The Classification and Evolution of Enzyme Function

Sergio Martínez Cuesta[1], Syed Asad Rahman[1], Nicholas Furnham[2] and Janet M. Thornton[1]

[1]European Molecular Biology Laboratory, European Bioinformatics Institute EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge, CB10 1SD, United Kingdom.

[2]Department of Pathogen Molecular Biology, London School of Hygiene & Tropical Medicine, London, WC1E 7HT, United Kingdom.

Abstract

Enzymes are the proteins responsible for the catalysis of life. Enzymes sharing a common ancestor as defined by sequence and structure similarity are grouped into families and superfamilies. The molecular function of enzymes is defined as their ability to catalyse biochemical reactions; it is manually classified by the Enzyme Commission (EC) and robust approaches to quantitatively compare catalytic reactions are just beginning to appear. Here we present an overview of studies at the interface of the evolution and function of enzymes.

Introduction

The notion of enzymes as biocatalysts was originally presented in 1833 with the discovery of the conversion of starch into sugars catalysed by diastase (1). However it was not until the 20th century that scientists realised their full potential in the context of medicine and technology. Major landmarks were the development of methods for enzyme isolation and purification, the realisation that enzymes are proteins with biochemical activity and their characterisation using X-ray diffraction techniques (2) (3). Studies on the dynamic nature of the structure of ribonuclease and efforts to decipher the catalytic mechanism of lysozyme revealed enzymology as an emerging scientific discipline.

Enzymes have many functional attributes. At the molecular level, enzymes catalyse biochemical reactions by accelerating the conversion of substrates into products in a buried pocket within the active site of the enzyme. Without enzyme catalysis, most reactions would be too slow to be useful for life, although not all reactions in nature require catalysis (4). From the pioneering studies by Krebs on the citric acid cycle (5) to the elaboration of comprehensive biochemical wall charts and databases, we have realised that enzymes do not act independently but modulate collectively metabolic pathways and networks. Enzymes perform their molecular function in a particular cell compartment. For instance, hexokinase turns D-glucose into α-D-glucose-6-phosphate in the glycolysis pathway, which takes place in the cytosol. Finally, there is great diversity in the fraction of enzymes in different organisms (6) and variations at the organelle, cell type and tissue levels have also been observed (7) (8) (9).

Classifying the function of enzymes

In a similar way to our present-day data deluge in genomics, the good old days of enzymology and biochemistry witnessed the growing accumulation of vast amounts of enzyme data: biochemical reactions, enzyme kinetics, crystallographic structures and mechanistic interpretations. However the means of data storage and dissemination were different at the time, databases did not exist as such and functional data was scattered through the literature making any form of overview analysis challenging. The nomenclature of enzymes was also problematic, enzymes were given trivial names in order to identify them. Some names were carefully chosen by groups of biochemists, however sometimes names were given to the same enzyme by different scientific schools, likewise different enzymes were named the same way. This led to confusing and ambiguous communication between researchers (10). For example, NADPH dehydrogenase was first known as “NADPH diaphorase” and “old yellow enzyme” due to its ability to reduce various dyes, both trivial names still persist today (11) (12). Soon after, D-amino acid oxidase was designated as “new yellow enzyme” and

distinction between both enzymes became even more difficult. The remarkable increase in the number of newly discovered enzymes called for the development of a system to name and classify them in a consistent manner.

Just as taxonomic classification proved so useful to identify and dissect the diversity of living organisms during the 18th century, in 1956 biochemists and enzymologists launched an initiative to gather all available information about the overall catalysed reactions in order to name and classify enzymes. This was led by experts from the Enzyme Commission (EC) of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB) who presented a framework to name old enzymes according to which new enzymes could be classified. Enzymes are now named and identified systematically with an EC number; this code is a four-level description that is used to classify enzymes depending on the overall chemical transformation of substrates into products (13). The first level corresponds to six different classes according to the type of chemistry being carried out. Oxidoreductases catalyse oxidation/reduction reactions (EC 1), transferases transfer a chemical group (EC 2), for example, a methyl or glycosyl moiety; hydrolases perform hydrolysis of chemical bonds (EC 3), lyases also cleave chemical bonds by other means than by oxidation or hydrolysis (EC 4), isomerases catalyse geometric and structural changes between isomers (EC 5) and lastly, ligases join two compounds with associated hydrolysis of a nucleoside triphosphate molecule (EC 6). These EC classes are further divided in subclasses and sub-subclasses (second and third level, respectively) in line with a variety of criteria such as the chemical bond cleaved or formed, the reaction centre, the transferred chemical group and the cofactor used for catalysis. The final level of classification defines substrate specificity. For example, alanine racemase is an isomerase (EC 5), in particular a racemase (EC 5.1) acting on the amino acid (EC 5.1.1) alanine (EC 5.1.1.1).

When the EC started to operate, the prevalent view was that enzymes were substrate specific, however as several enzymes were discovered to catalyse more than one reaction (enzyme promiscuity), EC numbers started to list additional reactions catalysed by the same enzyme. Many studies have characterised promiscuity in detail (14) (15) (16) (17) and revealed a possible role of promiscuous enzymes as intermediates in the evolution of enzyme function (18) (19) (20). However strictly speaking, the EC classification is still made on the basis of the “main” catalysed reaction therefore rendering a limited categorization to describe the full potential activity of enzymes. Nonetheless the assignment of EC numbers to enzymes is now a common routine in the functional annotation of proteins and protein-coding genes in databases such as UniprotKB (21) and Ensembl (22) and has been adopted by the widely used Gene Ontology (GO) (23).

The evolution of enzyme function

The ability of organisms to adapt to the changing conditions of their habitat is crucial to guarantee their survival and reproduction. Adaptation to chemical variations in the living environment drives the innovation, exchange and demise of enzyme function. At the metabolic level, this process of adaptation is related to the ability of enzymes to evolve beneficial functions in an environment of changing chemical conditions (24). For example, the capability of bacteria to acquire resistance to drugs and pesticides.

A) The emergence of enzymes

The discovery of non-enzymatic metal-catalysed metabolic reactions in laboratory experiments reproducing the chemistry of an early ocean suggests a potential abiotic origin of catalytic function (25). As the temperature of the Earth cooled down and stabilised, enzymes might have evolved from non-enzymatic precursors or proto-enzymes, which used to operate at higher temperatures (26). This evolution allowed better control of substrate specificity by preventing the synthesis of undesirable products and an enhanced regulation of metabolism (4). Although enzyme biosynthesis is energy demanding compared to its non-enzymatic counterpart, it led to the regulation of core metabolic pathways shared across most living organisms, e.g. glycolysis. An example of the synergy between the abiotic and living worlds are metalloenzymes. Nature has evolved the structure of these enzymes to wrap around metal catalysts endowing the complexes with the ability to harness and control diverse chemistry in biology.

B) Enzyme functionalisation

Multiple research studies have unveiled genetic mechanisms whereby innovation in enzyme function might have emerged. In its simplest form, the genetic diversity driving adaptation relies on the accumulation of point mutations. Although the majority of them are neutral or deleterious, gain-of-function mutations may create a new activity in an existing enzyme. Subsequently, beneficial mutations might either increase the level of the activity or when occurring in regulatory regions, they might enhance gene expression and therefore boost the cellular concentration of the enzyme up to physiological levels (27). The evolution of new activity is modulated by stability-function trade-offs (20) (28), and if it provides a selective advantage to the organism in its ability to perform a fundamental biological process such as the competition for resources, will become fixed within a population with beneficial mutations that enhance this activity gradually strengthening the fitness of the organism. Then the function might either remain as a new activity of a multifunctional enzyme or segregate as the primary activity of a new enzyme, through evolutionary processes such as gene

duplication and specialisation (29). The neo- and sub-functionalisation of enzymes is explained by a generally accepted model known as Innovation-Amplification-Divergence (IAD) (24), which is considered to be iterative so the newly evolved enzyme may develop new functions leading to further adaptive cycles.

C) Using sequence and structure to generate families of enzymes

The analysis of the sequences and structures of evolutionary related enzymes reveals relationships between catalytic activities. Approaches to capture fine details about the conservation of sequence and structural elements in protein domains use Hidden Markov Models (HMMs) (30), which facilitate the classification of proteins in families and superfamilies (31) (32). Although the three-dimensional location of active sites is frequently conserved within superfamilies (33), variations of the physicochemical properties of the residues lining the pockets and other patterns of structural change have been observed (34) (35). Enzymes accommodate alternative chemistries using a combination of chemistry-driven and substrate-driven evolution (36) (37). The overall chemical reaction is often changed whilst conserving at least one mechanistic step (38) (39), however binding similar substrates while conserving the reaction chemistry is also observed (40) (41). Analogous reactions are sometimes catalysed by different superfamilies (42) using similar active sites (43) or catalytic mechanisms (44). This illustrates how nature might evolve the same functional outcome independently using different structural solutions – convergent evolution.

In order to investigate the evolutionary routes whereby divergence in sequence and structure lead to new enzyme functions in superfamilies, researchers use phylogenetic analysis. For example, Furnham and colleagues developed FunTree (45), a resource containing phylogenetic trees decorated with structural, functional and mechanistic information that captures the evolution of enzyme function in superfamilies (Fig. 1A).

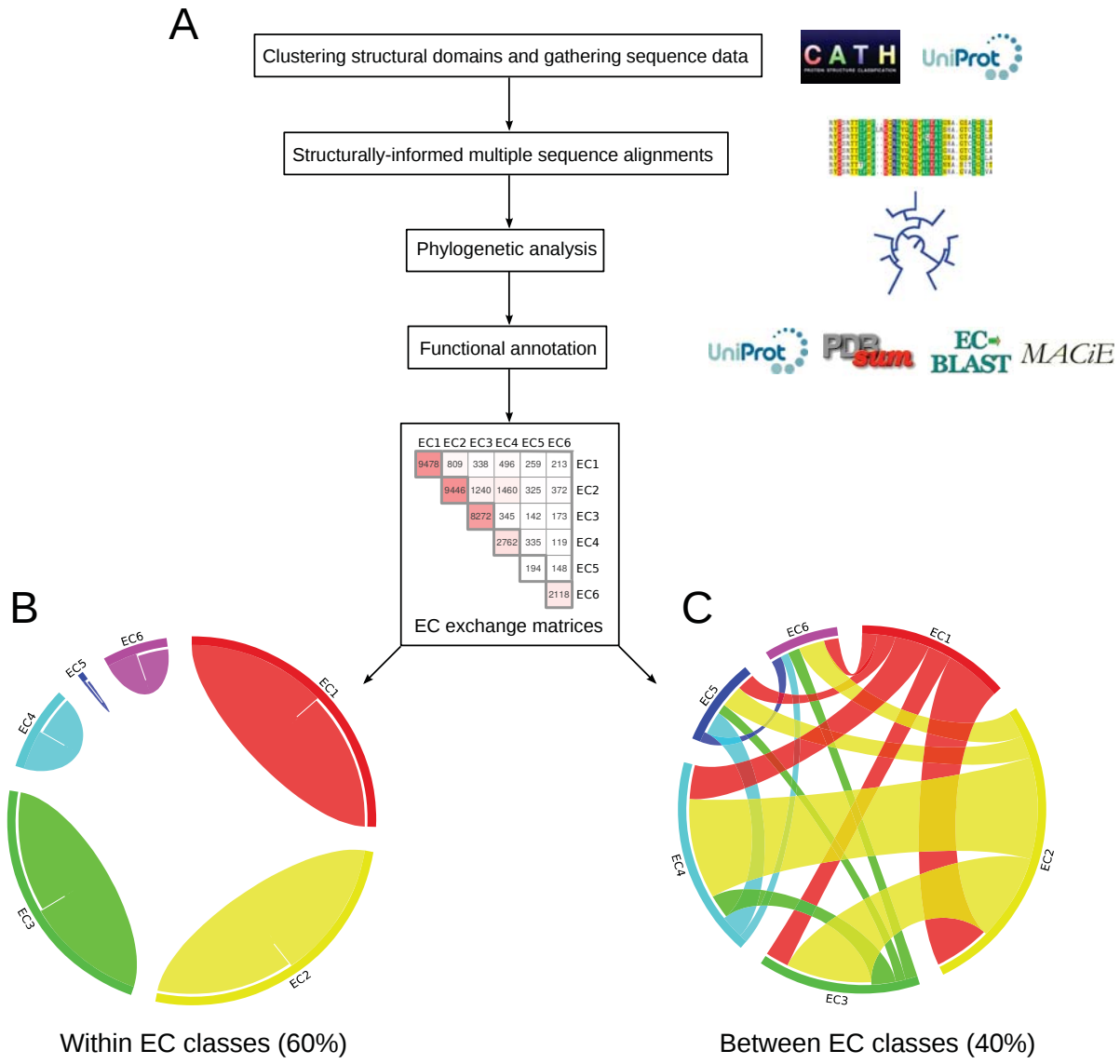

Figure 1. Exploring the evolution of enzyme function within 283 multifunctional CATH superfamilies. (A) FunTree approach (45): first, structural clusters of CATH domains involved in enzyme function are created, populated with sequence relatives and structurally-informed multiple sequence alignments are generated. Second, using alignments as starting point, species-guided phylogenetic trees are created. Last, functional annotations are retrieved from protein data resources and the frequency of all possible exchanges between different EC numbers within each superfamily is added to an EC exchange matrix. In order to visualise this matrix, circular diagrams are shown with ribbons representing the frequency of EC changes observed during evolution (band width) (B) within EC classes (diagonal of EC exchange matrix) and (C) between EC classes (off-diagonal). Although the ribbons were coloured according to the lowest EC primary class, they are bidirectional hence the frequency of changes ECX->ECY is the same to ECY->ECX. Data was obtained from CATH version 3.5 (32) and graphics were generated using Circos (56).

D) Cataloguing enzyme function evolution

During evolution, most enzymes evolve to become enzymes from the same EC class (60% of all EC

changes) (Fig. 1B) (e.g. one hydrolase will evolve a new hydrolase function). However, the remaining 40% of changes are between enzymes catalysing different overall chemistry (Fig. 1C). Remarkably, all possible changes between EC classes are observed. There are some preferences such as transferases (EC 2) becoming oxidoreductases (EC 1), hydrolases (EC 3) and lyases (EC 4). Isomerases (EC 5) are exceptional and evolve new overall chemistry more often than conserving the chemistry of isomerisation (46).

E) Enzyme reaction classification

The functional changes we observed within the same EC class would be expected with the existing classification of enzyme function, e.g. an enzyme could easily change to another catalysing similar chemistry thus changing the substrate specificity only. However, exchanges between different EC classes suggest that the chemistry of enzymes is more complex than previously classified, with close relationships between enzymes with radically different EC numbers. The chemistry of related enzyme functions can now be explored using robust computational approaches like EC-BLAST (47). This tool searches and compares reactions on the basis of bond changes, reaction centres and structures of substrates and products. Just like classical studies investigated the evolution of related protein sequences and structures (48), an accurate comparison of enzyme functions might facilitate the interpretation of the evolution of enzymes in the context of their chemistry, which is relevant for enzyme design and genomics (49).

F) Role of domain composition and allostery in enzyme function

Many enzymes are multidomain and acquire new functionality by changing the domain composition during evolution (50). Enzyme function is assigned on a whole-sequence basis without associating specific functions to the composite domains (51) (52). Therefore cataloguing the functional evolution of each individual domain is a complex process, which can lead to multiple different evolutionary routes. Laboratory experiments do also reveal how amino acid residues located far from the active site often make important contributions to binding and catalysis by inducing conformational changes (53) and acting allosterically (54). More experimental research (55) exploring how enzyme function changes with protein structure and domain architecture is necessary.

Conclusions

Enzymes have both biological and chemical attributes. Their sequences and structures delineate their role in the genome and proteome of all living organisms and their ability to catalyse chemical reactions extends their biological function to metabolic pathways and networks. Many enzymes are

promiscuous and perform multiple reactions and as protein sequence evolves, enzymes can change their “reaction profile”. Combining bond changes and reaction centres with structural information about the substrates, products and mechanisms is needed to capture the essence of enzyme chemistry in a functional classification. The development of tools to navigate through reaction space (e.g. EC-BLAST) paves a new way for an improved description of enzyme reactions, providing a deeper perspective of biological function.

Acknowledgements

The authors thank EMBL for funding.

References

1. Payen, A., and J. Persoz. 1833. Mémoire sur la Diastase, les principaux Produits de ses Réactions, et leurs applications aux arts industriels. Ann. chim. phys. 53: 73–92.

2. Sumner, J. 1926. The isolation and crystallization of the enzyme urease. J. Biol. Chem. 69: 435–441.

3. Blake, C.C., D.F. Koenig, G.A. Mair, A.C. North, D.C. Phillips, and V.R. Sarma. 1965. Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Angstrom resolution. Nature. 206: 757–61.

4. Keller, M.A., G. Piedrafita, and M. Ralser. 2015. The widespread role of non-enzymatic reactions in cellular metabolism. Curr. Opin. Biotechnol. 34C: 153–161.

5. Krebs, H.A. 1940. The citric acid cycle and the Szent-Györgyi cycle in pigeon breast muscle. Biochem. J. 34: 775–9.

6. Freilich, S., R. V. Spriggs, R.A. George, B. Al-Lazikani, M. Swindells, and J.M. Thornton. 2005. The complement of enzymatic sets in different species. J. Mol. Biol. 349: 745–63.

7. Lewis, N.E., G. Schramm, A. Bordbar, J. Schellenberger, M.P. Andersen, J.K. Cheng, N. Patel, A. Yee, R.A. Lewis, R. Eils, R. König, and B.Ø. Palsson. 2010. Large-scale in silico modeling of metabolic interactions between cell types in the human brain. Nat. Biotechnol. 28: 1279–85.

8. Mintz-Oron, S., S. Meir, S. Malitsky, E. Ruppin, A. Aharoni, and T. Shlomi. 2012. Reconstruction of Arabidopsis metabolic network models accounting for subcellular compartmentalization and tissue-specificity. Proc. Natl. Acad. Sci. U. S. A. 109: 339–44.

9. Stincone, A., A. Prigione, T. Cramer, M.M.C. Wamelink, K. Campbell, E. Cheung, V. Olin-Sandoval, N.-M. Grüning, A. Krüger, M. Tauqeer Alam, M.A. Keller, M. Breitenbach, K.M. Brindle, J.D. Rabinowitz, and M. Ralser. 2014. The return of metabolism: biochemistry and physiology of the pentose phosphate pathway. Biol. Rev. Camb. Philos. Soc. : 000–000.

10. Cornish-Bowden, A. 2014. Current IUBMB recommendations on enzyme nomenclature and kinetics. Perspect. Sci. 1: 74–87.

11. Daugherty, A.B., S. Govindarajan, and S. Lutz. 2013. Improved biocatalysts from a synthetic circular permutation library of the flavin-dependent oxidoreductase old yellow enzyme. J. Am. Chem. Soc. 135: 14425–32.

12. Savignon, T., E. Costa, F. Tenorio, A.C. Manhães, and P.C. Barradas. 2012. Prenatal hypoxic-ischemic insult changes the distribution and number of NADPH-diaphorase cells in the cerebellum. PLoS One. 7: e35786.

13. Tipton, K., and S. Boyce. 2000. History of the enzyme nomenclature system. Bioinformatics. 16: 34–40.

14. Duarte, F., B.A. Amrein, and S.C.L. Kamerlin. 2013. Modeling catalytic promiscuity in the alkaline phosphatase superfamily. Phys. Chem. Chem. Phys. 15: 11160–77.

15. Gatti-Lafranconi, P., and F. Hollfelder. 2013. Flexibility and reactivity in promiscuous enzymes. Chembiochem. 14: 285–92.

16. Pandya, C., J.D. Farelli, D. Dunaway-Mariano, and K.N. Allen. 2014. Enzyme promiscuity: engine of evolutionary innovation. J. Biol. Chem. 289: 30229–36.

17. Khanal, A., S. Yu McLoughlin, J.P. Kershner, and S.D. Copley. 2014. Differential Effects of a Mutation on the Normal and Promiscuous Activities of Orthologs: Implications for Natural and Directed Evolution. Mol. Biol. Evol. 32: 100–108.

18. Aharoni, A., L. Gaidukov, O. Khersonsky, S. McQ Gould, C. Roodveldt, and D.S. Tawfik. 2005. The evolvability of promiscuous protein functions. Nat. Genet. 37: 73–6.

19. Dellus-Gur, E., A. Toth-Petroczy, M. Elias, and D.S. Tawfik. 2013. What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs. J. Mol. Biol. 425: 2609–21.

20. Kaltenbach, M., and N. Tokuriki. 2014. Dynamics and constraints of enzyme evolution. J. Exp. Zool. B. Mol. Dev. Evol. 322: 468–87.

21. The Uniprot Consortium. 2013. Update on activities at the Universal Protein Resource (UniProt) in 2013. Nucleic Acids Res. 41: D43–7.

22. Kersey, P.J., J.E. Allen, M. Christensen, P. Davis, L.J. Falin, C. Grabmueller, D.S.T. Hughes, J. Humphrey, A. Kerhornou, J. Khobova, N. Langridge, M.D. McDowall, U. Maheswari, G. Maslen, M. Nuhn, C.K. Ong, M. Paulini, H. Pedro, I. Toneva, M.A. Tuli, B. Walts, G. Williams, D. Wilson, K. Youens-Clark, M.K. Monaco, J. Stein, X. Wei, D. Ware, D.M. Bolser, K.L. Howe, E. Kulesha, D. Lawson, and D.M. Staines. 2014. Ensembl Genomes 2013: scaling up access to genome-wide data. Nucleic Acids Res. 42: D546–52.

23. Ashburner, M., C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, and G. Sherlock. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat. Genet. 25: 25–9.

24. Copley, S.D. 2012. Toward a systems biology perspective on enzyme evolution. J. Biol. Chem. 287: 3–10.

25. Keller, M.A., A. V. Turchyn, and M. Ralser. 2014. Non-enzymatic glycolysis and pentose phosphate pathway-like reactions in a plausible Archean ocean. Mol. Syst. Biol. 10: 725.

26. Wolfenden, R. 2014. Massive thermal acceleration of the emergence of primordial chemistry, the incidence of spontaneous mutation, and the evolution of enzymes. J. Biol. Chem. 289: 30198–204.

27. Schulenburg, C., and B.G. Miller. 2014. Enzyme recruitment and its role in metabolic expansion. Biochemistry. 53: 836–45.

28. Studer, R.A., P.-A. Christin, M.A. Williams, and C.A. Orengo. 2014. Stability-activity tradeoffs constrain the adaptive evolution of RubisCO. Proc. Natl. Acad. Sci. U. S. A. 111: 2223–8.

29. Glasner, M.E., J.A. Gerlt, and P.C. Babbitt. 2006. Evolution of enzyme superfamilies. Curr. Opin. Chem. Biol. 10: 492–7.

30. Eddy, S.R. 2011. Accelerated Profile HMM Searches. PLoS Comput. Biol. 7: e1002195.

31. Punta, M., P.C. Coggill, R.Y. Eberhardt, J. Mistry, J. Tate, C. Boursnell, N. Pang, K. Forslund, G. Ceric, J. Clements, A. Heger, L. Holm, E.L.L. Sonnhammer, S.R. Eddy, A. Bateman, and R.D. Finn. 2012. The Pfam protein families database. Nucleic Acids Res. 40: D290–301.

32. Sillitoe, I., A.L. Cuff, B.H. Dessailly, N.L. Dawson, N. Furnham, D. Lee, J.G. Lees, T.E. Lewis, R.A. Studer, R. Rentzsch, C. Yeats, J.M. Thornton, and C.A. Orengo. 2013. New functional families (FunFams) in CATH to improve the mapping of conserved functional sites to 3D structures. Nucleic Acids Res. 41: D490–8.

33. Dessailly, B.H., N.L. Dawson, K. Mizuguchi, and C.A. Orengo. 2013. Functional site plasticity in domain superfamilies. Biochim. Biophys. Acta. 1834: 874–89.

34. Dawson, N.L., R. Studer, N. Furnham, D. Lees, S. Das, J. Thornton, and C. Orengo. 2015. What Can Comparative Genomics Reveal About the Mechanisms of Protein Function Evolution? In: New Chemistry and New Opportunities from the Expanding Protein Universe. World Scientific. pp. 16–19.

35. Brown, S.D., and P.C. Babbitt. 2014. New insights about enzyme evolution from large scale studies of sequence and structure relationships. J. Biol. Chem. 289: 30221–8.

36. Babbitt, P.C. 2003. Definitions of enzyme function for the structural genomics era. Curr. Opin. Chem. Biol. 7: 230–7.

37. Chiang, R.A., A. Sali, and P.C. Babbitt. 2008. Evolutionarily conserved substrate substructures for automated annotation of enzyme superfamilies. PLoS Comput. Biol. 4: e1000142.

38. Bartlett, G.J., N. Borkakoti, and J.M. Thornton. 2003. Catalysing new reactions during evolution: economy of residues and mechanism. J. Mol. Biol. 331: 829–60.

39. Gerlt, J.A., P.C. Babbitt, M.P. Jacobson, and S.C. Almo. 2012. Divergent evolution in enolase superfamily: strategies for assigning functions. J. Biol. Chem. 287: 29–34.

40. Furnham, N., I. Sillitoe, G.L. Holliday, A.L. Cuff, R.A. Laskowski, C.A. Orengo, and J.M. Thornton. 2012. Exploring the evolution of novel enzyme functions within structurally defined protein superfamilies. PLoS Comput. Biol. 8: e1002403.

41. Bastard, K., A.A.T. Smith, C. Vergne-Vaxelaire, A. Perret, A. Zaparucha, R. De Melo-Minardi, A. Mariage, M. Boutard, A. Debard, C. Lechaplais, C. Pelle, V. Pellouin, N. Perchat, J.-L. Petit, A. Kreimeyer, C. Medigue, J. Weissenbach, F. Artiguenave, V. De Berardinis, D. Vallenet, and M. Salanoubat. 2014. Revealing the hidden functional diversity of an enzyme family. Nat. Chem. Biol. 10: 42–9.

42. Omelchenko, M. V., M.Y. Galperin, Y.I. Wolf, and E. V. Koonin. 2010. Non-homologous isofunctional enzymes: a systematic analysis of alternative solutions in enzyme evolution. Biol. Direct. 5: 31.

43. Gherardini, P.F., M.N. Wass, M. Helmer-Citterich, and M.J.E. Sternberg. 2007. Convergent

evolution of enzyme active sites is not a rare phenomenon. J. Mol. Biol. 372: 817–45.

44. Almonacid, D.E., E.R. Yera, J.B.O. Mitchell, and P.C. Babbitt. 2010. Quantitative comparison of catalytic mechanisms and overall reactions in convergently evolved enzymes: implications for classification of enzyme function. PLoS Comput. Biol. 6: e1000700.

45. Furnham, N., I. Sillitoe, G.L. Holliday, A.L. Cuff, S.A. Rahman, R.A. Laskowski, C.A. Orengo, and J.M. Thornton. 2012. FunTree: a resource for exploring the functional evolution of structurally defined enzyme superfamilies. Nucleic Acids Res. 40: D776–82.

46. Martínez Cuesta, S., N. Furnham, S.A. Rahman, I. Sillitoe, and J.M. Thornton. 2014. The evolution of enzyme function in the isomerases. Curr. Opin. Struct. Biol. 26C: 121–130.

47. Rahman, S.A., S. Martínez Cuesta, N. Furnham, G.L. Holliday, and J.M. Thornton. 2014. EC-BLAST: a tool to automatically search and compare enzyme reactions. Nat. Methods. 11: 171–174.

48. Chothia, C., and A.M. Lesk. 1986. The relation between the divergence of sequence and structure in proteins. EMBO J. 5: 823–6.

49. Gabaldón, T., and E. V. Koonin. 2013. Functional and evolutionary implications of gene orthology. Nat. Rev. Genet. 14: 360–6.

50. Bashton, M., and C. Chothia. 2007. The generation of new protein functions by the combination of domains. Structure. 15: 85–99.

51. Lopez, D., and F. Pazos. 2009. Gene ontology functional annotations at the structural domain level. Proteins. 76: 598–607.

52. Fang, H., and J. Gough. 2013. DcGO: database of domain-centric ontologies on functions, phenotypes, diseases and more. Nucleic Acids Res. 41: D536–44.

53. Hilvert, D., J.-M. Frère, and W. Versèes. 2014. Discussions of Session 2. In: New Chemistry and New Opportunities from the Expanding Protein Universe. World Scientific. pp. 94–116.

54. Makhlynets, O. V., E.A. Raymond, and I. V. Korendovych. 2015. Design of allosterically regulated protein catalysts. Biochemistry. 54: 1444–56.

55. Baier, F., and N. Tokuriki. 2014. Connectivity between catalytic landscapes of the metallo-β-lactamase superfamily. J. Mol. Biol. 426: 2442–56.

56. Krzywinski, M., J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, and M.A. Marra. 2009. Circos: an information aesthetic for comparative genomics. Genome Res. 19: 1639–45.