

Excelzyme: A Swiss University–Industry Collaboration for Accelerated Biocatalyst Development

Sumire Honda Malca^{a§}, Peter Stockinger^{a§}, Nadine Duss^a, Daniela Milbredt^a, Hans Iding^b, and Rebecca Buller^{a*}

Abstract: Excelzyme, an enzyme engineering platform located at the Zurich University of Applied Sciences, is dedicated to accelerating the development of tailored biocatalysts for large-scale industrial applications. Leveraging automation and advanced computational techniques, including machine learning, efficient biocatalysts can be generated in short timeframes. Toward this goal, Excelzyme systematically selects suitable protein scaffolds as the foundation for constructing complex enzyme libraries, thereby enhancing sequence and structural biocatalyst diversity. Here, we describe applied workflows and technologies as well as an industrial case study that exemplifies the successful application of the workflow.

Keywords: Automation · Biocatalysis · Bioinformatics · Enzyme engineering · Machine learning · Pharmaceutical industry



The Excelzyme team. From left to right, back row: Rebecca Buller and Hans Iding; front row: Nadine Duss, Daniela Milbredt, Sumire Honda Malca, and Peter Stockinger.

1. Introduction

Biocatalysis applies enzymes, Nature's catalysts, for synthetic purposes. The technology has been explored scientifically for well over a century yielding key insights into the mechanisms that underlie the catalytic capability of various enzyme families. This

knowledge enabled the application of enzymes in the chemical industry, where product quality, stability and overall process efficiency are indispensable.^[1] Besides their exceptional chemo-, regio- and stereoselectivity, which makes them particularly promising for meeting specific product requirements,^[2,3] biocatalysts can render chemical syntheses and processes more environmentally benign by lowering the number of reaction steps, consumption of energy and raw materials, in addition to reducing byproduct and waste generation.^[4]

Over the last decades, biocatalytic approaches for the production of chiral building blocks has gained considerable prominence in organic synthesis, especially in the pharmaceutical sector.^[5–7] Engineering enzymes for asymmetric transformations and integrating them into chemo-enzymatic cascade reactions to produce the desired small molecules has become a valuable tool in drug development and manufacture. Many chemo-enzymatic reactions for the synthesis of small-molecule active pharmaceutical ingredients (APIs) on large scale have been implemented.^[8–13] Notable examples from the pharmaceutical industry include 1) the ketoreductase-mediated synthesis of a chiral diol intermediate of an antitumoral gamma secretase inhibitor;^[14] 2) the chemo-enzymatic processes involving transaminases to synthesize chiral amine precursors of antidiabetic sitagliptin^[15] or antihypertensive sacubitril;^[16] 3) the application of an imine reductase for reductive amination and consecutive kinetic amine resolution to yield an anticancer lysine-specific demethylase-1 inhibitor;^[17] and 4) the comprehensive synthesis of anti-HIV islatravir through a series of reactions involving five evolved and four auxiliary enzymes.^[18] In more recent years, enzymes like oxidases, transpeptidases and ligases have been utilized to access complex macromolecules such as modified peptides and oligonucleotides, as well as other biopolymers or bioconjugates with therapeutic applications.^[19–21]

However, sometimes enzymatic routes – even if relying on optimized enzymes – fall short of meeting expected key performance

*Correspondence: Prof. R. Buller, E-mail: rebecca.buller@zhaw.ch

^aCompetence Center for Biocatalysis, Institute of Chemistry and Biotechnology, Zurich University of Applied Sciences, Einsiedlerstrasse 31, CH-8820 Wädenswil, Switzerland; ^bProcess Chemistry & Catalysis, F. Hoffmann-La Roche Ltd., Grenzacherstrasse 124, CH-4070 Basel, Switzerland

[§]Both authors contributed equally

indicators (KPIs), such as acceptable conversion at high substrate load, satisfactory space-time yields and straightforward downstream processing, and thus might necessitate further improvements in order to be comparable to well-established chemical routes.^[22] While no fail-proof recipe exists, a key step in developing an effective enzyme for a technical application is the selection of a suitable protein scaffold as a starting point for enzyme engineering. The enzyme selection process must be approached on a case-by-case basis to fit the required process demands – while in some cases the most active enzyme will be chosen for further development, in other cases characteristics such as thermostability or selectivity define a given enzyme's suitability to serve as a starting scaffold.^[23,24] Once a biocatalyst with at least a minor ability to produce the desired product is identified, it then serves as a parent for subsequent enzyme engineering strategies aimed to tailor the catalyst's properties. For this purpose, many strategies and methods have been developed, including directed evolution, semi-rational enzyme engineering including computer-assisted design, and most recently, algorithm-aided enzyme engineering harnessing machine learning approaches (Fig. 1).^[25]

The Nobel-prize winning technology of directed evolution enables the enhancement of enzyme performance by mimicking the process of natural evolution in the laboratory: Following the generation of genetic diversity through random mutagenesis, the encoded enzyme variants are subjected to rigorous screening. Improved variants are selected based on specific criteria such as activity, selectivity or thermo- and solvent stability.^[26,27] Notably, directed evolution does not require an understanding of the enzyme's underlying structure-function relationship. However, in a typical directed evolution huge libraries of enzyme variants need to be screened to find viable solutions, making it a resource-intensive process.^[26,28]

In contrast, semi-rational enzyme design relies on structural information or is guided by conserved amino acid sequences. By using this additional information, more targeted enzyme variant libraries can be generated, in this way reducing the subsequent screening burden.^[25,29] Thanks to the significant expansion of computing power and advances in protein design algorithms,

computer-aided design has emerged as an additional tool to customize protein scaffolds for application. Computer-aided design primarily relies on the exploration of protein structure, substrate binding mode models as well as protein dynamics under different conditions (*e.g.* temperature, pH or solvent).^[25] By applying fundamental principles of chemistry, physics, and statistics, alterations in energy as well as interactions between individual amino acids and ligands are identified during structural adjustments. Along with molecular dynamic simulations, other techniques such as Monte Carlo algorithms have been applied in this context.^[30-32]

Lastly, artificial intelligence (AI) holds great promise to model complex relationships between protein sequence and function by finding sequence patterns which are important for enzymatic function.^[33] Trained by experiment, AI models account for any factors that contribute to specific properties, including those that are unknown.^[34,35]

2. Excelzyme Collaboration

Established in 2019 by the Competence Center for Biocatalysis (CCBIO) at the Zurich University of Applied Sciences (ZHAW) and the Process Chemistry & Catalysis Department at F. Hoffmann-La Roche Ltd. (Roche), Excelzyme aims to accelerate enzyme discovery and engineering campaigns to yield biocatalysts suitable for their implementation in drug substance manufacturing processes.

2.1. Excelzyme Enzyme Discovery Workflow

At the initiation of each project, suitable enzyme starting scaffolds are mined. Using available enzyme wildtype libraries at CCBIO and Roche, which are complemented with literature sourced biocatalysts, enzymes with initial activity for the desired transformation are identified. The sequences of active enzymes then serve as input for mining of public databases, such as NCBI and GenBank,^[36] PDB^[37] or UniProt^[38] to collect relevant evolutionary relatives around the promising starting enzyme candidates. The efficacy of this mining process can be improved through the use of sequence search and clustering tools like DIAMOND,^[39] OrthoFinder2^[40] or MMseqs2,^[41] which provide a better balance between speed and sensitivity compared to more established methods (*e.g.* PSI-BLAST). Combined with careful dataset curation, such as redundancy and length filtering, signal peptide detection, as well as in-depth evaluation of domain composition (*e.g.* transmembrane regions or domains), the mining process typically yields a diverse set of protein sequences. This curated sequence data is not only exploited for the discovery of functional orthologs, but also for the identification of conserved, diversifying amino acid residues and related protein functionalities,^[40,42] information which can be profitably employed in the subsequent engineering campaign. Thanks to advances in homology modeling^[43-45] and AI-enabled *de novo* structure prediction,^[46-50] sequences of interest for which no experimentally determined structure is available, can be modeled revealing the enzyme's tertiary and quaternary structures, oligomerization states, as well as cofactor dependencies. With this information, the capability of the orthologous enzyme to bind the substrate of interest can be preliminarily evaluated. In addition, generalized predictors of industrially relevant properties, such as solubility,^[51,52] thermostability,^[53-56] and pH stability^[57] may be included in the selection process. However, the bioinformatic workflow heavily relies on data-informed or experience-guided assessments of the models to avoid undesirable trade-offs, such as increased protein stability at the cost of significantly lower expression, solubility, or activity (Fig. 2, Table 1).^[58]

During an Excelzyme enzyme discovery phase, a first set of, up to, hundreds of orthologous candidates with a typical shared protein sequence identity of 30–80 % is screened for the desired reaction. Clearly, the actual ortholog library size and protein se-

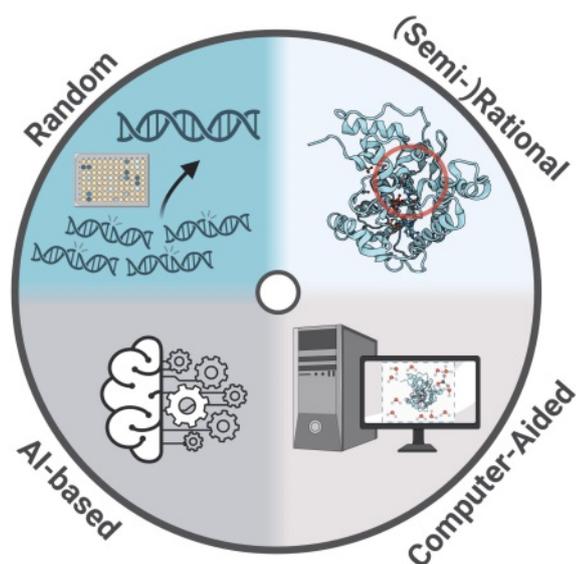


Fig. 1. Strategies for enzyme engineering. Directed evolution mimics natural evolution by random introduction of diversity and subsequent selection of the desired properties. Mutations are introduced more selectively in the context of (semi-)rational and computer-aided design. Trained by experimental data, artificial intelligence (AI) allows the identification of complex relationships between protein sequence and function by finding patterns relevant for enhanced enzyme activity, selectivity, or stability.

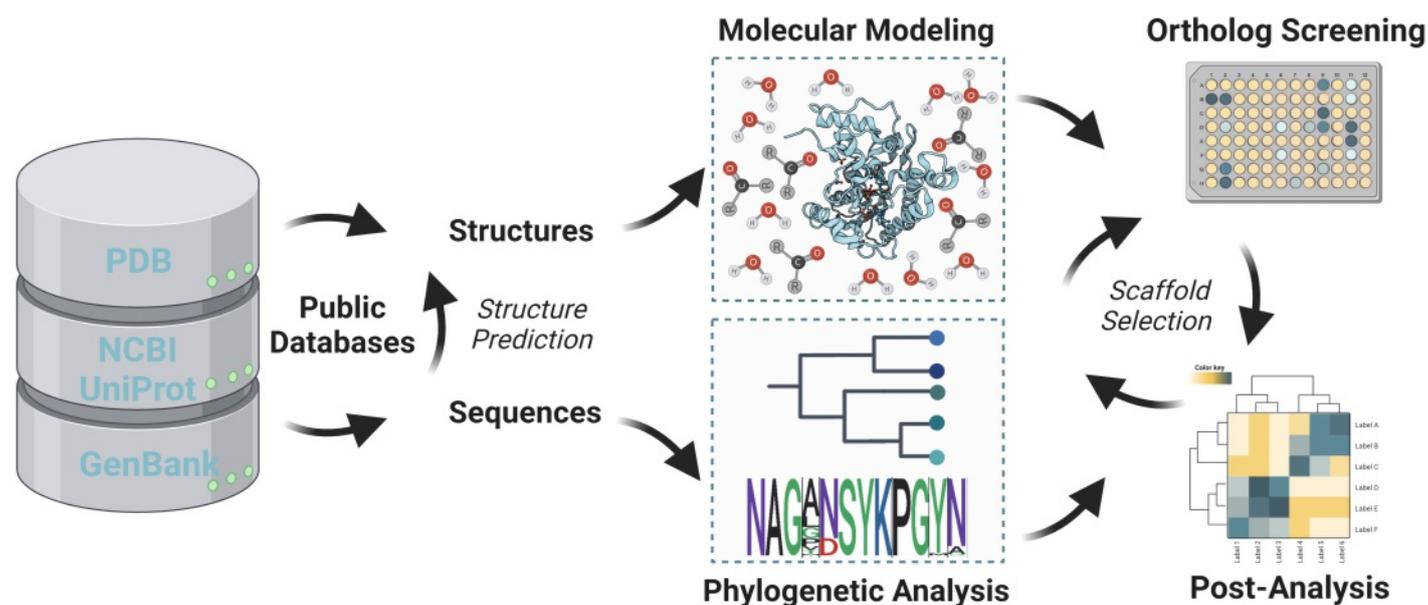


Fig. 2. Excelzyme's enzyme discovery workflow. Based on sequence and structural data from public databases, state-of-the-art methods in phylogenetic analysis, protein feature predictions, and molecular modeling are applied for the strategic design of a curated collection of orthologous enzyme candidates with putative function. Experimental data of the enzyme set towards the desired reaction serves as seed for a second fine-tuned collection. This iterative process typically leads to the identification of a protein scaffold with improved properties, well-suited for subsequent engineering efforts.

quence distribution is critically dependent on the size and diversity of the considered enzyme family. While some enzyme families, such as α -ketoglutarate dependent halogenases might only have a few representatives, distinctly larger enzyme families, such as the short-chain dehydrogenase/reductase (SDR) family,^[59] are a promising source for diverse screening panels sharing low sequence identity yet similar catalytic functionalities.^[60,61] These initial discovery rounds are usually followed by a more refined search, where tens of candidates with higher phylogenetical, structural, or motif-specific homology toward the seed enzyme are selected and screened to identify an even better-suited enzyme starting scaffold. In this way, not only bioinformatic tools and molecular modeling, but also statistical analysis and data science allow us to shed light on the molecular basis of enzymatic performance. Overall, the data-driven enzyme candidate selection usually leads to improved starting points for subsequent enzyme engineering.

2.2. Excelzyme Enzyme Engineering Workflow

Once the starting protein scaffold is defined, a set of methods for random or rational enzyme engineering are employed (Table 1). When information on the sequence-function relationship of the selected enzyme is insufficient, error-prone PCR-based methodologies^[28] or (deep) mutational scanning^[99] are applied. Clearly, the insights gained through these approaches can be leveraged in subsequent engineering rounds by analyzing the randomly generated diversity.^[100] In case of knowledge-driven enzyme engineering for increased activity or selectivity, library design relies on the identification of key amino acid residues involved in substrate binding and catalysis. These hotspots are identified *via* conservation analysis, enzyme modeling, molecular docking, and dynamics. Enhanced thermal or solvent stability is attained *via in silico* variant design based on evolutionary information, thermodynamic stability, and protein flexibility. In this context, it is important to assess and benchmark all applied computational tools to minimize the probability of spurious correlations, allowing us to build a repertoire of validated and complementary methods for future engineering cases with similar problem definitions.

To predict the protein fitness landscape from protein sequence characteristics using data from a fraction of a given library,^[34] we rely on machine learning (ML) algorithms. For example, Bayesian learning techniques such as Gaussian processes have been demonstrated as useful for the design of thermostable P450 variants,^[101] fluorescent proteins with altered fluorescence properties,^[102] more active and regio-divergent halogenase variants,^[103] and toluene-degrading monooxygenase mutants with increased substrate specificity.^[104] The relatively small number of variants screened in these studies confirms the remarkably high hit-rates of ML. More recently, deep learning and generative models, have found application in protein and enzyme design. For instance, the UniRep neural network utilizes deep learning to extract essential protein structure characteristics directly from amino acid sequences, accurately predicting the impact of amino acid replacements on protein function and saving significant resources.^[96] Furthermore, a method called RFDiffusion is a highly effective generative model for protein backbones, exhibiting exceptional performance in various protein design tasks and holding a great potential in the *de novo* design of artificial proteins or enzymes.^[97,98]

2.3. Excelzyme Technology Platform

To meet the often challenging project timelines of chemical process development, we rely on automation for the efficient production of enzyme variants in the desired formulation. Additionally, automated analysis of sequencing and function data speeds up the interpretation of the screening experiments.

An enzyme optimization round involves several key stages (Fig. 3). In the first step, variants or screening libraries are designed according to the selected enzymes engineering strategy. Hotspots can be tackled individually or in a combinatorial fashion, leading to *in silico* libraries of hundreds of thousands to millions of variants, which can be experimentally created by reduced library design with degenerate codons,^[105] design of oligo pools,^[91] or by full or partial gene synthesis with high uniformity of variant representation at reasonable costs.^[106] Subsequently, these gene constructs are directly utilized or further processed to generate

Table 1. Selection of techniques and bioinformatic tools, which – among others – are employed in Excelzyme’s enzyme discovery and engineering workflows.

Techniques or tools	Purpose in pipeline	Description, indicators	Ref.
Enzyme variant library generation (n: number of mutation sites)			
cepPCR	random mutagenesis	fragment-targeted or “casting” error-prone PCR	[62]
mutational scanning	(random) single-site saturation mutagenesis	n = 50 – 150; gene synthesis	[63]
NNK, 22c-trick, customized degenerate oligos	single-site or combinatorial saturation mutagenesis	n ≤ 3; degenerate oligos with reduced or no codon redundancy	[64,65]
oligo pools	combinatorial saturation or focused combinatorial site mutagenesis	n > 3; ≤ 300 bp-oligo synthesis, unlimited pool size	[66]
MEGAWHOP PCR	mutagenesis, gene assembly	whole plasmid amplification via a megaprimer	[67]
In-Fusion, TEDA	gene assembly	efficient ligase-free seamless cloning	[68–70]
Screening			
rapid achiral UHPLC-UV/MS	substrate conversion, product formation, regio- or diastereo-selectivity	% conversion, product distribution, % de (high throughput: ≤ 60 min per 96-well plate)	[71]
chiral LC/GC	stereoselectivity	% ee or % de (variable throughput)	[72,73]
UV/Vis-spectroscopy	initial reaction rate	min ⁻¹ (moderate throughput: 60 – 90 min per 96-well plate)	[74]
nanoDSF	thermostability	T _m (48 samples per run)	[75]
(Next-Generation) Sequencing			
Sanger	hit confirmation	plates or individual variants; outsourced service	[76]
Nanopore consensus NGS	multiplexed sequencing for ML applications	2.3k variants (2 single-use flow cells), ≥ 99 % accuracy at ≥ 50 reads per variant	[77,78]
Computational			
DIAMOND, OrthoFinder2, MMseqs2	ortholog search (discovery)	phylogenetic orthology (“orthogroup”) clustering	[40,41,79]
Protein-Sol	solubility prediction	sequence-based, for ortholog search	[52]
DeepSTABp, ProTstab2	stability prediction	sequence-based, for ortholog search	[53,55]
SWISS-Model	structure prediction	accurate homology modeling when known templates are available	[80]
AlphaFold (DB), OpenFold	structure prediction	accurate AI-based ligand-free structural prediction of novel enzymes	[49,50,81]
3DM	hotspot identification	multiple sequence alignment analysis of a superfamily	[82]
AutoDock Vina	hotspot identification	molecular docking (substrate, intermediate, product and/or cofactor)	[83]
MOE, Moloc	hotspot identification	ligand binding mode modeling	[84,85]
PLIP (PyMOL Plugin)	hotspot identification	protein-ligand interaction profiler	[86]
CAVER 3.0 (PyMOL Plugin)	hotspot identification	ligand access channel or tunnel analysis	[87]
AMBER, OpenMM, MDTraj	hotspot identification, post-analysis	MD simulations for analysis of protein conformational changes, protein-ligand interactions, solvent models	[88–90]
LibGENiE	deleterious mutation identification	sequence space reduction by excluding destabilizing mutations	[91]
Algorithm 2.1 GPML	activity & selectivity prediction	Gaussian processes for ML	[92,93]
EvoEF2, ACDC-NN	(thermo-) stability prediction	structure-based prediction of free energy changes in protein variants	[94,95]
UniRep	activity & stability prediction	sequence-based deep representation learning	[96]
RFdiffusion	de novo protein design	generative model for protein backbones	[97,98]

enzyme variants through appropriate mutagenesis and cloning strategies (Table 1).

To screen enzyme libraries, the Excelzyme platform employs a Tecan Fluent robotic platform allowing automated colony picking, cultivation, protein expression, cell harvesting and cell lysis

procedures. Automation of these steps support the rapid and uniform preparation of whole cells or cell lysates for downstream screening. In each case, screening conditions are continually updated, drawing insights from reaction- and process engineering outcomes from the industry partner. Depending on the product’s

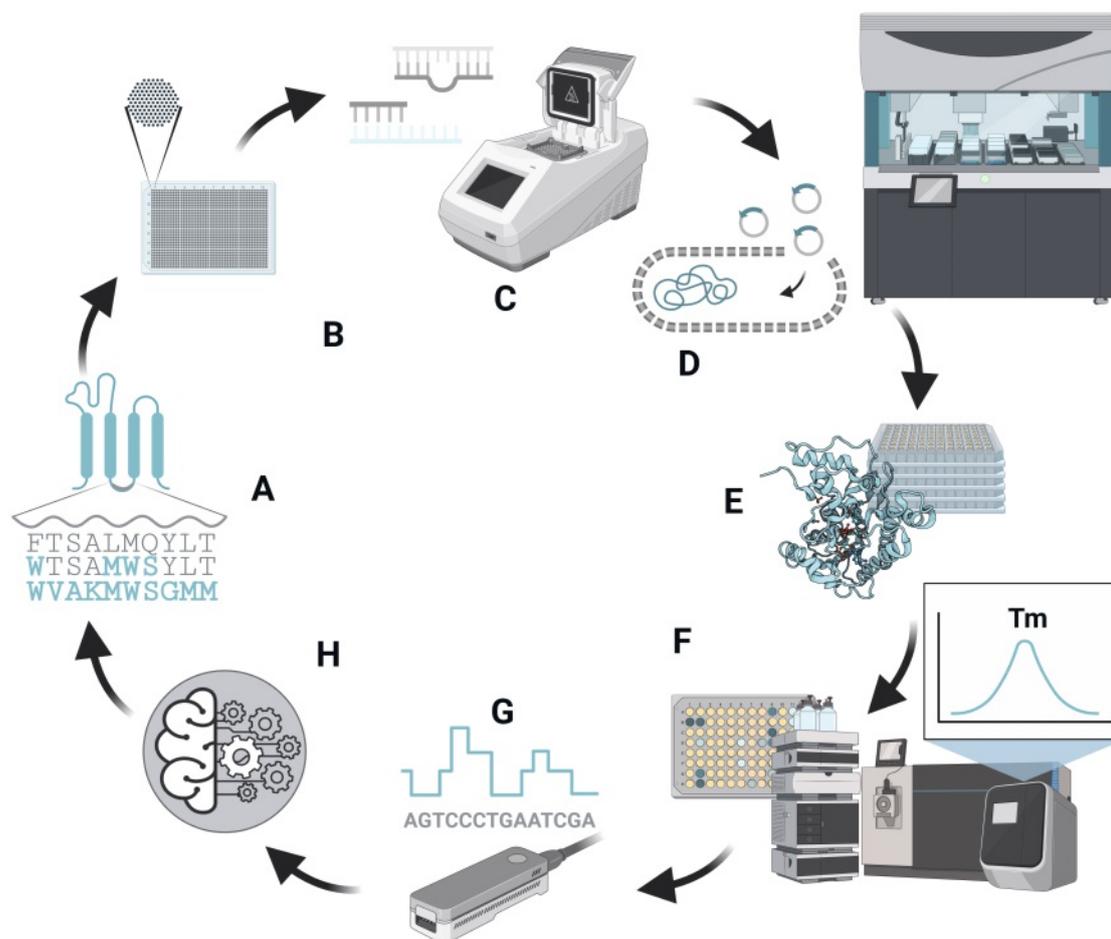


Fig. 3. Iterative enzyme engineering cycle enabled by the Excelzyme technology platform. Depending on the selected evolution strategy, enzyme variant libraries or individual variants are designed, (A). Mutated genes or gene fragments are generated by outsourced solid-phase oligonucleotide synthesis, (B), or by PCR amplification using oligo pools or individual mutagenic primers, (C), followed by suitable ligase-free cloning methods. Fully automated colony picking, cultivation, protein expression, cell harvesting, or cell lysis protocols, (D), are applied to obtain whole cells or enzyme lysates for high-throughput screening, (E). Based on the desired performance indicators and the target product's properties, rapid UHPLC-UV/MS (conversion), chiral LC-UV/GC-FID (selectivity), UV/Vis-spectroscopy (initial reaction rate), or nano differential scanning fluorimetry (thermostability) are usually applied, (F). Variants are sequenced by Sanger (hit confirmation) or in-house by Nanopore-based consensus NGS ($\leq 10\%$ of library size for ML applications), (G). The data generated during each optimization round is then curated and used for statistical analysis, modeling or ML to allow data-driven variant selection or library design for subsequent engineering rounds, (H).

properties and the enzyme engineering targets, various analytical techniques are deployed, such as rapid achiral UHPLC-UV/MS for conversion, chiral LC/GC for selectivity and nano differential scanning fluorimetry (nanoDSF) for protein melting temperature determination. Following these measurements, the function data of the generated enzyme variants is gathered and analyzed in an automated manner using scripts that report fold-improvement over the parent values, enable quality checks, and generate plate heatmaps for effective data visualization.

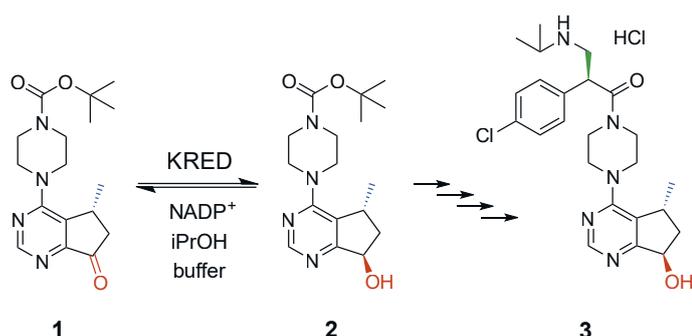
Verification of hit variant sequences is accomplished by outsourced Sanger sequencing. Sequencing data is analyzed using a custom script that evaluates sequencing read length and quality, enables manual checks of the DNA chromatograms, and generates output files of the variants and their corresponding amino acid substitutions. To obtain the sequences of hundreds or thousands of variants required to train ML algorithms, we additionally implemented multiplexed nanopore-based next-generation sequencing (NGS).^[107] Recently, the high error rate associated with nanopore sequencing has been overcome by introducing unique barcodes into individual sequences *via* PCR, followed by the processing of sequencing data through customized statistical analysis for consensus basecalling.^[77,78] Our in-house technology currently allows us to elucidate the sequences of 1'152 variants (12×96 -well plates) per single-use flow cell, requiring

a minimum of 50 reads per amplicon to achieve $\geq 99\%$ accuracy at the single nucleotide level in 72 h, including lab work and sequencing. Furthermore, nanopore sequencing can be applied to sequences beyond 1 kb reads, a limitation inherent to Sanger sequencing. This translates into more substantial cost savings when larger genes (> 1 kb) are sequenced. The generated sequence-function data is then curated and subjected to statistical analysis, modeling, or ML. These analytical tools help in data-driven variant selection and library design for subsequent engineering rounds, allowing for continuous enhancement of enzyme performance.^[34,100]

3. Case study: Data-driven KRED Evolution

3.1 KREDs in the Pharmaceutical Industry

Chiral secondary alcohols are key API intermediates that are accessible *via* different synthetic approaches, including asymmetric ketone hydrogenation,^[108-112] enzyme- or metal-catalyzed (dynamic) kinetic resolution of alcohol racemates,^[113,114] and selective ketone reduction by ketoreductases (KREDs).^[115-117] Nowadays, engineered KREDs are ubiquitous in the pharmaceutical industry and applied in reactions with NAD(P)H as hydride donor and generally isopropanol for cost-effective cofactor regeneration.^[115,116] These biocatalysts have enabled access to a wide



Scheme 1. Stereoselective KRED-catalyzed reduction of **1** to **2** in the chemo-enzymatic synthesis of ipatasertib (**3**). The API's three stereocenters are color-depicted.

array of chiral synthons leading to APIs like atorvastatin,^[118,119] montelukast,^[120] simeprevir,^[121,122] and ipatasertib.^[123,124]

3.2. Engineering of a KRED for the Synthesis of an Ipatasertib Precursor

Ipatasertib (**3**) is a potent protein kinase B inhibitor developed for the treatment of metastatic castration-resistant prostate cancer and triple-negative metastatic breast cancer.^[125–127] Its synthesis spans ten steps and involves eight isolated intermediates, leveraging chemical and enzyme catalysis for the incorporation of three stereocenters.^[123,124] One of these chiral centers is es-

tablished through the asymmetric KRED-catalyzed reduction of prochiral ketone **1** to the desired (*R,R*)-*trans* alcohol intermediate **2** (Scheme 1). We focused on this reaction step to showcase the efficacy of automation- and algorithm-aided enzyme engineering for the development of a high-performance biocatalyst for a potential second-generation process.^[74]

To identify a robust enzyme starting scaffold for engineering, we screened our in-house KRED toolbox consisting of 63 wild-type enzymes of bacterial, fungal and plant origin for the conversion of **1**.^[74] Following the screening, the NADP⁺-dependent aldehyde reductase II from *Sporidiobolus salmonicolor*,^[128,129] herein denoted as *Ssal*-KRED, was selected for further engineering owing to its absolute stereopreference for **2** (Fig. 4A). Former studies on this enzyme provided structural and functional data for different alkyl or aryl ketones.^[130–133]

At the outset of the study, we defined the campaign goal as improving *Ssal*-KRED's performance by a factor of 50 in the presence of isopropanol for NADPH regeneration. Enzyme variant performance was measured as fold-improvement over the wildtype enzyme (FIOWT), calculated from a kinetic UV assay based on consumption of **1**. Informed by mutational scanning data on every second amino acid within the protein in addition to a substrate-bound enzyme model and literature, six sites in the substrate entrance tunnel or substrate binding cavity (F97W, L174, A238, L241, M242, Q245; Fig. 4C) were chosen to be explored in single-site saturation mutagenesis (SSM) libraries. The screening of these libraries resulted in the identification of several beneficial

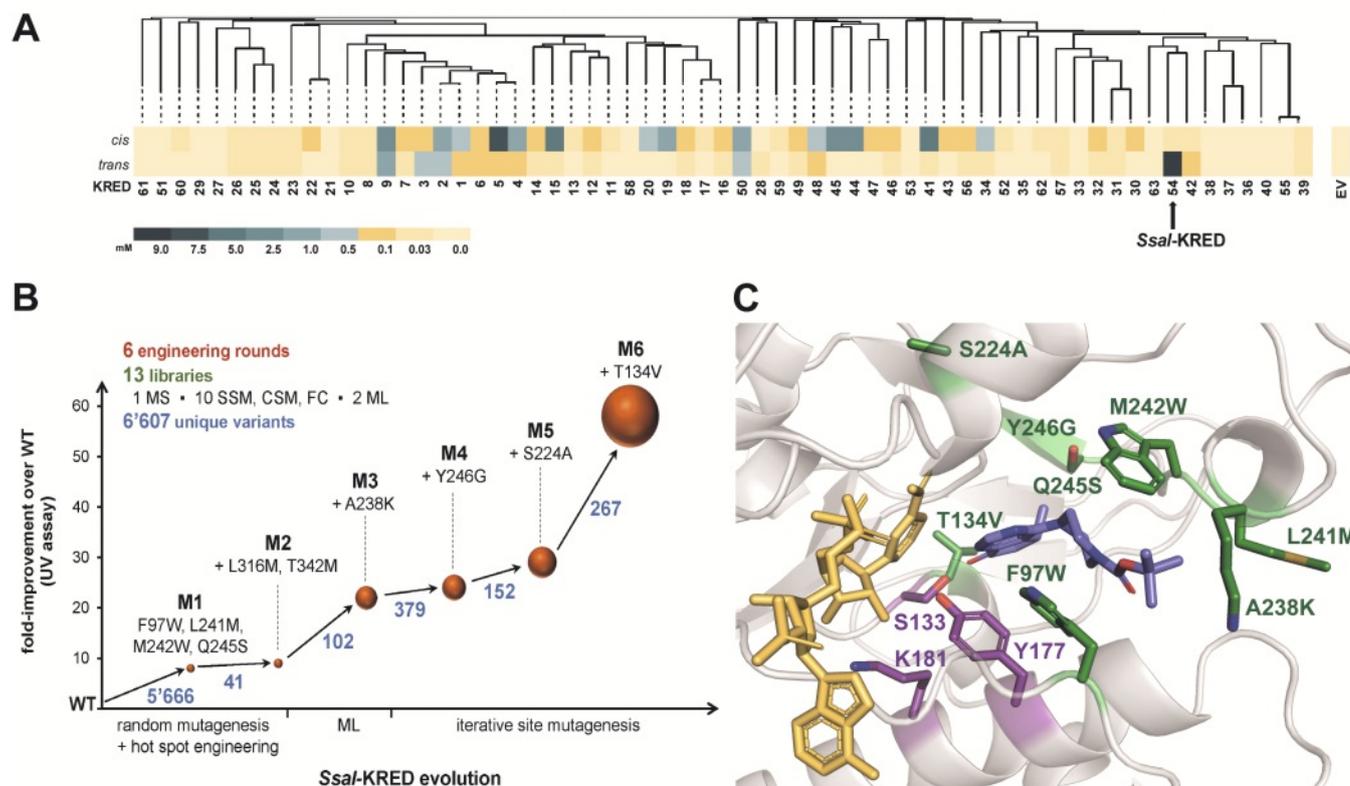


Fig. 4. *Ssal*-KRED discovery and engineering campaign for the efficient synthesis of **2**. (A) Discovery phase: screening of an in-house collection of 63 wildtype KRED enzymes against **1** in the presence of glucose/glucose dehydrogenase led to the identification of *Ssal*-KRED (92 % conversion, 99 % *de (R,R-trans)* at 0.3 % (w/v) substrate load); EV, empty vector. (B) Engineering phase: Overview of the fold-improvement of *Ssal*-KRED hit variants **M1** – **M6** over the wild-type enzyme (FIOWT, indicated in bold brown font) in terms of initial reaction rates determined *via* a UV assay that followed consumption of **1**. Introduced mutations are highlighted below each variant name. In total, 171 amino acid positions were investigated in the initial mutational scanning (MS) library while 23 amino acid positions were explored in the frame of 10 single-site saturation mutagenesis (SSM), combinatorial saturation mutagenesis (CSM) or focused combinatorial (FC), and 2 machine learning (ML) libraries. Overall, we investigated 6'600 unique on-target variants which required the screening of 15'000 transformants. (C) Model of **M6** binding **1** (blue) and NADPH (yellow) with highlighted catalytic residues (purple) and substrate environment residue mutations (green). Beneficial mutations on the protein surface (L316M and T342M) are not shown. The image was generated using PyMOL 2.5.5.

mutations, including F97W, M242F, and Q245T, with a FIOWT between 1.4 and 3.6.

Building on the obtained data, we set up four combinatorial 3-site and one combinatorial 5-site saturation mutagenesis (CSM) library by varying positions L174, A238, L241, M242, Q245 while fixing mutation F97W (all other substitutions at position 97 had exhibited a detrimental effect in the prior SSM library). With a ML application in mind, we then opted to screen a fraction of each library. Notably, ML can be applied at any stage of an enzyme engineering campaign provided sufficient and high-quality data is available for training the algorithm. To obtain the necessary sequence-function data, we thus covered ~ 7 % and 0.024 % of the theoretical size of the 3- and 5-site CSM libraries, respectively. In this screen, we identified variant **M1** (F97W/L241M/M242W/Q245S) as the best performing variant, which exhibited a FIOWT of 8. By combining **M1** with beneficial surface residue mutations L316M and T342M, identified in the initial mutational scanning library, variant **M2** was created. This variant exhibited a FIOWT of 9.

As a next step, the sequence-function data of the CSM libraries was used as input to train a Gaussian process-based ML algorithm.^[101,134] Toward this goal, we used, among others, the data derived from the multi-site CSM libraries to train the algorithm (2'600 datapoints). Instead of constructing specific variants individually, we opted to build a small variant library (ML filtered library) which contained the top amino acids the ML algorithm had predicted to be beneficial. This library consisted of 75 variants and gratifyingly its screening revealed variant **M3**, which contained additional mutation A238K and displayed a FIOWT of 22. While variant **M3** would have also been identifiable through combining beneficial mutations from the SSM libraries,^[135,136] it should be noted that the ML approach allowed us to obtain such results by screening only 75 variants instead of 29'400 variants. The latter number corresponds to the size of the hypothetical library if all beneficial mutations identified for the six hotspots were to be explored in a combinatorial manner with the respective wild-type amino acids. In practical terms, screening such a library size would have required the preparation and analysis of one thousand 96-well plates, considering the necessary 3-fold oversampling.

Using the ML-improved protein scaffold as parent, we further increased enzyme activity and stability by targeting experimentally determined and literature-based^[130–132,137,138] amino acid positions in the frame of iterative site mutagenesis as a complementary strategy.^[139] Upon tackling fourteen new positions located in the substrate- or NADPH-binding sites, we identified beneficial substitutions Y246G, S224A and T134V in three successive rounds and obtained hit variants **M4**, **M5** and **M6** with FIOWT values of 24, 29 and 58, respectively (Fig. 4B).

After completing six evolution rounds on *Ssal*-KRED, targeting 180 amino acid positions across 13 libraries, we successfully achieved our project's goal with the 10-amino acid variant **M6** (F97W/ T134V/S224A/A238K/L241M/M242W/Q245S/Y246G/L316M/T342M) (Fig. 4C). Notably, **M6** exhibited a 64-fold higher apparent k_{cat} and improved robustness under process conditions compared to the wildtype enzyme. Kinetic studies and modeling suggested that key mutations for the efficient reduction of **1** to **2** were substrate access tunnel or substrate binding mutations T134V, A238K, M242W and Q245S, while changes on the enzyme surface were likely responsible for enzyme stability. While reactions with the wild-type enzyme resulted in 26 % conversion and > 99.5 % *de* (*R,R-trans*) at 100 g/L of **1** after 24 h, preparative scale reactions with **M6** resulted in ≥ 98 % conversion and 99.7 % *de* (*R,R-trans*) at the same substrate loading after 30 h, showcasing the technical and commercial viability of the process based on the engineered enzyme variant.

4. Conclusions and Outlook

The Excelzyme enzyme discovery and engineering platform allows the fast development of biocatalysts, in line with the often challenging process development timelines. Typically, an Excelzyme project, including enzyme discovery and engineering, runs for 10 to 11 months, with one evolution round being limited to 4 to 5 weeks of library design, experimental work, and data analysis. In this fashion, Excelzyme has delivered additional engineered enzymes stemming from several different enzyme classes to be employed for drug synthesis. Importantly, as a collaborative venture between academia and industry, the platform equally profits from enzyme engineering and process development know-how, which need to go hand-in-hand to yield viable biocatalytic manufacturing routes.

Complementing semi-rational enzyme engineering principles, Excelzyme evaluates and employs bioinformatic tools. As more robust and powerful algorithms are being developed, we expect that computer- and AI-assisted enzyme engineering will be able to address a broader variety of enzyme design challenges, thus enabling more time- and resource- efficient development of versatile biocatalysts suited for various industrial applications.

Acknowledgements

We gratefully acknowledge former Excelzyme team members Michael Niklaus (bioinformatics) and Jasmin Meierhofer (molecular biology & analytics) for their valuable contributions to the development of the technology platform and their work on the KRED case study; David Patsch for his support in machine learning and the implementation of automated oligo design and sequencing analysis tools; Robin Küng, Takahiro Hayashi, Moritz Voss and Stefanie Reiter for their initial work on KRED discovery or characterization; Fabian Meyer for developing additional automated computational tools. We are grateful to Stefan Hildbrand, Kurt Püntener, Urs Hilber and Christian Hinderling for their continuous support; and to Bernd Kuhn, Steven Paul Hanlon, Dennis Wetzl, Patrik Meier, Veronika Kirrmann and Meinrad Birrer for their contributions to the case study at F. Hoffmann-La Roche Ltd. Figures 1–3 of this manuscript were created with BioRender.com (2023). The KRED study was performed as part of NCCR Catalysis, a National Center of Competence in Research funded by the Swiss National Science Foundation (Grant number 180544).

Received: December 21, 2023

- [1] A. Gutteridge, J. M. Thornton, *Elsevier Current Trends*, **2005**, 30, 11, 622, <https://doi.org/10.1016/j.tibs.2005.09.006>.
- [2] 'Hydrolases in Organic Synthesis', U. T. Bornscheuer, R. J. Kazlauskas, Wiley, **2005**, <https://doi.org/10.1002/3527607544>.
- [3] K. Faber, W. Kroutil, *Curr. Op. Chem Bio.* **2005**, 2, 181, <https://doi.org/10.1016/j.cbpa.2005.01.001>.
- [4] K. R. Jegannathan, P. H. Nielsen, *J. Cleaner Prod.* **2013**, 42, 228, <https://doi.org/10.1016/j.jclepro.2012.11.005>.
- [5] E. Siirola, F. Eggmann, C. Moore, K. Schroer, A. Vargas, T. Peschke, T. Schlama, R. Snajdrova, *CHIMIA* **2023**, 6, 376 <https://doi.org/10.2533/chimia.2023.376>.
- [6] U. T. Bornscheuer, G. W. Huisman, R. J. Kazlauskas, S. Lutz, J. C. Moore, K. Robins, *Nature* **2012**, 485, <https://doi.org/10.1038/nature11117>.
- [7] R. Buller, S. Lutz, R. J. Kazlauskas, R. Snajdrova, J. C. Moore, U. T. Bornscheuer, *Science* **2023**, 382, eadh8615, <https://doi.org/10.1126/SCIENCE.ADH8615>.
- [8] F. Rudroff, M. D. Mihovilovic, H. Gröger, R. Snajdrova, H. Iding, U. T. Bornscheuer, *Nat. Cat.* **2018**, 1, 12 <https://doi.org/10.1038/s41929-017-0010-4>.
- [9] D. Ghislieri, N. J. Turner, *Top Catal.* **2014**, 57, <https://doi.org/10.1007/s11244-013-0184-1>.
- [10] W. Zawodny, S. L. Montgomery, *Catal.* **2022**, 12, 595, <https://doi.org/10.3390/catal12060595>.
- [11] J. Nazor, J. Liu, G. Huisman, *Curr. Opin. Biotechnol.* **2021**, 69, 182, <https://doi.org/10.1016/j.copbio.2020.12.013>.
- [12] S. P. France, R. D. Lewis, C. A. Martinez, *JACS Au* **2023**, 3, 715, <https://doi.org/10.1021/jacsau.2c00712>.
- [13] S. Wu, R. Snajdrova, J. C. Moore, K. Baldenius, U. T. Bornscheuer, *Angew. Chem. Int. Ed.* **2021**, 60, 88, <https://doi.org/10.1002/anie.202006648>.

- [14] X. Linghu, N. Wong, H. Iding, V. Jost, H. Zhang, S. G. Koenig, C. G. Sowell, F. Gosselin, *Org. Process Res. Dev.* **2017**, *21*, 387, <https://doi.org/10.1021/acs.oprd.7b00006>.
- [15] C. K. Savile, J. M. Janey, E. C. Mundorff, J. C. Moore, S. Tam, W. R. Jarvis, J. C. Colbeck, A. Kriebber, F. J. Fleitz, J. Brands, P. N. Devine, G. W. Huisman, G. J. Hughes, *Science* **2010**, *329*, 305, <https://doi.org/10.1126/science.1188934>.
- [16] S. J. Novick, N. Dellas, R. Garcia, C. Ching, A. Bautista, D. Homan, O. Alvizo, D. Entwistle, F. Kleinbeck, T. Schlama, T. Ruch, *ACS Catal.* **2021**, *11*, 3762, <https://doi.org/10.1021/acscatal.0c05450>.
- [17] M. Schober, C. MacDermaid, A. A. Ollis, S. Chang, D. Khan, J. Hosford, J. Latham, A. F. Ihnken, M. J. B. Brown, D. Fuerst, M. J. Sanganee, G. D. Roiban, *Nat. Catal.* **2019**, *2*, 909, <https://doi.org/10.1038/s41929-019-0341-4>.
- [18] M. A. Huffman, A. Fryszkowska, O. Alvizo, M. Borra-Garske, K. R. Campos, K. A. Canada, P. N. Devine, D. Duan, J. H. Forstater, S. T. Grosser, H. M. Halsey, G. J. Hughes, J. Jo, L. A. Joyce, J. N. Kolev, J. Liang, K. M. Maloney, B. F. Mann, N. M. Marshall, M. McLaughlin, J. C. Moore, G. S. Murphy, C. C. Nawrat, J. Nazor, S. Novick, N. R. Patel, A. Rodriguez-Granillo, S. A. Robaire, E. C. Sherer, M. D. Truppo, A. M. Whittaker, D. Verma, L. Xiao, Y. Xu, H. Yang, *Science* **2019**, *366*, <https://doi.org/10.1126/science.aay8484>.
- [19] A. Debon, E. Siirola, R. Snajdrova, *JACS Au* **2023**, *3*, 1267, <https://doi.org/10.1021/jacsau.2c00617>.
- [20] K. J. D. Van Giesen, M. J. Thompson, Q. Meng, S. L. Lovelock, *JACS Au* **2023**, *3*, 13, <https://doi.org/10.1021/jacsau.2c00481>.
- [21] M. J. Lobba, C. Fellmann, A. M. Marmelstein, J. C. Maza, E. N. Kissman, S. A. Robinson, B. T. Staahl, C. Urnes, R. J. Lew, C. S. Mogilevsky, J. A. Doudna, M. B. Francis, *ACS Cent. Sci.* **2020**, *6*, 1564, <https://doi.org/10.1021/acscentsci.0c00940>.
- [22] D. R. Yazbeck, C. A. Martinez, S. Hu, J. Tao, *Tetrahedron: Asymmetry* **2004**, *18*, 2757, <https://doi.org/10.1016/j.tetasy.2004.07.050>.
- [23] S. Hanreich, E. Bonandi, I. Drienovská, *ChemBioChem* **2023**, *6*, <https://doi.org/10.1002/cbic.202200566>.
- [24] H. A. Bunzel, X. Garrabou, M. Pott, D. Hilvert, *Curr. Op. Struct. Bio.* **2018**, *48*, 149, <https://doi.org/10.1016/j.sbi.2017.12.010>.
- [25] W. Xiong, B. Liu, Y. Shen, K. Jing, T. R. Savage, *Biochem. Eng. J.* **2021**, *174* <https://doi.org/10.1016/j.bej.2021.108096>.
- [26] H. A. Bunzel, J. L. R. Anderson, A. J. Mulholland, *Curr. Op. Struct. Bio.* **2021**, *67*, 212, <https://doi.org/10.1016/j.sbi.2020.12.015>.
- [27] F. H. Arnold, *Angew. Chem. Int. Ed.* **2018**, *57*, 4143, <https://doi.org/10.1002/anie.201708408>.
- [28] J. N. Copp, P. Hanson-Manful, D. F. Ackerley, W. M. Patrick, *Methods mol. Biol.* **2014**, *1179*, https://doi.org/10.1007/978-1-4939-1053-3_1.
- [29] R. A. Sheldon, P. C. Pereira, *Chem. Soc. Rev.* **2017**, *46*, 2678, <https://doi.org/10.1039/c6cs00854b>.
- [30] A. Romero-Rivera, M. Garcia-Borrás, S. Osuna, *ChemComm.* **2017**, *53*, <https://doi.org/10.1039/C6CC06055B>.
- [31] J. Planas-Iglesias, S. M. Marques, G. P. Pinto, M. Musil, J. Stourac, J. Damborsky, D. Bednar, *Biotech. Adv.* **2021**, *47*, <https://doi.org/10.1016/j.biotechadv.2021.107696>.
- [32] M. C. Childers, V. Daggett, *Mol. Syst. Des. Eng.* **2017**, *2*, 9, <https://doi.org/10.1039/C6ME00083E>.
- [33] S. Osuna, G. Jiménez-Osés, E. L. Noey, K. N. Houk, *Acc. Chem. Res.* **2015**, *48*, 1080, <https://doi.org/10.1021/ar500452q>.
- [34] D. Patsch, R. Buller, *CHIMIA*, **2023**, *3*, 106 <https://doi.org/10.2533/chimia.2023.116>.
- [35] P. Kouba, P. Kohout, F. Haddadi, A. Bushuiev, R. Samusevich, J. Sedlar, J. Damborsky, T. Pluskal, J. Sivic, S. Mazurenko, *ACS Catal.* **2023**, *13*, 863, <https://doi.org/10.1021/acscatal.3c02743>.
- [36] 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, *Nucleic Acids Res.* **2022**, *50*, D1, <https://doi.org/10.1093/nar/gkab112>.
- [37] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, P. E. Bourne, *Nucleic Acid Research* **2000**, *28*, 235, <https://doi.org/10.1093/nar/28.1.235>.
- [38] A. Bateman, M. J. Martin, S. Orchard, M. Magrane, S. Ahmad, E. Alpi, E. H. Bowler-Barnett, R. Britto, H. Bye-A-Jee, A. Cukura, P. Denny, T. Dogan, T. G. Ebenezzer, J. Fan, P. Garmiri, L. J. da Costa Gonzales, E. Hatton-Ellis, A. Hussein, A. Ignatchenko, G. Insana, R. Ishtiaq, V. Joshi, D. Jyothi, S. Kandasamy, A. Lock, A. Luciani, M. Lugaric, J. Luo, Y. Lussi, A. MacDougall, F. Madeira, M. Mahmoudy, A. Mishra, K. Moulang, A. Nightingale, S. Pundir, G. Qi, S. Raj, P. Raposo, D. L. Rice, R. Saidi, R. Santos, E. Speretta, J. Stephenson, P. Tototo, E. Turner, N. Tyagi, P. Vasudev, K. Warner, X. Watkins, R. Zaru, H. Zellner, A. J. Bridge, L. Aimo, G. Argoud-Puy, A. H. Auchincloss, K. B. Axelsen, P. Bansal, D. Baratin, T. M. Batista Neto, M. C. Blatter, J. T. Bolleman, E. Boutet, L. Breuza, B. C. Gil, C. Casals-Casas, K. C. Echioukh, E. Coudert, B. Cuche, E. de Castro, A. Estreicher, M. L. Famiglietti, M. Feuermann, E. Gasteiger, P. Gaudet, S. Gehant, V. Gerritsen, A. Gos, N. Gruaz, C. Hulo, N. Hyka-Nouspikel, F. Junco, A. Kerhornou, P. Le Mercier, D. Lieberherr, P. Masson, A. Morgat, V. Muthukrishnan, S. Paesano, I. Pedruzzi, S. Pilbout, L. Pourcel, S. Poux, M. Pozzato, M. Pruess, N. Redaschi, C. Rivoire, C. J. A. Sigrist, K. Sonesson, S. Sundaram, C. H. Wu, C. N. Arighi, L. Arminski, C. Chen, Y. Chen, H. Huang, K. Laiho, P. McGarvey, D. A. Natale, K. Ross, C. R. Vinayaka, Q. Wang, Y. Wang, J. Zhang, *Nucleic Acids Res.* **2023**, *51*, D1, <https://doi.org/10.1093/nar/gkac1052>.
- [39] B. Buchfink, H. Ashkenazy, K. Reuter, J. A. Kennedy, H.-G. Drost, *bioRxiv* **2023**.
- [40] D. M. Emms, S. Kelly, *Genome Biol.* **2019**, *20*, 238, <https://doi.org/10.1186/s13059-019-1832-y>.
- [41] M. Steinegger, J. Söding, *Nat. Biotech.* **2017**, *35*, 1068, <https://doi.org/10.1038/nbt.3988>.
- [42] J. Hon, S. Borko, J. Stourac, Z. Prokop, J. Zendulka, D. Bednar, T. Martinek, J. Damborsky, *Nucleic Acids Res.* **2021**, *48*, <https://doi.org/10.1093/NAR/GKAA372>.
- [43] M. Biasini, S. Bienert, A. Waterhouse, K. Arnold, G. Studer, T. Schmidt, F. Kiefer, T. G. Cassarino, M. Bertoni, L. Bordoli, T. Schwede, *Nucleic Acids Res.* **2014**, *42*, <https://doi.org/10.1093/nar/gku340>.
- [44] B. Webb, A. Sali, *Curr. Protoc. Bioinformatics* **2016**, <https://doi.org/10.1002/cpb1.3>.
- [45] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. De Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, *Nucleic Acids Res.* **2018**, *46*, <https://doi.org/10.1093/nar/gky427>.
- [46] X. Zhou, W. Zheng, Y. Li, R. Pearce, C. Zhang, E. W. Bell, G. Zhang, Y. Zhang, *Nat. Protoc.* **2022**, *17*, <https://doi.org/10.1038/s41596-022-00728-0>.
- [47] Z. Lin, H. Akin, R. Rao, B. Hie, Z. Zhu, W. Lu, A. Dos, S. Costa, M. Fazel-Zarandi, T. Sercu, S. Candido, A. Rives, M. Ai, *bioRxiv* **2022**.
- [48] R. Wua, F. Dinga, R. Wanga, R. Shena, X. Zhanga, S. Luoa, C. Sua, Z. Wua, Q. Xieba, B. Bergerc, J. Maa, J. Penga, *bioRxiv* **2022**.
- [49] G. Ahdritz, N. Bouatta, S. Kadyan, Q. Xia, W. Gerecke, T. J. O. D. Berenberg, I. Fisk, N. Zanichelli, B. Zhang, A. Nowaczynski, B. Wang, M. M. Stepniwska-Dziubinska, S. Zhang, A. Ojewole, M. Efe Guney, S. Biderman, A. M. Watkins, S. Ra, P. Ribalta Lorenzo, L. Nivon, B. Weitzner, Y.-E. Andrew Ban, P. K. Sorger, E. Mostaque, Z. Zhang, R. Bonneau, M. AlQuraishi, B. Allen Hamilton, C. Bio, *bioRxiv* **2022**.
- [50] J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Židek, 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, *Nature* **2021**, *596*, 583, <https://doi.org/10.1038/s41586-021-03819-2>.
- [51] J. Hon, M. Marusiak, T. Martinek, A. Kunka, J. Zendulka, D. Bednar, J. Damborsky, *Bioinformatics* **2021**, *37*, 1, <https://doi.org/10.1093/bioinformatics/btaa1102>.
- [52] M. Hebditch, M. A. Carballo-Amador, S. Charonis, R. Curtis, J. Warwicker, *Bioinformatics* **2017**, *33*, 19, <https://doi.org/10.1093/bioinformatics/btx345>.
- [53] F. Jung, K. Frey, D. Zimmer, T. Mühlhaus, *Int. J. Mol. Sci.* **2023**, *24*, 8, <https://doi.org/10.3390/ijms24087444>.
- [54] J. Zhao, W. Yan, Y. Yang, *Int. J. Mol. Sci.* **2023**, *24*, 3, <https://doi.org/10.3390/ijms24032217>.
- [55] Y. Yang, J. Zhao, L. Zeng, M. Vihinen, *Int. J. Mol. Sci.* **2022**, *23*, 18, <https://doi.org/10.3390/ijms231810798>.
- [56] Y. Yang, X. Ding, G. Zhu, A. Niroula, Q. Lv, M. Vihinen, *BMC Genomics* **2019**, *20*, <https://doi.org/10.1186/s12864-019-6138-7>.
- [57] H. Lin, W. Chen, H. Ding, *PLoS One* **2013**, *8*, 10, <https://doi.org/10.1371/journal.pone.0075726>.
- [58] D. Patsch, M. Eichenberger, M. Voss, U. T. Bornscheuer, R. M. Buller, *Comput. Struct. Biotechnol. J.* **2023**, *21*, <https://doi.org/10.1016/j.csbj.2023.09.013>.
- [59] M. Gräff, P. C. F. Buchholz, P. Stockinger, B. Bommarius, A. S. Bommarius, J. Pleiss, *Proteins: Structure, Function, and Bioinformatics* **2019**, *87*, 443, <https://doi.org/10.1002/prot.25666>.
- [60] M. Voss, R. Küng, T. Hayashi, M. Jonczyk, M. Niklaus, H. Iding, D. Wetzl, R. Buller, *ChemCatChem* **2021**, *13*, 6, <https://doi.org/10.1002/cctc.202001871>.
- [61] S. Roth, P. Stockinger, J. Steff, S. Steimle, V. Sautner, K. Tittmann, J. Pleiss, M. Müller, *ChemBioChem* **2020**, *21*, 2615, <https://doi.org/10.1002/cbic.202000233>.
- [62] J. Yang, A. J. Ruff, M. Arlt, U. Schwaneberg, *Biotechnol. Bioeng.* **2017**, *114*, 1921, <https://doi.org/10.1002/bit.26327>.
- [63] D. M. Fowler, S. Fields, *Nat. Methods* **2014**, *11*, 801, <https://doi.org/10.1038/nmeth.3027>.
- [64] M. T. Reetz, D. Kahakeaw, R. Lohmer, *ChemBioChem* **2008**, *9*, 1797, <https://doi.org/10.1002/cbic.200800298>.
- [65] S. Kille, C. G. Acevedo-Rocha, L. P. Parra, Z. G. Zhang, D. J. Opperman, M. T. Reetz, J. P. Acevedo, *ACS Synth. Biol.* **2013**, *2*, 83, <https://doi.org/10.1021/sb300037w>.

- [66] B. P. Kuiper, R. C. Prins, S. Billerbeck, *ChemBioChem* **2022**, *23*, <https://doi.org/10.1002/cbic.202100507>.
- [67] T. M. Miyazaki K., *Biotechniques* **2002**, *33* (5) 1033, <https://doi.org/10.2144/02335st03>
- [68] N. S. Berrow, D. Alderton, S. Sainsbury, J. Nettleship, R. Assenberg, N. Rahman, D. I. Stuart, R. J. Owens, *Nucleic Acids Res.* **2007**, *35*, <https://doi.org/10.1093/nar/gkm047>.
- [69] B. Zhu, G. Cai, E. O. Hall, G. J. Freeman, *Biotechniques* **2007**, *43*, 354, <https://doi.org/10.2144/000112536>.
- [70] Y. Xia, K. Li, J. Li, T. Wang, L. Gu, L. Xun, *Nucleic Acids Res.* **2019**, *47*, 1, <https://doi.org/10.1093/nar/gky1169>.
- [71] M. W. Dong, K. Zhang, *TrAC - Trends in Analytical Chemistry* **2014**, *63*, 21, <https://doi.org/10.1016/j.trac.2014.06.019>.
- [72] D. C. Patel, M. F. Wahab, D. W. Armstrong, Z. S. Breitbach, *J. Chromatogr. A* **2016**, *1467*, 2, <https://doi.org/10.1016/j.chroma.2016.07.040>.
- [73] Y. Teng, C. Gu, Z. Chen, H. Jiang, Y. Xiong, D. Liu, D. Xiao, *Chirality* **2022**, *34*, 1094, <https://doi.org/10.1002/chir.23453>.
- [74] S. Honda Malca, N. Duss, J. Meierhofer, D. Patsch, M. Niklaus, S. Reiter, S. P. Hanlon, B. Kuhn, H. Iding, R. Buller, *Commun. Chem.* **2024**, <https://doi.org/10.1038/s42004-024-01130-5>.
- [75] A. O. Magnusson, A. Szekrenyi, H. J. Joosten, J. Finnigan, S. Charnock, W. D. Fessner, *FEBS Journal* **2019**, *286*, 184, <https://doi.org/10.1111/febs.14696>.
- [76] F. Sanger, S. Nicklen, A. R. Coulson, *Proc. Natl. Acad. Sci. U S A* **1977**, *74*, 5463, <https://doi.org/10.1073/pnas.74.12.5463>.
- [77] A. Currin, N. Swainston, M. S. Dunstan, A. J. Jervis, P. Mulherin, C. J. Robinson, S. Taylor, P. Carbonell, K. A. Hollywood, C. Yan, E. Takano, N. S. Scrutton, R. Breitting, *Synth. Biol.* **2019**, *4*, 1, <https://doi.org/10.1093/synbio/ysz025>.
- [78] P. J. Zurek, P. Knyphausen, K. Neufeld, A. Pushpanath, F. Hoffelder, *Nat. Commun.* **2020**, *11*, 1, <https://doi.org/10.1038/s41467-020-19687-9>.
- [79] B. Buchfink, H. Ashkenazy, K. Reuter, J. A. Kennedy, H.-G. Drost, *bioRxiv* **2023**.
- [80] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. De Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, *Nucleic Acids Res.* **2018**, *46*, W296, <https://doi.org/10.1093/NAR/GKY427>.
- [81] M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Zidek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, S. Velankar, *Nucleic Acids Res.* **2022**, *50*, D439, <https://doi.org/10.1093/NAR/GKAB1061>.
- [82] R. K. Kuipers, H. J. Joosten, W. J. H. Van Berkel, N. G. H. Leferink, E. Rooijen, E. Ittmann, F. Van Zimmeren, H. Jochens, U. Bornscheuer, G. Vriend, V. A. P. Martins Dos Santos, P. J. Schaap, *Proteins: Structure, Function and Bioinformatics* **2010**, *9*, 78, <https://doi.org/10.1002/prot.22725>.
- [83] O. Trott, A. J. Olson, *J. Comput. Chem.* **2010**, *31*, 455, <https://doi.org/10.1002/JCC.21334>.
- [84] 'Molecular Operating Environment 2019.01, (MOE)', Chemical Computing Group ULC, 1010 Sherbrooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, **2019**.
- [85] P. R. Gerber, K. Müller, *J. Comput. Aided Mol. Des.* **1995**, *9*, 251, <https://doi.org/10.1007/BF00124456>.
- [86] S. Salentin, S. Schreiber, V. J. Haupt, M. F. Adasme, M. Schroeder, *Nucleic Acids Res.* **2015**, *43*, W443, <https://doi.org/10.1093/NAR/GKV315>.
- [87] E. Chovancova, A. Pavelka, P. Benes, O. Strnad, J. Brezovsky, B. Kozlikova, A. Gora, V. Sustr, M. Klvana, P. Medek, L. Biedermannova, J. Sochor, J. Damborsky, *PLoS Comput. Biol.* **2012**, *8*, e1002708, <https://doi.org/10.1371/JOURNAL.PCBI.1002708>.
- [88] R. Salomon-Ferrer, D. A. Case, R. C. Walker, *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 2, <https://doi.org/10.1002/wcms.1121>.
- [89] P. Eastman, V. S. Pande, *Comput. Sci. Eng.* **2010**, *12*, 34, <https://doi.org/10.1109/MCSE.2010.27>.
- [90] R. T. McGibbon, K. A. Beauchamp, M. P. Harrigan, C. Klein, J. M. Swails, C. X. Hernández, C. R. Schwantes, L. P. Wang, T. J. Lane, V. S. Pande, *Biophys. J.* **2015**, *109*, 8, <https://doi.org/10.1016/j.bpj.2015.08.015>.
- [91] D. Patsch, M. Eichenberger, M. Voss, U. T. Bornscheuer, R. M. Buller, *Comput. Struct. Biotechnol. J.* **2023**, *21*, <https://doi.org/10.1016/j.csbj.2023.09.013>.
- [92] C. E. Rasmussen, H. M. De, *J. Mach. Learn. Res.* **2010**, *11*, 3011.
- [93] P. A. Romero, A. Krause, F. H. Arnold, *Proc. Natl. Acad. Sci. U S A* **2013**, *110*, E193, https://doi.org/10.1073/PNAS.1215251110/SUPPL_FILE/SD03.TXT.
- [94] X. Huang, R. Pearce, Y. Zhang, *Bioinformatics* **2020**, *36*, 1135, <https://doi.org/10.1093/BIOINFORMATICS/BTZ740>.
- [95] S. Benevenuto, C. Pancotti, P. Fariselli, G. Birolo, T. Sanavia, *J. Phys. D. Appl. Phys.* **2021**, *54*, 245403, <https://doi.org/10.1088/1361-6463/ABEDFB>.
- [96] E. C. Alley, G. Khimulya, S. Biswas, M. AlQuraishi, G. M. Church, *Nat. Methods* **2019**, *16*, <https://doi.org/10.1038/s41592-019-0598-1>.
- [97] J. L. Watson, D. Juergens, N. R. Bennett, B. L. Trippe, J. Yim, H. E. Eisenach, W. Ahern, A. J. Borst, R. J. Ragotte, L. F. Milles, B. I. M. Wicky, N. Hanikel, S. J. Pellock, A. Courbet, W. Sheffler, J. Wang, P. Venkatesh, I. Sappington, S. V. Torres, A. Lauko, V. De Bortoli, E. Mathieu, S. Ovchinnikov, R. Barzilay, T. S. Jaakkola, F. DiMaio, M. Baek, D. Baker, *Nature* **2023**, *620*, <https://doi.org/10.1038/s41586-023-06415-8>.
- [98] R. Krishna, J. Wang, W. Ahern, P. Sturmfels, P. Venkatesh, I. Kalvet, G. R. Lee, F. S. Morey-Burrows, I. Anishchenko, I. R. Humphreys, R. McHugh, D. Vafeados, X. Li, G. A. Sutherland, A. Hitchcock, C. N. Hunter, M. Baek, F. DiMaio, D. Baker, *bioRxiv* **2023**, <https://doi.org/10.1101/2023.10.09.561603>.
- [99] D. M. Fowler, S. Fields, *Nat. Methods* **2014**, *11*, 801, <https://doi.org/10.1038/nmeth.3027>.
- [100] J. Büchler, S. H. Malca, D. Patsch, M. Voss, N. J. Turner, U. T. Bornscheuer, O. Allemann, C. Le Chapelain, A. Lumbroso, O. Loiseleur, R. Buller, *Nat. Commun.* **2022**, *13*, <https://doi.org/10.1038/s41467-022-27999-1>.
- [101] P. A. Romero, A. Krause, F. H. Arnold, *Proc. Natl. Acad. Sci. U S A* **2013**, *110*, E193, <https://doi.org/10.1073/pnas.1215251110>.
- [102] Y. Saito, M. Oikawa, H. Nakazawa, T. Niide, T. Kameda, K. Tsuda, M. Umetsu, *ACS Synth. Biol.* **2018**, *7*, 2014, <https://doi.org/10.1021/acssynbio.8b00155>.
- [103] J. Büchler, S. H. Malca, D. Patsch, M. Voss, N. J. Turner, U. T. Bornscheuer, O. Allemann, C. Le Chapelain, A. Lumbroso, O. Loiseleur, R. Buller, *Nat. Commun.* **2022**, *13*, 1, <https://doi.org/10.1038/s41467-022-27999-1>.
- [104] Y. Ogawa, Y. Saito, H. Yamaguchi, Y. Katsuyama, Y. Ohnishi, *ACS Synth. Biol.* **2023**, *2*, 12, <https://doi.org/10.1021/acssynbio.2c00577>.
- [105] A. Currin, N. Swainston, P. J. Day, D. B. Kell, *Chem. Soc. Rev.* **2015**, *44*, 1172, <https://doi.org/10.1039/c4cs00351a>.
- [106] A. Li, Z. Sun, M. T. Reetz, *Chembiochem* **2018**, *19*, 2023, <https://doi.org/10.1002/cbic.201800339>.
- [107] Y. Wang, Y. Zhao, A. Bollas, A. Wang, K. F. Au, *Nat Biotechnol.* **2021**, *39*, 1348, <https://doi.org/10.1038/s41587-021-01108-x>.
- [108] R. Noyori, *Adv. Synth. Catal.* **2003**, *345*, 15, <https://doi.org/10.1002/adsc.200390002>.
- [109] S. Rodriguez, B. Qu, K. R. Fandrik, F. Buono, N. Haddad, Y. Xu, M. A. Herbage, X. Zeng, S. Ma, N. Grinberg, H. Lee, Z. S. Han, N. K. Yee, C. H. Senanayakea, *Adv. Synth. Catal.* **2014**, *356*, 301, <https://doi.org/10.1002/adsc.201901040>.
- [110] F. D. Klingler, *Acc. Chem. Res.* **2007**, *40*, 1367, <https://doi.org/10.1021/ar700100e>.
- [111] C. S. G. Seo, R. H. Morris, *Organometallics* **2019**, *38*, 47, <https://doi.org/10.1021/acs.organomet.8b00774>.
- [112] J. D. Hayler, D. K. Leahy, E. M. Simmons, *Organometallics* **2019**, *38*, 36, <https://doi.org/10.1021/acs.organomet.8b00566>.
- [113] S. Hashiguchi, A. Fujii, K.-J. Haack, K. Matsumura, T. Ikariya, R. Noyori, *Angew. Chem. Int. Ed.* **1997**, *36*, 288, <https://doi.org/10.1002/anie.199702851>.
- [114] M. J. Kim, Y. Il Chung, Y. K. Choi, H. K. Lee, D. Kim, J. Park, *J. Am. Chem. Soc.* **2003**, *125*, 11494, <https://doi.org/10.1021/ja036766r>.
- [115] G. W. Huisman, J. Liang, A. Krebber, *Curr. Opin. Chem. Biol.* **2010**, *14*, 122, <https://doi.org/10.1016/j.cbpa.2009.12.003>.
- [116] D. L. Hughes, *Org. Process Res. Dev.* **2018**, *22*, 1063, <https://doi.org/10.1021/acs.oprd.8b00232>.
- [117] R. N. Patel, *Biomolecules* **2013**, *3*, 741, <https://doi.org/10.3390/biom3040741>.
- [118] S. K. Ma, J. Gruber, C. Davis, L. Newman, D. Gray, A. Wang, J. Grate, G. W. Huisman, R. A. Sheldon, *Green Chem.* **2010**, *12*, 81, <https://doi.org/10.1039/b919115c>.
- [119] X. M. Gong, Z. Qin, F. L. Li, B. B. Zeng, G. W. Zheng, J. H. Xu, *ACS Catal.* **2019**, *9*, 147, <https://doi.org/10.1021/acscatal.8b03382>.
- [120] J. Liang, J. Lalonde, B. Borup, V. Mitchell, E. Mundorff, N. Trinh, D. A. Kochrekar, R. N. Cherat, G. Ganesh Pai, *Org. Process Res. Dev.* **2010**, *14*, 1, 193, <https://doi.org/10.1021/op900272d>.
- [121] D. P. M. Depré, D. J. Ormerod, A. Horvath, T. S. Moody, M. Brossat, O. Riant, N. Vriamont, S. F. E. Lemaire, S. N. J. Hermant, *US20180093943A1*, **2018**.
- [122] S. Slagman, W. D. Fessner, *Chem. Soc. Rev.* **2021**, *50*, 1968, <https://doi.org/10.1039/d0cs00763c>.
- [123] C. Han, S. Savage, M. Al-Sayah, H. Yajima, T. Remarchuk, R. Reents, B. Wirz, H. Iding, S. Bachmann, S. M. Fantasia, M. Scalone, A. Hell, P. Hidber, F. Gosselin, *Org. Lett.* **2017**, *19*, 4806, <https://doi.org/10.1021/acs.orglett.7b02228>.
- [124] S. Bachmann, H. Iding, C. Lautz, I. Thomé-Pfeiffer, C. Maierhofer, R. Mondière, P. Schmidt, C. Strasser, T. Bär, A. Aebi, A. Schuster, *CHIMIA* **2021**, *75*, 605, <https://doi.org/10.2533/chimia.2021.605>.
- [125] J. F. Blake, R. Xu, J. R. Bencsik, D. Xiao, N. C. Kallan, S. Schlachter, I. S. Mitchell, K. L. Spencer, A. L. Banka, E. M. Wallace, S. L. Gloor, M. Martinson, R. D. Woessner, G. P. A. Vigers, B. J. Brandhuber, J. Liang, B. S. Safina, J. Li, B. Zhang, C. Chabot, S. Do, L. Lee, J. Oeh, D. Sampath, B. B. Lee, K. Lin, B. M. Liederer, N. J. Skelton, *J. Med. Chem.* **2012**, *55*, 8110, <https://doi.org/10.1021/jm301024w>.

- [126] Z. Shi, C. Sweeney, S. Bracarda, C. N. Sternberg, K. N. Chi, D. Olmos, S. K. Sandhu, C. Massard, N. Matsubara, M.-L. Harle-Yge, M. Nowicka, M. Wongchenko, J. S. De Bono, *J. Clin. Oncol.* **2020**, *38*, 182, https://doi.org/10.1200/JCO.2020.38.6_suppl.182.
- [127] R. Dent, M. Oliveira, S. J. Isakoff, S. A. Im, M. Espi , S. Blau, A. R. Tan, C. Saura, M. J. Wongchenko, N. Xu, D. Bradley, S. J. Reilly, A. Mani, S. B. Kim, K. S. Lee, J. H. Sohn, J. H. Kim, J. H. Seo, J. S. Kim, S. Park, M. Velez, S. Dakhil, S. Hurvitz, V. Valero, G. Vidal, R. Figlin, M. A. K. Allison, D. Chan, M. Cobleigh, V. Hansen, N. Iannotti, W. Lawler, M. Salkini, L. Seigel, G. Romieu, M. Debled, C. Levy, A. Hardy-Bessard, S. Guiu, L. G. Estevez, R. Villanueva, A. G. Martin, P. S. Rovira, A. Montano, M. I. C. Plaza, J. A. G. Saenz, I. Garau, B. Bermejo, E. V. Alonso, H. C. Wang, C. S. Huang, S. C. Chen, Y. H. Chen, L. M. Tseng, A. Wong, C. S. P. Ang, M. De Laurentiis, P. F. Conte, F. De Braud, F. Montemurro, L. Gianni, L. Dirix, *Breast Cancer Res. Treat.* **2021**, *189*, 377, <https://doi.org/10.1007/s10549-021-06143-5>.
- [128] K. Kita, T. Fukura, K. I. Nakase, K. Okamoto, H. Yanase, M. Kataoka, S. Shimizu, *Appl. Environ. Microbiol.* **1999**, *65*, 5207, <https://doi.org/10.1128/aem.65.12.5207-5211.1999>.
- [129] K. Kita, K. I. Nakase, H. Yanase, M. Kataoka, S. Shimizu, *J. Mol. Catal. B Enzym.* **1999**, *6*, 305, [https://doi.org/10.1016/S1381-1177\(98\)00108-8](https://doi.org/10.1016/S1381-1177(98)00108-8).
- [130] S. Kamitori, A. Iguchi, A. Ohtaki, M. Yamada, K. Kita, *J. Mol. Biol.* **2005**, *352*, 551, <https://doi.org/10.1016/j.jmb.2005.07.011>.
- [131] H. Li, D. Zhu, L. Hua, E. R. Biehl, *Adv. Synth. Catal.* **2009**, *351*, 583, <https://doi.org/10.1002/adsc.200900045>.
- [132] H. Li, Y. Yang, D. Zhu, L. Hua, K. Kantardjieff, *J. Org. Chem.* **2010**, *75*, 7559, <https://doi.org/10.1021/jo101541n>.
- [133] D. Zhu, Y. Yang, S. Majkowicz, T. H. Y. Pan, K. Kantardjieff, L. Hua, *Org. Lett.* **2008**, *10*, 525, <https://doi.org/10.1021/ol702638j>.
- [134] K. K. Yang, Z. Wu, F. H. Arnold, *Nat. Methods* **2019**, *16*, 687, <https://doi.org/10.1038/s41592-019-0496-6>.
- [135] M. T. Reetz, *Angew. Chem. Int. Ed.* **2013**, *52*, 2658, <https://doi.org/10.1002/anie.201207842>.
- [136] E. J. Ma, E. Siirola, C. Moore, A. Kummer, M. Stoeckli, M. Faller, C. Bouquet, F. Eggmann, M. Ligibel, D. Huynh, G. Cutler, L. Siegrist, R. A. Lewis, A. C. Acker, E. Freund, E. Koch, M. Vogel, H. Schlingensiepen, E. J. Oakeley, R. Snajdrova, *ACS Catal.* **2021**, *11*, 12433, <https://doi.org/10.1021/acscatal.1c02786>.
- [137] X. Chen, H. Zhang, J. Feng, Q. Wu, D. Zhu, *ACS Catal.* **2018**, *8*, 3525, <https://doi.org/10.1021/acscatal.8b00591>.
- [138] J. Li, J. Feng, X. Chen, J. Gong, Y. Cui, H. Zhang, D. Bu, Q. Wu, D. Zhu, *Org. Lett.* **2020**, *22*, 3444, <https://doi.org/10.1021/acs.orglett.0c00892>.
- [139] M. T. Reetz, J. D. Carballeira, *Nat. Protoc.* **2007**, *2*, 891, <https://doi.org/10.1038/nprot.2007.72>.

License and Terms



This is an Open Access article under the terms of the Creative Commons Attribution License CC BY 4.0. The material may not be used for commercial purposes.

The license is subject to the CHIMIA terms and conditions: (<https://chimia.ch/chimia/about>).

The definitive version of this article is the electronic one that can be found at <https://doi.org/10.2533/chimia.2024.108>