of Glc-6P with Ser897 in human hexokinase I does not have an equivalent in yeast HK PI.

The putative ATP binding site

The ATP pattern recognition, identified on the basis of comparison of actin,[57] ATPase fragment of 70 K heat-shock cognate protein[58] and HK three-dimensional structures, was defined as consisting of five sequence motifs: Phosphate1 (residues 82:103), Connect1 (residues 203:223), Phosphate2 (residues 229:248), Adenosine (residues 411:439), and Connect2 (residues 452:472).[59] Residues Thr234 and Ser419 belong, respectively, to the Phosphate2 and Adenosine ATPase pattern recognition motifs. The α-phosphoryl group of ADP occupies the

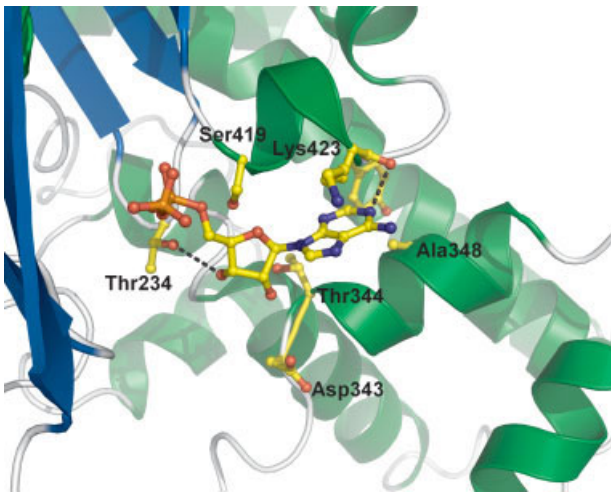

Figure 4

Interactions between PI hexokinase ADP, as suggested by modeling an ADP molecule by analogy to the human hexokinase I-ADP/Glu complex. The putative interactions shown in the figure as dashed lines are between 2.5 and 3.5 Å distance limits. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

same site as the 6-phosphoryl group of glucose-6-phosphate in human HK PI structure.[56,60]

An ADP molecule was modeled in yeast hexokinase PI by analogy to the position of an ADP molecule in the structure of the ADP/glucose complex of human hexokinase I (Fig. 4). The base moiety of ADP is sandwiched between helices α9 and α11′. The model suggests that the binding site of ADP is equivalent in both structures. In yeast hexokinase PI both helices are suitably positioned to accommodate the ADP molecule, and there are no steric clashes between the protein and the ADP molecule. However, to maintain the same interactions with the side chain residues observed in the human hexokinase I, helix α9 would have to move closer to the ADP molecule. This movement could be triggered by the binding of ATP to the molecule.

The modeling also shows that in yeast hexokinase PI the β-phosphoryl group points out in the direction of the 6-hydroxyl group of glucose in a position and orientation adequate for productive binding without being blocked by loop L1 as it happens in the human hexokinase I. A γ-phosphoryl group could equally be modeled into a productive conformation.

Cluster analysis of SCA data

Our SCA data for position coupling in the hexokinase family were treated with elimination of low signal rows and columns alternated with clusterization in MATLAB. The positions with high $\Delta\Delta G^{stat}$ were grouped together in two different clusters. Highly coupled clusters in SCA studies comprise the positions that evolved together in a protein family, a fact that might have happened for structural and functional reasons. The highest $\Delta\Delta G^{stat}$ cluster we have found comprises the regions of the protein that we described as crucial to our induced fit model, that is, the binding site and the hinges which allow the rearrangement of the two domains (Fig. 1). The fact that the active site residues show strong statistical coupling with the hinge regions is very consistent with the "induced fit" mechanism of enzymatic catalysis of the HK which

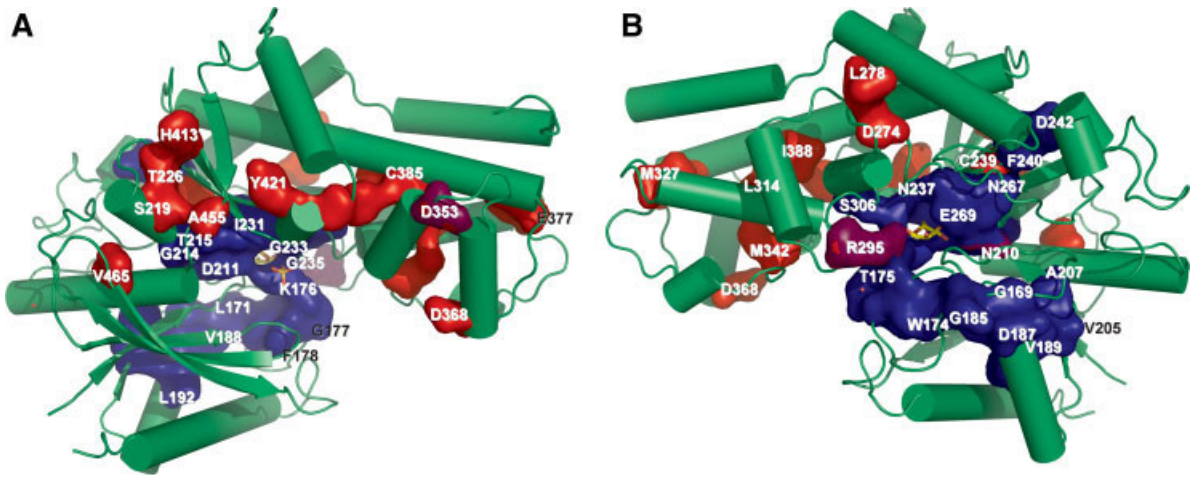

Figure 5

A: *A structural view of the clusters of highly evolutionarily coupled positions in hexokinases. The cluster of stronger coupled residues (mean $\Delta\Delta G^{stat}$ = 219) is shown in blue, while the cluster with the weaker coupling values (mean $\Delta\Delta G^{stat}$ = 182) is shown in red. The residues shown in mixed blue/red (R295, D353 N210) are present in both clusters;* ***B:*** *The same as (A) upon 180° rotation around Y axis.*

implies a necessity of simultaneous conservation of these important parts of the protein. Another cluster, with lower coupling values between positions, was also found, and consists of disperse protein positions [Figs. 5(A,B)]. The structural and functional role of this second cluster remains elusive.

CONCLUSIONS

Yeast hekokinase is a classic example of enzyme with the "induced fit" mechanism of reaction. The present crystallographic structure of the yeast HK PI bound to glucose reveals the details of substrate recognition and shows that conformational changes of this enzyme upon ligand binding ("induced fit") are somewhat larger than considered earlier (17° vs. 12°). The allosteric network involved in the induced fit mechanism, as suggested by our SCA analysis, is confined to the vicinity of the glucose binding site and the hinge region. A secondary network of residues which co-evolved with lower coupling values was also found, but its meaning remains to be determined. The fact that the residues in the binding site and in the mechanical hinges co-evolved together, as suggested by SCA, sheds light into the development of its mechanism of action – the ability to bind glucose and rotate the two domains were probably developed simultaneously during evolution of the protein family, suggesting that the induced fit mechanism was a necessity to protein function.

In addition, the X-ray structure of holo-PI HK allows for rationalization of the wealth of biochemical information accumulated to date. Moreover, the presence of a bound glucose molecule and, consequently, more compact conformation of the HK PI obtained from commercial preparations, requires a significant revision in the interpretation of past thermal denaturation experiments.

ACKNOWLEDGMENT

Authors acknowledge Rama Ranganathan's group for support with the statistical coupling analysis methods.

REFERENCES

1. Kopetzki E, Entian K, Mecke D. Complete nucleotide sequence of the hexokinase PI gene (HXK1) of *Saccharomyces cerevisiae*. Gene 1985;39:95–101
2. Fröhlich K, Entian K, Mecke D. The primary structure of the yeast hexokinase PII gene (HXK2) which is responsible for glucose repression. Gene 1985;36:105–111
3. Gancedo JM, Clifton D, Fraenkel DG. Yeast hexokinase mutants. J Biol Chem 1977;252:4443–4444.
4. Entian KD. Genetic biochemical evidence for hexokinase PII as a key enzyme involved in carbon catabolite repression in yeast. Mol Gen Genet 1980;178:633–637.
5. De Winde JH, Crauwels M, Hohmann S, Thevelein JM, Winderickx J. Differential requirement of the yeast sugar kinases for sugar sensing in establishing the catabolite-repressed state. Eur J Biochem 1996;241:633–643.
6. Fernandez R, Herrero P, Fernandez MT, Moreno F. Mechanism of inactivation of hexokinase PII of *Saccharomyces cerevisiae* by D-xylose. J Gen Microbiol 1986;132:3467–3472.
7. Vojtek AB, Fraenkel DG. Phosphorylation of yeast hexokinases. Eur J Biochem 1990;190:371–375.
8. Behlke J, Heidrich K, Naumann M, Muller EC, Otto A, Reuter R, Kriegel T. Hexokinase 2 from *Saccharomyces cerevisiae*: regulation of oligomeric structure by *in vivo* phosphorylation at serine-14. Biochemistry 1998;37:11989–11999.
9. Randez-Gil F, Sanz P, Entian KD, Prieto JA. Carbon source-dependent phosphorylation of hexokinase pii its role in the glucose-signaling response in yeast. Mol Cell Biol 1998;18:2940–2948.
10. Mayordomo I, Sanz P. Hexokinase PII: structural analysis glucose signaling in the yeast *Saccharomyces cerevisiae*. Yeast 2001;18:923–930.
11. Herrero P, Martinez-Campa C, Moreno F. The hexokinase II protein participates in regulatory DNA-protein complexes necessary for glucose repression of the SUC2 gene in *Saccharomyces cerevisiae*. FEBS Lett 1998;434:71–76.
12. Gancedo JM. Yeast carbon catabolite repression. Microbiol Mol Biol Rev 1998;62:334–361.
13. Rose M, Albig W, Entian KD. Glucose repression in *Saccharomyces cerevisiae* is directly associated with hexose phosphorylation by hexokinases PI PII. Eur J Biochem 1991;199:511–518.
14. Lazarus NR, Ramel AH, Rustum YM, Barnard EA. Yeast hexokinase. I. Preparation of the pure enzyme. Biochemistry 1966;5:4003.
15. Feldman I, Kramp DC. Fluorescence-quenching study of glucose binding by yeast hexokinase isoenzymes. Biochemistry 1978;17:1541–1547.
16. Wilson JE. Hexokinases. Rev Physiol Biochem Pharmacol 1995;126:65–198.
17. Colowick SP. The hexokinases. In: Boyer PD, editor. The Enzymes, Vol. 9. New York: Academic; 1973. pp 1–48.
18. Katzen HM, Schimke RT. Multiple forms of hexokinase in the rat: tissue distribution, age dependency, properties. Proc Natl Acad Sci 1965;54:1218–1225.
19. Easterby JS, O'Brien MJ. Purification properties of pig-heart hexokinase. Eur J Biochem 1973;38:201–211.
20. Holroyde MJ, Trayer IP. Purification properties of skeletal muscle hexokinase. FEBS Lett 1976;62:215–219.
21. Koshland DE. Mechanisms of transfer enzymes. In: Boyer PD, Lardy H, Myrbäck K, editors. The Enzymes, 2nd ed., Vol. 1. New York: Academic Press; 1959. pp 305–346.
22. Stryer L. Biochemistry. New York: WH Freeman Company; 1995. p 499.
23. Steitz TA, Fletterick RJ, Anderson WF, Anderson CM. High resolution X-ray structure of yeast hexokinase, an allosteric protein exhibiting a non-symmetric arrangement of subunits. J Mol Biol 1976;104:197–222.
24. Bennet WS, Jr, Steitz TA. Structure of a complex between yeast hexokinase A glucose. I. Structure determination refinement at 3.5 A resolution. J Mol Biol 1980;140:183–209, 211–230.
25. Anderson CM, Stenkamp RE, McDonald RC, Steitz TA. Sequencing a protein by X-ray crystallography. II. Refinement of yeast hexokinase B co-ordinates sequence at 2.1 A resolution. J Mol Biol 1978;123:15–34.
26. Anderson CM, Zucker FH, Steitz TA. Space-filling models of kinase clefts conformation changes. Science 1979;204:375–380.
27. Lockless SW, Ranganathan R. Evolutionarily conserved pathways of energetic connectivity in protein families. Science 1999;286:295–299.
28. Süel GM, Lockless SW, Wall MA, Ranganathan R. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat Struct Biol 2003;10:59–69.
29. Kuser PR, Krauchenco S, Antunes OAC, Polikarpov I. The high resolution crystal structure of yeast hexokinase PII with the correct primary sequence provides new insights into its mechanism of action. J Biol Chem 2000;275:20824–20831.

30. Kuser PR, Golubev AM, Polikarpov I. Crystallization preliminary crystal analysis of yeast hexokinase PI and PII. Acta Crystallogr D Biol Crystallogr 1999;55:2047–2048.
31. Polikarpov I, Perles LA, de Oliveira RT, Oliva G, Castellano EE, Garrat RC, Craievich A. Set-up experimental parameters of the protein crystallography beamline at the Brazilian National Synchrotron Laboratory. J Synchrotron Rad 1998;5:72–76.
32. Polikarpov I, Oliva G, Castellano EE, Garrat RC, Arruda P, Leite A, Craievich A. The protein crystallography beamline at LNLS, the Brazilian National Synchrotron Light Source. Nucl. Instrum Methods A 1998;405:159–164.
33. Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. Methods Enzymol 1997;276:307–326.
34. Matthews BW. Solvent content of protein crystals. J Mol Biol 1968;33:491–497.
35. Navaza J. Implementation of molecular replacement in AMoRe. Acta Crystallogr D Biol Crystallogr 2001;57:1367–1372.
36. Murshudov GN, Vagin AA, Dodson EJ. Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 1997;53:240–255.
37. Adams PD, Grosse-Kunstleve RW, Hung L-W, Ioerger TR, McCoy AJ, Moriarty NW, Read RJ, Sacchettini JC, Sauter NK, Terwilliger TC. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr 2002; 58:1948–1954.
38. Jones TA, Zou JY, Cowan SW, Kjeldgaard M. Improved methods for building protein models in electron density maps the location of errors in these models. Acta Crystallogr A 1991;47:110–119.
39. Ramachandran NG, Sasisekharen V. Conformation of polypeptides proteins. Adv Protein Chem 1968;23:283–437.
40. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.
41. Finn RD, Mistry J, Schuster-Böckler B, Griffiths-Jones S, Hollich V, Lassmann T, Moxon S, Marshall M, Khanna A, Durbin R, Eddy SR, Sonnhammer ELL, Bateman A. Pfam: clans, web tools services. Nucleic Acids Res 2006;34,D247–D251.
42. DeLano WL. The PyMol Molecular Graphics System. San Carlos, CA, USA: DeLano Scientific; 2002. Also available at http://www.pymol.org.
43. Mulichack AM, Wilson JE, Padmanabhan K, Garavito RM. The structure of mammalian hexokinase-1. Nat Struct Biol 1998;5:555–560.
44. Aleshin AE, Zeng C, Bourenkov GP, Bartunik HD, Fromm HJ, Honzatko RB. The mechanism of regulation of hexokinase: new insights from the crystal structure of recombinant human brain hexokinase complexed with glucose glucose-6-phosphate. Structure 1998;6:39–50.
45. Rosano C, Sabini E, Rizzi M, Deriu D, Murshudov G, Biachi M, Serafini G, Magnani M, Bolonesi M. Binding of non-catalytic ATP to human hexokinase I highlights the structural components for enzyme-membrane control. Structure 1999;7:1427–1437.
46. Hayward S, Berendsen HJC. Systematic analysis of domain motions in proteins from conformational change: New results on citrate synthase T4 lysozyme. Proteins 1998;30:144–154.
47. Sols A, Crane RK. Substrate specificity of brain hexokinase. J Biol Chem 1954;210:581–595.
48. Kraakman LS, Winderickx J, Thevelein J, de Winde JH. Structure-function analysis of yeast hexokinase: structural requirements for triggering cAMP triggering catabolite repression. J Biochem 1999; 343:159–168.
49. Arora KK, Filburn CR, Pedersen PL. Glucose phosphorylation. Site-directed mutations which impair the catalytic function of hexokinase. J Biol Chem 1991;266:5359–5362.
50. Xu LZ, Harrison RW, Weber IT, Pilkis SJ. Human β-cell glucokinase. Dual role of Ser-151 in catalysis hexose affinity. J Biol Chem 1995;270:9939–9946.
51. DelaFuente G, Sols A. The kinetics of yeast hexokinase in the light of the induced fit involved in the binding of its sugar substrate. Eur J Biochem 1970;16:234–239.
52. DelaFuente G. Specific inactivation of yeast hexokinase induced by xylose in the presence of a phosphoryl donor substrate. Eur J Biochem 1970;16:240–254.
53. Heidrich K, Otto A, Behlke J, Rush J, Wenzel KW, Kriegel T. Autophosphorylation-inactivation site of hexokinase 2 in *Saccharomyces cerevisiae*. Biochemistry 1997;36:1960–1964.
54. Shoham M, Steitz TA. Crystallographic studies model building of ATP at the active site of hexokinase. J Mol Biol 1980;140:1–14.
55. Shoham M, Steitz TA. The 6-hydroxymethyl group of a hexose is essential for the substrate-induced closure of the cleft in hexokinase. Biochim Biophys Acta 1980;705:380–384.
56. Aleshin AE, Kirby C, Liu X, Bourenkov GP, Bartunik HD, Fromm HJ, Honzatko RB. Crystal structures of mutant monomeric hexokinase I reveal multiple ADP binding sites conformational changes relevant to allosteric regulation. J Biol Mol 2000;296:1001–1015.
57. Kabsch W, Mannherz HG, Suck D, Pai EF, Holmes KC. Atomic structure of the actin:DNase I complex. Nature 1990;347: 37–44.
58. Flaherty KM, DeLuca-Flaherty C, McKay DB. Three-dimensional structure of the ATPase fragment of a 70K heat-shock cognate protein. Nature 1990;346:623–628.
59. Bork P, Sander C, Valencia A. An ATPase domain common to prokaryotic cell cycle proteins. Sugar kinases, actin, hsp70 heat shock proteins. Proc Natl Acad Sci 1992;89:7290–7294.
60. Aleshin AE, Zeng C, Bartunik HD, Fromm HJ, Honzatko RB. Regulation of hexokinase I: crystal structure of recombinant human brain hexokinase complexed with glucose phosphate. J Mol Biol 1998;282:345–357.