

STRUCTURAL BIOINFORMATICS STUDIES OF INTEGRAL TRANSMEMBRANE ENZYME SpMMO COMPLEX, C560, CYB, AND DHSD AND THEIR ALPHAFOLD3-PREDICTED WATER-SOLUBLE QTY VARIANTS

Jiayu Wang¹, Emily Pan² and Shuguang Zhang^{3,*}

¹Phillips Exeter Academy, 20 Main St, Exeter, NH 03833, USA

²Massachusetts Institute of Technology; 77 Massachusetts Avenue, Cambridge, MA 02139, USA

³Media Lab, Massachusetts Institute of Technology; 77 Massachusetts Avenue, Cambridge, MA 02139, USA

Abstract

The QTY (glutamine, threonine, tyrosine) code is a simple protein engineering and design tool that systematically replaces the hydrophobic amino acids leucine (L), isoleucine (I), valine (V), and phenylalanine (F) into the hydrophilic amino acids glutamine (Q), threonine (T), and tyrosine (Y), respectively, to enable the water-solubility of membrane proteins including membrane enzymes. In this study, we present the structural bioinformatics study of six membrane protein enzymes with experimentally determined CryoEM structures including the methane monooxygenase pMMO complex (pMOA, pMOB, pMOC), CYB, C560, DHSD, and their AlphaFold3-predicted water-soluble QTY variants. We applied the QTY code only to the transmembrane alpha-helices of the proteins. We then superposed structures of CryoEM-determined native proteins with their water-soluble QTY variants. The QTY code engineered water-soluble QTY variants demonstrate remarkable structural similarity with their native structures, with RMSD (Root Mean Square Deviation) values between 0.506Å and 0.887Å despite significant (33.33-86.61%) changes to the protein sequence of the transmembrane domains. To verify the effectiveness of our study, we show the changes in hydrophobicity surfaces between the native proteins and their QTY variants, and explain the rationale behind such change. Our structural bioinformatics studies provide key insight into differences between the hydrophobic alpha-helices and alpha-hydrophilic helices. The QTY code will likely further aid the future applications of methane oxidation enzymes. The water-soluble QTY variants could aid in studying and utilization of the enzymes in aqueous environments without detergent solubilization and stabilization. These QTY variant membrane enzymes may be particularly useful for future industrial-scale production in aqueous chassis such as filamentous algae.

Keywords

Alpha-helix, amino acid structural similarities • hydrophobic to hydrophilic conversion • membrane proteins • methane monooxygenase, QTY code

Introduction

On the planet earth, nature has evolved a set of twenty amino acids that are consisted of 10 hydrophilic amino acids and 10 hydrophobic amino acids (Figure 1, Figure S1) (1, 2, 3, 4). Hydrophilic amino acids form hydrogen bonds with water molecules, thus readily water-soluble. They include Aspartic acid (D), Glutamic acid (E), Asparagine (N), Glutamine (Q), Lysine (K), Arginine (R), Serine (S), Threonine (T), Histidine (H) and Tyrosine (Y). Conversely, hydrophobic amino acids do not form hydrogen bonds

with water molecules, thus, water-insoluble. They include Leucine (L), Isoleucine (I), Valine (V), Phenylalanine (F), Methionine (M), Tryptophan (W) and Alanine (A). Cysteine (C) and Glycine (G) are only weakly hydrophilic and hydrophobic, respectively. Proline (P) is a unique case: the nitrogen of Proline is part of the ring structure formed by the side chain. Proline is frequently found at the N-terminus of helices, but if it occurs in the middle of an alpha-helix, it forces the helix to bend (4).

* Corresponding author e-mail: shuguang@mit.edu; shuguang@media.mit.edu

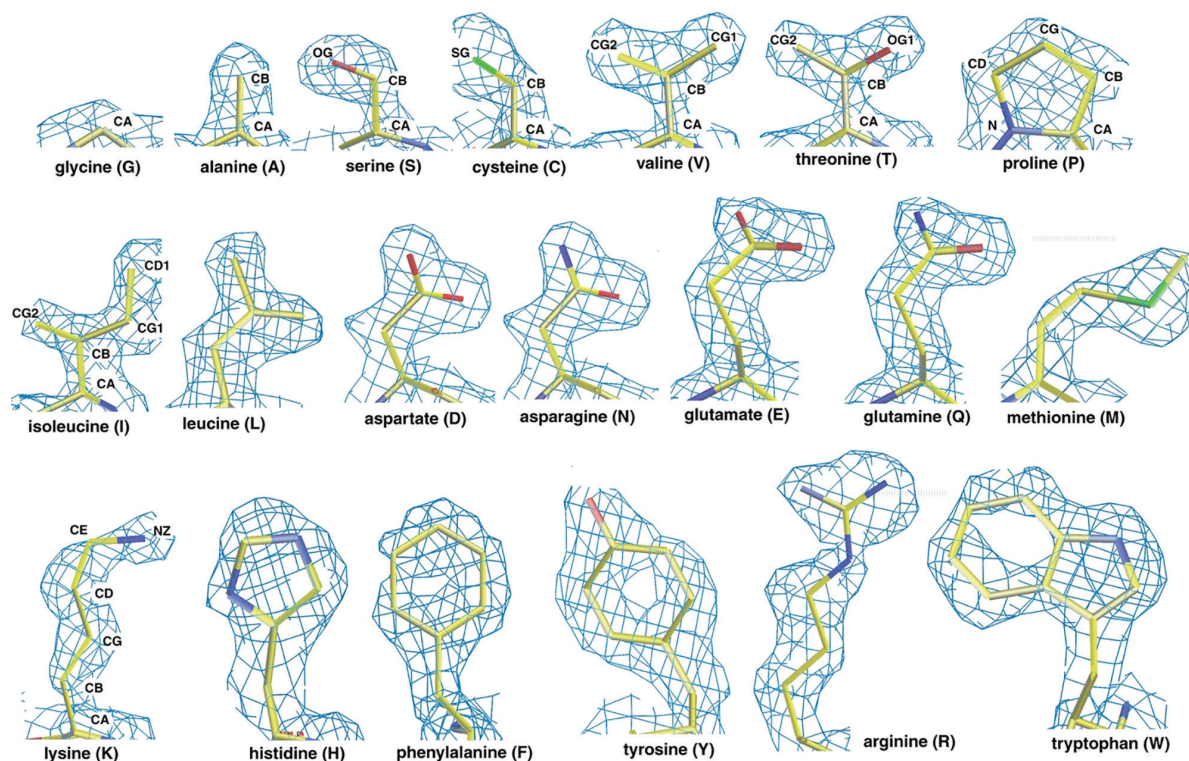


Figure 1. Experimental 1.5Å-resolution X-ray electron density maps of 20 amino acids arranged by size. This figure is provided by Dr. Mike Sawaya (UCLA), and used with permission in order to show the individual amino acid electron density maps at high resolution. The density maps demonstrate similar shapes of V and T; L, D, N, E and Q, and F and Y. Please see Dr. Mike Sawaya's original website: <http://people.mbi.ucla.edu/sawaya/m230d/Modelbuilding/modelbuilding.html> (courtesy of Dr. Michael R. Sawaya of University of California, Los Angeles, CA, USA).

Proteins can be generally divided into two classes: Class I is hydrophilic and Class II is hydrophobic (1, 2, 3, 4). Class I hydrophilic proteins include water-soluble proteins that reside in the cytoplasm, such as hemoglobin, as well as metabolic enzymes and circulating proteins outside cells, such as growth factors, hormones, and antibodies. On the other hand, Class II hydrophobic proteins comprise the integral membrane proteins that are embedded in cellular and other membranes. They include G protein-coupled receptors (GPCRs), membrane transporters and ion channels as well as the photosynthesis machinery. The Class I proteins are generally water-soluble, and the Class II proteins, as integral membrane proteins, are generally water-insoluble. In order to solubilize Class II proteins, various detergent or surfactants are required during isolation from their lipid-bilayer membrane environment and for stabilization after removal (5, 6, 7).

Alpha-helices can be classified into three chemically distinct types (Please see Table S1). Although they are significantly different in their chemical properties and water-solubility, they have nearly identical molecular structures, namely: i) a 1.5Å

rise per amino acid, ii) a 100° rotation per amino acid, iii) 3.6 amino acids per helical turn, iv) a 5.4Å rise per helical turn, and v) the key feature, i.e. each NH-group of an amino acid forms an H-bond with the C=O group of the amino acid 4 residues away. The latter, repeated i , $i + 3$, $i + 4$ H-bonding pattern is the most prominent characteristic of an alpha-helix (1, 2, 3, 4).

The Type I alpha-helix is mostly comprised of hydrophilic amino acids on the surface, including D, E, N, H, Q, K, R, S, T and Y, and is commonly found on the outer layer in water-soluble globular proteins, and in some cases, in the inner layer of membrane helices, away from the lipid bilayer. The Type II alpha-helix is mostly comprised of hydrophobic amino acids, including L, I, V, F, M, A, W and P, and is commonly found in the helical transmembrane segments of membrane proteins and buried in the interior of water-soluble proteins. The Type III amphiphilic alpha-helix is almost equally comprised of hydrophilic and hydrophobic amino acids that are partitioned into the hydrophobic face and the hydrophilic face. The Type III alpha-helix is sometimes attached to the surface of membrane lipid bilayers, or partially buried in the hydrophobic

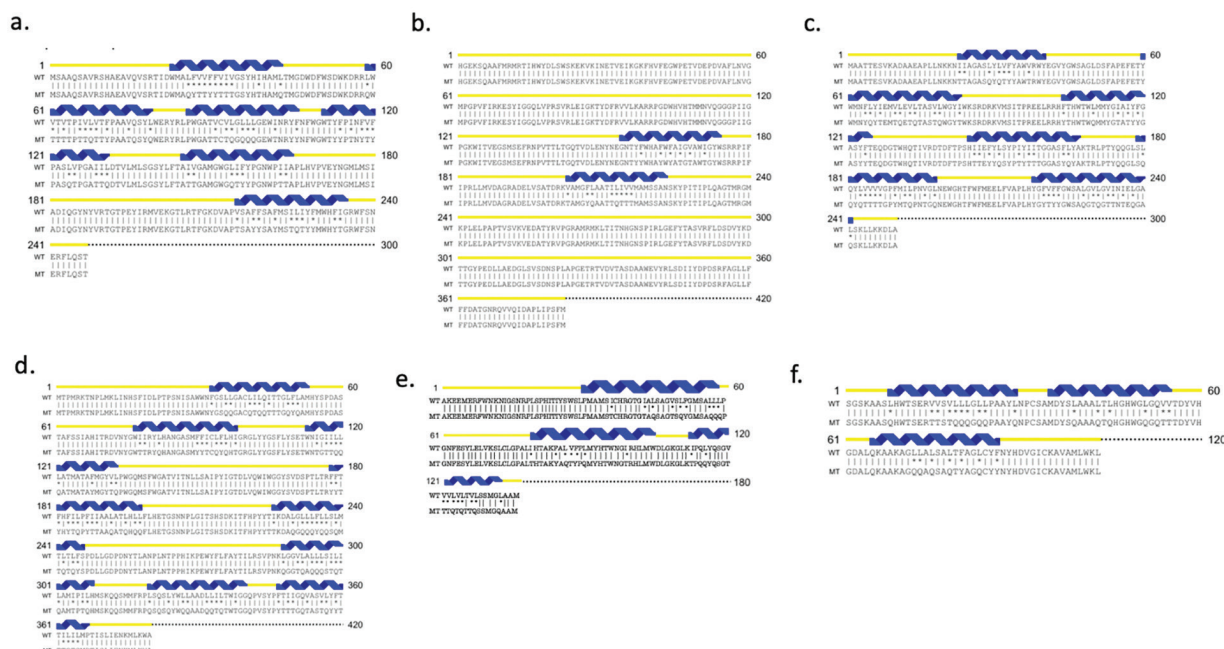


Figure 2. Protein sequence alignments of six native protein sequences with their water-soluble QTY variants. The symbols | and * indicate whether amino acids are identical or different, respectively. The alpha helices (blue) are shown above the sequences. Alignments shown include a) pMOA vs pMOA^{QTY}, b) pMOB vs pMOB^{QTY}, c) pMOC vs pMOC^{QTY}, d) CYB vs CYB^{QTY}, e) C560 vs C560^{QTY}, f) DHSD vs DHSD^{QTY}.

core and partially exposed on the surface of water-soluble globular proteins, or in the integral membrane pores that face away from the hydrophobic lipid bilayer (Table S1). Glycine (G) is referred as hydrophobic because its side chain hydrogen does not engage in H-bonds, but is only weakly hydrophobic at the same time.

Nature has evolved over time for the alpha-helices to be 3 chemically distinct types. After classifying them, we can now apply a simple QTY code to convert one type of alpha-helix, i.e., hydrophobic alpha-helix to hydrophilic ones and vice versa, apply the reverse QTY code (rQTY code) to convert hydrophilic alpha-helix to hydrophobic ones. The QTY code systematically replaces the hydrophobic amino acids L, V, I and F into hydrophilic Q, T, and Y. Here, we demonstrate the QTY code conversion of integral membrane proteins to their water-soluble counterparts.

Methanol holds potential as an alternative source of energy to petroleum and coal due to its combustion. It can be produced by the oxidation of methane, which is a greenhouse gas commonly found in the atmosphere. There are no current enzymatic catalysis at the industrial scale that efficiently perform this reaction, but naturally occurring enzymes are known to carry out this function in methanotrophs. In particular, the integral transmembrane enzyme, i.e., particulate methane monooxygenase (pMMO), which is a complex composing of

pMOA, pMOB, and pMOC found in methanotrophic bacteria, is known to perform this oxidation.

The membrane-bound pMMO is the predominant catalyst in nature to convert methane to methanol and is a three-protein complex that requires copper to carry out catalysis (8, 9, 10, 11, 12). Their X-ray crystal structures (8) and cryoEM structures (9) have been elucidated, and their 2D array structures have also been determined by cryo-electron tomography (10). pMMO is substantially more effective than natural soluble MMO (sMMO), the other of such naturally occurring catalyzers, since it has much higher affinity to molecular oxygen (O₂) and methane (CH₄) to catalyze the conversion of methane to methanol. pMMO has a lower apparent k_m than sMMO (12, 13, 14), most likely due to the high-density pMMO array structure that synergistically reinforce to capture methane (10) and carry out the subsequent enzymatic catalysis. pMMO enzyme catalytic assay has been well established (12, 13, 14, 15). In some methanotrophs under copper starvation conditions, pMMO can be expressed as ~20% total protein (16) and 80% of the pMMO is in the membranes that form high-density arrays (10, 17).

However, it is extremely difficult to express, purify and study membrane proteins like pMMO because they require detergent to stabilize them during removing from the cell membrane (18). It is impossible to produce membrane proteins in industrial

scale of kilograms. Thus, innovative methods must be found to express and purify membrane proteins in large scale.

One of us (S. Zhang) conceived and invented a simple QTY code to directly convert a hydrophobic alpha-helix to a hydrophilic alpha-helix (19). It was demonstrated that the QTY code can convert membrane proteins including G protein-coupled receptors and cytokine receptors into their water-soluble analogs that retain their biological function, namely ligand binding activities similar to the native receptors (20, 21, 22, 23). It was further demonstrated that the water-soluble QTY analog receptor CXCR4^{QTY} was used to design a highly sensitive detection device (24). Furthermore, we also carried out structural bioinformatic studies using AlphaFold2 of diverse integral membrane proteins. These studies also show that the water-soluble QTY analogs superposed very well with the corresponding native membrane proteins and yielded RMSDs of often less than 2Å including GPCRs (25), glucose transporters (26), SLC solute carrier transporters (27), ABC transporters (28), glutamate transporters (29) and monoamine transporters (30).

We recently demonstrated that a membrane-bound bacterial enzyme histidine kinase was converted to its water-soluble QTY analog, whereby the QTY enzyme histidine kinase retains its 4 different enzymatic activities (31). We therefore believe that we can also design pMMO to transform it into its water-soluble analog pMMO^{QTY} that may retain its enzymatic catalytic activity to convert methane into methanol.

Structural similarities exist between subunits of the particulate methane monooxygenase and cytochrome c oxidase II (terminal complex in the respiratory chain). Both contain a copper center, and such comparative study could yield important insights to electron transfer in oxidation (9, 31). Specifically, the subunits CYB, C560, and DHSD are known to be involved in electron transferring. CYB is a subunit of the ubiquinol-cytochrome c reductase complex (cytochrome b-c1 complex) responsible for transferring electrons to cytochrome c complex, and C560 and DHSD are both subunits of the receiving respiratory complex II succinate dehydrogenase (cytochrome c complex) (32, 33). Their structures have also been experimentally determined through CryoEM (33) and X Ray Crystallography (34, 35). However, similar problems regarding the difficulty of studying transmembrane regions exist. Thus, solubilizing the methane monooxygenase by QTY code can offer a target for further studies on electron transfer during oxidation.

AlphaFold2 and the latest version AlphaFold3 is an online program newly developed by Google DeepMind (36, 37, 38) capable of using deep learning to predict protein structures

from FASTA sequences. In comparison to traditional experimentally predicted structures, AlphaFold3 has a remarkable increase in efficiency while maintaining fairly high accuracy. Despite known limitations in flexible loops and the persisting need for experimental verification, it proves to significantly accelerate both protein structural studies and novo protein design.

Here, we present bioinformatic studies of six previously experimentally determined transmembrane proteins by others including pMOA, pMOB, pMOC, CYB, C560, and DHSD and their AlphaFold3-predicted water-soluble QTY variants. We provide superpositions of the insoluble native transporters and their soluble QTY variants. We also provide comparative structures of hydrophobic molecules with their hydrophilic QTY variants.

Results and Discussion

Rationale of the QTY code

The QTY code systematically replaces hydrophobic amino acids such as leucine (L), isoleucine (I), valine (V), and phenylalanine (F) into the hydrophilic amino acids glutamine (Q), threonine (T), and tyrosine (Y). Since the electron density maps of L vs Q, I/V vs T, and F vs Y are highly structurally similar, such conversion is able to preserve the functional structure of proteins while the transmembrane domain loses its hydrophobic characteristic.

Protein sequence alignments, isoelectric focusing (pI) and molecular weights

We aligned the protein sequences of the native pMOA, pMOB, pMOC, CYB, C560, and DHSD and their QTY variants (Figure 2). Despite overall mutations of amino acids in the protein sequence (3.66-22.89%) and notable variation in the transmembrane domain (33.33-86.61%), the isoelectric-focusing points (pI) of native proteins and engineered QTY variants remain alike, with unit differences ranging between 0-0.10. This is due to the lack of charges in amino acids Q, T, and Y at neutral pH, which minimizes impact on a protein's pI. Such differences are rather insignificant regarding surface charges and unlikely to disrupt delicate structures and functions.

Superposition of native transporters and their water-soluble QTY variants

After predicting the structures of water-soluble QTY variants with AlphaFold3, we superposed the molecular structure of six CryoEM-determined transporters and their respective QTY variants in order to directly compare their structures.

The molecular structures of native transporters are already experimentally determined for pMOA (PDB: 7EV9), pMOB (PDB: 7EV9), and pMOC (PDB: 7S4I), CYB (PDB: 8IOG), C560 (PDB: 8GS8), DHSD (PDB: 8GS8). Thus, the superpositions were carried out for pMOA vs pMOA^{QTY}, pMOB vs pMOB^{QTY}, pMOC vs pMOC^{QTY}, C560 vs C560^{QTY}, CYB vs CYB^{QTY}, and DHSD vs DHSD^{QTY}.

The experimentally determined native molecular structures and AlphaFold3-determined modified QTY structures superposed surprisingly well, with root mean square deviation (RMSD) values ranging below 1.00 Å at between 0.302 Å and 0.595 Å for all six protein pairs (Table 1 and Figure 3). Specifically, their RMSD values are as follows: pMOA vs pMOA^{QTY} (0.302Å), pMOB vs pMOB^{QTY} (0.460Å), pMOC vs pMOC^{QTY} (0.590Å), CYB vs CYB^{QTY} (0.595Å), C560 vs C560^{QTY} (0.489Å), and DHSD vs DHSD^{QTY} (0.343Å). Despite 33.33-86.61% amino acid substitutions in the transmembrane domain using the QTY code, these results demonstrate relatively similar and preserved 3-dimensional folds between the native transporters and their QTY variants.

Superpositions of AlphaFold3-predicted native transporters and their water-soluble QTY variants

We also ask how well the AlphaFold3-predicted native transporters would superpose with the AlphaFold3-predicted structures of their water-soluble QTY variants. Therefore, we carried out the superposition of the following pairs of protein structures (Figure 4) and yielded relatively low RMSD values for pMOA^{AF3} vs pMOA^{QTY}, pMOB^{AF3} vs pMOB^{QTY}, pMOC^{AF3} vs pMOC^{QTY}, CYB^{AF3} vs CYB^{QTY}, C560^{AF3} vs C560^{QTY}, and DHSD^{AF3} vs DHSD^{QTY}. These results suggest AlphaFold3 predicted and experimentally determined structures share similarity between the native proteins and their QTY variants, although the QTY

variant structures need to be experimentally verified. We also superposed the AlphaFold3-predicted native structures with their experimentally obtained CryoEM structures and calculated their RMSD to further test AlphaFold3's accuracy (Figure S8 in Supplementary Materials).

Superpositions of CryoEM structures with AlphaFold3-predicted native transporters and their water-soluble QTY variants

We also ask how the superposition will appear i) if we superpose the experimentally determined CryoEM transporter structures with ii) AlphaFold3-predicted native proteins and iii) AlphaFold3-predicted water-soluble QTY variant proteins. The superposition of the above three structures are shown in Figure 5. Structures in the superposition appear to align well, with minor differences in their residues and folding. To gain a clearer view of potential structural differences that would affect the proteins' abilities to function as transmembrane proteins, we also superposed only the transmembrane domains of the structures (Figure S9 in Supplementary Materials).

Analysis of the hydrophobic surface of native transporters and the water-soluble QTY variants

To validate the effectiveness of our protein engineering, we mapped the hydrophobicity of Cryo-EM-determined structures against AlphaFold3-predicted structures with the QTY code applied (Figure 6). The original proteins possess 2-8 transmembrane alpha-helical domains directly embedded in phospholipid bilayers, and their structures contain the highly hydrophobic amino acids L, I, V, and F. As desired, hydrophobic patches were largely reduced after systematic replacement of amino acids L, I, V, F with the more hydrophilic Q, T, and Y using the QTY code in said transmembrane regions, demonstrating the desired conversion.

Table 1. RMSD and protein characteristic comparison between six selected native proteins and their QTY variants.

	RMSD (Å)	pI	Mw (kDa)	TM Variation (%)	Overall Variation (%)
pMOA ^{CryoEM}	-	6.96	28.4252	-	-
pMOA ^{QTY}	0.302	6.95	28.7047	44.44	22.67
pMOB ^{CryoEM}	-	6.03	42.7860	-	-
pMOB ^{QTY}	0.460	6.03	42.8397	33.33	3.66
pMOC ^{CryoEM}	-	5.67	28.8514	-	-
pMOC ^{QTY}	0.590	5.67	29.1388	41.98	22.00
CYB ^{CryoEM}	-	7.82	42.7176	-	-
CYB ^{QTY}	0.595	7.76	43.3619	51.48	22.89
C560 ^{CryoEM}	-	9.30	14.9700	-	-
C560 ^{QTY}	0.489	9.20	15.1600	86.61	19.12
DHSD ^{CryoEM}	-	7.75	10.8887	-	-
DHSD ^{QTY}	0.343	7.73	11.1701	35.38	22.33

Abbreviations: RMSD = residue mean square distance; pI = isoelectric focusing Point; MW = molecular weight; TM = transmembrane; and (-) = not applicable.

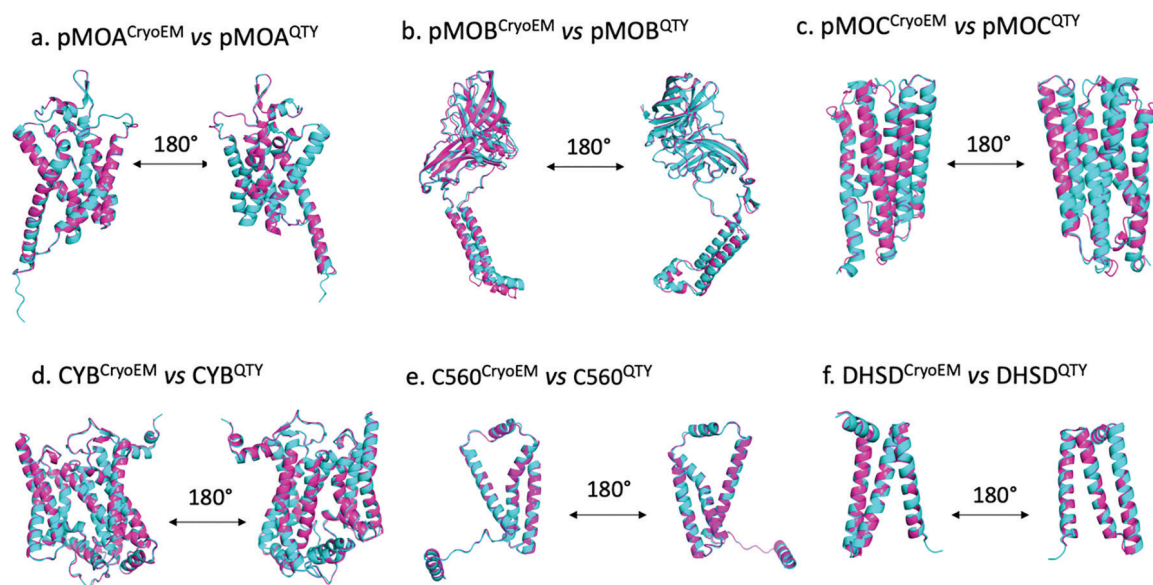


Figure 3. Superpositions of six Cryo-EM-determined structures of native oxidation enzymes and their AlphaFold3-predicted water-soluble QTY variants. The CryoEM-determined structures are obtained from the Protein Data Bank (PDB). The CryoEM structures (magenta) are superposed with their QTY variants (cyan) predicted by AlphaFold3. These superposed structures show that the native transporters and their QTY variants have very similar structures. For clarity of direct comparisons, unstructured loops in the CryoEM structures were removed in the QTY variants. a) pMOA vs pMOA^{QTY} (0.302 Å), b) pMOB vs pMOB^{QTY} (0.460 Å), c) pMOC vs pMOC^{QTY} (0.590 Å), d) CYB vs CYB^{QTY} (0.595 Å), e) C560 vs C560^{QTY} (0.489 Å), and f) DHSD vs DHSD^{QTY} (0.343 Å).

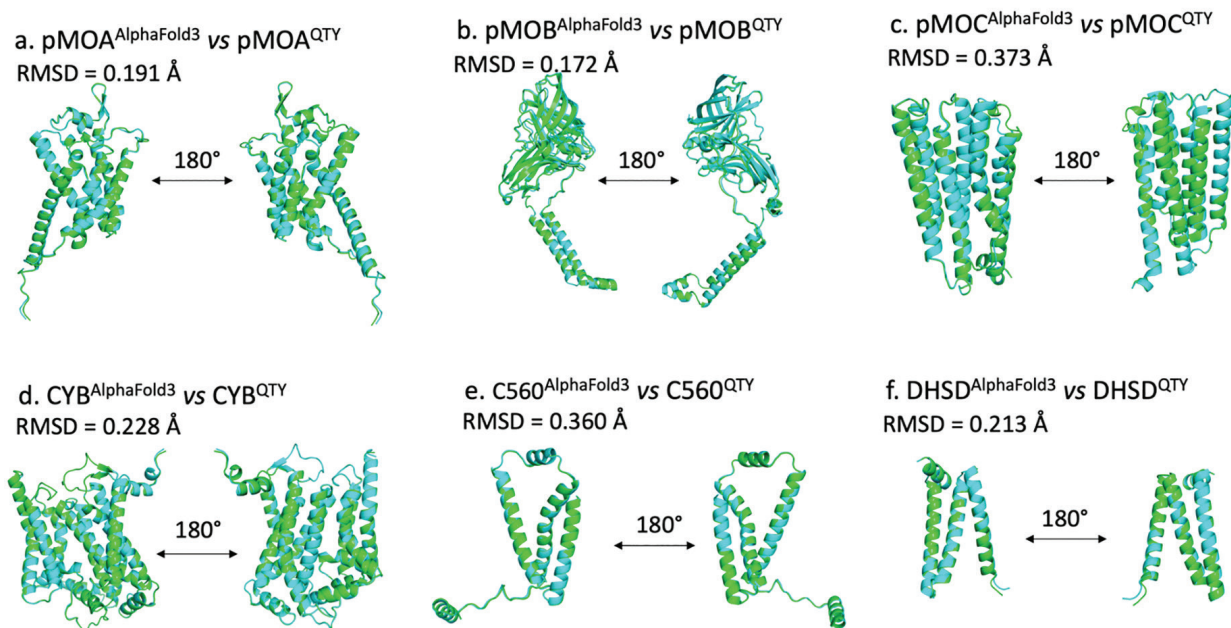


Figure 4. Superpositions of AlphaFold3-predicted structures of native and their QTY variants. Color code: green = AlphaFold3-predicted native structures; cyan = AlphaFold3-predicted water-soluble QTY variants. a) C560^{AlphaFold3} vs C560^{QTY}, b) CYB^{AlphaFold3} vs CYB^{QTY}, c) DHSD^{AlphaFold3} vs DHSD^{QTY}, d) pMOA^{AlphaFold3} vs pMOA^{QTY}, e) pMOB^{AlphaFold3} vs pMOB^{QTY}, and f) pMOC^{AlphaFold3} vs pMOC^{QTY}.

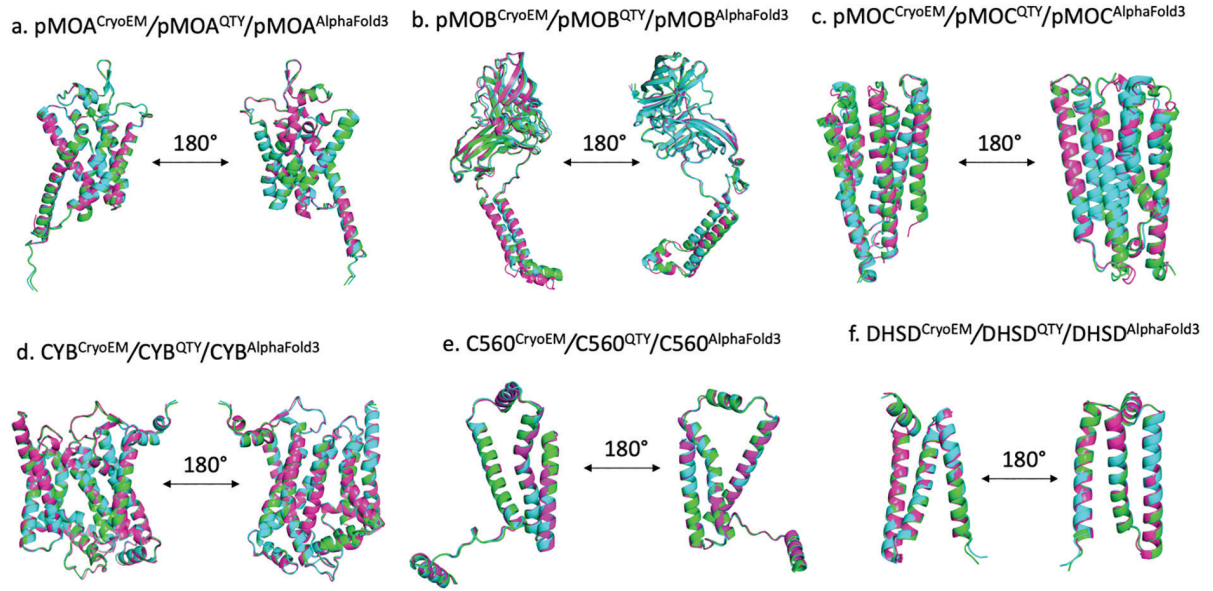


Figure 5. Superpositions of CryoEM structures with AlphaFold3-predicted native enzymes and their water-soluble QTY variants. Superposition of i) the experimentally determined CryoEM structures (magenta) with ii) AlphaFold3-predicted native transporters (green) and iii) AlphaFold3-predicted water-soluble QTY variant transporters (cyan). a) C560^{CryoEM} vs C560^{AlphaFold3} vs C560^{QTY}, b) CYB^{CryoEM} vs CYB^{AlphaFold3} vs CYB^{QTY}, c) DHSD^{CryoEM} vs DHSD^{AlphaFold3} vs DHSD^{QTY}, d) pMOA^{CryoEM} vs pMOA^{AlphaFold3} vs pMOA^{QTY}, e) pMOB^{CryoEM} vs pMOB^{AlphaFold3} vs pMOB^{QTY}, and f) pMOC^{CryoEM} vs pMOC^{AlphaFold3} vs pMOC^{QTY}.

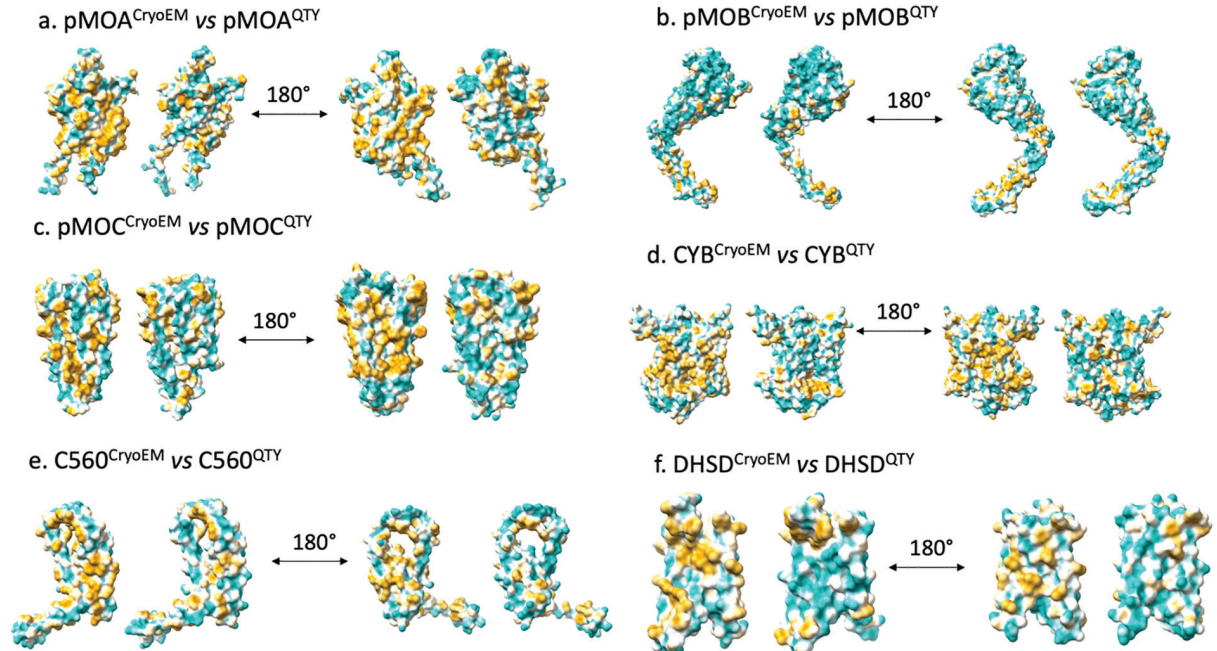


Figure 6. Hydrophobic surface of six native proteins and their water-soluble QTY variants. The native oxidation enzymes have many hydrophobic residues L, I, V, and F in the transmembrane helices. After Q, T, and Y substitutions of L, I and V, and F respectively, the hydrophobic surface patches (yellow) in the transmembrane helices become more hydrophilic (cyan). a) pMOA vs pMOA^{QTY}, b) pMOB vs pMOB^{QTY}, c) pMOC vs pMOC^{QTY}, d) CYB vs CYB^{QTY}, e) C560 vs C560^{QTY}, f) DHSD vs DHSD^{QTY}.

AlphaFold3 predictions

Structural biology researchers have sought to predict posttranslational protein folding accurately and efficiently for almost 70 years. With the advent of machine learning tool AlphaFold2 in 2021 (36, 37) and AlphaFold3 in 2024 (38), we are now able to study the structure of previously unattainable proteins, especially those with integral embedded transmembrane regions. Rather than undergo the difficult process of gene expression, protein production, detergent selection, purification, detergent exchange, maintaining stability and avoiding agglutination, scientists are able to generate fairly accurate predictions within hours, or even minutes depending on protein size.

In collaboration with the European Bioinformatics Institute (EBI), DeepMind has made over 214 million predicted protein structures freely available online (<https://alphafold.ebi.ac.uk>) within the three years since AlphaFold2's release. This speed and accuracy is unprecedented in the field of structural biology (36, 37, 38).

Rationale for selecting enzymes in this study

Particulate methane monooxygenase (pMMO) is one of the three enzymatic complexes known to carry out methane hydroxylation (8-15). The enzyme complex is very important for removing methane from the air and reducing the climate change damage. Its mechanism is not fully understood but thought to be related to the structurally similar cytochrome c oxidase II complex which also contains functionally important copper ions. The cytochrome c oxidase II complex found in the mitochondrial membrane is also involved in human diseases including mitochondrial encephalomyopathy, hypertrophic cardiomyopathy (HCM), sporadic mitochondrial myopathy (MM), pheochromocytoma syndrome, gastric stromal sarcoma, and nuclear type 3 mitochondrial complex II deficiency. We thus are interested to carry out the structural bioinformatic study of the six proteins in pMMO subunits and representative parts of the repository complex with known CryoEM structures to further validate the QTY code.

The water-soluble QTY variants may be useful in a variety of scenarios including i) using the purified soluble proteins to carry out ligand-binding drug discoveries, ii) determining the currently unknown detailed mechanisms and functions of methane monooxygenase and similar copper-center-containing proteins, and iii) potentially achieving industrial-scale use of methane as a biofuel by expressing -soluble pMMOs in filamentous algae which allows direct secretion of the protein into surrounding culture media without the need of cell break-down.

We previously showed that the QTY code can also be reversible because of their simple pairwise changes, $L \rightleftharpoons Q$, $I/V \rightleftharpoons T$, and $F \rightleftharpoons Y$. This is similar as the DNA double-helix, $A \rightleftharpoons T$, $G \rightleftharpoons C$ and vice versa $T \rightleftharpoons A$, $C \rightleftharpoons G$. These QTY or reverse QTY changes can be results of a single base pair mutation (Figure S1).

Karagöl et al. reported in several glutamate transporters and (29), serotonin, dopamine and norepinephrine transporters (30) that underwent such $L \rightleftharpoons Q$, $I \rightleftharpoons T$, and $F \rightleftharpoons Y$ mutations. Some of them result diseases and others are silent, all depending on the specific location of mutations.

Membrane transporter proteins are conserved from bacteria to mammals (39). This remarkable conservation makes them ideal subjects for evolutionary studies. For instance, evolutionary profiling potentially allows a better understanding of systematic amino acid substitutions, including the QTY code (29, 30). The phenomena perhaps also took place in other transmembrane helical proteins including pMOA, pMOB, pMOC, CYB, C560, and DHSD. Applied the same idea to the transmembrane helices of pMMO complex and others, QTY code-based studies are likely to provide insights into biological foundations of chemically distinct alpha-helices. Understanding the alpha-helical evolution has a potential to provide crucial information on protein molecular evolution and sequence-based structural predictions. Moreover, by combining this information with variant based approaches, we analyzed to what extent Darwinian natural selection and mutationally favored biases affected the diversification of QTY-code pairs (L/Q, I/T, V/T, and F/Y).

The chemical properties of the helical structures are governed by the genetic code, as genetic code's second position is determinant of the chemical nature of amino acids (1, 2, 3, 4, 19). For instance, i) amino acids with U at the second position are hydrophobic (Phe, Leu, Ile, Val, and Met); ii) amino acids with C at the second position are less hydrophobic (Pro and Ala), or with a hydroxyl -OH group (Ser and Thr); iii) amino acids with A at the second position are hydrophilic and water-soluble (Asp, Glu, Asn, Glu, Lys, His and Tyr), and 2 stop codons Ochre (UAA) and Amber (UAG); iv) amino acids (Arg and Ser) with G at the second position are water -soluble, Cys is partially water-soluble and Gly is achiral and has a hydrogen as the side chain (1, 2, 3, 4, 19). The stop codon is UGA. In general, pyrimidine U and C at the second position confer hydrophobicity; in contrast, purine A and G at the second position confer hydrophilicity (Figure S1).

The variations all occur in the second position of the genetic code, including transition mutation, i.e., purine to purine

(A->G, G->A) and pyrimidine to pyrimidine (C->U, or U->C); or transversion mutation (U->A, U->G, C->A, C->G, A->U, A->C, G->U, G->C). In the case of L->Q, I->T, and F->Y substitutions. **i**) In L (leucine), two codons are CUA and CUG, and in Q (glutamine), two codons are CAA and CAG; in these cases, the second position of U is mutated to A, which is a transversion mutation (U->A). **ii**) In I (isoleucine), three codons are AUU, AUC, and AUA, in T (threonine), four codons are ACU, ACC, ACA, and ACG; which is a transition mutation (U->C). **iii**) In F (phenylalanine), two codons are UUU and UUC, in Y (tyrosine), two codons are UAU and UAC, and the second position of U is mutated to A which is a transversion mutation (U->A). The same logic applies in the case of the reverse QTY mutations of Q->L, T->I, Y->F. Transitions are expected to be more frequent than transversions, with a ratio much higher in coding regions. F<=>Y and L<=>Q changes that required same mutations (U->A and A->U) observed in similar ratios.

Conclusion

Nature has evolved a set of 20 amino acids over billions of years with diverse chemical properties and some of which share remarkable structural similarities. These similarities are especially striking for L vs D, N, E, Q, V/I vs T and F vs Y. These amino acids are shared for all living systems on earth (perhaps elsewhere of living system on planets in other solar system in the universe). Understanding these fine structural similarities allows us to pairwise inter-change them and still keep the protein's structural integrity and retain the biological function.

Applying the simple QTY code, we bioengineered six water-soluble QTY variants of the pMMO complex and cytochrome c oxidase II proteins by systematically converting hydrophobic alpha-helices to hydrophilic ones. We then employed *in silico* bioinformatics tools to analyze the characteristics of native proteins and their QTY variants. After superposing in PyMOL, the structures of QTY variants showed a high overall similarity to both experimentally determined Cryo-EM structures and AlphaFold3-predicted structures of the native proteins, suggesting a high likelihood that their functions would be retained. To further verify the reliability, we calculated various characteristics of the two groups including root mean square deviation, isoelectric focusing points, molecular weight, variations and hydrophobicity. Despite significant changes to the amino acids in the integral transmembrane domain and reduction in hydrophobic patches, bioinformatic results demonstrate that the structures of native proteins and their QTY variants remain highly similar. This suggests the viability of the QTY code as an approach to modeling water-soluble

variants of various membrane proteins. We believe the bioengineered soluble QTY proteins may have real-world application potentials not only for structural and mechanistic studies of membrane enzymes, but also for removing methane from the air for combating climate change, and for developing a more sustainable alternative energy source.

Methods

Protein sequence alignments and other characteristics

The native protein sequences for transporters including pMOA, pMOB, and pMOC, CYB, C560, and DHSD, were obtained from UniProt (<https://www.uniprot.org>). The sequences for the QTY variants were aligned using the same methods as previously described. The MWs and pI values of the proteins were calculated using the ExPASy (https://web.expasy.org/compute_pi/).

AlphaFold3 predictions

The protein structures of the QTY variants were predicted using AlphaFold3 (<https://github.com/sokrypton/ColabFold>) and by following the instructions at the website. PDB files for the predicted native protein structures were obtained from The EBI (<https://alphafold.ebi.ac.uk>), which contains all AlphaFold3-predicted structures for native proteins. The UniProt website (<https://www.uniprot.org>) provided protein ID, entry name, description, and FASTA sequence for each native protein. The QTY code can be applied to FASTA sequences by manually replacing amino acids in the TM domains (found on the Protter 2D diagram-plotting website, <http://wlab.ethz.ch/protter/start/>) but can also be done through the QTY method website (<https://pss.sjtu.edu.cn/>). The website also provides MWs, pI values, TM variation, and overall variation.

Superposed structures

PDB files for native protein structures experimentally determined by CryoEM were taken from the PDB include pMOA, pMOB, and pMOC, CYB, C560, DHSD. Predictions for the QTY variants were carried out using the AlphaFold3 program, which can be found at <https://github.com/sokrypton/ColabFold>. These structures were superposed and the RMSDs were calculated using PyMOL (<https://pymol.org/3/>). For detailed method information, please see references 25-30.

Structure visualization

PyMOL (<https://pymol.org/3/>) was used to superpose the native protein structure and the QTY variant. UCSF Chimera (<https://www.rbvi.ucsf.edu/chimera>) was used to render each protein model with hydrophobicity patches. For detailed method information, please see references 25-30.

Ethics Approval

Ethics approval was not required for this computational study as it did not involve animal subjects, human participants, and identifiable data.

Consent for publication

Not applicable. This computational study did not involve human participants.

Competing financial interests

None.

Funding

The authors received no specific funding for this work.

Author Contributions

Conceptualization, S.Z.; methodology, J-Y.W., E.P.; formal analysis, J-Y.W.; investigation, J-Y.W., resources, S.Z.; data curation, J-Y.W., E.P.; writing—review and editing, J-Y.W., E.P. and S.Z.; visualization, J-Y.W., E.P., supervision, S.Z.; project administration, S.Z.; funding acquisition, S.Z. All authors have read and agreed to the published version of the manuscript.

References

1. Stryer L. Biochemistry. 2nd ed. San Francisco: WH Freeman and Company; 1981.
2. Brändén CI, Tooze J. Introduction to protein structure. 1st ed. Garland Science; 1991. p. 15.
3. Fersht A. Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding. San Francisco: WH Freeman and Company; 1998.
4. Liljas A, Liljas L, Ash M-R, Lindblom G, Nissen P, Kjeldgaard M. Textbook of Structural Biology. 2nd ed. Singapore: World Scientific Publishing Company; 2017.
5. Linke D. Detergents: an overview. *Methods Enzymol.* 2009;463:603-617.
6. Lin SH, Guidotti G. Purification of membrane proteins. *Methods Enzymol.* 2009;463:619-629.
7. Duquesne K, Sturgis JN. Membrane protein solubilization. *Methods Mol Biol.* 2010;601:205-217.
8. Lieberman RL, Rosenzweig AC. Crystal structure of a membrane-bound metalloenzyme that catalyses the biological oxidation of methane. *Nature.* 2005;434(7030):177-182.
9. Chang WH, et al. Copper centers in the Cryo-EM structure of particulate methane monooxygenase reveal the catalytic machinery of methane oxidation. *J Am Chem Soc.* 2021;143(26):9922-9932.
10. Zhu Y, et al. Structure and activity of particulate methane monooxygenase arrays in methanotrophs. *Nat Commun.* 2022;13(1):5221.
11. Wang W, Liang AD, Lippard SJ. Coupling oxygen consumption with hydrocarbon oxidation in bacterial multicomponent monooxygenases. *Acc Chem Res.* 2015;48(9):2632-2639.
12. DiSpirito AA, Gullledge J, Shiemke AK, et al. Trichloroethylene oxidation by the membrane-associated methane monooxygenase in type I, type II and type X methanotrophs. *Biodegradation.* 1991;2:151-164.
13. Choi DW, et al. The membrane-associated methane monooxygenase (pMMO) and pMMO-NADH:quinone oxidoreductase complex from *Methylococcus capsulatus* Bath. *J Bacteriol.* 2003;185(19):5755-5764.
14. Balasubramanian R, et al. Oxidation of methane by a biological dicopper centre. *Nature.* 2010;465(7294):115-119.
15. Sirajuddin S, et al. Effects of zinc on particulate methane monooxygenase activity and structure. *J Biol Chem.* 2014;289(31):21782-21794.
16. Martinho M, et al. Mössbauer studies of the membrane-associated methane monooxygenase from *Methylococcus capsulatus* Bath: Evidence for a diiron center. *J Am Chem Soc.* 2007;129:15783-15785.
17. Yu SS-F, et al. Production of high-quality particulate methane monooxygenase in high yields from *Methylococcus capsulatus* (Bath) with a hollow-fiber membrane bioreactor. *J Bacteriol.* 2003;185:5915-5924.
18. Vinothkumar KR, Henderson R. Structures of membrane proteins. *Q Rev Biophys.* 2010;43:65-158.
19. Zhang S, Egli M. Hiding in plain sight: three chemically distinct α -helix types. *Q Rev Biophys.* 2022;55:e7.
20. Zhang S, et al. QTY code enables design of detergent-free chemokine receptors that retain ligand-binding activities. *Proc Natl Acad Sci USA.* 2018;115(37):E8652-E8659. PMID: 30154163.
21. Qing R, et al. QTY code designed thermostable and water-soluble chimeric chemokine receptors with tunable ligand-binding activities. *Proc Natl Acad Sci USA.* 2019;116(51):25668-25676. PMID: 31776256.
22. Hao SL, et al. QTY code-designed water-soluble Fc-fusion cytokine receptors bind to their respective ligands. *QRB Discovery.* 2020;1:e4.
23. Tegler LT, et al. G protein-coupled receptor CXCR4 designed by the QTY code becomes more hydrophilic and retains cell-signaling activity. *Sci Rep.* 2020;10:21371.
24. Qing R, et al. Scalable biomimetic sensing system with membrane receptor dual-monolayer probe and graphene transistor arrays. *Sci Adv.* 2023;9(29):eadf1402. PMID: 37478177.
25. Skuhersky M, et al. Comparing native crystal structures and AlphaFold2 predicted water-soluble G protein-coupled receptor QTY variants. *Life.* 2021;11(12):10.3390/life11121285.
26. Smorodina E, et al. Comparing 2 crystal structures and 12 AlphaFold2 predicted human membrane glucose transporters and their water-soluble QTY variants. *QRB Discovery.* 2022;3:e5.
27. Smorodina E, et al. Structural informatic study of determined and AlphaFold2 predicted molecular structures of 13 human sol-

- ute carrier transporters and their water-soluble QTY variants. *Sci Rep.* 2022;12:20103.
28. Pan E, Smorodina E, Zhang S. Structural bioinformatics studies of six human ABC transporters and their AlphaFold2 predicted water-soluble QTY variants. *QRB Discovery.* 2024;4:e2. DOI: 10.1017/qrd.2024.2.
 29. Karagöl A, Karagöl T, Smorodina E, Zhang S. Structural bioinformatics studies of glutamate transporters and their AlphaFold2 predicted water-soluble QTY variants and uncovering the natural mutations of L->Q, V->T, F->Y and Q->L, T->V, Y->F. *PLoS ONE.* 2024;19(4):e0289644.
 30. Karagöl A, Karagöl T, Zhang S. Structural bioinformatics studies of serotonin, dopamine and norepinephrine transporters and their AlphaFold2 predicted water-soluble QTY variants and uncovering the natural mutations of L->Q, I->T, F->Y and Q->L, T->I and Y->F. *PLoS ONE.* 2024;19(3):e0300340. PMID: 38517879.
 31. Li M, et al. Design of a water-soluble transmembrane receptor kinase with intact molecular function by QTY code. *Nat Commun.* 2024;15(1):4293. PMID: 38858360.
 32. Finkelstein J. Methanol maker. *Nature.* 2005;434:151. DOI: 10.1038/434151b.
 33. Yano N, et al. The Mg²⁺-containing water cluster of mammalian cytochrome c oxidase collects four pumping proton equivalents in each catalytic cycle. *J Biol Chem.* 2016;291:23882-23894.
 34. Zong S, et al. Structure of the intact 14-subunit human cytochrome c oxidase. *Cell Res.* 2018;28:1026-1034.
 35. Gao X, et al. Structural basis for the quinone reduction in the bc1 complex: a comparative analysis of crystal structures of mitochondrial cytochrome bc1 with bound substrate and inhibitors at the Qi site. *Biochemistry.* 2003;42(30):9067-9080.
 36. Jumper J, Evans R, Pritzel A, Green T, Figurnov M, et al. Highly accurate protein structure prediction with AlphaFold. *Nature.* 2021;596(7873):583-589.
 37. Jumper J, Hassabis D. Protein structure predictions to atomic accuracy with AlphaFold. *Nat Methods.* 2022;19(1):11-12.
 38. Abramson J, Adler J, Dunger J, Evans R, Green T, et al. Accurate structure prediction of biomolecular interactions with AlphaFold 3. *Nature.* 2024;630(8016):493-500.
 39. Yan N. Structural advances for the major facilitator superfamily (MFS) transporters. *Trends Biochem Sci.* 2013;38:151-159. DOI: 10.1016/j.tibs.2013.01.003.