been engineered to have improved properties relative to the naturally occurring cis-4-proline hydroxylase (SEQ ID NO:2) of *Sinorhizobium meliloti*, a nitrogen fixing Gram negative bacterium. The improved biocatalyst properties of the engineered proline hydroxylases include, among others, activity, regioselectivity, substrate tolerance, and stability. The engineered proline hydroxylases are also found to hydroxylate a variety of substrate compounds, including the selective conversion of (2*S*)-piperidine-2-carboyxlic acid (*i.e.*, L-pipecolic acid) to (2*S*,5*S*)-5-hydroxylpiperidine-2-carboxylic acid (*i.e.*, cis-5-hydroxypipecolic acid). The engineered enzymes with improved properties have one or more residue differences as compared to the naturally occurring proline hydroxylase, where the residue differences occur at residue positions affecting the foregoing enzyme properties.

[0008a] In one aspect, there is provided an engineered polypeptide having proline hydroxylase activity, comprising an amino acid sequence having at least 80% sequence identity to SEQ ID NO:2 and one or more residue differences as compared to the sequence of SEQ ID NO:2 at residue positions selected from: X166; X2; X3; X4; X5; X9; X13; X25; X26; X29; X30; X36; X42; X52; X57; X58; X59; X66; X86; X92; X95; X103; X112; X113; X115; X116; X121; X131; X150; X151; X225; X230; X270; and X271.

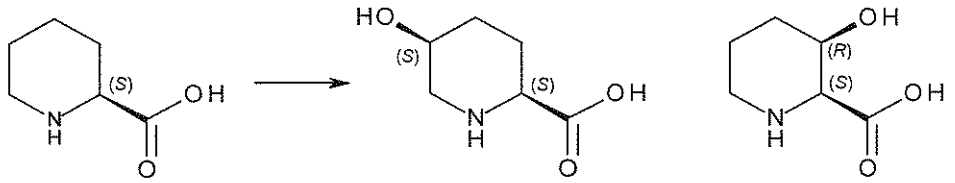
[0009] ~~Accordingly, i~~In one embodiment~~aspect~~, the present disclosure provides engineered polypeptides having proline hydroxylase activity, where the polypeptides comprises an amino acid sequence having at least 80% identity to SEQ ID NO:2 and one or more residue differences as compared to SEQ ID NO:2 at residue positions selected from: X2; X3; X4; X5; X9; X13; X17; X24; X25; X26; X29; X30; X36; X42; X52; X57; X58; X59; X62; X66; X86; X88; X92; X95; X98; X103; X112; X113; X114; X115; X116; X121; X131; X140; X150; X151; X166; X186; X188; X205; X225; X230; X270; and X271.

[0010] In some embodiments, the residue differences at residue positions X2; X3; X4; X5; X9; X13; X17; X24; X25; X26; X29; X30; X36; X42; X52; X57; X58; X59; X62; X66; X86; X88; X92; X95; X98; X103; X112; X113; X114; X115; X116; X121; X131; X140; X150; X151; X166; X186; X188; X205; X225; X230; X270; and X271 can be selected from X2K; X2T; X3S; X4Q; X4L; X4E; X4S; X5I; X5L; X5M; X9I; X13T; X17V, X24R; X24S; X25R; X26R; X26T; X26W; X29A; X30V; X30P; X36T; X42E; X52P; X57T; X57A; X58A; X59G; X62Q; X66Q; X86S; X88R; X92V; X95M; X98F; X98T; X103L; X103Q; X112T; X112V; X113E; X114N; X115E; X115H; X115D; X115G; X115S; X115A; X116L; X121F; X131Y; X131F; X140L; X150S; X151A; X151H; X151S; X166T; X166L; X166Q; X186G; X188G; X205V; X225L; X225Y; X225W; X230V; X270E; X271K; and X271R. The following detailed description provides guidance on the choices of the residue differences that can be used to prepare engineered proline hydroxylases with the desired improved biocatalytic properties.

[0011] In some embodiments, the engineered proline hydroxylase polypeptide has an amino acid sequence comprising at least a combination of features selected from: (a) X103L and X166Q; (b) X52P and X255Y; (c) X4E/L/S and X115A; (d) X25R and X58A; (e) X29A and X166T/Q/L; (f) X115H/D/G and X121F; (g) X3S, X103L, and X166Q; (h) X103L, X131Y/F, and X166T/Q/L; (i) X26T, X103L and X166T/Q/L; (j) X25R, X66Q, X92V and X115E; (k) X25R, X66Q, X92V, X103L, X115E, and X166Q; and (l) X3S, X25R, X66Q, X92V, X103L, X115E, and X166Q.

[0012] As noted above, the engineered polypeptides having proline hydroxylase activity, are also capable of converting substrate compound **(2)**, (2*S*)-piperidine-2-carboxylic acid to product compound **(1)**, (2*S*,5*S*)-5-hydroxylpiperidine, as shown in **Scheme 1**,

Scheme 1



(2) (1) (1a)

with improved properties as compared to the naturally occurring enzyme. In some embodiments, the engineered polypeptides are capable of converting substrate compound **(2)** to product compound **(1)** with at least 1.2 fold, 1.5 fold, 2 fold, 3 fold, 4 fold, 5 fold, or 10 fold or more the activity of the naturally occurring enzyme, and with greater than 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99 % or more diastereomeric excess of (2*S*,5*R*)-5-hydroxypiperidine-2-carboxylic acid.

[0013] In some embodiments, the engineered polypeptides having proline hydroxylase activity also display increased regioselectivity for forming (2*S*,5*S*)-5-hydroxypiperidine-2-carboxylic acid in excess of other regioisomers, for example (2*S*,3*R*)-3-hydroxypiperidine-2-carboxylic acid, shown in **Scheme 1** as compound **(1a)**. Thus, in some embodiments, the engineered proline hydroxylases are capable of converting substrate compound **(2)** to product compound **(1)** in excess of product compound **(1a)**, where the ratio of product compound **(1)** formed over product compound **(1a)** is at least 1.5, 2, 3, 4, 5 or 6 or more. In some embodiments, the engineered proline hydroxylases with increased selectivity for forming product compound **(1)** in excess of product compound **(1a)** comprises an amino acid sequence having one or more of the following features: X103L; X115E; X131Y and X166Q, particularly the combination of features X103L and X166Q.

[0014] In some embodiments, the engineered polypeptides having improved properties has an amino acid sequence comprising a sequence selected from the group consisting of SEQ ID NO: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 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, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, and 228.

[0014a] In one aspect, there is provided a polynucleotide encoding the engineered polypeptide as described herein, the polynucleotide optionally comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO: 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, and 227.

[0015] In another embodiment~~aspect~~, the present disclosure provides polynucleotides encoding the engineered proline

hydroxylase polypeptides with improved properties. Exemplary polynucleotide sequences are provided in the Sequence Listing incorporated by reference herein and include SEQ ID NO: 7, 9, 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 33, 35, 37, 39, 41, 43, 45, 47, 49, 51, 53, 55, 57, 59, 61, 63, 65, 67, 69, 71, 73, 75, 77, 79, 81, 83, 85, 87, 89, 91, 93, 95, 97, 99, 101, 103, 105, 107, 109, 111, 113, 115, 117, 119, 121, 123, 125, 127, 129, 131, 133, 137, 139, 141, 143, 145, 147, 149, 151, 153, 155, 157, 159, 161, 163, 165, 167, 169, 171, 173, 175, 177, 179, 181, 183, 185, 187, 189, 191, 193, 195, 197, 199, 201, 203, 205, 207, 209, 211, 213, 215, 217, 219, 221, 223, 225, and 227.

[0015a] In one aspect, there is provided a polynucleotide encoding the polypeptide of SEQ ID NO: 2, comprising a nucleic acid sequence optimized for expression in *E. coli*, optionally comprising a sequence having at least 80% or more identity to a nucleic acid sequence selected from SEQ ID NO: 1, 3, and 5.

[0016] In a further embodiment~~aspect~~, the present disclosure also provides codon optimized polynucleotides encoding a polypeptide that comprises an amino acid sequence of the wild-type proline hydroxylase (SEQ

ID NO:2). In some embodiments, the codon optimized polynucleotides have increased expression in a bacterial host cell as compared to the naturally occurring polynucleotide encoding the proline hydroxylase enzyme. In some embodiments, the codon optimized polynucleotides can have at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identity to the codon optimized nucleic acid sequence of SEQ ID NO: 1, 3, or 5, which encodes the identical polypeptide sequences of SEQ ID NO:2, 4, or 6, respectively. The codon optimized sequence of SEQ ID NO:1, 3, or 5 can enhance expression of the encoded, wild-type proline hydroxylase by at least 1.2 fold, 1.5 fold or 2 fold or greater as compared to the naturally occurring polynucleotide sequence.

[0016a] In one aspect, there is provided an expression vector comprising the polynucleotide as described herein, optionally further comprising a control sequence.

[0016b] In one aspect, there is provided a host cell comprising the polynucleotide as described herein, optionally wherein the host cell is *E. coli*.

[0016c] In one aspect, there is provided a method of preparing the engineered polypeptide as described herein, comprising culturing the host cell as described herein under conditions suitable for expression of the polypeptide, optionally the method further comprising a step of isolating the engineered polypeptide.

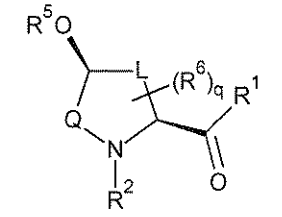
[0017] In another embodiment~~aspect~~, the polynucleotides of the disclosure can be incorporated into expression vectors and host cells for expression of the polynucleotides and the corresponding encoded proline hydroxylase polypeptides. As such, in some embodiments, the present disclosure provides methods of preparing the proline hydroxylase polypeptides by culturing a host cell comprising the polynucleotide or expression vector capable of expressing a proline hydroxylase of the disclosure under conditions suitable for expression of the engineered polypeptide. In some embodiments, the method of preparing the proline hydroxylase can comprise the additional step of isolating the expressed polypeptide.

[0018] In some embodiments, the present disclosure also provides methods of manufacturing an engineered proline hydroxylase polypeptide, where the method can comprise: (a) synthesizing a polynucleotide encoding a polypeptide comprising an amino acid sequence selected from SEQ ID NO: 8, 10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 32, 34, 36, 38, 40, 42, 44, 46, 48, 50, 52, 54, 56, 58, 60, 62, 64, 66, 68, 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, 138, 140, 142, 144, 146, 148, 150, 152, 154, 156, 158, 160, 162, 164, 166, 168, 170, 172, 174, 176, 178, 180, 182, 184, 186, 188, 190, 192, 194, 196, 198, 200, 202, 204, 206, 208, 210, 212, 214, 216, 218, 220, 222, 224, 226, and 228, and having one or more residue differences as compared to SEQ ID NO:2 at residue positions selected from: X2; X3; X4; X5; X9; X13; X17; X24; X25; X26; X29; X30; X36; X42; X52; X57; X58; X59; X62; X66; X86; X88; X92; X95; X98; X103; X112; X113; X114; X115; X116; X121; X131; X140; X150; X151; X166; X186; X188; X205; X225; X230; X270; and X271, and (b) expressing the proline hydroxylase polypeptide encoded by the polynucleotide. As noted above, the residue differences at positions X2; X3; X4; X5; X9; X13; X17; X24; X25; X26; X29; X30; X36; X42; X52; X57; X58; X59; X62; X66; X86; X88; X92; X95; X98; X103; X112; X113; X114; X115; X116; X121; X131; X140; X150; X151; X166; X186; X188; X205; X225;

X230; X270; and X271 are selected from X2K; X2T; X3S; X4Q; X4L; X4E; X4S; X5I; X5L; X5M; X9I; X13T; X17V, X24R; X24S; X25R; X26R; X26T; X26W; X29A; X30V; X30P; X36T; X42E; X52P; X57T; X57A; X58A; X59G; X62Q; X66Q; X86S; X88R; X92V; X95M; X98F; X98T; X103L; X103Q; X112T; X112V; X113E; X114N; X115E; X115H; X115D; X115G; X115S; X115A; X116L; X121F; X131Y; X131F; X140L; X150S; X151A; X151H; X151S; X166T; X166L; X166Q; X186G; X188G; X205V; X225L; X225Y; X225W; X230V; X270E; X271K; and X271R. As further provided in the detailed description, additional variations can be incorporated during the synthesis of the polynucleotide

to prepare engineered proline hydroxylases with corresponding differences in the expressed amino acid sequences.

[0018a] In one aspect, there is provided a process for preparing a product compound of formula **(Ia)**,



(Ia)

wherein

Q is selected from the group consisting of a $(C_1$-$C_5)$alkylene and $(C_2$-$C_5)$alkenylene;

L is selected from the group consisting of a bond, $(C_1$-$C_4)$alkylene and $(C_2$-$C_4)$alkenylene;

R^1 is selected from the group consisting of hydroxy, amino, $(C_1$-$C_6)$alkyloxy, aryloxy, $(C_1$-$C_6)$alkylthio and arylthio;

R^2 is selected from the group consisting of hydrogen and optionally substituted $(C_1$-$C_6)$alkyl, $(C_2$-$C_6)$alkenyl, and $(C_2$-$C_6)$alkynyl;

R^5 is hydrogen, or a direct bond to a carbon atom of L to form an epoxide;

each occurrence of R^6 is selected from the group consisting of halo, $(C_1$-$C_6)$alkyl, and $(C_1$-$C_6)$alkyloxy; and

q is an integer from 0 to 4;

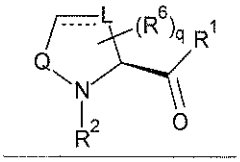
wherein the sum of ring carbon atoms for Q + L is an integer from 2 to 5;

with the provisos that

(i) when the sum of ring carbon atoms for Q + L is 2, then L is a methylene; and

(ii) when the sum of ring carbon atoms for Q + L is 3, then L is either a bond or ethylene;

the process comprising contacting substrate compound of formula **(IIa)**,



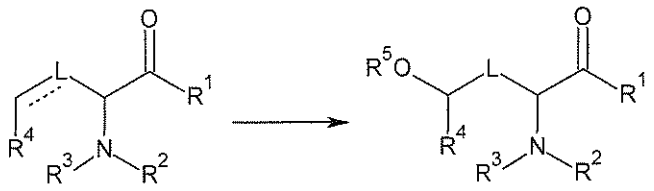
(IIa)

wherein L, Q, R^1, R^2, R^6, and q are as defined above for the compound of formula **(Ia)**; and

····· represents an optional bond to a carbon atom of L to form a double bond;

with an engineered polypeptide as described herein in presence of a co-substrate under suitable reaction conditions.

[0019] In another ~~aspect~~embodiment, the engineered proline hydroxylase polypeptides can be used in a process for preparing various hydroxylated compounds, such as hydroxylated proline or hydroxylated piperidine-2-carboxylic acids. Accordingly, in some embodiments, the engineered proline hydroxylase polypeptides can be used in a process for the conversion of a substrate compound of formula **(II)** to product compound of formula **(I)**, as shown below:



(II) (I)

wherein

L is selected from the group consisting of a bond, $(C_1\text{-}C_4)$alkylene and $(C_2\text{-}C_4)$alkenylene;

R^1 is selected from the group consisting of hydroxy, amino, $(C_1\text{-}C_6)$alkyloxy, aryloxy, $(C_1\text{-}C_6)$alkylthio and arylthio;

R^2 and R^3 are each independently selected from the group consisting of hydrogen and optionally substituted $(C_1\text{-}C_6)$alkyl, $(C_2\text{-}C_6)$alkenyl, and $(C_2\text{-}C_6)$alkynyl;

R^4 is selected from the group consisting of optionally substituted $(C_1\text{-}C_6)$alkyl, $(C_2\text{-}C_6)$alkenyl, $(C_2\text{-}C_6)$alkynyl, aryl, heteroaryl, cycloalkyl, heterocycloalkyl; or R^4 together with one of R^1 or R^2 is a $(C_1\text{-}C_5)$alkylene or $(C_2\text{-}C_5)$alkenylene and forms a 5- to 8-membered heterocyclic ring containing the nitrogen atom, wherein the ring is optionally substituted with 1 to 4 independently selected R^6 groups;

R^5 is hydrogen or a bond that forms an epoxide with a carbon atom of L;

each occurrence of R^6 is independently selected from the group consisting of halo, $(C_1\text{-}C_6)$alkyl, and $(C_1\text{-}C_6)$alkyloxy; and

····· represents an optional bond to a carbon atom of L to form a double bond;

with the provisos that

(i) when R^4 does not form a ring with one R^2 or R^3, or when R^4 forms a 5-membered heterocyclic ring containing the nitrogen atom with one of R^2 or R^3, then L is a methylene;

(ii) when R^4 forms a 6-membered heterocyclic ring containing the nitrogen atom with one of R^2 or R^3, then L is a bond or ethylene; and

(iii) when R^5 is a bond to a carbon atom of L to form an epoxide, then R^4 forms the heterocyclic ring containing the nitrogen atom with one of R^2 or R^3 and L is a $(C_1\text{-}C_4)$alkylene or $(C_2\text{-}C_4)$alkenylene.

[0020] In some embodiments, the engineered proline hydroxylase can be used for the conversion of a ring compound of structural formula **(IIa)** to the hydroxylated product compound of structural formula **(Ia)**;