RESEARCH

Analogous Enzymes: Independent Inventions in Enzyme Evolution

Michael Y. Galperin,¹ D. Roland Walker,^{1,2} and Eugene V. Koonin^{1,3}

¹National Center for Biotechnology Information (NCBI), National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894 USA; ²Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218 USA

It is known that the same reaction may be catalyzed by structurally unrelated enzymes. We performed a systematic search for such analogous (as opposed to homologous) enzymes by evaluating sequence conservation among enzymes with the same enzyme classification (EC) number using sensitive, iterative sequence database search methods. Enzymes without detectable sequence similarity to each other were found for 105 EC numbers (a total of 243 distinct proteins). In 34 cases, independent evolutionary origin of the suspected analogous enzymes was corroborated by showing that they possess different structural folds. Analogous enzymes were found in each class of enzymes, but their overall distribution on the map of biochemical pathways is patchy, suggesting multiple events of gene transfer and selective loss in evolution, rather than acquisition of entire pathways catalyzed by a set of unrelated enzymes. Recruitment of enzymes that catalyze a similar but distinct reaction seems to be a major scenario for the evolution of analogous enzymes, which should be taken into account for functional annotation of genomes. For many analogous enzymes, the bacterial form of the enzyme is different from the eukaryotic one; such enzymes may be promising targets for the development of new antibacterial drugs.

Enzymes that catalyze the same reaction typically show significant sequence and structural similarity. However, several notable exceptions from this rule have been described (e.g., Smith et al. 1992; Fothergill-Gilmore and Michels 1993; Doolittle 1994). Examples of apparently unrelated enzymes with the same specificity were noted as early as 1943 when Warburg and Christian (1943) described two distinct forms of fructose 1,6-bisphosphate aldolase in yeast and rabbit muscle, respectively. These two enzymes, referred to as class I and class II aldolases, were later shown to be associated with different phylogenetic lineages and have different catalytic mechanisms and little structural similarity (Rutter 1964; Perham 1990; Marsh and Lebherz 1992; Blom et al. 1996). Such enzymes are generally believed to have evolved independently of one another, rather than having descended from a common ancestral enzyme (Smith et al. 1992; Doolittle 1994), and are appropriately referred to as analogous, as opposed to homologous, enzymes (Fitch 1970; Florkin 1974).

On the other hand, a more careful comparison shows that fructose 1,6-bisphosphate aldolases of both classes share the same β/α [triosephosphate

isomerase (TIM)]-barrel fold and structurally similar active centers, indicating that they still could have a common ancestor (Cooper et al. 1996). Sequence comparisons alone cannot prove that two sequences are evolutionarily unrelated; common origin can be inferred from protein structure conservation even after sequence conservation has been completely washed out by divergence (Doolittle 1987; Holm and Sander 1996b; Murzin 1996). The possibility of a common origin can be ruled out only when candidate analogous enzymes have different three-dimensional (3D) folds, as it indeed has been shown for subtilisin and chymotrypsin families (Wright et al. 1969), Cu/Zn- and Mn/Fe-dependent superoxide dismutases (Stallings et al. 1983), and β -lactamases of type I (classes A and C) and type II (Lobkovsky et al. 1993; Carfi et al. 1995).

Despite a number of reports of newly sequenced enzymes that seemed to show no sequence similarity to previously known enzymes catalyzing the same biochemical reactions, analogous enzymes have been usually perceived as rare and exceptional. Recently, such cases gained higher visibility owing to the efforts to reconstruct metabolic pathways encoded in complete microbial genomes. In at least 11 instances, the same reaction in two bacteria, *Haemophilus influenzae* and *Mycoplasma genitalium*, has

³Corresponding author.
E-MAIL koonin@ncbi.nlm.nih.gov; FAX (301) 480 9241.

been found to be catalyzed by unrelated, or at least definitely not orthologous, enzymes (Koonin et al. 1996a). This phenomenon, termed nonorthologous gene displacement, turned out to be common when comparisons between archaeal, eukaryotic, and bacterial genomes have been performed (Ibba et al. 1997a; Koonin et al. 1997). We have undertaken a systematic comparison of the protein sequences of the enzymes stored in the GenBank database to identify candidate analogous enzymes and, whenever possible, corroborate their independent origin by connecting them to distinct structural folds. The functional and phylogenetic distribution of analogous enzymes and their relationships with other enzymes were examined in an attempt to infer their origin.

RESULTS

Newly cloned enzymes are often reported to lack sequence similarity to other enzymes catalyzing the same reactions. However, in many of these cases, iterative database search and detailed analysis of sequence motifs reveal underlying conservation, suggesting evolution from a common ancestor. For example, purine nucleoside phosphorylases [Enzyme Commission (EC) 2.4.2.1 classification] from bacteria (e.g., DEOD_ECOLI) and eukaryotes (e.g., PNPH_HUMAN) show no similarity to each other in a single-pass database search; furthermore, these enzymes have distinct oligomeric structures and only partially overlapping substrate specificities (Hammer-Jespersen 1983; Ealick and Bugg 1990). However, iterative database searches starting from either of these sequences retrieve the other one at a statistically significant level ($e < 10^{-4}$) within two or three iterations. The common ancestry of these enzymes is supported by the presence of a common nucleoside-binding motif (Mushegian and Koonin 1994) and similarities in the 3D structure (Mao et al. 1997). Similar motifs were also found in other distantly related enzymes with the same specificity, such as, for example, goose and chicken lysozymes or animal and plant glycogen (starch) synthases. A detailed discussion of the sequence motifs identified in the course of this project is beyond the scope of this paper and will be presented elsewhere.

For a number of distinct forms of enzymes that catalyze the same reaction, however, no sequence similarity could be detected, even at the level of subtle motifs. Overall, such apparently unrelated sequences were detected for 105 EC nodes out of the 1709 currently represented in GenBank (Tables 1–3). These figures do not include the numerous cases

of incorrectly assigned EC numbers, revealed in the course of this study (Galperin and Koonin 1998), and viral enzymes that may have a high substitution rate potentially hampering the detection of sequence similarity. For 18 of the 105 EC nodes that included candidate analogous enzymes, 3D structures were available for both forms. Structural comparisons using the SCOP and FSSP databases (Holm and Sander 1996a; Hubbard et al. 1997) showed that in 16 cases out of 18, the isoforms had different 3D folds, or even belonged to different structural classes [Table 1, and online supplement (Online Table 1, <http://www.genome.org>)]. Three enzymes, namely chloroperoxidase, cellulase, and lichenase, were each represented by three different structures (Table 1). Among the candidate analogous enzymes that actually had the same fold, only the two classes of aldolase belonged to the same SCOP family, and the two forms of peroxidase formed different families within a superfamily. In both cases, comparison of the respective 3D structures using the Dali method (Holm and Sander 1995) showed that the root mean square deviation (RMSD) of superimposed C α atoms in the structural alignment of the two isoforms was no less than 3 Å (Table 1).

In 18 additional cases, distinct structural folds for two different enzyme isoforms could be inferred on the basis of their sequence similarity to proteins with known 3D structures (Table 2). Same structural folds, suggesting possible common ancestry, were predicted for eight pairs of candidate analogous enzymes. To our knowledge, most of these predictions have not been reported previously. Some of them required multiple PSI-BLAST searches to demonstrate statistically significant similarity to structurally characterized proteins and relied to a large extent on the conservation of specific sequence motifs diagnostic of the respective protein families (Table 2; L. Aravind, M.Y. Galperin, and E.V. Koonin, unpubl.). Interestingly, inspection of the structural assignments for analogous enzymes shows recurrent patterns that involve some of the most common folds, such as several pairs of analogous oxidoreductases, in which one form has the Rossmann fold, and the other form has the TIM-barrel fold (Tables 1 and 2).

Altogether, for 34 of the 105 pairs of candidate analogous enzymes, it could be shown that different forms had distinct structural folds, which unequivocally shows that they are indeed evolutionarily unrelated; in contrast, for only 10 EC nodes, the same fold was detected. Thus, although our sequence comparisons could miss some important evolutionary relationships that potentially

Table 1. Dissimilar Enzymes Catalyzing the Same Biochemical Reactions^a

Enzyme activity (EC No.)	Taxonomic representation ^b			PDB entry	Structural folds ^c
	bacteria	archaea	eukaryotes		
Alcohol:NADP dehydrogenase (EC 1.1.1.2)	ADH_CLOBE DHSO_BACSU	ADH3_SULSO —	ADH1_ENTHI ALDX_HUMAN	1DEH 2ALR	different
Formate dehydrogenase (EC 1.2.1.2)	FDHF_ECOLI FDH_PESER	FDHA_METFO A64427	— FDH_NEUCR	1FDI 2NAD	
Dihydrofolate reductase (EC 1.5.1.3)	DYRA_ECOLI DYR2_ECOLI	DYR_HALVO —	DYR_HUMAN —	1DHF 1VIE	different
Peroxidase (EC 1.11.1.7)	— —	— —	PERM_HUMAN PER1_ARAHY	1MHL 1ARV	
Chloroperoxidase (EC 1.11.1.10)	PRXC_PSEPY — —	— — —	— PRXC_CALFU PRXC_CURIN	1BRO 1CPO 1VNC	different
Superoxide dismutase (EC 1.15.1.1)	SODC_ECOLI SODF_ECOLI	— SODF_SULAC	SODC_HUMAN SODM_HUMAN	1SPD 1ABM	
Protein-tyrosine phosphatase (EC 3.1.3.48)	PTPA_STRCO YOPH_YEREN	— —	PPAC_BOVIN PTN1_HUMAN	1PHR 2HNP	different
Cellulase (EC 3.2.1.4)	GUNA_CLOCE GUND_CLOTM —	— — —	GUNB_NEOPA GUN_PHAVU GUN1_TRIRE	1EDG 1CLC 1CEL	
Xylanase (EC 3.2.1.8)	XYNA_STRLI XYNA_BACCI	— —	S43846 XYN2_TRIRE	1XAS 1XNB	different
Chitinase (EC 3.2.1.14)	CHIA_SERMA YE15_HAEIN	— —	CHIT_BRUMA CHI1_ORYSA	1CTN 2BAA	
β -Galactosidase (EC 3.2.1.23)	BGAL_ECOLI BGLA_THEMA	— BGAM_SULSO	BGAL_KLULA BGLC_MAIZE	1BGL 1GOW	different
Lichenase (EC 3.2.1.73)	GUB_BACLI GUB_BACCI —	— — —	YG46_YEAST — GUB2_HORVU	1GBG 1CEM 1GHR	
β -Lactamase (EC 3.5.2.6)	AMPC_ENTCL BLAB_BACFR	— —	— —	2BLT 1ZNB	different
Fructose 1,6-bisphosphate aldolase (EC 4.1.2.13)	ALF_ECOLI ALF_STACA	— —	ALF_YEAST ALFA_HUMAN	1DOS 1FBA	
Carbonic anhydrase (EC 4.2.1.1)	CCMM_SYNP7 —	CAH_METTE —	— CAH1_HUMAN	1THJ 2CBA	different
Peptidyl-prolyl isomerase (EC 5.2.1.8)	FKBX_ECOLI CYPB_ECOLI	FKB1_METJA —	FKBP_HUMAN CYPB_HUMAN	1FKD 2CPL	
Chorismate mutase (EC 5.4.99.5)	PHEA_ECOLI CHMU_BACSU	Y246_METJA —	CHMU_YEAST —	1ECM 1COM	different
DNA topoisomerase I (EC 5.99.1.2)	TOP1_ECOLI —	TOPG_SULAC —	TOP3_YEAST TOP1_YEAST	1ECL 1OIS	

^aThe full version of the table, including homologs of the enzymes found in each of the sequenced genomes, is available as a WWW supplement at http://ncbi.nlm.nih.gov/Complete_Genomes.

^bThe proteins are listed under their SwissProt, GenBank, or Protein Data Base identifiers. The names of enzymes with experimentally demonstrated activity, shown in the first column, are in boldface type; the dash indicates absence of homologs in any of the sequenced genomes.

^cThe data are from SCOP [<http://scop.mrc-lmb.cam.ac.uk/scop> (Hubbard et al. 1997)] and FSSP [<http://www2.ebi.ac.uk/dali/fssp/fssp.html> (Holm and Sander 1996a)] databases. RMSD of superimposed C α atoms in the structural alignment of the two isoforms is from the FSSP database (Holm and Sander 1996a).

could be revealed by structural comparisons (Holm and Sander 1996b; Murzin 1996), the above observations indicate that with current methods,

the absence of detectable sequence similarity is a fairly reliable indicator of independent evolution.

Table 2. Fold Prediction for Analogous Enzymes^a

Enzyme activity (EC No.)	Taxonomic representation ^b			PDB ^b	Predicted structural folds ^c
	bacteria	archaea	eukaryotes		
3- α -hydroxysteroid dehydrogenase (EC 1.1.1.50)	JN0829 —	AF1207 —	HE27_HUMAN DIDH_RAT	1AHH 1RAL	Rossmann TIM-barrel
17- β -hydroxysteroid dehydrogenase (EC 1.1.1.62)	YBBO_ECOLI —	AF1207 —	DHB1_HUMAN A56424	1FDS 1RAL	Rossmann TIM-barrel
Protochlorophyllide reductase (EC 1.3.1.33)	BCHN_RHOCA slr0506	— —	CHLN_CHLRE PCR_ARATH	3MIN —	nitrogenase Fe-Mo Rossmann
Hydrogenase (EC 1.18.99.1)	PHFL_DESVH PHNL_DESVM CAT_ECOLI	FDHA_METFO FRHA_METTH —	1171117 — —	1FCA 1FRV 1CLA	ferredoxin-like Ni-Fe hydrogenase CoA-dependent acetyltransferases
Chloramphenicol acetyl-transferase (EC 2.3.1.28)	CAT4_ECOLI	MJ1064	YJV8_YEAST	1LXA	single-stranded β -helix
Gluconokinase (EC 2.7.1.12)	GNTK_ECOLI GNTK_BACSU	— AF1752	GNTK_SCHPO GLPK_YEAST	1DVR 1GLA	P-loop-containing actin-like ATPase
Diacylglycerol kinase (EC 2.7.1.107)	KDGL_ECOLI	—	—	—	not known; integral membrane protein
FAD synthase (EC 2.7.7.2)	RIBF_CORAM —	— MJ0973	KDGG_HUMAN YDR236c FAD1_YEAST	1CDL 1GSG 1GPM	phosphofructokinase Rossmann adenine nucleotide α -hydrolase
Glucan endo-1,3- β -glucosidase (EC 3.2.1.39)	E13B_BACCI —	— AF0876	1488257 E13L-TOBAC	1MAC 1GHR	ConA-like lectins/glucanases TIM-barrel
6-Phospho- β -glucosidase (EC 3.2.1.86)	BGLB_ECOLI CELF_ECOLI	BGAL_SULSO —	BGLC_MAIZE —	1PBG 1LDG	TIM-barrel Rossmann
Asparaginase (EC 3.5.1.1)	ASG1_ECOLI ASPG_FLAME	MJ0020 —	ASG1_YEAST ASPG_LUPAR	3ECA 1APY	glutaminase/asparaginase Ntn hydrolases
Apyrase, ATP-diphosphatase (EC 3.6.1.5)	USHA_ECOLI	AF0876	APY_AEDAE	1KBP	metallo-dependent phosphatases
Diadenosine tetra-phosphatase (EC 3.6.1.17)	— NTPA_ECOLI	— AF2200 AF2211	CD39_HUMAN 1546841 AP4A_HUMAN APH1_SCHPO	1DKG 1AKO 1MUT 1HXP	ribonuclease H-like DNase I-like NTP pyrophosphorylase HIT-like
Haloacetate dehalogenase (EC 3.8.1.3)	DEH1_MORSP	AF1706	HYES_HUMAN	1BRO	α/β -hydrolases
Prephenate dehydratase (EC 4.2.1.51)	DEH2_MORSP PHEA_ECOLI PHEC_PSEAE	MTH209 PHEA_METJA AF0231	1050822 PHA2_YEAST —	1JUD 1FCA 2LAO	haloacid dehalogenase ferredoxin-like periplasmic binding protein-like
DNA-(apurinic or apyrimidinic site) lyase (EC 4.2.99.18)	END3_ECOLI END4_ECOLI EX3_ECOLI	MJ0613 MTH1010 MTH212	END3_SCHPO APN1_YEAST APE1_HUMAN	2ABK 1DID 1AKO	DNA-glycosylase TIM-barrel DNase I-like
Phosphoglycerate mutase (EC 5.4.2.1)	PMG1_ECOLI PMGI_BACSU	— MTH1591	PMGE_HUMAN PMGI_TOBAC	3PGM 1ALK	phosphoglycerate mutase alkaline phosphatase
Lysine-tRNA ligase (EC 6.1.1.6)	SYK1_ECOLI BB0659	— 2645489	SYKC_YEAST —	1LYL 1GLN	class II aaRS synthetases adenine nucleotide α -hydrolase

^aAll designations are as in Table 1. Only the apparent orthologs (Tatusov et al. 1996, 1997) are included. Proteins from recently sequenced genomes, not yet included in SwissProt, are listed under their original identifiers. The source organisms are as follows: (JN0829) *Pseudomonas* sp.; (A56424), mouse; (1546841) *Rhodnius prolixus* (Sarkis et al. 1986); (2645489) *Methanococcus maripaludis* (Ibba et al. 1997b); (MJ) *Methanococcus jannaschii* (Bult et al. 1996); (MTH) *Methanobacterium thermoautotrophicum* (Smith et al. 1997); (AF) *Archaeoglobus fulgidus* (Klenk et al. 1997); (BB) *Borellia burgdorferi* (Fraser et al. 1997).

^bThe PDB codes in roman type indicate the structures that were used for fold prediction; italics indicate tentative fold predictions based on multiple iterative searches against the GenPept database. Only the folds of the catalytic domains are indicated.

^cFold names are from the SCOP database [(Hubbard et al. 1997) <http://scop.mrc-lmb.cam.ac.uk/scop>].

Table 3. Distribution of Analogous Enzymes Among Enzyme Classes

Enzyme class	Enzyme nodes in EC	Enzyme nodes in GenBank	Sequences with assigned EC numbers	Nodes containing dissimilar enzymes	
				two types of enzymes	more than two types of enzymes
Oxidoreductases	940	420	6754	20	3
Transferases	1029	439	6402	21	0
Hydrolases	1068	530	6600	19	15
Lyases	343	171	3192	11	5
Isomerases	149	70	990	6	1
Ligases	122	79	1265	3	1
Total	3651	1709	25,203	105	

Origin of Analogous Enzymes

The most likely mechanism for the evolution of analogous enzymes appears to be recruitment of existing enzymes that take over new functions by virtue of changed substrate specificity or a modified catalytic mechanism. Such a scenario could be inferred for about one half of the analogous enzyme sets by showing that at least one of them is homologous to a family of enzymes catalyzing a different, albeit related, reaction (Table 4). The argument for recruitment is particularly convincing when one of the analogous enzyme forms is found in a limited number of species, whereas the family to which it belongs is common. For example, gluconate kinase from *Bacillus subtilis* (GNTK_BACSU, EC 2.7.1.12) seems to have orthologs with the same activity only in other *Bacillus* species and is unrelated to gluconate kinases from other organisms. However, it belongs to a large kinase family that includes xylulose kinases and glycerol kinases from a variety of organisms (Galperin and Koonin 1998). Thus, the *Bacillus* gluconate kinase most likely evolved as the result of a duplication of a gene for xylulose or glycerol kinase in the Gram-positive lineage, which was followed by a shift in the substrate specificity. A similar mechanism is likely to account for the origin of the 2,3-bisphosphoglycerate-independent phosphoglycerate mutase from an alkaline phosphatase-like enzyme and of the second phosphofructokinase of *Escherichia coli* (PfkB) from an enzyme of the ribokinase family (Table 4).

Some of the analogous enzymes can be described as “second edition” (Doolittle et al. 1986), that is, enzymes that perform functions related to adaptation to new environments and life styles and usually have a limited phylogenetic distribution. In

cases like this, the recruitment hypothesis seems plausible for all analogous forms. For example, the three analogous apyrases (ATP-diphosphohydrolases, EC 3.6.1.5), typically extracellular or membrane-bound enzymes that are involved in such specific functions [e.g., prevention of blood clotting (Champagne et al. 1995)], seem to have been derived from three large, unrelated enzyme families, each with its own 3D fold (Tables 2 and 4).

When neither of the analogous isoforms has detectable sequence similarity to any enzymes of different specificity, the recruitment scenario may still apply, but the connection to the progenitor enzyme might have become undetectable owing to the recruitment having occurred very early in evolution and/or because of rapid change associated with the acquisition of the new function.

The Most Diverse Groups of Enzymes

Table 3 shows that analogous enzymes are more or less uniformly distributed among all enzyme classes. A closer examination of their functions, however, shows that the majority of analogous enzymes belongs to several functional groups. The most conspicuous one contains enzymes involved in synthesis and hydrolysis of polysaccharides (Davies and Henrissat 1995; Henrissat and Davies 1997). This group includes 51 different enzyme forms, belonging to 24 EC nodes, and contains glycosyl transferases, glycosyl hydrolases, and pectate and alginate lyases. The most striking examples are 6-phospho- β -glucosidase (EC 3.2.1.86), which is found in both *E. coli* and *B. subtilis* genomes in two unrelated forms, each represented by three paralogous genes, and cellulase (EC 3.2.1.4), found in six forms, belonging to at least three different folds (Table 1).

Table 4. Possible Origin of Analogous Enzymes by Recruitment of Enzymes with Related Activities

Enzyme (EC No.)	Analogous enzymes	Homologous enzymes with different activities (EC no., example)
Fructokinase (EC 2.7.1.4)	SCRK_ECOLI	ribokinase (EC 2.7.1.15, RBSK_ECOLI)
	SCRK_ZYMMO	glucokinase (EC 2.7.1.2, GLK_STRCO)
6-Phosphofructokinase (EC 2.7.1.11)	K6P1_ECOLI	PP _i -dependent 6-phosphofructokinase (EC 2.7.1.90, PFPB_SOLTU)
	K6P2_ECOLI	1-phosphofructokinase (EC 2.7.1.56, K1PF_ECOLI)
		ribokinase (EC 2.7.1.15, RBSK_ECOLI)
Gluconokinase (EC 2.7.1.12)	GNTK_ECOLI	—
	GNTK_BACSU	glycerol kinase (EC 2.7.1.30, GLPK_ECOLI)
Phosphatidylserine synthase (EC 2.7.8.8)	PSS_ECOLI	—
	PSS_YEAST	phosphatidylglycerophosphate synthase (EC 2.7.8.5, PGSA_ECOLI)
β-Galactosidase (EC 3.2.1.23)	BGAL_ECOLI	β-glucuronidase (EC 3.2.1.31, BGLR_HUMAN)
	BGAL_HUMAN	β-glucosidase (EC 3.2.1.21, BGLA_THEMEA)
Apyrase, ATP-diphosphatase (EC 3.6.1.5)	APY_AEDAE	5'-nucleotidase (EC 3.1.3.5, 5NTD_HUMAN)
	APY_SOLTU	nucleoside triphosphatase (EC 3.6.1.15, NTPA_PEA)
	1546841 ^a	inositol-1,4,5-triphosphate 5-phosphatase (EC 3.1.3.56, IT5P_HUMAN)
Diadenosine 5',5'''-tetrphosphatase (EC 3.6.1.17)	AP4A_HUMAN	NTP pyrophosphohydrolase (EC 3.6.1.-, NTPA_ECOLI)
	APH1_SCHPO	ATP adenylyltransferase (EC 2.7.7.53, APA1_YEAST)
Tagatose 1,6-bisphosphate aldolase (EC 4.1.2.40)	AGAY_ECOLI	fructose bisphosphate aldolase (EC 4.1.2.40, ALF_ECOLI)
	LACD_LACLA	—
Phosphoglycerate mutase (EC 5.4.2.1)	PMG1_ECOLI	fructose 2,6-bisphosphatase (EC 3.1.3.46, F26_YEAST)
	PMGI_BACSU	phosphopentomutase (EC 5.4.2.7, DEOB_BACSU)
	PMGI_TOBAC	alkaline phosphatase (EC 3.1.3.1, PPB_ECOLI)
Lysine-tRNA ligase (EC 6.1.1.6)	SYK1_ECOLI	aspartate-tRNA ligase (EC 6.1.1.12, SYD_HUMAN)
	2645489 ^b	cysteine-tRNA ligase (EC 6.1.1.16, SYC_ECOLI)

^aApyrase from *Rhodnius prolixus* (Sarkis et al. 1986).
^bLysine-tRNA ligase from *Methanococcus maripaludis* (Ibba et al. 1997b).

The second large group of analogous enzymes (nine EC nodes) deals with the effects of oxygen on cellular components. It includes, among others, cytochrome *c* peroxidase, catalase, peroxidase, chlo-

roperoxidase, superoxide dismutase, glutathione *S*-transferase, and glutathione synthetases I and II.

The third group (seven EC nodes) consists of enzymes involved in the synthesis and turnover of

bacterial and eukaryotic cell walls. It includes lysozyme (*N*-acetylmuramidase), hyaluronidase, penicillin acylase, *N*-acetylmuramoyl-L-alanine amidase, β -lactamase, phosphomannose isomerase, and phosphomannomutase.

Notably, these groups of enzymes are mostly involved in specific rather than universal cellular functions, suggesting that they might have evolved relatively recently.

Analogous Enzymes in Central Metabolism

Although less abundant than in functions like cell wall biosynthesis or cell defense, analogous enzymes can be found in a variety of metabolic pathways; however, no central pathways were detected that would consist exclusively of such enzymes. Rather, analogous enzymes typically are sandwiched between those that are universally conserved. Among glycolytic enzymes, phosphofructokinase, phosphoglycerate mutase, and lactate dehydrogenase have analogous forms, whereas glucokinase and aldolase are represented by highly divergent, albeit structurally similar, forms, in bacteria and eukaryotes. The presence of analogous enzymes in the early steps of glycolysis and the abundance of analogous enzymes catalyzing other reactions of hexose metabolism support the conclusion that only the lower part of glycolysis from glyceraldehyde 3-phosphate to pyruvate is ubiquitous and indispensable for life and that the original function of the Embden–Meyerhof pathway could have been biosynthetic, rather than glycolytic (Romano and Conway 1996).

The same patchy distribution of analogous enzymes can be seen in reactions of amino acid metabolism. Unrelated forms of asparaginase, 3-dehydroquinate dehydratase, and arginine decarboxylase are found in pathways that can serve both biosynthetic and biodegradative functions. Predictably, biosynthetic forms are mostly present in autotrophs, for example, cyanobacteria and plants, whereas biodegradative forms are typical of heterotrophs. In two of the three cases above, *E. coli* encodes both versions of the enzyme.

Analogous Enzymes in DNA Replication, Transcription, and Translation

DNA replication systems present prominent examples of analogous enzymes, although these are not easily captured by the automatic procedure used here owing to the multisubunit and/or multidomain structure of the key components (e.g., DNA

polymerases) and the difficulties of their placement within the framework of current enzyme classification (e.g., lack of an appropriate rule to distinguish DNA-dependent RNA polymerases involved in transcription and DNA primases). None of the essential DNA replication enzymes are orthologous in bacteria as compared with archaea and eukaryotes, leading to the speculation that the last common ancestor of all extant life forms might not have had a DNA genome at all (Mushegian and Koonin 1996). In particular, no sequence similarity can be detected between the dNTP polymerization domains of the replicative polymerases (DNA polymerase III α -subunit in bacteria and B family DNA polymerases in archaea and eukaryotes), suggesting that these central components of the DNA replication machinery are indeed analogous; the final conclusion, however, should await the determination of polymerase III 3D structure. A striking example of analogy in replicative enzymes that was automatically detected by our procedure is type I DNA topoisomerases, with two unrelated families including bacterial topoisomerases I together with bacterial and eukaryotic topoisomerase III, and eukaryotic topoisomerase I, respectively (Table 1). Another crucial step in DNA replication in bacteria and archaea/eukaryotes is catalyzed by apparently unrelated DNA ligases, though the EC numbers are different in this case as NAD and ATP, respectively, are used as substrates.

Although the transcription and translation machineries are generally uniform in all life forms and analogous enzymes are not typical, there are notable exceptions. Thus, the archaeal lysyl-tRNA synthetase, orthologs of which have been unexpectedly discovered in spirochaetes, belongs to aminoacyl-tRNA synthetase class I as opposed to the class II lysyl-tRNA synthetases found in the rest of bacteria and in eukaryotes (Ibba et al. 1997b). In this case, unrelated 3D folds are obvious, and evolution by recruitment of a class I enzyme appears most likely (Koonin and Aravind 1998).

Analogous Enzymes in Signal Transduction

As enzymes with functions in signal transduction may have originally evolved from metabolic enzymes, independent inventions in this area seem particularly likely. Because cAMP plays substantially different roles in bacterial and eukaryotic cells (Saier 1996), it is perhaps unsurprising that its synthesis is catalyzed by apparently unrelated enzymes in bacteria and eukaryotes. Pertussis and anthrax toxins also have adenylate cyclase activity, constituting yet another, third class of adenylate cyclases.

The two analogous enzymes that hydrolyze the intracellular signal molecule diadenosine tetraphosphate have been apparently recruited from two ancient classes of hydrolases, namely the HIT superfamily (APH1_SCHPO) and the NTP pyrophosphohydrolase (MutT) superfamily (AP4A_HUMAN) (Table 4).

The epitome of regulatory enzymes, serine-threonine protein kinases, exist in three unrelated forms. Although the great majority of them have the classical protein kinase fold (typified by Src), bacterial protein kinases containing P loops have been described recently (Galinier et al. 1998; Reizer et al. 1998), and furthermore, histidine kinase homologs have been described that phosphorylate specific serines in their target proteins (Popov et al. 1993; Yang et al. 1996). Although not detected automatically by our procedure owing to problems with assigning EC numbers, this is a striking example of apparent evolution of analogous enzymes by recruitment.

Bacteria and eukaryotes use analogous enzymes to synthesize glutathione, the universal regulator of the thiol-disulfide ratio in the cell. Both reactions of glutathione biosynthesis in *E. coli* are catalyzed by enzymes, namely glutamate-cysteine ligase (EC

6.3.2.2) and glutathione synthetase (EC 6.3.2.3), that are unrelated to the respective enzymes from yeast and humans. Remarkably, plants have a typical eukaryotic glutathione synthetase, but their glutamate-cysteine ligase is unrelated to either bacterial or yeast enzyme.

DISCUSSION

Phylogenetic Distribution of Analogous Enzymes

Comparison of microbial genomes has shown that species with larger genomes contain multiple paralogous genes, coding for enzymes with similar catalytic properties. In organisms with small genomes, not only the absolute number but also the fraction of proteins that belong to paralogous families is considerably lower (Koonin et al. 1997). In the same vein, the data in Table 5 show that organisms with small genomes, both parasitic and free-living ones, encode disproportionately small numbers of analogous enzymes. This observation is confirmed by the analysis of specific metabolic pathways such as glycolysis and purine biosynthesis, in which organisms with larger genomes have analogous enzymes for certain steps, whereas organisms with

Table 5. Analogous Enzymes in Completely Sequenced Genomes

Organism	Proteins encoded in the complete genome ^a	EC nodes with two analogous enzyme forms in the same genome
Bacteria		
<i>Escherichia coli</i>	4289	35
<i>Haemophilus influenzae</i>	1717	8
<i>Helicobacter pylori</i>	1566	4
<i>Synechocystis</i> sp.	3169	18
<i>Borrelia burgdorferi</i>	850	2
<i>Bacillus subtilis</i>	4100	30
<i>Mycoplasma genitalium</i>	467	0
<i>Mycoplasma pneumoniae</i>	677	0
Archaea		
<i>Methanococcus jannaschii</i>	1715	1
<i>Methanobacterium thermoautotrophicum</i>	1869	5
<i>Archaeoglobus fulgidus</i>	2407	3
Eukaryotes		
<i>Saccharomyces cerevisiae</i>	5932	22
<i>Caenorhabditis elegans</i> ^b	12178	17

^aThe numbers refer to the number of ORFs in the latest updates of the GenBank genomes division (ftp://ncbi.nlm.nih.gov/genbank/genomes).

^bThis genome is not yet completed; the data relate to the available portion of the genome as listed in wormpep12 database (http://www.sanger.ac.uk/Projects/C_elegans).

small genomes typically have only one form (Koonin et al. 1998). Thus, biochemical diversity, manifest in the variety of both analogs and paralogs, is a luxury enjoyed mostly by organisms with large genomes.

Analogous Enzymes and Enzyme Classification

We showed that numerous biochemical reactions may be catalyzed by enzymes without detectable sequence similarity and, in some cases, with demonstrably distinct 3D structures. In other words, many enzymatic activities have been independently invented in evolution on more than one occasion. This is not to say that these enzymes have nothing in common—although not sharing common ancestry, they still may have similar reaction mechanisms and even similar local active center geometries. Identification of such common features in analogous enzymes seems to be an interesting direction for future research that may shed new light on mechanisms of enzymatic catalysis.

The current classification of enzymes on the basis of the catalyzed reactions is an indispensable tool for enzymologists, but in the case of analogous enzymes, it inevitably fails. A hierarchical system of protein classification constructed by sequence and structure comparison, like the ones already proposed for peptidases (Barrett 1994) and glycosidases (Henrissat and Davies 1997), can handle these cases and will provide a wealth of information complementary to the information currently embodied in the EC system.

Implications for Genome Annotation and Drug Design

Functional annotations of proteins identified in the course of genome sequencing projects rely mostly on sequence similarities between newly identified proteins and those already in the databases. Although this process can produce many important insights into structure, evolution, and catalytic properties of various proteins (e.g., Galperin and Koonin 1997; Aravind et al. 1998), it can also lead to misannotations, which tend to spread as newly sequenced proteins are assigned functions based on their similarity to previously misannotated proteins (Bhatia et al. 1997). Enzyme recruitment, leading to a significant change of function accomplished by relatively minor sequence changes, is a phenomenon that can significantly affect the quality of similarity-based functional annotation. The ex-

amples of enzyme recruitment, identified here (Table 1) and available in the WWW supplement, may be useful to prevent such erroneous annotations.

In many cases, the phylogenetic distribution of analogous enzymes is such that one enzyme form is found in bacteria and the other one in eukaryotes. The enzyme forms that are absent in eukaryotes (particularly humans) can be used as targets for the development of new antibacterial drugs that would have a low chance of side effects. This strategy may be especially promising for targeting pathogenic bacteria as their genomes generally have a low level of enzyme redundancy and code for only one form of the analogous enzymes (Table 5).

METHODS

Identification of analogous enzymes was based on the fact that under the IUBMB Nomenclature Commission rules (1992), each complete EC number (node) specifies one particular reaction [peptidases, EC 3.4.*, classified on a different principle (Barrett 1994), were excluded from consideration]. Analogous enzymes were therefore identified as enzymes with the same EC numbers that had no detectable sequence similarity to each other.

Protein sequences with assigned complete, four-digit EC numbers were extracted from the GenBank using the Entrez search engine (Schuler et al. 1996). Sequences containing <100 amino acid residues were discarded as these typically are fragments, and each of the remaining sequences was compared to all the sequences with the same EC number using the gapped BLAST program (Altschul and Gish 1996). Sequences that had similarity scores above a cutoff of 120 (corresponding to the expectation value $-e < 0.001$ in a search of the complete nonredundant protein sequence database at the NCBI) were grouped by single-linkage clustering, and only one sequence from each such cluster was selected and considered further. The EC nodes with only one cluster were represented by a single sequence in the resulting data set and, accordingly, were excluded from subsequent analysis. The EC nodes represented by two or more proteins were further analyzed by comparing the sequences to the nonredundant protein database using the iterative gapped BLAST (PSI-BLAST) program (Altschul et al. 1997). The EC nodes for which PSI-BLAST (run to convergence) or motif analysis using the MoST program (Tatusov et al. 1994) detected any appreciable ($e < 0.1$) similarity between the included protein sequences were also removed from the data set.

The final analysis of the data was performed by manual elimination of the sequences that did not satisfy the criteria for analogous enzymes, primarily proteins with apparently incorrectly assigned EC numbers, undocumented enzymatic activity, and different subunits of the same enzyme (for additional details, see Galperin and Koonin 1998).

Structural folds were predicted by similarity to proteins with known 3D structure, which was detected by iteratively searching the nonredundant protein database using the PSI-BLAST program and extracting from the search output the sequences associated with the Protein Data Bank (PDB) acces-

GALPERIN ET AL.

sion numbers. In some cases, the prediction involved additional iterative searches from alternative starting points.

The taxonomic distribution of the analogous enzymes was deduced by analyzing the PSI-BLAST outputs using the BLATax program (Koonin et al. 1996b) and additionally, by comparing the sequences to the sets of proteins from complete microbial genomes (Fleischmann et al. 1995; Bult et al. 1996; Goffeau et al. 1996; Himmelreich et al. 1996; Kaneko et al. 1996; Blattner et al. 1997; Fraser et al. 1997; Klenk et al. 1997; Kunst et al. 1997; Smith et al. 1997; Tomb et al. 1997) and the available portion of the *Caenorhabditis elegans* genome.

Data processing and analysis were automated using the SEALS package (Walker and Koonin 1997). The complete listing of the analogous enzymes identified in this study is available on the Internet (www.ncbi.nlm.nih.gov/Complete_Genomes/).

ACKNOWLEDGMENTS

We thank L. Aravind for help with fold predictions and Drs. Amos Bairoch and Keith F. Tipton for helpful discussion.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Altschul, S.F. and W. Gish. 1996. Local alignment statistics. *Methods Enzymol.* 266: 460–480.
- Altschul, S.F., T.L. Madden, A.A. Schaffer, J. Zhang, Z. Zheng, W. Miller, and D.J. Lipman. 1997. Gapped BLAST and PSI-BLAST—A new generation of protein database search programs. *Nucleic Acids Res.* 25: 3389–3402.
- Aravind, L., M.Y. Galperin, and E.V. Koonin. 1998. The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. *Trends Biochem. Sci.* 23: 127–129.
- Barrett, A.J. 1994. Classification of peptidases. *Methods Enzymol.* 244: 1–15.
- Bhatia, U., K. Robison, and W. Gilbert. 1997. Dealing with database explosion: A cautionary note. *Science* 276: 1724–1725.
- Blattner, F.R., G. Plunkett III, C.A. Bloch, N.T. Perna, V. Burland, M. Riley, J. Collado-Vides, J.D. Glasner, C.K. Rode, G.F. Mayhew et al. 1997. The complete genome sequence of *Escherichia coli* K-12. *Science* 277: 1453–1474.
- Blom, N.S., S. Tetreault, R. Coulombe, and J. Sygusch. 1996. Novel active site in *Escherichia coli* fructose 1,6-bisphosphate aldolase. *Nat. Struct. Biol.* 3: 856–862.
- Bult, C.J., O. White, G.J. Olsen, L. Zhou, R.D. Fleischmann, G.G. Sutton, J.A. Blake, L.M. FitzGerald, R.A. Clayton, J.D. Gocayne et al. 1996. Complete genome sequence of the methanogenic archaeon, *Methanococcus jannaschii*. *Science* 273: 1058–1073.
- Carfi, A., S. Pares, E. Duee, M. Galleni, C. Duez, J.M. Frere, and O. Dideberg. 1995. The 3-D structure of a zinc metallo-beta-lactamase from *Bacillus cereus* reveals a new type of protein fold. *EMBO J.* 14: 4914–4921.
- Champagne, D.E., C.T. Smartt, J.M. Ribeiro, and A.A. James. 1995. The salivary gland-specific apyrase of the mosquito *Aedes aegypti* is a member of the 5'-nucleotidase family. *Proc. Natl. Acad. Sci.* 92: 694–698.
- Cooper, S.J., G.A. Leonard, S.M. McSweeney, A.W. Thompson, J.H. Naismith, S. Qamar, A. Plater, A. Berry, and W.N. Hunter. 1996. The crystal structure of a class II fructose-1,6-bisphosphate aldolase shows a novel binuclear metal-binding active site embedded in a familiar fold. *Structure* 4: 1303–1315.
- Davies, G. and B. Henrissat. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure* 3: 853–859.
- Doolittle, R.F. 1987. *Of URFs and ORFs: A primer on how to analyze derived amino acid sequences*. University Science Books, Mill Valley, CA.
- . 1994. Convergent evolution: The need to be explicit. *Trends Biochem. Sci.* 19: 15–18.
- Doolittle, R.F., D.F. Feng, M.S. Johnson, and M.A. McClure. 1986. Relationships of human protein sequences to those of other organisms. *Cold Spring Harbor Symp. Quant. Biol.* 51: 447–455.
- Ealick, S.E. and C.E. Bugg. 1990. Three-dimensional structure of human erythrocytic purine nucleoside phosphorylase at 3.2 Å resolution. *J. Biol. Chem.* 265: 1812–1820.
- Fitch, W.M. 1970. Distinguishing homologous from analogous proteins. *Syst. Zool.* 19: 99–113.
- Fleischmann, R.D., M.D. Adams, O. White, R.A. Clayton, E.F. Kirkness, A.R. Kerlavage, C.J. Bult, J.F. Tomb, B.A. Dougherty, and J.M. Merrick. 1995. Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269: 496–512.
- Florkin, M. 1974. Concepts of molecular biosemiotics and of molecular evolution. In *Comprehensive biochemistry* (ed. M. Florkin and E.H. Stolz), Vol. 29A, pp. 1–124. Elsevier, Amsterdam, Netherlands.
- Fothergill-Gilmore, L.A. and P.A. Michels. 1993. Evolution of glycolysis. *Prog. Biophys. Mol. Biol.* 59: 105–235.
- Fraser, C.M., S. Casjens, W.M. Huang, G.G. Sutton, R. Clayton, R. Lathigra, O. White, K.A. Ketchum, R. Dodson, E.K. Hickey et al. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 390: 580–586.
- Galinier, A., M. Kravanja, R. Engelmann, W. Hengstenberg, M.C. Kilhoffer, J. Deutscher, and J. Haiech. 1998. New protein kinase and protein phosphatase families mediate

- signal transduction in bacterial catabolite repression. *Proc. Natl. Acad. Sci.* 95: 1823–1828.
- Galperin, M.Y. and E.V. Koonin. 1997. A diverse superfamily of enzymes with ATP-dependent carboxylate-amine/thiol ligase activity. *Protein Sci.* 6: 2639–2643.
- . 1998. Sources of systematic error in functional annotation of genomes: Domain rearrangement, non-orthologous gene displacement, and operon disruption. *In Silico Biol.* 1: 0007. <http://www.bioinfo.de/isb/1998/01/0007>.
- Goffeau, A., B.G. Barrell, H. Bussey, R.W. Davis, B. Dujon, H. Feldmann, F. Galibert, J.D. Hoheisel, C. Jacq, M. Johnston et al. 1996. Life with 6000 genes. *Science* 274: 546, 563–567.
- Hammer-Jespersen, K. 1983. Nucleoside catabolism. In *Metabolism of nucleotides, nucleosides and nucleobases in microorganisms* (ed. A. Munch-Petersen), pp. 203–258. Academic Press, London, UK.
- Henrissat, B. and G. Davies. 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr. Opin. Struct. Biol.* 7: 637–644.
- Himmelreich, R., H. Hilbert, H. Plagens, E. Pirkil, B.C. Li, and R. Herrmann. 1996. Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*. *Nucleic Acids Res.* 24: 4420–4449.
- Holm, L. and C. Sander. 1995. Dali: A network tool for protein structure comparison. *Trends Biochem. Sci.* 20: 478–480.
- . 1996a. The FSSP database: Fold classification based on structure-structure alignment of proteins. *Nucleic Acids Res.* 24: 206–209.
- . 1996b. Mapping the protein universe. *Science* 273: 595–603.
- Hubbard, T.J.P., A.G. Murzin, S.E. Brenner, and C. Chothia. 1997. SCOP: A structural classification of proteins database. *Nucleic Acids Res.* 25: 236–239.
- Ibba, M., J.L. Bono, P.A. Rosa, and D. Soll. 1997a. Archaeal-type lysyl-tRNA synthetase in the Lyme disease spirochete *Borrelia burgdorferi*. *Proc. Natl. Acad. Sci.* 94: 14383–14388.
- Ibba, M., S. Morgan, A.W. Curnow, D.R. Pridmore, U.C. Vothknecht, W. Gardner, W. Lin, C.R. Woese, and D. Soll. 1997b. A euryarchaeal lysyl-tRNA synthetase: Resemblance to class I synthetases. *Science* 278: 1119–1122.
- IUBMB Nomenclature Commission. 1992. *Enzyme nomenclature 1992*. Academic Press, San Diego, CA.
- Kaneko, T., S. Sato, H. Kotani, A. Tanaka, E. Asamizu, Y. Nakamura, N. Miyajima, M. Hirose, M. Sugiura, S. Sasamoto 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: 185–209.
- Klenk, H.P., R.A. Clayton, J.-F. Tomb, O. White, K.E. Nelson, K.A. Ketchum, R.J. Dodson, M. Gwinn, E.K. Hickey, J.D. Peterson et al. 1997. The complete genome sequence of the hyperthermophilic, sulphate-reducing archaeon *Archaeoglobus fulgidus*. *Nature* 390: 364–370.
- Koonin, E.V. and L. Aravind. 1998. Re-evaluation of translation machinery evolution. *Curr. Biol.* 8: R266–R269.
- Koonin, E.V., A.R. Mushegian, and P. Bork. 1996a. Non-orthologous gene displacement. *Trends Genet.* 12: 334–336.
- Koonin, E.V., R.L. Tatusov, and K.E. Rudd. 1996b. Protein sequence comparison at genome scale. *Methods Enzymol.* 266: 295–322.
- Koonin, E.V., A.R. Mushegian, M.Y. Galperin, and D.R. Walker. 1997. Comparison of archaeal and bacterial genomes: Computer analysis of protein sequences predicts novel functions and suggests a chimeric origin for the archaea. *Mol. Microbiol.* 25: 619–637.
- Koonin, E.V., R.L. Tatusov, and M.Y. Galperin. 1998. Beyond the complete genomes: From sequences to structure and function. *Curr. Opin. Struct. Biol.* 8: 355–363.
- Kunst, F., N. Ogasawara, I. Moszer, A.M. Albertini, G. Alloni, V. Azevedo, M.G. Bertero, P. Bessieres, A. Bolotin, S. Borchert et al. 1997. The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* 390: 249–256.
- Lobkovsky, E., P.C. Moews, H. Liu, H. Zhao, J.M. Frere, and J.R. Knox. 1993. Evolution of an enzyme activity: Crystallographic structure at 2-Å resolution of cephalosporinase from the *ampC* gene of *Enterobacter cloacae* P99 and comparison with a class A penicillinase. *Proc. Natl. Acad. Sci.* 90: 11257–11261.
- Mao, C., W.J. Cook, M. Zhou, G.W. Kozalka, T.A. Krenitsky, and S.E. Ealick. 1997. The crystal structure of *Escherichia coli* purine nucleoside phosphorylase: A comparison with the human enzyme reveals a conserved topology. *Structure* 5: 1373–1383.
- Marsh, J.J. and H.G. Leberer. 1992. Fructose-bisphosphate aldolases: An evolutionary history. *Trends Biochem. Sci.* 17: 110–113.
- Murzin, A.G. 1996. Structural classification of proteins: New superfamilies. *Curr. Opin. Struct. Biol.* 6: 386–394.
- Mushegian, A.R. and E.V. Koonin. 1994. Unexpected sequence similarity between nucleosidases and phosphoribosyltransferases of different specificity. *Protein Sci.* 3: 1081–1088.
- . 1996. A minimal gene set for cellular life derived by

GALPERIN ET AL.

- comparison of complete bacterial genomes. *Proc. Natl. Acad. Sci.* 93: 10268–10273.
- Perham, R.N. 1990. The fructose-1,6-bisphosphate aldolases: Same reaction, different enzymes. *Biochem. Soc. Trans.* 18: 185–187.
- Popov, K.M., N.Y. Kedishvili, Y. Zhao, Y. Shimomura, D.W. Crabb, and R.A. Harris. 1993. Primary structure of pyruvate dehydrogenase kinase establishes a new family of eukaryotic protein kinases. *J. Biol. Chem.* 268: 26602–26606.
- Reizer, J., C. Hoischen, F. Titgemeyer, C. Rivolta, R. Rabus, J. Stulke, D. Karamata, M.H. Saier Jr., and W. Hillen. 1998. A novel protein kinase that controls carbon catabolite repression in bacteria. *Mol. Microbiol.* 27: 1157–1169.
- Romano, A.H. and T. Conway. 1996. Evolution of carbohydrate metabolic pathways. *Res. Microbiol.* 147: 448–455.
- Rutter, W.J. 1964. Evolution of aldolase. *Fed. Proc.* 23: 1248–1257.
- Saier, M.H., Jr. 1996. Regulatory interactions controlling carbon metabolism: An overview. *Res. Microbiol.* 147: 439–447.
- Sarkis, J.J., J.A. Guimaraes, and J.M. Ribeiro. 1986. Salivary apyrase of *Rhodnius prolixus*. Kinetics and purification. *Biochem. J.* 233: 885–891.
- Schuler, G.D., J.A. Epstein, H. Ohkawa, and J.A. Kans. 1996. Entrez: Molecular biology database and retrieval system. *Methods Enzymol.* 266: 141–162.
- Smith, D.R., L.A. Doucette-Stamm, C. Deloughery, H. Lee, J. Dubois, T. Aldredge, R. Bashirzadeh, D. Blakely, R. Cook, K. Gilbert et al. 1997. Complete genome sequence of *Methanobacterium thermoautotrophicum* deltaH: Functional analysis and comparative genomics. *J. Bacteriol.* 179: 7135–7155.
- Smith, M.W., D.F. Feng, and R.F. Doolittle. 1992. Evolution by acquisition: The case for horizontal gene transfers. *Trends Biochem. Sci.* 17: 489–493.
- Stallings, W.C., T.B. Powers, K.A. Pattridge, J.A. Fee, and M.L. Ludwig. 1983. Iron superoxide dismutase from *Escherichia coli* at 3.1-Å resolution: A structure unlike that of copper/zinc protein at both monomer and dimer levels. *Proc. Natl. Acad. Sci.* 80: 3884–3888.
- Tatusov, R.L., S.F. Altschul, and E.V. Koonin. 1994. Detection of conserved segments in proteins: Iterative scanning of sequence databases with alignment blocks. *Proc. Natl. Acad. Sci.* 91: 12091–12095.
- Tatusov, R.L., A.R. Mushegian, P. Bork, N.P. Brown, W.S. Hayes, M. Borodovsky, K.E. Rudd, and E.V. Koonin. 1996. Metabolism and evolution of *Haemophilus influenzae* deduced from a whole-genome comparison with *Escherichia coli*. *Curr. Biol.* 6: 279–291.
- Tatusov, R.L., E.V. Koonin, and D.J. Lipman. 1997. A genomic perspective on protein families. *Science* 278: 631–637.
- Tomb, J.-F., O. White, A.R. Kerlavage, R.A. Clayton, G.G. Sutton, R.F. Fleishmann, K.A. Ketchum, H.P. Klenk, S. Gill, B.A. Dougherty et al. 1997. The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* 388: 539–547.
- Walker, D.R. and E.V. Koonin. 1997. SEALS: A system for easy analysis of lots of sequences. *ISMB* 5: 333–339.
- Warburg, O. and W. Christian. 1943. Isolierung und kristallization des garungsferments zymohexase. *Biochem. Z.* 314: 149–176.
- Wright, C.S., R.A. Alden, and J. Kraut. 1969. Structure of subtilisin BPN' at 2.5 Å resolution. *Nature* 221: 235–242.
- Yang, X., C.M. Kang, M.S. Brody, and C.W. Price. 1996. Opposing pairs of serine protein kinases and phosphatases transmit signals of environmental stress to activate a bacterial transcription factor. *Genes & Dev.* 10: 2265–2275.

Received May 28, 1998; accepted in revised form July 10, 1998.

Analogous Enzymes: Independent Inventions in Enzyme Evolution

Michael Y. Galperin, D. Roland Walker and Eugene V. Koonin

Genome Res. 1998 8: 779-790

Access the most recent version at doi:[10.1101/gr.8.8.779](https://doi.org/10.1101/gr.8.8.779)

References

This article cites 61 articles, 21 of which can be accessed free at:
<http://genome.cshlp.org/content/8/8/779.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

**Affordable, Accurate
Sequencing.**

To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>