Review: Engineering of thermostable enzymes for industrial applications

Federica Rigoldi, Stefano Donini, Alberto Redaelli, Emilio Parisini, and Alfonso Gautieri

---

Articles you may be interested in

A tunable microfluidic 3D stenosis model to study leukocyte-endothelial interactions in atherosclerosis
APL Bioengineering **2**, 016103 (2018); 10.1063/1.4993762

Photolithographic patterned surface forms size-controlled lipid vesicles
APL Bioengineering **2**, 016104 (2018); 10.1063/1.5002604

Synthesis and targeting of gold-coated ^{177}Lu-containing lanthanide phosphate nanoparticles—A potential theranostic agent for pulmonary metastatic disease
APL Bioengineering **2**, 016101 (2018); 10.1063/1.5018165

An evaluation of irreversible electroporation thresholds in human prostate cancer and potential correlations to physiological measurements
APL Bioengineering **1**, 016101 (2017); 10.1063/1.5005828

Editorial: Welcome to APL Bioengineering
APL Bioengineering **1**, 010401 (2017); 10.1063/1.5007077

Review: Bioengineering strategies to probe T cell mechanobiology
APL Bioengineering **2**, 021501 (2018); 10.1063/1.5006599

Review: Engineering of thermostable enzymes for industrial applications

Federica Rigoldi,[1] Stefano Donini,[2] Alberto Redaelli,[1] Emilio Parisini,[2,a)] and Alfonso Gautieri[1,a)]

[1]*Biomolecular Engineering Lab, Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, Italy*
[2]*Center for Nano Science and Technology at Polimi, Istituto Italiano di Tecnologia, Via G. Pascoli 70/3, 20133 Milano, Italy*

The catalytic properties of some selected enzymes have long been exploited to carry out efficient and cost-effective bioconversions in a multitude of research and industrial sectors, such as food, health, cosmetics, agriculture, chemistry, energy, and others. Nonetheless, for several applications, naturally occurring enzymes are not considered to be viable options owing to their limited stability in the required working conditions. Over the years, the quest for novel enzymes with actual potential for biotechnological applications has involved various complementary approaches such as mining enzyme variants from organisms living in extreme conditions (extremophiles), mimicking evolution in the laboratory to develop more stable enzyme variants, and more recently, using rational, computer-assisted enzyme engineering strategies. In this review, we provide an overview of the most relevant enzymes that are used for industrial applications and we discuss the strategies that are adopted to enhance enzyme stability and/or activity, along with some of the most relevant achievements. In all living species, many different enzymes catalyze fundamental chemical reactions with high substrate specificity and rate enhancements. Besides specificity, enzymes also possess many other favorable properties, such as, for instance, cost-effectiveness, good stability under mild pH and temperature conditions, generally low toxicity levels, and ease of termination of activity. As efficient natural biocatalysts, enzymes provide great opportunities to carry out important chemical reactions in several research and industrial settings, ranging from food to pharmaceutical, cosmetic, agricultural, and other crucial economic sectors. © 2018 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). https://doi.org/10.1063/1.4997367

I. THERMOSTABLE ENZYMES IN INDUSTRIAL APPLICATIONS

Enzymes have long been used in detergents and for food production, especially cheese, sourdough, beer, and wine as well as for the production of leather, indigo, and linen.[1,2] Such processes are usually triggered by enzymes produced by specific microorganisms or by enzymes that are present in natural products, such as, for instance, the papaya fruit, hence, not used as pure isolated molecules. The development of fermentation processes has allowed the large-scale production, isolation, and purification of enzymes from selected production strains, thus making it possible to provide biocatalysts to be used as molecular tools in actual industrial processes, such as, for instance, in detergent, textile, and starch industries. In particular, a commercial breakthrough occurred in 1957 with the first commercial protease production by Novozymes.[1]

---

a)Author to whom correspondence should be addressed: emilio.parisini@iit.it, Telephone: +39-02-2399-9870 and alfonso.gautieri@polimi.it, Telephone: +39-02-2399-4142.

More recent advancements in protein engineering and directed evolution have made the continuous development of new and more efficient enzymes a reality. Indeed, mutant enzymes for established technical applications or new tailor-made enzymes for areas of application where enzymes had not been previously used have been successfully introduced. Of those enzymes that are used in industrial processes, over half are from fungi, over one-third is from bacteria, and the rest originate from animal (8%) and plant (4%) sources.[3] Recombinant DNA techniques have allowed the isolation and cloning of genes encoding for enzymes from all possible sources, including microbes and other microorganisms that are particularly difficult to manipulate, and high-yield heterologous protein expression. As a result, this convenient technology has increased production levels and has shifted enzyme production from bacterial strains that are not suited for industry into industry-friendly microorganisms such as *Aspergillus*, *Trichoderma*, *Kluyveromyces*, *Saccharomyces*, *Yarrowia*, and *Bacillus*.[4]

On a large scale, the enzymes produced by microbial strains are economical due to high production levels associated with standard expression, ease of growth, inexpensive culture media, and short fermentation cycles. Downstream processing allows for a rapid examination of thousands of cultures. Furthermore, different microbes produce somewhat different enzymes that catalyze the same reaction, providing high levels of flexibility.

The International Union of Biochemistry (IUB) categorizes enzymes in six different classes, based on the enzyme action mechanism. The six enzyme classes are ligases, isomerases, oxidoreductases, lyases, transferases, and hydrolases. Currently, more than 75% of the enzymes that are used commercially are members of the hydrolase family and are employed for the degradation of a number of different natural substances. Of all the commercial hydrolases, proteases are the major and most important sub-type. Indeed, they are widely used in both the detergent industry and the starch industry. Proteases are also important components in textile, animal feed, and dairy industries. The second largest sub-group is represented by carbohydrases, mostly amylases and cellulases. For instance, carbohydrases are extensively used in productive sectors such as starch, textile, detergent, and baking industries, which is where most industrial enzymes are used.

It has been estimated that by 2018, the global market for industrial enzymes will surpass the USD 7.1 billion mark and its five-year compound annual growth rate (CAGR) will be around 8.2%. The market for food enzymes alone is projected to reach USD 2.94 billion by 2021, at a CAGR of 7.4% between 2016 and 2021. Moreover, it is expected that the maximum growth rate will be observed in the detergent enzyme segment (CAGR of 11.3% in the 2016–2021 period).[5] Proteases were the prominent product segment in 2015, accounting for 27.4% of the global enzyme market; now, they are expected to show an even more profitable growth in light of their increasing application in pharmaceutical, detergent, and chemical sectors.[6] The most relevant industrial applications of enzymes are summarized in Table I.

At the industrial level, enzymes are mostly used as detergent additives. Indeed, detergents are supplemented with proteases, lipases, amylases, oxidases, peroxidases, and cellulases in order to breakdown different types of chemical bonds in water. To this end, it is essential that they maintain their activity at high temperatures (60 °C) and high pH values (pH 9–11), in particular when mixed with other washing powder components. Nearly 25% of the total worldwide sales of enzymes is represented by proteases that are added to laundry detergents. In Secs. I A–I H, the most important families of industrially relevant enzymes are described, together with key applications.

A. Pectinases

In paper and textile industries, enzymes are increasingly utilized not only to develop cleaner processes but also to reduce both raw material usage and waste production. For instance, an enzymatic process based on a pectate lyase that allows the low-temperature removal of pectin and other hydrophobic materials from cotton fabrics has been developed.[7] The food industry also takes advantage of pectinases, particularly for the clarification of fruit juices, for the degumming of fibers, and for wine making.

TABLE I. The most relevant industrial applications of bacterial/fungal enzymes: in the first column, the enzyme name is specified, followed by its International Union of Biochemistry (IUB) class, the typical substrate, and its main applications. The IUB classes are Oxido-Reductases (1), Transferases (2), Hydrolase (3), Lyases (4), Isomerases (5), and Ligases (6).

Enzyme	IUB classes	Substrate classes	Industry field (application)
α-amylase (or glycogenase)	3	Carbohydrates	Baking (softness and volume of bread), laundry and detergents (starch strain removal), paper and pulp (deinking, drainage improvement, and starch coating), starch and fuel (starch liquefaction and saccharification), textile (removal of starch from woven fabrics and de-sizing), and food (juice treatments, low calorie beer, glucose, and fructose syrup production)
α-galactosidase	3	Glycolipids and glycoproteins	Dairy
β-glucanase	3	Glucose polysaccharide	Animal feed (digestibility improvement) and food (mashing)
Acetolactate-decarboxylase	4	(S)-2-hydroxy-2-methyl-3-oxo-butanoate (C-C bond)	Food (beer maturation)
Acylase	3	Penicillin	Organic synthesis (synthesis of semisynthetic penicillin)
Amyloglucosidase	3	Maltooligo- and polysaccharides	Personal care (antimicrobial combined with glucose oxidase) and starch and fuel (saccharification)
Asparaginase	3	Asparagine	Pharmaceutical (treatment of acute leukemia) and food (decrease acrylamide)
Catalase	1	Hydrogen peroxide	Food and textile (bleach termination)
Cellulase	3	Cellulose and related polysaccharides	Plant waste treatment, paper and pulp (modification of fibers, deinking, and drainage), textile (denim finishing and softening of cotton), cleaning (removal of stains), color clarification, and anti-redeposition
Cyclodextrin-glycosyltransferase	2	Cyclic dextrins (starting from polysaccharides)	Cyclodextrin production
Dextranase	3	Dextran	Pharmaceutical and food
Glucose isomerase	5	D-glucose and D-xylose	Starch and fuel and fructose syrup (glucose to fructose conversion)
Glucose oxidase	1	Glucose	Baking (strengthening of dough) and personal care (bleaching and antimicrobial)
Invertase	3	Sucrose	Food
Lactase	3	Lactose	Food (juice clarification, beer flavor, cork stopper treatment, and milk lactose removal) and textile (bleaching)
Lipase	3	Lipids	Baking (stability of dough and conditioning), food (cheese flavoring), laundry and detergents (lipid stain removal), textile (de-inking, de-pickling, and cleaning), organic synthesis (resolution of chiral alcohols and amides), fats and oils (transesterification and de-gumming, lysolecithin production for phospholipases), and paper and pulp (pitch control and contaminant control)
Lipoxygenase	1	Polyunsaturated fatty acids	Baking (bread whitening and dough strengthening)
Mannase	3	β-mannose, galactomannan, and galactose	Laundry and detergent (reappearing stain removal) and animal feed additive
Naringinase	3	Naringin (flavanone-7-O-glycoside)	Food
Nitrilase	3	Nitriles	Organic synthesis (synthesis of enantiopure carboxylic acids)
Oxidoreductases	1	Various compounds: catalyzing the removal of hydrogen atoms and electrons	Food and detergents
Pectinase	3	Pectin	Food (clarification, mashing, and de-pectinization of fruit-based products) and textile (scouring)

TABLE I. (*Continued.*)

Enzyme	IUB classes	Substrate classes	Industry field (application)
Pectin methyl esterase	3	Pectin	Food (firming fruit-based products)
Penicillin amidase	3	Penicillin	Pharmaceutical
Phytase	3	Phytic acid	Animal feed (phytate digestibility-phosphorous release)
Peroxidase	1	Hydrogen peroxide and hydroperoxides	Textile (removal of excess dye) and personal care (antimicrobial)
Protease	3	Protein and peptide	Laundry and detergents (protein stain removal), baking, food (cheese making, milk clotting, low allergenic infant products, flavor, and brewing for clarification-low calorie beer), leather (de-hiding), pharmaceutical (treatment of blood clot), textile (unhearing and bating), paper and pulp (biofilm removal), and fuel (yeast nutrition)
Pullulanase	3	Polysaccharide	Baking, Starch, and Fuel (saccharification)
Rennin	3	Protein (k-casein)	Food (precipitation and curd formation in cheese-making)
Subtilisin	3	Protein and peptide	Pharmaceutical, laundry, and detergents
Transglutaminase	2	Protein and peptide	Baking (laminated dough strengths) and food (modification of visco-elastic properties)
Xylanase	3	Polysaccharide β-1,4-xylan	Baking (conditioning of dough), paper and pulp (bleach boosting), animal feed (digestibility improvement), and starch and fuel (viscosity reduction)

B. Cellulases

Cellulose is a renewable resource with great potential for bioconversion to value-added bioproducts. Cellulose can be degraded by cellulases produced by cellulolytic bacteria. Cellulases are among some of the most important industrial enzymes known to date.[8] For instance, they convert cellulose to sugars that are suitable for human consumption. On a large scale, these sugars can in turn be fermented to generate bioethanol and biobased products.[8] Cellulases also find application in the textile industry, where they are used for the polishing of fabrics, and in the laundry detergent industry.[9] A cellulase from *Streptomyces thermoviolaceus* with high thermal and pH stabilities has been shown to be more active than other commercial cellulases in the presence of detergents.[10]

C. Xylanases

Xylanases play a key role in the enzymatic depolymerization of hemicellulose to yield monomeric sugars. Traditionally, these enzymes are used in food and paper industries. In recent years, they have received growing attention for the production of sugars from lignocelluloses. Xylanases are also used for the bleaching of rice straw pulp.[11] Xylanases from *Actinomadura* sp. *FC7* and *Nonomuraea flexuosa* have been shown to have high thermostability.[12,13] Owing to their high thermal and pH stabilities, fused xylanases from fungi and actinomycetes have been used in paper and pulp industries.[14] *Streptomyces* spp. are able to produce high levels of xylanase and provide efficient biobleaching.[11] Similarly, they are able to hydrolyze straw waste and produce biogas.[15]

D. Amylases

Common applications for this family of enzymes are in bakery, brewing, and alcohol industries, where thermophilic and acidophilic amylases from *Streptomyces erumpens* are utilized.[16,17] Thermostable amylases from *Nocardiopsis* sp. are also used in bakery and paper industries.[18] The amylase from *Thermobifida* sp. is used for the production of maltotriose from starch.[19]

Moreover, in the starch processing industry, a number of other end product-specific amylases are commonly used for the synthesis of different maltooligosaccharides.[20] Finally, several actinomycetes are the source of cold-active α-amylases that find application in textile, detergent, and bioethanol industries.[21]

E. Proteases

Proteases are utilized in the dairy industry for the manufacturing of cheese. Due to its high specificity, calf rennin has generally been the protease of choice in cheese-making. However, rennin is gradually giving way to microbial proteases from microorganisms such as *Mucor miehei*, *Bacillus subtilis*, *Endothia parasitica*, and *Aspergillus oryzae MTCC 5341*. Among proteases, aminopeptidase hydrolyzes amino acid residues from the N-terminal portions of proteins. Aminopeptidases have a wide range of applications in various fields such as the pharmaceutical industry, where they represent an important molecular tool for protein sequence analysis,[22] and the food industry, for flavor enhancement.[23,27] When combined with other proteases, they lead to a complete degradation of proteins, such as casein, gluten, collagen, and gelatin, helping in nutrient utilization.[24]

F. Lipases

Lipases hydrolyze long chain triglycerides to form diglycerides, monoglycerides, fatty acids, and glycerol.[25,31] Besides their ability to hydrolyze carboxylic ester bonds, lipases can catalyze esterification reactions in non-aqueous media. Lipases find application in food, detergent, pharmaceutical, leather, textile, cosmetic, and paper industries.[26,27] In the food industry, lipases are used for fat and oil processing. Interestingly, lipases from different organisms provide different positional specificities, fatty acid specificities, thermal stabilities, and optimum pH values.[26] Detergent formulations also include lipases, which are of great help for the removal of lipid stains, fatty food stains, and sebum from fabrics. Alkaline yeast lipases can work at lower temperatures than bacterial and mold lipases. Cold-active lipases are used as components of detergents for cold washing, with clear advantages in both energy consumption and textile durability.

G. Laccases

Laccases are blue multicopper oxidases that participate in the degradation of polymers and ring cleavage of aromatic compounds. Owing to their ability to oxidize lignin-related compounds and highly resistant environmental pollutants, they are used in several biotechnological processes such as for instance wastewater treatment and detoxification. Their typical substrates are amines and phenols. These enzymes are also used as medical diagnostic tools and biosensors, in biofuel cells, for the bioremediation of herbicide- and pesticide-contaminated soil, as cleaning agents in water purification systems, as catalysts in drug manufacturing, and as ingredients in cosmetics.[28,29]

H. Phytases

Phytases are used both as an animal feed ingredient and in foods to improve plant phosphorus uptake by animals.[30] Phytases allow the release of phosphorus from plant feedstuffs, where about 2/3 of phosphorus is stored as phytate. Hydrolysis of phytate blocks the translocation of phosphorous into the soil, where it causes eutrophication. In the food industry, phytases are utilized to remove phytic acid. These enzymes are found in many bacteria, yeasts and fungi. New fungal phytases showing high specificity or thermostability have also been identified.[31]

II. THERMOSTABLE ENZYMES IN NATURE

A. Extremozymes

Thermophiles are organisms that have evolved to strive in extreme conditions such as temperatures ranging up to $120\,°C$, high pressure values (up to $250\,atm$), or extreme pH or salt

conditions (up to 5% of NaCl).[32] Their cellular components are also thermostable, including their enzymes, sometimes referred to as extremozymes, which are known to withstand high temperature and extremely acidic and alkaline conditions, and they generally exhibit increased resistance to denaturation[33] and proteolysis.[34] Thermophiles have long been considered of high industrial importance for their possible use in many technological processes, either as intact organisms or as a source of thermostable enzymes that can catalyze specific reactions at high temperatures. However, the majority of enzymes that are currently used in industry are obtained from fungi or mesophilic bacteria. To date, only a few extremozymes have been used for industrial applications, mostly involving DNA polymerase. Today, beside their use in DNA replication, new challenges have broadened their range of successful utilization. One of the major problems in the use of functional extremozymes is the establishment of proper and fine-regulated production conditions such as hosts, efficient transformation approaches, and adequate expression vectors. Due to differences in codon usage with respect to commonly used expression systems such as *Escherichia coli* or *Bacillus* sp., only a few systems have been successfully utilized, mostly members of the genus *Thermus* or the hyperthermophilic species *Sulfolobus solfataricus*.[35,36]

B. Thermophiles used in industry

There is hardly any example that can represent the impact of thermophiles in all aspects of our daily life better than the thermostable DNA polymerases that are used in the polymerase chain reaction (PCR). Nearly two and a half decades after the Nobel prize in chemistry given to Mullis and Smith in recognition of their pioneering work on the development of the technique, biomedical and biotechnology research have advanced far beyond levels that would have been unimaginable without it. Thermophilic organisms such as *Thermus aquaticus, Pyrococcus furiosus, and Thermococcus litoralis* have provided those stable and proofreading forms of DNA polymerases (usually referred to as Taq, Pfu, and Vent) that survive beyond the denaturation temperature of long DNA fragments and, hence, have allowed the technique to become routinely and efficiently used in all the molecular biology laboratories around the world.

A further example of the importance of extremophilic enzymes for industrial applications can be found in the starch field. The standard process for starch conversion into single glucose units occurs in two steps: (1) liquefaction of the raw starch granules followed by (2) saccharification.[37] The liquefaction of the raw starch granules is achieved through a necessary heating step (105 °C for 5 min and 95 °C for 1 h at pH 6.0) to facilitate liquefaction, and then, saccharification is done at 60 °C for 3 h at pH 4.5.[38] Currently, the key enzymes that are used for the production of glucose from starch are typically a bacterial amylase and a fungal glucoamylase combined with a pullulanase.[39] Since these enzymes are not active at high temperatures and low pH values (as needed in the second step of the process), cooling and pH adjustment is mandatory. This energy and time consuming procedure has been optimized by the use of more suitable extremophilic amylolytic enzymes.[40] The first archaeal amylase with an optimum temperature of about 100 °C and residual activity at 130 °C was found in *P. furiosus*, and it was characterized in 1990.[41] Recently, an acid-stable amylase with a half-life of 30 min at 80 °C was described.[42] To date, one of the most heat-active pullulanase, having an optimum temperature of 100 °C, has been discovered in *Thermococcus kodakarensis KOD1*.[43] One of the most heat-active starch-degrading enzymes known to date is an α-amylase from *Methanococcus jannaschii* (optimum temperature = 120 °C).[44]

Examples of the most common enzymes that exist in thermophilic microorganisms and that, due to their higher thermal stability compared to their mesophilic homologues, are commonly used in high temperature biotechnology processes include cellulases, amylases, xylanases, lipases, proteases, pectinases, and esterases. In recent years, the use of extremophiles as cellular biocatalysts for biotransformation, in particular for biofuel production, has attracted growing interest owing to the accelerated reaction rates, the reduced energy input, and the low contamination risk that are associated with the process. Moreover, their ability to exploit different carbohydrate sources (such as starch or hemicellulose) further favors the use of extremophilic species

in industrial bioprocesses. In this field, the current main focus is on the use of starch- and lignocellulose-degrading enzymes for the production of next-generation biofuels.

C. Nature's strategies to achieve thermostability

Several attempts to identify those key natural features, either at the sequence or at the structural level, that provide all proteins with their signature thermostability profile have failed to paint a clear picture and to define first principles of universal validity and general applicability. At most, consistent trends have been described when comparing different members of specific protein families, where a medium-to-high degree of sequence and structure similarity is often observed. In general, thermophilic proteins are mostly made of both hydrophobic and charged residues, showing a smaller proportion of uncharged polar residues compared to mesophilic proteins. Hydrophobic residues, which usually cluster in the core of the protein to minimize solvent exposure, form stabilizing van der Waals interactions with other hydrophobic residues, possibly exploiting the lack of directionality that is associated with these contacts and therefore adopting the conformation that minimizes voids in the structure while maximizing the contact surface area. Accordingly, thermophilic proteins usually have a higher than average content of valine (Val) and isoleucine (Ile) residues compared to mesophilic proteins.[45,46] On the other hand, properly oriented charged residues can allow the formation of stabilizing salt bridges, whose energy contribution to the stabilization of the protein fold often exceeds that of hydrophobic interactions. The stabilization energy provided by a single salt bridge is usually estimated to be in the order of 3–5 kcal/mol. The need for a proper orientation of charged residues to form stabilizing electrostatic interactions and, at the same time, to avoid the clustering of identical charges highlights the limitations that are intrinsic in any thermostability analysis approach that focuses only on residue counts and distribution within primary sequences without considering structure level comparisons between proteins. As previously mentioned, the tendency of hydrophobic residues to cluster and shield themselves from the aqueous environment is one of the main driving forces behind the folding and the thermal stability of a protein. Yet, within the class of hydrophobic residues, further stabilization is often provided when aromatic interactions (such as π-π, cation-π, and S/π interactions) can be formed, suggesting a complex interplay between the size, chemical nature, and electronic structure of the hydrophobic residues involved. Disulfide bond formation resulting from two spatially closed cysteines is also an important driving force in protein folding. In fact, disulfide bridges are common in nature. The strength of the covalent bond that is formed between two oxidized cysteines clearly provides a great deal of energy stabilization to the fold of the protein, either locally or globally. Interestingly, Fitter and colleagues[47] analyzed structural fluctuations of α-amylases from both mesophilic and thermophilic organisms, demonstrating that in the thermophilic protein, the folded state is more structurally flexible than the folded state of the mesophilic homologue. In contrast to the general rule that suggests that higher structural rigidity corresponds to higher thermostability,[48] in this case, higher thermostability could be the result of entropic stabilization. Altogether, these different strategies adopted by extremophiles to achieve their superior stability can be used to guide the design and engineering of novel enzymes.

III. ENZYME STABILIZATION THROUGH DIRECTED EVOLUTION

A. Directed evolution

Directed evolution has become a fundamental strategy in protein engineering for the production of more powerful and efficient biocatalysts.[49–54] The process is similar to natural evolution, albeit on a reduced timescale. In fact, directed evolution generates, and selects under a specific evolutionary pressure, enzymes with novel or improved features through an iterative process characterized by several rounds of mutagenesis and screening, starting from a parental protein. Generally, the creation of random mutants from a parental protein is done by error prone mutagenesis,[55–57] DNA shuffling,[58–60] site-saturation mutagenesis,[61,62] chemical mutagenesis,[63,64] or using different mutator strains.[65–67] However, all these methods generally require

reiterative manipulation of single genes and are not used for parallel and continuous directed evolution of gene networks or genomes. In this regard, Wang, Isaacs, and colleagues developed the so-called Multiplex Automated Genome Engineering (MAGE) approach that, by coupling parallel DNA synthesis and recombination in a single *E. coli* cell or across a population of cells, leads to the generation of multiple modifications (mismatches, insertions, and deletions), from single point mutations to the genome level.[68,69] Due to the large size of the resulting libraries, the ability to identify and isolate those mutants that feature the desired properties is a critical success factor in a directed evolution campaign. To face this burden, different strategies are adopted for library analysis and can be divided into two main methods: screening and selection. While the selection approach directly identifies the desired mutant, eliminating unproductive mutations based on a direct connection between cell growth and an optimized or acquired enzyme function, in a screening method, all the resulting variants from the library are individually tested for the desired function, minimizing the risk of false negatives. However, the weakness of this method lies in the reduced number of mutants that can be evaluated, which makes automation a necessary and key element for a rapid high-throughput screening approach. Microtitre plate- and agar plate-based screening procedures are the two most commonly adopted library screening formats. Compared to the latter, microtitre plate-based assays are suitable for handling large libraries of mutants and use a smaller quantity of the sample. Generally, an enzyme activity assay is performed on crude lysate from single colonies, or on purified proteins, into multi-well plates (e.g., from 96 to 9600-well). The use of classical colorimetric or fluorescence assays allows the verification of any small improvements in the desired function or property of the enzyme.[70–73] However, the use of these assays is restrained only to those cases where substrates, cofactors, or products are suitable for UV-vis absorbance or fluorescence measurements. In order to overcome this limitation, a variety of analytical techniques can be implemented within a high-throughput format. Examples include automated systems using on-line high performance liquid chromatography (HPLC)[74] or mass spectrometry (MC)[75] for direct product quantitation. Agar plate-based screening methods focus on the direct correlation between the growth of the host organism on selective agar plates for the screening of a specific enzyme function.[56,76,77] Here, colonies displaying a desired color within a certain time window are selected as active. Despite the outstanding performance of this approach for the direct detection of the desired mutant, its lower sensibility in the estimation of the enzymatic activity represents its main limitation compared to a microtitre plate-based screening method. Therefore, in order to identify significantly improved enzyme variants, it is common to couple an agar plate assay as a primary screen with a secondary activity assay, generally set up in a 96-well microtiter plate and based on a biochemical assay.[78] Microtitre plate- and agar plate-based screens are powerful methods for library analysis; however, relative to the size of complete libraries (10^8–10^9 mutants), their efficiency is relatively low ($\sim 10^4$/day for the microtitre plate and $\sim 10^5$/day for the agar plate). Interestingly, fluorescence-activated cell sorting (FACS) provides a high-throughput approach directly on a cell population ($\sim 10^8$/day), for the direct separation of mutants expressing the desired protein, where the link between the genotype (mutation) and the phenotype (desired property) is maintained through a fluorescence assay that detects those cells that display that particular enzymatic activity.[79] Different strategies can be adopted, generally based on the detection of the product either inside the cell (product entrapment)[80] or onto the surface of cells (cell surface display).[81] Both these methods can be classified as *in vivo* methods. A FACS-based *in vitro* approach, usually referred to as *in vitro* compartmentalization, involves compartmentalization of the gene encoding the mutant protein in small aqueous droplets, together with the fluorescent product of interest.[82] In practice, a directed evolution campaign can be separated into three well-defined steps: the identification of a good starting parental protein, the creation of a library of mutants, and the screening or selection step, based on the artificial selection imposed to identify improved mutants that carry out a specific function. This entire process of mutant generation and isolation is repeated until the desired change is observed, and it can be iterated over the resulting mutant from a previous cycle until no further change is elicited (Fig. 1). It is generally believed that the use of a consensus sequence that is either based on a protein family of interest or closer to the protein
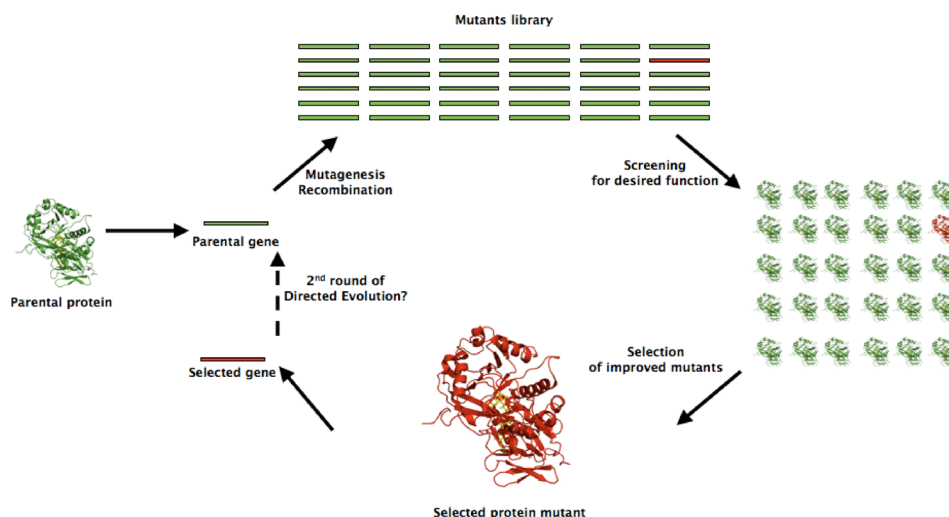
FIG. 1. Directed evolution campaign. An iterative mutagenesis and screening process is performed starting from the gene coding for the enzyme of interest. From the generated mutant library, mutants are screened for the desired function or property. The best performing mutant can then be used as the parental gene for the next iterative rounds of mutagenesis.

ancestor will likely define the most stable starting scaffold to be used in a directed evolution campaign, and it will be more tolerant to the deleterious effects of mutations or insertions. Indeed, protein families that are characterized by substrate ambiguity (i.e., are active on a wider range of substrates) or catalytic promiscuity (i.e., catalyze different types of reactions) are more suitable for a directed evolution campaign.[83–88] Clearly, directed evolution represents a powerful and effective approach for reshaping the basic characteristics of enzymes or for designing *de novo* enzymes in order to improve not only their catalytic features (e.g., optimization of kinetic parameters) but also their stability (e.g., thermal and pH stabilities) or protein yield production for subsequent industrial applications.

B. Thermostable enzymes by directed evolution

Several examples of thermal stabilization obtained by directed evolution can be found in the literature. For instance, improvements in the thermostability of galacto-N-biose/lacto-N-biose I phosphorylase (GLNBP) from *Bifidobacterium longum* JM1217 were achieved by directed evolution.[89] From the initial library of GLNBP mutants, two single mutants were selected, showing each a significant improvement relative to the wild type ($\sim$10 °C). Based on these results, the corresponding double mutant was generated and shown to exhibit a 20 °C higher thermostability than the wild type, allowing its use for the industrial production of LNB at high temperature, thus resulting in the shortening of the reaction time and in the prevention of microbial contamination. A study reporting the enhancement in the thermostability of an amylase from *Thermus* sp. strain IM6501 (ThMA) well represents the compromises that can be faced in a directed evolution experiment, showing how an improvement in thermostability can affect catalytic efficiency.[90,91] Indeed, the resulting thermostable mutant ThMA, which features a total of seven single mutations, exhibits a 15 °C increase in the optimal reaction temperature relative to the wild type enzyme. However, one of the mutations reduces the activity of the enzyme by 23% relative to the wild type form, still preserving significant thermoresistance.[92] Another example of thermoresistance improvement by directed evolution for industrial applications comes from the endo-β-1,4-xylanase (XynA) from *Thermomyces lanuginosus*.[93] Here, based on a first campaign of directed evolution, four mutants, which were selected based on the exhibited improvement of activity and stability, were subjected to further rounds of mutagenesis. The majority of the resultant mutants exhibited the expected compromise between stability and activity, and only one of these second generation mutants showed a significant improvement in activity and stability relative to the wild type enzyme. It is worth stressing that directed

evolution experiments can target not only high thermostability improvements, as described above. For instance, aggregation can significantly influence the yield and the biological activity of biopharmaceutical products.[94] For example, antibody stabilization is a critical issue for industry. Aggregation-resistant antibodies, showing a 2–3 °C increase compared to the wild type, have been developed by directed evolution in order to improve antigen-binding fragment (Fab) stability.[95] Another example involving a combination of rational design and directed evolution is that of an extremely stable green fluorescent protein, eCGP123, which was created by a consensus engineering approach or consensus green protein (CGP).[96] The process consisted of a recursive iteration involving the sequential introduction of three destabilizing heterologous inserts, a sequential mutagenesis step to overcome the destabilization, and the final removal of the destabilizing insert from the mutated gene.[97]

IV. ENZYME STABILIZATION THROUGH RATIONAL DESIGN

The vast majority of mutations in a natural enzyme destabilize.[98] Nonetheless, a naturally occurring enzyme is not necessarily the most stable form that is possible for that specific enzyme. Indeed, some mutations can further stabilize a protein, increasing the equilibrium population of the folded state.[83,99] In a directed evolution campaign, whereby random mutations are introduced in the coding gene and the large resulting libraries (typically consisting of more than 10^5 variants) are screened for a specific function, mutant enzymes featuring improved thermostability can most likely be identified.[100,101] Unfortunately, such a random approach is possible only with enzymes for which fast activity screens can be implemented. While directed evolution has yielded improved enzyme variants, it is a time-consuming and highly labor-intensive method, often leading to a dead-end, where further introduction of function-altering mutations is limited by low enzyme stability.[102]

A. Rational design

The modification of the properties of an enzyme could be achieved through the use of different rational approaches. Over the past few years, computational design has been successfully applied for the thermostabilization of noncatalytic proteins, the first notable examples being represented by the computational stabilization of a cytokine analog in 2002 by Luo and colleagues.[103] However, the stabilization of an enzyme presents additional challenges. Indeed, the geometry of the active site and the protein dynamic behavior during an enzymatic reaction are often crucial for providing the maximum catalytic efficiency. Therefore, rational design methods need to be able to predict stabilizing mutations in the context of a given fold and, at the same time, minimize any change in the backbone conformation that might disrupt the structure of the active site or reduce its flexibility. Recently, several different *in silico* methods to establish the effect of mutations on the stability of a protein have been developed.[104,105] Nonetheless, the reliability of these approaches is still unsatisfactory.[106] The techniques used to increase protein stability with a rational approach are often based on one or multiple methods including phylogenetic analysis,[107] comparison to homologous proteins (and particularly thermophiles),[108] optimization of charged interactions (salt bridges and hydrogen bonds), optimization of residues and loops showing unfavorable Ramachandran angles and high B-factors,[109] methods based on the calculation of free energies,[110] and structure-based computational design.[111,112]

Several rational design strategies can be implemented to improve enzyme thermal resistance (see Fig. 2). One of the most common methods used since the dawn of protein engineering is the introduction of disulfide bridges,[113,114] which usually provide considerable stability to proteins by locking their fold in a well-defined local or global conformation.[115,116] Other common strategies involve the introduction of surface hydrogen bonds[117] and salt-bridges[118] that enhance protein stability by increasing protein rigidity and decreasing free energy. Indeed, an increased number of surface hydrogen bonds and salt bridges are often observed in thermophilic proteins, where they contribute to thermal stability. In mobile loops, the introduction of prolines[119] provides an increase in stability by reducing the entropy of the denatured state. Further rational engineering strategies focus on the design of well-packed hydrophobic cores,[111]
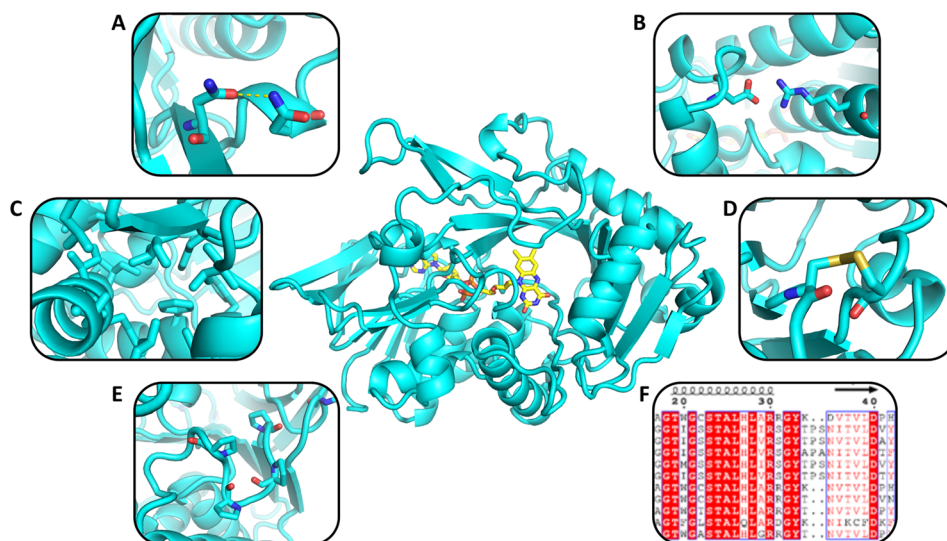
FIG. 2. Stabilizing strategies used in rational enzyme design. The most common strategies involve the introduction of surface hydrogen bonds (a) and salt bridges (b), the stabilization of the hydrophobic core (c), the introduction of disulfide bridges (d), and the stabilization of mobile loops using prolines (e). Phylogenetic analysis (f) can be used alone or in combination with previous strategies to guide the enzyme rational design process.

which play a central role in preserving enzyme stability and conformational specificity. Finally, phylogenetic analysis is often used to guide rational design.[120] Specifically, this kind of analysis helps in disfavoring uncommon residues for a specific position or favoring common ones (especially those found in thermophiles).

B. Computational screening

Proteins may often acquire an improved stability via the introduction of several mutations. However, sometimes, a single severely destabilizing mutation is sufficient to totally disrupt a protein fold, even in the presence of several stabilizing ones. Therefore, high prediction accuracy is essential. Unfortunately, all the existing methods have a relatively high probability of introducing unfavorable mutations.[111,121] Therefore, protein stabilization can more safely proceed only via the incorporation of a very limited number of predicted stabilizing mutations at each mutagenesis step. Alternatively, phylogenetic libraries can be used to identify optimal combinations of stabilizing mutations.[122–124] However, both these approaches are labor-intensive and applicable only to those proteins for which well-established medium-to-high throughput screens are available. Recent methods include computational high-throughput screening methods to evaluate libraries of potentially stabilizing mutations.[125,126]

C. Notable achievements

The earliest notable example of rational design for enzyme thermostabilization is the work of Baker and coworkers in 2005,[127] in which they used Rosetta to stabilize a cytosine deaminase, an enzyme with potential use for antitumoral strategies. In this work, an energy function was used to evaluate target sequences threaded onto a fixed backbone. Using an iterative heuristic procedure followed by an energy evaluation step, a comprehensive search of the space sequence was done. While the adoption of those sequences associated with the lower energy values was automatic, the higher energy sequences were assigned a probability based on the Rosetta score. The authors identified three mutations that, when combined, produced a remarkable increase in enzyme stability. Following this seminal work, several other researchers introduced various computational approaches aimed at improving the stability of enzymes that are relevant for different industrial applications. Table II summarizes the key achievements in the field.

TABLE II. Notable examples of rational design of enzymes with industrial applications.

Enzyme	Application	Method	Stabilization achieved	References
Cytosine deaminase	Possible antitumoral	Rosetta	T_m increases up to 10 °C	Korkegian[127]
Glucose dehydrogenase	Commodity chemical biosynthesis	Structure-guided consensus analysis	T_m increases up to 34 °C	Vázquez-Figueroa *et al.*[128]
Proteinase K	Molecular biology	Phylogenetic analysis and machine learning design	20X half-life at 68 °C	Liao *et al.*[129]
Cocaine esterase	Pharmaceutical industry	Molecular d and Rosetta	30X half-life at 37 °C	Gao *et al.*[130]
Xylanase	Production of paper	Flexible region stabilization and Rosetta	15X half-life at 50 °C	Joo *et al.*[131]
Lipase	Detergents, food, bioenergy, and pharmaceuticals	B-factor analysis and Rosetta	T_m increases by 2 °C	Kim *et al.*[132]
Terpene synthase	Production of terpenoids	Statistical, computationally assisted design strategy (SCADS) algorithm	T_m increases up to 40 °C	Diaz *et al.*[133]
Lipase	Detergents, food, bioenergy, and pharmaceuticals	Consensus analysis	2X half-life in organic solvents	Park *et al.*[134]
Lipase	Detergents, food, bioenergy, and pharmaceuticals	Disulfide by Design code	T_m increases by 7 °C	Yu *et al.*[135]
Methyl parathion hydrolase	Detoxification of pesticides	Unfolding free energy (Prethermut code)	T_m increases by 12 °C, and T_{50} increases by 10 °C	Tian *et al.*[136]
Pullulanase (α-amylase)	Production of high-glucose syrup	Structure-guided consensus analysis	4.3X half-life at 60 °C	Duan *et al.*[137]
Limonene epoxide hydrolase	Production of chiral building blocks	Rosetta, Dynamic Disulfide Discovery, and Molecular Dynamics	T_m increases up to 35 °C and increased activity	Wijma *et al.*[126]
Cellulase	Cellulose degradation	Salt bridge design	T_m increases by 16 °C	Lee *et al.*[138]
3-dehydroshikimate dehydratase	Commodity chemical biosynthesis	Visual inspection, Rosetta, and void identification and packing (VIP) server	10X higher half-life at 37 °C and increased expression	Harrington *et al.*[139]
Phytase	Phytate degradation	Disulfide by Design code	3X higher half-life at 60 °C	Tan *et al.*[140]
Cutinase	Polymer degradation	Rosetta	T_m increases by 6 °C, 10X higher half-life at 60 °C	Shirke *et al.*[141]
Acetylcholinesterase	Detoxification of pesticides and nerve agents	Rosetta and Phylogenetic analysis	T_m increases by 20 °C and increased expression	Goldenzweig *et al.*[125]
Transketolase	Synthesis of complex carbohydrates	Rosetta and consensus analysis	T_m increases by 5 °C and increased k_{cat}	Yu *et al.*[142]

D. Rational enzyme design algorithms

In the past few years, several procedures have been developed to design thermostable enzymes. Two of them appear to be particularly promising given their general applicability, integration of orthogonal methods, and ease of use.

E. FRESCO

The "Framework for Rapid Enzyme Stabilization by Computational libraries" (FRESCO),[126] developed by the Janssen group at the University of Groningen, follows five major steps. In the first step (1), a library of single point mutations is generated using Rosetta, FoldX, and an in-house code for disulfide discovery, eventually excluding the catalytic region from mutable residues. The mutations are retained in the library if they provide a stabilizing effect based on internal scoring. The second step (2) involves the elimination of unreasonable mutations upon visual inspection. The major reasons for elimination are hydrophobic side chains exposed to solvent or proline residues introduced inside an α-helix. The third step (3) involves the screening of the mutant library with molecular dynamics (MD) simulations and then filtering out mutations that lead to an increase in root mean square fluctuations. The assumption for this step is that mutations that increase structural flexibility with respect to the wild type are likely to be destabilizing. After the MD-based screening, in step (4), surviving mutations are experimentally tested to verify that they provide an actual increase in melting temperature while preserving catalytic activity. At this stage, a library of experimentally validated stabilizing single-point mutations is obtained. In step (5), the validated single-point mutations are combined, screened with MD simulations and then experimentally validated, providing the final stabilized variant(s) of the enzyme.

The two major advantages of FRESCO are the use of different strategies and methods to generate the initial library of mutations and the use of orthogonal methods to filter them (i.e., Rosetta/FoldX scoring followed by MD). On the other hand, usability and automation are limited, given that step (2) involves visual inspection. A second possible drawback is that the initial stage includes the screening of only single-point mutations. Indeed, it is possible that two single-point mutations, when considered alone, are found to be destabilizing, whereas, if combined, they could generate stabilizing interactions.

F. PROSS

A second promising framework for enzyme stabilization is the "Protein Repair One Stop Shop" (PROSS)[125] developed by the Fleishman Lab at Weizmann Institute of Science. In the first step (1), the algorithm performs a sequence alignment with homologous proteins. Rather than selecting the most promising mutations for each position, this step is performed to eliminate the mutations that are rare or not observed. The second step (2) involves the use of Rosetta to evaluate potentially stabilizing mutations, selecting only those that provide an energy decrease with respect to the wild type. Finally, in the third step (3), the single point mutations that have been selected in the previous step are combined and the enzyme variants are ranked based on their energy score.

A major advantage of PROSS is its ease of use, either reproducing the algorithm in-house or employing the web-server developed by the authors (http://pross.weizmann.ac.il). On the other hand, a possible major limitation of this approach is the lack of orthogonal methods (e.g., MD) in combination with Rosetta. A further limitation is the need for several homologs of the target protein in order to increase the reliability of the first step of the algorithm.

G. Outlook for rational enzyme engineering

In the last 60 years, protein science has moved from the first success in protein structure determination (the Myoglobin X-ray crystal structure in 1958)[143] to the early attempts at protein engineering (the design of a reduced ribonuclease in 1979).[144] The introduction of
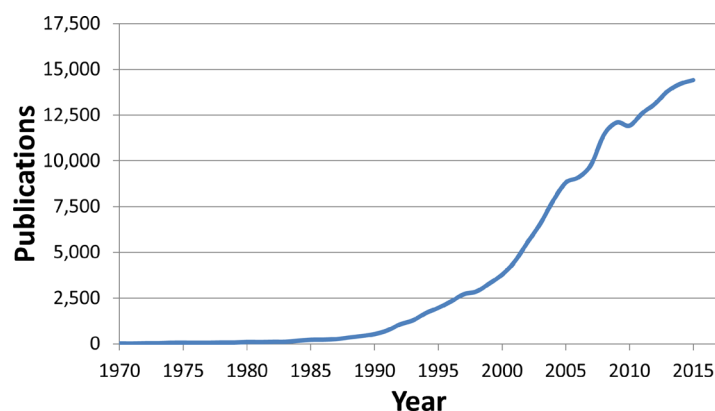
FIG. 3. Number of publications per year on the topic of "protein design" (source: Scopus).

computational analysis methods led to the *de novo* design of a protein motif in 1997,[145] boosting our ability to engineer proteins (see Fig. 3). The possibility of building new proteins, tuning enzyme catalytic activity, and extending their thermal stability is now increasingly exploited for different industrial applications. Protein design is often termed "the inverse folding problem" because when our ability to build or modify a protein in a predictable way will finally be attained, which will provide an indirect evidence that the protein folding problem is well-understood. Therefore, the methods and successes of enzyme engineering contribute directly to those of protein structure prediction.

A general major limitation of current computational enzyme design approaches is the lack of an objective assessment of the different available methods, similar to the one used in the Critical Assessment of Structure Prediction (CASP) competition.[146] Within the CASP challenge, research groups have the opportunity to test their structure prediction methods by identifying a protein structure from its amino acid sequence. The assessors are not part of the competitors, thus enabling objective analysis of achievements and challenges in a comparative manner. Without a similar community-wide objective assessment, the comparative analysis of computational design methods necessarily relies upon reports by the respective authors of each method, thus hampering the identification of advantages and disadvantages of each tool and the cross-dissemination of knowledge.

A further major challenge in the field of enzyme engineering is the accessibility to the general scientific community. Until now, computational enzyme design lacked standardization and solid reliability of results. As such, it has not been carried out by a large community, but rather it has clustered in a small number of labs leading the field. Most often, these are the laboratories where software packages are developed. As with other fields, it is expected that with time, more and more scientists will apply computational enzyme design, and rational design will become a standardized and common procedure carried out on a routine basis in biochemical laboratories.

ACKNOWLEDGMENTS

[1]M. Leisola, J. Jokela, O. Pastinen, O. Turunen, and H. Schoemaker, *Encyclopedia of Life Support Systems (EOLSS). Physiology and maintenance. Vol. II. Industrial use of enzymes* (UNESCO-EOLSS), available from: http://www.eolss.net/sample-chapters/ (Accessed: 4th February 2017).

[2]D. Prakash, N. Nawani, M. Prakash, M. Bodas, A. Mandal, M. Khetmalas, and B. Kapadnis, Biomed Res. Int. **2013**, 264020.

[3]See http://www.sbcs.qmul.ac.uk/iubmb/ for IUBMB and IUPAC biochemical and organic nomenclature recommendations, World Wide Web material prepared by G. P. Moss.

[4]Y. Galante and C. Formantici, Curr. Org. Chem. **7**, 1399 (2003).

[5]See   https://www.bccresearch.com/pressroom/bio/global-Market-Industrial-Enzymes-Reach-Nearly-\$7.1-Billion-2018 for "BBC research" (2017).

[6]See   http://www.marketsandmarkets.com/PressReleases/protein-Hydrolysis-Enzymes.asp   for "Markets and Markets" (2017).

[7]A. I. Solbak, T. H. Richardson, R. T. McCann, K. A. Kline, F. Bartnek, G. Tomlinson, X. Tan, L. Parra-Gessert, G. J. Frey, M. Podar, P. Luginbuhl, K. A. Gray, E. J. Mathur, D. E. Robertson, M. J. Burk, G. P. Hazlewood, J. M. Short, and J. Kerovuo, J. Biol. Chem. **280**, 9431 (2005).

[8]S. Sadhu, Br. Microbiol. Res. J. **3**, 235 (2013).

[9]J. Hill, E. Nelson, D. Tilman, S. Polasky, and D. Tiffany, Proc. Natl. Acad. Sci. U. S. A. **103**, 11206 (2006).

[10]B. E. Jones, W. A. H. Van Der Kleij, P. Van Solingen, and W. Weyler, European patent EP1034280 A2 (September 2000).

[11]H. M. Rifaat, Z. A. Nagieb, and Y. M. Ahmed, Appl. Ecol. Environ. Res. **4**, 151 (2006).

[12]S. H. Ryszard Brzezinski, C. V. Dery, C. Beaulieu, and J. F. Ethier, U.S. patent 8,759,033 B2 (June 24, 2014).

[13]J. Zhang, M. Siika-aho, T. Puranen, M. Tang, M. Tenkanen, and L. Viikari, Biotechnol. Biofuels **4**, 12 (2011).

[14]D. Haltrich, B. Nidetzky, K. D. Kulbe, W. Steiner, and S. Župančič, Bioresour. Technol. **58**, 137 (1996).

[15]P. B. Sathya, T. Stalin, and K. Selvam, Afr. J. Biotechnol. **11**, 14320 (2012), available at http://www.academicjournals.org/journal/AJB/article-stat/9E48D2F30299.

[16]Y. Ben Ammar, T. Matsubara, K. Ito, M. Iizuka, T. Limpaseni, P. Pongsawasdi, and N. Minamiura, J. Biochem. **35**, 568–575 (2002).

[17]S. Kar and R. C. Ray, Pol. J. Microbiol. **57**(1), 49–57 (2008).

[18]T. L. M. Stamford, N. P. Stamford, L. C. B. B. Coelho, and J. M. Araújo, Bioresour. Technol. **76**, 137 (2001).

[19]C. H. Yang and W. H. Liu, Enzyme Microb. Technol. **35**, 254 (2004).

[20]B. A. Kikani and S. P. Singh, Int. J. Biol. Macromol. **48**, 676 (2011).

[21]M. Kuddus, Roohi, J. M. Arif, and P. W. Ramteke, Biotechnology **10**, 246 (2011).

[22]Y. T. Wu, N. Di Zhou, Z. M. Zhou, X. X. Gao, and Y. P. Tian, J. Basic Microbiol. **54**, 1110 (2014).

[23]X. Gao, W. Cui, N. Ding, Z. Liu, Y. Tian, and Z. Zhou, Prion **7**, 328 (2013).

[24]Y. Méndez, K. Pérez-Labrada, J. González-Bacerio, G. Valdés, M. de los Chávez, J. Osuna, J. L. Charli, I. Pascual, and D. G. Rivera, ChemMedChem **9**, 2351 (2014).

[25]A. Houde, A. Kademi, and D. Leblanc, Appl. Biochem. Biotechnol. **118**, 155 (2004).

[26]F. Hasan, A. A. Shah, and A. Hameed, Enzyme Microb. Technol. **39**, 235 (2006).

[27]K.-E. Jaeger and T. Eggert, Curr. Opin. Biotechnol. **13**, 390 (2002).

[28]S. Rodríguez Couto and J. L. Toca Herrera, Biotechnol. Adv. **24**, 500 (2006).

[29]Shraddha, R. Shekher, S. Sehgal, M. Kamthania, and A. Kumar, Enzyme Res. **2011**, 1 (2011).

[30]A. Vohra and T. Satyanarayana, Crit. Rev. Biotechnol. **23**, 29 (2003).

[31]S. Haefner, A. Knietsch, E. Scholten, J. Braun, M. Lohscheidt, and O. Zelder, Appl. Microbiol. Biotechnol. **68**, 588 (2005).

[32]M. W. Adams, F. B. Perler, and R. M. Kelly, Nat. Biotechnol. **13**, 662 (1995).

[33]C. Vieille and G. J. Zeikus, Microbiol. Mol. Biol. Rev. **65**, 1 (2001).

[34]S. Kumar and R. Nussinov, Cell. Mol. Life Sci. **58**, 1216 (2001).

[35]S. V. Albers, M. Jonuscheit, S. Dinkelaker, T. Urich, A. Kletzin, R. Tampe, A. J. Driessen, and C. Schleper, Appl. Environ. Microbiol. **72**, 102 (2006).

[36]B. Leis, A. Angelov, and W. Liebl, Adv. Appl. Microbiol. **83**, 1 (2013).

[37]S. Pervez, A. Aman, S. Iqbal, N. N. Siddiqui, and S. A. Ul Qader, BMC Biotechnol. **14**, 49 (2014).

[38]C. Bertoldo and G. Antranikian, Curr. Opin. Chem. Biol. **6**, 151 (2002).

[39]P. Kumar and T. Satyanarayana, Crit. Rev. Biotechnol. **29**, 225 (2009).

[40]S. Elleuche and G. Antranikian, in *Thermophilic Microbes in Environmental and Industrial Biotechnol*ogy (Springer Netherlands, Dordrecht, 2013), pp. 509–533.

[41]R. Koch, P. Zablowski, A. Spreinat, and G. Antranikian, FEMS Microbiol. Lett. **71**, 21 (1990).

[42]A. Sharma and T. Satyanarayana, Extremophiles **16**, 515 (2012).

[43]T. Han, F. Zeng, Z. Li, L. Liu, M. Wei, Q. Guan, X. Liang, Z. Peng, M. Liu, J. Qin, S. Zhang, and B. Jia, Lett. Appl. Microbiol. **57**(4), 336–343 (2013).

[44]J. W. Kim, L. O. Flowers, M. Whiteley, and T. L. Peeples, Folia Microbiol. **46**, 467 (2001).

[45]S. Kumar, C.-J. Tsai, and R. Nussinov, Protein Eng., Des. Sel. **13**, 179 (2000).

[46]S. Chakravarty and R. Varadarajan, FEBS Lett. **470**, 65 (2000).

[47]J. Fitter, R. Herrmann, T. Hauß, R. E. Lechner, and N. A. Dencher, Physica B **301**, 1–7 (2001).

[48]V. L. Arcus, E. J. Prentice, J. K. Hobbs, A. J. Mulholland, M. W. Van Der Kamp, C. R. Pudney, E. J. Parker, and L. A. Schipper, Biochemistry **55**, 1681 (2016).

[49]F. H. Arnold, Cold Spring Harb. Symp. Quant. Biol. **74**, 41–46 (2010).

[50]M. S. Packer and D. R. Liu, Nat. Rev. Genet. **16**, 379 (2015).

[51]V. Tiwari, Front Chem. **4**, 39 (2016).

[52]M. Wang, T. Si, and H. Zhao, Bioresour. Technol. **115**, 117 (2012).

[53]A. Madhavan, R. Sindhu, P. Binod, R. K. Sukumaran, and A. Pandey, Bioresour. Technol. **245**, 1304–1313 (2017).

[54]B. O. B. D. P. Nannemann, W. R. Birmingham, and R. A. Scism, Futur. Med Chem. **3**, 809 (2011).

[55]J. Yang, A. J. Ruff, M. Arlt, and U. Schwaneberg, Biotechnol Bioeng. **114**(9), 1921–1927 (2017).

[56]J. Shim, Y. Kim, T. Kim, H. Chae, J. Park, H. Cha, J. Kim, Y. Kim, T. Schaefer, T. Spendler, T. Moon, and K. Park, Protein Engineering Design & Selections **17**(3), 205–211 (2004).

[57]D. S. Wilson and A. D. Keefe, "Random Mutagenesis by PCR," in Current Protocols in Molecular Biology (Wiley, 2001), Unit 8.3.

[58]I. Axarli, A. W. Muleta, E. G. Chronopoulou, A. C. Papageorgiou, and N. E. Labrou, Biochim. Biophys. Acta **1861**, 3416 (2017).

[59] J. Speck, S. C. Stebel, K. M. Arndt, and K. M. Müller, in *PCR Protocols*, Methods in Molecular Biology (Methods and Protocols) Series Vol. 687, edited by D. Park (Humana Press, 2011), pp. 333–344.

[60] S. I. Lehtonen, B. Taskinen, E. Ojala, S. Kukkurainen, R. Rahikainen, T. A. Riihimäki, O. H. Laitinen, M. S. Kulomaa, V. P. Hytönen, F. Laboratories, and F. Tampere, Protein Eng., Des. Sel. **28**, 23 (2015).

[61] K. D. Belsare, T. Horn, A. J. Ruff, R. Martinez, A. Magnusson, D. Holtmann, J. Schrader, and U. Schwaneberg, Protein Eng., Des. Sel. **30**, 119 (2017).

[62] C. Zhou, T. Saravanan, M. Lorilliere, D. Wei, F. Charmantray, L. Hecquet, W.-D. Fessner, and D. Yi, ChemBioChem **18**, 455–459 (2017).

[63] S. Dutta, B. Basak, B. Bhunia, A. Sinha, and A. Dey, World J. Microbiol. Biotechnol. **33**, 90 (2017).

[64] C. L. Windle, K. J. Simmons, J. R. Ault, C. H. Trinh, A. Nelson, A. R. Pearson, and A. Berry, Proc. Natl. Acad. Sci. U.S.A. **114**(10), 2610–2615 (2017).

[65] A. Greener, M. Callahan, and B. Jerpseth, Methods Mol. Biol. **57**, 375–385 (1996).

[66] G. Muteeb and R. Sen, in *In Vitro Mutagenesis Protocols*. Methods in Molecular Biology (Methods and Protocols) Series Vol. 634, edited by J. Braman (Humana Press, 2010), pp. 411–419.

[67] Y. Terao, K. Miyamoto, and H. Ohta, Appl. Microbiol. Biotechnol. **73**(3), 647–653 (2006).

[68] H. H. Wang, F. J. Isaacs, P. A. Carr, Z. Z. Sun, G. Xu, C. R. Forest, and G. M. Church, Nature **460**, 894 (2009).

[69] R. R. Gallagher, Z. Li, A. O. Lewis, and F. J. Isaacs, Nat. Protoc. **9**, 2301 (2014).

[70] R. M. Kelly, H. Leemhuis, J. Rozeboom, N. Van Oosterwijk, and B. W. Dijkstra, Biochem. J. **413**(3), 517–525 (2008).

[71] D. Böttcher and U. T. Bornscheuer, Nat. Protocols **1**, 2340–2343 (2006).

[72] N. Lülsdorf, L. Vojcic, H. Hellmuth, T. T. Weber, N. Mußmann, R. Martinez, and U. Schwaneberg, Appl. Microbiol. Biotechnol. **99**(12), 5237–5246 (2015).

[73] C. Schnepel, H. Minges, M. Frese, and N. Sewald, Angew. Chem. Int. Ed. **55**, 14159 (2016).

[74] O. J. Miller, E. G. Hibbert, C. U. Ingram, G. J. Lye, and P. A. Dalby, Biotechnol. Lett. **29**, 1759 (2007).

[75] B. S. Evans, Y. Chen, W. W. Metcalf, H. Zhao, and N. L. Kelleher, Chem. Biol. **18**, 601 (2011).

[76] C. Shi, X. Lu, C. Ma, Y. Ma, X. Fu, and W. Yu, Appl. Biochem. Biotechnol. **151**(1), 51–59 (2008).

[77] H. Li, L. Mei, V. B. Urlacher, and R. D. Schmid, Appl. Biochem. Biotechnol. **144**, 27 (2008).

[78] B. Van Loo, J. H. L. Spelberg, J. Kingma, T. Sonke, M. G. Wubbolts, and D. B. Janssen, Chem. Biol. **11**, 981 (2004).

[79] G. Yang and S. G. Withers, ChemBioChem **10**, 2704 (2009).

[80] A. Aharoni, K. Thieme, C. P. C. Chiu, S. Buchini, L. L. Lairson, H. Chen, N. C. J. Strynadka, W. W. Wakarchuk, and S. G. Withers, Nat. Methods **3**, 609 (2006).

[81] S. Becker, A. Michalczyk, S. Wilhelm, K. E. Jaeger, and H. Kolmar, ChemBioChem **8**, 943 (2007).

[82] D. S. Tawfik and A. D. Griffiths, Nat. Biotechnol. **16**, 652 (1998).

[83] J. D. Bloom, S. T. Labthavikul, C. R. Otey, and F. H. Arnold, Proc. Natl. Acad. Sci. **103**, 5869 (2006).

[84] J. L. England and E. I. Shakhnovich, Phys Rev Lett. **90**(21), 218101 (2003).

[85] T. L. O'Loughlin, W. M. Patrick, and I. Matsumura, Protein Eng. Des. Sel. **19**(10), 439–442 (2006).

[86] R. D. Socha and N. Tokuriki, FEBS J. **280**(22), 5582–5595 (2013).

[87] V. W. C. Soo, Y. Yosaatmadja, C. J. Squire, and W. M. Patrick, J. Biol. Chem. **291**(38), 19873–19887 (2016).

[88] P. J. O. Brien and D. Herschlag, Chem. Biol. **6**, R91 (1999).

[89] M. Kitaoka, J. Tian, and M. Nishimoto, Appl. Environ. Microbiol. **71**(6), 3158–3162 (2005).

[90] F. H. Arnold and A. A. Volkov, Curr. Opin. Chem. Biol. **3**(1), 54–59 (1999).

[91] B. K. Shoichet, W. A. Baase, R. Kuroki, and B. W. Matthews, Proc. Nat. Acad. Sci. **92**, 452 (1995).

[92] Y. Kim, J. Choi, J. Kim, C. Park, J. Kim, H. Cha, S. Lee, B. Oh, T. Moon, and K. Park, Appl. Environ. Microbiol. **69**, 4866–4874 (2003).

[93] D. Elizabeth, K. Rumbold, K. Permaul, B. Alexander, and S. Singh, J. Biotechnol. **27**(3), 348–354 (2006).

[94] M. Vázquez-Rey and D. A. Lang, Biotechnol. Bioeng. **108**(7), 1494–1508 (2011).

[95] K. C. Entzminger, J. L. Johnson, J. Hyun, R. L. Lieberman, and J. A. Maynard, Protein Eng. Des. Sel. **28**(10), 365–377 (2015).

[96] M. Dai, H. E. Fisher, J. Temirov, C. Kiss, M. E. Phipps, P. Pavlik, J. H. Werner, and A. R. M. Bradbury, Protein Eng., Des. Sel. **20**, 69 (2007).

[97] C. Kiss, J. Temirov, L. Chasteen, G. S. Waldo, and A. R. M. Bradbury, Protein Eng., Des. Sel. **22**, 313 (2009).

[98] K. B. Zeldovich, P. Chen, and E. I. Shakhnovich, Proc. Natl. Acad. Sci. **104**, 16152 (2007).

[99] N. Tokuriki, F. Stricher, J. Schymkowitz, L. Serrano, and D. S. Tawfik, J. Mol. Biol. **369**, 1318 (2007).

[100] N. J. Turner, Nat. Chem. Biol. **5**, 567 (2009).

[101] A. S. Bommarius, J. M. Broering, J. F. Chaparro-Riggers, and K. M. Polizzi, Curr. Opin. Biotechnol. **17**, 606 (2006).

[102] N. Tokuriki, F. Stricher, L. Serrano, and D. S. Tawfik, PLoS Comput. Biol. **4**(2), e1000002 (2008).

[103] P. Luo, R. J. Hayes, C. Chan, D. M. Stark, M. Y. Hwang, J. M. Jacinto, P. Juvvadi, H. S. Chung, A. Kundu, M. L. Ary, and B. I. Dahiyat, Protein Sci. **11**, 1218 (2002).

[104] S. Yin, F. Ding, and N. V. Dokholyan, Nat. Methods **4**, 466 (2007).

[105] D. Gilis and M. Rooman, Protein Eng., Des. Sel. **13**, 849 (2000).

[106] S. Khan and M. Vihinen, Hum. Mutat. **31**, 675 (2010).

[107] M. Lehmann, L. Pasamontes, S. F. Lassen, and M. Wyss, Biochim. Biophys. Acta **1543**, 408 (2000).

[108] K. A. Luke, C. L. Higgins, and P. Wittung-Stafshede, FEBS J. **274**, 4023 (2007).

[109] M. T. Reetz, J. D. Carballeira, and A. Vogel, Angew. Chem. Int. Ed. **45**, 7745 (2006).

[110] P. Alfarano, G. Varadamsetty, C. Ewald, F. Parmeggiani, R. Pellarin, O. Zerbe, A. Plückthun, and A. Caflisch, Protein Sci. **21**, 1298 (2012).

[111] B. Borgo and J. J. Havranek, Proc. Natl. Acad. Sci. **109**, 1494 (2012).

[112] R. Jacak, A. Leaver-Fay, and B. Kuhlman, Proteins: Struct., Funct., Bioinf. **80**, 825 (2012).

[113] A. A. Dombkowski, K. Z. Sultana, and D. B. Craig, FEBS Lett. **588**, 206 (2014).

[114] R. T. Sauer, K. Hehir, R. S. Stearman, M. a. Weiss, A. Jeitler-Nilsson, E. G. Suchanek, and C. O. Pabo, Biochemistry **25**, 5992 (1986).

[115] S. F. Betz, Protein Sci. **2**, 1551 (1993).

[116]V. I. Abkevich and E. I. Shakhnovich, J. Mol. Biol. **300**, 975 (2000).

[117]W. Zhang, E. J. Mullaney, and G. L. Xin, Appl. Environ. Microbiol. **73**, 3069 (2007).

[118]L. Jermutus, M. Tessier, L. Pasamontes, A. P. G. M. Van Loon, and M. Lehmann, J. Biotechnol. **85**, 15 (2001).

[119]H. Fu, G. R. Grimsley, A. Razvi, J. M. Scholtz, and C. N. Pace, Proteins: Struct. Funct. Bioinf. **77**, 491 (2009).

[120]A. Goldenzweig, M. Goldsmith, S. E. Hill, O. Gertman, P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck, J. Prilusky, R. L. Lieberman, A. Aharoni, I. Silman, J. L. Sussman, D. S. Tawfik, and S. J. Fleishman, Mol. Cell **63**, 337 (2016).

[121]T. J. Magliery, Curr. Opin. Struct. Biol. **33**, 161 (2015).

[122]O. Khersonsky, D. Röthlisberger, A. M. Wollacott, P. Murphy, O. Dym, S. Albeck, G. Kiss, K. N. Houk, D. Baker, and D. S. Tawfik, J. Mol. Biol. **407**, 391 (2011).

[123]B. J. Sullivan, T. Nguyen, V. Durani, D. Mathur, S. Rojas, M. Thomas, T. Syu, and T. J. Magliery, J. Mol. Biol. **420**, 384 (2012).

[124]D. L. Trudeau, T. M. Lee, and F. H. Arnold, Biotechnol. Bioeng. **111**, 2390 (2014).

[125]A. Goldenzweig, M. Goldsmith, S. E. Hill, J. L. Sussman, D. S. Tawfik, S. J. Fleishman, O. Gertman, P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck, J. Prilusky, R. L. Lieberman, A. Aharoni, and I. Silman, Mol. Cell **63**, 337 (2016).

[126]H. J. Wijma, R. J. Floor, P. A. Jekel, D. Baker, S. J. Marrink, and D. B. Janssen, Protein Eng. Des. Sel. **27**, 49 (2014).

[127]A. Korkegian, Science **308**, 857 (2005).

[128]E. Vázquez-Figueroa, J. Chapparro-Riggers, and A. S. Bommarius, ChemBioChem **8**, 2295 (2007).

[129]J. Liao, M. K. Warmuth, S. Govindarajan, J. E. Ness, R. P. Wang, C. Gustafsson, and J. Minshull, BMC Biotechnol. **7**, 16 (2007).

[130]D. Gao, D. L. Narasimhan, J. Macdonald, R. Brim, M.-C. Ko, D. W. Landry, J. H. Woods, R. K. Sunahara, and C.-G. Zhan, Mol. Pharmacol. **75**, 318 (2009).

[131]J. C. Joo, S. Pohkrel, S. P. Pack, and Y. J. Yoo, J. Biotechnol. **146**, 31 (2010).

[132]H. S. Kim, Q. A. T. Le, and Y. H. Kim, Enzyme Microb. Technol. **47**, 1 (2010).

[133]J. E. Diaz, C. S. Lin, K. Kunishiro, B. K. Feld, S. K. Avrantinis, J. Bronson, J. Greaves, J. G. Saven, and G. A. Weiss, Protein Sci. **20**, 1597 (2011).

[134]H. J. Park, J. C. Joo, K. Park, and Y. J. Yoo, Biotechnol. Bioprocess Eng. **17**, 722 (2012).

[135]X. W. Yu, N. J. Tan, R. Xiao, and Y. Xu, PLoS One **7**(10), e46388 (2012).

[136]J. Tian, P. Wang, L. Huang, X. Chu, N. Wu, and Y. Fan, Appl. Microbiol. Biotechnol. **97**, 2997 (2013).

[137]X. Duan, J. Chen, and J. Wu, Appl. Environ. Microbiol. **79**, 4072 (2013).

[138]C. W. Lee, H. J. Wang, J. K. Hwang, and C. P. Tseng, PLoS One **9**(11), e112751 (2014).

[139]L. B. Harrington, R. K. Jha, T. L. Kern, E. N. Schmidt, G. M. Canales, K. B. Finney, A. T. Koppisch, C. E. M. Strauss, and D. T. Fox, ACS Synth. Biol. **6**, 120 (2017).

[140]H. Tan, R. Miao, T. Liu, X. Cao, X. Wu, L. Xie, Z. Huang, W. Peng, and B. Gan, J. Microbiol. Biotechnol. **26**, 1717 (2016).

[141]A. N. Shirke, D. Basore, G. L. Butterfoss, R. Bonneau, C. Bystroff, and R. A. Gross, Proteins: Struct. Funct. Bioinf. **84**, 60 (2016).

[142]H. Yu, Y. Yan, C. Zhang, and P. A. Dalby, Sci. Rep. **7**, 41212 (2017).

[143]J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff, and D. C. Phillips, Nature **181**, 662 (1958).

[144]B. Gutte, M. Däumigen, and E. Wittschieber, Nature **281**, 650 (1979).

[145]B. I. Dahiyat, Science **278**, 82 (1997).

[146]See http://predictioncenter.org/index.cgi for information about CASP (Critical Assessment of protein Structure Prediction) (2017).