Unexpected catalytic site variation in phosphoprotein phosphatase homologues of cofactor-dependent phosphoglycerate mutase

Daniel J. Rigden*

Embrapa Genetic Resources and Biotechnology, Cenargen/Embrapa, Parque Estação Biológica, Final W3 Norte, 70770-900 Brasília, Brazil

Received 30 October 2002; revised 7 January 2003; accepted 7 January 2003

First published online 15 January 2003

Edited by Robert B. Russell

Abstract The cofactor-dependent phosphoglycerate mutase (dPGM) superfamily contains, besides mutases, a variety of phosphatases, both broadly and narrowly substrate-specific. Distant dPGM homologues, conspicuously abundant in microbial genomes, represent a challenge for functional annotation based on sequence comparison alone. Here we carry out sequence analysis and molecular modelling of two families of bacterial dPGM homologues, one the SixA phosphoprotein phosphatases, the other containing various proteins of no known molecular function. The models show how SixA proteins have adapted to phosphoprotein substrate and suggest that the second family may also encode phosphoprotein phosphatases. Unexpected variation in catalytic and substrate-binding residues is observed in the models.

Key words: Cofactor-dependent phosphoglycerate mutase; Fructose-2,6-bisphosphatase; Evolutionary relationship; Sequence analysis

1. Introduction

The cofactor-dependent phosphoglycerate mutase (dPGM) superfamily [1] is an interesting example of functional diversity associated with a single protein fold. Within the superfamily, several experimentally verified phosphatase activities on diverse substrates are known, some narrowly specific [2–4], others possessing broad specificity [5]. The best known of these enzymes, fructose-2,6,-bisphosphatase (F26BPase; [2]), catalyses an entirely different reaction from the founder member of the superfamily, dPGM [6] (see Table 1), but sequence conservation at the catalytic core common to both led to a predicted evolutionary relationship for the pair [7], as later confirmed by structural comparison [8]. The subsequent identification of the YhfR gene product (lately renamed PhoE) as a broad specificity phosphatase [5], not a dPGM as annotated in the databases, highlighted the difficulties in assigning function in the dPGM superfamily based on sequence comparison alone. Evolutionarily more distant relatives are also known – the phytases [9] and acid phosphatases [10]. Although modern computational methods readily identify these as dPGM homologues based on sequence analysis alone, at the time of the structural determination of the first representative [10], their evolutionary relationship with dPGMs was unsuspected.

The advent of the genome sequencing project has revealed the full sequence diversity of the dPGM superfamily. As pointed out [11,12], many homologues bear only very low sequence identity to the known structures, suggesting that much catalytic diversity remains to be uncovered. Here we study, by sequence analysis and molecular modelling, two further groups of dPGM homologues – the SixA and Ais families. In the former case, the presence of dPGM-like motifs has been previously noted [13]. In the latter case, we report the relationship with dPGMs for the first time. The SixA proteins act as phosphatases on phosphohistidyl–ArcB [14] and thereby influence the ArcB phosphorelay signaling system. The Ais group, of currently unknown molecular function, include aluminium-inducible proteins [15] and related proteins found in pilin- and antimicrobial peptide resistance-related operons [16,17]. Molecular modelling suggests that they too may act as phosphoprotein phosphatases. In both families, the presumed acid–base catalyst is unexpectedly different from the canonical glutamate present in dPGMs, F26BPases and PhoE.

2. Materials and methods

BLAST [18] and PSI-BLAST [19] were used to carry out sequence searches in the nr and unfinished genome databases (http://www.ncbi.nlm.nih.gov/BLAST/). The SixA and Ais families were aligned with T-coffee [20]. Alignment manipulations were carried out with Jalview (http://www.ebi.ac.uk/~michele/jalview/) and ESPRIPT [21]. Cellular localisation predictions were made with PSORT [22] and TARGETP [23]. Signal peptide cleavage sites were predicted with SignalP [24]. Convenient access to a variety of fold recognition methods through the Meta-server [25] enabled the determination of the most suitable templates for model construction. PSI-PRED [26] was used to predict secondary structure. Sequence relationships between families of the dPGM superfamily were further investigated using the SEQBOOT, PROTPARS and CONSENSE programs of the PHYLIP package [27].

Model construction was carried out using MODELLER-6 [28] within an iterative modelling scheme. Initial target–template alignments derived from fold recognition results were modified in the light of visual examination of template structures in order to best position insertions and deletions. In regions where the templates were significantly structurally divergent, the choice of which to include in the modelling procedure was made based on sequence length and characteristics in comparison with the corresponding region of the target. Models derived from initial alignments were subjected to packing analysis with PROSA II [29] and stereochemical analysis with PROCHECK [30]. Regions of atypical packing (positive stretches in the resulting PROSA II profiles) or stereochemical problems were subjected to alignment adjustments and new models made. The adjustments were either alignment shifts or changes in the template(s) used

*Corresponding author. Fax: (55)-61-340 3658.
E-mail address: daniel@cenargen.embrapa.br (D.J. Rigden).

0014-5793/03/$22.00 © 2003 Published by Elsevier Science B.V. on behalf of the Federation of European Biochemical Societies.
doi:10.1016/S0014-5793(03)00014-0

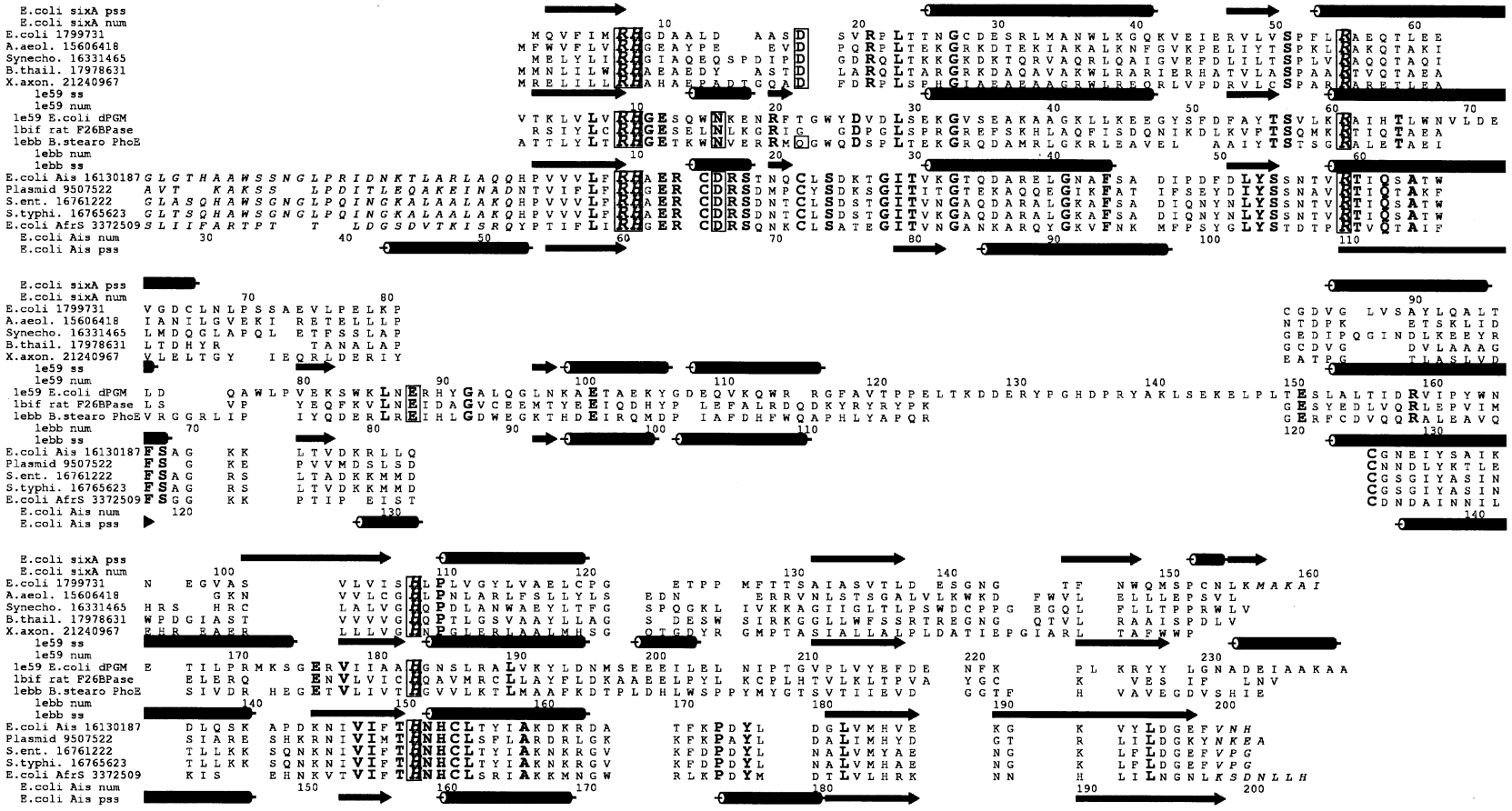

Fig. 1. ALSCRIPT [45] alignment of five maximally diverse SixA sequences, the five Ais sequences and the templates used for molecular modelling. Sequences are labelled with species name and GenBank accession numbers while templates are labelled with PDB code, species and abbreviated activity. Invariant residues in the entire SixA and Ais alignments are shown in bold. Boxes mark residues of the catalytic core and phosphate pocket with the former additionally italicised. Secondary structures of the templates were assigned with STRIDE [33] while PSIPRED [26] was used for secondary structure predictions of the targets.

Table 1
Comparison of final models with experimentally determined structures

	EC number of catalytic activity	Method	Size of structure (residues)	PROSA II score [29]	Residues in core area of Ramachadran plot (%) [30]	Number of residues in disallowed areas of Ramachadran plot [30]	χ_1/χ_2 outliers [30]
E. coli SixA	3.1.3.-	comparative modelling	156	−7.76	93.2	0	4
E. coli Ais	3.1.3.-(?)	comparative modelling	148	−8.53	84.3	0	1
B. stearothermophilus PhoE (1ebb)	3.1.3.-	X-ray crystallography	202	−8.88	89.2	0	1
E. coli dPGM (1e59)	5.4.2.1	X-ray crystallography	239	−9.93	93.9	0	1
Rat testis F26BPase domain (1bif)	3.1.3.46	X-ray crystallography	190	−9.45	88.8	0	1
Schizosaccharomyces pombe dPGM (1fzt; model 21)	5.4.2.1	nuclear magnetic resonance	211	−7.61	78.3	1	16

for a particular region. Alignment adjustments resulting in improved models were accepted and new improvements sought until no more could be obtained.

O [31] was used for visualisation of protein structures, LSQMAN [32] for protein structural superpositions and STRIDE [33] for the assignment of secondary structural elements in protein structures.

3. Results and discussion

3.1. Sequence analysis

Using *Escherichia coli* SixA and Ais as probes for database searches produced families of 30 and 5 non-identical sequences, respectively. Pairwise sequence identity in the SixA alignment was in the range of 15–99%, mean 32%. In the Ais alignment, 34–98% pairwise sequence identity was observed, mean 49%. This sequence diversity leads to just 10 and 40 invariant positions in the SixA and Ais alignments, respectively (Fig. 1). Remarkably, simple BLAST [18] searches with both SixA and Ais failed to assign significant *E*-values to members of other dPGM-related families. The best score obtained was for bullfrog F26BPase, which received an *E*-value of 0.036 in the Ais BLAST results. However, PSI-BLAST [19] (employing an iteration cut-off of 0.001) clearly demonstrated these relationships as early as the 2nd iteration for SixA (*E*-value of 3×10^{-6} for *Caenorhabditis elegans* F26BPase) and the 4th iteration in the case of Ais (*E*-value of 2×10^{-9} for *Bacillus stearothermophilus* PhoE).

The Ais family members, in contrast to SixA and indeed differently from dPGMs, F26BPase and PhoE, all contain clear predicted signal peptides specifying their secretion. Some members of the Ais family are annotated as containing N-terminal transmembrane domains but this probably reflects the known problem of distinguishing between signal peptides and transmembrane domains [34], since PSORT [22] and TARGETP [23] results favour periplasmic location. Secretion of the Ais family is also strongly supported by conservation of four cysteine residues suitably placed to form two disulphide bonds (see below). Analysis of the SixA and Ais families using the PHYLIP package [27], in comparison with other dPGM superfamily members of diverse activities, suggests that they have only a distant evolutionary relationship and bear no particularly close relationship to any well-characterised family (Fig. 2).

3.2. Model construction

In order to attempt to understand better the structure–function relationship in the SixA and Ais families we constructed molecular models of representative members – *E. coli* SixA (GenBank accession number 1799731) and *E. coli* Ais (GenBank accession number 16130187). Fold recognition experiments using the Meta-server [25] confirmed the dPGM fold for these two proteins. For example, the Pcons2 consensus method gave scores of up to 3.8 and 8.4 for SixA and Ais, respectively. In comparison, the top scoring false positive score for Pcons2, as monitored by the Livebench experiment [35], is currently 2.6. Although the fold assignments were clear, in neither case was a single existing structure strongly favoured over other related structures. We therefore selected three different structures – *B. stearothermophilus* PhoE (1ebb; [11]), rat testis F26BPase (1bif; [36]) and *E. coli* dPGM (1e59; [37]) – information from all of which was incorporated into the model construction process.

The target sequences, *E. coli* SixA and Ais, share only 14–

17% and 14–22% sequence identity, respectively, with the templates. Nevertheless, rigorous modelling methodology can produce useful models even in such difficult cases (e.g. [5]). As the data in Table 1 show, the final models compare favourably with experimentally determined structures. For example, the PROSA II [29] scores, analyzing residue packing characteristics and solvent exposure, approach those of the experimentally determined structures, even though the models are significantly smaller. Small C-terminal segments and 27 N-terminal Ais residues (following the predicted signal peptide cleavage site) were not modelled through the absence of suitable template structure (Fig. 1). In initial models of *E. coli* Ais, its four cysteine residues, each entirely conserved, were positioned in two spatially close pairs. The Sγ–Sγ separations of Cys65–Cys72 and Cys133–Cys161 were around 3.2 and 5.4 Å, respectively. This suggested the existence of two disulphide bonds in the Ais family, consistent with its predicted extracellular location, and evidence of an accurate target–template alignment. Subsequent models were therefore generated to include these two disulphide bonds.

3.3. Overall model analysis

The substrate of SixA, phosphorylated ArcB protein [14], is radically different from the substrates of better characterised members of the dPGM superfamily. dPGMs and F26BPase exhibit high specificity for their respective small, charged substrates, phosphoglycerate and F26BP. The more recently characterised PhoE is a broadly specific phosphatase, but its largest known substrate is naphthylphosphate [5]. In terms of molecular mass, therefore, phosphorylated ArcB is around

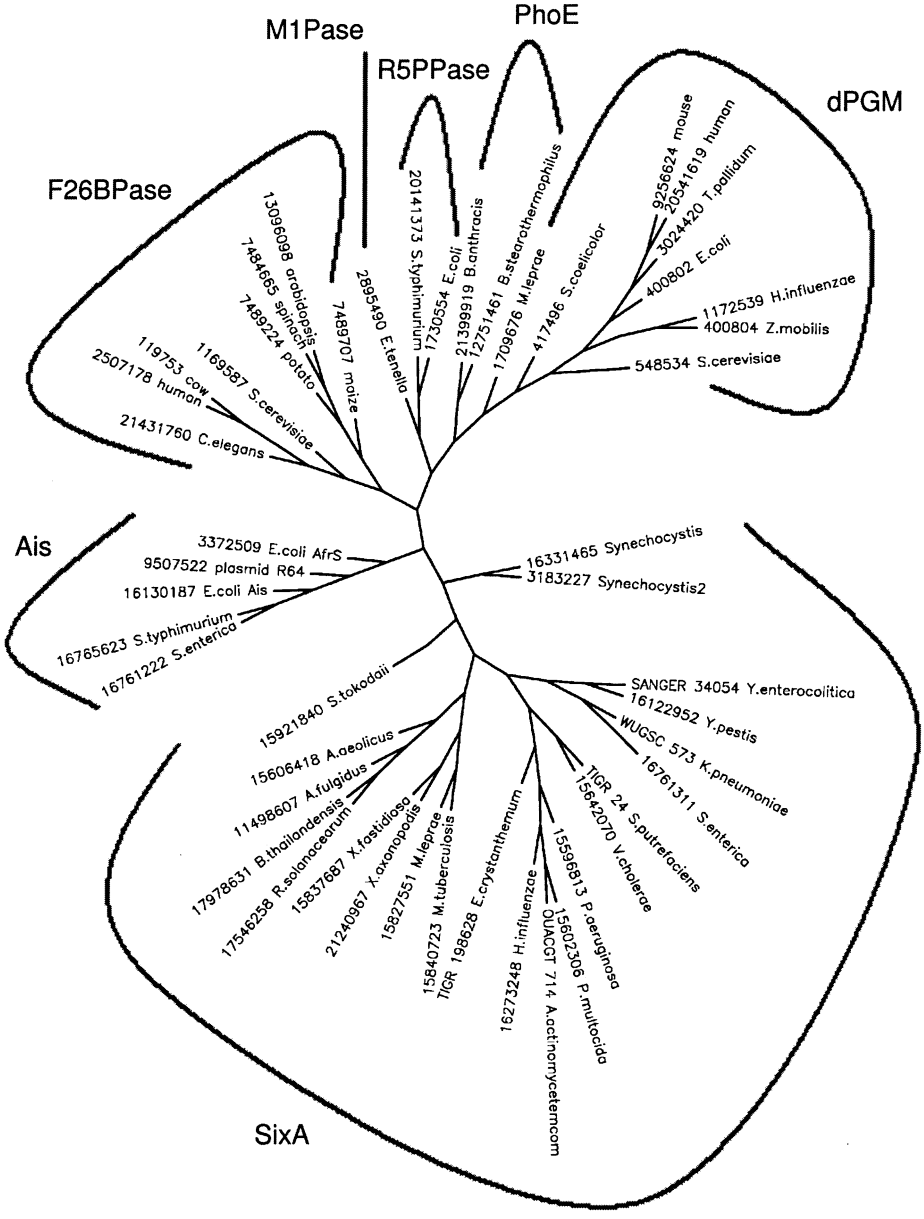

Fig. 2. Consensus maximum parsimony tree showing the distant evolutionary relationships between SixA, Ais and the other experimentally characterised families shown. Each sequence is given its GenBank accession number (or genome project code in the case of unfinished genome projects) and species name. The tree was calculated using the PHYLIP package [27] with the programs SEQBOOT (used to generate 100 bootstrapped datasets), PROTPARS and CONSENSE. Abbreviations not mentioned in the text are M1Pase (mannitol-1-phosphatase) and R5PPase (α-ribazole-5′-phosphate phosphatase).

350 times the size of the typical dPGM superfamily member substrate. How can the dPGM fold accommodate such radically differently sized substrates?

The alignment in Fig. 1 and the structural comparison in Fig. 3 show that SixA lacks the excursion from the main α/β domain, hitherto invariably present in dPGM members. Although substrate-binding sites differ between dPGMs, F26BPase and PhoE [11,37,38], it is notable that residues located within this excursion are invariably involved in substrate binding. The excursion, in general terms, functions to define the substrate-binding cleft on the side distal to the catalytic His. The absence of this excursion in SixA has dramatic consequences for the overall topography of the catalytic site (Fig. 3c,e). The small, deep cleft of dPGMs, for example, is replaced with a broad, shallow surface depression. The shallower site in SixA is clearly consistent with the location of the phosphorylable His of ArcB, which must bind at the SixA catalytic site, not in an extended loop, but within one of the four helices making up the phosphotransfer domain of the ArcB structure [39]. Indeed, preliminary docking studies suggest that the remoulded binding site of SixA can readily accommodate the four helical bundle of its substrate ArcB phosphotransfer domain (data not shown).

Understanding of the Ais family is much less complete than for the SixA family. The first determined sequence was of an *E. coli* aluminium-inducible protein [15]. Since then, the corresponding *Salmonella typhimurium* gene (named *pmrG*, only partially sequenced and therefore not included in our analysis) has been shown to form part of a putative operon [16], also present in *E. coli*. The regulation of PmrG by the PmrA–PmrB two-component system [16], along with the predicted functions of the other open reading frames of the operon, implicate it in the modification of bacterial lipopolysaccharide leading to resistance towards antimicrobial peptides. Later, a homologous sequence (coded by the *afrS* gene) was identified in the *E. coli* AF/R1 pilus operon [17] which is necessary for full *E. coli* RDEC-1 virulence. AfrS was implicated in the transcriptional activation of AfrA, the pilus structural subunit [17]. Both PmrG and AfrS therefore have medical significance, but their specific molecular roles are completely unknown. Sequence analysis (Fig. 1) and modelling (Fig. 3) show that the Ais group possesses an intact putative catalytic site with unimpeded substrate access. The Ais group is therefore likely to encode phosphatases. The periplasmic space, in which the Ais group of proteins is predicted to reside, contains a variety of phosphatase activities [40] and *E. coli* itself contains a periplasmic acid phosphatase [41] which is a distant relative of dPGMs and F26BPases. Intriguingly, the Ais proteins share with the SixA the lack of the excursion from the main α/β domain (Figs. 1 and 3c,e) leading to similarly shaped catalytic sites in both groups. This suggests that the unknown substrates of the Ais family members are not the small molecules more typically associated with the dPGM superfamily. Given the regulatory role of AfrS [17], one attractive possibility would be that, like SixA, these proteins act within signal transduction pathways as phosphoprotein phosphatases (though not necessarily phosphohistidyl protein phosphatases). Activity on another kind of large phosphorylated molecule cannot, however, be ruled out.

3.4. The modelled catalytic centres

Successive structural determinations of dPGMs, F26BPases and PhoE have revealed a catalytic core common to all and have enabled a good understanding of the catalytic mechanism. Although the dPGM catalyses a mutase reaction in contrast to the phosphatase activities of the others, the mutase reaction requires the initial phosphorylation of the catalytic His (the 'priming' of the enzyme) and the dPGMs can be inactivated by the slow, spontaneous hydrolysis of the 'primed' phospho-His enzyme. Phosphatase activity is therefore a side activity of dPGMs [6]. In all enzymes, the phosphogroup to be transferred to the enzyme binds in a complementary pocket largely composed of basic residues (His10, His151, Arg9 and Arg59 along with Asn14 and, in PhoE alone, Gln22). This numbering is for PhoE and will be used throughout. This pocket, here termed the phosphate pocket, is occupied variously by the phosphogroup of substrate, the phosphogroup after transfer to His10, or simply by phosphate (or isosteric sulphate) as seen in various crystal structures. Nucleophilic attack of His10 on the phosphocompound occurs with the involvement of conserved Glu83 which donates a proton to the departing product, having been induced into the protonated state by the proximity of the phosphogroup. The interactions of the phosphogroup with enzyme are similar before and after phosphotransfer [12], a factor which presumably enhances catalytic efficiency. The hydrolysis of the phosphorylated enzyme then proceeds by water-mediated nucleophilic attack of Glu83.

The clear evolutionary relationship of SixA and Ais with these enzymes would naturally lead to the expectation of conservation of all these catalytic core residues. Remarkably, sequence comparisons show this not to be the case. The phosphate pocket residues His10, His151, Arg9 and Arg59 are all conserved but neither Asn14, nor the PhoE-specific Gln22 are present in the SixA and Ais families. Most surprisingly, Glu83 is present in neither of these latter families. Since the acid–base catalysis function of Glu83 is an important part of the reaction scheme (mutagenesis of this residue results in dramatic loss of activity [42–44]), its absence in SixA and Ais presumably indicates that another residue acts as a functional substitute. A search for conserved acidic residues highlights Asp18 (*E. coli* SixA numbering) and Glu63 or Asp66 (*E. coli* Ais numbering) in these respective families. The molecular models show that Ais Glu63 is positioned on the protein surface, but that Asp18 and Asp66 are well positioned to effect acid–base catalysis in the SixA and Ais families, respectively (Fig. 3d,f). In each case, the Asp side chain is close to the phosphate pocket so that the latter's occupation would effect Asp protonation in a way analogous to that clearly inferred for Glu83 in PhoE [12].

In order to bring about the energetically unfavourable protonation of Asp18 or Asp66, the phosphogroup binding in the phosphate pocket must make multiple favourable interactions with the enzyme. The molecular models presented here show how ample compensation is offered for the lack of residues corresponding to PhoE Asn16 and Gln22. In each case, further conserved arginine residues are predicted to lie in the catalytic site, suitably positioned to make favourable interactions with the group occupying the phosphate pocket (Fig. 3d,f). In SixA this residue is Arg21, while in Ais it is Arg64. In fact, the charged head groups of these two Arg residues occupy approximately the same space when the models are superimposed. This same space is the location of the interaction of PhoE Gln22 with bound ligands [12] but is not used

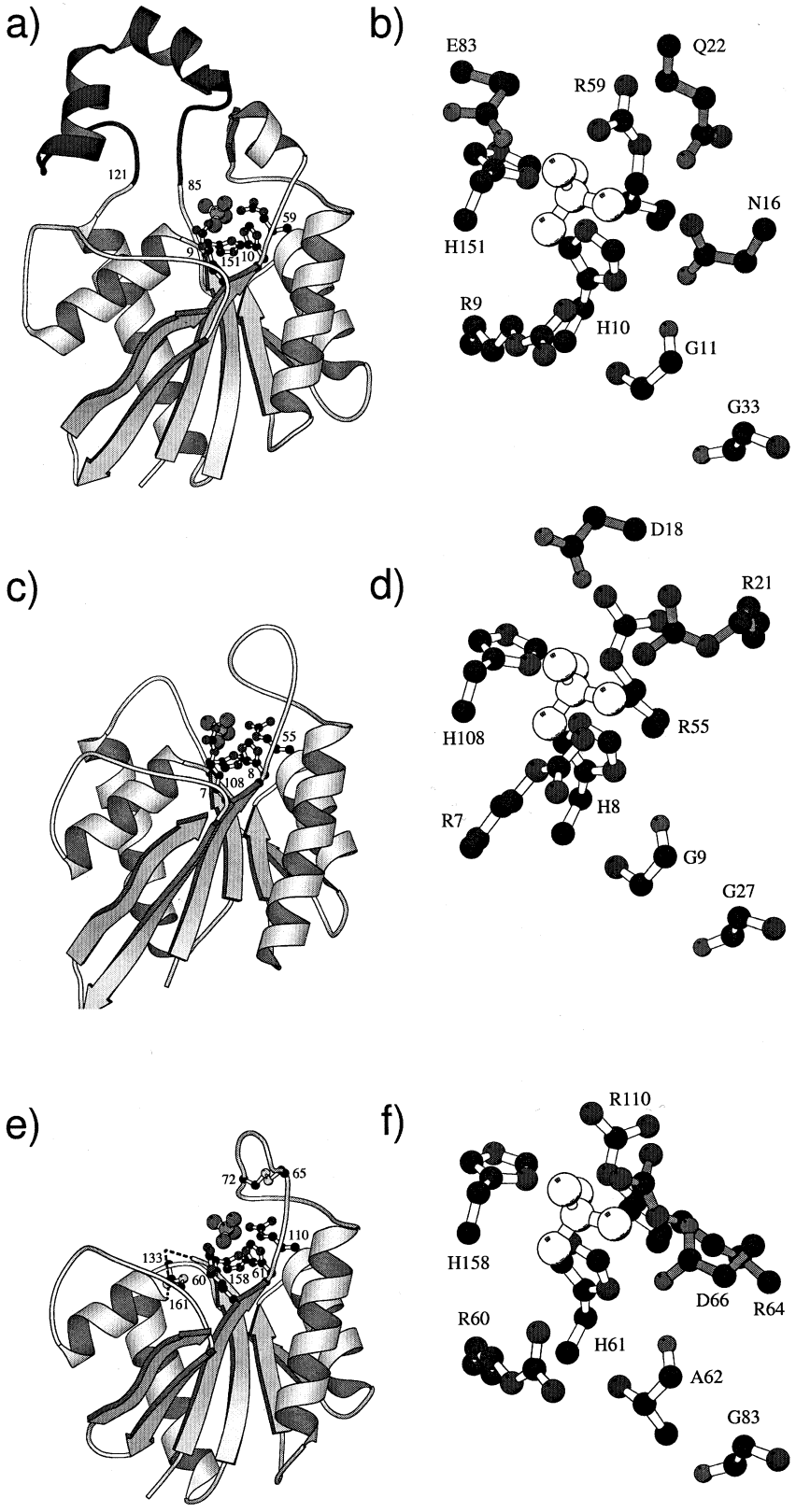

Fig. 3. MOLSCRIPT [46] diagrams of PhoE (panels a and b), the final *E. coli* SixA model (panels c and d) and the final *E. coli* Ais model (panels e and f). Overall molecular structures are shown on the left and close-ups of the catalytic sites on the right. In the overall figures the minimal catalytic four residues are shown along with a phosphate ion positioned in the phosphate pocket (see text). In panel e the disulphide bridges uniquely present in the Ais family are also shown, with part of a helix replaced with a dotted trace in order to reveal the Cys133–Cys161 bridge. In the detailed panels, the catalytic core is again shown, along with two other positions mentioned in the text (glycines 33 and 11 in PhoE). The acid–base catalyst and other phosphate pocket residues that differ between the three structures are shown with shaded bonds. A phosphate occupying the phosphate pocket is shown coloured white.

for phosphate interactions in either dPGMs or F26BPases due to functional constraints; corresponding regions are known or inferred to be involved in binding other parts of their respective substrates [12].

The surprising finding that the identity of the acid–base catalytic residue varies between SixA, Ais and dPGMs leads to a reduced definition of the catalytic core common to all. As examination of Fig. 1 shows, just five residues are identical among the three templates, all SixA family members and the Ais family. Four comprise the minimal catalytic core of two histidines and two arginines, residues Arg9, His10, Arg59 and His151 in the case of PhoE. The fifth conserved residue is surprisingly distant to the other four (Fig. 3b,d,f). Structural examination shows that Gly33 must be conserved since any side chain at that position would clash with Gly11, whose main chain carbonyl group in turn hydrogen bonds to the His10 side chain, defining the latter's orientation. Position 11 is nearly invariably occupied by a glycine, but is occasionally an alanine. Interestingly, the orientation of the His151 side chain is also maintained by hydrogen bonding, this time to the side chain of residue 55, invariably a serine or threonine.

As mentioned, structural determinations of acid phosphatases and phytases revealed their previously unexpected distant evolutionary relationship with the dPGMs and F26BPases [9,10]. It is interesting to note that the five conserved residues mentioned above – the catalytic core plus Gly33 – are all present in these distant relatives. The SixA and Ais families are clearly more closely related to the dPGM branch rather than the acid phosphatase branch since the members of the former, but not members of the latter, are present in PSI-BLAST results. It is therefore a measure of the structural divergence of SixA and Ais that they share the same number of invariant residues with dPGMs, F26BPases and PhoE as the latter group share with acid phosphatases and phytases.

In conclusion, two families of dPGM homologues were subjected to molecular modelling and sequence analysis. Although bearing only a distant evolutionary relationship, each family contains a dPGM fold lacking a domain excursion, previously assumed ubiquitous. In the case of the SixA proteins, with experimentally verified phosphoprotein phosphatase activity [13], the structural difference is clearly correlated with substrate specificity. The similar overall structure of the Ais family catalytic site suggests that they too may be phosphoprotein phosphatases. The positioning of a key acid–base catalytic residue near that domain excursion in better-understood members of the dPGM superfamily has led to unexpected reorganisation of the catalytic site in both SixA and Ais families. Each family has its own, different presumed acid–base catalytic aspartate residue, along with conserved arginines which likely contribute to the phosphate pockets. The remarkable diversity of the dPGM superfamily and, in particular, the apparent lack of a phosphorylable histidine in some sequences strongly hint at further structural and functional novelty still to be explored.

References

[1] Jedrzejas, M.J. (2000) Prog. Biophys. Mol. Biol. 73, 263–287.
[2] Okar, D.A., Manzano, A., Navarro-Sabate, A., Riera, L., Bartrons, R. and Lange, A.J. (2001) Trends Biochem. Sci. 26, 30–35.
[3] O'Toole, G.A., Trzebiatowski, J.R. and Escalante-Semerena, J.C. (1994) J. Biol. Chem. 269, 26503–26511.
[4] Liberator, P., Anderson, J., Feiglin, M., Sardana, M., Griffin, P., Schmatz, D. and Myers, R.W. (1998) J. Biol. Chem. 273, 4237–4244.
[5] Rigden, D.J., Bagyan, I., Lamani, E., Setlow, P. and Jedrzejas, M.J. (2001) Protein Sci. 10, 1835–1846.
[6] Fothergill-Gilmore, L.A. and Watson, H.C. (1989) Adv. Enzymol. Relat. Areas Mol. Biol. 62, 227–313.
[7] Bazan, J.F., Fletterick, R.J. and Pilkis, S.J. (1989) Proc. Natl. Acad. Sci. USA 86, 9642–9646.
[8] Lee, Y.H., Ogata, C., Pflugrath, J.W., Levitt, D.G., Sarma, R., Banaszak, L.J. and Pilkis, S.J. (1996) Biochemistry 35, 6010–6019.
[9] Kostrewa, D., Gruninger-Leitch, F., D'Arcy, A., Broger, C., Mitchell, D. and van Loon, A.P. (1997) Nat. Struct. Biol. 4, 185–190.
[10] Schneider, G., Lindqvist, Y. and Vihko, P. (1993) EMBO J. 12, 2609–2615.
[11] Rigden, D.J., Mello, L.V., Setlow, P. and Jedrzejas, M.J. (2002) J. Mol. Biol. 315, 1129–1143.
[12] Rigden, D.J., Littlejohn, J.E., Henderson, K. and Jedrzejas, M.J. (2002) J. Mol. Biol. 325, 411–420.
[13] Ogino, T., Matsubara, M., Kato, N., Nakamura, Y. and Mizuno, T. (1998) Mol. Microbiol. 27, 573–585.
[14] Matsubara, M. and Mizuno, T. (2000) FEBS Lett. 470, 118–124.
[15] Guzzo, A. and DuBow, M.S. (1994) FEMS Microbiol. Rev. 14, 369–374.
[16] Gunn, J.S., Lim, K.B., Krueger, J., Kim, K., Guo, L., Hackett, M. and Miller, S.I. (1998) Mol. Microbiol. 27, 1171–1182.
[17] Cantey, J.R., Blake, R.K., Williford, J.R. and Moseley, S.L. (1999) Infect. Immun. 67, 2292–2298.
[18] Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) J. Mol. Biol. 215, 403–410.
[19] Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[20] Notredame, C., Higgins, D.G. and Heringa, J. (2000) J. Mol. Biol. 302, 205–217.
[21] Gouet, P., Courcelle, E., Stuart, D.I. and Metoz, F. (1999) Bioinformatics 15, 305–308.
[22] Nakai, K. and Kanehisa, M. (1991) Proteins 11, 95–110.
[23] Emanuelsson, O., Nielsen, H., Brunak, S. and von Heijne, G. (2000) J. Mol. Biol. 300, 1005–1016.
[24] Nielsen, H., Engelbrecht, J., Brunak, S. and von Heijne, G. (1997) Protein Eng. 10, 1–6.
[25] Bujnicki, J.M., Elofsson, A., Fischer, D. and Rychlewski, L. (2001) Bioinformatics 17, 750–751.
[26] Jones, D.T. (1999) J. Mol. Biol. 292, 195–202.
[27] Felsenstein, J. (1989) Cladistics 5, 164–166.
[28] Sali, A. and Blundell, T.L. (1993) J. Mol. Biol. 234, 779–815.
[29] Sippl, M.J. (1993) Proteins 17, 335–362.
[30] Laskowski, R., Macarthur, M., Moss, D. and Thornton, J. (1993) J. Appl. Crystallogr. 26, 283–290.
[31] Jones, T.A., Zou, J.Y. and Cowan, S.W. (1991) Acta Crystallogr. A47, 110–119.
[32] Kleywegt, G.J., Zou, J.Y., Kjeldgaard, M. and Jones, T.A. (2001) in: International Tables for Crystallography, Vol. F. Crystallography of Biological Macromolecules (Rossmann, M.G. and Arnold, E., Eds.), chapter 17.1, pp. 353–356, 366–367, Kluwer Academic Publishers, Dordrecht.
[33] Frishman, D. and Argos, P. (1995) Proteins 23, 566–579.
[34] Chen, C.P. and Rost, B. (2002) Appl. Bioinform. 1, 21–35.
[35] Bujnicki, J.M., Elofsson, A., Fischer, D. and Rychlewski, L. (2001) Proteins 45, Suppl. 5, 184–191.
[36] Hasemann, C.A., Istvan, E.S., Uyeda, K. and Deisenhofer, J. (1996) Structure 4, 1017–1029.
[37] Bond, C.S., White, M.F. and Hunter, W.N. (2002) J. Mol. Biol. 316, 1071–1081.
[38] Yuen, M.H., Mizuguchi, H., Lee, Y.H., Cook, P.F., Uyeda, K. and Hasemann, C.A. (1999) J. Biol. Chem. 274, 2176–2184.
[39] Kato, M., Mizuno, T., Shimizu, T. and Hakoshima, T. (1997) Cell 88, 717–723.

[40] Rossolini, G.M., Schippa, S., Riccio, M.L., Berlutti, F., Macaskie, L.E. and Thaller, M.C. (1998) Cell. Mol. Life Sci. 54, 833–850.
[41] Ostanin, K., Harms, E.H., Stevis, P.E., Kuciel, R., Zhou, M.M. and Van Etten, R.L. (1992) J. Biol. Chem. 267, 22830–22836.
[42] Lin, K., Li, L., Correia, J.J. and Pilkis, S.J. (1992) J. Biol. Chem. 267, 6556–6562.
[43] Nairn, J., Duncan, D., Price, N.E., Kelly, S.M., Fothergill-Gilmore, L.A., Uhrinova, S., Barlow, P.N., Rigden, D.J. and Price, N.C. (2000) Eur. J. Biochem. 267, 7065–7074.
[44] Rigden, D.J., Walter, R.A., Phillips, S.E.V. and Fothergill-Gilmore, L.A. (1999) J. Mol. Biol. 286, 1507–1517.
[45] Barton, G.J. (1993) Protein Eng. 6, 37–40.
[46] Kraulis, J. (1991) J. Appl. Crystallogr. 24, 946–950.