이학박사 학위논문

Development of a general method for designing artificial metalloenzymes utilizing diverse protein scaffolds

다양한 단백질 골격을 활용한 인공 가수 분해 효소의 설계를 위한 방법 개발

2023년 8월

서울대학교 대학원

화학부 생화학 전공

정 우 재

Development of a general method for designing artificial metalloenzymes utilizing diverse protein scaffolds

지도 교수 송 윤 주

이 논문을 이학박사 학위논문으로 제출함

2023 년 8 월

서울대학교 대학원

화학부 생화학 전공

정 우 재

정우재의 이학박사 학위논문을 인준함

2023 년 6 월

위 원 장 김 석 희 (인)

부위원장 송 윤 주 (인)

위 원 이 연 (인)

위 원 이 형 호 (인)

위 원 홍 승 우 (인)

Abstract

Development of a general method for designing artificial metalloenzymes utilizing diverse protein scaffolds

Woo Jae Jeong
Department of Chemistry
The Graduate School
Seoul National University

Natural metalloenzymes exhibit essential chemical interaction between metal ions and amino acids. The molecular design of artificial metalloenzymes has been one of the routes to explore such interactions. However, previous studies of tailor-made artificial metalloenzymes have primarily utilized a limited number of protein scaffolds. Thus, this thesis focuses on developing a more general method for designing artificial metalloenzymes using diverse protein scaffolds, including non-α-helical proteins, such as outer membrane protein F (OmpF).

In **Chapter 1**, I describe recent advancements in artificial metalloenzymes and frequently adapted protein scaffolds. Although successful conversions of artificial metalloenzymes have been reported, these achievements have been limited to a narrow range of protein scaffolds. This restricted selection of protein scaffold is due to inherent limitations associated with design strategy, particularly the utilizing the pre-existing metal centers.

To overcome this limitation, in **Chapter 2**, I focus on developing a novel design method by extracting and analyzing the geometric parameters, such as the distances between zinc-ligating residues in native zinc metalloenzymes.

Then, I successfully applied these parameters to non-α helical protein structures, such as OmpF, creating several zinc-binding metalloproteins. By utilizing this novel approach, we aim to broaden the scope of protein scaffolds used to design artificial metalloenzymes. This expansion will enable the exploration of diverse protein architectures and enhance my ability to engineer enzymes with desired catalytic properties. Ultimately, my research seeks to contribute to developing more efficient and versatile artificial metalloenzymes that can be employed in a wide range of applications.

In **Chapter 3**, I describe the design of artificial metalloenzymes based on geometric parameters. I applied a tetrahedral Zn-binding motif to a non-α-helical protein scaffold, specifically the β-barrel outer membrane protein F (OmpF). Through this approach, I created three variants of artificial Zn-binding proteins and determined their crystal structures. These variants exhibited distinct first coordination spheres, leading to different catalytic properties. Notably, these enzymes demonstrated hydrolytic activities with various substrates, particularly β-glycosides. Furthermore, I successfully obtained evolved novel Zn-dependent artificial metalloenzymes using directed evolution. This research opens avenues for utilizing diverse protein scaffolds in artificial metalloenzyme design and expanding the chemical capabilities of inorganic cofactors in the biochemical space.

Keywords: Artificial metalloenzymes, Outer membrane protein F, Unsaturated tetrahedral Zn-binding site, Metallo-hydrolases, Glycosidases, Directed-evolution.

Student Number: 2016-25496

Contents

Figure Contents

Table Contents

CHAPTER 1

Introduction

Part of this chapter was adapted from W.J. Jeong*, J. Yu*, and W.J. Song. (2020) *Chem. Commun.* 56, 9586-9599.

1.1 Introduction

The discovery of numerous metalloenzymes has inspired us to synthesize inorganic compounds that resemble the structure or function of the active sites.[1-2] The design of biomimetic inorganic complexes often starts from the identification and construction of the active sites of natural enzymes, allowing for simplification of essential inorganic environments and evaluation of reaction mechanisms in catalysis. Unlike proteins, which have multiple amino acid residues comprising the active sites, however, synthesis of multiple functional groups around the catalytic sites is often a great synthetic challenge. In this regard, reengineering proteins into artificial metalloenzymes might be viewed as an efficient route to understand remarkably complex and tightly orchestrated actions of natural enzymes and to synthesize diverse metal-dependent biocatalysts.[3-5]

To design catalytically active units with only 20 canonical amino acids, well-defined hydrogen bonds, electrostatic, and/or hydrophobic interactions should be properly incorporated to accommodate a kinetically and thermodynamically tuned series of reactions during catalysis, which is not trivial even with our current state-of-the-art in the understanding of protein sequence structure-function relationship. Nevertheless, the difficulties in designing artificial metalloenzymes, remarkable progress has been made in the design of artificial metalloenzymes that harbor inorganic cofactors or metal ions in the protein matrix.[5-7]

I categorized the recent examples of artificial metalloenzymes based on their design strategies, resulting in three distinct categories. Through classification, I have observed that the design of artificial metalloenzymes has been limited to a narrow range of protein scaffolds. This restricted selection of proteins is due to inherent limitations associated with each design strategy.

The first strategy involves substituting metals at existing metal centers, which confines the design to proteins that already possess such centers.[8-9] The second strategy relies on proteins capable of specific binding through dative or covalent bonds. Consequently,[8-11] a small subset of proteins containing a few cysteine residues, is required or exhibits a high binding affinity with small molecules, which highly rely on streptavidin-biotin interaction.[12-15] The third strategy, which relies on computational design, predominantly focuses on proteins with α-helical structures. These secondary structures are beneficial for placing two amino acids within proximal distance and directionality, providing an easily designable bidentate ligand for Zn-coordination.[16-20] However, the computational design of artificial metalloenzymes using other protein scaffolds, such as β-sheets or β-barrels, has proven to be more challenging.[21] Consequently, the reliance on α-helical structures in computational design creates a more favorable but limited pool of protein scaffolds. These findings emphasize the importance of overcoming the limitations imposed by current design strategies for artificial metalloenzymes.

1.1.1 Pre-existing metal centers

One of the primary and long-standing methods to design novel metalloenzymes is to substitute inorganic elements, such as metal ions or metallo-cofactors, and/or to reshape the active site pockets. Then, natural metalloproteins that harbor pre-existing metal-active sites are the obvious starting materials for artificial metalloenzymes.

This efficient approach is applicable when apo-proteins are somewhat stable even after pre-existing metal ion/cofactors are chelated prior to reconstitution with novel inorganic elements. At least moderate binding affinity for metal ions and cofactors other than the native ones is also necessary for the preparation of artificial metalloproteins. Such metal promiscuity might sound counter-intuitive to the general understanding of natural enzymes as being highly selective for the specific metal ion to perform dedicated biochemical functions. It is, however, more prevalent than expected,[22-25] and protein promiscuity, in fact, can lead to the appearance and divergence of novel enzymes.[26-28]

To reconstitute metalloproteins, metal elements that favor similar inorganic environments to those of the native ones would be preferred, such as coordination number, geometry, and ligand properties (hard versus soft). The Metal-PDB database[29] can be a valuable resource to search prevalent coordination geometry and ligands per each metal ion within proteinaceous environments. It is, however, worthy to note that proteins are flexible ligands, which may undergo multiple conformational changes and/or adapt alternative rotameric states depending on the neighboring residues. Therefore, fine tailoring of proteins to construct novel metal-binding sites are still in great challenge.

Recent examples of replacing the native metal ions with novel ones have

been carried out with structurally and spectroscopically well-characterized metalloenzymes (Figure 1.1). A Cu in azurin (Figure 1.1a)[8] and Fe in rubredoxin (Figure 1.1b)[9] were replaced with Fe and Ni ion, yielding superoxide scavenging and hydrogenase activities, respectively. Notably, the Cu center in azurin is a distorted trigonal bipyramidal geometry, which can be accommodated by Fe. Indeed, the Metal-PDB database indicates that Fe is the most prevalent metal element that replaces Cu-binding sites, followed by Zn ion. Similarly, the Ni-substituted rubredoxin was prepared presumably because such coordination geometry and the 4 cysteine ligands can be generated with Ni, as observed in [NiFe] hydrogenase.

In addition, the structural resemblance between the native active sites and inorganic complexes can be adapted for repurposing metalloproteins. For example, a Mn-binding protein, TM1459 (Figure 1.1c), was successfully reconstituted with an Os ion in its native metal center.[30] This work was initiated owing to the geometric resemblance of the active site to that of a previously characterized Os complex, Os-TPA (TPA = tris(2-pyridyl-methyl)amine). The chemo-mimetic enzyme exhibits peroxygenase activities with alkenes, resembling the catalytic activities of synthetic Os complexes.

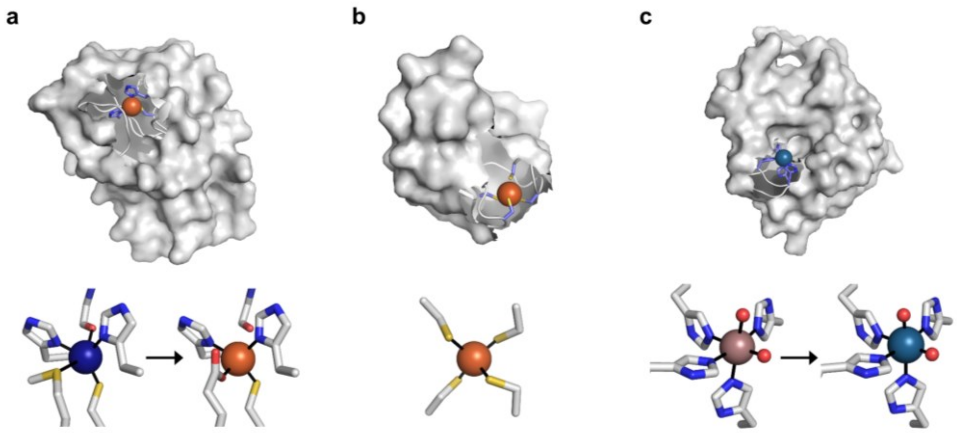


Figure 1.1 Transformation of pre-existing metal-binding sites into the active sites of artificial metalloenzymes

a, Fe-substituted azurin (4QLW) from Cu-azurin (4AZU). **b,** The wild-type rubredoxin with a Fe-center (6RXN) **c,** TM1459 substituted with an osmium complex (5WSF) from the native form with Mn-bound form (1VJ2). In **a** and **c,** the native metal coordination sites before the metal-substitution are depicted bottom left. PDB codes are given in parentheses. Metal ions and metal-ligating amino acids are depicted with spheres and sticks, respectively. Cu, Fe, Mn, Os, and Zn atoms are colored in navy, orange, purple, light blue, and light purple, respectively. Carbon, nitrogen, and oxygen atoms in metal-binding residues are colored in grey, blue, and red, respectively.

1.1.2 Native binding affinity

Selective introduction of inorganic elements becomes the most critical step when a non-metalloprotein is employed as a starting material for the synthesis of artificial metalloenzymes. One successful approach is to harness biological host proteins with the native binding affinity to guest-like small molecules (Figure 1.2). Ward *et al.* have employed streptavidin-biotin technology to incorporate inorganic catalysts at the desired position. The tight binding affinity of biotin to streptavidin was retained even upon structural modification of biotin and mutations in streptavidin, yielding a wide range of catalytic activities, such as transfer hydrogenation,[12] Suzuki–Miyaura cross-coupling,[13] and metathesis (Figure 1.2a).[14]

The catalytic reactivities are derived from inorganic complexes such as Ir d^6-piano stool complex (Figure 1.3a), monophosphine Pd complex, and Hoveyda–Grubbs catalyst, respectively, which are selectively introduced by the covalent linkage to biotin, whereas the surrounding protein environment contributes to the determination of product distribution and turnover number. Notably, a monomeric version of streptavidin, originally a homotetramer, was created recently,[15] diversifying the quaternary structure of the versatile protein scaffold for artificial metalloenzymes.

The streptavidin-biotin strategy enabled exploration of novel inorganic reactivity in proteinaceous environments. Borovik et al. recently introduced a mononuclear copper complex, Cu(biot-et-dpea), (Figure 1.3b) into streptavidin.[31] The chemical properties of the copper complex were previously investigated in non-aqueous solvents, and they are remarkably altered upon docking on the protein. In particular, hydrogen-bonding networks with adjacent residues significantly increase the stability of Cu–hydroperoxido species, indicating that proteins are evolvable platforms for

metal-dependent biocatalysis.

Metal-trafficking proteins have also been exploited along these lines. These proteins exhibit high selectivity and binding affinity for dedicated metal ions or metal complexes, which can be readily utilized to introduce abiological inorganic elements into proteins. Siderophore is an innate chelator and transporter of Fe, exhibiting high selectivity for the ferric state over ferrous ion. As a result, Fe-coordination of azotochelin, a siderophore from Azotobacter vinelandii (Figure 1.3c), was linked to a catalyst of interest, such as Ir imine reduction catalyst (Figure 1.3a), resulting in selective association and dissociation upon oxidation and reduction of Fe, respectively.[32] Additionally, when an Ir catalyst was attached to siderophore, periplasmic binding proteins (PBPs) were transformed into artificial hydrogenase (Figure 1.2b).

Analogously, a nickel transporter NikA was also utilized as a scaffold to dock a Fe complex such as Fe(EDTA)(H2O)– for catalysis (Figure 1.2c).[33] The NikA-Fe hybrid protein functions as an artificial oxygenase for sulfide oxidation.[34] Again, similarly to streptavidin, these examples indicate that catalytic activities are derived from metal complexes when the secondary coordination environments, and consequently, the enzymatic activities, are tuned by the protein matrix.

Proteins with inhibitors and strongly-binding ligands have also been utilized as biological hosts. Aryl-sulphonamide, for example, is an inhibitor of carbonic anhydrase (Figure 1.2d).[35] Human serum albumin is a carrier protein, exhibiting well-characterized interactions with coumarin-derivatives (Figure 1.2e).[36] After covalent linking of an IrCp* piano-stool complex or Hoveyda–Grubbs Ru catalyst to an inhibitor and a hydrophobic coumarin ligand, respectively, they became metalloenzymes with asymmetric transfer

hydrogenation and ring-closing metathesis activities, respectively.

For serine or cysteine-dependent hydrolases, the catalytically active serine residues can function as nucleophiles to be covalently linked to abiological inorganic elements so that catalytic units can be steered to the native active site pockets. Various organometallic catalysts, such as Hoveyda-Grubbs catalyst, Ir d^6-piano stool complex, Rh(NHC) phosphonate (Figure 1.3d), and pincer-type Pt d^8 complex have been attached to hydrolases such as α-chymotrypsin,[37] papain,[38] lipase,[39] and cutinase (Figure 1.2f),[40] respectively. Similarly, a serine hydrolase with two catalytic triads was converted into an artificial enzyme for oxidation and Friedel-Crafts alkylation.[41] One of its catalytic triads was conjugated with a Cu(bpy) complex, resulting in an artificial metalloenzyme with coupled catalytic activities.

In some cases, binding affinity to dock external inorganic components to novel catalytic environments can be derived from the native protein-protein interaction. Douglas et al. have created artificial metalloenzymes by encapsulating [NiFe]-hydrogenase within the capsid of bacteriophage P22.[42] The catalytic domain, [NiFe]-hydrogenase, was fused to a protein that exhibits protein-protein interaction with a coat protein. The self-assembled artificial metalloenzyme exhibits enhanced catalytic activity and stability against protease, thermal denaturation, and air exposure, indicating that a combination of two proteins can also give rise to artificial metalloenzymes.

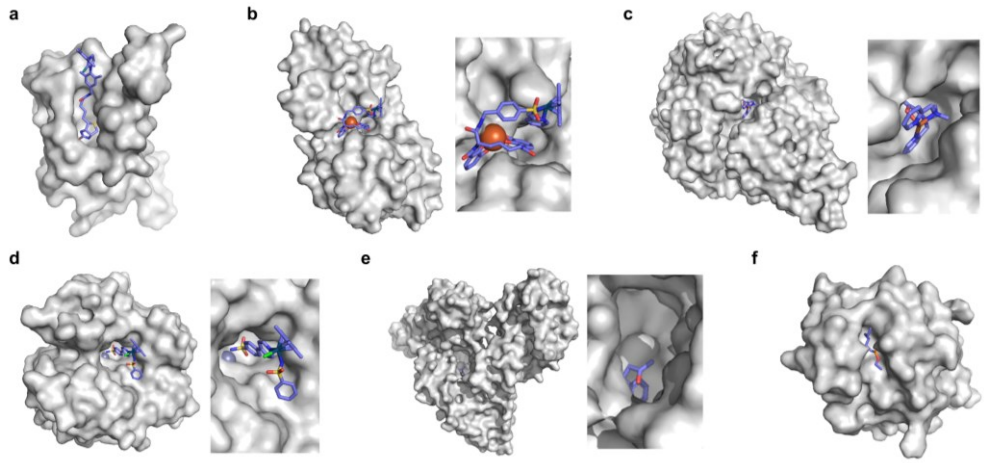


Figure 1.2 Representative proteins with the native binding properties

a, Biotinylated Hoveyda–Grubbs catalyst bound in streptavidin (PDB 5IRA) **b,** Periplasmic binding protein CeuE complexed with siderophore–Ir catalyst (5OD5) **c,** An Fe complex bound inside nickel binding protein, NikA (4I9D) **d,** α-carbonic anhydrase complexed with an inhibitor-Ir catalyst (PDB 3ZP9) **e,** A coumarin-derived molecule bound in human serum albumin (1H9Z) **f,** The wild-type cutinase-inhibitor complex (1XZL). The active site pockets are enlarged for clarity. The small molecules that bind to the active site of the proteins are depicted with blue sticks. PDB codes are given in parentheses.

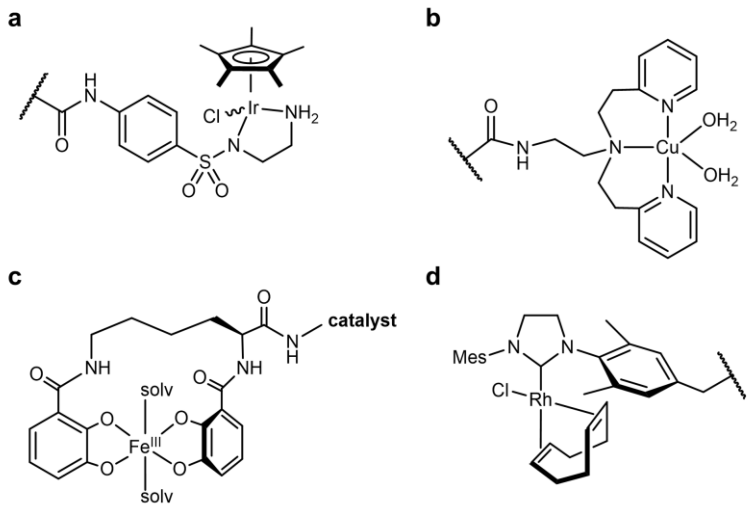


Figure 1.3 Structures of representative inorganic complexes described in sections 1.1.2

a, Ir d^6-piano stool complex for transfer hydrogenation **b,** Cu(biot-et-dpea)) complex, where biotin conjugated moiety is omitted for clarity. **c,** Fe azotochelin linked to a catalyst of interest. The conjugated Ir catalyst is omitted for clarity. **d,** Rh(NHC) phosphonate. The wavy lines in **a**–**b** and **d** indicate the anchoring point on the proteins.

1.1.3 α-Helical peptides and proteins

The α-helical domain is the most extensively investigated secondary structure. A single α-helical turn is composed of 3.6 amino acids, which allows us to translate one-dimensional sequence information into three-dimensional protein structures in a relatively predictable manner. Various α-helical coiled-coil oligomers have been synthesized by placing hydrophobic and charged residues at the interfacial and adjacent sites, respectively.[43] Because the hydrophobic residues are situated at the core of the oligomers, substitution for metal-binding residues has been an effective strategy to prepare catalytic metallo-peptides.[16]

Pecoraro *et al.* synthesized homo and hetero-α-helical trimers as artificial metalloenzymes (Figure 1.4a).[17] The peptides are templated by binding Hg^{2+}, Cd^{2+}, or Pb^{2+} where other metal ions such as Zn and Cu play catalytic roles, such as in CO_2 hydration/ester hydrolysis[44-45] and nitrite reduction,[46-47] respectively. Alternatively, Fe^{2+} can be incorporated into the α-helical structure, resulting in a rubredoxin-like protein, facilitating electron transfer.[48] Similarly, Tanaka *et al.* also have constructed an artificial dinuclear Cu-binding site by placing two histidine and two cysteine residues into a four-α-helical coiled-coil, yielding two copper atoms bridged with two cysteines (Figure 1.4b).[18] Spectroscopic characterization suggests that the artificial metalloenzyme is a structural and functional mimic of cytochrome c oxidase.

In addition, α-helical peptides have been versatile platforms to anchor a dirhodium center. Ball *et al.* created artificial metallopeptides with catalytic activity in enantioselective carbenoid insertion into Si-H bonds[19] and asymmetric cyclopropanation.[49] Kuhlman *et al.* have also designed metal-binding histidine residues on α-helical peptides, resulting in dimeric

metalloproteins with crevice-like architectures (Figure 1.4c).[20] These peptides were further engineered and evolved into a globular metallo-esterase with high enantiospecificity and catalytic efficiency (Figure 1.4d).[50]

Four α-helical bundle proteins have also been extensively employed for the design of artificial metalloenzymes. Dutton *et al.* have created maquettes that bind haem for O_2 transport[51] or exhibit oxidoreductase activity with haem, flavins, Fe-S clusters, Zn porphyrins, and Zn chlorines.[52] The proteins were further engineered into artificial oxidases that react with a diverse array of substrates using hydrogen peroxide as an oxidant.[53]

Tezcan *et al.* employed a four α-helical bundle protein, cytochrome cb_{562}, as a monomer to generate self-assembled tetramers as a 16 α-helical bundle protein (Figure 1.4e).[54-55] Rational and *in silico* redesign produced a tetrameric metalloprotein, creating protein-protein interfaces for the sites for metal-dependent hydrolysis even inside *E. coli* cells.[56] The α-helical protein scaffold is so versatile that various proteins that bind a wide array of first-row transition metal ions can be effectively generated,[57] and the diversion of evolutionary paths differed by local flexibility can be also explored.[58]

DeGrado and Lombardi *et al.* have designed a *de novo* due ferri protein (Figure 1.4f),[59] where the four α-helix bundles provide a dinuclear Fe-coordinating site for artificial phenol oxidase activity[60] and selective N-hydroxylation of arylamines.[59] Alternatively, the four α-helix bundle generated the sites for tetra-nuclear clusters[61] and for highly electron-deficient non-natural zinc porphyrin, the latter of which was highly stable even at 100 °C.[62]

Other than soluble α-helical proteins, DeGrado *et al.* reported the *de novo* design of a functional four α-helix transmembrane protein.[63] Despite great challenges in the design of transmembrane proteins,[64-65] the group has

designed two stable asymmetric states with a dynamic interface to transport selective metal ions, such as Zn^{2+} and Co^{2+} but not Ca^{2+}, across membranes. Although the protein exhibits no catalytic activity, this work suggests that the scope of α-helical scaffolds can be extended to membrane proteins. In addition, the sequence, structure, and function of α-helical proteins can be further expanded as exemplified with a combinatory library of α-helical proteins and their emergence of peroxidase activity.[66] Therefore, α-helical proteins are likely to contribute more versatile scaffolds for *de novo* artificial metalloenzymes in the future.

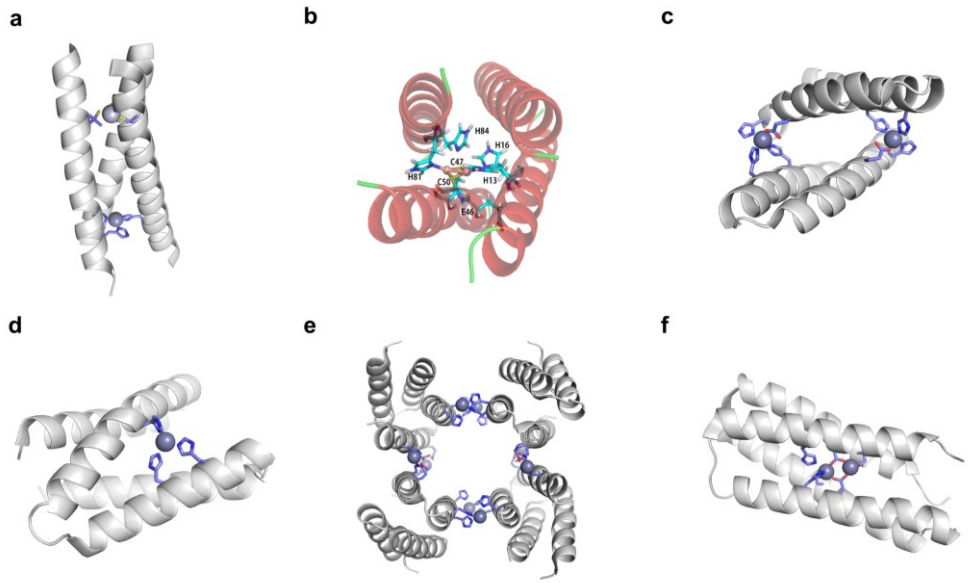


Figure 1.4 Representative structures of *de novo* metallo-α-helical peptides and proteins

a, Zn-dependent metallo-esterase (3PBJ). **b,** A structural model of the dicopper-binding protein. Reprinted with permission from ref 105. Copyright (2012) American Chemical Society. **c,** Zn-dependent hydrolase MID1-Zn (3V1C) **d,** Zn-dependent hydrolase MID1sc10 (5OD1) **e,** Zn-dependent β-lactamase (4U9E) **f,** Diiron artificial metalloenzyme 3His-G2DFsc (2LFD). PDB codes are given in parentheses. Metal atoms and metal-ligating residues are shown with navy spheres and sticks, respectively. In **c** and **d**, co-crystallized metal-binding molecules are also shown with sticks. The metal-bound chloride and water-derived molecule are colored in green and red, respectively.

1.2 Discussion

Significant advancements in the field of artificial metalloenzymes have demonstrated their capacity to enhance catalytic activity, manipulate chemical reactions through metal exchange, and uncover novel chemical reactivity. However, a predominant emphasis on the metal-binding portion of the protein has hindered the exploitation of the full potential of proteins. Considering only the metal binding site, ignoring the broader context of protein functionality is akin to neglecting the functional groups of ligands in metal-ligand design.

To fully capitalize on the versatile nature of proteins, it is imperative to integrate diverse protein scaffolds into the design of artificial metalloenzymes. By doing so, I can leverage the inherent characteristics and functionalities that proteins offer. For instance, thermo-stable proteins enable catalysis at high temperatures,[66-68] and the concave structure of protein scaffolds creates non-aqueous environments, impeding undesired chemical side reactions.[69] Moreover, adapting the native functions of protein scaffolds becomes a promising avenue for achieving desired catalytic activities in artificial metalloenzymes. Intrinsic features such as substrate binding affinity, electron/proton transfer pathways, and hydrogen-bonding networks can be harnessed to further enhance their catalytic performance. The utilization of proteins as privileged starting points in the design process underscores their potential to serve as key building blocks in the development of artificial metalloenzymes.

These considerations highlight the necessity for a new strategy in designing artificial metalloenzymes, one that applicable to diverse protein scaffolds. By moving beyond a narrow focus on the metal-binding portion and embracing the full spectrum of protein functionality, I can unlock new

opportunities for enhanced catalytic performance and expanded functional capabilities in artificial metalloenzymes.

CHAPTER 2

Design and directed evolution of noncanonical β-stereoselective metalloglycosidases

Part of this chapter was adapted from W.J. Jeong*, and W.J. Song. (2022) *Nat. Commun.* 13, 6844. copyright 2022. Nature Publishing Group.

2.1 Introduction

Hydrolases are ubiquitous and indispensable in every living organism. They catalyze the cleavage of various substrates via the association of a water molecule. Depending on the types of hydrolyzed chemical bonds, hydrolases are classified into thirteen enzyme commission numbers[70]. A large fraction of hydrolases are also categorized as metallohydrolases when metal elements are mechanistically involved with forming a nucleophile that reacts with the dedicated substrates[71].

Both metallo- and non-metallohydrolases have been discovered in nearly all subclasses of hydrolases, such as esterase, peptidase, and β-lactamase. However, glycosidase, which reacts with sugars, is an outstanding exception; very few glycosidases necessitate a metal ion[72], and even if they do, its role is limited to substrate binding[73-74] and stabilization of transition state[75] rather than to the direct activation of a water molecule and the cleavage of a glycosidic bond. Instead, canonical glycosidases primarily utilize a pair of acidic residues that function as nucleophile and/or general acid/base[76].

No discovery of metal-dependent glycosidase, namely metalloglycosidase, might be related to the intrinsic properties of metal elements; cationic metal ions may not accommodate a positively-charged oxocarbenium ion-like transition state[77], although such electrostatic repulsion perhaps can be overcome by the assistance of surrounding residues. In addition, inorganic complexes, metallopolymers, and metallopeptides showing glycosidase activities were reported[78-81]. This discrepancy led us to question whether metal-dependent glycosidases can be created within a biomacromolecular space. If so, the scope of artificial metalloenzymes and

whole-cell biocatalysts can expand against the odds of natural emergence and selection[5, 82].

Thus, I sought a versatile protein scaffold to build a metal-binding active site. I supposed that a protein containing a void space, such as β-barrel outer membrane protein F (OmpF) (Figure 2.1a)[83], would be suitable for my study; its overall shape resembles the structure of host compounds, such as cyclodextrin (Figure 2.1b), which provides a structural basis for various functions, including catalysis, by encompassing guest molecules and forming noncovalent interactions[84-85]. In addition, OmpF is an outer membrane suitable for whole-cell catalysis and directed evolution.

In this work, I demonstrate that the structure and mechanism-guided protein redesign and directed evolution of OmpF are transformed into noncanonical Zn-dependent glycosidases with even high β-stereoselectivity and catalytic activities under *in vitro* conditions and as whole-cell catalysts.

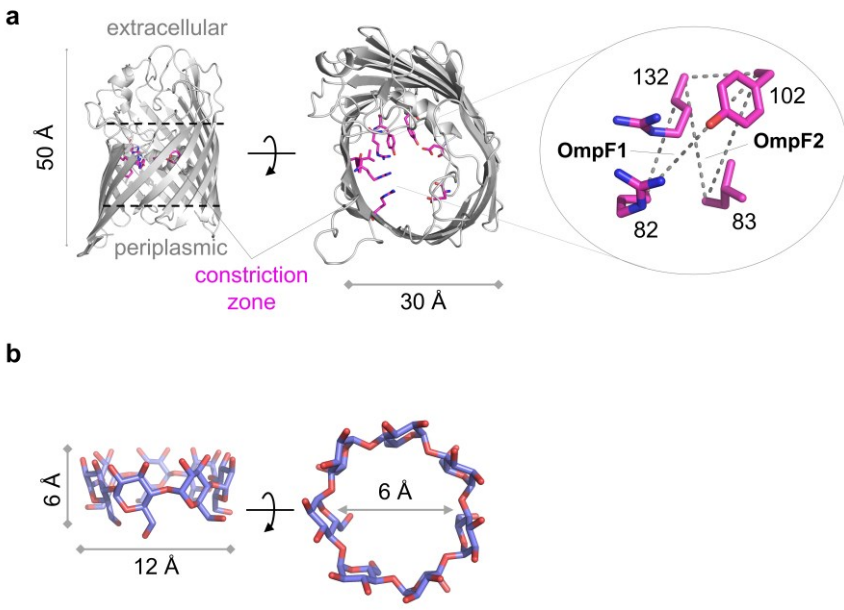


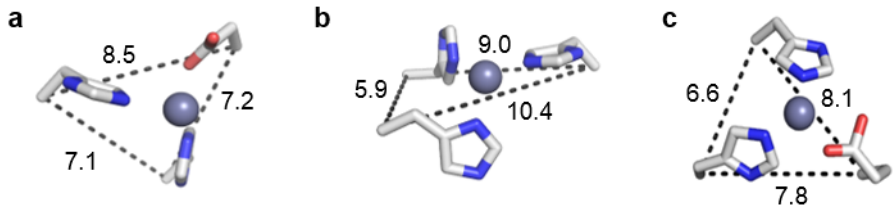
Figure 2.1 OmpF as a versatile scaffold protein

OmpF as a versatile scaffold protein. The structural resemblance of **a,** OmpF (PDB 2OMF) and **b,** cyclodextrin (CSD TEZZUV). Only one protomer of trimeric OmpF is shown for clarity. The horizontal dotted lines in a represent the location of lipid bilayers. The sets of three residues selected for putative Zn-binding motifs are depicted as dotted triangles.

2.2 Results

2.2.1 Structure-based design of artificial Zn-binding proteins

I inspected the constriction zone of OmpF, the narrowest region, approximately halfway inside the barrel-shaped protein. A coordinatively-unsaturated mononuclear Zn-binding site, a common catalytic motif in metallohydrolases, was created by placing three metal-coordinating histidines. R132H mutation was designed first because its side chain is oriented toward the eyelet and the local structure lies at the end of a β-strand; the latter may help to accommodate structural rigidity and flexibility for metal coordination. Then, I identified two positions whose the C_α atoms of residues are located within 3.8–10.6 Å of R132; the geometric boundaries are derived from the structural analysis of natural Zn-binding proteins (Figure 2.2, Figure 2.3, and Table 2.1). As a result, two triple mutants, R82H/Y102H/R132H (OmpF1) and L83H/Y102H/R132H (OmpF2), were designed as the parent templates.



.

Figure 2.2 Geometric analysis of representative natural metallo-hydrolases

a, carbohydrate esterase (PDB 2CC0), **b,** metallo-β-lactamase (PDB 2BC2), and **c,** thermolysin (PDB 4TLN). The distance between the C_α atoms of the Zn-binding residues is shown in Å. Nitrogen and oxygen atoms are colored in blue and red, respectively. Zn ions are shown as spheres.

Table 2.1 Geometric analysis of the representative natural Zn-binding metalloproteins

Proteins (PDB code)	Ligands	Zn−ligating atom (Å)	The $C_\alpha \cdots C_\alpha$ distance of ligating residues (Å)
Carbohydrate esterase (2CC0)	2His/Asp	2.1–2.8	7.1–8.5
Metallo-β-lactamase (2BC2)	3His	2.1–2.3	5.9–10.4
Thermolysin (4TLN)	2His/Glu	2.0–2.3	6.6–8.1
Carbonic anhydrase (3K34)	3His	2.0	5.5–8.1
Thermolysin (4N4E)	2His/Glu	2.0	6.6–8.4
Carboxypeptidase A (1M4L)	2His/Glu	2.0–2.3	6.0–7.2
Leucine aminopeptidase (1LAM)	Asp/Glu/Lys	2.0–2.2	7.6–10.5
	2Asp/Glu	2.0–2.1	5.2–8.7
Alkaline phosphatase (1ALK)	His/2Asp/Ser	1.8–2.2	3.8–7.8
	2His/Asp	2.0–2.3	6.1–8.7
Horse liver alcohol dehydrogenase (1HLD)	His/2Cys	2.2–2.3	7.4–8.1
	4Cys	2.3–2.4	5.1–7.4
Zinc finger protein (4M9V)	2His/2Cys	2.0–2.3	5.7–10.3
	2His/2Cys	2.0–2.3	5.7–9.8
Glyoxalase (2XF4)	2His/2Asp	2.1–2.4	3.8–10.6

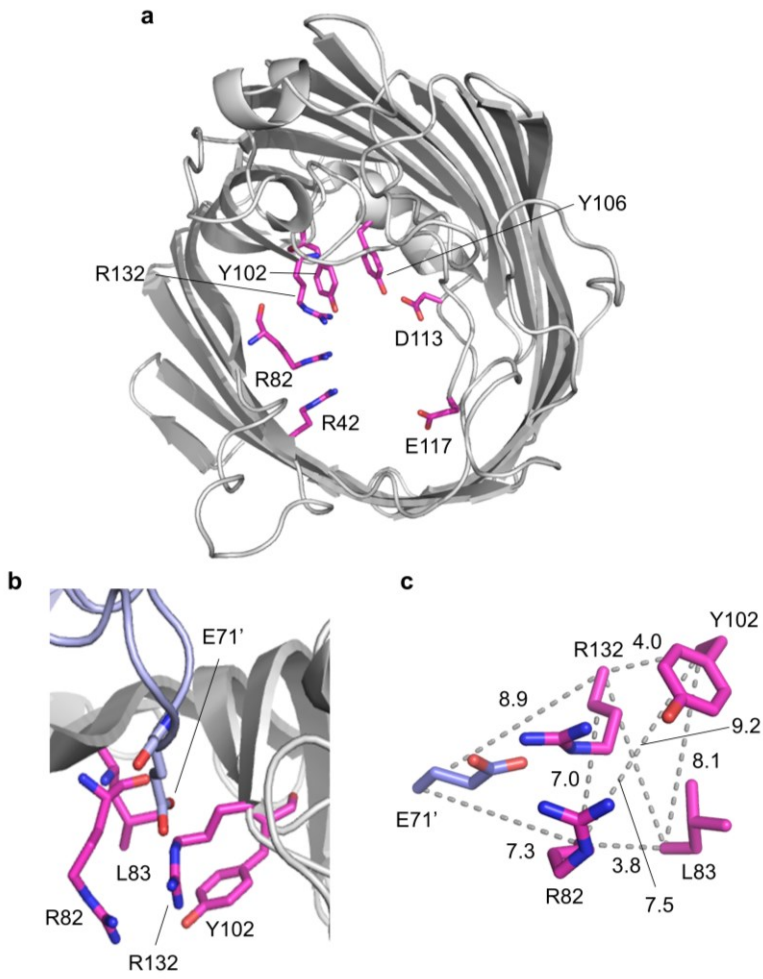


Figure 2.3 Geometric analysis of OmpF

a−b, Residues located in the constriction zone. Only the monomeric subunit is shown, for clarity. The adjacent protomer containing E71′ is colored in light purple. **c,** Measurements of the C_α-C_α distance (Å) between the selected residues for the metal-binding sites.

2.2.2 Structural characterization

Both OmpF variants were obtained from heterologous expression in *E. coli* as described previously[86]. After protein extraction, purification, and refolding, they were isolated as β-barrel trimers, retaining the native overall architecture of OmpF (Figure 2.4). It is also consistent with the X-ray single-crystal structures of OmpF1 and OmpF2 (Table 2.2 and Figure 2.5). More importantly, both proteins possess a mononuclear Zn-binding site in the constriction zone. In OmpF1, two designed residues, R82H and R132H, coordinated to Zn ion, as expected (Figure 2.5b). However, E71′ from another monomer, instead of Y102H, was ligated to the Zn ion, resulting in a 2His/1Glu triad (2.0–2.4 Å for Zn-O/δN bonds) instead of a 3His triad. I later found that E71′ is within the geometric range described above, and I dismissed that a latching loop from another protomer is accessible to the putative Zn-binding site. Regardless, the fourth site was ligated by a non-proteinaceous molecule, tentatively assigned as a water molecule. Consequently, OmpF1 possesses a coordinatively unsaturated Zn-site, satisfying the prerequisite for metal-dependent hydrolysis. Due to the E71 binding, the Y102H mutation becomes no longer needed for Zn coordination, while the residue may still influence the chemical properties of the Zn-site. Therefore, I prepared another variant that reverses Y102H mutation (OmpF1Y) for further investigation.

The Zn-ligating site in OmpF2 comprised of three histidine residues (3His) in a tetrahedral geometry, resulting in 2.1–2.3 Å Zn-εN bonds (Figure 2.5c). The fourth coordination site was coordinatively unsaturated, and therefore, suitable for my studies. Of note, the empty coordination site of OmpF2 pointed toward the periplasmic side, whereas that of OmpF1 is placed toward the extracellular region. Thus, they are likely to show distinct first-

coordination spheres (2His/1Glu versus 3His), secondary coordination spheres, and drastically altered microenvironments, particularly when embedded in the outer membrane of *E. coli* cells.

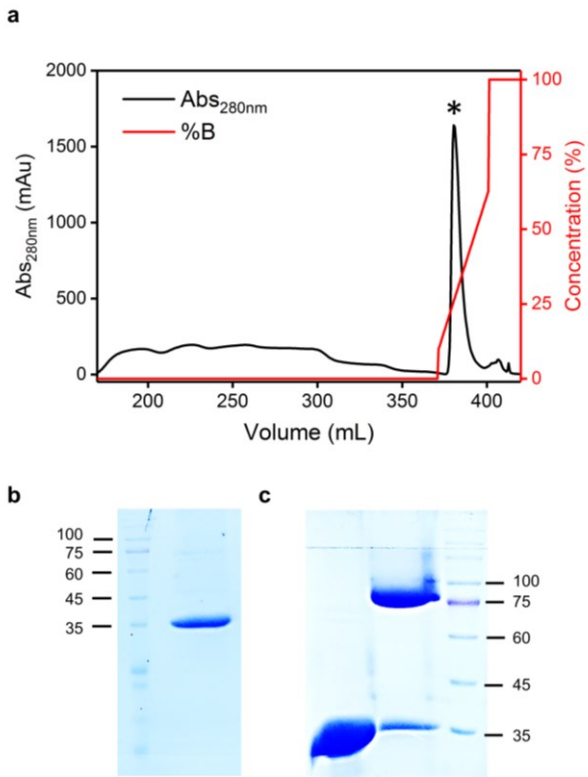


Figure 2.4 The preparation of OmpF

a, FPLC traces. An asterisk indicates the elution of the OmpF protein. **b,** Representative SDS-PAGE of the purified unfolded OmpF protein. **c,** OmpF refolding from an unfolded monomer (left) to a folded trimer (right). As reported previously[86], a properly-refolded OmpF trimer runs faster (80 kDa) than its molecular size (110 kDa) in SDS-gel. The fraction of unfolded monomer was removed by trypsin digestion for further characterization. The data shown in **b** and **c** are representatives of at least ten trials, and all attempts to replicate were successful.

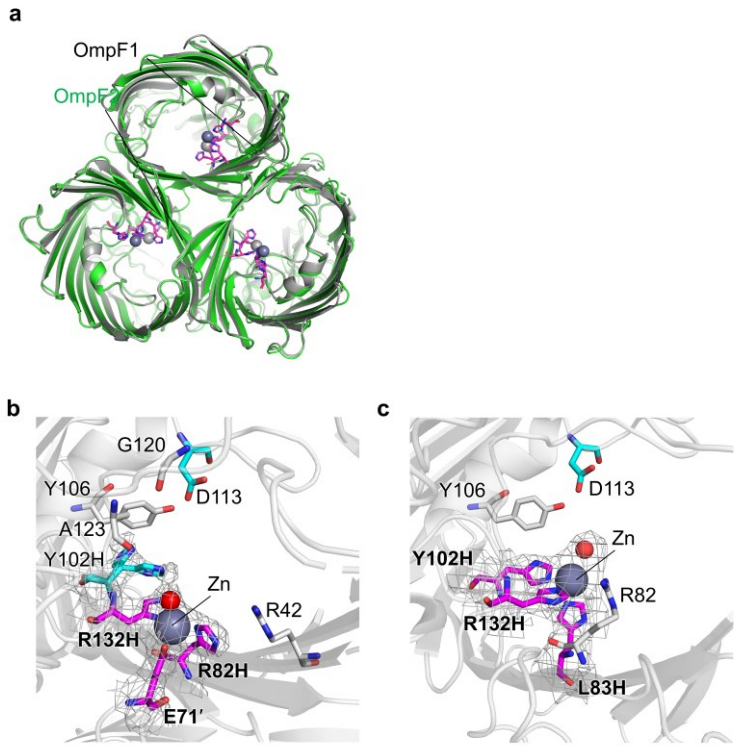


Figure 2.5 X-ray crystal structures

a, The overlaid structures shown from the periplasmic side. The active sites of **b,** OmpF1 and **c,** OmpF2. Zn atoms and a metal-bound water molecule are shown as light navy and red spheres, respectively. Metal-ligating residues are colored in magenta and labeled in bold. Rationally redesigned and iteratively optimized residues are shown with cyan and grey sticks, respectively. The grey grid represents $2F_o - F_c$ electron density of the Zn-ligating site contoured at 1.0 σ.

Table 2.2 Crystallographic data and refinement statistics

	OmpF1 (7FDY)	OmpF2 (7FF7)	OmpF2 (Zn SAD)
Data collection			
Wavelength (Å)	0.979	0.979	1.282
Space group	C2	C2	C2
Unit cell length (a, b, c, Å)	128.6, 140.4, 110.9	153.7, 112.0, 110.8	154.1, 112.4, 111.2
Unit cell angle (α, β, γ, °)	90.0, 120.0, 90.0	90.0, 110.9, 90.0	90.0, 110.9, 90.0
Resolution (Å)	30–3.10 (3.27–3.10)	50–3.38 (3.44–3.38)	50–3.98 (4.05–3.98)
R_{merge} (%)	10.9 (77.7)	12.5 (45.8)	19.8 (85.1)
I / I	10.33 (1.6)	32.5 (8.0)	23.6 (4.0)
Completeness (%)	98.3 (98.1)	99.7 (100.0)	99.6 (100.0)
Redundancy	3.7 (3.7)	7.2 (7.5)	3.8 (3.7)
$CC_{1/2}$ (%)	99.7 (83.9)	98 (98)	98 (82)
Refinement			
Resolution (Å)	29.13-3.1 (3.2–3.1)	38.05–3.38 (3.51–3.38)	
No. reflections	30643 (1532)	23204 (1206)	
R_{work} / R_{free}[b]	0.270/0.296	0.290/0.326	
No. atoms			
Protein	7599	7038	
Ligand/ion	3	3	
Water	3	14	
B-factors			
Protein	86.5	101.43	
Ligand/ion	82.2	80.99	
Water	58.7	62.44	
R.m.s. deviations			
Bond lengths (Å)	0.01	0.005	
Bond angles (°)	1.01	1.37	
Ramachandran plot (%)			
Favored	94.21	87.8	
Allowed	5.79	10.1	
Outliers	0.0	2.1	

A single crystal was used for each data collection. Values in parentheses are for the highest-resolution shell.

2.2.3 *In vitro* Zn-dependent hydrolytic activities of OmpF variants

To determine whether OmpF variants generate a Zn-mediated nucleophilic site for hydrolysis, I first measured esterase activities with a chromogenic substrate, *p*-nitrophenyl acetate (pNPA), with the purified proteins. Zn-dependent net activities were determined by observing the differences in the presence and absence of Zn ions (Figure 2.6a, Figure 2.7, and Table 2.3). The wild-type protein shows no detectible Zn-dependent hydrolytic activities. In contrast, all Zn-complexed OmpF variants exhibited considerably higher Zn-dependent esterase activities, demonstrating that they possess hydrolytically active Zn-binding sites, similar to synthetic, peptide-, and protein-based catalysts[44,56,87] and metalloesterases[71, 88.] Their steady-state kinetic parameters showed the following order of OmpF1 < OmpF1Y ≤ OmpF2 and OmpF1 ≈ OmpF1Y ≤ OmpF2 for k_{cat} and k_{cat}/K_M values, respectively. These values are presumably determined by the combination of their discrete first coordination spheres (2His/1Glu in OmpF1 versus 3His in OmpF2), the directionality of Zn-bound water molecules, and surrounding microenvironments, including the residues at the 102 position (Y102H in OmpF1 versus Y102 in OmpF1Y). They also showed Zn-dependent β-lactamase activities with nitrocefin in the order of OmpF1 < OmpF1Y ≈ OmpF2 (Figure 2.6b, Figure 2.7, and Table 2.3). Their catalytic efficiency or substrate-binding affinity were comparable to or higher than those of Zn-complexes[89-90] and artificial Zn-dependent β-lactamases[56]. In particular, all OmpF variants show saturation curves in Michaelis-Menten analysis, suggesting the OmpF protein scaffold may have a substantial binding affinity with the β-lactam analog even before any sequence optimization. It might be related to the native role of OmpF to be the native passage of antibiotics into the cells[91], implicating that OmpF is a versatile scaffold that functions as a host-like macromolecule and interacts

with guest-like molecules for catalysis.

Finally, I determined whether Zn-complexed OmpF variants are catalytically competent in the hydrolysis of glycosides. From this point, I modified the expression vectors to translocate OmpF variants to the outer membrane of *E. coli* cells to prepare folded membrane proteins directly. The extracted OmpF variants exhibited fluorescence increase upon the addition of a fluorogenic substrate, 4-methylumbelliferyl-β-D-glucopyranoside (4-β-MUG), only at Zn-bound states (Figure 2.10b). The formation of the hydrolyzed product, 4-methylumbelliferone, was also detected by HPLC, demonstrating that OmpF variants indeed hydrolyze the glycosidic bond of 4-β-MUG as Zn-dependent glycosidases.

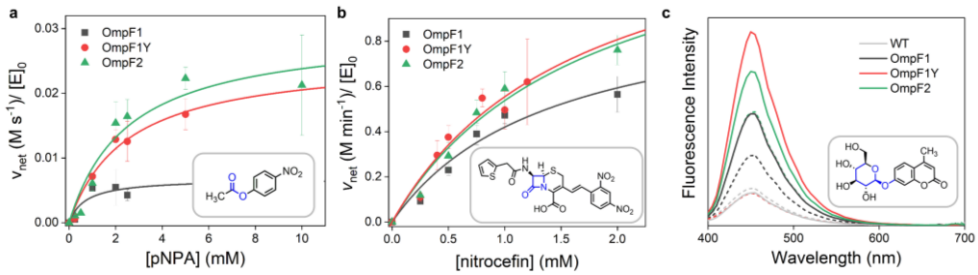


Figure 2.6 Zn-dependent hydrolytic activities

Kinetic analysis with **a,** pNPA and **b,** nitrocefin. **c,** Fluorescence changes upon the hydrolysis of 4-β-MUG. Solid and dotted lines in **c** indicate the presence and absence of Zn ions, respectively. The data shown in **a** and **b** represent the averages and standard deviations of three independent experiments.

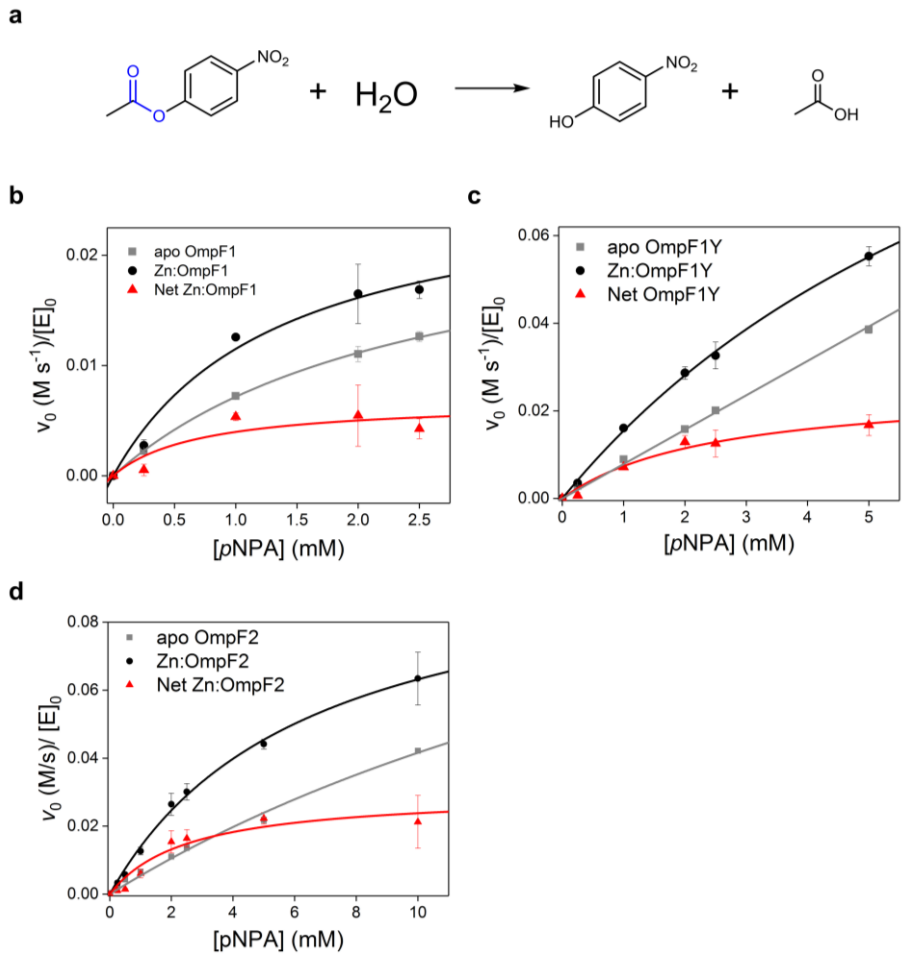


Figure 2.7 Steady-state esterase activities of the OmpF variants with pNPA

a, A reaction scheme. Michaelis-Menten kinetic analysis of **b,** OmpF1, **c,** OmpF1Y, and **d,** OmpF2. The metal-dependent net activities were obtained by the subtraction of the observed activities measured using the apo-protein from those in the presence of Zn ions, in which free Zn ion alone exhibited no detectible activity. The data shown in **b, c,** and **d** represent the average and standard deviation of three independent experiments.

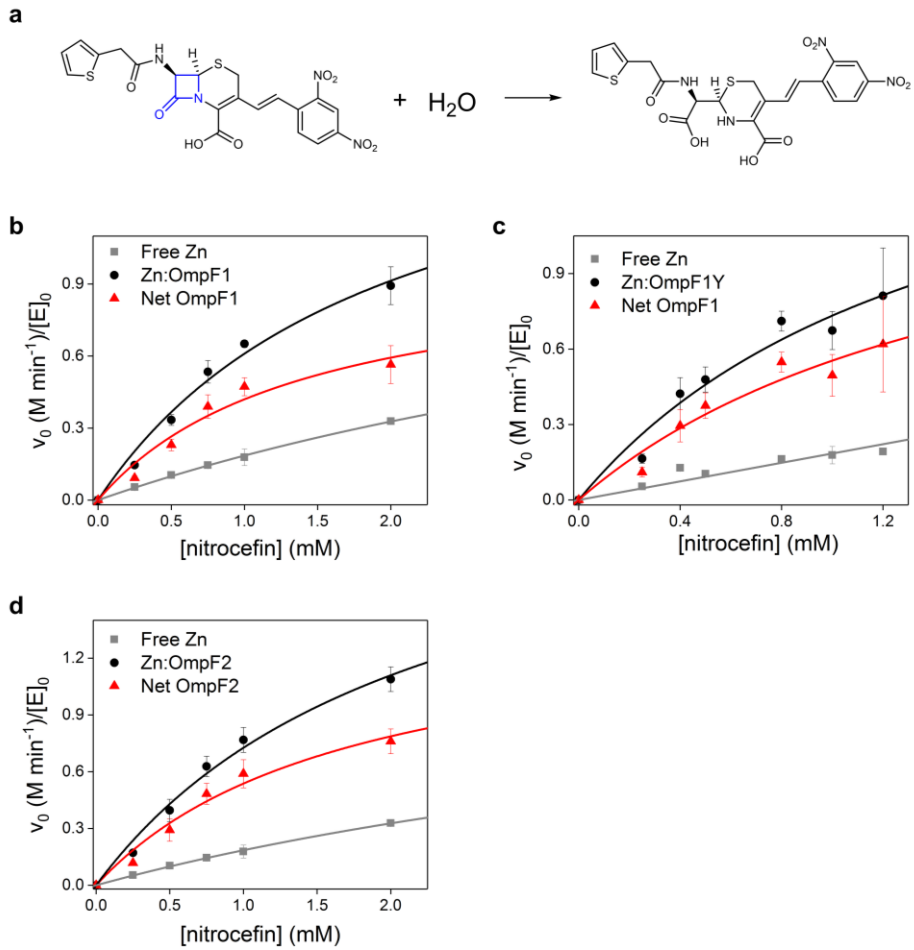


Figure 2.8 Steady-state β-lactamase activities of the OmpF variants with nitrocefin

a, A reaction scheme. Michaelis-Menten kinetic analysis of **b,** OmpF1, **c,** OmpF1Y, and **d,** OmpF2. The metal-dependent net activities were obtained by the subtraction of the observed activities measured with free Zn ions from those in the presence of Zn-bound protein, in which the apo-protein exhibits no detectible activity. The data shown in **b**, **c**, and **d** represent the average and standard deviation of three independent experiments.

Table 2.3 Steady-state kinetic analysis of the OmpF variants

Esterase activity with p-NPA

	k_{cat} (s^{-1})	K_M [97]	k_{cat}/K_M (s^{-1} M^{-1})
OmpF1	0.006 ± 0.001	0.5 ± 0.4	12 ± 8
OmpF1Y	0.026 ± 0.004	2.5 ± 0.8	10 ± 2
OmpF2	0.030 ± 0.004	3 ± 1	12 ± 3

β-lactamase activity with nitrocefin

	k_{cat} (s^{-1})	K_M [97]	k_{cat}/K_M (s^{-1} M^{-1})
OmpF1	0.017 ± 0.004	1.4 ± 0.6	12 ± 2
OmpF1Y	0.02 ± 0.01	2 ± 1	15 ± 3
OmpF2	0.025 ± 0.005	1.7 ± 0.6	14 ± 2

The rates for either apo or Zn-bound samples were measured in triplicate.

2.2.4 Structure- and mechanism-based redesign of OmpF variants

The canonical Koshland mechanism (Figure 2.13b) suggests that inverting and retaining β-glycosidases operate via a pair of acidic residues that are 6–11 Å and 5.1–5.5 Å apart, respectively[76, 93-94]. I surmised that the analogous mechanism is operative to OmpF variants. In that case, the glycosidase activity might arise from assisting at least one pre-existing acidic residue near the Zn-site (Figure 2.7 and Table 2.4). In particular, D113 shows proper orientation and interatomic distance between the terminal oxygen atom and Zn-bound exchangeable ligands as 7.9–9.5 Å and 4.7–5.1 Å in OmpF1 and OmpF2, respectively (Figure 2.7c-f). In addition, D113E mutation (denoted as /E) altered the glycosidase activities of OmpF variants; OmpF1/E and OmpF2/E showed substantially elevated activities, whereas that of OmpF1Y/E was somewhat reduced (Figure 2.10c). Although the impacts of D113E mutation are dissimilar with OmpF variants, these data suggest that the acidic residue is functionally coupled with the Zn-site, possibly playing an essential role in glycosidase activities.

Of note, OmpF1/E, but not others (OmpF1, OmpF1Y, OmpF2, OmpF1Y/E, and OmpF2/E), exhibited glycosidase activity even as the Zn-free apo-state (Figure 2.10c), implying that alternative catalytic pairs, possibly two acidic residues, might have co-emerged serendipitously along with Zn-mediated ones. Because wild-type protein or D113E single mutant showed no such activity with 4-β-MUG, the unexpected activities of OmpF1/E at the apo-states are likely to be associated with the introduction of mutations for Zn-ligation (R82H or R132H). Regardless, all OmpF variants are more active as Zn-bound forms than in the apo-states, indicating that they still primarily function as Zn-dependent glycosidases.

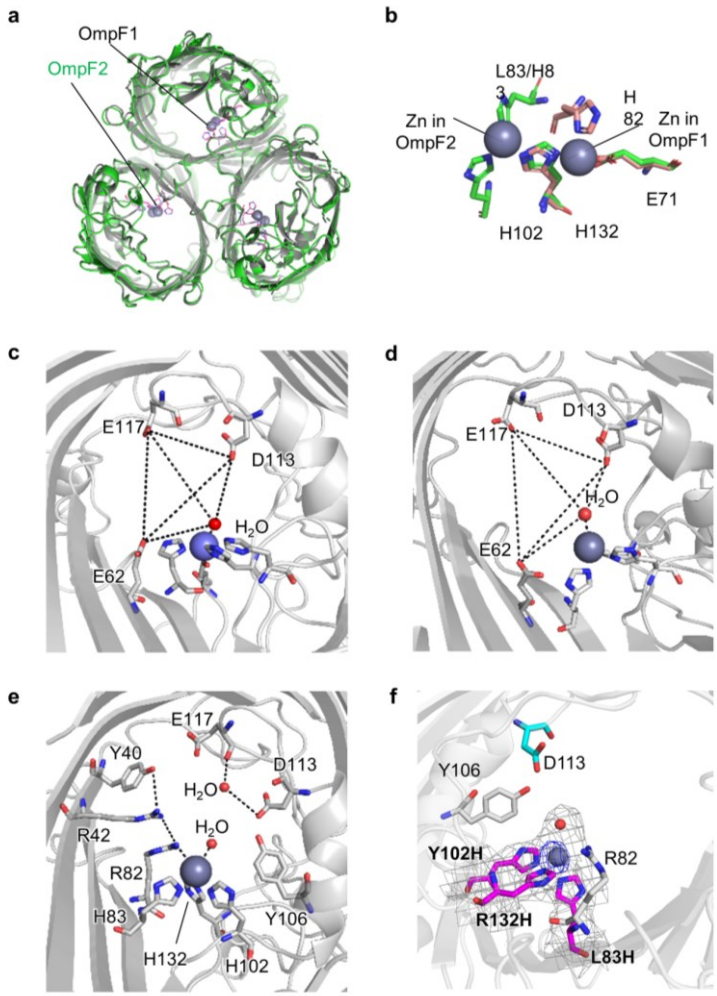


Figure 2.9 The X-ray crystal structures of the OmpF variants

a, The overlaid structures from the extracellular side. **b,** The Zn-binding sites of the OmpF1 and OmpF2 variants, colored in light magenta and green, respectively. The acidic residues located near the Zn-binding site of **c,** OmpF1 and **d,** OmpF2 variants. The interatomic distances between key residues and motifs are summarized in Table 2.4. **e,** The hydrogen-bonding network observed in the Zn-binding site of OmpF2. Their distances are within 2.6–3.4 Å. **f,** The overlaid electron density maps in OmpF2. The grey and blue grid represents the $2F_o - F_c$ electron density contoured at 1.0 σ and anomalous difference maps contoured at 2.0 σ, respectively.

Table 2.4 The geometric parameters of the Zn-binding sites in Om pF1 (PDB 7FDY) and OmpF2 (PDB 7FF7).

The distance of key residues

Interatomic distance (Å)	OmpF1	OmpF2
Zn−ligating atoms	2.0–2.4	2.1–2.3
C_α–C_α of ligating atoms	6.6–9.2	5.4–8.9
Zn-bound H_2O and D113	7.9–9.5	4.7–5.1
Zn-bound H_2O and E117	10.8–13.3	10.0–11.7
Zn-bound H_2O and E62	8.2–10.2	6.2–7.3
Oxygen atoms between D113 and E117	6.4–12.0	6.3–11.3
Oxygen atoms between D113 and E62	11.0–13.5	11.2–132
Oxygen atoms between D113 and E71	11.6–13.5	10.9–13.2
Oxygen atoms between E117 and E71	12.8–16.1	11.9–15.1
Oxygen atoms between E117 and E62	12.6–14.9	11.8–14.7
Oxygen atoms between E71and E62	8.0–10.4	9.4–10.6

The angles of key residues

	OmpF1	OmpF2
The angles of O/N−Zn−N of Zn-ligating residues (°)	94.9–119.4	71.3–99.4

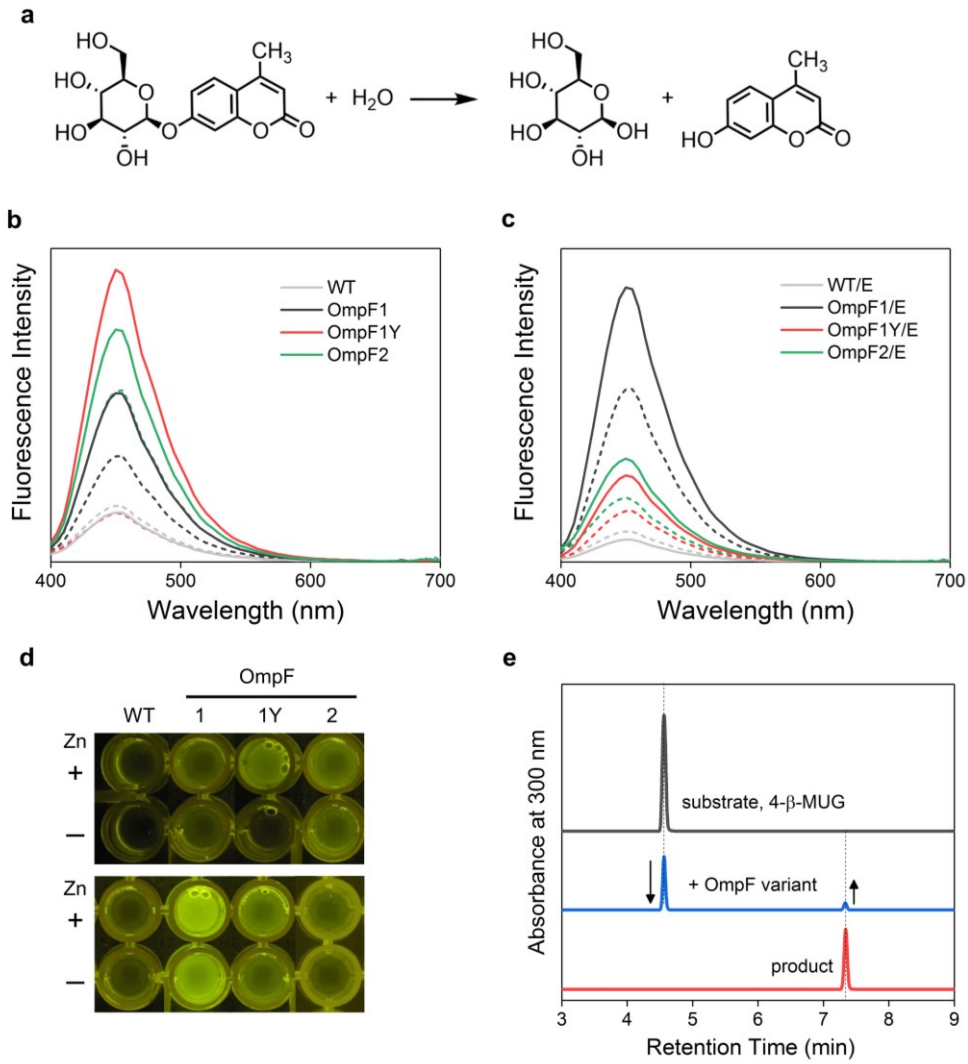


Figure 2.10 Steady-state glycosidase activities of the OmpF variants with 4-β-MUG

a, A reaction scheme. **b–c** The representative fluorescence spectra of the OmpF variants incubated with 4-β-MUG due to the formation of 4-methylumbelliferone as the hydrolyzed product. **b,** The wild-type protein and OmpF variants **c,** The D113E (/E) mutations of parent variants. The solid and dotted lines indicate the proteins in the presence and absence of Zn ions, respectively. **d,** Fluorescence images of the lysates. **e,** The representative HPLC trace of the reaction mixtures due to the consumption of 4-β-MUG and

the formation of the hydrolyzed product, 4-methylumbelliferone.

2.2.5 Optimization of the active site pockets by directed evolution

For directed evolution, I attached the N-terminal signal peptide of native OmpF into the N-terminus of our protein (Figure 2.11). With three D113E mutants (OmpF1/E, OmpF1Y/E, and OmpF2/E) as initial templates, I iteratively constructed whole-cell mutant libraries. Then, they were screened with two substrates with β-1,4-glycosidic linkages, cellobiose and 5-bromo-4-chloro-3-indolyl-β-D-glucopyranoside (X-Glu), simultaneously in Zn-supplemented growth medium. The growth and formation of blue-colored colonies indicated that the glucose and 5,5′-dibromo-4,4′-dichloro-indigo are produced sufficiently as a carbon source for cell growth and an insoluble blue pigment, respectively (Figure 2.12). After the iterative screening, the catalytically active cells were further sorted by the reactivities with 4-β-MUG. As a result, the following mutants were obtained: R42S/Y106A/G120C/A123N (OmpF1/E-R4), Y106R/R42E/G120S/A123V (OmpF1Y/E-R4), and Y106H/R82C (OmpF2/E-R2) (Figure 2.13). They showed significantly elevated whole-cell activities with the initial rates up to two orders of magnitude relative to those of their parent proteins (P), OmpF1, OmpF1Y, and OmpF2, respectively (Figure 2.14).

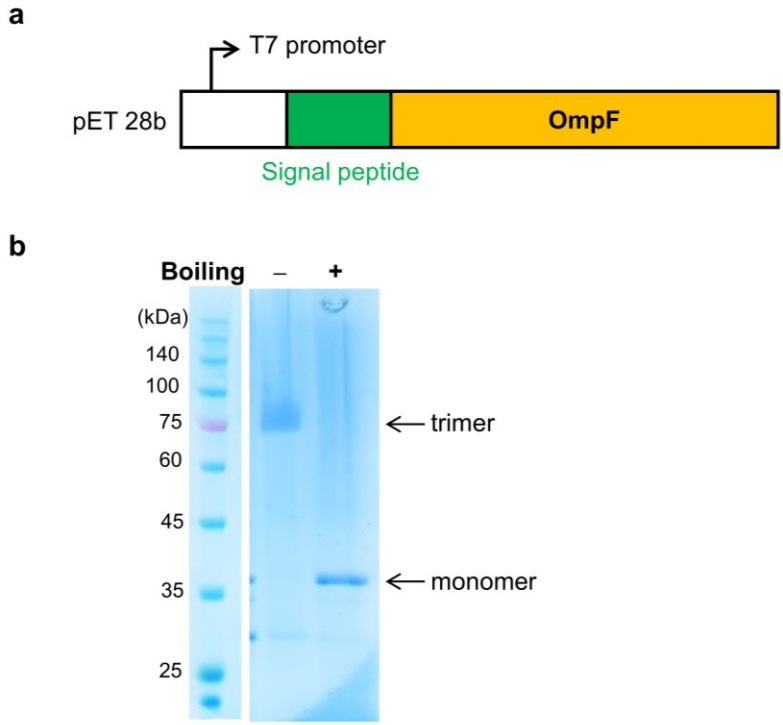


Figure 2.11 Expression and translocation of the OmpF variants to the outer membrane

a, The design of the OmpF expression vector by inserting a signal peptide at the N-terminus. **b,** Representative SDS-PAGE of the OmpF variants extracted from *E. coli* cells (left) before and (right) after boiling. The data shown in **b** are representatives of at least ten trials, and all attempts to replicate were successful.

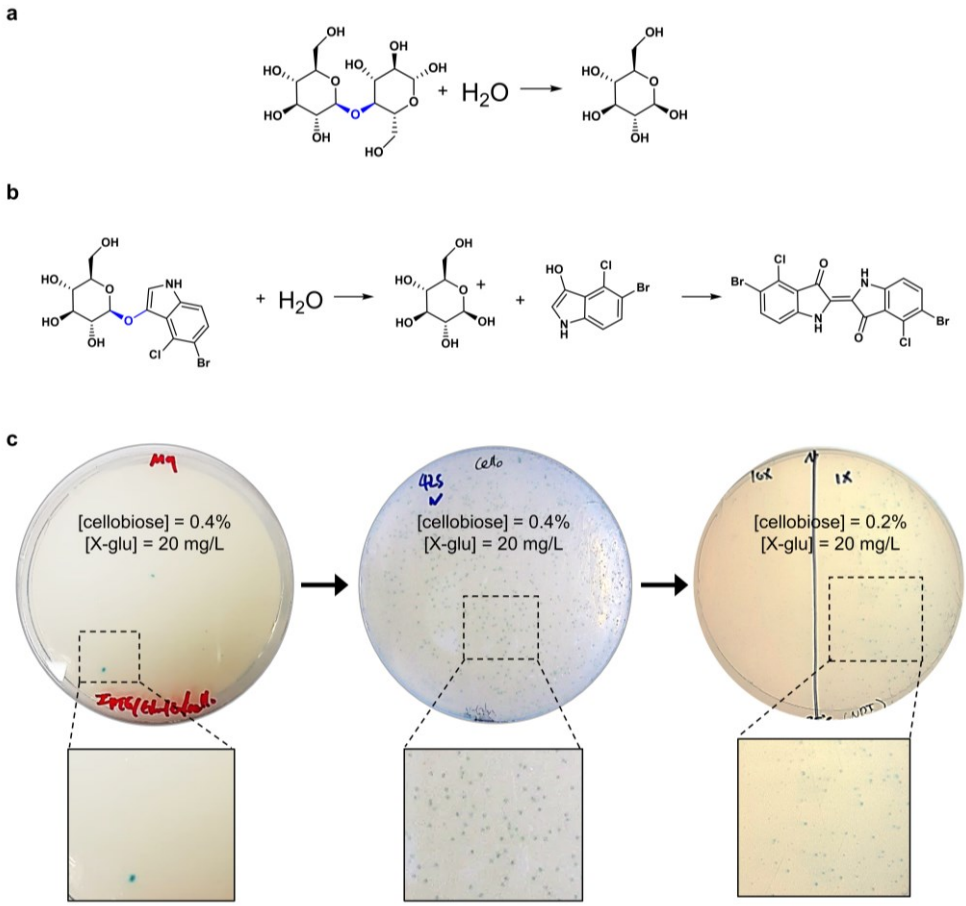


Figure 2.12 Screening of whole-cell mutant libraries

The reaction schemes of the hydrolysis of **a,** cellobiose and **b,** X-Glu. **c,** Representative agar plates to select artificial glycosidases with elevated glycosidase activity. Throughout the iterative rounds of screening, the concentration of cellobiose was adjusted gradually from 0.4% to 0.2%. The formation of blue colonies owing to the hydrolysis of two substrates were observed.

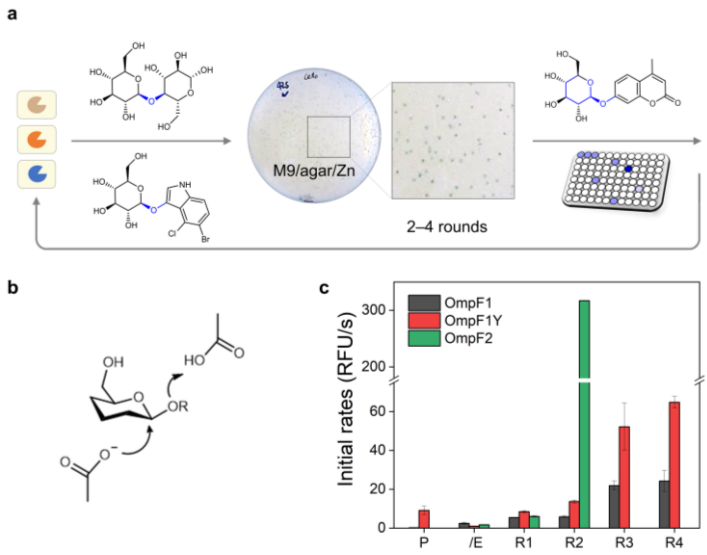
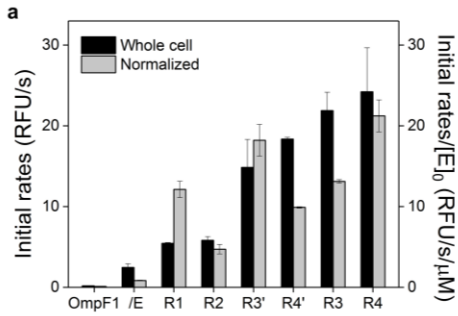
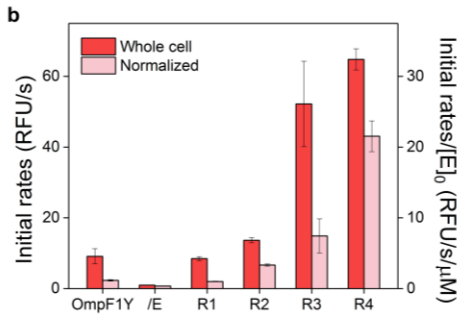


Figure 2.13 Representative scheme of directed evolution

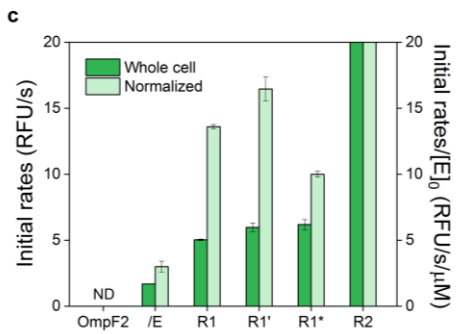
a, Directed evolution of Zn-dependent β-glucosidase. **b,** Canonical glycosidase mechanism. **c,** Whole-cell activity of the selected mutants on 4-β-MUG, where P and /E indicate the parent proteins (OmpF1, OmpF1Y, and OmpF2) and their D113E single-mutants, respectively. The screening data shown in **c** represent the averages and standard deviations of two technical replicates.



OmpF1	sequence
/E	D113E
R1	D113E/R42S
R2	D113E/R42S/Y106A
R3′	D113E/R42S/Y106A/A123E
R4′	D113E/R42S/Y106A/A123E/G120N
R3	D113E/R42S/Y106A/G120C
R4	D113E/R42S/Y106A/G120C/A123N



OmpF1Y	sequence
/E	D113E
R1	D113E/Y106R
R2	D113E/Y106R/R42E
R3	D113E/Y106R/R42E/G120S
R4	D113E/Y106R/R42E/G120S/A123V



OmpF2	sequence
/E	D113E
R1′	D113E/R82Y
R1	D113E/Y106H
R1*	D113E/R82Y/Y106H
R2	D113E/Y106H/R82C

Figure 2.14 The screening results of OmpF variants

The screening results of **a,** OmpF1 **b,** OmpF1Y and **c,** OmpF2 as parent templates. (left) The whole-cell activity values were measured using 4-β-MUG, and the normalized activity values were obtained by dividing the whole-cell activity values by the protein concentration values of the cell lysates determined using SDS-PAGE. (right) The sequence of the screening results. R1–R4 indicate the rounds of selection. The prime and asterisk indicate the output from alternatively constructed mutant libraries and the addition of two screening hits, respectively. The fluorescence changes with OmpF2/E-R2 in **c** exceeded the detection limit, and I adjusted the gain for

quantitative analysis. The whole cell data and normalized data shown in **a, b,** and **c** represent the average and standard deviation of at least two technical replicates of the experiments and three independent experiments, respectively.

2.2.6 Characterization of the evolved OmpF variants

For more accurate kinetic studies, I measured the glycosidase activities of the evolved variants under *in vitro* conditions. They are active in the order of OmpF1Y/E-R4 < OmpF1/E-R4 < OmpF2/E-R2 with the second-order rate constants, k_2 = 5.6–10.4 min^{-1} M^{-1} (Figure 2.15a). The kinetic parameters account for up to 2.8 x 10^9-fold enhancement from the uncatalyzed rate[95] (Figure 2.17 and Table 2.5). Intriguingly, the iterative mutations also elevated the glycosidase activities of their apo-forms, suggesting that the alternative Zn-independent reaction routes have developed simultaneously via sequence optimization. However, the inductively coupled plasma-mass spectrometry (ICP-MS) analysis demonstrated that OmpF2/E-R2 protein shows a Zn-bound state on the outer membrane of *E. coli* cells (Table 2.6), indicating that whole-cell activities are primarily derived from the coordinatively unsaturated Zn-sites.

When 4-α-MUG with an α-1,4-glycosidic linkage was employed instead of 4-β-MUG, all evolved OmpF variants forms displayed nearly no or negligible whole-cell activity (Figure 2.15b, Figure 2.16). Such high β-stereoselectivity is likely to be related to the structures of three substrates, cellobiose, X-Glu, and 4-β-MUG, used for iterative sequence optimization. Thus, their chiral active site pockets might have been optimized for β-glycosides via the orchestrated interactions with the adjacent residues such that 4-β-MUG, but not 4-α-MUG, can be positioned as a catalytically relevant orientation.

I also measured the turnover number (TON) of the evolved variants with 4-β-MUG, the mole of products per that of enzyme, using cell lysates (Figure 2.15c). The values were increased up to ~100-fold relative to those of the parents, again reflecting the chemical power of sequence optimization. The

removal of the Zn ion or the mutation of adjacently located acidic residues, such as E113, E117, and E62 (Figure 2.18), partially or entirely inactivated the enzymes, suggesting that a Zn-site and at least one acidic residue in proximity constitute a noncanonical catalytic motif of Zn-dependent glycosidases.

In addition, I observed that the OmpF2/E-R2 variant could hydrolyze the β-glycosidic bond in n-octyl-β-D-glucopyranoside (OG), which was initially added as a nonionic detergent for the preparation of membrane proteins. The formation of glucose and n-octanol was detected as the hydrolyzed products (Figure 2.19), indicating that the evolved OmpF variant accommodates a hydrophobic and bulky substrate that was not even used for selection.

All three evolved OmpF variants showed pH-dependent activities with 4-β-MUG (Figure 2.15d and Figure 2.20). Both Zn-complexed and apo-states show bell-shaped pH-dependence, revealing at least two ionizable side-chains to be essential for catalysis (pK_a = ~7.7 and 8.3). It is consistent with natural glycosidases having two discrete pK_a values[96-97]. However, Zn-complexed and the apo-state of OmpF variants yielded the maximal TONs at pH 8.0–8.5 and 7.5–8.0, respectively, whereas most natural glycosidases show the pH optimum at pH 5.0–6.5[97-99]. But glycosidases, such as alkaline xylanases, exhibit catalytic reactivity at more basic conditions[100-102]. The pH-dependence of OmpF variants might be derived from replacing one of an acidic pair with a Zn-OH_2/OH moiety. Besides, the unique chemical environments of OmpF might account for the pH-dependence; it was reported that the pK_a values of the side-chains in the constriction zone are considerably perturbed from those of amino acids in bulk solvents[103-105].

To further identify the catalytic motifs of OmpF variants, a mechanism-based covalent inhibitor, conduritol B epoxide (CBE)[106-107], was added, and

the resulting proteins were analyzed by trypsin digestion and tandem LC/MS spectrometry. CBE molecules are bound to at least two positions of Zn-complexed OmpF variants, one to the Zn(OH) complex ligated to either R82H or Y102H residues (Figure 2.21a and Figure 2.22a) and the other to the adjacently located acidic residues, such as E62, D113E, and E117 (Figure 2.21a, Figure 2.22b, and Table 2.7). In the apo-states, CBE was conjugated to D113E or additional positions, such as D92 or E117 (Figure 2.21b, Figure 2.22c, Figure 2.22d, and Table 2.7). These results contrast with the wild-type protein and D113E single mutant showing no attachment of CBE in the constriction zone (Table 2.7).

I also analyzed the reaction product with 4-β-MUG to identify whether the retaining versus inverting mechanism is operative. Because glucose shows too rapid mutarotation, I conducted activity assays in the presence of excess azide, which functions as an external nucleophile, as described previously[97, 108-109]. Although OmpF2/E-R2 still yielded glucose exclusively, OmpF2/E-R2* variant (Y106H/R82Y), which differs from OmpF2/E-R2 only by a single residue at the position 82, produced the mixtures of glucose and 1-azido-1-deoxy-glucose. The ^{1}H NMR analysis demonstrated that β-glycoside is isolated as the hydrolyzed product, indicating that I created a retaining β-glycosidase that operates via two sequential reactions, glycosylation and deglycosylation, from the actions of two catalytic motifs as a nucleophile and a Lewis acid/base (Figure 2.23).

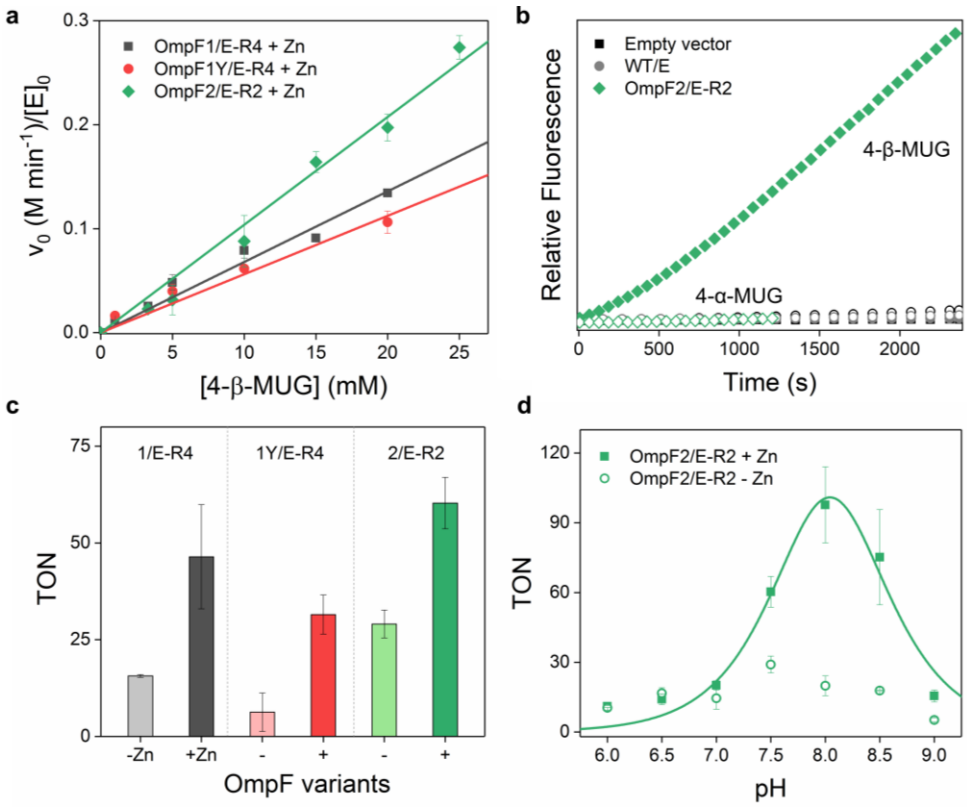


Figure 2.15 Glucosidase activities of OmpF variants with 4-β-MUG

a, Kinetic analysis. **b,** Whole-cell activities of OmpF2/E-R2. **c,** Turnover number (TON) in the absence and presence of Zn ions. **d,** The pH-dependent activities of OmpF2/E-R2. The data shown in **a**, **c**, and **d** represent the average and standard deviation of three independent experiments.

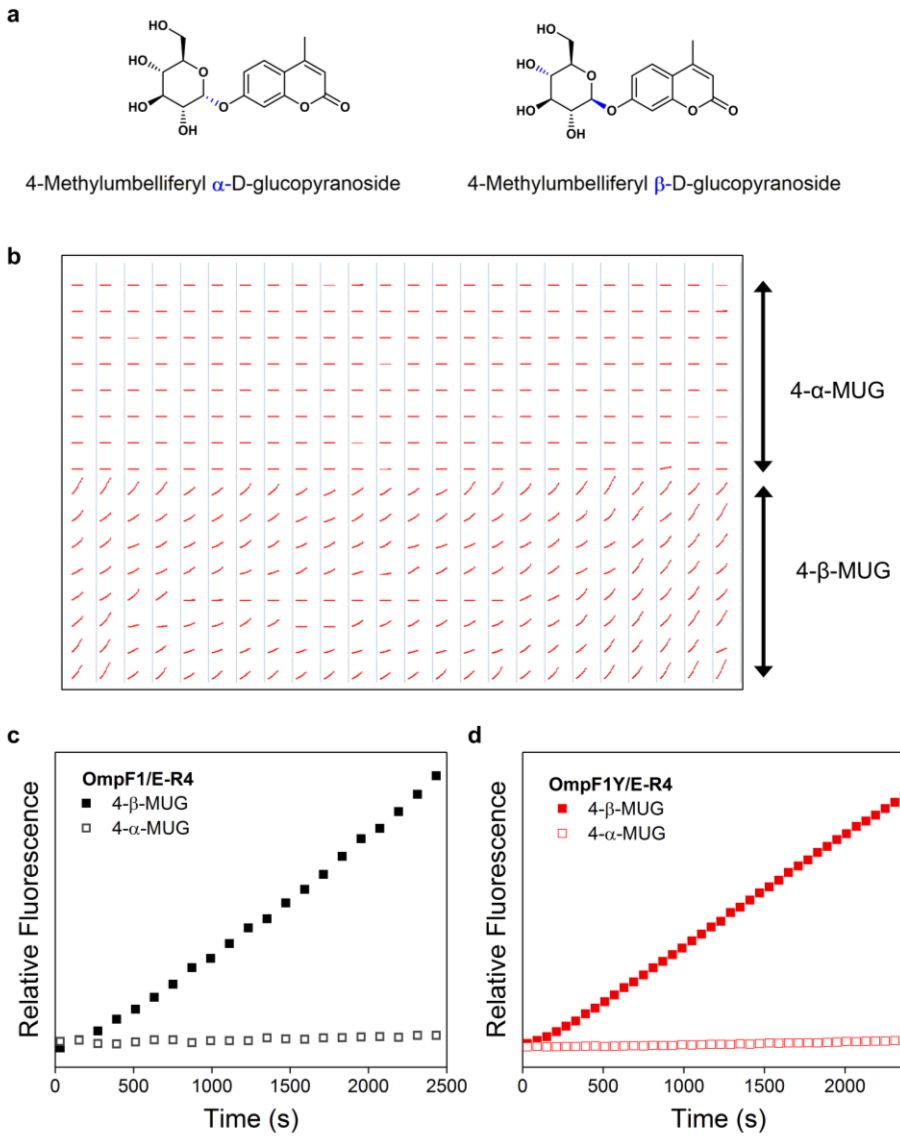


Figure 2.16 Representative measurements of the stereoselective glycosidase activities of the OmpF variants

a, Structures of 4-MUG having either a α- or β-glycosidic bond (4-α-MUG or 4-β-MUG, respectively). **b,** Time-dependent fluorescence assays of 4-α-MUG or 4-β-MUG added to the duplicates of whole-cell libraries. The results of **c,** OmpF1/E-R4 **d,** OmpF1Y/E-R4.

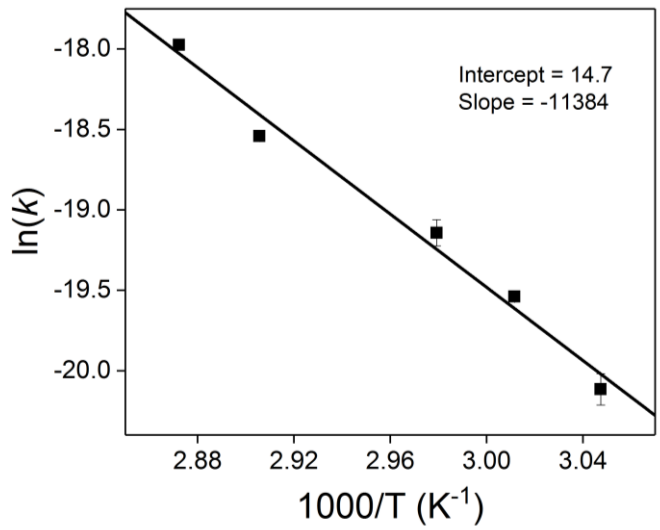


Figure 2.17 Measurements of the uncatalyzed hydrolytic rates of 4-β-MUG

The data represent the average and standard deviation of three independent experiments.

Table 2.5 Glycosidase activities of the OmpF variants with 4-β-MU G

	Zn	k_2 (min^{-1} M^{-1})	k_2/k_{uncat} (M^{-1})
OmpF1/E-R4	–	2.7 ± 0.1	0.99(± 0.1) x 10^9
	+	6.8 ± 0.3	1.83(± 0.3) x 10^9
OmpF1Y/E-R4	–	2.8 ± 0.7	0.76(± 0.1) x 10^9
	+	5.6 ± 0.3	1.51(± 0.2) x 10^9
OmpF2/E-R2	–	4.4 ± 0.1	1.2(± 0.2) x 10^9
	+	10.4 ± 0.3	2.8(± 0.3) x 10^9

The steady-state kinetic parameters were obtained as second-order rate constants (k_2) with the OmpF variants (10 μM). The uncatalyzed rate constant (k_{uncat}) was measured to be 3.72(± 0.63) × 10^{-9} min^{-1} as shown in Figure 2.17. The rates were measured in triplicate.

Table 2.6 ICP-MS results of OmpF2/E-R2

Molar ratio	Cu	Zn
WT	N.D./1.38	0.560/0.753
OmpF2/E-R2	N.D./1.40	2.105/2.118

The molar ratio of metal to protein were obtained from two independently prepared batches of the samples. (N.D.; not detected)

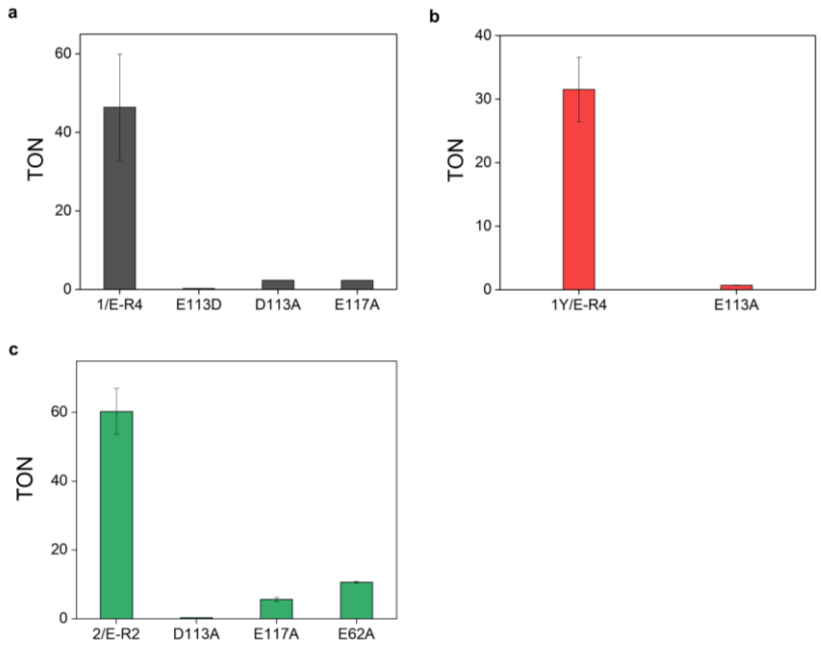


Figure 2.18 The glycosidase activities of OmpF variants

One of acidic residues nearby the Zn-site in **a,** OmpF1/E-R4 **b,** OmpF1Y/E-R4, and **c,** OmpF2/E-R2 variants are modified with alanine or aspartate. The data shown in **a**, **b**, and **c** represent the average and standard deviation of three independent experiments.

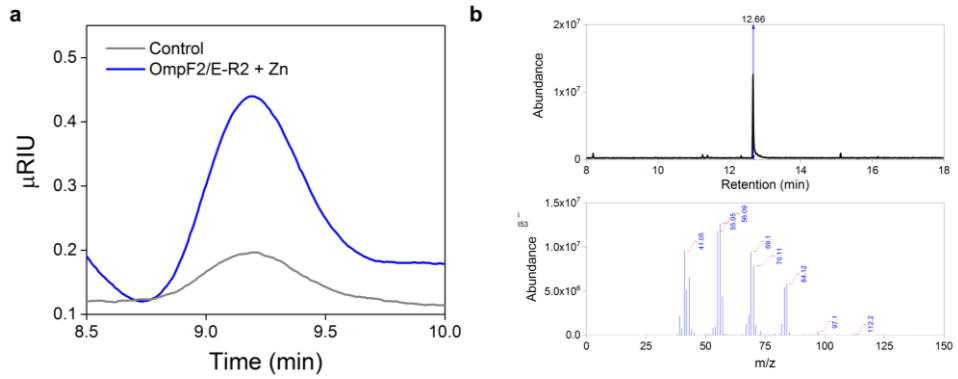


Figure 2.19 The catalytic activities of OmpF variants with OG

Detection of **a,** glucose by HPLC and **b,** 1-octanol by GC-MS.

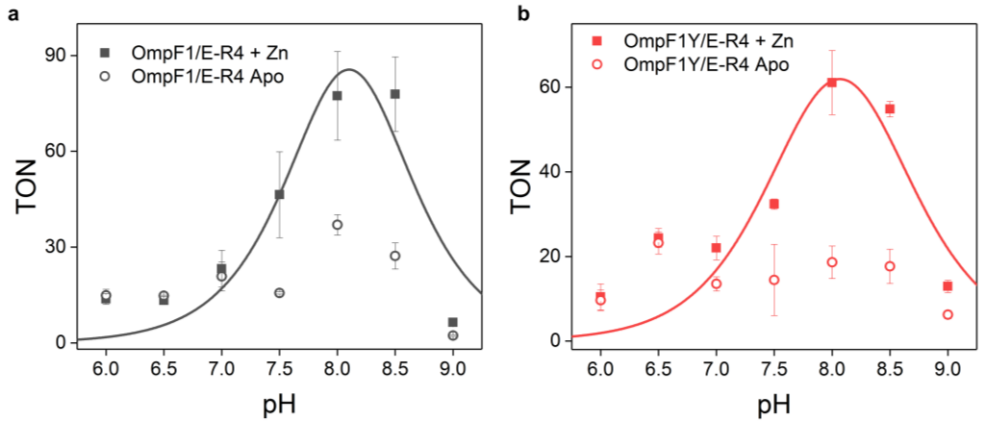


Figure 2.20 The pH-dependent glycosidase activities of OmpF variants

a, OmpF1/E-R4 **b,** OmpF1Y/E-R4. The data represent the average and standard deviation of three independent experiments.

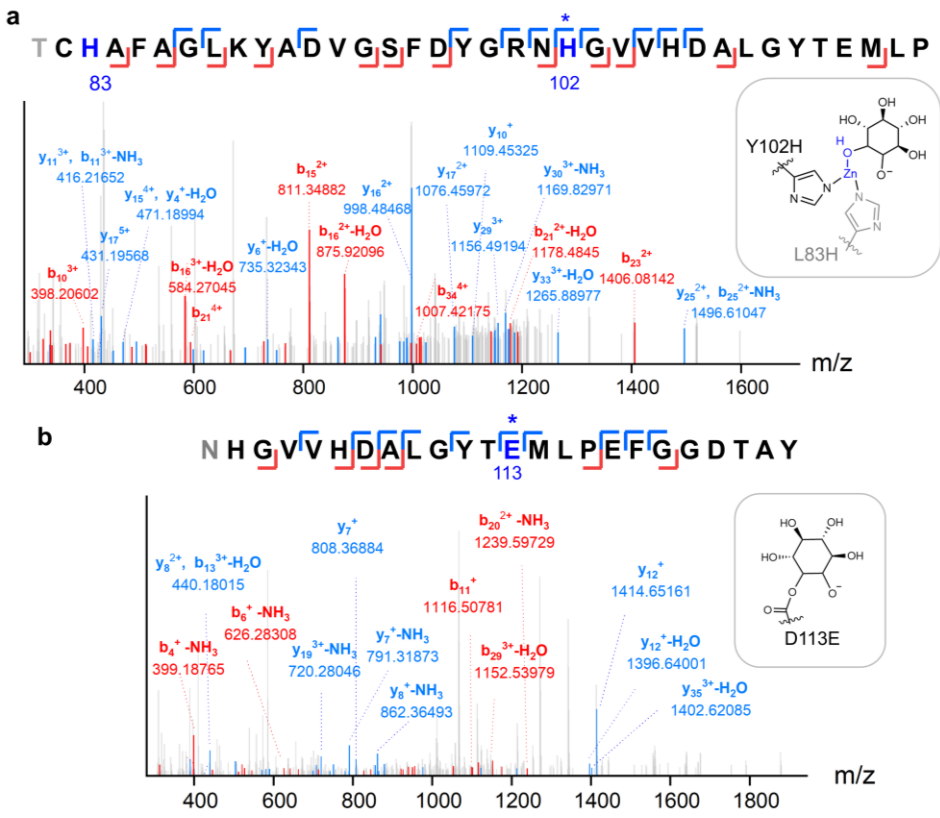
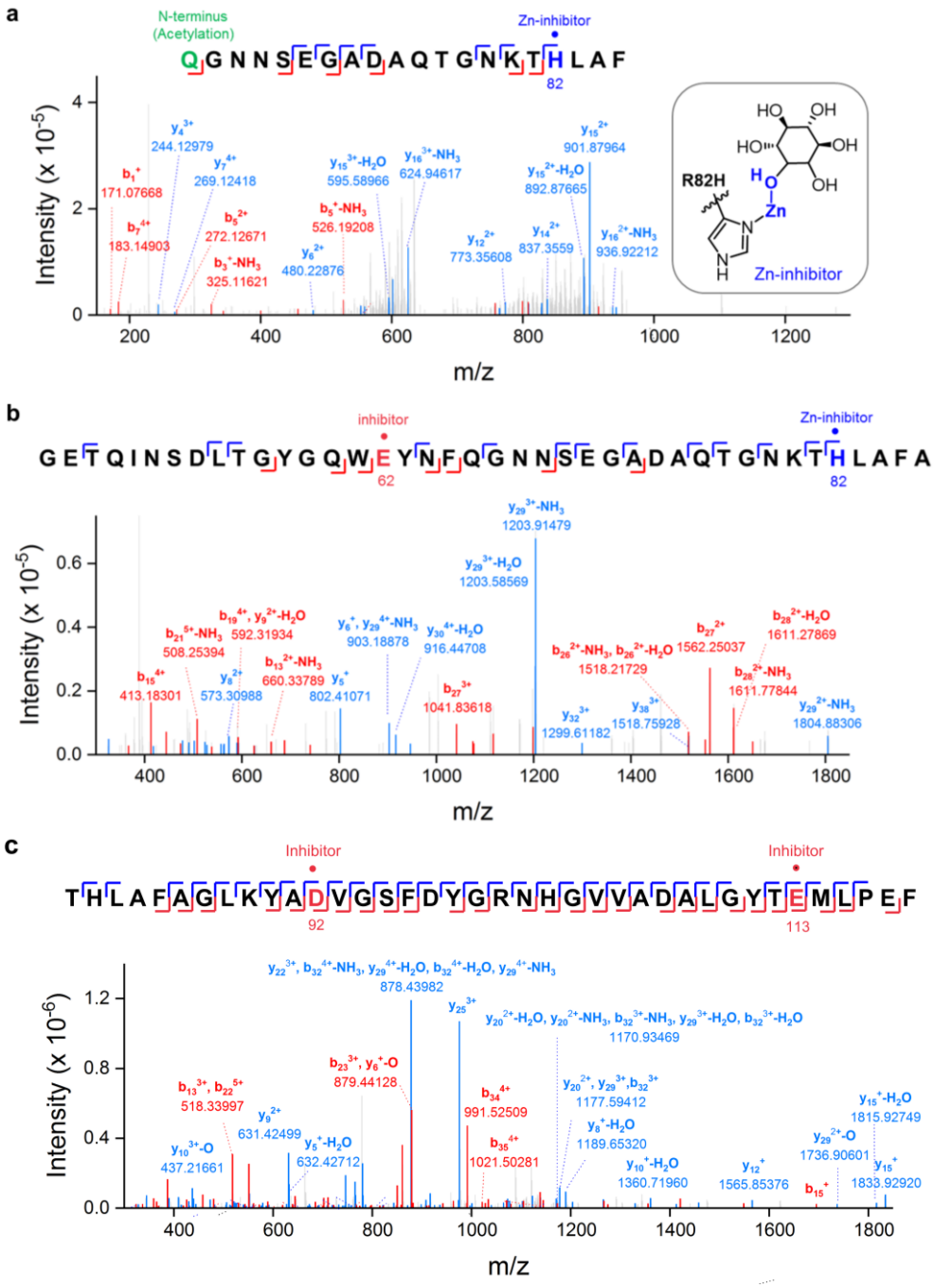


Figure 2.21 Mechanistic studies of OmpF2/E-R2 variant

The representative tandem LC/MS mass fragmentations after treating with CBE inhibitor with **a,** Zn-bound and **b,** apo proteins. Only major peaks are labeled for clarity. The proposed structures of Zn(OH)-CBE ligated to Y102H and L83H and CBE-conjugated D113E are shown as insets.


a
N-terminus
(Acetylation)
Zn-inhibitor
Q G N N S E G A D A Q T G N K T H L A F
82
Intensity (x 10-5)
m/z
R82H
Zn-inhibitor
b
inhibitor
Zn-inhibitor
G E T Q I N S D L T G Y G Q W E Y N F Q G N N S E G A D A Q T G N K T H L A F A
62
82
Intensity (x 10-5)
m/z
c
Inhibitor
Inhibitor
T H L A F A G L K Y A D V G S F D Y G R N H G V V A D A L G Y T E M L P E F
92
113
Intensity (x 10-6)
m/z

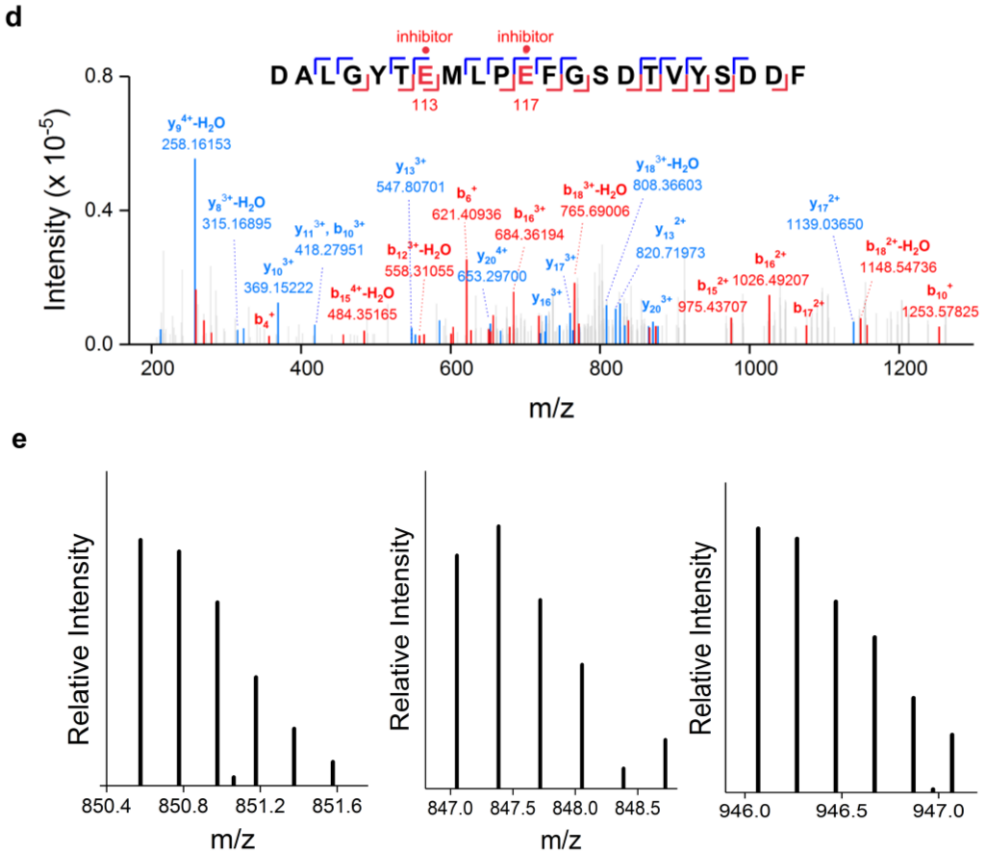


Figure 2.22 Tandem LC/MS analysis of the trypsin-digested OmpF variants after incubation with CBE

a, OmpF1/E-R4 and **b,** OmpF1Y/E-R4 at Zn-complexed states. **c,** OmpF1/E-R4 and **d,** OmpF1Y/E-R4 at the apo-states. Only the representative fragments containing CBE inhibitor at the constriction zone are shown for clarity. **e,** The fragments possessing Zn-moiety shows the characteristic patterns due to naturally occurring zinc isotopes (left: OmpF2/E-R2, middle: OmpF1/E-R4 right: OmpF1Y/E-R4).

Table 2.7 A list of fragments conjugated with CBE in tandem LC/ MS analysis

Zn-bound states

Protein	Proximal to Zn site	Non-constriction zone
WT	ND	D37, D92, E212, D221
D113E	ND	D221
OmpF1/E-R4	E62, D113E, E117, ZnOH:R82H	E29, D37, E48, D97, D107, D121, D149, D172, E181, E183, D195, E284, D290, E296, D312, D319
OmpF1Y/E-R4	E62, ZnOH:R82H	E29, D37, E42, D221
OmpF2/E-R2	E62, E71, ZnOH:Y102H	E29, D37, E48, D54, D74, D107, D121, D172, E181, E183, D195, E233, D266, E284, D288, E296, D312, D328

ND, Not detected

Apo-states

Protein	Proximal to Zn site	Non-constriction zone
WT	ND	D312
D113E	ND	ND
OmpF1/E-R4	E71, D113E	D92, D172, E181, E183, D195, E201, D221, D282, E284, D312
OmpF1Y/E-R4	D113E, E117	D92, E201, E212, D221, D290
OmpF2/E-R2	D113E, E117	E29, D37, D121, D172, E181, E212, D221, E233

ND, Not detected

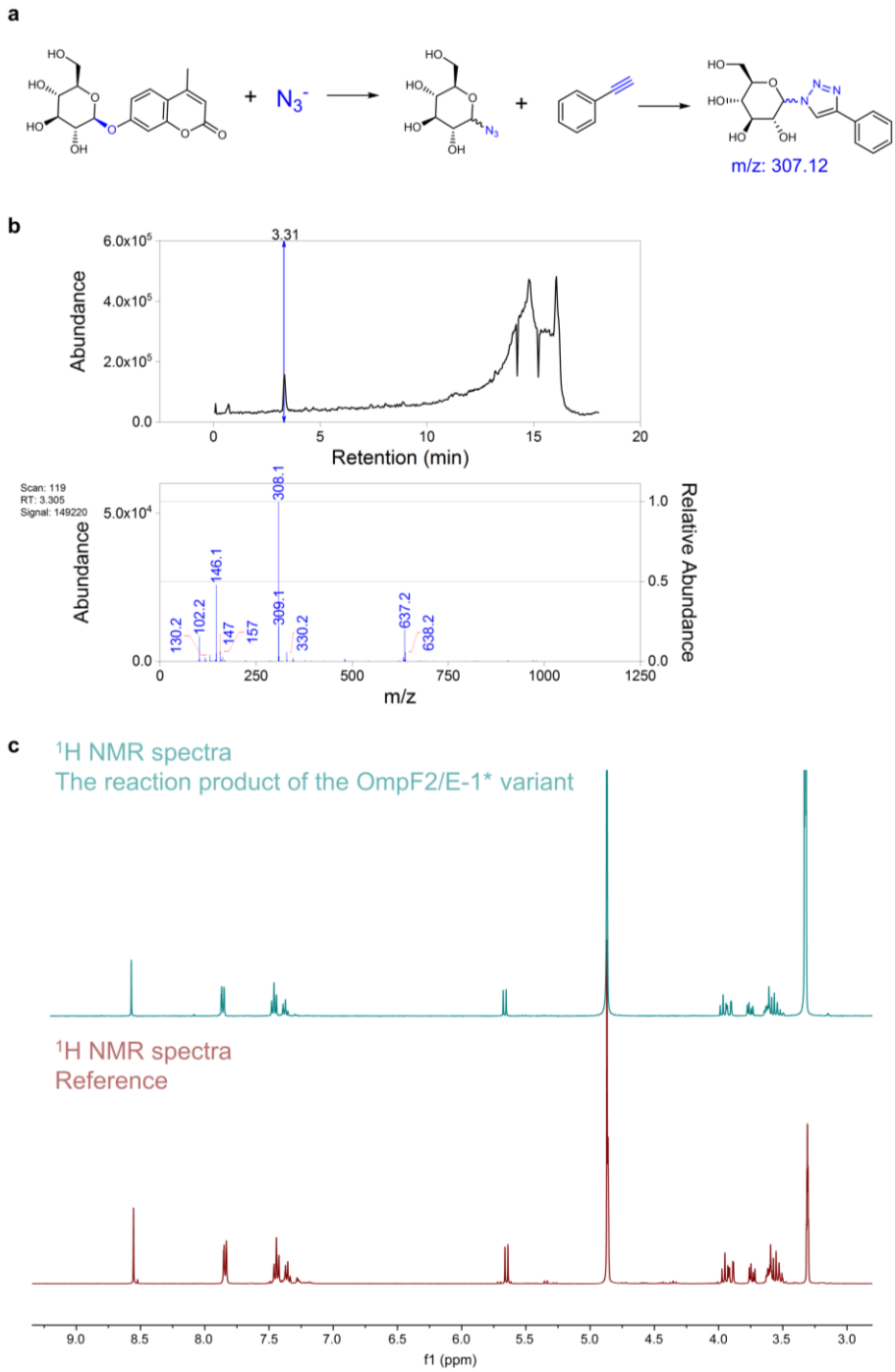


Figure 2.23 Characterization of the glycosidase activity

a, A reaction scheme. **b,** ES-API mass spectrum. Calculated *m/z* for $C_{14}H_{17}N_3O_5^+$ $[M+H]^+$: 308.12; observed: 308.1. **c,** ^1H NMR spectra. The reaction product of the OmpF2/E-1* variant (bottom) 1-β-D-glucopyranosyl-

4-phenyl-1H-1,2,3-triazole synthesized from 1-azido-1-deoxy-glucose as a reference. The dotted box highlights the chemical shift for a proton at the C1 anomeric carbon (5.65 ppm), indicative of a β-isomer. The asterisk in c, indicates the residual H_2O solvent.

2.2.7 Proposed mechanism of Zn-dependent glycosidases

Therefore, I proposed a reaction mechanism of Zn-dependent glycosidases by revising the Koshland mechanism, where one of the canonical acid pairs is replaced by a coordinatively unsaturated Zn-site (Figure 2.24). For retaining β-glycosidases, such as OmpF2/E-R2, an acidic residue, possibly D113E, may initiate glycosylation as a nucleophile when Zn-OH_2 species acts as a Lewis acid. Then, the resulting Zn-OH may become a direct nucleophile or activate a water molecule as a Lewis base, facilitating the deglycosylation step.

Docking simulations with Zn-complexed OmpF1/E-R4 and OmpF2/E-R2 model structures also suggest that the β-glycosidic bond of the 4-β-MUG may be positioned at the appropriate orientation to the coordinatively unsaturated Zn-site and D113E in the constriction zone as described above (Figure 2.25). Therefore, my work demonstrates that coordinatively unsaturated Zn-site can directly mediate the hydrolysis of glycosidic bonds, overcoming potential energy barriers with the formation of an oxocarbenium ion-like transition state during glycoside hydrolysis.

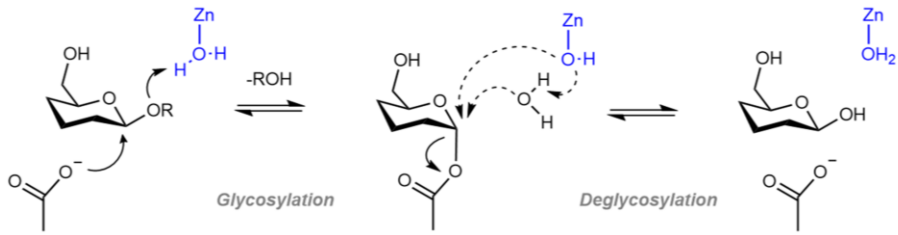


Figure 2.24 A proposed mechanism of Zn-dependent retaining β-glycosidase

The association of a water molecule is omitted for clarity.

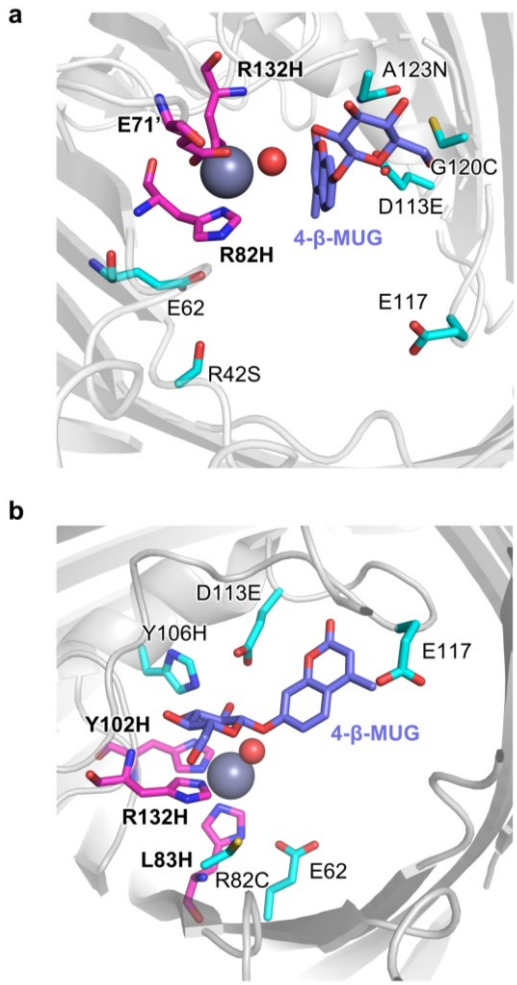


Figure 2.25 Docking simulation of OmpF variants with 4-β-MUG

A possible mode of substrate-binding in **a,** OmpF1/E-R4 and **b,** OmpF2/E-R2. Zn-binding residues are labeled in bold. The selected positions for iterative sequence optimization or adjacently located acidic residues are shown with cyan sticks. The Zn atoms and Zn-bound water molecules are shown with light navy and red spheres, respectively. The bound 4-β-MUG substrate is shown with navy sticks.

2.3 Discussion

The structure- and mechanism-based design of artificial metalloenzymes allows us to validate our level of understanding of the chemical interplay between metal elements and protein environments. This study demonstrated that the retrosynthetic construction of the active sites, reaction mechanism-based redesign, and directed evolution could create artificial metalloglycosidases, enabling inorganic cofactors to activate glycosidic bonds in protein environments. The successful conversion of neither catalytic nor metal-binding membrane protein OmpF into various metallohydrolases indicated that OmpF is apt for introducing an active site, reactions with biologically or non-biologically relevant substrates, and even whole-cell catalysis. Although membrane proteins have rarely been adapted for enzyme designs, my work demonstrated that OmpF could be a host macromolecule for various potential applications. Then, it leads to an intriguing question, why is there no precedence for such metal-dependent glycosidases in Nature? It might be related to the alternative metalloenzymes and relevant biomimetic complexes[110-112], which react with glycosides via Cu-dependent oxidative cleavage mechanism. Alternatively, there might still be unidentified metalloproteomes and metalloenzymes. Along this line, it is worthy to note that a novel glycosidase has been discovered and recently characterized, where Zn-coordinating cysteine functions as a direct nucleophile for glycosidase via a retaining mechanism[113]. They suggested that the Zn ion plays a direct role in tuning the nucleophilicity of the catalytic cysteine and the energetics of metalloenzymes, broadening the role of metallocofactors in glycosidases. In addition, the current study demonstrated that Zn-dependent and independent catalytic routes can co-emerge serendipitously. Such promiscuous reactivity might be related to divergent evolution.[114] If then,

artificially designed enzymes may resemble the states of primordial or ancestor enzymes[115].

2.4 Materials and methods

2.4.1 Structural analysis of OmpF

To create a coordinatively unsaturated mononuclear Zn-binding site on a protein scaffold, I manually curated and inspected the crystal structures of the proteins that natively possess a Zn-binding motif. The distance between the C_α atoms of the ligating residues was measured using the PyMOL program (Figure 2.1, Figure 2.2, and Table 2.1). The metal-ligating residues are located within 3.8–10.6 Å between their backbone C_α atoms, regardless of their sequence, overall structure, and function. Therefore, the values were applied as geometric restraints for my structure-guided design of the Zn-binding sites.

2.4.2 Design of triad mutants

The constriction zone was inspected to install a Zn-binding motif (Figure 2.1 and Figure 2.3). I first narrowed down the candidate positions into the following: Y102, Y106, D113, E117, and R132, due to the orientation of their side-chains, which pointed towards the void space of the cylindrical protein. The residues were prioritized depending on their secondary structures and local flexibility. R132 was the most desirable site as the first vertex of a Zn-binding triad, and the other two vertexes were determined based on the optimal range of the interatomic C_α distance, which was obtained as described above. As a result, two sets of triads, OmpF1 and OmpF2, comprising R132H/Y102H/R82H and R132H/Y102H/L83H, respectively, were determined.

2.4.3 Sample preparation

A plasmid responsible for OmpF heterologous expression (pET28b/kan^R/ompf) was a generous gift from Professor Seokhee Kim at Seoul National University. I deleted a His-tag and a thrombin cut-site prior to the N-terminus using polymerase chain reaction (PCR) with custom-designed primers (Table 2.8). The PCR products were transformed into E. coli DH5α competent cells for isolation and sequencing (pET28b/kan^R/ΔHis/ompf).

For the preparation of OmpF triad mutants, site-directed mutagenesis was carried out using custom-designed primers. After digestion with DpnI restriction enzyme for 1.5 h at 37 ºC, the plasmids were transformed into *E. coli* DH5α competent cells. Each colony was inoculated in 5 mL LB media with 50 mg/L kanamycin at 37 °C, and the plasmid was extracted for sequencing (Macrogen or Bionics).

OmpF variants were prepared as reported previously[86]. In short, the plasmids were transformed into *E. coli* BL21(DE3) competent cells. A few colonies were inoculated in LB medium with 50 mg/L kanamycin. After overnight growth at 37 ºC, the culture was inoculated in LB medium containing 50 mg/L kanamycin at 37 ºC in an orbital shaker. After the optical density of the culture (OD_{600}) reached 0.7, protein expression was induced by the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG) at a final concentration of 1 mM at 37 ºC for 4 h. The cells were harvested via centrifugation at 5,000 rpm (4,715 × *g*) and 4 ºC for 15 min, and the pellets were stored at −80 ºC for further use.

The cells were thawed in 50 mM Tris/HCl (pH 8.0) buffer and subjected to lysis via sonication. The lysates were centrifuged at 13,000 rpm (18,800 × *g*) and 4 ºC for 30 min, and the pellets were washed with 1% (v/v) Triton X-100 in 50 mM Tris-HCl (pH 8.0) buffer. The buffer with 8 M urea was added

to the pellet and incubated for 4 h at 37 ºC. The inclusion bodies were removed via centrifugation at 13,000 rpm (18,800 × *g*). The resulting supernatant was loaded on a Q anionic exchange column (HiTrap Q HP, GE Healthcare Life Sciences) using a protein purification system (ÄKTA Prime Plus). The OmpF protein was eluted by applying a linear gradient of 1 M NaCl, and the purity of the proteins was analyzed using 15% SDS-PAGE (Bio-rad). The pure fractions were concentrated at 4 ºC using a 30 kDa cutoff centrifugal filter (Amicon) or a stirred cell system (EMD, Millipore) with a 10 kDa cutoff membrane. The protein concentration was determined using a UV/vis spectrophotometer (Agilent, Cary 8454) using the extinction coefficients at 280 nm estimated from the protein sequence.

The detergents used for protein refolding were purchased from Anatrace, Thermofisher, Avanti Polar Lipids, and GoldBio. In short, non-ionic detergents, 1% (w/v) n-dodecyl-β-D-glucopyranoside and 1% (w/v) n-dodecyl-β-D-maltopyranoside were added to the unfolded OmpF protein (20 mg/mL), followed by the incubation at 37 ºC for 1–2 days. The protein folding and oligomerization were determined using 15% SDS-PAGE analysis (Figure 1.2c). The refolding efficiency was measured to be approximately 80%. The unfolded proteins (around 20%) were removed via trypsin digestion (Sigma Aldrich, 0.1 mg/mL) overnight at 37 ºC, followed by anion exchange column chromatography using a 50 mM Tris/HCl (pH 8.0) buffer containing 0.5% OG and NaCl (0–1 M). Metal-free protein samples were prepared by adding 5-fold molar excess of ethylenediaminetetraacetic acid (EDTA) and incubating overnight at 4 ºC. The excess EDTA was removed by washing with a metal-free 50 mM Tris/HCl (pH 8.0) buffer containing 0.5% OG several times, using a 30 kDa cutoff centrifugal filter.

Table 2.8 Custom-designed DNA primers for site-directed mutagenesis

Variants	Mutations	Primers
R82H/Y102H/R132H (OmpF1)	R82H	5′-GGTAACAAA ACGCATCTGGCATTCGCG G-3′ 5′-CCGCGAATGCCAGATGCGTTTTGTTACC-3′
	Y102H	5'-GATTACGGCCGTAACCACGGTGTGGTTTATG-3' 5'-CATAAACC CACCGTGGTTACGGCCGTAATC-3'
	R132H	5′-CTTCGTTGGTCATCATGGCGGCGTTGCTAC-3′ 5′-GTAGCAACGCCGCCATGATGACCAACGAAG-3′
L83H/Y102H/R132H (OmpF2)	L83H	5′-GTAACAAAACGCGTCATGCATTCGCGGGTC-3′ 5′-GACCCGCGAATGCATGACGCGTTTTGTTAC-3′
	Y102H	5'-GATTACGGCCGTAACCACGGTGTGGTTTATG-3' 5'-CATAAACCACACCGTGGTTACGGCCGTAATC-3'
	R132H	5'-CTTCTTCGTTGGTCATGTTGGCGGCGTTG-3′ 5'-CAACGCCGCCAACATGACCAACGAAGAAG-3'
R82H/R132H (OmpF1Y)	R82H	5′-GGTAACAAAACGCATCTGGCATTCGCGG-3′ 5′-CCGCGAATGCCAGATGCGTTTTGTTACC-3′
	R132H	5'-CTTCTTCGTTGGTCATGTTGGCGGCGTTG-3′ 5'-CAACGCCGCCAACATGACCAACGAAGAAG-3′
OmpF1/E, OmpF1Y/E, OmpF2/E	D113E	5′-GGGTTACACCGAGATGCTGCCAGAATTTGG-3′ 5′-CCAAATTCTGGCAGCATCTCGGTGTAACCC-3′
OmpF1H/E/R42S/Y106A	R42S	5′-GACATGACCTATGCCAGCCTTGGTTTTAAAG-3′ 5′-CTTTAAAACCAAGGCTGGCATAGGTCATGTC-3′
	Y106A	5′-GGTGTGGTTGCGGATGCACTGGGTTAC-3′ 5′-GTAACCCAGTGCATCCGCAACCACACC-3′
OmpF1Y/E/Y106R/R42E	Y106R	5′-GGTGTGGTTCGTGATGCACTGGGTTAC-3′ 5′-GTAACCCAGTGCATCACGAACCACACC-3′
	R42E	5′-GACATGACCTATGCCGAACTTGGTTTTAAAG-3′ 5′-CTTTAAAACCAAGTTCGGCATAGGTCATGTC-3′
OmpF1Y/E/R42H/Y106A	R42H	5′-GACATGACCTATGCCCATCTTGGTTTTAAAG-3′ 5′-CTTTAAAACCAAGATGGGCATAGGTCATGTC-3′
	Y106A	5′-GGTGTGGTTGCGGATGCACTGGGTTAC-3′ 5′-GTAACCCAGTGCATCCGCAACCACACC-3′
OmpF2	R82Y	5′-GTAACAAAACGTACCATGCATTCGCG-3′ 5′-CGCGAATGCATGGTACGTTTTGTTAC-3′
	Y106H	5′-GGTGTGGTTCACGATGCACTGGG-3′ 5′-CCCAGTGCATCGTGAACCACACC-3′
His-tag deletion (5’-phospholyration)		5′-CATATGGAAATCTATAACAAAGATGGCAAC-3′ 5′-TATATCTCCTTAAAGTTAAACAAAATTATTC-3′

2.4.4 Protein crystallization, structure determination, and refinement

Prior to crystallization, the protein samples were washed with a 50 mM Tris/HCl (pH 8.0) buffer containing 0.5% OG and 100 mM NaCl. $ZnCl_2$ solution was added to the resulting protein at the ratio of 1.2 equiv. of Zn ion to the protomer. The samples were used for crystallization using a sitting-drop method at 20 ºC, and the reservoir solution was a 0.1 mM sodium cacodylate (pH 8.0) buffer containing 43% PEG 200 and 0.12 M $MgCl_2$ as described previously[116]. The detailed information about crystallization, data collection, and refinement statistics are summarized in Table 2.2. The diffraction data were collected in 7A beamline at Pohang Accelerator Laboratory (PAL).

X-ray diffraction data were processed and scaled using the program suite HKL2000[117] and XDS[118.] All data were processed using CCP4, including Pointless, Matthews, and Scala5[119]. Molecular replacement was performed using either Molrep or Phaser[120-121] using the structure of the wild-type OmpF (PDB 2OMF or 2ZFG) as a search model. Rigid-body and restrained refinement using REFMAC5 and Phenix[122], along with manual model rebuilding and COOT[123], were used to obtain the structural models. Non-crystallographic symmetry restraints were also applied throughout the refinement[124]. Residues or side-chains with low electron density and B-factor (> 100 $Å^2$) were omitted. All the structural figures were produced using PyMOL. The geometric parameters of the Zn-binding sites in the OmpF1 and OmpF2 variants are summarized in Table 2.4.

2.4.5 *In vitro* esterase and β-lactamase activity assays

Chromogenic substrates, such as *p*-nitrophenyl acetate (pNPA) and nitrocefin, were purchased from Alfa Aesar and Cayman, respectively. The refolded OmpF variants at a final concentration of 10 μM in 50 mM Tris-HCl (pH 8) buffer were pre-mixed with 1.2 equiv. of Zn ions ($ZnCl_2$) to OmpF monomer on a 96-well plate. The reaction solution was prepared by mixing 90 μL of protein solution with 10 μL of the substrates dissolved in dimethyl sulfoxide (DMSO). The time-dependent optical changes were monitored at room temperature upon the addition of various concentrations of pNPA or nitrocefin at 410 or 486 nm, respectively, using a microplate reader (Biotek Synergy H1m). The kinetic data were collected using BioTek Gen5 software version 2.02 and analyzed with Origin 2021 (64-bit) 9.8.0.200 (academic). The net esterase activity of the Zn-dependent OmpF variants was determined by detecting the differences in the product formation rates in the presence and absence of Zn ions, in which free Zn ion exhibits no detectible activity (Figure 2.7). The net β-lactamase activity of the Zn-dependent OmpF variants was determined by detecting the differences in the product formation rates in the presence and absence of the OmpF variants, in which the apo-protein exhibits no detectible activity (Figure 2.8). The steady-state kinetic parameters were obtained via an iterative non-linear fit to a Michaelis-Menten equation or a linear fit using Origin 2021 (64-bit) 9.8.0.200 (academic).

2.4.6 Expression and extraction of the OmpF variants in a folded form

To bypass the protein refolding process, I expressed and extracted the OmpF variants in a folded form from the outer membrane of *E. coli*, as described previously[125]. For the translocation of the OmpF variants to the outer membrane, the native signal peptide (22 amino acids, MMKRNILAVIVPALLVAGTANA) was placed prior to the N-terminus of OmpF, resulting in plasmid, pET28b/kan^R/sig/ompf (Figure 2.11). For protein expression, *E. coli* BL21 (DE3) competent cells lacking LamB/OmpC/OmpF (ΔBCF, #102269 from Addgene)[126] were used. After overnight growth in LB medium, the cells were inoculated and grown in low salt LB medium with 50 mg/L kanamycin at 37 ºC in an orbital shaker at 150 rpm until the optical density (OD_{600}) reached 0.6. After the addition of IPTG at a final concentration of 0.1 mM and overnight incubation at 25 ºC, the cell pellets were harvested via centrifugation at 5,000 rpm and 4 ºC. Cell lysates were obtained via homogenization with a 50 mM Tris/HCl (pH 8.0) buffer, followed by centrifugation at 13,000 rpm (18,800 × *g*) and 4 ºC for 30 min. The pellets were resuspended and incubated with a 50 mM Tris/HCl (pH 8.0) buffer containing 2% (v/v) Triton X-100 for 2 h at 25 ºC. The solution was ultra-centrifuged (Optima L-100K) at 40,000 rpm (164,400 × *g*) for 1 h. The inner membrane fractions were removed by decanting the supernatants, and the remaining outer membrane fractions were incubated with a 50 mM Tris-HCl (pH 8.0) buffer containing 3% (w/v) OG at 25 ºC overnight. After ultracentrifugation, the supernatant was treated with 0.1 mg/mL trypsin at 37 ºC overnight, if necessary. The folding and oligomeric states of the proteins were analyzed using 15% SDS-PAGE analysis.

2.4.7 Initial glycosidase activity assays

As a fluorogenic substrate, 4-β-MUG (GoldBio) was added to the OmpF variants extracted directly from the cells (Figure 2.9b-c) when proteins (5–10 μM) in 50 mM Tris/HCl (pH 7.5) buffer containing 3% OG were pre-mixed with Zn ion ($ZnCl_2$, 20 μM) or the excess EDTA (1 mM). The excitation and emission wavelength were set as 370 and 450 nm, respectively, with a gain of 50 or 100. In addition, the consumption of 4-β-MUG and the subsequent formation of 4-methylumbelliferone was detected by HPLC (Agilent 1220 Infinity II), using Agilent ChemStation software version C.01.09. The reaction mixture (10 μL) was injected into an InfinityLab Proshell column (120 EC-C18 4.6 × 100 nm or 120 EC-C18 4.6 × 150 mm, 2.7 μm; particle size: 2.7 μm). A linear gradient of H_2O containing 0.05% trifluoroacetic acid (TFA) and CH_3CN containing 0.1% TFA solvents, from 9:1 to 0:10, was applied for 25 min, and the elution was monitored at 300 nm (Figure 2.10e).

2.4.8 Construction of mutant libraries

For sequence optimization, I selected two or four residues that are located in the vicinity of the Zn-binding sites: the R42, Y106, G120, and A123 residues of OmpF1/E; the Y106, R42, G120, and A123 residues of OmpF1Y/E; and Y106 and R82 residues of OmpF2/E. The mutant libraries were constructed via saturation mutagenesis using custom-designed primers containing NDT and VNK codons (N = A/G/C/T; D = A/G/T; V = A/C/G; K = G/T) for the selected sites, as shown previously[127] (Table 2.9). A combination of the NDT and VNK codons, instead of NNK, was used to exclude the stop codon and tryptophan. After transformation of the PCR products into *E. coli* DH5α competent cells, more than 100 colonies were pooled to generate a single-site randomized library with a 95% confidence level, as described previously[128].

Table 2.9 Custom-designed DNA primers for saturation mutagenesis.

Variants	Mutations	Primers
Y106X	NDT	5′-GGTGTGGTTNDTGATGCACTGGG-3′ 5′-CCCAGTGCATCAHNAACCACACC-3′
	VHG	5′-GGTGTGGTTVHGGATGCACTGG-3′ 5′-CCAGTGCATCCDBAACCACACC-3′
R42S	NDT	5′-GACATGACCTATGCCNDTCTTGGTTTTAAAG-3′ 5′-CTTTAA AACCAAGAHNGGCATAGGTCATGTC-3′
	VHG	5′-GACATGACCTATGCCVHGCTTGGTTTTAAAG-3′ 5′-CTTTAA AACCAAGCDBGGCATAGGTCATGTC-3′
R82X	NDT	5′-GGTAACAAAACGNDTCATGCATTCGCG-3′ 5′-CGCGAATGCATGAHNCGTTTTGTTACC-3′
	VHG	5′-GGTAACAAAACGVHGCATGCA TTC GCG-3′ 5′-CGCGAATGCATGCDBCGTTTT GTT ACC-3′
Y102X	NDT	5′-GATTACGGCCGTAACNDTGGTGTGG-3′ 5′-CCACACCAHNGTTACGGCCGTAATC-3′
	VHG	5′-GATTACGGCCGTAACVHGGGTGTGG-3′ 5′-CCACACCCDBGTTACGGCCGTAATC-3′
S125X	NDT	5′-GGTGGTGATACTGCATACNDTGATGACTTC-3′ 5′-GAAGTCATCAHNGTATGCAGTATCACCACC-3′
	VHG	5′-GGTGGTGATACTGCATACVHGGATGACTTC-3′ 5′-GAAGTCATCCDBGTATGCAGTATCACCACC-3′
G120X	NDT	5′-GAGATGCTGCCAGAATTTGGTNDTGATACT-3′ 5′-AGTATCAHNACCAAATTCTGGCAGCATCTC-3′
	VHG	5′-GAGATGCTGCCAGAATTTGGTVHGGATACT-3′ 5′-AGTATCCDBACCAAATTCTGGCAGCATCTC-3′
A123X	NDT	5′-GATACTNDTTACAGCGATGACTTCTTC-3′ 5′-GAAGAAGTCATCGCTGTAAHNAGTATC-3′
	VHG	5′-GATACTVHGTACAGCGATGACTTCTTC-3′ 5′-GAAGAAGTCATCGCTGTACDBAGTATC-3′
D113X	NDT	5′-GTTACACCNDTATGCTGCCAGAATTTGG-3′ 5′-CCAAATTCTGGCAGCATAHNGGTGTAAC-3′
	VHG	5′-GTTACACCVHGATGCTGCCAGAATTTGG-3′ 5′-CCAAATTCTGGCAGCATCDBGGTGTAAC-3′

2.4.9 Screening of mutant libraries

The mixture of plasmids containing a native leader sequence (pET28b/kan^R/leader/ompf) was transformed into *E. coli* BL21(DE3) ΔF (#102259, Addgene) and were grown in LB/agar plates containing 50 mg/L kanamycin at 37 ºC. The cultures were grown overnight in LB broth supplemented with kanamycin (50 mg/L) at 37 ºC in an orbital shaker at 150 rpm, and were inoculated in 100 mL of medium and grown until the OD_{600} reached 0.6. OmpF expression was induced by the addition of 0.1 mM IPTG and 50 μM $ZnCl_2$, followed by incubation for 16 h at 25 ºC. The cultures were spread on a M9/Agar plate containing 50 mg/L kanamycin, 0.2–0.4% cellobiose (Figure 2.12a), and 20 mg/L X-Glu (GoldBio) (Figure 2.12b), and inoculated for 1–2 days at 25 ºC. Blue colonies (Figure 2.12c) were selected and inoculated in 600 μL of LB medium containing 50 mg/L kanamycin on a 96 deep-well plate and grown at 37 ºC and 290 rpm (N-Biotek). After overnight incubation at 37 ºC, 540 μL of each culture was stored for DNA sequencing. The rest of the cultures were diluted with 540 μL of fresh LB and induced with 0.1 mM IPTG and 50 μM $ZnCl_2$ for 16 h at 25 ºC, until the OD_{600} reached 0.6. The catalytic activity for 4-β-MUG was measured at excitation and emission wavelengths of 370 and 450 nm, respectively, using a microplate reader (Biotek Synergy H1m). The measurements were carried out in duplicates. The colonies showing higher activity values than those obtained in the preceding round were regrown for isolating plasmid, sequencing, and further activity assays (Figure 2.14).

2.4.10 Whole-cell activity assays

The colonies from each round of screening showing the highest glycosidase activity were transformed into *E. coli* BL21(DE3) ΔBCF competent cells (#102269). Colonies were incubated in LB broth supplemented with kanamycin (50 mg/L) at 37 ºC for cell growth. The expression of the mutants was induced using 0.1 mM IPTG and 50 μM $ZnCl_2$ for 16 h at 25 ºC until an OD_{600} of 0.6 was obtained. After protein expression, the whole-cell activities were measured by mixing 45 μL of whole-cells with 5 μL of the substrates (4-α-MUG or 4-β-MUG at 10 mM dissolved in DMSO) (Figure 2.13c and Figure 2.14). Time-dependent fluorescence changes were monitored by using a microplate reader and a 384-well black plate (Corning 3575). The fluorescence changes were monitored at 450 nm upon excitation at 370 nm with a focal height of 8.5 mm and a gain of 50 or 100. The concentration values of the OmpF variants were quantified using densitometric analysis with SDS gels, where the protein samples were extracted from 100 mL of whole-cells using 1 mL of Tris/HCl (pH 8.0) buffer containing 3% OG. The normalized activity was determined by dividing the observed whole-cell activity by the protein concentration of the cell lysates (Figure 2.14).

2.4.11 Kinetic analysis of the evolved glycosidases under *in vitro* conditions

For kinetic analysis, the proteins extracted from whole-cells were further purified using an anion exchange column chromatography (HiTrap Q HP, GE Healthcare Life Sciences) with 50 mM Tris/HCl (pH 8.0) buffer containing 0.5% OG using a linear gradient of NaCl (0–1 M). For the preparation of a metal-free protein sample, a 5-fold molar excess of EDTA was added and incubated overnight 4 ºC. The excess EDTA was removed using a 30 kDa-cutoff centrifugal filter and a metal-free buffer. The reaction mixture was prepared by mixing 45 μL of protein pre-incubated with 1.2 equiv. of Zn ion ($ZnCl_2$) and 5 μL of 4-β-MUG (10 mM stock dissolved in DMSO).

Time-dependent fluorescence changes were monitored upon the addition of 4-β-MUG at various concentrations (0–20 mM) using a microplate reader (Biotek Synergy H1m) and a 384-well black plate (Corning 3575). The excitation and emission wavelengths were 370 and 450 nm, respectively, with a focal height of 8.5 mm and a gain of 100. Time-dependent fluorescence changes were measured and fit to a linear function to obtain the initial molar rates (Figure 2.15a and Table 2.5).

The rate constant of uncatalyzed hydrolysis (k_{uncat}) was measured by mixing 20 μL of 4-β-MUG (10 mM stock dissolved in DMSO) with 180 μL of 50 mM Tris (pH 7.5) buffer containing 3% OG at various temperature values (55–75 ºC) using a thermocycler (Bio-rad). Every 24 h, the formation of 4-methylumbelliferone was quantified using a microplate reader at excitation and emission wavelengths of 370 and 450 nm, respectively, with a gain of 58. The k_{uncat} value at 25 ºC was estimated by a linear fit to the Arrhenius equation.

2.4.12 ICP-MS analysis

The metal element and quantity of OmpF2/E-R2 were determined by ICP-MS. For sample preparation, the protein was extracted from whole-cells and purified using size exclusion chromatography (Superdex 200 10/300 GL) using 50 mM Tris-HCl (pH 8.0) buffer pre-treated with Chelex 100.

2.4.13 TON measurements

The turnover number (TON) was determined by mixing cell lysates (450 μL) containing 5–10 μM of the OmpF variants and 4-β-MUG (50 μL of 10 mM stock dissolved in DMSO). For the preparation of Zn-complexed and apo-proteins, $ZnCl_2$ and EDTA were added prior to the addition of 4-β-MUG. After incubation for 16 h, the concentration of 4-methylumbelliferone as the hydrolyzed product was quantified by using fluorescence and HPLC as described above. For fluorescence, the excitation and emission wavelength were set as 370 and 450 nm, respectively, with a gain of 50 or 100 (Figure 2.15c and Figure 2.18).

For the reactions with OG, Zn-complexed OmpF2/E-R2 was reacted for 48 h, and the solution was analyzed by HPLC and GC-MS (Trace 1310/ISQ LT single quadrupole mass spectrometer, Thermo Scientific) to detect the formation of glucose and 1-octanol, respectively. The GC-MS data were collected using Xcalibur software version 4.0.0.29. For pH-dependent assays, TONs were measured with 4-β-MUG by using the following buffers; 50 mM Bis-Tris (pH 6.0–6.5) and 50 mM Tris (pH 7.0–9.0).

2.4.14 Inhibitor assays

CBE (Carbosynth)[106-107] at a final concentration of 500 μM in ddH_2O was added to the extracted proteins at the final concentration of 10 μM. After incubation for 4 h at 25 ºC, 50 μL of 4-β-MUG (10 mM stock solution in DMSO) was added to the cell lysates (450 μL), and TON values were analyzed as described above.

2.4.15 Tandem LC/MS analysis

The extracted proteins were pre-mixed with or without Zn ions and treated overnight with the covalent inhibitor CBE at a 200-fold molar ratio to the proteins. The samples were heated at 70 ºC for 1.5 h using a thermocycler and treated with trypsin at 37 ºC overnight. After the SDS-PAGE experiment, the bands corresponding to the target proteins were excised and treated with DTT (10 mM) and iodoacetamide at the final concentrations of 10 mM and 55 mM, respectively. After digestion with trypsin/chymotrypsin (Thermo Fisher Scientific) and desalting using ZipTip C18 (Millipore), the resulting peptides were resuspended in 0.1% (v/v) formic acid (Merck) and analyzed using LC-MS/MS (Q Exactive Hybrid Quadrupole-Orbitrap, Thermo Fisher Scientific) and Acclaim PepMap 100 trap columns (100 μm × 2 cm, nanoViper C18, 5 μm, 100 Å or 75 μm × 15 cm, nanoViper C18, 3 μm, 100 Å), as published previously[129.] The injected samples were washed for 6 min with 98% solvent A (water and CH_3CN (98:2 v/v) and 0.1% formic acid) at a flow rate of 4 μL/min or 300 nL/min, respectively using the following elution program; a linear gradient of 2–40% solvent B for 45 min, a gradient of 40–95% for 5 min, elution with 95% solvent B (100% CH_3CN and 0.1% formic acid) for 10 min, and 2% solvent B for 20 min. The raw data were collected using Xcalibur software version 4.3. Subsequently, the raw data were processed using

Proteome Discoverer 2.3 (Thermo Fisher Scientific) using the sequence of OmpF (UniProt P02931) and the OmpF variants. Methionine oxidation, cysteine carbamidomethylation, and the covalent linkage of a CBE molecule to glutamate, aspartate, or Zn-hydroxide bound to 3 His or a 2 His/1 Glu site were included as specified modifications. A strict (1%) and a relaxed (5%) protein false discovery rates were applied using peptide spectrum matches validator node in Proteome Discoverer (Figure 2.17 and Table 2.7). Of note, only the fragments containing Zn-CBE moiety exhibited the distribution of natural zinc isotopes.

2.4.16 Characterization of evolved glycosidases

Activity assays were performed with OmpF2/E-R1* variant, which showed 14.3% of TON relative to OmpF2/E-R2, in the presence of sodium azide[96, 130]. After mixing 800 µL of 20 µM OmpF variants, 100 µL of 100 mM 4-β-MUG dissolved in DMSO, and 100 µL of 1 M sodium azide (Figure 2.23a), the formation of glycosyl azide was detected using LC-MS (Agilent Technologies 1200 Infinity Series equipped with Agilent Technologies 6120 Quadrupole LC/MS) using Agilent ChemStation software version B.04.03.

For structural characterization, glycosyl azide was purified by injecting the reaction mixture (4 mL in total) into an HPLC ZORBAX SB-C18 Semi-Preparative 9.4 × 250 mm, 5 µm column in a HPLC system (Agilent 1260 Infinity II). A linear gradient of two solvents, H_2O containing 0.05% TFA and CH_3CN containing 0.1% TFA, was applied at a ratio of 9:1 to 0:10 over 30 min at 2 mL/min. The eluted glycosyl azide was freeze-dried and subsequently reacted with Cu(I) (0.01 equiv.), 1,2-dimethylimidazole (0.01 equiv.), and phenylacetylene (~1.05 equiv.) for Cu-catalyzed azide-alkyne cycloaddition (CuAAC)[131].

The resulting solution was purified using HPLC with a ZORBAX SB-C18 Semi-Preparative 9.4 × 250 mm, 5 μm column, and freeze-dried for the LC-MS (ES-API) experiment, calculated for $C_{14}H_{17}N_3O_5^+$ $[M+H]^+$: m/z = 308.12, observed: 308.1 (Figure 2.23b) and ^{1}H NMR (Agilent) (Figure 2.23c). ^{1}H NMR (CD_3OD, 400 MHz) δ 8.57 (s, 1H), 7.87 (m, 2H), 7.46 (m, 2H), 7.37 (m, 1H), 5.65 (d, 1H), 3.99–3.90 (m, 2H), 3.74 (dd, 1H), and 3.64–3.50 (m, 3H) (Figure 2.18c, top). As a reference, 1-azido-1-deoxy-glucose (Synthose) was purchased and reacted with Cu(I) and phenylacetylene as described above to synthesize 1-β-D-glucopyranosyl-4-phenyl-1H-1,2,3-triazole (Figure 2.18c, bottom). The NMR data were acquired using VnmrJ software version 4.2 and analyzed using MestReNova x64-14.2.0.

2.4.17 Docking simulation using Autodock Vina

I obtained a substrate coordinate file from the PubChem database as reported previously[132]. Then, I manually mutated the X-ray crystal structures of OmpF1 and OmpF2 to account for the mutations acquired from rational redesign and directed evolution using the PyMOL program. After setting these residues to be flexible, I applied the files to the Autodock Vina software version 1.1.2 for docking simulation[133] to simulate and visualize the potential binding poses of 4-β-MUG in the evolved OmpF variants (Figure 2.25).

CHAPTER 3

Expanding the design strategy of artificial Zn-dependent metalloenzymes: embracing the full potential of protein structure

3.1 Introduction

In Chapter 1, I addressed the current state of the art and limitations in the design of tailor-made artificial metalloenzymes. There is an increasing demand for a more accurate and widely applicable design method that can be applied to diverse protein scaffolds, regardless of the presence of pre-existing metal-binding sites and native secondary or tertiary structures. To resolve this limitation and expand the scope of artificial metalloenzymes, I have developed and applied a new geometric analysis-based designing method to a β-barrel outer membrane protein F (OmpF), introducing a tetrahedral mononuclear Zn-binding site, as described in Chapter 2. X-ray crystallographic studies of OmpF variants demonstrated that out of six residues, three each metal-ligating site predicted for two OmpF variants, five amino acids correctly constitute for the Zn-coordinating first coordination spheres, validating the accuracy and versatility of our method.

However, these results also suggest that there is room for improvement; a glutamate instead of histidine residue was ligated to Zn ion, forming 2His/1Glu moiety, instead of 3His residues, for Zn-coordination. Thus, I report more precise design method in Chapter 3. To extract more precise geometric parameters for Zn-ligation sites, I scrutinized all zinc-binding sites in hydrolases available in the RCSB database. Then, I applied more specific geometric filters to sort Zn-binding proteins depending on the coordination number, composition of metal-binding residues, and the distribution of interatomic distances. Finally, I applied the revised Zn-site installing method, resulting in Zinc-binding metalloproteins. They were further characterized their structures by X-ray crystallography, demonstrating that this method shows improved accuracy and versatility in metalloprotein design.

3.2 Results

3.2.1 Design strategy

I analyzed the geometric parameters of Zn-hydrolases using the following procedure. First, I carefully curated a collection of Zn-hydrolase structures from the RCSB database. Then, I filtered the structures based on the following restraints, resolution and structure determination methods. Specifically, I selected Zn-hydrolase structures with crystallographic resolutions up to 3.0 Å. I focused on Zn-binding sites, exhibiting a coordination number of 3 or 4, the most frequently observed coordination geometry of zinc-hydrolase, resulting in a subset of structures suitable for further analysis.[134-136] To remove redundancy in the data set, I performed a sequence percent identity analysis to remove redundant sequences.[137] This step allowed us to avoid duplications and ensure the dataset represented comprehensive sequences of Zn-containing hydrolases.

With the non-redundant sequences of Zn-hydrolase, the next step involved calculating the distances between the C_α and C_β atoms of the Zn-ligating residues in the selected structures. I focused on analyzing the rigid backbone rather than the side chains because the distance between backbone carbon atoms remains unchanged regardless of whether they bind to the metal or not. All possible combinations of residues involved in zinc coordination were thoroughly examined, and the distances between them were measured (Figure 2.1). As a result, I gained valuable insights into the geometric parameters characterizing the Zn-binding sites in Zn-binding hydrolases by analyzing the distances.

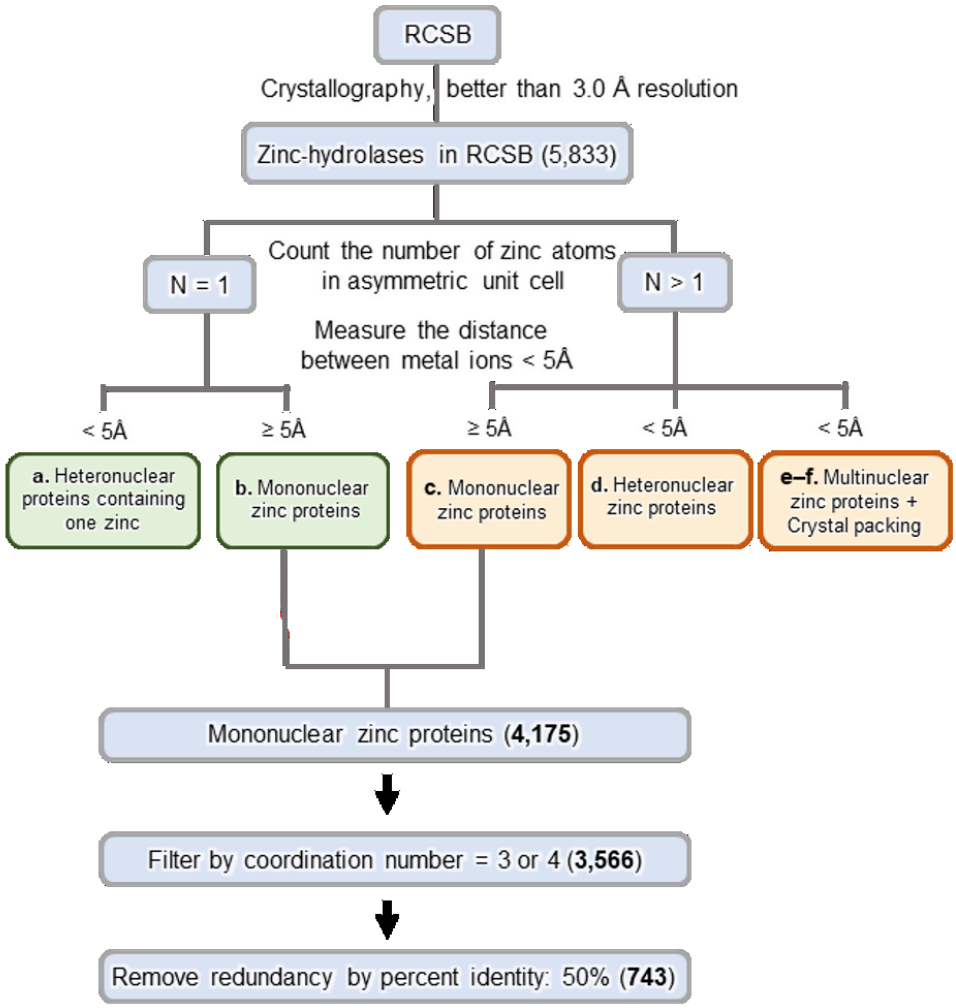


Figure 3.1 Flow chart for selection of mononuclear zinc proteins.

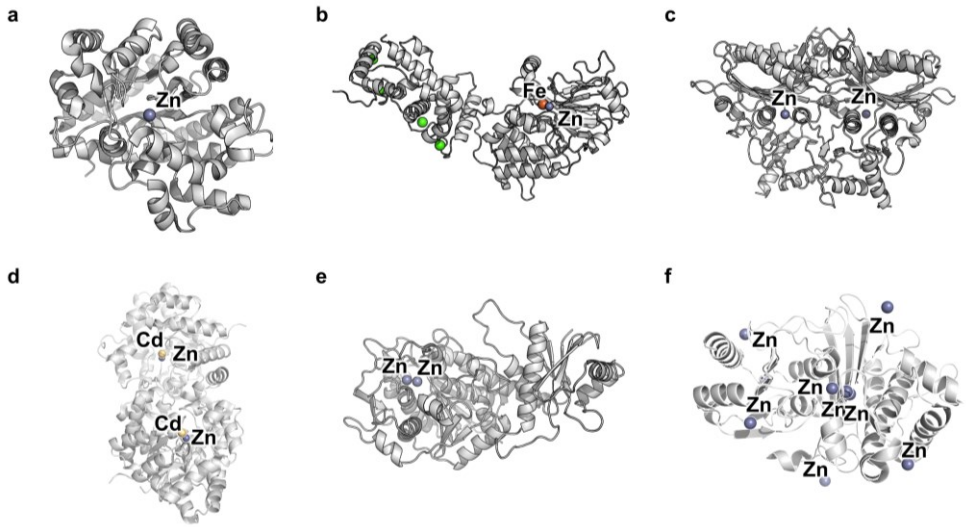


Figure 3.2 Representative crystal structure of zinc-binding proteins.

a–f. Crystal structure of catrgorized zinc-binding proteins. **a,** Heteronuclear proteins containing one zinc (PDB 1AUI). **b,** Mononuclear zinc proteins (PDB 1ADD). **c**, Mononuclear zinc proteins (PDB 1EYK). **d**, Heteronuclear zinc proteins (PDB:1I0D). **e,** Multinuclear zinc proteins (PDB 1BLL).

f, Crystal packing (PDB 1K9Z). Zn atoms and Fe atoms and Cd atoms are shown with light navy, orange and yellow, respectively.

3.2.2 Analysis of geometric parameters

I obtained geometric data from protein structures and analyzed various parameters based on residue combinations. I excluded the cysteine-containing combinations due to their potential for disulfide bond formation and protein instability. I classified the residue combinations based on the presence or absence of glutamate. This categorization revealed several zinc-binding motifs: 1His/2Asp, 2His/1Asp, and 3His motifs for coordination number 3 without glutamate. Additionally, the identified motifs for coordination number 4 without glutamate were 3His/1Asp and 2His/2Asp. Furthermore, I examined the minimum and maximum distances within the combinations to evaluate the range of potential metal binding configurations. My analysis revealed that the combinations with glutamate exhibited a narrower range of minimum and maximum distances of both C_α and C_β distances in both coordination numbers, suggesting the importance of finely tuning the side chain orientation of longer carbon chains containing glutamate for optimal metal binding. (Figure 3.3). Interestingly, combinations of three residues showed a narrower range of distances than those of four, indicating potential structural constraints (Figure 3.4). To achieve my objective of creating a coordinatively unsaturated tetrahedral site for catalysis, I focused on examining the geometric parameters derived from three specific residue combinations

Furthermore, I performed additional analysis by categorizing the C_α and C_β distances into minimum, median, and maximum distances. This categorization allowed me to apply the distance range for each distance, enabling more accurate sorting for the target residue (Figure 3.5). I applied three different distance parameters to investigate the catalytic activities of the β-barrel-shaped outer membrane protein F (OmpF). From a large dataset of

1.7×10^8 samples, I identified 20,000 datasets that met my predetermined threshold for the non-glutamate parameter and 13,000 for the glutamate parameter. Subsequently, as we did in Chapter 2, I focused on the constriction zone and mutated several residues that met the threshold for installing zinc-binding sites. Finally, we determined the crystal structure of OmpF3 and OmpF3/71E, which have a different first coordination sphere from OmpF1 and OmpF2. We observed 2His/1Asp metal binding motifs from OmpF3 and OmpF3/71E, but two variants have different combinations of residues, such as R132D/Y102H/L83H and R132D/Y102/R82H, respectively. (Figure 3.5). Additionally, all glutamate and aspartate residues acted as monodentate ligands. Notably, the location of the catalytic site in OmpF1 differed from that in the other variants (Figure 3.6). It is intriguing to observe that even a minor difference in mutation between OmpF3 and OmpF3/71E, specifically the substitution of single amino acid residue, E71'and E71'H, resulted in apparent changes in location and first coordination sphere. This data highlights that slight differences in protein environment can severely affect metal coordination events.

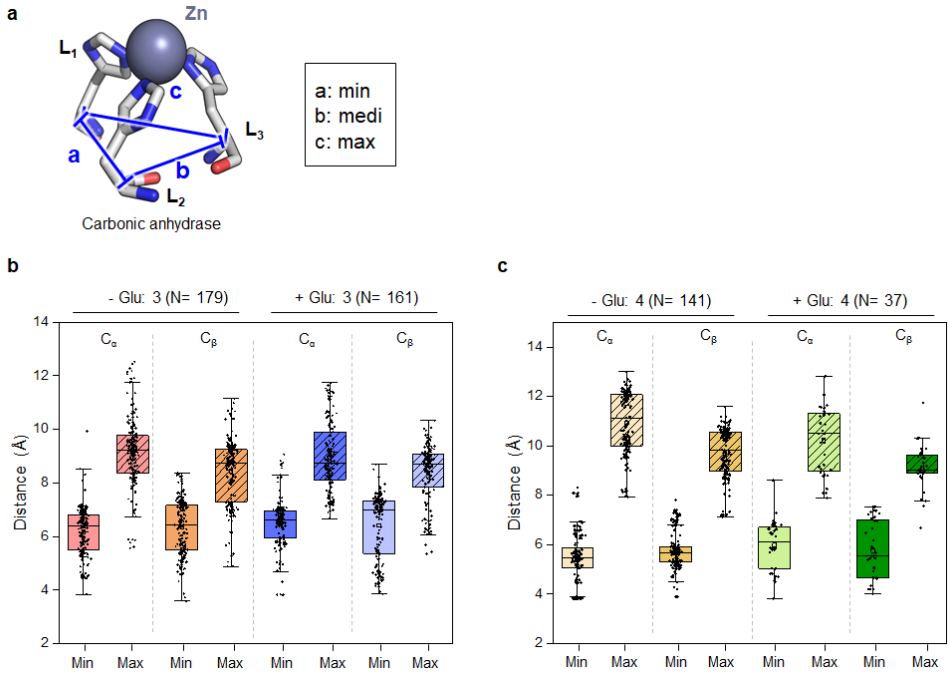


Figure 3.3 Cα and Cβ distance analysis by coordination number

a, Cα distance between residues. a: minimum distance, b: median distance, and c: maximum distance. **b,** Distance distribution of zinc-containing hydrolases with coordination number 3. **c,** Distance distribution of zinc-containing hydrolases with coordination number 4. Cα distance between -Glu: 3 colored in pink, Cβ distance between -Glu: 3 colored in orange, Cα distance between +Glu: 3 colored in blue, Cβ distance between GLUE-3 colored in sky blue, Cα distance between -Glu: 4 colored in wheat, Cβ distance between -Glu: 4 colored in light orange, Cα distance between +Glu: 4 colored in light green, and Cβ distance between +Glu: 4 colored in green. Maximum distance is marked with slashes.

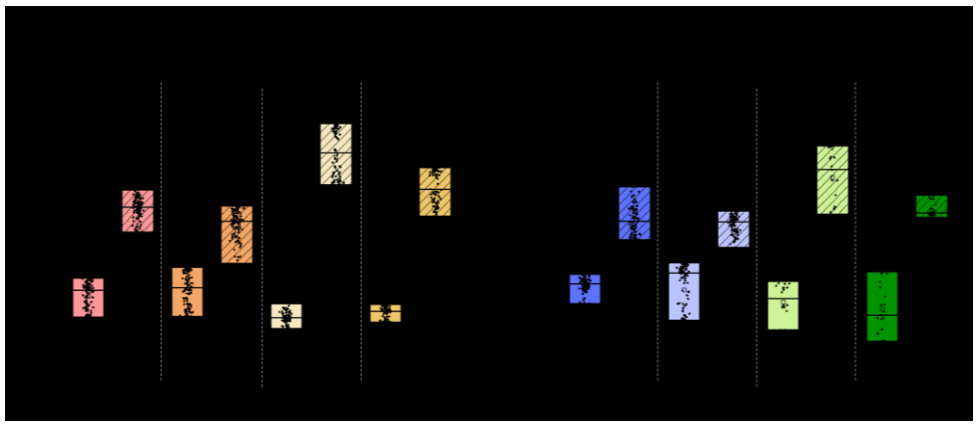
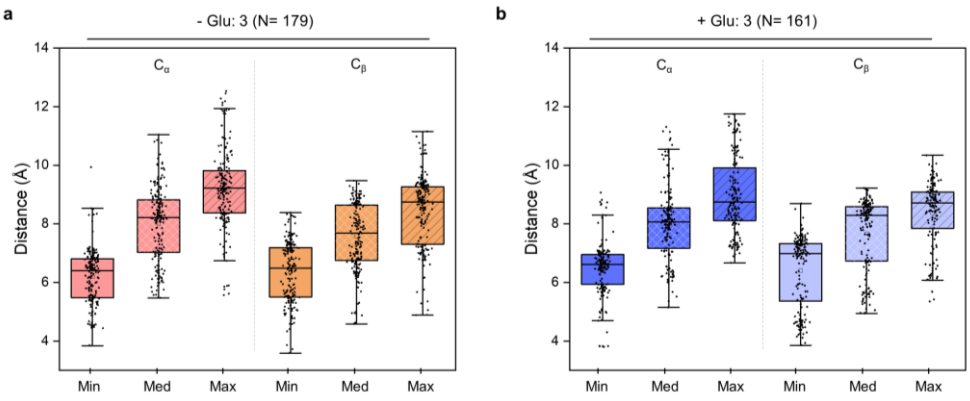


Figure 3.4 C_α and C_β distance analysis by composition

a, Distance distribution of zinc-containing hydrolases without glutamate residue. **b,** Distance distribution of zinc-containing hydrolases with coordination number 4. C_α distance between -Glu: 3 colored in pink, C_β distance between -Glu: 3 colored in orange, C_α distance between +Glu: 3 colored in blue, C_β distance between GLUE-3 colored in sky blue, C_α distance between -Glu: 4 colored in wheat, C_β distance between -Glu: 4 colored in light orange, C_α distance between +Glu: 4 colored in light green, and C_β distance between +Glu: 4 colored in green. Maximum distance is marked with slashes.



	C_α distance (Å)		C_β distance (Å)	
Min.	3.84	9.93	3.59	8.38
Med.	5.47	11.05	4.59	9.47
Max.	5.57	12.53	4.88	11.15

	C_α distance (Å)		C_β distance (Å)	
Min.	3.80	9.07	3.85	8.69
Med.	5.15	11.31	4.94	9.22
Max.	6.67	11.75	5.36	10.34

Figure 3.5 C_α and C_β distance analysis of coordination number: 3

Distance analysis of -Glu: 3 and +Glu: 3. **a,** Distance distribution of components in -Glu: 3. **b,** Distance distribution of components in +Glu: 3. Maximum distance is marked with slashes. The ranges of distances are shown in below table. All distances between three residues are shown; Min: minimum, Med: median, and Max: maximum distances, respectively. The median distance is marked with white slashes on both sides, and the maximum distance is marked with slashes.

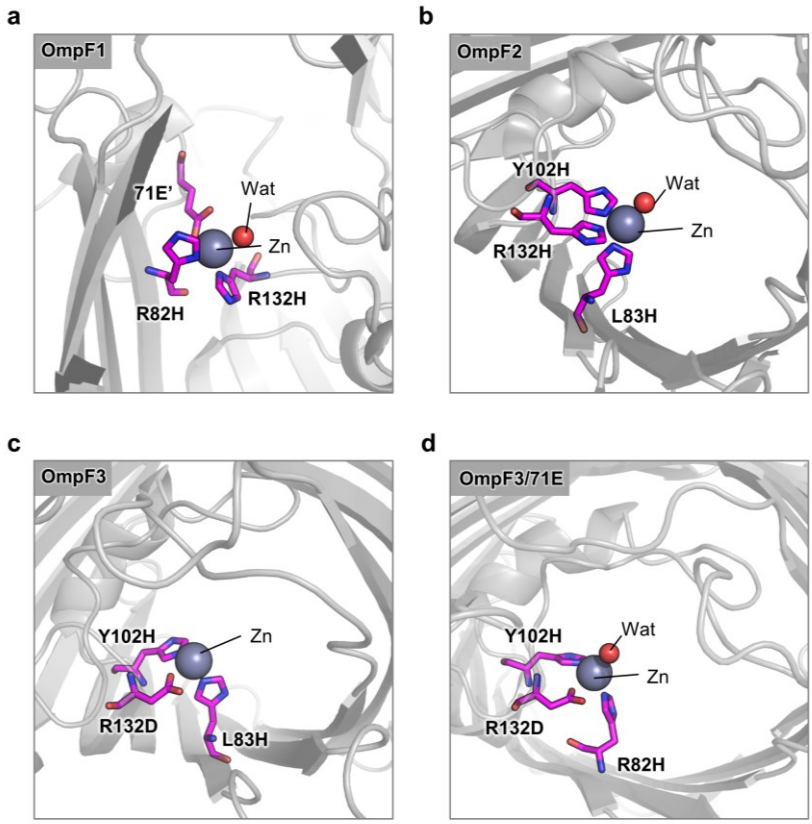


Figure 3.6 The X-ray crystal structures of the OmpF variants

The active sites of **a,** OmpF1, **b,** OmpF2, **c,** OmpF3 and **d,** OmpF3/71E. Zn atoms and Zn-bound water molecules (Wat) are shown with light navy and red spheres, respectively. Metal-ligating residues are colored in magenta.

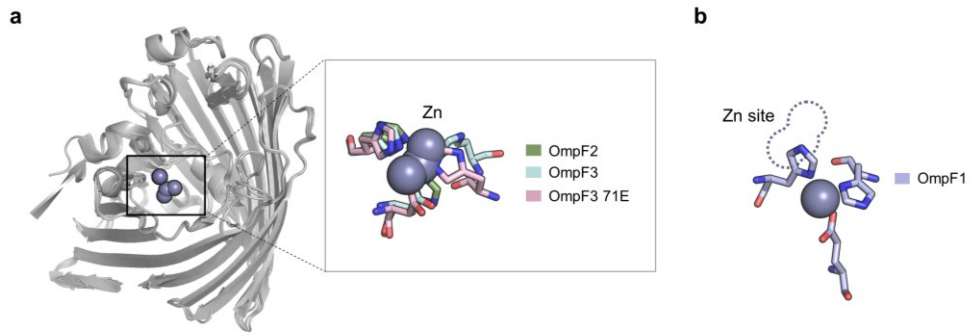


Figure 3.7 Structural analysis of the first coordination sphere of OmpF variants

a, The overlaid structures from the extracellular side. **b,** The relative location of the Zn-binding site in OmpF1 relative to other variants. The dotted-circle indicate the location of the active sites in OmpF2, OmpF3, and OmpF 71E mutants.

3.3 Discussion

This work introduced a novel method to design artificial metalloenzymes that are applicable to various protein scaffolds, irrespective of the secondary structures and protein topologies. Through the curation and analysis of a comprehensive set of Zn-binding proteins, we have gained valuable insights into the characteristics of Zn-binding motifs. I successfully converged non zinc-binding proteins into zinc-binding proteins based on this method.

My study highlights the importance of understanding the geometric parameters of Zn-binding motifs to converge non-zinc-binding proteins into zinc-binding ones successfully. This methodology demonstrates its applicability to diverse protein scaffolds, expanding the pool of potential protein scaffolds.

Furthermore, my findings hold substantial implications for the field of metalloenzymes, broadening the possibilities for tailored functionalities by utilizing the various inherent properties of proteins.

3.4 Materials and methods

3.4.1 Design of triad mutants

The constriction zone was inspected to install a Zn-binding motif. I first narrowed down the candidate positions into the following: Y102, Y106, D113, E117, and R132, due to the orientation of their side-chains, which pointed towards the void space of the cylindrical protein. The residues were prioritized depending on their secondary structures and local flexibility. R132 was the most desirable site as the first vertex of a Zn-binding triad, and the other two vertexes were determined based on the distance obtained from calculation. As a result, four sets of triads, OmpF1, OmpF2, OmpF3 and OmpF3/71E comprising R132H/Y102H/R82H, R132H/Y102H/L83H, R132D/Y102H/L83H/R82H/71H, and R132D/Y102H/L83H/R82H respectively, were determined.

3.4.2 Expression and extraction of the OmpF variants in a folded form

To bypass the protein refolding process, I expressed and extracted the OmpF variants in a folded form from the outer membrane of *E. coli*, as described previously[129]. For the translocation of the OmpF variants to the outer membrane, the native signal peptide (22 amino acids, MMKRNILAVIVPALLVAGTANA) was placed prior to the N-terminus of OmpF. Furthermore, we attached the TEV-protease cleavage site with 8 histidines between signal peptide and mature protein for purification of proteins by His-tag binding affinity, resulting in plasmid, pET28b/kan^{R}/sig/TEV-protease cleavage site/8xHis-Tag/OmpF. For protein expression, *E. coli* BL21 (DE3) competent cells lacking LamB/OmpC/OmpF (ΔBCF, #102269 from Addgene)[126] were used. After overnight growth in LB

medium, the cells were inoculated and grown in low salt LB medium with 50 mg/L kanamycin at 37 ºC in an orbital shaker at 150 rpm until the optical density (OD_{600}) reached 0.6. After the addition of IPTG at a final concentration of 0.2 mM and overnight incubation at 25 ºC, the cell pellets were harvested via centrifugation at 5,000 rpm and 4 ºC. Cell lysates were obtained via homogenization with a 0.1mM NaCl 20 mM Tris/HCl (pH 8.0) buffer, followed by centrifugation at 13,000 rpm (18,800 × *g*) and 4 ºC for 30 min. The pellets were resuspended and incubated with a 50 mM Tris/HCl (pH 8.0) buffer containing 2% (v/v) Triton X-100 for 2 h at 25 ºC. The solution was ultra-centrifuged (Optima L-100K) at 40,000 rpm (164,400 × *g*) for 1 h. The inner membrane fractions were removed by decanting the supernatants, and the remaining outer membrane fractions were incubated with a 50 mM Tris-HCl (pH 8.0) buffer containing 3% (w/v) OG at 25 ºC overnight. After ultracentrifugation, the supernatant was loaded on a His-Trap column (HisTrap HP, GE Healthcare Life Sciences) equilibrated with 0.8 % n-Octyl-oligo-oxyethylene (octyl-POE) (Santa Cruz Biotechnoly) using a protein purification system (ÄKTA Prime Plus). The OmpF protein was eluted by applying a linear gradient of 1 M imidazole, and the purity of the proteins was analyzed using 12% SDS-PAGE (Bio-rad). The pure fractions were concentrated at 4 ºC using a 50 kDa cutoff centrifugal filter (Amicon). The protein was desalted at final concentration 20 mM Tris-HCl, 0.1 M NaCl, and 0.8 % octyl-POE. Incubated with TEV-protease during overnight. The cleavage of TEV site were analyzed using 12% SDS-PAGE (Bio-rad) and fully cleaved protein was further purified by size exclusion chromatography (SEC).

3.4.3 Protein crystallization, structure determination, and refinement

Prior to crystallization, the protein samples were washed with a 20 mM Tris/HCl (pH 8.0) buffer containing 0.8% octyl-POE and 100 mM NaCl. $ZnCl_2$ solution was added to the resulting protein at the ratio of 1.2 equiv. of Zn ion to the protomer. The samples were used for crystallization using a sitting-drop method at 20 ºC, and the reservoir solution was a 0.3 M Magnesium chloride hexahydrate, 0.1 M BICINE 9.0, and 28 % w/v PEG 2000. The diffraction data were collected in 7A beamline at Pohang Accelerator Laboratory (PAL).

X-ray diffraction data were processed and scaled using the program suite HKL2000[117] and XDS[118]. All data were processed using CCP4, including Pointless, Matthews, and Scala5[119]. Molecular replacement was performed using either Molrep or Phaser[120-121] using the structure of the wild-type OmpF (PDB 2OMF or 2ZFG) as a search model. Rigid-body and restrained refinement using REFMAC5 and Phenix[122], along with manual model rebuilding and COOT[123], were used to obtain the structural models. Non-crystallographic symmetry restraints were also applied throughout the refinement[128]. Residues or side-chains with low electron density and B-factor (> 100 $Å^2$) were omitted. All the structural figures were produced using PyMOL.

3.5 References

1. Lippard, S. J.; Berg, J. M., *Principles of Bioinorganic Chemistry*. University Science Books: 1994.
2. Gray, H. B.; Stiefel, E. I.; Valentine, J. S.; Bertini, I., *Biological Inorganic Chemistry: Structure and Reactivity*. University Science Book: 2006.
3. Bornscheuer, U. T.; Huisman, G. W.; Kazlauskas, R. J.; Lutz, S.; Moore, J. C.; Robins, K., Engineering the third wave of biocatalysis. *Nature* **2012,** *485* (7397), 185-194.
4. Nanda, V.; Koder, R. L., Designing artificial enzymes by intuition and computation. *Nat. Chem.* **2010,** *2* (1), 15-24.
5. Schwizer, F.; Okamoto, Y.; Heinisch, T.; Gu, Y.; Pellizzoni, M. M.; Lebrun, V.; Reuter, R.; Köhler, V.; Lewis, J. C.; Ward, T. R., Artificial metalloenzymes: reaction scope and optimization strategies. *Chem. Rev.* **2018,** *118* (1), 142-231.
6. Yu, F.; Cangelosi, V. M.; Zastrow, M. L.; Tegoni, M.; Plegaria, J. S.; Tebo, A. G.; Mocny, C. S.; Ruckthong, L.; Qayyum, H.; Pecoraro, V. L., Protein design: toward functional metalloenzymes. *Chem. Rev.* **2014,** *114* (7), 3495-3578.
7. Nastri, F.; Chino, M.; Maglio, O.; Bhagi-Damodaran, A.; Lu, Y.; Lombardi, A., Design and engineering of artificial oxygen-activating metalloenzymes. *Chem. Soc. Rev.* **2016,** *45* (18), 5020-5054.
8. Liu, J.; Meier, K. K.; Tian, S.; Zhang, J. L.; Guo, H.; Schulz, C. E.; Robinson, H.; Nilges, M. J.; Münck, E.; Lu, Y., Redesigning the blue copper azurin into a redox-active mononuclear nonheme iron protein: preparation and study of Fe(II)-M121E azurin. *J. Am. Chem. Soc.* **2014,** *136* (35), 12337-44.

9. Slater, J. W.; Shafaat, H. S., Nickel-Substituted Rubredoxin as a Minimal Enzyme Model for Hydrogenase. *J. Phys. Chem. Lett.* **2015,** *6* (18), 3731-3736.
10. Doble, M. V.; Jarvis, A. G.; Ward, A. C. C.; Colburn, J. D.; Götze, J. P.; Bühl, M.; Kamer, P. C. J., Artificial Metalloenzymes as Catalysts for Oxidative Lignin Degradation. *ACS Sustain. Chem. Eng.* **2018,** *6* (11), 15100-15107.
11. Deuss, P. J.; Popa, G.; Slawin, A. M. Z.; Laan, W.; Kamer, P. C. J., Artificial Copper Enzymes for Asymmetric Diels–Alder Reactions. *ChemCatChem* **2013,** *5* (5), 1184-1191.
12. Köhler, V.; Wilson, Y. M.; Dürrenberger, M.; Ghislieri, D.; Churakova, E.; Quinto, T.; Knörr, L.; Häussinger, D.; Hollmann, F.; Turner, N. J.; Ward, T. R., Synthetic cascades are enabled by combining biocatalysts with artificial metalloenzymes. *Nat. Chem.* **2013,** *5* (2), 93-99.
13. Chatterjee, A.; Mallin, H.; Klehr, J.; Vallapurackal, J.; Finke, A. D.; Vera, L.; Marsh, M.; Ward, T. R., An enantioselective artificial Suzukiase based on the biotin–streptavidin technology. *Chem. Sci.* **2016,** *7* (1), 673-677.
14. Jeschek, M.; Reuter, R.; Heinisch, T.; Trindler, C.; Klehr, J.; Panke, S.; Ward, T. R., Directed evolution of artificial metalloenzymes for in vivo metathesis. *Nature* **2016,** *537* (7622), 661-665.
15. Wu, S.; Zhou, Y.; Rebelein, J. G.; Kuhn, M.; Mallin, H.; Zhao, J.; Igareta, N. V.; Ward, T. R., Breaking symmetry: engineering single-chain dimeric streptavidin as host for artificial metalloenzymes. *J. Am. Chem. Soc.* **2019,** *141* (40), 15869-15878.
16. Zozulia, O.; Dolan, M. A.; Korendovych, I. V., Catalytic peptide assemblies. *Chem. Soc. Rev.* **2018,** *47* (10), 3621-3639.

17. Pinter, T. B. J.; Koebke, K. J.; Pecoraro, V. L., Catalysis and electron transfer in *de novo* designed helical scaffolds. *Angew. Chem. Int. Ed. 59*, 2-24.
18. Shiga, D.; Funahashi, Y.; Masuda, H.; Kikuchi, A.; Noda, M.; Uchiyama, S.; Fukui, K.; Kanaori, K.; Tajima, K.; Takano, Y.; Nakamura, H.; Kamei, M.; Tanaka, T., Creation of a binuclear purple copper site within a *de novo* coiled-coil protein. *Biochemistry* **2012,** *51* (40), 7901-7907.
19. Sambasivan, R.; Ball, Z. T., Metallopeptides for asymmetric dirhodium catalysis. *J. Am. Chem. Soc.* **2010,** *132* (27), 9289-9291.
20. Der, B. S.; Machius, M.; Miley, M. J.; Mills, J. L.; Szyperski, T.; Kuhlman, B., Metal-mediated affinity and orientation specificity in a computationally designed protein homodimer. *J. Am. Chem. Soc.* **2012,** *134* (1), 375-385.
21. Chalkley, M. J.; Mann, S. I.; DeGrado, W. F., *De novo* metalloprotein design. *Nat. Rev. Chem.* **2022,** *6* (1), 31-50.
22. Valdez, C. E.; Smith, Q. A.; Nechay, M. R.; Alexandrova, A. N., Mysteries of metals in metalloenzymes. *Acc. Chem. Res.* **2014,** *47* (10), 3110-3117.
23. Eom, H.; Song, W. J., Emergence of metal selectivity and promiscuity in metalloenzymes. *J. Biol. Inorg. Chem.* **2019,** *24* (4), 517-531.
24. Pordea, A., Metal-binding promiscuity in artificial metalloenzyme design. *Curr. Opin. Chem. Biol.* **2015,** *25*, 124-132.
25. Lin, Y.-W., Rational design of metalloenzymes: From single to multiple active sites. *Coord. Chem. Rev.* **2017,** *336*, 1-27.
26. O'Brien, P. J.; Herschlag, D., Catalytic promiscuity and the evolution of new enzymatic activities. *Chem. Biol.* **1999,** *6* (4), R91-R105.
27. Nobeli, I.; Favia, A. D.; Thornton, J. M., Protein promiscuity and its

implications for biotechnology. *Nat. Biotechnol.* **2009,** *27* (2), 157-167.

28. Khersonsky, O.; Tawfik, D. S., Enzyme promiscuity: A mechanistic and evolutionary perspective. *Annu. Rev. Biochem.* **2010,** *79* (1), 471-505.
29. Putignano, V.; Rosato, A.; Banci, L.; Andreini, C., MetalPDB in 2018: a database of metal sites in biological macromolecular structures. *Nucleic Acids Res.* **2018,** *46* (D1), D459-d464.
30. Fujieda, N.; Nakano, T.; Taniguchi, Y.; Ichihashi, H.; Sugimoto, H.; Morimoto, Y.; Nishikawa, Y.; Kurisu, G.; Itoh, S., A well-defined osmium–cupin complex: Hyperstable artificial osmium peroxygenase. *J. Am. Chem. Soc.* **2017,** *139* (14), 5149-5155.
31. Mann, S. I.; Heinisch, T.; Ward, T. R.; Borovik, A. S., Peroxide activation regulated by hydrogen bonds within artificial Cu proteins. *J. Am. Chem. Soc.* **2017,** *139* (48), 17289-17292.
32. Raines, D. J.; Clarke, J. E.; Blagova, E. V.; Dodson, E. J.; Wilson, K. S.; Duhme-Klair, A.-K., Redox-switchable siderophore anchor enables reversible artificial metalloenzyme assembly. *Nat. Catal.* **2018,** *1* (9), 680-688.
33. Cherrier, M. V.; Martin, L.; Cavazza, C.; Jacquamet, L.; Lemaire, D.; Gaillard, J.; Fontecilla-Camps, J. C., Crystallographic and spectroscopic evidence for high affinity binding of FeEDTA(H2O)- to the periplasmic nickel transporter NikA. *J. Am. Chem. Soc.* **2005,** *127* (28), 10075-10082.
34. Esmieu, C.; Cherrier, M. V.; Amara, P.; Girgenti, E.; Marchi-Delapierre, C.; Oddon, F.; Iannello, M.; Jorge-Robin, A.; Cavazza, C.; Ménage, S., An artificial oxygenase built from scratch: Substrate binding site identified using a docking approach. *Angew. Chem. Int. Ed.* **2013,** *52* (14), 3922-3925.
35. Monnard, F. W.; Nogueira, E. S.; Heinisch, T.; Schirmer, T.; Ward, T. R.,

Human carbonic anhydrase II as host protein for the creation of artificial metalloenzymes: the asymmetric transfer hydrogenation of imines. *Chem. Sci.* **2013,** *4* (8), 3269-3274.

36. Eda, S.; Nasibullin, I.; Vong, K.; Kudo, N.; Yoshida, M.; Kurbangalieva, A.; Tanaka, K., Biocompatibility and therapeutic potential of glycosylated albumin artificial metalloenzymes. *Nat. Catal.* **2019,** *2* (9), 780-792.
37. Matsuo, T.; Imai, C.; Yoshida, T.; Saito, T.; Hayashi, T.; Hirota, S., Creation of an artificial metalloprotein with a Hoveyda–Grubbs catalyst moiety through the intrinsic inhibition mechanism of α-chymotrypsin. *ChemComm.* **2012,** *48* (11), 1662-1664.
38. Haquette, P.; Salmain, M.; Svedlung, K.; Martel, A.; Rudolf, B.; Zakrzewski, J.; Cordier, S.; Roisnel, T.; Fosse, C.; Jaouen, G., Cysteine-Specific, Covalent Anchoring of Transition Organometallic Complexes to the Protein Papain from Carica papaya. *ChemBioChem* **2007,** *8* (2), 224-231.
39. Basauri-Molina, M.; Riemersma, C. F.; Würdemann, M. A.; Kleijn, H.; Klein Gebbink, R. J. M., Lipase active site covalent anchoring of Rh(NHC) catalysts: towards chemoselective artificial metalloenzymes. *ChemComm.* **2015,** *51* (31), 6792-6795.
40. Kruithof, C. A.; Casado, M. A.; Guillena, G.; Egmond, M. R.; van der Kerk-van Hoof, A.; Heck, A. J.; Klein Gebbink, R. J.; van Koten, G., Lipase active-site-directed anchoring of organometallics: metallopincer/protein hybrids. *Chem. Eur. J* **2005,** *11* (23), 6869-77.
41. Alonso, S.; Santiago, G.; Cea-Rama, I.; Fernandez-Lopez, L.; Coscolín, C.; Modregger, J.; Ressmann, A. K.; Martínez-Martínez, M.; Marrero, H.; Bargiela, R.; Pita, M.; Gonzalez-Alfonso, J. L.; Briand, M. L.; Rojo,

D.; Barbas, C.; Plou, F. J.; Golyshin, P. N.; Shahgaldian, P.; Sanz-Aparicio, J.; Guallar, V.; Ferrer, M., Genetically engineered proteins with two active sites for enhanced biocatalysis and synergistic chemo- and biocatalysis. *Nat. Catal.* **2020,** *3* (3), 319-328.

42. Jordan, P. C.; Patterson, D. P.; Saboda, K. N.; Edwards, E. J.; Miettinen, H. M.; Basu, G.; Thielges, M. C.; Douglas, T., Self-assembling biomolecular catalysts for hydrogen production. *Nat. Chem.* **2016,** *8* (2), 179-185.
43. Beesley, J. L.; Woolfson, D. N., The *de novo* design of α-helical peptides for supramolecular self-assembly. *Curr. Opin. Biotechnol.* **2019,** *58*, 175-182.
44. Zastrow, M. L.; Peacock, A. F. A.; Stuckey, J. A.; Pecoraro, V. L., Hydrolytic catalysis and structural stabilization in a designed metalloprotein. *Nat. Chem.* **2012,** *4* (2), 118-123.
45. Tolbert, A. E.; Ervin, C. S.; Ruckthong, L.; Paul, T. J.; Jayasinghe-Arachchige, V. M.; Neupane, K. P.; Stuckey, J. A.; Prabhakar, R.; Pecoraro, V. L., Heteromeric three-stranded coiled coils designed using a Pb(II)(Cys)(3) template mediated strategy. *Nat. Chem.* **2020,** *12* (4), 405-411.
46. Tegoni, M.; Yu, F.; Bersellini, M.; Penner-Hahn, J. E.; Pecoraro, V. L., Designing a functional type 2 copper center that has nitrite reductase activity within α-helical coiled coils. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (52), 21234-21239.
47. Yu, F.; Penner-Hahn, J. E.; Pecoraro, V. L., *De novo*-designed metallopeptides with type 2 copper centers: modulation of reduction potentials and nitrite reductase activities. *J. Am. Chem. Soc.* **2013,** *135* (48), 18096-18107.

48. Tebo, A. G.; Pinter, T. B. J.; García-Serres, R.; Speelman, A. L.; Tard, C.; Sénéque, O.; Blondin, G.; Latour, J.-M.; Penner-Hahn, J.; Lehnert, N.; Pecoraro, V. L., Development of a rubredoxin-type center embedded in a *de novo*-designed three-helix Bundle. *Biochemistry* **2018,** *57* (16), 2308-2316.

49. Sambasivan, R.; Ball, Z. T., Screening rhodium metallopeptide libraries "On bead": asymmetric cyclopropanation and a solution to the enantiomer problem. *Angew. Chem. Int. Ed.* **2012,** *51* (34), 8568-8572.

50. Studer, S.; Hansen, D. A.; Pianowski, Z. L.; Mittl, P. R. E.; Debon, A.; Guffy, S. L.; Der, B. S.; Kuhlman, B.; Hilvert, D., Evolution of a highly active and enantiospecific metalloenzyme from short peptides. *Science* **2018,** *362* (6420), 1285-1288.

51. Koder, R. L.; Anderson, J. L. R.; Solomon, L. A.; Reddy, K. S.; Moser, C. C.; Dutton, P. L., Design and engineering of an O_2 transport protein. *Nature* **2009,** *458* (7236), 305-309.

52. Farid, T. A.; Kodali, G.; Solomon, L. A.; Lichtenstein, B. R.; Sheehan, M. M.; Fry, B. A.; Bialas, C.; Ennist, N. M.; Siedlecki, J. A.; Zhao, Z.; Stetz, M. A.; Valentine, K. G.; Anderson, J. L. R.; Wand, A. J.; Discher, B. M.; Moser, C. C.; Dutton, P. L., Elementary tetrahelical protein design for diverse oxidoreductase functions. *Nat. Chem. Biol.* **2013,** *9* (12), 826-833.

53. Watkins, D. W.; Jenkins, J. M. X.; Grayson, K. J.; Wood, N.; Steventon, J. W.; Le Vay, K. K.; Goodwin, M. I.; Mullen, A. S.; Bailey, H. J.; Crump, M. P.; MacMillan, F.; Mulholland, A. J.; Cameron, G.; Sessions, R. B.; Mann, S.; Anderson, J. L. R., Construction and *in vivo* assembly of a catalytically proficient and hyperthermostable *de novo* enzyme. *Nat. Commun.* **2017,** *8* (1), 358.

54. Salgado, E. N.; Radford, R. J.; Tezcan, F. A., Metal-directed protein self-assembly. *Acc. Chem. Res.* **2010,** *43* (5), 661-672.

55. Churchfield, L. A.; Tezcan, F. A., Design and construction of functional supramolecular metalloprotein assemblies. *Acc. Chem. Res.* **2019,** *52* (2), 345-355.

56. Song, W. J.; Tezcan, F. A., A designed supramolecular protein assembly with *in vivo* enzymatic activity. *Science* **2014,** *346* (6216), 1525-1528.

57. Rittle, J.; Field, M. J.; Green, M. T.; Tezcan, F. A., An efficient, step-economical strategy for the design of functional metalloproteins. *Nat. Chem.* **2019,** *11* (5), 434-441.

58. Song, W. J.; Yu, J.; Tezcan, F. A., Importance of scaffold flexibility/rigidity in the design and directed evolution of artificial metallo-β-lactamases. *J. Am. Chem. Soc.* **2017,** *139* (46), 16772-16779.

59. Reig, A. J.; Pires, M. M.; Snyder, R. A.; Wu, Y.; Jo, H.; Kulp, D. W.; Butch, S. E.; Calhoun, J. R.; Szyperski, T.; Solomon, E. I.; DeGrado, W. F., Alteration of the oxygen-dependent reactivity of *de novo* Due Ferri proteins. *Nat. Chem.* **2012,** *4* (11), 900-906.

60. Faiella, M.; Andreozzi, C.; de Rosales, R. T. M.; Pavone, V.; Maglio, O.; Nastri, F.; DeGrado, W. F.; Lombardi, A., An artificial di-iron oxo-protein with phenol oxidase activity. *Nat. Chem. Biol.* **2009,** *5* (12), 882-884.

61. Zhang, S.-Q.; Chino, M.; Liu, L.; Tang, Y.; Hu, X.; DeGrado, W. F.; Lombardi, A., *De novo* design of tetranuclear transition metal clusters stabilized by hydrogen-bonded networks in helical bundles. *J. Am. Chem. Soc.* **2018,** *140* (4), 1294-1304.

62. Polizzi, N. F.; Wu, Y.; Lemmin, T.; Maxwell, A. M.; Zhang, S.-Q.; Rawson, J.; Beratan, D. N.; Therien, M. J.; DeGrado, W. F., *De novo*

design of a hyperstable non-natural protein–ligand complex with sub-Å accuracy. *Nat. Chem.* **2017,** *9* (12), 1157-1164.

63. Joh, N. H.; Wang, T.; Bhate, M. P.; Acharya, R.; Wu, Y.; Grabe, M.; Hong, M.; Grigoryan, G.; DeGrado, W. F., *De novo* design of a transmembrane Zn^{2+}-transporting four-helix bundle. *Science* **2014,** *346* (6216), 1520-1524.
64. Perez-Aguilar, J. M.; Saven, J. G., Computational design of membrane proteins. *Structure* **2012,** *20* (1), 5-14.
65. Niitsu, A.; Heal, J. W.; Fauland, K.; Thomson, A. R.; Woolfson, D. N., Membrane-spanning α-helical barrels as tractable protein-design targets. *Philos. Trans. R. Soc. B* **2017,** *372* (1726), 20160213.
66. Patel, S. C.; Hecht, M. H., Directed evolution of the peroxidase activity of a *de novo*-designed protein. *Protein Eng. Des. Sel.* **2012,** *25* (9), 445-452.
67. Yang, H.; Swartz, A. M.; Park, H. J.; Srivastava, P.; Ellis-Guardiola, K.; Upp, D. M.; Lee, G.; Belsare, K.; Gu, Y.; Zhang, C.; Moellering, R. E.; Lewis, J. C., Evolving artificial metalloenzymes via random mutagenesis. *Nat. Chem.* **2018,** *10* (3), 318-324.
68. Fischer, J.; Renn, D.; Quitterer, F.; Radhakrishnan, A.; Liu, M.; Makki, A.; Ghorpade, S.; Rueping, M.; Arold, S. T.; Groll, M.; Eppinger, J., Robust and versatile host protein for the design and evaluation of artificial metal centers. *ACS Catal.* **2019,** *9* (12), 11371-11380.
69. Lee, J.; Song, W. J., Photocatalytic C–O coupling enzymes that operate via intramolecular electron transfer. *J. Amer. Chem. Soc.* **2023,** *145* (9), 5211-5221.
70. Jeske, L.; Placzek, S.; Schomburg, I.; Chang, A.; Schomburg, D., BRENDA in 2019: a European ELIXIR core data resource. *Nucleic*

Acids Res. **2018,** *47* (D1), D542-D549.

71. Lipscomb, W. N.; Sträter, N., Recent advances in zinc enzymology. *Chem. Rev.* **1996,** *96* (7), 2375-2434.
72. Lombard, V.; Golaconda Ramulu, H.; Drula, E.; Coutinho, P. M.; Henrissat, B., The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res.* **2013,** *42* (D1), D490-D495.
73. Yip, V. L. Y.; Varrot, A.; Davies, G. J.; Rajan, S. S.; Yang, X.; Thompson, J.; Anderson, W. F.; Withers, S. G., An unusual mechanism of glycoside hydrolysis involving redox and elimination steps by a family 4 β-glycosidase from *Thermotoga maritima*. *J. Am. Chem. Soc.* **2004,** *126* (27), 8354-8355.
74. Zhu, Y.; Suits, M. D. L.; Thompson, A. J.; Chavan, S.; Dinev, Z.; Dumon, C.; Smith, N.; Moremen, K. W.; Xiang, Y.; Siriwardena, A.; Williams, S. J.; Gilbert, H. J.; Davies, G. J., Mechanistic insights into a Ca^{2+}-dependent family of α-mannosidases in a human gut symbiont. *Nat. Chem. Biol.* **2010,** *6* (2), 125-132.
75. Jongkees, S. A. K.; Withers, S. G., Unusual enzymatic glycoside cleavage mechanisms. *Acc. Chem. Res.* **2014,** *47* (1), 226-235.
76. Zechel, D. L.; Withers, S. G., Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc. Chem. Res.* **2000,** *33* (1), 11-18.
77. Davidi, D.; Longo, L. M.; Jabłońska, J.; Milo, R.; Tawfik, D. S., A bird's-eye view of enzyme evolution: Chemical, physicochemical, and physiological considerations. *Chem. Rev.* **2018,** *118* (18), 8786-8797.
78. Striegler, S.; Barnett, J. D.; Dunaway, N. A., Glycoside hydrolysis with sugar-templated microgel catalysts. *ACS Catal.* **2012,** *2* (1), 50-55.
79. Striegler, S.; Dunaway, N. A.; Gichinga, M. G.; Barnett, J. D.; Nelson, A. G., Evaluating binuclear copper(II) complexes for glycoside

hydrolysis. *Inorg. Chem.* **2010,** *49* (6), 2639-48.

80. Yu, Z.; Cowan, J. A., Design of artificial glycosidases: Metallopeptides that remove H antigen from human erythrocytes. *Angew. Chem. Int. Ed.* **2017,** *56* (10), 2763-2766.
81. Sharma, B.; Striegler, S.; Whaley, M., Modulating the catalytic performance of an immobilized catalyst with matrix effects - A critical evaluation. *ACS Catal.* **2018,** *8* (8), 7710-7718.
82. Thompson, Z.; Cowan, J. A., Artificial metalloenzymes: Recent developments and innovations in bioinorganic catalysis. *Small* **2020,** *16* (27), 2000392.
83. Jeong, W. J.; Yu, J.; Song, W. J., Proteins as diverse, efficient, and evolvable scaffolds for artificial metalloenzymes. *Chem. Commun.* **2020,** *56* (67), 9586-9599.
84. Lehn, J.-M., Supramolecular chemistry: Receptors, catalysts, and carriers. *Science* **1985,** *227* (4689), 849-856.
85. Brown, C. J.; Toste, F. D.; Bergman, R. G.; Raymond, K. N., Supramolecular catalysis in metal–ligand cluster hosts. *Chem. Rev.* **2015,** *115* (9), 3012-3035.
86. Visudtiphole, V.; Thomas, M. B.; Chalton, D. A.; Lakey, J. H., Refolding of *Escherichia coli* outer membrane protein F in detergent creates LPS-free trimers and asymmetric dimers. *Biochem. J.* **2005,** *392* (Pt 2), 375-81.
87. Zastrow, M. L.; Peacock, A. F. A.; Stuckey, J. A.; Pecoraro, V. L., Hydrolytic catalysis and structural stabilization in a designed metalloprotein. *Nat. Chem.* **2011,** *4*, 118.
88. Koike, T.; Takamura, M.; Kimura, E., Role of Zinc(II) in beta-lactamase II: a model study with a Zinc(II)-macrocyclic tetraamine (1,4,7,10-

Tetraazacyclododecane, Cyclen) complex. *J. Am. Chem. Soc.* **1994,** *116* (19), 8443-8449.

89. Baier, F.; Chen, J.; Solomonson, M.; Strynadka, N. C. J.; Tokuriki, N., Distinct metal isoforms underlie promiscuous activity profiles of metalloenzymes. *ACS Chem. Biol.* **2015,** *10* (7), 1684-1693.
90. Daumann, L. J.; Gahan, L. R.; Comba, P.; Schenk, G., Cadmium(II) complexes: Mimics of organophosphate pesticide degrading enzymes and metallo-β-lactamases. *Inorg. Chem.* **2012,** *51* (14), 7669-7681.
91. Kaminskaia, N. V.; Spingler, B.; Lippard, S. J., Hydrolysis of β-Lactam antibiotics catalyzed by dinuclear Zinc(II) complexes: Functional mimics of metallo-β-lactamases. *J. Am. Chem. Soc.* **2000,** *122* (27), 6411-6422.
92. Ziervogel, Brigitte K.; Roux, B., The binding of antibiotics in OmpF porin. *Structure* **2013,** *21* (1), 76-87.
93. McCarter, J. D.; Stephen Withers, G., Mechanisms of enzymatic glycoside hydrolysis. *Curr. Opin. Struct. Biol.* **1994,** *4* (6), 885-892.
94. Davies, G.; Henrissat, B., Structures and mechanisms of glycosyl hydrolases. *Structure* **1995,** *3* (9), 853-859.
95. Wolfenden, R.; Lu, X.; Young, G., Spontaneous hydrolysis of glycosides. *J. Am. Chem. Soc.* **1998,** *120* (27), 6814-6815.
96. Vallmitjana, M.; Ferrer-Navarro, M.; Planell, R.; Abel, M.; Ausín, C.; Querol, E.; Planas, A.; Pérez-Pons, J.-A., Mechanism of the family 1 β-glucosidase from *Streptomyces sp*: Catalytic residues and kinetic studies. *Biochemistry* **2001,** *40* (20), 5975-5982.
97. Williams, S. J.; Mark, B. L.; Vocadlo, D. J.; James, M. N.; Withers, S. G., Aspartate 313 in the *Streptomyces plicatus* hexosaminidase plays a critical role in substrate-assisted catalysis by orienting the 2-acetamido

group and stabilizing the transition state. *J. Biol. Chem.* **2002,** *277* (42), 40055-65.

98. Park, J. K.; Wang, L. X.; Patel, H. V.; Roseman, S., Molecular cloning and characterization of a unique beta-glucosidase from *Vibrio cholerae*. *J. Biol. Chem.* **2002,** *277* (33), 29555-60.
99. McIntosh, L. P.; Hand, G.; Johnson, P. E.; Joshi, M. D.; Körner, M.; Plesniak, L. A.; Ziser, L.; Wakarchuk, W. W.; Withers, S. G., The pKa of the general acid/base carboxyl group of a glycosidase cycles during catalysis: A ^{13}C-NMR study of *Bacillus circulans* xylanase. *Biochemistry* **1996,** *35* (31), 9958-9966.
100. Gessesse, A., Purification and properties of two thermostable alkaline xylanases from an alkaliphilic *Bacillus sp*. *Appl. Environ. Microbiol.* **1998,** *64* (9), 3533-3535.
101. Nakamura, S.; Wakabayashi, K.; Nakai, R.; Aono, R.; Horikoshi, K., Purification and some properties of an alkaline xylanase from alkaliphilic *Bacillus sp.* strain 41M-1. *Appl. Environ. Microbiol.* **1993,** *59* (7), 2311-2316.
102. Mamo, G.; Thunnissen, M.; Hatti-Kaul, R.; Mattiasson, B., An alkaline active xylanase: Insights into mechanisms of high pH catalytic adaptation. *Biochimie* **2009,** *91* (9), 1187-1196.
103. Varma, S.; Jakobsson, E., Ionization states of residues in OmpF and mutants: Effects of dielectric constant and interactions between residues. *Biophys. J.* **2004,** *86* (2), 690-704.
104. Zachariae, U.; Koumanov, A.; Engelhardt, H.; Karshikoff, A., Electrostatic properties of the anion selective porin Omp32 from *Delftia acidovorans* and of the arginine cluster of bacterial porins. *Protein Sci.* **2002,** *11* (6), 1309-19.

105. Aguilella-Arzo, M.; Andrio, A.; Aguilella, V. M.; Alcaraz, A., Dielectric saturation of water in a membrane protein channel. *Phys. Chem. Chem. Phys.***2009,** *11* (2), 358-65.

106. Rempel, B. P.; Withers, S. G., Covalent inhibitors of glycosidases and their applications in biochemistry and biology. *Glycobiology* **2008,** *18* (8), 570-86.

107. Dan, S.; Marton, I.; Dekel, M.; Bravdo, B.-A.; He, S.; Withers, S. G.; Shoseyov, O., Cloning, expression, characterization, and nucleophile identification of family 3, *Aspergillus niger b*-Glucosidase. *J. Biol. Chem.* **2000,** *275* (7), 4973-4980.

108. Wang, Q.; Trimbur, D.; Graham, R.; Warren, R. A. J.; Withers, S. G., Identification of the acid/base catalyst in *Agrobacterium faecalis* beta-glucosidase by kinetic analysis of mutants. *Biochemistry* **1995,** *34* (44), 14554-14562.

109. Bravman, T.; Belakhov, V.; Solomon, D.; Shoham, G.; Henrissat, B.; Baasov, T.; Shoham, Y., Identification of the catalytic residues in family 52 glycoside hydrolase, a β-xylosidase from *Geobacillus stearothermophilus* T-6*. *J. Biol. Chem.* **2003,** *278* (29), 26742-26749.

110. Quinlan, R. J.; Sweeney, M. D.; Leggio, L. L.; Otten, H.; Poulsen, J.-C. N.; Johansen, K. S.; Krogh, K. B. R. M.; Jørgensen, C. I.; Tovborg, M.; Anthonsen, A.; Tryfona, T.; Walter, C. P.; Dupree, P.; Xu, F.; Davies, G. J.; Walton, P. H., Insights into the oxidative degradation of cellulose by a copper metalloenzyme that exploits biomass components. *Proc. Natl. Acad. Sci. U.S.A.* **2011,** *108* (37), 15079-15084.

111. Concia, A. L.; Beccia, M. R.; Orio, M.; Ferre, F. T.; Scarpellini, M.; Biaso, F.; Guigliarelli, B.; Réglier, M.; Simaan, A. J., Copper complexes as bioinspired models for lytic polysaccharide monooxygenases. *Inorg.*

Chem. **2017,** *56* (3), 1023-1026.

112. Yu, Z.; Thompson, Z.; Behnke, S. L.; Fenk, K. D.; Huang, D.; Shafaat, H. S.; Cowan, J. A., Metalloglycosidase mimics: Oxidative cleavage of saccharides promoted by multinuclear copper complexes under physiological conditions. *Inorg. Chem.* **2020,** *59* (16), 11218-11222.
113. McGregor, N. G. S.; Coines, J.; Borlandelli, V.; Amaki, S.; Artola, M.; Nin-Hill, A.; Linzel, D.; Yamada, C.; Arakawa, T.; Ishiwata, A.; Ito, Y.; van der Marel, G. A.; Codée, J. D. C.; Fushinobu, S.; Overkleeft, H. S.; Rovira, C.; Davies, G. J., Cysteine nucleophiles in glycosidase catalysis: Application of a covalent β-l-arabinofuranosidase inhibitor. *Angew. Chem. Int. Ed.* **2021,** *60* (11), 5754-5758.
114. Pandya, C.; Farelli, J. D.; Dunaway-Mariano, D.; Allen, K. N., Enzyme promiscuity: engine of evolutionary innovation. *J. Biol. Chem.* **2014,** *289* (44), 30229-30236.
115. Devamani, T.; Rauwerdink, A. M.; Lunzer, M.; Jones, B. J.; Mooney, J. L.; Tan, M. A. O.; Zhang, Z.-J.; Xu, J.-H.; Dean, A. M.; Kazlauskas, R. J., Catalytic promiscuity of ancestral esterases and hydroxynitrile lyases. *J. Am. Chem. Soc.* **2016,** *138* (3), 1046-1056.
116. Dhakshnamoorthy, B.; Ziervogel, B. K.; Blachowicz, L.; Roux, B., A structural study of ion permeation in OmpF porin from anomalous X-ray diffraction and molecular dynamics simulations. *J. Am. Chem. Soc.* **2013,** *135* (44), 16561-16568.
117. Otwinowski, Z.; Minor, W., [20] Processing of X-ray diffraction data collected in oscillation mode. In *Methods Enzymol.*, Academic Press: **1997**; Vol. 276, pp 307-326.
118. Kabsch, W., XDS. *Acta Crystallogr. D* **2010,** *66*, 125-132.
119. Collaborative, The CCP4 suite: programs for protein crystallography.

Acta Cryst. D **1994,** *50* (5), 760-763.

120. Vagin, A.; Teplyakov, A., Molecular replacement with MOLREP. *Acta Cryst. D* **2010,** *66* (1), 22-25.
121. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Crystallogr.* **2007,** *40* (Pt 4), 658-674.
122. Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L.-W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D., Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Cryst. D* **2019,** *75*, 861-877.
123. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Cryst. D* **2010,** *66* (4), 486-501.
124. Wiegels, T.; Lamzin, V. S., Use of noncrystallographic symmetry for automated model building at medium to low resolution. *Acta Cryst. D* **2012,** *68* (Pt 4), 446-53.
125. Yamashita, E.; Zhalnina, M. V.; Zakharov, S. D.; Sharma, O.; Cramer, W. A., Crystal structures of the OmpF porin: function in a colicin translocon. *EMBO J.* **2008,** *27* (15), 2171-80.
126. Meuskens, I.; Michalik, M.; Chauhan, N.; Linke, D.; Leo, J. C., A New strain collection for improved expression of outer membrane proteins. *Front.Cell. Infect. Microbiol.* **2017,** *7*, 464.
127. Kille, S.; Acevedo-Rocha, C. G.; Parra, L. P.; Zhang, Z.-G.; Opperman, D. J.; Reetz, M. T.; Acevedo, J. P., Reducing codon redundancy and

screening effort of combinatorial protein libraries created by saturation mutagenesis. *ACS Synth. Biol.* **2013,** *2* (2), 83-92.

128. Reetz, M. T.; Carballeira, J. D., Iterative saturation mutagenesis (ISM) for rapid directed evolution of functional enzymes. *Nat. Protoc.* **2007,** *2* (4), 891-903.
129. Shevchenko, A.; Tomas, H.; Havli, J.; Olsen, J. V.; Mann, M., In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* **2006,** *1* (6), 2856-2860.
130. Wang, Q.; Graham, R. W.; Trimbur, D.; Warren, R. A. J.; Withers, S. G., Changing enzymic reaction mechanisms by mutagenesis: Conversion of a retaining glucosidase to an inverting enzyme. *J. Am. Chem. Soc.* **1994,** *116* (25), 11594-11595.
131. Asano, K.; Matsubara, S., Effects of a flexible alkyl chain on a ligand for CuAAC Reaction. *Org. Lett.* **2010,** *12* (21), 4988-4991.
132. Rajoka, M. I.; Idrees, S.; Ashfaq, U. A.; Ehsan, B.; Haq, A., Determination of substrate specificities against β-glucosidase A (BglA) from *Thermotoga maritime*: a molecular docking approach. *J. Microbiol. Biiotechnol.* **2015,** *25* (1), 44-9.
133. Trott, O.; Olson, A. J., AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2010,** *31* (2), 455-461.
134. McCall, K. A.; Huang, C.-c.; Fierke, C. A., Function and mechanism of zinc metalloenzymes. *J. Nutr.* **2000,** *130* (5), 1437S-1446S.
135. Tamames, B.; Sousa, S. F.; Tamames, J.; Fernandes, P. A.; Ramos, M. J., Analysis of zinc-ligand bond lengths in metalloproteins: Trends and patterns. *Proteins: Struc. Funct. Bioinform.* **2007,** *69* (3), 466-475.
136. Zastrow, M. L.; Pecoraro, V. L., Designing hydrolytic zinc

metalloenzymes. *Biochemistry* **2014,** *53* (6), 957-978.

137. Raghava, G. P. S.; Barton, G. J., Quantification of the variation in percentage identity for protein sequence alignments. *BMC Bioinform.* **2006,** *7* (1), 415.

국문 초록

다양한 단백질 골격을 활용한 인공 가수 분해 효소의 설계를 위한 방법 개발

정우재

자연과학대학 화학부

서울대학교 대학원

자연에 존재하는 금속 효소는 금속 이온과 아미노산 간의 필수적인 상호작용을 보여준다. 인공 금속 효소의 분자 설계는 이러한 상호작용을 탐색 하기 위한 하나의 방법이었다. 그러나 이러한 맞춤형 인공 금속 효소의 연구들에서는 제한된 수의 단백질 골격들을 사용했다. 그래서, 본 논문에서는 외막 단백질 F (OmpF) 과 같은 비-알파-나선형 구조와 같은 다양한 단백질 골격들에 적용이 가능한 좀 더 일반적인 인공 금속 효소 설계 방법의 개발에 집중했다.

1 장에서는 인공 금속 효소 연구 분야와 자주 사용되는 단백질 골격에 대한 최신 연구 성과를 기술하였다. 인공 금속 효소로의 성공적인 전환이 많이 보고 되었지만, 이러한 성과들은 좁은 범위의 단백질 골격만을 사용하고 있다. 이러한 제한된 단백질 골격의 사용은 기존에 존재하는 금속 자리를 이용한 인공 금속 효소 설계 등과 같이 설계 전략 자체에서 오는 한계에 기인한다.

이러한 한계를 극복하기 위해서 2 장에서는 자연에 존재하는 아연 금속 효소의 금속 결합 잔기들의 길이와 같은 기하학적인 매개변수를 이용한 독창적인 인공 금속 효소의 설계 방법을 기술한다. 이러한 독창적인 방법을 통해, 비 알파-나선형 단백질 골격, 특히 베타-배럴구조를 가지는 외막 단백질 F(OmpF)에 개발한 방법을

적용하여 몇 가지의 변형된 아연 결합 단백질을 만들었고, 이를 그 결정 구조를 통해 확인했다. 이러한 연구를 통해, 다양한 단백질 구조의 사용이 가능해졌고 원하는 촉매 특성을 가진 효소를 다루는 능력을 향상시킬 수 있었다. 궁극적으로 본 연구는 광범위한 응용 분야에 사용될 수 있는 보다 효율적이고 다용도의 인공 금속 효소를 개발하는 데 기여하고자 한다.

3 장에서는 기하학적 매개변수에 기초한 인공 금속 효소의 설계를 기술한다. 비 알파-나선형 단백질 골격, 특히 베타 베럴 구조를 가지는 외막 단백질 F(OmpF)에 사면체 아연 결합 자리를 적용했다. 이러한 접근법을 통해 인공 아연 결합 단백질의 세 가지 변형체를 만들고 이의 결정 구조를 규명했다. 이러한 변형체들은 서로 다른 첫 번째 배위 자리를 가지고 있으며, 이로 인해 서로 다른 촉매 특성이 나타났다. 이 효소들은 다양한 기질, 특히 베타 결합을 가지는 당류에 대한 가수분해 활성을 보였다. 더 나아가, 유도 진화를 통해 진화된 새로운 아연 의존적인 인공 금속 효소를 성공적으로 얻었다. 본 연구를 통해 우리가 발견한 아연 결합 자리의 규칙이 다양한 단백질 골격에 적용되는 것을 확인 및 응용 가능함을 확인하였으며, 이를 통해 생화학 공간에서 무기 보조 인자의 화학적 능력을 확장할 수 있는 길을 열었다.

주요어: 인공 금속 효소, 외막 단백질 F, 불포화 사면체 아연 금속 자리, 금속 가수 분해 효소, 당 가수분해 효소, 유도진화

학번: 2016-25496