Metabolites: a helping hand for pathway evolution?

Steffen Schmidt[1,2,5], Shamil Sunyaev[1,3], Peer Bork[1,5] and Thomas Dandekar[1,4,5]

[1]European Molecular Biology Laboratory Heidelberg, Postfach 102209, D-69012 Heidelberg, Germany
[2]University of Heidelberg, Department of Parasitology, INF 324, D-69120 Heidelberg, Germany
[3]Present address: Genetics Division, Department of Medicine, Brigham & Women's Hospital and Harvard Medical School, 20 Shattuck Str. Thorn 10, Boston MA 02115, USA
[4]University of Würzburg, Department of Bioinformatics, Biozentrum, Am Hubland, D-97074 Würzburg, Germany
[5]Max Delbrück-Centrum für Molekulare Medizin, Robert-Rössle-Str. 10, D-13122 Berlin-Buch, Germany

The evolution of enzymes and pathways is under debate. Recent studies show that recruitment of single enzymes from different pathways could be the driving force for pathway evolution. Other mechanisms of evolution, such as pathway duplication, enzyme specialization, de novo invention of pathways or retro-evolution of pathways, appear to be less abundant. Twenty percent of enzyme superfamilies are quite variable, not only in changing reaction chemistry or metabolite type but in changing both at the same time. These variable superfamilies account for nearly half of all known reactions. The most frequently occurring metabolites provide a helping hand for such changes because they can be accommodated by many enzyme superfamilies. Thus, a picture is emerging in which new pathways are evolving from central metabolites by preference, thereby keeping the overall topology of the metabolic network.

It is well known that enzymes are specific and that they catalyze individual reactions with surprising accuracy and speed. However, for adaptation and evolution the opposite is required: new substrates must be recognized and new enzyme activities evolved. But where does the flexibility and plasticity of metabolic pathways and new enzyme activities come from? The majority of recent studies have concentrated on the level of pathways or enzymes as well as the variability of enzymes in their reaction chemistry [1–14]. However, the variability in substrates and products – metabolites – must also be considered.

Pathway evolution theories

On the level of pathway evolution, several hypotheses have been proposed (Fig. 1). First, pathways might have evolved spontaneously without adopting existing enzymes (Fig. 1a). For example, different tRNA synthetases seem to have initially evolved independently and then later have become involved in different pathways such as protein translation, tRNA dependent transamidation and non-discriminating acylation [15]. Second, the hypothesis of 'retro-evolution' of pathways [16,17] proposes that the selective pressure on a pathway mainly targets the successful production of its end-product (Fig. 1b). The formation of the required end-product from an intermediate metabolite increases the fitness of the organism. As the end-product can be derived from more and more 'distant' metabolites, fitness increases and the pathway evolves backwards. This retro-evolution has been proposed for both the glycolytic [18] and the mandelate pathway [19]. Third, pathways might have evolved from multifunctional enzymes [20] (Fig. 1c). Starting from a multifunctional enzyme catalyzing consecutive steps, the pathway might have then evolved by duplication and diversification of this precursor enzyme to the more specific and efficient enzymes known today, which catalyze only one step each in the pathway. O'Brien and Herschlag [21] analyzed several enzymes with alternative reactions distinct from their normal biological reaction to support the concept that broader substrates and reaction specificities are subsequently captured by adaptive evolution. Existing multifunctional enzymes, such as the carbamoyl phosphate synthase, are already used in diverse functions and pathways, such as β-D-glucan hydrolases in higher plants, and might be precursors to new pathways [6]. This hypothesis assumes that a single enzyme becomes specialized, but there is also the possibility that whole pathways (as a unit) became duplicated and diverted (Fig. 1d). This mechanism of acquiring new function has been examined for a long time [22] and can be readily identified using comparative genomics [10,23]. Examples include tryptophan and histidine biosynthesis [24,25]; these two pathways consist of several steps that have similar reaction chemistry and that are catalyzed by homologous enzymes – probably the result of early pathway duplication. Finally, pathways might have evolved by 'recruiting' enzymes from existing pathways, resulting in a mosaic or 'patchwork' of homologous enzymes that catalyze reactions in distinct pathways [25–27] (Fig. 1e). Observations indicate that one type of enzyme fold (e.g. TIM barrel [4]) or one enzyme superfamily [7] might catalyze similar reactions, but occur in different pathways owing to widespread recruitment. Such versatility has been found in many *Escherichia coli* small molecule metabolism enzymes [12,13].

Corresponding author: Thomas Dandekar (Dandekar@EMBL.de).

http://tibs.trends.com 0968-0004/03/$ - see front matter doi:10.1016/S0968-0004(03)00114-2

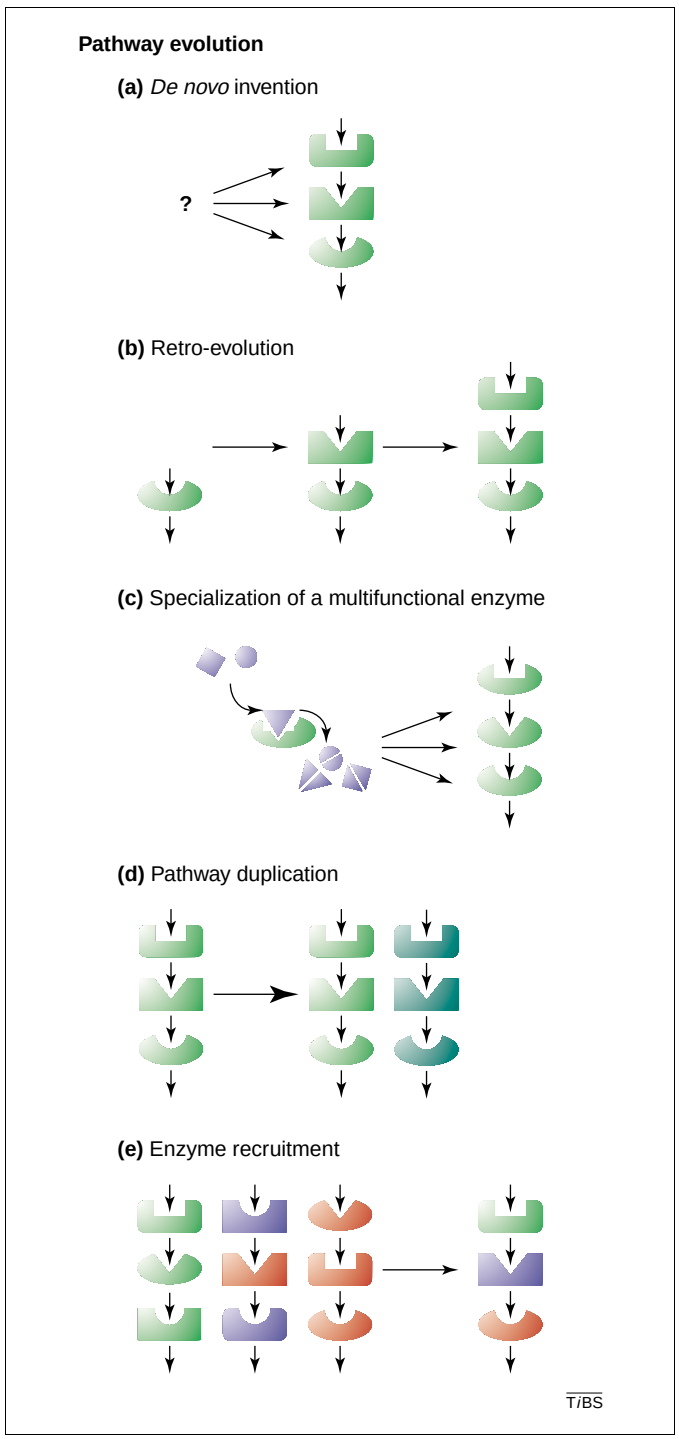

Fig. 1. Models of pathway evolution. Pathways might have evolved in different ways: (a) *de novo*, all reactions evolved independently from an unknown origin; (b) backwards, retro-evolution; (c) by specialization of a multifunctional enzyme, with multiple substrates shown in blue; (d) duplication of the complete pathway; or (e) by recruitment of enzymes from different pathways.

When comparing more recent studies [4,5,9,10,12–14,18,24,28,29], enzyme recruitment seems to be the main driving force for the evolution of new pathways, followed by specific pathway duplications and then by other, more rarely observed mechanisms. A good example for the evolution of pathways by enzyme variability involves two enzymes from two different amino-acid biosynthesis pathways. Sterner and colleagues [30] were able to show that by a single amino-acid substitution, an enzyme from the histidine biosynthesis (HisA) acquired the additional activity of an enzyme from the tryptophan biosynthesis (TrpF). This result clearly supports the hypothesis that those pathways are ancient duplicates of a common pathway. However, to appreciate the potential for variability in enzymes regarding long evolutionary time scales, a more comparative view of all enzymes is required.

Enzyme variability: reaction change or metabolite change?

Comparative studies on enzyme variability are often based on the definition of enzyme superfamilies – enzymes of common origin that can be identified by sequence and structural homology. The structural homology of an enzyme is important because the position of the catalytic residues in the structure, and the existence and form of different binding clefts is essential for its function and is therefore better conserved than its sequence. Hence, enzyme structure classification databases such as CATH (class, architecture, topology and homologous superfamilies) [31] or SCOP (structural classification of proteins) [32] are very helpful for the examination of enzyme evolution. For example, remote-sequence homologs such as actin (the ATPase domain of the heat shock protein) and hexokinase form an enzyme superfamily – the actin-like ATPase domain – owing to their common structure and probable common origin [33].

Using these defined enzyme superfamilies the variability of enzymes can be examined. The function of an enzyme can be altered either by changing its reaction chemistry but keeping substrate-binding pockets such as for the nucleotide cofactors NADH and FAD (Fig. 2a), or by changing its substrate specificity (Fig. 2b). The four-digit enzyme commission classification scheme (EC; Box 1) has been widely used to examine the variability in the reaction chemistry [34,35]. As shown in Fig. 3a, many enzyme

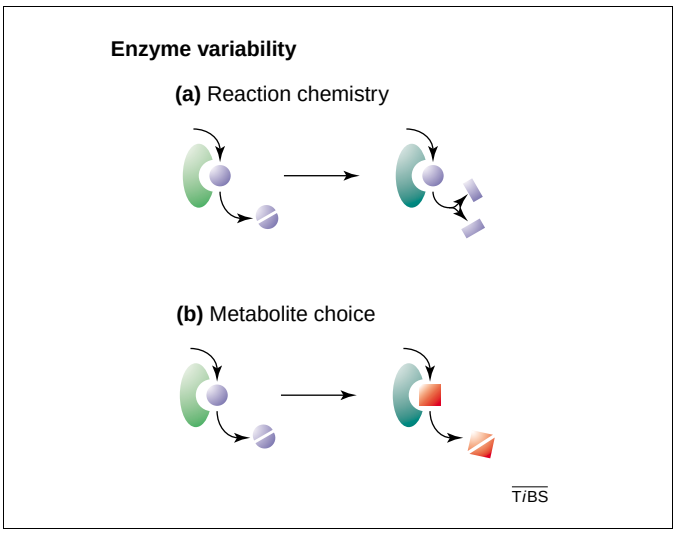

Fig. 2. Models of enzyme variability. The variability of an enzyme superfamily can occur in two different ways: either (a) by changing the reaction chemistry (substrate, blue circle; different blue shapes indicate modified reaction products) or (b) by metabolizing different substrates (substrate, blue circle, is replaced by red square). A pathway is shown with three enzymes (light green). Dark green indicates that the enzyme was modified in activity, red and blue indicate enzymes from different pathways.

Box 1. Enzyme classification

The classification of enzymes can be based on various systems such as the chemical nature of the enzyme, the chemical nature of the substrate, or, as introduced by the International Nomenclature Committee, on the overall reaction chemistry [35]. The committee introduced a unique four-digit code for each enzyme, the so-called Enzyme Commission (EC) number. The first digit of an EC number (A.-.-.-) represents one of the six categories of enzymes: (1) oxidoreductase, (2) transferases, (3) hydrolases, (4) lyases, (5) isomerases, and (6) ligases. Each of these classes are further subdivided; the second digit (A.B.-.-) – the subclass – describes the type of compound or group involved in the reaction. In the case of the oxidoreductases, the subclass indicates the group in the donor that undergoes oxidation (e.g. 1.1.-.- denotes a –CH–OH group). The third digit – the sub-subclass – further describes the type of reaction taking place. For oxidoreductase the third position in the code indicates the type of acceptor involved (e.g. in case of 1.-.1.- the acceptor is either NAD or NADP). The fourth digit is a serial number, which helps to distinguish enzymes of the same sub-subclass, for example, 1.1.1.1 is the alcohol dehydrogenase.

However, there are some caveats that should be considered. First, there are some enzymes that do not fit into this general scheme. For example, the proteases (3.4.-.-) all catalyze the same overall reaction, but are further subdivided based on the difference in the catalytical mechanism. Several authors [7,55] show further cases in which the reaction chemistry of homologous enzymes can be similar even though the EC numbers differ. By contrast, some enzymes with different mechanisms, such as the fructose-bisphosphate aldolase (4.1.2.13), are grouped together for historical reasons. Furthermore, enzymes of different evolutionary origin ('analogous' enzymes [56,57]) might be grouped together if their chemistry is similar. Another important point is that out of the 3541 chemically classified EC numbers (cf. ENZYME [58] database), the protein sequence of the enzyme is known for only 1662 (~47%).

superfamilies catalyze only one reaction (44%). In addition, the overall chemistry of the reactions catalyzed seems to be approximately conserved in a further ~36% of enzyme superfamilies: the EC class, at least, is kept. For example, all enzymes of an enzyme superfamily react as and are classified as hydrolases (always EC class 3). Twenty percent of all superfamilies seem to be able to catalyze reactions of different EC classes, which implies that the enzymes belonging to these superfamilies can catalyze reactions that have totally different chemistry. Interestingly, these superfamilies account for the enzymes that catalyze 45% of all known and classified reactions, whereas the enzymes of conserved superfamilies catalyze only 10% of the known reactions (Fig. 3b).

Alternatively, enzyme evolution can be driven by retention of the catalytic chemistry while different substrates and products are used, such as in proteases and nucleases (Fig. 2b). Evidence for this can be found in many protein families [36–38]. Substrate conservation is low in such families. To assess the metabolite variation, all metabolites (i.e. substrates and products) and cofactors that are involved must be considered. The conservation of metabolite usage in the reactions of a superfamily can be visualized using an association coefficient [39]. As shown in Fig. 4, most enzyme superfamilies that catalyze several enzymatic reactions tend to be conserved in their metabolite choice. However, for many superfamilies (20%; Fig. 3b) a radical change of both reaction type and substrate type is observed. When such a variable superfamily is studied in detail, we find that the metabolites remain well conserved among individual enzymes from a superfamily that catalyzes reactions from the same EC class. But there seems to be almost no conservation among the metabolites if enzymes of such a superfamily catalyze reactions of different EC classes (Fig. 4; see supplementary information, http://www.Bork.EMBL-Heidelberg.DE/~schmidt/metabolites/). Furthermore, recent biochemical studies have provided evidence that several enzymes of the central metabolism can bind additional substrates that are not commonly known (collected in enzyme databases such as BRENDA [8,40]). This 'underground metabolism' [41] might provide further variability for enzymes to explore new functions and to provide a 'toolset' for new pathways.

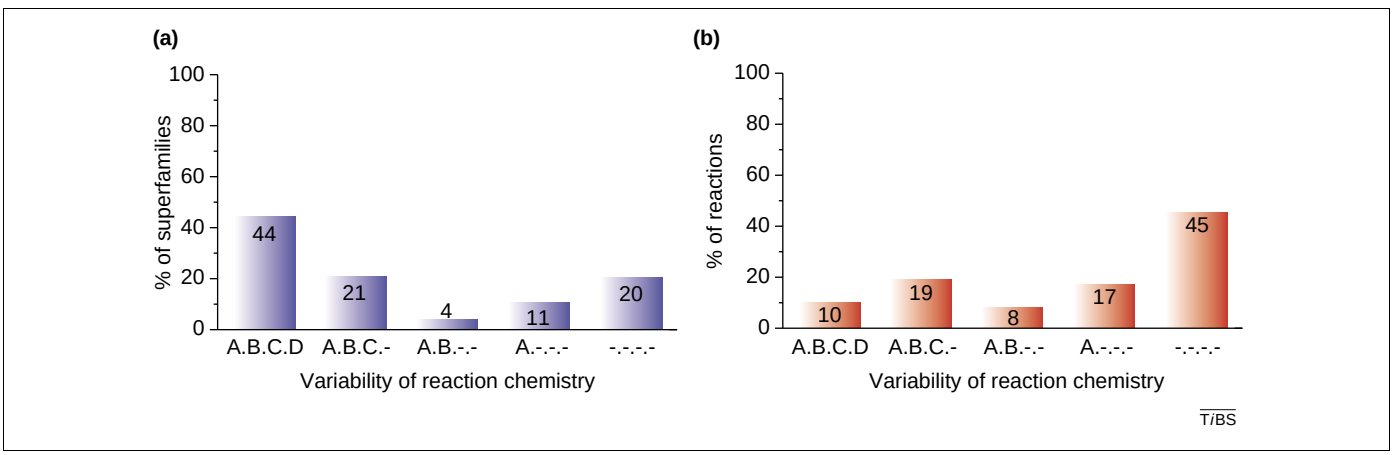

Fig. 3. Variability among enzyme superfamilies due to differences in reaction chemistry. (a) Variability of reaction chemistry (*x*-axis) and respective percentage from the total number of enzyme superfamilies (*y*-axis) according to the Enzyme Classification (EC) scheme [35]. Many (44%) of the enzyme superfamilies catalyze only one reaction (left bar). 36% of the superfamilies (three bars in the middle) are more (to the right) or less (to the left) variable in the chemistry of the catalyzed reaction but at least the EC class is conserved. In 20% of the cases, superfamilies seem to be extremely variable and able to catalyze reactions of different EC classes (right bar). (b) Variability of reaction chemistry (*x*-axis) and respective percentage from the total number of enzyme reactions (*y*-axis). The 20% of highly variable enzyme superfamilies account for 45% of all known reactions (right bar). Our dataset shown here is based on the SCOP [32] database. Similar data, published by Todd *et al.* [14], are based on the CATH [31] protein structure classification database.

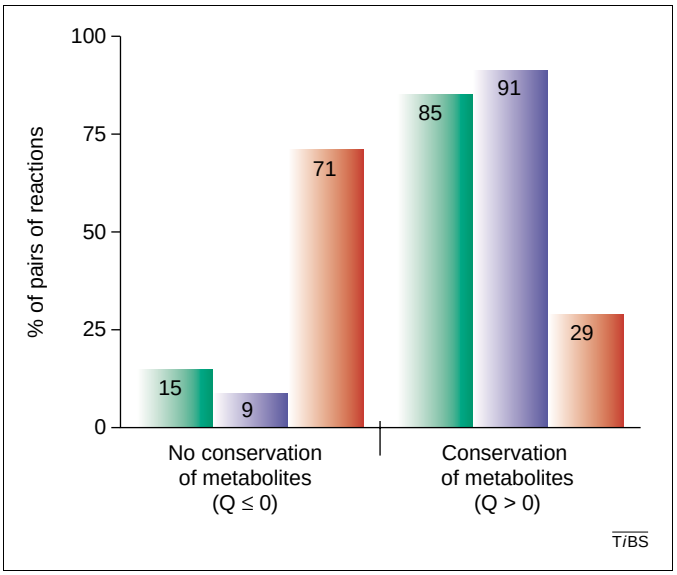

Fig. 4. Metabolite conservation in enzymatic reactions of the same enzyme superfamily. The conservation of metabolite usage in two enzymatic reactions can be quantified and demonstrated by their correlation [39] (coefficient of association Q; details in supplementary information, http://www.Bork.EMBL-Heidelberg.DE/~schmidt/metabolites/). When comparing two enzymatic reactions, a Q-value of −1 means that no metabolite is shared, whereas a value of 1 means that both reactions use the same metabolites. For each pair of reactions catalyzed by a common SCOP superfamily, the Q-values were calculated. Three categories are compared (shown in green, blue and red): the green bars contain the coefficients of SCOP superfamilies with 'conserved' enzymatic reactions (i.e. they catalyze reactions of only one EC class). For superfamilies catalyzing reactions of different EC main classes, the Q-values of reaction pairs within the same EC class (blue) are compared with those not sharing the same EC-class (red). The negative correlation in metabolite choice for enzymes of an enzyme superfamily that can catalyze different EC reactions (red) clearly stands out: these variable enzyme superfamilies easily change their metabolites. They account for >45% of all reactions (as shown in Fig. 2b). The resulting correlation values are collected and grouped into positive or negative (≤0) values. Each category is normalized to give a total value of 100%. In addition, similar metabolites are regarded as the same metabolite according to the KEGG [51] database. Detailed results for not pooling similar metabolites or including co-factors and all other combinations are given in the supplementary material, but show little effect on the overall tendency (as shown in this figure) these results are also not substantially changed using metabolites according to the ENZYME database.

Some superfamilies change more metabolites and reactions than others

Nature does not seem to favor one mechanism over the other; whether an enzyme superfamily turns out to be more variable or conservative depends on the specific enzyme superfamily. One explanation for the wide range of variability among enzyme superfamilies might be their differences in sequence divergence. Comparing conservative and variable enzyme superfamilies (Fig. 4), the average sequence identity within superfamilies of enzymes catalyzing reactions of one EC class is 38%. The sequence identity of highly variable superfamilies of enzymes that catalyze reactions of different EC classes is lower. As shown for the conservation of metabolites between reactions, the average sequence identity in a variable enzyme superfamily comparing enzyme sequences catalyzing reactions of different EC main classes is 13%, which is far less than in protein sequences of enzymes catalyzing reactions of the same main class (31%; see supplementary information for details of calculation).

The inherent variability of the members of some large enzyme superfamilies is high. Todd *et al.* [14] showed that for enzyme superfamilies, a variation in the reaction chemistry (i.e. changes in the first digit of EC numbers) is rare above 40% sequence identity. However, there are several larger enzyme superfamilies that include counter examples. For instance, the DNaseI-like superfamily (SCOP [32] 4.1.115) includes ExoDNase (EC 3.1.11.2) and DNA lyase (EC 4.2.99.18) sharing a sequence identity of 57% [11] but catalyzing reactions from different EC main classes. Thus, on average, even above 50% sequence identity <30% of the compared enzyme pairs in an enzyme superfamily catalyze entirely identical reactions [11].

By contrast, domain recombination does not boost enzyme evolution much. In recent studies, Apic *et al.* [2,3] analyzed protein domains (SCOP superfamilies) in seven genomes from all three kingdoms of life. 624 of the 764 superfamilies are found in these seven genomes and most of them are found in combination with one or two other domains, whereas only a few domains are versatile in their behavior. This indicates that the recombination of different enzyme domains among each other is limited. Furthermore, the biological repertoire of enzyme structures actually used is often small. Examples for strong biases in enzyme structure choice can be found in prokaryotes and yeast [24,42].

Impact of highly represented metabolites

Another factor that is important for the variability in metabolite choice of enzyme superfamilies and hence pathway evolution are highly represented metabolites. In a certain way they provide a hook for new pathways. On a note of caution, most of the current data are based on protein or genome information. Far less comes from direct experimental data on enzyme biochemistry or metabolites. It seems probable that a range of different scenarios might have occurred (and be occurring) during pathway evolution. One possibility is that the early stages of metabolic evolution, when pathways such as glycolysis and other housekeeping pathways were being established, occurred by enzyme-driven evolution, whereas more recent pathways are metabolite driven. Alternatively, constraints by structural and chemical properties of highly represented metabolites might have already biased the evolutionary space explored [43] in the early days of pathway evolution. Certainly, structural constraints in enzyme architecture currently influence pathway evolution as evidenced by strong preferences for certain folds such as the TIM barrel [4].

However, using the data currently available, we can quantify the usage of metabolites in enzymatic reactions. We have observed that, for many metabolites, they are rarely used and occur in individual enzymatic reactions, whereas only a few metabolites are used frequently (Fig. 5). For most of the metabolites the distribution can be fitted according to a power law, which was also noted by Alves *et al.* [1]. This kind of distribution can be shown for various biological systems such as the metabolic network [44] or the distribution of folds in genomes [42].

Interestingly, as demonstrated clearly in Fig. 5, we have found that the most frequently used metabolites occur even more often than expected from the power law. In addition, we observe that these often used metabolites,

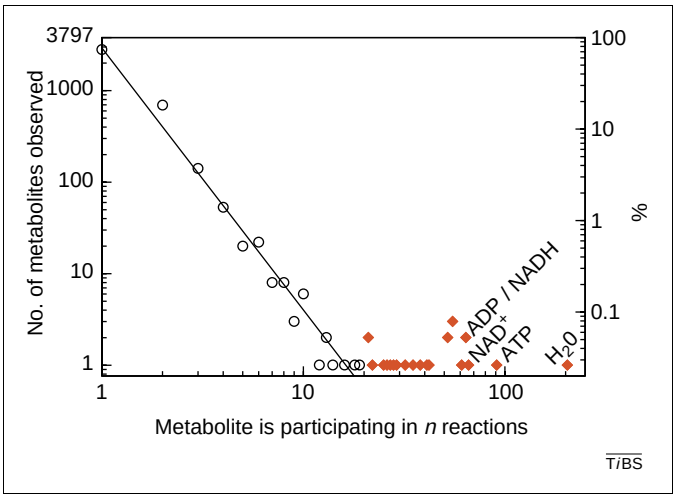

Fig. 5. Metabolite usage in enzymatic reaction. The occurrence of metabolites in enzymatic reactions (circles) follows a power-law distribution: both axes have logarithmic scale. On the *x*-axis the number (*n*) of enzymatic reactions each metabolite is involved in is shown. On the *y*-axis, we plot the number of metabolites or the percentage from the total number of metabolites (right) that are found to participate in a given number of enzymatic reactions. The straight (diagonal) line in this double logarithmic plot approximates the distribution. These data correspond to those recently published by Alves *et al.* [1]. This is called a power-law distribution; it has been shown for many natural and technical systems with growth by agglomeration, such as relations by mutual friends or the internet [52] and organism-wide organization of the metabolic network [44,46], protein folds [42] and protein interactions [53]. However, note that some metabolites do occur more often then expected by power-law (red diamonds). These metabolites, such as water and 13 other hydrophilic and mostly charged compounds, are accessible by many different superfamilies ('hub' metabolites). The ten right-most diamonds correspond to H_2O (203/92), ATP (90/32), NAD^+ (65/18), ADP (63/19), NADH (63/16), O_2 (60/30), CO_2 (54/35), NADP + (54/22), NADPH (54/22) and phosphate (P_i; 51/35); the first number in parentheses indicates the number of reactions the metabolite participates in and the second indicates the number of enzyme superfamilies using the metabolite. Only reactions with an assigned SCOP superfamily were taken into account. The metabolites were extracted from the LIGAND [54] database.

such as H_2O or ATP, are also used by many enzyme superfamilies (Fig. 5). Apparently, these metabolites seem to support enzyme variability; they help to change metabolites in the variable enzyme superfamilies and therefore stimulate pathway evolution. Pathways evolve and concentrate around these central metabolites. As they are connected to so many reactions, these are also called 'hub' metabolites [44]. Moreover, the well-connected central metabolites lead to short pathway distances in the network. This is called 'small-world behavior' because it is similar to global communication networks in which a hierarchy of well-connected central nodes enables messages to reach everybody and makes the world 'small' [45]. As a result, times required for perturbations to spread in the metabolic network are minimized and this might enable metabolism to react rapidly [46]. Work by different groups [46,47] confirms such a channeling effect of these hub metabolites.

There also seems to be a tendency for new pathways and the higher levels of metabolic organization to evolve with preference around such central metabolites. They provide a helping hand for pathway evolution and can be accommodated by many enzyme structures. Around central hub-metabolites new pathways evolve, often by enzyme recruitment from existing pathways. The evolution of the pathway for pentachlorophenol (PCP) degradation is a good example [48]. Now branded as a carcinogen, PCP was artificially introduced (as a 'xenobiotic' substance) in the biosphere in 1936 as a biocide for timber conservation. Since then, organisms such as *Sphingomonas chlorophenolica* rapidly developed a new pathway by recruiting enzymes from the tyrosine catabolism to a degrading pathway of dichloroquinones and alteration of the first enzyme in the pathway to accommodate and become resistant to PCP. Therefore, such man-made xenobiotics provide an ideal test suite for examining the extent to which, for example, plants can adapt to new metabolites [49] and which strategy regarding pathway evolution and enzyme adaptation is observed.

Interestingly, as a result of this accretion process, the same topology and hierarchical organization of modules in metabolic networks is preserved [47]. For example, the pathway organization of archaea and eukaryotes is more similar to each other then to bacteria, whereas individual eukaryotic metabolic enzymes are often more similar to (and have, since early evolutionary times, often been recruited from) bacterial enzymes [50].

Concluding remarks

Together, current analysis provides us with new insights into the evolution of enzymes and their pathways. Besides the observed limited variability of many enzyme superfamilies in their reaction chemistry and metabolite choice, certain superfamilies appear to have a broader substrate specificity and reaction variability. This variability provides a powerful 'toolset' for pathway evolution. Widespread recruitment of enzymes to new pathways becomes possible and is the most often observed mode of evolution for new pathways. Furthermore, we observe a strong bias among enzymes in the usage of metabolites. These frequently used metabolites play an important role because new pathways seem to evolve around those compounds with preference. Thus, these hub metabolites provide a helping hand for enzyme and pathway evolution. It will be interesting to extend our knowledge on biological systems to pathways involving regulation, transcription or differentiation to see how, in these other types of pathways, the balance between the requirements of enzyme conservation and pathway plasticity is achieved.

References

1 Alves, R. *et al.* (2002) Evolution of enzymes in metabolism: a network perspective. *J. Mol. Biol.* 320, 751–770
2 Apic, G. *et al.* (2001) An insight into domain combinations. *Bioinformatics* 17 (Suppl. 1), S83–S89
3 Apic, G. *et al.* (2001) Domain combinations in archaeal, eubacterial and eukaryotic proteomes. *J. Mol. Biol.* 310, 311–325
4 Copley, R.R. and Bork, P. (2000) Homology among (βα)8-barrels: implications for the evolution of metabolic pathways. *J. Mol. Biol.* 303, 627–641
5 Hegyi, H. and Gerstein, M. (1999) The relationship between protein structure and function: a comprehensive survey with application to the yeast genome. *J. Mol. Biol.* 288, 147–164
6 Hrmova, M. *et al.* (2002) Structural basis for broad substrate specificity in higher plant β-D-glucan glucohydrolases. *Plant Cell* 14, 1033–1052
7 Nahum, L.A. and Riley, M. (2001) Divergence of function in sequence-related groups of *Escherichia coli* proteins. *Genome Res.* 11, 1375–1381
8 Pollack, J.D. *et al.* (2002) Suspected utility of enzymes with multiple activities in the small genome Mycoplasma species: the replacement of the missing 'household' nucleoside diphosphate kinase gene and activity by glycolytic kinases. *OMICS* 6, 247–258

9 Rison, S.C.G. *et al.* (2002) Homology, pathway distance and chromosomal localization of the small molecule metabolism enzymes in *Escherichia coli*. *J. Mol. Biol.* 318, 911–932
10 Rison, S.C.G. and Thornton, J.M. (2002) Pathway evolution, structurally speaking. *Curr. Opin. Struct. Biol.* 12, 374–382
11 Rost, B. (2002) Enzyme function less conserved than anticipated. *J. Mol. Biol.* 318, 595–608
12 Teichmann, S.A. *et al.* (2001) Small-molecule metabolism: an enzyme mosaic. *Trends Biotechnol.* 19, 482–486
13 Teichmann, S.A. *et al.* (2001) The evolution and structural anatomy of the small molecule metabolic pathways in *Escherichia coli*. *J. Mol. Biol.* 311, 693–708
14 Todd, A.E. *et al.* (2001) Evolution of function in protein superfamilies, from a structural perspective. *J. Mol. Biol.* 307, 1113–1143
15 Min, B. *et al.* (2002) Transfer RNA-dependent amino acid biosynthesis: an essential route to asparagine formation. *Proc. Natl. Acad. Sci. U. S. A.* 99, 2678–2683
16 Horowitz, N.H. (1945) On the evolution of biochemical syntheses. *Proc. Natl. Acad. Sci. U. S. A.* 31, 153–157
17 Horowitz, N.H. (1965) The evolution of biochemical syntheses – retrospect and prospect. In *Evolving Genes and Proteins* (Bryson, V. and Vogel, J.H., eds) pp. 15–23, Academic Press
18 Fothergill-Gilmore, L.A. and Michels, P.A. (1993) Evolution of glycolysis. *Prog. Biophys. Mol. Biol.* 59, 105–235
19 Petsko, G.A. *et al.* (1993) On the origin of enzymatic species. *Trends Biochem. Sci.* 18, 372–376
20 Roy, S. (1999) Multifunctional enzymes and evolution of biosynthetic pathways: retro-evolution by jumps. *Proteins* 37, 303–309
21 O'Brien, P.J. and Herschlag, D. (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* 6, R91–R105
22 Fischer, R.A. (1958) *The genetical theory of natural selection (Revised and enlarged version of the work, originally published 1930)*, Dover Publ., New York
23 Huynen, M.A. and Snel, B. (2000) Gene and context: integrative approaches to genome analysis. *Adv. Protein Chem.* 54, 345–379
24 Gerlt, J.A. and Babbitt, P.C. (2001) Divergent evolution of enzymatic function: mechanistically diverse superfamilies and functionally distinct suprafamilies. *Annu. Rev. Biochem.* 70, 209–246
25 Jensen, R.A. (1976) Enzyme recruitment in evolution of new function. *Annu. Rev. Microbiol.* 30, 409–425
26 Jensen, R.A. (1996) Evolution of metabolic pathways in enteric bacteria. In *Escherichia coli and Salmonella* (Neidhardt, F.C. *et al.*, eds), pp. 2649–2662, ASM Press
27 Ycas, M. (1974) On earlier states of the biochemical system. *J. Theor. Biol.* 44, 145–160
28 Martin, A.C. *et al.* (1998) Protein folds and functions. *Structure* 6, 875–884
29 Nara, T. *et al.* (2000) Evolutionary implications of the mosaic pyrimidine-biosynthetic pathway in eukaryotes. *Gene* 257, 209–222
30 Jürgens, C. *et al.* (2000) Directed evolution of a $(\beta\alpha)8$-barrel enzyme to catalyze related reactions in two different metabolic pathways. *Proc. Natl. Acad. Sci. U. S. A.* 97, 9925–9930
31 Orengo, C.A. *et al.* (1997) CATH – a hierarchic classification of protein domain structures. *Structure* 5, 1093–1108
32 Murzin, A.G. *et al.* (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247, 536–540
33 Bork, P. *et al.* (1992) An ATPase domain common to prokaryotic cell cycle proteins, sugar kinases, actin, and hsp70 heat shock proteins. *Proc. Natl. Acad. Sci. U. S. A.* 89, 7290–7294
34 Tipton, K. and Boyce, S. (2000) History of the enzyme nomenclature system. *Bioinformatics* 16, 34–40
35 Tipton, K.F. (1994) Nomenclature committee of the international union of biochemistry and molecular biology (NC-IUBMB). enzyme nomenclature. Recommendations 1992. Supplement: corrections and additions. *Eur. J. Biochem.* 223, 1–5
36 Babbitt, P.C. and Gerlt, J.A. (1997) Understanding enzyme superfamilies. *Chemistry as the fundamental determinant in the evolution of new catalytic activities. J. Biol. Chem.* 272, 30591–30594
37 Eklund, H. and Fontecave, M. (1999) Glycyl radical enzymes: a conservative structural basis for radicals. *Structure Fold. Des.* 7, R257–R262
38 Lawrence, J.G. and Ochman, H. (1997) Amelioration of bacterial genomes: rates of change and exchange. *J. Mol. Evol.* 44, 383–397
39 Kendal, M. and Stuart, A. (1979) Categorized data. *The Advanced Theory of Statistics*, pp. 566–569, Charles Griffin & Company, London & High Wycombe
40 Schomburg, I. *et al.* (2002) BRENDA, enzyme data and metabolic information. *Nucleic Acids Res.* 30, 47–49
41 D'Ari, R. and Casadesús, J. (1998) Underground metabolism. *Bioessays* 20, 181–186
42 Qian, J. *et al.* (2001) Protein family and fold occurrence in genomes: power-law behaviour and evolutionary model. *J. Mol. Biol.* 313, 673–681
43 Gould, S.J. (1997) The exaptive excellence of spandrels as a term and prototype. *Proc. Natl. Acad. Sci. U. S. A.* 94, 10750–10755
44 Jeong, H. *et al.* (2000) The large-scale organization of metabolic networks. *Nature* 407, 651–654
45 Strogatz, S.H. (2001) Exploring complex networks. *Nature* 410, 268–276
46 Wagner, A. and Fell, D.A. (2001) The small world inside large metabolic networks. *Proc. R. Soc. Lond. B Biol. Sci.* 268, 1803–1810
47 Ravasz, E. *et al.* (2002) Hierarchical organization of modularity in metabolic networks. *Science* 297, 1551–1555
48 Copley, S.D. (2000) Evolution of a metabolic pathway for degradation of a toxic xenobiotic: the patchwork approach. *Trends Biochem. Sci.* 25, 261–265
49 Sandermann, H.J. (1994) Higher plant metabolism of xenobiotics: the 'green liver' concept. *Pharmacogenetics* 4, 225–241
50 Podani, J. *et al.* (2001) Comparable system-level organization of Archaea and Eukaryotes. *Nat. Genet.* 29, 54–56
51 Kanehisa, M. *et al.* (2002) The KEGG databases at GenomeNet. *Nucleic Acids Res.* 30, 42–46
52 Broder, A. *et al.* (2000) Graph structure in the web. *Comput. Netw.* 33, 309–320
53 Snel, B. *et al.* (2002) The identification of functional modules from the genomic association of genes. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5890–5895
54 Goto, S. *et al.* (2002) LIGAND: database of chemical compounds and reactions in biological pathways. *Nucleic Acids Res.* 30, 402–404
55 Gerlt, J.A. and Babbitt, P.C. (1998) Mechanistically diverse enzyme superfamilies: the importance of chemistry in the evolution of catalysis. *Curr. Opin. Chem. Biol.* 2, 607–612
56 Galperin, M.Y. and Koonin, E.V. (1999) Functional genomics and enzyme evolution. Homologous and analogous enzymes encoded in microbial genomes. *Genetica* 106, 159–170
57 Galperin, M.Y. *et al.* (1998) Analogous enzymes: independent inventions in enzyme evolution. *Genome Res.* 8, 779–790
58 Bairoch, A. (2000) The ENZYME database in 2000. *Nucleic Acids Res.* 28, 304–305