PROTEINS: Structure, Function, and Genetics 44:282–291 (2001)

Adenine Recognition: A Motif Present in ATP-, CoA-, NAD-, NADP-, and FAD-Dependent Proteins

Konstantin A. Denessiouk,[1] Ville-Veikko Rantanen,[1,2] and Mark S. Johnson[1]*
[1]*Department of Biochemistry and Pharmacy, Åbo Akademi University, Turku, Finland*
[2]*Department of Mathematics, University of Turku, Turku, Finland*

ABSTRACT **Adenosine triphosphate (ATP) plays an essential role in energy transfer within the cell. In the form of NAD, adenine participates in multiple redox reactions. Phosphorylation and ATP-hydrolysis reactions have key roles in signal transduction and regulation of many proteins, especially enzymes. In each cell, proteins with many different functions use adenine and its derivatives as ligands; adenine, of course, is present in DNA and RNA. We show that an adenine binding motif, which differs according to the backbone chain direction of a loop that binds adenine (and in one variant by the participation of an aspartate side-chain), is common to many proteins; it was found from an analysis of all adenylate-containing protein structures from the Protein Data Bank. Indeed, 224 protein–ligand complexes (86 different proteins) from a total of 645 protein structure files bind ATP, CoA, NAD, NADP, FAD, or other adenine-containing ligands, and use the same structural elements to recognize adenine, regardless of whether the ligand is a coenzyme, cofactor, substrate, or an allosteric effector. The common adenine-binding motif shown in this study is simple to construct. It uses only (1) backbone polar interactions that are not dependent on the protein sequence or particular properties of amino acid side-chains, and (2) nonspecific hydrophobic interactions. This is probably why so many different proteins with different functions use this motif to bind an adenylate-containing ligand. The adenylate-binding motif reported is present in "ancient proteins" common to all living organisms, suggesting that adenine-containing ligands and the common motif for binding them were exploited very early in evolution. The geometry of adenine binding by this motif mimics almost exactly the geometry of adenine base-pairing seen in DNA and RNA. Proteins 2001;44:282–291.** © 2001 Wiley-Liss, Inc.

Key words: adenine recognition; structural motif; ATP; CoA; NAD; FAD

INTRODUCTION

Adenine is a fundamental component of many key molecules in biology and is required for the function of many essential proteins common to all living creatures. At the same time, proteins that bind adenylate-containing nucleotides constitute a very large group of proteins and play a central role in the metabolism of the cell. Several sequence motifs for nucleotide binding have been identified,[1–5] but each of these motifs, with the exception of an adenine-binding motif found in a small group of ATP-binding proteins,[5] is responsible for binding a phosphate group, the sugar, or metal cations that coordinate the phosphates, and not the adenine moiety. For recognition of adenylate, it has been stated that there is probably no conserved hydrogen-bonding motif and the concept of a "fuzzy recognition template" was proposed.[6]

The first hint of a conserved adenine-binding motif was reported by Kobayashi and Go[7] as a four-residue loop that forms two hydrogen bonds with adenine of the ATP ligands bound to cyclic adenosine monophosphate (cAMP)-dependent protein kinase[8] (cAPK) and D-Ala:D-Ala ligase[9,10] (DD-ligase). In turn, we have identified a supersecondary structure organization amounting to 103 residues, shared by cAPK and DD-ligase,[11] showed that much of this super motif is present in the R1 subunit[12] of ribonucleotide reductase,[13] and found similarities for adenine recognition in 28% of the defined structures (deposited in the Protein Data Bank,[14] PDB) of ATP-binding proteins—12 different fold types.[15] The common adenine-binding motif includes a three-residue loop (adenine-binding loop) and an additional hydrophobic residue,[15] found consistently in the same position with respect to the adenine ring (Fig. 1). The loop forms two key hydrogen bonds with the N1 nitrogen and with the NH_2 group at the N6 position of the adenine ring. If the adenine-binding loop has the same backbone direction ["direct," Fig. 1(a)] as seen in cAPK [N-terminal → C-terminal, black molecule in Fig. 2(a,b)], the N1 nitrogen of adenine is hydrogen bonded to the backbone amide NH group of the third residue of the loop [designated site III, Fig. 1(a)], and the NH_2 group at position N6 is hydrogen bonded to the backbone carbonyl of the first residue of the loop (designated site I). When the adenine-binding loop has a chain direction opposite to this (the C-terminus corresponds to the N-terminus of cAPK and

Grant sponsor: Ministry of Education and the Graduate School of Informational and Structural Biochemistry; Grant sponsor: Technology Development Center of Finland; Grant sponsor: Academy of Finland; Grant sponsor: Erna and Victor Hasselblad Foundation.

*Correspondence to: Mark S. Johnson, Department of Biochemistry and Pharmacy, Åbo Akademi University, Tykistökatu 6, BioCity 3A, P.O. Box 66, FIN-20521 Turku, Finland. E-mail: johnson@abo.fi

Received 28 November 2000; Accepted 22 March 2001

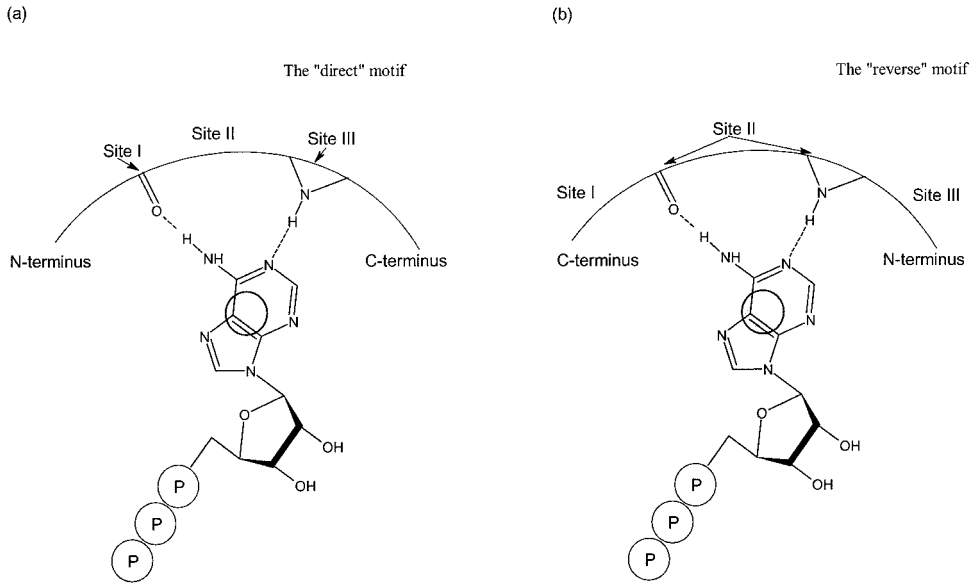

Fig. 1. Sketch of the (**a**) "direct" (N-terminus → C-terminus) and the (**b**) "reverse" (C-terminus → N-terminus) adenine-binding motifs found in many ATP-, CoA-, NAD-, NADP- and FAD-dependent proteins. The common adenine-binding motif includes a three-residue loop (sites I–III, adenine-binding loop) and an additional hydrophobic residue (gray circle), consistently found in the same positions with respect to the adenine ring. The loop forms two key hydrogen bonds with the N1 nitrogen and with the NH_2 group at the N6 position of the adenine ring. In the "direct" adenine-binding motif, the N1 nitrogen of adenine is hydrogen bonded to the backbone amide NH group of the third residue of the loop at site III, and the NH_2 group at position N6 is hydrogen bonded to the backbone carbonyl of the first residue of the loop at site I. In the "reverse" adenine-binding motif, the two hydrogen bonds to the adenine ring are both formed from the backbone amide NH group and the backbone carbonyl of a single residue at site II (middle residue) of the loop. The hydrophobic residue (gray circle) is positioned close to and behind the plane of the adenine ring. Drawn using the program ACD/ChemSketch v4.55, Advanced Chemistry Development, Toronto, Canada (http://www.acdlabs.com).

vice versa, the "reverse" orientation is present [Fig. 1(b)], for example, in aspartate carbamoyltransferase), the two hydrogen bonds to the adenine ring are both formed from the backbone amide NH group and the backbone carbonyl of a single residue at site II of the loop [gray molecule in Fig. 2(a,b)]. In both cases of the "direct" and "reverse" chain direction of the adenine-binding loop, the spatial position of the hydrogen bonds coincides [Fig. 2(a)]. A residue with a hydrophilic side-chain usually occupies site I of the loop, while residues at sites II and III are usually hydrophobic or nonpolar.

We have identified a unique adenine-binding motif in nearly one-third of the structurally defined ATP-dependent proteins and belonging to 12 different fold families.[15] The question arises whether this adenine-binding motif is present in other proteins that bind other adenylate-containing cofactors, such as nicotinamide adenine dinucleotide (NAD), nicotinamide adenine dinucleotide phosphate (NADP), flavin adenine dinucleotide (FAD), and coenzyme A (CoA), among others. In order to address this possibility, we have examined each of the available three-dimensional (3D) structures with bound adenine containing cofactors.

MATERIALS AND METHODS

The adenine ligand–protein complexes were taken from the August 2000 release of the PDB. The nucleotide-binding data set was constructed and analyzed as shown for ATP-dependent proteins.[15] Only those protein atoms that are strictly closer than 3.7 Å from any ligand atom were included in the analysis. To identify potential hydrogen bonds, the Ligand–Protein Contacts software[16] given on the web page for each PDB entry (http://www.rcsb.org/pdb/) and the molecular modeling software package SYBYL (Tripos Associates, St. Louis, MO) were used.

RESULTS AND DISCUSSION

CoA-Dependent Proteins

We have analyzed all available PDB entries containing CoA-like ligands, which include 24 different proteins. It was previously shown that some of the CoA-dependent proteins adopt topologies common to other proteins (e.g., the four-helix up-and-down bundle, the Rossmann fold, and the TIM barrel), while the others have unique folds (the fold of citrate synthase and the hydratase βα-spiral, among others).[17] CoA was shown to adopt either a bent or an extended conformation in these proteins. A common mode of CoA-binding has not been identified, suggesting that there are no conserved interactions between CoA-ligands and the proteins that bind them.[17] Nonetheless, we have found that 20 out of the 59 CoA–protein complexes in the PDB contain the adenine-binding loop that we previously identified in ATP-binding proteins. These 20 structures represent five different proteins and two different fold types (Table I): cold-active citrate synthase (PDB

code: 1A59) and citrate synthases from *Pyrococcus furiosus* (1AJ8), and chicken (representative structure, 1CSH). Two additional proteins form a second family with the βα-spiral fold: 2-enoyl-CoA hydratase (2DUB) and 4-Chlorobenzoyl-CoA dehalogenase (1NZY). In these five proteins, CoA is bound in the bent conformation, which has important implications for the structure of the ligand-binding site and on the binding motif itself (see below). 2-Enoyl-CoA hydratase and the 4-chlorobenzoyl-CoA dehalogenase have the "direct" form of the adenine-binding loop, like cAPK, where residues from sites I and III form hydrogen bonds to adenine (Table I). In contrast, all citrate synthases have the opposite "reverse" chain direction of the adenine-binding loop, and both hydrogen bonds are formed from the main-chain carbonyl oxygen and the main-chain amide NH group of the same residue located at site II.

With the exception of an aspartate present in 2-enoyl-CoA hydratase, the residue located at site II in CoA-dependent proteins is hydrophobic. Site III is hydrophobic or contains residues with side-chains that include long aliphatic parts, such as lysine. Site I is also hydrophobic, in contrast to the ATP-dependent proteins in which site I is hydrophilic. The hydrophobic residue at site I (Table I) is defined by the conformation of the ligand. Thus, when the structure of CoA and the attached substrate is bent into a U-like structure, the substrate is locked into position by one or two hydrogen bonds: between the N7 atom of the adenine ring and either the PN8 amide nitrogen and/or the PO10 hydroxyl oxygen of pantothenic acid [e.g., 4-chlorobenzoyl-CoA in 1NZY, Fig. 2(a)]. Consequently, the end of the ligand is close to the N6 atom of the adenine ring and to site I of the adenine-binding loop. As a result, in 1NZY, the ring of Phe64 (site I) is stacked parallel to the ring of the 4-chlorobenzoyl moiety of 4-chlorobenzoyl-CoA. In 2-enoyl-CoA hydratase, the hydrophobic methyl side-chain of Ala98 (site I) interacts equally with the aliphatic part of the β-mercaptoethylamine moiety and the octanoyl moiety of the CoA-ligand. In the cold-active citrate synthase and in the citrate synthase from *Pyrococcus furiosus*, the side-chain of methionine interacts directly with the aliphatic pantothenate moiety of coenzyme A, while in chicken citrate synthase site I is occupied by proline.

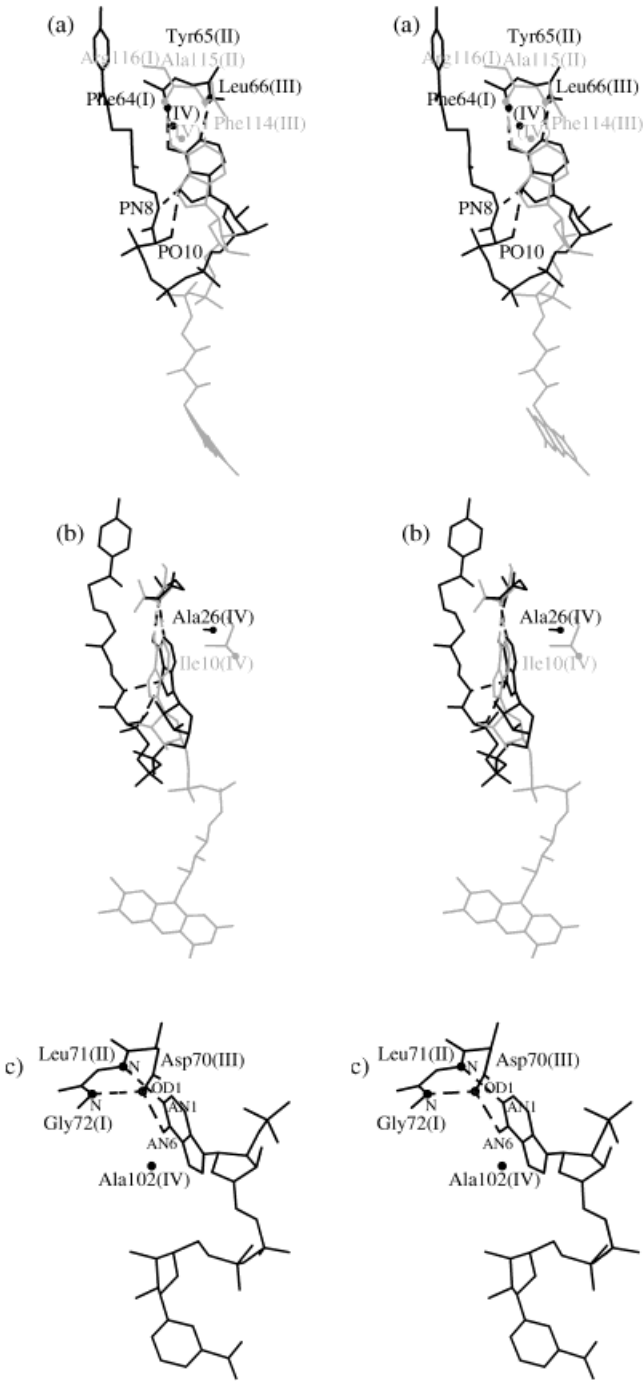

Fig. 2. **a,b:** Orthogonal views of the "direct" (CoA-dependent 4-chlorobenzoyl-CoA dehalogenase, 1NZY; black) and "reverse" (FAD-dependent glutathione reductase, 1GER; gray) orientations of the adenine-binding motif. Roman numerals in brackets indicate hydrophilic site I, aliphatic site III, and hydrophobic sites II and IV from Table I, respectively. **a:** Spatial position of the two key hydrogen bonds (dashed lines) coincides in both the "direct" and the "reverse" orientations of the adenine-binding motif. **b:** Site IV contains a hydrophobic residue consistently found close to the plane of the adenine ring. The bent conformation of CoA is stabilized by two hydrogen bonds between the N7 atom of the adenine ring and either the PN8 amide NH and/or the PO10 hydroxyl of pantothenic acid. **c:** Variation of the "reverse" adenine-binding motif found in many (56) NAD(P)-dependent short-chain alcohol dehydrogenases/reductases. In these proteins, the direction and conformation of the adenine-binding loop are in the "reverse" orientation of the adenine-binding motif; however, one of the two key hydrogen bonds is always formed through the side-chain oxygen atom of a bridging conserved aspartate (see Table II for the list of structures). Drawn using the program MOLSCRIPT[47] and rendered using Raster3D V2.3.[48]

FAD-Dependent Proteins

Several sequence fingerprints for FAD binding have been reported previously; however, none of them involves interactions with adenine. One is the well-known dinucleotide-binding motif (DBM), which consists of a βαβ structure, a part of the Rossmann fold.[18–20] DBM contains the characteristic GxGxxG sequence in the loop between the first β-strand and the α-helix (G, glycine; x, any residue), as well as an aspartate or glutamate at the end of the

second β-strand that is bound to ribose. Another motif in FAD-binding proteins is the TxxxxhϕhhGD amino acid fingerprint (T, threonine; D, aspartate; h, a nonpolar residue; ϕ, any aromatic residue). This motif contains a conserved aspartate, which forms hydrogen bonds with the O3 group of the flavin moiety of FAD.[21,22] In flavin-containing hydrolases a hydrophobic sequence motif, (F/A)TGY (F, phenylalanine; A, alanine; Y, tyrosine), has been observed,[23] while the RxxGGRxx(S/T) "GG doublet" motif (R, arginine; S, serine) was found near DBM and close to the ribose of FAD in L-amino acid oxidases, fumarate reductases, and a wide variety of other flavoprotein families.[24]

For the first time, we show that an adenine-binding motif exists in a large group of FAD-dependent proteins. Of 157 structures in the PDB that contain FAD as a ligand, 86 have the adenine-binding motif. The 86 PDB files include the 17 different proteins listed in Table I. In the first 13 groups, the FAD-binding domains are Rossmann folds, contain the DBM motif, and form several protein families according to their biological function. Glutathione reductases, thioredoxin reductases, trypanothione reductases, and dihydrolipoamide dehydrogenases (groups 3–6) form a family of closely related FAD-dependent disulfide oxidoreductases.[25] Adrenodoxin reductase (group 1), flavocytochrome C sulfide dehydrogenase (group 7), and the NADH peroxidases (group 8), reported to share 15–40% sequence identity, have significant similarities in the organization of their FAD-binding domain.[26–28] D-Amino acid oxidases, cholesterol oxidases, and glucose oxidases form another family of structurally related flavoenzymes known as GMC oxidoreductases;[29,30] and the folding topologies of polyamine oxidases and the monomeric sarcosine oxidase (groups 12 and 13) are similar to the GMC oxidoreductase family.[31,32] Groups 14–17 in Table I have the α+β FAD-binding fold of "covalent flavoproteins,"[33,34] but they do not have the DBM fingerprint. Instead, a flavin-binding motif is present in which flavin is covalently linked to the protein.[34]

Interestingly, independent of folds or families, all 86 FAD-dependent proteins have the "reverse" chain direction motif of the adenine-binding loop [Fig. 1(b)]. Hence, hydrogen bonds with the N1 and N6 nitrogens of adenine are formed by the carbonyl oxygen and amide nitrogen of the residue at hydrophobic site II (Table I), occupied mainly by alanine or valine, and in some cases by glycine, leucine, and isoleucine.

Of special interest is "hydrophobic" site IV (Table I), which is mainly occupied by valine and isoleucine in all FAD-dependent proteins. In those FAD-dependent proteins that are Rossmann folds, site IV corresponds to the residue adjacent to the first glycine of the GxGxxG fingerprint in the DBM motif. Structurally, site IV is positioned close to the adenine ring, with its hydrophobic side-chain perpendicular to the plane of the ring. The importance of aliphatic-aromatic interactions in packing of the adenine ring in proteins has been noted previously.[35] This type of interaction facilitates the proper positioning of the adenine moiety in the binding pocket, and we have observed this same type of interaction at site IV in the large group of unrelated ATP-dependent proteins.[15]

NAD- and NADP-Dependent Proteins

NADP and NAD differ by the presence of the phosphate group at the 2′ position of the adenylate ribose in NADP. The total number of NAD- and NADP-containing complexes in the PDB is 261. Many of these complexes represent NADP-dependent dehydrogenases and reductases, where the core of the coenzyme-binding domain has the Rossmann fold structure.[36] One would expect to see the adenine-binding motif in this case, since it was found in many FAD-dependent dehydrogenases and reductases with the Rossmann fold (families 1–13 in Table I). Surprisingly, only four proteins with two different folds do have the adenine-binding motif similar to that seen in cAPK (Table I): exotoxin A, domain III (1AER) and diphtheria A, domain III (1TOX) , as well as bacterial isocitrate dehydrogenases (1AI2) and 3-isopropylmalate dehydrogenase (1HEX). Interestingly, all four structures are NAD-dependent, and none has the Rossmann fold structure in their dinucleotide-binding domains.[37–42]

When we compared the other NAD- and NADP-dependent proteins, we have found that 56 structures, all members of the short-chain alcohol dehydrogenase/reductase (SDR) family also known as the single-domain reductases/epimerases/dehydrogenases (the "RED") family,[43] have the adenine-binding motif.

The core of the NAD(P)-binding domains in SDR proteins are typical Rossmann folds. The SDR proteins have the common GxxxGxG phosphate-binding pattern of glycine residues and the ribose-binding aspartate.[44] In addition, SDR proteins contain the YxxxK (K, lysine) motif that binds the nicotinamide ribose and a conserved serine residue that is part of the SDR active site.[45,46] These and several other amino acid signatures have been described in detail for the SDR protein family.[43] The SDR family show an interesting variation of the "reverse" adenine-binding motif [Fig. 2(c), Table II]. SDR proteins [i.e., sepiapterin reductases 1OAA, Fig. 2(c)] have the usual hydrogen bond formed between the main-chain nitrogen of the residue at site II and the N1 nitrogen of the adenine ring of NAD(P). The residue at site III is always an aspartate, except for two cases (GDP-fucose synthetase and trihydroxynaphthalene reductase where aspartate is replaced by an asparagine (Table II). The second characteristic hydrogen bond is, however, not formed between the N6 nitrogen of the adenine and the main-chain oxygen of the same residue from site II. Instead, it is formed between the N6 nitrogen of the adenine and the side-chain oxygen of aspartate (or asparagine) located at site III [Fig. 2(c)]. This oxygen of aspartate mediates the connection to the main-chain nitrogen of the residue at site I. In SDR proteins, site I is mostly hydrophilic or polar, and site II is strictly hydrophobic (Table II).

Other Adenylate-Binding Proteins

Besides ATP, ADP, AMP, CoA, FAD, NADP, or NAD, many crystal structures were found to contain bound

TABLE I. List of Consensus Interactions Between Nucleotide-Binding Site and Adenine for Different Adenine-Containing Ligands

Protein name, PDB code, chain	Source	Ligand	Site I	Site II	Site III	Hydrophobic site IV
CoA-dependent proteins						
1. Citrate synthases:	Antarctic bacterium DS2-3R	CoA	Met265	Val264 (N1, N6)	Lys263	Ile315
1.1. 1A59						
1.2. 1AJ8	*Pyrococcus furiosus*	CoA	Met258	Ile257 (N1, N6)	Lys256	Ile307
1.3. 1CSH,[a] 1AL6, 1AMZ, 1CSC, 1CSI, 1CSR, 1CSS, 2CSC, 2CTS, 3CSC, 3CTS, 4CSC, 5CTS, 6CSC, 6CTS	*Gallus gallus*	AMX	Pro316	Val315 (N1, N6)	Val314	Leu361
2. 2-Enoyl-CoA hydratase: 2DUB,[a] 1DUB	*Rattus norvegicus*	CO8	Ala98 (N6)	Asp99	Ile100 (N1)	Ala60
3. 4-Chlorobenzoyl-CoA dehalogenase: 1NZY	*Pseudomonas*	BCA	Phe64 (N6)	Tyr65	Leu66 (N1)	Ala26
FAD-dependent proteins						
1. Adrenodoxin reductase: 1CJC	*Bos taurus*	FAD	Gly83	Val82 (N1, N6)	Glu81	Val12
2. Fumarate reductases:	*Escherichia coli*	FAD	Leu158	Val157 (N1, N6)	Phe156	Val10
2.1. 1FUM (except M chain)						
2.2. 1D4D,[a] 1D4C, 1D4E	*Shewanella putrefaciens*	FAD	Val278	Val277 (N1, N6)	Arg276	Ile131
2.3. 1QJD	*Shewanella frigidimarina*	FAD	Ile279	Gly278 (N1, N6)	Arg277	Val132
2.4. 1QLA,[a] 1QLB (A-, and D-chains)	*Wolinella succinogenes*	FAD	Ile182	Ala181 (N1, N6)	Glu180	Ile11
3. Glutathione reductases:	*Homo sapiens*	FAD	Ala131	Ala130 (N1, N6)	His129	Ile26
3.1. 3GRS,[a] 1BWC, 1DNC, 1GSN, 1XAN, 1GRA, 1GRB, 1GRE, 1GRF, 1GRG, 4GR1, 1GRT, 2GRT, 3GRT, 4GRT, 5GRT						
3.2. 1GER,[a] 1GES, 1GET, 1GEU	*Escherichia coli*	FAD	Arg116	Ala115 (N1, N6)	Phe114	Ile10
4. Thioredoxin reductases:	*Escherichia coli*	FAD	Asn85	Ile84 (N1, N6)	His83	Leu11
4.1. 1TRB,[a] 1CL0, 1TDE, 1TDF						
4.2. 1VDC	*Arabidopsis thaliana*	FAD	Thr85	Val84 (N1, N6)	Thr83	Val11
5. Trypanothione reductases:	*Trypanosoma cruzi*	FAD	Ser129	Gly128 (N1, N6)	Trp127	Ile11
5.1. 1AOG,[a] 1BZL, 1NDA						
5.2. 1FEC,[a] 1FEA (except A-chain), 1FEB, 1TYP, 1TYT, 2TPR						
6. Dihydrolipoamide dehydrogenases:	*Neisseria meningitidis*	FAD	Gln234	Gly233 (N1, N6)	Asp232	Leu128
6.1. 1OJT,[a] 1BHY						
6.2. 1EBD	*Bacillus stearothermophilus*	FAD	Tyr120	Ala119 (N1, N6)	Glu118	Val15
6.3. 1LPF	*Pseudomonas fluorescens*	FAD	Lys122	Gly121 (N1, N6)	His120	Ile9
6.4. 1LVL	*Pseudomonas putida*	FAD	Lys119	Ala118 (N1, N6)	Trp117	Ile11
6.5. 3LAD	*Azotobacter vinelandii*	FAD	Lys122	Gly121 (N1, N6)	His120	Ile9
7. Flavocytochrome C sulfide dehydrogenase: 1FCD	*Chromatium vinosum*	FAD	Thr78	Ala77 (N1, N6)	Ser76	Val8
8. NADH peroxidases: 1NHP,[a] 1JOA, 1NHQ, 1NHR, 1NHS, 1NPX, 2NPX.	*Enterococcus faecalis*	FAD	Thr80	Ile79 (N1, N6)	Glu78	Leu6
9. D-Amino acid oxidases: 1AN9,[a] 1AA8, 1DAO, 1DDO, 1KIF	*Sus scrofa*	FAD	Glu165	Val164 (N1, N6)	Lys163	Ile6
10. Cholesterol oxidases:	*Streptomyces*	FAD	Lys251	Val250 (N1, N6)	Gln249	Ile16
10.1. 1B4V,[a] 1B8S, 1CBO, 1CC2						
10.2. 1COY,[a] 3COX	*Brevibacterium sterolicum*	FAD	Thr251	Val250 (N1, N6)	Arg249	Ile17
11. Glucose oxidases:	*Aspergillus niger*	FAD	Gly251	Val250 (N1, N6)	Tyr249	Ala25
11.1. 1CF3,[a] 1GAL						
11.2. 1GPE	*Penicillium amagasakiense*	FAD	Gly255	Val254 (N1, N6)	Met253	Ala30
12. Polyamine oxidases: 1B5Q,[a] 1B37	*Zea mays*	FAD	Arg238	Val237 (N1, N6)	Val236	Val10

TABLE I. (Continued)

Protein name, PDB code, chain	Source	Ligand	Site I	Site II	Site III	Hydrophobic site IV
13. Sarcosine oxidase: 1B3M	*Bacillus*	FAD	Glu174	Val173 (N1, N6)	Agr172	Val9
14. Carbon monoxide dehydrogenase: 1QJ2 (C-chain and I-chain)	*Pseudomonas carboxydovorans*	FAD	Thr168	Leu167 (N1, N6)	Leu166	Ala117
15. *p*-Cresol methylhydroxylases: 1DII,[1] 1DIQ	*Pseudomonas putida*	FAD	Thr232	Cys231 (N1, N6)	Ile230	Met166
16. Vanillyl-alcohol oxidases: 1VAO,[a] 1AHU, 1AHV, 1AHZ, 1QLT, 1QLU, 2VAO	*Penicillium simplicissimum*	FAD	Thr263	Val262 (N1, N6)	Ile261	Val181
17. Uridine diphospho-*N*-acetylpyruvylglucosamine reductases: 2MBR,[a] 1MBB, 1MBT, 1UXY	*Escherichia coli*	FAD	Val174	Ile173 (N1, N6)	Ala172	Ile119
NAD(P)(H)-dependent proteins						
1. Exotoxin A, domain III: 1AER	*Pseudomonas aeruginosa*	TAD	Gly454 (N6)	Val455	Arg456 (N1)	Ile450
2. Diphtheria toxin: 1TOX	*Candida albicans*	NAD	Gly34 (N6)	Ile35	Gln36 (N1)	Ile31
3. Isocitrate dehydrogenases: 1AI2,[a] 1AI3, 1BL5, 1IDE, 1ISO, 9ICD	*Escherichia coli*	NAD	Pro353	Asn352 (N1, N6)	Val351	Gly321
4. 3-Isopropylmalate dehydrogenase: 1HEX	*Thermus thermophilus*	NAD	Pro287	Asn286 (N1, N6)	Ala285	Gly255
Other adenylate-dependent proteins:						
cAMP-dependent protein kinase: 1FMO (standard direct motif)	*Mus musculus*	NAD	Glu121	Tyr122	Val123	Ala70
Diphtheria toxin: 1DDT,[a] 1MDT (standard direct motif)	*Corynebacterium diphtheriae*	APU	Gly34	Ile35	Gln36	Ile31
Momorcharins: 1AHA,[a] 1AHB(FMP) (standard reverse motif)	Bitter melon (*momordica charantia*) seeds	ADE	Met72	Ile71	Tyr70	Ile155
Trichosanthin: 1MRK,[a] 1MRJ(ADN) (standard reverse motif)	Tian hua fen (*cucurbitaceae trichosanthes kirilowii maxim*) root tuber	FMC	Met72	Ile71	Tyr70	Ile155
Chemotaxis receptor methyltransferase: 1AF7,[a] 1BC5 (standard reverse motif)	*Salmonella typhimurium*	SAH	Leu214	Leu213	Asn212	Val232
Cobalt precorrin-4-transmethylase: 1CBF,[a] 2CBF (standard reverse motif)	*Bacillus megaterium*	SAH	Thr214	Ala213	Lys212	Ser132
Poly(A) polymerase regulatory subunit: 1V39,[a] 1B42, 3MCT, 2VP3, 1BKY, 1P39, 1EQA, 1VP9, 1VP3, 1EAM, 1VPT(SAM), 1AV6, 3MAG, 4DCG (standard reverse motif)	*Vaccinia virus*	SAH	Asp117	Val116	Phe115	Val139
Pokeweed antiviral protein: 1QCI,[a] 1PAG(FMP)	*Phytolacca americana*	ADE	Met74	Val73	Tyr72	Ile171
Ricin: 1IFS,[a] 1IFU(FMC) (standard reverse motif)	*Ricinus communis*	ANE	Val82	Val81	Tyr80	Ile172
Adenine-*N*6-DNA-methyltransferases Taqi: 2ADM[a] (except B-chain), 1AQI(SAH), 1AQJ(SFG) (Asp-motif)	*Thermus aquaticus*	SAM	Leu91	Phe90	Asp89	Ala47
Adenine-specific methyltransferase Hhai: 2DPM (Asp-motif)	*Streptococcus pneumoniae*	SAM	Glu179	Phe178	Asp177	Phe43
Cytosine-specific methyltransferase Hhai: 6MHT[a](SAM), 10MH, 1MHT, 3MHT, 4MHT, 5MHT, 7MHT, 8MHT, 9MHT, 2HMY(SAM) (Asp-motif)	*Haemophilus haemolyticus*	SAH	Thr62	Ile61	Asp60	Phe18
*N*4-cytosine-specific methyltransferase PvuII: 1BOO (Asp-motif)	*Proteus vulgaris*	SAH	Leu36	Ser35	Asp34	Phe273
Glycine-*N*-methyltransferase: 1XVA (except B-chain) (Standard reverse motif), 1D2H (except D-chain) (Asp-motif)	*Escherichia coli*	SAM	Pro187	Pro188	Gly189	Tyr220
	Rattus norvegicus	SAH	Leu118	Trp117	Asn116	—
Adenine phosphoribosyltransferase: 1QB7,[a] 1QB8(AMP)	*Leishmania donovani*	ADE	Asp44	Ala43	Phe42	Val148

PDB, Protein Data Bank; AND, adenosine; ADE and ANE, adenine; APU, adenylyl-3′-5′-phospho-uridine-3′-monophosphate; SAH, S-adenosyl-L-homocysteine; SAM, S-adenosylmethionine; FMC, formycin.

[a]Among several available structures of the same protein from different organisms, ligand interactions are given for the one indicated.

TABLE II. Aspartate Variation of the Adenine-Binding Motif in NAD(P)-Binding Proteins of the SDR Family

Protein name, PDB code, chain, resolution	Source	Ligand	Site I	Site II (N1)	Site III (N6)	Hydrophobic site IV
1. 17β-Hydroxysteroid-dehydrogenases: 1A27,[a] 1EQU (but B-chain), 1FDT, 1FDU, 1FDV (but B-chain)	*Homo sapiens*	NAP	Arg67	Val66	Asp65	Ala91
2. 7α-Hydroxysteroid-dehydrogenases: 1FMC,[a] 1AHH, 1AHI	*Escherichia coli*	NAD	Thr70	Ile69	Asp68	Ala96
3. 3α-20β-Hydroxysteroid-dehydrogenases: 2HSD (but B chain)	*Streptomyces hydrogenans*	NAD	Thr62	Val61	Asp60	Ala88
4. *cis*-Biphenyl-2,3-dihydrodiol-2,3-dehydrogenase: 1BDB	*Pseudomonas*	NAD	Arg61	Val60	Asp59	Ala87
5. Glucose 6-phosphate dehydrogenase: 2DPG	*Leuconostoc mesenteroides*	NAP	Thr87	Val86	Asp85	Val118
6. Alcohol dehydrogenases: 1B16,[a] 1B14, 1B15, 1B2L	*Drosophila lebanonensis*	NAQ	Thr65	Val64	Asp63	Val92
7. Tropinone reductase-I: 1AE1	*Datura stramonium*	NAP	Leu80	Leu79	Asp78	Ala107
8. Tropinone reductase-II: 2AE2	*Datura stramonium*	NAP	Ser68	Leu67	Asp66	Ala95
9. Enoyl acyl carrier (ACP) protein reductases: 1QG6,[a] 1D8A, 1DFG, 1DFH, 1DFI, 1QSG	*Escherichia coli*	NAD	Ala66	Val65	Asp64	Ile92
10. Enoyl-ACP reductases: 1D7O,[a] 1CWU, 1ENO, 1ENP	*Brassica napus*	NAD	Val91	Ala90	Asp89	Leu137
11. Enoyl-ACP reductases: 1ENY,[a] 1ENZ, 1BVR (but E chain), 1ZID	*Mycobacterium tuberculosis*	NAD	Gln66	Val65	Asp64	Ile95
12. Carbonyl reductase: 1CYD	*Mus musculus*	NAP	Gly62	Leu61	Asp60	Ala84
13. Sepiapterin reductase: 1OAA,[a] 1NAS, 1SEP	*Mus musculus*	NAP	Gly72	Leu71	Asp70	Ala102
14. Trihydroxynaphthalene reductase: 1YBV	*Magnaporthe grisea*	NDP	Gly89	Val88	Asn87	Ser115
15. UDP-galactose 4-epimerases: 1UDB,[a] 1A9Y, 1A9Z, 1KVQ, 1KVR, 1KVS, 1KVT, 1KVU, 1NAH, 1NAI, 1UDA, 1UDC, 1XEL, 2UDP	*Escherichia coli*	NAD	Arg60	Ile59	Asp58	Ala81
16. GDP-fucose synthetases: 1BSV,[a] 1FXS and GDP-4-keto-6-deoxy-D-mannose epimerase/ reductase 1BWS	*Escherichia coli*	NAP	Leu42	Leu41	Asn40	Ala63
17. dTDP glucose 4,6-dehydratase: 1BXK	*Escherichia coli*	NAD	Cys61	Ile60	Asp59	Ala82
18. Sulfolipid biosynthesis (SqdI) protein: 1QRR	*Arabidopsis thaliana*	NAD	Cys77	Ile76	Asp75	Gly98
19. Deoxyhypusine synthase: 1DHS	*Homo sapiens*	NAD	Ser344	Ala343	Asp342	Gly282

[a]Among several available structures of the same protein from different organisms, ligand interactions are given for the one indicated.

adenine and other adenine-containing ligands, for example, S-adenosyl-L-homocysteine (SAH) or bound adenine analogues such as formycin-5′-monophosphate. Sixteen proteins (46 structures) have the adenine-binding motif. Among the 16, three proteins have the adenine-binding loop with the same chain direction seen in cAPK ["direct" motif, Fig. 1(a)]; eight proteins have the reverse chain direction ["reverse" motif, Fig. 1(b)]; and five proteins have the aspartate modification of the "reverse" motif [Fig. 2(c)] seen in NAD-dependent proteins with the Rossmann fold. These 16 proteins are listed in Table I.

A Simple Binding Motif for Adenine

The elegant example of adenine binding to adenine phosphoribosyltransferase demonstrates that the simple motif described here is very likely responsible for recognition and binding of adenine. The structure of adenine phosphoribosyltransferase has been solved with two bound ligands, adenine (1QB7) and AMP (1QB8). The ligand-binding site in adenine phosphoribosyltransferase has a unique symmetrical adenine-binding motif (Fig. 3). It contains a 4-residue adenine-binding loop, Arg41–Phe42–Ala43–Asp44 (a "hydrophilic–hydrophobic–hydrophobic–hydrophilic" structure), and two symmetrically placed hydrophobic residues, Val39 and Val148, which sandwich the adenine moiety from both sides. Initially, it was surprising to find that the ligands in 1QB7 and 1QB8 are bound such that the adenine rings are flipped 180° with respect to each other within the same binding site (Fig. 3). On visual inspection, however, adenine in 1QB7 was seen to bind to the main-chain carbonyl oxygen and amide nitrogen of Ala43, the standard "reverse" chain-direction motif, where hydrophobic site II corresponds to Ala43 in the loop Asp44–Ala43–Phe42 and site IV is Val148. In contrast, AMP in 1QB8 binds to the carbonyl oxygen of Arg41 and the main-chain nitrogen of Ala43 of the sequence Arg41–Phe42–Ala43, the "direct" chain-direction motif in which hydrophobic site II corresponds to Phe42 and site IV is Val39. If the structures 1QB7 and 1QB8 are superimposed using the atoms of the two adenine rings, then Val39 and Val148 at hydrophobic site IV are superposed too. This unique symmetrical ligand-binding site contains both the "direct" and "reverse" motifs within a single protein and at the same location, allowing the possibility for adenine to bind in either orientation.

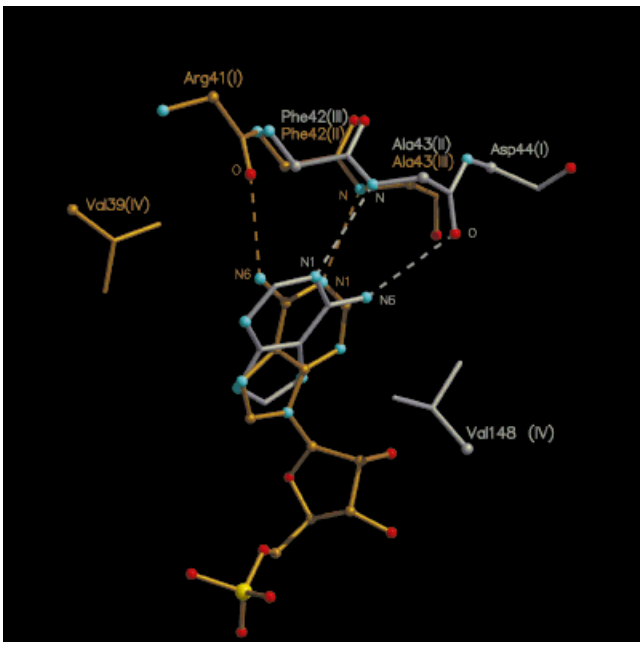

Fig. 3. Two structures of the same protein: adenine phosphoribosyltransferase 1QB7 (gray) and 1QB8 (yellow) are solved with bound adenine and AMP, respectively. The ligand-binding site in phosphoribosyltransferase has a unique symmetrical structure, which contains both the "direct" (yellow) and "reverse" (gray) adenine-binding motifs within a single protein, giving the possibility for adenine to bind in either orientation. Drawn using the program MOLSCRIPT[47] and rendered using Raster3D V2.3.[48]

When protein–ligand complexes containing adenine ligands are superimposed over their adenine rings, then we clearly see the clustering of nearby atoms from the protein binding site (Fig. 4): main-chain carbonyl oxygens (red) that interact with the N6 position of the adenine ring and the main-chain amide nitrogens (cyan) that hydrogen bond with the N1 position of adenine [Fig. 4(a), right side; 4(b), central clusters]. Hydrophobic side-chain carbon atoms are shown in white and cluster on either face of the adenine ring, which agrees with previous findings.[6] Site IV is located above and parallel to the adenine ring [Fig. 4(a); right side in Fig. 4(b)]. A striking feature of the motif is the similarity in binding, chemical and geometrical, shared with adenine base-pairing in DNA and RNA.

All told, given that a common motif is found in many different fold types, it is possible that similar adenine-binding sites in these folds arose independently on different occasions rather than from a more prosaic divergent mechanism. The common adenine-binding motif shown in this study is simple to construct. It uses only (1) backbone polar interactions that are not dependent on the protein sequence or particular properties of amino acid side-chains, and (2) nonspecific hydrophobic interactions. This is probably why so many different proteins with different functions use this motif to bind an adenylate-containing ligand. The motif common to many proteins and described in this study is not, however, the only way in which proteins bind adenine-containing ligands. Thus, although

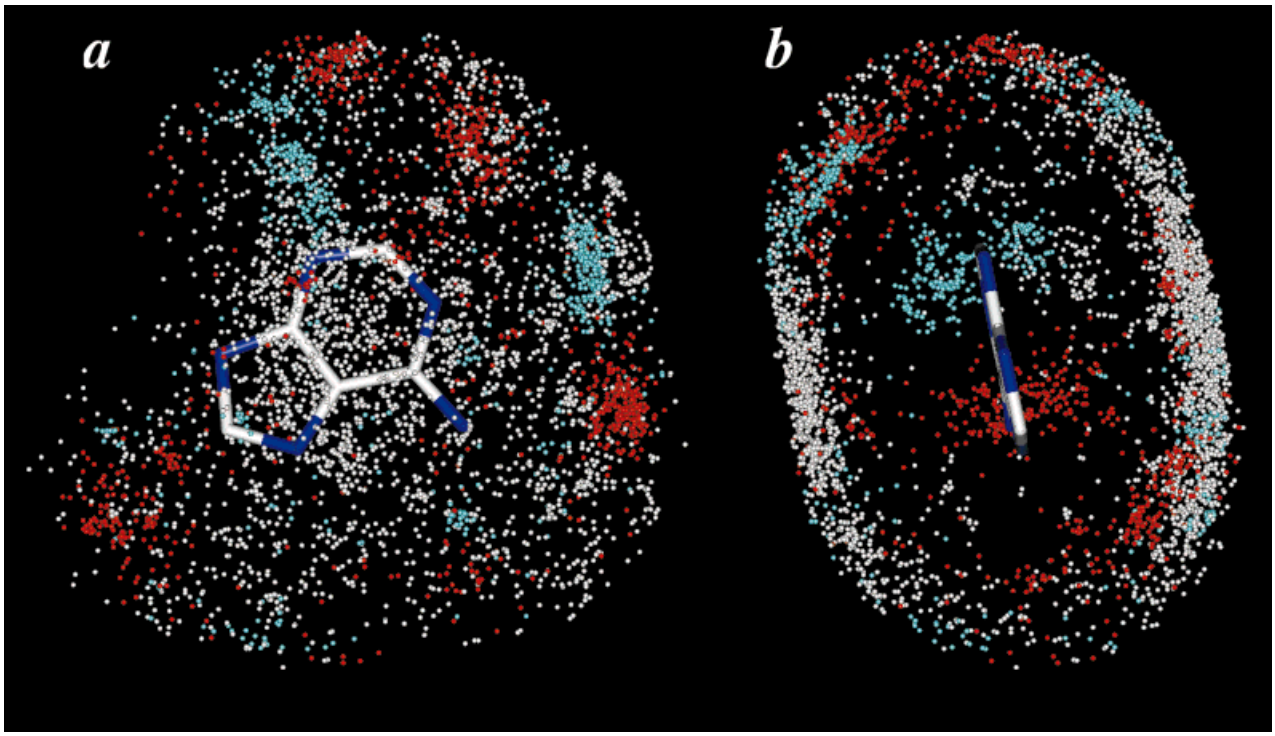

Fig. 4. Adenine rings of 540 protein–ligand complexes containing adenine ligands were superposed on top of each other and atoms within 4.0 Å of the ring are displayed. Only the main-chain oxygen (red) and nitrogen (cyan) atoms, and hydrophobic side-chain carbon atoms (white) are shown. Orthogonal views where the N1 nitrogen (upper) and N6 nitrogen (lower) are located. **a:** Right side of the adenine ring. **b:** Rear view. The ring is sandwiched between clusters of hydrophobic atoms; hydrophobic site IV is located parallel to and above the adenine ring in **a** and on the right-hand side in **b.** The software used in this analysis will be described elsewhere.

a wide variety of different folds bind adenine ligands in a similar way, many other proteins use alternative motifs. Nevertheless, the adenylate-binding motif reported in this article is present in "ancient proteins" common to all living organisms, suggesting that adenine-containing ligands and the common motif for binding them were exploited very early in evolution. The geometry of adenine binding by this motif mimics almost exactly the geometry of adenine base-pairing seen in DNA and RNA.

ACKNOWLEDGMENTS

Our colleague, Jukka V. Lehtonen, kindly provided several of the subroutines used in the analysis of the structures in this work.

REFERENCES

1. Walker JE, Saraste M, Runswick MJ, Gay NJ. Distantly related sequences in the α- and β-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. EMBO J 1982;1:945–951.
2. Schulz GE. Binding of nucleotides by proteins. Curr Opin Struct Biol 1992;2:61–67.
3. Traut TW. The functions and consensus motifs of nine types of peptide segments that form different types of nucleotide-binding sites. Eur J Biochem 1994;222:9–19.
4. Smith CA, Rayment I. Active site comparisons highlight structural similarities between myosin and other P-loop proteins. Biophys J 1996;70:1590–1602.
5. Krell T, Coggins JR, Lapthorn AJ. The three-dimensional structure of shikimate kinase. J Mol Biol 1998;278:983–997.
6. Moodie SL, Mitchell JBO, Thornton JM. Protein recognition of adenylate: an example of a fuzzy recognition template. J Mol Biol 1996;263:486–500.
7. Kobayashi N, Go N. ATP binding proteins with different folds share a common ATP-binding structural motif. Nature Struct Biol 1997;4:6–7.
8. Bossemeyer D, Engh RA, Kinzel V, Ponstingl H, Huber R. 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 1993;12:849–859.
9. Fan C, Moews PC, Walsh CT, Knox JR. Vancomycin resistance: structure of D-alanine: D-alanine ligase at 2.3 Å resolution. Science 1994;266:439–443.
10. Fan C, Park IS, Walsh CT, Knox JR. D-alanine: D-alanine ligase: phosphonate and phosphinate intermediates with wild type and the Y216F mutant. Biochemistry 1997;36:2531–2538.
11. Denessiouk KA, Lehtonen JV, Korpela T, Johnson MS. Two "unrelated" families of ATP-dependent enzymes share extensive structural similarities about their cofactor binding sites. Protein Sci 1998;7:1136–1146.
12. Eriksson M, Uhlin U, Ramaswamy S, Ekberg M Regnström K, Sjöberg B-M, Eklund H. Binding of allosteric effectors to ribonucleotide reductase protein R1: reduction of active-site cysteines promotes substrate binding. Structure 1997;5:1077–1092.
13. Denessiouk KA, Lehtonen JV, Johnson MS. Enzyme–mononucleotide interactions: three different folds share common structural elements for ATP recognition. Protein Sci 1998;7:1768–1771.
14. Bernstein FC, Koetzle TF, Williams GJB, Meyer EJ Jr, Brice MD, Rodgers JK, Kennard O, Shimanouchi I, Tasumi, M. The Protein Data Bank: a computer-based archival file for macromolecular structures. J Mol Biol 1977;112:535–542.
15. Denessiouk KA, Johnson MS. When fold is not important: a common structural framework for adenine and AMP binding in 12 unrelated protein families. Proteins 2000;38:310–326.
16. Sobolev V, Wade RC, Vriend G, Edelman M. Molecular docking using surface complementarity. Proteins 1996;25:120–129.
17. Engel C, Wierenga R. The diverse world of coenzyme A binding proteins. Curr Opin Struct Biol 1996;6:790–797.
18. Rossmann MG, Liljas A, Branden C-I, Banaszak LJ. Evolutionary and structural relationships among dehydrogenases. In: Boyer, PD, editor. The enzymes. Vol 11. New York: Academic Press; 1975. p 61–102.
19. Wierenga RK, De Maeyer MCH, Hol WGJ. Interaction of pyrophosphate moieties with α-helices in dinucleotide binding proteins. Biochemistry 1985;24:1346–1357.
20. Wierenga RK, Terpstra P, Hol WGJ. Prediction of the occurrence of the ADP-binding βαβ-fold in proteins, using an amino acid sequence fingerprint. J Mol Biol 1986;187:101–107.
21. Eggink G, Engel H, Vriend G, Terpstra P, Witholt B. Rubredoxin reductase of *Pseudomonas oleovorans*. Structural relationship to other flavoprotein oxidoreductases based on one NAD and two FAD fingerprints. J Mol Biol 1990;212:135–142.
22. Eppink MH, Schreuder HA, Van Berkel WJ. Identification of a novel conserved sequence motif in flavoprotein hydroxylases with a putative dual function in FAD/NAD(P)H binding. Protein Sci 1997;6:2454–2458.
23. Stehr M, Diekmann H, Smau L, Seth O, Ghisla S, Singh M, Macheroux P. A hydrophobic sequence motif common to N-hydroxylating enzymes. Trends Biochem Sci 1998;23:56–57.
24. Vallon O. New sequence motifs in flavoproteins: evidence for common ancestry and tools to predict structure. Proteins 2000;38: 95–114.
25. Pai EF. Variations on a theme: the family of FAD-dependent NAD(P)H-(disulphide)-oxidoreductases. Curr Opin Struct Biol 1991;1:796–803.
26. Ziegler GA, Vonrhein C, Hanukoglu I, Schulz GE. The structure of adrenodoxin reductase of mitochondrial P450 systems: electron transfer for steroid biosynthesis. J Mol Biol 1999;289:981–990.
27. Chen ZW, Koh M, Van Driessche G, Van Beeumen JJ, Bartsch RG, Meyer TE, Cusanovich MA, Mathews FS. The structure of flavocytochrome C sulfide dehydrogenase from a purple phototrophic bacterium. Science 1994;266:430–432.
28. Mande SS, Parsonage D, Claiborne A, Hol WG. Crystallographic analyses of NADH peroxidase Cys42Ala and Cys42Ser mutants: active site structures, mechanistic implications, and an unusual environment of Arg303. Biochemistry 1995;34:6985–6992.
29. Cavener DR. GMC oxidoreductases. A newly defined family of homologous proteins with diverse catalytic activities. J Mol Biol 1992;223:811–814.
30. Mattevi A. The PHBH fold: not only flavoenzymes. Biophys Chem 1998;70:217–222.
31. Binda C, Coda A, Angelini R, Federico R, Ascenzi P, Mattevi A. A 30 Å long U-shaped catalytic tunnel in the crystal structure of polyamine oxidase. Struct Fold Des 1999;7:265–276.
32. Trickey P, Wagner MA, Jorns MS, Mathews FS. Monomeric sarcosine oxidase: structure of a covalently flavinylated amine oxidizing enzyme. Struct Fold Des 1999;7:331–345.
33. Murzin AG. Structural classification of proteins: new superfamilies. Curr Opin Struct Biol 1996;6:386–394.
34. Fraaije MW, Van Berkel WJ, Benen JA, Visser J, Mattevi A. A novel oxidoreductase family sharing a conserved FAD-binding domain. Trends Biochem Sci 1998;23:206–207.
35. Chakrabarti P, Samanta, U. CH/π interaction in the packing of the adenine ring in protein structures. J Mol Biol 1995;251:9–14.
36. Richardson JS. The anatomy and taxonomy of protein structure. Adv Protein Chem 34:167–339.
37. Li M, Dyda F, Benhar I, Pastan I, Davies DR. Crystal structure of the catalytic domain of *Pseudomonas* exotoxin A complexed with a nicotinamide adenine dinucleotide analog: implications for the activation process and for ADP ribosylation. Proc Natl Acad Sci USA 1996;93:6902–6906.
38. Hurley JH, Dean AM. Structure of 3-isopropylmalate dehydrogenase in complex with NAD^+: ligand-induced loop closing and mechanism for cofactor specificity. Structure 1994;2:1007–1016.
39. Imada K, Sato M, Tanaka N, Katsube Y, Matsuura Y, Oshima T. Three-dimensional structure of a highly thermostable enzyme, 3-isopropylmalate dehydrogenase of *Thermus thermophilus* at 2.2 Å resolution. J Mol Biol 1991;222:725–738.
40. Hurley JH, Thorsness PE, Ramalingam V, Helmers NH, Koshland DE Jr, Stroud RM. Structure of a bacterial enzyme regulated by phosphorylation, isocitrate dehydrogenase. Proc Natl Acad Sci USA 1989;86:8635–8639.
41. Bell CE, Eisenberg D. Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide. Biochemistry 1996;35:1137–1149.
42. Stoddard BL, Dean A, Koshland DE Jr. Structure of isocitrate

dehydrogenase with isocitrate, nicotinamide adenine dinucleotide phosphate, and calcium at 2.5-Å resolution: a pseudo-Michaelis ternary complex. Biochemistry 1993;32:9310–9316.
43. Labesse G, Vidal-Cros A, Chomilier J, Gaudry M, Mornon JP. Structural comparisons lead to the definition of a new superfamily of NAD(P)(H)-accepting oxidoreductases: the single-domain reductases/epimerases/dehydrogenases (the "RED" family). Biochem J 1994;304:95–99.
44. Jörnvall H, Persson B, Krook M, Atrian S, Gonzàlez-Duarte R, Jeffery J, Ghosh D. Short-chain dehydrogenases/reductases (SDR). Biochemistry 1995;34:6003–6013.
45. Jörnvall H, Persson M, Jeffery J. Alcohol and polyol dehydrogenases are both divided into two protein types, and structural properties cross-relate the different enzyme activities within each type. Proc Natl Acad Sci USA 1981;78:4226–4230.
46. Ghosh D, Pletnev VZ, Zhu DW, Wawrzak Z, Duax WL, Pangborn W, Labrie F, Lin SX. Structure of human estrogenic 17β-hydroxysteroid dehydrogenase at 2.20 Å resolution. Structure 1995;3:503–513.
47. Kraulis PJ. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. J Appl Crystallogr 1991;24:945–949.
48. Merritt EA, Bacon DJ. Raster3D: photorealistic molecular graphics. Methods Enzymol 1997;277:505–524.