Advanced Drug Delivery Reviews 183 (2022) 114143

Contents lists available at ScienceDirect

Advanced Drug Delivery Reviews

journal homepage: www.elsevier.com/locate/adr

Tools for computational design and high-throughput screening of therapeutic enzymes

Michal Vasina [a,b,c,1], Jan Velecký [a,b,1], Joan Planas-Iglesias [a,b,c,1], Sergio M. Marques [a,b,c], Jana Skarupova [a,b], Jiri Damborsky [a,b,c,d], David Bednar [a,b,c,*], Stanislav Mazurenko [a,b,c,*], Zbynek Prokop [a,b,c,*]

[a] Loschmidt Laboratories, Department of Experimental Biology, Faculty of Science, Masaryk University, Kotlarska 2, Brno, Czech Republic
[b] Loschmidt Laboratories, RECETOX, Faculty of Science, Masaryk University, Kotlarska 2, Brno, Czech Republic
[c] International Clinical Research Centre, St. Anne's University Hospital, Pekarska 53, Brno, Czech Republic
[d] Enantis, INBIT, Kamenice 34, Brno, Czech Republic

ARTICLE INFO

Article history:
Received 26 December 2021
Revised 4 February 2022
Accepted 9 February 2022
Available online 12 February 2022

Keywords:
Big data
Bioinformatics
Biopharmaceuticals
Biocatalysts
Enzyme characterization
Enzyme diversity
Machine learning
Microfluidics
Rational design

ABSTRACT

Therapeutic enzymes are valuable biopharmaceuticals in various biomedical applications. They have been successfully applied for fibrinolysis, cancer treatment, enzyme replacement therapies, and the treatment of rare diseases. Still, there is a permanent demand to find new or better therapeutic enzymes, which would be sufficiently soluble, stable, and active to meet specific medical needs. Here, we highlight the benefits of coupling computational approaches with high-throughput experimental technologies, which significantly accelerate the identification and engineering of catalytic therapeutic agents. New enzymes can be identified in genomic and metagenomic databases, which grow thanks to next-generation sequencing technologies exponentially. Computational design and machine learning methods are being developed to improve catalytically potent enzymes and predict their properties to guide the selection of target enzymes. High-throughput experimental pipelines, increasingly relying on microfluidics, ensure functional screening and biochemical characterization of target enzymes to reach efficient therapeutic enzymes.

Contents

* Corresponding authors.
E-mail addresses: 222755@mail.muni.cz (D. Bednar), mazurenko@mail.muni.cz (S. Mazurenko), zbynek@chemi.muni.cz (Z. Prokop).
[1] Joint first authors.

https://doi.org/10.1016/j.addr.2022.114143
0169-409X/

1. Introduction

Enzymes are natural catalysts accelerating chemical reactions in multiple catalytic cycles, leading to the conversion of substrates to products. Enzymes play an important role in the pharmaceutical industry because they catalyze chemical reactions selectively and efficiently. Therapeutic enzymes stand out from non-enzymatic drugs as they provide an advantage of binding specifically to their respective targets, reduced toxicity, and minimal side effects. Moreover, enzymatic catalysis enables the conversion of multiple targets, including prodrugs, simultaneously into the desired products. Therefore, therapeutic enzymes have shown diverse applications as enzybiotics, digestive aids, anti-inflammatory, fibrinolytic and anti-cancer agents [1].

In the past hundred years, several types of therapeutic enzymes have been developed. The first genetically engineered drug was insulin in 1982. Activase (alteplase; recombinant human tissue plasminogen activator) was the second recombinant protein drug to be marketed. In 1987, Activase was approved by the Food and Drug Administration (FDA) as the first recombinant enzyme drug [2]. This clot-buster enzyme is used to treat heart attacks and acute stroke caused by the artery blockage by a clot. In 1990, Adagen (pegadamase bovine), a form of bovine adenosine deaminase, was approved to treat patients afflicted with a chronic deficiency of adenosine deaminase [2]. Since then, more therapeutic enzymes have gained importance in pharmaceutical industries, such as L-asparaginase, streptokinase, collagenase, ribonuclease, uricase, glucosidase, etc. [3].

These therapeutic enzymes are produced in large volumes at good manufacturing practice quality by employing various fermentation techniques and host organisms, ranging from bacteria to yeast, fungi, and mammalian cells. The key technology enabling the production of enzymes in sufficient amounts for therapeutic applications is recombinant DNA technology, which can be complemented with protein engineering, material sciences, and nanotechnologies [1]. Access to these modern biotechnologies is critical since therapeutic applications cannot be adjusted to the properties of a biocatalyst – the properties of enzymes must fit the application. Therapeutic biocatalysts need to be primarily stable, soluble and active, yet improving their resilience to non-natural conditions represents a promising engineering strategy [4].

Therefore, it is necessary to either discover new enzymes from specific organisms or improve existing enzymes through protein engineering. This article provides an overview of modern computational tools suitable for identifying potentially exciting enzymes and genomic/metagenomics databases and for enzyme engineering by molecular modeling, structural bioinformatics, and machine learning. We also highlight the need to combine computational approaches with efficient screening and biochemical characterization of target enzymes by high-throughput experimental technologies (Fig. 1). Whenever possible, we provide examples of utilization of a particular tool or experimental technique for the development of therapeutic enzymes.

2. Identification of novel enzymes

2.1. Rule-based methods

Rule-based methods aim to predict enzymatically catalyzed reactions and complete biochemical pathways. These methods are integrated into many software tools and ready-to-use websites (see Table 1). The data stored in the KEGG database developed by Kanehisa and co-workers [5] is the single most important resource typically used for deriving the rules, patterns and metabolic networks, and validation purposes.

PathPred is a web-based server that predicts plausible pathways of multi-step reactions starting from query compound [6]. This tool employs the match of biochemical structure transformation patterns and the global chemical structure alignment against the reactant pair library. Biochemical structure transformation patterns are derived from the KEGG PATHWAY database [5]. The server uses local biochemical structure transformation patterns in 947 and 1397 reactant pairs to predict plausible pathways for microbial biodegradation of environmental compounds and biosynthesis of secondary plant metabolites, respectively. The server displays all

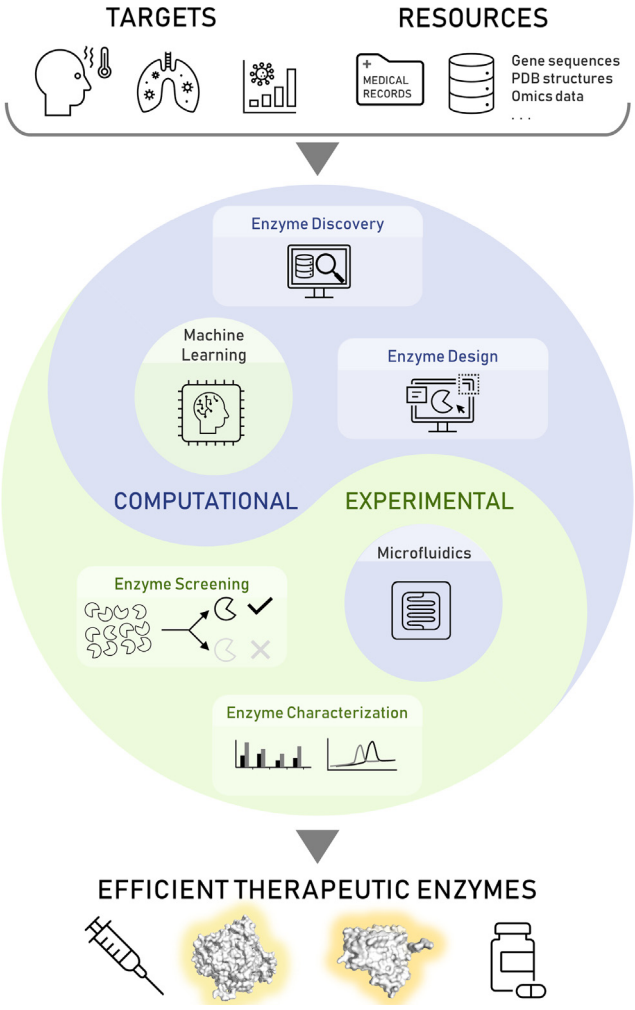

Fig. 1. Integration of computational and experimental approaches for the development of efficient therapeutic enzymes. The close cooperation between the computational and experimental approaches is depicted with a yin-yang scheme, where major objectives of each approach are in boxes, and the key techniques are in circles. These techniques are in the opposite color as they greatly depend on inputs from the counterpart approach.

Table 1
Overview of the sequence-based computational tools for novel enzymes design and discovery. They are available at the given locations – most of them as web tools and the rest as stand-alone programs. The table displays the tools discussed in Section 2 and therefore is not exhaustive.

Tool name	Objective	Prediction result	Ref.	Location
PathPred	Pathway prediction	Metabolic pathways of multi-step reactions	[6]	www.genome.jp/tools/pathpred/
PathSearch		Similarity paths for query reactions	[6]	www.genome.jp/tools/pathsearch/
PathComp		Putative reaction paths between two molecules	[6]	www.genome.jp/tools/pathcomp/
Retro-path	Retro-pathway design	Retrosynthesis based on generalized reaction rules	[7]	myexperiment.org/workflows/4987.html
METEOR		Knowledge-based metabolic fate of chemicals	[8]	lhasalimited.org/products/meteor-nexus.htm
BNICE		Generalized rules for retrosynthesis	[9]	lcsb-databases.epfl.ch/pathways/atlas/
BridgIT		Enzyme binding pockets and reaction similarities	[10]	lcsb-databases.epfl.ch/pathways/Bridgit
Transform-MinER		Substrates transformed into products at reaction centers	[13]	www.ebi.ac.uk/thornton-srv/transform-miner/
antiSMASH 6.0	Homology search	Homologous sequences based on Hidden Markov Models	[23]	antismash.secondarymetabolites.org
dbCAN2			[24]	bcb.unl.edu/dbCAN2/
DETECT v2		Homologous sequences based on density profiles	[26]	github.com/ParkinsonLab/DETECT-v2
HEC-Net	EC number prediction	EC by sequence similarity and pattern recognition	[32]	hecnet.cbrlab.org/
Bio2Rxn		EC by convolutional neural networks	[33]	design.rxnfinder.org/bio2rxn/
DEEPre		EC by convolutional and recurrent neural networks	[34]	cbrc.kaust.edu.sa/DEEPre/
ECPred		EC by ensemble of machine learning classifiers	[35]	ecpred.kansil.org/
DeepEC		EC by three convolutional neural networks	[36]	bitbucket.org/kaistsystemsbiology/deepec
Enzyme Miner	Novel enzyme search	Structurally and functionally diversified proteins	[37]	loschmidt.chemi.muni.cz/enzymeminer/
SoluProt	Solubility prediction	Protein solubility based on sequence information	[39]	loschmidt.chemi.muni.cz/soluprot/
Fireprot^{ASR}	Ancestor reconstruction	Reconstructed sequences of stable and robust ancestral proteins	[46]	loschmidt.chemi.muni.cz/fireprotasr/

predicted multi-step reaction pathways in a tree-shaped graph and provides transformed compounds and reference transformation patterns in each predicted reaction. PathSearch computes similar paths for a query reaction path using the RCLASS database for reaction patterns matching to the query reactions. The results can be mapped to KEGG pathway diagrams using the KEGG Mapper. PathComp computes possible reaction paths between two compounds using the binary relations of substrates and products in known enzymatic reactions. The list of binary relations is generated from the enzyme list that corresponds to the KEGG reference pathways.

Retro-path is an automated open-source workflow for retrosynthesis based on generalized reaction rules that perform the retrosynthesis search from chassis to target through an efficient and well-controlled protocol [7]. Retrosynthesis approaches explore the chemical biosynthetic space. The complexity associated with the large combinatorial retrosynthesis design space is the main challenge hindering this approach. Reaction rules used in retrosynthesis require a solved atom–atom mapping between the atoms of the substrates and those of the products to identify the reaction center. The atom–atom mapping problem is equivalent to the maximum common substructure problem. In some cases, avoiding atom–atom mapping to generate rules is possible, as addressed by the fingerprint subtraction. If encoding the reacting center is necessary, it may not be sufficient to properly define a reaction catalyzed by an enzyme since other atoms far from the reacting center could also be involved in the ligand binding. In this case, the definition of the reacting center is extended to neighboring atoms, either systematically at a predefined bond distance or based on manual curation.

A challenge for the retrosynthesis algorithms is the need to handle the processing of reactions of multiple substrates and multiple products. This effort requires more computational resources for modeling enzymatic promiscuity for each combination of promiscuous substrates. Retrosynthesis maps are subsequently constructed by iteratively applying reaction rules. Starting set of compounds is repeatedly assessed until a map of hits is generated. The application of this workflow can efficiently streamline retrosynthesis pathway design, reshape the design, build, test, and learn synthesis pipelines by driving the process towards an optimized bioproduction. There are several other systems available for the design of *retro*-pathways, e.g., METEOR [8], BNICE [9], BridgIT [10] and others. Machine learning (ML)-based approaches use Monte Carlo tree search and symbolic artificial intelligence (AI) to discover retrosynthetic routes [11]. Encoder-decoder architecture can also predict retrosynthetic pathways, consisting of two recurrent neural networks [12].

Transform-MinER (Transform Molecules in Enzyme Reactions) is an online computational tool that transforms query substrate molecules into products by applying known enzyme reactions at potential reaction centers and retrieves the most similar native enzyme reactions for each [13]. The server can be used in two modes. Molecule Search identifies potential enzyme transformations acting on a submitted query substrate. Path Search links submitted source and target molecules with enzyme transformations. The data behind Transform-MinER was obtained from the KEGG database [5]. Balanced Reaction files are generated for all reactions with available molecular structures. Reaction centers are atoms that change connectivity, neighbors, or stereochemistry. After performing the atom–atom mapping, these centers are identified from the mapped balanced reaction files. Canonical SMiles ARbitrary Target Specification SMARTS patterns are generated to represent these reaction centers, enabling query molecules to be searched for matching fragments. The results of Transform-MinER are presented to the user using two interactive views: the Path View shows molecules as nodes and their transformations as edges, and the Molecule View shows reaction centers as the interactive shapes. The users can control the number of transformations presented using a similarity slider that varies the similarity threshold between the submitted minimum similarity threshold and 1.0. Similar enzyme reactions data table is populated by showing matching native enzyme reactions and substrates in a descending reaction center similarity when selecting a transformation. Hyperlinks take the user to the KEGG reaction and KEGG compound, with the native KEGG compound structure in a hover-over box.

2.2. Sequence-based methods

In contrast to rule-based methods, sequence-based methods harness the fast-paced growth of genomic databases as a publicly available source for potential new biocatalysts. Strikingly, the ratio

of biochemically uncharacterized proteins is immense, as shown by the fact that only one in 450 proteins recorded in the NCBI nr database [14] is also present in the manually curated UniProtKB/Swiss-Prot database [15]. Properly assembled full-length protein sequences from metagenomic studies are also available in the database MGnify [16]. By mining these databases, bioinformatics approaches can significantly reduce the time and costs required by the available high-throughput experimental methods to biochemically characterize such gene expression products [17]. Metagenomic projects typically include bioinformatics pipelines aimed to automatically identify the function of the isolated gene products [18]. Such pipelines are based on the homology search strategies, where the well-known BLAST tool [19,20] still plays a major role in assessing sequence similarity: a sequence with unattributed activity is likely to perform a similar function to that of the most related sequence. Homology search was utilized for recent *in silico* identification of potential therapeutic L-asparaginases from hyperthermophilic organisms [21].

This rule of thumb has been extended for particular activities through narrowing the homology search, using patterns such as the hidden Markov model profiles (HMMs) [22]. Particularly antiSMASH 6.0, [23]) and dbCAN2, [24]) are popular tools relying on this technology. They encode profiles for specific activities against which genomic sequences can be scanned. Instances of specific activities covered by such profiles may include, for example, fungal functions such as bacteriocin, beta-lactam, or aminoglycoside biosynthesis, and carbohydrate-active enzyme sub-families as representatives of bacterial activities. Those profiles are built from sets of human-curated, well-annotated protein alignments such as those present in PFAM [25]. When an unannotated sequence from a genomic project matches one of such function-related profiles, the function encoded by the profile with a certain confidence probability can be assigned. DETECT v2 [26] explores the same concept, but instead of exploiting HMMs, it builds density profiles from homology alignments for each Enzyme Commission (EC) number. Each density profile is labeled as positive if it includes only one EC and negative otherwise. All the density profiles can be scanned over the query sequence(s), and a confidence score on the assigned EC number is derived from the ratio of positive and negative profiles. An earlier version of DETECT was used to detect novel enzymatic functions in *Plasmodium falciparum* [27].

Since the sequence space of genomic databases is vast, it is usual to find a large number of targets from the initial homology scan, especially if no dedicated profile was used to find the candidates. In order to avoid nearly identical results among the hits selected for further biochemical characterization, sequences are filtered out from the results based on an identity percentage threshold, which varies depending on the queried enzymatic activity. For example, a 90% threshold was used to design a synthetic pathway for 2-keto acid carbon chain elongation [28] and a 95% threshold for the discovery of novel imine reductases [29]. The sequence similarity measurements used for homology search can also be employed to represent the sequence/function space employing Sequence Similarity Networks [30]. They were pioneered by Babbitt and colleagues [31] and have become useful guidance when aiming to identify the most diverse candidates for further biochemical characterization [29].

The homology strategy described above approaches the problem of function assigning from the perspective of pattern recognition: patterns in a sequence lead to a functional assignment. Machine-learning approaches have since long contributed to the field within this domain of pattern recognition. Recently, a deep-learning tool named HEC-Net [32] has shown the ability to predict the fourth level of the EC classification with striking accuracy (over 90%). The prediction exploits sequence similarity and pattern recognition as well as individual amino-acid biochemical properties. Bio2Rxn [33] is another present-day tool for enzymatic activity inference that exploits convolutional neural networks and five other more traditional homology- and biochemical feature-based predictors. Moreover, Bio2Rxn offers a consensus annotation with over 90% precision and recall close to 60%. Other notable tools in this scope are DEEPre [34], ECPred [35], and DeepEC [36].

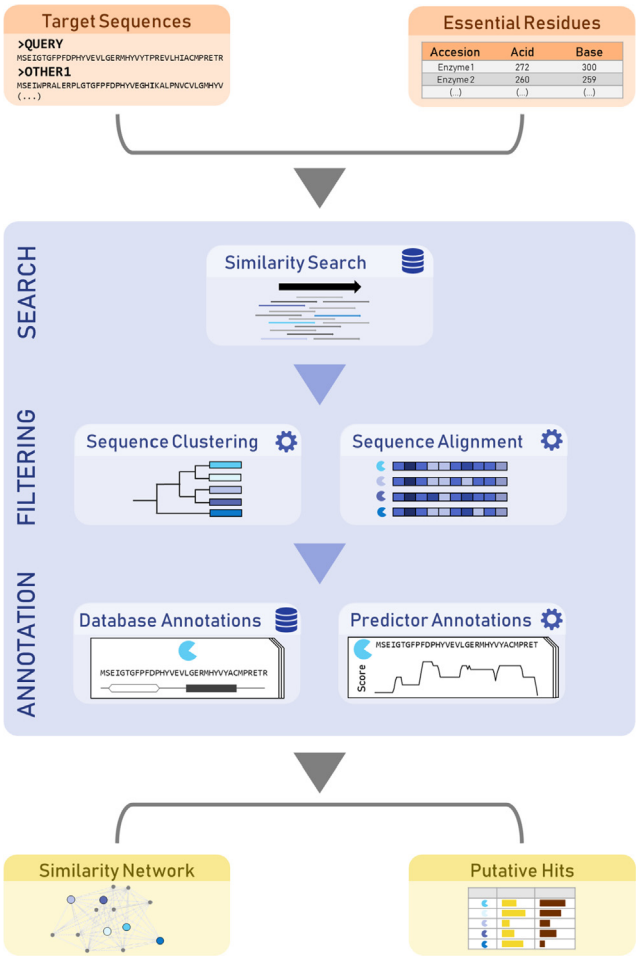

Fig. 2. Illustration of the EnzymeMiner workflow. The server accepts multiple target protein sequences as the input (orange). It can retrieve catalytic residues from the Catalytic Site Atlas, or the user can define them (allowing for degenerated positions). The mining (blue) starts with searching for homologs to query enzymes. The obtained hits are subsequently filtered based on clustering and alignment, ensuring the presence of the defined essential residues. Multiple annotations are retrieved to enrich the information of the filtered list of hits. The final results (yellow) are presented in two interactively integrated views: (i) the similarity network view presents the sequences clustered according to their sequence similarity, and (ii) the putative hits view allows for prioritization according to any of the retrieved annotations. The web service is provided free of charge for non-commercial use: https://loschmidt.chemi.muni.cz/enzymeminer/. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Opposite to the task of predicting the enzymatic activity from sequence and more related to the applications of the rule-based methods reviewed in the previous section is the exercise of comprehensively finding existing protein sequences that will execute a desired enzymatic activity. EnzymeMiner [37] tackles this challenge by inferring a number of activity-related constraints of the desired enzymatic function from its sequence inputs. The server accepts multiple sequences annotated with the desired enzymatic function as input, queries the Mechanism and Catalytic Site Atlas [38] to elucidate their catalytic and essential residues (which can also be given as input), and from them builds a protein sequence

enriched by essential residue profiles. EnzymeMiner automatically performs homology searches on a number of genomic-derived protein sequence databases using the protein sequence profile and filters out recovered hits that do not satisfy the derived essential residue profile. Thus, EnzymeMiner identifies only sequences that potentially carry out the desired enzymatic activity (Fig. 2). In addition, it ranks the results by the predicted solubility of the retrieved sequence (by SoluProt, [39]). It provides annotation on the source organism, extremophilic nature, structure availability, and other features such as sequence similarity networks visualization to assist target selection.

The ancestral sequence reconstruction is a complementary strategy for designing novel catalysts exploiting sequence similarity relationships. The rationale behind inferring the sequence composition of ancient variants of modern-day proteins is that such variants are often more thermostable [40] but may also exhibit interesting catalytic properties [41,42]. The method has proven successful in increasing the thermal stability of phenylalanine/tyrosine ammonia-lyase for complementary treatment of hereditary tyrosinemia type I [43] or for doubling the activity of iduronate-2-sulfatase for the treatment of Hunter syndrome [44]. The topic has recently been reviewed elsewhere [45], but it is worth noting here that the results obtained from EnzymeMiner can be used to feed such reconstruction methods. One of the latest tools available for ancestral sequence reconstruction is Fireprot^{ASR} [46]. It represents the first fully automated platform devoted to inferring the ancestral sequences of a given protein and achieves its goal by applying a multiple sequence alignment and phylogenetic tree reconstruction. Beyond sequence-derived information, knowledge derived from the three-dimensional structural arrangement of the target enzymes can be used to improve their catalytic properties further, as covered in the following section.

3. Rational design of highly active, stable, and soluble enzymes

3.1. Molecular modeling and bioinformatics

The applicability of any enzyme to a specific practical use depends mostly on several global properties: the catalytic activity and specificity towards a substrate or product, the solubility or bioavailability, and its stability. Over the last few decades, protein engineers have developed many methods and computational tools to enhance these aspects according to particular needs. Such tools have been recently reviewed [47–51]. Here we will focus only on some of the most popular and promising tools available to non-specialist users over user-friendly web servers.

ROSIE [52] is one of the web platforms for molecular modeling with the widest scope of applications. Developed in 2013 and updated recently, it hosts dozens of tools from the Rosetta family under the same roof for modeling and designing proteins, nucleic acids, and other biopolymers. In a uniform and user-friendly environment, enzyme engineers can perform, for example, molecular docking (*Ligand Docking* and others), predict and design stability (*Sequence Tolerance*, *RosettaVIP*), and improve solubility (*Supercharge*). It also allows the design of non-enzymatic proteins, such as antibodies (*RosettaAntibody*). RosettaDesign is another Rosetta-based software piece useful for enzyme design [53], also available as a web server. This tool can be applied for *de novo* design, modification of larger secondary structure elements, or to add mutations that can change the protein activity, specificity, or flexibility. This method has also demonstrated the ability to stabilize protein structures [54].

HotSpot Wizard 3.0 [55] is a web application for multiple analyses of proteins that allows a versatile identification of hot spots for mutagenesis. It can identify functional hot spots that may lead to improved catalytic function based on analyzing active sites, enzyme tunnels, and molecular docking. Furthermore, it evaluates correlated positions derived from the evolutionary analysis to prevent potentially deleterious mutations. HotSpot Wizard also allows the selection of hot spots for improving protein stability based on the analysis of flexibility (according to the respective B-factors) or by the back-to-consensus approach (phylogenetic analysis). The suggested hot spots are ranked by the respective mutability scores derived from a conservation analysis. The user can input either the sequence or three-dimensional structure of the target enzyme. As a result, the user can obtain an estimation of the effect of mutations on protein function and the effect of selected mutations on the enzyme stability, given by the changes in free energy. Effects on function and stability, together with evolutional variability on each position, enable the user to design optimal DNA codons to build smart libraries for screening the selected positions. HotSpot Wizard was used to identify mutable residues for improving flavin reductase FRase I, which improves activation of the anticancer prodrug CB1954 [56].

When it comes to improving the activity or specificity of an enzyme towards a substrate of interest, the most intuitive and common strategy is to modify its catalytic site and its immediate surroundings. FuncLib [57] was designed to modify the active site of an enzyme and add non-deleterious multiple-point mutations that will diversify the substrate specificity profiles while accounting for epistatic effects. It can also optimize the catalytic site towards a substrate of interest and improve the respective catalytic efficiency. It combines phylogenetic analysis and energy calculations to propose several multiple-point designs ranked based on the predicted stability. This tool was employed to improve protein yield and stability of an enzyme PodA, which has a therapeutic potential of eliminating biofilms of human pathogen Pseudomonas aeruginosa [58].

It is well-known that the modification of molecular tunnels can strongly impact several properties of the enzymes, such as their catalytic activity, substrate specificity, and stability to temperature or co-solvents [59]. CaverWeb [60] is a web-based tool that can calculate and analyze the access tunnels in enzymes and other proteins. Such analysis may reveal important structural features (e.g., the tunnel bottlenecks) and help identify hot spots to be targeted to improve the catalytic properties by ligand transport optimization. From the CaverWeb interface, the CaverDock software [61] can be also run to predict the trajectory of a small molecule (e.g., a substrate or a product) traveling through a given access tunnel and compute its (un)binding energetic profile. This prediction can reveal which ligands can (un)bind more easily to or from the enzyme active site. Moreover, it can be used to compare different enzymes (e.g., a wild-type and a mutant variant) and assess which ones are more prone to bind a certain substrate of interest, and hence predict the catalytic differences. Caver has recently been used to facilitate the engineering of nicotine oxidase, which is used for therapeutic enzymatic blockade of nicotine from the central neural system [62].

Loops are structural features common to nearly all proteins that are typically flexible. Often, they have a functional role, such as substrate recognition. Frequently, they are the dynamic elements that act as gateways to the active site and thus are important for catalytic regulation, determining the activity and substrate specificity [63,64]. LoopGrafter [65] is a tool designed to optimize the transplant of loops between two structurally related proteins. It starts by predicting the dynamic properties of the proteins and the correlations among the movements of different secondary structure elements (Fig. 3), which helps the user to select the relevant elements to be grafted. Next, it evaluates the geometric properties of the selected loops and proposes several solutions to find the ideal insertion points. The tool also provides a three-

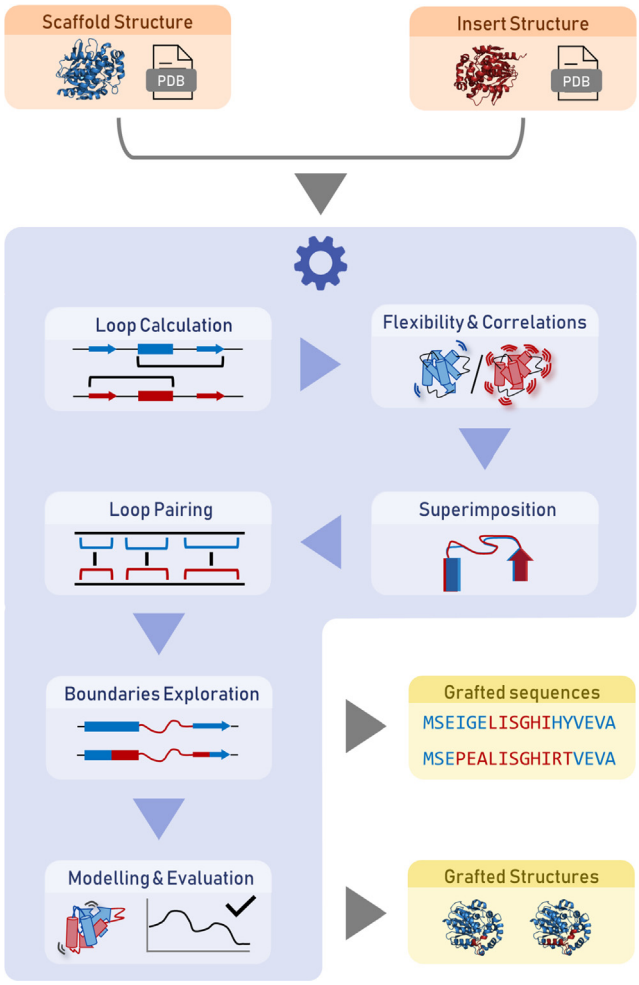

Fig. 3. Illustration of the LoopGrafter workflow. The server starts with the input of the PDB files or codes of two structures (orange): (i) the loop receptor (scaffold) and (ii) the loop donator (insert). The tool then calculates the secondary structures of all the loops, assesses the protein flexibility using normal mode analysis, and performs a superimposition of the two structures (blue). Once the user selects the loops of interest, the loops are paired. The suitable loop boundaries are then explored based on geometric restrictions so that maximal sequence diversity is probed in the designed grafted variants. The possible solutions are collected as a list of sequences, and the structures are then modeled and ranked by the respective stability scores obtained from Rosetta and MODELLER (yellow). The web service is provided free of charge for non-commercial use: https://loschmidt.chemi.muni.cz/loopgrafter/. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

dimensional model of the best resulting designs. The workflow has been applied successfully to graft loops between two structurally related (α/β hydrolase fold) but functionally very different enzymes, resulting in the transfer of activity from one to another [66]. The tool can potentially combine a highly efficient catalyst with a stable or soluble protein to create a new biocatalyst that is more suitable for practical applications.

The stability of proteins is an important factor for their applicability as it determines whether they resist the conditions required for their use (e.g., temperature, pH, co-solvents, etc.) without unfolding or for how long they survive in storage before degradation. Protein Repair One-Stop-Shop (PROSS) [67] is an automated web platform designed to improve protein thermostability and functional yields. A phylogenetic analysis is performed to exclude conserved positions or infrequent amino acids from the subsequent virtual mutagenesis screening. Rosetta [68] performs energy calculations and scans which mutations are not deleterious. The resulting substitutions are then combined to find the optimal multiple-point mutations to increase stability. PROSS was successfully involved in increasing the thermal stability of many proteins, including therapeutic enzymes [69]. Notably, for tyrosine phosphatase PTPN3 specifically dephosphorylating the epidermal growth factor substrate 15 [70], the melting temperature was increased up to 27 °C, while the activity was three times higher compared to wild type [69].

FireProt [71] was developed to find proteins with improved thermal stability. It is based on a hybrid method [72] combining energy- and evolution-based approaches in a complex workflow. In a fully automated process, FireProt integrates several tools and smart intermediate filters sequentially applied to find stabilizing single-point mutations. Individual mutations are eventually combined to generate several multiple-point mutants to predict the free energy stabilization and the mutant structures.

The term protein solubility comprises distinct biophysical or biological aspects. These include (i) the intrinsic solubility in water (which is a balance between the hydrophilic/hydrophobic character), (ii) the aggregation propensity (which is related to the solvent exposure of aggregation-prone regions), or even (iii) the expressibility (the soluble protein extract from the total proteins produced from a specific expression system). Regardless of which of these aspects is more limiting, low solubility is undesirable since it hinders protein production and availability for practical applications. Several computational tools are available to tackle this problem and help design more soluble proteins, including SolubiS, Aggrescan3D 2.0, and SoluProt (described in detail in Section 3.2, machine learning).

SolubiS [73,74] is a web-based tool that aims to increase protein solubility. The automated pipeline starts with identifying aggregation-prone regions of the protein structure, followed by predicting mutations that could reduce aggregation propensity. It can differentiate between the buried and solvent-exposed regions, as especially the solvent-exposed ones are critical for protein–protein aggregation. Free energy calculations guarantee that these mutations do not impair the overall thermodynamic stability. Intrinsic aggregation scores rank the different regions to help the user prioritize their sequences. SolubiS was successfully applied to reduce aggregation of human α-galactosidase, which is used in replacement therapy for Fabry disease [75].

Aggrescan3D 2.0 [76] is another well-established aggregation predictor that maps the intrinsic aggregation propensity on the protein structure. It focuses its prediction on the protein surface, taking into account the protein flexibility in its dynamic mode. The predictor then suggests solubility-improving mutations for the user to choose from. It models the structures of the mutants and predicts their aggregation propensity and the energetic effect of each mutation.

Recent advancement of machine-learning algorithms holds a considerable potential for improvement of these web tools, as further described in the following section.

3.2. Machine learning

Human cognitive abilities interpreting regularities in complex data, such as those collected in laboratory experiments, are limited to distinguishing patterns in 3–4 dimensional spaces at most. This limitation hampers the possibility of learning higher-order dependencies in the interpreted data. For instance, such higher-order dependencies can be non-linear relationships between seemingly unrelated variables or obscure evolutionary relationships between distant amino acids in a polypeptide chain. This lack of depth in understanding natural and experimental data might, in turn, limit the capacities of understanding what makes enzymes suitable biocatalysts and thus restrain our potential for engineering them.

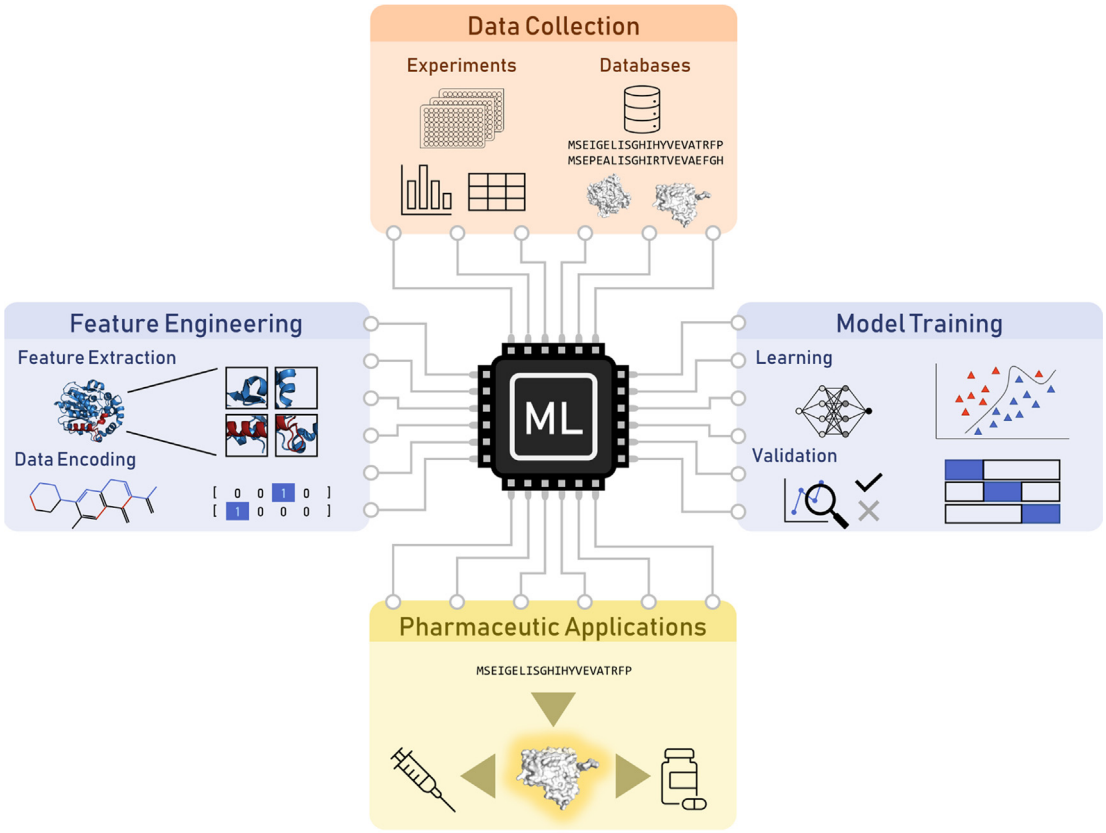

Fig. 4. Illustration of the machine learning workflow. Machine learning (ML) enables the discovery of hidden patterns in abundant biological data. Top: Abundant data may be collected from laboratory experiments or drawn from public databases. Left: Data features should be engineered to be used within a machine learning pipeline. Typically, the dimensionality is reduced by feature extraction. These features shall be suitably encoded to be correctly understood by the machine learning model. Right: A machine learning model is trained on the data until it achieves desired accuracy. The candidate model is then statistically evaluated on test data before its use in a real environment. Bottom: The final model is applied for useful structural- and functional-properties prediction of therapeutic enzymes.

Machine learning (ML) is a field in computer science devoted to creating learning models whose parameters are set automatically by training on input data, allowing these models to find patterns in such data with automated computational methods. Compared to human-driven data interpretation, ML allows finding complicated patterns consisting of multiple properties. ML methods represent soft computing methods as they naturally work with probabilities or imprecise values. These two considerations make ML suitable to study biology phenomena, which can be rarely explained both simply and precisely at the same time. We use the terms ML and AI interchangeably for this review as their methods overlap substantially.

ML workflows consist of several fundamental steps (Fig. 4). The data collection and treatment are crucial as the learning model that would be inferred can only be as good as the input data. The data collection is often the most time-consuming step as well. Since raw data may be too complex to be fed into an ML model directly, a set of features is commonly engineered from such data to reduce data dimensionality and avoid learning trivial rules between correlated properties. The aim of such features is to statistically represent the information encompassed in a whole set of data with a smaller amount of data. However, to speed up the learning, feature engineering can also introduce prior knowledge as these features can incorporate biological meaning (e.g., electrostatic charges of amino acids or any other biologically relevant property) rather than just statistical meaning. The computational time required in the learning phase is directly related to the complexity of the model represented by the number of parameters to be trained. In recent years, particularly complex ML systems, called deep neural networks or deep learning models, have been gaining popularity due to the high predictive power they can achieve. These intricate models can have billions of parameters, requiring long training times and making them prone to overfitting the training data, e.g., learning the distributions in the data perfectly and thus losing the ability to generalize [77]. However, once the model is trained, it can often predict new data almost instantly.

One of the most prominent success stories of implementing the ML workflow in biology is the recently published tool AlphaFold2 [78,79]. This deep learning tool was developed by DeepMind Technologies to solve one of the most challenging problems in biology: predicting protein structure from its sequence. AlphaFold won the two last Critical Assessments of protein Structure Prediction (CASP) – a biannual competition of the state-of-the-art predictors of protein 3D structure from an amino-acid sequence [80]. Its first appearance in the competition already brought the authors a win in the category of proteins that have no homologs with known structures. Two years later, AlphaFold2 dramatically increased its margin and approached the accuracy of experimental methods such as X-ray crystallography. The success would not be possible without almost two hundred thousand structures available in databases at that time. This breakthrough manifests the strength of ML when both big data and sufficient computational power are available. Heartened by these results, Alphabet, the parent company of DeepMind, established an Isomorphic Labs subsidiary in 2021, a venture focused on drug discovery with an AI-first approach [81]. Alphabet is not the first company to employ ML in the pharmaceutical sector. More than 40 established pharma companies are known to be employing ML methods [82], and over

230 start-ups use AI for drug discovery or development of products related to drugs or medicine [83].

Apart from structure prediction, ML can assist scientists with the design of enzymes by suggesting mutations in the existing enzymes to improve desired properties, such as stability, solubility, activity, specificity, and others. Such ML tools are focused on predicting the mutational effects on a property (how the mutation causes variation in the property) rather than the absolute value of the property. In addition, due to the breakthrough that AlphaFold represented in predicting protein structure, the use of ML tools devoted to predicting the effects of *de novo* designed drugs will soon increase. Finally, an important aspect that AlphaFold may bypass is the input type requirements for a particular ML tool to be used. While some predictors are built for sequence-only input, many others require three-dimensional structural data, and AlphaFold might help to bridge this gap for proteins whose structure has not been resolved experimentally yet.

As the available mutational data on protein stability is relatively abundant, multiple ML models to predict this property were recently reviewed by Marabotti et al. [84]. Two of the most recent ML models, DeepDDG [85] and DynaMut2 [86], represent different ML approaches – deep learning and random forest models, respectively. Both were trained on thousands of mutants to directly predict a Gibbs free energy ($\Delta\Delta G$) change upon introducing a mutation and are available as web tools. DeepDDG uses many features for each mutated position or for each pair of mutated positions and their neighboring amino acids. Such features include the amino-acid type, accessible surface area, or structural motifs. DynaMut2 uses features computed by normal mode analysis, graph-based signatures, or residue contacts for the training, which resulted in a more stable performance than its predecessors.

Protein solubility is a more complex phenomenon, especially in terms of its definition and quantification, and thus it has been a much lesser attractive target for ML predictive models. Nevertheless, this direction is also actively explored. One of the most recent predictors based on wild-type sequential data is SoluProt [39], available as a stand-alone program or as a web tool. It achieves better results than its predecessors and facilitates the prioritization of candidate sequences to increase the success rate of high-throughput experiments. The tool uses such features (including predicted ones) as the amino-acid content of the sequence, physicochemical properties, average flexibility, secondary structure content, and average disorder. SoluProt predicts the probability of a soluble protein expression in *E. coli* and simultaneously reflects the solubility and expressibility of the target proteins.

The extension of models trained on non-mutational data to predict the effect of mutations on protein solubility is challenging since the distances between wild-type sequences are typically much larger than those implied by mutations. And due to the lack of databases of mutational data on solubility, only a few simple ML models have been created to tackle this problem. The first such model was OptSolMut [87], trained on as few as 137 protein mutants. The authors brought the lack of mutational data for this problem to the community's attention in their work and manually collected the first dataset. In contrast, the authors of the most recent predictor PON-Sol2 [88] managed to collect a list of thousands of mutants on which to train their model. This shift in the magnitude of data availability can be attributed to the yield of high-throughput experiments in recent years. PON-Sol2 predicts three classes of solubility change (increased, decreased, not changed) but works as a pair of two stacked 2-class predictors. The first one predicts a decrease/non-decrease, and if the first prediction was non-decrease, the second decides between the remaining two choices (increased or not changed).

Apart from stability and solubility, several other enzyme properties became targets of ML analyses. EPSVR [89] is a web tool for predicting B-cell antigenic epitopes on a structure using support vector machines based on defining surface patches. Surface patches are scored according to the number of predicted epitope residues, and the best scoring ones are subsequently used to select the final predicted epitope residues. Features of epitope patches were learned on unbound structures of just 48 antibody-antigen complexes. EPSVR and other epitopes predictors may speed up and refine the discovery and development in epitope-based vaccine design. An in-silico study demonstrated the use of these predictors to find epitopes in SARS-CoV-2 [90]. NNTox [91] is a neural network to predict toxic proteins and their mode of toxicity. It was trained on Gene Ontology annotations of thousands of UniProt proteins, including or excluding the "toxin" keyword. Accordingly, protein toxicity is predicted given Gene Ontology terms. DeepBL [92] is a web tool based on deep learning of beta-lactamase identification and classification from a sequence and thus could help design its inhibitors. The tool was trained on beta-lactamase and random non-beta-lactamase sequences stored in the NCBI RefSeq database. The features are counts of all possible amino-acid pairs that appear in the input sequence at the sequential distance from 0 to 5.

The ML applications, as mentioned above, would not be possible without abundant data for the targets of interest. The more data is available, the more sophisticated models can be employed, and more complex patterns may be found accordingly. Another important criterion of source data is quality. When inaccurate or incorrect data is fed to a model, the model will likely learn inaccurate patterns. If not treated carefully, disbalances or underpopulated classes in data can project such biases to the final predictions.

Mutational and wild-type data availability for protein stability and solubility was reviewed by Musil et al. [49]. Two different databases were established to track various protein properties based on this work: FireProtDB [93] and SoluProtMutDB [94]. These are manually curated databases of stability and solubility changes, respectively. Both currently store experimental results in multiple thousands and follow the FAIR principles (Findable, Accessible, Interoperable and Reusable).

Another example of the importance of well-structured and curated data is the recent mutational database D3DistalMutation [95] that tracks the effects of mutations upon the enzymatic activity, comprising about a thousand entries. These differential protein databases store quantitative or qualitative effects of specific mutations on a property of interest in proteins. Apart from the selected differential databases, there are new databases of absolute values, such as MPTherm [96], storing thermodynamic data of membrane proteins.

The number of proteins with available mutational data is far behind the number of 3D structures available in the Protein Data Bank (PDB). The latter amounts to less than 1% of the known protein sequences. Also, the majority of the data in PDB are deposited by academia. This was demonstrated in an analysis of the structural biology contribution to new molecular entities (NMEs) development, where 4 of 5 relevant structures were uploaded by academia [97]. Alongside this, the private sector usually excludes structural biology insights from IP and patent filings for high-profile targets like antibodies or cancer therapeutics. Thus, the emerging high-throughput experimental techniques still hold great promise in unlocking the further potential of models to predict mutational effects. Furthermore, two classes of critical data are yet entirely missing: (i) data encompassing the mutational effect of insertions and deletions and (ii) data about the concomitant effect of mutational changes in two or more properties at the same time. The latter is especially relevant to addressing the multi-target prediction, i.e., the improvement of several properties at once. A trade-off between properties is often observed [41,98,99], e.g., antibody affinity *versus* stability or activity *versus* solubility. The task of

improving one target property without compromising the other is yet to be addressed in ML-aided design. In the following section, we cover the state of the art of experimental techniques that show considerable promise in generating the data necessary for addressing this task.

4. High-throughput enzyme screening

4.1. Starting a screening campaign

Enzyme screening resembles an adventurous journey through the immensely vast landscape of protein sequences [100] with the goal to find a treasure - enzymes with desired functional properties. In its actual biochemical meaning, however, screening requires laborious and costly experimental testing of the starting pool of enzymes. The size of this pool, which is feasible to be screened by the ultra-high throughput methods [101], reaches millions to billions of enzyme variants for directed evolution or metagenomic enzyme discovery campaigns. Nevertheless, this number is still only a drop in the ocean of the protein sequence space. Particularly, protein engineers aiming to use directed evolution need a good starting point – an enzyme to evolve [102]. These starting points can be provided by enzyme discovery or rational design studies (Sections 2 and 3). However, the hits from these studies can vary from tens to thousands of enzyme variants, so a high-throughput method to screen them is worth considering.

With any starting pool of enzymes, the screening procedure needs careful planning. Therefore, this section provides the key aspects to consider when starting a screening campaign. Any enzyme screening campaign involves the following general steps: (i) generation of a gene library encoding enzymes of interest; (ii) expression of the genes within the library; (iii) functional screening and selection of hits (requiring suitable screening technology and detection assay) and (iv) retrieving of the genotype-phenotype linkage by DNA sequencing. In the following subsections, we focus on (ii) and (iii), as these are the key variables for designing a screening campaign (Fig. 5). The required starting pool of enzymes, i.e., the library generation can vary tremendously with respect to the type of study. Therefore, we refer the readers to the recent reviews regarding directed evolution libraries [103] and metagenomic enzyme discoveries [104].

4.2. Suitable screening platforms

4.2.1. Throughput of screening technologies

Various enzyme screening technologies are available (Fig. 5, middle top). In order to select the most suitable one for a particular screening campaign, the key parameters to be considered are the throughput and the technology availability. One of the most common screening technologies is the microtiter plate format, where individual variants are compartmentalized in separate wells. Thus, this method provides a simple and robust link between genotype and phenotype. Conventional microtiter plate readers offer the throughput of hundreds to thousands of variants per day. This number can be elevated further by miniaturization to 1536-well formats or microcapillaries [105] and robotization (liquid handling robots) to reach thousands to hundreds of thousand variants per day [101]. Millions of variants can be screened when moving from solid support to fluids. Widely available flow cytometers enable the throughput of up to hundreds of millions of variants per day by conducting fluorescence-activated cell sorting (FACS) [106] or more recent image-activated cell sorting [107]. FACS has been successfully employed to evolve arginine deiminase for stronger inhibition of tumour growth [108] or more recently to engineer phenylalanine ammonia-lyase used to treat phenylketonuria [109].

A comparable throughput of up to billions of variants per day has been achieved by microfluidic droplet-sorting systems, combining the versatility of traditional microtiter plate screening with the high throughput achieved by FACS [101]. The ultra-high-throughput nature of droplet sorting devices makes microfluidics a cutting-edge technology, which has already been demonstrated on every class of enzymes [110], including therapeutic enzymes. For example, tissue plasminogen activator mutants have been screened by a microfluidic fluorescence-activated droplet sorting (FADS) coupled with a retroviral display, resulting in greater than 1300-fold enrichment of the active wild-type enzyme [111]. When microfluidic sorters are coupled with next-generation sequencing, complex sequence-functional relationships can be revealed in deep mutational scanning (DMS) studies [112,113]. A recent example of pharmaceutically relevant microfluidic DMS has been demonstrated on dissecting the structural, functional and regulatory differences of the executioner caspases CASP3 and CASP7, which are involved in cell death and inflammation responses [114].

All screening campaigns ultimately aim for the maximum possible throughput to make this process both sample- and time-efficient. In this context, ultra-high-throughput microfluidic sorters are becoming one of the most promising techniques. However, a trade-off can be observed between the highest possible throughput offered and the simplicity of operation [115]. Therefore, the challenge for microfluidic engineers lies in leveraging the microfluidic sorting technology to a standard commercially available laboratory procedure, and the first operational solutions are already approaching the market [116]. With respect to the potential of droplet microfluidic systems acknowledged by the scientific community, the following paragraphs focus on microfluidic enzyme screening platforms.

4.2.2. Expression systems

The selection of a suitable expression system for the production of the enzymes (encoded in the starting library) depends mostly on the detection assay (Fig. 5, middle left). The screening campaign is much more affordable if the enzyme purification step can be avoided and whole cells or cell extracts can be employed [117]. If the substrate can easily diffuse into the cell (which is often the case of fluorogenic substrates), the screening can be carried out in the whole-cell mode and is suitable for microfluidic FACS [118]. When the substrate can be tethered to the cell surface, a display system can be applied, as demonstrated on bacteria [119] or yeast [120]. However, for most microfluidic screening campaigns, cells need to be lysed prior to assaying [121]. Efficient lysis can be done electrically [122], enzymatically [121], or by heating [123]. The selection of a suitable substrate remains critical for lysate-based screening, as compatibility with the native proteins and cellular metabolites of the expression host needs to be considered [117].

To eliminate the challenge of varying cellular transformation efficiency, *in vitro* transcription and translation directly provide the biosynthesis of proteins of interest from their encoding genes [124]. Cell-free protein expression is an attractive way to produce recombinant enzymes, especially when the enzyme is toxic to the host cell or unnatural amino acids are introduced [110]. The potential of this approach was recently demonstrated by a droplet-based *in vitro* evolution study [125], improving the protease Savinase (toxic to *E. coli)* to a 5-fold faster enzyme.

4.2.3. Assay and detection modes

Every enzyme screening needs a sensitive detection of the reaction product (Fig. 5, middle right). Droplet sorters most commonly employ optical detection methods [110]. There is a need to employ either a fluorogenic or a chromogenic substrate. Otherwise, the enzymatic activity needs to be monitored by a coupled assay with a photometric readout [101]. Fluorescence detection is widely applied in FADS [126]. Absorbance detection [127] substantially

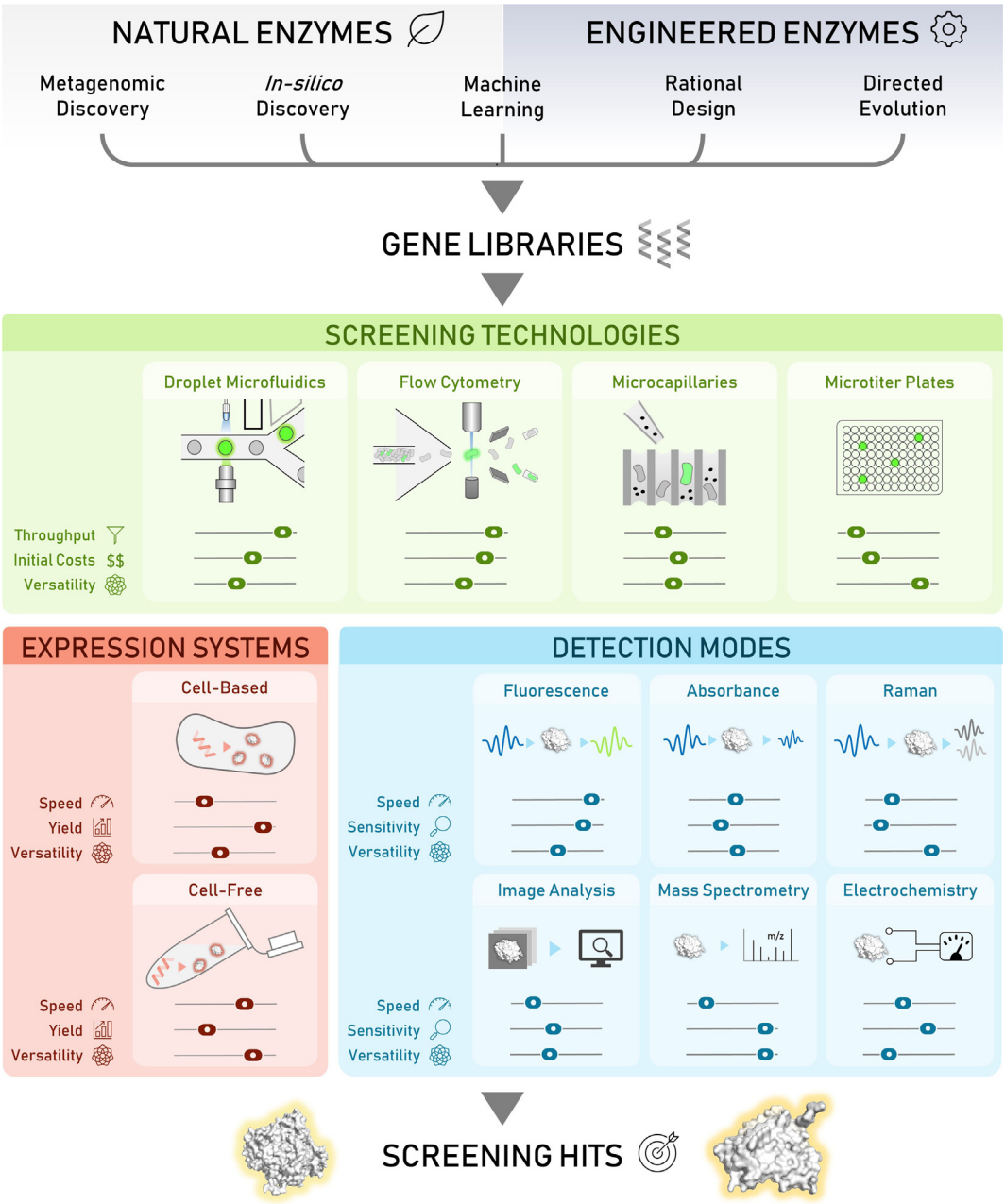

Fig. 5. Illustration of high-throughput screening of enzymes. Top: The library of genes encoding enzymes of interest can be generated from studies of both natural and engineered enzymes. More specifically, this starting pool of enzymes comes from studies such as metagenomic and *in-silico* discovery (natural enzymes) or rational design and directed evolution (engineered enzymes). Middle: The three panels guide the selection of suitable screening technology (green), expression system (red), and detection mode (blue). Each panel contains several options to choose from and parameters to evaluate their properties. Each parameter is described in the bottom left of each panel. The parameter score is depicted as a "slider" going from left to right (low to high, respectively). Bottom: The screening campaign results in enzyme variants, which require further characterization – see Fig. 6. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

extends the range of assays suitable for high-throughput screening. To reach a label-free detection mode, droplet microfluidic sorters have been successfully connected with the surface-enhanced Raman scattering [128], mass spectrometry [129], or electrochemical detection [130]. However, not all of these newly explored detection methods have been applied in an actual screening campaign [110].

4.3. Screening hits

When the link between genotype and phenotype is maintained, the result of the screening campaign is a set of genes encoding enzymes with desired or improved properties. Out of millions of variants in the libraries, the "hit" enzymes can be sorted without a precise understanding of how they work (i.e., without knowing the precise reaction mechanism). However, for the subsequent utilization of the hit biocatalysts, they first need to be biochemically characterized.

5. High-throughput characterization and validation of enzyme hits

The thorough biochemical characterization of the identified hits of a screening campaign is often the rate-limiting step, especially

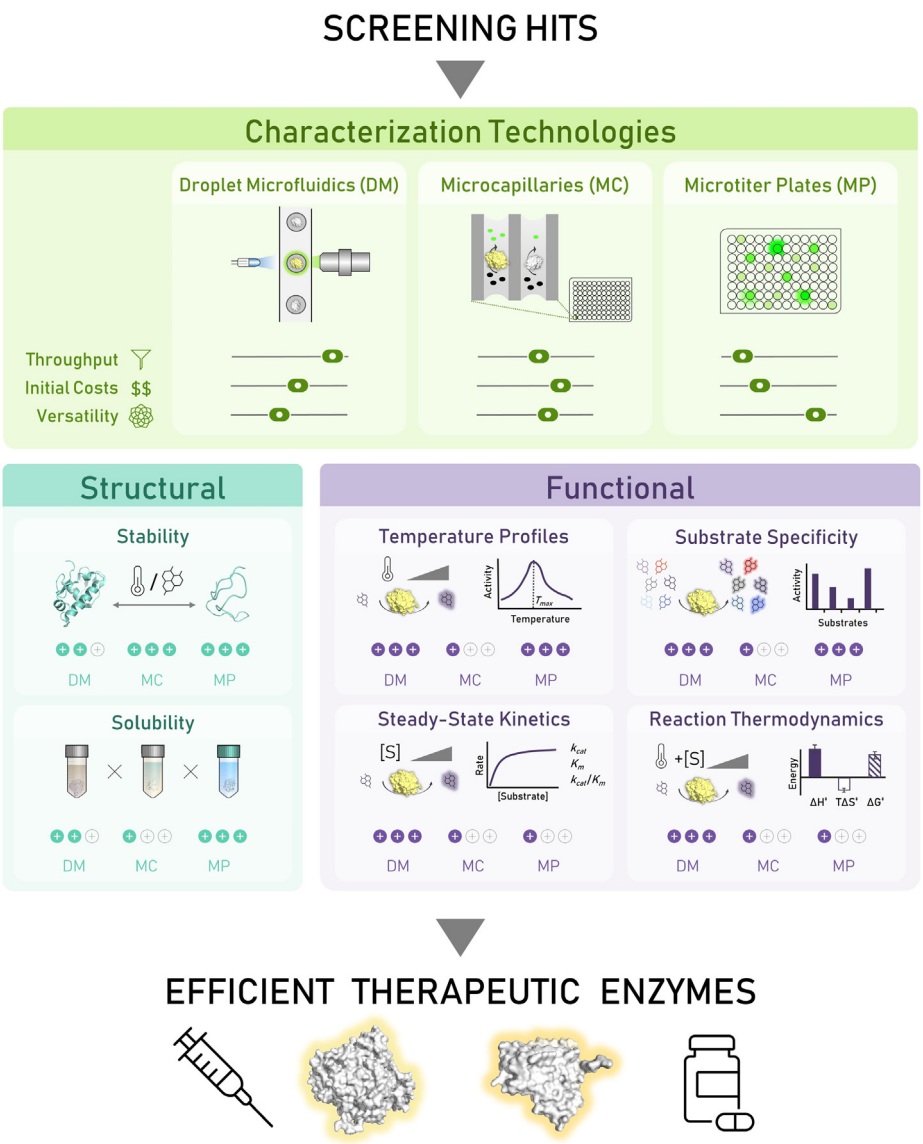

Fig. 6. Illustration of high-throughput characterization of enzymes. The hits from experimental or *in silico* screening need to be biochemically characterized. The three panels in the middle guide the selection of suitable characterization technologies (green) for structural (teal) and functional characterization (purple). Characterization technologies are described by parameters to evaluate their properties. Each parameter is described on the bottom left of each panel. The parameter score is depicted as a "slider" going from left to right (low to high, respectively). Both structural and functional panels highlight the key characteristics to be obtained for a target enzyme. In the bottom part of each characteristic, the applicability of each technique from the technology panel is depicted. DM = droplet microfluidics, MC = microcapillaries, MP = microtiter plates. One plus means potential applicability of the technique; two pluses denote the proof-of-concept studies using these techniques and three pluses stand for systematic application of the technique. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

when conventional biochemical technologies are used as they are low-throughput and time- and sample-demanding [131]. Similar to screening, new high-throughput experimental techniques are being developed to accelerate the biochemical characterization by miniaturization and automation [132]. The applications of microscale and microfluidic technologies for structural and functional characterization of enzymes (Fig. 6) are herein reviewed in the following two sections.

5.1. Structural stability characterization

Solubility and stability of the discovered or engineered enzymes are two crucial properties. Maintaining them ensures the protein product quality and guarantees the safety of therapeutic enzymes [133]. The critical structural characteristics to be investigated in a high-throughput manner are the enzyme solubility, stability and folding associated with secondary structure, as well as its quaternary structure or oligomerization state. High-throughput techniques to characterize protein solubility mainly depend on reporter-based assays, such as split-green fluorescence protein [134] or NanoLuc [135]. These can be fully integrated with microtiter plate formats, while recent coupling with DMS studies offers significantly higher throughput [136].

High-throughput technologies for enzyme stability characterization have been standardized to commercially available instruments, with the prominent example of capillary nano-DSF [137]. These technologies, combining intrinsic fluorescence measurement with static and dynamic light scattering, provide valuable information on thermal or chemical denaturation, ligand binding affinity, reaction buffer optimization, or protein aggregation [131]. Nano-DSF was utilized to characterise phenylalanine/tyrosine ammonia lyases obtained via ancestral sequence reconstruction [44], while

other therapeutic proteins were also characterized by this technique [138]. Some of the other benchmark techniques, including differential scanning calorimetry [139] and circular dichroism [140], have been recently adapted to higher throughput by miniaturization and microfluidics. These efforts are valuable, especially for circular dichroism, which is routinely performed to check proper enzyme folding and evaluate its secondary structure. Small-angle X-ray scattering (SAXS) is one of the methods to provide quaternary structure information and was also recently adapted to 96-well plate samplers [141] and microfluidic chips [142].

5.2. Functional characterization

Despite its benefits of time and sample efficiency, microfluidic systems have been surprisingly scarcely employed for the systematic kinetic characterization of enzymes [143]. To our best knowledge, microfluidic kinetic characterization of any therapeutic enzyme was demonstrated only by Baret et al. [119]. The authors tested bacterial expression systems and measured steady-state kinetics of L-asparaginase used as drugs in treating acute childhood lymphoblastic leukaemia. A significant step towards systematic functional characterization of enzymes was recently made by introducing a microfluidic platform named High-Throughput Microfluidic Enzyme Kinetics (HT-MEK) [144]. HT-MEK provides unprecedented capacities in the systematic kinetic and thermodynamic characterization of over a thousand enzyme mutants in parallel. By simultaneous *in vitro* expression and purification, more than 1500 mutants of a phosphate esterase PafA were subjected to multiple quantitative assays in days. For each mutant, the steady-state kinetic parameters, namely the rate constant (k_{cat}), Michaelis constant (K_m), and specificity constant (k_{cat}/K_m), were obtained for multiple substrates. Moreover, the inhibition constants (K_i) and the effects on folding were determined from over 670,000 reactions in total.

The team behind HT-MEK suggests that their device can be adapted to any enzyme system that can be tagged with a fluorescent protein (to monitor enzyme expression), expressed *in vitro,* and has a direct or coupled fluorogenic assay [144]. In particular, HT-MEK is applicable to detect inorganic phosphate (Pi), a compound readily soluble in water. However, many classes of enzymes perform valuable reactions on rather hydrophobic substrates, which are generally less suitable for droplet microfluidic systems due to their leakage to oil [145]. The challenge with the delivery of hydrophobic compounds into microfluidic droplets was recently addressed, where oil–water partitioning and microdialysis were utilized to deliver halogenated compounds for haloalkane dehalogenases within a microfluidic platform [146]. This concept enables the high-throughput characterization of dehalogenase activity, substrate specificity, temperature optima, and thermodynamics. This platform was systematically applied to characterize tens of dehalogenases within several different studies, including engineered variants [66,147] and enzymes discovered by *in silico* analysis [148].

Although steady-state analysis provides valuable initial information on the enzymatic reaction, the parameters obtained from such simplified models do not contain important mechanical understanding. However, the advanced transient kinetics providing such valuable information is extremely time-consuming and material-intensive. This demand has been recently addressed by a study featuring a droplet microfluidic platform, which was applied to assess the transient kinetics of three model enzymes in high throughput, namely β-galactosidase, horseradish peroxidase, and microperoxidase [149]. It was also used in a complex kinetic and thermodynamic study of engineered variants of haloalkane dehalogenases – all this in the throughput of 9000 reactions/min and with the reaction volumes reduced by as much as six orders of magnitude compared to conventional techniques.

Microfluidic enzyme characterization platforms have recently demonstrated their potential to meet the demands of the genomic era for the high-throughput collection of quantitative structural and functional characteristics of enzymes [132]. Similar to the prediction of enzyme structures via machine-learning-based AlphaFold2 [79], the functional data obtained by high-throughput microfluidic methods can facilitate progress in the long-term ambition of enzyme function prediction [132].

6. Conclusions and perspectives

Genomic and metagenomic databases serve as an ever-expanding supply of unexplored protein-encoding sequences. This natural diversity is a valuable source of potential new therapeutic enzymes. Alternatively, man-made diversity can be explored via protein engineering strategies, mainly by rational design and directed evolution. Structural bioinformatic methods enable efficient exploration of both of these sources of potential biocatalysts.

The accessibility of these computational methods used to be limited to specialized teams. Nowadays, web tools provide access to state-of-the-art bioinformatic methods without the need for tedious and time-consuming installations or special computational skills. Another advantage is that the server developers typically optimize parameters of individual computational tools and the end-users do not have to study every individual parameter to obtain meaningful results. Web tools are therefore appropriate for both experienced users and non-experts in the field.

The output of these bioinformatic methods is usually a pool of interesting enzyme-encoding genes, either promising newly discovered enzyme variants or suggested mutants to improve selected enzyme properties. Similarly, libraries of enzyme-encoding genes can be obtained from directed evolution or metagenomic studies. Microfluidics and FACS can efficiently screen these libraries and thus narrow down the number of potential therapeutic enzymes to the most promising candidates.

These hits from screening campaigns need to be further biochemically characterized to ensure their sufficient stability, solubility, and activity for biomedical applications. Microfluidics and deep mutational scanning are being increasingly utilized for the systematic characterization of various key enzyme properties. Furthermore, these high-throughput technologies can collect a desirable amount of high-quality data to feed machine learning algorithms.

We expect that machine learning algorithms will be further incorporated in both sequence- and structure-based bioinformatics. The prominent example of machine-learning achievements in this domain is the accurate prediction of tertiary protein structures by AlphaFold2. Thanks to tackling this long-term challenge in life sciences, models of many proteins potentially relevant as biopharmaceuticals have been predicted and deposited to the AlphaFold Protein Structure Database. The structures available in this steadily growing database represent an attractive source of information for further analysis.

In the coming years, new algorithms and computational workflows should be developed to assign a specific function and biological activity based on structural information. Such progress would move us closer to reaching another long-term challenge: the prediction of enzyme function from its sequence. Overall, the synergistic development of both advanced computational and high-throughput experimental methods will lead to higher chances of identification and characterization of valuable therapeutic enzymes.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to express their gratitude to the Czech Ministry of Education (INBIO - CZ.02.1.01/0.0/0.0/16_026/0008451, RECETOX research infrastructure no. LM2018121, ESFRI ELIXIR - LM2018131), and Czech Science Foundation (no. 20-15915Y) for financial support. This work has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 857560. This publication reflects only the authors' view and the European Commission is not responsible for any use that may be made of the information it contains. Michal Vasina acknowledges the financial support of his doctoral study by the scholarship Brno Ph.D. Talent and from Operational Programme Research, Development and Education-project "Internal Grant Agency of Masaryk University" (No.CZ.02.2.69/0.0/0.0/19_073/0016943).

References

[1] S. Tandon, A. Sharma, S. Singh, S. Sharma, S.J. Sarma, Therapeutic enzymes: discoveries, production and applications, J. Drug Deliv. Sci. Technol. 63 (2021) 102455, https://doi.org/10.1016/j.jddst.2021.102455.
[2] S. Bansal, K.S. Sangha, P. Khatri, Drug treatment of acute ischemic stroke, Am. J. Cardiovasc. Drugs Drugs Devices Interv. 13 (1) (2013) 57–69, https://doi.org/10.1007/s40256-013-0007-6.
[3] N. Labrou (Ed.), Therapeutic Enzymes: Function and Clinical Implications, Springer Singapore, Singapore, 2019. https://doi.org/10.1007/978-981-13-7709-9.
[4] N. Dellas, J. Liu, R.C. Botham, G.W. Huisman, Adapting protein sequences for optimized therapeutic efficacy, Curr. Opin. Chem. Biol. 64 (2021) 38–47, https://doi.org/10.1016/j.cbpa.2021.03.005.
[5] M. Kanehisa, Y. Sato, M. Kawashima, M. Furumichi, M. Tanabe, KEGG as a reference resource for gene and protein annotation, Nucleic Acids Res. 44 (D1) (2016) D457–D462, https://doi.org/10.1093/nar/gkv1070.
[6] Y. Moriya, D. Shigemizu, M. Hattori, T. Tokimatsu, M. Kotera, S. Goto, M. Kanehisa, PathPred: an enzyme-catalyzed metabolic pathway prediction server, Nucleic Acids Res. 38 (Web Server) (2010) W138–W143, https://doi.org/10.1093/nar/gkq318.
[7] B. Delépine, T. Duigou, P. Carbonell, J.-L. Faulon, RetroPath2.0: A retrosynthesis workflow for metabolic engineers, Metab. Eng. 45 (2018) 158–170, https://doi.org/10.1016/j.ymben.2017.12.002.
[8] C.A. Marchant, K.A. Briggs, A. Long, In silico tools for sharing data and knowledge on toxicity and metabolism: derek for windows, meteor, and vitic, Toxicol. Mech. Methods. 18 (2-3) (2008) 177–187, https://doi.org/10.1080/15376510701857320.
[9] N. Hadadi, J. Hafner, A. Shajkofci, A. Zisaki, V. Hatzimanikatis, ATLAS of biochemistry: a repository of all possible biochemical reactions for synthetic biology and metabolic engineering studies, ACS Synth. Biol. 5 (10) (2016) 1155–1166, https://doi.org/10.1021/acssynbio.6b00054.
[10] N. Hadadi, H. MohammadiPeyhani, L. Miskovic, M. Seijo, V. Hatzimanikatis, Enzyme annotation for orphan and novel reactions using knowledge of substrate reactive sites, Proc. Natl. Acad. Sci. 116 (15) (2019) 7298–7307, https://doi.org/10.1073/pnas.1818877116.
[11] M.H.S. Segler, M. Preuss, M.P. Waller, Planning chemical syntheses with deep neural networks and symbolic AI, Nature 555 (7698) (2018) 604–610, https://doi.org/10.1038/nature25978.
[12] B. Liu, B. Ramsundar, P. Kawthekar, J. Shi, J. Gomes, Q. Luu Nguyen, S. Ho, J. Sloane, P. Wender, V. Pande, Retrosynthetic reaction prediction using neural sequence-to-sequence models, ACS Cent Sci. 3 (2017) 1103–1113, https://doi.org/10.1021/acscentsci.7b00303.
[13] J.D. Tyzack, A.J.M. Ribeiro, N. Borkakoti, J.M. Thornton, Transform-MinER: transforming molecules in enzyme reactions, Bioinformatics 34 (2018) 3597–3599, https://doi.org/10.1093/bioinformatics/bty394.
[14] E.W. Sayers, E.E. Bolton, J.R. Brister, K. Canese, J. Chan, D.C. Comeau, R. Connor, K. Funk, C. Kelly, S. Kim, T. Madej, A. Marchler-Bauer, C. Lanczycki, S. Lathrop, Z. Lu, F. Thibaud-Nissen, T. Murphy, L. Phan, Y. Skripchenko, T. Tse, J. Wang, R. Williams, B.W. Trawick, K.D. Pruitt, S.T. Sherry, Database resources of the national center for biotechnology information, Nucleic Acids Res. (2021) gkab1112, https://doi.org/10.1093/nar/gkab1112.
[15] The UniProt Consortium, UniProt: a worldwide hub of protein knowledge, Nucleic Acids Res. 47 (2019) D506–D515, https://doi.org/10.1093/nar/gky1049.
[16] A.L. Mitchell, A. Almeida, M. Beracochea, M. Boland, J. Burgin, G. Cochrane, M. R. Crusoe, V. Kale, S.C. Potter, L.J. Richardson, E. Sakharova, M. Scheremetjew, A. Korobeynikov, A. Shlemov, O. Kunyavskaya, A. Lapidus, R.D. Finn, MGnify: the microbiome analysis resource in 2020, Nucleic Acids Res. (2019), https://doi.org/10.1093/nar/gkz1035.
[17] P. Vanacek, E. Sebestova, P. Babkova, S. Bidmanova, L. Daniel, P. Dvorak, V. Stepankova, R. Chaloupkova, J. Brezovsky, Z. Prokop, J. Damborsky, Exploration of enzyme diversity by integrating bioinformatics with expression analysis and biochemical characterization, ACS Catal. 8 (3) (2018) 2402–2412, https://doi.org/10.1021/acscatal.7b03523.
[18] F.A. Prayogo, A. Budiharjo, H.P. Kusumaningrum, W. Wijanarka, A. Suprihadi, N. Nurhayati, Metagenomic applications in exploration and development of novel enzymes from nature: a review, J. Genet. Eng. Biotechnol. 18 (2020) 39, https://doi.org/10.1186/s43141-020-00043-9.
[19] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (3) (1990) 403–410, https://doi.org/10.1016/S0022-2836(05)80360-2.
[20] C. Camacho, G. Coulouris, V. Avagyan, N. Ma, J. Papadopoulos, K. Bealer, T.L. Madden, BLAST+: architecture and applications, BMC Bioinf. 10 (2009) 421, https://doi.org/10.1186/1471-2105-10-421.
[21] M. Sajed, Sabeel un Naeem, N. Rashid, Chapter 13 - l-Asparaginases from hyperthermophilic archaea and their applications, in: M. Kuddus (Ed.), Microb. Extrem., Academic Press, 2022, pp. 177–184. https://doi.org/10.1016/B978-0-12-822945-3.00022-1.
[22] J. Mistry, R.D. Finn, S.R. Eddy, A. Bateman, M. Punta, Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions, Nucleic Acids Res. 41 (2013) e121–e121, https://doi.org/10.1093/nar/gkt263.
[23] K. Blin, S. Shaw, A.M. Kloosterman, Z. Charlop-Powers, G.P. van Wezel, M.H. Medema, T. Weber, antiSMASH 6.0: improving cluster detection and comparison capabilities, Nucleic Acids Res. 49 (2021) W29–W35, https://doi.org/10.1093/nar/gkab335.
[24] H. Zhang, T. Yohe, L. Huang, S. Entwistle, P. Wu, Z. Yang, P.K. Busk, Y. Xu, Y. Yin, dbCAN2: a meta server for automated carbohydrate-active enzyme annotation, Nucleic Acids Res. 46 (2018) W95–W101, https://doi.org/10.1093/nar/gky418.
[25] J. Mistry, S. Chuguransky, L. Williams, M. Qureshi, G.A. Salazar, E.L.L. Sonnhammer, S.C.E. Tosatto, L. Paladin, S. Raj, L.J. Richardson, R.D. Finn, A. Bateman, Pfam: The protein families database in 2021, Nucleic Acids Res. 49 (2021) D412–D419, https://doi.org/10.1093/nar/gkaa913.
[26] N. Nursimulu, L.L. Xu, J.D. Wasmuth, I. Krukov, J. Parkinson, Improved enzyme annotation with EC-specific cutoffs using DETECT v2, Bioinformatics 34 (2018) 3393–3395, https://doi.org/10.1093/bioinformatics/bty368.
[27] S.S. Hung, J. Wasmuth, C. Sanford, J. Parkinson, DETECT—a Density estimation tool for enzyme classification and its application to plasmodium falciparum, Bioinformatics 26 (2010) 1690–1698, https://doi.org/10.1093/bioinformatics/btq266.
[28] W.S. Mak, S. Tran, R. Marcheschi, S. Bertolani, J. Thompson, D. Baker, J.C. Liao, J.B. Siegel, Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway, Nat. Commun. 6 (1) (2015), https://doi.org/10.1038/ncomms10005.
[29] J.R. Marshall, P. Yao, S.L. Montgomery, J.D. Finnigan, T.W. Thorpe, R.B. Palmer, J. Mangas-Sanchez, R.A.M. Duncan, R.S. Heath, K.M. Graham, D.J. Cook, S.J. Charnock, N.J. Turner, Screening and characterization of a diverse panel of metagenomic imine reductases for biocatalytic reductive amination, Nat. Chem. 13 (2) (2021) 140–148, https://doi.org/10.1038/s41557-020-00606-w.
[30] J.N. Copp, E. Akiva, P.C. Babbitt, N. Tokuriki, Revealing unexplored sequence-function space using sequence similarity networks, Biochemistry 57 (31) (2018) 4651–4662, https://doi.org/10.1021/acs.biochem.8b00473.
[31] H.J. Atkinson, J.H. Morris, T.E. Ferrin, P.C. Babbitt, I.K. Jordan, Using sequence similarity networks for visualization of relationships across diverse protein superfamilies, PLoS ONE 4 (2) (2009) e4345, https://doi.org/10.1371/journal.pone.0004345.
[32] S.A. Memon, K.A. Khan, H. Naveed, HECNet: a hierarchical approach to enzyme function classification using a Siamese Triplet Network, Bioinformatics 36 (2020) 4583–4589, https://doi.org/10.1093/bioinformatics/btaa536.
[33] T. Zhang, Y. Tian, L. Yuan, F. Chen, A. Ren, Q.-N. Hu, Bio2Rxn: sequence-based enzymatic reaction predictions by a consensus strategy, Bioinformatics 36 (2020) 3600–3601, https://doi.org/10.1093/bioinformatics/btaa135.
[34] Y. Li, S. Wang, R. Umarov, B. Xie, M. Fan, L. Li, X. Gao, DEEPre: sequence-based enzyme EC number prediction by deep learning, Bioinformatics 34 (2018) 760–769, https://doi.org/10.1093/bioinformatics/btx680.
[35] A. Dalkiran, A.S. Rifaioglu, M.J. Martin, R. Cetin-Atalay, V. Atalay, T. Doğan, ECPred: a tool for the prediction of the enzymatic functions of protein sequences based on the EC nomenclature, BMC Bioinf. 19 (2018) 334, https://doi.org/10.1186/s12859-018-2368-y.
[36] J.Y. Ryu, H.U. Kim, S.Y. Lee, Deep learning enables high-quality and high-throughput prediction of enzyme commission numbers, Proc. Natl. Acad. Sci. 116 (28) (2019) 13996–14001, https://doi.org/10.1073/pnas.1821905116.
[37] J. Hon, S. Borko, J. Stourac, Z. Prokop, J. Zendulka, D. Bednar, T. Martinek, J. Damborsky, EnzymeMiner: automated mining of soluble enzymes with diverse structures, catalytic properties and stabilities, Nucleic Acids Res. 48 (2020) W104–W109, https://doi.org/10.1093/nar/gkaa372.
[38] A.J.M. Ribeiro, G.L. Holliday, N. Furnham, J.D. Tyzack, K. Ferris, J.M. Thornton, Mechanism and Catalytic Site Atlas (M-CSA): a database of enzyme reaction

mechanisms and active sites, Nucleic Acids Res. 46 (2018) D618–D623, https://doi.org/10.1093/nar/gkx1012.
[39] J. Hon, M. Marusiak, T. Martinek, A. Kunka, J. Zendulka, D. Bednar, J. Damborsky, SoluProt: prediction of soluble protein expression in *Escherichia coli*, Bioinformatics 37 (2021) 23–28, https://doi.org/10.1093/bioinformatics/btaa1102.
[40] D.L. Trudeau, M. Kaltenbach, D.S. Tawfik, On the potential origins of the high stability of reconstructed ancestral proteins, Mol. Biol. Evol. 33 (10) (2016) 2633–2641, https://doi.org/10.1093/molbev/msw138.
[41] M.A. Siddiq, G.K. Hochberg, J.W. Thornton, Evolution of protein specificity: insights from ancestral protein reconstruction, Curr. Opin. Struct. Biol. 47 (2017) 113–122, https://doi.org/10.1016/j.sbi.2017.07.003.
[42] R. Chaloupkova, V. Liskova, M. Toul, K. Markova, E. Sebestova, L. Hernychova, M. Marek, G.P. Pinto, D. Pluskal, J. Waterman, Z. Prokop, J. Damborsky, Light-emitting dehalogenases: reconstruction of multifunctional biocatalysts, ACS Catal. 9 (6) (2019) 4810–4823, https://doi.org/10.1021/acscatal.9b0103110.1021/acscatal.9b01031.s001.
[43] N.M. Hendrikse, A. Holmberg Larsson, S. Svensson Gelius, S. Kuprin, E. Nordling, P.-O. Syrén, Exploring the therapeutic potential of modern and ancestral phenylalanine/tyrosine ammonia-lyases as supplementary treatment of hereditary tyrosinemia, Sci. Rep. 10 (1) (2020), https://doi.org/10.1038/s41598-020-57913-y.
[44] N.M. Hendrikse, A. Sandegren, T. Andersson, J. Blomqvist, Å. Makower, D. Possner, C. Su, N. Thalén, A. Tjernberg, U. Westermark, J. Rockberg, S. Svensson Gelius, P.-O. Syrén, E. Nordling, Ancestral lysosomal enzymes with increased activity harbor therapeutic potential for treatment of Hunter syndrome, IScience. 24 (3) (2021) 102154, https://doi.org/10.1016/j.isci.2021.102154.
[45] M.A. Spence, J.A. Kaczmarski, J.W. Saunders, C.J. Jackson, Ancestral sequence reconstruction for protein engineers, Curr. Opin. Struct. Biol. 69 (2021) 131–141, https://doi.org/10.1016/j.sbi.2021.04.001.
[46] M. Musil, R.T. Khan, A. Beier, J. Stourac, H. Konegger, J. Damborsky, D. Bednar, FireProtASR: a web server for fully automated ancestral sequence reconstruction, Brief. Bioinform. 22 (2021) bbaa337, https://doi.org/10.1093/bib/bbaa337.
[47] S.M. Marques, J. Planas-Iglesias, J. Damborsky, Web-based tools for computational enzyme design, Curr. Opin. Struct. Biol. 69 (2021) 19–34, https://doi.org/10.1016/j.sbi.2021.01.010.
[48] J. Planas-Iglesias, S.M. Marques, G.P. Pinto, M. Musil, J. Stourac, J. Damborsky, D. Bednar, Computational design of enzymes for biotechnological applications, Biotechnol. Adv. 47 (2021) 107696, https://doi.org/10.1016/j.biotechadv.2021.107696.
[49] M. Musil, H. Konegger, J. Hon, D. Bednar, J. Damborsky, Computational design of stable and soluble biocatalysts, ACS Catal. 9 (2) (2019) 1033–1054, https://doi.org/10.1021/acscatal.8b03613.
[50] C.E. Sequeiros-Borja, B. Surpeta, J. Brezovsky, Recent advances in user-friendly computational tools to engineer protein function, Brief. Bioinform. 22 (2021) bbaa150, https://doi.org/10.1093/bib/bbaa150.
[51] J. Weinstein, O. Khersonsky, S.J. Fleishman, Practically useful protein-design methods combining phylogenetic and atomistic calculations, Curr. Opin. Struct. Biol. 63 (2020) 58–64, https://doi.org/10.1016/j.sbi.2020.04.003.
[52] R. Moretti, S. Lyskov, R. Das, J. Meiler, J.J. Gray, Web-accessible molecular modeling with Rosetta: the rosetta online server that includes everyone (ROSIE), Protein Sci. Publ. Protein Soc. 27 (1) (2018) 259–268, https://doi.org/10.1002/pro.3313.
[53] Y. Liu, B. Kuhlman, RosettaDesign server for protein design, Nucleic Acids Res. 34 (Web Server) (2006) W235–W238, https://doi.org/10.1093/nar/gkl163.
[54] G. Dantas, B. Kuhlman, D. Callender, M. Wong, D. Baker, A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins, J. Mol. Biol. 332 (2) (2003) 449–460, https://doi.org/10.1016/S0022-2836(03)00888-X.
[55] L. Sumbalova, J. Stourac, T. Martinek, D. Bednar, J. Damborsky, HotSpot Wizard 3.0: web server for automated design of mutations and smart libraries based on sequence input information, Nucleic Acids Res. 46 (2018) W356–W362, https://doi.org/10.1093/nar/gky417.
[56] P.M. Swe, J.N. Copp, L.K. Green, C.P. Guise, A.M. Mowday, J.B. Smaill, A.V. Patterson, D.F. Ackerley, Targeted mutagenesis of the Vibrio fischeri flavin reductase FRase I to improve activation of the anticancer prodrug CB1954, Biochem. Pharmacol. 84 (6) (2012) 775–783, https://doi.org/10.1016/j.bcp.2012.07.002.
[57] O. Khersonsky, R. Lipsh, Z. Avizemer, Y. Ashani, M. Goldsmith, H. Leader, O. Dym, S. Rogotner, D.L. Trudeau, J. Prilusky, P. Amengual-Rigo, V. Guallar, D.S. Tawfik, S.J. Fleishman, Automated design of efficient and functionally diverse enzyme repertoires, Mol. Cell. 72 (1) (2018) 178–186.e5, https://doi.org/10.1016/j.molcel.2018.08.033.
[58] C.M. VanDrisse, R. Lipsh-Sokolik, O. Khersonsky, S.J. Fleishman, D.K. Newman, Computationally designed pyocyanin demethylase acts synergistically with tobramycin to kill recalcitrant Pseudomonas aeruginosa biofilms, Proc. Natl. Acad. Sci. 118 (2021), https://doi.org/10.1073/pnas.2022012118.
[59] P. Kokkonen, D. Bednar, G. Pinto, Z. Prokop, J. Damborsky, Engineering enzyme access tunnels, Biotechnol. Adv. 37 (6) (2019) 107386, https://doi.org/10.1016/j.biotechadv.2019.04.008.
[60] J. Stourac, O. Vavra, P. Kokkonen, J. Filipovic, G. Pinto, J. Brezovsky, J. Damborsky, D. Bednar, Caver Web 1.0: identification of tunnels and channels in proteins and analysis of ligand transport, Nucleic Acids Res. 47 (2019) W414–W422, https://doi.org/10.1093/nar/gkz378.
[61] O. Vavra, J. Filipovic, J. Plhak, D. Bednar, S.M. Marques, J. Brezovsky, J. Stourac, L. Matyska, J. Damborsky, CaverDock: a molecular docking-based tool to analyse ligand transport through protein tunnels and channels, Bioinformatics 35 (2019) 4986–4993, https://doi.org/10.1093/bioinformatics/btz386.
[62] D.O. Deay, S. Seibold, K.P. Battaile, S. Lovell, M.L. Richter, P.A. Petillo, Improving the kinetic parameters of nicotine oxidizing enzymes by homologous structure comparison and rational design, Arch. Biochem. Biophys. 718 (2022) 109122, https://doi.org/10.1016/j.abb.2022.109122.
[63] E. Papaleo, G. Saladino, M. Lambrughi, K. Lindorff-Larsen, F.L. Gervasio, R. Nussinov, The role of protein loops and linkers in conformational dynamics and allostery, Chem. Rev. 116 (11) (2016) 6391–6423, https://doi.org/10.1021/acs.chemrev.5b00623.
[64] C. Moreira, A.R. Calixto, J.P. Richard, S.C.L. Kamerlin, The role of ligand-gated conformational changes in enzyme catalysis, Biochem. Soc. Trans. 47 (2019) 1449–1460, https://doi.org/10.1042/BST20190298.
[65] J. Planas-Iglesias, F. Opaleny, P. Ulbrich, J. Stourac, Z. Sanusi, G.P. Pinto, A. Schenkmayerova., J. Byska, J. Damborsky, B. Kozlikova, D. Bednar, LoopGrafter: A Web Tool for Transplanting Dynamical Loops for Protein Engineering. In preparation, (2022).
[66] A. Schenkmayerova, G.P. Pinto, M. Toul, M. Marek, L. Hernychova, J. Planas-Iglesias, V. Daniel Liskova, D. Pluskal, M. Vasina, S. Emond, M. Dörr, R. Chaloupkova, D. Bednar, Z. Prokop, F. Hollfelder, U.T. Bornscheuer, J. Damborsky, Engineering the protein dynamics of an ancestral luciferase, Nat. Commun. 12 (2021) 3616, https://doi.org/10.1038/s41467-021-23450-z.
[67] A. Goldenzweig, M. Goldsmith, S. Hill, O.r. Gertman, P. Laurino, Y. Ashani, O. Dym, T. Unger, S. Albeck, J. Prilusky, R. Lieberman, A. Aharoni, I. Silman, J. Sussman, D. Tawfik, S. Fleishman, Automated structure- and sequence-based design of proteins for high bacterial expression and stability, Mol. Cell. 63 (2) (2016) 337–346, https://doi.org/10.1016/j.molcel.2016.06.012.
[68] T.A. Whitehead, A. Chevalier, Y. Song, C. Dreyfus, S.J. Fleishman, C. De Mattos, C.A. Myers, H. Kamisetty, P. Blair, I.A. Wilson, D. Baker, Optimization of affinity, specificity and function of designed influenza inhibitors using deep sequencing, Nat. Biotechnol. 30 (6) (2012) 543–548, https://doi.org/10.1038/nbt.2214.
[69] Y. Peleg, R. Vincentelli, B.M. Collins, K.-E. Chen, E.K. Livingstone, S. Weeratunga, N. Leneva, Q. Guo, K. Remans, K. Perez, G.E.K. Bjerga, Ø. Larsen, O. Vaněk, O. Skořepa, S. Jacquemin, A. Poterszman, S. Kjær, E. Christodoulou, S. Albeck, O. Dym, E. Ainbinder, T. Unger, A. Schuetz, S. Matthes, M. Bader, A. de Marco, P. Storici, M.S. Semrau, P. Stolt-Bergner, C. Aigner, S. Suppmann, A. Goldenzweig, S.J. Fleishman, Community-wide experimental evaluation of the PROSS stability-design method, J. Mol. Biol. 433 (13) (2021) 166964, https://doi.org/10.1016/j.jmb.2021.166964.
[70] M.-Y. Li, P.-L. Lai, Y.-T. Chou, A.-P. Chi, Y.-Z. Mi, K.-H. Khoo, G.-D. Chang, C.-W. Wu, T.-C. Meng, G.-C. Chen, Protein tyrosine phosphatase PTPN3 inhibits lung cancer cell proliferation and migration by promoting EGFR endocytic degradation, Oncogene 34 (29) (2015) 3791–3803, https://doi.org/10.1038/onc.2014.312.
[71] M. Musil, J. Stourac, J. Bendl, J. Brezovsky, Z. Prokop, J. Zendulka, T. Martinek, D. Bednar, J. Damborsky, FireProt: web server for automated design of thermostable proteins, Nucleic Acids Res. 45 (2017) W393–W399, https://doi.org/10.1093/nar/gkx285.
[72] D. Bednar, K. Beerens, E. Sebestova, J. Bendl, S. Khare, R. Chaloupkova, Z. Prokop, J. Brezovsky, D. Baker, J. Damborsky, N.V. Dokholyan, FireProt: Energy- and evolution-based computational design of thermostable multiple-point mutants, PLOS Comput. Biol. 11 (11) (2015) e1004556, https://doi.org/10.1371/journal.pcbi.1004556.
[73] J. Van Durme, G. De Baets, R. Van Der Kant, M. Ramakers, A. Ganesan, H. Wilkinson, R. Gallardo, F. Rousseau, J. Schymkowitz, Solubis: a webserver to reduce protein aggregation through mutation, Protein Eng. Des. Sel. 29 (8) (2016) 285–289, https://doi.org/10.1093/protein/gzw019.
[74] R. van der Kant, J. van Durme, F. Rousseau, J. Schymkowitz, SolubiS: Optimizing protein solubility by minimal point mutations, in: C.M. Gomes (Ed.), Protein Misfolding Dis., Springer New York, New York, NY, 2019, pp. 317–333. https://doi.org/10.1007/978-1-4939-8820-4_21.
[75] A. Ganesan, A. Siekierska, J. Beerten, M. Brams, J. Van Durme, G. De Baets, R. Van der Kant, R. Gallardo, M. Ramakers, T. Langenberg, H. Wilkinson, F. De Smet, C. Ulens, F. Rousseau, J. Schymkowitz, Structural hot spots for the solubility of globular proteins, Nat. Commun. 7 (2016) 10816, https://doi.org/10.1038/ncomms10816.
[76] A. Kuriata, V. Iglesias, J. Pujols, M. Kurcinski, S. Kmiecik, S. Ventura, Aggrescan3D (A3D) 2.0: prediction and engineering of protein solubility, Nucleic Acids Res. 47 (2019) W300–W307, https://doi.org/10.1093/nar/gkz321.
[77] J. Lever, M. Krzywinski, N. Altman, Model selection and overfitting, Nat. Methods. 13 (9) (2016) 703–704, https://doi.org/10.1038/nmeth.3968.
[78] A.W. Senior, R. Evans, J. Jumper, J. Kirkpatrick, L. Sifre, T. Green, C. Qin, A. Žídek, A.W.R. Nelson, A. Bridgland, H. Penedones, S. Petersen, K. Simonyan, S. Crossan, P. Kohli, D.T. Jones, D. Silver, K. Kavukcuoglu, D. Hassabis, Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13), Proteins. 87 (12) (2019) 1141–1148, https://doi.org/10.1002/prot.25834.
[79] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A. A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M.

Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature 596 (7873) (2021) 583–589, https://doi.org/10.1038/s41586-021-03819-2.
[80] A. Kryshtafovych, T. Schwede, M. Topf, K. Fidelis, J. Moult, Critical assessment of methods of protein structure prediction (CASP)—Round XIII, Proteins Struct. Funct. Bioinforma. 87 (12) (2019) 1011–1020, https://doi.org/10.1002/prot.25823.
[81] D. Hassabis, Introducing Isomorphic Labs, (2021). https://www.isomorphiclabs.com/blog (accessed December 16, 2021).
[82] S. Smith, 43 Pharma Companies Using Artificial Intelligence in Drug Discovery, (2021). https://blog.benchsci.com/pharma-companies-using-artificial-intelligence-in-drug-discovery (accessed November 19, 2021).
[83] S. Smith, 230 Startups Using Artificial Intelligence in Drug Discovery, (2021). https://blog.benchsci.com/startups-using-artificial-intelligence-in-drug-discovery (accessed December 16, 2021).
[84] A. Marabotti, B. Scafuri, A. Facchiano, Predicting the stability of mutant proteins by computational approaches: an overview, Brief. Bioinform. 22 (2021), https://doi.org/10.1093/bib/bbaa074.
[85] H. Cao, J. Wang, L. He, Y. Qi, J.Z. Zhang, DeepDDG: predicting the stability change of protein point mutations using neural networks, J. Chem. Inf. Model. 59 (4) (2019) 1508–1514, https://doi.org/10.1021/acs.jcim.8b00697.
[86] C.H.M. Rodrigues, D.E.V. Pires, D.B. Ascher, DynaMut2: Assessing changes in stability and flexibility upon single and multiple point missense mutations, Protein Sci. 30 (2021) 60–69, https://doi.org/10.1002/pro.3942.
[87] Y. Tian, C. Deutsch, B. Krishnamoorthy, Scoring function to predict solubility mutagenesis, Algorithms Mol. Biol. 5 (2010) 33, https://doi.org/10.1186/1748-7188-5-33.
[88] Y. Yang, L. Zeng, M. Vihinen, PON-Sol2: prediction of effects of variants on protein solubility, Int. J. Mol. Sci. 22 (2021) 8027, https://doi.org/10.3390/ijms22158027.
[89] S. Liang, D. Zheng, D.M. Standley, B. Yao, M. Zacharias, C. Zhang, EPSVR and EPMeta: prediction of antigenic epitopes using support vector regression and multiple server results, BMC Bioinf. 11 (2010) 381, https://doi.org/10.1186/1471-2105-11-381.
[90] S.N. Crooke, I.G. Ovsyannikova, R.B. Kennedy, G.A. Poland, Immunoinformatic identification of B cell and T cell epitopes in the SARS-CoV-2 proteome, Sci. Rep. 10 (2020) 14179, https://doi.org/10.1038/s41598-020-70864-8.
[91] A. Jain, D. Kihara, NNTox: gene ontology-based protein toxicity prediction using neural network, Sci. Rep. 9 (2019) 17923, https://doi.org/10.1038/s41598-019-54405-6.
[92] Y. Wang, F. Li, M. Bharathwaj, N.C. Rosas, A. Leier, T. Akutsu, G.I. Webb, T.T. Marquez-Lago, J. Li, T. Lithgow, J. Song, DeepBL: a deep learning-based approach for *in silico* discovery of beta-lactamases, Brief. Bioinform. 22 (2021) bbaa301, https://doi.org/10.1093/bib/bbaa301.
[93] J. Stourac, J. Dubrava, M. Musil, J. Horackova, J. Damborsky, S. Mazurenko, D. Bednar, FireProtDB: database of manually curated protein stability data, Nucleic Acids Res. 49 (2021) D319–D324, https://doi.org/10.1093/nar/gkaa981.
[94] J. Velecký, J. Stourac, M. Jankujova, M. Musil, J. Damborsky, D. Bednar, S. Mazurenko, SoluProtMutDB: Manually Curated Database of Protein Solubility Changes upon Mutation. In preparation, 2022.
[95] X. Wang, X. Zhang, C. Peng, Y. Shi, H. Li, Z. Xu, W. Zhu, D3DistalMutation: a database to explore the effect of distal mutations on enzyme activity, J. Chem. Inf. Model. 61 (5) (2021) 2499–2508, https://doi.org/10.1021/acs.jcim.1c00318.
[96] A. Kulandaisamy, R. Sakthivel, M.M. Gromiha, MPTherm: database for membrane protein thermodynamics for understanding folding and stability, Brief. Bioinform. 22 (2021) 2119–2125, https://doi.org/10.1093/bib/bbaa064.
[97] J.D. Westbrook, S.K. Burley, How structural biologists and the protein data bank contributed to recent FDA new drug approvals, Structure 27 (2) (2019) 211–217, https://doi.org/10.1016/j.str.2018.11.007.
[98] L.A. Rabia, A.A. Desai, H.S. Jhajj, P.M. Tessier, Understanding and overcoming trade-offs between antibody affinity, specificity, stability and solubility, Biochem. Eng. J. 137 (2018) 365–374, https://doi.org/10.1016/j.bej.2018.06.003.
[99] J.R. Klesmith, J.-P. Bacik, E.E. Wrenbeck, R. Michalczyk, T.A. Whitehead, Trade-offs between enzyme fitness and solubility illuminated by deep mutational scanning, Proc. Natl. Acad. Sci. 114 (9) (2017) 2265–2270, https://doi.org/10.1073/pnas.1614437114.
[100] F.H. Arnold, Innovation by evolution: bringing new chemistry to life (nobel lecture), Angew. Chem. Int. Ed. 58 (41) (2019) 14420–14426, https://doi.org/10.1002/anie.201907729.
[101] H.A. Bunzel, X. Garrabou, M. Pott, D. Hilvert, Speeding up enzyme discovery and engineering with ultrahigh-throughput methods, Curr. Opin. Struct. Biol. 48 (2018) 149–156, https://doi.org/10.1016/j.sbi.2017.12.010.
[102] D.L. Trudeau, D.S. Tawfik, Protein engineers turned evolutionists—the quest for the optimal starting point, Curr. Opin. Biotechnol. 60 (2019) 46–52, https://doi.org/10.1016/j.copbio.2018.12.002.
[103] L. Alejaldre, J.N. Pelletier, D. Quaglia, Methods for enzyme library creation: Which one will you choose? A guide for novices and experts to introduce genetic diversity, BioEssays 43 (8) (2021) 2100052, https://doi.org/10.1002/bies.v43.810.1002/bies.202100052.
[104] S.L. Robinson, J. Piel, S. Sunagawa, A roadmap for metagenomic enzyme discovery, Nat. Prod. Rep. 38 (11) (2021) 1994–2023, https://doi.org/10.1039/D1NP00006C.
[105] B. Chen, S. Lim, A. Kannan, S.C. Alford, F. Sunden, D. Herschlag, I.K. Dimov, T. M. Baer, J.R. Cochran, High-throughput analysis and protein engineering using microcapillary arrays, Nat. Chem. Biol. 12 (2) (2016) 76–81, https://doi.org/10.1038/nchembio.1978.
[106] N. Varadarajan, J.R. Cantor, G. Georgiou, B.L. Iverson, Construction and flow cytometric screening of targeted enzyme libraries, Nat. Protoc. 4 (6) (2009) 893–901, https://doi.org/10.1038/nprot.2009.60.
[107] A. Isozaki, H. Mikami, H. Tezuka, H. Matsumura, K. Huang, M. Akamine, K. Hiramatsu, T. Iino, T. Ito, H. Karakawa, Y. Kasai, Y. Li, Y. Nakagawa, S. Ohnuki, T. Ota, Y. Qian, S. Sakuma, T. Sekiya, Y. Shirasaki, N. Suzuki, E. Tayyabi, T. Wakamiya, M. Xu, M. Yamagishi, H. Yan, Q. Yu, S. Yan, D. Yuan, W. Zhang, Y. Zhao, F. Arai, R.E. Campbell, C. Danelon, D. Di Carlo, K. Hiraki, Y.u. Hoshino, Y. Hosokawa, M. Inaba, A. Nakagawa, Y. Ohya, M. Oikawa, S. Uemura, Y. Ozeki, T. Sugimura, N. Nitta, K. Goda, Intelligent image-activated cell sorting 2.0, Lab. Chip. 20 (13) (2020) 2263–2273, https://doi.org/10.1039/D0LC00080A.
[108] F. Cheng, T. Kardashliev, C. Pitzler, A. Shehzad, H. Lue, J. Bernhagen, L. Zhu, U. Schwaneberg, A competitive flow cytometry screening system for directed evolution of therapeutic enzyme, ACS Synth. Biol. 4 (7) (2015) 768–775, https://doi.org/10.1021/sb500343g.
[109] K.J. Adolfsen, I. Callihan, C.E. Monahan, P.J. Greisen, J. Spoonamore, M. Momin, L.E. Fitch, M.J. Castillo, L. Weng, L. Renaud, C.J. Weile, J.H. Konieczka, T. Mirabella, A. Abin-Fuentes, A.G. Lawrence, V.M. Isabella, Improvement of a synthetic live bacterial therapeutic for phenylketonuria with biosensor-enabled enzyme engineering, Nat. Commun. 12 (2021) 6215, https://doi.org/10.1038/s41467-021-26524-0.
[110] S. Neun, P.J. Zurek, T.S. Kaminski, F. Hollfelder, Chapter Thirteen - Ultrahigh throughput screening for enzyme function in droplets, in: D.S. Tawfik (Ed.), Methods Enzymol, Academic Press, 2020, pp. 317–343, https://doi.org/10.1016/bs.mie.2020.06.002.
[111] L. Granieri, J.-C. Baret, A.D. Griffiths, C.A. Merten, High-throughput screening of enzymes by retroviral display using droplet-based microfluidics, Chem. Biol. 17 (3) (2010) 229–235, https://doi.org/10.1016/j.chembiol.2010.02.011.
[112] P.A. Romero, T.M. Tran, A.R. Abate, Dissecting enzyme function with microfluidic-based deep mutational scanning, Proc. Natl. Acad. Sci. 112 (23) (2015) 7159–7164, https://doi.org/10.1073/pnas.1422285112.
[113] D.M. Fowler, S. Fields, Deep mutational scanning: a new style of protein science, Nat. Methods. 11 (8) (2014) 801–807, https://doi.org/10.1038/nmeth.3027.
[114] H. Roychowdury, P.A. Romero, Microfluidic deep mutational scanning of the human executioner caspases reveals differences in structure and regulation, Cell Death Discov. 8 (2022) 1–8, https://doi.org/10.1038/s41420-021-00799-0.
[115] Y.V. Sheludko, W.-D. Fessner, Winning the numbers game in enzyme evolution – fast screening methods for improved biotechnology proteins, Curr. Opin. Struct. Biol. 63 (2020) 123–133, https://doi.org/10.1016/j.sbi.2020.05.003.
[116] C. Frey, J. Pfeil, T. Neckernuss, D. Geiger, K. Weishaupt, I. Platzman, O. Marti, J. P. Spatz, Label-free monitoring and manipulation of microfluidic water-in-oil droplets, View. 1 (4) (2020) 20200101, https://doi.org/10.1002/viw2.v1.410.1002/VIW.20200101.
[117] R.A. Sheldon, J.M. Woodley, Role of biocatalysis in sustainable chemistry, Chem. Rev. 118 (2) (2018) 801–838, https://doi.org/10.1021/acs.chemrev.7b00203.
[118] A. Dalili, E. Samiei, M. Hoorfar, A review of sorting, separation and isolation of cells and microbeads for biomedical applications: microfluidic approaches, The Analyst. 144 (1) (2019) 87–113, https://doi.org/10.1039/C8AN01061G.
[119] C.S. Karamitros, M. Morvan, A. Vigne, J. Lim, P. Gruner, T. Beneyton, J. Vrignon, J.-C. Baret, Bacterial expression systems for enzymatic activity in droplet-based microfluidics, Anal. Chem. 92 (7) (2020) 4908–4916, https://doi.org/10.1021/acs.analchem.9b04969.
[120] J.J. Agresti, E. Antipov, A.R. Abate, K. Ahn, A.C. Rowat, J.-C. Baret, M. Marquez, A.M. Klibanov, A.D. Griffiths, D.A. Weitz, Ultrahigh-throughput screening in drop-based microfluidics for directed evolution, Proc. Natl. Acad. Sci. 107 (9) (2010) 4004–4009, https://doi.org/10.1073/pnas.0910781107.
[121] B. Kintses, C. Hein, M. Mohamed, M. Fischlechner, F. Courtois, C. Lainé, F. Hollfelder, Picoliter cell lysate assays in microfluidic droplet compartments for directed enzyme evolution, Chem. Biol. 19 (8) (2012) 1001–1009, https://doi.org/10.1016/j.chembiol.2012.06.009.
[122] N. de Lange, T.M. Tran, A.R. Abate, Electrical lysis of cells for detergent-free droplet assays, Biomicrofluidics 10 (2) (2016) 024114, https://doi.org/10.1063/1.4944742.
[123] A.C. Larsen, M.R. Dunn, A. Hatch, S.P. Sau, C. Youngbull, J.C. Chaput, A general strategy for expanding polymerase function by droplet microfluidics, Nat. Commun. 7 (2016) 11235, https://doi.org/10.1038/ncomms11235.
[124] R. Galinis, G. Stonyte, V. Kiseliovas, R. Zilionis, S. Studer, D. Hilvert, A. Janulaitis, L. Mazutis, DNA nanoparticles for improved protein synthesis in vitro, Angew. Chem. Int. Ed. 55 (9) (2016) 3120–3123, https://doi.org/10.1002/anie.201511809.
[125] J.M. Holstein, C. Gylstorff, F. Hollfelder, Cell-free directed evolution of a protease in microdroplets at ultrahigh throughput, ACS Synth. Biol. 10 (2) (2021) 252–257, https://doi.org/10.1021/acssynbio.0c00538.
[126] J.-C. Baret, O.J. Miller, V. Taly, M. Ryckelynck, A. El-Harrak, L. Frenz, C. Rick, M. L. Samuels, J.B. Hutchison, J.J. Agresti, D.R. Link, D.A. Weitz, A.D. Griffiths, Fluorescence-activated droplet sorting (FADS): efficient microfluidic cell sorting based on enzymatic activity, Lab. Chip. 9 (13) (2009) 1850, https://doi.org/10.1039/b902504a.

[127] F. Gielen, R. Hours, S. Emond, M. Fischlechner, U. Schell, F. Hollfelder, Ultrahigh-throughput–directed enzyme evolution by absorbance-activated droplet sorting (AADS), Proc. Natl. Acad. Sci. 113 (47) (2016) E7383–E7389, https://doi.org/10.1073/pnas.1606927113.
[128] M. Hassoun, J. Rüger, T. Kirchberger-Tolstik, I.W. Schie, T. Henkel, K. Weber, D. Cialla-May, C. Krafft, J. Popp, A droplet-based microfluidic chip as a platform for leukemia cell lysate identification using surface-enhanced Raman scattering, Anal. Bioanal. Chem. 410 (3) (2018) 999–1006, https://doi.org/10.1007/s00216-017-0609-y.
[129] D.A. Holland-Moritz, M.K. Wismer, B.F. Mann, I. Farasat, P. Devine, E.D. Guetschow, I. Mangion, C.J. Welch, J.C. Moore, S. Sun, R.T. Kennedy, Mass activated droplet sorting (MADS) enables high-throughput screening of enzymatic reactions at nanoliter scale, Angew. Chem. Int. Ed. 59 (11) (2020) 4470–4477, https://doi.org/10.1002/anie.201913203.
[130] H. Goto, Y. Kanai, A. Yotsui, S. Shimokihara, S. Shitara, R. Oyobiki, K. Fujiwara, T. Watanabe, Y. Einaga, Y. Matsumoto, N. Miki, N. Doi, Microfluidic screening system based on boron-doped diamond electrodes and dielectrophoretic sorting for directed evolution of NAD(P)-dependent oxidoreductases, Lab. Chip. 20 (4) (2020) 852–861, https://doi.org/10.1039/C9LC01263J.
[131] M. Vasina, P. Vanacek, J. Damborsky, Z. Prokop, Chapter Three - Exploration of enzyme diversity: High-throughput techniques for protein production and microscale biochemical characterization, in: D.S. Tawfik (Ed.), Methods Enzymol, Academic Press, 2020, pp. 51–85, https://doi.org/10.1016/bs.mie.2020.05.004.
[132] D.A. Mokhtari, M.J. Appel, P.M. Fordyce, D. Herschlag, High throughput and quantitative enzymology in the genomic era, Curr. Opin. Struct. Biol. 71 (2021) 259–273, https://doi.org/10.1016/j.sbi.2021.07.010.
[133] S.A. Berkowitz, D.J. Houde, The complexity of protein structure and the challenges it poses in developing biopharmaceuticals, in: Biophys. Charact. Proteins Dev. Biopharm, Elsevier, 2015, pp. 1–21, https://doi.org/10.1016/B978-0-444-59573-7.00001-4.
[134] S. Cabantous, T.C. Terwilliger, G.S. Waldo, Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein, Nat. Biotechnol. 23 (1) (2005) 102–107, https://doi.org/10.1038/nbt1044.
[135] A.S. Dixon, M.K. Schwinn, M.P. Hall, K. Zimmerman, P. Otto, T.H. Lubben, B.L. Butler, B.F. Binkowski, T. Machleidt, T.A. Kirkland, M.G. Wood, C.T. Eggers, L.P. Encell, K.V. Wood, NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells, ACS Chem. Biol. 11 (2) (2016) 400–408, https://doi.org/10.1021/acschembio.5b00753.
[136] E.E. Wrenbeck, M.A. Bedewitz, J.R. Klesmith, S. Noshin, C.S. Barry, T.A. Whitehead, An automated data-driven pipeline for improving heterologous enzyme expression, ACS Synth. Biol. 8 (3) (2019) 474–481, https://doi.org/10.1021/acssynbio.8b00486.
[137] A.O. Magnusson, A. Szekrenyi, H.-J. Joosten, J. Finnigan, S. Charnock, W.-D. Fessner, nanoDSF as screening tool for enzyme libraries and biotechnology development, FEBS J. 286 (1) (2019) 184–204, https://doi.org/10.1111/febs.14696.
[138] J. Wen, H. Lord, N. Knutson, M. Wikström, Nano differential scanning fluorimetry for comparability studies of therapeutic proteins, Anal. Biochem. 593 (2020) 113581, https://doi.org/10.1016/j.ab.2020.113581.
[139] S. Yu, S. Wang, M. Lu, L. Zuo, Review of MEMS differential scanning calorimetry for biomolecular study, Front Mech. Eng. 12 (4) (2017) 526–538, https://doi.org/10.1007/s11465-017-0451-0.
[140] C. Moore-Kelly, J. Welsh, A. Rodger, T.R. Dafforn, O.R.T. Thomas, Automated high-throughput capillary circular dichroism and intrinsic fluorescence spectroscopy for rapid determination of protein structure, Anal. Chem. 91 (21) (2019) 13794–13802, https://doi.org/10.1021/acs.analchem.9b03259.
[141] G.L. Hura, A.L. Menon, M. Hammel, R.P. Rambo, F.L. Poole II, S.E. Tsutakawa, F. E. Jenney Jr, S. Classen, K.A. Frankel, R.C. Hopkins, S. Yang, J.W. Scott, B.D. Dillard, M.W.W. Adams, J.A. Tainer, Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS), Nat. Methods. 6 (2009) 606–612, https://doi.org/10.1038/nmeth.1353.
[142] I. Rodríguez-Ruiz, D. Radajewski, S. Charton, N. Phamvan, M. Brennich, P. Pernot, F. Bonneté, S. Teychené, Innovative high-throughput SAXS methodologies based on photonic lab-on-a-chip sensors: application to macromolecular studies, Sensors 17 (2017) 1266, https://doi.org/10.3390/s17061266.
[143] D. Hess, T. Yang, S. Stavrakis, Droplet-based optofluidic systems for measuring enzyme kinetics, Anal. Bioanal. Chem. 412 (14) (2020) 3265–3283, https://doi.org/10.1007/s00216-019-02294-z.
[144] C.J. Markin, D.A. Mokhtari, F. Sunden, M.J. Appel, E. Akiva, S.A. Longwell, C. Sabatti, D. Herschlag, P.M. Fordyce, Revealing enzyme functional architecture via high-throughput microfluidic enzyme kinetics, Science 373 (2021) eabf8761, https://doi.org/10.1126/science.abf8761.
[145] F. Courtois, L.F. Olguin, G. Whyte, A.B. Theberge, W.T.S. Huck, F. Hollfelder, C. Abell, Controlling the retention of small molecules in emulsion microdroplets for use in cell-based assays, Anal. Chem. 81 (8) (2009) 3008–3016, https://doi.org/10.1021/ac802658n.
[146] T. Buryska, M. Vasina, F. Gielen, P. Vanacek, L. van Vliet, J. Jezek, Z. Pilat, P. Zemanek, J. Damborsky, F. Hollfelder, Z. Prokop, Controlled oil/water partitioning of hydrophobic substrates extending the bioanalytical applications of droplet-based microfluidics, Anal. Chem. 91 (15) (2019) 10008–10015, https://doi.org/10.1021/acs.analchem.9b01839.
[147] K. Markova, A. Kunka, K. Chmelova, M. Havlasek, P. Babkova, S.M. Marques, M. Vasina, J. Planas-Iglesias, R. Chaloupkova, D. Bednar, Z. Prokop, J. Damborsky, M. Marek, Computational enzyme stabilization can affect folding energy landscapes and lead to catalytically enhanced domain-swapped dimers, ACS Catal. 11 (21) (2021) 12864–12885, https://doi.org/10.1021/acscatal.1c03343.
[148] M. Vasina, P. Vanacek, J. Hon, D. Kovar, H. Faldynova, A. Kunka, C. Badenhorst, T. Buryska, S. Mazurenko, D. Bednar, S. Stavros, U. Bornscheuer, A. deMello, J. Damborsky, Z. Prokop, Functional and Mechanistic Characterization of an Enzyme Family Combining Bioinformatics and High-Throughput Microfluidics, 2021. https://www.researchsquare.com/article/rs-1027271/v1 (accessed December 16, 2021).
[149] D. Hess, V. Dockalova, P. Kokkonen, D. Bednar, J. Damborsky, A. deMello, Z. Prokop, S. Stavrakis, Exploring mechanism of enzyme catalysis by on-chip transient kinetics coupled with global data analysis and molecular modeling, Chem. 7 (4) (2021) 1066–1079, https://doi.org/10.1016/j.chempr.2021.02.011.