

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/227806256>

# Two “unrelated” families of ATP-dependent enzymes share extensive structural similarities about their cofactor binding sites

ARTICLE *in* PROTEIN SCIENCE · MAY 1998

Impact Factor: 2.85 · DOI: 10.1002/pro.5560070507

---

CITATIONS

24

---

READS

18

## 4 AUTHORS, INCLUDING:



Konstantin Denessiouk

Åbo Akademi University

25 PUBLICATIONS 664 CITATIONS

[SEE PROFILE](#)



Mark S Johnson

Åbo Akademi University

152 PUBLICATIONS 7,514 CITATIONS

[SEE PROFILE](#)

## Two “unrelated” families of ATP-dependent enzymes share extensive structural similarities about their cofactor binding sites

KONSTANTIN A. DENESSIOUK,<sup>1,3,5</sup> JUKKA V. LEHTONEN,<sup>2,4</sup> TIMO KORPELA,<sup>1</sup>  
AND MARK S. JOHNSON<sup>2,4</sup>

<sup>1</sup>Joint Biotechnology Laboratory, Department of Biochemistry, University of Turku, BioCity 6A, FIN-20520 Turku, Finland

<sup>2</sup>Department of Biochemistry and Pharmacy, Åbo Akademi University, BioCity 3A, P.O. Box 66, FIN-20521 Turku, Finland

<sup>3</sup>Moscow Institute of Physics and Technology, 141700, Dolgoprudny, Russia

<sup>4</sup>Centre for Biotechnology, University of Turku and Åbo Akademi University, BioCity 5B, FIN-20521 Turku, Finland

(RECEIVED December 3, 1997; ACCEPTED January 22, 1998)

### Abstract

Two proteins, D-alanine:D-alanine ligase and cAMP-dependent protein kinase, share a remarkable degree of structural convergence despite having different three-dimensional folds and different enzymatic functions. Here we report that as many as 103 residues from 10 segments form two identical super-secondary structures between which the cofactor ATP is bound. The cofactor, two bound metal cations, and several water molecules form a large network of electrostatic and hydrophobic interactions common to both enzymes, and these are mediated by the similar placement of equivalent amino acids within the common supersecondary structures.

**Keywords:** ATP-dependent enzymes; convergent structural similarities; identical supersecondary structures; similar cofactor binding sites; structure alignment

Recently, it was reported that two families of ATP-binding proteins share a similar spatial organization of several residues comprising a “structural motif” surrounding the adenine moiety (Kobayashi & Go, 1997a, 1997b), yet they have different folds and are thus unlikely to be related evolutionarily. One family includes D-Ala:D-Ala ligase (DD-ligase) (Fan et al., 1994, 1997), glutathione synthetase (Yamaguchi et al., 1993; Matsuda et al., 1996), succinyl-CoA synthetase (Wolodko et al., 1994), the biotin carboxylase subunit of acetyl-CoA carboxylase (Waldrop et al., 1994), pyruvate phosphate dikinase (Herzberg et al., 1996), and carbamoyl phosphate synthetase (Thoden et al., 1997). All of these enzymes catalyze similar reactions, and the structural elements within the ATP-binding site pack together in an equivalent way to form the common family fold termed the “glutathione synthetase fold” (Matsuda et al., 1996). The fold consists of three structural domains, each having one  $\beta$ -sheet of at least four strands; and one molecule of the cofactor is bound to a

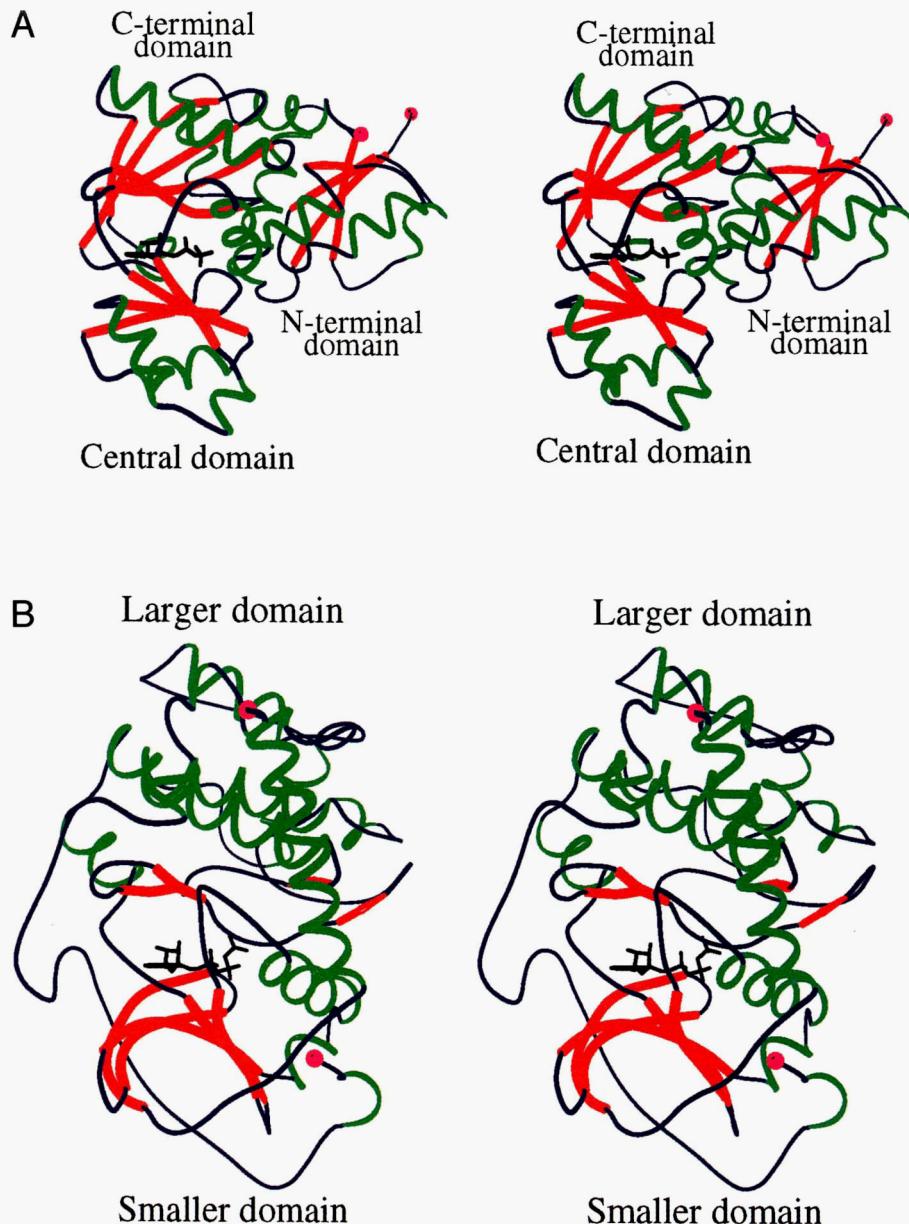
cleft located between the two antiparallel  $\beta$ -sheets, termed the “ATP-grasp” (Murzin, 1996), and formed from the central and the carboxy-terminal domains (Fig. 1A). The other family consists of a large number of related protein kinases (Hanks et al., 1988; Bossemeyer et al., 1993; Hubbard et al., 1994; Xu et al., 1995). The most pronounced similarity between these two “unrelated” families was found between DD-ligase and cAMP-dependent kinase (cAPK) (Kobayashi & Go, 1997a, 1997b). The fold of cAPK (Fig. 1B) consists of two lobes, one dominated by anti-parallel  $\beta$ -strands (smaller domain) and the other by  $\alpha$ -helices (larger domain), and the ATP binding site is located in a cleft between these two domains (Bossemeyer et al., 1993).

In binding to either of these enzymes, the cofactor forms a network of electrostatic interactions and hydrogen bonds with many residues within its binding site (Bossemeyer et al., 1993; Fan et al., 1994, 1997). The local similarity reported by Kobayashi and Go (1997a, 1997b) for DD-ligase and cAPK was limited to a four-residue segment (two of these residues form two hydrogen bonds with the adenine ring of the cofactor in both proteins) and three sequentially separated amino acids surrounding the adenine moiety of the cofactor. Considering the reports by Kobayashi and Go (1997a, 1997b) of a handful of structurally similar amino acids that interact with the cofactor in DD-ligase and cAPK, we have taken a more detailed look at the tertiary structures of these proteins. Our comparisons demonstrate that the topologically common segments

Reprint requests to: Mark S. Johnson, Department of Biochemistry and Pharmacy, Åbo Akademi University, BioCity 3A, P.O. Box 66, FIN-20521 Turku, Finland; e-mail: johnson@abo.fi

<sup>5</sup>Current address: Department of Biochemistry and Pharmacy, Åbo Akademi University, BioCity 3A, FIN-20521 Turku, Finland.

Abbreviations: DD-ligase, D-Ala:D-Ala ligase; cAPK, cAMP-dependent kinase; RMSD, root-mean-square deviation.



**Fig. 1.** A ribbon representation (**A**) of the three-domain structure of DD-ligase, and (**B**) of the two-domain structure of cAPK, in stereo. Helices are drawn in green,  $\beta$ -strands in red and the rest in blue. Bound cofactors are shown by stick models in black. Amino- and carboxy-terminal ends of the polypeptide chains are indicated by circles and highlighted in purple. Figures 1, 3, 6, and 7 were drawn with the program MOLSCRIPT (Kraulis, 1991).

of the polypeptide chains responsible for the interactions with ATP in these two enzymes are much more extensive than have been suggested previously, providing nearly identical orientation of the cofactor, bound metal ions, and water molecules.

## Results

### Comparison of cAPK and DD-ligase

A preliminary pairwise superposition of the catalytic subunits of cAPK (crystallized with AMP-PNP, a nonhydrolyzable analogue

of ATP) and DD-ligase (crystallized with ATP and D-Ala-D- $\alpha$ -hydroxybutyrate phosphonate, but that reacted at the enzyme's active site to form ADP plus 1(S)-aminoethyl-(2-carboxypropyl)phosphoryl-phosphinic acid as a transition state intermediate) was made by the "Fit Atoms" command of SYBYL computer program (Tripos Associates, Inc., St. Louis, MO) on the basis of the  $C_{\alpha}$ -atoms of the four consecutive residues reported by Kobayashi and Go (1997a, 1997b): **Met120**, **Glu121**, **Tyr122**, **Val123**, and Glu180, Lys181, Trp182, Leu183 in cAPK and DD-ligase, respectively (here and later all structural data relating to cAPK will be marked in bold). After this initial superposition, new equivalent

pairs of amino acids were defined on the edges of the **Met120–Val123** (Glu180–Leu183) segments by means of visual inspection. The segments were extended until the chains diverged completely. These elongated segments were super-imposed again and all equivalent segments of three or more residues were then identified and extended as before. As a result of several iterations of this procedure, a number of strands, helices, and loops totaling 103 C<sub>α</sub>-atoms from each of the two structures were superimposed with a RMSD of 3.19 Å (Figs. 2, 3A). We will refer to this alignment as the “original hand-made” one. Interestingly, as a result of the structural alignment, the two “cofactors” (**AMP–PNP** and ADP) were found to be similarly oriented in space with an RMSD between the adenine moieties equal to 0.98 Å.

Independently, the pairwise superposition of the same subunits of cAPK and DD-ligase was made automatically by the computer program GENFIT (the algorithm will be described elsewhere). Given only a pair of protein structures as input to the program, GENFIT finds the largest region of structural similarity (stretches of at least four contiguous C<sub>α</sub>-atom pairs where the C<sub>α</sub>–C<sub>α</sub> distance between each pair is ≤3.0 Å) without any a priori knowledge about any presumed structural similarity. Because GENFIT is a generic structural alignment procedure, it is able to pinpoint local structural similarities irrespective of the position of the similar segments in the molecule: thus, both the order of the segments along the sequence and the direction (N → C-terminus versus C → N-terminus) of the segments relative to each other in the compared structures is irrelevant (unlike programs designed to find similarities only among homologous protein structures). Three consecutive applications of this computer program each revealed from 9 to 12 segments composed of from 73 to 83 equivalent C<sub>α</sub>-atom pairs (the first comparison is shown in Fig. 3B). The RMSD (equivalent C<sub>α</sub>-atoms) among these three structural alignments varies from 2.13 to 2.22 Å; as a result of the super-positioning of the structures, the adenine moieties of **AMP–PNP** and ADP have RMSDs ranging from 1.45 Å to 1.57 Å. These data are summarized in Table 1, where they are compared with the result of our original hand-made alignment (Figs. 2, 3A).

For comparison, note that the pairwise superposition of cAPK with other related kinase structures indicates that casein kinase-1

(with bound Mg<sup>2+</sup>-ATP) and cyclin-dependent kinase 2 (with bound Mg<sup>2+</sup>-ATP) are the most closely related known three-dimensional structures to cAPK (De Bondt et al., 1993; Zheng et al., 1993a; Xu et al., 1995). In the first case, 179 C<sub>α</sub>-atoms of cAPK were superimposed with an RMSD of 1.8 Å, and in the second case the same set of atoms gave an RMSD of 2.5 Å (Xu et al., 1995).

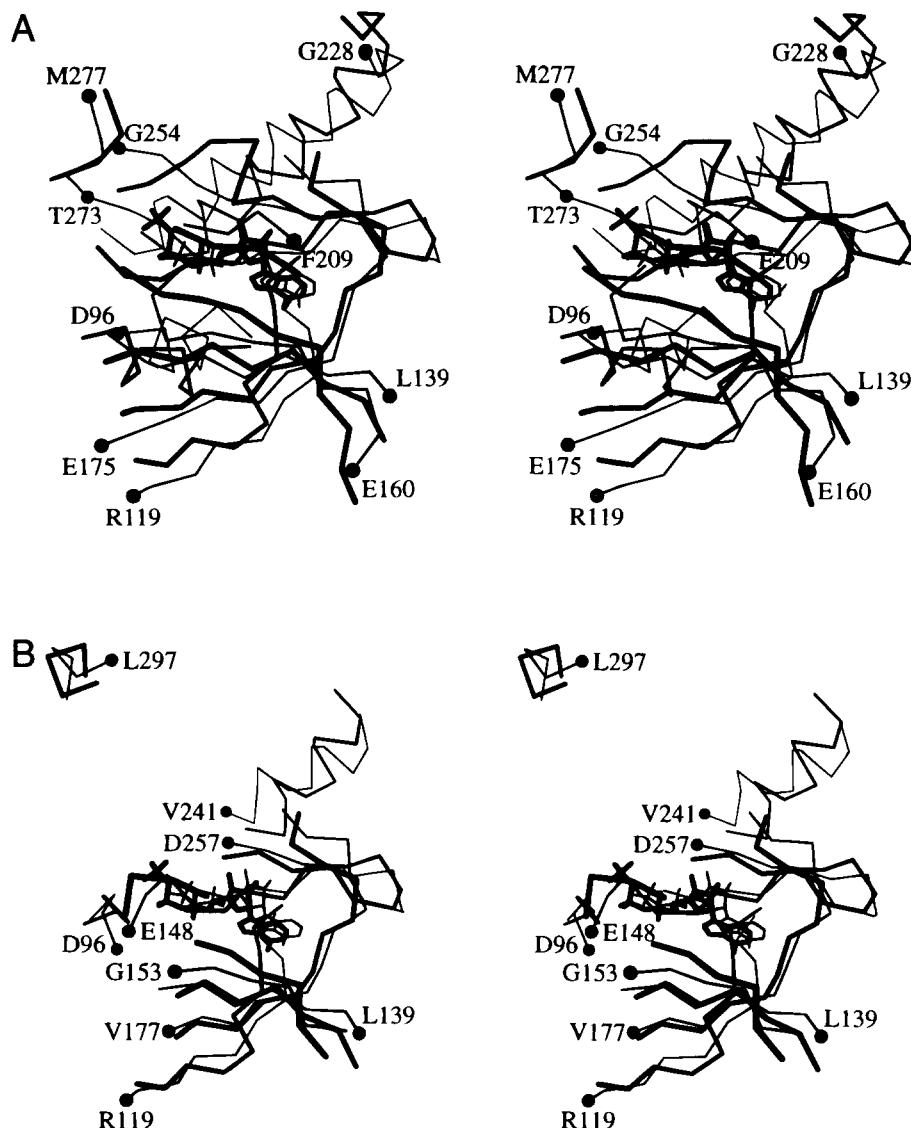
#### *Similar supersecondary structures surround the ATP-binding site*

DD-Ligase and cAPK have 10 segments of the polypeptide chain in common (Figs. 2, 3A). Note that all aligned segments have the same direction of the polypeptide chain. These 10 segments in cAPK include all seven principal β-strands existing in the protein: five antiparallel β-strands, **S1–S5**, from the smaller domain and two antiparallel β-strands, **S7** and **S8**, from the larger domain. Strands **S1–S5** form a β-sheet surrounding the cofactor, and strands **S7** and **S8** form a β-hairpin loop structure. The ATP-binding site is located between the **S1–S5** β-sheet and the **S7–S8** loop (Fig. 3A). The structure-based alignment also contains one α-helix from each of the two domains (helices **HC** and **HE**). Helix **HC** is positioned in the cofactor environment close to the **S1–S5** β-sheet and the **S7–S8** hairpin, while helix **HE**, on the contrary, is not situated in the vicinity of the **AMP–PNP** and is separated from the cofactor binding site by the **S7–S8** β-hairpin (Fig. 3A).

In DD-ligase, the super-secondary structures surrounding the cofactor binding site are constructed in a very similar way to that seen for cAPK (Figs. 2, 3A). Ten segments in the alignment contain all four antiparallel β-strands, **S5–S8**, one of the α-helices from the central domain, **H5**; and two antiparallel β-strands, **S12** and **S13**, and one α-helix, **H10**, from the carboxy-terminal domain. As with the corresponding strands **S2–S5** in cAPK, strands **S5–S8** in DD-ligase also form an anti-parallel β-sheet surrounding the cofactor (Fig. 3A), but a loop formed by residues Phe209 through Tyr212 is structurally equivalent to the extra strand, **S1**, seen in cAPK. Helix **H5** is adjacent to the cofactor and is situated close to the antiparallel β-sheet. Strands **S12–S13** form a β-hairpin loop near helix **H5**. Thus, as in the case of cAPK, strands **S5–S8** and the

DD-ligase	209 FYDY 212	151 SVGMSKVVAE 160	S7(155–157)	S6(141–145)	H5(97–106)
	■	*		■ *	*
CAPK	49 LGTG 52	54 FGRVMLVKHK 63		67 NHFAMKYILD 75	90 NEKRILQAVN 99
	S1(43–52)	S2(55–62)		S3(67–75)	HC(85–97)
DD-ligase		S5(114–118)	S8(176–181)		H10(231–248)
	110 PVAPWVALTR 119		175 EEVLIEKWLSGPEF 188	228 GLEASQEANLQALVLKAWTTLG 249	
			* & *		
CAPK	102 FLVKLEYSFK 111	115 NLYVMVMEYVPGGEM 128	137 RFSEPHARFYAAQIVLTFEYLH 158		
	S4(106–111)	S5(115–120)	HE(140–159)		
DD-ligase	S12(253–260)	S12(253–260), S13(266–272)			
	254 GRI 256	257 DVMLDSDGQFYLLLEAN 273	274 SPGM 277		
	#	§	#		
CAPK	165 RDL 167	171 NLLIDQQGYIQVTDFGF 187	199 CGTP 202		
		S7(172–174), S8(180–182)			

**Fig. 2.** Starting from the four equivalent pairs of residues reported by Kobayashi and Go (1997a, 1997b) for DD-ligase and cAPK, 103 structurally equivalent residues in 10 segments were found to be common to these two proteins (original hand-made alignment). The start-stop positions of the secondary structures are noted above and below the sequences, according to the X-ray crystallographic data (Bossemeyer et al., 1993; Fan et al., 1997). Equivalent residues forming hydrogen bonds and hydrophobic interactions with the cofactors are marked by “\*” and circle with four tips, respectively. One pair of residues, **Val123** (Leu183), participating in both hydrophobic (side-chain) and polar (main-chain) interactions with bound **AMP–PNP** and ADP is indicated by “&.” Two pairs of equivalent amino acids that coordinate Mn<sup>2+</sup> and Mg<sup>2+</sup> cations are labeled by “\$.” Two pairs of equivalent amino acids that are crucial to the mechanisms of the reactions catalyzed by cAMP-dependent protein kinase and DD-ligase are indicated by “#.”



**Fig. 3.** Stereoview of the pairwise superposition (A) of 10 segments in the original alignment of Figure 2, and (B) the 9 segments of the first GENFIT alignment (GENFIT1 in Table 1). Fragments of DD-ligase (with ADP) and cAPK (with the AMP-PNP analogue of ATP) are shown by thin and thick lines, respectively. The crystallographic numbering of the residues at the ends of each segment from DD-ligase is shown.

S12–S13  $\beta$ -hairpin from DD-ligase form a sandwich structure with the cofactor located in between them.

One of the most striking features of the structural alignment (Fig. 2) is the presence of seven equivalent, consecutively ordered segments in the polypeptide chain of both proteins (Fig. 4). These seven segments are formed from 80 amino acids within the segments Asn90–Pro202 and Asn96–Met277. Five of these seven segments incorporate six equivalent elements of secondary structure (Fig. 5), while the remaining two short segments, Arg165–Leu167 (Gly254–Ile256) and Cys199–Pro202 (Ser274–Met277), do not. Segment Asn96–Met277 in DD-ligase is 69 residues longer than segment Asn90–Pro202 in cAPK, and this is due mainly to two long insertions between  $\beta$ -strands S5 and S8 and  $\beta$ -strand S8 and  $\alpha$ -helix H10. The first long insertion (Fig. 5) includes three  $\alpha$ -helices: H6 (119–124), H7 (129–137), and H8 (160–172), and

two  $\beta$ -strands: S6 (141–145) and S7 (155–157) (secondary structure assignments according to Fan et al., 1997). Interestingly,  $\beta$ -strands S6 and S7 participate in the structure-based alignment but match  $\beta$ -strands S3 and S2, respectively. The second insertion contains three  $\beta$ -strands: S9 (187–193), S10 (201–204), and S11 (222–224), and one  $\alpha$ -helix H9 (212–216). In the case of cAPK (secondary structure assignments according to Bossemeyer et al., 1993), the alignment encompasses all helices and strands within Asn90–Pro202 with the exception of two small  $\beta$ -strands, S6 (162–163), and S9 (189–190), and an  $\alpha$ -helix, HD (128–135), which is possibly an analogue of  $\alpha$ -helix H9 (212–216).

It is interesting to note the disposition of elements of secondary structure within the segments of the amino acid sequence encoded by different exons in cAPK. The entire mouse heart gene of the catalytic subunit of cAPK (the porcine and mouse amino acid

**Table 1.** Equivalent segments, number of equivalences, and RMSD ( $C_{\alpha}$ -atoms only) values between cAPK (bold) and dd-ligase for original hand-made alignment and three consecutive runs of the computer program GENFIT<sup>a</sup>

Segments	Secondary structure <sup>b</sup>	Hand-made alignment	GENFIT1 <sup>c</sup>	GENFIT2 <sup>c</sup>	GENFIT3 <sup>c</sup>	COMMON
1		209–212 <b>S1(43–52)</b>	<b>49–52</b>			
2			151–148 <b>51–54</b>	151–148 <b>51–54</b>	151–148 <b>51–54</b>	
3	S7(155–157) <b>S2(55–62)</b>	151–160 <b>54–63</b>	153–158 <b>56–61</b>	153–158 <b>56–61</b>	153–158 <b>56–61</b>	153–158 <b>56–61</b>
4	S6(141–145) <b>S3(67–75)</b>	139–147 <b>67–75</b>	139–145 <b>67–73</b>	140–145 <b>68–73</b>	139–145 <b>67–73</b>	139–145 <b>67–73</b>
5	H5(97–106) <b>HC(85–97)</b>	96–105 <b>90–99</b>				
6	S5(114–118) <b>S4(106–111)</b>	110–119 <b>102–111</b>	110–119 <b>102–111</b>	110–119 <b>102–111</b>	110–119 <b>102–111</b>	110–119 <b>102–111</b>
7	S8(176–181) <b>S5(115–120)</b>	175–188 <b>115–128</b>	177–188 <b>117–128</b>	177–188 <b>117–128</b>	177–188 <b>117–128</b>	177–188 <b>117–128</b>
8				224–226 <b>135–137</b>		
9	H10(231–248) <b>HE(140–159)</b>	228–249 <b>137–158</b>	231–241 <b>140–150</b>	231–241 <b>140–150</b>	231–241 <b>140–150</b>	231–241 <b>140–150</b>
10	S12(253–260)	254–256 <b>165–167</b>				
11	S12(253–260) S13(266–272) <b>S7(172–174)</b> <b>S8(180–182)</b>	257–273 <b>171–187</b>	257–270 <b>171–184</b>	257–270 <b>171–184</b>	257–270 <b>171–184</b>	257–270 <b>171–184</b>
12			96–93 <b>186–189</b>	96–93 <b>186–189</b>	96–93 <b>186–189</b>	
13		274–277 <b>199–202</b>				
14				85–82 <b>209–212</b>	85–82 <b>209–212</b>	
15				197–200 <b>228–231</b>	197–200 <b>228–231</b>	
16			297–301 <b>246–250</b>	297–301 <b>246–250</b>	297–301 <b>246–250</b>	
Number of equivalences ( <i>n</i> )		103	73	83	81	60
RMSD on equivalences RMSD ( <i>n</i> ), Å		3.19	2.13	2.22	2.19	2.03
RMSD between adenine moieties, Å		0.98	1.45	1.57	1.47	1.48

<sup>a</sup>Start and end positions of the equivalent segments are listed.

<sup>b</sup>Taken from Fan et al. (1997) and Bossemeyer et al. (1993).

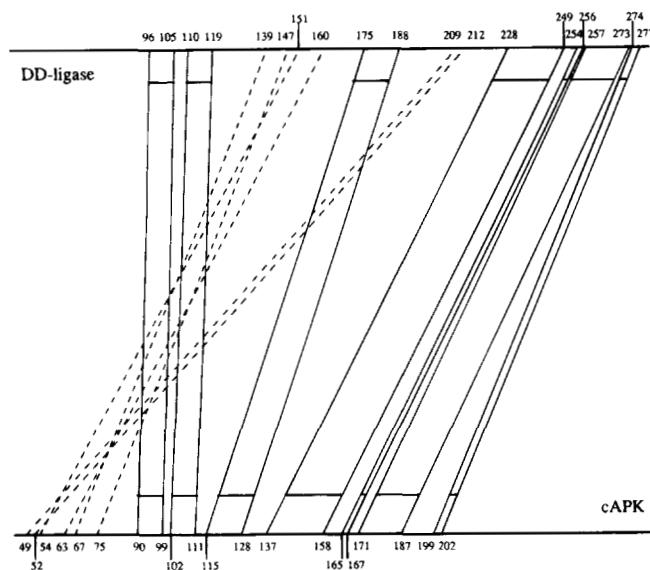
<sup>c</sup>A distance cutoff of 3.0 Å was used to define equivalences in GENFIT; this was not the case for the hand-made alignment.

sequences share 97% identity (Zheng et al., 1993b)) is divided into 10 exons (Chrivia et al., 1988). In particular, exon 3 (residues 36–78) encodes  $\beta$ -strands S1 (not shown), S2, and S3; exon 4 (79–111),  $\alpha$ -helix HC and  $\beta$ -strand S4; exon 5 (112–139),  $\beta$ -strand S5 and  $\alpha$ -helix HD; exon 6 (141–181),  $\alpha$ -helix HE and  $\beta$ -strands S6, S7, and S8; and exon 7 (182–213),  $\beta$ -strand S9 and  $\alpha$ -helix HE1 (207–211) (not shown) (Fig. 5). Thus, segment Asn90–Pro202 is encoded by the exons 4–7. Unfortunately, no data are available concerning dd-ligase at this time.

Automated comparisons of these proteins, made without any human intervention, were performed using GENFIT, which has as its basis a genetic algorithm-based comparison procedure (May &

Johnson, 1994, 1995). The key to the procedure is the random exploration of all possible solutions to the superposition, and it is no surprise that the results produced by GENFIT vary from one run of the GENFIT program to another—this is a characteristic of genetic algorithms. Nonetheless, in three consecutive runs about 70 pairs of residues from nine segments of secondary structure are present in all three alignments (columns 4–6 in Table 1). The differences between the three GENFIT runs consist only of short four-residue long sequences (rows 8, 14, and 15 in Table 1).

Despite these small differences among the GENFIT alignments, the original hand-made alignment and the three GENFIT alignments share six common segments containing 60 equivalent resi-



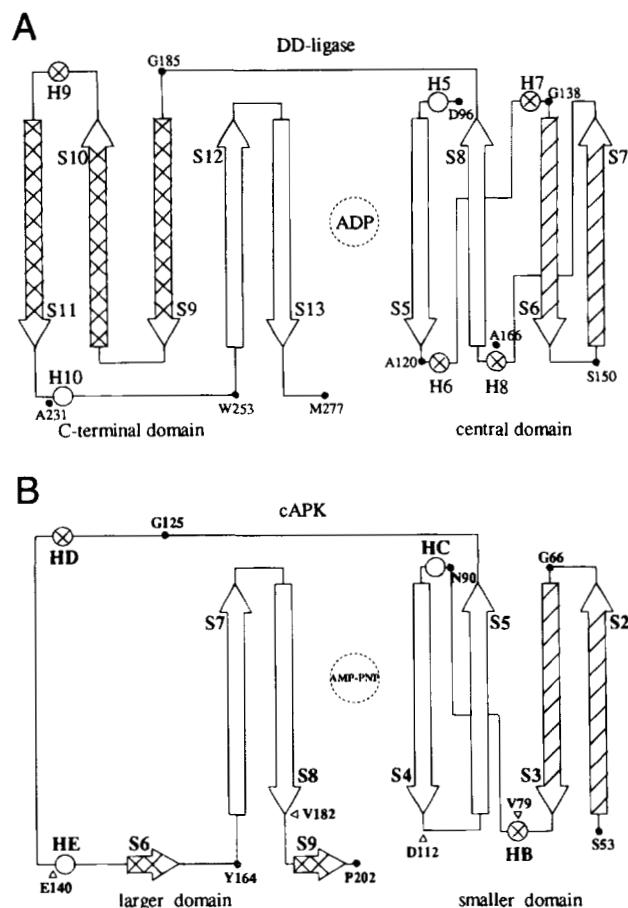
**Fig. 4.** A block diagram of the equivalent segments along the polypeptide chains of cAPK and dd-ligase. Ends of the 10 equivalent segments in each protein are given according to the original hand-made alignment shown in Figure 2. Seven of 10 matching segments, which are consecutively ordered, are connected by solid vertical lines and marked by short horizontal lines parallel to the segments. The last three inversely ordered segments are shown by dashed lines.

dues that superimpose with an RMSD of 2.03 Å (column 7 in Table 1). These six segments include six  $\beta$ -strands and one  $\alpha$ -helix (Table 1). In comparison with the original hand-made alignment, these six segments do not contain  $\alpha$ -helices HC and H5, which were not part of any of the GENFIT solutions, and this is most likely due to the 3.0 Å  $C_{\alpha}$ - $C_{\alpha}$  cutoff distance used to define equivalencies between the two structures in GENFIT. Interestingly, two of the three segments given in all GENFIT alignments, yet absent from the original hand-made alignment, have polypeptide chains that run in the opposite direction (rows 2 and 12 in Table 1).

Thus, although the folds of the enzymes are unquestionably different, we can conclude from this analysis that analogous segments of cAPK and dd-ligase form two equivalent supersecondary structures: (a) a  $\beta$ -sheet formed from four strands S2–S5 (S5–S8) and (b) an  $\alpha/\beta$  complex formed from another domain, including an  $\alpha$ -helix HE (H10) and two antiparallel  $\beta$ -strands S7–S8 (S12–S13) forming a  $\beta$ -hairpin structure (Fig. 6). In both enzymes the cofactor binds to a cleft located between the  $\beta$ -sheet and the two antiparallel  $\beta$ -strands of these two supersecondary structures. Surprisingly, the common four-residue segment found by Kobayashi and Go (1997a) is not entirely a part of any of these two supersecondary structures. Instead, it is a part of the segment Met120–Met128 (Gly180–Phe188) which, together with the segment Phe102–Leu106 (Pro110–Trp114), connects the two supersecondary structures and plays an important role in their mutual disposition.

## Discussion

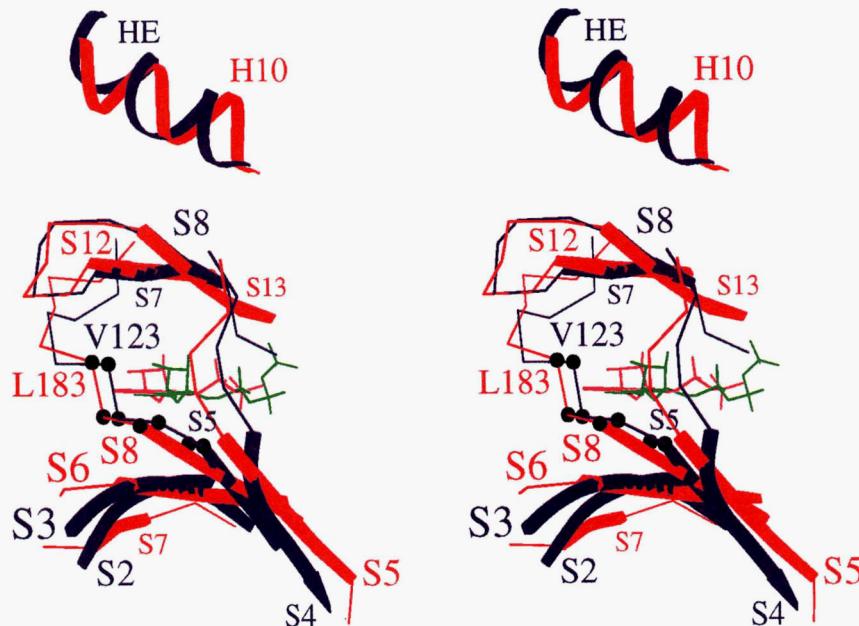
Because the conformations of AMP-PNP and ADP bound to the cofactor binding site in the crystal structures coincide well as a result of our superposition of the two proteins, we might expect that the amino acid residues responsible for key interactions be-



**Fig. 5.** Topological diagrams of the secondary structure elements within (A) segment Asp96–Met277 of dd-ligase and (B) segment Asn90–Pro202 together with  $\beta$ -strands S2 and S3 and  $\alpha$ -helix HB (76–81) of cAPK. Helices and  $\beta$ -strands are drawn by circles and arrows, respectively, and numbered after Fan et al. (1997) and Bossemeyer et al. (1993). The ends of the segments, Asp96–Met277 and Asn90–Pro202, and some amino acids are shown by small black circles. Amino acids are designated by single letter code. Six equivalent elements of the secondary structure within these segments are shown unshaded. Two  $\beta$ -strands, S2 and S3, which are outside the segment Asn90–Pro202 and structurally equivalent to them  $\beta$ -strands S7 and S6 are shaded by “/.” The other (nonequivalent) elements of the secondary structure are shaded by “X.” Triangles indicate position of the left ends of segments of the amino acid sequence encoded by 4, 5, 6, and 7 exons in cAPK. Two large circles represent general location of the “cofactors.”

tween the enzymes and AMP-PNP and ADP (and thus ATP) will coincide also, and participate in stabilizing contacts formed both through main-chain and side-chain interactions with these “cofactors.” Indeed, this is the case, and both invariant and conservatively varied residues, mainly among members of the kinase family, are involved in binding AMP-PNP and ADP to these enzymes. Below we consider those residues interacting with the purine base, the ribose ring, the phosphate groups, bound metal cations, and bound water.

The adenine bases are bound to cAPK and dd-ligase residues by equivalent hydrogen bonds between the ring nitrogens N1 of the bases and the backbone amides of Val123 and Leu183, and between nitrogens N6 of the bases and the backbone carbonyls of Glu121 and Lys181 (Fig. 7A). Equivalent nonpolar interactions



**Fig. 6.** Stereoview of the main-chain structure surrounding the cofactor binding site and common to both the original and GENFIT alignments. Six common segments for DD-ligase (red) and cAPK (blue) form two supersecondary structures: a  $\beta$ -sheet formed from four strands S2–S5 (S5–S8) (lying below the “cofactors”), and an  $\alpha/\beta$  complex, including an  $\alpha$ -helix HE (H10) and two antiparallel  $\beta$ -strands S7–S8 (S12–S13) (lying above the “cofactors”). Bound cofactors are shown by stick models in green for cAPK and in violet for DD-ligase. The other common  $\alpha$ -helix of the original hand-made alignment, HC (H5), is not shown. The cofactor binds to a cleft located between these two supersecondary structures. Four equivalent pairs of amino acids reported by Kobayashi and Go (1997a, 1997b) are drawn as black circles, and one pair Val123 (Leu183) is indicated.

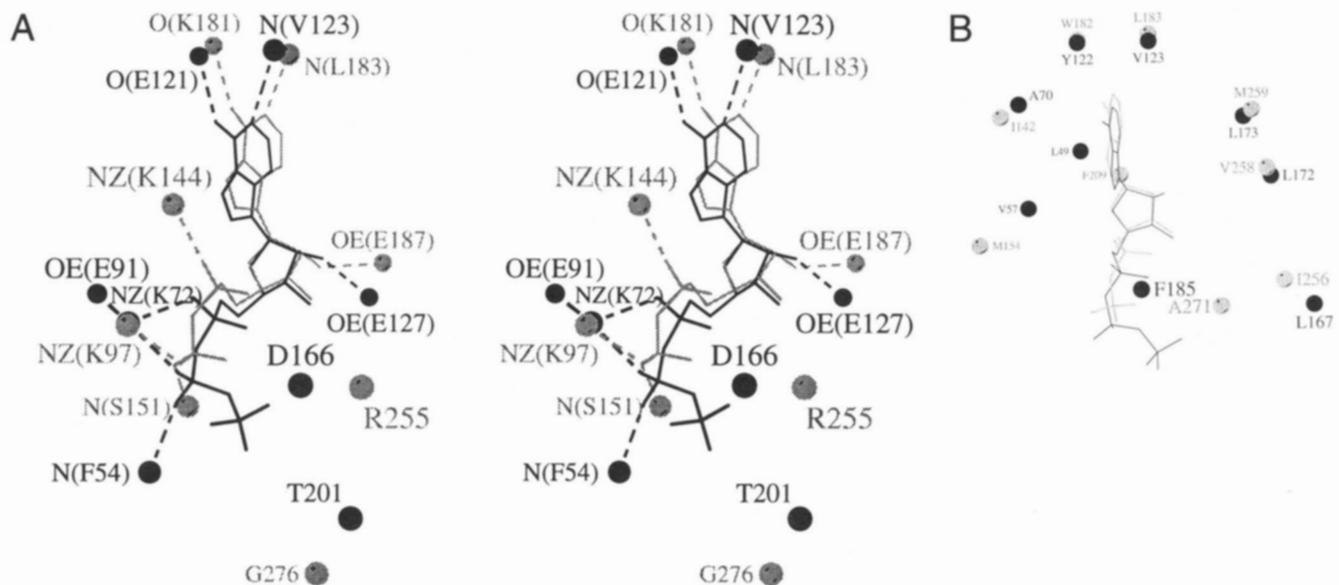
occur between one face of the purine rings and **Val123** and **Leu173** of cAPK and, correspondingly, Leu183 and Met259 of DD-ligase (Fig. 7B); the opposite face of the purine rings also form equivalent hydrophobic interactions with **Leu49**, **Val57** (invariant within the kinase family (Hanks et al., 1988)), **Ala70**, and **Tyr122** in cAPK and, correspondingly, Phe209, Met154, Ile142, and Trp182 in DD-ligase. As a result, both enzymes form a hydrophobic pocket into which they bind the adenine moiety of the cofactor.

The hydroxyl groups of ribose are tightly bound to these enzymes through hydrogen bonding with the side chains of **Glu127** and Glu187 (Fig. 7A). Glu187 also forms a second hydrogen bond with oxygen O3 of the ribose ring in DD-ligase (not shown in Fig. 7A), which does not have an equivalent partner in the cAPK complex but is instead hydrogen bonded to **Glu170**, to **Arg I-14** of a bound inhibitor peptide, and to one water molecule.

The  $\beta$ -phosphate groups form equivalent hydrogen bonds with the main-chain amides of **Phe54** and Ser151 and the  $\alpha$ -phosphate groups with the side chains of **Lys72** and Lys144 (residues invariant throughout both families; Hanks et al., 1988; Fan et al., 1994) (Fig. 7A). The  $\beta$ -phosphate group of DD-ligase also forms an electrostatic link with Lys97, which corresponds to **Glu91** in cAPK (invariant within the kinase family (Hanks et al., 1988)). However, the side chain of **Glu91** forms a hydrogen bond with the  $\beta$ -phosphate group through the side-chain amino group of **Lys72**, which in turn, forms two hydrogen bonds with the  $\alpha$ - and  $\beta$ -phosphate groups (Fig. 7A). In addition, Arg255 and Gly276 stabilize the high-energy phosphinate intermediate in DD-ligase (Fan et al., 1997). These residues correspond to **Asp166** and **Thr201** in cAPK, which are important for the release of product in the kinase (Bossemeyer et al., 1993). Bossemeyer et al. (1993) have stated that the appropriate orientation of

the invariant and conserved polar residues in cAPK is provided by neighboring residues with conserved hydrophobicity such as **Leu167**, **Leu172**, and **Phe185**. These residues (and the corresponding ones in DD-ligase: Ile256, Val258, and Ala271) form a hydrophobic cluster near the cofactor binding site (Fig. 7B).

Two invariant residues (Hanks et al., 1988) in cAPK, **Asn171**, and **Asp184**, and their corresponding amino acids in DD-ligase, Asp257, and Glu270, interact with oxygens of both  $\alpha$ - and  $\beta$ -phosphoryl groups through two atoms each of  $Mn^{2+}$  and  $Mg^{2+}$ . The network of electrostatic links thus formed is shown in Figure 7C. In addition, the two  $Mn^{2+}$  cations in cAPK also coordinate two oxygens of the  $\gamma$ -phosphoryl group of **AMP-PNP** (Fig. 7D). In the case of DD-ligase the  $\gamma$ -phosphate of ATP has been transferred to D-Ala-D- $\alpha$ -hydroxybutyrate phosphonate, and as a result, ADP and the phosphorylated intermediate are found bound to the ATP binding site: the phosphate group of the bound intermediate (1(S)-aminoethyl-(2-carboxypropyl)phosphoryl-phosphinic acid) mimics the  $\gamma$ -phosphoryl group of the cofactor in cAPK. Two oxygens of the phosphate group of this intermediate are coordinated by the two  $Mg^{2+}$  cations in DD-ligase and are indeed structurally equivalent to what is seen in cAPK for the nonhydrolyzable ATP analogue (Fig. 7D). Furthermore, the backbone amide groups of the amino acids **Ser53** and **Ser150**, adjacent to segment 3 in the original hand-made alignment (Table 1), form hydrogen bonds with a third pair of structurally equivalent oxygens of the  $\gamma$ -phosphoryl group of **AMP-PNP** in cAPK and the phosphate group of the phosphinate intermediate in DD-ligase. **Ser53** is highly conserved in Ser/Thr kinases (Hanks et al., 1988) and the hydroxyl group of **Ser150** forms a hydrogen bond with **Tyr216**, which is important for ligand specificity and binding in DD-ligase (Fan et al., 1997).



**Fig. 7.** Equivalent (A) polar (in stereo) and (B) hydrophobic environments around the “cofactors” bound in dd-ligase (light line) and cAPK (dark line). In (A), atoms of amino acids forming hydrogen bonds with the cofactors are drawn as circles. Dashed lines indicate the hydrogen bonding between the N6 (N6) of the adenine moiety and the backbone carbonyls of **Glu121** (Lys181), **N1** (N1) of the adenine moiety and the backbone amides of **Val123** (Leu183), **O2** (O2) of the ribose and the side chains of **Glu127** (Glu187), **O1A** (O1A) of the  $\alpha$ -phosphate and the side chains of **Lys72** (Lys144), and **O1B** (O1B) of the  $\beta$ -phosphate and the main-chain amides of **Phe54** (Ser151). The side chain of Lys97 forms a hydrogen bond with O3B of ADP in dd-ligase. A carboxyl group of **Glu91** in cAPK, structurally equivalent to Lys97, tightly binds to **O2B** of AMP-PNP through a hydrogen bond with the side-chain amino group of **Lys72**. Residue **Lys72**, invariant in the kinase family, forms two electrostatic links with the  $\alpha$ - and  $\beta$ -phosphate groups of AMP-PNP. Positions of the  $C_{\alpha}$ -atoms of **Asp166** (Arg255) and **Thr201** (Gly276), important amino acids for the function of these enzymes, are also indicated. In (B), nine pairs of equivalent amino acids forming a hydrophobic environment about the cofactors are drawn as circles: **Leu49** (Phe209), **Val57** (Met154), **Ala70** (Ile142), **Tyr122** (Trp182), **Val123** (Leu183), **Leu167** (Ile256), **Leu172** (Val258), **Leu173** (Met259), and **Phe185** (Ala271). In (C), equivalent interactions among amino acids surrounding the  $\alpha$ - and  $\beta$ -phosphate groups of AMP-PNP and ADP, and the two bound metal cations are shown in stereo; the carboxyl groups of two equivalent amino acid pairs, **Asn171** (Asp257) and **Asp184** (Glu270) (small circles) coordinate oxygens of the  $\alpha$ - ( $P_1$ ) and  $\beta$ -phosphoryl groups of the cofactors through the **Mn<sup>2+</sup>** and **Mg<sup>2+</sup>** cations (large circles). Dashed lines indicate the metal-ligand bonds and a hydrogen bond between **ND2** (OD1) atom of the side chain of **Asn171** (Asp257) and **OD2** (NH2) atom of the side chain of **Asp166** (Arg255). In (D), the phosphinate intermediate of the dd-ligase complex has been included, and is shown with respect to the super-position of ADP and AMP-PNP, in stereo. The backbone amides of a pair of equivalent amino acids, **Ser53** (Ser150), and the **Mn<sup>2+</sup>** and **Mg<sup>2+</sup>** cations are drawn by small and large circles, respectively. The  $\gamma$ -phosphoryl group ( $P_3$ ) of AMP-PNP and the phosphoryl group ( $P_2$ ) of the phosphinate intermediate are labeled. Dashed lines show the equivalent hydrogen bonds and electrostatic interactions. In (E), equivalent interactions among amino acids surrounding the phosphate groups of AMP-PNP and ADP with the phosphinate intermediate, the two bound metal cations (large circles), and two water molecules **WAT551** (WAT407) and **WAT644** (WAT401) (medium circles) are shown: the carboxyl groups of two equivalent amino acid pairs, **Asn171** (Asp257) and **Asp184** (Glu270) (small circles) coordinate oxygens of the phosphoryl groups of the cofactors through the water molecules. Dashed lines,  $P_1$  and  $P_2$  denote the same as in Figure 7D. (Figure continues on the following pages.)

Finally, there are two pairs of equivalently placed water molecules, **WAT551** (WAT407) and **WAT644** (WAT401), in the active sites (Fig. 7E). **WAT551** and **WAT407** are present within hydrogen bond distance (3.5 Å) to the O3 oxygens of the ribose rings; to the  $\alpha$ - and  $\gamma$ -phosphoryl groups of AMP-PNP and the  $\alpha$ -phosphoryl group of ADP and the  $P_2$  group of the bound phosphinate intermediate; to one pair of metal cations **Mn<sup>2+</sup>** and **Mg<sup>2+</sup>**; and to the side chains of **Asn171** and **Asp257**. **WAT644** and **WAT401** form four hydrogen bonds each with the  $\beta$ - and  $\gamma$ -phosphoryl groups of AMP-PNP and the  $\beta$ -phosphoryl group of ADP and the  $P_2$  group of the phosphinate intermediate; with the other pair of metal cations **Mn<sup>2+</sup>** and **Mg<sup>2+</sup>**; and with the side chains of **Asp184** and **Glu270**. Thus, in each protein there are two identical triads: (a) the **OD1** (OD2) oxygen of the side chain of **Asn171** (Asp257), water molecule **WAT551** (WAT407), and one metal cation **Mn<sup>2+</sup>** (**Mg<sup>2+</sup>**); and (b) the **OD1** (OE1) oxygen of the side chain of **Asp184** (Glu270), water molecule **WAT644** (WAT401), and another metal cation **Mn<sup>2+</sup>**

(**Mg<sup>2+</sup>**). Furthermore, oxygen **OD2** (OE2) of the side chain of **Asp184** (Glu270) connects these two triads together and lies between the metal cations, while the **ND2** (OD1) atom of the side chain of **Asn171** (Asp257) forms a hydrogen bond with atom **OD2** (NH2) of the side chain of **Asp166** (Arg255) (Fig. 7C). As a result, the triads, whose apparent function is to serve as a phosphate anchor, play a significant role in stabilization of the enzyme-cofactor complex.

Overall, the structure-based alignment shows the equivalent placement of eight amino acids participating in essential polar interactions with the cofactor, and of nine residues involved in hydrophobic contacts with it (Fig. 2). These 17 pairs of amino acids are located on 7 of the 10 common segments reported in Figure 2, and none of these “key” amino acids are located within  $\beta$ -strand **S4** (S5),  $\alpha$ -helix **HE** (H10), or the short segment **199–202** (274–277), due to their distant position from the cofactor (Fig. 3A).

Thus, we see that (a) the “cofactors” have nearly identical conformations when bound to either of these two enzymes; this in-

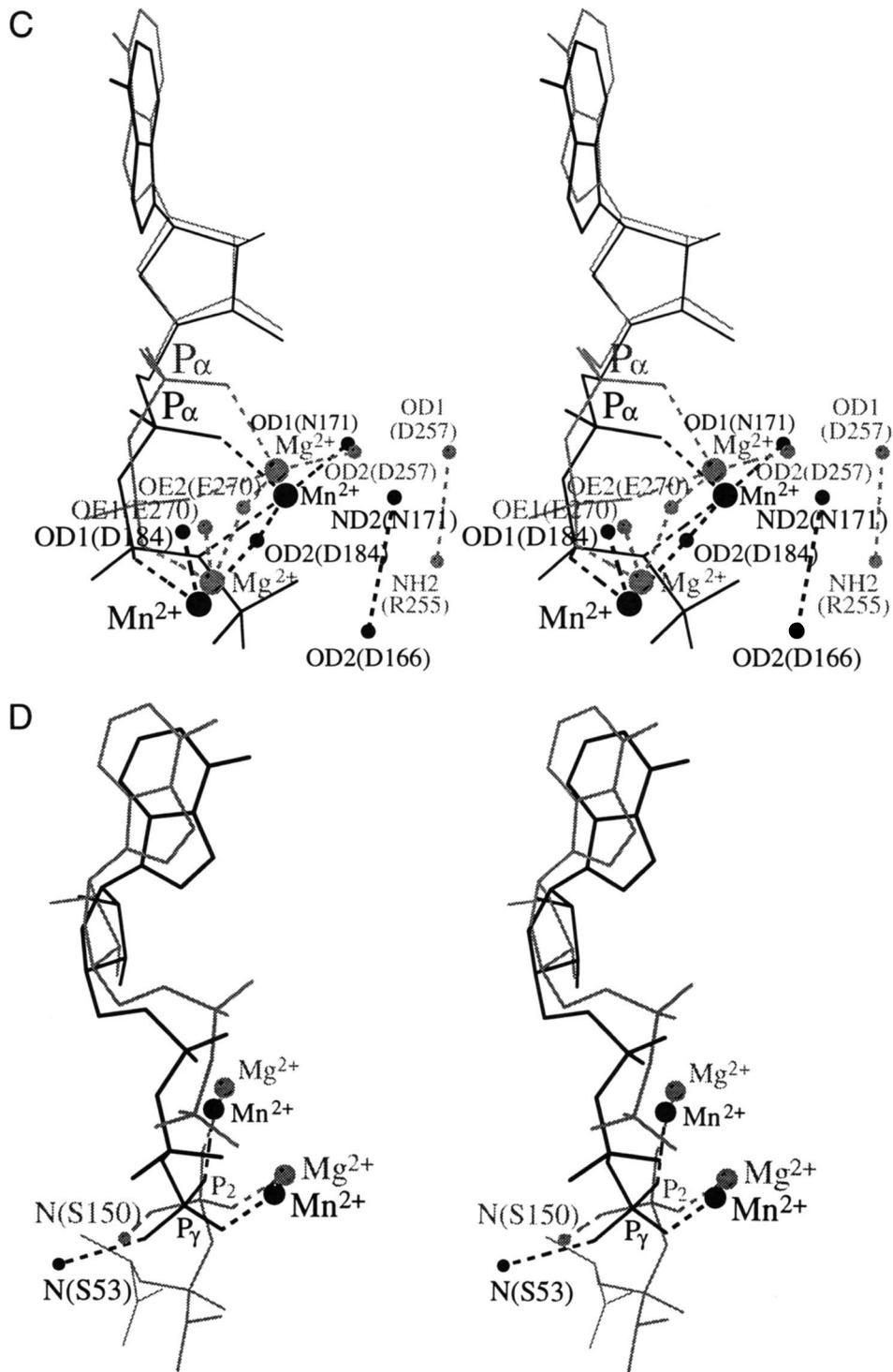


Fig. 7. Continues.

cludes the  $\alpha$ -,  $\beta$ -, and  $\gamma$ -phosphates, and both pairs of metal cations and bound water molecules, and strongly supports a similar mode of binding for ATP to these enzymes; (b) the equivalence of co-factor conformations is coupled with the equivalent disposition of key residues responsible for their specific interactions with the enzymes; (c) in order for these key residues to be appropriately placed about the cofactor metal ions and water molecules, rela-

tively long segments of polypeptide are required; (d) the equivalent disposition of these long segments is provided by the fact that they have identical secondary structures, are packed into two equivalent supersecondary structures ( $\beta$ -sheets and  $\alpha/\beta$ -complexes) and are linked together by two equivalent structures joining them.

All told, this study reveals a remarkable degree of similarity between two families of proteins whose folds are reported to be

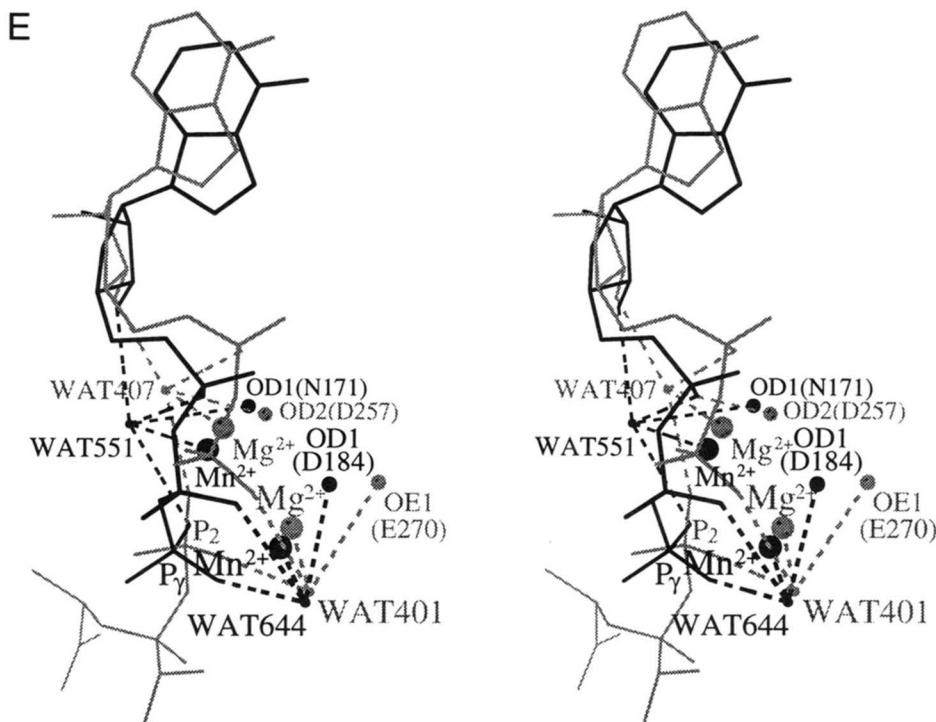


Fig. 7. Continued.

different (Fig. 1) and, thus, are not likely to share a common evolutionary history. Nature can and does converge toward similar local structure when function demands it. The extensive structural similarities seen here suggest that the proteins may be an example of the convergent evolution where the same cofactors are repeatedly exploited in completely different settings. However, it is impossible to rule out a more prosaic common origin for these structural similarities in these two families of enzymes.

#### Materials and methods

The coordinates of the 1.9 Å resolution *Escherichia coli* structure of dd-ligase (1IOW) and the 2.0 Å resolution structure of porcine heart cAMP-dependent kinase (1CDK) were obtained from the Brookhaven Protein Data Bank (Bernstein et al., 1977). Superposition between dd-ligase and cAPK was initially made with the molecular modeling software package SYBYL (Tripos Associates, Inc., St. Louis, MO) on a Silicon Graphics ONYX REALITY ENGINE 2 workstation. A parallel version of the computer program GENFIT was run on two dual-processor Pentium-Pro personal computers running the LINUX operating system. The parallel version of the program was written in GNU C to operate within the LAM parallel processing environment (Ohio Supercomputer Center, The Ohio State University, USA) that implements the Message-Passing Interface (MPI) standard.

#### Acknowledgments

Figure 4 is based on the suggestion of an anonymous referee whom we thank. This work was supported by grants from the Academy of Finland (M.S.J.) and the Technology Development Centre Finland "TEKES" (M.S.J.).

#### References

- Bernstein FC, Koetzle TF, Williams GJB, Meyer EJ Jr, Brice MD, Rodgers JK, Kennard O, Shimanouchi I, Tasumi M. 1977. The Protein Data Bank: A computer-based archival file for macromolecular structures. *J Mol Biol* 112:535–542.
- Bossemeyer D, Engh RA, Kinzel V, Ponstingl H, Huber R. 1993. Phosphotransferase and substrate binding mechanism of the cAMP-dependent protein kinase catalytic subunit from porcine heart as deduced from the 2.0 Å structure of the complex with Mn<sup>2+</sup> adenylyl imidodiphosphate and inhibitor peptide PKI(5–24). *EMBO J* 12:849–859.
- Chrivia JC, Uhler MD, McKnight GS. 1988. Characterization of genomic clones coding for the Cα and Cβ subunits of mouse cAMP-dependent protein kinase. *J Biol Chem* 263:5739–5744.
- De Bondt HL, Rosenblatt J, Jancarik J, Jones HD, Morgan DO, Kim SH. 1993. Crystal structure of cyclin-dependent kinase 2. *Nature* 363:595–602.
- Fan C, Moews PC, Walsh CT, Knox JR. 1994. Vancomycin resistance: Structure of d-alanine:d-alanine ligase at 2.3 Å resolution. *Science* 266:439–443.
- Fan C, Park IS, Walsh CT, Knox JR. 1997. d-Alanine:d-alanine ligase: Phosphonate and phosphinate intermediates with wild type and the Y216F mutant. *Biochemistry* 36:2531–2538.
- Hanks SK, Quinn AM, Hunter T. 1988. The protein kinase family: Conserved features and deduced phylogeny of the catalytic domains. *Science* 241:42–52.
- Herzberg O, Chen CC, Kapadia G, McGuire M, Carroll LJ, Noh SJ, Dunaway-Mariano D. 1996. Swiveling-domain mechanism for enzymatic phosphotransfer between remote reaction sites. *Proc Natl Acad Sci USA* 93:2652–2657.
- Hubbard SR, Wei L, Ellis L, Hendrickson WA. 1994. Crystal structure of the tyrosine kinase domain of the human insulin receptor. *Nature* 372:746–754.
- Kobayashi N, Go N. 1997a. ATP binding proteins with different folds share a common ATP-binding structural motif. *Nat Struct Biol* 4:6–7.
- Kobayashi N, Go N. 1997b. A method for similar protein local structures at ligand-binding sites and its application to adenine recognition. *Eur Biophys J* 26:135–144.
- Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr* 24:945–949.
- Matsuda K, Mizuguchi K, Nishioka T, Kato H, Go N, Oda J. 1996. Crystal structure of glutathione synthetase at optimal pH: Domain architecture and structural similarity with other proteins. *Protein Eng* 9:1083–1092.

- May ACW, Johnson MS. 1994. Protein structure comparisons using a combination of a genetic algorithm, dynamic programming and least-squares minimization. *Protein Eng* 7:475–485.
- May ACW, Johnson MS. 1995. Improved genetic algorithm-based protein structure comparisons: Pairwise and multiple superpositions. *Protein Eng* 9:873–882.
- Murzin AG. 1996. Structural classification of proteins: New superfamilies. *Curr Opin Struct Biol* 6:386–394.
- Thoden JB, Holden HM, Wessenberg G, Raushel FM, Rayment I. 1997. Structure of carbamoyl phosphate synthetase: A journey of 96 Å from substrate to product. *Biochemistry* 36:6305–6316.
- Waldrop GL, Rayment I, Holden HM. 1994. Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase. *Biochemistry* 33:10249–10256.
- Wolodko WT, Fraser ME, James MNG, Bridger WA. 1994. The crystal structure of succinyl-CoA synthetase from *Escherichia coli* at 2.5 Å resolution. *J Biol Chem* 269:10883–10890.
- Xu R-M, Carmel G, Sweet RM, Kuret J, Cheng X. 1995. Crystal structure of casein kinase-1, a phosphate-directed protein kinase. *EMBO J* 14:1015–1023.
- Yamaguchi H, Kato H, Hata Y, Nishioka T, Kimura A, Oda J, Katsube Y. 1993. Three-dimensional structure of the glutathione synthetase from *Escherichia coli* at 2.0 Å resolution. *J Mol Biol* 229:1083–1100.
- Zheng J, Knighton DR, Ten Eyck LF, Karlsson R, Xuong N, Taylor SS, Sowadski JM. 1993a. Crystal structure of the catalytic subunit of cAMP-dependent protein kinase complexed with MgATP and peptide inhibitor. *Biochemistry* 32:2154–2161.
- Zheng J, Knighton DR, Xuong N, Taylor SS, Sowadski JM, Ten Eyck LF. 1993b. Crystal structures of the myristylated catalytic subunit of cAMP-dependent protein kinase reveal open and closed conformations. *Protein Sci* 2:1559–1573.