Enzyme Function and its Evolution

John BO Mitchell

EaStCHEM School of Chemistry and Biomedical Sciences Research Complex, University of St Andrews, North Haugh, St Andrews, Scotland, KY16 9ST, United Kingdom.

Abstract

With rapid increases over recent years in the determination of protein sequence and structure, alongside knowledge of thousands of enzyme functions and hundreds of chemical mechanisms, it is now possible to combine breadth and depth in our understanding of enzyme evolution. Phylogenetics continues to move forward, though determining correct evolutionary family trees is not trivial. Protein function prediction has spawned a variety of promising methods that offer the prospect of identifying enzymes across the whole range of chemical functions and over numerous species. This knowledge is essential to understand antibiotic resistance, as well as in protein re-engineering and de novo enzyme design.

Introduction

Our picture of the natural history of proteins is based on reconstructing the evolutionary past of the protein domain folds as catalogued in databases such as CATH [1, 2], SCOP [3], its successor SCOPe [4], CDD [5], ECOD [6] – which is specifically designed primarily to reflect evolutionary relationships, and Pfam [7]. These databases provide the key to understanding the evolutionary past of the various cellular and molecular functions, especially enzymatic ones, associated with the catalogued protein folds. Given that protein domains widely found in the proteomes of diverse present-day organisms are more likely to be ancient than those present only in niches, it is possible to make inferences about the approximate ages of protein folds [8, 9, 10]. Cross-referencing with data from other fields of science such as geology can provide estimates of absolute fold ages [11]. One can similarly make suggestions about the folds, and indeed functions, which may have been present in the last universal common ancestor (LUCA) of extant life [8, 12], which constitutes an event horizon for bioinformatics.

This brief review will consider not only this broad sweep of the evolutionary history of enzymes, but also discuss studies capturing specific changes in function. We will look at a combination of experimental and computational approaches to unravel the mysteries of how enzymes manage to evolve novel functions, and consider recent progress in protein function prediction. Finally, we will discuss some priorities for future research.

Enzyme Evolution

Voordeckers et al. [13] carried out a beautifully-designed joint experimental and computational study in which they caught a family of fungal sugar-metabolising enzymes in the act of evolving. In addition to assaying the extant maltases (EC 3.2.1.20) and isomaltases (EC 3.2.1.10) for activity against a range of sugars, they also reconstructed the putative sequences of their common ancestors. They found that the reconstructed ancestral enzymes had broader but weaker activity, turning over a wider range of substrates. Gene duplications gave rise to paralogs which were able to

specialise on a narrower range of substrates and increase their catalytic power on these, while relinquishing the ability to turn over alternative substrates. Interestingly, at least one modern enzyme retains the ancestral breadth of catalytic capability. The study's authors note that textbook categories of evolutionary process, such as neofunctionalisation and subfunctionalisation, are inadequate to describe the shift from diverse to specific functionality. Their work looks at small changes in enzyme function, corresponding to changes only at the fourth level of the EC number. The picture of modern enzymes as having evolved from precursors with lower activity and broader specificity is consistent with that suggested by the Tawfik group [14], noting that a few mutations can improve the secondary activity of a moonlighting or promiscuous enzyme by several orders of magnitude without immediate and complete loss of the primary function.

The Babbitt group [15, 16] have created the Structure-Function Linkage Database (SFLD), which takes a bigger-picture view of protein evolution. They study superfamilies of evolutionarily related enzymes with whose chemical functions are related, but nonetheless diverse. While they catalogue only a few families, they do so in considerable detail. For example, the radical SAM superfamily contains 85 separate reactions. Several subsets of these reactions have very similar EC numbers, within the same third level subclass, and all members of the superfamily share a common mechanistic step. Nonetheless, the superfamily is still functionally broad enough to include examples from four of the six EC classes. This illustrates how very similar chemical mechanisms can be co-opted to catalyse reactions which are well-separated within the EC scheme. This ability of similar enzymes to catalyse diverse reactions provides support for Lazcano & Miller's patchwork model [17] of recruitment of enzymes to metabolic pathways.

In an ambitious project, Furnham et al. [18] have created the FunTree description of the evolution of function within each CATH homologous superfamily of protein domains. For this purpose, they have created hundreds of phylogenies describing the evolution of function. They consider 379 superfamilies within which enzymatic functions have evolved – many of which have more than one so-called structurally similar group (SSG), with a separate tree needed for each SSG. Producing that number of individual family trees of enzymes is not a trivial task, and the best option is using a consistent automated approach. The resulting trees give a protein-centric picture of evolution, but their construction is guided by an underlying tree of relationships between species. Inevitably, a tree generated by such a high-throughput approach may differ from the tree that would result for the same superfamily if a phylogenist were given months to fine-tune the selection of data, parameters and model-building software to their complete satisfaction. FunTree is a resource which allows one to look at the evolution of enzyme function in every annotated superfamily where catalytic capability is present, right across protein structure space. However, surprising or unexpected results from this analysis will require further investigation. We encountered such phylogenetic ambiguity when devising a methodology [19] to investigate the still-unresolved question of whether metallo-beta-lactamase activity (EC 3.5.2.6) has arisen twice independently in the same CATH superfamily 3.60.15.10, after we had used the FunTree phylogeny as a starting point. Phylogenetic trees of enzymes can identify presumptive evolutionary events, but they do not in themselves assign functions to the putative ancestors. While Voordeckers et al. [13] were able to do this by expressing reconstructed ancient sequences, typically the required resources to do this are unavailable; in our case we used homology modelling alongside protein function prediction software. In any case, the ancestral sequences are subject to uncertainty, as therefore are estimates of their catalytic power

and substrate specificity. Nonetheless, the potential for beta-lactamase activity to evolve anew is significant in the context of current concerns over the rapid spread of antibiotic resistance.

Martinez Cuesta et al. [20] carried out a detailed study of evolutionary events involving isomerases. Since this EC class is united most obviously by its members having reaction products that happen to by isomers of the substrates, it is not immediately clear how much shared chemistry there might be. For other EC classes such as oxidoreductases, hydrolases and ligases, likely similarity in reactions and mechanisms is more obvious. Indeed, those authors find that isomerases are very frequently involved in out-of-class evolutionary changes, just as might have been expected from the eclectic nature of the categorisation that defines the class. Their data show a number of evolutionary changes to isomerases where the change in reaction catalysed is small in chemoinformatics terms, but nonetheless sufficient to result in a change of top-level EC class, and hence they describe multiple examples of similar chemical reactions being far apart in the EC classification.

Smock et al. [21] have used a combination of bioinformatics and directed evolution experiments to look at the structural aspects of protein evolution. Although they carry out selection based on binding proteins, the insights into structural evolution of proteins are likely to apply equally to enzymes. Smock et al. identified beta-propeller sequences from Pfam [7] and used phylogenetic methods to reconstruct sequences of putative ancestral motifs. They used deliberately error-prone PCR to introduce diversity into their library of motifs. By means of duplication and fusion, lectin-like proteins were assembled from these motifs. Using iterations of directed evolution, the authors of the study were able to select variants with optimal ability to bind the glycoprotein mucin. Thus they found that beta-propeller proteins could be formed by duplication and fusion of small sequence segments of around 50 residues, and they argued that foldability is the main property being evolutionarily selected for in this case. The application of directed evolution approaches to artificially change or improve the properties of enzymes has been reviewed at some length by Currin et al. [22] Gilson et al. [23] used lattice models of protein folding and data from SCOP in their study of the relationship between the divergence of protein sequence and structure, and how fitness and foldability are preserved along evolutionary trajectories. They suggest that discovery of new structures by evolving proteins is likely to require traversal of regions of lower fitness. All these studies have clear applicability to protein re-engineering.

The Importance of Chemical Mechanism

The structural and evolutionary information in CATH [1], SCOP [3], or ECOD [6] and the chemical transformations inherent in EC numbers provide complementary ways of describing and categorising enzymes. A further dimension to the conceptual space of enzyme functions comes from considering the chemical mechanisms employed, that is the different routes through which the molecular transformations are brought about. These cannot be deduced directly from the substrates and products, but instead require specific experimental or computational studies to identify the sets of intermediates and transition states through which these routes pass. Such studies have traditionally been published in biochemistry, organic chemistry or computational chemistry journals, each of which may require some expertise to translate into a form comprehensible even to experts in the adjacent fields. To address this, the database MACiE [24] provides a catalogue of around 350 mechanisms in both human-readable and computer-readable forms. MACiE, like most approaches to

structural bioinformatics, was originally based on a non-homologous dataset, albeit with later additions. Given this, and also because of the experimental limitations on mechanism determination, MACiE mostly provides a zoomed-out overview of the totality of enzyme space, and only occasionally includes close neighbours with small differences. SFLD [15], in contrast, has very good coverage of a few specific regions of that space, corresponding to a few specific functional superfamilies. While not concentrating on mechanism to the extent of MACiE, the SFLD's superfamilies are partly defined by a shared mechanistic step common to their reactions. Thus the kind of divergent evolution described by SFLD involves mechanistic similarity. By way of contrast, convergently evolved instances of similar chemical transformations typically have mechanisms that are significantly less similar than are their overall reactions [25]. A third mechanistic database, EZCatDB [26], currently contains mechanistic data on 878 enzymes classified according to its own RLCP system. By combining the steps constituting MACiE mechanisms with the associated catalytic domains and fold ages, Nath et al. [27] produced a somewhat speculative account of the development of enzyme mechanistic and functional diversity over evolutionary time, see Figure 1.

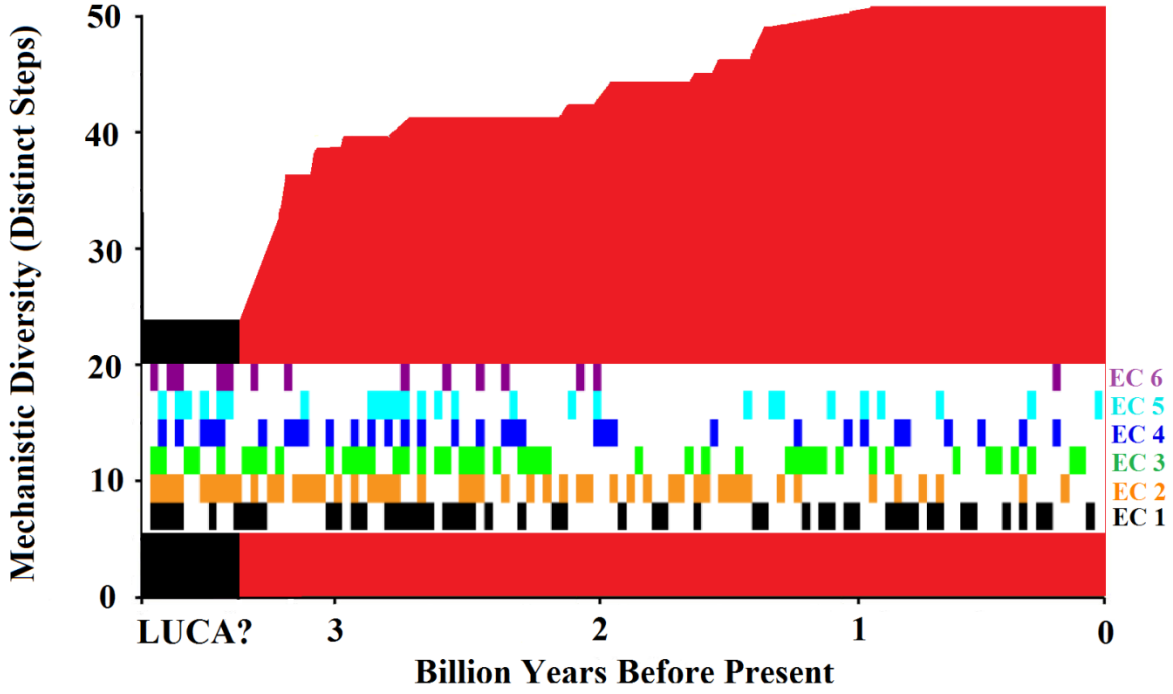

Figure 1. Growth of the diversity of enzyme chemistry over evolutionary time, created using data from Nath et al. [27]. This work uses fold ages from MANET [9] and mechanistic steps from MACiE [24]. The last universal common ancestor (LUCA) may possibly lie in the region indicated. The multi-coloured inset shows functions of different EC classes arising over time.

Protein Function Prediction

As suggested above, assignment of function to enzymes is ideally done by experimental means. Considering the extensive resources required to achieve this, however, it is more usual to utilise computer-based function prediction [28]. The difficulty of function prediction for a particular protein varies greatly, depending on the available sequence and structure information and on the identification of homologues, the available methods being based on one or both of sequence and

structure [29, 30, 31]. The majority of the predictive load is usually carried by sequence [32, 33]. Prediction of protein function on a large scale remains a significant challenge. As the volume of genomic data appearing each year far exceeds the capacity for manual annotation, let alone experiment, assignment of function to novel genes and proteins needs to be an automatic process. Unfortunately, an unknown but possibly significant proportion of such annotations in bioinformatics databases may be erroneous, with misannotations then propagating as they are transferred to fresh homologues and other databases [34]. Such misannotations could then be further propagated to related sequences in future prediction exercises. Indeed, the circularity of the combined process of propagating annotations and then predicting function, based on the same annotations and homologies, may be problematic. Sequence-based enzyme function predictions based on EC number annotations in databases can indeed give very impressive results [35] and such predictive exercises can be extended to include mechanism [36], both processes usually operating mostly via the detection of homology - although 3D structure-based methods also exist [37, 38, 39]. Using mechanisms and catalytic chains as defined in MACiE, the corresponding UniProt sequences are interrogated against InterPro signatures [29] to re-express the MACiE entries in terms of the signatures present in them. This information forms the input into a machine learning exercise [36] to associate test sequences with enzymatic mechanisms, as shown in Figure 2.

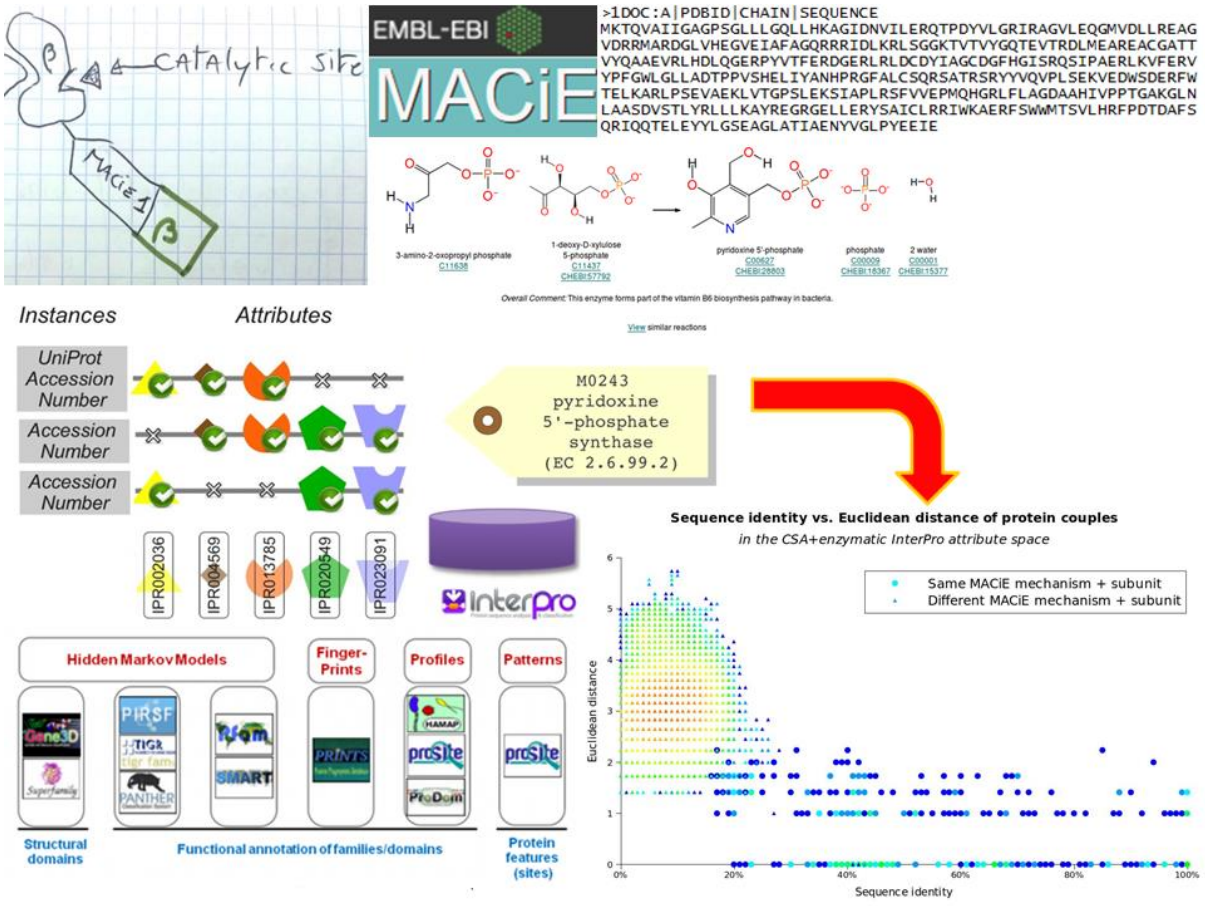

Figure 2. Clockwise from top right: Sequences, mechanisms at catalytic domain definitions are taken from MACiE and combined in a machine learning exercise with InterPro signatures, which are themselves derived from a diversity of source databases. All these data, bottom right, can be used to predict mechanisms for new query sequences [36].

Recently, the success of different groups' approaches to protein function prediction has been evaluated in the CAFA (Critical Assessment of Functional Annotation) exercises, of which the second [40] assessed predictions made in late 2013 and focussed on predicting the Gene Ontology (GO) [41] terms associated with proteins. This process was lengthy, and notably involved a period of several months in which new annotations on the many target proteins were allowed to accumulate in the literature before these freshly assigned labels were used in the assessment of the already-submitted entries. Given the large numbers of sequences and of ontological terms being predicted, the participants' freedom to predict only subsets, and the ever growing nature of the available experimental annotations, it was inevitable that submitted predictions would be both incomplete and partially incorrect. The process of assessment and criteria for evaluation were therefore not straightforward, and this complexity meant that CAFA2 had no clear 'winner'. Nevertheless, the official paper reporting the exercise convincingly argued that the quality of predictions had improved since the previous exercise [40,42].

Amongst the successful entries was the Orengo group's functional clustering of CATH superfamilies into functional families (FunFams) by the FunFHMMer method, as reported by Das et al. [43] The Gough group [44] made extensive use of SCOP data to predict functional annotations at the domain level by statistical inference. Also impressing in CAFA2, the FFPred3 method of Cozzetto et al. [45] assigns functional labels based on predicted biophysical attributes associated with protein secondary structure, and is especially useful in those hard-to-predict cases where no relevant information is available from homology. The Multi-Source k-Nearest Neighbor (MS-kNN) approach of Lan et al. [46] achieved its success by identifying proteins similar to the query as its neighbours, and then inferring its function from a weighted average of their functions. Another very successful approach was that of Gong et al. [47], who trained their algorithm to identify the functionally discriminating residues relevant to each GO term. Some of the methods in CAFA2 specialised in identifying particular functions, rather than being general purpose; for instance APRICOT [48] is a sequence signature approach designed specifically to identify RNA binding proteins. APRICOT makes substantial use of both InterPro [29] and CDD [5].

Conclusions and Future Priorities

While protein function prediction is a well-established field, more progress can be made by making databases more robust against propagation of erroneous information, and by describing both molecular and biological function in more specific and detailed ways. For enzyme reactions, more basic science is required to investigate if and how mechanism is affected by relatively modest evolutionary changes in sequence and structure. Alongside this, more enzyme mechanisms need to be determined and consistently recorded wherever possible. Applications such as protein re-engineering and even de novo enzyme design [49] will require a deep understanding of the interplay of chemistry with protein structure. Such advances promise major applications in fields as diverse as medicine, agriculture, food, laundry, deodorants and green energy. Further understanding of how enzyme functions evolve is another major priority, especially in the context of rapidly increasing antibiotic resistance [50].

References

1. Sillitoe I, Lewis TE, Cuff A, Das S, Ashford P, Dawson NL, et al. (2015) CATH: comprehensive structural and functional annotations for genome sequences. Nucleic Acids Research 43:D376-D381
2. * Das S, Dawson NL, Orengo CA (2015). Diversity in protein domain superfamilies. Current Opinion in Genetics & Development 35:40-49
A readable and up-to-date discussion of the interplay of protein structure and function, highly relevant to the study of enzyme evolution. Clearly described concepts are illustrated with specific examples.
3. Hubbard TJP, Murzin AG, Brenner SE, Chothia C (1997). SCOP: a structural classification of proteins database. Nucleic Acids Research 25:236-239
4. Fox NK, Brenner SE, Chandonia JM (2014). SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures. Nucleic Acids Research 42:D304-309
5. Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, et al. (2012) CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Research, 41:D348-D352
6. * Cheng H, Schaeffer RD, Liao Y, Kinch LN, Pei J, Shi S, et al. (2014) ECOD: An Evolutionary Classification of Protein Domains. PLOS Computational Biology 10:e1003926
Although originally based on SCOP, ECOD is specifically designed primarily to reflect evolutionary rather than structural relationships amongst protein domains.
7. Finn RD, Tate J, Mistry J, Coggill PC, Sammut SJ, Hotz HR, Ceric G, et al. (2008). The Pfam protein families database. Nucleic Acids Res. 36:D281–288
8. Kim KM, Caetano-Anolles G (2011). The proteomic complexity and rise of the primordial ancestor of diversified life. BMC Evolutionary Biology 11:140
9. Kim HSS, Mittenthal JE, Caetano-Anolles G (2006) MANET: tracing evolution of protein architecture in metabolic networks. BMC Bioinformatics 7:351
10. Winstanley HF, Abeln S, Deane CM (2005). How old is your fold? Bioinformatics 21:i449-i458
11. Dupont CL, Yang S, Palenik B, Bourne PE. Modern proteomes contain putative imprints of ancient shifts in trace metal geochemistry (2006) Proceedings of the National Academy of Sciences USA 103:17822-17827
12. Ranea JAG, Sillero A, Thornton JM, Orengo CA (2006). Protein Superfamily Evolution and the Last Universal Common Ancestor (LUCA). Journal of Molecular Evolution 63:513-525
13. Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, et al. (2012) Reconstruction of Ancestral Metabolic Enzymes Reveals Molecular Mechanisms Underlying Evolutionary Innovation through Gene Duplication. PLoS Biol. 10:e1001446
14. Khersonsky O, Roodveldt C, Tawfik D (2006) Enzyme promiscuity: evolutionary and mechanistic aspects. Current Opinion in Chemical Biology 10:498-508
15. Akiva E, Brown S, Almonacid DE, Alan, Custer AF, Hicks MA, et al. (2014) The Structure–Function Linkage Database. Nucleic Acids Research 42:D521-D530
16. Mashiyama ST, Malabanan MM, Akiva E, Bhosle R, Branch MC, Hillerich B, et al. (2014) Large-Scale Determination of Sequence, Structure, and Function Relationships in Cytosolic Glutathione Transferases across the Biosphere. PLOS Biology. 12:e1001843
17. Lazcano A, Miller SL. (1999) On the Origin of Metabolic Pathways. J Mol Evol 49:424-431
18. ** Furnham N, Dawson NL, Rahman SA, Thornton JM, Orengo CA. (2016) Large-Scale Analysis Exploring Evolution of Catalytic Machineries and Mechanisms in Enzyme Superfamilies. Journal of Molecular Biology. 428:253-267
A substantial and important paper giving an overview picture of enzyme evolution. While some other papers are essentially individual case studies, this one gives a statistically meaningful description of the overall features of enzyme evolution across the full diversity of structures and functions.
19. Alderson RG, Barker D, Mitchell JBO (2014). One origin for metallo-β-lactamase activity, or two? An investigation assessing a diverse set of reconstructed ancestral sequences based on a sample of phylogenetic trees. Journal of Molecular Evolution 79:117-129
20. Martinez Cuesta S, Furnham N, Rahman SA, Sillitoe I, Thornton JM (2014) The evolution of enzyme function in the isomerases. Current Opinion in Structural Biology 26:121-130
21. * Smock RG, Yadid I, Dym O, Clarke J, Tawfik DS (2016). De Novo Evolutionary Emergence of a Symmetrical Protein Is Shaped by Folding Constraints. Cell, 164:476-486
A combination of bioinformatics and directed evolution is used to show that beta-propellers can be formed from duplication and fusion of short segments of around 50 residues.

22. Currin A, Swainston N, Day PJ, Kell DB (2015) Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently. Chem Soc Rev 44:1172-1239
23. * Gilson AI, Marshall-Christensen A, Choi JM, Shakhnovich EI (2017) The Role of Evolutionary Selection in the Dynamics of Protein Structure Evolution. Biophysical Journal, 112:1350-1365
Study using lattice models of protein folding and data from SCOP to explore the relationship between sequence identity and structure similarity along evolutionary trajectories. They study the onset of the twilight zone of limiting sequence divergence where structural similarity is no longer maintained, and consider the requirements of foldability and fitness along evolutionary trajectories that could lead to new structures.
24. Holliday GL, Almonacid DE, Bartlett GJ, O'Boyle NM, Torrance JW, Murray-Rust P, et al. (2007) MACiE (Mechanism, Annotation and Classification in Enzymes): novel tools for searching catalytic mechanisms. Nucleic Acids Research 35:D515-D520
25. Almonacid DE, Yera ER, Mitchell JBO, Babbitt PC. Quantitative Comparison of Catalytic Mechanisms and Overall Reactions in Convergently Evolved Enzymes: Implications for Classification of Enzyme Function (2010) PLoS Comput Biol. 6:e1000700
26. Nagano N (2005) EzCatDB: the Enzyme Catalytic-mechanism Database. Nucleic Acids Research, 33(suppl 1):D407-D412
27. Nath N, Mitchell JBO, Caetano-Anolles G (2014) The Natural History of Biocatalytic Mechanisms. PLoS Comput Biol. 10:e1003642
28. Friedberg I (2006) Automated protein function prediction-the genomic challenge. Briefings in Bioinformatics 7:225-242
29. Mitchell A, Chang HYY, Daugherty L, Fraser M, Hunter S, Lopez R, et al. (2015) The InterPro protein families database: the classification resource after 15 years. Nucleic Acids Research 43:D213-D221
30. Barker JA, Thornton JM (2013) An algorithm for constraint-based structural template matching: application to 3D templates with statistical analysis. Bioinformatics 19:1644–1649
31. Laskowski RA, Watson JD, Thornton JM (2005) ProFunc: a server for predicting protein function from 3D structure. Nucleic Acids Research 33:W89-W93
32. Beattie KE, De Ferrari L, Mitchell JBO (2015) Why do Sequence Signatures Predict Enzyme Mechanism? Homology versus Chemistry. Evolutionary Bioinformatics 11:267-274
33. Zvacek C, Friedrichs G, Heizinger L, Merkl R. An assessment of catalytic residue 3D ensembles for the prediction of enzyme function (2015) BMC Bioinformatics 16:359
34. Furnham N, Garavelli JS, Apweiler R, Thornton JM (2009) Missing in action: enzyme functional annotations in biological databases. Nat Chem Biol. 5:521-525
35. De Ferrari L, Aitken S, van Hemert J, Goryanin I. (2012) EnzML: multi-label prediction of enzyme classes using InterPro signatures. BMC Bioinformatics 13:61
36. De Ferrari L, Mitchell JBO (2014) From sequence to enzyme mechanism using multi-label machine learning. BMC Bioinformatics 15:150
37. Hermann JC, Marti-Arbona R, Fedorov AA, Fedorov E, Almo SC, Shoichet BK, et al. (2007) Structure-based activity prediction for an enzyme of unknown function. Nature 448:775-779
38. Fan H, Hitchcock DS, Seidel RD, Hillerich B, Lin H, Almo SC, et al. (2013) Assignment of Pterin Deaminase Activity to an Enzyme of Unknown Function Guided by Homology Modeling and Docking. J Am Chem Soc.135:795-803
39. Nilmeier JP, Kirshner DA, Wong SE, Lightstone FC (2013) Rapid catalytic template searching as an enzyme function prediction procedure. PloS One 8:e62535
40. Jiang Y, Oron TR, Clark WT, Bankapur AR, D'Andrea D, Lepore R, et al. (2016) An expanded evaluation of protein function prediction methods shows an improvement in accuracy. Genome Biology 17:184
41. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. (2000) Gene Ontology: tool for the unification of biology. Nature Genetics, 25:25-29
42. Gillis J, Pavlidis P (2013) Characterizing the state of the art in the computational assignment of gene function: lessons from the first critical assessment of functional annotation (CAFA). BMC Bioinformatics 14(Suppl 3):S15
43. * Das S, Lee D, Sillitoe I, Dawson NL, Lees JG, Orengo CA (2015) Functional classification of CATH superfamilies: a domain-based approach for protein function annotation. Bioinformatics, 31:3460-3467
A method based on functional sub- classification of the CATH homologous superfamilies. The paper extensively compares its results with those of other protein domain classifications.
44. Fang H, Gough J (2013) A domain-centric solution to functional genomics via dcGO Predictor, BMC Bioinformatics 14(Suppl 3):S9.

45. * Cozzetto D, Minneci F, Currant H, Jones DT (2016) FFPred 3: feature-based function prediction for all Gene Ontology domains, Scientific Reports, 6:31865
The Jones group's FFPred3 method assigns GO labels based on the predicted biophysical attributes of the protein's secondary structure, and is especially useful where no predictively useful information is available from homology to proteins of known function.
46. ** Lan L, Djuric N, Guo Y, Vucetic S (2013) MS-kNN: protein function prediction by integrating multiple data sources, BMC Bioinformatics 14(Suppl 3):S8
This approach uses various measures of protein similarity to identify neigbours of a given query protein. It then invokes weighted averaging of the neighbours' properties in order to predict functional annotations. Given that the k-Nearest Neighbours method combines locality of prediction with global coverage of the search space, this is a very promising method.
47. Gong Q, Ning W, Tian W (2016) GoFDR: A sequence alignment based method for predicting protein functions. Methods, 93:3-14
48. Sharan M, Forstner KU, Eulalio A, Vogel J (2017) APRICOT: an integrated computational pipeline for the sequence-based identification and characterization of RNA-binding proteins. Nucleic Acids Research, 45:e96
49. Jiang L, Althoff EA, Clemente FR, Doyle L, Röthlisberger D, Zanghellini A, et al. (2008) De Novo Computational Design of Retro-Aldol Enzymes. Science 319:1387-1391
50. * Perry J, Waglechner N, Wright G. (2016) The Prehistory of Antibiotic Resistance. Cold Spring Harbor Perspectives in Medicine 6:a025197
A useful perspective on the history of antibiotic resistance; required reading for those who want to be well informed about perhaps the most timely practical application of enzyme evolution.