The evolution of an allosteric site in phosphorylase

Virginia L Rath†*, Kai Lin, Peter K Hwang and Robert J Fletterick

Background: Glycogen phosphorylases consist of a conserved catalytic core onto which different regulatory sites are added. By comparing the structures of isozymes, we hope to understand the structural principles of allosteric regulation in this family of enzymes. Here, we focus on the differences in the glucose 6-phosphate (Glc-6-P) binding sites of two isozymes.

Results: We have refined the structure of Glc-6-P inhibited yeast phosphorylase *b* to 2.6 Å and compared it with known structures of muscle phosphorylase. Glc-6-P binds in a novel way, interacting with a distinct set of secondary elements. Structural links connecting the Glc-6-P binding sites and catalytic sites are conserved, although the specific contacts are not.

Conclusions: Our comparison reveals that the Glc-6-P binding site was modified over the course of evolution from yeast to vertebrates to become a bi-functional switch. The additional ability of muscle phosphorylase to be activated by AMP required the recruitment of structural elements into the binding site and sequence changes to create a binding subsite for adenine, whilst maintaining links to the catalytic site.

Address: Department of Biochemistry and Biophysics, University of California, San Francisco, CA 94143-0448, USA.

†Present address: Pfizer Central Research, Eastern Point Road, Groton, CT 06340, USA.

*Corresponding author.

Key words: allosteric, crystal structure, glucose 6-phosphate, molecular replacement, phosphorylase, yeast

Received: **6 Nov 1995**
Revisions requested: **24 Nov 1995**
Revisions received: **17 Jan 1996**
Accepted: **2 Feb 1996**

Structure 15 April 1996, **4**:463–473

Introduction

The glycogen phosphorylases provide a highly conserved enzymatic function which is diversely regulated in different cell types. By comparing their kinetic parameters and structures, this family of enzymes provides a unique opportunity to understand the structural details underlying these different methods of regulation. Phosphorylases catalyze the phosphorylytic degradation of glycogen, a storage polymer, to yield glucose 1-phosphate for glycolysis. Some 14 sequences from diverse organisms are known. Each has a conserved catalytic core onto which different regulatory structures have been added. These regulatory features have evolved to meet the specific needs of different cell types and include phosphorylation and allosteric control by metabolites.

The enzyme from rabbit muscle has been studied in the most depth, both biochemically and structurally. In the absence of effectors, the unphosphorylated enzyme, phosphorylase *b* (P*b*), is inactive. The covalent phosphorylation of a single serine residue in the N terminus results in the active form of the enzyme, phosphorylase *a* (P*a*). P*b* may also be activated by AMP binding to an allosteric effector site, termed the nucleotide-binding site, adjacent to the phosphorylation site. AMP activation of P*b* is completely reversed by ATP or glucose 6-phosphate (Glc-6-P) binding to the nucleotide-binding site. The muscle enzyme is inhibited by glucose, which binds in the catalytic site, and by diverse heterocycles, such as caffeine, which bind just outside the active site.

The regulation of yeast phosphorylase is comparatively simple. The phosphorylation site is unrelated to that of the muscle enzyme and resides in a 39 residue N-terminal extension unique to the yeast enzyme. Neither glucose, caffeine, AMP or ATP alter the activity of the enzyme. Yeast phosphorylase is inhibited solely by Glc-6-P which binds to the nucleotide effector site identified in the muscle enzyme.

During the course of evolution from a unicellular organism to vertebrate muscle, the Glc-6-P inhibitor site became a bi-functional switch, capable of binding both activators and inhibitors. What structural changes were required to accomplish this functional change? We attempt to answer this question by comparing the refined 2.6 Å structure of the unphosphorylated, inactive form of phosphorylase from *Saccharomyces cerevisiae* complexed with the inhibitor Glc-6-P with known structures of rabbit muscle phosphorylase.

The purification of the enzyme and crystallization of the complex with Glc-6-P have already been described [1]. In addition, we have published a preliminary description of the structure and speculated on the mechanism of activation by phosphorylation [2]. Here we provide the details of the structure solution and refinement and describe the Glc-6-P binding site with reference to the binding modes of Glc-6-P and AMP to muscle phosphorylase. We omit a discussion of the active site and domain interface, as these will be reported elsewhere (KL, PKH and RJF, unpublished data).

Results and discussion

Structure solution by molecular replacement

We have previously described the data collection and reduction [2]. Briefly, the protein crystallizes in the space

group $P4_12_12$, with cell constants a=b=161.12 Å, and c=175.49 Å. The asymmetric unit contains the 205 kDa functional homodimer; a crystallographic twofold axis results in a tetramer. Each subunit binds one molecule of Glc-6-P. We used the known structure of rabbit muscle phosphorylase *b* as a model for molecular replacement [3]. Because of the high symmetry space group and alterations in the packing arrangement of the domains and subunits, the solution of the structure by the molecular replacement method was challenging and we therefore describe it in some detail.

Based on the measured density (50% solvent content) we expected the asymmetric unit to contain two subunits of yeast phosphorylase (the functional dimer) related by a non-crystallographic symmetry operator [1]. The real-space Patterson search method in X-PLOR [4] was used to calculate the self- and cross-rotation functions [5]. We were unable to determine the position of the twofold rotation axis using the self-rotation function and vectors 5–25 Å in length, presumably because of the weak signal.

We expected to find two solutions to the cross-rotation function, one for each of the orientations of the twofold symmetric search model. By systematically varying the resolution range and vector sets, we learned that the function was particularly sensitive to the upper and lower Patterson vector limits. Surprisingly, the best results were obtained from a cross-rotation function calculated using vector radii between 20 Å and 60 Å in length. Results were consistent over several resolution ranges (8–4.5 Å, 10–4.5 Å and 10–6 Å), meaning that one of the twofold-related solutions was the highest or second highest value of the rotation function. For example, using data from 10–6 Å resolution, one (θ_1=57°, θ_2=57°, θ_3=39°, 4.2σ above the mean) was the highest and the other (θ_1=33°, θ_2=57°, θ_3=51°, 4.1σ above the mean) was the third highest value of the rotation function.

Using a single subunit as a search model and the conditions given above, the correct solution ranked tenth in the rotation function. Using higher resolution data (8–4 Å resolution, same vector set) the correct solution had the top value of the rotation function (7.1σ above the mean; next highest solution, 6.1σ above the mean). Although the monomeric search model accounted for only one third of the atoms in the asymmetric unit, the correct solution was more easily identified using it. This result supported the idea that the subunits in the yeast enzyme were oriented differently from those in the muscle enzyme. A comparison of the amino acid sequences from yeast and muscle phosphorylase shows that residues involved in contacts between the two subunits are not conserved, and further, that the domain interface within each subunit is poorly conserved. It was therefore likely that the relative orientation between the two domains and subunits in the yeast phosphorylase dimer would differ significantly from the search model.

Using the optimized value of 20 Å for the minimum vector length, we were subsequently able to obtain a solution to the self-rotation function which was consistent with the results of the cross-rotation function. For data in the 10–6 Å resolution shell and vector radii 20–45 Å, the correct solution was 1.0σ above the mean. The next highest solution was 0.9σ above the mean. Located at ψ=46.4°, ϕ=17.7°, κ=180°, the non-crystallographic operator is very close to a crystallographic twofold symmetry operator, x=y at z=½. We conclude that medium length interatomic vectors provide the best discrimination in these searches and note that the approximate dimensions of the dimer are 65 Å×45 Å×120 Å.

To compensate for the differences in the domain and subunit orientations between the search model and the target, Patterson correlation (PC) refinement was used to improve and confirm the solution to the cross-rotation function, by minimization of the relative orientations of the two domains comprising each subunit and the two subunits. The optimized solution to the rotation function (θ_1=57.41°, θ_2=58.02°, θ_3=37.73°) is slightly less than 2° from the starting value, and the corresponding correlation coefficients before and after PC refinement were 0.010 and 0.063, respectively.

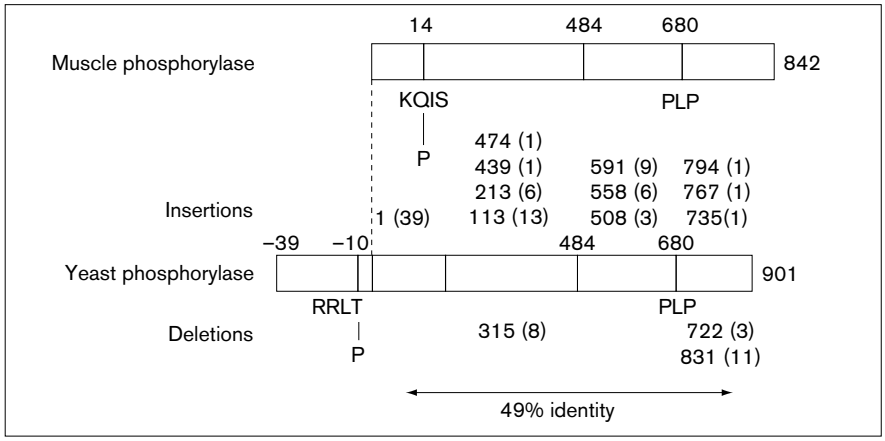
PC refinement of the peaks from the rotation function was clearly advantageous. Refinement of the top 200 peaks resulting from an arbitrarily chosen cross-rotation function confirmed the results obtained by consensus after varying the function parameters. Peaks ranked 149 and 157 before PC refinement (i.e. with rotation function values well below the mean) refined to the correct solution. The independent confirmation of the solution by PC refinement was important because it was difficult to determine parameters which gave consistent results for the cross-rotation function.

The translation search was carried out by computing the standard linear correlation coefficient between the squares of the normalized and calculated structure factors. To resolve the ambiguity in the handedness of the fourfold symmetry operator, it was necessary to calculate a translation function for both space groups $P4_12_12$ and $P4_32_12$. Using a 1 Å grid and data from 10–4 Å resolution, the solution peak in space group $P4_12_12$ was 38σ above the mean and 7σ above the next solution. The overall R-factor for the correct solution was 0.47 for all data between 3.5 Å and 8 Å resolution. A solution (16σ above the mean and 1σ above the next highest solution) was present in nearly the same place in the enantiomer but gave an overall R-factor of 0.51.

We also calculated a translation search using the unrefined orientation derived from the cross-rotation function, using data from 10–6 Å and a grid of 2 Å. The top peak in the resulting map corresponded to a solution (x=0.013, y=0.500, z=0.256), which was close to the result obtained

Figure 1

Schematic comparison of yeast and rabbit muscle phosphorylase amino acid sequences. (To facilitate comparison with the rabbit muscle crystal structure, we use the rabbit muscle numbering system with insertions indicated by alphabetic suffixes starting with the letter C (e.g. 113C).) Insertions and deletions in the yeast sequence relative to rabbit muscle phosphorylase are listed above and below the box, respectively, by residue number followed by the total number of amino acids involved. The N-terminal extension of yeast phosphorylase is designated by numbering backwards from −1 to the yeast N terminus at −39. Lys680 forms a Schiff base with the cofactor pyridoxal phosphate (PLP) and marks the location of the active site. Residue 484 divides the N- and C-terminal domains. The phosphorylation sites are indicated (P). The two enzymes are 49% identical over alignable sequences.



from the refined model (x=0.019, y=0.491, z=0.256). The peak was less significant (8.2σ above the mean) and the R-factor was correspondingly higher (0.59).

Refinement

The structure was refined using X-PLOR [4], omitting 10% of the data for a free R-factor analysis [6]. Refinement consisted of alternating rounds of simulated annealing as described [7], and manual rebuilding of the model using the molecular graphics programs FRODO [8] and O [9] into Sim-weighted $2F_{obs}–F_{calc}$ electron-density maps using a twofold averaged Sim-weighted $2F_{obs}–F_{calc}$ electron-density map for reference. The model has undergone seven cycles of rebuilding and eight cycles of simulated annealing refinement. The location of the twofold axis was determined after cycle 2 by ignoring the non-crystallographic symmetry and subjecting the phosphorylase dimer to positional refinement. The twofold axis relating the subunits in the dimer was strictly maintained thereafter. The refined model consists of 7012 non-hydrogen atoms with a root mean square (rms) deviation of 0.015 Å for bond lengths, and of 2.24° for bond angles. The R-factor is 0.191 and the corresponding free R-factor is 0.265.

While rebuilding the structure of yeast glycogen phosphorylase, we noticed several discrepancies between the electron density and the published sequence [1,10]. We have confirmed the following corrections to Figure 2 in reference [1] by sequence analysis; Val49→Ala, Glu114→Asp, Gly115→Leu, Leu254→Phe, Asn255→Ala, Pro267→Ala, Glu412→Gln, Asn447→Val, Val456→Ala, Ile472→Val, Tyr537→Val, Arg596→Lys and Leu805→Val.

Overall structure

The fold of the enzyme core is nearly identical to that of rabbit muscle phosphorylase. Differences between the two enzymes involve insertions and deletions on the surface of the protein and at the N and C termini (Fig. 1). Relative to the rabbit muscle enzyme, yeast phosphorylase is 39 residues longer at the N terminus and 11 residues shorter at the C terminus.

In addition to the differences in the N and C termini, yeast phosphorylase has 10 insertions and 2 deletions of varying lengths. Eight of these insertions are ordered, but are not associated with any known function of yeast phosphorylase. These are the single-residue insertions after positions 439, 474, 735, 767 and 794, and the multiple insertions after residues 508 (3 residues), 558 (6 residues) and 591 (9 residues). The remaining two insertions are either partially or wholly disordered in the electron-density map. The first, a 13-residue insertion after amino acid 113, is discussed below. An insertion of 6 residues after position 213 occurs within a loop which extends out from the body of the protein into the solvent and has no known function. The inserted residues, as well as those flanking position 213 (209–212 and 214), are disordered.

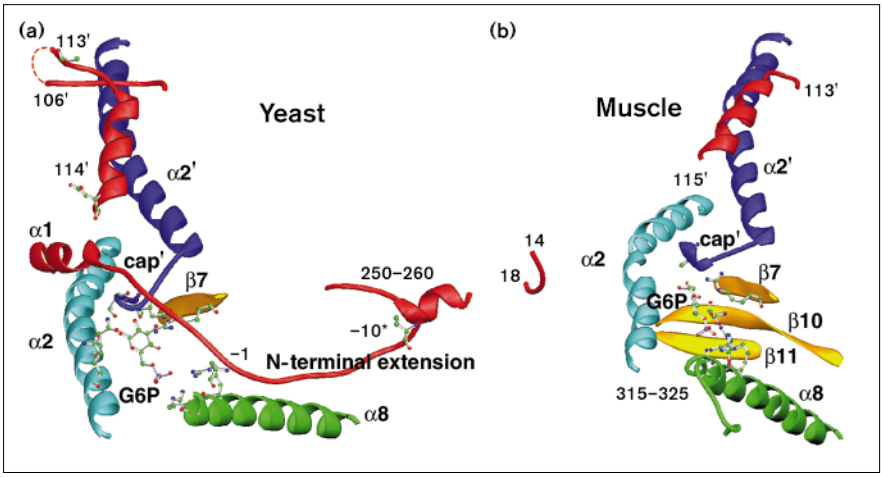
The two deletions (316–323 and 723–725) eliminate surface loops without altering the secondary structure of adjacent elements. Residues 316–326 have been implicated in nucleotide discrimination and the functional effect of this loss is discussed below. Residues 723–725 participate in tetramer contacts in muscle phosphorylase and these contacts are altered in the yeast tetramer [11,12].

Two parts of the chain, which are more or less disordered in all crystal structures of muscle phosphorylase, are ordered in the yeast phosphorylase structure. These are residues 1–4 and a loop (residues 250–260). Both are located near the Glc-6-P site and will be discussed in that context.

Dual versus single function of the nucleotide binding site in phosphorylases

Glc-6-P inhibits rabbit muscle phosphorylase by binding cooperatively to the site defined as the nucleotide-binding

Figure 2



Architecture of the Glc-6-P site. **(a)** A ribbon diagram of the secondary structural elements and ligand-binding residues in yeast phosphorylase. Glc-6-P and the side chains that interact with it are colored by atom type. A prime denotes a residue of the symmetry-related subunit in the functional dimer. The phosphorylation site, Thr(–10), is marked by an asterisk. The two symmetry-related α2 helices are shown in light blue and purple; the cap (residues 41′–47′) which precedes α2′ is also shown for one subunit (purple); strand β7 is in yellow and helix α8 is in green. Insertions in the polypeptide or regions of ordered chain not found in muscle phosphorylase are colored red and are discussed in the text. **(b)** Corresponding view of the nucleotide-binding site in muscle phosphorylase *b* [18]. Secondary structural elements are colored as in (a). The region of the polypeptide housing the 113 insertion in yeast is shown in red. Residues 1–14 are disordered in this structure.

site. This bi-functional site binds both activators (AMP) and inhibitors (ATP and Glc-6-P). In muscle tissue, phosphorylase responds sensitively to the energy level of the cell, determined by the intracellular ratio of AMP to ATP. AMP activates P*b* to 80% of its highest activity. AMP also increases the activity of P*a* by an additional 10–20%. Both ATP and Glc-6-P compete with AMP for binding at this site and stabilize an inactive conformation. The association constant (K_a) of muscle phosphorylase for AMP is 20 mM, the dissociation constant (K_i) for ATP is 5 mM and the K_i for Glc-6-P is in the low millimolar range [13–16]. From crystallographic studies, we now understand the differences in the binding modes of activators and inhibitors at this site [11,17–19].

The nucleotide-binding site is located at the subunit interface and consists of residues from both subunits. Studies of rabbit muscle phosphorylase have shown that activation of the enzyme by AMP or by metal ions when AMP binding residues are replaced by histidines [20] causes the two subunits to draw together in this region. In both cases this is accomplished by knitting together the two subunits through mutual interactions with ligand. The nucleotide or the metal binds between helix α2 (47–78) of one subunit and a loop, termed the cap (41–47), of the opposite subunit.

In contrast, the nucleotide-binding site in yeast phosphorylase binds a single inhibitor, Glc-6-P. AMP and ATP have no effect on catalytic activity. The apparent K_i of yeast P*b* for Glc-6-P is 0.3 mM; for P*a*, it is 5 mM and binding is cooperative (Hill coefficient 1.3) [21].

Novel architecture of the Glc-6-P site

Although the binding residues are largely conserved, the architecture of the Glc-6-P binding site in the yeast enzyme differs from that of muscle phosphorylase (Fig. 2). These differences result from the N-terminal extension (–39 to –1) unique to the yeast isozyme; a difference in ordered chain (1–18); an insertion after residue 113 (113C–113O); a second difference in ordered chain (250–260) and a deletion (316–323) which contains residues required for specific recognition of the adenine ring of AMP in the muscle enzyme (315–325).

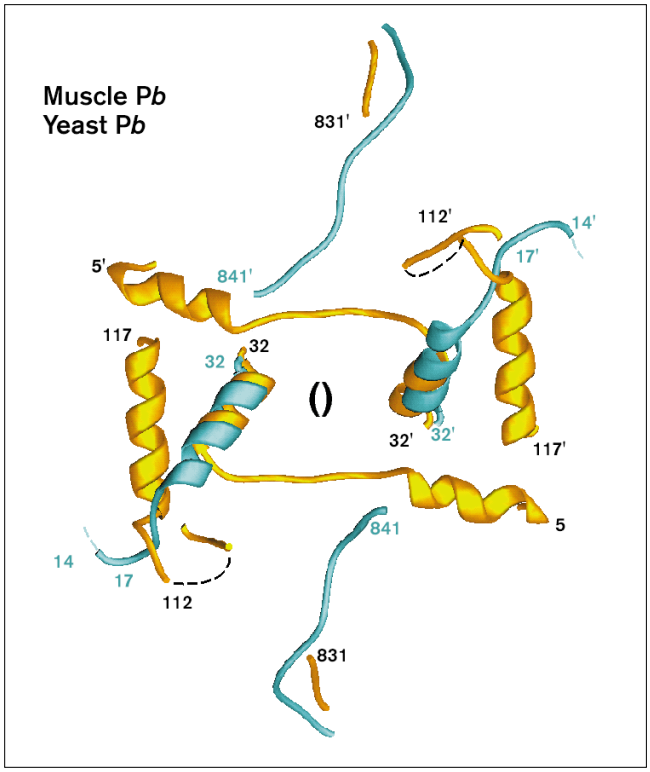
Like the ligand Glc-6-P, residues –11 to –1 of the yeast N-terminal extension, followed by residues 1–13, extend across the subunit interface. Residues –11 to –1 of the yeast extension stretch from the phosphate-binding subsite (309–310 on α8) of Glc-6-P to the ordered 250–260 loop in one subunit. Residues 1–13 pass over the cap′ (41′–47′) of the Glc-6-P binding site which emanates from the second, symmetry-related subunit.

Residues –11 to 14 are bound to the subunit interface by weak main-chain and van der Waals interactions. The binding site for the unphosphorylated peptide in rabbit muscle phosphorylase is similarly non-specific. Relatively weak binding is characteristic of segments whose structure and position is altered by phosphorylation and may be required for efficient interaction with the kinase. The temperature factors for this part of the N terminus are higher than for the rest of the protein.

One stabilizing interaction with the phosphorylation site, Thr(–10), involves residues 250–260 which are disordered in all structures of rabbit muscle phosphorylase. In the yeast structure, residues 250–260 form a strand which terminates in a two-turn helix. Thr(–10) interacts weakly with the first turn of the helix.

As has been described, residues 5–21 of yeast P*b* resemble muscle P*a* rather than P*b* [2]. The N terminus in muscle P*b* (residues 14–32) binds to its own subunit. When Ser14

Figure 3



Superposition of the N-terminal domains of one subunit from inactive yeast (yellow with black numbering) and inactive muscle (blue) phosphorylase structures at the subunit interface. The twofold axis relating the subunits in the dimer is indicated ().

is phosphorylated, residues 13–21 are displaced from the intrasubunit binding site, fold up into a 3_{10} helix and bind across the subunit interface [22]. Part of this intrasubunit binding site is formed by repositioning residues 113′–117′, which flank the 13 residue insertion in yeast phosphorylase [17,22]. Although the yeast phosphorylation site is at Thr(–10) and residue 14 is not subject to phosphorylation, this region of the polypeptide adopts the same conformation as in muscle P*a*.

The 13-residue insertion at position 113 is ordered but is preceded by a gap in the electron density. The secondary structures of muscle and yeast phosphorylase match up to residue 102. In muscle phosphorylase, residues 103 and 104 make a tight turn between two α helices. In the yeast enzyme, this turn is broken. There is electron density for residues 103–106 on one side of the turn and the first four residues of the insertion (113C–113F) on the other. Between 106 and 113C, the electron density for residues 107–112 is missing in the yeast enzyme. Following this gap, the secondary structures of the two isozymes again overlap as residues 105–114 of the muscle enzyme align with residues 113G–114 of the yeast enzyme.

A first step in the evolution of the allosteric switch involving phosphorylation of Ser14 in the mammalian enzymes may have been the elimination of this insertion from the muscle enzyme to create a binding site for the unphosphorylated peptide (Fig. 3). The structural transition that accompanies phosphorylation in muscle phosphorylase requires a distinct and separate binding site for the phosphorylated peptide on the surface of the protein, one that it cannot bind to unless it is covalently modified. Modeling shows that residues 113C and 113D of the yeast insertion overlap with residues 17–19 of muscle P*b* (Fig. 3). Thus, the alternative binding site for the unphosphorylated peptide observed in muscle is not available in the yeast enzyme.

Creation of such a binding site may also explain the longer C terminus of muscle phosphorylase. The C termini of the two isozymes differ in length by 11 residues (Fig. 3; residue 842 of muscle phosphorylase is disordered). When the muscle enzyme is phosphorylated, residues 839–841 become disordered or mobile to allow room for the intersubunit position of the phosphopeptide. Because the yeast C terminus is shorter, there is no steric conflict and therefore no need to displace it to create a binding site analogous to that of the phosphopeptide in muscle P*a*.

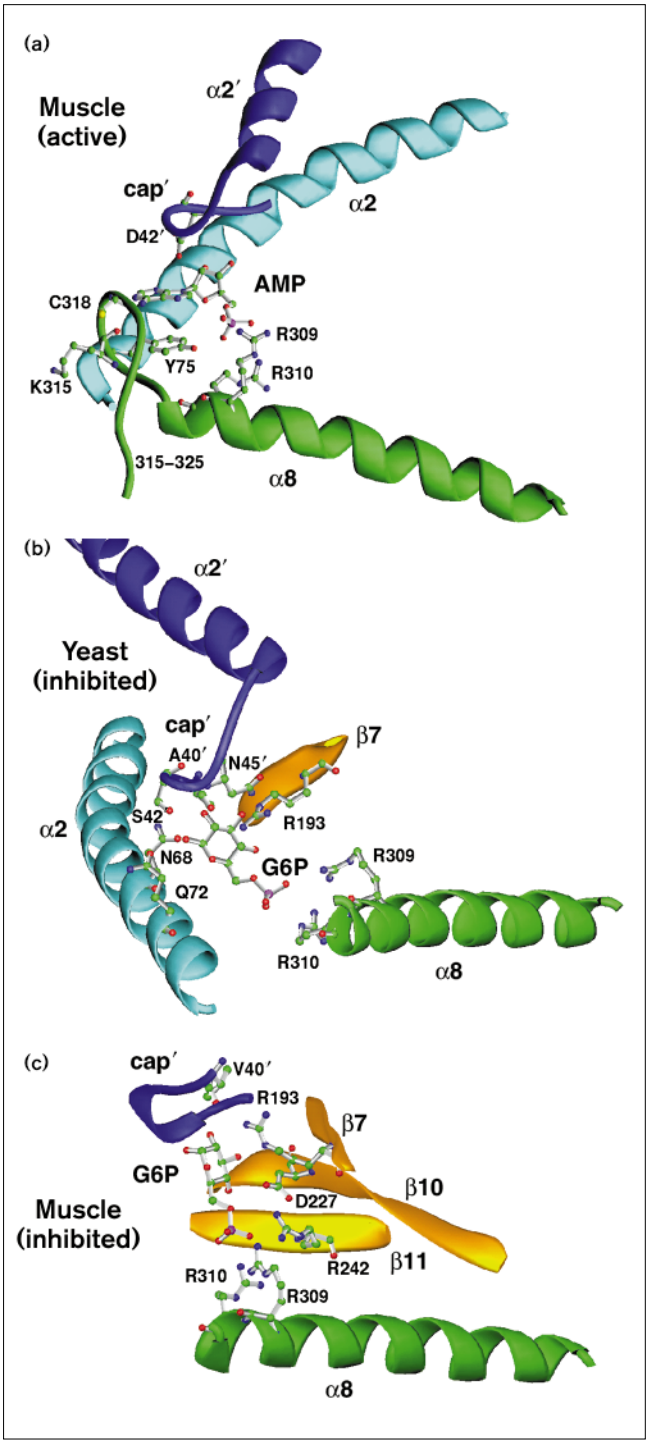
Why AMP does not activate yeast phosphorylase

AMP does not activate yeast phosphorylase because the enzyme lacks the adenine-specific interactions required to do so (Fig. 4a). The residues that interact with the adenine of AMP in muscle phosphorylase are located in two structural elements that belong to symmetry-related subunits. When bound, the adenine moiety draws together the cap from one subunit and helix α2 from the other. Correct positioning of the cap relative to helix α2 is required for activation. ATP fails to activate muscle phosphorylase because the diphosphate moiety displaces the nucleotide away from its binding site between the cap and Tyr75 in α2, with the result that the adenine ring is disordered. Glc-6-P inhibits muscle P*a* in a similar fashion by stabilizing the inactive conformation.

The nucleotide-binding site in muscle phosphorylase can be divided up into three subsites for the ribose, phosphate and adenine moieties. One of three residues that binds ribose in muscle phosphorylase is identical in the yeast amino acid sequence, suggesting that a sugar-binding site is partially retained. Four of the five residues that interact with the phosphate in muscle phosphorylase are identical in the yeast sequence.

Residues that bind the adenine ring of AMP or ATP are not conserved in yeast phosphorylase and no opportunistic candidates appear in the three-dimensional structure. In muscle phosphorylase, the adenine ring forms stacking interactions with Tyr75 and the adenine N3 forms a

Figure 4



How to build a switch. Ligands and binding residues for each complex are colored by atom type. Secondary structural elements involved in binding are shown; helix α2 from one subunit (light blue); cap′ and part of helix α2′ from symmetry-related subunit (purple); core strands β7, β10 and β11 (yellow); helix α8 (green). For clarity, views of each complex are slightly altered. **(a)** Rabbit muscle P*b* complexed with AMP [11]; **(b)** yeast P*b* complexed with Glc-6-P; **(c)** rabbit muscle P*b* complexed with Glc-6-P [18].

hydrogen bond to Asp42′ Oδ2. In yeast phosphorylase, Tyr75 is replaced by a threonine, which is unable to form significant stacking interactions, and Asp42 is substituted by a serine, which is one carbon shy of reaching the adenine ring.

The yeast enzyme also lacks a sequence which may be required for specific recognition of AMP over other nucleotides. The main-chain carbonyl oxygens of residues 315 and 318 form hydrogen bonds with the N6 of adenine in the complex of AMP with muscle P*b* [11]. The loop bearing these residues is disordered in all other structures of muscle phosphorylase and is deleted from the yeast sequence (316–323). The addition of AMP activation at the effector site in the muscle enzyme is the result of sequence changes that have created an adenine-specific binding site from the added 316–323 loop and contacts to helix α2.

Comparison of the binding sites for Glc-6-P in yeast and muscle phosphorylase

The binding mode of Glc-6-P to the yeast enzyme involves structural elements which are distinct from those used by the muscle enzyme (Fig. 4b,c). In both enzymes, Glc-6-P makes contacts to the cap structure and to helix α8. In muscle phosphorylase, additional contacts are formed to strands β10 and β11. Strand β11 in muscle phosphorylase links the Glc-6-P binding site to the active site through a long helix (residues 260–274, the tower helix) which terminates in the active site. In yeast phosphorylase, Glc-6-P lacks the contacts to strands β10 and β11 and forms additional contacts to the long helix, α2.

Glc-6-P is an effective inhibitor with similar K_i values in both rabbit muscle and yeast phosphorylase, but is bound in very different ways by the two enzymes (Table 1). Rabbit muscle phosphorylase binds α-D-Glc-6-P [18], whereas yeast P*b* appears to bind the β conformer, although we cannot be certain of this without higher resolution data. At equilibrium, a solution of either anomer consists of 36% α and 64% β, owing to the greater stability of the latter. The glucose moiety has the characteristic ^{1}C4 chair conformation. In the yeast structure, the O1 hydroxyl forms hydrogen bonds with the side chains of Gln72 and Asn68. Neither of these hydrogen bonds is formed in the complex between Glc-6-P and muscle P*b* [18]. In muscle phosphorylase, Asn68 is an isoleucine and therefore unable to form a hydrogen bond to the glucose O1 hydroxyl.

Glc-6-P also occupies a very different position in the yeast structure compared with the muscle enzyme. After superposition of the two binding sites (Cα carbons of residues 40′, 42′, 45′, 68, 72, 193, 309 and 310), the glucose moieties are rotated by 90° relative to one another. The displacement of Glc-6-P from its binding site in muscle phosphorylase is the result of destablizing contacts with the side

Table 1

Intermolecular contacts to Glc-6-P.

Glc-6-P	Yeast		Distance (Å)	Muscle		Distance (Å)
O1	Ser42′	Oγ	3.0			
	Asn68	Oδ1	3.3			
	Gln72	Oϵ1	3.0			
O2	Ser42′	N	3.0	Val40′	O	3.0
	Ser42′	Oγ	3.1	Arg193	NH1	3.0
	Asn68	Oδ1	3.0			
O3	Ala40′	O	2.9	Arg193	Nϵ	3.4
	Asn45′	Nδ2	2.9	Arg193	NH1	2.8
	Arg193	NH1	3.4	Asn227	Oδ1	3.0
O4	Arg193	NH1	3.4	Asn227	Oδ1	2.9
	Arg193	NH2	3.4			
OP1	Arg310	Nϵ	3.1	Arg309	NH1	3.0
OP2	Arg309	NH2	2.7	Arg309	NH2	3.4
	Arg310	NH2	2.9	Arg310	NH2	2.5
	Arg310	Nϵ	3.0	Arg242	NH1	2.5
OP3	Arg309	NH1	3.2	Arg310	NH2	2.6

Hydrogen bond contacts within 3.4 Å of Glc-6-P in yeast and rabbit muscle phosphorylase are compared.

chains of Trp67 and Asn68 and optimization of contacts with residues of the cap′ (which are more extensive in the yeast isozyme than in muscle phosphorylase), as well as with Gln72, Arg309 and Arg310.

In the complex of muscle P*b* with Glc-6-P, Trp67 stabilizes the inhibited conformation of the enzyme by forming a wall against which the O2′ and O3′ hydroxyls of the glucose moiety pack [18]. When the activator AMP is bound. Trp67 rotates to form van der Waal contacts with the AMP ribose. In its activated position, Trp67 blocks the Glc-6-P binding site. In the yeast structure, Trp67 is positioned as it would be in activated muscle phosphorylase, sterically occluding binding of Glc-6-P.

The contacts to the phosphate oxygens are similar in the yeast and muscle enzymes. In all AMP and Glc-6-P complexes of both active and inactive muscle phosphorylase, binding is dominated by the conserved Arg309 and Arg310 interactions with the phosphate oxygens. These are present in the yeast structure, but differ in detail.

The most notable difference in the phosphate oxygen contacts in the two isozymes involves Arg242, which links Glc-6-P to strand β11 in muscle phosphorylase. In the muscle isozyme, both 242 (β11) and Asn227 (β10) interact with Glc-6-P. In the yeast enzyme, these contacts to the core β strands are not made. Strand β11 is shifted back, such that 242 cannot extend to Glc-6-P and instead forms intrasubunit hydrogen bonds to residues 227 and 306. This strand movement may be partly caused by packing interactions with the extended yeast N terminus, because residues –1 to –10 pack against strands β8 and β10. Phosphorylation of Thr(–10) may alter these interactions, as this region of the polypeptide is likely to adopt a different conformation when phosphorylated.

Making the switch

The secondary structures used for binding ligands at the nucleotide site in yeast and muscle phosphorylases may be thought of as components of a switch (Table 2). In muscle phosphorylase, contacts to helix α2 and the 315–325 loop activate (AMP binding), whereas the alternative set of contacts to strands β7, β10 and β11 inhibit (Glc-6-P binding). The evolution of the effector site from an inhibitory locus (yeast) into a site that can both activate and inhibit (muscle) required turning the contacts to helix α2, strand β7, and the 315–325 loop into a switch and recruiting the inhibitory contacts to strands β10 and β11.

In our theoretical ancestral phosphorylase, Glc-6-P inhibition requires contacts to the cap, α2, β7 and α8 (yeast). Interactions with helix α2 were lost and additional contacts to the core strands β10 and β11 were added to stabilize the inactive conformation (Glc-6-P inhibition, muscle phosphorylase). Gain (α2) or loss (β10–β11) of these contacts would be expected to destabilize the inactive conformation and thereby promote the active state. The active conformation, lacking contacts to β10 and β11, was then further stabilized by AMP-specific interactions to helix α2 and the 315–325 loop (AMP activation, muscle phosphorylase).

Conservation of subunit interface contacts specific for Glc-6-P inhibition

A comparison of 14 phosphorylase sequences from human, rat, *Dictyostelium*, yeast, potato and *Escherichia coli* identified conserved residues of the subunit interface. The interacting pairs of residues can be grouped into three networks. The first comprises residues from the phosphorylation site and the AMP-binding site; the second, Glc-6-P binding residues and the third, active-site residues including the tower helix and active-site gate (tower/gate network) [23]. These networks are independent of each

Table 2

Structural elements involved in ligand binding in yeast and muscle phosphorylase.

Element	Residues	Yeast Glc-6-P	Muscle Glc-6-P	Muscle ATP	Muscle AMP	Switch elements
α1	23–40	–	–	–	–	–
Cap	41–47	+	+	–	+	–
α2	48–78	+	–	–	+	A
β7	190–194	+	+	–	–	I
β8	196–210	–	–	–	–	–
β10	221–230	–	+	–	–	I
β11	239–244	–	+	–	–	I
α8	290–313	+	+	+	+	–
Loop	315–325	–	–	–	+	A

The van der Waals contacts and hydrogen bonds within 3.4 Å of the ligand are listed for muscle phosphorylase and Glc-6-P [18]; muscle phosphorylase and ATP [33]; muscle phosphorylase and AMP [11]. Elements of the switch described in the text which result in activation (A) or inhibition (I) of enzyme activity are listed. A (+) indicates an interaction between the specified structural element and the ligand.

other. The phosphorylation site/AMP and tower/gate networks are connected by a set of conserved dimer interface pairs referred to as link 1; the Glc-6-P and tower/gate networks are joined by link 2 [23].

All the mammalian enzymes are subject to regulation by phosphorylation (of the same amino acid) and AMP. The non-mammalian phosphorylases — from *E. coli*, potato and *Dictyostelium* phosphorylase (II) — are unregulated. Yeast phosphorylase and phosphorylase (I) from *Dictyostelium* are regulated by phosphorylation (different amino acids) and either Glc-6-P or AMP, respectively. Overall, the interface pairs are highly conserved in all the mammalian enzymes. In the non-mammalian enzymes, the network associated with phosphorylation and regulation by AMP is poorly conserved; the most highly conserved network is the one associated with Glc-6-P binding residues. The conservation of binding residues for the phosphorylation site/AMP or Glc-6-P correlates with the conservation of dimer contact residues for the corresponding network, indicating a functional link between the two. Hudson *et al.* [23] proposed that Glc-6-P regulation evolved prior to activation by phosphorylation/AMP and that enzyme inhibition by Glc-6-P might use an intersubunit route distinct from the one used for activation.

The structure of yeast phosphorylase validates the conclusions from this sequence analysis. In yeast phosphorylase, contacts between the secondary structural elements within the Glc-6-P network, link 2 and part of the tower/gate network are preserved, although the specific contacts differ. It is striking that the contacts between secondary structural elements are conserved because the subunit interface is not. Only a small proportion (25%) of the interface residues (those within 4.0 Å of the second subunit within the dimer) in any of the muscle phosphorylase crystal structures is conserved in the yeast enzyme (6–8 out of 23–32, depending on the structure). Only individual residues are conserved; there are no conserved pairs of hydrogen bonds or contacts.

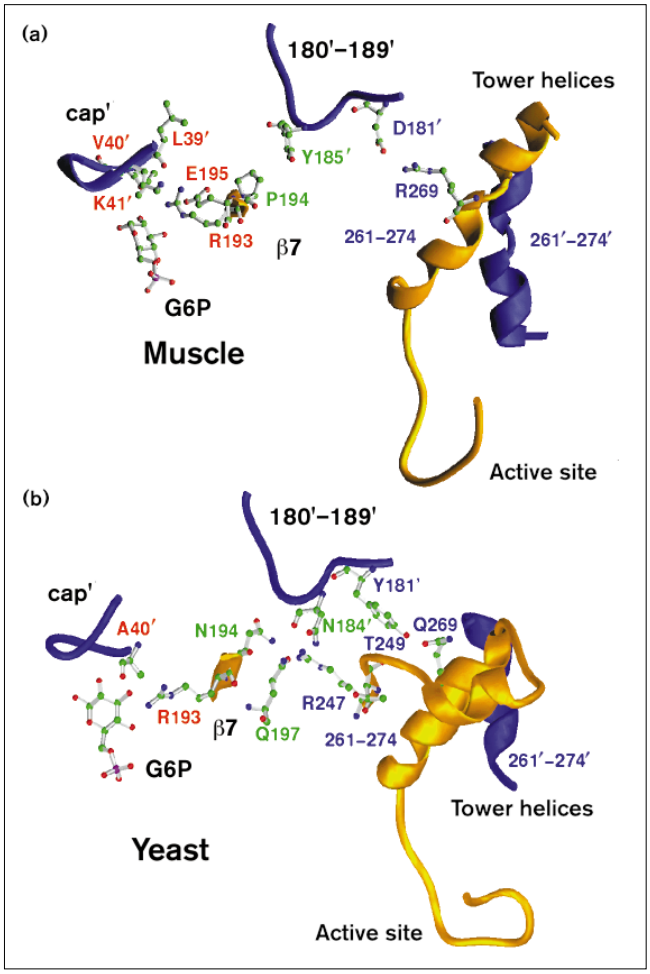
Conserved path between the effector and active sites

In muscle phosphorylase, the signal initiated by ligand binding to the nucleotide site is transmitted to the active site some 30 Å away. In yeast and muscle phosphorylases, residues which form the catalytic site are identical. Given that the binding modes of the ligands are not conserved, how is the function (inhibition) preserved? We suggest that the nucleotide-binding site and the active site are linked though circuitry intrinsic to the enzyme core. These transmission lines may be manipulated through interactions initiated by residues that bind Glc-6-P. It is this core circuitry that is exploited differentially by the regulatory structures in each isozyme.

In muscle phosphorylase, the interactions between Glc-6-P, residues of the cap and the strands β7 and β8 are critical components of the subunit interface and part of a pathway which links the allosteric site and the active site (Fig. 5a). The specific contacts between the two sites identified in the muscle isozyme are disrupted in yeast phosphorylase because of sequence changes. We have identified an alternative set of contacts between the same structural elements that are likely to serve the same purpose.

The contacts between the allosteric and active sites in muscle phosphorylase have been described in detail [3]. Briefly, the cap is linked to the turn joining strands β7 and β8 by hydrogen bonds between the main-chain carbonyls of residues 39′ and 40′ and the side chain of Arg193 and

Figure 5



Conserved path between effector and active sites. Structural elements of one subunit are colored purple; those belonging to the symmetry-related subunit are colored yellow. Ribbon diagram of structural elements with ligand and residues colored by atom type. From left to right, the figure traces the specific interactions linking the Glc-6-P site to the symmetry-related tower helices which terminate in the active-site gate in **(a)** muscle phosphorylase and **(b)** yeast phosphorylase. Residue labels are colored according to the network or link they belong to; Glc-6-P network (red), link 2 (green), tower/gate network (purple).

between Lys41′ and Glu195 (Glc-6-P network). The turn contacts a second loop (180′–189′, between β6 and β7) of the symmetry-related subunit by a hydrophobic contact between Pro194 and Tyr185′ (link 2). Neighboring residue Asp181′ is linked by a water molecule to Arg269 of the tower helix (residues 261–274, tower/gate network). The symmetry-related tower helices pack against each other and connect the two active sites in the physiological dimer. The tower helix terminates in a turn called the active-site gate (residues 278–289). In muscle P*b* or when the inhibitor glucose is bound to muscle P*a*, the gate prevents substrates from entering the catalytic site. When the enzyme is activated, the contacts between the two tower helices are altered dramatically and the gate becomes disordered.

In the yeast enzyme, all of these contacts are altered (Fig. 5b). Arg193 is conserved but forms a single hydrogen bond to residue 40′ of the cap (Glc-6-P network). Glu195 does not contact 41′, but instead makes an intrasubunit hydrogen bond with Arg193. The van der Waals interaction between Asn194 and Ser185′ is missing. As Ser185′ forms no other contacts, the circuit is disrupted.

However, an alternative route is present linking these same structural elements to the tower helix in the yeast isozyme (Fig. 5b). Asn194 forms a hydrogen bond with the main-chain carbonyl of Asn184′ as does the side chain of Arg247. The side chain of Asn184′ forms a hydrogen bond to Gln197 in strand β8 (link 2). Tyr181′ is linked to both ends of the tower helix through hydrogen bonds to Thr249 and Gln269 (tower/gate network). This proposed route of transmission remains to be confirmed by the determination and analysis of structures of the yeast enzyme in different conformational states.

The tetramer interface

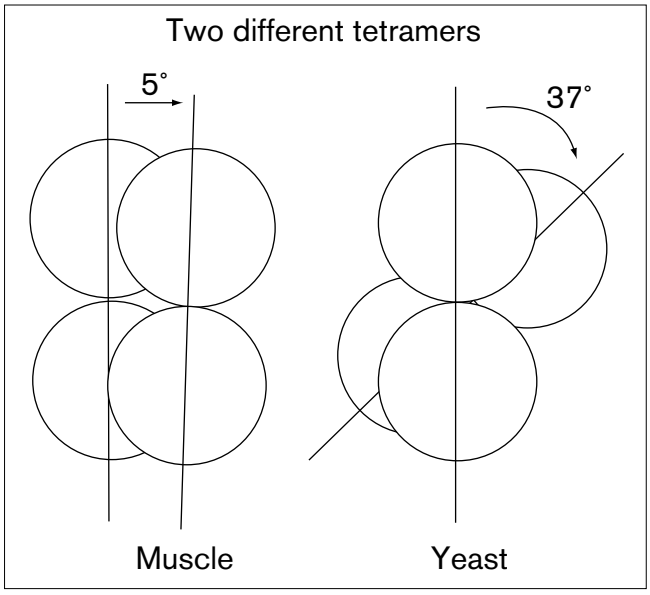
The physiological tetramer can be located in the yeast phosphorylase crystal by applying a crystallographic symmetry operator to the dimer in the asymmetric unit. Under different conditions, both yeast and muscle phosphorylases form tetramers in solution, but the tetramers differ structurally and serve different functional roles. Rabbit muscle P*a* forms tetramers in solution, whereas yeast P*b* is tetrameric. The tetramer interface is more extensive in yeast than in muscle phosphorylase and may serve a regulatory role. The alternative yeast tetramer is a direct result of alterations in the subunit interface.

The structural difference can be described in terms of the way the two dimers pack together (Fig. 6). An axis perpendicular to the non-crystallographic twofold operator is drawn for each dimer. The axis connects the centers of mass of the pyridoxal phosphate cofactor in each subunit. The difference between the two dimer axes within the tetramer can be defined by three angles, which can be summed to give an overall angular difference. The two dimers in the yeast phosphorylase tetramer differ by 37° whereas those in the rabbit muscle tetramers differ by 5° [11,17,24]. This difference in packing angle means that different sets of residues are recruited into the tetramer interface in the two isozymes.

The yeast and muscle phosphorylase tetramers assume different roles in the two cell types. In muscle, in the absence of glycogen, the activated state forms a crystallizable tetramer in solution, in which the glycogen-binding face of the enzyme is buried within the tetramer interface [24–26]. The glycogen-binding subdomain (residues 377–439) is distinct from the active site and has been identified crystallographically using oligosaccharides to mimic glycogen binding [27]. Inside the muscle cell, the enzyme binds to glycogen as a dimer.

In yeast, the activity of phosphorylase is regulated in part by controlling the equilibrium between dimers and tetramers. Sedimentation experiments have shown that the unphosphorylated, inactive yeast phosphorylase forms a tetramer in solution [21]. The tetramer interface masks the presumed binding site for the glycogen particle as identified in rabbit muscle phosphorylase. When phosphorylated, the enzyme becomes activated and dimeric. Phosphorylation disrupts the tetramer to permit the enzyme to bind substrate. *In vivo*, the tetramers may serve an additional protective function for the N terminus. We superimposed the glycogen-binding domains of the yeast

Figure 6



Two different tetramers. Yeast and muscle phosphorylase dimers pack together at different angles to form tetramers. The cartoon illustrates the overall angular difference between axes linking the active sites in the dimers of muscle and yeast phosphorylases.

enzyme and the structure of phosphorylated muscle phosphorylase complexed with maltopentaose [27]. The superposition shows that the extended yeast N terminus overlaps partially with oligosaccharides bound to the muscle enzyme, suggesting that the yeast N terminus may compete with glycogen for its binding site in the tetramer.

Biological implications

Glycogen phosphorylases catalyze the phosphorylytic degradation of glycogen, a storage polymer, into glucose 1-phosphate which feeds into glycolysis to meet the energy demands of the cell. Phosphorylases are regulated in different ways to meet the specific needs of the cell type in which they are expressed. Sequence analysis of this family of enzymes suggests that regulatory features were added as structural modules onto a conserved catalytic core. Regulation of enzyme activity can be thought of as manipulating circuitry intrinsic to the atomic structure of the protein.

To understand the design criteria for regulatory features, we have solved the structure of inactive, glucose 6-phosphate (Glc-6-P) inhibited, yeast phosphorylase *b* and compared it with the known crystal structures of muscle phosphorylase. The catalytic cores of the two enzymes are conserved. In this paper, we focus on the allosteric site which binds only the inhibitor Glc-6-P in the yeast enzyme but is regulated by both inhibitors (ATP and Glc-6-P) and activators (AMP) in the muscle enzyme.

The addition of AMP activation at the effector site in the muscle enzyme is the result of sequence changes which create an adenine-specific binding site. In muscle phosphorylase, contacts between the effector and helix α2 and loop 315–325 activate the enzyme (AMP binding), whereas the alternative set of contacts to strands β7, β10 and β11 inhibit it (Glc-6-P binding). The evolution of the effector site from an inhibitory locus (yeast) into a site that can both activate and inhibit (muscle) required turning the contacts to helix α2, strands β7, and the 315–325 loop into a switch and adding the inhibitory contacts to strands β10 and β11. It appears that the Glc-6-P site of yeast phosphorylase evolved into an allosteric switch in muscle phosphorylase by altering the binding residues such that bound ligands contact different structural elements at the subunit interface. These elements are linked to their respective catalytic sites by a common route.

Materials and methods

Data measurement and reduction

Crystals of the dephosphorylated form of yeast phosphorylase were grown as described [1]. Diffraction data were measured from five crystals using a Mark II Xuong/Hamlin multiwire detector with graphite-monochromated CuKα X-rays. Data from each crystal were merged and scaled separately [28]. Orientations or crystals with high merging R-factors were omitted from the final data set. The data sets of three crystals were then merged and scaled together using the CCP4 programs LSCALE, ROTAVATA and AGROVATA [29]. The final data set consisted of 557 407 reflections which reduced to 67 571 unique reflections from 23.0–2.47 Å resolution. The data set was 80.1% complete for the 2.64–2.7 Å resolution shell and 44% complete in the 2.59–2.64 Å resolution shell. The overall I/σ was 7.4, the merging R-factor was 8.8% on intensities.

Molecular replacement model

The structure of yeast glycogen phosphorylase was solved by the method of molecular replacement [30] using a model derived from the atomic coordinates of the dephosphorylated form of rabbit muscle phosphorylase (kindly provided by LN Johnson [3]). Solvent molecules, ligands and the cofactor, pyridoxal phosphate, were removed from the model. The side chains of non-glycine residues which were not identical between the yeast and rabbit muscle sequences were replaced with alanine. Regions of the protein which were poorly ordered in the rabbit muscle phosphorylase crystal structure (residues 250–260, 315–325 and 839–841; [3]) were deleted from the model. Insertions and deletions in the yeast sequence relative to that of rabbit muscle phosphorylase (including several residues bracketing such sequences) were omitted from the model (residues 111–116, 212–213, 439–440, 475–476, 507–509, 555–560, 591–599, 722–726, 736–737, 766–767, 795–796 and 831–841). The poorly conserved N-terminal 60 residues were also omitted from the search molecule. The search model consisted of 9398 atoms or 65% of the total atoms in the yeast phosphorylase homodimer. The program X-PLOR [4] was used for both molecular replacement and refinement of the structure with rebuilding in FRODO [8] and O [9]. All figures were made with the program RIBBONS [31].

Solvent accessibility was calculated using the program ACCESS (B Lee, FM Richards, TJ Richmond and M Handschumacher, personal communication). A probe radius of 1.4 Å was used for water; van der Waals radii of the protein atoms were as given in [32].

The atomic coordinates for yeast phosphorylase *b* complexed with Glc-6-P are being deposited in the Brookhaven Protein Data Bank.

Acknowledgements

This work was supported by grants from the San Diego Super Computer Center (CSF232), the Pittsburgh Supercomputing Center (DMB910006P) and the NIH (DK32822 and DK20681).

References

1. Rath, V.L., Hwang, P.K. & Fletterick, R.J. (1992). Purification and crystallization of glycogen phosphorylase from *Saccharomyces cerevisiae*. *J. Mol. Biol.* **225**, 1027–1034.
2. Rath, V.L. & Fletterick, R.J. (1994). Parallel evolution in two homologues of phosphorylase. *Nature Struct. Biol.* **1**, 681–690.
3. Acharya, K.R., Stuart, D.I., Varvill, K.M. & Johnson, L.N. (1991). *Glycogen Phosphorylase* b*: Description of the Protein Structure*. World Scientific Publishing Co., Inc., Teaneck, NJ.
4. Brünger, A.T. (1992). *X-PLOR Manual, Version 3.1*. Yale University Press, New Haven and London.
5. Huber, R. (1985). Molecular replacement. In *Proceedings of the Daresbury Study Weekend*. SERC Daresbury Laboratory, Warrington, UK.
6. Brünger, A.T. (1992). Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature* **355**, 472–475.
7. Brünger, A.T., Krukowski, A. & Erickson, J.W. (1990). Slow cooling protocols for crystallographic refinement by simulated annealing. *Acta Cryst. A* **46**, 583–593.
8. Jones, T.A. (1978). A graphics model building and refinement system for macromolecules. *J. Appl. Cryst.* **11**, 268–272.
9. Jones, T.A., Zou, J.-Y., Cowan, S.W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Cryst. A* **47**, 110–119.
10. Hwang, P.K. & Fletterick, R.J. (1986). Convergent and divergent evolution of regulatory sites in eukaryotic phosphorylases. *Nature* **324**, 80–84.

11. Sprang, S.R., Withers, S.G., Goldsmith, E.J., Fletterick, R.J. & Madsen, N.B. (1991). Structural basis for the activation of glycogen phosphorylase *b* by adenosine monophosphate. *Science* **254**, 1367–1371.
12. Barford, D. & Johnson, L.N. (1992). The molecular mechanism for the tetrameric association of glycogen phosphorylase promoted by protein phosphorylation. *Protein Sci.* **1**, 472–493.
13. Madsen, N.B. (1964). Allosteric properties of phosphorylase *b*. *Biochem. Biophys. Res. Commun.* **15**, 390–395.
14. Morgan, H.E. & Parmeggiani, A. (1964). Regulation of glycogenolysis in muscle. *J. Biol. Chem.* **239**, 2440–2445.
15. Wang, J.H., Tu, J.I. & Lo, F.M. (1970). Effect of glucose-6-phosphate on the nucleotide site of phosphorylase *b*. *J. Biol. Chem.* **245**, 3115–3121.
16. Leonidas, D.D., Oikonomakos, N.G., Papageorgiou, A.C., Xenakis, A., Cazianis, C.T. & Bem, F. (1990). The ammonium sulfate activation of phosphorylase *b*. *FEBS Lett.* **261**, 23–27.
17. Barford, D., Hu, S.-H. & Johnson, L.N. (1991). Structural mechanism for glycogen phosphorylase control by phosphorylation and AMP. *J. Mol. Biol.* **218**, 233–260.
18. Johnson, L.N., Snape, P., Martin, J.L., Acharya, K.R., Barford, D. & Oikonomakos, N.G. (1993). Crystallographic binding studies on the allosteric inhibitor glucose-6-phosphate to T state glycogen phosphorylase *b*. *J. Mol. Biol.* **232**, 253–267.
19. Sprang, S., Goldsmith, E. & Fletterick, R. (1987). Structure of the nucleotide activation switch in glycogen phosphorylase *a*. *Science* **237**, 1012–1019.
20. Browner, M.F., Hackos, D. & Fletterick, R.J. (1994). Identification of the molecular trigger for allosteric activation in glycogen phosphorylase. *Nature Struct. Biol.* **1**, 327–333.
21. Lin, K., Hwang, P.K. & Fletterick, R.J. (1995). Mechanism of regulation in yeast glycogen phosphorylase. *J. Biol. Chem.* **270**, 26833–26839.
22. Sprang, S.R., *et al.*, & Johnson, L.N. (1988). Structural changes in glycogen phosphorylase induced by phosphorylation. *Nature* **336**, 215–221.
23. Hudson, J.W., Golding, G.B. & Crerar, M.M. (1993). Evolution of allosteric control in glycogen phosphorylase. *J. Mol. Biol.* **234**, 700–721.
24. Barford, D. & Johnson, L.N. (1989). The allosteric transition of glycogen phosphorylase. *Nature* **340**, 609–616.
25. Wang, J.H., Shonka, M.L. & Graves, D.J. (1965). Influence of carbohydrates on phosphorylase structure and activity. I. Activation by preincubation with glycogen. *Biochemistry* **11**, 2296–2301.
26. Wang, J.H., Shonka, M.L. & Graves, D.J. (1965). The effect of glucose on the sedimentation and catalytic activity of glycogen phosphorylase. *Biochem. Biophys. Res. Commun.* **18**, 131–135.
27. Goldsmith, E.J., Sprang, S.R., Hamlin, R., Xuong, N.-H. & Fletterick, R.J. (1989). Domain separation in the activation of glycogen phosphorylase *a*. *Science* **245**, 528–532.
28. Howard, A.J., Nielsen, C. & Xuong, N.-H. (1985). Software for a diffractometer with multiwire area detector. *Methods Enzymol.* **114**, 452–472.
29. Collaborative Computational Project, Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Cryst. D* **50**, 760–763.
30. Rossmann, M.G., Ed. (1972). *The Molecular Replacement Method.* Gordon and Breach, New York.
31. Carson, M. (1991). Ribbons 2.0. *J. Appl. Cryst.* **24**, 958–961.
32. Lee, B. & Richards, F.M. (1971). The interpretation of protein structures: estimation of static inaccessibility. *J. Mol. Biol.* **55**, 379–400.
33. Sprang, S.R., Fletterick, R.J., Stern, M., Yang, D., Madsen, N.B. & Sturtevant, J. (1982). Analysis of an allosteric binding site: the nucleoside inhibitor site of phosphorylase *a*. *Biochemistry* **21**, 2036–2048.