ChemCatChem

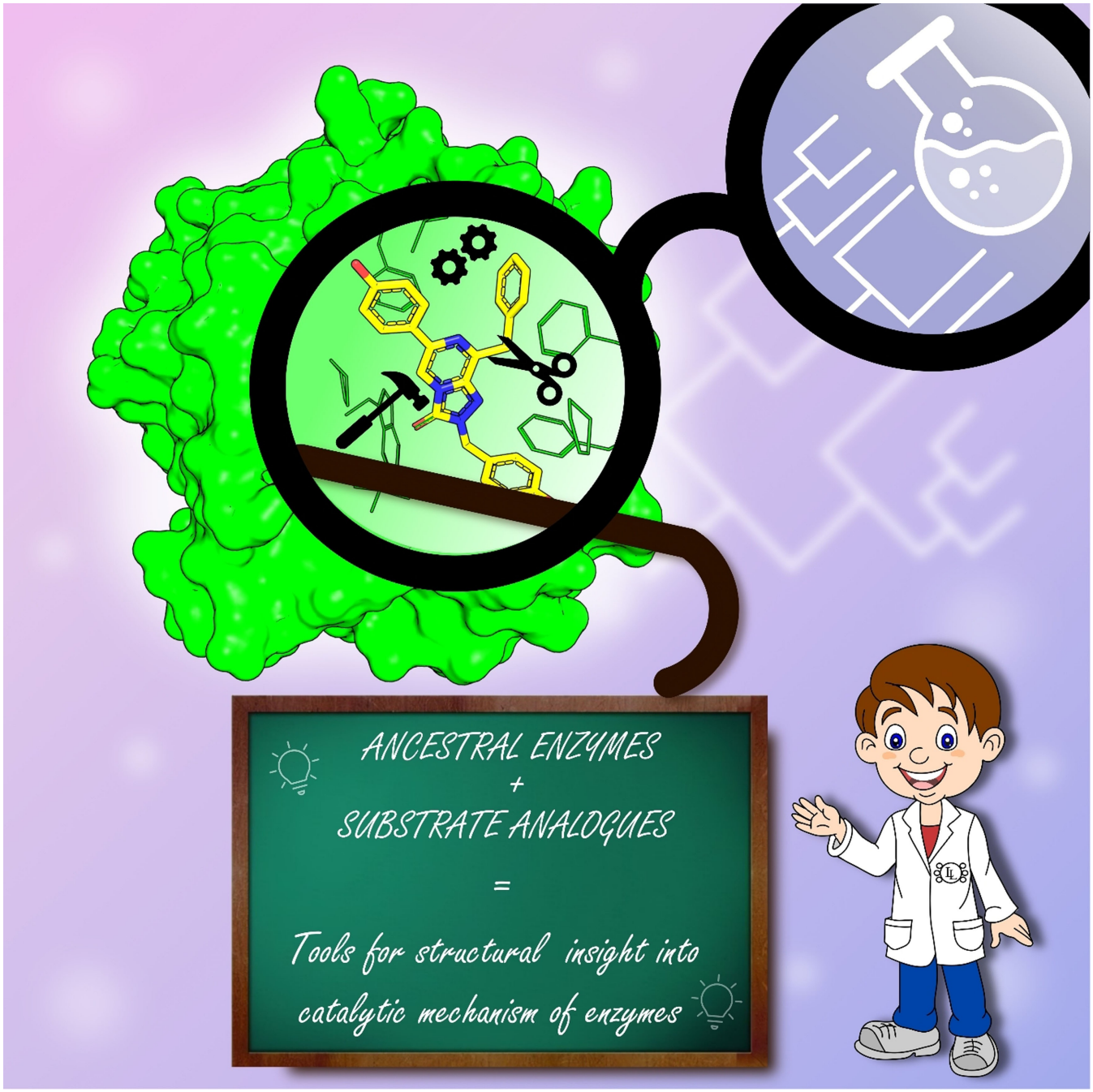
Review
doi.org/10.1002/cctc.202300745

Chemistry Europe
European Chemical Societies Publishing

www.chemcatchem.org

Deciphering Enzyme Mechanisms with Engineered Ancestors and Substrate Analogues

Tadeja Gao,[a, b] Jiri Damborsky,[a, b] Yves L. Janin,*[c] and Martin Marek*[a, b]



Environmentally friendly industrial and biotech processes greatly benefit from enzyme-based technologies. Their use is often possible only when the enzyme-catalytic mechanism is thoroughly known. Thus, atomic-level knowledge of a Michaelis enzyme-substrate complex, revealing molecular details of substrate recognition and catalytic chemistry, is crucial for understanding and then rationally extending or improving enzyme-catalyzed reactions. However, many known enzymes sample huge protein conformational space, often preventing complete structural characterization by X-ray crystallography. Moreover, using a cognate substrate is problematic since its conversion into a reaction product in the presence of the enzyme will prevent the capture of the enzyme-substrate conformation in an activated state. Here, we outlined how to deal with such obstacles, focusing on the recent discovery of a *Renilla*-type bioluminescence reaction mechanism facilitated by a combination of engineered ancestral enzyme and the availability of a non-oxidizable luciferin analogue. The automated ancestral sequence reconstructions using FireProt^{ASR} provided a thermostable enzyme suited for structural studies, and a stable luciferin analogue azacoelenterazine provided a structurally cognate chemical incapable of catalyzed oxidation. We suggest that an analogous strategy can be applied to various enzymes with unknown catalytic mechanisms and poor crystallizability.

1. Introduction

Enzymes are biological catalysts for reactions that in their absence would proceed very slowly or not at all. They exist in all living organisms, from bacteria and archaea to plants and animals, encompassing diverse biological processes, including digestion, metabolism, and DNA replication. Their function and substrate specificity depend on their unique three-dimensional (3D) structure together with an active site where their substrate is bound and then catalytically converted into a product.[1]

The efforts to understand the catalytic mechanisms of enzymes are not driven solely by pure scientific curiosity, but also by the motivation to develop and improve various fields of "real life" including drug design, industrial processes, biotechnology, and environmental remediation. Drugs such as statins and ACE inhibitors, used by millions of patients around the world, target specific enzymes in our body and cause inhibition or enhancement of enzymatic activity.[2] Understanding the mechanism of how enzymes catalyze reactions is crucial for the design of drugs that specifically target certain enzymes. Enzymes are also widely used in various industrial processes, such as food production, paper manufacturing, and biofuel production.[3] They are also important tools in biotechnology, such as DNA sequencing, protein engineering, and biocatalysis.[4] In environmental remediation, enzymes are used for bioremediation of oil spills, heavy metals, pesticides, and chlorinated contaminations.[5] The use of enzymes is rather abundant and there is a constant need for new as well as improved enzyme-based technologies. However, this can only be effectively achieved when the knowledge of enzymatic catalytic mechanisms is available.

Enzymatic catalysis is a dynamic process in which an enzyme undergoes several conformational changes. To obtain a complete picture of the catalytic mechanism, one often needs to employ a combination of several different techniques, each providing a certain piece of information. Structural changes upon interaction with ligands in a solution can be measured in real-time by single-molecule fluorescence resonance energy transfer (smFRET).[6] The fluorescent signal of two fluorophores attached to one molecule of an enzyme is sensitive to conformational changes that are detected by smFRET. Hydrogen/deuterium exchange mass spectrometry (HDX-MS),[7] provides information of a higher resolution compared to smFRET, namely resolution on a peptide or even on a single amino acid level. The principle of HDX-MS is an exchange of labile hydrogen atoms with deuterium in the enzyme's solvent-accessible regions. The difference in deuteration between enzymes with and without a ligand can reveal key protein motions and conformational changes induced by enzyme-ligand interaction. Among scattering techniques, Bio-SAXS (Small-Angle X-ray Scattering) can provide information about the overall shape, size, and flexibility of an enzyme-ligand complex in solution.[8] A big advantage of the above-mentioned techniques is the possibility of studying interactions and dynamics in solution and native conditions. However, they offer relatively low to medium resolution of enzyme structure compared to X-ray crystallography, cryo-electron microscopy (cryo-EM), and nuclear magnetic resonance (NMR). Therefore, such low to medium resolution techniques are often employed as a support to other experimental studies,[9] and they are particularly useful in setting constraints in molecular dynamics simulations and docking studies.[10]

[a] *Dr. T. Gao, Prof. J. Damborsky, Dr. M. Marek*
Loschmidt Laboratories
Department of Experimental Biology and RECETOX
Faculty of Science
Masaryk University
Kamenice 5, Bld. C13
625 00 Brno (Czech Republic)
E-mail: martin.marek@recetox.muni.cz

[b] *Dr. T. Gao, Prof. J. Damborsky, Dr. M. Marek*
International Clinical Research Center
St. Anne's University Hospital Brno
Pekarska 53
656 91 Brno (Czech Republic)

[c] *Dr. Y. L. Janin*
Structure and Instability of Genomes (StrInG)
Museum of National Natural History
INSERM, CNRS
Alliance Sorbonne University
43 Rue Cuvier - CP N° 26
75231 Paris Cedex 05 (France)
E-mail: yves.janin@cnrs.fr

A high-resolution 3D structure of the enzyme-substrate complex is required to understand the catalytic mechanism of particular enzyme, since this will reveal the interactions important for the binding and catalytic conversion of the substrate. Historically, structural biology relied only on experimental methods, including X-ray crystallography, NMR, and cryo-EM. Nowadays, various computational tools can predict a protein structure with a certain reliability using template-based or template-free modeling approaches. Homology modeling, the template-based approach, requires an experimentally determined 3D structure of one or more homologous proteins to build a model of the target structure.[11] The underlying assumption is that the 3D structure of the target protein is similar to one of the templates because of their high degree of sequence similarity. On the other hand, template-free modeling (*de novo* modeling) does not rely on similarity to a known structure and therefore it can be applied to proteins with rarer sequences/folds. Today, machine learning-based tools such as AlphaFold[12] and RoseTTAFold[13] are popular *de novo* modeling tools in the structural biology community.[14] In addition, a newly presented KarmaLoop[15] is a deep learning tool designed to predict loops in protein structure. However, one should keep in mind that the aforementioned models are only predictions. Recently, Terwilliger and co-workers reported that only some AlphaFold predictions match experimental maps closely, while most differ on a global scale.[16] Moreover, it is crucial to bear in mind that only an enzyme structure is not enough to understand functional and catalytic mechanisms. A case in point is the structure of *Renilla* luciferase (RLuc) which has been known for over 15 years[17] but a catalytic mechanism was only deciphered recently,[18] when co-crystallized with a non-oxidizable luciferin analogue called azacoelenterazine (azaCTZ). Another example is NanoLuc, an even smaller and superior luciferase, which structure has been known for 7 years[19] but a catalytic mechanism is yet to be reported. On the other hand, molecular docking and molecular dynamics simulations (MDs) are computational and cost-effective methods which provide insight into enzyme-ligand binding mode.[20] Molecular docking can screen thousands of compounds in a short period of time. However, both molecular docking and MDs may not accurately consider solvent effects, which can significantly affect ligand's binding mode and affinity for a catalytic site. Many times, they also require binding constraints and structural input of an enzyme, which are mostly provided by experimental studies.[21] Therefore, structures and binding modes obtained by computational tools should be treated as hypotheses and their experimental validations are essential to conduct proper scientific investigations. Table 1 summarizes the advantages and disadvantages of experimental and computational tools for studying structural insight into enzyme-ligand interaction.

Experimentally, the 3D structure of the enzyme-ligand complex can be obtained by employing X-ray crystallography,[22] NMR,[23] and cryo-EM.[24] Most structures deposited in RCSB PDB[25] are solved by X-ray diffraction (~85%), while only ~8%

Dr. Tadeja Gao obtained her Ph.D. in physical chemistry in 2018 under the supervision of Prof. Dr. Miha Lukšič at the University of Ljubljana, Slovenia. Following her doctorate, she pursued postdoctoral research in Physico-chimie des Electrolytes et Nanosystèmes Interfaciaux (PHENIX) group at Sorbonne University in Paris. Later, she joined Protein Structure and Dynamics group at Masaryk University, Brno, Czech Republic. Currently, she continues as a postdoctoral researcher at Loschmidt Laboratories at Masaryk University under the guidance of Dr. Martin Marek and Prof. Dr. Jiri Damborsky. She is interested in biophysical characterization of proteins and their structural features.

Prof. Dr. Jiri Damborsky holds the Josef Loschmidt Chair Professorship of Biochemistry at the Faculty of Science, Masaryk University, in Brno, Czech Republic. His research group focuses on protein engineering, developing new concepts, microfluidic devices, and software tools for protein engineering (Caver, CaverDock, Hotspot Wizard, FireProt, Calfitter, EnzymeMiner, SoluProt, LoopGrafter, etc.). They utilize these tools for the computational design of enzymes and regularly organize international courses. His group published over 300 original articles, 20 book chapters, and filed 11 international patents. He co-founded the first biotechnology spin-off from Masaryk University, Enantis Ltd.

Dr. Yves L. Janin obtained a Ph.D. in organic chemistry in 1993 under the guidance of Dr. Emile Bisagni at the Institut Curie. Following two years of postdoc at the ICSN, Gif/Yvette and a year in the Danish School of Pharmacy in Copenhagen, he then worked six years at the Institut Curie. After a sabbatical in the Vitry/Seine Aventis laboratories he then joined the Institut Pasteur until 2022. He is now working at the Muséum National d'Histoire Naturelle in Paris. His interest lies in medicinal chemistry and design of original bioluminescence-based reporting systems.

Dr. Martin Marek received his Ph.D. degree in 2007 from the Czech University of Life Sciences in Prague. After two years of postdoc at the Wageningen University, he worked six years at the Institute of Genetics and Molecular and Cellular Biology (IGBMC) in Strasbourg. Since 2017, he has worked at the Masaryk University and the International Clinical Research Center of St. Anne's University Hospital in Brno (FNUSA-ICRC). He wishes to understand how diverse protein catalysts catalyse chemical reactions that evolved in living creatures. To this end, he combines molecular biology, biochemistry, and structural biology techniques to obtain atomic-level insights into enzyme structure, function, and regulation.

Table 1. Comparison between experimental and computational tools for atomic-level structural characterization of enzyme-substrate complexes.

Methods	Advantages	Disadvantages
Experimental tools: X-ray crystallography, cryo-EM, and NMR	– Providing high-resolution structures. – Possibility to mimic native conditions. – High possibility to capture catalytically relevant conformation. – Screening of conformational space of the enzyme.	– Time-consuming: Optimization of sample preparation and measurement conditions. – Financially-consuming: Materials, chemicals, instruments. – Very flexible or disordered regions are not detected by X-ray crystallography. – Molecular weight limitations: NMR $<$ 40-60 kDa; cryo-EM $>$ 50 kDa.
Computational tools: Molecular docking and molecular dynamics simulations (MDs)	– Speed and efficiency: quick screening of a large number of ligands by molecular docking. – Insights into binding modes. – Possibility to apply to the systems that may be difficult to study experimentally.	– Dependence on force fields in MDs. -They may not always fully capture the complexity of enzyme-ligand interactions. – Predictions from computational methods need to be validated with experimental data. – Usually, they need constraints that are provided by experimental data. – Larger the system more computational- and time-consuming (especially MDs). – Extremely short time scale in MDs (usually between 10 ns and 10 μs) compared with the actual experimental time scale (s, min).

and ~7% of the structures are solved by EM and NMR, respectively. Nevertheless, the selection of a technique depends on the specific characteristics of the enzyme-ligand complex being studied, as well as the available resources and expertise of the research team. Advantages, disadvantages, requirements, and directions of technical development of all three techniques are listed in Table 2. The combination of two techniques in pursuit of solving a structure of one enzyme is possible as well.[26] When studying the catalytic mechanism of enzymes, a high-resolution structure is required. X-ray crystallography was often preferred over NMR and cryo-EM, but recent progress in cryo-EM makes it suitable for the determination of atomic-level structures of enzyme-ligand complexes too.[24,27] Solution NMR is limited to relatively small enzymes. However, technological advances, especially in solid-state NMR and cryo-NMR, might in the future elevate this limit. The use of NMR is expanding to new applications as recently demonstrated by Bhattacharya *et al.*[28] They identified mutagenic hot spots in enzymes by NMR and used the information for NMR-guided protein engineering. In any case, X-ray crystallography can easily achieve atomic-level resolutions, which is critical for understanding the precise interactions between the enzyme and ligand and has been used for decades as a reliable technique. It is appropriate for the study of a wide range of enzyme-ligand complexes, including small and large enzymes, soluble and membrane enzymes, and enzyme complexes with multiple subunits.[22,29] Moreover, recent developments in time-resolved serial crystallography allow capturing the enzyme catalysis states on a time scale, covering the whole catalytic cycle.[30]

Nevertheless, X-ray crystallography is a proper approach only when diffraction-quality crystals of the enzyme-ligand complex can be obtained. In other words, such a complex has to be crystallizable. The nightmare of a structural biologist begins when despite his/her best efforts to find the proper crystallization conditions, crystals do not grow. This report describes how to tackle this challenge by employing ancestral sequence reconstruction (ASR) followed by a resurrection of ancestral enzymes. It has been shown that ancestral enzymes often exhibit high thermal stability, high expression yield, and high crystallizability.[31] Although their sequence similarity with extant enzymes might be low, structure and catalytic machinery are highly conserved. All these traits make the use of ancestral enzymes a suitable approach to probe catalytic mechanism by X-ray crystallography.[18,31b]

A blueprint of enzymatic catalytic mechanism usually provides a detailed description of the following steps: (i) substrate recognition and binding (pre-catalytic), (ii) stabilization of a transition state to lower the activation energy required for the reaction to occur, (iii) catalysis, during which enzyme actively participates in the reaction, either by donating or accepting protons, electrons, or other chemical groups, and (iv) product release. To get a full repertoire of these steps, one should, ideally, "freeze" each of them into a crystal. This task might be easier when in a particular step co-factors (e.g., metal ions, vitamins, coenzymes) and/or co-substrates (e.g., ATP, NADH) are needed. Indeed, with the addition and withdrawal of such additives, one may navigate the outcome.[32] The task becomes more complex when there are no requirements for any co-factors and co-substrates or when these are not easily removed from the medium/protein. In this case, the conversion of a substrate pool into a product by the enzyme occurs very quickly (often in a second or minute time frame), therefore, the pre-catalytic step (substrate recognition and binding) may not be possible to capture. To tackle this challenge, one has two options: (i) inactivation of catalytic residue(s) of the enzyme by mutagenesis, or (ii) use of a substrate analogue that cannot be catalytically converted. However, there exists a certain risk when mutating the enzyme, since even a single point mutation can affect properties of the enzyme beyond the catalytic activity, such as structure, stability, and even the binding mode.[33] Nevertheless, when one decides on the second approach, the design of a substrate analogue should be

Table 2. Key factors to consider when selecting an experimental technique for atomic-level structural characterization of enzyme-substrate complexes together with directions of development for these methods to overcome specific technical challenges.

Technique	Advantages	Disadvantages	Requirements	Challenges and directions of developments
X-ray crystallography	– High-resolution structures can be obtained, typically at atomic resolution. – A large number of structures have been solved, enabling comparative studies. – Crystallography is relatively straightforward and robust. – No limit is applied regarding the molecular mass of an enzyme.	– Relatively large amount of target enzyme is needed. – Obtaining diffraction-quality crystals of enzyme-ligand complexes can be a challenging task with sampling many different crystallization conditions. – Enzyme conformational changes induced by crystal packing can lead to artifacts. – Not applicable for intrinsically disordered enzymes or intrinsically disordered regions. – Although the enzymes-ligand complex is crystalized, sometimes very flexible loops cannot be resolved.	– Access to a high-intensity X-ray source. – Production of diffraction-quality crystals of the enzyme-ligand complex.	Challenges: Static structure of an enzyme, radiation-sensitive enzymes, and non-native conditions. Directions of development: – Time-resolved crystallography: Capturing the structural changes during the reaction. – Free-Electron Lasers (FELs): Snapshots of proteins at various stages of the catalytic cycle. – Serial crystallography: Study of multiple microcrystals of unstable or radiation-sensitive enzyme intermediates. – In situ crystallography: Analyzing protein crystals within their native environment or relevant conditions.
NMR	– Structures of the enzyme-ligand complex in the solution (liquid) can be obtained. – The technique can be used for studying enzyme dynamics and flexibility. – NMR can be used for ligand screening and fragment-based drug discovery. – No limitations regarding ordered or disordered structures.	– The achievable resolution is typically lower than that of X-ray crystallography. – NMR is relatively insensitive, requiring a high concentration of enzyme (around 1 mM). – Signal-to-noise ratio can be a limiting factor. – Limited to enzymes with lower molecular mass (up to 40–60 kDa). – Expression of isotopically labeled enzyme can be expensive and challenging.	– Access to high-field NMR spectrometers. – Production of isotopically labeled enzyme-ligand complex.	Challenges: Low resolution for enzymes with high molecular weight. Directions of development: – Cryo-NMR: T below −100 °C enhances spectral resolution by reducing conformational dynamics. – Incorporation of non-natural amino acids in enzyme sequence. – Solid-State NMR: Protein is immobilized in a solid matrix, reducing the rapid tumbling.
Cryo-EM	– Structures can be obtained without the need for crystallization. – Cryo-EM can be used for studying large enzyme complexes and membrane enzymes. – Imaging can be performed at near-physiological conditions. – Low amount of enzyme is needed.	– The achievable resolution is typically lower than that of X-ray crystallography. – The technique is relatively new and requires specialized equipment and expertise. – Cryo-EM requires the preparation of homogeneous samples. – Small complexes with low molecular mass (below 50 kDa) might not be applicable for cryo-EM.	– Access to high-resolution electron microscopes. – Preparation of high-quality cryo-samples of the enzyme-ligand complex.	Challenges: Not applicable for proteins with low molecular mass (<50 kDa). Directions of development: – Direct Electron Detectors improve the sensitivity and efficiency of cryo-EM data collection. – Defocus Gradient Cryo-EM improves the resolution in regions with low contrast by varying defocus across the micrographs. – Volta Phase Plates (VPP) improve the contrast of low-molecular-weight samples by phase-shifting the electron wave before it reaches the specimen.

carefully planned to avoid altering the binding mode. Here, we focus on a luciferase substrate coelenterazine (CTZ), one of the most abundant marine luciferin.[34] CTZ is in the presence of a luciferase (an enzyme) and molecular oxygen (O_2) converted into a product called coelenteramide, followed by an emission of a blue photon.[35] This phenomenon is called bioluminescence.

In this review, we introduce ancestral enzymes and show how different studies benefited from their higher crystallizability, to solve diverse fundamental questions. Then, we focus on *Renilla*-type luciferase reaction with an emphasis on non-oxidizable substrate analogue, azaCTZ, which enables the molecular view of a pre-catalytic step of *Renilla* luciferase (RLuc). Collectively, we describe how Schenkmayerova and co-workers[18] exploited a combination of an ancestral enzyme and substrate analogue (azaCTZ) to decipher the catalytic mechanism of *Renilla*-type bioluminescence.

2. Engineering of ancestors as an approach to overcome poor crystallizability of extant enzymes

Ancestral enzymes were encoded by organisms that existed thousands, millions, or even billions of years ago. Nowadays, these organisms (and consequently their enzymes) are extinct. Although the resurrection of a woolly mammoth is still a work in progress,[36] the resurrection of ancestral enzymes is already possible and accessible to a broad scientific community.[37]

The process of acquiring the sequence of ancestral enzymes is known as ancestral sequence reconstruction (ASR). This is a probabilistic method that determines sequences of the extinct genes from which the extant genes evolved. ASR requires several steps: (i) creating a set of biologically relevant homologous sequences of extant enzymes from genomic databases, (ii) creating a multiple-sequence alignment (MSA) and a phylogenetic tree, and (iii) employing maximum-likelihood[38] and/or Bayesian inference[39] algorithms to infer the ancestral sequence from MSA and phylogenetic tree. The maximum-likelihood and Bayesian inference algorithms are accessible to the community via many tools developed over the years. However, for non-expert users, the limiting steps are acquiring the MSA of selected homologs and the rooted phylogenetic tree. Freely available web server FireProt^{ASR}[40] combines all these steps in a fully automated workflow in which the only input of the user is the sequence of the queried enzyme, and it does not require any particular expertise from a user.

The general workflow of ASR is shown in Figure 1, with the example of the reconstruction of a common ancestor from structurally similar but functionally distinct enzymes: haloalkane dehalogenases (HLDs) and *Renilla* luciferase (RLuc).[41] Phylogenetically, RLuc belongs to subfamily II of HLDs (HLD-II). Structurally, HLDs and RLuc share a common fold typical for α/β hydrolases. However, RLuc has developed a catalytically distinct activity towards monooxygenation, which results in bioluminescence in the presence of cognate luciferin substrate known as coelenterazine (CTZ). On the other hand, all HLD family members possess dehalogenase activity, in which the carbon-halogen bond is cleaved in halogenated hydrocarbons. To predict a common ancestor of extant RLuc and HLDs-II, Chaloupkova and co-workers employed the ASR strategy.[41] The resulting ancestral enzyme, named Anc$^{HLD-RLuc}$, shares 48% and 56% identity with sequences of the extant enzymes LinB (typical member of HLD-II) and RLuc8 (a stabilized eight-point mutant of RLuc), respectively. Although Anc$^{HLD-RLuc}$ has a low sequence identity, its structure closely resembles that of its descendants. Additionally, Anc$^{HLD-RLuc}$ shows a promiscuous catalytic nature by the ability to perform both monooxygenation (luciferase) and dehalogenation reactions, although its luciferase activity is low. Many other studies reported similar observations that ancestral enzymes often exhibit substrate or even catalytic promiscuity.[42] Ancestral enzymes can also exhibit a higher k_{cat} and enzymatic efficiency than extant enzymes, as Lin *et al*[43] demonstrated for ancestral Rubiscos.

The second striking trait of ancestral enzymes is their higher thermal stability compared to their extant enzymes.[42b,44] Moreover, they often show a higher solubility and expression yield.[31a,b, 45] These traits increase the probability of a successful crystallization and this is reflected in the rising number of deposited crystal structures of ancestral enzymes (Table 3). Schriever and co-workers[31a] illustrated the strategy to overcome the non-crystallizable nature of extant spiroviolene synthase (SvS). By employing ASR, they obtained a sequence of an ancestral enzyme which was more soluble, expressible, and crystallizable than extant SvS and it resulted in a high-resolution crystal structure of ancestral SvS. Although the first attempt to use this crystal structure for a homology modelling failed, a six-point mutant of the ancestral enzyme resulted in a crystal structure that was successfully employed in homology modeling to gain the 3D structure of the extant SvS.

Similarly, Nicoll and co-workers[31b] employed ASR to overcome the non-crystallizable nature of extant mammalian flavin-containing monooxygenases (FMOs). Extant mammalian FMOs are insoluble and reside in the membranes of the endoplasmic reticulum. Despite countless failed crystallization attempts of extant human FMO3 and FMO5, they successfully crystallized and determined structures of ancestral enzymes, AncFMOs. With this approach, they unveiled the structural basis of function in mammalian FMOs. They concluded that ancestral proteins should be routinely considered a protein crystallization tool. Furthermore, the recently published review by Nicoll *et al*[46] depicts the impact of ancestral enzymes on unveiling catalytic mechanisms.

The promiscuity and crystallizability of ancestral haloalkane dehalogenases (HLDs) were thoroughly investigated by Babkova and co-workers.[42b,47] In two studies, they resurrected five ancestral HLD enzymes in order to study substrate promiscuity and thermal stability,[42b] structural features by X-ray crystallography and conformational dynamics using molecular dynamics simulations.[47] They could demonstrate that promiscuous ancestral HLDs show significantly improved thermodynamic stability and restricted conformational dynamics compared to extant HLDs, which is in favor of crystallization.

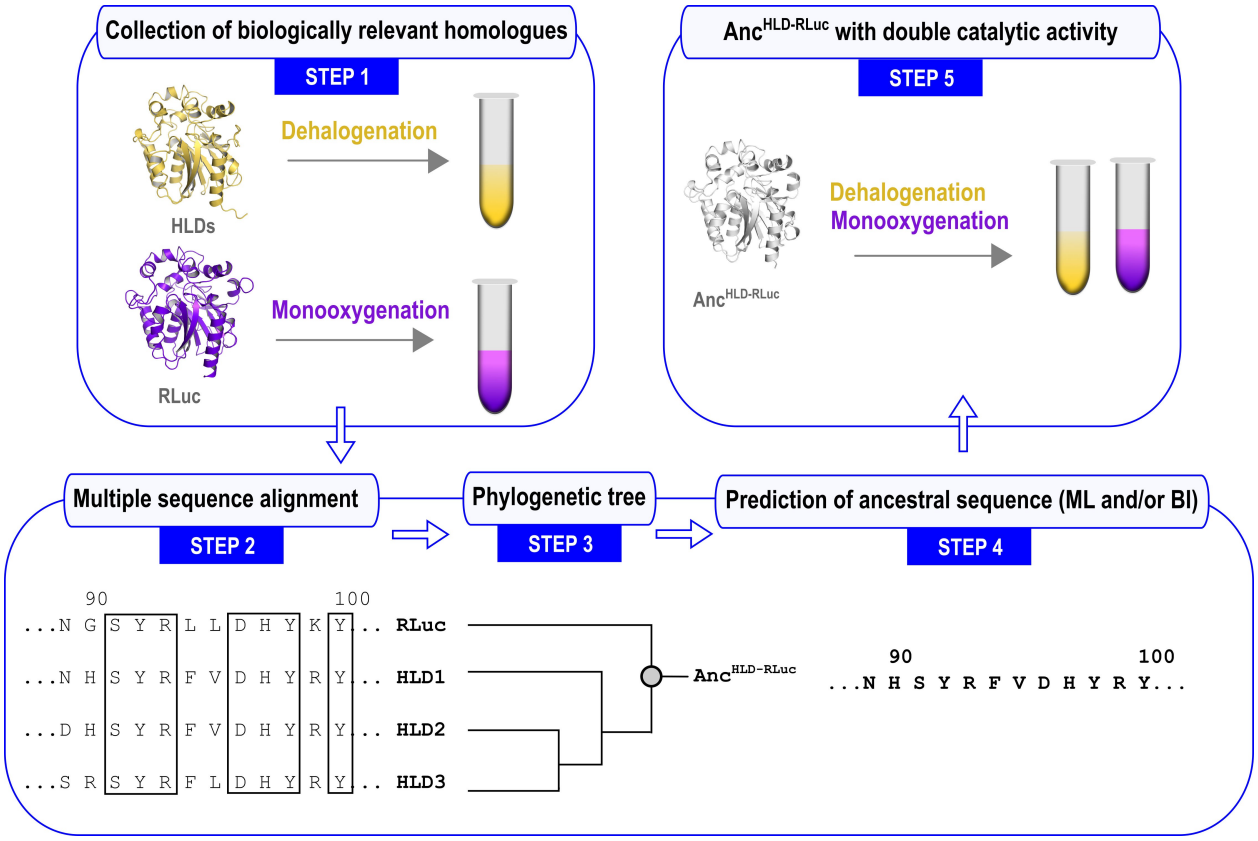


Figure 1. Workflow of ancestral sequence reconstruction (ASR). The depicted example represents a reconstruction of Anc$^{HLD-RLuc}$, a putative common ancestral enzyme of modern-day haloalkane dehalogenases (HLDs) and RLuc luciferase that exhibits dual-function catalytic activity (dehalogenation and monooxygenation).[41] The following steps are depicted: Step 1 (creating a set of biologically relevant homologous sequences of extant enzymes from genomic databases), step 2 (multiple-sequence alignment), step 3 (creating a phylogenetic tree), step 4 (employing maximum-likelihood (ML) and/or Bayesian inference (BI) algorithms to infer the ancestral sequence), and step 5 (production of an ancestral enzyme in the lab).

Although the sequence similarity of ancestral enzymes might be low compared to extant enzymes, their structure and catalytic machinery are typically conserved.[41,47] Together with their high thermal stability, high expression yield, high solubility, and crystallizability, reconstructed ancestral enzymes thus represent excellent scaffolds to probe the underlying catalytic mechanism by X-ray crystallography. However, when the enzyme of interest is an intrinsically disordered protein (IDP), X-ray crystallography is not the right choice. IDPs lack ordered 3D structure and they explore a large conformational space, which makes them non-crystallizable. Usually, IDPs have evolved so rapidly that ancestral sequence reconstruction (ASR) cannot be employed.[48] Other limitations of ASR for IDPs include difficulty in identifying functional regions within disordered regions, and limited availability of homologous sequences due to high sequence variability. Strikingly, IDPs represent over one-third of the eukaryotic proteome[49] and they participate in many physiological and pathophysiological functions, therefore, they are the focus of various studies, especially as drug targets.[50] On the other hand, nuclear magnetic resonance (NMR) has been successfully applied in many IDP-ligand interaction studies, making NMR the method of choice for IDPs.[51]

3. Chemical tools for structural studies of coelenterazine-using luciferases

Bioluminescence, a phenomenon involving the emission of light by living creatures, has fascinated scientists for a long time. The emitted light is generated via oxidation of a substrate (luciferin) and the process is catalyzed by an enzyme (luciferase). Nowadays, there are around 40 known luciferase-luciferin systems, many of which are widely exploited in biochemistry, agriculture, hygiene control, medicine, art, etc.[58] Futuristic predictions suggest that bioluminescent plants will replace streetlights and night lamps in our homes.[58a,59]

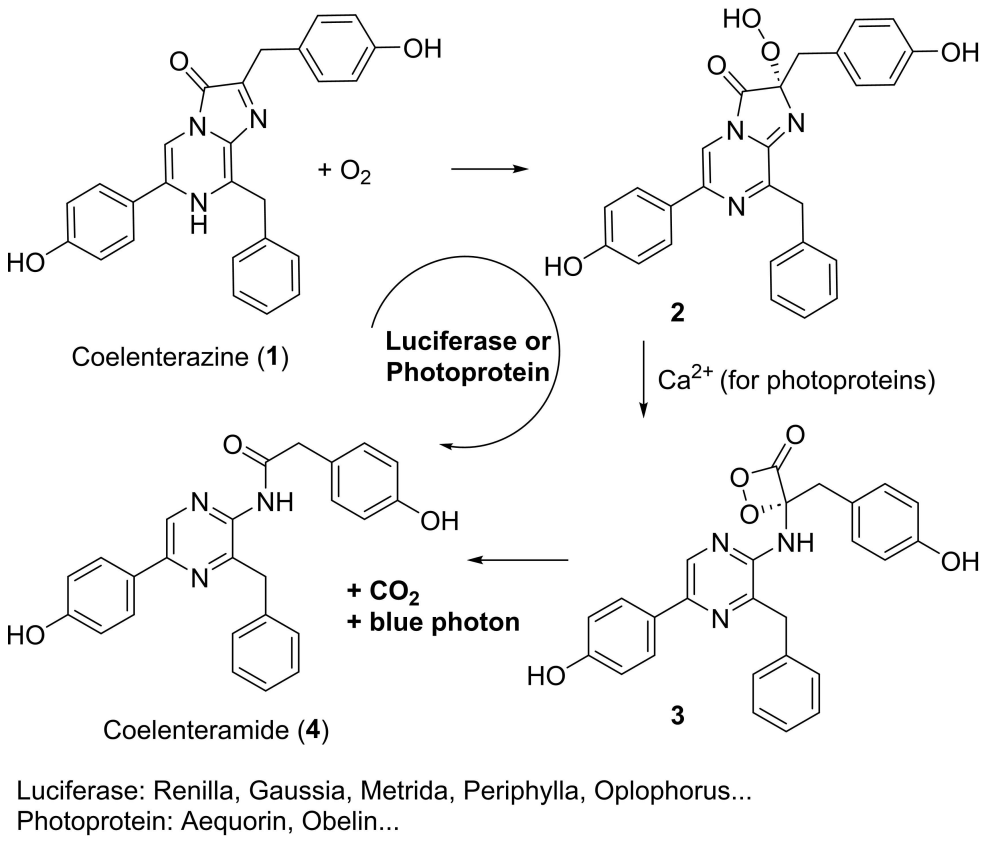
As mentioned above, the structural and mechanistic studies of many enzymes have greatly benefited from the use of chemical probes, which enable to go beyond the conformations these can adopt in their apo forms. Such tools were usually designed to mimic the substrates without undergoing the reaction catalyzed by the enzyme and actually inhibit its process. Other compounds can affect the function of the enzyme via allosteric interactions. In both instances, NMR-based or, when successfully co-crystallized, diffraction-based investigations often provided precious mechanistic insights. In the case of coelenterazine (**1**) using enzymes, the X-ray-based resolution of the structure of the photoprotein aequorin, which is binding

Table 3. Collection of ancestral enzymes whose crystal structures were exploited in various research fields.

Extant enzyme/ Extant enzyme family	Ancestral enzyme (PDB ID)	Exploitation of structure	Reference
Spiroviolene synthase (SvS)	SvSA2 (6TBD) SvS-A2 surface variant (6TIV)	Obtaining crystallizable ancestral enzyme. Homology model. Mechanistic basis of enzyme's promiscuity. Rational design.	[31a]
Haloalkane dehalogenases (subfamily HLD-II)	AncHLD2 (6Y9E) AncHLD3 (6Y9F) AncHLD5 (6Y9G)	Insight into restricted conformational dynamics of ancestral enzymes.	[47]
Haloalkane dehalogenase (HLD-II) *Renilla* luciferase (RLuc)	$Anc^{HLD\text{-}RLuc}$ (6G75)	Structural insight of catalytically promiscuous ancestral enzyme.	[41]
	AncFT (6S97)	Improvement of luciferase activity by transplantation of dynamic fragment from extant RLuc to ancestral $Anc^{HLD\text{-}RLuc}$.	[9a]
		Structural insight into catalytic mechanism of RLuc-CTZ system.	[18]
TEM-1 β-lactamase	GNCA (4B88) ENCA (3ZDJ)	Insight into thermal hyperstability and substrate promiscuity of ancestral enzymes.	[42a]
ADP-dependent sugar kinases	ancMT (5K27) ancMT E72A (5KKG)	Insights into the evolution of enzyme function and substrate specificity.	[52]
	AncMsPFK-F6P (6XIO)	Obtaining structure of ancestral enzyme-F6P complex. Revealing key residues for binding, catalysis, and ligand-induced conformational changes.	[42d]
Flavin-containing monooxygenases (FMOs)	AncFMO2 (6SEM) AncFMO3-6 (6SE3) AncFMO5 (6SEK)	Obtaining soluble and crystallizable ancestral enzyme. Structural basis of function in mammalian FMOs.	[31b]
Firefly luciferase	AncLamp (6K4C)	Evolution of luminescent color in fireflies.	[53]
LeuB/3-isopropylmalate dehydrogenase	Anc4 (3U1H)	Investigation of the evolution of thermophily.	[44b]
Metallo-β-lactamase family	Igni18 (6HRG)	Identification of enzyme variable regions accounting for modulation of activity, specificity, and oligomerization patterns.	[42e]
Cytochrome P450 1B subfamily	Ancestral CYP1B1 (6OYU, 6OYV)	Investigation of key features that contribute to thermostability.	[54]
Bacterial ribonuclease H1 (RNH)	AncC (4LY7)	The evolution of diverse thermostability in extant RNHs.	[55]
Lactate dehydrogenases (LDHs)	AncMDH2 (4PLW) AncLDH* (4PLG) AncMDH2-INS (4PLY)	Evolution of apicomplexan lactate dehydrogenases.	[56]
Adenylate kinase	ANC1 (5G3Y) ANC3 (5G3Z) ANC4 (5G41)	Mechanism of thermo-adaptation of enzyme catalysis in adenylate kinases.	[57]

the hydroperoxycoelenterazine (**2**) was one of the achievements of Osamu Shimomura research group (PDB ID: 1EJ3).[60] This structure, which demonstrates the occurrence of this hydroxyperoxide intermediate, nicely confirmed the oxidative decarboxylation process, catalyzed by many types of luciferases.[61] As depicted in Scheme 1 in the case of the many coelenterazine-using marine luciferases it proceeds via the two per-oxidated intermediates **2** and **3** and thus transforms coelenterazine (**1**) into coelenteramide (**4**) and carbon dioxide along with the emission of a blue photon.

Many structures of semisynthetic aequorins binding hydroperoxydated coelenterazine analogues depicted in Figure 2, PDB ID: 1UHH-binding **5**, 1UHI-binding **6**, 1UHJ-binding **7**, 1UHK-binding **8**[32a] and PDB ID: 5ZAB-binding **9**[32b], were then reported as well as structures of other photoproteins binding the hydroperoxide **2**, such as Mitracona (PDB ID: 4NQG), Clytin (PDB ID: 3KPX)[62] or Obelin (PDB ID: 1JF0 or 1SL9, 1EL4,[63] 1JF2,[64] 1QV0 and 1QV1,[65] 4N1F and 4MRX).[66] All these provided remarkable insights in the binding mode of coelenterazine to the photoproteins, but a crucial aspect is that these enzymes are a special case of luciferase which "freezes" after the peroxidation of coelenterazine (**1**). Then, the trapped hydroperoxide (**2**) will not undergo the oxidative decarboxylation process leading to coelenteramide (**4**) and a photon until a calcium salt is added. This property was instrumental in the success of these co-crystallization experiments, which were



Scheme 1. Two stages of the catalytic oxidative decarboxylation of coelenterazine (**1**) into coelenteramide (**4**) via per-oxidated intermediates (**2** and **3**).

conducted in the absence of calcium salts. In fact, more recent work led to an Obelin structure binding compound **10** (PDB ID: 7O3U), and upon the addition of a calcium salt, structures of Obelin binding the fully hydrolyzed reaction products **11** (or coelenteramide (**4**), from coelenterazine) were obtained (PDB ID: 8A9S and 1S36).[67] On the other hand, the structural studies of "true" coelenterazine-using luciferases have been greatly hampered by the fact that these are instantly consuming coelenterazine (**1**). Indeed, co-crystallization experiments with the *Renilla* luciferase (RLuc) and coelenterazine (**1**) led to structures in which the luciferase is binding coelenteramide (**4**) (see PDB ID: 2PSJ,[68] 6YN2,[9a] and 7OMR[18]) or the fully hydrolyzed reaction product coelenteramine (**12**) (PDB ID: 7OMO[18]). Concerning apo luciferases, the growing interest in bioluminescence-based reporting systems provided the impetus leading to X-ray-based structures of the RLuc (PDB ID: 2PSE, 2PSF, 2PSD, 2PSH, and 2RH7)[68] and the NanoLuc/NanoKAZ luciferase, which is derived from the catalytic subunit of *Oplophorus* luciferase, (PDB ID: 5IBO, 7MJB, 7SNR, 7SNS, 5B0 U[19] or 7VSX[69]). Moreover, the NMR-derived structure of the apo *Gaussia* luciferase was also reported (PDB ID: 7D2O).[70] One of the potential interests of these studies is to provide clues on how to modify the luciferin and/or the catalytic site of these luciferases in order to improve their bioluminescence properties. Securing apo structures is a good start toward such a goal. However, obtaining structures of luciferase binding small molecules is also of real importance, especially to provide information on the proceeding of these catalyzed reactions. Accordingly, attempts were made to undertake structure-based studies using inhibitors of the corresponding luciferase as well as compounds more closely related to coelenterazine (**1**). For the first approach, the earliest report was made in the course of a conference which described the X-ray-based structure of NanoLuc binding compound **13** (Figure 3).[71] Compound **13** is an example in a series of NanoLuc inhibitors patented[72] and then described for their potential use as cell-permeable or impermeable tools in bioluminescence-based multiplexing experiments.[73] Unfortunately, the X-ray-based structure of triamide **13** bound to NanoLuc has yet to be published. Interestingly, NanoLuc inhibition by various compounds, sulfamides being recurrent, has been described[73–74], and one sulfamide series, illustrated by compound **14**, was patented.[75] Finally, a recently posted X-ray-based structure of NanoLuc binding the sulfonic acid derivative **15** (PDB ID: 7SNW) should lead to an interesting publication. In the absence of such reports, we can at least state that these structures could provide crucial information, not only on the light production mechanism of this luciferase but also on the class of compounds which may inhibit this enzyme and thus hamper high throughput screenings and/or diagnostics based on this technology. Concerning inhibitors structurally related to coelenterazine (**1**), an early report actually described the strong inhibition of RLuc luciferase by coelenterazine sulfate **16**.[76] Further research aiming at probing the catalytic site of coelenterazine-using luciferase, led to the synthesis and

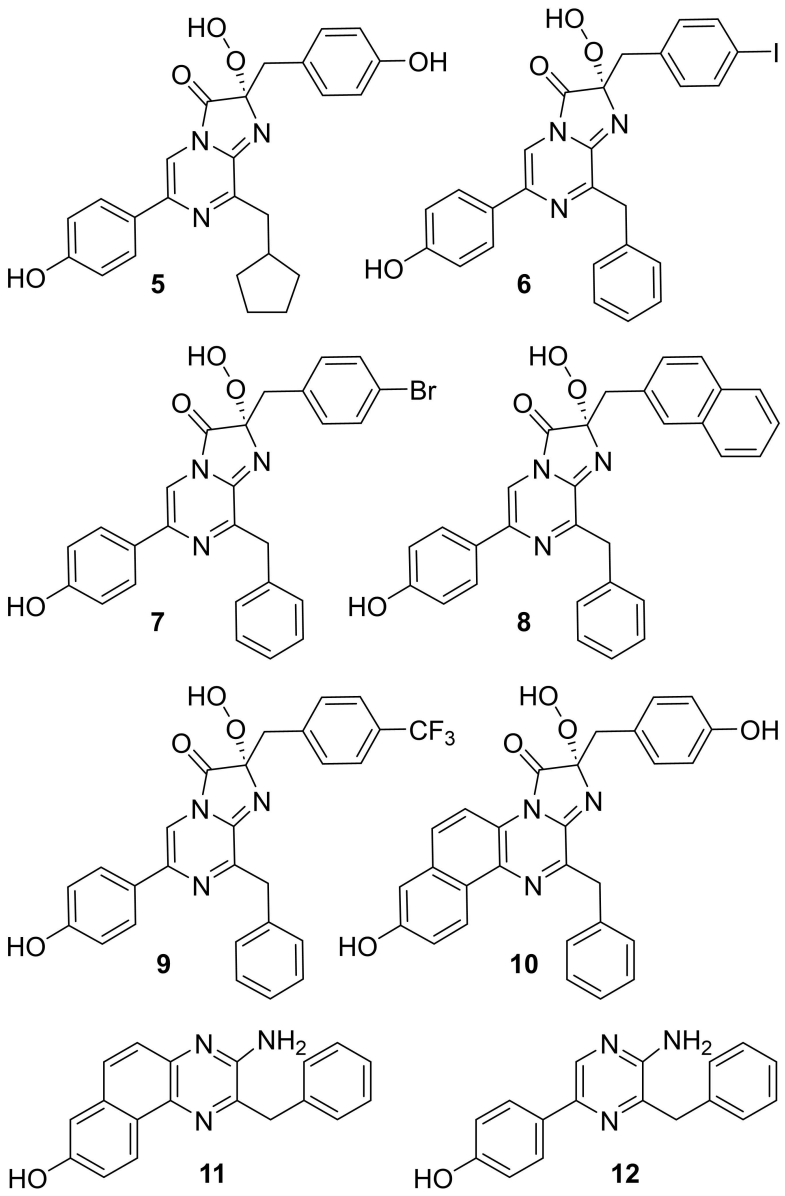


Figure 2. Structures of compounds **5–12**.

enantiomer separation of stable transition state analogues such as the ketones and hydroxymethyl-bearing derivatives **17** and **18**.[77] The ketones turned out to be strong inhibitors of the RLuc[77a], and more recent work reported the stereoselective inhibition of the regeneration step of the photoprotein aequorin by both (S) enantiomers of **17** and **18**. Moreover, the crystal structures of aequorin binding these compounds were resolved (PDB ID: 7EG3-binding (S)-**17** and 7EG2-binding (S)-**18**).[77c] Further research with these chemical probes then demonstrated that these coelenterazine analogues were also inhibiting various luciferases derived from the catalytic subunit of *Oplophorus* luciferase as well as *Gaussia* or RLuc.[69] However, attempts to co-crystallize these compounds with the reverse mutant of NanoLuc/nanoKAZ (QL-nanoKAZ) were, so far, not successful.[69]

In the course of a similar strategy, we designed azacoelenterazine (azaCTZ, **18**) which features nitrogen in place of the carbon subject of oxidation in coelenterazine (**1**). This compound turned out to be quite insoluble, and this probably explained its rather modest inhibition of *Gaussia* luciferase.[78] However, NMR-based studies of its interaction with this luciferase provided some insights into their dynamic interaction's complexity.[78] Indeed, this chemical shift perturbations-based approach not only detected amino acid residue near or part of a hydrophobic cavity which is likely to be the catalytic site of *Gaussia* luciferase but also highlighted the involvement of more remote regions, which hints at large protein movements taking place upon an initial luciferin binding. Moreover, as described in far more detail in this review, azaCTZ (**18**) turned out to be quite useful for diffraction-based investigations of RLuc as well as NanoLuc/nanoKAZ luciferases (PDB ID:

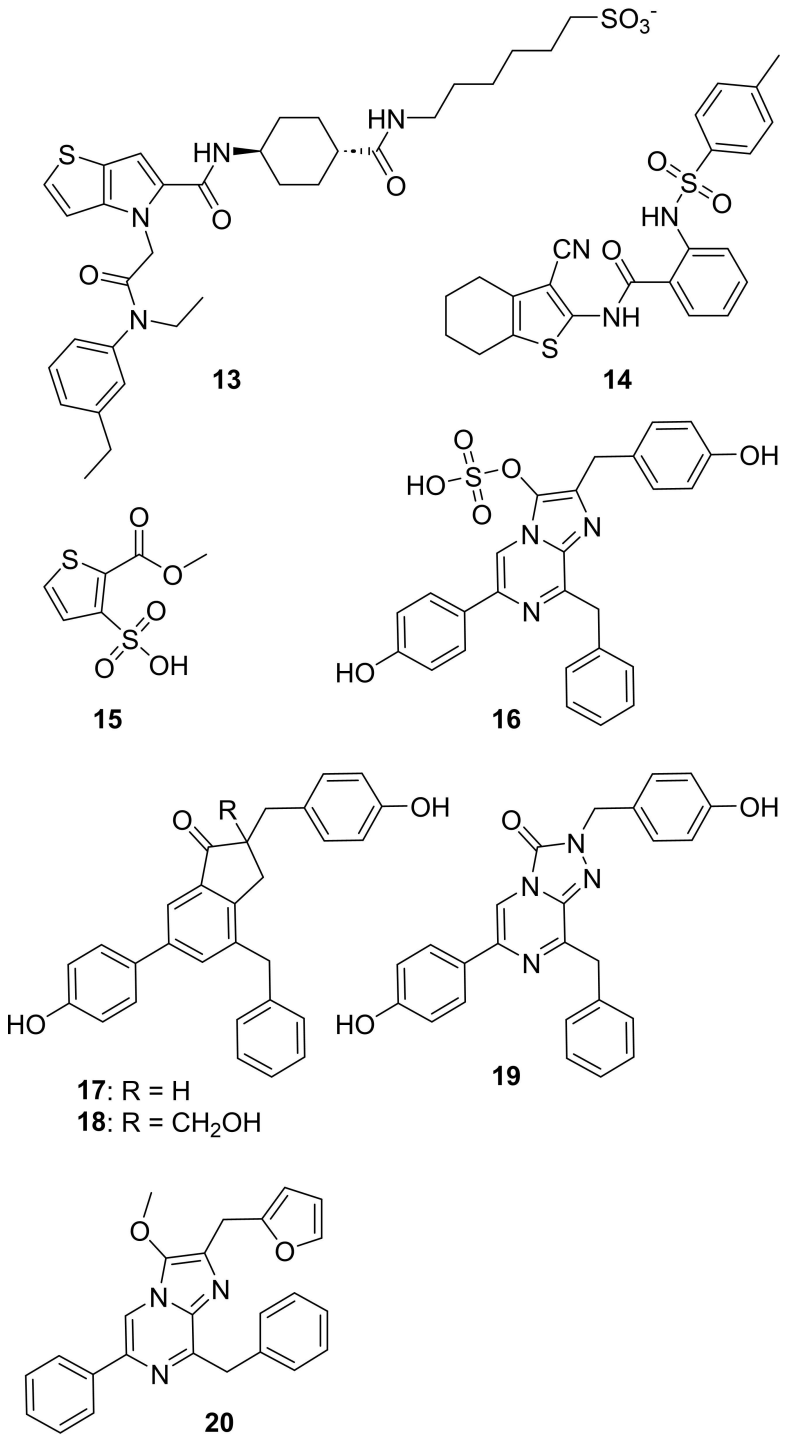


Figure 3. Structures of compounds **13–20**.

7QXR, 7OMD, or 7OME).[18,79] Finally, as a conclusion of this section, the recently released X-ray-based structure of NanoLuc/nanoKAZ binding the 3-methoxy-furimazine derivative **20** (PDB ID: 7SNT) is certainly heralding even more results in this research domain.

4. Case study: the delineation of the molecular mechanism of *Renilla*-type catalysis

One of the widely exploited bioluminescent systems is based on a luciferase isolated from the sea pansy, *Renilla reniformis*.[80] As illustrated above (Scheme 1), *Renilla* luciferase (RLuc) is a cofactor-independent monooxygenase which catalyzes the conversion of coelenterazine (CTZ, **1**) to coelenteramide (CEI, **4**),

and as a result, it emits blue light (~480 nm).[81] The RLuc8,[17,82] an eight-point mutant amenable to crystallization as opposed to wild-type RLuc, is commonly used in research labs.[58a] Even though the RLuc-CTZ system has already been known for over 40 years, the catalytic mechanism of RLuc was deciphered only recently by Schenkmayerova and colleagues.[18] The two major obstacles that have been keeping scientists from unraveling the *Renilla*-type catalytic mechanism were: (i) the lack of stable non-oxidizable CTZ analogue prevented the acquisition of the pre-catalytic enzyme-substrate complex in catalytically favored state and (ii) intrinsic flexibility and malleability of the catalytic pocket of extant RLuc enzymes complicated the crystallization attempts. These two obstacles were addressed and overcome in the recent ground-breaking study by Schenkmayerova and co-workers.[18] They successfully captured crystallographic snap-shots of pre- and post-catalytic ligand-bound luciferase complexes, which were crucial to unravel the underlying catalytic mechanism.

This was achieved with the use of the non-oxidizable CTZ analogue, azaCTZ (**19**) which was already described in the previous section. In CTZ the C2 core carbon atom is attacked by O_2 during the monooxygenation reaction. Therefore, C2 is replaced by a nitrogen atom in azaCTZ to prevent an O_2 attack, and consequently, it is protected from catalytic conversion. The synthetic pathway of azaCTZ was published in ref.[18]. After a successful design of azaCTZ, the co-crystallization of RLuc8 mutants in the presence of excess azaCTZ or CTZ was undertaken. CTZ was predictably converted into the post-catalytic product CEI (coelenteramide, **4**) or CMN (coelenteramine, **12**, a less common product) while azaCTZ (**19**) remained bound in a pre-catalytic state. Crystallization studies with extant RLuc variants highlighted its active-site pocket's high flexibility and malleability. This was the most pronounced in the case of the azaCTZ-bound enzyme complex, where up to three azaCTZ molecules bound to the same active pocket were observed (Figure 4A). All in all, those structures demonstrated that extant RLuc explores a large conformational space due to its intrinsic flexibility. This feature might be important in further steps of catalysis (e.g., unbinding of the product), but it complicated the attempts to obtain the structure of the catalytically favored Michaelis enzyme-substrate complex. To overcome this flexibility, they employed in-lab engineered ancestral enzyme AncFT, which was proven to have a more rigid structure than extant variants.[9a] AncFT has been engineered from the ancestral enzyme $Anc^{HLD-RLuc}$ (see **section 2** and ref.[41]). The loop-helix fragment (L9-α4) of thermostable ancestor $Anc^{HLD-RLuc}$ was replaced (transplanted) with the corresponding fragment L9-α4 from the extant RLuc. This fragment transplantation resulted in an engineered ancestor AncFT which exhibited a substantially improved luciferase activity.[9a] Then, co-crystallization of AncFT with azaCTZ or CTZ provided structures of substrate- and product-bound enzyme complexes in catalytically favored conformation (Figure 4A).[18] These structures revealed the key molecular interactions between enzyme and luciferin molecules. With the help of complementary kinetic and molecular dynamics studies, they were able to propose a full picture of CTZ-to-CEI conversion. As summed up in Figure 4B, four major steps taken in order to discover this catalytic mechanism were: (i) reconstruction of $Anc^{HLD-RLuc}$ scaffold,[41] (ii) engineering high-performance AncFT luciferase from $Anc^{HLD-Rluc}$ [9a] (iii) design and synthesis of non-oxidizable CTZ analogue (azaCTZ),[18] and (iv) co-crystallization of AncFT with azaCTZ.[18]

The benefits expected when employing ancestral enzymes and/or substrate analogues are summarized in Figure 5. Many ancestral enzymes show their promiscuous nature by binding various substrates[42] or even catalyzing various reactions,[9a,41] unlike their descendants who developed substrate and catalytic specificity. However, a disagreement between studies can be noticed when their promiscuity is explained based on their intrinsic flexibility. On the one hand, studies on promiscuous ancestral HLDs (AncHLDs)[42b] show that AncHLDs exhibit less flexible structures compared to their descendants.[47] This was further confirmed by the low intrinsic flexibility seen for ancestral AncFT in comparison with its more flexible descendant counterpart (RLuc).[18] On the other hand, studies on the promiscuous ancestral β-lactamases[42c,83] show higher intrinsic flexibility than their descendants. However, this did not affect the growth of crystals which were obtained with sufficient quality to solve the 3D structure.[42a] In any case, the degree of confusion provided by these somehow different observations is of interest and points out a direction for future studies.

5. Conclusions

A famous quote: "If you want to understand function, study structure" by a co-discoverer of DNA structure Francis Crick shows the direction for protein engineers and researchers in their pursuit to decipher the catalytic mechanism of enzymes. However, knowing only the molecular structure of enzymes is not enough to achieve this goal. Therefore, one must acquire structural knowledge of the enzyme-ligand complex, which reveals key molecular interactions important for catalysis. X-ray crystallography is the most exploited technique for this matter, as it routinely provides structures at atomic-level resolution. However, a bottleneck can be obtaining diffraction-quality crystals, which might be impossible when the enzyme is not prone to crystallize. Another obstacle in obtaining a complete blueprint of the catalytic mechanism is the uncontrollable conversion of a substrate into a product in the presence of an enzyme, which disables studying pre-catalytic conformations.

In this review, we outlined how to deal with these two challenges and demonstrated that (i) highly stable and crystallizable ancestral enzymes can replace a non-crystallizable extant enzyme of interest to study catalytic mechanism, and (ii) a substrate analogue, mimicking as much as possible the cognate substrate, can lead to pictures of the pre-catalytic step. Ancestral sequence reconstruction is now accessible to non-experts thanks to fully automated web servers, like $FireProt^{ASR}$. We suggest that the outlined strategy generally applies to many enzymes with unknown catalytic mechanisms.

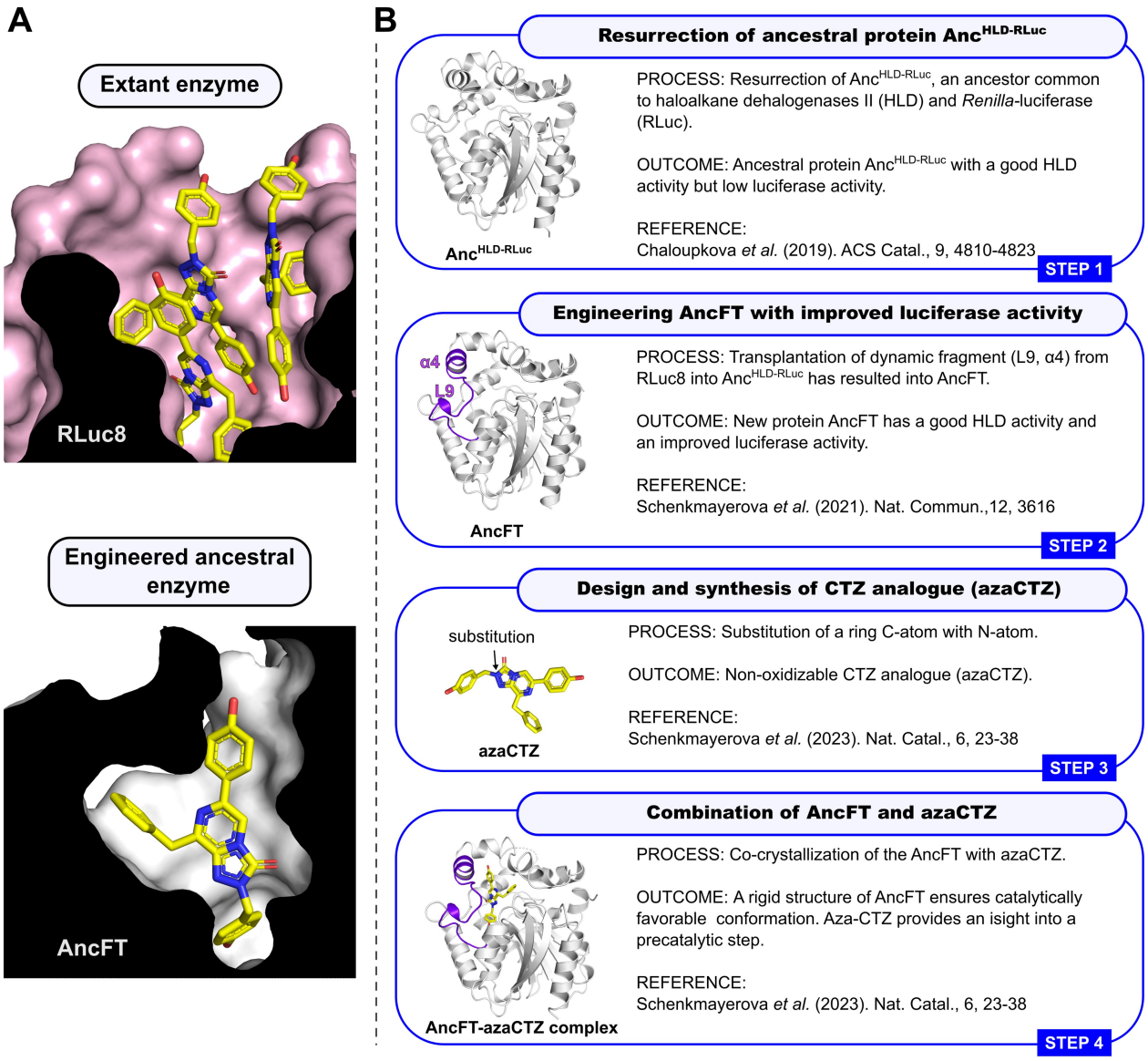


Figure 4. (A) Structure of the extant enzyme (RLuc8, PDB ID: 7OMD) and the engineered ancestral enzyme (AncFT, PDB ID: 7QXR) in a complex with azaCTZ. Three azaCTZ molecules are bound in a voluminous pocket of the extant enzyme RLuc8. On the other hand, the rigid nature of the engineered ancestral enzyme AncFT provides a catalytically favorable complex. (B) The workflow of the RLuc-CTZ case. Four major steps were taken in order to acquire structural insight into RLuc-CTZ catalytic activity. The resurrection of the ancestral enzyme using FireProt^{ASR} (step 1[41]), an improvement of luciferase activity of the ancestral enzyme (step 2[9a]), a design and synthesis of azaCTZ (step 3[18]), and a co-crystallization of the ancestral enzyme AncFT with azaCTZ to obtain the biologically relevant pre-catalytic state (step 4[18]).

Acknowledgements

The work on this paper was supported by the Czech Science Foundation grant GA22-09853S, the Czech Ministry of Education (TEAMING CZ.02.1.01/0.0/0.0/17_043/0009632, RECETOX RI LM2018121 and LM2023069, e-INFRA LM2018140), the Technology Agency of Czech Republic (RETEMED TN02000122). This project has received funding from the project National Institute for Neurology Research (Programme EXCELES, ID Project No. LX22NPO5107) funded by the European Union - Next Generation EU. This project has received funding from the European Union's Horizon Europe program under grant agreement No 101087124 and was supported from the European Union's Horizon 2020 research and innovation program under grant agreement No 857560. This publication reflects only the author's view, and the European Commission is not responsible for any use that may be made of the information it contains. The design and synthesis of azaCTZ also benefited from the Valoexpress funding calls of the Institut Pasteur.

Conflict of Interests

The authors declare no conflict of interest.

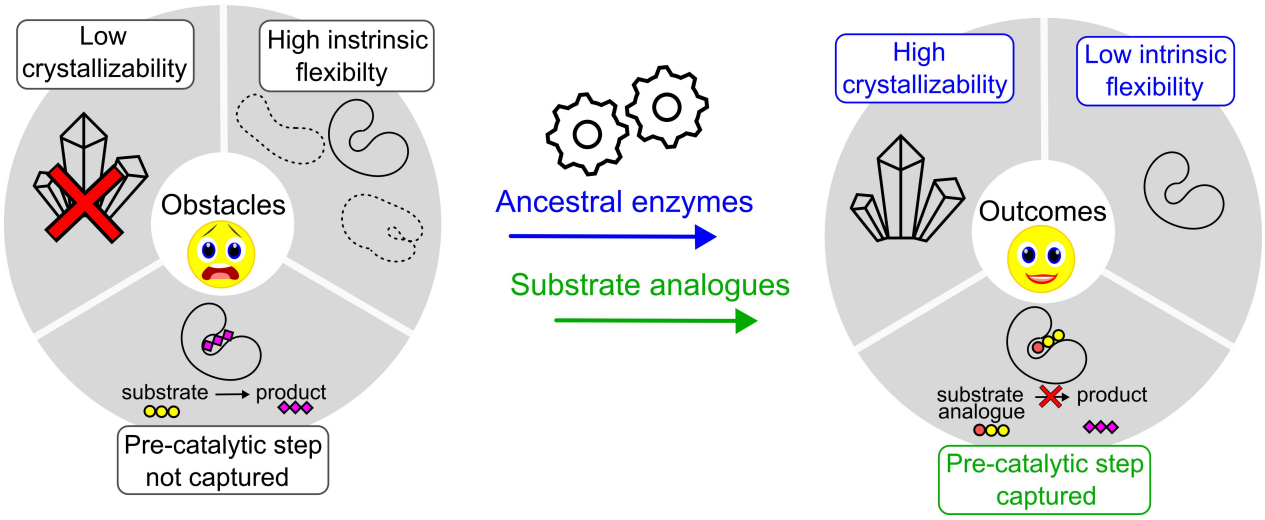


Figure 5. Combining ancestral enzymes and substrate analogues can tackle low crystallizability, high conformational flexibility, and undesirable substrate conversion into a product.

Keywords: ancestral sequence reconstruction · azacoelenterazine · biocatalysis · bioluminescence · coelenterazine · luciferase · enzymes · FireProt ASR · reaction mechanism · substrate analogues

[1] a) H. Suzuki, *How enzymes work: from structure to function*, 2nd ed., Crc Press, **2019**; b) L. J. Kingsley, M. A. Lill, *Proteins* **2015**, *83*, 599–611.
[2] a) C. Stancu, A. Sima, *J. Cell. Mol. Med.* **2001**, *5*, 378–387; b) N. J. Brown, D. E. Vaughan, *Circulation* **1998**, *97*, 1411–1420; c) M. Shahriari, M. Zahiri, K. Abnous, S. M. Taghdisi, M. Ramezani, M. Alibolandi, *J. Controlled Release* **2019**, *308*, 172–189.
[3] a) S. Raveendran, B. Parameswaran, S. B. Ummalyma, A. Abraham, A. K. Mathew, A. Madhavan, S. Rebello, A. Pandey, *Food Technol. Biotechnol.* **2018**, *56*, 16–30; b) G. K. Gupta, M. Dixit, R. K. Kapoor, P. Shukla, *Mol. Biotechnol.* **2022**, *64*, 130–143; c) M. D. N. Ramos, S. Thais Milessi, R. Garcia Candido, A. Aguiar Mendes, A. Aguiar, *Energy Sustain. Dev.* **2022**, *68*, 103–119.
[4] B. Wiltschi, T. Cernava, A. Dennig, M. Galindo Casas, M. Geier, S. Gruber, M. Haberbauer, P. Heidinger, E. Herrero Acero, R. Kratzer, C. Luley-Goedl, C. A. Müller, J. Pitzer, D. Ribitsch, M. Sauer, K. Schmölzer, W. Schnitzhofer, C. W. Sensen, J. Soh, K. Steiner, C. K. Winkler, M. Winkler, T. Wriessnegger, *Biotechnol. Adv.* **2020**, *40*, 107520.
[5] a) C. M. Quintella, A. M. T. Mata, L. C. P. Lima, *J. Environ. Manage.* **2019**, *241*, 156–166; b) P. Sharma, S. Kumar, *Bioresour. Technol.* **2021**, *339*, 125589; c) H. Mali, C. Shah, B. H. Raghunandan, A. S. Prajapati, D. H. Patel, U. Trivedi, R. B. Subramanian, *J. Environ. Sci.* **2023**, *127*, 234–250; d) A. Inobeme, C. O. Adetunji, M. J. Tsado, A. I. Ajai, J. Inobeme, B. Oyedolapo, in *Modern Approaches in Waste Bioremediation* (Ed.: M. P. Shah), Springer, Cham, **2023**, pp. 101–115.
[6] E. Lerner, T. Cordes, A. Ingargiola, Y. Alhadid, S. Chung, X. Michalet, S. Weiss, *Science* **2018**, *359*, 288.
[7] L. Konermann, J. Pan, Y. H. Liu, *Chem. Soc. Rev.* **2011**, *40*, 1224–1234.
[8] B. Vestergaard, *Arch. Biochem. Biophys.* **2016**, *602*, 69–79.
[9] a) 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, *Nat. Commun.* **2021**, *12*, 3616; b) L. Uhrik, T. Henek, J. Planas-Iglesias, J. Kucera, J. Damborsky, M. Marek, L. Hernychova, in *Advances Methods in Structural Biology*, Springer US, New York, **2023**, pp. 293–318.
[10] a) A. Nath, M. Sammalkorpi, D. C. DeWitt, A. J. Trexler, S. Elbaum-Garfinkle, C. S. O'Hern, E. Rhoades, *Biophys. J.* **2012**, *103*, 1940–1949; b) B. Rozycki, E. Boura, *J. Phys. Condens. Matter* **2014**, *26*, 463103.
[11] T. Hameduh, Y. Haddad, V. Adam, Z. Heger, *Comput. Struct. Biotechnol. J.* **2020**, *18*, 3494–3506.
[12] 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, *Proteins* **2019**, *87*, 1141–1148.
[13] M. Baek, F. DiMaio, I. Anishchenko, J. Dauparas, S. Ovchinnikov, G. R. Lee, J. Wang, Q. Cong, L. N. Kinch, R. D. Schaeffer, C. Millán, H. Park, C. Adams, C. R. Glassman, A. DeGiovanni, J. H. Pereira, A. V. Rodrigues, A. A. van Dijk, A. C. Ebrecht, D. J. Opperman, T. Sagmeister, C. Buhlheller, T. Pavkov-Keller, M. K. Rathinaswamy, U. Dalwadi, C. K. Yip, J. E. Burke, K. C. Garcia, N. V. Grishin, P. D. Adams, R. J. Read, D. Baker, *Science* **2021**, *373*, 871–876.
[14] a) J. Durairaj, D. de Ridder, A. D. J. van Dijk, *Comput. Struct. Biotechnol. J.* **2023**, *21*, 630–643; b) A. David, S. Islam, E. Tankhilevich, M. J. E. Sternberg, *J. Mol. Biol.* **2022**, *434*, 167336.
[15] T. Wang, X. Zhang, O. Zhang, P. Pan, G. Chen, Y. Kang, C.-Y. Hsieh, T. Hou, *arXiv:2306.12754* **2023**.
[16] T. C. Terwilliger, D. Liebschner, T. I. Croll, C. J. Williams, A. J. McCoy, B. K. Poon, P. V. Afonine, R. D. Oeffner, J. S. Richardson, R. J. Read, P. D. Adams, *bioRxiv (preprint)* **2022**, DOI: 10.1101/2022.11.21.517405 .
[17] A. M. Loening, T. D. Fenn, A. M. Wu, S. S. Gambhir, *Protein Eng. Des. Sel.* **2006**, *19*, 391–400.
[18] A. Schenkmayerova, M. Toul, D. Pluskal, R. Baatallah, G. Gagnot, G. P. Pinto, V. T. Santana, M. Stuchla, P. Neugebauer, P. Chaiyen, J. Damborsky, D. Bednar, Y. L. Janin, Z. Prokop, M. Marek, *Nat. Catal.* **2023**, *6*, 23–38.
[19] Y. Tomabechi, T. Hosoya, H. Ehara, S. I. Sekine, M. Shirouzu, S. Inouye, *Biochem. Biophys. Res. Commun.* **2016**, *470*, 88–93.
[20] a) J. J. Irwin, B. K. Shoichet, *J. Med. Chem.* **2016**, *59*, 4103–4120; b) K. Liu, E. Watanabe, H. Kokubo, *J. Comput.-Aided Mol. Des.* **2017**, *31*, 201–211; c) J. Mortier, C. Rakers, M. Bermudez, M. S. Murgueitio, S. Riniker, G. Wolber, *Drug Discovery Today* **2015**, *20*, 686–702.
[21] a) C. E. Dempsey, D. Wright, C. K. Colenso, R. B. Sessions, J. C. Hancox, *J. Chem. Inf. Model.* **2014**, *54*, 601–612; b) J. Nedergaard Pedersen, P. W. J. M. Frederix, J. Skov Pedersen, S. J. Marrink, D. E. Otzen, *ChemBioChem* **2018**, *19*, 263–271; c) Z. F. Brotzakis, M. Vendruscolo, P. G. Bolhuis, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, e2012423118.
[22] L. Maveyraud, L. Mourey, *Molecules* **2020**, *25*, 1030.
[23] I. P. Gerothanassis, *Arab. J. Chem.* **2023**, *16*, 104536.
[24] F. Merino, S. Raunser, *Angew. Chem. Int. Ed.* **2017**, *56*, 2846–2860.
[25] wwPDB Consortium, *Nucleic Acids Res.* **2019**, *47*, D520–D528.
[26] a) J. A. Geraets, K. R. Pothula, G. F. Schröder, *Curr. Opin. Struct. Biol.* **2020**, *61*, 173–181; b) L. Cerofolini, M. Fragai, E. Ravera, C. A. Diebolder, L. Renault, V. Calderone, *Biomol. Eng.* **2019**, *9*, 370.
[27] a) M. J. Belousoff, H. Venugopal, A. Wright, S. Seoner, I. Stuart, C. Stubenrauch, R. S. Bamert, D. W. Lupton, T. Lithgow, *ChemMedChem* **2019**, *14*, 527–531; b) E. P. Morris, P. C. A. da Fonseca, *Acta Crystallogr. Sect. D* **2017**, *73*, 522–533.

[28] S. Bhattacharya, E. G. Margheritis, K. Takahashi, A. Kulesha, A. D'Souza, I. Kim, J. H. Yoon, J. R. H. Tame, A. N. Volkov, O. V. Makhlynets, I. V. Korendovych, *Nature* **2022**, *610*, 389–393.
[29] A. A. Kermani, *FEBS J.* **2021**, *288*, 5788–5804.
[30] a) T. Tosha, T. Nomura, T. Nishida, N. Saeki, K. Okubayashi, R. Yamagiwa, M. Sugahara, T. Nakane, K. Yamashita, K. Hirata, G. Ueno, T. Kimura, T. Hisano, K. Muramoto, H. Sawai, H. Takeda, E. Mizohata, A. Yamashita, Y. Kanematsu, Y. Takano, E. Nango, R. Tanaka, O. Nureki, O. Shoji, Y. Ikemoto, H. Murakami, S. Owada, K. Tono, M. Yabashi, M. Yamamoto, H. Ago, S. Iwata, H. Sugimoto, Y. Shiro, M. Kubo, *Nat. Commun.* **2017**, *8*, 1585; b) A. Butryn, P. S. Simon, P. Aller, P. Hinchliffe, R. N. Massad, G. Leen, C. L. Tooke, I. Bogacz, I. S. Kim, A. Bhowmick, A. S. Brewster, N. E. Devenish, J. Brem, J. Kamps, P. A. Lang, P. Rabe, D. Axford, J. H. Beale, B. Davy, A. Ebrahim, J. Orlans, S. L. S. Storm, T. Zhou, S. Owada, R. Tanaka, K. Tono, G. Evans, R. L. Owen, F. A. Houle, N. K. Sauter, C. J. Schofield, J. Spencer, V. K. Yachandra, J. Yano, J. F. Kern, A. M. Orville, *Nat. Commun.* **2021**, *12*, 4461; c) J. A. R. Worrall, M. A. Hough, *Curr. Opin. Struct. Biol.* **2022**, *77*, 102486.
[31] a) K. Schriever, P. Saenz-Mendez, R. S. Rudraraju, N. M. Hendrikse, E. P. Hudson, A. Biundo, R. Schnell, P. O. Syrén, *J. Am. Chem. Soc.* **2021**, *143*, 3794–3807; b) C. R. Nicoll, G. Bailleul, F. Fiorentini, M. L. Mascotti, M. W. Fraaije, A. Mattevi, *Nat. Struct. Mol. Biol.* **2020**, *27*, 14–24; c) R. E. S. Thomson, S. E. Carrera-Pacheco, E. M. J. Gillam, *J. Biol. Chem.* **2022**, *298*, 102435.
[32] a) S. Toma, K. T. Chong, A. Nakagawa, K. Teranishi, S. Inouye, O. Shimomura, *Protein Sci.* **2005**, *14*, 409–416; b) S. Inouye, Y. Tomabechi, T. Hosoya, S. I. Sekine, M. Shirouzu, *J. Biochem.* **2018**, *164*, 247–255.
[33] a) L. Quan, Q. Lv, Y. Zhang, *Bioinformatics* **2016**, *32*, 2936–2946; b) M. G. Lepre, S. I. Omar, G. Grasso, U. Morbiducci, M. A. Deriu, J. A. Tuszynski, *Molecules* **2017**, *22*, 1358; c) A. Kumar, P. Biswas, *Proteins* **2019**, *87*, 760–773.
[34] S. H. Haddock, M. A. Moline, J. F. Case, *Annu. Rev. Mar. Sci.* **2010**, *2*, 443–493.
[35] S. V. Markova, E. S. Vysotski, *Biochemistry* **2015**, *80*, 714–732.
[36] P. Loi, J. Saragusty, G. Ptak, in *Reproductive Sciences in Animal Conservation, Vol. 753* (Eds.: W. V. Holt, J. L. Brown, P. Comizzoli), Springer, New York, NY, **2014**, pp. 489–502.
[37] a) G. K. A. Hochberg, J. W. Thornton, *Annu. Rev. Biophys.* **2017**, *46*, 247–269; b) J. M. Gardner, M. Biler, V. A. Risso, J. M. Sanchez-Ruiz, S. C. L. Kamerlin, *ACS Catal.* **2020**, *10*, 4863–4870.
[38] a) Z. Yang, *Mol. Biol. Evol.* **2007**, *24*, 1586–1591; b) A. Stamatakis, *Bioinformatics* **2014**, *30*, 1312–1313.
[39] F. Ronquist, M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M. A. Suchard, J. P. Huelsenbeck, *Syst. Biol.* **2012**, *61*, 539–542.
[40] M. Musil, R. T. Khan, A. Beier, J. Stourac, H. Konegger, J. Damborsky, D. Bednar, *Briefings Bioinf.* **2021**, *22*, 1–11.
[41] 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, *ACS Catal.* **2019**, *9*, 4810–4823.
[42] a) V. A. Risso, J. A. Gavira, D. F. Mejia-Carmona, E. A. Gaucher, J. M. Sanchez-Ruiz, *J. Am. Chem. Soc.* **2013**, *135*, 2899–2902; b) P. Babkova, E. Sebestova, J. Brezovsky, R. Chaloupkova, J. Damborsky, *ChemBioChem* **2017**, *18*, 1448–1456; c) T. Zou, V. A. Risso, J. A. Gavira, J. M. Sanchez-Ruiz, S. B. Ozkan, *Mol. Biol. Evol.* **2015**, *32*, 132–143; d) S. M. Muñoz, V. Castro-Fernandez, V. Guixé, *J. Biol. Chem.* **2021**, *296*, 100219; e) P. Perez-Garcia, S. Kobus, C. G. W. Gertzen, A. Hoeppner, N. Holzscheck, C. H. Strunk, H. Huber, K. E. Jaeger, H. Gohlke, F. Kovacic, S. H. J. Smits, W. R. Streit, J. Chow, *Commun. Biol.* **2021**, *4*, 132.
[43] M. T. Lin, H. Salihovic, F. K. Clark, M. R. Hanson, *Sci. Adv.* **2022**, *8*, eabm6871.
[44] a) R. Perez-Jimenez, A. Inglés-Prieto, Z. M. Zhao, I. Sanchez-Romero, J. Alegre-Cebollada, P. Kosuri, S. Garcia-Manyes, T. J. Kappock, M. Tanokura, A. Holmgren, J. M. Sanchez-Ruiz, E. A. Gaucher, J. M. Fernandez, *Nat. Struct. Mol. Biol.* **2011**, *18*, 592–597; b) J. K. Hobbs, C. Shepherd, D. J. Saul, N. J. Demetras, S. Haaning, C. R. Monk, R. M. Daniel, V. L. Arcus, *Mol. Biol. Evol.* **2012**, *29*, 825–835; c) D. L. Trudeau, M. Kaltenbach, D. S. Tawfik, *Mol. Biol. Evol.* **2016**, *33*, 2633–2641; d) M. A. Spence, J. A. Kaczmarski, J. W. Saunders, C. J. Jackson, *Curr. Opin. Struct. Biol.* **2021**, *69*, 131–141; e) K. Watanabe, A. Yamagishi, *FEBS Lett.* **2006**, *580*, 3867–3871; f) V. A. Risso, F. Manssour-Triedo, A. Delgado-Delgado, R. Arco, A. Barroso-delJesus, A. Ingles-Prieto, R. Godoy-Ruiz, J. A. Gavira, E. A. Gaucher, B. Ibarra-Molero, J. M. Sanchez-Ruiz, *Mol. Biol. Evol.* **2014**, *32*, 440–455; g) C. H. Chiang, T. Wymore, A. Rodriguez Benitez, A. Hussain, J. L. Smith, C. L. Brooks, 3rd, A. R. H. Narayan, *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2218248120.
[45] F. Gonzalez-Ordenes, P. A. Cea, N. Fuentes-Ugarte, S. M. Muñoz, R. A. Zamora, D. Leonardo, R. C. Garratt, V. Castro-Fernandez, V. Guixé, *Front. Microbiol.* **2018**, *9*, 1305.
[46] C. R. Nicoll, M. Massari, M. W. Fraaije, M. L. Mascotti, A. Mattevi, *Curr. Opin. Struct. Biol.* **2023**, *82*, 102669.
[47] P. Babkova, Z. Dunajova, R. Chaloupkova, J. Damborsky, D. Bednar, M. Marek, *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1497–1508.
[48] a) C. J. Brown, S. Takayama, A. M. Campen, P. Vise, T. W. Marshall, C. J. Oldfield, C. J. Williams, A. K. Dunker, *J. Mol. Evol.* **2002**, *55*, 104–110; b) C. J. Brown, A. K. Johnson, G. W. Daughdrill, *Mol. Biol. Evol.* **2010**, *27*, 609–621.
[49] J. J. Ward, J. S. Sodhi, L. J. McGuffin, B. F. Buxton, D. T. Jones, *J. Mol. Biol.* **2004**, *337*, 635–645.
[50] a) M. Biesaga, M. Frigole-Vivas, X. Salvatella, *Curr. Opin. Chem. Biol.* **2021**, *62*, 90–100; b) P. Santofimia-Castaño, B. Rizzuti, Y. Xia, O. Abian, L. Peng, A. Velázquez-Campoy, J. L. Neira, J. Iovanna, *Cell. Mol. Life Sci.* **2020**, *77*, 1695–1707; c) S. E. Bondos, A. K. Dunker, V. N. Uversky, *Cell Commun. Signaling* **2022**, *20*, 20; d) P. Chakrabarti, D. Chakravarty, *Biophys. Chem.* **2022**, *283*, 106769.
[51] a) H. J. Dyson, P. E. Wright, *J. Biomol. NMR* **2019**, *73*, 651–659; b) H. J. Dyson, P. E. Wright, *Curr. Opin. Struct. Biol.* **2021**, *70*, 44–52; c) R. Schneider, M. Blackledge, M. R. Jensen, *Curr. Opin. Struct. Biol.* **2019**, *54*, 10–18.
[52] V. Castro-Fernandez, A. Herrera-Morande, R. Zamora, F. Merino, F. Gonzalez-Ordenes, F. Padilla-Salinas, H. M. Pereira, J. Brandao-Neto, R. C. Garratt, V. Guixe, *J. Biol. Chem.* **2017**, *292*, 15598–15610.
[53] Y. Oba, K. Konishi, D. Yano, H. Shibata, D. Kato, T. Shirai, *Sci. Adv.* **2020**, *6*, eabc5705.
[54] A. G. Bart, K. L. Harris, E. M. J. Gillam, E. E. Scott, *J. Biol. Chem.* **2020**, *295*, 5640–5653.
[55] K. M. Hart, M. J. Harms, B. H. Schmidt, C. Elya, J. W. Thornton, S. Marqusee, *PLoS Biol.* **2014**, *12*, e1001994.
[56] J. I. Boucher, J. R. Jacobowitz, B. C. Beckett, S. Classen, D. L. Theobald, *eLife* **2014**, *3*, e02304.
[57] V. Nguyen, C. Wilson, M. Hoemberger, J. B. Stiller, R. V. Agafonov, S. Kutter, J. English, D. L. Theobald, D. Kern, *Science* **2017**, *355*, 289–294.
[58] a) A. J. Syed, J. C. Anderson, *Chem. Soc. Rev.* **2021**, *50*, 5668–5705; b) H.-M. Ke, I. J. Tsai, *Curr. Opin. Green Sustain. Chem.* **2022**, *33*, 100570.
[59] T. Mitiouchkina, A. S. Mishin, L. G. Somermeyer, N. M. Markina, T. V. Chepurnyh, E. B. Guglya, T. A. Karataeva, K. A. Palkina, E. S. Shakhova, L. I. Fakhranurova, S. V. Chekova, A. S. Tsarkova, Y. V. Golubev, V. V. Negrebetsky, S. A. Dolgushin, P. V. Shalaev, D. Shlykov, O. A. Melnik, V. O. Shipunova, S. M. Deyev, A. I. Bubyrev, A. S. Pushin, V. V. Choob, S. V. Dolgov, F. A. Kondrashov, I. V. Yampolsky, K. S. Sarkisyan, *Nat. Biotechnol.* **2020**, *38*, 944–946.
[60] J. F. Head, S. Inouye, K. Teranishi, O. Shimomura, *Nature* **2000**, *405*, 372–376.
[61] O. Shimomura, *Bioluminescence: Chemical Principles and Methods*, World Scientific, Singapore, **2012**.
[62] M. S. Titushin, Y. Feng, G. A. Stepanyuk, Y. Li, S. V. Markova, S. Golz, B. C. Wang, J. Lee, J. Wang, E. S. Vysotski, Z. J. Liu, *J. Biol. Chem.* **2010**, *285*, 40891–40900.
[63] Z. J. Liu, E. S. Vysotski, C. J. Chen, J. P. Rose, J. Lee, B. C. Wang, *Protein Sci.* **2000**, *9*, 2085–2093.
[64] E. S. Vysotski, Z.-J. Liu, S. V. Markova, J. R. Blinks, L. Deng, L. A. Frank, M. Herko, N. P. Malikova, J. P. Rose, B.-C. Wang, J. Lee, *Biochemistry* **2003**, *42*, 6013–6024.
[65] Z. J. Liu, E. S. Vysotski, L. Deng, J. Lee, J. Rose, B.-C. Wang, *Biochem. Biophys. Res. Commun.* **2003**, *311*, 433–439.
[66] P. V. Natashin, S. V. Markova, J. Lee, E. S. Vysotski, Z.-J. Liu, *FEBS J.* **2014**, *281*, 1432–1445.
[67] M. D. Larionova, L. Wu, E. V. Eremeeva, P. V. Natashin, D. V. Gulnov, E. V. Nemtseva, D. Liu, Z. J. Liu, E. S. Vysotski, *Protein Sci.* **2022**, *31*, 454–469.
[68] A. M. Loening, T. D. Fenn, S. S. Gambhir, *J. Mol. Biol.* **2007**, *374*, 1017–1028.
[69] S. Inouye, J. I. Sato, Y. Sahara-Miura, Y. Tomabechi, Y. Sumida, S. I. Sekine, M. Shirouzu, T. Hosoya, *PLoS One* **2022**, *17*, e0272992.
[70] N. Wu, N. Kobayashi, K. Tsuda, S. Unzai, T. Saotome, Y. Kuroda, T. Yamazaki, *Sci. Rep.* **2020**, *10*, 20069.
[71] G. Gotthard, J. Walker, M. Saidi, J. Hiblot, A. Royant, in *20th International Symposium on Bioluminescence and Chemiluminescence*, Nantes, France, **2018**.

[72] S. Duellman, M. B. Robers, J. R. Walker, W. Zhou, C. Zimprich, *Vol. WO 2016210294*, **2016**.
[73] J. R. Walker, M. P. Hall, C. A. Zimprich, M. B. Robers, S. J. Duellman, T. Machleidt, J. Rodriguez, W. Zhou, *ACS Chem. Biol.* **2017**, *12*, 1028–1037.
[74] a) P. I. Ho, K. Yue, P. Pandey, L. Breault, F. Harbinski, A. J. McBride, B. Webb, J. Narahari, N. Karassina, K. V. Wood, A. Hill, D. S. Auld, *ACS Chem. Biol.* **2013**, *8*, 1009–1017; b) T. N. Cartwright, S. K. Meyer, J. M. G. Higgins, *SLAS Discov.* **2022**, *27*, 471–475.
[75] J. R. Walker, M. Hall, C. T. Eggers, H. Lazaro, *Vol. WO 2019232284*, **2019**.
[76] J. C. Matthews, K. Hori, M. J. Cornier, *Biochemistry* **1977**, *16*, 5217–5220.
[77] a) C. Wu, H. Nakamura, A. Murai, S. Inouye, *Tetrahedron* **2001**, *57*, 9575–9583; b) H. Nakamura, C. Wu, S. Inouye, A. Murai, *J. Am. Chem. Soc.* **2001**, *123*, 1523–1524; c) S. Inouye, Y. Sumida, Y. Tomabechi, J. Taguchi, M. Shirouzu, T. Hosoya, *PLoS One* **2021**, *16*, e0251743.
[78] K. Takatsu, N. Kobayashi, N. Wu, Y. L. Janin, T. Yamazaki, Y. Kuroda, *BBA Adv.* **2023**, *3*, 100068.
[79] M. Nemergut, D. Pluskal, J. Horackova, T. Sustrova, J. Tulis, T. Barta, R. Baatallah, G. Gagnot, V. Novakova, M. Majerova, S. M. Marques, M. Toul, J. Damborsky, D. Bednar, Z. Prokop, Y. L. Janin, M. Marek, *bioRxiv (preprint)* **2022**, DOI: 10.1101/2022.12.05.519101.
[80] W. W. Lorenz, R. McCann, M. Longiaru, M. J. Cormier, *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 4438–4442.
[81] J. C. Matthews, K. Hori, M. J. Cormier, *Biochemistry* **1977**, *16*, 85–91.
[82] S. Rahnama, B. Saffar, Z. F. Kahrani, M. Nazari, R. Emamzadeh, *Enzyme Microb. Technol.* **2017**, *96*, 60–66.
[83] V. A. Risso, S. Martinez-Rodriguez, A. M. Candel, D. M. Krüger, D. Pantoja-Uceda, M. Ortega-Muñoz, F. Santoyo-Gonzalez, E. A. Gaucher, S. C. L. Kamerlin, M. Bruix, J. A. Gavira, J. M. Sanchez-Ruiz, *Nat. Commun.* **2017**, *8*, 16113.

Manuscript received: June 10, 2023
Revised manuscript received: August 10, 2023
Accepted manuscript online: August 21, 2023
Version of record online: September 8, 2023