

## Editorial overview: Catalysis and regulation

Janet Thornton, Christine Orengo

### Introduction

Catalysis and its regulation have long been recognised as essential components of life. In this special issue of Current Opinions we have brought together opinions from leaders in the field, who provide an overview of our increasing understanding of enzyme promiscuity and its importance in biological systems, and who discuss how new functions could have emerged during evolution, considering both experimental and computational approaches. In some of the reviews, experimental analyses of specificity and promiscuity are considered for specific systems. Whilst other reviews provide a broad overview of enzyme evolution by applying computational approaches, based on sequence, structure and functional data, illustrating that many new catalytic functions emerged early, during bacterial evolution, and that most reactions have evolved from other 'older' functions. More specialised reviews cover enzyme evolution in plant metabolism and the emergence of pseudoenzymes, where catalytic activity has been lost, yet the genes are retained to play a different biological role. We also include reviews on enzyme fitness and design and the role of promiscuity in evolving new functions. Finally, some selected examples of allostery are reviewed that bring catalysis and regulation together, since one is rarely found without the other.

Shelley Copley shines a light on enzyme promiscuity highlighting the fact that most, if not all, enzymes are capable of catalysing secondary reactions. Most of these reactions are not physiologically important, but make enzymes very suitable for protein engineering through directed evolution, facilitating the emergence of synthetic biology. New high throughput methods, including droplet screening, reveal multiple activities, but many of these activities are rather weak usually due to imperfect geometry for catalysis. Shelley distinguishes enzymes that have evolved the ability to act on many substrates as part of their biological *raison d'être* (e.g. detoxification) from enzymes which are indiscriminate, catalysing many physiologically irrelevant reactions in addition to their primary reaction. Many enzymes have a huge activity space, with tens of potential substrates, making it difficult to identify those which are physiologically relevant. Most microbes contain 1000–2000 enzymes; if each has 10 promiscuous activities (a modest guess), over 20 000 promiscuous activities may be available to fuel evolutionary innovation within a single microbe and Shelley considers the potential for green biotechnology. New approaches, like activity-based metabolic profiling can recognize new metabolites, but are still limited by the challenges of identifying components in complex metabolic mixtures. High-throughput screening of enzymes for promiscuous activity has improved in recent years due to advances in robotics and microfluidics but it is still difficult to find suitable fluorescent molecules, as these are often only applicable for certain activities.

Unlike most other organisms, plants cannot move and so have evolved unique defense and attraction strategies, using small metabolites as their agents. Last et al. present examples of evolution of plant specialised metabolism, that is, the enzymes, reactions and pathways evolved to make novel metabolites. Many of these compounds are lineage-specific and this review explores factors involved in such enzyme evolution and discusses how these results

in metabolite diversity. They discuss new enzymes involved in producing different insecticidal sugars in tomatoes, metabolites for attracting insect pollinators and metabolites for defense against microbes. For each example, they describe the changes in sequence and structure which lead to these new functions and identify a range of mechanisms by which such changes in function are achieved, from single amino acid changes to conformational flexibility and open pockets, which allow many promiscuous interactions. Interestingly they describe an example where the control of gene expression determines the dominant activity. It is clear that promiscuity and gene duplication play major roles in the evolution of specialised novel metabolic enzymes and product diversity, leading to new pathways and new metabolites which are used according to the ecological niche of the plant.

In some cases, evolution leads to loss of function. Evers et al.'s review considers pseudoenzymes - a specific sub-group of enzyme evolution, wherein despite loss of enzyme activity, the associated genes are retained. This useful and comprehensive review presents our current knowledge of pseudoenzymes - highlighting that they constitute about 10% of the proteome and perform many different functions. The field (and the review) is dominated by the pseudokinases, for which there are many examples. The review focuses on the new functions of the pseudoenzymes and presents numerous recent examples. New functions include allosteric regulation, acting as a molecular switch and integrator, forming the hub for assembling protein complexes and as modulators of substrate availability or holoenzyme assembly. The authors use structural information to reveal how these functional changes have been achieved.

Two contributions focus on computational analysis of protein families to provide an overview of the evolution of new functions from old. Mitchell et al. consider the evolutionary history of enzymes and highlight the role of enzyme 3D structure to identify ancient proteins (and their likely folds and chemistries) likely to be present in the last universal ancestor (LUCA). Reconstruction of ancestral relatives combined with experimental testing reveals how enzymes evolved from relatives with broad enzymatic profiles but generally became more specific, though there are examples of modern proteins that retain the breadth of specificities and activities. They highlight useful computational resources, which integrate structure, sequence and experimental data to support detailed evolutionary studies, for example, SFLD data had revealed relatives sharing mechanistic similarities despite diverse chemistries and co-option to different pathways whilst FunTree based analyses showed that metallo-beta lactamases - responsible for antibiotic resistance - had emerged in evolution multiple times. Another useful resource, MACIE, providing information on mechanism, was used together with evolutionary data on protein folds, to highlight the emergence of mechanistic diversity since LUCA. Tyzak et al. also use computational approaches and focus on recent developments in the application of structural bioinformatics methods to understand the evolution of specificity and guide de-novo enzyme design. They attempt to describe the different types of enzyme evolution most frequently observed and the structural basis for changes in function. Bringing together protein sequence, structure and function they illustrate the amazing ability of many enzyme families to evolve new functions from old ones.

Thornton et al.'s excellent opinion piece on protein specificity considers whether specific enzymes evolve from multifunctional ancestors and whether promiscuity of an enzyme is

essential for the evolution of new functions. They review the considerable evidence for the widely accepted hypothesis that ancestral proteins were generalists. However, three observations have led the authors to query this hypothesis - moonlighting, non-natural processes found in 'strong-arm' directed evolution' experiments and the fact that many of the functions evolved in distant evolutionary times, so that many of the residues in the proteins have changed, obscuring those that are responsible for the new function. Recently published experimental studies suggest that although promiscuity is often a factor, there is evidence that some new functions have evolved de novo in some protein families from ancestral proteins that did not already have those functions. They highlight the use of ancestral protein reconstruction to characterise the ancient enzymes which can then be tested experimentally to characterise their catalytic properties. They present three examples of very different enzymes, in which new functions can arise by simple evolutionary changes, with just one or a few mutations responsible for changing the specificity. These include the key metabolic enzymes, malate and lactate dehydrogenases, which diverged between 700 and 900 million years ago; allosteric regulation in vertebrate steroid receptors and DNA specificity of transcription factors. They conclude by stating that acquisition of new functions during evolution is neither difficult nor rare.

Exploiting evolutionary analyses for protein design, Goldsmith et al. describe how such studies provide insights to guide protein design and offer clues on how to avoid dead-ends. Although most enzymes are not close to the maximum, diffusion controlled limit, natural selection can ensure rapid exploration of alternative routes. For example, bacterial lactonases have evolved over a few decades to act on a previously non-cognate substrate with a significant increase in activity (105) - up to the diffusion limit. It is much harder to achieve these improvements in the lab, where typically at least 10 rounds of mutations are required to get an improvement of 3 orders of magnitude or more. Although early rounds of mutation can yield substantial improvements (e.g. up to 103), subsequent optimisations are subject to diminishing returns (as low as 0.18) as they are often in the secondary shell, compensating for the destabilizing effects of the early mutations in the primary shell, and their contribution to catalytic efficiency is typically low. A common cause for a local optimisation plateau is destabilizing effects - these often need to be rescued by combinations of mutations. One possible solution is to swap entire secondary structure elements to survey a larger region of sequence space. This review cites many helpful ways for 'pulling out' of local optimization plateaus. Ancestral inference can suggest highly stable, mutation tolerant starting points and computational strategies can identify large numbers (e.g. >50) of potentially stabilizing mutations as stability boosts appear critical for escaping these plateaus.

Finally, the control of catalysis, which is essential in all organisms, is reviewed by Loria et al. who present selected examples of allostery - the mechanism by which information, transferred between spatially distinct sites in the protein, affects catalysis - demonstrating how small perturbations can lead to large disruptions in the catalytic activity. In protein tyrosine phosphatases, implicated in many pathologies, inhibitor binding more than 20 Å from the active site has been shown to lock the catalytic acid loop in an open (inactive) conformation, with MD analyses showing dramatic changes. In contrast, in the kinase PKA, nucleotide binding triggers an allosteric mechanism altering substrate affinity by affecting the extent of closure of a Gly-rich loop, essential for catalysis. Mutation of a single tyrosine,

at the interface of the small and large lobes in this protein, can result in up to 400 fold losses in catalytic efficiency and affect millisecond motions that control the open to closed transition. Similarly, with Imidazole Glycerol Phosphate Synthase, a range of effectors induce millisecond motions, each invoking unique pathways of allosteric information transfer. Disruption of these pathways suggests increased rigidity of the active site which needs to rearrange to stabilise the catalytic intermediate.