discloses esomeprazole magnesium trihydrate prepared from the corresponding potassium salt, precipitated with acetone, and treated with water.

[0006] However, it is desirable to increase the efficiency of manufacture as well as reducing the number of processing steps for forming esomeprazole preparations of high enantiopurity. It is also desirable to identify processes that are applicable to preparation of other prazole compounds in addition to esomeprazole.

4. SUMMARY

[0007] The present disclosure provides polypeptides, polynucleotides encoding the polypeptides and methods of using the polypeptides for the biocatalytic sulfoxidation of prazole compounds, in particular the conversion of the sulfide substrate of 5-methoxy-2-((4-methoxy-3,5-dimethylpyridin-2-yl)methylthio)-1*H*-benzo[d]imidazole (also referred to herein as "compound **(1)**") to the *(R)*- and *(S)*-omeprazole product of 5-methoxy-2-((4-methoxy-3,5-dimethylpyridin-2-yl)methylsulfinyl)-1*H*-benzo[d]imidazole (also referred to herein as "compound **(2)**").

[0008] While naturally occurring cyclohexanone monooxygenase (CHMO) polypeptides do not efficiently convert compound **(1)** to compound **(2)**, the non-naturally occurring monooxygenase polypeptides (also referred to herein as "engineered CHMO polypeptides") of the present disclosure have been designed to efficiently carryout the conversion. Moreover, in some embodiments, the non-naturally occurring monooxygenase polypeptides of the disclosure can carry out the reaction stereoselectively to the form the *S*-isomer, *(S)*-5-methoxy-2-((4-methoxy-3,5-dimethylpyridin-2-yl)methylsulfinyl)-1*H*-benzo[d]imidazole (also referred to herein as "compound **(2b)**") in enantiomeric excess over the *R*-isomer, or in some embodiments, to form *(R)*-5-methoxy-2-((4-methoxy-3,5-dimethylpyridin-2-yl)methylsulfinyl)-1*H*-benzo[d]imidazole (also referred to herein as "compound **(2a)**") in enantiomeric excess over the *S*-isomer. In some embodiments, the monooxygenase polypeptides can produce compound **(2b)** as a substantially enantiomerically pure preparation.

[0008a] In one aspect, there is provided a non-naturally occurring monooxygenase comprising a polypeptide capable of converting 5-methoxy-2-((4-methoxy-3,5-dimethylpyridin-2-yl)methylthio)-1*H*-benzo[d]imidazole ("compound **(1)**") to 5-methoxy-2-((4-methoxy-3,5-dimethylpyridin-2-yl)methylsulfinyl)-1*H*-benzo[d]imidazole ("compound **(2)**"), and which polypeptide amino acid sequence:

(a) is at least 80% identical to SEQ ID NO: 124; and

(b) has two or more residue differences as compared to SEQ ID NO:2 at the following residue positions: position corresponding to X246; and position corresponding to position corresponding to X248.

[0009] In some embodiments, the polypeptide is capable of converting compound **(1)** to compound **(2a)** in enantiomeric excess. In some embodiments, the *R*-enantioselective monooxygenase polypeptide comprises an amino acid sequence that is at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identical to a reference sequence selected from the group consisting of SEQ ID NO: 4, 6, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, and 208. In particular, the reference sequence is SEQ ID NO: 6, 166, 170, 174, 190, 192, 196, 200, 204, or 206.

[0010] In some embodiments, the polypeptide is capable of converting compound **(1)** to compound **(2b)** in enantiomeric excess. In some embodiments, the *S*-enantioselective monooxygenase polypeptide comprises an amino acid sequence that is at least 80%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more identical to a reference sequence selected

at position corresponding to X516 is V; residue at position corresponding to X526 is V; residue at position corresponding to X537 is T; and residue at position corresponding to X540 is Q.

[0031] In addition to the residue positions specified above, various other residue differences relative to SEQ ID NO:2 can be present at other residue positions in the non-naturally occurring monooxygenase polypeptides disclosed herein. These can be conservative or non-conservative differences, including conservative substitutions and non-conservative substitutions. Guidance on these other residue positions and the choices of amino acid residues at these other specified positions is provided in the detailed description.

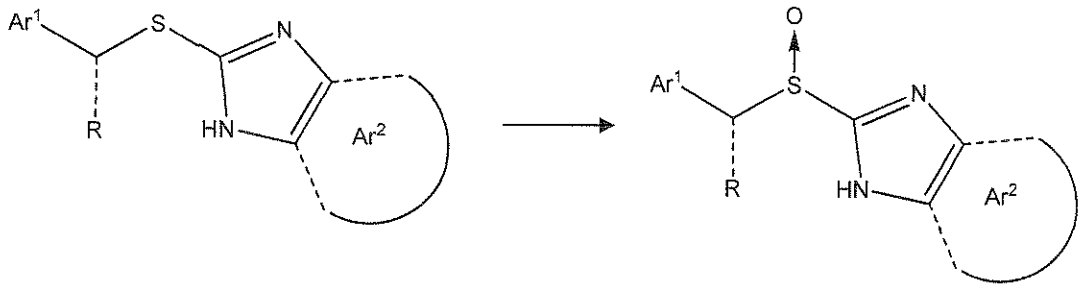
[0032] Also~~In another aspect,~~ provided herein are polynucleotides encoding the monooxygenase polypeptides, expression vectors comprising the polynucleotides, and host cells capable of expressing the polypeptides.

[0032a] In one aspect, there is provided a polynucleotide encoding the polypeptide as defined above.

[0032b] In one aspect, there is provided an expression vector comprising the polynucleotide as defined above.

[0032c] In one aspect, there is provided a host cell comprising a polynucleotide as defined above or an expression vector as defined above.

[0033] Further~~In a further aspect,~~ provided herein are processes of using the polypeptides of the disclosure for converting a substrate compound of structural formula **(I)** to the product compound of formula **(II)**:



(I) **(II)**

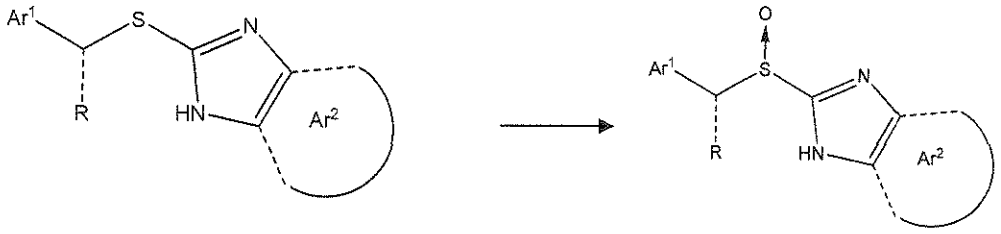
wherein,

Ar^1 is an optionally substituted aryl or heteroaryl ring;

R is H, a lower alkyl, a heteroalkyl, or forms a 5 to 8 membered cycloalkyl, heteroalkyl, aryl or heteroaryl fused ring with a ring carbon of Ar^1; and

Ar^2 is an optionally substituted cycloalkyl, heterocycloalkyl, aryl, or heteroaryl ring fused to the imidazole ring.

[0033a] In one aspect, there is provided a process for converting a substrate compound of structural formula **(I)** to the product compound of formula **(II)**:



(I) **(II)**

wherein,

Ar^1 is an optionally substituted aryl or heteroaryl ring;

R is H, a lower alkyl, a heteroalkyl, or forms a 5 to 8 membered cycloalkyl, heteroalkyl, aryl or heteroaryl fused ring with a ring carbon of Ar^1; and

Ar^2 is an optionally substituted cycloalkyl, heterocycloalkyl, aryl, or heteroaryl ring fused to the imidazole ring; the process comprising:

contacting the compound of formula **(I)** with a polypeptide as defined above in presence of a cofactor under suitable reaction conditions for converting the substrate to the product compound of formula **(II).**

[0034] Generally, the process comprises contacting the compound of formula **(I)** with an engineered monooxygenase polypeptide described herein in presence of a cofactor under suitable reaction conditions for converting the substrate to the product compound of formula **(II)**. The suitable reaction conditions typically include a source of molecular oxygen O_2, and the cofactor is NADPH or NADH.

[0035] processes of using the polypeptides of the disclosure can be used in the preparation of omeprazole analog compounds of structural formula **(II)** in enantiomeric excess, wherein the compounds of structural formula **(II)** are selected from: (*R*) or (*S*)-lansoprazole, (*R*) or (*S*)-tenatoprazole, (*R*) or (*S*)-rabeprazole, (*R*) or (*S*)-pantoprazole, (*R*) or (*S*)-ilaprazole, (*R*) or (*S*)-leminoprazole, (*R*) or (*S*)-saviprazole, and (*R*) or (*S*)-TY-11345.

70, 72, 74, 76, 78, 80, 82, 84, 86, 88, 90, 92, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 242, 244, 248, 250, 254, 256, 258, 262, and 264.

[0042] In some embodiments, the process comprises contacting compound **(1)** with a polypeptide of the disclosure in presence of a cofactor under suitable reaction conditions to form compound **(2b)** in at least 99% enantiomeric excess. Exemplary polypeptides for preparing compound **(2b)** in at least 99% enantiomeric excess can comprise an amino acid sequence selected from the group consisting of SEQ ID NO: 16, 18, 20, 22, 24, 26, 28, 30, 32, 42, 44, 86, 94, 96, 98, 100, 102, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, and 158.

[0043] In some embodiments, the reaction condition in the process comprises a temperature of 10°C to 50°C, and in particular 25°C to 40°C.

[0044] In some embodiments, the reaction condition in the process comprises a pH of about 8.5 to a pH of about 10, in particular a pH of about 8.5 to about 9.0.

[0045] In some embodiments, the reaction condition in the process comprises a partial pressure of O_2 at greater than atmospheric pressure. Additionally, dissolved molecular oxygen in the process can be increased by sparging the reaction solution with O_2-containing gas or by use of bubble-free aeration with O_2-containing gas.

[0046] In some embodiments of the process, the reaction condition comprises a co-solvent, such as for example, MeOH, EtOH, isopropanol (IPA), acetone, toluene, MeCN, methyl tert-butyl ether (MTBE), N-methyl-2-pyrrolidone (NMP), dimethylacetamide (DMAc), dimethylformamide (DMF), propylene glycol, polyethylene glycol (PEG), tetramethylurea, N-ethylpyrollidinone, tetraglyme, 1,3-Dimethyl-3,4,5,6-tetrahydro-2(1*H*)-pyrimidinone (DMPU), DMIU, hexamethylphosphoramide (HMPA) and dimethylsulfoxide (DMSO). The co-solvent can reduce the formation of aggregates which can affect the rate and scalability of the process.

[0047] In some embodiments, the process further comprises converting $NADP^+$ or NAD+ formed from the NADPH or NADH, respectively, with a cofactor regenerating system. The cofactor regenerating system can use an appropriate dehydrogenase, such a glucose dehydrogenase, glucose-phosphate dehydrogenase, formate dehydrogenase, phosphite dehydrogenase, and ketoreductase/alcohol dehydrogenase and corresponding substrate, for example, glucose, glucose-6-phosphate, formate, phosphite, or alcohol, respectively. In some embodiments, the co-factor regenerating system is a ketoreductase and a secondary alcohol, particularly isopropanol (IPA).

[0048] In a further embodimentaspect, the engineered monooxygenase polypeptides can be provided in the form of kits or arrays, particularly for use in the screening for activity on other prazole compounds of interest.

5. DETAILED DESCRIPTION

residue at position corresponding to X209 is P;
residue at position corresponding to X240 is F, or K;
residue at position corresponding to X246 is E, H, Q, S, R, W, or Y;
residue at position corresponding to X248 is C, I, N, V, or S;
residue at position corresponding to X277 is L, M, or Q;
residue at position corresponding to X278 is G, N, or S;
residue at position corresponding to X280 is G, or R;
residue at position corresponding to X281 is A, or S;
residue at position corresponding to X282 is S;
residue at position corresponding to X288 is I;
residue at position corresponding to X307 is C or R;
residue at position corresponding to X341 is E;
residue at position corresponding to X373 is L;
residue at position corresponding to X377 is G;
residue at position corresponding to X388 is K;
residue at position corresponding to X390 is R or I;
residue at position corresponding to X400 is I;
residue at position corresponding to X415 is A;
residue at position corresponding to X426 is F;
residue at position corresponding to X449 is M, F or L;
residue at position corresponding to X464 is D;
residue at position corresponding to X481 is K;
residue at position corresponding to X488 is F, K, or L;
residue at position corresponding to X489 is a C;
residue at position corresponding to X499 is L;
residue at position corresponding to X505 is W or L;
residue at position corresponding to X516 is V;
residue at position corresponding to X526 is V;
residue at position corresponding to X537 is T; and
residue at position corresponding to X540 is Q, or A.

11. The non-naturally occurring monooxygenase of claim 3 in which the amino acid sequence comprises a sequence selected from the group consisting of SEQ ID NO: 92, 94, 96, 98, 100, 104, 106, 108, 110, 112, 114, 116, 118, 120, 122, 124, 126, 128, 130, 132, 134, 136, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156 and 158.

39. The process of claim 22 in which the compound of formula **(I)** is 2-((3-methyl-4-(2,2,2-trifluoroethoxy)pyridin-2-yl)methylthio)-1H-benzo[d]imidazole and the compound of formula **(II)** is (*R*)-lansoprazole which is produced in enantiomeric excess.

40. The process of claim 39 in which the polypeptide comprises an amino acid sequence having the set of amino acid differences relative to SEQ ID NO: 2 that are present in SEQ ID NO: 6.

41. The process of any one of claims 22 to 40 in which the reaction condition comprises a temperature of 25°C to 40°C.

42. The process of any one of claims 22 to 40 in which the reaction condition comprises a pH of about 8.5 to a pH of about 10.

43. The process of any one of claims 22 to 40 in which the reaction condition comprises a partial pressure of O_2 at greater than atmospheric pressure.

44. The process of any one of claims 22 to 40 in which the reaction condition comprises phosphate at a concentration of about 5 mM to about 50 mM.

45. The process of any one of claims 22 to 40 in which the cofactor is NADPH or NADH, and the process further comprises converting NAD+ or NADP+ formed form the NADH or NADPH, respectively, to NADH or NADPH with a cofactor regenerating system.

46. The process of claim 45 in which the cofactor regenerating system is selected from a ketoreductase and an alcohol, or glucose dehydrogenase and glucose.

47. The process of claim 46 in which the cofactor regenerating system is a ketoreductase and alcohol, wherein the alcohol is a secondary alcohol.

48. The process of claim 47 in which the secondary alcohol is isopropanol.

49. The process of any one of claims 22 to 48 which is carried out in the presence of a co-solvent.