

ALKANE OXIDATION BY MODIFIED HYDROXYLASES**CROSS REFERENCE TO RELATED APPLICATIONS**

[0001] This application claims priority to U.S. Provisional Application Serial No. 60/665,903 filed March 28, 2005, and to U.S. Provisional Application Serial No. 60/698,872 filed July 13, 2005, and also to U.S. Provisional Application Serial No. 60/700,781 filed July 20, 2005, the disclosures of which are incorporated herein by reference.

STATEMENTS REGARDING FEDERALLY SPONSORED RESEARCH

[0002] The invention was funded in part by Grant No. BES-0313567 awarded by National Science Foundation (NSF). The government may have certain rights in the invention.

TECHNICAL FIELD

[0003] This invention relates to modified hydroxylases. The invention further relates to cells expressing such modified hydroxylases and methods of producing hydroxylated alkanes by contacting a suitable substrate with such cells. Also included are modified hydroxylases comprising unique regioselectivities.

BACKGROUND

[0004] Hydroxylation of linear alkanes has the important practical implication of providing valuable intermediates for chemical synthesis. Nevertheless, selective oxyfunctionalization of hydrocarbons remains one of the great challenges for contemporary chemistry. Many chemical methods for hydroxylation require severe conditions of temperature or pressure, and the reactions are prone to over-oxidation, producing a range of products, many of which are not desired.

[0005] Enzymes are an attractive alternative to chemical catalysts. In particular, monooxygenases have the ability to catalyze the specific hydroxylation of non-activated C-H

bonds. These cofactor-dependent oxidative enzymes have multiple domains and function via complex electron transfer mechanisms to transport a reduction equivalent to the catalytic center. Exemplary monooxygenases include the cytochrome P450 monooxygenases ("P450s"). The P450s are a group of widely-distributed heme-containing enzymes that insert one oxygen atom from diatomic oxygen into a diverse range of hydrophobic substrates, often with high regio- and stereoselectivity. Their ability to catalyze these reactions with high specificity and selectivity makes P450s attractive catalysts for chemical synthesis and other applications, including oxidation chemistry.

[0006] Despite the ability of these enzymes to selectively hydroxylate a wide range of compounds, including fatty acids, aromatic compounds, alkanes, alkenes, and natural products, only a few members of this large superfamily of proteins are capable of hydroxylating alkanes. Accordingly, there is a need for modified hydroxylases that have the ability to efficiently hydroxylate alkanes in vivo. In addition, there is a need for cells that can express such modified hydroxylases while producing recoverable quantities of alkane-derived alcohols. Such hydroxylases and cells can be used to convert ethane to ethanol and/or methane to methanol.

SUMMARY

[0007] Provided herein are polypeptides that convert alkanes to alcohols. Also provided are nucleic acid molecules that encode such polypeptides, cells expressing such polypeptides, and methods of synthesizing alcohols from a suitable alkane substrate. Accordingly, in various embodiments, isolated or recombinant polypeptides comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:1, 2, 3, 4, 5 or 6, are provided. The polypeptides include up to 50, 25, 10, or 5 conservative amino

acid substitutions excluding residues: (a) 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, and 353 of SEQ ID NO:1; (b) 48, 79, 83, 95, 143, 176, 185, 206, 227, 238, 254, 257, 292, 330, and 355 of SEQ ID NO:2 or SEQ ID NO:3; (c) 49, 80, 84, 96, 144, 177, 186, 207, 228, 239, 255, 258, 293, 331, and 357 of SEQ ID NO:4; (d) 54, 85, 89, 101, 149, 182, 191, 212, 232, 243, 259, 262, 297, 336, and 361 of SEQ ID NO:5; and (e) 51, 82, 86, 98, 146, 179, 188, 208, 231, 242, 258, 262, 296, 337, and 363 of SEQ ID NO:6. The amino acid sequence includes the following residues:

(1) a Z1 amino acid residue at positions: (a) 47, 82, 142, 205, 236, 252, and 255 of SEQ ID NO:1; (b) 48, 83, 143, 206, 238, 254, and 257 of SEQ ID NO:2 or SEQ ID NO:3; (c) 49, 84, 144, 207, 239, 255, and 258 of SEQ ID NO:4; (d) 54, 89, 149, 212, 243, 259, and 262 of SEQ ID NO:5; and (e) 51, 86, 146, 208, 242, 258, and 262 of SEQ ID NO:6;

(2) a Z2 amino acid residue at positions: (a) 94, 175, 184, 290, and 353 of SEQ ID NO:1; (b) 95, 176, 185, 292, and 355 of SEQ ID NO:2 or SEQ ID NO:3; (c) 96, 177, 186, 293, and 357 of SEQ ID NO:4; (d) 101, 182, 191, 297, and 361 of SEQ ID NO:5; and (e) 98, 179, 188, 296, and 363 of SEQ ID NO:6;

(3) a Z3 amino acid residue at position: (a) 226 of SEQ ID NO:1; (b) 227 of SEQ ID NO:2 or SEQ ID NO:3; (c) 228 of SEQ ID NO:4; (d) 232 of SEQ ID NO:5; and (e) 231 of SEQ ID NO:6; and

(4) a Z4 amino acid residue at positions: (a) 78 and 328 of SEQ ID NO:1; (b) 79 and 330 of SEQ ID NO:2 or SEQ ID NO:3; (c) 80 and 331 of SEQ ID NO:4; (d) 85 and 336 of SEQ ID NO:5; and (e) 82 and 337 of SEQ ID NO:6.

[0008] In general, a Z1 amino acid residue includes glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), or cysteine (C). A Z2 amino acid residue includes alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), or methionine (M). A Z3 amino acid residue

includes lysine (K), or arginine (R). A Z4 amino acid residue includes tyrosine (Y), phenylalanine (F), tryptophan (W), or histidine (H).

[0009] In general, polypeptides provided herein display hydroxylase activity that converts an alkane to an alcohol.

[0010] In other embodiments, the polypeptide further includes a Z3 amino acid residue at position: (a) 285 of SEQ ID NO:1; (b) 287 of SEQ ID NO:2 or 3; (c) 288 of SEQ ID NO:4; (d) 292 of SEQ ID NO:5; and (e) 291 of SEQ ID NO:6. A Z3 amino acid residue includes lysine (K), arginine (R), or histidine (H). In some aspects, the amino acid residue at this position is an arginine (R).

[0011] In general, alkanes include methane (CH₄), ethane (C₂H₆), propane (C₃H₈), butane (C₄H₁₀), pentane (C₅H₁₂), hexane (C₆H₁₄), heptane (C₇H₁₆), octane (C₈H₁₈), nonane (C₉H₂₀), decane (C₁₀H₂₂), undecane (C₁₁H₂₄), and dodecane (C₁₂H₂₆). Also in general, alcohols include methanol, ethanol, propanol, butanol, pentanol, hexanol, heptanol, octanol, nonanol, decanol, undecanol, and dodecanol.

[0012] In other embodiments, the amino acid sequence of the polypeptide includes residues at the following positions:

(1) a glycine (G), glutamine (Q), serine (S), threonine (T), or cysteine (C) amino acid residue at position: (a) 47, 82, 142, 205, 236, 252, and 255 of SEQ ID NO:1; (b) 48, 83, 143, 206, 238, 254, and 257 of SEQ ID NO:2 or SEQ ID NO:3; (c) 49, 84, 144, 207, 239, 255, and 258 of SEQ ID NO:4; (d) 54, 89, 149, 212, 243, 259, and 262 of SEQ ID NO:5; and (e) 51, 86, 146, 208, 242, 258, and 262 of SEQ ID NO:6;

(2) a valine (V) or isoleucine (I) amino acid residue at position: (a) 94, 175, 184, 290, and 353 of SEQ ID NO:1; (b) 95, 176, 185, 292, and 355 of SEQ ID NO:2 or SEQ ID NO:3; (c) 96, 177, 186, 293, and 357 of SEQ ID NO:4; (d) 101, 182, 191, 297, and 361 of SEQ ID NO:5; and (e) 98, 179, 188, 296, and 363 of SEQ ID NO:6;

(3) an arginine amino acid residue at position: (a) 226 of SEQ ID NO:1; (b) 227 of SEQ ID NO:2 or SEQ ID NO:3; (c) 228 of SEQ ID NO:4; (d) 232 of SEQ ID NO:5; and (e) 231 of SEQ ID NO:6; and

(4) a phenylalanine (F) or histidine (H) amino acid residue at position: (a) 78 and 328 of SEQ ID NO:1; (b) 79 and 330 of SEQ ID NO:2 or SEQ ID NO:3; (c) 80 and 331 of SEQ ID NO:4; (d) 85 and 336 of SEQ ID NO:5; and (e) 82 and 337 of SEQ ID NO:6.

[0013] In other embodiments, the amino acid sequence of the polypeptide includes residues at the following positions:

(1) a serine (S) residue at position: (a) 82, 142, and 255 of SEQ ID NO:1; (b) 83, 143 and 257 of SEQ ID NO:2 or SEQ ID NO:3; (c) 84, 144, and 258 of SEQ ID NO:4; (d) 89, 149, and 262 of SEQ ID NO:5; (e) 86, 146 and 262 of SEQ ID NO:6;

(2) a cysteine (C) amino acid residue at position: (a) 47 and 205 of SEQ ID NO:1; (b) 48 and 206 of SEQ ID NO:2 or SEQ ID NO:3; (c) 49 and 207 of SEQ ID NO:4; (d) 54 and 212 of SEQ ID NO:5; (e) 51 and 208 of SEQ ID NO:6;

(3) a glutamine (Q) amino acid residue at position: (a) 236 of SEQ ID NO:1; (b) 238 of SEQ ID NO:2 or SEQ ID NO:3; (c) 239 of SEQ ID NO:4; (d) 243 of SEQ ID NO:5; (e) 242 of SEQ ID NO:6;

(4) a glycine (G) amino acid residue at position: (a) 252 of SEQ ID NO:1; (b) 254 of SEQ ID NO:2 or SEQ ID NO:3; (c) 255 of SEQ ID NO:4; (d) 259 of SEQ ID NO:5; (e) 258 of SEQ ID NO:6;

(5) a valine (V) amino acid residue at position: (a) 184, 290 and 353 of SEQ ID NO:1; (b) 185, 292, and 355 of SEQ ID NO:2 or SEQ ID NO:3; (c) 186, 293 and 357 of SEQ ID NO:4; (d) 191, 297, and 361 of SEQ ID NO:5; (e) 188, 296, and 363 of SEQ ID NO:6;

(6) an isoleucine (I) amino acid residue at position: (a) 94 and 175 of SEQ ID NO:1; (b) 95 and 176 of SEQ ID NO:2 or

SEQ ID NO:3; (c) 96 and 177 of SEQ ID NO:4; (d) 101 and 182 of SEQ ID NO:5; (e) 98 and 179 of SEQ ID NO:6; and

(7) a phenylalanine (F) amino acid residue at position: (a) 78 and 328 of SEQ ID NO:1; (b) 79 and 330 of SEQ ID NO:2 or SEQ ID NO:3; (c) 80 and 331 of SEQ ID NO:4; (d) 85 and 336 of SEQ ID NO:5; (e) 82 and 337 of SEQ ID NO:6.

[0014] In another embodiment, an isolated or recombinant polypeptide that includes residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:1, 2, 3, 4, 5 or 6 is provided. The polypeptide includes amino acid residues in the amino acid sequence that correspond to the following positions, at least 80%, 85%, 90%, or 95% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as follows:

(1) a serine (S) residue at position: (a) 82, 142, and 255 of SEQ ID NO:1; (b) 83, 143 and 257 of SEQ ID NO:2 or SEQ ID NO:3; (c) 84, 144, and 258 of SEQ ID NO:4; (d) 89, 149, and 262 of SEQ ID NO:5; (e) 86, 146 and 262 of SEQ ID NO:6;

(2) a cysteine (C) amino acid residue at position: (a) 47 and 205 of SEQ ID NO:1; (b) 48 and 206 of SEQ ID NO:2 or SEQ ID NO:3; (c) 49 and 207 of SEQ ID NO:4; (d) 54 and 212 of SEQ ID NO:5; (e) 51 and 208 of SEQ ID NO:6;

(3) a glutamine (Q) amino acid residue at position: (a) 236 of SEQ ID NO:1; (b) 238 of SEQ ID NO:2 or SEQ ID NO:3; (c) 239 of SEQ ID NO:4; (d) 243 of SEQ ID NO:5; (e) 242 of SEQ ID NO:6;

(4) a glycine (G) amino acid residue at position: (a) 252 of SEQ ID NO:1; (b) 254 of SEQ ID NO:2 or SEQ ID NO:3; (c) 255 of SEQ ID NO:4; (d) 259 of SEQ ID NO:5; (e) 258 of SEQ ID NO:6;

(5) a valine (V) amino acid residue at position: (a) 184, 290 and 353 of SEQ ID NO:1; (b) 185, 292, and 355 of SEQ ID NO:2 or SEQ ID NO:3; (c) 186, 293 and 357 of SEQ ID NO:4; (d)

191, 297, and 361 of SEQ ID NO:5; (e) 188, 296, and 363 of SEQ ID NO:6;

(6) an isoleucine (I) amino acid residue at position: (a) 94 and 175 of SEQ ID NO:1; (b) 95 and 176 of SEQ ID NO:2 or SEQ ID NO:3; (c) 96 and 177 of SEQ ID NO:4; (d) 101 and 182 of SEQ ID NO:5; (e) 98 and 179 of SEQ ID NO:6; and

(7) a phenylalanine (F) amino acid residue at position: (a) 78 and 328 of SEQ ID NO:1; (b) 79 and 330 of SEQ ID NO:2 or SEQ ID NO:3; (c) 80 and 331 of SEQ ID NO:4; (d) 85 and 336 of SEQ ID NO:5; (e) 82 and 337 of SEQ ID NO:6; and

(8) an arginine (R) amino acid residue at position: (a) 226 of SEQ ID NO:1; b) 227 of SEQ ID NO:2 or SEQ ID NO:3; (c) 228 of SEQ ID NO:4; (d) 232 of SEQ ID NO:5; (e) 231 of SEQ ID NO:6.

[0015] In other embodiments, the polypeptide further includes a Z3 amino acid residue at position: (a) 285 of SEQ ID NO:1; (b) 287 of SEQ ID NO:2 or 3; (c) 288 of SEQ ID NO:4; (d) 292 of SEQ ID NO:5; and (e) 291 of SEQ ID NO:6. A Z3 amino acid residue includes lysine (K), arginine (R), or histidine (H). In some aspects, the amino acid residue at this position is an arginine (R).

[0016] In yet another embodiment, an isolated or recombinant polypeptide that includes residues 456-1048 of SEQ ID NO:1, 456-1059 of SEQ ID NO:2, 456-1053 of SEQ ID NO:3, 456-1064 of SEQ ID NO:4, 456-1063 of SEQ ID NO:5, or 456-1077 of SEQ ID NO:6, is provided. The polypeptide includes an amino acid sequence with up to 65, 40, 25, or 10 conservative amino acid substitutions excluding residues: (a) 464, 631, 645, 710 and 968 of SEQ ID NO:1; (b) 475, 641, 656, 721 and 980 of SEQ ID NO:2; (c) 467, 634, 648, 713 and 972 of SEQ ID NO:3; (d) 477, 644, 659, 724 and 983 of SEQ ID NO:4; (e) 472, 640, 656, 723 and 985 of SEQ ID NO:5; and (f) 480, 648, 664, 733 and 997 of SEQ ID NO:6. The amino acid sequence includes the following residues:

(1) a Z1, Z3, Z4, or Z5 amino acid residue at position: (a) 464 of SEQ ID NO:1; (b) 475 of SEQ ID NO:2; (c) 467 of SEQ ID NO:3; (d) 477 of SEQ ID NO:4; (e) 472 of SEQ ID NO:5; and (f) 480 of SEQ ID NO:6;

(2) a Z1 amino acid residue at position: (a) 631 and 710 of SEQ ID NO:1; (b) 641 and 721 of SEQ ID NO:2; (c) 634 and 713 of SEQ ID NO:3; (d) 644 and 724 of SEQ ID NO:4; (e) 640 and 723 of SEQ ID NO:5; and (f) 648 and 733 of SEQ ID NO:6;

(3) a Z3 amino acid residue at position: (a) 645 of SEQ ID NO:1; (b) 656 of SEQ ID NO:2; (c) 648 of SEQ ID NO:3; (d) 659 of SEQ ID NO:4; (e) 656 of SEQ ID NO:5; and (f) 664 of SEQ ID NO:6; and

(4) a Z2 amino acid residue at position: (a) 968 of SEQ ID NO:1; (b) 980 of SEQ ID NO:2; (c) 972 of SEQ ID NO:3; (d) 983 of SEQ ID NO:4; (e) 985 of SEQ ID NO:5; and (f) 997 of SEQ ID NO:6. Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine (C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and arginine (R); Z4 is an amino acid residue selected from the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H); and is an amino acid residue selected from the group consisting of threonine (T), valine (V), and isoleucine (I).

[0017] In other embodiments, the amino acid sequence of the polypeptide includes residues at the following positions:

(1) a glycine (G), arginine (R), tyrosine (Y), or threonine (T) amino acid residue at position: (a) 464 of SEQ ID NO:1; (b) 475 of SEQ ID NO:2; (c) 467 of SEQ ID NO:3; (d) 477 of SEQ ID NO:4; (e) 472 of SEQ ID NO:5; and (f) 480 of SEQ ID NO:6;

(2) an asparagine (N) amino acid residue at position: (a) 631 of SEQ ID NO:1; (b) 641 of SEQ ID NO:2; (c) 634 of SEQ ID NO:3; (d) 644 of SEQ ID NO:4; (e) 640 of SEQ ID NO:5; and (f) 648 of SEQ ID NO:6;

(3) an arginine (R) amino acid residue at position: (a) 645 of SEQ ID NO:1; (b) 656 of SEQ ID NO:2; (c) 648 of SEQ ID NO:3; (d) 659 of SEQ ID NO:4; (e) 656 of SEQ ID NO:5; and (f) 664 of SEQ ID NO:6;

(4) a threonine (T) amino acid residue at position: (a) 710 of SEQ ID NO:1; (b) 721 of SEQ ID NO:2; (c) 713 of SEQ ID NO:3; (d) 724 of SEQ ID NO:4; (e) 723 of SEQ ID NO:5; and (f) 733 of SEQ ID NO:6; and

(5) a lysine (L) amino acid residue at position: (a) 968 of SEQ ID NO:1; (b) 980 of SEQ ID NO:2; (c) 972 of SEQ ID NO:3; (d) 983 of SEQ ID NO:4; (e) 985 of SEQ ID NO:5; and (f) 997 of SEQ ID NO:6.

[0018] In another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in SEQ ID NO:7 with up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710, is provided. The polypeptide can display hydroxylase activity that converts an alkane to an alcohol.

[0019] In yet another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in SEQ ID NO:8 with up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710, is provided.

[0020] In another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in SEQ ID NO:9 with up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142,

175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710, is provided.

[0021] In another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in SEQ ID NO:10 with up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710, is provided.

[0022] The polypeptides of SEQ ID NO:7, 8, 9 or 10, further optionally include an arginine (R) at amino acid residue position 285.

[0023] As previously noted, polypeptides provided herein can display hydroxylase activity that converts an alkane to an alcohol. In general an alkane includes methane (CH_4), ethane (C_2H_6), propane (C_3H_8), butane (C_4H_{10}), pentane (C_5H_{12}), hexane (C_6H_{14}), heptane (C_7H_{16}), octane (C_8H_{18}), nonane (C_9H_{20}), decane ($\text{C}_{10}\text{H}_{22}$), undecane ($\text{C}_{11}\text{H}_{24}$), and dodecane ($\text{C}_{12}\text{H}_{26}$). In general, an alcohol includes methanol, ethanol, propanol, butanol, pentanol, hexanol, heptanol, octanol, nonanol, decanol, undecanol, and dodecanol.

[0024] In another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in SEQ ID NO:11 with up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 285, 290, 328, 353, 464, and 710, is provided.

[0025] In another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in SEQ ID NO:12 with up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 645, and 710, is provided.

[0026] In another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in

SEQ ID NO:13 with up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 631, and 710, is provided.

[0027] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:7; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:7; (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:7 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:7; or (d) a polypeptide comprising an amino acid sequence that can be optimally aligned with the sequence of SEQ ID NO:7 to generate a similarity score of at least 1830, using the BLOSUM62 matrix, a gap existence penalty of 11, and a gap extension penalty of 1, excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:7.

[0028] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:8; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:8; or (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:8 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:8.

[0029] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:9; (b) a polypeptide consisting of the amino acid sequence set

forth in SEQ ID NO:9; or (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:9 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:9.

[0030] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:10; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:10; or (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:10 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:10.

[0031] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:11; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:11; or (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:11 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 285, 290, 328, 353, 464, and 710 of SEQ ID NO:11.

[0032] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:12; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:12; or (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:12 excluding amino acid residues 47, 78, 82, 94, 142,

175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 645, and 710 of SEQ ID NO:12.

[0033] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:13; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:13; or (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:13 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 631, and 710 of SEQ ID NO:13.

[0034] In other embodiments, isolated nucleic acid molecules are provided. Such nucleic acid molecules include: (a) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:7, 8, 9, 10, 11, 12, or 13; (b) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:7, 8, 9, 10, 11, 12, or 13; (c) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:7 or 11; (d) a nucleic acid molecule which encodes a polypeptide comprising residues about 456 to about 1088 of the amino acid sequence set forth in SEQ ID NO:7, 8, 9, 10, 12, or 13; (e) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:15; (f) a nucleic acid molecule consisting of the nucleotide sequence set forth in SEQ ID NO:15; (g) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:17; (h) a nucleic acid molecule consisting of the nucleotide sequence set forth in SEQ ID NO:17.

[0035] In other embodiments, nucleic acid molecules of the invention include: (a) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino

acid sequence set forth in SEQ ID NO:2 or 3 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227 is arginine (R); 238 is glutamine (Q); and 254 is glycine (G);

(b) a nucleic acid molecule which encodes a polypeptide consisting of residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:2 or 3 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227 is arginine (R); 238 is glutamine (Q); and 254 is glycine (G);

(c) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:4 with the following amino acid residues: 49 and 207 are cysteine (C); 80 and 331 are phenylalanine (F); 84, 144, and 258 are serine (S); 96 and 177 are isoleucine (I); 186, 293, and 357 are valine (V); 228 is arginine (R); 239 is glutamine (Q); and 255 is glycine (G);

(d) a nucleic acid molecule which encodes a polypeptide consisting of residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:4 with the following amino acid residues: 49 and 207 are cysteine (C); 80 and 331 are phenylalanine (F); 84, 144, and 258 are serine (S); 96 and 177 are isoleucine (I); 186, 293, and 357 are valine (V); 228 is arginine (R); 239 is glutamine (Q); and 255 is glycine (G);

(e) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:5 with the following amino acid residues: 54 and 212 are cysteine (C); 85 and 336 are phenylalanine (F); 89, 149, and 262 are serine (S); 101 and 182 are isoleucine (I); 191, 297, and 361 are valine (V); 232 is arginine (R); 243 is glutamine (Q); and 259 is glycine (G);

(f) a nucleic acid molecule which encodes a polypeptide

consisting of residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:5 with the following amino acid residues: 54 and 212 are cysteine (C); 85 and 336 are phenylalanine (F); 89, 149, and 262 are serine (S); 101 and 182 are isoleucine (I); 191, 297, and 361 are valine (V); 232 is arginine (R); 243 is glutamine (Q); and 259 is glycine (G); (g) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:6 with the following amino acid residues: 51 and 208 are cysteine (C); 82 and 337 are phenylalanine (F); 86, 146, and 262 are serine (S); 98 and 179 are isoleucine (I); 188, 296, and 363 are valine (V); 231 is arginine (R); 243 is glutamine (Q); and 258 is glycine (G); and (h) a nucleic acid molecule which encodes a polypeptide consisting of residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:6 with the following amino acid residues: 51 and 208 are cysteine (C); 82 and 337 are phenylalanine (F); 86, 146, and 262 are serine (S); 98 and 179 are isoleucine (I); 188, 296, and 363 are valine (V); 231 is arginine (R); 243 is glutamine (Q); and 258 is glycine (G).

[0036] In other embodiments, nucleic acid molecules of the invention include: (a) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227, 287, and 656 are arginine (R); 238 is glutamine (Q); 254 is glycine (G); 475 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 641 is asparagine (N); 721 is threonine (T); and 980 is leucine; (b) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:2 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine

(S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227, 287, and 656 are arginine (R); 238 is glutamine (Q); 254 is glycine (G); 475 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 641 is asparagine (N); 721 is threonine (T); and 980 is leucine; (c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:3 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227, 287, and 648 are arginine (R); 238 is glutamine (Q); 254 is glycine (G); 467 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 634 is asparagine (N); 713 is threonine (T); and 972 is leucine; (d) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:3 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227, 287, and 648 are arginine (R); 238 is glutamine (Q); 254 is glycine (G); 467 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 634 is asparagine (N); 713 is threonine (T); and 972 is leucine; (e) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:4 with the following amino acid residues: 49 and 207 are cysteine (C); 80 and 331 are phenylalanine (F); 84, 144, and 258 are serine (S); 96 and 177 are isoleucine (I); 186, 293, and 357 are valine (V); 228, 288, and 659 are arginine (R); 239 is glutamine (Q); and 255 is glycine (G); 477 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 644 is asparagine (N); 724 is threonine (T); and 983 is leucine; (f) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:4 with the

following amino acid residues: 49 and 207 are cysteine (C); 80 and 331 are phenylalanine (F); 84, 144, and 258 are serine (S); 96 and 177 are isoleucine (I); 186, 293, and 357 are valine (V); 228, 288, and 659 are arginine (R); 239 is glutamine (Q); and 255 is glycine (G); 477 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 644 is asparagine (N); 724 is threonine (T); and 983 is leucine; (g) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:5 with the following amino acid residues: 54 and 212 are cysteine (C); 85 and 336 are phenylalanine (F); 89, 149, and 262 are serine (S); 101 and 182 are isoleucine (I); 191, 297, and 361 are valine (V); 232, 292, and 656 are arginine (R); 243 is glutamine (Q); and 259 is glycine (G); 472 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 640 is asparagine (N); 723 is threonine (T); and 985 is leucine; (h) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:5 with the following amino acid residues: 54 and 212 are cysteine (C); 85 and 336 are phenylalanine (F); 89, 149, and 262 are serine (S); 101 and 182 are isoleucine (I); 191, 297, and 361 are valine (V); 232, 292, and 656 are arginine (R); 243 is glutamine (Q); and 259 is glycine (G); 472 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 640 is asparagine (N); 723 is threonine (T); and 985 is leucine; (i) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:6 with the following amino acid residues: 51 and 208 are cysteine (C); 82 and 337 are phenylalanine (F); 86, 146, and 262 are serine (S); 98 and 179 are isoleucine (I); 188, 296, and 363 are valine (V); 231, 291, and 664 are arginine (R); 243 is glutamine (Q); and 258 is glycine (G); 480 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 648 is asparagine (N); 733 is threonine (T); and 997 is leucine; and

(j) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:6 with the following amino acid residues: 51 and 208 are cysteine (C); 82 and 337 are phenylalanine (F); 86, 146, and 262 are serine (S); 98 and 179 are isoleucine (I); 188, 296, and 363 are valine (V); 231, 291, and 664 are arginine (R); 243 is glutamine (Q); and 258 is glycine (G); 480 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 648 is asparagine (N); 733 is threonine (T); and 997 is leucine.

[0037] In one embodiment, an isolated nucleic acid molecule that includes a nucleic acid molecule of the invention and a nucleotide sequence encoding a heterologous polypeptide, is provided.

[0038] In other embodiments, vectors that include a nucleic acid molecule of the invention are provided.

[0039] In other embodiments, host cells transfected with a nucleic acid molecule of the invention, or a vector that includes a nucleic acid molecule of the invention, are provided. Host cells include eucaryotic cells such as yeast cells, insect cells, or animal cells. Host cells also include procaryotic cells such as bacterial cells.

[0040] In other embodiments, methods for producing a cell that converts an alkane to alcohol, are provided. Such methods generally include: (a) transforming a cell with an isolated nucleic acid molecule encoding a polypeptide that includes an amino acid sequence set forth in SEQ ID NO: 7, 8, 9, 10, 11, 12, or 13; (b) transforming a cell with an isolated nucleic acid molecule encoding a polypeptide of the invention; or (c) transforming a cell with an isolated nucleic acid molecule of the invention.

[0041] In other embodiments, methods for selecting a cell that converts an alkane to an alcohol, are provided. The methods generally include: (a) providing a cell containing a nucleic acid construct that includes a nucleotide sequence

that encodes a modified cytochrome P450 polypeptide, the nucleotide sequence selected from: (i) a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 7, 8, 9, 10, 11, 12, or 13; (ii) a nucleic acid molecule encoding a polypeptide of the invention; or (iii) a nucleic acid molecule of the invention. The methods further include (b) culturing the cell in the presence of a suitable alkane and under conditions where the modified cytochrome P450 is expressed at a level sufficient to convert an alkane to an alcohol. Such conditions are met when an alcohol is produced at a level detectable by a method provided herein, or a method known to one skilled in the art of enzymology.

[0042] In other embodiments, methods for producing an alcohol, are provided. The methods include: (a) providing a cell containing a nucleic acid construct comprising a nucleotide sequence that encodes a modified cytochrome P450 polypeptide, the nucleotide sequence selected from: (i) a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 7, 8, 9, 10, 11, 12, or 13; (ii) a nucleic acid molecule encoding a polypeptide of the invention; or (iii) a nucleic acid molecule of the invention. The methods further include (b) culturing the cell in the presence of a suitable alkane and under conditions where the modified cytochrome P450 is expressed at an effective level; and (c) producing an alcohol by hydroxylation of the suitable alkane such as methane (CH_4), ethane (C_2H_6), propane (C_3H_8), butane (C_4H_{10}), pentane (C_5H_{12}), hexane (C_6H_{14}), heptane (C_7H_{16}), octane (C_8H_{18}), nonane (C_9H_{20}), decane ($\text{C}_{10}\text{H}_{22}$), undecane ($\text{C}_{11}\text{H}_{24}$), and dodecane ($\text{C}_{12}\text{H}_{26}$). In general an alcohol produced by a method the invention can include methanol, ethanol, propanol, butanol, pentanol, hexanol, heptanol, octanol, nonanol, decanol, undecanol, and dodecanol.

[0043] In yet another embodiment, methods for producing a cytochrome P450 variant that hydroxylates an alkane are provided. Such methods include selecting a parent cytochrome P450 polypeptide and modifying at least one amino acid residue positioned in or near the active site of the heme domain of the parent cytochrome P450 polypeptide. In general the modification reduces the volume of the active site. The method further includes contacting the polypeptide comprising the modified amino acid with at least one alkane under conditions suitable for hydroxylation of the alkane and detecting a hydroxylated alkane. The modification may include a substitution of the parent amino acid for a different amino acid or it may include modifying the parent amino acid to include an additional group. If the modification is a substitution, such substitutions can include a phenylalanine, tyrosine, histidine, or serine for the parent amino acid residue positioned in or near the active site of the heme domain of the parent cytochrome P450 polypeptide.

[0044] The details of one or more embodiments of the disclosure are set forth in the accompanying drawings and the description below. Other features, objects, and advantages will be apparent from the description and drawings, and from the claims.

BRIEF DESCRIPTION OF DRAWINGS

[0045] **Figure 1** depicts the reduction of substrate size and increase in C-H bond dissociation energy achieved by directed evolution. Directed evolution was used to convert wild-type P450 BM-3 stepwise from a fatty acid hydroxylase into an enzyme capable of activating the higher-energy C-H bond of ethane. The figure shows the substrate range of wild-type P450 BM-3 and its mutant 53-5H and, for comparison, the range of substrates of other alkane monooxygenases.

[0046] **Figure 2** depicts a gas chromatogram of ethanol reaction mixtures using variant 53-5H as the catalyst. Prior to analysis, ethanol was derivatized to form ethyl nitrite for detection via an electron capture detector (ECD).

[0047] **Figure 3** depicts the conversion of ethane to ethanol using 200 nM protein corresponds to 50 and 250 turnovers per enzyme active site of 53-5H and 35-E11, respectively. Halving or doubling the enzyme concentration yielded approximately correspondingly lower or higher product concentrations, respectively. Control reactions, either without ethane or with inactivated protein, did not contain ethanol concentrations above the background level of 1 μ M. Error bars are the standard deviation of three experiments.

[0048] **Figure 4** depicts ^{13}C -NMR spectrum of singly ^{13}C -labeled ethanol incubated with 53-5H and NADPH regeneration system. No ^{13}C -labeled carbonyl peaks due to overoxidation of the ethanol were detectable.

[0049] **Figure 5** depicts an absorption spectrum of 53-5H. The ferric heme iron is in low-spin form. A shift to high-spin ferric iron upon addition of ethane could not be detected. Addition of octane did not induce an increase at 390 nm. The spectrum of 35-E11 is substantially similar.

[0050] **Figures 6A - 6F** depict an amino acid sequence alignment of 6 CYP102 homologs and BM-3 mutant 35-E11. Corresponding mutations in homologs are provided in Table 4.

[0051] **Figure 7** depicts the spectra of methane reactions of BM-3 variant 35-E11 and the variant 35-E11-E464R after treatment with alcohol oxidase and Purpald. The peak at 550 nm corresponds to methanol produced. Solutions containing all reactants except methane do not absorb in this region when similarly treated.

[0052] **Figure 8** depicts SEQ ID NO: 1 (CYP102A1 from *Bacillus megaterium*).

- [0053] **Figure 9** depicts SEQ ID NO: 2 (CYP102A2 from *Bacillus subtilis*, 59% identity to CYP102A1).
- [0054] **Figure 10** depicts SEQ ID NO: 3 (CYP102A3 from *Bacillus subtilis*, 58% identity to CYP102A1).
- [0055] **Figure 11** depicts SEQ ID NO: 4 (CYP102A5 from *Bacillus cereus*, 60% identity to CYP102A1).
- [0056] **Figure 12** depicts SEQ ID NO: 5 (CYP102E1 from *Ralstonia metallidurans*, 38% identity to CYP102A1).
- [0057] **Figure 13** depicts SEQ ID NO: 6 (CYP102A6 from *Bradyrhizobium japonicum*, 46% identity to CYP102A1).
- [0058] **Figure 14** depicts SEQ ID NO: 7 (35-E11).
- [0059] **Figure 15** depicts SEQ ID NO: 8 (35-E11-E464R).
- [0060] **Figure 16** depicts SEQ ID NO: 9 (35-E11-E464Y).
- [0061] **Figure 17** depicts SEQ ID NO: 10 (35-E11-E464T).
- [0062] **Figure 18** depicts SEQ ID NO: 11 (20-D3).
- [0063] **Figure 19** depicts SEQ ID NO: 12 (23-1D).
- [0064] **Figure 20** depicts SEQ ID NO: 13 (21-4G).
- [0065] **Figure 21** depicts SEQ ID NO: 15 (53-5H).
- [0066] **Figures 22A and 22B** depict SEQ ID NO: 16 (expression plasmid pCWORI-53-H containing the heme and reductase domain of variant 53-5H).
- [0067] **Figure 23** depicts SEQ ID NO: 17 (nucleic acid sequence encoding variant 35-E11).
- [0068] **Figures 24A and 24B** depict SEQ ID NO: 18 (expression plasmid pCWORI-35-E11 containing heme and reductase domain of variant 35-E11).
- [0069] **Figure 25** depicts alkyl methyl ether assays. Panel a) shows hydroxylation of the propane surrogate dimethyl ether produces formaldehyde which can easily be detected using the dye Purpald. Panel b) shows hexyl methyl ether can be used as an octane surrogate. Panel c) shows a typical screening plate with P450-containing cell lysate, which, upon hydroxylation of the methoxy group of DME or HME and addition of Purpald, forms

a purple color. Each well represents a different BM-3 variant.

[0070] **Figure 26** depicts standard curves for formaldehyde (■) and hexanal (♦) using Purpald. The assay was performed in a 96-well microtiter plate and the final volume was 250 μ L per well.

[0071] **Figure 27** depicts a colorimetric reaction after addition of Purpald to reaction mixtures containing BM-3 mutants, chloromethane and cofactor regeneration system. Panel A and Panel B refer to independent experiments.

[0072] **Figure 28** depicts total turnover numbers for the reaction of CM dehalogenation catalyzed by BM-3 mutants.

[0073] **Figure 29** depicts a colorimetric response for the reaction BM-3 variants substituted at position A328.

[0074] **Figure 30** depicts the chemical structures of products from the hydroxylation/dehalogenation reaction catalyzed by P450 mutants on different kinds of halogenated alkanes.

[0075] **Figures 31A - 31C** depict an amino acid sequence alignment of the heme domain (residues 1-455) of 6 CYP102 homologs and BM-3 mutant 35-E11. Corresponding mutations in homologs are provided in Table 4.

[0076] **Figure 32** depicts protein similarity scores for polypeptides provided herein. Panel A shows exemplary similarity scores for full length polypeptides (e.g., SEQ ID NOs: 1 - 7). Panel B shows exemplary similarity scores for residues 1-455 of of SEQ ID NOs: 1 - 7 corresponding to the heme domain of P450 polypeptides.

[0077] **Figure 33** depicts protein percent (%) identities for polypeptides provided herein. Panel A shows exemplary percent (%) identities for full length polypeptides (e.g., SEQ ID NOs: 1 - 7). Panel B shows exemplary percent (%) identities for residues 1-455 of of SEQ ID NOs: 1 - 7 corresponding to the heme domain of P450 polypeptides.

[0078] Like reference symbols in the various drawings indicate like elements.

DETAILED DESCRIPTION

[0079] Before describing the invention in detail, it is to be understood that this invention is not limited to particular compositions or biological systems, which can, of course, vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting. As used in this specification and the appended claims, the singular forms "a," "an," and "the" include plural referents unless the content clearly dictates otherwise. Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which the invention pertains. Although any methods and materials similar or equivalent to those described herein can be used in the practice for testing of the invention(s), specific examples of appropriate materials and methods are described herein.

[0080] As will be described in more detail below, the invention is based, at least in part, on the generation and expression of novel enzymes that catalyze the hydroxylation of alkanes to alcohols. In one embodiment, novel polypeptides that have been engineered to convert an alkane to an alcohol are provided. Such polypeptides include P450 variants that have been altered to include amino acid substitutions at specified residues. While these variants will be described in more detail below, it is understood that polypeptides of the invention may contain one or more modified amino acids. The presence of modified amino acids may be advantageous in, for example, (a) increasing polypeptide in vivo half-life, (b) reducing or increasing polypeptide antigenicity, and (c) increasing polypeptide storage stability. Amino acid(s) are

modified, for example, co-translationally or post-translationally during recombinant production (e.g., N-linked glycosylation at N--X--S/T motifs during expression in mammalian cells) or modified by synthetic means. Accordingly, A "mutant", "variant" or "modified" protein, enzyme, polynucleotide, gene, or cell, means a protein, enzyme, polynucleotide, gene, or cell, that has been altered or derived, or is in some way different or changed, from a parent protein, enzyme, polynucleotide, gene, or cell. A mutant or modified protein or enzyme is usually, although not necessarily, expressed from a mutant polynucleotide or gene.

[0081] A "mutation" means any process or mechanism resulting in a mutant protein, enzyme, polynucleotide, gene, or cell. This includes any mutation in which a protein, enzyme, polynucleotide, or gene sequence is altered, and any detectable change in a cell arising from such a mutation. Typically, a mutation occurs in a polynucleotide or gene sequence, by point mutations, deletions, or insertions of single or multiple nucleotide residues. A mutation includes polynucleotide alterations arising within a protein-encoding region of a gene as well as alterations in regions outside of a protein-encoding sequence, such as, but not limited to, regulatory or promoter sequences. A mutation in a gene can be "silent", i.e., not reflected in an amino acid alteration upon expression, leading to a "sequence-conservative" variant of the gene. This generally arises when one amino acid corresponds to more than one codon.

[0082] Non-limiting examples of a modified amino acid include a glycosylated amino acid, a sulfated amino acid, a prenylated (e.g., farnesylated, geranylgeranylated) amino acid, an acetylated amino acid, an acylated amino acid, a pegylated amino acid, a biotinylated amino acid, a carboxylated amino acid, a phosphorylated amino acid, and the like. References adequate to guide one of skill in the

modification of amino acids are replete throughout the literature. Example protocols are found in Walker (1998) Protein Protocols on CD-ROM (Humana Press, Towata, N.J.).

[0083] Recombinant methods for producing and isolating modified P450 polypeptides of the invention are described herein. In addition to recombinant production, the polypeptides may be produced by direct peptide synthesis using solid-phase techniques (e.g., Stewart et al. (1969) Solid-Phase Peptide Synthesis (WH Freeman Co, San Francisco); and Merrifield (1963) J. Am. Chem. Soc. 85: 2149-2154). Peptide synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer, Foster City, Calif.) in accordance with the instructions provided by the manufacturer.

[0084] "Cytochrome P450 monooxygenase" or "P450 enzyme" means an enzyme in the superfamily of P450 haem-thiolate proteins, which are widely distributed in bacteria, fungi, plants and animals. The enzymes are involved in metabolism of a plethora of both exogenous and endogenous compounds. Usually, they act as terminal oxidases in multicomponent electron transfer chains, called here P450-containing monooxygenase systems. The unique feature which defines whether an enzyme is a cytochrome P450 enzyme is traditionally considered to be the characteristic absorption maximum ("Soret band") near 450 nm observed upon binding of carbon monoxide (CO) to the reduced form of the heme iron of the enzyme. Reactions catalyzed by cytochrome P450 enzymes include epoxidation, N-dealkylation, O-dealkylation, S-oxidation and hydroxylation. The most common reaction catalyzed by P450 enzymes is the monooxygenase reaction, i.e., insertion of one atom of oxygen into a substrate while the other oxygen atom is reduced to water.

[0085] "Heme domain" refers to an amino acid sequence within an oxygen carrier protein, which sequence is capable of binding an iron-complexing structure such as a porphyrin. Compounds of iron are typically complexed in a porphyrin (tetrapyrrole) ring that may differ in side chain composition. Heme groups can be the prosthetic groups of cytochromes and are found in most oxygen carrier proteins. Exemplary heme domains include that of P450 BM-3 (P450.sub.BM-P), SEQ ID NO:3, as well as truncated or mutated versions of these that retain the capability to bind the iron-complexing structure. The skilled artisan can readily identify the heme domain of a specific protein using methods known in the art.

[0086] A "protein" or "polypeptide", which terms are used interchangeably herein, comprises one or more chains of chemical building blocks called amino acids that are linked together by chemical bonds called peptide bonds. An "enzyme" means any substance, preferably composed wholly or largely of protein, that catalyzes or promotes, more or less specifically, one or more chemical or biochemical reactions. The term "enzyme" can also refer to a catalytic polynucleotide (e.g., RNA or DNA). A "native" or "wild-type" protein, enzyme, polynucleotide, gene, or cell, means a protein, enzyme, polynucleotide, gene, or cell that occurs in nature.

[0087] Accordingly, in various embodiments, isolated or recombinant polypeptides comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:1, 2, 3, 4, 5 or 6, are provided. The polypeptides include up to 50, 25, 10, or 5 conservative amino acid substitutions excluding residues: (a) 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, and 353 of SEQ ID NO:1; (b) 48, 79, 83, 95, 143, 176, 185, 206, 227, 238, 254, 257, 292, 330, and 355 of SEQ ID NO:2 or SEQ ID NO:3; (c) 49, 80, 84, 96, 144, 177, 186, 207, 228, 239, 255, 258, 293, 331, and 357 of SEQ ID NO:4; (d) 54, 85, 89, 101, 149, 182, 191, 212, 232, 243, 259,

262, 297, 336, and 361 of SEQ ID NO:5; and (e) 51, 82, 86, 98, 146, 179, 188, 208, 231, 242, 258, 262, 296, 337, and 363 of SEQ ID NO:6. The amino acid sequence includes the following residues:

(1) a Z1 amino acid residue at positions: (a) 47, 82, 142, 205, 236, 252, and 255 of SEQ ID NO:1; (b) 48, 83, 143, 206, 238, 254, and 257 of SEQ ID NO:2 or SEQ ID NO:3; (c) 49, 84, 144, 207, 239, 255, and 258 of SEQ ID NO:4; (d) 54, 89, 149, 212, 243, 259, and 262 of SEQ ID NO:5; and (e) 51, 86, 146, 208, 242, 258, and 262 of SEQ ID NO:6;

(2) a Z2 amino acid residue at positions: (a) 94, 175, 184, 290, and 353 of SEQ ID NO:1; (b) 95, 176, 185, 292, and 355 of SEQ ID NO:2 or SEQ ID NO:3; (c) 96, 177, 186, 293, and 357 of SEQ ID NO:4; (d) 101, 182, 191, 297, and 361 of SEQ ID NO:5; and (e) 98, 179, 188, 296, and 363 of SEQ ID NO:6;

(3) a Z3 amino acid residue at position: (a) 226 of SEQ ID NO:1; (b) 227 of SEQ ID NO:2 or SEQ ID NO:3; (c) 228 of SEQ ID NO:4; (d) 232 of SEQ ID NO:5; and (e) 231 of SEQ ID NO:6; and

(4) a Z4 amino acid residue at positions: (a) 78 and 328 of SEQ ID NO:1; (b) 79 and 330 of SEQ ID NO:2 or SEQ ID NO:3; (c) 80 and 331 of SEQ ID NO:4; (d) 85 and 336 of SEQ ID NO:5; and (e) 82 and 337 of SEQ ID NO:6.

[0088] In other embodiments, the polypeptide further includes a Z3 amino acid residue at position: (a) 285 of SEQ ID NO:1; (b) 287 of SEQ ID NO:2 or 3; (c) 288 of SEQ ID NO:4; (d) 292 of SEQ ID NO:5; and (e) 291 of SEQ ID NO:6. A Z3 amino acid residue includes lysine (K), arginine (R), or histidine (H). In some aspects, the amino acid residue at this position is an arginine (R).

[0089] An "amino acid sequence" is a polymer of amino acids (a protein, polypeptide, etc.) or a character string representing an amino acid polymer, depending on context. The

terms "protein," "polypeptide," and "peptide" are used interchangeably herein.

[0090] "Conservative amino acid substitution" or, simply, "conservative variations" of a particular sequence refers to the replacement of one amino acid, or series of amino acids, with essentially identical amino acid sequences. One of skill will recognize that individual substitutions, deletions or additions which alter, add or delete a single amino acid or a percentage of amino acids in an encoded sequence result in "conservative variations" where the alterations result in the deletion of an amino acid, addition of an amino acid, or substitution of an amino acid with a chemically similar amino acid.

[0091] Conservative substitution tables providing functionally similar amino acids are well known in the art. For example, one conservative substitution group includes Alanine (A), Serine (S), and Threonine (T). Another conservative substitution group includes Aspartic acid (D) and Glutamic acid (E). Another conservative substitution group includes Asparagine (N) and Glutamine (Q). Yet another conservative substitution group includes Arginine (R) and Lysine (K). Another conservative substitution group includes Isoleucine, (I) Leucine (L), Methionine (M), and Valine (V). Another conservative substitution group includes Phenylalanine (F), Tyrosine (Y), and Tryptophan (W).

[0092] Thus, "conservative amino acid substitutions" of a listed polypeptide sequence (e.g., SEQ ID NOs: 1 - 6) include substitutions of a percentage, typically less than 10%, of the amino acids of the polypeptide sequence, with a conservatively selected amino acid of the same conservative substitution group. Accordingly, a conservatively substituted variation of a polypeptide of the invention can contain 100, 75, 50, 25, or 10 substitutions with a conservatively substituted variation of the same conservative substitution group.

[0093] It is understood that the addition of sequences which do not alter the encoded activity of a nucleic acid molecule, such as the addition of a non-functional or non-coding sequence, is a conservative variation of the basic nucleic acid. The "activity" of an enzyme is a measure of its ability to catalyze a reaction, i.e., to "function", and may be expressed as the rate at which the product of the reaction is produced. For example, enzyme activity can be represented as the amount of product produced per unit of time or per unit of enzyme (e.g., concentration or weight), or in terms of affinity or dissociation constants. As used interchangeably herein a "cytochrome P450 activity", "biological activity of cytochrome P450" or "functional activity of cytochrome P450", refers to an activity exerted by a cytochrome P450 protein, polypeptide or nucleic acid molecule on a cytochrome P450 polypeptide substrate, as determined in vivo, or in vitro, according to standard techniques. The biological activity of cytochrome P450 is described herein.

[0094] One of skill in the art will appreciate that many conservative variations of the nucleic acid constructs which are disclosed yield a functionally identical construct. For example, as discussed above, owing to the degeneracy of the genetic code, "silent substitutions" (i.e., substitutions in a nucleic acid sequence which do not result in an alteration in an encoded polypeptide) are an implied feature of every nucleic acid sequence which encodes an amino acid. Similarly, "conservative amino acid substitutions," in one or a few amino acids in an amino acid sequence are substituted with different amino acids with highly similar properties, are also readily identified as being highly similar to a disclosed construct. Such conservative variations of each disclosed sequence are a feature of the polypeptides provided herein.

[0095] "Conservative variants" are proteins or enzymes in which a given amino acid residue has been changed without

altering overall conformation and function of the protein or enzyme, including, but not limited to, replacement of an amino acid with one having similar properties, including polar or non-polar character, size, shape and charge. Amino acids other than those indicated as conserved may differ in a protein or enzyme so that the percent protein or amino acid sequence similarity between any two proteins of similar function may vary and can be, for example, at least 30%, at least 50%, at least 70%, at least 80%, or at least 90%, as determined according to an alignment scheme. As referred to herein, "sequence similarity" means the extent to which nucleotide or protein sequences are related. The extent of similarity between two sequences can be based on percent sequence identity and/or conservation. "Sequence identity" herein means the extent to which two nucleotide or amino acid sequences are invariant. "Sequence alignment" means the process of lining up two or more sequences to achieve maximal levels of identity (and, in the case of amino acid sequences, conservation) for the purpose of assessing the degree of similarity. Numerous methods for aligning sequences and assessing similarity/identity are known in the art such as, for example, the Cluster Method, wherein similarity is based on the MEGALIGN algorithm, as well as BLASTN, BLASTP, and FASTA (Lipman and Pearson, 1985; Pearson and Lipman, 1988). When using all of these programs, the preferred settings are those that results in the highest sequence similarity.

[0096] Non-conservative modifications of a particular polypeptide are those which substitute any amino acid not characterized as a conservative substitution. For example, any substitution which crosses the bounds of the six groups set forth above. These include substitutions of basic or acidic amino acids for neutral amino acids, (e.g., Asp, Glu, Asn, or Gln for Val, Ile, Leu or Met), aromatic amino acid for basic or acidic amino acids (e.g., Phe, Tyr or Trp for Asp, Asn, Glu

or Gln) or any other substitution not replacing an amino acid with a like amino acid. Basic side chains include lysine (K), arginine (R), histidine (H); acidic side chains include aspartic acid (D), glutamic acid (E); uncharged polar side chains include glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), cysteine (C); nonpolar side chains include alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), phenylalanine (F), methionine (M), tryptophan (W); beta-branched side chains include threonine (T), valine (V), isoleucine (I); aromatic side chains include tyrosine (Y), phenylalanine (F), tryptophan (W), histidine (H).

[0097] Accordingly, some amino acid residues at specific positions in a polypeptide are "excluded" from conservative amino acid substitutions. Instead, these restricted amino acids are generally chosen from a "Z" group. "Z" groups are amino acid groups designated Z1, Z2, Z3, Z4, Z5, and Z6. These amino acid residues can be substituted at a designated position to obtain a modified or variant polypeptide. While some overlap may occur, the members of each "Z" group are not "conservative amino acid substitutions" as defined above. The "Z" group members are established based upon the substitutions set forth in Table 4. In general, these mutations represent non-conservative substitutions at the indicated position in the designated sequence. For example, as shown in the first column at the first row, an arginine (R) to cysteine (C) substitution is made in SEQ ID NO:1 at position 47. This substitution is generally not considered a "conservative" substitution. Similar substitutions are made throughout the various sequences at the indicated positions in order to modify the activity of the polypeptide. Such modifications are made in order to generate polypeptides that are active on substrates such as alkanes. Referring again to Table 4, a parent polypeptide (e.g., SEQ ID NO:1, 2, 3, 4, 5 or 6) can be

modified to include an amino acid substitution at a particular residue by modifying the nucleic acid sequence that encodes the parent polypeptide (see below for additional information regarding modifying nucleic acid sequences). The region of P450 modified is identified on the left side of Table 4. The specific substitutions are identified throughout Table 4. Accordingly, "Z" groups are identified as follows: a Z1 amino acid residue includes glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), or cysteine (C); a Z2 amino acid residue includes alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), or methionine (M); a Z3 amino acid residue includes lysine (K), or arginine (R); a Z4 amino acid residue includes tyrosine (Y), phenylalanine (F), tryptophan (W), or histidine (H); a Z5 amino acid residue includes threonine (T), valine (V), and isoleucine (I); and a Z6 amino acid residue includes aspartic acid (D) and glutamic acid (E).

[0098] Accordingly, a polypeptide provided herein can include amino acids that are "restricted" to particular amino acid substitutions. For example, residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, and 353 can be restricted to substitutions set forth in a "Z" group as defined below and throughout the specification. It is understood that not all of the identified restricted residues need be altered in the same polypeptide. In some embodiments, the invention encompasses polypeptides where only about 80%, 85%, 90% or 95% of the restricted amino acid residues are altered in a given polypeptide.

[0099] Member residues of each "Z" group can be included at the appropriate position in a designated polypeptide. Accordingly, in other embodiments, the amino acid sequence of the polypeptide includes residues at the following positions:

(1) a glycine (G), glutamine (Q), serine (S), threonine (T), or cysteine (C) amino acid residue at position: (a) 47,

82, 142, 205, 236, 252, and 255 of SEQ ID NO:1; (b) 48, 83, 143, 206, 238, 254, and 257 of SEQ ID NO:2 or SEQ ID NO:3; (c) 49, 84, 144, 207, 239, 255, and 258 of SEQ ID NO:4; (d) 54, 89, 149, 212, 243, 259, and 262 of SEQ ID NO:5; and (e) 51, 86, 146, 208, 242, 258, and 262 of SEQ ID NO:6;

(2) a valine (V) or isoleucine (I) amino acid residue at position: (a) 94, 175, 184, 290, and 353 of SEQ ID NO:1; (b) 95, 176, 185, 292, and 355 of SEQ ID NO:2 or SEQ ID NO:3; (c) 96, 177, 186, 293, and 357 of SEQ ID NO:4; (d) 101, 182, 191, 297, and 361 of SEQ ID NO:5; and (e) 98, 179, 188, 296, and 363 of SEQ ID NO:6;

(3) an arginine amino acid residue at position: (a) 226 of SEQ ID NO:1; (b) 227 of SEQ ID NO:2 or SEQ ID NO:3; (c) 228 of SEQ ID NO:4; (d) 232 of SEQ ID NO:5; and (e) 231 of SEQ ID NO:6; and

(4) a phenylalanine (F) or histidine (H) amino acid residue at position: (a) 78 and 328 of SEQ ID NO:1; (b) 79 and 330 of SEQ ID NO:2 or SEQ ID NO:3; (c) 80 and 331 of SEQ ID NO:4; (d) 85 and 336 of SEQ ID NO:5; and (e) 82 and 337 of SEQ ID NO:6.

[00100] The term "substrate" or "suitable substrate" means any substance or compound that is converted or meant to be converted into another compound by the action of an enzyme catalyst. The term includes alkanes, and includes not only a single compound, but also combinations of compounds, such as solutions, mixtures and other materials which contain at least one substrate. Substrates for hydroxylation using the cytochrome P450 enzymes of the invention include para-nitrophenoxy-carboxylic acids ("pNCAs") such as 12-pNCA, as well as decanoic acid, myristic acid, lauric acid, and other fatty acids and fatty acid-derivatives. For alkane/alkene-substrates, propane, propene, ethane, ethene, butane, butene, pentane, pentene, hexane, hexene, cyclohexane, octane, octene, styrene, p-nitrophenoxyoctane (8-pnpene), and various

derivatives thereof, can be used. The term "derivative" refers to the addition of one or more functional groups to a substrate, including, but not limited, alcohols, amines, halogens, thiols, amides, carboxylates, etc.

[00101] A "parent" protein, enzyme, polynucleotide, gene, or cell, is any protein, enzyme, polynucleotide, gene, or cell, from which any other protein, enzyme, polynucleotide, gene, or cell, is derived or made, using any methods, tools or techniques, and whether or not the parent is itself native or mutant. A parent polynucleotide or gene encodes for a parent protein or enzyme.

[00102] In other embodiments, the amino acid sequence of the polypeptide includes residues at the following positions:

(1) a serine (S) residue at position: (a) 82, 142, and 255 of SEQ ID NO:1; (b) 83, 143 and 257 of SEQ ID NO:2 or SEQ ID NO:3; (c) 84, 144, and 258 of SEQ ID NO:4; (d) 89, 149, and 262 of SEQ ID NO:5; (e) 86, 146 and 262 of SEQ ID NO:6;

(2) a cysteine (C) amino acid residue at position: (a) 47 and 205 of SEQ ID NO:1; (b) 48 and 206 of SEQ ID NO:2 or SEQ ID NO:3; (c) 49 and 207 of SEQ ID NO:4; (d) 54 and 212 of SEQ ID NO:5; (e) 51 and 208 of SEQ ID NO:6;

(3) a glutamine (Q) amino acid residue at position: (a) 236 of SEQ ID NO:1; (b) 238 of SEQ ID NO:2 or SEQ ID NO:3; (c) 239 of SEQ ID NO:4; (d) 243 of SEQ ID NO:5; (e) 242 of SEQ ID NO:6;

(4) a glycine (G) amino acid residue at position: (a) 252 of SEQ ID NO:1; (b) 254 of SEQ ID NO:2 or SEQ ID NO:3; (c) 255 of SEQ ID NO:4; (d) 259 of SEQ ID NO:5; (e) 258 of SEQ ID NO:6;

(5) a valine (V) amino acid residue at position: (a) 184, 290 and 353 of SEQ ID NO:1; (b) 185, 292, and 355 of SEQ ID NO:2 or SEQ ID NO:3; (c) 186, 293 and 357 of SEQ ID NO:4; (d) 191, 297, and 361 of SEQ ID NO:5; (e) 188, 296, and 363 of SEQ ID NO:6;

(6) an isoleucine (I) amino acid residue at position: (a) 94 and 175 of SEQ ID NO:1; (b) 95 and 176 of SEQ ID NO:2 or SEQ ID NO:3; (c) 96 and 177 of SEQ ID NO:4; (d) 101 and 182 of SEQ ID NO:5; (e) 98 and 179 of SEQ ID NO:6; and

(7) a phenylalanine (F) amino acid residue at position: (a) 78 and 328 of SEQ ID NO:1; (b) 79 and 330 of SEQ ID NO:2 or SEQ ID NO:3; (c) 80 and 331 of SEQ ID NO:4; (d) 85 and 336 of SEQ ID NO:5; (e) 82 and 337 of SEQ ID NO:6.

[00103] Referring to the sequence comparison of various cytochromes P450 in Figure 6A - 6F, SEQ ID NO:1 includes the amino acid sequence of cytochrome P450 BM-3 isolated from *Bacillus megaterium*. This wild-type P450 BM-3 is also designated CYP102A1. SEQ ID NO:2 provides the amino acid sequence of wild-type cytochrome P450 from *Bacillus subtilis* strain 1A1.15. This wild-type P450 is also designated CYP102A2 and shares 59% amino acid sequence identity to CYP102A1 (SEQ ID NO:1). SEQ ID NO:3 includes the amino acid sequence of wild-type cytochrome P450 from *Bacillus subtilis* strain 1A1.15). This wild-type P450 is also designated CYP102A3 and shares 58% amino acid sequence identity to CYP102A1 (SEQ ID NO:1). SEQ ID NO:4 includes the amino acid sequence of wild-type cytochrome P450 from *Bacillus cereus*. This wild-type P450 is also designated CYP102A5 and shares 60% amino acid sequence identity to CYP102A1 (SEQ ID NO:1). SEQ ID NO:5 includes the amino acid sequence of wild-type cytochrome P450 from *Ralstonia metallidurans*. This wild-type P450 is also designated CYP102E1 and shares 38% amino acid sequence identity to CYP102A1 (SEQ ID NO:1). SEQ ID NO:6 includes the amino acid sequence of wild-type cytochrome P450 from *Bradyrhizobium japonicum*. This wild-type P450 is also designated CYP102A6 and shares 38% amino acid sequence identity to CYP102A1 (SEQ ID NO:1).

[00104] A polynucleotide, polypeptide, or other component is "isolated" when it is partially or completely separated from

components with which it is normally associated (other proteins, nucleic acids, cells, synthetic reagents, etc.). A nucleic acid or polypeptide is "recombinant" when it is artificial or engineered, or derived from an artificial or engineered protein or nucleic acid. For example, a polynucleotide that is inserted into a vector or any other heterologous location, e.g., in a genome of a recombinant organism, such that it is not associated with nucleotide sequences that normally flank the polynucleotide as it is found in nature is a recombinant polynucleotide. A protein expressed in vitro or in vivo from a recombinant polynucleotide is an example of a recombinant polypeptide. Likewise, a polynucleotide sequence that does not appear in nature, for example a variant of a naturally occurring gene, is recombinant.

[00105] The cytochromes P450 set forth in SEQ ID NOs:1 - 6 are closely related to one another and show a high degree of sequence identity. The sequences can be aligned based on the sequence homology. The alignment provided in Figure 6A - 6F identifies "equivalent positions" in the sequences. An equivalent position denotes a position which, on the basis of the alignment of the sequence of the parent cytochrome P450 in question with the "reference" cytochrome P450 amino acid sequence in question (e.g. SEQ ID NO: 1) so as to achieve juxtapositioning of amino acid residues which are common to both, corresponds most closely to a particular position in the reference sequence in question. This process can cause gaps or insertions to appear in the sequences. In the alignment of Figures 6A - 6F or Figures 31A - 31C, equivalent positions are shown lined up vertically with one another. For example, position 87 in SEQ ID NO: 1 is equivalent to position 88 in SEQ ID NO: 2 and 3 and position 89 in SEQ ID NO: 4. As shown in Figure 6A - 6F and more particularly in Figures 31A - 31C, the identity in the heme domain is often higher than reported

because of large differences in the linker region (~residue 470) between the heme and reductase domains (see Gonindaraj & Poulos, J. Biol. Chem. 272:7915).

[00106] It is envisioned that modifying the regions of a P450 polypeptide associated with the "linker" region and "reductase" domain. The "linker" region generally includes those residues from positions about 450 to about 480 of, for example, SEQ ID NO:1 - 6. In addition, the "reductase" domain generally includes those residues from positions about 470 to about 1050 of SEQ ID NOs:1-6. As shown in Table 4, polypeptides that include amino acid substitutions at these residues are provided. These substitutions can be made in conjunction with substitutions in the "heme" domain or they can be made in isolation from those in the heme domain.

[00107] Accordingly, in yet another embodiment, an isolated or recombinant polypeptide that includes residues 456-1048 of SEQ ID NO:1, 456-1059 of SEQ ID NO:2, 456-1053 of SEQ ID NO:3, 456-1064 of SEQ ID NO:4, 456-1063 of SEQ ID NO:5, or 456-1077 of SEQ ID NO:6, is provided. The polypeptide includes an amino acid sequence with up to 65, 40, 25, or 10 conservative amino acid substitutions excluding residues: (a) 464, 631, 645, 710 and 968 of SEQ ID NO:1; (b) 475, 641, 656, 721 and 980 of SEQ ID NO:2; (c) 467, 634, 648, 713 and 972 of SEQ ID NO:3; (d) 477, 644, 659, 724 and 983 of SEQ ID NO:4; (e) 472, 640, 656, 723 and 985 of SEQ ID NO:5; and (f) 480, 648, 664, 733 and 997 of SEQ ID NO:6. The amino acid sequence includes the following residues:

(1) a Z1, Z3, Z4, or Z5 amino acid residue at position: (a) 464 of SEQ ID NO:1; (b) 475 of SEQ ID NO:2; (c) 467 of SEQ ID NO:3; (d) 477 of SEQ ID NO:4; (e) 472 of SEQ ID NO:5; and (f) 480 of SEQ ID NO:6;

(2) a Z1 amino acid residue at position: (a) 631 and 710 of SEQ ID NO:1; (b) 641 and 721 of SEQ ID NO:2; (c) 634 and 713

of SEQ ID NO:3; (d) 644 and 724 of SEQ ID NO:4; (e) 640 and 723 of SEQ ID NO:5; and (f) 648 and 733 of SEQ ID NO:6;

(3) a Z3 amino acid residue at position: (a) 645 of SEQ ID NO:1; (b) 656 of SEQ ID NO:2; (c) 648 of SEQ ID NO:3; (d) 659 of SEQ ID NO:4; (e) 656 of SEQ ID NO:5; and (f) 664 of SEQ ID NO:6; and

(4) a Z2 amino acid residue at position: (a) 968 of SEQ ID NO:1; (b) 980 of SEQ ID NO:2; (c) 972 of SEQ ID NO:3; (d) 983 of SEQ ID NO:4; (e) 985 of SEQ ID NO:5; and (f) 997 of SEQ ID NO:6. Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine (C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and arginine (R); Z4 is an amino acid residue selected from the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H); and is an amino acid residue selected from the group consisting of threonine (T), valine (V), and isoleucine (I).

[00108] In other embodiments, the amino acid sequence of the polypeptide includes residues at the following positions:

(1) a glycine (G), arginine (R), tyrosine (Y), or threonine (T) amino acid residue at position: (a) 464 of SEQ ID NO:1; (b) 475 of SEQ ID NO:2; (c) 467 of SEQ ID NO:3; (d) 477 of SEQ ID NO:4; (e) 472 of SEQ ID NO:5; and (f) 480 of SEQ ID NO:6;

(2) an asparagine (N) amino acid residue at position: (a) 631 of SEQ ID NO:1; (b) 641 of SEQ ID NO:2; (c) 634 of SEQ ID NO:3; (d) 644 of SEQ ID NO:4; (e) 640 of SEQ ID NO:5; and (f) 648 of SEQ ID NO:6;

(3) an arginine (R) amino acid residue at position: (a) 645 of SEQ ID NO:1; (b) 656 of SEQ ID NO:2; (c) 648 of SEQ ID

NO:3; (d) 659 of SEQ ID NO:4; (e) 656 of SEQ ID NO:5; and (f) 664 of SEQ ID NO:6;

(4) a threonine (T) amino acid residue at position: (a) 710 of SEQ ID NO:1; (b) 721 of SEQ ID NO:2; (c) 713 of SEQ ID NO:3; (d) 724 of SEQ ID NO:4; (e) 723 of SEQ ID NO:5; and (f) 733 of SEQ ID NO:6; and

(5) a lysine (L) amino acid residue at position: (a) 968 of SEQ ID NO:1; (b) 980 of SEQ ID NO:2; (c) 972 of SEQ ID NO:3; (d) 983 of SEQ ID NO:4; (e) 985 of SEQ ID NO:5; and (f) 997 of SEQ ID NO:6.

[00109] As noted above, polypeptides provided herein generally include a "heme domain" which includes amino acid residues 1 to about 455 of the various P450 sequences provided in, for example, SEQ ID NO:1 through SEQ ID NO:13. As used herein, the term "heme domain" refers to a catalytically functional region of the polypeptide that exhibits monooxygenase/hydroxylase activity. A heme domain is a redox domain capable of binding an iron-complexing structure such as a porphyrin. Compounds of iron are typically complexed in a porphyrin (tetrapyrrole) ring that may differ in side chain composition. Heme groups can be the prosthetic groups of cytochromes and are found in most oxygen carrier proteins. Exemplary heme domains include that of P450 BM-3 (e.g., amino acid residues 1 to about 455 of SEQ ID NO:1), as well as truncated or mutated versions of these that retain the capability to bind the iron-complexing structure. The skilled artisan can readily identify the heme domain of a specific protein using methods known in the art. Accordingly, the skilled artisan will recognize that amino acid residues 1 to "about" or "approximately" 430 means within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which will depend in part on how the value is measured or determined, i.e., the limitations of the measurement system. For example, "about" can mean a range

of up to 20%, up to 10%, up to 5%, up to 1% of a given value, such as the number of amino acid residues in a heme domain or a reductase domain.

[00110] Accordingly, the term "hydroxylase" or "monooxygenase" should be considered to include any enzyme that can insert one oxygen atom from diatomic oxygen into a substrate. Exemplary enzymes include the cytochrome P450 monooxygenases. "Cytochrome P450 monooxygenase" or "P450 enzyme" means an enzyme in the superfamily of P450 heme-thiolate proteins, which are widely distributed in bacteria, fungi, plants and animals. The unique feature which defines whether an enzyme is a cytochrome P450 enzyme is traditionally considered to be the characteristic absorption maximum ("Soret band") near 450 nm observed upon binding of carbon monoxide (CO) to the reduced form of the heme iron of the enzyme. Reactions catalyzed by cytochrome P450 enzymes include epoxidation, N-dealkylation, O-dealkylation, S-oxidation and hydroxylation. The most common reaction catalyzed by P450 enzymes is the monooxygenase reaction, i.e., insertion of one atom of oxygen into a substrate while the other oxygen atom is reduced to water.

[00111] The invention envisions the synthesis of chimeric polypeptides that include a heme catalytic domain derived from a first source and an electron transfer domain (e.g., a reductase domain) derived from a second source. The first source and the second source may differ by the genus, or the species from which they are derived, or they may be derived from the same species as one another but from different organelles or compartments in the same species. Generally the electron transfer domain is a heme reductase domain.

[00112] Accordingly, in addition to providing variants of full-length P450 polypeptides, chimeric polypeptides that comprise: 1) a variant heme domain isolated from a first organism and modified to include a new activity; and 2) a

variant reductase domain isolated from a second organism and modified to include a new activity or an activity that complements the heme domain, are provided. Methods for engineering a chimeric polypeptide of the invention are disclosed in U.S. Patent Application Publication Number 20050124025, the contents of which are incorporated herein by reference.

[00113] In another embodiment, an isolated or recombinant polypeptide that includes residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:1, 2, 3, 4, 5 or 6 is provided. The polypeptide includes amino acid residues in the amino acid sequence that correspond to the following positions, at least 80%, 85%, 90%, or 95% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as follows:

(1) a serine (S) residue at position: (a) 82, 142, and 255 of SEQ ID NO:1; (b) 83, 143 and 257 of SEQ ID NO:2 or SEQ ID NO:3; (c) 84, 144, and 258 of SEQ ID NO:4; (d) 89, 149, and 262 of SEQ ID NO:5; (e) 86, 146 and 262 of SEQ ID NO:6;

(2) a cysteine (C) amino acid residue at position: (a) 47 and 205 of SEQ ID NO:1; (b) 48 and 206 of SEQ ID NO:2 or SEQ ID NO:3; (c) 49 and 207 of SEQ ID NO:4; (d) 54 and 212 of SEQ ID NO:5; (e) 51 and 208 of SEQ ID NO:6;

(3) a glutamine (Q) amino acid residue at position: (a) 236 of SEQ ID NO:1; (b) 238 of SEQ ID NO:2 or SEQ ID NO:3; (c) 239 of SEQ ID NO:4; (d) 243 of SEQ ID NO:5; (e) 242 of SEQ ID NO:6;

(4) a glycine (G) amino acid residue at position: (a) 252 of SEQ ID NO:1; (b) 254 of SEQ ID NO:2 or SEQ ID NO:3; (c) 255 of SEQ ID NO:4; (d) 259 of SEQ ID NO:5; (e) 258 of SEQ ID NO:6;

(5) a valine (V) amino acid residue at position: (a) 184, 290 and 353 of SEQ ID NO:1; (b) 185, 292, and 355 of SEQ ID NO:2 or SEQ ID NO:3; (c) 186, 293 and 357 of SEQ ID NO:4; (d)

191, 297, and 361 of SEQ ID NO:5; (e) 188, 296, and 363 of SEQ ID NO:6;

(6) an isoleucine (I) amino acid residue at position: (a) 94 and 175 of SEQ ID NO:1; (b) 95 and 176 of SEQ ID NO:2 or SEQ ID NO:3; (c) 96 and 177 of SEQ ID NO:4; (d) 101 and 182 of SEQ ID NO:5; (e) 98 and 179 of SEQ ID NO:6; and

(7) a phenylalanine (F) amino acid residue at position: (a) 78 and 328 of SEQ ID NO:1; (b) 79 and 330 of SEQ ID NO:2 or SEQ ID NO:3; (c) 80 and 331 of SEQ ID NO:4; (d) 85 and 336 of SEQ ID NO:5; (e) 82 and 337 of SEQ ID NO:6; and

(8) an arginine (R) amino acid residue at position: (a) 226 of SEQ ID NO:1; b) 227 of SEQ ID NO:2 or SEQ ID NO:3; (c) 228 of SEQ ID NO:4; (d) 232 of SEQ ID NO:5; (e) 231 of SEQ ID NO:6.

[00114] In other embodiments, the polypeptide further includes a Z3 amino acid residue at position: (a) 285 of SEQ ID NO:1; (b) 287 of SEQ ID NO:2 or 3; (c) 288 of SEQ ID NO:4; (d) 292 of SEQ ID NO:5; and (e) 291 of SEQ ID NO:6. A Z3 amino acid residue includes lysine (K), arginine (R), or histidine (H). In some aspects, the amino acid residue at this position is an arginine (R).

[00115] Accordingly, in some embodiments, a polypeptide provided herein includes amino acid residue substitutions that correspond to positions in a particular sequence at least 80%, 85%, 90%, or 95% of the time. In other words, the invention encompasses polypeptides that contain the recited amino acid substitutions at 80%, 85%, 90%, or 95% of the recited positions in a given sequence. The skilled artisan will recognize that not every substitution from a group of substitutions is necessary to obtain a modified polypeptide that is active on an alkane substrate.

[00116] In another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in SEQ ID NO:7, is provided. SEQ ID NO:7 (see Figure 14) provides

an amino acid sequence of a P450 BM-3 variant. This variant is also designated 35-E11. The polypeptide may contain up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710.

[00117] In yet another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in SEQ ID NO:8, is provided. SEQ ID NO:8 (see Figure 15) provides an amino acid sequence of a P450 BM-3 variant. This variant is also designated 35-E11-E464R. The polypeptide may contain up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710.

[00118] In another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in SEQ ID NO:9, is provided. SEQ ID NO:9 (see Figure 16) provides an amino acid sequence of a P450 BM-3 variant. This variant is also designated 35-E11-E464Y. The polypeptide may contain up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710.

[00119] In another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in SEQ ID NO:10, is provided. SEQ ID NO:10 (see Figure 17) provides an amino acid sequence of a P450 BM-3 variant. This variant is also designated 35-E11-E464T. The polypeptide may contain up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710.

[00120] In yet another embodiment, the polypeptides of SEQ ID NO:7, 8, 9 or 10, further optionally include an arginine (R) at amino acid residue position 285.

[00121] In another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in

SEQ ID NO:11, is provided. SEQ ID NO:11 (see Figure 18) provides an amino acid sequence of a P450 BM-3 variant. This variant is also designated 20-D3. The polypeptide may contain up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 285, 290, 328, 353, 464, and 710.

[00122] In another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in SEQ ID NO:12, is provided. SEQ ID NO:12 (see Figure 19) provides an amino acid sequence of a P450 BM-3 variant. This variant is also designated 23-1D. The polypeptide may contain up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 645, and 710.

[00123] In another embodiment, an isolated or recombinant polypeptide that includes the amino acid sequence set forth in SEQ ID NO:13, is provided. SEQ ID NO:13 (see Figure 20) provides an amino acid sequence of a P450 BM-3 variant. This variant is also designated 21-4G. The polypeptide may contain up to 75, 50, 25, or 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 631, and 710.

[00124] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:7; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:7; (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:7 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:7; or (d) a polypeptide comprising an amino acid sequence that can be optimally aligned with the sequence of SEQ ID NO:7 to generate a similarity score of at least 1830,

using the BLOSUM62 matrix, a gap existence penalty of 11, and a gap extension penalty of 1, excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:7.

[00125] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:8; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:8; or (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:8 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:8.

[00126] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:9; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:9; or (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:9 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:9.

[00127] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:10; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:10; or (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:10 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:10.

[00128] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:11; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:11; or (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:11 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 285, 290, 328, 353, 464, and 710 of SEQ ID NO:11.

[00129] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:12; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:12; or (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:12 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 645, and 710 of SEQ ID NO:12.

[00130] In other embodiments, isolated or recombinant polypeptides of the invention include: (a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:13; (b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:13; or (c) a polypeptide comprising an amino acid sequence having at least 60%, 70%, 80%, 90%, or 98% sequence identity to the amino acid sequence set forth in SEQ ID NO:13 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 631, and 710 of SEQ ID NO:13.

[00131] "Sequence identity" herein means the extent to which two nucleotide or amino acid sequences are invariant.

"Sequence alignment" means the process of lining up two or more sequences to achieve maximal levels of identity (and, in

the case of amino acid sequences, conservation) for the purpose of assessing the degree of similarity. Numerous methods for aligning sequences and assessing similarity/identity are known in the art such as, for example, the Cluster Method, wherein similarity is based on the MEGALIGN algorithm, as well as BLASTN, BLASTP, and FASTA (Lipman and Pearson, 1985; Pearson and Lipman, 1988). When using all of these programs, the preferred settings are those that results in the highest sequence similarity. For example, the "identity" or "percent identity" with respect to a particular pair of aligned amino acid sequences can refer to the percent amino acid sequence identity that is obtained by ClustalW analysis (version W 1.8 available from European Bioinformatics Institute, Cambridge, UK), counting the number of identical matches in the alignment and dividing such number of identical matches by the greater of (i) the length of the aligned sequences, and (ii) 96, and using the following default ClustalW parameters to achieve slow/accurate pairwise alignments--Gap Open Penalty: 10; Gap Extension Penalty: 0.10; Protein weight matrix: Gonnet series; DNA weight matrix: IUB; Toggle Slow/Fast pairwise alignments=SLOW or FULL Alignment.

[00132] Two sequences are "optimally aligned" when they are aligned for similarity scoring using a defined amino acid substitution matrix (e.g., BLOSUM62), gap existence penalty and gap extension penalty so as to arrive at the highest score possible for that pair of sequences. Amino acid substitution matrices and their use in quantifying the similarity between two sequences are well-known in the art and described, e.g., in Dayhoff et al. (1978) "A model of evolutionary change in proteins" in "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (ed. M. O. Dayhoff), pp. 345-352. Natl. Biomed. Res. Found., Washington, D.C. and Henikoff et al. (1992) Proc. Nat'l. Acad. Sci. USA 89: 10915-10919. The BLOSUM62 matrix (FIG. 10) is often used as a default scoring substitution

matrix in sequence alignment protocols such as Gapped BLAST 2.0. The gap existence penalty is imposed for the introduction of a single amino acid gap in one of the aligned sequences, and the gap extension penalty is imposed for each additional empty amino acid position inserted into an already opened gap. The alignment is defined by the amino acids positions of each sequence at which the alignment begins and ends, and optionally by the insertion of a gap or multiple gaps in one or both sequences so as to arrive at the highest possible score. While optimal alignment and scoring can be accomplished manually, the process is facilitated by the use of a computer-implemented alignment algorithm, e.g., gapped BLAST 2.0, described in Altschul et al. (1997) Nucl. Acids Res. 25: 3389-3402, and made available to the public at the National Center for Biotechnology Information (NCBI) Website (www.ncbi.nlm.nih.gov). Optimal alignments, including multiple alignments, can be prepared using, e.g., PSI-BLAST, available through the NCBI website and described by Altschul et al. (1997) Nucl. Acids Res. 25:3389-3402.

[00133] With respect to an amino acid sequence that is optimally aligned with a reference sequence, an amino acid residue "corresponds to" the position in the reference sequence with which the residue is paired in the alignment. The "position" is denoted by a number that sequentially identifies each amino acid in the reference sequence based on its position relative to the N-terminus. For example, in SEQ ID NO:1, position 1 is T, position 2 is I, position 3 is K, etc. When a test sequence is optimally aligned with SEQ ID NO:1, a residue in the test sequence that aligns with the K at position 3 is said to "correspond to position 3" of SEQ ID NO:1. Owing to deletions, insertion, truncations, fusions, etc., that must be taken into account when determining an optimal alignment, in general the amino acid residue number in a test sequence as determined by simply counting from the N-

terminal will not necessarily be the same as the number of its corresponding position in the reference sequence. For example, in a case where there is a deletion in an aligned test sequence, there will be no amino acid that corresponds to a position in the reference sequence at the site of deletion. Where there is an insertion in an aligned reference sequence, that insertion will not correspond to any amino acid position in the reference sequence. In the case of truncations or fusions there can be stretches of amino acids in either the reference or aligned sequence that do not correspond to any amino acid in the corresponding sequence.

[00134] In other embodiments, isolated nucleic acid molecules are provided. In one aspect, the invention provides a novel family of isolated or recombinant polynucleotides referred to herein as "P450 polynucleotides" or "P450 nucleic acid molecules." P450 polynucleotide sequences are characterized by the ability to encode a P450 polypeptide. In general, the invention includes any nucleotide sequence that encodes any of the novel P450 polypeptides described herein. In some aspects of the invention, a P450 polynucleotide that encodes a P450 variant polypeptide with activity on alkanes is provided. The terms "polynucleotide," "nucleotide sequence," and "nucleic acid molecule" are used to refer to a polymer of nucleotides (A, C, T, U, G, etc. or naturally occurring or artificial nucleotide analogues), e.g., DNA or RNA, or a representation thereof, e.g., a character string, etc., depending on the relevant context. A given polynucleotide or complementary polynucleotide can be determined from any specified nucleotide sequence.

[00135] In one aspect, the P450 polynucleotides comprise recombinant or isolated forms of naturally occurring nucleic acids isolated from an organism, e.g., a bacterial strain. Exemplary P450 polynucleotides include those that encode the wild-type polypeptides set forth in SEQ ID NO: 1, 2, 3, 4, 5,

or 6. In another aspect of the invention, P450 polynucleotides are produced by diversifying, e.g., recombining and/or mutating one or more naturally occurring, isolated, or recombinant P450 polynucleotides. As described in more detail elsewhere herein, it is often possible to generate diversified P450 polynucleotides encoding P450 polypeptides with superior functional attributes, e.g., increased catalytic function, increased stability, or higher expression level, than a P450 polynucleotide used as a substrate or parent in the diversification process. Exemplary polynucleotides include those that encode the P450 variant polypeptides set forth in SEQ ID NO: 7, 8, 9, 10, 11, 12, or 13.

[00136] The polynucleotides of the invention have a variety of uses in, for example recombinant production (i.e., expression) of the P450 polypeptides of the invention and as substrates for further diversity generation, e.g., recombination reactions or mutation reactions to produce new and/or improved P450 homologues, and the like.

[00137] It is important to note that certain specific, substantial and credible utilities of P450 polynucleotides do not require that the polynucleotide encode a polypeptide with substantial P450 activity or even variant P450 activity. For example, P450 polynucleotides that do not encode active enzymes can be valuable sources of parental polynucleotides for use in diversification procedures to arrive at P450 polynucleotide variants, or non-P450 polynucleotides, with desirable functional properties (e.g., high k_{cat} or k_{cat}/K_m , low K_m , high stability towards heat or other environmental factors, high transcription or translation rates, resistance to proteolytic cleavage, etc.).

[00138] P450 polynucleotides, including nucleotide sequences that encode P450 polypeptides and variants thereof, fragments of P450 polypeptides, related fusion proteins, or functional equivalents thereof, are used in recombinant DNA molecules

that direct the expression of the P450 polypeptides in appropriate host cells, such as bacterial cells. Due to the inherent degeneracy of the genetic code, other nucleic acid sequences which encode substantially the same or a functionally equivalent amino acid sequence can also be used to clone and express the P450 polynucleotides. The term "host cell", as used herein, includes any cell type which is susceptible to transformation with a nucleic acid construct. The term "transformation" means the introduction of a foreign (i.e., extrinsic or extracellular) gene, DNA or RNA sequence to a host cell, so that the host cell will express the introduced gene or sequence to produce a desired substance, typically a protein or enzyme coded by the introduced gene or sequence. The introduced gene or sequence may include regulatory or control sequences, such as start, stop, promoter, signal, secretion, or other sequences used by the genetic machinery of the cell. A host cell that receives and expresses introduced DNA or RNA has been "transformed" and is a "transformant" or a "clone." The DNA or RNA introduced to a host cell can come from any source, including cells of the same genus or species as the host cell, or cells of a different genus or species.

[00139] As will be understood by those of skill in the art, it can be advantageous to modify a coding sequence to enhance its expression in a particular host. The genetic code is redundant with 64 possible codons, but most organisms preferentially use a subset of these codons. The codons that are utilized most often in a species are called optimal codons, and those not utilized very often are classified as rare or low-usage codons (see, e.g., Zhang et al. (1991) *Gene* 105:61-72). Codons can be substituted to reflect the preferred codon usage of the host, a process sometimes called "codon optimization" or "controlling for species codon bias."

[00140] Optimized coding sequences containing codons preferred by a particular prokaryotic or eukaryotic host (see also, Murray et al. (1989) Nucl. Acids Res. 17:477-508) can be prepared, for example, to increase the rate of translation or to produce recombinant RNA transcripts having desirable properties, such as a longer half-life, as compared with transcripts produced from a non-optimized sequence. Translation stop codons can also be modified to reflect host preference. For example, preferred stop codons for *S. cerevisiae* and mammals are UAA and UGA, respectively. The preferred stop codon for monocotyledonous plants is UGA, whereas insects and *E. coli* prefer to use UAA as the stop codon (Dalphin et al. (1996) Nucl. Acids Res. 24: 216-218). Methodology for optimizing a nucleotide sequence for expression in a plant is provided, for example, in U.S. Pat. No. 6,015,891, and the references cited therein.

[00141] Accordingly, in some embodiments, nucleic acid molecules of the invention include: (a) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:7, 8, 9, 10, 11, 12, or 13; (b) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:7, 8, 9, 10, 11, 12, or 13; (c) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:7 or 11; (d) a nucleic acid molecule which encodes a polypeptide comprising residues about 456 to about 1088 of the amino acid sequence set forth in SEQ ID NO:7, 8, 9, 10, 12, or 13; (e) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:15; (f) a nucleic acid molecule consisting of the nucleotide sequence set forth in SEQ ID NO:15; (g) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:17; (h) a nucleic acid molecule consisting of the nucleotide sequence set forth in SEQ ID NO:17.

[00142] In other embodiments, nucleic acid molecules of the invention include: (a) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:2 or 3 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227 is arginine (R); 238 is glutamine (Q); and 254 is glycine (G); (b) a nucleic acid molecule which encodes a polypeptide consisting of residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:2 or 3 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227 is arginine (R); 238 is glutamine (Q); and 254 is glycine (G); (c) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:4 with the following amino acid residues: 49 and 207 are cysteine (C); 80 and 331 are phenylalanine (F); 84, 144, and 258 are serine (S); 96 and 177 are isoleucine (I); 186, 293, and 357 are valine (V); 228 is arginine (R); 239 is glutamine (Q); and 255 is glycine (G); (d) a nucleic acid molecule which encodes a polypeptide consisting of residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:4 with the following amino acid residues: 49 and 207 are cysteine (C); 80 and 331 are phenylalanine (F); 84, 144, and 258 are serine (S); 96 and 177 are isoleucine (I); 186, 293, and 357 are valine (V); 228 is arginine (R); 239 is glutamine (Q); and 255 is glycine (G); (e) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:5 with the following amino acid residues: 54 and 212 are cysteine (C); 85 and 336 are phenylalanine (F); 89, 149, and 262 are serine (S); 101 and

182 are isoleucine (I); 191, 297, and 361 are valine (V); 232 is arginine (R); 243 is glutamine (Q); and 259 is glycine (G);

(f) a nucleic acid molecule which encodes a polypeptide consisting of residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:5 with the following amino acid residues: 54 and 212 are cysteine (C); 85 and 336 are phenylalanine (F); 89, 149, and 262 are serine (S); 101 and 182 are isoleucine (I); 191, 297, and 361 are valine (V); 232 is arginine (R); 243 is glutamine (Q); and 259 is glycine (G);

(g) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:6 with the following amino acid residues: 51 and 208 are cysteine (C); 82 and 337 are phenylalanine (F); 86, 146, and 262 are serine (S); 98 and 179 are isoleucine (I); 188, 296, and 363 are valine (V); 231 is arginine (R); 243 is glutamine (Q); and 258 is glycine (G);

and (h) a nucleic acid molecule which encodes a polypeptide consisting of residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:6 with the following amino acid residues: 51 and 208 are cysteine (C); 82 and 337 are phenylalanine (F); 86, 146, and 262 are serine (S); 98 and 179 are isoleucine (I); 188, 296, and 363 are valine (V); 231 is arginine (R); 243 is glutamine (Q); and 258 is glycine (G).

[00143] In other embodiments, nucleic acid molecules of the invention include: (a) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227, 287, and 656 are arginine (R); 238 is glutamine (Q); 254 is glycine (G); 475 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 641 is asparagine (N); 721 is threonine (T); and 980 is leucine; (b) a nucleic acid molecule which encodes a polypeptide consisting

of the amino acid sequence set forth in SEQ ID NO:2 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227, 287, and 656 are arginine (R); 238 is glutamine (Q); 254 is glycine (G); 475 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 641 is asparagine (N); 721 is threonine (T); and 980 is leucine; (c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:3 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227, 287, and 648 are arginine (R); 238 is glutamine (Q); 254 is glycine (G); 467 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 634 is asparagine (N); 713 is threonine (T); and 972 is leucine; (d) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:3 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227, 287, and 648 are arginine (R); 238 is glutamine (Q); 254 is glycine (G); 467 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 634 is asparagine (N); 713 is threonine (T); and 972 is leucine; (e) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:4 with the following amino acid residues: 49 and 207 are cysteine (C); 80 and 331 are phenylalanine (F); 84, 144, and 258 are serine (S); 96 and 177 are isoleucine (I); 186, 293, and 357 are valine (V); 228, 288, and 659 are arginine (R); 239 is glutamine (Q); and 255 is glycine (G); 477 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 644 is

asparagine (N); 724 is threonine (T); and 983 is leucine; (f) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:4 with the following amino acid residues: 49 and 207 are cysteine (C); 80 and 331 are phenylalanine (F); 84, 144, and 258 are serine (S); 96 and 177 are isoleucine (I); 186, 293, and 357 are valine (V); 228, 288, and 659 are arginine (R); 239 is glutamine (Q); and 255 is glycine (G); 477 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 644 is asparagine (N); 724 is threonine (T); and 983 is leucine; (g) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:5 with the following amino acid residues: 54 and 212 are cysteine (C); 85 and 336 are phenylalanine (F); 89, 149, and 262 are serine (S); 101 and 182 are isoleucine (I); 191, 297, and 361 are valine (V); 232, 292, and 656 are arginine (R); 243 is glutamine (Q); and 259 is glycine (G); 472 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 640 is asparagine (N); 723 is threonine (T); and 985 is leucine; (h) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:5 with the following amino acid residues: 54 and 212 are cysteine (C); 85 and 336 are phenylalanine (F); 89, 149, and 262 are serine (S); 101 and 182 are isoleucine (I); 191, 297, and 361 are valine (V); 232, 292, and 656 are arginine (R); 243 is glutamine (Q); and 259 is glycine (G); 472 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 640 is asparagine (N); 723 is threonine (T); and 985 is leucine; (i) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:6 with the following amino acid residues: 51 and 208 are cysteine (C); 82 and 337 are phenylalanine (F); 86, 146, and 262 are serine (S); 98 and 179 are isoleucine (I); 188, 296, and 363 are valine (V); 231, 291, and 664 are arginine (R); 243 is

glutamine (Q); and 258 is glycine (G); 480 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 648 is asparagine (N); 733 is threonine (T); and 997 is leucine; and (j) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:6 with the following amino acid residues: 51 and 208 are cysteine (C); 82 and 337 are phenylalanine (F); 86, 146, and 262 are serine (S); 98 and 179 are isoleucine (I); 188, 296, and 363 are valine (V); 231, 291, and 664 are arginine (R); 243 is glutamine (Q); and 258 is glycine (G); 480 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 648 is asparagine (N); 733 is threonine (T); and 997 is leucine.

[00144] In one embodiment, an isolated nucleic acid molecule that includes a nucleic acid molecule of the invention and a nucleotide sequence encoding a heterologous polypeptide, is provided.

[00145] It will be appreciated by those skilled in the art that due to the degeneracy of the genetic code, a multitude of nucleotide sequences encoding modified P450 polypeptides of the invention may be produced, some of which bear substantial identity to the nucleic acid sequences explicitly disclosed herein. For instance, codons AGA, AGG, CGA, CGC, CGG, and CGU all encode the amino acid arginine. Thus, at every position in the nucleic acids of the invention where an arginine is specified by a codon, the codon can be altered to any of the corresponding codons described above without altering the encoded polypeptide. It is understood that U in an RNA sequence corresponds to T in a DNA sequence.

[00146] "Silent variations" are one species of "conservatively modified variations." One of skill will recognize that each codon in a nucleic acid (except AUG, which is ordinarily the only codon for methionine) can be modified by standard techniques to encode a functionally identical polypeptide. Accordingly, each silent variation of a nucleic

acid which encodes a polypeptide is implicit in any described sequence. The invention provides each and every possible variation of nucleic acid sequence encoding a polypeptide of the invention that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleic acid sequence encoding a P450 homologue polypeptide of the invention. All such variations of every nucleic acid herein are specifically provided and described by consideration of the sequence in combination with the genetic code. Any variant can be produced as noted herein.

[00147] In general, the invention includes any polypeptide encoded by a modified P450 polynucleotide derived by mutation, recursive sequence recombination, and/or diversification of the polynucleotide sequences described herein. In some aspects of the invention, a P450 polypeptide, such as a heme domain and/or a reductase domain, is modified by single or multiple amino acid substitutions, a deletion, an insertion, or a combination of one or more of these types of modifications. Substitutions can be conservative or non-conservative, can alter function or not, and can add new function. Insertions and deletions can be substantial, such as the case of a truncation of a substantial fragment of the sequence, or in the fusion of additional sequence, either internally or at N or C terminal.

[00148] One aspect of the invention pertains to isolated nucleic acid molecules that encode modified P450 polypeptides or biologically active portions thereof. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

[00149] An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. For example, with regards to genomic DNA, the term "isolated" includes nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kb, 4kb, 3kb, 2kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

[00150] A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule that encodes a polypeptide set forth in any of SEQ NOs:1-13, or having the nucleotide sequence of set forth in any of SEQ ID NOs:15-18, or a portion thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein.

[00151] A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to nucleotide sequences can be prepared by standard synthetic techniques, e.g., using an automated DNA

synthesizer. In some embodiments, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of a nucleotide sequence encoding a polypeptide set forth in any of SEQ NOs:1-13, or having the nucleotide sequence of set forth in any of SEQ ID NOs:15-18. In still another embodiment, an isolated nucleic acid molecule of the present invention comprises a nucleotide sequence which is at least about 50%, 54%, 55%, 60%, 62%, 65%, 70%, 75%, 78%, 80%, 85%, 86%, 90%, 95%, 97%, 98% or more homologous to the nucleotide sequence encoding a polypeptide set forth in any of SEQ NOs:1-13, or having the nucleotide sequence set forth in any of SEQ ID NOs:15-18, or a portion of any of these nucleotide sequences.

[00152] In addition to the nucleotide sequences encoding a polypeptide set forth in any of SEQ NOs:1-13, or having the nucleotide sequence set forth in any of SEQ ID NOs:15-18, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequences of the proteins may exist within a population. Such genetic polymorphisms may exist among individuals within a population due to natural allelic variation. Such natural allelic variations include both functional and non-functional proteins and can typically result in 1-5% variance in the nucleotide sequence of a gene. Any and all such nucleotide variations and resulting amino acid polymorphisms in genes that are the result of natural allelic variation and that do not alter the functional activity of a protein are intended to be within the scope of the invention.

[00153] Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention hybridizes under stringent conditions to a nucleic acid molecule comprising the nucleotide sequence encoding a polypeptide set forth in any of SEQ NOs:1-13, or having the nucleotide sequence set forth in any of SEQ ID NOs:15-18. In other embodiments, the nucleic

acid is at least 30, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, or 600 nucleotides in length. Nucleic acid molecules are "hybridizable" to each other when at least one strand of one polynucleotide can anneal to another polynucleotide under defined stringency conditions. Stringency of hybridization is determined, e.g., by (a) the temperature at which hybridization and/or washing is performed, and (b) the ionic strength and polarity (e.g., formamide) of the hybridization and washing solutions, as well as other parameters. Hybridization requires that the two polynucleotides contain substantially complementary sequences; depending on the stringency of hybridization, however, mismatches may be tolerated. Typically, hybridization of two sequences at high stringency (such as, for example, in an aqueous solution of 0.5 X SSC at 65°C) requires that the sequences exhibit some high degree of complementarity over their entire sequence. Conditions of intermediate stringency (such as, for example, an aqueous solution of 2 X SSC at 65°C) and low stringency (such as, for example, an aqueous solution of 2 X SSC at 55°C), require correspondingly less overall complementarity between the hybridizing sequences (1 X SSC is 0.15 M NaCl, 0.015 M Na citrate). Nucleic acid molecules that hybridize include those which anneal under suitable stringency conditions and which encode polypeptides or enzymes having the same function, such as the ability to catalyze the conversion of an alkane (e.g., methane) to an alcohol (e.g., methanol), of the invention. Further, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 30%, 40%, 50%, or 60% homologous to each other typically remain hybridized to each other. Preferably, the conditions are such that sequences at least about 70%, more preferably at least about 80%, even more preferably at least about 85% or 90% homologous to each other typically remain hybridized to

each other. In some cases, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a nucleic acid sequence encoding a polypeptide set forth in any of SEQ NOs:1-13, or having the nucleotide sequence set forth in any of SEQ ID NOs:15-18, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

[00154] The skilled artisan will appreciate that changes can be introduced by mutation into the nucleotide sequences of any nucleic acid sequence encoding a polypeptide set forth in any of SEQ NOs:1-13, or having the nucleotide sequence set forth in any of SEQ ID NOs:15-18, thereby leading to changes in the amino acid sequence of the encoded proteins. In some cases the alteration will lead to altered function of the polypeptide. In other cases the change will not alter the functional ability of the encoded polypeptide. In general, substitutions that do not alter the function of a polypeptide include nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. Generally these substitutions can be made in, for example, in the sequence encoding a polypeptide set forth in any of SEQ NOs:7-13, or having the nucleotide sequence set forth in any of SEQ ID NOs:15-18, without altering the ability of the enzyme to catalyze the oxidation of an alkane to an alcohol. A "non-essential" amino acid residue is a residue that can be altered from the parent sequence without altering the biological activity of the resulting polypeptide, e.g., catalyzing the conversion of methane to methanol.

[00155] Also contemplated are those situations where it is desirable to alter the activity of a parent polypeptide such that the polypeptide has new or increased activity on a particular substrate. It is understood that these amino acid

substitutions will generally not constitute "conservative" substitutions. Instead, these substitutions constitute non-conservative substitutions introduced in to a sequence in order to obtain a new or improved activity. For example, the polypeptides set forth in Table 4 (SEQ ID NOs: 1-6) describe specific amino acid substitutions that contribute to the alteration of the activity of a parent polypeptide. SEQ ID NOs: 1-6 are P450 wild-type polypeptides. These polypeptides may also be referred to as "parent" amino acid sequences because substitutions made to their amino acid sequence give rise to modified P450 polypeptides that can have altered or modified activity as compared to the parent sequence (see Table 4). For example, SEQ ID NO:1 provided the parent sequence for the modified P450 polypeptide set forth in SEQ ID NO:7. SEQ ID NO:7 includes amino acid substitutions that impart activities to the SEQ ID NO:7 polypeptide not measurable in the parent polypeptide. Accordingly, the nucleic acid molecule encoding the amino acid sequence of SEQ ID NO:1 provides a "parent" nucleic acid molecule from which mutations can be made to obtain a nucleic acid molecule that encodes a modified polypeptide that includes amino acid substitutions. In general, these mutations provide non-conservative amino acid substitutions at indicated position(s) in a designated sequence.

[00156] It is also understood that a modified polypeptide can constitute a "parent" polypeptide from which additional substitutions can be made. Accordingly, a parent polypeptide, and a nucleic acid molecule that encodes a parent polypeptide, includes modified polypeptides and not just "wild-type" sequences. For example, the polypeptide of SEQ ID NO:7 is a modified polypeptide with respect to SEQ ID NO:1 (i.e., the "parent" polypeptide). Similarly, the polypeptide of SEQ ID NO:8 is a modified polypeptide with respect to SEQ ID NO:7. Accordingly, SEQ ID NO:7 is the parent sequence of SEQ ID

NO:8. Further, the nucleic acid sequence encoding the polypeptide of SEQ ID NO:7 provides the parent sequence from which mutations can be made to obtain a nucleic acid sequence that encodes the polypeptide of SEQ ID NO:8.

[00157] It is also understood that an isolated nucleic acid molecule encoding a polypeptide homologous to the polypeptides of SEQ ID NOs:1-13 can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence encoding the particular polypeptide, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced into the nucleic acid sequence by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. In contrast to those positions where it may be desirable to make a non-conservative amino acid substitutions (see above), in some positions it is preferable to make conservative amino acid substitutions. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine).

[00158] Mutational methods of generating diversity include, for example, site-directed mutagenesis (Ling et al. (1997) "Approaches to DNA mutagenesis: an overview" Anal Biochem. 254(2): 157-178; Dale et al. (1996) "Oligonucleotide-directed

random mutagenesis using the phosphorothioate method" *Methods Mol. Biol.* 57:369-374; Smith (1985) "In vitro mutagenesis" *Ann. Rev. Genet.* 19:423-462; Botstein & Shortle (1985) "Strategies and applications of in vitro mutagenesis" *Science* 229:1193-1201; Carter (1986) "Site-directed mutagenesis" *Biochem. J.* 237:1-7; and Kunkel (1987) "The efficiency of oligonucleotide directed mutagenesis" in *Nucleic Acids & Molecular Biology* (Eckstein, F. and Lilley, D. M. J. eds., Springer Verlag, Berlin)); mutagenesis using uracil containing templates (Kunkel (1985) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Proc. Natl. Acad. Sci. USA* 82:488-492; Kunkel et al. (1987) "Rapid and efficient site-specific mutagenesis without phenotypic selection" *Methods in Enzymol.* 154, 367-382; and Bass et al. (1988) "Mutant Trp repressors with new DNA-binding specificities" *Science* 242:240-245); oligonucleotide-directed mutagenesis (*Methods in Enzymol.* 100: 468-500 (1983); *Methods in Enzymol.* 154: 329-350 (1987); Zoller & Smith (1982) "Oligonucleotide-directed mutagenesis using M13-derived vectors: an efficient and general procedure for the production of point mutations in any DNA fragment" *Nucleic Acids Res.* 10:6487-6500; Zoller & Smith (1983) "Oligonucleotide-directed mutagenesis of DNA fragments cloned into M13 vectors" *Methods in Enzymol.* 100:468-500; and Zoller & Smith (1987) "Oligonucleotide-directed mutagenesis: a simple method using two oligonucleotide primers and a single-stranded DNA template" *Methods in Enzymol.* 154:329-350); phosphorothioate-modified DNA mutagenesis (Taylor et al. (1985) "The use of phosphorothioate-modified DNA in restriction enzyme reactions to prepare nicked DNA" *Nucl. Acids Res.* 13: 8749-8764; Taylor et al. (1985) "The rapid generation of oligonucleotide-directed mutations at high frequency using phosphorothioate-modified DNA" *Nucl. Acids Res.* 13: 8765-8787; Nakamaye & Eckstein (1986) "Inhibition of restriction endonuclease Nci I

cleavage by phosphorothioate groups and its application to oligonucleotide-directed mutagenesis" Nucl. Acids Res. 14: 9679-9698; Sayers et al. (1988) "Y-T Exonucleases in phosphorothioate-based oligonucleotide-directed mutagenesis" Nucl. Acids Res. 16:791-802; and Sayers et al. (1988) "Strand specific cleavage of phosphorothioate-containing DNA by reaction with restriction endonucleases in the presence of ethidium bromide" Nucl. Acids Res. 16: 803-814); mutagenesis using gapped duplex DNA (Kramer et al. (1984) "The gapped duplex DNA approach to oligonucleotide-directed mutation construction" Nucl. Acids Res. 12: 9441-9456; Kramer & Fritz (1987) Methods in Enzymol. "Oligonucleotide-directed construction of mutations via gapped duplex DNA" 154:350-367; Kramer et al. (1988) "Improved enzymatic in vitro reactions in the gapped duplex DNA approach to oligonucleotide-directed construction of mutations" Nucl. Acids Res. 16: 7207; and Fritz et al. (1988) "Oligonucleotide-directed construction of mutations: a gapped duplex DNA procedure without enzymatic reactions in vitro" Nucl. Acids Res. 16: 6987-6999).

[00159] Additional suitable methods include point mismatch repair (Kramer et al. (1984) "Point Mismatch Repair" Cell 38:879-887), mutagenesis using repair-deficient host strains (Carter et al. (1985) "Improved oligonucleotide site-directed mutagenesis using M13 vectors" Nucl. Acids Res. 13: 4431-4443; and Carter (1987) "Improved oligonucleotide-directed mutagenesis using M13 vectors" Methods in Enzymol. 154: 382-403), deletion mutagenesis (Eghtedarzadeh & Henikoff (1986) "Use of oligonucleotides to generate large deletions" Nucl. Acids Res. 14: 5115), restriction-selection and restriction-purification (Wells et al. (1986) "Importance of hydrogen-bond formation in stabilizing the transition state of subtilisin" Phil. Trans. R. Soc. Lond. A 317: 415-423), mutagenesis by total gene synthesis (Nambiar et al. (1984) "Total synthesis and cloning of a gene coding for the ribonuclease S protein"

Science 223: 1299-1301; Sakamar and Khorana (1988) "Total synthesis and expression of a gene for the α -subunit of bovine rod outer segment guanine nucleotide-binding protein (transducin)" Nucl. Acids Res. 14: 6361-6372; Wells et al. (1985) "Cassette mutagenesis: an efficient method for generation of multiple mutations at defined sites" Gene 34:315-323; and Grundstrom et al. (1985) "Oligonucleotide-directed mutagenesis by microscale 'shot-gun' gene synthesis" Nucl. Acids Res. 13: 3305-3316); double-strand break repair (Mandecki (1986); Arnold (1993) "Protein engineering for unusual environments" Current Opinion in Biotechnology 4:450-455; and "Oligonucleotide-directed double-strand break repair in plasmids of Escherichia coli: a method for site-specific mutagenesis" Proc. Natl. Acad. Sci. USA, 83:7177-7181). Additional details on many of the above methods can be found in Methods in Enzymology Volume 154, which also describes useful controls for trouble-shooting problems with various mutagenesis methods.

[00160] Additional details regarding various diversity generating methods can be found in the following U.S. patents, PCT publications, and EPO publications: U.S. Pat. No. 5,605,793 to Stemmer (Feb. 25, 1997), "Methods for In vitro Recombination;" U.S. Pat. No. 5,811,238 to Stemmer et al. (Sep. 22, 1998) "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" U.S. Pat. No. 5,830,721 to Stemmer et al. (Nov. 3, 1998), "DNA Mutagenesis by Random Fragmentation and Reassembly;" U.S. Pat. No. 5,834,252 to Stemmer, et al. (Nov. 10, 1998) "End-Complementary Polymerase Reaction;" U.S. Pat. No. 5,837,458 to Minshull, et al. (Nov. 17, 1998), "Methods and Compositions for Cellular and Metabolic Engineering;" WO 95/22625, Stemmer and Cramer, "Mutagenesis by Random Fragmentation and Reassembly;" WO 96/33207 by Stemmer and Lipschutz "End Complementary Polymerase Chain Reaction;" WO

97/20078 by Stemmer and Cramer "Methods for Generating Polynucleotides having Desired Characteristics by Iterative Selection and Recombination;" WO 97/35966 by Minshull and Stemmer, "Methods and Compositions for Cellular and Metabolic Engineering;" WO 99/41402 by Punnonen et al. "Targeting of Genetic Vaccine Vectors;" WO 99/41383 by Punnonen et al. "Antigen Library Immunization;" WO 99/41369 by Punnonen et al. "Genetic Vaccine Vector Engineering;" WO 99/41368 by Punnonen et al. "Optimization of Immunomodulatory Properties of Genetic Vaccines;" EP 752008 by Stemmer and Cramer, "DNA Mutagenesis by Random Fragmentation and Reassembly;" EP 0932670 by Stemmer "Evolving Cellular DNA Uptake by Recursive Sequence Recombination;" WO 99/23107 by Stemmer et al., "Modification of Virus Tropism and Host Range by Viral Genome Shuffling;" WO 99/21979 by Apt et al., "Human Papillomavirus Vectors;" WO 98/31837 by del Cardayre et al. "Evolution of Whole Cells and Organisms by Recursive Sequence Recombination;" WO 98/27230 by Patten and Stemmer, "Methods and Compositions for Polypeptide Engineering;" WO 98/13487 by Stemmer et al., "Methods for Optimization of Gene Therapy by Recursive Sequence Shuffling and Selection;" WO 00/00632, "Methods for Generating Highly Diverse Libraries;" WO 00/09679, "Methods for Obtaining in vitro Recombined Polynucleotide Sequence Banks and Resulting Sequences;" WO 98/42832 by Arnold et al., "Recombination of Polynucleotide Sequences Using Random or Defined Primers;" WO 99/29902 by Arnold et al., "Method for Creating Polynucleotide and Polypeptide Sequences;" WO 98/41653 by Vind, "An in vitro Method for Construction of a DNA Library;" WO 98/41622 by Borchert et al., "Method for Constructing a Library Using DNA Shuffling;" WO 98/42727 by Pati and Zarling, "Sequence Alterations using Homologous Recombination;" WO 00/18906 by Patten et al., "Shuffling of Codon-Altered Genes;" WO 00/04190 by del Cardayre et al. "Evolution of Whole Cells and Organisms by Recursive Recombination;" WO 00/42561 by Cramer et al.,

"Oligonucleotide Mediated Nucleic Acid Recombination;" WO 00/42559 by Selifonov and Stemmer "Methods of Populating Data Structures for Use in Evolutionary Simulations;" WO 00/42560 by Selifonov et al., "Methods for Making Character Strings, Polynucleotides & Polypeptides Having Desired Characteristics;" WO 01/23401 by Welch et al., "Use of Codon-Variied Oligonucleotide Synthesis for Synthetic Shuffling;" and WO 01/64864 "Single-Stranded Nucleic Acid Template-Mediated Recombination and Nucleic Acid Fragment Isolation" by Affholter.

[00161] Also provided are recombinant constructs comprising one or more of the nucleic acid sequences as broadly described above. The constructs comprise a vector, such as, a plasmid, a cosmid, a phage, a virus, a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), or the like, into which a nucleic acid sequence of the invention has been inserted, in a forward or reverse orientation. In a preferred aspect of this embodiment, the construct further comprises regulatory sequences, including, for example, a promoter, operably linked to the sequence. Large numbers of suitable vectors and promoters are known to those of skill in the art, and are commercially available.

[00162] Accordingly, in other embodiments, vectors that include a nucleic acid molecule of the invention are provided. In other embodiments, host cells transfected with a nucleic acid molecule of the invention, or a vector that includes a nucleic acid molecule of the invention, are provided. Host cells include eucaryotic cells such as yeast cells, insect cells, or animal cells. Host cells also include procaryotic cells such as bacterial cells.

[00163] In other embodiments, methods for producing a cell that converts an alkane to alcohol, are provided. Such methods generally include: (a) transforming a cell with an isolated nucleic acid molecule encoding a polypeptide comprising an

amino acid sequence set forth in SEQ ID NO: 7, 8, 9, 10, 11, 12, or 13; (b) transforming a cell with an isolated nucleic acid molecule encoding a polypeptide of the invention; or (c) transforming a cell with an isolated nucleic acid molecule of the invention.

[00164] In other embodiments, methods for selecting a cell that converts an alkane to an alcohol, are provided. The methods generally include: (a) providing a cell containing a nucleic acid construct that includes a nucleotide sequence that encodes a modified cytochrome P450 polypeptide, the nucleotide sequence selected from: (i) a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 7, 8, 9, 10, 11, 12, or 13; (ii) a nucleic acid molecule encoding a polypeptide of the invention; or (iii) a nucleic acid molecule of the invention. The methods further include (b) culturing the cell in the presence of a suitable alkane and under conditions where the modified cytochrome P450 is expressed at an effective level; and (c) detecting the production of an alcohol.

[00165] In other embodiments, methods for producing an alcohol, are provided. The methods include: (a) providing a cell containing a nucleic acid construct comprising a nucleotide sequence that encodes a modified cytochrome P450 polypeptide, the nucleotide sequence selected from: (i) a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 7, 8, 9, 10, 11, 12, or 13; (ii) a nucleic acid molecule encoding a polypeptide of the invention; or (iii) a nucleic acid molecule of the invention. The methods further include (b) culturing the cell in the presence of a suitable alkane and under conditions where the modified cytochrome P450 is expressed at an effective level; and (c) producing an alcohol by hydroxylation of the suitable alkane such as methane (CH₄), ethane (C₂H₆), propane (C₃H₈), butane (C₄H₁₀), pentane (C₅H₁₂), hexane (C₆H₁₄),

heptane (C_7H_{16}), octane (C_8H_{18}), nonane (C_9H_{20}), decane ($C_{10}H_{22}$), undecane ($C_{11}H_{24}$), and dodecane ($C_{12}H_{26}$). In general an alcohol produced by a method the invention can include methanol, ethanol, propanol, butanol, pentanol, hexanol, heptanol, octanol, nonanol, decanol, undecanol, and dodecanol.

[00166] As previously discussed, general texts which describe molecular biological techniques useful herein, including the use of vectors, promoters and many other relevant topics, include Berger and Kimmel, Guide to Molecular Cloning Techniques, Methods in Enzymology Volume 152, (Academic Press, Inc., San Diego, Calif.) ("Berger"); Sambrook et al., Molecular Cloning--A Laboratory Manual, 2d ed., Vol. 1-3, Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1989 ("Sambrook") and Current Protocols in Molecular Biology, F. M. Ausubel et al., eds., Current Protocols, a joint venture between Greene Publishing Associates, Inc. and John Wiley & Sons, Inc., (supplemented through 1999) ("Ausubel"). Examples of protocols sufficient to direct persons of skill through in vitro amplification methods, including the polymerase chain reaction (PCR), the ligase chain reaction (LCR), Q β -replicase amplification and other RNA polymerase mediated techniques (e.g., NASBA), e.g., for the production of the homologous nucleic acids of the invention are found in Berger, Sambrook, and Ausubel, as well as in Mullis et al. (1987) U.S. Pat. No. 4,683,202; Innis et al., eds. (1990) PCR Protocols: A Guide to Methods and Applications (Academic Press Inc. San Diego, Calif.) ("Innis"); Arnheim & Levinson (Oct. 1, 1990) C&EN 36-47; The Journal Of NIH Research (1991) 3: 81-94; Kwoh et al. (1989) Proc. Natl. Acad. Sci. USA 86: 1173; Guatelli et al. (1990) Proc. Nat'l. Acad. Sci. USA 87: 1874; Lomell et al. (1989) J. Clin. Chem 35: 1826; Landegren et al. (1988) Science 241: 1077-1080; Van Brunt (1990) Biotechnology 8: 291-294; Wu and Wallace (1989) Gene 4:560; Barringer et al. (1990) Gene 89:117; and Sooknanan and Malek (1995) Biotechnology 13: 563-

564. Improved methods for cloning in vitro amplified nucleic acids are described in Wallace et al., U.S. Pat. No.

5,426,039. Improved methods for amplifying large nucleic acids by PCR are summarized in Cheng et al. (1994) Nature 369: 684-685 and the references cited therein, in which PCR amplicons of up to 40 kb are generated. One of skill will appreciate that essentially any RNA can be converted into a double stranded DNA suitable for restriction digestion, PCR expansion and sequencing using reverse transcriptase and a polymerase. See, e.g., Ausubel, Sambrook and Berger, all supra.

[00167] Also provided are engineered host cells that are transduced (transformed or transfected) with a vector provided herein (e.g., a cloning vector or an expression vector), as well as the production of polypeptides of the invention by recombinant techniques. The vector may be, for example, a plasmid, a viral particle, a phage, etc. The engineered host cells can be cultured in conventional nutrient media modified as appropriate for activating promoters, selecting transformants, or amplifying the P450 homologue gene. Culture conditions, such as temperature, pH and the like, are those previously used with the host cell selected for expression, and will be apparent to those skilled in the art and in the references cited herein, including, e.g., Sambrook, Ausubel and Berger, as well as e.g., Freshney (1994) Culture of Animal Cells: A Manual of Basic Technique, 3rd ed. (Wiley-Liss, New York) and the references cited therein.

[00168] The terms "vector", "vector construct" and "expression vector" mean the vehicle by which a DNA or RNA sequence (e.g. a foreign gene) can be introduced into a host cell, so as to transform the host and promote expression (e.g. transcription and translation) of the introduced sequence. Vectors typically comprise the DNA of a transmissible agent, into which foreign DNA encoding a protein is inserted by restriction enzyme technology. A common type of vector is a

"plasmid", which generally is a self-contained molecule of double-stranded DNA that can readily accept additional (foreign) DNA and which can readily introduced into a suitable host cell. A large number of vectors, including plasmid and fungal vectors, have been described for replication and/or expression in a variety of eukaryotic and prokaryotic hosts. Non-limiting examples include pKK plasmids (Clonetech), pUC plasmids, pET plasmids (Novagen, Inc., Madison, Wis.), pRSET or pREP plasmids (Invitrogen, San Diego, Calif.), or pMAL plasmids (New England Biolabs, Beverly, Mass.), and many appropriate host cells, using methods disclosed or cited herein or otherwise known to those skilled in the relevant art. Recombinant cloning vectors will often include one or more replication systems for cloning or expression, one or more markers for selection in the host, e.g., antibiotic resistance, and one or more expression cassettes.

[00169] The terms "express" and "expression" mean allowing or causing the information in a gene or DNA sequence to become manifest, for example producing a protein by activating the cellular functions involved in transcription and translation of a corresponding gene or DNA sequence. A DNA sequence is expressed in or by a cell to form an "expression product" such as a protein. The expression product itself, e.g. the resulting protein, may also be said to be "expressed" by the cell. A polynucleotide or polypeptide is expressed recombinantly, for example, when it is expressed or produced in a foreign host cell under the control of a foreign or native promoter, or in a native host cell under the control of a foreign promoter.

[00170] Polynucleotides provided herein can be incorporated into any one of a variety of expression vectors suitable for expressing a polypeptide. Suitable vectors include chromosomal, nonchromosomal and synthetic DNA sequences, e.g., derivatives of SV40; bacterial plasmids; phage DNA;

baculovirus; yeast plasmids; vectors derived from combinations of plasmids and phage DNA, viral DNA such as vaccinia, adenovirus, fowl pox virus, pseudorabies, adenovirus, adeno-associated viruses, retroviruses and many others. Any vector that transduces genetic material into a cell, and, if replication is desired, which is replicable and viable in the relevant host can be used.

[001711] Vectors can be employed to transform an appropriate host to permit the host to express an inventive protein or polypeptide. Examples of appropriate expression hosts include: bacterial cells, such as *E. coli*, *B. subtilis*, *Streptomyces*, and *Salmonella typhimurium*; fungal cells, such as *Saccharomyces cerevisiae*, *Pichia pastoris*, and *Neurospora crassa*; insect cells such as *Drosophila* and *Spodoptera frugiperda*; mammalian cells such as CHO, COS, BHK, HEK 293 or Bowes melanoma; or plant cells or explants, etc.

[001721] In bacterial systems, a number of expression vectors may be selected depending upon the use intended for the P450 polypeptide. For example, when large quantities of P450 polypeptide or fragments thereof are needed for commercial production or for induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified can be desirable. Such vectors include, but are not limited to, multifunctional *E. coli* cloning and expression vectors such as BLUESCRIPT (Stratagene), in which the P450 polypeptide coding sequence may be ligated into the vector in-frame with sequences for the amino-terminal Met and the subsequent 7 residues of beta-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke & Schuster (1989) *J. Biol. Chem.* 264: 5503-5509); pET vectors (Novagen, Madison Wis.); and the like.

[001731] Similarly, in the yeast *Saccharomyces cerevisiae* a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase and PGH may be

used for production of the P450 polypeptides of the invention. For reviews, see Ausubel (supra) and Grant et al. (1987) Methods in Enzymology 153:516-544.

EXAMPLES

[00174] The invention is illustrated in the following examples, which are provided by way of illustration and are not intended to be limiting.

[00175] Example 1: Ethane Conversion

[00176] A catalyst for the selective oxidation under mild conditions of gaseous alkanes into easily transported alcohols allows for the economical exploitation of both local, small natural gas resources and vast, remote reserves. Prior to the present studies, the selective conversion of ethane and methane mainly to their corresponding alcohols was not demonstrated. Provided herein are enzyme-based catalysts that convert ethane to ethanol and methane to methanol with high selectivity and productivity. The enzymes are function at ambient temperature and pressure, use dioxygen from the air as the oxidant, and produce little or no hazardous wastes.

[00177] Biological systems have evolved efficient and productive metalloenzymes that convert alkanes into alcohols using dioxygen from the air. Methane monooxygenases (MMO) catalyze the conversion of methane to methanol and thereby enable methanotrophic bacteria to use methane as a source of carbon and energy. Microorganisms that grow on larger gaseous alkanes (ethane, propane, butane) have also been reported and putatively use similar enzymes for the first, alkane oxidation step. The relatively well-studied MMOs have long been a source of inspiration for designers of chemical catalysts. Unfortunately, these structurally complex enzymes have never been functionally expressed in a heterologous organism suitable for bioconversion and process optimization and

therefore have proven to be of little practical use themselves for production of alcohols.

[00178] The present studies use directed evolution techniques to convert heme-containing medium-chain fatty acid hydroxylases in to hydroxylating catalysts for small alkanes. An exemplary hydroxylase includes cytochrome P450 BM-3 (CYP102A1) which is the third P450 identified in the bacterium *Bacillus megaterium*. P450 BM-3 is a soluble, catalytically self-sufficient fatty acid monooxygenase isolated from *Bacillus megaterium*. It has a multi-domain structure, composed of three domains: one FAD, one FMN and one heme domain, fused on the same 119 kDa polypeptide chain of 1048 residues. Furthermore, despite its bacterial origin, P450 BM-3 has been classified as a class II P450 enzyme, typical of microsomal eukaryotic P450s: it shares 30% sequence identity with microsomal fatty acid ω -hydroxylase, 35% sequence identity with microsomal NADPH:P450 reductase, and only 20% homology with other bacterial P450s. Homologues of P450 BM-3 (CYP102A1) have been recognized in other bacteria, including two characterized systems in *Bacillus subtilis*: CYP102A2 and CYP102A3.

[00179] Communication between the heme and reductase domains of BM-3 is substrate-dependent and fine-tuned to transfer a pair of electrons from an NADPH cofactor through two reductase-bound flavins and then shuttle the electrons, one at a time, into the hydroxylase domain during catalysis. Engineering the protein for activity on a new substrate such as ethane and/or methane must conserve this regulated communication, and poses significant technical challenges. The engineered enzyme must bind ethane and/or methane, a substrate considerably smaller than a fatty acid, directly above the heme in the active site of the hydroxylase domain. Substrate binding must initiate electron transfer from the reductase domain to the heme domain during catalysis. Once

electron transfer occurs, the active iron-oxo species must be capable of breaking the high energy C-H bond in ethane (101.1 kcal mol⁻¹ vs. 95-99 kcal mol⁻¹ for the secondary C-H bonds of the fatty acid substrates of wild type BM-3). Finally, the singly-hydroxylated product must be released from the active site before further oxidation can occur.

[00180] Crystal structures of BM-3 heme domain with and without substrate are available while a structure of the reductase domain is not. Furthermore, large conformational changes in both domains during catalysis make it difficult to identify amino acid substitutions which can address all these issues. In the present studies, ethane and/or methane hydroxylases were engineered using an 'evolutionary' strategy in which mutations are accumulated in multiple generations of random mutagenesis and screening. The approach has been to use directed evolution methods to adapt the enzyme to exhibit higher turnover on smaller and smaller alkanes (see e.g., Figure 1). Through successive rounds of mutagenesis of the heme domain (residues 1-429 of SEQ ID NO:1) the activity of BM-3 towards octane was increased. Screening was initially based on the hydroxylation of the octane surrogate substrate p-nitrophenyl octyl ether. The resulting mutant, 139-3, contained 11 amino acid substitutions distributed throughout the hydroxylase domain and exhibited good activity on octane. It had also acquired a small but appreciable activity on propane, producing 2-propanol and a very small (less than 3%) amount of 1-propanol. Subsequently, the propane-hydroxylating activity of this mutant was increased by randomly mutating its heme domain and screening the resulting libraries for activity towards the (propane surrogate) substrate dimethyl ether. Mutant 9-10A obtained in this way was much more active towards propane than 139-3 and produced significantly more 1-propanol (8%). The increase in 1-propanol was particularly interesting

because the propane terminal C-H bond energy (100.4 kcal/mol) is similar to the C-H bond energy of ethane (101.1 kcal/mol).

[00181] To obtain an ethane-hydroxylating P450, the directed evolution of 9-10A was initiated. The mutations were targeted to the active site rather than the entire heme domain. Most of these targeted substitutions were not accessible by random point mutagenesis due to the conservative nature of the genetic code. The BM-3 active site consists of a hydrophobic channel located directly above the heme. Guided by the high-resolution crystal structure of P450 BM-3 with bound palmitoylglycine substrate eleven amino acid residues in this channel that lie within five Å of the terminal eight carbons of the bound substrate were chosen for mutagenesis. These residues were targeted because they are likely to contact small alkanes in the active site during catalysis. Saturation mutagenesis libraries constructed for each of these residues were screened for mutants showing improved activity towards dimethyl ether. Single mutants selected from these libraries were then recombined to form a large library containing all possible combinations of the beneficial active-site mutations, and this library was screened for activity towards dimethyl ether. The mutant with the highest activity on propane, 53-5H, was isolated and purified and shown to catalyse thousands of turnovers of propane to propanol at a rate of 370 min⁻¹ (see Table 1).

[00182] Notably, P450 BM-3 mutant 53-5H also hydroxylates ethane to generate ethanol (see Table 1). 53-5H contains three active-site mutations, A78F, A82S, and A328F, all of which replace alanine with a larger side chain which reduce the volume of the active site and position small alkanes above the heme during catalysis. In addition to its activity towards ethane, 53-5H exhibits the highest regioselectivity (89% 2-octanol) and enantioselectivity (65% S-2-octanol) towards octane yet encountered in any BM-3 variant (see Table 1),

further evidence of tighter substrate binding in the engineered active site.

[00183] Ethane hydroxylation by 53-5H was measured after 12-hour incubations of the enzyme in ethane-saturated buffer containing an ethanol-free NADPH regeneration system (commercial NADPH contains 2-3% ethanol). While the reaction is uncoupled to NADPH oxidation in the presence of ethane (660 min^{-1} , see Table 2), 53-5H nevertheless consistently produces at least 50 equivalents of ethanol independent of the starting concentration of enzyme (see Figures 2 and 3). Overoxidation was tested by supplying ethanol as a substrate and monitoring ethanol depletion by gas chromatography and the production of acetaldehyde and acetic acid by ^{13}C NMR using ^{13}C -labeled ethanol. No (i.e. less than 10%, the detection limit of these experiments) overoxidation products were detected in these experiments (see Figure 4). In general, the hydrophobic active sites in P450s assist in the release of singly-hydroxylated products. Alkane-hydroxylating mutants overoxidize longer-chain alkanes such as decane to a higher extent (< 7%) than shorter ones such as hexane (< 1%).¹⁸ Mutant 53-5H also catalyses the rapid (660 min^{-1}) and efficient (80% coupled) conversion of octane to primarily 2-octanol in the presence of 1% ethanol, demonstrating that ethanol does not inhibit the catalyst. These results demonstrate that the 101.1 kcal/mol C-H bond dissociation energy of ethane does not pose a fundamental barrier to a cytochrome P450 and that a P450-based biocatalyst can cleanly convert ethane into ethanol without undesired side reactions.

[00184] A further round of directed evolution was performed to increase the ethane activity of 53-5H, this time targeting mutations to the reductase domain, including the polypeptide linker that connects it to the heme domain. The amino acid substitution E464G in the linker region increases total turnover in selected BM-3 mutants. Alone this mutation does

not enhance the production of ethanol by 53-5H, but further improvement was found upon screening a library of 53-5H containing this mutation and random mutations in the reductase domain for high activity towards dimethyl ether and accompanied by reduced NADPH consumption rates in the absence of substrate. The resulting mutant 35-E11 contains all 15 of the hydroxylase domain substitutions found in 53-5H, the linker mutation E464G, and a new mutation, leading to the substitution I710T, located in the reductase domain. Compared to 53-5H, mutant 35-E11 exhibits a five-fold increase in total turnover of ethane to ethanol (total turnover number (ttn) = 250). In a typical reaction with ethane, 200 nM of mutant 35-E11 produces over 50 μ M of ethanol (Figure 3). The rate of product formation by 35-E11 equals that of 53-5H, while the NADPH consumption rate in the presence of ethane has dropped to $\sim 520 \text{ min}^{-1}$ (Table 2). The increased productivity likely reflects a prolonged catalyst lifetime, achieved by reducing non-productive cofactor oxidation, which inactivates the protein by forming various reactive species. Amino acid residue I710 by comparison to the crystal structure of the homologous rat P450 reductase, is located near the FAD cofactor.

[00185] Accordingly, provided herein are variant P450s that have been adapted to hydroxylate ever smaller substrates. For example, the wildtype cytochrome P450 BM-3 enzyme is inactive on propane. Evolving it to become an octane hydroxylase, however, generated a small amount of activity towards propane. Similarly, increasing the newly acquired propane activity produced measurable activity towards ethane. Directed evolution has generated a biocatalyst with activity towards ethane. Further, as provided below, additional rounds of directed evolution has generated a BM-3 variant with the ability to convert methane to methanol. This enzyme provides

the foundation for a novel biocatalytic route to methanol production from methane.

[00186] Construction of Variants: Unless noted otherwise, chemicals were purchased from Sigma-Aldrich (St. Louis, MO). Ethane (99.95%), propane and dimethyl ether (DME) were purchased from Special Gas Services, Inc. (Warren, NJ). Ethanol was purchased from AAPER (Shelbyville, KY). NADP⁺ and NADPH were purchased from Roche (Indianapolis, IN) or Biocatalytics, Inc. (Pasadena, CA). Hexanes (99.94 %) were purchased from EMD (Gibbstown, NJ). Restriction enzymes were purchased from Roche (Indianapolis, IN) or New England Biolabs (Beverly, MA), Taq DNA polymerase from Roche (Indianapolis, IN), Pfu turbo DNA polymerase from Stratagene (La Jolla, CA), and T4 DNA ligase from Invitrogen (Carlsbad, CA).

[00187] Expression and Purification of P450 BM-3: The genes encoding P450 BM-3 mutants 53-5H (with amino acid substitutions R47C, V78F, A82S, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328F, L353V) and 35-E11 (with additional amino acid substitutions E464G and I710T) were expressed and purified as described (Peters et al., J. Am. Chem. Soc. 125:13442 (2003)). Enzyme concentrations were measured in triplicate from the CO-difference spectra.

[00188] Construction of saturation mutagenesis libraries: Mutations were introduced into mutant 9-10A (with amino acid substitutions R47C, V78A, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, L353V) by PCR overlap extension mutagenesis using Pfu turbo DNA polymerase (Stratagene). The primers for each saturation library contained all possible combinations of bases, NNN (N = A, T, G, or C), at the codon for a particular residue. The recitation of "for" indicates a "forward" primer. Accordingly, the primers in the forward direction for each library were:

74NNN-for (5'-GTCAANNNCTTAAATTTGCACG-3') (SEQ ID NO:19)
 75NNN-for (5'-GTCAAGCGNNNAATTTGCACG-3') (SEQ ID NO:20)
 78NNN-for (5'-GCTTAAATTTNNNCGTGATTTTGCAGG-3') (SEQ ID NO:21)
 81NNN-for (5'-CGTGATNNNGCAGGAGAC-3') (SEQ ID NO:22)
 82NNN-for (5'-CGTGATTTTNNNGGAGAC-3') (SEQ ID NO:23)
 87NNN-for (5'-GAGACGGGTTANNNACAAGCTGGAC-3') (SEQ ID NO:24)
 88NNN-for (5'-GGAGACGGGTTATTTNNNAGCTGGACG-3') (SEQ ID NO:25)
 260NNN-for (5'-CAAATTATTNNNTTCTTAATTGCGGGAC-3') (SEQ ID NO:26)
 263NNN-for (5'-ACATTCTTANNNGCGGGACACGAAAC-3') (SEQ ID NO:27)
 264NNN-for (5'-ACATTCTTAATTNNNGGACACGAAAC-3') (SEQ ID NO:28)
 328NNN-for (5'-CCAACCTNNNCCTGCGTTTCC-3') (SEQ ID NO:29)

[00189] The reverse primers for each of these libraries complement their corresponding forward primers. For each mutation, two separate PCRs were performed, each using a perfectly complementary primer, BamHI-forw (5'-GGAAACAGGATCCATCGATGC-3') (SEQ ID NO:30) and SacI-rev (5'-GTGAAGGAATACCGCCAAGC-3') (SEQ ID NO:31), at the end of the sequence and a mutagenic primer. The resulting two overlapping fragments that contain the mutations were then annealed in a second PCR to amplify the complete mutated gene. The full gene was then cut with BamHI and SacI restriction enzymes and ligated with T4 ligase (Invitrogen) into pBM-3-WT18-6, previously cut with BamHI and SacI to remove the wild-type gene. The ligation mixtures were then transformed into E. coli DH5alpha electrocompetent cells and plated onto Luria-Bertani (LB) agar plates to form single colonies for picking.

[00190] Construction of the active site recombination

library: Improved mutations that were picked up from the saturation mutagenesis libraries were recombined in a combinatorial fashion. Saturation libraries of amino acid positions 74, 81, 263 and 264 did not yield improved mutants so only beneficial mutations at the remaining seven sites were recombined. Mutations were added sequentially using overlap extension PCR with degenerate primers (see above) or a mixture

of degenerate primers (mutation sites in the primers are in bold letters). Starting with mutant 9-10A, mutations at amino acid position 78 and 82 were introduced using primers:

78A/T 82A/T/S/F/I/C/G/L/V-for

(5'GCTTAAATTCTRCGCGTGATTTTDBC GGAGACG) (SEQ ID NO:32), and

78F 82A/T/S/F/I/C/G/L/V-for

5'CTTAAATTCTTTCGTGATTTTDBC GGAGACG) (SEQ ID NO:33).

[00191] Next, mutations at position 260 were introduced using primers:

260T-for (5'CAAATTATTACATTCTTAATTGCGGGAC) (SEQ ID NO:34),

260N-for (5'CAAATTATTA ACTTCTTAATTGCGGGAC) (SEQ ID NO:35) and

260L-for (5'CAAATTATTCTTTTCTTAATTGCGGGAC) (SEQ ID NO:36).

[00192] Mutations at position 87 and 88 were introduced using primers:

87F/I/V/L 88T for (5'GAGACGGGTTANTYACAAGCTGGAC) (SEQ ID NO:37)

and 87F/I/V/L 88C for (5'GAGACGGGTTANTYTG TAGCTGGAC) (SEQ ID NO:38).

[00193] Mutations at position 328 were introduced using primers:

328A-for (5'CCA ACTGCTCCTGCGTTTTCC) (SEQ ID NO:39),

328L/F/V-for (5'CCA ACTBTTCTGCGTTTTCC) (SEQ ID NO:40), and

328M-for (5'CCA ACTATGCCTGCGTTTTCC) (SEQ ID NO:41)

[00194] Finally, mutations at position 75 were introduced using primers:

75L/I-for (5'CTTAAGTCAAGCGMTTAAATTC) (SEQ ID NO:42), and

75W-for (5'CTTAAGTCAAGCGTGGAATTC) (SEQ ID NO:43).

[00195] Sequences of resulting ethane hydroxylating mutant

53-5H: Nucleic acid mutations compared to wild-type P450 BM-3 include C142T, T234C, G235T, A237T, G247A, C248G, A249C, A284T, C324T, C427T, C527T, C554T, T617G, C681G, T711G, A758G, C766A, C872T, G985T, C986T, C1060G, and A1119G. Amino acid mutations compared to wild-type P450 BM-3 include R47C, V78F, A82S, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328F, and L353V.

[00196] Construction of the reductase library: The mutation E465G was found by directed evolution of a different BM-3 variant not reported in this work. To make the reductase library of 53-5H containing this mutation, a random library of the reductase portion of the E465G BM-3 gene (1285 bp to 3147 bp) was PCR amplified using the primers SacI-for (5'-GAC TTTGAAGATCATA CAAACTACGAGCTCG-3') (SEQ ID NO:44) and EcoRI-rev (5'-AGATCTGCTCATGTTTGACAGCTTATCATCGATGAATTC-3') (SEQ ID NO:45), with various concentrations of MnCl₂ (0, 50, 100, and 150 µM) added. The amplified DNA was restriction digested using SacI and EcoRI and then ligated with T4 ligase into the pCWori plasmid containing the gene for 53-5H with its reductase portion removed with SacI and EcoRI. A small library (~100 mutants) from each PCR condition was tested for mutation rate. The 100 µM MnCl₂ condition produced the optimal mutation rate of 1-2 mutations per gene, and a larger library containing 2610 members from this condition was picked and screened.

[00197] Nucleic acid sequences of resulting ethane hydroxylating mutant 35-E11 (positions given with regard to the coding sequence of wild-type P450 BM-3 nucleic acid sequence) include C142T, T234C, G235T, A237T, G247A, C248G, A249C, A284T, C324T, C427T, C527T, C554T, T617G, C681G, T711G, A758G, C766A, C872T, G985T, C986T, C1060G, A1119G, A1394G, T2132C, and A2313G. Amino acid sequence mutations compared to wild-type P450 BM-3 include R47C, V78F, A82S, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328F, L353V, E464G, I710T.

[00198] Cell lysates for high throughput screening were prepared as described previously (Peters et al., J. Am. Chem. Soc. 125:13442 (2003)). High-throughput dimethyl ether (DME) screens were performed in 96 well plates as described previously. Improved mutants were selected based on an increased signal compared to 9-10A (for the saturation

mutagenesis libraries) and 1-12G (for the recombination library). Reactions on octane and propane were analyzed by gas chromatography as described.

[00199] Reactions on ethane were carried out similarly to those described for propane (Peters et al., J. Am. Chem. Soc. 125:13442 (2003)). Because commercially-available NADPH contains 2-3% ethanol, an NADPH regeneration system was used. Thereby, the background ethanol concentration was reduced to approximately 1 μ M. Ethane hydroxylation reactions were performed in 20 mL glass vials. A 5.0 mL reaction mixture contained 100, 200 or 500 nM purified protein in 0.1 M potassium phosphate buffer (pH 8.0) saturated with ethane and dioxygen. 300 μ L of the regeneration system (1.66 mM NADP⁺, 167 U/mL isocitrate dehydrogenase, 416 mM isocitrate) was added. The vial was immediately topped with a septum, pressurized with 20 psi of ethane and stirred at 25°C. For determination of total turnover, the reaction was allowed to proceed for four hours. For rate determination the reaction was stopped at 0, 15 and 30 min. 50 μ L of 5 mM 1-butanol was added to the reaction mixture as an internal standard. Ethanol and 1-butanol were subsequently derivatized to their corresponding alkyl nitrites. Control reactions were performed by repeating these steps without substrate and with inactivated protein, to correct for background levels of ethanol. Reaction buffer saturated with ethane and oxygen did not contain ethanol concentrations above this background. Reactions with ethane and corresponding control reactions without ethane or with inactivated protein were carried out in triplicate. Taking into account the background ethanol concentration of around 1 μ M (1.2 ± 0.1 μ M for the control reaction without ethane or 0.8 ± 0.1 μ M for the control reaction with inactivated P450), the reactions yielded final ethanol concentrations corresponding to total turnover numbers

of 50 and 250 for 53-5H and 35-E11, respectively (see e.g., Figures 2 and 3).

[00200] Gas chromatographic analysis of the ethanol was performed using purchased standards and a five-point calibration curve with 1-butanol as an internal standard. The samples were injected at a volume of 1 uL and analyses were performed at least in triplicate. Ethanol analysis was performed on a Hewlett-Packard 5890 Series II Plus gas chromatograph using an HP-1 column (Agilent Technologies, 30m length, 0.53 mm ID, 2.65 uM film thickness) connected to an electron capture detector (ECD). The temperature program was: 60°C injector, 150°C detector, 35° oven for 3 min, 10°C/min gradient to 60°C, 60°C gradient to 200°C, and then 200°C for 5 min.

[00201] In addition, overoxidation of ethanol by 53-5H was tested. First, ethanol (20 µM), instead of ethane was used as the substrate under the reaction conditions described above, and no decrease in ethanol concentration was observed within experimental error (± 3 uM) over the course of the reaction. Second, NMR (Varian Mercury 300 MHz) spectroscopy was used to monitor reactions of 53-5H with singly ^{13}C -labeled ethanol ($\text{HO-}^{13}\text{CH}_2\text{CH}_3$, 57 ppm). Concentrations of ^{13}C -labeled ethanol as low as 50 µM in the presence of 10 µM P450 (in 0.1M potassium phosphate buffer, pH=8, with 10% deuterium oxide added) could be detected using the 13C channel of the NMR. The longer relaxation rates of the carbonyl carbon atoms in acetaldehyde (206 ppm) and acetic acid (176 ppm) decrease their carbon NMR signal heights by 25-40% in our experiments, making their detection limit approximately 100 µM. Reactions using 10 µM 53-5H, 1 mM ^{13}C -ethanol, 100 µM NADP+, 50 mM isocitrate, and 10 U/mL isocitrate dehydrogenase were stirred 12 hours at room temperature. Deuterium oxide (10% final concentration) was then added to each reaction, and their NMR spectra (averaging ~20,000 scans each) were recorded. Control reactions using 50

nM of 53-5H from the same batch, 4 mM of octane, 100 μ M NADP⁺, 50 mM isocitrate, and 10 U/mL isocitrate dehydrogenase were also performed and the products characterized to verify that the P450 was fully active. No overoxidation products were detected in these experiments, indicating that, in the presence of 1 mM ethanol, 53-5H catalyzes less than 10 total turnovers of ethanol (see Figure 4).

[00202] The absorption spectra for 53-5H and 35-E11 were taken using a 3 μ M solution of purified P450 in a quartz cuvette. Upon incubation with ethane, no shift in equilibrium from low- (A_{max} at 418 nm) to high-spin (A_{max} at 390 nm) configuration of the ferric P450 was observed. Octane, a good substrate, also did not induce a shift. Mutant 35-E11 behaved identically (see Figure 5). Previously presented mutants, e.g. 9-10A exhibit a significant spin shift upon incubation with good substrates, such as octane.

[00203] NADPH oxidation rates were measured over 20 s at 25°C using a BioSpec-1601 UV-Vis spectrophotometer (Shimadzu, Columbia, MD) and 1 cm pathlength cuvettes. A 1 mL reaction mixture contained 100 nM P450, 166 μ M NADPH and either methane-, ethane- or propane-saturated potassium phosphate buffer (0.1 M, pH 8.0) or 4 mM octane and 1% ethanol in potassium phosphate buffer (0.1M, pH 8.0). The reaction was initiated by the addition of NADPH and the decrease in absorption at 340 nm was monitored. NADPH consumption rates were also measured without substrate (see Table 2).

[00204] Octane reactions were performed in presence of 1% ethanol and the results shown in Table 1. Rates of ethanol formation were measured over 30 min using gas chromatography (GC) coupled to an electron capture detector (ECD) and are reported as nmol ethanol/min/nmol enzyme. Errors are at most 10%. Total turnover number (ttn) was measured using GC after completion of the reaction and is reported as nmol product/nmol protein. Errors are at most 10%. Initial rates

of propanol and octanol formation were measured over 1 min by GC and are reported as nmol product/min/nmol protein. Errors are at most 15%. The "ee" of 2-octanol (main product) was measured by GC; the favored enantiomer is listed in parenthesis. Wild-type P450 BM-3 primarily produces 3- and 4-octanol. The yields were not sufficient for the determination of ee for wt BM-3. Mutant 9-10A contains amino acid substitutions R47C, V78A K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, and L353V.

[00205] Table 1

Enzyme	Number of amino acid substitutions	Active site amino acid substitutions	Ethane		Propane		Octane			
			Rate	ttn	Rate	ttn	Rate	ttn	% 2-octanol	% ee
Wild-type BM-3	---	---	---	---	---	---	30	150	17	n.d
9-10A	13	V78A	---	---	23	1100	540	3000	82	50 (S)
53-5H	15	V78F, A82S, A328F	0.4	50	370	5000	660	8000	89	65 (S)
35-E11	17	V78F, A82S, A328F	0.4	250	210	6000	420	8000	89	65 (S)

[00206] Table 2 provides NADPH oxidation rates of 53-5H and 35-E11 in presence and absence of substrate. The coupling efficiency is estimated from product formation rate / NADPH oxidation rate. In general, NADPH oxidation rates cannot be measured on the same time scale as ethanol formation rates.

[00207] Table 2

	53-5H		35-E11	
	NADPH oxidation rate [min ⁻¹]	Coupling efficiency [%]	NADPH oxidation rate [min ⁻¹]	Coupling efficiency [%]
no substrate	540		390	
ethane	660	0.06	520	0.08
propane	930	40	800	26
octane	820	80	660	64

[00208] Example 2: Methane Conversion

[00209] World reserves of methane in natural gas are more abundant than oil. Methane in methane hydrates increases this by orders of magnitude. Methane is also emitted from coal beds and renewably produced by the anaerobic decomposition of organic matter in landfills and human and animal waste streams. Most of these sources, deemed "stranded" by the energy industries, are not used because they are either too remote to transport the gas to a power generating facility or too small to economically convert with existing technology into the more easily transported methanol. A practical catalyst for the conversion of methane to methanol could allow these sources of methane gas to be converted into easily transported liquid fuel. Accessing these reserves with a new technology will have a dramatic effect on both the energy and chemical commodity industries.

[00210] The wide scale production of methanol from these sources would also have a favorable impact on the concentration of greenhouse gases in the atmosphere. Methane has a 20 fold higher global warming potential than carbon dioxide. Human activities release about 300 MMT/year of methane into the atmosphere, accounting for about a quarter of the measured increase in global temperatures. Conversion of this wasted methane into methanol could have a dramatic effect on the concentrations of global warming gases in the atmosphere, since in addition to removing methane from the

atmosphere by this process, the methanol when burned as a fuel releases less ($\sim 1/2$) carbon dioxide than is released by burning an equivalent amount of gasoline. Converting methane released into the environment from human activities into a usable alternative fuel is therefore one of the most effective means of curbing global warming.

[00211] The engineered cytochromes P450 mutants described throughout the present disclosure were acquired by accumulating point mutations in directed evolution experiments. An alternative method for making libraries for directed evolution to obtain P450s with new or altered properties is recombination, or chimeragenesis, in which portions of homologous P450s are swapped to form functional chimeras. Recombining equivalent segments of homologous proteins generates variants in which every amino acid substitution has already proven to be successful in one of the parents. Therefore, the amino acid mutations made in this way are less disruptive, on average, than random mutations. A structure-based algorithm, such as SCHEMA, identifies fragments of proteins that can be recombined to minimize disruptive interactions that would prevent the protein from folding into its active form. SCHEMA has been used to design chimeras of P450 BM-3 and its homolog CYP102A2, sharing 63% amino acid sequence identity. Fourteen of the seventeen constructed hybrid proteins were able to fold correctly and incorporate the heme cofactor, as determined by CO difference spectra. Half of the chimeras had altered substrate specificities, while three mutants acquired activities towards a new substrate not hydroxylated by either parent. A large library of chimeras made by SCHEMA-guided recombination of BM-3 with its homologs CYP102A2 and CYP102A3 contains more than 3000 new properly folded variants of BM-3. In addition to creating new biocatalysts, these chimeragenesis studies demonstrate that the homologous P450 enzymes CYP102A2 and

CYP102A3 are similar enough to BM-3 to recombine in this fashion and still retain a high probability of folding.

[00212] Provided herein are variants of cytochrome P450 BM-3 that catalyze the direct conversion of methane to methanol. The reaction uses dioxygen from the air and the reduced form of the biological cofactor nicotinamide adenine dinucleotide phosphate (NADPH). The stoichiometry of the reactions catalyzed by the BM-3 variant is $\text{CH}_4 + \text{O}_2 + \text{NADPH} + \text{H}^+ \rightarrow \text{CH}_3\text{OH} + \text{H}_2\text{O} + \text{NADP}^+$.

[00213] A "variant" cytochrome P450 contains one or more amino acid substitutions, insertions or deletions relative to its parent cytochrome P450. The term "variant" is often used interchangeably with "mutant". A cytochrome P450 variant is suitably prepared on the basis of a parent cytochrome P450 having one of the sequences shown, for example, in SEQ ID NO: 1, 2 or 3. The parent cytochromes P450 shown in SEQ NO: 2 and 3 are obtainable from *Bacillus subtilis* strain 1A1.15. These cytochromes P450 are known as CYP102A2 and CYP102A3, respectively. The cytochrome P450 of SEQ ID NO: 1 is known as the cytochrome P450 BM-3, from *Bacillus megaterium*. It is also known in the CYP nomenclature as CYP102A1. The parent cytochromes P450 shown in SEQ ID NO: 4, SEQ ID NO: 5, and SEQ ID NO: 6 are obtainable from *Bacillus cereus* (CYP102A4), *Ralstonia metallidurans* (CYP102E1) and *Bradyrhizobium japonicum* (CYP102A6), respectively. In Figure 6A - 6F the % identity between the parent sequences and CYP102A1 are reported from alignments made with standard in the art software. As shown in Figure 6A - 6F, the identity in the heme domain is often higher than reported because of large differences in the linker region (~residue 470) between the heme and reductase domains.

[00214] These six cytochromes P450 are closely related to one another and show a high degree of sequence identity. The sequences can be aligned based on the sequence homology, and

that alignment is shown in Figure 6A - 6F. This alignment defines "equivalent positions" in the sequences. An equivalent position denotes a position which, on the basis of the alignment of the sequence of the parent cytochrome P450 in question with the "reference" cytochrome P450 amino acid sequence in question (e.g. SEQ ID NO: 1) so as to achieve juxtapositioning of amino acid residues which are common to both, corresponds most closely to a particular position in the reference sequence in question. This process can cause gaps or insertions to appear in the sequences. In the alignment of Figure 6A - 6F, equivalent positions are shown lined up vertically with one another. For example, position 87 in SEQ ID NO: 1 is equivalent to position 88 in SEQ ID NO: 2 and 3 and position 89 in SEQ ID NO: 4.

[00215] The term "total turnover number" (TTN) is the total number of substrate molecules converted to product (or turned over) by a single enzyme over its lifetime or during a specified time period. TTN is an important figure of merit for a catalyst because it allows for the calculation of the total amount of product that can be made from a given quantity of catalyst. The term "coupling" means the ratio, in percent, of product formed to cofactor (NAD(P)H) consumed during the enzyme-catalyzed reaction. If 2 moles of cofactor are consumed for every mole of product made, then the coupling for the reaction is 50%. Coupling is also an important figure of merit because it allows one to assess how much cofactor is required to produce a given amount of product.

[00216] The BM-3-based catalyst functions at ambient temperature and pressure. Unlike methane monooxygenase enzymes, this BM-3 based biocatalyst can be expressed in host organisms that do not consume the methanol as it is produced. Thus it may be incorporated into a whole-cell biocatalysis system that can accumulate methanol in the presence of methane and air.

[00217] Methane hydroxylase variants were generated by an 'evolutionary' strategy in which mutations were accumulated in multiple generations of random mutagenesis and screening, in essence adapting the enzyme to exhibit higher turnover on smaller and smaller alkanes. In the first generations, the activity of BM-3 towards octane was increased and a small but appreciable activity on propane was acquired. Further evolution led to a variant which is much more active towards propane and, significantly, produced more 1-propanol. The increase in the terminal hydroxylation product was particularly interesting because the propane terminal C-H bond energy (100.4 kcal/mol) is similar to the C-H bond energy of ethane (101.1 kcal/mol).

[00218] Continued directed evolution then led to the first ethane-hydroxylating P450 BM-3 as described above. The ethane-hydroxylating P450 BM-3 variant 35-E11 was obtained by improving the electron transfer in the triple active site variant 53-5H as described. Briefly, the active site mutations in 53-5H (V78F, A82S, and A328F) were obtained by screening a library containing all possible combinations of single-site mutations of a BM-3 variant previously evolved for high activity towards propane. The 15 mutations in 53-5H (12 are not in the active site) are located in the heme domain: R47C, V78F, A82S, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, A328F, L353V. While not bound to a particular mechanism of action, it is believed that the active site mutations serve to decrease the size of the active site so that ethane is bound and hydroxylated. Mutant 53-5H has only limited activity towards ethane, producing only about 50 total turnovers before it is inactivated. Variant 35-E11 was obtained by modifying the reductase domain of 53-5H to more efficiently transfer electrons to the active site during catalysis. This variant contains two mutations, one in the linker region connecting the two domains (E464G) and one in

the reductase domain (I710T), that improve coupling and increase the total turnovers on ethane to more than 250.

[00219] The 35-E11 variant was tested for the ability to convert methane to methanol using the method of Anthon and Barrett (Agric. Food Chem., 52:3749 (2004)). The method was optimized so that it could detect methanol concentrations in water as low as 10 μ M. Using this assay the 35-E11 variant was identified as producing methanol from methane.

[00220] Libraries of mutants of 35-E11 were screened for phenotypes indicative of either increased product formation on small gaseous substrates slightly larger than methane. Additionally, they were screened for increased coupling of the consumption of NADPH reducing equivalents to product formation. Product formation was determined by measuring total amount of dimethyl ether hydroxylation while increased coupling was determined primarily by decreased background NADPH consumption rates (measured spectrophotometrically at 340 nm in the absence of substrate).

[00221] Four libraries of mutants of 35-E11 were constructed. Two saturation mutagenesis libraries were prepared in 35-E11, in which all possible amino acid substitutions were introduced at the positions of two non-heme domain mutations, E464G and I710T. Additionally, two random mutagenesis libraries of 35-E11 were prepared by error-prone PCR such that each library contained an average of 1-2 amino acid mutations. One of these random mutagenesis libraries contained only mutations in the heme domain (residues 1 to about 455) while the other library contained mutations only in the linker between the two domains and the reductase domain (residues about 470 to 1048). These four libraries were screened as described above, and several improved variants (described below) were isolated, sequenced, and purified for further analysis.

[00222] The mutant BM-3 enzymes described below were purified using published procedures and used as catalysts in small reactions with methane. In glass head space vials containing magnetic stirring bars, BM-3 catalyst was added to 0.1 M potassium phosphate buffer (pH=8) saturated with methane. The vials were then crimped shut and charged to 20 psi with methane. To start the biotransformation, an NADPH regeneration system consisting of NADP⁺, isocitrate, and isocitric dehydrogenase was added through a syringe. The reaction was allowed to stir 12-15 hours. Methanol production was measured using a coupled enzyme assay. Aliquots of the reaction were transferred to 96-well microtitre plates and reacted with the enzyme alcohol oxidase for 5 minutes at room temperature. Alcohol oxidase converts the methanol produced by the BM-3 variants into formaldehyde which can be detected with the dye Purpald. To detect formaldehyde and quench the alcohol oxidase reaction, a solution of Purpald in 2 M NaOH was added and allowed to react for one hour. The purple color that Purpald forms in the presence of formaldehyde was quantified one hour later by spectrophotometrically reading the well at 550 nm and comparing the absorbance against similarly-treated solutions of BM-3, NADPH regeneration system, and methane-free buffer spiked with known concentrations of methanol. The results of these biotransformations are reported in Table 3 and sample spectra from the colorimetric assay are shown in Figure 7.

[00223] As expected from their relative increase in activity towards the screening substrates, several of the mutants isolated from these libraries have increased activity towards methane. Three of these mutants were isolated from the saturation library of the linker position 464: E464R, E464Y, and E464T. None of these mutants exhibited higher activity towards the screening substrate DME, but all of them had a 30-40% lower NADPH background consumption rate in the screen,

indicating that the roughly equivalent production of oxidized DME they produced compared to the parent 35-E11 required less NADPH. This increase in coupling on DME translated to an increase in TTN of methane (from 8 to 11-13). From the heme domain library, mutant 20-D3 was identified for its increased activity towards DME. Even though the background consumption rate of NADPH was increased by an equivalent amount in the screen, more of these reducing equivalents are transferred to the heme domain of this mutant in the presence of substrate, leading to an increase in methanol formation (from 8 to 10 TTN). This mutation, H285R, is located on the surface of the heme domain, near the putative interface between the heme and reductase domains. From the reductase domain library, two mutants, 23-1D and 21-4G, were identified for their decreased background NADPH consumption rates, leading to an increase in methanol formation. The mutations in these variants, Q645R/P968L and D631N, respectively, are located outside of the electron transfer pathway. Their presence, however, gives 35-E11 increased TTN on methane (from 8 to 10-11).

[00224] The methane hydroxylation activity reported here is not sufficiently high for a practical process for converting methane to methanol. TTN must be increased, and coupling of methanol production to consumption of the cofactor must be increased, or the enzyme will use too much. However, the methods of making mutations and screening the libraries described here can be used to further improve the activity, TTN and coupling.

[00225] Engineering Other Cytochromes P450 to Convert Methane to Methanol. Provided herein are novel modified cytochrome P450 enzymes capable of hydroxylating methane to produce methanol. Because the heme domain is capable of this reaction, protein engineering of cytochrome P450s from other sources can be expected to lead to a similar result. It is well known in the art that amino acid substitutions having a

particular effect (e.g. that confer activity towards a new substrate) can have the same effect in closely related proteins. Three close homologs of cytochrome P450 BM-3, also known as CYP102A1, are known: CYP102A2 and CYP102A3 and the P450 from *B. cereus*, CYP102A4.^{34,35} The alignment of these four homologs and 35-E11 shown in Figures 6A - 6F illustrate the high degree of sequence similarity among the four cytochromes P450, particularly in the heme domain. It has been shown on multiple occasions that amino acid substitutions at equivalent positions in these enzymes have equivalent effects on function. For example, the substitution of F87 by A in CYP102A1 increases the peroxygenase activity of the enzyme.^{18,19,37} The same substitution of the equivalent position in CYP102A2 and CYP102A3, which is F88A, has the same effect: it increases the peroxygenase.³¹ Appel et al. showed that the substitutions A74G, F87A and L188Q in CYP102A1 together broaden the enzyme's substrate range, so that it hydroxylates octane and naphthalene.²¹ Lentz et al. demonstrated that substituting the equivalent positions in CYP102A3 (no substitution was required at A74G because there was a G at this position already), F88 to V and S188 to Q, had a similar effect of conferring the ability to hydroxylate octane and naphthalene.³⁸ Thus on the basis of this experience, it is reasonable to assume that substitutions equivalent to those described here in CYP102A2 as conferring methane hydroxylation activity will have the same effect in CYP102A2, CYP102A3 as well as other cytochromes P450 that show high identity (55% or greater) to CYP102A such as the one from *B. cereus*. Additionally, these P450 enzymes can be subjected to rounds of directed evolution using the techniques and screens described herein to obtain and/or increase methane hydroxylation activity. Less closely-related cytochromes P450 with similar domain architecture to BM-3 have also been reported in the literature, including P450s from the unrelated

β -proteobacterium *Ralstonia metallidurans* (CYP102E1) and the gram negative bacterium *Bradyrhizobium japonicum* (CYP102A6), and could potentially be similarly engineered for methane oxidation (see Figures 6A - 6F and Table 4 for corresponding mutations in these two P450s).

[00226] Whole Cell Processes for the Conversion of Methane to Methanol: The cytochrome P450 methane hydroxylase is expressed in its functional form in *Escherichia coli* and the whole cell conversion of methane to methanol using optimized versions of these enzymes in *E. coli* will be possible. As demonstrated in the examples provided herein, *Escherichia coli* over expressing BM-3 enzymes will act as whole cell biocatalysts with relative activities equivalent to the activities of the individual BM-3 catalyst. Substrates used for these whole cell experiments include alkanes, alkenes, and heterocyclic compounds. In the case of methane hydroxylation, the maintenance of the gaseous substrate levels in cultures of whole cells will complicate the process and constitute a large portion of future work with this system. Further, the metabolism of host organisms can be engineered to further increase the levels of NADPH (or NADH for BM-3 variants that accept NADH) available to biocatalysts provided herein.

[00227] Examples of whole cell systems that could be used to convert methane to methanol will now be described. The alkane content of natural gas generally comprises about 80% methane and 20% other small alkanes, primarily ethane and propane. Naturally occurring strains of gram negative (e.g. *Pseudomonas*) and gram positive (e.g. *Rhodococcus*), and even engineered laboratory strains of *E. coli*, can grow using small alcohols as their sole carbon source. Expression of optimized BM-3 based biocatalysts in one of these bacterial strains will allow the bacteria to use the alkane impurities in natural gas as their sole source of carbon and energy (and cofactor regeneration) while converting the methane into methanol.

Expression of our BM-3 enzymes in any of these bacteria can be done by standard in the art techniques using multi-host plasmids and shuttle vectors.

[00228] Methane generated by the decomposition of biomass in landfills and agricultural waste does not contain other alkanes. To utilize these methane sources with a whole cell system containing our BM-3 biocatalysts, an external carbon and energy source such as glucose must be provided to the host cells to regenerate NADPH. Glucose or similar compounds (with greater NADPH-producing potentials) can be added directly to these whole cell reactions or generated from biomass present in the methane-producing waste. The amount of cofactor a host organism generates from these carbon sources can also be increased by streamlining the cell strain for methane production. Expression of optimized BM-3 enzymes in these bacteria can be done by standard in the art techniques using strain specific plasmids. Modification of *E. coli* strains to optimize NADPH availability can also be performed using standard in the art metabolic engineering techniques such as chromosomal insertion of phage-delivered DNA and recombination of homologous gene fragments from plasmids.

[00229] Alternatively, methylotrophs, i.e. methanol-metabolizing yeast (e.g. strains of *Pichia*) and bacterial cell strains (e.g. strains of *Methylobacterium*), can serve as hosts for BM-3 biocatalysts. In this scenario, the engineered BM-3 enzymes supply all of the methanol the hosts need to grow and live while generating excess methanol for production. These cell strains could also be modified genetically to optimize the amount of methanol they can produce. Additionally, since the growth of these methylotrophic strains are dependent on methanol, these strains can be used to identify new enzyme variants with improved activity towards methane. Expression of BM-3 in *Pichia* is possible with standard in the art *Pichia*-specific gene expression plasmids. Expression in

methylootrophs is possible with standard in the art techniques using multi-host shuttle vectors.

[00230] The above examples utilize liquid cultures of specific microorganisms for the whole cell bioconversion of methane to methanol, but other host organisms including but not limited to other bacteria, other yeast, and even plants, algae, and fungi could also be used. Plants, algae, and even photosynthesizing bacteria such as cyanobacteria that can convert carbon dioxide and light into sugars (and then NAD(P)H) would be especially valuable as hosts for a BM-3 methane oxidation biocatalyst since these organisms would not require any carbon source other than carbon dioxide. Because methane streams derived from biomass contain about 50% carbon dioxide, these hosts are ideal for methane to methanol processes using landfill and waste generated methane. Expression of BM-3 in plant and fungi is possible with standard in the art DNA transfer techniques specific to these types of cells, such as the use of viral vectors for plants.

[00231] A Process for Bioconversion of Methane to Methanol Using Cytochromes P450. Provided herein are novel cytochrome P450-based variants that are capable of ethane and/or methane hydroxylation. Accordingly, processes that use such catalysts for the bioconversion of ethane into ethanol or methane into methanol can be performed in procedures described below. However, it is understood that the following procedures are exemplary only. The novel enzymes of the invention allow those skilled in the art of bioconversion to modify the exemplary procedures to suit their particular needs.

[00232] First, in a large vessel, cell mass is generated using a carbon source such as glucose and nitrogen and other nutrients supplied to the cells. At an optimum cell density, expression of the P450 biocatalyst is induced using strategies well known in the art. After optimum biocatalyst expression levels are reached, nitrogen is removed from the medium to

prevent further growth of the cell, ensuring that any additional carbon source supplied to the non-growing cells is converted into reducing equivalents for methanol production. Methane and oxygen are then supplied to the culture at saturating levels assisted by mixing with either propeller-type mixing or air-lift mixing. As methanol is accumulated in the culture, portions of the culture are separated from the main reactor, centrifuged to remove cells (which are added back to the main reactor), and the supernatant treated to recover the methanol. The process can be constructed as either a batch or continuous process.

[00233] Specifically, the first demonstration of whole cell methanol production using a catalyst based on BM-3 or one of its homologs expressed in a laboratory strain of *E. coli* will occur in a small fermentor where oxygen and glucose levels can be easily monitored and adjusted. Oxygen levels are particularly important, because in the absence of oxygen, the facultative aerobe *E. coli* will begin to degrade glucose into ethanol and carbon dioxide. In our experience this causes the cell to up-regulate the production of alcohol dehydrogenases that will consume the trace levels of methanol produced by our catalyst. Additionally, the ethanol that accumulates from anaerobic metabolism interferes with our methanol detection assays. To perform the whole cell reaction in the fermentor, a 1 L culture of cells will be grown to high density ($OD_{600} = 1-2$) in Terrific Broth containing 1 ug/mL ampicillin (for retention of the PCWori plasmid containing the biocatalyst gene). Catalyst expression will be induced by the addition of 1 mM IPTG and 0.5 mM δ -amino levulinic acid. After 12-24 hours, the culture will be centrifuged to collect the cells containing biocatalyst. After removal of the supernatant, the cells will be resuspended and washed in a nitrogen free minimal medium to remove traces of the terrific broth. The cells in nitrogen free minimal medium will then be added to a

fermentor and fed 0.5% glucose. During the reaction, additional glucose will be added to maintain this level. In the fermentor, oxygen will be added to the culture at saturating levels and will be monitored and added during the reaction as needed to maintain these levels. To initiate the reaction, methane will be bubbled through the culture. The lack of nitrogen in the culture forces the cells into a non-growing state, allowing most of the glucose aerobically consumed to be converted into NADPH for the biocatalyst.³⁹ An *E. coli* cell functioning as an aerobe also has a limited ability to metabolize alcohols, including the methanol product.⁴⁰ Methanol will be measured in this culture as described for the *in vitro* reactions in this work.

[00234] The present data demonstrates that methane hydroxylation activity can be achieved through modification of various cytochrome P450 enzymes by directed evolution. Since this enzyme can be expressed in heterologous hosts that do not metabolize methanol as it is produced, the invention of this biocatalyst also makes the production of methanol using a whole cell process possible.

[00235] All reagents and enzymes except for NADPH were purchased from Sigma-Aldrich (St. Louis, MO). NADPH was purchased from BioCatalytics, Inc. (Pasadena, CA). Protein purification, library construction, and screening methods were performed by standard techniques known to the skilled artisan. Also, as previously published, a PCWori plasmid containing an ampicillin resistance gene and a double Tac promoter for the biocatalyst gene was used in all of this work.¹¹

[00236] Reactions for enzyme-catalyzed conversion of methane to methanol were carried out in 10 mL glass headspace vials. A 470 μ L reaction mixture containing 5 μ M purified BM-3 variant in 0.1 M potassium phosphate buffer (pH 8.0) was first added to the vial. The vial was immediately topped with a septum, pressurized with 30 psi of methane and shaken at 4°C for 1

hour. Then 30 uL of an NADPH regeneration system (1.66 mM NADP⁺, 167 U/mL isocitric dehydrogenase, 416 mM isocitrate) was added to initiate the reaction. For determination of total turnover, the reaction was allowed to proceed for 20 hours at room temperature.

[00237] The amount of methanol product formed during the 20 hr methane reaction was quantified colorimetrically with the use of alcohol oxidase and Purpald. 190 µL of the reaction mixture is transferred onto a single cell of a 96 well plate. 10 µL of 0.05U/µL solution of alcohol oxidase (from *Pichia pastoris*, purchased from Sigma-Aldrich) was added to cell and allowed to incubate at room temperature for 5 minutes. During this time, the alcohol oxidase converts the methanol into formaldehyde. Purpald (168 mM in 2 M NaOH) was then added to form a purple product with the formaldehyde in solution. The purple color was read approximately 1 hour later at 550 nm using a Spectramax Plus microtiter plate reader. The concentration of methanol was determined based on the measure absorbance in comparison to a methanol calibration curve. The methanol calibration curve was generated with several standards. Each standard contained 100 µL of a methanol solution (10, 20, 40, 60, 100 µM) and 90 µL of the methane reaction mixture described above, without the methane pressurization.

[00238] Table 3 provides amino acid substitution, screening and activity data for second-generation 35-E11 variants exhibiting methane hydroxylation activity. The relative dimethyl ether (DME) activity was measured, 1 hour after treatment with Purpald, as the average net and absorbance at 550 nm from 8 microtitre plate DME reactions of the mutant compared to 8 equivalent reactions of 35-E11. All values were normalized by measuring the P450 concentration in each well with CO binding. "Asteriks" indicate criteria used to select mutants from libraries for further characterization. The

oxidation rate is measured as absorbance change at 340 nm over 1 minute upon the addition of NADPH to 8 microtitre plate wells containing P450-containing cell lysate and compared to 8 equivalent reactions of 35-E11. All values were normalized by measuring the P450 concentration in each well with CO binding. The total turnover number (TTN) was measured as the absorbance at 550 nm of methane reactions with each mutant after treatment with alcohol oxidase and Purpald.

[00239] Table 3

Mutant	AA substitutions to 35-E11	Relative DME activity in screen	Relative background NADPH oxidation rate	Methane TTN
35-E11	-	1	1	8
35-E11-E464R	E464R	0.9	0.6*	12
35-E11-E464Y	E464Y	0.8	0.7*	13
35-E11-E464T	E464T	0.8	0.7*	11
20-D3	H285R	1.7*	1.6	10
23-1D	Q645R, P968L	0.7	0.8*	11
21-4G	D631N	1.1	0.7*	10

[00240] Table 4 provides an extensive list of amino acid substitutions in various parental P450 sequences (e.g., SEQ ID NO: 1, 2, 3, 4, 5, and 6).

[00241] Table 4

Mutation	Seq. ID No. 1	Seq. ID No. 2	Seq. ID No. 3	Seq. ID No. 4	Seq. ID No. 5	Seq. ID No. 6
Heme Domain Mutation in 35-E11	R47C	G48C	V48C	D49C	K54C	S51C
Active Site Mutation in 35-E11	V78F	V79F	V79F	V80F	L85F	V82F
Active Site Mutation in 35-E11	A82S	A83S	S83S	G84S	A89S	G86S
Heme Domain Mutation in 35-E11	K94I	P95I	P95I	P96I	P101I	P98I
Heme Domain Mutation in 35-E11	P142S	P143S	A143S	P144S	A149S	V146S
Heme Domain Mutation in 35-E11	T175I	N176I	T176I	T177I	M182I	E179I
Heme Domain Mutation in 35-E11	A184V	A185V	A185V	A186V	A191V	T188V
Heme Domain Mutation in 35-E11	F205C	F206C	F206C	F207C	A212C	L208C
Heme Domain Mutation in 35-E11	S226R	N227R	N227R	S228R	V232R	E231R
Heme Domain Mutation in 35-E11	H236Q	R238Q	L238Q	R239Q	L243Q	A242Q
Heme Domain Mutation in 35-E11	E252G	E254G	E254G	E255G	A259G	V258G
Heme Domain Mutation in 35-E11	R255S	R257S	R257S	R258S	R262S	R262S
Heme Domain Mutation in 20-D3	H285R	D287R	E287R	D288R	G292R	D291R
Heme Domain Mutation in 35-E11	A290V	Y292V	Q292V	Y293V	Y297V	Y296V
Active Site Mutation in 35-E11	A328F	A330F	A330F	A331F	A336F	A337F
Heme Domain Mutation in 35-E11	L353V	I355V	V355V	I357V	I361V	I363V
Linker Mutation in 35-E11 and Derivatives	E464G/R/Y/ T	E475G/R/Y/ T	D467G/R/Y/ T	I477 G/R/Y/T	S472G/R/Y/ T	S480G/R/Y/ T
Reductase Domain Mutation in 21-4G	D631N	E641N	N634N	E644N	T640N	D648N
Reductase Domain Mutation in 23-1D	Q645R	Q656R	T648R	Q659R	R656R	E664R
Reductase Domain Mutation in 35-E11	I710T	N721T	L713T	N724T	L723T	L733T
Reductase Domain Mutation in 23-1D	P968L	E980L	E972L	E983L	P985L	D997L

[00242] Example 3: Regiospecific Activity

[00243] Biocatalysis provides a useful alternative to classic chemical synthesis due to the inherent specificity of enzymes. Enantioselective catalysts that accept only a single enantiomer or regioisomer of a substrate have been created using directed evolution. The enzymatic, regio- and enantioselective functionalization of achiral compounds, however, is a greater challenge. In particular, the regio- and enantioselective hydroxylation of saturated hydrocarbons remains a problem in chemical synthesis.

[00244] As previously noted, cytochrome P450 BM-3 is a soluble protein that contains an hydroxylase domain and a reductase domain on a single polypeptide. Also as previously noted, cytochrome P450 BM-3 variants are provided that exhibit efficient alkane hydroxylase activity. One of these variants, 9-10A, was further modified at eleven active site residues that are in close contact to the substrate and screened for activity. Beneficial mutations were recombined, generating variants that catalyze the hydroxylation of linear alkanes with varying regioselectivities.

[00245] Mutant 9-10A exhibits high activity towards alkanes as small as propane. The alkane hydroxylation properties of this mutant are detailed in Tables 5 and 6.

[00246] In Table 5 the product distribution is determined as ratio of a specific alcohol product to the total amount of all alcohol products (given in %). The product distribution for ketones was similar to alcohol product distribution. The numbers reported here are the (%) total of all ketones relative to total products (alcohols + ketones). The favored enantiomer is listed in parentheses.

[00247] Table 5

Mutant	Product	Propane	Hexane	%ee	heptane	%ee	Octane	%ee	Nonane	%ee	Decane	%ee
9-10A	1-alcohol	8	0		1		1		0		1	
	2-alcohol	92	6	14(s)	26	7(S)	53	50(S)	39	60(S)	16	
	3-alcohol		95	41(s)	41	8(S)	20		59		32	
	4-alcohol				33		26		3		51	
	Ketones		<1		3		5		5		6	

[00248] For the data shown in Table 6 the initial NADPH consumption rates were measured over 15 seconds at 340 nm as nmol NADPH/min/nmol protein. Octane reactions contained 100 nM P450, 166 uM NADPH, and 4 mM octane in 1% ethanol and potassium phosphate buffer. Propane reactions contained 200 nM P450, 166 uM NADPH, and propane-saturated potassium phosphate buffer. Errors are at most 10%. Initial rates were measured by GC over 1 minute as nmol total products/min/nmol protein. Octane reactions contained 100nM P450, 500 uM NADPH, and 4 mM octane in 1% ethanol and potassium phosphate buffer. Propane reactions contained 1 uM P450, 500 uM NADPH, and propane-saturated potassium phosphate buffer. Coupling was determined by ratio of product formation rate to NADPH consumption rate. Total turnover numbers were determined as nmol product/nmol enzyme. Octane reactions contained 10-25 nM P450, 500 uM NADPH, and 4 mM octane in 1% ethanol and potassium phosphate buffer. Propane reactions contained 10-25 nM protein, potassium phosphate buffer saturated with propane, and an NADPH regeneration system containing 100 uM NADP+, 2 U/mL isocitrate dehydrogenase, and 10 mM isocitrate.

[00249] Table 6

Mutant	Substrate	Rate of NADPH Consumption (min ⁻¹) ^(a)	Rate of Product Formation (min ⁻¹) ^(b)	Coupling to NADPH (%) ^(c)	Total Turnovers ^(d)
9-10A	octane	2600	540	21	3000
	propane	420	23	5	1100

[00250] As shown in Table 5 mutant 9-10A hydroxylates octane into a mixture of octanols and ketones. The distribution of these products is different from the distribution produced by

wild type BM-3. 9-10A contains 14 mutations in the heme domain, only one of which, V78A, is located in the active site. This data indicates that active site changes, especially multiple changes, may alter product distributions and increase activity. For example, two other active-site mutations, A82L and A328V, shifted the regioselectivity of alkane hydroxylation predominantly to a single, subterminal ($\omega-1$) position. Based on these results, active site variants were generated in order to obtain enzymes with altered regioselectivity of hydroxylation and altered substrate specificities.

[00251] Screening libraries by spectroscopically monitoring NADPH consumption rates in the presence of the alkane often selects for mutants with NADPH oxidation rates highly uncoupled to product formation. This can be overcome through the use of surrogate substrates that mimic desired substrates, yet yield species when hydroxylated that are easily detected either directly or with the addition of a dye. For example, the screen for the high-throughput identification of mutant BM-3 enzymes with activity towards propane as described previously uses dimethyl ether (DME) as a propane surrogate substrate. Upon hydroxylation, it decomposes into methanol and dye-sensitive formaldehyde. In addition, hexyl methyl ether (HME) may be used as a surrogate for octane. It shares most of the physical properties of octane, including solubility in buffer, size, shape, and strength of its C-H bonds. When HME is hydroxylated at its methoxy carbon, it also produces an equivalent of formaldehyde which is visually detected at concentrations as low as 10 μ M with the dye Purpald (see Figure 25). Hexanal, the aldehyde product formed when the α -carbon of the hexyl group is hydroxylated, is practically unreactive towards Purpald at some concentrations used in the present screens. Formaldehyde is approximately 42 times more sensitive to Purpald than hexanal (see Figure 26).

This disproportionate response makes HME, in contrast to the symmetric ether DME, a regio-selective screen for terminal alkane hydroxylation.

[00252] Small libraries of changes at each individual active site position in 9-10A were made to identify changes that either increase its activity towards alkanes or alter its regioselectivity. The 11 key active site residues were chosen for their proximity, i.e. within 5 Angstroms of the 10 carbon atoms closest to the heme cofactor, to the bound fatty acid substrates co-crystallized with BM-3 heme domain in published crystal structures. These active site residues are A74, L75, V78, F81, A82, F87, T88, T260, I263, A264, and A328. For example, crystal structures of wildtype P450 BM-3 with and without substrate reveal large conformational changes upon substrate binding at the active site. The substrate free structure displays an open access channel with 17 to 21 ordered water molecules. Substrate recognition serves as a conformational trigger to close the channel, which dehydrates the active site, increases the redox potential, and allows dioxygen to bind to the heme. In the present studies saturation mutagenesis libraries were made at each of these 11 positions. The libraries were screened for activity towards dimethyl ether (DME) and hexylmethyl ether (HME).

[00253] Twenty-one single active site mutants with increased activity towards either of these substrates were isolated. Mutations that improved the activity were found at amino acid positions L75 (I, W), A78 (T, F, S), A82 (T, S, F, I, C, G, L), F87 (I, V, L), T88 (C), T260 (N, L, S) and A328 (L, F, M). The single active site mutants of 9-10A were characterized by their altered regioselectivity towards octane. Each of these mutations produces a characteristic octanol product distribution as a result of their presence altering the active site geometry above the heme cofactor. These 21 different mutants and their regioselectivity towards octane are listed

in Table 7. Transformations were carried out using cell lysates. The mutations were generated in variant 9-10A (with amino acid mutations R47C, V78A, K94I, P142S, T175I, A184V, F205C, S226R, H236Q, E252G, R255S, A290V, L353V). "Others" indicates overoxidation products (ketones and aldehydes) with a similar product distribution as the alcohols. "Rate" was measured as nmol octanol/nmol P450/minute. "TTN" was measured as nmol octanol/nmol P450 produced in 4 hour reaction.

[00254] Table 7

mutation ^(b)	1-ol [%]	2-ol [%]	3-ol [%]	4-ol [%]	Others ^(c)	rate ^(d)	ttn ^(e)
	0	51	21	27	1	540	3000
L75I	3	42	18	17	20	210	4500
L75W	3	6	18	24	9	200	4000
A78T	4	22	27	45	2	500	4000
A78F	4	37	17	31	11	380	3000
A78S	5	48	18	20	9	260	2500
A82T	5	23	25	44	3	340	2700
A82S	6	40	19	26	9	300	2600
A82F	4	26	20	48	2	140	3000
A82I	4	11	21	59	5	600	3500
A82C	5	26	26	42	1	200	3000
A82G	7	52	17	23	1	360	2000
F87I	8	70	6	3	13	60	2600
F87V	5	52	13	4	26	240	3600
F87L	7	55	22	12	4	60	2300
T88C	4	51	21	22	2	460	4000
T260L	7	67	11	10	5	260	2500
T260S	6	39	26	29	0	160	3000
T260N	6	29	23	42	0	300	2500
A328F	6	88	4	0	2	40	2800
A328M	24	71	0	0	5	10	700
A328L	13	87	0	0	0	20	1000

[00255] The original amino acids and the beneficial mutations described above were recombined to form a library of mutants containing all possible combinations of these

mutations. Mutations A82L and A328V were included in this library as they had proven to be important determinants of regioselectivity in previous experiments. Many of the active site mutations are too close to each other on the encoding gene such that standard methods such as DNA shuffling or StEP could not be used to recombine them. Therefore, the recombination library was assembled stepwise using designed oligonucleotide primers containing the mutations. Mutant 1-12G2 was used as a control in the screening procedure. A total of 16 mutants with improved activity towards DME or HME relative to 1-12G were selected for purification and further characterization. Octane was used as a model substrate for alkane hydroxylation and rates of product formation, total turnover number and regioselectivity were determined (Table 8). Mutations were found at amino acid positions 75, 78, 82, 87 and 328. While building the library ten improved members of a partial library were screened and sequenced. All possible amino acids at position 75 and 260 were identified, and four out of eight amino acids at position 78.

[00256] Table 8 provides a listing of mutations, product distribution, rates and total turnover numbers of octane hydroxylation reactions catalyzed by BM-3 active site mutants.

[00257] Mutations are relative to variant 9-10A (described above). Wild-type has a valine at position 78. The product distribution determined as ratio of a specific alcohol product to the total amount of alcohol products (given in %). The formation of ketones was also observed but is less than 5% of the total amount of product. The initial rates of product formation were measured by GC over 60 s as nmol total products/min/nmol protein. The reactions contained P450 (100 nM), NADPH (500 μ M), and octane (4 mM) in ethanol (1%) and potassium phosphate buffer (0.1 M, pH=8.0). The total turnover numbers are determined as nmol product/nmol protein. The reactions contained P450 (25 nM), NADPH (500 μ M), and

octane (4 mM) in ethanol (1%) and potassium phosphate buffer (0.1 M, pH=8.0).

[00258] Table 8

9-10A	A75	A78 ^[a]	A82	F87	A328	4-ol [%]	3-ol [%]	2-ol [%]	1-ol [%]	ttn ^[c]	rate ^[d]
9-10A						27	21	51	0	3000	540
77-9H		T	G		L	0	3	45	52	1300	160
1-7D	I	F	G		L	1	6	55	38	1300	220
68-8F		F	G		L	2	8	58	31	1400	180
49-9B			G	V	L	0	2	76	20	3800	310
35-7F		F	S		L	1	6	74	18	2000	200
1-5G	I	F	S		L	1	7	74	17	2200	380
13-7C		T			L	1	8	74	17	2600	310
49-1A		T	G	V	L	0	2	81	15	3700	350
2-4B	I	T			L	1	8	76	14	3900	320
12-10C			G	V	V	1	6	77	12	3100	240
11-8E				V	L	0	2	87	8	5700	550
41-5B		F	G		V	5	15	68	7	1800	370
29-3E		F	G		F	3	10	79	6	3200	160
7-11D			F		V	2	13	78	6	2200	150
29-10E			F		F	1	8	84	5	3200	140
53-5H		F	S		F	1	5	87	3	7400	660

[00259] The HME screening procedure was used to identify BM-3 variants with a regioselectivity shifted towards the terminal methyl carbon of the alkane chain. In general, mutants with high total turnover of HME in the screen produced more 1-octanol than compared to 1-12G. The total amount of 1-octanol formed (percent of 1-octanol multiplied by ttn) correlates well with the amount of HME converted. For example, mutant 77-9H (see Table 8 above) is not improved relative to 1-12G when considering rates and total turnovers on octane, but this mutant converts octane to 52% 1-octanol, making it highly active towards HME in the screen. The proportion of 1-octanol in the product mixture varies from three to 51%. Single active site mutants that showed significantly different regioselectivities were recombined. Upon recombination and screening for shifts towards terminal hydroxylation all the identified variants are selective for the terminal methyl or penultimate subterminal methylene group. Screening the recombination library e.g. for

hydroxylation of the third or fourth carbon of a linear alkane chain may identify variants that exhibit this specific regioselectivity.

[00260] The recombination library contains approximately 9,000 different mutants of P450 BM-3, each of which contains a characteristic active site. More than 70% of these were active towards DME or HME. Data provided herein indicate that mutations in the active site can yield variants that hydroxylate complex substrates such as 2-cyclopentyl-benzoxazole with high regio- and enantioselectivity, yielding potentially valuable intermediates for chemical synthesis (see example below).

[00261] Application of active site mutations to homologous proteins: Mutant 9-10A described above was acquired by accumulating point mutations in directed evolution experiments. An alternative method for making libraries for directed evolution to obtain P450s with new or altered properties is recombination, or chimeragenesis, in which portions of homologous P450s are swapped to form functional chimeras. Recombining equivalent segments of homologous proteins generates variants in which every amino acid substitution has already proven to be successful in one of the parents. Therefore, the amino acid mutations made in this way are less disruptive, on average, than random mutations. Recently, the structure-based algorithm "SCHEMA" was developed. The algorithm identifies fragments of proteins that can be recombined to minimize disruptive interactions that would prevent the protein from folding into its active form. P450 BM-3 and homologs thereof are soluble fusion proteins consisting of a catalytic heme-domain and an FAD- and FMN-containing NADPH reductase. They require dioxygen and an NADPH cofactor to catalyze substrate hydroxylation. However, the heme domain can utilize hydrogen peroxide via the peroxide "shunt" pathway for catalysis. While this 'peroxygenase'

activity is low in CYP102A1, it is enhanced by the amino acid substitution F87A. A similar effect has been shown for the equivalent F88A mutation in CYP102A2.

[00262] SCHEMA was used to design chimeras of P450 BM-3 and its homolog CYP102A2 sharing 63% amino acid sequence identity. Fourteen of the seventeen constructed hybrid proteins were able to fold correctly and incorporate the heme cofactor, as determined by CO difference spectra. The folded chimeras exhibited peroxygenase activity. A large library of chimeras made by SCHEMA-guided recombination of BM-3 with its homologs CYP102A2 and CYP102A3 contains more than 3000 new properly folded variants of BM-3. In addition to creating new biocatalysts, these chimeragenesis studies demonstrate that the homologous P450 enzymes CYP102A2 and CYP102A3 are similar enough to BM-3 to recombine in this fashion and still retain a high probability of folding.

[00263] Regioselectivity of variant 77-9H for the hydroxylation of n-alkanes: Mutant 77-9H with the highest selectivity for terminal hydroxylation was characterized in reactions with alkanes of chain length C6 to C10. The product distributions summarized in Table 9 show that mutant 77-9H hydroxylates only octane primarily at the terminal position. Alkanes shorter or longer than octane are hydroxylated mainly at the 2-position. Wildtype BM-3 hydroxylates lauric and palmitic acid at subterminal positions. Lauric acid is hydroxylated mainly (48%) at the ω -1 position, while palmitic acid is hydroxylated at the ω -2 position (48%). On fatty acids, mutant 77-9H shows a similar trend, with selectivity for specific subterminal positions (see Table 10). Residues that contact the terminal eight carbons of the substrate were altered based on the crystal structure with bound palmitoylglycine. Amino acid residues outside this region have not been modified, particularly those interacting with the carboxyl moiety.

[00264] Table 9

Substrate	1-ol [%]	2-ol [%]	3-ol [%]	4-ol [%]	5-ol [%]
hexane	11.4	83.8	4.9		
heptane	21.3	74.3	4.2	0.2	
octane	51.8	45.1	2.8	0.3	
nonane	9.7	88.1	1.5	0.7	0.0
decane	5.5	91.8	2.0	0.7	0.0

[00265] For Table 9 shows product distribution determined as ratio of a specific alcohol product to the total amount of alcohol products (given in %). The formation of ketones was also observed but is less than 10% of the total amount of product.

[00266] Table 10

Substrate	ω [%]	ω -1 [%]	ω -2 [%]	ω -3 [%]
lauric acid	7.5	52.1	2.8	37.5
palmitic acid	8.9	5.2	83.0	3.0

[00267] Table 10 shows product distributions for hydroxylation of fatty acids catalyzed by cytochrome P450 BM-3 77-9H. Product distribution was determined as ratio of a specific hydroxylation product to the total amount of hydroxylation products (given in %).

[00268] P450 BM-3 mutant F87A has been reported to support ω -hydroxylation of lauric acid. However, data provided herein shows that the F87A mutation broadens regioselectivity and shifts hydroxylation away from the terminal position. In the present studies, the F87A substitution was engineered into wild-type P450 BM-3 and 9-10A. The distribution of alkane and fatty acid hydroxylation product regioisomers were analyzed. Neither variant supports terminal hydroxylation activity on alkanes (Table 11) or fatty acids (Table 12). Table 11 shows product distributions for hydroxylation of alkanes catalyzed by cytochrome P450 BM-3 variants wt F87A and 9-10A F87A.

[00269] Table 11

Mutant	Substrate	1-ol [%]	2-ol [%]	3-ol [%]	4-ol [%]	5-ol [%]
wt F87A	hexane	0	54	46		
	heptane	0	36	61	3	
	octane	0	14	43	44	
	nonane	0	7	20	2	72
	decane	0	10	9	0	81
9-10A F87A	hexane	0	42	57		
	heptane	2	62	34	3	
	octane	1	55	37	7	
	nonane	0	24	46	5	25
	decane	0	14	23	1	62

[00270] Table 12 shows product distributions for fatty acid hydroxylation catalyzed by cytochrome P450 BM-3 variants wt F87A and 9-10A F87A.

[00271] Table 12

Mutant	Substrate	ω [%]	ω -1 [%]	ω -2 [%]	ω -3 [%]	ω -4 [%]	ω -5 [%]	ω -6 [%]	ω -7 [%]
wt-F87A	lauric acid	0	5	9	39	20	27	0	0
	palmitic acid	0	11	36	18	22	13	0	0
9-10A-F87A	lauric acid	0	5	6	28	14	47	0	0
	palmitic acid	0	0	1	1	2	8	20	68

[00272] Total turnover numbers (ttn) and rates of product formation were measured for hydroxylation of the different alkanes by 77-9H (see Table 13). The rates of product formation range from 112 min⁻¹ (for heptane) to 742 min⁻¹ (for nonane). The highest turnover rates of cytochromes P450 that carry out terminal hydroxylation reactions, however, are an order of magnitude lower than those of 77-9H. Wildtype BM-3 hydroxylates alkanes at rates that are an order of magnitude lower (e.g., 30 min⁻¹ for octane). To determine ttn, the enzyme was diluted to a concentration at which the P450 is neither oxygen-limited nor inactivated due to dialysis of the flavins. The total turnover number was found to be substrate-dependent and generally increases with the alkane chain

length. Rates of substrate hydroxylation appear to follow the same trend. It is well known that unproductive dissociation of reduced oxygen species can lead to the formation of superoxide or peroxide, which inactivate the enzyme.

Consistent with this is the fact that the enzyme supports the most turnovers on nonane, the substrate for which the coupling efficiency is the highest (66%, Table 13).

[00273] Table 13

Substrate	ttn ^(a)	rate of product formation [min ⁻¹] _j ^(b)	rate of NADPH oxidation [min ⁻¹] _j ^(c)	coupling efficiency [%] ^(d)
hexane	1800 ± 130	251 ± 29	1560 ± 60	16.1 ± 2.0
heptane	1730 ± 110	112 ± 9	1630 ± 20	6.9 ± 0.6
octane	3040 ± 150	236 ± 24	1630 ± 30	14.5 ± 1.7
nonane	4240 ± 680	742 ± 30	1150 ± 100	65.6 ± 9.3
decane	2831 ± 225	318 ± 47	940 ± 50	33.4 ± 7.2

[00274] For Table 13, total turnover numbers were determined as nmol product/nmol protein. The reactions contained P450 (25 nM), an NADPH regeneration system (166 μM NADP⁺, 6.6 U/mL isocitrate dehydrogenase, 41.6 mM isocitrate) and octane (4 mM) in ethanol (2 %) and potassium phosphate buffer (0.1 M, pH 8.0). The rates of product formation were measured by GC over 20 s as nmol total products/min/nmol protein. The reactions contained P450 (200 nM), NADPH (500 μM) and octane (2 mM) in ethanol (2%) and potassium phosphate buffer (0.1 M, pH 8.0). The rates of NADPH oxidation were measured over 20 s at 340 nm as nmol NADPH/min/nmol protein. The reactions contained P450 (200 nM), NADPH (160 μM), and octane (2 mM) in ethanol (2 %) and potassium phosphate buffer (0.1 M, pH 8.0). The background NADPH oxidation rate (without substrate) rate was 220/min. The coupling efficiency is the ratio of product formation rate to NADPH oxidation rate.

[00275] Wildtype P450 BM-3 tightly regulates electron transfer from the cofactor (NADPH) to the heme. In the absence of substrate, a weakly-bound water molecule acts as

the sixth, axial ligand of the heme iron. Substrate replaces this water molecule, perturbing the spin-state equilibrium of the heme iron in favor of the high-spin form and also increases the heme iron reduction potential by approximately 130 mV. These events trigger electron transfer to the substrate bound heme and start the catalytic cycle. However, with mutant BM-3 enzymes and nonnatural substrates, NADPH consumption is not necessarily coupled to the formation of product. Instead, electrons from NADPH reduce heme-bound dioxygen to water or the reactive oxygen species peroxide and superoxide. To determine the coupling efficiency (the fraction of NADPH used for product formation) of mutant 77-9H, rates of NADPH oxidation upon addition of the alkane substrates were measured and compared to product formation rates (Table 13). The rate of product formation is correlated to coupling efficiency, i.e. the more the reaction is coupled to NADPH oxidation, the higher the rate of product formation. The highest coupling efficiency (66%) was observed for nonane.

[00276] Linear alkanes are challenging to hydroxylate selectively because they contain no functional groups that help fix their orientation for reaction with the active heme iron-oxo species of cytochrome P450. Model studies with metalloporphyrins show that, absent substrate binding effects, the regioselectivity of n-alkane hydroxylation is determined by the relative bond dissociation energies. Engineering a P450 for terminal hydroxylase activity must therefore override the inherent specificity of the catalytic species for the methylene groups so that only the terminal methyl carbon is activated.

[00277] Wildtype cytochrome P450 BM-3 hydroxylates its preferred substrates, medium-chain fatty acids, at subterminal positions. The terminal carbon is not hydroxylated. By screening saturation mutagenesis libraries at 11 active site positions and recombining beneficial mutations a set of P450

BM-3 mutants that exhibit a range of regioselectivities for fatty acid and alkane hydroxylation are provided herein. One mutant, 77-9H, hydroxylates octane at the terminal position with 51% selectivity, but exhibits this selectivity only for octane. The active site of 77-9H is therefore not restricted near the activated oxygen to prevent subterminal hydroxylation. The specificity for the terminal methyl group of octane probably reflects specific interactions between active site residues and octane methylene groups and/or the methyl group at the unreacted end. The hexyl methyl ether substrate used for screening terminal hydroxylation activity has the same chain length as octane. The mutations that enable terminal hydroxylation of HME likely act in a similar fashion to promote terminal hydroxylation of octane.

[00278] Recombining active site mutations for linear, terminal alkene epoxidation: The target reaction was the selective epoxidation of linear, terminal alkenes. The same saturation mutagenesis libraries were screened to identify single active site mutants that exhibit this activity. Upon recombination of beneficial active site mutations, two mutants were identified that epoxidize linear terminal alkenes with high regioselectivity and opposite enantioselectivity.

[00279] Potential uses of active site recombination variants for the regioselective hydroxylation of larger compounds: The target compound 2-cyclopentylbenzoxazole was hydroxylated by active site mutant 1-12G with high regio- and enantioselectivity (see 'Regio- and Enantio-Selective Alkane Hydroxylation with Engineered Cytochromes P450' - PCT/US04/18832, incorporated herein by reference). The work presented herein demonstrates that active site mutants are well suited for the preparation of chiral intermediates for chemical synthesis. In this case, the hydroxylated product can be used to as an intermediate in the synthesis of, for

example, Carbovir, a carboxylic nucleoside potentially active against human HIV.

[00280] The active site mutants 53-5H and 77-9H, as provided herein, were tested for the hydroxylation of 2-cyclopentylbenzoxazole. The regioselectivity for the formation of the 3-(benzoxazol-2-yl)cyclopentanol was even higher than reported for 1-12G. The observed selectivities of 53-5H and 77-9H were 99.4 % and 99.9 %, respectively. The enantioselectivities were not as high with measured ee values of 55.8 % for 53-5H and 1.5 % for 77-9H. However, mutants were selected for high regioselectivity and not enantioselectivity.

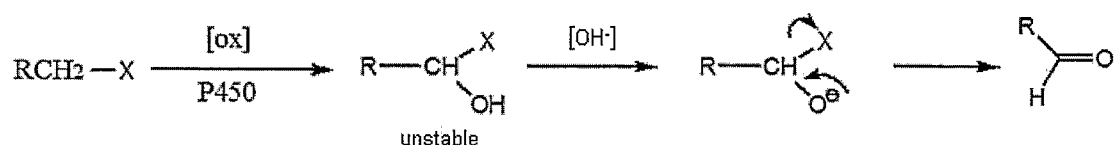
[00281] Direct conversion of gaseous alkanes, such as ethane or propane to their corresponding alcohols: The mutants provided herein clearly indicate that single active site mutants, and combinations thereof, may be used to generate active sites in P450 BM-3 capable of binding a diverse array of substrates, including gaseous alkanes.

[00282] The mutant with the highest activity on propane, 53-5H, was shown to catalyze thousands of turnovers of propane to propanol at a rate of 370 min⁻¹. This mutant also hydroxylates ethane to generate ethanol 53-5H contains three active-site mutations, A78F, A82S, and A328F, all of which replace alanine with a larger side chain and presumably reduce the volume of the active site and position small alkanes above the heme during catalysis. In addition to its activity towards ethane, 53-5H exhibits the highest regioselectivity (89% 2-octanol) and enantioselectivity (65% S-2-octanol) towards octane of any BM-3 variant identified herein. This information further indicates that tighter substrate binding in the engineered active site occurs in the modified enzymes.

[00283] Dehalogenation of halogenated alkanes by P450 BM-3 mutants: The hydroxylation activity of BM-3 active site mutants on short- and medium-chain alkanes could be

potentially useful to carry out the dehalogenation of halogenated alkanes. Accordingly, wild-type P450 BM-3 was used as the starting point for generating active site variants. Alanine at amino acid position 328 was substituted with asparagine, leucine, isoleucine, proline, and valine using standard DNA manipulation techniques. The amino acid sequence of the wild-type enzyme is provided in SEQ ID NO: 1 (see Figure 8).

[00284] The dehalogenation reaction is based on the hydroxylation of the carbon atom carrying the halogen by a spontaneous (base catalyzed) alpha-elimination reaction which leads to the formation of an aldehyde or ketone depending on the starting haloalkane. The expected reaction mechanism for the dehalogenation of a general alpha-haloalkane is described by the scheme:



[00285] To test the feasibility of this approach, chloromethane was chosen as substrate for the dehalogenation reaction. Compared to other haloalkanes, chloromethane (CM) has the advantage to be moderately soluble in water (0.5g/100 ml) and to lead to formaldehyde as end product of the hydroxylation/alpha-elimination reaction. Formaldehyde can be then readily detected by colorimetric reaction with Purpald as described previously. To test the reaction, two BM-3 mutants from later rounds of random mutagenesis, 35-E1141 and 20-D3, were incubated in the presence of CM and a cofactor regeneration system based on NADP⁺, glucose-6-phosphate and glucose-6-phosphate dehydrogenase. Reactions with wild-type P450 BM-3 and without enzyme were included as controls. A specific colorimetric response (see Figure 27A) was observed only in the presence of the BM-3 mutants.

[00286] The reaction was repeated in the presence of additional BM-3 mutants, including 1-8H, 1-12G2 and the wild-type P450 BM-3 in which A328 was substituted with valine (BM-3-A328V) (Figure 27B). The total turnover number of BM-3-A328V on CM is considerably higher than that of the other mutants considered in the study, while the wild-type enzyme exhibits no detectable activity (Figure 28). Wild-type BM-3 was substituted with other amino acid residues at position V328. These mutants, however, are much less active than BM-3-A328V on CM (Figure 29).

[00287] The activity of BM-3-A328V on CM and other substrates such as HME, DME, and propane was further investigated by measuring product formation rate, cofactor consumption and coupling efficiency (Table 14). This mutant was also found to be capable to hydroxylate ethane to ethanol, although only for a total of about 5-10 turnovers. In Table 14 "consumption" rates are expressed as nmol NADPH/nmol enzyme/min and "product formation" rates are expressed as nmol product/nmol enzyme/min.

[00288] Table 14

rates/substrates	no substrate	HME	DME	Propane	CH ₃ Cl	ethane
NADPH consumption ^[a]	19.9	36.3	12.5	n.d.	16.2	n.d.
product formation ^[b]	-	4.3	3.0	7.0	1.1	n.d.
coupling (%)	-	11.8	24.2	8.0	6.8	n.d.
TTN	-	n.d.	n.d.	2300	130	7

[00289] It is envisioned that the type of dehalogenation reaction described for chloromethane is extendable to other halomethanes such as bromo- and iodomethane as well as longer haloalkanes (e.g., haloethane, 1- and 2-halopropane). The resulting haloalcohols are reactive intermediates which can form the corresponding carbonyl compounds (aldehyde, ketones)

or epoxides depending on the regioselectivity of the hydroxylation reaction and the nature of starting compounds (Figure 30). These reactions are useful for bioremediation as well as for synthetic purposes.

[00290] The examples set forth above are provided to give those of ordinary skill in the art a complete disclosure and description of how to make and use the embodiments of the devices, systems and methods of the invention, and are not intended to limit the scope of what the inventors regard as their invention. Modifications of the above-described modes for carrying out the invention that are obvious to persons of skill in the art are intended to be within the scope of the following claims. All patents and publications mentioned in the specification are indicative of the levels of skill of those skilled in the art to which the invention pertains. All references cited in this disclosure are incorporated by reference to the same extent as if each reference had been incorporated by reference in its entirety individually.

[00291] A number of embodiments of the invention have been described. Nevertheless, it will be understood that various modifications may be made without departing from the spirit and scope of the invention. Accordingly, other embodiments are within the scope of the following claims.

WHAT IS CLAIMED IS

1. An isolated or recombinant polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:1 with up to 50 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, and 353, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at positions 47, 82, 142, 205, 236, 252, and 255, a Z1 amino acid residue;

(b) at positions 94, 175, 184, 290, and 353, a Z2 amino acid residue;

(c) at position 226, a Z3 amino acid residue; and

(d) at positions 78 and 328, a Z4 amino acid residue,

wherein: Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine (C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and arginine (R); and Z4 is an amino acid residue selected from the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H).

2. The polypeptide of claim 1, further comprising a Z3 amino acid residue at position 285 of SEQ ID NO:1, wherein Z3 is an amino acid residue selected from the group consisting of lysine (K), arginine (R), and histidine (H).

3. The polypeptide of claim 2, wherein the amino acid residue at position 285 of SEQ ID NO:1 is an arginine (R).

4. The polypeptide of claim 1, comprising up to 25 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, and 353 of SEQ ID NO:1.
5. The polypeptide of claim 1, comprising up to 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, and 353 of SEQ ID NO:1.
6. The polypeptide of claim 1, comprising up to 5 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, and 353 of SEQ ID NO:1.
7. The polypeptide of claim 1, wherein the amino acid sequence comprises residues selected from the group consisting of:
- (a) at positions 47, 82, 142, 205, 236, 252, and 255, a glycine (G), glutamine (Q), serine (S), threonine (T), or cysteine (C) amino acid residue;
 - (b) at positions 94, 175, 184, 290, and 353, a valine (V) or isoleucine (I) amino acid residue;
 - (c) at position 226, an arginine (R) amino acid residue;
- and
- (d) at positions 78 and 328, a phenylalanine (F) or histidine (H) amino acid residue.
8. The polypeptide of claim 7, wherein the amino acid sequence comprises residues selected from the group consisting of:
- (a) at positions 82, 142 and 355, a serine (S) amino acid residue;

(b) at positions 47 and 205, a cysteine (C) amino acid residue;

(c) at position 236, a glutamine (Q) amino acid residue;

(d) at position 252, a glycine (G) amino acid residue;

(e) at positions 184, 290 and 353, a valine (V) amino acid residue;

(f) at positions 94 and 175, an isoleucine (I) amino acid residue; and

(g) at positions 78 and 328, a phenylalanine (F) amino acid residue.

9. An isolated or recombinant polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:1, wherein at least 80% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as follows:

(a) at positions 82, 142 and 255 the amino acid residue is serine (S);

(b) at positions 184, 290 and 353 the amino acid residue is valine (V);

(c) at positions 94 and 175 the amino acid residue is isoleucine (I);

(d) at positions 47 and 205 the amino acid residue is cysteine (C);

(e) at positions 78 and 328 the amino acid residue is phenylalanine (F);

(f) at position 252 the amino acid residue is glycine (G);

(g) at position 236 the amino acid residue is glutamine (Q); and

(h) at position 226 the amino acid residue is arginine (R).

10. The polypeptide of claim 9, wherein at least 90% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as specified in (a) - (h).
11. The polypeptide of claim 9, wherein at least 95% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as specified in (a) - (h).
12. The polypeptide of claim 9, further comprising an arginine (R) at position 285 of SEQ ID NO:1.
13. An isolated or recombinant polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:2 or SEQ ID NO:3 with up to 50 conservative amino acid substitutions excluding residues 48, 79, 83, 95, 143, 176, 185, 206, 227, 238, 254, 257, 292, 330, and 355, wherein the amino acid sequence comprises residues selected from the group consisting of:
- (a) at positions 48, 83, 143, 206, 238, 254, and 257, a Z1 amino acid residue;
 - (b) at positions 95, 176, 185, 292, and 355, a Z2 amino acid residue;
 - (c) at position 227, a Z3 amino acid residue; and
 - (d) at positions 79 and 330, a Z4 amino acid residue,
- wherein: Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine (C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and arginine (R); and Z4 is an amino acid residue selected from

the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H).

14. The polypeptide of claim 13, further comprising a Z3 amino acid residue at position 287 of SEQ ID NO:2 or SEQ ID NO:3, wherein Z3 is an amino acid residue selected from the group consisting of lysine (K), arginine (R), and histidine (H).

15. The polypeptide of claim 14, wherein the amino acid residue at position 287 of SEQ ID NO:2 or SEQ ID NO:3 is an arginine (R).

16. An isolated or recombinant polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:2 or SEQ ID NO:3, wherein at least 80% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as follows:

(a) at positions 83, 143 and 257 the amino acid residue is serine (S);

(b) at positions 185, 292 and 355 the amino acid residue is valine (V);

(c) at positions 95 and 176 the amino acid residue is isoleucine (I);

(d) at positions 48 and 206 the amino acid residue is cysteine (C);

(e) at positions 79 and 330 the amino acid residue is phenylalanine (F);

(f) at position 254 the amino acid residue is glycine (G);

(g) at position 238 the amino acid residue is glutamine (Q); and

(h) at position 227 the amino acid residue is arginine (R).

17. The polypeptide of claim 16, wherein at least 90% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as specified in (a) - (h).

18. The polypeptide of claim 16, wherein at least 95% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as specified in (a) - (h).

19. The polypeptide of claim 16, further comprising an arginine (R) at position 287 of SEQ ID NO:2 or SEQ ID NO:3.

20. An isolated or recombinant polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:4 with up to 50 conservative amino acid substitutions excluding residues 49, 80, 84, 96, 144, 177, 186, 207, 228, 239, 255, 258, 293, 331, and 357, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at positions 49, 84, 144, 207, 239, 255, and 258, a Z1 amino acid residue;

(b) at positions 96, 177, 186, 293, and 357, a Z2 amino acid residue;

(c) at position 228, a Z3 amino acid residue; and

(d) at positions 80 and 331, a Z4 amino acid residue,

wherein: Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine (C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and

arginine (R); and Z4 is an amino acid residue selected from the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H).

21. The polypeptide of claim 20, further comprising a Z3 amino acid residue at position 288 of SEQ ID NO:4, wherein Z3 is an amino acid residue selected from the group consisting of lysine (K), arginine (R), and histidine (H).

22. The polypeptide of claim 20, wherein the amino acid residue at position 288 of SEQ ID NO:4 is an arginine (R).

23. An isolated or recombinant polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:4, wherein at least 80% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as follows:

(a) at positions 84, 144 and 258 the amino acid residue is serine (S);

(b) at positions 186, 293 and 357 the amino acid residue is valine (V);

(c) at positions 96 and 177 the amino acid residue is isoleucine (I);

(d) at positions 49 and 207 the amino acid residue is cysteine (C);

(e) at positions 80 and 331 the amino acid residue is phenylalanine (F);

(f) at position 255 the amino acid residue is glycine (G);

(g) at position 239 the amino acid residue is glutamine (Q); and

(h) at position 228 the amino acid residue is arginine (R).

24. The polypeptide of claim 23, wherein at least 90% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as specified in (a) - (h).

25. The polypeptide of claim 23, wherein at least 95% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as specified in (a) - (h).

26. The polypeptide of claim 23, further comprising an arginine (R) at position 288 of SEQ ID NO:4.

27. An isolated or recombinant polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:5 with up to 50 conservative amino acid substitutions excluding residues 54, 85, 89, 101, 149, 182, 191, 212, 232, 243, 259, 262, 297, 336, and 361, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at positions 54, 89, 149, 212, 243, 259, and 262, a Z1 amino acid residue;

(b) at positions 101, 182, 191, 297, and 361, a Z2 amino acid residue;

(c) at position 232, a Z3 amino acid residue; and

(d) at positions 85 and 336, a Z4 amino acid residue,

wherein: Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine (C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and arginine (R); and Z4 is an amino acid residue selected from

the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H).

28. The polypeptide of claim 27, further comprising a Z3 amino acid residue at position 292 of SEQ ID NO:5, wherein Z3 is an amino acid residue selected from the group consisting of lysine (K), arginine (R), and histidine (H).

29. The polypeptide of claim 27, wherein the amino acid residue at position 292 of SEQ ID NO:5 is an arginine (R).

30. An isolated or recombinant polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:5, wherein at least 80% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as follows:

(a) at positions 85, 149 and 259 the amino acid residue is serine (S);

(b) at positions 191, 297 and 361 the amino acid residue is valine (V);

(c) at positions 101 and 182 the amino acid residue is isoleucine (I);

(d) at positions 54 and 212 the amino acid residue is cysteine (C);

(e) at positions 85 and 336 the amino acid residue is phenylalanine (F);

(f) at position 259 the amino acid residue is glycine (G);

(g) at position 243 the amino acid residue is glutamine (Q); and

(h) at position 232 the amino acid residue is arginine (R).

31. The polypeptide of claim 30, wherein at least 90% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as specified in (a) - (h).

32. The polypeptide of claim 30, wherein at least 95% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as specified in (a) - (h).

33. The polypeptide of claim 30, further comprising an arginine (R) at position 292 of SEQ ID NO:5.

34. An isolated or recombinant polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:6 with up to 50 conservative amino acid substitutions excluding residues 51, 82, 86, 98, 146, 179, 188, 208, 231, 242, 258, 262, 296, 337, and 363, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at positions 51, 86, 146, 208, 242, 258, and 262, a Z1 amino acid residue;

(b) at positions 98, 179, 188, 296, and 363, a Z2 amino acid residue;

(c) at position 231, a Z3 amino acid residue; and

(d) at positions 82 and 337, a Z4 amino acid residue,

wherein: Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine (C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and arginine (R); and Z4 is an amino acid residue selected from

the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H).

35. The polypeptide of claim 34, further comprising a Z3 amino acid residue at position 291 of SEQ ID NO:6, wherein Z3 is an amino acid residue selected from the group consisting of lysine (K), arginine (R), and histidine (H).

36. The polypeptide of claim 34, wherein the amino acid residue at position 291 of SEQ ID NO:6 is an arginine (R).

37. An isolated or recombinant polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:6, wherein at least 80% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as follows:

(a) at positions 82, 146 and 258 the amino acid residue is serine (S);

(b) at positions 188, 296 and 363 the amino acid residue is valine (V);

(c) at positions 98 and 179 the amino acid residue is isoleucine (I);

(d) at positions 51 and 208 the amino acid residue is cysteine (C);

(e) at positions 82 and 337 the amino acid residue is phenylalanine (F);

(f) at position 258 the amino acid residue is glycine (G);

(g) at position 242 the amino acid residue is glutamine (Q); and

(h) at position 231 the amino acid residue is arginine (R).

38. The polypeptide of claim 37, wherein at least 90% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as specified in (a) - (h).

39. The polypeptide of claim 37, wherein at least 95% of the amino acid residues in the amino acid sequence at the positions specified in (a) - (h) are as specified in (a) - (h).

40. The polypeptide of claim 37, further comprising an arginine (R) at position 291 of SEQ ID NO:6.

41. An isolated or recombinant polypeptide comprising residues 456-1048 of the amino acid sequence set forth in SEQ ID NO:1 with up to 65 conservative amino acid substitutions excluding residues 464, 631, 645, 710 and 968, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at position 464, a Z1, Z3, Z4, or Z5 amino acid residue;

(b) at positions 631 and 710, a Z1 amino acid residue;

(c) at position 645, a Z3 amino acid residue; and

(d) at position 968, a Z2 amino acid residue,

wherein: Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine (C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and arginine (R); Z4 is an amino acid residue selected from the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H); and is an amino acid

residue selected from the group consisting of threonine (T), valine (V), and isoleucine (I).

42. The polypeptide of claim 41, comprising up to 40 conservative amino acid substitutions excluding residues 464, 631, 645, 710 and 968 of SEQ ID NO:1.

43. The polypeptide of claim 41, comprising up to 25 conservative amino acid substitutions excluding residues 464, 631, 645, 710 and 968 of SEQ ID NO:1.

44. The polypeptide of claim 41 comprising up to 10 conservative amino acid substitutions excluding residues 464, 631, 645, 710 and 968 of SEQ ID NO:1.

45. The polypeptide of claim 41, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at position 464, a glycine (G), arginine (R), tyrosine (Y), or threonine (T) amino acid residue;

(b) at position 631, an asparagine (N) amino acid residue;

(c) at position 645, an arginine (R) amino acid residue;

(d) at position 710, a threonine (T) amino acid residue;

and

(e) at position 968, a lysine (L) amino acid residue.

46. An isolated or recombinant polypeptide comprising residues 456-1059 of the amino acid sequence set forth in SEQ ID NO:2 with up to 65 conservative amino acid substitutions excluding residues 475, 641, 656, 721 and 980, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at position 475, a Z1, Z3, Z4, or Z5 amino acid residue;

(b) at positions 641 and 721, a Z1 amino acid residue;

(c) at position 656, a Z3 amino acid residue; and

(d) at position 980, a Z2 amino acid residue,

wherein: Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine (C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and arginine (R); Z4 is an amino acid residue selected from the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H); and is an amino acid residue selected from the group consisting of threonine (T), valine (V), and isoleucine (I).

47. The polypeptide of claim 46, comprising up to 40 conservative amino acid substitutions excluding residues 475, 641, 656, 721 and 980 of SEQ ID NO:2.

48. The polypeptide of claim 46, comprising up to 25 conservative amino acid substitutions excluding residues 475, 641, 656, 721 and 980 of SEQ ID NO:2.

49. The polypeptide of claim 46, comprising up to 10 conservative amino acid substitutions excluding residues 475, 641, 656, 721 and 980 of SEQ ID NO:2.

50. The polypeptide of claim 46, wherein the amino acid sequence comprises residues selected from the group consisting of:

- (a) at position 475, a glycine (G), arginine (R), tyrosine (Y), or threonine (T) amino acid residue;
 - (b) at position 641, an asparagine (N) amino acid residue;
 - (c) at position 656, an arginine (R) amino acid residue;
 - (d) at position 721, a threonine (T) amino acid residue;
- and
- (e) at position 980, a lysine (L) amino acid residue.

51. An isolated or recombinant polypeptide comprising residues 456-1053 of the amino acid sequence set forth in SEQ ID NO:3 with up to 65 conservative amino acid substitutions excluding residues 467, 634, 648, 713 and 972, wherein the amino acid sequence comprises residues selected from the group consisting of:

- (a) at position 467, a Z1, Z3, Z4, or Z5 amino acid residue;
- (b) at positions 634 and 713, a Z1 amino acid residue;
- (c) at position 648, a Z3 amino acid residue; and
- (d) at position 972, a Z2 amino acid residue,

wherein: Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine (C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and arginine (R); Z4 is an amino acid residue selected from the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H); and Z5 is an amino acid residue selected from the group consisting of threonine (T), valine (V), and isoleucine (I).

52. The polypeptide of claim 51, comprising up to 40 conservative amino acid substitutions excluding residues 467, 634, 648, 713 and 972 of SEQ ID NO:3.

53. The polypeptide of claim 51, comprising up to 25 conservative amino acid substitutions excluding residues 467, 634, 648, 713 and 972 of SEQ ID NO:3.

54. The polypeptide of claim 51, comprising up to 10 conservative amino acid substitutions excluding residues 467, 634, 648, 713 and 972 of SEQ ID NO:3.

55. The polypeptide of claim 51, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at position 467, a glycine (G), arginine (R), tyrosine (Y), or threonine (T) amino acid residue;

(b) at position 634, an asparagine (N) amino acid residue;

(c) at position 648, an arginine (R) amino acid residue;

(d) at position 713, a threonine (T) amino acid residue;

and

(e) at position 972, a lysine (L) amino acid residue.

56. An isolated or recombinant polypeptide comprising residues 456-1064 of the amino acid sequence set forth in SEQ ID NO:4 with up to 65 conservative amino acid substitutions excluding residues 477, 644, 659, 724 and 983, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at position 477, a Z1, Z3, Z4, or Z5 amino acid residue;

(b) at positions 644 and 724, a Z1 amino acid residue;

(c) at position 659, a Z3 amino acid residue; and

(d) at position 983, a Z2 amino acid residue,
wherein: Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine (C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and arginine (R); Z4 is an amino acid residue selected from the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H); and is an amino acid residue selected from the group consisting of threonine (T), valine (V), and isoleucine (I).

57. The polypeptide of claim 56, comprising up to 40 conservative amino acid substitutions excluding residues 477, 644, 659, 724 and 983 of SEQ ID NO:4.

58. The polypeptide of claim 56, comprising up to 25 conservative amino acid substitutions excluding residues 477, 644, 659, 724 and 983 of SEQ ID NO:4.

59. The polypeptide of claim 56, comprising up to 10 conservative amino acid substitutions excluding residues 477, 644, 659, 724 and 983 of SEQ ID NO:4.

60. The polypeptide of claim 56, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at position 477, a glycine (G), arginine (R), tyrosine (Y), or threonine (T) amino acid residue;

(b) at position 644, an asparagine (N) amino acid residue;

(c) at position 659, an arginine (R) amino acid residue;

(d) at position 724, a threonine (T) amino acid residue;
and

(e) at position 983, a lysine (L) amino acid residue.

61. An isolated or recombinant polypeptide comprising residues 456-1063 of the amino acid sequence set forth in SEQ ID NO:5 with up to 65 conservative amino acid substitutions excluding residues 472, 640, 656, 723 and 985, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at position 472, a Z1, Z3, Z4, or Z5 amino acid residue;

(b) at positions 640 and 723, a Z1 amino acid residue;

(c) at position 659, a Z3 amino acid residue; and

(d) at position 985, a Z2 amino acid residue,

wherein: Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine (C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and arginine (R); Z4 is an amino acid residue selected from the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H); and Z5 is an amino acid residue selected from the group consisting of threonine (T), valine (V), and isoleucine (I).

62. The polypeptide of claim 61, comprising up to 40 conservative amino acid substitutions excluding residues 472, 640, 656, 723 and 985 of SEQ ID NO:5.

63. The polypeptide of claim 61, comprising up to 25 conservative amino acid substitutions excluding residues 472, 640, 656, 723 and 985 of SEQ ID NO:5.

64. The polypeptide of claim 61, comprising up to 10 conservative amino acid substitutions excluding residues 472, 640, 656, 723 and 985 of SEQ ID NO:5.

65. The polypeptide of claim 61, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at position 472, a glycine (G), arginine (R), tyrosine (Y), or threonine (T) amino acid residue;

(b) at position 640, an asparagine (N) amino acid residue;

(c) at position 656, an arginine (R) amino acid residue;

(d) at position 723, a threonine (T) amino acid residue;

and

(e) at position 985, a lysine (L) amino acid residue.

66. An isolated or recombinant polypeptide comprising residues 456-1077 of the amino acid sequence set forth in SEQ ID NO:6 with up to 65 conservative amino acid substitutions excluding residues 480, 648, 664, 733 and 997, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at position 480, a Z1, Z3, Z4, or Z5 amino acid residue;

(b) at positions 648 and 733, a Z1 amino acid residue;

(c) at position 664, a Z3 amino acid residue; and

(d) at position 997, a Z2 amino acid residue,

wherein: Z1 is an amino acid residue selected from the group consisting of glycine (G), asparagine (N), glutamine (Q), serine (S), threonine (T), tyrosine (Y), and cysteine

(C); Z2 is an amino acid residue selected from the group consisting of alanine (A), valine (V), leucine (L), isoleucine (I), proline (P), and methionine (M); Z3 is an amino acid residue selected from the group consisting of lysine (K), and arginine (R); Z4 is an amino acid residue selected from the group consisting of tyrosine (Y), phenylalanine (F), tryptophan (W), and histidine (H); and is an amino acid residue selected from the group consisting of threonine (T), valine (V), and isoleucine (I).

67. The polypeptide of claim 66, comprising up to 40 conservative amino acid substitutions excluding residues 480, 648, 664, 733 and 997 of SEQ ID NO:6.

68. The polypeptide of claim 66, comprising up to 25 conservative amino acid substitutions excluding residues 480, 648, 664, 733 and 997 of SEQ ID NO:6.

69. The polypeptide of claim 66, comprising up to 10 conservative amino acid substitutions excluding residues 480, 648, 664, 733 and 997 of SEQ ID NO:6.

70. The polypeptide of claim 66, wherein the amino acid sequence comprises residues selected from the group consisting of:

(a) at position 480, a glycine (G), arginine (R), tyrosine (Y), or threonine (T) amino acid residue;

(b) at position 648, an asparagine (N) amino acid residue;

(c) at position 664, an arginine (R) amino acid residue;

(d) at position 733, a threonine (T) amino acid residue;

and

(e) at position 997, a lysine (L) amino acid residue.

71. An isolated or recombinant polypeptide comprising the amino acid sequence set forth in SEQ ID NO:7 with up to 75 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710.

72. An isolated or recombinant polypeptide comprising the amino acid sequence set forth in SEQ ID NO:8 with up to 75 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710.

73. An isolated or recombinant polypeptide comprising the amino acid sequence set forth in SEQ ID NO:9 with up to 75 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710.

74. An isolated or recombinant polypeptide comprising the amino acid sequence set forth in SEQ ID NO:10 with up to 75 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710.

75. The polypeptide of claim 71, 72, 73, or 74, further comprising an arginine (R) at amino acid residue position 285 of SEQ ID NO:7, 8, 9, or 10.

76. The polypeptide of claim 71, 72, 73, or 74, comprising up to 50 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710.

77. The polypeptide of claim 71, 72, 73, or 74, comprising up to 25 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710.

78. The polypeptide of claim 71, 72, 73, or 74, comprising up to 10 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710.

79. An isolated or recombinant polypeptide comprising the amino acid sequence set forth in SEQ ID NO:11 with up to 75 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 285, 290, 328, 353, 464, and 710.

80. An isolated or recombinant polypeptide comprising the amino acid sequence set forth in SEQ ID NO:12 with up to 75 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 645, and 710.

81. An isolated or recombinant polypeptide comprising the amino acid sequence set forth in SEQ ID NO:13 with up to 75 conservative amino acid substitutions excluding residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 631, and 710.

82. The polypeptide of claim 79, 80, or 81, comprising up to 50 conservative amino acid substitutions excluding residues:

(a) 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 285, 290, 328, 353, 464, and 710 of SEQ ID NO:11;

(b) 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 645, and 710 of SEQ ID NO:12; and

(c) 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 631, and 710 of SEQ ID NO:13.

83. The polypeptide of claim 79, 80, or 81, comprising up to 25 conservative amino acid substitutions excluding residues:

(a) 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 285, 290, 328, 353, 464, and 710 of SEQ ID NO:11;

(b) 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 645, and 710 of SEQ ID NO:12; and

(c) 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 631, and 710 of SEQ ID NO:13.

84. The polypeptide of claim 79, 80, or 81, comprising up to 10 conservative amino acid substitutions excluding residues:

(a) 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 285, 290, 328, 353, 464, and 710 of SEQ ID NO:11;

(b) 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 645, and 710 of SEQ ID NO:12; and

(c) 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 631, and 710 of SEQ ID NO:13.

85. An isolated or recombinant polypeptide selected from the group consisting of:

(a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:7;

(b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:7;

(c) a polypeptide comprising an amino acid sequence having at least 60% sequence identity to the amino acid sequence set forth in SEQ ID NO:7 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:7; and

(d) a polypeptide comprising an amino acid sequence that can be optimally aligned with the sequence of SEQ ID NO:7 to

generate a similarity score of at least 1830, using the BLOSUM62 matrix, a gap existence penalty of 11, and a gap extension penalty of 1, excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:7.

86. An isolated or recombinant polypeptide selected from the group consisting of:

(a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:8;

(b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:8;

(c) a polypeptide comprising an amino acid sequence having at least 60% sequence identity to the amino acid sequence set forth in SEQ ID NO:8 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:8.

87. An isolated or recombinant polypeptide selected from the group consisting of:

(a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:9;

(b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:9; and

(c) a polypeptide comprising an amino acid sequence having at least 60% sequence identity to the amino acid sequence set forth in SEQ ID NO:9 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:9.

88. An isolated or recombinant polypeptide selected from the group consisting of:

(a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:10;

(b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:10; and

(c) a polypeptide comprising an amino acid sequence having at least 60% sequence identity to the amino acid sequence set forth in SEQ ID NO:10 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, and 710 of SEQ ID NO:10.

89. An isolated or recombinant polypeptide selected from the group consisting of:

(a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:11;

(b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:11; and

(c) a polypeptide comprising an amino acid sequence having at least 60% sequence identity to the amino acid sequence set forth in SEQ ID NO:11 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 285, 290, 328, 353, 464, and 710 of SEQ ID NO:11.

90. An isolated or recombinant polypeptide selected from the group consisting of:

(a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:12;

(b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:12; and

(c) a polypeptide comprising an amino acid sequence having at least 60% sequence identity to the amino acid sequence set forth in SEQ ID NO:12 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 645, and 710 of SEQ ID NO:12.

91. An isolated or recombinant polypeptide selected from the group consisting of:

(a) a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:13; and

(b) a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:13;

(c) a polypeptide comprising an amino acid sequence having at least 60% sequence identity to the amino acid sequence set forth in SEQ ID NO:13 excluding amino acid residues 47, 78, 82, 94, 142, 175, 184, 205, 226, 236, 252, 255, 290, 328, 353, 464, 631, and 710 of SEQ ID NO:13.

92. The polypeptide of claim 85, 86, 87, 88, 89, 90, or 91, wherein the amino acid sequence comprises at least 70% sequence identity to the amino acid sequence set forth in part (c) of each of claims 110, 111, 112, 113, 114, 115, and 116.

93. The polypeptide of claim 85, 86, 87, 88, 89, 90, or 91, wherein the amino acid sequence comprises at least 80% sequence identity to the amino acid sequence set forth in part (c) of each of claims 110, 111, 112, 113, 114, 115, and 116.

94. The polypeptide of claim 85, 86, 87, 88, 89, 90, or 91, wherein the amino acid sequence comprises at least 90% sequence identity to the amino acid sequence set forth in part (c) of each of claims 110, 111, 112, 113, 114, 115, and 116.

95. The polypeptide of claim 85, 86, 87, 88, 89, 90, or 91, wherein the amino acid sequence comprises at least 95% sequence identity to the amino acid sequence set forth in part (c) of each of claims 110, 111, 112, 113, 114, 115, and 116.

96. The polypeptide of claims 1, 9, 13, 16, 20, 23, 27, 30, 34, 37, 41, 45, 46, 51, 56, 61, 66, 71, 72, 73, 74, 79, 80, 81, 85, 86, 87, 88, 89, 90, or 91, wherein the polypeptide

displays hydroxylase activity that converts an alkane to an alcohol.

97. The polypeptide of claim 96, wherein the alkane is selected from the group consisting of methane (CH_4), ethane (C_2H_6), propane (C_3H_8), butane (C_4H_{10}), pentane (C_5H_{12}), hexane (C_6H_{14}), heptane (C_7H_{16}), octane (C_8H_{18}), nonane (C_9H_{20}), decane ($\text{C}_{10}\text{H}_{22}$), undecane ($\text{C}_{11}\text{H}_{24}$), and dodecane ($\text{C}_{12}\text{H}_{26}$).

98. The polypeptide of claim 97, wherein the alkane is ethane.

99. The polypeptide of claim 97, wherein the alkane is methane.

100. The polypeptide of claim 96, wherein the alcohol is selected from the group consisting of methanol, ethanol, propanol, butanol, pentanol, hexanol, heptanol, octanol, nonanol, decanol, undecanol, and dodecanol.

101. The polypeptide of claim 100, wherein the alcohol is ethanol.

102. The polypeptide of claim 100, wherein the alcohol is methanol.

103. An isolated nucleic acid molecule selected from the group consisting of:

(a) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:7, 8, 9, 10, 11, 12, or 13;

(b) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:7, 8, 9, 10, 11, 12, or 13;

(c) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:7 or 11;

(d) a nucleic acid molecule which encodes a polypeptide comprising residues 456-1088 of the amino acid sequence set forth in SEQ ID NO:7, 8, 9, 10, 12, or 13;

(e) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:15;

(f) a nucleic acid molecule consisting of the nucleotide sequence set forth in SEQ ID NO:15;

(g) a nucleic acid molecule comprising the nucleotide sequence set forth in SEQ ID NO:17; and

(h) a nucleic acid molecule consisting of the nucleotide sequence set forth in SEQ ID NO:17.

104. An isolated nucleic acid molecule selected from the group consisting of:

(a) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:2 or 3 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227 is arginine (R); 238 is glutamine (Q); and 254 is glycine (G);

(b) a nucleic acid molecule which encodes a polypeptide consisting of residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:2 or 3 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227 is arginine (R); 238 is glutamine (Q); and 254 is glycine (G);

(c) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:4 with the following amino acid

residues: 49 and 207 are cysteine (C); 80 and 331 are phenylalanine (F); 84, 144, and 258 are serine (S); 96 and 177 are isoleucine (I); 186, 293, and 357 are valine (V); 228 is arginine (R); 239 is glutamine (Q); and 255 is glycine (G);

(d) a nucleic acid molecule which encodes a polypeptide consisting of residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:4 with the following amino acid residues: 49 and 207 are cysteine (C); 80 and 331 are phenylalanine (F); 84, 144, and 258 are serine (S); 96 and 177 are isoleucine (I); 186, 293, and 357 are valine (V); 228 is arginine (R); 239 is glutamine (Q); and 255 is glycine (G);

(e) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:5 with the following amino acid residues: 54 and 212 are cysteine (C); 85 and 336 are phenylalanine (F); 89, 149, and 262 are serine (S); 101 and 182 are isoleucine (I); 191, 297, and 361 are valine (V); 232 is arginine (R); 243 is glutamine (Q); and 259 is glycine (G);

(f) a nucleic acid molecule which encodes a polypeptide consisting of residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:5 with the following amino acid residues: 54 and 212 are cysteine (C); 85 and 336 are phenylalanine (F); 89, 149, and 262 are serine (S); 101 and 182 are isoleucine (I); 191, 297, and 361 are valine (V); 232 is arginine (R); 243 is glutamine (Q); and 259 is glycine (G);

(g) a nucleic acid molecule which encodes a polypeptide comprising residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:6 with the following amino acid residues: 51 and 208 are cysteine (C); 82 and 337 are phenylalanine (F); 86, 146, and 262 are serine (S); 98 and 179 are isoleucine (I); 188, 296, and 363 are valine (V); 231 is arginine (R); 243 is glutamine (Q); and 258 is glycine (G);
and

(h) a nucleic acid molecule which encodes a polypeptide consisting of residues 1 to about 455 of the amino acid sequence set forth in SEQ ID NO:6 with the following amino acid residues: 51 and 208 are cysteine (C); 82 and 337 are phenylalanine (F); 86, 146, and 262 are serine (S); 98 and 179 are isoleucine (I); 188, 296, and 363 are valine (V); 231 is arginine (R); 243 is glutamine (Q); and 258 is glycine (G).

105. An isolated nucleic acid molecule selected from the group consisting of:

(a) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:2 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227, 287, and 656 are arginine (R); 238 is glutamine (Q); 254 is glycine (G); 475 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 641 is asparagine (N); 721 is threonine (T); and 980 is leucine;

(b) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:2 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227, 287, and 656 are arginine (R); 238 is glutamine (Q); 254 is glycine (G); 475 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 641 is asparagine (N); 721 is threonine (T); and 980 is leucine;

(c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:3 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227, 287, and 648 are arginine (R);

238 is glutamine (Q); 254 is glycine (G); 467 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 634 is asparagine (N); 713 is threonine (T); and 972 is leucine;

(d) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:3 with the following amino acid residues: 48 and 206 are cysteine (C); 79 and 330 are phenylalanine (F); 83, 143, and 257 are serine (S); 95 and 176 are isoleucine (I); 185, 292, and 355 are valine (V); 227, 287, and 648 are arginine (R); 238 is glutamine (Q); 254 is glycine (G); 467 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 634 is asparagine (N); 713 is threonine (T); and 972 is leucine;

(e) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:4 with the following amino acid residues: 49 and 207 are cysteine (C); 80 and 331 are phenylalanine (F); 84, 144, and 258 are serine (S); 96 and 177 are isoleucine (I); 186, 293, and 357 are valine (V); 228, 288, and 659 are arginine (R); 239 is glutamine (Q); and 255 is glycine (G); 477 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 644 is asparagine (N); 724 is threonine (T); and 983 is leucine;

(f) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:4 with the following amino acid residues: 49 and 207 are cysteine (C); 80 and 331 are phenylalanine (F); 84, 144, and 258 are serine (S); 96 and 177 are isoleucine (I); 186, 293, and 357 are valine (V); 228, 288, and 659 are arginine (R); 239 is glutamine (Q); and 255 is glycine (G); 477 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 644 is asparagine (N); 724 is threonine (T); and 983 is leucine;

(g) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:5 with the following amino acid residues: 54 and 212 are cysteine (C); 85 and 336 are phenylalanine (F); 89, 149, and

262 are serine (S); 101 and 182 are isoleucine (I); 191, 297, and 361 are valine (V); 232, 292, and 656 are arginine (R); 243 is glutamine (Q); and 259 is glycine (G); 472 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 640 is asparagine (N); 723 is threonine (T); and 985 is leucine;

(h) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:5 with the following amino acid residues: 54 and 212 are cysteine (C); 85 and 336 are phenylalanine (F); 89, 149, and 262 are serine (S); 101 and 182 are isoleucine (I); 191, 297, and 361 are valine (V); 232, 292, and 656 are arginine (R); 243 is glutamine (Q); and 259 is glycine (G); 472 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 640 is asparagine (N); 723 is threonine (T); and 985 is leucine;

(i) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence set forth in SEQ ID NO:6 with the following amino acid residues: 51 and 208 are cysteine (C); 82 and 337 are phenylalanine (F); 86, 146, and 262 are serine (S); 98 and 179 are isoleucine (I); 188, 296, and 363 are valine (V); 231, 291, and 664 are arginine (R); 243 is glutamine (Q); and 258 is glycine (G); 480 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 648 is asparagine (N); 733 is threonine (T); and 997 is leucine; and

(j) a nucleic acid molecule which encodes a polypeptide consisting of the amino acid sequence set forth in SEQ ID NO:6 with the following amino acid residues: 51 and 208 are cysteine (C); 82 and 337 are phenylalanine (F); 86, 146, and 262 are serine (S); 98 and 179 are isoleucine (I); 188, 296, and 363 are valine (V); 231, 291, and 664 are arginine (R); 243 is glutamine (Q); and 258 is glycine (G); 480 is glycine (G), arginine (R), tyrosine (Y), or threonine (T); 648 is asparagine (N); 733 is threonine (T); and 997 is leucine.

106. An isolated nucleic acid molecule of claims 103, 104, or 105, further comprising a nucleotide sequence encoding a heterologous polypeptide.

107. A vector comprising the nucleic acid molecule of any one of claims 103, 104, or 105.

108. The vector of claim 107, which is an expression vector.

109. A host cell transfected with the vector of claim 108.

110. The host cell of claim 109, wherein the cell is procaryotic or eucaryotic.

111. The host cell of claim 110, wherein the eucaryotic cell is a yeast cell, an insect cell, or an animal cell.

112. The host cell of claim 110, wherein the procaryotic cell is a bacterial cell.

113. A method of producing a cell that converts an alkane to alcohol, the method comprising:

(a) transforming a cell with an isolated nucleic acid molecule encoding a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 7, 8, 9, 10, 11, 12, and 13; or

(b) transforming a cell with an isolated nucleic acid molecule encoding a polypeptide set forth in claim 1, 9, 13, 16, 20, 23, 27, 30, 34, 37, 41, 45, 46, 51, 56, 61, 66, 71, 72, 73, 74, 79, 80, 81, 85, 86, 87, 88, 89, 90, or 91; or

(c) transforming a cell with an isolated nucleic acid molecule set forth in claim 103, 104, or 105.

114. A method for selecting a cell that converts an alkane to an alcohol, the method comprising:

(a) providing a cell containing a nucleic acid construct comprising a nucleotide sequence that encodes a modified cytochrome P450 polypeptide, the nucleotide sequence selected from the group consisting of:

(i) a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 7, 8, 9, 10, 11, 12, or 13;

(ii) a nucleic acid molecule encoding a polypeptide set forth in claim 1, 9, 13, 16, 20, 23, 27, 30, 34, 37, 41, 45, 46, 51, 56, 61, 66, 71, 72, 73, 74, 79, 80, 81, 85, 86, 87, 88, 89, 90, or 91; or

(iii) a nucleic acid molecule set forth in claim 103, 104, or 105;

(b) culturing the cell in the presence of a suitable alkane and under conditions where the modified cytochrome P450 is expressed at a level sufficient to convert said alkane to an alcohol; and

(c) detecting the production of said alcohol.

115. A method for producing an alcohol, the method comprising:

(a) providing a cell containing a nucleic acid construct comprising a nucleotide sequence that encodes a modified cytochrome P450 polypeptide, the nucleotide sequence selected from the group consisting of:

(i) a nucleic acid molecule encoding a polypeptide comprising an amino acid sequence set forth in SEQ ID NO: 7, 8, 9, 10, 11, 12, or 13;

(ii) a nucleic acid molecule encoding a polypeptide set forth in claim 1, 9, 13, 16, 20, 23, 27, 30, 34, 37, 41, 45, 46, 51, 56, 61, 66, 71, 72, 73, 74, 79, 80, 81, 85, 86, 87, 88, 89, 90, or 91; or

(iii) a nucleic acid molecule set forth in claim 103, 104, or 105;

(b) culturing the cell in the presence of a suitable alkane and under conditions where the modified cytochrome P450 is expressed at a level sufficient to convert said alkane to an alcohol; and

(c) producing said alcohol by hydroxylation of the suitable alkane.

116. The method of claim 113, 114, or 115, wherein the cell is procaryotic or eucaryotic.

117. The method of claim 116, wherein the eucaryotic cell is a yeast cell, an insect cell, or an animal cell.

118. The method of claim 116, wherein the procaryotic cell is a bacterial cell.

119. A method for producing a cytochrome P450 variant that hydroxylates an alkane, the method comprising:

(a) modifying at least one amino acid residue positioned in or near the active site of the heme domain of the parent cytochrome P450 polypeptide, wherein the modification reduces the volume of the active site;

(b) contacting the polypeptide comprising the modified amino acid with at least one alkane under conditions suitable for hydroxylation of the alkane; and

(c) detecting a hydroxylated alkane, wherein said detection is indicative of a cytochrome P450 variant that hydroxylates an alkane.

120. The method of claim 119, wherein the modification is a substitution.

121. The method of claim 120, wherein the substitution comprises substituting a phenylalanine, tyrosine, histidine, or serine for the amino acid residue positioned in or near the active site of the heme domain of the parent cytochrome P450 polypeptide.

122. The method of claim 119, wherein the residue positioned in or near the active site of the heme domain of the parent cytochrome P450 polypeptide is an alanine.

123. The method of claim 122, wherein the alanine is residue 78, 82, and/or 328 of SEQ ID NO:1.

124. The method of claim 119, wherein the alkane is selected from the group consisting of methane (CH_4), ethane (C_2H_6), propane (C_3H_8), butane (C_4H_{10}), pentane (C_5H_{12}), hexane (C_6H_{14}), heptane (C_7H_{16}), octane (C_8H_{18}), nonane (C_9H_{20}), decane ($\text{C}_{10}\text{H}_{22}$), undecane ($\text{C}_{11}\text{H}_{24}$), and dodecane ($\text{C}_{12}\text{H}_{26}$).

125. The method of claim 124, wherein the alkane is ethane.

126. The method of claim 124, wherein the alkane is methane.

127. The method of claim 119, wherein the hydroxylated alkane is an alcohol.

128. The method of claim 127, wherein the alcohol is selected from the group consisting of methanol, ethanol, propanol, butanol, pentanol, hexanol, heptanol, octanol, nonanol, decanol, undecanol, and dodecanol.

129. The method of claim 128, wherein the alcohol is ethanol.

130. The method of claim 128, wherein the alcohol is methanol.