Chapter One

Sequencing and Analysis of the *Methylococcus capsulatus* (Bath) Soluble Methane Monooxygenase Genes

Introduction

Non-heme carboxylate-bridged diiron centers have emerged as widespread and chemically diverse units in proteins (*1-5*). The prototypical members of this family of proteins are hemerythrin, ribonucleotide reductase R2 subunit (R2) and the soluble methane monooxygenase (sMMO) hydroxylase. A hallmark of these cofactors is their reactivity with dioxygen. For example, hemerythrin is a dioxygen transport protein, utilizing its iron center to reversibly bind O_2. The R2 subunit uses its iron center and one O_2 molecule to generate a tyrosine-based radical, which is subsequently used catalytically by the ribonucleotide reductase R1 subunit to reduce ribonucleotides to deoxyribonucleotides.

A common chemistry catalyzed by carboxylate-bridged diiron centers is monooxygenase activity. This function is exemplified by the soluble methane monooxygenase enzyme system (sMMO), the best studied member of this subclass (*3, 6-8*). It catalyzes the oxidation of methane to methanol, utilizing NADH as an electron source and releasing water as a byproduct. There are three known proteins in the sMMO system. The hydroxylase (H) contains two copies each of three subunits, α, β and γ, with the α subunits housing the diiron active sites. The reductase (R) contains a FAD, an [2Fe-2S] center and a NADH binding site, and is responsible for accepting electrons from NADH and transferring them to the hydroxylase. The third member of the sMMO system, the coupling

protein or protein B (B), is presumably a regulatory protein and modulates the activity of the system depending upon the B:H (vide infra). The sMMO systems from three different methylotrophs, *Methylococcus capsulatus* (Bath), *Methylosinus trichosporium* OB3b OB3B and *Methylocystis* sp. strain M have been characterized, and the proteins from the first two organisms are best understood. The structure of the sMMO hydroxylase from *M. capsulatus* (Bath) as determined by X-ray crystallography is shown in Figure 1-1, together with a schematic of the active site residues (*9-11*). The most striking element of the tertiary structure is a large canyon located in the center of the molecule formed by the α and β subunits.

The genes for the five polypeptides making up the three sMMO proteins are arranged in one operon, along with an open reading frame, designated *orfY*, for a heretofore unknown and unisolated protein (*12-15*). Figure 1-2 shows the sMMO operon organization for all three organisms.

The role of the two ancillary proteins is of current interest, for the chemistry (eq. 1) that takes place at the hydroxylase protein does not occur

$$CH_4 + NADH + H^+ + O_2 \xrightarrow{MMO} CH_3OH + NAD^+ + H_2O \qquad (1)$$

without them. The reductase supplies electrons to the hydroxylase. Characterizing the interactions between the reductase and the hydroxylase is important to the general understanding of the system and may be of use in designing industrially useful variants of the sMMOH. Protein B also plays an important role. Without the presence of protein B, the hydroxylase reduces dioxygen to water or hydrogen peroxide without concomitant oxidation of substrate (*16, 17*). As the protein B:hydroxylase ratio is increased, the coupling of dioxygen to substrate consumption increases. At a 1:1 ratio of protein

B:hydroxylase, the system is fully coupled, and activity is optimal. Protein B has an inhibitory effect when this ratio rises above 1. Understanding the effects of protein B on the active site of the hydroxylase is crucial to the elucidation of the mechanism; the structural consequences of B binding are unknown at this time.

The sMMO genes from all three organisms have been sequenced, but several mistakes were made in the gene sequence of both the *M. capsulatus* (Bath) and *M. trichosporium* OB3b operons. A base insertion that caused a frame shift was found at the end of the *M. capsulatus* (Bath) *mmoY* during X-ray crystallographic work on the enzyme (*9*). More recently, mass spectrometric studies of the hydroxylase from *M. capsulatus* (Bath) disagreed with the molecular mass predicted by the gene sequence (*18*). Many mistakes was found in the *M. trichosporium* OB3b hydroxylase peptide sequence as a result of X-ray crystallographic work (*19*).

There are several other monooxygenases that are quite similar to the methane monooxygenases. The toluene monooxygenases (toluene-2-monooxygenase (*20*), toluene-3-monooxygenase (*21, 22*) and toluene-4-monooxygenase (*23-26*)) are capable of specifically hydroxylating toluene to the corresponding *o*-, *m*- or *p*-cresol. Phenol hydroxylase (PH) from *Pseudomonas* sp. CF600 catalyzes the hydroxylation of phenol and certain methyl- substituted derivatives of phenol to catechol (*27-30*). The alkene monooxygenases (AMO) are capable of epoxidizing various alkenes and alkynes, a reaction that the sMMOs can also perform (*31-34*). All of these proteins have core similarities. They all contain a hydroxylase protein with three subunits and a non-heme carboxylate-bridged diiron center, a reductase and a coupling protein. They also catalyze hydroxylation or epoxidation reactions by using dioxygen as an oxidant. The operons for the phenol hydroxylase and toluene monooxygenase systems are shown in Figure 1-3.

Sequence analysis can be a useful tool in identifying residues that are important in structure and function, especially when combined with crystallographic information about the enzymes in question. There have been several sequence analyses of the MMO proteins in the literature. One of the first was a comparison of the protein sequence of the sMMOH, the structure of which was then unknown, with that of the structurally characterized R2 subunit (*35*). This remarkably prescient paper correctly deduced much of the detail of the diiron active site of sMMO by comparison with R2. A second report compares protein sequence data from several Δ-9 desaturases with those from several ribonucleotide reductases and the soluble methane monooxygenase (*36*). This alignment established that the desaturases were members of a carboxylate-bridged diiron center family of proteins distinct from both the sMMO family of monooxygenases and the ribonucleotide reductase R2 family. Homology between the hydroxylase alpha subunits from the toluene-4-monooxygenase, toluene-3-monooxygenase, toluene-2-monooxygenase, phenol hydroxylase, alkene monooxygenase and sMMO systems has been helpful to establish that all these proteins contain carboxylate-bridged diiron centers (*21*, *23*, *37*).

The reductase sequence from the sMMO family has been analyzed extensively. It is a member of the FNR family of oxidoreductases, which include ferredoxin-$NADP^+$ reductase (*38*) and phthalate dioxygenase reductase (*39*). Sequence comparisons of the reductase proteins can be found in several places in the literature (*24*, *40*, *41*). The coupling proteins have also been compared (*42*).

The sequence analyses thus far reported focus mainly on the diiron center, its ligands, and the substrate binding region. No comparisons of the entire protein sequences have appeared. Such a comparison of all of the proteins of the multicomponent sMMO monooxygenase family would be of value in determining which areas of the protein are essential for activity. Similarities

between these proteins could provide important clues about structure, function and relationships for this subclass of enzymes, including areas contributing to interactions between the components. The docking site for protein B would be of particular interest as a first step to understanding its ability to affect on the iron active site.

Before the sequences can be compared, it is necessary to know that they are correct. Crystallographic work on the hydroxylase indicated that the published gene sequence was inaccurate.

This chapter reports the determination of the correct sequence of the *M. capsulatus* (Bath) soluble methane monooxygenase hydroxylase and sequence comparisons between the methane monooxygenases, alkene monooxygenases, toluene monooxygenases and phenol hydroxylase. Possible roles for several of the universally conserved residues are suggested based upon their locations in the *M. capsulatus* (Bath) sMMOH structure.

Materials and Methods

DNA Sequencing. Sequenase 2.0 kits and enzymes were purchased from United States Biochemicals. TEMED and 29:1 acrylamide:bis were purchased from Biorad. Long Ranger™ acrylamide was purchased from J. T. Baker and $\alpha[^{32}P]$dATP and $\alpha[^{35}S]$dATP from DuPont-NEN. All other chemicals were of the highest quality available.

Oligonucleotide primers were either synthesized on a Cruachem or Expedite DNA synthesizer or purchased from the MIT Biopolymers lab. All oligonucleotides were PAGE-purified according to the Cruachem manufacturer's instructions. Primer sequences are shown in Table 1-1, and the binding locations of those primers relative to the sMMO operon genes are given in Figure 1-4.

Plasmid DNA was purified by using Qiagen purification kits. Plasmid pCH4, which contains a genomic clone of the sMMO operon from *M. capsulatus* (Bath), was a gift from Prof. J. Colin Murrell (Warwick).

Plasmid DNA was sequenced by using the suggested protocol for double-stranded DNA from USB, except that the sequence extension steps were carried out at 45 °C. Both $\alpha[^{32}P]$dATP and $\alpha[^{35}S]$dATP were used in the sequencing reactions; 8% 29:1 acrylamide:bis and 5% Long Ranger™ gels were used in these experiments.

Sequence Analysis. The Wisconsin GCG sequence analysis package (version 7.3) was used to perform most of the sequence analysis (*43*). Initial sequence alignment was made by using the GCG PileUp program with a gap weight of 3 and a gap length weight of 0.1. Maligned (*44*) and BoxShade (*45*) were used to generate protein homology plots. The *M. capsulatus* (Bath) sMMOH oxidized Form I crystal structure was analyzed by using a combination of Quanta97 (*46*) and MolScript 2.02 (*47*). A search profile of an alignment of the three sMMO OrfY peptides was prepared by using the GCG ProfileMake program. Both the GenEMBL (peptide-translated) and SwissProt databases (release 107.0 with updates as of 7/20/98) were searched with the OrfY profile with a gap weight of 4.5 and gap length weight of 0.05.

The DNA sequences used in this chapter were obtained from the GenEMBL database on 7/20/98, through either the GCG Fetch program or the world wide web *Entrez* interface at the National Center for Biotechnology Information <http://www.ncbi.nlm.nih.gov/Entrez/nucleotide.html>. The sequences and accession numbers used are as follows: *Methylococcus capsulatus* (Bath) soluble methane monooxygenase (*13*, *14*) (accession numbers M90050 and M32314, NID g149833, EC 1.14.13.25), *Methylosinus trichosporium* OB3b OB3b soluble methane monooxygenase (*12*) (accession number X55394, NID g44613,

EC 1.14.13.25), *Methylocystis* sp. strain M soluble methane monooxygenase (*15*) (accession number U81594, NID g2098694, EC 1.14.13.25), *Pseudomonas* sp. strain CF600 phenol hydroxylase (EC official name phenol-2-monooxygenase) (*27*) (accession numbers M60276, M37764, M37775 and M57564, NID g151449, EC 1.14.13.7), *Xanthobacter* sp. strain Py2 alkene monooxygenase (*48*) (accession number AJ006979, NID g3250923), *Nocardia corallina* B-276 alkene monooxygenase (*32*) (accession number D37875, NID g529091), *Pseudomonas mendocina* KR1 toluene-4-monooxygenase (*23, 24*) (accession numbers M65106 and M95045, NID's g151590 and g151596), *Pseudomonas pickettii* PKO1 toluene-3-monooxygenase (*21*) (accession number U04052, NID g1590790), and *Pseudomonas* sp. strain JS150 toluene/benzene-2-monooxygenase (*20*) (accession number L40033, NID g1008895). Note that the sequences for the two AMO systems are incomplete at this time. Only the α and β epoxygenase subunits, reductase and coupling sequences are known for the *N. corallina* AMO; only the epoxygenase α subunit sequence from *Xanthobacter* AMO has been reported. Corrections to the *M. trichosporium* OB3b sMMO hydroxylase amino acid sequence were applied as described (*19*). Corrections to the *M. capsulatus* (Bath) sMMO protein sequences were applied as reported herein.

Results

Corrections to the sMMO gene sequences. The DNA sequences of the six genes in the sMMO operon are shown in Figure 1-5, including the changes found as a result of the sequencing described in this chapter. The errors found in the published sequence are shown in Figure 1-6. Change A is a simple base pair switch that alters residue Val84 of the α subunit to Glu. Change B is a double frame shift that inserts a G at base pair 423 of *mmoY* and deletes a G at base pair

434. The net result in the β peptide sequence is that four residues are altered. Change C is a frame shift caused by the deletion of C at position 1086. This frame shift alters the last 27 residues of the β subunit. Change D is a base pair inversion in the *mmoZ* gene, changing residue H21 to Q and residue V22 to L.

An additional two changes to the *M. capsulatus* (Bath) sMMO operon have been determined. An error in the sequence of protein B was discovered by amino terminal sequencing, altering residue F12 to I (*49*). Electrospray mass spectrometry (*50*), MALDI mass spectrometry (*51*) and DNA sequencing (*52*) have confirmed this change, and that the protein sequence is otherwise correct. An error in the *mmoC* gene was found recently by DNA sequencing (*52*). Both of these errors are shown in Figure 1-7. Corrected DNA sequences for all of the sMMO genes have been submitted to the GenEMBL gene database.

Changes to the *M. trichosporium* OB3b gene sequence as suggested by the crystallographically determined electron density of the hydroxylase have also been incorporated as part of the sequence alignment in this chapter (*19*). One of the changes, D209E, is quite clearly correct from both a crystallographic and a sequence alignment viewpoint, as that position is a glutamic acid in every other sMMO family monooxygenase. Forty amino acid changes are made to the *M. trichosporium* OB3b sequence, including a 26-residue stretch at the end of the β subunit. Twenty of those alterations change the residue to its corresponding residue in the *M. capsulatus* (Bath) sequence. The *M. capsulatus* (Bath) structure was used as a starting model for the *M. trichosporium* OB3b enzyme structure determination, due to problems encountered obtaining the phases. The possibility therefore exists that the *M. trichosporium* OB3b model is biased. The changes reported in the literature unfortunately were not confirmed by DNA sequencing. For the purposes of this chapter, all of the changes made to the *M. trichosporium* OB3b hydroxylase sequence are taken as correct, and many of them

probably are, but it is impossible to determine which are correct and which are artifacts. Clearly, confirmation of the *M. trichosporium* OB3b sequence is needed.

Roles for genes from the sMMO family operons. The relationships between the genes and gene products of the nine proteins examined in this chapter, *M. capsulatus* (Bath) sMMO, *M. trichosporium* OB3b sMMO, *M.* sp. strain M sMMO, *P. mendocina* toluene-4-monooxygenase (T4M), *P. pickettii* PKO1 toluene-3-monooxygenase (T3M), *P.* sp. toluene-2-monooxygenase (T2M), *P. pudita* phenol hydroxylase (PH), *N. corallina* alkene monooxygenase (AMO) and *Xanthobacter* sp. Py2 alkene monooxygenase, are shown in Figure 1-8. All of these systems contain a hydroxylase that comprises three subunits. The homologies between the sMMOHα analogues are all very high, and it is clear which genes correspond to this important subunit. The same is true of the reductase and coupling protein analogues. Although the sMMOβ analogues all have homology to each other, in some cases (*dmpL*, *tbmB*, *tbuA2* and *tmoE*), they have higher homologies to the *mmoX* genes of sMMO (*20, 21*). With knowledge of the crystal structure of the sMMO hydroxylase, this is understandable. Although the β subunit of the *M. capsulatus* (Bath) hydroxylase is not homologous to the α subunit, they have similar folding topologies (*9*). It is possible that a protein with sequence homologous to sMMOHα might have a tertiary structure like that of the sMMOβ subunit, and would fulfill its function in the quaternary structure of the hydroxylase. It seems likely that the β subunit is evolutionarily related to the α subunit, and has retained structural aspects from its precursor. Note that the iron binding ligands are not present in any of the β subunit analogues. The identity of the γ subunit in each hydroxylase is not clear from homology in many of the cases. Because the P4 peptide of the phenol hydroxylase has been identified as the γ subunit in the hydroxylase of that system (*28*), and because its gene, *dmpO*, is homologous to *tbmE* of the T2M system (35.3% identity between the two) (*20*),

we can assign that role to those two proteins. Amino-terminal sequencing of the purified hydroxylase from T4M indicates that the TmoB peptide is its third subunit (*25*), and its homology with *tbuU* from T3M (38.8% identity) (*20*) completes the set of known γ analogues. The T3M and T4M systems both contain a ferredoxin (TbuB and TmoC) with a Rieske-type [2Fe-2S] center, which is necessary for turnover in the T4M system along with the reductase (*25*). This ferredoxin is not present in the other systems. Finally, there is a gene of unknown function in many of the systems: *orfY* in the sMMOs, *dmpK* in PH and *tbmA* in T2M. There is homology between the *orfY*s in the sMMOs, and between dmpK and tbmA. Interestingly, although the PH system is fully functional *in vitro* without the dmpK gene product (P0), it is necessary *in vivo* for growth on phenol (*30*). Note that the alkene monooxygenase operons are only partially sequenced. The assignment of the AMO genes to functions in the AMO system is clear from homology. Only the sMMOHα, sMMOHβ, sMMOB and sMMOR homologies are high enough to align the peptide sequences from all nine proteins.

Comparisons between the sMMO family protein sequences. Percent identity and percent dissimilarity between MMOHα, MMOHβ, MMOB and MMOR analogues are shown in Figure 1-9. The proteins can be divided into three groups that appear related: the sMMO systems, the T3M and T4M systems and the PH and T2M systems. The *N. corallina* AMO hydroxylase is quite similar to the sMMOH, but the reductase and coupling protein from that system are much different from any of the other systems. Surprisingly, the *Xanthobacter* AMO hydroxylase α subunit is very different from its homologues.

Sequence alignments for the α and β subunits of the hydroxylases/epoxidases, coupling proteins and reductases from the nine enzyme systems are shown in Figures 1-10 through 1-13 respectively.

Search for OrfY analogues. The three OrfY putative protein sequences from the sMMO systems were aligned, and a search profile based on the consensus sequence was prepared. Both a peptide-translated GenEMBL database and the SwissProt protein database were searched by using this profile. The top candidates were a putative RNA polymerase σ factor in the *Streptomyces coelicolor* rpoX operon, the E6 DNA-binding zinc-finger protein from Human Papillomavirus type 30, insertion element IS402 from *Pseudomonas cepacia* and *Rattus norvegicus* cortexin. None of these proteins exhibited significant homology with the OrfY sequences.

The sMMOHα analogues. The sequence alignment of the MMOHα analogues is shown in Figure 1-10. Clearly, the MMOHα subunit has the most homology to its counterparts in the other monooxygenases. The α subunit is mostly comprised of α-helices. It is apparent that many of the identical residues are spaced every 3-4 residues, indicating conservation of one side of an α-helix.

The MMOHα residues that are conserved absolutely between all the proteins can be divided into four groups. The first group is a series of residues in and around the active site of the enzyme. A view of the active site showing the residues discussed here is shown in Figure 1-14. The iron ligands (E114, E144, H147, E209, E243 and H246) are, of course, absolutely conserved. The iron ligand residue sequence pattern E•••EX_2H has been noted to be a hallmark of proteins containing carboxylate-bridged non-heme diiron centers (*53*), and is in fact the only sequence that is conserved all across the sMMO, R2 and stearoyl-ACP desaturase families (*36*).

The active site can be divided into two regions. The first lies above the iron atoms as depicted in Figure 1-14, where there are few conserved residues. The top side of the active site might be where substrates bind, and where the monooxygenase chemistry takes place. This hypothesis is supported by a

number of observations. In the R2 family, a closely related set of enzymes, this side of the active site is where the tyrosine radical is located (*54-56*). The iron ligands on this side of the iron site are quite mobile (*10, 11*); such lability might indicate that this location is where dioxygen binds. Finally, the lack of conserved residues in this area is to be expected if it is the site of substrate binding. A phenol binding pocket is likely to be very different in size and shape from a methane binding pocket. The only two conserved residues in this area are both protic, T213 and N214. Threonine 213 is in a position homologous to a serine residue in the R2 subunit thought to be responsible for proton flux into the active site (*57*). Either one of these residues might assist in the delivery of protons during the hydroxylation reaction.

The bottom half of the active site is, by contrast, quite well conserved. A set of residues involved in hydrogen bonding between the F and C helices (D143, R146, S238, D242 and R245) are absolutely conserved among all sMMO family proteins. A schematic showing residues within hydrogen bonding distance of one another is presented in Figure 1-15. These residues might be part of a framework to hold the iron center in place, or possibly to deliver protons to the active site. There are also two residues that are conserved for clear steric reasons. Both A117 and G250 are located in positions where the packing is very tight, and a larger residue would not fit. Glycine 250 in particular is located where helices E and F come very close to one another. Finally, there is a triad of surface accessible residues, comprising A224, G228 and D229, located at the turn between helices E and F. It is possible that they form a "handle" on one end of the E/F helix half of the active site. One of the ancillary proteins might interact with the "handle", causing the E/F helices to shift, thereby altering the coordination environment at the iron center.

The conserved residues in other parts of the α subunit are shown in Figure 1-16. A group of aromatic residues located at the top of the hydroxylase (Y292, W371, Y376) in close proximity to each other are shown in Figure 1-16 in red. Along with those aromatics, a proline (P377) is absolutely conserved and may be important structurally. This entire cluster of residues may be some kind of docking site for another protein, or part of an electron transfer path. It has been suggested that this area of the hydroxylase is the site where the reductase binds (*58*).

Another set of residues is located in the second domain of the hydroxylase, shown in white in Figure 1-16. They comprise residues P424, G443, P461 and Y464. They are located slightly beneath the surface of the protein, near the γ subunit interface. No function for these residues is obvious. It is possible that they are involved in interactions with the γ subunit.

The fourth and final set of residues is located on the surface of the protein in the "canyon" area, above the active site. This residue group comprises Y67, K74, L321, G325, and P329, which can be seen in yellow in Figure 1-16. It has been hypothesized that the canyon is a docking site for some other protein, possibly protein B (*9*). These residues may be important in mediating the interactions between the two proteins. In particular, K74 and Y67 are very close to the surface and are located in the canyon. Combined with the E/F helix "handle" described above, these residues might be key interaction points between the coupling protein and the hydroxylase.

There are a number of residues that show up in the alignment that are probably artifacts of the alignment algorithm, due to proximity to gaps and isolation from other conserved residues. These residues are Y292, H344 and G492.

The sMMOHβ analogues. There is less identity in the MMOHβ subunit analogues than for the MMOHα subunit. It is therefore difficult to guess at the function of any of the conserved residues, which for MMOHβ are shown in white in Figure 1-17. They can be divided into three general areas. The first is at the interface between the α and β subunits, and consists of D100, P101 and E185. These residues might be important in intersubunit interactions, although there are no conserved hydrogen bonding or salt bridge partners in the α subunit. A second group of residues can be found under the surface of the β subunit, comprising W218, R228 and A331. The third group of residues, D240, E243, Q313 and W320, is very near the surface and consists of mostly polar residues. Lack of knowledge about the function of the β subunit precludes making conclusions about the functions of these residues. They might be involved in protein folding, interactions with other ancillary proteins, or in electron transfer. Cross-linking experiments in the *M. trichosporium* OB3b sMMO system indicate that the reductase has a binding site on the β subunit (*59*). One or more of the conserved areas listed above may be part of a reductase binding site.

The coupling protein analogues. There is little identity between the coupling proteins from the sMMO family. The five absolutely conserved residues (V38, E53, I79, G97, G144) are mapped in white onto the NMR structure of the P2 protein (*42*) from the phenol hydroxylase system in Figure 1-18. There are also a number of closely homologous residues of interest near these absolutely conserved residues, including I/V52, I/L85, D/E94, R/K98 and I/V107, shown in yellow in Figure 1-18. The charged residues in this group might have partners in the charged surface residues in the hydroxylase canyon. This idea is consistent with cross-linking studies of the *M. trichosporium* OB3b sMMO system, where protein B was shown to have a binding site on the α subunit of the hydroxylase (*59*).

The reductase analogues. Conserved residues in the reductase have been discussed previously, as the sMMOR is a member of the FNR family of oxidoreductases, which contain well-characterized [2Fe-2S] and FAD cofactor sites and pockets for binding NADH (*24*, *38*, *40*). No reductase from a sMMO-family multicomponent monooxygenase has yet been characterized crystallographically. The T3M and T4M systems have electron delivery systems that are different from the other seven. They contain ferredoxins that are necessary for multiple turnover. Presumably, these ferredoxins ferry electrons from the reductase to the hydroxylase, but such a function has not been shown conclusively. If so, then identification of the reductase-hydroxylase interaction region from this alignment is problematic, since the site does not exist in two of the proteins.

Discussion

Proteins housing carboxylate-bridged non-heme diiron centers can be divided into two structurally distinct classes, the hemerythrin-like class and R2/sMMO-like class (*36*). The latter can be further divided into families of proteins, including the R2, stearoyl-ACP desaturase and sMMO families. The sMMO family has a number of distinguishing characteristics, including the ability to catalyze hydroxylation/epoxidation chemistry, the presence of a three-subunit hydroxylase that houses diiron centers in its α subunits, and the presence of a coupling protein that is necessary for optimal activity. This chapter compares the sequences of known members of this family in an effort to determine which areas of the protein are important for its unique features.

Implications of the corrected M. capsulatus *(Bath) DNA sequence.* The corrected sequence of the sMMO operon from *M. capsulatus* (Bath) presented

here is consistent with the crystallographically-determined electron density for the sMMO hydroxylase. All errors in the reported sequence can be easily attributed to gel-reading errors or typographical mistakes. It can also explain recent electrospray-ionization (ESI) mass spectrometric results (*18*), which were clearly not consistent with the published sequence. The unexplained ESI data obtained are listed in Table 1-2. Based on the mass spectrometric data obtained, the authors correctly deduced the actual sequence of the β and γ subunit, although they could not distinguish between isoleucine and leucine. Due to apparent misinterpretation of MS/MS sequencing of the α subunit peptides (data not shown in the published MS work), however, they misidentified L82 as T82, Q83 as R83 and failed to correct V84 to D84. These sequence changes, if correct, would require several changes to the DNA sequence of the *mmoX* gene, and these changes would be difficult to imagine as typographical errors. The sequence as reported in Figure 1-6, however, is consistent with the electrospray mass spectrometric data, clear from the cDNA sequencing and easy to imagine as a simple typographical TA base inversion at base 251 of *mmoX*. Although mass spectrometry is a useful tool, it is best used in concert with other complementary techniques to obtain meaningful, clear results.

Sequence alignment of sMMO family proteins. Comparison of the α and β hydroxylase subunits, coupling proteins and reductases from the known sMMO family proteins has revealed an interesting set of conserved residues. Much of the data can be interpreted in light of what is known about the sMMO function and structure.

Active site of the hydroxylase. A set of conserved residues corresponding to the immediate area of the active site was identified in the sMMOHα subunit. The residues comprised the iron ligands, a network of hydrogen bonds, a possible proton delivery site, and sterically necessary conservations. The conserved

residues were restricted to the area "below" the active site as it is usually drawn, which is not surprising, given that the substrate is thought to bind "above" the active site, and the proteins examined have very different substrates.

Understanding protein-protein interactions in sMMO. The structural and functional effects of binding of the ancillary sMMO proteins, protein B and the reductase, to the hydroxylase are important for understanding the mechanism of methane oxidation. Determining how and where the reductase binds is crucial to identifying the electron transfer pathway from the [2Fe-2S] site to the diiron active site. Characterizing the effects of protein B is also critical. The hydroxylation reaction does not occur without protein B, yet at high levels, protein B inhibits the reaction. The binding of protein B must have a major effect on the active site of the enzyme, without knowledge of which, an understanding of the sMMO systems cannot be complete.

A possible protein B binding site was identified on the surface of the hydroxylase canyon region. It is located just above the active site and could have a direct effect on the four helix bundle that surrounds the diiron center. Part of this binding site might include the residues A224, G228 and D229, which are located at the bend where helix E and helix F meet, as depicted in Figure 1-14. Crystallography reveals residues E209 and E243 to be the most mobile of all the iron ligands, and they are located on helices E and F, respectively. Protein B could affect the coordination of the active site irons through use of the E/F helix turn as a "handle." Additionally, a set of residues that may be part of a binding site at or near the surface of the canyon was identified as depicted in Figure 1-16.

A set of aromatic residues where the electron transfer pathway might begin was also identified at the top of the hydroxylase molecule. The location of these residues is consistent with a model placing the reductase binding site on the top of the α subunit. Another possible binding site for the reductase was

located on the surface of the β subunit. Two possible models of the binding sites of the ancillary sMMO proteins are shown in Figure 1-19.

The sequence alignments described in this chapter provide information about functionally essential areas of the sMMO proteins. These areas are obvious targets for mutagenesis. The proposed binding sites for the reductase and coupling proteins are particularly attractive in this regard. Initial studies in this vein might involve mutating individual residues in the proposed binding sites and measuring the effects on protein-protein interactions, electron transfer or hydroxylase activity.

Many of the areas that are well conserved in the sMMO family have obscure functionality. Perhaps as more is learned about the structure and function of these interesting and useful protein systems, the functions of these residues will be understood.

References:

1. Lippard, S. J., and Berg, J. M. (1994) *Principles of Bioinorganic Chemistry*, 1st ed., University Science Books, Mill Valley, California.

2. Feig, A. L., and Lippard, S. J. (1994) *Chem. Rev. 94*, 759-805.

3. Wallar, B. J., and Lipscomb, J. D. (1996) *Chem. Rev. 96*, 2625-2657.

4. Kurtz, D. M. (1990) *Chem. Rev. 90*, 585-606.

5. Vincent, J. B., Olivier-Lilley, G. L., and Averill, B. A. (1990) *Chem. Rev. 90*, 1447-1467.

6. Valentine, A. M., and Lippard, S. J. (1997) *J. Chem. Soc., Dalton Trans.*, 3925-3931.

7. Liu, K. E., and Lippard, S. J. (1995) in *Adv. Inorg. Chem.* (Sykes, A. G., Ed.) pp 263-289, Academic Press, San Diego.

8. Lipscomb, J. D. (1994) *Annu. Rev. Microbiol. 48*, 371-399.

9. Rosenzweig, A. C., Frederick, C. A., Lippard, S. J., and Nordlund, P. (1993) *Nature 366*, 537-543.

10. Rosenzweig, A. C., Nordlund, P., Takahara, P. M., Frederick, C. A., and Lippard, S. J. (1995) *Chemistry & Biology 2*, 409-418.

11. Rosenzweig, A. C., Frederick, C. A., and Lippard, S. J. (1996) in *Microbial Growth on C1 Compounds* (Lidstrom, M. E., and Tabita, F. R., Eds.) pp 141-149, Kluwer Academic Publishers, Dordrecht.

12. Cardy, D. L. N., Laidler, V., Salmond, G. P. C., and Murrell, J. C. (1991) *Molec. Microbiol. 5*, 335-342.

13. Stainthorpe, A. C., Murrell, J. C., Salmond, G. P. C., Dalton, H., and Lees, V. (1989) *Arch. Microbiol. 152*, 154-159.

14. Stainthorpe, A. C., Lees, V., Salmond, G. P. C., Dalton, H., and Murrell, J. C. (1990) *Gene 91*, 27-34.

15. McDonald, I. R., Uchiyama, H., Kambe, S., Yagi, O., and Murrell, J. C. (1997) *App. Env. Microbiol. 63*, 1898-1904.

16. Green, J., and Dalton, H. (1985) *J. Biol. Chem. 260*, 15795-15801.

17. Gassner, G., and Lippard, S. J. (1998) *Manuscript in Preparation*.

18. Buzy, A., Millar, A. L., Legros, V., Wilkins, P. C., Dalton, H., and Jennings, K. R. (1998) *Eur. J. Biochem. 254*, 602-609.

19. Elango, N., Radhakrishnan, R., Froland, W., Wallar, B. J., Earhart, C. A., Lipscomb, J. D., and Ohlendorf, D. H. (1997) *Protein Science 6*, 556-568.

20. Johnson, G. R., and Olsen, R. H. (1995) *App. Env. Microbiol. 61*, 3336-3346.

21. Byrne, A. M., Kukor, J. J., and Olsen, R. H. (1995) *J. Bact. 178*, 65-70.

22. Olsen, R. H., Kukor, J. J., and Kaphammer, B. (1994) *J. Bact. 176*, 3749-3756.

23. Yen, K.-M., Karl, M. R., Blatt, L. M., Simon, M. J., Winter, R. B., Fausset, P. R., Lu, H. S., Harcourt, A. A., and Chen, K. K. (1991) *J. Bact. 173*, 5315-5327.

24. Yen, K.-M., and Karl, M. R. (1992) *J. Bact. 174*, 7253-7261.

25. Pikus, J. D., Studts, J. M., Achim, C., Kauffmann, K. E., Münck, E., Steffan, R. J., McClay, K., and Fox, B. G. (1996) *Biochemistry 35*, 9106-9119.

26. Whited, G. M., and Gibson, D. T. (1991) *J. Bact. 173*, 3010-3016.

27. Nordlund, I., Powlowski, J., and Shingler, V. (1990) *J. Bact. 172*, 6826-6833.

28. Nordlund, I., Powlowski, J., Hagström, Å., and Shingler, V. (1993) *J. Gen. Microbiol. 139*, 2695-2703.

29. Shingler, V., Franklin, F. C. H., Tsuda, M., Holroyd, D., and Bagdasarian, M. (1989) *Journal of General Microbiology 135*, 1083-1092.

30. Powlowski, J., and Shingler, V. (1990) *J. Bact.*, 6834-6840.

31. Gallagher, S. C., Cammack, R., and Dalton, H. (1997) *Eur. J. Biochem. 247*, 635-641.

32. Saeki, H., and Furuhashi, K. (1994) *J. Ferment. Bioeng. 78*, 399-406.

33. Miura, A., and Dalton, H. (1995) *Biosci. Biotech. Biochem. 263*, 853-859.

34. Small, F. J., and Ensign, S. A. (1997) *J. Biol. Chem. 272*, 24913-24920.

35. Nordlund, P., Dalton, H., and Eklund, H. (1992) *FEBS Lett. 307*, 257-262.

36. Fox, B. G., Shanklin, J., Ai, J., Loehr, T. M., and Sanders-Loehr, J. (1994) *Biochemistry 33*, 12776-12786.

37. Gallagher, S. C., George, A., and Dalton, H. (1998) *Eur. J. Biochem. 254*, 480-489.

38. Karplus, P. A., Daniels, M. J., and Herriot, J. R. (1991) *Science 251*, 60-66.

39. Correll, C. C., Batie, C. J., Ballou, D. P., and Ludwig, M. L. (1992) *Science 258*, 1604-1610.

40. Andrews, S. C., Shipley, D., Keen, J. N., Findlay, J. B. C., Harrison, P. M., and Guest, J. R. (1992) *FEBS Letters 302*, 247-252.

41. Neidle, E. L., Hartnett, C., Ornston, L. N., Bairoch, A., Rekik, M., and Harayama, S. (1991) *J. Bact. 173*, 5385-5395.

42. Qian, H., Edlund, U., Powlowski, J., Shingler, V., and Sethson, I. (1997) *Biochemistry 36*, 495-504.

43. Genetics Computer Group (1991) , The Wisconsin Package, Madison, Wisc.

44. Clark, S. (1991) Maligned.

45. Hofmann, K., and Baron, M. D. (1998) BoxShade , Pirbright, Surrey UK.

46. Molecular Simulations, Inc. (1997) Quanta97 , Molecular Simulations, Inc., San Diego, CA.

47. Kraulis, P. J. (1991) *J. Appl. Cryst. 24*, 946-950.

48. Zhou, N., Jenkins, A., Chion, C. C. K., and Leak, D. J. (1998) *Submitted for Publication*.

49. Wu, W., Rosenzweig, A., and Lippard, S. J. (1989) *Unpublished Results*.

50. Brandstetter, H., and Lippard, S. J. (1997) *Unpublished Results*.

51. Coufal, D., Wolf, S., Biemann, K., and Lippard, S. J. (1998) *Unpublished Results*.

52. Blazyk, J., and Lippard, S. J. (1998) *Unpublished Results*.

53. Nordlund, P., Sjöberg, B.-M., and Eklund, H. (1990) *Nature 345*, 593-598.

54. Nordlund, P., Åberg, A., Uhlin, U., and Eklund, H. (1993) *Biochem. Soc. Trans. 21*, 735-738.

55. Nordlund, P., and Eklund, H. (1993) *J. Mol. Biol. 232*, 123-164.

56. Kauppi, B., Nielsen, B. B., Ramaswamy, S., Larsen, I. K., Thelander, M., Thelander, L., and Eklund, H. (1996) *J. Mol. Biol. 262*, 706-720.

57. Regnström, K., Åberg, A., Ormö, M., Sahlin, M., and Sjöberg, B.-M. (1994) *J. Biol. Chem. 269*, 6355-6361.

58. Rosenzweig, A. C., Brandstetter, H., Whittington, D. A., Nordlund, P., Lippard, S. J., and Frederick, C. A. (1997) *Proteins 29*, 141-152.

59. Fox, B. G., Liu, Y., Dege, J. E., and Lipscomb, J. D. (1991) *J. Biol. Chem. 266*, 540-550.

Table 1-1. sMMO Hydroxylase Sequencing Primer Names and Binding Locations.

Table 1-2. Electrospray-Ionization Mass Spectrometric Results. Taken from Buzy et al. (*18*). Only selected results that exhibited differences from the expected, published sequence are shown. The amino-terminal methionine was cleaved off in all cases. All results can be explained by the revised sequence as described herein. The α′ refers to α subunit that is cleaved between T7 and K8, and the α″ refers to α subunit cleaved between A9 and A10. The αT3a, βT2, βCB8 and γT4 peptides are cyanogen bromide/trypsin cleavage products; the parenthetical numbers identify the residue range that they cover. These peptides cover the areas of the sMMOH peptides that were not consistent with the published sequence.

Figure 1-1. ***Methylococcus capsulatus*** **(Bath) sMMO Hydroxylase Structure.** a. Structure of the *Methylococcus capsulatus* (Bath) sMMO hydroxylase (crystal form I, diferric state) (*9*). The α-subunits are shown in purple, the β-subunits are in aqua, and the γ-subunits are in green. The iron atoms are colored green, and one of the two protomers is partially transparent to emphasize the location of the iron center within the α subunit. The canyon is visible between the two α subunits. This Figure was prepared by using MolScript (*47*). b. Structure of the active site, showing all ligands, including protein residues, a hydroxide bridge, a terminal water and an exogenous acetate bridge.

Figure 1-2. Organization of the sMMO operons from three different organisms. Gene arrangement is identical between the organisms.

Figure 1-3. Organization of other sMMO family multicomponent monooxygenase operons. The gene arrangement is quite different from the sMMO operons and from each other. The alkene monooxygenases are not included, as neither of them are fully sequenced.

Figure 1-4. Binding locations of primers used in sequencing the *M. capsulatus* (Bath) sMMO hydroxylase genes. Primers beginning with a D were the gift of Thanos Salifoglou.

Figure 1-5. DNA sequences and deduced peptide sequences of the *M. capsulatus* (Bath) sMMO genes. The hydroxylase, protein B, and reductase genes can be found in the GenEMBL database. All errors in the original sequence as listed in Figures 1-6 and 1-7 are incorporated in this listing.

M. capsulatus (Bath) *mmoX* sequence

```
        10                  30                  50
         .         .         .         .         .         .
ATGGCACTTAGCACCGCAACCAAGGCCGCGACGGACGCGCTGGCTGCCAATCGGGCACCC
MetAlaLeuSerThrAlaThrLysAlaAlaThrAspAlaLeuAlaAlaAsnArgAlaPro
  1

        70                  90                  110
         .         .         .         .         .         .
ACCAGCGTGAATGCACAGGAAGTGCACCGTTGGCTCCAGAGCTTCAACTGGGATTTCAAG
ThrSerValAsnAlaGlnGluValHisArgTrpLeuGlnSerPheAsnTrpAspPheLys
 21

       130                 150                 170
         .         .         .         .         .         .
AACAACCGGACCAAGTACGCCACCAAGTACAAGATGGCGAACGAGACCAAGGAACAGTTC
AsnAsnArgThrLysTyrAlaThrLysTyrLysMetAlaAsnGluThrLysGluGlnPhe
 41

       190                 210                 230
         .         .         .         .         .         .
AAGCTGATCGCCAAGGAATATGCGCGCATGGAGGCAGTCAAGGACGAAAGGCAGTTCGGT
LysLeuIleAlaLysGluTyrAlaArgMetGluAlaValLysAspGluArgGlnPheGly
 61

       250                 270                 290
         .         .         .         .         .         .
AGCCTGCAGGATGCGCTGACCCGCCTCAACGCCGGTGTTCGCGTTCATCCGAAGTGGAAC
SerLeuGlnAspAlaLeuThrArgLeuAsnAlaGlyValArgValHisProLysTrpAsn
 81

       310                 330                 350
         .         .         .         .         .         .
GAGACCATGAAAGTGGTTTCGAACTTCCTGGAAGTGGGCGAATACAACGCCATCGCCGCT
GluThrMetLysValValSerAsnPheLeuGluValGlyGluTyrAsnAlaIleAlaAla
101

       370                 390                 410
         .         .         .         .         .         .
ACCGGGATGCTGTGGGATTCCGCCCAGGCGGCGGAACAGAAGAACGGCTATCTGGCCCAG
ThrGlyMetLeuTrpAspSerAlaGlnAlaAlaGluGlnLysAsnGlyTyrLeuAlaGln
121

       430                 450                 470
         .         .         .         .         .         .
GTGTTGGATGAAATCCGCCACACCCACCAGTGTGCCTACGTCAACTACTACTTCGCGAAG
ValLeuAspGluIleArgHisThrHisGlnCysAlaTyrValAsnTyrTyrPheAlaLys
141
```

```
       490                 510                 530
         .         .         .         .         .         .
AACGGCCAGGACCCGGCCGGTCACAACGATGCTCGCCGCACCCGTACCATCGGTCCGCTG
AsnGlyGlnAspProAlaGlyHisAsnAspAlaArgArgThrArgThrIleGlyProLeu
161
       550                 570                 590
         .         .         .         .         .         .
TGGAAGGGCATGAAGCGCGTGTTTTCCGACGGCTTCATTTCCGGCGACGCCGTGGAATGC
TrpLysGlyMetLysArgValPheSerAspGlyPheIleSerGlyAspAlaValGluCys
181

       610                 630                 650
         .         .         .         .         .         .
TCCCTCAACCTGCAGCTGGTGGGTGAGGCCTGCTTCACCAATCCGCTGATCGTCGCAGTG
SerLeuAsnLeuGlnLeuValGlyGluAlaCysPheThrAsnProLeuIleValAlaVal
201

       670                 690                 710
         .         .         .         .         .         .
ACCGAATGGGCTGCCGCCAACGGCGATGAAATCACCCCGACGGTGTTCCTGTCGATCGAG
ThrGluTrpAlaAlaAlaAsnGlyAspGluIleThrProThrValPheLeuSerIleGlu
221

       730                 750                 770
         .         .         .         .         .         .
ACCGACGAACTGCGCCACATGGCCAACGGTTACCAGACCGTCGTTTCCATCGCCAACGAT
ThrAspGluLeuArgHisMetAlaAsnGlyTyrGlnThrValValSerIleAlaAsnAsp
241

       790                 810                 830
         .         .         .         .         .         .
CCGGCTTCCGCCAAGTATCTCAACACGGACCTGAACAACGCCTTCTGGACCCAGCAGAAG
ProAlaSerAlaLysTyrLeuAsnThrAspLeuAsnAsnAlaPheTrpThrGlnGlnLys
261

       850                 870                 890
         .         .         .         .         .         .
TACTTCACGCCGGTGTTGGGCATGCTGTTCGAGTATGGCTCCAAGTTCAAGGTCGAGCCG
TyrPheThrProValLeuGlyMetLeuPheGluTyrGlySerLysPheLysValGluPro
281

       910                 930                 950
         .         .         .         .         .         .
TGGGTCAAGACGTGGAACCGCTGGGTGTACGAGGACTGGGGCGGCATCTGGATCGGCCGT
TrpValLysThrTrpAsnArgTrpValTyrGluAspTrpGlyGlyIleTrpIleGlyArg
301
```

```
       970                 990                1010
         .         .         .         .         .         .
CTGGGCAAGTACGGGGTGGAGTCGCCGCGCAGCCTCAAGGACGCCAAGCAGGACGCTTAC
LeuGlyLysTyrGlyValGluSerProArgSerLeuLysAspAlaLysGlnAspAlaTyr
321

      1030                1050                1070
         .         .         .         .         .         .
TGGGCTCACCACGACCTGTATCTGCTGGCTTATGCGCTGTGGCCGACCGGCTTCTTCCGT
TrpAlaHisHisAspLeuTyrLeuLeuAlaTyrAlaLeuTrpProThrGlyPhePheArg
341

      1090                1110                1130
         .         .         .         .         .         .
CTGGCGCTGCCGGATCAGGAAGAAATGGAGTGGTTCGAGGCCAACTACCCCGGCTGGTAC
LeuAlaLeuProAspGlnGluGluMetGluTrpPheGluAlaAsnTyrProGlyTrpTyr
361

      1150                1170                1190
         .         .         .         .         .         .
GACCACTACGGCAAGATCTACGAGGAATGGCGCGCCCGCGGTTGCGAGGATCCGTCCTCG
AspHisTyrGlyLysIleTyrGluGluTrpArgAlaArgGlyCysGluAspProSerSer
381

      1210                1230                1250
         .         .         .         .         .         .
GGCTTCATCCCGCTGATGTGGTTCATCGAAAACAACCATCCCATCTACATCGATCGCGTG
GlyPheIleProLeuMetTrpPheIleGluAsnAsnHisProIleTyrIleAspArgVal
401

      1270                1290                1310
         .         .         .         .         .         .
TCGCAAGTGCCGTTCTGCCCGAGCTTGGCCAAGGGCGCCAGCACCCTGCGCGTGCACGAG
SerGlnValProPheCysProSerLeuAlaLysGlyAlaSerThrLeuArgValHisGlu
421

      1330                1350                1370
         .         .         .         .         .         .
TACAACGGCCAGATGCACACCTTCAGCGACCAGTGGGGCGAGCGCATGTGGCTGGCCGAG
TyrAsnGlyGlnMetHisThrPheSerAspGlnTrpGlyGluArgMetTrpLeuAlaGlu
441

      1390                1410                1430
         .         .         .         .         .         .
CCGGAGCGCTACGAGTGCCAGAACATCTTCGAACAGTACGAAGGACGCGAACTGTCGGAA
ProGluArgTyrGluCysGlnAsnIlePheGluGlnTyrGluGlyArgGluLeuSerGlu
461
```

```
      1450                1470                1490
         .          .          .          .          .          .
GTGATCGCCGAACTGCACGGGCTGCGCAGTGATGGCAAGACCCTGATCGCCCAGCCGCAT
ValIleAlaGluLeuHisGlyLeuArgSerAspGlyLysThrLeuIleAlaGlnProHis
481

      1510                1530                1550
         .          .          .          .          .          .
GTCCGTGGCGACAAGCTGTGGACGTTGGACGATATCAAACGCCTGAACTGCGTCTTCAAG
ValArgGlyAspLysLeuTrpThrLeuAspAspIleLysArgLeuAsnCysValPheLys
501

      1570
         .          .
AACCCGGTGAAGGCATTCAATTGA
AsnProValLysAlaPheAsnEnd
521
```

M. capsulatus (Bath) *mmoY* sequence

```
        10                  30                  50
         .         .         .         .         .         .
ATGAGCATGTTAGGAGAAAGACGCCGCGGTCTGACCGATCCGGAAATGGCGGCCGTCATT
MetSerMetLeuGlyGluArgArgArgGlyLeuThrAspProGluMetAlaAlaValIle
  1

        70                  90                 110
         .         .         .         .         .         .
TTGAAGGCGCTTCCTGAAGCTCCGCTGGACGGCAACAACAAGATGGGTTATTTCGTCACC
LeuLysAlaLeuProGluAlaProLeuAspGlyAsnAsnLysMetGlyTyrPheValThr
 21

       130                 150                 170
         .         .         .         .         .         .
CCCCGCTGGAAACGCTTGACGGAATATGAAGCCCTGACCGTTTATGCGCAGCCCAACGCC
ProArgTrpLysArgLeuThrGluTyrGluAlaLeuThrValTyrAlaGlnProAsnAla
 41

       190                 210                 230
         .         .         .         .         .         .
GACTGGATCGCCGGCGGCCTGGACTGGGGCGACTGGACCCAGAAATTCCACGGCGGCCGC
AspTrpIleAlaGlyGlyLeuAspTrpGlyAspTrpThrGlnLysPheHisGlyGlyArg
 61

       250                 270                 290
         .         .         .         .         .         .
CCTTCCTGGGGCAACGAGACCACGGAGCTGCGCACCGTCGACTGGTTCAAGCACCGTGAC
ProSerTrpGlyAsnGluThrThrGluLeuArgThrValAspTrpPheLysHisArgAsp
 81

       310                 330                 350
         .         .         .         .         .         .
CCGCTCCGCCGTTGGCATGCGCCGTACGTCAAGGACAAGGCCGAGGAATGGCGCTACACC
ProLeuArgArgTrpHisAlaProTyrValLysAspLysAlaGluGluTrpArgTyrThr
101

       370                 390                 410
         .         .         .         .         .         .
GACCGCTTCCTGCAGGGTTACTCCGCCGACGGTCAGATCCGGGCGATGAACCCGACCTGG
AspArgPheLeuGlnGlyTyrSerAlaAspGlyGlnIleArgAlaMetAsnProThrTrp
121

       430                 450                 470
         .         .         .         .         .         .
CGGGACGAGTTCATCAACCGGTATTGGGGCGCCTTCCTGTTCAACGAATACGGATTGTTC
ArgAspGluPheIleAsnArgTyrTrpGlyAlaPheLeuPheAsnGluTyrGlyLeuPhe
141
```

```
       490                 510                 530
         .         .         .         .         .         .
AACGCTCATTCGCAGGGCGCCCGGGAGGCGCTGTCGGACGTAACCCGCGTCAGCCTGGCT
AsnAlaHisSerGlnGlyAlaArgGluAlaLeuSerAspValThrArgValSerLeuAla
161
       550                 570                 590
         .         .         .         .         .         .
TTCTGGGGCTTCGACAAGATCGACATCGCCCAGATGATCCAACTCGAACGGGGTTTCCTC
PheTrpGlyPheAspLysIleAspIleAlaGlnMetIleGlnLeuGluArgGlyPheLeu
181

       610                 630                 650
         .         .         .         .         .         .
GCCAAGATCGTACCCGGTTTCGACGAGTCCACAGCGGTGCCGAAGGCCGAATGGACGAAC
AlaLysIleValProGlyPheAspGluSerThrAlaValProLysAlaGluTrpThrAsn
201

       670                 690                 710
         .         .         .         .         .         .
GGGGAGGTCTACAAGAGCGCCCGTCTGGCCGTGGAAGGGCTGTGGCAGGAGGTGTTCGAC
GlyGluValTyrLysSerAlaArgLeuAlaValGluGlyLeuTrpGlnGluValPheAsp
221

       730                 750                 770
         .         .         .         .         .         .
TGGAACGAGAGCGCTTTCTCGGTGCACGCCGTCTATGACGCGCTGTTCGGTCAGTTCGTC
TrpAsnGluSerAlaPheSerValHisAlaValTyrAspAlaLeuPheGlyGlnPheVal
241

       790                 810                 830
         .         .         .         .         .         .
CGCCGCGAGTTCTTTCAGCGGCTGGCTCCCCGCTTCGGCGACAATCTGACGCCATTCTTC
ArgArgGluPhePheGlnArgLeuAlaProArgPheGlyAspAsnLeuThrProPhePhe
261

       850                 870                 890
         .         .         .         .         .         .
ATCAACCAGGCCCAGACATACTTCCAGATCGCCAAGCAGGGCGTACAGGATCTGTATTAC
IleAsnGlnAlaGlnThrTyrPheGlnIleAlaLysGlnGlyValGlnAspLeuTyrTyr
281

       910                 930                 950
         .         .         .         .         .         .
AACTGTCTGGGTGACGATCCGGAGTTCAGCGATTACAACCGTACCGTGATGCGCAACTGG
AsnCysLeuGlyAspAspProGluPheSerAspTyrAsnArgThrValMetArgAsnTrp
301
```

```
       970                 990                1010
         .         .         .         .         .         .
ACCGGCAAGTGGCTGGAGCCCACGATCGCCGCTCTGCGCGACTTCATGGGGCTGTTTGCG
ThrGlyLysTrpLeuGluProThrIleAlaAlaLeuArgAspPheMetGlyLeuPheAla
321

      1030                1050                1070
         .         .         .         .         .         .
AAGCTGCCGGCGGGCACCACTGACAAGGAAGAAATCACCGCGTCCCTGTACCGGGTGGTC
LysLeuProAlaGlyThrThrAspLysGluGluIleThrAlaSerLeuTyrArgValVal
341

      1090                1110                1130
         .         .         .         .         .         .
GACGACTGGATCGAGGACTACGCCAGCAGGATCGACTTCAAGGCGGACCGCGATCAGATC
AspAspTrpIleGluAspTyrAlaSerArgIleAspPheLysAlaAspArgAspGlnIle
361

      1150                1170
         .         .         .
GTTAAAGCGGTTCTGGCAGGATTGAAATAA
ValLysAlaValLeuAlaGlyLeuLysEnd
381
```

M. capsulatus (Bath) *mmoB* sequence

```
        10                  30                  50
         .         .         .         .         .         .
ATGAGCGTAAACAGCAACGCATACGACGCCGGCATCATGGGCCTGAAAGGCAAGGACTTC
MetSerValAsnSerAsnAlaTyrAspAlaGlyIleMetGlyLeuLysGlyLysAspPhe
  1

        70                  90                  110
         .         .         .         .         .         .
GCCGATCAGTTCTTTGCCGACGAAAACCAAGTGGTCCATGAAAGCGACACGGTCGTTCTG
AlaAspGlnPhePheAlaAspGluAsnGlnValValHisGluSerAspThrValValLeu
 21

       130                 150                 170
         .         .         .         .         .         .
GTCCTCAAGAAGTCGGACGAGATCAATACCTTTATCGAGGAGATCCTTCTGACGGACTAC
ValLeuLysLysSerAspGluIleAsnThrPheIleGluGluIleLeuLeuThrAspTyr
 41

       190                 210                 230
         .         .         .         .         .         .
AAGAAGAACGTCAATCCGACGGTAAACGTGGAAGACCGCGCGGGTTACTGGTGGATCAAG
LysLysAsnValAsnProThrValAsnValGluAspArgAlaGlyTyrTrpTrpIleLys
 61

       250                 270                 290
         .         .         .         .         .         .
GCCAACGGCAAGATCGAGGTCGATTGCGACGAGATTTCCGAGCTGTTGGGGCGGCAGTTC
AlaAsnGlyLysIleGluValAspCysAspGluIleSerGluLeuLeuGlyArgGlnPhe
 81

       310                 330                 350
         .         .         .         .         .         .
AACGTCTACGACTTCCTCGTCGACGTTTCCTCCACCATCGGCCGGGCCTATACCCTGGGC
AsnValTyrAspPheLeuValAspValSerSerThrIleGlyArgAlaTyrThrLeuGly
 101

       370                 390                 410
         .         .         .         .         .         .
AACAAGTTCACCATTACCAGTGAGCTGATGGGCCTGGACCGCAAGCTCGAAGACTATCAC
AsnLysPheThrIleThrSerGluLeuMetGlyLeuAspArgLysLeuGluAspTyrHis
121


GCTTAA
AlaEnd
141
```

M. capsulatus (Bath) *mmoZ* sequence

```
        10                  30                  50
         .         .         .         .         .         .
ATGGCGAAACTGGGTATACACAGCAACGACACCCGCGACGCCTGGGTGAACAAGATCGCG
MetAlaLysLeuGlyIleHisSerAsnAspThrArgAspAlaTrpValAsnLysIleAla
  1

        70                  90                  110
         .         .         .         .         .         .
CAGCTCAACACCCTGGAAAAAGCGGCCGAGATGCTGAAGCAGTTCCGGATGGACCACACC
GlnLeuAsnThrLeuGluLysAlaAlaGluMetLeuLysGlnPheArgMetAspHisThr
 21

       130                 150                 170
         .         .         .         .         .         .
ACGCCGTTCCGCAACAGCTACGAACTGGACAACGACTACCTCTGGATCGAGGCCAAGCTC
ThrProPheArgAsnSerTyrGluLeuAspAsnAspTyrLeuTrpIleGluAlaLysLeu
 41

       190                 210                 230
         .         .         .         .         .         .
GAAGAGAAGGTCGCCGTCCTCAAGGCACGCGCCTTCAACGAGGTGGACTTCCGTCATAAG
GluGluLysValAlaValLeuLysAlaArgAlaPheAsnGluValAspPheArgHisLys
 61

       250                 270                 290
         .         .         .         .         .         .
ACCGCTTTCGGCGAGGATGCCAAGTCCGTTCTGGACGGCACCGTCGCGAAGATGAACGCG
ThrAlaPheGlyGluAspAlaLysSerValLeuAspGlyThrValAlaLysMetAsnAla
 81

       310                 330                 350
         .         .         .         .         .         .
GCCAAGGACAAGTGGGAGGCGGAGAAGATCCATATCGGTTTCCGCCAGGCCTACAAGCCG
AlaLysAspLysTrpGluAlaGluLysIleHisIleGlyPheArgGlnAlaTyrLysPro
101

       370                 390                 410
         .         .         .         .         .         .
CCGATCATGCCGGTGAACTATTTCCTGGACGGCGAGCGTCAGTTGGGGACCCGGCTGATG
ProIleMetProValAsnTyrPheLeuAspGlyGluArgGlnLeuGlyThrArgLeuMet
121

       430                 450                 470
         .         .         .         .         .         .
GAACTGCGCAACCTCAACTACTACGACACGCCGCTGGAAGAACTGCGCAAACAGCGCGGT
GluLeuArgAsnLeuAsnTyrTyrAspThrProLeuGluGluLeuArgLysGlnArgGly
141
```

```
       490                 510
         .         .         .
GTGCGGGTGGTGCATCTGCAGTCGCCGCACTGA
ValArgValValHisLeuGlnSerProHisEnd
161
```

M. capsulatus (Bath) *orfY* sequence

```
        10                  30                  50
         .         .         .         .         .         .
ATGGTCGAATCGGCATTTCAGCCATTTTCGGGCGACGCAGACGAATGGTTCGAGGAACCA
MetValGluSerAlaPheGlnProPheSerGlyAspAlaAspGluTrpPheGluGluPro
  1

        70                  90                  110
         .         .         .         .         .         .
CGGCCCCAGGCCGGTTTCTTCCCTTCCGCGGACTGGCATCTGCTCAAACGGGACGAGACC
ArgProGlnAlaGlyPhePheProSerAlaAspTrpHisLeuLeuLysArgAspGluThr
 21

       130                 150                 170
         .         .         .         .         .         .
TACGCAGCCTATGCCAAGGATCTCGATTTCATGTGGCGGTGGGTCATCGTCCGGGAAGAA
TyrAlaAlaTyrAlaLysAspLeuAspPheMetTrpArgTrpValIleValArgGluGlu
 41

       190                 210                 230
         .         .         .         .         .         .
AGGATCGTCCAGGAGGGTTGCTCGATCAGCCTGGAGTCGTCGATCCGCGCCGTGACGCAC
ArgIleValGlnGluGlyCysSerIleSerLeuGluSerSerIleArgAlaValThrHis
 61

       250                 270                 290
         .         .         .         .         .         .
GTACTGAATTATTTTGGTATGACCGAACAACGCGCCCCGGCAGAGGACCGGACCGGCGGA
ValLeuAsnTyrPheGlyMetThrGluGlnArgAlaProAlaGluAspArgThrGlyGly
 81

       310
         .
GTTCAACATTGA
ValGlnHisEnd
101
```

M. capsulatus (Bath) *mmoC* sequence

```
        10                  30                  50
         .         .         .         .         .         .
ATGCAGCGAGTTCACACTATCACGGCGGTGACGGAGGATGGCGAATCGCTCCGCTTCGAA
MetGlnArgValHisThrIleThrAlaValThrGluAspGlyGluSerLeuArgPheGlu
  1

        70                  90                 110
         .         .         .         .         .         .
TGCCGTTCGGACGAGGACGTCATCACCGCCGCCCTGCGCCAGAACATCTTTCTGATGTCG
CysArgSerAspGluAspValIleThrAlaAlaLeuArgGlnAsnIlePheLeuMetSer
 21

       130                 150                 170
         .         .         .         .         .         .
TCCTGCCGGGAGGGCGGCTGTGCGACCTGCAAGGCCTTGTGCAGCGAAGGGGACTACGAC
SerCysArgGluGlyGlyCysAlaThrCysLysAlaLeuCysSerGluGlyAspTyrAsp
 41

       190                 210                 230
         .         .         .         .         .         .
CTCAAGGGCTGCAGCGTTCAGGCGCTGCCGCCGGAAGAGGAGGAGGAAGGGTTGGTGTTG
LeuLysGlyCysSerValGlnAlaLeuProProGluGluGluGluGlyLeuValLeu
 61

       250                 270                 290
         .         .         .         .         .         .
TTGTGCCGGACCTACCCGAAGACCGACCTGGAAATCGAACTGCCCTATACCCATTGCCGC
LeuCysArgThrTyrProLysThrAspLeuGluIleGluLeuProTyrThrHisCysArg
 81

       310                 330                 350
         .         .         .         .         .         .
ATCAGTTTTGGTGAGGTCGGCAGTTTCGAGGCGGAGGTCGTCGGCCTCAACTGGGTTTCG
IleSerPheGlyGluValGlySerPheGluAlaGluValValGlyLeuAsnTrpValSer
101

       370                 390                 410
         .         .         .         .         .         .
AGCAACACCGTCCAGTTTCTTTTGCAGAAGCGGCCCGACGAGTGCGGCAACCGTGGCGTG
SerAsnThrValGlnPheLeuLeuGlnLysArgProAspGluCysGlyAsnArgGlyVal
121

       430                 450                 470
         .         .         .         .         .         .
AAATTCGAACCCGGTCAGTTCATGGACCTGACCATCCCCGGCACCGATGTCTCCCGCTCC
LysPheGluProGlyGlnPheMetAspLeuThrIleProGlyThrAspValSerArgSer
141
```

```
       490                 510                 530
         .         .         .         .         .         .
TACTCGCCGGCGAACCTTCCTAATCCCGAAGGCCGCCTGGAGTTCCTGATCCGCGTGTTA
TyrSerProAlaAsnLeuProAsnProGluGlyArgLeuGluPheLeuIleArgValLeu
161

       550                 570                 590
         .         .         .         .         .         .
CCGGAGGGACGGTTTTCGGACTACCTGCGCAATGACGCGCGTGTCGGACAGGTCCTCTCG
ProGluGlyArgPheSerAspTyrLeuArgAsnAspAlaArgValGlyGlnValLeuSer
181

       610                 630                 650
         .         .         .         .         .         .
GTCAAAGGGCCACTGGGCGTGTTCGGTCTCAAGGAGCGGGGCATGGCGCCGCGCTATTTC
ValLysGlyProLeuGlyValPheGlyLeuLysGluArgGlyMetAlaProArgTyrPhe
201

       670                 690                 710
         .         .         .         .         .         .
GTGGCCGGCGGCACCGGGTTGGCGCCGGTGGTCTCGATGGTGCGGCAGATGCAGGAGTGG
ValAlaGlyGlyThrGlyLeuAlaProValValSerMetValArgGlnMetGlnGluTrp
221

       730                 750                 770
         .         .         .         .         .         .
ACCGCGCCGAACGAGACCCGCATCTATTTCGGTGTGAACACCGAGCCGGAATTGTTCTAC
ThrAlaProAsnGluThrArgIleTyrPheGlyValAsnThrGluProGluLeuPheTyr
241

       790                 810                 830
         .         .         .         .         .         .
ATCGACGAGCTCAAATCCCTGGAACGATCGATGCGCAATCTCACCGTGAAGGCCTGTGTC
IleAspGluLeuLysSerLeuGluArgSerMetArgAsnLeuThrValLysAlaCysVal
261

       850                 870                 890
         .         .         .         .         .         .
TGGCACCCGAGCGGGGACTGGGAAGGCGAGCAGGGCTCGCCCATCGATGCGTTGCGGGAA
TrpHisProSerGlyAspTrpGluGlyGluGlnGlySerProIleAspAlaLeuArgGlu
281

       910                 930                 950
         .         .         .         .         .         .
GACCTGGAGTCCTCCGACGCCAACCCGGACATTTATTTGTGCGGTCCGCCGGGCATGATC
AspLeuGluSerSerAspAlaAsnProAspIleTyrLeuCysGlyProProGlyMetIle
301
```

```
       970                 990                1010
         .         .         .         .         .         .
GATGCCGCCTGCGAGCTGGTACGCAGCCGCGGTATCCCCGGCGAACAGGTCTTCTTCGAA
AspAlaAlaCysGluLeuValArgSerArgGlyIleProGlyGluGlnValPhePheGlu
321

      1030
         .         .
AAATTCCTGCCGTCCGGGGCGGCCTGA
LysPheLeuProSerGlyAlaAlaEnd
341
```

Figure 1-6. Mistakes in the published hydroxylase subunit gene sequences and their corrections. Changes are shown in bold face. All mistakes were confirmed by DNA sequencing.

Figure 1-7. Mistakes found in the published DNA sequence for the sMMO genes *mmoB* and *mmoC*. The mistake in the *mmoB* gene was discovered by N-terminal sequencing of the protein B peptide (*49*) and confirmed via DNA sequencing (*52*). The mistake in the *mmoC* gene was discovered by DNA sequencing (*52*). Changes are shown in bold face.

Figure 1-8. Relationships between the genes and proteins of the sMMO multicomponent monooxygenase family. The T3M and T4M systems have ferredoxins encoded in their operons, where the others do not. The gray background behind the *orfY* homologues for T2M and PH indicate that sequence alignments and homology analyses do not indicate that these proteins are related to *orfY*. They are included in the *orfY* category because they are all small open reading frames that have no similarities with any other genes in the GenEMBL database.

Figure 1-9. Identity statistics between the closely related proteins of the sMMO multicomponent monooxygenase family. These include the α and β subunits of the hydroxylase, protein B and the reductase. Three numbers are given in each comparison; the top is the number of identical residues, the middle is the number of differing residues, and the bottom number is the number of gap placeholders. Numbers right of the diagonal are percentages and left of the diagonal are raw numbers. Percentages do not add up to 100% due to mismatches at the ends of the peptides. Note that two randomly paired proteins would be expected to have identity percentages of 10-15%, depending upon the size of the protein.

Figure 1-10. Sequence alignments of the peptides corresponding to the α subunit of the hydroxylase. Residues conserved across all peptides are shown in reverse. The top number in each bar is the consensus numbering, the bottom number is the *M. capsulatus* (Bath) residue number. Residues belonging to helices in domain 1 of the hydroxylase are marked.

Figure 1-11. Sequence alignments of the peptides corresponding to the β subunit of the hydroxylase. Residues conserved across all peptides are shown in reverse. The top number in each bar is the consensus numbering, the bottom number is the *M. capsulatus* (Bath) residue number.

Figure 1-12. Sequence alignments of the peptides corresponding to protein B. Residues conserved across all peptides are shown in reverse. Residues with conservative substitutions are shown with gray background. The top number in each bar is the consensus numbering, the bottom numbers are the *M. capsulatus* (Bath) sMMOB residue number and *P. pudita* P2 residue number.

Figure 1-13. Sequence alignments of the peptides corresponding to the reductase. Residues conserved across all peptides are shown in reverse. The top number in each bar is the consensus numbering, the bottom number is the *M. capsulatus* (Bath) residue number. Domains corresponding to [2Fe-2S], FAD and NADH binding regions are marked (*24*).

Figure 1-14. View of the absolutely conserved α subunit residues near the active site of the *M. capsulatus* (Bath) sMMO hydroxylase. Carbons are marked gray, oxygens are red, and nitrogens are blue. The four helices that surround the active site are shown as partially transparent ribbons. The irons are shown in pink, and iron 1 is marked as such. The hydroxide bridge, terminal water and acetate bridge are all omitted for clarity. This Figure was generated by using the program MolScript (*47*).

Figure 1-15. Schematic of the hydrogen bonding network in the absolutely conserved residues near the active site. Distances as shown in the Figure do not necessarily correspond to actual distances.

Figure 1-16. View of the absolutely conserved α subunit residues not near the active site of the *M. capsulatus* (Bath) sMMO hydroxylase. All peptides but the rightmost α subunit are partially transparent for clarity. Residues lining the canyon are shown in yellow, the aromatic cluster residues on top of the hydroxylase are in red, and the cluster of residues under the γ subunit are shown in white. This Figure was generated by using the program MolScript (*47*).

Figure 1-17. View of the absolutely conserved β subunit residues in the *M. capsulatus* (Bath) sMMO hydroxylase. All peptides but the rightmost β subunit are partially transparent for clarity. Conserved residues are shown in white. This Figure was generated by using the program MolScript (*47*).

Figure 1-18. View of the absolutely conserved residues in the *P. pudita* CF600 phenol hydroxylase P2 protein (model 1). Conserved residues are shown in white. Conservative substitutions are shown in yellow. This Figure was generated by using the program MolScript (*47*).

Figure 1-19. Cartoon of two possible sMMO family interprotein interactions.