Non-Canonical Cytochrome P450 Enzymes in Nature

Andy K. L. Nguy[1,6], Kendra A. Ireland[2,6], Chase M. Kayrouz[1,6], Juan Carlos Cáceres[3], Jonathan Z. Huang[1], Vanessa Y. Ying[1], Joanna A. Quaye[1], Brandon L. Greene[3,4], Katherine M. Davis[2], and Mohammad R. Seyedsayamdost[1,5,*]

[1]Department of Chemistry, Princeton University, Princeton, NJ, USA
[2]Department of Chemistry, Emory University, Atlanta, GA, USA
[3]Interdepartmental Program in Quantitative Biosciences, University of California Santa Barbara, Santa Barbara, CA, USA
[4]Department of Chemistry and Biochemistry, University of California Santa Barbara, Santa Barbara, CA, USA
[5]Department of Molecular Biology, Princeton University, Princeton, NJ, USA
[6]These authors contributed equally
[*]Corresponding author. Email: mrseyed@princeton.edu

Abstract

Cytochrome P450s (CYPs) are a superfamily of thiolate-ligated heme metalloenzymes principally responsible for the hydroxylation of unactivated C–H bonds. The lower-axial cysteine is an obligatory and universally conserved residue for the CYP enzyme class. Herein, we challenge this paradigm by systematically identifying non-canonical CYPs (ncCYPs) that do not harbor a proximal cysteine ligand. Our bioinformatic search reveals 20 distinct ncCYP families encoded in diverse microbial genomes with alternative residues at this position. We characterize a native serine-ligated CYP with a high-spin ferric resting state that catalyzes azide reduction and nitrene insertion reactions. Its crystal structure clearly shows a typical CYP fold and a serine alkoxide as a proximal heme ligand. In addition, we report the discovery and characterization of the first native selenocysteine-ligated CYP in nature. Our findings expand the CYP metalloenzyme family and provide opportunities for future enzymatic and biocatalytic discoveries.

Introduction

Hemoproteins constitute a broad class of metalloenzymes that perform a diverse range of biological functions, including oxygen transport, electron transfer, and various essential enzymatic reactions in the cell[1-5]. The identity of the lower axial ligand to the heme, known as the proximal ligand, dictates the function and reactivity of hemoproteins by modulating the electronic properties of the cofactor[6]. Several heme ligands are found in nature, including histidine in myoglobin (for oxygen binding)[7,8], tyrosine in catalase (for peroxide disproportionation)[9,10], cysteine in aromatic peroxygenase[REF], and methionine in cytochrome c (involved in the electron transport chain)[1,12]. Among hemoproteins, cytochrome P450 enzymes (CYPs) form a large and distinct superfamily, members of which carry a specific protein architecture as well as a universally conserved proximal cysteine (Cys) ligand[13]. They are named for the sharp absorption peak at 450 nm of the Fe^{II}-CO complex, when the resting ferric-heme enzyme is reduced and reacted with CO. The proximal Cys ligand governs P450 reactivity by providing the thermodynamic driving force, via the so-called 'push-and-pull' effect, to facilitate formation of the highly oxidizing compound-I intermediate, which has been detected by time-resolved spectroscopic studies[23,24]. CYPs catalyze myriad challenging transformations, notably the remarkable stereospecific hydroxylation of chemically inert C–H bonds via oxygen rebound chemistry, as well as epoxidation, desaturation, and nitration reactions (**Fig. 1a, 1b**)[14-16]. CYPs are ubiquitous in all domains of life and play essential roles in cholesterol biosynthesis, antibiotic biogenesis, and drug metabolism[17-22].

Because of their catalytic versatility, CYP enzymes have been adapted for a variety of industrial applications. Recently, they have been engineered to incorporate various non-natural proximal ligands, such as serine (Ser), which have been shown to catalyze various new modifications that expand the reaction repertoire of the protein superfamily. Interestingly, however, natural enzymes that incorporate an alternative proximal ligand within the CYP superfamily have not been reported to date. Given the intimate relationship between the proximal heme ligand and the properties of the cofactor, we wondered whether nature had selected for alternative residues at this position. Herein, we report an expansive bioinformatic and biochemical study that unearthed >800 CYP superfamily members with an alternative proximal ligand. We validated several of these, notably a Ser-ligated enzyme for which we determined a crystal structure as well as new reactions, including nitrene insertion and azide reduction. We also report naturally-occurring selenocysteine (Sec)-ligated CYPs, one of which we characterized experimentally, as well as hundreds of variants with nearly every proteinogenic amino acid substituted at the position of the canonical Cys ligand. These findings expand the CYP superfamily and provide rich avenues for future research into the properties and reactions of alternative CYP cofactors.

Results and Discussion

Bioinformatic Search for Non-canonical CYPs. To assess whether nature had selected for alternative proximal ligands to modulate CYP function, we first employed a simple bioinformatic search strategy in which a small set of diverse CYP representatives was used as a query for BLAST alignment against the entire NCBI RefSeq sequence database (**Fig. 1c**, **Supplementary Table 1**). The CYP sequences were randomly selected from the NCBI database, with the criterion that no two sequences share greater than 60% sequence identity. This query resulted in over 200,000

BLAST hits, each of which represents a distinct CYP homolog from a sequenced organism. For each homolog, the residue at the putative axial position was inferred from its top scoring BLAST alignment. This approach relies on the assumption that the alignment position, but perhaps not the identity, of the axial ligand is invariant among all homologs. It almost certainly holds for the vast majority of CYPs, and we indeed find that >99.5% of all BLAST hits feature an expected Cys residue at the aligned proximal ligand site. Of the remaining sequences, 301 were removed due to their strong similarity to the PDGAF (P450-derived glycosyltransferase activating factor) family. These close homologs of CYPs carry the canonical protein fold but lack a heme cofactor and only serve as chaperones (**Fig. 1d**, **1e**)[25,26]. Strikingly, upon removal of PDGAFs, we arrived at a list of 880 novel CYP homologs with an alternative residue at the aligned proximal ligand site (**Supplementary Table 2**). Genes encoding these unusual proteins are harbored by taxonomically diverse bacteria, spanning six phyla and over forty genera. The non-canonical CYPs (ncCYPs) were subsequently categorized into 20 subfamilies (designated as A through T) using sequence similarity analysis, revealing a rich assortment of putative axial ligands never before observed (**Fig. 1f, Supplementary Fig. 1**). While these enzymes are not CYPs *per se*, we refer to them as ncCYPs because they belong to the CYP superfamily based on the conserved amino acid sequence, save for the identity of the proximal ligand.

The two largest subfamilies (A and B) feature putative glutamate (Glu) and arginine (Arg) ligands, respectively, in place of Cys (**Fig. 1f**). Others exhibit serine (Ser, subfamilies G, R, and T) or selenocysteine (Sec, subfamily M) residues at the proximal ligand site, representing intriguing isosteric replacements of the canonical thiolate ligand. Residues with hydrophobic sidechains are found in several subfamilies as well, including alanine (subfamily C), phenylalanine (subfamily D), leucine (subfamilies F and K), and isoleucine/valine (subfamily Q). In some cases, no ligand is predicted (e.g. subfamilies E, H, L, O, and S). The fact that some residues appear in multiple clades suggests convergent evolution from disparate ancestors. Further analysis of the genetic context surrounding the ncCYPs revealed that several of the subfamilies are situated in operons containing other biosynthetic genes, suggesting their potential involvement in natural product biosynthetic pathways (**Fig. 1g**). A detailed analysis of each subfamily and its genomic neighborhood is provided in the SI (**Supplementary Fig. 2–21**).

Validation of a Naturally-occurring Ser-ligated CYPs. The proximal ligand has been at the forefront of CYP research, with detailed spectroscopic and kinetic studies providing mechanistic underpinnings for the remarkable reactivity of the cofactor and the role that the key Cys residue plays. We therefore initially focused our validation efforts on Cys isosteres, namely the Ser- and Sec-ligated CYPs, which directly modulate the electronic properties of the cofactor. Ser-ligated CYP variants have been generated by site-directed mutagenesis from Cys-containing parental wild-type sequences,[27-29] but they have never been characterized in native form. We recombinantly expressed and purified the ncCYP from family G (termed P450-G9) encoded by *Actinokineospora terrae*. In close agreement with literature precedent on engineered Ser-ligated P450s by the Arnold group[27-29], P450-G9 exhibits a strong Soret peak at 403 nm with a broad Q band at 605 nm and an additional feature at 485 nm (**Fig. 2a**). The dithionite-reduced ferrous P450-G9 shows a strong Soret peak at 423 nm. Upon formation of the ferrous-CO complex, this feature shifts to an intense band at 410 nm with well-defined Q bands at 534 nm and 567, in agreement with the engineered 'P411' enzyme[27]. Moreover, analysis by low-temperature (10 K)

X-band electron paramagnetic resonance (EPR) spectroscopy revealed two high-spin ferric heme signals (S = 5/2): a major component with simulated effective g-values of (7.1, 4.8 and 1.9), and a minor component with effective g-values of (5.95, 2.05, 1.99) (**Fig. 2b**). These features are distinct from substrate-displaced high-spin CYP systems that have characteristic g-values near (8, 4, 1.8)[30-32]. Based on the S = 5/2 rhombogram, P450-G9 displays a rhombicity parameter (E/D) of approximately 0.06, as inferred from the observed g-values, compared to a higher E/D value of 0.087 for $P450_{cam}$. This suggests that the proximal serine ligand in P450-G9 leads to reduced rhombic anisotropy relative to the cysteine-ligated $P450_{cam}$ counterpart, consistent with the weaker π-donating ability of the oxygen donor compared to the softer and more polarizable sulfur atom. Within this framework, the major EPR signal arises from the lowest $|\pm 1/2\rangle$ Kramers doublet, while the minor signal originates from the intermediate $|\pm 3/2\rangle$ doublet. The spectroscopic data are thus consistent with the presence of a Ser-ligated heme.

We subsequently solved the structure of P450-G9 via X-ray crystallography to further corroborate this conclusion (PDB accession code 9CJG; **Supplementary Table 3**). The resultant model, refined to 1.85 Å resolution, displays the hallmark CYP-fold (**Fig. 2c**, **Supplementary Table 4**), with clear electron density confirming the presence of a non-canonical alkoxide Ser ligand (residue S341) that forms a dative bond with the ferric heme center (**Fig. 2d**). Further inspection of the secondary coordination sphere reveals an unusual tryptophan residue (W229) lying near-perpendicular to the heme plane (**Fig. 2e**). Curiously, the indole side chain is positioned 4.7 Å from the heme iron and appears to block substrate binding. It adopts an unusual rotamer in the structure and appears to function as an active site lid, sealing the heme-binding pocket from the external environment (**Fig. 2f**). We modeled in the most common Trp rotamer observed in protein structures (the so-called m95 rotamer) at W229, which results in an open active site, thus facilitating the creation of a narrow solvent channel to the heme cofactor (**Fig. 2g**). Volume calculations indicate that this putative side chain rotation results in a nearly four-fold increase in the active site volume (from 523 $Å^3$ to 1965 $Å^3$). Analysis of P450-G9's binding pocket thus indicates that conformational changes may be imperative in enabling substrate ingress to the active site.

While uncommon, P450-G9 is not the only CYP that hosts a tryptophan residue distal to the heme cofactor. Prostacyclin synthase, a class III CYP that catalyzes the isomerization of prostaglandin H_2, features a tryptophan at the same location in its I helix[33,34]. Unlike P450-G9, however, prostacyclin synthase's analogous tryptophan does not block substrate binding, lying parallel with the heme plane and positioned 8.3–8.5 Å from the heme iron in substrate-free structures (PDB accession codes 2IAG and 3B98; **Supplementary Fig. 22**). Furthermore, the active site remains solvent accessible, and minimal changes to the protein structure are required for substrate binding. More analogous to P450-G9, CYP8B1, an important enzyme involved in the conversion of cholesterol to bile acids, also harbors a Trp residue (W281) above the heme cofactor. In this case, the W281 is thought to be important for substrate positioning in the active[35]. Collectively, these findings confirm identification of the first natural Ser-ligated CYP and implicate W229 and other structural features (see **Supplementary Figs. 23-24**) in regulating entry to the P450-G9 heme-binding pocket.

With the Ser-ligated cofactor verified, we next sought to examine the catalytic capability of P450-G9. Not surprisingly, the enzyme was an effective peroxidase, reducing H_2O_2 to water with

an apparent 2nd order rate constant of 2.6 x 10^5 $M^{-1}s^{-1}$ (**Supplementary Fig. 25**). Reactions for engineered Ser-ligated CYPs have focused on carbene and nitrene insertions (**Fig. 3a**). We conducted the carbene-based reaction that has been demonstrated for the engineered P411 enzymes upon reaction of a diazoester with styrene and found the corresponding cyclopropanation adduct, although in low yields (**Fig. 3b**, **Supplementary Fig. 26**). The active site of P450-G9 shows some homology to enzymes that bind vitamin D-like substrates, notably to vitamin D hydroxylase. We therefore screened nine steroidal substrates, but again found no product (see SI Methods). A nitrene source, in the form of tosyl azide, when co-incubated with these nine substrates did not lead to product formation either, though we did detect the reduced tosyl amine, indicating the enzyme can catalyze azide reduction, presumably via a nitrene intermediate (**Supplementary Fig. 27**). Given the lack of success with intermolecular reactions, we screened an intramolecular nitrene insertion reaction and found efficient sulfonamidation of triisopropyltosyl azide, a new product for the P411 compounds, which was fully characterized by NMR analysis (**Fig. 3b, Supplementary Figs. 28-30**).

Azide and diazo compounds are rather rare in nature. However, the mechanism of carbene or nitrene insertion can be extended to natural substrates. For example, one can envision carbene formation with a natural carbanion source, such β-ketoesters, followed by capture via electrophilic addition through an olefin such as succinate (**Fig. 3c**). We therefore screened a variety of potential carbene and olefins, but again did not observe product formation (see SI Methods). While the native reaction for P450-G9 remains to be determined, our studies indicate the enzyme can carry out peroxide and azide reduction as well as nitrene insertion on route to a cyclic sulfonamide (**Fig. 3b**). These results set the stage for further examination of naturally occurring Ser-ligated CYPs in the future.

Validation of a Naturally-occurring Sec-ligated CYPs. As a second case study, we further analyzed ncCYP subfamily M, putative selenoproteins featuring a Sec residue at the CYP axial position. Sec, often termed the 21st amino acid[36], is encoded by the opal stop codon (UGA) and ribosomally installed into proteins with the assistance of a specific elongation factor (SelB)[37], which concurrently binds a structural mRNA element (selenocysteine insertion sequence, SECIS) downstream of the opal codon (**Fig. 4a**)[38]. There have only been three examples of artificially Sec-substituted CYPs reported by Ortiz de Montellano[39], Hilvert[40], and Green[41], either through cysteine auxotrophy or an engineered SECIS-based strategy. To date, however, there are no known examples of a natural Sec-ligated CYP.

Intriguingly, ncCYP family M, specific to a small clade of *Entotheonella* (marine sponge endosymbionts)[42,43], appears to consist of Sec-ligated CYPs, evidenced by the presence of a conserved opal codon at the proximal site as well as the full set of Sec-incorporation genes elsewhere on the chromosomes. Further *in silico* analysis through RNAfold[44], an RNA secondary structure prediction tool, uncovered downstream stem-loop regions consistent with SECIS mRNA regulatory elements (**Fig. 4b**, **4c**, **Supplementary Fig. 31**). Moreover, structural prediction by AlphaFold2[45] reveals an overall topology consistent with canonical CYPs, including a near perfect superimposition of Sec with the proximal cysteine ligand (**Fig. 4d**). Though conclusive validation of native opal codon readthrough is challenging due to the uncultivability of *Entotheonella*, our bioinformatic analyses strongly suggest Sec incorporation in these systems.

To gain more insights into this interesting subfamily, we selected one of the members, termed P450-M4, for biochemical characterization. Given the complex biochemical machinery required for Sec installation[46], standard recombinant expression in *E. coli* is not possible. The Söll group recently pioneered an amber suppressor-based strategy to site-specifically install Sec in recombinant proteins, thereby eliminating the necessity for SelB/SECIS-mediated extension[47]. This method relies upon an engineered allo-tRNA^{Sec} that proficiently engages with *E. coli* Ser aminoacyl-tRNA synthetase and the canonical elongation factor Tu. Co-expression of SelA and SelD converts serine-acylated allo-tRNA^{Sec} to Sec[48]. To our delight, we observed robust selenoprotein expression when P450-M4 with a requisite amber stop codon was simultaneously co-expressed with a plasmid encoding SelA, SelD and allo-tRNA (**Supplementary Fig. 32**). Selenium quantification by ICP-OES elemental analysis revealed a yield of 0.63 ± 0.07 Se/mol protein and HPLC/HR-MS analysis confirmed the expected high-resolution mass of a tryptic fragment containing the Sec residue, concomitant with its distinctive selenium isotope envelope (**Fig. 4e**). Additionally, high-resolution tandem MS/MS fragmentation unequivocally demonstrated the site-specific incorporation of Sec at the expected position within P450-M4 (**Fig. 4f**, **4g**). A UV/Vis absorbance spectrum of resting state ferric P450-M4 revealed a Soret peak at 414 nm and, upon reduction and reaction with CO, a ferrous-CO band at 454 nm (**Fig. 4h**). Given the enhanced nucleophilicity of Sec vs Cys, or perhaps steric constraints, P450-M4 proved difficult to reduce even with excess strong reductants; therefore, a significant amount of ferric P450-M4 remained after reduction and reaction with CO.

The X-band EPR spectrum of P450-M4 exhibits a ferric, low-spin rhombic signature with g-values at (2.65, 2.27, 1.87) (**Fig. 4i**). These are in contrast to the values reported by Ortiz de Montellano (2.46, 2.27, 1.98)[39] and Hilvert (2.47, 2.27, 2.00)[40], perhaps due to alterations in the secondary coordination sphere of the native Sec-ligated P450 that likely affects the properties and EPR parameters of the cofactor. As further validation of Sec-ligation, we determined the Fe^{III}/Fe^{II} reduction potential of the resting cofactor using protein film voltammetry, as the greater electron donating ability of Sec, relative to Cys and Ser, should lower the reduction potential of the cofactor. As control, we used the Cys-ligated enzyme OxyB, involved in vancomycin biosynthesis, as well as the Ser-ligated P450-G9. The results indeed showed P450-M4 to be the most reducing with a peak reduction potential at -148 mV (vs. NHE) (**Fig. 4j, Supplementary Fig. 33**). OxyB and the Ser-ligated P450-G9 showed peak potentials at -123 mV and -68 mV, respectively, in line with the expected trend, based on the electron donating ability of the proximal residues and the effect of these ligands on the heme cofactor. Interestingly, the Fe^{III}/Fe^{II} potential for the natural P450-M4 (-148 mV) is significantly more negative than the engineered Sec-P450cam (2 mV)[40].

The active site of P450-M4 shows homology to orsellinic acid oxidase, and we therefore conducted reactions with orsellinic acid but did not detect product in these reactions (see SI Methods). To assess whether the enzyme can generate reactive intermediates, we monitored its reaction with the oxidant *meta*-chloroperoxybenzoic acid (*m*CPBA). Diode array spectra show a transition with a clear isosbestic point from the ferric resting state to a new species upon reaction with *m*CPBA (**Fig. 4k**). The diagnostic ~690 nm feature, indicative of compound-I, was not observed as part of this new species (**Supplementary Fig. 34**); instead we observed a new shift at 600 nm and a blue-shifted Soret (from 420 to 408 nm); the nature of this intermediate needs

to be further examined[49]. The discovery of this native Sec-ligated P450 will facilitate further downstream studies to interrogate cofactor properties, reaction scope and mechanism, as well as possible biocatalytic applications.

Other Non-canonical CYPs. Finally, we explored two enzymes from the network that are situated in obvious biosynthetic gene clusters, CYP-L2 and CYP-Q1, which are adjacent to a polyketide synthase in *Kutzneria* sp. and a spermidine synthase in *Acanthopleuribacter* sp., respectively (**Supplementary Fig. 35**). However, heterologous expression of the L2-containing cluster in *Streptomyces albus* and the Q1-containing operon in *E. coli* did not reveal any new metabolites compared to the empty vector control, when monitored by HPLC-MS (see **SI Methods**). Enzymatic assays with P450-Q1 with several alkyalamine, given that it is situated adjacent to spermidine synthase, did not reveal product formation either. We subsequently explored 20 additional ncCYPs from the network and expressed enzymes with Glu (CYP-A58, A164, and A180), Arg (CYP-B19, B64, B110, I4, J5, P1, P2), Ala (C101), Phe (D2), no residue (E11, L2), Leu (F1, K2), Gly (N1), and Ile/Val (Q1, Q2) at the proximal site as hexaHis-tagged constructs in *E. coli*. Fourteen of the 20 constructs proved insoluble (**Supplementary Fig. 35**), even when co-expressed with GroEL/GroES, which can facilitate folding and solubility. Six proteins (B19, B64, C101, F1, L2, and Q1) were soluble and were purified using metal ion affinity chromatography. All enzymes purified without a bound heme cofactor (**Supplementary Fig. 36**). The successful expression and purification of these and the arrangement of some ncCYPs in obvious biosynthetic clusters provide rich avenues for further research into the functions of these newly identified proteins.

Conclusions

Herein, we have bioinformatically identified a collection of ncCYP enzymes, which carry a non-Cys residue at the proximal site. Biochemical characterization confirmed the natural occurrence of Ser- and Sec-ligated CYPs. Others remain to be studied. Our results open a new field of inquiry in understanding the evolution, structure, and perhaps most importantly, the function of these alternative CYPs. Together, our findings challenge the dogma that CYPs must always possess a conserved Cys thiolate ligand. Electronic perturbations of the heme cofactor, via substitution of Ser, Sec, and other residues, are likely to modulate reactivity through push-pull chemistry. However, the exact function of uncharged residues, which abolish the protein's ability to bind the heme cofactor, remains unclear. Our findings further suggest that the CYP fold is a privileged protein scaffold, for which the axial ligand has been diversified throughout evolution to tune reactivity and modulate enzyme function. Future directions include elucidating the small molecules associated with the biosynthetic gene clusters featuring ncCYPs as well as unveiling the roles of the ncCYPs in these pathways. These and other studies are now enabled by the expanded chemical landscape of CYPs with exotic P450 heme ligands.

References

1. A. W. Munro, H. M. Girvan, K. J. McLean, M. R. Cheesman, D. Ley, "Heme and Hemoproteins". *Tetrapyrroles: Birth, Life and Death*; M. J. Warren, A. G. Smith, Ed.; Springer: 2009, 160–183.
2. Sono, M., Roach, M. P., Coulter, E. D. & Dawson, J. H. Heme-Containing Oxygenases. *Chem. Rev.* **96**, 2841-2888 (1996). https://doi.org/10.1021/cr9500500

3. Dawson, J. H. Probing Structure-Function Relations in Heme-Containing Oxygenases and Peroxidases. *Science* **240**, 433-439 (1988). https://doi.org/doi:10.1126/science.3358128
4. Poulos, T. L. Heme Enzyme Structure and Function. *Chem. Rev.* **114**, 3919-3962 (2014). https://doi.org/10.1021/cr400415k
5. J. Wyman Jr., *Adv. Protein Chem.* **4**, 407–531 (1948). https://doi.org/10.1016/S0065-3233(08)60011-X
6. Roderick, S. L. & Matthews, B. W. Structure of the cobalt-dependent methionine aminopeptidase from Escherichia coli: a new type of proteolytic enzyme. *Biochemistry* **32**, 3907-3912 (1993). https://doi.org/10.1021/bi00066a009
7. Kendrew, J. C. *et al.* A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature* **181**, 662-666 (1958). https://doi.org/10.1038/181662a0
8. Kendrew, J. C. *et al.* Structure of Myoglobin: A Three-Dimensional Fourier Synthesis at 2 Å. Resolution. *Nature* **185**, 422-427 (1960). https://doi.org/10.1038/185422a0
9. Murthy, M. R. N., Reid, T. J., Sicignano, A., Tanaka, N. & Rossmann, M. G. Structure of beef liver catalase. *J. Mol. Biol.* **152**, 465-499 (1981). https://doi.org/https://doi.org/10.1016/0022-2836(81)90254-0
10. Reid, T. J. *et al.* Structure and heme environment of beef liver catalase at 2.5 Å resolution. *Proc. Natl. Acad. Sci. USA* **78**, 4767-4771 (1981). https://doi.org/doi:10.1073/pnas.78.8.4767
11. Takano, T. & Dickerson, R. E. Conformation change of cytochrome c: I. Ferrocytochrome c structure refined at 1·5 Å resolution. *J. Mol. Biol.* **153**, 79-94 (1981). https://doi.org/https://doi.org/10.1016/0022-2836(81)90528-3
12. Takano, T. & Dickerson, R. E. Conformation change of cytochrome c: II. Ferricytochrome c refinement at 1·8 Å and comparison with the ferrocytochrome structure. *J. Mol. Biol.* **153**, 95-115 (1981). https://doi.org/https://doi.org/10.1016/0022-2836(81)90529-5
13. P. R. Ortiz de Montellano, "Cytochrome P450: Structure, Mechanism, and Biochemistry" Springer: 2015.
14. Guengerich, F. P. & Munro, A. W. Unusual Cytochrome P450 Enzymes and Reactions*. *J. Biol. Chem.* **288**, 17065-17073 (2013). https://doi.org/https://doi.org/10.1074/jbc.R113.462275
15. Huang, X. & Groves, J. T. Beyond ferryl-mediated hydroxylation: 40 years of the rebound mechanism and C–H activation. *J. Biol. Inorg. Chem.* **22**, 185-207 (2017). https://doi.org/10.1007/s00775-016-1414-3
16. Zhang, X. & Li, S. Expansion of chemical space for natural products by uncommon P450 reactions. *Nat. Prod. Rep.* **34**, 1061-1089 (2017). https://doi.org/10.1039/C7NP00028F
17. Lamb, D. C. *et al.* On the occurrence of cytochrome P450 in viruses. *Proc. Natl. Acad. Sci. USA* **116**, 12343-12352 (2019). https://doi.org/doi:10.1073/pnas.1901080116
18. McLean, Kirsty J., Hans, M. & Munro, Andrew W. Cholesterol, an essential molecule: diverse roles involving cytochrome P450 enzymes. *Biochem. Soc. Trans.* **40**, 587-593 (2012). https://doi.org/10.1042/bst20120077
19. Thompson, E. A. & Siiteri, P. K. The Involvement of Human Placental Microsomal Cytochrome P-450 in Aromatization. *J. Biol. Chem.* **249**, 5373-5378 (1974). https://doi.org/https://doi.org/10.1016/S0021-9258(20)79736-X
20. Simpson, E. R. *et al.* Aromatase Cytochrome P450, The Enzyme Responsible for Estrogen Biosynthesis*. *Endocr. Rev.* **15**, 342-355 (1994). https://doi.org/10.1210/edrv-15-3-342
21. Guengerich, F. P. Roles of Individual Human Cytochrome P450 Enzymes in Drug Metabolism. *Pharmacol. Rev.*, PHARMREV-AR-2024-001173 (2024). https://doi.org/10.1124/pharmrev.124.001173
22. Wrighton, S. A. & Stevens, J. C. The Human Hepatic Cytochromes P450 Involved in Drug Metabolism. *Crit. Rev. Toxicol.* **22**, 1-21 (1992). https://doi.org/10.3109/10408449209145319
23. Yosca, T. H. *et al.* Iron(IV)hydroxide pKa and the Role of Thiolate Ligation in C–H Bond Activation by Cytochrome P450. *Science* **342**, 825-829 (2013). https://doi.org/doi:10.1126/science.1244373
24. Groves, J. T. Using push to get pull. *Nat. Chem.* **6**, 89-91 (2014). https://doi.org/10.1038/nchem.1855
25. Oberthür, M. *et al.* A Systematic Investigation of the Synthetic Utility of Glycopeptide Glycosyltransferases. *J. Am. Chem. Soc.* **127**, 10747-10752 (2005). https://doi.org/10.1021/ja052945s
26. Moncrieffe, M. C. *et al.* Structure of the Glycosyltransferase EryCIII in Complex with its Activating P450 Homologue EryCII. *J. Mol. Biol.* **415**, 92-101 (2012). https://doi.org/https://doi.org/10.1016/j.jmb.2011.10.036

27. Coelho, P. S. *et al.* A serine-substituted P450 catalyzes highly efficient carbene transfer to olefins in vivo. *Nat. Chem. Biol.* **9**, 485-487 (2013). https://doi.org/10.1038/nchembio.1278
28. Yang, Y. & Arnold, F. H. Navigating the Unnatural Reaction Space: Directed Evolution of Heme Proteins for Selective Carbene and Nitrene Transfer. *Acc Chem. Res.* **54**, 1209-1225 (2021).
29. Liu, Z. & Arnold, F. H. New-to-nature chemistry from old protein machinery: carbene and nitrene transferases. *Curr. Opin. Biotechnol.* **69**, 43-51 (2021).
30. Luthra, A., Denisov, I. G. & Sligar, S. G. Spectroscopic features of cytochrome P450 reaction intermediates. *Arch. Biochem. Biophys.* **507**, 26-35 (2011). https://doi.org/https://doi.org/10.1016/j.abb.2010.12.008
31. Tsai, R. *et al.* Spin-state changes in cytochrome P-450$_{cam}$ on binding of specific substrates. *Proc. Natl. Acad. Sci. USA* **66**, 1157-1163 (1970). https://doi.org/doi:10.1073/pnas.66.4.1157
32. Lipscomb, J. D. Electron paramagnetic resonance detectable states of cytochrome P-450cam. *Biochemistry* **19**, 3590-3599 (1980). https://doi.org/10.1021/bi00556a027
33. Chiang, C.-W., Yeh, H.-C., Wang, L.-H. & Chan, N.-L. Crystal Structure of the Human Prostacyclin Synthase. *J. Mol. Biol.* **364**, 266-274 (2006). https://doi.org/https://doi.org/10.1016/j.jmb.2006.09.039
34. Li, Y.-C. *et al.* Structures of Prostacyclin Synthase and Its Complexes with Substrate Analog and Inhibitor Reveal a Ligand-specific Heme Conformation Change*. *J. Biol. Chem.* **283**, 2917-2926 (2008). https://doi.org/https://doi.org/10.1074/jbc.M707470200
35. Böck, A. *et al.* Selenocysteine: the 21st amino acid. *Mol. Microbiol.* **5**, 515-520 (1991). https://doi.org/10.1111/j.1365-2958.1991.tb00722.x
36. Liu, J., Carlson, H. A. & Scott, E. E. The structure and characterization of human cytochrome P450 8B1 supports future drug design for nonalcoholic fatty liver disease and diabetes. *J. Biol. Chem.* **209**, 102344 (2022).
37. Leibundgut, M., Frick, C., Thanbichler, M., Böck, A. & Ban, N. Selenocysteine tRNA-specific elongation factor SelB is a structural chimaera of elongation and initiation factors. *EMBO* **24**, 11-22 (2005). https://doi.org/https://doi.org/10.1038/sj.emboj.7600505
38. Walczak, R., Westhof, E., Carbon, P. & Krol, A. A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA* **2**, 367-379 (1996).
39. Jiang, Y. *et al.* Calculated and experimental spin state of seleno cytochrome P450. *Angew. Chem., Int. Ed. Engl.* **48**, 7193 (2009).
40. Aldag, C. *et al.* Probing the role of the proximal heme ligand in cytochrome P450cam by recombinant incorporation of selenocysteine. *Proc. Natl. Acad. Sci. USA* **106**, 5481-5486 (2009). https://doi.org/doi:10.1073/pnas.0810503106
41. Onderko, E. L., Silakov, A., Yosca, T. H. & Green, M. T. Characterization of a selenocysteine-ligated P450 compound I reveals direct link between electron donation and reactivity. *Nat. Chem.* **9**, 623-628 (2017). https://doi.org/10.1038/nchem.2781
42. Freeman, M. F., Vagstad, A. L. & Piel, J. Polytheonamide biosynthesis showcasing the metabolic potential of sponge-associated uncultivated 'Entotheonella' bacteria. *Curr. Opin. Chem. Biol.* **31**, 8-14 (2016). https://doi.org/https://doi.org/10.1016/j.cbpa.2015.11.002
43. Lackner, G., Peters, E. E., Helfrich, E. J. N. & Piel, J. Insights into the lifestyle of uncultured bacterial natural product factories associated with marine sponges. *Proc. Natl. Acad. Sci. USA* **114**, E347-E356 (2017). https://doi.org/doi:10.1073/pnas.1616234114
44. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA Websuite. *Nucleic Acids Res.* **36**, W70-W74 (2008). https://doi.org/10.1093/nar/gkn188
45. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021). https://doi.org/10.1038/s41586-021-03819-2
46. Stadtman, T. C. Selenocysteine. *Annu. Rev. Biochem.* **65**, 83-100 (1996). https://doi.org/10.1146/annurev.bi.65.070196.000503
47. Mukai, T., Sevostyanova, A., Suzuki, T., Fu, X. & Söll, D. A Facile Method for Producing Selenocysteine-Containing Proteins. *Angew. Chem. Int. Ed.* **57**, 7215-7219 (2018). https://doi.org/https://doi.org/10.1002/anie.201713215
48. Kayrouz, C. M. & Seyedsayamdost, M. R. Enzymatic strategies for selenium incorporation into biological molecules. *Curr. Opin. Chem. Biol.* **81**, 102495 (2024). https://doi.org/https://doi.org/10.1016/j.cbpa.2024.102495

49. Wang, X., Ullrich, R., Hofrichter, M. & Groves, J. T. Heme-thiolate ferryl of aromatic peroxygenase is basic and reactive. *Proc. Natl. Acad. Sci. USA* **112**, 3686-3691 (2015). https://doi.org/doi:10.1073/pnas.1503340112

Reporting Summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data Availability Statement

X-ray data have been deposited to the Protein Data Bank (accession code 9CJG). All other data are available in the main text or the supplementary materials.

Acknowledgments: The authors thank the National Science Foundation (GRFP No. 1937971 to K.A.I.), the Eli Lilly-Edward C. Taylor Fellowship in Chemistry (to C.M.K.), the Anid/Subdirección De Capital Humano/Beca de Doctorado Becas Chile/7220044 (to J.C.C.), and the National Institutes of Health (grants R35 GM154993 to B.L.G., R35 GM147557 to K.M.D, and R35 GM152049 to M.R.S.) for financial support.

Author contributions:

Conceptualization: AKL, CMK, MRS
Methodology: AKL, KAI, CMK, JCC
Investigation: AKL, KAI, CMK, JCC, BLG, KMD, MRS
Funding acquisition: BLG, KMD, MRS
Writing – original draft: AKL, KAI, CMK, MRS
Writing – review & editing: AKL, KAI, CMK, JCC, BLG, KMD, MRS

Competing interests: The authors declare no competing financial interests.

Additional information

Supplementary Materials

Supplementary Information is available for this paper. This report contains Supplementary text, Supplementary Tables S1 to S4, and Supplementary Figs. S1 to S36.

Correspondence and requests for materials should be addressed to Mohammad R. Seyedsayamdost

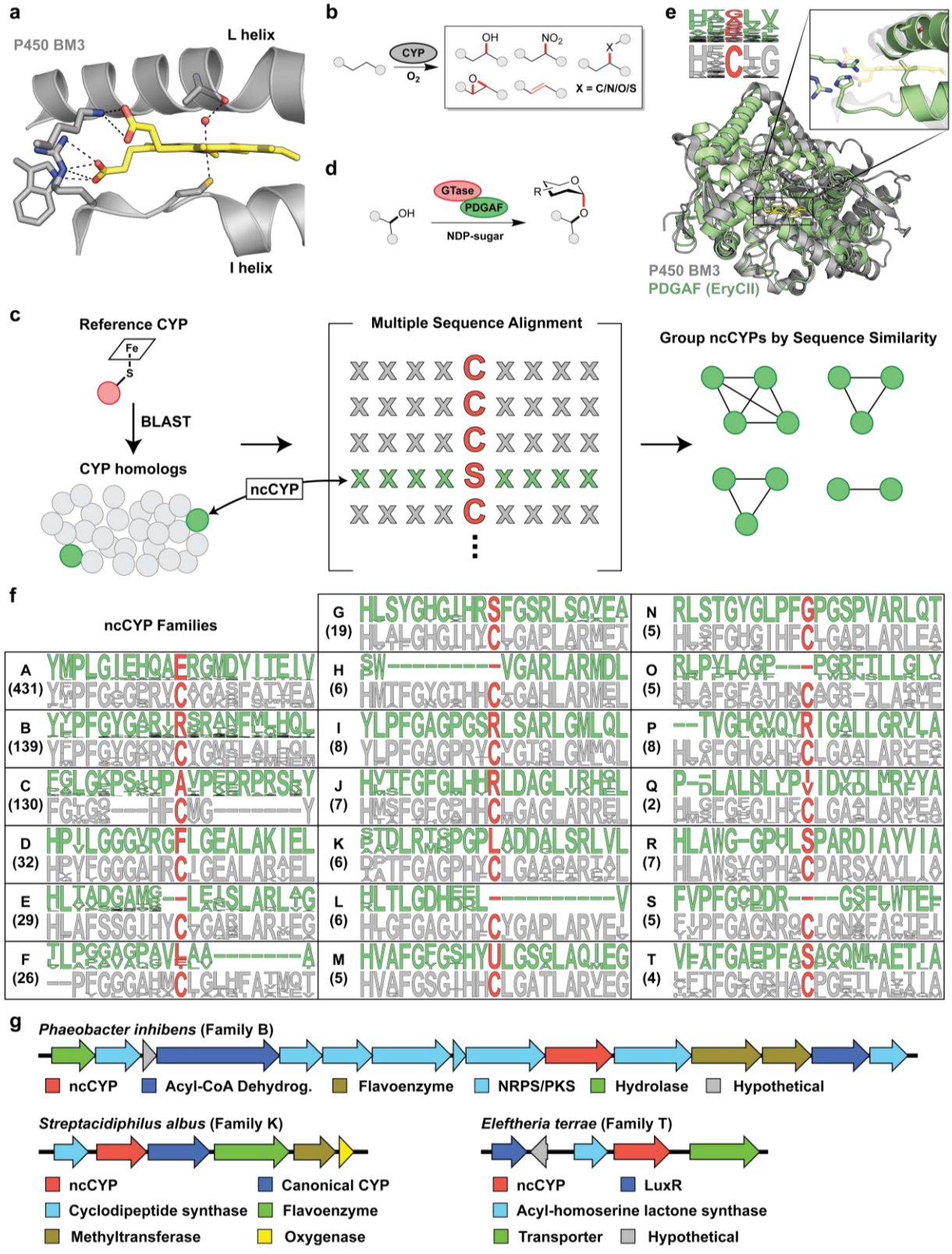

Fig. 1. Bioinformatic analysis of non-canonical cytochrome P450 enzymes. (a) Representative CYP crystal structure (P450 BM3, PDB accession code 2IJ2) with lower-axial thiolate heme ligand. **(b)**

Chemical transformations catalyzed by CYPs. **(c)** Bioinformatic strategy to identify ncCYPs through simple BLAST search, multiple sequence alignment, and sequence similarity analysis. **(d)** PDGAF acts as protein chaperone for glycosyltransferases. **(e)** PDGAF (EryCII, PDB accession code 2YJN) in green overlaid with heme-bound P450 BM3 in gray, and logo plots demonstrating sequence conservation surrounding the axial position (in red). **(f)** ncCYP families (in green), represented as logo plots illustrating sequence conservation surrounding the axial position, along with outgroups (in gray) representing the subfamily of canonical CYPs most closely related to each ncCYP family. Dashes represent alignment gaps. Family sizes are listed in parentheses. **(g)** Putative biosynthetic gene clusters that feature ncCYPs. Genes are color-coded as indicated.

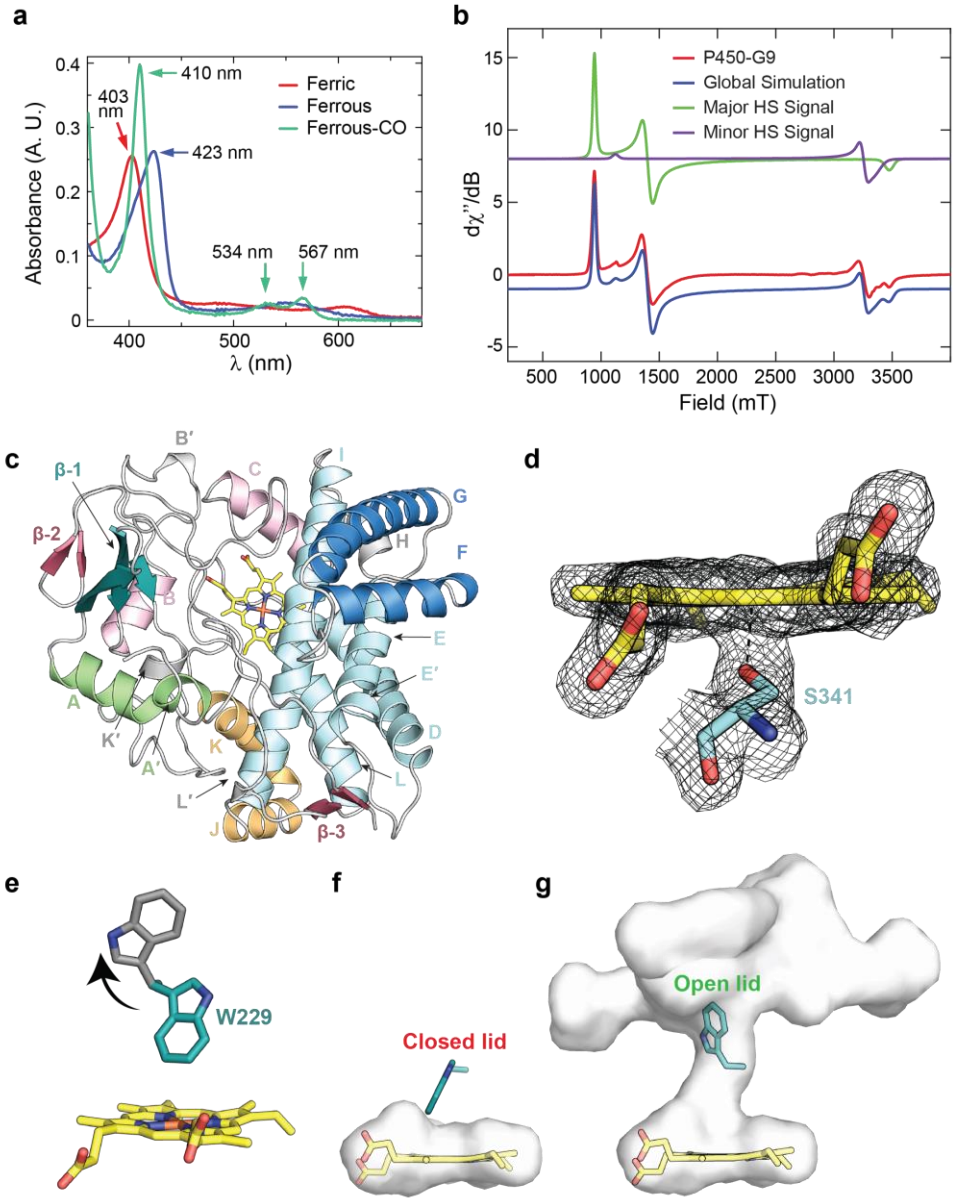

Fig. 2. Spectroscopic and structural analysis of P450-G9, a Ser-ligated CYP. (a) UV/Vis spectra of P450-G9 in resting-state ferric form, reduced ferrous form, and ferrous-CO form. Key spectral features are marked. **(b)** X-band EPR spectrum at 10 K and simulation of ferric high-spin P450-G9. **(c)** Structural overview of P450-G9 (PDB accession code 9CJG) with labeled secondary structures and the heme cofactor shown in yellow. **(d)** Alkoxide ligation to the heme cofactor via S341, as demonstrated by the 2Fo-Fc electron density map contoured at 1.0 sigma. **(e)** P450-G9 features a tryptophan residue in the center of the I helix, preceding a conserved alanine that typically hydrogen bonds to the water molecule occupying the sixth ligand position in cysteine-ligated CYPs. The crystallographically observed conformation of W229 (dark teal) blocks substrate binding and closes the distal solvent channel to heme. Rotation to the most common tryptophan rotamer (gray), however, would allow access to the iron center. **(f)** P450-G9's solvent-inaccessible heme-binding pocket (523 Å^3) with W229 in the observed closed-lid position. **(g)** Upon rotation of W229 to the open lid position, the heme-binding pocket becomes solvent accessible and increases nearly four-fold in volume (1965 Å^3).

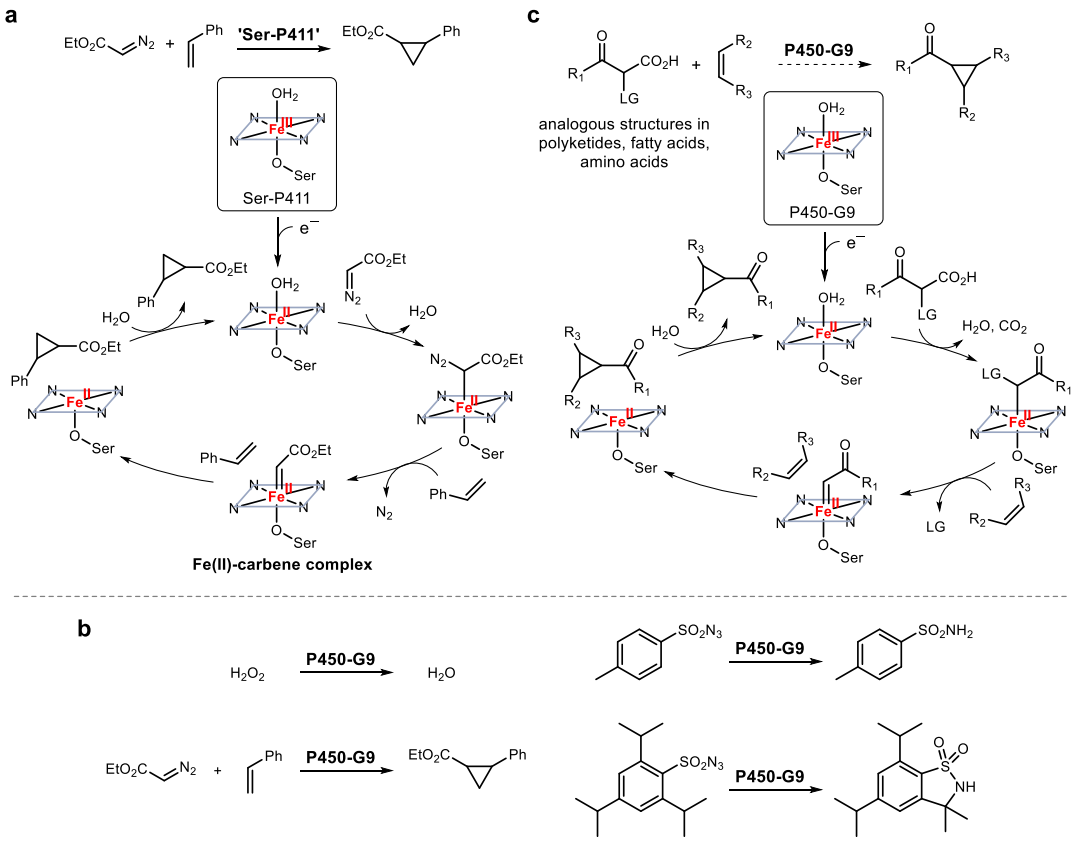

Fig. 3. Reactions of the Ser-ligated P450-G9. (a) Carbene insertion mechanism for an engineered P411 enzyme. (b) Reactions carried out by P450-G9 include H_2O_2 reduction, carbene insertion, tosyl azide reduction and nitrene insertion. (c) Hypothetical carbene insertion reaction with natural substrates.

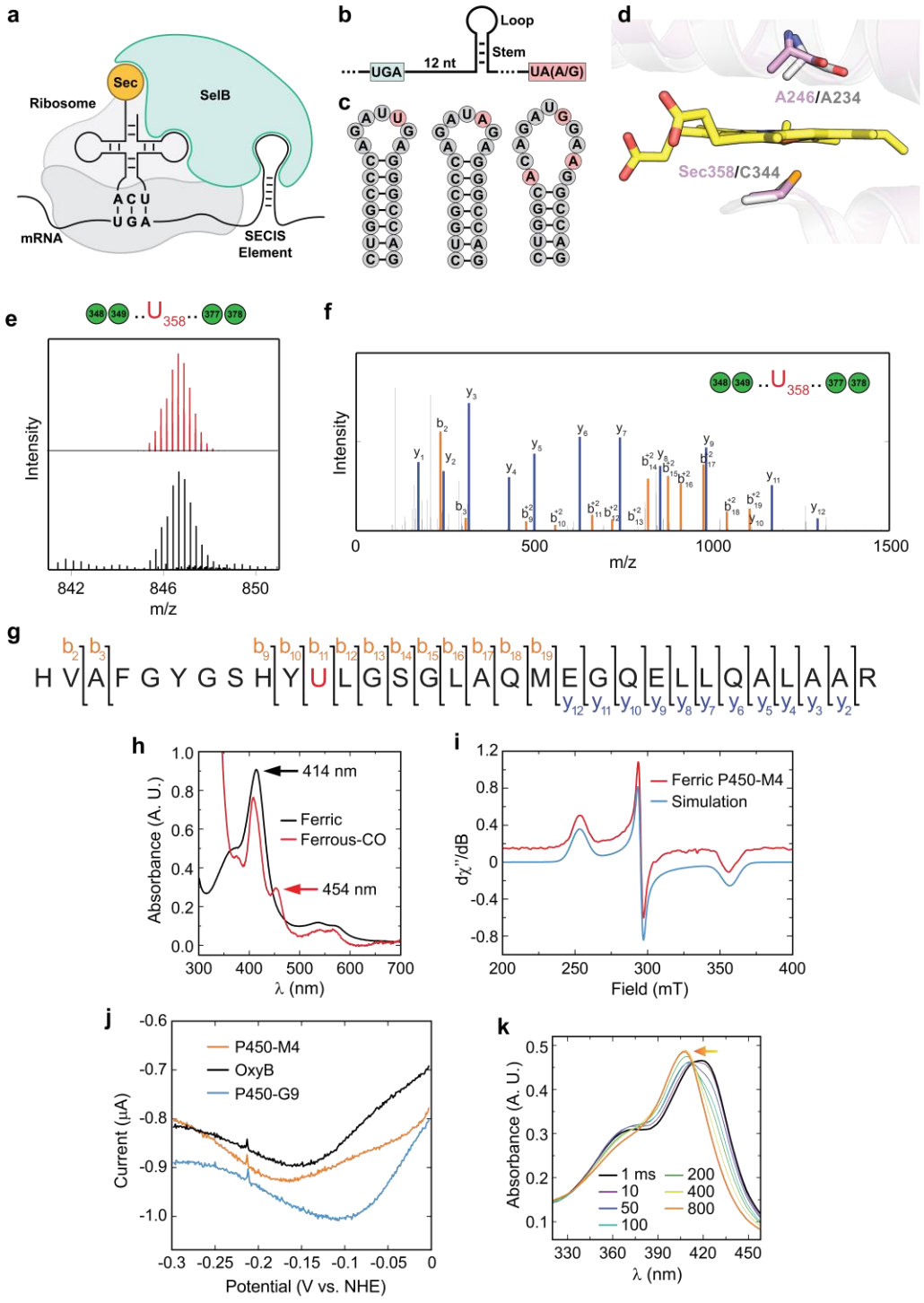

Fig. 4. Selenoprotein biosynthesis and spectroscopic analysis of P450-M4, a Sec-ligated CYP. **(a)** Schematic representation of opal stop codon readthrough during selenoprotein biosynthesis, requiring the elongation factor SelB and downstream mRNA SECIS element. **(b)** Overall mRNA architecture of ncCYP family M transcripts, inferred from genome sequences. For each family member, the SECIS element is located 12 nt downstream of the opal stop codon (UGA). Open reading frames terminate with a non-opal stop codon (UAA or UAG). **(c)** Predicted mRNA SECIS elements within ncCYP family M transcripts

exhibiting stem-loop topology. **(d)** Structural overlay of P450 107H1 (PDB 3EJD) with an AlphaFold2-generated model of P450-M4. The predicted Sec358 overlays with C344, a lower axial ligand in P450 107H1. Note, this figure is only shown to locate the position of Sec358. **(e)** Observed (black) and simulated (red) isotope envelopes for the P450-M4 tryptic fragment containing Sec. **(f)** HR-MS/MS spectra exhibiting b and y ions for the P450-M4 tryptic fragment. **(g)** Inferred MS/MS fragments for the P450-M4 tryptic product from the data in panel **f**. **(h)** UV/Vis spectra of P450-M4 in the resting-state ferric form and the ferrous-CO form. Only partial reduction of Sec-ligated P450-M4 is observed under our experimental conditions. Key spectral features are marked. **(i)** X-band EPR spectrum and simulation of resting state ferric low-spin P450-M4. The upper axial ligand is not known but is likely water. **(j)** Protein film voltammetry to derive redox properties of three enzymes, the Sec-ligated P450-M4, the Cys-ligated OxyB, and the Ser-ligated P450-G9; peak potentials are observed at -148 mV, -123 mV, and -68 mV in the reductive trace, respectively. **(k)** SF UV/Vis spectra after reaction of P450-M4 with *m*CPBA. The spectrum after 1 ms (black) and 800 ms (orange) are shown as thick lines. The transition from the resting state to the new species is marked by the orange arrow.