Chem Soc Rev

Metalloprotein Design Using Genetic Code Expansion

Journal:	*Chemical Society Reviews*
Manuscript ID:	CS-TRV-01-2014-000018.R1
Article Type:	Tutorial Review
Date Submitted by the Author:	05-Mar-2014
Complete List of Authors:	Wang, Jiangyun; Institute of Biophysics, Sawyer, Elizabeth; Institute of Biophysics, Chan, Sunney; Academia Sinica, Institute of Chemistry; California Institute of Technology, Chemistry Hu, Cheng; Institute of Biophysics, Yu, Yang; Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences,

Metalloprotein Design Using Genetic Code Expansion

Cheng Hu[2,4], Sunney I. Chan[3,4], Elizabeth B. Sawyer[2,4], Yang Yu[1], Jiangyun Wang[2]*

[1]Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, Tianjin 300308, China

[2]Institute of Biophysics, Chinese Academy of Sciences, 15 Datun Road, Chaoyang District, Beijing, 100101, China

[3]Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 911525, USA

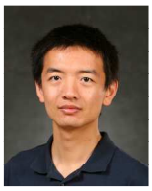
[4]These authors contributed equally to this work

E-mail: jwang@ibp.ac.cn

Mr. Cheng Hu obtained his B.S. degree from the College of Chemistry and Molecular Engineering, Peking University (China) in 2011. He is currently a graduate student under the direction of Prof. Jiangyun Wang at the Institute of Biophysics, Chinese Academy of Sciences, working on the development of artificial metalloenzymes.

Dr. Beth Sawyer studied MSci Biochemistry at Imperial College London. In 2009 she obtained her PhD in Chemistry from the University of Cambridge, where she researched bacterial heme transfer proteins under the supervision of Dr Paul Barker. Following one year's post-doctoral experience at The University of Melbourne, Australia, she moved to the Institute of Biophysics, Chinese Academy of Sciences, where she has been studying protein folding and engineering under the supervision of Professor Sarah Perrett since 2010.

Dr. Sunney I. Chan. B.S., University of California, 1957; Ph.D., 1961. Assistant Professor of Chemical Physics, Caltech, 1963-64; Associate Professor, 1964-68; Professor, 1968-76; Professor of Chemical Physics and Biophysical Chemistry, 1976-92; Hoag Professor, 1992-2001; Hoag Professor Emeritus, 2002-. His research interests include protein folding, membrane proteins, and bioinorganic chemistry. His work on metalloproteins, which include cytochrome c oxidase and the particulate methane monooxygenase, has spanned several decades.



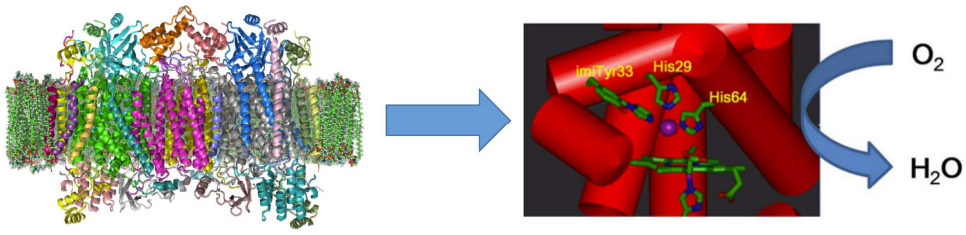
Dr. Yang Yu received his B. Sc. degree from Peking University in 2008 and his Ph. D. degree from University of Illinois at Urbana-Champaign in 2014 under the direction of Professor Yi Lu. After that, He joined Tianjin Institute of Industrial Biotechnology, Chinese Academy of Sciences, where he is working on metalloprotein design and engineering using unnatural amino acids

Dr. Jiangyun Wang learned physical chemistry at University of Science and Technology of China and graduated with a B.S. degree in 1998. In 2003, he finished his Ph.D study in metalloprotein chemistry under the supervision of Professor Kenneth Suslick in the University of Illinois at Urbana-Champaign. Between 2003 and 2007, he worked in the laboratory of Professor Peter Schultz at the Scripps Research Institute as a postdoctoral fellow. In 2008, he started his independent lab at the Institute of Biophysics, Chinese Academy of Sciences.

Abstract: More than one third of all proteins are metalloproteins. They catalyze important reactions such as photosynthesis, nitrogen fixation and CO_2 reduction. Metalloproteins such as the olfactory receptors also serve as highly elaborate sensors. Here we review recent developments in

functional metalloprotein design using the genetic code expansion approach. We show that, through the site-specific incorporation of metal-chelating unnatural amino acids (UAA), proton and electron transfer mediators, and UAAs bearing bioorthgonal reaction groups, small soluble proteins can recapitulate and expand the important functions of complex metalloproteins. Further developments along this route may result in cell factories and live-cell sensors with unprecedented efficiency and selectivity.

Table of contents entry:



Genetic code expansion has become an essential new tool for designing functional small protein models for complex metalloenzymes.

Key Learning Points

Genetic code expansion: genetic code expansion is a strategy to rewire the protein translation machinery to recode blank codons as unnatural amino acids (UAA). Since UAA can have intrinsic properties such as metal-chelating, redox activity, chemists' ability to probe protein mechanism, and to design new enzymes is much improved.

Bioorthogonal reaction: bioorthogonal reactions require high selectivity, high yield, and high reaction rate, and that both reactants do not react with natural biomolecules.

Cytochrome *c* oxidase (CcO): CcO is a membrane protein residing in the inner membrane of the mitochondrion, the powerhouse of eukaryotic cells. CcO's main function is to convert oxygen to water, and generate pH gradient across the membrane, which drives ATP synthesis.

Metalloenzyme design using small-soluble protein scaffold: many metalloenzymes, such as photosystem II and CcO are very complex membrane protein complexes harboring multiple metal ions and cofactors, and are difficult to study. By recapitulating the metal clusters which are essential for native enzyme's activity into a small-soluble protein scaffold (wolf in sheep's clothing), it becomes much simpler to carry out mechanistic studies.

Introduction

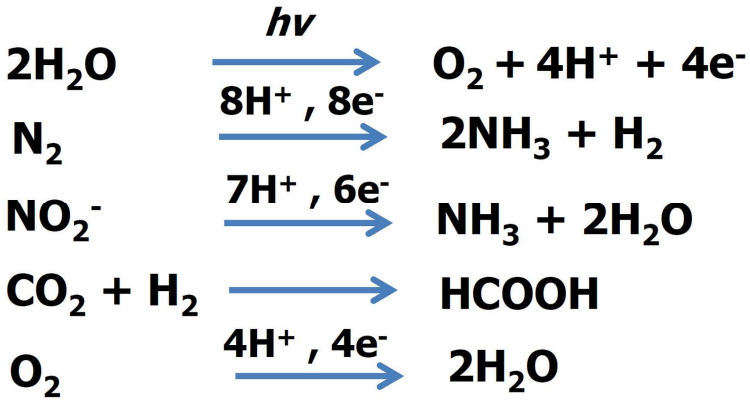
Chemists have always been interested in certain benchmark chemical reactions. Some of these include: (1) the H and H_2 exchange reaction; (2) the proton-coupled four-electron reduction of O_2 to water; (3) the conversion of N_2 to NH_3; (4) the proton-coupled six-electron reduction of nitrite

to ammonia; (5) reduction of CO_2 to yield formic acid; and (6) the photooxidation of water to produce O_2 (scheme 1).

While the H and H_2 exchange reaction is relatively simple, the remaining reactions in Scheme 1 are significantly more complex. For example, it is not trivial to bring four electrons and four protons together for the concerted reduction of O_2 to two water molecules without the formation of reactive oxygen species (ROS). Nature carries out the reactions highlighted in Scheme 1 with high efficiency and selectivity with the help of metalloenzymes. Accordingly, we can uncover the mechanisms of the catalytic chemistry mediated by these enzymes and develop biomimetics of these systems. The development of artificial catalysts based on knowledge gleaned from natural systems has been a major focus of research among chemists for many years. Chemists are, of course, at a disadvantage here. The reaction potential surface that a metalloenzyme uses to catalyze one of these reactions is extremely complex. Biomacromolecules have many more degrees of freedom than an inorganic complex and can therefore use the protein scaffold supporting the catalytic metal centers to tune reaction specificities and reaction rates. Decades of research have revealed that the precise location of cofactors, efficient electron and proton transfer pathways through mediators such as tyrosine, and elaborate arrangement of metal ion clusters hold the key for the high catalytic activity and selectivity of metalloenzymes.

To design biomimetics of metalloproteins, two kinds of complementary methods have been used. The first approach relies on synthetic chemistry techniques to synthesize small molecule models, which permit atomic control and faithful recapitulation of the coordination shell. However, it remains difficult to install secondary shell elements, which are often key to control electron and proton transfer to substrates. While this approach has been successful in revealing some of the fundamental principles which the metalloenzymes employ, small molecule model compounds can be difficult and costly to prepare, and recapitulate only part of the elaborate properties of native metalloenzymes. The second approach use molecular biology techniques to synthesize small protein models of complex metalloenzymes.[1] Advantages of this method include the ability to install both coordination shell and secondary shell elements around metal centers, the economic production of gram quantities of protein using well-established protocols and the ability to mutate and optimize proteins through directed evolution. However, it has remained difficult to introduce metal-binding sites to proteins with defined coordination spheres, due to the lack of genetically encodable metal-chelating amino acids. Furthermore, it is challenging to fine tune the pK_a and reduction potential of a particular amino acid to facilitate the precise delivery of electrons and protons to substrates. Thus, the availability of only twenty natural amino acids poses a stringent limit on the chemist's ability to probe and expand the function of metalloproteins.

Using the genetic code expansion technique[2], the advantages of two approaches described above can be combined, and their disadvantages overcome. In the following sections, we will first describe recent progresses on the genetic incorporation of metal-chelating unnatural amino acids,[3-7] and the developments of bioorthogonal reactions[8, 9] for the specific ligation of cofactors. We will then highlight how these approaches have been successfully employed in the synthesis of small-protein models of complex metalloenzymes such as cytochrome *c* oxidase (CcO), nitrite reductase, and the design of metalloprotein sensors.



Scheme 1. Examples of important reactions catalyzed by metalloenzymes. While the metal clusters for each metalloenzyme are different, they all require the precise delivery of protons and electrons for substrate activation.

General metalloprotein design strategies: heme protein as an example

Metalloprotein design strategies without using genetic code expansion have recently been reviewed.[1] Since then, significant progress has been made in heme protein design. Here, we will first review general heme protein design strategies, and then we will discuss recent work on heme protein design using genetic code expansion.

The Protein Data Bank (PDB) contains over 3200 structures of heme-binding proteins. In fact, the first protein structures ever solved were those of hemoglobin and myoglobin, for which Max Perutz and Sir John Kendrew were awarded the 1962 Nobel Prize in Chemistry. The structures of many other classes of heme-binding proteins have followed, and we now appreciate how vast and elaborate the range of structural associations between proteins and heme is. It would appear that evolutionary pressures to use this versatile and reactive cofactor in a range of biochemical functions have led to the evolution of a wide range of distinct protein folds and structures even amongst proteins of similar function.

What is it that makes heme such a suitable cofactor for biological catalysis? Heme (iron protoporhyrin IX) is a planar, cyclic tetrapyrrole with the four pyrrole rings are linked by methene (meso) bridges. The porphyrin ring is a tetradentate ligand and can bind a variety of metals (in the case of heme, iron) through its imino (=N-) and pyrrole (-NH-) nitrogens. The sidechains of the porphyrin ring can also be substituted to yield porphyrins with different properties. The association of heme with proteins not only confers a wide range of properties upon the proteins, but also enables the unique chemistry of the heme to be harnessed for use in biological systems. Iron is a versatile metal whose properties can be tightly defined and controlled by the local protein environment: the redox potential spans almost the whole biological range, from very low (nitrogenases) to very high (cytochrome P450s) potential, and a variety of spin states and geometries are accessible to both ferrous and ferric forms of the iron depending on the ligands available.

Heme may be covalently or non-covalently attached to proteins. Non-covalent binding is mediated through a combination of hydrophobic packing to the porphyrin and its substituents, π-π stacking to the extended system of delocalized electrons in the porphyrin, ionic or hydrogen bonding interactions between the heme propionate groups and polar residues, and coordination of the heme iron through one or more axial ligand(s) supplied by the protein (usually histidine, methionine or cysteine residues). In the case where only one axial ligand is provided, there is often space for water, CO, O_2 or another small molecule to bind on the other (distal) face of the heme.

In addition to the heme group, other metal-binding site are often found in heme proteins, such as in cytochrome *c* oxidase[10-13] (CcO) manganese peroxidase[14] and nitric oxide reductase[15] (NOR) which contain a copper (Cu_B), manganese or iron (Fe_B) binding site situated in close proximity to the heme group. As both CcO and NOR are large membrane proteins that harbor multiple chromophores and metal ions, it is difficult to investigate their catalytic mechanisms through conventional mutational and spectroscopic methods. One approach to counteract this difficulty is to reconstitute their active sites in small and easy-to-produce proteins, such as myoglobin (Mb), to mimic these complex membrane proteins, thereby addressing crucial factors that are responsible for the activity and selectivity of the native enzyme. Recently, Lu and co-workers rationally designed a metal binding site in myoglobin by introducing two histidines in the heme active site of Mb by Leu29His and Phe43His mutations.[15] His29 and His43, together with the native His64 residue, formed a 3-His metal-chelating site (Cu_B site). This Mb mutant is termed Cu_BMb, and functions as a copper-dependent heme oxygenase, degrading heme to verdoheme.[16] To mimic the NOR active site, a 3-His-1-Glu metal-chelating site is designed in Cu_BMb through an additional Val68Glu mutation (termed Fe_BMb). Direct evidence for Fe^{2+} binding to Fe_BMb, comes from a crystal structure[15] (Figure 1). Indeed, Fe(II)-FeBMb exhibits NOR activity, which demonstrates that it is the first protein model both structurally and functionally resemble native NOR. Not surprisingly, the crystal structure of a bacterial NOR from *Pseudomonas aeruginosa* at 2.7 Å resolution (PDB entry 3O0R),[17] shows that the non-heme iron (Fe_B) is coordinated by one Glu and three His ligands, validating the myoglobin NOR model. Importantly, the active site ligand and metal center of Fe_BMb overlays closely with that of the native NOR.

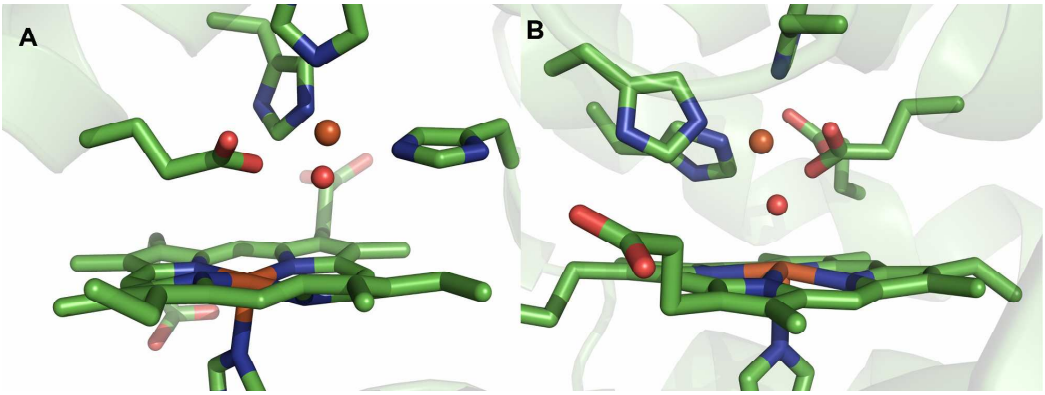


Figure 1. Crystal structures of bacteria nitric oxide reductase (left, PDB code 3O0R) and a myoglobin model of NOR (right, PDB code 3K9Z).

Modifying the native metalloenzymes is another approach to obtain desirable catalytic activity. Great progress has recently been made in engineering cytochrome P450 for olefin

cyclopropanation reactions.[18] Asymmetric cyclopropanation of olefins with carbene is an important reaction in the synthesis of natural products and pharmaceuticals, but cyclopropanation of olefins with carbene is not known to occur in native biological systems. The authors screened a P450 library containing random mutations in the close proximity to the heme center, and selected several engineered P450BM3 variants with enhanced turnover numbers and selectivity. They found that one variant P450BM3-T268A, is an active cyclopropanation catalyst with *trans*-selectivity. Another variant, named BM3-CIS, exhibits enhanced *cis*-selectivity. This work demonstrated that the function of metalloenzymes can be expanded to catalyze important reactions not known to occur in native biological systems, and optimized through random library screening and directed evolution.

Due to the strong interest in metalloporphyrin chemistry, great progress has been made in designing metalloenzymes harboring metalloporphyrin other than heme.[1, 18-21] While many metalloporphyrins have intrinsic catalytic abilities, incorporating them into protein scaffolds can not only improve solubility in aqueous solution, but also modulate catalytic activity and selectivity. One major difficulty in designing novel metalloporphyrin protein is the inability of metalloporphyrins to cross the cell membranes of nonpathogenic *E.coli* cells. Pathogenic *E.coli* strains are capable of utilizing external heme in the growth medium, while common lab strains such as BL21 strain do not have the outer membrane heme transporter, ChuA. As a result, the reconstitution of metalloporphyrins and apo-protein has to be done in *vitro*,[19] which requires harsh denaturing conditions to first remove the heme and then lengthy procedures for metalloporphyrin reconstitution and purification. This significantly limits the application of novel metalloprophyrins and makes it impossible to carry out directed evolution and high-throughput screening to obtain desirable properties. Much progress has been made towards solving this problem. Marletta and coworkers[21] used a special *E. coli* strain RP523, which contains a *hemB* gene deletion and an unknown permeability mutation that renders the bacteria heme-permeable. Using this strain, they expressed a ruthenium porphyrin substituted H-NOX (Heme Nitric oxide/Oxygen binding) as an oxygen sensor. Ruthenium porphyrin is an ideal cofactor for O_2 sensing because it exhibits O_2 sensitive phosphorescence. Jasanoff and co-workers[20] substituted a manganese porphyrin in P450 BM3 by co-expressing the heme transporter ChuA. The manganese porphyrin substituted P450 exhibits enhanced relaxivity and ligand sensitivity, and is used as a MRI contrast agent. These strategies for the *in vivo* incorporation of metalloporphyrins into proteins will expand the application of metalloporphyrin chemistry and facilitate functional metalloprotein design.

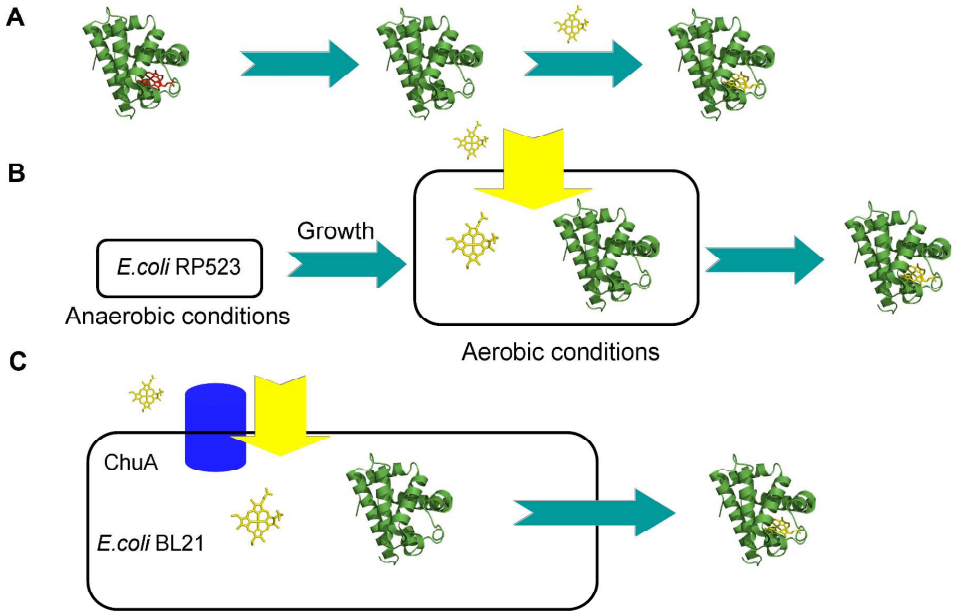


Figure 2. Methods for metalloporphyrin incorporation into proteins. A: Heme proteins are first denatured to remove heme, then they are refolded in the presence of unnatural metalloporphyrins. B: The usage of a special *E. coli* strain RP523 for metalloporphyrin incorporation into proteins. C: Co-expression of the protein of interest with the heme transporter, ChuA, renders the cell membrane permeable to heme and metalloprophyrin.

Site-specific incorporation of metal-chelating unnatural amino acid into proteins

While tremendous progress has been made in metalloprotein mechanistic studies and engineering through rational and computational design of metalloproteins, the genetic code expansion method can significantly improve our ability to design metalloproteins through site-specific incorporation of metal-chelating unnatural amino acids (UAA), UAA with optimal pK_a and reduction potential, and UAA bearing boorthogonal groups to facilitate site-specific cofactor ligation. Here, we briefly review these methods.

Nearly all proteins are translated by ribosomes using amino acids as building blocks. Each amino acid of a protein is specified by nucleotide triplet, or codon, in mRNA. In most creatures, the standard genetic code includes 61 sense codons specifying 20 natural amino acids, and 3 nonsense codons (or stop codons) serving as termination signals. A sense codon can be read by an appropriate aminoacyl-tRNA molecule, which carries a corresponding amino acid to the ribosome. Stop codons cannot be read by tRNA molecules. Instead, they can be recognized by release factors, which terminate translation processes. However, in some archaea, a special amino acid, pyrrolysine (Pyl), is genetically encoded by the amber stop codon TAG, using a unique aminoacyl-tRNA synthetase (aaRS)/tRNA pair.[22, 23] In order to site-specifically incorporate UAA in a host organism in response to the TAG codon, an exogenous aaRS/tRNA pair which does not interact with endogenous aaRS/tRNA pairs needs to be used[2, 22-25] (Figure 3). The exogenous aaRS can selectively aminoacylate an amber suppressor tRNA, which then decodes the TAG codon as a unique UAA. The exogenous aaRS must not recognize any other endogenous tRNA or natural

amino acids, and the exogenous amber suppressor tRNA must not be reorganized by any other endogenous aaRS. If all the above conditions are satisfied, the exogenous aaRS/tRNA can be defined as an "orthogonal pair", and can be used for site-specific UAA incorporation *in vivo*. The first engineered orthogonal pair is a tyrosyl-tRNA synthetase and cognate nonsense suppressor tRNA pair derived from the archaea *Methanocaldococcus jannaschii* (*Mj*TyrRS/*Mj*$tRNA_{CUA}^{Tyr}$).[24] Later, pyrrolysine aaRS/tRNA pairs from *Methanosarcina barkeri* were demonstrated to be orthogonal in *E. coli* and a host of other organisms.[23] In addition, orthogonal leucine tRNA synthetase, phosphoserine tRNA synthetase, and tryptophan tRNA synthetase pair have been used for UAA incorporation.[2, 23, 25] Overall, more than 100 UAAs have been genetically encoded in response to TAG amber codon.[2, 23, 25] This expanded toolkit has greatly empowered us to carry out more sophisticated metalloprotein design.

Genetically encoding metal-chelating amino acids is an important strategy for metalloprotein design. The natural amino acids such as histidine, cysteine, aspartate, glutamate, tyrosine and methionine usually bind metal ions in a mono-dentate manner. To ensure strong and selective metal ion binding, three to five amino acid ligands are necessary. Due to the great variability of metal ion valence state, coordination sphere, and complexity in amino acid sidechain packing, designing metal-binding sites in proteins is a trial-and-error process requiring multiple rounds of optimization.[1]

By contrast, due to the high affinity between metal chelating UAA and metal ions, it is possible to easily introduce a metal ion at any site of a protein with minimal perturbation to protein structure. Since the metal binding site is genetically encoded by one single codon, it also allows for high-throughput screening and directed evolution in living cells. To date, several metal binding UAA have been added to the genetic code of *E.coli*, including (2,2-bipyridin-5yl) alanine(Bpy-Ala), (8-hydroxyquinolin-3-yl) alanine, 2-amino-3-[4-hydroxy-3-(1H-pyrazol-1-yl)phenyl] propanoic acid (pyTyr) and 2-amino-3-(8-hydroxyquinolin-5-yl) propanoic acid (HqAla)[3-6, 26, 27]. Their variance in metal selectivity, binding affinity and chemical properties greatly expand the toolkit for metalloenzyme design.

Bpy-Ala (Figure 4) was the first metal chelating UAA introduced into the genetic code of *E. coli*,[5] using an evolved tyrosine tRNA/tyrosyl-tRNA synthetase pair. The bipyridine side chain has strong binding affinity to transition-metal ions such as $Fe^{2+/3+}$,Cu^{2+}, $Co^{2+/3+}$, and $Ru^{2+/3+}$. Using this method, a DNA-cleaving enzyme with sequence specificity was designed[3] by site-specifically incorporating Bpy-Ala into a DNA binding protein CAP. While Bpy-Ala incorporation does not affect the DNA binding affinity or specificity of CAP, on addition of Cu^{2+} or Fe^{2+} and reducing regent to protein, the Fenton reaction occurs near the CAP binding site and a double-stranded DNA substrate can be cleaved with sequence specificity. This study demonstrates that the site-specific incorporation of metal-chelating UAA provide simple and efficient routes to bestow proteins with novel functions, without significantly perturbation of their structures.

A computational method for designing metalloprotein harboring metal-chelating UAA was recently developed.[27] In this study, Mills et al. extended the Rosetta design methodology to design metalloproteins which contain Bpy-Ala as a ligand for metal ions. The resulting

Bpy-Ala-containing protein contains a high affinity binding site with octahedral geometry that binds divalent metal ions. X-ray crystallographic analysis demonstrated that the designed protein folds as expected. Further development along this route should allow for the more precise computational design of more complex metalloenzymes.

8-hydroxylquinoline (8HQ) is an important metal chelator in both laboratory research and industry. While the genetic incorporation of (8-hydroxyquinolin-3-yl) alanine was first demonstrated by Schultz and colleagues,[4] the synthesis of this amino acid required more than 5 steps, including heavy metal catalysts, strong acid and base, carcinogenic solvents and multiple column purification steps, which limited its application. Recently, our group reported the efficient synthesis and genetic incorporation of 2-amino-3-(8-hydroxyquinolin-5-yl)-propanoic acid (HqAla).[7] HqAla is synthesized through a one-step enzymatic reaction, catalyzed by a *Citrobacter freundii* tyrosine phenol lyase (TPL) mutant. The evolved TPL mutant can efficiently catalyze the synthesis of HqAla using only cheap starting materials and water as solvent, at 40% yield. We have also recently reported the efficient synthesis of the metal-chelating amino acid pyTyr,[6] which requires only two steps and no heavy metal catalysts or column purification. We are well aware that most laboratories which specialize in protein biochemistry and protein engineering lack organic synthesis facilities. Due to the great ease of HqAla and pyTyr synthesis, they can be easily prepared in any biology or biochemistry laboratory. Combined with computational design and directed evolution methods, many more novel metalloproteins can be designed using these two metal-chelating UAA.

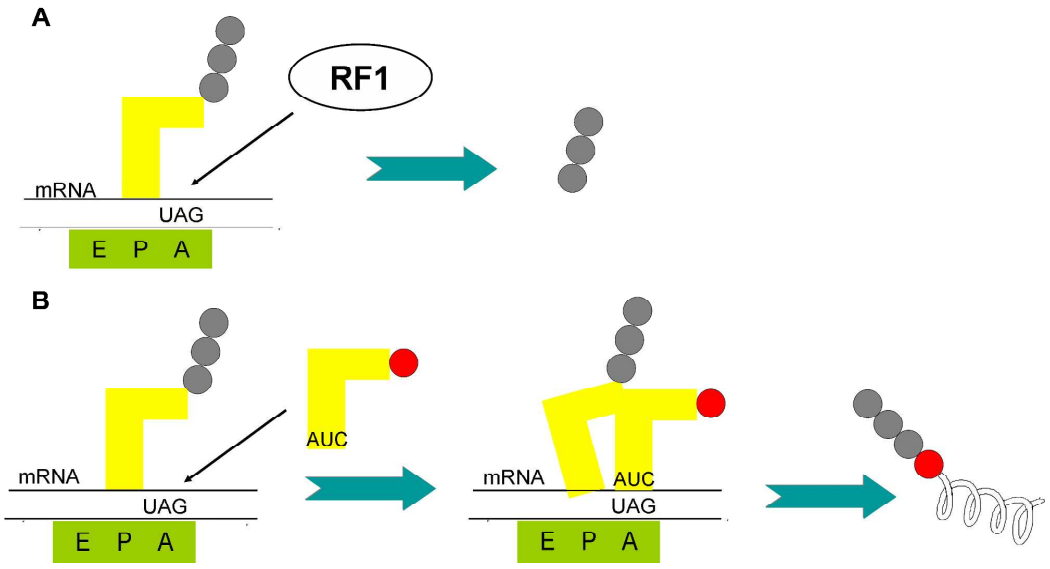


Figure 3. Site-specific UAA incorporation using the cellular protein translation system and genetic code expansion. A: In the absence of evolved tRNA/aaRS or UAA, the amber codon is recognized by release factor 1 (RF1), and the translation process is terminated. B: In the presence of evolved tRNA/aaRS and UAA, amino-acylated amber suppressor tRNA can recognize the UAG codon, allowing the ribosome to decode UAG codon as an UAA.

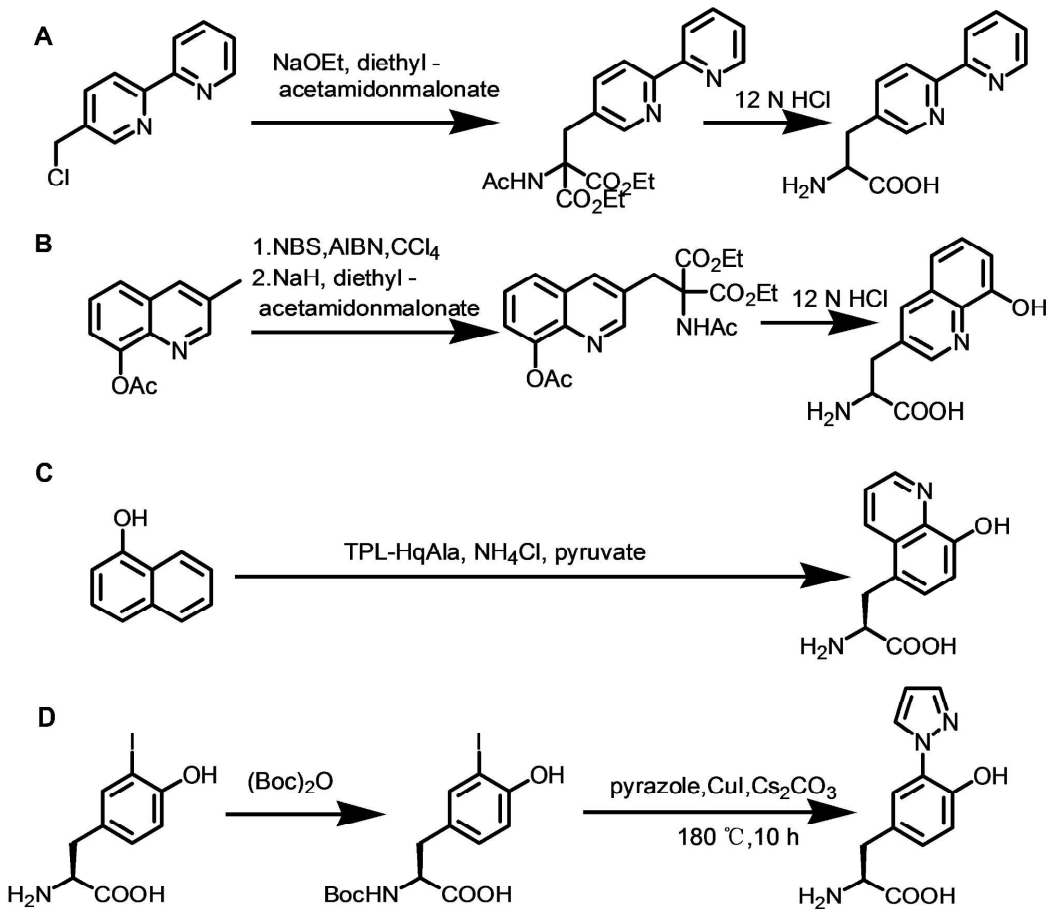


Figure 4. Synthesis routes of metal-chelating UAAs. A: (2,2-bipyridin-5yl) alanine (Bpy-Ala); B: (8-hydroxyquinolin-3-yl) alanine; C: 2-amino-3-(8-hydroxyquinolin-5-yl) propanoic acid (Hq-Ala); D: 2-amino-3-[4-hydroxy-3- (1H-pyrazol-1-yl)phenyl] propanoic acid (pyTyr).

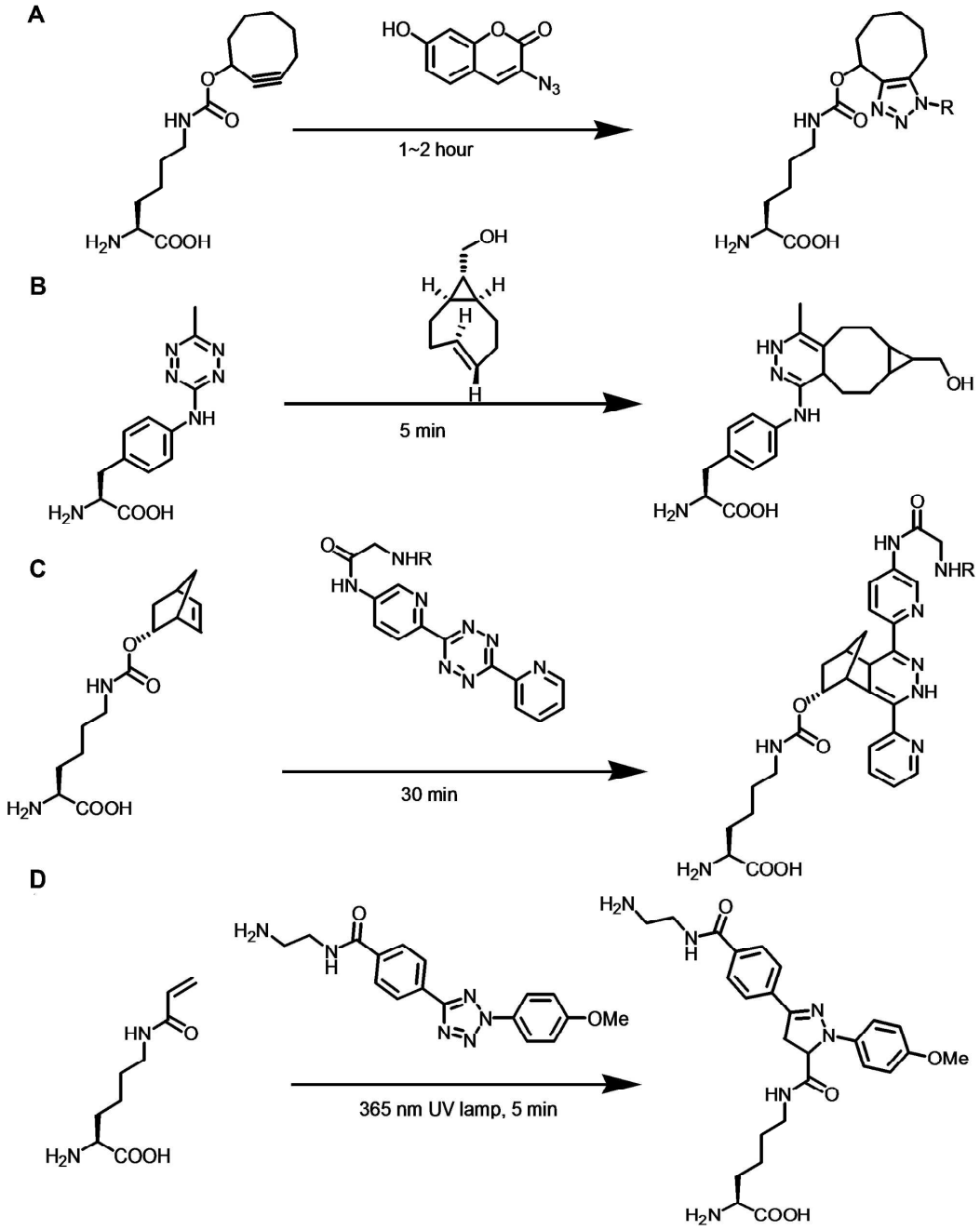


Figure 5. Genetically encoded UAAs facilitate bioorthogonal chemical ligations. A: Genetically encoded UAA bearing strained cyclooctyne group facilitate azide/alkyne ligation reaction. B: Genetically encoded UAA bearing tetrazine group facilitate tetrazine/trans-cyclooctene ligation reaction. C: Genetically encoded UAA bearing norbornene group facilitate norbornene/tetrazine ligation reaction. D: Genetically encoded UAA bearing cyclopropene or electron-deficient alkene group facilitate photoclick reaction.

Development of bioorthgonal chemical reactions for cofactor incorporation

While several non-covalent and covalent methods for introducing cofactors into metalloprotein have been reported,[19, 28-30] such methods lack specificity *in vivo*, and not all protein scaffolds can be used. In order to perform site-specific ligation to any target protein *in vivo*, numerous unnatural amino acids bearing bioorthogonal functional groups have been genetically encoded (Figure 5).

The term click reaction, first introduced by Sharpless and coworkers[31], refers to reactions with high selectivity, high yield and fast reaction rate. Bioorthogonal reactions[8, 32, 33] also require high selectivity, high yield, and high reaction rate, and that both reactants do not react with natural biomolecules. Bioorthogonal reactions are essential tools for biomolecule site-selective modification. One popular bioorthogonal pair is azide and alkyne. Both azide and alkyne are rare in natural biological systems, and do not react with natural bio-molecules. They can perform a [3+2] cycloaddition reaction in the presence of Cu(I) ion, in aqueous solution at room temperature and neutral pH. Alternatively, azide/alkyne click chemistry can also be driven by ring-strain, without requirment for the toxic copper catalyst.[8]

Bioorthogonal reactions together with the genetic code expansion technique have provided exciting new means for protein labeling in general, and site-specific ligations of cofactors to metalloproteins in particular. Toward these goals, UAA bearing small bioorthogonal groups such as azide, alkyne or cyclopropene moieties, and larger bioorthogonal groups, such as cyclooctyne, aryltetrazole, and norbornene, *trans*-cyclooctene and aryltetrazine have been genetically encoded,[23, 25, 33] allowing for selective conjugation through tetrazole-alkene photoclick chemistry, azide-alkyne click chemistry, or reverse-electron demand Diels-Alder reactions. The main advantages of the photoclick reaction are its fast rate (up to 50 $M^{-1}s^{-1}$), and the ease of alkene and tetrazole synthesis. We have reported the site-specific incorporation of *p*-(2-tetrazole)phenylalanine (*p*-Tpa),[34] N-ε-(1-methylcycloprop-2-enecarboxamido)lysine (CpK)[35] and N-ε-acryllysine (AcrK)[36] in *E. coli*, mammalian cells and *Arabidopsis thaliana*. Subsequent photoirradiation of labeled proteins with UV lamps facilitates selective conjugation with dimethyl fumarate or diaryltetrazole, respectively. Cyclopropene is a small chemical group with inherent high strain energy, but is stable in the biological environment. Much of the strain energy is released in the cycloaddition reaction, and therefore the photoclick reaction is significantly accelerated. Another route to accelerate the photoclick reaction is to use an electron-deficient alkene, such as N-ε-acryllysine (AcrK). The genetically encoded AcrK provide an especially attractive means for metalloprotein ligation since the synthesis of AcrK require only two steps.

Probing the function of Tyr-His crosslink in CcO using a myoglobin model

Cytochrome *c* oxidase (CcO)[10-13] is a membrane protein residing in the inner membrane of the mitochondrion, the powerhouse of the eukaryotic cell. CcO mediates the transfer of four electrons from the four molecules of ferrocytochrome c from the inner membrane space of the mitochondrion to O_2 bound to the enzyme embedded in the inner membrane, reducing the dioxygen to form two molecules of water with four protons derived from the matrix. In addition,

up to four protons can be pumped across the matrix to the inner membrane space during the turnover. This process is electrogenic and contributes to the electrostatic potential and pH gradient across the inner membrane of the mitochondrion.

CcO is a metalloenzyme with four redox-active metal cofactors (Figure 6): (i) a low-potential heme A, with two axial imidazole nitrogen ligands from histidines; (ii) Cu_A, a low-potential dinuclear $Cu^{1.5}Cu^{1.5}$ dicopper site bridged by two cysteine sulfurs, with each Cu coordinated to an additional nitrogen from an imidazole of histidine and a more remote atom, a methionine sulfur or a peptide carbonyl from a glutamate residue; (iii) a second heme *a* center, usually called cytochrome a_3, with on axial histidine ligand and an open coordination site for axial ligands such as O_2 and CO; and (iv) Cu_B, ligated to three imidazole nitrogen from histidines, one of which is covalently linked to a tyrosine. The location and ligand structures of these metal cofactors are highlighted in the X-ray structure of bovine heart CcO shown in Figure 6. The bovine enzyme is a highly complex protein structure with 13 subunits, with the 3 largest subunits encoded by mitochrondria genes. All the redox-active metal cofactors are associated with subunits 1 and 2. The X-ray structures of CcO have been now been determined for many species, and the ligand structures of the redox-active centers are absolutely conserved.

As a redox-linked proton pump, the flow of electrons from one redox-active metal center to another together with the dioxygen chemistry are key to the functioning of the enzyme as a molecular machine. Initially, two electrons are passed from two ferrocytochrome *c* molecules through the Cu_A and cytochrome a sites to the cytochrome a_3-Cu_B binuclear center, reducing Fe_{a3}^{3+} and Cu_B^{2+} to Fe_{a3}^{2+} and Cu_B^{+}, respectively. The hydroxide ligand bridging the binuclear center is then protonated and the resulting water molecule is displaced by O_2. The dioxygen is rapidly reduced with two electrons from the binuclear site and the Fe of the cytochrome a_3 is converted to the oxyferryl (Fe^{4+}=O) species upon cleavage of the O-O bond with input of an additional electron and a proton from the histidine-linked tyrosine forming a tyrosyl radical and a hydroxide ion that is bound to Cu_B. To continue the cycle, a third and a fourth electron from two additional ferrocytochrome *c*s are passed through Cu_A and cytochrome *a* to the cytochrome a_3-Cu_B binuclear center to reduce the Fe^{4+}=O species to Fe^{3+}-OH and the tyrosyl radical back to tyrosine in two successive proton-coupled electron transfer steps, followed by protonation of the Cu_B bound OH^- to release the water molecule. The net process is that four ferrocytochrome *c*s are used, along with four protons to reduce O_2 to two water molecules.

Due to the great importance of CcO in aerobic respiration, and its potential application in fuel cells, tremendous work has been directed towards the construction inorganic complex and macromolecular models of CcO.[1, 10-12] While tremendous progress has been made, important questions about the CcO mechanism still remain. We are particularly interested in the conserved covalent cross-link between C6 of Tyr244 and Nε2 of His240 (Tyr-His crosslink，Figure 6C), revealed first by X-ray structure and then confirmed by biochemical studies.[10-12] While the presence of Tyr-His crosslink is firmly established in CcO, its exact function is still not understood. Studies using synthetic model compounds[37] suggest that the Tyr-His crosslink decreases the pK_a and the reduction potential of phenol, facilitating proton and electron delivery to the oxygen substrate, thus accelerating the reaction rates and specificity by preventing ROS formation.

Despite these advances, the precise function of Tyr-His crosslink remains unknown, since no study has been able to compare the functions of His/Tyr residues in the same protein with and without cross-linking.

The genetic code expansion strategy is especially suitable for addressing this important question. We synthesized (S)-2-amino-3-(4-hydroxy-3-(1H-imidazol-1-yl)phenyl)propanoic acid (imiTyr，Figure 7B)[38] which contains the ortho-imidazole–phenol linkage, and genetically encoded it in myoglobin, at the 33rd position. In addition, we introduced Leu29His mutation, which together with the native His64 residue form a copper ion-binding site resembling that of Cu_B in CcO. This mutant is termed imiTyrCu_BMb. We found that imiTyrCu_BMb is capable of performing more than 1000 cycles of O_2 reduction to water, while producing less than 6% of reactive oxygen species (ROS).[38] By contrast, removal of the copper ion, or mutation of imiTyr33 to Tyr33, drastically lowers the O_2 reduction activity and results in a loss of specificity, and the generation of ROS (Figure 7C). These results demonstrate that by genetically incorporating the UAA imiTyr into Mb, we have successfully designed a functional CcO model which catalyzes the selective and efficient reduction of O_2 to water. Importantly, we demonstrate that both the Tyr-His crosslink and copper ion is important for the selective and efficient O2 reduction to water. Our designed enzyme harbors UAA imiTyr which is highly similar to the Tyr-His ligand found in Cu_B site of CcO, and thus serves as an ideal model for a more detailed understanding of CcO. We can use the imiTyrCu_BMb to develop strategies to ascertain how CcO uses the protein scaffold to tune the dioxygen chemistry of the dinuclear site toward generation of the high-potential dioxygen intermediates required for proton pumping. Regardless, our experiment has already provided insights into the chemistry of dioxygen reduction with potential applications for the design of fuel cells.

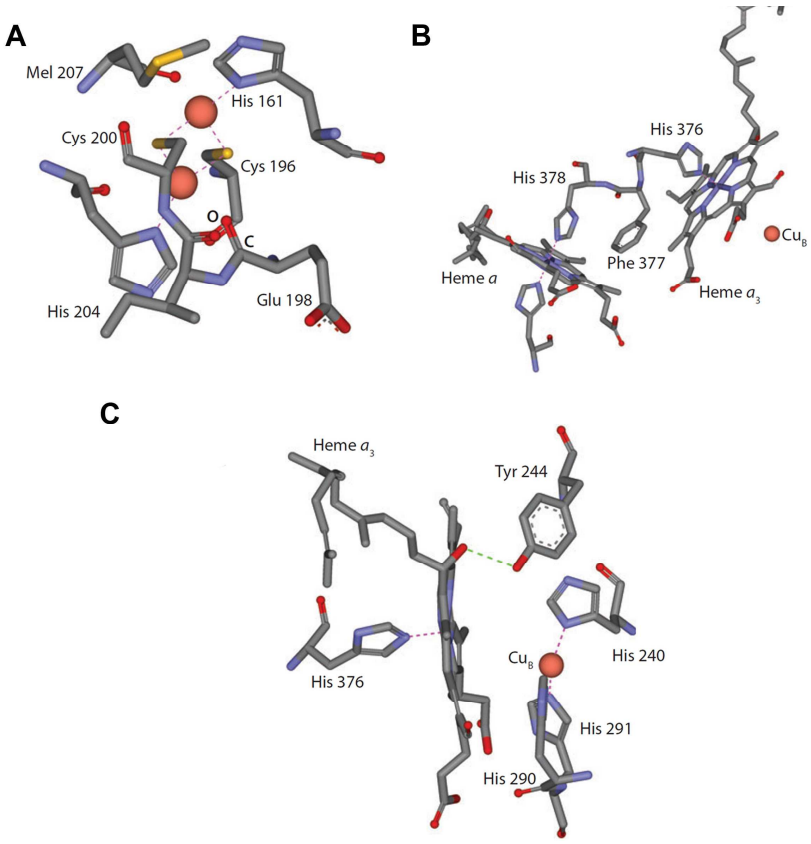


Figure 6. Ligand structures of (A) CuA, (B) cytochrome a, and cytochrome a_3/Cu_B (C) in bovine heart CcO.[10]

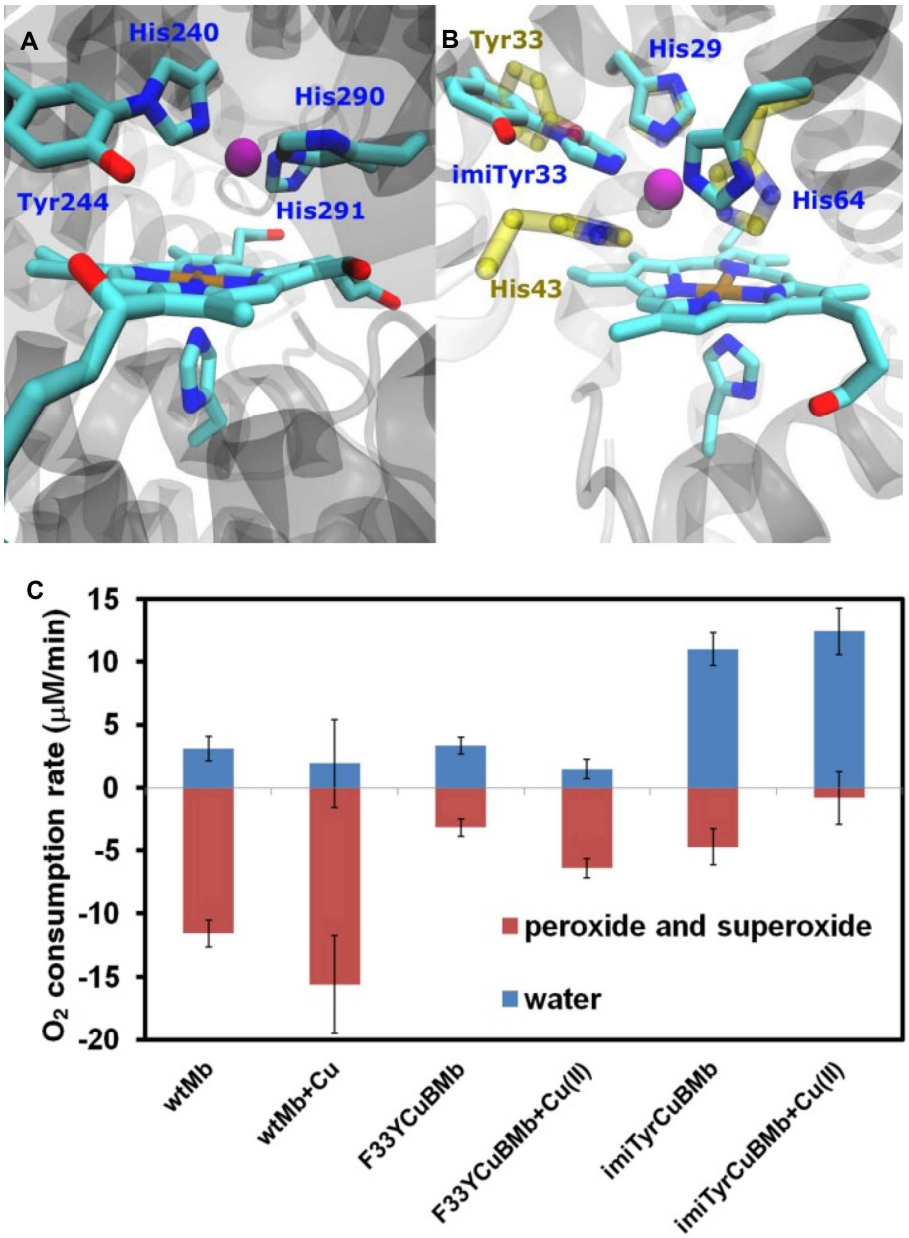


Figure 7. A: Cytochrome a_3/Cu_B site of cytochrome c oxidase (bovine numbering). B: Structure model overlay of imiTyrCu_BMb (cyan) and F33YCu_BMb (yellow). C: Rates of oxygen reduction to form either water (blue) or H_2O_2 (red) catalyzed by 6 μM wtMb, F33YCu_BMb, or imiTyrCu_BMb, in the presence or absence of 6 μM of Cu^{2+}, with 0.6 mM TMPD and 6 mM ascorbic acid as reductants.[38] Reproduced from ref[38] with permission.

Probing the function of Tyr-Cys crosslink in metalloenzymes using a myoglobin model

The thioether-bonded tyrosine-cysteine crosslink (Tyr-Cys, Figure 8 and 9) is essential for many metalloenzymes, including galactoase oxidase (GO) and glyoxal oxidase, cysteine dioxygenase (CDO), sulfite reductase (NirA), and *T. nitratireducens* cytochrome *c* nitrite reductase (TvNiR).[39-42] In these enzymes, a covalent bond is present between the C3 ring carbon atom of a tyrosine sidechain and the Sγ atom of a neighboring cysteine sidechain.

While the presence of the Tyr-Cys cofactor is firmly established in GO, CDO, NirA and TvNiR, its function is still not fully understood. Synthetic model compounds for Tyr-Cys have been made,[41] and studies on these model compounds have indicated that the Tyr-Cys crosslink significantly decreases the pK_a and reduction potential of the phenol group, thereby facilitating proton and electron transfer between the substrates and the Tyr-Cys cofactor, which is essential for optimizing enzyme activity.

Cytochrome *c* nitrite reductase (NiR)[40] catalyzes the 6-electron reduction of nitrite to ammonia. The presence of hydroxylamine as an intermediate is demonstrated by the results that NiR catalyzes hydroxylamine reduction to ammonia with high efficiency. Recent structural studies[41] revealed that in TvNiR, a Tyr-Cys crosslink between residues Tyr303 and Cys305 is present, which is responsible for donating an electron and a proton to the hydroxylamine intermediate. While this crosslink does not cause structural changes in the TvNiR active site, formation of the Tyr-Cys crosslink results in significant lowering of the pK_a and reduction potential of Tyr303. It was suggested that these changes are responsible for the higher activity of TvNiR compared to NiR which does not contain the Tyr-Cys crosslink.

To directly test this hypothesis, we have constructed a functional model of TvNiR, called MtTyrMb,[42] using myoglobin as a scaffold protein (Figure 8). MtTyrMb harbors the UAA MtTyr at position 33, and a His29 mutation (these residues correspond to similar positions relative to heme as residues Tyr303 and His316 of TvNiR) (Figure 8). Our results show that MtTyrMb and TyrMb (which has a normal Tyr residue at position 33) exhibited similar K_m values for the hydroxylamine substrate, but that the activity of MtTyrMb was four-fold higher than that of TyrMb. These results provide direct evidence that 3-methylthio substitution on a tyrosine residue can increase enzyme activity. Since myoglobin is much easier to produce, and easier to characterize than aforementioned metallo enzymes, MtTyrMb serves as an ideal model to further elucidate the mechanism of NiR activity. There is strong interest in evolving galactose oxidase to obtain mutants with selective activity towards monosaccharides other than galactose in order to facilitate polysaccharide characterization. The genetic incorporation of MtTyr should facilitate evolution of metalloenzymes with specific activity towards various monosaccharides, by using well-characterized protein scaffolds as a copper ligand and by applying directed evolution methods.

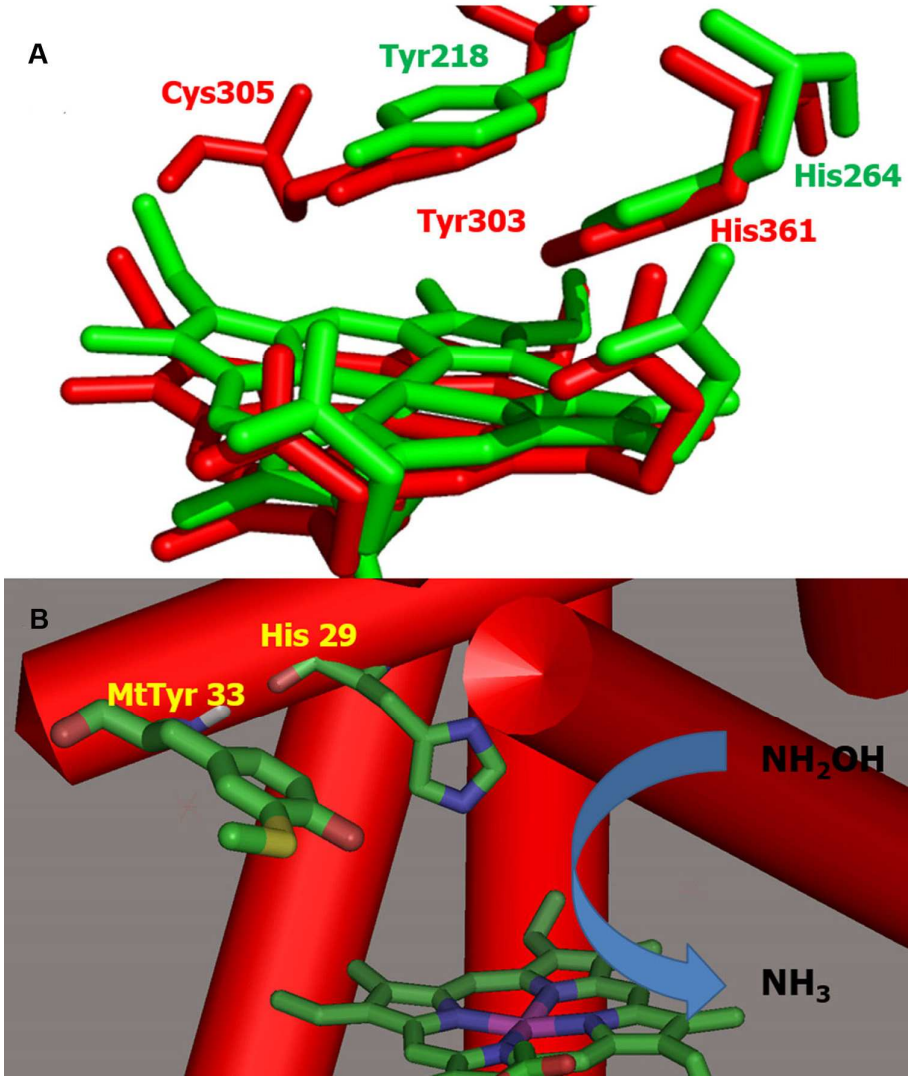


Figure 8. A. Structural alignment of *Wolinella succinogenes* cytochrome *c* nitrite reductase (green，PDB code 1FS7) and TvNiR (red, PDB code 3F29). Note the presence of a Tyr-Cys crosslink between Tyr303 and Cys305 in TvNiR. B. Structural model of MtTyrMb, constructed based on the crystal structure of the F33Y/L29H/F43H mutant myoglobin (PDB code 4FWX). Residues MtTyr33 and His29 are in similar positions in the MtTyrMb active site as Tyr303 and His361 in TvNiR.[43]

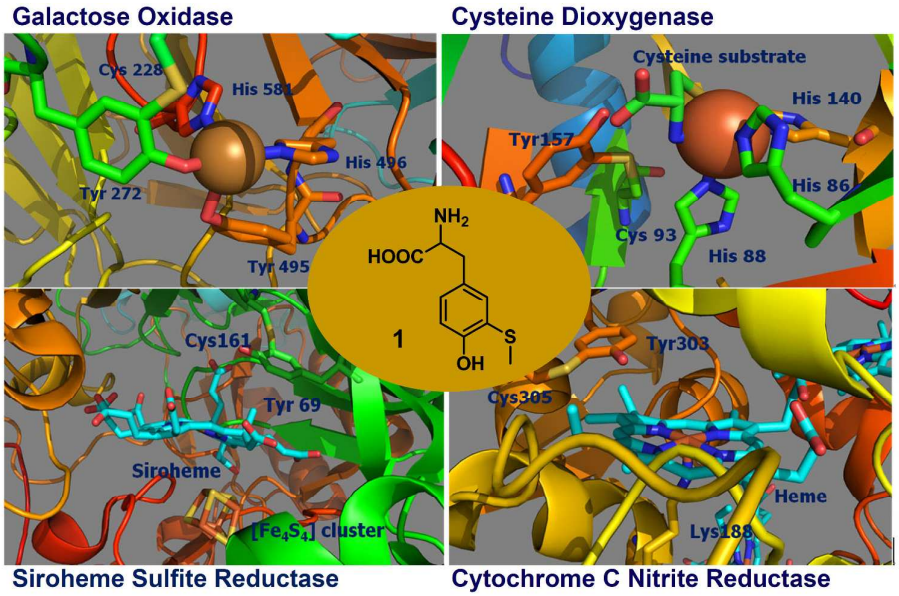


Figure 9. Tyr-cys crosslink found in various metalloenzymes.

Metalloprotein sensor design

The two key methods of metalloprotein sensor design are the incorporation of metal-binding sites in fluorescent proteins and the modification of natural metalloprotein sensors, such as the olfactory receptors.

Fluorescent protein (FP) sensors are important components of the chemical tool box for biological studies. FP sensors for metal ions, pH, and post-translational modifications (PTM) have been developed, which are essential for untangling their function *in vivo*. These FP sensors mostly rely on fluorescence resonance energy transfer (FRET) or protonation/deprotonation of the of the GFP fluorophore.[43] Despite their broad applications, these sensors typically exhibit less than two-fold fluorescence intensity change before and after analyte binding. By contrast, photoinduced electron transfer (PET) is the most widely used mechanism for fluorescent sensor design, due to the significantly larger fluorescence change (10- to 200-fold) upon analyte binding. While PET is essential for photosynthesis and small molecule sensor design, to date there have been no reports of PET protein sensors, due to the lack of methods to install PET probes on proteins.

To circumvent these limitations, we genetically incoporated the metal-chelating amino acid (S)-2-amino-3-(4-hydroxy-3-(1H-pyrazol-1-yl)phenyl)propanoic acid (pyTyr) into green fluorescent protein (GFP).[6] We demonstrated that Cu^{2+} selectively and strongly binds to pyTyr with sub-nanomolar affinity, forming a pyTyr/Cu^{2+} complex (termed pyTyrCu) wherein the pyrazole ring, copper ion, and phenol ring are co-planar (Figure 10). We show that pyTyrCu has a redox potential of 250 mV vs. NHE, which is much lower than that of tyrosine or tryptophan. In contrast to natural amino acid or dopa, 3-aminotyrosine, difluorotyorsine which are used as electron *donors* in ET experiments, pyTyrCu is an electron *acceptor*. These unique properties should allow us to design metalloprotein sensors in ways that were previously unachievable.

Using this method, we showed that PET between the GFP chromophore and pyTyrCu occurs within one nanosecond and in a distance-dependent manner, consistent with the Marcus equation. We also show that fluorescent dyes such as Cy5 and Cy7 are selectively quenched by pyTyrCu, but not by any natural amino acids. Since GFP and Cy5 are among the most important labels used for fluorescence spectroscopy, we believe that the GFP:pyTyrCu and Cy5:pyTyrCu fluorophore/quencher pairs will facilitate the design of protein PET sensors and conformational dynamics in complex biomolecules.

Another route to bestow fluorescent proteins with metal binding ability is to modify the GFP chromophore through genetic code expansion (Figure 11). By substituting Tyr of the fluorophore in diverse fluorescent proteins (FP) with the metal-chelating UAA HqAla, we showed that UAA incorporation red-shifts excitation and emission spectra by more than 30 nm.[7] The X-ray structure of superfolder GFP (sfGFP) bearing HqAla in 66^{th} position reveals a novel 8-hydroxyquinolin-imidazolinone (HQI) chromophore (Figure 11) which has a much expanded conjugated π-system. Our results show that HqAla incorporation into the FP fluorophore bestows it with metal-chelating and metal ion sensing abilities. Among all biologically relevant metal ions, only zinc binding to HQI causes a significant increase (7.2 fold) in fluorescence. This Zn^{2+} selective FP sensor was then used for Zn^{2+} sensing *in vivo*. This is the first report showing that a genetically encoded metal-chelating UAA can bind transition metal ions in living cells. In the future, it would be of great interest to use HqAla to perform directed evolution in living cells in order to obtain novel metalloenzymes. Since HqAla can be synthesized through an enzymatic route in one step and in high yield, and it can bind strongly to most transition metal ions including lanthanide ions, the technological barriers which have restricted the application of the genetic encoded metal-chelating UAA have now been overcome. With this method, the future is now even brighter for metalloprotein sensor engineering, metalloenzyme design, and protein NMR using paramagnetic metal ions.

It would also be of great interest to directly modify natural metalloprotein sensors for the direct visualization of analytes. Mammals have around one thousand olfactory receptor (OR) genes and can detect more than ten thousand scents.[44] The ORs are 7-helix membrane proteins, with an odorant-binding site in the periplasmic side and a G protein-binding site on the cytoplasmic side. Once odorants bind to an OR, it undergoes structural changes which trigger G protein activation, eventually leading to neuronal activity. We previously reported identification of a consensus metal-binding site between the fourth and fifth helix (4–5 loop) of the ORs[45] and proposed that Cu^{2+} or Zn^{2+} binding is required for OR's high sensitivity and selectivity to amines and thiols (which can be detected at parts per billion). Zhuang and coworkers[46] later demonstrated that ORs do indeed bind metals. We have synthesized a pentapeptide that contains the consensus metal binding site from OR and have found that it not only has high affinity for Cu^{2+} and Zn^{2+} ions, but also that it undergoes a transition to an α-helical structure upon metal ion binding. Based on these results, we proposed a "shuttlecock" mechanism for the structural change in ORs upon odorant binding. This mechanism involves membrane penetration of the 4–5 loop after residue charge neutralization by metal ion binding (Figure 12). This mechanism could be adapted to convert ORs into artificial metalloprotein sensors through genetic code expansion, allowing direct visualization of analytes. By site-specifically incorporating AcrK or CpK into the 4^{th} helix of OR, we can insert

solvent polarity sensitive fluorescent dyes at specific positions in OR. Upon odorant binding, the 4th helix is expelled from the membrane. This environmental change can be directly sensed by the solvent polarity sensitive fluorescent dye attached to the 4th helix, eliciting a fluorescence signal.

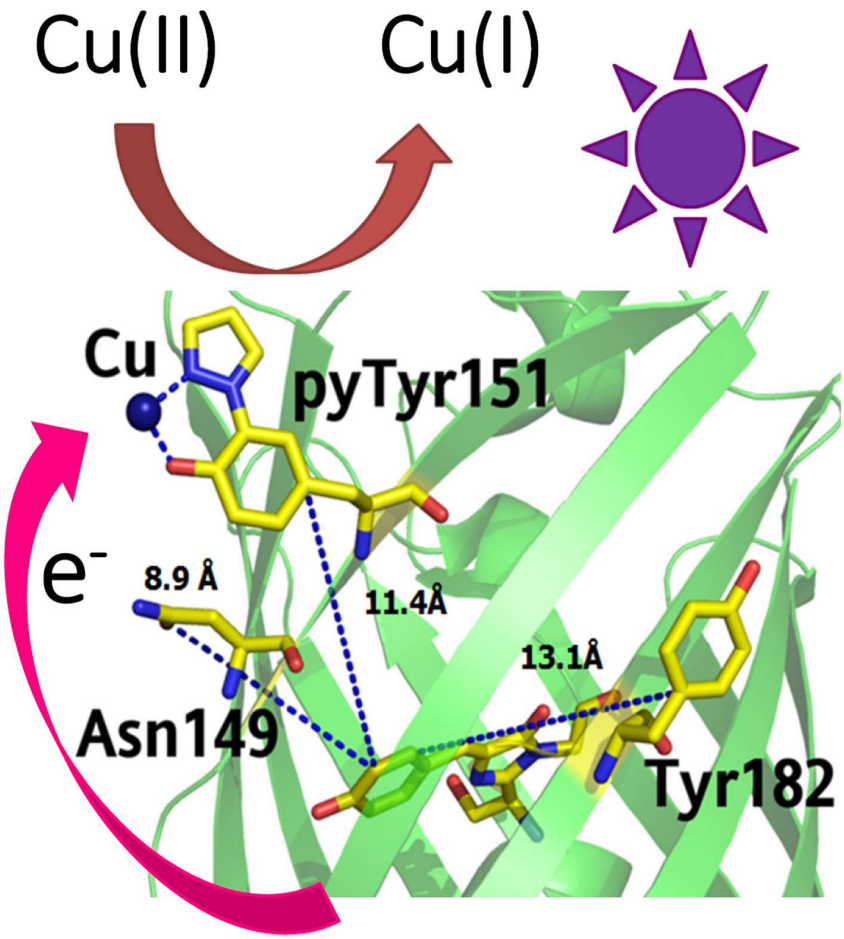


Figure 10: Efficient PET from the GFP chromophore to the Cu(II) ion bound to pyTyr151 site.

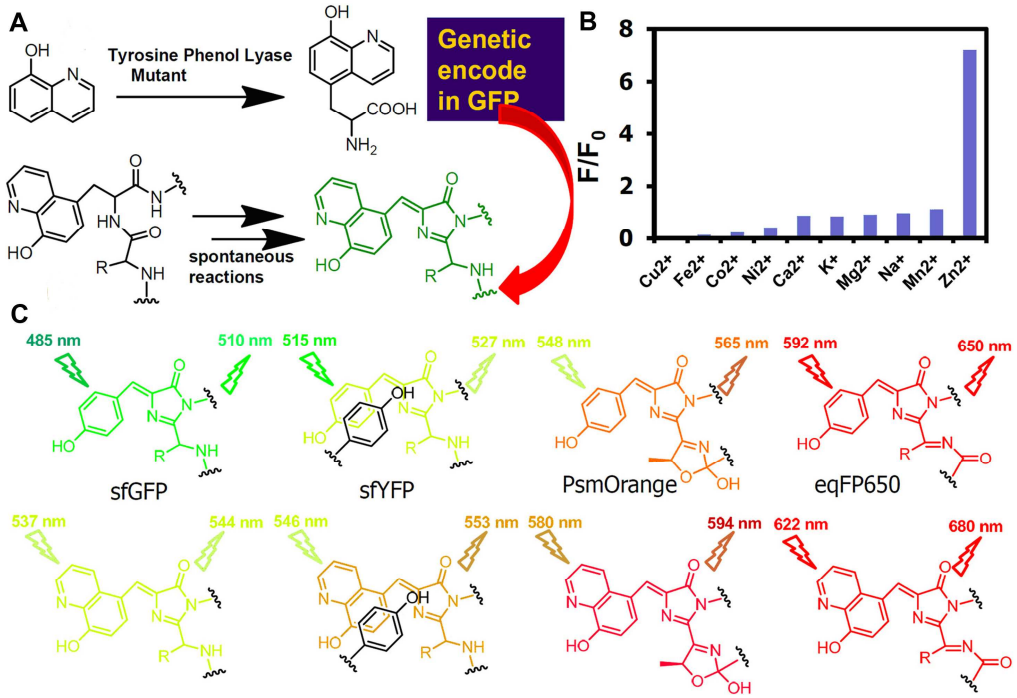


Figure 11. A: Synthesis route of the HqAla unnatural amino acid, and maturation mechanism of the sfGFP-66-HqAla chromophore. B: Relative fluorescence emission intensity cpsfGFP-66-HqAla(λex = 537 nm; λem = 544 nm), in the presence of various transition metal ions. C: Chemical structures sfGFP-66-HqAla, sfYFP-66-HqAla, PsmOrange-72-HqAla, and eqFP650-67-HqAla chromophores. Excitation and emission maxima of each chromophore are shown on left and right sides, respectively.

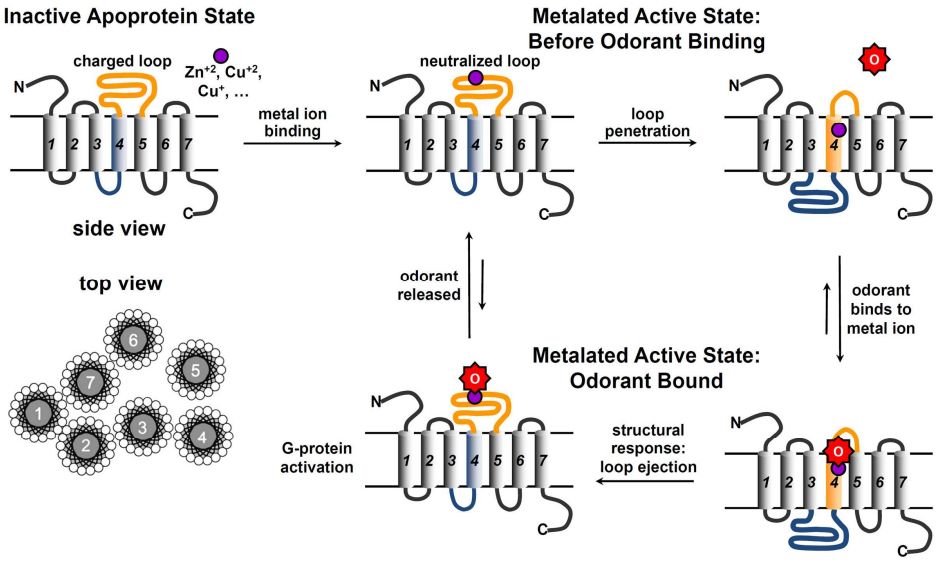


Figure 12. The proposed mechanism of olfactory receptor response *via* a shuttlecock mechanism. In the absence of odorant, the metal binding site is in helical conformation. Upon odorant binding, the primary response is helix ejection (Lower Left).[45] Reproduced from ref[45] with permission.

Conclusions and outlook

Metalloproteins are especially adept at catalyzing reactions requiring multiple electron and proton transfers as shown in Scheme 1. The key for high activity and selectivity, as previous studies have shown, is the precise arrangement of metal ions, cofactors and amino acids with optimal pK_as and reduction potentials to serve as proton and electron relays for substrate activation. However, as shown in Figure 13, metalloproteins such as photosystem II (PSII) can be extremely complex. They can have molecular weight of a few hundred thousand kilo Daltons, contain many cofactors and metal ions, and require multiple proteins for cofactor insertion. This complexity has greatly hindered our understanding of the basic catalytic mechanism of reactions exemplified in Scheme 1, and our ability to expand the functions of metalloproteins for their applications in synthetic biology.

As shown in Figure 13, the basic mechanism for PSII's ability to photo-catalytically oxidize water into O_2 is the following:[47, 48] upon light-absorption, the P680 pigment is promoted to a photo-excited state, it then transfer electrons to protein-bound quinones, resulting in a highly oxidative $P680^+$ cation. Electrons are then transferred rapidly from the Mn_4CaO_5 cluster (oxygen evolution center or OEC) via the Y_z-His190D hopping station. Repetition of the above process four times result in the loss of four electrons from OEC, giving rise to the highly oxidative S4 state, which is capable of oxidizing two water molecules to O_2.

With the genetic code expansion strategies we have discussed, designing a miniature PSII protein mimic might be within reach. It should be much easier to assemble the metal clusters using the

metal-chelating UAA pyTyr and HqAla and a suitable protein scaffold than to design these clusters *de novo*. While the proton and electron delivery properties of the Y_z-His190D hopping station are difficult to recapitulate in small molecule model compounds, it should be possible to install UAA to mimic the proton-couple electron transfer (PCET) property in a model protein. Finally, by genetically encoding Acrk or CpK, it should be possible to install photosensitizers site-specifically on a model protein, allowing it to perform efficient photo-induced charge separation. If all the above elements can be put together in a model protein in a proper manner, a small protein mimic with activity equal or greater than PSII may be designed.

Another area of interest is the multiple-heme redox enzymes and electron-conducting nanowires.[49] As shown in Figure 13, multiple-heme proteins are used by microorganisms such as *Shewanella oneidensis* MR-1 and *Geobacter* species to reduce minerals such as iron oxide or manganese oxide, and to perform multiple electron reductions such as nitrite reduction to ammonia, and tetrathionate reduction. They can even form electron-conducting nanowires allow electrons to flow from extracellular minerals all the way to the inner memebrane (Figure 13B). Here, heme groups sever as “capacitors” to facilitate the rapid multiple electron redox reaction, and mediators to allow for electrons to transfer the long distance across the periplasmic space.

It would be of great interest to use the genetic code expansion strategy to probe the mechanism of these multiple-heme redox enzymes, and engineer microbes to convert organic and inorganic matter into electricity, with potential applications in microbial fuel cells. Microbial fuel cells, devices in which microorganisms catalyse the release of electrons from organic matter and transfer them to electrochemically active electron carriers, are attractive sources of energy because they are carbon-neutral, can oxidise a diverse range of inexpensive and impure fuels (such as organic waste) and can be operated at any temperature at which microbial life is possible.

With the genetic code expansion method, chemists are no longer limited by the 20 natural amino acids to construct model metalloproteins for mechanistic studies, catalytic and sensing applications. Through the site-specific genetic incorporation of metal-chelating UAAs, UAAs with fine-tunable pK_as and reduction potentials, and selective ligation of cofactors using bioorthgonal reactions, our ability to design metalloenzymes to mimic and expand the function of natural enzymes is much improved. The concerted efforts of bioinorganic chemists and synthetic biologists raise the likelihood of further exciting developments in this rapidly progressing field in the near future, resulting in the production of cell factories and live-cell sensors with unprecedented efficiency and selectivity.

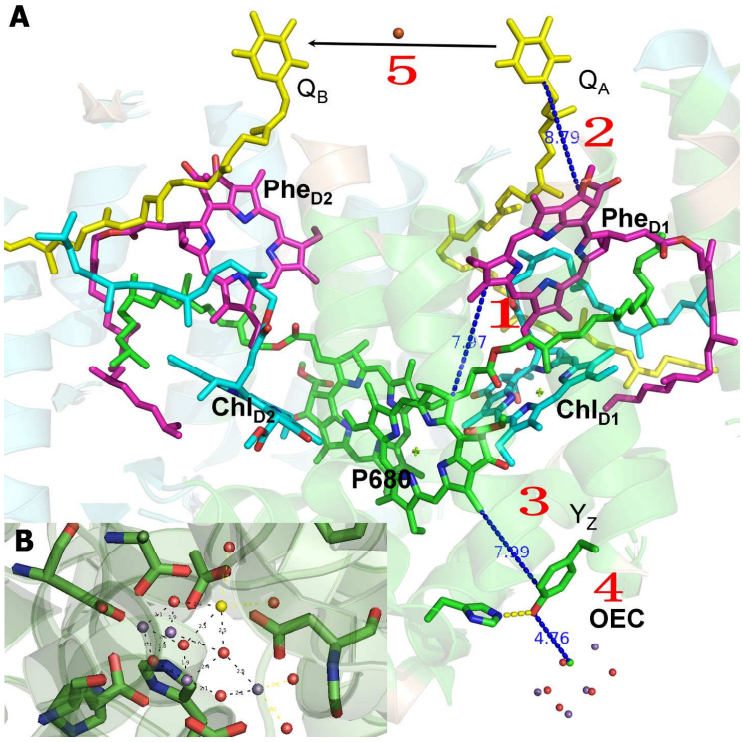


Figure 13. A: Photo-induced electron transfer pathway of PSII. Black arrows represent electron transfer directions. Numbers represent the sequence of each electron transfer event. (PDB code 3arc). The Y_z phenol oxygen atom and His190D imidazole nitrogen atom form strong hydrogen bonds.（B）Structure of Mn_4CaO_5 cluster. Purple sphere represents magnesium atom; red sphere represents oxygen atom; yellow sphere represents calcium atom.

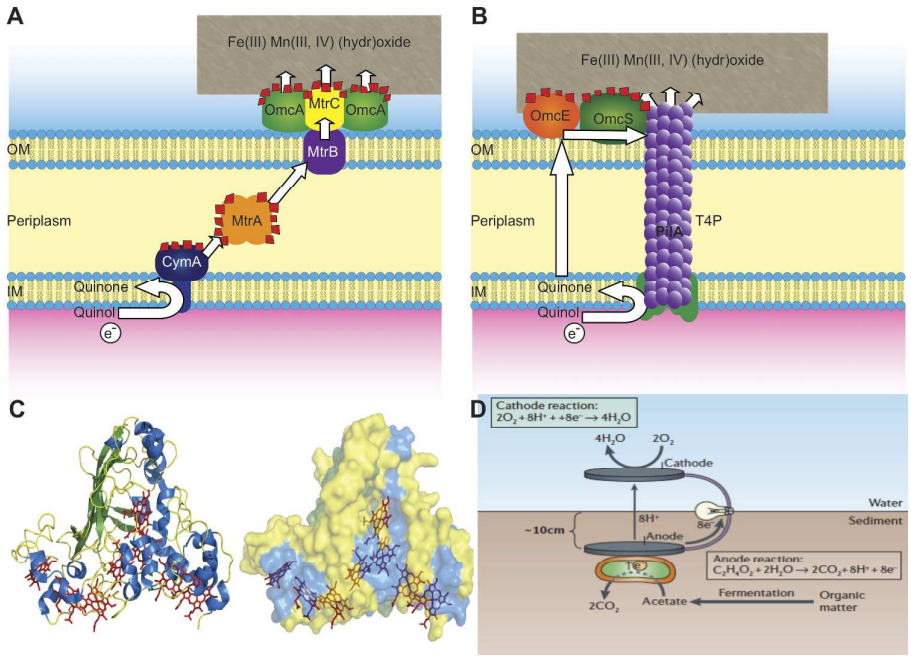


Figure 14. Multi-heme cytochromes from Gram-negative bacteria. A: *S. oneidensis* MR-1 transfer

electrons from the inner membrane quinone/quinol pool to solid substrates outside the cell via multi-heme cytochromes. The key proteins involved in this process are shown here with red squares indicating the number of heme groups in each protein. B: *G. sulfurreducens* produce conducting type IV pili that function alongside multi-heme cytochromes to deliver electrons to extracellular acceptors. Again, the number of heme groups in each cytochrome is indicated with red squares. C: Crystal structure of OTR (Octahaem tetrathionate reductase) from *S. oneidensis* (PDB code 1SP3) colored by secondary structure (α-helices, blue; β-sheets, green; loops/coils, yellow). Left, cartoon representation showing the unusual fold of OTR. Right, surface representation showing relative positions and orientations of heme groups. D: Conceptual model showing how *Geobacter* species could be used to create sediment bacterial fuel cells.[49] Reproduced from ref[49] with permission.

Acknowledgement: We gratefully acknowledge the Major State Basic Research Program of China (2010CB912301, 2011CBA00800), National Science Foundation of China (91313301, 21325211), CAS grant (KJZD-EW-L01) to Y. Y and J. Y. W.

Reference

1. Y. Lu, N. Yeung, N. Sieracki and N. M. Marshall, *Nature*, 2009, **460**, 855-862.
2. C. C. Liu and P. G. Schultz, *Ann. Rev. Biochem.*, 2010, **79**, 413-444.
3. H. S. Lee and P. G. Schultz, *J. Am. Chem. Soc*, 2008, **130**, 13194-13196.
4. H. S. Lee, G. Spraggon, P. G. Schultz and F. Wang, *J. Am. Chem. Soc*., 2009, **131**, 2481-2483.
5. J. Xie, W. Liu and P. G. Schultz, *Angew.Chem. Intl. Ed.*, 2007, **119**, 9399-9402.
6. X. H. Liu, J. S. Li, J. S. Dong, C. Hu, W. M. Gong and J. Y. Wang, *Angew. Chem. Intl. Ed.*, 2012, **51**, 10261-10265.
7. X. H. Liu, J. S. Li, C. Hu, Q. Zhou, W. Zhang, M. R. Hu, J. Z. Zhou and J. Y. Wang, *Angew. Chem. Intl. Ed.*, 2013, **52**, 4805-4809.
8. J. C. Jewett and C. R. Bertozzi, *Chem. Soc. Rev.*, 2010, **39**, 1272-1275.
9. R. K. Lim and Q. Lin, *Acc. Chem. Res.*, 2011, **44**, 828-839.
10. S. I. Chan, *Annual review of biophysics*, 2009, **38**, 1-27.
11. V. R. Kaila, M. I. Verkhovsky and M. Wikstrom, *Chem. Rev.*, 2010, **110**, 7062-7081.
12. E. Kim, E. E. Chufa´n, K. Kamaraj and K. D. Karlin, *Chem. Rev.*, 2004, **104**, 1077-1135.
13. J. P. Collman, N. K. Devaraj, R. A. Decreau, Y. Yang, Y. L. Yan, W. Ebina, T. A. Eberspacher and C. E. D. Chidsey, *Science*, 2007, **315**, 1565-1568.
14. T. D. Pfister, A. Y. Mirarefi, A. J. Gengenbach, X. Zhao, C. Danstrom, N. Conatser, Y. G. Gao, H. Robinson, C. F. Zukoski, A. H. J. Wang and Y. Lu, *J. Biol. Inorg. Chem.*, 2007, **12**, 126-137.
15. N. Yeung, Y. W. Lin, Y. G. Gao, X. Zhao, B. S. Russell, L. Lei, K. D. Miner, H. Robinson and Y. Lu, *Nature*, 2009, **462**, 1079-1082.
16. J. A. Sigman, B. C. Kwok and Y. Lu, *J. Am. Chem. Soc.*, 2000, **122**, 8192-8196.
17. T. Hino, Y. Matsumoto, S. Nagano, H. Sugimoto, Y. Fukumori, T. Murata, S. Iwata and Y. Shiro, *Science*, 2010, **330**, 1666-1670.
18. P. S. Coelho, E. M. Brustad, A. Kannan and F. H. Arnold, *Science*, 2013, **339**, 307-310.

19. K. Oohora, Y. Kihira, E. Mizohata, T. Inoue and T. Hayashi, *J. Am. Chem. Soc.*, 2013, **135**, 17282-17285.
20. V. S. Lelyveld, E. Brustad, F. H. Arnold and A. Jasanoff, *J. Am. Chem. Soc*, 2011, **133**, 649-651.
21. M. B. Winter, E. J. McLaurin, S. Y. Reece, C. Olea, D. G. Nocera and M. A. Marletta, *J. Am. Chem. Soc.*, 2010, **132**, 5582-+.
22. B. Hao, W. M. Gong, T. K. Ferguson, C. M. James, J. A. Krzycki and M. K. Chan, *Science*, 2002, **296**, 1462-1466.
23. L. Davis and J. W. Chin, *Nat. Rev. Mol. Cell Bio.*, 2012, **13**, 168-182.
24. L. Wang, A. Brock, B. Herberich and P. G. Schultz, *Science*, 2001, **292**, 498-500.
25. C. H. Kim, J. Y. Axup and P. G. Schultz, *Curr. Opin. Chem. Bio.*, 2013, **17**, 412-419.
26. T. S. Young, I. Ahmad, J. A. Yin and P. G. Schultz, *J. Mol. Biol.*, 2010, **395**, 361-374.
27. J. H. Mills, S. D. Khare, J. M. Bolduc, F. Forouhar, V. K. Mulligan, S. Lew, J. Seetharaman, L. Tong, B. L. Stoddard and D. Baker, *J. Am. Chem. Soc*, 2013, **135**, 13393-13399.
28. J. R. Carey, S. K. Ma, T. D. Pfister, D. K. Garner, H. K. Kim, J. A. Abramite, Z. L. Wang, Z. J. Guo and Y. Lu, *J. Am. Chem. Soc.*, 2004, **126**, 10812-10813.
29. N. H. Tran, D. Nguyen, S. Dwaraknath, S. Mahadevan, G. Chavez, A. Nguyen, T. Dao, S. Mullen, T. A. Nguyen and L. E. Cheruzel, *J. Am. Chem. Soc*, 2013, **135**, 14484-14487.
30. T. K. Hyster, L. Knorr, T. R. Ward and T. Rovis, *Science*, 2012, **338**, 500-503.
31. H. C. Kolb, M. G. Finn and K. B. Sharpless, *Angew. Chem. Int. Ed.*, 2001, **40**, 2005-2022.
32. T. Plass, S. Milles, C. Koehler, C. Schultz and E. A. Lemke, *Angew. Chem. Intl. Ed.*, 2011, **50**, 3878-3881.
33. Z. Y. Hao, S. L. Hong, X. Chen and P. R. Chen, *Acc. Chem. Res.*, 2011, **44**, 742-751.
34. J. Y. Wang, W. Zhang, W. J. Song, Y. Z. Wang, Z. P. Yu, J. S. Li, M. H. Wu, L. Wang, J. Y. Zang and Q. Lin, *J. Am. Chem. Soc.*, 2010, **132**, 14812-14818.
35. Z. P. Yu, Y. C. Pan, Z. Y. Wang, J. Y. Wang and Q. Lin, *Angew. Chem. Intl. Ed.*, 2012, **51**, 10600-10604.
36. F. H. Li, H. Zhang, Y. Sun, Y. C. Pan, J. Z. Zhou and J. Y. Wang, *Angew. Chem. Intl. Ed.*, 2013, **52**, 9700-9704.
37. K. M. McCauley, J. M. Vrtis, J. Dupont and W. A. v. d. Donk, *J. Am. Chem. Soc.*, 2000, **122**, 2403-2404.
38. X. Liu, Y. Yu, C. Hu, W. Zhang, Y. Lu and J. Wang, *Angew. Chem. Intl. Ed.*, 2012, **51**, 4312-4316.
39. M. S. Rogers, R. n. Hurtado-Guerrero, S. J. Firbank, M. A. Halcrow, D. M. Dooley, S. E. V. Phillips, P. F. Knowles and M. J. McPherson, *Biochemistry*, 2008, **47**, 10428-10439.
40. K. M. Polyakov, K. M. Boyko, T. V. Tikhonova, A. Slutsky, A. N. Antipov, R. A. Zvyagilskaya, A. N. Popov, G. P. Bourenkov, V. S. Lamzin and V. O. Popov, *J. Mol. Biol.*, 2009, **389**, 846-862.
41. P. Verma, R. C. Pratt, T. Storr, E. C. Wasinger and T. D. Stack, *Proc. Natl. Acad. Sci.*, 2011, **108**, 18600-18605.
42. Q. Zhou, M. Hu, W. Zhang, L. Jiang, S. Perrett, J. Zhou and J. Wang, *Angew. Chem. Intl. Ed.*, 2013, **52**, 1203-1207.
43. R. H. Newman, M. D. Fosbrink and J. Zhang, *Chem. Rev.*, 2011, **111**, 3614-3666.
44. L. Buck and R. Axel, *Cell*, 1991, **65**, 175-187.

45. J. Wang, Z. A. Luthey-Schulten and K. S. Suslick, *Proc. Natl. Acad. Sci.*, 2003, **100**, 3035-3039.
46. X. Duan, E. Block, Z. Li, T. Connelly, J. Zhang, Z. Huang, X. Su, Y. Pan, L. Wu, Q. Chi, S. Thomas, S. Zhang, M. Ma, H. Matsunami, G. Q. Chen and H. Zhuang, *Proc. Natl. Acad. Sci.*, 2012, **109**, 3492-3497.
47. T. J. Wydrzynski and K. Satoh, *Photosystem II : the light-driven water:plastoquinone oxidoreductase*, Springer, Dordrecht, 2005.
48. Y. Umena, K. Kawakami, J. R. Shen and N. Kamiya, *Nature*, 2011, **473**, 55-60.
49. D. R. Lovley, *Nat. Rev. Microbio.*, 2006, **4**, 497-508.