(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
12 January 2017 (12.01.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/007547 A1

(51) International Patent Classification:
C07K 14/80 (2006.01) *C12N 15/09* (2006.01)
C12N 9/02 (2006.01) *C12N 5/10* (2006.01)

(21) International Application Number:
PCT/US2016/035399

(22) International Filing Date:
2 June 2016 (02.06.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/189,281 7 July 2015 (07.07.2015) US

(71) Applicant: **CODEXIS, INC.** [US/US]; 200 Penobscot Drive, Redwood City, California 94063 (US).

(72) Inventors: **OSBORNE, Robert**; 6311 Terra Verde Dr Apt 215, Raleigh, North Carolina 27609 (US). **MITCHELL, Vesna**; 200 Penobscot Drive, Redwood City, California 94063 (US). **HTWE, Khin Yu Naing**; 48 Larkspur Ave, Daly City, California 94015 (US). **ZHANG, Xiyun**; 34279 Maybird Circle, Fremont, California 94555 (US). **MILCZEK, Erika, M.**; 126 East Lincoln Avenue, Rahway, New Jersey 07065 (US). **MOORE, Jeffrey, C.**; 367 Orenda Circle, Westfield, New Jersey 07090 (US).

(74) Agent: **MACKNIGHT, Kamrin, T.**; CODEXIS, INC., 200 Penobscot Drive, Redwood City, California 94063 (US).

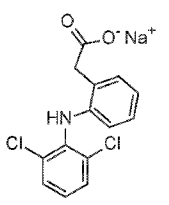
(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE,

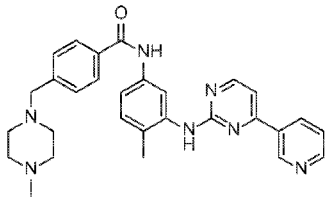
[Continued on next page]

(54) Title: NOVEL P450-BM3 VARIANTS WITH IMPROVED ACTIVITY

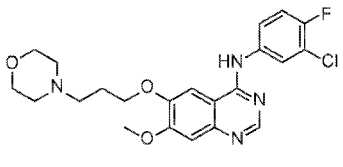
(57) Abstract: The present invention provides improved P450-BM3 variants with improved activity. In some embodiments, the P450-BM3 variants exhibit improved activity over a wide range of substrates.



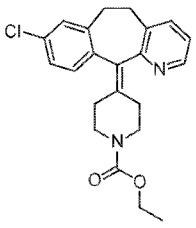
Diclofenac



Imatinib



Gefitinib



Loratadine

FIG. 1

WO 2017/007547 A1

DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

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 monooxgenases ("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, geftinib, 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.

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 "Xn," 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 "XnY" where "Xn" 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 "AnB", 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 this 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 thermoactivity, 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 is 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 10, 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 adenyltransferase), 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 (ΔfhuA) 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, 5,811238, 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]; Crameri et al., Nature, 391:288-291 [1998]; Crameri, et al., Nat. Biotechnol., 15:436-438 [1997]; Zhang et al., Proc. Nat. Acad. Sci. U.S.A., 94:4504-4509 [1997]; Crameri 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 μM (micromolar); nM (nanomolar); mol (moles); gm and g (gram); mg (milligrams); ug and μg (micrograms); L and l (liter); ml and mL (milliliter); cm

(centimeters); mm (millimeters); um 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 ug/L ammonium molybdate tetrahydrate, 5.8 mgs/L zinc sulfate heptahydrate, 620 ug/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 OD_{600} of 0.8 – 1.2, P450 expression was induced by addition of IPTG (500 uM) and the heme precursor 5-aminolevulinic acid (5-ALA) to a final concentration of 500 uM. 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), $MgSO_4$ (10 mM), DTT (1 mM), lysozyme (1 mg/mL), PMBS (0.5 mg/mL), and DNAseI (3 μgs/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 OD_{600} of 0.8 – 1.2, P450 expression was induced by addition of IPTG (500 uM) and the heme precursor 5-aminolevulinic acid (5-ALA) to a final concentration of 500 uM. 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 μM 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 μL) followed by enzyme solution (50 μL) was added to the plates. The substrate stock solution was added to the enzyme solutions (15 μL 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 μL 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 μL 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 H_2O) 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 H_2O) 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 nth 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 uL of quenched sample was injected onto an Acquity UPLC HSS T3 column, 100Å, (1.8 um; 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 H_2O) 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 nth 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, [a]TON is calculated as ([substrate]*% conversion/[P450]) and[b]FIOPC 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	ǂ	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)

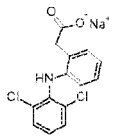
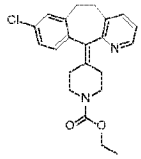
SF Validation Substrate(s)	Parent (SEQ ID NO:2)	Variant 6 (V51S;D118R;L219Y; K236H;P351T;S353T; N579T;T582A;E624D; A632S;D727S;Q931K)	Variant 7 (V51S;D118R;T179V; L219Y;P351T;N579T; T582A;E624D;D727S)	Variant 8 (D118R;L219Y;K236H ; S353T;N579T;T582A; E624D;D727S)
 Diclofenac	[Diclofenac] = 1 mM TON[a] = ǂ % 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 = ǂ	[Loratadine] = 1 mM TON = # % Conv. = ++	[Loratadine] = 1 mM TON = # % Conv. = ++	Loratadine] = 1 mM TON = # % Conv. = ++

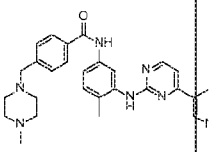
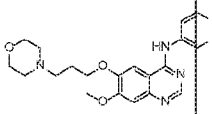
Table 2. Results for Variants 6, 7, and 8 (with substitutions shown with reference to SEQ ID NO:2)				
SF Validation Substrate(s)	**Parent (SEQ ID NO:2)**	**Variant 6** (V51S;D118R;L219Y; K236H;P351T;S353T; N579T;T582A;E624D; A632S;D727S;Q931K)	**Variant 7** (V51S;D118R;T179V; L219Y;P351T;N579T; T582A;E624D;D727S)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

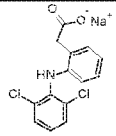
Table 3. Results for Variants 14 and 15 (with substitutions shown with reference to SEQ ID NO:4)			
SF Validation Substrate(s)	**Backbone (SEQ ID NO:4)**	**Variant 14** (C48S;I95P;G115R; L216Y;D232H;M491A; N574T;D722S)	**Variant 15** (K32R;C48S;I95P; L216Y;D232H;E349T; M491A;N574T;T577A; E619D;D722S)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

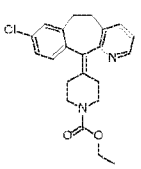
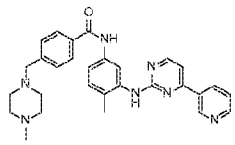
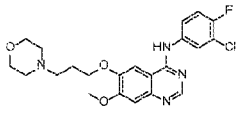
Table 3. Results for Variants 14 and 15 (with substitutions shown with reference to SEQ ID NO:4)			
SF Validation Substrate(s)	**Backbone (SEQ ID NO:4)**	**Variant 14** (C48S;I95P;G115R; L216Y;D232H;M491A; N574T;D722S)	**Variant 15** (K32R;C48S;I95P; L216Y;D232H;E349T; M491A;N574T;T577A; E619D;D722S)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

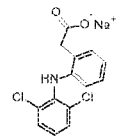
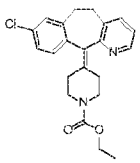
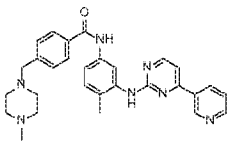
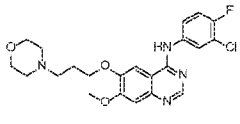
SF Validation Substrate(s)	Backbone (SEQ ID NO:6)	Variant 16 (K32R;C48S;Y52F; G115R;L216Y;D232H; E349K;M491A;T577A; E619D;A627S;D722S)	Variant 17 (K32R;C48S;Y52F; I95P;Q111R;G115R; I176V;L216Y;D232H; E349T;M491A;T577A; E619D;D722S)
 Diclofenac	[Diclofenac] = 1 mM TON = # % Conv. = ++	[Diclofenac] = 1 mM TON = ### % Conv. = ++++ FIOPC (A12) = *	[Diclofenac] = 1 mM TON = ### % Conv. = ++++ FIOPC (A12) = *
 Loratadine	[Loratadine] = 1 mM TON = ǂ % Conv. = +	[Loratadine] = 1 mM TON = ǂ % Conv. = + FIOPC (A12) = *	[Loratadine] = 1 mM TON = ǂ % 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)

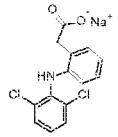
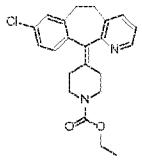
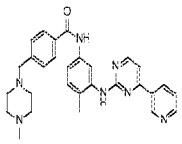
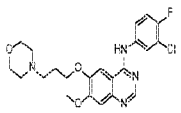
SF Validation Substrate(s)	Backbone	Variant 9 (Y52F;D232H;E349K; E619D;D722S;R762H)	Variant 10 (K32R;Y52F;K95P; Q111R;G115R;I176V; L216Y;D232H;P347T; E349T;M491A;T577A; E619D;D722S)	Variant 11 (K95P;Q111R;G115R; I176V;L216Y;E349T; M491A;T577A;E619D; D722S)
 Diclofenac	[Diclofenac] = 1 mM TON = # % Conv. = ++	[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 = ǂ % Conv. = +	[Gefitinib] = 1 mM TON = ǂ % Conv. = + FIOPC (B02) = *	[Gefitinib] = 1 mM TON = # % Conv. = ++ FIOPC (B02) = ****	[Gefitinib] = 1 mM TON = ǂ % Conv. = + FIOPC (B02) = *

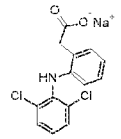
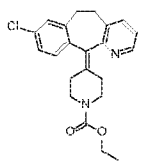
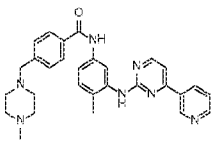
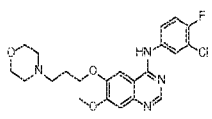
Table 6. Results for Variants 3, 4, and 5 (with substitutions shown with reference to SEQ ID NO:10)				
SF Validation Substrate(s)	**Backbone**	**Variant 3** (K32R;C48S;Y52F; I95P;L216Y;D232H; E349K;M491A;N574T; D722S)	**Variant 4** (K32R;C48S;Y52F; I95P;Q111R;L216Y; D232H;P347T;E349K; M491A;N574T;E619D; D722S)	**Variant 5** (C48S;I95P;G115R; L216Y;S231R;D232H; M491A;N574T;T577A; E619D;D722S)
 Diclofenac	[Diclofenac] = 1 mM TON = ###	[Diclofenac] = 1 mM TON = ###	[Diclofenac] = 1 mM TON = ###	[Diclofenac] = 1 mM TON = ###
	% Conv. = ++++	% Conv. = ++++	% Conv. = ++++	% Conv. = ++++
	FIOPC (B12) = *	FIOPC (B12) = *	FIOPC (B12) = *	FIOPC (B12) = *
 Loratadine	[Loratadine] = 1 mM TON = ###	[Loratadine] = 1 mM TON = ###	[Loratadine] = 1 mM TON = ###	[Loratadine] = 1 mM TON = ###
	% Conv. = ++++	% Conv. = ++++	% Conv. = ++++	% Conv. = ++++
	FIOPC (B12) = *	FIOPC (B12) = *	FIOPC (B12) = *	FIOPC (B12) = *
 Imatinib	[Imatinib] = 1 mM TON = #	[Imatinib] = 1 mM TON = ###	[Imatinib] = 1 mM TON = ###	[Imatinib] = 1 mM TON = ###
	% Conv. = ++	% Conv. = ++++	% Conv. = ++++	% Conv. = ++++
		FIOPC (B12) = ***	FIOPC (B02) = ***	FIOPC (B02) = ***
 Gefitinib	[Gefitinib] = 1 mM TON = #	[Gefitinib] = 1 mM TON = ###	[Gefitinib] = 1 mM TON = ###	[Gefitinib] = 1 mM TON = ###
	% Conv. =	% Conv. = ++++	% Conv. = ++++	% 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) = ***

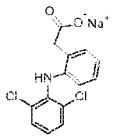
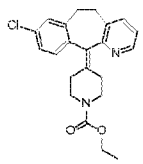
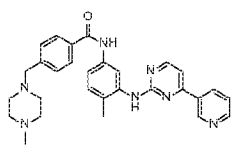
Table 7. Results for Variants 12 and 13 (with substitutions shown with reference to SEQ ID NO:12)			
SF Validation Substrate(s)	**Backbone**	**Variant 12** (K32R;C48S;Y52F; I95P;Q111R;G115R; D232H;M491A;T577A; A627S;D722S)	**Variant 13** (K32R;C48S;Y52F; I95P;G115R;D232H; M491A;N574T;T577A; E619D;A627S;D722S)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

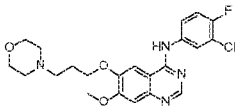
Table 7. Results for Variants 12 and 13 (with substitutions shown with reference to SEQ ID NO:12)			
SF Validation Substrate(s)	**Backbone**	**Variant 12** (K32R;C48S;Y52F; I95P;Q111R;G115R; D232H;M491A;T577A; A627S;D722S)	**Variant 13** (K32R;C48S;Y52F; I95P;G115R;D232H; M491A;N574T;T577A; E619D;A627S;D722S)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

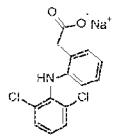
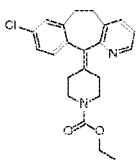
Table 8. Results for Variants 1 and 2 (with substitutions shown with reference to SEQ ID NO:14)			
SF Validation Substrate(s)	**Backbone**	**Variant 1** (Y52F;K95P;G115R; T176V;L216Y;D232H; E349T;M491A;T577A; E619D)	**Variant 2** (Y52F;G115R;T176V; L216Y;D232H;P347T; E349T;T577A;D722S)
 Diclofenac	[Diclofenac]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

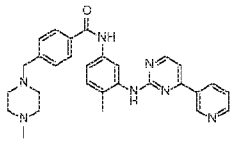
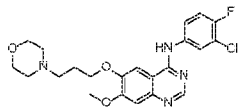
Table 8. Results for Variants 1 and 2 (with substitutions shown with reference to SEQ ID NO:14)			
SF Validation Substrate(s)	**Backbone**	**Variant 1** (Y52F;K95P;G115R; T176V;L216Y;D232H; E349T;M491A;T577A; E619D)	**Variant 2** (Y52F;G115R;T176V; L216Y;D232H;P347T; E349T;T577A;D722S)
 Imatinib	[Imatinib]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

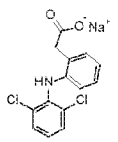
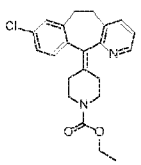
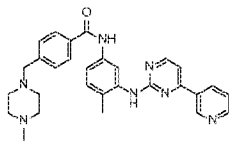
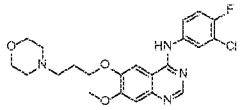
Table 9. Results for Variants 18 and 19 (with substitutions shown with reference to SEQ ID NO:16)			
SF Validation Substrate(s)	**Backbone**	**Variant 18** (K32R;C48S;Y52F; I95P;I176V;L216Y; D232H;E349K;M491A; N574T;E619D;D722S; A767T)	**Variant 19** (C48S;Y52F;I95P; Q111R;I176V;L216Y; D232H;P347T;M491A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

CLAIMS

WHAT IS CLAIMED IS:

1. 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.

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

3. The recombinant cytochrome P450-BM3 variant of Claim 2, wherein said organic substrate is selected from loratadine, imatinib, geftinib, and diclofenac.

4. A recombinant polynucleotide sequence encoding the recombinant cytochrome P450-BM3 variant of Claim 1.

5. The recombinant polynucleotide sequence of Claim 4, wherein said sequence comprises SEQ ID NO:1, 3, 5, 7, 9, 11, 13, or 15.

6. An expression vector comprising the polynucleotide sequence of Claim 4 or 5.

7. The vector of Claim 6, wherein said polynucleotide sequence is operably linked with regulatory sequences suitable for expression of said polynucleotide sequence in a suitable host cell.

8. The vector of Claim 6 and/or 7, wherein said host cell is a prokaryotic or eukaryotic cell.

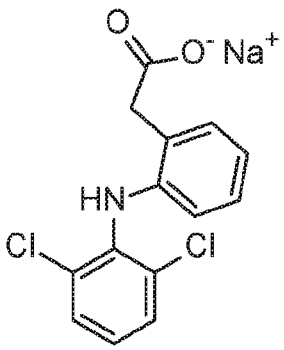
9. The vector of Claim 8, wherein said host cell is a prokaryotic cell.

10. The vector of Claim 8 and/or 9, wherein said host cell is *E. coli*.

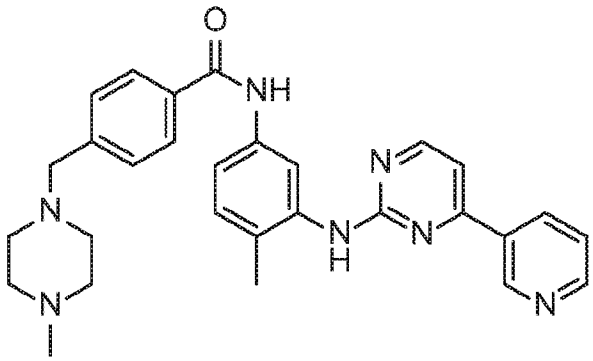
11. A host cell comprising the vector of Claim 6 and/or 7.

12. A method for producing at least one recombinant cytochrome P450-BM3 variant comprising culturing the host cell of Claim 11 under conditions such that the recombinant cytochrome P450-BM3 variant of any of Claims 1-3 is produced by said host cell.

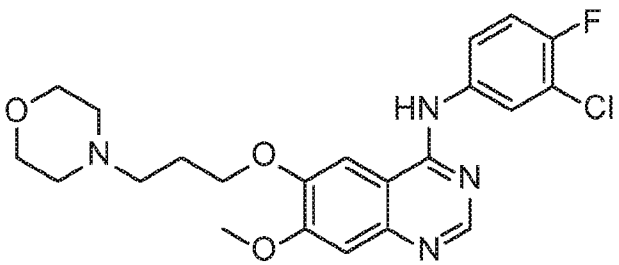
13. The method of Claim 12, further comprising the step of recovering said at least one recombinant cytochrome P450 variant.



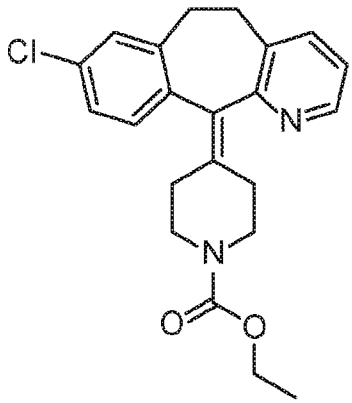
Diclofenac



Imatinib



Gefitinib



Loratadine

FIG. 1

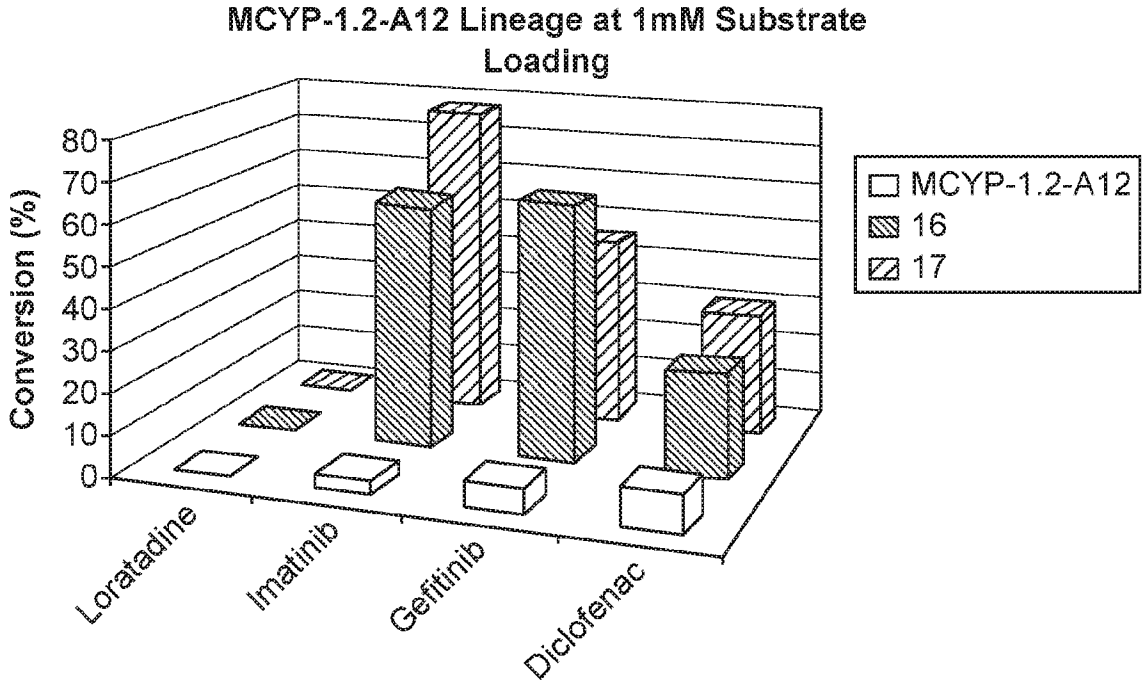

MCYP-1.2-A12 Lineage at 1mM Substrate Loading
Conversion (%)
80
70
60
50
40
30
20
10
0
Loratadine
Imatinib
Gefitinib
Diclofenac
MCYP-1.2-A12
16
17

FIG. 2

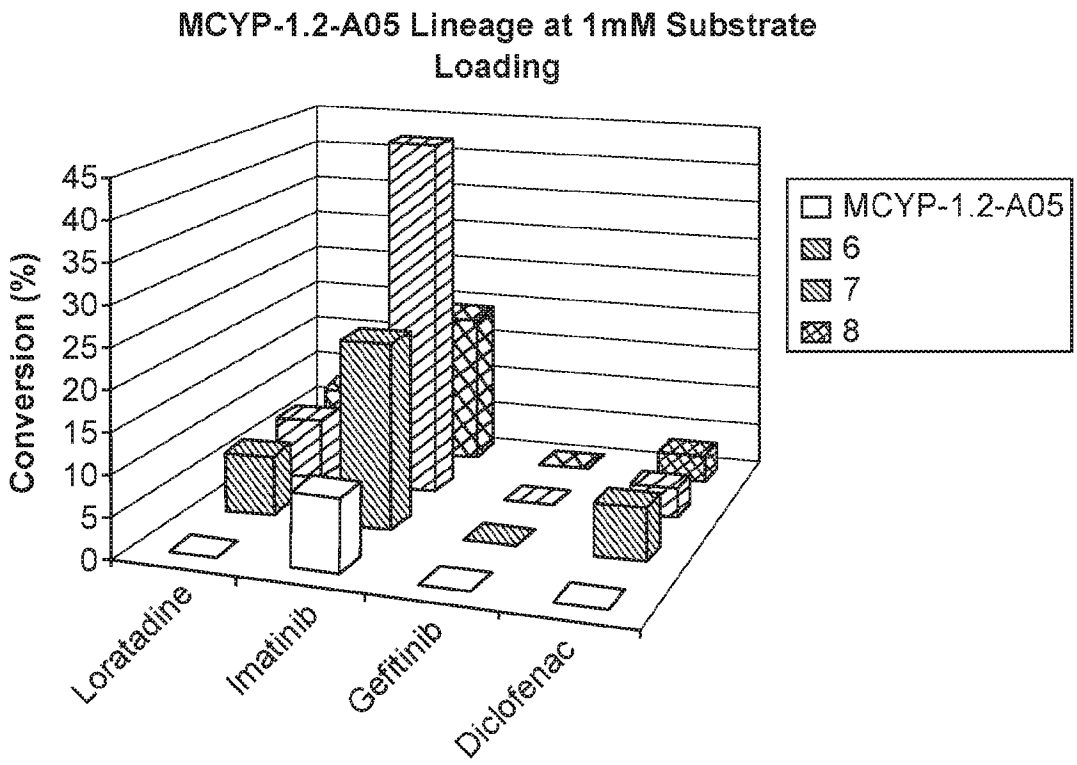

MCYP-1.2-A05 Lineage at 1mM Substrate Loading
Conversion (%)
45
40
35
30
25
20
15
10
5
0
Loratadine
Imatinib
Gefitinib
Diclofenac
MCYP-1.2-A05
6
7
8

FIG. 3

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US16/35399

A. CLASSIFICATION OF SUBJECT MATTER
IPC(8) - C07K 14/80; C12N 9/02 15/09, 5/10 (2016.01)
CPC - C07K 14/80; C12N 15/09, 5/10
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC(8): C07K 14/80; C12N 9/02 15/00, 15/09, 5/10 (2016.01)
CPC: C07K 14/80; C12N 15/00, 15/09, 5/10

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA, INPADOC Data); EBSCO Discovery; PubMed; Google; Google Scholar; Google Patents; The Lens; ENA; NCBI Blast; recombinant, cytochrome, P450-BM3, variant

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2008/0187983 A1 (DIETRICH, J et al.) Aug 7, 2008; paragraphs [0021], [0084], [0162], [0221]	1-4, 6/4, 7
A	US 2008/0044882 A1 (HILL, H et al.) February 21, 2008; paragraphs [0015], [0076]	1-4, 6/4, 7
A	US 2005/0059128 A1 (ARNOLD, F et al.) March 17, 2005; abstract	1-4, 6/4, 7

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
13 October 2016 (13.10.2016)	17 NOV 2016
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Form PCT/ISA/210 (second sheet) (January 2015)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US16/35399

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 8-13
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

-***-Please See Supplemental Page-***-

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

1-4, 6/4, 7, SEQ ID NO: 2

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.
PCT/US16/35399

-***-Continued from Box III-***-

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Groups I+: Claims 1-7 are directed toward a nucleic acid complex comprising a first single-stranded nucleic acid molecule and a method for determining pH in a sample comprising providing a nucleic acid complex comprising a first single-stranded nucleic acid molecule.

A recombinant cytochrome P450-BM3 variant having at least 90% sequence identity to a polypeptide sequence will be searched to the extent that it encompasses SEQ ID NO: 2 (P450 BM3 amino acid sequence). It is believed that Claims 1 (in-part), 2 (in-part), 3 (in-part), 4 (in-part), 6 (in-part) and 7 (in-part) encompass this first named invention and thus these claims will be searched without fee to the extent that they encompass 2 (P450 BM3 amino acid sequence). Applicant is invited to elect additional sequence(s), with specified SEQ ID NO(s): for each, to be searched. Additional sequence(s) will be searched upon the payment of additional fees. Applicants must specificy the claims that encompass any additionally elected sequence(s). Applicants must further indnicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined. An Exemplary Election would be: SEQ ID NO: 4 (P450 BM3 amino acid sequence).

Groups I+ share the technical features including a recombinant cytochrome P450-BM3 variant having at least 90% sequence identity to a polypeptide sequence comprising the sequence set forth in SEQ ID NO: 4; a recombinant polynucleotide sequence; and an expression vector.

However, these shared technical features are previously disclosed by US 2005/0059128 A1 to Arnold, et al. (hereinafter 'Arnold').

Arnold discloses a recombinant cytochrome P450-BM3 variant having at least 90% sequence identity to a polypeptide sequence comprising the sequence set forth in SEQ ID NO: 4 (P450-BM3 amino acid sequence) (SEQ ID NO: 40 of the Arnold reference is 99.9% identical to SEQ ID NO: 4 of the instant PCT application; Abstract; Example 5 of the Arnold reference, paragraph [0145]); a recombinant polynucleotide sequence (generating mutations in the parent polynucleotide sequence; paragraph [0070]); and an expression vector (inserting expression vectors; paragraph [0071]).

Since none of the special technical features of the Groups I+ inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Arnold reference, unity of invention is lacking.