Evolutionary relationship between the bacterial HPr kinase and the ubiquitous PEP-carboxykinase: expanding the P-loop nucleotidyl transferase superfamily

Robert B. Russell[a,*], José Antonio Márquez[a], Wolfgang Hengstenberg[b], Klaus Scheffzek[a]

[a] *Structural and Computational Biology Programme, EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany*
[b] *Department of Biology, Ruhr-Universität Bochum, Universitätstrasse 150, 44780 Bochum, Germany*

Received 18 January 2002; revised 21 February 2002; accepted 21 February 2002

First published online 21 March 2002

Edited by Gunnar von Heijne

Abstract Similarities between protein three-dimensional structures can reveal evolutionary and functional relationships not apparent from sequence comparison alone. Here we report such a similarity between the metabolic enzymes histidine phosphocarrier protein kinase (HPrK) and phosphoenolpyruvate carboxykinase (PCK), suggesting that they are evolutionarily related. Current structure classifications place PCK and other P-loop containing nucleotidyl-transferases into different folds. Our comparison of both HPrK and PCK to other P-loop containing proteins reveals that all share a common structural motif consisting of an αβ segment containing the P-loop flanked by an additional β-strand that is adjacent in space, but far apart along the sequence. Analysis also shows that HPrK/PCK differ from other P-loop containing structures no more than they differ from each other. We thus suggest that HPrK and PCK should be classified with other P-loop containing proteins, and that all probably share a common ancestor that probably contained a simple P-loop motif with different protein segments being added or lost over the course of evolution. We used the structure-based sequence alignment containing residues specific to HPrK/PCK to identify additional members of this P-loop containing family.

Key words: Protein structure similarity; Histidine phosphocarrier protein kinase; Phosphoenolpyruvate carboxykinase; P-loop; Evolution

1. Introduction

Proteins can adopt similar protein three-dimensional (3D) structures in the absence of significant sequence similarity. Such similarities can imply an evolutionary relationship and often similarities in function (e.g. [1]). Here we report such a similarity between the structures of phosphoenolpyruvate carboxykinase (PCK; [2]) and histidine phosphocarrier protein kinase (HPrK; [3]) and discuss the associated evolutionary and functional implications.

Both PCK and HPrK play important roles in metabolic regulation by catalysing the ATP-dependent transfer of a phosphoryl group to the substrates oxaloacetate and HPr, respectively. PCK transfers phosphate from ATP or GTP to a compound at the crossroads of the tricarboxylic acid cycle and glycolysis. It is ubiquitous in bacteria and eukaryotes where it plays a crucial role in gluconeogenesis. In contrast, HPrK has thus far only been found in bacteria where its major role appears to be the regulation of carbon source uptake in the absence/presence of glucose. Although both enzymes are similar with respect to the type of reaction they catalyse, they differ markedly in their substrates. Whereas in PCK the phosphoryl acceptor is a small metabolite, the HPrK substrate is HPr, a 9 kDa protein, which is phosphorylated at Ser46.

After its identification, HPrK was not found to be related to any known eukaryotic protein kinases [4]. The recent elucidation of 3D structures for HPrK provided the first hints as to similarities with other proteins [3,5], which we discuss in detail here.

2. Materials and methods

We compared the HPrK structure [5] to the database of known 3D structures search using the DALI server (http://www2.embl-ebi.ac.uk/dali/ [1]), and aligned 3D structures with the STAMP package [6]. Structures were taken from the protein databank (PDB; www.rcsb.org). Alignments of homologues of HPrK and PCK were taken from the Pfam database (http://www.sanger.ac.uk/Software/Pfam/ [7] families HPr_kinase and PEPCK_ATP respectively) and merged with the structure alignment.

The region around the P-loop was used to construct profiles for HMMer (S. Eddy, unpublished) and PSI-blast [8] using default parameters, and to derive a motif (regular expression) described in Section 3. These profiles/motifs were then used to search the NCBI non-redundant protein database (www.ncbi.nlm.nih.gov). Results of these searches were compared to results obtained when searching with profiles derived separately for the PCK and HPrK families and only those sequences not found with significant scores by the separate searches are given. Only those sequences with expectation (*E*)-value thresholds of ≥ 0.1, or those that matched the motif are shown, and only the region around the P-loop is shown in Fig. 2. In the case of PSI-blast, the *E*-value reported is that from the final iteration. Alignments with PCK/HPrK homologues were constructed initially with HMMer and edited manually.

3. Results and discussion

3.1. HPrK/PCK similarity

The similarity between HPrK and PCK was initially detected by a DALI [1] search with a *Z*-score of 7.8 [5] suggesting a homologous relationship [9,10]. The similarity covers the

*Corresponding author.
E-mail address: russell@embl-heidelberg.de (R.B. Russell).

PII: S0014-5793(02)02518-8

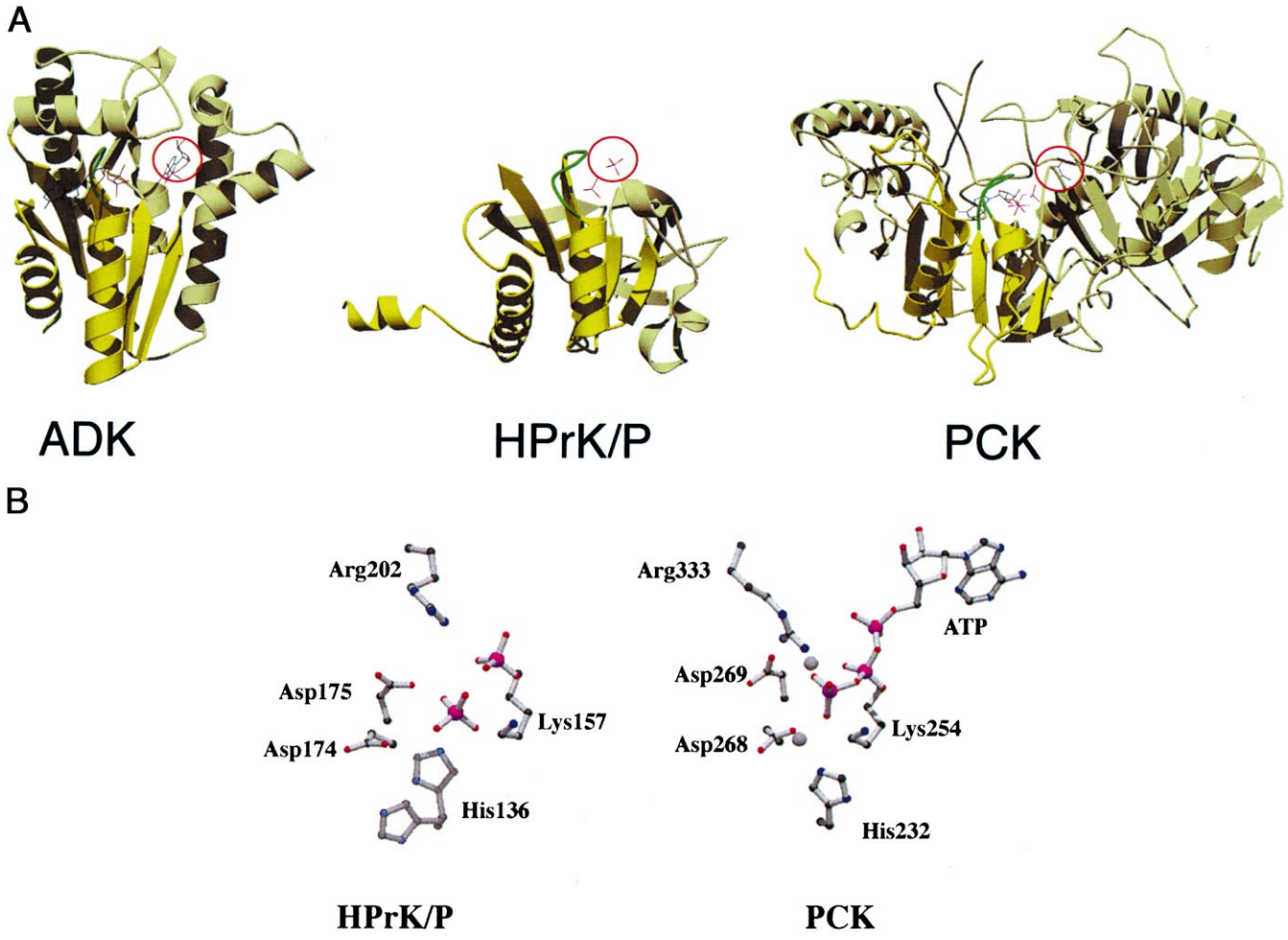

Fig. 1. A: Ribbon diagrams of *E. coli* adenylate kinase (ADK) bound to ATP and AMP (left), *Staphylococcus xylosus* HPrK bound to two phosphate molecules (centre) and *E. coli* PCK bound to ATP and pyruvate (right). Ligands are represented as stick models. Equivalent secondary structure elements are depicted in yellow. Non-conserved elements are coloured in pale green. ATP and pyruvate (red circles) are bound to the respective enzymes in deeply buried pockets. In contrast, the second phosphate in HPrK (red circle) that represents Ser46 of the HPr protein docking site is highly accessible from the solvent. These pictures were produced by Molscript [25] and Raster3D [26]. B: Detail of the active sites of *S. xylosus* HPrK (left; PDB code 1ko7) and *E. coli* PCK (right; 1ayl) showing structurally equivalent residues.

C-terminal domains of both enzymes (PCK, PDB code 1ayl, residues 227–469; HPrK, residues 135–286) and includes a core of 9 β-strands along with two α-helices (Fig. 1A). A total of 89 Cα atoms can be superimposed with a root-mean-square deviation of 1.6 Å (relaxing the requirement for similarity gives 128 atoms with an root-mean-square of 2.8 Å). The common structural core includes the P-loop (or Walker-A motif; [11]) and much of the active site from both enzymes. The sequence identity observed in the structure-based sequence alignment is 19% (17 identities over 89 residues), which is indicative of a homologous relationship [12]. Moreover, the observed number of identities is associated with a statistical significance *P*-value of 10^{-20} [13,14], suggesting that the degree of sequence identity is unlikely to be due to a coincidental similarity between folds. Fig. 2 shows a structure-based alignment [6] augmented with various homologues of the N-terminal regions of PCK and HPrK.

Several conserved features make HPrK and PCK unique among P-loop containing enzymes, in particular, the N-terminal histidine (HPrK, His136; PCK, His232) and two conserved aspartates (HPrK, Asp174/175; PCK, Asp268/269). Comparison of crystal structures from *Escherichia coli* PCK in complex with substrates and transition state analogues has given insights into the catalytic mechanism of the phosphotransfer reaction [15–17]. PCK contains a Mg^{2+}–Mn^{2+} bimetal cluster at the active site that is stablised by Asp268 and Asp269, which are part of the PCK kinase 2 motif (Fig. 1B) [15,16]. His232 and the highly conserved Arg333 contact the γ-phosphate of ATP from opposite sides and are proposed to play a role in either substrate activation or the stabilisation of the transition state [15,18]. Corresponding residues in HPrK, Asp174, Asp175 and His136, are located in equivalent positions in the alignment and are likely to play similar roles in catalysis [5]. Remarkably, the side chain of the invariant residue Arg202 in HPrK is structurally equivalent to the conserved Arg333 from PCK although the residues come from topologically different regions of PCK and HPrK, suggesting that this feature has evolved convergently.

3.2. Comparison of HPrK and PCK to other P-loop proteins

The structure database search also showed less significant

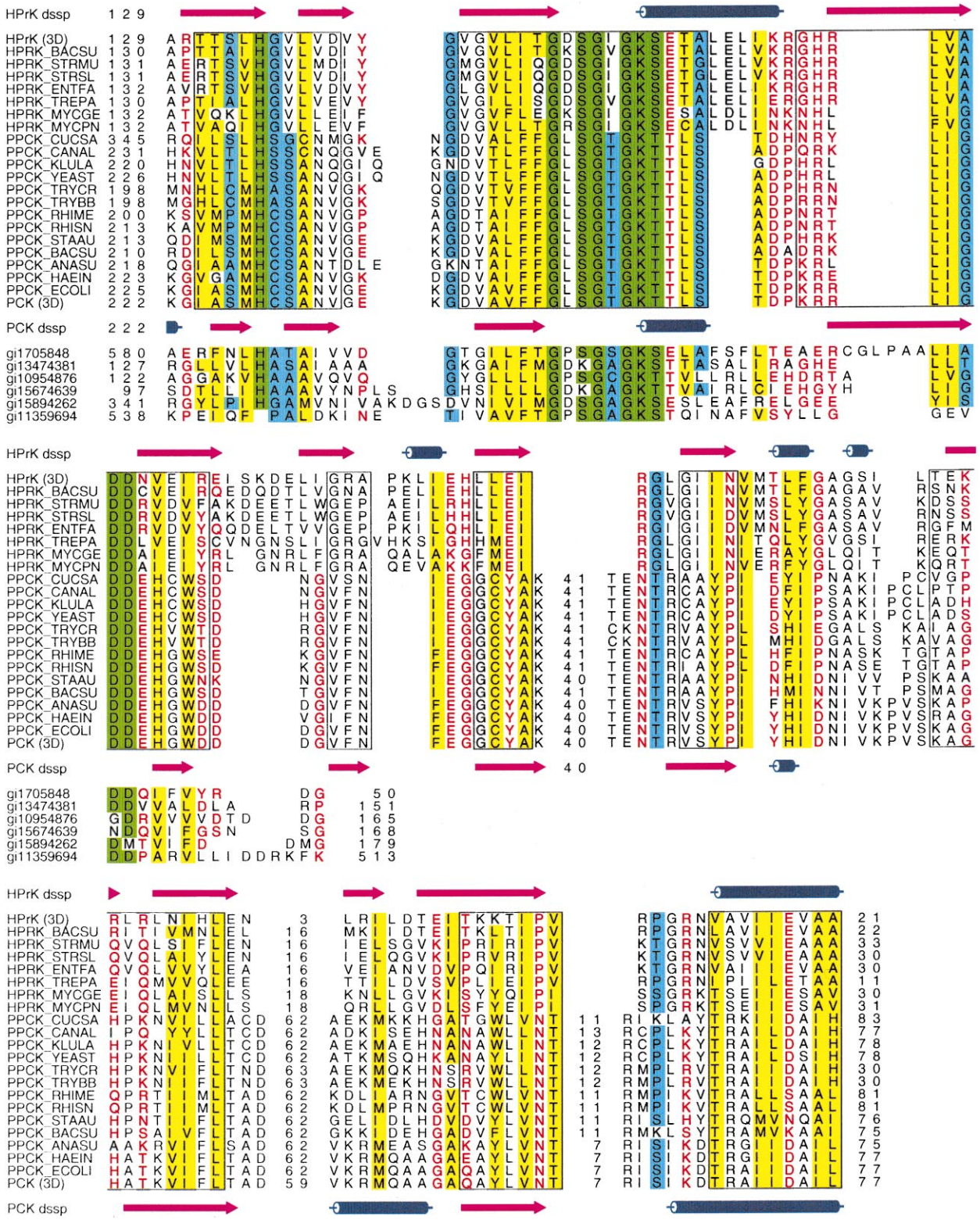

Fig. 2. Alscript [27] figure showing an alignment of HPrK and PCK sequences based on 3D structure comparison. Sequences are given by Swiss-Prot identifiers, apart from HPrK (3D), and PCK (3D), which are from the known structures (PDB codes 1ko7 and 1ayl). Secondary structures, as defined by DSSP [28], and are shown above and below the alignment for HPrK and PCK: α- or 3_{10}-helices, blue cylinders; β-strands, magenta arrows. Regions deemed structurally equivalent by STAMP [6] are boxed. Residues are coloured according to conservation of properties (i.e. only conserved positions highlighted): yellow background, hydrophobic; blue background, small; red, polar, apart from those regions important to HPrK or PCK function (including the P-loop), which are shown by a green background. Note that glycine and proline are considered neutral and can thus occur in different conservation contexts (e.g. small and polar). Numbers within the aligned sequences give the lengths of long insertions with no sequence or structure similarity between HPrKs and PCKs that have been deleted for clarity.

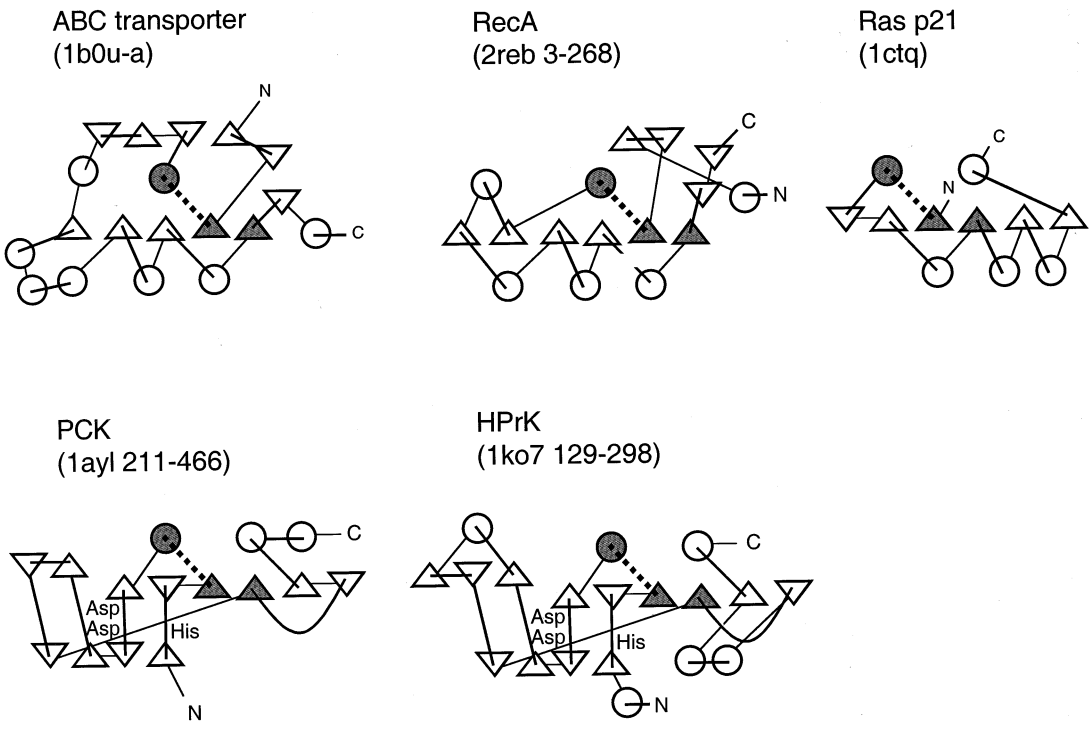

Fig. 3. Topology diagrams showing the structural diversity within P-loop containing proteins. β-Strands are represented by triangles and α-helices by circles, and connecting lines are shown in front or behind the circles/triangles to indicate the relative connectivity of the loops. Shaded elements correspond to the structurally conserved core found in all P-loop containing proteins. Boxes denote groupings based on key topological differences as discussed in the text. Structures are denoted by their names, and by their PDB codes. The approximate location of PCK/HPrK specific Asp and His residues is also shown.

similarities between HPrK and a number of other P-loop containing proteins. Indeed, a similarity between PCK and other P-loop containing proteins has long been known [2]. However, a number of important topological differences between the PCK structure and those of other P-loop containing enzymes mean that they are currently classified as the different folds (e.g. in SCOP [19] or CATH [20]).

We compared HPrK and PCK to representatives from other P-loop containing protein families in SCOP, and found that all contain a structurally conserved core with many long insertions that can differ substantially (Fig. 3). This core comprises an αβ segment containing the P-loop and an additional β-strand that is adjacent in space, but often far apart along the sequence. These structures are embedded in a β-sheet that can differ greatly in topology and that is normally flanked by additional α-helices or β-sheets. This conserved core can apparently tolerate long insertions or additions of many secondary structures, and in some cases entirely different domains. Structure comparison shows that HPrK and PCK deviate no more from the conserved P-loop core than do other members of the superfamily. We suggest that all of the known P-loop containing proteins probably descended from a common ancestor and should be classified in the same fold and superfamily. The P-loop containing superfamily is clearly one that has been subject to drastic fold changes over the course of evolution [21].

Table 1
Details of novel HPrK/PCK homologues identified

Accessions	Length	Region	Species	HMM	PSI	Mot	Comments
gi1705848 (CHVG_AGRTU)	690	581–637	*A. tumefaciens*	7.8e–06	6.0e–11	Y	Transmembrane helix: 34–53; HAMP domain: 267–338; HisKA domain: 342–407; HATPase_c domain: 459–572; Weak match to HPrK
gi13474381 (NP_105949.1)	333	6–62	*M. loti*	5.7e–09	0.037	Y	Weak match to HPrK
gi10954876 (NP_053296.1)	344	123–174	*A. tumefaciens*	1.6e–05	0.004	Y	Weak match to HPrK
gi15674639 (NP_268813.1)	322	97–150	*Streptococcus pyogenes*	0.056	0.08	N	–
gi15894262 (NP_347611.1)	579	342–397	*Clostridium acetobutylicum*	7.9	0.27	Y	Weak match to PCK (previously identified)
gi11359694 (T51227)	1270	714–754	*N. crassa*	–	0.081	N	FN3 domain: 517–617; Weak match to AAA domain

Accessions are from the NCBI non-redundant database (www.ncbi.nlm.nih.gov), with additional accessions given in parentheses. Values reported for HMM and PSI are the *E*-values. 'Y' or 'N' in the Mot (for motif) column give a yes/no as to whether the sequence contains the motif discussed in the text. Domains (and abbreviations) given under comments were found using the SMART database [24].

3.3. Additional members of the HPrK/PCK family

We used features specific to HPrK/PCK to search for additional homologues of these two enzymes. We included the region between the conserved histidine and the pair of aspartates discussed above (i.e. including the P-loop). The region around the conserved histidine lies within a β-strand, which is anti-parallel to another that is in the conserved core (see Fig. 3), and is not seen in any of the other P-loop containing structures. The loop containing the aspartates often has a topological equivalent in other P-loop containing proteins, though inspection of known structures shows that none of the others contain two aspartates in an equivalent location. We searched the protein sequence database using HMMer (S. Eddy, unpublished), PSI-blast [8] and a simple motif search (i.e. with the motif [LVIMF]Hx{9,21}GxxGK[TS]x{7,27}DD, where brackets denote alternative amino acids, 'x' denotes any amino acid and regions in braces the range of variable loops). We considered only the N-terminal portion of the alignment because methods of protein sequence database searching (such as PSI-blast or HMMer) have difficulties dealing with long insertions like those seen when comparing PCK to HPrK (Fig. 2).

Table 1 shows those sequences that match these profiles and/or the motif and do not reside in a different domain, and which could not be identified by PSI-blast or HMMer searches using HPrK or PCK homologues alone. These sequences are aligned to the HPrK and PCK homologues in Fig. 2. Both PSI-blast and HMMer searches detected weak similarities between HPrK/PCK and other proteins containing P-loops, including sugar transporters, and exonucleases (results not shown). While not significant, this hints at an ancient evolutionary relationship among these proteins as we suggest above.

With the exception of one protein from *Neurospora crassa*, all proteins reported in Table 1 are from bacteria. All sequences are singletons, by which we imply that they do not have any obvious sequence homologues when searching the database (e.g. with blast). However, in some cases marginal matches suggest a closer kinship to either HPrK or PCK (see 'Weak match' in Table 1). All have total sequence lengths that differ substantially from known HPrK or PCK sequences. This suggests that they may have functions that are distinct from either of these enzymes. Probably they are kinases/phosphatases, but the substrate may differ, as is often the case for proteins that reside in the same family but have limited overall sequence similarity (i.e. when only catalytic residues are conserved).

In *E. coli* PCK, ATP and oxalacetate bind at a deep crevice formed between the N- and C-terminal domains [18]. The buried character of substrate binding sites is very common in enzymes and has been proposed to favour catalysis by trapping the substrates and excluding water molecules from the active site that may produce spurious hydrolytic reactions if accessible from the solvent environment. In nucleoside monophosphate kinases such as adenylate kinase (a member of the P-loop containing superfamily) binding of substrates is associated with domain closure upon the bound substrates. In particular, the so-called lid domain covers the bound ATP in a hinge movement that shields the substrate binding site [22,23] (Fig. 1A). In contrast, the environment of the catalytic site in HPrK is open towards the solvent leaving the space required to accommodate HPr or P-Ser46-HPr. HPrK also has an N-terminal domain, however, it forms a separate entity, leaving the catalytic site in the C-terminal domain completely exposed to solvent [5]. This feature may be of help in predicting functions in the PCK/HPrK protein family. A lack of sequences capable of forming a potential lid on the active site could indicate that an enzyme acts on a protein rather than a small-molecule substrate. For example, the HPrK/PCK-like sequences in *Agrobacterium tumefaciens* and *Mesorhizobium loti* lack a PCK- (or HPrK-) like N-terminal domain, and also appear to lack any long insertions within the kinase domain that could act as a lid (see Table 1)

A key goal of structural genomics projects is first to use 3D structure comparison to identify likely evolutionary and/or functional similarities that are not apparent when comparing sequences alone. An associated goal is to then use the similarity between structures to identify additional protein sequences that could not be found when using sequences or structures separately. The HPrK/PCK similarity provides an excellent illustration of how these goals can be reached, and how determination of a new structure can provide insights into the functional and evolutionary relationships within a diverse protein family.

Acknowledgements: J.A.M gratefully acknowledges support from the Fundación Ramón Areces Foundation (Madrid). We thank Domenico Bordo (IST-CBA, Genoa) for a critical reading of the manuscript.

References

[1] Holm, L. and Sander, C. (1996) Science 273, 595–603.
[2] Matte, A., Goldie, H., Sweet, R.M. and Delbaere, L.T. (1996) J. Mol. Biol. 256, 126–143.
[3] Fieulaine, S. et al. (2001) EMBO J. 20, 3917–3927.
[4] Galinier, A., Kravanja, M., Engelmann, R., Hengstenberg, W., Kilhoffer, M.C., Deutscher, J. and Haiech, J. (1998) Proc. Natl. Acad. Sci. USA 95, 1823–1828.
[5] Marquez, J.A., Hasenbein, S., Koch, B., Fieulaine, S., Nessler, S., Russell, R.B., Hengstenberg, W. and Scheffzek, K. (2002) Proc. Natl. Acad. Sci. USA 99, 3458–3463.
[6] Russell, R.B. and Barton, G.J. (1992) Proteins 14, 309–323.
[7] Bateman, A., Birney, E., Durbin, R., Eddy, S.R., Howe, K.L. and Sonnhammer, E.L. (2000) Nucleic Acids Res. 28, 263–266.
[8] Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Nucleic Acids Res. 25, 3389–3402.
[9] Holm, L. and Sander, C. (1997) Proc. Int. Conf. Intell. Syst. Mol. Biol. 5, 140–146.
[10] Dietmann, S. and Holm, L. (2001) Nat. Struct. Biol. 8, 953–957.
[11] Saraste, M., Sibbald, P.R. and Wittinghofer, A. (1990) Trends Biochem. Sci. 15, 430–434.
[12] Russell, R.B., Saqi, M.A., Sayle, R.A., Bates, P.A. and Sternberg, M.J. (1997) J. Mol. Biol. 269, 423–439.
[13] Murzin, A.G. (1993) J. Mol. Biol. 230, 689–694.
[14] Aloy, P., Oliva, B., Querol, E., Aviles, F.X. and Russell, R.B. (2002) Protein Sci., in press.
[15] Tari, L.W., Matte, A., Goldie, H. and Delbaere, L.T. (1997) Nat. Struct. Biol. 4, 990–994.
[16] Matte, A., Tari, L.W., Goldie, H. and Delbaere, L.T. (1997) J. Biol. Chem. 272, 8105–8108.
[17] Sudom, A.M., Prasad, L., Goldie, H. and Delbaere, L.T. (2001) J. Mol. Biol. 314, 83–92.
[18] Tari, L.W., Matte, A., Pugazhenthi, U., Goldie, H. and Delbaere, L.T. (1996) Nat. Struct. Biol. 3, 355–363.
[19] Lo Conte, L., Ailey, B., Hubbard, T.J., Brenner, S.E., Murzin, A.G. and Chothia, C. (2000) Nucleic Acids Res. 28, 257–259.
[20] Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B. and Thornton, J.M. (1997) Structure 5, 1093–1108.
[21] Grishin, N.V. (2001) J. Struct. Biol. 134, 167–185.

[22] Gerstein, M., Schulz, G. and Chothia, C. (1993) J. Mol. Biol. 229, 494–501.
[23] Schlauderer, G.J., Proba, K. and Schulz, G.E. (1996) J. Mol. Biol. 256, 223–227.
[24] Schultz, J., Copley, R.R., Doerks, T., Ponting, C.P. and Bork, P. (2000) Nucleic Acids Res. 28, 231–234.
[25] Kraulis, P.J. (1991) J. Appl. Cryst. 24, 946–950.
[26] Merritt, E.A. and Bacon, D.J. (1997) Methods Enzymol. 277, 505–524.
[27] Barton, G.J. (1993) Protein Eng. 6, 37–40.
[28] Kabsch, W. and Sander, C. (1983) Biopolymers 22, 2577–2637.