Protein Science (1997), 6:2639–2643. Cambridge University Press. Printed in the USA.

FOR THE RECORD

A diverse superfamily of enzymes with ATP-dependent carboxylate–amine/thiol ligase activity

MICHAEL Y. GALPERIN AND EUGENE V. KOONIN

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894

(RECEIVED June 12, 1997; ACCEPTED September 8, 1997)

Abstract: The recently developed PSI-BLAST method for sequence database search and methods for motif analysis were used to define and expand a superfamily of enzymes with an unusual nucleotide-binding fold, referred to as palmate, or ATP-grasp fold. In addition to D-alanine-D-alanine ligase, glutathione synthetase, biotin carboxylase, and carbamoyl phosphate synthetase, enzymes with known three-dimensional structures, the ATP-grasp domain is predicted in the ribosomal protein S6 modification enzyme (RimK), urea amidolyase, tubulin–tyrosine ligase, and three enzymes of purine biosynthesis. All these enzymes possess ATP-dependent carboxylate-amine ligase activity, and their catalytic mechanisms are likely to include acylphosphate intermediates. The ATP-grasp superfamily also includes succinate–CoA ligase (both ADP-forming and GDP-forming variants), malate–CoA ligase, and ATP–citrate lyase, enzymes with a carboxylate–thiol ligase activity, and several uncharacterized proteins. These findings significantly extend the variety of the substrates of ATP-grasp enzymes and the range of biochemical pathways in which they are involved, and demonstrate the complementarity between structural comparison and powerful methods for sequence analysis.

Keywords: ATP binding site; ATP-grasp fold; biotin carboxylase; glutathione synthetase; purine biosynthesis; succinate thiokinase; tubuline–tyrosine ligase

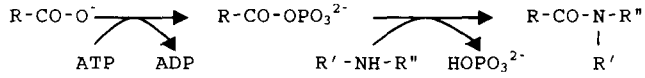
With the rapid accumulation of three-dimensional (3D) protein structures and the complementary development of structure-to-structure comparison methods, there has been lately a remarkable growth in the number of protein superfamilies delineated through structural conservation alone, in the absence of detectable sequence similarity (Holm & Sander, 1996). One of such structural superfamilies unites two groups of peptide synthetases, namely D-alanine:D-alanine ligase (DD-ligase) and glutathione synthetase (GSHase), with biotin carboxylases (BCases) and carbamoyl phosphate synthase (Fan et al., 1995; Artymiuk et al., 1996; Thoden et al., 1997). All these enzymes catalyze a reaction that involves an ATP-dependent ligation of a carboxyl group carbon of one substrate with an amino or imino group nitrogen of the second one and includes, in each case, the formation of acylphosphate intermediates (Gushima et al., 1983; Ogita & Knowles, 1988; Meister, 1989; Fan et al., 1994). Structural alignment of DD-ligase, GSHase, and BCase revealed three conserved motifs, corresponding to the phosphate-binding loop and the Mg^{2+}-binding site of the ATP-binding domain (Artymiuk et al., 1996). In each of these enzymes, ATP binds in a cleft formed by two structural elements, each containing two antiparallel β-strands and a loop (Hibi et al., 1996). A similar ATP-binding fold, referred to as GSHase fold, palmate (β-sheet) fold (Yamaguchi et al., 1993), or ATP-grasp fold (Murzin, 1996) has been detected in succinyl–CoA synthetase (SCS) (Wolodko et al., 1994; Matsuda et al., 1996). Sequence similarity with BCases indicates that this superfamily additionally includes the biotin-dependent carboxylase domains of pyruvate carboxylase and propionyl–CoA carboxylase (Artymiuk et al., 1996). Here, using recently developed sensitive methods for sequence database search and sequence motif analysis, we further expand the ATP-grasp superfamily to include the enzyme involved in ribosomal protein S6 modification, urea amidolyase, tubulin–tyrosine ligase, three enzymes of purine biosynthesis, and several uncharacterized proteins. These findings significantly extend the range of the biochemical pathways, in which ATP-grasp enzymes are involved, the variety of their substrates, and emphasize the complementarity between structural comparison and powerful methods for sequence analysis.

In the course of detailed comparative analysis of the protein sequences encoded in complete bacterial and archaeal genomes (Koonin et al., 1997), we observed that the *Escherichia coli* RimK protein, which is involved in post-translational modification of the ribosomal protein S6 (Reeh & Pedersen, 1979; Kang et al., 1989), had highly conserved homologs in all completely sequenced bacterial and archaeal genomes and also showed a significant similarity to GSHases. When the non-redundant protein sequence database at the National Center for Biotechnology Information was searched using BLASTGP program, which is an extension of the BLAST method (Altschul et al., 1990) incorporating statistical analysis of local alignments with gaps (Altschul & Gish, 1996;

Reprint requests to: Michael Y. Galperin, National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894; e-mail: galperin@ncbi.nlm.nih.gov.

Altschul et al., 1997), alignments of the RimK sequence with GSHases were detected with a probability of occurring by chance, $P < 10^{-8}$. We further iterated this search using the recently developed PSI-BLAST (Position-Specific Iterative BLAST) program, which converts local alignment produced by BLASTGP into position-specific weight matrices that are then used for iterative database scanning (Altschul et al., 1997). This search detected the known proteins with the ATP-grasp fold, GSHases, DD-ligases, and BCases, with a high statistical significance ($P < 10^{-4}$). It also revealed, at the same significance level, a similar domain in urea amidolyase, phosphoribosylamine–glycine ligase, phosphoribosylglycinamide formyltransferase, and phosphoribosylaminoimidazole carboxylase. In addition, marginal similarity was detected with the sequences of SCS and tubulin–tyrosine ligase (TTL). Finally, when the alignment block containing the phosphate-binding site with a flexible glycine-rich loop flanked by two anti-parallel β-strands from these enzymes (residues 137–157 in DD-ligase) was used in a motif search using MoST program (Tatusov et al., 1994), a total of 125 different sequences were retrieved, of which 122 were considered members of the ATP-grasp superfamily.

The extended ATP-grasp superfamily currently includes 15 groups of enzymes, catalyzing ATP-dependent ligation of a carboxylate-containing molecule to an amino or thiol group-containing molecule (Table 1). The list of reactions catalyzed by these enzymes demonstrates their flexibility with respect to both carboxyl and amino/thiol group-containing substrates. Thus, phosphoribosylglycinamide formyltransferase uses formic acid as a substrate, showing that the moiety at the carboxyl group can be as simple as H atom. On the other hand, in case of RimK and TTL, the carboxyl-containing substrates are proteins. In carbamoyl phosphate synthetase, the amino group containing substrate is simply ammonia (derived from glutamine), while in biotin carboxylases this substrate is N′ atom of enzyme-bound biotin molecule. This shows that primary and secondary amines can both be used by enzymes of this family. The reaction catalyzed by ATP-dependent carboxylate-amine ligases can be summarized as follows:



In this scheme, R can be a hydrogen atom, hydroxyl group, an organic molecule, or even a protein; R′ can be either a hydrogen atom or a part of a biotin ring, and R″ can be an amino-group

Table 1. *Carboxylate–amine/thiol ligases containing ATP-grasp domains*

Enzyme	Function or pathway	SWISS-PROT symbol	Active form	Carboxylate substrate	Amine or thiol substrate
Ribosomal protein S6 modification protein[a]	Ribosome biogenesis	RIMK_ECOLI	Monomer?	Ribosomal protein S6	Glutamate
Glutathione synthetase (EC 6.3.2.3)	Glutathione biosynthesis	GSHB_ECOLI	Tetramer	γ-Glutamyl-cysteine	Glycine
D-Alanine-D-alanine ligase (EC 6.3.2.4)	Peptidoglycan biosynthesis	DDLA_ECOLI DDLB_ECOLI	Dimer	D-Alanine	D-Alanine
Phosphoribosylamineglycine ligase[a] (EC 6.3.4.13)	Purine biosynthesis	PUR2_ECOLI	Monomer	Glycine	5-Phosphoribosylamine
Phosphoribosylglycinamide formyltransferase[a] (EC 2.1.2.-)	Purine biosynthesis	PURT_ECOLI	Monomer	$HCOO^-$	5′-Phosphoribosylglycinamide
Phosphoribosylaminoimidazole carboxylase[a] (EC 4.1.1.21)	Purine biosynthesis	PURK_ECOLI PUR6_YEAST	Dimer	HCO_3^-	5′-Phosphoribosyl-5-aminoimidazole
Acetyl–CoA carboxylase, biotin carboxylase subunit (EC 6.3.4.14)	Fatty acid biosynthesis	ACCC_ECOLI COAC_YEAST	Heterohexamer Tetramer	HCO_3^-	Biotin-enzyme
Propionyl–CoA carboxylase (EC 6.4.1.3)	Amino acid catabolism	PCCA_HUMAN	Heterodimer	HCO_3^-	Biotin-enzyme
Pyruvate carboxylase (EC 6.4.1.1)	Gluconeogenesis	PYC_HUMAN	Tetramer	HCO_3^-	Biotin-enzyme
Urea amidolyase[a] (EC 6.3.4.6)	Urea hydrolysis	DUR1_YEAST	Monomer	HCO_3^-	Biotin-enzyme
Carbamoyl-phosphate synthetase, large chain (EC 6.3.5.5)	Arginine biosynthesis pyrimidine biosynthesis	CARB_ECOLI PYR1_HUMAN	Heterodimer Hexamer	HCO_3^- NH_2COO^-	NH_3 —
Tubulin-tyrosine ligase[a] (EC 6.3.2.25)	Microtubules assembly	TTL_PIG	Monomer	α-Tubulin	Tyrosine
Succinyl-CoA synthetase, β subunit (EC 6.2.1.5, 6.2.1.4)	Citric acid cycle	SUCC_ECOLI SUCB_PIG	Heterotetramer Heterodimer	Succinate	Coenzyme A
Malate-CoA ligase, β subunit[a] (EC 6.2.1.9)	Growth on C-1 compounds	MTKB_METEX	Heterodimer	Malate, succinate	Coenzyme A
ATP-citrate lyase[a] (EC 4.1.3.8)	Lipid biosynthesis	ACLY_HUMAN	Monomer	Citrate	Coenzyme A

[a]Newly identified members of the superfamily.

containing a molecule, or a part of a biotin ring (Table 1). In several cases, substrates of ATP-grasp enzymes do not have an amino group. One of such enzymes is the enterococcal DD-ligase (vancomycin-resistance protein), which ligates D-alanine with D-lactate (Fan et al., 1994; Evers et al., 1996). Another variant of the same catalytic mechanism works in succinyl–CoA synthetase, where it is the thiol group of HS-CoA that performs the nucleophilic attack on the succinyl phosphate intermediate. This probably also occurs in related enzymes, such as malate–CoA ligase and ATP–citrate lyase (Wells, 1991):

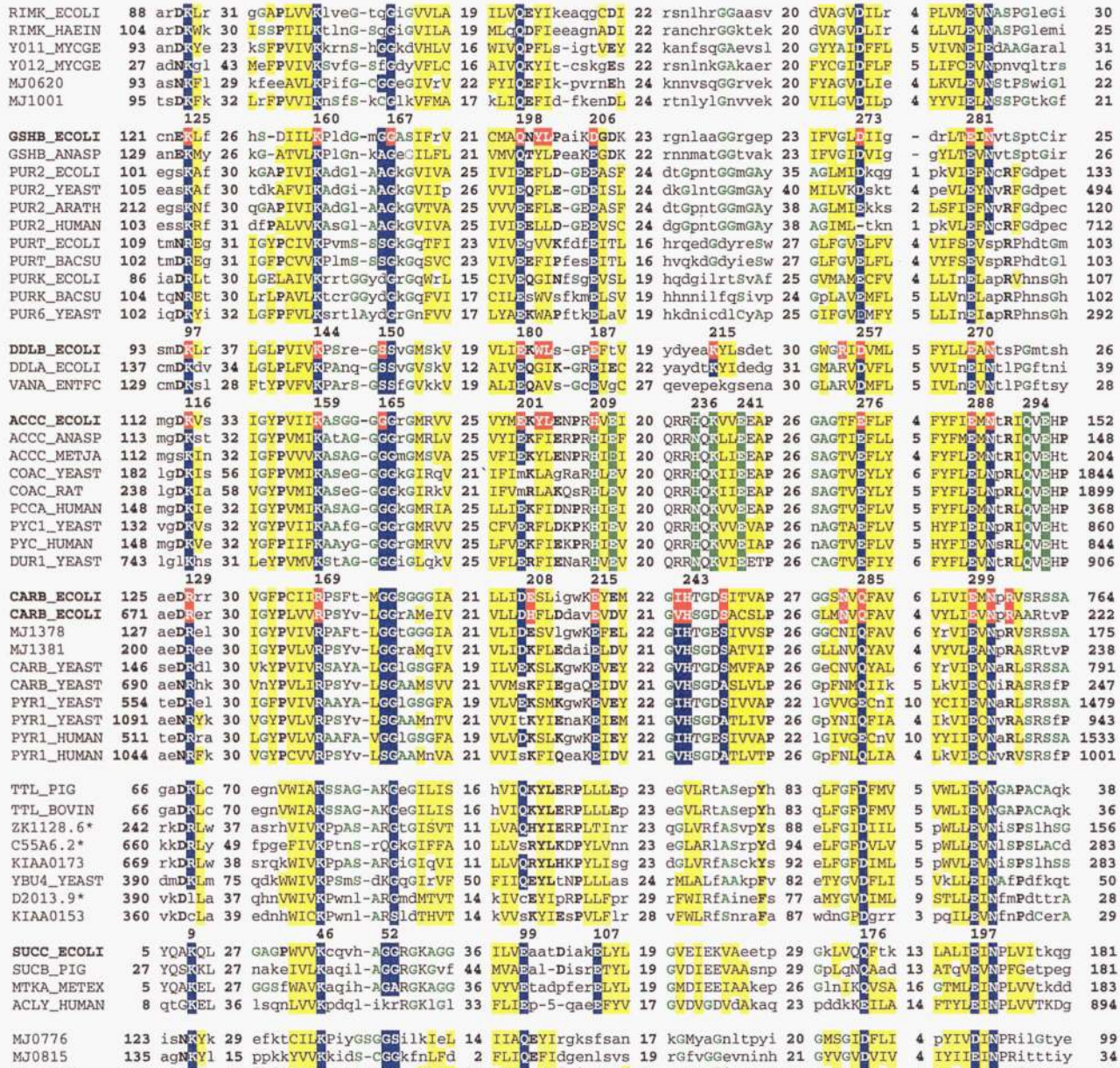
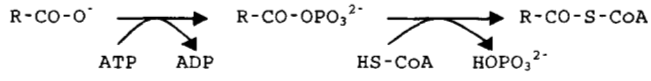


Fig. 1. Multiple alignment of the conserved regions in the ATP-grasp proteins. The proteins are listed under their unique SWISS-PROT or GenBank identifiers; proteins from *Caenorhabditis elegans* are marked with an asterisk. The numbers indicate distances to the ends of each protein and the sizes of the gaps between aligned segments. The names of proteins with known 3D structures are shown in bold; the positions of the conserved residues are indicated above such residues. Red shading indicates the amino acid residues that were shown to be involved in ATP binding by X-ray analysis; corresponding amino acid residues in other sequences are shaded blue. Magenta shading indicates the residues identified by site-specific mutagenesis. Conserved amino acid residues of the active center of biotin carboxylases are shaded green. Yellow shading indicates uncharged amino acid residues (A, I, L, V, M, F, Y, or W) with a propensity to form β-strands. Conserved small residues (G, A, S, or C) are shown in green, the residues conserved within a protein family are in caps, the ones conserved among several protein families are in bold. The consensus includes amino acid residues conserved in all sequences (upper case) and those conserved in the majority of the sequences (lower case). U stands for a bulky hydrophobic residue (I, L, V, M, F, Y, W), O stands for a small residue (G, A, S, C), + stands for K or R, − stands for D or E, $ indicates any charged residue (D, E, K, R, N, Q), and dot stands for any residue.



However, these exceptions appear to be the only ones where an ATP-grasp enzyme works without an amino group-containing substrate. Remarkably, SCS of *E. coli* can use GTP and ITP in addition to ATP (Murakami et al., 1972; Kelly & Cha, 1977), while the enzymes from pig hearts and *Dictyostelium discoideum* are GTP specific (discussed by Nishimura, 1986; Anschutz et al., 1993).

The multiple alignment of the ATP-grasp superfamily generated from the PSI-BLAST data using CLUSTALW (Thompson et al., 1994) and MACAW (Schuler et al., 1991) programs (Fig. 1) shows a clear pattern of conservation in the three motifs described by Artymiuk et al. (1996). These motifs are also detectable in the sequences of SCS, in accordance with the structural classification (Murzin, 1996), and TTL, confirming that TTL contains an ATP-grasp domain (Fig. 1). The alignment also includes several uncharacterized proteins, for which no function could be deduced from the sequence data.

This alignment demonstrates that most of the amino acid residues that interact with ATP in DD-ligase (Fan et al., 1994, 1995) are conserved in all ATP-grasp domains, and the few allowed substitutions are consistent with their predicted role in ATP binding. Thus, Lys-97 and Lys-144 of DD-ligase, interacting electrostatically with α- and β-phosphates of ATP, are conserved in a majority of sequences and are only substituted by Arg in carbamoyl phosphate synthetases. Glu-180, hydrogen bonded to the amino group of adenine, can be substituted by Gln or Asp, while Glu-187, hydrogen bonded to ribose OH groups, can be substituted by His, Asp, or Asn. Amino acid residues Asp-257, Glu-270, and Asn-272, participating in coordination of Mg^{2+}, and Trp-182, Leu-183, and Leu-269, providing hydrophobic interactions for adenine and ribose rings, are also highly conserved. In accordance with the site mutagenesis data (Shi & Walsh, 1995), Tyr-216 can be substituted by other hydrophobic residues. On the other hand, most of the amino acid residues forming the active center of BCase (Waldrop et al., 1994), such as Tyr-82, His-236, Lys-238, Glu-241, Gln-294, and Glu-296, are conserved only among BCases. These residues are not conserved even in phosphoribosylaminoimidazole carboxylase (Fig. 1), which also uses bicarbonate as a substrate, but is likely biotin-independent (Mueller et al., 1994). Finally, we found no sequence conservation around Glu-15, which in DD-ligase binds amino groups of the substrate. This may reflect the variety of the amine-containing substrates of the enzymes of this family.

The ATP-grasp domain is ubiquitous, with multiple representatives of the superfamily encoded in each of the completely sequenced genomes (Table 2). It is of interest, however that, unlike metabolic enzymes, the two groups of ATP-grasp proteins involved in protein modification, while highly conserved, are limited in their phylogenetic distribution to prokaryotes (RimK) or eukaryotes (TTL) (Table 2). Conceivably, these enzymes could have evolved from ancestral metabolic enzymes. Phylogenetic distribution of the ATP-grasp enzymes suggests some interesting functional clues. *Methanococcus jannaschii* encodes two paralogous proteins that are both orthologous to RimK, yet there is no gene for

Table 2. *ATP-grasp domains encoded in prokaryotic and eukaryotic genomes*

Organism	RimK	Glutathione synthetase	D-Alanine-D-alanine ligase	Purine biosynthesis enzymes			Biotin carboxylases	Carbamoyl phosphate synthetase	Tubulin-tyrosin ligase	Unknown
Escherichia coli	RimK	GshB	DdlA DdlB	PurD	PurT	PurK	AccC	CarB	—	—
Hemophilus influenzae	HI1531	—	HI1140	HI0888	—	HI1616	—	—	—	—
Mycoplasma genitalium	MG011 MG012	—	—	—	—	—	—	—		
Synechocystis sp.	—	slr1238 slr2002	slr1874	slr1159	slr0861	sll0578	sll0053	sll0370	—	slr1616
Methanococcus jannaschii	MJ0620 MJ1001	—	—	MJ0937	MJ1486	—	MJ1229	MJ1378 MJ1381	—	MJ0776 MJ0815
Saccharomyces cerevisiae	—	—[b]	—	YGL234w	—	YOR128c	YBR208c, YBR218c, YGL062w, YNR016c, YM8261.01c	YJL130c YJR109c	YBR094w	YSCL9931
Caenorhabditis elegans	—	—[b]	—	F38B6.4	—	—	D2023.2, F26D10.2, F27D9.5, F32B6.e	D2085.1	C55A6.2 ZK1128.6	F46F11.1
Homo sapiens	—	—[b]	—	P22102	—	—	P05165, P11498, S41121	P27708 P31327	KIAA0153 KIAA0173	

[a]Proteins are indicated by their original authors' designations; *E. coli* proteins are from SWISS-PROT database, *H. influenzae*—from Fleischmann et al. (1995), *M. genitalium* —from Fraser et al. (1995), *Synechocystis sp.*—from Kaneko et al. (1996), *M. jannaschii*—from Bult et al. (1996), *S. cerevisiae*—from the *Saccharomyces* genome database (Stanford University). Yeast, human, and worm proteins are listed under their GenBank identifiers.

[b]Eukaryotic glutathione synthetase has no detectable sequence similarity with the bacterial enzyme.

the bacterial ribosomal protein S6 in the *M. jannaschii* genome. A similar pattern can be seen in the recently sequenced genomes of two other archaea, *Methanobacterium thermoautotrophicum* and *Archaeoglobus fulgidus*. Thus, it seems likely that RimK actually has another, conserved, but not yet discovered activity. Of further interest is the finding of a group of uncharacterized eukaryotic proteins containing an ATP-grasp domain and showing the highest, though limited, similarity to RimK. These putative enzymes may be involved in an as yet unknown protein modification mechanism.

A substantial number of enzymes that are similar in function to ATP-dependent carboxylate-amine ligases do not show any detectable sequence similarity to the ATP-grasp superfamily. These include such peptide synthetases as γ-glutamyl-cysteine synthetase, eukaryotic GSHase, and bacterial peptidoglycan biosynthesis proteins MurC, MurD, and MurE. MurD was shown recently to have a typical Rossman fold (Bertrand et al., 1997), rather than ATP-grasp. Pyruvate phosphate dikinase, which reportedly has the ATP-grasp fold (Herzberg et al., 1996; Murzin, 1996), shows no apparent sequence similarity to DD-ligase, GSHase, BCase, or SCS. Finally, in spite of certain similarities in the reaction mechanism, there is no indication that glutamine synthetases belong to the ATP-grasp superfamily.

On a general note, the findings presented here show that with refinement of sequence comparison methods their sensitivity may match that of methods based on structure comparison. Such developments are particularly important, given the parallel rapid growth of sequence and structure databases as sequence analysis methods complement structure analysis by assigning subtly similar sequences to superfamilies with known folds.

Acknowledgments: We thank Stephen Altschul, David Lipman, Tom Madden, and Alejandro Schaffer for providing us with the pre-release version of the program PSI-BLAST, and James Knox and Alex Kuzin for helpful discussion.

References

Altschul SF, Gish W. 1996. Local alignment statistics. *Methods Enzymol 266*:460–480.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol 215*:403–410.

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucl Acids Res 25*:3389–3402.

Anschutz AL, Um HD, Siegel NR, Veron M, Klein C. 1993. P36, a *Dictyostelium discoideum* protein whose phosphorylation is stimulated by GDP, is homologous to the alpha-subunit of succinyl–CoA synthetase. *Biochim Biophys Acta 1162*:40–46.

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

Bertrand JA, Auger G, Fanchon E, Martin L, Blanot D, van Heijenoort J, Dideberg O. 1997. Crystal structure of UDP-*N*-acetylmuramoyl-L-alanine:D-glutamate ligase from *Escherichia coli*. *EMBO J 16*:3416–3425.

Bult CJ, White O, Olsen GJ, Zhou L, Fleischmann RD, Sutton GG, Blake JA, FitzGerald LM, Clayton RA, Gocayne JD, et al. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science 273*:1058–1073.

Evers S, Casadewall B, Charles M, Dutka-Malen S, Galimand M, Courvalin P. 1996. Evolution of structure and substrate specificity in D-alanine:D-alanine ligases and related enzymes. *J Mol Evol 42*:706–712.

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, 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.

Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, Bult CJ, Tomb JF, Dougherty BA, Merrick JM, et al. 1995. Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. *Science 269*:496–512.

Fraser CM, Gocayne JD, White O, Adams MD, Clayton RA, Fleischmann RD, Bult CJ, Kerlavage AR, Sutton G, Kelley JM, et al. 1995. The minimal gene complement of *Mycoplasma genitalium*. *Science 270*:397–403.

Gushima H, Miya T, Murata K, Kimura A. 1983. Purification and characterization of glutathione synthetase from *Escherichia coli* B. *J Appl Biochem 5*:210–218.

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.

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. *Nat Struct Biol 3*:16–18.

Holm L, Sander C. 1996. Alignment of three-dimensional protein structures: Network server for database searching. *Methods Enzymol 266*:653–662.

Kaneko T, Sato S, Kotani H, Tanaka A, Asamizu E, Nakamura Y, Miyajima N, Hirosawa M, Sugiura M, Sasamoto S, et al. 1996. Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis* sp. strain PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions. *DNA Res 3*:109–136.

Kang WK, Icho T, Isono S, Kitakawa M, Isono K. 1989. Characterization of the gene rimK responsible for the addition of glutamic acid residues to the C-terminus of ribosomal protein S6 in *Escherichia coli* K12. *Mol Gen Genet 217*:281–288.

Kelly CJ, Cha S. 1977. Nucleotide specificity of succinate thiokinases from bacteria. *Arch Biochem Biophys 178*:208-217.

Koonin EV, Mushegian AR, Galperin MY, Walker DR. 1997. Comparison of archaeal and bacterial genomes: Computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for archaea. *Mol Microbiol 25*:619–637.

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.

Meister A. 1989. Mechanism and regulation of the glutamine-dependent carbamyl phosphate synthetase of *Escherichia coli*. *Adv Enzymol Relat Areas Mol Biol 62*:315–374.

Mueller EJ, Meyer E, Rudolph J, Davisson VJ, Stubbe J. 1994. N5-carboxyaminoimidazole ribonucleotide: Evidence for a new intermediate and two new enzymatic activities in the de novo purine biosynthetic pathway of *Escherichia coli*. *Biochemistry 33*:2269–2278.

Murakami K, Mitchell T, Nishimura JS. 1972. Nucleotide specificity of *Escherichia coli* succinic thiokinase. Succinyl coenzyme A-stimulated nucleoside diphosphate kinase activity of the enzyme. *J Biol Chem 247*:6247–6252.

Murzin AG. 1996. Structural classification of proteins: New superfamilies. *Curr Opin Struct Biol 6*:386–394.

Nishimura JS. 1986. Succinyl–CoA synthetase structure–function relationships and other considerations. *Adv Enzymol Relat Areas Mol Biol 58*:141–172.

Ogita T, Knowles JR. 1988. On the intermediacy of carboxyphosphate in biotin-dependent carboxylations. *Biochemistry 27*:8028–8033.

Reeh S, Pedersen S. 1979. Post-translational modification of *Escherichia coli* ribosomal protein S6. *Mol Gen Genet 173*:183–187.

Schuler GD, Altschul SF, Lipman DJ. 1991. A workbench for multiple alignment construction and analysis. *Proteins 9*:180–190.

Shi Y, Walsh CT. 1995. Active site mapping of *Escherichia coli* D-Ala-D-Ala ligase by structure-based mutagenesis. *Biochemistry 34*:2768–2776.

Tatusov RL, Altschul SF, Koonin EV. 1994. Detection of conserved segments in proteins: Iterative scanning of sequence databases with alignment blocks. *Proc Natl Acad Sci USA 91*:12091–12095.

Thoden JB, Holden HM, Wesenberg G, Raushel FM, Rayment I. 1997. Structure of carbamoyl phosphate synthetase: A journey of 96 Å from substrate to product. *Biochemistry 36*:6305–6316.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res 22*:4673–4680.

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

Wells TN. 1991. ATP-citrate lyase from rat liver. Characterisation of the citryl–enzyme complexes. *Eur J Biochem 199*:163–168

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.

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* B at 2.0 Å resolution. *J Mol Biol 229*:1083–1100.