Eur Biophys J (1997) 26: 135–144

ARTICLE

Nobuo Kobayashi · Nobuhiro Go

A method to search for similar protein local structures at ligand-binding sites and its application to adenine recognition

Received: 9 December 1996 / Accepted: 7 February 1997

Abstract We have developed a method of searching for similar spatial arrangements of atoms around a given chemical moiety in proteins that bind a common ligand. The first step in this method is to consider a set of atoms that closely surround a given chemical moiety. Then, to compare the spatial arrangements of such surrounding atoms in different proteins, they are translated and rotated so that the chemical moieties are superposed on each other. Spatial arrangements of surrounding atoms in a pair of proteins are judged to be similar, when there are many corresponding atoms occupying similar spatial positions. Because the method focuses on the arrangements of surrounding atoms, it can detect structural similarities of binding sites in proteins that are dissimilar in their amino acid sequences or in their chain folds. We have applied this method to identify modes of nucleotide base recognition by proteins. An all-against-all comparison of the arrangements of atoms surrounding adenine moieties revealed an unexpected structural similarity between protein kinases, cAMP-dependent protein kinase (cAPK), and casein kinase-1 (CK1), and D-Ala:D-Ala ligase (DD-ligase) at their adenine-binding sites, despite a lack of similarity in their chain folds. The similar local structure consists of a four-residue segment and three sequentially separated residues. In particular the four-residue segments of these enzymes were found to have nearly identical conformations in their backbone parts, which are involved in the recognition of adenine. This common local structure was also found in substrate-free three-dimensional structures of other proteins that are similar to DD-ligase in the chain fold and of other protein kinases. As the proteins with different folds were found to share a common local structure, these proteins seem to constitute a remarkable example of convergent evolution for the same recognition mechanism.

N. Kobayashi · N. Go (✉)
Department of Chemistry, Graduate School of Science,
Kyoto University, Kitashirakawa-Oiwakecho, Kyoto 606-01, Japan
(Fax 81-75-711-6083; e-mail: go@qchem.kuchem.kyoto-u.ac.jp)

Key words Molecular recognition · Search for similar arrangements of atoms · Adenine recognition · Database search

Abbreviations cAPK, cAMP-dependent protein kinase; DD-ligase, D-Ala:D-Ala ligase; CK1, casein kinase-1; IRK, tyrosine kinase domain of the insulin receptor; GSHase, glutathione synthetase; BNC, biotin carboxylase subunit of acetyl-CoA synthetase; SCS, succinyl-CoA synthetase; PPDK, pyruvate phosphate dikinase; glnRS, glutaminyl-tRNA synthetase; r.m.s.d., root-mean-square deviation; PDB, Protein Data Bank.

1. Introduction

Enzymes specifically recognize their substrates. Identification of a set of molecular interactions responsible for recognition of a specific substrate is an important step towards understanding the molecular mechanism of protein functions. In this paper we will first develop a computerized method for this identification, and then apply the method to find modes of recognition of the adenine moiety in nucleotide binding proteins.

The strategy we adopted in this paper for this identification is to search for similar spatial arrangements of atoms around a given chemical moiety among proteins that bind a common ligand. The underlying hypothesis here is that very similar spatial arrangements of atoms in different proteins imply a functional importance of such atoms.

The catalytic triad shared by serine proteases with different folds, such as chymotrypsin and subtilisin, is a remarkable example to support this hypothesis. In these enzymes, three amino acid residues, Asp, His and Ser, are arranged similarly in space to perform a specific function. The same triad, involving the same amino acid residues, has also been found in lipase (Brady et al. 1990), and a similar triad containing Glu unstead of Asp has been found in acetylcholine-esterase (Sussman et al. 1991).

A similar situation has also been observed for β-lactamase (Strynadka et al. 1992) and aspartic proteinases (Pearl and Blundell 1984), which are a pair of proteins very different in function and fold. In these enzymes, the arrangements of side-chains at their active sites are very similar: the side-chains of Glu-166 and Asp-170 and their bound water molecule in the β-lactamase, and those of Asp-215 and Asp-32 and the bound water molecule of aspartic proteinases (Pearl 1993). This similarity strongly suggests that the mechanism of a part of the catalytic step, activating a bound water molecule, is essentially the same in both enzymes.

In order to carry out the required search, we have developed a computerized method. The method first considered a set of atoms that closely surround a given chemical moiety in a three-dimensional structure of a protein-ligand complex. In order to choose these surrounding atoms, we use Delaunay tessellation (Coxeter 1961). Delaunay tessellation, which is a method mathematically dual to Voronoi tessellation, is a geometrical method to divide space occupied by atoms into tetrahedra whose four vertices are neighboring atoms. As such, a set of Delaunay tetrahedra involving atoms of the chemical moiety defines space near the moiety. We select all protein and solvent atoms in the set as surrounding atoms. Next, the method evaluates similarity between arrangements of surrounding atoms in a pair of proteins, say, A and B. To compare arrangements of surrounding atoms, their coordinates are translated and rotated so that the coordinates of atoms of the chemical moieties are the same. Then, we define such a pair of surrounding atoms, one from protein A and the other from protein B, that occupy similar spatial positions as *corresponding atoms*. We quantify the degree of similarity of a pair of sets of surrounding atoms in terms of the number and the degrees of proximity of such *corresponding atom pairs.* The method is independent of the chain folds, and thus can detect similarities in proteins without similarities in their chain folds. This method, however, is not intended to be a fully-automated searching method without human intervention, but rather to provide a convenient tool to examine the modes of recognition. We have applied the method to search for similar arrangements of atoms around an adenine moiety of bound mononucleotide.

ATP, the most abundant mononucleotide in living cells, works as the universal currency of free energy in biological systems. Other mononucleotides work in different biological contexts; some proteins involved in gene translation use GTP, those involved in sugar metabolism use UTP, and those involved in lipid metabolism use CTP. These enzymes distinguish the slight differences between the base moieties.

There are a large number of structures of protein-ATP/ATP-analogue complexes solved by X-ray crystallography or NMR. Several classes of fold in ATP-binding proteins have been recognized (Schulz 1992) such as: the classical mononucleotide-binding fold (Schulz et al. 1986), the actin fold (Flaherty et al. 1991; Hurley et al. 1993), the fold of the protein kinase family (Bossemeyer et al. 1993; Hanks et al. 1988), the glutathione synthetase fold (also referred to as the ATP-grasp fold) (Artymiuk et al. 1996; Fan et al. 1995; Hibi et al. 1996; Murzin 1996), the class II tRNA synthetase fold (Moras 1992), and the nucleoside diphosphate kinase fold (Swindells and Alexandrov 1994). The mononucleotide-binding sites in proteins of the last four folds are located on the surface of an anti-parallel β-sheet, whereas in the first two folds, they are located at the edge of a β-sheet.

We have previously reported the unexpected structural similarity at the adenine binding sites in proteins with different folds (Kobayashi and Go 1996a), i.e., protein kinases, cAMP-dependent protein kinase (cAPK) (Bossemeyer et al. 1993) and casein kinase-1 (CK1) (Xu et al. 1995), and D-Ala:D-Ala ligase (DD-ligase) (Fan et al. 1994). This finding illuminates a common binding structural motif shared in proteins with different folds.

cAPK (Bossemeyer et al. 1993), one of the serine/threonine kinases, is a member of a large family of evolutionarily related protein kinases, in which some key residues are conserved in their catalytic domain (Hanks et al. 1988). The structure of the catalytic subunit of cAPK consists of two domains: one is a larger, predominantly α-helical domain, formed by most of the C-terminal portion (except for the C-terminal long loop) and the N-terminal α-helix; and the other is a smaller domain – which is made up largely by an anti-parallel β-sheet – formed by most of the N-terminal portion (except for the first α-helix) and the C-terminal long loop (Fig. 1a). In addition to cAPK, structures of both CK1 (Xu et al. 1995), which is also a serine/threonine kinase, and the tyrosine kinase domain of the insulin receptor (IRK) (Hubbard et al. 1994) have also been determined. These three protein kinases show a high degree of structural similarity.

DD-ligase (Fan et al. 1994) is very similar to glutathione synthetase (GSHase) (Yamaguchi et al. 1993) in chain fold. The structure consists of three domains: the N-terminal domain having an α/β topology and the central and the C-terminal domains both containing predominantly an anti-parallel β-sheet (Fig. 1b). In addition to GSHase and DD-ligase, the biotin carboxylase subunit of acetyl-CoA synthetase (BNC) (Waldrop et al. 1994) has been recognized as a protein with the GSHase fold (Artymiuk et al. 1996). The ATP-grasp fold (Murzin 1996) is a part of the GSHase fold consisting of two domains: the central and the C-terminal domains, that grasp the nucleotide molecule between them. This fold has been recognized in recently determined structures of two ADP-binding ligases: succinyl-CoA synthetase (SCS) (Wolodko et al. 1994) and pyruvate phosphate dikinase (PPDK) (Herzberg et al. 1996). All of these enzymes with the GSHase fold catalyze a common reaction, i.e., coupling the conversion of ATP to ADP with the formation of a carbon-nitrogen bond between a carboxyl group and an amino group. SCS and PPDK catalyze the reaction by way of an intermediate involving a phosphorylated histidine residue of the enzyme itself.

In the Methods section, details of the algorithm for searching similar arrangements of surrounding atoms will be described. In the Results and Discussion sections, we

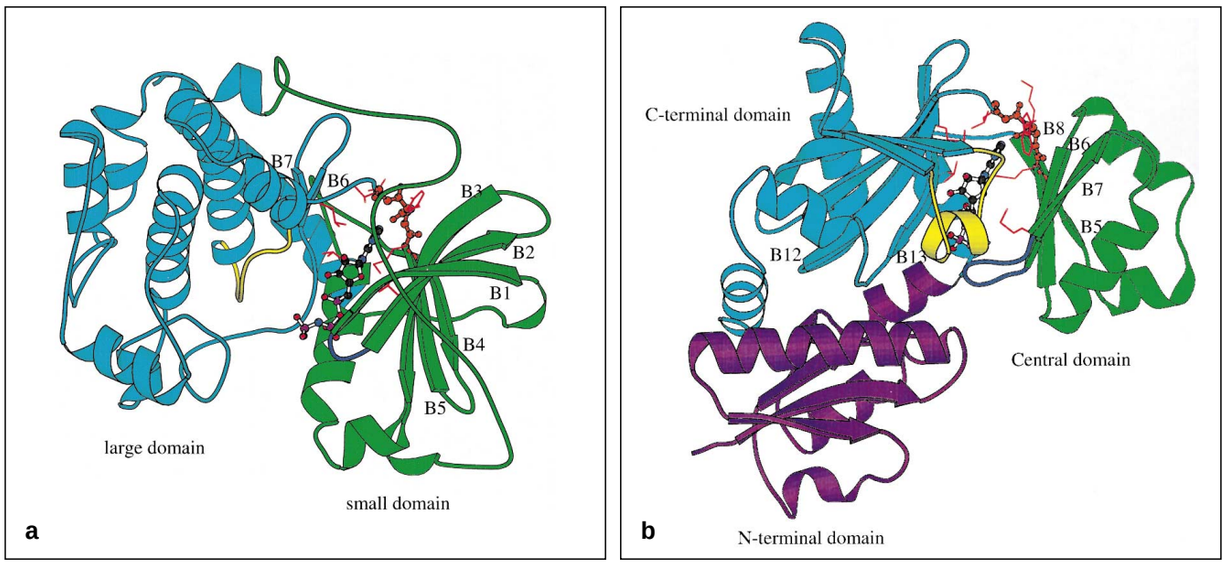

Fig. 1 Ribbon diagrams of **a** cAPK and of **b** DD-ligase. The backbone part of the four-residue segment in each of the proteins is shown by a ball-and-stick model in orange. **a** cAPK consists of two domains: small domain shown in *green* and large domain in *cyan*. The glycine-rich loop is shown in *blue* and the catalytic loop in *yellow.* The bound ATP analogue is shown by a ball-and-stick model. Side-chains in residues participating in the adenine-binding are shown in *red*. β-strands are labeled: B1 (residues 43–52), B2 (residues 55–62), B3 (residues 67–75), B4 (residues 106–111), B5 (residues 115–120), B6 (residues 172–174), and B7 (residues 180–182). **b** DD-ligase consists of three domains: N-terminal domain shown in purple, central domain in green, and C-terminal domain in cyan. The small loop is shown in *blue* and the partially helical large loop shown in *yellow*. The bound ADP is shown by a ball-and-stick model. Side-chains in residues participating in adenine-binding are shown in red. Some β-strands are labeled: B5 (residues 113–117), B6 (residues 141–145), B7 (residues 154–158), B8 (residues 176–181), B12 (residues 253–260) and B13 (residues 265–272). This figure was drawn using the program MOLSCRIPT (Kraulis 1991)

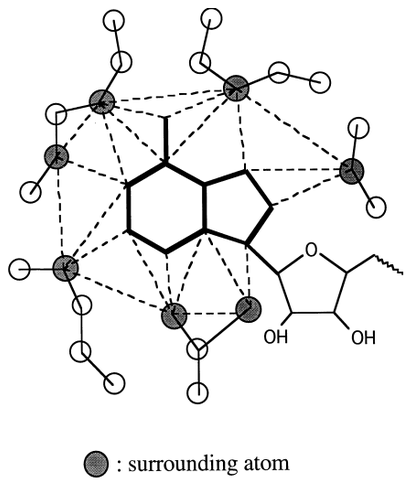

Fig. 2 Schematic representation of Delaunay tessellation around an adenine moiety. Dotted lines connect the neighboring atoms in Delaunay tessellation. Those atoms that are neighbors in Delaunay tessellation to atoms of the adenine are enlisted as surrounding atoms

will report the results of an all-against-all comparison of adenine mononucleotide-protein complexes. In particular, common local structures around the adenine moiety in cAPK and DD-ligase were examined in detail. This examination is further extended to other proteins with the ATP-grasp fold and also to other protein kinases.

2. Method

2.1 List of surrounding atoms

Both Delaunay and Voronoi tessellations have been applied to study various geometrical properties of protein structures (Finney 1975; Kobayashi and Go 1995b; Richards 1974, 1977). We use Delaunay tessellation to define space near a given chemical moiety. All protein and solvent atoms in a set of Delaunay tetrahedra involving atoms of the chemical moiety are listed as surrounding atoms (see Fig. 2). In this method we do not use any parameters such as a cut-off distance to determine surrounding atoms. All heavy atoms in the complex, including bound water molecules, are considered in the selection of the surrounding atoms. We use the algorithm of Delaunay tessellation developed by Tanemura et al. (1983).

2.2 Definition of the similarity score

The second step is evaluation of the similarity between arrangements of surrounding atoms around a common ligand

in a pair of proteins. To compare the surrounding atoms their coordinates are translated and rotated so that the coordinates of the atoms of the chemical moieties are the same. After this treatment, spatial similarity of surrounding atoms can be simply evaluated from their coordinates. Then, we define such a pair of surrounding atoms, one from protein A and the other from protein B, that occupy similar spatial positions as *corresponding atoms.* We quantify the degree of similarity of a pair of sets of surrounding atoms as a similarity score in terms of the number and the degrees of proximity of such *corresponding atom pairs.*

In the definition of *corresponding atom pairs,* we consider only those pairs of atoms with the same atom type, i.e., carbon, oxygen, nitrogen, or sulfur atom. The distance of a pair of surrounding atoms, denoted by $D(a_i, b_j)$, is used to measure the degree of proximity between atom a_i of protein A and atom b_j of protein B. We consider only such b_j that give the smallest $D(a_i, b_j)$ for a given a_i, and only such a_i that give the smallest $D(a_i, b_j)$ for a given b_j. Further we consider only pairs of atoms, a_i and b_j, $D(a_i, b_j)$ smaller than a cut-off value $D_{cut\text{-}off}$, here taken to be 1.5 Å. Later we will discuss the reason for the choice of this cut-off value. Pairs of atoms selected in this way are defined as *corresponding atoms.*

We employ a similarity score given by a product of two factors; one is the number N_C of the *corresponding atom pairs,* and the other is a sum of $|D_{cut\text{-}off} - D(a_i, b_j)|$ over *corresponding atom* pairs. Each *corresponding atom pair* contributes to both factors. The score we employed is "quadratic" in this sense. We will discuss later the reason for this choice of the similarity score.

2.3 Choice of the similarity score

The similarity score defined above has been chosen so as to reliably reproduce well-established cases of substrate recognition by proteins. As such well-established cases, we have chosen four GTP-binding proteins (Traut 1994), cH-ras p21 (PDB code 5p21), elongation factor Tu (PDB code left), the α subunit of transducin (PDB code 1tada) and ADP-ribosylation factor-1 (PDB code 1hura), all of which recognize the guanine moiety by the consensus sequence motif NKxD (Dever et al. 1987) and whose three-dimensional structures are known in complex with guanine mononucleotide. For the four protein-complex structures, we select atoms surrounding the guanine moiety by the method of 2.1. In the list of surrounding atoms, we identify non-hydrogen atoms belonging to residues N, K, D in the above mentioned sequence motif. These atoms were found to occupy very similar positions with respect to the guanine moiety in the four complex structures. The maximum value of $D(a_i, b_j)$ was found to be 1.42 Å. Therefore, we have chosen $D_{cut\text{-}off}$ to be 1.5 Å so that all the corresponding atom pairs identified here are *corresponding atom pairs* in the sense of 2.2.

In the next step, we have applied our method for all proteins whose three-dimensional structures in complex with guanine mononucleotide are known. In this calculation we tried similar but slightly different score functions from the one described in 2.2. For example, similarity scores consisting of only N_C, or only of the sum of $|D_{cut\text{-}off} - D(a_i, b_j)|$ were also examined. For these various choices of the similarity score functions, protein pairs giving high values of the score were listed. The score function described in 2.2 was found to be the best in the sense that protein pairs from the four protein complexes studied above were always top in the list.

3. Results

We applied the method to identify modes of adenine recognition by mononucleotide-binding proteins. The base moiety is an aromatic compound and thus has no conformational flexibility. This character of the base moiety is well suited for the method developed here, because common base moieties can be superposed exactly. We report here results of searching for similar spatial arrangements of atoms around an adenine moiety.

3.1 Initial and representative datasets

The dataset of adenine mononucleotide-protein complexes are taken from the three-dimensional structures deposited in the Brookhaven Protein Data Bank (PDB, January 1996, Release #75) (Bernstein et al. 1977). The bound adenine monunucleotides are identified by examining all ligands in those PDB entries that are determined by X-ray crystallography and by accepting those ligands that include some phosphates, one sugar, and one adenine base. Therefore, not only ATP but also all the other adenine mononucleotides are selected for comparison of arrangements of atoms around the adenine moiety. Those entries that contain only Cα coordinates of protein atoms are removed. This gives 121 bound adenine mononucleotides in 71 entries. Thus, an initial dataset consists of 121 adenine-binding sites.

We first carried out an all-against-all comparison of the 121 sets of arrangements of surrounding atoms. The result of this comparison showed that the method properly recognized distinctly similar pairs such as the same binding sites in different subunits of a homo-polymer, or the same binding sites of the same protein in different PDB entries. Based on this result, we defined the initial dataset by manually removing such essentially identical binding sites, and then we obtained a representative dataset containing 38 different adenine-binding sites.

3.2 A common local structure shared by proteins with different folds

Next, we re-examined pairs of proteins having significantly higher scores in the representative dataset. Such pairs are, in general, found to be only in the binding sites of similar proteins, such as adenylate kinase and uridylate

Table 1 Results of a search for similar arrangements of surrounding atoms to those of cAPK (1cdka). The number of surrounding atoms in 1cdka is 27. SA denotes surrounding atoms, and CA denotes *corresponding atoms*

PDB code	Name	No. SAs	Score	No. CAs
1gtr	Glutaminyl-tRNA synthetase (glnRS)	38	175.3	16
2dln	D-Ala:D-Ala ligase (DD-ligase)	34	139.6	14
1csn	Casein kinase-1 (CK1)	30	127.8	15
1ses	Seryl-tRNA synthetase	36	83.1	14
1lgr	Glutamine synthetase	43	79.1	14

kinase both of which bind ATP at similar positions of the classical mononucleotide-binding fold. Nevertheless, from close examinations of high scoring pairs, we found an unexpected similarity between proteins with different folds, i.e., cAMP-dependent protein kinase (cAPK) and D-Ala:D-Ala ligase (DD-ligase).

Table 1 shows the result of a search for cAPK (PDB code 1cdka) against 38 representative datasets. The three highest scoring proteins are glutaminyl-tRNA synthetase (glnRS, PDB code 1gtr, a member of class I tRNA synthetase) (Rould et al. 1991), DD-ligase (PDB code 2dln, a protein with the GSHase fold) and CK1 (PDB code 1csn, a protein kinase). The similarity between cAPK and CK1 is, in a sense, trivial because both of them are in the protein kinase family, but neither glnRS nor DD-ligase has been known to have similarities to the protein kinase family. Therefore, we should ask the following question. How similar are the local structures of the adenine recognition sites in these proteins to that of cAPK?

To answer the above question, we examined the similarity of local structures of residues containing the *corresponding atoms*. The similarities of these local structures were visually inspected. Figure 3 shows the local structures of (a) DD-ligase and (b) glnRS superposed to that of cAPK. These figures clearly show that DD-ligase is extensively similar to cAPK even in the arrangements of the residues containing the *corresponding atoms*, while glnRS is similar to cAPK only in the arrangements of the *corresponding atoms*. Among seven corresponding residues between DD-ligase and cAPK (Fig. 3a), the backbone structures of corresponding four-residue segments (residue numbers from 180 to 183 in DD-ligase, and from 120 to 123 in cAPK, respectively) are spatially very similar. These results indicate that the similarity beteen proteins with the protein kinase fold and with the GSHase fold is significant, wheras the similarity between proteins with the protein kinase fold and with the class I tRNA synthetase fold is superficial.

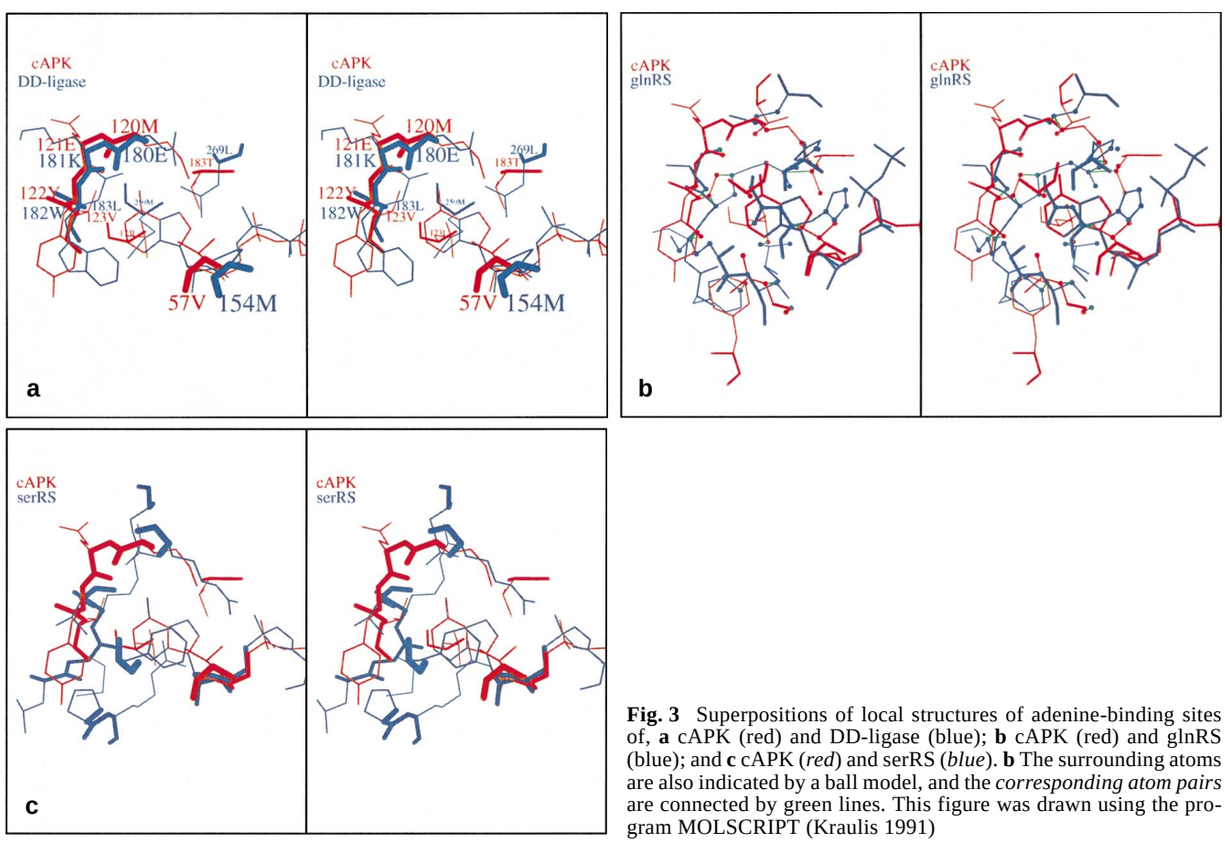

Fig. 3 Superpositions of local structures of adenine-binding sites of, **a** cAPK (red) and DD-ligase (blue); **b** cAPK (red) and glnRS (blue); and **c** cAPK (*red*) and serRS (*blue*). **b** The surrounding atoms are also indicated by a ball model, and the *corresponding atom pairs* are connected by green lines. This figure was drawn using the program MOLSCRIPT (Kraulis 1991)

As we have discussed in the Introduction, proteins with the GSHase fold are different from the protein kinase family in their chain fold, but the mononucleotide-binding sites of both proteins have been shown to be located on the surface of an anti-parallel β-sheet. There are two other fold groups, the class II tRNA synthetase fold and the nucleoside diphosphate kinase fold, that also bind the mononucleotide on an anti-parallel β-sheet. It is reasonable to expect that proteins with the above two folds have similar local structures to those of DD-ligase and cAPK. Seryl-tRNA synthetase (PDB code 1ses) (Belrhali et al. 1994) is a class II tRNA synthetase and is the fourth highest scoring protein in Table 1. We also examined the local structure of those residues containing the *corresponding atoms*, but this protein does not show any significant structural similarity to cAPK (Fig. 3c). We also examined the local structure of a nucleoside diphosphate kinase (PDB code 1nlk) (Williams et al. 1993), whose similarity score with cAPK is 23.1 (the 21st score in 38 data), and found no significant similarity to cAPK. These examinations indicate that only the protein kinase family (cAPK and CK1) and DD-ligase have common local structures at the adenine recognition site among the proteins whose binding sites are located on the surface of an anti-parallel β-sheet. Note that there are no proteins with the GSHase fold or from the protein kinase family in the representative dataset except for the above three proteins. (Recently the structure of GSHase-ADP complex has been solved and deposited in the PDB (Hara et al. 1996).)

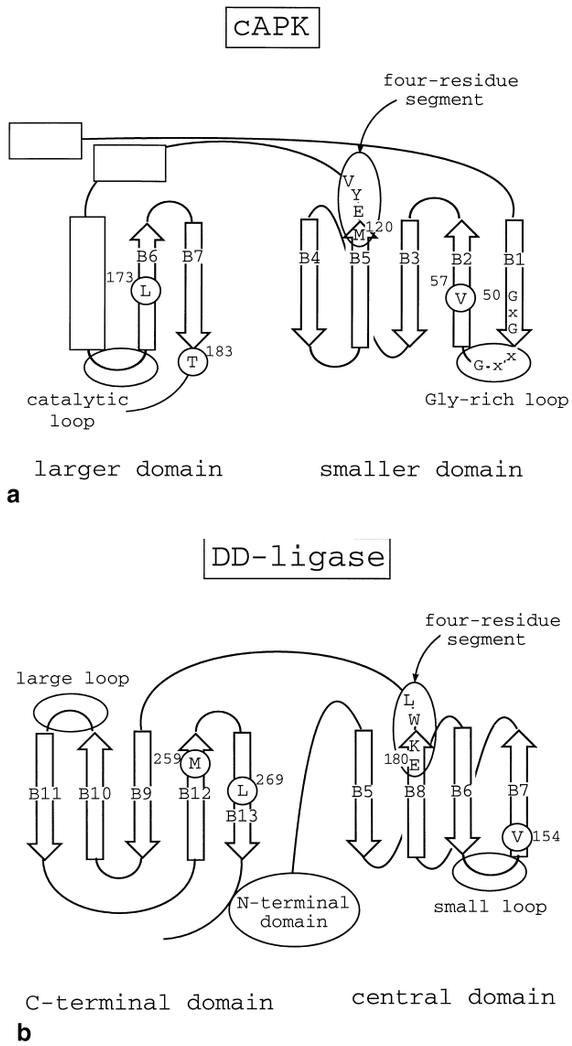

Fig. 4 Topological diagrams of **a** cAPK and **b** DD-ligase. *Arrows* represent β-strands and *boxes* represent α-helices. β-strands are numbered, according to their order in the sequences. Seven amino acids involving the adenine-binding sites are also shown. **a** All β-strands and some α-helices in the large domain, and the finger print sequences are shown, and **b** only β-strands in the central and C-terminal domains are shown

3.3 Characteristic features of the common local structure

As we indicated above, the local structures common between cAPK and DD-ligase consist of seven residues: in cAPK, residue of the four-residue segment are Met-120, Glu-121, Tyr-122 and Val-123, and the other three residues are Val-57, Leu-173 and Thr-183, and in DD-ligase, residues of the four-residue segment are Glu-180, Lys-181, Trp-182 and Leu-183, and the other three residues are Met-154, Met-259 and Leu-269. The order of these corresponding residues are found to be the same. The positions of these residues in the tertiary structures are shown schematically in Figs. 4a and b. The residue first in sequence is in the domain having predominantly an anti-parallel β-sheet, i.e., the smaller domain in cAPK and the central domain in DD-ligase, respectively. The four-residue segment, next in sequence, is located from the end of the last β-strand of the first domain to the beginning of the successive loop connecting the two domains. The other two residues are in the next domain, i.e., the larger domain in cAPK and the C-terminal domain in DD-ligase, respectively. Although the arrangements of the seven residues are very similar, they are located on different secondary structural elements (Fig. 4a, b).

Further examination of the four-residue segment indicated the importance of the backbone part of the segment for adenine recognition. There is a significant structural similarity of the backbone part in the four-residue segments of cAPK and DD-ligase with a root-mean-square deviation, r.m.s.d., of 0.372 Å. Another key point is that the adenine moiety is bound by two hydrogen bonds to the backbone part of the four-residue segment. One is between N6 of the adenine ring and the backbone carbonyl of the second residue, and the other is between N1 and the backbone amide of the fourth residue; e.g., the backbone carbonyl of Lys-181 binds to N6 and the backbone amide of Leu-183 binds to N1 in DD-ligase (Fig. 5). The nitrogen atoms, N1 and N6, are characteristic of an adenine base. Therefore, the backbone part of the four-residue segment plays an es-

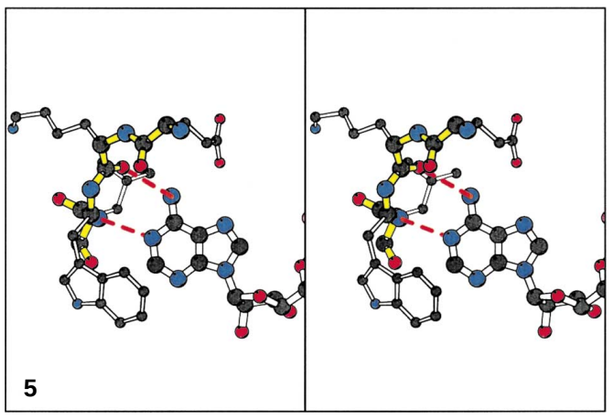

Fig. 5 Specific interactions between the adenine base and the four-residue segment in DD-ligase are shown. The adenine base is bound by two hydrogen bonds: one is between the N6 and the backbone carbonyl of Lys-181 (the second residue), and the other is between the N1 and the backbone amide of Leu-183 (the fourth residue). The conserved backbone structure of the segment is shown in yellow. This figure was drawn using the program MOLSCRIPT (Kraulis 1991)

Table 2 Amino acids constituting the two phosphate binding loop regions of proteins with the GSHase fold: DD-ligase, GSHase, BNC, SCS and PPDK, and of protein kinases: cAPK, CK1 and IRK

	Small loop	Large loop
DD-ligase	^{149}GSSVG	^{208}TFYDYEAKYLSDETQY
GSHase	^{164}GMGG	^{226}IPQGGETRGNLAAGR
BNC	^{165}GGRG	^{230}CSMQRRHQKV
SCS	^{52}GGRG	^{127}TEGGVEIEKVAEETPHLI
PPDK	^{100}PGMM	^{275}NAQGEDVVAGVRTP
	Gly-rich loop	**Catalytic loop**
cAPK	^{50}GTGSFG	^{164}YRDLKPEN
CK1	^{19}GEGSFG	^{129}YRDIKPEN
IRK	^{1003}GQGSFG	^{1130}HRDLAARN

sential role in distinguishing between adenine and other base moieties. (Because these two atoms participate in the Watson-Crick base pairs in the double-stranded DNA, this mode of adenine recognition should be clearly different from modes of double-stranded DNA-binding proteins.) In contrast, the side-chain parts are not involved in any specific interactions with the adenine moiety. These results suggest that only the backbone conformation is required for the adenine recognition and the role of side-chains is to stabilize and maintain the backbone conformation of this segment. A rather large amino acid sequence variation in this segment is tolerated, probably for this reason.

3.4 Similarities with other proteins with the ATP-grasp fold and with other protein kinases

As mentioned in the Introduction, GSHase, BNC, SCS and PPDK, as well as DD-ligase, are proteins with the ATP-grasp fold. In addition to the similarity of their chain folds, these proteins share another common feature that two loop regions are involved in phosphate binding. In DD-ligase, two loops consisting of residues 149–153 and of residues 208–224, the small and the large loops, respectively, cover the phosphate group of ATP (Fan et al. 1994) (Fig. 1b). The small loop is in the central domain and connects the sixth and seventh anti-parallel β-strands, and the large loop is in the C-terminal domain and connects the tenth and eleventh anti-parallel β-strands (Fig. 4b). In GSHase, two loops consisting of residues 164–167 and of residues 226–241 are thought to form the catalytic site (Tanaka et al. 1992; Tanaka et al. 1993). Fan et al. (1995) have suggested that these two loops in GSHase correspond to the two loops in DD-ligase. This correspondence has been confirmed in the structure determined recently (Hibi et al. 1996). In BNC, the ATP-binding site is also believed to occupy an approximately equivalent position to those of DD-ligase and of GSHase (Waldrop et al. 1994). In contrast, SCS, which phosphorylates its own histidine residue, has been thought to bind ATP at a different position, because the site of phosphorylation is far from the two loops. However, Matsuda et al. (1996) have suggested that SCS also binds ATP at the position corresponding to those in DD-ligase and GSHase, because the structure of the domains containing the corresponding two loops as well as the amino acid residues of the small loop in SCS are similar to those of the two proteins. Therefore, the two loops are likely to be involved in binding the phosphate group of ATP in the above for enzymes. Amino acid sequences of the two loops are listed in Table 2 for the five proteins with the ATP-grasp fold. (We assigned the large loops of BNC, SCS and PPDK by examining the topological correspondences.) Summarizing the above, the five proteins with the ATP-grasp fold, DD-ligase, GSHase, BNC, SCS and PPDK, are characterized by the similarities in the chain fold and the phosphate binding abilities of the two loops. Moreover, the structural similarity of the four-residue segment around the adenine moiety in DD-ligase and cAPK suggests the possibility that the four-residue segment might be a third common feature in proteins with the ATP-grasp fold.

Even through the structures of GSHase, BNC, SCS and PPDK deposited in the PDB (PDB code: 2glt, 1bnca, 1scub and 1dik, respectively) do not contain bound mononucleotide, the four-residue segment corresponding topologically to the segment of DD-ligase can be identified (amino acid sequences listed in Table 3). We found that each of the four-residue segments has very similar backbone conformation with a very low r.m.s.d. of 0.315 Å in GSHase, 0.459 Å in BNC, 0.309 Å in SCS, or 0.395 Å in PPDK, for superposition of backbone atoms to those of DD-ligase. Not only the four-residue segment but also the other three residues in the close neighborhead of the adenine moiety were also found (for the case of SCS shown in Fig. 6). The significant structural similarity in the backbone part strongly suggests that ATP is bound at the same site as that

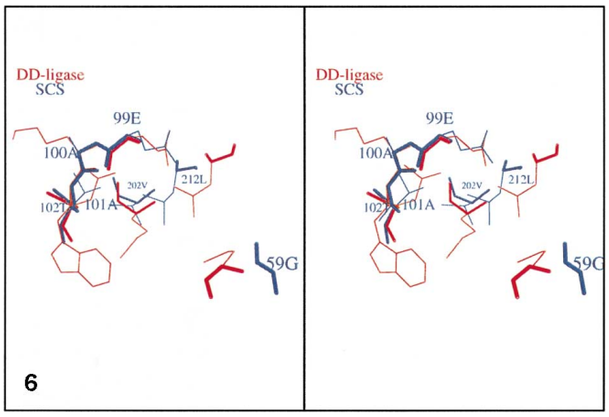

Fig. 6 Superposition of adenine-binding sites of DD-ligase (*red*) and SCS (*blue*). This figure was drawn using the program MOLSCRIPT (Kraulis 1991)

Table 3 Amino acids constituting the adenine binding sites of DD-ligase, cAPK and CK1, and the corresponding residues in GSHase, BNC, SCS, PPDK and IRK. The secondary structure assignments for DD-ligase and cAPK are also shown

	←B7	B8→loop	←B12→	←B13→
DD-ligase	^{154}M	^{180}EKWL	^{259}M	^{269}L
GSHase	^{170}I	^{198}QNYL	^{275}I	^{280}T
BNC	^{169}M	^{201}EKYL	^{278}L	^{287}I
SCS	^{59}G	^{99}EAAT	^{202}V	^{212}L
PPDK	^{105}T	^{240}QTMV	^{325}T	^{334}V
	←B2→	B5→loop	←B6→	B7→loop
cAPK	^{57}V	^{120}MEYV	^{173}L	^{183}T
CK1	^{26}I	^{85}IDLL	^{138}L	^{153}V
IRK	^{1010}V	^{1076}MELM	^{1139}M	^{1149}G

of DD-ligase in the other four enzymes. Whether with or without bound mononucleotide, the adenine-binding site is found to be very similar among all proteins with the ATP-grasp fold. The recently deposited GSHase-ADP complex structure (PDB code 1gsa) (Hara et al. 1996) again confirmed the rigidity of this segment. The local structures consisting of the seven residues in mononucleotide free- and ADP bound-GSHases were found to be almost identical with an r.m.s.d. of 0.076 Å for the superposition of backbone atoms of the four-residue segment. This finding indicates that the common local structures in proteins with the ATP-grasp fold form a rather rigid *lock* structure for the key of the adenine moiety.

We also examined the other protein kinase, IRK, that contains no bound mononucleotide in the PDB (PDB code 1irk). The local structure of IRK was also found to be conserved. The four-residue segment of IRK has an r.m.s.d. of 0.361 Å for superposition of the backbone atoms to those of cAPK.

In summary, we have found the common local structures around the adenine moiety in proteins with different chain folds, the ATP-grasp fold and the protein kinase fold. The same local structure is found even when there is no bound mononucleotide. The backbone part of the four-residue segment is found to play an important role in the recognition of the adenine moiety.

4. Discussion

As in the case of proteins with the ATP-grasp fold, protein kinases also contain two loops that are involved in binding of the phosphate group of the mononucleotide. One is called a glycine-rich loop containing a consensus sequence motif, GxGxxG, and the other is called a catalytic loop. As the protein kinases are closely related evolutionarily, amino acids in these loops are well conserved (Table 2). Locations of these loops are indicated in Fig. 4a. From the visual inspection of the three-dimensional structures (Fig. 1a, b), the Gly-rich loop and the catalytic loop in cAPK correspond roughly to the small loop and the large loop of DD-ligase, respectively (the Gly-rich loop and the small loop shown in blue; the catalytic loop and the large loop shown in yellow). Nevertheless, significant structural similarities were not found between the corresponding loops, and the Gly-rich loop and the corresponding small loop are in opposite directions, i.e., the Gly-rich loop comes from the end of the β-sheet while the small loop goes to the end of the β-sheet, respectively (Figs. 4a, b) (Matsuda et al. 1996). Moreover, the interactions between the two loops existing in DD-ligase were not found in cAPK. These differences in the phosphate-binding sites present a striking contrast to the adenine-binding sites that are found to be very similar. The proteins with the ATP-grasp fold and with the protein kinase fold bind ATP at the similar adenine-binding sites and at the somewhat dissimilar phosphate-binding sites. On the other hand, in the ATP- or GTP-binding proteins with the clinical mononucleotide-binding fold (Schulz 1992), the phosphate-binding sites containing the consensus G/AxxxxGKT/S sequence motif are very similar while the adenine- or guanine-binding sites are quite different. Even the sites for adenine base binding are different in adenylate kinase and recA protein (Story and Steitz 1992). Binding of the base and of the phosphate group are thus found to be independent of each other in these mononucleotide-binding proteins.

In spite of the high degree of structural similarity, amino acid sequences of the local structures are not conserved in proteins with the ATP-grasp fold or with the protein kinase fold (Table 3). As we have indicated, only the backbone conformation of the four-residue segment is required to recognize the adenine base, and thus the rather large amino acid sequence variation appears to be tolerated. This sequence variation makes it difficult to identify the conservation of the adenine-recognition segment even in evolutionarily related protein kinases. We have found the segment that is structurally conserved in the protein kinase

family of both serine/threonine kinases (cAPK and CK1) and a tyrosine kinase (IRK) by the structure search for the adenine-binding site.

5. Conclusion

We have developed a method of searching for similar spatial arrangements of atoms around a given chemical moiety. We applied this method to identify modes of adenine base recognition by mononucleotide-binding proteins. An all-against-all comparison of the arrangements of surrounding atoms around adenine moieties revealed an unexpected similarity between proteins with different folds, DD-ligase and cAPK, at their adenine-binding sites. The local structure common in DD-ligase and cAPK is found to be conserved in all the proteins in the PDB with the ATP-grasp or the protein kinase fold, even when the structure is in the substrate (mononucleotide) free state. This finding illustrates the third common feature in proteins with the ATP-grasp fold, and suggests strongly that ATP does actually bind in the same corresponding sites in these proteins. Taking the difference of their chain folds into consideration, we conclude that these proteins with the ATP-grasp fold and with the protein kinase fold are likely to be a good example of convergent evolution. Nature seems to have independently discovered the same recognition mechanism.

Among the common local structures around the adenine moiety, the four-residue segment shows a high degree of structural similarity in its backbone part. The backbone part of this segment plays an essential role in recognition of the adenine moiety, i.e., the backbone part binds the adenine moiety by two hydrogen bonds with the two atoms specific to the adenine base. We describe this four-residue segment as having a *structural motif* whose presence could not be detected by a sequence motif search among protein kinases.

Although this paper has focused on the study of adenine recognition, our method can also be applied to study the mechanism of recognition of other molecules. Such applications may reveal structural strategies adopted by proteins to perform various functions.

Acknowledgements We thank Dr. K. Mizuguchi for reading the manuscript carefully, and Ms. K. Matsuda and Dr. T. Nishioka for helpful discussions. This work has been suported by research grants from Ministry of Education, Science and Culture, Japan.

References

Artymiuk PJ, Poirrette AR, Rice DW, Willett P (1996) Biotin carboxylase comes into the fold. Nature Struct Biol 3:128–132

Belrhali H, Yaremchuk A, Tukalo M, Larsen K, Berthet-Colominas C, Leberman R, Beijer B, Sproat B, Als-Nielsen J, Grubel G, Legrand JF, Lehmann M, Cusack S (1994) Crystal structures at 2.5 Å resolution of seryl-tRNA synthetase complexed with two analogs of seryl adenylate. Science 263:1432–1436

Bernstein FC, Koetzle TF, Williams GJB, Meyer EF, Brice MD, Rodgers JR, Kennard O, Shimanouchi T, Tasumi M (1977) The Protein Data Bank: a computer based archival file for macromolecular structures. J Mol Biol 122:535–542

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^{2+} adenylyl imidodiphosphate and inhibitor peptide PKI (5-24). EMBO J 12:849–859

Brady L, Brzozowski AM, Derewenda ZS, Dodson E, Dodson G, Tolley S, Turkenburg JP, Christiansen L, Huge-Jensen B, Norskov L (1990)A serine protease triad forms the catalytic centre of a triacylglycerol lipase. Nature 343:767–770

Coxeter HSM (1961) Introduction to Geometry. Wiley, New York

Dever TE, Glynias MJ, Merrick WC (1987) GTP-binding domain: three consensus sequence elements with distinct spacing. Proc Natl Acad Sci USA 84:1814–1818

Fan C, Moews PC, Shi Y, Walsh CT, Knox JR (1995) A common fold for peptide synthetases cleaving ATP to ADP: glutathione synthetase and D-alanine:D-alanine ligase of Escherichia coli. Proc Natl Acad Sci USA 92:1172–1176

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

Finney JL (1975) Volume occupation, environment, and accessibility in proteins. The problem of the protein surface. J Mol Biol 96:721–732

Flaherty KM, McKay DB, Kabsch W, Holmes KC (1991) Similarity of the three-dimensional structures of actin and the ATPase fragment of a 70-kDa heat shock cognate protein. Proc Natl Acad Sci USA 88:5041–5045

Hanks SK, Quinn AM, Hunter T (1988) The protein kinase family: conserved features and deduced phylogeny of the catalytic domains. Science 241:42–52

Hara T, Kato H, Katsube Y, Oda J (1996) A pseudo-Michaelis quaternary complex in the reverse reaction of a ligase; structure of *Escherichia coli* B glutathione synthetase complexed with ADP, glutathione and sulfate at 2.0-Å resolution. Biochemistry (in press)

Herzberg O, Chen CC, Kapadia G, McGuire M, Carroll LJ, Noh SJ, Dunaway-Mariano D (1996) Swiveling-domina mechanism for enzymatic phosphotransfer between remote reaction sites. Proc Natl Acad Sci USA 93:2652–2657

Hibi T, Nishioka T, Kato H, Tanizawa K, Fukui T, Katsube Y, Oda J (1996) Structure of the multifunctional loops in the nonclassical ATP-binding fold of glutathione synthetase. Nature Struct Biol 3:16–18

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

Hurley JH, Fabar HR, Worthylake D, Meadow ND, Roseman S, Pettigrew DW, Remington SJ (1993) Structure of the regulatory complex of *Escherichia coli* III^{Glc} with glycerol kinase. Science 259:673–677

Kobayashi N, Go N (1996a) ATP binding proteins with different folds share a common ATP-binding structural motif. Nature Struct Biol (in press)

Kobayashi N, Go N (1996b) Mechanical property of a TIM-barrel protein. Proteins (in press)

Kraulis PJ (1991) MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 24:946–950

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 Engineering (in press)

Moras D (1992) Structural and functional relationships between aminoacyl-tRNA synthetase. Trends Biochem Sci 17:159–164

Murzin AG (1996) Structural classification of proteins: new superfamilies. Curr Opin Struct Biol 6:386–394

Pearl L (1993) Similarity of active-site structures. Nature 362:24

Pearl L, Blundell T (1984) The active site of aspartic proteinases. FEBS Lett 174:96–101

Richards FM (1974) The interpretation of protein structures: total volume, group volume distributions and packing density. J Mol Biol 82:1–14

Richards FM (1977) Areas, volumes, packing, and protein structure. Ann Rev Biophys Bioeng 6:151–176

Rould MA, Perona JJ, Steitz TA (1991) Structural basis of anticodon loop recognition by glutaminyl-tRNA synthetase. Nature 352:213–218

Schulz GE (1992) Binding of nucleotides by proteins. Curr Opin Struct Biol 2:61–67

Schulz GE, Schiltz E, Tomasselli AG, Frank R, Brune M, Wittinghofer A, Schirmer RH (1986) Structural relationships in the adenylate kinase family. Eur J Biochem 161:127–132

Story RM, Steitz TA (1992) Structure of the recA protein-ADP complex. Nature 355:374–376

Strynadka NC, Adachi H, Jensen SE, Johns K, Sielecki A, Betzel C, Sutoh K, James MN (1992) Molecular structure of the acyl-enzyme intermediate in β-lactam hydrolysis at 1.7 Å resolution. Nature 359:700–705

Sussman JL, Harel M, Frolow F, Oefner C, Goldman A, Toker L, Silman I (1991) Atomic structure of acetylcholinesterase from Torpedo californica: a prototypic acetylcholine-binding protein. Science 253:872–879

Swindells MG, Alexandrov NN (1994) Nucleotide binding in $\beta\alpha\beta$–$\beta\alpha\beta$ topologies. Nature Struct Biol 1:677–678

Tanaka T, Kato H, Nishioka T, Oda J (1992) Mutational and proteolytic studies on a flexible loop in glutathione synthetase from *Escherichia coli* B: the loop and arginine 233 are critical for the catalytic reaction. Biochemistry 31:2259–2265

Tanaka T, Yamaguchi H, Kato H, Nishioka T, Katsube Y, Oda J (1993) Flexibility impaired by mutations revealed the multifunctional roles of the loop in glutathione synthetase. Biochemistry 32: 12398–12404

Tanemura M, Ogawa T, Ogita N (1983) A new algorithm for three-dimensional Voronoi tessellation. J Comp Phys 51:191–207

Traut TW (1994) The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. Eur J Biochem 222:9–19

Waldrop GL, Rayment I, Holden HM (1994) Three-dimensional structure of the biotin carboxylase subunit of acetyl-CoA carboxylase. Biochemistry 33:10249–10256

Williams RL, Oren DA, Munoz-Dorado J, Inouye S, Inouye M, Arnold E (1993) Crystal structure of Myxococcus xanthus nucleoside diphosphate kinase and its interaction with a nucleotide substrate at 2.0 Å resolution. J Mol Biol 234:1230–1247

Wolodko WT, Fraser ME, James MN, 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 RM, 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, Hara Y, Nishioka T, Kumura A, Oda J, Katsube Y (1993) Three-dimensional structure of the glutathione synthetase from *Escherichia coli* B at 2.0 Å resolution. J Mol Biol 229:1083–1100