Chem Soc Rev

REVIEW ARTICLE

View Article Online
View Journal | View Issue

Cite this: *Chem. Soc. Rev.*, 2015, **44**, 1172

Synthetic biology for the directed evolution of protein biocatalysts: navigating sequence space intelligently

Andrew Currin,[abc] Neil Swainston,[acd] Philip J. Day[ace] and Douglas B. Kell*[abc]

The amino acid sequence of a protein affects both its structure and its function. Thus, the ability to modify the sequence, and hence the structure and activity, of individual proteins in a systematic way, opens up many opportunities, both scientifically and (as we focus on here) for exploitation in biocatalysis. Modern methods of synthetic biology, whereby increasingly large sequences of DNA can be synthesised *de novo*, allow an unprecedented ability to engineer proteins with novel functions. However, the number of possible proteins is far too large to test individually, so we need means for navigating the 'search space' of possible protein sequences efficiently and reliably in order to find desirable activities and other properties. Enzymologists distinguish binding (K_d) and catalytic (k_{cat}) steps. In a similar way, judicious strategies have blended design (for binding, specificity and active site modelling) with the more empirical methods of classical directed evolution (DE) for improving k_{cat} (where natural evolution rarely seeks the highest values), especially with regard to residues distant from the active site and where the functional linkages underpinning enzyme dynamics are both unknown and hard to predict. Epistasis (where the 'best' amino acid at one site depends on that or those at others) is a notable feature of directed evolution. The aim of this review is to highlight some of the approaches that are being developed to allow us to use directed evolution to improve enzyme properties, often dramatically. We note that directed evolution differs in a number of ways from natural evolution, including in particular the available mechanisms and the likely selection pressures. Thus, we stress the opportunities afforded by techniques that enable one to map sequence to (structure and) activity *in silico*, as an effective means of modelling and exploring protein landscapes. Because known landscapes may be assessed and reasoned about as a whole, simultaneously, this offers opportunities for protein improvement not readily available to natural evolution on rapid timescales. Intelligent landscape navigation, informed by sequence-activity relationships and coupled to the emerging methods of synthetic biology, offers scope for the development of novel biocatalysts that are both highly active and robust.

Received 20th October 2014

DOI: 10.1039/c4cs00351a

www.rsc.org/csr

Introduction

Much of science and technology consists of the search for desirable solutions, whether theoretical or realised, from an enormously larger set of possible candidates. The design, selection and/or improvement of biomacromolecules such as proteins represents a particularly clear example.[1] This is because natural molecular evolution is caused by changes in protein primary sequence that (leaving aside other factors such as chaperones and post-translational modifications) can then fold to form higher-order structures with altered function or activity; the protein then undergoes selection (positive or negative) based on its new function (Fig. 1). Bioinformatic analyses can trace the path of protein evolution at the sequence level[2–4] and match this to the corresponding change in function.

Proteins are nature's primary catalysts, and as the unsustainability of the present-day hydrocarbon-based petrochemicals industry becomes ever more apparent, there is a move towards carbohydrate feedstocks and a parallel and burgeoning interest in the use of proteins to catalyse reactions of non-natural as well as of natural chemicals. Thus, as well as observing the products of natural evolution we can now also initiate changes, whether

[a] *Manchester Institute of Biotechnology, The University of Manchester, 131, Princess St, Manchester M1 7DN, UK. E-mail: dbk@manchester.ac.uk; Web: http://dbkgroup.org/; @dbkell; Tel: +44 (0)161 306 4492*
[b] *School of Chemistry, The University of Manchester, Manchester M13 9PL, UK*
[c] *Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM), The University of Manchester, 131, Princess St, Manchester M1 7DN, UK*
[d] *School of Computer Science, The University of Manchester, Manchester M13 9PL, UK*
[e] *Faculty of Medical and Human Sciences, The University of Manchester, Manchester M13 9PT, UK*

in vivo or *in vitro*, for any target sequence. When the experimenter has some level of control over what sequence is made, variations can be introduced, screened and selected over several iterative cycles ('generations'), in the hope that improved variants can be created for a particular target molecule, in a process usually referred to as directed evolution (Fig. 2) or DE. Classically this is achieved in a more or less random manner or by making a small number of specific changes to an existing sequence (see below); however, with the emergence of 'synthetic biology' a greater diversity of sequences can be created by assembling the desired sequence *de novo* (without a starting template to amplify from). Hence, almost any bespoke DNA sequence can be created, thus permitting the engineering of biological molecules and systems with novel functions. This is possible largely due to the reducing cost of DNA oligonucleotide synthesis and improvements in the methods that assemble these into larger fragments and even genomes.[5,6] Therefore, the question arises as to what sequences one should make for a particular purpose, and on what basis one might decide these sequences.

In this intentionally wide-ranging review, we introduce the basis of protein evolution (sequence spaces, constraints and conservation), discuss the methodologies and strategies that can be utilised for the directed evolution of individual biocatalysts, and reflect on their applications in the recent literature. To restrict our scope somewhat, we largely discount questions of the directed evolution of pathways (*i.e.* series of reactions) or gene clusters

Andrew Currin

Andrew Currin is a research associate at the Manchester Institute of Biotechnology, University of Manchester. He received his undergraduate degree (First Class Honours) in Biomedical Science in 2008 from the University of Birmingham, followed by a PhD in Biochemistry in 2012 at the University of Manchester. His work now focuses on the engineering of biocatalysts and developing DNA technology as a tool for synthetic biology. His interests lie in protein engineering, molecular biology, synthetic biology and drug discovery, with a particular focus on protein structure–function relationships and developing improved methodologies to investigate them.

Neil Swainston

Neil Swainston is a Research Fellow at the Manchester Institute of Biotechnology, University of Manchester. Following several years of industrial experience in proteomics bioinformatics software development with the Waters Corporation, he began his research career 8 years ago in the Manchester Centre for Integrative Systems Biology. His research interests span 'omics data analysis and management, genome-scale metabolic modelling, and enzyme optimisation through synthetic biology, and he has published over 25 papers covering these subjects. Driving all of these interests is a continued commitment to software development, data standardisation and reusability, and the development of novel informatics approaches.

Philip J. Day

Philip Day is Reader in Quantitative Analytical Genomics and Synthetic Biology, Manchester University, Philip leads interdisciplinary research for developing innovative tools in genomics and for pediatric cancer studies. Current research focuses on closed loop strategies for directed evolution gene synthesis and aptamer developments, and the development of active drug uptake using membrane transporters. Philip applies miniaturization for single cell analyses to decipher molecules per cell activities across heterogeneous cell populations. His research aims to providing exquisite quantitative data for systems biology applications and pathway analysis as a central theme for enabling personalised healthcare.

Douglas B. Kell

Douglas Kell is Research Professor in Bioanalytical Science at the University of Manchester, UK. His interests lie in systems biology, iron metabolism and dysregulation, cellular drug transporters, synthetic biology, e-science, chemometrics and cheminformatics. He was Director of the Manchester Centre for Integrative Systems Biology prior to a 5 year secondment (2008–2013) as Chief Executive of the UK Biotechnology and Biological Sciences Research Council. He is a Fellow of the Learned Society of Wales and of the American Association for the Advancement of Science, and was awarded a CBE for services to Science and Research in the New Year 2014 Honours list.

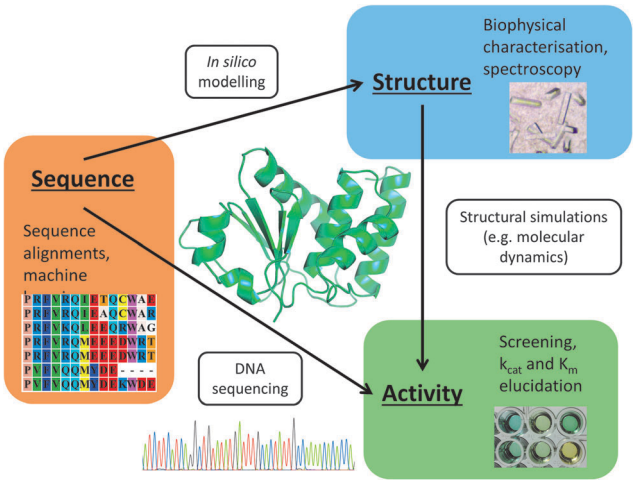


Fig. 1 Relationship between amino acid sequence, 3D structure (and dynamics) and biocatalytic activity. Implicitly, there is a host in which these manipulations take place (or they may be done entirely *in vitro*). This is not a major focus of the review. Typically, a directed evolution study concentrates on the relationships between protein sequence, structure and activity, and the usual means for assessing these are outlined (within the boxes). Many methods are available to connect and rationalise these relationships and some examples are shown (grey boxes). Thorough directed evolution studies require understanding of each of these parameters so that the changes in protein function can be rationalised, thereby to allow effective search of the sequence space. The key is to use emerging knowledge from multiple sources to navigate the search spaces that these represent. Although the same principles apply to multi-subunit proteins and protein complexes, most of what is written focuses on single-domain proteins that, like ribonuclease,[1342,1343] can fold spontaneously into their tertiary structures without the involvement of other proteins, chaperones, *etc.*

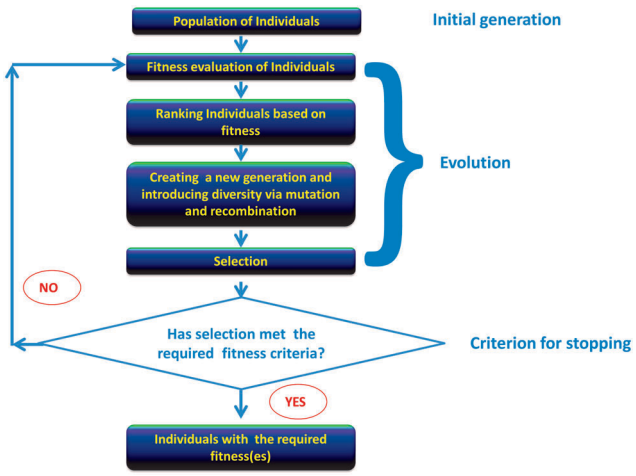


Fig. 2 The essential components of an evolutionary system. At the outset, a starting individual or population is selected, and one or more fitness criteria that reflect the objective of the process are determined. Next, the ability to rank these fitnesses and to select for diversity is created (by breeding individuals with variant sequences, introduced typically by mutation and/or recombination) in a way that tends to favour fitter individuals, this is repeated iteratively until a desired criterion is met.

(*e.g.* ref. 7 and 8) and of the choice[9] or optimization of the host organism or expression conditions in which such directed evolution might be performed or its protein products expressed, nor the process aspects of any fermentation or biotransformation. We also focus on catalytic rate constants, albeit we recognize the importance of enzyme stability as well. Most of the strategies we describe can equally well be applied to proteins whose function is not directly catalytic, such as vaccines, binding agents, and the like. Consequently we intend this review to be a broadly useful resource or portal for the entire community that has an interest in the directed evolution of protein function. A broad summary is given as a mind map in Fig. 3, while the various general elements of a modern directed evolution program, on which we base our development of the main ideas, appears as Fig. 4.

The size of sequence space

An important concept when considering a protein's amino acid sequence is that of (its) sequence space, *i.e.* the number of variations of that sequence that can possibly exist. Straightforwardly, for a protein that contains just the 20 main natural amino acids, a sequence length of N residues has a total number of possible sequences of 20^N. For $N = 100$ (a rather small protein) the number 20^{100} ($\sim 1.3 \times 10^{130}$) is already far greater than the number of atoms in the known universe. Even a library with the mass of the Earth itself – 5.98×10^{27} g – would comprise at most 3.3×10^{47} different sequences, or a miniscule fraction of such diversity.[10] Extra complexity, even for single-subunit proteins, also comes with incorporation of additional structural features beyond the primary sequence, like disulphide linkages, metal ions,[11] cofactors and post-translational modifications, and the use of non-standard amino acids (outwith the main 20). Beyond this, there may be 'moonlighting' activities[12] by which function is modified *via* interaction with other binding partners.

Considering sequence variation, using only the 20 'common' amino acids, the number of sequence variants for M substitutions in a given protein of N amino acids is $\frac{19^M \cdot N!}{(N-M)!M!}$.[13] For a protein of 300 amino acids with random changes in just 1, 2 or 3 amino acids in the whole protein this is 5700, *ca.* 16 million and *ca.* 30 billion, while even for a comparatively small protein of $N = 100$ amino acids, the number of variants exceeds 10^{15} when $M = 10$. Insertions can be considered as simply increasing the length of N and the number of variants to 21 (a 'gap' being coded as a 21st amino acid), respectively.

Consequently, the search for variants with improved function in these large sequence spaces is best treated as a combinatorial optimization problem,[1] in which a number of parameters must be optimised simultaneously to achieve a successful outcome. To do this, heuristic strategies (that find good but not provably optimal solutions) are appropriate; these include algorithms based on evolutionary principles.

The 'curse of dimensionality' and the sparseness or 'closeness' of strings in sequence space

One way to consider protein sequences (or any other strings of this type) is to treat each position in the string as a dimension in a discrete and finite space. In an elementary way, an amino

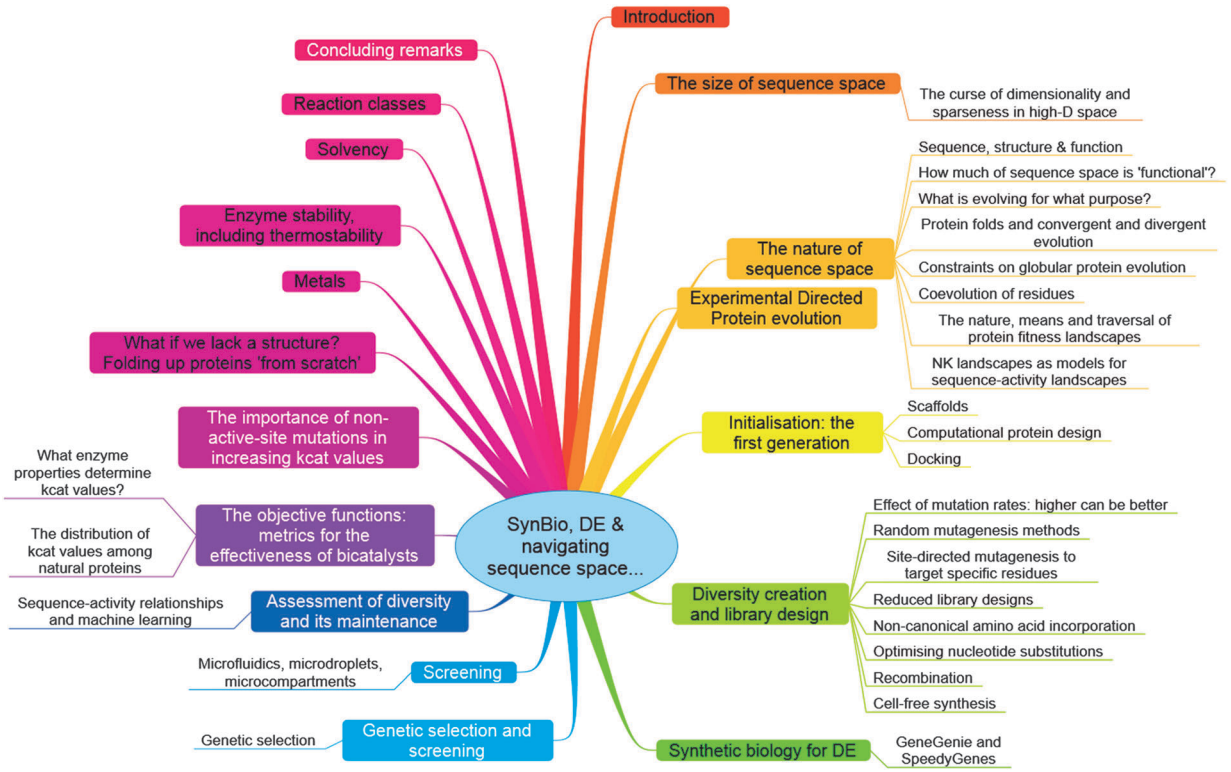


Fig. 3 A 'mind map'[1344] of the contents of this paper; to read this start at "twelve o'clock" and read clockwise.

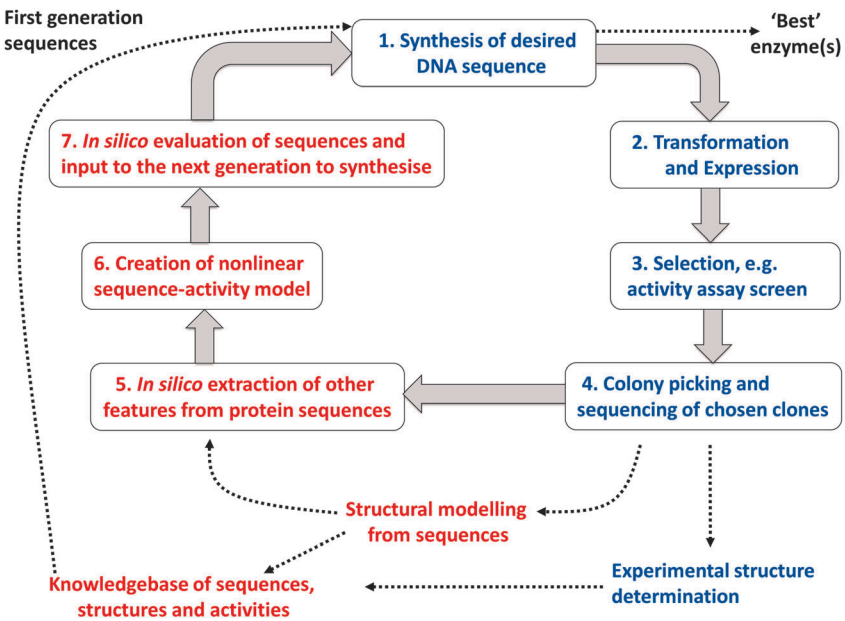


Fig. 4 An example of the basic elements of a mixed computational and experimental programme in directed evolution. Implicit are the choice of objective function (*e.g.* a particular catalytic activity with a certain turnover number) and the starting sequences that might be used with an initial or 'wild type' activity from which one can evolve improved variants. The core experimental (blue) and computational (red) aspects are shown as seven steps of an iterative cycle involving the creation and analysis of appropriate protein sequences and their attendant activities. Additional facets that can contribute to the programme are also shown (connected using dotted lines).

acid X has one of 20 positions in 1-dimensional space, an individual dimer X_kY_l has a specified position or represents a point (from 400 discrete possibilities) in 2D space, a trimer $X_kY_lZ_m$ a specified location (from 8000) in 3D space, and so on. Various difficulties arise, however ('the curse of dimensionality'[14,15]) as the number of dimensions increases, even for quite small

numbers of dimensions or string length, since the dimensionality increases exponentially with the number of residues being changed. One in particular is the potential 'closeness' to each other of various randomly selected sequences, and how this effectively diverges extremely rapidly as their length is increased.

Imagine (as in ref. 16) that we have examples uniformly distributed in a p-dimensional hypercube, and wish to surround a target point with a hypercubical 'neighbourhood' to capture a fraction r of all the samples. The edge length of the (hyper)cube will be $e_p(r) = r^{(1/p)}$. In just 10 dimensions $e_{10}(0.01) = 0.63$ and $e_{10}(0.1) = 0.79$ while the range (of a unit hypercube) for each dimension is just 1. Thus to capture even just 1% or 10% of the observations we need to cover 63% or 80% of the range (*i.e.* values) of each individual dimension. Two consequences for any significant dimensionality are that even large numbers of samples cover the space only very sparsely indeed, and that most samples are actually close to the edge of the n-dimensional hypercube. We shall return later to the question of metrics for the effective distance between protein strings and for the effectiveness of protein catalysts; for the latter we shall assume (and discuss below) that the enzyme catalytic rate constant or turnover number (with units of s^{-1}, or in less favourable cases min^{-1}, h^{-1}, or d^{-1}) is a reasonable surrogate for most functional purposes.

Overall, it is genuinely difficult to grasp or to visualise the vastness of these search spaces,[17] and the manner in which even very large numbers of examples populate them only extremely sparsely. One way to visualise them[18–22] is to project them into two dimensions. Thus, if we consider just 30mers of nucleic acid sequences, and in which each position can be A, T, G or C, the number of possible variants is 4^{30}, which is $\sim 10^{18}$, and even if arrayed as 5 μm spots the array would occupy 29 km^2![23] The equivalent array for proteins would contain only 14mers, in that there are more than 10^{18} possible proteins containing the 20 natural amino acids when their length is just 14 amino acids.

The nature of sequence space

Sequence, structure and function

One of the fundamental issues in the biosciences is the elucidation of the relationship between a protein's primary sequence, its structure and its function. Difficulties arise because the relationship between a protein's sequence and structure is highly complex, as is the relationship between structure and function. Even single mutations at an individual residue can change a protein's activity completely – hence the discovery of 'inborn errors of metabolism'.[24,25] (The same is true in pharmaceutical drug discovery, with quite small changes in small molecule structure often leading to a dramatic change in activity – so-called 'activity cliffs'[26–33] – and with similar metaphors of structure–activity relationships, rather than those of sequence-activity, being equally explicit.[34–37]) Annotation of putative function from unknown sequences is largely based upon sequence homology (similarity) to proteins of known characterised function and particularly the presence of specific sequence/structure motifs (such as the Rossmann fold[38] or the P-loop motif[39]). While there have been great advances in predicting protein structure from primary sequence (see later), the prediction of function from structure (let alone sequence) remains an important (if largely unattained) aim.[40–54]

How much of sequence space is 'functional'?

The relationship between sequence and function is often considered in terms of a metaphor in which their evolution is seen as akin to

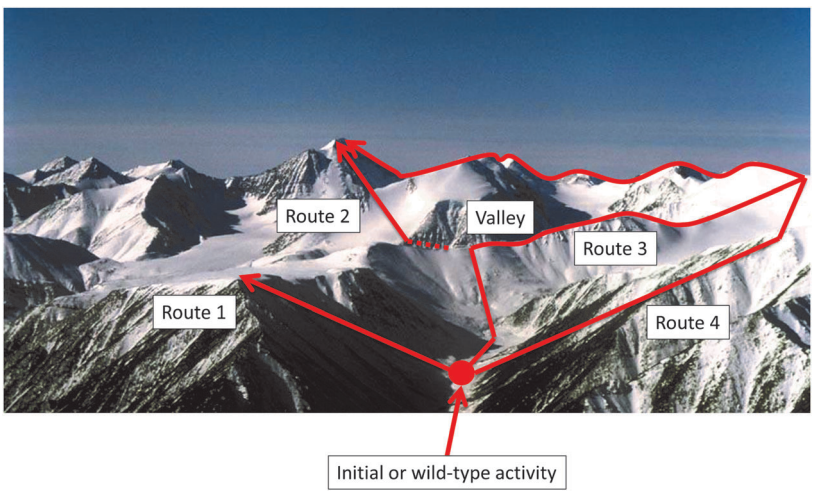


Fig. 5 A fitness landscape and its navigation. The initial or wild-type activity denotes the starting point (initialisation) for a directed evolution study (red circle). Accumulation of mutations that increase activity is represented by four routes to different positions in the landscape. Route 1 successfully increases activity through a series of additive mutations, but becomes stuck in a local optimum. Due to the nature of rugged fitness landscapes some of the shorter paths to a maximum possible (global optimum) fitness (activity) can require movement into troughs before navigating a new higher peak (route 2). Alternatively, one can arrive at the global optimum using longer but typically less steep routes without deep valleys (equivalent over flat ground to neutral mutations – routes 3 and 4).

traversing a 'landscape',[55] that may be visualised in the same way as one considers the topology of a natural landscape,[56,57] with the 'position' reflecting the sequence and the desirable function(s) or fitness reflected in the 'height' at that position in the landscape (Fig. 5).

Given the enormous numbers for populating sequence space, and the present impossibility of computing or sampling function from sequence alone, it is clear that natural evolution cannot possibly have sampled all possible sequences that might have biological function.[58] Hence, the strategy of a DE project faces the same questions as those faced in nature: how to navigate sequence space effectively while maintaining at least some function, but introducing sufficient variation that is required to improve that function. For DE there are also the practical considerations: how many variants can be screened (and/or selected for) and analysed with our current capabilities?

The first general point to be made is that most completely random proteins are practically non-functional.[10,56,59–66] Indeed, many are not even soluble,[67,68] although they may be evolved to become so.[69] Keefe and Szostak noted that *ca.* 1 in 10^{11} of random sequences have measurable ATP-binding affinity.[70] Consistent with this relative sparseness of functional protein space is the fact that even if one does have a starting structure-(/function), one typically need not go 'far' from such a structure to lose structure quite badly,[71] albeit that with a 'density' of only 1 in 10^{11} proteins being functional this implies that all such functional sequences are connected by trajectories involving changes in only a single amino acid[72] (and see ref. 58). This is also consistent with the fact that sequence space is vast, and only a tiny fraction of possible sequences tend to be useful and hence selected for by natural evolution. One may note[70,73] that at least some degree of randomness will be accompanied by some structure,[74,75] functionality or activity. For proteins, secondary structure is understood to be a strong evolutionary driver,[76] particularly through the binary-patterning (arrangement of hydrophilic/hydrophobic residues),[64,77–84] and so is the (somewhat related) packing density.[85–89] In a certain sense, proteins must at some point have begun their evolution as more or less random sequences.[90] Indeed ''Folded proteins occur frequently in libraries of random amino acid sequences'',[91] but quite small changes can have significantly negative effects.[92] Harms and Thornton give a very thoughtful account of evolutionary biochemistry,[4] recognizing that the ''physical architecture {of proteins both} facilitates and constrains their evolution''. This means that it will be hard (but not impossible), especially without plenty of empirical data,[93] to make predictions about the best trajectories. Fortunately, such data are now beginning to appear.[57,94] Indeed, the leitmotiv of this review is that understanding such (sequence-structure–activity) landscapes better will assist us considerably in navigating them.

What is evolving and for what purpose?

In a simplistic way, it is easy to assume that protein sequences are being selected for on the basis of their contribution to the host organism's fitness, without normally having any real knowledge of what is in fact being implied or selected for. However, a profound and interesting point has been made by Keiser *et al.*[95] to the effect that once a metabolite has been 'chosen' (selected) to be part of a metabolic or biochemical network, proteins are somewhat constrained to evolve as 'slaves', to learn to bind and react with the metabolites that exist. Thus, in evolution, the proteins follow the metabolites as much as *vice versa*, making knowledge of ligand binding[96,97] and affinity[98] to protein binding sites a matter of primary interest, especially if (as in the DE of biocatalysts) we wish to bind or evolve catalysts for novel (and xenobiotic) small molecule substrates. In DE we largely assume that the experimenter has determined what should be the objective function(s) or fitness(es), and we shall indicate the nature of some of the choices later; notwithstanding, several aspects of DE do tend to differ from those selected by natural evolution (Table 1). Thus, most mutations are pleiotropic *in vivo*,[99,100] for instance. As DNA sequencing becomes increasingly economical and of higher throughput[101,102] a greater provenance of sequence data enables a more thorough knowledge of the entire evolutionary landscape to be obtained. In the case of short sequences most[103] or all[104] of the entire genotype-fitness landscape may be measured experimentally. We note too (and see later) that there are equivalent issues in the optimization and algorithms of evolutionary computing (*e.g.* ref. 105–107), where strategies such as uniform cross-over,[108] with no real counterpart in natural or experimental evolution, have been shown to be very effective.

However, in the case of multi-objective optimisation (*e.g.* seeking to optimise two objectives such as both k_{cat} and thermostability, or activity *vs.* immunogenicity[109]), there is normally no individual preferred solution that is optimal for all objectives,[110] but a set of them, known as the Pareto front (Fig. 6), whose members are optimal in at least one objective while not being bettered (not 'dominated') in any other property by any other individual. The Pareto front is thus also known as the non-dominated front or 'set' of solutions. A variety of algorithms in multi-objective evolutionary optimisation (*e.g.* ref. 111–116) use members of the Pareto front as the choice of which 'parents' to use for mutation and recombination in subsequent rounds.

Protein folds and convergent and divergent evolution

What is certain, given that form follows function, is that natural evolution has selected repeatedly for particular kinds of secondary and tertiary structure 'domains' and 'folds'.[128,129] It is uncertain as to how many more are 'common' and are to be found *via* the methods of structural genomics,[130] but many have been expertly classified,[131] *e.g.* in the CATH,[132–134] SCOP[135–137] or InterPro[138,139] databases, and do occur repeatedly.

Given that structural conservation of protein folds can occur for sequences that differ markedly from each other, it is desirable that these analyses are done at the structural (rather than sequence) level (although there is a certain arbitrariness about where one fold ends and another begins[140,141]). Some folds have occurred and been selected *via* divergent evolution (similar sequences with different functions)[142] and some *via* convergent evolution (different sequences with similar functions).[143,144] This latter in particular makes the nonlinear mapping of sequence to

Table 1 Some features by which natural evolution, classical DE of biocatalysts, and directed evolution of biocatalysts using synthetic biology differ from each other. Population structures also differ in natural evolution *vs*. DE, but in the various strategies for DE they follow from the imposed selection in ways that are difficult to generalize

Feature	Natural evolution	Classical DE	DE with synthetic biology
Objective function and selection pressure	Unclear; there is only a weak relation of a protein's function with organismal fitness;[117] k_{cat} is not strongly selected for. Although presumably multi-objective, actual selection and fitness are 'composites'. If there is no redundancy, organisms must retain function during evolution.[58,118]	Typically strong selection weak mutation (rarely was sequencing done so selection was based on fitness only). Can select explicitly for multiple outputs (*e.g.* k_{cat}, thermostability).	Much as with classical DE, but diversity maintenance can be much enhanced *via* high-throughput methods of DNA synthesis and sequencing.
Mutation rates	Varies with genome size over orders of magnitude,[119] but typically (for organisms from bacteria to humans) $<10^{-8}$ per base per generation.[120,121] Can itself be selected for.[122]	Mutation rates are controlled but often limited to only a few residues per generation, *e.g.* to $1/L$ where L is the aa length of the protein; much more can lead to too many stop codons.	Library design schemes that permit stop codons only where required mean that mutation rates can be almost arbitrarily high.
Recombination rates	Very low in most organisms (though must have occurred in cases of 'horizontal gene transfer'); in some cases almost non-existent.[123]	Could be extremely high in the various schemes of DNA shuffling, including the creation of chimaeras from different parents.	Again it can be as high or low as desired; the experimenter has (statistically) full control.
Randomness of mutation	Although there are 'hot spots', mutations in natural evolution are considered to be random and not 'directed'.[124]	In error-prone PCR, mutations are seen as essentially random. Site-directed methods offer control over mutations at a small number of specified positions.	As much or as little randomness may be introduced as the experimenter desires by using defined mixtures of bases for each codon, *e.g.* NNN or NNK as alternatives to specific subsets such as polar or apolar.
Evolutionary 'memory'	For individuals (*cf.* populations[125]) there is no 'memory' as such, although the sequence reflects the evolutionary 'trace' (but not normally the pathway – *cf.* ref. 126 and 127).	Again, there is no real 'memory' in the absence of large-scale sequencing, but there is potential for it.[56]	With higher-throughput sequencing we can create an entire map of the landscape as sampled to date, to help guide the informed assessment of which sequences to try next.
Degree of epistasis	It exists, but only when there is a more or less neutral pathway joining the epistatic sites.	It is comparatively hard to detect at low mutation rates.	Potentially epistasis is much more obvious as sites can be mutated pairwise or in more complex designed patterns.
Maintenance of individuals of lower or similar fitness in population	They are soon selected out in a 'strong selection, weak mutation' regime; this limits jumps *via* lower fitness, and enforces at least neutral mutations.	It is in the hands of the experimenter, and usually not done when only fitnesses are measured.	Again it is entirely up to the experimenter; diversity may be maintained to trade exploration against exploitation.

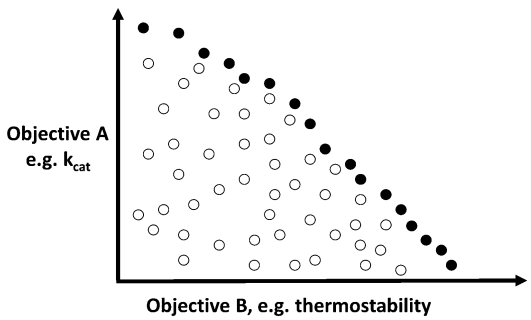


Fig. 6 A two-objective optimisation problem, illustrating the non-dominated or Pareto front. In this case we wish to maximise both objectives. Each individual symbol is a candidate solution (*i.e.* protein sequence), with the filled ones denoting an approximation to the Pareto front.

function extremely difficult, and there are roughly two unrelated sequences for each E.C. (Enzyme Commission classification) number.[145] As phrased by Ferrada and colleagues,[146] "two proteins with the same structure and/or function in our data. . .{have} a median amino acid divergence of no less than 55 percent". However, normally information is available only for extant molecules but not their history and precise evolutionary path (in contrast to DE). One conclusion might be that conventional means of phylogenetic analysis are not necessarily best placed to assist the processes of directed evolution, and we argue later (because a protein has no real 'memory' of its full evolutionary pathway) that modern methods of machine learning that can take into account ensembles of sequences and activities may prove more suitable. However, we shall first look at natural evolution.

Constraints on globular protein evolution structure in natural evolution

In gross terms, a major constraint on protein evolution is provided by thermodynamics, in that proteins will have a tendency to fold up to a state of minimum free energy.[147–149] Consequently, the composition of the amino acids has a major influence over protein folding because this means satisfying, so far as is possible, the preference of hydrophilic or polar amino acids to bind to each other and the equivalent tendency of hydrophobic residues to do so.[150–152] Alteration of residues, especially non-conservatively, often leads to a lowering of thermodynamic folding stability,[153] which may of course be compensated by changes in other locations. Naturally, at one level proteins need to have a certain stability to function, but

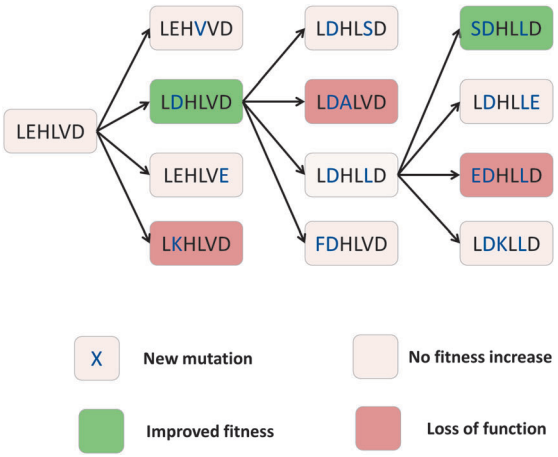


Fig. 7 Some evolutionary trajectories of a peptide sequence undergoing mutation. Mutations in the peptide sequence can cause an increase in fitness (*e.g.* enzyme activity, green), loss of fitness (salmon pink) or no change in fitness (grey). Typically, improved fitness mutations are selected for and subjected to further modification and selection. Neutral mutations keep sequences 'alive' in the series, and these can often be required for further improvements in fitness, as shown in steps 2 and 3 of this trajectory.

they also need to be flexible to effect catalysis. This is coupled to the idea that proteins are marginally stable objects in the face of evolution.[154–159] Overall, this is equivalent to 'evolution to the edge of chaos',[160,161] a phenomenon recognizing the importance of trading off robustness with evolvability that can also be applied[162,163] to biochemical networks.[164–170] Thermostability (see later) may also sometimes (but not always[171–173]) correlate with evolvability.[174,175]

Given the thermodynamic and biophysical[157,176,177] constraints, that are related to structural contacts, various models (*e.g.* ref. 147 and 178) have been used to predict the distribution of amino acids in known proteins. As regards to specific mechanisms, it has been stated that "solvent accessibility is the primary structural constraint on amino acid substitutions and mutation rates during protein evolution.",[148] while "satisfaction of hydrogen bonding potential influences the conservation of polar sidechains".[179] Overall, given the tendency in natural evolution for strong selection, it is recognized that a major role is played by neutral mutations[180–182] or neutral evolution[183–188] (see Fig. 5 and 7). Gene duplication provides another strategy, allowing redundancy followed by evolution to new functions.[189]

Coevolution of residues

Thus far, we have possibly implied that residues evolve (*i.e.* are selected for) independently, but that is not the case at all.[190–192] There can be a variety of reasons for the conservation of sequence (including correlations between 'distant' regions[193]), but the importance to structure and function, and functional linkage between them, underlie such correlations.[194–209] Covariation in natural evolution reflects the fact that, although not close in primary sequence, distal residues can be adjacent in the tertiary structure and may represent an interaction favourable to protein function. Covariation also provides an important computational approach to protein folding more generally (see below).

The nature, means of analysis and traversal of protein fitness landscapes

Since John Holland's brilliant and pioneering work in the 1970s (reprinted as ref. 210), it has been recognized that one can search large search spaces very effectively using algorithms that have a more or less close analogy to that of natural evolution. Such algorithms are typically known as genetic or evolutionary algorithms (*e.g.* ref. 106 and 211–213, and their implementation is referred to as evolutionary computing.[106,214–216] The algorithms can be classified according to whether one knows only the fitnesses (phenotypes) of the population or also the genotypes (sequences).[107]

Since we cannot review the very large literature, essentially amounting to that of the whole of molecular protein evolution, on the nature of (natural) protein landscapes, we shall therefore seek to concentrate on a few areas where an improved understanding of the nature of the landscape may reasonably be expected to help us traverse it. Importantly, even for single objectives or fitnesses, a number of important concepts of ruggedness, additivity, promiscuity and epistasis are inextricably intertwined; they become more so where multiple and often incommensurate objectives are considered.

Additivity. Additivity implies simple continuing fixing of improved mutations,[217–220] and follows from a model in which selection in natural evolution quite badly disfavours lower fitnesses,[221] a circumstance known from Gillespie[222,223] as 'strong selection, weak mutation' (SSWM, see also ref. 224–229). For small changes (close to neutral in a fitness or free energy sense), additivity may indeed be observed,[230,231] and has been exploited extensively in DE.[232–236] If additivity alone were true, however (and thus there is no epistasis for a given protein at all) then a rapid strategy for DE would be to synthesise all 20L amino acid variants at each position (of a starting protein of length L) and pick the best amino acid at each position. However, the very existence of convergent and divergent evolution implies that landscapes are rugged[237] (and hence epistatic), so at the very least additivity and epistasis must coexist.[236,238]

Epistasis. The term 'epistasis' in DE covers a concept in which the 'best' amino acid at a given position depends on the amino acid at one or more other positions. In fact, we believe that one should start with an assumption of rather strong epistasis,[238–248] as did Wright.[55] Indeed the rugged fitness landscape is itself a necessary reflection of epistasis and *vice versa*. Thus, epistasis may be both cryptic and pervasive,[249] the demonstrable coevolution goes hand in hand with epistasis, and "to understand evolution and selection in proteins, knowledge of coevolution and structural change must be integrated".[250]

Promiscuity. The concept of enzyme promiscuity mainly implies that some enzymes may bind, or catalyse reactions with, more than one substrate, and this is inextricably linked to how one can traverse evolutionary landscapes.[251–270] It clearly bears strongly on how we might seek to effect the directed evolution of biocatalysts.

NK landscapes as models for sequence-activity landscapes

A very important class of conceptual (and tunable) landscapes are the so-called NK landscapes devised by Kauffman[161,271] and

developed by many other workers (*e.g.* ref. 220, 221, 237 and 272–278). The 'ruggedness' of a given landscape is a slightly elusive concept,[279] but can be conceptualized[56,220] in a manner that implies that for a smooth landscape (like Mt Fuji[280,281]) fitness and distance tend to be correlated, while for a very 'rugged' landscape the correlation is much weaker (since as one moves away from a starting sequence one may pass through many peaks and troughs of fitness). In NK landscapes, K is the parameter that tunes the extent of ruggedness, and it is possible to seek landscapes whose ruggedness can be approximated by a particular value of K, since one of the attractions of NK is that they can reproduce (in a statistical sense) any kind of landscape.[282] Indeed, we can use the comparatively sparse data presently available to determine that experimental sequence-fitness landscapes reflect NK landscapes that are fairly considerably (but not pathologically) rugged,[23,57,104,241,251,274,276,283] and that there is likely to be one or more optimal mutation rates that themselves depend on the ruggedness (see later). Note too that the landscapes for individual proteins, as discussed here, are necessarily more rugged than are those of pathways or organisms, due to the more profound structural constraints in the former.[57,157] (Parenthetically, NK-type landscapes and the evolutionary metaphor have also proved useful in a variety of other 'complex' spheres, such as business, innovation and economics (*e.g.* ref. 278 and 284–295, though a disattraction of NK landscapes in evolutionary biology itself is that they do not obey evolutionary rules.[224])

Experimental directed protein evolution

A number of excellent books and review articles have been devoted to DE, and a sampling with a focus on biocatalysis includes.[296–334] As indicated above, DE begins with a population that we hope contains at least one member that displays some kind of activity of interest, and progresses through multiple rounds of mutation, selection and analysis (as per the steps in Fig. 4).

Initialisation; the first generation

During the preliminary design of a DE project the main objective and required fitness criteria must be defined and these criteria influence the experimental design and screening strategy.

We consider in this review that a typical scenario is that one has a particular substrate or substrate class in mind, as well as the chemical reaction type (oxidation, hydroxylation, amination and so on) that one wishes to catalyse. If any activity at all can be detected then this can be a starting point. In some cases one does not know where to start at all because there are no proteins known either to catalyse a relevant reaction or to bind the substrate of interest. For pharmaceutical intermediates, it can still be useful to look for reactions involving metabolites, as most drugs do bear significant structural similarities to known metabolites,[335,336] and it is possible to look for reactions involving the latter. A very useful starting point may be the structure-function linkage database http://sfld.rbvi.ucsf.edu/django/.[337] There are also 'hub' sequences that can provide useful starting points,[338] while Verma,[330] Nov[339] and Zaugg[340] list various computational approaches. If one has a structure in the form of a PDB file one can try HotSpotWizard http://loschmidt.chemi.muni.cz/hotspotwizard/.[341] Analysing the diversity of known enzyme sequences is also a very sensible strategy.[342,343] Nowadays, an increasing trend is to seek relevant diversity, aligned using tools such as Clustal Omega,[344,345] MUSCLE,[346] PROMALS,[347,348] or other methods based on polypharmacology,[141,349,350] that one may hope contains enzymes capable of effecting the desired reaction. Another strategy is to select DNA from environments that have been exposed to the substrate of interest, using the methods of functional metagenomics.[351,352] More commonly, however, one does have a very poor protein (clone) with at least some measurable activity, and the aim is to evolve this into a much more active variant.

In general, scientific advance is seen in a Popperian view (see *e.g.* ref. 353–357) as an iterative series of 'conjectures' and 'refutations' by which the search for scientific truth is 'narrowed' by finding what is not true (may be falsified) *via* predictions based on hypothetico-deductive reasoning and their anticipated and experimental outcomes. However, Popper was purposely coy about where hypotheses actually came from, and we prefer a variant[358–362] (see also ref. 363 and 364) that recognises the equal contribution of a more empirical 'data-driven' arc to the 'cycle of knowledge' (Fig. 8).

In a similar vein, many commentators (*e.g.* ref. 365–368) consider the best strategy for both the starting population and the subsequent steps to be a judicious blend between the more empirical approaches of (semi-)directed evolution and strategies more formally based on attempts to design[369] (somewhat in the absence of fully established principles) sequences or structures based on what is known of molecular interactions. We concur with this, since at the present time it is simply not possible to design enzymes with high activities *de novo* (from scratch, or from sequence alone), despite progress in simple 4-helix-bundle and related 'maquettes'.[370–373] David Baker, probably the leading

The cycle of knowledge in directed evolution

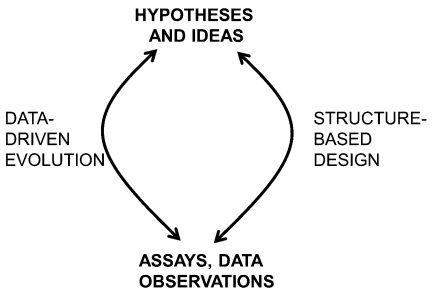


Fig. 8 The 'cycle of knowledge' in modern directed evolution. Both structure-based design and a more empirical data-driven approach can contribute to the evolution of a protein with improved properties, in a series of iterative cycles.

expert in protein design, considers that design is still incapable of predicting active enzymes even when the chemistry and active sites appear good.[374,375] Several reviews attest to this,[329,376–379] but crowdsourcing approaches have been shown to help,[380] and computational design (and see below) certainly beats random sequences.[381] Overall, the fairest comment is probably that we can benefit from design for binding, specificity and active site modelling, but that for improving k_{cat} we need the more empirical methods of DE, especially (see below) of residues distant from the active site.

Scaffolds

Because natural evolution has selected for a variety of motifs that have been shown in general terms to admit a wide range of possible enzyme activities, a number of approaches have exploited these motifs or 'scaffolds'.[382] Triose phosphate isomerase (TIM) has proved a popular enzyme since the pioneering work of Albery and Knowles[383] and more recent work on TIM energetics,[384] and TIM $(\beta\alpha)_8$ barrels can be found in 5 of 6 EC classes.[146] TIM and many (but not all) such natural enzymes are most active as dimers,[385,386] caused by a tight interaction of 32 residues of each subunit in the wild type, though functional monomers can be created.[387,388]

Thus, $(\beta\alpha)_8$ barrel enzymes[389–402] have proven particularly attractive as scaffolds for DE.[403–407] Some use or need cofactors like PLP, FMN, *etc.*,[392,408] and their folding mechanisms are to some degree known.[385,409–411] We note, however, that virtual screening of substrates against these[412] has shown a relative lack of effectiveness of consensus design because of the importance of correlations (*i.e.* epistasis).[386]

α/β and $(\alpha/\beta)_2$ barrels have also been favoured as scaffolds,[395,413–417] while attempts at automated scaffold selection can also be found.[374,418,419] A very interesting suggestion[420] is that the polarity of a fold may determine its evolvability.

Although not focused on biocatalysis, other scaffolds such as lipocalins[421–426] and affibodies[427–437] have proved useful for combinatorial biosynthesis and directed evolution.

Computational protein design

While computational protein design completely from scratch (*in silico*) is not presently seen as reasonable, probably (as we stress later) because we cannot yet use it to predict dynamics effectively, significant progress continues to be made in a number of areas,[373,438–456] including 'fold to function',[457] combinatorial design,[458] and a maximum likelihood framework for protein design.[459] Notable examples include a metalloenzyme for organophosphate hydrolysis,[460,461] aldolase[462,463] and others.[464–468] Theozymes[469–472] (theoretical catalysts, constructed by computing the optimal geometry for transition state stabilization by model functional) groups represent another approach.

Arguably the most advanced strategies for protein design and manipulation *in silico* are Rosetta[374,473–483] and Rosetta-Backrub,[484,485] while more 'bottom-up' approaches, based on some of the ideas of synthetic biology, are beginning to appear.[486–492] It is an easy prediction that developments in synthetic biology will have highly beneficial effects on *de novo* design, and *vice versa*.

Docking

If one is to find an enzyme that catalyses a reaction, one might hope to be able to predict that it can at least bind that substrate using the methods of *in silico* docking.[493] To date, methods based on Autodock,[494–499] APoC,[500] Glide[501–503] or other programs[504–511] have been proposed, but this strategy is not yet considered mainstream for the DE of a first generation of biocatalysts (and indeed is subject to considerable uncertainty[512]). Our experience is that one must have considerable knowledge of the approximate answer (the binding site or pocket) before one tries these methods for DE of a biocatalyst.

Having chosen a member (or a population) as a starting point, the next step in any DE program is the important one of diversity creation. Indeed, the means of creating and exploiting suitable libraries that focus on appropriate parts of the protein landscape lies at the heart of any intelligent search method.[513]

Diversity creation and library design

A diversity of sequences can be created in many ways,[514] but mutation or recombination methods are most commonly used in DE. Some are purely empirical and statistical (*e.g.* N mutations per sequence), while others are more focused to a specific part of the sequence (Fig. 9). Strategies may also be discriminated in terms of the degree of randomness of the changes and their extensiveness (Fig. 10). Two useful reviews include[515] and,[516] while others[334,517–519] cover computational approaches. A DE library creation bibliography is maintained at http://openwetware.org/wiki/Reviews:Directed_evolution/Library_construction/bibliography.

Effect of mutation rates, implying that higher can be better

In classical evolutionary computing, the recognition that most mutations were or are deleterious meant that mutation rates were kept low. If only one in 10^3 sequences is an improvement when the mutation rate is $1/L$ per position (L being the length of the string), then (in the absence of epistasis) only 1 in 10^6 is at $2/L$. (Of course $1/L$ is far greater than the mutation rates common in natural evolution, which scales inversely with genome size,[119] may depend on cell–cell interactions,[520] and is normally below 10^{-8} per base per generation for organisms from bacteria to humans.[119–121]) This logic is persuasive but limited, since it takes into account only the frequency but not the quality of the improvement (and as mentioned essentially does not consider epistasis). Indeed there is evidence that higher mutation rates are favoured both *in silico*[220,521–524] and experimentally.[525–528] This is especially the case for directed mutagenesis methods (especially those of synthetic biology), where stop codons can be avoided completely. We first discuss the more classical methods.

Random mutagenesis methods

Error-prone PCR (epPCR) is probably the most commonly used method for introducing random mutations. PCR amplification using *Taq* polymerase is performed under suboptimal conditions by altering the components of the reaction (in particular polymerase concentration, $MgCl_2$ and dNTP concentration, or

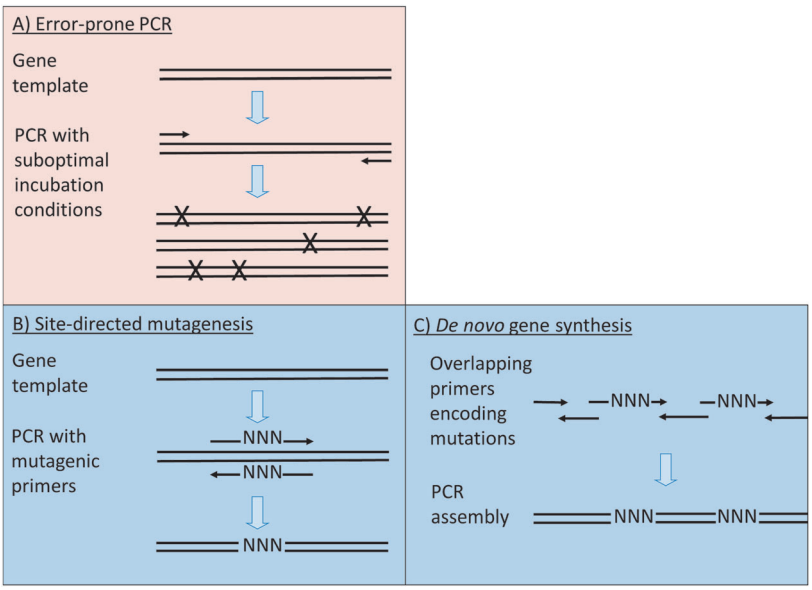


Fig. 9 Overview of the different mutagenesis strategies commonly employed to create variant protein libraries. Random methods (pink background) can create the greatest diversity of sequences in an uncontrolled manner. Mutations during error-prone PCR (A) are typically introduced by a polymerase amplifying sequences imperfectly (by being used under non-optimal conditions). In contrast, directed mutagenesis methods (blue background) introduce mutations at defined positions and with a controlled outcome. Site-directed mutagenesis (B) introduces a mutation, encoded by oligonucleotides, onto a template gene sequence in a plasmid. However, gene synthesis (C) can encode mutations on the oligonucleotides used to synthesise the sequence *de novo*, hence multiple mutations can be introduced simultaneously. X = random mutation, N = controlled mutation. → = PCR primer.

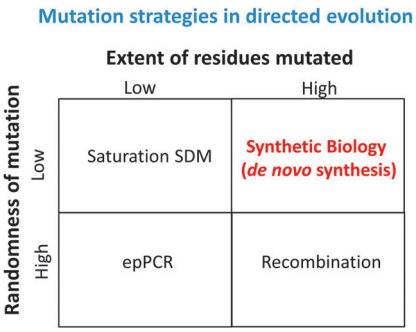


Fig. 10 A Boston matrix of the different strategies for variant libraries. Methods are identified in terms of the randomness of the mutations they create and the number of residues that can be targeted.

supplementation with $MnCl_2$ (ref. 529)) or cycling conditions (increased extension times).[530] Although epPCR is the simplest to implement and most commonly used method for library creation, it is limited by its failure to access all possible amino acid changes with just one mutation,[339,531–533] a strong bias towards transition mutations (AT to GC mutations),[531] and an aversion to consecutive nucleotide mutations.[532,534]

Refinement of these methods has allowed greater control over the mutation bias, rate of mutations[530,535–537] and the development of alternative methodologies like Mutagenic Plasmid Amplification,[538] replication,[539] error-prone rolling circle[540] and indel[541–543] mutagenesis. Typically, for reasons indicated above, the epPCR mutation rate is tuned to produce a small number of mutations per gene copy (although orthogonal replication *in vivo* may improve this[544]), since entirely random epPCR produces multiple stop codons (3 in every 64 mutations) and a large proportion of non-functional, truncated or insoluble proteins.[545] The library size also dictates that a large number of mutants must be screened to test for all possibilities, which may also be impractical depending on the screening strategy available. While random methods for library design can be successful, intelligent searching of the sequence space, as per the title of this review, does not include *purely* random methods.[546] In particular, these methods do not allow information about which parts of the sequence have been mutated or whether all possible mutations for a particular region of interest have been screened.

Site-directed mutagenesis to target specific residues

Since the combinatorial explosion means that one cannot try every amino acid at every residue, one obvious approach is to restrict the number of target residues (in the following sections we will discuss why we do not think this is the best strategy for making faster biocatalysts). Indeed, mutagenesis directed at specific residues, usually referred to as site-directed mutagenesis,[547,548] dates from the origins of modern protein engineering itself.[549]

In site-directed mutagenesis, an oligonucleotide encoding the desired mutation is designed with flanking sequences either side that are complementary to the target sequence and these direct its binding to the desired sequence on a template. This oligomer is used as a PCR primer to amplify the template sequence, hence all amplicons encode the desired mutation. This control over the mutation enables particular types of mutation to be made by using mixed base codons, *i.e.*

Degenerate codon	Mixed base sequence	Encoded codons	Stop codons	Encoded amino acids	Properties
NNN	(A,T,G,C) (A,T,G,C) (A,T,G,C)	64	TAA, TAG, TGA	All	Fully randomised codon
NNK	(A,T,G,C) (A,T,G,C) (G,T)	32	TAG*	All	All 20 amino acids
NNS	(A,T,G,C) (A,T,G,C) (G,C)	32	TAG**	All	All 20 amino acids
NDT	(A,T,G,C) (A,T,G) T	12	No	Phe, Leu, Ile, Val, Tyr, His, Asn, Asp, Cys, Arg, Ser, Gly	Mixture of polar, nonpolar, positive and negative charge (Reetz 2008)
NTN	(A,T,G,C) T (A,T,G,C)	16	No	Met, Phe, Leu, Ile, Val	Nonpolar residues
NAN	(A,T,G,C) A (A,T,G,C)	16	TAA, TAG	Tyr, His, Gln, Asn, Lys, Asp, Glu	Charged, larger side chains
NCN	(A,T,G,C) C (A,T,G,C)	16	No	Ser, Pro, Thr, Ala	Smaller side chains, polar and nonpolar residues
RST	(A,G) (G,C) T	4	No	Ala, Gly, Ser, Thr	Small side chains

Fig. 11 Examples of some of the common degenerate codons used in DE studies. A codon containing specific mixed bases is used to encode a particular set of amino acids, ranging from all twenty amino acids (NNN or NNK) to those with particular properties. Hence, choice of degenerate codons to use depends on the design and objective of the study. In the IUPAC terminology[590] K = G/T, M = A/C, R = A/G, S = C/G, W = A/T, Y = C/T, B = C/G/T, D = A/G/T, H = A/C/T, V = A/C/G, N = A/C/G/T. (*Typically with low codon usage; suppressor mutation may be used to block it. **Typically with low codon usage, especially in yeast; suppressor mutation may be used to block it).

codons that contain a mixture of bases at a specified position (*e.g.* N denotes an equal mixture of A, T, G or C at a single position). Fig. 11 shows a compilation of the more common types of mixed codons used. These range from those capable of encoding all 20 amino acids (*e.g.* NNK) to a small subset of residues with a particular physicochemical property (*e.g.* NTN for nonpolar residues only).

The most common method (QuikChange and derivatives thereof) uses mutagenic oligonucleotides complementary to both strands of a target sequence, which are used as primers for a PCR amplification of the plasmid encoding the gene. Following DpnI digestion of the template, the PCR product is transformed into *E. coli* and the nicked plasmid is repaired *in vivo*.[550,551] Despite its popularity, QuikChange is somewhat limited by aspects like primer design and efficiency, and a variety of derivatives have been published that improve upon the original method.[552,553]

Given that site-directed mutagenesis provides a way of mutating a small number of residues with high levels of accuracy, several approaches have been developed to identify possible positions to target to increase the hit rate and success. Combinatorial alanine scanning[554,555] is well known, while other flavours include the Mutagenesis Assistant Program,[531,556] and the semi-rational CASTing and B-FIT approaches[323,557] that employ a Mutagenic Plasmid Amplification method.[558]

In addition to these more conventional methods, new approaches are continually being developed to improve efficiency and to reduce the number of steps in the workflow, for example Mutagenic Oligonucleotide-Directed PCR Amplification (MOD-PCR),[559] Overlap Extension PCR (OE-PCR),[560–564] Sequence Saturation Mutagenesis,[565–571] Iterative Saturation Mutagenesis,[557,572–579] and a variety of transposon-based methods.[580–583] However, a common issue with site-directed mutagenesis methods is the large number of steps involved and the limited number of positions that can be efficiently targeted at a time. The ability to mutate residues in multiple positions in a sequence is of particular interest as this can be used to address the question of combinatorial mutations simultaneously. Hence, methods like those by Liu *et al.*,[584] Seyfang *et al.*,[585] Fushan *et al.*[586] and Kegler-Ebo *et al.*[587] are important developments in mutagenesis strategies. Rational approaches have been reviewed,[588] including from the perspective of the necessary library size.[589] As a result, there is significant interest in the development of novel methodologies that can address these issues to produce accurate variant libraries, with larger numbers of simultaneous mutations in an economical workflow.

Optimising nucleotide substitutions

Following the selection of residues to target for mutation an important choice is the type of mutation to create. This choice is not obvious but determines the type of mutations that are made and the level of screening required. The experimenter needs to consider the nature of the mutations that they want to introduce for each position and this relates to the objective of the study. Using the common mixed base IUPAC terminology[590] (Fig. 11) there are a large number of codons that can be chosen, ranging from those encoding all 20 amino acids (the NNK or

NNS codons), to a particular characteristic (*e.g.* NTN encodes just nonpolar residues[64]) and a limited number of defined residues (GAN encoding just aspartate or glutamate). Importantly, choosing to use these specified mixed base codons in mutagenesis can reduce the possibility of premature stop codons and increase the chance of creating functional variants. For example, if a wild-type sequence encodes a nonpolar residue at a particular position then the number of functional variants is likely to be higher if the nonpolar codon NTN is used, encoding what are conserved substitutions, compared to encoding all possible residues with the NNK codon.[591,592]

Indeed, it is known to be better to search a large library sparsely than a small library thoroughly.[593] Thus, a general strategy that seeks to move the trade-off between numbers of changes and numbers of clones to be assessed recognizes that one can design libraries that cover different general amino acid properties (such as charged, hydrophobic) while not encoding all 20 amino acids, thereby reducing (somewhat) the size of the search space. These are known as reduced library designs (see Fig. 11).

Reduced library designs

One limitation with the use of single degenerate codons is that for some sequences not all amino acids are equally represented and sometimes rare codons or stop codons are encoded. To circumvent this issue ''small-intelligent'' or ''smart'' libraries have been developed to provide equal frequency of each amino acid without bias.[594] Using a mixture of oligonucleotides, Kille *et al.*[595] created a restricted library with three codons NDT, VHG and TGG that encode 12, 9 and 1 codon, respectively. Together these encode 22 codons for all 20 amino acids in equal frequency, which provides good coverage of possible mutations but reduces the screening effort required to cover the sequence space. Alternative methods with the same objective include the MAX randomisation strategy[596] and using ratios of different degenerate codons designed by software (DC-Analyzer[597]). Alternatively, the use of a reduced amino acid alphabet can also search a relevant sequence space whilst reducing the screening effort further. For example, the NDT codon encodes 12 amino acids of different physicochemical properties without encoding stop codons and has been shown to increase the number of positive hits (*versus* full randomization) in directed evolution studies.[324] Overall, a considerable number of such strategies have been used (*e.g.* ref. 64, 67, 68, 81, 82, 324, 513, 556, 592 and 596–603).

The opposite strategy to reduced library designs is to increase them by modifying the genetic code. While one may think that there is enough potential in the very large search spaces using just 20 amino acids, such approaches have led to some exceptionally elegant work that bears description.

Non-canonical amino acid incorporation

If the existing protein synthetic machinery of the host cell is able to recognise a novel amino acid, it is possible to take an auxotroph and add the non-canonical amino acid (NCAA)[604] that is thereby incorporated non-selectively. If one wishes to have site specificity, there are two main ways to increase the number of amino acids that can be incorporated into proteins.[605] First, the specificity of a tRNA molecule (*e.g.* one encoding a stop codon) can be modified to accommodate non-canonical amino acids; in this way, the use of the relevant codon can introduce an NCAA at the specified position.[606,607] Using this method, eight NCAAs were incorporated into the active site of nitroreductase (NTR, at Phe124) and screened for activity. One Phe analogue, *p*-nitrophenylalanine (*p*NF), exhibited more than a two-fold increase in activity over the best mutant containing a natural amino acid substitution (P124K), showing that NCAAs can produce higher enzyme activity than is possible with natural amino acids.[608]

The other, considerably more radical and potentially ground-breaking, is effectively to evolve the genetic code and other apparatus such that instead of recognising triplets a subset of mRNAs and the relevant translational machinery can recognise and decode quadruplets.[609–619] To date, some 100 such NCAAs have been incorporated. However, the incorporation of NCAAs can often impact negatively on protein folding and thermostability, an issue that can be addressed through further rounds of directed evolution.[620]

Recombination

In contrast to the mutagenesis methods of library creation outlined above, but entirely consistent with our knowledge from strategies used in evolutionary computing (*e.g.* ref. 106), recombination is an alternative (or complementary) and effective strategy for DE (Fig. 12). Recombination techniques offer several advantages that reflect aspects of natural evolution that differ from random mutagenesis methods, not least because such changes can be combinatorial and hence able to search more areas of the sequence space in a given experiment. Recombination for the purposes of DE was popularized by Stemmer and his colleagues under the term 'DNA shuffling'.[621–625] This used a pool of parental genes with point mutations that were randomly fragmented by DNAseI and then reassembled using OE-PCR. Since then, a variety of further methods have been developed using different fragmentation and assembly protocols.[626–629] Parental genes for DNA shuffling can be generated by random mutagenesis (epPCR) or from homologous gene families; such chimaeras may be particularly effective.[630–633]

Despite its advantages for searching wider sequence space, however, such recombination does not yield chimaeric proteins with balanced mutation distribution. Bias occurs in crossover regions of high sequence identity because the assembly of these sequences is more favourable during OE-PCR.[634,635] As a result, this reduces the diversity of sequences in the variant library. Alternative methods like SCRATCHY[636,637] generate chimaeras from genes of low sequence homology and so may help to reduce the extent of bias at the crossover points.

In addition to these more traditional methods of DNA shuffling, a number of variations have been developed (often with a penchant for a quirky acronym), such as Iterative Truncation for the Creation of HYbrid enzymes (ITCHY[638,639]), RAndom CHImeragenesis on Transient Templates (RACHITT),[640] Recombined Extension on Truncated Templates (RETT),[641]

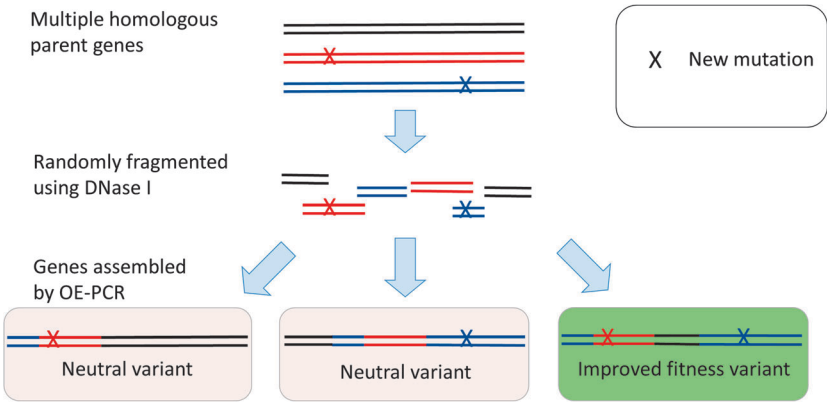


Fig. 12 The traditional recombination method for diversity creation. Recombination requires a sample of different variants of a gene (parents), which can be derived from a family of homologous genes or generated by random mutagenesis methods. The random fragmentation of these genes (using DNase I or other method) cleaves them into small constituent parts. Importantly, as the parental genes are all homologous, the fragments overlap in sequence thus allowing them to be reassembled by overlap extension PCR (OE-PCR) producing products that encode a random mixture of the parental genes. A key advantage of recombination methods is the improved ability to create combinatorial mutations. This is illustrated using two mutations (present in two different parental sequences) that when recombined separately produce no fitness improvement, but when combined together produce a variant with improved fitness.

One-pot Simple methodology for CAssette Randomization and Recombination) OSCARR,[642,643] DNA shuffling Frame shuffling,[644] Synthetic shuffling,[645] Degenerate Oligonucleotide Gene Shuffling (DOGS),[646] USERec,[647,648] SCOPE[649–651] and Incorporation of Synthetic Oligos duRing gene shuffling (ISOR).[652,653] Other methods of recombination that have been used for the improvement of proteins include the Protamax approach,[654] DNA assembler,[655,656] homologous recombination *in vitro*[657] and Recombineering (*e.g.* ref. 658 and 659). Circular permutation, in which the beginning and end of a protein are effectively recombined in different places, provides a (perhaps surprisingly) effective strategy.[17,660–667]

There has long been a recognition that the better kind of chimaeragenesis strategies are those that maintain major structural elements,[668,669] by ensuring that crossover occurs mainly or solely in what are seen as structurally the most 'suitable' locations. This is the basis of the OPTCOMB,[670,671] RASPP,[672] SCHEMA (*e.g.* ref. 673–683) and other types of approach.[109,684–691]

Thus, in the directed evolution of a cytochrome P450, Otey *et al.*[674] utilized the SCHEMA algorithm to approximate the effect of recombination with different parent P450s on the protein structure. SCHEMA provided a prediction of preferred positions for crossovers, which enabled the creation of a mutant with a 40-fold higher peroxidase activity.[673,678] Similarly, the recombination of stabilizing fragments was also able to increase the thermostability of P450s using the same approach.[692]

Cell-free synthesis

Although the majority of the mutations and recombinations described above have been performed *in vitro*, the actual expression of the proteins themselves, and the analysis of their functionalities, is usually done *in vivo*. However, we should mention a series of purely *in vitro* strategies that have also been used to identify good sequences when coupled to suitable *in vitro* translation systems with functional assays.[693–700]

Synthetic biology for directed evolution

With the recent improvements in DNA synthesis technology and reducing costs it is becoming increasingly feasible to synthesise sequences on a large scale. The most widely used methods for DNA synthesis continue to be short single-stranded oligodeoxyribonucleotides (typically 10–100 nt in length, often abbreviated to oligonucleotides or oligos) using phosphoramidite chemistry,[701,702] although syntheses from microarrays have particular promise.[546,703–708] Following synthesis, these oligonucleotides are assembled into larger constructs using enzymatic methods.

Hence, the foundation of synthetic biology is based on the ability to design and assemble novel biological systems 'from the ground up', *i.e.* synthetically at the DNA level.[709–713] As a result, gene synthesis and genome assembly methods have been developed to create novel sequences of several kilobases in length.[714] In particular, Gibson *et al.* recently assembled sections of the *Mycoplasma genitalium* genome (each 136 to 166 kb) using overlapping synthesised oligonucleotides.[5,6]

These developments in DNA synthesis technology (and lowered cost) can greatly benefit directed evolution studies. In particular, gene synthesis using overlapping oligonucleotides presents a particularly promising method for introducing controlled mutations into a gene sequence. As these methods assemble the gene *de novo*, multiple mutations at different positions in the gene can be introduced simultaneously in a single workflow, decreasing the need for iterative rounds of mutagenesis.

In this process, oligonucleotide sequences are designed to be overlapping and span the length of the gene of interest,

following synthesis they are assembled by either PCR-based[715,716] or ligation-based[717–720] methods. Variant libraries can be created using this process by encoding mixed base codons on the oligonucleotides and at multiple positions if required.[721] However, a limitation of the conventional gene synthesis procedure is the inherent error rate (primarily single base inserts or deletions),[722,723] which arises from errors in the phosphoramidite synthesis of the oligonucleotides. As a result, clones encoding the desired sequence must be verified by DNA sequencing and an error-correction procedure is often required. Several error-correction methods are used, including site-directed mutagenesis,[724] mismatch binding proteins[725] and mismatch cleaving endonucleases.[726,727] Of these, mismatch endonucleases are the most commonly used, and they are amenable to high throughput and automation.

SpeedyGenes and GeneGenie: tools for synthetic biology applied to the directed evolution of biocatalysts

Mismatch endonucleases recognise and cleave heteroduplexes in a DNA sequence. Consequently, they can be used as an effective method for the removal of errors during gene synthesis. However, when using mixed-base codons in directed evolution this is problematic, as these mixed sequences will form heteroduplexes and so will be heavily cleaved, thus preventing assembly of the required full-length sequence. Hence, we have developed an improved gene synthesis method, SpeedyGenes, which both improves the accurate synthesis of larger genes and can also accommodate mixed-base codon sequences.[728] SpeedyGenes integrates a mismatch endonuclease step to cleave mismatched bases and, anticipating complete digestion of the mixed-base sequences, then restores these mixed base sequences by reintroducing the oligonucleotides encoding the mutation back into the PCR ("spiking in") to allow the full length, error corrected gene to be synthesised. Importantly, multiple variant codons can be encoded at different positions of the gene simultaneously, enabling greater search of the sequence space through combinatorial mutations. This was illustrated[728] by the synthesis of a monoamine oxidase (MAO-N) with three contiguous mixed-base codons mutated at two different positions in the gene. The known structure of MAO-N showed that the side chains of these residues were known to interact, hence these libraries could be screened for combinatorial coevolutionary mutations.

As with most synthetic biology methods, the use of sequence design *in silico* is crucial to the successful synthesis *in vitro*. In the case of SpeedyGenes, a parallel, online software design tool, GeneGenie, was developed to automate the design of DNA sequences and the desired variant library.[729] By calculating the melting temperature (T_m) of the overlapping sequences, and minimising the potential mis-annealing of oligomers, GeneGenie greatly improves the success rate of assembly by PCR *in vitro*. In addition, codons are selected according to the codon usage of the expression host organism, and cloning sequences can be encoded *ab initio* to facilitate downstream cloning. Importantly, any mixed base codon can be added to incorporate into the designed sequence, hence automating the design of the variant library. As an example, a limited library of enhanced green fluorescent protein (EGFP) were designed to encode two variant codons (YAT at Y66 and TWT at Y145), the product of which would encode a limited variant library of green and blue variants of EGFP[728] (Fig. 13).

Genetic selection and screening

An important aspect of any experiment exploiting directed evolution for the development of improved biocatalysts is how one determines which of the many millions (or more) of the different clones that are created is worth testing further and/or retaining for subsequent generations. If it is possible to include a (genetic) selection step prior to any screening, this is always likely to prove valuable.[303,730–732]

Genetic selection

Most strategies for selection are unique to the protein of interest, and hence need to be designed empirically. Generally, this entails selection of a clone containing a desirable protein because it leads the cell to have a higher fitness.[599,733] Examples including those based on enantioselectivity,[734,735] substrate utilisation,[736] chemical complementation,[737,738] riboswitches,[739–743] and counter-selection[744] can be given. An ideal is when the selection rescues cells from a toxic insult that would otherwise kill them[745] (see Fig. 14) or repairs a growth defect[746–748]). Two such examples[749,750] of genetic selection are based on transporter engineering. However, most of the time it is quite difficult to develop such a genetic selection assay, so one must resort to screening.

Screening

Microtitre plates are the standard in biomolecular screening, and this is no different in DE.[751] Herein, clones are seeded such that one clone per well is cultured, the substrates added, and the activity or products screened, primarily using chromogenic or fluorogenic substrates. This said, flow cytometry and fluorescence-activated cell sorting (FACS) have the benefit of much higher throughputs and have been widely applied (*e.g.* ref. 415 and 752–790) (and see below for microchannels and picodroplets). 2D arrays using immobilized proteins may also be used.[791,792] However, not all products of interest are fluorescent, and these therefore need alternative methods of detection.

Thus, other techniques have included Raman spectroscopy for the chemical imaging of productive clones,[793,794] while IR spectroscopy has been used to assess secondary structure (*i.e.* folding).[795] Various constructs have been used to make non-fluorescent substrates or product produce a fluorescence signal.[796] These include substrate-induced gene expression screening[797–799] and product-induced gene expression,[800] fluorescent RNAs,[801] reporter bacteria,[773,802] the detection of metabolites by fluorogenic RNA aptamers,[803–811] colourimetric aptamers and Au particles,[812] or appropriate transcription factors.[787] Riboswitches that respond to product formation,[742,743] chemical tags,[813,814] and chemical proteomics[815] have also been used as reporters for the production of small molecules.

Solid-phase screening with digital imaging is another alternative used for the engineering of biocatalysts. These methods generally

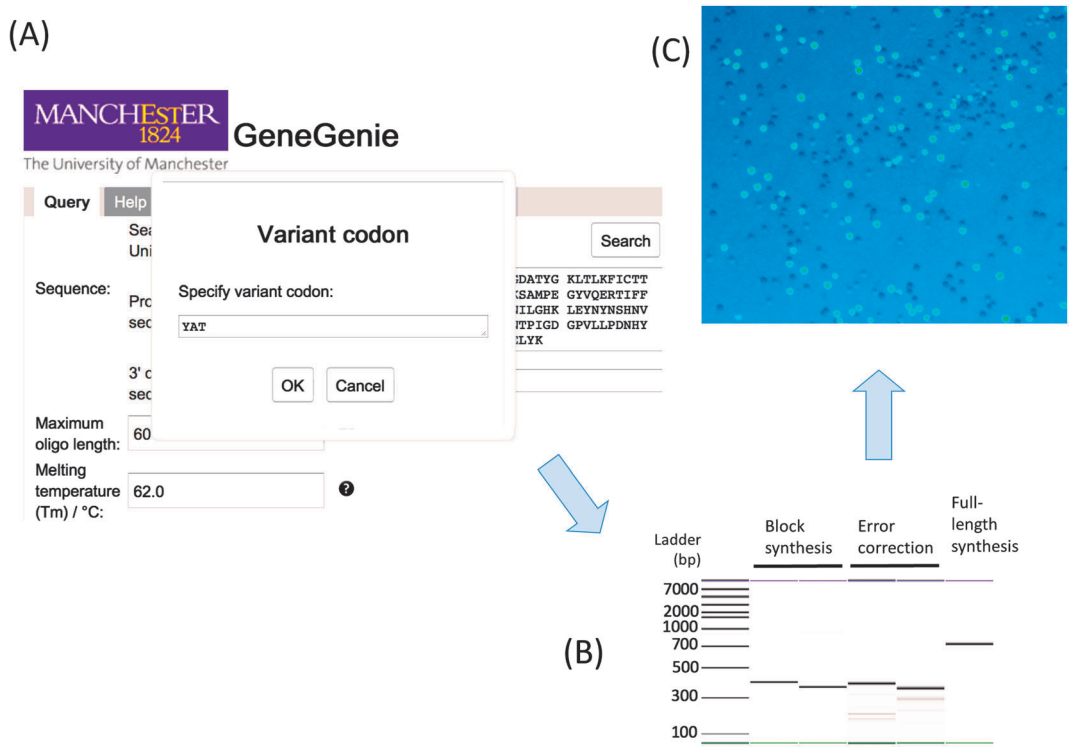


Fig. 13 GeneGenie and SpeedyGenes: synthetic biology tools for the purposes of directed evolution. The integration of computational design and accurate gene synthesis methodology provide a strong platform that can be utilised for directed evolution. As an example, the design, synthesis and screening of a small library of EGFP variants is shown. Mixed base codons are used to encode the green and blue variants of EGFP in a single library. (A) GeneGenie (www.gene-genie.org/) designs overlapping oligonucleotides for a given protein together with any specific mixed base codon (here YAT denoting C/T,A,T). (B) SpeedyGenes assembles the gene sequence using these oligonucleotides, accurately (using error correction) producing variant libraries with the desired mutations. (C) Direct expression (no pre-selection) of the library in *E. coli* yielded colonies with the desired mutations (green or blue fluorescence).

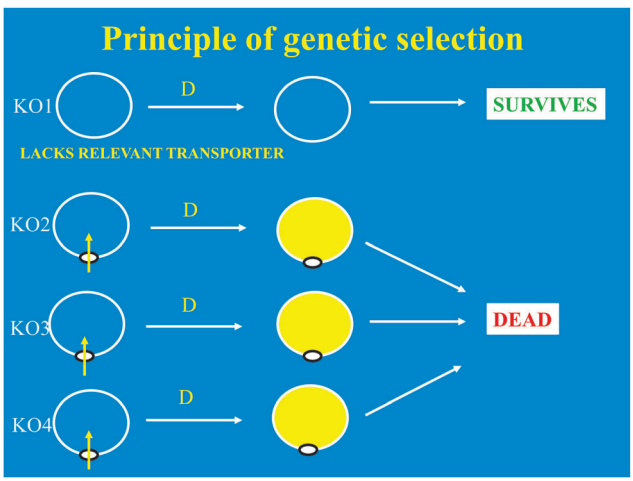


Fig. 14 The principle of genetic selection, here illustrated with a transporter gene knockout mutant in competition with others[749] that does not take up toxic levels of an otherwise cytotoxic drug D.

use microbial colonies expressing the protein of interest to screen for activity directly *in situ*.[816–818] Advantages to this include the ability to use enzyme-coupled assays (like HRP)[819,820] or substrates of poor solubility or viscosity.[821]

Microfluidics, microdroplets and microcompartments

Sometimes the 'host' and the screen are virtually synonymous, as this kind of miniaturisation can also offer considerable speeds.[822–825] Thus, there are trends towards the analysis of directed evolution experiments in microcompartments,[766,826–831] using suitable microfluidics[777,832–838] or picodroplets.[831,839–843] Agresti *et al.*[844] have shown that microfluidics using picolitre-volume droplets can screen a library of 10^8 HRP mutants in 10 hours. Although further refinement of microfluidics-based screening is required before its use becomes commonplace, it is clear that it has the capability to process the larger and more diverse libraries that one wishes to investigate.

Assessment of diversity and its maintenance

By now we have acquired a population of clones that are 'better' in some sense(s) than those of their parents. If we measure only fitnesses, however, as we have implicitly done thus far, we have only half the story, and we now return to the question of using knowledge of where we are or have been in a search space to optimize how we navigate it. There is of course a considerable literature on the role of 'genetic' and related searches in all

kinds of single and multi-objective optimisation (see *e.g.* ref. 106, 107, 110, 113, 116, 210–213 and 845–858), all of which recognises that there is a trade-off between 'exploration' (looking for productive parts of the landscape) and 'exploitation' (performing more local searches in those parts). Methods or algorithms such as 'efficient global optimisation'[859] calculate these explicitly. Of course 'where' we are in the search space is simply encoded by the protein's sequence.

There is thus an increasing recognition that for the assessment[860–863] and maintenance[864] of diversity under selection one needs to study sequence-activity relationships. When DNA sequencing was much more expensive, methods were focused on assessing functionally important residues (*e.g.* ref. 865–868). As sequences became more readily available, methods such as PROSAR[219,232,233,869] were used to fix favourable amino acids, a strategy that proved rather effective (albeit that it does not consider epistasis). Now (although sequence-free methods are also possible[340,870–872]), as large-scale DNA (including 'next-generation') sequencing becomes commonplace in DE,[873–876] we may hope to see large and rich datasets becoming openly available to those who care to analyse them.

Sequence-activity relationships and machine learning

A historically important development in what is nowadays usually known as machine learning (ML)[877–879] was the recognition that it is possible to learn relationships (in the form of mathematical models) between paired inputs and outputs – in the classical case between mass spectra and the structures of the molecules that had generated them[880–884] – and more importantly that one could apply such models successfully in a predictive manner to molecules and spectra not used in the generation of the model. Such models are thus said to 'learn', or to 'generalise' to unseen samples (Fig. 15).

In a similar vein, the first implementation of the idea that one could learn a mathematical model that captured the (normally rather nonlinear) relationships between a macromolecule's sequence and its activity in an assay of some kind, and thereby use that model to predict (*in silico*) the activities of untested sequences, seems to be that of Jonsson *et al.*[885] These authors[885] used partial least squares regression (a statistical model rather than ML – for the key differences see ref. 886) to establish a 'quantitative sequence-activity model' (QSAM) between (a numerical description of) 68-base-pair fragments of 25 *E. coli* promoters and their corresponding promoter strengths. The QSAM was then used to predict two 68 bp fragments that it was hoped would be more potent promoters than any in the training set. While extrapolation, to 'fitnesses' beyond what had been seen thus far, was probably a little optimistic, this work showed that such kinds of mappings were indeed possible (*e.g.* ref. 887–891). We have used such methods for a variety of protein-related problems, including predicting the nature and visibility of protein mass spectra.[892–894]

As a separate example, we used another ML method known as 'random forests'[895] to learn the relationship between features of some 40 000 macromolecular (DNA aptamer) sequences and their activities,[23] and could use this to predict (from a search space some 14 orders of magnitude greater) the activities of previously untested sequences. While considerable work is going on in structural biology, we are always going to have very many more (indeed increasingly more) sequences than we have structures; thus we consider that approaches such as this are going to be very important in speeding up DE in biocatalysis and improving the functional annotation of proteins. In particular, those performing directed evolution can have simultaneous access to all sequences and activities for a given protein.[896,897] In contrast, an individual protein undergoing natural evolution cannot in any sense have a detailed 'memory' of its evolutionary past or pathway and in any event cannot (so far as is known, but *cf.* ref. 122 and 898) itself determine where to make mutations (only what to select on the basis of a poorly specified fitness). Machine learning methods seem extremely well suited for searching landscapes of this type.[23,56,107,677,899] Overall, this is a very important difference between natural evolution and (Experimenter-) Directed Evolution.

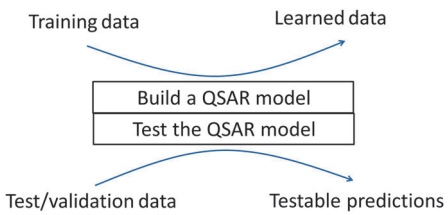


Fig. 15 The principles of building and testing a machine learning model, illustrated here with a QSAR model. We start with paired inputs and outputs (here sequences and activities) and learn a nonlinear mapping between the two. Methods for doing this that we have found effective include genetic programming[1345] and random forests.[23] In a second phase, the learned model is used to make predictions on an unseen validation and/or test set[1346] to establish that the model has generalized well.

The objective function(s): metrics for the effectiveness of biocatalysts

This is not a review of enzyme kinetics and kinetic mechanisms,[549,900–902] and for our purposes we shall mainly assume that we are dealing with enzymes that catalyse thermodynamically favourable reactions, operating *via* a Michaelis–Menten type of reaction whose kinetic properties can largely be characterized *via* binding or Michaelis constants plus a (slower) catalytic rate constant k_{cat} that is equivalent to the enzyme's turnover number (with units of reciprocal time). Much literature (*e.g.* ref. 549, 902 and 903) summarises the view that an appropriate measure of the effectiveness of an enzyme is a high value of k_{cat}/K_m, effected *via* the transduction of the initial energy of substrate/cofactor binding.[903–905] Certainly the lowering of K_m alone is a very poor target for most purposes in directed evolution where initial substrate concentrations are large. Better (as an objective function) than enantiomeric excess for chiral reactions producing a preferred *R* form (preferred over the *S* form) is a *P* factor or *E*

factor $(k_{cat,R}/K_{m,R})/(k_{cat,S}/K_{m,S})$[906] of a product. For industrial purposes, we are normally much more interested in the overall conversion in a reactor, rather than any specific enzyme kinetic parameter. Hence, the space-time yield (STY) or volume–time output (VTO) over a specified period, whose units are expressed in amount × (volume × time)$^{-1}$ (*e.g.* ref. 907–911) has also been preferred as an objective function. This is clearly more logical from the engineering point of view, but for understanding how best to drive directed evolution at the molecular level, it is arguably best to concentrate on k_{cat}, *i.e.* the turnover number, which is what we do here.

The distribution of k_{cat} values among natural proteins

Not least because of the classic and beautiful work on triose phosphate isomerase, an enzyme that is operating almost at the diffusion-controlled limit,[383,912] there is a quite pervasive view that natural evolution has taken enzymes 'as far as they can go' to make 'proficient' enzymes (*e.g.* ref. 913–915). Were this to be the case, there would be little point in developing directed evolution save for artificial substrates. However, it is not; most enzymes operate *in vivo* (and *in vitro*) at rates much lower than diffusion-controlled limits[916,917] (online databases of enzyme kinetic parameters include BRENDA[918] and SABIO-RK[919]). One assumes that this is largely because evolution simply had no need (*i.e.* faster enzymes did not confer sufficient evolutionary advantage)[920] to select for them to increase their rates beyond that necessary to lose substantial flux control (a systems property[921–925]). It is this in particular that makes it both desirable and possible to improve k_{cat} or k_{cat}/K_m values over what Nature thus far has commonly achieved.

In biotransformations studies, most papers appear to report processes in terms of g product × (g enzyme × day)$^{-1}$; while process parameters are important,[907] this serves (and is probably designed) to hide the very poor molecular kinetic parameters that actually pertain. K_m is largely irrelevant because the concentrations in use are huge; thus our focus is on k_{cat}. While DE has been shown to be capable of improving enzyme turnover numbers significantly, calculations show that even the 'poster child' examples (prositagliptin ketone transaminase,[926] ~0.03 s^{-1}; halohydrin dehydrogenase,[219] ~2 s^{-1}; isopropylmalate dehydrogenase,[927] ~5 s^{-1}; lovD,[368] ~2 s^{-1}) have turnover numbers that are very poor compared to those typical of primary metabolism, let alone the diffusion-controlled rates (depending on K_m) of nearer 10^6–10^7 s^{-1}.[916,917]

What enzyme properties determine k_{cat} values?

Almost since the beginning of molecular enzymology, scientists have come to wonder what particular features of enzymes are 'responsible' for enzyme catalytic power (*i.e.* can be used to explain it, from a mechanistic point of view).[928–930] It is implausible that there will be a unitary answer, as different sources will contribute differently in different cases. Scientifically, one may assume from the many successes of protein engineering that comparing various related sequences (and structures) by homology will be productive for our understanding of enzymology. Directed evolution studies increase the opportunities massively.

In general terms (*e.g.* ref. 930–933), preferred contributions to mechanisms have their different proponents, with such contributions being ascribed variously to a 'Circe effect',[904] strain or distortion,[934] electrostatic pre-organisation,[933,935–939] hydrogen tunneling,[940–947] reorganization energy,[948] and in particular various kinds of fluctuations and enzyme dynamics.[254,379,930,942,944–946,949–978] Less well-known flavours of dynamics include the idea that solitons may be involved.[979,980] Overall, we consider that the 'dynamics' view of enzyme action is especially attractive for those seeking to increase the turnover number of an enzyme.

This is because what is not in doubt is that following substrate binding at the active site (that is dominantly responsible for substrate affinity and the degree of specificity), the binding energy has been 'used up' (Fig. 16) and is not thereby available to drive the catalytic step in a thermodynamic sense. This means that the protein must explore its energy landscape *via* conformational fluctuations that are essentially isoenergetic,[951,966,981] before finding a configuration that places the active site residues into positions appropriate to effect the chemical catalysis itself, that happens as a 'protein quake'[981,982] in picoseconds or less.[983] The source of these motions, whether normal mode or otherwise,[984,985] can only be the protein and solvent fluctuations in the heat bath, and this means that their origins can lie in any parts of the protein, not just the few amino acids at the active site. Two exceedingly important corollaries of this follow. The first is that one may hope to predict or reflect this through the methods of molecular dynamics even when the active site is essentially unchanged, and this has recently been shown[368]). The second corollary is that one should expect it to be found that successful directed evolution programs that increase k_{cat} lead to

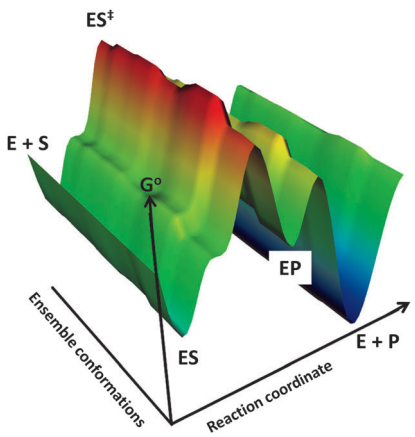


Fig. 16 A standard representation of an energy diagram for enzyme catalysis. Substrate binding is thermodynamically favourable, but to effect the catalytic reaction thermal energy is used to take the reaction to the right, often shown as a barrier represented by one or more 'transition states'. Changes in the K_m and K_d (affecting substrate affinity) can be influenced most directly by mutagenesis of the residues at the active site whilst changes in the k_{cat} occur primarily from mutagenesis of residues away from the active site (which can affect the fluctuations in enzyme structure required either for crossing the transition state ES‡ or by tunnelling under the barrier). At all points there are multiple roughly isoenergetic conformational (sub)states. Figure based on elements of those in ref. 930 and 966.

many mutations that are very distant from the active site. This can also serve, at least in part, to account for why surface post-translational modifications such as glycosylation can have significant effects on turnover (*e.g.* ref. 986).

The importance of non-active-site mutations in increasing k_{cat} values

In general terms, it is known that there is a considerable amount of long-range allostery in proteins,[987] such that distant mutations couple to the active site.[988] Indeed, most mutations (and amino acid residues) are necessarily distant from the active site, and there is a lack of correlation between a mutation's influence on k_{cat} and its proximity to the active site[989] (by contrast, specificity is determined much more by the closeness to the active site[990]). We still do not have that much data, since we require 3D structural information on many related variants; however, a number of excellent examples (Table 2) are indeed consistent with this recognition that effective DE strategies that raise k_{cat} require that we spread our attention throughout the protein, and do not simply concentrate on the active site (Fig. 17). Indeed mutants with major improvements in k_{cat} may display only minor changes in active site structure.[368,938,974,991]

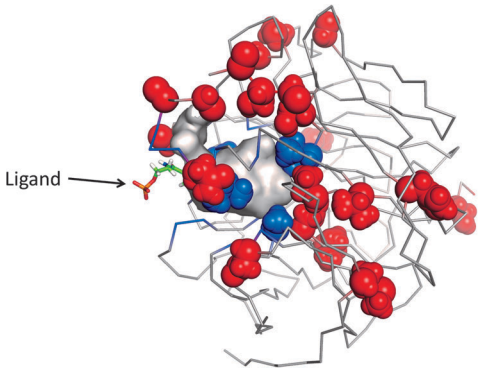


Fig. 17 The residues that influence k_{cat} tend to be distributed throughout an enzyme. The amino acid side chains of each of the 24 mutations obtained[993] by the directed evolution of a Dielsalderase (PDB: 4O5T) are highlighted. The active site pocket is shown in grey, while all mutated residues within 5 Å of the ligand (blue) are differentiated from those more than 5 Å away (red). This illustrates that the majority of mutations influencing k_{cat} are not in close proximity to the active (substrate-binding) site. The figure was prepared using PyMol.

What if we lack the structure? Folding up proteins 'from scratch'

A very particular kind of dynamics is that which leads a protein to fold into its tertiary structure in the first place, and the purpose of this brief section is to draw attention to some recent advances that might allow us to do this computationally. Advances in specialist computer hardware (albeit not yet widely available) can now make a prediction of how a given (smallish) protein sequence will fold up '*ab initio*'.[998–1002] However, there are many more protein sequences than there are structures, and this gap is destined to become considerably wider[1003] as sequencing methods continue to increase in speed.[102] The need for methods that can fold up proteins accurately '*de novo*' (from their sequences) is thus acute.[1004] However, despite a number of advances (*e.g.* ref. 1000, 1005 and 1006) this is not yet routine. The problem is, of course, that the search space of possible structures is enormous,[1] and largely unconstrained. As well as using more powerful hardware, the real key is finding suitable constraints. An important recognition is that the covariance of residues in a series of homologous functional proteins provides a massive constraint on the inter-residue contacts and thus what structures they might adopt, and substantial advances have recently been made by a number of groups[197,199,1007–1011] in this regard. Directed evolution supplies an obvious means of creating and assessing suitable sequences.

Metals

As mentioned above, many proteins use metals and cofactors to aid the chemistry that they can catalyse, and while we shall

Table 2 Some examples of improvements in biocatalytic activities that have been achieved using directed evolution, focusing on examples where most relevant mutations are in amino acids that are distal to the active site

Target	Fold improvement over starting point	Ref.	Other notes
Cytochrome P450	9000	56 and 992	20 from generation 5, more than 15 away from active site
Diels–Alderase	k_{cat} 108-fold; catalytic power 9000-fold	993	21 aa, 16 outside active site
Glycerol dehydratase	336	994	2 aa, both very distant from active site
Glyphosate acyltransferase	200 in k_{cat}	995	21 mutations, only 4 at active site
Halohydrin dehalogenase	4000 in volumetric productivity	219	35 mutations, only 8 at active site
3-Isopropylmalate dehydrogenase	65	927	8 mutations, 6 distant from active site
LovD	>1000	368	29 mutations, 18 on enzyme surface
Phosphotriesterase	25	996	7, only 1 at active site
Prositagliptin ketone transaminase	∞ (no starting activity)	926	27 mutations, 17 binding substrate. 200 g L^{-1}, >99.5 ee
Triose phosphate isomerase	>10 000	386	36 mutations, only 1 at active site (NB effects on dimerisation, also implying distant effects)
Valine aminotransferase	21 000 000	997	17 mutations, only 1 at active site

not discuss cofactors, a short section on metalloenzymes is warranted, not least since nearly half of natural enzymes contain metals,[1012] albeit that free metals can be quite toxic.[1013–1015]

To this end, if one wishes to keep open the possibility of incorporating metals into proteins undergoing DE (sometimes referred to as hybrid enzymes[1016–1019]), it is necessary to understand the common mechanisms, residues and structures involved.[460,461,1020–1042]

Some specific and unusual examples include high-valent metal catalysis,[1043] multi-metal designs as in a di-iron hydrylation reaction,[1044] a protein whose fluorescence is metal-dependent[1045] and various chelators, quantum dots and so on[1046–1050] and metallo-enzymes based on (strept)avidin–biotin technology.[1051–1053]

A particular attraction of DE is that it becomes possible to incorporate metal ions that are rarely (or never) used in living organisms, to provide novel functions. Examples include iridium,[1054] rhodium[1055] and uranium (uranyl).[1056,1057]

Enzyme stability, including thermostability

In general, the rates of chemical reactions increase with temperature, and if we evolve k_{cat} to high levels we may create processes in which temperature may rise naturally anyway (and some processes may simply require it[1058]). In a similar vein, protein stability tends to decrease with increasing temperature, and there is commonly[1059–1061] (though not always[1062]) a trade-off between k_{cat} and thermostability, including at the cellular level.[1063] This relationship depends effectively on the evolutionary pathway followed.[1062] As discussed above, thermostability may also sometimes (but not always[171–173]) correlate with evolvability,[175] and is the result of multiple mutations each contributing a small amount.[1064–1069]

Of course the 'first law' of directed evolution is that you get what you select for (even if you did not mean to). Thus if thermostability is important one must incorporate it into one's selection regime, typically by screening for it.[1070,1071] Of course if one uses a thermophile such as *T. thermophilus* then *in vivo* selection is possible, too.[1072]

As rehearsed above, protein flexibility (a somewhat ill-defined concept[87,1073]) is related to k_{cat}, and most residues involved in improving k_{cat} are away from the active site, at the protein surface (where they are bombarded by solvent thermal fluctuations). The connection between flexibility and thermostability is not well understood, and it does not always follow that less flexibility provides greater stability.[1074,1075] However, one might suppose that some residues that contribute flexibility are most important for (*i.e.* contribute significantly to) thermostability too. This is indeed the case.[989,1076,1077] Indeed, the same blend of design and focused (thus semi-empirical) DE that has proved valuable for improving k_{cat} values seems to be the best strategy for enhancing thermostability too.[1078–1080]

Some aspects of thermostability[1081–1087] can be related to individual amino acids (*e.g.* an ionic or H-bond formed by an arginine is of greater strength than that formed by a corresponding lysine, or thermophilic enzymes have more charged and hydrophobic but fewer polar residues[1088,1089]). However, some aspects are best based on analyses of the 3D structure,[456,1090,1091] *e.g.* intra-helix ion pairs[1092,1093] and packing density.[1068,1094] Thus, Greaves and Warwicker[1095] conclude that "charge number relates to solubility, whereas protein stability is determined by charge location". The choice of which residues to focus on can be assisted (if a structure is available) by looking at the local flexibility *via* methods such as mutability[1096,1097] or *via* *B*-factors,[1062,1098,1099] or *via* certain kinds of mass spectrometry.[1100–1108] Constraint Network Analysis[1109] provides a useful strategy for choosing which residues might be most important for thermostability. Unnatural amino acids may be beneficial too; thus fluoro-aminoacids can increase stability.[1110–1112]

To disentangle the various contributions to k_{cat} and thermostability, what we need are detailed studies of sequences and structures as they relate to both of these, and published ones remain largely lacking. However, the goal of finding sequence changes that improve both k_{cat} and thermostability is exceptionally desirable. It should also be attainable, on the grounds that protein structural constraints that increase the rate of desirable conformational fluctuations while minimizing those that do not help the enzyme to its catalytically active confirmations must exist and will tend to have this precise effect.

Finally, thermal stresses are not the only stresses that may pertain during a biocatalytic process, albeit sometimes the same mutations can be beneficial in both (*e.g.* in permitting resistance to oxidation[1113,1114] or catalysis in organic solvents[1115]).

Solvency

While our focus is on evolving proteins, those that are catalyzing reactions are always immersed in a solvent, and we cannot ignore this completely. Although 'bulk' measurements of solvent properties are typically unsuitable for molecular analyses of transport across membranes,[269,335,356,362,1116–1119] it is the case that some of the binding energy used in enzyme catalysis is effectively used in transferring a substrate from a usually hydrophilic aqueous phase to a usually more 'hydrophobic' protein phase. In general, the increased mass/hydrophobicity is also accompanied by a changed value for K_m.[916] This can lead to some interesting effects of organic solvents, and solvent mixtures,[1120] on the specificity,[1121–1126] equilibria[1127] and catalytic rate constants[1128,1129] of enzymes, for reasons that are still not entirely understood. However, because the intention of many DE programs is the production of enzymes for use in industrial processes, the ability to function in organic solvents is often another important objective function, and can be solved *via* the above strategies.[577] One recent trend of note is the exploitation of ionic liquids[1130,1131] and 'deep eutectic solvents'[1132–1135] in biocatalysis.

Reaction classes

Apart from circumstances involving extremely reactive substrates and products, there is no known reason of principle why one might not be able to evolve a biocatalyst for any more-or-less simple (*i.e.* one-step, mono- or bi-molecular) chemical reaction. Thus, one's imagination is limited only by the reactions chosen (nowadays, for a more complex pathway, *via* retrosynthetic and related strategies (ref. 1136–1148)). Given that these are practically limitless (even if one might wish to start with 'core' molecules[1149,1150]), we choose to be illustrative, and thereby provide a table of some of the kinds of reaction, reaction class or products for which the methods of DE have been used, with a slight focus on non-mainstream reactions. (Curiously, a convenient online database for these is presently lacking.) Our main point is that there seems no obvious limitation on reactions, beyond the case of very highly reactive substrates, intermediates or products, for which an enzymatic reaction cannot be evolved. Since the search space of possible enzymes can never be tested exhaustively, it is a safe prediction that we should expect this to hold for many more, and more complex, chemistries than have been tried to date, provided that the thermodynamics are favourable.

While the focus of this review is about how best to navigate the very large search spaces that pertain in directed enzyme evolution, we recognize that a number of processes including enzymes evolved by DE are now operated industrially.[327,328] Examples include sitagliptin,[926] generic chiral amine APIs,[1151] bio-isoprene,[1152] and atorvastatin.[1153]

Concluding remarks and future prospects

In our review above, we have developed the idea that the most appropriate strategy for improving biocatalysts involves a judicious interplay between design and empiricism, the former more focused at the active site that determines binding and specificity, while the latter might usefully be focussed more on other surface and non-active-site residues to improve k_{cat} and (in part) (thermo)stability. As our knowledge improves, design may begin to play a larger role in optimising k_{cat}, but we consider that this will still require a considerable improvement in our understanding of the relationships between enzyme sequence, structure and dynamics. Thus, protein improvement is likely to involve the creation of increasingly 'smart' variant libraries over larger parts of the protein.

Another such interplay relates to the combination of experimental ('wet') and computational ('dry') approaches. We detect a significant trend towards more of the latter,[519] for instance in the use[368,1235,1296,1297] of molecular dynamics to calculate properties that suggest which residues might be creating internal friction[1298,1299] and hence lowering k_{cat}. These examples help to illustrate that predictions and simulations *in silico* are likely to play an increasingly important role in predicting strategies for mutagenesis *in vitro*.

The increasing availability of genomic and metagenomic data, coupled to improvements in the design and prediction of protein structures (and maybe activities) will certainly contribute to improving the initialisation steps of DE. The availability of large sets of protein homologues and analogues will lead to a greater understanding of the relationships (Fig. 1) between protein sequence, structure, dynamics and catalytic activities, all of which can contribute to the design of DE experiments. Together with the development of improved synthetic biology methodology for DNA synthesis and variation, the tools for designing and initialising DE experiments are increasing greatly.

Specifically, the availability of large numbers of sequence-activity pairs may be used to learn to predict where mutations might best be tested. This decreases the empiricism of entirely random mutations in favour of synthetic biology strategies in which one has (at least statistically) more or less complete control over which sequences to make and test. Thus we see a considerable role for modern versions of sequence-activity mapping based on appropriate machine learning methods as a means of predicting where searches might optimally be conducted; this can be done *in silico* before creating the sequences themselves.[23] No doubt many useful datasets of this type exist in the databases of commercial organisations, but they need to become public as the likelihood is that crowdsourcing analyses would add value for their originators[1300] as well as for the common good.[1301]

In terms of optimisation algorithms, we have already pointed out that very few of the modern algorithms of evolutionary optimisation have been applied to the DE problem,[107] and the advent of synthetic biology now makes their development and comparison (given that no one size will fit all[1302–1306]) a worthwhile and timely endeavour. Complex DE algorithms that have no real counterpart in natural evolution can also now be carried out using the methods of synthetic biology.

Searching our empirical knowledge of reactions is becoming increasingly straightforward as it becomes digitised. As implied above, we expect to see an increasing cross-fertilisation between the fields of bioinformatics and cheminformatics[1307,1308] and text mining;[1309–1311] a very interesting development in this direction is that of Cadeddu *et al.*[1136]

Conspicuous by their absence in Table 3 are the members of one important set of reactions that are widely ignored (because they do not always involve actual chemical transformations). These are the transmembrane transporters, and they make up fully one third of the reactions in the reconstructed yeast[1312] and human[25,1313] metabolic networks. Despite a widespread and longstanding assumption (*e.g.* ref. 1314) that xenobiotics simply tend to 'float' across biological membranes according to their lipophilicity, it is here worth highlighting the considerable literature (that we have reviewed elsewhere, *e.g.* ref. 269, 335, 356, 357, 362, 1116–1119 and 1315), including a couple of experimental examples (ref. 749 and 750), that implies that the diffusion of xenobiotics through phospholipid bilayers in intact cells is normally negligible. It is now clear that transporters enhance (and are probably required for) the transmembrane transport even of hydrophobic molecules such as alkanes,[1316–1321] terpenoids,[1319,1322,1323] long-chain,[1324–1328] and short-chain[1329–1332]

Table 3 Some reactions, reaction classes or product types for which DE has proved successful. We largely exclude the very well-established programmes such as ketone and other stereoselective reductases, which along with various other reactions aimed at pharmaceutical intermediates have recently been reviewed in *e.g.* ref. 326, 328 and 1154–1162. Chiralities are implicit

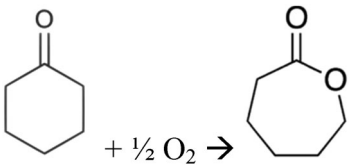
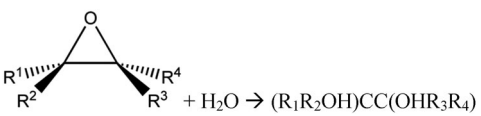
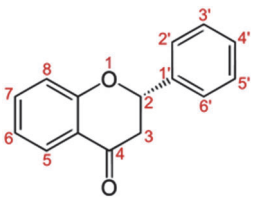
Reaction (class) or substrate/product	Illustrative ref.
Aldolases *e.g.* $R_1CHO + R_2C(=O)R_3 \rightleftharpoons R_1C(=O)CH_2C(O)R_3$	462, 1163 and 1164
Alkenyl and arylmalonate decarboxylases *e.g.* $HOOCC(R_1R_2)COOH \rightarrow HC(R_1R_2)COOH$	1165
Amines	1166–1169
Amine dehydrogenase $RC(=O)Me + NH_3 + NADH + H^+ \rightarrow RCHNH_2Me + H_2O + NAD^+$	1167
Antifreeze proteins	1170
Azidation $RH \rightarrow RN_3$	1171 and 1172
Baeyer–Villiger monooxygenases	1173 and 1174
	
Beta-keto adipate $HOOCCH_2CH_2C(=O)CH_2COOH$	1175
Carotenoid biosynthesis	1176
Chlorinase $Ar–H \rightarrow Ar–Cl$	1177–1180
Chloroperoxidase $RH + Cl^- + H_2O_2 \rightarrow RCl + H_2O + OH^-$	1181–1187
CO groups	1157
Cytochromes P450 *e.g.* $R–H \rightarrow R–OH$	56, 366, 632, 633, 674, 692, 992 and 1188–1208
Diels–Alderases *e.g.* $CH_2=CHCH=CH_2 + CH_2=CH_2 \rightarrow$ cyclohexene	378, 380, 993, 1209 and 1210
DNA polymerase	1211 and 1212
Endopeptidases	769
Esterase enantioselectivity	1213
Epoxide hydrolase	947
	
Flavanones	1214–1221
	
Fluorinase	1178 and 1222 (and see ref. 1223)
Fatty acids	1224–1226
Glyphosate acyltransferase $HOOCCH_2NHCH_2PO_3^{2-} +$ $AcCoA \rightarrow HOOCCH_2N(CH_3C=O)CH_2PO_3^{2-} + CoA$	995 and 1227–1229

Table 3 (*continued*)

Reaction (class) or substrate/product	Illustrative ref.
Glycine (glyphosate) oxidase $HOOCCH_2NHCH_2PO_3^{2-} + O_2 \rightarrow$ $OHC–COOH + H_2NCH_2\ PO_3^{2-} + H_2O_2$	1229–1231
Haloalkane dehalogenase $R_1C(HBr)R_2 + H_2O \rightarrow R_1C(HOH)R_2 + H^+ + Br^-$	1232–1235
Halogenase $Ar–H \rightarrow Ar–Hal$	1236–1239
Hydroxytyrosol $2–NO_2–Ph–CH_2CH_2OH + O_2 \rightarrow 2–OH, 3–OH–Ph–CH_2CH_2OH + NO_2$	1240
Kemp eliminase	377 and 1241–1247
Ketone reductions $R_1–C(=O)–R_2 \rightarrow R_1–CH(OH)–R_2$	1248 and 1249
Laccase	1250–1252
Michael addition $R–CH_2CHO + Ph–CH=CHNO_2 \rightarrow OHC–CH(R)–CH(Ph)CH_2NO_2$	1253
Monoamine oxidase $R_1R_2CHCH_2NR_3R_4 + 1/2O_2 \rightarrow$ $R_1R_2C=NR_3$ (R_4 = H) or $R_1R_2CH=N^+R_3R_4$ (R_4 = alkyl) + H_2O	728, 1151 and 1254–1256
Nitrogenase $N_2 + 3H_2 \rightarrow 2NH_3$	1257
Nucleases	1258
Old yellow enzyme (activated alkene reductions)	664, 1259 and 1260
Paraoxonase $R_1(R_2O)(R_3O)–P=O \rightarrow R_1(R_2O)(HO)P=O + R_3OH$	1261–1263
Peroxidase	1264 and 1265
Phospho(mono/di/tri)esterases	255, 831 and 1266–1271
Polyketides	1272–1275
Polylactate	1276
Redox enzymes	1277–1279
Reductive cyclisation	1280
Restriction protease	1281
Retro-aldolase *e.g.* $R_1C(=O)CH_2C(O)R_3 \rightleftharpoons R_1CHO + R_2C(=O)R_3$	463 and 1282
Sesquiterpene synthases	1283
Tautomerases $Ar–CH=C(OH)COOH \rightleftharpoons Ar–CH_2COCOOH$	1284
Terpene synthase/cyclase	1285–1287
Transaldolase erythrose-4-phosphate + fructose-6-phosphate → glyceraldehyde-3-phosphate + sedoheptulose-7-phosphate	1288 and 1289
Transketolase $RCHO + HOCH_2COCOOH \rightarrow RCH(OH)COCH_2OH$	1290–1293
Zinc finger proteins	1294 and 1295

fatty acids, and even CO_2.[1333,1334] This may imply a significantly enhanced role for transporter engineering in whole cell biocatalysis.

The recent introduction of the community standard Synthetic Biology Open Language (SBOL) will certainly facilitate the sharing and re-use of synthetic biology designed sequences and modules. Beginning in 2008, the development of SBOL has been driven by an international community of computational synthetic biologists, and has led to the introduction of an initial standard for the sharing of synthetic DNA sequences[1335] and also for their visualisation. A recent proposal has introduced a more complete extension to the language, covering interactions between synthetic sequences, the design of modules and specification of their overall function.[1336] Just as with the Systems Biology Markup Language,[1337] the Systems Biology Graphical Notation,[1338] and related controlled vocabularies, metadata and ontologies for knowledge exchange in systems biology[1339,1340] and metabolomics,[1341] the availability of these kinds of standards will help move the field forward considerably.

Overall, we conclude that existing and emerging knowledge-based methods exploiting the strategies and capabilities of synthetic biology and the power of e-science will be a huge driver for the improvement of biocatalysts by directed evolution. We have only just begun.

Acknowledgements

We thank Chris Knight, Rainer Breitling, Nigel Scrutton and Nick Turner for very useful comments on the manuscript, Colin Levy for some material for the figures, and the Biotechnology and Biological Sciences Research Council for financial support (grant BB/M017702/1). This is a contribution from the Manchester Centre for Synthetic Biology of Fine and Speciality Chemicals (SYNBIOCHEM).

References

1 D. B. Kell, Scientific discovery as a combinatorial optimisation problem: how best to navigate the landscape of possible experiments? *BioEssays*, 2012, **34**, 236–244.
2 *Phylogenetic analysis of DNA sequences*, ed. M. M. Miyamoto and J. Cracraft, Oxford University Press, Oxford, 1991.
3 R. D. M. Page and E. C. Holmes, *Molecular evolution: a phylogenetic approach*, Blackwell Science, Oxford, 1998.
4 M. J. Harms and J. W. Thornton, Evolutionary biochemistry: revealing the historical and physical causes of protein properties, *Nat. Rev. Genet.*, 2013, **14**, 559–571.
5 D. G. Gibson, G. A. Benders, K. C. Axelrod, J. Zaveri, M. A. Algire, M. Moodie, M. G. Montague and J. C. Venter, H. O. Smith and C. A. Hutchison, 3rd, One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic Mycoplasma genitalium genome, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 20404–20409.
6 D. G. Gibson, G. A. Benders, C. Andrews-Pfannkoch, E. A. Denisova, H. Baden-Tillson, J. Zaveri, T. B. Stockwell, A. Brownley, D. W. Thomas, M. A. Algire, C. Merryman, L. Young, V. N. Noskov, J. I. Glass, J. C. Venter, C. A. Hutchison 3rd and H. O. Smith, Complete chemical synthesis, assembly, and cloning of a *Mycoplasma genitalium* genome, *Science*, 2008, **319**, 1215–1220.
7 H. J. Frasch, M. H. Medema, E. Takano and R. Breitling, Design-based re-engineering of biosynthetic gene clusters: plug-and-play in practice, *Curr. Opin. Biotechnol.*, 2013, **24**, 1144–1150.
8 S. E. Ongley, X. Bian, B. A. Neilan and R. Muller, Recent advances in the heterologous expression of microbial natural product biosynthetic pathways, *Nat. Prod. Rep.*, 2013, **30**, 1121–1138.
9 A. Pourmir and T. W. Johannes, Directed evolution: selection of the host organism, *Comput. Struct. Biotechnol. J.*, 2012, **2**, e201209012.
10 S. V. Taylor, K. U. Walter, P. Kast and D. Hilvert, Searching sequence space for protein catalysts, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 10596–10601.
11 K. J. Waldron, J. C. Rutherford, D. Ford and N. J. Robinson, Metalloproteins and metal sensing, *Nature*, 2009, **460**, 823–830.
12 C. J. Jeffery, Moonlighting proteins—an update, *Mol. BioSyst.*, 2009, **5**, 345–350.
13 J. C. Moore, H. M. Jin, O. Kuchner and F. H. Arnold, Strategies for the *in vitro* evolution of protein function: Enzyme evolution by random recombination of improved sequences, *J. Mol. Biol.*, 1997, **272**, 336–347.
14 R. Bellman, *Adaptive control processes: a guided tour*, Princeton University Press, Princeton, NJ, 1961.
15 C. D. Manning, P. Raghavan and H. Schütze, *Introduction to information retrieval*, CUP, Cambridge, 2009.
16 T. Hastie, R. Tibshirani and J. Friedman, *The elements of statistical learning: data mining, inference and prediction*, Springer-Verlag, Berlin, 2001.
17 F. H. Arnold, Fancy footwork in the sequence space shuffle, *Nat. Biotechnol.*, 2006, **24**, 328–330.
18 W. S. Cleveland, *Visualizing data*, Hobart Press, Summit, NJ, 1993.
19 *Beautiful visualization: looking at data through the eyes of experts*, ed. J. Steele and N. Iliinsky, O'Reilly, Sebastopol, CA, 2010.
20 C. Ware, *Information visualization*, Morgan Kaufmann, San Francisco, 2000.
21 D. M. McCandlish, Visualizing fitness landscapes, *Evolution*, 2011, **65**, 1544–1558.
22 N. Yau, *Visualize this: the FlowingData guide to design, visualization and statistics*, Wiley, Indianapolis, IN, 2011.
23 C. G. Knight, M. Platt, W. Rowe, D. C. Wedge, F. Khan, P. Day, A. McShea, J. Knowles and D. B. Kell, Array-based evolution of DNA aptamers allows modelling of an explicit sequence-fitness landscape, *Nucleic Acids Res.*, 2009, **37**, e6.
24 S. Sahoo, L. Franzson, J. J. Jonsson and I. Thiele, A compendium of inborn errors of metabolism mapped onto the human metabolic network, *Mol. BioSyst.*, 2012, **8**, 2545–2558.
25 I. Thiele, N. Swainston, R. M. T. Fleming, A. Hoppe, S. Sahoo, M. K. Aurich, H. Haraldsdottír, M. L. Mo, O. Rolfsson, M. D. Stobbe, S. G. Thorleifsson, R. Agren, C. Bölling, S. Bordel, A. K. Chavali, P. Dobson, W. B. Dunn, L. Endler, I. Goryanin, D. Hala, M. Hucka, D. Hull, D. Jameson, N. Jamshidi, J. Jones, J. J. Jonsson, N. Juty, S. Keating, I. Nookaew, N. Le Novère, N. Malys, A. Mazein, J. A. Papin, Y. Patel, N. D. Price, E. Selkov Sr, M. I. Sigurdsson, E. Simeonidis, N. Sonnenschein, K. Smallbone, A. Sorokin, H. V. Beek, D. Weichart, J. B. Nielsen, H. V. Westerhoff, D. B. Kell, P. Mendes and B. Ø. Palsson, A community-driven global reconstruction of human metabolism, *Nat. Biotechnol.*, 2013, **31**, 419–425.
26 D. Dimova, M. Wawer, A. M. Wassermann and J. Bajorath, Design of multitarget activity landscapes that capture hierarchical activity cliff distributions, *J. Chem. Inf. Model.*, 2011, **51**, 258–266.
27 G. M. Maggiora, On outliers and activity cliffs—why QSAR often disappoints, *J. Chem. Inf. Model.*, 2006, **46**, 1535.

28 R. Guha and J. H. Van Drie, Structure–activity landscape index: identifying and quantifying activity cliffs, *J. Chem. Inf. Model.*, 2008, **48**, 646–658.
29 J. L. Medina-Franco, Activity cliffs: facts or artifacts? *Chem. Biol. Drug Des.*, 2013, **81**, 553–556.
30 D. Stumpfe and J. Bajorath, Frequency of occurrence and potency range distribution of activity cliffs in bioactive compounds, *J. Chem. Inf. Model.*, 2012, **52**, 2348–2353.
31 D. Stumpfe and J. Bajorath, Exploring activity cliffs in medicinal chemistry, *J. Med. Chem.*, 2012, **55**, 2932–2942.
32 D. Stumpfe, Y. Hu, D. Dimova and J. Bajorath, Recent progress in understanding activity cliffs and their utility in medicinal chemistry, *J. Med. Chem.*, 2014, **57**, 18–28.
33 R. Guha and J. L. Medina-Franco, On the validity *versus* utility of activity landscapes: are all activity cliffs statistically significant? *J. Cheminf.*, 2014, **6**, 11.
34 R. Guha, What makes a good structure activity landscape? Network metrics and structure representations as a way of exploring activity landscapes, *Abstracts of Papers of the American Chemical Society*, 2010, **240**.
35 R. Guha, The ups and downs of structure-activity landscapes, *Methods Mol. Biol.*, 2011, **672**, 101–117.
36 R. Guha, Exploring uncharted territories: predicting activity cliffs in structure-activity landscapes, *J. Chem. Inf. Model.*, 2012, **52**, 2181–2191.
37 R. Guha, Exploring Structure-Activity Data Using the Landscape Paradigm, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2012, **2**, 829–841, DOI: 10.1002/wcms.1087.
38 S. T. Rao and M. G. Rossmann, Comparison of super-secondary structures in proteins, *J. Mol. Biol.*, 1973, **76**, 241–256.
39 M. Saraste, P. R. Sibbald and A. Wittinghofer, The P-loop: a common motif in ATP- and GTP-binding proteins, *Trends Biochem. Sci.*, 1990, **15**, 430–434.
40 P. D. Dobson and A. J. Doig, Distinguishing enzyme structures from non-enzymes without alignments, *J. Mol. Biol.*, 2003, **330**, 771–783.
41 P. D. Dobson and A. J. Doig, Predicting enzyme class from protein structure without alignments, *J. Mol. Biol.*, 2005, **345**, 187–199.
42 J. Minshull, J. E. Ness, C. Gustafsson and S. Govindarajan, Predicting enzyme function from protein sequence, *Curr. Opin. Chem. Biol.*, 2005, **9**, 202–209.
43 L. Han, J. Cui, H. Lin, Z. Ji, Z. Cao, Y. Li and Y. Chen, Recent progresses in the application of machine learning approach for predicting protein functional class independent of sequence similarity, *Proteomics*, 2006, **6**, 4023–4037.
44 Z. Q. Tang, H. H. Lin, H. L. Zhang, L. Y. Han, X. Chen and Y. Z. Chen, Prediction of functional class of proteins and peptides irrespective of sequence homology by support vector machines, *Bioinf. Biol. Insights*, 2007, **1**, 19–47.
45 J. L. Faulon, M. Misra, S. Martin, K. Sale and R. Sapra, Genome scale enzyme-metabolite and drug-target interaction predictions using the signature molecular descriptor, *Bioinformatics*, 2008, **24**, 225–233.
46 H. Strömbergsson, P. Daniluk, A. Kryshtafovych, K. Fidelis, J. E. Wikberg, G. J. Kleywegt and T. R. Hvidsten, Interaction model based on local protein substructures generalizes to the entire structural enzyme-ligand space, *J. Chem. Inf. Model.*, 2008, **48**, 2278–2288.
47 T. Bray, P. Chan, S. Bougouffa, R. Greaves, A. J. Doig and J. Warwicker, Sites Identify: a protein functional site prediction tool, *BMC Bioinf.*, 2009, **10**, 379.
48 T. R. Hvidsten, A. Lægreid, A. Kryshtafovych, G. Andersson, K. Fidelis and J. Komorowski, A comprehensive analysis of the structure-function relationship in proteins based on local structure similarity, *PLoS One*, 2009, **4**, e6266.
49 D. M. Fowler, C. L. Araya, S. J. Fleishman, E. H. Kellogg, J. J. Stephany, D. Baker and S. Fields, High-resolution mapping of protein sequence-function relationships, *Nat. Methods*, 2010, **7**, 741–746.
50 T. Lee, H. Min, S. J. Kim and S. Yoon, Application of maximin correlation analysis to classifying protein environments for function prediction, *Biochem. Biophys. Res. Commun.*, 2010, **400**, 219–224.
51 C. R. Shyu, B. Pang, P. H. Chi, N. Zhao, D. Korkin and D. Xu, ProteinDBS v2.0: a web server for global and local protein structure search, *Nucleic Acids Res.*, 2010, **38**, W53–W58.
52 S. D. Brown and P. C. Babbitt, Inference of functional properties from large-scale analysis of enzyme superfamilies, *J. Biol. Chem.*, 2012, **287**, 35–42.
53 L. De Ferrari, S. Aitken, J. van Hemert and I. Goryanin, EnzML: multi-label prediction of enzyme classes using InterPro signatures, *BMC Bioinf.*, 2012, **13**, 61.
54 Y. Qi, M. Oja, J. Weston and W. S. Noble, A unified multitask architecture for predicting local protein properties, *PLoS One*, 2012, **7**, e32235.
55 S. Wright, The roles of mutation, inbreeding, crossbreeding and selection in evolution, in *Proc. Sixth Int. Conf. Genetics*, ed. D. F. Jones, Genetics Society of America, Austin TX, Ithaca, NY, 1932, pp. 356–366.
56 P. A. Romero and F. H. Arnold, Exploring protein fitness landscapes by directed evolution, *Nat. Rev. Mol. Cell Biol.*, 2009, **10**, 866–876.
57 J. A. G. M. de Visser and J. Krug, Empirical fitness landscapes and the predictability of evolution, *Nat. Rev. Genet.*, 2014, **15**, 480–490.
58 J. Maynard Smith, Natural selection and the concept of a protein space, *Nature*, 1970, **225**, 563–564.
59 A. R. Davidson, K. J. Lumb and R. T. Sauer, Cooperatively folded proteins in random sequence libraries, *Nat. Struct. Biol.*, 1995, **2**, 856–864.
60 J. R. Beasley and M. H. Hecht, Protein design: the choice of *de novo* sequences, *J. Biol. Chem.*, 1997, **272**, 2031–2034.
61 T. Matsuura, A. Ernst and A. Plückthun, Construction and characterization of protein libraries composed of secondary structure modules, *Protein Sci.*, 2002, **11**, 2631–2643.
62 T. Matsuura and A. Plückthun, Strategies for selection from protein libraries composed of *de novo* designed

secondary structure modules, *Origins Life Evol. Biospheres*, 2004, **34**, 151–157.
63 D. D. Axe, Estimating the prevalence of protein sequences adopting functional enzyme folds, *J. Mol. Biol.*, 2004, **341**, 1295–1315.
64 L. H. Bradley, P. P. Thumfort and M. H. Hecht, *De novo* proteins from binary-patterned combinatorial libraries, *Methods Mol. Biol.*, 2006, **340**, 53–69.
65 J. J. Graziano, W. Liu, R. Perera, B. H. Geierstanger, S. A. Lesley and P. G. Schultz, Selecting folded proteins from a library of secondary structural elements, *J. Am. Chem. Soc.*, 2008, **130**, 176–185.
66 T. Schmidt-Goenner, A. Guerler, B. Kolbeck and E. W. Knapp, Circular permuted proteins in the universe of protein folds, *Proteins*, 2010, **78**, 1618–1630.
67 J. Tanaka, N. Doi, H. Takashima and H. Yanagawa, Comparative characterization of random-sequence proteins consisting of 5, 12, and 20 kinds of amino acids, *Protein Sci.*, 2010, **19**, 786–795.
68 M. H. Hecht, M. W. West, J. Patterson, J. D. Mancias, J. R. Beasley, B. M. Broome and W. Wang, Designed combinatorial libraries of novel amyloid-like proteins, *Self-Assem. Pept. Syst. Biol., Med. Eng., [Workshop]*, 2001, 127–138.
69 Y. Ito, T. Kawama, I. Urabe and T. Yomo, Evolution of an arbitrary sequence in solubility, *J. Mol. Evol.*, 2004, **58**, 196–202.
70 A. D. Keefe and J. W. Szostak, Functional proteins from a random-sequence library, *Nature*, 2001, **410**, 715–718.
71 B. Kuhlman and D. Baker, Native protein sequences are close to optimal for their structures, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 10383–10388.
72 H. P. Yockey, *Information Theory and Molecular Biology*, Cambridge University Press, Cambridge, 1992.
73 Y. Wei and M. H. Hecht, Enzyme-like proteins from an unselected library of designed amino acid sequences, *Protein Eng., Des. Sel.*, 2004, **17**, 67–75.
74 G. Minervini, G. Evangelista, L. Villanova, D. Slanzi, D. De Lucrezia, I. Poli, P. L. Luisi and F. Polticelli, Massive non-natural proteins structure prediction using grid technologies, *BMC Bioinf.*, 2009, **10**(suppl 6), S22.
75 K. Prymula, M. Piwowar, M. Kochanczyk, L. Flis, M. Malawski, T. Szepieniec, G. Evangelista, G. Minervini, F. Polticelli, Z. Wisniowski, K. Salapa, E. Matczynska and I. Roterman, *In silico* structural study of random amino acid sequence proteins not present in nature, *Chem. Biodiversity*, 2009, **6**, 2311–2336.
76 M. Scalley-Kim and D. Baker, Characterization of the folding energy landscapes of computer generated proteins suggests high folding free energy barriers and cooperativity may be consequences of natural selection, *J. Mol. Biol.*, 2004, **338**, 573–583.
77 S. Kamtekar, J. M. Schiffer, H. Xiong, J. M. Babik and M. H. Hecht, Protein design by binary patterning of polar and nonpolar amino acids, *Science*, 1993, **262**, 1680–1685.
78 M. W. West and M. H. Hecht, Binary patterning of polar and nonpolar amino acids in the sequences and structures of native proteins, *Protein Sci.*, 1995, **4**, 2032–2039.
79 H. Xiong, B. L. Buckwalter, H. M. Shieh and M. H. Hecht, Periodicity of polar and nonpolar amino acids is the major determinant of secondary structure in self-assembling oligomeric peptides, *Proc. Natl. Acad. Sci. U. S. A.*, 1995, **92**, 6349–6353.
80 D. A. Moffet, J. Foley and M. H. Hecht, Midpoint reduction potentials and heme binding stoichiometries of *de novo* proteins from designed combinatorial libraries, *Biophys. Chem.*, 2003, **105**, 231–239.
81 M. H. Hecht, A. Das, A. Go, L. H. Bradley and Y. Wei, *De novo* proteins from designed combinatorial libraries, *Protein Sci.*, 2004, **13**, 1711–1723.
82 L. H. Bradley, Y. Wei, P. Thumfort, C. Wurth and M. H. Hecht, Protein design by binary patterning of polar and nonpolar amino acids, *Methods Mol. Biol.*, 2007, **352**, 155–166.
83 S. C. Patel, L. H. Bradley, S. P. Jinadasa and M. H. Hecht, Cofactor binding and enzymatic activity in an unevolved superfamily of *de novo* designed 4-helix bundle proteins, *Protein Sci.*, 2009, **18**, 1388–1400.
84 M. A. Fisher, K. L. McKinley, L. H. Bradley, S. R. Viola and M. H. Hecht, *De novo* designed proteins from a library of artificial sequences function in *Escherichia coli* and enable cell growth, *PLoS One*, 2011, **6**, e15364.
85 C. H. Shih, C. M. Chang, Y. S. Lin, W. C. Lo and J. K. Hwang, Evolutionary information hidden in a single protein structure, *Proteins*, 2012, **80**, 1647–1657.
86 C. M. Chang, Y. W. Huang, C. H. Shih and J. K. Hwang, On the relationship between the sequence conservation and the packing density profiles of the protein complexes, *Proteins*, 2013, **81**, 1192–1199.
87 T. T. Huang, M. L. D. Marcos, J. K. Hwang and J. Echave, A mechanistic stress model of protein evolution accounts for site-specific evolutionary rates and their relationship with packing density and flexibility, *BMC Evol. Biol.*, 2014, **14**, 78.
88 S. W. Yeh, T. T. Huang, J. W. Liu, S. H. Yu, C. H. Shih, J. K. Hwang and J. Echave, Local Packing Density Is the Main Structural Determinant of the Rate of Protein Sequence Evolution at Site Level, *BioMed Res. Int.*, 2014, 572409.
89 S. W. Yeh, J. W. Liu, S. H. Yu, C. H. Shih, J. K. Hwang and J. Echave, Site-Specific Structural Constraints on Protein Sequence Evolutionary Divergence: Local Packing Density *versus* Solvent Exposure, *Mol. Biol. Evol.*, 2014, **31**, 135–139.
90 A. Yamauchi, T. Nakashima, N. Tokuriki, M. Hosokawa, H. Nogami, S. Arioka, I. Urabe and T. Yomo, Evolvability of random polypeptides through functional selection within a small library, *Protein Eng.*, 2002, **15**, 619–626.
91 A. R. Davidson and R. T. Sauer, Folded proteins occur frequently in libraries of random amino acid sequences, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 2146–2150.

92 H. H. Guo, J. Choe and L. A. Loeb, Protein tolerance to random amino acid change, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 9205–9210.
93 N. Silver, *The signal and the noise: the art and science of prediction*, Penguin, New York, 2012.
94 F. J. Poelwijk, D. J. Kiviet, D. M. Weinreich and S. J. Tans, Empirical fitness landscapes reveal accessible evolutionary paths, *Nature*, 2007, **445**, 383–386.
95 M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin and B. K. Shoichet, Relating protein pharmacology by ligand chemistry, *Nat. Biotechnol.*, 2007, **25**, 197–206.
96 M. Bashton, I. Nobeli and J. M. Thornton, PROCOGNATE: a cognate ligand domain mapping for enzymes, *Nucleic Acids Res.*, 2008, **36**, D618–D622.
97 R. Adams, C. L. Worth, S. Guenther, M. Dunkel, R. Lehmann and R. Preissner, Binding sites in membrane proteins - diversity, druggability and prospects, *Eur. J. Cell Biol.*, 2012, **91**, 326–339.
98 A. M. Wassermann and J. Bajorath, BindingDB and ChEMBL: online compound databases for drug discovery, *Expert Opin. Drug Discovery*, 2011, **6**, 683–687.
99 D. E. Featherstone and K. Broadie, Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network, *BioEssays*, 2002, **24**, 267–274.
100 M. Soskine and D. S. Tawfik, Mutational effects and the evolution of new protein functions, *Nat. Rev. Genet.*, 2010, **11**, 572–582.
101 J. Shendure and H. Ji, Next-generation DNA sequencing, *Nat. Biotechnol.*, 2008, **26**, 1135–1145.
102 J. Shendure and E. L. Aiden, The expanding scope of DNA sequencing, *Nat. Biotechnol.*, 2012, **30**, 1084–1094.
103 J. I. Jiménez, R. Xulvi-Brunet, G. W. Campbell, R. Turk-MacLeod and I. A. Chen, Comprehensive experimental fitness landscape and evolutionary network for small RNA, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 14984–14989.
104 W. Rowe, M. Platt, D. Wedge, P. J. Day, D. B. Kell and J. Knowles, Analysis of a complete DNA-protein affinity landscape, *J. R. Soc., Interface*, 2010, **7**, 397–408.
105 Á. E. Eiben, R. Hinterding and Z. Michalewicz, Parameter control in evolutionary algorithms, *IEEE Trans. Evol. Comput.*, 1999, **3**, 124–141.
106 *Handbook of evolutionary computation.*, ed. T. Bäck, D. B. Fogel and Z. Michalewicz, IOP Publishing/Oxford University Press, Oxford, 1997.
107 S. O'Hagan, J. Knowles and D. B. Kell, Exploiting genomic knowledge in optimising molecular breeding programmes: algorithms from evolutionary computing, *PLoS One*, 2012, **7**, e48862.
108 G. Syswerda, Uniform crossover in genetic algorithms, in *Proc 3rd Int Conf on Genetic Algorithms*, ed. J. Schaffer, Morgan Kaufmann, 1989, pp. 2–9.
109 L. He, A. M. Friedman and C. Bailey-Kellogg, A divide-and-conquer approach to determine the Pareto frontier for optimization of protein engineering experiments, *Proteins*, 2012, **80**, 790–806.
110 J. Handl, D. B. Kell and J. Knowles, Multiobjective optimization in bioinformatics and computational biology, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 2007, **4**, 279–292.
111 J. D. Knowles and D. W. Corne, Approximating the nondominated front using the Pareto Archived Evolution Strategy, *Evol. Comput.*, 2000, **8**, 149–172.
112 J. D. Knowles and D. W. Corne, M-PAES: a memetic algorithm for multiobjective optimization, *Proc. 2000 Congr. Evol. Computation*, 2000, vol. 1 and 2, pp. 325–332.
113 K. Deb, *Multi-objective optimization using evolutionary algorithms*, Wiley, New York, 2001.
114 E. Zitzler, L. Thiele, M. Laumanns, C. M. Fonseca and V. G. da Fonseca, Performance assessment of multiobjective optimizers: An analysis and review, *IEEE Trans. Evol. Comput.*, 2003, **7**, 117–132.
115 J. Knowles, ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems, *IEEE Trans. Evol. Comput.*, 2006, **10**, 50–66.
116 *Multiobjective Problem Solving from Nature*, ed. J. Knowles, D. Corne and K. Deb, Springer, Berlin, 2008.
117 B. Maher, The case of the missing heritability, *Nature*, 2008, **456**, 18–21.
118 D. M. Weinreich, N. F. Delaney, M. A. Depristo and D. L. Hartl, Darwinian evolution can follow only very few mutational paths to fitter proteins, *Science*, 2006, **312**, 111–114.
119 W. Sung, M. S. Ackerman, S. F. Miller, T. G. Doak and M. Lynch, Drift-barrier hypothesis and mutation-rate evolution, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 18488–18492.
120 M. W. Nachman and S. L. Crowell, Estimate of the mutation rate per nucleotide in humans, *Genetics*, 2000, **156**, 297–304.
121 P. D. Keightley, Rates and fitness consequences of new mutations in humans, *Genetics*, 2012, **190**, 295–304.
122 P. D. Sniegowski, P. J. Gerrish and R. E. Lenski, Evolution of high mutation rates in experimental populations of *E. coli*, *Nature*, 1997, **387**, 703–705.
123 M. V. Rockman and L. Kruglyak, Recombinational landscape and population genomics of Caenorhabditis elegans, *PLoS Genet.*, 2009, **5**, e1000419.
124 P. D. Sniegowski and R. E. Lenski, Mutation and Adaptation - the Directed Mutation Controversy in Evolutionary Perspective, *Annu. Rev. Ecol. Evol. Syst.*, 1995, **26**, 553–578.
125 J. R. Peck and D. Waxman, Is life impossible? Information, sex, and the origin of complex organisms, *Evolution*, 2010, **64**, 3300–3309.
126 J. Franke, A. Klözer, J. A. G. M. de Visser and J. Krug, Evolutionary accessibility of mutational pathways, *PLoS Comput. Biol.*, 2011, **7**, e1002134.
127 B. Papp, R. A. Notebaart and C. Pál, Systems-biology approaches for predicting genomic evolution, *Nat. Rev. Genet.*, 2011, **12**, 591–602.
128 C. A. Orengo and J. M. Thornton, Protein families and their evolution-a structural perspective, *Annu. Rev. Biochem.*, 2005, **74**, 867–900.

129 A. D. Moore, S. Grath, A. Schuler, A. K. Huylmans and E. Bornberg-Bauer, Quantification and functional analysis of modular protein evolution in a dense phylogenetic tree, *Biochim. Biophys. Acta*, 2013, **1834**, 898–907.

130 T. E. Lewis, I. Sillitoe, A. Andreeva, T. L. Blundell, D. W. Buchan, C. Chothia, A. Cuff, J. M. Dana, I. Filippis, J. Gough, S. Hunter, D. T. Jones, L. A. Kelley, G. J. Kleywegt, F. Minneci, A. Mitchell, A. G. Murzin, B. Ochoa-Montano, O. J. Rackham, J. Smith, M. J. Sternberg, S. Velankar, C. Yeats and C. Orengo, Genome3D: a UK collaborative project to annotate genomic sequences with predicted 3D structures based on SCOP and CATH domains, *Nucleic Acids Res.*, 2013, **41**, D499–D507.

131 D. Xu, Protein databases on the internet, *Curr. Protoc. Mol. Biol.*, 2012, ch. 19, Unit 19.14.

132 L. H. Greene, T. E. Lewis, S. Addou, A. Cuff, T. Dallman, M. Dibley, O. Redfern, F. Pearl, R. Nambudiry, A. Reid, I. Sillitoe, C. Yeats, J. M. Thornton and C. A. Orengo, The CATH domain structure database: new protocols and classification levels give a more comprehensive resource for exploring evolution, *Nucleic Acids Res.*, 2007, **35**, D291–D297.

133 A. L. Cuff, I. Sillitoe, T. Lewis, O. C. Redfern, R. Garratt, J. Thornton and C. A. Orengo, The CATH classification revisited—architectures reviewed and new ways to characterize structural divergence in superfamilies, *Nucleic Acids Res.*, 2009, **37**, D310–D314.

134 A. L. Cuff, I. Sillitoe, T. Lewis, A. B. Clegg, R. Rentzsch, N. Furnham, M. Pellegrini-Calace, D. Jones, J. Thornton and C. A. Orengo, Extending CATH: increasing coverage of the protein structure universe and linking structure with function, *Nucleic Acids Res.*, 2011, **39**, D420–D426.

135 A. G. Murzin, S. E. Brenner, T. Hubbard and C. Chothia, SCOP: a structural classification of proteins database for the investigation of sequences and structures, *J. Mol. Biol.*, 1995, **247**, 536–540.

136 A. Andreeva, D. Howorth, J. M. Chandonia, S. E. Brenner, T. J. Hubbard, C. Chothia and A. G. Murzin, Data growth and its impact on the SCOP database: new developments, *Nucleic Acids Res.*, 2008, **36**, D419–D425.

137 N. K. Fox, S. E. Brenner and J. M. Chandonia, SCOPe: Structural Classification of Proteins—extended, integrating SCOP and ASTRAL data and classification of new structures, *Nucleic Acids Res.*, 2014, **42**, D304–D309.

138 S. Hunter, P. Jones, A. Mitchell, R. Apweiler, T. K. Attwood, A. Bateman, T. Bernard, D. Binns, P. Bork, S. Burge, E. de Castro, P. Coggill, M. Corbett, U. Das, L. Daugherty, L. Duquenne, R. D. Finn, M. Fraser, J. Gough, D. Haft, N. Hulo, D. Kahn, E. Kelly, I. Letunic, D. Lonsdale, R. Lopez, M. Madera, J. Maslen, C. McAnulla, J. McDowall, C. McMenamin, H. Mi, P. Mutowo-Muellenet, N. Mulder, D. Natale, C. Orengo, S. Pesseat, M. Punta, A. F. Quinn, C. Rivoire, A. Sangrador-Vegas, J. D. Selengut, C. J. Sigrist, M. Scheremetjew, J. Tate, M. Thimmajanarthanan, P. D. Thomas, C. H. Wu, C. Yeats and S. Y. Yong, InterPro in 2011: new developments in the family and domain prediction database, *Nucleic Acids Res.*, 2011, **40**, D306–D312.

139 S. Burge, E. Kelly, D. Lonsdale, P. Mutowo-Muellenet, C. McAnulla, A. Mitchell, A. Sangrador-Vegas, S. Y. Yong, N. Mulder and S. Hunter, Manual GO annotation of predictive protein signatures: the InterPro approach to GO curation, *Database*, 2012, **2012**, bar068.

140 A. Cuff, O. C. Redfern, L. Greene, I. Sillitoe, T. Lewis, M. Dibley, A. Reid, F. Pearl, T. Dallman, A. Todd, R. Garratt, J. Thornton and C. Orengo, The CATH hierarchy revisited-structural divergence in domain superfamilies and the continuity of fold space, *Structure*, 2009, **17**, 1051–1062.

141 L. Xie and P. E. Bourne, Detecting evolutionary relationships across existing fold space, using sequence order-independent profile-profile alignments, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 5441–5446.

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

143 G. J. Bartlett, N. Borkakoti and J. M. Thornton, Catalysing new reactions during evolution: economy of residues and mechanism, *J. Mol. Biol.*, 2003, **331**, 829–860.

144 P. F. Gherardini, M. N. Wass, M. Helmer-Citterich and M. J. Sternberg, Convergent evolution of enzyme active sites is not a rare phenomenon, *J. Mol. Biol.*, 2007, **372**, 817–845.

145 G. L. Holliday, C. Andreini, J. D. Fischer, S. A. Rahman, D. E. Almonacid, S. T. Williams and W. R. Pearson, MACiE: exploring the diversity of biochemical reactions, *Nucleic Acids Res.*, 2011, **40**, D783–D789.

146 E. Ferrada and A. Wagner, Evolutionary innovations and the organization of protein functions in genotype space, *PLoS One*, 2010, **5**, e14172.

147 U. Bastolla, M. Porto, H. E. Roman and M. Vendruscolo, A protein evolution model with independent sites that reproduces site-specific amino acid distributions from the Protein Data Bank, *BMC Evol. Biol.*, 2006, **6**, 43.

148 C. L. Worth, S. Gong and T. L. Blundell, Structural and functional constraints in the evolution of protein families, *Nat. Rev. Mol. Cell Biol.*, 2009, **10**, 709–720.

149 C. L. Worth, R. Preissner and T. L. Blundell, SDM—a server for predicting effects of mutations on protein stability and malfunction, *Nucleic Acids Res.*, 2011, **39**, W215–W222.

150 J. Overington, D. Donnelly, M. S. Johnson, A. Šali and T. L. Blundell, Environment-specific amino acid substitution tables: tertiary templates and prediction of protein folds, *Protein Sci.*, 1992, **1**, 216–226.

151 A. Šali and T. L. Blundell, Comparative protein modelling by satisfaction of spatial restraints, *J. Mol. Biol.*, 1993, **234**, 779–815.

152 M. E. Peterson, F. Chen, J. G. Saven, D. S. Roos, P. C. Babbitt and A. Sali, Evolutionary constraints on structural similarity in orthologs and paralogs, *Protein Sci.*, 2009, **18**, 1306–1315.

153 R. Mendez, M. Fritsche, M. Porto and U. Bastolla, Mutation bias favors protein folding stability in the evolution

of small populations, *PLoS Comput. Biol.*, 2010, **6** e1000767.
154 D. M. Taverna and R. A. Goldstein, The distribution of structures in evolving protein populations, *Biopolymers*, 2000, **53**, 1–8.
155 D. M. Taverna and R. A. Goldstein, Why are proteins so robust to site mutations? *J. Mol. Biol.*, 2002, **315**, 479–484.
156 D. M. Taverna and R. A. Goldstein, Why are proteins marginally stable? *Proteins*, 2002, **46**, 105–109.
157 M. A. DePristo, D. M. Weinreich and D. L. Hartl, Missense meanderings in sequence space: a biophysical view of protein evolution, *Nat. Rev. Genet.*, 2005, **6**, 678–687.
158 P. D. Williams, D. D. Pollock and R. A. Goldstein, Functionality and the evolution of marginal stability in proteins: inferences from lattice simulations, *Evol. Bioinf. Online*, 2006, **2**, 91–101.
159 R. A. Goldstein, The evolution and evolutionary consequences of marginal thermostability in proteins, *Proteins*, 2011, **79**, 1396–1407.
160 C. G. Langton, Life at the Edge of Chaos, *SFI S Sci. C*, 1992, **10**, 41–91.
161 S. A. Kauffman, *The origins of order*, Oxford University Press, Oxford, 1993.
162 P. Csermely, K. S. Sandhu, E. Hazai, Z. Hoksza, H. J. Kiss, F. Miozzo, D. V. Veres, F. Piazza and R. Nussinov, Disordered proteins and network disorder in network descriptions of protein structure, dynamics and function: hypotheses and a comprehensive review, *Curr. Protein Pept. Sci.*, 2012, **13**, 19–33.
163 P. Csermely, T. Korcsmáros, H. J. M. Kiss, G. London and R. Nussinov, Structure and dynamics of molecular networks: A novel paradigm of drug discovery. A comprehensive review, *Pharmacol. Therapeut.*, 2013, **138**, 333–408.
164 J. M. Carlson and J. Doyle, Highly optimized tolerance: a mechanism for power laws in designed systems, *Phys. Rev. E: Stat. Phys., Plasmas, Fluids, Relat. Interdiscip. Top.*, 1999, **60**, 1412–1427.
165 J. M. Carlson and J. Doyle, Complexity and robustness, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**(suppl 1), 2538–2545.
166 M. E. Csete and J. C. Doyle, Reverse engineering of biological complexity, *Science*, 2002, **295**, 1664–1669.
167 M. Csete and J. Doyle, Bow ties, metabolism and disease, *Trends Biotechnol.*, 2004, **22**, 446–450.
168 T. Zhou, J. M. Carlson and J. Doyle, Evolutionary dynamics and highly optimized tolerance, *J. Theor. Biol.*, 2005, **236**, 438–447.
169 F. J. Doyle 3rd and J. Stelling, Systems interface biology, *J. R. Soc., Interface*, 2006, **3**, 603–616.
170 P. Ao, Global view of bionetwork dynamics: adaptive landscape, *J. Genet. Genomics*, 2009, **36**, 63–73.
171 L. Giver, A. Gershenson, P. O. Freskgard and F. H. Arnold, Directed evolution of a thermostable esterase, *Proc. Natl. Acad. Sci. U. S. A.*, 1998, **95**, 12809–12813.
172 F. H. Arnold, P. L. Wintrode, K. Miyazaki and A. Gershenson, How enzymes adapt: lessons from directed evolution, *Trends Biochem. Sci.*, 2001, **26**, 100–106.
173 N. Tokuriki, F. Stricher, L. Serrano and D. S. Tawfik, How protein stability and new functions trade off, *PLoS Comput. Biol.*, 2008, **4**, e1000002.
174 J. D. Bloom, D. A. Drummond, F. H. Arnold and C. O. Wilke, Structural determinants of the rate of protein evolution in yeast, *Mol. Biol. Evol.*, 2006, **23**, 1751–1761.
175 J. D. Bloom, S. T. Labthavikul, C. R. Otey and F. H. Arnold, Protein stability promotes evolvability, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 5869–5874.
176 C. O. Wilke, J. D. Bloom, D. A. Drummond and A. Raval, Predicting the tolerance of proteins to random amino acid substitution, *Biophys. J.*, 2005, **89**, 3714–3720.
177 C. O. Wilke and D. A. Drummond, Signatures of protein biophysics in coding sequence evolution, *Curr. Opin. Struct. Biol.*, 2010, **20**, 385–389.
178 U. Bastolla, M. Porto, H. E. Roman and M. Vendruscolo, Looking at structure, stability, and evolution of proteins through the principal eigenvector of contact matrices and hydrophobicity profiles, *Gene*, 2005, **347**, 219–230.
179 C. L. Worth and T. L. Blundell, Satisfaction of hydrogen-bonding potential influences the conservation of polar sidechains, *Proteins*, 2009, **75**, 413–429.
180 J. D. Bloom, A. Raval and C. O. Wilke, Thermodynamics of neutral protein evolution, *Genetics*, 2007, **175**, 255–266.
181 J. D. Bloom, P. A. Romero, Z. Lu and F. H. Arnold, Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution, *Biol. Direct*, 2007, **2**, 17.
182 R. D. Gupta and D. S. Tawfik, Directed enzyme evolution *via* small and effective neutral drift libraries, *Nat. Methods*, 2008, **5**, 939–942.
183 D. L. Hartl, D. E. Dykhuizen and A. M. Dean, Limits of adaptation: the evolution of selective neutrality, *Genetics*, 1985, **111**, 655–674.
184 M. A. Huynen, P. F. Stadler and W. Fontana, Smoothness within ruggedness: The role of neutrality in adaptation, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 397–401.
185 C. M. Reidys and P. F. Stadler, Neutrality in fitness landscapes, *Appl. Math. Comput.*, 2001, **117**, 321–350.
186 G. Amitai, R. D. Gupta and D. S. Tawfik, Latent evolutionary potentials under the neutral mutational drift of an enzyme, *HFSP J.*, 2007, **1**, 67–78.
187 J. Noirel and T. Simonson, Neutral evolution of proteins: The superfunnel in sequence space and its relation to mutational robustness, *J. Chem. Phys.*, 2008, **129**, 185104.
188 A. Wagner, Neutralism and selectionism: a network-based reconciliation, *Nat. Rev. Genet.*, 2008, **9**, 965–974.
189 S. G. Peisajovich, L. Rockah and D. S. Tawfik, Evolution of new protein topologies through multistep gene rearrangements, *Nat. Genet.*, 2006, **38**, 168–174.
190 L. Pritchard, P. Bladon, J. M. O. Mitchell and M. J. Dufton, Evaluation of a novel method for the identification of coevolving protein residues, *Protein Eng.*, 2001, **14**, 549–555.
191 S. Govindarajan, J. E. Ness, S. Kim, E. C. Mundorff, J. Minshull and C. Gustafsson, Systematic variation of

amino acid substitutions for stringent assessment of pairwise covariation, *J. Mol. Biol.*, 2003, **328**, 1061–1069.
192 A. F. Neuwald, Surveying the Manifold Divergence of an Entire Protein Class for Statistical Clues to Underlying Biochemical Mechanisms, *Stat. Appl. Genet. Mol. Biol.*, 2011, **10**, 36.
193 L. Pritchard and M. J. Dufton, Do proteins learn to evolve? The Hopfield network as a basis for the understanding of protein evolution, *J. Theor. Biol.*, 2000, **202**, 77–86.
194 V. Chelliah, L. Chen, T. L. Blundell and S. C. Lovell, Distinguishing structural and functional restraints in evolution in order to identify interaction sites, *J. Mol. Biol.*, 2004, **342**, 1487–1504.
195 V. Chelliah and T. L. Blundell, Quantifying structural and functional restraints on amino acid substitutions in evolution of proteins, *Biochemistry*, 2005, **70**, 835–840.
196 V. Chelliah, T. Blundell and K. Mizuguchi, Functional restraints on the patterns of amino acid substitutions: application to sequence-structure homology recognition, *Proteins*, 2005, **61**, 722–731.
197 D. S. Marks, L. J. Colwell, R. Sheridan, T. A. Hopf, A. Pagnani, R. Zecchina and C. Sander, Protein 3D structure computed from evolutionary sequence variation, *PLoS One*, 2011, **6**, e28766.
198 T. A. Hopf, L. J. Colwell, R. Sheridan, B. Rost, C. Sander and D. S. Marks, Three-dimensional structures of membrane proteins from genomic sequencing, *Cell*, 2012, **149**, 1607–1621.
199 F. Morcos, A. Pagnani, B. Lunt, A. Bertolino, D. S. Marks, C. Sander, R. Zecchina, J. N. Onuchic, T. Hwa and M. Weigt, Direct-coupling analysis of residue coevolution captures native contacts across many protein families, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, E1293–E1301.
200 R. N. McLaughlin Jr, F. J. Poelwijk, A. Raman, W. S. Gosal and R. Ranganathan, The spatial architecture of protein function and adaptation, *Nature*, 2012, **491**, 138–142.
201 P. E. Tomatis, S. M. Fabiane, F. Simona, P. Carloni, B. J. Sutton and A. J. Vila, Adaptive protein evolution grants organismal fitness by improving catalysis and flexibility, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 20605–20610.
202 M. I. Sadowski and D. T. Jones, An automatic method for assessing structural importance of amino acid positions, *BMC Struct. Biol.*, 2009, **9**, 10.
203 L. A. Abriata, M. S. ML and P. E. Tomatis, Sequence-function-stability relationships in proteins from datasets of functionally annotated variants: The case of TEM beta-lactamases, *FEBS Lett.*, 2012, **586**, 3330–3335.
204 L. Burger and E. van Nimwegen, Accurate prediction of protein-protein interactions from sequence alignments using a Bayesian method, *Mol. Syst. Biol.*, 2008, **4**, 165.
205 M. Weigt, R. A. White, H. Szurmant, J. A. Hoch and T. Hwa, Identification of direct residue contacts in protein-protein interaction by message passing, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 67–72.
206 L. Burger and E. van Nimwegen, Disentangling Direct from Indirect Co-Evolution of Residues in Protein Alignments, *PLoS Comput. Biol.*, 2010, **6**, e1000633.
207 A. Rausell, D. Juan, F. Pazos and A. Valencia, Protein interactions and ligand binding: from protein subfamilies to functional specificity, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 1995–2000.
208 J. Strafford, P. Payongsri, E. G. Hibbert, P. Morris, S. S. Batth, D. Steadman, M. E. B. Smith, J. M. Ward, H. C. Hailes and P. A. Dalby, Directed evolution to re-adapt a co-evolved network within an enzyme, *J. Biotechnol.*, 2012, **157**, 237–245.
209 D. de Juan, F. Pazos and A. Valencia, Emerging methods in protein co-evolution, *Nat. Rev. Genet.*, 2013, **14**, 249–261.
210 J. H. Holland, *Adaption in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*, MIT Press, 1992.
211 D. E. Goldberg, *Genetic algorithms in search, optimization and machine learning*, Addison-Wesley, 1989.
212 D. E. Goldberg, *The design of innovation: lessons from and for competent genetic algorithms*, Kluwer, Boston, 2002.
213 J. Knowles, Closed-Loop Evolutionary Multiobjective Optimization, *IEEE Computational Intelligence Magazine*, 2009, **4**, 77–91.
214 D. B. Fogel, *Evolutionary computation: toward a new philosophy of machine intelligence*, IEEE Press, Piscataway, 1995.
215 *Evolutionary computation in bioinformatics*, ed. G. B. Fogel and D. W. Corne, Morgan Kaufmann, Amsterdam, 2003.
216 D. Ashlock, *Evolutionary computation for modeling and optimization*, Springer, New York, 2006.
217 N. Hamamatsu, T. Aita, Y. Nomiya, H. Uchiyama, M. Nakajima, Y. Husimi and Y. Shibanaka, Biased mutation-assembling: an efficient method for rapid directed evolution through simultaneous mutation accumulation, *Protein Eng., Des. Sel.*, 2005, **18**, 265–271.
218 N. Hamamatsu, Y. Nomiya, T. Aita, M. Nakajima, Y. Husimi and Y. Shibanaka, Directed evolution by accumulating tailored mutations: thermostabilization of lactate oxidase with less trade-off with catalytic activity, *Protein Eng., Des. Sel.*, 2006, **19**, 483–489.
219 R. J. Fox, S. C. Davis, E. C. Mundorff, L. M. Newman, V. Gavrilovic, S. K. Ma, L. M. Chung, C. Ching, S. Tam, S. Muley, J. Grate, J. Gruber, J. C. Whitman, R. A. Sheldon and G. W. Huisman, Improving catalytic function by ProSAR-driven enzyme evolution, *Nat. Biotechnol.*, 2007, **25**, 338–344.
220 D. Wedge, W. Rowe, D. B. Kell and J. Knowles, *In silico* modelling of directed evolution: implications for experimental design and stepwise evolution, *J. Theor. Biol.*, 2009, **257**, 131–141.
221 M. Carneiro and D. L. Hartl, Adaptive landscapes and protein evolution, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **107**(suppl 1), 1747–1751.
222 J. H. Gillespie, A simple stochastic gene substitution model, *Theor. Popul. Biol.*, 1983, **23**, 202–215.
223 J. H. Gillespie, Molecular Evolution over the Mutational Landscape, *Evolution*, 1984, **38**, 1116–1129.
224 H. A. Orr, The genetic theory of adaptation: a brief history, *Nat. Rev. Genet.*, 2005, **6**, 119–127.

225 H. A. Orr, The population genetics of adaptation on correlated fitness landscapes: the block model, *Evolution*, 2006, **60**, 1113–1124.
226 H. A. Orr, The distribution of fitness effects among beneficial mutations in Fisher's geometric model of adaptation, *J. Theor. Biol.*, 2006, **238**, 279–285.
227 H. A. Orr, Fitness and its role in evolutionary genetics, *Nat. Rev. Genet.*, 2009, **10**, 531–539.
228 R. L. Unckless and H. A. Orr, The population genetics of adaptation: multiple substitutions on a smooth fitness landscape, *Genetics*, 2009, **183**, 1079–1086.
229 I. G. Szendro, J. Franke, J. A. G. M. de Visser and J. Krug, Predictability of evolution depends nonmonotonically on population size, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 571–576.
230 J. A. Wells, Additivity of mutational effects in proteins, *Biochemistry*, 1990, **29**, 8509–8517.
231 M. Lunzer, S. P. Miller, R. Felsheim and A. M. Dean, The biochemical architecture of an ancient adaptive landscape, *Science*, 2005, **310**, 499–501.
232 R. Fox, A. Roy, S. Govindarajan, J. Minshull, C. Gustafsson, J. T. Jones and R. Emig, Optimizing the search algorithm for protein engineering by directed evolution, *Protein Eng.*, 2003, **16**, 589–597.
233 R. Fox, Directed molecular evolution by machine learning and the influence of nonlinear interactions, *J. Theor. Biol.*, 2005, **234**, 187–199.
234 M. Iwakura, K. Maki, H. Takahashi, T. Takenawa, A. Yokota, K. Katayanagi, T. Kamiyama and K. Gekko, Evolutional design of a hyperactive cysteine- and methionine-free mutant of Escherichia coli dihydrofolate reductase, *J. Biol. Chem.*, 2006, **281**, 13234–13246.
235 C. A. Tracewell and F. H. Arnold, Directed enzyme evolution: climbing fitness peaks one amino acid at a time, *Curr. Opin. Chem. Biol.*, 2009, **13**, 3–9.
236 M. T. Reetz, The importance of additive and non-additive mutational effects in protein engineering, *Angew. Chem., Int. Ed.*, 2013, **52**, 2658–2666.
237 Y. Hayashi, T. Aita, H. Toyota, Y. Husimi, I. Urabe and T. Yomo, Experimental rugged fitness landscape in protein sequence space, *PLoS One*, 2006, **1**, e96.
238 J. D. Bloom, F. H. Arnold and C. O. Wilke, Breaking proteins with mutations: threads and thresholds in evolution, *Mol. Syst. Biol.*, 2007, **3**, 76.
239 K. Jain and S. Seetharaman, Multiple adaptive substitutions during evolution in novel environments, *Genetics*, 2011, **189**, 1029–1043.
240 J. D. Bloom, L. I. Gong and D. Baltimore, Permissive secondary mutations enable the evolution of influenza oseltamivir resistance, *Science*, 2010, **328**, 1272–1275.
241 T. Aita, N. Hamamatsu, Y. Nomiya, H. Uchiyama, Y. Shibanaka and Y. Husimi, Surveying a local fitness landscape of a protein with epistatic sites for the study of directed evolution, *Biopolymers*, 2002, **64**, 95–105.
242 B. Østman, A. Hintze and C. Adami, Impact of epistasis and pleiotropy on evolutionary adaptation, *Proc. R. Soc. B*, 2011, **279**, 247–256.
243 M. S. Breen, C. Kemena, P. K. Vlasov, C. Notredame and F. A. Kondrashov, Epistasis as the primary factor in molecular evolution, *Nature*, 2012, **490**, 535–538.
244 C. Natarajan, N. Inoguchi, R. E. Weber, A. Fago, H. Moriyama and J. F. Storz, Epistasis among adaptive mutations in deer mouse hemoglobin, *Science*, 2013, **340**, 1324–1327.
245 J. L. Rummer, D. J. McKenzie, A. Innocenti, C. T. Supuran and C. J. Brauner, Root effect hemoglobin may have evolved to enhance general tissue oxygen delivery, *Science*, 2013, **340**, 1327–1329.
246 S. Mirceta, A. V. Signore, J. M. Burns, A. R. Cossins, K. L. Campbell and M. Berenbrink, Evolution of mammalian diving capacity traced by myoglobin net surface charge, *Science*, 2013, **340**, 1234192.
247 P. A. Alexander, Y. He, Y. Chen, J. Orban and P. N. Bryan, A minimal sequence code for switching protein structure and function, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 21149–21154.
248 L. I. Gong and J. D. Bloom, Epistatically Interacting Substitutions Are Enriched during Adaptive Protein Evolution, *PLoS Genet.*, 2014, **10**, e1004328.
249 M. Lunzer, G. B. Golding and A. M. Dean, Pervasive cryptic epistasis in molecular evolution, *PLoS Genet.*, 2010, **6**, e1001162.
250 S. G. Williams and S. C. Lovell, The effect of sequence evolution on protein structural divergence, *Mol. Biol. Evol.*, 2009, **26**, 1055–1065.
251 Y. Yoshikuni, T. E. Ferrin and J. D. Keasling, Designed divergent evolution of enzyme function, *Nature*, 2006, **440**, 1078–1082.
252 K. Hult and P. Berglund, Enzyme promiscuity: mechanism and applications, *Trends Biotechnol.*, 2007, **25**, 231–238.
253 I. Nobeli, A. D. Favia and J. M. Thornton, Protein promiscuity and its implications for biotechnology, *Nat. Biotechnol.*, 2009, **27**, 157–167.
254 N. Tokuriki and D. S. Tawfik, Protein dynamism and evolvability, *Science*, 2009, **324**, 203–207.
255 A. Babtie, N. Tokuriki and F. Hollfelder, What makes an enzyme promiscuous? *Curr. Opin. Chem. Biol.*, 2010, **14**, 200–207.
256 O. Khersonsky and D. S. Tawfik, Enzyme promiscuity: a mechanistic and evolutionary perspective, *Annu. Rev. Biochem.*, 2010, **79**, 471–505.
257 O. Khersonsky and D. S. Tawfik, Enzyme promiscuity: evolutionary and mechanistic aspects, in *Comprehensive Natural Products II Chemistry and Biology*, ed. L. Mander and H. -W. Lui, Elsevier, Oxford, 2010, pp. 48–90.
258 A. Aharoni, L. Gaidukov, O. Khersonsky, Q. G. S. Mc, C. Roodveldt and D. S. Tawfik, The 'evolvability' of promiscuous protein functions, *Nat. Genet.*, 2005, **37**, 73–76.
259 M. S. Humble and P. Berglund, Biocatalytic Promiscuity, *Eur. J. Org. Chem.*, 2011, 3391–3401.

260 R. Huang, F. Hippauf, D. Rohrbeck, M. Haustein, K. Wenke, J. Feike, N. Sorrelle, B. Piechulla and T. J. Barkman, Enzyme functional evolution through improved catalysis of ancestrally nonpreferred substrates, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 2966–2971.

261 D. R. Burton, P. Poignard, R. L. Stanfield and I. A. Wilson, Broadly neutralizing antibodies present new prospects to counter highly antigenically diverse viruses, *Science*, 2012, **337**, 183–186.

262 H. Garcia-Seisdedos, B. Ibarra-Molero and J. M. Sanchez-Ruiz, Probing the mutational interplay between primary and promiscuous protein functions: a computational-experimental approach, *PLoS Comput. Biol.*, 2012, **8**, e1002558.

263 H. Nam, N. E. Lewis, J. A. Lerman, D. H. Lee, R. L. Chang, D. Kim and B. Ø. Palsson, Network context and selection in the evolution to enzyme specificity, *Science*, 2012, **337**, 1101–1104.

264 J. H. Luo, B. van Loo and S. C. L. Kamerlin, Catalytic promiscuity in *Pseudomonas aeruginosa* arylsulfatase as an example of chemistry-driven protein evolution, *FEBS Lett.*, 2012, **586**, 1622–1630.

265 S. Chakraborty, An automated flow for directed evolution based on detection of promiscuous scaffolds using spatial and electrostatic properties of catalytic residues, *PLoS One*, 2012, **7**, e40408.

266 P. Gatti-Lafranconi and F. Hollfelder, Flexibility and reactivity in promiscuous enzymes, *ChemBioChem*, 2013, **14**, 285–292.

267 P. Carbonell, G. Lecointre and J. L. Faulon, Origins of specificity and promiscuity in metabolic networks, *J. Biol. Chem.*, 2011, **286**, 43994–44004.

268 P. Carbonell and J. L. Faulon, Molecular signatures-based prediction of enzyme promiscuity, *Bioinformatics*, 2010, **26**, 2012–2019.

269 D. B. Kell, P. D. Dobson, E. Bilsland and S. G. Oliver, The promiscuous binding of pharmaceutical drugs and their transporter-mediated uptake into cells: what we (need to) know and how we can do so, *Drug Discovery Today*, 2013, **18**, 218–239.

270 S. Chakraborty, R. Minda, L. Salaye, A. M. Dandekar, S. K. Bhattacharjee and B. J. Rao, Promiscuity-based enzyme selection for rational directed evolution experiments, *Methods Mol. Biol.*, 2013, **978**, 205–216.

271 S. A. Kauffman and E. D. Weinberger, The NK model of rugged fitness landscapes and its application to maturation of the immune response, *J. Theor. Biol.*, 1989, **141**, 211–245.

272 L. Barnett, Ruggedness and neutrality: the NKp family of fitness landscapes, *Proc. 6th Int'l Conf. on Artificial Life*, MIT Press, 1998, pp. 17–27.

273 T. Aita, Hierarchical distribution of ascending slopes, nearly neutral networks, highlands, and local optima at the dth order in an NK fitness landscape, *J. Theor. Biol.*, 2008, **254**, 252–263.

274 T. Aita and Y. Husimi, Fitting protein-folding free energy landscape for a certain conformation to an NK fitness landscape, *J. Theor. Biol.*, 2008, **253**, 151–161.

275 B. Østman, A. Hintze and C. Adami, Critical Properties of Complex Fitness Landscapes, *Proc XII Conf Alife*, 2010, 126–132.

276 W. Rowe, D. C. Wedge, M. Platt, D. B. Kell and J. Knowles, Predictive models for population performance on real biological fitness landscapes, *Bioinformatics*, 2010, **26**, 2125–2142.

277 N. A. Rosenberg, A sharp minimum on the mean number of steps taken in adaptive walks, *J. Theor. Biol.*, 2005, **237**, 17–22.

278 D. B. Kell and E. Lurie-Luke, The Virtue of Innovation: Innovation through the Lenses of Biological Evolution, *J. R. Soc., Interface*, 2015, in the press.

279 C. R. Reeves and J. E. Rowe, *Genetic algorithms – principles and perspectives: a guide to GA theory*, Kluwer Academic Publishers, Dordrecht, 2002.

280 T. Aita and Y. Husimi, Adaptive walks by the fittest among finite random mutants on a Mt. Fuji-type fitness landscape - II. Effect of small non- additivity, *J. Math. Biol.*, 2000, **41**, 207–231.

281 T. Aita, H. Uchiyama, T. Inaoka, M. Nakajima, T. Kokubo and Y. Husimi, Analysis of a local fitness landscape with a model of the rough Mt. Fuji-type landscape: Application to prolyl endopeptidase and thermolysin, *Biopolymers*, 2000, **54**, 64–79.

282 E. D. Weinberger, *NP completeness of Kauffman's NK model: a tuneably rugged fitness landscape*, Santa Fe Institute Technical Report, 1996, 96-02-003.

283 N. Tokuriki, C. J. Jackson, L. Afriat-Jurnou, K. T. Wyganowski, R. Tang and D. S. Tawfik, Diminishing returns and tradeoffs constrain the laboratory optimization of an enzyme, *Nat. Commun.*, 2012, **3**, 1257.

284 D. T. Campbell, Blind variation and selective retention in creative thought as in other knowledge processes, *Psychol. Rev.*, 1960, **67**, 380–400.

285 J. W. Rivkin and N. Siggelkow, Organisational sticking points on NK landscapes, *Complexity*, 2002, **7**, 31–43.

286 K. Frenken, Technological innovation and complexity theory, *Econ. Innov. New Technol.*, 2006, **15**, 137–155.

287 S. Geisendorf, Searching NK Fitness Landscapes: On the Trade Off Between Speed and Quality in Complex Problem Solving, *Comput. Econ.*, 2010, **35**, 395–406.

288 S. Johnson, *Where good ideas come from: the seven patterns of innovation*, Penguin, London, 2011.

289 P. Auerswald, S. Kauffman, J. Lobo and K. Shell, The production recipes approach to modeling technological innovation: An application to learning by doing, *J. Econ. Dyn. Control*, 2000, **24**, 389–450.

290 S. Kauffman, J. Lobo and W. G. Macready, Optimal search on a technology landscape, *J. Econ. Behav. Organ.*, 2000, **43**, 141–166.

291 M. Ganco and G. Hoetker, NK Modeling Methodology in the Strategy Literature: Bounded Search on a Rugged Landscape, *Res Methodol Strat Mangement*, 2009, **5**, 237–268.

292 G. M. Hodgson and T. Knudsen, Balancing inertia, innovation, and imitation in complex environments, *J. Econ. Issues*, 2006, **40**, 287–295.

293 L. Gabora, An evolutionary framework for cultural change: Selectionism *versus* communal exchange, *Phys. Life Rev.*, 2013, **10**, 117–145.
294 R. V. Solé, S. Valverde, M. R. Casals, S. A. Kauffman, D. Farmer and N. Eldredge, The Evolutionary Ecology of Technological Innovations, *Complexity*, 2013, **18**, 15–27.
295 A. Wagner and W. Rosen, Spaces of the possible: universal Darwinism and the wall between technological and biological innovation, *J. R. Soc., Interface*, 2014, **11**, 20131190.
296 *Directed evolution library creation: methods and protocols*, ed. F. H. Arnold and G. Georgiou, Springer, Berlin, 1996.
297 *Directed molecular evolution of proteins*, ed. S. Brakmann and K. Johnsoon, Wiley-VCH, Weinheim, 2002.
298 C. A. Voigt, S. Kauffman and Z. G. Wang, Rational evolutionary design: The theory of *in vitro* protein evolution, in *Adv. Protein Chem.*, ed. F. M. Arnold, 2001, vol. 55, pp. 79–160.
299 F. H. Arnold and A. A. Volkov, Directed evolution of biocatalysts, *Curr. Opin. Biotechnol.*, 1999, **3**, 54–59.
300 *Evolutionary protein design*, ed. F. H. Arnold, Academic Press, San Diego, 2001.
301 K. A. Powell, S. W. Ramer, S. B. Del Cardayré, W. P. C. Stemmer, M. B. Tobin, P. F. Longchamp and G. W. Huisman, Directed Evolution and Biocatalysis, *Angew. Chem., Int. Ed.*, 2001, **40**, 3948–3959.
302 C. Schmidt-Dannert, Directed evolution of single proteins, metabolic pathways, and viruses, *Biochemistry*, 2001, **40**, 13125–13136.
303 S. V. Taylor, P. Kast and D. Hilvert, Investigating and engineering enzymes by genetic selection, *Angew. Chem., Int. Ed.*, 2001, **40**, 3311–3335.
304 H. Tao and V. W. Cornish, Milestones in directed enzyme evolution, *Curr. Opin. Chem. Biol.*, 2002, **6**, 858–864.
305 P. A. Dalby, Optimising enzyme function by directed evolution, *Curr. Opin. Struct. Biol.*, 2003, **13**, 500–505.
306 N. J. Turner, Directed evolution of enzymes for applied biocatalysis, *Trends Biotechnol.*, 2003, **21**, 474–478.
307 C. Neylon, Chemical and biochemical strategies for the randomization of protein encoding DNA sequences: library construction methods for directed evolution, *Nucleic Acids Res.*, 2004, **32**, 1448–1459.
308 H. Leemhuis, V. Stein, A. D. Griffiths and F. Hollfelder, New genotype-phenotype linkages for directed evolution of functional proteins, *Curr. Opin. Struct. Biol.*, 2005, **15**, 472–478.
309 L. Yuan, I. Kurek, J. English and R. Keenan, Laboratory-directed protein evolution, *Microbiol. Mol. Biol. Rev.*, 2005, **69**, 373–392.
310 T. Matsuura and T. Yomo, *In vitro* evolution of proteins, *J. Biosci. Bioeng.*, 2006, **101**, 449–456.
311 P. A. Dalby, Engineering enzymes for biocatalysis, *Recent Pat. Biotechnol.*, 2007, **1**, 1–9.
312 S. Sen, V. Venkata Dasu and B. Mandal, Developments in directed evolution for improving enzyme functions, *Appl. Biochem. Biotechnol.*, 2007, **143**, 212–223.
313 C. C. Akoh, S. W. Chang, G. C. Lee and J. F. Shaw, Biocatalysis for the production of industrial products and functional foods from rice and other agricultural produce, *J. Agric. Food Chem.*, 2008, **56**, 10445–10451.
314 J. C. Chaput, N. W. Woodbury, L. A. Stearns and B. A. Williams, Creating protein biocatalysts as tools for future industrial applications, *Expert Opin. Biol. Ther.*, 2008, **8**, 1087–1098.
315 R. N. Patel, Synthesis of chiral pharmaceutical intermediates by biocatalysis, *Coord. Chem. Rev.*, 2008, **252**, 659–701.
316 C. Jäckel, P. Kast and D. Hilvert, Protein design by directed evolution, *Annu. Rev. Biophys.*, 2008, **37**, 153–173.
317 F. H. Arnold, How proteins adapt: lessons from directed evolution, *Cold Spring Harbor Symp. Quant. Biol.*, 2009, **74**, 41–46.
318 S. C. Stebel, A. Gaida, K. M. Arndt and K. M. Müller, Directed Protein Evolution, in *Molecular Biomethods Handbook*, ed. J. M. Walker and R. Rapley, Humana Press, Totowa, NJ, 2nd edn, 2008, pp. 631–656.
319 N. J. Turner, Directed evolution drives the next generation of biocatalysts, *Nat. Chem. Biol.*, 2009, **5**, 567–573.
320 C. Jäckel and D. Hilvert, Biocatalysts by evolution, *Curr. Opin. Biotechnol.*, 2010, **21**, 753–759.
321 P. A. Dalby, Strategy and success for the directed evolution of enzymes, *Curr. Opin. Struct. Biol.*, 2011, **21**, 473–480.
322 N. J. Turner and M. D. Truppo, Biocatalysis enters a new era, *Curr. Opin. Chem. Biol.*, 2013, **17**, 212–214.
323 M. T. Reetz, J. D. Carballeira, J. Peyralans, H. Hobenreich, A. Maichele and A. Vogel, Expanding the substrate scope of enzymes: combining mutations obtained by CASTing, *Chemistry*, 2006, **12**, 6031–6038.
324 M. T. Reetz, D. Kahakeaw and R. Lohmer, Addressing the numbers problem in directed evolution, *ChemBioChem*, 2008, **9**, 1797–1804.
325 G. A. Behrens, A. Hummel, S. K. Padhi, S. Schätzle and U. T. Bornscheuer, Discovery and Protein Engineering of Biocatalysts for Organic Synthesis, *Adv. Synth. Catal.*, 2011, **353**, 2191–2215.
326 M. T. Reetz, Laboratory evolution of stereoselective enzymes: a prolific source of catalysts for asymmetric reactions, *Angew. Chem., Int. Ed.*, 2011, **50**, 138–174.
327 G. A. Strohmeier, H. Pichler, O. May and M. Gruber-Khadjawi, Application of designed enzymes in organic synthesis, *Chem. Rev.*, 2011, **111**, 4141–4164.
328 U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore and K. Robins, Engineering the third wave of biocatalysis, *Nature*, 2012, **485**, 185–194.
329 M. Goldsmith and D. S. Tawfik, Directed enzyme evolution: beyond the low-hanging fruit, *Curr. Opin. Struct. Biol.*, 2012, **22**, 406–412.
330 R. Verma, U. Schwaneberg and D. Roccatano, Computer-Aided Protein Directed Evolution: a Review of Web Servers, Databases and other Computational Tools for Protein Engineering, *Comput. Struct. Biotechnol. J.*, 2012, **2**, e201209008.
331 T. Davids, M. Schmidt, D. Bottcher and U. T. Bornscheuer, Strategies for the discovery and engineering of enzymes for biocatalysis, *Curr. Opin. Chem. Biol.*, 2013, **17**, 215–220.

332 A. Kumar and S. Singh, Directed evolution: tailoring biocatalysts for industrial applications, *Crit. Rev. Biotechnol.*, 2013, **33**, 365–378.
333 M. T. Reetz, Biocatalysis in organic chemistry and biotechnology: past, present, and future, *J. Am. Chem. Soc.*, 2013, **135**, 12480–12496.
334 J. Damborsky and J. Brezovsky, Computational tools for designing and engineering enzymes, *Curr. Opin. Chem. Biol.*, 2014, **19**, 8–16.
335 P. D. Dobson, Y. Patel and D. B. Kell, ''Metabolite-likeness'' as a criterion in the design and selection of pharmaceutical drug libraries, *Drug Discovery Today*, 2009, **14**, 31–40.
336 S. O'Hagan, N. Swainston, J. Handl and D. B. Kell, A 'rule of 0.5' for the metabolite-likeness of approved pharmaceutical drugs, *Metabolomics*, 2015, DOI: 10.1007/s11306-11014-10733-z.
337 E. Akiva, S. Brown, D. E. Almonacid, A. E. Barber, 2nd, A. F. Custer, M. A. Hicks, C. C. Huang, F. Lauck, S. T. Mashiyama, E. C. Meng, D. Mischel, J. H. Morris, S. Ojha, A. M. Schnoes, D. Stryke, J. M. Yunes, T. E. Ferrin, G. L. Holliday and P. C. Babbitt, The structure-function linkage database, *Nucleic Acids Res.*, 2014, **42**, D521–D530.
338 K. A. Armstrong and B. Tidor, Computationally mapping sequence space to understand evolutionary protein engineering, *Biotechnol. Prog.*, 2008, **24**, 62–73.
339 Y. Nov, Probabilistic methods in directed evolution: library size, mutation rate, and diversity, *Methods Mol. Biol.*, 2014, **1179**, 261–278.
340 J. Zaugg, Y. Gumulya, E. M. Gillam and M. Boden, Computational tools for directed evolution: a comparison of prospective and retrospective strategies, *Methods Mol. Biol.*, 2014, **1179**, 315–333.
341 A. Pavelka, E. Chovancova and J. Damborsky, Hot Spot Wizard: a web server for identification of hot spots in protein engineering, *Nucleic Acids Res.*, 2009, **37**, W376–W383.
342 M. Höhne, S. Schätzle, H. Jochens, K. Robins and U. T. Bornscheuer, Rational assignment of key motifs for function guides *in silico* enzyme identification, *Nat. Chem. Biol.*, 2010, **6**, 807–813.
343 H. Jochens and U. T. Bornscheuer, Natural Diversity to Guide Focused Directed Evolution, *ChemBioChem*, 2010, **11**, 1861–1866.
344 F. Sievers, A. Wilm, D. Dineen, T. J. Gibson, K. Karplus, W. Li, R. Lopez, H. McWilliam, M. Remmert, J. Soding, J. D. Thompson and D. G. Higgins, Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega, *Mol. Syst. Biol.*, 2011, **7**, 539.
345 F. Sievers and D. G. Higgins, Clustal Omega, accurate alignment of very large numbers of sequences, *Methods Mol. Biol.*, 2014, **1079**, 105–116.
346 R. C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, *Nucleic Acids Res.*, 2004, **32**, 1792–1797.
347 J. Pei, B. H. Kim, M. Tang and N. V. Grishin, PROMALS web server for accurate multiple protein sequence alignments, *Nucleic Acids Res.*, 2007, **35**, W649–W652.
348 J. Pei and N. V. Grishin, PROMALS3D: multiple protein sequence alignment enhanced with evolutionary and three-dimensional structural information, *Methods Mol. Biol.*, 2014, **1079**, 263–271.
349 L. Xie, L. Xie and P. E. Bourne, A unified statistical model to support local sequence order independent similarity searching for ligand-binding sites and its application to genome-based drug discovery, *Bioinformatics*, 2009, **25**, i305–312.
350 L. Xie, L. Xie and P. E. Bourne, Structure-based systems biology for analyzing off-target binding, *Curr. Opin. Struct. Biol.*, 2011, **21**, 189–199.
351 T. Uchiyama and K. Miyazaki, Functional metagenomics for enzyme discovery: challenges to efficient screening, *Curr. Opin. Biotechnol.*, 2009, **20**, 616–622.
352 M. M. Schofield and D. H. Sherman, Meta-omic characterization of prokaryotic gene clusters for natural product biosynthesis, *Curr. Opin. Biotechnol.*, 2013, **24**, 1151–1158.
353 P. Medawar, *Pluto's republic*, Oxford University Press, Oxford, 1982.
354 K. R. Popper, *Conjectures and refutations: the growth of scientific knowledge*, Routledge & Kegan Paul, London, 5th edn, 1992.
355 A. F. Chalmers, *What is this thing called Science? An assessment of the nature and status of science and its methods*, Open University Press, Maidenhead, 1999.
356 D. B. Kell and S. G. Oliver, How drugs get into cells: tested and testable predictions to help discriminate between transporter-mediated uptake and lipoidal bilayer diffusion, *Front. Pharmacol.*, 2014, **5**, 231.
357 D. B. Kell, What would be the observable consequences if phospholipid bilayer diffusion of drugs into cells is negligible? *Trends Pharmacol. Sci.*, 2015, DOI: 10.1016/j.tips.2014.10.005.
358 D. B. Kell and S. G. Oliver, Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era., *BioEssays*, 2004, **26**, 99–105.
359 D. B. Kell, Metabolomics and systems biology: making sense of the soup, *Curr. Opin. Microbiol.*, 2004, **7**, 296–307.
360 D. B. Kell and J. D. Knowles, The role of modeling in systems biology, in *System modeling in cellular biology: from concepts to nuts and bolts*, ed. Z. Szallasi, J. Stelling and V. Periwal, MIT Press, Cambridge, 2006, pp. 3–18.
361 D. B. Kell, Metabolomics, modelling and machine learning in systems biology: towards an understanding of the languages of cells. The 2005 Theodor Bücher lecture, *FEBS J.*, 2006, **273**, 873–894.
362 D. B. Kell, Finding novel pharmaceuticals in the systems biology era using multiple effective drug targets, phenotypic screening, and knowledge of transporters: where drug discovery went wrong and how to fix it, *FEBS J.*, 2013, **280**, 5957–5980.
363 L. R. Franklin, Exploratory experiments, *Philos. Sci.*, 2005, **72**, 888–899.
364 K. C. Elliott, Epistemic and methodological iteration in scientific research, *Stud. Hist. Philos Sci.*, 2012, **43**, 376–382.

365 S. A. Doyle, S. Y. Fung and D. E. Koshland, Jr., Redesigning the substrate specificity of an enzyme: isocitrate dehydrogenase, *Biochemistry*, 2000, **39**, 14348–14355.

366 Q. S. Li, U. Schwaneberg, M. Fischer, J. Schmitt, J. Pleiss, S. Lutz-Wahl and R. D. Schmid, Rational evolution of a medium chain-specific cytochrome P-450 BM-3 variant, *Biochim. Biophys. Acta*, 2001, **1545**, 114–121.

367 C. Gustafsson, S. Govindarajan and J. Minshull, Putting engineering back into protein engineering: bioinformatic approaches to catalyst design, *Curr. Opin. Biotechnol.*, 2003, **14**, 366–370.

368 G. Jiménez-Osés, S. Osuna, X. Gao, M. R. Sawaya, L. Gilson, S. J. Collier, G. W. Huisman, T. O. Yeates, Y. Tang and K. N. Houk, The role of distant mutations and allosteric regulation on LovD active site dynamics, *Nat. Chem. Biol.*, 2014, **10**, 431–436.

369 P. Tian, Computational protein design, from single domain soluble proteins to membrane proteins, *Chem. Soc. Rev.*, 2010, **39**, 2071–2082.

370 B. R. Lichtenstein, T. A. Farid, G. Kodali, L. A. Solomon, J. L. Anderson, M. M. Sheehan, N. M. Ennist, B. A. Fry, S. E. Chobot, C. Bialas, J. A. Mancini, C. T. Armstrong, Z. Zhao, T. V. Esipova, D. Snell, S. A. Vinogradov, B. M. Discher, C. C. Moser and P. L. Dutton, Engineering oxidoreductases: maquette proteins designed from scratch, *Biochem. Soc. Trans.*, 2012, **40**, 561–566.

371 T. A. Farid, G. Kodali, L. A. Solomon, B. R. Lichtenstein, M. M. Sheehan, B. A. Fry, C. Bialas, N. M. Ennist, J. A. Siedlecki, Z. Zhao, M. A. Stetz, K. G. Valentine, J. L. Anderson, A. J. Wand, B. M. Discher, C. C. Moser and P. L. Dutton, Elementary tetrahelical protein design for diverse oxidoreductase functions, *Nat. Chem. Biol.*, 2013, **9**, 826–833.

372 C. W. Wood, M. Bruning, A. A. Ibarra, G. J. Bartlett, A. R. Thomson, R. B. Sessions, R. L. Brady and D. N. Woolfson, CCBuilder: an interactive web-based tool for building, designing and assessing coiled-coil protein assemblies, *Bioinformatics*, 2014, **30**, 3029–3035.

373 A. R. Thomson, C. W. Wood, A. J. Burton, G. J. Bartlett, R. B. Sessions, R. L. Brady and D. N. Woolfson, Computational design of water-soluble alpha-helical barrels, *Science*, 2014, **346**, 485–488.

374 F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic and D. Baker, *De novo* enzyme design using Rosetta3, *PLoS One*, 2011, **6**, e19230.

375 L. Wang, E. A. Althoff, J. Bolduc, L. Jiang, J. Moody, J. K. Lassila, L. Giger, D. Hilvert, B. Stoddard and D. Baker, Structural analyses of covalent enzyme-substrate analog complexes reveal strengths and limitations of *de novo* enzyme design, *J. Mol. Biol.*, 2012, **415**, 615–625.

376 L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Rothlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, 3rd, D. Hilvert, K. N. Houk, B. L. Stoddard and D. Baker, *De novo* computational design of retro-aldol enzymes, *Science*, 2008, **319**, 1387–1391.

377 D. Röthlisberger, O. Khersonsky, A. M. Wollacott, L. Jiang, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D. S. Tawfik and D. Baker, Kemp elimination catalysts by computational enzyme design, *Nature*, 2008, **453**, 190–195.

378 J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael and D. Baker, Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction, *Science*, 2010, **329**, 309–313.

379 J. Sykora, J. Brezovsky, T. Koudelakova, M. Lahoda, A. Fortova, T. Chernovets, R. Chaloupkova, V. Stepankova, Z. Prokop, I. K. Smatanova, M. Hof and J. Damborsky, Dynamics and hydration explain failed functional transformation in dehalogenase design, *Nat. Chem. Biol.*, 2014, **10**, 428–430.

380 C. B. Eiben, J. B. Siegel, J. B. Bale, S. Cooper, F. Khatib, B. W. Shen, F. Players, B. L. Stoddard, Z. Popovic and D. Baker, Increased Diels-Alderase activity through backbone remodeling guided by Foldit players, *Nat. Biotechnol.*, 2012, **30**, 190–192.

381 Y. Kipnis and D. Baker, Comparison of designed and randomly generated catalysts for simple chemical reactions, *Protein Sci.*, 2012, **21**, 1388–1395.

382 Y. L. Boersma and A. Pluckthun, DARPins and other repeat protein scaffolds: advances in engineering and applications, *Curr. Opin. Biotechnol.*, 2011, **22**, 849–857.

383 W. J. Albery and J. R. Knowles, Evolution of enzyme function and the development of catalytic efficiency, *Biochemistry*, 1976, **15**, 5631–5640.

384 E. B. Nickbarg and J. R. Knowles, Triosephosphate isomerase: energetics of the reaction catalyzed by the yeast enzyme expressed in *Escherichia coli*, *Biochemistry*, 1988, **27**, 5939–5947.

385 F. Zárate-Pérez, M. E. Chánez-Cárdenas and E. Vázquez-Contreras, The folding pathway of triosephosphate isomerase, *Prog. Mol. Biol. Transl. Sci.*, 2008, **84**, 251–267.

386 B. J. Sullivan, V. Durani and T. J. Magliery, Triosephosphate isomerase by consensus design: dramatic differences in physical properties and activity of related variants, *J. Mol. Biol.*, 2011, **413**, 195–208.

387 M. Alahuhta, M. Salin, M. G. Casteleijn, C. Kemmer, I. El-Sayed, K. Augustyns, P. Neubauer and R. K. Wierenga, Structure-based protein engineering efforts with a monomeric TIM variant: the importance of a single point mutation for generating an active site with suitable binding properties, *Protein Eng., Des. Sel.*, 2008, **21**, 257–266.

388 T. V. Borchert, R. Abagyan, R. Jaenicke and R. K. Wierenga, Design, creation, and characterization of a stable, monomeric triosephosphate isomerase, *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 1515–1518.

389 M. Holmquist, Alpha/Beta-hydrolase fold enzymes: structures, functions and mechanisms, *Curr. Protein Pept. Sci.*, 2000, **1**, 209–235.

390 M. Henn-Sax, B. Hocker, M. Wilmanns and R. Sterner, Divergent evolution of (betaalpha)8-barrel enzymes, *Biol. Chem.*, 2001, **382**, 1315–1320.

391 R. K. Wierenga, The TIM-barrel fold: a versatile framework for efficient enzymes, *FEBS Lett.*, 2001, **492**, 193–198.

392 N. Nagano, C. A. Orengo and J. M. Thornton, One fold with many functions: the evolutionary relationships between TIM barrel families based on their sequences, structures and functions, *J. Mol. Biol.*, 2002, **321**, 741–765.

393 D. M. Z. Schmidt, E. C. Mundorff, M. Dojka, E. Bermudez, J. E. Ness, S. Govindarajan, P. C. Babbitt, J. Minshull and J. A. Gerlt, Evolutionary potential of (beta/alpha)8-barrels: functional promiscuity produced by single substitutions in the enolase superfamily, *Biochemistry*, 2003, **42**, 8387–8393.

394 J. E. Vick, D. M. Schmidt and J. A. Gerlt, Evolutionary potential of (beta/alpha)8-barrels: *in vitro* enhancement of a ''new'' reaction in the enolase superfamily, *Biochemistry*, 2005, **44**, 11722–11729.

395 H. S. Park, S. H. Nam, J. K. Lee, C. N. Yoon, B. Mannervik, S. J. Benkovic and H. S. Kim, Design and evolution of new catalytic activity with an existing protein scaffold, *Science*, 2006, **311**, 535–538.

396 J. E. Vick and J. A. Gerlt, Evolutionary potential of (beta/alpha)8-barrels: stepwise evolution of a ''new'' reaction in the enolase superfamily, *Biochemistry*, 2007, **46**, 14589–14597.

397 S. Leopoldseder, J. Claren, C. Jurgens and R. Sterner, Interconverting the catalytic activities of (betaalpha)(8)-barrel enzymes from different metabolic pathways: sequence requirements and molecular analysis, *J. Mol. Biol.*, 2004, **337**, 871–879.

398 R. Sterner and B. Hocker, Catalytic versatility, stability, and evolution of the (betaalpha)8-barrel enzyme fold, *Chem. Rev.*, 2005, **105**, 4038–4055.

399 T. Seitz, M. Bocola, J. Claren and R. Sterner, Stabilisation of a (betaalpha)8-barrel protein designed from identical half barrels, *J. Mol. Biol.*, 2007, **372**, 114–129.

400 J. Claren, C. Malisi, B. Höcker and R. Sterner, Establishing wild-type levels of catalytic activity on natural and artificial (beta alpha)8-barrel protein scaffolds, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 3704–3709.

401 B. Höcker, A. Lochner, T. Seitz, J. Claren and R. Sterner, High-resolution crystal structure of an artificial (betaalpha)(8)-barrel protein designed from identical half-barrels, *Biochemistry*, 2009, **48**, 1145–1147.

402 X. Yang, S. V. Kathuria, R. Vadrevu and C. R. Matthews, Betaalpha-hairpin clamps brace betaalphabeta modules and can make substantive contributions to the stability of TIM barrel proteins, *PLoS One*, 2009, **4**, e7179.

403 J. A. Gerlt, New wine from old barrels, *Nat. Struct. Biol.*, 2000, **7**, 171–173.

404 T. A. Bharat, S. Eisenbeis, K. Zeth and B. Höcker, A beta alpha-barrel built by the combination of fragments from different folds, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 9942–9947.

405 B. Höcker, Directed evolution of (betaalpha)(8)-barrel enzymes, *Biomol. Eng.*, 2005, **22**, 31–38.

406 A. Fischer, T. Seitz, A. Lochner, R. Sterner, R. Merkl and M. Bocola, A fast and precise approach for computational saturation mutagenesis and its experimental validation by using an artificial (betaalpha)8-barrel protein, *ChemBioChem*, 2011, **12**, 1544–1550.

407 S. Eisenbeis, W. Proffitt, M. Coles, V. Truffault, S. Shanmugaratnam, J. Meiler and B. Höcker, Potential of Fragment Recombination for Rational Design of Proteins, *J. Am. Chem. Soc.*, 2012, **134**, 4019–4022.

408 X. Deng, J. Lee, A. J. Michael, D. R. Tomchick, E. J. Goldsmith and M. A. Phillips, Evolution of substrate specificity within a diverse family of beta/alpha-barrel-fold basic amino acid decarboxylases: X-ray structure determination of enzymes with specificity for L-arginine and carboxynorspermidine, *J. Biol. Chem.*, 2010, **285**, 25708–25719.

409 S. Deechongkit, H. Nguyen, M. Jager, E. T. Powers, M. Gruebele and J. W. Kelly, Beta-sheet folding mechanisms from perturbation energetics, *Curr. Opin. Struct. Biol.*, 2006, **16**, 94–101.

410 W. R. Forsyth, O. Bilsel, Z. Gu and C. R. Matthews, Topology and sequence in the folding of a TIM barrel protein: global analysis highlights partitioning between transient off-pathway and stable on-pathway folding intermediates in the complex folding mechanism of a (betaalpha)8 barrel of unknown function from B. subtilis, *J. Mol. Biol.*, 2007, **372**, 236–253.

411 Z. Gu, M. K. Rao, W. R. Forsyth, J. M. Finke and C. R. Matthews, Structural analysis of kinetic folding intermediates for a TIM barrel protein, indole-3-glycerol phosphate synthase, by hydrogen exchange mass spectrometry and Go model simulation, *J. Mol. Biol.*, 2007, **374**, 528–546.

412 C. Kalyanaraman, K. Bernacki and M. P. Jacobson, Virtual screening against highly charged active sites: identifying substrates of alpha-beta barrel enzymes, *Biochemistry*, 2005, **44**, 2059–2071.

413 A. Sakai, A. A. Fedorov, E. V. Fedorov, A. M. Schnoes, M. E. Glasner, S. Brown, M. E. Rutter, K. Bain, S. Chang, T. Gheyi, J. M. Sauder, S. K. Burley, P. C. Babbitt, S. C. Almo and J. A. Gerlt, Evolution of enzymatic activities in the enolase superfamily: stereochemically distinct mechanisms in two families of cis,cis-muconate lactonizing enzymes, *Biochemistry*, 2009, **48**, 1445–1453.

414 R. Kourist, H. Jochens, S. Bartsch, R. Kuipers, S. K. Padhi, M. Gall, D. Böttcher, H. J. Joosten and U. T. Bornscheuer, The alpha/beta-hydrolase fold 3DM database (ABHDB) as a tool for protein engineering, *ChemBioChem*, 2010, **11**, 1635–1643.

415 H. Jochens, M. Hesseler, K. Stiba, S. K. Padhi, R. J. Kazlauskas and U. T. Bornscheuer, Protein engineering of alpha/beta-hydrolase fold enzymes, *ChemBioChem*, 2011, **12**, 1508–1517.

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

417 T. Lukk, A. Sakai, C. Kalyanaraman, S. D. Brown, H. J. Imker, L. Song, A. A. Fedorov, E. V. Fedorov, R. Toro, B. Hillerich, R. Seidel, Y. Patskovsky, M. W. Vetting, S. K. Nair, P. C. Babbitt, S. C. Almo, J. A. Gerlt and M. P. Jacobson,

Homology models guide discovery of diverse enzyme specificities among dipeptide epimerases in the enolase superfamily, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 4122–4127.

418 A. Zanghellini, L. Jiang, A. M. Wollacott, G. Cheng, J. Meiler, E. A. Althoff, D. Rothlisberger and D. Baker, New algorithms and an *in silico* benchmark for computational enzyme design, *Protein Sci.*, 2006, **15**, 2785–2794.

419 C. Malisi, O. Kohlbacher and B. Höcker, Automated scaffold selection for enzyme design, *Proteins*, 2009, **77**, 74–83.

420 E. Dellus-Gur, A. Toth-Petroczy, M. Elias and D. S. Tawfik, What makes a protein fold amenable to functional innovation? Fold polarity and stability trade-offs, *J. Mol. Biol.*, 2013, **425**, 2609–2621.

421 M. Gebauer and A. Skerra, Anticalins small engineered binding proteins based on the lipocalin scaffold, *Methods Enzymol.*, 2012, **503**, 157–188.

422 M. Gebauer, A. Schiefner, G. Matschiner and A. Skerra, Combinatorial design of an Anticalin directed against the extra-domain B for the specific targeting of oncofetal fibronectin, *J. Mol. Biol.*, 2012, **425**, 780–802.

423 A. M. Hohlbaum and A. Skerra, Anticalins: the lipocalin family as a novel protein scaffold for the development of next-generation immunotherapies, *Expert Rev. Clin. Immunol.*, 2007, **3**, 491–501.

424 S. Schlehuber and A. Skerra, Anticalins as an alternative to antibody technology, *Expert Opin. Biol. Ther.*, 2005, **5**, 1453–1462.

425 A. Skerra, Anticalins as alternative binding proteins for therapeutic use, *Curr. Opin. Mol. Ther.*, 2007, **9**, 336–344.

426 A. Skerra, Alternative binding proteins: anticalins - harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities, *FEBS J.*, 2008, **275**, 2677–2683.

427 E. Gunneriusson, K. Nord, M. Uhlen and P. A. Nygren, Affinity maturation of a Taq DNA polymerase specific affibody by helix shuffling, *Environ. Prot. Eng.*, 1999, **12**, 873–878.

428 E. Gunneriusson, P. Samuelson, J. Ringdahl, H. Gronlund, P. A. Nygren and S. Stahl, Staphylococcal surface display of immunoglobulin A (IgA)- and IgE-specific in vitro-selected binding proteins (affibodies) based on Staphylococcus aureus protein A, *Appl. Environ. Microbiol.*, 1999, **65**, 4134–4140.

429 G. Kronvall and K. Jonsson, Receptins: a novel term for an expanding spectrum of natural and engineered microbial proteins with binding properties for mammalian proteins, *J. Mol. Recognit.*, 1999, **12**, 38–44.

430 K. Nord, E. Gunneriusson, J. Ringdahl, S. Stahl, M. Uhlen and P. A. Nygren, Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain, *Nat. Biotechnol.*, 1997, **15**, 772–777.

431 K. Nord, O. Nord, M. Uhlen, B. Kelley, C. Ljungqvist and P. A. Nygren, Recombinant human factor VIII-specific affinity ligands selected from phage-displayed combinatorial libraries of protein A, *Eur. J. Biochem.*, 2001, **268**, 4269–4277.

432 J. Feldwisch and V. Tolmachev, Engineering of affibody molecules for therapy and diagnostics, *Methods Mol. Biol.*, 2012, **899**, 103–126.

433 J. Löfblom, J. Feldwisch, V. Tolmachev, J. Carlsson, S. Ståhl and F. Y. Frejd, Affibody molecules: engineered proteins for therapeutic, diagnostic and biotechnological applications, *FEBS Lett.*, 2010, **584**, 2670–2680.

434 P.-Å. Nygren, Alternative binding proteins: affibody binding proteins developed from a small three-helix bundle scaffold, *FEBS J.*, 2008, **275**, 2668–2676.

435 A. Orlova, J. Feldwisch, L. Abrahmsén and V. Tolmachev, Update: affibody molecules for molecular imaging and therapy for cancer, *Cancer Biother. Radiopharm.*, 2007, **22**, 573–584.

436 V. Tolmachev, A. Orlova, F. Y. Nilsson, J. Feldwisch, A. Wennborg and L. Abrahmsén, Affibody molecules: potential for *in vivo* imaging of molecular targets for cancer therapy, *Expert Opin. Biol. Ther.*, 2007, **7**, 555–568.

437 L. A. Fernández, Prokaryotic expression of antibodies and affibodies, *Curr. Opin. Biotechnol.*, 2004, **15**, 364–373.

438 R. Sterner and F. X. Schmid, *De novo* design of an enzyme, *Science*, 2004, **304**, 1916–1917.

439 G. Cheng, B. Qian, R. Samudrala and D. Baker, Improvement in protein functional site prediction by distinguishing structural and functional constraints on protein family evolution using computational design, *Nucleic Acids Res.*, 2005, **33**, 5861–5867.

440 R. A. Chica, N. Doucet and J. N. Pelletier, Semi-rational approaches to engineering enzyme activity: combining the benefits of directed evolution and rational design, *Curr. Opin. Biotechnol.*, 2005, **16**, 378–384.

441 C. T. Saunders and D. Baker, Recapitulation of protein family divergence using flexible backbone protein design, *J. Mol. Biol.*, 2005, **346**, 631–644.

442 C. A. Floudas, H. K. Fung, S. R. McAllister, M. Monnigmann and R. Rajgaria, Advances in protein structure prediction and *de novo* protein design: A review, *Chem. Eng. Sci.*, 2006, **61**, 966–988.

443 O. Alvizo, B. D. Allen and S. L. Mayo, Computational protein design promises to revolutionize protein engineering, *BioTechniques*, 2007, **42**, 3133, 35 passim.

444 J. L. Anderson, R. L. Koder, C. C. Moser and P. L. Dutton, Controlling complexity and water penetration in functional *de novo* protein design, *Biochem. Soc. Trans.*, 2008, **36**, 1106–1111.

445 T. R. Ward, Artificial enzymes made to order: combination of computational design and directed evolution, *Angew. Chem., Int. Ed.*, 2008, **47**, 7802–7803.

446 M. L. Bellows and C. A. Floudas, Computational methods for *de novo* protein design and its applications to the human immunodeficiency virus 1, purine nucleoside phosphorylase, ubiquitin specific protease 7, and, histone demethylases, *Curr. Drug Targets*, 2010, **11**, 264–278.

447 H. C. Fry, A. Lehmann, J. G. Saven, W. F. DeGrado and M. J. Therien, Computational design and elaboration of a *de novo* heterotetrameric alpha-helical protein that

selectively binds an emissive abiological (porphinato)zinc chromophore, *J. Am. Chem. Soc.*, 2010, **132**, 3997–4005.
448 S. Lutz, Beyond directed evolution–semi-rational protein engineering and design, *Curr. Opin. Biotechnol.*, 2010, **21**, 734–743.
449 R. J. Pantazes, M. J. Grisewood and C. D. Maranas, Recent advances in computational protein design, *Curr. Opin. Struct. Biol.*, 2011, **21**, 467–472.
450 J. Pleiss, Protein design in metabolic engineering and synthetic biology, *Curr. Opin. Biotechnol.*, 2011, **22**, 611–617.
451 I. Samish, C. M. MacDermaid, J. M. Perez-Aguilar and J. G. Saven, Theoretical and computational protein design, *Annu. Rev. Phys. Chem.*, 2011, **62**, 129–149.
452 D. Hilvert, Design of protein catalysts, *Annu. Rev. Biochem.*, 2013, **82**, 447–470.
453 H. Kries, R. Blomberg and D. Hilvert, *De novo* enzymes by computational design, *Curr. Opin. Chem. Biol.*, 2013, **17**, 221–228.
454 H. K. Privett, G. Kiss, T. M. Lee, R. Blomberg, R. A. Chica, L. M. Thomas, D. Hilvert, K. N. Houk and S. L. Mayo, Iterative approach to computational enzyme design, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 3790–3795.
455 J. G. Saven, Computational protein design: engineering molecular diversity, nonnatural enzymes, nonbiological cofactor complexes, and membrane proteins, *Curr. Opin. Chem. Biol.*, 2011, **15**, 452–457.
456 P. S. Huang, G. Oberdorfer, C. Xu, X. Y. Pei, B. L. Nannenga, J. M. Rogers, F. DiMaio, T. Gonen, B. Luisi and D. Baker, High thermodynamic stability of parametrically designed helical bundles, *Science*, 2014, **346**, 481–485.
457 B. A. Smith and M. H. Hecht, Novel proteins: from fold to function, *Curr. Opin. Chem. Biol.*, 2011, **15**, 421–426.
458 A. Bhattacherjee and P. Biswas, Combinatorial design of protein sequences with applications to lattice and real proteins, *J. Chem. Phys.*, 2009, **131**, 125101.
459 C. L. Kleinman, N. Rodrigue, C. Bonnard, H. Philippe and N. Lartillot, A maximum likelihood framework for protein design, *BMC Bioinf.*, 2006, **7**, 326.
460 S. D. Khare, Y. Kipnis, P. Greisen Jr, R. Takeuchi, Y. Ashani, M. Goldsmith, Y. Song, J. L. Gallaher, I. Silman, H. Leader, J. L. Sussman, B. L. Stoddard, D. S. Tawfik and D. Baker, Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis, *Nat. Chem. Biol.*, 2012, **8**, 294–300.
461 B. Höcker, A metalloenzyme reloaded, *Nat. Chem. Biol.*, 2012, **8**, 224–225.
462 E. A. Althoff, L. Wang, L. Jiang, L. Giger, J. K. Lassila, Z. Wang, M. Smith, S. Hari, P. Kast, D. Herschlag, D. Hilvert and D. Baker, Robust design and optimization of retroaldol enzymes, *Protein Sci.*, 2012, **21**, 717–726.
463 L. Giger, S. Caner, R. Obexer, P. Kast, D. Baker, N. Ban and D. Hilvert, Evolution of a designed retro-aldolase leads to complete active site remodeling, *Nat. Chem. Biol.*, 2013, **9**, 494–498.
464 K. Feldmeier and B. Höcker, Computational protein design of ligand binding and catalysis, *Curr. Opin. Chem. Biol.*, 2013, **17**, 929–933.
465 S. Božič, T. Doles, H. Gradišar and R. Jerala, New designed protein assemblies, *Curr. Opin. Chem. Biol.*, 2013, **17**, 940–945.
466 J. Simms and P. J. Booth, Membrane proteins by accident or design, *Curr. Opin. Chem. Biol.*, 2013, **17**, 976–981.
467 M. A. Hallen, D. A. Keedy and B. R. Donald, Dead-end elimination with perturbations (DEEPer): a provable protein design algorithm with continuous sidechain and backbone flexibility, *Proteins*, 2013, **81**, 18–39.
468 G. Kiss, N. Çelebi-ölçüm, R. Moretti, D. Baker and K. N. Houk, Computational enzyme design, *Angew. Chem., Int. Ed.*, 2013, **52**, 5700–5725.
469 D. J. Tantillo, How an enzyme might accelerate an intramolecular Diels-Alder reaction: theozymes for the formation of salvileucalin B, *Org. Lett.*, 2010, **12**, 1164–1167.
470 X. Zhang, J. DeChancie, H. Gunaydin, A. B. Chowdry, F. R. Clemente, A. J. Smith, T. M. Handel and K. N. Houk, Quantum mechanical design of enzyme active sites, *J. Org. Chem.*, 2008, **73**, 889–899.
471 J. Dechancie, F. R. Clemente, A. J. Smith, H. Gunaydin, Y. L. Zhao, X. Zhang and K. N. Houk, How similar are enzyme active site geometries derived from quantum mechanical theozymes to crystal structures of enzyme–inhibitor complexes? Implications for enzyme design, *Protein Sci.*, 2007, **16**, 1851–1866.
472 D. J. Tantillo, J. Chen and K. N. Houk, Theozymes and compuzymes: theoretical models for biological catalysis, *Curr. Opin. Chem. Biol.*, 1998, **2**, 743–750.
473 G. Bouvignies, P. Vallurupalli, D. F. Hansen, B. E. Correia, O. Lange, A. Bah, R. M. Vernon, F. W. Dahlquist, D. Baker and L. E. Kay, Solution structure of a minor and transiently formed state of a T4 lysozyme mutant, *Nature*, 2011, **477**, 111–114.
474 S. Cooper, F. Khatib, A. Treuille, J. Barbero, J. Lee, M. Beenen, A. Leaver-Fay, D. Baker, Z. Popović and F. Players, Predicting protein structures with a multiplayer online game, *Nature*, 2010, **466**, 756–760.
475 F. DiMaio, A. Leaver-Fay, P. Bradley, D. Baker and I. Andre, Modeling symmetric macromolecular structures in Rosetta3, *PLoS One*, 2011, **6**, e20450.
476 S. J. Fleishman, A. Leaver-Fay, J. E. Corn, E. M. Strauch, S. D. Khare, N. Koga, J. Ashworth, P. Murphy, F. Richter, G. Lemmon, J. Meiler and D. Baker, RosettaScripts: a scripting language interface to the Rosetta macromolecular modeling suite, *PLoS One*, 2011, **6**, e20161.
477 J. Handl, J. Knowles, R. Vernon, D. Baker and S. C. Lovell, The dual role of fragments in fragment-assembly methods for *de novo* protein structure prediction, *Proteins*, 2012, **80**, 490–504.
478 A. Leaver-Fay, M. Tyka, S. M. Lewis, O. F. Lange, J. Thompson, R. Jacak, K. Kaufman, P. D. Renfrew, C. A. Smith, W. Sheffler, I. W. Davis, S. Cooper, A. Treuille, D. J. Mandell, F. Richter, Y. E. Ban, S. J. Fleishman, J. E. Corn, D. E. Kim, S. Lyskov, M. Berrondo, S. Mentzer, Z. Popovic, J. J. Havranek, J. Karanicolas, R. Das, J. Meiler, T. Kortemme, J. J. Gray, B. Kuhlman, D. Baker and P. Bradley, ROSETTA3:

an object-oriented software suite for the simulation and design of macromolecules, *Methods Enzymol.*, 2011, **487**, 545–574.

479 O. F. Lange, P. Rossi, N. G. Sgourakis, Y. Song, H. W. Lee, J. M. Aramini, A. Ertekin, R. Xiao, T. B. Acton, G. T. Montelione and D. Baker, Determination of solution structures of proteins up to 40 kDa using CS-Rosetta with sparse NMR data from deuterated samples, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 10873–10878.

480 T. C. Terwilliger, F. Dimaio, R. J. Read, D. Baker, G. Bunkóczi, P. D. Adams, R. W. Grosse-Kunstleve, P. V. Afonine and N. Echols, phenix.mr_rosetta: molecular replacement and model rebuilding with Phenix and Rosetta, *J. Struct. Funct. Genomics*, 2012, **13**, 81–90.

481 C. Schmitz, R. Vernon, G. Otting, D. Baker and T. Huber, Protein structure determination from pseudocontact shifts using ROSETTA, *J. Mol. Biol.*, 2012, **416**, 668–677.

482 D. Gront, D. W. Kulp, R. M. Vernon, C. E. Strauss and D. Baker, Generalized fragment picking in Rosetta: design, protocols and applications, *PLoS One*, 2011, **6**, e23294.

483 J. Thompson and D. Baker, Incorporation of evolutionary information into Rosetta comparative modeling, *Proteins*, 2011, **79**, 2380–2388.

484 F. Lauck, C. A. Smith, G. F. Friedland, E. L. Humphris and T. Kortemme, RosettaBackrub—a web server for flexible backbone protein structure modeling and design, *Nucleic Acids Res.*, 2010, **38**, W569–W575.

485 C. A. Smith and T. Kortemme, Predicting the tolerated sequences for proteins and protein interfaces using RosettaBackrub flexible backbone design, *PLoS One*, 2011, **6**, e20451.

486 E. H. C. Bromley, K. Channon, E. Moutevelis and D. N. Woolfson, Peptide and protein building blocks for synthetic biology: from programming biomolecules to self-organized biomolecular systems, *ACS Chem. Biol.*, 2008, **3**, 38–50.

487 C. T. Armstrong, A. L. Boyle, E. H. Bromley, Z. N. Mahmoud, L. Smith, A. R. Thomson and D. N. Woolfson, Rational design of peptide-based building blocks for nanoscience and synthetic biology, *Faraday Discuss.*, 2009, **143**, 305–317, discussion 359–372.

488 E. H. C. Bromley, R. B. Sessions, A. R. Thomson and D. N. Woolfson, Designed alpha-helical tectons for constructing multicomponent synthetic biological systems, *J. Am. Chem. Soc.*, 2009, **131**, 928–930.

489 S. R. Gordon, E. J. Stanley, S. Wolf, A. Toland, S. J. Wu, D. Hadidi, J. H. Mills, D. Baker, I. S. Pultz and J. B. Siegel, Computational design of an alpha-gliadin peptidase, *J. Am. Chem. Soc.*, 2012, **134**, 20513–20520.

490 D. Bhattacharya and J. Cheng, 3Drefine: Consistent protein structure refinement by optimizing hydrogen bonding network and atomic-level energy minimization, *Proteins*, 2013, **81**, 119–131.

491 J. A. Davey and R. A. Chica, Multistate approaches in computational protein design, *Protein Sci.*, 2012, **21**, 1241–1252.

492 K. J. M. Hanf, Protein design for diversity of sequences and conformations using dead-end elimination, *Methods Mol. Biol.*, 2012, **899**, 127–144.

493 R. Kim and J. Skolnick, Assessment of programs for ligand binding affinity prediction, *J. Comput. Chem., Jpn*, 2008, **29**, 1316–1331.

494 P. Cunningham, I. Afzal-Ahmed and R. J. Naftalin, Docking studies show thatD-glucose and quercetin slide through the transporter GLUT1, *J. Biol. Chem.*, 2006, **281**, 5797–5803.

495 C. Hetényi, U. Maran, A. T. García-Sosa and M. Karelson, Structure-based calculation of drug efficiency indices, *Bioinformatics*, 2007, **23**, 2678–2685.

496 S. Cosconati, S. Forli, A. L. Perryman, R. Harris, D. S. Goodsell and A. J. Olson, Virtual Screening with AutoDock: Theory and Practice, *Expert Opin. Drug Discovery*, 2010, **5**, 597–607.

497 O. Trott and A. J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, *J. Comput. Chem.*, 2010, **31**, 455–461.

498 G. Sandeep, K. P. Nagasree, M. Hanisha and M. M. Kumar, AUDocker LE: A GUI for virtual screening with AUTODOCK Vina, *BMC Res. Notes*, 2011, **4**, 445.

499 S. D. Handoko, X. Ouyang, C. T. Su, C. K. Kwoh and Y. S. Ong, QuickVina: accelerating AutoDock Vina using gradient-based heuristics for global optimization, *IEEE/ACM Trans. Comput. Biol. Bioinf.*, 2012, **9**, 1266–1272.

500 M. Gao and J. Skolnick, APoc: large-scale identification of similar protein pockets, *Bioinformatics*, 2013, **29**, 597–604.

501 M. P. Repasky, R. B. Murphy, J. L. Banks, J. R. Greenwood, I. Tubert-Brohman, S. Bhat and R. A. Friesner, Docking performance of the glide program as evaluated on the Astex and DUD datasets: a complete set of glide SP results and selected results for a new scoring function integrating WaterMap and glide, *J. Comput.-Aided Mol. Des.*, 2012, **26**, 787–799.

502 T. A. Halgren, R. B. Murphy, R. A. Friesner, H. S. Beard, L. L. Frye, W. T. Pollard and J. L. Banks, Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, *J. Med. Chem.*, 2004, **47**, 1750–1759.

503 R. A. Friesner, J. L. Banks, R. B. Murphy, T. A. Halgren, J. J. Klicic, D. T. Mainz, M. P. Repasky, E. H. Knoll, M. Shelley, J. K. Perry, D. E. Shaw, P. Francis and P. S. Shenkin, Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem.*, 2004, **47**, 1739–1749.

504 K. T. Schomburg, I. Ardao, K. Gotz, F. Rieckenberg, A. Liese, A. P. Zeng and M. Rarey, Computational biotechnology: Prediction of competitive substrate inhibition of enzymes by buffer compounds with protein-ligand docking, *J. Biotechnol.*, 2012, **161**, 391–401.

505 M. Vass, A. Tarcsay and G. M. Keserü, Multiple ligand docking by Glide: implications for virtual second-site screening, *J. Comput.-Aided Mol. Des.*, 2012, **26**, 821–834.

506 P. A. Greenidge, C. Kramer, J. C. Mozziconacci and R. M. Wolf, MM/GBSA Binding Energy Prediction on the PDBbind Data Set: Successes, Failures, and Directions for Further Improvement, *J. Chem. Inf. Model.*, 2012, **53**, 201–209.

507 D. Plewczynski, M. Laźniewski, R. Augustyniak and K. Ginalski, Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database, *J. Comput. Chem., Jpn*, 2011, **32**, 742–755.

508 R. Wang, X. Fang, Y. Lu, C. Y. Yang and S. Wang, The PDBbind database: methodologies and updates, *J. Med. Chem.*, 2005, **48**, 4111–4119.

509 Y. X. Yuan, J. F. Pei and L. H. Lai, LigBuilder 2: A practical *de novo* drug design approach, *J. Chem. Inf. Model.*, 2011, **51**, 1083–1091.

510 S. Sirin, R. Kumar, C. Martinez, M. J. Karmilowicz, P. Ghosh, Y. A. Abramov, V. Martin and W. Sherman, A Computational Approach to Enzyme Design: Predicting omega-Aminotransferase Catalytic Activity Using Docking and MM-GBSA Scoring, *J. Chem. Inf. Model.*, 2014, **54**, 2334–2346.

511 H. Y. Zhou and J. Skolnick, FINDSITEcomb: A Threading/Structure-Based, Proteomic-Scale Virtual Ligand Screening Approach, *J. Chem. Inf. Model.*, 2013, **53**, 230–240.

512 J. C. Faver, M. L. Benson, X. A. He, B. P. Roberts, B. Wang, M. S. Marshall, M. R. Kennedy, C. D. Sherrill and K. M. Merz, Formal Estimation of Errors in Computed Absolute Interaction Energies of Protein–Ligand Complexes, *J. Chem. Theory Comput.*, 2011, **7**, 790–797.

513 M. Goldsmith and D. S. Tawfik, Enzyme engineering by targeted libraries, *Methods Enzymol.*, 2013, **523**, 257–283.

514 A. J. Ruff, A. Dennig and U. Schwaneberg, To get what we aim for: progress in diversity generation methods, *FEBS J.*, 2013, **280**, 2961–2978.

515 T. Zhang, Z. F. Wu, H. Chen, Q. Wu, Z. Z. Tang, J. B. Gou, L. H. Wang, W. W. Hao, C. M. Wang and C. M. Li, Progress in strategies for sequence diversity library creation for directed evolution, *Afr. J. Biotechnol.*, 2010, **9**, 9277–9285.

516 *Directed Evolution Library Creation: Methods and Protocols*, ed. E. M. J. Gillam, J. N. Copp and D. Ackerley, Springer, Berlin, 2014.

517 J. Damborsky and J. Brezovsky, Computational tools for designing and engineering biocatalysts, *Curr. Opin. Chem. Biol.*, 2009, **13**, 26–34.

518 J. Brezovsky, E. Chovancova, A. Gora, A. Pavelka, L. Biedermannova and J. Damborsky, Software tools for identification, visualization and analysis of protein tunnels and channels, *Biotechnol. Adv.*, 2013, **31**, 38–49.

519 E. Sebestova, J. Bendl, J. Brezovsky and J. Damborsky, Computational tools for designing smart libraries, *Methods Mol. Biol.*, 2014, **1179**, 291–314.

520 R. Krašovec, R. V. Belavkin, J. A. D. Aston, A. Channon, E. Aston, B. M. Rash, M. Kadirvel, S. Forbes and C. G. Knight, Mutation rate plasticity in rifampicin resistance depends on *Escherichia coli* cell-cell interactions, *Nat. Commun.*, 2014, **5**, 3742.

521 M. Oates, D. Corne and R. Loader, Investigation of a characteristic bimodal convergence-time/mutation-rate feature in evolutionary search, *Proc. Congr. Evol. Comput., IEEE*, 1999, pp. 2175–2182.

522 M. Oates and D. W. Corne, Overcoming fitness barriers in multi-modal search spaces, in *Foundation of Genetic Algorithms 6*, ed. W. N. Martin and W. M. Spears, Academic Press, London, 2001, pp. 5–26.

523 M. J. Oates, D. Corne and R. Loader, Tri-phase performance profile of evolutionary search on uni- and multi-modal search spaces, in *Proc. IEEE Congr. Evol. Computation*, IEEE Neural Networks Council, San Diego, 2000, pp. 357–364.

524 M. J. Oates, D. W. Corne and D. B. Kell, The bimodal feature at large population sizes and high selection pressure: implications for directed evolution, in *Recent advances in simulated evolution and learning*, ed. K. C. Tan, M. H. Lim, X. Yao and L. Wang, World Scientific, Singapore, 2003, pp. 215–240.

525 M. Zaccolo and E. Gherardi, The effect of high-frequency random mutagenesis on *in vitro* protein evolution: A study on TEM-1 β-lactamase, *J. Mol. Biol.*, 1999, **285**, 775–783.

526 P. S. Daugherty, G. Chen, B. L. Iverson and G. Georgiou, Quantitative analysis of the effect of the mutation frequency on the affinity maturation of single chain Fv antibodies, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 2029–2034.

527 D. A. Drummond, B. L. Iverson, G. Georgiou and F. H. Arnold, Why high-error-rate random mutagenesis libraries are enriched in functional and improved proteins, *J. Mol. Biol.*, 2005, **350**, 806–816.

528 A. M. Leconte, B. C. Dickinson, D. D. Yang, I. A. Chen, B. Allen and D. R. Liu, A population-based experimental model for protein evolution: effects of mutation rate and selection stringency on evolutionary outcomes, *Biochemistry*, 2013, **52**, 1490–1499.

529 D. W. Leung, E. Chen and D. V. Goeddel, A method for random mutagenesis of a defined DNA segment using a modified polymerase chain reaction., *Technique*, 1989, **1**, 11–15.

530 E. O. McCullum, B. A. Williams, J. Zhang and J. C. Chaput, Random mutagenesis by error-prone PCR, *Methods Mol. Biol.*, 2010, **634**, 103–109.

531 T. S. Wong, D. Roccatano, M. Zacharias and U. Schwaneberg, A statistical analysis of random mutagenesis methods used for directed protein evolution, *J. Mol. Biol.*, 2006, **355**, 858–871.

532 T. S. Wong, D. Roccatano and U. Schwaneberg, Challenges of the genetic code for exploring sequence space in directed protein evolution, *Biocatal. Biotransform.*, 2007, **25**, 229–241.

533 T. S. Rasila, M. I. Pajunen and H. Savilahti, Critical evaluation of random mutagenesis by error-prone polymerase chain reaction protocols, *Escherichia coli* mutator strain, and hydroxylamine treatment, *Anal. Biochem.*, 2009, **388**, 71–80.

534 J. Zhao, T. Kardashliev, A. Joelle Ruff, M. Bocola and U. Schwaneberg, Lessons from diversity of directed evolution experiments by an analysis of 3,000 mutations, *Biotechnol. Bioeng.*, 2014, **111**, 2380–2389.

535 R. C. Cadwell and G. F. Joyce, Randomization of genes by PCR mutagenesis, *PCR Methods Appl.*, 1992, **2**, 28–33.

536 M. Camps, J. Naukkarinen, B. P. Johnson and L. A. Loeb, Targeted gene evolution in *Escherichia coli* using a highly error-prone DNA polymerase I, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 9727–9732.

537 J. N. Copp, P. Hanson-Manful, D. F. Ackerley and W. M. Patrick, Error-prone PCR and effective generation of gene variant libraries for directed evolution, *Methods Mol. Biol.*, 2014, **1179**, 3–22.

538 I. Matsumura and L. A. Rowe, Whole plasmid mutagenic PCR for directed protein evolution, *Biomol. Eng.*, 2005, **22**, 73–79.

539 D. L. Alexander, J. Lilly, J. Hernandez, J. Romsdahl, C. J. Troll and M. Camps, Random mutagenesis by error-prone pol plasmid replication in *Escherichia coli*, *Methods Mol. Biol.*, 2014, **1179**, 31–44.

540 R. Fujii, M. Kitaoka and K. Hayashi, One-step random mutagenesis by error-prone rolling circle amplification, *Nucleic Acids Res.*, 2004, **32**, e145.

541 R. Fujii, M. Kitaoka and K. Hayashi, RAISE: a simple and novel method of generating random insertion and deletion mutations, *Nucleic Acids Res.*, 2006, **34**, e30.

542 Y. Kipnis, E. Dellus-Gur and D. S. Tawfik, TRINS: a method for gene modification by randomized tandem repeat insertions, *Protein Eng., Des. Sel.*, 2012, **25**, 437–444.

543 R. Fujii, M. Kitaoka and K. Hayashi, Random insertional-deletional strand exchange mutagenesis (RAISE): a simple method for generating random insertion and deletion mutations, *Methods Mol. Biol.*, 2014, **1179**, 151–158.

544 A. Ravikumar, A. Arrieta and C. C. Liu, An orthogonal DNA replication system in yeast, *Nat. Chem. Biol.*, 2014, **10**, 175–177.

545 L. Pritchard, D. W. Corne, D. B. Kell, J. J. Rowland and M. K. Winson, A general model of error-prone PCR, *J. Theor. Biol.*, 2004, **234**, 497–509.

546 S. Hoebenreich, F. E. Zilly, C. G. Acevedo-Rocha, M. Zilly and M. T. Reetz, Speeding up Directed Evolution: Combining the Advantages of Solid-Phase Combinatorial Gene Synthesis with Statistically Guided Reduction of Screening Effort, *ACS Synth. Biol.*, 2014, DOI: 10.1021/sb5002399.

547 D. Wei, M. Li, X. Zhang and L. Xing, An improvement of the site-directed mutagenesis method by combination of megaprimer, one-side PCR and DpnI treatment, *Anal. Biochem.*, 2004, **331**, 401–403.

548 D. L. Steffens and J. G. Williams, Efficient site-directed saturation mutagenesis using degenerate oligonucleotides, *J. Biomol. Tech.*, 2007, **18**, 147–149.

549 A. Fersht, *Enzyme structure and mechanism*, W.H. Freeman, San Francisco, 2nd edn, 1977.

550 J. Braman, C. Papworth and A. Greener, Site-directed mutagenesis using double-stranded plasmid DNA templates, *Methods Mol. Biol.*, 1996, **57**, 31–44.

551 C. Papworth, J. C. Bauer, J. Braman and D. A. Wright, Site-directed mutagenesis in one day with >80% efficiency, *Strategies*, 1996, **9**, 3–4.

552 L. Zheng, U. Baumann and J. L. Reymond, An efficient one-step site-directed and site-saturation mutagenesis protocol, *Nucleic Acids Res.*, 2004, **32**, e115.

553 J. Sanchis, L. Fernandez, J. D. Carballeira, J. Drone, Y. Gumulya, H. Hobenreich, D. Kahakeaw, S. Kille, R. Lohmer, J. J. Peyralans, J. Podtetenieff, S. Prasad, P. Soni, A. Taglieber, S. Wu, F. E. Zilly and M. T. Reetz, Improved PCR method for the creation of saturation mutagenesis libraries in directed evolution: application to difficult-to-amplify templates, *Appl. Microbiol. Biotechnol.*, 2008, **81**, 387–397.

554 K. L. Morrison and G. A. Weiss, Combinatorial alanine-scanning, *Curr. Opin. Chem. Biol.*, 2001, **5**, 302–307.

555 G. A. Weiss, C. K. Watanabe, A. Zhong, A. Goddard and S. S. Sidhu, Rapid mapping of protein functional epitopes by combinatorial alanine scanning, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 8950–8954.

556 T. S. Wong, D. Roccatano and U. Schwaneberg, Steering directed protein evolution: strategies to manage combinatorial complexity of mutant libraries, *Environ. Microbiol.*, 2007, **9**, 2645–2659.

557 M. T. Reetz, L. W. Wang and M. Bocola, Directed evolution of enantioselective enzymes: iterative cycles of CASTing for probing protein-sequence space, *Angew. Chem., Int. Ed.*, 2006, **45**, 1236–1241.

558 R. D. Kirsch and E. Joly, An improved PCR-mutagenesis strategy for two-site mutagenesis or sequence swapping between related genes, *Nucleic Acids Res.*, 1998, **26**, 1848–1850.

559 L. W. Chiang, I. Kovari and M. M. Howe, Mutagenic oligonucleotide-directed PCR amplification (Mod-PCR): an efficient method for generating random base substitution mutations in a DNA sequence element, *PCR Methods Appl.*, 1993, **2**, 210–217.

560 S. N. Ho, H. D. Hunt, R. M. Horton, J. K. Pullen and L. R. Pease, Site-directed mutagenesis by overlap extension using the polymerase chain reaction, *Gene*, 1989, **77**, 51–59.

561 R. M. Horton, H. D. Hunt, S. N. Ho, J. K. Pullen and L. R. Pease, Engineering hybrid genes without the use of restriction enzymes: gene splicing by overlap extension, *Gene*, 1989, **77**, 61–68.

562 R. H. Peng, A. S. Xiong and Q. H. Yao, A direct and efficient PAGE-mediated overlap extension PCR method for gene multiple-site mutagenesis, *Appl. Microbiol. Biotechnol.*, 2006, **73**, 234–240.

563 K. L. Heckman and L. R. Pease, Gene splicing and mutagenesis by PCR-driven overlap extension, *Nat. Protoc.*, 2007, **2**, 924–932.

564 E. M. Williams, J. N. Copp and D. F. Ackerley, Site-saturation mutagenesis by overlap extension PCR, *Methods Mol. Biol.*, 2014, **1179**, 83–101.

565 T. S. Wong, K. L. Tee, B. Hauer and U. Schwaneberg, Sequence saturation mutagenesis (SeSaM): a novel method for directed evolution, *Nucleic Acids Res.*, 2004, **32**, e26.

566 T. S. Wong, D. Roccatano, D. Loakes, K. L. Tee, A. Schenk, B. Hauer and U. Schwaneberg, Transversion-enriched sequence saturation mutagenesis (SeSaM-Tv+): a random mutagenesis method with consecutive nucleotide exchanges that complements the bias of error-prone PCR, *Biotechnol. J.*, 2008, **3**, 74–82.

567 H. Mundhada, J. Marienhagen, A. Scacioc, A. Schenk, D. Roccatano and U. Schwaneberg, SeSaM-Tv-II generates a protein sequence space that is unobtainable by epPCR, *ChemBioChem*, 2011, **12**, 1595–1601.

568 A. J. Ruff, T. Kardashliev, A. Dennig and U. Schwaneberg, The Sequence Saturation Mutagenesis (SeSaM) method, *Methods Mol. Biol.*, 2014, **1179**, 45–68.

569 A. Dennig, A. V. Shivange, J. Marienhagen and U. Schwaneberg, OmniChange: the sequence independent method for simultaneous site-saturation of five codons, *PLoS One*, 2011, **6**, e26222.

570 A. Dennig, J. Marienhagen, A. J. Ruff and U. Schwaneberg, OmniChange: simultaneous site saturation of up to five codons, *Methods Mol. Biol.*, 2014, **1179**, 139–149.

571 A. V. Shivange, A. Dennig and U. Schwaneberg, Multi-site saturation by OmniChange yields a pH- and thermally improved phytase, *J. Biotechnol.*, 2014, **170**, 68–72.

572 C. G. Acevedo-Rocha, S. Hoebenreich and M. T. Reetz, Iterative saturation mutagenesis: a powerful approach to engineer proteins by systematically simulating Darwinian evolution, *Methods Mol. Biol.*, 2014, **1179**, 103–128.

573 L. P. Parra, R. Agudo and M. T. Reetz, Directed evolution by using iterative saturation mutagenesis based on multi-residue sites, *ChemBioChem*, 2013, **14**, 2301–2309.

574 M. T. Reetz, J. D. Carballeira and A. Vogel, Iterative saturation mutagenesis on the basis of B factors as a strategy for increasing protein thermostability, *Angew. Chem., Int. Ed.*, 2006, **45**, 7745–7751.

575 M. T. Reetz and J. D. Carballeira, Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes, *Nat. Protoc.*, 2007, **2**, 891–903.

576 M. T. Reetz, D. Kahakeaw and J. Sanchis, Shedding light on the efficacy of laboratory evolution based on iterative saturation mutagenesis, *Mol. BioSyst.*, 2009, **5**, 115–122.

577 M. T. Reetz, P. Soni, L. Fernandez, Y. Gumulya and J. D. Carballeira, Increasing the stability of an enzyme toward hostile organic solvents by directed evolution based on iterative saturation mutagenesis using the B-FIT method, *Chem. Commun.*, 2010, **46**, 8657–8658.

578 M. T. Reetz, S. Prasad, J. D. Carballeira, Y. Gumulya and M. Bocola, Iterative saturation mutagenesis accelerates laboratory evolution of enzyme stereoselectivity: rigorous comparison with traditional methods, *J. Am. Chem. Soc.*, 2010, **132**, 9144–9152.

579 H. Zheng and M. T. Reetz, Manipulating the stereoselectivity of limonene epoxide hydrolase by directed evolution based on iterative saturation mutagenesis, *J. Am. Chem. Soc.*, 2010, **132**, 15744–15751.

580 A. J. Baldwin, K. Busse, A. M. Simm and D. D. Jones, Expanded molecular diversity generation during directed evolution by trinucleotide exchange (TriNEx), *Nucleic Acids Res.*, 2008, **36**, e77.

581 A. J. Baldwin, J. A. Arpino, W. R. Edwards, E. M. Tippmann and D. D. Jones, Expanded chemical diversity sampling through whole protein evolution, *Mol. BioSyst.*, 2009, **5**, 764–766.

582 W. R. Edwards, K. Busse, R. K. Allemann and D. D. Jones, Linking the functions of unrelated proteins using a novel directed evolution domain insertion method, *Nucleic Acids Res.*, 2008, **36**, e78.

583 D. D. Jones, J. A. J. Arpino, A. J. Baldwin and M. C. Edmundson, Transposon-based approaches for generating novel molecular diversity during directed evolution, *Methods Mol. Biol.*, 2014, **1179**, 159–172.

584 H. Liu and J. H. Naismith, An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol, *BMC Biotechnol.*, 2008, **8**, 91.

585 A. Seyfang and J. H. Jin, Multiple site-directed mutagenesis of more than 10 sites simultaneously and in a single round, *Anal. Biochem.*, 2004, **324**, 285–291.

586 A. A. Fushan and D. T. Drayna, MALS: an efficient strategy for multiple site-directed mutagenesis employing a combination of DNA amplification, ligation and suppression PCR, *BMC Biotechnol.*, 2009, **9**, 83.

587 D. M. Kegler-Ebo, C. M. Docktor and D. DiMaio, Codon cassette mutagenesis: a general method to insert or replace individual codons by using universal mutagenic cassettes, *Nucleic Acids Res.*, 1994, **22**, 1593–1599.

588 K. Steiner and H. Schwab, Recent advances in rational approaches for enzyme engineering, *Comput. Struct. Biotechnol. J.*, 2012, **2**, e201209010.

589 Y. Nov, Fitness loss and library size determination in saturation mutagenesis, *PLoS One*, 2013, **8**, e68069.

590 Nomenclature Committee of the International Union of Biochemistry (NC-IUB), Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984, *Eur. J. Biochem.*, 1985, **150**, 1–5.

591 M. A. Mena and P. S. Daugherty, Automated design of degenerate codon libraries, *Protein Eng., Des. Sel.*, 2005, **18**, 559–561.

592 L. Tang, X. Wang, B. Ru, H. Sun, J. Huang and H. Gao, MDC-Analyzer: A novel degenerate primer design tool for the construction of intelligent mutagenesis libraries with contiguous sites, *BioTechniques*, 2014, **56**, 301–310.

593 E. Muñoz and M. W. Deem, Amino acid alphabet size in protein evolution experiments: better to search a small library thoroughly or a large library sparsely? *Protein Eng., Des. Sel.*, 2008, **21**, 311–317.

594 K. L. Tee and T. S. Wong, Polishing the craft of genetic diversity creation in directed evolution, *Biotechnol. Adv.*, 2013, **31**, 1707–1721.

595 S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z. G. Zhang, D. J. Opperman, M. T. Reetz and J. P. Acevedo, Reducing Codon Redundancy and Screening Effort of Combinatorial Protein Libraries Created by Saturation Mutagenesis, *ACS Synth. Biol.*, 2013, **2**, 83–92.

596 M. D. Hughes, D. A. Nagel, A. F. Santos, A. J. Sutherland and A. V. Hine, Removing the redundancy from randomised gene libraries, *J. Mol. Biol.*, 2003, **331**, 973–979.

597 L. X. Tang, H. Gao, X. C. Zhu, X. Wang, M. Zhou and R. X. Jiang, Construction of ''small-intelligent'' focused mutagenesis libraries using well-designed combinatorial degenerate primers, *BioTechniques*, 2012, **52**, 149–157.

598 S. Akanuma, T. Kigawa and S. Yokoyama, Combinatorial mutagenesis to restrict amino acid usage in an enzyme to a reduced set, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 13549–13553.

599 L. H. Bradley, R. E. Kleiner, A. F. Wang, M. H. Hecht and D. W. Wood, An intein-based genetic selection allows the construction of a high-quality library of binary patterned *de novo* protein sequences, *Protein Eng., Des. Sel.*, 2005, **18**, 201–207.

600 J. F. Chaparro-Riggers, K. M. Polizzi and A. S. Bommarius, Better library design: data-driven protein engineering, *Biotechnol. J.*, 2007, **2**, 180–191.

601 J. Tanaka, H. Yanagawa and N. Doi, Comparison of the frequency of functional SH3 domains with different limited sets of amino acids using mRNA display, *PLoS One*, 2011, **6**, e18034.

602 A. G. Sandström, Y. Wikmark, K. Engström, J. Nyhlén and J. E. Bäckvall, Combinatorial reshaping of the *Candida antarctica* lipase A substrate pocket for enantioselectivity using an extremely condensed library, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 78–83.

603 J. Bacardit, M. Stout, J. D. Hirst, A. Valencia, R. E. Smith and N. Krasnogor, Automated alphabet reduction for protein datasets, *BMC Bioinf.*, 2009, **10**, 6.

604 S. Zheng and I. Kwon, Manipulation of enzyme properties by noncanonical amino acid incorporation, *Biotechnol. J.*, 2012, **7**, 47–60.

605 M. G. Hoesl and N. Budisa, *In vivo* incorporation of multiple noncanonical amino acids into proteins, *Angew. Chem., Int. Ed.*, 2011, **50**, 2896–2902.

606 L. Wang, A. Brock, B. Herberich and P. G. Schultz, Expanding the genetic code of *Escherichia coli*, *Science*, 2001, **292**, 498–500.

607 Q. Wang, A. R. Parrish and L. Wang, Expanding the genetic code for biological studies, *Chem. Biol.*, 2009, **16**, 323–336.

608 J. C. Jackson, S. P. Duffy, K. R. Hess and R. A. Mehl, Improving Nature's enzyme active site with genetically encoded unnatural amino acids, *J. Am. Chem. Soc.*, 2006, **128**, 11124–11127.

609 J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang and P. G. Schultz, An expanded eukaryotic genetic code, *Science*, 2003, **301**, 964–967.

610 J. C. Anderson, N. Wu, S. W. Santoro, V. Lakshman, D. S. King and P. G. Schultz, An expanded genetic code with a functional quadruplet codon, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 7566–7571.

611 H. Neumann, K. Wang, L. Davis, M. Garcia-Alai and J. W. Chin, Encoding multiple unnatural amino acids *via* evolution of a quadruplet-decoding ribosome, *Nature*, 2010, **464**, 441–444.

612 J. W. Chin, Reprogramming the genetic code, *Science*, 2012, **336**, 428–429.

613 L. Davis and J. W. Chin, Designer proteins: applications of genetic code expansion in cell biology, *Nat. Rev. Mol. Cell Biol.*, 2012, **13**, 168–182.

614 K. Wang, W. H. Schmied and J. W. Chin, Reprogramming the genetic code: from triplet to quadruplet codes, *Angew. Chem., Int. Ed.*, 2012, **51**, 2288–2297.

615 C. C. Liu and P. G. Schultz, Adding new chemistries to the genetic code, *Annu. Rev. Biochem.*, 2010, **79**, 413–444.

616 C. C. Liu, A. V. Mack, M. L. Tsao, J. H. Mills, H. S. Lee, H. Choe, M. Farzan, P. G. Schultz and V. V. Smider, Protein evolution with an expanded genetic code, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 17688–17693.

617 J. Xie and P. G. Schultz, A chemical toolkit for proteins—an expanded genetic code, *Nat. Rev. Mol. Cell Biol.*, 2006, **7**, 775–782.

618 L. Wang, J. Xie and P. G. Schultz, Expanding the genetic code, *Annu. Rev. Biophys. Biomol. Struct.*, 2006, **35**, 225–249.

619 K. M. Bradley and S. A. Benner, OligArch: A software tool to allow artificially expanded genetic information systems (AEGIS) to guide the autonomous self-assembly of long DNA constructs from multiple DNA single strands, *Beilstein J. Org. Chem.*, 2014, **10**, 1826–1833.

620 J. Xie and P. G. Schultz, Adding amino acids to the genetic repertoire, *Curr. Opin. Chem. Biol.*, 2005, **9**, 548–554.

621 W. P. C. Stemmer, Rapid evolution of a protein *in vivo* by DNA shuffling, *Nature*, 1994, **370**, 389–391.

622 W. P. C. Stemmer, DNA shuffling by random fragmentation and reassembly: *in vitro* recombination for molecular evolution., *Proc. Natl. Acad. Sci. U. S. A.*, 1994, **91**, 10747–10751.

623 H. Zhao and F. H. Arnold, Optimization of DNA shuffling for high fidelity recombination, *Nucleic Acids Res.*, 1997, **25**, 1307–1308.

624 W. M. Coco, W. E. Levinson, M. J. Crist, H. J. Hektor, A. Darzins, P. T. Pienkos, C. H. Squires and D. J. Monticello, DNA shuffling method for generating highly recombined genes and evolved enzymes, *Nat. Biotechnol.*, 2001, **19**, 354–359.

625 W. M. Coco, L. P. Encell, W. E. Levinson, M. J. Crist, A. K. Loomis, L. L. Licato, J. J. Arensdorf, N. Sica, P. T. Pienkos and D. J. Monticello, Growth factor engineering by degenerate homoduplex gene family recombination, *Nat. Biotechnol.*, 2002, **20**, 1246–1250.

626 M. Kikuchi, K. Ohnishi and S. Harayama, Novel family shuffling methods for the *in vitro* evolution of enzymes, *Gene*, 1999, **236**, 159–167.

627 M. Kikuchi, K. Ohnishi and S. Harayama, An effective family shuffling method using single-stranded DNA, *Gene*, 2000, **243**, 133–137.

628 K. Miyazaki, Random DNA fragmentation with endonuclease V: application to DNA shuffling, *Nucleic Acids Res.*, 2002, **30**, e139.
629 Y. An, W. Wu and A. Lv, A convenient and robust method for construction of combinatorial and random mutant libraries, *Biochimie*, 2010, **92**, 1081–1084.
630 A. Crameri, S. A. Raillard, E. Bermudez and W. P. C. Stemmer, DNA shuffling of a family of genes from diverse species accelerates directed evolution, *Nature*, 1998, **391**, 288–291.
631 J. B. Y. H. Behrendorff, W. A. Johnston and E. M. J. Gillam, Restriction enzyme-mediated DNA family shuffling, *Methods Mol. Biol.*, 2014, **1179**, 175–187.
632 J. B. Y. H. Behrendorff, W. A. Johnston and E. M. J. Gillam, DNA shuffling of cytochrome P450 enzymes, *Methods Mol. Biol.*, 2013, **987**, 177–188.
633 N. N. Rosic, W. Huang, W. A. Johnston, J. J. DeVoss and E. M. J. Gillam, Extending the diversity of cytochrome P450 enzymes by DNA family shuffling, *Gene*, 2007, **395**, 40–48.
634 J. M. Joern, P. Meinhold and F. H. Arnold, Analysis of shuffled gene libraries, *J. Mol. Biol.*, 2002, **316**, 643–656.
635 J. F. Chaparro-Riggers, B. L. Loo, K. M. Polizzi, P. R. Gibbs, X. S. Tang, M. J. Nelson and A. S. Bommarius, Revealing biases inherent in recombination protocols, *BMC Biotechnol.*, 2007, **7**, 77.
636 Y. Kawarasaki, K. E. Griswold, J. D. Stevenson, T. Selzer, S. J. Benkovic, B. L. Iverson and G. Georgiou, Enhanced crossover SCRATCHY: construction and high-throughput screening of a combinatorial library containing multiple non-homologous crossovers, *Nucleic Acids Res.*, 2003, **31**, e126.
637 S. Lutz and M. Ostermeier, Preparation of SCRATCHY hybrid protein libraries: size- and in-frame selection of nucleic acid sequences, *Methods Mol. Biol.*, 2003, **231**, 143–151.
638 M. Ostermeier and S. Lutz, The creation of ITCHY hybrid protein libraries, *Methods Mol. Biol.*, 2003, **231**, 129–141.
639 W. M. Patrick and M. L. Gerth, ITCHY: Incremental Truncation for the Creation of Hybrid enzYmes, *Methods Mol. Biol.*, 2014, **1179**, 225–244.
640 W. M. Coco, RACHITT: Gene family shuffling by Random Chimeragenesis on Transient Templates, *Methods Mol. Biol.*, 2003, **231**, 111–127.
641 S. H. Lee, E. J. Ryu, M. J. Kang, E. S. Wang, Z. Piao, Y. J. Choi, K. H. Jung, J. Y. J. Jeon and Y. C. Shin, A new approach to directed gene evolution by recombined extension on truncated templates (RETT), *J. Mol. Catal.*, 2003, **26**, 119–129.
642 A. Hidalgo, A. Schliessmann, R. Molina, J. Hermoso and U. T. Bornscheuer, A one-pot, simple methodology for cassette randomisation and recombination for focused directed evolution, *Protein Eng., Des. Sel.*, 2008, **21**, 567–576.
643 A. Hidalgo, A. Schliessmann and U. T. Bornscheuer, One-pot Simple methodology for CAssette Randomization and Recombination for focused directed evolution (OSCARR), *Methods Mol. Biol.*, 2014, **1179**, 207–212.
644 K. Kashiwagi, Y. Isogai, K. Nishiguchi and K. Shiba, Frame shuffling: a novel method for *in vitro* protein evolution, *Protein Eng., Des. Sel.*, 2006, **19**, 135–140.
645 J. E. Ness, S. Kim, A. Gottman, R. Pak, A. Krebber, T. V. Borchert, S. Govindarajan, E. C. Mundorff and J. Minshull, Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently, *Nat. Biotechnol.*, 2002, **20**, 1251–1255.
646 P. L. Bergquist, R. A. Reeves and M. D. Gibbs, Degenerate oligonucleotide gene shuffling (DOGS) and random drift mutagenesis (RNDM): two complementary techniques for enzyme evolution, *Biomol. Eng.*, 2005, **22**, 63–72.
647 B. R. Villiers, V. Stein and F. Hollfelder, USER friendly DNA recombination (USERec): a simple and flexible near homology-independent method for gene library construction, *Protein Eng., Des. Sel.*, 2010, **23**, 1–8.
648 B. Villiers and F. Hollfelder, USER friendly DNA recombination (USERec): gene library construction requiring minimal sequence homology, *Methods Mol. Biol.*, 2014, **1179**, 213–224.
649 P. E. O'Maille, M. Bakhtina and M. D. Tsai, Structure-based combinatorial protein engineering (SCOPE), *J. Mol. Biol.*, 2002, **321**, 677–691.
650 P. E. O'Maille, M. D. Tsai, B. T. Greenhagen, J. Chappell and J. P. Noel, Gene library synthesis by structure-based combinatorial protein engineering, *Methods Enzymol.*, 2004, **388**, 75–91.
651 M. Dokarry, C. Laurendon and P. E. O'Maille, Automating gene library synthesis by structure-based combinatorial protein engineering: examples from plant sesquiterpene synthases, *Methods Enzymol.*, 2012, **515**, 21–42.
652 A. Herman and D. S. Tawfik, Incorporating Synthetic Oligonucleotides *via* Gene Reassembly (ISOR): a versatile tool for generating targeted libraries, *Protein Eng., Des. Sel.*, 2007, **20**, 219–226.
653 L. Rockah-Shmuel, D. S. Tawfik and M. Goldsmith, Generating targeted libraries by the combinatorial incorporation of synthetic oligonucleotides during gene shuffling (ISOR), *Methods Mol. Biol.*, 2014, **1179**, 129–137.
654 M. D. Hughes, Z. R. Zhang, A. J. Sutherland, A. F. Santos and A. V. Hine, Discovery of active proteins directly from combinatorial randomized protein libraries without display, purification or sequencing: identification of novel zinc finger proteins, *Nucleic Acids Res.*, 2005, **33**, e32.
655 Z. Shao, H. Zhao and H. Zhao, DNA assembler, an *in vivo* genetic method for rapid construction of biochemical pathways, *Nucleic Acids Res.*, 2009, **37**, e16.
656 Z. Shao, Y. Luo and H. Zhao, Rapid characterization and engineering of natural product biosynthetic pathways *via* DNA assembler, *Mol. BioSyst.*, 2011, **7**, 1056–1059.
657 M. Z. Li and S. J. Elledge, Harnessing homologous recombination *in vitro* to generate recombinant DNA *via* SLIC, *Nat. Methods*, 2007, **4**, 251–256.

658 J. A. Mosberg, C. J. Gregg, M. J. Lajoie, H. H. Wang and G. M. Church, Improving lambda red genome engineering in *Escherichia colivia* rational removal of endogenous nucleases, *PLoS One*, 2012, **7**, e44638.

659 E. M. Nordwald, A. Garst, R. T. Gill and J. L. Kaar, Accelerated protein engineering for chemical biotechnology *via* homologous recombination, *Curr. Opin. Biotechnol.*, 2013, **24**, 1017–1022.

660 Z. Qian and S. Lutz, Improving the catalytic activity of *Candida antarctica* lipase B by circular permutation, *J. Am. Chem. Soc.*, 2005, **127**, 13466–13467.

661 Z. Qian, C. J. Fields and S. Lutz, Investigating the structural and functional consequences of circular permutation on lipase B from *Candida antarctica*, *ChemBioChem*, 2007, **8**, 1989–1996.

662 S. Lutz, A. B. Daugherty, Y. Yu and Z. Qian, Generating random circular permutation libraries, *Methods Mol. Biol.*, 2014, **1179**, 245–258.

663 E. Fischereder, D. Pressnitz, W. Kroutil and S. Lutz, Engineering strictosidine synthase: Rational design of a small, focused circular permutation library of the beta-propeller fold enzyme, *Bioorg. Med. Chem.*, 2014, **22**, 5633–5637.

664 A. B. Daugherty, S. Govindarajan and S. Lutz, Improved biocatalysts from a synthetic circular permutation library of the flavin-dependent oxidoreductase old yellow enzyme, *J. Am. Chem. Soc.*, 2013, **135**, 14425–14432.

665 Y. Yu and S. Lutz, Circular permutation: a different way to engineer enzyme structure and function, *Trends Biotechnol.*, 2011, **29**, 18–25.

666 E. Haglund, M. O. Lindberg and M. Oliveberg, Changes of protein folding pathways by circular permutation. Overlapping nuclei promote global cooperativity, *J. Biol. Chem.*, 2008, **283**, 27904–27915.

667 M. Lindberg, J. Tangrot and M. Oliveberg, Complete change of the protein folding transition state upon circular permutation, *Nat. Struct. Biol.*, 2002, **9**, 818–822.

668 M. A. Smith, P. A. Romero, T. Wu, E. M. Brustad and F. H. Arnold, Chimeragenesis of distantly-related proteins by noncontiguous recombination, *Protein Sci.*, 2013, **22**, 231–238.

669 D. L. Trudeau, M. A. Smith and F. H. Arnold, Innovation by homologous recombination, *Curr. Opin. Chem. Biol.*, 2013, **17**, 902–909.

670 M. C. Saraf, A. Gupta and C. D. Maranas, Design of combinatorial protein libraries of optimal size, *Proteins*, 2005, **60**, 769–777.

671 R. J. Pantazes, M. C. Saraf and C. D. Maranas, Optimal protein library design using recombination or point mutations based on sequence-based scoring functions, *Protein Eng., Des. Sel.*, 2007, **20**, 361–373.

672 J. B. Endelman, J. J. Silberg, Z. G. Wang and F. H. Arnold, Site-directed protein recombination as a shortest-path problem, *Protein Eng., Des. Sel.*, 2004, **17**, 589–594.

673 C. A. Voigt, C. Martinez, Z. G. Wang, S. L. Mayo and F. H. Arnold, Protein building blocks preserved by recombination, *Nat. Struct. Biol.*, 2002, **9**, 553–558.

674 C. R. Otey, J. J. Silberg, C. A. Voigt, J. B. Endelman, G. Bandara and F. H. Arnold, Functional evolution and structural conservation in chimeric cytochromes p450: calibrating a structure-guided approach, *Chem. Biol.*, 2004, **11**, 309–318.

675 P. Heinzelman, C. D. Snow, M. A. Smith, X. Yu, A. Kannan, K. Boulware, A. Villalobos, S. Govindarajan, J. Minshull and F. H. Arnold, SCHEMA recombination of a fungal cellulase uncovers a single mutation that contributes markedly to stability, *J. Biol. Chem.*, 2009, **284**, 26229–26233.

676 P. Heinzelman, R. Komor, A. Kanaan, P. Romero, X. Yu, S. Mohler, C. Snow and F. Arnold, Efficient screening of fungal cellobiohydrolase class I enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination, *Protein Eng., Des. Sel.*, 2010, **23**, 871–880.

677 P. Heinzelman, P. A. Romero and F. H. Arnold, Efficient sampling of SCHEMA chimera families to identify useful sequence elements, *Methods Enzymol.*, 2013, **523**, 351–368.

678 M. M. Meyer, J. J. Silberg, C. A. Voigt, J. B. Endelman, S. L. Mayo, Z. G. Wang and F. H. Arnold, Library analysis of SCHEMA-guided protein recombination, *Protein Sci.*, 2003, **12**, 1686–1693.

679 M. M. Meyer, L. Hochrein and F. H. Arnold, Structure-guided SCHEMA recombination of distantly related beta-lactamases, *Protein Eng., Des. Sel.*, 2006, **19**, 563–570.

680 P. A. Romero, E. Stone, C. Lamb, L. Chantranupong, A. Krause, A. E. Miklos, R. A. Hughes, B. Fechtel, A. D. Ellington, F. H. Arnold and G. Georgiou, SCHEMA-designed variants of human Arginase I and II reveal sequence elements important to stability and catalysis, *ACS Synth. Biol.*, 2012, **1**, 221–228.

681 J. J. Silberg, J. B. Endelman and F. H. Arnold, SCHEMA-guided protein recombination, *Methods Enzymol.*, 2004, **388**, 35–42.

682 M. A. Smith and F. H. Arnold, Designing libraries of chimeric proteins using SCHEMA recombination and RASPP, *Methods Mol. Biol.*, 2014, **1179**, 335–343.

683 M. A. Smith and F. H. Arnold, Noncontiguous SCHEMA protein recombination, *Methods Mol. Biol.*, 2014, **1179**, 345–352.

684 A. S. Parker, K. E. Griswold and C. Bailey-Kellogg, Optimization of Combinatorial Mutagenesis, *Research in Computational Molecular Biology*, 2011, **6577**, 321–335, 580.

685 H. Zhao, K. Blazanovic, Y. Choi, C. Bailey-Kellogg and K. E. Griswold, Gene and protein sequence optimization for high-level production of fully active and aglycosylated lysostaphin in *Pichia pastoris*, *Appl. Environ. Microbiol.*, 2014, **80**, 2746–2753.

686 L. He, A. M. Friedman and C. Bailey-Kellogg, Algorithms for optimizing cross-overs in DNA shuffling, *BMC Bioinf.*, 2012, **13**(suppl 3), S3.

687 W. Zheng, A. M. Friedman and C. Bailey-Kellogg, Algorithms for joint optimization of stability and diversity in planning combinatorial libraries of chimeric proteins, *J. Comput. Biol.*, 2009, **16**, 1151–1168.

688 X. Ye, A. M. Friedman and C. Bailey-Kellogg, Optimizing Bayes error for protein structure model selection by stability mutagenesis, *Comput. Syst. Bioinf., CSB2007 Conf. Proc., 6th*, 2008, **7**, 99–108.

689 W. Zheng, X. Ye, A. M. Friedman and C. Bailey-Kellogg, Algorithms for selecting breakpoint locations to optimize diversity in protein engineering by site-directed protein recombination, *Comput. Syst. Bioinf., CSB2007 Conf. Proc., 6th*, 2007, **6**, 31–40.

690 X. Ye, A. M. Friedman and C. Bailey-Kellogg, Hypergraph model of multi-residue interactions in proteins: sequentially-constrained partitioning algorithms for optimization of site-directed protein recombination, *J. Comput. Biol.*, 2007, **14**, 777–790.

691 L. Saftalov, P. A. Smith, A. M. Friedman and C. Bailey-Kellogg, Site-directed combinatorial construction of chimaeric genes: general method for optimizing assembly of gene fragments, *Proteins*, 2006, **64**, 629–642.

692 Y. Li, D. A. Drummond, A. M. Sawayama, C. D. Snow, J. D. Bloom and F. H. Arnold, A diverse family of thermostable cytochrome P450s created by recombination of stabilizing fragments, *Nat. Biotechnol.*, 2007, **25**, 1051–1056.

693 D. Lipovsek and A. Plückthun, *In vitro* protein evolution by ribosome display and mRNA display, *J. Immunol. Methods*, 2004, **290**, 51–67.

694 M. Y. He, Cell-free protein synthesis: applications in proteomics and biotechnology, *New Biotechnol.*, 2008, **25**, 126–132.

695 T. Okano, T. Matsuura, H. Suzuki and T. Yomo, Cell-free Protein Synthesis in a Microchamber Revealed the Presence of an Optimum Compartment Volume for High-order Reactions, *ACS Synth. Biol.*, 2014, **3**, 347–352.

696 T. Nishikawa, T. Sunami, T. Matsuura and T. Yomo, Directed Evolution of Proteins through *In Vitro* Protein Synthesis in Liposomes, *J. Nucleic Acids*, 2012, **2012**, 923214.

697 T. Okano, T. Matsuura, Y. Kazuta, H. Suzuki and T. Yomo, Cell-free protein synthesis from a single copy of DNA in a glass microchamber, *Lab Chip*, 2012, **12**, 2704–2711.

698 K. Nishimura, T. Matsuura, T. Sunami, H. Suzuki and T. Yomo, Cell-free protein synthesis inside giant unilamellar vesicles analyzed by flow cytometry, *Langmuir*, 2012, **28**, 8426–8432.

699 H. Yanagida, T. Matsuura and T. Yomo, Ribosome display for rapid protein evolution by consecutive rounds of mutation and selection, *Methods Mol. Biol.*, 2010, **634**, 257–267.

700 M. T. Smith, K. M. Wilding, J. M. Hunt, A. M. Bennett and B. C. Bundy, The emerging age of cell-free synthetic biology, *FEBS Lett.*, 2014, **588**, 2755–2761.

701 M. H. Caruthers, A. D. Barone, S. L. Beaucage, D. R. Dodds, E. F. Fisher, L. J. McBride, M. Matteucci, Z. Stabinsky and J. Y. Tang, Chemical synthesis of deoxyoligonucleotides by the phosphoramidite method, *Methods Enzymol.*, 1987, **154**, 287–313.

702 J. D. Tian, K. S. Ma and I. Saaem, Advancing high-throughput gene synthesis technology, *Mol. BioSyst.*, 2009, **5**, 714–722.

703 E. M. LeProust, B. J. Peck, K. Spirin, H. B. McCuen, B. Moore, E. Namsaraev and M. H. Caruthers, Synthesis of high-quality libraries of long (150mer) oligonucleotides by a novel depurination controlled process, *Nucleic Acids Res.*, 2010, **38**, 2522–2540.

704 K. E. Richmond, M. H. Li, M. J. Rodesch, M. Patel, A. M. Lowe, C. Kim, L. L. Chu, N. Venkataramaian, S. F. Flickinger, J. Kaysen, P. J. Belshaw, M. R. Sussman and F. Cerrina, Amplification and assembly of chip-eluted DNA (AACED): a method for high-throughput gene synthesis, *Nucleic Acids Res.*, 2004, **32**, 5011–5018.

705 A. Y. Borovkov, A. V. Loskutov, M. D. Robida, K. M. Day, J. A. Cano, T. Le Olson, H. Patel, K. Brown, P. D. Hunter and K. F. Sykes, High-quality gene assembly directly from unpurified mixtures of microarray-synthesized oligonucleotides, *Nucleic Acids Res.*, 2010, **38**, e180.

706 M. Matzas, P. F. Stähler, N. Kefer, N. Siebelt, V. Boisguérin, J. T. Leonard, A. Keller, C. F. Stähler, P. Häberle, B. Gharizadeh, F. Babrzadeh and G. M. Church, High-fidelity gene synthesis by retrieval of sequence-verified DNA identified using high-throughput pyrosequencing, *Nat. Biotechnol.*, 2010, **28**, 1291–1294.

707 H. Kim, H. Han, J. Ahn, J. Lee, N. Cho, H. Jang, H. Kim, S. Kwon and D. Bang, 'Shotgun DNA synthesis' for the high-throughput construction of large DNA molecules, *Nucleic Acids Res.*, 2012, **40**, e140.

708 G. M. Church, M. B. Elowitz, C. D. Smolke, C. A. Voigt and R. Weiss, Realizing the potential of synthetic biology, *Nat. Rev. Mol. Cell Biol.*, 2014, **15**, 289–294.

709 I. Tabuchi, S. Soramoto, S. Ueno and Y. Husimi, Multiline split DNA synthesis: a novel combinatorial method to make high quality peptide libraries, *BMC Biotechnol.*, 2004, **4**, 19.

710 J. Liang, Y. Luo and H. Zhao, Synthetic biology: putting synthesis into biology, *Wiley Interdiscip. Rev.: Syst. Biol. Med.*, 2011, **3**, 7–20.

711 S. A. Lynch and R. T. Gill, Synthetic biology: New strategies for directing design, *Metab. Eng.*, 2011, **14**, 205–211.

712 J. L. Foo, C. B. Ching, M. W. Chang and S. S. Leong, The imminent role of protein engineering in synthetic biology, *Biotechnol. Adv.*, 2012, **30**, 541–549.

713 D. W. Watkins, C. T. Armstrong and J. L. Anderson, *De novo* protein components for oxidoreductase assembly and biological integration, *Curr. Opin. Chem. Biol.*, 2014, **19**, 90–98.

714 S. Ma, N. Tang and J. Tian, DNA synthesis, assembly and applications in synthetic biology, *Curr. Opin. Chem. Biol.*, 2012, **16**, 260–267.

715 W. P. C. Stemmer, A. Crameri, K. D. Ha, T. M. Brennan and H. L. Heyneker, Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides, *Gene*, 1995, **164**, 49–53.

716 A. S. Xiong, Q. H. Yao, R. H. Peng, X. Li, H. Q. Fan, Z. M. Cheng and Y. Li, A simple, rapid, high-fidelity and cost-effective PCR-based two-step DNA synthesis method for long gene sequences, *Nucleic Acids Res.*, 2004, **32**, e98.

717 H. O. Smith, C. A. Hutchison, 3rd, C. Pfannkoch and J. C. Venter, Generating a synthetic genome by whole genome assembly: phiX, 174, bacteriophage from synthetic oligonucleotides, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 15440–15445.

718 G. H. Yang, S. Q. Wang, H. L. Wei, J. Ping, J. Liu, L. M. Xu and W. W. Zhang, Patch oligodeoxynucleotide synthesis (POS): a novel method for synthesis of long DNA sequences and full-length genes, *Biotechnol. Lett.*, 2012, **34**, 721–728.

719 S. de Kok, L. H. Stanton, T. Slaby, M. Durot, V. F. Holmes, K. G. Patel, D. Platt, E. B. Shapland, Z. Serber, J. Dean, J. D. Newman and S. S. Chandran, Rapid and reliable DNA assembly *via* ligase cycling reaction, *ACS Synth. Biol.*, 2014, **3**, 97–106.

720 T. L. Roth, L. Milenkovic and M. P. Scott, A rapid and simple method for DNA engineering using cycled ligation assembly, *PLoS One*, 2014, **9**, e107329.

721 J. Cherry, B. W. Nieuwenhuijsen, E. J. Kaftan, J. D. Kennedy and P. K. Chanda, A modified method for PCR-directed gene synthesis from large number of overlapping oligodeoxyribonucleotides, *J. Biochem. Biophys. Methods*, 2008, **70**, 820–822.

722 A. S. Xiong, R. H. Peng, J. Zhuang, F. Gao, Y. Li, Z. M. Cheng and Q. H. Yao, Chemical gene synthesis: strategies, softwares, error corrections, and applications, *FEMS Microbiol. Rev.*, 2008, **32**, 522–540.

723 B. F. Binkowski, K. E. Richmond, J. Kaysen, M. R. Sussman and P. J. Belshaw, Correcting errors in synthetic DNA through consensus shuffling, *Nucleic Acids Res.*, 2005, **33**, e55.

724 A. S. Xiong, Q. H. Yao, R. H. Peng, H. Duan, X. Li, H. Q. Fan, Z. M. Cheng and Y. Li, PCR-based accurate synthesis of long DNA sequences, *Nat. Protoc.*, 2006, **1**, 791–797.

725 P. A. Carr, J. S. Park, Y. J. Lee, T. Yu, S. Zhang and J. M. Jacobson, Protein-mediated error correction for *de novo* DNA synthesis, *Nucleic Acids Res.*, 2004, **32**, e162.

726 M. Fuhrmann, W. Oertel, P. Berthold and P. Hegemann, Removal of mismatched bases from synthetic genes by enzymatic mismatch cleavage, *Nucleic Acids Res.*, 2005, **33**, e58.

727 I. Saaem, S. Ma, J. Quan and J. Tian, Error correction of microchip synthesized genes using Surveyor nuclease, *Nucleic Acids Res.*, 2012, **40**, e23.

728 A. Currin, N. Swainston, P. J. Day and D. B. Kell, SpeedyGenes: a novel approach for the efficient production of error-corrected, synthetic gene libraries, *Protein Evol Design Sel*, 2014, **27**, 273–280.

729 N. Swainston, A. Currin, P. J. Day and D. B. Kell, GeneGenie: optimised oligomer design for directed evolution, *Nucleic Acids Res.*, 2014, **12**, W395–W400.

730 H. Lin and V. W. Cornish, Screening and selection methods for large-scale analysis of protein function, *Angew. Chem., Int. Ed.*, 2002, **41**, 4402–4425.

731 Y. L. Boersma, M. J. Dröge and W. J. Quax, Selection strategies for improved biocatalysts, *FEBS J.*, 2007, **274**, 2181–2195.

732 C. Troll, D. Alexander, J. Allen, J. Marquette and M. Camps, Mutagenesis and functional selection protocols for directed evolution of proteins in *E. coli*, *J. Visualized Exp.*, 2011, 49.

733 M. R. Parikh, D. N. Greene, K. K. Woods and I. Matsumura, Directed evolution of RuBisCO hypermorphs through genetic selection in engineered E.coli, *Protein Eng., Des. Sel.*, 2006, **19**, 113–119.

734 M. T. Reetz, H. Hobenreich, P. Soni and L. Fernandez, A genetic selection system for evolving enantioselectivity of enzymes, *Chem. Commun.*, 2008, 5502–5504.

735 C. G. Acevedo-Rocha, R. Agudo and M. T. Reetz, Directed evolution of stereoselective enzymes based on genetic selection as opposed to screening systems, *J. Biotechnol.*, 2014, **191C**, 3–10, DOI: 10.1016/j.jbiotec.2014.1004.1009.

736 Y. L. Boersma, M. J. Dröge, A. M. van der Sloot, T. Pijning, R. H. Cool, B. W. Dijkstra and W. J. Quax, A novel genetic selection system for improved enantioselectivity of *Bacillus subtilis* lipase A, *ChemBioChem*, 2008, **9**, 1110–1115.

737 P. Peralta-Yahya, B. T. Carter, H. Lin, H. Tao and V. W. Cornish, High-throughput selection for cellulase catalysts using chemical complementation, *J. Am. Chem. Soc.*, 2008, **130**, 17446–17452.

738 H. Tao, P. Peralta-Yahya, H. Lin and V. W. Cornish, Optimized design and synthesis of chemical dimerizer substrates for detection of glycosynthase activity *via* chemical complementation, *Bioorg. Med. Chem.*, 2006, **14**, 6940–6953.

739 S. K. Desai and J. P. Gallivan, Genetic screens and selections for small molecules based on a synthetic riboswitch that activates protein translation, *J. Am. Chem. Soc.*, 2004, **126**, 13247–13254.

740 W. C. Winkler and R. R. Breaker, Regulation of bacterial gene expression by riboswitches, *Annu. Rev. Microbiol.*, 2005, **59**, 487–517.

741 Y. Nomura and Y. Yokobayashi, Reengineering a natural riboswitch by dual genetic selection, *J. Am. Chem. Soc.*, 2007, **129**, 13814–13815.

742 N. Dixon, J. N. Duncan, T. Geerlings, M. S. Dunstan, J. E. McCarthy, D. Leys and J. Micklefield, Reengineering orthogonally selective riboswitches, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 2830–2835.

743 N. Dixon, C. J. Robinson, T. Geerlings, J. N. Duncan, S. P. Drummond and J. Micklefield, Orthogonal riboswitches for tuneable coexpression in bacteria, *Angew. Chem., Int. Ed.*, 2012, **51**, 3620–3624.

744 L. M. Wingler and V. W. Cornish, A library approach for the discovery of customized yeast three-hybrid counter selections, *ChemBioChem*, 2011, **12**, 715–717.

745 Y. Yokobayashi and F. H. Arnold, A dual selection module for directed evolution of genetic circuits, *Nat. Comput.*, 2005, **4**, 245–254.

746 W. Besenmatter, P. Kast and D. Hilvert, New enzymes from combinatorial library modules, *Methods Enzymol.*, 2004, **388**, 91–102.

747 E. G. Hibbert and P. A. Dalby, Directed evolution strategies for improved enzymatic performance, *Microb. Cell Fact.*, 2005, **4**, 29.
748 M. M. Müller, H. Kries, E. Csuhai, P. Kast and D. Hilvert, Design, selection, and characterization of a split chorismate mutase, *Protein Sci.*, 2010, **19**, 1000–1010.
749 K. Lanthaler, E. Bilsland, P. Dobson, H. J. Moss, P. Pir, D. B. Kell and S. G. Oliver, Genome-wide assessment of the carriers involved in the cellular uptake of drugs: a model system in yeast, *BMC Biol.*, 2011, **9**, 70.
750 G. E. Winter, B. Radic, C. Mayor-Ruiz, V. A. Blomen, C. Trefzer, R. K. Kandasamy, K. V. M. Huber, M. Gridling, D. Chen, T. Klampfl, R. Kralovics, S. Kubicek, O. Fernandez-Capetillo, T. R. Brummelkamp and G. Superti-Furga, The solute carrier SLC35F2 enables YM155-mediated DNA damage toxicity, *Nat. Chem. Biol.*, 2014, **10**, 768–773.
751 M. L. Geddie, L. A. Rowe, O. B. Alexander and I. Matsumura, High throughput microplate screens for directed protein evolution, *Methods Enzymol.*, 2004, **388**, 134–145.
752 G. An, J. Bielich, R. Auerbach and E. A. Johnson, Isolation and characterization of carotenoid hyperproducing mutants of yeast by flow cytometry and cell sorting, *Bio/Technology*, 1991, **9**, 69–73.
753 T. Azuma, G. I. Harrison and A. L. Demain, Isolation of gramicidin S hyperproducing strain of *Bacillus brevis* by use of a fluorescence activated cell sorting system., *Appl. Microbiol. Biotechnol.*, 1992, **38**, 173–178.
754 B. P. Cormack, R. H. Valdivia and S. Falkow, FACS-optimized mutants of the green fluorescent protein (GFP), *Gene*, 1996, **173**, 33–38.
755 M. Reckermann, Flow sorting in aquatic ecology, *Sci. Mar.*, 2000, **64**, 235–246.
756 C. J. Hewitt and G. Nebe-Von-Caron, An industrial application of multiparameter flow cytometry: assessment of cell physiological state and its application to the study of microbial fermentations, *Cytometry*, 2001, **44**, 179–187.
757 M. Rieseberg, C. Kasper, K. F. Reardon and T. Scheper, Flow cytometry in biotechnology, *Appl. Microbiol. Biotechnol.*, 2001, **56**, 350–360.
758 J. Vidal-Mas, P. Resina, E. Haba, J. Comas, A. Manresa and J. Vives-Rego, Rapid flow cytometry–Nile red assessment of PHA cellular content and heterogeneity in cultures of Pseudomonas aeruginosa 47T2 (NCIB 40044) grown in waste frying oil, *Antonie van Leeuwenhoek*, 2001, **80**, 57–63.
759 S. W. Santoro and P. G. Schultz, Directed evolution of the site specificity of Cre recombinase, *Proc. Natl. Acad. Sci. U. S. A.*, 2002, **99**, 4185–4190.
760 K. Bernath, M. Hai, E. Mastrobattista, A. D. Griffiths, S. Magdassi and D. S. Tawfik, *In vitro* compartmentalization by double emulsions: sorting and gene enrichment by fluorescence activated cell sorting, *Anal. Biochem.*, 2004, **325**, 151–157.
761 A. R. Buskirk, Y. C. Ong, Z. J. Gartner and D. R. Liu, Directed evolution of ligand dependence: small-molecule-activated protein splicing, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 10505–10510.
762 C. J. Hewitt and G. Nebe-Von-Caron, The application of multi-parameter flow cytometry to monitor individual microbial cell physiological state, *Adv. Biochem. Eng./Biotechnol.*, 2004, **89**, 197–223.
763 A. Aharoni, K. Thieme, C. P. Chiu, S. Buchini, L. L. Lairson, H. Chen, N. C. Strynadka, W. W. Wakarchuk and S. G. Withers, High-throughput screening methodology for the directed evolution of glycosyltransferases, *Nat. Methods*, 2006, **3**, 609–614.
764 H. M. Davey and D. B. Kell, Flow cytometry and cell sorting of heterogeneous microbial populations: the importance of single-cell analysis, *Microbiol. Rev.*, 1996, **60**, 641–696.
765 D. Mattanovich and N. Borth, Applications of cell sorting in biotechnology, *Microb. Cell Fact.*, 2006, **5**, 12.
766 O. J. Miller, K. Bernath, J. J. Agresti, G. Amitai, B. T. Kelly, E. Mastrobattista, V. Taly, S. Magdassi, D. S. Tawfik and A. D. Griffiths, Directed evolution by *in vitro* compartmentalization, *Nat. Methods*, 2006, **3**, 561–570.
767 M. Valli, M. Sauer, P. Branduardi, N. Borth, D. Porro and D. Mattanovich, Improvement of lactic acid production in *Saccharomyces cerevisiae* by cell sorting for high intracellular pH, *Appl. Environ. Microbiol.*, 2006, **72**, 5492–5499.
768 S. Becker, H. Hobenreich, A. Vogel, J. Knorr, S. Wilhelm, F. Rosenau, K. E. Jaeger, M. T. Reetz and H. Kolmar, Single-cell high-throughput screening to identify enantioselective hydrolytic enzymes, *Angew. Chem., Int. Ed.*, 2008, **47**, 5085–5088.
769 N. Varadarajan, S. Rodriguez, B. Y. Hwang, G. Georgiou and B. L. Iverson, Highly active and selective endopeptidases with programmed substrate specificities, *Nat. Chem. Biol.*, 2008, **4**, 290–294.
770 L. Liu, Y. Li, D. Liotta and S. Lutz, Directed evolution of an orthogonal nucleoside analog kinase *via* fluorescence-activated cell sorting, *Nucleic Acids Res.*, 2009, **37**, 4472–4481.
771 G. Yang and S. G. Withers, Ultrahigh-throughput FACS-based screening for directed enzyme evolution, *ChemBioChem*, 2009, **10**, 2704–2715.
772 M. Díaz, M. Herrero, L. A. García and C. Quirós, Application of flow cytometry to industrial microbial bioprocesses, *Biochem. Eng. J.*, 2010, **48**, 385–407.
773 J. A. Dietrich, A. E. McKee and J. D. Keasling, High-throughput metabolic engineering: advances in small-molecule screening and selection, *Annu. Rev. Biochem.*, 2010, **79**, 563–590.
774 S. Iijima, Y. Shimomura, Y. Haba, F. Kawai, A. Tani and K. Kimbara, Flow cytometry-based method for isolating live bacteria with meta-cleavage activity on dihydroxy compounds of biphenyl, *J. Biosci. Bioeng.*, 2010, **109**, 645–651.
775 S. Ishii, K. Tago and K. Senoo, Single-cell analysis and isolation for microbiology and biotechnology: methods and applications, *Appl. Microbiol. Biotechnol.*, 2010, **86**, 1281–1292.
776 M. E. Lidstrom and M. C. Konopka, The role of physiological heterogeneity in microbial population behavior, *Nat. Chem. Biol.*, 2010, **6**, 705–712.

777 S. W. Lim and A. R. Abate, Ultrahigh-throughput sorting of microfluidic drops with flow cytometry, *Lab Chip*, 2013, **13**, 4563–4572.
778 S. Müller and G. Nebe-von-Caron, Functional single-cell analyses: flow cytometry and cell sorting of microbial populations and communities, *FEMS Microbiol. Rev.*, 2010, **34**, 554–587.
779 S. G. Rhee, T. S. Chang, W. Jeong and D. Kang, Methods for detection and measurement of hydrogen peroxide inside and outside of cells, *Mol. Cells*, 2010, **29**, 539–549.
780 G. Stadlmayr, K. Benakovitsch, B. Gasser, D. Mattanovich and M. Sauer, Genome-scale analysis of library sorting (GALibSo): Isolation of secretion enhancing factors for recombinant protein production in *Pichia pastoris*, *Biotechnol. Bioeng.*, 2010, **105**, 543–555.
781 W. Throndset, S. Kim, B. Bower, S. Lantz, B. Kelemen, M. Pepsin, N. Chow, C. Mitchinson and M. Ward, Flow cytometric sorting of the filamentous fungus *Trichoderma reesei* for improved strains, *Enzyme Microb. Technol.*, 2010, **47**, 335–341.
782 W. Throndset, B. Bower, R. Caguiat, T. Baldwin and M. Ward, Isolation of a strain of *Trichoderma reesei* with improved glucoamylase secretion by flow cytometric sorting, *Enzyme Microb. Technol.*, 2010, **47**, 342–347.
783 B. P. Tracy, S. M. Gaida and E. T. Papoutsakis, Flow cytometry for bacteria: enabling metabolic engineering, synthetic biology and the elucidation of complex phenotypes, *Curr. Opin. Biotechnol.*, 2010, **21**, 85–99.
784 Y. J. Eun, A. S. Utada, M. F. Copeland, S. Takeuchi and D. B. Weibel, Encapsulating bacteria in agarose microparticles using microfluidics for high-throughput cell analysis and isolation, *ACS Chem. Biol.*, 2011, **6**, 260–266.
785 E. Fernández-Álvaro, R. Snajdrova, H. Jochens, T. Davids, D. Böttcher and U. T. Bornscheuer, A combination of *in vivo* selection and cell sorting for the identification of enantioselective biocatalysts, *Angew. Chem., Int. Ed.*, 2011, **50**, 8584–8587.
786 T. T. Y. Doan, B. Sivaloganathan and J. P. Obbard, Screening of marine microalgae for biodiesel feedstock, *Biomass Bioenergy*, 2011, **35**, 2534–2544.
787 S. Binder, G. Schendzielorz, N. Stäbler, K. Krumbach, K. Hoffmann, M. Bott and L. Eggeling, A high-throughput approach to identify genomic variants of bacterial metabolite producers at the single-cell level, *Genome Biol.*, 2012, **13**, R40.
788 R. Lönneborg, E. Varga and P. Brzezinski, Directed evolution of the transcriptional regulator DntR: isolation of mutants with improved DNT-response, *PLoS One*, 2012, **7**, e29994.
789 T. H. Yoo, M. Pogson, B. L. Iverson and G. Georgiou, Directed Evolution of Highly Selective Proteases by Using a Novel FACS-Based Screen that Capitalizes on the p53 Regulator MDM2, *ChemBioChem*, 2012, **13**, 649–653.
790 T. Lopes da Silva, J. C. Roseiro and A. Reis, Applications and perspectives of multi-parameter flow cytometry to microbial biofuels production processes, *Trends Biotechnol.*, 2012, **30**, 225–232.
791 M. Uttamchandani, X. Huang, G. Y. J. Chen and S. Q. Yao, Nanodroplet profiling of enzymatic activities in a microarray, *Bioorg. Med. Chem. Lett.*, 2005, **15**, 2135–2139.
792 A. A. Gordeev, T. R. Samatov, H. V. Chetverina and A. B. Chetverin, 2D format for screening bacterial cells at the throughput of flow cytometry, *Biotechnol. Bioeng.*, 2011, **108**, 2682–2690.
793 W. E. Huang, M. Li, R. M. Jarvis, R. Goodacre and S. A. Banwart, Shining light on the microbial world the application of Raman microspectroscopy, *Adv. Appl. Microbiol.*, 2010, **70**, 153–186.
794 L. Peng, G. Wang, W. Liao, H. Yao, S. Huang and Y. Q. Li, Intracellular ethanol accumulation in yeast cells during aerobic fermentation: a Raman spectroscopic exploration, *Lett. Appl. Microbiol.*, 2010, **51**, 632–638.
795 J. G. Lees and R. W. Janes, Combining sequence-based prediction methods and circular dichroism and infrared spectroscopic data to improve protein secondary structure determinations, *BMC Bioinf.*, 2008, **9**, 24.
796 T. van Rossum, S. W. Kengen and J. van der Oost, Reporter-based screening and selection of enzymes, *FEBS J.*, 2013, **280**, 2979–2996.
797 T. Uchiyama and K. Watanabe, The SIGEX scheme: high throughput screening of environmental metagenomes for the isolation of novel catabolic genes, *Biotechnol. Genet. Eng. Rev.*, 2007, **24**, 107–116.
798 T. Uchiyama and K. Watanabe, Substrate-induced gene expression (SIGEX) screening of metagenome libraries, *Nat. Protoc.*, 2008, **3**, 1202–1212.
799 T. Uchiyama and K. Miyazaki, Substrate-induced gene expression screening: a method for high-throughput screening of metagenome libraries, *Methods Mol. Biol.*, 2010, **668**, 153–168.
800 T. Uchiyama and K. Miyazaki, Product-induced gene expression, a product-responsive reporter assay used to screen metagenomic libraries for enzyme-encoding genes, *Appl. Environ. Microbiol.*, 2010, **76**, 7029–7035.
801 J. Dictenberg, Genetic encoding of fluorescent RNA ensures a bright future for visualizing nucleic acid dynamics, *Trends Biotechnol.*, 2012, **30**, 621–626.
802 J. R. van der Meer and S. Belkin, Where microbiology meets microengineering: design and applications of reporter bacteria, *Nat. Rev. Microbiol.*, 2010, **8**, 511–522.
803 M. N. Stojanovic, P. de Prada and D. W. Landry, Aptamer-based folding fluorescent sensor for cocaine, *J. Am. Chem. Soc.*, 2001, **123**, 4928–4931.
804 M. N. Stojanovic and D. M. Kolpashchikov, Modular aptameric sensors, *J. Am. Chem. Soc.*, 2004, **126**, 9266–9270.
805 K. Kikuchi, Design, synthesis and biological application of chemical probes for bio-imaging, *Chem. Soc. Rev.*, 2010, **39**, 2048–2053.
806 K. Kikuchi, Design, synthesis, and biological application of fluorescent sensor molecules for cellular imaging, *Adv. Biochem. Eng./Biotechnol.*, 2010, **119**, 63–78.

807 S. Okumoto, A. Jones and W. B. Frommer, Quantitative Imaging with Fluorescent Biosensors: Advanced Tools for Spatiotemporal Analysis of Biodynamics in Cells, *Annu. Rev. Plant Biol.*, 2012, **63**, 663–706.

808 S. Okumoto, Quantitative imaging using genetically encoded sensors for small molecules in plants, *Plant J.*, 2012, **70**, 108–117.

809 J. S. Paige, K. Y. Wu and S. R. Jaffrey, RNA mimics of green fluorescent protein, *Science*, 2011, **333**, 642–646.

810 J. S. Paige, T. Nguyen-Duc, W. Song and S. R. Jaffrey, Fluorescence imaging of cellular metabolites with RNA, *Science*, 2012, **335**, 1194.

811 R. L. Strack and S. R. Jaffrey, New approaches for sensing metabolites and proteins in live cells using RNA, *Curr. Opin. Chem. Biol.*, 2013, **17**, 651–655.

812 X. Xu, J. Zhang, F. Yang and X. Yang, Colorimetric logic gates for small molecules using split/integrated aptamers and unmodified gold nanoparticles, *Chem. Commun.*, 2011, **47**, 9435–9437.

813 R. Wombacher and V. W. Cornish, Chemical tags: applications in live cell fluorescence imaging, *J. Biophotonics*, 2011, **4**, 391–402.

814 C. Jing and V. W. Cornish, Chemical tags for labeling proteins inside living cells, *Acc. Chem. Res.*, 2011, **44**, 784–792.

815 *Chemical proteomics*, ed. G. Drewes and M. Bantscheff, Springer, Berlin, 2012.

816 A. P. Arkin and D. C. Youvan, Digital imaging spectroscopy, in *The photosynthetic reaction center*, ed. J. Deisenhofer and J. R. Norris, Academic Press, New York, 1993, vol. 1, pp. 133–155.

817 E. R. Goldman and D. C. Youvan, An algorithmically optimized combinatorial library screened by digital imaging spectroscopy, *Bio/Technology*, 1992, **10**, 1557–1561.

818 H. Joo, A. Arisawa, Z. L. Lin and F. H. Arnold, A high-throughput digital imaging screen for the discovery and directed evolution of oxygenases, *Chem. Biol.*, 1999, **6**, 699–706.

819 M. Alexeeva, A. Enright, M. J. Dawson, M. Mahmoudian and N. J. Turner, Deracemization of alpha-methylbenzylamine using an enzyme obtained by *in vitro* evolution, *Angew. Chem., Int. Ed.*, 2002, **41**, 3177–3180.

820 M. Alexeeva, R. Carr and N. J. Turner, Directed evolution of enzymes: new biocatalysts for asymmetric synthesis, *Org. Biomol. Chem.*, 2003, **1**, 4133–4137.

821 S. Delagrave, D. J. Murphy, J. L. Pruss, A. M. Maffia, 3rd, B. L. Marrs, E. J. Bylina, W. J. Coleman, C. L. Grek, M. R. Dilworth, M. M. Yang and D. C. Youvan, Application of a very high-throughput digital imaging screen to evolve the enzyme galactose oxidase, *Protein Eng.*, 2001, **14**, 261–267.

822 J. C. Weaver, Gel Microdroplets for Microbial Measurement and Screening: Basic Principles, *Biotechnol. Bioeng. Symp.*, 1987, **17**, 185–195.

823 J. C. Weaver, G. B. Williams, A. Klibanov and A. L. Demain, Gel Microdroplets: Rapid Detection and Enumeration of Individual Microorganisms by their Metabolic Activity, *Bio/Technology*, 1988, **6**, 1084–1089.

824 J. C. Weaver, J. G. Bliss, K. T. Powell, G. I. Harrison and G. B. Williams, Rapid Clonal Growth Measurements at the Single-Cell Level: Gel Microdroplets and Flow Cytometry, *Bio/Technology*, 1991, **9**, 873–876.

825 J. C. Weaver, J. G. Bliss, G. I. Harrison, K. T. Powell and G. B. Williams, Microdrop Technology: A General Method for Separating Cells by Function and Composition, *Methods*, 1991, **2**, 234–247.

826 D. S. Tawfik and A. D. Griffiths, Man-made cell-like compartments for molecular evolution, *Nat. Biotechnol.*, 1998, **16**, 652–656.

827 A. D. Griffiths and D. S. Tawfik, Man-made enzymes - from design to *in vitro* compartmentalisation, *Curr. Opin. Biotechnol.*, 2000, **11**, 338–353.

828 F. Courtois, L. F. Olguin, G. Whyte, A. B. Theberge, W. T. Huck, F. Hollfelder and C. Abell, Controlling the retention of small molecules in emulsion microdroplets for use in cell-based assays, *Anal. Chem.*, 2009, **81**, 3008–3016.

829 S. R. A. Devenish, M. Kaltenbach, M. Fischlechner and F. Hollfelder, Droplets as reaction compartments for protein nanotechnology, *Methods Mol. Biol.*, 2013, **996**, 269–286.

830 W. C. Lu and A. D. Ellington, *In vitro* selection of proteins *via* emulsion compartments, *Methods*, 2013, **60**, 75–80.

831 M. Fischlechner, Y. Schaerli, M. F. Mohamed, S. Patil, C. Abell and F. Hollfelder, Evolution of enzyme catalysts caged in biomimetic gel-shell beads, *Nat. Chem.*, 2014, **6**, 791–796.

832 A. Fallah-Araghi, J. C. Baret, M. Ryckelynck and A. D. Griffiths, A completely *in vitro* ultrahigh-throughput droplet-based microfluidic screening system for protein engineering and directed evolution, *Lab Chip*, 2012, **12**, 882–891.

833 A. Grünberger, N. Paczia, C. Probst, G. Schendzielorz, L. Eggeling, S. Noack, W. Wiechert and D. Kohlheyer, A disposable picolitre bioreactor for cultivation and investigation of industrially relevant bacteria on the single cell level, *Lab Chip*, 2012, **12**, 2060–2068.

834 I. Levin and A. Aharoni, Evolution in microfluidic droplet, *Chem. Biol.*, 2012, **19**, 929–931.

835 Y. P. Bai, E. Weibull, H. N. Joensson and H. Andersson-Svahn, Interfacing picoliter droplet microfluidics with addressable microliter compartments using fluorescence activated cell sorting, *Sens. Actuators, B*, 2014, **194**, 249–254.

836 F. Ma, Y. Xie, C. Huang, Y. Feng and G. Yang, An improved single cell ultrahigh throughput screening method based on *in vitro* compartmentalization, *PLoS One*, 2014, **9**, e89785.

837 L. Rosenfeld, T. Lin, R. Derda and S. K. Y. Tang, Review and analysis of performance metrics of droplet microfluidics systems, *Microfluid. Nanofluid.*, 2014, **16**, 921–939.

838 S. L. Sjostrom, Y. P. Bai, M. T. Huang, Z. H. Liu, J. Nielsen, H. N. Joensson and H. A. Svahn, High-throughput screening for industrial enzyme production hosts by droplet microfluidics, *Lab Chip*, 2014, **14**, 806–813.

839 B. Kintses, L. D. van Vliet, S. R. A. Devenish and F. Hollfelder, Microfluidic droplets: new integrated workflows for biological experiments, *Curr. Opin. Chem. Biol.*, 2010, **14**, 548–555.

840 B. Kintses, C. Hein, M. F. Mohamed, M. Fischlechner, F. Courtois, C. Laine and F. Hollfelder, Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution, *Chem. Biol.*, 2012, **19**, 1001–1009.

841 J. U. Shim, R. T. Ranasinghe, C. A. Smith, S. M. Ibrahim, F. Hollfelder, W. T. Huck, D. Klenerman and C. Abell, Ultrarapid generation of femtoliter microfluidic droplets for single-molecule-counting immunoassays, *ACS Nano*, 2013, **7**, 5955–5964.

842 C. A. Smith, X. Li, T. H. Mize, T. D. Sharpe, E. I. Graziani, C. Abell and W. T. S. Huck, Sensitive, high throughput detection of proteins in individual, surfactant-stabilized picoliter droplets using nanoelectrospray ionization mass spectrometry, *Anal. Chem.*, 2013, **85**, 3812–3816.

843 A. Zinchenko, S. R. Devenish, B. Kintses, P. Y. Colin, M. Fischlechner and F. Hollfelder, One in a million: flow cytometric sorting of single cell-lysate assays in monodisperse picolitre double emulsion droplets for directed evolution, *Anal. Chem.*, 2014, **86**, 2526–2533.

844 J. J. Agresti, E. Antipov, A. R. Abate, K. Ahn, A. C. Rowat, J. C. Baret, M. Marquez, A. M. Klibanov, A. D. Griffiths and D. A. Weitz, Ultrahigh-throughput screening in drop-based microfluidics for directed evolution, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 4004–4009.

845 J. Sacks, W. Welch, T. Mitchell and H. Wynn, Design and analysis of computer experiments (with discussion), *Statist Sci*, 1989, **4**, 409–435.

846 J. R. Koza, *Genetic programming: on the programming of computers by means of natural selection*, MIT Press, Cambridge, Mass, 1992.

847 J. R. Koza, *Genetic programming II: automatic discovery of reusable programs*, MIT Press, Cambridge, Mass, 1994.

848 T. Bäck, *Evolutionary algorithms in theory and practice*, Oxford University Press, Oxford, 1996.

849 W. B. Langdon, *Genetic programming and data structures: genetic programming+data structures=automatic programming!*, Kluwer, Boston, 1998.

850 *New ideas in optimization*, ed. D. Corne, M. Dorigo and F. Glover, McGraw Hill, London, 1999.

851 J. R. Koza, F. H. Bennett, M. A. Keane and D. Andre, *Genetic Programming III: Darwinian Invention and Problem Solving*, Morgan Kaufmann, San Francisco, 1999.

852 *Evolutionary Computation 1: basic algorithms and operators*, ed. T. Bäck, D. B. Fogel and Z. Michalewicz, IOP Publishing, Bristol, 2000.

853 *Evolutionary Computation 2: advanced algorithms and operators*, ed. T. Bäck, D. B. Fogel and Z. Michalewicz, IOP Publishing, Bristol, 2000.

854 W. B. Langdon and R. Poli, *Foundations of genetic programming*, Springer-Verlag, Berlin, 2002.

855 J. R. Koza, M. A. Keane and M. J. Streeter, Evolving inventions, *Sci Am*, 2003, **288**, 52–59.

856 J. R. Koza, M. A. Keane, M. J. Streeter, W. Mydlowec, J. Yu and G. Lanza, *Genetic programming: routine human-competitive machine intelligence*, Kluwer, New York, 2003.

857 J. Handl and J. Knowles, An evolutionary approach to multiobjective clustering, *IEEE Trans. Evol. Comput.*, 2007, **11**, 56–76.

858 R. Poli, W. B. Langdon and N. F. McPhee, A Field Guide to Genetic Programming, http://www.lulu.com/product/file-download/a-field-guide-to-genetic-programming/2502914, 2009.

859 D. R. Jones, M. Schonlau and W. J. Welch, Efficient global optimization of expensive black-box functions, *J. Global. Opt.*, 1998, **13**, 455–492.

860 W. M. Patrick, A. E. Firth and J. M. Blackburn, User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries, *Protein Eng.*, 2003, **16**, 451–457.

861 A. E. Firth and W. M. Patrick, Statistics of protein library construction, *Bioinformatics*, 2005, **21**, 3314–3315.

862 W. M. Patrick and A. E. Firth, Strategies and computational tools for improving randomized protein libraries, *Biomol. Eng.*, 2005, **22**, 105–112.

863 A. E. Firth and W. M. Patrick, GLUE-IT and PEDEL-AA: new programmes for analyzing protein diversity in randomized libraries, *Nucleic Acids Res.*, 2008, **36**, W281–W285.

864 J. Starrfelt and H. Kokko, Bet-hedging-a triple trade-off between means, variances and correlations, *Biol. Rev. Cambridge Philos. Soc.*, 2012, **87**, 742–755.

865 A. del Sol Mesa, F. Pazos and A. Valencia, Automatic methods for predicting functionally important residues, *J. Mol. Biol.*, 2003, **326**, 1289–1302.

866 S. Herrgard, S. A. Cammer, B. T. Hoffman, S. Knutson, M. Gallina, J. A. Speir, J. S. Fetrow and S. M. Baxter, Prediction of deleterious functional effects of amino acid mutations using a library of structure-based function descriptors, *Proteins*, 2003, **53**, 806–816.

867 G. Lopez, P. Maietta, J. M. Rodriguez, A. Valencia and M. L. Tress, firestar—advances in the prediction of functionally important residues, *J. Chem. Inf. Model.*, 2011, **39**, W235–W241.

868 T. A. Addington, R. W. Mertz, J. B. Siegel, J. M. Thompson, A. J. Fisher, V. Filkov, N. M. Fleischman, A. A. Suen, C. S. Zhang and M. D. Toney, Janus: Prediction and Ranking of Mutations Required for Functional Interconversion of Enzymes, *J. Mol. Biol.*, 2013, **425**, 1378–1389.

869 R. J. Fox and G. W. Huisman, Enzyme optimization: moving from blind evolution to statistical exploration of sequence-function space, *Trends Biotechnol.*, 2008, **26**, 132–138.

870 F. Liang, X. J. Feng, M. Lowry and H. Rabitz, Maximal use of minimal libraries through the adaptive substituent

reordering algorithm, *J. Phys. Chem. B*, 2005, **109**, 5842–5854.

871 S. R. McAllister, X. J. Feng, P. A. DiMaggio, Jr., C. A. Floudas, J. D. Rabinowitz and H. Rabitz, Descriptor-free molecular discovery in large libraries by adaptive substituent reordering, *Bioorg. Med. Chem. Lett.*, 2008, **18**, 5967–5970.

872 X. Feng, J. Sanchis, M. T. Reetz and H. Rabitz, Enhancing the efficiency of directed evolution in focused enzyme libraries by the adaptive substituent reordering algorithm, *Chemistry*, 2012, **18**, 5646–5654.

873 C. L. Araya and D. M. Fowler, Deep mutational scanning: assessing protein function on a massive scale, *Trends Biotechnol.*, 2011, **29**, 435–442.

874 N. Fischer, Sequencing antibody repertoires: the next generation, *MAbs*, 2011, **3**, 17–20.

875 F. Luciani, R. A. Bull and A. R. Lloyd, Next generation deep sequencing and vaccine design: today and tomorrow, *Trends Biotechnol.*, 2012, **30**, 443–452.

876 T. A. Whitehead, A. Chevalier, Y. Song, C. Dreyfus, S. J. Fleishman, C. De Mattos, C. A. Myers, H. Kamisetty, P. Blair, I. A. Wilson and D. Baker, Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing, *Nat. Biotechnol.*, 2012, **30**, 543–548.

877 P. Baldi and S. Brunak, *Bioinformatics: the machine learning approach*, MIT Press, Cambridge, MA, 1998.

878 *Machine learning and data mining. Methods and applications*, ed. R. S. Michalski, I. Bratko and M. Kubat, Wiley, Chichester, 1998.

879 T. M. Mitchell, *Machine learning*, McGraw Hill, New York, 1997.

880 B. G. Buchanan and E. A. Feigenbaum, DENDRAL and META-DENDRAL: their application dimensions, *Artif. Intell. Mater. Process., Proc. Int. Symp.*, 1978, **11**, 5–24.

881 B. G. Buchanan, E. A. Feigenbaum and J. Lederberg, On Gray interpretation of the DENDRAL project and programs - myth or mythunderstanding, *Chemom. Intell. Lab. Syst.*, 1988, **5**, 33–35.

882 E. A. Feigenbaum and B. G. Buchanan, DENDRAL and META-DENDRAL: Roots of knowledge systems and expert system applications, *Artif. Intell.*, 1993, **59**, 223–240.

883 J. Lederberg, How DENDRAL was conceived and born, ACM Symp Hist Med Informatics, 1987, http://profiles.nlm.nih.gov/ps/access/BBALYP.pdf.

884 R. K. Lindsay, B. G. Buchanan, E. A. Feigenbaum and J. Lederberg, DENDRAL – a Case study of the first expert system for scientific hypothesis formation, *Artif. Intell. Mater. Process., Proc. Int. Symp.*, 1993, **61**, 209–261.

885 J. Jonsson, T. Norberg, L. Carlsson, C. Gustafsson and S. Wold, Quantitative sequence-activity models (QSAM)—tools for sequence design, *J. Chem. Inf. Model.*, 1993, **21**, 733–739.

886 L. Breiman, Statistical modeling: The two cultures, *Stat Sci*, 2001, **16**, 199–215.

887 J. Liao, M. K. Warmuth, S. Govindarajan, J. E. Ness, R. P. Wang, C. Gustafsson and J. Minshull, Engineering proteinase K using machine learning and synthetic genes, *BMC Biotechnol.*, 2007, **7**, 16.

888 G. Liang and Z. Li, Scores of generalized base properties for quantitative sequence-activity modelings for E. coli promoters based on support vector machine, *J. Mol. Graphics Modell.*, 2007, **26**, 269–281.

889 B. Petersen, T. N. Petersen, P. Andersen, M. Nielsen and C. Lundegaard, A generic method for assignment of reliability scores applied to solvent accessibility predictions, *BMC Struct. Biol.*, 2009, **9**, 51.

890 P. Zhou, X. Chen, Y. Wu and Z. Shang, Gaussian process: an alternative approach for QSAM modeling of peptides, *Amino Acids*, 2010, **38**, 199–212.

891 B. A. van den Berg, M. J. T. Reinders, M. Hulsman, L. Wu, H. J. Pel, J. A. Roubos and D. de Ridder, Exploring Sequence Characteristics Related to High-Level Production of Secreted Proteins in *Aspergillus niger*, *PLoS One*, 2012, **7**, e45869.

892 S. Vaidyanathan, D. I. Broadhurst, D. B. Kell and R. Goodacre, Explanatory optimisation of protein mass spectrometry *via* genetic search, *Anal. Chem.*, 2003, **75**, 6679–6686.

893 S. Vaidyanathan, D. B. Kell and R. Goodacre, Selective detection of proteins in mixtures using electrospray ionization mass spectrometry: influence of instrumental settings and implications for proteomics, *Anal. Chem.*, 2004, **76**, 5024–5032.

894 D. Wedge, S. J. Gaskell, S. Hubbard, D. B. Kell, K. W. Lau and C. Eyers, Peptide detectability following ESI mass spectrometry: prediction using genetic programming, in *GECCO 2007*, ed. D. Thierens *et al.*, ACM, New York, 2007, pp. 2219–2225.

895 L. Breiman, Random forests, *Machine Learning*, 2001, **45**, 5–32.

896 D. M. Fowler, J. J. Stephany and S. Fields, Measuring the activity of protein variants on a large scale using deep mutational scanning, *Nat. Protoc.*, 2014, **9**, 2267–2284.

897 D. M. Fowler and S. Fields, Deep mutational scanning: a new style of protein science, *Nat. Methods*, 2014, **11**, 801–807.

898 J. Cairns, J. Overbaugh and S. Miller, The origin of mutants, *Nature*, 1988, **335**, 142–145.

899 P. A. Romero, A. Krause and F. H. Arnold, Navigating the protein fitness landscape with Gaussian processes, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, E193–E201.

900 T. Keleti, *Basic enzyme kinetics*, Akadémiai Kiadó, Budapest, 1986.

901 A. Cornish-Bowden, *Fundamentals of enzyme kinetics*, Portland Press, London, 2nd edn, 1995.

902 A. Fersht, *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*, W.H. Freeman, San Francisco, 1999.

903 A. R. Fersht, Catalysis, binding and enzyme-substrate complementarity, *Proc. R. Soc. London, Ser. B*, 1974, **187**, 397–407.

904 W. P. Jencks, Binding energy, specificity, and enzymic catalysis: the Circe effect, *Adv. Enzymol. Relat. Areas Mol. Biol.*, 1975, **43**, 219–410.

905 A. Whitty, C. A. Fierke and W. P. Jencks, Role of binding energy with coenzyme A in catalysis by 3-oxoacid coenzyme A transferase, *Biochemistry*, 1995, **34**, 11678–11689.
906 K. Liebeton, A. Zonta, K. Schimossek, M. Nardini, D. Lang, B. W. Dijkstra, M. T. Reetz and K. E. Jaeger, Directed evolution of an enantioselective lipase, *Chem. Biol.*, 2000, **7**, 709–718.
907 L. Greiner, S. Laue, A. Liese and C. Wandrey, Continuous homogeneous asymmetric transfer hydrogenation of ketones: lessons from kinetics, *Chemistry*, 2006, **12**, 1818–1823.
908 R. J. Fox and M. D. Clay, Catalytic effectiveness, a measure of enzyme proficiency for industrial applications, *Trends Biotechnol.*, 2009, **27**, 137–140.
909 D. Porro, B. Gasser, T. Fossati, M. Maurer, P. Branduardi, M. Sauer and D. Mattanovich, Production of recombinant proteins and metabolites in yeasts: when are these systems better than bacterial production systems? *Appl. Microbiol. Biotechnol.*, 2011, **89**, 939–948.
910 J. Becker and C. Wittmann, Systems and synthetic metabolic engineering for amino acid production - the heartbeat of industrial strain development, *Curr. Opin. Biotechnol.*, 2012, **23**, 718–726.
911 R. Dach, J. H. J. Song, F. Roschangar, W. Samstag and C. H. Senanayake, The Eight Criteria Defining a Good Chemical Manufacturing Process, *Org. Process Res. Dev.*, 2012, **16**, 1697–1706.
912 J. R. Knowles and W. J. Albery, Perfection in enzyme catalysis – energetics of triosephosphate isomerase, *Acc. Chem. Res.*, 1977, **10**, 105–111.
913 B. G. Miller and R. Wolfenden, Catalytic proficiency: the unusual case of OMP decarboxylase, *Annu. Rev. Biochem.*, 2002, **71**, 847–885.
914 A. Radzicka and R. Wolfenden, A proficient enzyme, *Science*, 1995, **267**, 90–93.
915 B. G. Miller, A. M. Hassell, R. Wolfenden, M. V. Milburn and S. A. Short, Anatomy of a proficient enzyme: the structure of orotidine 5'-monophosphate decarboxylase in the presence and absence of a potential transition state analog, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 2011–2016.
916 A. Bar-Even, E. Noor, Y. Savir, W. Liebermeister, D. Davidi, D. S. Tawfik and R. Milo, The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters, *Biochemistry*, 2011, **50**, 4402–4410.
917 R. Milo and R. L. Last, Achieving diversity in the face of constraints: lessons from metabolism, *Science*, 2012, **336**, 1663–1667.
918 I. Schomburg, A. Chang, S. Placzek, C. Söhngen, M. Rother, M. Lang, C. Munaretto, S. Ulas, M. Stelzer, A. Grote, M. Scheer and D. Schomburg, BRENDA in 2013: integrated reactions, kinetic data, enzyme function data, improved disease classification: new options and contents in BRENDA, *Nucleic Acids Res.*, 2013, **41**, D764–D772.
919 U. Wittig, R. Kania, M. Golebiewski, M. Rey, L. Shi, L. Jong, E. Algaa, A. Weidemann, H. Sauer-Danzwith, S. Mir, O. Krebs, M. Bittkowski, E. Wetsch, I. Rojas and W. Müller, SABIO-RK–database for biochemical reaction kinetics, *J. Chem. Inf. Model.*, 2012, **40**, D790–D796.
920 H. Kacser and J. A. Burns, The molecular basis of dominance., *Genetics*, 1981, **97**, 639–666.
921 H. Kacser and J. A. Burns, The control of flux, in *Rate Control of Biological Processes. Symposium of the Society for Experimental Biology*, ed. D. D. Davies, Cambridge University Press, Cambridge, 1973, vol. 27, pp. 65–104.
922 D. B. Kell and H. V. Westerhoff, Metabolic control theory: its role in microbiology and biotechnology., *FEMS Microbiol. Rev.*, 1986, **39**, 305–320.
923 D. B. Kell and H. V. Westerhoff, Towards a rational approach to the optimization of flux in microbial biotransformations., *Trends Biotechnol.*, 1986, **4**, 137–142.
924 D. A. Fell, *Understanding the control of metabolism*, Portland Press, London, 1996.
925 R. Heinrich and S. Schuster, *The regulation of cellular systems.*, Chapman & Hall, New York, 1996.
926 C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Krebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman and G. J. Hughes, Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture, *Science*, 2010, **329**, 305–309.
927 Y. Suzuki, K. Asada, J. Miyazaki, T. Tomita, T. Kuzuyama and M. Nishiyama, Enhancement of the latent 3-isopropylmalate dehydrogenase activity of promiscuous homoisocitrate dehydrogenase by directed evolution, *Biochem. J.*, 2010, **431**, 401–410.
928 S. J. Benkovic and S. Hammes-Schiffer, A perspective on enzyme catalysis, *Science*, 2003, **301**, 1196–1202.
929 M. Garcia-Viloca, J. Gao, M. Karplus and D. G. Truhlar, How enzymes work: analysis by modern rate theory and computer simulations, *Science*, 2004, **303**, 186–195.
930 G. G. Hammes, S. J. Benkovic and S. Hammes-Schiffer, Flexibility, diversity, and cooperativity: pillars of enzyme catalysis, *Biochemistry*, 2011, **50**, 10422–10430.
931 T. C. Bruice, Computational approaches: Reaction trajectories, structures, and atomic motions. Enzyme reactions and proficiency, *Chem. Rev.*, 2006, **106**, 3119–3139.
932 J. L. Gao, S. H. Ma, D. T. Major, K. Nam, J. Z. Pu and D. G. Truhlar, Mechanisms and free energies of enzymatic reactions, *Chem. Rev.*, 2006, **106**, 3188–3209.
933 M. H. Olsson, W. W. Parson and A. Warshel, Dynamical contributions to enzyme catalysis: critical tests of a popular hypothesis, *Chem. Rev.*, 2006, **106**, 1737–1756.
934 P. R. Carey, Spectroscopic characterization of distortion in enzyme complexes, *Chem. Rev.*, 2006, **106**, 3043–3054.
935 A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. B. Liu and M. H. M. Olsson, Electrostatic basis for enzyme catalysis, *Chem. Rev.*, 2006, **106**, 3210–3235.
936 A. V. Pisliakov, J. Cao, S. C. Kamerlin and A. Warshel, Enzyme millisecond conformational dynamics do not catalyze the chemical step, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 17359–17364.

937 S. C. Kamerlin and A. Warshel, At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins*, 2010, **78**, 1339–1375.
938 A. J. Adamczyk, J. Cao, S. C. Kamerlin and A. Warshel, Catalysis by dihydrofolate reductase and other enzymes arises from electrostatic preorganization, not conformational motions, *Proc. Natl. Acad. Sci. U. S. A.*, 2011, **108**, 14115–14120.
939 M. P. Frushicheva, M. J. Mills, P. Schopf, M. K. Singh, R. B. Prasad and A. Warshel, Computer aided enzyme design and catalytic concepts, *Curr. Opin. Chem. Biol.*, 2014, **21C**, 56–62.
940 Z. D. Nagel and J. P. Klinman, Tunneling and dynamics in enzymatic hydride transfer, *Chem. Rev.*, 2006, **106**, 3095–3118.
941 J. Z. Pu, J. L. Gao and D. G. Truhlar, Multidimensional tunneling, recrossing, and the transmission coefficient for enzymatic reactions, *Chem. Rev.*, 2006, **106**, 3140–3169.
942 S. Hay and N. S. Scrutton, Incorporation of hydrostatic pressure into models of hydrogen tunneling highlights a role for pressure-modulated promoting vibrations, *Biochemistry*, 2008, **47**, 9880–9887.
943 S. Hay, L. O. Johannissen, M. J. Sutcliffe and N. S. Scrutton, Barrier compression and its contribution to both classical and quantum mechanical aspects of enzyme catalysis, *Biophys. J.*, 2010, **98**, 121–128.
944 C. R. Pudney, L. O. Johannissen, M. J. Sutcliffe, S. Hay and N. S. Scrutton, Direct analysis of donor-acceptor distance and relationship to isotope effects and the force constant for barrier compression in enzymatic H-tunneling reactions, *J. Am. Chem. Soc.*, 2010, **132**, 11329–11335.
945 S. Hay and N. S. Scrutton, Good vibrations in enzyme-catalysed reactions, *Nat. Chem.*, 2012, **4**, 161–168.
946 S. Hay, L. O. Johannissen, P. Hothi, M. J. Sutcliffe and N. S. Scrutton, Pressure effects on enzyme-catalyzed quantum tunneling events arise from protein-specific structural and dynamic changes, *J. Am. Chem. Soc.*, 2012, **134**, 9749–9754.
947 M. Widersten, Protein engineering for development of new hydrolytic biocatalysts, *Curr. Opin. Chem. Biol.*, 2014, **21C**, 42–47.
948 M. Fuxreiter and L. Mones, The role of reorganization energy in rational enzyme design, *Curr. Opin. Chem. Biol.*, 2014, **21C**, 34–41.
949 B. Gavish and M. M. Werber, Viscosity-dependent structural fluctuations in enzyme catalysis, *Biochemistry*, 1979, **18**, 1269–1275.
950 B. Gavish, Position-dependent viscosity effects on rate coefficients, *Phys. Rev. Lett.*, 1980, **44**, 1160–1163.
951 D. Beece, L. Eisenstein, H. Frauenfelder, D. Good, M. C. Marden, L. Reinisch, A. H. Reynolds, L. B. Sorensen and K. T. Yue, Solvent viscosity and protein dynamics, *Biochemistry*, 1980, **19**, 5147–5157.
952 D. B. Kell, Enzymes As Energy Funnels, *Trends Biochem. Sci.*, 1982, **7**, 349.
953 G. R. Welch, B. Somogyi and S. Damjanovich, The role of protein fluctuations in enzyme action: a review, *Prog. Biophys. Mol. Biol.*, 1982, **39**, 109–146.
954 B. Somogyi, G. R. Welch and S. Damjanovich, The dynamic basis of energy transduction in enzymes, *Biochim. Biophys. Acta*, 1984, **768**, 81–112.
955 D. Vitkup, D. Ringe, G. A. Petsko and M. Karplus, Solvent mobility and the protein 'glass' transition, *Nat. Struct. Biol.*, 2000, **7**, 34–38.
956 R. M. Daniel, R. V. Dunn, J. L. Finney and J. C. Smith, The role of dynamics in enzyme activity, *Annu. Rev. Biophys. Biomol. Struct.*, 2003, **32**, 69–92.
957 J. E. Basner and S. D. Schwartz, How enzyme dynamics helps catalyze a reaction in atomic detail: a transition path sampling study, *J. Am. Chem. Soc.*, 2005, **127**, 13822–13831.
958 D. Antoniou, J. Basner, S. Nunez and S. D. Schwartz, Computational and theoretical methods to explore the relation between enzyme dynamics and catalysis, *Chem. Rev.*, 2006, **106**, 3170–3187.
959 R. Callender and R. B. Dyer, Advances in time-resolved approaches to characterize the dynamical nature of enzymatic catalysis, *Chem. Rev.*, 2006, **106**, 3031–3042.
960 I. J. Finkelstein, A. M. Massari and M. D. Fayer, Viscosity-dependent protein dynamics, *Biophys. J.*, 2007, **92**, 3652–3662.
961 H. Frauenfelder, P. W. Fenimore and R. D. Young, Protein dynamics and function: Insights from the energy landscape and solvent slaving, *IUBMB Life*, 2007, **59**, 506–512.
962 K. Henzler-Wildman and D. Kern, Dynamic personalities of proteins, *Nature*, 2007, **450**, 964–972.
963 K. A. Henzler-Wildman, M. Lei, V. Thai, S. J. Kerns, M. Karplus and D. Kern, A hierarchy of timescales in protein dynamics is linked to enzyme catalysis, *Nature*, 2007, **450**, 913–916.
964 K. A. Henzler-Wildman, V. Thai, M. Lei, M. Ott, M. Wolf-Watz, T. Fenn, E. Pozharski, M. A. Wilson, G. A. Petsko, M. Karplus, C. G. Hubner and D. Kern, Intrinsic motions along an enzymatic reaction trajectory, *Nature*, 2007, **450**, 838–844.
965 H. Frauenfelder, G. Chen, J. Berendzen, P. W. Fenimore, H. Jansson, B. H. McMahon, I. R. Stroe, J. Swenson and R. D. Young, A unified model of protein dynamics, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 5129–5134.
966 S. J. Benkovic, G. G. Hammes and S. Hammes-Schiffer, Free-energy landscape of enzyme catalysis, *Biochemistry*, 2008, **47**, 3317–3321.
967 R. K. Eppler, E. P. Hudson, S. D. Chase, J. S. Dordick, J. A. Reimer and D. S. Clark, Biocatalyst activity in non-aqueous environments correlates with centisecond-range protein motions, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 15672–15677.
968 D. D. Boehr, R. Nussinov and P. E. Wright, The role of dynamic conformational ensembles in biomolecular recognition, *Nat. Chem. Biol.*, 2009, **5**, 789–796.
969 S. D. Schwartz and V. L. Schramm, Enzymatic transition states and dynamic motion in barrier crossing, *Nat. Chem. Biol.*, 2009, **5**, 551–558.
970 I. Bahar, T. R. Lezon, L. W. Yang and E. Eyal, Global dynamics of proteins: bridging between structure and function, *Annu. Rev. Biophys.*, 2010, **39**, 23–42.

971 P. Csermely, R. Palotai and R. Nussinov, Induced fit, conformational selection and independent dynamic segments: an extended view of binding events, *Trends Biochem. Sci.*, 2010, **35**, 539–546.
972 A. E. Sitnitsky, Solvent viscosity dependence for enzymatic reactions, *Phys. A*, 2008, **387**, 5483–5497.
973 A. E. Sitnitsky, Model for solvent viscosity effect on enzymatic reactions, *Chem. Phys.*, 2010, **369**, 37–42.
974 G. Bhabha, J. Lee, D. C. Ekiert, J. Gam, I. A. Wilson, H. J. Dyson, S. J. Benkovic and P. E. Wright, A Dynamic Knockout Reveals That Conformational Fluctuations Influence the Chemical Step of Enzyme Catalysis, *Science*, 2011, **332**, 234–238.
975 M. Shushanyan, D. E. Khoshtariya, T. Tretyakova, M. Makharadze and R. van Eldik, Diverse role of conformational dynamics in carboxypeptidase A-driven peptide and ester hydrolyses: disclosing the ''perfect induced fit'' and ''protein local unfolding'' pathways by altering protein stability, *Biopolymers*, 2011, **95**, 852–870.
976 A. R. Jones, C. Levy, S. Hay and N. S. Scrutton, Relating localized protein motions to the reaction coordinate in coenzyme B_{12}-dependent enzymes, *FEBS J.*, 2013, **280**, 2997–3008.
977 *Dynamics in enzyme catalysis*, ed. J. P. Klinman and S. Hammes-Schiffer, Springer, Berlin, 2013.
978 K. Świderek, J. Javier Ruiz-Pernía, V. Moliner and I. Tuñon, Heavy enzymes-experimental and computational insights in enzyme dynamics, *Curr. Opin. Chem. Biol.*, 2014, **21C**, 11–18.
979 A. S. Davydov, Solitons and energy transfer along protein molecules, *J. Theor. Biol.*, 1977, **66**, 379–387.
980 A. S. Davydov, Excitons and solitons in molecular systems, *Int. Rev. Cytol.*, 1987, **106**, 183–225.
981 A. Ansari, J. Berendzen, S. F. Bowne, H. Frauenfelder, I. E. Iben, T. B. Sauke, E. Shyamsunder and R. D. Young, Protein states and proteinquakes, *Proc. Natl. Acad. Sci. U. S. A.*, 1985, **82**, 5000–5004.
982 G. Dadusc, J. P. Ogilvie, P. Schulenberg, U. Marvet and R. J. Miller, Diffractive optics-based heterodyne-detected four-wave mixing signals of protein motion: from ''protein quakes'' to ligand escape for myoglobin, *Proc. Natl. Acad. Sci. U. S. A.*, 2001, **98**, 6110–6115.
983 D. Arnlund, L. C. Johansson, C. Wickstrand, A. Barty, G. J. Williams, E. Malmerberg, J. Davidsson, D. Milathianaki, D. P. DePonte, R. L. Shoeman, D. Wang, D. James, G. Katona, S. Westenhoff, T. A. White, A. Aquila, S. Bari, P. Berntsen, M. Bogan, T. B. van Driel, R. B. Doak, K. S. Kjaer, M. Frank, R. Fromme, I. Grotjohann, R. Henning, M. S. Hunter, R. A. Kirian, I. Kosheleva, C. Kupitz, M. Liang, A. V. Martin, M. M. Nielsen, M. Messerschmidt, M. M. Seibert, J. Sjohamn, F. Stellato, U. Weierstall, N. A. Zatsepin, J. C. Spence, P. Fromme, I. Schlichting, S. Boutet, G. Groenhof, H. N. Chapman and R. Neutze, Visualizing a protein quake with time-resolved X-ray scattering at a free-electron laser, *Nat. Methods*, 2014, **11**, 923–926.
984 E. Fuglebakk, J. Echave and N. Reuter, Measuring and comparing structural fluctuation patterns in large protein datasets, *Bioinformatics*, 2012, **28**, 2431–2440.
985 J. E. Jimenez-Roldan, R. B. Freedman, R. A. Romer and S. A. Wells, Rapid simulation of protein motion: merging flexibility, rigidity and normal mode analyses, *Phys. Biol.*, 2012, **9**, 016008.
986 R. M. Pelis, W. M. Suhre and S. H. Wright, Functional influence of N-glycosylation in OCT2-mediated tetraethylammonium transport, *Am. J. Physiol.: Renal, Fluid Electrolyte Physiol.*, 2006, **290**, F1118–F1126.
987 G. M. Süel, S. W. Lockless, M. A. Wall and R. Ranganathan, Evolutionarily conserved networks of residues mediate allosteric communication in proteins, *Nat. Struct. Biol.*, 2003, **10**, 59–69.
988 K. F. Wong, T. Selzer, S. J. Benkovic and S. Hammes-Schiffer, Impact of distal mutations on the network of coupled motions correlated to hydride transfer in dihydrofolate reductase, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 6807–6812.
989 K. L. Morley and R. J. Kazlauskas, Improving enzyme properties: when are closer mutations better? *Trends Biotechnol.*, 2005, **23**, 231–237.
990 J. Paramesvaran, E. G. Hibbert, A. J. Russell and P. A. Dalby, Distributions of enzyme residues yielding mutants with improved substrate specificities from two different directed evolution strategies, *Protein Eng., Des. Sel.*, 2009, **22**, 401–411.
991 N. G. H. Leferink, S. V. Antonyuk, J. A. Houwman, N. S. Scrutton, R. R. Eady and S. S. Hasnain, Impact of residues remote from the catalytic centre on enzyme catalysis of copper nitrite reductase, *Nat. Commun.*, 2014, **5**, 4395.
992 R. Fasan, Y. T. Meharenna, C. D. Snow, T. L. Poulos and F. H. Arnold, Evolutionary history of a specialized P450 propane monooxygenase, *J. Mol. Biol.*, 2008, **383**, 1069–1080.
993 N. Preiswerk, T. Beck, J. D. Schulz, P. Milovnik, C. Mayer, J. B. Siegel, D. Baker and D. Hilvert, Impact of scaffold rigidity on the design and evolution of an artificial Diels-Alderase, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, 8013–8018.
994 X. Qi, Y. Chen, K. Jiang, W. Zuo, Z. Luo, Y. Wei, L. Du, H. Wei, R. Huang and Q. Du, Saturation-mutagenesis in two positions distant from active site of a Klebsiella pneumoniae glycerol dehydratase identifies some highly active mutants, *J. Biotechnol.*, 2009, **144**, 43–50.
995 D. L. Siehl, L. A. Castle, R. Gorton and R. J. Keenan, The molecular basis of glyphosate resistance by an optimized microbial acetyltransferase, *J. Biol. Chem.*, 2007, **282**, 11446–11455.
996 C. M. Cho, A. Mulchandani and W. Chen, Bacterial cell surface display of organophosphorus hydrolase for selective screening of improved hydrolysis of organophosphate nerve agents, *Appl. Environ. Microbiol.*, 2002, **68**, 2026–2030.

997 S. Oue, A. Okamoto, T. Yano and H. Kagamiyama, Redesigning the substrate specificity of an enzyme by cumulative effects of the mutations of non-active site residues, *J. Biol. Chem.*, 1999, **274**, 2344–2349.

998 K. Lindorff-Larsen, S. Piana, R. O. Dror and D. E. Shaw, How fast-folding proteins fold, *Science*, 2011, **334**, 517–520.

999 S. Piana, K. Sarkar, K. Lindorff-Larsen, M. Guo, M. Gruebele and D. E. Shaw, Computational design and experimental testing of the fastest-folding beta-sheet protein, *J. Mol. Biol.*, 2011, **405**, 43–48.

1000 S. Piana, K. Lindorff-Larsen and D. E. Shaw, Atomic-level description of ubiquitin folding, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 5915–5920.

1001 A. Raval, S. Piana, M. P. Eastwood, R. O. Dror and D. E. Shaw, Refinement of protein structure homology models *via* long, all-atom molecular dynamics simulations, *Proteins*, 2012, **80**, 2071–2079.

1002 D. E. Shaw, M. M. Deneroff, R. O. Dror, J. S. Kuskin, R. H. Larson, J. K. Salmon, C. Young, B. Batson, K. J. Bowers, J. C. Chao, M. P. Eastwood, J. Gagliardo, J. P. Grossman, C. R. Ho, D. J. Ierardi, I. Kolossvary, J. L. Klepeis, T. Layman, C. Mcleavey, M. A. Moraes, R. Mueller, E. C. Priest, Y. B. Shan, J. Spengler, M. Theobald, B. Towles and S. C. Wang, Anton, a special-purpose machine for molecular dynamics simulation, *Commun. ACM*, 2008, **51**, 91–97.

1003 T. Schwede, Protein modeling: what happened to the ''protein structure gap''? *Structure*, 2013, **21**, 1531–1540.

1004 K. A. Dill and J. L. MacCallum, The protein-folding problem, 50 years on, *Science*, 2012, **338**, 1042–1046.

1005 F. Khatib, F. Dimaio, S. Cooper, M. Kazmierczyk, M. Gilski, S. Krzywda, H. Zabranska, I. Pichova, J. Thompson, Z. Popovic, M. Jaskolski and D. Baker, Crystal structure of a monomeric retroviral protease solved by protein folding game players, *Nat. Struct. Mol. Biol.*, 2011, **18**, 1175–1177.

1006 E. H. Kellogg, O. F. Lange and D. Baker, Evaluation and optimization of discrete state models of protein folding, *J. Phys. Chem. B*, 2012, **116**, 11405–11413.

1007 D. S. Marks, T. A. Hopf and C. Sander, Protein structure prediction from sequence variation, *Nat. Biotechnol.*, 2012, **30**, 1072–1080.

1008 W. R. Taylor, D. T. Jones and M. I. Sadowski, Protein topology from predicted residue contacts, *Protein Sci.*, 2012, **21**, 299–305.

1009 D. T. Jones, D. W. Buchan, D. Cozzetto and M. Pontil, PSICOV: Precise structural contact prediction using sparse inverse covariance estimation on large multiple sequence alignments, *Bioinformatics*, 2012, **28**, 184–190.

1010 T. Nugent and D. T. Jones, Accurate *de novo* structure prediction of large transmembrane protein domains using fragment-assembly and correlated mutation analysis, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, E1540–E1547.

1011 T. Kosciolek and D. T. Jones, *De novo* structure prediction of globular proteins aided by sequence variation-derived contacts, *PLoS One*, 2014, **9**, e92197.

1012 C. Andreini, I. Bertini, G. Cavallaro, G. L. Holliday and J. M. Thornton, Metal ions in biological catalysis: from enzyme databases to general principles, *JBIC, J. Biol. Inorg. Chem.*, 2008, **13**, 1205–1218.

1013 D. B. Kell, Iron behaving badly: inappropriate iron chelation as a major contributor to the aetiology of vascular and other progressive inflammatory and degenerative diseases, *BMC Med. Genomics*, 2009, **2**, 2.

1014 D. B. Kell, Towards a unifying, systems biology understanding of large-scale cellular death and destruction caused by poorly liganded iron: Parkinson's, Huntington's, Alzheimer's, prions, bactericides, chemical toxicology and others as examples, *Arch. Toxicol.*, 2010, **577**, 825–889.

1015 D. B. Kell and E. Pretorius, Serum ferritin is an important disease marker, and is mainly a leakage product from damaged cells, *Metallomics*, 2014, **6**, 748–773.

1016 M. T. Reetz, J. J. Peyralans, A. Maichele, Y. Fu and M. Maywald, Directed evolution of hybrid enzymes: Evolving enantioselectivity of an achiral Rh-complex anchored to a protein, *Chem. Commun.*, 2006, 4318–4320.

1017 M. T. Reetz, M. Rentzsch, A. Pletsch, M. Maywald, P. Maiwald, J. J. P. Peyralans, A. Maichele, Y. Fu, N. Jiao, F. Hollmann, R. Mondière and A. Taglieber, Directed evolution of enantioselective hybrid catalysts: a novel concept in asymmetric catalysis, *Tetrahedron*, 2007, **63**, 6404–6414.

1018 M. T. Reetz, Directed evolution of selective enzymes and hybrid catalysts, *Tetrahedron*, 2002, **58**, 6595–6602.

1019 M. T. Reetz, M. Rentzsch, A. Pletsch and M. Maywald, Towards the directed evolution of hybrid catalysts, *Chimia*, 2002, **56**, 721–723.

1020 D. E. Benson, M. S. Wisz and H. W. Hellinga, Rational design of nascent metalloenzymes, *Proc. Natl. Acad. Sci. U. S. A.*, 2000, **97**, 6292–6297.

1021 J. D. Bridgewater, J. Lim and R. W. Vachet, Transition metal-peptide binding studied by metal-catalyzed oxidation reactions and mass spectrometry, *Anal. Chem.*, 2006, **78**, 2432–2438.

1022 A. K. Petros, A. R. Reddi, M. L. Kennedy, A. G. Hyslop and B. R. Gibney, Femtomolar Zn(II) affinity in a peptide-based ligand designed to model thiolate-rich metalloprotein active sites, *Inorg. Chem.*, 2006, **45**, 9941–9958.

1023 A. Mantion, L. Massuger, P. Rabu, C. Palivan, L. B. McCusker and A. Taubert, Metal-peptide frameworks (MPFs): ''bioinspired'' metal organic frameworks, *J. Am. Chem. Soc.*, 2008, **130**, 2517–2526.

1024 G. D. Pirngruber, L. Frunz and M. Lüchinger, The characterisation and catalytic properties of biomimetic metal-peptide complexes immobilised on mesoporous silica, *Phys. Chem. Chem. Phys.*, 2009, **11**, 2928–2938.

1025 J. T. Pedersen, K. Teilum, N. H. Heegaard, J. Ostergaard, H. W. Adolph and L. Hemmingsen, Rapid formation of a preoligomeric peptide-metal-peptide complex following copper(II) binding to amyloid beta peptides, *Angew. Chem., Int. Ed.*, 2011, **50**, 2532–2535.

1026 T. Tanaka, T. Mizuno, S. Fukui, H. Hiroaki, J. Oku, K. Kanaori, K. Tajima and M. Shirakawa, Two-metal ion, Ni(II) and Cu(II), binding alpha-helical coiled coil peptide, *J. Am. Chem. Soc.*, 2004, **126**, 14023–14028.

1027 C. Tamerler, D. Khatayevich, M. Gungormus, T. Kacar, E. E. Oren, M. Hnilova and M. Sarikaya, Molecular biomimetics: GEPI-based biological routes to technology, *Biopolymers*, 2010, **94**, 78–94.

1028 R. L. Koder, J. L. Anderson, L. A. Solomon, K. S. Reddy, C. C. Moser and P. L. Dutton, Design and engineering of an O_2 transport protein, *Nature*, 2009, **458**, 305–309.

1029 A. F. Peacock, O. Iranzo and V. L. Pecoraro, Harnessing natures ability to control metal ion coordination geometry using *de novo* designed peptides, *Dalton Trans.*, 2009, 2271–2280.

1030 O. Iranzo, S. Chakraborty, L. Hemmingsen and V. L. Pecoraro, Controlling and Fine Tuning the Physical Properties of Two Identical Metal Coordination Sites in De Novo Designed Three Stranded Coiled Coil Peptides, *J. Am. Chem. Soc.*, 2011, **133**, 239–251.

1031 P. Braun, E. Goldberg, C. Negron, M. von Jan, F. Xu, V. Nanda, R. L. Koder and D. Noy, Design principles for chlorophyll-binding sites in helical proteins, *Proteins*, 2010, **79**, 463–476.

1032 F. V. Cochran, S. P. Wu, W. Wang, V. Nanda, J. G. Saven, M. J. Therien and W. F. DeGrado, Computational *de novo* design and characterization of a four-helix bundle protein that selectively binds a nonbiological cofactor, *J. Am. Chem. Soc.*, 2005, **127**, 1346–1347.

1033 K. Kuroda and M. Ueda, Molecular design of the microbial cell surface toward the recovery of metal ions, *Curr. Opin. Biotechnol.*, 2011, **22**, 427–433.

1034 J. M. González, M. R. Meini, P. E. Tomatis, F. J. Medrano Martín, J. A. Cricco and A. J. Vila, Metallo-beta-lactamases withstand low Zn(II) conditions by tuning metal-ligand interactions, *Nat. Chem. Biol.*, 2012, **8**, 698–700.

1035 I. Sóvágó, C. Kállay and K. Várnagy, Peptides as complexing agents: Factors influencing the structure and thermodynamic stability of peptide complexes, *Coord. Chem. Rev.*, 2012, **256**, 2225–2233.

1036 Y. Lu, N. Yeung, N. Sieracki and N. M. Marshall, Design of functional metalloproteins, *Nature*, 2009, **460**, 855–862.

1037 K. L. Harris, S. Lim and S. J. Franklin, Of folding and function: understanding active-site context through metalloenzyme design, *Inorg. Chem.*, 2006, **45**, 10002–10012.

1038 K. E. Sapsford, W. R. Algar, L. Berti, K. B. Gemmill, B. J. Casey, E. Oh, M. H. Stewart and I. L. Medintz, Functionalizing nanoparticles with biological molecules: developing chemistries that facilitate nanotechnology, *Chem. Rev.*, 2013, **113**, 1904–2074.

1039 T. Happe and A. Hemschemeier, Metalloprotein mimics - old tools in a new light, *Trends Biotechnol.*, 2014, **32**, 170–176.

1040 M. Dürrenberger and T. R. Ward, Recent achievments in the design and engineering of artificial metalloenzymes, *Curr. Opin. Chem. Biol.*, 2014, **19**, 99–106.

1041 J. Bos and G. Roelfes, Artificial metalloenzymes for enantioselective catalysis, *Curr. Opin. Chem. Biol.*, 2014, **19**, 135–143.

1042 I. D. Petrik, J. Liu and Y. Lu, Metalloenzyme design and engineering through strategic modifications of native protein scaffolds, *Curr. Opin. Chem. Biol.*, 2014, **19**, 67–75.

1043 A. J. Hickman and M. S. Sanford, High-valent organometallic copper and palladium in catalysis, *Nature*, 2012, **484**, 177–185.

1044 A. J. Reig, M. M. Pires, R. A. Snyder, Y. Wu, H. Jo, D. W. Kulp, S. E. Butch, J. R. Calhoun, T. Szyperski, E. I. Solomon and W. F. DeGrado, Alteration of the oxygen-dependent reactivity of *de novo* Due Ferri proteins, *Nat. Chem.*, 2012, **4**, 900–906.

1045 T. Mizuno, K. Murao, Y. Tanabe, M. Oda and T. Tanaka, Metal-ion-dependent GFP emission *in vivo* by combining a circularly permutated green fluorescent protein with an engineered metal-ion-binding coiled-coil, *J. Am. Chem. Soc.*, 2007, **129**, 11378–11383.

1046 W. Bae, A. Mulchandani and W. Chen, Cell surface display of synthetic phytochelatins using ice nucleation protein for enhanced heavy metal bioaccumulation, *J. Inorg. Biochem.*, 2002, **88**, 223–227.

1047 C. S. Cutler, H. M. Hennkens, N. Sisay, S. Huclier-Markai and S. S. Jurisson, Radiometals for combined imaging and therapy, *Chem. Rev.*, 2012, **113**, 858–883.

1048 R. Ferreirós-Martínez, D. Esteban-Gómez, C. Platas-Iglesias, A. de Blas and T. Rodríguez-Blas, Zn(II), Cd(II) and Pb(II) complexation with pyridinecarboxylate containing ligands, *Dalton Trans.*, 2008, 5754–5765.

1049 E. Boros, C. L. Ferreira, J. F. Cawthray, E. W. Price, B. O. Patrick, D. W. Wester, M. J. Adam and C. Orvig, Acyclic chelate with ideal properties for ^{68}Ga PET imaging agent elaboration, *J. Am. Chem. Soc.*, 2010, **132**, 15726–15733.

1050 S. R. Stürzenbaum, M. Hockner, A. Panneerselvam, J. Levitt, J. S. Bouillard, S. Taniguchi, L. A. Dailey, R. Ahmad Khanbeigi, E. V. Rosca, M. Thanou, K. Suhling, A. V. Zayats and M. Green, Biosynthesis of luminescent quantum dots in an earthworm, *Nat. Nanotechnol.*, 2013, **8**, 57–60.

1051 C. Lo, M. R. Ringenberg, D. Gnandt, Y. Wilson and T. R. Ward, Artificial metalloenzymes for olefin metathesis based on the biotin-(strept)avidin technology, *Chem. Commun.*, 2011, **47**, 12065–12067.

1052 T. R. Ward, Artificial metalloenzymes based on the biotin-avidin technology: enantioselective catalysis and beyond, *Acc. Chem. Res.*, 2011, **44**, 47–57.

1053 T. Heinisch and T. R. Ward, Design strategies for the creation of artificial metalloenzymes, *Curr. Opin. Chem. Biol.*, 2010, **14**, 184–199.

1054 Y. You, Phosphorescence bioimaging using cyclometalated Ir(III) complexes, *Curr. Opin. Chem. Biol.*, 2013, **17**, 699–707.

1055 T. K. Hyster, L. Knorr, T. R. Ward and T. Rovis, Biotinylated Rh(III) complexes in engineered streptavidin for

accelerated asymmetric C–H activation, *Science*, 2012, **338**, 500–503.

1056 S. V. Wegner, H. Boyaci, H. Chen, M. P. Jensen and C. He, Engineering a uranyl-specific binding protein from NikR, *Angew. Chem., Int. Ed.*, 2009, **48**, 2339–2341.

1057 L. Zhou, M. Bosscher, C. Zhang, S. özçubukçu, L. Zhang, W. Zhang, C. J. Li, J. Liu, M. P. Jensen, L. Lai and C. He, A protein engineered to bind uranyl selectively and with femtomolar affinity, *Nat. Chem.*, 2014, **6**, 236–241.

1058 P. Turner, G. Mamo and E. N. Karlsson, Potential and utilization of thermophiles and thermostable enzymes in biorefining, *Microb. Cell Fact.*, 2007, **6**, 9.

1059 P. S. Low and G. N. Somero, Temperature adaptation of enzymes: a proposed molecular basis for the different catalytic efficiencies of enzymes from ectotherms and endotherms, *Comp. Biochem. Physiol., Part B: Biochem. Mol. Biol.*, 1974, **49**, 307–312.

1060 P. L. Wintrode and F. H. Arnold, Temperature adaptation of enzymes: lessons from laboratory evolution, *Adv. Protein Chem.*, 2000, **55**, 161–225.

1061 R. D. Socha and N. Tokuriki, Modulating protein stability: directed evolution strategies for improved protein function, *FEBS J.*, 2013, **280**, 5582–5595.

1062 Y. Gumulya and M. T. Reetz, Enhancing the thermal robustness of an enzyme by directed evolution: least favorable starting points and inferior mutants can map superior evolutionary pathways, *ChemBioChem*, 2011, **12**, 2502–2510.

1063 R. L. Chang, K. Andrews, D. Kim, Z. Li, A. Godzik and B. Ø. Palsson, Structural systems biology evaluation of metabolic thermotolerance in *Escherichia coli*, *Science*, 2013, **340**, 1220–1223.

1064 M. Lehmann, D. Kostrewa, M. Wyss, R. Brugger, A. D'Arcy, L. Pasamontes and A. van Loon, From DNA sequence to improved functionality: using protein sequence comparisons to rapidly design a thermostable consensus phytase, *Protein Eng.*, 2000, **13**, 49–57.

1065 M. Lehmann, L. Pasamontes, S. F. Lassen and M. Wyss, The consensus concept for thermostability engineering of proteins, *Biochim. Biophys. Acta*, 2000, **1543**, 408–415.

1066 M. Lehmann and M. Wyss, Engineering proteins for thermostability: the use of sequence alignments *versus* rational design and directed evolution, *Curr. Opin. Biotechnol.*, 2001, **12**, 371–375.

1067 M. Lehmann, C. Loch, A. Middendorf, D. Studer, S. F. Lassen, L. Pasamontes, A. P. G. M. van Loon and M. Wyss, The consensus concept for thermostability engineering of proteins: further proof of concept, *Protein Eng.*, 2002, **15**, 403–411.

1068 R. G. Coleman and K. A. Sharp, Shape and evolution of thermostable protein structure, *Proteins*, 2010, **78**, 420–433.

1069 C. A. Hokanson, G. Cappuccilli, T. Odineca, M. Bozic, C. A. Behnke, M. Mendez, W. J. Coleman and R. Crea, Engineering highly thermostable xylanase variants using an enhanced combinatorial library method, *Protein Eng., Des. Sel.*, 2011, **24**, 597–605.

1070 J. P. Aucamp, A. M. Cosme, G. J. Lye and P. A. Dalby, High-throughput measurement of protein stability in microtiter plates, *Biotechnol. Bioeng.*, 2005, **89**, 599–607.

1071 J. P. Aucamp, R. J. Martinez-Torres, E. G. Hibbert and P. A. Dalby, A microplate-based evaluation of complex denaturation pathways: structural stability of *Escherichia coli* transketolase, *Biotechnol. Bioeng.*, 2008, **99**, 1303–1310.

1072 T. Schwab and R. Sterner, Stabilization of a metabolic enzyme by library selection in *Thermus thermophilus*, *ChemBioChem*, 2011, **12**, 1581–1588.

1073 C. Pfleger, S. Radestock, E. Schmidt and H. Gohlke, Global and local indices for characterizing biomolecular flexibility and rigidity, *J. Comput. Chem., Jpn*, 2013, **34**, 220–233.

1074 P. L. Wintrode, D. Zhang, N. Vaidehi, F. H. Arnold and W. A. Goddard 3rd, Protein dynamics in a family of laboratory evolved thermophilic enzymes, *J. Mol. Biol.*, 2003, **327**, 745–757.

1075 T. J. Kamerzell and C. R. Middaugh, The complex inter-relationships between protein flexibility and stability, *J. Pharm. Sci.*, 2008, **97**, 3494–3517.

1076 E. Bae, R. M. Bannen and G. N. Phillips, Bioinformatic method for protein thermal stabilization by structural entropy optimization, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 9594–9597.

1077 F. X. Schmid, Lessons about protein stability from *in vitro* selections, *ChemBioChem*, 2011, **12**, 1501–1507.

1078 E. Vázquez-Figueroa, J. Chaparro-Riggers and A. S. Bommarius, Development of a thermostable glucose dehydrogenase by a structure-guided consensus concept, *ChemBioChem*, 2007, **8**, 2295–2301.

1079 K. M. Polizzi, J. F. Chaparro-Riggers, E. Vazquez-Figueroa and A. S. Bommarius, Structure-guided consensus approach to create a more thermostable penicillin G acylase, *Biotechnol. J.*, 2006, **1**, 531–536.

1080 H. J. Wijma, R. J. Floor and D. B. Janssen, Structure- and sequence-analysis inspired engineering of proteins for enhanced thermostability, *Curr. Opin. Struct. Biol.*, 2013, **23**, 588–594.

1081 C. Vieille, D. S. Burdette and J. G. Zeikus, Thermozymes, *Biotechnol. Annu. Rev.*, 1996, **2**, 1–83.

1082 J. G. Zeikus, C. Vieille and A. Savchenko, Thermozymes: biotechnology and structure-function relationships, *Extremophiles*, 1998, **2**, 179–183.

1083 R. Maheshwari, G. Bharadwaj and M. K. Bhat, Thermophilic fungi: their physiology and enzymes, *Microbiol. Mol. Biol. Rev.*, 2000, **64**, 461–488.

1084 D. Sriprapundh, C. Vieille and J. G. Zeikus, Molecular determinants of xylose isomerase thermal stability and activity: analysis of thermozymes by site-directed mutagenesis, *Protein Eng.*, 2000, **13**, 259–265.

1085 C. Vieille and G. J. Zeikus, Hyperthermophilic enzymes: sources, uses, and molecular mechanisms for thermostability, *Microbiol. Mol. Biol. Rev.*, 2001, **65**, 1–43.

1086 M. E. Bruins, A. E. Janssen and R. M. Boom, Thermozymes and their applications: a review of recent literature and patents, *Appl. Biochem. Biotechnol.*, 2001, **90**, 155–186.

1087 W. F. Li, X. X. Zhou and P. Lu, Structural features of thermozymes, *Biotechnol. Adv.*, 2005, **23**, 271–281.
1088 L. C. Wu, J. X. Lee, H. D. Huang, B. J. Liu and J. T. Horng, An expert system to predict protein thermostability using decision trees, *Expert Syst. Appl.*, 2009, **36**, 668–674.
1089 I. N. Berezovsky, The diversity of physical forces and mechanisms in intermolecular interactions, *Phys. Biol.*, 2011, **8**, 035002.
1090 T. Imanaka, Molecular bases of thermophily in hyperthermophiles, *Proc. Jpn. Acad., Ser. B*, 2011, **87**, 587–602.
1091 L. D. Unsworth, J. van der Oost and S. Koutsopoulos, Hyperthermophilic enzymes—stability, activity and implementation strategies for high temperature applications, *FEBS J.*, 2007, **274**, 4044–4056.
1092 H. Hashimoto, T. Inoue, M. Nishioka, S. Fujiwara, M. Takagi, T. Imanaka and Y. Kai, Hyperthermostable protein structure maintained by intra and inter-helix ion-pairs in archaeal O6-methylguanine-DNA methyltransferase, *J. Mol. Biol.*, 1999, **292**, 707–716.
1093 I. Matsui and K. Harata, Implication for buried polar contacts and ion pairs in hyperthermostable enzymes, *FEBS J.*, 2007, **274**, 4012–4022.
1094 C. H. Chan, H. K. Liang, N. W. Hsiao, M. T. Ko, P. C. Lyu and J. K. Hwang, Relationship between local structural entropy and protein thermostability, *Proteins*, 2004, **57**, 684–691.
1095 R. B. Greaves and J. Warwicker, Stability and solubility of proteins from extremophiles, *Biochem. Biophys. Res. Commun.*, 2009, **380**, 581–585.
1096 P. C. Rathi, S. Radestock and H. Gohlke, Thermostabilizing mutations preferentially occur at structural weak spots with a high mutation ratio, *J. Biotechnol.*, 2012, **159**, 135–144.
1097 S. Radestock and H. Gohlke, Protein rigidity and thermophilic adaptation, *Proteins*, 2011, **79**, 1089–1108.
1098 C. P. Lin, S. W. Huang, Y. L. Lai, S. C. Yen, C. H. Shih, C. H. Lu, C. C. Huang and J. K. Hwang, Deriving protein dynamical properties from weighted protein contact number, *Proteins: Struct., Funct., Bioinf.*, 2008, **72**, 929–935.
1099 J. K. Blum, M. D. Ricketts and A. S. Bommarius, Improved thermostability of AEH by combining B-FIT analysis and structure-guided consensus method, *J. Biotechnol.*, 2012, **160**, 214–221.
1100 J. R. Engen, Analysis of protein conformation and dynamics by hydrogen/deuterium exchange MS, *Anal. Chem.*, 2009, **81**, 7870–7875.
1101 I. A. Kaltashov, C. E. Bobst and R. R. Abzalimov, H/D exchange and mass spectrometry in the studies of protein conformation and dynamics: is there a need for a top-down approach? *Anal. Chem.*, 2009, **81**, 7892–7899.
1102 S. T. Esswein, H. V. Florance, L. Baillie, J. Lippens and P. E. Barran, A comparison of mass spectrometry based hydrogen deuterium exchange methods for probing the cyclophilin A cyclosporin complex, *J. Chromatogr. A*, 2010, **1217**, 6709–6717.
1103 L. Konermann, J. Pan and Y. H. Liu, Hydrogen exchange mass spectrometry for studying protein structure and dynamics, *Chem. Soc. Rev.*, 2011, **40**, 1224–1234.
1104 E. Jurneczko, F. Cruickshank, M. Porrini, P. Nikolova, I. D. Campuzano, M. Morris and P. E. Barran, Intrinsic disorder in proteins: a challenge for (un)structural biology met by ion mobility-mass spectrometry, *Biochem. Soc. Trans.*, 2012, **40**, 1021–1026.
1105 S. Nakazawa, J. Ahn, N. Hashii, K. Hirose and N. Kawasaki, Analysis of the local dynamics of human insulin and a rapid-acting insulin analog by hydrogen/deuterium exchange mass spectrometry, *Biochim. Biophys. Acta*, 2013, **1834**, 1210–1214.
1106 D. Resetca and D. J. Wilson, Characterizing rapid, activity-linked conformational transitions in proteins *via* sub-second hydrogen deuterium exchange mass spectrometry, *FEBS J.*, 2013, **280**, 5616–5625.
1107 D. Goswami, C. Callaway, B. D. Pascal, R. Kumar, D. P. Edwards and P. R. Griffin, Influence of domain interactions on conformational mobility of the progesterone receptor detected by hydrogen/deuterium exchange mass spectrometry, *Structure*, 2014, **22**, 961–973.
1108 D. P. Marciano, V. Dharmarajan and P. R. Griffin, HDX-MS guided drug discovery: small molecules and biopharmaceuticals, *Curr. Opin. Struct. Biol.*, 2014, **28C**, 105–111.
1109 C. Pfleger, P. C. Rathi, D. L. Klein, S. Radestock and H. Gohlke, Constraint Network Analysis (CNA): a Python software package for efficiently linking biomacromolecular structure, flexibility, (thermo-)stability, and function, *J. Chem. Inf. Model.*, 2013, **53**, 1007–1015.
1110 B. C. Buer, B. J. Levin and E. N. G. Marsh, Influence of Fluorination on the Thermodynamics of Protein Folding, *J. Am. Chem. Soc.*, 2012, **134**, 13027–13034.
1111 B. C. Buer and E. N. G. Marsh, Fluorine: a new element in protein design, *Protein Sci.*, 2012, **21**, 453–462.
1112 B. C. Buer, J. L. Meagher, J. A. Stuckey and E. N. G. Marsh, Structural basis for the enhanced stability of highly fluorinated proteins, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 4810–4815.
1113 K. H. Oh, S. H. Nam and H. S. Kim, Improvement of oxidative and thermostability of N-carbamyl-d-amino Acid amidohydrolase by directed evolution, *Protein Eng.*, 2002, **15**, 689–695.
1114 K. H. Oh, S. H. Nam and H. S. Kim, Directed evolution of N-carbamyl-D-amino acid amidohydrolase for simultaneous improvement of oxidative and thermal stability, *Biotechnol. Prog.*, 2002, **18**, 413–417.
1115 E. Vázquez-Figueroa, V. Yeh, J. M. Broering, J. F. Chaparro-Riggers and A. S. Bommarius, Thermostable variants constructed *via* the structure-guided consensus method also show increased stability in salts solutions and homogeneous aqueous-organic media, *Protein Eng., Des. Sel.*, 2008, **21**, 673–680.
1116 P. D. Dobson and D. B. Kell, Carrier-mediated cellular uptake of pharmaceutical drugs: an exception or the rule? *Nat. Rev. Drug Discovery*, 2008, **7**, 205–220.

1117 P. Dobson, K. Lanthaler, S. G. Oliver and D. B. Kell, Implications of the dominant role of cellular transporters in drug uptake, *Curr. Top. Med. Chem.*, 2009, **9**, 163–184.

1118 D. B. Kell, P. D. Dobson and S. G. Oliver, Pharmaceutical drug transport: the issues and the implications that it is essentially carrier-mediated only., *Drug Discovery Today*, 2011, **16**, 704–714.

1119 D. B. Kell and R. Goodacre, Metabolomics and systems pharmacology: why and how to model the human metabolic network for drug discovery, *Drug Discovery Today*, 2014, **19**, 171–182.

1120 G. J. Salter and D. B. Kell, Solvent selection for whole cell biotransformations in organic media., *CRC Crit. Rev. Biotechnol.*, 1995, **15**, 139–177.

1121 C. R. Wescott and A. M. Klibanov, Predicting the solvent dependence of enzymatic substrate specificity using semiempirical thermodynamic calculations, *J. Am. Chem. Soc.*, 1993, **115**, 10362–10363.

1122 C. R. Wescott and A. M. Klibanov, The solvent dependence of enzyme specificity, *Biochim. Biophys. Acta*, 1994, **1206**, 1–9.

1123 G. Carrea, G. Ottolina and S. Riva, Role of solvents in the control of enzyme selectivity in organic media, *Trends Biotechnol.*, 1995, **13**, 63–70.

1124 Y. Sardessai and S. Bhosle, Tolerance of bacteria to organic solvents, *Res. Microbiol.*, 2002, **153**, 263–268.

1125 Y. N. Sardessai and S. Bhosle, Industrial Potential of Organic Solvent Tolerant Bacteria, *Biotechnol. Prog.*, 2004, **20**, 655–660.

1126 C. Liu, G. Yang, L. Wu, G. Tian, Z. Zhang and Y. Feng, Switch of substrate specificity of hyperthermophilic acylaminoacyl peptidase by combination of protein and solvent engineering, *Protein Cell*, 2011, **2**, 497–506.

1127 P. J. Halling, Solvent selection for biocatalysis in mainly organic systems: predictions of effects on equilibrium position, *Biotechnol. Bioeng.*, 1990, **35**, 691–701.

1128 K. Xu, K. Griebenow and A. M. Klibanov, Correlation between catalytic activity and secondary structure of subtilisin dissolved in organic solvents, *Biotechnol. Bioeng.*, 1997, **56**, 485–491.

1129 M. N. Gupta and I. Roy, Enzymes in organic media. Forms, functions and applications, *Eur. J. Biochem.*, 2004, **271**, 2575–2583.

1130 E. M. Nordwald and J. L. Kaar, Stabilization of Enzymes in Ionic Liquids Via Modification of Enzyme Charge, *Biotechnol. Bioeng.*, 2013, **110**, 2352–2360.

1131 E. M. Nordwald and J. L. Kaar, Mediating Electrostatic Binding of 1-Butyl-3-methylimidazolium Chloride to Enzyme Surfaces Improves Conformational Stability, *J. Phys. Chem. B*, 2013, **117**, 8977–8986.

1132 Z. Maugeri, W. Leitner and P. D. de Maria, Practical separation of alcohol-ester mixtures using Deep-Eutectic-Solvents, *Tetrahedron Lett.*, 2012, **53**, 6968–6971.

1133 Z. Maugeri and P. D. de Maria, Novel choline-chloride-based deep-eutectic-solvents with renewable hydrogen bond donors: levulinic acid and sugar-based polyols, *RSC Adv.*, 2012, **2**, 421–425.

1134 P. Domínguez de María and Z. Maugeri, Ionic liquids in biotransformations: from proof-of-concept to emerging deep-eutectic-solvents, *Curr. Opin. Chem. Biol.*, 2011, **15**, 220–225.

1135 Q. H. Zhang, K. D. Vigier, S. Royer and F. Jérôme, Deep eutectic solvents: syntheses, properties and applications, *Chem. Soc. Rev.*, 2012, **41**, 7108–7146.

1136 A. Cadeddu, E. K. Wylie, J. Jurczak, M. Wampler-Doty and B. A. Grzybowski, Organic chemistry as a language and the implications of chemical linguistics for structural and retrosynthetic analyses, *Angew. Chem., Int. Ed.*, 2014, **53**, 8108–8112.

1137 P. Carbonell, A. G. Planson, D. Fichera and J. L. Faulon, A retrosynthetic biology approach to metabolic pathway design for therapeutic production, *BMC Syst. Biol.*, 2011, **5**, 122.

1138 P. Carbonell, A. G. Planson and J. L. Faulon, Retrosynthetic design of heterologous pathways, *Methods Mol. Biol.*, 2013, **985**, 149–173.

1139 Q. Huang, L. L. Li and S. Y. Yang, RASA: a rapid retrosynthesis-based scoring method for the assessment of synthetic accessibility of drug-like molecules, *J. Chem. Inf. Model.*, 2011, **51**, 2768–2777.

1140 J. Law, Z. Zsoldos, A. Simon, D. Reid, Y. Liu, S. Y. Khew, A. P. Johnson, S. Major, R. A. Wade and H. Y. Ando, Route Designer: a retrosynthetic analysis tool utilizing automated retrosynthetic rule generation, *J. Chem. Inf. Model.*, 2009, **49**, 593–602.

1141 X. Q. Lewell, D. B. Judd, S. P. Watson and M. M. Hann, RECAP–retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 511–522.

1142 J. Gonzalez-Lergier, L. J. Broadbelt and V. Hatzimanikatis, Theoretical considerations and computational analysis of the complexity in polyketide synthesis pathways, *J. Am. Chem. Soc.*, 2005, **127**, 9930–9938.

1143 V. Hatzimanikatis, C. Li, J. A. Ionita, C. S. Henry, M. D. Jankowski and L. J. Broadbelt, Exploring the diversity of complex metabolic networks, *Bioinformatics*, 2005, **21**, 1603–1609.

1144 C. S. Henry, L. J. Broadbelt and V. Hatzimanikatis, Discovery and analysis of novel metabolic pathways for the biosynthesis of industrial chemicals: 3-hydroxypropanoate, *Biotechnol. Bioeng.*, 2010, **106**, 462–473.

1145 K. C. Soh and V. Hatzimanikatis, DREAMS of metabolism, *Trends Biotechnol.*, 2010, **28**, 501–508.

1146 A. G. Planson, P. Carbonell, I. Grigoras and J. L. Faulon, A retrosynthetic biology approach to therapeutics: from conception to delivery, *Curr. Opin. Biotechnol.*, 2012, **23**, 948–956.

1147 N. J. Turner and E. O'Reilly, Biocatalytic retrosynthesis, *Nat. Chem. Biol.*, 2013, **9**, 285–288.

1148 M. A. Campodonico, B. A. Andrews, J. A. Asenjo, B. O. Palsson and A. M. Feist, Generation of an atlas for

commodity chemical production in Escherichia coli and a novel pathway prediction algorithm, *GEM-Path, Metab Eng*, 2014, **25**, 140–158.

1149 K. J. Bishop, R. Klajn and B. A. Grzybowski, The core and most useful molecules in organic chemistry, *Angew. Chem., Int. Ed.*, 2006, **45**, 5348–5354.

1150 M. Fialkowski, K. J. Bishop, V. A. Chubukov, C. J. Campbell and B. A. Grzybowski, Architecture and evolution of organic chemistry, *Angew. Chem., Int. Ed.*, 2005, **44**, 7263–7269.

1151 D. Ghislieri, A. P. Green, M. Pontini, S. C. Willies, I. Rowles, A. Frank, G. Grogan and N. J. Turner, Engineering an enantioselective amine oxidase for the synthesis of pharmaceutical building blocks and alkaloid natural products, *J. Am. Chem. Soc.*, 2013, **135**, 10863–10869.

1152 G. M. Whited, F. J. Feher, D. A. Benko, M. A. Cervin, G. K. Chotani, J. C. McAuliffe, R. J. LaDuca, E. A. Ben-Shoshan and K. J. Sanford, Development of a gas-phase bioprocess for isoprene-monomer production using metabolic pathway engineering, *Ind. Biotechnol.*, 2010, **6**, 152–163.

1153 G. DeSantis, K. Wong, B. Farwell, K. Chatman, Z. Zhu, G. Tomlinson, H. Huang, X. Tan, L. Bibbs, P. Chen, K. Kretz and M. J. Burk, Creation of a productive, highly enantioselective nitrilase through gene site saturation mutagenesis (GSSM), *J. Am. Chem. Soc.*, 2003, **125**, 11476–11477.

1154 A. S. Bommarius, J. K. Blum and M. J. Abrahamson, Status of protein engineering for biocatalysts: how to design an industrially useful biocatalyst, *Curr. Opin. Chem. Biol.*, 2011, **15**, 194–200.

1155 K. Faber, *Biotransformations in organic chemistry. A textbook.*, Springer, Berlin, 2011.

1156 R. N. Patel, Biocatalysis: Synthesis of Key Intermediates for Development of Pharmaceuticals, *ACS Catal.*, 2011, **1**, 1056–1074.

1157 M. Schrewe, M. K. Julsing, B. Buhler and A. Schmid, Whole-cell biocatalysis for selective and productive C-O functional group introduction and modification, *Chem. Soc. Rev.*, 2013, **42**, 6346–6377.

1158 R. C. Simon, F. G. Mutti and W. Kroutil, Biocatalytic synthesis of enantiopure building blocks for pharmaceuticals, *Drug Discovery Today: Technol.*, 2013, **10**, e37–44.

1159 M. Wang, T. Si and H. Zhao, Biocatalyst development by directed evolution, *Bioresour. Technol.*, 2012, **115**, 117–125.

1160 N. Bhan, P. Xu and M. A. G. Koffas, Pathway and protein engineering approaches to produce novel and commodity small molecules, *Curr. Opin. Biotechnol.*, 2013, **24**, 1137–1143.

1161 G. W. Huisman and S. J. Collier, On the development of new biocatalytic processes for practical pharmaceutical synthesis, *Curr. Opin. Chem. Biol.*, 2013, **17**, 284–292.

1162 B. M. Nestl, S. C. Hammer, B. A. Nebel and B. Hauer, New generation of biocatalysts for organic synthesis, *Angew. Chem., Int. Ed.*, 2014, **53**, 3070–3095.

1163 A. Bolt, A. Berry and A. Nelson, Directed evolution of aldolases for exploitation in synthetic organic chemistry, *Arch. Biochem. Biophys.*, 2008, **474**, 318–330.

1164 C. L. Windle, M. Muller, A. Nelson and A. Berry, Engineering aldolases as biocatalysts, *Curr. Opin. Chem. Biol.*, 2014, **19**, 25–33.

1165 K. Okrasa, C. Levy, M. Wilding, M. Goodall, N. Baudendistel, B. Hauer, D. Leys and J. Micklefield, Structure-guided directed evolution of alkenyl and arylmalonate decarboxylases, *Angew. Chem., Int. Ed.*, 2009, **48**, 7691–7694.

1166 M. J. Abrahamson, E. Vazquez-Figueroa, N. B. Woodall, J. C. Moore and A. S. Bommarius, Development of an Amine Dehydrogenase for Synthesis of Chiral Amines, *Angew. Chem., Int. Ed.*, 2012, **51**, 3969–3972.

1167 M. J. Abrahamson, J. W. Wong and A. S. Bommarius, The Evolution of an Amine Dehydrogenase Biocatalyst for the Asymmetric Production of Chiral Amines, *Adv. Synth. Catal.*, 2013, **355**, 1780–1786.

1168 J. H. Sattler, M. Fuchs, K. Tauber, F. G. Mutti, K. Faber, J. Pfeffer, T. Haas and W. Kroutil, Redox Self-Sufficient Biocatalyst Network for the Amination of Primary Alcohols, *Angew. Chem., Int. Ed.*, 2012, **51**, 9156–9159.

1169 H. Kohls, F. Steffen-Munsberg and M. Höhne, Recent achievements in developing the biocatalytic toolbox for chiral amine synthesis, *Curr. Opin. Chem. Biol.*, 2014, **19**, 180–192.

1170 K. Meister, S. Ebbinghaus, Y. Xu, J. G. Duman, A. Devries, M. Gruebele, D. M. Leitner and M. Havenith, Long-range protein-water dynamics in hyperactive insect antifreeze proteins, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 1617–1622.

1171 M. L. Matthews, W. C. Chang, A. P. Layne, L. A. Miles, C. Krebs and J. M. Bollinger, Jr., Direct nitration and azidation of aliphatic carbons by an iron-dependent halogenase, *Nat. Chem. Biol.*, 2014, **10**, 209–215.

1172 E. M. Brustad, C-H activation: New recipes for biocatalysis, *Nat. Chem. Biol.*, 2014, **10**, 170–171.

1173 Z. G. Zhang, L. P. Parra and M. T. Reetz, Protein engineering of stereoselective Baeyer-Villiger monooxygenases, *Chemistry*, 2012, **18**, 10160–10172.

1174 I. Polyak, M. T. Reetz and W. Thiel, Quantum mechanical/molecular mechanical study on the enantioselectivity of the enzymatic Baeyer–Villiger reaction of 4-hydroxycyclohexanone, *J. Phys. Chem. B*, 2013, **117**, 4993–5001.

1175 T. Wells, Jr. and A. J. Ragauskas, Biotechnological opportunities with the beta-ketoadipate pathway, *Trends Biotechnol.*, 2012, **30**, 627–637.

1176 C. Schmidt-Dannert, D. Umeno and F. H. Arnold, Molecular breeding of carotenoid biosynthetic pathways, *Nat. Biotechnol.*, 2000, **18**, 750–753.

1177 A. Butler and M. Sandy, Mechanistic considerations of halogenating enzymes, *Nature*, 2009, **460**, 848–854.

1178 H. Deng and D. O'Hagan, The fluorinase, the chlorinase and the duf-62 enzymes, *Curr. Opin. Chem. Biol.*, 2008, **12**, 582–592.

1179 G. W. Gribble, Biohalogenation, *Prog. Chem. Org. Nat. Prod.*, 2010, **91**, 349–365.
1180 W. Runguphan, X. Qu and S. E. O'Connor, Integrating carbon-halogen bond formation into medicinal plant metabolism, *Nature*, 2010, **468**, 461–464.
1181 R. Vázquez-Duhalt, M. Ayala and F. J. Márquez-Rocha, Biocatalytic chlorination of aromatic hydrocarbons by chloroperoxidase of *Caldariomyces fumago*, *Phytochemistry*, 2001, **58**, 929–933.
1182 R. De Mot, A. De Schrijver, G. Schoofs and A. H. Parret, The thiocarbamate-inducible *Rhodococcus* enzyme ThcF as a member of the family of alpha/beta hydrolases with haloperoxidative side activity, *FEMS Microbiol. Lett.*, 2003, **224**, 197–203.
1183 Z. Hasan, R. Renirie, R. Kerkman, H. J. Ruijssenaars, A. F. Hartog and R. Wever, Laboratory-evolved vanadium chloroperoxidase exhibits 100-fold higher halogenating activity at alkaline pH: catalytic effects from first and second coordination sphere mutations, *J. Biol. Chem.*, 2006, **281**, 9738–9744.
1184 M. Hofrichter and R. Ullrich, Heme-thiolate haloperoxidases: versatile biocatalysts with biotechnological and environmental significance, *Appl. Microbiol. Biotechnol.*, 2006, **71**, 276–288.
1185 J. M. Winter and B. S. Moore, Exploring the Chemistry and Biology of Vanadium-dependent Haloperoxidases, *J. Biol. Chem.*, 2009, **284**, 18577–18581.
1186 M. Hofrichter, R. Ullrich, M. J. Pecyna, C. Liers and T. Lundell, New and classic families of secreted fungal heme peroxidases, *Appl. Microbiol. Biotechnol.*, 2010, **87**, 871–897.
1187 P. Bernhardt, T. Okino, J. M. Winter, A. Miyanaga and B. S. Moore, A Stereoselective Vanadium-Dependent Chloroperoxidase in Bacterial Antibiotic Biosynthesis, *J. Am. Chem. Soc.*, 2011, **133**, 4268–4270.
1188 C. R. Otey, M. Landwehr, J. B. Endelman, K. Hiraga, J. D. Bloom and F. H. Arnold, Structure-guided recombination creates an artificial family of cytochromes P450, *PLoS Biol.*, 2006, **4**, e112.
1189 S. L. Kelly and D. E. Kelly, Microbial cytochromes P450: biodiversity and biotechnology. Where do cytochromes P450 come from, what do they do and what can they do for us? *Philos. Trans. R. Soc. London, Ser. B*, 2013, **368**, 20120476.
1190 C. Gavira, R. Höfer, A. Lesot, F. Lambert, J. Zucca and D. Werck-Reichhart, Challenges and pitfalls of P450-dependent (+)-valencene bioconversion by *Saccharomyces cerevisiae*, *Metab. Eng.*, 2013, **18**, 25–35.
1191 D. C. Lamb and M. R. Waterman, Unusual properties of the cytochrome P450 superfamily, *Philos. Trans. R. Soc. London, Ser. B*, 2013, **368**, 20120434.
1192 J. M. Caswell, M. O'Neill, S. J. C. Taylor and T. S. Moody, Engineering and application of P450 monooxygenases in pharmaceutical and metabolite synthesis, *Curr. Opin. Chem. Biol.*, 2013, **17**, 271–275.
1193 G. A. Roberts, A. Celik, D. J. B. Hunter, T. W. B. Ost, J. H. White, S. K. Chapman, N. J. Turner and S. L. Flitsch, A self-sufficient cytochrome P450 with a primary structural organization that includes a flavin domain and a [2Fe-2S] redox center, *J. Biol. Chem.*, 2003, **278**, 48914–48920.
1194 D. J. B. Hunter, G. A. Roberts, T. W. B. Ost, J. H. White, S. Muller, N. J. Turner, S. L. Flitsch and S. K. Chapman, Analysis of the domain properties of the novel cytochrome P450RhF, *FEBS Lett.*, 2005, **579**, 2215–2220.
1195 E. O'Reilly, M. Corbett, S. Hussain, P. P. Kelly, D. Richardson, S. L. Flitsch and N. J. Turner, Substrate promiscuity of cytochrome P450 RhF, *Catal. Sci. Technol.*, 2013, **3**, 1490–1492.
1196 E. O'Reilly, S. J. Aitken, G. Grogan, P. P. Kelly, N. J. Turner and S. L. Flitsch, Regio- and stereoselective oxidation of unactivated C-H bonds with *Rhodococcus rhodochrous*, *Beilstein J. Org. Chem.*, 2012, **8**, 496–500.
1197 A. Robin, V. Kohler, A. Jones, A. Ali, P. P. Kelly, E. O'Reilly, N. J. Turner and S. L. Flitsch, Chimeric self-sufficient P450cam-RhFRed biocatalysts with broad substrate scope, *Beilstein J. Org. Chem.*, 2011, **7**, 1494–1498.
1198 A. Robin, G. A. Roberts, J. Kisch, F. Sabbadin, G. Grogan, N. Bruce, N. J. Turner and S. L. Flitsch, Engineering and improvement of the efficiency of a chimeric [P450cam-RhFRed reductase domain] enzyme, *Chem. Commun.*, 2009, 2478–2480.
1199 R. Fasan, M. M. Chen, N. C. Crook and F. H. Arnold, Engineered alkane-hydroxylating cytochrome P450(BM3) exhibiting nativelike catalytic properties, *Angew. Chem., Int. Ed.*, 2007, **46**, 8414–8418.
1200 A. Trefzer, V. Jungmann, I. Molnar, A. Botejue, D. Buckel, G. Frey, D. S. Hill, M. Jorg, J. M. Ligon, D. Mason, D. Moore, J. P. Pachlatko, T. H. Richardson, P. Spangenberg, M. A. Wall, R. Zirkle and J. T. Stege, Biocatalytic conversion of avermectin to 4''-oxo-avermectin: improvement of cytochrome p450 monooxygenase specificity by directed evolution, *Appl. Environ. Microbiol.*, 2007, **73**, 4317–4325.
1201 J. C. Lewis, S. M. Mantovani, Y. Fu, C. D. Snow, R. S. Komor, C. H. Wong and F. H. Arnold, Combinatorial alanine substitution enables rapid optimization of cytochrome P450BM3 for selective hydroxylation of large substrates, *ChemBioChem*, 2010, **11**, 2502–2505.
1202 V. B. Urlacher and M. Girhard, Cytochrome P450 monooxygenases: an update on perspectives for synthetic application, *Trends Biotechnol.*, 2012, **30**, 26–36.
1203 A. Seifert, M. Antonovici, B. Hauer and J. Pleiss, An efficient route to selective bio-oxidation catalysts: an iterative approach comprising modeling, diversification, and screening, based on CYP102A1, *ChemBioChem*, 2011, **12**, 1346–1351.
1204 F. E. Zilly, J. P. Acevedo, W. Augustyniak, A. Deege, U. W. Hausig and M. T. Reetz, Tuning a p450 enzyme for methane oxidation, *Angew. Chem., Int. Ed.*, 2011, **50**, 2720–2724.
1205 S. T. Jung, R. Lauchli and F. H. Arnold, Cytochrome P450: taming a wild type enzyme, *Curr. Opin. Biotechnol.*, 2011, **22**, 809–817.

1206 G. D. Roiban and M. T. Reetz, Enzyme promiscuity: using a P450 enzyme as a carbene transfer catalyst, *Angew. Chem., Int. Ed.*, 2013, **52**, 5439–5440.
1207 D. Jiang, R. Tu, P. Bai and Q. Wang, Directed evolution of cytochrome P450 for sterol epoxidation, *Biotechnol. Lett.*, 2013, **35**, 1663–1668.
1208 G. D. Roiban, R. Agudo, A. Ilie, R. Lonsdale and M. T. Reetz, CH-activating oxidative hydroxylation of 1-tetralones and related compounds with high regio- and stereoselectivity, *Chem. Commun.*, 2014, **50**, 14310–14313.
1209 H. J. Kim, M. W. Ruszczycky, S. H. Choi, Y. N. Liu and H. W. Liu, Enzyme-catalysed [4+2] cycloaddition is a key step in the biosynthesis of spinosyn A, *Nature*, 2011, **473**, 109–112.
1210 C. A. Townsend, A ''Diels-Alderase'' at Last, *ChemBioChem*, 2011, **12**, 2267–2269.
1211 S. Obeid, A. Schnur, C. Gloeckner, N. Blatter, W. Welte, K. Diederichs and A. Marx, Learning from directed evolution: *Thermus aquaticus* DNA polymerase mutants with translesion synthesis activity, *ChemBioChem*, 2011, **12**, 1574–1580.
1212 D. Loakes, J. Gallego, V. B. Pinheiro, E. T. Kool and P. Holliger, Evolving a polymerase for hydrophobic base analogues, *J. Am. Chem. Soc.*, 2009, **131**, 14827–14837.
1213 S. Park, K. L. Morley, G. P. Horsman, M. Holmquist, K. Hult and R. J. Kazlauskas, Focusing mutations into the *P. fluorescens* esterase binding site increases enantioselectivity more effectively than distant mutations, *Chem. Biol.*, 2005, **12**, 45–54.
1214 Z. L. Fowler and M. A. G. Koffas, Biosynthesis and biotechnological production of flavanones: current state and perspectives, *Appl. Microbiol. Biotechnol.*, 2009, **83**, 799–808.
1215 Z. L. Fowler, K. Shah, J. C. Panepinto, A. Jacobs and M. A. G. Koffas, Development of non-natural flavanones as antimicrobial agents, *PLoS One*, 2011, **6**, e25681.
1216 Z. L. Fowler, W. W. Gikandi and M. A. G. Koffas, Increased malonyl coenzyme A biosynthesis by tuning the *Escherichia coli* metabolic network and its application to flavanone production, *Appl. Environ. Microbiol.*, 2009, **75**, 5831–5839.
1217 E. Leonard, Y. Yan, Z. L. Fowler, Z. Li, C. G. Lim, K. H. Lim and M. A. G. Koffas, Strain improvement of recombinant *Escherichia coli* for efficient production of plant flavonoids, *Mol. Pharmaceutics*, 2008, **5**, 257–265.
1218 P. Xu, S. Ranganathan, Z. L. Fowler, C. D. Maranas and M. A. Koffas, Genome-scale metabolic network modeling results in minimal interventions that cooperatively force carbon flux towards malonyl-CoA, *Metab. Eng.*, 2011, **13**, 578–587.
1219 M. Mora-Pale, S. P. Sanchez-Rodriguez, R. J. Linhardt, J. S. Dordick and M. A. G. Koffas, Metabolic engineering and *in vitro* biosynthesis of phytochemicals and non-natural analogues, *Plant Sci.*, 2013, **210**, 10–24.
1220 C. X. Wu, R. Liu, M. Gao, G. Zhao, S. Wu, C. F. Wu and G. H. Du, Pinocembrin protects brain against ischemia/reperfusion injury by attenuating endoplasmic reticulum stress induced apoptosis, *Neurosci. Lett.*, 2013, **546**, 57–62.
1221 J. Wu, G. Du, J. Zhou and J. Chen, Metabolic engineering of *Escherichia coli* for (2*S*)-pinocembrin production from glucose by a modular metabolic strategy, *Metab. Eng.*, 2013, **16**, 48–55.
1222 H. Deng, S. M. Cross, R. P. McGlinchey, J. T. Hamilton and D. O'Hagan, *In vitro* reconstituted biotransformation of 4-fluorothreonine from fluoride ion: application of the fluorinase, *Chem. Biol.*, 2008, **15**, 1268–1276.
1223 W. Liu, X. Huang, M. J. Cheng, R. J. Nielsen, W. A. Goddard, 3rd and J. T. Groves, Oxidative aliphatic C-H fluorination with fluoride ion catalyzed by a manganese porphyrin, *Science*, 2012, **337**, 1322–1325.
1224 R. M. Lennen, M. G. Politz, M. A. Kruziki and B. F. Pfleger, Identification of transport proteins involved in free fatty acid efflux in *Escherichia coli*, *J. Bacteriol.*, 2013, **195**, 135–144.
1225 R. M. Lennen and B. F. Pfleger, Engineering *Escherichia coli* to synthesize free fatty acids, *Trends Biotechnol.*, 2012, **30**, 659–667.
1226 J. T. Youngquist, R. M. Lennen, D. R. Ranatunga, W. H. Bothfeld, W. D. Marner 2nd and B. F. Pfleger, Kinetic modeling of free fatty acid production in *Escherichia coli* based on continuous cultivation of a plasmid free strain, *Biotechnol. Bioeng.*, 2012, **109**, 1518–1527.
1227 L. A. Castle, D. L. Siehl, R. Gorton, P. A. Patten, Y. H. Chen, S. Bertain, H. J. Cho, N. Duck, J. Wong, D. Liu and M. W. Lassner, Discovery and directed evolution of a glyphosate tolerance gene, *Science*, 2004, **304**, 1151–1154.
1228 D. L. Siehl, L. A. Castle, R. Gorton, Y. H. Chen, S. Bertain, H. J. Cho, R. Keenan, D. Liu and M. W. Lassner, Evolution of a microbial acetyltransferase for modification of glyphosate: a novel tolerance strategy, *Pest Manage. Sci.*, 2005, **61**, 235–240.
1229 L. Pollegioni, E. Schonbrunn and D. Siehl, Molecular basis of glyphosate resistance-different approaches through protein engineering, *FEBS J.*, 2011, **278**, 2753–2766.
1230 M. Pedotti, E. Rosini, G. Molla, T. Moschetti, C. Savino, B. Vallone and L. Pollegioni, Glyphosate resistance by engineering the flavoenzyme glycine oxidase, *J. Biol. Chem.*, 2009, **284**, 36415–36423.
1231 T. Zhan, K. Zhang, Y. Chen, Y. Lin, G. Wu, L. Zhang, P. Yao, Z. Shao and Z. Liu, Improving glyphosate oxidation activity of glycine oxidase from *Bacillus cereus* by directed evolution, *PLoS One*, 2013, **8**, e79175.
1232 Z. Prokop, Y. Sato, J. Brezovsky, T. Mozga, R. Chaloupkova, T. Koudelakova, P. Jerabek, V. Stepankova, R. Natsume, J. G. van Leeuwen, D. B. Janssen, J. Florian, Y. Nagata, T. Senda and J. Damborsky, Enantioselectivity of haloalkane dehalogenases and its modulation by surface loop engineering, *Angew. Chem., Int. Ed.*, 2010, **49**, 6111–6115.
1233 T. Koudelakova, E. Chovancova, J. Brezovsky, M. Monincova, A. Fortova, J. Jarkovsky and J. Damborsky, Substrate specificity of haloalkane dehalogenases, *Biochem. J.*, 2011, **435**, 345–354.

1234 J. G. E. van Leeuwen, H. J. Wijma, R. J. Floor, J. M. van der Laan and D. B. Janssen, Directed evolution strategies for enantiocomplementary haloalkane dehalogenases: from chemical waste to enantiopure building blocks, *ChemBioChem*, 2012, **13**, 137–148.

1235 R. J. Floor, H. J. Wijma, D. I. Colpa, A. Ramos-Silva, P. A. Jekel, W. Szymanski, B. L. Feringa, S. J. Marrink and D. B. Janssen, Computational library design for increasing haloalkane dehalogenase stability, *ChemBioChem*, 2014, **15**, 1660–1672.

1236 W. S. Glenn, E. Nims and S. E. O'Connor, Reengineering a tryptophan halogenase to preferentially chlorinate a direct alkaloid precursor, *J. Am. Chem. Soc.*, 2011, **133**, 19346–19349.

1237 L. N. Herrera-Rodriguez, H. P. Meyer, K. T. Robins and F. Khan, Perspectives on biotechnological halogenation Part II: Prospecting for future biohalogenases, *Chim. Oggi*, 2011, **29**, 47–49.

1238 L. N. Herrera-Rodriguez, F. Khan, K. T. Robins and H. P. Meyer, Perspectives on biotechnological halogenation Part I: Halogenated products and enzymatic halogenation, *Chim Oggi*, 2011, **29**, 31–33.

1239 S. D. Wong, M. Srnec, M. L. Matthews, L. V. Liu, Y. Kwak, K. Park, C. B. Bell 3rd, E. E. Alp, J. Zhao, Y. Yoda, S. Kitao, M. Seto, C. Krebs, J. M. Bollinger, Jr. and E. I. Solomon, Elucidation of the Fe(IV)=O intermediate in the catalytic cycle of the halogenase SyrB2, *Nature*, 2013, **499**, 320–323.

1240 K. Bernath-Levin, J. Shainsky, L. Sigawi and A. Fishman, Directed evolution of nitrobenzene dioxygenase for the synthesis of the antioxidant hydroxytyrosol, *Appl. Microbiol. Biotechnol.*, 2014, **98**, 4975–4985.

1241 O. Khersonsky, D. Röthlisberger, O. Dym, S. Albeck, C. J. Jackson, D. Baker and D. S. Tawfik, Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the KE07 series, *J. Mol. Biol.*, 2010, **396**, 1025–1042.

1242 O. Khersonsky, D. Röthlisberger, A. M. Wollacott, P. Murphy, O. Dym, S. Albeck, G. Kiss, K. N. Houk, D. Baker and D. S. Tawfik, Optimization of the in-silico-designed Kemp eliminase KE70 by computational design and directed evolution, *J. Mol. Biol.*, 2010, **407**, 391–412.

1243 R. Blomberg, H. Kries, D. M. Pinkas, P. R. Mittl, M. G. Grutter, H. K. Privett, S. L. Mayo and D. Hilvert, Precision is essential for efficient catalysis in an evolved Kemp eliminase, *Nature*, 2013, **503**, 418–421.

1244 A. Labas, E. Szabo, L. Mones and M. Fuxreiter, Optimization of reorganization energy drives evolution of the designed Kemp eliminase KE07, *Biochim. Biophys. Acta*, 2013, **1834**, 908–917.

1245 O. Khersonsky, G. Kiss, D. Rothlisberger, O. Dym, S. Albeck, K. N. Houk, D. Baker and D. S. Tawfik, Bridging the gaps in design methodologies by evolutionary optimization of the stability and proficiency of designed Kemp eliminase KE59, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 10358–10363.

1246 M. P. Frushicheva, J. Cao and A. Warshel, Challenges and advances in validating enzyme design proposals: the case of Kemp eliminase catalysis, *Biochemistry*, 2011, **50**, 3849–3858.

1247 M. P. Frushicheva, J. Cao, Z. T. Chu and A. Warshel, Exploring challenges in rational enzyme design by simulating the catalysis in artificial kemp eliminase, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 16869–16874.

1248 J. C. Moore, D. J. Pollard, B. Kosjek and P. N. Devine, Advances in the enzymatic reduction of ketones, *Acc. Chem. Res.*, 2007, **40**, 1412–1419.

1249 R. Agudo, G. D. Roiban and M. T. Reetz, Induced axial chirality in biocatalytic asymmetric ketone reduction, *J. Am. Chem. Soc.*, 2013, **135**, 1665–1668.

1250 S. Camarero, I. Pardo, A. I. Cañas, P. Molina, E. Record, A. T. Martínez, M. J. Martínez and M. Alcalde, Engineering platforms for directed evolution of Laccase from *Pycnoporus cinnabarinus*, *Appl. Environ. Microbiol.*, 2012, **78**, 1370–1384.

1251 J. R. Jeon and Y. S. Chang, Laccase-mediated oxidation of small organics: bifunctional roles for versatile applications, *Trends Biotechnol.*, 2013, **31**, 335–341.

1252 D. M. Mate, D. Gonzalez-Perez, M. Falk, R. Kittl, M. Pita, A. L. De Lacey, R. Ludwig, S. Shleev and M. Alcalde, Blood tolerant laccase by directed evolution, *Chem. Biol.*, 2013, **20**, 223–231.

1253 Y. Miao, E. M. Geertsema, P. G. Tepper, E. Zandvoort and G. J. Poelarends, Promiscuous catalysis of asymmetric Michael-type additions of linear aldehydes to beta-nitrostyrene by the proline-based enzyme 4-oxalocrotonate tautomerase, *ChemBioChem*, 2013, **14**, 191–194.

1254 K. E. Atkin, R. Reiss, V. Koehler, K. R. Bailey, S. Hart, J. P. Turkenburg, N. J. Turner, A. M. Brzozowski and G. Grogan, The structure of monoamine oxidase from *Aspergillus niger* provides a molecular context for improvements in activity obtained by directed evolution, *J. Mol. Biol.*, 2008, **384**, 1218–1231.

1255 K. R. Bailey, A. J. Ellis, R. Reiss, T. J. Snape and N. J. Turner, A template-based mnemonic for monoamine oxidase (MAO-N) catalyzed reactions and its application to the chemo-enzymatic deracemisation of the alkaloid (+/-)-crispine A, *Chem. Commun.*, 2007, 3640–3642.

1256 I. Rowles, K. J. Malone, L. L. Etchells, S. C. Willies and N. J. Turner, Directed evolution of the enzyme monoamine oxidase (MAO-N): highly efficient chemoenzymatic deracemisation of the alkaloid (+/-)-Crispine A, *ChemCatChem*, 2012, **4**, 1259–1261.

1257 J. S. Anderson, J. Rittle and J. C. Peters, Catalytic conversion of nitrogen to ammonia by an iron model complex, *Nature*, 2013, **501**, 84–87.

1258 A. Pingoud and W. Wende, Generation of novel nucleases with extended specificity by rational and combinatorial strategies, *ChemBioChem*, 2011, **12**, 1495–1500.

1259 H. S. Toogood and N. S. Scrutton, Enzyme engineering toolbox - a 'catalyst' for change, *Catal. Sci. Technol.*, 2013, **3**, 2182–2194.

1260 H. S. Toogood and N. S. Scrutton, New developments in 'ene'-reductase catalysed biological hydrogenations, *Curr. Opin. Chem. Biol.*, 2014, **19**, 107–115.

1261 Y. Ashani, R. D. Gupta, M. Goldsmith, I. Silman, J. L. Sussman, D. S. Tawfik and H. Leader, Stereospecific synthesis of analogs of nerve agents and their utilization for selection and characterization of paraoxonase (PON1) catalytic scavengers, *Chem.-Biol. Interact.*, 2010, **187**, 362–369.

1262 U. Alcolombri, M. Elias and D. S. Tawfik, Directed evolution of sulfotransferases and paraoxonases by ancestral libraries, *J. Mol. Biol.*, 2011, **411**, 837–853.

1263 R. D. Gupta, M. Goldsmith, Y. Ashani, Y. Simo, G. Mullokandov, H. Bar, M. Ben-David, H. Leader, R. Margalit, I. Silman, J. L. Sussman and D. S. Tawfik, Directed evolution of hydrolases for prevention of G-type nerve agent intoxication, *Nat. Chem. Biol.*, 2011, **7**, 120–125.

1264 E. Garcia-Ruiz, D. Gonzalez-Perez, F. J. Ruiz-Dueñas, A. T. Martínez and M. Alcalde, Directed evolution of a temperature-peroxide- and alkaline pH-tolerant versatile peroxidase, *Biochem. J.*, 2012, **441**, 487–498.

1265 S. C. Patel and M. H. Hecht, Directed evolution of the peroxidase activity of a *de novo*-designed protein, *Protein Eng., Des. Sel.*, 2012, **25**, 445–452.

1266 L. F. Olguin, S. E. Askew, A. C. O'Donoghue and F. Hollfelder, Efficient catalytic promiscuity in an enzyme superfamily: an arylsulfatase shows a rate acceleration of 10^{13} for phosphate monoester hydrolysis, *J. Am. Chem. Soc.*, 2008, **130**, 16547–16555.

1267 A. C. Babtie, S. Bandyopadhyay, L. F. Olguin and F. Hollfelder, Efficient catalytic promiscuity for chemically distinct reactions, *Angew. Chem., Int. Ed.*, 2009, **48**, 3692–3694.

1268 B. van Loo, S. Jonas, A. C. Babtie, A. Benjdia, O. Berteau, M. Hyvonen and F. Hollfelder, An efficient, multiply promiscuous hydrolase in the alkaline phosphatase superfamily, *Proc. Natl. Acad. Sci. U. S. A.*, 2010, **107**, 2740–2745.

1269 L. Afriat-Jurnou, C. J. Jackson and D. S. Tawfik, Reconstructing a missing link in the evolution of a recently diverged phosphotriesterase by active-site loop remodeling, *Biochemistry*, 2012, **51**, 6047–6055.

1270 M. F. Mohamed and F. Hollfelder, Efficient, crosswise catalytic promiscuity among enzymes that catalyze phosphoryl transfer, *Biochim. Biophys. Acta*, 2013, **1834**, 417–424.

1271 H. Wiersma-Koch, F. Sunden and D. Herschlag, Site-directed mutagenesis maps interactions that enhance cognate and limit promiscuous catalysis by an alkaline phosphatase superfamily phosphodiesterase, *Biochemistry*, 2013, **52**, 9167–9176.

1272 H. Fu and C. Khosla, Antibiotic activity of polyketide products derived from combinatorial biosynthesis: implications for directed evolution, *Mol. Diversity*, 1996, **1**, 121–124.

1273 S. M. Ma, J. W. Li, J. W. Choi, H. Zhou, K. K. Lee, V. A. Moorthie, X. Xie, J. T. Kealey, N. A. Da Silva, J. C. Vederas and Y. Tang, Complete reconstitution of a highly reducing iterative polyketide synthase, *Science*, 2009, **326**, 589–592.

1274 W. Zha, S. B. Rubin-Pitel and H. Zhao, Exploiting genetic diversity by directed evolution: molecular breeding of type III polyketide synthases improves productivity, *Mol. BioSyst.*, 2008, **4**, 246–248.

1275 H. Y. Lee, C. J. Harvey, D. E. Cane and C. Khosla, Improved precursor-directed biosynthesis in *E. colivia* directed evolution, *J. Antibiot.*, 2011, **64**, 59–64.

1276 T. H. Yang, T. W. Kim, H. O. Kang, S. H. Lee, E. J. Lee, S. C. Lim, S. O. Oh, A. J. Song, S. J. Park and S. Y. Lee, Biosynthesis of polylactic acid and its copolymers using evolved propionate CoA transferase and PHA synthase, *Biotechnol. Bioeng.*, 2010, **105**, 150–160.

1277 F. Hollmann, I. W. C. E. Arends and K. Buehler, Biocatalytic Redox Reactions for Organic Synthesis: Nonconventional Regeneration Methods, *ChemCatChem*, 2010, **2**, 762–782.

1278 F. Hollmann, I. W. C. E. Arends, K. Buehler, A. Schallmey and B. Buhler, Enzyme-mediated oxidations for the chemist, *Green Chem.*, 2011, **13**, 226–265.

1279 F. Hollmann, I. W. C. E. Arends and D. Holtmann, Enzymatic reductions for the chemist, *Green Chem.*, 2011, **13**, 2285–2314.

1280 F. Geu-Flores, N. H. Sherden, V. Courdavault, V. Burlat, W. S. Glenn, C. Wu, E. Nims, Y. Cui and S. E. O'Connor, An alternative route to cyclic terpenes by reductive cyclization in iridoid biosynthesis, *Nature*, 2012, **492**, 138–142.

1281 M. Schöpfel, A. Tziridis, U. Arnold and M. T. Stubbs, Towards a restriction proteinase: construction of a self-activating enzyme, *ChemBioChem*, 2011, **12**, 1523–1527.

1282 R. Obexer, S. Studer, L. Giger, D. M. Pinkas, M. G. Grütter, D. Baker and D. Hilvert, Active Site Plasticity of a Computationally Designed RetroAldolase Enzyme, *ChemCatChem*, 2014, **6**, 1043–1050.

1283 T. Wymore, B. Y. Chen, H. B. Nicholas, A. J. Ropelewski and C. L. Brooks, A Mechanism for Evolving Novel Plant Sesquiterpene Synthase Function, *Mol. Inf.*, 2011, **30**, 896–906.

1284 B. J. Baas, E. Zandvoort, E. M. Geertsema and G. J. Poelarends, Recent advances in the study of enzyme promiscuity in the tautomerase superfamily, *ChemBioChem*, 2013, **14**, 917–926.

1285 J. E. Diaz, C. S. Lin, K. Kunishiro, B. K. Feld, S. K. Avrantinis, J. Bronson, J. Greaves, J. G. Saven and G. A. Weiss, Computational design and selections for an engineered, thermostable terpene synthase, *Protein Sci.*, 2011, **20**, 1597–1606.

1286 R. Lauchli, K. S. Rabe, K. Z. Kalbarczyk, A. Tata, T. Heel, R. Z. Kitto and F. H. Arnold, High-throughput screening for terpene-synthase-cyclization activity and directed evolution of a terpene synthase, *Angew. Chem., Int. Ed.*, 2013, **52**, 5571–5574.

1287 D. T. Major, Y. Freud and M. Weitman, Catalytic control in terpenoid cyclases: multiscale modeling of thermodynamic, kinetic, and dynamic effects, *Curr. Opin. Chem. Biol.*, 2014, **21C**, 25–33.

1288 S. H. Chen, D. R. Hwang, G. H. Chen, N. S. Hsu, Y. T. Wu, T. L. Li and C. H. Wong, Engineering transaldolase in *Pichia stipitis* to improve bioethanol production, *ACS Chem. Biol.*, 2012, **7**, 481–486.

1289 A. K. Samland, M. Rale, G. A. Sprenger and W. D. Fessner, The transaldolase family: new synthetic opportunities from an ancient enzyme scaffold, *ChemBioChem*, 2011, **12**, 1454–1474.

1290 E. G. Hibbert, T. Senussi, S. J. Costelloe, W. Lei, M. E. B. Smith, J. M. Ward, H. C. Hailes and P. A. Dalby, Directed evolution of transketolase activity on non-phosphorylated substrates, *J. Biotechnol.*, 2007, **131**, 425–432.

1291 E. G. Hibbert, T. Senussi, M. E. B. Smith, S. J. Costelloe, J. M. Ward, H. C. Hailes and P. A. Dalby, Directed evolution of transketolase substrate specificity towards an aliphatic aldehyde, *J. Biotechnol.*, 2008, **134**, 240–245.

1292 A. Cázares, J. L. Galman, L. G. Crago, M. E. B. Smith, J. Strafford, L. Ríos-Solís, G. J. Lye, P. A. Dalby and H. C. Hailes, Non-alpha-hydroxylated aldehydes with evolved transketolase enzymes, *Org. Biomol. Chem.*, 2010, **8**, 1301–1309.

1293 P. Payongsri, D. Steadman, J. Strafford, A. MacMurray, H. C. Hailes and P. A. Dalby, Rational substrate and enzyme engineering of transketolase for aromatics, *Org. Biomol. Chem.*, 2012, **10**, 9021–9029.

1294 M. Minczuk, P. Kolasinska-Zwierz, M. P. Murphy and M. A. Papworth, Construction and testing of engineered zinc-finger proteins for sequence-specific modification of mtDNA, *Nat. Protoc.*, 2010, **5**, 342–356.

1295 M. Papworth, P. Kolasinska and M. Minczuk, Designer zinc-finger proteins and their applications, *Gene*, 2006, **366**, 27–38.

1296 H. J. Wijma, S. J. Marrink and D. B. Janssen, Computationally efficient and accurate enantioselectivity modeling by clusters of molecular dynamics simulations, *J. Chem. Inf. Model.*, 2014, **54**, 2079–2092.

1297 H. J. Wijma and D. B. Janssen, Computational design gains momentum in enzyme catalysis engineering, *FEBS J.*, 2013, **280**, 2948–2960.

1298 A. Á. Rauscher, Z. Simon, G. J. Szöllösi, L. Gráf, I. Derényi and A. Málnási-Csizmadia, Temperature dependence of internal friction in enzyme reactions, *FASEB J.*, 2011, **25**, 2804–2813.

1299 A. Rauscher, I. Derényi, L. Gráf and A. Málnási-Csizmadia, Internal friction in enzyme reactions, *IUBMB Life*, 2013, **65**, 35–42.

1300 D. Tapscott and A. Williams, *Wikinomics: how mass collaboration changes everything*, New Paradigm, 2007.

1301 A. Rinaldi, Science wikinomics. Mass networking through the web creates new forms of scientific collaboration, *EMBO Rep.*, 2009, **10**, 439–443.

1302 D. Corne and J. Knowles, No free lunch and free leftovers theorems for multiobjecitve optimisation problems., in *Evolutionary Multi-criterion Optimization (EMO 2003)*, LNCS, ed. C. Fonseca *et al.*, Springer, Berlin, 2003, vol. 2632, pp. 327–341.

1303 J. C. Culberson, On the futility of blind search: an algorithmic view of 'no free lunch', *Evol. Comput.*, 1998, **6**, 109–127.

1304 N. J. Radcliffe and P. D. Surry, Fundamental limitations on search algorithms: evolutionary computing in perspective, *Computer Science Today*, 1995, **1995**, 275–291.

1305 D. H. Wolpert and W. G. Macready, No Free Lunch theorems for optimization, *IEEE Trans. Evol. Comput.*, 1997, **1**, 67–82.

1306 J. E. Rowe, M. D. Vose and A. H. Wright, Reinterpreting no free lunch, *Evol. Comput.*, 2009, **17**, 117–129.

1307 J. G. Zalatan and D. Herschlag, The far reaches of enzymology, *Nat. Chem. Biol.*, 2009, **5**, 516–520.

1308 *Computational approaches in cheminformatics and bioinformatics*, ed. R. Guha and A. Bender, Wiley, Hoboken, NJ, 2012.

1309 S. Ananiadou, D. B. Kell and J.-i. Tsujii, Text Mining and its potential applications in Systems Biology, *Trends Biotechnol.*, 2006, **24**, 571–579.

1310 S. Ananiadou, S. Pyysalo, J. i. Tsujii and D. B. Kell, Event extraction for systems biology by text mining the literature, *Trends Biotechnol.*, 2010, **28**, 381–390.

1311 S. Ananiadou, P. Thompson, R. Nawaz, J. McNaught and D. B. Kell, Event Based Text Mining for Biology and Functional Genomics, *Briefings Funct. Genomics*, 2014, DOI: 10.1093/bfgp/elu1015.

1312 M. J. Herrgård, N. Swainston, P. Dobson, W. B. Dunn, K. Y. Arga, M. Arvas, N. Blüthgen, S. Borger, R. Costenoble, M. Heinemann, M. Hucka, N. Le Novère, P. Li, W. Liebermeister, M. L. Mo, A. P. Oliveira, D. Petranovic, S. Pettifer, E. Simeonidis, K. Smallbone, I. Spasić, D. Weichart, R. Brent, D. S. Broomhead, H. V. Westerhoff, B. Kırdar, M. Penttilä, E. Klipp, B. Ø. Palsson, U. Sauer, S. G. Oliver, P. Mendes, J. Nielsen and D. B. Kell, A consensus yeast metabolic network obtained from a community approach to systems biology, *Nat. Biotechnol.*, 2008, **26**, 1155–1160.

1313 N. Swainston, P. Mendes and D. B. Kell, An analysis of a 'community-driven' reconstruction of the human metabolic network, *Metabolomics*, 2013, **9**, 757–764.

1314 P. Seeman, The membrane actions of anesthetics and tranquilizers, *Pharmacol. Rev.*, 1972, **24**, 583–655.

1315 D. B. Kell and P. D. Dobson, The cellular uptake of pharmaceutical drugs is mainly carrier-mediated and is thus an issue not so much of biophysics but of systems biology, in *Proc. Int. Beilstein Symposium on Systems Chemistry*, ed. M. G. Hicks and C. Kettner, Logos Verlag, Berlin, 2009, pp. 149–168, http://www.beilstein-institut.de/Bozen2008/Proceedings/Kell/Kell.pdf.

1316 R. Doshi, T. Nguyen and G. Chang, Transporter-mediated biofuel secretion, *Proc. Natl. Acad. Sci. U. S. A.*, 2013, **110**, 7642–7647.

1317 H. Ling, B. Chen, A. Kang, J. M. Lee and M. W. Chang, Transcriptome response to alkane biofuels in *Saccharomyces cerevisiae*: identification of efflux pumps involved in alkane tolerance, *Biotechnol. Biofuels*, 2013, **6**, 95.

1318 B. Chen, H. Ling and M. W. Chang, Transporter engineering for improved tolerance against alkane biofuels in *Saccharomyces cerevisiae*, *Biotechnol. Biofuels*, 2013, **6**, 21.

1319 J. L. Foo and S. S. J. Leong, Directed evolution of an *E. coli* inner membrane transporter for improved efflux of biofuel molecules, *Biotechnol. Biofuels*, 2013, **6**, 81.

1320 N. Nishida, N. Ozato, K. Matsui, K. Kuroda and M. Ueda, ABC transporters and cell wall proteins involved in organic solvent tolerance in *Saccharomyces cerevisiae*, *J. Biotechnol.*, 2013, **165**, 145–152.

1321 C. Grant, D. Deszcz, Y. C. Wei, R. J. Martinez-Torres, P. Morris, T. Folliard, R. Sreenivasan, J. Ward, P. Dalby, J. M. Woodley and F. Baganz, Identification and use of an alkane transporter plug-in for applications in biocatalysis and whole-cell biosensing of alkanes, *Sci. Rep.*, 2014, **4**, 5844.

1322 M. Jasiński, Y. Stukkens, H. Degand, B. Purnelle, J. Marchand-Brynaert and M. Boutry, A plant plasma membrane ATP binding cassette-type transporter is involved in antifungal terpenoid secretion, *Plant Cell*, 2001, **13**, 1095–1107.

1323 K. Yazaki, ABC transporters involved in the transport of plant secondary metabolites, *FEBS Lett.*, 2006, **580**, 1183–1191.

1324 Q. Wu, M. Kazantzis, H. Doege, A. M. Ortegon, B. Tsang, A. Falcon and A. Stahl, Fatty acid transport protein 1 is required for nonshivering thermogenesis in brown adipose tissue, *Diabetes*, 2006, **55**, 3229–3237.

1325 Q. Wu, A. M. Ortegon, B. Tsang, H. Doege, K. R. Feingold and A. Stahl, FATP1 is an insulin-sensitive fatty acid transporter involved in diet-induced obesity, *J. Mol. Cell Biol.*, 2006, **26**, 3455–3467.

1326 D. Khnykin, J. H. Miner and F. Jahnsen, Role of fatty acid transporters in epidermis: Implications for health and disease, *Dermatoendocrinol*, 2011, **3**, 53–61.

1327 M. H. Lin and D. Khnykin, Fatty acid transporters in skin development, function and disease, *Biochim. Biophys. Acta*, 2014, **1841**, 362–368.

1328 M. S. Villalba and H. M. Alvarez, Identification of a novel ATP-binding cassette transporter involved in long-chain fatty acid import and its role in triacylglycerol accumulation in *Rhodococcus jostii* RHA1, *Microbiology*, 2014, **160**, 1523–1532.

1329 R. Gimenez, M. F. Nuñez, J. Badia, J. Aguilar and L. Baldoma, The gene yjcG, cotranscribed with the gene acs, encodes an acetate permease in *Escherichia coli*, *J. Bacteriol.*, 2003, **185**, 6448–6455.

1330 R. Islam, N. Anzai, N. Ahmed, B. Ellapan, C. J. Jin, S. Srivastava, D. Miura, T. Fukutomi, Y. Kanai and H. Endou, Mouse organic anion transporter 2 (mOat2) mediates the transport of short chain fatty acid propionate, *J. Pharmacol. Sci.*, 2008, **106**, 525–528.

1331 I. Moschen, A. Bröer, S. Galić, F. Lang and S. Bröer, Significance of short chain fatty acid transport by members of the monocarboxylate transporter family (MCT), *Neurochem. Res.*, 2012, **37**, 2562–2568.

1332 J. Sá-Pessoa, S. Paiva, D. Ribas, I. J. Silva, S. C. Viegas, C. M. Arraiano and M. Casal, SATP (YaaH), a succinate-acetate transporter protein in *Escherichia coli*, *Biochem. J.*, 2013, **454**, 585–595.

1333 R. Kaldenhoff, L. Kai and N. Uehlein, Aquaporins and membrane diffusion of CO_2 in living organisms, *Biochim. Biophys. Acta*, 2014, **1840**, 1592–1595.

1334 L. Kai and R. Kaldenhoff, A refined model of water and CO_2 membrane diffusion: Effects and contribution of sterols and proteins, *Sci. Rep.*, 2014, 6665.

1335 M. Galdzicki, K. P. Clancy, E. Oberortner, M. Pocock, J. Y. Quinn, C. A. Rodriguez, N. Roehner, M. L. Wilson, L. Adam, J. C. Anderson, B. A. Bartley, J. Beal, D. Chandran, J. Chen, D. Densmore, D. Endy, R. Grunberg, J. Hallinan, N. J. Hillson, J. D. Johnson, A. Kuchinsky, M. Lux, G. Misirli, J. Peccoud, H. A. Plahar, E. Sirin, G. B. Stan, A. Villalobos, A. Wipat, J. H. Gennari, C. J. Myers and H. M. Sauro, The Synthetic Biology Open Language (SBOL) provides a community standard for communicating designs in synthetic biology, *Nat. Biotechnol.*, 2014, **32**, 545–550.

1336 N. Roehner, E. Oberortner, M. Pocock, J. Beal, K. Clancy, C. Madsen, G. Misirli, A. Wipat, H. Sauro and C. J. Myers, A Proposed Data Model for the Next Version of the Synthetic Biology Open Language, *ACS Synth. Biol.*, 2014, DOI: 10.1021/sb500176h.

1337 M. Hucka, A. Finney, H. M. Sauro, H. Bolouri, J. C. Doyle, H. Kitano, A. P. Arkin, B. J. Bornstein, D. Bray, A. Cornish-Bowden, A. A. Cuellar, S. Dronov, E. D. Gilles, M. Ginkel, V. Gor, I. I. Goryanin, W. J. Hedley, T. C. Hodgman, J. H. Hofmeyr, P. J. Hunter, N. S. Juty, J. L. Kasberger, A. Kremling, U. Kummer, N. Le Novere, L. M. Loew, D. Lucio, P. Mendes, E. Minch, E. D. Mjolsness, Y. Nakayama, M. R. Nelson, P. F. Nielsen, T. Sakurada, J. C. Schaff, B. E. Shapiro, T. S. Shimizu, H. D. Spence, J. Stelling, K. Takahashi, M. Tomita, J. Wagner and J. Wang, The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models, *Bioinformatics*, 2003, **19**, 524–531.

1338 N. Le Novère, M. Hucka, H. Mi, S. Moodie, F. Schreiber, A. Sorokin, E. Demir, K. Wegner, M. Aladjem, S. M. Wimalaratne, F. T. Bergman, R. Gauges, P. Ghazal, K. Hideya, L. Li, Y. Matsuoka, A. Villéger, S. E. Boyd, L. Calzone, M. Courtot, U. Dogrusoz, T. Freeman, A. Funahashi, S. Ghosh, A. Jouraku, S. Kim, F. Kolpakov, A. Luna, S. Sahle, E. Schmidt, S. Watterson, G. Wu, I. Goryanin, D. B. Kell, C. Sander, H. Sauro, J. L. Snoep, K. Kohn and H. Kitano, The systems biology graphical notation, *Nat. Biotechnol.*, 2009, **27**, 735–741.

1339 N. Le Novère, A. Finney, M. Hucka, U. S. Bhalla, F. Campagne, J. Collado-Vides, E. J. Crampin, M. Halstead, E. Klipp, P. Mendes, P. Nielsen, H. Sauro, B. Shapiro, J. L. Snoep, H. D. Spence and B. L. Wanner, Minimum information requested in the annotation of biochemical models (MIRIAM), *Nat. Biotechnol.*, 2005, **23**, 1509–1515.

1340 M. Courtot, N. Juty, C. Knüpfer, D. Waltemath, A. Zhukova, A. Dräger, M. Dumontier, A. Finney, M. Golebiewski, J. Hastings, S. Hoops, S. Keating, D. B. Kell, S. Kerrien, J. Lawson, A. Lister, J. Lu, R. Machne, P. Mendes, M. Pocock, N. Rodriguez, A. Villéger, D. J. Wilkinson, S. Wimalaratne, C. Laibe, M. Hucka and N. Le Novère, Controlled vocabularies and semantics in Systems Biology, *Mol. Syst. Biol.*, 2011, **7**, 543.

1341 R. Goodacre, D. Broadhurst, A. Smilde, B. S. Kristal, J. D. Baker, R. Beger, C. Bessant, S. Connor, G. Capuani, A. Craig, T. Ebbels, D. B. Kell, C. Manetti, J. Newton, G. Paternostro, R. Somorjai, M. Sjöström, J. Trygg and F. Wulfert, Proposed minimum reporting standards for data analysis in metabolomics, *Metabolomics*, 2007, **3**, 231–241.

1342 C. B. Anfinsen, E. Haber, M. Sela and F. H. White, The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain, *Proc. Natl. Acad. Sci. U. S. A.*, 1961, **47**, 1309–1314.

1343 C. B. Anfinsen, Principles that govern the folding of protein chains, *Science*, 1973, **181**, 223–230.

1344 T. Buzan, *How to mind map*, Thorsons, London, 2002.

1345 D. B. Kell, Genotype:phenotype mapping: genes as computer programs, *Trends Genet.*, 2002, **18**, 555–559.

1346 D. Broadhurst and D. B. Kell, Statistical strategies for avoiding false discoveries in metabolomics and related experiments, *Metabolomics*, 2006, **2**, 171–196.