

NOVEL P450-BM3 VARIANTS WITH IMPROVED ACTIVITY

[0001] The present application claims priority to US Prov. Pat. Appln. Ser. No. 62/189,281, filed July 7, 2015, hereby incorporated by reference in its entirety for all purposes.

FIELD OF THE INVENTION

[0002] The present invention provides improved P450-BM3 variants with improved activity. In some embodiments, the P450-BM3 variants exhibit improved activity over a wide range of substrates.

REFERENCE TO SEQUENCE LISTING, TABLE OR COMPUTER PROGRAM

[0003] The official copy of the Sequence Listing is submitted concurrently with the specification as an ASCII formatted text file via EFS-Web, with a file name of "CX2-150USP1_ST25.txt", a creation date of July 7, 2015, and a size of 99.6 kilobytes. The Sequence Listing filed via EFS-Web is part of the specification and is incorporated in its entirety by reference herein.

BACKGROUND OF THE INVENTION

[0004] The cytochrome P450 monooxygenases ("P450s") comprise a large group of widely-distributed heme enzymes that are ubiquitous in the natural world. Cytochrome P450-BM3 ("P450-BM3"), obtained from *Bacillus megaterium* catalyzes the NADPH-dependent hydroxylation of long-chain fatty acids, alcohols, and amides, as well as the epoxidation of unsaturated fatty acids (See e.g., Narhi and Fulco, *J. Biol. Chem.*, 261:7160-7169 [1986]; and Capdevila et al., *J. Biol. Chem.*, 271:2263-22671 [1996]). P450-BM3 is unique, in that the reductase (65 kDa) and monooxygenase (55 kDa) domains of the enzyme are fused and produced as a catalytically self-sufficient 120 kDa enzyme. Although these enzymes have been the subject of numerous studies, there remains a need in the art for improved P450s that exhibit high levels of enzymatic activity over a wide range of substrates.

SUMMARY OF THE INVENTION

[0005] The present invention provides improved P450-BM3 variants with improved activity. In some embodiments, the P450-BM3 variants exhibit improved activity over a wide range of substrates. A recombinant cytochrome P450-BM3 variant having at least 90% sequence identity to a polypeptide sequence comprising the sequence set forth in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, or 16. In some embodiments, the recombinant cytochrome P450-BM3 variants of the present invention oxidize at least three organic substrates. In some further embodiments, the recombinant cytochrome P450-BM3 variants oxidize at least one organic substrate selected from loratadine, imatinib, gefitinib, and diclofenac.

[0006] The present invention further provides isolated recombinant polynucleotide sequences encoding the recombinant cytochrome P450-BM3 polypeptide variants provided herein. In some embodiments, the isolated recombinant polynucleotide sequence comprises SEQ ID NO:1, 3, 5, 7, 9, 11, 13, or 15.

[0007] The present invention also provides expression vectors comprising at least one polynucleotide sequence provided herein. In some additional embodiments, the vector comprises at least one polynucleotide sequence that is operably linked with at least one regulatory sequence suitable for expression of the polynucleotide sequence in a suitable host cell. In some embodiments, the host cell is a prokaryotic or eukaryotic cell. In some additional embodiments, the host cell is a prokaryotic cell. In some further embodiments, the host cell is *E. coli*. The present invention also provides host cells comprising the vectors provided herein.

[0008] The present invention also provides methods for producing at least one recombinant cytochrome P450-BM3 variant comprising culturing the host cell provided herein under conditions such that at least one of the recombinant cytochrome P450-BM3 variants provided herein is produced by the host cell. In some additional embodiments, the methods further comprise the step of recovering at least one recombinant cytochrome P450 variant.

[0008A] The present invention also provides a recombinant cytochrome P450-BM3 variant having at least 80% sequence identity to a polypeptide sequence comprising the sequence set forth in SEQ ID NO:2.

[0008B] The present invention also provides a recombinant cytochrome P450-BM3 variant having at least 80% sequence identity to a polypeptide sequence comprising the sequence set forth in SEQ ID NO:2, wherein the variant has at least 90% sequence identity to a polypeptide sequence comprising the sequence set forth in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, or 16.

[0008C] The present invention also provides a recombinant cytochrome P450-BM3 variant having at least 80% sequence identity to a polypeptide sequence comprising the sequence set forth in SEQ ID NO:2, wherein the variant has at least 90% sequence identity to a polypeptide sequence comprising the sequence set forth in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, or 16, and wherein the variant has a polypeptide sequence comprising the sequence selected from the group comprising:

SEQ ID NO:2 comprising substitutions V51S; D118R; L219Y; K236H; P351T; S353T; N579T; T582A; E624D; A632S; D727S; and Q931K;

SEQ ID NO:2 comprising substitutions V51S; D118R; T179V; L219Y; P351T; N579T; T582A; E624D; D727S;

SEQ ID NO:2 comprising substitutions D118R; L219Y; K236H; S353T; N579T; T582A; E624D; D727S;

SEQ ID NO:4 comprising substitutions C48S; I95P; G115R; L216Y; D232H; M491A; N574T; and D722S;

SEQ ID NO:4 comprising substitutions K32R; C48S; I95P; L216Y; D232H; E349T; M491A; N574T; T577A; E619D; and D722S;

SEQ ID NO:6 comprising substitutions K32R; C48S; Y52F; G115R; L216Y; D232H; E349K; M491A; T577A; E619D; A627S; and D722S;

SEQ ID NO:6 comprising substitutions K32R; C48S; Y52F; I95P; Q111R; G115R; I176V; L216Y; D232H; E349T; M491A; T577A; E619D; and D722S;

SEQ ID NO:8 comprising substitutions Y52F; D232H; E349K; E619D; D722S; R762H;

SEQ ID NO:8 comprising substitutions K32R; Y52F; K95P; Q111R; G115R; I176V; L216Y; D232H; P347T; E349T; M491A; T577A; E619D; and D722S;

SEQ ID NO:8 comprising substitutions K95P; Q111R; G115R; I176V; L216Y; E349T; M491A; T577A; E619D; and D722S;

SEQ ID NO:10 comprising substitutions K32R; C48S; Y52F; I95P; L216Y; D232H; E349K; M491A; N574T; and D722S;

SEQ ID NO:10 comprising substitutions K32R; C48S; Y52F; I95P; Q111R; L216Y; D232H; P347T; E349K; M491A; N574T; E619D; and D722S;

SEQ ID NO:10 comprising substitutions C48S; I95P; G115R; L216Y; S231R; D232H; M491A; N574T; T577A; E619D; and D722S;

SEQ ID NO:12 comprising substitutions K32R; C48S; Y52F; I95P; Q111R; G115R; D232H; M491A; T577A; A627S; D722S;

SEQ ID NO:12 comprising substitutions K32R; C48S; Y52F; I95P; G115R; D232H; M491A; N574T; T577A; E619D; A627S; and D722S;

SEQ ID NO:14 comprising substitutions Y52F; K95P; G115R; T176V; L216Y; D232H; E349T; M491A; T577A; and E619D;

SEQ ID NO:14 comprising substitutions Y52F; G115R; T176V; L216Y; D232H; P347T; E349T; T577A; D722S;

SEQ ID NO:16 comprising substitutions K32R; C48S; Y52F; I95P; I176V; L216Y; D232H; E349K; M491A; N574T; E619D; D722S; A767T; and

SEQ ID NO:16 comprising substitutions C48S; Y52F; I95P; Q111R; I176V; L216Y; D232H; P347T; M491A; N574T; and D722S.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] Figure 1 provides the structures of the substrates used in the screening methods described herein. Diclofenac was used for HTP screening to detect/rank beneficial diversity. The remaining substrates were used to validate that the evolved BM3 variants were active on substrates that were not used for HTP screening.

[0010] Figure 2 provides a graphical summary of the results obtained for the MCYP-1.2-A12 lineage.

[0011] Figure 3 provides a graphical summary of the results obtained for the MCYP-1.2.-A05 lineage.

DESCRIPTION OF THE INVENTION

[0012] The present invention provides improved P450-BM3 variants with improved activity. In some embodiments, the P450-BM3 variants exhibit improved activity over a wide range of substrates. P450-BM3 enzymes exhibit the highest rate of catalysis amongst P450 monooxygenases due to the efficient electron transfer between the fused reductase and heme domains (See e.g., Noble et al., *Biochem. J.*, 339:371-379 [1999]; and Munro et al., *Eur. J. Biochem.*, 239:403-409 [2009]). Thus, P450-BM3 is a highly desirable enzyme for the manipulation of biotechnological processes (See e.g., Sawayama et al., *Chem.*, 15:11723-11729 [2009]; Otey et al., *Biotechnol. Bioeng.*, 93:494-499 [2006]; Damsten et al., *Biol. Interact.*, 171:96-107 [2008]; and Di Nardo and Gilardi, *Int. J. Mol. Sci.*, 13:15901-15924). However, there still remains a need in the art for P450 enzymes that exhibit

activity over a broad range of substrates. The present invention provides P450-BM3 variants that have improved enzymatic activity over a broad range of substrates, as compared to a parental P450-BM3 sequence (i.e., SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16).

[0013] In some embodiments, the present invention provides P450-BM3 variants that provide improved total percent conversion/turnover number for the oxidation of multiple organic substrates (See e.g., Figure 1). In particular, during the development of the present invention, beneficial diversity was identified and recombined based on HTP screening results.

Abbreviations and Definitions:

[0014] Unless defined otherwise, all technical and scientific terms used herein generally have the same meaning as commonly understood by one of ordinary skill in the art to which this invention pertains. Generally, the nomenclature used herein and the laboratory procedures of cell culture, molecular genetics, microbiology, organic chemistry, analytical chemistry and nucleic acid chemistry described below are those well-known and commonly employed in the art. Such techniques are well-known and described in numerous texts and reference works well known to those of skill in the art. Standard techniques, or modifications thereof, are used for chemical syntheses and chemical analyses. All patents, patent applications, articles and publications mentioned herein, both *supra* and *infra*, are hereby expressly incorporated herein by reference.

[0015] Although any suitable methods and materials similar or equivalent to those described herein find use in the practice of the present invention, some methods and materials are described herein. It is to be understood that this invention is not limited to the particular methodology, protocols, and reagents described, as these may vary, depending upon the context they are used by those of skill in the art. Accordingly, the terms defined immediately below are more fully described by reference to the application as a whole. All patents, patent applications, articles and publications mentioned herein, both *supra* and *infra*, are hereby expressly incorporated herein by reference.

[0016] Also, as used herein, the singular "a", "an," and "the" include the plural references, unless the context clearly indicates otherwise.

[0017] Numeric ranges are inclusive of the numbers defining the range. Thus, every numerical range disclosed herein is intended to encompass every narrower numerical range that falls within such broader numerical range, as if such narrower numerical ranges were all expressly written herein. It is also intended that every maximum (or minimum) numerical limitation disclosed herein includes every lower (or higher) numerical limitation, as if such lower (or higher) numerical limitations were expressly written herein.

[0018] The term "about" means an acceptable error for a particular value. In some instances "about" means within 0.05%, 0.5%, 1.0%, or 2.0%, of a given value range. In some instances, "about" means within 1, 2, 3, or 4 standard deviations of a given value.

[0019] Furthermore, the headings provided herein are not limitations of the various aspects or embodiments of the invention which can be had by reference to the application as a whole. Accordingly, the terms defined immediately below are more fully defined by reference to the application as a whole. Nonetheless, in order to facilitate understanding of the invention, a number of terms are defined below.

[0020] Unless otherwise indicated, nucleic acids are written left to right in 5' to 3' orientation; amino acid sequences are written left to right in amino to carboxy orientation, respectively.

[0021] As used herein, the term “comprising” and its cognates are used in their inclusive sense (*i.e.*, equivalent to the term “including” and its corresponding cognates).

[0022] “EC” number refers to the Enzyme Nomenclature of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (NC-IUBMB). The IUBMB biochemical classification is a numerical classification system for enzymes based on the chemical reactions they catalyze.

[0023] “ATCC” refers to the American Type Culture Collection whose biorepository collection includes genes and strains.

[0024] “NCBI” refers to National Center for Biological Information and the sequence databases provided therein.

[0025] As used herein “cytochrome P450-BM3” and “P450-BM3” refer to the cytochrome P450 enzyme obtained from *Bacillus megaterium* that catalyzes the NADPH-dependent hydroxylation of long-chain fatty acids, alcohols, and amides, as well as the epoxidation of unsaturated fatty acids.

[0026] “Protein,” “polypeptide,” and “peptide” are used interchangeably herein to denote a polymer of at least two amino acids covalently linked by an amide bond, regardless of length or post-translational modification (e.g., glycosylation or phosphorylation).

[0027] “Amino acids” are referred to herein by either their commonly known three-letter symbols or by the one-letter symbols recommended by IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single letter codes.

[0028] The term “engineered,” “recombinant,” “non-naturally occurring,” and “variant,” when used with reference to a cell, a polynucleotide or a polypeptide refers to a material or a material corresponding to the natural or native form of the material that has been modified in a manner that would not otherwise exist in nature or is identical thereto but produced or derived from synthetic materials and/or by manipulation using recombinant techniques.

[0029] As used herein, “wild-type” and “naturally-occurring” refer to the form found in nature. For example a wild-type polypeptide or polynucleotide sequence is a sequence present in an organism that can be isolated from a source in nature and which has not been intentionally modified by human manipulation.

[0030] “Coding sequence” refers to that part of a nucleic acid (e.g., a gene) that encodes an amino acid sequence of a protein.

[0031] The term “percent (%) sequence identity” is used herein to refer to comparisons among polynucleotides and polypeptides, and are determined by comparing two optimally aligned sequences over a comparison window, wherein the portion of the polynucleotide or polypeptide sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) as compared to the reference sequence for optimal alignment of the two sequences. The percentage may be calculated by determining the number of positions at which the identical nucleic acid base or amino acid residue occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Alternatively, the percentage may be calculated by determining the number of positions at which either the identical nucleic acid base or amino acid residue occurs in both sequences or a nucleic acid base or amino acid residue is aligned with a gap to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the window of comparison and multiplying the result by 100 to yield the percentage of sequence identity. Those of skill in the art appreciate that there are many established algorithms available to align two sequences. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith and Waterman (Smith and Waterman, *Adv. Appl. Math.*, 2:482 [1981]), by the homology alignment algorithm of Needleman and Wunsch (Needleman and Wunsch, *J. Mol. Biol.*, 48:443 [1970]), by the search for similarity method of Pearson and Lipman (Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444 [1988]), by computerized implementations of these algorithms (*e.g.*, GAP, BESTFIT, FASTA, and TFASTA in the GCG Wisconsin Software Package), or by visual inspection, as known in the art. Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity include, but are not limited to the BLAST and BLAST 2.0 algorithms, which are described by Altschul et al. (See, Altschul et al., *J. Mol. Biol.*, 215: 403-410 [1990]; and Altschul et al., 1977, *Nucl. Acids Res.*, 3389-3402 [1977], respectively). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information website. This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length *W* in the query sequence, which either match or satisfy some positive-valued threshold score *T* when aligned with a word of the same length in a database sequence. *T* is referred to as, the neighborhood word score threshold (See, Altschul et al, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters *M* (reward score for a pair of matching residues; always >0) and *N* (penalty score for mismatching residues; always <0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity *X* from its maximum achieved value; the cumulative score goes to zero or below, due to the accumulation of one or more negative-scoring

residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (See, Henikoff and Henikoff, Proc. Natl. Acad. Sci. USA 89:10915 [1989]). Exemplary determination of sequence alignment and % sequence identity can employ the BESTFIT or GAP programs in the GCG Wisconsin Software package (Accelrys, Madison WI), using default parameters provided.

[0032] “Reference sequence” refers to a defined sequence used as a basis for a sequence comparison. A reference sequence may be a subset of a larger sequence, for example, a segment of a full-length gene or polypeptide sequence. Generally, a reference sequence is at least 20 nucleotide or amino acid residues in length, at least 25 residues in length, at least 50 residues in length, at least 100 residues in length or the full length of the nucleic acid or polypeptide. Since two polynucleotides or polypeptides may each (1) comprise a sequence (*i.e.*, a portion of the complete sequence) that is similar between the two sequences, and (2) may further comprise a sequence that is divergent between the two sequences, sequence comparisons between two (or more) polynucleotides or polypeptides are typically performed by comparing sequences of the two polynucleotides or polypeptides over a “comparison window” to identify and compare local regions of sequence similarity. In some embodiments, a “reference sequence” can be based on a primary amino acid sequence, where the reference sequence is a sequence that can have one or more changes in the primary sequence. “Comparison window” refers to a conceptual segment of at least about 20 contiguous nucleotide positions or amino acid residues wherein a sequence may be compared to a reference sequence of at least 20 contiguous nucleotides or amino acids and wherein the portion of the sequence in the comparison window may comprise additions or deletions (*i.e.*, gaps) of 20 percent or less as compared to the reference sequence (which does not comprise additions or deletions) for optimal alignment of the two sequences. The comparison window can be longer than 20 contiguous residues, and includes, optionally 30, 40, 50, 100, or longer windows.

[0033] “Corresponding to”, “reference to” or “relative to” when used in the context of the numbering of a given amino acid or polynucleotide sequence refers to the numbering of the residues of a specified reference sequence when the given amino acid or polynucleotide sequence is compared to the reference sequence. In other words, the residue number or residue position of a given polymer is designated with respect to the reference sequence rather than by the actual numerical position of the residue within the given amino acid or polynucleotide sequence. For example, a given amino acid sequence, such as that of an engineered P450-BM3, can be aligned to a reference sequence by introducing gaps to optimize residue matches between the two sequences. In these cases, although the gaps are present, the numbering of the residue in the given amino acid or polynucleotide sequence is made with respect to the reference sequence to which it has been aligned.

[0034] “Amino acid difference” or “residue difference” refers to a difference in the amino acid residue at a position of a polypeptide sequence relative to the amino acid residue at a corresponding position in a reference sequence. The positions of amino acid differences generally are referred to herein as “X_n,” where n refers to the corresponding position in the reference sequence upon which the residue difference is based. For example, a “residue difference at position X₉₃ as compared to SEQ ID NO:2” refers to a difference of the amino acid residue at the polypeptide position corresponding to position 93 of SEQ ID NO:2. Thus, if the reference polypeptide of SEQ ID NO:2 has a serine at position 93, then a “residue difference at position X₉₃ as compared to SEQ ID NO:2” an amino acid substitution of any residue other than serine at the position of the polypeptide corresponding to position 93 of SEQ ID NO:2. In most instances herein, the specific amino acid residue difference at a position is indicated as “X_nY” where “X_n” specified the corresponding position as described above, and “Y” is the single letter identifier of the amino acid found in the engineered polypeptide (*i.e.*, the different residue than in the reference polypeptide). In some instances (*e.g.*, in Tables 2-9), the present disclosure also provides specific amino acid differences denoted by the conventional notation “A_nB”, where A is the single letter identifier of the residue in the reference sequence, “n” is the number of the residue position in the reference sequence, and B is the single letter identifier of the residue substitution in the sequence of the engineered polypeptide. In some instances, a polypeptide of the present disclosure can include one or more amino acid residue differences relative to a reference sequence, which is indicated by a list of the specified positions where residue differences are present relative to the reference sequence. In some embodiments, where more than one amino acid can be used in a specific residue position of a polypeptide, the various amino acid residues that can be used are separated by a “/” (*e.g.*, X₃₀₇H/X₃₀₇P or X₃₀₇H/P). The present application includes engineered polypeptide sequences comprising one or more amino acid differences that include either/or both conservative and non-conservative amino acid substitutions.

[0035] “Conservative amino acid substitution” refers to a substitution of a residue with a different residue having a similar side chain, and thus typically involves substitution of the amino acid in the polypeptide with amino acids within the same or similar defined class of amino acids. By way of example and not limitation, an amino acid with an aliphatic side chain may be substituted with another aliphatic amino acid (*e.g.*, alanine, valine, leucine, and isoleucine); an amino acid with hydroxyl side chain is substituted with another amino acid with a hydroxyl side chain (*e.g.*, serine and threonine); an amino acids having aromatic side chains is substituted with another amino acid having an aromatic side chain (*e.g.*, phenylalanine, tyrosine, tryptophan, and histidine); an amino acid with a basic side chain is substituted with another amino acid with a basis side chain (*e.g.*, lysine and arginine); an amino acid with an acidic side chain is substituted with another amino acid with an acidic side chain (*e.g.*, aspartic acid or glutamic acid); and/or a hydrophobic or hydrophilic amino acid is replaced with another hydrophobic or hydrophilic amino acid, respectively.

[0036] “Non-conservative substitution” refers to substitution of an amino acid in the polypeptide with an amino acid with significantly differing side chain properties. Non-conservative substitutions may use amino acids between, rather than within, the defined groups and affects (a) the structure of the peptide backbone in the area of the substitution (e.g., proline for glycine) (b) the charge or hydrophobicity, or (c) the bulk of the side chain. By way of example and not limitation, an exemplary non-conservative substitution can be an acidic amino acid substituted with a basic or aliphatic amino acid; an aromatic amino acid substituted with a small amino acid; and a hydrophilic amino acid substituted with a hydrophobic amino acid.

[0037] “Deletion” refers to modification to the polypeptide by removal of one or more amino acids from the reference polypeptide. Deletions can comprise removal of 1 or more amino acids, 2 or more amino acids, 5 or more amino acids, 10 or more amino acids, 15 or more amino acids, or 20 or more amino acids, up to 10% of the total number of amino acids, or up to 20% of the total number of amino acids making up the reference enzyme while retaining enzymatic activity and/or retaining the improved properties of an engineered P450-BM3 enzyme. Deletions can be directed to the internal portions and/or terminal portions of the polypeptide. In various embodiments, the deletion can comprise a continuous segment or can be discontinuous.

[0038] “Insertion” refers to modification to the polypeptide by addition of one or more amino acids from the reference polypeptide. Insertions can be in the internal portions of the polypeptide, or to the carboxy or amino terminus. Insertions as used herein include fusion proteins as is known in the art. The insertion can be a contiguous segment of amino acids or separated by one or more of the amino acids in the naturally occurring polypeptide.

[0039] A “functional fragment” or a “biologically active fragment” used interchangeably herein refers to a polypeptide that has an amino-terminal and/or carboxy-terminal deletion(s) and/or internal deletions, but where the remaining amino acid sequence is identical to the corresponding positions in the sequence to which it is being compared (e.g., a full-length engineered P450-BM3 of the present invention) and that retains substantially all of the activity of the full-length polypeptide.

[0040] “Isolated polypeptide” refers to a polypeptide which is substantially separated from other contaminants that naturally accompany it, e.g., protein, lipids, and polynucleotides. The term embraces polypeptides which have been removed or purified from their naturally-occurring environment or expression system (e.g., host cell or *in vitro* synthesis). The recombinant P450-BM3 polypeptides may be present within a cell, present in the cellular medium, or prepared in various forms, such as lysates or isolated preparations. As such, in some embodiments, the recombinant P450-BM3 polypeptides can be an isolated polypeptide.

[0041] “Substantially pure polypeptide” refers to a composition in which the polypeptide species is the predominant species present (i.e., on a molar or weight basis it is more abundant than any other individual macromolecular species in the composition), and is generally a substantially purified composition when the object species comprises at least about 50 percent of the macromolecular

species present by mole or % weight. However, in some embodiments, the composition comprising P450-BM3 comprises P450-BM3 that is less than 50% pure (e.g., about 10%, about 20%, about 30%, about 40%, or about 50%). Generally, a substantially pure P450-BM3 composition comprises about 60% or more, about 70% or more, about 80% or more, about 90% or more, about 95% or more, and about 98% or more of all macromolecular species by mole or % weight present in the composition. In some embodiments, the object species is purified to essential homogeneity (i.e., contaminant species cannot be detected in the composition by conventional detection methods) wherein the composition consists essentially of a single macromolecular species. Solvent species, small molecules (<500 Daltons), and elemental ion species are not considered macromolecular species. In some embodiments, the isolated recombinant P450-BM3 polypeptides are substantially pure polypeptide compositions.

[0042] “Improved enzyme property” refers to an engineered P450-BM3 polypeptide that exhibits an improvement in any enzyme property as compared to a reference P450-BM3 polypeptide and/or a wild-type P450-BM3 polypeptide or another engineered P450-BM3 polypeptide. Improved properties include but are not limited to such properties as increased protein expression, increased thermostability, increased pH activity, increased stability, increased enzymatic activity, increased substrate specificity or affinity, increased specific activity, increased resistance to substrate or end-product inhibition, increased chemical stability, improved chemoselectivity, improved solvent stability, increased tolerance to acidic pH, increased tolerance to proteolytic activity (i.e., reduced sensitivity to proteolysis), reduced aggregation, increased solubility, and altered temperature profile.

[0043] “Increased enzymatic activity” or “enhanced catalytic activity” refers to an improved property of the engineered P450-BM3 polypeptides, which can be represented by an increase in specific activity (e.g., product produced/time/weight protein) or an increase in percent conversion of the substrate to the product (e.g., percent conversion of starting amount of substrate to product in a specified time period using a specified amount of P450-BM3) as compared to the reference P450-BM3 enzyme. Exemplary methods to determine enzyme activity are provided in the Examples. Any property relating to enzyme activity may be affected, including the classical enzyme properties of K_m , V_{max} or k_{cat} , changes of which can lead to increased enzymatic activity. Improvements in enzyme activity can be from about 1.1 fold the enzymatic activity of the corresponding wild-type enzyme, to as much as 2-fold, 5-fold, 10-fold, 20-fold, 25-fold, 50-fold, 75-fold, 100-fold, 150-fold, 200-fold or more enzymatic activity than the naturally occurring P450-BM3 or another engineered P450-BM3 from which the P450-BM3 polypeptides were derived.

[0044] “Conversion” refers to the enzymatic conversion (or biotransformation) of a substrate(s) to the corresponding product(s). “Percent conversion” refers to the percent of the substrate that is converted to the product within a period of time under specified conditions. Thus, the “enzymatic

activity” or “activity” of a P450-BM3 polypeptide can be expressed as “percent conversion” of the substrate to the product in a specific period of time.

[0045] Enzymes with “generalist properties” (or “generalist enzymes”) refer to enzymes that exhibit improved activity for a wide range of substrates, as compared to a parental sequence. Generalist enzymes do not necessarily demonstrate improved activity for every possible substrate. In particular, the present invention provides P450-BM3 variants with generalist properties, in that they demonstrate similar or improved activity relative to the parental gene for a wide range of sterically and electronically diverse substrates. In addition, the generalist enzymes provided herein were engineered to be improved across a wide range of diverse API-like molecules to increase the production of metabolites/products.

[0046] “Hybridization stringency” relates to hybridization conditions, such as washing conditions, in the hybridization of nucleic acids. Generally, hybridization reactions are performed under conditions of lower stringency, followed by washes of varying but higher stringency. The term “moderately stringent hybridization” refers to conditions that permit target-DNA to bind a complementary nucleic acid that has about 60% identity, preferably about 75% identity, about 85% identity to the target DNA, with greater than about 90% identity to target-polynucleotide. Exemplary moderately stringent conditions are conditions equivalent to hybridization in 50% formamide, 5× Denhart's solution, 5×SSPE, 0.2% SDS at 42°C, followed by washing in 0.2×SSPE, 0.2% SDS, at 42°C. “High stringency hybridization” refers generally to conditions that are about 10°C or less from the thermal melting temperature T_m as determined under the solution condition for a defined polynucleotide sequence. In some embodiments, a high stringency condition refers to conditions that permit hybridization of only those nucleic acid sequences that form stable hybrids in 0.018M NaCl at 65°C (*i.e.*, if a hybrid is not stable in 0.018M NaCl at 65°C, it will not be stable under high stringency conditions, as contemplated herein). High stringency conditions can be provided, for example, by hybridization in conditions equivalent to 50% formamide, 5× Denhart's solution, 5×SSPE, 0.2% SDS at 42°C, followed by washing in 0.1×SSPE, and 0.1% SDS at 65°C. Another high stringency condition is hybridizing in conditions equivalent to hybridizing in 5X SSC containing 0.1% (w:v) SDS at 65°C and washing in 0.1x SSC containing 0.1% SDS at 65°C. Other high stringency hybridization conditions, as well as moderately stringent conditions, are described in the references cited above.

[0047] “Codon optimized” refers to changes in the codons of the polynucleotide encoding a protein to those preferentially used in a particular organism such that the encoded protein is more efficiently expressed in the organism of interest. Although the genetic code is degenerate in that most amino acids are represented by several codons, called “synonyms” or “synonymous” codons, it is well known that codon usage by particular organisms is nonrandom and biased towards particular codon triplets. This codon usage bias may be higher in reference to a given gene, genes of common function or ancestral origin, highly expressed proteins versus low copy number proteins, and the aggregate

protein coding regions of an organism's genome. In some embodiments, the polynucleotides encoding the P450-BM3 enzymes may be codon optimized for optimal production from the host organism selected for expression.

[0048] "Control sequence" refers herein to include all components, which are necessary or advantageous for the expression of a polynucleotide and/or polypeptide of the present application. Each control sequence may be native or foreign to the nucleic acid sequence encoding the polypeptide. Such control sequences include, but are not limited to, a leader, polyadenylation sequence, propeptide sequence, promoter sequence, signal peptide sequence, initiation sequence and transcription terminator. At a minimum, the control sequences include a promoter, and transcriptional and translational stop signals. The control sequences may be provided with linkers for the purpose of introducing specific restriction sites facilitating ligation of the control sequences with the coding region of the nucleic acid sequence encoding a polypeptide.

[0049] "Operably linked" is defined herein as a configuration in which a control sequence is appropriately placed (i.e., in a functional relationship) at a position relative to a polynucleotide of interest such that the control sequence directs or regulates the expression of the polynucleotide and/or polypeptide of interest.

[0050] "Promoter sequence" refers to a nucleic acid sequence that is recognized by a host cell for expression of a polynucleotide of interest, such as a coding sequence. The promoter sequence contains transcriptional control sequences, which mediate the expression of a polynucleotide of interest. The promoter may be any nucleic acid sequence which shows transcriptional activity in the host cell of choice including mutant, truncated, and hybrid promoters, and may be obtained from genes encoding extracellular or intracellular polypeptides either homologous or heterologous to the host cell.

[0051] "Suitable reaction conditions" refers to those conditions in the enzymatic conversion reaction solution (e.g., ranges of enzyme loading, substrate loading, temperature, pH, buffers, co-solvents, etc.) under which a P450-BM3 polypeptide of the present application is capable of converting a substrate to the desired product compound. Exemplary "suitable reaction conditions" are provided in the present application and illustrated by the Examples. "Loading", such as in "compound loading" or "enzyme loading" refers to the concentration or amount of a component in a reaction mixture at the start of the reaction. "Substrate" in the context of an enzymatic conversion reaction process refers to the compound or molecule acted on by the P450-BM3 polypeptide. "Product" in the context of an enzymatic conversion process refers to the compound or molecule resulting from the action of the P450-BM3 polypeptide on a substrate.

[0052] As used herein the term "culturing" refers to the growing of a population of microbial cells under any suitable conditions (e.g., using a liquid, gel or solid medium).

[0053] Recombinant polypeptides can be produced using any suitable methods known in the art. Genes encoding the wild-type polypeptide of interest can be cloned in vectors, such as plasmids, and expressed in desired hosts, such as *E. coli*, etc. Variants of recombinant polypeptides can be generated

by various methods known in the art. Indeed, there is a wide variety of different mutagenesis techniques well known to those skilled in the art. In addition, mutagenesis kits are also available from many commercial molecular biology suppliers. Methods are available to make specific substitutions at defined amino acids (site-directed), specific or random mutations in a localized region of the gene (regio-specific), or random mutagenesis over the entire gene (*e.g.*, saturation mutagenesis).

Numerous suitable methods are known to those in the art to generate enzyme variants, including but not limited to site-directed mutagenesis of single-stranded DNA or double-stranded DNA using PCR, cassette mutagenesis, gene synthesis, error-prone PCR, shuffling, and chemical saturation mutagenesis, or any other suitable method known in the art. Non-limiting examples of methods used for DNA and protein engineering are provided in the following patents: US Pat. No. 6,117,679; US Pat. No. 6,420,175; US Pat. No. 6,376,246; US Pat. No. 6,586,182; US Pat. No. 7,747,391; US Pat. No. 7,747,393; US Pat. No. 7,783,428; and US Pat. No. 8,383,346. After the variants are produced, they can be screened for any desired property (*e.g.*, high or increased activity, or low or reduced activity, increased thermal activity, increased thermal stability, and/or acidic pH stability, etc.). In some embodiments, "recombinant P450-BM3 polypeptides" (also referred to herein as "engineered P450-BM3 polypeptides," "variant P450-BM3 enzymes," and "P450-BM3 variants") find use.

[0054] As used herein, a "vector" is a DNA construct for introducing a DNA sequence into a cell. In some embodiments, the vector is an expression vector that is operably linked to a suitable control sequence capable of effecting the expression in a suitable host of the polypeptide encoded in the DNA sequence. In some embodiments, an "expression vector" has a promoter sequence operably linked to the DNA sequence (*e.g.*, transgene) to drive expression in a host cell, and in some embodiments, also comprises a transcription terminator sequence.

[0055] As used herein, the term "expression" includes any step involved in the production of the polypeptide including, but not limited to, transcription, post-transcriptional modification, translation, and post-translational modification. In some embodiments, the term also encompasses secretion of the polypeptide from a cell.

[0056] As used herein, the term "produces" refers to the production of proteins and/or other compounds by cells. It is intended that the term encompass any step involved in the production of polypeptides including, but not limited to, transcription, post-transcriptional modification, translation, and post-translational modification. In some embodiments, the term also encompasses secretion of the polypeptide from a cell.

[0057] As used herein, an amino acid or nucleotide sequence (*e.g.*, a promoter sequence, signal peptide, terminator sequence, etc.) is "heterologous" to another sequence with which it is operably linked if the two sequences are not associated in nature.

[0058] As used herein, the terms "host cell" and "host strain" refer to suitable hosts for expression vectors comprising DNA provided herein (*e.g.*, the polynucleotides encoding the P450-BM3 variants).

In some embodiments, the host cells are prokaryotic or eukaryotic cells that have been transformed or transfected with vectors constructed using recombinant DNA techniques as known in the art.

[0059] The term “analogue” means a polypeptide having more than 70% sequence identity but less than 100% sequence identity (e.g., more than 75%, 78%, 80%, 83%, 85%, 88%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% sequence identity) with a reference polypeptide. In some embodiments, analogues means polypeptides that contain one or more non-naturally occurring amino acid residues including, but not limited, to homoarginine, ornithine and norvaline, as well as naturally occurring amino acids. In some embodiments, analogues also include one or more D-amino acid residues and non-peptide linkages between two or more amino acid residues.

[0060] The term “effective amount” means an amount sufficient to produce the desired result. One of general skill in the art may determine what the effective amount by using routine experimentation.

[0061] The terms “isolated” and “purified” are used to refer to a molecule (e.g., an isolated nucleic acid, polypeptide, etc.) or other component that is removed from at least one other component with which it is naturally associated. The term “purified” does not require absolute purity, rather it is intended as a relative definition.

Engineered P450-BM3 Polypeptides:

[0062] In some embodiments, engineered P450-BM3 polypeptides are produced by cultivating a microorganism comprising at least one polynucleotide sequence encoding at least one engineered P450-BM3 polypeptide under conditions which are conducive for producing the engineered P450-BM3 polypeptide(s). In some embodiments, the engineered P450-BM3 polypeptide is recovered from the resulting culture medium and/or cells.

[0063] The present invention provides exemplary engineered P450-BM3 polypeptides having P450-BM3 activity (i.e., P450-BM3 variants). The Examples provide Tables showing sequence structural information correlating specific amino acid sequence features with the functional activity of the engineered P450-BM3 polypeptides. This structure-function correlation information is provided in the form of specific amino acid residues differences relative to a reference engineered polypeptide, as indicated in the Examples. The Examples further provide experimentally determined activity data for the exemplary engineered P450-BM3 polypeptides.

[0064] In some embodiments, the engineered P450-BM3 polypeptides of the invention having P450-BM3 activity comprise: a) an amino acid sequence having at least 85% sequence identity to reference sequence SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16; b) an amino acid residue difference as compared to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16 at one or more amino acid positions; and c) which exhibits an improved property selected from i) enhanced catalytic activity, ii) reduced proteolytic sensitivity, iii) increased tolerance to acidic pH, iv) reduced aggregation, v) increased

activity on a range of substrates (i.e., enzymes with a broad substrate range), or a combination of any of i), ii), iii) or iv), as compared to the reference sequence.

[0065] In some embodiments the engineered P450-BM3 which exhibits an improved property has at least about 85%, at least about 88%, at least about 90%, at least about 91%, at least about 92%, at least about 93%, at least about 94%, at least about 95%, at least about 96%, at least about 97%, at least about 98%, at least about 99%, or at about 100% amino acid sequence identity with SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16, and an amino acid residue difference as compared to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16, at one or more amino acid positions (such as at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 20 or more amino acid positions compared to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16, or a sequence having at least 85%, at least 88%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or greater amino acid sequence identity with SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16). In some embodiment the residue difference as compared to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16, at one or more positions will include at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more conservative amino acid substitutions. In some embodiments, the engineered P450-BM3 polypeptide is a polypeptide listed in any of Tables 2-9.

[0066] In some embodiments the engineered P450-BM3 which exhibits an improved property has at least 85%, at least 88%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% amino acid sequence identity with SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16.

[0067] In some embodiments, the engineered P450-BM3 polypeptide comprises a functional fragment of an engineered P450-BM3 polypeptide encompassed by the invention. Functional fragments have at least 95%, 96%, 97%, 98%, or 99% of the activity of the engineered P450-BM3 polypeptide from which it was derived (i.e., the parent engineered P450-BM3). A functional fragment comprises at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% and even 99% of the parent sequence of the engineered P450-BM3. In some embodiments the functional fragment is truncated by less than 5, less than 10, less than 15, less than 20, less than 25, less than 30, less than 35, less than 40, less than 45, and less than 50 amino acids.

Variants with Improved Activity:

[0068] In some embodiments, the engineered P450-BM3 polypeptides of the invention having P450-BM3 activity comprise: a) an amino acid sequence having at least 85% sequence identity to reference sequence SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16, or a fragment thereof; b) an amino acid residue difference as compared to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16, at one or more amino acid positions; and c) which exhibits improved activity, as compared to SEQ ID NO:2.

[0069] In some embodiments, the engineered P450-BM3 that exhibits improved activity has at least 85%, at least 88%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at

least 96%, at least 97%, at least 98%, at least 99% or greater amino acid sequence identity with SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16, and an amino acid residue difference as compared to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16, at one or more amino acid positions (such as at 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15, 20 or more amino acid positions compared to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16, or a sequence having at least 85%, at least 88%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99% or greater amino acid sequence identity with SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16.

[0070] In some embodiments, when all other assay conditions are essentially the same, the engineered P450-BM3 polypeptide has improved activity as compared to a reference P450-BM3 polypeptide. In some embodiments this activity can be measured under conditions that monitor enzymatic activity using any suitable assay system to assess the maximum activity of the enzyme (e.g., the k_{cat}). In other embodiments this activity can be measured under substrate concentrations resulting in one-half, one-fifth, one-tenth or less of maximal activity. Under either method of analysis, the engineered polypeptide has improved activity levels about 1.0 fold, 1.5-fold, 2-fold, 5-fold, 10-fold, 20-fold, 25-fold, 50-fold, 75-fold, 100-fold, or more of the enzymatic activity of the reference P450-BM3. In some embodiments, the engineered P450-BM3 polypeptide having improved activity as compared to a reference P450-BM3 when measured by any standard assay, including, but not limited to the assays described in the Examples.

[0071] In light of the guidance provided herein, it is further contemplated that any of the exemplary engineered polypeptides can be used as the starting amino acid sequence for synthesizing other engineered P450-BM3 polypeptides, for example by subsequent rounds of evolution by adding new combinations of various amino acid differences from other polypeptides and other residue positions described herein. Further improvements may be generated by including amino acid differences at residue positions that had been maintained as unchanged throughout earlier rounds of evolution.

Polynucleotides Encoding Engineered Polypeptides, Expression Vectors and Host Cells:

[0072] The present invention provides polynucleotides encoding the engineered P450-BM3 polypeptides described herein. In some embodiments, the polynucleotides are operatively linked to one or more heterologous regulatory sequences that control gene expression to create a recombinant polynucleotide capable of expressing the polypeptide. Expression constructs containing a heterologous polynucleotide encoding the engineered P450-BM3 polypeptides can be introduced into appropriate host cells to express the corresponding P450-BM3 polypeptide.

[0073] As will be apparent to the skilled artisan, availability of a protein sequence and the knowledge of the codons corresponding to the various amino acids provide a description of all the polynucleotides capable of encoding the subject polypeptides. The degeneracy of the genetic code, where the same amino acids are encoded by alternative or synonymous codons, allows an extremely

large number of nucleic acids to be made, all of which encode the engineered P450-BM3 polypeptide. Thus, having knowledge of a particular amino acid sequence, those skilled in the art could make any number of different nucleic acids by simply modifying the sequence of one or more codons in a way which does not change the amino acid sequence of the protein. In this regard, the present invention specifically contemplates each and every possible variation of polynucleotides that could be made encoding the polypeptides described herein by selecting combinations based on the possible codon choices, and all such variations are to be considered specifically disclosed for any polypeptide described herein, including the variants provided in Tables 2-9, as well as SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, and/or 16.

[0074] In various embodiments, the codons are preferably selected to fit the host cell in which the protein is being produced. For example, preferred codons used in bacteria are used for expression in bacteria. Consequently, codon optimized polynucleotides encoding the engineered P450-BM3 polypeptides contain preferred codons at about 40%, 50%, 60%, 70%, 80%, or greater than 90% of codon positions of the full length coding region.

[0075] In some embodiments, as described above, the polynucleotide encodes an engineered polypeptide having P450-BM3 activity with the properties disclosed herein, wherein the polypeptide comprises an amino acid sequence having at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to a reference sequence (e.g., SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16), or the amino acid sequence of any variant as disclosed in any of Tables 2-9, and one or more residue differences as compared to the reference polypeptide of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16, or the amino acid sequence of any variant as disclosed in any of Tables 2-9 (for example 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more amino acid residue positions). In some embodiments, the reference sequence is selected from SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16.

[0076] In some embodiments, the polynucleotides are capable of hybridizing under highly stringent conditions to a reference polynucleotide sequence selected from SEQ ID NO:1, 3, 5, 7, 9, 11, 13, and/or 15, or a complement thereof, or a polynucleotide sequence encoding any of the variant P450-BM3 polypeptides provided herein. In some embodiments, the polynucleotide capable of hybridizing under highly stringent conditions encodes a P450-BM3 polypeptide comprising an amino acid sequence that has one or more residue differences as compared to SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16.

[0077] In some embodiments, an isolated polynucleotide encoding any of the engineered P450-BM3 polypeptides provided herein is manipulated in a variety of ways to provide for expression of the polypeptide. In some embodiments, the polynucleotides encoding the polypeptides are provided as expression vectors where one or more control sequences is present to regulate the expression of the polynucleotides and/or polypeptides. Manipulation of the isolated polynucleotide prior to its insertion into a vector may be desirable or necessary depending on the expression vector. The techniques for

modifying polynucleotides and nucleic acid sequences utilizing recombinant DNA methods are well known in the art.

[0078] In some embodiments, the control sequences include among other sequences, promoters, leader sequences, polyadenylation sequences, propeptide sequences, signal peptide sequences, and transcription terminators. As known in the art, suitable promoters can be selected based on the host cells used. For bacterial host cells, suitable promoters for directing transcription of the nucleic acid constructs of the present application, include, but are not limited to the promoters obtained from the *E. coli* lac operon, *Streptomyces coelicolor* agarase gene (dagA), *Bacillus subtilis* levansucrase gene (sacB), *Bacillus licheniformis* alpha-amylase gene (amyL), *Bacillus stearothermophilus* maltogenic amylase gene (amyM), *Bacillus amyloliquefaciens* alpha-amylase gene (amyQ), *Bacillus licheniformis* penicillinase gene (penP), *Bacillus subtilis* xylA and xylB genes, and prokaryotic beta-lactamase gene (See e.g., Villa-Kamaroff et al., Proc. Natl Acad. Sci. USA 75: 3727-3731 [1978]), as well as the *tac* promoter (See e.g., DeBoer et al., Proc. Natl Acad. Sci. USA 80: 21-25 [1983]). Exemplary promoters for filamentous fungal host cells, include promoters obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Rhizomucor miehei* aspartic proteinase, *Aspergillus niger* neutral alpha-amylase, *Aspergillus niger* acid stable alpha-amylase, *Aspergillus niger* or *Aspergillus awamori* glucoamylase (glaA), *Rhizomucor miehei* lipase, *Aspergillus oryzae* alkaline protease, *Aspergillus oryzae* triose phosphate isomerase, *Aspergillus nidulans* acetamidase, and *Fusarium oxysporum* trypsin-like protease (See e.g., WO 96/00787), as well as the NA2-tpi promoter (a hybrid of the promoters from the genes for *Aspergillus niger* neutral alpha-amylase and *Aspergillus oryzae* triose phosphate isomerase), and mutant, truncated, and hybrid promoters thereof. Exemplary yeast cell promoters can be from the genes can be from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* galactokinase (GAL1), *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP), and *Saccharomyces cerevisiae* 3-phosphoglycerate kinase. Other useful promoters for yeast host cells are known in the art (See e.g., Romanos et al., Yeast 8:423-488 [1992]).

[0079] In some embodiments, the control sequence is a suitable transcription terminator sequence, a sequence recognized by a host cell to terminate transcription. The terminator sequence is operably linked to the 3' terminus of the nucleic acid sequence encoding the polypeptide. Any terminator which is functional in the host cell of choice finds use in the present invention. For example, exemplary transcription terminators for filamentous fungal host cells can be obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Aspergillus niger* alpha-glucosidase, and *Fusarium oxysporum* trypsin-like protease. Exemplary terminators for yeast host cells can be obtained from the genes for *Saccharomyces cerevisiae* enolase, *Saccharomyces cerevisiae* cytochrome C (CYC1), and *Saccharomyces cerevisiae* glyceraldehyde-3-phosphate dehydrogenase. Other useful terminators for yeast host cells are known in the art (See e.g., Romanos et al., *supra*).

[0080] In some embodiments, the control sequence is a suitable leader sequence, a non-translated region of an mRNA that is important for translation by the host cell. The leader sequence is operably linked to the 5' terminus of the nucleic acid sequence encoding the polypeptide. Any leader sequence that is functional in the host cell of choice may be used. Exemplary leaders for filamentous fungal host cells are obtained from the genes for *Aspergillus oryzae* TAKA amylase and *Aspergillus nidulans* triose phosphate isomerase. Suitable leaders for yeast host cells include, but are not limited to those obtained from the genes for *Saccharomyces cerevisiae* enolase (ENO-1), *Saccharomyces cerevisiae* 3-phosphoglycerate kinase, *Saccharomyces cerevisiae* alpha-factor, and *Saccharomyces cerevisiae* alcohol dehydrogenase/glyceraldehyde-3-phosphate dehydrogenase (ADH2/GAP).

[0081] The control sequence may also be a polyadenylation sequence, a sequence operably linked to the 3' terminus of the nucleic acid sequence and which, when transcribed, is recognized by the host cell as a signal to add polyadenosine residues to transcribed mRNA. Any polyadenylation sequence which is functional in the host cell of choice may be used in the present invention. Exemplary polyadenylation sequences for filamentous fungal host cells include, but are not limited to those from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* glucoamylase, *Aspergillus nidulans* anthranilate synthase, *Fusarium oxysporum* trypsin-like protease, and *Aspergillus niger* alpha-glucosidase. Useful polyadenylation sequences for yeast host cells are also known in the art (See e.g., Guo and Sherman, Mol. Cell. Bio., 15:5983-5990 [1995]).

[0082] In some embodiments, the control sequence is a signal peptide coding region that codes for an amino acid sequence linked to the amino terminus of a polypeptide and directs the encoded polypeptide into the cell's secretory pathway. The 5' end of the coding sequence of the nucleic acid sequence may inherently contain a signal peptide coding region naturally linked in translation reading frame with the segment of the coding region that encodes the secreted polypeptide. Alternatively, the 5' end of the coding sequence may contain a signal peptide coding region that is foreign to the coding sequence. Any signal peptide coding region that directs the expressed polypeptide into the secretory pathway of a host cell of choice finds use for expression of the engineered P450-BM3 polypeptides provided herein. Effective signal peptide coding regions for bacterial host cells include, but are not limited to the signal peptide coding regions obtained from the genes for *Bacillus* NCIB 11837 maltogenic amylase, *Bacillus stearothermophilus* alpha-amylase, *Bacillus licheniformis* subtilisin, *Bacillus licheniformis* beta-lactamase, *Bacillus stearothermophilus* neutral proteases (nprT, nprS, nprM), and *Bacillus subtilis* prsA. Further signal peptides are known in the art (See e.g., Simonen and Palva, Microbiol. Rev., 57:109-137 [1993]). Effective signal peptide coding regions for filamentous fungal host cells include, but are not limited to the signal peptide coding regions obtained from the genes for *Aspergillus oryzae* TAKA amylase, *Aspergillus niger* neutral amylase, *Aspergillus niger* glucoamylase, *Rhizomucor miehei* aspartic proteinase, *Humicola insolens* cellulase, and *Humicola lanuginosa* lipase. Useful signal peptides for yeast host cells include, but are not limited to those from the genes for *Saccharomyces cerevisiae* alpha-factor and *Saccharomyces cerevisiae* invertase.

[0083] In some embodiments, the control sequence is a propeptide coding region that codes for an amino acid sequence positioned at the amino terminus of a polypeptide. The resultant polypeptide is referred to as a “proenzyme,” “propolypeptide,” or “zymogen,” in some cases). A propolypeptide can be converted to a mature active polypeptide by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide. The propeptide coding region includes, but is not limited to the genes for *Bacillus subtilis* alkaline protease (aprE), *Bacillus subtilis* neutral protease (nprT), *Saccharomyces cerevisiae* alpha-factor, *Rhizomucor miehei* aspartic proteinase, and *Myceliophthora thermophila* lactase (See e.g., WO 95/33836). Where both signal peptide and propeptide regions are present at the amino terminus of a polypeptide, the propeptide region is positioned next to the amino terminus of a polypeptide and the signal peptide region is positioned next to the amino terminus of the propeptide region.

[0084] In some embodiments, regulatory sequences are also utilized. These sequences facilitate the regulation of the expression of the polypeptide relative to the growth of the host cell. Examples of regulatory systems are those which cause the expression of the gene to be turned on or off in response to a chemical or physical stimulus, including the presence of a regulatory compound. In prokaryotic host cells, suitable regulatory sequences include, but are not limited to the *lac*, *tac*, and *trp* operator systems. In yeast host cells, suitable regulatory systems include, but are not limited to the ADH2 system or GAL1 system. In filamentous fungi, suitable regulatory sequences include, but are not limited to the TAKA alpha-amylase promoter, *Aspergillus niger* glucoamylase promoter, and *Aspergillus oryzae* glucoamylase promoter.

[0085] In another aspect, the present invention also provides a recombinant expression vector comprising a polynucleotide encoding an engineered P450-BM3 polypeptide, and one or more expression regulating regions such as a promoter and a terminator, a replication origin, etc., depending on the type of hosts into which they are to be introduced. In some embodiments, the various nucleic acid and control sequences described above are joined together to produce a recombinant expression vector which includes one or more convenient restriction sites to allow for insertion or substitution of the nucleic acid sequence encoding the variant P450-BM3 polypeptide at such sites. Alternatively, the polynucleotide sequence(s) of the present invention are expressed by inserting the polynucleotide sequence or a nucleic acid construct comprising the polynucleotide sequence into an appropriate vector for expression. In creating the expression vector, the coding sequence is located in the vector so that the coding sequence is operably linked with the appropriate control sequences for expression.

[0086] The recombinant expression vector may be any vector (e.g., a plasmid or virus), that can be conveniently subjected to recombinant DNA procedures and can result in the expression of the variant P450-BM3 polynucleotide sequence. The choice of the vector will typically depend on the compatibility of the vector with the host cell into which the vector is to be introduced. The vectors may be linear or closed circular plasmids.

[0087] In some embodiments, the expression vector is an autonomously replicating vector (i.e., a vector that exists as an extra-chromosomal entity, the replication of which is independent of chromosomal replication, such as a plasmid, an extra-chromosomal element, a minichromosome, or an artificial chromosome). The vector may contain any means for assuring self-replication. In some alternative embodiments, the vector may be one which, when introduced into the host cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. Furthermore, a single vector or plasmid or two or more vectors or plasmids which together contain the total DNA to be introduced into the genome of the host cell, or a transposon may be used.

[0088] In some embodiments, the expression vector preferably contains one or more selectable markers, which permit easy selection of transformed cells. A "selectable marker" is a gene the product of which provides for biocide or viral resistance, resistance to heavy metals, prototrophy to auxotrophs, and the like. Examples of bacterial selectable markers include, but are not limited to the *dal* genes from *Bacillus subtilis* or *Bacillus licheniformis*, or markers, which confer antibiotic resistance such as ampicillin, kanamycin, chloramphenicol or tetracycline resistance. Suitable markers for yeast host cells include, but are not limited to ADE2, HIS3, LEU2, LYS2, MET3, TRP1, and URA3. Selectable markers for use in a filamentous fungal host cell include, but are not limited to, *amdS* (acetamidase), *argB* (ornithine carbamoyltransferases), *bar* (phosphinothricin acetyltransferase), *hph* (hygromycin phosphotransferase), *niaD* (nitrate reductase), *pyrG* (orotidine-5'-phosphate decarboxylase), *sC* (sulfate adenylyltransferase), and *trpC* (anthranilate synthase), as well as equivalents thereof. In another aspect, the present invention provides a host cell comprising a polynucleotide encoding at least one engineered P450-BM3 polypeptide of the present application, the polynucleotide being operatively linked to one or more control sequences for expression of the engineered P450-BM3 enzyme(s) in the host cell. Host cells for use in expressing the polypeptides encoded by the expression vectors of the present invention are well known in the art and include but are not limited to, bacterial cells, such as *E. coli*, *Vibrio fluvialis*, *Streptomyces* and *Salmonella typhimurium* cells; fungal cells, such as yeast cells (e.g., *Saccharomyces cerevisiae* and *Pichia pastoris* [ATCC Accession No. 201178]); insect cells such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, BHK, 293, and Bowes melanoma cells; and plant cells. Exemplary host cells are *Escherichia coli* strains (such as W3110 (Δ fhuA) and BL21).

[0089] Accordingly, in another aspect, the present invention provides methods for producing the engineered P450-BM3 polypeptides, where the methods comprise culturing a host cell capable of expressing a polynucleotide encoding the engineered P450-BM3 polypeptide under conditions suitable for expression of the polypeptide. In some embodiments, the methods further comprise the steps of isolating and/or purifying the P450-BM3 polypeptides, as described herein.

[0090] Appropriate culture media and growth conditions for the above-described host cells are well known in the art. Polynucleotides for expression of the P450-BM3 polypeptides may be introduced into cells by various methods known in the art. Techniques include, among others, electroporation,

biolistic particle bombardment, liposome mediated transfection, calcium chloride transfection, and protoplast fusion.

[0091] The engineered P450-BM3 with the properties disclosed herein can be obtained by subjecting the polynucleotide encoding the naturally occurring or engineered P450-BM3 polypeptide to mutagenesis and/or directed evolution methods known in the art, and as described herein. An exemplary directed evolution technique is mutagenesis and/or DNA shuffling (See e.g., Stemmer, Proc. Natl. Acad. Sci. USA 91:10747-10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767 and U.S. Pat. 6,537,746). Other directed evolution procedures that can be used include, among others, staggered extension process (StEP), in vitro recombination (See e.g., Zhao et al., Nat. Biotechnol., 16:258-261 [1998]), mutagenic PCR (See e.g., Caldwell et al., PCR Methods Appl., 3:S136-S140 [1994]), and cassette mutagenesis (See e.g., Black et al., Proc. Natl. Acad. Sci. USA 93:3525-3529 [1996]).

[0092] For example, mutagenesis and directed evolution methods can be readily applied to polynucleotides to generate variant libraries that can be expressed, screened, and assayed.

Mutagenesis and directed evolution methods are well known in the art (See e.g., US Patent Nos. 5,605,793, 5,830,721, 6,132,970, 6,420,175, 6,277,638, 6,365,408, 6,602,986, 7,288,375, 6,287,861, 6,297,053, 6,576,467, 6,444,468, 5,811,238, 6,117,679, 6,165,793, 6,180,406, 6,291,242, 6,995,017, 6,395,547, 6,506,602, 6,519,065, 6,506,603, 6,413,774, 6,573,098, 6,323,030, 6,344,356, 6,372,497, 7,868,138, 5,834,252, 5,928,905, 6,489,146, 6,096,548, 6,387,702, 6,391,552, 6,358,742, 6,482,647, 6,335,160, 6,653,072, 6,355,484, 6,303,344, 6,319,713, 6,613,514, 6,455,253, 6,579,678, 6,586,182, 6,406,855, 6,946,296, 7,534,564, 7,776,598, 5,837,458, 6,391,640, 6,309,883, 7,105,297, 7,795,030, 6,326,204, 6,251,674, 6,716,631, 6,528,311, 6,287,862, 6,335,198, 6,352,859, 6,379,964, 7,148,054, 7,629,170, 7,620,500, 6,365,377, 6,358,740, 6,406,910, 6,413,745, 6,436,675, 6,961,664, 7,430,477, 7,873,499, 7,702,464, 7,783,428, 7,747,391, 7,747,393, 7,751,986, 6,376,246, 6,426,224, 6,423,542, 6,479,652, 6,319,714, 6,521,453, 6,368,861, 7,421,347, 7,058,515, 7,024,312, 7,620,502, 7,853,410, 7,957,912, 7,904,249, and all related non-US counterparts; Ling et al., Anal. Biochem., 254:157-78 [1997]; Dale et al., Meth. Mol. Biol., 57:369-74 [1996]; Smith, Ann. Rev. Genet., 19:423-462 [1985]; Botstein et al., Science, 229:1193-1201 [1985]; Carter, Biochem. J., 237:1-7 [1986]; Kramer et al., Cell, 38:879-887 [1984]; Wells et al., Gene, 34:315-323 [1985]; Minshull et al., Curr. Op. Chem. Biol., 3:284-290 [1999]; Christians et al., Nat. Biotechnol., 17:259-264 [1999]; Cramer et al., Nature, 391:288-291 [1998]; Cramer et al., Nat. Biotechnol., 15:436-438 [1997]; Zhang et al., Proc. Nat. Acad. Sci. U.S.A., 94:4504-4509 [1997]; Cramer et al., Nat. Biotechnol., 14:315-319 [1996]; Stemmer, Nature, 370:389-391 [1994]; Stemmer, Proc. Nat. Acad. Sci. USA, 91:10747-10751 [1994]; WO 95/22625; WO 97/0078; WO 97/35966; WO 98/27230; WO 00/42651; WO 01/75767; WO 2009/152336, and U.S. Pat. No. 6,537,746, all of which are incorporated herein by reference).

[0093] In some embodiments, the enzyme clones obtained following mutagenesis treatment are screened by subjecting the enzymes to a defined temperature (or other assay conditions, such as

testing the enzyme's activity over a broad range of substrates) and measuring the amount of enzyme activity remaining after heat treatments or other assay conditions. Clones containing a polynucleotide encoding a P450-BM3 polypeptide are then sequenced to identify the nucleotide sequence changes (if any), and used to express the enzyme in a host cell. Measuring enzyme activity from the expression libraries can be performed using any suitable method known in the art (e.g., standard biochemistry techniques, such as HPLC analysis).

[0094] For engineered polypeptides of known sequence, the polynucleotides encoding the enzyme can be prepared by standard solid-phase methods, according to known synthetic methods. In some embodiments, fragments of up to about 100 bases can be individually synthesized, then joined (e.g., by enzymatic or chemical ligation methods, or polymerase mediated methods) to form any desired continuous sequence. For example, polynucleotides and oligonucleotides disclosed herein can be prepared by chemical synthesis using the classical phosphoramidite method (See e.g., Beaucage et al., *Tetra. Lett.*, 22:1859-69 [1981]; and Matthes et al., *EMBO J.*, 3:801-05 [1984]), as it is typically practiced in automated synthetic methods. According to the phosphoramidite method, oligonucleotides are synthesized (e.g., in an automatic DNA synthesizer), purified, annealed, ligated and cloned in appropriate vectors.

[0095] Accordingly, in some embodiments, a method for preparing the engineered P450-BM3 polypeptide can comprise: (a) synthesizing a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from the amino acid sequence of any variant provided in any of Tables 2-9, as well as SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and/or 16, and (b) expressing the P450-BM3 polypeptide encoded by the polynucleotide. In some embodiments of the method, the amino acid sequence encoded by the polynucleotide can optionally have one or several (e.g., up to 3, 4, 5, or up to 10) amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally 1-2, 1-3, 1-4, 1-5, 1-6, 1-7, 1-8, 1-9, 1-10, 1-15, 1-20, 1-21, 1-22, 1-23, 1-24, 1-25, 1-30, 1-35, 1-40, 1-45, or 1-50 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 30, 30, 35, 40, 45, or 50 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the amino acid sequence has optionally 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 18, 20, 21, 22, 23, 24, or 25 amino acid residue deletions, insertions and/or substitutions. In some embodiments, the substitutions can be conservative or non-conservative substitutions.

[0096] The expressed engineered P450-BM3 polypeptide can be measured for any desired improved property (e.g., activity, selectivity, stability, acid tolerance, protease sensitivity, etc.), using any suitable assay known in the art, including but not limited to the assays and conditions described herein.

[0097] In some embodiments, any of the engineered P450-BM3 polypeptides expressed in a host cell are recovered from the cells and/or the culture medium using any one or more of the well-known

techniques for protein purification, including, among others, lysozyme treatment, sonication, filtration, salting-out, ultra-centrifugation, and chromatography.

[0098] Chromatographic techniques for isolation of the P450-BM3 polypeptides include, among others, reverse phase chromatography high performance liquid chromatography, ion exchange chromatography, hydrophobic interaction chromatography, gel electrophoresis, and affinity chromatography. Conditions for purifying a particular enzyme depends, in part, on factors such as net charge, hydrophobicity, hydrophilicity, molecular weight, molecular shape, etc., and will be apparent to those having skill in the art. In some embodiments, affinity techniques may be used to isolate the improved variant P450-BM3 enzymes. In some embodiments utilizing affinity chromatography purification, any antibody which specifically binds the variant P450-BM3 polypeptide finds use. For the production of antibodies, various host animals, including but not limited to rabbits, mice, rats, etc., are immunized by injection with a P450-BM3 polypeptide (e.g., a P450-BM3 variant), or a fragment thereof. In some embodiments, the P450-BM3 polypeptide or fragment is attached to a suitable carrier, such as BSA, by means of a side chain functional group or linkers attached to a side chain functional group.

[0099] In some embodiments, the engineered P450-BM3 polypeptide is produced in a host cell by a method comprising culturing a host cell (e.g., an *E. coli* strain) comprising a polynucleotide sequence encoding an engineered P450-BM3 polypeptide as described herein under conditions conducive to the production of the engineered P450-BM3 polypeptide and recovering the engineered P450-BM3 polypeptide from the cells and/or culture medium.

[0100] In some embodiments, the engineered P450-BM3 polypeptides are recovered from the recombinant host cells or cell culture and they are further purified by any suitable method(s) known in the art. In some additional embodiments, the purified P450-BM3 polypeptides are combined with other ingredients and compounds to provide compositions and formulations comprising the engineered P450-BM3 polypeptide as appropriate for different applications and uses (e.g., pharmaceutical compositions).

[0101] The foregoing and other aspects of the invention may be better understood in connection with the following non-limiting examples. The examples are provided for illustrative purposes only and are not intended to limit the scope of the present invention in any way.

EXPERIMENTAL

[0102] The following Examples, including experiments and results achieved, are provided for illustrative purposes only and are not to be construed as limiting the present invention.

[0103] In the experimental disclosure below, the following abbreviations apply: ppm (parts per million); M (molar); mM (millimolar), uM and μ M (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and μ g (micrograms); L and l (liter); ml and mL (milliliter); cm

(centimeters); mm (millimeters); μm and μm (micrometers); sec. (seconds); min(s) (minute(s)); h(s) and hr(s) (hour(s)); U (units); MW (molecular weight); rpm (rotations per minute); °C (degrees Centigrade); CDS (coding sequence); DNA (deoxyribonucleic acid); RNA (ribonucleic acid); NA (nucleic acid; polynucleotide); AA (amino acid; polypeptide); *E. coli* W3110 (commonly used laboratory *E. coli* strain, available from the Coli Genetic Stock Center [CGSC], New Haven, CT); HPLC (high pressure liquid chromatography); SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis); PES (polyethersulfone); CFSE (carboxyfluorescein succinimidyl ester); IPTG (isopropyl beta-D-1-thiogalactopyranoside); PMBS (polymyxin B sulfate); NADPH (nicotinamide adenine dinucleotide phosphate); GDH (glucose dehydrogenase); TON (turnover number); FIOPC (fold improvement over positive control); TON (turnover number); ESI (electrospray ionization); LB (Luria broth); TB (terrific broth); MeOH (methanol); Athens Research (Athens Research Technology, Athens, GA); ProSpec (ProSpec Tany Technogene, East Brunswick, NJ); Sigma-Aldrich (Sigma-Aldrich, St. Louis, MO); Ram Scientific (Ram Scientific, Inc., Yonkers, NY); Pall Corp. (Pall, Corp., Pt. Washington, NY); Millipore (Millipore, Corp., Billerica MA); Difco (Difco Laboratories, BD Diagnostic Systems, Detroit, MI); Molecular Devices (Molecular Devices, LLC, Sunnyvale, CA); Kuhner (Adolf Kuhner, AG, Basel, Switzerland); Cambridge Isotope Laboratories, (Cambridge Isotope Laboratories, Inc., Tewksbury, MA); Applied Biosystems (Applied Biosystems, part of Life Technologies, Corp., Grand Island, NY); Agilent (Agilent Technologies, Inc., Santa Clara, CA); Thermo Scientific (part of Thermo Fisher Scientific, Waltham, MA); Fisher (Fisher Scientific, Waltham, MA); Corning (Corning, Inc., Palo Alto, CA); Waters (Waters Corp., Milford, MA); GE Healthcare (GE Healthcare Bio-Sciences, Piscataway, NJ); Pierce (Pierce Biotechnology (now part of Thermo Fisher Scientific), Rockford, IL); Phenomenex (Phenomenex, Inc., Torrance, CA); Optimal (Optimal Biotech Group, Belmont, CA); and Bio-Rad (Bio-Rad Laboratories, Hercules, CA).

EXAMPLE 1

P450-BM3 Evolution and Construction of Expression Vectors

[0104] Libraries of P450-BM3 variants were produced using standard methods known in the art, based on eight parental sequences of SEQ ID NO:2, 4, 6, 8, 10, 12, 14, and 16. These parental strains were used to generate combinatorial libraries by recombining beneficial diversity. These parental backbone strains and their Sequence IDs are listed in Table 1 below, with the polynucleotide sequence listed first, followed by the polypeptide sequence.

Table 1. P450-BM3 Parental Backbone Sequences

Backbone Names	SEQ ID NOS:
MCYP-1.2-A05	1/2

Table 1. P450-BM3 Parental Backbone Sequences

Backbone Names	SEQ ID NOS:
MCYP-1.2-A07	3/4
MCYP-1.2-A12	5/6
MCYP-P1.2-B02	7/8
MCYP-P1.2-B12	9/10
MCYP-P1.2-D06	11/12
MCYP-P1.2-F02	13/14
MCYP-1.2-F12	15/16

[0105] These variants were cloned into an IPTG-inducible vector, transformed into *E. coli* strain BL21, and plated on LB agar plates supplemented with chloramphenicol (30 µg/mL). The plates were grown at 37 °C for 16 hrs before single clones were picked and added to 96-well Axygen® plates (Corning), containing LB medium (250 µL/well) supplemented with chloramphenicol (30 µg/mL). After the plates were shaken at 250 rpm, 30 °C, and 85% humidity for 20-24 h to grow the cultures to saturation, an aliquot (50 µL) was used to inoculate 2 mL 96-well Costar® deep plates (Corning) containing TB medium (900 µL) supplemented with chloramphenicol (30 µg/mL), trace element solution (740 µg/L ammonium molybdate tetrahydrate, 5.8 mgs/L zinc sulfate heptahydrate, 620 µg/L boric acid anhydrous, 1 mg/L copper sulfate pentahydrate, and 4 mgs/L manganese chloride tetrahydrate), and 0.05 g/L ammonium iron (III) citrate. After being shaken at 250 rpm, 30 °C, and 85% humidity to an OD₆₀₀ of 0.8 – 1.2, P450 expression was induced by addition of IPTG (500 µM) and the heme precursor 5-aminolevulinic acid (5-ALA) to a final concentration of 500 µM. The cultures were shaken at 250 rpm, 26 °C, 85% humidity for 24 hrs before the cells were centrifuged and stored at -80 °C.

[0106] Cell lysis was accomplished by resuspending cell pellets in 96-well Costar® plates (Corning) with lysis buffer (300 µL/well) containing potassium phosphate, pH 8.0 (100 mM), MgSO₄ (10 mM), DTT (1 mM), lysozyme (1 mg/mL), PMBS (0.5 mg/mL), and DNaseI (3 µg/mL). The lysis reactions were shaken using a table top shaker (setting 8-10) at room temperature for 1.25 hrs. The lysis reaction was centrifuged to pellet cellular debris and the supernatant was used in the activity assays described in Example 2.

[0107] For the production of lyophilized protein powders, LB agar plates supplemented with chloramphenicol (30 µg/mL) were streaked with *E. coli* containing a desired *B. megaterium* P450-BM3 variant in an IPTG-inducible vector. The plates were grown at 37 °C for 16 hrs before single

clones were selected to inoculate a 15 mL Falcon™ tube (Fisher) containing TB media (3 mL) supplemented with chloramphenicol (30 µg/mL). The tube was shaken at 200 rpm, 30 °C, and 85% humidity for 20-24 h to grow the cultures to saturation. Then, 2.5 mL of the overnight culture was used to inoculate sterile 1L flasks containing TB medium (250 mL) supplemented with chloramphenicol (30 µg/mL), trace element solution (as described above), and 0.05 g/L ammonium iron (III) citrate. After being shaken at 250 rpm, 30 °C, and 85% humidity to an OD₆₀₀ of 0.8 – 1.2, P450 expression was induced by addition of IPTG (500 µM) and the heme precursor 5-aminolevulinic acid (5-ALA) to a final concentration of 500 µM. The cultures were grown for additional 20-24 hours and centrifuged in pre-weighed 250 mL centrifuge bottles for 20 minutes at 4000 rpm, 4 °C. The supernatants were discarded, and the centrifuge bottles containing cell pellets were weighed. The pellets were resuspended in 50 mM potassium phosphate buffer with 2 mM DTT, pH 8.0 (5 mL of buffer per gram of cell pellet). The cells were lysed using a microfluidizer homogenizer, and the suspensions of cells and lysate were collected in sterile 50 mL centrifuge tubes. The samples were centrifuged for 30 minutes at 10,000 rpm, 4 °C. The clarified lysate was collected into a plastic petri plate and frozen at -80 °C prior to lyophilization. The enzyme-containing lysates were lyophilized using standard methods known in the art.

EXAMPLE 2

ASSAY SYSTEMS & RESULTS

[0108] In this Example, the test systems used to assess the activities and generalist properties (i.e., activity on a broad substrate range) are described.

I. Activity-Based High Throughput Screening (HTP) for Enzymatic Activity:

[0109] Diclofenac (See, Figure 1) was used as a substrate for high throughput (HTP) screening assays to detect variants with improved activity. Enzymatic activity screens were initiated by adding 60 µL lysate and 120 µL of the reaction mixture to each well of a 96-well (2 mL) plate. The reaction mixture contained the recycling system (120 mM potassium phosphate, 1.2 mM NADP⁺, 30 mM glucose, and 0.6 mg/mL glucose dehydrogenase), co-solvent (7.5% DMSO), and substrate (3 mM diclofenac). The reactions were shaken at 250 rpm, 30 °C, 85% humidity for 4-24 hrs. The reactions were quenched by the addition of acetonitrile (400 µL to 1 mL) to each well. The quenched reactions were centrifuged to remove precipitated proteins prior to analysis with HPLC and LCMS, as described below.

II. Validation of Generalist Properties:

[0110] Enzyme stocks (~12 µM) were prepared by dissolving ~20 mg of each enzyme in 100 mM potassium phosphate buffer, pH 8.0 (1 mL). The concentration of each enzyme stock solution was

determined by the UV-visible absorption spectroscopy (after centrifugation to remove particulates) and diluted to standardize at 12 μ M heme protein. Substrate solutions were prepared by dissolving each substrate in DMSO to reach a final concentration of 20 mM or 40 mM. Reaction mix (235 μ L) followed by enzyme solution (50 μ L) was added to the plates. The substrate stock solution was added to the enzyme solutions (15 μ L at the two different concentrations). Each reaction contained 100 mM potassium phosphate, 1.0 mM NADP⁺, 25 mM glucose, 0.5 mg/mL glucose dehydrogenase, 5% DMSO and substrate at 1 or 2 mM. Loratadine, imatinib, and gefitinib (See, Figure 1) were selected as substrates in addition to diclofenac to validate improved BM3 variants. The reactions were shaken at 450 rpm at 30 °C over 24 hours. All reactions were diluted with acetonitrile to a final concentration of 0.5 mM substrate. The plates were then centrifuged at 3,000g for 1 hour at 20 °C. The supernatant was diluted 1:1 with acetonitrile, filtered using a 0.4 micron filter, and analyzed by UPLC-MS.

III. HPLC, LCMS and UPLC-MS Analysis:

[0111] For HPLC and LCMS analysis, 150 μ L of each quenched reaction sample was transferred to 96-well round bottom plates for analysis by HPLC on an Agilent Technologies 1200 series equipped with an autosampler. 10 μ L of quenched sample was injected onto an Onyx Monolithic C18 column (2 x 50 mm). The column was eluted at a constant flow rate of 0.5 mL/min; conditions with solvent A (0.1% formic acid v/v, in H₂O) and solvent B (0.1% formic acid v/v, in acetonitrile) used to elute the products of the reaction were: 0–1 min, A/B 90:10; 1-2 min, A/B 80:20; 2-4 min, A/B 70:30; 4-4.5 min, A/B 60:40; 4.5-4.9 min, A/B 10:90, and 4.9–5.3 min, A/B 90:10. Column eluent was monitored by UV at 270 nm. Alternatively, analysis by LC-UV-MS was performed for some substrates on a Thermo LXQ ion trap system using a Surveyor Plus LC-PDA system for sample separation.

Quenched sample (0.01 ml) was injected onto a Waters Xbridge C18 column (3 x 50 mm, 5 μ). The column was eluted at a constant flow rate of 0.5 mL min; the conditions with solvent A (0.1% formic acid v/v, in H₂O) and solvent B (0.1% formic acid v/v, in acetonitrile) used to elute the products of the reaction were: 0–1.5 min, A/B 90:10; 1.5-5.5 min, A/B 20:80; 5.5-6.0 min, A/B 1:99; 6.0-6.25 min, A/B 90:10; 6.25-7.5 min, A/B 90:10. Column eluent was monitored by PDA (200-600 nm) prior to MS analysis in positive ESI mode (capillary temperature 350°C, 5kV spray voltage). The column eluent was diverted to waste for the first 1.5 minutes of the run. For the remainder of the LC run, both MS (*m/z* 125-1000 scan range) and MS/MS were collected. MS/MS spectra were acquired in a data-dependent manner for the *n*th most intense ions employing dynamic exclusion for dominate ions after the 5th occurrence with an exclusion duration of 30 seconds. Data were analyzed using Xcalibur software for substrate and product base peaks and MS/MS transitions.

[0112] For UPLC-MS analysis, the quenched and filtered reactions were transferred to 96-well HTP plates for analysis by UPLC on a Waters Acquity H-class system equipped with an autosampler. 1 μ L of quenched sample was injected onto an Acquity UPLC HSS T3 column, 100Å, (1.8 μ m; 2.1 x 100 mm). The column was eluted at a constant flow rate of 0.6 mL/min; conditions with solvent A (0.05%

trifluoroacetic acid v/v, in H₂O) and solvent B (0.05% trifluoroacetic acid v/v, in acetonitrile) used to elute the products of the reaction were: 0–2 min, A/B 95:5; 2–2.9 min, A/B 5:95; 2.9–3 min, A/B 95:5; 3–3.5 min, A/B 95:5. Column eluent was monitored by PDA (200–600 nm) prior to MS analysis in positive ESI mode (desolvation temperature 350°C, 3.25kV spray voltage, cone voltage 25V). The column eluent was diverted to waste for the first 0.7 minutes of the run. For the remainder of the LC run, both MS (*m/z* 95–600 scan range) and MS/MS were collected. MS/MS spectra were acquired in a data-dependent manner for the *n*th most intense ions employing dynamic exclusion for dominate ions after the 5th occurrence with an exclusion duration of 30 seconds. Data were analyzed using MassLynx and Virscidian software for substrate and product base peaks and MS/MS transitions.

IV. Results:

[0113] The ability to generate enzymes that have improved activity for a vast range of substrates requires an evolution approach that optimizes both substrate binding and rate-limiting electron transfer. Eighteen mutations identified previously that generally optimize both parameters (See, U.S. Pat. Appln. Publ. No. 2016/0010065, the contents of which are incorporated herein by reference in its entirety and for all purposes) were recombined using the eight alternate backbones listed in Table 1-1, from the 96-well commercially available MCYP panel (MCYP-0343; Codexis). The purpose of this approach was to screen a combinatorial library from each lineage built on an alternate backbone using diclofenac as a screening substrate to identify and select improved variants. Lyophilized powders of the most improved variants were screened against a suite of compounds to determine the magnitude of improvements on multiple substrates (e.g., generalist properties). This approach has been referred to the “transferability of generalist diversity.” The variants, substrates screened, and mutations are summarized in Tables 2 through 9. Figure 2 is a graphical summary for the MCYP-1.2-A12 lineage (one of the eight lineages summarized in Tables 2-9). In Figure 2, the percent conversion for each substrate screened is plotted as a function of each enzyme screened. In this Figure, the performance (% conversion at 1mM substrate loading) of two evolved P450s (variants 16 and 17) is compared to the parental backbone, MCYP-1.2-A12. As shown in Figure 2 and summarized in Table 4, the parental backbone (MCYP-1.2-A12) exhibited low activity for each substrate screened. The evolved variants (variants 16 and 17) have improved and significant activity for three out of four substrates screened. Similar trends were observed for the remaining lineages. The performance (% conversion at 1mM substrate loading) of variants 14 and 15 was compared to their parental backbone, MCYP-1.2-A07. The performance (% conversion at 1mM substrate loading) of the parental backbone (MCYP-1.2-A07) is summarized in Table 3 and exhibits low activity for imatinib and gefitinib, and moderate activity for diclofenac and loratadine. The evolved variants exhibit moderate activity for all four substrates. For each lineage, an evolved variant exhibits activity for at least one substrate that the corresponding parental backbone showed little to no activity and/or exhibits improved performance

(% conversion at 1mM substrate loading) for at least one substrate for which the parental backbone exhibits activity.

[0114] The same trend is observed for the MCYP-1.2-A05 lineage (See, Figure 3), although MCYP-1.2-A05 is a chimera. The P450 domain is 86% identical to *Bacillus subtilis* P450. These results indicate that the combination of analogous mutations/positions should impart improvements in P450 enzymes from other organisms.

[0115] In Tables 2-9, ^aTON is calculated as ([substrate]*% conversion/[P450]) and ^bFIOPC is calculated as either the TON (variant)/TON (parent) or % conversion (variant)/% conversion (parent). The following Table provides the key to the remaining entries.

% Conversion	Notation	TON	Notation	FIOPC	Notation
0 - 5.00	+	0-1000	+	0.0 - 3.0	*
5.01 - 10.00	++	1001-5000	#	3.01 - 5.0	**
10.01 -15.00	+++	5001-10000	##	5.01 - 10.0	***
15.01 - 20.00	++++	10000-50000	###	10.01 - 15.00	****

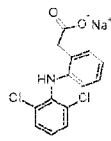
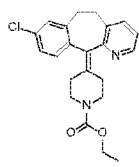
Table 2. Results for Variants 6, 7, and 8 (with substitutions shown with reference to SEQ ID NO:2)				
SF Validation Substrate(s)	Parent (SEQ ID NO:2)	Variant 6 (V51S;D118R;L219Y; K236H;P351T;S353T; N579T;T582A;E624D; A632S;D727S;Q931K)	Variant 7 (V51S;D118R;T179V; L219Y;P351T;N579T; T582A;E624D;D727S)	Variant 8 (D118R;L219Y;K236H; S353T;N579T;T582A; E624D;D727S)
 Diclofenac	[Diclofenac] = 1 mM TON ^a = ‡ % Conv. = +	[Diclofenac] = 1 mM TON = # % Conv. = ++ FIOPC ^b (A05) = ****	[Diclofenac] = 1 mM TON = # % Conv. = + FIOPC (A05) = ****	[Diclofenac] = 1 mM TON = # % Conv. = + FIOPC (A05) = ****
 Loratadine	[Loratadine] = 1 mM TON = ‡	[Loratadine] = 1 mM TON = # % Conv. = ++	[Loratadine] = 1 mM TON = # % Conv. = ++	Loratadine] = 1 mM TON = # % Conv. = ++

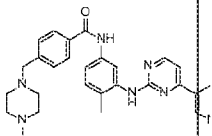
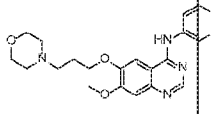
Table 2. Results for Variants 6, 7, and 8 (with substitutions shown with reference to SEQ ID NO:2)				
SF Validation Substrate(s)	Parent (SEQ ID NO:2)	Variant 6 (V51S;D118R;L219Y; K236H;P351T;S353T; N579T;T582A;E624D; A632S;D727S;Q931K)	Variant 7 (V51S;D118R;T179V; L219Y;P351T;N579T; T582A;E624D;D727S)	Variant 8 (D118R;L219Y;K236H ; S353T;N579T;T582A; E624D;D727S)
	% Conv. = +	FIOPC (A05) = ****	FIOPC (A05) = ****	FIOPC (A05) = ****
 Imatinib	[Imatinib] = 1 mM TON = # % Conv. = ++	[Imatinib] = 1 mM TON = ### % Conv. = +++++ FIOPC (A05) = *	[Imatinib] = 1 mM TON = ### % Conv. = +++++ FIOPC (A05) = **	[Imatinib] = 1 mM TON = ## % Conv. = +++++ FIOPC (A05) = *
 Gefitinib	[Gefitinib] = 1 mM TON = + % Conv. = +	[Gefitinib] = 1 mM TON = + % Conv. = + FIOPC (A05) = *	[Gefitinib] = 1 mM TON = + % Conv. = + FIOPC (A05) = *	[Gefitinib] = 1 mM TON = + % Conv. = + FIOPC (A05) = *

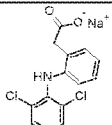
Table 3. Results for Variants 14 and 15 (with substitutions shown with reference to SEQ ID NO:4)			
SF Validation Substrate(s)	Backbone (SEQ ID NO:4)	Variant 14 (C48S;I95P;G115R; L216Y;D232H;M491A; N574T;D722S)	Variant 15 (K32R;C48S;I95P; L216Y;D232H;E349T; M491A;N574T;T577A; E619D;D722S)
 Diclofenac	[Diclofenac] = 1 mM	[Diclofenac] = 1 mM TON = ###	[Diclofenac] = 1 mM TON = ###

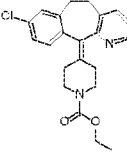
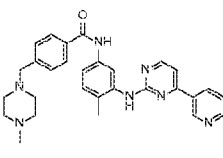
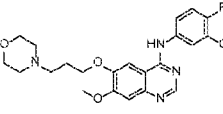
Table 3. Results for Variants 14 and 15 (with substitutions shown with reference to SEQ ID NO:4)			
SF Validation Substrate(s)	Backbone (SEQ ID NO:4)	Variant 14 (C48S;I95P;G115R; L216Y;D232H;M491A; N574T;D722S)	Variant 15 (K32R;C48S;I95P; L216Y;D232H;E349T; M491A;N574T;T577A; E619D;D722S)
Diclofenac	TON = ## % Conv. = +++	% Conv. = ++++ FIOPC (A07) = **	% Conv. = ++++ FIOPC (A07) = *
 Loratadine	[Loratadine] = 1 mM TON = ### % Conv. = ++++	[Loratadine] = 1 mM TON = ## % Conv. = +++ FIOPC (A07) = *	[Loratadine] = 1 mM TON = ### % Conv. = ++++ FIOPC (A07) = *
 Imatinib	[Imatinib] = 1 mM TON = # % Conv. = ++	[Imatinib] = 1 mM TON = ## % Conv. = +++ FIOPC (A07) = *	[Imatinib] = 1 mM TON = ## % Conv. = ++++ FIOPC (A07) = **
 Gefitinib	[Gefitinib] = 1 mM TON = ‡ % Conv. = +	[Gefitinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (A07)= ****	[Gefitinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (A07)= ****

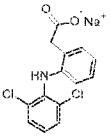
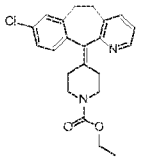
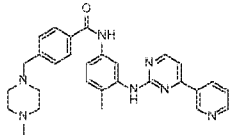
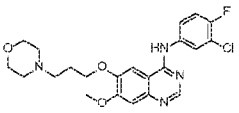
Table 4. Results for Variants 14 and 15 (with substitutions shown with reference to SEQ ID NO:6)			
SF Validation Substrate(s)	Backbone (SEQ ID NO:6)	Variant 16 (K32R;C48S;Y52F; G115R;L216Y;D232H; E349K;M491A;T577A; E619D;A627S;D722S)	Variant 17 (K32R;C48S;Y52F; I95P;Q111R;G115R; I176V;L216Y;D232H; E349T;M491A;T577A; E619D;D722S)
 Diclofenac	[Diclofenac] = 1 mM TON = # % Conv. = ++	[Diclofenac] = 1 mM TON = ### % Conv. = ++++ FIOPC (A12) = *	[Diclofenac] = 1 mM TON = ### % Conv. = ++++ FIOPC (A12) = *
 Loratadine	[Loratadine] = 1 mM TON = ‡ % Conv. = +	[Loratadine] = 1 mM TON = ‡ % Conv. = + FIOPC (A12) = *	[Loratadine] = 1 mM TON = ‡ % Conv. = + FIOPC (A12) = *
 Imatinib	[Imatinib] = 1 mM TON = # % Conv. = +	[Imatinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (A12) = ****	[Imatinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (A12) = ****
 Gefitinib	[Gefitinib] = 1 mM TON = # % Conv. = ++	[Gefitinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (A12) = ****	[Gefitinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (A12) = ***

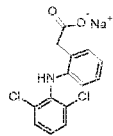
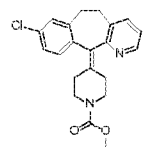
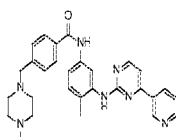
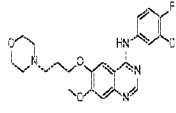
Table 5. Results for Variants 9, 10, and 11 (with substitutions shown with reference to SEQ ID NO:8)				
SF Validation Substrate(s)	Backbone	Variant 9 (Y52F;D232H;E349 K; E619D;D722S;R762 H)	Variant 10 (K32R;Y52F;K95P; Q111R;G115R;I176V; L216Y;D232H;P347T; E349T;M491A;T577A; E619D;D722S)	Variant 11 (K95P;Q111R;G115R; I176V;L216Y;E349T; M491A;T577A;E619 D; D722S)
 Diclofenac	[Diclofenac] = 1 mM TON = # % Conv. = ++	[Diclofenac] = 1 mM TON = ## % Conv. = ++++ FIOPC (B02) = *	[Diclofenac] = 1 mM TON = ## % Conv. = +++ FIOPC (B02) = *	[Diclofenac] = 1 mM TON = ## % Conv. = ++++ FIOPC (B02) = *
 Loratadine	[Loratadine] = 1 mM TON = # % Conv. = ++	[Loratadine] = 1 mM TON = # % Conv. = ++ FIOPC (B02) = *	[Loratadine] = 1 mM TON = # % Conv. = ++ FIOPC (B02) = *	[Loratadine] = 1 mM TON = ## % Conv. = +++ FIOPC (B02) = *
 Imatinib	[Imatinib] = 1 mM TON = ### % Conv. = ++++	[Imatinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (B02) = *	[Imatinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (B02) = **	[Imatinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (B02) = *
 Gefitinib	[Gefitinib] = 1 mM TON = + % Conv. = +	[Gefitinib] = 1 mM TON = + % Conv. = + FIOPC (B02) = *	[Gefitinib] = 1 mM TON = # % Conv. = ++ FIOPC (B02) = ****	[Gefitinib] = 1 mM TON = + % Conv. = + FIOPC (B02) = *

Table 6. Results for Variants 3, 4, and 5
(with substitutions shown with reference to SEQ ID NO:10)

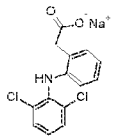
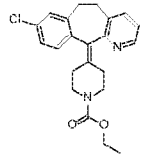
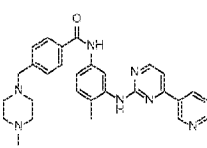
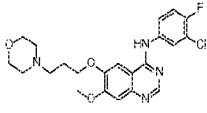
SF Validation Substrate(s)	Backbone	Variant 3 (K32R;C48S;Y52F; I95P;L216Y;D232H; E349K;M491A;N574T; D722S)	Variant 4 (K32R;C48S;Y52F; I95P;Q111R;L216Y; D232H;P347T;E349K; M491A;N574T;E619D; D722S)	Variant 5 (C48S;I95P;G115R; L216Y;S231R;D232H; M491A;N574T;T577A; E619D;D722S)
 Diclofenac	[Diclofenac] = 1 mM TON = ### % Conv. = ++++	[Diclofenac] = 1 mM TON = ### % Conv. = ++++ FIOPC (B12) = *	[Diclofenac] = 1 mM TON = ### % Conv. = ++++ FIOPC (B12) = *	[Diclofenac] = 1 mM TON = ### % Conv. = ++++ FIOPC (B12) = *
 Loratadine	[Loratadine] = 1 mM TON = ### % Conv. = ++++	[Loratadine] = 1 mM TON = ### % Conv. = ++++ FIOPC (B12) = *	[Loratadine] = 1 mM TON = ### % Conv. = ++++ FIOPC (B12) = *	[Loratadine] = 1 mM TON = ### % Conv. = ++++ FIOPC (B12) = *
 Imatinib	[Imatinib] = 1 mM TON = # % Conv. = ++	[Imatinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (B12) = ***	[Imatinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (B02) = ***	[Imatinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (B02) = ***
 Gefitinib	[Gefitinib] = 1 mM TON = # % Conv. =	[Gefitinib] = 1 mM TON = ### % Conv. = ++++	[Gefitinib] = 1 mM TON = ### % Conv. = ++++	[Gefitinib] = 1 mM TON = ### % Conv. = ++++

Table 6. Results for Variants 3, 4, and 5 (with substitutions shown with reference to SEQ ID NO:10)				
SF Validation Substrate(s)	Backbone	Variant 3 (K32R;C48S;Y52F; I95P;L216Y;D232H; E349K;M491A;N574T; D722S)	Variant 4 (K32R;C48S;Y52F; I95P;Q111R;L216Y; D232H;P347T;E349K; M491A;N574T;E619D; D722S)	Variant 5 (C48S;I95P;G115R; L216Y;S231R;D232H; M491A;N574T;T577A; E619D;D722S)
	++	FIOPC (B12) = ***	FIOPC (B12) = ***	FIOPC (B12) = ***

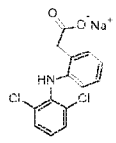
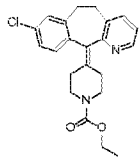
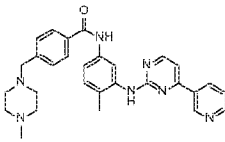
Table 7. Results for Variants 12 and 13 (with substitutions shown with reference to SEQ ID NO:12)			
SF Validation Substrate(s)	Backbone	Variant 12 (K32R;C48S;Y52F; I95P;Q111R;G115R; D232H;M491A;T577A; A627S;D722S)	Variant 13 (K32R;C48S;Y52F; I95P;G115R;D232H; M491A;N574T;T577A; E619D;A627S;D722S)
 Diclofenac	[Diclofenac] = 1 mM TON = # % Conv. = +	[Diclofenac] = 1 mM TON = ## % Conv. = +++ FIOPC (D06) = **	[Diclofenac] = 1 mM TON = ## % Conv. = ++++ FIOPC (D06) = ***
 Loratadine	[Loratadine] = 1 mM TON = # % Conv. = +	[Loratadine] = 1 mM TON = # % Conv. = ++ FIOPC (D06) = *	[Loratadine] = 1 mM TON = # % Conv. = + FIOPC (D06) = *
 Imatinib	[Imatinib] = 1 mM TON = # % Conv. =	[Imatinib] = 1 mM TON = ## % Conv. = ++++	[Imatinib] = 1 mM TON = ## % Conv. = ++++

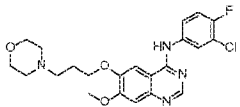
Table 7. Results for Variants 12 and 13 (with substitutions shown with reference to SEQ ID NO:12)			
SF Validation Substrate(s)	Backbone	Variant 12 (K32R;C48S;Y52F; I95P;Q111R;G115R; D232H;M491A;T577A; A627S;D722S)	Variant 13 (K32R;C48S;Y52F; I95P;G115R;D232H; M491A;N574T;T577A; E619D;A627S;D722S)
	++	FIOPC (D06) = ***	FIOPC (D06) = ***
 Gefitinib	[Gefitinib] = 1 mM TON = ‡ % Conv. = +	[Gefitinib] = 1 mM TON = ‡ % Conv. = + FIOPC (D06) = *	[Gefitinib] = 1 mM TON = # % Conv. = + FIOPC (D06) = ****

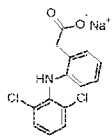
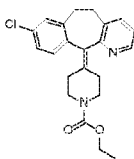
Table 8. Results for Variants 1 and 2 (with substitutions shown with reference to SEQ ID NO:14)			
SF Validation Substrate(s)	Backbone	Variant 1 (Y52F;K95P;G115R; T176V;L216Y;D232H; E349T;M491A;T577A; E619D)	Variant 2 (Y52F;G115R;T176V; L216Y;D232H;P347T; E349T;T577A;D722S)
 Diclofenac	[Diclofenac] = 1 mM TON = ‡ % Conv. = +	[Diclofenac] = 1 mM TON = # % Conv. = + FIOPC (F02) = ****	[Diclofenac] = 1 mM TON = # % Conv. = + FIOPC (F02) = ****
 Loratadine	[Loratadine] = 1 mM TON = # % Conv. = +	[Loratadine] = 1 mM TON = # % Conv. = + FIOPC (F02) = *	[Loratadine] = 1 mM TON = # % Conv. = + FIOPC (F02) = *

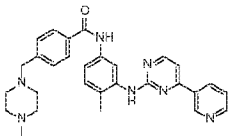
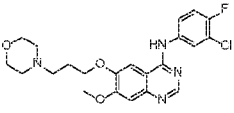
Table 8. Results for Variants 1 and 2 (with substitutions shown with reference to SEQ ID NO:14)			
SF Validation Substrate(s)	Backbone	Variant 1 (Y52F;K95P;G115R; T176V;L216Y;D232H; E349T;M491A;T577A; E619D)	Variant 2 (Y52F;G115R;T176V; L216Y;D232H;P347T; E349T;T577A;D722S)
 Imatinib	[Imatinib] = 1 mM TON = # % Conv. = +	[Imatinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (F02) = ****	[Imatinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (F02) = ****
 Gefitinib	[Gefitinib] = 1 mM TON = † % Conv. = +	[Gefitinib] = 1 mM TON = † % Conv. = + FIOPC (F02) = *	[Gefitinib] = 1 mM TON = † % Conv. = + FIOPC (F02) = *

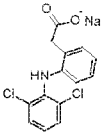
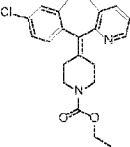
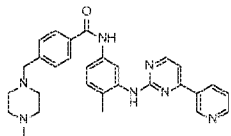
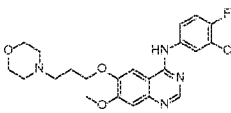
Table 9. Results for Variants 18 and 19 (with substitutions shown with reference to SEQ ID NO:16)			
SF Validation Substrate(s)	Backbone	Variant 18 (K32R;C48S;Y52F; I95P;I176V;L216Y; D232H;E349K;M491A; N574T;E619D;D722S; A767T)	Variant 19 (C48S;Y52F;I95P; Q111R;I176V;L216Y; D232H;P347T;M491A; N574T;D722S)
 Diclofenac	[Diclofenac] = 1 mM TON = ### % Conv. =	[Diclofenac] = 1 mM TON = ### % Conv. = ++++	[Diclofenac] = 1 mM TON = ### % Conv. = ++++

Table 9. Results for Variants 18 and 19 (with substitutions shown with reference to SEQ ID NO:16)			
SF Validation Substrate(s)	Backbone	Variant 18 (K32R,C48S;Y52F; I95P;I176V;L216Y; D232H;E349K;M491A; N574T;E619D;D722S; A767T)	Variant 19 (C48S;Y52F;I95P; Q111R;I176V;L216Y; D232H;P347T;M491A; N574T;D722S)
	++++	FIOPC (F12) = *	FIOPC (F12) = *
 Loratadine	[Loratadine] = 1 mM TON = ## % Conv. = +++	[Loratadine] = 1 mM TON = ## % Conv. = ++ FIOPC (F12) = *	[Loratadine] = 1 mM TON = ## % Conv. = +++ FIOPC (F02) = *
 Imatinib	[Imatinib] = 1 mM TON = # % Conv. = +	[Imatinib] = 1 mM TON = ## % Conv. = +++ FIOPC (F12) = **	[Imatinib] = 1 mM TON = ## % Conv. = +++ FIOPC (F12) = **
 Gefitinib	[Gefitinib] = 1 mM TON = # % Conv. = +	[Gefitinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (F12) = ****	[Gefitinib] = 1 mM TON = ### % Conv. = ++++ FIOPC (F12) = ****

[0116] While the invention has been described with reference to the specific embodiments, various changes can be made and equivalents can be substituted to adapt to a particular situation, material, composition of matter, process, process step or steps, thereby achieving benefits of the invention without departing from the scope of what is claimed.

[0117] For all purposes in the United States of America, each and every publication and patent document cited in this disclosure is incorporated herein by reference as if each such publication or

document was specifically and individually indicated to be incorporated herein by reference. Citation of publications and patent documents is not intended as an indication that any such document is pertinent prior art, nor does it constitute an admission as to its contents or date.

The Claims defining the Invention are as follows:

1. A recombinant cytochrome P450-BM3 variant having at least 80% sequence identity to a polypeptide sequence comprising the sequence set forth in SEQ ID NO:2;

wherein the variant has a polypeptide sequence comprising the sequence selected from the group comprising:

SEQ ID NO:2:

SEQ ID NO:2 comprising substitutions V51S; D118R; L219Y; K236H; P351T; S353T; N579T; T582A; E624D; A632S; D727S; and Q931K;

SEQ ID NO:2 comprising substitutions V51S; D118R; T179V; L219Y; P351T; N579T; T582A; E624D; D727S;

SEQ ID NO:2 comprising substitutions D118R; L219Y; K236H; S353T; N579T; T582A; E624D; D727S;

SEQ ID NO:4:

SEQ ID NO:4 comprising substitutions C48S; I95P; G115R; L216Y; D232H; M491A; N574T; and D722S;

SEQ ID NO:4 comprising substitutions K32R; C48S; I95P; L216Y; D232H; E349T; M491A; N574T; T577A; E619D; and D722S;

SEQ ID NO:6:

SEQ ID NO:6 comprising substitutions K32R; C48S; Y52F; G115R; L216Y; D232H; E349K; M491A; T577A; E619D; A627S; and D722S;

SEQ ID NO:6 comprising substitutions K32R; C48S; Y52F; I95P; Q111R; G115R; I176V; L216Y; D232H; E349T; M491A; T577A; E619D; and D722S;

SEQ ID NO:8:

SEQ ID NO:8 comprising substitutions Y52F; D232H; E349K; E619D; D722S; R762H;

SEQ ID NO:8 comprising substitutions K32R; Y52F; K95P; Q111R; G115R; I176V; L216Y; D232H; P347T; E349T; M491A; T577A; E619D; and D722S;

SEQ ID NO:8 comprising substitutions K95P; Q111R; G115R; I176V; L216Y; E349T; M491A; T577A; E619D; and D722S;

SEQ ID NO:10:

SEQ ID NO:10 comprising substitutions K32R; C48S; Y52F; I95P; L216Y; D232H; E349K; M491A; N574T; and D722S;

SEQ ID NO:10 comprising substitutions K32R; C48S; Y52F; I95P; Q111R; L216Y; D232H; P347T; E349K; M491A; N574T; E619D; and D722S;

SEQ ID NO:10 comprising substitutions C48S; I95P; G115R; L216Y; S231R; D232H; M491A; N574T; T577A; E619D; and D722S;

SEQ ID NO:12:

SEQ ID NO:12 comprising substitutions K32R; C48S; Y52F; I95P; Q111R; G115R; D232H; M491A; T577A; A627S; D722S;

SEQ ID NO:12 comprising substitutions K32R; C48S; Y52F; I95P; G115R; D232H; M491A; N574T; T577A; E619D; A627S; and D722S;

SEQ ID NO:14:

SEQ ID NO:14 comprising substitutions Y52F; K95P; G115R; T176V; L216Y; D232H; E349T; M491A; T577A; and E619D;

SEQ ID NO:14 comprising substitutions Y52F; G115R; T176V; L216Y; D232H; P347T; E349T; T577A; D722S; or

SEQ ID NO:16:

SEQ ID NO:16 comprising substitutions K32R; C48S; Y52F; I95P; I176V; L216Y; D232H; E349K; M491A; N574T; E619D; D722S; A767T; and

SEQ ID NO:16 comprising substitutions C48S; Y52F; I95P; Q111R; I176V; L216Y; D232H; P347T; M491A; N574T; and D722S.

2. The recombinant cytochrome P450-BM3 variant of Claim 1, wherein the variant has at least 90% sequence identity to a polypeptide sequence comprising the sequence set forth in SEQ ID NOS:2, 4, 6, 8, 10, 12, 14, or 16.

3. The recombinant cytochrome P450-BM3 variant of Claim 1 or Claim 2, wherein said variant oxidizes at least three organic substrates.

4. The recombinant cytochrome P450-BM3 variant of Claim 3, wherein said organic substrate is selected from the group comprising loratadine, imatinib, gefitinib, and diclofenac.

5. The recombinant cytochrome P450-BM3 variant of any one of Claims 1 to 4, wherein the variant exhibits improved oxidizing activity relative to the oxidizing activity exhibited by SEQ ID NO:2, with respect to an organic substrate selected from the group comprising loratadine, imatinib, gefitinib, and diclofenac.

6. A recombinant polynucleotide sequence encoding the recombinant cytochrome P450-BM3 variant of any one of Claims 1 to 5.

7. The recombinant polynucleotide sequence of Claim 6, wherein said sequence comprises SEQ ID NO:1, 3, 5, 7, 9, 11, 13, or 15.

8. An expression vector comprising the polynucleotide sequence of Claim 6 or 7.

9. The vector of Claim 8, wherein said polynucleotide sequence is operably linked with regulatory sequences suitable for expression of said polynucleotide sequence in a suitable non-human host cell.

10. The vector of Claim 8 or Claim 9, wherein said non-human host cell is a prokaryotic or eukaryotic cell.

11. The vector of Claim 10, wherein said non-human host cell is a prokaryotic cell.

12. The vector of Claim 10 or Claim 11, wherein said non-human host cell is *E. coli*.

13. A non-human host cell comprising the vector of Claim 8 or Claim 9.

14. A method for producing at least one recombinant cytochrome P450-BM3 variant comprising culturing the non-human host cell of Claim 13 under conditions such that the recombinant cytochrome P450-BM3 variant of any of Claims 1 to 5 is produced by said non-human host cell.

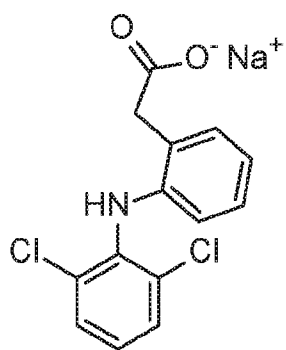
15. The method of Claim 14, further comprising the step of recovering said at least one recombinant cytochrome P450 variant.

ABSTRACT

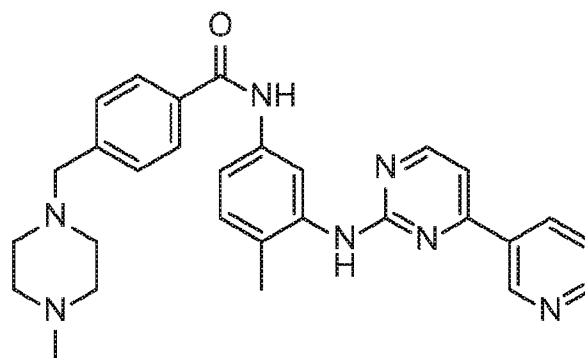
The present invention provides improved P450-BM3 variants with improved activity. In some embodiments, the P450-BM3 variants exhibit improved activity over a wide range of substrates.

FIGURE 1

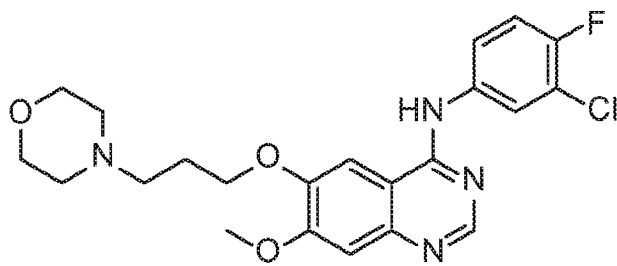
1 / 2



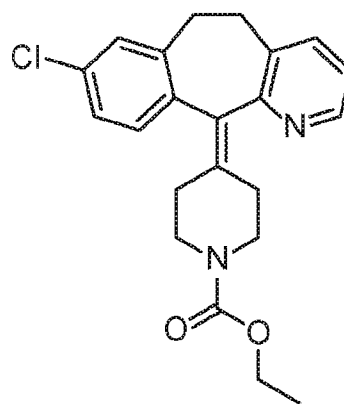
Diclofenac



Imatinib



Gefitinib



Loratadine

FIG. 1

2 / 2

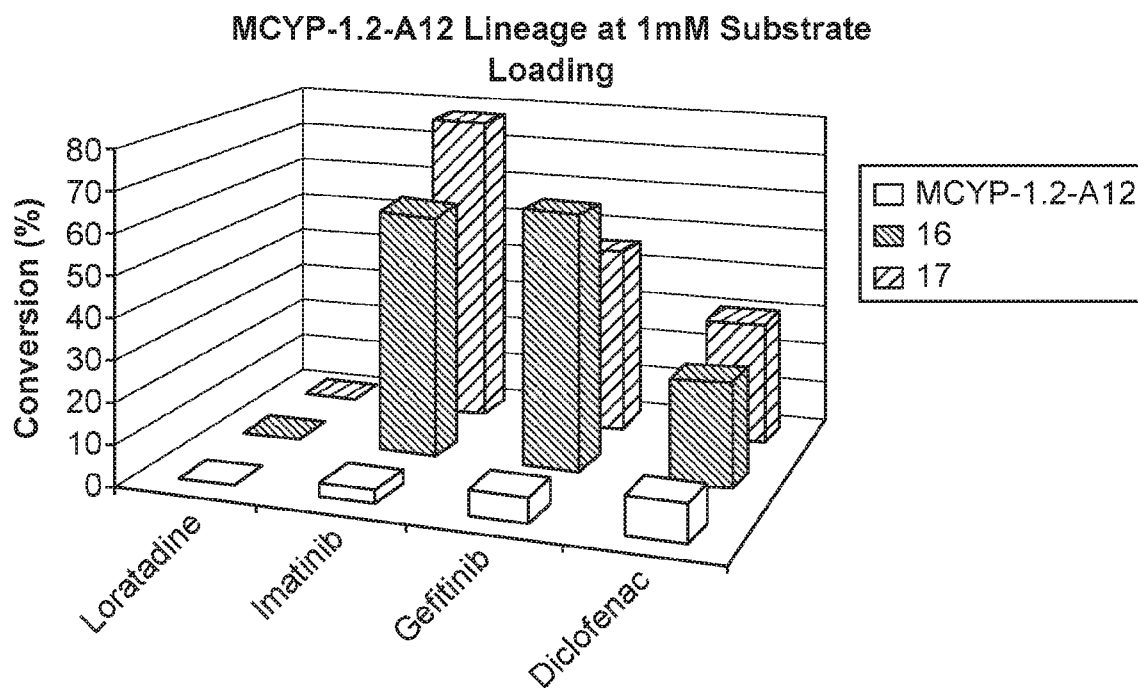


FIG. 2

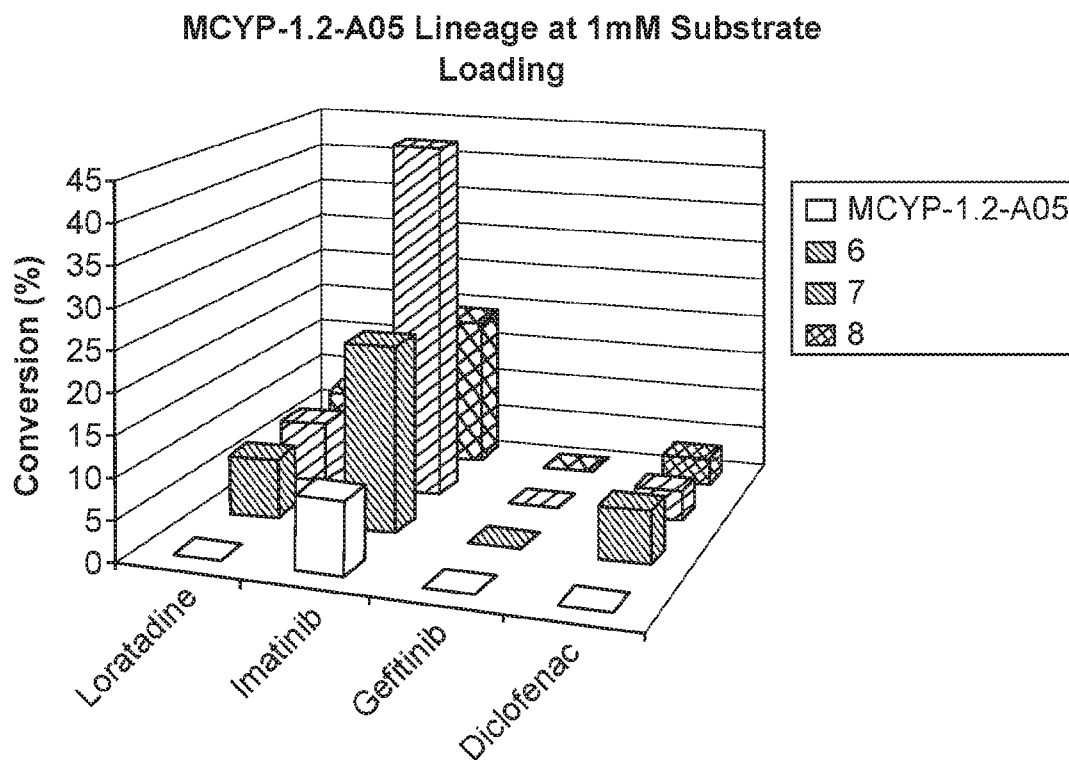


FIG. 3

hp3238sgw-sp.sequence listing_ST25
SEQUENCE LISTING

<110> CODEXIS, INC.
OSBORNE, Robert
MITCHELL, Vesna
HTWE, Khin Yu Naing
ZHANG, Xiyun
MILCZEK, Erika M.
MOORE, Jeffrey C.

<120> NOVEL P450-BM3 VARIANTS WITH IMPROVED ACTIVITY

<130> CX2-150USP1

<160> 16

<210> 1

<211> 3165

<212> DNA

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 1

```
atgttaatga aacaggcaag cgcaatacct cagcccaaaa catacggacc tttaaaaaat 60
cttccgcatac tggaaaaaga acagctttct caatccttat ggcggatagc tgatgaattg 120
ggaccgattt tccgttttga ttttccggga gtatccagtg tttttgtgtc cggccacaat 180
cttgtggctg aagtgtgtga tgaagaacgg tttgataaaa gcattgaagg cgccttgga 240
aaggttcgcg cattttccgg tgacggattg gccactagtt ggacgcatga gcctaactgg 300
agaaaagcgc acaacattct gatgccgacg ttcagccagc gggccatgaa ggactatcat 360
gagaaaatgg tcgatatcgc aaccagctg attcaaaagt ggagccggtt aaaccccaat 420
gaagaaattg atgtagcgga cgatatgaca cgtctgacgc ttgatacgat tgggttatgc 480
gggtttaact atcgattcaa cagcttttac cgagatcagc ctcatccatt tattacaagt 540
atggtccgtg cactggatga agcaatgaac aagctgcagc gagcaaatcc agacgaccca 600
gcttatgatg aaaacaagcg ccagtttcaa gaagatatca aggtgatgaa cgacctagtc 660
gacagcatta ttgcagagcg cagggcgaat ggagaccagg atgaaaaaga tttgctcgcc 720
cgcatgctga atgtggaaga tccggaaact ggtgaaaagc tcgacgacga aaatatccgc 780
tttcagatca tcacgttttt gattgccggc catgaaacaa caagcgggtt gctatccttt 840
gcgattttatt gtctgcttac acatccggaa aaactgaaaa aagctcagga ggaagcggat 900
cgcgtgttaa cggatgacac gcctgaatat aaacaaatcc agcagctcaa atacattcgg 960
atggttttaa atgaaaccct cagactgtat ccaacagctc cggctttttc tctatatgctg 1020
aaggaggata ctgttcttgg cggggaatat ccgatcagca aagggcagcc agtcactgtt 1080
ttaattccaa aactgcaccg ggatcaaaac gcttggggac cggatgcgga agatttccgt 1140
ccggaacggt ttgaagatcc ttcaagtatc cctcaccatg cgtataagcc gtttggaac 1200
ggacagcgcg cttgtattgg catgcagttt gctcttcaag aagcgacaat ggttctcgg 1260
cttgatttaa agcattttga attgataaac catactggct acgaactaaa aatcaaagaa 1320
gcattaacga tcaagccgga tgattttaaa attactgtga aaccgcgaaa aacagcggca 1380
atcaatgtac agagaaaaga acaggcagaa cagtctgcta aaaaagtacg caaaaaggca 1440
gaaaacgctc ataatacgcc gctgcttggt ctatacggtt caaatatggg aacagctgaa 1500
ggaacggcgc gtgatttagc agatattgca atgagcaaag gatttgcacc gcaggtcgca 1560
acgcttgatt cacacgccgg aaatcttccg cgcaaggag ctgtattaat tgtaacggcg 1620
```

hp3238sgw-sp.sequence listing_ST25

```

tcttataacg gtcacccgcc tgataacgca aagcaatttg tcgactgggt agaccaagcg 1680
tctgctgatg aagtaaaagg cgttcgctac tccgtatttg gatgcggcga taaaaactgg 1740
gctactacgt atcaaaaagt gcctgctttt atcgatgaaa cgcttgccgc taaaggggca 1800
gaaaacatcg ctgaccgcgg tgaagcagat gcaagcgacg actttgaagg cacatatgaa 1860
gaatggcgtg aacatatgtg gagtgacgta gcagcctact ttaacctcga cattgaaaac 1920
agtgaagata ataaatctac tctttcactt caatttgtcg acagcgccgc ggatatgccg 1980
cttgcgaaaa tgcacgggtg gttttcaacg aacgtcgtag caagcaaaga acttcaacag 2040
ccaggcagtg cacgaagcac gcgacatctt gaaattgaac ttccaaaaga agcttcttat 2100
caagaaggag atcatttagg tgttattcct cgcaactatg aaggaatagt aaaccgtgta 2160
acagcaaggt tcggcctaga tgcatacacag caaatccgtc tggaagcaga agaagaaaaa 2220
ttagctcatt tgccactcgc taaaacagta tccgtagaag agcttctgca atacgtggag 2280
cttcaagatc ctgttacgcg cacgcagctt cgcgcaatgg ctgctaaaac ggtctgcccg 2340
ccgcataaag tagagcttga agccttgctt gaaaagcaag cctacaaaga acaagtgtcg 2400
gcaaaacgtt taacaatgct tgaactgctt gaaaaatacc cggcgtgtga aatgaaattc 2460
agcgaattta tcgcccttct gccaagcata cgcccgcgct attactcgat ttcttcatca 2520
cctcgtgtcg atgaaaaaca agcaagcatc acggtcagcg ttgtctcagg agaagcgtgg 2580
agcggatatg gagaatataa aggaattgcg tcgaactatc ttgccgagct gcaagaagga 2640
gatacgatta cgtgctttat ttccacaccg cagtcagaat ttacgctgcc aaaagaccct 2700
gaaacgccgc ttatcatggt cggaccggga acaggcgtcg cgccgtttag aggctttgtg 2760
caggcgcgca aacagctaaa agaacaagga cagtcacttg gagaagcaca ttatacttc 2820
ggctgccgtt cacctcatga agactatctg tatcaagaag agcttgaaaa cgcccaaagc 2880
gaaggcatca ttacgcttca taccgctttt tctcgcatgc caaatcagcc gaaaacatac 2940
gttcagcacg taatggaaca agacggcaag aaattgattg aacttcttga tcaaggagcg 3000
cacttctata tttcgggaga cggaagccaa atggcacctg ccgttgaagc aacgcttatg 3060
aaaagctatg ctgacgttca ccaagtgagt gaagcagacg ctcgcttatg gctgcagcag 3120
ctagaagaaa aaggccgata cgcaaaagac gtgtgggctg ggtaa 3165

```

<210> 2

<211> 1054

<212> PRT

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 2

```

Met Leu Met Lys Gln Ala Ser Ala Ile Pro Gln Pro Lys Thr Tyr Gly
1          5          10          15
Pro Leu Lys Asn Leu Pro His Leu Glu Lys Glu Gln Leu Ser Gln Ser
20          25          30
Leu Trp Arg Ile Ala Asp Glu Leu Gly Pro Ile Phe Arg Phe Asp Phe
35          40          45
Pro Gly Val Ser Ser Val Phe Val Ser Gly His Asn Leu Val Ala Glu
50          55          60
Val Cys Asp Glu Glu Arg Phe Asp Lys Ser Ile Glu Gly Ala Leu Glu
65          70          75          80
Lys Val Arg Ala Phe Ser Gly Asp Gly Leu Ala Thr Ser Trp Thr His
85          90          95
Glu Pro Asn Trp Arg Lys Ala His Asn Ile Leu Met Pro Thr Phe Ser
100         105         110
Gln Arg Ala Met Lys Asp Tyr His Glu Lys Met Val Asp Ile Ala Thr

```

hp3238sgw-sp.sequence listing ST25

		115					120				125				
Gln	Leu	Ile	Gln	Lys	Trp	Ser	Arg	Leu	Asn	Pro	Asn	Glu	Glu	Ile	Asp
	130					135					140				
Val	Ala	Asp	Asp	Met	Thr	Arg	Leu	Thr	Leu	Asp	Thr	Ile	Gly	Leu	Cys
145					150					155					160
Gly	Phe	Asn	Tyr	Arg	Phe	Asn	Ser	Phe	Tyr	Arg	Asp	Gln	Pro	His	Pro
				165					170					175	
Phe	Ile	Thr	Ser	Met	Val	Arg	Ala	Leu	Asp	Glu	Ala	Met	Asn	Lys	Leu
			180					185				190			
Gln	Arg	Ala	Asn	Pro	Asp	Asp	Pro	Ala	Tyr	Asp	Glu	Asn	Lys	Arg	Gln
		195					200					205			
Phe	Gln	Glu	Asp	Ile	Lys	Val	Met	Asn	Asp	Leu	Val	Asp	Ser	Ile	Ile
	210					215					220				
Ala	Glu	Arg	Arg	Ala	Asn	Gly	Asp	Gln	Asp	Glu	Lys	Asp	Leu	Leu	Ala
225					230					235					240
Arg	Met	Leu	Asn	Val	Glu	Asp	Pro	Glu	Thr	Gly	Glu	Lys	Leu	Asp	Asp
			245					250						255	
Glu	Asn	Ile	Arg	Phe	Gln	Ile	Ile	Thr	Phe	Leu	Ile	Ala	Gly	His	Glu
			260					265				270			
Thr	Thr	Ser	Gly	Leu	Leu	Ser	Phe	Ala	Ile	Tyr	Cys	Leu	Leu	Thr	His
		275					280					285			
Pro	Glu	Lys	Leu	Lys	Lys	Ala	Gln	Glu	Glu	Ala	Asp	Arg	Val	Leu	Thr
	290					295					300				
Asp	Asp	Thr	Pro	Glu	Tyr	Lys	Gln	Ile	Gln	Gln	Leu	Lys	Tyr	Ile	Arg
305					310					315					320
Met	Val	Leu	Asn	Glu	Thr	Leu	Arg	Leu	Tyr	Pro	Thr	Ala	Pro	Ala	Phe
			325					330						335	
Ser	Leu	Tyr	Ala	Lys	Glu	Asp	Thr	Val	Leu	Gly	Gly	Glu	Tyr	Pro	Ile
			340					345				350			
Ser	Lys	Gly	Gln	Pro	Val	Thr	Val	Leu	Ile	Pro	Lys	Leu	His	Arg	Asp
		355					360					365			
Gln	Asn	Ala	Trp	Gly	Pro	Asp	Ala	Glu	Asp	Phe	Arg	Pro	Glu	Arg	Phe
	370					375					380				
Glu	Asp	Pro	Ser	Ser	Ile	Pro	His	His	Ala	Tyr	Lys	Pro	Phe	Gly	Asn
385					390					395					400
Gly	Gln	Arg	Ala	Cys	Ile	Gly	Met	Gln	Phe	Ala	Leu	Gln	Glu	Ala	Thr
				405					410					415	
Met	Val	Leu	Gly	Leu	Val	Leu	Lys	His	Phe	Glu	Leu	Ile	Asn	His	Thr
			420					425					430		
Gly	Tyr	Glu	Leu	Lys	Ile	Lys	Glu	Ala	Leu	Thr	Ile	Lys	Pro	Asp	Asp
		435					440					445			
Phe	Lys	Ile	Thr	Val	Lys	Pro	Arg	Lys	Thr	Ala	Ala	Ile	Asn	Val	Gln
	450					455					460				
Arg	Lys	Glu	Gln	Ala	Glu	Gln	Ser	Ala	Lys	Lys	Val	Arg	Lys	Lys	Ala
465															

hp3238sgw-sp.sequence listing_ST25

	530					535					540				
His 545	Pro	Pro	Asp	Asn	Ala 550	Lys	Gln	Phe	Val	Asp 555	Trp	Leu	Asp	Gln	Ala 560
Ser	Ala	Asp	Glu	Val 565	Lys	Gly	Val	Arg	Tyr 570	Ser	Val	Phe	Gly	Cys	Gly 575
Asp	Lys	Asn	Trp 580	Ala	Thr	Thr	Tyr	Gln 585	Lys	Val	Pro	Ala	Phe	Ile	Asp 590
Glu	Thr	Leu 595	Ala	Ala	Lys	Gly	Ala 600	Glu	Asn	Ile	Ala	Asp 605	Arg	Gly	Glu 610
Ala	Asp 610	Ala	Ser	Asp	Asp	Phe 615	Glu	Gly	Thr	Tyr	Glu 620	Glu	Trp	Arg	Glu 625
His 625	Met	Trp	Ser	Asp 630	Val	Ala	Ala	Tyr	Phe	Asn 635	Leu	Asp	Ile	Glu	Asn 640
Ser	Glu	Asp	Asn 645	Lys	Ser	Thr	Leu	Ser	Leu 650	Gln	Phe	Val	Asp	Ser	Ala 655
Ala	Asp	Met 660	Pro	Leu	Ala	Lys	Met	His 665	Gly	Ala	Phe	Ser	Thr 670	Asn	Val 675
Val	Ala 675	Ser	Lys	Glu	Leu	Gln	Gln 680	Pro	Gly	Ser	Ala	Arg 685	Ser	Thr	Arg 690
His 690	Leu	Glu	Ile	Glu	Leu	Pro 695	Lys	Glu	Ala	Ser	Tyr 700	Gln	Glu	Gly	Asp 705
His 705	Leu	Gly	Val	Ile 710	Pro	Arg	Asn	Tyr	Glu	Gly 715	Ile	Val	Asn	Arg	Val 720
Thr	Ala	Arg	Phe 725	Gly	Leu	Asp	Ala	Ser	Gln 730	Gln	Ile	Arg	Leu	Glu	Ala 735
Glu	Glu	Glu 740	Lys	Leu	Ala	His	Leu	Pro 745	Leu	Ala	Lys	Thr	Val	Ser	Val 750
Glu	Glu 755	Leu	Leu	Gln	Tyr	Val	Glu 760	Leu	Gln	Asp	Pro	Val	Thr	Arg	Thr 765
Gln 770	Leu	Arg	Ala	Met	Ala 775	Ala	Lys	Thr	Val	Cys	Pro 780	Pro	His	Lys	Val 785
Glu 785	Leu	Glu	Ala	Leu 790	Leu	Glu	Lys	Gln	Ala	Tyr 795	Lys	Glu	Gln	Val	Leu 800
Ala	Lys	Arg	Leu 805	Thr	Met	Leu	Glu	Leu 810	Leu	Glu	Lys	Tyr	Pro	Ala	Cys 815
Glu	Met	Lys 820	Phe	Ser	Glu	Phe	Ile	Ala 825	Leu	Leu	Pro	Ser	Ile	Arg	Pro 830
Arg	Tyr	Tyr 835	Ser	Ile	Ser	Ser	Ser 840	Pro	Arg	Val	Asp	Glu	Lys	Gln	Ala 845
Ser 850	Ile	Thr	Val	Ser	Val	Val	Ser 855	Gly	Glu	Ala	Trp	Ser	Gly	Tyr	Gly 860
Glu 865	Tyr	Lys	Gly	Ile 870	Ala	Ser	Asn	Tyr	Leu	Ala 875	Glu	Leu	Gln	Glu	Gly 880
Asp	Thr	Ile	Thr 885	Cys	Phe	Ile	Ser	Thr	Pro 890	Gln	Ser	Glu	Phe	Thr	Leu 895
Pro	Lys	Asp 900	Pro	Glu	Thr	Pro	Leu	Ile 905	Met	Val	Gly	Pro	Gly	Thr	Gly 910
Val	Ala 915	Pro	Phe	Arg	Gly	Phe	Val	Gln 920	Ala	Arg	Lys	Gln	Leu	Lys	Glu 925
Gln	Gly 930	Gln	Ser	Leu	Gly	Glu	Ala 935	His	Leu	Tyr	Phe 940	Gly	Cys	Arg	Ser 945
Pro	His	Glu	Asp	Tyr	Leu	Tyr	Gln	Glu	Glu	Leu	Glu	Asn	Ala	Gln	Ser 950

hp3238sgw-sp.sequence listing_ST25

945		950		955		960
Glu Gly Ile Ile Thr	Leu His Thr Ala Phe	Ser Arg Met Pro Asn Gln				
	965		970		975	
Pro Lys Thr Tyr Val	Gln His Val Met Glu Gln	Asp Gly Lys Lys Leu				
	980		985		990	
Ile Glu Leu Leu Asp	Gln Gly Ala His Phe Tyr	Ile Cys Gly Asp Gly				
	995		1000		1005	
Ser Gln Met Ala Pro	Ala Val Glu Ala Thr	Leu Met Lys Ser Tyr Ala				
	1010		1015		1020	
Asp Val His Gln Val	Ser Glu Ala Asp Ala	Arg Leu Trp Leu Gln Gln				
	1025		1030		1035	
Leu Glu Glu Lys Gly	Arg Tyr Ala Lys Asp	Val Trp Ala Gly				
	1045		1050			

<210> 3

<211> 3150

<212> DNA

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 3

```

atgacaataa aggaaatgcc gcagccgaaa acgtttggag aacttaaaaa tttaccgtta 60
ttaaacacag ataaaccggt tcaagctttg atgaaaattg cggatgaatt aggagaaatc 120
tttaaattcg aggcgcctgg ttgtgtaacg cgctacttat caagtcagcg tctaattaaa 180
gaagcatgcg atgaatcacg ctttgataaa aacttaagtc aagcgcttaa atttgcacgt 240
gattttgcag gagacgggtt ggtgacaagc tggacgcagc aaataaattg gaaaaaagcg 300
cataatatct tacttccaag ctttagtcag caggcaatga aaggctatca tgcgatgatg 360
gtcgatatcg ccgtgcagct tgttcaaaaag tgggagcgtc taaatgcaga tgagcatatt 420
gaagtatcgg aagacatgac acgtttaacg cttgatacaa ttggtctttg cggctttaac 480
tatcgcttta acagctttta ccgagatcag cctcatccat ttattataag tatggtccgt 540
gcactggatg aagtaatgaa caagctgcag cgagcaaadc cagacgaccc agcttatgat 600
gaaaacaagc gccagtgtca agaagatatc aaggtgatga acgacctagt agataaaatt 660
attgcagatc gcaaagcaag ggggtgaaca agcgatgatt tattaacgca gatgctaaac 720
ggaaaagatc cagaaacggg tgagccgctt gatgacggga acattagcta tcaaattatt 780
acattcttaa ttgcgggaca cgaaacaaca agtggtcttt tatcatttgc gctgtatttc 840
ttagtgaaaa atccacatgt attacaaaaa gtagcagaag aagcagcacg agttctagta 900
gatcctgttc caagctacaa acaagtcaaa cagcttaaat atgtcggcat ggtcttaaac 960
gaagcgctgc gcttatggcc aactgctcct gcgttttccc tatatgcaaa agaagatacg 1020
gtgcttggag gagaatatcc tttagaaaaa ggcgacgaag taatggttct gattcctcag 1080
cttcaccgtg ataaaacaat ttggggagac gatgtggagg agttccgtcc agagcgtttt 1140
gaaaatccaa gtgcgattcc gcagcatgcg tttaaaccgt ttggaaacgg tcagcgctgcg 1200
tgtatcggtc agcagttcgc tcttcatgaa gcaacgtgg tacttggtat gatgctaaaa 1260
cactttgact ttgaagatca tacaaactac gagctcgata ttaaagaaac tttaacgtta 1320
aaacctgaag gctttgtggt aaaagcaaaa tcgaaaaaaa ttccgcttgg cggtatctct 1380
tcacctagca ctgaacagtc tgctaaaaaa gtacgcaaaa aggcagaaaa cgctcataat 1440
acgccgtgc ttgtgctata cggttcaaat atgggaacag ctgaaggaaac ggcgctgat 1500
ttagcagata ttgcaatgag caaaggattt gcaccgcagg tcgcaacgct tgattcacac 1560
gccggaaatc ttccgcgcga aggagctgta ttaattgtaa cggcgtctta taacggtcac 1620
ccgctgata acgcaaagca atttgtcgac tggtagacc aagcgtctgc tgatgaagta 1680

```

hp3238sgw-sp.sequence listing_ST25

```

aaaggcggttc gctactccgt atttggatgc ggcgataaaa actgggctac tacgtatcaa 1740
aaagtgcctg cttttatcga tgaaacgctt gccgctaaag gggcagaaaa catcgctgac 1800
cgcggtgaag cagatgcaag cgacgacttt gaaggcacat atgaagaatg gcgtgaacat 1860
atgtggagtg acgtagcagc ctactttaac ctcgacattg aaaacagtga agataataaa 1920
tctactcttt cacttcaatt tgtcgacagc gccgcggata tgccgcttgc gaaaatgcac 1980
ggtgcgtttt caacgaacgt cgtagcaagc aaagaacttc aacagccagg cagtgcacga 2040
agcacgcgac atcttgaaat tgaacttcca aaagaagctt cttatcaaga aggagatcat 2100
ttaggtgtta ttccctcgaa ctatgaagga atagtaaacc gtgtaacagc aaggttcggc 2160
ctagatgcat cacagcaaat ccgtctggaa gcagaagaag aaaaattagc tcatttgcca 2220
ctcgctaaaa cagtatccgt agaagagctt ctgcaatacg tggagcttca agatcctgtt 2280
acgcgcacgc agcttcgcgc aatggctgct aaaacggtct gcccgccgca taaagtagag 2340
cttgaagcct tgcttgaaaa gcaagcctac aaagaacaag tgctggcaaa acgtttaaca 2400
atgcttgaaac tgcttgaaaa ataccggcg tgtgaaatga aattcagcga atttatcgcc 2460
cttctgccaa gcatacgccc gcgctattac tcgatttctt catcacctcg tgtcgatgaa 2520
aaacaagcaa gcatcacggt cagcgttgct tcaggagaag cgtggagcgg atatggagaa 2580
tataaaggaa ttgcgtcgaa ctatcttgcc gagctgcaag aaggagatac gattacgtgc 2640
tttattttcca caccgcagtc agaatttacg ctgccaaaag accctgaaac gccgcttattc 2700
atggctcggac cgggaacagg cgtcgcgccg tttagaggct ttgtgcaggc gcgcaaacag 2760
ctaaaagaac aaggacagtc acttgagaa gcacatttat acttcggctg ccgttcacct 2820
catgaagact atctgtatca agaagagctt gaaaacgccc aaagcgaagg catcattacg 2880
cttcataccg ctttttctcg catgccaaat cagccgaaaa catacgttca gcacgtaatg 2940
gaacaagacg gcaagaaatt gattgaactt cttgatcaag gagcgcactt ctatatttgc 3000
ggagacggaa gccaaatggc acctgccgtt gaagcaacgc ttatgaaaag ctatgctgac 3060
gttcaccaag tgagtgaagc agacgctgc ttatggctgc agcagctaga agaaaaaggc 3120
cgatacgcaa aagacgtgtg ggctgggtaa 3150

```

<210> 4

<211> 1049

<212> PRT

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 4

```

Met Thr Ile Lys Glu Met Pro Gln Pro Lys Thr Phe Gly Glu Leu Lys
1          5          10          15
Asn Leu Pro Leu Leu Asn Thr Asp Lys Pro Val Gln Ala Leu Met Lys
20        25        30
Ile Ala Asp Glu Leu Gly Glu Ile Phe Lys Phe Glu Ala Pro Gly Cys
35        40        45
Val Thr Arg Tyr Leu Ser Ser Gln Arg Leu Ile Lys Glu Ala Cys Asp
50        55        60
Glu Ser Arg Phe Asp Lys Asn Leu Ser Gln Ala Leu Lys Phe Ala Arg
65        70        75        80
Asp Phe Ala Gly Asp Gly Leu Val Thr Ser Trp Thr His Glu Ile Asn
85        90        95
Trp Lys Lys Ala His Asn Ile Leu Leu Pro Ser Phe Ser Gln Gln Ala
100       105       110
Met Lys Gly Tyr His Ala Met Met Val Asp Ile Ala Val Gln Leu Val
115       120       125

```

hp3238sgw-sp.sequence listing_ST25

Gln	Lys	Trp	Glu	Arg	Leu	Asn	Ala	Asp	Glu	His	Ile	Glu	Val	Ser	Glu	130	135	140
Asp	Met	Thr	Arg	Leu	Thr	Leu	Asp	Thr	Ile	Gly	Leu	Cys	Gly	Phe	Asn	145	150	155
Tyr	Arg	Phe	Asn	Ser	Phe	Tyr	Arg	Asp	Gln	Pro	His	Pro	Phe	Ile	Ile	165	170	175
Ser	Met	Val	Arg	Ala	Leu	Asp	Glu	Val	Met	Asn	Lys	Leu	Gln	Arg	Ala	180	185	190
Asn	Pro	Asp	Asp	Pro	Ala	Tyr	Asp	Glu	Asn	Lys	Arg	Gln	Cys	Gln	Glu	195	200	205
Asp	Ile	Lys	Val	Met	Asn	Asp	Leu	Val	Asp	Lys	Ile	Ile	Ala	Asp	Arg	210	215	220
Lys	Ala	Arg	Gly	Glu	Gln	Ser	Asp	Asp	Leu	Leu	Thr	Gln	Met	Leu	Asn	225	230	235
Gly	Lys	Asp	Pro	Glu	Thr	Gly	Glu	Pro	Leu	Asp	Asp	Gly	Asn	Ile	Ser	245	250	255
Tyr	Gln	Ile	Ile	Thr	Phe	Leu	Ile	Ala	Gly	His	Glu	Thr	Thr	Ser	Gly	260	265	270
Leu	Leu	Ser	Phe	Ala	Leu	Tyr	Phe	Leu	Val	Lys	Asn	Pro	His	Val	Leu	275	280	285
Gln	Lys	Val	Ala	Glu	Glu	Ala	Ala	Arg	Val	Leu	Val	Asp	Pro	Val	Pro	290	295	300
Ser	Tyr	Lys	Gln	Val	Lys	Gln	Leu	Lys	Tyr	Val	Gly	Met	Val	Leu	Asn	305	310	315
Glu	Ala	Leu	Arg	Leu	Trp	Pro	Thr	Ala	Pro	Ala	Phe	Ser	Leu	Tyr	Ala	325	330	335
Lys	Glu	Asp	Thr	Val	Leu	Gly	Gly	Glu	Tyr	Pro	Leu	Glu	Lys	Gly	Asp	340	345	350
Glu	Val	Met	Val	Leu	Ile	Pro	Gln	Leu	His	Arg	Asp	Lys	Thr	Ile	Trp	355	360	365
Gly	Asp	Asp	Val	Glu	Glu	Phe	Arg	Pro	Glu	Arg	Phe	Glu	Asn	Pro	Ser	370	375	380
Ala	Ile	Pro	Gln	His	Ala	Phe	Lys	Pro	Phe	Gly	Asn	Gly	Gln	Arg	Ala	385	390	395
Cys	Ile	Gly	Gln	Gln	Phe	Ala	Leu	His	Glu	Ala	Thr	Leu	Val	Leu	Gly	405	410	415
Met	Met	Leu	Lys	His	Phe	Asp	Phe	Glu	Asp	His	Thr	Asn	Tyr	Glu	Leu	420	425	430
Asp	Ile	Lys	Glu	Thr	Leu	Thr	Leu	Lys	Pro	Glu	Gly	Phe	Val	Val	Lys	435	440	445
Ala	Lys	Ser	Lys	Lys	Ile	Pro	Leu	Gly	Gly	Ile	Pro	Ser	Pro	Ser	Thr	450	455	460
Glu	Gln	Ser	Ala	Lys	Lys	Val	Arg	Lys	Lys	Ala	Glu	Asn	Ala	His	Asn	465	470	475
Thr	Pro	Leu	Leu	Val	Leu	Tyr	Gly	Ser	Asn	Met	Gly	Thr	Ala	Glu	Gly	485	490	495
Thr	Ala	Arg	Asp	Leu	Ala	Asp	Ile	Ala	Met	Ser	Lys	Gly	Phe	Ala	Pro	500	505	510
Gln	Val	Ala	Thr	Leu	Asp	Ser	His	Ala	Gly	Asn	Leu	Pro	Arg	Glu	Gly	515	520	525
Ala	Val	Leu	Ile	Val	Thr	Ala	Ser	Tyr	Asn	Gly	His	Pro	Pro	Asp	Asn	530	535	540

hp3238sgw-sp.sequence listing_ST25

Ala	Lys	Gln	Phe	Val	Asp	Trp	Leu	Asp	Gln	Ala	Ser	Ala	Asp	Glu	Val	545	550	555	560
Lys	Gly	Val	Arg	Tyr	Ser	Val	Phe	Gly	Cys	Gly	Asp	Lys	Asn	Trp	Ala	565	570	575	
Thr	Thr	Tyr	Gln	Lys	Val	Pro	Ala	Phe	Ile	Asp	Glu	Thr	Leu	Ala	Ala	580	585	590	
Lys	Gly	Ala	Glu	Asn	Ile	Ala	Asp	Arg	Gly	Glu	Ala	Asp	Ala	Ser	Asp	595	600	605	
Asp	Phe	Glu	Gly	Thr	Tyr	Glu	Glu	Trp	Arg	Glu	His	Met	Trp	Ser	Asp	610	615	620	
Val	Ala	Ala	Tyr	Phe	Asn	Leu	Asp	Ile	Glu	Asn	Ser	Glu	Asp	Asn	Lys	625	630	635	640
Ser	Thr	Leu	Ser	Leu	Gln	Phe	Val	Asp	Ser	Ala	Ala	Asp	Met	Pro	Leu	645	650	655	
Ala	Lys	Met	His	Gly	Ala	Phe	Ser	Thr	Asn	Val	Val	Ala	Ser	Lys	Glu	660	665	670	
Leu	Gln	Gln	Pro	Gly	Ser	Ala	Arg	Ser	Thr	Arg	His	Leu	Glu	Ile	Glu	675	680	685	
Leu	Pro	Lys	Glu	Ala	Ser	Tyr	Gln	Glu	Gly	Asp	His	Leu	Gly	Val	Ile	690	695	700	
Pro	Arg	Asn	Tyr	Glu	Gly	Ile	Val	Asn	Arg	Val	Thr	Ala	Arg	Phe	Gly	705	710	715	720
Leu	Asp	Ala	Ser	Gln	Gln	Ile	Arg	Leu	Glu	Ala	Glu	Glu	Glu	Lys	Leu	725	730	735	
Ala	His	Leu	Pro	Leu	Ala	Lys	Thr	Val	Ser	Val	Glu	Glu	Leu	Leu	Gln	740	745	750	
Tyr	Val	Glu	Leu	Gln	Asp	Pro	Val	Thr	Arg	Thr	Gln	Leu	Arg	Ala	Met	755	760	765	
Ala	Ala	Lys	Thr	Val	Cys	Pro	Pro	His	Lys	Val	Glu	Leu	Glu	Ala	Leu	770	775	780	
Leu	Glu	Lys	Gln	Ala	Tyr	Lys	Glu	Gln	Val	Leu	Ala	Lys	Arg	Leu	Thr	785	790	795	800
Met	Leu	Glu	Leu	Leu	Glu	Lys	Tyr	Pro	Ala	Cys	Glu	Met	Lys	Phe	Ser	805	810	815	
Glu	Phe	Ile	Ala	Leu	Leu	Pro	Ser	Ile	Arg	Pro	Arg	Tyr	Tyr	Ser	Ile	820	825	830	
Ser	Ser	Ser	Pro	Arg	Val	Asp	Glu	Lys	Gln	Ala	Ser	Ile	Thr	Val	Ser	835	840	845	
Val	Val	Ser	Gly	Glu	Ala	Trp	Ser	Gly	Tyr	Gly	Glu	Tyr	Lys	Gly	Ile	850	855	860	
Ala	Ser	Asn	Tyr	Leu	Ala	Glu	Leu	Gln	Glu	Gly	Asp	Thr	Ile	Thr	Cys	865	870	875	880
Phe	Ile	Ser	Thr	Pro	Gln	Ser	Glu	Phe	Thr	Leu	Pro	Lys	Asp	Pro	Glu	885	890	895	
Thr	Pro	Leu	Ile	Met	Val	Gly	Pro	Gly	Thr	Gly	Val	Ala	Pro	Phe	Arg	900	905	910	
Gly	Phe	Val	Gln	Ala	Arg	Lys	Gln	Leu	Lys	Glu	Gln	Gly	Gln	Ser	Leu	915	920	925	
Gly	Glu	Ala	His	Leu	Tyr	Phe	Gly	Cys	Arg	Ser	Pro	His	Glu	Asp	Tyr	930	935	940	
Leu	Tyr	Gln	Glu	Glu	Leu	Glu	Asn	Ala	Gln	Ser	Glu	Gly	Ile	Ile	Thr	945	950	955	960

hp3238sgw-sp.sequence listing_ST25

Leu	His	Thr	Ala	Phe	Ser	Arg	Met	Pro	Asn	Gln	Pro	Lys	Thr	Tyr	Val
				965					970					975	
Gln	His	Val	Met	Glu	Gln	Asp	Gly	Lys	Lys	Leu	Ile	Glu	Leu	Leu	Asp
			980					985					990		
Gln	Gly	Ala	His	Phe	Tyr	Ile	Cys	Gly	Asp	Gly	Ser	Gln	Met	Ala	Pro
		995					1000					1005			
Ala	Val	Glu	Ala	Thr	Leu	Met	Lys	Ser	Tyr	Ala	Asp	Val	His	Gln	Val
	1010					1015					1020				
Ser	Glu	Ala	Asp	Ala	Arg	Leu	Trp	Leu	Gln	Gln	Leu	Glu	Glu	Lys	Gly
1025					1030					1035					1040
Arg	Tyr	Ala	Lys	Asp	Val	Trp	Ala	Gly							
				1045											

<210> 5

<211> 3150

<212> DNA

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 5

atgacaataa	aggaaatgcc	gcagccgaaa	acgtttggag	aacttaaaaa	tttaccgtta	60
ttaaacacag	ataaacgggt	tcaagctttg	atgaaaattg	cggatgaatt	aggagaaatc	120
tttaaattcg	aggcgcctgg	ttgtgtaacg	cgctacttat	caagtcagcg	tctaattaaa	180
gaagcatgcg	atgaatcacg	ctttgataaa	aacttaagtc	aagcgcttaa	atttgcacgt	240
gattttcttg	gagacgggtt	atttacaagc	tggacgcatg	aaataaattg	gaaaaaagcg	300
cataatatct	tacttccaag	ctttagtcag	caggcaatga	aaggctatca	tgcgatgatg	360
gtcgatatcg	ccgtgcagct	tgttcaaaaag	tgggagcgtc	taaatgcaga	tgagcatatt	420
gaagtatcgg	aagacatgac	acgtttaacg	cttgatacaa	ttggtctttg	cggctttaac	480
tatcgcttta	acagctttta	ccgagatcag	cctcatccat	ttattataag	tatggtccgt	540
gacttggtatg	aagtaatgaa	caagctgcag	cgagcaaadc	cagacgaccc	agcttatgat	600
gaaaacaagc	gccagtgtca	agaagatatc	aaggtgatga	acgacctagt	agataaaaatt	660
attgcagatc	gcaaagcaag	gggtgaacaa	agcgatgatt	tattaacgca	gatgctaaac	720
ggaaaagatc	cagaaacggg	tgagccgctt	gatgacggga	acattagcta	tcaaattatt	780
acattcttaa	ttgcgggaca	cgaacaaca	agtggctctt	tatcatttgc	gctgtatttc	840
ttagtgaaaa	atccacatgt	attacaaaaa	gtagcagaag	aagcagcacg	agttctagta	900
gatcctgttc	caagctacaa	acaagtcaaa	cagcttaaat	atgtcggcat	ggtcttaaac	960
gaagcgctgc	gcttatggcc	aactgctcct	gcgttttccc	tatatgcaaa	agaagatacg	1020
gtgcttggag	gagaatatcc	tttagaaaaa	ggcgacgaag	taatggttct	gattcctcag	1080
cttcaccgtg	ataaaacaat	ttggggagac	gatgtggagg	agttccgtcc	agagcgtttt	1140
gaaaatccaa	gtgcgattcc	gcagcatgcg	tttaaaccgt	ttggaaacgg	tcagcggtgcg	1200
tgtatcggtc	agcagttcgc	tcttcatgaa	gcaacgctgg	tacttggtat	gatgctaaaa	1260
cactttgact	ttgaagatca	tacaaactac	gagctcgata	ttaaagaaac	tttaacgtta	1320
aaacctgaag	gctttgtggt	aaaagcaaaa	tcgaaaaaaa	ttccgcttgg	cggatttcct	1380
tcacctagca	ctgaacagtc	tgctaaaaaa	gtacgcaaaa	aggcagaaaa	cgctcataat	1440
acgccgctgc	ttgtgctata	cggttcaaat	atgggaacag	ctgaagggaac	ggcgcggtgat	1500
ttagcagata	ttgcaatgag	caaaggattt	gcaccgcagg	tcgcaacgct	tgattcacac	1560
gccggaaatc	ttccgcgcga	aggagctgta	ttaattgtaa	cggcgtctta	taacgggtcat	1620
ccgcctgata	acgcaaagca	atttgtcgac	tggttagacc	aagcgtctgc	tgatgaagta	1680
aaaggcgttc	gctactccgt	atttggatgc	ggcgataaaa	actgggctac	tacgtatcaa	1740

hp3238sgw-sp.sequence listing_ST25

```

aaagtgcctg cttttatcga tgaaacgctt gccgctaaag gggcagaaaa catcgctgac 1800
cgcggtgaag cagatgcaag cgacgacttt gaaggcacat atgaagaatg gcgtgaacat 1860
atgtggagtg acgtagcagc ctactttaac ctcgacattg aaaacagtga agataataaa 1920
tctactcttt cacttcaatt tgtcgacagc gccgcggata tgccgcttgc gaaaatgcac 1980
gggtgcgtttt caacgaacgt cgtagcaagc aaagaacttc aacagccagg cagtgcacga 2040
agcacgcgac atcttgaaat tgaacttcca aaagaagctt cttatcaaga aggagatcat 2100
ttaggtgtta ttccctgcaa ctatgaagga atagtaaacc gtgtaacagc aaggttcggc 2160
ctagatgcat cacagcaaat ccgtctggaa gcagaagaag aaaaattagc tcatttgcca 2220
ctcgctaaaa cagtatccgt agaagagctt ctgcaatacg tggagcttca agatcctgtt 2280
acgcgcacgc agcttcgcgc aatggctgct aaaacggtct gcccgccgca taaagtagag 2340
cttgaagcct tgcttgaaaa gcaagcctac aaagaacaag tgctggcaaa acgtttaaca 2400
atgcttgaaac tgcttgaaaa ataccggcg tgtgaaatga aattcagcga atttatcgcc 2460
cttctgccaa gcatacgccc gcgctattac tcgatttctt catcacctcg tgtcgatgaa 2520
aaacaagcaa gcatcacggt cagcgttgct tcaggagaag cgtggagcgg atatggagaa 2580
tataaaggaa ttgcgtcgaa ctatcttgcc gagctgcaag aaggagatac gattacgtgc 2640
tttattttcca caccgcagtc agaatttacg ctgccaaaag accctgaaac gccgcttattc 2700
atggctcggac cgggaacagg cgtcgcgccg tttagaggct ttgtgcaggc gcgcaaacag 2760
ctaaaagaac aaggacagtc acttgagaa gcacatttat acttcggctg ccgttcacct 2820
catgaagact atctgtatca agaagagctt gaaaacgccc aaagcgaagg catcattacg 2880
cttcataccg ctttttctcg catgccaat cagccgaaaa catacgttca gcacgtaatg 2940
gaacaagacg gcaagaaatt gattgaactt cttgatcaag gagcgcactt ctatatttgc 3000
ggagacggaa gccaaatggc acctgccgtt gaagcaacgc ttatgaaaag ctatgctgac 3060
gttcaccaag tgagtgaagc agacgctgc ttatggctgc agcagctaga agaaaaaggc 3120
cgatacgcaa aagacgtgtg ggctgggtaa 3150

```

<210> 6

<211> 1049

<212> PRT

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 6

```

Met Thr Ile Lys Glu Met Pro Gln Pro Lys Thr Phe Gly Glu Leu Lys
1          5          10          15
Asn Leu Pro Leu Leu Asn Thr Asp Lys Pro Val Gln Ala Leu Met Lys
20        25        30
Ile Ala Asp Glu Leu Gly Glu Ile Phe Lys Phe Glu Ala Pro Gly Cys
35        40        45
Val Thr Arg Tyr Leu Ser Ser Gln Arg Leu Ile Lys Glu Ala Cys Asp
50        55        60
Glu Ser Arg Phe Asp Lys Asn Leu Ser Gln Ala Leu Lys Phe Ala Arg
65        70        75        80
Asp Phe Leu Gly Asp Gly Leu Phe Thr Ser Trp Thr His Glu Ile Asn
85        90        95
Trp Lys Lys Ala His Asn Ile Leu Leu Pro Ser Phe Ser Gln Gln Ala
100       105       110
Met Lys Gly Tyr His Ala Met Met Val Asp Ile Ala Val Gln Leu Val
115       120       125
Gln Lys Trp Glu Arg Leu Asn Ala Asp Glu His Ile Glu Val Ser Glu

```

hp3238sgw-sp.sequence listing_ST25

	130					135					140				
Asp 145	Met	Thr	Arg	Leu	Thr	Leu	Asp	Thr	Ile	Gly	Leu	Cys	Gly	Phe	Asn 160
Tyr	Arg	Phe	Asn	Ser 165	Phe	Tyr	Arg	Asp	Gln 170	Pro	His	Pro	Phe	Ile	Ile
Ser	Met	Val	Arg 180	Ala	Leu	Asp	Glu	Val 185	Met	Asn	Lys	Leu	Gln	Arg	Ala
Asn	Pro	Asp 195	Asp	Pro	Ala	Tyr	Asp 200	Glu	Asn	Lys	Arg	Gln	Cys	Gln	Glu
Asp 210	Ile	Lys	Val	Met	Asn	Asp 215	Leu	Val	Asp	Lys	Ile	Ile	Ala	Asp	Arg
Lys 225	Ala	Arg	Gly	Glu	Gln	Ser	Asp	Asp	Leu	Leu	Thr	Gln	Met	Leu	Asn 240
Gly	Lys	Asp	Pro	Glu 245	Thr	Gly	Glu	Pro	Leu	Asp	Asp	Gly	Asn	Ile	Ser
Tyr	Gln	Ile	Ile 260	Thr	Phe	Leu	Ile	Ala 265	Gly	His	Glu	Thr	Thr	Ser	Gly
Leu	Leu	Ser 275	Phe	Ala	Leu	Tyr	Phe	Leu 280	Val	Lys	Asn	Pro	His	Val	Leu
Gln	Lys 290	Val	Ala	Glu	Glu	Ala 295	Ala	Arg	Val	Leu	Val	Asp	Pro	Val	Pro
Ser 305	Tyr	Lys	Gln	Val	Lys 310	Gln	Leu	Lys	Tyr	Val	Gly	Met	Val	Leu	Asn 320
Glu	Ala	Leu	Arg	Leu 325	Trp	Pro	Thr	Ala	Pro 330	Ala	Phe	Ser	Leu	Tyr	Ala
Lys	Glu	Asp	Thr 340	Val	Leu	Gly	Gly	Glu	Tyr 345	Pro	Leu	Glu	Lys	Gly	Asp
Glu	Val	Met 355	Val	Leu	Ile	Pro	Gln	Leu	His 360	Arg	Asp	Lys	Thr	Ile	Trp
Gly	Asp 370	Asp	Val	Glu	Glu	Phe	Arg	Pro	Glu	Arg	Phe	Glu	Asn	Pro	Ser
Ala 385	Ile	Pro	Gln	His	Ala 390	Phe	Lys	Pro	Phe	Gly	Asn	Gly	Gln	Arg	Ala
Cys	Ile	Gly	Gln	Gln 405	Phe	Ala	Leu	His	Glu	Ala	Thr	Leu	Val	Leu	Gly
Met	Met	Leu	Lys 420	His	Phe	Asp	Phe	Glu	Asp 425	His	Thr	Asn	Tyr	Glu	Leu
Asp	Ile	Lys 435	Glu	Thr	Leu	Thr	Leu	Lys	Pro	Glu	Gly	Phe	Val	Val	Lys
Ala	Lys 450	Ser	Lys	Lys	Ile	Pro	Leu	Gly	Gly	Ile	Pro	Ser	Pro	Ser	Thr
Glu 465	Gln	Ser	Ala	Lys	Lys 470	Val	Arg	Lys	Lys	Ala	Glu	Asn	Ala	His	Asn
Thr	Pro	Leu	Leu	Val 485	Leu	Tyr	Gly	Ser	Asn	Met	Gly	Thr	Ala	Glu	Gly
Thr	Ala	Arg	Asp 500	Leu	Ala	Asp	Ile	Ala	Met	Ser	Lys	Gly	Phe	Ala	Pro
Gln	Val	Ala 515	Thr	Leu	Asp	Ser	His	Ala	Gly	Asn	Leu	Pro	Arg	Glu	Gly
Ala	Val	Leu	Ile	Val	Thr	Ala 535	Ser	Tyr	Asn	Gly	His	Pro	Pro	Asp	Asn
Ala	Lys	Gln	Phe	Val	Asp	Trp	Leu	Asp	Gln	Ala	Ser	Ala	Asp	Glu	Val

hp3238sgw-sp.sequence listing_ST25

545					550					555				560
Lys	Gly	Val	Arg	Tyr	Ser	Val	Phe	Gly	Cys	Gly	Asp	Lys	Asn	Trp
				565					570					575
Thr	Thr	Tyr	Gln	Lys	Val	Pro	Ala	Phe	Ile	Asp	Glu	Thr	Leu	Ala
			580					585						590
Lys	Gly	Ala	Glu	Asn	Ile	Ala	Asp	Arg	Gly	Glu	Ala	Asp	Ala	Ser
		595					600					605		
Asp	Phe	Glu	Gly	Thr	Tyr	Glu	Glu	Trp	Arg	Glu	His	Met	Trp	Ser
	610					615					620			
Val	Ala	Ala	Tyr	Phe	Asn	Leu	Asp	Ile	Glu	Asn	Ser	Glu	Asp	Asn
625					630					635				640
Ser	Thr	Leu	Ser	Leu	Gln	Phe	Val	Asp	Ser	Ala	Ala	Asp	Met	Pro
				645						650				655
Ala	Lys	Met	His	Gly	Ala	Phe	Ser	Thr	Asn	Val	Val	Ala	Ser	Lys
			660					665						670
Leu	Gln	Gln	Pro	Gly	Ser	Ala	Arg	Ser	Thr	Arg	His	Leu	Glu	Ile
		675					680							685
Leu	Pro	Lys	Glu	Ala	Ser	Tyr	Gln	Glu	Gly	Asp	His	Leu	Gly	Val
	690					695					700			
Pro	Arg	Asn	Tyr	Glu	Gly	Ile	Val	Asn	Arg	Val	Thr	Ala	Arg	Phe
705					710					715				720
Leu	Asp	Ala	Ser	Gln	Gln	Ile	Arg	Leu	Glu	Ala	Glu	Glu	Glu	Lys
				725						730				735
Ala	His	Leu	Pro	Leu	Ala	Lys	Thr	Val	Ser	Val	Glu	Glu	Leu	Leu
			740					745						750
Tyr	Val	Glu	Leu	Gln	Asp	Pro	Val	Thr	Arg	Thr	Gln	Leu	Arg	Ala
		755					760							765
Ala	Ala	Lys	Thr	Val	Cys	Pro	Pro	His	Lys	Val	Glu	Leu	Glu	Ala
		770				775					780			
Leu	Glu	Lys	Gln	Ala	Tyr	Lys	Glu	Gln	Val	Leu	Ala	Lys	Arg	Leu
785					790					795				800
Met	Leu	Glu	Leu	Leu	Glu	Lys	Tyr	Pro	Ala	Cys	Glu	Met	Lys	Phe
				805					810					815
Glu	Phe	Ile	Ala	Leu	Leu	Pro	Ser	Ile	Arg	Pro	Arg	Tyr	Tyr	Ser
			820					825						830
Ser	Ser	Ser	Pro	Arg	Val	Asp	Glu	Lys	Gln	Ala	Ser	Ile	Thr	Val
		835					840					845		
Val	Val	Ser	Gly	Glu	Ala	Trp	Ser	Gly	Tyr	Gly	Glu	Tyr	Lys	Gly
	850					855					860			
Ala	Ser	Asn	Tyr	Leu	Ala	Glu	Leu	Gln	Glu	Gly	Asp	Thr	Ile	Thr
865					870					875				880
Phe	Ile	Ser	Thr	Pro	Gln	Ser	Glu	Phe	Thr	Leu	Pro	Lys	Asp	Pro
				885					890					895
Thr	Pro	Leu	Ile	Met	Val	Gly	Pro	Gly	Thr	Gly	Val	Ala	Pro	Phe
		900						905						910
Gly	Phe	Val	Gln	Ala	Arg	Lys	Gln	Leu	Lys	Glu	Gln	Gly	Gln	Ser
		915					920							925
Gly	Glu	Ala	His	Leu	Tyr	Phe	Gly	Cys	Arg	Ser	Pro	His	Glu	Asp
	930					935					940			
Leu	Tyr	Gln	Glu	Glu	Leu	Glu	Asn	Ala	Gln	Ser	Glu	Gly	Ile	Ile
945					950					955				960
Leu	His	Thr	Ala	Phe	Ser	Arg	Met	Pro	Asn	Gln	Pro	Lys	Thr	Tyr

hp3238sgw-sp.sequence listing_ST25

				965				970				975			
Gln	His	Val	Met	Glu	Gln	Asp	Gly	Lys	Lys	Leu	Ile	Glu	Leu	Leu	Asp
980				985				990							
Gln	Gly	Ala	His	Phe	Tyr	Ile	Cys	Gly	Asp	Gly	Ser	Gln	Met	Ala	Pro
995				1000				1005							
Ala	Val	Glu	Ala	Thr	Leu	Met	Lys	Ser	Tyr	Ala	Asp	Val	His	Gln	Val
1010				1015				1020							
Ser	Glu	Ala	Asp	Ala	Arg	Leu	Trp	Leu	Gln	Gln	Leu	Glu	Glu	Lys	Gly
1025				1030				1035				1040			
Arg	Tyr	Ala	Lys	Asp	Val	Trp	Ala	Gly							
1045															

<210> 7

<211> 3150

<212> DNA

<213> Artificial Sequence

 $\langle 220 \rangle$

<223> P450 BM3

<400> 7

atgacaataa	aggaaatgcc	gcagccgaaa	acgttttgag	aacttaaaaa	tttaccgtta	60
ttaaaccacag	ataaacccggt	tcaagctttg	atgaaaattg	cggatgaatt	aggagaaatc	120
tttaaattcg	aggcgcctgg	tcgtgtaacg	cgctacttat	caagtcagcg	tctaattaaa	180
gaagcatgcg	atgaatcacg	ctttgataaa	aacttaagtc	aagcgcttaa	atttgcacgt	240
gattttgcag	gagacgggtt	atttacaagc	tggacgcatg	aaaaaaattg	gaaaaaagcg	300
cataatatct	tacttccaag	ctttagtcag	caggcaatga	aaggctatca	tgcgatgatg	360
gtcgatatcg	ccgtgcagct	tgttcaaaaag	tgggagcgtc	taaatgcaga	tgagtatatatt	420
gaagtaccgg	aagacatgac	acgtttaacg	cttgatacaa	ttggctcttg	cggctttaac	480
tatcgcttta	acagctttta	ccgagatcag	cctcatccat	ttattataag	tatgatccgt	540
gcactggatg	aagtaatgaa	caagctgcag	cgagcaaatc	cagacgaccc	agcttatgat	600
gaaaacaagc	gccagtttca	agaagatatc	aaggtgatga	acgacctagt	agataaaaatt	660
attgcagatc	gcaaagcaag	cgggtgaacaa	agcgatgatt	tattaacgca	gatgctaaac	720
ggaaaagatc	cagaaacggg	tgagccgctt	gatgacggga	acattagcta	tcaaattatt	780
acatttcttaa	ttgcgggaca	cgaacaaca	agtggctctt	tatcatttgc	gctgtatttc	840
ttagtgaaaa	atccacatgt	attacaaaaa	gtagcagaag	aagcaacacg	agttctagta	900
gatcctgttc	caagctacaa	acaagtcaaa	cagcttaaatt	atgtcggcat	ggctcttaaac	960
gaagcgctgc	gcttatggcc	aactgctcct	gcgttttccc	tatatgcaa	agaagatacg	1020
gtgcttggag	gagaatatcc	tttagaaaaa	ggcgacgaag	taatggttct	gattcctcag	1080
cttcaccgtg	ataaaacaat	ttggggagac	gatgtggaag	agttccgtcc	agagcgtttt	1140
gaaaatccaa	gtgcgattcc	gcagcatgcg	tttaaaccgt	ttggaaacgg	tcaacgtgcg	1200
tgtatcggtc	agcagttcgc	tcttcatgaa	gcaacgctgg	tacttggtat	gatgctaaaa	1260
cactttgact	ttgaagatca	tacaaactac	gagctcgata	ttaaagaaac	tttaacgtta	1320
aaacctgaag	gctttgtggt	aaaagcaaaa	tcgaaaaaaa	ttccgcttgg	cggtattcct	1380
tcacctagca	ctgaacagtc	tgctaaaaaa	gtacgcaaaa	aggcagaaaa	cgctcataat	1440
acgccgctgc	ttgtgctata	cggttcfaat	atgggaacag	ctgaagggaac	ggcgcgtgat	1500
ttagcagata	ttgcaatgag	caaaggattt	gcaccgcagg	tcgcaacgct	tgattcacac	1560
gccggaaatc	ttccgcgcga	aggagctgta	ttaattgtaa	cggcgtctta	taacggtcac	1620
ccgcctgata	acgcaaagca	atttgtcgac	tggttagacc	aagcgtctgc	tgatgaagta	1680
aaaggcgttc	gctactccgt	atttggatgc	ggcgataaaa	actgggctac	tacgtatcaa	1740
aaagtgcctg	cttttatcga	tgaaacgctt	gccgctaaag	gggcagaaaa	catcgctgac	1800

hp3238sgw-sp.sequence listing_ST25

```

cgcggtgaag cagatgcaag cgacgacttt gaaggcacat atgaagaatg gcgtgaacat 1860
atgtggagtg acgtagcagc ctactttaac ctcgacattg aaaacagtga agataataaa 1920
tctactcttt cacttcaatt tgtcgacagc gccgcggata tgccgcttgc gaaaatgcac 1980
gggtgcgtttt caacgaacgt cgtagcaagc aaagaacttc aacagccagg cagtgcacga 2040
agcacgcgac atcttgaaat tgaacttcca aaagaagctt cttatcaaga aggagatcat 2100
ttaggtgtta ttcctcgcaa ctatgaagga atagtaaacc gtgtaacagc aaggttcggc 2160
ctagatgcat cacagcaaat ccgtctggaa gcagaagaag aaaaattagc tcatttgcca 2220
ctcgctaaaa cagtatccgt agaagagctt ctgcaatacg tggagcttca agatcctgtt 2280
acgcgcacgc agcttcgcgc aatggctgct aaaacggtct gcccgccgca taaagtagag 2340
cttgaagcct tgcttgaaaa gcaagcctac aaagaacaag tgctggcaaa acgtttaaca 2400
atgcttgaac tgcttgaaaa atacccggcg tgtgaaatga aattcagcga atttatcgcc 2460
cttctgccaa gcatacgccc gcgctattac tcgatttctt catcacctcg tgtcgatgaa 2520
aaacaagcaa gcatcacggt cagcgttgtc tcaggagaag cgtggagcgg atatggagaa 2580
tataaaggaa ttgcgtcgaa ctatcttgcc gagctgcaag aaggagatac gattacgtgc 2640
tttattttcca caccgcagtc agaatttacg ctgccaaaag accctgaaac gccgcttattc 2700
atggctcggac cgggaacagg cgtcgcgccg tttagaggct ttgtgcaggc gcgcaaacag 2760
ctaaaagaac aaggacagtc acttgagaa gcacatttat acttcggctg ccgttcacct 2820
catgaagact atctgtatca agaagagctt gaaaacgccc aaagcgaagg catcattacg 2880
cttcataccg ctttttctcg catgccaaat cagccgaaaa catacgttca gcacgtaatg 2940
gaacaagacg gcaagaaatt gattgaactt cttgatcaag gagcgcactt ctatatttgc 3000
ggagacggaa gccaaatggc acctgccgtt gaagcaacgc ttatgaaaag ctatgctgac 3060
gttcaccaag tgagtgaagc agacgctcgc ttatggctgc agcagctaga agaaaaaggc 3120
cgatacgcaa aagacgtgtg ggctgggtaa 3150

```

<210> 8

<211> 1049

<212> PRT

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 8

```

Met Thr Ile Lys Glu Met Pro Gln Pro Lys Thr Phe Gly Glu Leu Lys
1          5          10          15
Asn Leu Pro Leu Leu Asn Thr Asp Lys Pro Val Gln Ala Leu Met Lys
          20          25          30
Ile Ala Asp Glu Leu Gly Glu Ile Phe Lys Phe Glu Ala Pro Gly Arg
          35          40          45
Val Thr Arg Tyr Leu Ser Ser Gln Arg Leu Ile Lys Glu Ala Cys Asp
          50          55          60
Glu Ser Arg Phe Asp Lys Asn Leu Ser Gln Ala Leu Lys Phe Ala Arg
65          70          75          80
Asp Phe Ala Gly Asp Gly Leu Phe Thr Ser Trp Thr His Glu Lys Asn
          85          90          95
Trp Lys Lys Ala His Asn Ile Leu Leu Pro Ser Phe Ser Gln Gln Ala
          100          105          110
Met Lys Gly Tyr His Ala Met Met Val Asp Ile Ala Val Gln Leu Val
          115          120          125
Gln Lys Trp Glu Arg Leu Asn Ala Asp Glu Tyr Ile Glu Val Pro Glu
          130          135          140

```

hp3238sgw-sp.sequence listing_ST25

Asp	Met	Thr	Arg	Leu	Thr	Leu	Asp	Thr	Ile	Gly	Leu	Cys	Gly	Phe	Asn
145					150					155					160
Tyr	Arg	Phe	Asn	Ser	Phe	Tyr	Arg	Asp	Gln	Pro	His	Pro	Phe	Ile	Ile
				165					170					175	
Ser	Met	Ile	Arg	Ala	Leu	Asp	Glu	Val	Met	Asn	Lys	Leu	Gln	Arg	Ala
				180				185					190		
Asn	Pro	Asp	Asp	Pro	Ala	Tyr	Asp	Glu	Asn	Lys	Arg	Gln	Phe	Gln	Glu
		195					200					205			
Asp	Ile	Lys	Val	Met	Asn	Asp	Leu	Val	Asp	Lys	Ile	Ile	Ala	Asp	Arg
	210					215					220				
Lys	Ala	Ser	Gly	Glu	Gln	Ser	Asp	Asp	Leu	Leu	Thr	Gln	Met	Leu	Asn
225					230					235					240
Gly	Lys	Asp	Pro	Glu	Thr	Gly	Glu	Pro	Leu	Asp	Asp	Gly	Asn	Ile	Ser
				245					250					255	
Tyr	Gln	Ile	Ile	Thr	Phe	Leu	Ile	Ala	Gly	His	Glu	Thr	Thr	Ser	Gly
			260					265						270	
Leu	Leu	Ser	Phe	Ala	Leu	Tyr	Phe	Leu	Val	Lys	Asn	Pro	His	Val	Leu
		275					280					285			
Gln	Lys	Val	Ala	Glu	Glu	Ala	Thr	Arg	Val	Leu	Val	Asp	Pro	Val	Pro
	290					295					300				
Ser	Tyr	Lys	Gln	Val	Lys	Gln	Leu	Lys	Tyr	Val	Gly	Met	Val	Leu	Asn
305					310					315					320
Glu	Ala	Leu	Arg	Leu	Trp	Pro	Thr	Ala	Pro	Ala	Phe	Ser	Leu	Tyr	Ala
				325				330						335	
Lys	Glu	Asp	Thr	Val	Leu	Gly	Gly	Glu	Tyr	Pro	Leu	Glu	Lys	Gly	Asp
			340					345					350		
Glu	Val	Met	Val	Leu	Ile	Pro	Gln	Leu	His	Arg	Asp	Lys	Thr	Ile	Trp
		355					360					365			
Gly	Asp	Asp	Val	Glu	Glu	Phe	Arg	Pro	Glu	Arg	Phe	Glu	Asn	Pro	Ser
	370					375					380				
Ala	Ile	Pro	Gln	His	Ala	Phe	Lys	Pro	Phe	Gly	Asn	Gly	Gln	Arg	Ala
385					390					395					400
Cys	Ile	Gly	Gln	Gln	Phe	Ala	Leu	His	Glu	Ala	Thr	Leu	Val	Leu	Gly
			405						410					415	
Met	Met	Leu	Lys	His	Phe	Asp	Phe	Glu	Asp	His	Thr	Asn	Tyr	Glu	Leu
		420						425					430		
Asp	Ile	Lys	Glu	Thr	Leu	Thr	Leu	Lys	Pro	Glu	Gly	Phe	Val	Val	Lys
		435					440					445			
Ala	Lys	Ser	Lys	Lys	Ile	Pro	Leu	Gly	Gly	Ile	Pro	Ser	Pro	Ser	Thr
	450					455					460				
Glu	Gln	Ser	Ala	Lys	Lys	Val	Arg	Lys	Lys	Ala	Glu	Asn	Ala	His	Asn
465					470					475					480
Thr	Pro	Leu	Leu	Val	Leu	Tyr	Gly	Ser	Asn	Met	Gly	Thr	Ala	Glu	Gly
				485					490					495	
Thr	Ala	Arg	Asp	Leu	Ala	Asp	Ile	Ala	Met	Ser	Lys	Gly	Phe	Ala	Pro
			500					505					510		
Gln	Val	Ala	Thr	Leu	Asp	Ser	His	Ala	Gly	Asn	Leu	Pro	Arg	Glu	Gly
		515					520					525			
Ala	Val	Leu	Ile	Val	Thr	Ala	Ser	Tyr	Asn	Gly	His	Pro	Pro	Asp	Asn
	530					535					540				
Ala	Lys	Gln	Phe	Val	Asp	Trp	Leu	Asp	Gln	Ala	Ser	Ala	Asp	Glu	Val
545					550				555						560

hp3238sgw-sp.sequence listing_ST25

Lys	Gly	Val	Arg	Tyr	Ser	Val	Phe	Gly	Cys	Gly	Asp	Lys	Asn	Trp	Ala		
				565					570					575			
Thr	Thr	Tyr	Gln	Lys	Val	Pro	Ala	Phe	Ile	Asp	Glu	Thr	Leu	Ala	Ala		
			580					585					590				
Lys	Gly	Ala	Glu	Asn	Ile	Ala	Asp	Arg	Gly	Glu	Ala	Asp	Ala	Ser	Asp		
		595					600					605					
Asp	Phe	Glu	Gly	Thr	Tyr	Glu	Glu	Trp	Arg	Glu	His	Met	Trp	Ser	Asp		
	610				615					620							
Val	Ala	Ala	Tyr	Phe	Asn	Leu	Asp	Ile	Glu	Asn	Ser	Glu	Asp	Asn	Lys		
625					630					635					640		
Ser	Thr	Leu	Ser	Leu	Gln	Phe	Val	Asp	Ser	Ala	Ala	Asp	Met	Pro	Leu		
			645					650						655			
Ala	Lys	Met	His	Gly	Ala	Phe	Ser	Thr	Asn	Val	Val	Ala	Ser	Lys	Glu		
			660					665						670			
Leu	Gln	Gln	Pro	Gly	Ser	Ala	Arg	Ser	Thr	Arg	His	Leu	Glu	Ile	Glu		
		675					680					685					
Leu	Pro	Lys	Glu	Ala	Ser	Tyr	Gln	Glu	Gly	Asp	His	Leu	Gly	Val	Ile		
	690					695				700							
Pro	Arg	Asn	Tyr	Glu	Gly	Ile	Val	Asn	Arg	Val	Thr	Ala	Arg	Phe	Gly		
705					710					715					720		
Leu	Asp	Ala	Ser	Gln	Ile	Arg	Leu	Glu	Ala	Glu	Glu	Glu	Lys	Leu			
			725					730						735			
Ala	His	Leu	Pro	Leu	Ala	Lys	Thr	Val	Ser	Val	Glu	Glu	Leu	Leu	Gln		
			740					745						750			
Tyr	Val	Glu	Leu	Gln	Asp	Pro	Val	Thr	Arg	Thr	Gln	Leu	Arg	Ala	Met		
		755				760						765					
Ala	Ala	Lys	Thr	Val	Cys	Pro	Pro	His	Lys	Val	Glu	Leu	Glu	Ala	Leu		
		770				775					780						
Leu	Glu	Lys	Gln	Ala	Tyr	Lys	Glu	Gln	Val	Leu	Ala	Lys	Arg	Leu	Thr		
785					790					795					800		
Met	Leu	Glu	Leu	Leu	Glu	Lys	Tyr	Pro	Ala	Cys	Glu	Met	Lys	Phe	Ser		
			805					810						815			
Glu	Phe	Ile	Ala	Leu	Leu	Pro	Ser	Ile	Arg	Pro	Arg	Tyr	Tyr	Ser	Ile		
			820					825						830			
Ser	Ser	Ser	Pro	Arg	Val	Asp	Glu	Lys	Gln	Ala	Ser	Ile	Thr	Val	Ser		
		835					840					845					
Val	Val	Ser	Gly	Glu	Ala	Trp	Ser	Gly	Tyr	Gly	Glu	Tyr	Lys	Gly	Ile		
	850					855					860						
Ala	Ser	Asn	Tyr	Leu	Ala	Glu	Leu	Gln	Glu	Gly	Asp	Thr	Ile	Thr	Cys		
865					870					875					880		
Phe	Ile	Ser	Thr	Pro	Gln	Ser	Glu	Phe	Thr	Leu	Pro	Lys	Asp	Pro	Glu		
			885					890						895			
Thr	Pro	Leu	Ile	Met	Val	Gly	Pro	Gly	Thr	Gly	Val	Ala	Pro	Phe	Arg		
			900					905						910			
Gly	Phe	Val	Gln	Ala	Arg	Lys	Gln	Leu	Lys	Glu	Gln	Gly	Gln	Ser	Leu		
		915					920							925			
Gly	Glu	Ala	His	Leu	Tyr	Phe	Gly	Cys	Arg	Ser	Pro	His	Glu	Asp	Tyr		
		930				935					940						
Leu	Tyr	Gln	Glu	Glu	Leu	Glu	Asn	Ala	Gln	Ser	Glu	Gly	Ile	Ile	Thr		
945					950					955					960		
Leu	His	Thr	Ala	Phe	Ser	Arg	Met	Pro	Asn	Gln	Pro	Lys	Thr	Tyr	Val		
			965					970						975			

hp3238sgw-sp.sequence listing_ST25

Gln His Val Met Glu Gln Asp Gly Lys Lys Leu Ile Glu Leu Leu Asp
980 985 990
Gln Gly Ala His Phe Tyr Ile Cys Gly Asp Gly Ser Gln Met Ala Pro
995 1000 1005
Ala Val Glu Ala Thr Leu Met Lys Ser Tyr Ala Asp Val His Gln Val
1010 1015 1020
Ser Glu Ala Asp Ala Arg Leu Trp Leu Gln Gln Leu Glu Glu Lys Gly
1025 1030 1035 1040
Arg Tyr Ala Lys Asp Val Trp Ala Gly
1045

<210> 9

<211> 3150

<212> DNA

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 9

```

atgacaatta aagaaatgcc tcagccaaaa acgtttggag agcttaaaaa tttaccgtta 60
ttaaacacag ataaaccggt tcaagctttg atgaaaattg cggatgaatt aggagaaatc 120
tttaaattcg aggcgcctgg ttgtgtaacg cgctacttat caagtcagcg tctaattaaa 180
gaagcatgcg atgaatcacg ctttgataaa aacttaagtc aagcgcttaa atttgcacgt 240
gattttgacg gagacgggtt agcaacaagc tggacgcatg aaataaattg gaaaaaagcg 300
cataatatct tacttccaag ctttagtcag caggcaatga aaggctatca tgcgatgatg 360
gtcgatatcg ccgtgcagct tgttcaaaaag tgggagcgtc taaatgcaga tgagcatatt 420
gaagtatcgg aagacatgac acgtttaacg cttgatacaa ttggtctttg cggctttaac 480
tatcgcttta acagctttta ccgagatcag cctcatccat ttattataag tatggtccgt 540
gcaactggatg aagtaatgaa caagctgcag cgagcaaadc cagacgacc agcttatgat 600
gaaaacaagc gccagtgtca agaagatadc aaggtgatga acgacctagt agataaaatt 660
attgcagatc gcaaagcaag ggggtgaacaa agcgatgatt tattaacgca gatgctaaac 720
ggaaaagatc cagaaacggg tgagccgctt gatgacggga acattagcta tcaaatattt 780
acattcttaa ttgcgggaca cgaaacaaca agtggctctt tatcatttgc gctgtatttc 840
ttagtgaata atccacatgt attacaaaaa gtagcagaag aagcagcacg agttctagta 900
gatcctgttc caagctacaa acaagtcaaa cagcttaaat atgtcggcat ggtcttaaac 960
gaagcgctgc gcttatggcc aactgctcct gcgttttccc tatatgcaaa agaagatacg 1020
gtgcttggag gagaatatcc tttagaaaaa ggcgacgaag taatggttct gattcctcag 1080
cttcaccgtg ataaaacaat ttggggagac gatgtggagg agttccgtcc agagcgtttt 1140
gaaaatccaa gtgcgattcc gcagcatgcg tttaaaccgt ttggaaacgg tcagcggtgcg 1200
tgtatcggtc agcagttcgc tcttcatgaa gcaacgctgg tacttggtat gatgctaaaa 1260
cactttgact ttgaagatca tacaactac gagctcgata ttaaagaaac tttaacgtta 1320
aaacctgaag gctttgtggt aaaagcaaaa tcgaaaaaaa ttccgcttgg cggatttcct 1380
tcacctagca ctgaacagtc tgctaaaaaa gtacgcaaaa aggcagaaaa cgctcataat 1440
acgccgctgc ttgtgctata cggttcaaat atgggaacag ctgaaggaac ggcgcggtgat 1500
ttagcagata ttgcaatgag caaaggattt gcaccgcagg tcgcaacgct tgattcacac 1560
gccggaaatc ttccgcgcga aggagctgta ttaattgtaa cggcgctcta taacggtcat 1620
ccgcctgata acgcaaagca atttgtcgac tggttagacc aagcgtctgc tgatgaagta 1680
aaaggcggtc gctactccgt atttggatgc ggcgataaaa actgggctac tacgtatcaa 1740
aaagtgcctg cttttatcga tgaaacgctt gccgctaaag gggcagaaaa catcgctgac 1800
cgcggtgaag cagatgcaag cgacgacttt gaaggcacat atgaagaatg gcgtgaacat 1860

```

hp3238sgw-sp.sequence listing_ST25

```

atgtggagtg acgtagcagc ctactttaac ctcgacattg aaaacagtga agataataaa 1920
tctactcttt cacttcaatt tgtcgacagc gccgcggata tgccgcttgc gaaaatgcac 1980
gggtgcgtttt caacgaacgt cgtagcaagc aaagaacttc aacagccagg cagtgcacga 2040
agcacgcgac atcttgaaat tgaacttcca aaagaagctt cttatcaaga aggagatcat 2100
ttaggtgtta ttcttcgcaa ctatgaagga atagtaaacc gtgtaacagc aaggttcggc 2160
ctagatgcat cacagcaaat ccgtctggaa gcagaagaag aaaaattagc tcatttgcca 2220
ctcgctaaaa cagtatccgt agaagagctt ctgcaatacg tggagcttca agatcctgtt 2280
acgcgcacgc agcttcgcgc aatggctgct aaaacggtct gcccgccgca taaagtagag 2340
cttgaagcct tgcttgaaaa gcaagcctac aaagaacaag tgctggcaaa acgtttaaca 2400
atgcttgaac tgcttgaaaa atacccggcg tgtgaaatga aattcagcga atttatcgcc 2460
cttctgccaa gcatacgccc gcgctattac tcgatttctt catcacctcg tgtcgatgaa 2520
aaacaagcaa gcatcacggt cagcgttgtc tcaggagaag cgtggagcgg atatggagaa 2580
tataaaggaa ttgcgtcgaa ctatcttgcc gagctgcaag aaggagatac gattacgtgc 2640
tttattttcca caccgcagtc agaatttacg ctgccaaaag accctgaaac gccgcttattc 2700
atggtcggac cgggaacagg cgtcgcgccg tttagaggct ttgtgcaggc gcgcaaacag 2760
ctaaaagaac aaggacagtc acttgagaa gcacatttat acttcggctg ccgttcacct 2820
catgaagact atctgtatca agaagagctt gaaaacgccc aaagcgaagg catcattacg 2880
cttcataccg ctttttctcg catgccaaat cagccgaaaa catacgttca gcacgtaatg 2940
gaacaagacg gcaagaaatt gattgaactt cttgatcaag gagcgcactt ctatatattgc 3000
ggagacggaa gccaaatggc acctgccgtt gaagcaacgc ttatgaaaag ctatgctgac 3060
gttcaccaag tgagtgaagc agacgctcgc ttatggctgc agcagctaga agaaaaaggc 3120
cgatacgcaa aagacgtgtg ggctgggtaa 3150

```

<210> 10

<211> 1049

<212> PRT

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 10

```

Met Thr Ile Lys Glu Met Pro Gln Pro Lys Thr Phe Gly Glu Leu Lys
1          5          10          15
Asn Leu Pro Leu Leu Asn Thr Asp Lys Pro Val Gln Ala Leu Met Lys
20        25        30
Ile Ala Asp Glu Leu Gly Glu Ile Phe Lys Phe Glu Ala Pro Gly Cys
35        40        45
Val Thr Arg Tyr Leu Ser Ser Gln Arg Leu Ile Lys Glu Ala Cys Asp
50        55        60
Glu Ser Arg Phe Asp Lys Asn Leu Ser Gln Ala Leu Lys Phe Ala Arg
65        70        75        80
Asp Phe Ala Gly Asp Gly Leu Ala Thr Ser Trp Thr His Glu Ile Asn
85        90        95
Trp Lys Lys Ala His Asn Ile Leu Leu Pro Ser Phe Ser Gln Gln Ala
100       105       110
Met Lys Gly Tyr His Ala Met Met Val Asp Ile Ala Val Gln Leu Val
115       120       125
Gln Lys Trp Glu Arg Leu Asn Ala Asp Glu His Ile Glu Val Ser Glu
130       135       140
Asp Met Thr Arg Leu Thr Leu Asp Thr Ile Gly Leu Cys Gly Phe Asn

```

hp3238sgw-sp.sequence listing_ST25

145					150					155				160
Tyr	Arg	Phe	Asn	Ser	Phe	Tyr	Arg	Asp	Gln	Pro	His	Pro	Phe	Ile
				165					170					175
Ser	Met	Val	Arg	Ala	Leu	Asp	Glu	Val	Met	Asn	Lys	Leu	Gln	Arg
			180					185					190	
Asn	Pro	Asp	Asp	Pro	Ala	Tyr	Asp	Glu	Asn	Lys	Arg	Gln	Cys	Gln
		195					200					205		
Asp	Ile	Lys	Val	Met	Asn	Asp	Leu	Val	Asp	Lys	Ile	Ile	Ala	Asp
	210					215					220			
Lys	Ala	Arg	Gly	Glu	Gln	Ser	Asp	Asp	Leu	Leu	Thr	Gln	Met	Leu
225					230					235				240
Gly	Lys	Asp	Pro	Glu	Thr	Gly	Glu	Pro	Leu	Asp	Asp	Gly	Asn	Ile
				245					250					255
Tyr	Gln	Ile	Ile	Thr	Phe	Leu	Ile	Ala	Gly	His	Glu	Thr	Thr	Ser
			260					265						270
Leu	Leu	Ser	Phe	Ala	Leu	Tyr	Phe	Leu	Val	Lys	Asn	Pro	His	Val
		275					280					285		
Gln	Lys	Val	Ala	Glu	Glu	Ala	Ala	Arg	Val	Leu	Val	Asp	Pro	Val
	290					295					300			
Ser	Tyr	Lys	Gln	Val	Lys	Gln	Leu	Lys	Tyr	Val	Gly	Met	Val	Leu
305					310					315				320
Glu	Ala	Leu	Arg	Leu	Trp	Pro	Thr	Ala	Pro	Ala	Phe	Ser	Leu	Tyr
				325					330					335
Lys	Glu	Asp	Thr	Val	Leu	Gly	Gly	Glu	Tyr	Pro	Leu	Glu	Lys	Gly
			340					345					350	
Glu	Val	Met	Val	Leu	Ile	Pro	Gln	Leu	His	Arg	Asp	Lys	Thr	Ile
		355					360					365		
Gly	Asp	Asp	Val	Glu	Glu	Phe	Arg	Pro	Glu	Arg	Phe	Glu	Asn	Pro
	370					375					380			
Ala	Ile	Pro	Gln	His	Ala	Phe	Lys	Pro	Phe	Gly	Asn	Gly	Gln	Arg
385					390					395				400
Cys	Ile	Gly	Gln	Gln	Phe	Ala	Leu	His	Glu	Ala	Thr	Leu	Val	Leu
			405						410					415
Met	Met	Leu	Lys	His	Phe	Asp	Phe	Glu	Asp	His	Thr	Asn	Tyr	Glu
			420					425					430	
Asp	Ile	Lys	Glu	Thr	Leu	Thr	Leu	Lys	Pro	Glu	Gly	Phe	Val	Val
		435					440					445		
Ala	Lys	Ser	Lys	Lys	Ile	Pro	Leu	Gly	Gly	Ile	Pro	Ser	Pro	Ser
	450					455				460				
Glu	Gln	Ser	Ala	Lys	Lys	Val	Arg	Lys	Lys	Ala	Glu	Asn	Ala	His
465					470					475				480
Thr	Pro	Leu	Leu	Val	Leu	Tyr	Gly	Ser	Asn	Met	Gly	Thr	Ala	Glu
				485					490					495
Thr	Ala	Arg	Asp	Leu	Ala	Asp	Ile	Ala	Met	Ser	Lys	Gly	Phe	Ala
			500					505					510	
Gln	Val	Ala	Thr	Leu	Asp	Ser	His	Ala	Gly	Asn	Leu	Pro	Arg	Glu
		515					520					525		
Ala	Val	Leu	Ile	Val	Thr	Ala	Ser	Tyr	Asn	Gly	His	Pro	Pro	Asp
	530					535					540			
Ala	Lys	Gln	Phe	Val	Asp	Trp	Leu	Asp	Gln	Ala	Ser	Ala	Asp	Glu
545					550					555				560
Lys	Gly	Val	Arg	Tyr	Ser	Val	Phe	Gly	Cys	Gly	Asp	Lys	Asn	Trp

hp3238sgw-sp.sequence listing ST25

Thr	Thr	Tyr	Gln	Lys	Val	Pro	Ala	Phe	Ile	Asp	Glu	Thr	Leu	Ala	Ala
			580					585					590		
Lys	Gly	Ala	Glu	Asn	Ile	Ala	Asp	Arg	Gly	Glu	Ala	Asp	Ala	Ser	Asp
		595					600					605			
Asp	Phe	Glu	Gly	Thr	Tyr	Glu	Glu	Trp	Arg	Glu	His	Met	Trp	Ser	Asp
	610					615					620				
Val	Ala	Ala	Tyr	Phe	Asn	Leu	Asp	Ile	Glu	Asn	Ser	Glu	Asp	Asn	Lys
625					630					635					640
Ser	Thr	Leu	Ser	Leu	Gln	Phe	Val	Asp	Ser	Ala	Ala	Asp	Met	Pro	Leu
				645					650					655	
Ala	Lys	Met	His	Gly	Ala	Phe	Ser	Thr	Asn	Val	Val	Ala	Ser	Lys	Glu
			660					665					670		
Leu	Gln	Gln	Pro	Gly	Ser	Ala	Arg	Ser	Thr	Arg	His	Leu	Glu	Ile	Glu
		675					680					685			
Leu	Pro	Lys	Glu	Ala	Ser	Tyr	Gln	Glu	Gly	Asp	His	Leu	Gly	Val	Ile
	690					695					700				
Pro	Arg	Asn	Tyr	Glu	Gly	Ile	Val	Asn	Arg	Val	Thr	Ala	Arg	Phe	Gly
705					710					715					720
Leu	Asp	Ala	Ser	Gln	Gln	Ile	Arg	Leu	Glu	Ala	Glu	Glu	Glu	Lys	Leu
				725					730					735	
Ala	His	Leu	Pro	Leu	Ala	Lys	Thr	Val	Ser	Val	Glu	Glu	Leu	Leu	Gln
			740					745					750		
Tyr	Val	Glu	Leu	Gln	Asp	Pro	Val	Thr	Arg	Thr	Gln	Leu	Arg	Ala	Met
		755				760					765				
Ala	Ala	Lys	Thr	Val	Cys	Pro	Pro	His	Lys	Val	Glu	Leu	Glu	Ala	Leu
		770				775					780				
Leu	Glu	Lys	Gln	Ala	Tyr	Lys	Glu	Gln	Val	Leu	Ala	Lys	Arg	Leu	Thr
785					790					795					800
Met	Leu	Glu	Leu	Leu	Glu	Lys	Tyr	Pro	Ala	Cys	Glu	Met	Lys	Phe	Ser
				805					810					815	
Glu	Phe	Ile	Ala	Leu	Leu	Pro	Ser	Ile	Arg	Pro	Arg	Tyr	Tyr	Ser	Ile
			820					825					830		
Ser	Ser	Ser	Pro	Arg	Val	Asp	Glu	Lys	Gln	Ala	Ser	Ile	Thr	Val	Ser
		835				840						845			
Val	Val	Ser	Gly	Glu	Ala	Trp	Ser	Gly	Tyr	Gly	Glu	Tyr	Lys	Gly	Ile
	850					855					860				
Ala	Ser	Asn	Tyr	Leu	Ala	Glu	Leu	Gln	Glu	Gly	Asp	Thr	Ile	Thr	Cys
865					870					875					880
Phe	Ile	Ser	Thr	Pro	Gln	Ser	Glu	Phe	Thr	Leu	Pro	Lys	Asp	Pro	Glu
				885					890					895	
Thr	Pro	Leu	Ile	Met	Val	Gly	Pro	Gly	Thr	Gly	Val	Ala	Pro	Phe	Arg
			900					905					910		
Gly	Phe	Val	Gln	Ala	Arg	Lys	Gln	Leu	Lys	Glu	Gln	Gly	Gln	Ser	Leu
		915													

hp3238sgw-sp.sequence listing ST25

980										985					990				
Gln	Gly	Ala	His	Phe	Tyr	Ile	Cys	Gly	Asp	Gly	Ser	Gln	Met	Ala	Pro				
995							1000					1005							
Ala	Val	Glu	Ala	Thr	Leu	Met	Lys	Ser	Tyr	Ala	Asp	Val	His	Gln	Val				
1010				1015					1020										
Ser	Glu	Ala	Asp	Ala	Arg	Leu	Trp	Leu	Gln	Gln	Leu	Glu	Glu	Lys	Gly				
1025					1030				1035						1040				
Arg	Tyr	Ala	Lys	Asp	Val	Trp	Ala	Gly											
1045																			

<210> 11
<211> 3150
<212> DNA
<213> Artificial Sequence

<220>
<223> P450 BM3

<400>	11					
atgacaataa	aggaaatgcc	gcagccgaaa	acgttttgag	aacttaaaaa	tttaccgtta	60
ttaaacacag	ataaaccggt	tcaagctttg	atgaaaattg	cggatgaatt	aggagaaatc	120
tttaaattcg	aggcgcctgg	ttgtgtaacg	cgctacttat	caagtcagcg	tctaattaaa	180
gaagcatgcg	atgaatcacg	ctttgataaa	aacttaagtc	aagcgcttaa	attttcccgt	240
gattttgcag	gagacgggtt	atttacaagc	tggacgcatg	aaataaattg	gaaaaaagcg	300
cataatatct	tacttccaag	ctttagtcag	caggcaatga	aaggctatca	tgcgatgatg	360
gtcgatatcg	ccgtgcagct	tgttcaaaaag	tgggagcgtc	taaatgcaga	tgagcatatt	420
gaagtatcgg	aagacatgac	acgtttaacg	cttgatacaa	ttggctcttg	cggctttaac	480
tatcgcttta	acagctttta	ccgagatcag	cctcatccat	ttattataag	tatgggtccgt	540
gcactggatg	aagtaatgaa	caagctgcag	cgagcaaadc	cagacgaccc	agcttatgat	600
gaaaacaagc	gccagtgtca	agaagatatc	aaggatgatg	acgacctagt	agataaaaatt	660
attgcagatc	gcaaagcaag	gggtgaacaa	agcgatgatt	tattaacgca	gatgctaaac	720
ggaaaagatc	cagaaacggg	tgagccgctt	gatgacggga	acattagcta	tcaaatattt	780
acattcttaa	ttgcgggaca	cgaacaaca	agtggctctt	tatcatttgc	gctgtatttc	840
ttagtgaaaa	atccacatgt	attacaaaaa	gtagcagaag	aagcagcacg	agttctagta	900
gatcctgttc	caagctacaa	acaagtcaaa	cagcttaaat	atgtcggcat	ggtcttaaac	960
gaagcgctgc	gcttatggcc	aactgctcct	gcgttttccc	tatatgcaa	agaagatacg	1020
gtgcttgagg	gagaatatcc	tttagaaaaa	ggcgacgaag	taatggttct	gattcctcag	1080
cttcaccgtg	ataaaacaat	ttggggagac	gatgtggagg	agttccgtcc	agagcgtttt	1140
gaaaatccaa	gtgcgattcc	gcagcatgcg	tttaaaccgt	ttggaaacgg	tcagcgtgcg	1200
tgtatcggtc	agcagttcgc	tcttcatgaa	gcaacgctgg	tacttggtat	gatgctaaaa	1260
cactttgact	ttgaagatca	tacaaactac	gagctcgata	ttaaagaaac	tttaacgtta	1320
aaacctgaag	gctttgtggt	aaaagcaaaa	tcgaaaaaaa	ttccgcttgg	cggtattcct	1380
tcacctagca	ctgaacagtc	tgctaaaaaa	gtacgcaaaa	aggcagaaaa	cgctcataat	1440
acgccgctgc	ttgtgctata	cggttcaaat	atgggaacag	ctgaagggaac	ggcgcgtgat	1500
ttagcagata	ttgcaatgag	caaaggattt	gcaccgcagg	tcgcaacgct	tgattcacac	1560
gccggaaatc	ttccgcgcga	aggagctgta	ttaattgtaa	cggcgtctta	taacgggtcat	1620
ccgcctgata	acgcaaagca	atthgtcgac	tggttagacc	aagcgtctgc	tgatgaagta	1680
aaaggcgctt	gctactccgt	atthggatgc	ggcgataaaa	actgggctac	tacgtatcaa	1740
aaagtgcctg	cttttatcga	tgaaacgctt	gccgctaaag	gggcagaaaa	catcgctgac	1800
cgcggtgaag	cagatgcaag	cgacgacttt	gaaggcacat	atgaagaatg	gcgtgaacat	1860
atgtggagtg	acgtagcagc	ctactttaac	ctcgacattg	aaaacagtga	agataataaa	1920

hp3238sgw-sp.sequence listing_ST25

```

tctactcttt cacttcaatt tgtcgacagc gccgcggata tgccgcttgc gaaaatgcac 1980
gggtgcgtttt caacgaacgt cgtagcaagc aaagaacttc aacagccagg cagtgcacga 2040
agcacgcgac atcttgaaat tgaacttcca aaagaagctt cttatcaaga aggagatcat 2100
ttaggtgtta ttctctgcaa ctatgaagga atagtaaacc gtgtaacagc aaggttcggc 2160
ctagatgcat cacagcaaat ccgtctggaa gcagaagaag aaaaattagc tcatttgcca 2220
ctcgctaaaa cagtatccgt agaagagctt ctgcaatacg tggagcttca agatcctgtt 2280
acgcgcacgc agcttcgcgc aatggctgct aaaacggtct gcccgccgca taaagtagag 2340
cttgaagcct tgcttgaaaa gcaagcctac aaagaacaag tgctggcaaa acgtttaaca 2400
atgcttgaac tgcttgaaaa ataccggcg tgtgaaatga aattcagcga atttatcgcc 2460
cttctgccaa gcatacgccc gcgctattac tcgatttctt catcacctcg tgtcgatgaa 2520
aaacaagcaa gcatcacggt cagcgttgtc tcaggagaag cgtggagcgg atatggagaa 2580
tataaaggaa ttgcgtcgaa ctatcttgcc gagctgcaag aaggagatac gattacgtgc 2640
tttatttcca caccgcagtc agaatttacg ctgcaaaaag accctgaaac gccgcttata 2700
atggtcggac cgggaacagg cgtcgcgccg tttagaggct ttgtgcaggc gcgcaaacag 2760
ctaaaagaac aaggacagtc acttgagaa gcacatttat acttcggctg ccgttcacct 2820
catgaagact atctgtatca agaagagctt gaaaacgccc aaagcgaagg catcattacg 2880
cttcataccg ctttttctcg catgccaaat cagccgaaaa catacgttca gcacgtaatg 2940
gaacaagacg gcaagaaatt gattgaactt cttgatcaag gagcgcactt ctatatattgc 3000
ggagacggaa gccaaatggc acctgccgtt gaagcaacgc ttatgaaaag ctatgctgac 3060
gttcaccaag tgagtgaagc agacgctcgc ttatggctgc agcagctaga agaaaaaggc 3120
cgatacgcaa aagacgtgtg ggctgggtaa 3150

```

<210> 12

<211> 1049

<212> PRT

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 12

```

Met Thr Ile Lys Glu Met Pro Gln Pro Lys Thr Phe Gly Glu Leu Lys
1           5           10          15
Asn Leu Pro Leu Leu Asn Thr Asp Lys Pro Val Gln Ala Leu Met Lys
          20          25          30
Ile Ala Asp Glu Leu Gly Glu Ile Phe Lys Phe Glu Ala Pro Gly Cys
          35          40          45
Val Thr Arg Tyr Leu Ser Ser Gln Arg Leu Ile Lys Glu Ala Cys Asp
          50          55          60
Glu Ser Arg Phe Asp Lys Asn Leu Ser Gln Ala Leu Lys Phe Ser Arg
65          70          75          80
Asp Phe Ala Gly Asp Gly Leu Phe Thr Ser Trp Thr His Glu Ile Asn
          85          90          95
Trp Lys Lys Ala His Asn Ile Leu Leu Pro Ser Phe Ser Gln Gln Ala
          100         105         110
Met Lys Gly Tyr His Ala Met Met Val Asp Ile Ala Val Gln Leu Val
          115         120         125
Gln Lys Trp Glu Arg Leu Asn Ala Asp Glu His Ile Glu Val Ser Glu
          130         135         140
Asp Met Thr Arg Leu Thr Leu Asp Thr Ile Gly Leu Cys Gly Phe Asn
145         150         155         160

```

hp3238sgw-sp.sequence listing_ST25

Tyr	Arg	Phe	Asn	Ser	Phe	Tyr	Arg	Asp	Gln	Pro	His	Pro	Phe	Ile	Ile	165	170	175
Ser	Met	Val	Arg	Ala	Leu	Asp	Glu	Val	Met	Asn	Lys	Leu	Gln	Arg	Ala	180	185	190
Asn	Pro	Asp	Asp	Pro	Ala	Tyr	Asp	Glu	Asn	Lys	Arg	Gln	Cys	Gln	Glu	195	200	205
Asp	Ile	Lys	Val	Met	Asn	Asp	Leu	Val	Asp	Lys	Ile	Ile	Ala	Asp	Arg	210	215	220
Lys	Ala	Arg	Gly	Glu	Gln	Ser	Asp	Asp	Leu	Leu	Thr	Gln	Met	Leu	Asn	225	230	235
Gly	Lys	Asp	Pro	Glu	Thr	Gly	Glu	Pro	Leu	Asp	Asp	Gly	Asn	Ile	Ser	245	250	255
Tyr	Gln	Ile	Ile	Thr	Phe	Leu	Ile	Ala	Gly	His	Glu	Thr	Thr	Ser	Gly	260	265	270
Leu	Leu	Ser	Phe	Ala	Leu	Tyr	Phe	Leu	Val	Lys	Asn	Pro	His	Val	Leu	275	280	285
Gln	Lys	Val	Ala	Glu	Glu	Ala	Ala	Arg	Val	Leu	Val	Asp	Pro	Val	Pro	290	295	300
Ser	Tyr	Lys	Gln	Val	Lys	Gln	Leu	Lys	Tyr	Val	Gly	Met	Val	Leu	Asn	305	310	315
Glu	Ala	Leu	Arg	Leu	Trp	Pro	Thr	Ala	Pro	Ala	Phe	Ser	Leu	Tyr	Ala	325	330	335
Lys	Glu	Asp	Thr	Val	Leu	Gly	Gly	Glu	Tyr	Pro	Leu	Glu	Lys	Gly	Asp	340	345	350
Glu	Val	Met	Val	Leu	Ile	Pro	Gln	Leu	His	Arg	Asp	Lys	Thr	Ile	Trp	355	360	365
Gly	Asp	Asp	Val	Glu	Glu	Phe	Arg	Pro	Glu	Arg	Phe	Glu	Asn	Pro	Ser	370	375	380
Ala	Ile	Pro	Gln	His	Ala	Phe	Lys	Pro	Phe	Gly	Asn	Gly	Gln	Arg	Ala	385	390	395
Cys	Ile	Gly	Gln	Gln	Phe	Ala	Leu	His	Glu	Ala	Thr	Leu	Val	Leu	Gly	405	410	415
Met	Met	Leu	Lys	His	Phe	Asp	Phe	Glu	Asp	His	Thr	Asn	Tyr	Glu	Leu	420	425	430
Asp	Ile	Lys	Glu	Thr	Leu	Thr	Leu	Lys	Pro	Glu	Gly	Phe	Val	Val	Lys	435	440	445
Ala	Lys	Ser	Lys	Lys	Ile	Pro	Leu	Gly	Gly	Ile	Pro	Ser	Pro	Ser	Thr	450	455	460
Glu	Gln	Ser	Ala	Lys	Lys	Val	Arg	Lys	Lys	Ala	Glu	Asn	Ala	His	Asn	465	470	475
Thr	Pro	Leu	Leu	Val	Leu	Tyr	Gly	Ser	Asn	Met	Gly	Thr	Ala	Glu	Gly	485	490	495
Thr	Ala	Arg	Asp	Leu	Ala	Asp	Ile	Ala	Met	Ser	Lys	Gly	Phe	Ala	Pro	500	505	510
Gln	Val	Ala	Thr	Leu	Asp	Ser	His	Ala	Gly	Asn	Leu	Pro	Arg	Glu	Gly	515	520	525
Ala	Val	Leu	Ile	Val	Thr	Ala	Ser	Tyr	Asn	Gly	His	Pro	Pro	Asp	Asn	530	535	540
Ala	Lys	Gln	Phe	Val	Asp	Trp	Leu	Asp	Gln	Ala	Ser	Ala	Asp	Glu	Val	545	550	555
Lys	Gly	Val	Arg	Tyr	Ser	Val	Phe	Gly	Cys	Gly	Asp	Lys	Asn	Trp	Ala	565	570	575

hp3238sgw-sp.sequence listing_ST25

Thr	Thr	Tyr	Gln	Lys	Val	Pro	Ala	Phe	Ile	Asp	Glu	Thr	Leu	Ala	Ala		
			580					585					590				
Lys	Gly	Ala	Glu	Asn	Ile	Ala	Asp	Arg	Gly	Glu	Ala	Asp	Ala	Ser	Asp		
		595					600					605					
Asp	Phe	Glu	Gly	Thr	Tyr	Glu	Glu	Trp	Arg	Glu	His	Met	Trp	Ser	Asp		
	610					615					620						
Val	Ala	Ala	Tyr	Phe	Asn	Leu	Asp	Ile	Glu	Asn	Ser	Glu	Asp	Asn	Lys		
625					630					635					640		
Ser	Thr	Leu	Ser	Leu	Gln	Phe	Val	Asp	Ser	Ala	Ala	Asp	Met	Pro	Leu		
			645					650						655			
Ala	Lys	Met	His	Gly	Ala	Phe	Ser	Thr	Asn	Val	Val	Ala	Ser	Lys	Glu		
			660					665						670			
Leu	Gln	Gln	Pro	Gly	Ser	Ala	Arg	Ser	Thr	Arg	His	Leu	Glu	Ile	Glu		
		675					680							685			
Leu	Pro	Lys	Glu	Ala	Ser	Tyr	Gln	Glu	Gly	Asp	His	Leu	Gly	Val	Ile		
	690					695					700						
Pro	Arg	Asn	Tyr	Glu	Gly	Ile	Val	Asn	Arg	Val	Thr	Ala	Arg	Phe	Gly		
705					710					715					720		
Leu	Asp	Ala	Ser	Gln	Gln	Ile	Arg	Leu	Glu	Ala	Glu	Glu	Glu	Lys	Leu		
				725					730					735			
Ala	His	Leu	Pro	Leu	Ala	Lys	Thr	Val	Ser	Val	Glu	Glu	Leu	Leu	Gln		
			740					745						750			
Tyr	Val	Glu	Leu	Gln	Asp	Pro	Val	Thr	Arg	Thr	Gln	Leu	Arg	Ala	Met		
	755						760					765					
Ala	Ala	Lys	Thr	Val	Cys	Pro	Pro	His	Lys	Val	Glu	Leu	Glu	Ala	Leu		
	770					775					780						
Leu	Glu	Lys	Gln	Ala	Tyr	Lys	Glu	Gln	Val	Leu	Ala	Lys	Arg	Leu	Thr		
785					790					795					800		
Met	Leu	Glu	Leu	Leu	Glu	Lys	Tyr	Pro	Ala	Cys	Glu	Met	Lys	Phe	Ser		
			805						810					815			
Glu	Phe	Ile	Ala	Leu	Leu	Pro	Ser	Ile	Arg	Pro	Arg	Tyr	Tyr	Ser	Ile		
			820					825						830			
Ser	Ser	Ser	Pro	Arg	Val	Asp	Glu	Lys	Gln	Ala	Ser	Ile	Thr	Val	Ser		
		835					840						845				
Val	Val	Ser	Gly	Glu	Ala	Trp	Ser	Gly	Tyr	Gly	Glu	Tyr	Lys	Gly	Ile		
	850					855					860						
Ala	Ser	Asn	Tyr	Leu	Ala	Glu	Leu	Gln	Glu	Gly	Asp	Thr	Ile	Thr	Cys		
865					870					875					880		
Phe	Ile	Ser	Thr	Pro	Gln	Ser	Glu	Phe	Thr	Leu	Pro	Lys	Asp	Pro	Glu		
				885					890					895			
Thr	Pro	Leu	Ile	Met	Val	Gly	Pro	Gly	Thr	Gly	Val	Ala	Pro	Phe	Arg		
			900					905						910			
Gly	Phe	Val	Gln	Ala	Arg	Lys	Gln	Leu	Lys	Glu	Gln	Gly	Gln	Ser	Leu		
		915					920						925				
Gly	Glu	Ala	His	Leu	Tyr	Phe	Gly	Cys	Arg	Ser	Pro	His	Glu	Asp	Tyr		
	930					935					940						
Leu	Tyr	Gln	Glu	Glu	Leu	Glu	Asn	Ala	Gln	Ser	Glu	Gly	Ile	Ile	Thr		
945					950					955					960		
Leu	His	Thr	Ala	Phe	Ser	Arg	Met	Pro	Asn	Gln	Pro	Lys	Thr	Tyr	Val		
				965					970						975		
Gln	His	Val	Met	Glu	Gln	Asp	Gly	Lys	Lys	Leu	Ile	Glu	Leu	Leu	Asp		
			980					985						990			

hp3238sgw-sp.sequence listing_ST25

Gln Gly Ala His Phe Tyr Ile Cys Gly Asp Gly Ser Gln Met Ala Pro
 995 1000 1005
 Ala Val Glu Ala Thr Leu Met Lys Ser Tyr Ala Asp Val His Gln Val
 1010 1015 1020
 Ser Glu Ala Asp Ala Arg Leu Trp Leu Gln Gln Leu Glu Glu Lys Gly
 1025 1030 1035 1040
 Arg Tyr Ala Lys Asp Val Trp Ala Gly
 1045

<210> 13
 <211> 3150
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> P450_BM3

<400> 13
 atgacaataa aggaaatgcc gcagccgaaa acgtttggag aacttaaaaa tttaccgtta 60
 ttaaacacag ataaaccggt tcaagctttg atgaaaattg cggatgaatt aggagaaatc 120
 tttaaattcg aggcgcttgg tcttgtaacg cgctacttat caagtcagcg tctaattaaa 180
 gaagcatgcg atgaatcacg ctttgataaa aacttaagtc aagcgcttaa atttgcacgt 240
 gattttgcag gagacgggtt atttacaagc tggacgcag aaaaaaattg gaaaaaagcg 300
 cataatatct tacttccaag cttcagtcag caggcaatga aaggctatca tgcgatgatg 360
 gtcgatatcg ccgtgcagct tgttcaaaaag tgggagcgtc taaatgcaga tgagcatatt 420
 gaagtaccgg aagacatgac acgtttaacg cttgatacaa ttggtctttg cggctttaac 480
 tatcgcttta acagctttta ccgagatcag cctcatccat ttattacaag tatggtccgt 540
 gcaactggatg aagcaatgaa caagcagcag cgagcaaatc cagacgacc agcttatgat 600
 gaaaacaagc gccagtttca agaagatatc aaggtgatga acgacctagt agataaaatt 660
 attgcagatc gcaaagcaag cgggtgaaca agcgatgatt tattaacgca tatgctaaac 720
 ggaaaagatc cagaaacggg tgagccgctt gatgacgaga acattcgcta tcaaattatt 780
 acattcttaa ttgcggggaca cgaaacaaca agtgggtctt tatcatttgc gctgtatttc 840
 ttagtgaaaa atccacatgt attacaaaaa gcagcagaag aagcagcacg agttctagta 900
 gatcctgttc caagctacaa acaagtcaaa cagcttaaat atgtcggcat ggtcttaaac 960
 gaagcgctgc gcttatggcc aactgctcct gcgttttccc tatatgcaaa agaagatacg 1020
 gtgcttgag gagaatatcc tttagaaaaa ggcgacgaac taatggttct gattcctcag 1080
 cttcaccgtg ataaaacaat ttggggagac gatgtggaag agttccgtcc agagcgtttt 1140
 gaaaatccaa gtgcgattcc gcagcatgcg tttaaaccgt ttggaaacgg tcagcggtgcg 1200
 tgtatcggtc agcagttcgc tcttcatgaa gcaacgctgg tacttggtat gatgctaaaa 1260
 cactttgact ttgaagatca tacaaactac gagctcgata ttaaagaaac tttaacgtta 1320
 aaacctgaag gctttgtggt aaaagcaaaa tcgaaaaaaa ttccgcttgg cggtattcct 1380
 tcacctagca ctgaacagtc tgctaaaaaa gtacgcaaaa aggcagaaaa cgctcataat 1440
 acgccgtgc ttgtgctata cggttcaaat atgggaacag ctgaaggaac ggcgcgatgat 1500
 ttagcagata ttgcaatgag caaaggattt gcaccgcagg tcgcaacgct tgattcacac 1560
 gccggaaatc ttccgcgcga aggagctgta ttaattgtaa cggcgtctta taacgggtcat 1620
 ccgcctgata acgcaaagca atttgtcgac tggttagacc aagcgtctgc tgatgaagta 1680
 aaaggcgttc gctactccgt atttggatgc ggcgataaaa actgggctac tacgtatcaa 1740
 aaagtgcctg cttttatcga tgaaacgctt gccgctaaag gggcagaaaa catcgctgac 1800
 cgcggtgaag cagatgcaag cgacgacttt gaaggcacat atgaagaatg gcgtgaacat 1860
 atgtggagtg acgtagcagc ctactttaac ctgcacattg aaaacagtga agataataaa 1920
 tctactcttt cacttcaatt tgtcgacagc gccgcggata tgccgcttgc gaaaatgcac 1980

hp3238sgw-sp.sequence listing_ST25

```

ggtgcgtttt caacgaacgt cgtagcaagc aaagaacttc aacagccagg cagtgcacga 2040
agcacgcgac atcttgaaat tgaacttcca aaagaagctt cttatcaaga aggagatcat 2100
ttaggtgtta ttcctcgcaa ctatgaagga atagtaaacc gtgtaacagc aaggttcggc 2160
ctagatgcat cacagcaaat ccgtctggaa gcagaagaag aaaaattagc tcatttgcca 2220
ctcgtaaaaa cagtatccgt agaagagctt ctgcaatacg tggagcttca agatcctgtt 2280
acgcgcacgc agcttcgcgc aatggctgct aaaacggtct gcccgccgca taaagtagag 2340
cttgaagcct tgcttgaaaa gcaagcctac aaagaacaag tgctggcaaa acgtttaaca 2400
atgcttgaac tgcttgaaaa ataccggcg tgtgaaatga aattcagcga atttatcgcc 2460
cttctgccaa gcatacgccc gcgctattac tcgatttctt catcacctcg tgtcgatgaa 2520
aaacaagcaa gcatcacggt cagcgttgtc tcaggagaag cgtggagcgg atatggagaa 2580
tataaaggaa ttgcgtcgaa ctatcttgcc gagctgcaag aaggagatac gattacgtgc 2640
tttatttcca caccgcagtc agaatttacg ctgcaaaaag accctgaaac gccgcttattc 2700
atggctcggac cgggaacagg cgtcgcgccg tttagaggct ttgtgcaggc gcgcaaacag 2760
ctaaaagaac aaggacagtc acttgagaa gcacatttat acttcggctg ccgttcacct 2820
catgaagact atctgtatca agaagagctt gaaaacgccc aaagcgaagg catcattacg 2880
cttcataccg ctttttctcg catgccaaat cagccgaaaa catacgttca gcacgtaatg 2940
gaacaagacg gcaagaaatt gattgaactt cttgatcaag gagcgcactt ctatatttgc 3000
ggagacggaa gccaaatggc acctgccgtt gaagcaacgc ttatgaaaag ctatgctgac 3060
gttcaccaag tgagtgaagc agacgctcgc ttatggctgc agcagctaga agaaaaaggc 3120
cgatacgcaa aagacgtgtg ggctgggtaa 3150

```

<210> 14

<211> 1049

<212> PRT

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 14

```

Met Thr Ile Lys Glu Met Pro Gln Pro Lys Thr Phe Gly Glu Leu Lys
1           5           10          15
Asn Leu Pro Leu Leu Asn Thr Asp Lys Pro Val Gln Ala Leu Met Lys
          20          25          30
Ile Ala Asp Glu Leu Gly Glu Ile Phe Lys Phe Glu Ala Pro Gly Leu
          35          40          45
Val Thr Arg Tyr Leu Ser Ser Gln Arg Leu Ile Lys Glu Ala Cys Asp
          50          55          60
Glu Ser Arg Phe Asp Lys Asn Leu Ser Gln Ala Leu Lys Phe Ala Arg
65          70          75          80
Asp Phe Ala Gly Asp Gly Leu Phe Thr Ser Trp Thr His Glu Lys Asn
          85          90          95
Trp Lys Lys Ala His Asn Ile Leu Leu Pro Ser Phe Ser Gln Gln Ala
          100         105         110
Met Lys Gly Tyr His Ala Met Met Val Asp Ile Ala Val Gln Leu Val
          115         120         125
Gln Lys Trp Glu Arg Leu Asn Ala Asp Glu His Ile Glu Val Pro Glu
          130         135         140
Asp Met Thr Arg Leu Thr Leu Asp Thr Ile Gly Leu Cys Gly Phe Asn
145         150         155         160
Tyr Arg Phe Asn Ser Phe Tyr Arg Asp Gln Pro His Pro Phe Ile Thr

```

hp3238sgw-sp.sequence listing_ST25

				165					170					175	
Ser	Met	Val	Arg	Ala	Leu	Asp	Glu	Ala	Met	Asn	Lys	Gln	Gln	Arg	Ala
			180					185					190		
Asn	Pro	Asp	Asp	Pro	Ala	Tyr	Asp	Glu	Asn	Lys	Arg	Gln	Phe	Gln	Glu
		195					200					205			
Asp	Ile	Lys	Val	Met	Asn	Asp	Leu	Val	Asp	Lys	Ile	Ile	Ala	Asp	Arg
	210					215					220				
Lys	Ala	Ser	Gly	Glu	Gln	Ser	Asp	Asp	Leu	Leu	Thr	His	Met	Leu	Asn
225					230					235					240
Gly	Lys	Asp	Pro	Glu	Thr	Gly	Glu	Pro	Leu	Asp	Asp	Glu	Asn	Ile	Arg
				245					250					255	
Tyr	Gln	Ile	Ile	Thr	Phe	Leu	Ile	Ala	Gly	His	Glu	Thr	Thr	Ser	Gly
			260					265						270	
Leu	Leu	Ser	Phe	Ala	Leu	Tyr	Phe	Leu	Val	Lys	Asn	Pro	His	Val	Leu
		275					280					285			
Gln	Lys	Ala	Ala	Glu	Glu	Ala	Ala	Arg	Val	Leu	Val	Asp	Pro	Val	Pro
	290					295					300				
Ser	Tyr	Lys	Gln	Val	Lys	Gln	Leu	Lys	Tyr	Val	Gly	Met	Val	Leu	Asn
305					310					315					320
Glu	Ala	Leu	Arg	Leu	Trp	Pro	Thr	Ala	Pro	Ala	Phe	Ser	Leu	Tyr	Ala
				325					330					335	
Lys	Glu	Asp	Thr	Val	Leu	Gly	Gly	Glu	Tyr	Pro	Leu	Glu	Lys	Gly	Asp
			340					345					350		
Glu	Leu	Met	Val	Leu	Ile	Pro	Gln	Leu	His	Arg	Asp	Lys	Thr	Ile	Trp
		355					360					365			
Gly	Asp	Asp	Val	Glu	Glu	Phe	Arg	Pro	Glu	Arg	Phe	Glu	Asn	Pro	Ser
	370					375					380				
Ala	Ile	Pro	Gln	His	Ala	Phe	Lys	Pro	Phe	Gly	Asn	Gly	Gln	Arg	Ala
385					390					395					400
Cys	Ile	Gly	Gln	Gln	Phe	Ala	Leu	His	Glu	Ala	Thr	Leu	Val	Leu	Gly
				405					410					415	
Met	Met	Leu	Lys	His	Phe	Asp	Phe	Glu	Asp	His	Thr	Asn	Tyr	Glu	Leu
			420					425					430		
Asp	Ile	Lys	Glu	Thr	Leu	Thr	Leu	Lys	Pro	Glu	Gly	Phe	Val	Val	Lys
		435					440					445			
Ala	Lys	Ser	Lys	Lys	Ile	Pro	Leu	Gly	Gly	Ile	Pro	Ser	Pro	Ser	Thr
	450					455					460				
Glu	Gln	Ser	Ala	Lys	Lys	Val	Arg	Lys	Lys	Ala	Glu	Asn	Ala	His	Asn
465					470					475					480
Thr	Pro	Leu	Leu	Val	Leu	Tyr	Gly	Ser	Asn	Met	Gly	Thr	Ala	Glu	Gly
				485					490					495	
Thr	Ala	Arg	Asp	Leu	Ala	Asp	Ile	Ala	Met	Ser	Lys	Gly	Phe	Ala	Pro
			500					505					510		
Gln	Val	Ala	Thr	Leu	Asp	Ser	His	Ala	Gly	Asn	Leu	Pro	Arg	Glu	Gly

hp3238sgw-sp.sequence listing ST25

Lys	Gly	Ala	Glu	Asn	Ile	Ala	Asp	Arg	Gly	Glu	Ala	Asp	Ala	Ser	Asp
		595					600					605			
Asp	Phe	Glu	Gly	Thr	Tyr	Glu	Glu	Trp	Arg	Glu	His	Met	Trp	Ser	Asp
	610					615					620				
Val	Ala	Ala	Tyr	Phe	Asn	Leu	Asp	Ile	Glu	Asn	Ser	Glu	Asp	Asn	Lys
625					630					635					640
Ser	Thr	Leu	Ser	Leu	Gln	Phe	Val	Asp	Ser	Ala	Ala	Asp	Met	Pro	Leu
				645					650				655		
Ala	Lys	Met	His	Gly	Ala	Phe	Ser	Thr	Asn	Val	Val	Ala	Ser	Lys	Glu
			660					665					670		
Leu	Gln	Gln	Pro	Gly	Ser	Ala	Arg	Ser	Thr	Arg	His	Leu	Glu	Ile	Glu
		675					680					685			
Leu	Pro	Lys	Glu	Ala	Ser	Tyr	Gln	Glu	Gly	Asp	His	Leu	Gly	Val	Ile
	690					695					700				
Pro	Arg	Asn	Tyr	Glu	Gly	Ile	Val	Asn	Arg	Val	Thr	Ala	Arg	Phe	Gly
705					710					715					720
Leu	Asp	Ala	Ser	Gln	Gln	Ile	Arg	Leu	Glu	Ala	Glu	Glu	Glu	Lys	Leu
				725					730					735	
Ala	His	Leu	Pro	Leu	Ala	Lys	Thr	Val	Ser	Val	Glu	Glu	Leu	Leu	Gln
			740					745					750		
Tyr	Val	Glu	Leu	Gln	Asp	Pro	Val	Thr	Arg	Thr	Gln	Leu	Arg	Ala	Met
		755					760					765			
Ala	Ala	Lys	Thr	Val	Cys	Pro	Pro	His	Lys	Val	Glu	Leu	Glu	Ala	Leu
	770					775					780				
Leu	Glu	Lys	Gln	Ala	Tyr	Lys	Glu	Gln	Val	Leu	Ala	Lys	Arg	Leu	Thr
785					790					795					800
Met	Leu	Glu	Leu	Leu	Glu	Lys	Tyr	Pro	Ala	Cys	Glu	Met	Lys	Phe	Ser
				805					810					815	
Glu	Phe	Ile	Ala	Leu	Leu	Pro	Ser	Ile	Arg	Pro	Arg	Tyr	Tyr	Ser	Ile
			820					825					830		
Ser	Ser	Ser	Pro	Arg	Val	Asp	Glu	Lys	Gln	Ala	Ser	Ile	Thr	Val	Ser
		835					840					845			
Val	Val	Ser	Gly	Glu	Ala	Trp	Ser	Gly	Tyr	Gly	Glu	Tyr	Lys	Gly	Ile
	850					855					860				
Ala	Ser	Asn	Tyr	Leu	Ala	Glu	Leu	Gln	Glu	Gly	Asp	Thr	Ile	Thr	Cys
865					870					875					880
Phe	Ile	Ser	Thr	Pro	Gln	Ser	Glu	Phe	Thr	Leu	Pro	Lys	Asp	Pro	Glu
				885					890					895	
Thr	Pro	Leu	Ile	Met	Val	Gly	Pro	Gly	Thr	Gly	Val	Ala	Pro	Phe	Arg
			900					905					910		
Gly	Phe	Val	Gln	Ala	Arg	Lys	Gln	Leu	Lys	Glu	Gln	Gly	Gln	Ser	Leu
		915					920					925			
Gly	Glu	Ala	His	Leu	Tyr	Phe	Gly	Cys	Arg	Ser	Pro	His	Glu	Asp	Tyr
	930					935</									

hp3238sgw-sp.sequence listing_ST25

995	1000	1005
Ala Val Glu Ala Thr Leu Met Lys Ser Tyr Ala Asp Val His Gln Val		
1010	1015	1020
Ser Glu Ala Asp Ala Arg Leu Trp Leu Gln Gln Leu Glu Glu Lys Gly		
1025	1030	1035
Arg Tyr Ala Lys Asp Val Trp Ala Gly		1040
	1045	

<210> 15

<211> 3150

<212> DNA

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 15

```

atgacaataa aggaaatgcc gcagccgaaa acgtttggag aacttaaaaa tttaccgtta 60
ttaaacacag ataaaccggt tcaagctttg atgaaaattg cggatgaatt aggagaaatc 120
tttaaattcg aggcgcctgg ttgtgtaacg cgctacttat caagtcagcg tctaattaaa 180
gaagcatgcg atgaatcacg ctttgataaa aacttaagtc aagcgcttaa atttgcacgt 240
gattttctgg gagacgggtt agtgacaagc tggacgcacg aaataaattg gaaaaaagcg 300
cataatatct tacttccaag ctttagtcag caggcaatga aaggctatca tgcgatgatg 360
gtcgatatcg ccgtgcagct tgttcaaaaag tgggagcgtc taaatgcaga tgagcatatt 420
gaagtatcgg aagacatgac acgtttaacg cttgatacaa ttggtctttg cggctttaac 480
tatcgcttta acagctttta ccgagatcag cctcatccat ttattataag tatggtccgt 540
gcactggatg aagtaatgaa caagctgcag cgagcaaadc cagacgacc agcttatgat 600
gaaaacaagc gccagtgta agaagatatc aaggtgatga acgacctagt agataaaatt 660
attgcagatc gcaaagcaag ggggtgaaca agcgatgatt tattaacgca gatgctaaac 720
ggaaaagatc cagaaacggg tgagccgctt gatgacggga acattagcta tcaaattatt 780
acattcttaa ttgcgggaca cgaaacaaca agtgggtctt tatcatttgc gctgtatttc 840
ttagtgaaaa atccacatgt attacaaaaa gtagcagaag aagcagcacg agttctagta 900
gatcctgttc caagctacaa acaagtcaaa cagcttaaat atgtcggcat ggtcttaaac 960
gaagcgtgct gcttatggcc aactgctcct gcgttttccc tatatgcaaa agaagatacg 1020
gtgcttgagg gagaatatcc tttagaaaaa ggcgacgaag taatggttct gattcctcag 1080
cttcaccgtg ataaaacaat ttggggagac gatgtggagg agttccgtcc agagcgtttt 1140
gaaaatccaa gtgcgattcc gcagcatgcg tttaaaccgt ttggaaacgg tcagcgtgctg 1200
tgtatcggtc agcagttcgc tcttcatgaa gcaacgctgg tacttggtat gatgctaaaa 1260
cactttgact ttgaagatca tacaactac gagctcgata ttaaagaaac tttaacgtta 1320
aaacctgaag gctttgtggt aaaagcaaaa tcgaaaaaaa ttccgcttgg cgttattcct 1380
tcacntagca ctgaacagtc tgctaaaaaa gtacgcaaaa aggcagaaaa cgctcataat 1440
acgccgctgc ttgtgctata cggttcaa atgggaacag ctgaaggaaac ggcgctgat 1500
ttagcagata ttgcaatgag caaaggattt gcaccgcagg tcgcaacgct tgattcacac 1560
gccggaaatc ttccgcgcga aggagctgta ttaattgtaa cggcgtctta taacgggtcat 1620
ccgcctgata acgcaaagca atttgtcgac tggttagacc aagcgtctgc tgatgaagta 1680
aaaggcgttc gctactccgt atttggatgc ggcgataaaa actgggctac tacgtatcaa 1740
aaagtgcctg cttttatcga tgaaacgctt gccgctaaag gggcagaaaa catcgctgac 1800
cgcggtgaag cagatgcaag cgacgacttt gaaggcacat atgaagaatg gcgtgaacat 1860
atgtggagtg acgtagcagc ctactttaac ctcgacattg aaaacagtga agataataaa 1920
tctactcttt cacttcaatt tgtcgacagc gccgcggata tgccgcttgc gaaaatgcac 1980
ggtgcgtttt caacgaacgt cgtagcaagc aaagaacttc aacagccagg cagtgcacga 2040

```

hp3238sgw-sp.sequence listing_ST25

```

agcacgcgac atcttgaaat tgaacttcca aaagaagctt cttatcaaga aggagatcat 2100
ttaggtgtta ttctctgcaa ctatgaagga atagtaaacc gtgtaacagc aaggttcggc 2160
ctagatgcat cacagcaaat ccgtctggaa gcagaagaag aaaaattagc tcatttgcca 2220
ctcgctaaaa cagtatccgt agaagagctt ctgcaatacg tggagcttca agatcctggt 2280
acgcgcacgc agcttcgcgc aatggctgct aaaacggtct gcccgccgca taaagtagag 2340
cttgaagcct tgcttgaaaa gcaagcctac aaagaacaag tgctggcaaa acgtttaaca 2400
atgcttgaac tgcttgaaaa ataccggcg tgtgaaatga aattcagcga atttatcgcc 2460
cttctgccaa gcatacgccc gcgctattac tcgatttctt catcacctcg tgtcgatgaa 2520
aaacaagcaa gcatcacggt cagcgttgtc tcaggagaag cgtggagcgg atatggagaa 2580
tataaaggaa ttgcgtcgaa ctatcttgcc gagctgcaag aaggagatac gattacgtgc 2640
tttattttcca caccgcagtc agaattttac ctgccaaaag accctgaaac gccgcttatt 2700
atggctcggac cgggaacagg cgctcgcgccg tttagaggct ttgtgcaggc gcgcaaacag 2760
ctaaaagaac aaggacagtc acttgagaa gcacatttat acttcggctg ccgttcacct 2820
catgaagact atctgtatca agaagagctt gaaaacgccc aaagcgaagg catcattacg 2880
cttcataccg ctttttctcg catgccaaat cagccgaaaa catacgttca gcacgtaatg 2940
gaacaagacg gcaagaaatt gattgaactt cttgatcaag gagcgcactt ctatatattgc 3000
ggagacggaa gccaaatggc acctgccgtt gaagcaacgc ttatgaaaag ctatgctgac 3060
gttcaccaag tgagtgaagc agacgctcgc ttatggctgc agcagctaga agaaaaaggc 3120
cgatacgcaa aagacgtgtg ggctgggtaa 3150

```

<210> 16

<211> 1049

<212> PRT

<213> Artificial Sequence

<220>

<223> P450_BM3

<400> 16

```

Met Thr Ile Lys Glu Met Pro Gln Pro Lys Thr Phe Gly Glu Leu Lys
1           5           10           15
Asn Leu Pro Leu Leu Asn Thr Asp Lys Pro Val Gln Ala Leu Met Lys
          20           25           30
Ile Ala Asp Glu Leu Gly Glu Ile Phe Lys Phe Glu Ala Pro Gly Cys
          35           40           45
Val Thr Arg Tyr Leu Ser Ser Gln Arg Leu Ile Lys Glu Ala Cys Asp
          50           55           60
Glu Ser Arg Phe Asp Lys Asn Leu Ser Gln Ala Leu Lys Phe Ala Arg
65           70           75           80
Asp Phe Leu Gly Asp Gly Leu Val Thr Ser Trp Thr His Glu Ile Asn
          85           90           95
Trp Lys Lys Ala His Asn Ile Leu Leu Pro Ser Phe Ser Gln Gln Ala
          100          105          110
Met Lys Gly Tyr His Ala Met Met Val Asp Ile Ala Val Gln Leu Val
          115          120          125
Gln Lys Trp Glu Arg Leu Asn Ala Asp Glu His Ile Glu Val Ser Glu
          130          135          140
Asp Met Thr Arg Leu Thr Leu Asp Thr Ile Gly Leu Cys Gly Phe Asn
145          150          155          160
Tyr Arg Phe Asn Ser Phe Tyr Arg Asp Gln Pro His Pro Phe Ile Ile
          165          170          175

```

hp3238sgw-sp.sequence listing_ST25

Ser	Met	Val	Arg	Ala	Leu	Asp	Glu	Val	Met	Asn	Lys	Leu	Gln	Arg	Ala	180	185	190
Asn	Pro	Asp	Asp	Pro	Ala	Tyr	Asp	Glu	Asn	Lys	Arg	Gln	Cys	Gln	Glu	195	200	205
Asp	Ile	Lys	Val	Met	Asn	Asp	Leu	Val	Asp	Lys	Ile	Ile	Ala	Asp	Arg	210	215	220
Lys	Ala	Arg	Gly	Glu	Gln	Ser	Asp	Asp	Leu	Leu	Thr	Gln	Met	Leu	Asn	225	230	235
Gly	Lys	Asp	Pro	Glu	Thr	Gly	Glu	Pro	Leu	Asp	Asp	Gly	Asn	Ile	Ser	245	250	255
Tyr	Gln	Ile	Ile	Thr	Phe	Leu	Ile	Ala	Gly	His	Glu	Thr	Thr	Ser	Gly	260	265	270
Leu	Leu	Ser	Phe	Ala	Leu	Tyr	Phe	Leu	Val	Lys	Asn	Pro	His	Val	Leu	275	280	285
Gln	Lys	Val	Ala	Glu	Glu	Ala	Ala	Arg	Val	Leu	Val	Asp	Pro	Val	Pro	290	295	300
Ser	Tyr	Lys	Gln	Val	Lys	Gln	Leu	Lys	Tyr	Val	Gly	Met	Val	Leu	Asn	305	310	315
Glu	Ala	Leu	Arg	Leu	Trp	Pro	Thr	Ala	Pro	Ala	Phe	Ser	Leu	Tyr	Ala	325	330	335
Lys	Glu	Asp	Thr	Val	Leu	Gly	Gly	Glu	Tyr	Pro	Leu	Glu	Lys	Gly	Asp	340	345	350
Glu	Val	Met	Val	Leu	Ile	Pro	Gln	Leu	His	Arg	Asp	Lys	Thr	Ile	Trp	355	360	365
Gly	Asp	Asp	Val	Glu	Glu	Phe	Arg	Pro	Glu	Arg	Phe	Glu	Asn	Pro	Ser	370	375	380
Ala	Ile	Pro	Gln	His	Ala	Phe	Lys	Pro	Phe	Gly	Asn	Gly	Gln	Arg	Ala	385	390	395
Cys	Ile	Gly	Gln	Gln	Phe	Ala	Leu	His	Glu	Ala	Thr	Leu	Val	Leu	Gly	405	410	415
Met	Met	Leu	Lys	His	Phe	Asp	Phe	Glu	Asp	His	Thr	Asn	Tyr	Glu	Leu	420	425	430
Asp	Ile	Lys	Glu	Thr	Leu	Thr	Leu	Lys	Pro	Glu	Gly	Phe	Val	Val	Lys	435	440	445
Ala	Lys	Ser	Lys	Lys	Ile	Pro	Leu	Gly	Gly	Ile	Pro	Ser	Pro	Ser	Thr	450	455	460
Glu	Gln	Ser	Ala	Lys	Lys	Val	Arg	Lys	Lys	Ala	Glu	Asn	Ala	His	Asn	465	470	475
Thr	Pro	Leu	Leu	Val	Leu	Tyr	Gly	Ser	Asn	Met	Gly	Thr	Ala	Glu	Gly	485	490	495
Thr	Ala	Arg	Asp	Leu	Ala	Asp	Ile	Ala	Met	Ser	Lys	Gly	Phe	Ala	Pro	500	505	510
Gln	Val	Ala	Thr	Leu	Asp	Ser	His	Ala	Gly	Asn	Leu	Pro	Arg	Glu	Gly	515	520	525
Ala	Val	Leu	Ile	Val	Thr	Ala	Ser	Tyr	Asn	Gly	His	Pro	Pro	Asp	Asn	530	535	540
Ala	Lys	Gln	Phe	Val	Asp	Trp	Leu	Asp	Gln	Ala	Ser	Ala	Asp	Glu	Val	545	550	555
Lys	Gly	Val	Arg	Tyr	Ser	Val	Phe	Gly	Cys	Gly	Asp	Lys	Asn	Trp	Ala	565	570	575
Thr	Thr	Tyr	Gln	Lys	Val	Pro	Ala	Phe	Ile	Asp	Glu	Thr	Leu	Ala	Ala	580	585	590

hp3238sgw-sp.sequence listing_ST25

Lys	Gly	Ala	Glu	Asn	Ile	Ala	Asp	Arg	Gly	Glu	Ala	Asp	Ala	Ser	Asp	595	600	605
Asp	Phe	Glu	Gly	Thr	Tyr	Glu	Glu	Trp	Arg	Glu	His	Met	Trp	Ser	Asp	610	615	620
Val	Ala	Ala	Tyr	Phe	Asn	Leu	Asp	Ile	Glu	Asn	Ser	Glu	Asp	Asn	Lys	625	630	635
Ser	Thr	Leu	Ser	Leu	Gln	Phe	Val	Asp	Ser	Ala	Ala	Asp	Met	Pro	Leu	645	650	655
Ala	Lys	Met	His	Gly	Ala	Phe	Ser	Thr	Asn	Val	Val	Ala	Ser	Lys	Glu	660	665	670
Leu	Gln	Gln	Pro	Gly	Ser	Ala	Arg	Ser	Thr	Arg	His	Leu	Glu	Ile	Glu	675	680	685
Leu	Pro	Lys	Glu	Ala	Ser	Tyr	Gln	Glu	Gly	Asp	His	Leu	Gly	Val	Ile	690	695	700
Pro	Arg	Asn	Tyr	Glu	Gly	Ile	Val	Asn	Arg	Val	Thr	Ala	Arg	Phe	Gly	705	710	715
Leu	Asp	Ala	Ser	Gln	Gln	Ile	Arg	Leu	Glu	Ala	Glu	Glu	Glu	Lys	Leu	725	730	735
Ala	His	Leu	Pro	Leu	Ala	Lys	Thr	Val	Ser	Val	Glu	Glu	Leu	Leu	Gln	740	745	750
Tyr	Val	Glu	Leu	Gln	Asp	Pro	Val	Thr	Arg	Thr	Gln	Leu	Arg	Ala	Met	755	760	765
Ala	Ala	Lys	Thr	Val	Cys	Pro	Pro	His	Lys	Val	Glu	Leu	Glu	Ala	Leu	770	775	780
Leu	Glu	Lys	Gln	Ala	Tyr	Lys	Glu	Gln	Val	Leu	Ala	Lys	Arg	Leu	Thr	785	790	795
Met	Leu	Glu	Leu	Leu	Glu	Lys	Tyr	Pro	Ala	Cys	Glu	Met	Lys	Phe	Ser	805	810	815
Glu	Phe	Ile	Ala	Leu	Leu	Pro	Ser	Ile	Arg	Pro	Arg	Tyr	Tyr	Ser	Ile	820	825	830
Ser	Ser	Ser	Pro	Arg	Val	Asp	Glu	Lys	Gln	Ala	Ser	Ile	Thr	Val	Ser	835	840	845
Val	Val	Ser	Gly	Glu	Ala	Trp	Ser	Gly	Tyr	Gly	Glu	Tyr	Lys	Gly	Ile	850	855	860
Ala	Ser	Asn	Tyr	Leu	Ala	Glu	Leu	Gln	Glu	Gly	Asp	Thr	Ile	Thr	Cys	865	870	875
Phe	Ile	Ser	Thr	Pro	Gln	Ser	Glu	Phe	Thr	Leu	Pro	Lys	Asp	Pro	Glu	885	890	895
Thr	Pro	Leu	Ile	Met	Val	Gly	Pro	Gly	Thr	Gly	Val	Ala	Pro	Phe	Arg	900	905	910
Gly	Phe	Val	Gln	Ala	Arg	Lys	Gln	Leu	Lys	Glu	Gln	Gly	Gln	Ser	Leu	915	920	925
Gly	Glu	Ala	His	Leu	Tyr	Phe	Gly	Cys	Arg	Ser	Pro	His	Glu	Asp	Tyr	930	935	940
Leu	Tyr	Gln	Glu	Glu	Leu	Glu	Asn	Ala	Gln	Ser	Glu	Gly	Ile	Ile	Thr	945	950	955
Leu	His	Thr	Ala	Phe	Ser	Arg	Met	Pro	Asn	Gln	Pro	Lys	Thr	Tyr	Val	965	970	975
Gln	His	Val	Met	Glu	Gln	Asp	Gly	Lys	Lys	Leu	Ile	Glu	Leu	Leu	Asp	980	985	990
Gln	Gly	Ala	His	Phe	Tyr	Ile	Cys	Gly	Asp	Gly	Ser	Gln	Met	Ala	Pro	995	1000	1005

hp3238sgw-sp.sequence listing_ST25

Ala	Val	Glu	Ala	Thr	Leu	Met	Lys	Ser	Tyr	Ala	Asp	Val	His	Gln	Val
1010						1015					1020				
Ser	Glu	Ala	Asp	Ala	Arg	Leu	Trp	Leu	Gln	Gln	Leu	Glu	Glu	Lys	Gly
1025					1030				1035						1040
Arg	Tyr	Ala	Lys	Asp	Val	Trp	Ala	Gly							
				1045											