A Novel Ligand-binding Domain Involved in Regulation of Amino Acid Metabolism in Prokaryotes*

A combination of sequence profile searching and structural protein analysis has revealed a novel type of small molecule binding domain that is involved in the allosteric regulation of prokaryotic amino acid metabolism. This domain, designated RAM, has been found to be fused to the DNA-binding domain of Lrp-like transcription regulators and to the catalytic domain of some metabolic enzymes, and has been found as a stand-alone module. Structural analysis of the RAM domain of Lrp reveals a $\beta\beta\alpha\beta\beta\alpha$-fold that is strikingly similar to that of the recently described ACT domain, a ubiquitous allosteric regulatory domain of many metabolic enzymes. However, structural alignment and re-evaluation of previous mutagenesis data suggest that the effector-binding sites of both modules are significantly different. By assuming that the RAM and ACT domains originated from a common ancestor, these observations suggest that their ligand-binding sites have evolved independently. Both domains appear to play analogous roles in controlling key steps in amino acid metabolism at the level of gene expression as well as enzyme activity.

Allosteric regulation is a general mechanism that enables a tight control of both enzyme activity and gene expression. A textbook example of this mechanism is the feedback inhibition of enzymes that catalyze key steps in amino acid biosynthesis. Important insight in the molecular basis of modulated enzyme activity has been provided by the crystal structure of 3-phosphoglycerate dehydrogenase (SerA) (1, 2), an enzyme that catalyzes the rate-limiting first step in serine biosynthesis. The SerA structure has revealed separated catalytic and regulatory domains that are connected by a flexible hinge. The structural motif of the regulatory domain consists of a four-stranded anti-parallel $\beta$-sheet with two $\alpha$-helices packed on one side. When the serine effector molecule binds to this $\alpha\beta$-sandwich, the hinge region allows allosteric regulation: a slight interdomain rearrangement that down-regulates the catalytic activity of the enzyme (2, 3).

A thorough sequence profile analysis has shown that the SerA regulatory domain is an ancient small molecule binding domain (SMBD) that is conserved in a wide variety of enzymes as well as in some transcriptional regulators that are involved in the control of amino acid and purine metabolism (4). As predicted in the latter study, the presence of this ancient “ACT domain” (for review see Ref. 5) was indeed demonstrated in the recent crystal structures of rat phenylalanine hydroxylase (6), an enzyme that catalyzes the conversion of phenylalanine to tyrosine. Interestingly, the structure of the Escherichia coli threonine deaminase (7), the enzyme catalyzing the first step of the isoleucine biosynthesis pathway, revealed two domains that resembled the ACT domain of SerA at the structural level rather than at the sequence level (4). This observation demonstrates that the sequence divergence of SMBDs like ACT can expand beyond the detection limits of the sequence-based algorithm of PSI-BLAST (8). For this reason, the threonine deaminase-ACT domains have been referred to as “ACT-like” domains (5).

In the present study we describe a novel ligand-binding module that we named the RAM domain because of its general involvement in the allosteric Regulation of Amino acid Metabolism. This domain is mainly found in association with a class of prokaryotic transcriptional regulators but also as a module in enzymes and in some instances as stand-alone SMBD.

EXPERIMENTAL PROCEDURES

PSI-BLAST Analysis and Multiple Sequence Alignments, Domain Analysis of Proteins—In order to verify and characterize the relationship between distant RAM domains at the sequence level, we performed several PSI-BLAST searches (8) at the National Center of Biotechnology Information. When a PSI-BLAST search was seeded with the C-terminal domain of the Pyrococcus furiosus LrpA (residues 62–141), using a BLOSUM60 matrix and an expect value threshold of 0.001, the first stand-alone versions of RAM (lacking the HTH domain) were retrieved within the first iteration (10580664; $E = 6 \times 10^{-12}$); the RAM domains within the Sulfolobus solfataricus and Sulfolobus tokodaii 2-isopropylmalate synthase (13814162, respectively, 15623321) were recovered in iteration 3 ($E = 7 \times 10^{-6}$). A reverse PSI-BLAST using the S. solfataricus 2-isopropylmalate synthase sequence (residues 342–461) recovered HTH-RAM proteins (e.g. 13813287 at iteration 1, $E = 3 \times 10^{-3}$) and stand-alone versions of the RAM domain (e.g. 13813398 at iteration 2, $E = 5 \times 10^{-3}$), thereby connecting the most distant RAM domain containing proteins with each other on a statistical basis. A multiple alignment of RAM domains was constructed by collecting the “highest scoring pairs of sequence segments” using PSI-BLAST, which were re-aligned using ClustalW (9), followed by minor manual adjustment based on the secondary structure. Domain analysis of RAM- and ACT-containing proteins was performed using SMART (10) an PFAM (11). All protein sequences were extracted from ENTREZ.

Structural Alignment and Superimposition Three-dimensional Structures of SerA ACT Domain and LrpA RAM Domain—In order to

* The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The atomic coordinates and structure factors (code 1I1G) have been deposited in the Protein Data Bank, Research Collaboratory for Structural Bioinformatics, Rutgers University, New Brunswick, NJ (http://www.rcsb.org/).

§ To whom correspondence should be addressed. Tel.: 31-317-483110; Fax: 31-317-483829; E-mail: Thijs.Ettema@algemeen.micr.wau.nl.

1 The abbreviations used are: SMBD, small molecule binding domain; PDB, Protein Data Bank; IPMS, 2-isopropylmalate synthase; HTH, helix-turn-helix.
detect the structural similarity of ACT and RAM, a superimposition of the C-terminal domain of the *P. furiosus* LrpA, the ACT domain present in the SerA of *E. coli* (residues 335–410, PDB entry 1PSD) was constructed using the Swiss PDB viewer (12) using the “Iterative fit” option. The superimposition was constructed with 49 α-carbon atoms, which composes about 65% of the domains. In addition, the position of the negative effector of SerA ACT domain, serine, is confirmed by superimposition of both structures (Fig. 1). The obtained root mean square deviation value between SerA-ACT and LrpA-ACT is 1.8 Å, comprising about 65% of the domains. From the superimposed structures, a structural alignment was deduced using the Structural alignment tool of the Swiss PDB viewer (12).

**RESULTS**

The Structure of the C-terminal Domain of LrpA Resembles Structure of the ACT Domain—The Lrp family of transcriptional regulators plays a crucial role in the control of amino acid metabolism in prokaryotes. Although the leucine-responsive regulatory protein (Lrp) from *E. coli* is a global transcriptional regulator (13), most Lrp homologs act as specific regulators of the expression of several amino acid metabolism genes (14). The interaction of specific amino acid effectors with Lrp-like regulators may lead to modulation of (i) DNA affinity, (ii) DNA bending, (iii) Lrp oligomeric state (dimer/tetramer/octamer/hexadecamer), and (iv) Lrp tertiary structure (15, 16, 17). All these changes most likely reflect an allosteric regulation of Lrp activity by a (minor) structural rearrangement.

Recently, the structure of an archaeal Lrp homolog, the *P. furiosus* LrpA octamer (or tetramer of dimers), has been resolved (18). The structure of the LrpA monomer revealed an N-terminal DNA-binding helix-turn-helix (HTH) domain, an extended hinge, and a C-terminal globular domain with a βαββαββ-fold. Previous mutagenesis analyses were in perfect agreement with the N-terminal domain being involved in the interaction with DNA and predicted the C-terminal αβ-sandwich to have a regulatory function (18, 19). What was not noted during the initial analysis of the LrpA structure is the interesting fact that the C-terminal regulatory domain of LrpA appears to resemble the ACT domains (4, 5) with respect to structure and function; both consist of a typical αβ-sandwich and are anticipated to be regulatory domains involved in allosteric modulation of the activity of enzymes and DNA-binding proteins that are involved in amino acid metabolism. Despite the similarity in structure with the ACT domain, we propose that, for reasons discussed below, the C-terminal regulatory domain of LrpA is part of a novel, distinct class of regulatory domains.

The overall structural resemblance between the RAM domain of the *P. furiosus* LrpA and the ACT domain of the *E. coli* SerA is confirmed by superimposition of both structures (Fig. 1a). The obtained root mean square deviation value between the LrpA-RAM and the SerA-ACT (1.8 Å) compares with a value obtained with a superimposition between SerA-ACT and the ACT domain of the rat phenylalanine hydroxylase (1.7 Å).

**Ligand Response Mutations in RAM and ACT Suggest a Different Location of the Ligand-binding Sites in RAM and ACT**—Despite the structural similarities between the RAM domain and the ACT domain, however, the effector-binding sites in these domains seem to be different. In the ACT domain of SerA, the loop that links the first β-strand (β1) and the first...
A Novel Regulatory Domain in Amino Acid Metabolism

Fig. 2. a, an alignment of the most diverse members of the ACT domain. The secondary structure assignment that is indicated above the alignment was derived from the crystal structure of the E. coli 3-phosphoglycerate dehydrogenase (SerA; PDB code 1PSD). The amino acid residues that are involved in the binding of the effector serine in SerA, are indicated above the alignment (s). In addition, the ligand response mutations that were determined for the E. coli small subunit of the acetolactate synthase (IlvH) and aspartokinase (LysC) are indicated with * above the alignment. The 80% consensus shown below the alignments was obtained as described above, and the position numbers on the left side indicate the limits of the domains. Also a structural alignment of the two regulatory ACT-like domains of the E. coli threonine deaminase (THDI) is included, with a mapped ligand response mutation (+1), indicating the divergence on the sequence level between these domains and the genuine ACT domains. Secondary structure from the first repeat of THDI is indicated above the alignment. The 80% consensus shown below the alignments was obtained using the following amino acid classes (4): small (s), Aea, Aquifex aeolicus; Afu, Archaeoglobus fulgidus; Atu, Agrobacterium tumefaciens; Bsu, Bacillus subtilis; Ccr, Caulobacter crescentus; Eco, E. coli; Mlo, Mesorhizobium loti; Mtu, Mycobacterium tuberculosis; Pho, Pyrococcus abyssi; Pmu, Proteus mirabilis; Pvi, Pasteurella multocida; Psa, Pseudomonas putida; Sco, Streptomyces coelicolor; Sso, Streptomyces venezuelae; Sso, S. solfataricus; Sto, Streptomyces sp.; S. tokodaii; Tvu, Thermoplasma vulcanium.

b, alignment of the most diverse members of the ACT domain. The secondary structure assignment that is indicated above the alignment was derived from the crystal structure of the E. coli 3-phosphoglycerate dehydrogenase (SerA; PDB code 1PSD). The amino acid residues that are involved in the binding of the effector serine in SerA, are indicated above the alignment (s). In addition, the ligand response mutations that were determined for the E. coli small subunit of the acetolactate synthase (IlvH) and aspartokinase (LysC) are indicated with * above the alignment. The 80% consensus shown below the alignments was obtained as described above, and the position numbers on the left side indicate the limits of the domains. Also a structural alignment of the two regulatory ACT-like domains of the E. coli threonine deaminase (THDI) is included, with a mapped ligand response mutation (+1), indicating the divergence on the sequence level between these domains and the genuine ACT domains. Secondary structure from the first repeat of THDI is indicated above the alignment. The 80% consensus shown below the alignments was obtained using the following amino acid classes (4): small (s), Aea, Aquifex aeolicus; Afu, Archaeoglobus fulgidus; Atu, Agrobacterium tumefaciens; Bsu, Bacillus subtilis; Ccr, Caulobacter crescentus; Eco, E. coli; Mlo, Mesorhizobium loti; Mtu, Mycobacterium tuberculosis; Pho, Pyrococcus abyssi; Pmu, Proteus mirabilis; Pvi, Pasteurella multocida; Psa, Pseudomonas putida; Sco, Streptomyces coelicolor; Sso, Streptomyces venezuelae; Sso, S. solfataricus; Sto, Streptomyces sp.; S. tokodaii; Tvu, Thermoplasma vulcanium.

α-helix (α1) of the βαβαββ-fold, makes up the binding pocket of the serine effector. The important role of this loop is in agreement with the fact that it is very well conserved in ACT-containing enzymes and regulators (Fig. 2a). An invariant glycine residue, which is also conserved in the ACT-like domains of the E. coli threonine deaminase, and an adjacent hydrophobic residue have been proposed to be involved in maintenance of the strand-helix interface; two additional conserved polar residues are involved in binding the ligand with hydrogen bonds (1, 4). The importance of the conserved region was also confirmed by mutation analysis of other ACT-containing enzymes from E. coli, i.e. the valine-binding regulatory subunit of the acetolactate synthase (IlvH) (20) and the lysine-sensitive aspartokinase (LysC) (21) (Fig. 2a). Mutations that resulted in a strongly reduced or abolished response when placed at their respective effector (ligand response mutations) all cluster within the conserved loop region, suggesting that binding of the effector resembles the interaction of SerA with serine. Mapping the ACT ligand response mutations into the structure of the ACT dimer of SerA (Fig. 1c) confirms this idea. The ligand
response mutations cluster at the dimer interface in general and at the ligand (serine)-binding site in particular (Fig. 1c).

The dimer structure of the SerA-ACT differs significantly from the LrpA-RAM dimer. Whereas the contact between the two ACT domains seems to be mediated via the α2 helix and β3 interface resulting in an eight-stranded anti-parallel β-sheet (1), in LrpA the RAM dimer is mainly formed by interactions between the antiparallel β-sheets that are facing each other, forming an anti-parallel β-barrel-like structure (Fig. 1c) (18).

Because of this structural difference, and given the fact that the ligand-binding site for ACT domains is located at the dimer interface, this might imply that the ligand-binding site is different in RAM domains. Indeed, the region that is involved in interaction with the ligand (Fig. 2a) is well conserved in ACT domains, whereas the corresponding loop that links strand β1 and helix α1 in RAM domains displays only poor sequence conservation (Fig. 2b). In RAM, the best conserved region, again including an invariant glycine residue, appears to be the region surrounding the loop connecting strands β2 and β3. Moreover, extensive mutagenesis studies that have been performed with the E. coli Lrp (19) confirm the importance of this region with respect to ligand response. The Lrp leucine response mutations that were obtained in this study apparently lost the capacity to bind their ligand (19). When the equivalents of these E. coli Lrp ligand response mutations are mapped into the structure of the LrpA-RAM dimer of P. furiosus, it becomes clear that five of seven mapped mutations belonging to this class (Leu-95, Met-101, Ala-134, Ile-135, and Ile-136) are clustered in a region across the dimer interface (Fig. 1c). The remaining two ligand response mutations (Gly-111 and Gly-123) are located in close proximity to the other mutations, albeit in adjacent dimers of the octamer rather than within the same dimer (18) (not shown in Fig. 1c). These observations suggest that the ligand-binding site of RAM is located at a different position than that of ACT. Based on the relatively high sequence conservation and to some extent on the mutation data, it is tempting to speculate that the ligand-binding site of RAM is located between the β2 and β3. Lrp ligand co-crystallization experiments are required to confirm this hypothesis.

The RAM Domain Has a Wide Phylogenetic Distribution and Is Present in Transcriptional Regulators, Metabolic Enzymes, and as a Stand-alone Version—In order to verify and characterize the relationship between distant RAM domains at the sequence level, we performed several iterative data base searches (PSI-BLAST) (8) at the National Center of Biotechnology Information, revealing the ubiquitous phylogenetic distribution of this do-

![Diagram](image-url)

**Fig. 3. Domain architectures of ACT and RAM containing proteins, subdivided into three classes according to their function.** For each class of proteins, the phylogenetic distribution is depicted below the domain structure. Domain analysis was performed using SMART (20). Apart from the regulatory domains (ACT or RAM), the other domains that are part of the proteins are described as follows. 1, transcription regulation-associated regulatory domains; 2, duplicated form of Lrp-like transcriptional regulator, consisting of a tandem repeat of the transcriptional regulator; 3, transcriptional regulator of aromatic amino acid biosynthesis, containing an N-terminal ACT domain, followed by a PAS domain which is possibly involved in signal sensing. The C-terminal part of these regulators contain a sigma54 interacting domain (PF00098) and a DNA-binding helix-turn-helix domain fused to a C-terminal RAM domain; 2, duplicated form of Lrp-like transcriptional regulator, consisting of a tandem repeat of the transcriptional regulator; 3, transcriptional regulator of aromatic amino acid biosynthesis, containing an N-terminal ACT domain, followed by a PAS domain which is possibly involved in signal sensing. Below each class of proteins, the phyletic distribution is depicted.
A Novel Regulatory Domain in Amino Acid Metabolism

main among Archaea and Bacteria (>250 proteins after 5 iterations). The proteins that were retrieved during this search were dominated by the Lrp-like regulators, typically consisting of an N-terminal DNA-binding helix-turn-helix domain fused to a C-terminal RAM domain (HTH-RAM). An interesting variant of HTH-RAM is a duplicated version, found in some bacteria (Streptomyces sp.); this supports the view that the native form of Lrp-like transcriptional regulators is at least a dimer configuration. In addition, examples were found of a RAM domain that was fused to the C terminus of 2-isopropylmalate synthase (IPMS) in the Crenarchaea S. solfataricus (SS09777), S. tokodaii (ST1301), and Pyrobaculum aerophilum (PAE1986). Also several stand-alone RAM domains were detected with the PSI-BLAST search (Fig. 3). This variable domain architecture of RAM again resembles that of the ACT domain. ACT is found as an allosteric regulatory domain associated with metabolic enzymes (like SerA, phenylalanine hydroxylase and threonine deaminase) and transcriptional regulators (like TyrR and PhxR) as well as a stand-alone version (IlvH) (Fig. 3). It is noted that the RAM domains are mainly associated with transcriptional regulators, whereas the ACT domain is most often found as a regulatory module of metabolic enzymes.

DISCUSSION

The βαββαβ-motif appears to be a common regulatory structure in amino acid metabolic enzymes and transcriptional regulators; both the RAM and the ACT domains share this fold and are associated with proteins that are involved with amino acid metabolism either as part of enzymes, as part of transcriptional regulators, or as stand-alone SMBD. Apart from the structural and functional similarity between the two domains, another connection is the fact that the expression of some bacterial ACT-containing enzymes (e.g. SerA and IlvH) is under control of RAM-containing transcriptional regulators of the Lrp family (9). These observed analogies between RAM and ACT may be useful for speculating about the function of uncharacterized RAM and ACT domains. For example, the function of the stand-alone versions of the RAM domains that are present in several bacterial and archaeal genomes is yet unclear. However, the function of stand-alone ACT domains might suggest the possible function of their stand-alone RAM counterparts. For example, the acetolactate synthase in E. coli is a key enzyme in branched chain amino acid biosynthesis that is subjected to valine feedback inhibition (19). The heterotetrameric holoenzyme is made up of the large catalytic subunit (IlvI) and a small regulatory subunit that consists of a single ACT domain (IlvH) and accounts for the valine-mediated feedback repression. It is possible that a stand-alone RAM domain performs a function that is analogous to that of the IlvH subunit, allosteric regulation of enzymes (or possibly transcriptional regulators) involved in amino acid metabolism via protein-protein interactions. Interestingly, profile-based analysis of prokaryal genomes (e.g. S. solfataricus) failed to identify a gene encoding the small regulatory subunit of the acetolactate synthase (IlvH), whereas homologs of the gene encoding the large catalytic subunit of the acetolactate synthase could be identified on the genome (e.g. ilvB-1). Possibly, the large subunit of the acetolactate synthase present in these organisms are subjected to allosteric regulation by the stand-alone RAM domains that are encoded on the genome.

A more clear-cut example is the anticipated role for the RAM domain that is present at the C terminus of the crenarchaeal IPMS. Generally, IPMS catalyzes the first step in leucine biosynthesis that is subjected to leucine-mediated feedback inhibition. Leucine has the following two known effects on this enzyme in Salmonella typhimurium: (i) it controls the catalytic activity of the enzyme by feedback inhibition (22), and (ii) it causes a dissociation of the tetrameric enzyme into its monomeric subunits (23). Genetic mapping of leucine-insensitive mutants (24, 25) revealed that the mutation resulting in the affected end product inhibition is located at the C terminus of the S. typhimurium IPMS. So the domain that is probably responsible for the allosteric control of IPMS is located in the C-terminal part of the enzyme. Sequence analysis of the C-terminal part of the S. typhimurium IPMS did not reveal the presence of a RAM or ACT domain, possibly suggesting the presence of yet another alternative regulatory domain. Although the exact mechanism of inhibition remains to be elucidated, the presence of a C-terminal RAM domain in the crenarchaeal RAM-IPMS strongly suggests that these enzymes are subjected to RAM-mediated feedback regulation.

The data presented here indicate that structure-based alignments of the regulatory domains as well as extensive (reverse) PSI-BLAST analyses fail to close the apparent gap in sequence divergence between RAM and ACT. This might suggest that ACT and RAM both have independently evolved into allosteric regulatory domains. However, the most parsimonious scenario is that the two domains have emerged from a common ancestor and only evolved different interaction specificity with respect to their ligands. Possibly, the diversification in binding sites is the underlying reason of the dramatic sequence divergence between RAM and ACT, with the overall structure being well conserved. This appears to be yet another example of domain evolution in which sequence similarity has been lost while retaining their structural similarity (26).

Acknowledgment—We thank Dr. Eugene V. Koonin (National Center of Biotechnology Information, Bethesda) for stimulating discussions.

REFERENCES

1. Schaller, D. J., Grant, G. A., and Banaszak, L. J. (1995) Nat. Struct. Biol. 2, 69–76
2. Grant, G. A., Schlifer, D. J., and Banaszak, L. J. (1996) Protein Sci. 5, 34–41
3. Al-Rabiee, R., Zhang, Y., and Grant, G. A. (1996) J. Biol. Chem. 271, 23235–23241
4. Aravind, L., and Koonin, E. V. (1999) J. Mol. Biol. 287, 1023–1040
5. Shipman, D. M., and Shaanan, B. (2001) Curr. Opin. Struct. Biol. 11, 694–700
6. Kobe, B., Jennings, I. G., House, C. M., Michell, B. J., Goodwill, K. E., Santarsiero, B. D., Stevens, R. C., Cotton, R. G., and Kemp, D. B. (1999) Nat. Struct. Biol. 6, 44048
7. Gallagher, D. T., Gilliland, G. L., Xiao, G., Zondlo, J., Fisher, K. E., Chinchilla, D., and Eisenstein, E. (1998) Structure 6, 465–475
8. Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Nucleic Acids Res. 25, 3389–3402
9. Thompson, J. D., Higgins, D. G., and Gibson, T. J. (1994) Nucleic Acids Res. 22, 4676–4680
10. Schultz, J., Milpetz, F., Bork, P., and Ponting, C. P. (1998) Proc. Natl. Acad. Sci. U. S. A. 95, 5578–5584
11. Bateman, A., Birney, E., Cerruti, L., Durbin, R., Eddy, S., Eddy, S. R., Griffiths-Jones, S., Howe, K. L., Marshall, M., and Sonnhammer, E. L. L. (2002) Nucleic Acids Res. 30, 276–280
12. Schwede, T., Diemand, A., Guex, N., and Peitsch, M. C. (2000) Rev. Physiol. Biochem. Pharmacol. 131, 107–112
13. Calvo, J. M., and Matthews, R. G. (1994) Microbiol. Rev. 58, 466–490
14. Kolling, R., Gielow, A., Seufert, W., Kucherer, C., and Messer, W. (1988) J. Biol. Chem. 263, 29–36
15. Madhusudhan, K. T., Huang, N., Braswell, E. H., and Kemp, D. B. (1997) Nat. Struct. Biol. 14, 586–593
16. Tama, A., Kimura, Y., Koga, H., and Tanaka, T. (1999) J. Biol. Chem. 274, 997–1007
17. Plakout, J. V., and Calvo, J. M. (1993) J. Bacteriol. 175, 1110–1117
18. Mendel, S., Elkanay, T., Sella, C., Vinogradov, V., Vayatsenysky, M., Shipman, D. M., and Barak, Z. (2003) J. Mol. Biol. 327, 2445–2477
19. Kikuchi, Y., Kojima, H., and Tanaka, T. (1999) FEBS Microbiol. Lett. 173, 211–215
20. Kohlhaw, G., Leary, T. R., and Umberger, H. E. (1969) J. Biol. Chem. 244, 2218–2225
21. Leary, T. R., and Kohlhaw, G. (1970) Biochem. Biophys. Res. Commun. 39, 584–591
22. Calvo, R. A., and Calvo, J. M. (1967) Science 156, 1107–1109
23. Bartholomew, J. C., and Calvo, J. M. (1971) Biochim. Biophys. Acta 250, 568–576
24. Anantharaman, V., Koonin, E. V., and Aravind, L. (2001) J. Mol. Biol. 307, 1271–1292