A cofactor-dependent phosphoglycerate mutase homolog from *Bacillus stearothermophilus* is actually a broad specificity phosphatase

DANIEL J. RIGDEN,[1] IRINA BAGYAN,[2] EJVIS LAMANI,[3] PETER SETLOW,[2] AND MARK J. JEDRZEJAS[3]

[1]National Centre of Genetic Resources and Biotechnology, Cenargen/Embrapa, S.A.I.N. Parque Rural, Final W5, Asa Norte, 70770–900, Brasília, Brazil
[2]Department of Biochemistry, University of Connecticut Health Center, Farmington, Connecticut 06032, USA
[3]Department of Microbiology, University of Alabama at Birmingham, Birmingham, Alabama 35294, USA

(RECEIVED April 25, 2001; FINAL REVISION June 13, 2001; ACCEPTED June 13, 2001)

Abstract

The distribution of phosphoglycerate mutase (PGM) activity in bacteria is complex, with some organisms possessing both a cofactor-dependent and a cofactor-independent PGM and others having only one of these enzymes. Although *Bacillus* species contain only a cofactor-independent PGM, genes homologous to those encoding cofactor-dependent PGMs have been detected in this group of bacteria, but in at least one case the encoded protein lacks significant PGM activity. Here we apply sequence analysis, molecular modeling, and enzymatic assays to the cofactor-dependent PGM homologs from *B. stearothermophilus* and *B. subtilis*, and show that these enzymes are phosphatases with broad substrate specificity. Homologs from other gram-positive bacteria are also likely to possess phosphatase activity. These studies clearly show that the exploration of genomic sequences through three-dimensional modeling is capable of producing useful predictions regarding function. However, significant methodological improvements will be needed before such analysis can be carried out automatically.

Keywords: *Bacillus*; functional genomics; molecular modeling; phosphatase; phosphoglycerate mutase homolog; structure–-function relationship; structure prediction

Phosphoglycerate mutase (PGM) primarily interconverts 3-phosphoglyceric acid (3PGA) and 2-phosphoglyceric acid (2PGA) in both glycolysis and gluconeogenesis. Two different types of PGMs have been identified, one which is dependent on the cofactor 2,3-bisphosphoglyceric acid (BPG) for activity (dPGMs) and the other whose activity is independent of BPG (iPGMs) (Fothergill-Gilmore and Watson 1989). The iPGMs are monomers of ~60 kD and are found in plants, whereas the dPGMs are oligomers of 25-kD subunits and are found in animals and fungi (Fothergill-Gilmore and Watson 1989; Jedrzejas 2001). There is no amino acid sequence similarity between these two types of PGMs and their structures are also quite different as are their catalytic mechanisms (Jedrzejas 2000).

Surprisingly, both types of PGMs are found in some bacteria, such as *Escherichia coli*, where both enzymes are expressed and are active PGMs, although the dPGM appears to be the predominant form active in vivo (Fraser et al. 1999). The situation is different in gram-positive bacteria related to *Bacillus* species, in which the iPGM is the predominant PGM, and mutation of the gene coding for this enzyme has an extremely deleterious effect on growth on glucose (Chander et al. 1998; Leyva-Vazquez and Setlow

Reprint requests to: Dr. Mark J. Jedrzejas, Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr Way, Oakland, CA 94609-1673, USA; e-mail: mjedrzejas@chori.org; fax: (510) 450-7910.

Abbreviations: 2PGA, 2-phosphoglyceric acid; 3PGA, 3-phosphoglyceric acid; BPG, 2,3-bisphosphoglyceric acid; PGM, phosphoglycerate mutase; dPGM, BPG-dependent PGM; iPGM, BPG-independent PGM; DTT, dithiothreitol; F26BPase, fructose-2,6-bisphosphatase; LB, Luria-Bertani; R5PPase, α-ribazole-5-phosphate phosphatase.

Article and publication are at http://www.proteinscience.org/cgi/doi/10.1101/ps.15701.

1994). Nevertheless, *Bacillus* species and their close relatives do contain a gene that encodes a protein with significant homology to dPGMs. In *B. subtilis* this gene, termed *yhfR*, is expressed, albeit at a rather low level (Pearson et al. 2000). However, deletion of *yhfR* had no phenotypic effect and assays of overexpressed and purified *B. subtilis* YhfR showed that this protein had no PGM activity (Pearson et al. 2000). Although this latter result was somewhat surprising, *B. subtilis* YhfR and its homologs in related bacteria such as those of various *Clostridium* species exhibited much lower amino acid sequence identity to dPGMs than might have been expected (Pearson et al. 2000). These results suggested that YhfR might not be a PGM, but might instead catalyze some other enzymatic reaction.

Structural determination has confirmed an evolutionary relationship between dPGMs and fructose-2,6-bisphosphatases (F26BPases), which was previously inferred from sequence similarities (Pilkis et al. 1987; Jedrzejas 2000). One hundred ninety of the ~235 located residues of *Saccharomyces cerevisiae* dPGM can be superimposed on the F26BPase structures with a root-mean-square deviation of matched Cα–-Cα distances of 2.5Å and a pairwise sequence identity of 20%. In the SCOP database (Murzin et al. 1995), dPGMs and F26BPase cluster together at the "family" level. They share a common catalytic core centered on a histidine residue (residue 8 in the *S. cerevisiae* dPGM numbering), which is transiently phosphorylated during the reaction (Pilkis et al. 1987; Han and Rose 1979). Nearby conserved residues are Glu86 and His181, which may participate as proton acceptor and donor, respectively, during catalysis (Rigden et al. 1999; Bond et al. 2001), along with Arg7, Asn14, and Arg59 that serve to stabilize the phosphohistidine intermediate (Bond et al. 2001).

A previously unsuspected relationship of dPGMs with the larger acid phosphatases and phytases has also been revealed by analysis of structures determined for the latter enzymes (Schneider et al. 1993). For example, 146 Cα atoms can be matched between *S. cerevisiae* dPGM and rat prostatic acid phosphatase (342 residues) with a root-mean-square difference of 4Å. However, the sequence similarity between dPGMs and acid phosphatases is insignificant at ~13%. In the SCOP classification dPGMs and acid phosphatases are grouped together at the "superfamily" level. The distant relationship between the two types of enzymes was indicated by the mechanistic importance of residues conserved in one family that are not present in the other (e.g., Ostanin et al. 1994). However, a phosphohistidine intermediate is also involved in the acid phosphatase mechanism (van Etten 1982). Another notable difference between dPGMs and F26BPases on the one hand and acid phosphatases on the other, is the well-defined substrate specificity seen for the first two types of enzymes (Fothergill-Gilmore and Watson 1989), which contrasts with the very broad specificity of the acid phosphatases (see, for example, van Etten and Waymack 1991; Apostol et al. 1985).

Given the large amount of information available on the structure and function of proteins clearly related to the YhfR of *Bacillus* species, we decided to carry out sequence analysis and modeling studies with YhfR to gain insight into the likely structure of this protein, in particular its active site, as this might give some indication of the reaction catalyzed by YhfR.

Results and Discussion

Sequence relationships of B. stearothermophilus *YhfR*

Database searches with YhfR showed this protein to be a member of the dPGM/F26BPase family, but failed to group it consistently with either dPGMs, F26Bpases, or α-ribazole-5-phosphate phosphatases (R5PPases). For example, FASTA3 gave an order of degree of sequence similarity for YhfR of dPGMs > R5PPases > F26Bpases, whereas BLAST2 and PSI-BLAST gave dPGMs > F26BPases > R5PPases. In contrast, Hidden Markov Modeling methods suggested that the closest kinship of YhfR was with F26BPases. The maximum pairwise sequence identity of the *B. stearothermophilus* YhfR sequence with a dPGM was 34%, but with F26BPases and R5PPases the maximum figures were 35% and 29%, respectively. These figures are lower than those typically observed between proteins having a similar catalytic activity. Rigorous phylogenetic analysis also showed YhfR to sit well outside the three main groups of dPGMs, F26Bpases, and R5PPases (Fig. 2, below). Bootstrapping values attached to nodes involving *B. stearothermophilus* YhfR and other bacterial YhfR homologs (as discussed in more detail later) are low indicating a lack of clear evolutionary positioning with respect to the dPGM, F26Bpase, and R5PPase groups.

Although these analyses were inconclusive, it seemed possible that detailed analysis of the most conserved regions of the individual families might enable a more confident assignment of function. Consequently, the MEME (Grundy et al. 1996) motifs were calculated for the dPGM and F26BPase families individually, although this analysis could not be carried out for R5Ppases, as only two sequences are present in the ENZYME database. These more highly conserved regions are shown, along with their rankings, in Figure 1. Detailed comparison of these motifs in representative members of both families of enzymes and YhfR again yields an interestingly mixed picture. For example, the second highest ranked MEME motif covers dPGMs from Arg7 to Trp22 (5PGM numbering scheme). In this region the YhfR sequence is much closer to the dPGM sequence consensus than to the F26BPase consensus, sharing with the former both Trp13 and Trp22. These are aligned, respectively, with an unconserved position and a

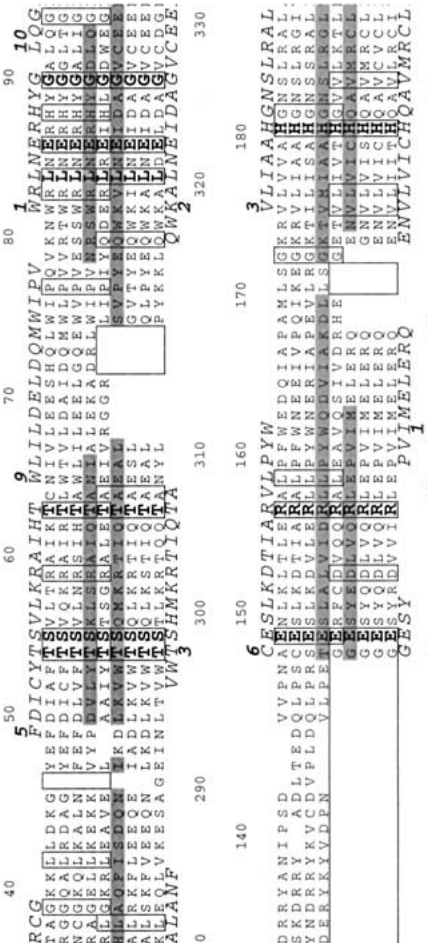

Fig. 1. Sequence alignment of the *B. stearothermophilus* sequence with representative members of the dPGM and F26BPase families (above and below, respectively). Residues conserved between all sequences are in bold and those shared by the *B. stearothermophilus* NGB101 YhfR sequence with either dPGMs or F26BPases are boxed. The templates used in model construction are immediately above and below the YhfR sequence and are shaded in regions used for model building. Numbering for the templates is shown above and below the alignment for 5PGM and 1BIF, respectively. Motifs associated with the dPGM and F26BPase families are shown above and below the alignment, respectively, and are labeled in approximate order of significance. This figure was made with ALSCRIPT (Barton 1993) as was Figure 5.

deletion in the F26BPase family. In contrast, residues 125–149 in dPGMs comprise an insertion not present in F26BPases. Here the YhfR sequence resembles more that of the F26BPases in lacking the insertion and largely matching their sequence consensus. In still other regions, the YhfR sequence differs from both the dPGM and F26BPase sequences. For example, residues 169–173 of the dPGMs comprise a five-residue insertion relative to F26BPases, whereas YhfR has a three-residue loop. A neighbor-joining tree representation of the *B. stearothermophilus* YhfR sequence with dPGAM/F26BPase family members present in the ENZYME database (Fig. 2) confirms the distant relationship of YhfR to known sequences. Sequences not grouping with the three main branches—dPGMs, R5Ppases, and F26Bpases—are few, comprising PMG2 and PMG3 of yeast that lack PGM activity and are probably pseudogenes

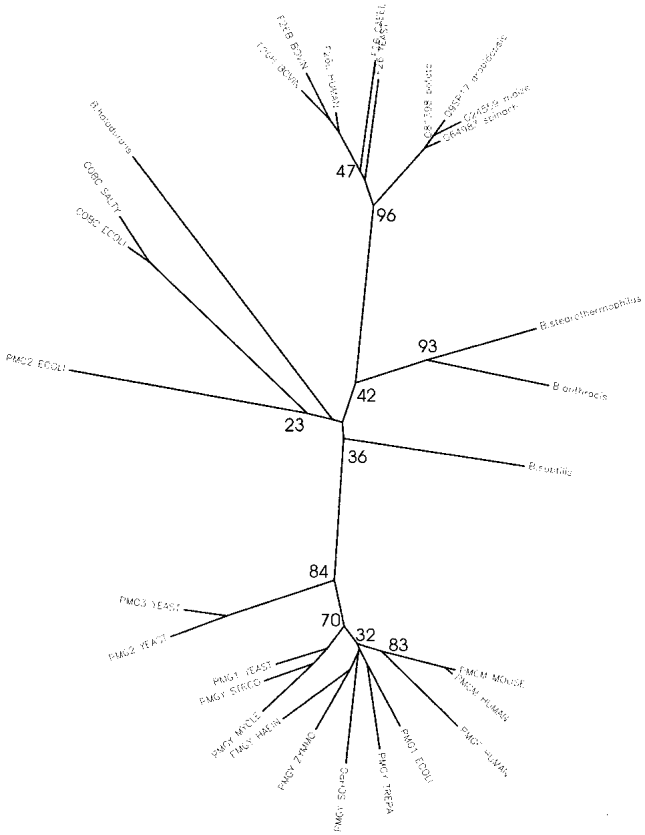

Fig. 2. Neighbor-joining tree produced using programs of the PHYLIP package (Felsenstein 1989). Sequences are dPGM homologs from the ENZYME database (Bairoch 2000), with some near identical members not shown for clarity, supplemented with YhfR sequences from *B. stearothermophilus* NGB101, *B. halodurans* and *B. subtilis* (Takami et al. 2000; Kunst et al. 1997), and four plant F26BPases. SWISSPROT identifiers are shown for ENZYME database members, SWISSPROT code plus species name given for the plant F26BPases and the others referred to by their species name. Bootstrapping values derived from a consensus tree (100 replicates) have been added to some major nodes. The branching position of the *B. halodurans* sequence is different in the consensus tree.

(Heinisch et al. 1998) and PMG2 of *E. coli*, an uncharacterized genomic sequence (Burland et al. 1995).

Individual positions of known importance in the dPGM/F26BPase family were then analyzed in the YhfR sequence. Information about which residues are most important for catalysis and substrate binding has come from mutagenesis studies (e.g., Lin et al. 1992; White et al. 1993; Mizuguchi et al. 1999), crystal structures with bound substrates (Yuen et al. 1999) or ions (Rigden et al. 1999), and from the high-resolution crystal structure of phosphorylated *E. coli* dPGM (Bond et al. 2001). These studies have revealed a catalytic machinery conserved between dPGMs and F26Bpases, as well as substrate-binding residues specific to each family. Using *S. cerevisiae* dPGM numbering, the catalytic machinery comprises His8, His181, Glu86, Arg59, Asn14, and Arg7. The YhfR sequence shares all of these key residues. Among the family-specific positions suggested to be important in substrate binding are Tyr336, Arg350, Lys354, Tyr365, and Arg395 in the F26BPases (Haseman et al. 1996; 1BIF numbering) and Arg87, Tyr89, Arg113, Arg114, and Asn183 in the dPGMs (Rigden et al. 1999; 5PGM numbering). These residues are well positioned to hydrogen bond to the 6-phosphate group of F26BP in the F26BPases or the 3-phosphate group of BPG or 1,3-bisphosphoglyceric acid in the dPGMs. The YhfR sequence is strikingly different from both F26BPases and dPGMs at these positions, as in YhfR just two of the five positions implicated in F26BPase catalysis retain hydrogen bonding capability, whereas the comparison with dPGMs reveals only a single position in the YhfR sequence with hydrogen bonding capability.

The conclusion from these various sequence analyses is that the YhfR sequence is only distantly related to the dPGMs, F26Bpases, and RBPpases, as YhfR seems to share some structural characteristics with both dPGMs and F26Bpases, while also having a number of unique features. Functionally, the YhfR sequence has an apparently intact catalytic machinery but a substrate-binding site clearly different from that of both dPGMs and F26BPases.

Modeling of the B. stearothermophilus *YhfR sequence*

As noted, the analyses of the YhfR sequence were somewhat inconclusive as to a possible catalytic function for this protein. Because protein function ultimately depends on structure, modeling of the YhfR structure was carried out to shed further light on this protein's possible catalytic activity. Threading studies were used to determine the best templates to use in model construction. Once again, neither dPGM nor F26BPase was highlighted as a single favorable template. GenTHREADER rated dPGMs and F26BPases as equally suitable templates, the Bioinbgu methods favored F26BPase, whereas 3D-PSSM gave dPGMs better scores. Both a dPGM (PDB code 5PGM) and a F26BPase (PDB

code 1BIF) were used in model construction. In general the use of multiple templates improves the accuracy of models (Bates and Sternberg 1999).

Although structurally similar over most of their length, these templates also have regions of high structural divergence. Because simultaneous use of divergent templates degrades model quality (Sali et al. 1997), a single template was used in regions in which corresponding Cα positions in superimposed templates were >2Å distant. The choice of template in these regions was based on local sequence comparison. Where this proved inconclusive, different sets of models were produced and analyzed using either template.

PROSA II profile results were used to highlight regions of unusual packing or solvent exposure characteristics. In this way decisions could be made about the choice of template and positions of insertions and deletions. The template regions used for construction of the final model are shown shaded in Figure 1. From the initial alignment of the YhfR sequence with templates, three changes involving switches from dPGM to F26BPase template were found to be favorable. These were coordinated, as steric considerations prohibited the use of a combination of different templates in these three regions. For construction of the final model, both templates were used for 124 residues, dPGM alone for 24, F26BPase alone for 54, and 7 residues were modeled ab initio. The greater presence of F26BPase is in agreement with the slightly closer positioning of the YhfR sequence with F26BPases (Fig. 2) compared to dPGMs. When the best template had been determined, the best scoring model by PROSA II was stereochemically analyzed in detail. Residues occupying disallowed or generously allowed areas of the Ramachandran plot (Laskowski et al. 1993; Kleywegt and Jones 1996) were remodeled, in some cases by peptide bond flipping, in others through ab initio regeneration with MODELLER. In this way, two rounds of model "breeding" led to the final model (Fig. 3A,B).

Analysis of the final model of the YhfR structure

The *B. stearothermophilus* YhfR sequence is only 26% identical to the dPGM template and 24% identical to the F26BPase template. The overall PROSA II score of the final model of −7.8, therefore, compares very favorably with the corresponding scores for the templates, both −9.6. The final model also has good stereochemical characteristics, containing no residues in disallowed areas of the Ramachandran plot, and scoring an overall PROCHECK G-factor of −0.11 (compared to 0.3 and 0.2 for the templates). Three regions in the final model have positive peaks in the PROSA II profile, possibly suggestive of incorrectly modeled portions, but also encountered in well-determined crystal structures (Sippl 1993) and sometimes associated with important functional parts of a protein (D.J. Rigden, unpubl. results). These three regions are situated at the amino terminus, the carboxyl terminus, and in the residues 103–120 of the model (corresponding to residues 106–123 of the dPGM template used for modeling). The amino-terminal positive peak is shared with the F26BPase template (data not shown), whereas the carboxy-terminal positive peak may be due to this part of the YhfR sequence adopting a conformation not seen in either template. The much shorter carboxy-terminal region of the YhfR model appears not to be associated with function as, unlike in the templates, it does not reach the vicinity of the active site cleft. The reason for the positive PROSA II profile of the third region seems to be the unusual solvent exposure of Ile103, Trp109, and Leu114 of YhfR (corresponding to positions 106, 112, and 117, respectively, in the dPGM template). Confidence in the correct modeling of this area is increased by the observation of the conservation of the Trp in the dPGM template and by the generally clear model–template alignment in the area. However, conformational differences with the template cannot be ruled out.

Inference of possible catalytic activity for YhfR

With the high quality of the structural model of YhfR established by objective protein structure validation methods, some tentative conclusions about the activity of YhfR were made. Sequence analyses already suggested that the YhfR sequence had key amino acid differences in the active site cleft compared to the dPGM and F26BPase, thereby suggesting some altered or novel catalytic activity for YhfR. However, it seemed possible that compensatory changes could have lead to retention of either dPGM or F26BPase activity. Therefore, possible activities were reexamined in the light of the modeled structure (Fig. 3A,B). Preliminary analysis showed that none of the unique insertions or deletions in the model had structural consequences for the active site cleft capable of entirely ruling out catalytic activity.

The model confirmed the lack of a 3-phosphate binding site that would be necessary for dPGM activity; the basic residues and hydrogen bond donors lost in comparison with the dPGM structure are not compensated for by other replacements. Furthermore, some features of the model were incompatible with the substrate reorientation proposed to occur during the dPGM reaction (Rigden et al. 1999) and presumably, would be involved in any PGM activity. For example, of the two arginine residues (Arg87 and Arg59; 5pgm numbering scheme) conserved in dPGMs and proposed to interact with phospho groups during reorientation, only one is conserved, the other being replaced by an isoleucine. It was also noted that residues from three different portions of the YhfR model (Asn16, Met21, Tyr175, and Glu207) would sterically clash with the reorientation of BPG during dPGM catalysis. Although a

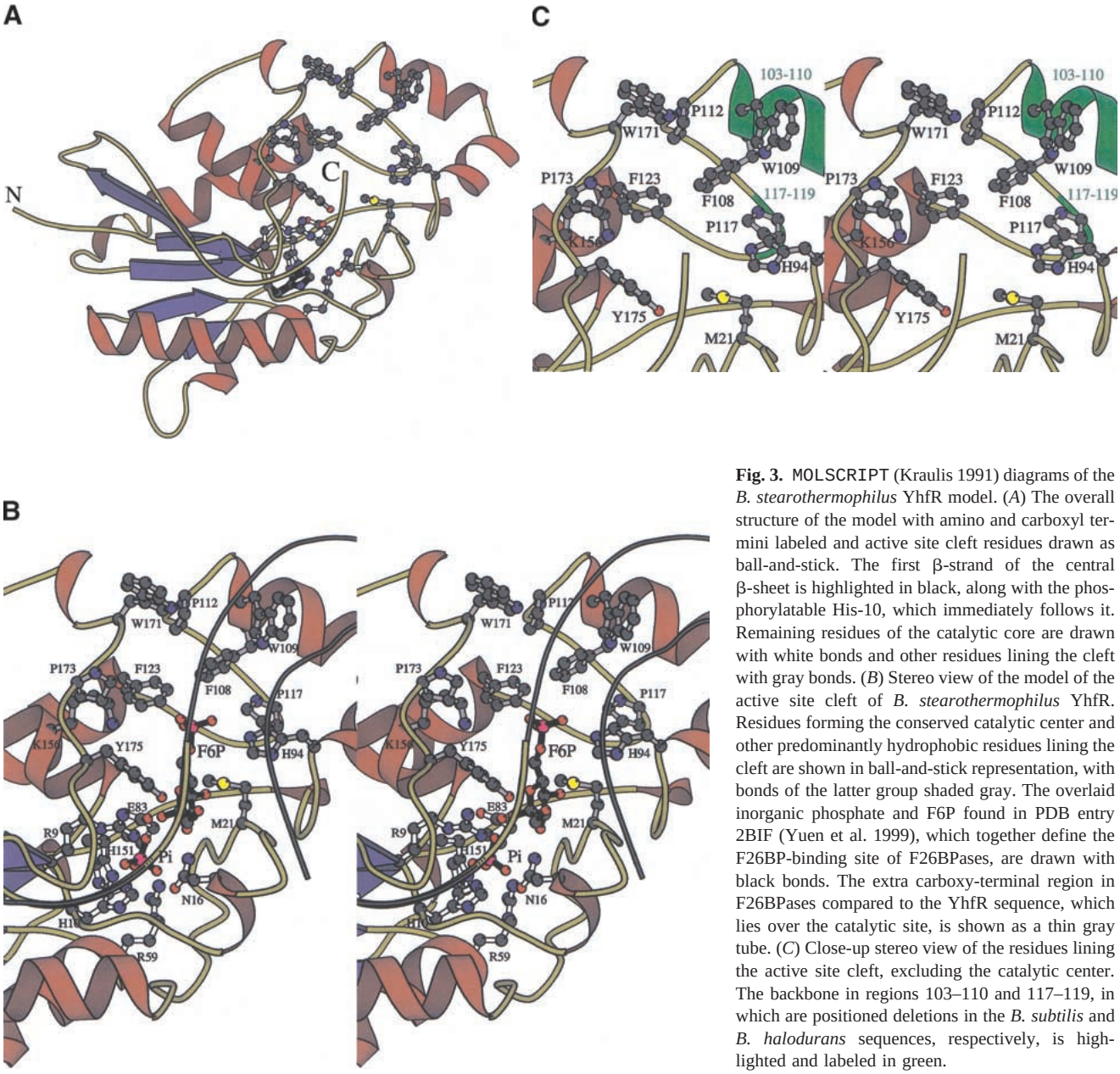

Fig. 3. MOLSCRIPT (Kraulis 1991) diagrams of the *B. stearothermophilus* YhfR model. (*A*) The overall structure of the model with amino and carboxyl termini labeled and active site cleft residues drawn as ball-and-stick. The first β-strand of the central β-sheet is highlighted in black, along with the phosphorylatable His-10, which immediately follows it. Remaining residues of the catalytic core are drawn with white bonds and other residues lining the cleft with gray bonds. (*B*) Stereo view of the model of the active site cleft of *B. stearothermophilus* YhfR. Residues forming the conserved catalytic center and other predominantly hydrophobic residues lining the cleft are shown in ball-and-stick representation, with bonds of the latter group shaded gray. The overlaid inorganic phosphate and F6P found in PDB entry 2BIF (Yuen et al. 1999), which together define the F26BP-binding site of F26BPases, are drawn with black bonds. The extra carboxy-terminal region in F26BPases compared to the YhfR sequence, which lies over the catalytic site, is shown as a thin gray tube. (*C*) Close-up stereo view of the residues lining the active site cleft, excluding the catalytic center. The backbone in regions 103–110 and 117–119, in which are positioned deletions in the *B. subtilis* and *B. halodurans* sequences, respectively, is highlighted and labeled in green.

model built on distant homology will inevitably contain errors, the above analysis clearly disfavored a PGM activity for YhfR.

As shown, whereas detailed sequence analysis ruled out simple dPGM or F26BPase activity, the structural model for YhfR shows that none of the insertions or deletions unique to YhfR have structural consequences for the active site cleft capable of completely ruling out any type of catalytic activity. Comparison of the structural model for YhfR with the structure of F26BPase, with which YhfR seems to share a slightly closer relationship, reveals a conserved catalytic machinery but a modified binding site (Fig. 3B). Among the residues contacting the sugar moiety of the substrate in F26BPase (Yuen et al. 1999) a reasonable degree of conservation is observed; F26BPase residue Glu325 (binding to atom O2) of the template is conserved in YhfR and the backbone H-bond from residue Gly268 of the template to O3 may be made. Gln391 of F26BPase is replaced with Gly in YhfR, but a replacement H-bond donor Tyr175 (replacing a Cys in F26BPase) is favorably placed. In addition, the F26BPase residue making hydrophobic contacts with the sugar ring Ile267 is replaced with Met in YhfR. In contrast, the 6-phosphate-binding region of F26BPase changes character dramatically in the YhfR model, retaining just two hydrophilic residues, His94 and Lys156. Among the noteworthy replacements are Arg350 (F26BPase) → Phe, Tyr365 → Phe, Lys354 → Pro, Tyr359 → Pro, Asp351 → Trp, and Tyr411 → Trp.

Compared to F26BPase, the active site of the YhfR model is more open overall, principally due to the absence in the YhfR model of the carboxy-terminal region that lies over the active site cleft. The active site cleft is also larger because of the use of the dPGM template alone for modeling of the region aligning with residues 405–411 of the F26BPase. This choice is supported by the highly favorable PROSA II profile of this region (data not shown). Also contributing to the large cleft in the active site in the YhfR model is the striking preponderance of the replacement of F26Bpase residues with smaller amino acids (Arg395 in F26BPase → Cys in the *B. stearothermophilus* sequence, Gln391 → Gly, Lys354 → Pro, Lys350 → Phe, Tyr359 → Pro, Tyr365 → Phe), compared to those positions where the side chain in F26Bpase is changed to a larger residue in YhfR (Cys414 → Tyr, Ala392 → Val). This situation is reminiscent of key differences previously noted between dPGMs and F26BPases and the more distantly related acid phosphatases. A conserved proline present in the latter but not in the former two enzymes impedes the formation of an α-helix present only in dPGMs and F26BPases. This helix and its adjoining loop serve to narrow the active site, thereby imparting the specificity of dPGMs and F26Bpases for smaller substrates (LaCount et al. 1998). Therefore, the more open active site of the acid phosphatases seems to be correlated with their broad substrate specificity and ability to bind to large substrates. Another factor involved in the definition of broad versus narrow substrate specificity has been identified from the structural comparison of the broad substrate specificity at pH 2.5 of acid phosphatase (Kostrewa et al. 1999) and the narrowly specific phytase (Kostrewa et al. 1997), both from *Aspergillus ficuum.* In the former enzyme only two negatively charged residues are present in the part of the catalytic site defining specificity, contrasting with the six charged residues in the equivalent part of the phytase. The conclusion drawn was that the more neutral electrostatic field of the pH 2.5 acid phosphatase places fewer constraints on the substrate's charge distribution, thereby broadening substrate specificity (Kostrewa et al. 1999). An analogous comparison can be made between the YhfR model and the F26BPase structure. They contain, respectively, two and four charged residues in the 6-phosphate-binding region of the F26BPase.

The conclusion from these analyses is that the YhfR protein seemed to possess the potential for catalytic activity, in accord with its likely expression in this bacterium by analogy with the situation in *B. subtilis* where *yhfR* is expressed, albeit at a relatively low level (Pearson et al. 2000). Sequence analyses and modeling studies showed that the activity could not be attributable to those currently known to be associated with the dPGM/F26BPase family—dPGM, F26Bpase, or R5PPase. Because the substrate reorientation associated with mutase activity (Rigden et al. 1999) seemed impossible for YhfR, its most likely activity is suggested to be a phosphatase, and the character of the binding site cleft is more suggestive of monophosphatase activity rather than bisphosphatase activity. The reasonable conservation of the sugar-binding residues of F26BPase suggested that compounds containing a sugar–phosphate group might be substrates, whereas the more open, neutral, and hydrophobic nature of the pocket suggested that a variety of hydrophobic substrates, even large ones, might also be capable of binding.

Phosphatase activity of B. stearothermophilus *YhfR*

Given the conclusions from the modeling analyses described, the *B. stearothermophilus* YhfR was purified as described in Materials and Methods to >98% homogeneity (Fig. 4); the molecular mass of the purified protein determined by SDS-polyacrylamide gel electrophoresis was ~25 kD, which agreed well with the calculated molecular mass for YhfR of 23,732 dalton (Fig. 4). The purified protein was assayed for phosphatase activity, and was found to exhibit this activity on nucleoside monophosphates, 3PGA, sugar phosphates, and two aromatic phosphomonoesters (Table 1). With *p*-nitrophenylphosphate as the substrate the phosphatase activity had a pH optimum of 6.2, a K_m of 3 mM, and was inhibited ≥95% by 50 mM phosphate. It is notable that the phosphatase specific activity of this enzyme with saturating levels of *p*-nitrophenylphosphate was 35 μmol/min/mg protein (note that the assays in Table 1 used 2.5 mM substrate), a value quite similar to that for *E. coli* alkaline phosphatase with this substrate (50 μmol/min/mg) (Malamy

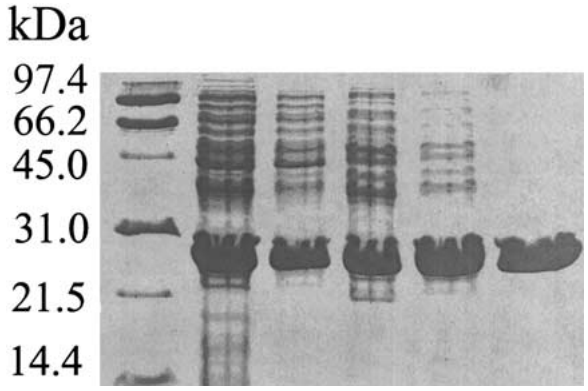

Fig. 4. SDS-polyacrylamide gel electrophoresis of *B. stearothermophilus* YhfR. Electrophoresis was performed in a 11% polyacrylamide gel under reducing conditions as described in Materials and Methods, and the gels were stained with Coomassie blue. The samples in the various lanes are: (lane *1*) molecular mass markers; (lane *2*) induced cell lysate (10 μL); (lane *3*) supernatant fraction from cell lysate (protein loaded on the DEAE-Sepharose column) (10 μL); (lane *4*) pooled sample after DEAE-Sepharose chromatography (10 μg); (lane *5*) pooled sample after Superdex 75 chromatography (10 μg); and (lane *6*) pooled sample after Mono-Q chromatography (final purified YhfR) (10 μg). The values to the left of the figure give the molecular mass of the markers in kD.

Table 1. *Phosphatase activity of purified* B. stearothermophilus *YhfR on various phosphomonoesters*[a]

Substrate	Specific activity (μmol/min/mg protein)
3PGA	27
α-napthylphosphate	21
p-nitrophenylphosphate	16
AMP	5
fructose-6-phosphate	1
ribose-5-phosphate	0.3
CMP	0.1

[a] Assays were carried out with purified *B. stearothermophilus* YhfR and 2.5 mM substrate by measuring the release of inorganic phosphate at pH 6.2 as described in Materials and Methods. A temperature of 65°C was used in the case of *B. stearothermophilus* YhfR and 37°C for *B. subtilis* YhfR.

and Horecker 1966), but only about one-tenth of the activity of the rat acid phosphatase with the same substrate (Himeno et al. 1988). The K_m value is within the range seen for acid phosphatases of 0.025–5.0 mM (van Etten 1982). In contrast to the high phosphatase activity of purified *B. stearothermophilus* YhfR, the specific activity of this protein in the dPGM assay was ≤70 nmol/min/mg protein; these latter assays were done under conditions where not all the 3PGA was dephosphorylated.

The experimental results agree remarkably well with the predictions made on the basis of molecular modeling. *B. stearothermophilus* YhfR is a phosphatase and is capable of hydrolyzing large hydrophobic substrates such as α-napthylphosphate. In addition, its ability to hydrolyze 3PGA highlights a wide substrate specificity. Interestingly, compounds containing a sugar–phosphate group are also substrates, as predicted by the reasonable conservation of the sugar-binding region of F26BPase in the *B. stearothermophilus* YhfR model. Which of the compounds tested, if any, is the physiological substrate of YhfR in vivo is an interesting question, although, as the example of the otherwise well-characterized prostatic acid phosphatase shows (LaCount et al. 1998), not one that is always easy to answer.

Other YhfR homologs in Bacilli

Sequence database searches show YhfR homologs to be present in the genomes of *Bacilli* and related gram-positive species such as *Clostridia*, despite these organisms containing functional iPGM homologues (Chander et al. 1998; Pearson et al. 2000). Comparison of the YhfR sequences from *B. subtilis* (Kunst et al. 1997; Pearson et al. 2000), *B. halodurans* (Takami et al. 2000), *B. anthracis* (unfinished genome project at http://www.tigr.org/), and two *B. stearothermophilus* strains (Fig. 5; Table 2) shows them to be, with the exception of the two *B. stearothermophilus* strains, highly diverse with only 28 positions conserved in all five sequences. (Note that the *B. halodurans* sequence present in the database lacks some amino-terminal residues compared with other sequences, but the use of an alternative Met upstream of the original start leads to a typical amino-terminal YhfR sequence; see Fig. 5.) The maximum sequence identity is 42% between the *B. stearothermophilus* sequences and that from *B. anthracis*. The *B. subtilis* and *B. halodurans* sequences are only 25–32% identical with the other sequences. With the exception of a clear evolutionary relationship between the *B. subtilis* and *B. halodurans* sequences, bootstrapping values related to bacterial YhfR homolog positioning are low (Fig. 2), indicating a lack of clear evolutionary positioning with respect to the dPGM, F26Bpase, and R5PPase groups. Similar conclusions could be drawn from phylogenies derived from the maximum parsimony method (data not shown), although this latter analysis suggested a closer evolutionary relationship between the *B. subtilis* and *B. halodurans* sequences. Therefore, these

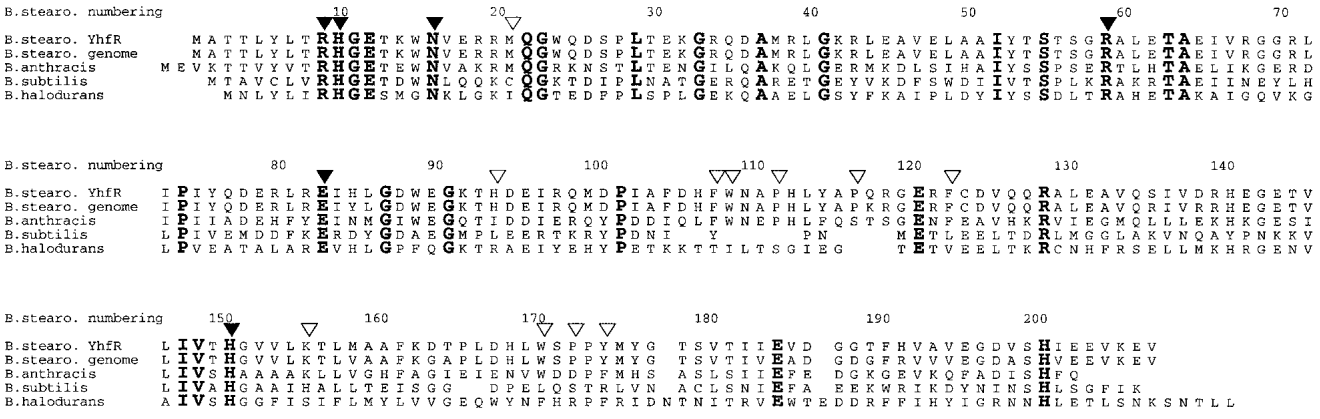

Fig. 5. Sequence alignment of *B. stearothermophilus* NGB101 YhfR with sequences of YhfR homologs in other *Bacillus* strains and species. Conserved residues are in bold and symbols mark residues lining the active site cleft (Fig. 3B); filled symbols denote the conserved catalytic core; open symbols denote other positions.

Table 2. *Pairwise identities between* B. stearothermophilus *NGB101YhfR and related sequences from other strains and species of* Bacilli[a]

	B. stearothermophilus NGB101	*B. stearothermophilus* 10	*B. anthracis*	*B. subtilis*
B. stearothermophilus 10	93			
B. anthracis	42	42		
B. subtilis	25	26	32	
B. halodurans	27	28	25	30

[a] Values given are the percent amino acid identities between YhfR homologs, as determined by the MODELLER algorithm (Sali and Blundell 1993).

bacterial sequences do not constitute a new group of closely related enzymes such as the dPGMs or F26BPases. Modeling suggested and assays proved that *B. stearothermophilus* NGB101 YhfR is a broad specificity phosphatase. Because the genomic sequence of *B. stearothermophilus* strain 10 *yhfR* predicts no changes in amino acid residues lining the active site cleft (Figs. 3 and 5), conclusions drawn for *B. stearothermophilus* strain NGB101 YhfR catalytic activity may safely be applied to the YhfR from strain 10.

In contrast, examination of residues in the conserved catalytic core as well as other residues predicted to line the active site cleft (Figs. 3 and 5) in the YhfR sequences from other *Bacillus* species yields different conclusions. The *B. anthracis* YhfR sequence is 42% identical to *B. stearothermophilus* YhfR. Comparison shows only three amino acid replacements in residues lining the active site cleft, none disturbing the core catalytic apparatus: His94 → Ile, Pro117 → Ser, and Pro173 → Asp. A phosphatase activity may be confidently predicted for this protein, with a similar, but perhaps not identical, substrate specificity profile to that of *B. stearothermophilus* YhfR. The catalytic core of *B. subtilis* YhfR also appears intact. In the remainder of the active site cleft, the residues lining the pocket seem to be slightly more hydrophilic than corresponding residues in the *B. stearothermophilus* protein, but still much more hydrophobic than F26BPases or dPGMs. Relative to the *B. stearothermophilus* sequence, the *B. subtilis* sequence has a deletion of residues 103–110. Because this region forms part of the binding pocket lining (Fig. 3), the *B. subtilis* active site cleft will have a very different structure to that of the *B. stearothermophilus* sequence. This analysis suggests that *B. subtilis* YhfR may also have phosphatase activity, as previously suggested (Pearson et al. 2000), but perhaps with a different substrate specificity than the *B. stearothermophilus* protein. Initial results have shown that *B. subtilis* YhfR purified as described previously and lacking detectable PGM activity (Pearson et al. 2000) does have phosphatase activity on *p*-nitrophenylphosphate, with a specific activity at saturating substrate concentration of ~15 μmol/min/mg protein, which is similar to that of *B. stearothermophilus* YhfR. The *B. halodurans* sequence has a three-residue deletion in comparison to the model structure residues 117–119, which will also change the shape of the pocket. In addition, comparison of residues predicted to line the active site cleft reveals not a single residue conserved between *B. stearothermophilus* and *B. halodurans* sequences. Once again the catalytic core is completely preserved, suggesting a probable phosphatase activity for the *B. halodurans* sequence, but with dramatically different substrate specificity.

Conclusions

The main conclusion of this work is that a hitherto unsuspected activity, broad specificity phosphatase, is present in the dPGM/F26BPase family of enzymes. Detailed sequence analysis and molecular modeling proved capable of making a strong prediction, subsequently borne out by experiment. This work adds further weight to the argument that extrapolation of sequence data into three-dimensional models yields additional insights. For this reason, bulk molecular modeling is often carried out for completed genomes (e.g., Sanchez and Sali 1998). However, it seems unlikely that an automatically generated model of the *B. stearothermophilus* YhfR sequence, using the distantly related templates available and produced without the rigorous modeling procedure applied here, would have led to such a strong conclusion; errors would likely have obscured important details. A challenge lies ahead in the construction of automatic methods of modeling more difficult cases. However, it is worth noting that our detailed sequence analysis was capable of casting serious doubt on the annotation of the *yhfR* gene of *B. stearothermophilus* as encoding either a dPGM or F26BPase. Cross-links between the Protein Data Base, in particular for important residues identified by contributors, and sequence databases also offer a route to improved genomic annotation.

Materials and methods

Cloning of B. stearothermophilus yhfR

The homolog of *B. subtilis yhfR* was cloned from *B. stearothermophilus* NGB101 (obtained from H.F. Foerster 1983) by PCR. The PCR primers were designed based on the sequence of the *B.*

subtilis yhfR homolog in *B. stearothermophilus* 10 (available at www.genome.ou.edu/bstearo.html) and were: yhfR-Nco, 5′-CAT GCC**ATG**GCGACACCTTGTATTTG-3′; and yhfr-Nde, 5′-GGA ATTCCATATG**CTA**TACTTCTTTTACTTCCTCC-3′. The primers also contained either *Nco*I or *Nde*I sites (underlined) with the ATG in boldface lettering in the *Nco*I site being the translational initiation codon, and the CTA in boldface lettering adjacent to the *Nde*I site being complementary to the translational stop codon. Note that these primers encompass regions encoding the first seven amino-terminal amino acids and the last six carboxy-terminal amino acids of the *B. stearothermophilus* 10 YhfR sequence, and this ensured that the residues in the YhfR sequence in strain NGB101 were identical to those in strain 10. The PCR fragment was digested with *Nco*I and *Nde*I, ligated into plasmid pET15b (Novagen), digested in the same manner, and the ligation mix was used to transform *E. coli* DH5α to ampicillin resistance, giving strain PS3296. The insert in the plasmid from strain PS3296 was sequenced and found to encode a protein of 208 amino acids, the same size as the protein encoded by the *yhfR* homolog from *B. stearothermophilus* 10. However, there were 15 amino acid differences between the proteins encoded by the *yhfR* homologs of the two *B. stearothermophilus* strains (see below) and 35 differences at the DNA sequence level. The large number of differences between the *yhfR* homologs of the two strains was not the result of PCR errors in the isolation of *yhfR* from strain NGB101, as identical DNA sequences were obtained from two independent PCR amplifications. The DNA sequence determined for *B. stearothermophilus* NGB101 *yhfR* has been submitted to GenBank under accession number AF343668. The plasmid from strain PS3297 was used to transform *E. coli* BL21 (DE3) to ampicillin resistance, giving strain PS3297, which was used to overexpress *B. stearothermophilus* YhfR.

Expression and purification of B. stearothermophilus *YhfR*

E. coli strain PS3297 was grown overnight at 37°C in 25 mL of Luria-Bertani (LB) medium (Sambrook et al. 1989) with 50 μg/mL ampicillin. These cells were used to inoculate a 1-L culture of LB medium containing 50 μg/mL ampicillin, which was grown at 37°C to an OD_{600} of 0.8 and isopropyl-thio-β-D-galactopyranoside then added to 1 mM to induce YhfR synthesis. After 2.5 h of further growth the cells were harvested by centrifugation (30 min; 6000*g*) at 4°C. The cell pellet was resuspended in 200 mL of cold buffer A [20 mM Tris-HCl at pH 7.4, 1 mM dithiothreitol (DTT), and 2 mM EDTA] containing 15% glycerol. The cells were disrupted by sonication with 5 × 20-sec pulses with 1-min intervals to allow cooling, and then centrifuged (20 min; 30,000*g*). All subsequent steps were at 4°C. The supernatant fluid was loaded at 0.7 mL/min onto a 100-mL DEAE–Sepharose ion-exchange column equilibrated in buffer A. The column was washed with four column volumes of buffer A, and protein eluted with a 1-L gradient of 0 to 0.5 M NaCl gradient in buffer A. Fractions containing YhfR were identified by SDS-polyacrylamide gel electrophoresis as described below, pooled, concentrated to 10 mL using an Ultrafree-15 centrifugal filter device (Millipore), and loaded and run on a 26/60 Superdex 75 size exclusion column (Amersham Pharmacia Biotech) using buffer B (10 mM Tris-HCl at pH 7.4, 1 mM DTT, 2 mM EDTA, 50 mM NaCl). The YhfR fractions were again pooled and loaded at 0.8 mL/min onto a 10/10 Mono-Q ion-exchange column (Amersham Pharmacia Biotech) equilibrated in buffer B, and YhfR was eluted using a 100-mL gradient of 50 to 500 mM NaCl in buffer B. Fractions containing homogeneous YhfR were pooled and concentrated to 1 mg/mL using an Ultrafree-15 centrifugal filter device (Millipore). The yield of purified YhfR was 18 mg/L of the original culture.

Enzyme assays

dPGM activity was measured using the two-step assay with 3PGA and BPG present in the first step as described (Chander et al. 1998); the first step of the reaction (conversion of 3PGA to 2PGA) was run at 65°C and the second step (assay of 2PGA) was run at 37°C. Phosphatase activity was measured by the release of either *p*-nitrophenol from *p*-nitrophenylphosphate or inorganic phosphate from various phosphomonoesters. Phosphatase assays were in 1 mL of 0.1 M 2-[*N*-morpholino]ethanesulfonic acid at pH 6.2 unless otherwise noted. Reaction mixes without enzyme or substrate were preincubated at 65°C (*B. stearothermophilus* YhfR) or 37°C (*B. subtilis* YhfR) for 10 min, reactions started by the addition of the substrate (routinely to 2.5 mM) and enzyme incubated another 10–15 min at the preincubation temperature. For assay of *p*-nitrophenol release, 0.2 mL of 1 M NaOH was added to stop the reaction and the OD_{405} measured; for analysis of phosphate release 0.3-ml aliquots of the reaction mix were assayed for inorganic phosphate as described (Ames 1966). In all cases, enzyme activities were corrected for nonenzymatic substrate hydrolysis.

Sequence analysis

To characterize the relationships of *B. stearothermophilus* YhfR with other sequences, database searches were carried out using `FASTA` (Pearson and Lipman 1988), `BLAST2` (Altschul et al. 1990), `PSI-BLAST` (Altschul et al. 1997), and Hidden Markov Model methods (Karplus et al. 1998). Analyses with threading programs were carried out to determine the compatibility of the sequence of *B. stearothermophilus* YhfR with known protein folds. The methods used were `GenTHREADER` (Jones 1999), `3D-PSSM` (Kelley et al. 2000), and the consensus analysis carried out by the Bioinbgu server (Fischer 2000). The alignment of YhfR and related sequences was produced using only the better characterized enzymes in the ENZYME database (Bairoch 2000) supplemented with four plant F26BPases whose functional annotation seems clear (Villadsen et al. 2000). Conserved sequence motifs within groups of dPGMs and F26BPases were obtained using the MEME software (Grundy et al. 1996). The program `Jalview` (available at http://circinus.ebi.ac.uk:6543/jalview/) was used to manipulate alignments.

Phylogenetic analysis

Phylogenetic relationships were inferred by distance matrix and maximum parsimony methods using the `PHYLIP` package (Felsenstein 1989). The distance method was used in conjunction with the neighbor-joining algorithm (programs `Protdist` and `Neighbor`). The maximum parsimony method used the program `Protpars`. Boostrapping (100 replicates) was carried out in both cases.

Model construction

The limited sequence identity between YhfR and the available templates (~26% with yeast dPGM and 24% with F26BPase) necessitated the application of a rigorous modeling strategy in which construction and evaluation of multiple models is used to validate

the target-templates alignment (e.g., Rigden and Carneiro 1999; Rigden et al. 2000). In this way, 15 models were built for each target-templates alignment tested with YhfR. Protein models were constructed using the MODELLER-4 package (Sali and Bundell 1993). A 3-Å coordinate randomization step was applied before refinement of the models to sample coordinate space. The set of models was then analyzed for solvent exposure and packing with PROSA II (Sippl 1993) and for stereochemical properties with PROCHECK (Laskowski et al. 1993). Structural superpositions were made with LSQMAN (Kleywegt 1996) and structures were visualized with O (Jones et al. 1991).

Other Methods

Electrophoresis was carried out under reducing conditions in 11% SDS-polyacrylamide gels using the buffer system described by Laemmli (1970) and a Mini Protein II gel system (Bio-Rad). The gels were stained with Coomassie blue. Protein concentrations were determined by UV absorption at 280 nm using a molar extinction coefficient for *B. stearothermophilus* YhfR calculated based on the protein sequence (Pace et al. 1995).

Acknowledgments

Sequencing of the *B. anthracis* genome is being carried out by the Institute for Genome Research with support from the Office of Naval Research, the Department of Energy, and the National Institute of Allergy and Infectious Disease. The *B. stearothermophilus* Genome Sequencing Project at the University of Oklahoma is funded by a National Science Foundation EPSCoR Program grant, EPS-9550478. Work in the authors' laboratories was supported by grants GM19698 (PS), AI44079 (MJJ) from the National Institutes of Health, and N66001-01-C-8013 from Defense Advanced Research Projects Agency.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389–3402.

Ames, B.N. 1966. Assay of inorganic phosphate, total phosphate and phosphatases. *Method Enzymol.* **8:** 115–118.

Apostol, I., Kuciel, R., Wasylewska, E., and Ostrowski, W.S. 1985. Phosphotyrosine as a substrate of acid and alkaline phosphatases. *Acta Biochim. Pol.* **32:** 187–197.

Bairoch, A. 2000. The ENZYME database in 2000. *Nucleic Acids Res.* **28:** 304–305.

Barton, G.J. 1993. ALSCRIPT, a tool to format multiple sequence alignments. *Prot. Eng.* **6:** 37–40.

Bates, P.A. and Sternberg, M.J.E. 1999. Model building by comparison at CASP3: Using expert knowledge and computer automation. *Proteins* (Suppl.) **3:** 47–54.

Bond, C., White, M.F., and Hunter, W.N. 2001. High-resolution structure of the phosphohistidine-activated form of *Escherichia coli* cofactor-dependent phosphoglycerate mutase. *J. Biol. Chem.* **276:** 3247–3253.

Burland, V.D., Plunkett, G. III, Sofia, H.J., Daniels, D.L., and Blattner F.R. 1995. Analysis of the *Escherichia coli* genome VI: DNA sequence of the region from 92.8 through 100 minutes. *Nucleic Acids Res.* **23:** 2105–2119.

Chander, M. Setlow, B., and Setlow, P. 1998. The enzymatic activity of phosphoglycerate mutase from gram-positive endospore-forming bacteria requires Mn^{2+} and is pH sensitive. *Can. J. Microbiol.* **44:** 759–767.

Felsenstein, J. 1989. PHYLIP—Phylogeny Inference Package (version 3.2). *Cladistics* **5:** 164–166.

Fischer, D. 2000. Hybrid fold recognition: Combining sequence derived properties with evolutionary information. In *Proceedings of the Pacific Symposium on Biocomputing, 2000,* pp. 119–130. Pacific Symp. Biocomputing, Hawaii.

Foerster, H.F. 1983. Activation and germination characteristics observed in endospores of thermophilic strains of *Bacillus*. *Arch. Microbiol.* **134:** 175–181.

Fothergill-Gilmore, L.A. and Watson, H.C. 1989. The phosphoglycerate mutases. *Advan. Enzymol. Rel. Areas Mol. Biol.* **62:** 227–313.

Fraser, H.I., Kvaratskhelia, M., and White, M.F. 1999. The two analagous phosphoglycerate mutases of *Escherichia coli*. *FEBS Lett.* **455:** 344–348.

Grundy, W.N., Bailey, T.L., and Elkan, C.P. 1996. ParaMEME: A parallel implementation and a web interface for a DNA and protein motif discovery tool. *Comput. Appl. Biosci.* **12:** 303–310.

Han, C.H. and Rose, Z.B. 1979. Active site phosphohistidine peptides from red cell bisphosphoglycerate synthase and yeast phosphoglycerate mutase. *J. Biol. Chem.* **254:** 8836–8840.

Hasemann, C.A., Istvan, E.S., Uyeda, K., and Deisenhofer, J. 1996. The crystal structure of the bifunctional enzyme 6-phosphofructo-2-kinase/fructose-2,6-bisphosphatase reveals distinct domain homologies. *Structure* **4:** 1017–1029.

Heinisch, J.J., Mueller, S., Schlueter, E., Jacoby, J., and Rodicio R. 1998. Investigation of two yeast genes encoding putative isoenzymes of phosphoglycerate mutase. *Yeast* **14:** 203–213

Himeno, M., Koutoku, H., Tsuji, H., and Kato, K. 1988. Purification and characterization of acid phosphatase in rat liver lysosomal contents. *J. Biochem.* **104:** 773–776.

Jedrzejas M.J. 2000. Structure, function, and evolution of phosphoglycerate mutases: comparison with fructose-2,6-bisphosphatase, acid phosphatase, and alkaline phosphatase. *Prog. Biophys. Mol. Biol.* **73:** 263–287.

Jedrzejas, M.J. and Setlow, P. 2001. Comparison of the binuclear metalloenzymes diphosphoglycerate-independent phosphoglycerate mutase and alkaline phosphatase: their mechanism of catalysis via a phosphoserine intermediate *Chem. Rev.* **101:** 607–618.

Jones, D.T. 1999. GenTHREADER: An efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* **287:** 797–815.

Jones, T.A., Zou, J-Y., Cowan, S.W., and Kjeldgaard, M. 1991. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta. Crystallogr.* **A47:** 110–119.

Karplus, K., Barrett, C., and Hughey, R. 1998. Hidden Markov models for detecting remote protein homologies. *Bioinformatics* **14:** 846–856.

Kelley, L.A., MacCallum, R.M., and Sternberg, M.J.E. 2000. Enhanced genome annotation using structural profiles in the program 3D-PSSM. *J. Mol. Biol.* **299:** 501–522.

Kleywegt, G.J. 1996. Use of non-crystallographic symmetry in protein structure refinement. *Acta Crystallogr.* **D52:** 842–857.

Kleywegt, G.J. and Jones, T.A. 1996. Phi/Psi-chology: Ramachandran revisited. *Structure* **4:** 1395–1400.

Kostrewa, D., Gruninger-Leitch, F., D'Arcy, A., Broger, C., Mitchell, D., and van Loon, A.P. 1997. Crystal structure of phytase from *Aspergillus ficuum* at 2.5 A resolution. *Nat. Struct. Biol.* **4:** 185–190.

Kostrewa, D., Wyss, M., D'Arcy, A., and van Loon, A.P. 1999. Crystal structure of *Aspergillus niger* pH 2.5 acid phosphatase at 2.4 Å resolution. *J. Mol. Biol.* **288:** 965–974.

Kraulis J. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24:** 946–950.

Kunst, F., Ogasawara, N., Moszer, I., Albertini, A.M., Alloni, G., Azevedo, V., Bertero, M.G., Bessieres, P., Bolotin, A., Borchert, S., et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390:** 249–256.

LaCount, M.W., Handy, G., and Lebioda, L. 1998. Structural origins of L(+)-tartrate inhibition of human prostatic acid phosphatase. *J. Biol. Chem.* **273:** 30406–30409.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227:** 680–685.

Laskowski, R., MacArthur, M., Moss, D., and Thornton, J. 1993. PROCHECK: A program to check stereochemical quality of protein structures. *J. Appl. Crystallog.* **26:** 283–290.

Leyva-Vazquez, M. and Setlow, P. 1994. Cloning and nucleotide sequence of the genes encoding triose phosphate isomerase, phosphoglycerate mutase and enolase from *Bacillus subtilis*. *J. Bacteriol.* **176:** 3903–3910.

Lin, K., Li, L., Correia, J.J., and Pilkis, S.J. 1992. Glu327 is part of the catalytic

triad in rat liver fructose-2,6-bisphosphatase. *J. Biol. Chem.* **267:** 6556–6562.

Malamy, M. and Horecker, B.L. 1966. Alkaline phosphatase (crystalline). *Method Enzymol.* **9:** 639–642.

Mizuguchi, H., Cook, P.F., Tai, C.-H., Hasemann, C.A., and Uyeda, K. 1999. Reaction mechanism of fructose-2,-6-bisphosphatase. A mutation of nucleophilic catalyst, histidine 256, induces an alteration in the reaction pathway. *J. Biol. Chem.* **274:** 2166–2175.

Murzin, A.G., Brenner, S.E., Hubbard, T., and Chothia, C. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247:** 536–540.

Ostanin K., Saeed A., and Van Etten R.L. 1994. Heterologous expression of human prostatic acid phosphatase and site-directed mutagenesis of the enzyme active site. *J. Biol. Chem.* **269:** 8971–8978.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray T. 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci.* **4:** 2411–2423.

Pearson, C. L., Loshon, C. A., Pedersen, L.B., Setlow, B., and Setlow P. 2000. Analysis of the function of a putative 2,3-diphosphoglyceric acid-dependent phosphoglycerate mutase from *Bacillus subtilis*. *J. Bacteriol.* **182:** 4121–4123.

Pearson, W.R. and Lipman, D.J. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci.* **85:** 2444–2448.

Pilkis, S.J., Lively, M.O., and el-Maghrabi, M.R. 1987. Active site sequence of hepatic fructose-2,6-bisphosphatase. Homology in primary structure with phosphoglycerate mutase. *J. Biol. Chem.* **262:** 12672–12765.

Rigden, D.J. and Carneiro, M. 1999. A structural model for the rolA protein and its interaction with DNA. *Proteins* **37:** 697–708.

Rigden, D.J., Walter, R.A., Phillips, S.E.V., and Fothergill-Gilmore, L.A. 1999. Sulphate ions observed in the 2.12Å structure of a new crystal form of *Saccharomyces cerevisiae* phosphoglycerate mutase provide insights into understanding the catalytic mechanism. *J. Mol. Biol.* **286:** 1507–1517.

Rigden, D.J., Mello, L.V., and Bertioli, D.J. 2000. Structural modeling of a plant disease resistance gene product domain. *Proteins* **41:** 133–143.

Sali, A. and Blundell, T.L. 1993. Comparative protein modeling by satisfaction of spatial restraints. *J. Mol. Biol.* **234:** 779–815.

Sali, A., Sanchez, R., and Badretdinov, A. 1997. *MODELLER:* A program for *protein structure prediction.* Release 4 Manual. The Rockefeller University, New York, NY.

Sambrook, J., Fritsch, E.T., and Maniatis, T. 1989. *Molecular cloning, A laboratory manual.* Cold Spring Harbor Press, Cold Spring Harbor, NY.

Sanchez, R. and Sali, A. 1998. Large-scale protein structure modeling of the *Saccharomyces cerevisiae* genome. *Proc. Natl. Acad. Sci.* **95:** 13597–13602.

Schneider, G., Lindqvist, Y., and Vihko, P. 1993. Three-dimensional structure of rat acid phosphatase. *EMBO J.* **12:** 2609–2615.

Sippl, M.J. 1993. Recognition of errors in three-dimensional structures of proteins. *Proteins* **17:** 355–362.

Takami, H., Nakasone, K., Takaki, Y., Maeno, G., Sasaki, R., Masui, N., Fuji, F., Hirama, C., Nakamura, Y., Ogasawara, N., Kuhara, S., and Horikoshi, K. 2000. Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*. *Nucleic Acids Res.* **28:** 4317–4331.

van Etten, R.L. 1982. Human prostatic acid phosphatase: a histidine phosphatase. *Ann. N. Y. Acad. Sci.* **390:** 27–51.

van Etten, R.L. and Waymack, P.P. 1991. Substrate specificity and pH dependence of homogeneous wheat germ acid phosphatase. *Arch. Biochem. Biophys.* **288:** 634–645.

Villadsen D., Rung J.H., Draborg H., and Nielsen T.H. 2000. Structure and heterologous expression of a gene encoding fructose-6-phosphate,2-kinase/fructose-2,6-bisphosphatase from *Arabidopsis thaliana*. *Biochim. Biophys. Acta* **1492:** 406–413.

White, M.F., Fothergill-Gilmore, L.A., Kelly, S.M., and Price, N.C. 1993. Substitution of His-181 by alanine in yeast phosphoglycerate mutase leads to cofactor-induced dissociation of the tetrameric structure. *Biochem. J.* **291:** 479–483.

Yuen, M.H., Mizuguchi, H., Lee, Y.H., Cook, P.F., Uyeda, K., and Hasemann, C.A. 1999. Crystal structure of the H256A mutant of rat testis fructose-6-phosphate,2-kinase/fructose-2,6-bisphosphatase. Fructose 6-phosphate in the active site leads to mechanisms for both mutant and wild type bisphosphatase activities. *J. Biol. Chem.* **274:** 2176–2184.