



# Biocatalysis

Elizabeth L. Bell<sup>1</sup>, William Finnigan<sup>1</sup>, Scott P. France<sup>2</sup>, Anthony P. Green<sup>1</sup>, Martin A. Hayes<sup>3</sup>, Lorna J. Hepworth<sup>1</sup>, Sarah L. Lovelock<sup>1</sup>, Haruka Niikura<sup>4</sup>, Silvia Osuna<sup>5,6</sup>, Elvira Romero<sup>3</sup>, Katherine S. Ryan<sup>4</sup>, Nicholas J. Turner<sup>1</sup> and Sabine L. Flitsch<sup>1</sup> ✉

**Abstract** | Biocatalysis has become an important aspect of modern organic synthesis, both in academia and across the chemical and pharmaceutical industries. Its success has been largely due to a rapid expansion of the range of chemical reactions accessible, made possible by advanced tools for enzyme discovery coupled with high-throughput laboratory evolution techniques for biocatalyst optimization. A wide range of tailor-made enzymes with high efficiencies and selectivities can now be produced quickly and on a gram to kilogram scale, with dedicated databases and search tools aimed at making these biocatalysts accessible to a broader scientific community. This Primer discusses the current state-of-the-art methodology in the field, including route design, enzyme discovery, protein engineering and the implementation of biocatalysis in industry. We highlight recent advances, such as de novo design and directed evolution, and discuss parameters that make a good reproducible biocatalytic process for industry. The general concepts will be illustrated by recent examples of applications in academia and industry, including the development of multistep enzyme cascades.

Enzymes have been employed for a wide variety of chemical processes for decades (FIG. 1). For example, nitrile hydratases are used to make acrylamide on the thousands of tons scale, and enzymes have been added to detergents for more than 30 years<sup>1,2</sup>. More recently, the use of proteins as catalysts for chemical synthesis of more complex molecules, such as pharmaceuticals, has become increasingly widespread. Enzymes are particularly powerful because they merge the advantages of a directing group controlling selectivity and a catalyst in a single reagent<sup>3</sup>, which can also be used with other enzymes in a one-pot reaction. Over the past 20 years, combined synthetic–enzymatic systems have enabled multiple total synthesis endeavours, and the use of enzymes is becoming routine in some process chemistry groups in industry<sup>4</sup>. Until recently, only a subset of enzymes, such as lipases or ketoreductases (KREDs), were available for chemical synthesis applications<sup>5</sup>. However, the growth of potential sources of enzymes for process chemistry applications has accelerated, resulting in a diverse toolkit of enzymes now available to researchers. In 2014, the development of a total enzymatic synthesis of the nucleoside didanosine highlighted the possibility of ‘bio-retrosynthesis’<sup>6</sup>. Based on the principles of retrosynthesis, where the target molecule is transformed into simple precursors by ‘breaking’ bonds that can be formed from synthetic transformations, ‘bio-retrosynthesis’ involves

the design of an artificial enzyme cascade — a synthetic biochemical pathway — that offers a possible route towards the desired target molecule by choosing enzymes as catalysts for the required chemistry. The fully biocatalyst-driven synthesis of the HIV inhibitor islatravir (FIG. 1), which will be discussed in more detail in the Applications section, demonstrates the power of combining modern approaches towards designing new enzyme cascades, including repurposing of known biosynthetic pathways, screening of saturation mutagenesis libraries of enzyme variants and directed evolution against selected residues towards increased enzyme stability and turnover<sup>7</sup>.

In this Primer, we discuss the different development stages (reaction design, biocatalyst choice and optimization, and bioprocess development) that can lead to a range of industrial products as shown in FIG. 1. These stages are interdependent and need to be closely integrated. Starting with a target molecule, a single or multi-step biocatalytic process needs to be designed, often by manual design using expertise and precedent literature from organic synthesis and biocatalysis. More recently, programmes such as *RetroBioCat*<sup>8</sup> including biocatalyst databases are being developed to speed up this step and enable the automatic design of de novo biosynthetic pathways. Once a process has been designed, suitable enzymes need to be selected for each step and tested (Experimentation). The increasing adoption of

<sup>1</sup>School of Chemistry and MIB, The University of Manchester, Manchester, UK.

<sup>2</sup>Pfizer Worldwide Research and Development, Groton, CT, USA.

<sup>3</sup>Compound Synthesis and Management, Discovery Sciences, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden.

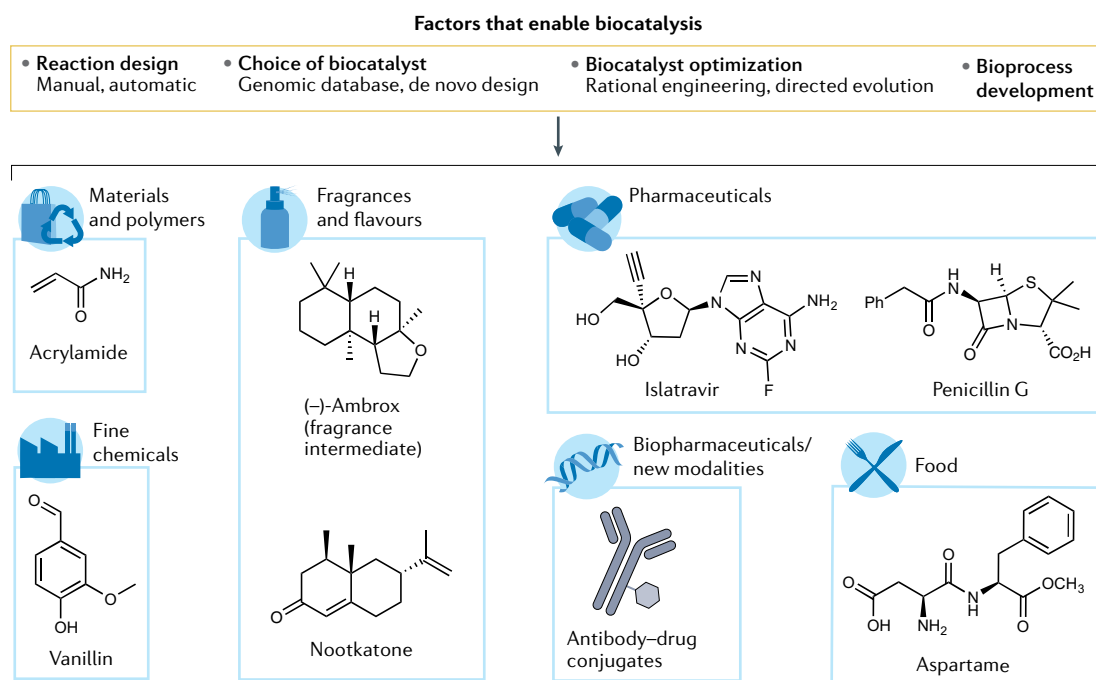
<sup>4</sup>Department of Chemistry, The University of British Columbia, Vancouver, British Columbia, Canada.

<sup>5</sup>CompBioLab Group, Institut de Química Computacional i Catalisi, Departament de Química, Universitat de Girona, Girona, Spain.

<sup>6</sup>ICREA, Barcelona, Spain.

✉e-mail: [sabine.flitsch@manchester.ac.uk](mailto:sabine.flitsch@manchester.ac.uk)

<https://doi.org/10.1038/s43586-021-00044-z>



**Fig. 1 | Examples of different products synthesized using biocatalysis.** Biocatalysis is enabled by various factors including reaction design, biocatalyst choice and optimization as well as bioprocess development. It has been used for the synthesis of a wide range of chemicals and pharmaceuticals, including those shown.

biocatalysis by the pharmaceutical industry has been driven by innovative tools in protein engineering, which allow fast optimization of catalyst activity, including laboratory evolution and computational design (Experimentation). As a result, strict reaction parameters (Results) can now be met at reasonable timescales for successful bioprocess development (Applications). These parameters include non-physiological reaction conditions such as high activity on non-natural substrates, high temperature, high concentration of substrates and tolerance of organic solvents and wide pH ranges. Alongside protein engineering tools, databases of available biocatalysts with their reaction profiles are starting to be established (Reproducibility and data deposition). We also detail the current limitations of biocatalysis and areas of importance for further advancing this method to expand the breadth of applications (Limitations and optimizations). Finally, we highlight what the future holds for biocatalysis and the impact it will likely have in the next decade (Outlook).

### Experimentation

In this section, we highlight several sources available to scientists looking for an enzyme as a starting point to develop a new biocatalyst. We discuss how one can optimize biocatalysts using directed evolution and computational design as well as how to incorporate non-canonical amino acids to enable novel chemistries.

### Sources of biocatalysts

Enzymes can be sourced from a few outlets that include commercial sources, adaptation of enzymes from biosynthesis, screening of metagenomic libraries and in silico mining of databases.

**Commercial sources of biocatalysts.** Purified enzymes or lyophilized crude cell lysates are often available for direct purchase through chemical vendors (FIG. 2a). For example, one can purchase specific dehydrogenases, reductases or carbohydrate-active enzymes, and directly employ them in chemical synthesis. Libraries of commercially produced enzymes can be screened against specific substrates to identify candidate biocatalysts. Available libraries include oxidoreductases (KREDs, imine reductases (IREDs), ene reductases, Baeyer–Villiger monooxygenases, monoamine oxidases), transferases (transaminases), lyases (nitrilases, halohydrin dehalogenases, acylases), carbon–carbon bond-forming enzymes and carbon–nitrogen bond-forming enzymes<sup>9</sup> (BOX 1).

In addition to these types of commercial sources and kits, the wide availability of commercial gene synthesis means that, in principle, any enzyme with a known amino acid sequence can be obtained through gene synthesis. Researchers can order the synthetic gene corresponding to the protein, recombinantly express it in a desirable host organism or by cell-free protein synthesis and, then, purify the protein for testing as one would similarly do with a catalyst in a chemical synthesis. Databases ranging from SciFinder (with a licence) to UniProt<sup>10</sup> and BioCatNet<sup>11</sup> allow researchers to identify enzymes that catalyse desired chemical transformations. Thus, through the combination of publicly available sequence data and commercial gene synthesis, any enzyme reported is available to the researcher, for a cost.

**Natural product biosynthetic enzymes.** Enzymes involved in the synthesis of specialized metabolites, or natural products, are particularly useful as starting points for biocatalysis. Natural products tend to have diverse chemical

#### Enzyme cascade

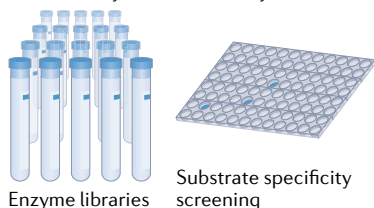
Within the biocatalysis community this term is used broadly for concurrent, multienzyme one-pot biocatalytic reactions as well as reactions in which components are added sequentially or process steps are telescoped.

#### Metagenomic libraries

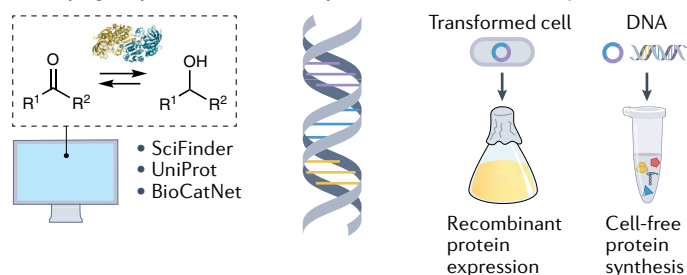
Genomic libraries constructed by the direct cloning of the large fragments of the environmental DNA into an appropriate vector, transformed into the host bacteria.

**a Commercial sources of biocatalysts****Chemical vendors**

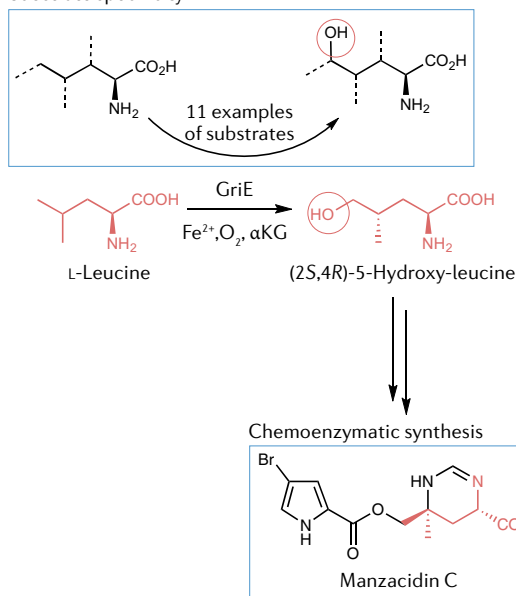
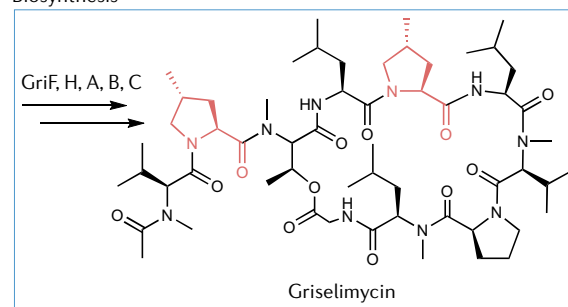
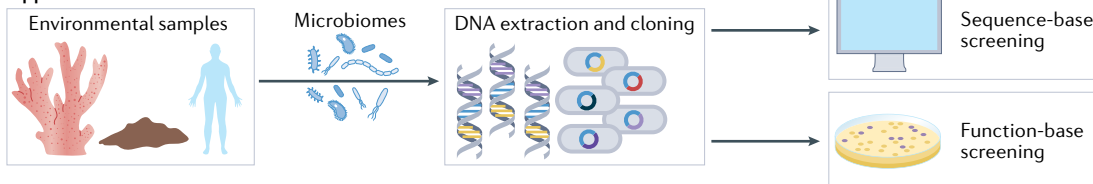
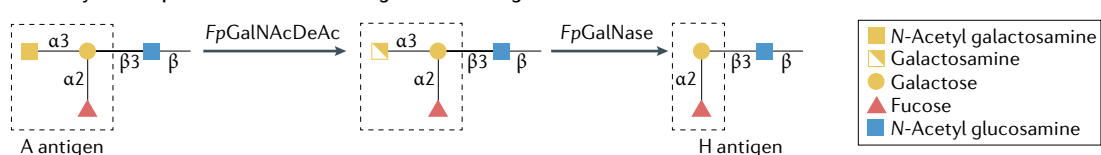
Purified enzymes, crude cell lysates

**Enzymes from the synthesized gene**

Identifying enzymes → Gene synthesis → Protein expression

**b Natural product biosynthetic enzymes****Fe(II) and 2-oxoglutarate-dependent enzyme, GriE**

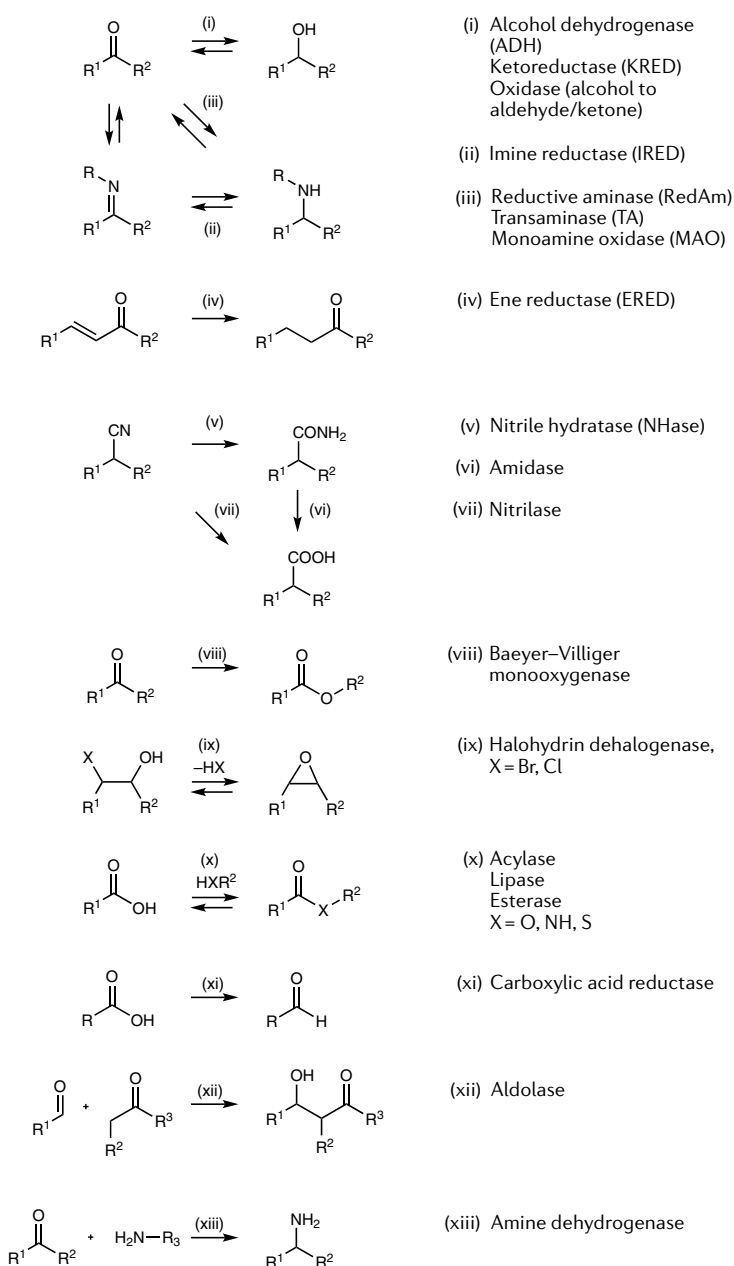
Substrate specificity

**Biosynthesis****c Metagenomic and in silico screening****Approaches****Biocatalyst example: conversion of A antigen into H antigen**

**Fig. 2 | Sources of biocatalysts.** **a** | Types of biocatalyst useful in chemoenzymatic synthesis and biochemical applications. **b** | Natural product biosynthetic enzymes. In the biosynthetic pathway of griselimycin, the Fe(II) and 2-oxoglutarate-dependent enzyme, GriE, reacts with L-leucine to give (2S,4R)-5-hydroxy-leucine. This regioselective and stereoselective hydroxylation enables a chemoenzymatic synthesis of manzacidin C. **c** | Metagenomic screening approaches. Top: first step of metagenomics screening is extraction of total DNA from environmental samples such as marine sponges or soil microorganism communities. This DNA is constructed into a metagenomic library and is commonly screened using sequence-based or function-based approaches. Bottom: enzymatic conversion of A/B antigens into H antigen. All blood group antigens depicted are type I.

### Box 1 | Common biocatalysis enzymes and their associated chemical transformations

The interconversion between alcohols and aldehydes or ketones is catalysed by alcohol dehydrogenases, ketoreductases (KREDs) and alcohol oxidases (reaction i). Imine reductases catalyse the reduction of imines to amines (reaction ii). Reductive aminases and transaminases transfer amino groups to prochiral ketones or aldehydes, giving rise to chiral amine products. Monoamine oxidases catalyse the oxidation of compounds containing one amino group and the resulting imine spontaneously forms a ketone in water (reaction iii). Ene reductases reduce C–C double bonds (reaction iv). Nitrile hydratases catalyse the addition of water across nitriles producing amides (reactions v, vi). The hydrolysis of nitriles to carboxylic acids and ammonia is catalysed by nitrilases (reaction vii). Baeyer–Villiger monoxygenases oxidize ketones to esters (reaction viii). Halohydrin dehalogenases produce enantiopure epoxides and corresponding ring-opening products (reaction ix). Acylases hydrolyse amide bonds, whereas lipases and esterases do so for ester bonds (reaction x). Part of the esterase group, thioesterases catalyse the hydrolysis of the thioester bonds. Carboxylic acid reductases reduce aromatic or aliphatic acids to aldehydes (reaction xi). Aldolases catalyse Aldol reactions (reaction xii). Amine dehydrogenases catalyse the reductive amination of carbonyl compounds (reaction xiii).



structures, and studies on the biosynthesis of such natural products have unveiled a correspondingly diverse set of biosynthetic enzymes. Therefore, natural product biosynthetic enzymes are a potential source for diverse catalysts. A recent review discusses the wide-ranging chemical and enzymatic diversity found in natural product biosynthesis<sup>12</sup>. From a biocatalytic point of view, the most important criteria in selecting a potential biosynthetic enzyme include its substrate specificity, cofactor dependence, turnover, stability, functional recombinant expression and ability to perform a stand-alone function outside its natural pathway within a cell.

One group of biosynthetic enzymes widely found in natural product biosynthesis are oxidative Fe(II) and 2-oxoglutarate-dependent enzymes, which can catalyse challenging reactions such as hydroxylation, halogenation and oxidative cyclizations, typically for C(sp<sup>3</sup>)–H functionalization, due to the reactive intermediates generated in the catalytic cycle<sup>13</sup>. These powerful enzymes have emerged as useful biocatalysts for chemical synthesis<sup>14</sup>. One example is GriE — an enzyme involved in the synthesis of the peptide griselimycin in *Streptomyces*<sup>15</sup> that hydroxylates the  $\delta$ -carbon of L-leucine and was employed for a key step in the total synthesis of manzacidin C<sup>16</sup> (FIG. 2b). Other examples include halogenases such as BesD, which chlorinate free amino acids to generate non-proteinogenic amino acids<sup>17</sup>. A final example is KabC, which catalyses an oxidative cyclization and was employed in a biocatalytic synthesis of the neurochemical kainic acid<sup>18</sup>. These types of Fe(II) and 2-oxoglutarate-dependent enzymes represent a promising group of biocatalysts for further development as they can be reconstituted either using *Escherichia coli* cell lysates or in vitro<sup>19</sup>.

Other known natural product biosynthetic enzymes with demonstrated use in chemical synthesis include methyltransferases<sup>20</sup>, Diels–Alderses that catalyse cycloadditions<sup>21</sup>, halogenases<sup>22</sup>, uridine diphosphate-dependent glycosyltransferases<sup>23</sup>, laccases<sup>24</sup> and pyridoxal phosphate-dependent enzymes<sup>25</sup>. However, biosynthetic enzymes are typically sluggish catalysts, with Michaelis constant  $k_{\text{cat}}$  values typically about 30-fold lower than those of primary metabolism<sup>26</sup>, which is probably owing to the constraints at work in their evolutionary histories<sup>26,27</sup>. To this end, biosynthetic enzymes are viewed as starting points for directed evolution efforts to improve rates of catalysis, substrate tolerance as well as stability and solubility in desired solvents, which include ionic liquids and water-soluble organic solvents<sup>23</sup>.

**Metagenomic and in silico screening.** Metagenomic libraries are an additional source for new biocatalysts<sup>28</sup>; these are genomic libraries of DNA obtained from environmental samples such as marine sponges, soil or faeces<sup>29</sup>. Functional-based screening or sequence-based screening is often used to search through a genomic library to find enzymes of interest (FIG. 2c). Functional-based screening can incorporate the use of colorimetric assays<sup>30–32</sup>, mechanism-based probes<sup>33</sup> and/or droplet microfluidics<sup>34</sup>. Researchers are looking for enzymes in these libraries that demonstrate catalysis of the desired

**C(sp<sup>3</sup>)-H functionalization**

A type of reaction in which a C-H bond, in which the carbon is sp<sup>3</sup> hybridized, is cleaved and a new C-X bond is formed (where X is usually carbon, oxygen, nitrogen or a halide).

**Diels-Alderases**

Enzymes that catalyse a [4 + 2] cycloaddition reaction between a conjugated diene and a substituted alkene forming a cyclohexene derivative.

**Rates of catalysis**

The rates by which substrates are converted into products in catalytic reactions.

**Saturation mutagenesis**

A method that allows the randomization of a target codon or set of codons in a gene.

**Iterative combinatorial active site testing**

A method that allows the generation of DNA libraries where active site positions are randomized in pairs

target reaction within these screening platforms. For example, mammalian microbiomes are a rich source of carbohydrate-active enzymes, which are often encoded in polysaccharide utilization loci in specific bacterial genomes. Screening of a human metagenomic library using a fluorogenic substrate identified a pair of enzymes that convert the A antigen into the H antigen of O-type blood, enabling a biocatalytic approach to produce universal donor O-type blood from A-type blood<sup>35</sup>.

By contrast, sequence-based screening is based on finding new enzymes with sequence similarity to known enzymes. One approach involves the use of PCR amplification of genomic sequences with degenerate primers. These primers are a mixture of oligonucleotide sequences that code for highly conserved regions of a desired type of catalyst, and therefore allow amplification of genes encoding for this catalyst from naturally derived DNA libraries. For example, sponge microbiomes are known for their biosynthetic capabilities, and in a case of sequence-based screening with degenerate primers, a new halogenase *Krml* was isolated from a sponge microbiome and employed for the regioselective halogenation of tryptophan and a range of indole-derived peptide substrates<sup>36</sup>.

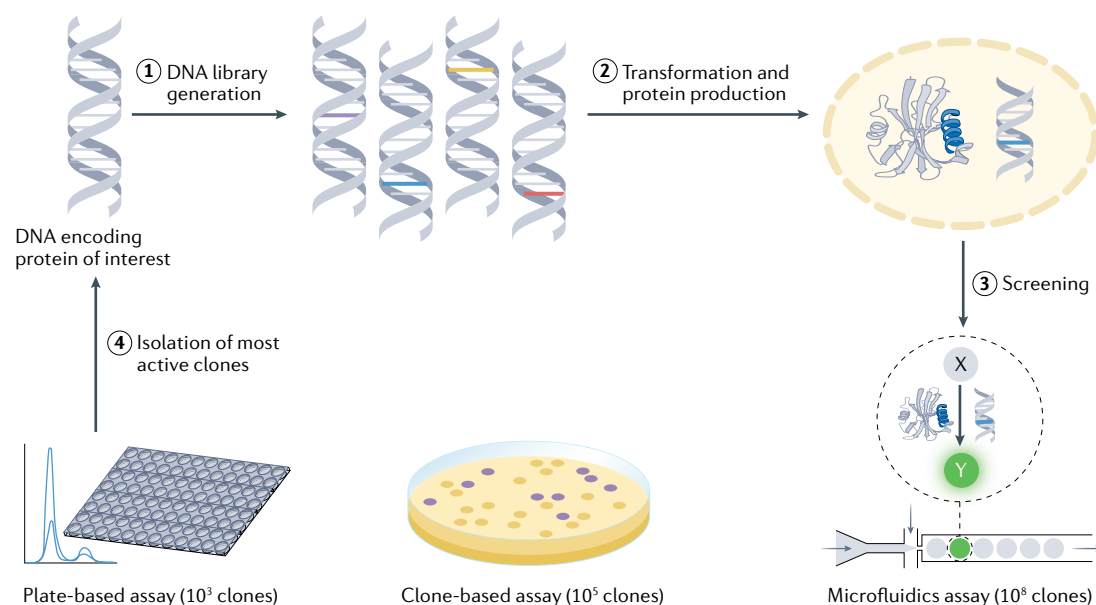
Another approach is the direct analysis of sequencing data, where specific types of desired enzymes can be found through *in silico* analysis<sup>37</sup>. For example, the sequencing of a domestic drain metagenome was analysed *in silico* to find transaminase candidates, which were then tested for their ability to carry out transaminations on diverse substrates. One enzyme from this set of candidates retained activity in 50% dimethyl sulfoxide (DMSO), a solvent tolerance that had not been

reported in a transaminase previously. As these examples demonstrate, once a gene encoding a desired enzyme is identified, the gene can be cloned and the enzyme can be expressed recombinantly to establish its function<sup>38</sup>.

**Directed evolution**

Wild-type enzymes are often not suitable for direct use in industrial applications and must first undergo optimization to improve properties such as substrate specificity and selectivity as well as catalytic efficiency and stability. Directed evolution is a powerful and versatile technology for adapting these enzymes to perform new functions (as highlighted by the award of the 2018 Nobel Prize for Chemistry)<sup>39–41</sup>. The directed evolution cycle involves iterative rounds of DNA library design and generation, gene expression and screening of enzyme library members (FIG. 3). Multiple properties can be optimized in parallel and improved variants can be isolated, characterized and used as templates for further rounds of evolution.

Following identification of a suitable starting template, DNA libraries are generated using numerous standard molecular biology techniques, such as random mutagenesis or site saturation mutagenesis. The chosen method of library generation depends on factors such as the availability of structural information and screening capacity. The design of smaller, more focused libraries (10<sup>2</sup>–10<sup>4</sup> variants) often employs computational modelling and bioinformatics to guide the selection of amino acid residues for randomization. These libraries are generated using techniques such as saturation mutagenesis or iterative combinatorial active site testing<sup>42</sup> and often employ reduced codons<sup>43,44</sup>. Larger library diversity



**Fig. 3 | Directed evolution cycle.** First step in the evolution cycle involves creating a library of DNA variants using standard molecular biology techniques (step 1). Host cells are transformed with the DNA library and target proteins are produced (step 2). Proteins are evaluated using high-throughput screening techniques such as plate-based assays coupled with high performance liquid chromatography analysis, colony-based assays coupled with colorimetric detection or microfluidics-based assays coupled with fluorescence detection, where X represents the starting material and Y represents the fluorescent product or by-product (step 3). Final step of the evolution cycle involves isolation and characterization of the most active clones. The most active variant will serve as a template for subsequent rounds of evolution (step 4).



**Error-prone PCR**

A PCR (polymerase chain reaction) that is run under reaction conditions that introduce random mutations into the target DNA sequence.

**Gene shuffling**

A method that allows for the generation of chimeric libraries of genes.

**DNA shuffling**

A method that allows the recombination of beneficial mutations in a directed evolution experiment.

**High performance liquid chromatography**

An analytical technique that allows for the rapid separation and quantification of compound mixtures using pressurized liquid solvent passed through chromatographic columns.

( $10^5$ – $10^8$  variants) can be generated using techniques such as error-prone PCR<sup>45</sup> and gene shuffling<sup>46</sup>. It is common to use multiple mutagenesis techniques during enzyme evolution to target different regions of the protein structure. For example, focused active site mutations can be beneficial for reshaping substrate binding pockets and improving activity towards non-native substrates, and mutations to the protein surface and flexible loop regions can often result in improved solvent tolerance and thermostability. Beneficial mutations are typically combined during evolutionary optimization using DNA shuffling and can be guided by computational algorithms.

Transforming cells with DNA libraries leads to spatial separation of library members and establishes a link between genotype and phenotype that must be maintained during protein production and screening to allow characterization of individual library members. Arraying colonies into multiwell plates for protein production and screening offers the greatest versatility, as variants can be evaluated using a wide range of chromatographic, spectrophotometric and spectroscopic techniques. Although chromatographic methods are of relatively low throughput, they are commonly employed for applications in industrial biocatalysis as the assays are compatible with screening under process conditions, which often employ high substrate loadings ( $>100\text{ g l}^{-1}$ ), co-solvents (for example, up to 50% DMSO in aqueous media) and high temperatures (40–50 °C). This workflow can also be automated to improve speed, accuracy and throughput. For example, colony pickers allow users to array  $10^3$  colonies per hour into 96-well plates, liquid handling robots accelerate aliquoting and transfer steps required for library generation and protein production, and reaction analysis using state-of-the-art ultra-high performance liquid chromatography systems allows evaluation of  $10^3$  clones per instrument per day. GSK used this approach to engineer an enantioselective IRED with a 38,000-fold improvement over the wild-type enzyme<sup>47</sup>. This IRED variant was employed in a reductive amination and kinetic resolution step to manufacture the lysine-specific demethylase 1 (LSD1) inhibitor GSK2879552, a treatment for small cell lung cancer and acute leukaemia. Following a similar approach, Codexis and Merck have engineered five different enzymes, which form part of an impressive nine-enzyme cascade process to manufacture the HIV treatment islatravir<sup>7</sup> (see Applications section for a more detailed discussion and reaction scheme).

In order to evaluate larger libraries of  $10^5$ – $10^8$  variants, more specialized screening approaches can be employed, including colony-based assays<sup>48–50</sup>, fluorescence-activated cell sorting<sup>51</sup>, phage display<sup>52</sup>, microfluidic-based screening<sup>53,54</sup>, selection-based approaches<sup>55</sup> and continuous evolution<sup>56,57</sup>. For example, monoamine oxidase from *Aspergillus niger* (MAO-N) has been extensively engineered for the selective oxidation of a wide range of amine substrates using a colorimetric colony-based assay that relies on the detection of the hydrogen peroxide co-product by horseradish peroxidase (HRP) and a reactive dye<sup>58</sup>. The throughput of oxidase evolution based on hydrogen peroxide detection can be further increased by screening

variants in picolitre droplets. Indeed, in a recent study, an ultra-high-throughput microfluidic assay was used to screen a library of  $10^7$  cyclohexylamine oxidase variants and, after only a single round of evolution, the most improved variant isolated had a 960-fold improvement in catalytic efficiency<sup>59</sup>. Although the screening capacity in this example is greatly increased, picolitre droplet sorting is currently restricted to fluorescence as a detection method, which limits the versatility of this approach. Ongoing research is focused on coupling microfluidics with alternative methods of detection, such as mass spectrometry, to provide approaches for a wide host of chemistries<sup>60</sup>.

Selection-based approaches including continuous evolution platforms<sup>53–55</sup>, where improved catalyst performance is linked to cell viability, offer ultra-high-throughput screening capabilities ( $10^6$ – $10^{10}$  variants). These methods are highly specialized as improvements in the enzyme activity of interest must be linked to cell survival. Optimization of these multicomponent systems requires considerable effort and can take up to several years; it is important to have control over the stringency of the selection pressure and to ensure that the host organism is not able to evolve alternative mechanisms of survival. However, for enzymes whose native activities can be associated with organism fitness, this type of assay is particularly powerful as it allows rapid evaluation of broad sequence space. A key example is the development of a selection-based method for engineering pyrrolysyl-tRNA synthetases (PylRS) for the genetic incorporation of non-canonical amino acids into proteins<sup>55</sup>. This versatile approach has been applied by numerous research groups to the evolution of a panel of PylRS variants, which are now able to accept more than 150 different non-canonical amino acids<sup>61</sup>.

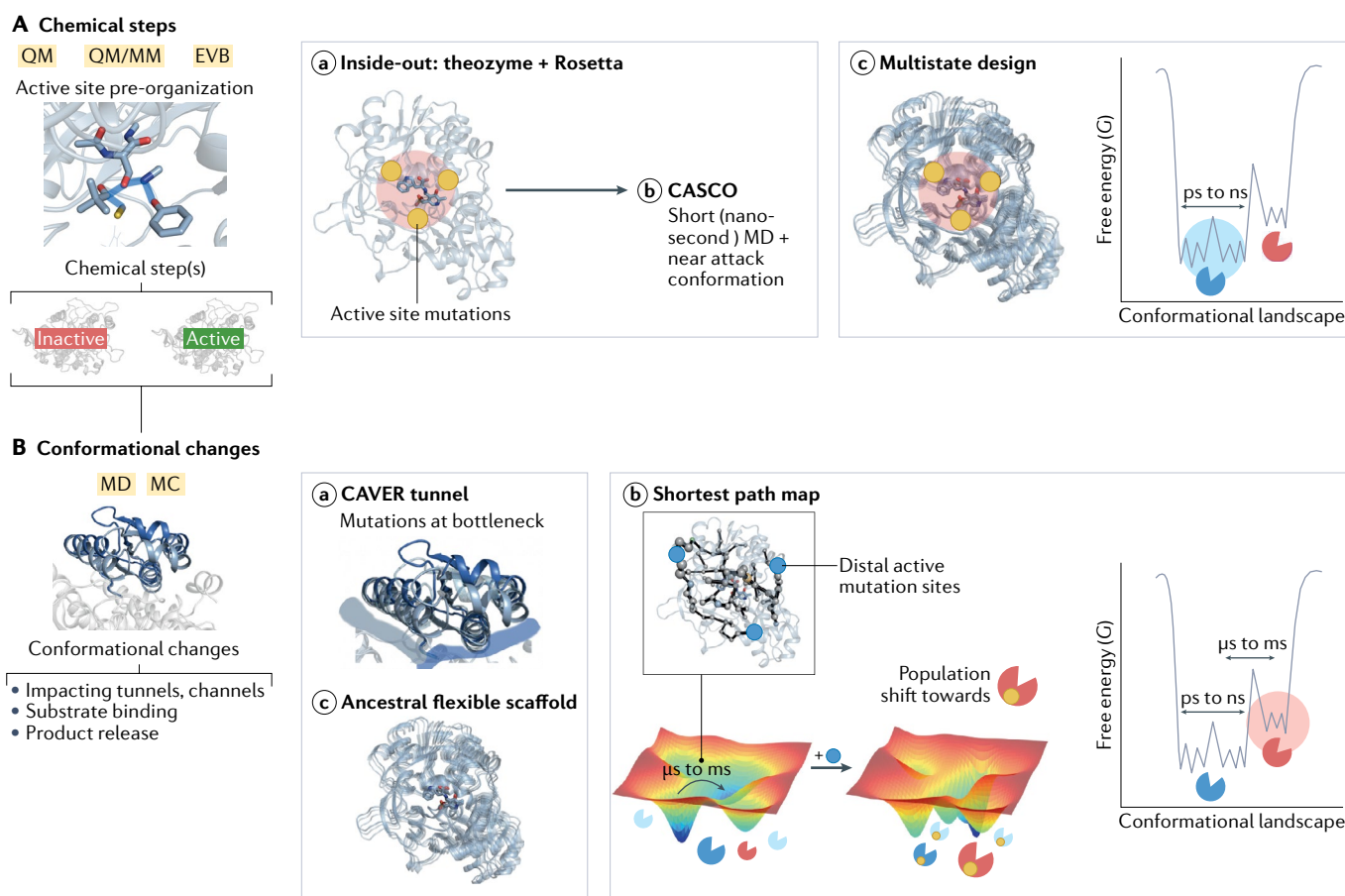
**Computational enzyme design and engineering**

For the most part, the high efficiency of enzymes in accelerating chemical reactions has been attributed to their highly pre-organized active site pockets that precisely position the catalytic residues for transition state stabilization<sup>62</sup>. This precise arrangement in the active site pocket to optimize the chemical steps is complemented by the inherent flexibility of the enzyme structure. Enzymes can adopt multiple conformations that often play critical roles in equally important processes, such as substrate binding and/or product release for restarting the catalytic cycle<sup>63</sup>. To this end, computational enzyme design protocols should propose specific amino acid changes (located in the active site but also in remote positions) to achieve highly pre-organized active site pockets for transition state stabilization, and optimize the enzyme conformational ensemble to favour substrate binding and product release<sup>64</sup>.

In practice, available computational protocols focus only on a selected set of the complex features of enzyme catalysis; that is, they design enzymes based on either the chemical steps of the desired chemical transformations (see FIG. 4A), the substrate binding/product release process or the enzyme conformational dynamics (FIG. 4B). Different computational techniques are needed for each of the above features (see FIG. 4).

Initial attempts to rationally design enzymes were focused on the chemical steps of the process (FIG. 4A, and selected examples FIG. 4Aa–c). The transition states of the desired transformation in the theoretical enzyme or ‘theozyme’ active site pocket is modelled with quantum mechanics to assess the potential rate acceleration and the ideal geometric constraints for optimal transition state stabilization (FIG. 4Aa). This optimal arrangement that contains only a few active site residues owing to the high computational cost of quantum mechanics calculations is then grafted onto an existing protein scaffold, and further optimized by means of Rosetta or other related protein design software<sup>65,66</sup>. Further refinements to this original formulation (named inside-out) can be made by incorporating data on protein conformational dynamics by means of short molecular dynamics simulations. In these simulations, the enzyme variant

is immersed in a water solvent box, and whether the optimal arrangement of the catalytic residues — also known as the near attack conformation — is maintained throughout the simulation time is assessed. A higher number of near attack conformations explored during the molecular dynamics is attributed to a higher catalytic activity and/or selectivity as the catalytic residues are properly arranged for catalysis most of the simulation time. These observations resulted in the development of some computational methodologies based on Rosetta and molecular dynamics simulations for enhancing the enzyme activity and selectivity (catalytic selectivity by computational design (CASCO), as shown in FIG. 4Ab) or thermostability (framework for rapid enzyme stabilization by computational libraries (FRESCO))<sup>67</sup>. Additional refinements such as the use of an ensemble of closely related enzyme conformations



**Fig. 4 | Computational approaches to enzyme design. A |** Methods that consider the chemical steps of the desired chemical transformations: quantum mechanics (QM), quantum mechanics/molecular mechanics (QM/MM) and empirical valence bond (EVB). Quantum mechanics (QM) is used to model the active site and associated transition states of the desired transformation in the theozyme to evaluate potential changes for rate acceleration. This optimal arrangement is then grafted onto an existing protein scaffold and further optimized by protein design software, such as Rosetta (part Aa). Protein conformational dynamics can be incorporated using short molecular dynamics (MD) simulations to refine this estimation, ultimately enhancing enzyme activity and selectivity (catalytic selectivity by computational design (CASCO)) (part Ab). Additional refinements for multistate design allow for considering a set of related structures often coming from short-timescale MD simulations in the design process (part Ac). **B |** Methods that consider the

conformational changes key for substrate binding or product release: MD simulations and Monte Carlo (MC) simulations. CAVER software is used to model the different tunnels that are available for substrate binding to the active site and product release (part Ba). Tunnel analysis allows identification of the narrowest regions, that is the bottleneck, where mutations are usually introduced (part Ba). Recent approaches take advantage of the information gained from extensive MD simulations for identifying key conformationally relevant residues (not strictly located at the active site), which by mutation can induce a shift in the relative populations of the conformational states stabilizing the catalytically relevant ones (shortest path map, SPM) (part Bb). Selection of an ancestral enzyme scaffold displaying high levels of flexibility can be also used for the design of new enzymatic activities (part Bc). Images of proteins throughout the figure were generated using PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC.)

from either normal mode analysis, short molecular dynamics simulations or small perturbations in the enzyme backbone angles for multi-state design (FIG. 4Ac) were proposed to include some limited protein flexibility in the design process<sup>68</sup>. Although these strategies include some protein flexibility during the design, the ensemble of conformations used is rather similar as they come from usually short (picosecond to nanosecond) molecular dynamics simulations. Other strategies are based on computing the direct effect of the included mutations on the activation barrier of the enzyme-catalysed process (with the computationally more demanding quantum mechanics/molecular mechanics or empirical valence bond (EVB) strategies, as shown in FIG. 4A), rather than estimating the effect by means of some key geometrical constraints (as in the near attack conformation analysis)<sup>69</sup>.

The importance of enzyme conformational dynamics for enzyme design gained popularity in recent years<sup>70–72</sup> (FIG. 4B). Conformationally flexible loops adjacent to the active site pocket can regulate substrate binding and/or product release, and some studies have shown these loops as crucial for enhanced enzymatic activity in many enzyme families<sup>73</sup>. Bioinformatic tools such as CAVER have been developed to identify tunnels and channels, and to suggest potential mutational hotspots for novel catalytic activity<sup>74</sup> (FIG. 4Ba). The analysis of some natural and laboratory evolution pathways demonstrated that increased enzymatic activity is often achieved by introducing mutations that alter the enzyme's conformational ensemble<sup>75</sup>. These mutations can be located at the active site or may be located at distal positions and induce a long-range effect that impacts the enzyme active site pocket and, thus, catalysis. This impact on enzymatic activity is often achieved by favouring the enzyme conformational states that are key for the novel functionality (catalytically productive conformations), while disfavoring non-productive conformational states, thus converting computational enzyme design into a population shift problem<sup>64</sup>. In this direction, some conformationally driven computational approaches focused on identifying such long-range allosteric networks of interactions include the shortest path map (SPM) tool and have been used recently for predicting distal and active site mutations<sup>76</sup> (FIG. 4Bb). Multistate computational design based on ensembles of enzyme conformations taken from room-temperature X-ray crystallography corresponds to a successful strategy for efficient computational enzyme design<sup>77</sup>. The reconstruction of ancestral enzymes that display a higher degree of flexibility than their modern counterparts and their use as initial scaffolds for enzyme design has additionally yielded interesting new insights<sup>78</sup>. The higher flexibility observed in many ancestral variants was key for achieving high catalytic activity with only a few mutations located at the active site. Ancestrally reconstructed enzymes are usually less specialized than their modern counterparts, thus often presenting higher levels of substrate and catalytic promiscuity<sup>79</sup>, which makes them excellent starting protein scaffolds for enzyme design (FIG. 4Bc). These examples indicate that both the selection of a conformationally rich scaffold and the consideration of multiple enzyme

conformations is crucial for successful computational enzyme design.

### Biocatalysts with non-canonical amino acids

The design and engineering of enzymes with an expanded amino acid alphabet is a nascent and rapidly developing area of biocatalysis. Enzymes are exceptionally powerful catalysts capable of promoting chemical transformations with efficiencies and selectivities that are difficult to achieve with small-molecule systems. However, enzymes are typically biosynthesized from the 20 canonical amino acids that contain a limited number of functional groups, restricting the range of catalytic mechanisms that can be installed into designed active sites. The emergence of powerful genetic code expansion methodologies has enabled the site-specific installation of hundreds of structurally and functionally diverse non-canonical amino acids into proteins<sup>80,81</sup>. Careful selection of a suitable non-natural amino acid and its positioning within the target protein scaffold is required to address the application of interest (FIG. 5). For instance, a key active site residue is often replaced with a non-canonical amino acid that is a close structural analogue to modulate catalytic function for mechanistic investigations of natural enzymes<sup>82–86</sup>. Alternatively, to design enzymes with new functions, the selection of amino acid takes inspiration from structural motifs present in small-molecule catalysts with positioning within the protein guided by computation<sup>87,88</sup>.

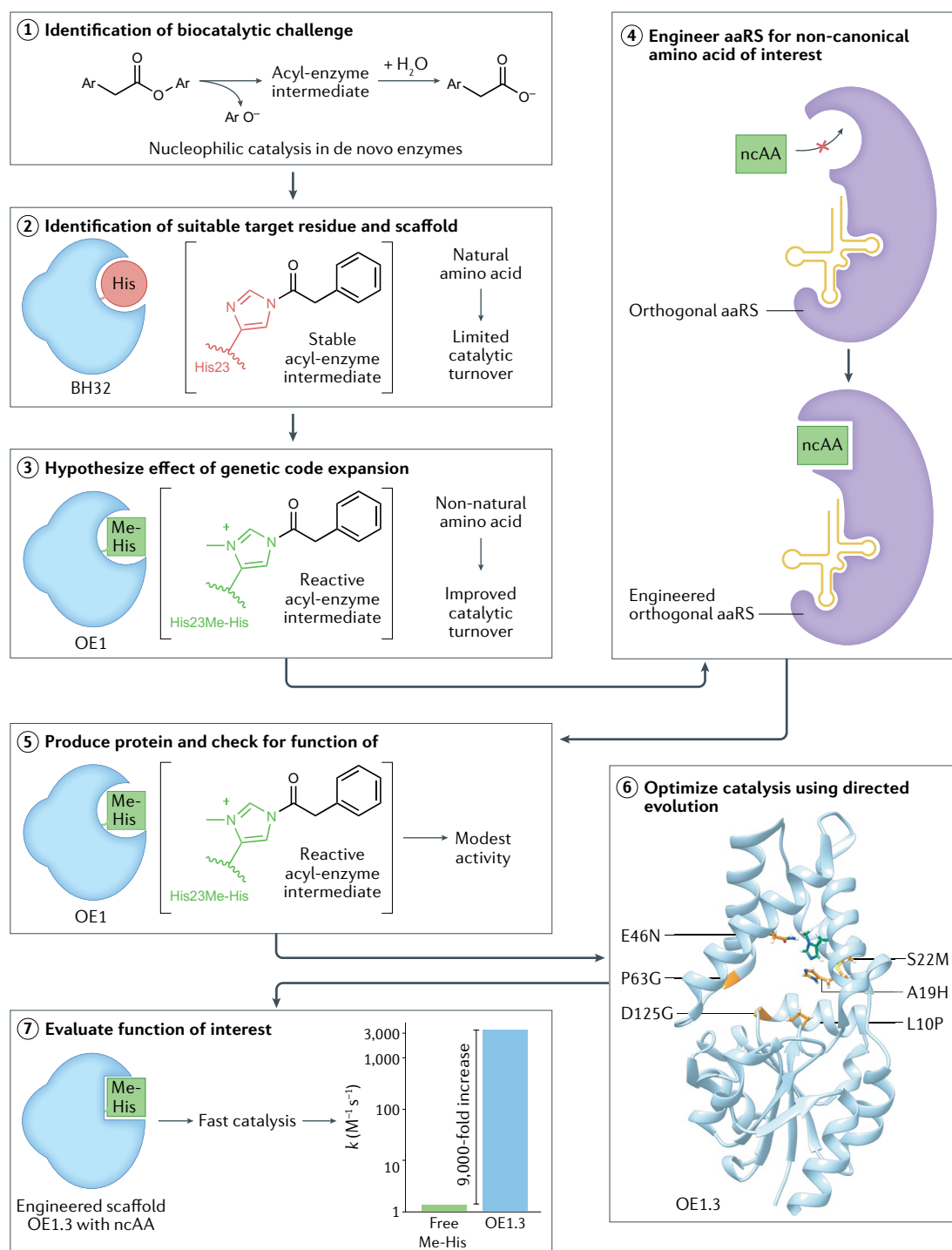
The favoured method for encoding a non-canonical amino acid exploits an engineered aminoacyl-tRNA synthetase/tRNA pair that is orthogonal to the host's translation machinery to direct the incorporation of the non-canonical amino acid by suppression of a nonsense codon, which is most commonly the UAG stop codon. The aminoacyl-tRNA synthetase of the orthogonal translation component pair is typically engineered towards a desired non-canonical amino acid through iterative rounds of positive and negative selections, which link cell viability to aminoacyl-tRNA synthetase activity and selectivity<sup>80,89</sup>. Introducing the non-canonical functionality directly through the cellular translation machinery offers significant advantages over alternative methods of chemically modifying protein structures. For instance, this approach facilitates the homogeneous production of precisely edited proteins, enables the introduction of aminoacyl-tRNA synthetase at diverse sites in any protein scaffold and, perhaps most significantly, allows for rapid optimization of enzyme properties using directed evolution workflows adapted to an expanded genetic code<sup>87,90</sup>.

The availability of an expanded set of amino acid building blocks offers exciting new opportunities for biocatalysis. Genetically encoded non-canonical amino acids have been used to improve both biocatalyst activity and stability<sup>91</sup> as well as provide new tools to understand how enzymes function at the molecular level<sup>82–84</sup>. Key recent examples include the replacement of serine and cysteine catalytic nucleophiles with 2,3-diaminopropionic acid as a means of trapping acyl-enzyme intermediates for structural characterization<sup>85</sup>, and the use of non-canonical axial haem

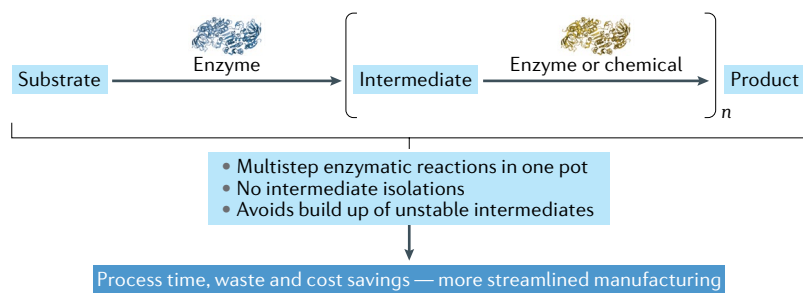
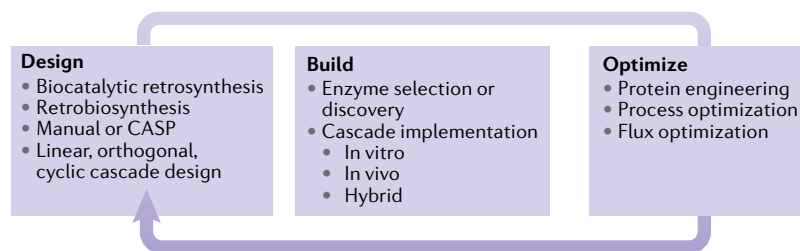
#### Nonsense codon

A codon within the genetic code that does not encode an amino acid but is recognized as a stop codon in transcription and translation of DNA.





**Fig. 5 | Design and engineering of biocatalysts using non-canonical amino acids.** Representative workflow for the use of genetic code expansion to enhance biocatalytic function, using the design and engineering of a de novo enzyme with a non-natural catalytic nucleophile as an example<sup>82</sup>. (Step 1) Biocatalytic challenge to be addressed must be defined. (Step 2) Identify a protein scaffold and a suitable target residue for modification. (Step 3) If the challenge cannot be addressed through standard mutagenesis, identify a functional non-canonical amino acid (ncAA) to achieve the desired properties. For example, replacement of the His23 nucleophile of the de novo protein BH32 with N $\delta$ -methylhistidine (Me-His) should increase the activity of acyl-enzyme intermediates. (Step 4) Are translation components already available to incorporate the ncAA of interest? If not, an aminoacyl-tRNA synthetase (aaRS) can be evolved to accept the target ncAA using directed evolution<sup>75,84</sup>. (Step 5) Use engineered translation components to produce the catalytically modified protein, which is then evaluated for the activity of interest. (Step 6) Catalytic function can be optimized using directed evolution workflows adapted for an expanded genetic code. Ribbon representation of the evolved OE1.3 enzyme structure (light blue) has a Me-His residue shown as a ball and stick representation (green) along with the sites of mutation revealed through directed evolution (orange) (protein image generated using Chimera). (Step 7) The optimized enzyme is characterized. In the case of OE1.3, ester hydrolysis was accelerated by >9,000-fold compared with the free ncAA (Me-His) in solution. Steps 2–5 and 7 adapted from the original figure created with BioRender.com.

**a Biocatalytic cascades****b Cascade development**

**Fig. 6 | Biocatalytic cascades and their development.** **a** | Biocatalytic cascades are reaction sequences in which each chemical step is catalysed by an enzyme. Reaction conditions for enzymatic reactions are often compatible with multistep one-pot protocols. **b** | Design–build–optimize cycle is a useful concept for generating optimized biocatalytic cascades — for each step, several complementary methodologies are now available. CASP, computer-aided synthesis planning.

ligands to unravel the active site features that control the reactivities of high-energy metal-oxo intermediates<sup>86</sup>. The availability of an increased repertoire of covalently embedded functional groups also provides exciting opportunities to design *de novo* enzymes with catalytic mechanisms inspired by small-molecule catalysis. This approach was recently showcased through the design of an artificial hydrolase OE1 (REF.<sup>87</sup>) that employs *N*δ-methyl histidine (Me-His) as a catalytic nucleophile, which operates with a similar mode of action to the widely employed small-molecule catalyst dimethylaminopyridine (DMAP)<sup>92</sup>. Histidine methylation was integral to catalytic function and leads to the generation of reactive acyl-imidazolium intermediates, which are readily hydrolysed to regenerate the catalytic nucleophile (FIG. 5). By contrast, the catalytic function of *de novo* hydrolases employing canonical histidine, serine or cysteine nucleophiles was compromised by the formation of unreactive acyl-enzyme intermediates<sup>93–96</sup>. The modest initial hydrolysis activity of OE1 was subsequently enhanced via iterative rounds of directed evolution giving rise to variant OE1.3 containing six active site mutations<sup>87</sup>. OE1.3 accelerates ester hydrolysis beyond 9,000-fold and 2,800-fold as compared with free Me-His and DMAP, respectively. Further rounds of evolution lead to the enantioselective hydrolase OE1.4. This study showcases how the interplay of genetic code expansion, computational design and directed evolution can provide a truly versatile platform for building *de novo* biocatalysts with new and improved catalytic functions.

**Regioselectivity**

The property that favours bond formation or breaking at a particular atom over all other possible atoms in a molecule.

**Cascade development**

Combining multiple enzyme-catalysed steps in the same pot is a very important research area. Biocatalysis is particularly well suited to these cascade processes as enzymes possess inherent chemoselectivity, regioselectivity and stereoselectivity and operate in a common aqueous media. Akin to natural biosynthetic pathways, fully *de novo* non-natural biocatalytic cascades can be designed and developed for the synthesis of complex targets. It should be pointed out that biocatalytic cascades have only become more commonly used because of the advances in biocatalyst design and build discussed elsewhere in this Primer.

Biocatalytic cascades<sup>97,98</sup> typically feature two or more steps (functional group interconversions or bond forming) with at least one enzymatic transformation and without intermediate isolations (FIG. 6a). The definition of ‘cascade’ is generally broadly applied within the biocatalysis community to describe not only concurrent, multienzyme processes in one pot but also reactions in which components are added sequentially or process steps are telescoped despite attempts to impart order on the nomenclature<sup>99–102</sup>.

The development of novel biocatalytic cascades can be broadly described by a design–build–optimize cycle (FIG. 6b) until a final process is achieved<sup>3,103</sup>. Initially, retrosynthetic analysis is performed using the principles of biocatalytic retrosynthesis<sup>104–107</sup> and/or retrobiosynthesis<sup>108</sup> to make key bond disconnections and plan the forward route. This can be performed manually or complemented by more recent computer-aided synthesis planning tools that are becoming the focus of increased interest<sup>8,109,110</sup>. Additionally, selecting a cascade design that will enable the planned synthesis is required, which can range from simple linear to orthogonal or cyclic processes<sup>99</sup>. Any cofactor requirements or potential compound incompatibilities should be considered at this stage.

Once a process design is in place, enzymes need to be identified to fulfil each cascade step. Enzymes can be identified from the literature, from screening of enzyme libraries or from enzyme discovery efforts. When it comes to building the cascade, there is a choice of operating the enzymatic steps with purified or crude cell-free extracts (*in vitro*), with viable whole cells (*in vivo*) or with a combination of the two (hybrid)<sup>103</sup>. A multitude of factors will help determine which system is best to use, such as enzyme availability, cofactor recycling requirements and reactor/facility infrastructure. Often, each step in the cascade is validated individually before any single-pot combinations are tested.

Finally, optimization of the process can help maximize throughput and product titre. Several rounds of protein engineering are typically required, especially for industrial application, to improve enzyme activity and stability to overcome any bottlenecks in the pathway and maximize pathway flux. General process engineering optimizations also complement cascade development; for example, enzyme immobilization strategies to simplify the workup and/or improve a biocatalyst’s lifetime, recoverability and reuse<sup>111,112</sup>.

Further understanding of the full process can then influence subsequent design–build–optimize cycles in

Enzyme operational stability  
Retention of enzyme activity  
when the enzyme is in use.

an iterative fashion to streamline the entire synthetic route to the desired compound.

## Results

### What makes a good industrial biocatalyst?

Before scientists embark on the challenge of discovering a good (or ideally excellent) industrial biocatalyst, they need to define which properties the biocatalyst must have for efficiently performing a commercially interesting target reaction under select industrial conditions. Here, we describe the beneficial characteristics that are usually found in industrial biocatalysts, metrics to assess their performance in industrial processes and a few exciting examples. Other illustrative examples can be found in excellent recent articles<sup>4,113–115</sup>. New biocatalytic processes aim to generate new molecules of considerable commercial interest. They may also be designed to replace or complement existing non-optimal chemical or biocatalytic syntheses in industry. In either case, the viability and possible bottlenecks of a biocatalytic process can be assessed using both economic and green chemistry process performance metrics<sup>116–119</sup> (TABLE 1). Both high substrate concentration and conversion are desired in industrial reactions to achieve high product concentrations and so reduce the product recovery cost. Reactions resulting in low product concentrations may require additional concentration steps or large volumes of extraction solvent, which will increase costs associated with a rise in energy consumption and/or waste production. Thus, an ideal biocatalyst's activity should not be inhibited by high substrate concentrations ( $>50\text{ g l}^{-1}$ ) or the amount of co-solvent required for substrate solubilization. It is worth mentioning that substrate loadings as high as  $>1\text{ kg l}^{-1}$  for aldolase dehydratase, which catalyses the synthesis of linear aliphatic nitriles, have been reported<sup>120</sup>. Nevertheless, frequently observed detrimental effects at high substrate concentration or by organic

solvents on biocatalysts can be alleviated using fed-batch strategies<sup>121</sup>. In examples where inhibition of the enzyme by-product, unfavourable thermodynamic reaction equilibria or product side reactions are problematic, in situ product removal can be applied<sup>122</sup>. To remove a product resulting from an ongoing enzyme-catalysed reaction, various techniques can be used such as in situ product crystallization, adsorption, distillation and extraction<sup>122,123</sup>. For example, in situ product crystallization can be achieved by forming a product salt via inclusion of an appropriate counter-ion in the reaction media. Similarly, another option is to perform the bioconversion in the presence of a resin that selectively adsorbs the product from the solution.

High stability under industrial process conditions is an essential property of a good biocatalyst. Numerous robust enzymes of industrial interest have been discovered or redesigned over the past decade by enzyme engineering, computational methods, genome mining, ancestral sequence reconstruction or combinations thereof. A recent example using FRESCO generated an alcohol dehydrogenase mutant with a melting temperature — the temperature at which half of the protein is unfolded at equilibrium — of  $94^\circ\text{C}$  (close to water's boiling point) and this has previously been applied successfully to other enzyme classes<sup>124</sup>. Sequence reconstruction of a robust ancestor has been achieved for an increasing number of biocatalysts including cytochrome P450 monooxygenases<sup>125</sup>, carboxylic acid reductases<sup>126</sup>, flavin-containing monooxygenases<sup>127</sup> and laccases<sup>128</sup>, made available in a recently created database of resurrected proteins with 211 members (Revenant)<sup>129</sup>. Enzyme immobilization, which facilitates repeated enzyme reuse, has also been used to enhance enzyme operational stability in industrial processes<sup>130–132</sup>. As there is great interest in the utility of enzyme immobilization, especially in continuous flow systems<sup>133</sup>, tolerance to immobilization without significant loss of activity or selectivity is an appealing property for a biocatalyst<sup>134</sup>.

Biocatalytic processes outcompete their chemical counterparts regarding sustainability, as illustrated when comparing the chemical and biocatalytic synthetic routes for pregabalin, atorvastatin intermediate, sitagliptin and ambrox<sup>135</sup>. In contrast to chemical catalysts, biocatalysts are derived from renewable resources, are biodegradable, act in aqueous solvent under mild reaction conditions and generate low amounts of waste by-products. Furthermore, biocatalytic synthetic routes obviate the need for hazardous chemicals, high energy usage and additional reagents for functional group activation, protection or deprotection steps.

Biocatalytic processes requiring whole-cell fermentations (for either enzyme production or substrate conversion) generate waste biomass, which can be reused as a source of energy or animal feed. To reduce water usage and carbon feedstocks required for cell growth, biotransformations with isolated enzymes or cell lysates can be performed instead of whole-cell fermentations at increased concentrations. A reduction in biocatalyst loading, without reducing productivity as measured by yield and speed, can be accomplished by using engineered biocatalysts that offer improved properties such

Table 1 | Metrics used to evaluate efficacy of biocatalytic processes

Metric	Definition
<b>Economic metrics</b>	
Productivity ( $\text{g l}^{-1}\text{ h}^{-1}$ )	Amount of product made as a function of the reactor volume and time
Yield of product on substrate ( $\text{g g}^{-1}$ )	Amount of product (grams) made as a function of the amount of substrate (grams) added in the reaction
Turnover ( $\text{mol mol}^{-1}$ )	Amount of product made as a function of the amount of biocatalyst added in the reaction
<b>Green chemistry metrics</b>	
Atom economy ( $\text{g mol}^{-1}$ per $\text{g mol}^{-1}$ )	Molecular weight of the desired product as a function of the sum of the molecular weight of all products
E factor ( $\text{g g}^{-1}$ )	Amount of waste made as a function of the amount of product made
C factor ( $\text{g g}^{-1}$ )	Amount of $\text{CO}_2$ equivalents as a function of the amount of product made
Process mass intensity ( $\text{g g}^{-1}$ )	Amount of raw material input as a function of the amount of desired product made

Information based on REF.<sup>116</sup>.

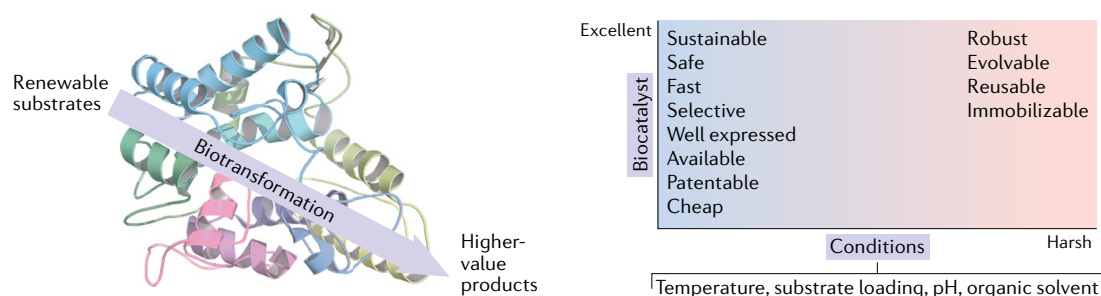


Fig. 7 | **Beneficial properties of an excellent biocatalyst under industrial process conditions.** Ideal industrial biocatalysts are able to efficiently convert renewable raw materials into higher-value molecules. Their most desired characteristics are indicated in the plot. Under highly demanding industrial conditions, biocatalyst properties (top right, orange background) such as robustness or tolerance to mutations enhancing stability often become essential. Other beneficial biocatalyst properties are desired under either mild (blue background) or more demanding operational conditions.

as higher turnover rates and/or stability for reuse. Energy consumption due to biocatalyst recovery from reaction solutions can be minimized by enzyme immobilization. Importantly, inexpensive renewable carriers for enzyme immobilization, such as rice husk, are being developed to replace organic fossil-based carriers<sup>136</sup>. However, a significant expansion of enzyme-based technologies in the production of bulk chemicals (high volume, low priced) must be achieved to increase the impact of biocatalysis on sustainability<sup>137</sup>. So far, biocatalysts are more frequently used to synthesize high-price low-volume products such as pharmaceuticals.

Various companies (for example, Merck, Pfizer, GlaxoSmithKline and AstraZeneca) have become active in the development of new biocatalytic processes and often collaborate with academic groups to accelerate progress in this research area. Examples of some of the enzymatic processes developed by industry with biocatalysts including KREDs, transaminases, hydroxylases and IREDs are described in recent review articles<sup>138,139</sup>. When selecting a biocatalyst for process development, it is often desirable to select enzymes that will enable freedom to operate to avoid infringing intellectual property rights or to access desired patented biocatalysts during the early stages of process design. To this end, industries and universities often provide experts in the complex and rapidly evolving field of intellectual property to guide research scientists.

A good industrial biocatalyst should combine numerous beneficial properties to deliver higher-value molecules under demanding industrial conditions while achieving satisfactory economic and green metrics for various applications (FIG. 7). A few of the most desired characteristics of efficient industrial biocatalysts have been highlighted above, which include high activity, stability, ease of immobilization, environmental sustainability and accessibility. The importance of other relevant properties of a good biocatalyst, such as substrate selectivity, evolvability and affordability, will be illustrated through various examples in the following sections.

### Applications

An ideal catalyst converts renewable, cheap and readily available raw materials such as plant-derived feedstocks, generates few to no undesired by-products, is safe and

exhibits a reduced environmental footprint (low energy consumption and waste). These characteristics are not often observed for industrial chemical catalysts. Also, biocatalysts usually act under mild reaction conditions and can be engineered towards the desired substrate scope. Thus, biocatalysis paves the way for a bio-based economy, less reliant on fossil fuels<sup>117</sup>. Here, we highlight the utility of biocatalysis in various applications, first according to different reaction metrics or enzyme properties that are of importance in biocatalysis followed by an overview of enzyme cascade development.

### Activity and productivity

Biocatalysts should exhibit high activity under the desired industrial conditions to achieve a high reaction productivity. Chemically heterogeneous catalysts are challenging rivals for biocatalysts in terms of productivity, often reaching production rates of 1–10 and 0.001–0.3 kg l<sup>-1</sup> h<sup>-1</sup>, respectively<sup>140</sup>. High productivities of 50–100 g l<sup>-1</sup> h<sup>-1</sup> have been achieved using free-resting *Rhodococcus* cells containing nitrile hydratase for the synthesis of acrylamide from acrylonitrile, considered to be one of the most successful industrial biocatalytic processes<sup>140,141</sup>. Acrylamide is used to produce polyacrylamide, which is used in water treatment, oil exploitation and the textile industry sector, as well as many others. The potential of nitrile hydratase as an industrial biocatalyst for the hydration of nitriles to form higher-value amides was demonstrated in the 1980s<sup>1</sup>. The vast market for acrylamide and the lack of an efficient chemical process for its production have propelled the improvement of the biocatalytic process over the past few years. The selection and optimization of a robust microbial host for nitrile hydratase was instrumental in preventing enzyme inactivation, owing to the high acrylamide concentrations required in the industrial process (300–500 g l<sup>-1</sup>) and the underlying exothermic nature of the reaction<sup>141</sup>. A selective robust transaminase was obtained, by combining rational mutagenesis, directed evolution and a substrate walking approach, for the large-scale manufacturing of the antidiabetic drug sitagliptin under demanding industrial conditions (200 g l<sup>-1</sup> substrate loading, 50% DMSO and 45 °C)<sup>142</sup>. This is an impressive example of an excellent industrial biocatalytic approach that outcompeted the previously used rhodium-catalysed

### Evolvability

Capacity of an enzyme to acquire beneficial properties or functions through genetic modification.



### Design of experiments

A statistical approach to analyse the influence of various factors in a system to predict the optimal operating conditions.

sitagliptin synthesis in terms of selectivity, productivity, sustainability and cost.

A relatively high productivity ( $13 \text{ g l}^{-1} \text{ h}^{-1}$ ) was recently achieved for IREDs by testing a commercially available IRED collection and various reaction conditions at the pilot plant scale, which was facilitated by a design of experiments strategy<sup>121</sup>. IREDs are of great interest for the industrial synthesis of cyclic and acyclic amines via the reduction of C=N bonds. This study identified reaction bottlenecks (for example, enzyme stability) and exposed possible strategies to overcome them (for example, using a fed-batch process) for a model reaction. Importantly, the first industrial synthesis catalysed by an IRED (on a 20-l scale) was recently reported<sup>47</sup>, highlighting an excellent industrial biocatalyst after three rounds of directed evolution, which outcompeted the corresponding chemical process with respect to green metrics such as lower catalyst requirement. This engineered IRED is used for the industrial synthesis of the LSD1 inhibitor GSK2879552. In contrast to the IRED used as starting point in this study, the engineered IRED is an excellent biocatalyst due to its increased stability under the required reaction conditions (moderately acidic pH and  $20 \text{ g l}^{-1}$  substrate concentration) showing a 38,000-fold improvement in turnover. In this case, the selectivity — another requirement for a good biocatalyst — needed no further improvement. The preparation of the fragrance ingredient (–)-ambrox using an engineered squalene hopene cyclase is another example of a successful industrial biocatalytic process, which achieved relatively high productivity ( $12 \text{ g l}^{-1} \text{ h}^{-1}$ ) for catalysing the cyclization of (*E,E*)-homofarnesol to yield (–)-ambrox<sup>143</sup>. The enzyme variant used in this study, which exhibited a 10-fold increase in productivity over the wild type, was discovered by random mutagenesis. This cyclase whole-cell biotransformation in *E. coli* was carried out under conditions that were optimized using a design of experiments strategy, in which the optimized parameters included the cell, sodium dodecyl sulfate (SDS) and (*E,E*)-homofarnesol concentrations, temperature and pH. SDS was required in this process to ensure substrate solubilization and access to the enzyme through the cell membrane.

### Selectivity and substrate scope

Enzymes with excellent regioselectivity, chemoselectivity and/or stereoselectivity and the desired substrate scope for industrial applications can be obtained by either mining the enormous diversity evolved by nature or performing protein engineering campaigns in the laboratory. Studies that have uncovered the extraordinary diversity of enzymes involved in natural product biosynthetic pathways have provided promising industrial biocatalysts with complementary selectivity as well as substrate scope. For example, the recent comparison of three similar FAD-dependent monooxygenases, which catalyse the oxidative dearomatization of phenol and resorcinol in different biosynthetic pathways, has revealed their complementary site selectivities and stereoselectivities by testing a diverse panel of unnatural substrates<sup>144</sup>. This approach enabled the identification of an optimal biocatalyst for specific asymmetric

transformations of phenols into *ortho*-quinols, a chemical reaction of great value in the synthesis of various bioactive natural products<sup>144</sup>. In another example that highlights the importance of enzyme discovery and characterization, the substrate scope of 87 putative flavin-dependent halogenases was determined using a high-throughput mass spectrometry-based screen<sup>22</sup>. Various halogenases discovered in this study exhibited complementary regioselectivity on relatively complex substrates. Thus, this enzyme library is attractive for late-stage C–H functionalization of drug leads, leading to diverse drug candidates from common intermediates. Furthermore, this study enabled the discovery of new halogenases for biotechnology applications, which exhibited beneficial properties such as regioselectivity, substrate scope and stability that were engineered in other previously discovered halogenases<sup>22</sup>.

An increasing number of studies demonstrate that required selectivities can be readily engineered into different enzyme classes<sup>145</sup>. A recent example is the synthesis of a Janus kinase (JAK) inhibitor, which involved engineering IRED variants with markedly improved selectivity and activity compared with the wild type<sup>146</sup>. Synthesis of enantiomerically pure compounds is a key driver for the implementation of enzymes in the pharmaceutical industry<sup>3</sup>. Enantioselective enzymes are also used industrially for the production of target molecules required in food supplements, flavourings, fragrances and agrochemicals<sup>147</sup>. To this end, a wild-type cytochrome P450 monooxygenase catalyses the enantioselective and regioselective C5 hydroxylation of decanoic acid to form (*S*)-5-hydroxydecanoic acid, which is subsequently converted by chemical lactonization into the high-value fragrance compound (*S*)- $\delta$ -decalactone<sup>148</sup>. In the food industry, small-scale reactions using an engineered ethylenediamine-*N,N'*-disuccinic acid lyase have demonstrated its utility for the enantioselective synthesis of chiral synthons for artificial dipeptide sweeteners<sup>149</sup>. The lyase used as a starting scaffold exhibited excellent enantioselectivity for the target substrate but had low activity, which was increased 1,140-fold by rational protein engineering.

### Enzyme cascades

From an industrial perspective, biocatalytic cascade processes are especially attractive as they eliminate the need for intermediate isolation steps, reducing waste, saving time and costs as well as streamlining the overall synthesis<sup>150</sup>. Some intermediates can be unstable to isolation or have inhibitory effects on the enzymes present in the system, and therefore the use of a cascade process can be beneficial to overcome these challenges and avoid the build up of problematic intermediates.

Several recent reviews have been published on enzymatic cascades that reveal the potential and scope of these processes<sup>99,101,102,151,152</sup>. Some examples<sup>97</sup> of industrially applied systems are highlighted here (FIG. 8). Evonik Degussa GmbH described a whole-cell cascade to produce diamines — which are valuable building blocks in the polymer industry — from renewably sourced dicarboxylic acids (FIG. 8a). The patented process<sup>153</sup> details the co-expression of a carboxylic acid reductase and



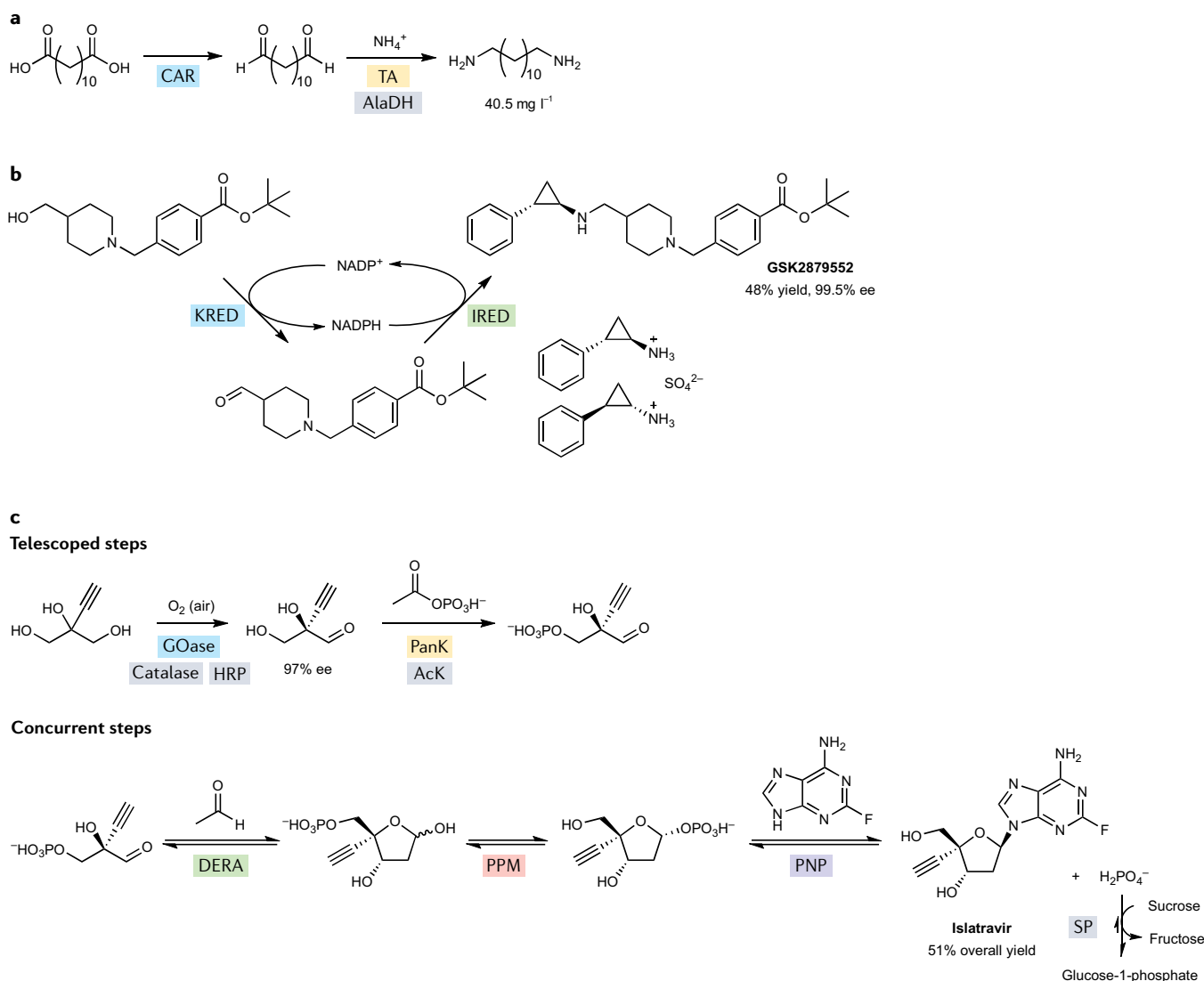


Fig. 8 | **Industrial examples of biocatalytic cascades.** **a** | Evonik Degussa GmbH described a whole-cell cascade to produce diamines from renewably sourced dicarboxylic acids<sup>153</sup>. **b** | Hydrogen-borrowing, redox-neutral cascade developed by GSK for production of GSK2879552 (REF.<sup>47</sup>). **c** | Merck & Co.<sup>7</sup> developed a total enzymatic synthesis of the HIV drug isotravir built on five key enzymatic steps. AcK, acetate kinase; AlaDH, alanine dehydrogenase; CAR, carboxylic acid reductase; DERA, deoxyribose phosphate aldolase; ee, enantiomeric excess; GOase, galactose oxidase; HRP, horseradish peroxidase; IRED, imine reductase; KRED, ketoreductase; PanK, pantothenate kinase; PPM, phosphopentamutase; PNP, purine nucleoside phosphorylase; SP, sucrose phosphorylase; TA, transaminase.

a transaminase to enable the desired cascade. An alanine dehydrogenase was also incorporated to provide a source of L-alanine, required for the transaminase step, from ammonia as an input nitrogen source. Additional process considerations for the in vivo implementation of the cascade included co-expression of fatty acid transporters to improve substrate uptake or the incorporation of an initial esterase step enabling the use of esters as starting materials.

A hydrogen-borrowing, redox-neutral cascade was developed by GSK for the production of GSK2879552 (REF.<sup>47</sup>) (FIG. 8b). A KRED IRED system was evaluated to take the desired alcohol to the chiral amine via an aldehyde, with internal cofactor recycling between the two enzymes. The main synthetic focus of the work was the engineering of the IRED step, involving reductive

amination and concurrent resolution of the racemic amine substrate. The cascade synthesis enabled generation of the desired product in 48% yield with high enantiopurity (99.5% enantiomeric excess). Although the IRED step can operate as a stand-alone process and achieve higher yields, the proof of concept for the cascade was established. Process development and a more active KRED were highlighted as areas of potential focus to further improve the cascade and realize its potential for manufacturing.

Recently, Merck & Co.<sup>7</sup> developed a total enzymatic synthesis of the HIV drug isotravir built on five key enzymatic steps (FIG. 8c). The selected enzymes were subjected to multiple rounds of protein engineering to achieve either the desired activity, stability or selectivity for operation of the cascade. A single aqueous reaction

**BLAST**

(Basic Local Alignment Search Tool). A tool that compares nucleotide or protein sequences of interest (most commonly to sequences within a database), and finds regions of statistically significant similarity.

stream was employed throughout the entire process, in which the galactose oxidase (GOase) and pantothenate kinase (PanK) steps operated sequentially to avoid cross-reactivity between substrates. The final deoxyribose phosphate aldolase (DERA), phosphopentamutase (PPM) and purine nucleoside phosphorylase (PNP) steps were then run concurrently, and the equilibrium of these steps was pulled through to product formation by an orthogonal sucrose phosphorylase (SP) step that removed phosphate from the reaction mixture. The cascade synthesis of islatravir (and, more recently, of molnupiravir)<sup>154</sup> replaced alternative chemical routes to this drug that required more than double the step count with protecting group manipulations, thereby vastly improving the efficiency of synthesis.

**Reproducibility and data deposition****Databases for biocatalysis**

Over the past decade, the cost of DNA sequencing and synthesis has fallen rapidly; a trend commonly referred to as the Carlson curve<sup>155</sup>.

This associated abundance of protein sequence data provides a rich seam for mining for new biocatalysts. The National Center for Biotechnology Information (NCBI) maintains databases of both DNA and protein sequences, regularly updated with new sequencing data, and with the option to search for sequences of interest using tools such as BLAST (Basic Local Alignment Search Tool)<sup>156</sup>. Other databases, such as UniProt, InterPro or Pfam, offer further analysis of protein sequences, structures or families.

As the amount of data collected for an increasing number of enzymes and enzymatic transformations rises, it becomes prohibitive for interested researchers to efficiently scour the literature in search of ideal/appropriate candidates to analyse. Catalyst and enzyme selection, for use in organic chemistry syntheses or synthetic biology pathways, respectively, already benefit from numerous well-developed databases. Reaxys<sup>157</sup> and SciFinder<sup>158</sup> contain a plethora of searchable

information related to reaction conditions, choice of catalyst, substrate scope, percentage conversions and analytical information, among others, for use when designing a synthetic chemistry route towards a target molecule, whereas BRENDA<sup>159</sup> and KEGG<sup>160</sup> hold data on the natural substrate specificity, and sequence information, of biosynthetic enzymes to be used in a synthetic biology pipeline. A comparable repository, comprising information collected for synthetic enzyme reactions in biocatalysis, would be of great use for the biocatalysis community.

Despite the fact that several databases for the biocatalysis community have been developed, none of them contains information related to the whole biocatalytic toolbox, and the majority do not provide such critical information as the substrate scope of specific enzymes, successful reaction conditions or reaction yields (TABLE 2). In general, the majority of the resources listed in TABLE 2 rely on data extracted from pre-existing databases such as BRENDA and PDB (Protein Data Bank) and, as such, are restricted to solely utilizing the sequence and/or structural information contained within them. Additionally, the curation and maintenance of substantial databases is often laborious and challenging, and so most biocatalyst databases focus on a specific reaction type or enzyme type of interest, rather than compiling data on the field as a whole. One of the few examples of a database recording information related to substrates, products and reaction outcomes in a biocatalysis context has been developed for the prenyltransferase enzyme class (PrenDB)<sup>161</sup>. PrenDB aims to collect data in the literature concerned with prenyltransferase enzymes and use them in various algorithms to achieve wider application of this family of synthetically useful enzymes. The compilation of a biocatalysis database, similar in scope to PrenDB but covering a broad spectrum of the different enzyme classes available in the biocatalytic toolbox, would unquestionably enhance the development of new enzymatic (cascade) reactions.

An ideal database dedicated to biocatalytic transformations would capture both successful and unsuccessful transformations on an enzyme by enzyme basis and would broadly collect both enzyme activity data and enzyme sequence data. For example, data regarding the substrate scope, reaction temperature and length, buffer choice and pH, cofactor use, co-solvent use, substrate concentrations, reaction outcomes including percentage conversions and selectivities would all need to be collected to maximize the applicability of such a database. Enzyme homologue information, such as the amino acid sequence, structural information, mutant information and accession codes, would also need to be obtained. Additionally, integration with existing databases, such as those outlined above for chemistry and synthetic biology applications, would allow for extremely powerful synthesis planning towards target molecules. A fully functioning and searchable biocatalyst database could be used to augment tools designed to automate synthesis planning and would, ultimately, benefit researchers from both the chemistry and biocatalysis fields. In related fields such as natural product discovery, crowdsourcing has been successfully utilized for the construction of similar

Table 2 | **Databases containing information specifically on enzymes for biocatalysis**

Database	Description	Substrate specificity
PrenDB <sup>161</sup> ( <a href="http://prendb.pharmazie.uni-marburg.de/prendb/home/">http://prendb.pharmazie.uni-marburg.de/prendb/home/</a> )	Contains substrate, product and reaction information for prenyltransferase enzymes	Yes
Lipase database <sup>193</sup> ( <a href="http://www.led.uni-stuttgart.de/">http://www.led.uni-stuttgart.de/</a> )	Contains sequence and structural information for lipase enzymes	No
BioCatNet databases <sup>11</sup> ( <a href="https://biocatnet.de/">https://biocatnet.de/</a> )	Collection of databases containing sequence and structural information for transaminase, P450 and imine reductase (IRED) enzymes, among other enzyme classes	No
CAZy <sup>194</sup> ( <a href="http://www.cazy.org/">http://www.cazy.org/</a> )	Contains sequence and structural information for carbohydrate-active enzymes	No
RedoxiBase <sup>195</sup> ( <a href="http://peroxidase.toulouse.inra.fr/">http://peroxidase.toulouse.inra.fr/</a> )	Contains sequence and structural information for peroxidase and other oxidoreductase enzymes	No

databases<sup>162</sup>. Indeed, a platform for curation of biocatalysis data has recently been made available to the community with this in mind, as part of the computer-aided synthesis planning tool RetroBioCat<sup>8</sup>.

### **Reproducibility issues in the field**

A successful biocatalyst database requires a system that captures all useful information on biocatalyst performance reported in the literature. However, the diverse scientific communities that work with and characterize biocatalysts have varying standards when it comes to recording reaction parameters and outcomes, with some favouring kinetic data and others preferring to record percentage conversions and overall yields, for example. These different approaches have resulted in a wealth of information for many different enzyme classes and homologues that may not be directly comparable with one another, and so it becomes necessary to standardize the data collected in order to obtain a better overall picture of where select developments stand. One such way of standardizing data reported in the literature would be to categorize reactions qualitatively for enzyme activities with respect to a given substrate (for example, high, medium, low, none). Different data sources, such as percentage conversions and specific activities, could be categorized in this way and then compared against each other.

Alternatively, biocatalytic experiments could be standardized in the laboratory prior to data deposition. For numerous years, the STRENDa (Standards for Reporting Enzyme Data) commission has sought to provide guidelines on the experimental detail required when reporting enzyme activities and kinetics<sup>163</sup>. Recently, these guidelines have been incorporated into an online storage and validation tool, where enzyme data can be deposited and checked for compliance with the STRENDa guidelines<sup>164</sup>. This serves as a useful blueprint for reporting biotransformations in biocatalysis, but likely must be extended to include the additional datatypes often reported in biocatalysis papers, for example percentage conversion.

Recently, numerous start-up companies have emerged across biology and chemistry to develop smart-laboratory infrastructures, aiming to make research more reproducible by capturing data on all of the possible variables in an experiment<sup>165</sup>. Others offer platforms to structure the collected data in a process, allowing machine learning to pull insight out of the vast data sets that smart-laboratories might produce<sup>166</sup>. Experimentally, this can allow trends to be observed that might otherwise be missed — for example, a new batch of a reagent causing a drop in yields, or a shift in pH causing improved enantioselectivity. The digitization of experimental procedures and data collection should greatly improve the reproducibility of experiments across biology and chemistry. In particular, this may allow methods sections in journals to offer links to a more atomized record of the experimental procedures carried out. However, uptake by academia may be slow in comparison with industry laboratories, where electronic laboratory notebooks are more commonly employed.

## **Limitations and optimizations**

### **Cost and accessibility**

Biocatalyst cost usually has an influence on the viability of an industrial process, but especially in the synthesis of low-priced products. Currently, a wide variety of affordable enzyme collections (kits) are accessible from various vendors (for example, Prozomix, Almac, Codexis and Gecco). Enzyme discovery and production in-house is the alternative approach. Advantages and limitations of these options, ‘the buy or build operating models’, have been previously discussed<sup>167</sup>. The choice of biocatalyst format (for example, purified, whole-cell or crude preparations) varies, depending on the particular application and enzyme class. Obviously, well-expressed enzymes are highly desired to reduce costs and effort. Access to an increasingly diverse platform of molecular biology tools allows the tailoring of enzyme performance to meet demanding industrial requirements and to efficiently convert non-natural substrates. Generation of improved biocatalysts by enzyme engineering is possible simply because enzymes are able to tolerate *in vitro* mutation. Thus, evolvability is another highly desirable property of a good biocatalyst. Engineering one enzyme may take only a few months, but building complex cellular metabolic networks may take years and demand considerable economic investment<sup>168</sup>. These timescales are not fast enough to meet ‘the need for speed’ in industry<sup>169</sup>. A recent example of a three-step route including two enzymatic steps, which was developed in just 6 months, is the synthesis of the COVID-19 direct-acting antiviral molnupiravir<sup>154</sup>. Development of highly efficient biocatalysts by either rational or evolution techniques will be accelerated in the near future by expanding both the use of machine learning<sup>170</sup> and ultra-high-throughput screening<sup>171</sup> technologies for protein engineering.

Machine-learning algorithms use the sequence-function data resulting from experimental work to predict which new enzyme mutants may exhibit the desired property. Thus, DNA sequences of both improved and unimproved variants are valuable in generating initial data sets. Importantly, machine-learning methods allow a reduction in the number of mutants that have to be produced and tested in the laboratory to discover a significant fraction of improved enzymes, and are particularly interesting in cases that require expensive or labour-intensive screening methods<sup>170</sup>. The additional costs of implementing machine-learning algorithms in a traditional protein engineering laboratory include computation and DNA sequencing, costs that are decreasing and, thus, are affordable for numerous research groups in both academia and industry. To explore a vast protein sequence space (library sizes >10<sup>6</sup> variants), ultra-high-throughput screening technologies have been developed. Many academic or industrial researchers have access to flow cytometers to perform fluorescence-based screenings of up to 10<sup>8</sup> enzyme variants per day<sup>172</sup>. Complementary or improved technologies are rapidly emerging in this field. Miniaturization of the reaction volume is generally pursued because it increases the speed of screening and reduces associated costs and waste. Label-free detection methods are also highly desired, allowing for screening without

a reporter molecule. A recent example meeting both objectives allowed the analysis of around 15,000 samples in 6 h using droplet microfluidics (nanolitre scale) coupled to electrospray ionization mass spectrometry for detection<sup>60</sup>. Development and wide access to novel technologies for biocatalysis has been propelled by recent investments from, for example, the European Commission and the UK Biotechnology and Biological Sciences Research Council (BBSRC) to facilitate collaborations between industry and academia. The recent establishment of a Global Biofoundry Alliance represents another example<sup>173</sup>. Biofoundries are facilities to automate the design–build–test iteration cycle for engineering biology, which allows the fast delivery of genetically reprogrammed organisms for biotechnology<sup>77</sup>. Access to biocatalysts is also facilitated by other strategies such as Science Exchange (an online marketplace of research services) and collaborations established between the Centre of Excellence for Biocatalysis, Biotransformations and Biocatalytic Manufacture (CoEBio3) and various companies.

### Expanding the range of biocatalysis

Biocatalytic transformations, particularly those routinely applied in industry, often effect functional group interconversions with high conversion and selectivity. However, one of the biggest gaps is broader enzyme platforms that perform C–C bond formation. Despite a plethora of enzymes used in nature for C–C bond formation in primary and secondary metabolism, they are often challenging to repurpose for non-natural substrates. Only a handful of enzymatic C–C bond-forming enzymes have been utilized for industrial applications, mainly limited to aldol reactions, acyloin condensations or cyanohydrin formation catalysed by lyases<sup>174,175</sup>. A recent review highlights progress made in this space to diversify the toolbox of enzymes and the C–C bond-forming reactions they catalyse<sup>176</sup>. Another industrial gap is scalable and robust oxidative enzymes. Despite the potential to catalyse remote and unactivated C–H oxidations, which are chemically challenging, enzymes such as cytochrome P450s and other oxygenases are problematic to scale up due to low activity, instability and promiscuity, resulting in a mixture of products<sup>177</sup>. However, these features are well-suited for small-scale, late-stage diversification of biologically active compounds in which the enzyme promiscuity is advantageous to generate new libraries of compounds for evaluation<sup>178,179</sup>. The synthetic utility of the transformation afforded by these enzymes encourages continued efforts to find solutions and realize the potential of these biocatalysts for large-scale manufacture<sup>180</sup>.

### Speeding up synthesis

In the pharmaceutical industry, the acceleration of the drug development process is crucial to be able to deliver new medicines to patients as quickly as possible as well as maximize patent lifetimes for approved drugs. As such, time pressures for synthetic development are increasingly tight, which is driving advancements in the speed of rounds of protein engineering and the establishment of biocatalysis earlier in synthetic route planning<sup>3,169</sup>.

These advances include improvements in DNA library syntheses, smart library design and high-throughput screening. On the horizon are technologies such as cell-free expression, which enables skipping the need for growing and harvesting cells that contain enzyme mutants, to further reduce cycle times<sup>181,182</sup>. Although the acceleration of development timelines is often associated with the pharmaceutical industry, these improvements are also beneficial to the wider chemical industry, making development more efficient and cost-effective<sup>140</sup>.

### Outlook

Advances in protein engineering, genomic database mining and computational methods have enabled a step change in biocatalysis over the past 20 years, and have led to its increasing application in the chemical and pharmaceutical industries as highlighted in this Primer. Adoption of biocatalysis is also driven by reduced cost, the need to develop environmentally friendly processes and use of renewable resources<sup>183</sup>.

The number of chemical reactions realized as amenable to biocatalysis has dramatically increased, as new enzyme classes become accessible and non-natural biocatalytic reactions are being developed<sup>184</sup>. However, the range of reactions compared with those used in organic synthesis is still small and there are some obvious gaps in the repertoire of biocatalytic reactions that are currently being identified, including halogenation, amide-bond formation, C–C bond formation and cleavage, ether formation, carbonylation, C=C bond functionalization, isomerization and reduction of isolated C=C bonds<sup>140</sup>. Some of these issues are being addressed by combining chemical and enzymatic reactions<sup>185</sup>. Biocatalysis also offers opportunities to develop reactions that are chemically difficult, such as remote and selective C–H activation, which is often observed in nature<sup>148,178</sup>, but the scale and substrate scope remain limitations for biocatalytic C–H activation<sup>186</sup>.

The use of enzyme cascades is a particularly attractive aspect of biocatalysis, because of general reaction compatibility and the ability to telescope several reactions either in cell-free systems or whole cells<sup>103</sup>, akin to biosynthetic pathways. The design of such cascades is already starting to become automated using dedicated computational tools and databases that provide rich resources to the scientific community<sup>8</sup>. The accessibility of obtaining biocatalysts through commercial sources or from synthetic genes is continuing to lower the barrier to entry for biosynthesis and the bottleneck for reaction screening is now often at the assay stage, where more label-free high-throughput analytical methods are needed<sup>50</sup>. Current successes for compounds such as islatravir<sup>7</sup> and molnupiravir<sup>154</sup> have demonstrated the application of biocatalysis to multistep syntheses of small molecules. The next challenges will be to extend the application scope to targets of increasing molecular complexity and size, as well as to decrease the time required to develop efficient biocatalytic industrial processes. Examples of production of bulk chemicals and polymers by biocatalysis are still rare and offer a rich opportunity in terms of green chemistries. Biocatalysis also has a role to play in generating new modalities more



efficiently and selectively for the biopharmaceutical industries, such as producing biomacromolecules and antibody–drug conjugates.

Looking to the future, there are numerous key trends and scientific breakthroughs that are promising to have a significant impact on accelerating the discovery, development and application of biocatalysts. First of all, the range of chemical, new to nature biocatalytic reactions is rapidly expanding using *de novo* design<sup>187</sup> and/or directed evolution<sup>188</sup>. Increasingly powerful computational tools will allow for better *de novo* design but will also provide better selection tools for identifying suitable biocatalysts from the rich protein primary sequence information already accessible in databanks. Advances in computational methods to predict the protein structure from sequences through artificial intelligence<sup>189</sup> and subsequent prediction of function and physico-chemical properties will provide access to biocatalysts

that are finely tuned to the requirements of a desired target reaction and/or product<sup>190–192</sup>. To maximize synthetic utility, these tools will need to be integrated with the design of new biocatalytic cascade processes. Many individual steps of biocatalyst development can already be automated at the implementation stage, including desktop DNA printing, cell-free protein expression, enzyme immobilization and analysis, which hints at the potential for ‘fully automated biocatalytic synthesizers’ being available to individual laboratories<sup>3</sup> within the next decade.

In conclusion, biocatalysis has enabled essential contributions to the safe, cheap and sustainable production of high-value chemicals and pharmaceuticals, but still provides many exciting challenges for potential advancements.

Published online: 24 June 2021

- Yamada, H. & Kobayashi, M. Nitrile hydratase and its application to industrial production of acrylamide. *Biosci. Biotechnol. Biochem.* **60**, 1391–1400 (1996).
- Kirk, O., Borchert, T. V. & Fuglsang, C. C. Industrial enzyme applications. *Curr. Opin. Biotechnol.* **13**, 345–351 (2002).
- Devine, P. N. et al. Extending the application of biocatalysis to meet the challenges of drug development. *Nat. Rev. Chem.* **2**, 409–421 (2018).
- Wu, S., Snajdrova, R., Moore, J. C., Baldenius, K. & Bornscheuer, U. T. Biocatalysis: enzymatic synthesis for industrial applications. *Angew. Chem. Int. Ed.* **60**, 88–119 (2021).
- Sheldon, R. A., Brady, D. & Bode, M. L. The hitchhiker's guide to biocatalysis: recent advances in the use of enzymes in organic synthesis. *Chem. Sci.* **11**, 2587–2605 (2020).  
**This article presents an excellent recent overall review of biocatalysis.**
- Birmingham, W. R. et al. Bioretrosynthetic construction of a didanosine biosynthetic pathway. *Nat. Chem. Biol.* **10**, 392–399 (2014).  
**This article develops the concept of ‘bio-retrosynthesis’ and its application to afford an important biomolecule.**
- Huffman, M. A. et al. Design of an in vitro biocatalytic cascade for the manufacture of islatravir. *Science* **366**, 1255–1259 (2019).  
**This article presents the very impressive development of a multistep enzyme cascade with multiple enzyme engineering challenges by an industrial team.**
- Finnigan, W. et al. RetroBioCat as a computer-aided synthesis planning tool for biocatalytic reactions and cascades. *Nat. Catal.* **4**, 98–104 (2021).  
**This article establishes a database and computer-aided synthesis planning tool that allows scientists to design enzyme cascades.**
- Charnock, S., Bernardini, A. M., Monza, E., Lucas, M. F. & Sutton, P. W. in *Applied Biocatalysis* (eds Whittall, J. & Sutton, P. W.) 27–133 (Wiley, 2020).
- Bateman, A. et al. UniProt: the Universal Protein Knowledgebase. *Nucleic Acids Res.* **45**, D158–D169 (2017).
- Buchholz, P. C. F. et al. BioCatNet: a database system for the integration of enzyme sequences and biocatalytic experiments. *ChemBioChem* **17**, 2093–2098 (2016).
- Scott, T. A. & Piel, J. The hidden enzymology of bacterial natural product biosynthesis. *Nat. Rev. Chem.* **3**, 404–425 (2019).
- Martinez, S. & Hausinger, R. P. Catalytic mechanisms of Fe(II)- and 2-oxoglutarate-dependent oxygenases. *J. Biol. Chem.* **290**, 20702–20711 (2015).
- Zwick, C. R. & Renata, H. Harnessing the biocatalytic potential of iron- and  $\alpha$ -ketoglutarate-dependent dioxygenases in natural product total synthesis. *Nat. Prod. Rep.* **37**, 1065–1079 (2020).
- Lukat, P. et al. Biosynthesis of methyl-proline containing griselimycins, natural products with anti-tuberculosis activity. *Chem. Sci.* **8**, 7521–7527 (2017).
- Zwick, C. R. & Renata, H. Remote C–H hydroxylation by an  $\alpha$ -ketoglutarate-dependent dioxygenase enables efficient chemoenzymatic synthesis of manzacidin C and proline analogs. *J. Am. Chem. Soc.* **140**, 1165–1169 (2018).
- Neugebauer, M. E. et al. A family of radical halogenases for the engineering of amino-acid-based products. *Nat. Chem. Biol.* **15**, 1009–1016 (2019).
- Chekan, J. R. et al. Scalable biosynthesis of the seaweed neurochemical, kainic acid. *Angew. Chem. Int. Ed.* **58**, 8454–8457 (2019).
- Chakrabarty, S., Wang, Y., Perkins, J. C. & Narayan, A. R. H. Scalable biocatalytic C–H oxyfunctionalization reactions. *Chem. Soc. Rev.* **49**, 8137–8155 (2020).  
**This article presents an excellent review on the current state of the art of C–H oxyfunctionalizations of organic molecules using biocatalysis.**
- Liao, C. & Seebeck, F. P. S-Adenosylhomocysteine as a methyl transfer catalyst in biocatalytic methylation reactions. *Nat. Catal.* **2**, 696–701 (2019).
- Marsh, C. O. et al. A natural Diels–Alder biocatalyst enables efficient [4 + 2] cycloaddition under harsh reaction conditions. *ChemCatChem* **11**, 5027–5031 (2019).
- Fisher, B. F., Snodgrass, H. M., Jones, K. A., Andorfer, M. C. & Lewis, J. C. Site-selective C–H halogenation using flavin-dependent halogenases identified via family-wide activity profiling. *ACS Cent. Sci.* **5**, 1844–1856 (2019).
- Galanie, S., Entwistle, D. & Lalonde, J. Engineering biosynthetic enzymes for industrial natural product synthesis. *Nat. Prod. Rep.* **37**, 1122–1145 (2020).
- Schultz, B. J., Kim, S. Y., Lau, W. & Sattely, E. S. Total biosynthesis for milligram-scale production of etoposide intermediates in a plant chassis. *J. Am. Chem. Soc.* **141**, 19231–19235 (2019).
- Chen, M., Liu, C. T. & Tang, Y. Discovery and biocatalytic application of a PLP-dependent amino acid  $\gamma$ -substitution enzyme that catalyzes C–C bond formation. *J. Am. Chem. Soc.* **142**, 10506–10515 (2020).
- Bar-Even, A. et al. The moderately efficient enzyme: evolutionary and physicochemical trends shaping enzyme parameters. *Biochemistry* **50**, 4402–4410 (2011).
- Goldsmith, M. & Tawfik, D. S. Enzyme engineering: reaching the maximal catalytic efficiency peak. *Curr. Opin. Struct. Biol.* **47**, 140–150 (2017).
- Handelsman, J. Metagenomics: application of genomics to uncultured microorganisms. *Microbiol. Mol. Biol. Rev.* **68**, 669–685 (2004).
- Iqbal, H. A., Feng, Z. & Brady, S. F. Biocatalysts and small molecule products from metagenomic studies. *Curr. Opin. Chem. Biol.* **16**, 109–116 (2012).
- Coscolin, C. et al. Bioprospecting reveals class  $\omega$ -transaminases converting bulky ketones and environmentally relevant polyamines. *Appl. Environ. Microbiol.* **85**, e02404-18 (2019).
- Green, A. P., Turner, N. J. & O'Reilly, E. Chiral amine synthesis using  $\omega$ -transaminases: an amine donor that displaces equilibria and enables high-throughput screening. *Angew. Chem. Int. Ed.* **53**, 10714–10717 (2014).
- Baud, D., Ladkau, N., Moody, T. S., Ward, J. M. & Hailes, H. C. A rapid, sensitive colorimetric assay for the high-throughput screening of transaminases in liquid or solid-phase. *Chem. Commun.* **51**, 17225–17228 (2015).
- Nasseri, S. A., Betschart, L., Opaleva, D., Rahfeld, P. & Withers, S. G. A mechanism-based approach to screening metagenomic libraries for discovery of unconventional glycosidases. *Angew. Chem. Int. Ed.* **57**, 11359–11364 (2018).
- Colin, P. Y. et al. Ultrahigh-throughput discovery of promiscuous enzymes by picodroplet functional metagenomics. *Nat. Commun.* **6**, 1–12 (2015).
- Rahfeld, P. et al. An enzymatic pathway in the human gut microbiome that converts A to universal O type blood. *Nat. Microbiol.* **4**, 1475–1485 (2019).  
**This article identifies a two-enzyme system from a large gut microbiome library that enables generation of universal O-type blood from A-type donors.**
- Smith, D. R. M. et al. An unusual flavin-dependent halogenase from the metagenome of the marine sponge *Theonella swinhoei* WA. *ACS Chem. Biol.* **12**, 1281–1287 (2017).
- Baud, D., Jeffries, J. W. E., Moody, T. S., Ward, J. M. & Hailes, H. C. A metagenomics approach for new biocatalyst discovery: application to transaminases and the synthesis of allylic amines. *Green. Chem.* **19**, 1134–1143 (2017).
- Armstrong, Z. et al. Metagenomics reveals functional synergy and novel polysaccharide utilization loci in the *Castor canadensis* fecal microbiome. *ISME J.* **12**, 2757–2769 (2018).
- Bornscheuer, U. T. et al. Engineering the third wave of biocatalysis. *Nature* **485**, 185–194 (2012).
- Arnold, F. H. Directed evolution: bringing new chemistry to life. *Angew. Chem. Int. Ed.* **57**, 4143–4148 (2018).  
**This article presents a review of directed evolution by the 2018 Nobel Prize Laureate for Chemistry.**
- Qu, G., Li, A., Acevedo-Rocha, C. G., Sun, Z. & Reetz, M. T. The crucial role of methodology development in directed evolution of selective enzymes. *Angew. Chem. Int. Ed.* **59**, 13204–13231 (2020).
- Reetz, M. T., Bocola, M., Carballeira, J. D., Zha, D. & Vogel, A. Expanding the range of substrate acceptance of enzymes: combinatorial active-site saturation test. *Angew. Chem. Int. Ed.* **44**, 4192–4196 (2005).
- Reetz, M. T., Kahakeaw, D. & Lohmer, R. Addressing the numbers problem in directed evolution. *ChemBioChem* **9**, 1797–1804 (2008).
- Kille, S. et al. Reducing codon redundancy and screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2**, 83–92 (2013).
- Cadwell, R. C. & Joyce, G. F. Mutagenic PCR. *CSH Protoc.* <https://doi.org/10.1101/pdb.prot4143> (2006).



46. Stemmer, W. P. C. Rapid evolution of a protein in vitro by DNA shuffling. *Nature* **370**, 389–391 (1994). **This article establishes DNA shuffling as a method for generation of gene libraries for directed evolution.**
47. Schober, M. et al. Chiral synthesis of LSD1 inhibitor GSK2879552 enabled by directed evolution of an imine reductase. *Nat. Catal.* **2**, 909–915 (2019).
48. Heath, R. S., Pontini, M., Bechi, B. & Turner, N. J. Development of an *R*-selective amine oxidase with broad substrate specificity and high enantioselectivity. *ChemCatChem* **6**, 996–1002 (2014).
49. Weiß, M. S., Pavlidis, I. V., Vickers, C., Hohne, M. & Bornscheuer, U. T. Glycine oxidase based high-throughput solid-phase assay for substrate profiling and directed evolution of (*R*)- and (*S*)-selective amine transaminases. *Anal. Chem.* **86**, 11847–11853 (2014).
50. Yan, C. et al. Real-time screening of biocatalysts in live bacterial colonies. *J. Am. Chem. Soc.* **139**, 1408–1411 (2017). **This article presents recent approaches to using label-free screening methods based on mass spectrometry.**
51. Becker, S., Schmoldt, H. U., Adams, T. M., Wilhelm, S. & Kolmar, H. Ultra-high-throughput screening based on cell-surface display and fluorescence-activated cell sorting for the identification of novel biocatalysts. *Curr. Opin. Biotechnol.* **15**, 323–329 (2004).
52. Chen, T. et al. Evolution of thermophilic DNA polymerases for the recognition and amplification of C2'-modified DNA. *Nat. Chem.* **8**, 556–562 (2016).
53. Agresti, J. J. et al. Ultrahigh-throughput screening in drop-based microfluidics for directed evolution. *Proc. Natl Acad. Sci. USA* **107**, 4004–4009 (2010).
54. Obexer, R. et al. Emergence of a catalytic tetrad during evolution of a highly active artificial aldolase. *Nat. Chem.* **9**, 50–56 (2017).
55. Wang, L. & Schultz, P. G. A general approach for the generation of orthogonal tRNAs. *Chem. Biol.* **8**, 883–890 (2001).
56. Bryson, D. I. et al. Continuous directed evolution of aminoacyl-tRNA synthetases. *Nat. Chem. Biol.* **13**, 1253–1260 (2017).
57. Ravikumar, A., Arrieta, A. & Liu, C. C. An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* **10**, 175–177 (2014).
58. Ghislieri, D. et al. Engineering an enantioselective amine oxidase for the synthesis of pharmaceutical building blocks and alkaloid natural products. *J. Am. Chem. Soc.* **135**, 10863–10869 (2013).
59. Debon, A. et al. Ultrahigh-throughput screening enables efficient single-round oxidase remodelling. *Nat. Catal.* **2**, 740–747 (2019).
60. Holland-Moritz, D. A. et al. Mass activated droplet sorting (MADS) enables high-throughput screening of enzymatic reactions at nanoliter scale. *Angew. Chem. Int. Ed.* **59**, 4470–4477 (2020). **This article applies droplet sorting as one of the most successful methods for ultra-high-throughput screening of enzyme libraries.**
61. Wan, W., Sharp, J. M. & Liu, W. P. Pyrrolysyl-tRNA synthetase: an ordinary enzyme but an outstanding genetic code expansion tool. *Biochim. Biophys. Acta* **1844**, 1059–1070 (2014).
62. Warshel, A. et al. Electrostatic basis for enzyme catalysis. *Chem. Rev.* **106**, 3210–3235 (2006).
63. Boehr, D. D., Nussinov, R. & Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol.* **5**, 789–796 (2009).
64. Osuna, S. The challenge of predicting distal active site mutations in computational enzyme design. *WIREs Comput. Mol. Sci.* **11**, e1502 (2021).
65. Kiss, G., Çelebi-Ölcüm, N., Moretti, R., Baker, D. & Houk, K. N. Computational enzyme design. *Angew. Chem. Int. Ed.* **52**, 5700–5725 (2013).
66. Privett, H. K. et al. Iterative approach to computational enzyme design. *Proc. Natl Acad. Sci. USA* **109**, 3790–3795 (2012).
67. Wijma, H. J. et al. Enantioselective enzymes by computational design and in silico screening. *Angew. Chem.* **127**, 3797–3801 (2015).
68. Davey, J. A. & Chica, R. A. Multistate approaches in computational protein design. *Protein Sci.* **21**, 1241–1252 (2012).
69. Mondal, D., Kolev, V. & Warshel, A. Combinatorial approach for exploring conformational space and activation barriers in computer-aided enzyme design. *ACS Catal.* **10**, 6002–6012 (2020).
70. Maria-Solano, M. A., Serrano-Hervás, E., Romero-Rivera, A., Iglesias-Fernández, J. & Osuna, S. Role of conformational dynamics in the evolution of novel enzyme function. *Chem. Commun.* **54**, 6622–6634 (2018).
71. Campbell, E. C. et al. Laboratory evolution of protein conformational dynamics. *Curr. Opin. Struct. Biol.* **50**, 49–57 (2018).
72. Crean, R. M., Gardner, J. M. & Kamerlin, S. C. L. Harnessing conformational plasticity to generate designer enzymes. *J. Am. Chem. Soc.* **142**, 11324–11342 (2020).
73. Kreß, N., Halder, J. M., Rapp, L. R. & Hauer, B. Unlocked potential of dynamic elements in protein structures: channels and loops. *Curr. Opin. Chem. Biol.* **47**, 109–116 (2018).
74. Vavra, O. et al. CaverDock: a molecular docking-based tool to analyse ligand transport through protein tunnels and channels. *Bioinformatics* **35**, 4986–4993 (2019).
75. Otten, R. et al. How directed evolution reshapes the energy landscape in an enzyme to boost catalysis. *Science* **370**, 1442–1446 (2020).
76. Romero-Rivera, A., García-Borrás, M. & Osuna, S. Role of conformational dynamics in the evolution of retro-aldolase activity. *ACS Catal.* **7**, 8524–8532 (2017).
77. Casini, A. et al. A pressure test to make 10 molecules in 90 days: external evaluation of methods to engineer biology. *J. Am. Chem. Soc.* **140**, 4302–4316 (2018).
78. Gardner, J. M., Biler, M., Rizzo, V. A., Sanchez-Ruiz, J. M. & Kamerlin, S. C. L. Manipulating conformational dynamics to repurpose ancient proteins for modern catalytic functions. *ACS Catal.* **10**, 4863–4870 (2020).
79. Pabis, A., Rizzo, V. A., Sanchez-Ruiz, J. M. & Kamerlin, S. C. Cooperativity and flexibility in enzyme evolution. *Curr. Opin. Struct. Biol.* **48**, 83–92 (2018).
80. Liu, C. C. & Schultz, P. G. Adding new chemistries to the genetic code. *Annu. Rev. Biochem.* **79**, 413–444 (2010). **This article reviews the methods to expand the genetic code for the introduction of non-natural amino acids into proteins.**
81. Chin, J. W. Expanding and reprogramming the genetic code. *Nature* **550**, 53–60 (2017).
82. Seyedsayamdost, M. R., Xie, J., Chan, C. T. Y., Schultz, P. G. & Stubbe, J. Site-specific insertion of 3-aminotyrosine into subunit  $\alpha 2$  of *E. coli* ribonucleotide reductase: direct evidence for involvement of Y730 and Y731 in radical propagation. *J. Am. Chem. Soc.* **129**, 15060–15071 (2007).
83. Faraldos, J. A. et al. Probing eudesmane cation- $\pi$  interactions in catalysis by aristolochene synthase with non-canonical amino acids. *J. Am. Chem. Soc.* **133**, 13906–13909 (2011).
84. Wu, Y. & Boxer, S. G. A critical test of the electrostatic contribution to catalysis with noncanonical amino acids in ketosteroid isomerase. *J. Am. Chem. Soc.* **138**, 11890–11895 (2016).
85. Huguenin-Dezot, N. et al. Trapping biosynthetic acyl-enzyme intermediates with encoded 2,3-diaminopropionic acid. *Nature* **565**, 112–117 (2019).
86. Ortmayer, M. et al. Rewiring the 'push-pull' catalytic machinery of a heme enzyme using an expanded genetic code. *ACS Catal.* **10**, 2735–2746 (2020).
87. Burke, A. J. et al. Design and evolution of an enzyme with a non-canonical organocatalytic mechanism. *Nature* **570**, 219–225 (2019). **This article is a recent example of using non-canonical amino acids in protein evolution to obtain new enzyme activities.**
88. Drienovská, I., Mayer, C., Dulson, C. & Roelfes, G. A designer enzyme for hydrazine and oxime formation featuring an unnatural catalytic aniline residue. *Nat. Chem.* **10**, 946–952 (2018).
89. Santoro, S. W., Wang, L., Herberich, B., King, D. S. & Schultz, P. G. An efficient system for the evolution of aminoacyl-tRNA synthetase specificity. *Nat. Biotechnol.* **20**, 1044–1048 (2002).
90. Pott, M. et al. A noncanonical proximal heme ligand affords an efficient peroxidase in a globin fold. *J. Am. Chem. Soc.* **140**, 1535–1543 (2018).
91. Li, J. C., Liu, T., Wang, Y., Mehta, A. P. & Schultz, P. G. Enhancing protein stability with genetically encoded noncanonical amino acids. *J. Am. Chem. Soc.* **140**, 15997–16000 (2018).
92. Wur, R. P. Chiral dialkylaminopyridine catalysts in asymmetric synthesis. *Chem. Rev.* **107**, 5570–5595 (2007).
93. Bolon, D. N. & Mayo, S. L. Enzyme-like proteins by computational design. *Proc. Natl Acad. Sci. USA* **98**, 14274–14279 (2001).
94. Richter, F. et al. Computational design of catalytic dyads and oxyanion holes for ester hydrolysis. *J. Am. Chem. Soc.* **134**, 16197–16206 (2012).
95. Moroz, Y. S. et al. New tricks for old proteins: single mutations in a nonenzymatic protein give rise to various enzymatic activities. *J. Am. Chem. Soc.* **137**, 14905–14911 (2015).
96. Burton, A. J., Thomson, A. R., Dawson, W. M., Brady, R. L. & Woolfson, D. N. Installing hydrolytic activity into a completely de novo protein framework. *Nat. Chem.* **8**, 837–844 (2016).
97. Nazor, J., Liu, J. & Huisman, G. Enzyme evolution for industrial biocatalytic cascades. *Curr. Opin. Biotechnol.* **69**, 182–190 (2021). **This review focuses on recent industrial biocatalytic cascades.**
98. McIntosh, J. A. & Owens, A. Enzyme engineering for biosynthetic cascades. *Curr. Opin. Green Sustain. Chem.* **29**, 100448 (2021).
99. Schrittwieser, J. H., Velikogne, S., Hall, M. & Kroutil, W. Artificial biocatalytic linear cascades for preparation of organic molecules. *Chem. Rev.* **118**, 270–348 (2018).
100. Mayer, S. F., Kroutil, W. & Faber, K. Enzyme-initiated domino (cascade) reactions. *Chem. Soc. Rev.* **30**, 332–339 (2001).
101. García-Junceda, E., Lavandera, I., Rother, D. & Schrittwieser, J. H. (Chemo)enzymatic cascades — nature's synthetic strategy transferred to the laboratory. *J. Mol. Catal. B Enzym.* **114**, 1–6 (2015).
102. Rudroff, F. et al. Opportunities and challenges for combining chemo- and biocatalysis. *Nat. Catal.* **1**, 12–22 (2018).
103. France, S. P., Hepworth, L. J., Turner, N. J. & Flitsch, S. L. Constructing biocatalytic cascades: in vitro and in vivo approaches to de novo multi-enzyme pathways. *ACS Catal.* **7**, 710–724 (2017). **This article reviews the literature on a wide range of multienzyme de novo cascades using isolated enzymes and whole-cell systems.**
104. Turner, N. J. & O'Reilly, E. Biocatalytic retrosynthesis. *Nat. Chem. Biol.* **9**, 285–288 (2013). **This review develops the concept of biocatalytic retrosynthesis.**
105. Hönig, M., Söndermann, P., Turner, N. J. & Carreira, E. M. Enantioselective chemo- and biocatalysis: partners in retrosynthesis. *Angew. Chem. Int. Ed.* **56**, 8942–8973 (2017).
106. Green, A. P. & Turner, N. J. Biocatalytic retrosynthesis: redesigning synthetic routes to high-value chemicals. *Perspect. Sci.* **9**, 42–48 (2016).
107. de Souza, R. O. M. A., Miranda, L. S. M. & Bornscheuer, U. T. A retrosynthetic approach for biocatalysis in organic synthesis. *Chem. A Eur. J.* **23**, 12040–12063 (2017).
108. Bachmann, B. O. Biosynthesis: is it time to go retro? *Nat. Chem. Biol.* **6**, 390–393 (2010).
109. Hadadi, N. & Hatzimanikatis, V. Design of computational retrobiosynthesis tools for the design of de novo synthetic pathways. *Curr. Opin. Chem. Biol.* **28**, 99–104 (2015).
110. Coley, C. W., Green, W. H. & Jensen, K. F. Machine learning in computer-aided synthesis planning. *Acc. Chem. Res.* **51**, 1281–1289 (2018).
111. Mohamad, N. R., Marzuki, N. H. C., Buang, N. A., Huyop, F. & Wahab, R. A. An overview of technologies for immobilization of enzymes and surface analysis techniques for immobilized enzymes. *Biotechnol. Biotechnol. Equip.* **29**, 205–220 (2015).
112. Basso, A. & Serban, S. Industrial applications of immobilized enzymes — a review. *Mol. Catal.* **479**, 110607 (2019).
113. Fryszkowska, A. & Devine, P. N. Biocatalysis in drug discovery and development. *Curr. Opin. Chem. Biol.* **55**, 151–160 (2020).
114. Latham, J. et al. in *Applied Biocatalysis* (eds Whittall, J. & Sutton, P. W.) 1–25 (Wiley, 2020).
115. Prier, C. K. & Kosjek, B. Recent preparative applications of redox enzymes. *Curr. Opin. Chem. Biol.* **49**, 105–112 (2019).
116. Woodley, J. M. New frontiers in biocatalysis for sustainable synthesis. *Curr. Opin. Green. Sustain. Chem.* **21**, 22–26 (2020).
117. Sheldon, R. A. & Woodley, J. M. Role of biocatalysis in sustainable chemistry. *Chem. Rev.* **118**, 801–838 (2018).

118. Sheldon, R. A. Metrics of green chemistry and sustainability: past, present, and future. *ACS Sustain. Chem. Eng.* **6**, 32–48 (2018).
119. Tieves, F. et al. Energising the E-factor: the E-factor. *Tetrahedron* **75**, 1311–1314 (2019).
120. Hinzmann, A., Glinski, S., Worm, M. & Gröger, H. Enzymatic synthesis of aliphatic nitriles at a substrate loading of up to 1.4 kg/l: a biocatalytic record achieved with a heme protein. *J. Org. Chem.* **84**, 4867–4872 (2019).
121. Bornadel, A. et al. Technical considerations for scale-up of imine-reductase-catalyzed reductive amination: a case study. *Org. Process. Res. Dev.* **23**, 1262–1268 (2019).
122. Hülsewede, D., Meyer, L. & von Langermann, J. Application of in situ product crystallization and related techniques in biocatalytic processes. *Chem. A Eur. J.* **25**, 4871–4884 (2019).
123. Fellechner, O., Blatkiewicz, M. & Smirnova, I. Reactive separations for in situ product removal of enzymatic reactions: a review. *Chem. Ing. Tech.* **91**, 1522–1543 (2019).
124. Aalbers, F. S. et al. Approaching boiling point stability of an alcohol dehydrogenase through computationally-guided enzyme engineering. *eLife* **9**, e54639 (2020).
125. Gumulya, Y. et al. Engineering highly functional thermostable proteins using ancestral sequence reconstruction. *Nat. Catal.* **1**, 878–888 (2018).
126. Thomas, A., Cutlan, R., Finnigan, W., van der Giezen, M. & Harmer, N. Highly thermostable carboxylic acid reductases generated by ancestral sequence reconstruction. *Commun. Biol.* **2**, 1–12 (2019).
127. Nicoll, C. R. et al. Ancestral-sequence reconstruction unveils the structural basis of function in mammalian FMOs. *Nat. Struct. Mol. Biol.* **27**, 14–24 (2020).
128. Gomez-Fernandez, B. J., Rizzo, V. A., Rueda, A., Sanchez-Ruiz, J. M. & Alcalde, M. Ancestral resurrection and directed evolution of fungal mesozoic laccases. *Appl. Environ. Microbiol.* **86**, e00778-20 (2020).
129. Carletti, M. S. et al. Revenant: a database of resurrected proteins. *Database*. **2020**, 31 (2020).
130. Truppo, M. D., Strotman, H. & Hughes, G. Development of an immobilized transaminase capable of operating in organic solvent. *ChemCatChem* **4**, 1071–1074 (2012).
131. Matthey, A. P. et al. Natural heterogeneous catalysis with immobilized oxidase biocatalysts. *RSC Adv.* **10**, 19501–19505 (2020).
132. Böhmer, W. et al. Highly efficient production of chiral amines in batch and continuous flow by immobilized  $\omega$ -transaminases on controlled porosity glass metal-ion affinity carrier. *J. Biotechnol.* **291**, 52–60 (2019).
133. Britton, J., Majumdar, S. & Weiss, G. A. Continuous flow biocatalysis. *Chem. Soc. Rev.* **47**, 5891–5918 (2018).
134. Rodrigues, R. C., Ortiz, C., Berenguer-Murcia, A., Torres, R. & Fernández-Lafuente, R. Modifying enzyme activity and selectivity by immobilization. *Chem. Soc. Rev.* **42**, 6290–6307 (2013).
135. Sheldon, R. A. in *Green Biocatalysis* (ed. Patel, R. N.) 1–15 (Wiley, 2016).
136. Cespuogli, M. et al. Rice husk as an inexpensive renewable immobilization carrier for biocatalysts employed in the food, cosmetic and polymer sectors. *Catalysts* **8**, 471 (2018).
137. Woodley, J. M. Towards the sustainable production of bulk-chemicals using biotechnology. *N. Biotechnol.* **59**, 59–64 (2020).
138. Hughes, D. L. Biocatalysis in drug development — highlights of the recent patent literature. *Org. Process. Res. Dev.* **22**, 1063–1080 (2018).
139. de María, P., de Gonzalo, G. & Alcántara, A. Biocatalysis as useful tool in asymmetric synthesis: an assessment of recently granted patents (2014–2019). *Catalysts* **9**, 802 (2019).
140. Hauer, B. Embracing nature's catalysts: a viewpoint on the future of biocatalysis. *ACS Catal.* **10**, 8418–8427 (2020).  
**This article presents an insightful review of the future challenges of biocatalysis in academia and industry.**
141. Jiao, S., Li, F., Yu, H. & Shen, Z. Advances in acrylamide bioproduction catalyzed with *Rhodococcus* cells harboring nitrile hydratase. *Appl. Microbiol. Biotechnol.* **104**, 1001–1012 (2020).
142. Savile, C. K. et al. Biocatalytic asymmetric synthesis of chiral amines from ketones applied to sitagliptin manufacture. *Science* **329**, 305–309 (2010).
143. Eichhorn, E. et al. Biocatalytic process for (–)-ambrox production using squalene hopene cyclase. *Adv. Synth. Catal.* **360**, 2359–2351 (2018).
144. Baker Dockrey, S. A., Lukowski, A. L., Becker, M. R. & Narayan, A. R. H. Biocatalytic site- and enantioselective oxidative dearomatization of phenols. *Nat. Chem.* **10**, 119–125 (2018).
145. Li, G., Wang, J.-B. & Reetz, M. T. Biocatalysts for the pharmaceutical industry created by structure-guided directed evolution of stereoselective enzymes. *Bioorg. Med. Chem.* **26**, 1241–1251 (2018).
146. Arora, K. K. et al. Manufacturing Process and Intermediates for a Pyrrolo[2,3-D]Pyrimidine Compound and use Thereof. US Patent 10,815,240 (2020).
147. Jaeger, K. E., Eggert, T., Eipper, A. & Reetz, M. T. Directed evolution and the creation of enantioselective biocatalysts. *Appl. Microbiol. Biotechnol.* **55**, 519–530 (2001).
148. Manning, J. et al. Regio- and enantio-selective chemo-enzymatic C–H-lactonization of decanoic acid to (S)- $\delta$ -decalactone. *Angew. Chem. Int. Ed.* **58**, 5668–5671 (2019).
149. Zhang, J. et al. Engineered C–N lyase: enantioselective synthesis of chiral synthons for artificial dipeptide sweeteners. *Angew. Chem.* **132**, 437–443 (2020).
150. Bruggink, A., Schoevaart, R. & Kieboom, T. Concepts of nature in organic synthesis: cascade catalysis and multistep conversions in concert. *Org. Process. Res. Dev.* **7**, 622–640 (2003).
151. Sperl, J. M. & Sieber, V. Multienzyme cascade reactions — status and recent advances. *ACS Catal.* **8**, 2385–2396 (2018).
152. Lenz, M., Borlinghaus, N., Weinmann, L. & Nestl, B. M. Recent advances in imine reductase-catalyzed reactions. *World J. Microbiol. Biotechnol.* **33**, 199 (2017).
153. Schaffer, S. et al. Producing amines and diamines from a carboxylic acid or dicarboxylic acid or a monoester thereof. US Patent 9,725,746 (2017).
154. Benkovics, T. et al. Evolving to an ideal synthesis of molnupiravir, an investigational treatment for COVID-19. Preprint at <https://doi.org/10.26434/chemrxiv.13472373.v1> (2020).
155. Yin, Z. et al. Computing platforms for big biological data analytics: perspectives and challenges. *Comput. Struct. Biotechnol. J.* **15**, 403–411 (2017).
156. Agarwala, R. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **44**, D7–D19 (2016).
157. Goodman, J. Computer SSoftware review: Reaxys Reaxys. Elsevier Properties SA 360 Park Avenue South, New York, NY 10010-1710. [www.info.reaxys.com](http://www.info.reaxys.com). *J. Chem. Inf. Model.* **49**, 2897–2898 (2009).
158. Garritano, J. R. Evolution of SciFinder, 2011–2013: new features, new content. *Sci. Technol. Libr.* **32**, 346–371 (2013).
159. Jeske, L., Placzek, S., Schomburg, I., Chang, A. & Schomburg, D. BRENDA in 2019: a European ELIXIR core data resource. *Nucleic Acids Res.* **47**, D542–D549 (2019).
160. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361 (2017).
161. Gunera, J., Kindinger, F., Li, S. M. & Kolb, P. PrendB, a substrate prediction database to enable biocatalytic use of prenyltransferases. *J. Biol. Chem.* **292**, 4003–4021 (2017).
162. Van Santen, J. A. et al. The natural products atlas: an open access knowledge base for microbial natural products discovery. *ACS Cent. Sci.* **5**, 1824–1833 (2019).
163. Tipton, K. F. et al. Standards for reporting enzyme data: the STREND consortium: what it aims to do and why it should be helpful. *Perspect. Sci.* **1**, 131–137 (2014).
164. Swainston, N. et al. STREND DB: enabling the validation and sharing of enzyme kinetics data. *FEBS J.* **285**, 2193–2204 (2018).
165. Perkel, J. M. The Internet of Things comes to the lab. *Nature* **542**, 125–126 (2017).
166. Jennings-Antipov, L. D. & Gardner, T. S. Digital publishing isn't enough: the case for 'blueprints' in scientific communication. *Emerg. Top. Life Sci.* **2**, 755–758 (2018).
167. Goodwin, N. C., Morrison, J. P., Fuerst, D. E. & Hadi, T. Biocatalysis in medicinal chemistry: challenges to access and drivers for adoption. *ACS Med. Chem. Lett.* **10**, 1363–1366 (2019).
168. Nielsen, J. & Keasling, J. D. Engineering cellular metabolism. *Cell* **164**, 1185–1197 (2016).
169. Truppo, M. D. Biocatalysis in the pharmaceutical industry: the need for speed. *ACS Med. Chem. Lett.* **8**, 476–480 (2017).  
**This article presents an excellent review of the adaptation and challenges of biocatalysis in the pharmaceutical industry, with particular focus on timescales of process development.**
170. Yang, K. K., Wu, Z. & Arnold, F. H. Machine-learning-guided directed evolution for protein engineering. *Nat. Methods* **16**, 687–694 (2019).
171. Markel, U. et al. Advances in ultrahigh-throughput screening for directed enzyme evolution. *Chem. Soc. Rev.* **49**, 233–262 (2020).
172. Bunzel, H. A., Garabou, X., Pott, M. & Hilvert, D. Speeding up enzyme discovery and engineering with ultrahigh-throughput methods. *Curr. Opin. Struct. Biol.* **48**, 149–156 (2018).
173. Hillson, N. et al. Building a global alliance of biofoundries. *Nat. Commun.* **10**, 1–4 (2019).
174. Windle, C. L., Müller, M., Nelson, A. & Berry, A. Engineering aldolases as biocatalysts. *Curr. Opin. Chem. Biol.* **19**, 25–33 (2014).
175. Brovetto, M., Gamena, D., Saenz Méndez, P. & Seoane, G. A. C–C bond-forming lyases in organic synthesis. *Chem. Rev.* **111**, 4346–4403 (2011).
176. Zetzsch, L. E. & Narayan, A. R. H. Broadening the scope of biocatalytic C–C bond formation. *Nat. Rev. Chem.* **4**, 334–346 (2020).
177. Li, Z. et al. Engineering cytochrome P450 enzyme systems for biomedical and biotechnological applications. *J. Biol. Chem.* **295**, 833–849 (2020).
178. Fessner, N. D. P450 monooxygenases enable rapid late-stage diversification of natural products via C–H bond activation. *ChemCatChem* **11**, 2226–2242 (2019).
179. Lall, M. S. et al. Late-stage lead diversification coupled with quantitative nuclear magnetic resonance spectroscopy to identify new structure–activity relationship vectors at nanomole-scale synthesis: application to loratadine, a human histamine H1 receptor inverse agonist. *J. Med. Chem.* **63**, 7268–7292 (2020).
180. Dong, J. J. et al. Biocatalytic oxidation reactions: a chemist's perspective. *Angew. Chem. Int. Ed.* **57**, 9238–9261 (2018).
181. Silverman, A. D., Karim, A. S. & Jewett, M. C. Cell-free gene expression: an expanded repertoire of applications. *Nat. Rev. Genet.* **21**, 151–170 (2020).
182. Khambhati, K. et al. Exploring the potential of cell-free protein synthesis for extending the abilities of biological systems. *Front. Bioeng. Biotechnol.* **7**, 248 (2019).
183. Zimmerman, J. B., Anastas, P. T., Erythropel, H. C. & Leitner, W. Designing for a green chemistry future. *Science* **367**, 397–400 (2020).
184. Hammer, S. C., Knight, A. M. & Arnold, F. H. Design and evolution of enzymes for non-natural chemistry. *Curr. Opin. Green Sustain. Chem.* **7**, 23–30 (2017).
185. DeHovitz, J. S. et al. Static to inducibly dynamic stereocontrol: the convergent use of racemic  $\beta$ -substituted ketones. *Science* **369**, 1113–1118 (2020).
186. O'Reilly, E., Köhler, V., Flitsch, S. L. & Turner, N. J. Cytochromes P450 as useful biocatalysts: addressing the limitations. *Chem. Commun.* **47**, 2490–2501 (2011).
187. Basler, S. et al. Efficient Lewis acid catalysis of an abiological reaction in a de novo protein scaffold. *Nat. Chem.* **13**, 231–235 (2021).
188. Liu, Z. & Arnold, F. H. New-to-nature chemistry from old protein machinery: carbene and nitrene transferases. *Curr. Opin. Biotechnol.* **69**, 43–51 (2021).  
**This review describes recent developments of new to nature reactions catalysed by engineered enzymes.**
189. Callaway, E. 'It will change everything': DeepMind's AI makes gigantic leap in solving protein structures. *Nature* **582**, 203–204 (2020).  
**This article presents a recent breakthrough in computational protein structure prediction from the primary sequence.**
190. Dou, J. et al. De novo design of a fluorescence-activating  $\beta$ -barrel. *Nature* **561**, 485–491 (2018).
191. Senior, A. W. et al. Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).

192. Huang, P. S., Boyken, S. E. & Baker, D. The coming of age of de novo protein design. *Nature* **537**, 320–327 (2016).
193. Fischer, M. & Pleiss, J. The Lipase Engineering Database: a navigation and analysis tool for protein families. *Nucleic Acids Res.* **31**, 319–321 (2003).
194. Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P. M. & Henrissat, B. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **42**, D490–D495 (2014).  
**This article discusses the curated comprehensive database of carbohydrate-active enzymes (CAZymes) that has been a useful tool to the glycoscience community.**
195. Savelli, B. et al. RedoxiBase: a database for ROS homeostasis regulated proteins. *Redox Biol.* **26**, 101247 (2019).

#### Acknowledgements

The authors are grateful for funding from the European Research Council (ERC) (S.L.F., 788231; A.P.G., 757991; S.O., 679001; N.J.T., 742987), the Engineering and Physical Sciences Research Council (EPSRC) (S.L.F. and N.J.T., EP/S005226/1), the Biotechnology and Biological Sciences Research Council

(BBSRC) (S.L.F. and N.J.T., BB/M027791/1, BB/M028836/1; A.P.G., BB/M027023/1), the Spanish Ministry of Economy and Competitiveness (MINECO) (S.O., PGC2018-102192-B-I00), Generalitat de Catalunya (S.O., SGR 2017 1707) and the University of Manchester (Presidential Fellowship to S.L.L.).

#### Author contributions

Introduction (S.L.F., H.N. and K.S.R.); Experimentation (E.L.B., A.P.G., H.N., S.O., K.S.R., N.J.T., S.L.L., L.J.H. and W.F.); Results (M.A.H. and E.R.); Applications (S.P.F., M.A.H. and E.R.); Reproducibility and data deposition (W.F. and L.J.H.); Limitations and optimizations (W.F., L.J.H., M.A.H. and E.R.); Outlook (S.L.F.); overview of Primer (S.L.F.). All authors contributed equally to planning and revision of the manuscript as described. Please note that co-authors have been listed in alphabetical order.

#### Competing interests

The authors declare no competing interests.

#### Peer review information

*Nature Reviews Methods Primers* thanks L. Betancor, A. Fryszkowska and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

#### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

#### RELATED LINKS

**BioCatNet:** <https://www.biocatnet.de/>  
**Basic Local Alignment Search Tool:** <https://blast.ncbi.nlm.nih.gov/Blast.cgi>  
**BRENDA:** <https://www.brenda-enzymes.org/>  
**CAVER:** <http://www.caver.cz/>  
**KEGG:** <https://www.genome.jp/kegg/>  
**InterPro:** <https://www.ebi.ac.uk/interpro/>  
**Pfam:** <https://pfam.xfam.org/>  
**PrenDB:** <http://prendb.pharmazie.uni-marburg.de/prendb/home/>  
**Protein Data Bank:** <https://www.rcsb.org/>  
**RetroBioCat:** <https://retrobiocat.com/>  
**Revenant:** <https://revenant.inf.pucp.edu.pe/>  
**Rosetta:** <https://www.rosettacommons.org/>  
**UniProt:** <https://www.uniprot.org/>

© Springer Nature Limited 2021